ACADEMIA

Москва
Издательский центр «Академия»
2013

Данный учебник является частью учебно-методического комплекта по специальностям укрупненной группы «Экономика и управление».

Учебник предназначен для изучения общепрофессиональной дисциплины «Статистика».

Учебно-методические комплекты нового поколения включают в себя традиционные и инновационные учебные материалы, позволяющие обеспечить изучение общеобразовательных и общепрофессиональных дисциплин и профессиональных модулей. Каждый комплект содержит учебники и учебные пособия, средства обучения и контроля, необходимые для освоения общих и профессиональных компетенций, в том числе и с учетом требований работодателя.

Учебные издания дополняются электронными образовательными ресурсами. Электронные ресурсы содержат теоретические и практические модули с интерактивными упражнениями и тренажерами, мультимедийные объекты, ссылки на дополнительные материалы и ресурсы в Интернете. В них включен терминологический словарь и электронный журнал, в котором фиксируются основные параметры учебного процесса: время работы, результат выполнения контрольных и практических заданий. Электронные ресурсы легко встраиваются в учебный процесс и могут быть адаптированы к различным учебным программам.

Учебно-методический комплект разработан на основании Федерального государственного образовательного стандарта среднего профессионального образования с учетом его профиля.

Рыночная экономика существенно повышает требования к качеству подготовки экономистов. Чтобы быть конкурентоспособным на рынке труда, сегодня необходимо владеть современным статистическим инструментарием анализа экономической информации. Предлагаемый учебник знакомит с рядом важнейших разделов статистики, формирует основы статистического мышления.

Что же такое статистика? За последние два столетия это понятие претерпело значительные изменения. Семантически «статистика» имеет один корень со словом «государство» (*state*). Первоначальный его смысл: «искусство и наука управления». Преподаватели статистики университетов Германии XVIII в. сегодня назывались бы специалистами по общественным (политическим) наукам. Понятно, что решения правительств во многом основывались на данных о населении, промышленности (ремеслах), сельском хозяйстве и т. д. Поэтому обществоведы-статистики, поставлявшие информацию своим сюзеренам, курфюрстам, правительствам, стали интересоваться такими сведениями, и постепенно слово «статистика» стало означать сбор данных о государстве, экономике, населении, а затем вообще сбор и обработку всяких цифровых данных. Подобное понимание задач статистики и до сих пор во многом сохраняет свою силу. Однако такая характеристика относится лишь к той части предмета, которая именуется «описательной статистикой».

Ныне наблюдается дальнейшее усложнение понятия «статистика». Современный представитель этой науки наряду со сбором, обработкой и анализом данных все больше внимания уделяет вопросам интерпретации получаемых результатов. Начало такому подходу было положено Гауссом, применившим статистические приемы для вычисления планетных орбит по несовершенным данным. В настоящее время можно считать, что задачей статистики стала обработка количественных результатов научных экспериментов.

Статистика довольно трудно вошла в духовную культуру человечества, поскольку оно сравнительно недавно стало размышлять над понятиями «случай и случайность». Буквально на наших глазах статистические методы стали применяться в химии и геологии, лингвистике и истории, социологии и психологии, экономике и промышленности, при обработке результатов летных испытаний. В нашем учебнике основной упор сделан на применении статистических методов анализа и обобщения в экономике и бизнесе.

Независимо от уровня и стадии экономического развития, характера политической системы статистика на протяжении сотен лет своего существования всегда оставалась 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, она всегда играла роль главного поставщика фактов для управленческих, научно-исследовательских организаций и различного рода иных структур, а также для населения.

Роль статистики в нашей жизни настолько значительна, что люди, часто не задумываясь и не осознавая, постоянно используют элементы статистической методологии не только в трудовых процессах, но и в повседневном быту. Работая и отдыхая, делая покупки, знакомясь с другими людьми, принимая какие-то решения, человек пользуется определенной системой имеющихся у него сведений, сложившихся вкусов и привычек, фактов, систематизирует, сопоставляет эти факты, анализирует их, делает необходимые для себя выводы и принимает определенные решения, предпринимает конкретные действия. Таким образом, в каждом человеке заложены элементы статистического мышления, представляющего собой способности к анализу и синтезу информации об окружающем нас мире. Это так называемая обыденная компонента статистического мышления.

Цель настоящего учебника — развить статистическое мышление. Речь идет о постижении множества специальных правил, методов и приемов количественного анализа разного рода информации. Что же необходимо для этого? Не очень много и не так уж мало. Нужны серьезность и вдумчивость, неплохое знание математики, истории, экономической теории, основ предпринимательства и информатики. Чем шире кругозор и эрудиция в самых различных областях знаний, тем заметнее будут ваши успехи в изучении статистики.

Основными объектами приложения статистической теории и методологии в учебнике выступают хозяйственная деятельность, народонаселение, условия жизни людей, управление экономическими и общественными процессами. Учебник призван дать знание теории и методологии статистики экономистам, финансистам, менеджерам, коммерсантам и бухгалтерам, демографам и социологам, а также лицам иных профессиональных интересов, самостоятельно изучающим предмет. Настоящий учебник рассчитан на развитие исследовательских и предпринимательских навыков у указанных кругов населения в условиях современной российской действительности.

Основной задачей курса статистики является овладение знаниями общих основ статистической науки, искусством организации и проведения статистических исследований, анализа и обобщения их результатов, навыками прогнозирования. Результатом изучения курса статистики должно быть *знание* принципов современной организации национальных и зарубежных статистических служб, категорий и понятий статистики, методов организации сбора, обработки данных (материалов) статистического наблюдения, их анализа с помощью обобщающих показателей, методов статистического моделирования и прогнозирования. Учащиеся, изучившие курс, *обязаны уметь* организовать сплошное и несплошное наблюдения, строить статистические графики и таблицы, анализировать массивы статистических данных, исчислять и интерпретировать статистические показатели, формулировать выводы, вытекающие из проведенного анализа.

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

Описательным методам статистики посвящены гл. 1—5. В них рассматриваются методы сбора статистической информации, вопросы методологии и практики осуществления статистической сводки и группировки, построения статистических таблиц и графиков, расчета абсолютных, относительных и средних величин, их использования в анализе социально-экономических явлений.

Аналитические методы статистики рассматриваются в гл. 6—10. Они включают методы анализа вариации, методологию индексного анализа, вопросы теории и практики выборочного наблюдения, оценки взаимосвязей признаков, а также методы статистического анализа рядов динамики и прогнозирования. В гл. 8 «Выборочные наблюдения» приводятся некоторые сведения из теории вероятности, которые непосредственно не входят в курс статистики, но необходимы для понимания выборочного метода.

Учебник подготовлен авторским коллективом Московского государственного университета экономики, статистики и информатики под руководством доктора экономических наук профессора Национального исследовательского университета «Высшая школа экономики» В. С. Мхитаряна.

В романе И. Ильфа и Е. Петрова «Двенадцать стульев» сказано: «Статистика знает все». Слегка перефразировав этот афоризм, можно сказать: «О статистике слышали все». Например, сообщения Федеральной службы государственной статистики о выполнении федеральных программ развития национального хозяйства, сведения о количестве родившихся и умерших, данные о браках и разводах и т. д. К рубрике в газетах «Статистика» мы так привыкли, что часто и не задумываемся, а знаем ли мы, собственно, что это за наука, как не задумываемся над смыслом многих ежедневно употребляемых слов и выражений.

Итак, что же означает понятие «статистика»? Оно многолико, многозначно и в полном соответствии с одним из статистических терминов многомерно. В настоящее время насчитывается около тысячи определений статистики. Первое из них относится к 1749 г. Затем на протяжении 250 лет оно уточнялось и дополнялось. Дать определение статистики как науки пытались философы, математики, экономисты, социологи, государственные деятели и конечно же сами статистики.

Прежде всего, следует сказать, что статистика — это ***эффективное орудие***, ***инструмент познания***, используемый в естественных и общественных науках для установления тех специфических закономерностей, которые действуют в конкретных массовых условиях. Вместе с тем статистика — одна из форм *практической деятельности* людей, цель которой — сбор, обработка и анализ данных о тех или иных явлениях. Когда мы говорим о государственной или ведомственной статистике РФ, об организации статистики в России, то как раз и имеем в виду особую форму общественно-полезной деятельности людей. Статистикой называют также различного рода ***цифровые данные***, характеризующие, например, различные стороны жизни конкретной страны: политические отношения, культуру, население, производство и т. д.

Нередко слово «статистика» употребляется в качестве более короткого эквивалента для словосочетания «***статистические методы***». Статистические методы применяют тогда, когда из больших массивов данных требуется выделить полезную для нас информацию. Иногда термин «статистика» может употребляться одновременно в *нескольких значениях*. Известный английский статистик У. Дж. Рейхман (р. 1920 г.) 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»[1].

Сила статистики 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, которой 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»[2].

Развитие статистики сходно с развитием языка и счета. Эта наука имеет древние корни. Она зародилась как результат обобщения уже достаточно развитой статистической практики, вызванной потребностями развития общества.

Вот лишь некоторые сведения. В Китае более чем за две тысячи лет до нашей эры производились 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.

---

[1] *Рейхман У. Дж.* Применение статистики. — М. : Статистика, 1969. — С. 11.

[2] *Успенский Г.* ПСС. Т. 10. Кн. 2. — М. : Изд-во Академии наук СССР, 1954. — С. 155—156.

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

В Европе в конце IX в. проводились первые учетные операции: инвентаризация королевских имений, учет жителей, пригодных 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 несовершенен. В тот период статистические операции, как правило, проводились в исключительных случаях и в основном в военной и финансовой сферах. Позднее потребность в статистических операциях порождалась необходимостью стимулирования роста народонаселения, производительных сил страны, регулирования потребления.

Однако если сбор статистических данных начался в самой глубокой древности, то их обработка и анализ, 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) с 1746 г. впервые вначале в Марбургском, а затем в Геттингенском университете стал читать новую учебную дисциплину, которую он назвал *статистикой*. Школа просуществовала более 150 лет, не меняя своих теоретических основ. Предмет и метод этой науки не были четко определены. Собирался в основном описательно-информационный материал, который впоследствии почти не анализировался. Описывался, как правило, последний период, иначе, по мнению представителей школы, в статистической работе не было смысла.

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

Гораздо ближе к сегодняшнему пониманию статистики была английская ***школа политических арифметиков.*** Она возникла на 100 лет раньше немецкой описательной школы. Основоположниками школы «Политическая арифметика» были Д. Граунт (1620—1674), Э. Галлей (1656—1742) и В. Петти (1623—1687). В их трудах преобладали 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 лондонского населения Д. Граунт впервые открыл некоторые закономерности массовых общественных явлений и показал, как следует обрабатывать и анализировать множественный первичный материал. Он впервые попытался построить таблицу смертности для стационарного населения. Теоретическую разработку проблемы смертности продолжил Э. Галлей. Знаменитый английский астроном выдвинул 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 явную 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 массовом материале. Цели и задачи, которые ставили перед собой эти 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 сосредоточилась 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-математического***. Особый вклад в развитие этого направления внес бельгийский статистик А. Кетле (1796—1874). Он называл статистику «социальной физикой», т. е. наукой, изучающей законы общественной системы с помощью количественных методов. Важнейшей его заслугой является 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. К. Пирсон также провел много плодотворных исследований в статистике. Наряду с Ф. Гальтоном он внес 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.

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 ценным источником сведений по экономической теории России с древних времен до XVIII в.

Превращению статистики из науки описательной в науку теоретическую, т. е. формированию статистики как подлинной науки, положили начало 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 большая роль принадлежит трудам 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 уделял большое внимание проблеме достоверности данных, методу группировок, раскрыл принцип единства количественного и качественного анализа.

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 это время значительный вклад в теорию индексного метода был внесен учеными С. М. Югенбергом, В. Е. Адамовым, Г. И. Баклановым, Л. С. Казинцом, И. Г. Венецким и др. Заслуживают серьезного внимания труды по изучению статистической связи Я. И. Лукомского. Большим шагом вперед в развитии 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.

Для сбора, обработки и анализа статистической информации в настоящее время в Российской Федерации функционирует единая централизованная система государственной статистики. Центральным органом этой системы является Федеральная служба государственной статистики (Росстат). Статистическая деятельность в России регулируется Федеральным законом № 282-ФЗ от 29 ноября 2007 г. «Об официальном статистическом учете и системе государственной статистики в Российской Федерации». В субъектах Российской Федерации — республиках, краях, областях и районах — статистическая работа осуществляется территориальными органами государственной статистики, комитетами или отделами.

Непосредственная обработка поступающих из регионов статистических данных осуществляется в Главном межрегиональном центре обработки и распространения статистической информации, который обладает необходимыми для этих целей мощными вычислительными ресурсами.

На Федеральную службу государственной статистики возложено как методологическое, так и практическое руководство всеми работами по сбору, обработке и анализу статистических данных на государственном уровне. Для реализации этих задач в структуре Федеральной службы государственной статистики выделены Управления:

организации статистического наблюдения и контроля;

национальных счетов;

статистики предприятий;

сводных статистических работ и общественных связей;

статистики цен и финансов;

статистики сельского хозяйства и окружающей природной среды;

статистики строительства, инвестиций и жилищно-коммунального хозяйства;

статистики торговли и услуг;

статистики труда, науки, образования и культуры;

статистики уровня жизни и обследований домашних хозяйств;

статистики населения и здравоохранения;

статистики зарубежных стран и международного сотрудничества;

организации поведения переписей и сплошных обследований;

административное и др.

Федеральная служба государственной статистики ежегодно разрабатывает и утверждает Федеральную программу статистических работ на календарный год, которая согласовывается на заседании Правительства Российской Федерации.

Работа по сбору статистической информации проводится не только Федеральной службой государственной статистики. В соответствии с Федеральной программой отдельные виды статистических работ осуществляются другими органами государственного управления — Банком России, Минфином России, МВД России и др.

Получаемые Федеральной службой государственной статистики статистические данные прежде всего предоставляются органам федеральной власти, а также публикуются для широкого использования в аналитических целях научными и практическими работниками, руководителями и специалистами предприятий и организаций всех форм собственности. Основными статистическими ежегодными изданиями являются:

Российский статистический ежегодник;
Россия в цифрах;
Регионы России;
Демографический ежегодник России;
Промышленность России;
Строительство в России;
Сельское хозяйство, охота и лесоводство в России;
Малое и среднее предпринимательство в России;
Жилищное хозяйство и бытовое обслуживание населения в России;
Финансы России;
Цены в России;
Транспорт в России;
Инвестиции в России;
Наука в России;
Россия и страны мира и др.

Среди периодических — ежемесячных и ежеквартальных статистических изданий следует отметить:

«Статистический бюллетень»;
«Статистическое обозрение»;
«Информация о социально-экономическом положении России» (ежемесячный краткий доклад);
«Социально-экономическое положение России»;
«Вопросы статистики» (ежемесячный научно-информационный журнал).

С важнейшими социально-экономическими показателями Российской Федерации, а также с изданиями Росстата в электронном виде можно познакомиться через сеть Интернет на сайте Федеральной службы государственной статистики — http://www.gks.ru.

Статистическое исследование представляет собой изучение социально-экономических явлений и процессов посредством системы статистических методов и количественных характеристик. Оно проходит следующие стадии: а) сбор статистической информации и формирование информационной базы исследования (статистическое наблюдение); б) сводка и группировка данных статистического наблюдения; в) обобщение и анализ результатов обработки статистических данных, формулировка выводов и рекомендаций по итогам статистического исследования в целом.

Первым и исходным этапом статистического исследования является этап статистического наблюдения. Именно в процессе наблюдения формируется исходная статистическая информация, являющаяся основой статистического исследования.

Статистическая информация — это совокупность сведений социального и экономического характера, полученных в результате статистического наблюдения, на основе которых осуществляются такие функции, как учет, контроль, планирование, статистический анализ и управление.

Статистическое наблюдение представляет собой научно организованный, планомерный и систематический сбор массовых сведений о социально-экономических явлениях и процессах путем регистрации заранее намеченных существенных признаков.

Рассмотрим составляющие данного определения.

Планомерность статистического наблюдения заключается в том, что оно подготавливается и проводится по заранее четко разработанному плану, который включает в себя вопросы методологии, организации и техники сбора статистической информации, контроля за ее качеством и достоверностью.

Систематичность статистического наблюдения определяется тем, что оно должно проводиться не от случая к случаю, а постоянно через равные промежутки времени. Это обусловливается тем, что изучаемые социально-экономические явления и процессы изменяются во времени и пространстве. Изучение тенденций и закономерностей этих изменений возможно лишь на основе систематической регистрации значений признаков, т. е. систематического наблюдения.

Массовый характер статистического наблюдения означает, что оно направлено на охват достаточно большого объема информации о социально-экономических явлениях и процессах, чтобы тенденции и закономерности их изменения проявились достаточно полно и достоверно. Статистическое наблюдение может проводиться органами государственной статистики, научно-исследовательскими институтами, экономическими и аналитическими службами различных организационно-правовых структур. Главная задача проведения статистического наблюдения состоит в получении достоверных, объективных и точных данных, необходимых для различных социально-экономических исследований, выявления и изучения структуры, взаимосвязей, закономерностей и тенденций в развитии общественных явлений и процессов.

Процесс проведения статистического наблюдения включает в себя ряд этапов. Первый из них — программно-методологическая подготовка проведения наблюдения. В нее входят: определение цели и объекта наблюдения, состава признаков, подлежащих регистрации; разработка документов для сбора данных; выбор отчетной единицы и единицы, относительно которой будет проводиться наблюдение; определение методов и средств получения данных. Далее следует организационная подготовка проведения наблюдения. Она включает следующие виды работ:

подбор и подготовка кадров;

составление календарного плана работ по подготовке и проведению статистического наблюдения, обработке его материалов;

подготовка технической документации и оборудования.

После этого выбирают форму, способ и вид статистического наблюдения.

Статистическое наблюдение включает следующие стадии:

1) сбор данных наблюдения, накапливание статистической информации;

2) синтаксический, логический и арифметический контроль данных статистического наблюдения, которые основываются на знании документооборота, логических и арифметических взаимосвя-

зей между показателями, их количественными и качественными характеристиками;

3) формулировка выводов и предложений по проведению статистического наблюдения, анализ точности и достоверности полученных данных и причин возможного возникновения ошибок наблюдения.

От правильной, научно обоснованной организации и проведения статистического наблюдения зависят результаты статистического исследования в целом. На основе материалов статистического наблюдения, их обработки и анализа разрабатываются решения и предложения для оценки состояния и прогнозирования социально-экономического развития страны, отдельных регионов, видов экономической деятельности, организационно-правовых структур и т. д.

Важнейшим этапом подготовки статистического наблюдения является разработка плана его проведения, который содержит формулировку и решение программно-методологических и организационных вопросов. К программно-методологическим вопросам статистического наблюдения относятся: установление цели и задач наблюдения; определение объекта и единицы наблюдения; разработка программы наблюдения; выбор вида и способа наблюдения.

Установление цели и задач является исходным этапом при организации и проведении любого статистического наблюдения, а следовательно, и статистического исследования в целом. Цель проводимого статистического наблюдения должна быть сформулирована ясно, четко и развернуто. В целом конечной целью любого статистического наблюдения является получение достоверной и научно обоснованной информации о состоянии и закономерностях развития реальных социально-экономических явлений и процессов. Но это общее требование всегда имеет воплощение в виде конкретно обозначенной цели. Она определяет прямые предметные задачи наблюдения, которые также должны быть квалифицированно и строго сформулированы. Задачи статистического наблюдения предопределяют программу и организационные формы его проведения. В соответствии с целями и задачами определяются объект и единица статистического наблюдения.

Объект статистического наблюдения — это совокупность социально-экономических явлений и процессов, которая подвергается статистическому наблюдению. Объектом наблюдения может быть, например, народонаселение страны, группа коммерческих банков, совокупность страховых компаний, финансово-промышленных групп и т. д.

Для успешного проведения статистического наблюдения большое значение имеет точное и научно обоснованное определение объекта наблюдения, т. е. конкретной совокупности явлений или процессов, их границ и направления изучения.

Выделение объекта статистического наблюдения, определение его границ предполагает, что из всей совокупности существенных признаков, характеризующих обследуемый объект, выделяются те из них, которые отличают его от других, сходных или близких с ним по характеру объектов. Например, недостаточно указать, что объектом наблюдения является совокупность банков. Необходимо обязательно указать форму собственности, организационно-правовую форму банков. Следует предписать, какие именно банки подлежат наблюдению: крупные, мелкие, средние или все банки, банки Москвы, Санкт-Петербурга, другого региона или страны в целом и т. д. Столь же важно определить, какие именно финансово-кредитные институты относятся к банкам, какие признаки их отличают от тех, которые к банкам не относятся.

Всякий объект статистического наблюдения состоит из отдельных элементов — единиц наблюдения. Определяя объект наблюдения, важно четко определить единицу статистического наблюдения и единицу статистической совокупности.

Единица статистического наблюдения — это первичный, составной элемент объекта статистического наблюдения, который является носителем регистрируемых при наблюдении признаков. В качестве единицы наблюдения могут выступать банк, страховая компания, финансово-промышленная группа, высшее учебное заведение, в зависимости от того, какой объект обследуется. Определение единицы наблюдения должно содержать указание ее основных отличительных признаков. Единицы наблюдения называются отчетными единицами.

Отчетная единица — это единица статистического наблюдения, от которой поступают отчетные данные по утвержденным для нее в установленном порядке формам. Так, в системе статистической отчетности в строительстве отчетными единицами являются проектные и строительно-монтажные организации.

Единица статистической совокупности — это отдельно взятый первичный, составной элемент статистической совокупности, который служит основой счета и обладает признаками, подлежащими регистрации при проведении статистического наблюдения. Единица совокупности — это то, что при проведении статистического наблюдения подвергается обследованию.

Единицы совокупности и единицы наблюдения могут совпадать. Если, скажем, необходимо определить объем строительно-монтажных работ, то единицей совокупности и единицей наблюдения в обоих случаях будет строительно-монтажная организация. Но, как правило, единица совокупности отличается от единицы наблюдения. Так, при обследовании ввода в действие объектов производственного и непроизводственного назначения единицей совокупности будет выступать отдельный объект, поскольку признаки, регистрируемые в данном обследовании, относятся не к строительно-монтажной организации, а к каждому отдельно взятому строительному объекту производственного и непроизводственного назначения. Единицей же наблюдения в данном обследовании будут строительно-монтажные организации.

Единицы наблюдения и состоящий из них объект наблюдения в целом обладают множеством характерных черт, специфических особенностей, различных свойств, называемых в статистике признаками, которые невозможно учесть полностью в одном исследовании. Поэтому при организации статистического наблюдения важен отбор из всей совокупности признаков только тех, которые являются существенными и информативными для характеристики изучаемого объекта. Именно эти признаки регистрируются в процессе наблюдения. Таким образом, исходя из конкретного содержания объекта, цели и задач предпринимаемого статистического исследования разрабатывается программа статистического наблюдения.

Программа статистического наблюдения — это перечень признаков единицы наблюдения, регистрируемых в процессе проведения статистического наблюдения. В практическом выражении программа наблюдения представляет перечень наиболее значимых в практическом и теоретическом аспекте вопросов, на которые должны быть получены ответы от каждой единицы наблюдения.

Разработка программы наблюдения является одной из самых важных теоретических и практических проблем статистического наблюдения, предполагающих знание экономических, социальных и статистических основ и специфики изучаемых явлений и процессов. От качества разработки программы зависят результаты статистического наблюдения и статистического исследования в целом, его ценность для решения социально-экономических задач, пригодность для выработки конкретных производственных и управленческих предложений.

Программа статистического наблюдения должна строиться с учетом следующих требований, предъявляемых к ее разработке:

программа должна содержать существенные признаки, непосредственно характеризующие изучаемое социально-экономическое явление, процесс, их тип, основные черты, специфические особенности и свойства, которые будут использованы при разработке материалов статистического наблюдения. Программа наблюдения не должна включать второстепенные или косвенные вопросы, так как в многообразии малозначимой информации может потеряться главное и существенное в изучаемом явлении или процессе;

программа наблюдения должна включать только те признаки, которые являются отличительными для всех единиц изучаемой совокупности;

от качества программы зависят полнота и достоверность сведений, получаемых в ходе статистического наблюдения. При этом необходимо учитывать и контролировать количество вопросов программы. Если отсутствуют уверенность и предпосылки в получении полных и достоверных данных по широкой программе, то лучше ограничить объем собираемых сведений, чтобы получить небольшой, но достоверный статистический материал;

вопросы программы наблюдения должны быть сформулированы четко, кратко и ясно, без допущения двусмысленности их толкования;

программа наблюдения должна содержать только те вопросы, на которые реально можно получить объективные и достаточно точные ответы;

в программу наблюдения целесообразно включать вопросы контрольного характера, чтобы ответы на одни из них могли использоваться для контроля ответов на другие. Например, для выяснения времени работы строительной фирмы на рынке в исследовании может быть поставлено два вопроса: о дате создания фирмы и о числе лет работы фирмы;

вопросы в программе наблюдения должны быть изложены в логической последовательности, а не в хаотичном порядке.

Для проведения наблюдения в каждом конкретном случае разрабатывается инструментарий статистического наблюдения, который включает в себя формуляр и инструкцию.

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

В зависимости от конкретного содержания и специфических особенностей проводимого статистического наблюдения формуляр наблюдения может иметь различные формы выражения и наименования: бланк, переписной лист, опросный лист, форма отчетности, анкета и т.д. В статистике различают две системы статистического формуляра: индивидуальную (формуляр-карточка) и списочную (формуляр-список). ***Индивидуальный формуляр*** (***формуляр-карточка***) — это формуляр, предназначенный для регистрации в нем ответов на вопросы программы наблюдения только об одной единице наблюдения. ***Списочный формуляр*** (***формуляр-список***) — это формуляр, предназначенный для регистрации в нем ответов на вопросы программы наблюдения о нескольких единицах наблюдения. Примером списочной системы формуляров служат переписные листы переписи населения страны, каждый из которых предназначается для регистрации нескольких лиц и характеризующих их признаков.

Для обеспечения единообразного толкования вопросов программы наблюдения и облегчения их понимания в статистических формулярах может быть дан статистический подсказ, который выражается либо в перечислении возможных ответов на вопросы, либо в указании методики или способа расчета того или иного показателя. ***Статистический подсказ*** — это перечень возможных ответов на поставленный вопрос. Например, при обследовании строительных компаний города после вопроса о величине рентабельности в скобках необходимо указать методику расчета рентабельности как отношение прибыли к себестоимости работ. Полным называется статистический подсказ, содержащий исчерпывающий перечень ответов на поставленный в статистическом формуляре вопрос. Полный подсказ предполагает выбор заполняющим формуляр только одного из перечисленных ответов на вопрос. Неполным называется статистический подсказ, содержащий некоторые (но не все) из возможных ответов на поставленный в статистическом формуляре вопрос.

С целью обеспечения более полного и правильного процесса организации и проведения наблюдения к программе статистического наблюдения составляется подробная инструкция. ***Статистическая инструкция*** — это документ, разъясняющий вопросы программы статистического наблюдения, порядок заполнения статистического формуляра и частично планово-организационные вопросы. В инструкции отражаются цели и задачи наблюдения, сведения об объекте и единицах наблюдения, о времени и сроках проведения наблюдения, об оформлении результатов и сроков их представления в соответствующие организации. В инструкции, так же как и в формуляре наблюдения, но в более развернутом виде, могут быть представлены толкование того или иного вопроса программы, примерные ответы на вопросы, типичные случаи заполнения формуляров и т. д. Инструкция статистического наблюдения должна быть составлена кратко, просто и ясно.

В целях успешной подготовки и проведения статистического наблюдения важно решить не только программно-методологические, но и организационные вопросы, для реализации которых разрабатывается организационный план статистического наблюдения.

Организационный план статистического наблюдения — это основной документ, в котором излагается порядок подготовки, организации и проведения статистического наблюдения, фиксируется перечень мероприятий, необходимых для успешного проведения работы по сбору данных, с указанием конкретных сроков проведения намеченных мероприятий. В организационном плане статистического наблюдения указываются объект наблюдения (его описание, определение, отличительные признаки), цели и задачи наблюдения, органы наблюдения, место, время и сроки наблюдения, круг лиц, отвечающих за проведение наблюдения, подготовительные работы к проведению наблюдения, в том числе порядок комплектования и обучения кадрового состава, необходимого для проведения наблюдения, сроки и порядок приема, оформления и сдачи материалов наблюдения, порядок получения и представления промежуточных и окончательных итогов наблюдения. Могут быть и другие практические вопросы.

Обязательным элементом организационного плана статистического наблюдения является указание органа наблюдения. ***Орган***

наблюдения — это организация, учреждение или их подразделение, осуществляющие подготовку, проведение статистического наблюдения и несущие ответственность за эту работу. Должны быть четко определены обязанности каждого органа с четким указанием сферы его деятельности, прав и круга обязанностей, за которые он несет ответственность, а также порядок взаимоотношений органов наблюдения между собой. При проведении статистических наблюдений в масштабе страны органом наблюдения выступает Государственный комитет по статистике РФ и его местные органы. Органами, проводящими локальные статистические наблюдения, могут выступать министерства, ведомства, различные организационно-правовые структуры и т. д.

При подготовке статистического наблюдения с целью обеспечения полноты охвата объекта наблюдения и достоверности собираемых сведений в организационном плане четко устанавливается место наблюдения. ***Место статистического наблюдения*** — это место, где непосредственно производится регистрация наблюдаемых фактов и заполнение статистических формуляров. Место наблюдения нередко совпадает с местонахождением единицы наблюдения. Например, статистическая отчетность о деятельности строительной фирмы, коммерческого банка, страховой компании составляется по месту их нахождения. Более сложен вопрос о месте наблюдения при проведении специально организованного наблюдения, в случае если единицы наблюдения обследуемого объекта меняют или могут изменить место своего пребывания. Например, при проведении переписи населения страны зафиксировано, что оно учитывается строго по месту жительства, а не по месту работы или службы.

Важное место в составлении организационного плана занимает установление времени, к которому относятся регистрируемые в ходе статистического наблюдения сведения. ***Время наблюдения*** — это время, по состоянию на которое или за которое регистрируются сведения в процессе статистического наблюдения. Дело в том, что объект статистического наблюдения, его объем и состав изменяются во времени. В зависимости от характера объекта и его специфики, а также сущности показателей, описывающих данный объект, сведения могут регистрироваться по состоянию на определенную дату (на начало или конец года, на конкретное число) или за определенный промежуток времени (месяц, квартал, год и т. д.).

Так, например, данные о производстве продукции, о материальных ресурсах предприятия регистрируются за определенный период времени, а численность населения страны, курс иностранной валюты фиксируются на определенную дату. Иными словами, при организации статистического наблюдения должен быть решен вопрос о времени производства наблюдения, который включает в себя установление срока (периода) наблюдения, а в некоторых случаях и момента времени, на который должны учитываться единицы наблюдения и признаки, их характеризующие, т. е. установлен критический момент наблюдения.

Срок (период) наблюдения — это время, в течение которого производится заполнение статистических формуляров, т. е. осуществляется регистрация единиц наблюдения по установленной программе.

Срок наблюдения определяется рядом факторов. В первую очередь он зависит от специфики и особенностей объекта наблюдения. Так, чем крупнее объект наблюдения, тем, при прочих равных условиях, требуется больше времени для проведения статистического наблюдения над ним. Срок диктуется также программой наблюдения, ее объемом и сложностью признаков, подлежащих регистрации. Более развернутая, детализированная программа, содержащая сложные вопросы, требует больше времени для проведения наблюдения. Наконец, на сроки влияет наличие кадров, которые могут быть привлечены к проведению наблюдения, степень их квалификации и численность. Достаточное число высококвалифицированных работников способствует сокращению сроков проведения наблюдения.

Срок наблюдения, как правило, предполагает указание даты начала и завершения статистического наблюдения. Например, срок проведения обследования страховых компаний в одном из регионов РФ составил 30 дней — с 01.01.2012 г. по 30.01.2012 г. Это и было время, в течение которого оно проводилось.

Критический момент статистического наблюдения — это момент времени (конкретный год, день и час), по состоянию на который производится регистрация собираемых сведений в процессе статистического наблюдения. В качестве критического момента времени обычно выбирают 00.00 часов — полночь, т. е. момент перехода от одних суток к другим. Все сведения независимо от времени регистрации должны фиксироваться такими, какими они были в критический момент. Все изменения, которые происходят с единицами наблюдения после критического момента времени, во внимание не принимаются. Так, например, при проведении опроса студентов вуза о качестве преподавания дисциплины «Статистика» в режиме on-line в качестве критического момента были выбраны 00.00 часов с 20 на 21 января 2012 г. и все анкеты, отправленные студентами после 00.00 часов 21.01.2012 г., для обработки и анализа не принимались. Установление критического момента наблюдения способствует обеспечению сопоставимости статистических данных обо всех единицах наблюдения.

Важной задачей в ходе проведения статистического наблюдения является получение достоверных и объективных данных о состоянии обследуемых объектов. В этих целях желательно, чтобы срок наблюдения был не очень удален во времени от критического момента наблюдения. Чем короче такой период, тем меньше изменений происходит в объекте наблюдения и тем проще, в случае необходимости, восстановить истинное состояние объекта наблюдения на тот или иной момент времени.

Существенное значение для успешного проведения статистического наблюдения имеет определение кадрового состава. Численность лиц, проводящих наблюдение, зависит от объекта наблюдения, количества единиц наблюдения, составляющих его, от срока, программы и способа проведения наблюдения, от территориального распределения единиц наблюдения и т. д. Подготовка кадров для проведения статистического наблюдения включает обучение и инструктирование кадрового состава с целью выработки практических навыков правильного заполнения статистических формуляров.

Подготовительный этап организации статистического наблюдения предполагает подготовку и размножение бланков, инструкций, другой документации и рассылку их на места. К числу важнейших подготовительных работ по проведению статистического наблюдения относится его пропаганда через средства массовой информации и локально. Успешное проведение статистического наблюдения обеспечивается четкой структурой и разработанностью его организационного плана.

Статистическое наблюдение различается по организационным формам, видам, источникам сведений и способам их собирания.

Организационные формы статистического наблюдения многообразны, но все они могут быть сведены к двум основным: отчетность (предприятий, организаций, учреждений различных организационно-правовых форм) и специально организованное наблюдение (переписи, единовременные учеты, обследования сплошного и несплошного характера).

Отчетность — это организационная форма статистического наблюдения, при которой в установленные сроки сведения поступают в соответствующие статистические органы от предприятий, организаций и учреждений различных организационно-правовых форм в виде установленных в законном порядке отчетных документов (статистических отчетов), заполненных на основании данных первичного учета и подписанных лицами, ответственными за представление и достоверность содержащихся в них сведений. Отчетность является важнейшей формой статистического наблюдения. В ней содержатся основные учетно-статистические данные о состоянии и деятельности предприятий, организаций и других субъектов экономики. Все формы статистической отчетности утверждают органы государственной статистики.

Отчетность составляется на основе данных первичного учета и является их обобщением. ***Первичный учет*** — это первоначальная, систематическая регистрация в формах первичных учетных документов фактов, событий, актов производственно-хозяйственной деятельности предприятий, организаций различных организационно-правовых форм, производимая по мере совершения.

Различают общегосударственную и внутриведомственную отчетность. Главная особенность ***общегосударственной отчетности*** состоит в том, что она обязательна для всех без исключения предприятий, учреждений и организаций различных организационно-правовых форм и представляется в сводном виде в органы государственной статистики. ***Внутриведомственная отчетность*** действует в пределах отдельного министерства, ведомства. Она устанавливается для подведомственных предприятий, организаций и учреждений различных организационно-правовых форм и используется для своих оперативных потребностей.

Формы отчетности могут быть типовыми и специализированными. ***Типовая отчетность*** — это отчетность, содержащая показатели, одинаковые для всех предприятий, организаций и учреждений различных организационно-правовых форм, а также для разных производств и видов деятельности.

Специализированная отчетность вводится для предприятий, организаций и учреждений, имеющих определенные особенности. Она содержит наряду с общими показателями, имеющимися в соответствующей типовой отчетности, специфические показатели, характерные для определенных организационно-правовых форм, видов деятельности и производства.

По периодичности представления сведений отчетность подразделяется на периодическую и единовременную. ***Периодическая***

отчетность — это отчетность, представляемая через одинаковые промежутки времени или в точно определенные даты. Например, не позднее 5-го числа каждого месяца. Периодическая отчетность подразделяется на текущую, период представления которой составляет менее года (квартал, месяц, неделя и т. д.), и годовую, период представления которой составляет календарный год. ***Единовременная отчетность*** — это отчетность, представляемая только один раз либо по мере необходимости, либо без определенной периодичности. Чем больше временной период, за который отчитывается предприятие, организация, тем шире программа отчетности. Так, ежемесячная отчетность содержит более ограниченный круг показателей, чем годовая.

По способу представления отчетных данных различают ***отчетность почтовую*** и ***срочную***, представляемую по телеграфу, телетайпу, факсу и другими быстрыми способами.

Отчетность является одним из основных источников получения сведений о социально-экономическом состоянии и развитии страны в целом.

Специально организованное статистическое наблюдение — это наблюдение, организуемое с определенно заданной целью на определенную, как правило, дату для получения данных, которые в силу тех или иных причин не собираются посредством отчетности, или для проверки и уточнения данных отчетности, а также для глубокого и всестороннего анализа конкретных социально-экономических явлений и процессов. Примером такой организационной формы статистического наблюдения являются переписи населения, бюджетные обследования домашних хозяйств, инвентаризации и переоценки основных фондов в экономике страны в целом, социологические обследования и т. д.

В зависимости от степени охвата наблюдением единиц изучаемого объекта статистическое наблюдение подразделяется на сплошное и несплошное. ***Сплошное статистическое наблюдение*** — это наблюдение, при котором обследованию подвергаются все без исключения единицы изучаемой совокупности явлений и процессов. Типичным примером сплошного наблюдения являются переписи населения, при проведении которых регистрации по основной программе подлежат все без исключения жители страны. К сплошному наблюдению относится также текущая отчетность предприятий и организаций, которая охватывает все подотчетные объекты различных организационно-правовых форм, и т. д.

Несплошное статистическое наблюдение предполагает, что обследованию подвергаются не все единицы изучаемой совокупности, а только часть из них. При проведении несплошного наблюдения заранее определяется факт его проведения, принцип и методы отбора и формирования части совокупности, которая будет подвергнута наблюдению. Примером несплошного наблюдения является регистрация цен на отдельные виды товаров длительного пользования. В практике статистического исследования применяется несколько видов несплошного наблюдения: выборочное, монографическое, метод основного массива.

Выборочное наблюдение — это вид несплошного наблюдения, основанный на принципе случайного отбора тех единиц совокупности, которые должны быть подвергнуты статистическому наблюдению. Правильная организация и проведение выборочного наблюдения позволяют получить достаточно достоверные статистические данные для характеристики изучаемой совокупности в целом.

Монографическое наблюдение — это вид несплошного наблюдения, предполагающий изучение и описание отдельных, характерных в каком-либо отношении, единиц совокупности с целью характеристики всей совокупности социально-экономических явлений и процессов в целом. Практическое применение монографического наблюдения ставит задачу выявления реально существующих или намечающихся тенденций и закономерностей в развитии социально-экономических явлений и процессов с целью определения резервов их состояния и развития, изучения опыта отдельных предприятий и организаций различных организационно-правовых форм.

Метод основного массива — это вид несплошного наблюдения, при котором обследованию подвергаются наиболее существенные, крупные единицы изучаемой совокупности, где объем изучаемого признака составляет наибольший, преобладающий удельный вес. Единицы совокупности, обладающие незначительной величиной изучаемого признака, обследованию не подвергаются. Данный вид наблюдения применяется, например, при регистрации цен на рынках продовольственных и непродовольственных товаров города.

В зависимости от временно́го фактора, т. е. по частоте регистрации сведений, наблюдение бывает текущее и прерывное. ***Текущее наблюдение*** осуществляется систематически, путем непрерывной регистрации фактов по мере их возникновения. Примером текущего наблюдения является регистрация актов гражданского состояния. ***Прерывное наблюдение*** означает, что регистрация фактов производится регулярно, через определенные промежутки времени или по мере необходимости.

Различают периодическое и единовременное прерывное статистическое наблюдение. ***Периодическим наблюдением*** называют наблюдение, которое проводится регулярно, через определенные, равные промежутки времени. Примером могут служить переписи населения, которые проводятся каждые 10 лет. ***Единовременное наблюдение*** проводится по мере необходимости, время от времени, без соблюдения строгой периодичности или вообще проводится один раз и больше никогда не повторяется. Примером данного вида наблюдения является учет товарных остатков и денежной наличности на момент денежной реформы.

В зависимости от источников собираемых сведений различают наблюдение непосредственное, документальное и опрос. Иными словами, сведения в ходе статистического наблюдения могут быть получены либо непосредственно от работника на месте наблюдения, либо в виде соответствующих документов, либо путем регистрации показаний опрашиваемых.

Непосредственное наблюдение осуществляют сами регистраторы с помощью осмотра, путем непосредственного замера, взвешивания или подсчета признаков изучаемого явления. Тем самым устанавливают факт и регистрируют его в формуляре статистического наблюдения. Примером может служить инвентаризация имущества предприятия или учреждения. ***Документальное наблюдение*** — это такое наблюдение, когда запись ответов на вопросы формуляра наблюдения производится на основании соответствующих документов. Например, для составления статистической отчетности инвестиционно-финансовой компании используются данные, взятые из документов бухгалтерского учета. ***Опрос*** — это наблюдение, при котором ответы на вопросы формуляра наблюдения записываются со слов опрашиваемого. Данный вид наблюдения характерен для переписей населения, социологических обследований. Наибольшая точность собираемых сведений достигается при непосредственном и документальном наблюдении.

В статистической практике применяются различные способы статистического наблюдения. Это отчетный, экспедиционный, метод самоисчисления, а также корреспондентский, анкетный и явочный.

Отчетный способ статистического наблюдения заключается в представлении предприятиями, организациями и учреждениями различных организационно-правовых форм статистических отчетов о своей деятельности в строго установленном порядке (сроки, адреса, формы). Это наиболее распространенный способ сбора статистических данных в нашей стране.

Экспедиционный способ статистического наблюдения заключается в том, что специально привлеченные и обученные работники, именуемые счетчиками или регистраторами, посещают каждую единицу наблюдения и сами регистрируют сведения о наблюдаемом явлении в формуляре статистического наблюдения. Данный способ применяется только при специально организованном наблюдении (например, во время переписи населения).

Способ самоисчисления (***саморегистрации***) предполагает, что формуляры статистического наблюдения заполняют сами опрашиваемые, а специально привлеченные работники обеспечивают опрашиваемых формулярами наблюдения, инструктируют их, собирают заполненные формуляры и проверяют правильность их заполнения. Данный способ статистического наблюдения применяется, например, при изучении маятниковой миграции — передвижения населения от места жительства до места работы или учебы.

Анкетный способ статистического наблюдения сводится к сбору статистических данных с помощью специальных вопросников (анкет), рассылаемых определенному кругу лиц или публикуемых в периодической печати. Заполнение анкет носит добровольный характер и осуществляется, как правило, анонимно. Данный способ применяется в обследованиях, где не требуется получения результатов, отличающихся высокой точностью. Поэтому заполненные бланки в присутствии опрашиваемого не проверяются.

Корреспондентский способ статистического наблюдения состоит в том, что статистические органы договариваются с определенными лицами, которые берут на себя обязательство вести наблюдение за социально-экономическими явлениями и процессами, составляющими объект наблюдения, и сообщать его результаты статистическим органам. Статистические органы снабжают корреспондентов формулярами наблюдения, инструкциями и другими необходимыми материалами для проведения статистического наблюдения. Корреспонденты не состоят в штатах статистических органов. Данный способ статистического наблюдения применяется, как правило, при проведении обследований экспертным методом.

Явочный способ статистического наблюдения предполагает представление сведений в органы, которые ведут наблюдение, в явочном порядке. Применяется, например, при регистрации актов гражданского состояния.

Выбор вида и способа статистического наблюдения зависит от характера объектов статистического наблюдения, от предполагаемой и ожидаемой степени точности получаемых сведений, а также от финансовых возможностей при организации наблюдения.

Всякое статистическое наблюдение предполагает получение данных, которые бы точно и полно отражали действительность. При этом, как бы тщательно ни было подготовлено наблюдение, в процессе его проведения могут возникнуть погрешности, которые приводят к снижению достоверности статистического наблюдения. Точность данных является основным требованием, предъявляемым к статистическому наблюдению.

Точность наблюдения — это степень соответствия значения какого-либо признака или показателя, полученного посредством статистического наблюдения, действительному его значению. Чем меньше расхождения между показателями, полученными в результате статистического наблюдения, и фактическими их значениями, тем выше степень точности статистического наблюдения. Таким образом, точность статистического наблюдения может быть охарактеризована абсолютной величиной — разностью данных наблюдения и действительных значений изучаемых признаков или показателей, или относительной величиной — их соотношением. Расхождение между установленными статистическим наблюдением и действительными значениями изучаемых величин называется ***ошибками наблюдения***. В зависимости от характера, степени влияния на конечные результаты наблюдения, источников и причин возникновения различают несколько видов таких ошибок.

Ошибки регистрации — это расхождение между зафиксированным при статистическом наблюдении значении признака и действительным его значением в результате неправильной, ошибочной регистрации ответа на вопрос статистического формуляра. Речь идет об ошибках регистрации, которые возникают вследствие неправильного установления фактов в процессе наблюдения, или ошибочной их записи, или и того и другого вместе. Этот вид ошибок может быть и при сплошном, и при несплошном наблюдении. Ошибки регистрации могут быть случайными и систематическими.

Случайные ошибки регистрации — это ошибки, которые возникают вследствие различных случайных причин. Случайные ошибки могут быть допущены как опрашиваемыми при ответах на вопросы, так и регистраторами при заполнении формуляра наблюдения, а также в результате ошибок в подсчете. Такого рода ошибки обычно возникают, если опрашиваемый оговорился или при записи результата опроса допустил описку. Кроме того, регистратор может ослышаться, случайно переставить местами цифры. Бывает и так, что при заполнении статистического формуляра регистратор перепутал строки или графы и т. д.

Такого рода ошибки направлены как в сторону преуменьшения, так и в сторону преувеличения значения регистрируемого количественного признака. Случайные ошибки регистрации действуют в противоположных направлениях и при сводной обработке результатов достаточно большого числа наблюдений они, как правило, взаимопогашаются и на конечном результате не отражаются.

Систематические ошибки регистрации — это неточности, возникающие в силу определенных и постоянно действующих на протяжении всего статистического наблюдения в одном направлении факторов. Это существенно искажает итоги наблюдения. В каждом конкретном случае систематические ошибки регистрации имеют тенденцию или к увеличению, или к уменьшению значения показателя по каждой единице наблюдения. В конечном итоге ошибка значения показателя накапливается. Это обусловливает однонаправленное искажение конечного результата статистического наблюдения. Примером систематической ошибки регистрации при проведении социологических опросов населения могут служить округления населением возраста, как правило, на цифрах, оканчивающихся на 0 и 5. Часто опрашиваемые, например, вместо 39 лет или 42 лет говорят, что им 40 лет.

Систематические ошибки регистрации могут быть преднамеренными и непреднамеренными.

Преднамеренные систематические ошибки регистрации — это ошибки, являющиеся результатом того, что опрашиваемый умышленно, сознательно сообщает регистратору неправильные данные. К преднамеренному искажению данных относятся занижения величины прибыли в формах отчетности некоторых предприятий и организаций.

Непреднамеренные систематические ошибки регистрации — это ошибки, которые носят нечаянный характер, допускаются неумышленно. Такие ошибки могут быть результатом недостаточной квалификации работников, их небрежности в работе. Сюда относятся пропуски в записях, использование неисправных измерительных приборов и т. д.

При несплошном наблюдении зачастую возникают ошибки репрезентативности. ***Ошибки репрезентативности*** — это расхождение между значениями изучаемого признака или показателя в отобранной и обследованной части совокупности (выборочной) и его значениями во всей исходной (генеральной) совокупности. Причина возникновения данного рода ошибок заключается в том, что отобранная и обследованная часть изучаемой совокупности недостаточно точно отражает состав всей совокупности в целом.

Ошибки репрезентативности могут быть случайными и систематическими. ***Случайные ошибки репрезентативности*** — это ошибки, возникающие в силу несплошного характера статистического наблюдения, когда совокупность отобранных на основе принципа беспристрастного, непреднамеренного, случайного отбора единиц наблюдения недостаточно полно и точно воспроизводит совокупность в целом. ***Систематические ошибки репрезентативности*** — это ошибки, возникающие вследствие нарушения принципов беспристрастного, непреднамеренного отбора единиц изучаемой (генеральной) совокупности, которые должны быть подвергнуты наблюдению.

Для обеспечения необходимой достоверности данных статистического наблюдения необходимо предусмотреть тщательную проверку их качества с точки зрения полноты охвата изучаемого объекта статистическим наблюдением, качества и полноты заполнения формуляров и других документов наблюдения. Целью таких проверок является установление достоверности данных наблюдения, правильности ведения учета, выявление и устранение причин несвоевременного представления информации и фактов искажения данных.

Проверка данных статистического наблюдения на достоверность предполагает проведение синтаксического, логического и арифметического (счетного) контроля.

Полученные в результате наблюдения данные прежде всего подвергаются ***синтаксическому контролю***, который заключается в проверке правильности структуры документа, наличия необходимых реквизитов, в проверке оформления документов на предмет наличия и четкости всех необходимых записей, предусмотренных инструкцией, а также в анализе полноты материала и охвата всех отчетных единиц наблюдения.

Логический контроль основывается на знании логических взаимосвязей между показателями. Он заключается в сопоставлении ответов на взаимосвязанные вопросы статистического формуляра или другого документа статистического или социологического обследования с целью выявления логически несовместимых ответов. В случае обнаружения логически несовместимых ответов истинный ответ определяется путем дальнейших сопоставлений с ответами на другие вопросы. Например, в ходе социологического обследования установлено, что мужчина 13 лет, работающий в строительной фирме, зарегистрирован как имеющий высшее образование. Тут очевидна ошибка в регистрации возраста.

Арифметический (счетный) контроль основывается на знании количественных связей между значениями показателей. Он заключается в проверке правильности арифметических итогов показателей, содержащихся в отчетности или в формулярах обследования и в увязке тех показателей, которые могут быть выведены один из другого. В ряде случаев при арифметическом контроле данных статистического наблюдения и особенно отчетности можно привлекать материалы из других источников по тем же показателям. Например, данные годовых отчетов можно проверять данными квартальных отчетов и т. д. Научно обоснованная и полная проверка данных статистического наблюдения и объективная оценка его точности позволяют получить достаточно полную и достоверную информацию о состоянии и тенденциях изучаемого социально-экономического явления или процесса.

1. Объект статистического наблюдения — это:
   а) единица наблюдения;
   б) статистическая совокупность;
   в) единица статистической совокупности;
   г) совокупность признаков изучаемого явления.

2. Инструментарий статистического наблюдения содержит:
   а) инструкцию;
   б) формуляр;
   в) инструкцию и формуляр;
   г) макет разработочных таблиц;
   д) нет точного ответа.

3. Ошибки статистического наблюдения бывают:
   а) только случайные;
   б) случайные и систематические;
   в) только ошибки репрезентативности.

4. Отчетной единицей выступает:
   а) единица наблюдения;
   б) единица совокупности;
   в) субъект, представляющий данные.

5. Программа статистического наблюдения включает:
   а) время наблюдения;
   б) критический момент;

в) способ и метод наблюдения;
г) систему признаков, подлежащих статистическому наблюдению.

6. Срок статистического наблюдения — это время, в течение которого:
   а) заполняются статистические формуляры;
   б) обучается кадровый состав для проведения наблюдения;
   в) обрабатывается полученный в ходе наблюдения материал.

7. Статистическая отчетность — это:
   а) вид статистического наблюдения;
   б) организационная форма статистического наблюдения;
   в) форма статистического наблюдения.

8. По времени регистрации фактов статистическое наблюдение бывает:
   а) специально организованное;
   б) единовременное;
   в) выборочное;
   г) непосредственное.

9. По охвату единиц совокупности статистическое наблюдение бывает:
   а) периодическое;
   б) в виде отчетности;
   в) документальное;
   г) монографическое.

10. Опрос предполагает использование в качестве источника информации:
   а) различные документы;
   б) слова респондентов;
   в) штат добровольных корреспондентов;
   г) анкеты.

11. При методе основного массива обследованию подвергаются:
   а) все единицы совокупности;
   б) самые существенные, наиболее крупные единицы совокупности, имеющие по основному признаку наибольший удельный вес в совокупности;
   в) самые существенные, наиболее мелкие единицы совокупности, имеющие по основному признаку наименьший удельный вес в совокупности;
   г) отдельные единицы совокупности, представители новых типов явлений.

12. Монографическое обследование предполагает, что обследованию подвергаются:
   а) все без исключения единицы совокупности;
   б) самые существенные, наиболее крупные единицы совокупности, имеющие по основному признаку наибольший удельный вес в совокупности;
   в) отдельные единицы совокупности, представители новых типов явлений.

13. Ошибки регистрации возникают:
   а) только при сплошном наблюдении;
   б) только при несплошном наблюдении;
   в) как при сплошном, так и при несплошном наблюдении.

14. Ошибки репрезентативности возникают:
   а) только при сплошном наблюдении;
   б) только при несплошном наблюдении;
   в) как при сплошном, так и при несплошном наблюдении.

15. Перепись населения России — это:
   а) единовременное, специально организованное, сплошное наблюдение;
   б) периодическое, специально организованное, сплошное наблюдение;
   в) периодическое, регистровое, сплошное наблюдение;
   г) периодическое, специально организованное, несплошное наблюдение.

16. Инвентаризация основных средств на предприятии — это:
   а) текущее наблюдение;
   б) периодическое наблюдение;
   в) единовременное обследование.

17. Расхождение между расчетными и действительными значениями изучаемых величин называется:
   а) ошибкой наблюдения;
   б) ошибкой регистрации;
   в) ошибкой репрезентативности.

18. Сформулируйте объект, единицу и цель наблюдения и разработайте программу:
   а) обследования страховых компаний;
   б) обследования предприятий общественного питания;
   в) обследования промышленной фирмы.

19. Предполагается провести единовременное обследование коммерческих банков России. Каким из известных вам способом следовало бы статистическим органам провести это обследование? Мотивируйте ваш ответ.

20. Проверьте с помощью счетного контроля следующие данные: всего учащихся в гимназии — 850, в том числе: в 1—3 классах — 300, в 4—8 классах — 388, в 9—11 классах — 170. Из всего числа учащихся: мальчиков — 418, девочек — 442.

В результате первой стадии статистического исследования — стадии статистического наблюдения — исследователь получает сведения о каждой единице анализируемой совокупности. Эти сведения характеризуют ее с различных сторон, поскольку обладают многочисленными признаками и свойствами, изменяющимися во времени и пространстве. Возникает необходимость в систематизации и обобщении результатов статистического наблюдения для получения сводной характеристики всего объекта при помощи обобщающих показателей. Это дает возможность выявить характерные особенности, специфические черты статистической совокупности в целом и отдельных ее составляющих, обнаружить закономерности изучаемых социально-экономических явлений и процессов. Такую систематизацию называют сводкой первичного статистического материала.

Статистическая сводка является очередным, вторым, этапом статистического исследования социально-экономических явлений и процессов. ***Статистическая сводка*** — это обработка первичных данных в целях получения обобщенных характеристик изучаемого явления или процесса по ряду существенных для него признаков для выявления типичных черт и закономерностей, присущих явлению или процессу в целом. По глубине и точности обработки материала различают сводку простую и сложную.

Простая статистическая сводка — это операция по подсчету общих итоговых и групповых данных непосредственно по совокупности единиц наблюдения и оформление этого материала в таблицах. ***Сложная статистическая сводка*** — это комплекс операций, включающих распределение единиц наблюдения изучаемого социально-экономического явления или процесса на группы, составление системы показателей для характеристики типичных групп и подгрупп изучаемой совокупности явлений, подсчет числа единиц и итогов в каждой группе и подгруппах и оформление результатов этой работы в виде статистических таблиц.

Статистическая сводка проводится на основе всестороннего теоретического анализа сущности и содержания изучаемых явлений и процессов. Полнота, достоверность и обоснованность результатов сводки обеспечиваются программой и планом ее проведения.

Программа статистической сводки содержит перечень групп, на которые может быть разбита или разбивается совокупность единиц наблюдения, а также систему показателей, характеризующих изучаемую совокупность явлений и процессов как в целом, так и отдельных ее частей.

Программа статистической сводки зависит от целей и задач исследования. Ее разработка включает следующие этапы. Прежде всего выбирают группировочный признак для образования однородных групп. Затем определяют порядок формирования и число групп. Далее разрабатывают систему статистических показателей для характеристики групп и объекта в целом. Завершается работа над программой созданием макетов статистических таблиц для представления результатов сводки.

Наряду с программой статистической сводки составляется план ее проведения. ***План статистической сводки*** содержит информацию о последовательности, сроках и технике проведения сводки, ее исполнителях, о порядке и правилах оформления ее результатов в виде таблиц. По форме обработки материала сводка бывает децентрализованной и централизованной.

Децентрализованная статистическая сводка — это специфический способ организации сводки статистических данных. Он состоит в том, что обработка таких данных производится на местах, т. е. отчеты предприятий сводятся статистическими органами субъектов Российской Федерации. Полученные итоги поступают в Росстат, а там выводятся итоговые показатели в целом по экономике страны.

Централизованная статистическая сводка — это способ организации сводки статистических данных, при котором все первичные данные, полученные в результате статистического наблюдения, сосредотачиваются в одной, как правило, центральной организации и подвергаются в ней обработке от начала до конца.

По технике выполнения статистическая сводка бывает ***автоматизированная*** (с использованием электронно-вычислительной техники) и ***ручная***.

Простая сводка статистических данных без распределения единиц совокупности на группы по определенным существенным признакам не позволяет получить полную, достоверную характеристику изучаемого объекта. Недостаточно, скажем, ограничиться знанием общей численности студентов вуза. Для организации учебного процесса важно знать распределение студеитов по факультетам, специальностям, успеваемости, посещаемости и т. д. Для исследования представляет интерес знание не только всей совокупности в целом, но и отдельных ее частей, групп. Обобщение данных об изменениях, происходящих в группах и подгруппах изучаемых социально-экономических явлений и процессов, дает возможность получить картину состояния и характера развития объекта в целом. Это значит, что статистическая группировка является одним из основных этапов проведения статистического исследования.

Статистическая группировка — это разделение общей совокупности единиц объекта наблюдения по одному или нескольким существенным признакам на однородные группы, различающиеся между собой в качественном и количественном отношении и позволяющие выделить социально-экономические типы явлений, изучить структуру совокупности или проанализировать взаимосвязи и взанмозависимости между признаками. Иными словами, группировки являются важнейшим статистическим методом обобщения статистических данных, основой для правильного исчисления статистических показателей.

С помощью метода группировок решаются три основные задачи: во-первых, выделение социально-экономических типов явлений; во-вторых, изучение структуры явления и структурных сдвигов, происходящих в нем; в-третьих, выявление взаимосвязей и взаимозависимостей между явлениями и признаками, характеризующими этн явления. В соответствии с этими задачами различают следующие виды статистических группировок: типологические, структурные, аналитические.

Типологическая группировка — это расчленение разнородной совокупности на отдельные качественно одиородные группы, социально-экономические классы и выявление на этой основе экономических типов явлений. При построении группировки данного вида основное внимание должно быть уделено идеитификации 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, 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 группировочных признаков могут выступать как количественные, так и атрибутивные (качественные) признаки.

Группировка по атрибутивному признаку предполагает, что число выделенных групп строго соответствует фактическому числу градаций этого признака. Рассмотрим этот вопрос на данных табл. 3.1. Из нее видно, что группировка музеев произведена по видам по классификации ЮНЕСКО (искусствоведческие, исторические и археологические, краеведческие, естественно-научные, научно-технические, комплексные, отраслевые, специализированные и пр.).

При построении типологической группировки ***по количественному признаку*** необходимо правильно установить интервал группировки, определить необходимое число групп. Границы интервалов устанавливаются таким образом, чтобы внутри каждой выде-

Вид музея	Число музеев, ед.	Удельный вес музеев, % к итогу
Искусствоведческие	309	12,17
Исторические и археологические	532	20,95
Краеведческие	1 294	50,96
Естественно-научные	39	1,54
Научно-технические	20	0,79
Комплексные	190	7,49
Отраслевые, специализированные и прочие	155	6,10
Итого	2 539	100,00

ленной группы единицы совокупности были однородными в качественном и количественном отношении, а сами группы при этом существенно отличались друг от друга. Другими словами, проблема определения интервалов типологической группировки решается на основании выделения таких количественных границ изменения группировочного признака, при которых явление измеияет или приобретает новое качество. Так, при анализе успеваемости студентов учебного заведения в зимнюю сессию по предмету возможна типологическая группировка на успевающих и неуспевающих. В качестве группировочного признака в данном случае выступает успеваемость студентов. Уровнем варьирующего признака, который разделяет эти два типа студентов, будет удовлетворнтельная оценка «3». Студенты, которые не получили эту оценку, составили один тип, а студеиты, получившие оценки «удовлетворительно» и выше, — другой тип.

Число групп в типологической группировке зависит от числа реально существующих социально-экономических типов. Социально-экономические типы явлений, выделенные типологической группировкой, изучаются с точки зрения их состава, структуры типически однородных групп и изучения вариации признаков внутри однотипной совокупностн и однотипных групп на основе построения структурной группировки.

Структурная группировка — это расчленение однородной в качественном отношении совокупностн единиц по определенным, существенным признакам на группы, характеризующие ее состав и структуру. Структурные группировки применяются практически в изучении всех социально-экономических процессов и явлений. В качестве группировочных признаков, так же как и при построении типологической группировки, могут рассматриваться количественные и атрибутивные признаки.

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. 3.2.

Структурная группировка по ***количественному признаку*** также предполагает определение числа групп и ширину интервала (более подробно это рассмотрено в разд. 3.3). Пример структурной группировки по количественному признаку представлен в табл. 3.3.

Применение структурных группировок позволяет на локальном уровне раскрыть структуру совокупности, проанализировать изучаемые явления и процессы, изменение их во времени и зако-

Отделения	Число образовательных учреждений		Численность обучающихся		Подготовлено квалифицированных рабочих	
	Всего, ед.	% к итогу	Всего, тыс. чел.	% к итогу	Всего, тыс. чел.	% к итогу
Отделения на базе основного общего образования	1 940	44,15	740	77,65	270	63,23
Отделения на базе среднего (полного) общего образования	1 407	32,02	120	12,60	103	24,12
Группы, в которых молодежь не получает среднего (полного) общего образования	1 047	23,83	93	9,75	54	12,65
Итого	4 394	100,00	953	100,00	427	100,00

Группы городов и поселков городского типа по числу жителей, тыс. чел.	Число городов и поселков городского типа		Численность населения	
	Всего, ед.	% к итогу	Всего, тыс. чел.	% к итогу
До 3,0	349	14,58	604	0,58
3,0—4,9	315	13,16	1 257	1,21
5,0—9,9	561	23,43	4 002	3,86
10,0—19,9	473	19,76	6 582	6,35
20,0—49,9	375	15,66	12 057	11,63
50,0—99,9	157	6,56	10 928	10,54
100,0—499,9	129	5,39	27 085	26,12
500,0—999,9	24	1,00	15 430	14,88
1 000,0 и более	11	0,46	25 760	24,83
Итого	2 394	100,00	103 705	100,00

номерности изменения состава совокупности во времени, если они прослеживаются за ряд последовательных периодов времени.

Одна из основных задач статистических группировок состоит в исследовании связей и зависимостей между признаками единиц статистической совокупности, которая решается с помощью построения аналитических группировок. ***Аналитическая группировка*** — это группировка, выявляющая взаимосвязи и взаимозависимости между изучаемыми социально-экономическими явлениями и признаками, их характеризующими. В статистике все признаки делятся на факторные и результативные. ***Факторные признаки*** — это признаки, которые оказывают влияние на изменение результативных признаков. ***Результативные признаки*** — это признаки, которые изменяются под влиянием факторных признаков. Взаимосвязь проявляется в том, что с возрастанием роли факторного признака и под его влиянием более интенсивно изменяется результативный признак.

Особенности аналитической группировки состоят в том, что единицы совокупности группируются по факторному признаку, а расчет групповых средних производится по значениям результативного признака. Иначе говоря, каждая выделенная группа характеризуется средними величинами результативного признака. По изменению этих величин и определяется наличие связей и зависимостей между признаками.

Как и в случае со структурной группировкой, важной задачей при построении аналитической группировки является выбор числа групп, на которые необходимо разбить изучаемую совокупность единиц наблюдения, и определение их границ. Это определяется целью, задачами, условиями конкретного исследования, спецификой изучаемого социально-экономического явления или процесса.

Процесс построения аналитических группировок предполагает соблюдение определенных требований. Назовем два основных. Каждая выделенная группа должна содержать статистически однородные единицы совокупности по группировочному признаку. Количество единиц совокупности в каждой выделенной группе должно быть достаточным для получения типичных, надежных статистических характеристик изучаемого явления или процесса. В качестве примера аналитической группировки рассмотрим материалы табл. 3.4.

Все рассмотренные группировки объединяет то обстоятельство, что единицы объекта разделены на группы по какому-либо общему признаку. В одном случае — это вид музея (см. табл. 3.1), в другом — отделений образовательных учреждений (см. табл. 3.2), в тре-

Группы организаций по коэффициенту текущей ликвидности	Число организаций, ед.	Кредиторская задолженность, млн руб.		Дебиторская задолженность, млн руб.	
		Всего	В среднем на одну организацию	Всего	В среднем на одну организацию
0—100	28 036	7 011 555	250,09	3 749 744	133,75
101—200	26 527	7 673 865	289,29	7 014 197	264,42
Свыше 200	24 064	2 820 840	117,22	5 973 338	248,23
Итого	78 627	17 506 260	222,65	16 737 279	212,87

тьем — число жителей (см. табл. 3.3), в четвертом — коэффициент текущей ликвидности (см. табл. 3.4). Каждый раз группы образованы по одиому признаку. Такая группировка называется ***простой***.

Комбинационной считается группировка, когда разбиение совокупности на группы производится по двум и более группировочным признакам, взятым в сочетании (комбинации) друг с другом.

Номер группы	Группы областей по числу малых предприятий, ед.	Подгруппы областей по величине оборота малых предприятий, млрд руб.	Число областей, ед.	Средняя численность работников малых предприятий, тыс. чел.	Оборот малых предприятий, млрд руб.	Число малых предприятий, тыс. ед.
1	2	3	4	5	6	7
1	7,0—15,3	45,6—99,2	10	690,4	748,5	89,4
		99,2—152,8	1	86,7	117,1	11,0
		Итого	11	777,1	865,6	100,4
2	15,3—23,6	45,6—99,2	2	234,4	171,0	32,7
		99,2—152,8	3	373,5	414,0	59,0
		Итого	5	607,9	585,0	91,7
Итого по подгруппам		45,6—99,2	12	924,8	919,5	122,1
		99,2—152,8	4	460,2	531,1	70,0
		Всего	16	1 385,0	1 450,6	192,1

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. Комбинационные группировки позволяют изучать единицы совокупности одновременно по нескольким признакам. При построении комбинационной группировки возникает вопрос о последовательности разбиения 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 этого вида группировок является наличие достаточно большого числа наблюдений. Особенно актуально это условие при исследовании зависимости результативных признаков от двух и более факторных. 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 придет к малообоснованным выводам.

Примером комбинационной группировки является табл. 3.5.

Построение статистических группировок предполагает решение ряда основных задач. Прежде всего следует 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 существенных и сложных вопросов теории статистической группировки и статистического исследования в целом. ***Группировочный признак*** — это основание, по которому проводится разбиение единиц совокупности на отдельные группы. От выбора группировочного признака зависит правильность выводов статистического исследования. В качестве основания группировки необходимо использовать наиболее существенные, теоретически обоснованные признаки, отражающие сущность изучаемых социально-экономических явлений и процессов в условиях поставленных целей и задач.

В основание группировки могут быть положены как количественные, так и атрибутивные (качественные) признаки. ***Количественные признаки*** обычно имеют числовое выражение (например, объем выпускаемой продукции, возраст человека, доход семьи и т. д.). ***Атрибутивные признаки*** не предполагают цифрового выражения. Они дают качественную характеристику единицы совокупности. Таковы, например, пол, семейное положение, форма собственности и т. д.

Построение группировки по атрибутивному признаку требует безошибочного определения градаций, на основе которых устанавливается принадлежность изучаемых явлений и процессов к определенному типу. Выделенные группы должны отличаться друг от друга не по количественной, а исключительно по качественной характеристике признака. Число групп, на которые разбивается изучаемая совокупность, как правило, зависит от числа градаций атрибутивного признака. Так, группировка населения по полу допускает образование только двух групп — мужская и женская. Группировка промышленных предприятий по формам собственности предполагает образование стольких групп, сколько существует ее различных форм, но не более числа видов установленных форм собственности.

При построении группировки по количественному признаку необходимо тщательно изучить экономическую (социальную) сущность исследуемого явления или процесса. Лишь после этого в соответствии с задачами исследования можно решать вопрос о числе групп, близких по значению к варьирующему признаку единиц совокупности. Определив таким образом основание группировки, приступают к решению проблемы числа групп, на которое нужно разбить исследуемую совокупность.

Следует помнить, что число групп во многом зависит от задач исследования и вида показателя, положенного в основание группировки, от объема совокупности (числа единиц) исследуемого объекта, от вариативности группировочного признака. Количество групп должно быть достаточным, чтобы четко и определенно выявились характер и особенности изучаемого явления. Плохо, если их мало, но также плохо, если их избыток. Так, при небольшом объеме совокупности нельзя образовывать большое число групп. Они будут включать недостаточное число единиц объекта. Поэтому показатели, рассчитанные при такой группировке, не будут представительными и типичными, не позволят получить адекватную характеристику исследуемого социально-экономического явления.

Число групп зависит от степени вариативности группировочного признака: чем она выше, тем больше можно образовать групп.

Чтобы не встать на ошибочный путь, нужно внимательно присмотреться к размаху колебаний признака, который определяется как разность между его максимальным и минимальным значениями. Важно постоянно учитывать, что чем больше размах варьирования признака, положенного в основание группировки, тем больше может быть образовано групп, но тем больше и вероятиость построения «пустых» групп.

С одной стороны, чем больше образовано групп, тем точнее воспроизводится характер исследуемого объекта. С другой стороны, как уже говорилось, слишком большое число групп затрудняет выявление закономерностей при исследовании социально-экономических явлений и процессов. В них растворяются особенности, выражающие взаимосвязь и взаимовлияние признаков. При неоправданно малом числе групп в одну и ту же группу попадут единицы совокупности с различным выражением особенности. Поэтому в каждом конкретном случае при определении числа групп следует исходить не только из степени вариации признака. Важно учитывать особенности объекта, а главное — цели и задачи исследования.

В практической работе для определения числа групп можно воспользоваться формулой Стерджесса:

$$n = 1 \quad 3{,}322 \lg N, \tag{3.1}$$

где n — число групп; N — число единиц совокупности; $\lg N$ — десятичный логарифм от N.

Данная формула подтверждает, что выбор числа групп зависит от объема совокупности. Недостаток состоит в том, что ее применение дает хорошие результаты лишь в тех случаях, когда совокупность состоит из большого числа единиц, а 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 предиолагается равным. Значит, если величина интервала равна $0{,}5\sigma$, то совокупность разбивается на 12 групп, а когда она равна $2/3\sigma$ или σ, то совокупность делится, соответственно, на 9 и 6 групп. Однако при определении числа групп данным методом существует большая вероятность получения «пустых» или малочисленных групп. «Пустыми» называются группы, которые не содержат ни одной единицы совокупности. Поэтому данными формулами нельзя пользоваться механически. Их показания требуют корректировки.

После того как установлено число групп, решается задача определения интервалов группировки. ***Интервал группировки*** — это интервал значений варьирующего признака, лежащих в определенных пределах. Каждый интервал имеет свою длину (ширину), верхнюю и нижнюю границы или хотя бы одну из них.

Нижняя граница интервала — это наименьшее значение признака в интервале, а ***верхняя граница*** интервала — его наибольшее значение. За нижнюю границу первого интервала принимается, как правило, наименьшее значение признака в совокупности единиц наблюдения. Верхняя граница последиего интервала не может быть меньше наибольшего значения признака в совокупности единиц наблюдения.

Ширина интервала — это разность между верхней и нижней границами.

Интервалы группировки в зависимости от их ширины 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 равномерный характер, то строят группировку с ***равными интервалами***.

Величина равного интервала определяется по следующей формуле:

$$h = \frac{R}{n} = \frac{x_{max} - x_{min}}{n}, \qquad (3.2)$$

где x_{max}, x_{min} — соответственно максимальное и минимальное значения признака в совокупности; n — число групп.

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

Если размах вариации признака в совокупности велик и значения признака варьируют неравномерно, то надо использовать группировку с ***неравными интервалами***. Неравные интервалы могут быть получены несколькими способами. В случае, если построенная группировка с равными интервалами содержит группы, не отражающие определенные типы изучаемого явления или процесса или не содержащие ни одной еднницы совокупности, возникает необходимость увеличения — объединения двух или нескольких малочисленных или «пустых» последовательных равных интервалов. Другим способом построения неравных иитервалов является применение прогрессивно-возрастающих и прогрессивно-убывающих иитервалов. В основе их построения лежит принцип арифметической или геометрической прогрессии. Величина изменяющихся интервалов определяется по следующим формулам:

$$\begin{aligned} &\text{арифметической прогрессии} \quad h_{i+1} = h_i + a; \\ &\text{геометрической прогрессии} \quad h_{i+1} = h_i q, \end{aligned} \tag{3.3}$$

где h_i — величина i-го интервала; a — константа (для прогрессивно-возрастающих интервалов имеет знак «плюс», а при прогрессивно-убывающих — знак «минус»); q — константа (она больше 1 — для прогрессивно-возрастающих и меньше 1 — в противоположном случае).

Применение неравных интервалов обусловлено спецификой и особенностями изучаемого социально-экономического явления или процесса, когда в первых группах небольшая разница в показателях имеет большое значение, а в последних группах эта разница существенного значения не имеет. Например, при построении группировки предприятий связи по показателю численности работников, который варьирует от 200 чел. до 2 000 чел., нецелесообразно применять равные интервалы. Ведь совокупность представлена как малыми, так и крупнейшими предприятиями. Поэтому следует образовывать неравные интервалы такие, как 200—500, 500—1 100, 1 100—2 000. В нашем примере величина каждого последующего интервала больше предыдущего на 300 чел., т.е. увеличивается в арифметической прогрессии.

Решение вопроса о выборе равных или неравных интервалов зависит от числа единиц совокупности, попавших в каждую выделенную группу, т.е. от степени заполнения интервалов.

Интервалы группировок могут быть закрытыми и открытыми. ***Закрытыми интервалами*** называют такие, в которых указана верхняя и нижняя границы. ***Открытые интервалы*** имеют только одну границу (верхнюю — у первого, нижнюю — у последнего). Например, группы коммерческих банков по числу работающих в них сотрудников, чел.: до 200, 200—300, 300—400, 400 и более.

При группировке единиц совокупности по количественному признаку границы интервалов могут быть обозначены по-разному, в зависимости от того, непрерывный это признак или дискретный.

Допустим, основанием группировки служит ***непрерывный признак*** (например, группы строительных фирм по величине прибыли, тыс. руб.: 4 200—4 400, 4 400—4 600, 4 600—4 800, 4 800—5 000). В этом случае одно и то же значение признака выступает и верхней и нижней границами двух смежных интервалов. Так, величина прибыли 4 400 тыс. руб. составляет верхнюю границу первого интервала и нижнюю границу второго, 4 600 тыс. руб. — соответственно второго и третьего. Следовательно, верхняя граница i-го интервала равна нижней границе $(i + 1)$-го интервала.

При таком обозначении границ может возникнуть вопрос, в какую группу включать единицы совокупности, значения признака у которых совпадают с границами интервалов. Например, во вторую или третью группу должна войти строительная фирма с прибылью 4 600 тыс. руб.? Если верхняя граница формируется по принципу «исключительно», то фирма должна быть отнесена к третьей группе, а если «включительно» — то ко второй.

Чтобы правильно отнести к той или иной группе единицу совокупности, значение признака которой совпадает с границами интервалов, можно использовать открытые интервалы (по нашему примеру группы строительных фирм по прибыли преобразуются в следующие: до 4 400, 4 400 — 4 600, 4 600 — 4 800, 4 800 и более). В данном случае вопрос отнесения отдельных единиц совокупности, значения которых являются граничными, к той или иной группе решается на основе анализа последнего открытого интервала. Возможны два случая обозначения последнего открытого интервала: 1) 4 800 тыс. руб. и более; 2) более 4 800 тыс. руб. В первом случае строительные фирмы с величиной прибыли 4 600 тыс. руб. попадут в третью группу; во втором случае — во вторую группу.

Если в основании группировки лежит ***дискретный признак***, то нижняя граница i-го интервала равна верхней границе $(i - 1)$-го интервала, увеличенной на 1. Например, группы страховых компаний по числу занятого персонала (чел.) будут иметь вид: 100 — 150, 151 — 200, 201 — 300.

При определении границ интервалов статистических группировок иногда 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. В группировках, имеющих целью отобразить качественные особенности и специфику выделяемых групп единиц изучаемой совокупности по признаку, применяются специализированные интервалы. ***Специализированные интервалы*** — это интервалы, которые применяются для выделения из совокупности одних и тех же типов по одному и тому же признаку у явлений, находящихся в различных условиях.

При изучении социально-экономических явлений на макроуровне часто применяют группировки, где интервалы не бывают ни прогрессивно-возрастающими, ни прогрессивно-убывающими. Такие интервалы называются ***произвольными***. Как правило, они используются при группировке предприятий по уровню рентабельности, прибыльности и др.

Рассмотрим комплексный пример построения всех перечисленных в данном подразделе видов статистических группировок на основе данных, представленных в табл. 3.6.

№ п/п	Субъект	Число малых предприятий, тыс.	Средняя численность работников малых предприятий, тыс. чел.	Оборот малых предприятий, млрд руб.
1	Белгородская область	16,3	82,4	77,4
2	Брянская область	7,1	71,9	74,4
3	Владимирская область	16,4	152,0	93,6
4	Воронежская область	18,7	149,6	152,8
5	Ивановская область	8,7	74,2	94,9
6	Калужская область	11,0	86,7	117,1
7	Костромская область	7,1	54,8	45,7
8	Курская область	8,7	67,2	64,9
9	Липецкая область	10,1	73,9	98,9
10	Орловская область	7,0	48,1	45,6
11	Рязанская область	11,4	78,5	97,0
12	Смоленская область	9,9	67,5	74,3
13	Тамбовская область	7,1	65,8	72,8
14	Тверская область	12,3	88,5	80,0
15	Тульская область	16,7	112,1	118,2
16	Ярославская область	23,6	111,8	143,0
17	Республика Карелия	8,7	48,1	60,0
18	Республика Коми	10,6	82,4	100,2
19	Архангельская область	11,1	64,7	80,7
20	Вологодская область	6,4	73,6	77,6
21	Калининградская область	22,2	109,4	246,0
22	Ленинградская область	15,5	137,0	110,8
23	Мурманская область	8,5	42,8	99,5
24	Новгородская область	5,7	45,6	37,5
25	Псковская область	4,3	45,6	39,4

* Из рассмотрения исключены показатели по Москве, Санкт-Петербургу и Московской области.

Группы субъектов по числу малых предприятий, тыс. ед.	Число субъектов	Число малых предприятий, тыс.	Средняя численность работников малых предприятий, тыс. чел.	Оборот малых предприятий, млрд руб.
А	1	2	3	4
4,3—9,2	11	79,3	637,7	712,3
9,2—14,1	7	76,4	542,2	648,2
14,1—19,0	5	83,6	633,1	552,8
19,0—23,9	2	45,8	221,1	389,0
Итого	25	285,1	2 034,2	2 302,3

В качестве группировочного признака взято число малых предприятнй. Поскольку объем изучаемой совокупности единиц наблюдения невелик — всего 25 единиц, представляется целесообразным строить небольшое число групп. Образуем четыре группы субъектов с равными интервалами. Величину интервала определим по формуле

$$h = \frac{x_{\max} - x_{\min}}{n} = \frac{23{,}6 - 4{,}3}{4} = 4{,}9.$$

Обозначим границы групп: 4,3—9,2 — 1-я группа; 9,2—14,1 — 2-я группа; 14,1—19,0 — 3-я группа; 19,0—23,9 — 4-я группа.

После того как определен группировочный признак — число малых предприятий, задано число групп — 4 и образованы сами группы, необходимо отобрать показатели, которые характеризуют группы, и определить их величины по каждой группе. Показатели, характеризующие субъекты РФ, разносятся по четырем указанным группам и подсчитываются групповые итоги. Результаты группировки заносятся в таблицу и определяются общие итоги в абсолютных величинах по совокупности единиц наблюдения по каждому показателю (табл. 3.7).

Теперь эти абсолютиые показатели пересчитываем в «проценты к нтогу». Таким образом получаем табл. 3.8.

Из табл. 3.8 видно, что в основном преобладают субъекты с числом малых предприятий от 4,3 до 9,2 тыс. Их доля составляет 44 %. На долю малых предприятий данной группы приходится 31,3 % средней численности работников всех малых предприятий и 30,9 % оборота. Субъекты Центрального и Северо-Западного федераль-

Группы субъектов по числу малых предприятий, тыс.	Удельный вес субъектов	Структура числа малых предприятий	Структура средней численности работников малых предприятий	Структура оборота малых предприятий
4,3—9,2	44	27,8	31,3	30,9
9,2—14,1	28	26,8	26,7	28,2
14,1—19,0	20	29,3	31,1	24,0
19,0—23,9	8	16,1	10,9	16,9
Итого	100	100,0	100,0	100,0

ных округов распределены по числу малых предприятий не равномерно. Так субъекты по числу малых предприятий от 4,3 до 9,2 тыс. и от 1,1 до 19,0 тыс. имеют приблизительно одинаковый удельный вес численности работников: 31,1—31,3 %, при том, что и удельный вес малых предприятий в них не сильно различается: 27,8 % и 29,3 %.

Более конкретный анализ взаимосвязи показателей можно сделать на основе аналитической группировки (табл. 3.9).

Группы субъектов по числу малых предприятий, тыс.	Число субъектов	Число малых предприятий, тыс.		Средняя численность работников малых предприятий, тыс. чел.		Оборот малых предприятий, млрд руб.	
		Всего	В среднем на один субъект	Всего	В среднем на один субъект	Всего	В среднем на один субъект
4,3—9,2	11	73,9	7,2	637,7	58,0	712,3	64,8
9,2—14,1	7	76,4	10,9	542,2	77,5	648,2	92,6
14,1—19,0	5	83,6	16,7	633,1	126,6	552,8	110,6
19,0—23,9	2	45,8	22,9	221,2	110,6	389,0	194,5
Итого	25	285,1	—	2 034,2	—	2 302,3	—
В среднем на один субъект	—	—	11,4	—	81,4	—	92,1

Анализ табл. 3.9 свидетельствует о том, что число малых предприятий, их средняя численность работников и оборот находятся в прямой зависимости: чем крупнее субъект по числу малых предприятий, тем больше средняя численность работников и тем больше оборот. Исключение составляет группа субъектов с числом малых предприятий от 19,0 до 23,9 тыс., для которой характерны большие обороты при меньшей численности работников.

Мы рассмотрели примеры группировок по одному признаку. Однако в ряде случаев для решения поставленных задач такая группировка является недостаточной. В этих случаях переходят к группировке исследуемой совокупности по двум и более существенным признакам в их взаимосвязи, т. е. к комбинационной группировке.

В результате обработки и систематизации первичных данных статистического наблюдения получают простейший вид группировки, называемый рядом распределения. В них известна численность единиц наблюдения в группах, представленная в абсолютиом и относительном выражении.

Ряд распределения — это упорядоченные по определенному варьирующему признаку однородные группы единиц совокупности. В зависимости от признака, положенного в основание построения ряда распределения, различают атрибутивные и вариационные ряды распределения.

Атрибутивный ряд распределения — это ряд распределения, построенный по качественным признакам, не имеющим числового выражения и характеризующим свойство, качество изучаемого социально-экономического явления. Атрибутивные ряды распределения характеризуют состав совокупности по тем или иным существенным признакам. Взятые за несколько периодов, этн данные позволяют исследовать изменение структуры. Число групп атрибутивного ряда распределения адекватно числу градаций, разновидностей атрибутивного признака (табл. 3.10).

Элементами данного ряда распределения являются градации атрибутивного признака «Успеваемость» («успевают» — «не успевают») и численность каждой группы в абсолютном (человек) и относительном (%) выражении. Студентов, сдавших экзамен по статистике, было 46 чел. Их удельный вес составил 92 %.

Успеваемость	Число студентов, чел.	Удельный вес в общей численности студентов, %
Успевают	46	92
Не успевают	4	8
Итого	50	100

Вариационный ряд распределения строится по количественному признаку. Любой такой ряд состоит из вариантов числовых значений количественного признака, т. е. частоты (численности) отдельных вариантов или каждой группы вариационного ряда. Этн числа показывают, как часто встречаются те или иные варианты (значения признака) в ряду распределения. Сумма всех частот определяет численность всей совокупности.

Численности групп могут быть выражены как в абсолютиых величинах, т. е. числом единиц совокупности в каждой выделенной группе, так и в относительных величинах — в виде долей, удельных весов, представленных в процентах к итогу. Частость — это отношение численности группы к общей численности, выраженное в отиосительных еднницах или процентах к итогу. Соответственно, сумма частостей равна 1, если они выражены в долях единицы, или 100 %, если они выражены в процентах.

В зависимости от характера вариации признака различают дискретные и интервальные вариационные ряды распределения.

Дискретный вариационный ряд распределения — это ряд, в котором группы составлены по признаку, изменяющемуся дискретио и принимающему только целые значения.

Примером дискретного вариационного ряда распределения является распределение студентов по оценке, полученной на экзаменах (табл. 3.11).

В гр. 1 табл. 3.11 представлены варианты дискретного вариационного ряда, в гр. 2 — частоты, а в гр. 3 — частости. В случае непрерывной вариации величина признака у единиц совокупности может принимать в определенных пределах любые значения, отличающиеся друг от друга на сколь угодно малую величину.

Интервальный вариационный ряд распределения — это ряд, в котором группировочный признак, составляющий основание группировки, может принимать в определенном иитервале любые значения. Интервальный ряд распределения целесообразно строить

Экзаменационный балл	Число студентов, чел.	Удельный вес студентов, % к итогу
1	2	3
5	16	32
4	23	46
3	7	14
2	4	8
Итого	50	100

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. С целью проведения сравнительного анализа заполненности интервалов определяется показатель, характеризующий ***плотность распределения***. Это отношение числа единиц совокупности к ширине интервала.

Численность работающих, чел.	Число строительных фирм	Удельный вес, % к итогу
100—200	12	15,00
200—300	18	22,50
300—400	25	31,25
400—500	14	17,50
500—600	11	13,75
Итого	80	100,00

Примером интервального вариационного ряда распределения может служить табл. 3.12.

Представленный ряд распределения является интервальным, в основании образования групп которого лежит непрерывный признак.

Анализ рядов распределения можно для наглядности проводить на основе их графического изображения. Для этой цели строят полигон, гистограмму, кумуляту и огиву распределения.

Полигон используется при изображении дискретных вариационных рядов. Для его построения в прямоугольной системе координат по оси абсцисс в однаковом масштабе откладываются ранжированные значения варьирующего признака, а по оси ординат наносится шкала для выражения численности каждого варианта, т. е. величины частот. Полученные на пересечении абсцисс и орднат точки соединяются прямыми линиями, в результате чего получают ломаную линию, называемую полигоном частот (рис. 3.1). Иногда для замыкания полигона предлагается крайние точки (слева и справа на ломаной линии) соединить с точками на оси абсцисс, в результате чего получается многоугольник. Графически распределение студентов по экзаменационному баллу представлено на рис. 3.1 (см. табл. 3.11).

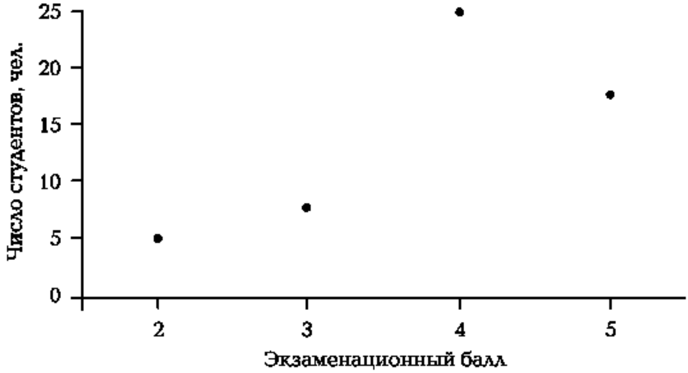


Рис. 3.1. Полигон распределения студентов по экзаменационному баллу

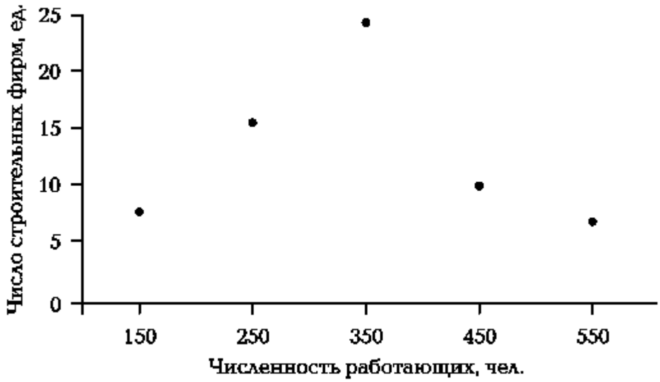


Рис. 3.2. Гистограмма распределения строительных фирм по среднесписочной численности работающих

Гистограмма применяется для изображения интервального вариационного ряда. При построении гистограммы на оси абсцисс откладываются величины интервалов, а частоты изображаются прямоугольниками, построенными на соответствующих интервалах. Высота столбиков должна быть пропорциональна частотам. В результате получается график, на котором ряд распределения изображен в виде столбиковой диаграммы (рис. 3.2). Гистограмма может быть преобразована в полигон распределения, если середины верхних сторон прямоугольников соединить прямыми.

При построении гистограммы распределения вариационного ряда с неравными интервалами по оси ординат наносят не частоты, а отношение частоты к ширине соответствующего интервала. Это необходимо сделать для устранения влияния величины интервала на распределение интервала и получения возможности сравнивать частоты. На рис. 3.2 представлена гистограмма распределения строительных фирм по среднесписочной численности работающих (см. табл. 3.12).

Для графического изображения вариационных рядов может использоваться кумулятнвная кривая. При помощи ***кумуляты*** изображается ряд накопленных частот (рис. 3.3). Накопленные частоты определяются путем последовательного суммирования частот по группам. Накопленные частоты показывают, сколько единиц совокупности имеют значения признака не больше, чем рассматриваемое значение.

При построении кумуляты интервального вариационного ряда по оси абсцисс откладываются варианты ряда, а по оси ординат на-

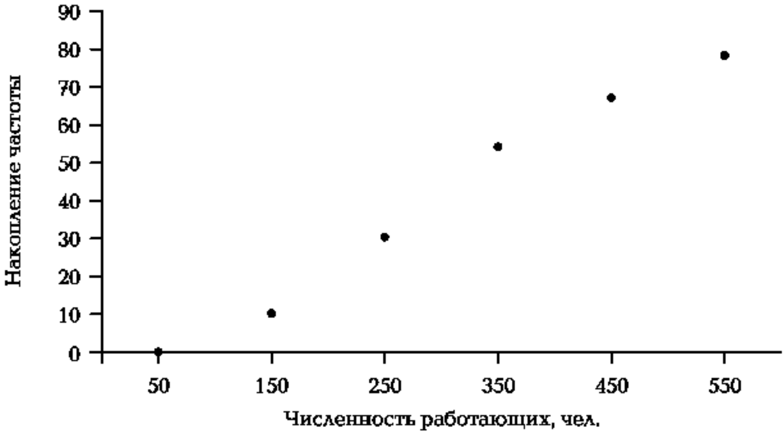


Рис. 3.3. Кумулята распределения строительных фирм по среднесписочной численности работающих

копленные частоты, которые наносят на поле графика в виде перпендикуляров к оси абсцисс в верхних границах интервалов. Затем эти перпендикуляры соединяют прямыми и получают ломаную линию, т. е. кумуляту.

Пример кумуляты, построенной по данным табл. 3.12, представлен на рис. 3.3.

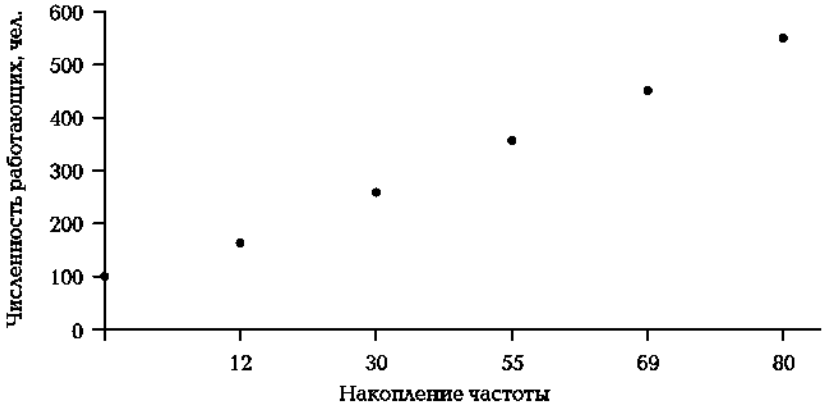


Рис. 3.4. Огива распределения строительных фирм по среднесписочной численности работающих

Если при графическом изображении вариационного ряда в виде кумуляты оси поменять местами, то получим ***огиву***. На рис. 3.4 приведена огива, построенная на основе данных табл. 3.12.

Графическое изображение рядов распределения позволяет наглядно представить распределение данных статистического наблюдения.

1. Группировка — это:
   а) упорядочение единиц совокупности по признаку;
   б) разбиение единиц совокупности на группы по признаку;
   в) обобщение единичных фактов.

2. Группировка, выявляющая взаимосвязи между явлениями и их признаками, называется:
   а) типологической;
   б) структурной;
   в) аналитической.

3. Группировка, в которой разнородная совокупность разбивается на однородные группы, называется:
   а) типологической;
   б) структурной;
   в) аналитической.

4. Группировка, построенная по двум признакам, называется:
   а) рядом распределения;
   б) простой;
   в) комбинационной.

5. Группировочным признаком при построении аналитической группировки выступает:
   а) факторный;
   б) результативный;
   в) факторный и результативный.

6. Основанием группировки может быть:
   а) качественный признак;
   б) количественный признак;
   в) как качественный, так и количественный признаки.

7. Ряд распределения, построенный по качественному признаку, называется:
   а) атрибутивным;
   б) дискретным;
   в) вариационным.

8. Вариационный ряд распределения — это ряд, построенный:
   а) по качественному признаку;
   б) по количественному признаку;
   в) как по качественному, так и по количественному признаку.
9. При непрерывной вариации признака целесообразно построить:
   а) атрибутивный ряд распределения;
   б) дискретный ряд распределения;
   в) интервальный ряд распределения.
10. Для изображения дискретных рядов распределения используется:
   а) полигон;
   б) гистограмма;
   в) кумулята.
11. Охарактеризуйте вид ряда распределения абитуриентов по результатам сдачи вступительных экзаменов:

Группы абитуриентов по результатам сдачи экзаменов	Число абитуриентов	Удельный вес, % к итогу
Не поступившие	100	28,6
Поступившие	250	71,4
Итого	350	100,0

   а) дискретный вариационный;
   б) интервальный вариационный;
   в) атрибутивный.
12. Охарактеризуйте вид ряда распределения коммерческих банков по величине работающих активов:

Величина работающих активов банка, млн руб.	Число банков	Удельный вес, % к итогу
До 150	7	10,9
150—250	10	15,6
250—350	32	50,0
350—450	11	17,2
450 и более	4	6,3
Итого	64	100,0

   а) дискретный вариационный;
   б) интервальный вариационный;
   в) атрибутивный.
13. Представлен макет статистической таблицы, характеризующий группировку промышленных предприятий по среднегодовой стоимости основных фондов:

Стоимость основных фондов предприятия, млн руб.	Число предприятий	Объем выпускаемой продукции, млн руб.		Численность промышленно-производственного персонала, чел.	
		Всего	В среднем на одно предприятие	Всего	В среднем на одно предприятие
100—120					
120—140					
140—160					
Итого					

Какой вид группировки отражает данный макет:

а) типологическую;
б) структурную;
в) аналитическую.

14. Распределение предприятий по тарифному разряду характеризуется следующими данными:

Тарифный разряд	Число рабочих	Удельный вес, в % к итогу
2	5	10
3	6	12
4	5	10
5	12	24
6	22	44
Итого	50	100

Определите вид ряда распределения:

а) интервальный вариационный;
б) дискретный вариационный;
в) атрибутивный.

15. Распределение строительных компаний по численности работающих характеризуется следующими данными:

Группы компаний по численности работающих, чел.	Число компаний
70—100	17
100—130	40
130 и более	23
Итого	80

Определите удельный вес числа строительных компаний в каждой выделенной группе:
   а) 17, 10; 40, 20; 23, 30;
   б) 21, 25; 50, 00; 28, 75;
   в) 23, 00; 17, 00; 40, 00.

16. Распределение фермерских хозяйств по размеру земельных угодий характеризуется следующими данными:

Группы фермерских хозяйств по размеру земельных угодий, тыс. га	Число фермерских хозяйств
до 40	21
40—60	13
60—80	9
80 и более	7
Итого	50

   В какую группу попали фермерские хозяйства с размером земельных угодий 40 тыс. га:
   а) в первую;
   б) во вторую;
   в) в третью.

17. Пользуясь формулой Стерджесса, определите интервал группировки сотрудников фирмы по уровню доходов, если общая численность сотрудников составляет 100 чел., а минимальный и максимальный доход соответственно равен 5 000 и 10 000 руб.

18. Известны следующие данные об объеме реализованной продукции 20 предприятиями (млн руб.):

42,31	42,75	42,97	53,67	52,12
32,82	42,37	43,01	53,22	43,65
42,45	42,84	52,46	42,77	47,99
42,84	42,55	43,12	53,86	49,00

   Постройте интервальный вариационный ряд распределения предприятий по объему реализованной продукции, предварительно выделив не более 4 групп.

19. Имеются следующие данные об успеваемости 30 студентов: 5, 4, 4, 5, 3, 3, 4, 4, 4, 5, 5, 2, 5, 4, 4, 2, 3, 3, 4, 4, 4, 5, 4, 5, 4, 4, 3, 4, 5, 5. Постройте дискретный ряд распределения студентов по баллам, полученным в сессию.

20. По данным задачи 19 постройте ряд распределения студентов по уровню успеваемости, выделив в нем две группы студентов: не успевающие и успевающие.

Статистические таблицы являются средством оформления результатов сводки и группировки, а также орудием анализа статистических данных и их графического представления. Таблица есть наиболее рациональная форма изложения и изображения результатов сводки и группировки, которая существенно облегчает их чтение и анализ. Без статистических таблиц пришлось бы сопровождать каждый показатель громоздкими пояснениями. С помощью же таблиц статистические материалы располагаются в определенном порядке, удобном для их сравнения между собой и для исчисления производных показателей. Чтобы соответствовать своему назначению, статистическая таблица должна быть по возможности небольшой, компактной и удобообозримой.

Наименование таблицы (общий заголовок)

Содержание строк	Наименование граф (верхние заголовки)						
А	1	2	3	4	5	• • • •	Нумерация ← граф
Наименование строк (боковые заголовки) **Подлежащее таблицы**				**Сказуемое таблицы**			← Строки таблицы ←
Итоговая строка						Итоговая графа	

Статистическую таблицу от других табличных форм отличают следующие особенности. Во-первых, в ней дается сводная характеристика единиц статистической совокупности, подводится один или несколько итогов. Во-вторых, она содержит результаты подсчета эмпирических данных. В-третьих, характеризуемые в ней объекты и показатели располагаются так, чтобы их наименование приводилось лишь однажды в виде общего заголовка.

В целом статистическая таблица представляет собой компактное и удобочитаемое изложение ряда экономических показателей, отражающих то или иное социально-экономическое явление или процесс. Составные части и элементы статистической таблицы, составляющие как бы ее остов, показаны на с. 64.

Основа (остов) статистической таблицы — это ряд взаимопересекающихся горизонтальных и вертикальных линий, образующих по горизонтали строки, а по вертикали — графы.

Остов таблицы, заполненный заголовками граф и строк, образует ***макет таблицы*** (табл. 4.1). Перед статистической сводкой необходимо составить макеты таблиц, которые являются выражением плана сводки. Если на пересечении граф и строк, образующих клетки, записать соответствующие цифры и показатели, то получится полная статистическая таблица.

Статистическая таблица содержит три вида заголовков: общий, верхние и боковые. ***Общий заголовок*** отражает содержание всей таблицы (к какому месту и времени она относится). Он располагается над ее макетом по центру и является внешним заголовком. ***Верхние заголовки*** характеризуют содержание граф (заголовки сказуемого), а ***боковые*** (заголовки подлежащего) — строк. Они являются внутренними заголовками. Как видим, по своему содержанию статистическая таблица напоминает грамматическое предложение:

№ п/п	Группы студентов по посещаемости занятий, %	Число студентов, чел.	Сумма баллов в зачетную сессию	Средний балл студента
	А	1	2	3
1	0—20			
2	20—40			
3	40—60			
4	60—80			
5	80—100			
	Итого			

в ней принято выделять подлежащее и сказуемое. Содержание таблицы заключается в системе суждений, выраженных не словами, а цифрами, и оно требует определенного порядка в расположении материала. Эти суждения касаются обобщенной характеристики групп и совокупностей единиц при помощи их численности и сводных характеристик их признаков.

Подлежащим в статистической таблице называется объект, который характеризуется цифрами. Это может быть одна или несколько совокупностей, отдельные единицы совокупностей (фирмы, объединения) в порядке их перечня, территориальные единицы или временны́е периоды, а также единицы, сгруппированные по каким-либо признакам. Обычно подлежащее таблицы дается в левой части, в наименовании строк. ***Сказуемое*** статистической таблицы образует система показателей, которыми характеризуется 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. Так, при большом перечне элементов подлежащего и небольшом сказуемого иногда целесообразно переставить местами подлежащее и сказуемое, чтобы получить характеристику групп более компактно. Что помещать в подлежащем, а что в сказуемом таблицы, зависит от характера материала, задач, поставленных перед статистической таблицей.

В зависимости от построения подлежащего статистические таблицы подразделяют на три вида: простые, групповые и комбинационные, а по разработке сказуемого — с простой и сложной разработкой. ***Простыми*** называются такие статистические таблицы, в подлежащем которых нет группировки, а дается перечисление территориальных единиц, единиц времени или какой-либо другой перечень. Простые таблицы имеют самое широкое распространение. Статистические данные по стране и отдельным ее зонам с перечислением в подлежащем всех районов, всех областей или республик, всех предприятий данного города можно представить в виде простых таблиц. Рассмотрим пример такой таблицы (табл. 4.2). Она по своему содержанию является простой (территориальной) таблицей, так как ее подлежащее содержит название стран, а не группировку данных.

Страны	Тыс. долл. США	% к итогу
Азербайджан	39 713	0,5
Армения	108 395	1,5
Белорусь	4 125 440	55,0
Казахстан	1 205 872	16,1
Киргизия	210 237	2,8
Республика Молдова	26 649	0,4
Таджикистан	182 846	2,4
Туркмения	1 616	0,0
Узбекистан	204 039	2,7
Украина	1 393 113	18,6
Итого	7 497 920	100

Еще один пример простой таблицы — табл. 4.3. Она по своему построению также принадлежит к числу простых, поскольку в подлежащем ее нет группировки, а дается перечень показателей, характеризующих медицинскую помощь населению.

Временну́ю характеристику развития явления также можно представить с помощью простой таблицы. При этом в подлежащем приводится перечень дат или периодов изменяющихся во времени данных.

Статистический анализ простой таблицы ограничивается параллельным сравнением приведенных показателей сказуемого. Более глубокий анализ простой таблицы невозможен в силу отсутствия причинной связи между показателями подлежащего и сказуемого. Комбинация признаков сказуемого не дает основания для отнесе-

Показатели	2000	2005	2009	2010
Численность врачей, всего, тыс. чел.	680	690	711	716
Число врачей на 10 тыс. чел. населения	46,8	48,8	50,1	50,1
Число больничных коек на 10 тыс. чел. населения	115	111	97	94

ния простых таблиц к комбинированным, так как такая комбинация все равно не позволяет исследовать зависимость между признаками.

Групповыми статистическими таблицами называются такие таблицы, в которых изучаемый объект разделен в подлежащем на группы по тому или иному признаку. Иначе говоря, групповые таблицы возникают в результате применения метода группировок при сводке статистических данных. Рассмотрим табл. 4.4.

Данная таблица является групповой, так как в подлежащем ее дана группировка банков по кредитным вложениям. Анализ таблицы говорит о том, что 46,6 % коммерческих банков имеют объем кредитных вложений от 139 до 231 млн руб., чистые активы этих банков составляют 4 655 млн руб.

Часто в сказуемом групповых таблиц показатели располагаются по периодам, поскольку в изменении соотношения групп во времени и проявляется конкретная закономерность. Так, в различные периоды времени изменялось соотношение городского и сельского населения в общем составе жителей страны (табл. 4.5).

По построению подлежащего табл. 4.5 относится к групповым таблицам, поскольку в ней используется группировка населения на городское и сельское.

Комбинационной называется такая таблица, в подлежащем которой дана группировка единиц совокупности по двум или более признакам, взятым в комбинации. Следовательно, комбинационная таблица в подлежащем содержит группы, образованные

№ п/п	Группы банков по кредитным вложениям, млн руб.	Число банков		Объем кредитных вложений, млн руб.	Чистые активы, млн руб.
		Единиц	% к итогу		
	А	1	2	3	4
1	До 139	6	20,0	339	1 250
2	139—185	7	23,3	1 078	2 087
3	185—231	7	23,3	1 427	2 568
4	231—277	3	10,0	761	1 577
5	277—323	3	10,0	865	1 922
6	323 и более	4	13,4	1 584	3 149
	Итого	30	100,0	6 054	12 553

№ п/п	Группы населения по типу проживания	1989	2002	2009	2010
	А	1	2	3	4
1	Городское	73,0	73,0	73,0	74,00
2	Сельское	27,0	27,0	27,0	26,00
	Итого	100,00	100,00	100,00	100,00

по одному признаку, и подгруппы (внутри групп), образованные по другому признаку. Познавательная сторона комбинационных таблиц заключается в том, что появляется возможность проследить влияние на признаки сказуемого не одного, а двух или более факторов, т. е. признаков, которые легли в основание комбинационной группировки (табл. 4.6).

По построению подлежащего табл. 4.6 является комбинационной. В подлежащем таблицы дана группировка банков по объему вложений в ценные бумаги (первый признак). Эта группировка разбита на подгруппы по второму признаку — величине кредитных вложений. Данные таблицы говорят о том, что большинство банков вкладывает в ценные бумаги менее 1 000 тыс. руб., а в общей численности банков они составляют 44 %.

Что касается капитальных вложений, то наибольшую долю в общем капитале имеют банки 5-й группы. Она составляет 34,6 % от общего объема капитальных вложений. К тому же в банках 5-й группы суммарная величина чистых активов в четыре раза больше, чем в банках 1-й группы, и в восемь раз больше, чем в 4-й. В 5-й же группе наибольшая величина чистых активов у банков 4-й подгруппы (с величиной кредитных вложений от 24 750 млн руб. до 32 950 млн руб.).

По структурному построению сказуемого различают статистические таблицы с простой и сложной его разработкой. Так, при большом числе признаков, характеризующих каждую группу подлежащего, можно по-разному составить таблицу.

Положим, что подлежащим простой таблицы является территориальный признак — область и район, а каждая территориальная единица указывает общую численность населения, количество мужчин и женщин и численность в возрасте до 17 лет и 18 лет и старше.

№ п/п	Группы банков по объему вложений в ценные бумаги, тыс. руб.	Подгруппы банков по величине кредитных вложений, тыс. руб.	Число банков	Капитал	Чистые активы	Вклады граждан
	А	Б	1	2	3	4
1	До 1 000	150 — 8 350	11	10 096	31 191	1 682
		8 350 — 16 550	—	—	—	—
		16 550 — 24 750	—	—	—	—
		24 750 — 32 950	—	—	—	—
		Итого по группе	11	10 096	31 191	1 682
2	1 000 — 2 000	150 — 8 350	2	1 439	7 483	699
		8 350 — 16 550	1	1 328	17 408	852
		16 550 — 24 750	2	6 088	45 939	894
		24 750 — 32 950	—	—	—	—
		Итого по группе	5	8 855	70 830	2 445
3	2 000 — 3 000	150 — 8 350	3	2 866	25 298	4 156
		8 350 — 16 550	—	—	—	—
		16 550 — 24 750	—	—	—	—
		24 750 — 32 950	—	—	—	—
		Итого по группе	3	2 866	25 298	4 156
4	3 000 — 4 000	150 — 8 350	2	4 522	15 350	121
		8 350 — 16 550	—	—	—	—
		16 550 — 24 750	—	—	—	—
		24 750 — 32 950	—	—	—	—
		Итого по группе	2	4 522	15 350	121
5	4 000 и более	150 — 8 350	1	2 711	14 038	134
		8 350 — 16 550	—	—	—	—
		16 550 — 24 750	2	9 583	69 713	2 299
		24 750 — 32 950	1	1 612	41 886	9 017
		Итого по группе	4	13 906	125 637	11 450
	Итого по подгруппам	150 — 8 350	19	21 634	43 360	6 092
		8 350 — 16 550	1	1 328	17 408	852
		16 550 — 24 750	4	15 671	11 562	3 193
		24 750 — 32 950	1	1 612	41 886	9 017
		Всего	25	40 245	268 306	19 854

Тогда заголовки граф таблицы будут такими:

Области, районы	Все население	В том числе			
		мужчины	женщины	в возрасте	
				до 17 лет	18 лет и старше
....	....	....	....	....	....
....	....	....	....	....	....
....	....	....	....	....	....
....	....	....	....	....	....

Такая разработка сказуемого в статистике называется *простой*, поскольку каждый признак в сказуемом подсчитывается отдельно. В самом деле, по каждой области и району можно получить сведения о распределении населения по полу и возрасту. Углубленную статистическую характеристику населения может дать ***сложная*** (***комбинационная***) разработка сказуемого. Она предполагает деление признака на подгруппы, когда оба признака сказуемого связаны между собой. При этом заголовки граф таблицы будут выглядеть уже по-другому:

Области, районы	Все население	В том числе распределение по полу и возрасту			
		до 17 лет		18 лет и старше	
		женщины	мужчины	женщины	мужчины
....	....	....	....	....	....
....	....	....	....	....	....
....	....	....	....	....	....
....	....	....	....	....	....

В такой таблице можно проследить не только численность населения в целом, но и оценить число мужчин и женщин указанных возрастных групп.

Если в таблице с простой разработкой сказуемого может быть столько сведений о каждой группе, сколько признаков отмечено при статистическом наблюдении каждой единицы, то при сложной разработке сказуемого каждая группа может быть охарактеризована различным сочетанием этих признаков.

Число сочетаний должно быть строго ограничено и соответствовать задаче статистического исследования. При сложной, комбинационной разработке сказуемого разноска данных наблюдения в таблице затруднена. Поэтому построению таблиц предшествует подбор макета для нее, состоящего из строк и граф, не заполненных цифрами.

При составлении макетов статистических таблиц (проектирования) необходимо тщательно взвесить, какие признаки подвергнуть простой, какие — сложной разработке сказуемого. Надо учитывать, что комбинированный подсчет сложнее и дороже, что при таком подсчете могут появиться таблицы с большим количеством граф, труднодоступные для обозрения.

С другой стороны, необходимо учитывать и бо́льшие возможности статистического анализа, осуществляемого с помощью таблиц со сложной разработкой сказуемого. Подсчет итогов целесообразно начинать по таблицам со сложной разработкой сказуемого, из которых легко получаются таблицы с простой разработкой сказуемого.

Вид таблицы не зависит от способа разработки его сказуемого. При сложной или простой разработке сказуемого таблица может быть простой, групповой или комбинационной. Следовательно, вид таблицы зависит исключительно от разработки ее подлежащего.

Практикой выработаны рациональные правила составления и оформления статистических таблиц. Перечислим их.

1. Таблица должна быть по возможности компактной, небольшой по размеру. Иногда целесообразнее построить две-три небольшие таблицы, чем одну большую. Краткую таблицу легче проанализировать. Цифровой материал необходимо располагать 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 таблиц и граф. Это облегчит чтение таблиц. В заголовках граф допускаются точки только при необходимых сокращениях.

В заголовке таблицы должны найти отражение объект, признак, время и место совершения события. И чем короче и лаконичнее заголовок, тем он доходчивее для чтения и анализа. Естественно, делается это в меру, не в ущерб точности и познавательности. Заголовки таблицы, граф и строк пишутся полностью, без сокращений.

3. Информация, располагаемая в графах таблицы, как правило, завершается итоговой строкой. В групповых и комбинационных таблицах всегда необходимо давать итоговые графы и строки. Су-

ществуют различные способы соединения слагаемых граф с их итогом. Так, строка «Итого» или «Всего» может завершать статистическую таблицу. Но она может располагаться первой, соединяясь с совокупностью ее слагаемых словами «В том числе».

4. 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 как для подлежащего, так и для сказуемого таблиц.

5. Строки и графы в таблице нумеруются для того, чтобы удобнее было ссылаться на цифры в таблице. При этом графы, содержащие подлежащее, нумеруются заглавными буквами алфавита («А», «Б», «В» и т. д.), графы, содержащие сказуемое, — арабскими цифрами.

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

7. 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.). Если все графы имеют единую единицу измерения, то она выносится в заголовок таблицы.

8. Лучше всего располагать в таблицах сопоставляемую в ходе анализа цифровую информацию в одной и той же графе, одну под другой. Это значительно облегчает процесс их сравнения. Поэтому в групповых таблицах, например, группы по изучаемому признаку правильнее всего располагать в порядке убывания или возрастания его значений при сохранении логической связи между подлежащим и сказуемым таблицы.

9. Для более удобной работы с цифровым материалом числа в таблицах следует расставлять в середине граф, одно под другим: единицы под единицами, запятая под запятой, четко соблюдая при этом их разрядность.

10. Числа по возможности целесообразно округлять. При этом округление в пределах одной и той же графы или строки следует проводить с одинаковой степенью точности.

11. Отсутствие данных об анализируемом социально-экономическом явлении может быть обусловлено различными причинами и по-разному фиксируется в таблице. Если данная позиция (на пересечении соответствующих графы и строки) вообще не подлежит заполнению или не имеет экономического смысла, то ставится знак «x».

Если по какой-либо причине отсутствуют сведения, то ставится многоточие или пишут «нет сведений». Если сведения имеются, но числовые значения меньше принятой в таблице точности, то ставится число 0,0. Если сведения о данном явлении отсутствуют, то клетка заполняется с помощью тире.

12. Таблица может сопровождаться примечаниями, в которых указываются источники данных, более подробное содержание показателей и другие необходимые пояснения.

Нужно уметь пользоваться статистическими таблицами. Прежде чем приступить к анализу ее данных, необходимо ознакомиться с названием таблицы, заголовками граф и строк, установить, к какому периоду или на какую дату, к какой территории относятся данные, обратить внимание на единицы измерения, уяснить, какие процессы характеризуются средними и относительными величинами.

Анализ статистической таблицы следует начинать с итогов. Ознакомление с ними дает общее представление о содержании таблицы. Затем необходимо перейти к изучению данных отдельных строк и граф. Но читать их нужно не подряд, а выбирать сначала частные итоги и наиболее характерные данные, а затем анализировать все остальное.

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

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.

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 они являются незаменимым средством обобщения статистических данных, подведения итогов сложных исследований и выявления связи между явлениями. Поэтому так необходимо уметь строить и читать статистические графики.

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

Полем графика называют место, на котором он 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 близким к оптимальному для зрительного восприятия является график, выполненный на поле прямоугольной формы с соотношением сторон от 1 : 1,3 до 1 : 1,5. Этот вариант именуется правилом «золотого сечения». Иногда используется и поле графика с равными сторонами, т. е. имеющее форму квадрата.

Графический образ — это символические знаки, с помощью которых изображаются статистические данные: линии, точки, плоские геометрические фигуры (прямоугольники, квадраты, круги и т. д.). В качестве графического образа выступают и объемные фигуры. Иногда в графиках используются и негеометрические фигуры (силуэты, изображения каких-либо предметов и т. п.). В принципе одни и те же статистические данные можно выразить графически с помощью самых разных средств. Поэтому-то так важен правильный подбор графического образа. Он должен в наибольшей мере соответствовать основному предназначению графика и наиболее доходчиво отображать изучаемые показатели.

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

В практике графического изображения применяются также полярные координаты. Они необходимы для наглядного изображения циклического движения во времени. В полярной системе координат (рис. 4.1) один из лучей, обычно правый горизонтальный, принимается за ось координат, относительно которой определяется угол луча. Второй координатой считается ее расстояние от центра сетки,

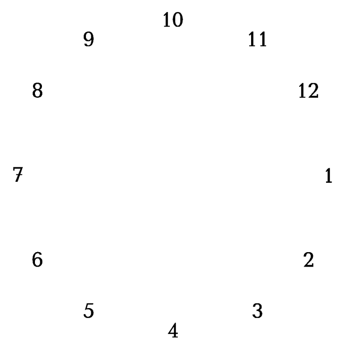


Рис. 4.1. Числовые интервалы в полярной системе координат

называемое *радиусом*. В радиальных графиках лучи обозначают моменты времени, а окружности — величины изучаемого явления.

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

Масштабные ориентиры статистического графика определяются масштабом и системой масштабных шкал. Масштаб статистического графика — это мера перевода числовой величины в графическую. ***Масштабной шкалой*** называется линия, отдельные точки которой могут быть прочитаны как определенные числа. Шкала имеет большое значение в графике и включает три элемента: линию (или носитель шкалы), определенное число помеченных черточками точек, которые расположены на носителе шкалы в 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 интервале. По правилам числовое значение необходимо помещать строго против соответствующих точек, а не между ними (рис. 4.2).

Носитель шкалы может представлять собой как прямую, так и кривую линии. Поэтому различают шкалы ***прямолинейные*** (например, миллиметровая линейка) и ***криволинейные*** — дуговые

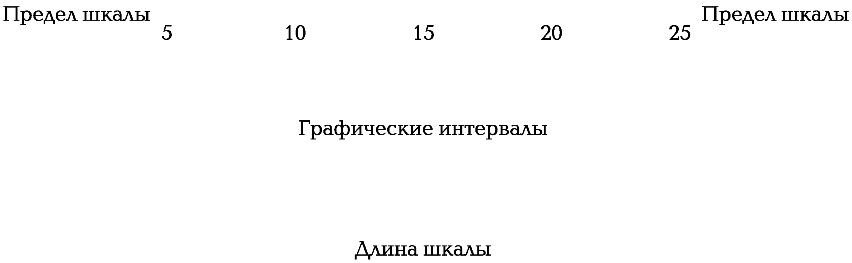


Рис. 4.2. Длина отрезка равномерной масштабной шкалы

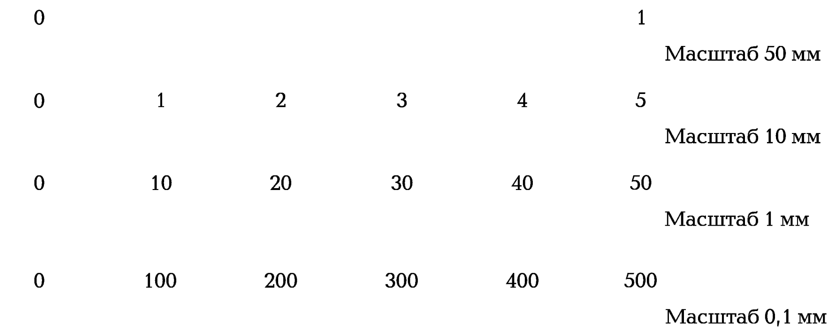


Рис. 4.3. Масштабы шкалы

и круговые (циферблат часов) (см. рис. 4.1). Графические и числовые интервалы бывают равными и неравными. Если на всем протяжении шкалы равным графическим интервалам соответствуют равные числовые, то такая шкала называется ***равномерной***. Когда же равным числовым интервалам соответствуют неравные графические интервалы, и наоборот, то перед нами ***неравномерная*** шкала. ***Масштабом равномерной шкалы*** называется длина отрезка (графический интервал), принятого за единицу и измеренного в каких-либо мерах. Чем меньше масштаб (рис. 4.3), 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.

Как правило, масштаб определяется примерной прикидкой возможной длины шкалы и ее пределов. Скажем, на поле в 20 клеток надо построить шкалу от 0 до 850. Так как 850 не делится удобно на 20, то округляем это число до ближайшего удобного числа, в данном случае 1 000 (1 000 : 20 = 50), т. е. в одной клетке 50, в двух клетках 100; следовательно, масштаб — 100 в двух клетках.

Из неравномерных наибольшее распространение имеет логарифмическая шкала. Методика ее построения несколько иная, так как на этой шкале отрезки пропорциональны не изображаемым величинам, а их логарифмам. Так, при основании логарифма 10 имеем $\lg 1 = 0$; $\lg 10 = 1$; $\lg 100 = 2$ и т. д. (рис. 4.4).

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

0 0,5 1

0 1 2 1

а

0 10 100 1000 Числа

0 1 2 3 Логарифмы чисел

б

Рис. 4.4. Масштабные шкалы:
а — равномерные шкалы; *б* — неравномерные шкалы

Построение графика — всегда творческий процесс. Здесь необходим некоторый поиск. Лишь после составления и сравнения нескольких черновых вариантов можно наметить правильную композицию графика, установить масштабы и расположение знаков на поле графика.

Наконец, о заголовке графика. Он должен кратко и точно раскрыть содержание графического изображения, дать ответ на три вопроса — что, где, когда?

Существует множество видов графических изображений. Их классификация (рис. 4.5 и 4.6) основана на ряде признаков: а) способ построения графического образа; б) геометрические знаки, изображающие статистические показатели; в) задачи, решаемые с помощью графического изображения.

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, временно́м и др.) независимых друг от друга совокупностей. При этом сравнение исследуемых совокупностей производится по какому-либо существующему варьирующему признаку.

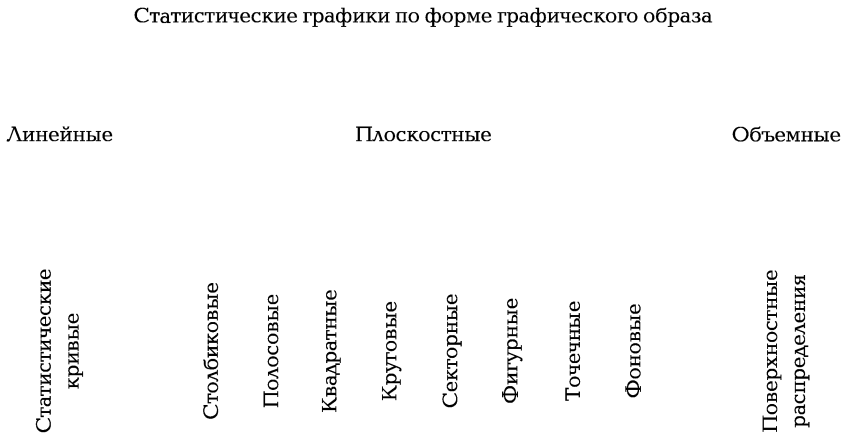


Рис. 4.5. Классификация статистических графиков по форме графического образа

Статистические карты — графики количественного распределения по конкретной территории. По своей основной характеристике они близко примыкают к диаграммам и специфичны лишь в том отношении, что представляют собой условные изображения статистических данных на контурной географической карте. Их задача — отражать пространственное размещение или про-

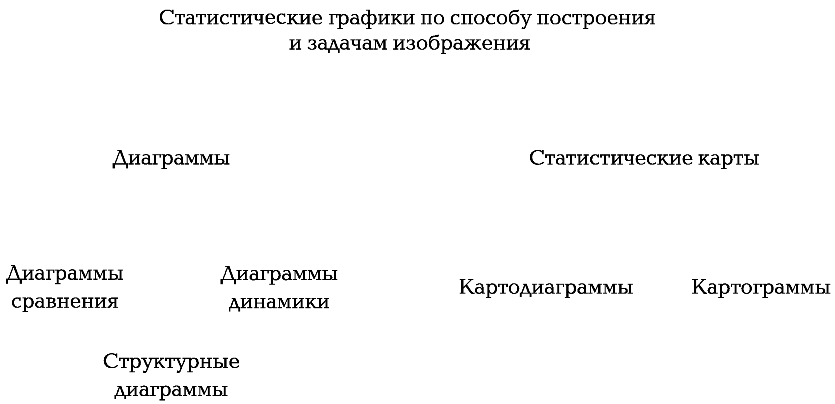


Рис. 4.6. Классификация статистических графиков по способу построения и задачам изображения

странственную распространенность статистических показателей. По графическому образу статистические карты делятся на картограммы и картодиаграммы.

Геометрические знаки, как было сказано выше, представляют собой точки, линии, плоскости, геометрические тела. 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 плоскостных диаграмм сводится к тому, что статистические величины 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. Это гистограмма, полигон, огива, кумулята. В данной главе рассматриваются наиболее часто применяемые в статистической практике графики, на которые можно ориентироваться в известной степени при составлении аналогичных графиков.

Наиболее распространенными диаграммами сравнения являются ***столбиковые диаграммы*** — графическое изображение статистических данных в виде столбиков — прямоугольников. Эти 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 располагаются столбики. На горизонтальной оси располагаются основания столбиков. Размер основания столбиков определяется произвольно, но он должен быть одинаковым для всех.

Шкала, определяющая масштаб столбиков по высоте, расположена по вертикальной оси. Величина каждого столбика по вертикали соответствует размеру изображаемого на графике статистического показателя. Таким образом, у всех столбиков, составляющих диаграмму, переменной величиной является только одно измерение. Размещение столбиков в поле графика производится на оди-

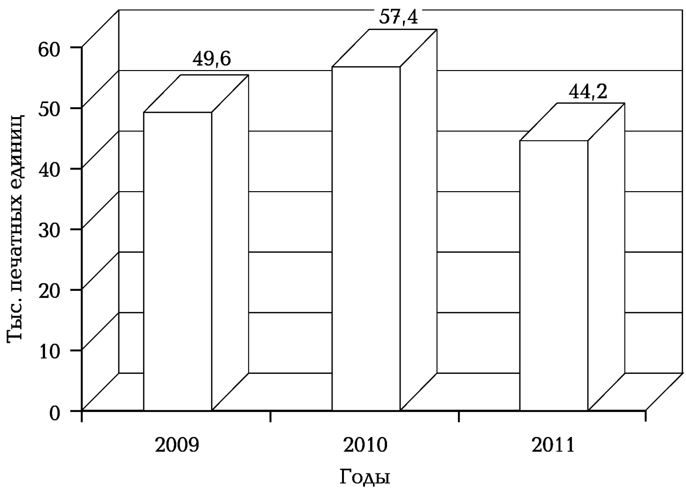


Рис. 4.7. Динамика выпуска книг и брошюр в одном из регионов России за 2009—2011 гг. (цифры условные).

наковом расстоянии друг от друга, вплотную друг к другу, при частичном наложении друг на друга.

Технику построения столбиковой диаграммы рассмотрим на примере данных о динамике выпуска книг и брошюр в N-м регионе России. Предположим, что это были 2009—2011 гг. Итак: 2009 г. — 49,6 тыс. печатных единиц; 2010 г. — 57,4 тыс. печатных единиц; 2011 г. — 44,2 тыс. печатных единиц.

Для построения диаграммы (рис. 4.7) берем систему прямоугольных координат. На оси абсцисс на одинаковом расстоянии друг от друга наносим три отрезка равной длины — основания для столбиков. Высота столбиков определяется в соответствии с принятым масштабом по оси ординат и значениями показателей. Учитывая размер поля графика и максимальное значение показателя, установим масштаб. Допустим, что каждым 10 тыс. печатных единиц соответствует отрезок по оси ординат 1 см. Тогда высота столбиков составляет для 2009 г. — 4,96 см (49,6 : 10); для 2010 г. — 5,74 см и для 2011 г. — 4,42 см.

Наглядность данной диаграммы достигается сравнением высоты столбиков.

Столбиковые диаграммы целесообразно применять для сравнения нескольких показателей. На рис. 4.8 посредством столбиковой диаграммы показана динамика инвестиций в основной капитал

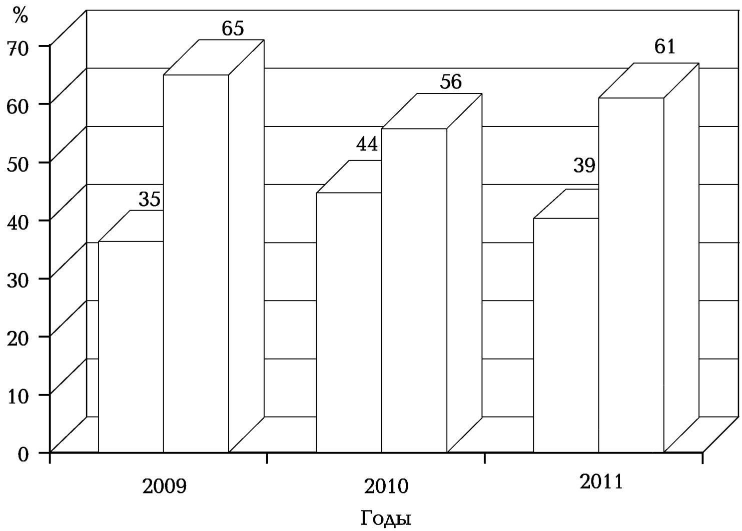


Рис. 4.8. Динамика инвестиций в основной капитал по объектам производственного и непроизводственного назначения за 2009—2011 гг., % к итогу (цифры условные):

□ — объекты производственного назначения; □ — объекты непроизводственного назначения

по объектам производственного и непроизводственного назначения в Москве за 2009—2011 гг. На этой диаграмме столбики располагаются вплотную по группам объектов. Масштаб принят такой, что каждым 10 % соответствует отрезок на оси ординат в 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 суточный расход энергии у людей разных профессий.

Для построения данной диаграммы на поле графика откладываем полосы, длина которых соответствует значениям изображаемых данных на масштабной шкале (рис. 4.9). Из диаграммы следует, что среди представителей названных профессий наибольшее количество калорий за сутки затрачивают штукатуры.

Столбиковые и полосовые диаграммы хорошо подходят для характеристики структуры 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.

Для простого сравнения не зависимых друг от друга показателей могут также использоваться диаграммы, в которых сравниваемые величины изображаются в виде ***правильных геометрических фигур***. Они строятся так, чтобы площади их соотносились между собой как значения величин, этими фигурами изображаемые. Иными словами, эти диаграммы выражают величину изображаемого явления размером своей площади. Для создания диаграмм такого типа используют разнообразные геометрические фигуры — круги, квадраты, реже — прямоугольники.

Для построения ***квадратных*** и ***круговых диаграмм*** необходимо сначала из статистических данных извлечь квадратные корни. Затем на базе полученных результатов определить сторону квадрата или радиус круга соответственно принятому масштабу. Например, необходимо изобразить в виде квадратов добычу полезных ископаемых в 2009 г. в России.

Для построения квадратной диаграммы сначала извлечем квадратные корни из чисел: $\sqrt{301} = 17{,}3$; $\sqrt{494} = 22{,}2$; $\sqrt{583} = 24{,}14$. Затем установим масштаб, примем 1 см = 8 млн т. Тогда сторона первого квадрата составит 2,1 см (17,3 : 8), второго — 2,7 см (22,2 : 8), третьего — 3 см (24,14 : 8). Далее строим квадраты на одинаковом

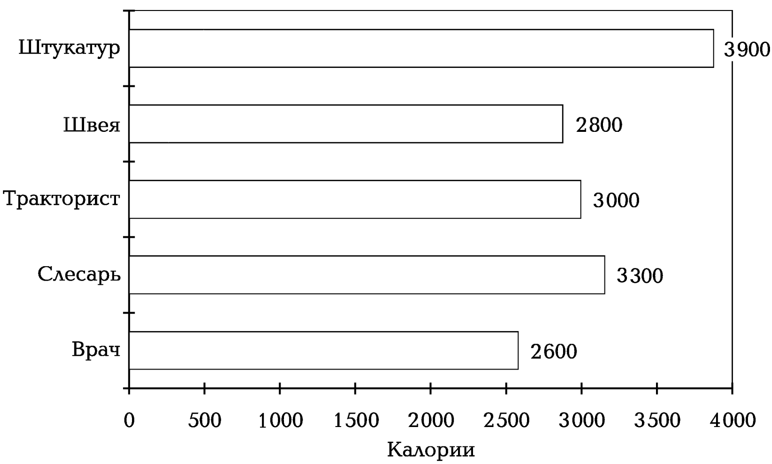


Рис. 4.9. Суточный расход энергии людей разных профессий

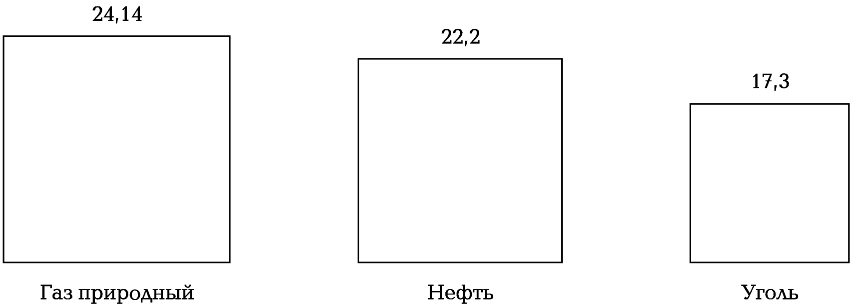


Рис. 4.10. Добыча полезных ископаемых в 2009 г. в России

расстоянии друг от друга. Квадратная диаграмма представлена на рис. 4.10.

Круговая диаграмма строится аналогично квадратной с той разницей, что находим значение радиуса для каждого круга.

Для правильного построения диаграмм квадраты или круги необходимо расположить на одинаковом расстоянии друг от друга (см. рис. 4.10), а в каждой фигуре указать числовое значение, которое она изображает, не приводя масштаба измерения.

К рассматриваемому виду диаграмм относится графическое изображение, полученное путем построения один в другом квадратов, кругов или прямоугольников. Такие диаграммы также позволяют сравнить между собой ряд исследуемых величин.

Круг часто используется в качестве геометрической формы при построении диаграмм. Следует различать два вида применения круга. В одном случае сравниваются площади кругов друг с другом. Такого рода диаграммы называются ***круговыми***. В другом случае круг используется для сравнения площади отдельных секторов друг с другом. Такая диаграмма именуется ***секторной***. Секторная диаграмма применяется для наглядной иллюстрации структуры какого-либо явления, характеристики удельных весов отдельных частей целого, выявления структурных сдвигов.

Построим секторную диаграмму, характеризующую распределение занятого населения одной из областей по отраслям народного хозяйства в 2011 г. Удельный вес населения, занятого в отдельных отраслях, к общей численности занятого населения области составил: промышленность — 31,8 %; торговля и сфера услуг — 34,2 %; сельское хозяйство — 4,6 %; здравоохранение, образование, наука и культура — 18,7 %; другие отрасли — 10,7 %. Возьмем круг произвольного радиуса. Площадь круга нужно разделить на пять секторов. Известно, что площади секторов пропорциональны их центральным углам. Следовательно, для определения площади секторов нужно 360° распределить пропорционально величинам удельных весов. Поскольку 3,6° соответствует 1 %, то центральные углы составят: промышленность — $3{,}6° \cdot 31{,}8 = 114{,}48°$, торговля — $3{,}6° \cdot 34{,}2 = 123{,}12°$, сельское хозяйство — $3{,}6° \cdot 4{,}6 = 16{,}56°$, здравоохранение — $3{,}6° \cdot 18{,}7 = 67{,}3°$, другие отрасли — $3{,}6° \cdot 10{,}7 = 38{,}52°$.

На основании полученных данных делим круг на сектора и получаем диаграмму (рис. 4.11).

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

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 изучаемого явления. Достоинство такого изображения заключается в высокой степени наглядности, в получении подобного изображения, отражающего содержание сравниваемых явлений.

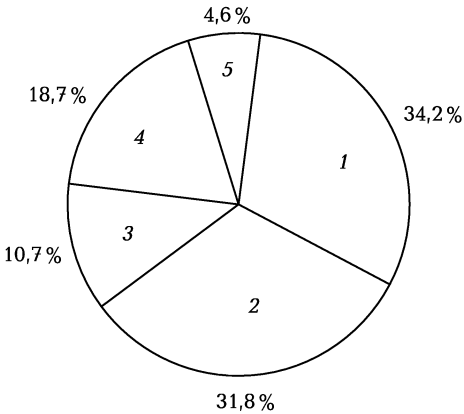


Рис. 4.11. Распределение занятого населения одной из областей по отраслям народного хозяйства в 2011 г., %:

1 — торговля и сфера услуг; *2* — промышленность; *3* — другие отрасли; *4* — здравоохранение, образование, наука и культура; *5* — сельское хозяйство

Фигурные диаграммы можно строить, используя различную численность фигур одинакового размера или фигуры неодинаковых размеров.

Предположим, что на рынках города продано населению за 5 мес яблок, ц: в январе — 85,1; в феврале — 100,7; в марте — 37,3; в апреле — 29,5; в мае — 30,2.

При построении графика путем различной численности ***фигур одинакового размера*** прежде всего необходимо установить масштабный знак с таким расчетом, чтобы не получилось фигур слишком много, но в то же время и не слишком мало. Иначе график будет невыразителен. В нашем примере хорошо подходит такой масштаб, когда каждой фигуре соответствует продажа 10 ц яблок. Разделим приведенные показатели на 10 и получим число фигур для января — 8,5; для февраля — 10,1; для марта — 3,7; для апреля — 2,95; для мая — 3,01. Построим диаграмму (рис. 4.12).

Для построения диаграмм с ***фигурами различного размера*** необходимо предварительно построить соответствующие по величине квадраты, а затем уже внутри каждого квадрата рисовать фигуру изучаемого явления.

При построении фигурных диаграмм следует иметь в виду, что на одном графике может быть отражена лишь одна тема. Не следует применять сложные рисунки, а также слишком яркие краски. Фигурам-символам надлежит быть лаконичными, простыми и понятными. Они не должны требовать дополнительных объяснений.

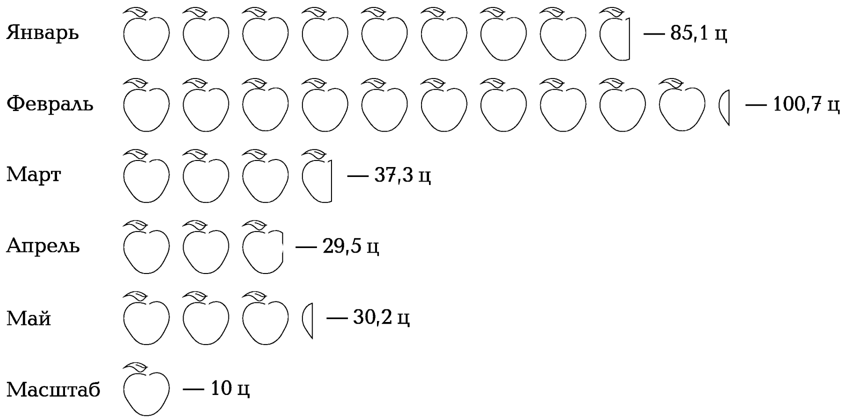


Рис. 4.12. Динамика реализации яблок на рынках одного из городов за 5 мес

Прямоугольные диаграммы (не квадраты!) находят применение, главным образом, для двумасштабных сравнений: один масштаб для основания, другой — для высоты. Эти диаграммы называют знаками Варзара. Обычно они применяются в тех случаях, где показатель является произведением двух других (например, валовой сбор есть произведение посевной площади на урожайность, численность населения является произведением плотности населения на территорию и т.д.). Такой показатель можно графически изобразить в виде сомножителей. Для этого поступают следующим образом: один множитель принимают за основание, другой — за высоту. Затем устанавливают масштабы: один для основания, другой для высоты. Далее, располагая значением статистического показателя основания и высоты, строят прямоугольники. Покажем этот способ на примере данных по сбору яровой пшеницы в одном из регионов России, в котором при посевной площади 14,5 млн га урожайность составила 1,16 т/га (рис. 4.13).

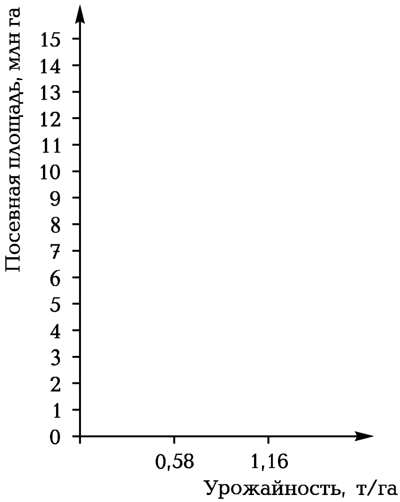


Рис. 4.13. Зависимость валового сбора яровой пшеницы от урожайности и посевной площади в одном из регионов России

В нашем случае в основание прямоугольника положена урожайность яровой пшеницы, высота — посевная площадь, а площадью прямоугольника является валовой сбор яровой пшеницы $1{,}16 \cdot 14{,}5 = 16{,}8$ млн т.

Широкое распространение имеют диаграммы, которые можно назвать ***координатными***, поскольку они основаны на системе прямоугольных координат. В отличие от столбиковых и полосовых диаграмм диаграммы этого вида требуют не одного, а двух масштабов: одного по оси абсцисс, другого по оси ординат. Среди координатных диаграмм наиболее распространены ***линейные***, когда характеризуются изменения явлений во времени. Они незаменимы в тех случаях, когда на одном графике нужно показать динамику нескольких явлений. В статистической практике чаще всего применяются графические изображения динамики 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.

Покажем построение линейного графика на следующем примере. Динамика инвестиций в основной капитал в одной из областей за 2002—2011 гг. (в % к 2002 г.) характеризуется следующими данными: 2002 г. — 100; 2003 г. — 51; 2004 г. — 50,5; 2005 г. — 61,6; 2006 г. — 63,7; 2007 г. — 54,9; 2008 г. — 44,3; 2009 г. — 44,3; 2010 г. — 52,1 и 2011 г. — 54,2.

На оси абсцисс прямоугольной системы координат откладываем десять точек с учетом одинаковой продолжительности периодов времени между приведенными годами. Учитывая, что максимальное значение уровня 100 %, по оси ординат принимаем масштаб: 20 % соответствует 1,0 см. На вертикальную шкалу наносим числа масштабов. Из точек на оси абсцисс восстанавливаем перпендикуляры, высота которых пропорциональна динамике инвестиций и принятому масштабу по оси ординат. Вершины перпендикуляров соединяем отрезками прямой и получаем ломаную линию, характеризующую динамику инвестиций в основной капитал в одной из областей за 2002—2011 гг. (рис. 4.14).

Нередко на одном линейном графике проводится несколько кривых, которые дают сравнительную характеристику динамики различных показателей или одного и того же показателя на разных территориях. Примером графического изображения сразу нескольких показателей является рис. 4.15.

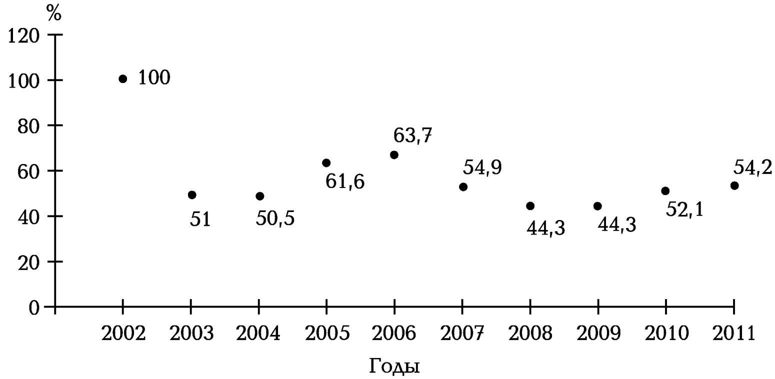


Рис. 4.14. Динамика инвестиций в основной капитал в одной из областей за 2002—2011 гг. (в % к 2002 г.)

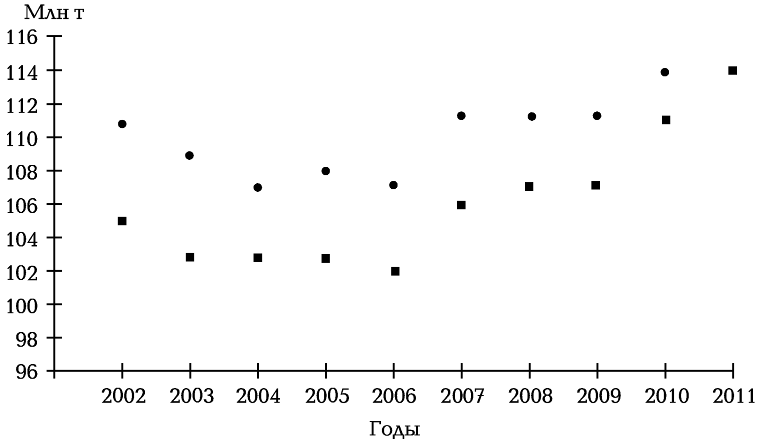


Рис. 4.15. Динамика производства чугуна и готового проката в регионе за период с 2002 по 2011 гг.

● чугун; ■ готовый прокат

На одном графике не следует помещать более трех-четырех кривых. Большое их количество неизбежно осложняет чертеж и линейная диаграмма теряет наглядность.

В некоторых случаях нанесение 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. 4.16.

Линейные диаграммы с равномерной шкалой имеют один недостаток, снижающий их познавательную ценность. Равномерная шкала позволяет измерять и сравнивать только отраженные на диаграмме абсолютные приросты или уменьшения показателей на про-

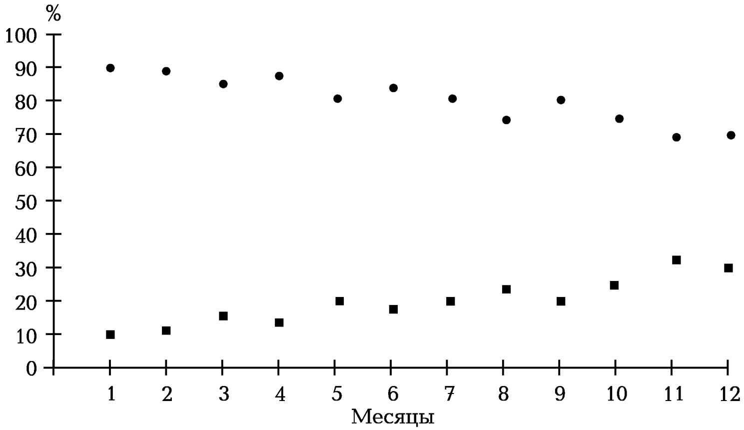


Рис. 4.16. Удельный вес вкладов граждан в Сбербанк и коммерческие банки города по месяцам:

● Сбербанк; ■ коммерческие банки

тяжении исследуемого периода. Однако при изучении динамики важно знать относительные изменения исследуемых показателей по сравнению с достигнутым уровнем или темпы их изменения. Но 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 и даже становится невозможным изображение 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 свое преимущество. Возникает возможность уменьшения 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.

Год	y_i	$\lg y_i$	Год	y_i	$\lg y_i$
1980	170	2,23	2000	1039	3,02
1985	292	2,46	2005	1294	3,11
1990	507	2,70	2010	1544	3,19
1995	741	2,84			

Полулогарифмической сеткой называется сетка, где 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.). Покажем построение линейной диаграммы на полулогарифмической сетке на основе данных динамики производства электроэнергии в регионе (табл. 4.7).

Определив минимальное и максимальное значение логарифмов производства электроэнергии, построим масштаб с таким расчетом, чтобы все данные разместились на графике. Учитывая масштаб, находим соответствующие точки, которые соединим прямыми линиями, в результате получим график (рис. 4.17) с использованием логарифмического масштаба на оси ординат. Он называется

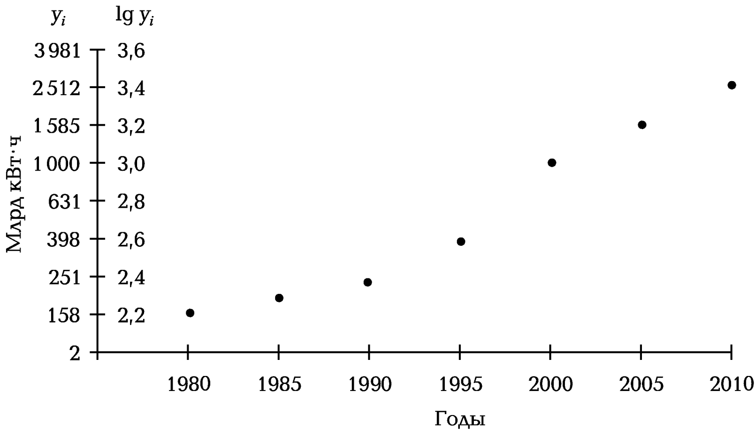


Рис. 4.17. Динамика производства электроэнергии в регионе за 1980— 2010 гг.

диаграммой на полулогарифмической сетке. Полной логарифмической диаграммой он станет в том случае, если по оси абсцисс будет построен логарифмический масштаб. Однако в рядах динамики это неприменимо, поскольку 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. Если она отклоняется от прямой в сторону, выпуклую к оси абсцисс, изучаемое явление имеет тенденцию к росту с увеличивающимися темпами. Когда кривая в своем течении приближается к прямой, наблюдается стабильность темпов.

Статистические карты представляют собой вид графических изображений на схематичной географической карте статистических данных, характеризующих 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 (рис. 4.18, 4.19).

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 — такой вид картограммы, где штриховкой различной густоты или окраской определенной степени насыщенности показывают интенсивность какого-либо показателя в пределах территориальной единицы. ***Картограмма точечная*** изображает уровень какого-либо явления с помощью точек. Точка изображает одну единицу совокупности или некоторое их количество. Путем нанесения точек на географической карте показывается плотность или частота появления определенного явления. Фоновые картограммы, как правило, используются для изображения средних или относительных показателей, а точечные — для

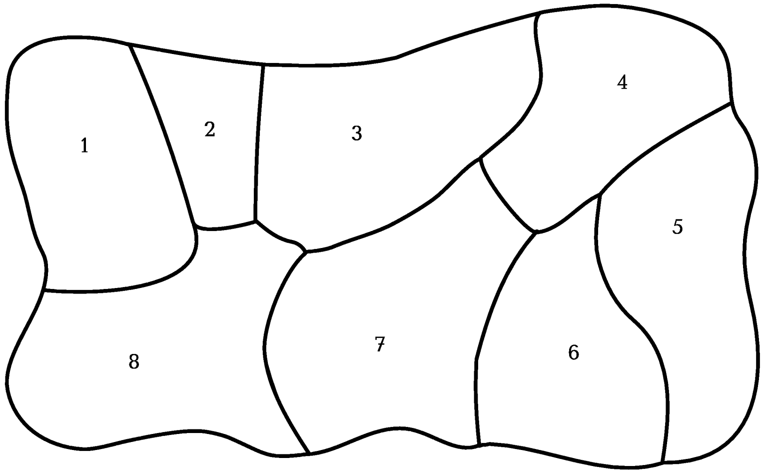


Рис. 4.18. Картограмма плотности населения восьми районов области на 1 000 км^2 площади:

☐ — до 4 тыс. чел.; ☐ — 4—12 тыс. чел.; ☐ — 12—17 тыс. чел.

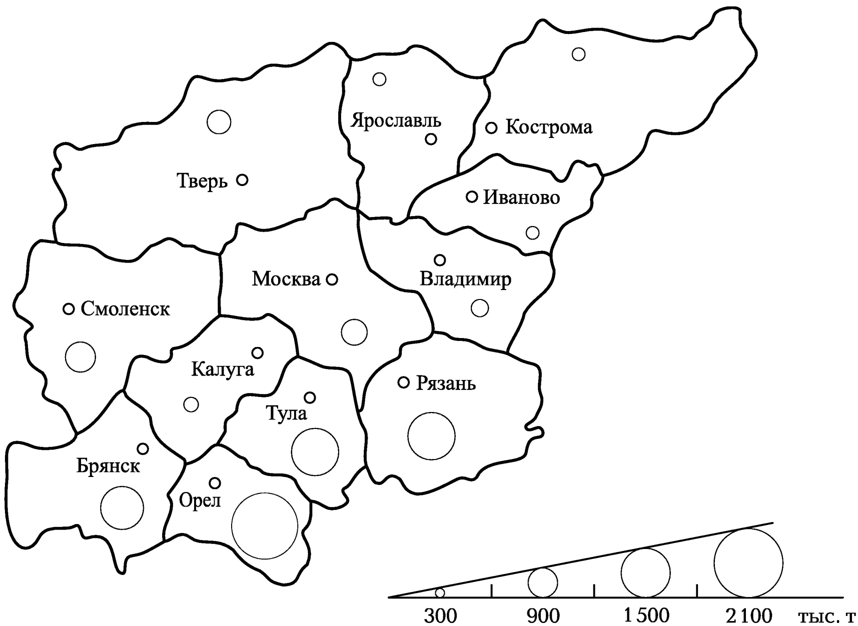


Рис. 4.19. Картограмма валового сбора зерна Центрального района России (данные условные)

Показатель	Район							
	1	2	3	4	5	6	7	8
Плотность населения на 1 000 м^2, тыс. чел.	3,0	4,0	11,0	14,0	17,0	13,0	11,0	3,0

объемных (количественных) показателей (численность населения, поголовье скота и т. д.). Рассмотрим построение картограммы, используя данные табл. 4.8.

Прежде чем приступить к построению картограммы, необходимо разбить районы на группы по плотности населения, а затем установить для каждой группы определенную окраску или штриховку.

Согласно данным табл. 4.8 все районы по плотности населения можно разбить на три группы: 1) районы, имеющие плотность населения до 4 тыс. чел.; 2) районы — от 4 до 12 тыс. чел.; 3) районы — от 12 до 17 тыс. чел. Тогда к первой группе относятся районы № 1, 8; ко второй — № 2, 3, 7; к третьей — № 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 (см. рис. 4.18).

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 отразить более сложные статистико-географические построения, чем картограммы. Среди картодиаграмм следует выделить картодиаграммы простого сравнения, пространственных перемещений и изолиний.

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 обычно присутствуют на политической карте, где города отмечаются различными геометрическими фигурами в зависимости от численности жителей.

В качестве примера на рис. 4.19 приведена картодиаграмма валового сбора зерна Центрального района России, где в качестве условных фигур выступают круги.

Изолинии (от греч. *isol* — равный, одинаковый, подобный) — линии равного значения какой-либо величины в ее распространении на поверхности, в частности на географической карте или графике. Изолинии отражают непрерывное изменение исследуемой величины в зависимости от двух других переменных. Они применяются при картографировании природных и социально-экономических явлений; могут быть использованы для получения их количественной характеристики и для анализа корреляционных связей между ними.

Перечисленные виды графиков не являются исчерпывающими, но остаются наиболее часто употребляемыми.

1. Статистическая таблица представляет собой:
   а) форму наиболее рационального изображения результатов статистического наблюдения;
   б) сведения о чем-нибудь, расположенные по строкам и графам.
2. Статистической таблицей является:
   а) таблица логарифмов;
   б) таблица умножения;
   в) таблица, в которой обобщаются итоги экзаменационной сессии по институту.
3. Статистической таблицей является:
   а) таблица расписания поездов;
   б) таблица квадратов;
   в) таблица, в которой обобщаются результаты финансовой работы банка.
4. Статистическим подлежащим называется:
   а) статистические совокупности, которые характеризуются различными показателями;
   б) показатели, характеризующие совокупности;
   в) сведения, расположенные в боковых заголовках таблицы;
   г) числовые характеристики, размещенные в графах таблицы.
5. Статистическим сказуемым называется:
   а) статистические совокупности, которые характеризуются различными показателями;
   б) показатели, характеризующие совокупности;
   в) сведения, расположенные в боковых заголовках таблицы;
   г) числовые характеристики, размещенные в графах таблицы.

6. Основными элементами статистического графика являются:
   а) поле графика;
   б) масштабные ориентиры;
   в) геометрические знаки;
   г) экспликация графика;
   д) рисунок.

7. Какие виды диаграмм используются в форме геометрического образа:
   а) линейные;
   б) плоскостные;
   в) объемные;
   г) статистические карты;
   д) диаграммы?

8. Какие виды статистических графиков существуют по экономическим задачам изображения социально-экономических явлений:
   а) диаграммы сравнения;
   б) диаграммы динамики;
   в) плоскостные диаграммы;
   г) диаграммы структуры;
   д) объемные диаграммы?

9. При изображении данных рядов распределения на графике применяются диаграммы:
   а) гистограммы;
   б) знаки Варзара;
   в) полигоны;
   г) кумуляты.

10. Известна динамика числа родившихся в целом по стране. Выберите подходящее графическое изображение этого процесса:
   а) статистическая кривая;
   б) картодиаграмма;
   в) картограмма;
   г) секторная диаграмма.

В результате сводки и группировки статистических данных получают обобщающие показатели, которые отражают результаты познания количественной стороны изучаемых явлений. Важной задачей статистики является построение статистических показателей.

Статистический показатель — это количественная характеристика социально-экономических явлений и процессов. Статистический показатель непосредственно выражает внутреннее содержание изучаемого явления или процесса, его сущность. Статистика знает большое количество разнообразных показателей, относящихся ко всем сторонам общественной жизни. К их числу принадлежат: показатели продукции различных отраслей; показатели, характеризующие с разных сторон население; показатели, относящиеся к характеристике кредитных учреждений страны и к продаже товаров населению; показатели медицинского обслуживания населения; показатели посевных площадей и поголовья скота; показатели запасов сырья и топлива; показатели доходов и расходов населения и т. д.

Достаточно обратиться к любому статистическому справочнику, скажем, к ежегоднику «Россия в цифрах. 2011», насчитывающему свыше 500 страниц и содержащему только небольшую часть имеющихся статистических показателей, чтобы составить себе представление об их количестве и разнообразии. Совокупность взаимосвязанных показателей, имеющая одноуровневую или многоуровневую структуру и нацеленная на решение конкретной статистической задачи, образует систему ***статистических показателей***.

Следует различать показатель-категорию и конкретный статистический показатель. ***Показатель-категория*** отражает сущность, общие отличительные свойства конкретных статистических показателей одного и того же вида без указания места, времени и его числового значения, т.е. дает чисто качественную характеристику явления. Но после привязки к конкретному месту (объекту) и времени, получив количественное выражение, он становится ***конкретным статистическим показателем***. Например, «численность населения» — это качественное определение, т.е. показатель-категория. Но если мы скажем, что на 31 октября 2011 г. численность населения Земли достигла 7,0 млрд чел., то это будет уже конкретный статистический показатель.

Статистические показатели обладают целым рядом свойств, характеризующих различные стороны понятия «показатель» как единого целого. Эти свойства классифицируются следующим образом.

Так, по охвату единиц совокупности показатели разделяются на индивидуальные и сводные (общие). ***Индивидуальные показатели*** отражают отдельные явления или отдельную единицу совокупности (банк, предприятие, ферма, отдельный человек и т.п.). ***Сводные (общие) показатели*** характеризуют группу единиц, представляющих собой часть статистической совокупности или всю совокупность в целом (совокупность предприятий, совокупность банков, совокупность ферм и т.п.).

Эти показатели в свою очередь подразделяются на объемные и расчетные. ***Объемные показатели*** получаются путем сложения значений признака отдельных единиц совокупности (например, объем реализованной продукции промышленной фирмы и т.д.). ***Расчетные показатели*** получаются как функции нескольких величин. Они вычисляются по различным формулам и служат для решения отдельных статистических задач анализа — измерения взаимосвязи, вариации, характеристики структурных сдвигов и т.д.

По временно́му фактору показатели делятся на моментные и интервальные. Дело в том, что социально-экономические явления и процессы находят свое выражение в статистических показателях либо по состоянию на определенный момент времени, как правило, на определенную дату, начало или конец месяца, года (численность населения, дебиторская задолженность, остатки товаров в магазинах), либо за определенный период — день, месяц, квартал, год (число заключенных браков, число вкладов населения, производство продукции). Соответственно, в одном случае показатели называются ***моментными***, в другом — ***интервальными***.

С точки зрения пространственной определенности показатели подразделяются на ***федеральные***, характеризующие изучаемый объект или явление в целом по стране, ***региональные*** и ***местные*** (***локальные***), относящиеся к какой-либо части территории или отдельному объекту.

С точки зрения свойств конкретных объектов и формы выражений показатели делятся на абсолютные, относительные и средние.

В статистическом ежегоднике Федеральной службы государственной статистики «Россия в цифрах. 2011 г.» приводятся многочисленные данные по национальному хозяйству нашей страны в виде абсолютных и относительных величин. Так, в частности, в нем мы читаем, что на 14 октября 2010 г. численность населения России составила 142,9 млн чел., в том числе городское население — 105,3 млн чел., сельское — 37,6 млн чел.; число зарегистрированных браков в 2010 г. составило 1 053,7 тыс.

Вот такого рода статистические показатели, выражающие размеры (объемы, уровни) социально-экономических явлений в единицах меры, веса, объема, протяженности, площади, стоимости называются ***абсолютными статистическими величинами***. Они всегда имеют определенную размерность, определенные единицы измерения. Вопрос о единицах измерения, в которых выражаются абсолютные статистические величины чрезвычайно важен для статистического исследования. Выбор единиц измерения абсолютных величин определяется сущностью, свойствами изучаемого явления, а также задачами исследования. В статистике применяется большое число самых разнообразных единиц измерения. В самой общей классификации их можно свести к трем типам: натуральные, денежные (стоимостные) и трудовые.

Натуральными принято называть такие единицы измерения, которые выражаются в мерах веса, объема, длины, площади и т. д. Такими единицами измерения пользуются для характеристики объема различных видов продукции, размеров продажи товаров, мощности электростанций и т. д. Таковы производство тканей — в погонных и (или) квадратных метрах, производство газа — в кубических метрах, электроэнергии — в киловатт-часах.

В ряде случаев применяются ***условно натуральные*** единицы измерения. Они используются для сведения воедино нескольких разновидностей одной и той же потребительской стоимости. Одну из них принимают за эталон, а другие пересчитываются с помощью специальных коэффициентов в единицы меры этого эталона. Так, в практике нашей статистики пересчитываются все виды топлива в условное топливо с теплотой сгорания 29,3 МДж/кг (7 000 ккал/кг). Мыло с различным содержанием жировых кислот пересчитывается на 40%-ное содержание жировых кислот, консервы разного объема — в условные консервные банки объемом 353,4 см^3, грузовые вагоны — в двухосные и т. д. Если, допустим, имеется 100 т мыла с содержанием жировых кислот в 40 % и 100 т с содержанием жировых кислот в 60 %, то, пересчитав на 40%-е мыло, получим $100 + 100 \cdot 60/40 = 250$ условных тонн мыла.

Трудовые единицы измерения такие, как человеко-часы, человеко-дни, используются для определения затрат труда на производство продукции, на выполнение какой-нибудь работы, на учет трудоемкости отдельных операций технологического процесса.

В условиях рыночной экономики большое значение и широкое применение имеют ***стоимостные*** единицы измерения, дающие денежную оценку социально-экономическим явлениям и процессам. Таковыми являтся: валовой внутренний продукт, товарооборот, доходы и расходы населения и др.

Абсолютные статистические показатели подразделяются на показатели объема и показатели уровня.

Показатели объема позволяют характеризовать величину всей совокупности или ее частей. Так, численность экономически активного населения в России в 2010 г. составила 75 448 тыс. чел., в том числе мужчин — 36 870 тыс. чел., женщин — 34 217 тыс. чел. Они могут также выражать суммарную величину какого-либо признака всей совокупности или ее части. В 2010 г. производство электроэнергии всеми электростанциями России составило 992 млрд кВт · ч, в том числе атомными электростанциями 164 млрд кВт · ч.

Показатели уровня характеризуют величину нагрузки единицы одной совокупности элементами другой совокупности (например, в России в 2010 г. число жителей на 1 км^2 территории составило 8,4 чел.). Они могут определять и степень насыщенности конкретной совокупности элементами какого-то признака данной или другой совокупности (в России в 2010 г. величина прожиточного минимума в среднем на душу населения в месяц составила в IV квартале 2010 г. 5 907 руб.).

Существуют также ***разностные*** абсолютные показатели. Они представляют собой абсолютный размер в различии двух абсолютных показателей во времени или в пространстве. Примером абсолютного показателя разности во времени (называемого абсолютным показателем прироста) может служить разность между производством кондитерских изделий в России в 2009 г. (2 779 тыс. т) и в 2005 г. (2 419 тыс. т), равная 360 тыс. т. На эту величину за пять лет увеличился абсолютный размер производства кондитерских изделий в России.

Относительными показателями называются статистические показатели, определяемые как отношение сравниваемой абсолютной величины к базе сравнения. Величина, с которой производится сравнение (знаменатель дроби) обычно называется основанием, базой сравнения или базисной величиной. Числитель — сравниваемая величина. Ее называют также текущей или отчетной величиной. Например, разделив численность городского населения на всю численность населения страны, получаем показатель «доля городского населения». Сопоставляемые величины могут быть одноименными и разноименными. Если сравниваются одноименные величины, то относительные показатели выражаются в отвлеченных числах. Как правило, базу сравнения принимают равной 1, 100, 1 000 или 10 000. Если основание равно 1, то относительная величина показывает, какую долю от базисной составляет текущая величина. Если база сравнения равна 100, то относительная величина выражена в процентах (%), если база сравнения равна 1 000 — в промилле (‰), 10 000 — в продецимилле (‱).

При сопоставлении разноименных величин наименования относительных величин образуются от наименований сравниваемых величин (плотность населения страны: чел./км^2; урожайность: ц/га и т. д.).

В зависимости от задач, содержания и значения выражаемых количественных соотношений различают относительные показатели планового задания, выполнения плана, динамики, структуры, координации, сравнения, интенсивности, уровня экономического развития.

Относительные показатели планового задания (ОППЗ) используются в целях перспективного планирования деятельности субъектов финансово-хозяйственной сферы, а также для сравнения реально достигнутых результатов с ранее намеченными:

$$\text{ОППЗ} = \frac{\begin{array}{c}\text{Уровень показателя,}\\ \text{запланированный на предстоящий период } (i+1)\end{array}}{\begin{array}{c}\text{Уровень показателя,}\\ \text{достигнутый в предыдущем периоде } (i)\end{array}}.$$

Пример 5.1. В I квартале розничный товарооборот торгового объединения составил 250 млн руб., во II квартале планируется розничный товарооборот в 350 млн руб. Определить относительную величину планового задания.

Решение. ОППЗ $= \frac{350}{250} \cdot 100\,\% = 140\,\%$. Таким образом, во II квартале планируется увеличение розничного товарооборота торгового объединения на 40 %.

Относительные показатели выполнения плана (ОПВП) выражают соотношение между фактическим и плановым уровнями показателя. Обычно они выражаются в процентах. Способ вычисления относительных показателей выполнения плана зависит от того, в каком виде и в какой форме даны показатели плана. Плановые показатели могут быть установлены в виде абсолютных и средних величин. Если плановое задание установлено в виде абсолютных и средних величин, степень выполнения плана определяется путем деления фактически достигнутой величины показателя на величину, предусмотренную планом:

$$\text{ОПВП} = \frac{\begin{array}{c}\text{Уровень, фактически достигнутый}\\ \text{в отчетном периоде}\end{array}}{\begin{array}{c}\text{Уровень, запланированный}\\ \text{на отчетный период}\end{array}} \cdot 100\,\%.$$

Пример 5.2. Фирма согласно плану должна была выпустить продукции в течение квартала на сумму 200 тыс. руб. Фактически же выпустила продукции на 220 тыс. руб. Определить степень выполнения плана выпуска продукции фирмой за квартал.

Решение. ОПВП $= \frac{220}{200} \cdot 100\,\% = 110\,\%$. Следовательно, план выполнен на 110 %, т. е. перевыполнение плана составило 10 %.

Когда план задан в виде относительного показателя (по сравнению с базисным уровнем), выполнение плана определяется из соотношения относительной величины динамики и относительной величины планового задания.

Пример 5.3. Производительность труда в промышленности региона по плану на 2011 г. должна была возрасти на 2,9 %. Фактически производительность труда увеличилась на 3,6 %. Определить степень выполнения плана по производительности труда регионом.

Решение. ОПВП $= \frac{1{,}036}{1{,}029} \cdot 100\,\% = 100{,}7\,\%$. Следовательно, достигнутый в 2011 г. уровень производительности труда выше запланированного на 0,7 %.

Если плановое задание предусматривает снижение уровня показателя, то результат сравнения фактического уровня с запланированным, составивший по своей величине менее 100 %, будет свидетельствовать о перевыполнении плана.

Относительными показателями динамики (ОПД) называют статистические величины, характеризующие степень изменения изучаемого явления во времени. Они представляют собой отношение уровня исследуемого процесса или явления за данный период времени и уровня этого же процесса или явления в прошлом:

$$\text{ОПД} = \frac{\begin{array}{c}\text{Уровень, фактически сложившийся}\\ \text{в текущем периоде}\end{array}}{\begin{array}{c}\text{Уровень, фактически сложившийся}\\ \text{в предшествующем или базисном периоде}\end{array}}.$$

Рассчитанная таким образом величина показывает, во сколько раз текущий уровень превышает предшествующий (базисный) или какую долю от последнего он составляет. Данный показатель может быть выражен в долях или в процентах.

Пример 5.4. Число телефонных станций в России в 2008 г. составило 53 тыс., а в 2009 — 54,7 тыс. Определить относительную величину динамики.

Решение. $\text{ОПД} = \frac{54{,}7}{53{,}2} = 1{,}028$ раза, или 102,8 %. Следовательно, число телефонных станций в 2009 г. увеличилось по сравнению с 2008 г. на 2,8 %.

При наличии данных за несколько периодов времени сравнение каждого данного уровня может производиться либо с уровнем предшествующего периода, либо с каким-то другим, принятым за базу сравнения (базисным уровнем). Первые называются относительными показателями динамики ***с переменной базой сравнения***, или ***цепными***, вторые — относительными показателями динамики ***с постоянной базой сравнения***, или ***базисными***. Относительные показатели динамики иначе называются темпами роста и коэффициентами роста.

Между относительными показателями планового задания, выполнения плана и динамики существует следующая взаимосвязь: ОППЗ · ОПВП = ОПД. Основываясь на этой взаимосвязи, по любым двум известным показателям всегда можно определить третью неизвестную величину.

Относительные показатели структуры (ОПС) представляют собой отношение части и целого. Они характеризуют структуру, состав той или иной совокупности социально-экономических явлений. Из определения относительных показателей структуры следует, что при их исчислении в качестве базы сравнения берется величина целого (общий итог по какому-либо показателю), а сравниваемыми являются значения показателей отдельных частей этого целого:

$$\text{ОПС} = \frac{\text{Уровень части совокупности}}{\text{Суммарный уровень совокупности в целом}}.$$

Обычно относительные показатели этого вида выражают в долях единицы или процентах.

Пример 5.5. По представленным в первых двух графах табл. 5.1 данным рассчитать относительные показатели структуры.

Наименование	Число станций, тыс. шт.	Удельный вес каждой сети в общем итоге, %
Телефонные станции		
В том числе:	54,7	100,0
городские сети	49,0	89,6
сельские сети	5,7	10,4

Рассчитанные в последней графе табл. 5.1 проценты представляют собой относительные показатели структуры (в данном случае удельные веса). Сумма всех удельных весов всегда должна быть строго равна 100 %.

Относительные показатели координации (ОПК) представляют собой соотношение одной части совокупности к другой части этой же совокупности:

$$\text{ОПК} = \frac{\text{Уровень, характеризующий } i\text{-ю часть совокупности}}{\begin{array}{c}\text{Уровень, характеризующий часть совокупности,}\\ \text{выбранную в качестве базы сравнения}\end{array}}.$$

В результате этого деления получают, во сколько раз данная часть совокупности больше (меньше) базисной, или сколько процентов от нее составляет, или сколько единиц данной структурной части приходится на 1 единицу, на 100, на 1 000 и т. д. единиц другой части, принятой за базу сравнения. Так, на основе данных, приведенных в табл. 5.1, мы можем вычислить, что на одну сельскую телефонную сеть приходится 49 : 5,7 = 8,6 городских телефонных сетей. Можно также определить, сколько на 100 рабочих приходится служащих, инженерно-технических работников; сколько на 10 (или 100) инженеров приходится техников.

Относительные показатели интенсивности (ОПИ) характеризуют степень насыщенности или развития данного явления и представляют собой отношение исследуемого показателя к размеру присущей ему среды:

$$\text{ОПИ} = \frac{\text{Уровень, характеризующий явление А}}{\begin{array}{c}\text{Уровень, характеризующий}\\ \text{среду распространения явления А}\end{array}}.$$

Данный показатель получают сопоставлением разноименных, но взаимосвязанных в своем развитии величин. Поэтому, как правило, данный показатель представляет собой именованную величину, но может также быть выражен в долях, процентах, промилле или продецимилле.

Обычно относительный показатель интенсивности рассчитывается в тех случаях, когда абсолютная величина оказывается недостаточной для формулировки обоснованных выводов о масштабах явления, его размерах, насыщенности, плотности распространения. Например, для определения плотности населения рассчитывается число людей, приходящихся на 1 км^2, или для определения уровня обеспеченности населения легковыми автомобилями рассчитывается число автомашин, приходящихся на 100 семей.

Пример 5.6. Среднегодовая численность населения региона в 2011 г. составила 146,5 тыс. чел., численность врачей всех специальностей — 1,1 тыс. Определить число врачей, приходящихся на каждые 10 000 чел. населения.

Решение.

$$\text{ОПИ} = \frac{\text{Число врачей}}{\text{Среднегодовая численность населения}} \cdot 10\,000\,‱ =$$

$$= \frac{1{,}1}{146{,}5}\quad 10\,000\,‱ = 75{,}1\,‱.$$

Разновидностью относительных показателей интенсивности являются ***относительные показатели уровня экономического развития*** (ОПУЭР). Они характеризуют выпуск продукции в расчете на душу населения и весьма значимы при оценке состояния экономики государства.

Поскольку объемные показатели производства по своей природе являются интервальными, а показатель численности населения —

моментным, в расчете используют среднюю за период численность населения (например, среднегодовую):

$$\text{ОПУЭР} = \frac{\text{Объем производства какой-либо продукции за год}}{\text{Среднегодовая численность населения}}.$$

Пример 5.7. Имеются данные о производстве некоторых видов промышленной продукции в Российской Федерации в натуральном выражении в 2009 г.: стиральные машины, млн шт. — 2,8; электропылесосы, тыс. шт. — 203; электрочайники, тыс. шт. — 162. Требуется определить относительные показатели уровня экономического развития, зная, что численность постоянного населения России на начало 2009 г. составила 141,9 млн чел. и на начало 2010 г. — 142,9 млн чел.

Решение. Сначала определим среднегодовую численность населения в 2009 г.: $\frac{141{,}9 + 142{,}9}{2} = 142{,}4$ млн чел. Далее, разделив объем производства каждого вида продукции на 146,5 млн чел. и умножив на 1 000, получим, что на 1 000 жителей в 2010 г. произведено: 20 стиральных машин, 1 электропылесос, 1 электрочайник.

Относительные показатели сравнения (ОПСр) представляют собой отношение одноименных величин, относящихся к разным объектам (предприятиям, фирмам, районам, областям, странам и т. п.):

$$\text{ОПСр} = \frac{\text{Абсолютный показатель, характеризующий объект А}}{\text{Абсолютный показатель, характеризующий объект Б}}.$$

С помощью такого показателя можно сравнивать численность населения, размеры территории, величину посевных площадей по странам, областям, районам и т. д.

Пример 5.8. Требуется сравнить численность населения Великобритании, Австрии и Бельгии. Как видим из табл. 5.2, в 2010 г. в Великобритании численность населения была в 7,4 раза больше, чем в Австрии (61,8/8,4), и в 5,7 раза больше, чем в Бельгии (61,8/10,8).

Страны	Численность населения, млн чел.	Отношение численности населения Великобритании к численности населения других стран, раз
Австрия	8,4	7,3
Бельгия	10,8	5,8
Великобритания	61,8	—

1. Показатели, выражающие размеры, объем, уровни социально-экономических явлений и процессов, являются величинами:
   а) абсолютными;
   б) относительными.

2. Абсолютные величины могут выражаться в единицах измерения:
   а) натуральных и условно-натуральных;
   б) трудовых и денежных;
   в) отвлеченных.

3. Абсолютные величины выражаются в единицах измерения:
   а) килограммах, штуках, метрах, тоннах, километрах и т. д.;
   б) коэффициентах, процентах, промилле, продецимилле.

4. Виды абсолютных величин:
   а) индивидуальные, общие;
   б) выполнение плана, планового задания, динамики, структуры, координации, сравнения, интенсивности.

5. Объемные абсолютные величины получаются в результате:
   а) сложения индивидуальных абсолютных величин;
   б) подсчета числа единиц, входящих в каждую группу или совокупность в целом.

6. Относительные величины выполнения плана исчисляются как:
   а) отношение планового задания на предстоящий период к фактически достигнутому уровню, являющемуся базисным для плана;
   б) отношение фактически достигнутого уровня к плановому заданию за тот же период времени.

7. Относительные величины динамики получаются в результате сопоставления показателей каждого последующего периода:
   а) с предыдущим;
   б) с первоначальным;
   в) со средним.

8. Относительные величины структуры:
   а) характеризуют состав явления и показывают, какой удельный вес в общем итоге составляет каждая его часть;
   б) показывают соотношение отдельных составных частей целого явления.

9. Относительные величины интенсивности представляют собой:
   а) отношение двух разноименных показателей, находящихся в определенной взаимосвязи;
   б) отношение двух одноименных показателей, относящихся к разным объектам или территориям за один и тот же период или момент времени.

10. Укажите относительную величину уровня экономического развития:
    а) в одном из регионов на душу населения было произведено 760 м^3 газа;
    б) производство хлопчатобумажных тканей на душу населения в одном из регионов в 2,3 раза больше, чем в другом.

11. Имеются следующие данные о производстве муки в одном из регионов:

Показатель	2008	2009	2010	2011
Произведено муки, млн т	11,5	9,6	10,9	11,2

Вычислите относительные показатели динамики с переменной и постоянной базой сравнения. Проверьте их взаимосвязь.

12. Известны следующие данные о производстве легковых автомашин в I полугодии 2011 г. в одном из регионов:

Показатель	январь	февраль	март	апрель	май	июнь
Объем производства, в % к декабрю 2010 г.	100,6	116,1	120,8	125,6	112,9	120,6

Вычислите относительные показатели динамики с переменной базой сравнения. Сделайте выводы.

13. Произведенные затраты металлургического комбината за год составили:

Статья затрат	Объем затрат, млн руб.
Сырье и материалы	280,5
Топливо и энергия	110,5
Оплата труда	34,0
Амортизация	85,0
Прочие расходы	340,0
Итого	850,0

Вычислите относительные показатели структуры и координации.

14. Имеются данные о ценах на спортивные детские товары за отчетный период в условных единицах за штуку:

Вид товара	Зарубежное производство	Отечественное производство
Костюм спортивный	170,5	92
Футболка	20,6	15,1
Куртка спортивная	160,2	142,5

Определите относительные показатели сравнения условных цен по каждому виду товаров.

Наиболее распространенной формой статистических показателей, используемой в экономических исследованиях, является ***средняя величина***, представляющая собой обобщенную количественную характеристику признака в статистической совокупности. Средняя величина дает обобщающую характеристику однотипных явлений по одному из варьирующих признаков. Она отражает уровень этого признака, отнесенный к единице совокупности. Широкое применение средних объясняется тем, что они имеют ряд положительных свойств, делающих их незаменимым инструментом анализа явлений и процессов в экономике.

Важнейшее свойство средней величины заключается в том, что она отражает то общее, что присуще всем единицам исследуемой совокупности. Значения признака отдельных единиц совокупности колеблются в ту или иную сторону под влиянием множества факторов, среди которых могут быть как основные, так и случайные. Например, курс акций корпорации в целом определяется ее финансовым положением. В то же время в отдельные дни и на отдельных биржах эти акции в силу сложившихся обстоятельств могут продаваться по более высокому или заниженному курсу. 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 основных факторов. Это позволяет средней абстрагироваться от индивидуальных особенностей, присущих отдельным единицам.

Средняя величина только тогда будет отражать типичный уровень признака, когда она рассчитана по качественно однородной совокупности. Так, если мы рассчитаем средний курс по акциям всех предприятий, реализуемых в данный день на данной бирже, то получим фиктивную среднюю. Это объясняется тем, что используемая для расчета совокупность 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.

Категорию средней можно раскрыть через понятие ее ***определяющего свойства***. Согласно этому понятию средняя, будучи обобщающей характеристикой всей совокупности, должна ориентироваться на определенную величину, связанную со всеми единицами этой совокупности. Эту величину можно представить в виде функции: $f(x_1, x_2, ..., x_n)$.

Так как данная величина в большинстве случаев отражает реальную экономическую категорию, понятие определяющего свойства средней иногда заменяют понятием определяющего показателя.

Если в приведенной выше функции все величины $x_1, x_2, ..., x_n$ заменить их средней величиной $\bar{x}$, то значение этой функции должно остаться прежним:

$$f(x_1, x_2, ..., x_n) = f(\bar{x}, \bar{x}, ..., \bar{x}). \quad (6.1)$$

Исходя из данного равенства, и определяется средняя. На практике 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:

$$\text{ИСС} = \frac{\text{Фонд заработной платы (тыс. руб.)}}{\text{Число работников (чел.)}}.$$

Числитель исходного соотношения средней представляет собой ее определяющий показатель. Для средней заработной платы таким определяющим показателем является фонд заработной платы. Независимо от того, какой первичной информацией мы располагаем — известен ли нам общий фонд заработной платы или заработная плата и численность работников, занятых на отдельных должностях, или какие-либо другие исходные данные — в любом случае среднюю заработную плату можно получить только через данное исходное соотношение средней.

Для каждого показателя, используемого в 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:

$$\text{ИСС} = \frac{\text{Сумма всех вкладов (тыс. руб.)}}{\text{Число вкладов}}.$$

Если же необходимо определить среднюю процентную ставку по кредитам, выданным на один и тот же срок, то потребуется следующее исходное соотношение:

$$\text{ИСС} = \frac{\begin{array}{c}\text{Общая сумма выплат по процентам}\\ \text{(из расчета за год, тыс. руб.)}\end{array}}{\text{Общая сумма предоставленных кредитов (тыс. руб.)}}.$$

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 один из видов средней величины. Это может быть средняя арифметическая, средняя гармоническая, средняя геометрическая, средняя квадратическая, кубическая и т. д.

Перечисленные средние объединяются в общей формуле ***средней степенной*** (при различной величине c):

$$\bar{x} = \sqrt[c]{\frac{\sum_{i=1}^{k} x_i^c m_i}{\sum_{i=1}^{k} m_i}}, \qquad (6.2)$$

где x_i — i-й вариант рассматриваемого признака ($i = \overline{1, k}$); m_i — удельный вес i-го варианта.

Помимо степенных средних в экономической практике также используются средние структурные, среди которых наиболее распространены мода и медиана. При осреднении уровней динамических рядов применяются различные виды средней хронологической.

Наиболее распространенным видом средних величин является средняя арифметическая, которая, как и все средние, в зависимости от характера имеющихся данных может быть простой или взвешенной.

Экономический показатель	Торговый центр (i)				
	1	2	3	4	5
Товарооборот (млн руб.) x_i	130	142	125	164	127

Средняя арифметическая простая. Эта форма средней используется в тех случаях, когда расчет осуществляется по несгруппированным данным. Предположим, пять торговых центров фирмы имеют следующий объем товарооборота за месяц (табл. 6.1).

Для того чтобы определить средний месячный товарооборот в расчете на один центр, необходимо воспользоваться следующим исходным соотношением:

$$\text{ИСС} = \frac{\text{Общий объем товарооборота (млн руб.)}}{\text{Число торговых центров}}.$$

Используя приведенные в предыдущем параграфе условные обозначения, запишем формулу данной средней ($i = \overline{1, n}$):

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{6.3}$$

С учетом имеющихся данных получим:

$$\bar{x} = \frac{130 + 142 + 125 + 164 + 127}{5} = 137{,}6 \text{ млн руб.}$$

В этом примере 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.

Рассмотрим следующий условный пример (табл. 6.2).

Сделка	Количество проданных акций, шт. (m_i)	Курс продажи, руб. (x_i)
11	500	1 080
2	300	1 050
3	1 100	1 145

Определим по данному дискретному вариационному ряду средний курс продажи одной акции, что можно сделать, только используя следующее исходное соотношение:

$$\text{ИСС} = \frac{\text{Общая сумма сделок (руб.)}}{\text{Количество проданных акций (шт.)}}.$$

Чтобы получить общую сумму сделок, необходимо по каждой сделке курс продажи умножить на количество проданных акций и полученные произведения сложить. В конечном итоге мы будем иметь следующий результат:

$$\bar{x} = \frac{1\,080 \cdot 500 + 1\,050 \cdot 300 + 1\,145 \cdot 1100}{500 + 300 + 1100} = \frac{2\,114\,500}{1\,900} = 1\,112{,}9 \text{ руб.}$$

Расчет среднего курса продажи произведен по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i m_i}{\sum_{i=1}^{k} m_i}, \qquad (6.4)$$

где k — число вариантов ($i = 1, 2, ..., k$).

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 приведенной выше формулы получим:

$$\bar{x} = \sum_{i=1}^{k} \left(x_i \frac{m_i}{\sum_{i=1}^{k} m_i} \right) \qquad (6.5)$$

или $\bar{x} = 1\,080 \cdot 0{,}263 + 1\,050 \cdot 0{,}158 + 1\,145 \cdot 0{,}579 = 1\,112{,}9$ руб.

На практике наиболее часто встречаемая при расчете средних ошибка заключается в игнорировании весов в тех случаях, когда эти веса в действительности необходимы. Предположим, имеются следующие данные (табл. 6.3).

Оптовый рынок	Средняя цена (руб./шт.)
1	43
2	42

Можно ли по имеющимся данным определить среднюю цену данного товара по двум рынкам, вместе взятым? Можно, но только в том случае, когда объемы реализации этого товара на двух рынках совпадают. Тогда средняя цена реализации составит 42 руб. (доказательство этого правила будет приведено ниже). Однако на первом рынке может быть реализовано, например, 100 единиц товара, а на втором — 1 000 единиц. Тогда для расчета средней цены потребуется уже средняя арифметическая взвешенная:

$$\bar{x} = \frac{43 \cdot 100 + 41 \cdot 1\,000}{100 + 1\,000} = 41{,}2 \text{ руб.}$$

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).

Для определения средней прибыли в расчете на одно предприятие найдем середины интервалов. При этом величины открытых интервалов (первого и последнего) условно приравниваются к величинам интервалов, примыкающих к ним (второго и предпоследнего). С учетом этого середины интервалов будут следующими:

$$15, \quad 25, \quad 35, \quad 50, \quad 70, \quad 90.$$

Используя среднюю арифметическую взвешенную, определим среднюю прибыль предприятий отрасли:

$$\bar{x} = \frac{15 \cdot 7 + 25 \cdot 13 + 35 \cdot 38 + 50 \cdot 42 + 70 \cdot 16 + 90 \cdot 5}{7 + 13 + 38 + 42 + 16 + 5} = 44{,}9 \text{ млн руб.}$$

Прибыль, млн руб.	Число предприятий
До 20	7
20—30	13
30—40	38
40—60	42
60—80	16
80 и более	5
Итого	121

Свойства средней арифметической. Средняя арифметическая обладает некоторыми математическими свойствами, более полно раскрывающими ее сущность. В ряде случаев они используются при ее расчете. Рассмотрим эти свойства.

1. Произведение средней на сумму частот равно сумме произведений отдельных вариантов на соответствующие частоты (i-й группы):

$$\bar{x} \cdot \sum_{i=1}^{k} m_i = \sum_{i=1}^{k} x_i m_i. \qquad (6.6)$$

Действительно, если мы обратимся к приведенному выше примеру расчета среднего курса продажи акций, то получим следующее равенство (за счет округления среднего курса правая и левая части равенства в данном случае будут незначительно отличаться):

$$1\,112{,}9 \cdot 1\,900 = 1\,080 \cdot 500 + 1\,050 \cdot 300 + 1\,145 \cdot 1\,100.$$

2. Сумма отклонений индивидуальных значений признака от средней арифметической равна нулю:

$$\sum_{i=1}^{k} (x_i - \bar{x}) m_i = 0. \qquad (6.7)$$

Для нашего примера:

$$(1\,080 - 1\,112{,}9) \cdot 500 + (1\,050 - 1\,112{,}9) \cdot 300 + (1\,145 - 1\,112{,}9) \cdot 1\,100 = 0.$$

Математическое доказательство данного свойства сводится к следующему:

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

3. Сумма квадратов отклонений индивидуальных значений признака от средней арифметической меньше, чем сумма квадратов их отклонений от любой другой произвольной величины C:

$$\sum_{i=1}^{k} (x_i - C)^2 m_i = \sum_{i=1}^{k} (x_i - \bar{x} + \bar{x} - C)^2 m_i = \sum_{i=1}^{k} \left[(x_i - \bar{x}) + (\bar{x} - C) \right]^2 m_i =$$

$$= \sum_{i=1}^{k} \left[(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - C) + (\bar{x} - C)^2 \right] m_i =$$

$$= \sum_{i=1}^{k} (x_i - \bar{x})^2 m_i + 2(\bar{x} - C) \sum_{i=1}^{k} (x_i - \bar{x}) m_i + \sum_{i=1}^{k} (\bar{x} - C)^2 m_i =$$

$$= \sum_{i=1}^{k} (x_i - \bar{x})^2 m_i + 2(\bar{x} - C) \cdot 0 + \sum_{i=1}^{k} (\bar{x} - C) m_i. \qquad (6.8)$$

Следовательно, сумма квадратов отклонений индивидуальных значений признака от произвольной величины C больше суммы квадратов их отклонений от своей средней на величину

$$\sum_{i=1}^{k}(\bar{x}-C)^2 m_i = (\bar{x}-C)^2\sum_{i=1}^{k}m_i.$$

На использовании этого свойства базируется расчет центральных моментов, представляющих собой характеристики вариационного ряда при $C = \bar{x}$[1]:

$$\mu_k = \frac{\sum_{i=1}^{l}(x_i-\bar{x})^k m_i}{\sum_{i=1}^{l}m_i}, \qquad (6.9)$$

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:

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

Так, если все курсы продажи акций увеличить на 100 руб., то средний курс также увеличится на 100 руб.:

$$\bar{x} = \frac{1180 \cdot 500 + 1150 \cdot 300 + 1\,245 \cdot 1100}{1\,900} = 1\,212{,}9 \text{ руб.}$$

5. Если все варианты значений признака изменить в A раз, то средняя также изменится в A раз:

$$\frac{\sum_{i=1}^{k}\frac{x_i}{A}m_i}{\sum_{i=1}^{k}m_i} = \frac{\frac{1}{A}\sum_{i=1}^{k}x_i m_i}{\sum_{i=1}^{k}m_i} = \frac{1}{A}\bar{x}. \qquad (6.10)$$

Предположим, что курс продажи в каждом случае возрастет в 1,5 раза. Тогда и средний курс также увеличится в 1,5 раза:

---

[1] При $C = 0$ получают начальные моменты (начальный момент 1-го порядка — средняя арифметическая и т. д.).

$$\bar{x} = \frac{1\,800 \cdot 1{,}5 \cdot 500 + 1\,050 \cdot 1{,}5 \cdot 300 + 1\,145 \cdot 1{,}5 \cdot 1100}{1\,900} =$$

$$= 1\,112{,}9 \cdot 1{,}5 = 1\,669{,}4 \text{ руб.}$$

Необходимо также отметить, что если все веса равны между собой, то средняя арифметическая взвешенная совпадает со средней арифметической простой.

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. Данная форма используется, когда известен числитель исходного соотношения средней, но неизвестен его знаменатель. Рассмотрим расчет средней урожайности, являющейся одним из основных показателей эффективности сельскохозяйственного производства (табл. 6.5).

Средняя урожайность любой сельскохозяйственной культуры по нескольким территориям, агрофирмам, фермерским хозяйствам может быть определена только на основе следующего исходного соотношения:

$$\text{ИСС} = \frac{\text{Общий валовый сбор (тыс. ц)}}{\text{Общая посевная площадь (тыс. га)}}.$$

Область	Валовый сбор, тыс. т	Урожайность, ц/га
Белгородская	97	16,1
Воронежская	204	9,5
Курская	0,5	4,8
Липецкая	16	10,9
Тамбовская	69	7,0

* Цифры приблизительные.

Общий валовой сбор мы получим простым суммированием валового сбора по областям. Данные же о посевной площади 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:

$$\bar{x} = \frac{970 + 2\,040 + 5 + 160 + 690}{\dfrac{970}{16{,}1} + \dfrac{2\,040}{9{,}5} + \dfrac{5}{4{,}8} + \dfrac{160}{10{,}9} + \dfrac{690}{7{,}0}} = \frac{3\,865}{389{,}3} = 9{,}9 \text{ ц/га}.$$

Таким образом, общая посевная площадь подсолнечника по Центрально-Черноземному району составляла 389,3 тыс. га, а средняя урожайность — 9,9 ц с одного гектара.

В данном примере расчет произведен по формуле средней гармонической взвешенной:

$$\bar{x} = \frac{\sum_{i=1}^{k} w_i}{\sum_{i=1}^{k} \dfrac{w_i}{x_i}}, \quad \text{где } w_i = x_i m_i. \tag{6.11}$$

Данная формула используется для расчета средних показателей не только в статике, но и в динамике, когда известны индивидуальные значения признака и веса w за ряд временны́х интервалов.

Средняя гармоническая невзвешенная. Эта форма средней имеет следующий вид:

$$\bar{x} = \frac{n}{\sum_{i=1}^{k} \dfrac{1}{x_i}}. \tag{6.12}$$

Для иллюстрации области ее применения воспользуемся условным примером. Пусть упаковкой и отправкой товаров занимаются два работника фирмы, специализирующейся на торговле по почте. Первый из них на обработку одного заказа затрачивает 8 мин, второй — 14 мин. Каковы средние затраты времени на 1 заказ, если общая продолжительность рабочего времени у работников равна?

На первый взгляд, ответ на этот вопрос заключается в осреднении индивидуальных значений затрат времени на 1 заказ, т. е. (8 + 14) : 2 = 11 мин. Проверим обоснованность такого подхода на примере одного часа работы. За этот час первый работник обрабатывает 7,5 заказов (60 : 8), второй — 4,3 заказа (60 : 14). В сумме это составляет 11,8 заказа. Если же заменить индивидуальные значения их предполагаемым средним значением, то общее число обработанных обоими работниками заказов в данном случае уменьшится:

$$\frac{60}{11}+\frac{60}{11}=10,9.$$

Подойдем к решению через исходное соотношение средней. Для определения средних затрат времени необходимо общие затраты времени разделить на общее число обработанных за этот интервал двумя работниками заказов:

$$\bar{x}=\frac{60+60}{\frac{60}{8}+\frac{60}{14}}=\frac{1+1}{\frac{1}{8}+\frac{1}{14}}=\frac{2}{0{,}125+0{,}071}=10{,}2 \text{ мин.}$$

Если теперь мы заменим индивидуальные значения их средней величиной, то общее количество обработанных за 1 час заказов не изменится:

$$\frac{60}{10{,}2}+\frac{60}{10{,}2}=11{,}8.$$

Таким образом, средняя гармоническая невзвешенная может использоваться вместо взвешенной в тех случаях, когда значения w_i для единиц совокупности равны (рабочий день у сотрудников одинаковый).

Средняя геометрическая. Еще одной формулой, по которой может осуществляться расчет среднего показателя, является средняя геометрическая. Сначала обратимся к формуле невзвешенной средней геометрической. Она выглядит следующим образом:

$$\bar{x}=\sqrt[k]{x_1 x_2 x_3 \ldots x_k}=\sqrt[k]{\prod_{i=1}^{k} x_i}. \qquad (6.13)$$

Соответственно средняя геометрическая взвешенная приобретает следующее выражение:

$$\bar{x}=\sqrt[\sum m]{x_1^{m_1} x_2^{m_2} x_3^{m_3} \ldots x_k{}^{m_k}}=\sqrt[\sum m]{\prod_{i=1}^{k} x_i^{m_i}}.$$

Наиболее широкое применение этот вид средней получил в анализе динамики для определения среднего темпа роста, что будет рассмотрено в гл. 10.

Средняя квадратическая. В основе вычислений ряда сводных расчетных показателей лежит средняя квадратическая. Формула невзвешенной средней квадратической достаточно проста:

$$\bar{x} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}. \qquad (6.14)$$

Взвешенная средняя квадратическая:

$$\bar{x} = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 m_i}{\sum_{i=1}^{k} m_i}}.$$

Наиболее широко этот вид средней используется при расчете показателей вариации.

В статистическом анализе также применяются степенные средние 3-го порядка и более высоких порядков.

Наиболее часто используемыми в экономической практике структурными характеристиками являются мода и медиана. ***Мода*** представляет собой значение изучаемого признака, повторяющееся с наибольшей частотой. ***Медианой*** называется значение признака, приходящееся на середину ранжированной (упорядоченной) совокупности. Основное свойство медианы заключается в том, что сумма абсолютных отклонений значений признака от медианы меньше, чем от любой другой величины:

$$\sum_{i=1}^{n} \left| x_i - \mathrm{M}_e \right| = \min.$$

Определим моду и медиану по ***несгруппированным данным***.

Предположим, что 9 торговых фирм города реализуют товар А по следующим оптовым ценам (тыс. руб.): 4,4; 4,3; 4,4; 4,5; 4,3; 4,3; 4,6; 4,2; 4,6. Как видим, чаще всего встречается цена 4,3 тыс. руб. Она и будет модальной. Для определения медианы необходимо провести ранжирование приведенного цифрового ряда: 4,2; 4,3; 4,3; 4,3; 4,4; 4,4; 4,5; 4,6; 4,6.

Центральной в этом ряду является цена 4,4 тыс. руб. Следовательно, данная цена и будет медианой. Если ранжированный ряд

№ п/п	1	2	3	4	5	6	7	8	9	10
Доход, тыс. руб.	10	10	11	12	14	15	15	17	20	500

включает четное число единиц, то медиана определяется как средняя из двух центральных значений.

Если мода отражает типичный, наиболее распространенный вариант значения признака, то медиана выполняет функции средней для неоднородной совокупности. В этих случаях средняя не позволяет объективно оценить исследуемую совокупность вследствие сильного влияния аномальных максимальных или минимальных значений. Проиллюстрируем сказанное следующим примером.

Допустим, нам необходимо дать обобщающую характеристику среднедушевых доходов группы людей, насчитывающей десять человек, из которых девять имеют доходы в интервале от 10 до 20 тыс. руб. в месяц, а месячные доходы последнего составляют 500 тыс. руб. (табл. 6.6).

Если мы воспользуемся средней арифметической, то получим средний доход, равный примерно 62,4 руб., что не только почти в 8 раз меньше дохода 10-го человека, но и имеет мало общего с доходами остальной части группы. Медиана же, равная в данном случае 14,5 руб., позволит дать объективную характеристику уровня доходов 90 % данной совокупности людей.

Теперь рассмотрим определение моды и медианы по ***сгруппированным данным*** (рядам распределения). Предположим, распределение торговых предприятий города по уровню розничных цен на товар А имеет следующий вид (табл. 6.7).

Определение моды по дискретному вариационному ряду не составляет большого труда — наибольшую частоту (60 предприятий) имеет цена 55 руб., следовательно, она и является модальной.

Цена, руб.	Число торговых предприятий
52	12
53	48
54	56
55	60
56	14
Итого	190

Для определения медианного значения признака по следующей формуле находят номер медианной единицы ряда:

$$N_{me} = \frac{n+1}{2},$$

где n — объем совокупности.

В нашем случае $N_{me} = \frac{190+1}{2} = 95{,}5$.

Полученное дробное значение, всегда имеющее место при четном числе единиц в совокупности, указывает, что точная середина находится между 95 и 96 предприятиями. Необходимо определить, в какой группе находятся предприятия с этими порядковыми номерами. Это можно сделать, рассчитав накопленные частоты. Очевидно, что магазинов с этими номерами нет в первой группе, где всего лишь 12 торговых предприятий, нет их и во второй группе (12 + 48 = 60). Что касается 95-го и 96-го предприятий, то они находятся в третьей группе (12 + 48 + 56 = 116) и, следовательно, медианой является цена 54 руб.

В отличие от дискретных вариационных рядов определение моды и медианы по ***интервальным рядам*** требует проведения определенных расчетов на основе следующих формул.

Первая из них:

$$M_0 = x_0 + \frac{h(m_{M_0} - m_{M_0-1})}{(m_{M_0} - m_{M_0-1}) + (m_{M_0} - m_{M_0+1})}, \tag{6.15}$$

где x_0 — нижняя граница модального интервала (модальным называется интервал, имеющий наибольшую частоту); h — ширина модального интервала; m_{M_0} — частота модального интервала; m_{M_0-1} — частота интервала, предшествующего модальному; m_{M_0+1} — частота интервала, следующего за модальным.

Вторая:

$$M_e = x_0 + \frac{h\left(\frac{1}{2}\sum_{i=1}^{k} m_i - S_{M_{e-1}}\right)}{m_{M_e}}, \tag{6.16}$$

где x_0 — нижняя граница медианного интервала (медианным называется первый интервал, накопленная частота которого превышает половину общей суммы частот); h — ширина медианного интервала; $S_{M_{e-1}}$ — накопленная частота интервала, предшествующего медианному; m_i — частота i-го интервала, i = 1, 2, ..., k; m_{M_e} — частота медианного интервала.

Проиллюстрируем применение этих формул, используя данные табл. 6.8.

Среднедушевой денежный доход (в среднем за месяц), руб.	Численность населения, млн чел.
До 4 000	22,1
4 000 — 6 000	27,8
6 000 — 8 000	25,2
8 000 — 10 000	19,6
10 000 — 12 000	14,3
12 000 — 16 000	17,6
16 000 — 20 000	9,0
20 000 и более	11,1
Итого	146,7

Информация, подобная представленной в табл. 6.8, необходима для получения обоснованного представления об уровне жизни населения страны и региона, о его покупательной способности, для оценки эластичности спроса и, в конечном итоге, для выбора того или иного метода ценообразования и обоснования окончательной цены на товар.

Интервал с границами 4 000 — 6 000 в данном распределении будет модальным, так как он имеет наибольшую частоту. Используя формулу (6.15), определим моду:

$$M_0 = 4\,000 + \frac{2\,000 \cdot (27{,}8 - 22{,}1)}{(27{,}8 - 22{,}1) + (27{,}8 - 25{,}2)} = 5\,373 \text{ руб.}$$

Для определения медианного интервала необходимо рассчитывать накопленную частоту каждого последующего интервала до тех пор, пока она не превысит 1/2 суммы накопленных частот (в нашем случае — 73,35) (табл. 6.9).

Интервал	Накопленная частота, млн чел.
До 4 000	22,1
4 000 — 6 000	49,9
6 000 — 8 000	75,1

Таким образом, медианным является интервал с границами 6 000 — 8 000. Тогда медиана равна:

$$M_e = 6\,000 \quad 2\,000 \frac{73{,}35 - 49{,}9}{25{,}2} = 7\,861 \text{ руб.}$$

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

На основе полученных в последнем примере значений структурных средних можно заключить, что наиболее распространенным, типичным является среднедушевой доход порядка 5 373 руб. в месяц. В то же время более половины населения располагает доходом свыше 7 861 руб. при среднем уровне 9 363 руб. (средняя арифметическая взвешенная). Из соотношения этих показателей следует вывод о правосторонней асимметрии распределения населения по уровню среднедушевых денежных доходов.

Информация о средних уровнях исследуемых показателей обычно бывает недостаточной для глубокого анализа изучаемого процесса или явления. Поэтому необходимо учитывать и вариацию значений отдельных единиц относительно средней, которая является важной характеристикой изучаемой совокупности. Значительной вариации подвержены курсы акций, объемы спроса и предложения, процентные ставки в разные периоды времени.

Основными показателями, характеризующими вариацию, являются размах, дисперсия, среднее квадратическое отклонение и коэффициент вариации.

Для иллюстрации расчетов этих показателей воспользуемся следующими данными (табл. 6.10).

Простейшим показателем, уже использованным выше при группировке данных, является ***размах вариации***. Он представляет собой разность максимального и минимального значений признака:

$$R = x_{\max} - x_{\min} = 22{,}83 - 22{,}40 = 0{,}43 \text{ руб.}$$

Недостатком данного показателя является то, что он оценивает только границы варьирования признака и не отражает его колебание внутри этих границ. Этого недостатка лишена ***дисперсия***, рассчитываемая как средний квадрат отклонений значений признака от их

Биржа	Курс, руб./долл. США	Оборот, млн долл. США
ММВБ	22,73	158,0
СПВБ	22,63	10,0
УРВБ	22,42	3,0
СМВБ	22,40	2,9
АТМВБ	22,64	0,7
СВМБ	22,83	1,6
НФВБ	22,56	0,7

средней величины. Как и средняя величина, дисперсия может рассчитываться по-разному. Здесь перед нами невзвешенная формула:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}. \qquad (6.17)$$

Взвешенная формула используется в тех случаях, когда варианты значений изучаемого признака повторяются. Она имеет следующий вид:

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \overline{x})^2 m_i}{\sum_{i=1}^{k} m_i}. \qquad (6.18)$$

По данным табл. 6.10 определим средневзвешенный курс доллара по итогам всех торгов и рассчитаем дисперсию:

$$\overline{x} = \frac{22{,}73 \cdot 158{,}0 + 22{,}63 \cdot 10{,}0 + \ldots + 22{,}56 \cdot 0{,}7}{158{,}0 + 10{,}0 + \ldots + 0{,}7} = 22{,}71 \text{ руб};$$

$$s^2 = \frac{(22{,}73 - 22{,}71)^2 \cdot 158{,}0 + (22{,}63 - 22{,}71)^2 \cdot 10{,}0 + \ldots + (22{,}56 - 22{,}71)^2 \cdot 0{,}7}{158{,}0 + 10{,}0 + \ldots + 0{,}7} =$$

$$= 0{,}004.$$

Дисперсию в отдельных случаях удобнее рассчитывать по другой формуле, представляющей собой алгебраическое преобразование приведенных выше выражений:

$$s^2 = \overline{x^2} \quad (\overline{x})^2, \text{ где } \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n} \qquad (6.19)$$

или

$$\overline{x^2} = \frac{\sum_{i=1}^{k} x_i^2 m_i}{\sum_{i=1}^{k} m_i}. \qquad (6.20)$$

Другим наиболее широко распространенным показателем является ***среднее квадратическое отклонение***. Оно определяется как квадратный корень из дисперсии и имеет ту же размерность, что и изучаемый признак. Поэтому

невзвешенная формула:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}}; \qquad (6.21)$$

взвешенная формула:

$$s = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 m_i}{\sum_{i=1}^{k} m_i}}. \qquad (6.22)$$

В нашем случае получим: $s = \sqrt{0{,}004}$ 0,06 руб.

Полученная величина показывает, что курсы доллара на биржах отклонялись от средневзвешенного курса в среднем на 6 коп.

Рассмотренные показатели позволяют получить абсолютное значение вариации, т. е. оценивать ее в единицах измерения исследуемого признака. В отличие от них, *коэффициент вариации* измеряет относительную колеблемость (относительно среднего уровня), что во многих случаях намного предпочтительнее:

$$V = \frac{s}{\overline{x}}\ 100\,\%. \qquad (6.23)$$

Определим значение этого показателя по нашим данным:

$$V = \frac{0{,}06}{22{,}71}\ 100\,\% \cdot = 0{,}26\,\%.$$

Рассчитанная величина свидетельствует об очень незначительном относительном уровне колеблемости курса доллара. Если V не превышает 33 %, то совокупность по рассматриваемому признаку можно считать однородной.

Информативность показателей вариации повышается, если они рассчитываются для целей сравнительного анализа. При этом показатели, рассчитанные по одной совокупности, сопоставляются с показателями, рассчитанными по другой аналогичной совокупности или по той же самой, но относящейся к другому периоду времени. Например, исследуется динамика вариации курса доллара по недельным или месячным данным.

Показатели вариации также являются составной частью или служат основой для расчетов других статистических показателей. Они используются в анализе взаимосвязей между признаками, в измерении структурных сдвигов в экономике, в оценке рисков.

1. Возможны ли случаи, когда взвешенные и невзвешенные средние приводят к одному и тому же результату:
   а) возможны;
   б) нет?

2. Могут ли веса средней быть выражены относительными показателями:
   а) могут;
   б) не могут?

3. Может ли одно и то же исходное соотношение быть реализовано на основе различных форм средней:
   а) может;
   б) не может?

4. Можно ли вместо средней арифметической невзвешенной использовать среднюю гармоническую невзвешенную:
   а) нельзя;
   б) можно при отсутствии весов;
   в) можно при равенстве весов?

5. Как изменится средняя величина, если все варианты признака уменьшить в 1,5 раза, а все веса в 1,5 раза увеличить:
   а) не изменится;
   б) уменьшится;
   в) возрастет?

6. Изменится ли средняя величина, если все веса уменьшить на 20 %:
   а) изменится;
   б) не изменится?

7. Изменится ли средняя величина, если все веса уменьшить на некоторую постоянную величину:
   а) изменится;
   б) не изменится?

8. Могут ли мода, медиана и средняя арифметическая совпадать:
   а) могут;
   б) могут совпадать только средняя и медиана;
   в) не могут?.
9. Может ли ряд распределения характеризоваться двумя и более модами:
   а) не может;
   б) может двумя;
   в) может двумя и более?
10. В каких границах изменяется коэффициент вариации:
   а) от 0 до 100 %;
   б) от 0 до 200 %;
   в) нижняя граница — 0 %, верхняя — практически отсутствует?
11. Распределение рабочих предприятия по тарифному разряду имеет следующий вид:

Тарифный разряд	Число рабочих, чел.	Тарифный разряд	Число рабочих, чел.
1	2	4	74
2	3	5	18
3	26	6	4

Определите средний уровень квалификации рабочих предприятия.

12. Имеются следующие данные по фермерским хозяйствам области:

Группы хозяйств по себестоимости 1 ц сахарной свеклы, руб.	Число хозяйств	Валовой сбор в среднем на 1 хозяйство, ц
До 400	32	111,3
400—450	58	89,7
450—500	124	113,5
500 и более	17	130,1

Определите среднюю себестоимость 1 ц свеклы в целом по фермерским хозяйствам области.

13. Качество продукции предприятия характеризуется следующими данными (за месяц):

Вид продукции	Процент брака	Стоимость бракованной продукции, руб.
А	1,3	21 350
Б	0,9	35 600
В	2,4	9 800

Определите средний процент брака в целом по предприятию.

14. Площадь складских помещений города характеризуется следующими данными:

Группы складских помещений по площади, тыс. $м^2$	Число помещений
До 5	3
15—10	21
10—15	17
15—20	9
20—25	5
23—30	4
30—35	4
35 и более	2

Определите модальный и медианный размер складского помещения.

15. Распределение предприятий отрасли по объему полученной за год прибыли имеет следующий вид:

Группы предприятий по прибыли, млн руб.	Число предприятий
До 50	7
50—100	24
100—150	11
150 и более	3

Рассчитайте среднее квадратическое отклонение и коэффициент вариации прибыли предприятий.

Индекс — это относительный показатель, который характеризует изменение исследуемого явления во времени, в пространстве или по сравнению с некоторым эталоном (планируемым, нормативным уровнем и т. п.).

Если изучаемая совокупность включает соизмеримые элементы, то оценить изменение обобщающих показателей можно и без использования индексов. Например, располагая данными о прибыли всех предприятий отрасли за 2010 и 2011 гг., можно рассчитать среднюю прибыль в расчете на одно предприятие в каждом году и вычислить темп роста средней прибыли. Располагая данными о доходах населения, в анализе динамики логично использовать среднедушевые доходы. Список подобных примеров можно продолжить. Индексы же являются незаменимым инструментом исследования в тех случаях, когда необходимо сравнить во времени или в пространстве две совокупности, элементы которых являются несоизмеримыми величинами.

Например, при анализе динамики цен нельзя рассчитать среднюю цену, так как на потребительском рынке реализуются совершенно различные товары — продукты питания, одежда, мебель, транспортные средства, недвижимость. Нельзя рассчитать и среднюю цену продуктов, по крайней мере, из-за различных единиц измерения (килограммы, десятки, штуки, литры). Даже если рассматривать только продукты питания, измеряемые в килограммах, то любому человеку понятно, что «средняя цена 1 кг еды» — очень абстрактная категория, объединяющая мясо, рыбу, масло, картофель, фрукты, овощи и другие подчас несопоставимые продукты питания. Для анализа динамики показателей, характеризующих разнородные совокупности, и используются индексы.

Различают индексы динамические и пространственные (территориальные). Динамические индексы позволяют исследовать изменение одной и той же совокупности во времени, на основе сравнения показателей за два периода и более. Пространственные индексы используются для сравнения показателей по двум совокупностям в пространстве. Это могут быть два предприятия, два региона, две страны. Если в качестве базы сравнения используется уровень за какой-либо предшествующий период — получают динамический индекс, если же базой является уровень того же явления по другой территории — индекс пространственный.

По охвату единиц совокупности индексы делятся на индивидуальные и сводные. Индивидуальные индексы рассчитываются по одной единице — одному товару, одному виду продукции. Сводные же индексы вычисляются по товарным группам или нескольким видам продукции, выпускаемым одним предприятием или всеми предприятиями отрасли. Сводные индексы могут быть представлены в агрегатной, среднеарифметической или среднегармонической формах.

Простейшим показателем, используемым в индексном анализе, является индивидуальный индекс. Он характеризует изменение во времени (или в пространстве) характеристик отдельных элементов той или иной совокупности. Так, ***индивидуальный индекс цены*** рассчитывается по формуле

$$i_p = \frac{p_1}{p_0}, \tag{7.1}$$

где p_1 — цена товара в текущем периоде; p_0 — цена товара в базисном периоде.

Например, если цена товара А в текущем периоде составляла 30 руб., а в базисном — 25 руб., то индивидуальный индекс цены будет равен:

$$i_p = \frac{30}{25} = 1{,}2 \text{ или } 120{,}0\,\%.$$

В данном примере цена товара А возросла по сравнению с базисным уровнем в 1,2 раза или на 20 %.

Оценить изменение объемов продажи товара в натуральных единицах измерения позволяет ***индивидуальный индекс физического объема реализации***:

$$i_q = \frac{q_1}{q_0}, \qquad (7.2)$$

где q_1, q_0 — количество товара, реализованное соответственно в текущем и базисном периодах.

Изменение объема реализации товара в стоимостном выражении отражает ***индивидуальный индекс товарооборота***:

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0}. \qquad (7.3)$$

Приведенные в качестве примеров три индивидуальных индекса взаимосвязаны между собой:

$$i_{pq} = i_p i_q. \qquad (7.4)$$

Данная взаимосвязь показывает, что изменение товарооборота складывается под воздействием динамики цены и изменения объема продажи данного товара.

Индивидуальные индексы, по существу, представляют собой относительные показатели динамики или темпы роста, которые по данным за несколько периодов времени могут рассчитываться в цепной или базисной формах.

Агрегатная форма является исходной формой выражения сводного индекса. При расчете агрегатного индекса для разнородной совокупности находят такой общий показатель, в котором можно объединить все ее элементы. Вернемся к примеру с розничными ценами. Цены различных товаров, реализуемых в розничной торговле, складывать неправомерно, однако с экономической точки зрения вполне допустимо суммировать товарооборот по этим товарам. Если мы сравним товарооборот по n видам товаров в текущем периоде с его величиной в базисном периоде, то получим ***сводный индекс товарооборота***:

$$I_{pq} = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i0}}, \qquad (7.5)$$

где p_{i1} и p_{i0} — цена, а q_{i1} и q_{i0} — объем продаж i-го товара соответственно в текущем и базисном периодах.

Числитель данного индекса представляет собой товарооборот текущего периода (сумма цен различных товаров, умноженных на объемы их реализации), знаменатель — товарооборот предшествующего периода.

На величину индекса товарооборота оказывают влияние как изменение цен на товары, так и изменение объемов их реализации. Для того чтобы оценить изменение только цен (индексируемой величины), необходимо количество проданных товаров (веса индекса) зафиксировать на каком-либо постоянном уровне. При исследовании динамики таких показателей как цена, себестоимость, производительность труда, урожайность количественный показатель обычно фиксируют на уровне текущего периода. Таким способом получают ***сводный индекс цен*** (индекс цен Пааше):

$$I_p = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}}. \qquad (7.6)$$

Числитель данного индекса содержит фактический товарооборот текущего периода. Знаменатель же представляет собой условную величину, показывающую, каким был бы товарооборот в текущем периоде при условии сохранения цен на базисном уровне. Поэтому соотношение этих двух категорий и отражает имевшее место изменение цен[1].

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

Третьим индексом в данной индексной системе является ***сводный индекс физического объема реализации***. Он характеризует изменение количества проданных товаров не в денежных, а в физических единицах измерения:

$$I_q = \frac{\sum_{i=1}^{n} q_{i1} p_{i0}}{\sum_{i=1}^{n} q_{i0} p_{i0}}. \qquad (7.7)$$

[1] Сводный индекс цен можно получить и методом Ласпейерса, фиксируя количество проданного товара на базисном уровне.

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

Между рассчитанными индексами также существует взаимосвязь:

$$I_p I_q = I_{pq}. \tag{7.8}$$

Пример 7.1. Имеются данные (табл. 7.1) о реализации плодово-ягодной продукции в области. Требуется определить индекс товарооборота.

Рассчитаем сводный индекс товарооборота по формуле (7.5):

$$I_{pq} = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \quad \frac{6\,180}{6\,380} = 0{,}969, \text{ или } 96{,}9\,\%.$$

Мы получили, что товарооборот в целом по данной товарной группе в текущем периоде по сравнению с базисным уменьшился на 3,1 % (100 % — 96,9 %).

Вычислим сводный индекс цен по формуле (7.6):

$$I_p = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}} \quad \frac{6\,180}{6\,930} = 0{,}892, \text{ или } 89{,}2\,\%.$$

По данной товарной группе цены в августе по сравнению с июлем в среднем снизились на 10,8 %.

Числитель и знаменатель сводного индекса цен можно интерпретировать с точки зрения потребителей. Числитель представляет собой сумму денег, фактически уплаченных покупателями за приобретенные в текущем периоде товары. Знаменатель же показывает, какую сумму покупатели заплатили бы за те же товары, если бы цены не изменились. Разность

Наименование товара (i)	Июль		Август		Расчетные графы, тыс. руб.		
	Цена за 1 кг, руб. (p_{i0})	Продано, т (q_{i0})	Цена за 1 кг, руб. (p_{i1})	Продано, т (q_{i1})	$p_{i0}q_{i0}$	$p_{i1}q_{i1}$	$p_{i0}q_{i1}$
Черешня(1)	120	18	120	15	2 160	1 800	1 800
Персики (2)	110	22	100	27	2 420	2 700	2 970
Виноград (3)	90	20	70	24	1 800	1 680	2 160
Итого	x	x	x	x	6 380	6 180	6 930

числителя и знаменателя (Е) будет отражать величину экономии («–») или перерасхода («+») покупателей от изменения цен:

$$\text{Е} = \sum_{i=1}^{n} p_{i1}q_{i1} - \sum_{i=1}^{n} p_{i0}q_{i1} = 6\,180 - 6\,930 = -750 \text{ тыс. руб.}$$

Индекс физического объема реализации рассчитывается по формуле (7.7):

$$I_q = \frac{\sum_{i=1}^{n} q_{i1}p_{i1}}{\sum_{i=1}^{n} q_{i0}p_{i0}} \quad \frac{6\,930}{6\,380} = 1{,}086, \text{ или } 108{,}6\,\%.$$

Физический объем реализации (товарооборота) увеличился на 8,6 %.

Используя взаимосвязь индексов, проверим правильность вычислений:

$$I_{pq} = I_p I_q = 0{,}892 \cdot 1{,}086 = 0{,}969, \text{ или } 96{,}9\,\%.$$

Отметим, что объем товарной группы при расчете этих и последующих индексов значения не имеет. Аналогичные расчеты могут быть выполнены для любой товарной группы.

Мы рассмотрели применение агрегатных индексов в анализе товарооборота и цен. При анализе результатов производственной деятельности промышленного предприятия приведенные выше сводные индексы соответственно называются индексом стоимости продукции, индексом оптовых цен и индексом физического объема продукции.

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

$$I_z = \frac{\sum_{i=1}^{n} z_{i1}q_{i1}}{\sum_{i=1}^{n} z_{i0}q_{i1}}, \tag{7.9}$$

где z_{i1} и z_{i0} — себестоимость i-го вида продукции соответственно в текущем и базисном периодах.

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

$$\text{Е} = \sum_{i=1}^{n} z_{i1}q_{i1} - \sum_{i=1}^{n} z_{i0}q_{i1}. \qquad (7.10)$$

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

$$I_q = \frac{\sum_{i=1}^{n} q_{i1}z_{i0}}{\sum_{i=1}^{n} q_{i0}z_{i0}}. \qquad (7.11)$$

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

$$I_{zq} = \frac{\sum_{i=1}^{n} z_{i1}q_{i1}}{\sum_{i=1}^{n} z_{i0}q_{i0}}. \qquad (7.12)$$

Все три индекса взаимосвязаны между собой соотношением:

$$I_z I_q = I_{zq}.$$

Еще одна область применения индексного метода — анализ изменений в производительности труда. При этом возможны два подхода к расчету индексов. Первый основан на учете количества продукции, вырабатываемого в единицу времени (w). При таких расчетах необходимо решить ряд методологических проблем — какой именно показатель продукции использовать, как оценивать продукцию работников непроизводственных отраслей и пр.

При втором подходе производительность труда определяется затратами рабочего времени на единицу продукции (t). На практике эти расчеты также сопряжены с определенными трудностями, так как не всегда имеется возможность оценить вклад конкретного работника в производство того или иного изделия.

Количество продукции w, вырабатываемой в единицу времени (в натуральном выражении), и затраты времени t на единицу продукции взаимосвязаны между собой:

$$w = \frac{1}{t}.$$

Например, если работник на каждое изделие затрачивает 15 мин ($t = 0{,}25$ ч), то за час его выработка составит 4 изделия. Отметим, что выработка может измеряться не только в натуральном, но и в стоимостном выражении (pq).

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

$$i_w = \frac{w_1}{w_0} \quad \frac{q_1}{T_1} \div \frac{q_0}{T_0}; \tag{7.13}$$

$$i_t = \frac{t_0}{t_1} \quad \frac{T_0}{q_0} : \frac{T_1}{q_1}, \tag{7.14}$$

где T — суммарные затраты времени на выпуск данной продукции в чел.-ч, чел.-днях или чел.-мес (в последнем случае соответствует общей численности работников).

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

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

$$I_t = \frac{\sum_{i=1}^{n} t_{i0} q_{i1}}{\sum_{i=1}^{n} t_{i1} q_{i1}}. \tag{7.15}$$

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

Пример 7.2. По данным о производительности труда на предприятии (табл. 7.2) определить индекс производительности труда.

Рассчитаем сводный индекс производительности труда по трудоемкости по формуле (7.15):

$$I_t = \frac{\sum_{i=1}^{n} t_{i0} q_{i1}}{\sum_{i=1}^{n} t_{i1} q_{i1}} \quad \frac{1\,515{,}6}{1\,330{,}6} = 1{,}139 \text{ или } 113{,}9\,\%.$$

Вид продукции (i)	Затраты времени на 1 изделие, чел.-ч		Произведено, шт.		Расчетные графы, чел.-ч	
	январь (t_{i0})	февраль (t_{i1})	январь (q_{i0})	февраль (q_{i1})	$t_{i0}q_{i1}$	$t_{i1}q_{i1}$
Изделие 1	1,0	0,9	458	450	450,0	405,0
Изделие 2	1,2	1,0	311	324	388,8	324,0
Изделие 3	0,9	0,8	765	752	676,8	601,6
Итого					1 515,6	1 330,6

Мы получили, что прирост производительности труда в целом по предприятию составил 13,9 %.

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

$$I_w = \frac{\sum_{i=1}^{n} q_{i1} p_i}{\sum_{i=1}^{n} T_{i1}} : \frac{\sum_{i=1}^{n} q_{i0} p_i}{\sum_{i=1}^{n} T_{i0}}. \tag{7.16}$$

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

Пример 7.3. Требуется определить индекс производительности труда по данным (табл. 7.3) о производстве продукции и отпускных ценах предприятия.

Вычислим индекс производительности труда по формуле (7.16):

$$I_w = \frac{\sum_{i=1}^{n} q_{i1} p_i}{\sum_{i=1}^{n} T_{i1}} : \frac{\sum_{i=1}^{n} q_{i0} p_i}{\sum_{i=1}^{n} T_{i0}} = \frac{217\,350}{3\,302} : \frac{211\,700}{3\,289} = 65{,}8 : 64{,}4 = 1{,}022, \text{ или } 102{,}2\,\%.$$

Мы получили, что в текущем периоде за 1 чел.-ч вырабатывалось 65,8 тыс. руб. продукции, а в базисном — 64,4 тыс. руб. Поэтому прирост производительности труда составил 2,2 %.

Вид продукции (i)	Сентябрь		Октябрь		Отпускная цена, тыс. руб. (p_i)	Расчетные графы, тыс. руб.	
	Произведено, шт. (q_{i0})	Трудовые затраты, чел.-ч (T_{i0})	Произведено, шт. (q_{i1})	Трудовые затраты, чел.-ч (T_{i1})		$q_{i0}p_i$	$q_{i1}p_i$
Изделие 1	370	1 024	390	1 032	200	74 000	78 000
Изделие 2	210	965	205	960	210	44 100	43 050
Изделие 3	520	1 300	535	1 310	180	93 600	96 300
Итого	х	3 289	х	3 302	х	211 700	217 350

Умножение индекса производительности труда по выработке на индекс затрат рабочего времени дает ***индекс физического объема продукции***, ***взвешенный по цене***:

$$I_w I_T = I_q \text{ или } \left(\frac{\sum_{i=1}^{n} q_{i1}p_i}{\sum_{i=1}^{n} T_{i1}} : \frac{\sum_{i=1}^{n} q_{i0}p_i}{\sum_{i=1}^{n} T_{i0}} \right) \cdot \frac{\sum_{i=1}^{n} T_{i1}}{\sum_{i=1}^{n} T_{i0}} = \frac{\sum_{i=1}^{n} q_{i1}p_i}{\sum_{i=1}^{n} q_{i0}p_i}. \qquad (7.17)$$

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

Предположим, мы располагаем данными о стоимости проданной продукции в текущем периоде (p_1q_1) и индивидуальными индексами цен ($i_p = p_1/p_0$), полученными, скажем, в результате выборочного наблюдения. Тогда в знаменателе сводного индекса цен ($I_p = \frac{\sum_{i=1}^{n} p_{i1}q_{i1}}{\sum_{i=1}^{n} p_{i0}q_{i1}}$) можно использовать следующую замену:

$$p_{i0} = \frac{1}{i_{ip}} p_{i1}.$$

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

$$I_p = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} \frac{1}{i_{ip}} p_{i1} q_{i1}}. \qquad (7.18)$$

Пример 7.4. По данным табл. 7.4 требуется получить сводную оценку изменения цен.

Товар (i)	Реализация в текущем периоде, тыс. руб. ($p_{i1}q_{i1}$)	Изменение цен в текущем периоде по сравнению с базисным, % ($i_{ip} \cdot 100\,\% - 100\,\%$)	Расчетные графы	
			i_{ip}	$\frac{p_{i1}q_{i1}}{i_{ip}}$
Морковь (1)	23 000	+4,0	1,040	22 115
Свекла (2)	21 000	+2,3	1,023	20 528
Лук (3)	29 000	–0,8	0,992	29 234
Итого	73 000			71 877

Вычислим средний гармонический индекс (7.18):

$$I_p = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} \frac{1}{i_{ip}} p_{i1} q_{i1}} = \frac{73\,000}{71\,877} = 1{,}016, \text{ или } 101{,}6\,\%.$$

Цены по данной товарной группе в текущем периоде по сравнению с базисным в среднем возросли на 1,6 %.

При расчете ***сводного индекса физического объема товарооборота*** ($I_q = \frac{\sum_{i=1}^{n} q_{i1} p_{i0}}{\sum_{i=1}^{n} q_{i0} p_{i0}}$) можно использовать среднеарифметическую форму. При этом в числителе производится замена:

$$q_{i1} = i_{iq} q_{i0}.$$

Тогда индекс примет вид:

$$I_q = \frac{\sum_{i=1}^{n} i_{iq} q_{i0} p_{i0}}{\sum_{i=1}^{n} q_{i0} p_{i0}}. \qquad (7.19)$$

Пример 7.5. По данным табл. 7.5 о реализации трех товаров в натуральном и стоимостном выражении рассчитать индекс физического объема товарооборота.

Товар (i)	Реализация в базисном периоде, тыс. руб. ($q_{i0}p_{i0}$)	Изменение физического объема реализации в текущем периоде по сравнению с базисным, % ($i_{ip} \cdot 100\,\% - 100\,\%$)	Расчетные графы	
			i	$i_{iq}q_{i0}p_{i0}$
Мандарины (1)	46 000	–6,4	0,936	43 056
Грейпфруты (2)	27 000	–8,2	0,918	24 786
Апельсины (3)	51 000	+1,3	1,013	51 663
Итого	124 000			119 505

Рассчитаем средний арифметический индекс согласно (7.19):

$$I_q = \frac{\sum_{i=1}^{n} i_{iq} q_{i0} p_{i0}}{\sum_{i=1}^{n} q_{i0} p_{i0}} = \frac{119\,505}{124\,000} = 0{,}964, \text{ или } 96{,}4\,\%.$$

Физический объем реализации данных товаров в среднем снизился на 3,6 %.

В среднеарифметической форме также может рассчитываться и индекс производительности труда по трудоемкости, известный как ***индекс С. Г. Струмилина***:

$$I_w = \frac{\sum_{i=1}^{n} i_{it} T_{i1}}{\sum_{i=1}^{n} T_{i1}} = \frac{\sum_{i=1}^{n} \left(\frac{T_{i0}}{q_{i0}} : \frac{T_{i1}}{q_{i1}} \right) \cdot T_{i1}}{\sum_{i=1}^{n} T_{i1}}. \qquad (7.20)$$

Как правило, индексы используются для анализа динамики социально-экономических явлений за ряд последовательных периодов. Поэтому для достижения сопоставимости они должны рассчитываться по единой схеме. Такая схема расчета индексов за несколько временны́х периодов называется ***системой индексов***.

Входящие в систему индексы могут быть цепными или базисными. При построении цепных индексов цены каждого периода сравниваются с ценами предшествующего периода. В базисных индексах цены каждого периода сравниваются с ценами базисного (как правило, первого) периода.

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

А. Цепные индексы цен с переменными весами:

$$I_{p^{1}/_{0}} = \frac{\sum_{i=1}^{n} p_{i1}q_{i1}}{\sum_{i=1}^{n} p_{i0}q_{i1}}; \quad I_{p^{2}/_{1}} = \frac{\sum_{i=1}^{n} p_{i2}q_{i2}}{\sum_{i=1}^{n} p_{i1}q_{i2}}; \quad \ldots \quad I_{p^{m}/_{m-1}} = \frac{\sum_{i=1}^{n} p_{im}q_{im}}{\sum_{i=1}^{n} p_{i,m-1}q_{i,m}}.$$

Б. Цепные индексы цен с постоянными весами:

$$I_{p^{1}/_{0}} = \frac{\sum_{i=1}^{n} p_{i1}q_{i0}}{\sum_{i=1}^{n} p_{i0}q_{i0}}; \quad I_{p^{2}/_{1}} = \frac{\sum_{i=1}^{n} p_{i2}q_{i0}}{\sum_{i=1}^{n} p_{i1}q_{i0}}; \quad \ldots \quad I_{p^{m}/_{m-1}} = \frac{\sum_{i=1}^{n} p_{im}q_{i0}}{\sum_{i=1}^{n} p_{m-1}q_{i0}}.$$

В. Базисные индексы цен с переменными весами:

$$I_{p^{1}/_{0}} = \frac{\sum_{i=1}^{n} p_{i1}q_{i1}}{\sum_{i=1}^{n} p_{i0}q_{i1}}; \quad I_{p^{2}/_{0}} = \frac{\sum_{i=1}^{n} p_{i2}q_{i2}}{\sum_{i=1}^{n} p_{i0}q_{i2}}; \quad \ldots \quad I_{p^{m}/_{0}} = \frac{\sum_{i=1}^{n} p_{im}q_{im}}{\sum_{i=1}^{n} p_{i0}q_{im}}.$$

Г. Базисные индексы цен с постоянными весами:

$$I_{p\,1/0}=\frac{\sum_{i=1}^{n}p_{i1}q_{i0}}{\sum_{i=1}^{n}p_{i0}q_{i0}};\quad I_{p\,2/0}=\frac{\sum_{i=1}^{n}p_{i2}q_{i0}}{\sum_{i=1}^{n}p_{i0}q_{i0}};\quad \ldots\quad I_{p\,m/0}=\frac{\sum_{i=1}^{n}p_{im}q_{i0}}{\sum_{i=1}^{n}p_{i0}q_{i0}}.$$

При построении индексов с постоянными весами более удобен подход Ласпейерса, т. е. использование базисных весов начального периода. Если же данные индексы построены по принципу Пааше, то на практике они могут быть рассчитаны только по завершению всего рассматриваемого временнóго интервала (например, при расчете индексов по месяцам — всего года), когда уже известны веса последнего периода (декабря).

Отметим также, что из рассмотренных индексных систем определенные преимущества имеет вариант Б, так как составляющие эту систему индексы мультипликативны — их последовательное перемножение позволяет получить базисный индекс в целом за исследуемый временной интервал.

Все рассмотренные выше индексы рассчитывались по нескольким товарам, реализуемым в пределах одной территории, или видам продукции, производимым на одном предприятии. Рассмотрим теперь случай, когда один товар реализуется в нескольких местах или вид продукции производится на ряде предприятий.

Если реализуется только один вид продукции, вполне правомерно рассчитать его среднюю цену в каждом временнóм периоде. ***Индекс цен переменного состава*** представляет собой отношение полученных средних значений:

$$I_p^{\text{п.с}}=\frac{\sum_{i=1}^{n}p_{i1}q_{i1}}{\sum_{i=1}^{n}q_{i1}}:\frac{\sum_{i=1}^{n}p_{i0}q_{i0}}{\sum_{i=1}^{n}q_{i0}}. \tag{7.21}$$

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

$$I^{\text{стр}} = \frac{\sum_{i=1}^{n} p_{i0} q_{i1}}{\sum_{i=1}^{n} q_{i1}} : \frac{\sum_{i=1}^{n} p_{i0} q_{i0}}{\sum_{i=1}^{n} q_{i0}}. \tag{7.22}$$

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

$$I_p^{\text{ф.с}} = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}}. \tag{7.23}$$

Между данными индексами существует следующая взаимосвязь:

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

Пример 7.6. По данным табл. 7.6 провести анализ цен реализации товара в двух регионах.

Вычислим индекс цен переменного состава согласно (7.21):

$$I_p^{\text{п.с}} = \frac{405\,000}{27\,000} : \frac{460\,000}{30\,000} = 15{,}00 : 15{,}33 = 0{,}978, \text{ или } 97{,}8\,\%.$$

Из табл. 7.6 видно, что цена в каждом регионе в июле по сравнению с июнем возросла. В целом же средняя цена снизилась на 2,2 % (97,8 % — 100 %). Такое несоответствие объясняется влиянием изменения структуры реализации товаров по регионам: в июне по более высокой цене продавали товара вдвое больше, в июле ситуация принципиально изменилась (в дан-

Регион (i)	Июнь: цена, тыс. руб. (p_{i0})	Июнь: продано, шт. (q_{i0})	Июль: цена, тыс. руб. (p_{i1})	Июль: продано, шт. (q_{i1})	Расчетные графы, тыс. руб.: $p_{i0}q_{i0}$	$p_{i1}q_{i1}$	$p_{i0}q_{i1}$
1	12	10 000	13	8 000	120 000	234 000	216 000
2	17	20 000	19	9 000	340 000	171 000	153 000
Итого	—	30 000	—	27 000	460 000	405 000	369 000

* Цифры условные.

ном условном примере для наглядности числа подобраны таким образом, чтобы это различие в структуре продаж было очевидным). Рассчитаем индекс структурных сдвигов согласно (7.22):

$$I^{\text{стр}} = \frac{369\,000}{27\,000} : \frac{460\,000}{30\,000} = 0{,}891, \text{ или } 89{,}1\,\%.$$

Первая часть этого выражения позволяет ответить на вопрос, какой была бы средняя цена в июле, если бы цены в каждом регионе сохранились на прежнем июньском уровне. Вторая часть отражает фактическую среднюю цену июня. В целом по полученному значению индекса мы можем сделать вывод, что за счет структурных сдвигов цены снизились на 10,9 %.

Рассчитанный индекс цен фиксированного состава равен 1,098 или 109,8 %. Отсюда следует вывод: если бы структура реализации товара А по регионам не изменилась, средняя цена возросла бы на 9,8 %. Однако влияние на среднюю цену первого фактора оказалось сильнее, что отражается в следующей взаимосвязи:

$$1{,}098 \cdot 0{,}891 = 0{,}978.$$

Аналогично строятся индексы структурных сдвигов, переменного и фиксированного состава для анализа изменения себестоимости, урожайности и других показателей.

Территориальные индексы служат для сравнения показателей в пространстве, т. е. по предприятиям, округам, городам, районам и пр. Построение территориальных (пространственных) индексов требует решения ряда методологических вопросов, связанных с выбором базы сравнения и весов, или уровня, на котором фиксируются веса.

При двусторонних сравнениях каждая территория может быть и сравниваемой (числитель индекса), и базой сравнения (знаменатель). Веса как первой, так и второй территории, в принципе, также имеют равные основания использоваться при расчете индекса. Однако это может привести к различным или даже противоречивым результатам.

Избежать подобной неопределенности можно несколькими способами. Один из них заключается в том, что в качестве весов принимаются объемы проданных товаров i-го вида ($i = 1, 2, \ldots, n$) по двум регионам, вместе взятым:

$$Q_i = q_{ia} \quad q_{ib}+$$

Территориальный индекс цен в этом случае рассчитывается по следующей формуле:

$$I_{pb/a} = \frac{\sum_{i=1}^{n} p_{ib} Q_i}{\sum_{i=1}^{n} p_{ia} Q_i}. \tag{7.24}$$

Пример 7.7. По данным табл. 7.7 о цене и объеме реализации товаров по двум регионам рассчитать территориальный индекс цен.

Товар (i)	Регион А		Регион Б		Расчетные графы		
	Цена, тыс. руб. (p_{ia})	Реали-зация, ц (q_{ia})	Цена, тыс. руб. (p_{ib})	Реали-зация, ц (q_{ib})	Реализа-ция, ц ($Q_i = q_{ia} + q_{ib}$)	$p_{ia}Q_i$, руб.	$p_{ib}Q_i$, руб.
1	11,0	30	12,0	35	65	715,0	780,0
2	8,5	45	9,0	50	95	807,5	855,0
3	17,0	15	16,0	90	105	1 785,0	1 680,0
Итого						3 307,5	3 315,0

Рассчитаем территориальный индекс цен согласно (7.24):

$$I_{pb/a} = \frac{\sum_{i=1}^{n} p_{ib} Q_i}{\sum_{i=1}^{n} p_{ia} Q_i} \quad \frac{3\,315{,}0}{3\,307{,}5} = 1{,}002, \text{ или } 100{,}2\ \%.$$

Цены в регионе Б на 0,2 % превышают цены в регионе А. Этому выводу не противоречит и обратный индекс:

$$I_{pa/b} = \frac{\sum_{i=1}^{n} p_{ia} Q_i}{\sum_{i=1}^{n} p_{ib} Q_i} \quad \frac{3\,307{,}5}{3\,315{,}0} = 0{,}998, \text{ или } 99{,}8\ \%.$$

В формуле данного территориального индекса вместо суммарных иногда используются стандартизованные веса (стандартизованная структура). В качестве таких весов может выступать структура продажи данных видов продукции по более крупному территориальному образованию, например республике. В этом случае индекс имеет вид:

$$I_p = \frac{\sum_{i=1}^{n} p_{ia} Q_{i\,\text{респ}}}{\sum_{i=1}^{n} p_{ib} Q_{i\,\text{респ}}}, \tag{7.25}$$

где $Q_{i\,\text{респ}}$ — объем продаж i-го товара.

Второй способ расчета территориальных индексов учитывает соотношение весов на каждой из сравниваемых территорий. При этом способе первый шаг заключается в расчете средней цены каждого товара по двум территориям, вместе взятым:

$$\bar{p}_i = \frac{p_{ia} q_{ia} + p_{ib} q_{ib}}{q_{ia} + q_{ib}}. \tag{7.26}$$

После этого непосредственно рассчитывается территориальный индекс:

$$I_{pb/a} = \frac{\sum_{i=1}^{n} \boldsymbol{p_{ib}} q_{ib}}{\sum_{i=1}^{n} \bar{p}_i q_{ib}} : \frac{\sum_{i=1}^{n} \boldsymbol{p_{ia}} q_{ia}}{\sum_{i=1}^{n} \bar{p}_i q_{ia}}. \tag{7.27}$$

Пример 7.8. По данным примера 7.7 определить территориальный индекс вторым способом. По формуле (7.26) получим:

$$\bar{p}_1 = \frac{11{,}0 \cdot 30 + 12{,}0 \cdot 35}{65} = 11{,}54; \quad \bar{p}_2 = \frac{8{,}5 \cdot 45 + 9{,}0 \cdot 50}{95} = 8{,}76;$$

$$\bar{p}_3 = \frac{17{,}0 \cdot 15 + 16{,}0 \cdot 90}{150} = 16{,}14.$$

С учетом рассчитанных средних цен вычислим индекс согласно (7.27):

$$I_{pb/a} = \frac{12{,}0 \cdot 35 + 9{,}0 \cdot 50 + 16{,}0 \cdot 90}{11{,}54 \cdot 35 + 8{,}76 \cdot 50 + 16{,}14 \cdot 90} : \frac{11{,}0 \cdot 30 + 8{,}5 \cdot 45 + 17{,}0 \cdot 15}{11{,}54 \cdot 30 + 8{,}76 \cdot 45 + 16{,}14 \cdot 15} =$$
$$= 1{,}022, \text{ или } 102{,}2\,\%.$$

Данный подход к расчету территориального индекса обеспечивает известную взаимосвязь:

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

Индекс физического объема реализации при этом строится следующим образом:

$$I_{qb/a} = \frac{\sum_{i=1}^{n} q_{ib}\overline{p}_i}{\sum_{i=1}^{n} q_{ia}\overline{p}_i}.$$

Аналогично строятся индексы для сравнения цен территории А с ценами территории Б.

1. Индексы позволяют соизмерить социально-экономические явления:
   а) в пространстве;
   б) во времени;
   в) в пространстве и во времени.
2. Можно ли утверждать, что индивидуальные индексы по методологии исчисления адекватны темпам роста:
   а) можно;
   б) нельзя?
3. Сводные индексы позволяют получить обобщающую оценку изменения:
   а) по товарной группе;
   б) одного товара за несколько периодов.
4. Является ли средний арифметический индекс разновидностью агрегатной формы индексов:
   а) является;
   б) не является?
5. Может ли в отдельных случаях средний гармонический индекс рассчитываться по средней гармонической невзвешенной:
   а) может;
   б) не может?
6. Может ли средний гармонический индекс быть меньше минимального из осредняемых индивидуальных индексов?
   а) да;
   б) нет?
7. Свойством мультипликативности обладают индексы:
   а) цепные с переменными весами;
   б) цепные с постоянными весами;
   в) базисные с переменными весами.
8. Являются ли цепные индексы с переменными весами индексами Пааше:
   а) являются;
   б) не являются?

9. Индексы переменного состава рассчитываются:
   а) по товарной группе;
   б) по одному товару.

10. Может ли индекс переменного состава превышать индекс фиксированного состава:
   а) может;
   б) не может?

11. По имеющимся в таблице данным о цене на товар определите недостающие значения показателей:

Месяц	Цена, руб.	Индивидуальные индексы цен	
		цепные	базисные
Январь	?	?	100,0
Февраль	250	102,0	?
Март	?	?	104,5

12. Имеются следующие данные о реализации мясных продуктов на городском рынке:

Продукт	Сентябрь		Октябрь	
	Цена за 1 кг, руб.	Продано, т	Цена за 1 кг, руб.	Продано, т
А	70	6,3	73	4,1
Б	62	1,8	64	1,2
В	85	4,5	86	3,3

Рассчитайте сводные индексы цен, физического объема реализации и товарооборота.

13. Определите, как изменился физический объем реализации потребительских товаров предприятиями розничной торговли города в текущем периоде по сравнению с предшествующим, если товарооборот возрос на 12,3 %, а цены повысились на 3,7 %.

14. Имеются следующие данные о реализации молочных продуктов предприятиями розничной торговли:

Продукт	Товарооборот, млн руб.		Изменение цены в декабре по сравнению с ноябрем, %
	Ноябрь	Декабрь	
Молоко	9,7	6,3	+2,1
Сметана	4,5	4,0	+3,5
Творог	12,9	11,5	+4,2

Рассчитайте сводные индексы цен, товарооборота и физического объема реализации.

15. Имеются следующие данные о реализации картофеля на рынках города:

Рынок	Июль		Август	
	Цена за кг, руб.	Продано, ц	Цена за кг, руб.	Продано, ц
1	13	24,5	9	21,9
2	16	18,7	12	37,8
3	10	32,0	10	33,4

Рассчитайте: а) индекс цен переменного состава; б) индекс цен фиксированного состава; в) индекс структурных сдвигов.

Современные методы статистического анализа, применяемые в самых разнообразных социально-экономических исследованиях, а также для принятия управленческих решений, опираются на теорию вероятностей.

Одним из основных понятий теории вероятностей является случайное событие. ***Случайным событием*** называется событие, которое должно либо произойти, либо не произойти при выполнении некоторого комплекса условий. В дальнейшем вместо «выполнение некоторого комплекса условий» и «случайное событие» будем употреблять выражения «произведено испытание», «событие» и «результат испытания».

Укажем некоторые соотношения, которые могут существовать между событиями.

1. ***Произведением событий*** A и B называется такое событие AB, которое заключается в совместном наступлении этих событий.

2. ***Суммой событий*** A и B называется такое событие $A + B$, которое заключается в наступлении, по крайней мере, одного из этих событий.

3. Событие U называется ***достоверным***, если оно обязательно должно произойти при каждом испытании. Событие V называется ***невозможным***, если оно не может произойти ни при каком испытании.

4. Событие $\overline{A}$ называется ***противоположным*** событию A (и наоборот), если для них одновременно выполняются равенства $\overline{A} + A = U$; $A\overline{A} = V$.

5. События A и B называются ***несовместимыми***, если их совместное наступление неосуществимо, т. е. если $AB = V$.

6. События A_1, A_2, ... A_n образуют полную группу попарно несовместимых событий, если события A_i и A_j при $i \neq j$ несовместимы

и хотя бы одно из событий $A_1, A_2, \ldots A_n$ непременно должно произойти. Иными словами, полная группа попарно несовместимых событий $A_1, A_2, \ldots A_n$ удовлетворяет двум условиям:

$$A_1 + A_2 + \ldots + A_n = U \text{ (полная группа);}$$

$$A_i A_j = V,\ i \neq j \text{ (попарная несовместимость).}$$

Степень возможности появления события определяет его вероятность. В зависимости от метода вычисления известны классическое и статистическое определение вероятности, которые отражают два разных подхода к этому понятию. Основное различие между ними связано с моментом вычисления вероятности. ***Классический подход*** предполагает предварительное вычисление вероятности, т. е. до наступления каких-либо событий. При ***статистическом подходе*** вероятности вычисляют после наступления исходов множества испытаний.

Согласно классическому определению, под вероятностью события A понимают отношения числа благоприятных исходов M к общему числу N равновозможных элементарных исходов, т. е. $P(A) = \dfrac{M}{N}$.

Например, вероятность появления цифры 5 при подбрасывании игральной кости равна 1/6. Поскольку игральная кость это куб, имеющий шесть оцифрованных граней 1, 2, 3, 4, 5, 6, то $N = 6$, причем только одна грань содержит цифру 5, поэтому $M = 1$.

Для невозможного события A число благоприятных исходов равно нулю ($M = 0$), а отсюда равна нулю и вероятность $P(A) = 0$. Например, к числу невозможных относится и событие, заключающееся в появлении цифры семь при подбрасывании игральной кости ($P(7) = 0$).

Для достоверного события A число благоприятных исходов равно общему числу возможных элементарных исходов ($M = N$) и вероятность равна единице $P(A) = 1$.

Например, достоверным является событие, заключающееся в том, что число очков при подбрасывании игральной кости будет менее 7. Очевидно, что для этого события $M = 6$, а вероятность $P(< 7) = 1$.

Таким образом, вероятность события A меняется в пределах от 0 до 1, т. е. $0 \leq P(A) \leq 1$, причем для случайного события A вероятность находится в интервале $0 < P(A) < 1$.

Классическое определение вероятности имеет тот недостаток, что в практических задачах далеко не всегда удается выделить равновозможность элементарных событий. Например, при бросании игральной кости классическое определение вероятности применимо, если только принять, что вероятности появления одной из цифр 1, 2, ..., 6 равны. При этом предполагается, что кость имеет форму куба и сделана из однородного материала. В этом случае центр тяжести игральной кости совпадает с ее центром. Если центр тяжести смещен относительно центра куба, то вероятность отдельных исходов (цифр 1, 2, 3, 4, 5 или 6) можно определить, используя только статистическое определение вероятности.

Статистически вероятность определяется как отношение $\frac{m}{n}$, полученное при неограниченном увеличении числа наблюдений n. Величина m называется частотой появления события A.

Чтобы показать, насколько согласуются вероятности событий, полученные на основании классического и статистического определения вероятности, рассмотрим следующий пример.

Пусть испытание состоит в подбрасывании монеты, а событием является появление герба.

В табл. 8.1 приведены результаты трех серий испытаний, проведенных известными статистиками Ж.-Л. Л. Бюффоном и К. Пирсоном.

Как видно из табл. 8.1, относительные частоты $\frac{m}{n}$ незначительно отличаются от вероятности $P(A) = 0{,}5$, вычисленной на основании классического определения вероятности. При этом чем больше число наблюдений, тем меньше расхождение.

В практической деятельности нам часто приходится иметь дело не только со случайными событиями, но и со случайными величинами. ***Случайной*** называют величину X, которая в результате испытания принимает одно из возможных значений, заранее неизвестное и зависящее от многих причин, учесть которые мы не в состоянии.

В зависимости от характера области возможных значений выделяют два вида случайных величин: дискретные и непрерывные.

Например, число очков, выпавших при бросании игральной кости, есть дискретная случайная величина X с шестью возможны-

Число наблюдений (n)	Частота (m)	Относительная частота $\left(\frac{m}{n}\right)$
4 040	2 048	0,5080
12 000	6 019	0,5016
21 000	12 012	0,5005

x_i	x_1	x_2	………	x_n
p_i	p_1	p_2	………	p_n

ми значениями: 1, 2, 3, 4, 5, 6. Причем одно возможное отличается от другого на величину, не меньшую единицы.

Интервал времени между двумя последовательными появлениями автобуса на данной остановке есть пример непрерывной случайной величины X, так как один интервал времени может отличаться от другого на любую малую величину.

Случайная величина X задается ***законом распределения вероятностей***, который определяет соотношение между возможными значениями случайной величины и соответствующими вероятностями.

Если обозначить возможные значения дискретной случайной величины X через $x_1, x_2, \ldots, x_n$, а через $p_i = P\,(X = x_i)$ вероятность появления значения x_i, то случайная величина полностью определяется законом распределения, заданным табл. 8.2.

Поскольку в табл. 8.2 рассмотрены все возможные значения случайной величины X, то вероятности p_i обладают следующим свойством:

$$p_1 + p_2 + \ldots + p_n = \sum_{i=1}^{n} p_i = 1. \tag{8.1}$$

Например, если X — число очков, появляющихся при бросании игральной кости, то закон распределения вероятностей этой случайной величины можно представить в виде табл. 8.3.

Из табл. 8.3 следует, что

$$\sum_{i=1}^{6} p_i = 6\ \frac{1}{6} = 1.$$

Закон распределения вероятностей случайной величины X может быть задан функцией распределения $F(x)$ и плотностью распределения $f(x)$ — для непрерывных случайных величин. ***Функцией***

x_i	1	2	3	4	5	6
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

распределения $F(x)$ случайной величины X называется вероятность того, что величина X примет значение, меньшее чем x:

$$F(x) = P(X < x). \tag{8.2}$$

Для дискретной случайной величины $F(x)$ вычисляется по формуле

$$F(x) = \sum_{x_i < x} p_i, \tag{8.3}$$

где суммирование ведется по всем значениям i, для которых $x_i < x$.

Например, из табл. 8.3 следует, что вероятность того, что при бросании игральной кости выпадает число очков, меньшее $x = 4$, равна $\frac{3}{6}$, т. е. $F(4) = p_1 + p_2 + p_3 = 3/6$.

График функции распределения для непрерывной случайной величины X представлен на рис. 8.1.

Из рис. 8.1 видно, что $F(x)$ неубывающая функция, меняющаяся в пределах от 0 до 1. По мере возрастания X функция $F(x)$ стремится к единице (при $x \to \infty$ $F(x) \to 1$), а по мере уменьшения X функция $F(x)$ приближается к 0 (при $x \to -\infty$ $F(x) \to 0$). При этом вероятность попадания случайной величины в интервал от x_1 до x_2 определяется формулой

$$P(x_1 < X < x_2) = F(x_2) - F(x_1). \tag{8.4}$$

В справедливости данной формулы можно легко убедиться. По определению функции распределения $F(x) = P(X < x)$, а из (8.4) следует, что $P(x_1 < X < x_2) = P(X < x_2) - P(X < x_1)$, что согласуется с рис. 8.1.

Закон распределения непрерывной случайной величины X может быть задан ***функцией плотности*** $f(x)$, которая определяется как первая производная функции распределения $F(x)$:

$$f(x) = F'(x) = \lim_{\Delta_x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x}. \tag{8.5}$$

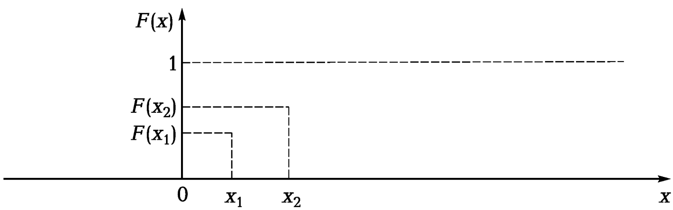


Рис. 8.1. График функции распределения непрерывной величины X

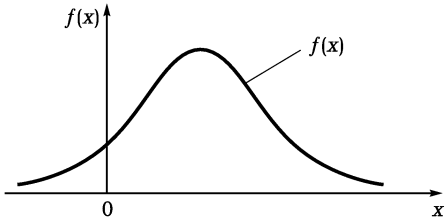


Рис. 8.2. График функции плотности распределения

График функции плотности распределения представлен на рис. 8.2.

Из приведенного определения вытекают следующие свойства функции плотности распределения. Во-первых, плотность есть неотрицательная функция, т. е. $f(x) \geq 0$. Во-вторых, вероятность попадания случайной величины X в интервал от x_1 до x_2 равна $P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$, т. е. площади, ограниченной сверху графиком плотности $f(x)$ на интервале от x_1 до x_2 (рис. 8.3).

Подынтегральное выражение $f(x)dx$ называется элементом вероятности. Оно определяет вероятность попадания случайной величины в интервал от x до $x + dx$, где dx — бесконечно малая величина. Площадь, ограниченная графиком плотности на всем интервале возможных значений X, равна единице (см. рис. 8.2), т. е. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Числовые характеристики случайной величины. Для практического применения не всегда необходимо иметь полное представление о случайной величине, заложенное в законе распределения вероятностей, а достаточно знать некоторые ее числовые характеристики, дающие представление об отдельных свойствах случайной величины. К таким характеристикам прежде всего относится

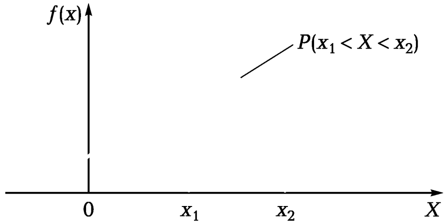


Рис. 8.3. Вероятность $P(x_1 < X < x_2)$ попадания X в интервал от x_1 до x_2

математическое ожидание MX и дисперсия DX случайной величины X.

Математическое ожидание MX характеризует среднее значение, центр, вокруг которого группируются возможные значения величины X.

Для дискретной случайной величины X при n возможных значениях x_i и соответствующих вероятностях p_i математическое ожидание определяется по формуле

$$MX = \sum_{i=1}^{n} x_i p_i. \tag{8.6}$$

Чтобы убедиться в справедливости определения математического ожидания как среднего значения случайной величины, рассмотрим дискретную величину X — число очков, появляющуюся при подбрасывании игральной кости, закон распределения которой представлен в табл. 8.3. «Здравый смысл» подсказывает, что среднее должно находиться между цифрами 3 и 4. С учетом данных табл. 8.3 будем иметь:

$$MX = \sum_{i=1}^{n} x_i p_i \quad 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6} = \frac{1}{6}\cdot 21 = 3{,}5.$$

Как видим, полученное значение математического ожидания $MX = 3{,}5$ согласуется с нашим представлением о среднем.

Рассмотрим свойства математического ожидания, которые нам будут полезны для дальнейших статистических исследований.

1. Математическое ожидание постоянной величины $M(c)$ равно этой постоянной. Так, пусть $X = c$, где c — постоянная, тогда $M(c) = c$. Здесь c можно рассматривать как дискретную случайную величину, принимающую единственное значение с вероятностью единица. Поэтому $M(c) = c \cdot 1 = c$.

2. Если ко всем возможным значениям величины X прибавить постоянную c, то математическое ожидание изменится на эту постоянную:

$$M(X + c) = MX + c. \tag{8.7}$$

3. Постоянный множитель c можно вынести за закон математического ожидания:

$$M(c \cdot X) = c \cdot MX. \tag{8.8}$$

4. Математическое ожидание суммы случайных величин X_1, X_2, ..., X_n равно сумме математических ожиданий этих величин. $M(X_1 + X_2 + ... + X_n) = MX_1 + MX_2 + ... + MX_n$, т. е.

$$M\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} MX_i. \tag{8.9}$$

Дисперсией DX называется математическое ожидание квадрата отклонения случайной величины X от ее математического ожидания MX, т. е.

$$DX = M(X - MX)^2. \tag{8.10}$$

Дисперсия DX характеризует средний разброс, рассеяние значений случайной величины X около математического ожидания. Положительный корень из дисперсии $\sqrt{DX}$ называют ***средним квадратичным отклонением***.

Для дискретной случайной величины X с n возможными значениями x_i и соответствующими вероятностями p_i, где i = 1, 2, ..., n, дисперсия определяется по формуле

$$DX = \sum_{i=1}^{n}(x_i \quad MX)^2 p_i. \tag{8.11}$$

Рассмотрим основные свойства дисперсии, которые достаточно просто выводятся из (8.10) с учетом свойств математического ожидания.

1. Дисперсия постоянной величины равна нулю. В самом деле, если $X = C$, то согласно свойству 1 $MC = C$. Тогда

$$DC = M(C - C)^2 = 0. \tag{8.12}$$

2. Дисперсия не изменится, если все значения случайной величины X изменить на постоянную C. Так,

$$D(X + C) = DX. \tag{8.13}$$

3. Постоянную величину C можно вынести за знак дисперсии, предварительно возведя ее в квадрат. Тогда

$$D(C \cdot X) = C^2 \cdot DX. \tag{8.14}$$

4. Дисперсия суммы независимых случайных величин $X_1, X_2, ..., X_n$ равна сумме дисперсий этих величин:

$$D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} DX_i. \tag{8.15}$$

Моменты случайной величины. Для характеристики случайной величины кроме математического ожидания и дисперсии применяются и моменты. ***Моментом*** k-порядка называется математическое ожидание k-й степени отклонения случайной величины X от некоторой постоянной c.

Если в качестве c берется нуль, моменты называют начальными. Тогда

$$v_k = M(X)^k. \tag{8.16}$$

Если $c = M(X)$, то моменты называются центральными.

Иными словами

$$\mu_k = [X - M(X)]^k. \tag{8.17}$$

В формулах, определяющих начальные и центральные моменты, нижние индексы указывают порядок момента. С помощью свойств математического ожидания легко показать, что

$$v_0 = \mu_0 = 1; \quad \mu_1 = 0; \quad v_1 = MX; \quad \mu_2 = DX. \tag{8.17'}$$

Центральные моменты связаны с начальными следующими соотношениями:

$$\begin{gathered} \mu_0 = 1; \quad \mu_1 = 0; \quad \mu_2 = v_2 \quad v_1^2; \quad \mu_3 = v_3 \quad 3v_1 v_2 + 2v_1^3; \\ \mu_4 = v_4 \quad 4v_1 v_3 + 6v_1^2 v_2 - 3v_1^4. \end{gathered} \tag{8.18}$$

Первые моменты играют важную роль в статистике при нахождении параметров функции распределения. Формула $D(X) = = \mu_2 \quad \nu_2 - \nu_2^1$ употребляется для вычисления дисперсии. Мерой скошенности графика плотности относительно математического ожидания является коэффициент асимметрии

$$A_c = \frac{\mu_3}{\sqrt{\mu_2^3}}, \tag{8.19}$$

который для симметричных законов распределения равен нулю.

Пример 8.1. Вычислить начальный и центральный моменты третьего порядка для случайной величины X — число появлений герба при трех подбрасываниях монеты.

Решение. Для вычисления моментов удобно воспользоваться вспомогательной таблицей (табл. 8.4).

Из табл. 8.4 следует:

$$v_1 = MX \quad \sum_{i=0}^{3} x_i p_i = 1{,}5; \quad v_2 = \sum_{i=0}^{3} x_i^2 p_i \quad 3{,}0; \quad v_3 = \sum_{i=0}^{3} x_i^3 p_i \quad 6{,}750. =$$

Теперь воспользуемся формулой

$$\mu_3 = v_3 \quad 2v_1 v_2 + 2v_1^3$$

и получим $\mu_3 = 6{,}75 \quad 3 \cdot 1{,}5 \cdot 3 + 2 \cdot (1{,}5)^3 = 0$.

x_i	p_i	$x_i p_i$	$x_i^2 p_i$	$x_i^3 p_i$
0	0,125	0	0	0
11	0,375	0,375	0,375	0,375
2	0,375	0,750	1,500	3,000
3	0,125	0,375	1,125	0,375
Итого	1	1,500	3,000	6,750

Рассмотрим модели законов распределения вероятностей, наиболее распространенные в статистической практике.

Нормальный закон распределения. Нормальное распределение — наиболее часто встречающийся вид распределения. С ним приходится сталкиваться при анализе погрешностей измерений, контроле технологических процессов и режимов, а также при анализе и прогнозировании различных явлений в экономике, социологии, демографии и других областях знаний. Наиболее важным условием возникновения нормального закона распределения является формирование признака как суммы большого числа взаимно независимых слагаемых, ни одно из которых не характеризуется исключительно большой по сравнению с другими дисперсией. В экономической и производственной практике такие предпосылки нередко соблюдаются.

Главная особенность нормального распределения состоит в том, что оно является предельным, к которому приближаются другие распределения. Непрерывная случайная величина X подчиняется нормальному закону распределения, если ее функция плотности вероятности имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{8.20}$$

Нормальный закон распределения зависит от двух параметров. Во-первых, от $\mu = MX$ — математического ожидания величины X. Во-вторых, от $\sigma^2 = DX$ — дисперсии величины X. Отсюда следует, что σ — среднее квадратичное отклонение величины X, которое, как и дисперсия, характеризует рассеяние значений X около математического ожидания μ. В этом случае говорят, что X подчиняется нормальному закону с параметрами μ, σ, т. е. $X \in N(\mu, \sigma)$.

Из формулы (8.20) вытекают следующие свойства функции плотности нормального распределения. Функция плотности $f(x)$ симме-

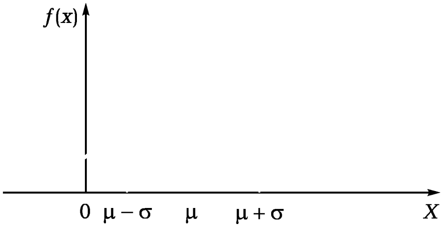


Рис. 8.4. График плотности нормального распределения $X \in N$ (μ; σ)

трична относительно математического ожидания μ. Она достигает своего максимума при $x = \mu$, который равен $\frac{1}{\sigma\sqrt{2\pi}}$. Функция плотности стремится к нулю по мере отдаления x от μ, при ($x \to -\infty$ и $x \to +\infty$). Она всегда положительна, т. е. $f(x) > 0$. Добавим к этому, что площадь, ограниченная сверху графиком плотности $f(x)$, равна 1.

Из указанных свойств следует, что график функции нормального распределения $f(x)$ представляет собой колоколообразную кривую, которая представлена на рис. 8.4.

Для построения таблиц нормального распределения переходят от случайной величины X, имеющей нормальное распределение с параметрами μ и σ, т. е. $X \in N(\mu, \sigma)$, к нормальной (стандартизированной) случайной величине

$$T = \frac{X - \mu}{\sigma}, \tag{8.21}$$

которая имеет нормированное нормальное распределение с математическим ожиданием, равным нулю $MT = 0$, и дисперсией, равной единице $DT = 1$, т. е. $T \in N(0; 1)$.

Воспользовавшись свойствами математического ожидания, можно показать, что

$$MT = M\left(\frac{X - M}{\sigma}\right) = \frac{1}{\sigma}[M(X \quad \mu)] = \frac{1}{\sigma}[MX \quad \mu] = 0, -$$

а на основании свойств дисперсии, что

$$DT = D\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2}D(X \quad \mu) = \frac{1}{\sigma^2}DX \quad \frac{1}{\sigma^2}\sigma^2 = 1.$$

Плотность нормированного нормального распределения описывается функцией

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}. \tag{8.22}$$

Эта функция, как следует из (8.22), симметрична относительно начала координат. Имеются таблицы плотности нормированного нормального распределения $f(t)$, табл. П.7, и интегральной функции Лапласа $\Phi(t)$, табл. П.1. Интегральная функция Лапласа определяет вероятность местонахождения случайной величины T в интервале от $-t$ до t:

$$\Phi(t) = P(\ \ t < T < t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-\frac{t^2}{2}} dt. \qquad (8.23)$$

Графическое представление интегральной функции Лапласа дано на рис. 8.5.

Из (8.22) следует, что плотность нормированного нормального распределения является четной функцией, т. е. $f(-t) = f(t)$.

Интегральная функция Лапласа является нечетной функцией, т. е. $\Phi(-t) = \Phi(t)$.

Таблица интегральной функции Лапласа (табл. П.1) позволяет достаточно просто решать задачи определения вероятности попадания нормальной случайной величины в заданный интервал.

В частности, из табл. П.1 следует, что:

$$\begin{aligned} &\text{при } t = 1 \ \Phi(t = 1) = 0{,}6827; \\ &\text{при } t = 2 \ \Phi(t = 2) = 0{,}9545; \qquad (8.24) \\ &\text{при } t = 3 \ \Phi(t = 3) = 0{,}9973. \end{aligned}$$

Таким образом, вероятность того, что случайная величина t, имеющая нормальное распределение, отклонится от своего математического ожидания ($Mt = 0$) на величину, не превышающую одно ($t = 1$) среднее квадратическое отклонение $\sigma_t = 1$ (см. рис. 8.5), равна 0,6827; два ($t = 2$) среднего квадратического отклонения равна 0,9545 и три ($t = 3$) равна 0,9973.

Отсюда следует правило трех сигм (3σ), согласно которому практически равна нулю вероятность (0,0027) того, что нормальная слу-

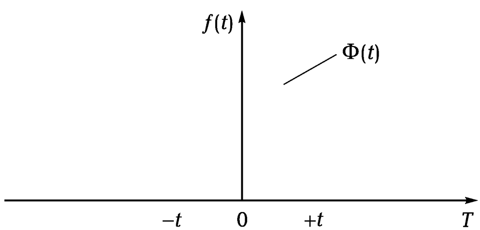


Рис. 8.5. Графическое представление функции $\Phi(t)$

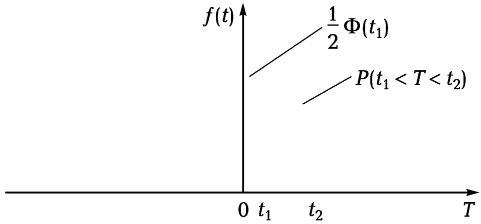


Рис. 8.6. Определение вероятности $P(t_1 < T < t_2)$

чайная величина отклонится от своего математического ожидания на величину, превышающую 3σ.

Вероятность попадания нормированной нормальной величины T в интервал от t_1 до t_2 определяется по формуле

$$P(t_1 < T < t_2) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)]. \tag{8.25}$$

В справедливости этой формулы достаточно просто убедиться, обратившись к рис. 8.6.

Поскольку $f(t)$ симметрична относительно начала координат, то $P(0 < T < t_1) = \frac{1}{2}\Phi(t_1)$. Тогда согласно рис. 8.6 получаем выражение

$$P(t_1 < T < t_2) = \frac{1}{2}\Phi(t_2) - \frac{1}{2}\Phi(t_1) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)]. \tag{8.26}$$

Вероятность попадания случайной величины X, имеющей нормальное распределение с параметрами μ, σ, т. е. $X \in N(\mu; \sigma)$, в интервал от x_1 до x_2 определяется по формуле

$$P(x_1 < X < x_2) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)], \tag{8.27}$$

где $t_1 = \dfrac{x_1 - \mu}{\sigma}$ и $t_2 = \dfrac{x_2 - \mu}{\sigma}$.

Для симметричного относительно μ интервала $(\mu - \delta; \mu + \delta)$ вероятность попадания в него случайной величины X определяется формулой

$$P\left(|X - \mu| < \delta\right) = \Phi\left(\frac{\delta}{\sigma}\right) = \Phi(t),$$

где $\delta = t\sigma$.

Пример 8.2. Известно, что средняя стоимость разовой покупки одного покупателя продуктового магазина есть случайная величина, имеющая нормальное распределение с математическим ожиданием $\mu = 876$ руб. и средним квадратическим отклонением $\sigma = 150$ руб. Определим, какая часть покупателей приобретает продуктов на сумму в пределах от 580 руб. до 1 200 руб., а какая — ниже 580 руб. Определим также, доход какой части жителей отклоняется от среднего μ на величину, не превышающую 150 руб.

Таким образом, требуется установить вероятность того, что случайная величина X, имеющая нормальное распределение, попадет в интервал от $x_1 = 580$ руб. до $x_2 = 1\,200$ руб., будет меньше 580 руб., т. е. попадает в интервал от 0 до 580 руб., и попадет в интервал от $(\mu - 150)$ до $(\mu + 150)$ руб.

Решение. Определим вероятность $P(580 < X < 1\,200)$.

Согласно (8.26) имеем

$$P(580 < X < 1\,200) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)],$$

где $t_1 = \dfrac{x_1 - \mu}{\sigma} = \dfrac{580 - 876}{150} = -1{,}97$ и $t_2 = \dfrac{x_2 - \mu}{\sigma} = \dfrac{1\,200 - 876}{150} = 2{,}16.$

По табл. П.1 для $t_1 = -1{,}973$ найдем $\Phi(t_1) = -\Phi(1{,}973) = -0{,}9512$, а для $t_2 = 2{,}16$ найдем $\Phi(t_2) = 0{,}9692$. Тогда искомая вероятность равна

$$P(580 < X < 1\,200) = \frac{1}{2}[0{,}9692 - (-0{,}9512)] = 0{,}9602.$$

Таким образом, 96,02 % покупателей продуктового магазина совершают разовые покупки на сумму от 580 руб. до 1 200 руб.

Далее определим долю долю покупателей, разовые покупки которых стоят менее 580 руб., т. е. вероятность того, что

$$P(0 < X < 580) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)],$$

где $t_1 = \dfrac{0 - 876}{150} = -5{,}84$ и $t_2 = \dfrac{580 - 876}{150} = -1{,}97.$

По табл. П.1 для $t_1 = -5{,}84$ найдем $\Phi(5{,}84) = -1$, а для $t_2 = -1{,}973$ найдем $\Phi(-1{,}973) = -0{,}9512$. Тогда искомая вероятность равна

$$P(X < 580) = \frac{1}{2}[-0{,}9512 - (-1)] = 0{,}0244.$$

Таким образом, 2,44 % покупателей совершают разовые покупки на сумму менее 580 руб.

Результаты решения примера 8.2 представлены на рис. 8.7.

Теперь определим вероятность попадания X в симметричный относительно μ интервал от $(\mu - 150)$ до $(\mu + 150)$ руб., т. е. от 726 руб. до 1 026 руб.

Согласно (8.27) имеем, учитывая, что $\sigma = 150$ руб.,

$$P(|X - \mu| < 150) = \Phi\left(\frac{150}{\sigma}\right) = \Phi(1).$$

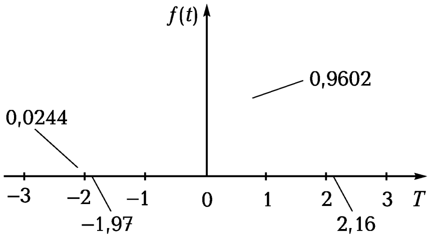


Рис. 8.7. Определение вероятностей

По табл. П.1 для $t = 1$ найдем $Ф(1) = 0{,}6827$. Таким образом, 68,27 % покупателей совершают разовые покупки на сумму от 726 руб. до 1 026 руб.

Биноминальный закон распределения. Биноминальный закон определяет вероятную числовую возможность (m) появления события A при n независимых испытаний, в каждом из которых событие A может осуществиться с одной и той же вероятностью $P(A) = p = \text{const}$. Кроме события A может произойти также противоположное событие $\overline{A}$, вероятность которого $P(\overline{A}) = 1 - p = q$.

Возможность того, что событие A возникает m раз при n независимых испытаниях с одинаковой вероятностью p, можно рассчитать по формуле Бернулли, которая и определяет биномиальный закон распределения:

$$P_n(m) = C_n^m p^m (1 \quad p)^{n-m}, \ - \tag{8.28}$$

где $C_n^m = \dfrac{n!}{m!(n-m)!}$ — число сочетаний из n элементов по m элементов, а факториал $n! = n(n-1) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$.

Из (8.28) следует, что биномиальный закон распределения зависит от двух параметров: n и p. Математическое ожидание частоты m появления события A при n независимых испытаний равно:

$$M(m) = np. \tag{8.29}$$

Дисперсия частоты m появления события A равна $D(m) = np(1-p)$, а среднее квадратическое отклонение равно:

$$\sigma(m) = \sqrt{np(1 \quad p)}. \quad - \tag{8.30}$$

При достаточно большом числе наблюдений ($n > 50$) биномиальное распределение сводится к нормальному распределению с математическим ожиданием $M(m) = np$ и средним квадратическим отклонением $\sigma(m) = \sqrt{npq}$, где $q = 1 - p$, т. е. $m \sim N(np; \sqrt{npq})$.

В этом случае вероятность того, что в n испытаниях событие A наступит m раз, определяется с помощью локальной функции Лапласа:

$$P_n(m) = \frac{1}{\sqrt{npq}} f(t) \quad \text{и} \quad t = \frac{m - np}{\sqrt{npq}}, \tag{8.31}$$

где $f(t)$ — плотность нормированной нормальной случайной величины, определяется по табл. П.7.

При $n > 50$ вероятность того, что в n независимых испытаниях Бернулли событие A наступит от m_1 до m_2 раз, определяется с помощью интегральной функции Лапласа:

$$P(m_1 < m < m_2) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)], \tag{8.32}$$

где $t_1 = \dfrac{m_1 - np}{\sqrt{npq}}$ и $t_2 = \dfrac{m_2 - np}{\sqrt{npq}}$.

Значение $\Phi(t)$ см. в табл. П.1.

С учетом свойств математического ожидания и дисперсии из (8.29) и (8.30) следует, что

$$M\left(\frac{m}{n}\right) = p \quad \text{и} \quad D\left(\frac{m}{n}\right) = \frac{p(1-p)}{n}. \tag{8.33}$$

При достаточно большом числе испытаний n ($n > 100$) относительная частота $\omega = \dfrac{m}{n}$ имеет нормальное распределение. Поэтому вероятность того, что относительная частота ω попадает в интервал от ω_1 до ω_2, определяется по формуле

$$P(\omega_1 < \omega < \omega_2) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)], \tag{8.34}$$

где $t_1 = \dfrac{\omega_1 - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ и $t_2 = \dfrac{\omega_2 - p}{\sqrt{\dfrac{p(1-p)}{n}}}$.

Значение $\Phi(t)$ см. в табл. П.1.

Пример 8.3. Известно, что 10 сотрудников фирмы ($n = 10$) имеют автомобили, причем вероятность, что сотрудник поедет на работу на своей машине, равна $p = 0{,}2$. Фирма решила для своих сотрудников арендовать четыре места на платной парковке. Определите вероятность того, что на своих машинах на работу приедут в один день не более четырех сотрудников.

Решение. Искомая вероятность $P_n(m \le 4)$, где $n = 10$, представляет собой сумму вероятностей пяти несовместных событий:

$$P_{10}(m \le 4;\ p = 0{,}2) = P_{10}(0) \quad P_{10}(1) ++ P_{10}(2) + P_{10}(3) + P_{10}(4),$$

где

$$P_{10}(0;\ p = 0{,}2) = C_{10}^{0} \cdot 0{,}2^{0} \cdot 0{,}8^{10} = 1 \cdot 0{,}8^{10} = 0{,}1074;$$

$$P_{10}(1;\ p = 0{,}2) = C_{10}^{1} \cdot 0{,}2^{1} \cdot 0{,}8^{9} = 10 \cdot 2 \cdot 0{,}8^{9} = 0{,}2684;$$

$$P_{10}(2;\ p = 0{,}2) = C_{10}^{2} \cdot 0{,}2^{2} \cdot 0{,}8^{8} = \frac{10 \cdot 9}{2 \cdot 1} \cdot 0{,}2^{2} \cdot 0{,}8^{8} = 0{,}3020;$$

$$P_{10}(3;\ p = 0{,}2) = C_{10}^{3} \cdot 0{,}2^{3} \cdot 0{,}8^{7} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} \cdot 0{,}2^{3} \cdot 0{,}8^{7} = 0{,}2013;$$

$$P_{10}(4;\ p = 0{,}2) = C_{10}^{4} \cdot 0{,}2^{4} \cdot 0{,}8^{6} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} \cdot 0{,}2^{4} \cdot 0{,}8^{6} = 0{,}0881.$$

Отсюда

$$P_{10}\ (m \le 4;\ p = 0{,}2) = 0{,}1074 + 0{,}2684 + 0{,}3020 + 0{,}2013 + 0{,}0881 = 0{,}9672.$$

Таким образом, с вероятностью 0,9672 можно утверждать, что в один день не более четырех сотрудников приедут на работу на своих автомобилях и четырех мест на парковке будет достаточным.

Пример 8.4. По данным прошлых лет известно, что из общего количества проданных фирмой холодильников 40 % приходится на отечественные образцы. Определить вероятность того, что из 500 шт., которые фирма планирует продать: 1) будет реализовано 200 отечественных холодильников; 2) отечественных холодильников будет реализовано от 190 до 230 шт.; 3) доля отечественных холодильников будет находиться в интервале от 0,36 до 0,42.

Решение. По условию примера $p = 0{,}4$ и $n = 500$.

1) Требуется определить вероятность $P_n(m)$, где $n = 500$, а $m = 200$. Поскольку $n = 500$ достаточно велико, то можно воспользоваться нормальным распределением. Согласно (8.31) имеем

$$P_n(m) = \frac{1}{\sqrt{npq}} f(t),$$

где $f(t)$ — значение плотности нормированного нормального распределения, определяется по табл. П.7 для $t = \dfrac{m - np}{\sqrt{npq}} = \dfrac{200 - 500 \cdot 0{,}4}{\sqrt{500 \cdot 0{,}4 \cdot 0{,}6}} = 0.$

Тогда

$$f(t) = f(0) = 0{,}3989 \text{ и } P_{500}(200) = \frac{1}{\sqrt{500 \cdot 4 \cdot 0{,}6}} \cdot 0{,}3989 = 0{,}0364.$$

Таким образом, вероятность того, что из 500 проданных холодильников отечественных будет ровно 200, равна 0,0364.

2) Определим вероятность того, что из $n = 500$ проданных холодильников отечественных будет от 190 до 230 штук, т. е. $P_{500}(190 < m < 230)$.

$$P_{500}(190 < m < 230) = \frac{1}{2}[\Phi(t_2) \quad \Phi(t_1)], -$$

где $t_1 = \dfrac{190 - np}{\sqrt{npq}} = \dfrac{190 - 500 \cdot 0,4}{\sqrt{500 \cdot 0,4 \cdot 0,6}} = -0,91$ и $t_2 = \dfrac{230 - 500 \cdot 0,4}{\sqrt{500 \cdot 0,4 \cdot 0,6}} = 2,74.$

По табл. П.1 найдем

$$\Phi(-0,91) = -0,6372 \text{ и } \Phi(2,74) = 0,9939.$$

Тогда

$$P_{500}(190 < m < 230) = \frac{1}{2}[0,9939 - (-0,6372)] = 0,816.$$

Таким образом, с вероятностью 0,816 можно утверждать, что среди 500 проданных холодильников отечественных будет от 190 до 230 шт.

3) Определим вероятность того, что из $n = 500$ проданных холодильников доля отечественных будет находиться в интервале от $\omega_1 = 0,36$ до $\omega_2 = 0,42$, т. е. $P_{500}(0,36 < \omega < 0,42)$.

$$P_{500}(0,36 < \omega < 0,42) = \frac{1}{2}[\Phi(t_2) \quad \Phi(t_1)], -$$

где $t_1 = \dfrac{\omega_1 - p}{\sqrt{p(1-p)}} = \dfrac{0,36 - 0,4}{\sqrt{\dfrac{0,4 \cdot 0,6}{500}}} = -1,83$ и $t_2 = \dfrac{\omega_2 - p}{\sqrt{p(1-p)}} = \dfrac{0,42 - 0,4}{\sqrt{\dfrac{0,4 \cdot 0,6}{500}}} = 0,91.$

По табл. П.1 найдем:

$$\Phi(-1,83) = -0,9327; \ \Phi(0,91) = 0,6372.$$

Тогда

$$P_{500}(0,36 < \omega < 0,42) = \frac{1}{2}[0,6372 - (-0,927)] = 0,785.$$

Таким образом, с вероятностью 0,785 можно гарантировать, что среди 500 проданных холодильников доля отечественных будет находиться в интервале от 0,36 до 0,42.

Давно было замечено, что результаты отдельных наблюдений (экономических, демографических, физических, метеорологических или иных), хотя и произведенных в относительно однородных условиях, сильно колеблются. В то же время средние из большого числа наблюдений обнаруживают замечательную устойчивость. Математическим обоснованием этого факта служат различные формы так называемого закона больших чисел. *Законом больших чисел* можно назвать общий принцип, в силу которого совокупное действие большого числа случайных факторов приводит при некоторых весьма общих условиях к результату, почти не зависящему от случая. Также хорошо известен «принцип практической невозможности», заключающийся в том, что маловероятные события на практике рассматриваются как невозможные.

Значительная часть задач статистики связана с необходимостью описать большую совокупность объектов. Обычно эту совокупность называет генеральной. Генеральной совокупностью, например, могут быть все жители Москвы, месячная продукция завода, производящего телевизоры, популяция рыб, живущих в данном заливе. Но генеральная совокупность — это не просто множество. Эти слова применимы лишь к тем случаям, когда множество изучается выборочным методом. Если интересующая нас совокупность объектов слишком многочисленна или ее объекты труднодоступны, или имеются другие причины, не позволяющие изучить все объекты, прибегают к изучению какой-то части объектов. Эта выбранная для полного исследования часть называется выборкой. Естественно желание выбрать ее так, чтобы она наилучшим образом представляла целое, была, как говорят, ***репрезентативной***. Как этого добиться? Если целое, т. е. генеральная совокупность, нам мала или совсем неизвестна, не удается предложить ничего лучшего, чем чисто случайный выбор. Бóльшая осведомленность позволяет действовать лучше, но все равно на некоторой стадии наступает незнание и как результат — случайный выбор.

Допустим, на большом поле мы хотим оценить урожай, прежде чем он будет собран. На различных участках растения развивались по-разному. На возвышениях испытывали недостаток влаги, а во впадинах — избыток. Удобрения были рассеяны тоже не вполне равномерно. По краям поля растения больше угнетались сорняками и т. д. Теоретически возможно было бы разделить наше поле на ряд однородных участков и изучить каждый участок отдельно. Однако разбивка на совершенно однородные участки невозможна. Выборочный метод предлагает другой подход. Мысленно разделим наше поле на квадраты площадью, скажем, 1 м^2. Допустим, площадь нашего поля 12 тыс. м^2. Генеральная совокупность состоит теперь из 12 тыс. малых участков по 1 м^2. Допустим, мы решаем обследовать 50 малых участков. (Вопрос о том, сколько участков надо обследовать и как это влияет на точность результата, мы рассмотрим ниже.)

Как их выбрать? Занумеруем все 12 тыс. участков, безразлично, по какому принципу. Важно лишь, чтобы каждый участок имел номер и различные участки получили бы разные номера. Пятьдесят участков для обследования выберем чисто случайно. Этот выбор может быть сделан с помощью таблиц случайных чисел.

Сейчас проводится много социальных обследований. Цели их различны, но всегда связаны с описанием распределения в обществе значений каких-либо признаков, будь то политические симпатии или использование свободного времени. Генеральной совокупностью в зависимости от задач могут быть население города или страны, служащие определенного предприятия или учащиеся высших учебных заведений. Основной метод исследования выборочный, на сплошное обследование не хватает ни времени, ни сил. Но как произвести случайный выбор?

Как правило, случайный отбор идет по легко наблюдаемым признакам, относительно которых мы уверены, что они не связаны с интересующим нас признаком, ради изучения которого ведется исследование. Например, имея в руках список всех сотрудников данного предприятия, можно выделить из этого списка каждого десятого; они-то и составят выборку.

Нарушение принципов случайного выбора обычно приводит к серьезным ошибкам. Стал знаменитым своей неудачей опрос, проведенный американским журналом «Литературное обозрение» относительно исхода президентских выборов в 1936 г. Кандидатами на этих выборах были Ф.Д. Рузвельт и А.М. Ландон. В качестве генеральной совокупности редакция журнала использовала телефонные книги. Отобрав случайно 4 млн адресов, она разослала открытки с вопросами об отношении к кандидатам в президенты по всей стране. Затратив большую сумму на рассылку и обработку открыток, журнал объявил, что на предстоящих выборах президентом США с большим перевесом будет избран А.М. Ландон. Результат выборов оказался противоположным этому прогнозу.

Здесь были совершены сразу две ошибки. Во-первых, телефонные книги не могли дать репрезентативную выборку из населения страны, хотя бы потому, что абоненты в 1936 г. были в основном зажиточные главы семейств. Во-вторых, прислали ответы не все, а люди, не только достаточно уверенные в своем мнении, но и привыкшие отвечать на письма, т.е. в значительной части представители делового мира, которые и поддерживали А.М. Ландона. Явление, подобное только что описанному, когда выборка представляет не всю генеральную совокупность, а лишь какой-то ее слой, какую-то ее часть, называется *смещением выборки*. Смещение — один из основных источников ошибок при использовании выборочного метода.

Такой ошибки избежали социологи Дж. Гэллап и Э. Роупер. Они правильно предсказали победу Ф. Д. Рузвельта, основываясь всего лишь на 4 тыс. анкет. Причиной этого успеха, сделавшего славу его авторам, было не только правильное составление выборки. Они учли, что общество распадается на социальные группы, которые более однородны по отношению к кандидатам в президенты. Значит, выборка из слоя может быть относительно малочисленной с высоким результатом точности. Имея результаты обследования по слоям, можно характеризовать общество в целом.

Из сказанного выше следуют следующие выводы.

Генеральной совокупностью (X) называют множество результатов всех мыслимых наблюдений над значениями одного или нескольких признаков, которые могут быть сделаны при данном комплексе условий. При этом комплекс условий определяет вариацию признаков генеральной совокупности. Синонимом генеральной совокупности в статистике является случайная величина X. ***Выборочной совокупностью*** (выборкой) x_1, x_2, ..., x_n называют множество результатов, случайно отобранных из генеральной совокупности. Выборка должна быть репрезентативной, т. е. правильно отражать пропорции генеральной совокупности. Это достигается случайностью отбора, когда все объекты генеральной совокупности имеют одинаковую вероятность быть отобранными. Задача статистики практически сводится к обоснованному суждению об объективных свойствах генеральной совокупности по результатам случайной выборки.

В основе объяснения перехода от характеристик случайной величины X, которые вычисляются на основе точного знания исследуемого закона распределения, к эмпирическим (выборочным) лежит интерпретация выборки как модели генеральной совокупности, в которой возможными значениями являются наблюдавшиеся (т. е. практически реализованные) значения x_1, x_2, ..., x_n, а в качестве вероятностей берутся соответствующие относительные частоты их появления в выборке, т. е. величины, равные $1/n$. Таким образом, выборку можно представить в табличном виде:

x_1	x_2	...	x_n
$1/n$	$1/n$	...	$1/n$

Условно рассматривая выборку как табличную форму задания дискретной случайной величины, возможные значения которой

x_1, x_2, ..., x_n появляются с одними и теми же вероятностями $p_1 = p_2 = \ldots = p_n = \frac{1}{n}$, легко представить эмпирические аналоги рассмотренных выше начальных (8.16) и центральных (8.17) моментов.

Сказанное проиллюстрируем на примере наиболее часто используемых начального момента первого порядка $\nu_1 = MX$ и центрального момента второго порядка $\mu_2 = DX$.

Согласно (8.6) математическое ожидание дискретной случайной величины с n возможными значениями x_1, x_2, ..., x_n и соответствующими вероятностями $p_1 = p_2 = \ldots = p_n = \frac{1}{n}$ равно:

$$MX = \sum_{i=1}^{n} x_i P_I = \sum_{i=1}^{n} x_i \frac{1}{n} \quad \frac{1}{n}\sum_{i=1}^{n} x_i = \quad \bar{x}. \qquad =$$

Исходя из формулы математического ожидания, мы пришли к формуле средней арифметической (выборочной средней), основной и наиболее употребительной характеристики центра группирования:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{8.35}$$

Таким образом, средняя арифметическая $\bar{x}$ является ***выборочным аналогом математического ожидания*** MX. Дисперсия дискретной случайной величины согласно (8.11) равна:

$$DX = \sum_{i=1}^{n} (x_i \quad MX)^2 p_i.$$

Учитывая, что для нашего примера $MX = \bar{x}$, а $p_i = \frac{1}{n}$, получим

$$DX = \frac{1}{n}\sum_{i=1}^{n} (x_i \quad \bar{x})^2 = s^2.$$

Итак, мы пришли к формуле ***выборочной дисперсии***:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i \quad \bar{x})^2, \quad - \tag{8.36}$$

которая является выборочным аналогом генеральной дисперсии DX.

Рассуждая аналогично, можно получить выборочные аналоги и других моментов генеральной совокупности, а также показать, что относительная частота $\frac{m}{n}$ есть выборочный аналог вероятности p появления некоторого события A в отдельном испытании, если m есть число появления события A в n независимых испытаниях. Выборочным аналогом теоретической функции распределения $F(x)$ (6.2) является функция $\hat{F}(x)$, построенная по выборке объемом n и определяемая соотношением

$$\hat{F}(x) = \frac{\nu(x)}{n}, \quad (8.37)$$

где $\nu(x)$ — число выявленных значений в выборке x_1, x_2, ..., x_n, меньших x.

Из определения эмпирической функции распределения непосредственно следует объяснение часто используемого ее другого названия — «накопленная относительная частота».

Для построения выборочной функции плотности $\hat{f}(\boldsymbol{x})$ по выборке объема n из непрерывной генеральной совокупности X используют предварительно сгруппированные данные (см. п. 2.4) и полагают, что

$$\hat{f}(x) = \frac{\nu_k(\boldsymbol{x})}{n\Delta_k}, \quad (8.38)$$

где k — порядковый номер интервала группирования, в который попала точка x; $\nu(x)$ — число наблюдений, попавших в этот интервал; Δ_k — длина интервала.

В статистике используются два различных варианта интерпретации выборки и ее отдельных элементов.

При первом (практическом) варианте интерпретации ***под выборкой*** x_1, x_2, ..., x_n понимаются фактически выявленные значения исследуемой случайной величины, т. е. конкретные числа.

В соответствии со вторым вариантом интерпретации ***под выборкой*** x_1, x_2, ..., x_n понимается последовательность независимых, одинаково распределенных случайных величин, закон распределения которых совпадает с распределением генеральной совокупности. Таким образом, если генеральная совокупность X имеет нормальное распределение, т. е. $X \sim N(\mu; \sigma)$, то x_i также принадлежит для всех $i = 1, 2, ..., n$ к нормальному распределению с математическим ожиданием $Mx_i = \mu$ и дисперсией $Dx_i = \sigma^2$, т. е. $x_i \sim N(\mu; \sigma)$ для $i = 1, 2, ..., n$.

В соответствии со вторым определением выборки все выборочные характеристики $(\bar{x}, s^2, s, \frac{m}{n})$ являются ***случайными величинами***, как функции от случайных x_i.

В отличие от выборочных, параметры генеральной совокупности $(\mu, \sigma^2, \sigma, p)$ являются ***неслучайными величинами***.

Как уже отмечалось, средние из большого числа наблюдений обнаруживают замечательную устойчивость. К такого рода средним относятся и все рассмотренные выше выборочные характеристики. Математическим обоснованием этого факта служат различные формы закона больших чисел. Закон позволяет теоретически обосновать устойчивость основных выборочных характеристик распределения (среднего значения, дисперсии, функции распределения и плотности, построенных по выборке x_1, x_2, ..., x_n). При этом показателем устойчивости служит дисперсия соответствующей выборочной характеристики.

Проиллюстрируем сказанное на примере выборочной средней $\overline{x}$, полученной по выборке x_1, x_2, ..., x_n, взятой из генеральной совокупности x с математическим ожиданием $Mx_i = \mu$ и дисперсией $Dx_i = \sigma^2$.

Отсюда, в соответствии со вторым определением выборки, следует, что ее i-й элемент x_i есть случайная величина с математическим ожиданием $Mx_i = \mu$ и дисперсией $Dx_i = \sigma^2$ для всех $i = 1, 2, ..., n$. При этом элементы выборки x_1, x_2, ..., x_n взаимно независимы. Тогда дисперсия средней арифметической x равна $D\overline{x} = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$.

В соответствии со свойствами дисперсии (8.14) и (8.15) можно вынести множитель $\frac{1}{n}$ за знак дисперсии, возведя предварительно его в квадрат, и поменять местами знаки дисперсии и суммирования. Тогда будем иметь

$$D\overline{x} = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i \quad \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 \quad \frac{\sigma^2}{n} \quad =$$

и окончательно

$$D\overline{x} = \frac{\sigma^2}{n}. \tag{8.39}$$

Таким образом, согласно (8.39) по мере увеличения объема выборки дисперсия $\overline{x}$ будет уменьшаться и средняя будет стремиться к постоянной величине, определяемой генеральной средней μ.

Чтобы убедиться в этом, определим математическое ожидание $\overline{x}$. С учетом свойств (8.8) и (8.9) математического ожидания будем иметь

$$M\overline{x} = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Mx_i \quad \frac{1}{n}n\mu = \mu.$$

В приведенном виде

$$M\bar{x} = \mu \qquad (8.40)$$

Поэтому по мере увеличения n величина $\bar{x}$ будет приближаться к $M\bar{x} = \mu$ со все меньшей вариацией.

Аналогично, относительная частота $w = \dfrac{m}{n}$ какого-либо события по мере увеличения объема выборки n будет более точно характеризовать вероятность p этого события, так как согласно (8.33) математическое ожидание w равно $Mw = p$, а среднее квадратическое отклонение w равно $\sigma_w = \sqrt{\dfrac{p(1-p)}{n}}$.

Среднее квадратическое отклонение выборочных характеристик $\bar{x}$ и w с одной стороны характеризует их вариацию относительно математического ожидания, а с другой — *ошибку выборки* при оценке соответствующего параметра генеральной совокупности (μ или p). В этой связи среднее квадратическое отклонение называют ***стандартной ошибкой*** соответственно средней $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ или относительной частоты $\sigma_w = \sqrt{\dfrac{p(1-p)}{n}}$.

Статистические оценки стандартной ошибки выборки. Во многих случаях параметры генеральной совокупности μ, σ и p неизвестны, а известны лишь полученные по выборке их оценки, значения средней арифметической $\bar{x}$, выборочного среднего квадратического отклонения s или относительной частоты $w = \dfrac{m}{n}$. Тогда оценка значения средней квадратической ошибки средней арифметической $\sigma_{\bar{x}}$ определяют по формуле

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}, \text{ где } s \quad \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i \equiv \bar{x})^2}. \qquad (8.41)$$

Оценку значения средней квадратической ошибки относительной частоты σ_w находят с помощью формулы

$$s_w = \sqrt{\frac{w(1-w)}{n}}. \qquad (8.42)$$

При достаточно больших объемах выборки ($n > 30$ для $\bar{x}$ и $n > 100$ для w) можно считать с учетом (8.41) и (8.42), что нормальный закон распределения имеет выборочные характеристики:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{s}\sqrt{n} \in \underset{n>30}{N}(0;1), \tag{8.43}$$

т. е. $Mt = 0$; $Dt = 1$, $s_{\bar{x}} = \frac{s}{\sqrt{n}}$, а также

$$t = \frac{w - p}{s_w} = \frac{w - p}{\sqrt{\frac{w(1-w)}{n}}} \in \underset{n>100}{N}(0;1), \text{ где } s_w = \sqrt{\frac{w(1-w)}{n}}. \tag{8.44}$$

Из (8.23) следует, что вероятность попадания случайной величины t в интервал от $-t_\gamma$ до t_γ равна:

$$P(-t_\gamma < t < t_\gamma) = \Phi(t_\gamma) = \gamma, \tag{8.45}$$

где γ — заданная вероятность, а значения t_γ определяются по табл. П.1 из условия $\Phi(t_\gamma) = \gamma$.

Подставив в (8.45) значения t согласно выражениям (8.43) и (8.44) и решив неравенства относительно параметров μ и p соответственно, получим первый результат:

$$P(\bar{x} - t_\gamma s_{\bar{x}} < \mu < \bar{x} + t_\gamma s_{\bar{x}}) = \gamma, \tag{8.46}$$

где $\delta_{\bar{x}} = t_\gamma s_{\bar{x}} = t_\gamma \frac{s}{\sqrt{n}}$, $\delta_{\bar{x}}$ — предельная ошибка средней $\bar{x}$.

Второй результат будет таким:

$$P(w - t_\gamma s_w < p < w + t_\gamma s_w) = \gamma, \tag{8.47}$$

где $\delta_w = t_\gamma s_w = t_\gamma \sqrt{\frac{w(1-w)}{n}}$, а δ_w — предельная ошибка относительной частоты w.

Формулы (8.46) и (8.47) определяют также интервал, в котором с вероятностью γ будут находиться неизвестные параметры генеральной совокупности, соответственно математическое ожидание μ и вероятность p.

Достоверность рассчитанных по выборочным данным характеристик в значительной степени определяется репрезентативностью выборочной совокупности, которая в свою очередь зависит от способа отбора единиц из генеральной совокупности. В каждом конкретном случае в зависимости от сущности исследуемого явления, объема совокупности, вариации и распределения наблюдаемых признаков, материальных и трудовых ресурсов, а также иных условий выбирают наиболее предпочтительную систему организации отбора, которая определяется видом, методом и способом отбора.

По виду различают индивидуальный, групповой и комбинированный отбор. При ***индивидуальном отборе*** в выборочную совокупность отбираются отдельные единицы генеральной совокупности, при ***групповом*** — группы единиц, при ***комбинированном*** — сочетается групповой и индивидуальный отбор.

Метод отбора определяет возможность продолжения участия отобранной единицы в процедуре отбора.

Бесповторным называется такой отбор, при котором попавшая в выборку единица не возвращается в совокупность, из которой осуществляется дальнейший отбор. При ***повторном отборе*** попавшая в выборку единица после регистрации наблюдаемых признаков возвращается в исходную (генеральную) совокупность для участия в дальнейшей процедуре отбора. При этом методе отбора объем генеральной совокупности на всем протяжении процедуры выборки остается неизменным, что обусловливает постоянную вероятность попадания в выборку всех единиц совокупности.

Повторный метод отбора применяется в тех случаях, когда характер исследуемого явления предполагает возможность повторной регистрации единиц. Такая возможность, прежде всего, может иметь место в выборочных обследованиях населения в качестве покупателей, пациентов, избирателей, абитуриентов и т. д. К повторному приравнивается и отбор из совокупности, значительно превышающий объем выборки, а также из совокупности, границы которой не определены, например, вследствие непрерывного производственного процесса. В подобных случаях значения отобранных единиц рассматривают как гипотетические величины, не исключающие возможности многократного повторения.

Способ отбора определяет конкретный механизм или процедуру выборки единиц из генеральной совокупности. В практике выборочных обследований наибольшее распространение получили собственно-случайная, механическая, типическая, серийная, комбинированная выборки.

Собственно-случайная выборка заключается в отборе единиц из генеральной совокупности наугад, наудачу, без каких-либо элементов системности. Однако, прежде чем производить собственно-случайный отбор, необходимо убедиться, что все без исключения единицы генеральной совокупности имеют абсолютно равные шансы попадания в выборку, а в списках или перечне отсутствуют пропуски, игнорирования отдельных единиц и т. п. Следует также установить четкие границы генеральной совокупности, чтобы включение или невключение в нее отдельных единиц не вызывало сомнений. Так, при обследовании студентов необходимо указать, будут ли приниматься во внимание лица, находящиеся в академическом отпуске, студенты негосударственных вузов, военных училищ и т. п. При обследовании торговых предприятий важно определиться, включит ли генеральная совокупность торговые павильоны, коммерческие палатки и прочие подобные объекты. Технически собственно-случайный отбор проводят методом жеребьевки или по таблице случайных чисел.

Для ***жеребьевки*** необходимо подготовить достаточное количество жребиев — фишек, шаров, карточек в соответствии с объемом генеральной совокупности. Каждый жребий должен содержать информацию об отдельной единице совокупности — номер, фамилию лица или адрес, название или какой-либо другой отличительный признак. Необходимое в соответствии с установленным процентом отбора количество жребиев извлекается из общей их совокупности в случайном порядке.

При отборе ***по таблицам случайных чисел*** каждая единица генеральной совокупности должна иметь порядковый номер. Таблицы случайных чисел получаются с помощью датчика случайных чисел на ЭВМ и представляют собой абсолютно произвольные столбцы цифр. В соответствии с объемом генеральной совокупности выбирается любой столбец с числами необходимой значимости. Например, если генеральная совокупность включает 5 000 единиц, потребуются четырехзначные столбцы, при этом числа больше 5 000 не будут приниматься во внимание. В выборочную совокупность отбираются единицы с порядковыми номерами, соответствующими числам выбранного столбца.

Собственно-случайный отбор может быть как повторным, так и бесповторным. Для проведения бесповторного отбора в процессе жеребьевки выпавшие жребии обратно в исходную совокупность не возвращаются и в дальнейшем отборе не участвуют. При использовании таблиц случайных чисел бесповторность отбора достигается пропуском чисел в случае их повторения в выбранном столбце или столбцах.

После проведения отбора рассчитываются средняя и предельная ошибка выборочных характеристик $\bar{x}$ и w.

При повторной собственно-случайной выборке стандартная ошибка средней $\bar{x}$ и относительной частоты $w = \dfrac{m}{n}$ определяется по формулам (8.41) и (8.42):

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \quad \text{и} \quad s_w = \sqrt{\frac{w(1-w)}{n}}.$$

Предельные ошибки этих характеристик с вероятностью γ выявляются по формулам (8.46) и (8.47): $s_{\bar{x}} = t_\gamma s_{\bar{x}}$ и $s_w = t_\gamma s_w$, где t_γ находят по табл. П.1 из условия $\Phi(t_\gamma) = \gamma$.

При бесповторном отборе в формулу выборочной средней квадратической ошибки средней ($\bar{x}$) и относительной частоты (w) необходимо ввести поправки на численность генеральной совокупности. После этого формулы (8.41) и (8.42) приводят к виду:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N}} \tag{8.48}$$

и

$$s_w = \sqrt{\frac{w(1-w)}{n}} \cdot \sqrt{\frac{N-n}{N}}. \tag{8.49}$$

Пример 8.5. В табл. 8.5 представлены результаты обследования жилищных условий 200 жителей ($n = 200$) города, полученные на основе собственно-случайной повторной выборки.

Требуется определить с доверительной вероятностью 0,95: 1) предельную ошибку выборочной средней $\bar{x}$ и интервал, в котором может находиться математическое ожидание μ, средний размер общей площади жилья, приходящегося на одного человека; 2) предельную ошибку относительной частоты w, долю жителей, обеспеченность жильем которых составляет менее 15 м^2 и интервал, в котором может находиться вероятность p этого события.

Общая площадь жилищ, приходящихся на одного человека, м^2 $(a_i — b_i)$	Число жителей (m_i)	Общая площадь жилищ, приходящихся на одного человека, м^2 $(a_i — b_i)$	Число жителей (m_i)
До 5,0	2	20,0—25,0	42
5,0—10,0	19	25,0—30,0	26
10,0—15,0	41	30,0—35,0	16
15,0—20,0	54		

Решение. 1. Определим предварительно стандартную ошибку средней $s_{\bar{x}}$. С этой целью рассчитаем среднюю арифметическую $\bar{x}$ и выборочную дисперсию по данным табл. 8.5. Предварительные вычисления представим в виде расчетной табл. 8.6.

Далее, учитывая, что $n = \sum_{i=1}^{l} m_i = 200$, где $l = 7$ — число интервалов, получим

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{l} x_i m_i = \frac{1}{200} \cdot 3\,785 = 18{,}925 \text{ м}^2,$$

а

$$s^2 = \overline{x^2} - (\bar{x})^2 = \frac{81\,850}{200} - (18{,}925)^2 = 51{,}094.$$

Отсюда следует, что $s = 7{,}148$ м^2. Тогда стандартная ошибка средней согласно (8.41) равна:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{7{,}148}{\sqrt{200}} = 0{,}505.$$

Для определения предельной ошибки $\bar{x}$ из условия $\Phi(t_\gamma) = \gamma = 0{,}95$ по табл. П.1 найдем $t_\gamma = 1{,}96$. Тогда, согласно (8.46):

$$\delta_{\bar{x}} = t_\gamma s_{\bar{x}} = 1{,}96 \cdot 0{,}507 = 0{,}994,$$

а интервальная оценка генеральной средней равна

$$\bar{x} - \delta_{\bar{x}} \le \mu \le \bar{x} + \delta_{\bar{x}},$$

или с учетом полученных значений $17{,}93 \le \mu \le 19{,}92$.

Границы интервалов a_i—b_i	Число жителей m_i	Середина интервала x_i	$x_i m_i$	$x_i^2 m_i$
До 5,0	2	2,5	5,0	12,50
05,0—10,0	19	7,5	142,5	1 068,75
10,0—15,0	41	12,5	512,5	6 406,25
15,0—20,0	54	17,5	945,0	16 537,50
20,0—25,0	42	22,5	945,0	21 262,50
25,0—30,0	26	27,5	715,0	19 662,50
30,0—35,0	16	32,5	520,0	16 900,00
Итого	200	—	3 785	81 850,00

Таким образом, с вероятностью 0,95 можно утверждать, что средний размер общей площади, приходящейся на одного человека, в целом по городу находится в пределах от 17,9 до 19,9 м^2.

При расчете стандартной ошибки собственно-случайной бесповторной выборки согласно (8.48) имеем

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \quad \sqrt{1-\frac{n}{N}}.$$

Если предположить, что данные табл. 8.5 являются результатом 5%-го бесповторного отбора (т. е. генеральная совокупность включает 4 000 тыс. единиц), то стандартная ошибка $\bar{x}$ будет несколько меньше:

$$s_{\bar{x}} = \frac{7{,}148}{\sqrt{200}} \quad \sqrt{1-\frac{200}{4\,000}} = 0{,}493 \text{ м}^2.$$

Соответственно уменьшится и предельная ошибка выборки, что вызовет сужение границ генеральной средней μ. Особенно ощутимое влияние поправки на бесповторность отбора при большом проценте выборки.

2. Теперь решим задачу для доли лиц, обеспеченность жильем которых составляет менее 15 м^2. Согласно табл. 8.5, численность таких лиц составляет $m = 62$ чел., а относительная частота

$$w = \frac{m}{n} \quad \frac{62}{200} = 0{,}31.$$

Тогда стандартная ошибка относительной частоты согласно (8.42) равна:

$$s_w = \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0{,}31 \cdot 0{,}69}{200}} = 0{,}0327,$$

а предельная ошибка относительной частоты (8.47) с вероятностью 0,95 составит

$$\delta_w = t_\gamma s_w \quad 1{,}96 = 0{,}0327 = 0{,}0641.$$

Определим интервальную оценку генеральной доли, вероятности p. С учетом (8.47) получим: $w - \delta_w \leq p \leq w + \delta_w$, а с учетом полученных значений $0{,}31 - 0{,}064 \leq p \leq 0{,}31 + 0{,}064$ или $0{,}246 \leq p \leq 0{,}374$.

Следовательно, с вероятностью 0,95 можно утверждать, что доля лиц, имеющих общую площадь менее 15 м^2 на человека, по городу в целом находится в пределах от 24,6 до 37,4 %.

Механическая выборка применяется в 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 такой выборки устанавливается пропорция отбора, которая определяется соотнесением объемов выборочной и генеральной совокупностей. Допустим, из совокупности в 500 000 единиц предполагается получить 2%-ю выборку, т. е. отобрать 10 000 единиц. 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.

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 значения изучаемого признака (если из каждого интервала регистрируется первое значение) или его завышением (если из каждого интервала регистрируется последнее значение). Поэтому целесообразно отбор начинать с середины первого интервала, например при 5%-й выборке отобрать 10, 30, 50, 70 и с таким же интервалом последующие единицы.

Опасность систематической ошибки при механической выборке также может появиться вследствие случайного совпадения выбранного интервала и циклических закономерностей в расположении единиц генеральной совокупности. Так, при Всесоюзной переписи населения 1989 г. в ходе 25%-го выборочного обследования семей имела место опасность попадания в выборку квартир только одного типа (например, только однокомнатных или только трехкомнатных), так как на лестничных площадках многих типовых домов располагается именно по 4 квартиры. Чтобы избежать систематической ошибки, в каждом новом подъезде счетчик менял начало отбора.

Типическая выборка используется в тех случаях, когда все единицы генеральной совокупности можно разбить на несколько типических групп. При обследованиях населения такими группами могут быть, например, районы, социальные, возрастные или образовательные группы, при обследовании предприятий — отрасль и подотрасль, форма собственности и т. п. Типический отбор предполагает выборку единиц из каждой типической группы 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, которая в этом случае определяется только внутригрупповой вариацией.

Отбор единиц в типическую выборку может быть организован либо пропорционально объему типических групп, либо пропорционально внутригрупповой дифференциации признака. При выборке, пропорциональной объему типических групп, число единиц, подлежащих отбору из каждой группы, определяется следующим образом:

$$n_j = n\frac{N_j}{N}, \tag{8.50}$$

где N_j — объем j-й группы, а N — объем генеральной совокупности, $N = \sum_{j=1}^{l} N_j$; l — число типических групп; n_j — объем выборки из j-й группы, а n — объем суммарной выборки, $n = \sum_{j=1}^{l} n_j$.

Тогда выборочная средняя квадратическая ошибка средней $(\bar{x})$ определяется по конкретным формулам. При повторном отборе

$$s_{\bar{x}} = \sqrt{\frac{\bar{s}^2}{n}}, \tag{8.51}$$

где $\bar{s}^2$ — средняя арифметическая групповых дисперсий:

$$\bar{s}^2 = \frac{\sum_{j=1}^{l} s_j^2 n_j}{\sum_{j=1}^{l} n_j}; \tag{8.51′}$$

$s_j^2 = \frac{1}{n}\sum_{i=1}^{n_j}(x_{ij}\quad \bar{x}_j)^2$; x_{ij} — результат i-го наблюдения, полученного по выборке из j-й группы.

При бесповторном отборе дело сводится к формуле

$$s_{\bar{x}} = \sqrt{\frac{\bar{s}^2}{n}\left(1 \quad \frac{n}{N}\right)}. \; - \tag{8.52}$$

Выборочная средняя квадратическая ошибка относительной частоты определяется следующим образом.

При повторном отборе:

$$s_{\omega} = \sqrt{\frac{1}{n}\sum_{j=1}^{l} s_{\omega_j}^2 n_j} = \sqrt{\frac{1}{n}\sum_{j=1}^{l} \omega_j(1-\omega_j)}, \tag{8.53}$$

где $s_{\omega_j}^2 = \frac{\omega_j(1-\omega_j)}{n_j}$; ω_j — относительная частота, полученная из j-й группы.

При бесповторном отборе:

$$s_{\omega} = \sqrt{\left(\frac{1}{n} \quad \frac{1}{N}\right)\sum_{j=1}^{l} \omega_j (1-\omega_j)}. \tag{8.54}$$

Пример 8.6. С целью оценки величины потерь от временной нетрудоспособности работников на машиностроительном заводе проведено выборочное обследование, в ходе которого была взята 10%-я бесповторная типическая выборка. Отбор проводится пропорционально численности работников цехов завода (табл. 8.7).

Требуется определить: 1) среднюю и предельную (с вероятностью 0,954) ошибку выборочной средней $\bar{x}$; 2) с вероятностью 0,954 определить интервальную оценку среднего времени нетрудоспособности (μ) одного работника завода за год.

Решение. 1. Предварительно определим среднюю из внутригрупповых дисперсий (8.51′), учитывая, что $n = n_1 + n_2 + n_3 = 320$:

$$\bar{s}^2 = \frac{1}{n}\sum_{j=1}^{l} s_j^2 n_j = \frac{49\cdot 100 + 25\cdot 140 + 16\cdot 80}{100 + 140 + 80} = 30{,}25.$$

Тогда средняя квадратическая ошибка средней $\bar{x}$ согласно (8.52) равна:

$$s_{\bar{x}} = \sqrt{\frac{\bar{s}^2}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{30{,}25}{320}\left(1-\frac{320}{3\,200}\right)} = 0{,}29.$$

Для определения предельной ошибки средней по табл. П.1 из условия $\Phi(t_\gamma) = 0{,}954$ найдем $t = 2$. Тогда $\delta_{\bar{x}} = t_\gamma s_{\bar{x}} = 2\cdot 0{,}29 = 0{,}58$.

2. Интервальная оценка генеральной средней μ согласно (8.46) равна:

$$\bar{x} - \delta_{\bar{x}} \le \mu \le \bar{x} + \delta_{\bar{x}},$$

где $\bar{x} = \frac{1}{n}\sum_{j=1}^{l} \bar{x}_j n_j = \frac{1}{320}(18\cdot 100 + 12\cdot 140 + 15\cdot 80) = 14{,}6.$

Отсюда $14{,}6 - 0{,}58 \le \mu \le 14{,}6 + 0{,}58$ и окончательно $14{,}02 \le \mu \le 15{,}18$.

Таким образом, с вероятностью $\gamma = 0{,}954$ можно утверждать, что среднее число дней временной нетрудоспособности одного работника в год находится в интервале от 14,02 до 15,18 дней.

Номер цеха, i	Численность работников, N_i	Объем выборки n_i	Число дней временной нетрудоспособности за год	
			Средняя ($\bar{x}_j$)	Дисперсия (s_j^2)
11	1 000	100	18	49
2	1 400	140	12	25
3	800	80	15	16

Серийная выборка удобна в тех случаях, когда единицы совокупности объединены в небольшие группы или серии. В качестве таких серий могут рассматриваться партии товара, студенческие группы, бригады и другие объединения. Сущность серийной отборки заключается в собственно-случайном либо механическом отборе серий, внутри которых проводится сплошное обследование единиц.

Пусть генеральная совокупность состоит из R серий одинакового объема n, из которых случайным образом отобрано r серий. Тогда оценка средней квадратической ошибки средней определяется по следующим формулам.

При повторном отборе:

$$s_{\bar{x}} = \sqrt{\frac{s^2_{\text{м.г}}}{r}},$$

где $s^2_{\text{м.г}} = \frac{1}{r}\sum_{j=1}^{r}(\bar{x}_j \quad \bar{x})^2$ ——межгрупповая дисперсия; $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ — средняя арифметическая j-й серии; x_{ij} — результат i-го наблюдения в j-й серии; $\bar{x} = \frac{1}{r}\sum_{j=1}^{r}\bar{x}_j$ — общая средняя.

При бесповторном отборе:

$$s_{\bar{x}} = \sqrt{\frac{s^2_{\text{м.г}}}{r}\left(1 \quad \frac{r}{R}\right)}. \quad -$$

В практике статистических обследований помимо рассмотренных выше способов отбора применяется и их комбинация — ***комбинированная выборка***. Так бывает, когда серии отбираются в установленном порядке из нескольких типических групп. Возможна также комбинация серийного и собственно случайного отборов. Тогда отдельные единицы отбираются внутри серии в собственно-случайном порядке. 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.

Многофазная выборка в отличие от многоступенчатой предполагает сохранение одной и той же единицы отбора на всех этапах его проведения. При этом отобранные на каждой стадии единицы подвергаются обследованию (на каждой последующей стадии отбора программа обследования расширяется).

При проектировании выборочных обследований возникает вопрос о необходимой численности выборки n. Объем выборки обычно определяют по формуле предельной ошибки соответствующей выборочной характеристики $\delta_{\bar{x}}$ или δ_w, которая характеризует точность оценивания параметра генеральной совокупности (μ или p).

Задача определения n возникает тогда, когда предварительно проведенное выборочное обследование, позволившее определить значение выборочной дисперсии s^2, не обеспечило при заданной вероятности γ требуемого значения предельной ошибки.

В табл. 8.8 приведены формулы необходимого объема выборки для наиболее часто используемых на практике способов формирования выборочной совокупности.

Вид выборочного наблюдения	Отбор	
	повторный	бесповторный
Собственно-случайная или механическая выборка для частоты:		
а) средней	$n = \dfrac{t^2 s^2}{\delta_{\bar{x}}^2}$	$n = \dfrac{t^2 s^2 N}{\delta_{\bar{x}}^2 N + t^2 s^2}$
б) относительной	$n = \dfrac{t^2 w(1-w)}{\delta_w^2}$	$n = \dfrac{t^2 w(1-w)}{\delta_w^2 N + t^2 w(1-w)}$
Типическая выборка для частоты:		
а) средней	$n = \dfrac{t^2 \overline{s}^2}{\delta_{\bar{x}}^2}$	$n = \dfrac{t^2 \overline{s}^2 N}{\delta_{\bar{x}}^2 N + t^2 \overline{s}^2}$
б) относительной	$n = \dfrac{t^2 \sum_{j=1}^{1} w_j (1-w_j)}{\delta_{\bar{w}}^2}$	$n = \dfrac{t^2 N \sum_{j=1}^{1} w_j (1-w_j)}{N\delta_{\bar{w}}^2 + t^2 \sum_{j=1}^{1} w_j (1-w_j)}$
Серийная выборка для средней частоты	$r = \dfrac{t^2 s_{\text{м.г}}^2}{\delta_{\bar{x}}^2}$	$r = \dfrac{t^2 s_{\text{м.г}}^2 R}{\delta_{\bar{x}}^2 R + t^2 s_{\text{м.г}}^2}$

Пример 8.7. В микрорайоне проживает 5 000 семей. Требуется определить минимальный объем случайной бесповторной выборки, который позволит с вероятностью $p = 0{,}95$ оценить средний размер семьи с предельной ошибкой средней, равной 0,4 при среднем квадратическим отклонением $s = 2$ чел., найденном по данным предварительного обследования.

Решение. По табл. П.1 для $p = 0{,}95$ найдем $t = 1{,}96$. Тогда численность выборки (см. табл. 8.7) равна: $n = \dfrac{1{,}96^2 \cdot 2^2 \cdot 5\,000}{5\,000 \cdot 0{,}16 + 1{,}96^2 \cdot 2^2} \approx 956$ семей.

Пример 8.8. В фермерских хозяйствах области 10 000 коров. Из них в районе А — 5 000, в районе Б — 3 000, в районе В — 2 000. С целью определения средней удойности предполагается провести типическую выборку коров с пропорциональным отбором внутри групп (механическим). Какое количество коров следует отобрать, чтобы с вероятностью 0,954 ошибка выборки не превышала 5 л, если на основе предыдущих обследований известно, что дисперсия типической выборки равна 1 600?

Решение. Рассчитаем необходимую численность типической выборки.

По табл. П.1 для $p = 0{,}954$ найдем $t = 2{,}0$. Тогда согласно табл. 8.8

$$n = \frac{2^2 \cdot 1\,600 \cdot 10\,000}{5^2 \cdot 10\,000 + 2^2 \cdot 1\,600} = \frac{64\,000\,000}{250\,000 + 6\,400} = 249{,}6 \approx 250 \text{ коров.}$$

Следовательно, необходимо отобрать 250 коров, из них в районе А: $n_1 = 250 \cdot \dfrac{5\,000}{10\,000} = 125$ коров; в районе Б: $n_2 = 250 \cdot \dfrac{3\,000}{10\,000} = 75$ коров; в районе В: $n_3 = 250 \cdot \dfrac{2\,000}{10\,000} = 50$ коров.

Пример 8.9. На склад АО «Машиностроитель» поступило 100 ящиков готовых изделий по 80 шт. в каждом. Для установления среднего веса деталей следует провести серийную выборку деталей методом механического отбора так, чтобы с вероятностью 0,954 ошибка выборки не превышала 2 г. На основе предыдущих обследований известно, что межгрупповая дисперсия равна 5. Предстоит определить необходимый объем выборки.

Решение. $r = \dfrac{2^2 \cdot 4 \cdot 100}{100 \cdot 2^2 + 2^2 \cdot 4} = \dfrac{1\,600}{416} \approx 5$ ящиков.

Подробное рассмотрение вопросов определения дисперсии для нахождения объема выборки не исключает использование в этих целях других показателей вариации. В последние годы были проведены исследования по разработке методики определения необходимой численности выборки при заданных значениях отдельных параметров и, в частности, коэффициентов вариации.

1. Отклонение выборочных характеристик от соответствующих характеристик генеральной совокупности, возникающее вследствие нарушения принципа случайности отбора, называется:
   а) систематической ошибкой репрезентативности;
   б) случайной ошибкой репрезентативности.

2. Отклонение выборочных характеристик от соответствующих характеристик генеральной совокупности, возникающее вследствие несплошного характера наблюдения, называется:
   а) систематической ошибкой репрезентативности;
   б) случайной ошибкой репрезентативности.

3. Чтобы уменьшить ошибку выборки, рассчитанную в условиях механического отбора, можно:
   а) уменьшить численность выборочной совокупности;
   б) увеличить численность выборочной совокупности;
   в) применить серийный отбор;
   г) применить типичный отбор.

4. Средняя из групповых дисперсий в генеральной совокупности составляет 64 % общей дисперсии. Средняя ошибка выборки при механическом отборе из этой совокупности будет при одном и том же объеме выборки больше ошибки типической выборки:
   а) на 36 %;
   б) на 64 %;
   в) на 25 %;
   г) предсказать результат невозможно.

5. Проведено собственно-случайное бесповторное обследование заработной платы сотрудников аппарата управления двух финансовых корпораций. Обследовано одинаковое число сотрудников. Дисперсия заработной платы для двух финансовых корпораций одинакова, а численность аппарата управления больше на первой корпорации. Средняя ошибка выборки:
   а) больше на первой корпорации;
   б) больше на второй корпорации;
   в) на обеих корпорациях одинакова;
   г) данные не позволяют сделать вывод.

6. Проведено обследование: 1) восьми кафе района с целью изучения их санитарного состояния; 2) шести магазинов из 40, переведенных на новый график работы, с целью определения эффективности внедрения нового графика в магазинах города. Выборочным обследованием является:
   а) никакое из перечисленных;
   б) 1; 2;

в) 1;
г) 2.

7. По данным 10%-го выборочного обследования дисперсия средней заработной платы сотрудников первого туристического агентства 225, а второго — 100. Численность сотрудников первого туристического агентства в четыре раза больше, чем второго. Ошибка выборки больше:
   а) в первом туристическом агентстве;
   б) во втором туристическом агентстве;
   в) ошибки одинаковы;
   г) предсказать результат невозможно.

8. При выборочном обследовании продуктивности скота в фермерских хозяйствах вначале отбирались группы фермерских хозяйств определенного производственного направления, а в отобранных группах — отдельные хозяйства. Этот отбор:
   а) серийный;
   б) типический;
   в) двухступенчатый;
   г) двухфазный.

9. При отборе рабочих экспедиторских фирм для обследования причин потерь рабочего времени были заведомо исключены рабочие, имеющие сокращенный рабочий день. Результаты обследования содержат:
   а) систематическую ошибку регистрации;
   б) систематическую ошибку репрезентативности.

10. На таможенном посту проверено 36 % ручной клади пассажиров. Ошибка собственно-случайной бесповторной выборки меньше ошибки повторной выборки:
    а) на 10 %;
    б) на 19 %;
    в) на 1 %;
    г) предсказать результат невозможно.

11. С целью определения трудоемкости изготовления деталей на предприятии произведен хронометраж работы 50 рабочих, отобранных в случайном порядке. По данным обследований получили $\bar{x} = 10$ мин, при $s = 1$ мин. Определите:
    а) как изменится ошибка выборки, если объем выборочной совокупности увеличить в 1,5 раза;
    б) как скажется на ошибке выборки увеличение дисперсии в 2 раза;
    в) как изменится ошибка выборки, если с увеличением дисперсии в 1,44 раза объем выборочной совокупности увеличить в 2,56 раза;
    г) как изменится ошибка выборки, если численность генеральной совокупности будет в 3 раза больше?

12. Из партии импортируемой продукции на посту Московской региональной таможни было взято в порядке случайной повторной выборки 20 проб продукта А. В результате проверки установлена средняя влажность продукта А в выборке, которая оказалась равной 6 % при среднем квадратическом отклонении 1 %. С вероятностью 0,683 определите пределы средней влажности продукта во всей партии импортируемой продукции.

13. В порядке механической выборки обследован возраст 100 студентов вуза из общего числа 2 000 чел. Результаты обработки материалов наблюдения приведены в таблице:

Возраст, лет	17	18	19	20	21	22	23
Число студентов, чел.	11	13	18	23	17	10	8

Установите: а) средний возраст студентов вуза по выборке; б) величину стандартной ошибки при определении возраста студентов на основе выборки; в) вероятные пределы колебания возраста для всех студентов при вероятности 0,997.

14. Для выборочного обследования занятости мужского населения сельских районов республики имеются следующие данные:

Район	Численность мужчин трудоспособного возраста, тыс. чел.	Удельный вес занятости мужчин, %
1	3,5	75
2	56	80
3	1,7	70
4	2,8	85

С вероятностью 0,954 определите необходимый объем типической пропорциональной выборки для установления границ генеральной доли: а) при повторном отборе; б) при бесповторном отборе в районах, чтобы ошибка выборки не превышала 5 %.

15. Генеральная совокупность содержит данные по 2 400 семьям области, в том числе проживающим в областном центре — 960, городах — 720, поселках городского типа — 480, сельской местности — 240. С целью исследования дифференциации населения по уровню доходов планируется организовать типическую пропорциональную выборку семей с механическим отбором внутри групп. По результатам аналогичного обследования в других регионах среднее квадратическое отклонение составило 60. Сколько семей необходимо отобрать из каждой группы, чтобы ошибка не превышала 20 единиц при вероятности 0,997?

Исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами играет в развитии экономики значительную роль. Оно позволяет глубже понять сложный механизм причинно-следственных отношений. В настоящее время важно уметь количественно измерить тесноту причинно-следственных связей и выявить форму связи между экономическими процессами. Для исследования интенсивности, вида и формы причинных связей широко применяется корреляционный и регрессионный анализ. Выявление количественных соотношений дает возможность лучше понять природу исследуемого явления. Это в свою очередь позволяет воздействовать на изученные факторы, вмешиваться в соответствующий процесс с целью получения нужных результатов.

Но чтобы глубоко и основательно проникнуть в суть явления, необходимо исследовать и раскрыть его причинные связи, его отношения с другими явлениями. Под причинной связью понимают такую связь, когда изменение одних процессов есть следствие изменения других. Обычно одно и то же экономическое явление выступает как результат, следствие, эффект одной или нескольких причин. Вместе с тем оно служит причиной наступления других явлений или процессов. Раскрытие объективно существующих причинных зависимостей приводит исследователя к источнику зарождения отдельных процессов.

Признание факта множественности причин и следствий в реальной действительности нашло свое отражение и при исследовании закономерностей в экономике. Так, на величину себестоимости единицы продукции влияют объем производства, используемая технология и уровень производительности труда. Производительность труда, которая служит причиной формирования себестоимости, в свою очередь является следствием таких причин, как уровень развития техники и подготовки работников, эффективность использования парка оборудования и т. д. Урожайность сельскохозяйственных культур зависит от состояния почвы, состава и количества внесенных удобрений, метеорологических условий и других не менее важных причин.

Один из важных признаков причинной связи — это соблюдение ***временно́й последовательности*** причины и следствия. Причина всегда предшествует следствию. Однако не всякое предшествующее событие служит подлинной причиной появления последующего. Поэтому для правильного понимания причинно-следственных отношений большую опасность представляют совпадения явлений и одновременно развивающиеся процессы. Например, увеличение числа онкологических заболеваний за последние 10 лет ни в коей мере не является причиной спада промышленного производства за тот же период времени.

Следует также отметить, что статистический анализ требует такого обязательного условия, как повторяемость явления. Ведь только наличие достаточно большого числа наблюдений обеспечивает практическую возможность выявления связи. Это обусловлено тем, что причинному действию и определяемому им следствию присуща в той или иной степени случайность. Большинство экономических процессов представляют собой результат множества одновременно действующих причин. Каждый процесс при повторении его причинного комплекса за счет случайности реализуется с отклонением от закона, лежащего в его основе.

Различают два вида зависимости между экономическими явлениями: функциональную и статистическую. Зависимость между двумя величинами X и Y, отображающими соответственно два явления, называется ***функциональной***, если каждому значению величины X соответствует единственное значение величины Y, и наоборот. Примером функциональной связи в экономике может служить зависимость производительности труда от объема произведенной продукции и затрат рабочего времени. При этом следует отметить, что если X — детерминированная, не случайная величина, то и функционально зависящая от нее величина Y тоже является детерминированной. Если же X — величина случайная, то и Y также случайная величина.

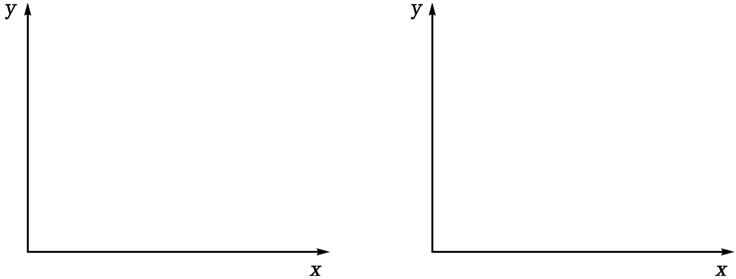
Однако гораздо чаще в экономике имеет место не функциональная, а ***статистическая*** зависимость, когда каждому фиксированному значению независимой переменой X соответствует не одно, а множество значений зависимой переменной Y, причем заранее нельзя сказать, какое именно значение примет Y. Это связано с тем, что на Y кроме переменной X влияют и многочисленные неконтролируемые случайные факторы. В этой ситуации Y является случайной величиной, а переменная X может быть как детерминированной, так и случайной величиной. Частным случаем статистической зависимости является ***корреляционная зависимость***, при которой функциональной зависимостью связаны фактор X и среднее значение (математическое ожидание) результативного показателя Y.

Статистическая зависимость может быть выявлена лишь по результатам достаточно большого числа наблюдений. Графически статистическая зависимость двух признаков может быть представлена с помощью ***поля корреляции***, при построении которого на оси абсцисс откладывается значение факторного признака X, а по оси ординат — результирующего Y.

В качестве примера на рис. 9.1 представлены данные, иллюстрирующие прямую зависимость между x и y (рис. 9.1, *а*) и обратную зависимость (рис. 9.1, *б*). В случае «*а*» это прямая зависимость между, например, среднедушевым доходом (x) и сбережением (y) в семье. В случае «*б*» речь идет об обратной зависимости. Такова, например, зависимость между производительностью труда (x) и себестоимостью единицы продукции (y). На рис. 9.1 каждая точка характеризует объект наблюдения со своими значениями x и y.

На рис. 9.1 также представлены прямые линии, линейные уравнения регрессии типа $\tilde{y} = \beta_0 \quad \beta_1 x_1$, характеризующие функциональную зависимость между независимой переменной x и средним значением результативного показателя y. Таким образом, по уравнению регрессии, зная x, можно восстановить лишь среднее значение y.

Ставя задачу статистического исследования зависимостей, важно хорошо представлять конечную прикладную цель построе-



а *б*

Рис. 9.1. Поле корреляции

ния моделей статистической зависимости между результативным показателем y с одной стороны и объясняющими переменными $x_1, x_2, ..., x_k$, с другой (до сих пор рассматривалась только одна объясняющая переменная x). Отметим две основных цели подобных исследований.

Первая из них состоит в установлении самого факта ***наличия*** (или ***отсутствия***) статистически значимой ***связи*** между Y и X. При такой постановке задачи статистический вывод имеет альтернативную природу — «связь есть» или «связи нет». Он обычно сопровождается лишь численной характеристикой — измерителем ***степени тесноты*** исследуемой зависимости. Задача оценки степени тесноты связи между показателями решается ***методами корреляционного анализа***. При этом выбор *формы связи* между результативным показателем y и объясняющими переменными $x_1, x_2, ..., x_k$, а также выбор состава последних играет вспомогательную роль, призванную максимизировать характеристику степени тесноты связи.

Вторая цель сводится ***к прогнозу, восстановлению неизвестных индивидуальных или средних значений*** результативного показателя «y» по заданным значениям объясняющих переменных.

Задача восстановления средних значений результативного показателя «y» по заданным значениям объясняющих переменных решается методами ***регрессионного анализа***. При этом выбор формы и вида зависимости «y» от объясняющих переменных $x_1, x_2, ..., x_k$ нацелен на минимализацию суммарной ошибки, т. е. отклонений наблюдаемых значений y от значений, полученных по регрессионной модели.

Таким образом, в задачах исследования зависимостей используются методы корреляционного и регрессионного анализов. При этом методы корреляционного анализа применяют на этапе предварительной обработки информации, результаты которого используют в регрессионном анализе при построении и анализе свойств уравнения регрессии. Выбор тех или иных методов анализа во многом определяется природой изучаемых переменных, шкалой в которой они измерены.

Количественные переменные позволяют измерять степень проявления изучаемого свойства объекта (денежный доход и сбережения семьи, объем валовой продукции, численность работников на предприятии и т. п.). ***Порядковые*** (***или ординальные***) ***переменные*** позволяют упорядочивать анализируемые объекты по степени проявления в них изучаемого свойства (уровень жилищных условий семьи, квалификационный разряд рабочего, уровень образования работника и т. п.). Наконец, ***классификационные*** (***или***

номинальные) ***переменные*** дают возможность разбивать обследованную совокупность объектов на не поддающиеся упорядочиванию однородные классы (профессия работника, мотив миграции семьи, отрасль промышленности и т. п.).

Теперь рассмотрим приемы и методы, позволяющие установить наличие связи между исследуемыми переменными, выявить структуру этих связей и измерить их тесноту. Поскольку перечисленные задачи решаются с помощью вычисления и анализа соответствующих корреляционных характеристик, совокупность используемых для этих целей методов называют корреляционным анализом.

Корреляционный анализ разработан К. Пирсоном и Дж. Юлом. Он призван прежде всего ответить на вопрос, как выбрать с учетом специфики и природы анализируемых переменных подходящий *измеритель статистической связи* (коэффициент корреляции, корреляционное отношение, ранговый коэффициент корреляции и т. д.). Далее предстоит решить задачу, как оценить его числовые значения по имеющимся выборочным данным. Корреляционный анализ позволяет найти методы проверки того, что полученное числовое значение анализируемого измерителя связи действительно свидетельствует о наличии статистической связи. Наконец, он помогает определить структуру связей между исследуемыми k признаками x_1, x_2, ..., x_k, сопоставив каждой паре признаков ответ («связь есть» или «связи нет»).

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

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

$$\rho(x,y)=\rho=\frac{M[(x-Mx)(y-My)]}{\sigma_x\sigma_y}, \qquad (9.1)$$

где Mx и My — математические ожидания величин x и y, а σ_x и σ_y — их среднеквадратические отклонения.

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

Содержательная интерпретация коэффициента корреляции приведена в табл. 9.1.

Коэффициент корреляции, определяемый (9.1), относится к генеральной совокупности и как всякий параметр генеральной совокупности нам не известен. Его можно лишь оценить по результатам выборочных наблюдений.

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

$$r_{xy} = r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}, \tag{9.2}$$

где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$,

а

$$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}; \quad s_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{9.3}$$

Значение $\rho(x, y)$	Связь	Интерпретация связи
$\rho = 0$	Отсутствует	Отсутствует линейная связь между величинами x и y
$0 < \rho < 1$	Прямая	С увеличением x величина y в среднем увеличивается, и наоборот
$-1 < \rho < 0$	Обратная	С увеличением x величина y в среднем уменьшается, и наоборот
$\rho = +1$ $\rho = -1$	Функциональная	Каждому значению x соответствует одно строго определенное значение величины y, и наоборот

Значение r	Связь
От 0 до $\pm 0{,}3$	Практически отсутствует
От $\pm 0{,}3$ до $\pm 0{,}5$	Слабая
От $\pm 0{,}5$ до $\pm 0{,}7$	Умеренная
От $\pm 0{,}7$ до ± 1	Сильная

Формула (9.2) симметрична, т. е. $r_{xy} = r_{yx} = r$. Если в ее числителе раскрыть скобки, то после несложных преобразований получим формулу, которую широко используют при вычислении коэффициента корреляции:

$$r = \frac{\overline{xy} - \overline{x}\,\overline{y}}{s_x s_y}, \tag{9.4}$$

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

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i. \tag{9.5}$$

Выборочный коэффициент корреляции r, как всякая выборочная характеристика, является случайной величиной, и по отдельным его значениям нельзя делать окончательные выводы о степени тесноты линейной связи между двумя величинами. Здесь речь может идти о некоторых практических, качественных рекомендациях (табл. 9.2) при достаточно больших n ($n > 40$).

В табл. 9.2 значения r рассматриваются по модулю, так как степень тесноты связи зависит от близости r к единице без учета знака. Степень зависимости между x и y существенно выше в случае, когда $r = -0{,}8$ по сравнению со случаем когда $r = 0{,}5$.

Пример 9.1. На основании выборочных данных (табл. 9.3) о деятельности $n = 6$ коммерческих фирм оценить тесноту связи между прибылью (млн руб.) (y) и затратами на 1 руб. произведенной продукции (x).

Используем формулу (9.4): $r = \dfrac{\overline{xy} - \overline{xy}}{s_x s_y}$. Прежде всего определим s_x и s_y:

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

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

Номер наблюдения (i)	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	96	0,22	21,12	9 216	0,049
2	78	1,07	83,46	6 084	1,145
3	77	1,00	77,00	5 929	1,000
4	89	0,61	54,29	7 921	0,372
5	81	0,78	63,18	6 561	0,608
6	82	0,79	64,78	6 724	0,624
Сумма	503	4,47	363,83	42 435	3,798
Средняя	83,833	0,745	60,638	7 072,5	0,633

Тогда

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

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

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

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

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{i1} & x_{i2} & x_{i3} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{pmatrix}.$$

В ней каждая i-я строка ($i = 1, 2, ..., n$) характеризует i-й экономический объект, а столбец, например первый, содержит значение для 1-го показателя для всех n объектов. По данным первого столбца матрицы X можно определить среднее значение $\bar{x}_1$ и выборочную дисперсию s_1^2 первого показателя.

$$\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{i1}; \quad s_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2.$$

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

Отсюда, согласно (9.4), рассчитаем выборочные парные коэффициенты корреляции r_{12}, r_{13}, r_{23}.

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

Выборочный частный коэффициент корреляции, как выборочный аналог $\rho_{12/3}$, определяется по формуле

$$r_{12/3} = r(x_1, x_2/x_3) = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}, \tag{9.6}$$

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

В трехмерной модели имеются еще два частных коэффициента корреляции $r_{13/2}$ и $r_{23/1}$, которые рассчитываются аналогично.

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

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

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

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

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

$$r_{1/23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}. \qquad (9.7)$$

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

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

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

Пример 9.2. Деятельность коммерческих фирм ($n = 6$) характеризуется тремя показателями: x_1 — прибыль (млн руб.), x_2 — затраты на 1 руб. произведенной продукции (коп./руб.) и x_3 — стоимость основных фондов (млн руб.). По данным табл. 9.4 требуется определить частный $r_{12/3}$ и множественный $r_{1/23}$ коэффициенты корреляции.

Воспользовавшись результатами решения примера 9.1, будем иметь: $s_1 = 0{,}279$; $s_2 = 6{,}673$ и $r_{12} = -0{,}976$.

Номер фирмы (i)	x_{i1} (млн руб.)	x_{i2} (коп./руб.)	x_{i3} (млн руб.)	x_{i3}^2	$x_{i1}x_{i2}$	$x_{i1}x_{i3}$	$x_{i2}x_{i3}$
1	0,22	96	4,3	18,49	21,12	0,946	412,8
2	1,07	78	5,9	34,81	83,46	6,313	460,2
3	1,00	77	5,9	34,81	77,00	5,900	454,3
4	0,61	89	3,9	15,21	54,29	2,379	347,1
5	0,78	81	4,9	24,01	63,18	3,822	396,9
6	0,79	82	4,3	18,49	64,78	3,397	352,6
Сумма	4,47	503	29,2	145,82	363,83	22,757	2 423,9
Средняя	0,745	83,833	4,867	24,303	60,638	3,793	403,983

Найдем $s_3 = \sqrt{\overline{x_3^2} - (\bar{x}_3)^2} = \sqrt{24{,}303 - (4{,}867)^2} = 0{,}784$. С учетом (9.4) определим $r_{13} = \dfrac{\overline{x_1 x_3} - \bar{x}_1 \bar{x}_3}{s_1 \cdot s_3} = \dfrac{3{,}793 - 0{,}745 \cdot 4{,}867}{0{,}279 \cdot 0{,}784} = 0{,}764$,

$$r_{23} = \frac{\overline{x_2 x_3} - \bar{x}_2 \bar{x}_3}{s_2 \cdot s_3} = \frac{403{,}983 - 83{,}833 \cdot 4{,}867}{6{,}673 \cdot 0{,}784} = -0{,}771.$$

Согласно (9.6) частный коэффициент корреляции равен:

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

Сравнивая значения парного $r_{12} = -0{,}976$ и частного $r_{12/3} = -0{,}948$ коэффициентов корреляции (они достаточно близки), можно утверждать, что x_3 слабо влияет на степень зависимости между величинами x_1 и x_2.

Определим теперь с учетом (9.7) множественный коэффициент корреляции:

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

и окончательно $r_{1/23} = 0{,}976$.

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

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

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

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

Для измерения степени тесноты связи ранжировками $x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}$ и $x_1^{(j)}, x_2^{(j)}, \ldots, x_n^{(j)}$. К. Спирмен в 1904 г. предложил показатель

$$\hat{\tau}_{kj}^{(s)} = 1 \quad \frac{6}{n^3 - n} \sum_{i=1}^{n} (x_i^{(k)} - x_i^{(j)})^2, \tag{9.8}$$

названный в последствии ***ранговым коэффициентом Спирмена***. Прямым подсчетом нетрудно убедиться, что для совпадающих ранжировок (т. е. при $x_i^{(k)} = x_i^{(j)}$ для всех $i = 1, 2, \ldots, n$) $\hat{\tau}_{kj}^{(s)} = 1$, а для противоположных (при $x_i^{(k)} = n \quad x_i^{(j)} + 1$, $i = 1, 2, \ldots, n$) $\hat{\tau}_{kj}^{(s)} = -1$. Можно показать, что во всех остальных случаях $|\hat{\tau}_{kj}^{(s)}| < 1$. Формула пригодна лишь в случае объединенных рангов в обеих исследуемых ранжировках.

В общем случае для каждой ранжировки по k-му признаку определяют величину

$$T^{(k)} = \frac{1}{2} \sum_{t=1}^{m^{(k)}} \left[(n_t^{(k)})^3 - n_t^{(k)} \right], \tag{9.9}$$

где $m^{(k)}$ — число групп неразличимых рангов у переменных x_k, а $n_t^{(k)}$ — число элементов (рангов), входящих в t-ю группу неразличных рангов (в случае отсутствия объединенных рангов $m^{(k)} = n$, а $n_i^{(k)} = \ldots = n_n^{(k)} = 1$ и $T^{(k)} = 0$).

Тогда ранговый коэффициент Спирмена определяется по формуле

$$r_{kj}^{(s)} = \frac{\frac{1}{6}(n^3 - n) - \sum_{i=1}^{n} \left(x_i^{(k)} - x_i^{(j)} \right)^2 - T^{(k)} - T^{(j)}}{\sqrt{\left[\frac{1}{6}(n^3 - n) - 2T^{(k)} \right] \left[\frac{1}{6}(n^3 - n) - 2T^{(j)} \right]}}. \tag{9.10}$$

Если $T^{(k)}$ и $T^{(j)}$ значительно меньше $\frac{1}{6}(n^3 - n)$, то можно воспользоваться приближенным соотношением:

$$r_{kj}^{(s)} = 1 \quad \frac{\sum_{i=1}^{n} \left(x_i^{(k)} - x_i^{(j)} \right)^2}{\frac{1}{6}(n^3 - n) - (T^{(k)} - T^{(j)})}. \tag{9.11}$$

Пример 9.3. Два эксперта проранжировали 10 предложенных проектов реорганизации НПО с точки зрения их эффективности. Ранжировка 1-ого эксперта: (1; 2; 3; 4; 5; 6; 7; 8; 9; 10). Ранжировка 2-го эксперта: (2; 3; 1; 4; 6; 5; 9; 7; 8; 10). Вычисления согласно (9.8) дают результат:

$$r_{12}^{(s)} = 1 \quad \frac{6}{1\,000-10}(1+1+2^2+0+1+1+2^2+1+1+0) = 1 \quad \frac{6}{990}\cdot 14 = 0{,}915,$$

что свидетельствует о положительной ранговой связи между переменными.

Пример 9.4. Десять предприятий подотрасли были проранжированы вначале по степени прогрессивности их оргструктур (признак x_1), а затем по эффективности их функционирования в отчетном году — x_2. В результате получены ранжировки: (1; 2,5; 2,5; 4,5; 4,5; 6,5; 6,5; 8; 9,5; 9,5) и (1; 2; 4,5; 4,5; 4,5; 4,5; 8; 8; 8; 10).

В 1-й ранжировке — четыре группы неразличимых рангов, а во 2-й — две такие группы.

Согласно (9.9) получаем:

$$T^{(1)} = \frac{1}{12}[(2^3 \quad 2)+(2^3-2)+(2^3-2)+(2^3-2)] = \frac{24}{12} \quad 2{,}00,$$

$$T^{(2)} = \frac{1}{12}[(4^3 \quad 4)+(3^3-3)] = 7{,}00.$$

Поскольку $T^{(1)}$ и $T^{(2)}$ значительно меньше $\frac{1}{6}(n^3-n) = \frac{1}{6}\cdot 900 = 165$, то воспользуемся формулой (9.11). Тогда:

$$r_{12}^{(s)} = 1 \quad \frac{(0+0{,}25+4+0+0+4+2{,}25+0+2{,}25+0{,}25)}{165-(2+7)} = 1 \quad \frac{13}{156} = 0{,}917.$$

Ранговый коэффициент корреляции Кендела определяется по формуле

$$r_{kj}^{(k)} = 1 \quad \frac{4I(x^{(k)}, x^{(j)})}{n(n-1)}, \tag{9.12}$$

где $I(x^{(k)}, x^{(j)})$ — число инверсий, равное числу нарушений ранжировки пар элементов последовательностей $x^{(k)}$ и $x^{(j)}$.

Число инверсий является естественной мерой нарушения порядка объектов в одной последовательности относительно другой. Число инверсий определяется по формуле

$$I(x^{(k)}, x^{(j)}) = \sum_{q=1}^{n-1}\sum_{l=q+1}^{n} v_{ql}^{(j,k)}, \tag{9.13}$$

$$
\text{где } v_{ql}^{(j,k)} = \begin{cases} 1, \text{ если } \tilde{x}_q^{(k)} \geq \tilde{x}_l^{(j)} \text{ (т.е. нарушен порядок} \\ \quad \text{последовательности } \tilde{x}^{(k)}); \\ 0, \text{ в противном случае.} \end{cases}
$$

Легко подсчитать, что число инверсий $I(x^{(k)}, x^{(j)})$ может меняться от 0 (при совпадающих ранжировках) до $\frac{1}{2}n(n-1)$ — в случае противоположных ранжировок.

Пример 9.5. Получены две ранжировки $n = 10$ объектов (см. данные задачи 9.3): (1; 2; 3; 4; 5; 6; 7; 8; 9; 10) и (2; 3; 1; 4; 6; 5; 9; 7; 8; 10). Требуется рассчитать ранговый коэффициент корреляции Кенделла.

Решение. Рассчитаем число инверсий:

$v_{12} = 0$, $v_{13} = 1$, $v_{14} = v_{15} = v_{16} = v_{17} = v_{18} = v_{19} = v_{10} = 0$;

$v_{23} = 1$, $v_{2j} = 0$ для $j = 4, 5, ..., 10$;

$v_{3j} = 0$ для $j = 4, 5, ..., 10$;

$v_{4j} = 0$ для $j = 5, 6, ..., 10$;

$v_{56} = 1$ и $v_{5j} = 0$ для $j = 7, 8, 9, 10$;

$v_{6j} = 0$ для $j = 7, 8, 9, 10$;

$v_{78} = 1$; $v_{79} = 0$; $v_{7,10} = 0$; $v_{89} = 1$ для $v_{8,10} = 1$; $v_{9,10} = 0$.

Таким образом, $I(x^{(k)}, x^{(j)}) = 1 + 1 + 1 + 1 + 1 + 0 = 5$ и коэффициент корреляции Кенделла равен $\hat{r}_{12}^{(k)} = 1 \quad \frac{4 \cdot 5}{10 \cdot 9} = 1 \quad 0{,}222 = 0{,}778$.

После того как с помощью корреляционного анализа выявлено наличие статистических связей между переменными и оценена степень их тесноты, обычно переходят к математическому описанию конкретного вида зависимостей с использованием регрессионного анализа. С этой целью подбирают класс функций, связывающий результативный показатель y и аргументы $x_1, x_2, ..., x_k$, отбирают наиболее информативные аргументы, вычисляют оценки неизвестных значений параметров уравнения связи и анализируют свойства полученного уравнения.

Функция $f(x_1, x_2, ..., x_k)$, описывающая зависимость среднего значения результативного признака y от заданных значений аргументов, называется ***функцией (уравнением) регрессии***. Термин «регрессия» (лат. — *regression* — отступление, возврат к чему-либо) введен английским психологом и антропологом Ф. Гальтоном и связан исключительно со спецификой одного из первых конкретных примеров, в котором это понятие было использовано. Так, обрабатывая статистические данные в связи с анализом наследственности роста, Ф. Гальтон нашел, что если отцы отклоняются от среднего роста всех отцов на x дюймов, то их сыновья отклоняются от среднего роста всех сыновей меньше, чем на x дюймов. Выявленная тенденция была названа ***регрессией к среднему состоянию***. С тех пор термин «регрессия» широко используется в статистической литературе, хотя во многих случаях он недостаточно точно характеризует понятие статистической зависимости.

Для точного описания уравнения регрессии необходимо знать закон распределения результативного показателя y. В статистической практике обычно приходится ограничиваться поиском подходящих аппроксимаций для неизвестной истинной функции регрессии $f(x)$, так как исследователь не располагает точным знанием условного закона распределения вероятностей анализируемого результирующего показателя y при заданных значениях аргумента x.

Рассмотрим взаимоотношение между истинной $f(x) = M(y/x)$, модельной регрессией $\tilde{y}$ и оценкой $\hat{y}$ регрессии. Пусть результативный показатель y связан с аргументом x соотношением:

$$y = 2x^{1,5} + \varepsilon,$$

где ε — случайная величина, имеющая нормальный закон распределения, причем $M\varepsilon = 0$ и $D\varepsilon = \sigma^2$. Истинная функция регрессии в этом случае имеет вид: $f(x) = M(y/x) = 2x^{1,5}$.

Предположим, что точный вид истинного уравнения регрессии нам не известен, но мы располагаем девятью наблюдениями над двумерной случайной величиной, связанной соотношением $y_i = 2x_i^{1,5} + \varepsilon_i$ и представленной на рис. 9.2.

Расположение точек на рис. 9.2 позволяет ограничиться классом линейных зависимостей вида: $\hat{y} = \beta_0 + \beta_1 x$. С помощью метода наименьших квадратов найдем оценку уравнения регрессии $\hat{y} = b_0 + b_1 x$. Для сравнения на рис. 9.2 приводятся графики истинной функции регрессии $f(x) = 2x^{1,5}$, теоретической аппроксимирующей функции регрессии $\hat{y} = \beta_0 + \beta_1 x$.

Поскольку мы ошиблись в выборе класса функции регрессии, а это достаточно часто встречается в практике статистических исследований, то наши статистические выводы и оценки окажутся ошибочными. И как бы мы ни увеличивали объем наблюдений, наша выборочная оценка $\hat{y}$ не будет близка к истинной функции регрессии $f(x)$. Если бы мы правильно выбрали класс функций ре-

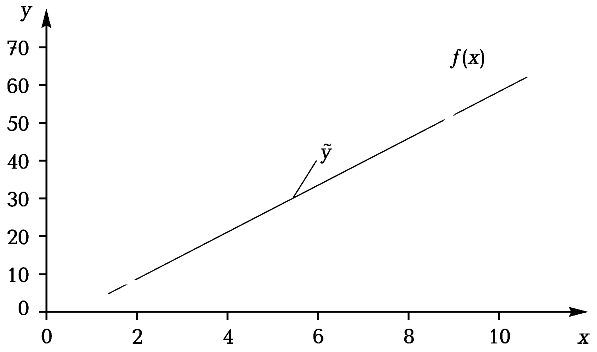


Рис. 9.2. Взаимное расположение истинной $f(x)$ и теоретической $\tilde{y}$ модели регрессии

грессии, то неточность в описании $f(x)$ с помощью $\hat{y}$ объяснялась бы только ограниченностью выборки.

С целью наилучшего восстановления по исходным статистическим данным условного значения результативного показателя $y(x)$ и неизвестной функции регрессии $f(x) = M(y/x)$ наиболее часто используют следующие критерии адекватности (функции потерь).

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

Получаемая регрессия называется ***среднеквадратической***.

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

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

Обычно предполагается, что случайная величина y имеет нормальный закон распределения с условным математическим ожиданием $\tilde{y}$, являющимся функцией от аргументов x_j $(j = 1, 2, ..., k)$ и постоянной, не зависящей от аргументов, дисперсией σ^2.

В общем виде ***линейная модель*** регрессионного анализа имеет вид

$$y=\sum_{j=0}^{k}\beta_j\varphi_j(x_1,x_2,\dots,x_k)+\varepsilon,$$

где φ_j — некоторая функция его переменных $x_1, x_2, \dots, x_k$; ε — случайная величина с нулевым математическим ожиданием и дисперсией σ^2.

В регрессионном анализе вид уравнения регрессии выбирают исходя из физической сущности изучаемого явления и результатов наблюдения.

Чаще других встречаются следующие виды уравнений регрессии:

линейное многомерное: $\tilde{y}=\beta_0+\beta_1x_1+\beta_2x_2+\dots+\beta_kx_k$;

полиномиальное: $\tilde{y}=\beta_0+\beta_1x+\beta_2x^2+\dots+\beta_kx^k$;

гиперболическое: $\tilde{y}=\beta_0+\beta_1\dfrac{1}{x}$;

степенное: $\tilde{y}=\beta_0'x_1^{\beta_1}x_2^{\beta_2}\dots x_k^{\beta_k}$.

Путем логарифмирования степенные уравнения регрессии могут быть преобразованы в линейные уравнения относительно параметров β_j.

Логарифмируя, получим

$$\ln\tilde{y}=\ln\beta_0'+\beta_1\ln x_1+\beta_2\ln x_2+\dots+\beta_k\ln x_k.$$

Пусть $\ln x_j=u_j$ для $j=1, 2, \dots, k$; $\ln\tilde{y}=\tilde{z}$ и $\ln\beta_0'=\beta_0$, тогда после подстановки будем иметь линейные уравнения регрессии:

$$\tilde{z}=\beta_0+\beta_1u_1+\beta_2u_2+\dots+\beta_ku_k.$$

Путем подстановок $\dfrac{1}{x}=u$ и $x^j=u_j$ гиперболическое и полиномиальное уравнения могут быть преобразованы в линейные, теория которых разработана наиболее полно.

Оценки неизвестных параметров уравнения регрессии находят обычно методом наименьших квадратов. Ниже остановимся более подробно на этой проблеме.

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

$$\tilde{y}=M(y/x)=\beta_0+\beta_1x, \qquad (9.14)$$

где $M(y/x)$ — условное математическое ожидание случайной величины y при заданном x; β_0 и β_1 — неизвестные параметры генеральной совокупности, которые надлежит оценить по результатам выборочных наблюдений.

Предположим, что для оценки параметров β_0 и β_1 из двухмерной генеральной совокупности (x, y) взята выборка объемом n, где (x_i, y_i) результат i-го наблюдения ($i = 1, 2, \ldots, n$). В этом случае модель регрессионного анализа имеет вид

$$y_i = \beta_0 + \beta_1 x + \varepsilon_i, \tag{9.15}$$

где ε_i — независимые нормально распределенные случайные величины с нулевым математическим ожиданием и дисперсией σ^2, т. е. $M\varepsilon_i = 0$; $D\varepsilon_i = \delta^2$; для всех $i = 1, 2, \ldots, n$.

Согласно методу наименьших квадратов в качестве оценок неизвестных параметров β_0 и β_1 следует брать такие значения выборочных характеристик b_0 и b_1, которые минимизируют сумму квадратов отклонений значений результативного признака y_i от условного математического ожидания $\tilde{y}_i$, т. е.

$$Q = \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_i x_i)^2 = \sum_{i=1}^{n}\varepsilon_i^2. \tag{9.16}$$

Так как Q дифференцируема по β_0 и β_1, то для отыскания минимума функции (9.16) найдем частные производные по β_0 и β_1:

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_0} = -2\sum\limits_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i); \\ \dfrac{\partial Q}{\partial \beta_1} = -2\sum\limits_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i. \end{cases} \tag{9.17}$$

Приравняв производные нулю и подставив в (9.17) вместо β_0 и β_1 их оценки b_0 и b_1, получим

$$\begin{cases} \sum\limits_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0; \\ \sum\limits_{i=1}^{n}(y_i - b_0 - b_1 x_i)x_i = 0 \end{cases} \quad \text{или} \quad \begin{cases} b_0 n + b_1\sum\limits_{i=1}^{n}x_i = \sum\limits_{i=1}^{n}y_i; \\ b_0\sum\limits_{i=1}^{n}x_i + b_1\sum\limits_{i=1}^{n}x_i^2 = \sum\limits_{i=1}^{n}x_i y_i. \end{cases} \tag{9.18}$$

В результате возникает *система нормальных уравнений*. Решая систему (9.18) относительно b_0 и b_1, получим

$$b_1 = \frac{\sum\limits_{i=1}^{n}x_i y_i - \frac{1}{n}\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{\sum\limits_{i=1}^{n}x_i^2 - \frac{1}{n}\left(\sum\limits_{i=1}^{n}x_i\right)^2}; \quad b_0 = \frac{1}{n}\sum_{i=1}^{n}y_i - b_1\frac{1}{n}\sum_{i=1}^{n}x_i = \bar{y} - b_1\bar{x}. \tag{9.19}$$

Таким образом, имеем оценку уравнения регрессии:

$$\hat{y} = b_0 \quad b_1 x. \tag{9.20}$$

О качестве полученной модели можно судить по остаточной дисперсии:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i \quad b_0 - b_1 x_i)^2. \tag{9.21}$$

Иными словами, исправленная выборочная дисперсия равна:

$$\hat{s}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2. \tag{9.22}$$

Теперь исследуем свойства b_0 и b_1. Среднее значение, математическое ожидание $\hat{s}^2$ равно σ^2, т. е. $M\hat{s}^2 = \sigma^2$.

Пример 9.6. По данным годовых отчетов десяти ($n = 10$) машиностроительных предприятий провести регрессионный анализ зависимости производительности труда y (тыс. руб. на чел.) от объема производства x (млн руб.). Предполагается, что уравнение регрессии линейно и имеет вид $\tilde{y} = \beta_0 \quad \beta_1 x$. Исходные данные для анализа представлены в табл. 9.5.

Номер предприятия (i)	y_i	x_i	x_i^2	$(\mathbf{x}_i - \bar{\mathbf{x}}_i)^2$	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$	$\bar{\delta}_i$
1	2,1	3	9	20,25	2,77	–0,67	–31,9
2	2,8	4	16	12,25	3,52	–0,72	–25,7
3	3,2	5	25	6,25	4,27	–1,07	–33,4
4	4,5	5	25	6,25	4,27	0,23	5,1
5	4,8	5	25	6,25	4,27	0,53	11,0
6	4,9	5	25	6,25	4,27	0,63	12,9
7	5,5	6	36	2,25	5,02	0,48	8,7
8	6,5	7	49	0,25	5,77	0,73	11,2
9	12,1	15	225	56,25	11,75	0,35	2,9
10	15,1	20	400	156,25	15,50	–0,4	–2,6
Сумма	61,5	75	835	272,5	—	—	—
Средняя	6,15	7,5	83,5	—	—	—	—

Решение. Согласно (9.19), учитывая, что $\sum_{i=1}^{10} x_i y_i = 666,5$, получим

$$b_1 = \frac{666,5 - \frac{1}{10}\cdot 75\cdot 61,5}{835 - \frac{1}{10}(75)^2} = \frac{205,25}{272,5} \quad 0,753; = b_0 = 615 - 0,753\cdot 7,5 = 0,502.$$

Таким образом, оценка уравнения регрессии будет иметь вид: $\hat{y} = b_0 + b_1 x$. После подстановки окончательно получим $\hat{y} = 0,502 + 0,753x$.

Из уравнения регрессии следует, что при увеличении объема производства на единицу его измерения производительность труда в среднем увеличивается на 0,753 тыс. руб.

Для интерпретации модели можно также воспользоваться коэффициентом эластичности, значение которого $e_1 = b_1\frac{\bar{x}}{\bar{y}} \quad 0,753\frac{7,5}{6,15} = = 0,918$ показывает, что при увеличении объема производства x на 1 % производительность труда y в среднем увеличится на 0,918 %.

Перейдем к статистическому анализу полученного уравнения регрессии и рассчитаем исправленную выборочную дисперсию $\hat{s}^2$, абсолютные $e_i = y_i \quad \hat{y}_i$ и относительные $\delta_i = \frac{e_i}{y_i}100\,\%$ ошибки аппроксимации. Выборочная дисперсия равна: $\hat{s}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 0,486$.

Теперь среднюю относительную ошибку аппроксимации определим по формуле

$$\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}|\delta_i| \quad \frac{1}{10}145,4 = 14,54\,\%,$$

где $|\delta_i|$ — абсолютное значение относительной ошибки аппроксимации. Среднее значение относительной ошибки 14,54 % говорит о том, что наша модель достаточно хорошо согласуется с исходными данными.

Самую низкую эффективность по производительности труда, как следует из табл. 9.5, имеет третье предприятие. У этого предприятия производительность труда $y_3 = 3,2$ тыс. руб. на человека, что на 33,4 % ниже того, что имело бы «среднее» предприятие с объемом производства $x_3 = 5,0$ млн руб.

Лучшим по критерию производительности труда является шестое предприятие, у которого этот показатель на 12,9 % выше среднего значения по совокупности рассматриваемых предприятий при $x_5 = 5$.

1. В каких пределах изменяется парный коэффициент корреляции:
   а) $0 \le r_{xy} \le 1$;
   б) $-1 \le r_{xy} \le 1$;
   в) $-\infty < r_{xy} < +\infty$;
   г) $0 \le r_{xy} < \infty$?

2. В каких пределах изменяется множественный коэффициент корреляции:
   а) $0 \le r_{y/xz} \le 1$;
   б) $-1 \le r_{y/xz} \le 1$;
   в) $-\infty \le r_{y/xz} < +\infty$;
   г) $0 < r_{y/xz} < \infty$?

3. Если парный коэффициент корреляции по модулю больше модуля соответствующего частного (например, $|r_{xy}| > |r_{xy/z}|$) и коэффициенты не имеют разных знаков, то это означает, что:
   а) фиксируемая переменная z ослабляет корреляционную связь;
   б) фиксируемая переменная усиливает связь между x и y;
   в) фиксируемая переменная не связана с факторами x и y;
   г) возможен любой из первых трех исходов.

4. Коэффициент детерминации между x и y характеризует:
   а) долю дисперсии y, обусловленную влиянием не входящих в модель факторов;
   б) долю дисперсии y, обусловленную влиянием x;
   в) долю дисперсии x, обусловленную влиянием не входящих в модель факторов;
   г) направление зависимости между x и y.

5. Парный коэффициент корреляции между факторами равен 1. Это означает:
   а) наличие нелинейной функциональной связи;
   б) отсутствие связи;
   в) наличие функциональной связи;
   г) отрицательную линейную связь.

6. На основании 20 наблюдений выяснено, что выборочная доля дисперсии случайной величины y, вызванной вариацией x, составит 64 %. Чему равен выборочный парный коэффициент корреляции:
   а) 0,64;
   б) 0,36;
   в) 0,8;
   г) 0,8 или –0,8?

7. Уравнение регрессии имеет вид $\tilde{y} = 5{,}1 \quad 1{,}7x$. На сколько единиц своего измерения в среднем изменится y при увеличении x на 1 единицу своего измерения:
   а) увеличится на 1,7;
   б) не изменится;
   в) уменьшится на 1,7;
   г) увеличится на 3,4.

8. Согласно методу наименьших квадратов в качестве оценок параметров β_0 и β_1 следует использовать такие значения b_0 и b_1, которые минимизируют сумму квадратов отклонений:

а) фактических значений зависимой переменной от ее среднего значения;
б) фактических значений объясняемой переменной от ее среднего значения;
в) расчетных значений зависимой переменной от ее среднего значения;
г) фактических значений зависимой переменной от ее расчетных значений.

9. Какой коэффициент указывает в среднем процент изменения результативного показателя y при увеличении аргумента x на 1 %:
а) бета-коэффициент;
б) коэффициент эластичности;
в) коэффициент детерминации;
г) коэффициент регрессии?

10. Какая из следующих формул минимизируется в методе наименьших квадратов:
а) $\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$;
б) $\sum_{i=1}^{n}(y_i - \bar{y})^2$;
в) $\sum_{i=1}^{n}|y_i - \bar{y}|$;
г) $\sum_{i=1}^{n}(y_i - \tilde{y}_i)$?

11. Темпы прироста (%) следующих макроэкономических показателей $n = 10$ развитых стран мира за 1992 г.: x_1 — внутренний национальный продукт (ВНП); x_2 — промышленное производство; x_3 — индекс цен, представлены в таблице:

№ п/п (i)	Страны	x_{i1}	x_{i2}	x_{i3}
1	Япония	3,5	4,3	2,1
2	США	3,1	4,6	3,9
3	Германия	2,2	2,0	3,4
4	Франция	2,7	3,1	2,9
5	Италия	2,7	3,0	5,6
6	Великобритания	1,6	1,4	4,0
7	Канада	3,1	3,4	3,0
8	Австралия	1,8	2,6	4,0
9	Бельгия	2,3	2,6	3,4
10	Нидерланды	2,3	2,4	3,5

Требуется найти выборочные коэффициенты корреляции между темпами прироста: а) ВНП (x_1) и промышленного производства (x_2); б) ВНП (x_1) и индекса цен (x_3); в) между темпами прироста промышленного производства (x_2) и индекса цен (x_3); г) выборочный частный коэффициент корреляции $r_{12/3}$ и, сравнив его с парным коэффициентом корреляции r_{12}, сделать вывод о влиянии x_3 на степень зависимости между x_1 и x_2; д) выборочный частный коэффициент корреляции $r_{13/2}$ и, сравнив его с парным коэффициентом корреляции r_{13}, сделать вывод о влиянии x_2 на степень зависимости между x_1 и x_3; е) выборочный множественный коэффициент корреляции $r_{1/23}$ и определить, на сколько процентов вариация x_1 обусловлена влиянием x_2 и x_3; ж) выборочный множественный коэффициент детерминации $r^2_{2/13}$ и определить, на сколько процентов вариация x_2 обусловлена влиянием x_1 и x_3.

12. При исследовании взаимосвязи цен на следующие виды продовольственных товаров: растительное масло (x_1), сахар-песок (x_2) и хлеб белый (x_3) в $n = 22$ городах Центрального района России получены выборочные парные коэффициенты корреляции: $r_{12} = 0{,}82$; $r_{13} = 0{,}24$; $r_{23} = -0{,}05$.
Требуется определить: а) выборочный частный коэффициент корреляции $r_{12/3}$ и, сравнив его с парным коэффициентом корреляции r_{12}, сделать вывод о влиянии x_3 на степень зависимости между x_1 и x_2; б) выборочный частный коэффициент корреляции $r_{13/2}$; в) выборочный множественный коэффициент корреляции $r_{1/23}$; г) выборочный множественный коэффициент детерминации $r^2_{2/13}$ и определить, на сколько процентов вариация цен на сахар-песок зависит от изменения цен на растительное масло и хлеб белый высшего сорта.

13. С целью исследования зависимости усушки формового хлеба (y) от продолжительности хранения (x) было проведено $n = 5$ наблюдений:

Продолжительность хранения (г) (x)	1	3	6	8	10
Усушка, % к массе горячего хлеба	1,6	2,4	2,8	3,2	3,3

В предположении о линейной зависимости y от x требуется: а) вычислить оценки b_0 и b_1 параметров уравнения регрессии $\hat{y} = b_0 \quad b_1 x$; б) построить на графике поле корреляции и уравнение регрессии.

14. Данные о темпах прироста курса акций строительной компании (y) за пять месяцев приведены в таблице:

Месяцы (x)	0	1	2	3	4
Темп прироста курса акций (y), %	10	8	5	3	4

На их основании в предположении линейной зависимости y от x требуется: а) вычислить оценки b_0 и b_1 параметров уравнения регрессии $\hat{y} = b_0 \quad b_1 x$; б) графически представить пояс корреляции и уравнение регрессии.

15. Данные корпорации о темпах прироста объема инвестиций (x) и объема выпуска продукции (y) представлены в таблице:

Темп прироста инвестиций (x), %	2	4	6	9	11
Темп прироста выпуска продукции (y), %	1	2	2	5	6

В предположении, что уравнение регрессии имеет вид $\hat{y} = b_0 \quad b_1 x^2$, требуется: а) определить оценки параметров уравнения регрессии b_0, b_1; б) графически представить поле корреляции и уравнение регрессии.

Среди основных задач статистики видное место занимает описание изменений показателей во времени, изучение динамики развития социально-экономических процессов. Как изменяется уровень оплаты труда? Каковы колебания курса доллара? Какая тенденция прослеживается в изменении важнейших макроэкономических показателей? Ответы на эти и аналогичные вопросы могут быть получены с помощью специальных статистических методов, анализирующих ряды динамики.

Рядом динамики (динамическим рядом, временны́м рядом) называется последовательность значений статистического показателя (признака), упорядоченная в хронологическом порядке, т. е. в порядке возрастания временно́го параметра. Отдельные наблюдения временно́го ряда называются уровнями этого ряда. В англоязычной литературе для временны́х рядов используется термин «*time series*».

Каждый ряд динамики содержит ***значения времени*** и соответствующие им ***значения уровней ряда***. В качестве показателя времени в рядах динамики могут указываться либо определенные моменты (даты) времени, либо отдельные периоды (сутки, месяцы, кварталы, полугодия, годы и т. д.). В зависимости от характера временно́го параметра ряды делятся на моментные и интервальные.

В ***моментных рядах*** динамики уровни характеризуют значения показателя по состоянию на определенные моменты времени. Например, моментными являются временны́е ряды цен на определенные виды товаров, ряды курсов акций, уровни которых фиксируются для конкретных чисел. Примерами моментных рядов динамики могут служить также ряды численности населения или стоимости основных фондов, так как значения уровней этих рядов определяются ежегодно на одно и то же число.

В *интервальных* рядах уровни характеризуют значение показателя за определенные интервалы (периоды) времени. Примерами могут служить ряды годовой (месячной, квартальной) динамики производства продукции в натуральном или стоимостном выражении.

В табл. 10.1 и 10.2 приведены моментные ряды динамики, а в табл. 10.3 и 10.4 — интервальные, причем у последнего ряда динамики в отличие от других примеров уровни не считаются равноотстоящими во времени.

Показатель	2007	2008	2009	2010	2011
Число кредитных организаций, зарегистрированных Банком России	1 345	1 296	1 228	1 178	1 146

* Россия в цифрах. 2011: крат. стат. сб. / Росстат. — М., 2011.

Показатель	2006	2007	2008	2009	2010
Численность безработных	1 742,0	1 553,0	1 521,8	2 147,3	1 589,9

* Россия в цифрах. 2011: крат. стат. сб. / Росстат. — М., 2011.

Показатель	Январь	Февраль	Март	Апрель	Май	Июнь
Фонд заработной платы работников предприятия, тыс. руб.	990,0	984,1	985,5	988,5	989,9	979,5

Показатель	2000	2005	2008	2009	2010
Объем экспорта, тыс. долл. США	1 200	1 350	1 400	1 370	1 325

Месяц	Производство продукции (тыс. шт.)	Количество рабочих дней в месяце	Среднесуточное производство (тыс. шт.)
1	2	3	4
Январь	195,0	25	7,8
Февраль	204,0	24	8,5
Март	210,6	26	8,1
Апрель	195,0	26	7,5
Май	207,5	25	8,3
Июнь	205,4	26	7,9

Уровни рядов динамики могут представлять собой абсолютные, относительные и средние величины. Если уровни ряда представляют собой не непосредственно наблюдаемые значения, а производные величины: средние или относительные, то такие ряды называются ***производными***. Уровни этих рядов получаются с помощью некоторых вычислений на основе абсолютных показателей. Примером производного ряда динамики может служить ряд среднесуточного производства промышленной продукции (табл. 10.5).

Данные графы 4 (см. табл. 10.5) получены путем деления данных графы 2 на данные графы 3.

Важной особенностью интервальных рядов динамики абсолютных величин является возможность суммирования их уровней. В результате этой процедуры получаются накопленные итоги, имеющие осмысленное содержание благодаря отсутствию повторного счета. Например, суммируя фонд заработной платы работников предприятия за первые три месяца и три последующих месяца (см. табл. 10.3), получаем, соответственно, фонд заработной платы за первый и второй кварталы, а сумма этих квартальных данных дает фонд заработной платы за полугодие.

Суммирование уровней моментного ряда динамики не практикуется, так как накопленные итоги лишены всякого смысла. Например, уровни моментного ряда «Число кредитных организаций, зарегистрированных Банком России» (см. табл. 10.1), содержат элементы повторного счета. Второй уровень содержит кредитные организации, учтенные первым уровнем, и т. д. Таким образом, моментные ряды динамики, в отличие от интервальных, не обладают свойством аддитивности (от английского глагола *to add* — добавлять).

При исследовании моментного ряда динамики определенный смысл имеет расчет разностей уровней, характеризующих изменение показателя за некоторый отрезок времени. Например, за 2010 г. число кредитных организаций, зарегистрированных Банком России, сократилось на 32 ед. (см. табл. 10.1).

На практике часто требуется проанализировать динамику показателя не только за данный отрезок времени, но и с учетом ряда предшествующих периодов. Для этого строится ***ряд динамики с нарастающими итогами***, уровни которого дают обобщающий результат развития показателя с начала отчетного периода (квартала, полугодия, года и т.д.). В качестве примера рассмотрим данные о производстве холодильников и морозильников на предприятии (табл. 10.6). Данные графы 3 получены последовательным суммированием смежных уровней.

Уровни ряда могут принимать детерминированные или случайные значения. Примером ряда с детерминированными значениями уровней служит ряд последовательных данных о количестве дней в месяцах. Естественно, анализу, а в дальнейшем и прогнозированию, подвергаются ряды со случайными значениями уровней.

Успешность статистического анализа развития процессов во времени во многом зависит от правильного построения рядов динамики. Так, большое значение имеет выбор интервалов между соседними уровнями ряда. Удобнее всего иметь дело с равноотстоящими друг от друга уровнями. При этом нежелательно брать слишком большой интервал времени, ибо можно упустить существенные закономерности в динамике показателя. Например, по квартальным данным невозможно судить о месячных сезонных колебаниях. Но инфор-

Месяц	Произведено холодильников и морозильников (шт.)	
	за месяц	с начала года
1	2	3
Январь	146	146
Февраль	181	146 + 181 = 327
Март	174	327 + 174 = 501
Апрель	152	501 + 152 = 653
Май	160	653 + 160 = 813
Июнь	179	813 + 179 = 992

* Цифры условные.

мация может оказаться и слишком «короткой» для использования некоторых методов анализа и прогнозирования динамики, предъявляющих «жесткие» требования к длине рядов. К тому же слишком малые интервалы между наблюдениями увеличивают объем вычислений, а также могут приводить к появлению ненужных деталей в динамике процесса, засоряющих общую тенденцию. Разумеется, вопрос о выборе интервала времени между уровнями ряда должен решаться исходя из целей каждого конкретного исследования.

Одним из важнейших условий, необходимым для правильного отражения временны́м рядом реального процесса развития, является ***сопоставимость уровней ряда***. Для несопоставимых величин неправомерно проводить исследование динамики. Между тем появление несопоставимых уровней может быть обусловлено разными причинами: изменением методики расчета показателя, изменением классификации, терминологии и т. д. Например, уровни временно́го ряда, характеризующие количество малых предприятий, могут оказаться несопоставимыми из-за изменения самого понятия «малое предприятие». Подразумевается, что это понятие должно быть одинаковым для всего исследуемого периода.

Чаще всего несопоставимость встречается в стоимостных показателях. Это вызвано изменением цен в разные периоды времени. Поэтому на практике осуществляют пересчет уровней в сопоставимые цены (цены одного периода). Несопоставимость может возникнуть и вследствие территориальных модификаций (изменение границ области, района, страны). Следует также иметь в виду, что вопрос о сопоставимости может зависеть от целей исследования. Например, при описании военной, экономической мощи страны следует учитывать данные в изменяющихся границах территории, а при сопоставлении темпов развития промышленности следует производить сравнение в рамках одной и той же территории. Несопоставимость может возникнуть и в результате структурных изменений. Например, произошло укрупнение нескольких ведомств путем слияния их в единое целое или укрупнение производства за счет слияния нескольких предприятий в одно объединение.

В большинстве случаев удается устранить несопоставимость, вызванную указанными причинами, путем пересчета более ранних значений показателей. Правда, проведение такой обработки не всегда обеспечивает требуемую точность, что приводит к снижению ценности исходной информации, а, следовательно, и к затруднению дальнейшего анализа.

Для успешного изучения динамики процесса важно, чтобы информация была полной, чтобы временной ряд имел достаточную длину (с учетом конкретных целей исследования). Так, при изучении периодических колебаний желательно иметь информацию не менее чем за три полных периода колебания. Поэтому при анализе сезонных колебаний на базе рядов месячной или квартальной динамики желательно иметь информацию, как правило, не менее чем за три года. Заметим, что применение определенного математического аппарата влияет на допустимую длину временны́х рядов. Например, для использования регрессионного анализа требуется иметь временны́е ряды, длина которых в несколько раз превосходит количество независимых переменных.

Уровни рядов динамики могут содержать аномальные значения или «выбросы». Появление последних нередко объясняется ошибками при сборе, записи и передаче информации. Возможными источниками появления ошибочных значений могут стать сдвиг запятой при перенесении информации из документа, занесение данных в другую графу и т. д.

Выявление и исключение таких значений, замена их истинными или расчетными является необходимым этапом первичной обработки данных, поскольку применение математических методов к «засоренной» информации приводит к искажению результатов анализа. Однако аномальные значения могут отражать и реальное развитие процесса. Таков, например, был «скачок» курса доллара в «черный вторник» 1998 г. Как правило, эти значения также заменяются расчетными при построении моделей, но учитываются при расчете возможной величины отклонений фактических значений от полученных по модели.

Соответствие исходной информации всем указанным требованиям проверяется на этапе предварительного анализа временны́х рядов. Лишь после этого переходят к расчету и анализу основных показателей динамики развития, построению моделей прогнозирования, получению прогнозных оценок.

При анализе изменений явления во времени на практике часто определяют средние показатели, в том числе средний уровень ряда. ***Средний уровень*** является важной обобщающей характеристикой для рядов динамики, изменение которых стабилизировалось в исследуемом периоде и при этом подвержено ощутимым случайным колебаниям. Например, средний уровень урожайности за ряд лет лучше опишет урожайность, чем уровень одного года, значение которого формируется под действием множества случайных факторов. Если же в исследуемом периоде приходится выделять неоднородные этапы, в течение которых условия развития существенно менялись, то нецелесообразно рассчитывать общую среднюю, следует построить анализ динамики по отдельным этапам.

Средний уровень ряда определяется по-разному для моментных и интервальных рядов. При этом ряд динамики может содержать как равноотстоящие, так и не равноотстоящие во времени уровни.

Для интервальных рядов динамики с равноотстоящими во времени уровнями расчет среднего уровня производится по формуле простой средней арифметической:

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}, \tag{10.1}$$

где n — число уровней или длина ряда; y_t — текущий уровень ряда динамики ($t = 1, 2, ..., n$).

В качестве примера рассмотрим определение среднего уровня для интервального ряда динамики, представленного в табл. 10.3. Среднемесячный фонд заработной платы в первом полугодии составил:

$$\bar{y} = \frac{990{,}0 + 984{,}1 + 985{,}5 + 988{,}5 + 989{,}9 + 979{,}5}{6} = 986{,}25 \text{ (тыс. руб.)}.$$

В случае интервальных рядов динамики с не равноотстоящими во времени уровнями для расчета среднего уровня используется формула взвешенной средней арифметической, где в качестве весовых коэффициентов используется продолжительность интервалов времени между уровнями (число периодов времени, при которых значение уровня не изменяется). Для моментных рядов динамики с равноотстоящими во времени уровнями средний уровень (так называемая средняя хронологическая) находится по формуле

$$\bar{y} = \frac{\frac{y_1 + y_n}{2} + \sum_{t=2}^{n-1} y_t}{n - 1}, \tag{10.2}$$

где $y_1, y_2, ..., y_n$ — уровни ряда динамики; y_1 и y_n — соответственно начальный и конечный уровни ряда; n — число уровней или длина ряда.

Проиллюстрируем использование данной формулы на примере. Требуется рассчитать среднегодовой запас сырья на складе предприятия на основе данных о запасах сырья, представленных на начало кварталов (тыс. руб.): на 01.01.2010 — 386; на 01.04.2010 — 392; на 01.07.2010 — 390; на 01.10.2010 — 395; на 01.01.2011 — 384. Согласно (10.2):

$$\bar{y} = \frac{\frac{386 + 384}{2} + 392 + 390 + 395}{4} = 390{,}5 \text{ (тыс. руб.).}$$

В случае моментных рядов динамики с не равноотстоящими во времени уровнями средний уровень определяется по формуле средней хронологической взвешенной:

$$\bar{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \ldots + (y_{n-1} + y_n)t_{n-1}}{2\sum_{i=1}^{n-1} t_i}, \qquad (10.3)$$

где $y_1, y_2, \ldots, y_n$ — уровни ряда динамики; t_i — продолжительность интервала времени между соседними уровнями.

Рассмотрим применение этой формулы на следующем примере.

Требуется определить средний уровень для моментного ряда динамики, характеризующего численность официально зарегистрированных безработных в городе (табл. 10.7).

Ряд динамики имеет не равноотстоящие во времени уровни, поэтому применим формулу (10.3). Средний уровень ряда равен:

$$\bar{y} = \frac{(1\,800 + 1\,950)2 + (1\,950 + 1\,899)3 + (1\,899 + 1\,600)3 + (1\,600 + 1\,850)4}{2(2 + 3 + 3 + 4)} =$$

$$= 1\,806 \text{ (чел.).}$$

На практике для количественной оценки динамики явлений широко применяется ряд основных аналитических показателей. Таковы абсолютные приросты, темпы роста, темпы прироста. Каждый

Показатель	Дата				
	1.01.2010	1.03.2010	1.06.2010	1.09.2010	1.01.2011
Численность безработных (чел.)	1 800	1 950	1 899	1 600	1 850

* Цифры условные.

из указанных показателей бывает трех видов: цепной, базисный, средний.

В основе расчета этих показателей динамики лежит сравнение уровней временнóго ряда. Если сравнение осуществляется с одним и тем же уровнем, принятым за базу сравнения, то эти показатели называются ***базисными***. В качестве базы сравнения выбирается либо начальный уровень динамического ряда, либо уровень, с которого начинается новый этап развития. Например, при анализе динамики производства основных видов промышленной продукции в натуральном выражении часто за базу сравнения выбирали 1990 г. Это объяснялось тем, что до этого года во многих отраслях российской промышленности наблюдался замедлявшийся подъем, перешедший затем в спад производства. Поэтому последующий рост производства желательно было оценивать не только по отношению к предыдущему году, но и в сравнении с 1990 г. Если сравнение осуществляется при переменной базе и каждый последующий уровень сравнивается с предыдущим, то вычисленные таким образом показатели называются ***цепными***.

Абсолютный прирост равен разности двух сравниваемых уровней и характеризует величину изменения показателя за определенный промежуток времени. В общем случае абсолютный прирост может быть представлен в виде:

$$\Delta y_t = y_t \quad y_{t-k}, \tag{10.4}$$

где y_t — текущий уровень ряда динамики; $t = 2, 3, ..., n$; $k = 1, 2, ..., n - 1$.

При $k = 1$ от текущего уровня y_t вычитается предыдущий уровень y_{t-1} и получается формула для расчета цепного абсолютного прироста:

$$\Delta y_t = y_t \quad y_{t-1}. \tag{10.5}$$

При $k = t - 1$ из формулы (10.4) вытекает выражение для базисного абсолютного прироста, определяемого относительно начального уровня ряда:

$$\Delta y_t^{б} = y_t \quad y_1. \tag{10.6}$$

Для записи формулы базисного абсолютного прироста в более общем виде уровень y_1 в (10.6) может быть заменен на уровень временнóго ряда, принятый за базу сравнения — $y_б$:

$$\Delta y_t^{б} = y_t \quad y_б. \tag{10.7}$$

Средний абсолютный прирост является обобщающей характеристикой скорости изменения исследуемого показателя во времени (скоростью будем называть прирост в единицу времени). Для его определения за весь период наблюдения используется формула простой средней арифметической:

$$\overline{\Delta y} = \frac{\sum_{t=2}^{n} \Delta y_t}{n-1}, \qquad (10.8)$$

где Δy_t — цепной абсолютный прирост; n — длина временно́го ряда.

Подставив в числитель выражения для цепных абсолютных приростов, получим более удобную форму записи для среднего абсолютного прироста:

$$\overline{\Delta y} = \frac{y_2 - y_1 + y_3 - y_2 + \ldots + y_n - y_{n-1}}{n-1} = \frac{y_n - y_1}{n-1}, \qquad (10.9)$$

где y_n и y_1 — соответственно конечный и начальный уровни ряда динамики.

Темп роста T характеризует отношение двух сравниваемых уровней ряда, как правило, выраженное в процентах. Темп роста может быть представлен в виде:

$$T_t = \frac{y_t}{y_{t-k}} 100\ \%, \qquad (10.10)$$

где y_t — текущий уровень ряда динамики; $t = 2, 3, \ldots, n$; $k = 1, 2, \ldots, n - 1$.

Отметим, что индекс уровня y_{t-k}, находящегося в знаменателе, определяется так же, как и в случае абсолютного прироста. Следовательно, из выражения (10.10) в зависимости от значений индекса k получаются формулы для расчета цепных и базисных темпов роста.

Цепной темп роста равен:

$$T_t = \frac{y_t}{y_{t-1}} 100\ \%. \qquad (10.11)$$

Базисный темп роста может быть представлен в виде:

$$T_t = \frac{y_t}{y_б} 100\ \%, \qquad (10.12)$$

где $y_б$ — уровень временно́го ряда, принятый за базу сравнения.

Темп роста всегда положителен. Если темп роста равен 100 %, то значение уровня не изменилось, если меньше 100 %, то значение уровня понизилось, больше 100 % — повысилось.

Средний темп роста является обобщающей характеристикой динамики и отражает интенсивность изменения уровней ряда. Он показывает, сколько в среднем процентов последующий уровень составляет от предыдущего в течение всего периода наблюдения. Этот показатель рассчитывается по формуле средней геометрической из цепных темпов роста:

$$\bar{T} = \sqrt[n-1]{T_2 T_3 \ldots T_n}. \qquad (10.13)$$

Выразив цепные темпы роста T_2, T_3, ..., T_n через соответствующие уровни ряда, получим

$$\bar{T} = \sqrt[n-1]{\frac{y_2}{y_1}\frac{y_3}{y_2}\frac{y_4}{y_3}\ldots\frac{y_n}{y_{n-1}}}100\ \% = \sqrt[n-1]{\frac{y_n}{y_1}}100\ \%. \qquad (10.14)$$

Темп прироста характеризует абсолютный прирост в относительных величинах. Определенный в процентах темп прироста показывает, на сколько процентов изменился сравниваемый уровень по отношению к уровню, принятому за базу сравнения. Темп прироста есть выраженное в процентах отношение абсолютного прироста к уровню, принятому за базу сравнения:

$$K_t = \frac{y_t - y_{t-k}}{y_{t-k}}100\ \%, \qquad (10.15)$$

где y_t — текущий уровень ряда динамики; $t = 2, 3, ..., n$; $k = 1, 2, ..., n - 1$.

Очевидно, что темп прироста может быть положительным, отрицательным или равным нулю.

При $k = 1$ получаем цепной темп прироста:

$$K_t = \frac{y_t - y_{t-1}}{y_{t-1}}100\ \%. \qquad (10.16)$$

Преобразовав выражение (10.16), можно показать зависимость цепного темпа прироста от соответствующего темпа роста:

$$K_t = \frac{y_t}{y_{t-1}}100\ \% \quad 100\ \% = T_t \quad 100\ \%, \qquad (10.17)$$

где T_t — цепной темп роста.

Базисный темп прироста равен отношению базисного абсолютного прироста к уровню ряда, принятому за базу сравнения:

$$K_t^{б} = \frac{\Delta y_t^{б}}{y_{б}}. \qquad (10.18)$$

По аналогии с (10.17) получаем

$$K_t^{б} = T_t^{б} \quad 100\ \%, \qquad (10.19)$$

где $T_t^{б}$ — базисный темп роста.

Тип показателя	Абсолютный прирост	Темп роста	Темп прироста
Цепной	$\Delta y_t = y_t \quad y_{t-1}$	$T_t = \dfrac{y_t}{y_{t-1}}\ 100$, %	$K_t = T_t \quad 100$, %
Базисный	$\Delta y_t^{б} = y_t \quad y_{б}$	$T_t^{б} = \dfrac{y_t}{y_{б}}\ 100$, %	$K_t^{б} = T_t^{б} \quad 100$, %
Средний	$\overline{\Delta y} = \dfrac{y_n - y_1}{n-1}$	$\overline{T} = \sqrt[n-1]{\dfrac{y_n}{y_1}}\ 100$, %	$\overline{K} = \overline{T} \quad 100$, %

Соответственно ***средний темп прироста*** может быть выражен через средний темп роста:

$$\overline{K} = \overline{T} \quad 100\,\%. \tag{10.20}$$

Сравнение абсолютного прироста и темпа прироста за одни и те же периоды времени показывает, что в реальных экономических процессах замедление темпов прироста не всегда сопровождается уменьшением абсолютных приростов. Поэтому на практике часто проводят сопоставление этих показателей. Для этого рассчитывают *абсолютное значение одного процента прироста*, определяемое как отношение абсолютного прироста к соответствующему темпу прироста:

$$A_t = \frac{y_t - y_{t-1}}{\dfrac{y_t - y_{t-1}}{y_{t-1}}100} = 0{,}01 y_{t-1}. \tag{10.21}$$

В табл. 10.8 приведены выражения для вычисления рассмотренных аналитических показателей динамики.

Проиллюстрируем расчет и анализ статистических показателей динамики на следующем примере.

Пример 10.1. По данным о вводе в действие жилых домов (млн м^2) рассчитать цепные, базисные и средние: а) абсолютные приросты; б) темпы роста; в) темпы прироста (табл. 10.9).

В качестве базисного уровня взять начальный уровень ряда.

Текущий номер года t	1	2	3	4	5
Общая площадь, млн м^2	7,0	6,5	5,9	5,5	4,9

* Цифры условные.

t	y_t (млн м²)	Абсолютный прирост, млн м²		Темп роста, %		Темп прироста, %	
		Цепной	Базисный	Цепной	Базисный	Цепной	Базисный
1	7,0	—	—	—	—	—	—
2	6,5	$6{,}5 - 7{,}0 = -0{,}5$	$6{,}5 - 7{,}0 = -0{,}5$	$\frac{6{,}5}{7{,}0}100 = 92{,}86$	$\frac{6{,}5}{7{,}0}100 = 92{,}86$	$92{,}86 - 100 = -7{,}14$	$92{,}86 - 100 = -7{,}14$
3	5,9	$5{,}9 - 6{,}5 = -0{,}6$	$5{,}9 - 7{,}0 = -1{,}1$	$\frac{5{,}9}{6{,}5}100 = 90{,}77$	$\frac{5{,}9}{7{,}0}100 = 84{,}29$	$90{,}77 - 100 = -9{,}23$	$84{,}29 - 100 = -15{,}71$
4	5,5	$5{,}5 - 5{,}9 = -0{,}4$	$5{,}5 - 7{,}0 = -1{,}5$	$\frac{5{,}5}{5{,}9}100 = 93{,}22$	$\frac{5{,}5}{7{,}0}100 = 78{,}57$	$93{,}22 - 100 = -6{,}78$	$78{,}57 - 100 = -21{,}43$
5	4,9	$4{,}9 - 5{,}9 = -0{,}4$	$4{,}9 - 7{,}0 = -2{,}1$	$\frac{4{,}9}{5{,}5}100 = 89{,}09$	$\frac{4{,}9}{7{,}0}100 = 70{,}00$	$89{,}09 - 100 = -10{,}91$	$70{,}00 - 100 = -30{,}00$

Решение. Представим расчет цепных и базисных абсолютных приростов, темпов роста, темпов прироста в табл. 10.10.

Для получения обобщающих показателей динамики развития необходимы средние характеристики: средний абсолютный прирост, средний темп роста и средний темп прироста.

Средний абсолютный прирост равен:

$$\overline{\Delta y} = \frac{y_n - y_1}{n-1} = \frac{y_5 - y_1}{4} = \frac{4{,}9 - 7{,}0}{4} = -0{,}525 \text{ млн м}^2,$$

т. е. в среднем ежегодно общая площадь вводимого жилья уменьшалась на 0,525 млн м².

Средний темп роста определим по формуле

$$\overline{T} = \sqrt[n-1]{\frac{y_n}{y_1}}\ 100\ \% = \sqrt[4]{\frac{4{,}9}{7{,}0}}\ 100\ \% = 91{,}47\ \%,$$

т. е. в среднем ежегодно строительство жилья составляло 91,47 % от уровня предыдущего года.

Средний темп прироста $\overline{K} = \overline{T}\quad 100\ \% = -8{,}53\ \%$, т. е. в среднем ежегодно строительство жилья снижалось на 8,53 %.

Наибольший интерес для статистического анализа представляют средний абсолютный прирост, средний темп роста, средний темп прироста. Эти показатели являются обобщающими характеристиками динамики. С их помощью можно строить прогнозы исследуемых показателей. Однако необходимо отметить, что их применение требует определенной осторожности.

Описание динамики ряда с помощью среднего абсолютного прироста соответствует его представлению в виде прямой, проведенной через две крайние точки. В этом случае, чтобы получить прогноз на i шагов вперед (***i*-*период упреждения***), достаточно воспользоваться следующей формулой:

$$\hat{y}_{n+i} = y_n \quad i \cdot \overline{\Delta y}, \tag{10.22}$$

где y_n — фактическое значение в последней n-й точке ряда (конечный уровень ряда); $\hat{y}_{n+i}$ — прогнозная оценка значения (n+i) уровня ряда; $\overline{\Delta y}$ — значение среднего абсолютного прироста, рассчитанное для временнóго ряда $y_1, y_2, \ldots, y_n$.

Очевидно, что такой подход к получению прогнозного значения корректен, если характер развития близок к линейному. На такой равномерный характер развития могут указывать примерно одинаковые значения цепных абсолютных приростов. Продемонстрируем использование рассмотренного приема на следующем примере.

Пример 10.2. Данные табл. 10.11 описывают динамику объема продаж продукции фирмы за последние пять лет.

Текущий номер года, t	1	2	3	4	5
Объем продаж, y_t (тыс. шт.)	24,9	25,4	26,2	26,8	27,7

Требуется: а) обосновать правомерность использования среднего абсолютного прироста для получения прогнозного значения объема продаж в следующем году (при $t = 6$); б) рассчитать прогноз объема продаж в следующем году с помощью среднего абсолютного прироста.

Решение.

1. Рассчитаем цепные абсолютные приросты:

$\Delta y_2 = 25{,}4 - 24{,}9 = 0{,}5$ (тыс. шт.);

$\Delta y_3 = 26{,}2 - 25{,}4 = 0{,}8$ (тыс. шт.);

$\Delta y_4 = 26{,}8 - 26{,}2 = 0{,}6$ (тыс. шт.);

$\Delta y_5 = 27{,}7 - 26{,}8 = 0{,}9$ (тыс. шт.).

Легко заметить, что цепные абсолютные приросты примерно одинаковы. Их значения варьируют незначительно, что свидетельствует о близости процесса развития к линейному. Поэтому для определения прогнозного значения показателя в следующем году ($\hat{y}_6$) можно использовать средний абсолютный прирост, если по оценкам специалистов инерционность развития показателя сохранится.

2. Найдем значение среднего абсолютного прироста, воспользовавшись формулой (10.9):

$$\overline{\Delta y} = \frac{y_5 - y_1}{4} = \frac{27{,}7 - 24{,}9}{4} = 0{,}7 \text{ (тыс. шт.)}.$$

Рассчитаем прогнозное значение объема продаж с помощью выражения (10.22):

$$\hat{y}_6 = y_5 + \overline{\Delta y} = 27{,}7 + 0{,}7 = 28{,}4 \text{ (тыс. шт.)}.$$

Применение среднего темпа роста (и среднего темпа прироста) для описания динамики ряда соответствует его представлению в виде показательной или экспоненциальной кривой, проведенной через две крайние точки. Поэтому использование этого показателя в качестве обобщающего целесообразно для тех процессов, изменение динамики которых происходит примерно с постоянным темпом роста.

В этом случае прогнозное значение на i шагов вперед может быть получено по формуле

$$\hat{y}_{n+i} = y_n \overline{T}^{\,i}, \qquad (10.23)$$

где $\hat{y}_{n+i}$ — прогнозная оценка значения $(n+i)$-го уровня ряда; y_n — фактическое значение в последней n-й точке ряда (конечный уровень ряда); $\overline{T}$ — средний темп роста, рассчитанный для ряда $y_1, y_2, ..., y_n$ (не в % выражении).

Следующий пример иллюстрирует данный подход.

Пример 10.3. Анализ квартальной динамики процентной ставки банка показал, что ее изменение происходило примерно с постоянным темпом роста в течение шести кварталов. Процентная ставка банка в первом квартале составила 10,2 %, в шестом — 14,4 %. По мнению экспертов, сложившийся характер изменения процентной ставки банка сохранится еще полгода. Используя средний темп роста, рассчитать прогноз процентной ставки банка в седьмом и восьмом кварталах.

Решение. Известно, что изменение процентной ставки банка происходило примерно с постоянным темпом роста в течение шести кварталов. Следовательно, вполне правомерно использовать средний темп роста для расчета прогноза этого показателя в седьмом и восьмом кварталах. Средний темп роста согласно (10.14) равен:

$$\overline{T} = \sqrt[n-1]{\frac{y_n}{y_1}}\,100\,\% = \sqrt[6]{\frac{14{,}4}{10{,}2}}\,100\,\%; \quad \overline{T} = 107{,}1\,\%.$$

Рассчитаем прогноз процентной ставки банка в седьмом и восьмом кварталах в соответствии с (10.23): $\hat{y}_7 = y_6\overline{T}$, где $\overline{T}$ — не в процентном выражении; $\hat{y}_7 = 14{,}4 \;\; 1{,}071 = 15{,}4\,\%$; $\hat{y}_8 = 14{,}4 \cdot 1{,}071^2 = 16{,}5\,\%$.

К недостаткам среднего прироста и среднего темпа роста следует отнести то, что они учитывают лишь конечный и начальный уровни ряда, исключая влияние промежуточных уровней. Тем не менее эти показатели имеют весьма широкую область применения, что объясняется чрезвычайной простотой их вычисления. Они могут быть использованы как приближенные, простейшие способы прогнозирования, предшествующие более глубокому количественному и качественному анализу.

В практике исследования динамики явлений и прогнозирования принято считать, что значения уровней временны́х рядов могут содержать следующие компоненты (составные части или структурообразующие элементы): тренд, сезонную компоненту, циклическую компоненту, случайную составляющую. Под ***трендом*** понимают изменение, определяющее общее направление развития, основную тенденцию временнóго ряда. Это систематическая составляющая долговременного действия.

Наряду с долговременными тенденциями во временны́х рядах часто имеют место более или менее регулярные колебания — периодические составляющие рядов динамики. Если период колебаний не превышает одного года, их называют ***сезонными***. Чаще всего причиной их возникновения бывают природно-климатические условия. Примером могут служить колебания цен на сельскохозяйственную продукцию, в частности на картофель. Из года в год наблюдается снижение цен в период после уборки урожая и последующее повышение цен, связанное с необходимостью хранения продукции. Своего «пика» цены достигают перед следующим урожаем. Таким образом, в колебаниях цен прослеживается устойчивая годовая периодичность. Иногда причины сезонных колебаний имеют социальный характер, например, увеличение закупок в предпраздничный период, увеличение платежей в конце квартала и т. д.

При большем периоде колебания считается, что во временны́х рядах имеет место ***циклическая*** составляющая. Примерами могут служить циклы деловой активности, исследованные Кондратьевым, демографические, инвестиционные и другие циклы. В экономических временны́х рядах редко предоставляется возможность для выделения и дальнейшего анализа циклической компоненты, так как для этого желательно иметь ряды динамики, охватывающие не менее чем три полных периода колебаний. Поэтому ряды динамики экономических показателей подчас оказываются слишком «короткими» для проведения такого исследования. Хорошо изучены циклические составляющие в рядах динамики, относящихся к естественным наукам. Например, по многолетним наблюдениям установлена цикличность солнечной активности (с периодом колебаний примерно в одиннадцать лет).

Если из временнóго ряда удалить тренд и периодические составляющие, то останется ***нерегулярная*** компонента. Экономисты разделяют факторы, под действием которых формируется нерегулярная компонента, на два вида. Это факторы резкого, внезапного действия и текущие факторы. Первый тип факторов (например, стихийные бедствия, эпидемии, война, кризис и др.), как правило, вызывает более значительные отклонения, иногда такие отклонения называют катастрофическими колебаниями. Факторы второго типа вызывают случайные колебания, являющиеся результатом действия большого числа побочных причин. Влияние каждого из текущих факторов незначительно, но ощущается их суммарное воздействие.

Если временной ряд представляется в виде суммы соответствующих компонент, то полученная модель носит название *аддитивной* (10.24), если в виде произведения — мультипликативной (10.25). Существует и модель смешанного типа (10.26). Итак:

$$y_t = u_t + s_t + v_t + \varepsilon_t; \tag{10.24}$$

$$y_t = u_t s_t v_t \varepsilon_t; \tag{10.25}$$

$$y_t = u_t s_t v_t + \varepsilon_t, \tag{10.26}$$

где y_t — уровни временно́го ряда; u_t — трендовая составляющая; s_t — сезонная компонента; v_t — циклическая компонента; ε_t — случайная компонента.

Обратимся к рис. 10.1 и 10.2.

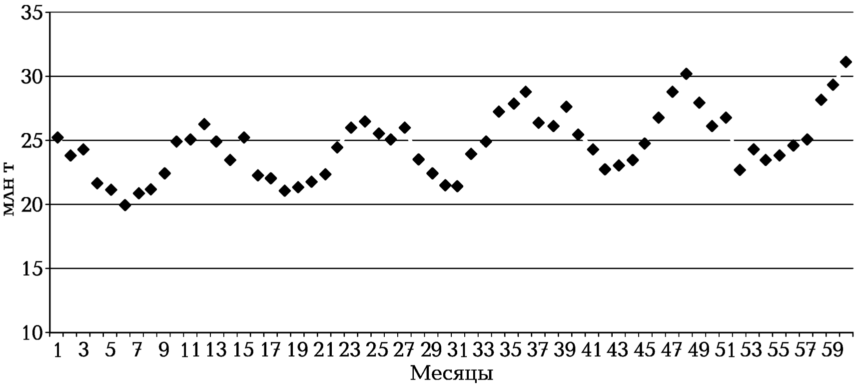


Рис.10.1. Месячная динамика добычи угля

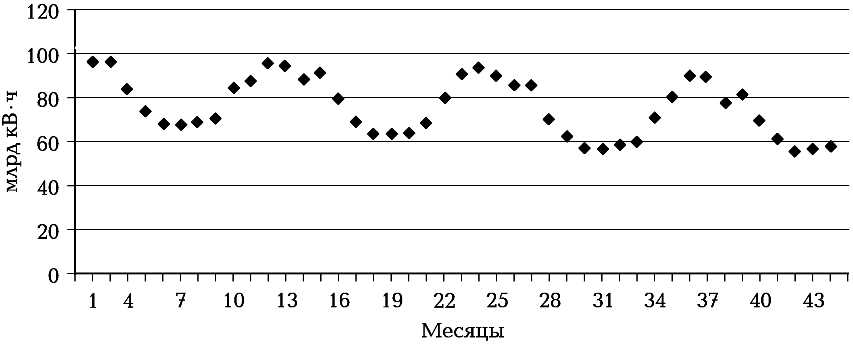


Рис. 10.2. Месячная динамика производства электроэнергии

На рисунках приведены примеры временны́х рядов, иллюстрирующие присутствие в них указанных компонент. Графики месячных временны́х рядов производства электроэнергии и добычи угля наглядно демонстрируют устойчивые сезонные колебания при наличии тренда.

Отметим, что не обязательно в процессе формирования значений уровней каждого временно́го ряда должны участвовать одновременно все компоненты. В изменении значений одного показателя может отсутствовать трендовая компонента, в изменении другого — периодические составляющие. Динамика третьего показателя может описываться лишь случайной компонентой. Однако во всех случаях проведения анализа предполагается наличие случайной, нерегулярной составляющей.

Решение любой задачи по анализу и прогнозированию временны́х рядов начинается с построения графика исследуемого показателя, тем более что современные программные средства предоставляют пользователю большие возможности для этого. Иногда на стадии графического анализа можно определить характер сезонных колебаний: аддитивный или мультипликативный. Отличительной особенностью аддитивной модели является то, что амплитуда сезонных колебаний, отражающая отклонения от тренда или среднего, остается примерно постоянной, неизменной во времени.

На рис. 10.3 изображена динамика объемов пассажирских перевозок, выполненных на авиалиниях Великобритании. На графике отчетливо прослеживаются сезонные колебания, наслаивающиеся на монотонно возрастающий тренд. Ежегодно повторяющиеся пики активности в авиаперевозках приходятся на праздники (рож-

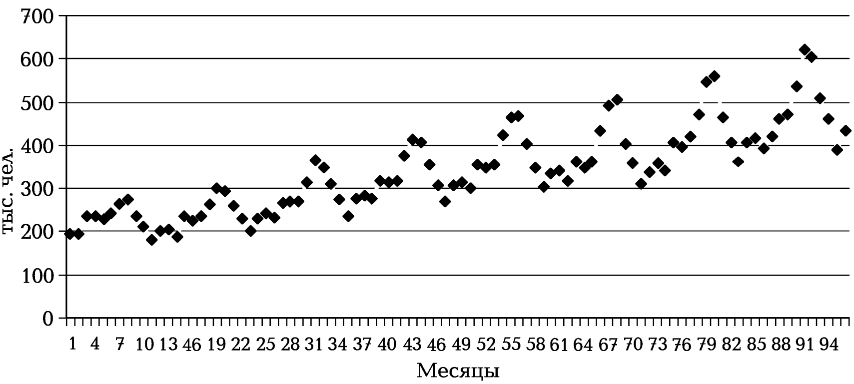


Рис.10.3. Месячная динамика перевозок авиапассажиров

дественские и пасхальные каникулы), а также на время летних отпусков. Причем амплитуда сезонных колебаний возрастает с ростом объемов перевозок, что приводит к выводу о мультипликативном характере сезонности.

Важной задачей, возникающей при анализе рядов динамики, является определение основной тенденции в развитии исследуемого явления. В некоторых случаях общая тенденция ясно прослеживается в динамике показателя, в других ситуациях она может не просматриваться из-за ощутимых случайных колебаний. Например, в отдельные моменты времени сильные колебания в курсах акций могут заслонить наличие тенденции к росту или снижению этого показателя.

На практике для обнаружения общей тенденции часто используют простой прием ***укрупнения интервалов***. Например, ряд недельных данных можно преобразовать в ряд месячной динамики, ряд квартальных данных заменить годовыми уровнями. Уровни нового ряда могут быть получены суммированием уровней исходного ряда либо могут представлять средние значения. Распространенным приемом при выявлении тенденции развития является ***сглаживание временно́го ряда***. Суть различных приемов сглаживания сводится к замене фактических уровней временно́го ряда расчетными уровнями, которые в меньшей степени подвержены колебаниям. Это способствует более четкому проявлению тенденции развития.

Скользящие средние позволяют сгладить как случайные, так и периодические колебания, выявить имеющуюся тенденцию в развитии процесса. Они являются важным инструментом при фильтрации компонент временно́го ряда.

Алгоритм сглаживания ***по простой скользящей средней*** может быть представлен в виде следующей последовательности шагов.

1. Определяют длину интервала сглаживания g, включающего в себя g последовательных уровней ряда ($g < n$). При этом нужно иметь в виду, что чем шире интервал сглаживания, тем в большей степени поглощаются колебания, и тенденция развития носит более плавный, сглаженный характер. Чем сильнее колебания, тем шире должен быть интервал сглаживания.

2. Разбивают весь период наблюдений на участки, при этом интервал сглаживания как бы скользит по ряду с шагом, равным 1.

3. Рассчитывают средние арифметические из уровней ряда, образующих каждый участок.

4. Заменяют фактические значения ряда, стоящие в центре каждого участка, на соответствующие средние значения.

При этом удобно брать длину интервала сглаживания g в виде нечетного числа: $g = 2p + 1$, ибо в этом случае полученные значения скользящей средней приходятся на средний член интервала. Наблюдения, которые берутся для расчета среднего значения, называются ***активным участком сглаживания***. При нечетном значении $g = 2p + 1$ все уровни активного участка могут быть представлены в виде:

$$y_{t-p}, y_{t-p+1}, \ldots, y_{t-1}, y_t, y_{t+1}, \ldots, y_{t+p-1}, y_{t+p},$$

где y_t — центральный уровень активного участка; $y_{t-p}, y_{t-p+1}, \ldots, y_{t-1}$ — последовательность из p уровней активного участка, предшествующих центральному; $y_{t+1}, \ldots, y_{t+p-1}, y_{t+p}$ — последовательность из p уровней активного участка, следующих за центральным.

Тогда скользящая средняя может быть определена по формуле

$$\hat{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \ldots + y_{t+p-1} + y_{t+p}}{2p+1}, \tag{10.27}$$

где y_i — фактическое значение i-го уровня; $\hat{y}_t$ — значение скользящей средней в момент t; $2p + 1$ — длина интервала сглаживания.

Процедура сглаживания приводит к устранению периодических колебаний во временно́м ряду, если длина интервала сглаживания берется равной или кратной периоду колебаний.

Для устранения сезонных колебаний на практике часто требуется использовать скользящие средние с длиной интервала сглаживания равной 4 или 12, но при этом не будет выполняться условие нечетности. Поэтому при четном числе уровней принято первое и последнее наблюдение на активном участке брать с половинными весами:

$$\hat{y}_t = \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + \ldots + y_{t-1} + y_t + y_{t+1} + \ldots + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p} =$$

$$= \frac{\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p}. \tag{10.28}$$

Тогда для сглаживания сезонных колебаний при работе с временными рядами квартальной или месячной динамики можно использовать 4-членную (10.29) и 12-членную (10.30) скользящую среднюю:

$$\hat{y}_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4} \qquad (10.29)$$

и

$$\hat{y}_t = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \ldots + y_t + \ldots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}. \qquad (10.30)$$

В первом случае (10.29) каждый активный участок содержит 5 уровней, во втором (10.30) — 13, при этом крайние уровни имеют половинные весовые коэффициенты.

При использовании скользящей средней с длиной активного участка $g = 2p + 1$ первые и последние p уровней ряда сгладить нельзя, их значения теряются. Очевидно, что потеря значений последних точек является существенным недостатком, так как для исследователя «свежие» данные обладают наибольшей информационной ценностью.

Рассмотрим один из приемов, позволяющих восстановить потерянные значения временнóго ряда. Для этого необходимо:

1) вычислить средний абсолютный прирост на последнем активном участке;

2) получить p сглаженных значений в конце временнóго ряда путем последовательного прибавления среднего абсолютного прироста к последнему сглаженному значению. Аналогичную процедуру можно реализовать для оценивания первых уровней временнóго ряда.

Метод простой скользящей средней применим, если графическое изображение динамического ряда напоминает прямую. Если для процесса характерно нелинейное развитие, то простая скользящая средняя может привести к существенным искажениям. Когда тренд выравниваемого ряда имеет изгибы и для исследователя желательно сохранить мелкие волны, то целесообразно использовать ***взвешенную скользящую среднюю***.

В этом случае на каждом активном участке значение центрального уровня заменяется на расчетное, определяемое по формуле средней арифметической взвешенной:

$$\hat{y}_t = \frac{\sum\limits_{i=t-p}^{t+p} y_i w_i}{\sum\limits_{i=t-p}^{t+p} w_i}, \qquad (10.31)$$

где w_i — весовые коэффициенты.

Простая скользящая средняя учитывает все уровни ряда, входящие в активный участок сглаживания, с равными весами, а взвешенная средняя приписывает каждому уровню вес, зависящий от удаления данного уровня до уровня, стоящего в середине активного участка. Это вызвано тем, что при простой скользящей средней выравнивание на каждом активном участке производится по прямой (по полиному первого порядка). При сглаживании же по взвешенной скользящей средней используются полиномы чаще всего 2-го или 3-го порядков. Поэтому метод простой скользящей средней может рассматриваться как частный случай метода взвешенной скользящей средней.

Весовые коэффициенты определяются методом наименьших квадратов. При этом нет необходимости каждый раз заново вычислять весовые коэффициенты при уровнях ряда, входящих в активный участок сглаживания, поскольку они будут одинаковыми для каждого активного участка.

В табл. 10.12 представлены весовые коэффициенты в зависимости от длины интервала сглаживания (при сглаживании по полиному 2-го или 3-го порядка).

Так как веса симметричны относительно центрального уровня, то в таблице использована символическая запись: приведены веса для половины уровней активного участка; выделен (полужирным шрифтом) вес, относящийся к уровню, стоящему в центре участка сглаживания. Для оставшихся уровней веса не приводятся, так как они могут быть симметрично отражены.

Интервал сглаживания	Весовые коэффициенты
5	$\frac{1}{35}$ [–3, +12, +**17**]
7	$\frac{1}{21}$ [–2, +3, +6, +**7**]
9	$\frac{1}{231}$ [–21, +14, +39, +54, +**59**]
11	$\frac{1}{429}$ [–36, +9, +44, +69, +84, +**89**]
13	$\frac{1}{143}$ [–11, 0, +9, +16, +21, +24, +**25**]

Отметим важные свойства весовых коэффициентов: 1) они симметричны относительно центрального уровня; 2) сумма весов с учетом общего множителя, вынесенного за скобки, равна единице; 3) наличие как положительных, так и отрицательных весов, позволяет сглаженной кривой сохранять различные изгибы кривой тренда.

Проиллюстрируем использование табл. 10.12 на примере вычисления пятичленной взвешенной скользящей средней. В этом случае центральное значение на каждом активном участке y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2} будет оцениваться по формуле

$$\hat{y}_t = \frac{1}{35}(\ 3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}),$$

где соответствующие весовые коэффициенты уровней – 3/35, 12/35, 17/35 — взяты из первой строки табл. 10.12.

Существуют приемы, позволяющие с помощью дополнительных вычислений получить сглаженные значения для p потерянных начальных и конечных уровней ряда при длине интервала сглаживания $g = 2p + 1$.

Пример 10.4. По данным об урожайности за 16 лет (табл. 10.13) рассчитать: 1) трех-, семилетние скользящие средние и графически сравнить результаты; 2) пятилетнюю взвешенную скользящую среднюю.

Решение. 1. Результаты расчетов представлены в табл. 10.14.

При трехлетней скользящей средней (гр. 3 табл. 10.14):

$$\hat{y}_2 = \frac{12{,}4 + 16{,}4 + 14{,}1}{3} = 14{,}3;$$

$$\hat{y}_3 = \frac{16{,}4 + 14{,}1 + 14{,}6}{3} = 15{,}0$$

и т. д.

Текущий номер года t	1	2	3	4	5	6	7	8
y_t	12,4	16,4	14,1	14,6	16,5	18,8	17,4	22,3

Текущий номер года t	9	10	11	12	13	14	15	16
y_t	19,2	15,9	19,6	19,6	22,0	17,5	20,8	22,8

* Данные условного региона.

t	y_t	Скользящие средние		Взвешенная скользящая средняя $g = 5$
		$g = 3$	$g = 7$	
1	2	3	4	5
1	12,4	—	—	—
2	16,4	14,3	—	—
3	14,1	15,0	—	15,0
4	14,6	15,1	15,7	14,6
5	16,5	16,6	17,2	16,8
6	18,8	17,6	17,6	17,6
7	17,4	19,5	17,8	19,5
8	22,3	19,6	18,5	20,4
9	19,2	19,1	19,0	19,3
10	15,9	18,2	19,4	17,4
11	19,6	18,4	19,4	18,2
12	19,6	20,4	19,2	20,9
13	22,0	19,7	19,7	19,9
14	17,5	20,1	—	19,5
15	20,8	20,4	—	—
16	22,8	—	—	—

* На основе данных табл. 10.13.

При семилетней скользящей средней (гр. 4 табл. 10.14):

$$\hat{y}_4 = \frac{12,4 + 16,4 + 14,1 + 14,6 + 16,5 + 18,8 + 17,4}{7} = 15,7;$$

$$\hat{y}_5 = \frac{16,4 + 14,1 + 14,6 + 16,5 + 18,8 + 17,4 + 22,3}{7} = 17,2$$

и т. д.

Графический анализ (рис. 10.4) показывает, что ряд, сглаженный по семилетней скользящей средней, имеет более гладкий характер. Это объясняется тем, что чем больше длина интервала сглаживания, тем более гладкий ряд на выходе модели.

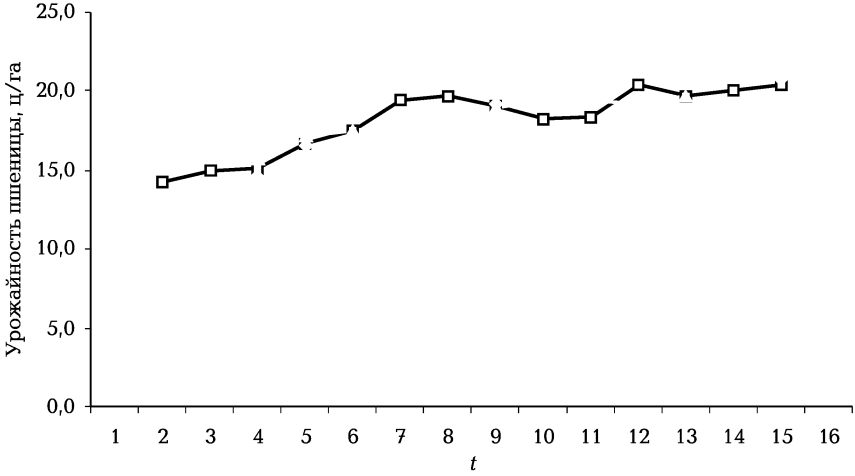


Рис. 10.4. Сглаживание временного ряда урожайности с помощью скользящих средних:

фактические уровни y_t; —□— скользящая средняя, g = 3; скользящая средняя, g = 7

2. Для вычисления значений пятилетней взвешенной скользящей средней воспользуемся табл. 10.12. Тогда:

$$\hat{y}_3 = \frac{1}{35}\ (-3 \cdot 12{,}4 + 12 \cdot 16{,}4 + 17 \cdot 14{,}1 + 12 \cdot 14{,}6 - 3 \cdot 16{,}5) = 15{,}0;$$

$$\hat{y}_4 = \frac{1}{35}\ (-3 \cdot 16{,}4 + 12 \cdot 14{,}1 + 17 \cdot 14{,}6 + 12 \cdot 16{,}5 - 3 \cdot 18{,}8) = 14{,}6$$

и т. д.

Результаты дальнейших расчетов отражены в гр. 5 табл. 10.14.

На практике для описания тенденции развития явления широко используются ***модели кривых роста***, представляющие собой различные функции времени $y = f(t)$. При таком подходе изменение исследуемого показателя связывают лишь с течением времени, считается, что влияние других факторов несущественно или косвенно сказывается через фактор времени. Правильно выбранная модель кривой роста должна соответствовать характеру изменения тенденции исследуемого явления. Кривая роста позволяет получить выровненные или теоретические значения уровней динамического ряда. Это те уровни, которые наблюдались бы в случае полного совпадения динамики явления с кривой.

Прогнозирование на основе модели кривой роста базируется на экстраполяции, т. е. на продлении в будущее тенденции, наблюдавшейся в прошлом. При этом предполагается, что во временно́м ряду присутствует тренд, характер развития показателя обладает свойством инерционности. Сложившаяся же тенденция не должна претерпевать существенных изменений в течение периода упреждения.

Процедура разработки прогноза с использованием кривых роста включает в себя следующие этапы: 1) выбор одной или нескольких кривых, форма которых соответствует характеру изменения временно́го ряда; 2) оценка параметров выбранных кривых; 3) проверка адекватности выбранных кривых прогнозируемому процессу и окончательный выбор кривой роста; 4) расчет точечного и интервального прогнозов.

К настоящему времени в литературе описано несколько десятков кривых роста. Эти модели условно могут быть разделены на три класса в зависимости от того, какой тип динамики развития они лучше описывают.

К I классу относятся функции, используемые для описания процессов с ***монотонным характером развития*** и отсутствием пределов роста. Эти условия справедливы для многих экономических показателей, например, для большинства показателей промышленного производства в натуральном выражении.

Ко II классу относятся кривые, описывающие процесс, который имеет предел роста в исследуемом периоде. С такими процессами часто сталкиваются в демографии, при изучении потребностей в товарах и услугах (в расчете на душу населения), при исследовании эффективности использования ресурсов и т. д. Примерами показателей, для которых могут быть указаны пределы роста, являются среднедушевое потребление определенных продуктов питания, расход удобрений на единицу площади и т. п. Функции, относящиеся ко II классу, называются ***кривыми насыщения***.

Если кривые насыщения имеют точки перегиба, то они относятся к III классу кривых роста — к ***S-образным кривым***. Эти кривые описывают как бы два последовательных лавинообразных процесса (когда прирост зависит от уже достигнутого уровня): один с ускорением развития, другой — с замедлением. *S*-образные кривые находят применение в демографических исследованиях, в страховых расчетах, при решении задач прогнозирования научно-технического прогресса, при определении спроса на новый вид продукции.

Вопрос о выборе кривой является основным при выравнивании ряда.

Существует несколько подходов к решению этой задачи, и все они предполагают знакомство с основными свойствами используемых кривых роста. Поэтому остановимся на характеристике отдельных типов кривых, наиболее часто применяемых на практике.

Среди кривых роста I класса прежде всего следует выделить класс полиномов:

$$y_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_p t^p, \tag{10.32}$$

где a_i ($i = 0, 1, \ldots, p$) — параметры многочлена; t — независимая переменная (время), $t = 1, 2, \ldots, n$.

Коэффициенты полиномов невысоких степеней могут иметь конкретную интерпретацию в зависимости от содержания динамического ряда. Например, их можно трактовать как скорость роста (a_1), ускорение роста (a_2), изменение ускорения (a_3).

Обычно в экономических исследованиях применяются полиномы не выше третьего порядка. Использовать для определения тренда полиномы высоких степеней нецелесообразно, поскольку полученные таким образом аппроксимирующие функции будут отражать случайные отклонения (что противоречит смыслу тенденции). Полином первой степени $y_t = a_0 + a_1 t$ на графике изображается прямой и используется для описания процессов, развивающихся во времени равномерно.

Полином второй степени (парабола) $y_t = a_0 + a_1 t + a_2 t^2$ применим в тех случаях, когда процесс развивается равноускоренно (т. е. имеется равноускоренный рост или равноускоренное снижение уровней). Как известно, если параметр $a_2 > 0$, то ветви параболы направлены вверх, если же $a_2 < 0$, то вниз. Параметры a_0 и a_1 не влияют на форму параболы, а лишь определяют ее положение.

Полином третьей степени имеет вид $y_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$. У этого полинома знак прироста ординат может изменяться один или два раза (см. рис. 10.5).

Оценки параметров в модели (10.32) находятся методом наименьших квадратов. Как известно, суть его состоит в определении таких коэффициентов (параметров), при которых сумма квадратов отклонений расчетных значений уровней от фактических значений была бы минимальной. Таким образом, эти оценки находятся в результате минимизации выражения:

$$\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \to \min, \tag{10.33}$$

где y_t — фактическое значение уровня временнóго ряда; $\hat{y}_t$ — расчетное значение; n — длина временнóго ряда.

В результате минимизации выражения (10.33) получается система нормальных уравнений:

$$\begin{cases} \sum_{t=1}^{n} y_t = a_0 n + a_1 \sum_{t=1}^{n} t + a_2 \sum_{t=1}^{n} t^2 + \ldots + a_p \sum_{t=1}^{n} t^p; \\ \sum_{t=1}^{n} y_t t = a_0 \sum_{t=1}^{n} t + a_1 \sum_{t=1}^{n} t^2 + a_2 \sum_{t=1}^{n} t^3 + \ldots + a_p \sum_{t=1}^{n} t^{p+1}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots ; \\ \sum_{t=1}^{n} y_t t^{p-1} = a_0 \sum_{t=1}^{n} t^{p-1} + a_1 \sum_{t=1}^{n} t^p + a_2 \sum_{t=1}^{n} t^{p+1} + \ldots + a_p \sum_{t=1}^{n} t^{2p-1}; \\ \sum_{t=1}^{n} y_t t^p = a_0 \sum_{t=1}^{n} t^p + a_1 \sum_{t=1}^{n} t^{p+1} + a_2 \sum_{t=1}^{n} t^{p+2} + \ldots + a_p \sum_{t=1}^{n} t^{2p}. \end{cases} \tag{10.34}$$

Возможность упрощения вычислений за счет переноса начала отсчета t рассмотрим ниже. Система (10.34) состоит из $(p + 1)$ уравнений, содержащих в качестве неизвестных величин $(p + 1)$ коэффициентов $a_0, a_1, \ldots, a_p$. Решение этой системы позволяет вычислить оценки искомых коэффициентов.

Системы для оценивания полиномов невысоких степеней выглядят намного проще. Например, нормальные уравнения для оценивания параметров линейной модели имеют следующий вид:

$$\begin{cases} \sum_{t=1}^{n} y_t = a_0 n + a_1 \sum_{t=1}^{n} t; \\ \sum_{t=1}^{n} y_t t = a_0 \sum_{t=1}^{n} t + a_1 \sum_{t=1}^{n} t^2. \end{cases} \tag{10.35}$$

Решение этой системы относительно искомых параметров дает следующие выражения:

$$a_1 = \frac{\sum_{t=1}^{n} y_t t - \frac{\sum_{t=1}^{n} y_t}{n} \sum_{t=1}^{n} t}{\sum_{t=1}^{n} t^2 - \frac{\left(\sum_{t=1}^{n} t\right)^2}{n}}; \qquad a_0 = \frac{\sum_{t=1}^{n} y_t}{n} - a_1 \frac{\sum_{t=1}^{n} t}{n}. \tag{10.36}$$

Для параболической модели (полинома 2-го порядка) получим аналогичную систему нормальных уравнений:

$$\begin{cases} \sum\limits_{t=1}^{n} y_t = a_0 n + a_1 \sum\limits_{t=1}^{n} t + a_2 \sum\limits_{t=1}^{n} t^2; \\ \sum\limits_{t=1}^{n} y_t t = a_0 \sum\limits_{t=1}^{n} t + a_1 \sum\limits_{t=1}^{n} t^2 + a_2 \sum\limits_{t=1}^{n} t^3; \\ \sum\limits_{t=1}^{n} y_t t^2 = a_0 \sum\limits_{t=1}^{n} t^2 + a_1 \sum\limits_{t=1}^{n} t^3 + a_2 \sum\limits_{t=1}^{n} t^4. \end{cases} \qquad (10.37)$$

Эта система содержит три уравнения, позволяющих найти оценки трех неизвестных коэффициентов a_0, a_1, a_2.

Составление нормальных уравнений можно упростить, воспользовавшись тем, что величины $\sum t$, $\sum t^2$, ... не зависят от конкретных уровней динамического ряда. Эти суммы являются функциями только числа членов в динамическом ряду. Для них получены следующие формулы:

$$\sum t = \frac{n(n+1)}{2}; \quad \sum t^2 = \frac{n(n+1)(2n+1)}{6}; \quad \sum t^3 = \frac{n^2(n+1)^2}{4};$$

$$\sum t^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$$ (суммирование по $t = 1, 2, ..., n$).

Другой подход к упрощению расчетов заключается в переносе начала координат в середину ряда динамики. Это позволяет упростить сами нормальные уравнения, а также уменьшить абсолютные значения величин, участвующих в расчете. Если до переноса начала координат t было равно 1, 2, 3, ..., то после переноса:

а) для четного числа членов ряда $t = ..., -5; -3; -1; 1; 3; 5; ...$;

б) для нечетного числа членов ряда $t = ..., -3; -2; -1; 0; 1; 2; 3; ...$.

В формулах (10.38 — 10.41) суммирование производится по t, полученному после переноса начала координат. При этом учитывается, что в этом случае $\sum t^k = 0$ (где k — нечетное число). Такой подход существенно упрощает систему (10.34). В частности, система нормальных уравнений для линейной модели (10.35) за счет того, что $\sum t = 0$, примет следующий вид:

$$\begin{cases} \sum y_t = a_0 n; \\ \sum y_t t = a_1 \sum t^2. \end{cases} \qquad (10.38)$$

Для параболической модели система (10.37) также существенно упрощается, так как после переноса начала координат в середину ряда динамики $\sum t = 0$ и $\sum t^3 = 0$. Получаются следующие формулы:

$$\begin{cases} \sum y_t = a_0 n \quad a_2 \sum t^2; \\ \sum y_t t = a_1 \sum t^2; \\ \sum y_t t^2 = a_0 \sum t^2 \quad a_2 \sum t^4. \end{cases} \tag{10.39}$$

В этом случае оценки параметров соответствующих моделей имеют вид:

для линейной модели:

$$a_0 = \frac{\sum y_t}{n}; \qquad a_1 = \frac{\sum y_t t}{\sum t^2}; \tag{10.40}$$

для параболической модели:

$$\begin{gathered} a_0 = \frac{\sum y_t}{n} \quad \frac{\sum t^2}{n} \left[\frac{n \sum y_t t^2 - \sum t^2 \sum y_t}{n \sum t^4 - \left(\sum t^2 \right)^2} \right]; \\ a_1 = \frac{\sum y_t t}{\sum t^2}; \qquad a_2 = \frac{n \sum y_t t^2 - \sum t^2 \sum y_t}{n \sum t^4 - \left(\sum t^2 \right)^2}. \end{gathered} \tag{10.41}$$

Для класса экспоненциальных кривых, в отличие от полиномов, характерной является зависимость приростов от величины самой функции. Эти кривые хорошо описывают процессы, имеющие «лавинообразный» характер, когда прирост зависит от достигнутого уровня функции.

Простая показательная (экспоненциальная) кривая имеет вид

$$y_t = ab^t. \tag{10.42}$$

Если $b > 1$, то кривая растет вместе с ростом t. Но она падает, если $b < 1$ (при $a > 0$).

Параметр a характеризует начальные условия развития, а параметр b — постоянный темп роста.

Действительно, темп роста равен

$$T_t = \frac{y_t}{y_{t-1}} \quad 100\ \%.$$

В данном случае

$$T_t = \frac{ab^t}{ab^{t-1}} \quad 100\ \% = b \quad 100\ \% = \text{const}.$$

Соответственно и темпы прироста — постоянны:

$$K_t = T_t - 100\,\% = \text{const}.$$

Можно показать, что логарифм ординаты этой функции линейно зависит от t, для этого прологарифмируем выражение (10.42): $\log y_t = \log a + t \cdot \log b$. Пусть $\log a = A$; $\log b = B$. Тогда $\log y_t = A + tB$.

Теперь для оценки неизвестных параметров можно использовать систему нормальных уравнений для прямой (10.35).

Иначе говоря, нормальные уравнения строятся исходя из минимизации:

$$\sum_{t=1}^{n} (\log y_t - \log \hat{y}_t)^2 \to \min.$$

Соответственно в нормальных уравнениях вместо фактических уровней выступают их логарифмы:

$$\sum_{t=1}^{n} \log y_t = nA + B\sum_{t=1}^{n} t; \quad \sum_{t=1}^{n} (\log y_t \cdot t) = A\sum_{t=1}^{n} t + B\sum_{t=1}^{n} t^2. \qquad (10.43)$$

Найдем неизвестные параметры A и B. Зная значения $A = \log a$ и $B = \log b$, определим значения a и b и с помощью потенцирования получим показательную функцию, служащую для выравнивания ряда. Такой подход к оцениванию неизвестных параметров привлекает своей универсальностью.

Все рассмотренные типы кривых используются для описания монотонно возрастающих или убывающих процессов без «насыщения». Если же процесс характеризуется «насыщением», его следует описывать при помощи кривой, имеющей отличную от нуля асимптоту. Примером такой кривой может служить модифицированная экспонента:

$$y_t = k + ab^t, \qquad (10.44)$$

где $y = k$ является горизонтальной асимптотой.

Если параметр a отрицателен, то асимптота находится выше кривой, если a положителен, то ниже. При решении экономических задач чаще всего приходится иметь дело с кривой, у которой $a < 0$, $b < 1$. В этом случае рост уровней происходит с замедлением и стремится к некоторому пределу.

При решении экономических задач нередко приходится определять значение асимптоты, исходя из свойств прогнозируемого процесса (например, коэффициент использования оборудования не может превышать 1). Иногда значение асимптоты задается экспертным путем. Например, главный инженер предприятия указывает, что производственные мощности не позволят наращивать объемы производства выше определенного уровня. Этот уровень является оценкой значения асимптоты k при прогнозировании производства продукции.

Таким образом, модифицированная экспонента хорошо описывает процесс, на развитие которого воздействует ограничивающий фактор, причем влияние этого воздействия растет вместе с ростом достигнутого уровня.

Если воздействие ограничивающего фактора начинает сказываться только после определенного момента (точки перегиба), до которого процесс развивался по некоторому экспоненциальному закону, то для выравнивания используют S-образные кривые.

Наиболее известными являются кривая Гомперца (10.45) и логистическая кривая (кривая Перла — Рида). Кривая Гомперца описывается следующей формулой:

$$y_t = k \quad a^{b^t}. \tag{10.45}$$

Кривая несимметрична. Если $\log a < 0$, кривая имеет S-образный вид, при этом асимптота, равная k, проходит выше кривой. Если $\log a > 0$, асимптота, равная k, лежит ниже кривой, а сама кривая изменяется монотонно. При $b < 1$ она монотонно убывает, а при $b > 1$ — монотонно возрастает.

Для решения экономических задач наибольший интерес представляет вариант кривой, когда $\log a < 0$ и $b < 1$ (рис. 10.5).

Уравнение логистической кривой получается путем замены в модифицированной экспоненте y_t обратной величиной $\frac{1}{y_t}$. В результате получаем

$$\frac{1}{y_t} = k \quad ab^t. \tag{10.46}$$

Используется и другая форма записи уравнения логистической кривой: $y_t = \frac{k}{1 + be^{-at}}$.

При $t \to -\infty$ ордината стремится к нулю, а при $t \to \infty$ — к асимптоте, равной значению параметра k. Кривая симметрична относительно точки перегиба с координатами: $t = \ln b : a$; $y_t = k : 2$.

Как видно из графика (см. рис. 10.5), логистическая функция возрастает сначала ускоренным темпом, затем темп роста замедляется и, наконец, рост почти полностью прекращается, о чем свидетельствует тот факт, что кривая асимптотически приближается к некоторой прямой, параллельной оси абсцисс.

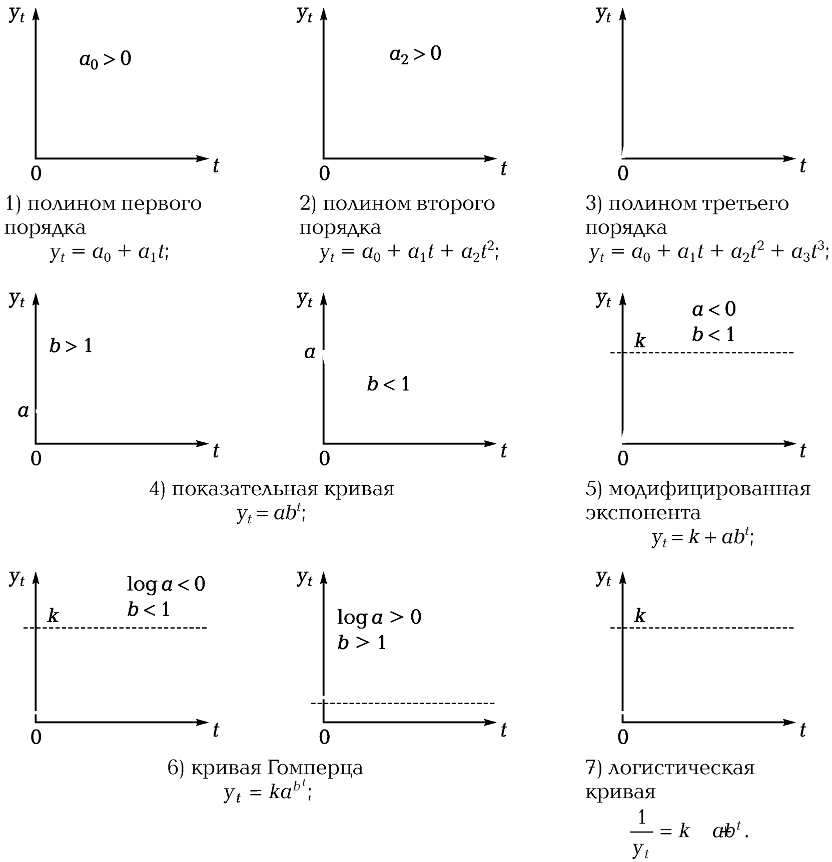


Рис. 10.5. Кривые роста

С помощью этой функции хорошо описывается развитие новой отрасли (нового производства). Сначала технические методы производства еще недостаточно разработаны, издержки производства высоки и спрос на рынке на данный товар еще очень мал, поэтому производство развивается медленно. В дальнейшем, благодаря усовершенствованию технических методов изготовления, переходу к массовому производству и увеличению емкости рынка для данного товара производство растет быстрее. Затем наступает период насыщения рынка, рост производства все более замедляется и, наконец, почти прекращается. Наступает стабилизация производства на определенном уровне. Однако выявленные закономерности развития следует обобщать с определенной осторожностью, причем для коротких периодов. Выявленная тенденция развития производства может быть нарушена, например, вследствие технического переворота в данной отрасли или в связанной с ней.

Таким образом, мы рассмотрели наиболее часто используемые виды кривых роста. Выявленные особенности и свойства этих кривых могут существенно помочь при решении задачи выбора типа кривой. Теперь остановимся подробнее на некоторых практических подходах, облегчающих процесс выбора кривой роста.

Наиболее простой путь — это визуальный, опирающийся на графическое изображение временно́го ряда. Подбирают такую кривую роста, форма которой соответствует фактическому развитию процесса. Если на графике исходного ряда тенденция развития недостаточно четко просматривается, то можно провести некоторые стандартные преобразования ряда (например, сглаживание), а потом подобрать функцию, отвечающую графику преобразованного ряда. В современных пакетах статистической обработки имеется богатый арсенал стандартных преобразований данных и широкие возможности для графического изображения, в том числе в различных масштабах. Все это позволяет существенно упростить для исследователя проведение данного этапа.

В статистической литературе описан метод последовательных разностей, помогающий при выборе кривых полиномиального типа. Этот метод применим при выполнении следующего условия: уровни временно́го ряда могут быть представлены в виде суммы систематической составляющей и случайной компоненты, подчиненной нормальному закону распределения с математическим ожиданием, равным 0, и постоянной дисперсией. Метод предполагает вычисление первых, вторых и т. д. разностей уровней ряда:

$$\Delta y_t = y_t \quad y_{t-1}; \; \Delta^2 y_t = \Delta y_t \quad \Delta y_{t-1} \text{ и т. д.}$$

Расчет ведется до тех пор, пока разности не станут примерно равными. Порядок разностей принимается за степень выравнивающего полинома.

Однако чаще всего на практике к выбору формы кривой подходят, исходя из значений критерия, в качестве которого принимают сумму квадратов отклонений фактических значений уровней от расчетных, получаемых выравниванием.

Из рассматриваемых кривых предпочтение будет отдано той, которой соответствует минимальное значение критерия, ибо чем меньше значение критерия, тем ближе к кривой ложатся данные наблюдений.

Используя этот подход, следует иметь в виду ряд моментов. Во-первых, к ряду, состоящему из m точек, можно подобрать многочлен (полином) степени (m – 1), проходящий через все m точек. Кроме того, существует множество многочленов более высоких степеней, также проходящих через все эти точки. Для этих многочленов значение критерия будет равно 0, однако очевидно, что такая кривая не слишком пригодна как для выделения тенденции, так и для целей прогнозирования.

Применение данного подхода должно проходить в два этапа. На первом происходит ограничение приемлемых функций, исходя из содержательного анализа задачи. На втором осуществляется расчет значений критерия и выбор на его основе наиболее подходящей кривой роста. Необходимость содержательного анализа изучаемого процесса развития может быть проиллюстрирована следующими примерами.

Предположим, что на ретроспективном участке ряд динамики может быть хорошо описан с помощью экспоненциальной кривой. Однако первая половина логистической кривой также представлена экспонентой. Поэтому принять гипотезу об экспоненциальной тенденции ряда в будущем можно только после проведения содержательного анализа. В его ходе следует дать ответ на вопрос: возможно ли наступление «насыщения» при данной совокупности условий. Например, процесс производства может быть ограничен материальными ресурсами или производственными мощностями.

Не исключена ситуация, когда наилучшей функцией по данному критерию будет признана прямая, однако, полученное на ее основе прогнозное значение будет отрицательным. Если из экономической сути показателя вытекает невозможность отрицательных значений (например, при прогнозировании объема выпускаемой продукции), то, естественно, следует отказаться от этой функции, выбрав менее «удачную» по данному критерию, но соответствующую содержательному смыслу показателя. Например, более подходящей в этом случае может оказаться показательная кривая при значении параметра $b < 1$ (см. рис. 10.5).

Иногда в качестве критерия выбирается средняя квадратическая ошибка:

$$s = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n-k}}, \tag{10.47}$$

где y_t — фактическое значение уровня ряда; $\hat{y}_t$ — расчетное значение уровня ряда, полученное по модели; k — число оцениваемых коэффициентов модели; n — длина ряда.

В заключение отметим, что не существует «жестких» рекомендаций для выбора кривых роста. Особенно осторожно следует подходить к решению этой задачи при использовании полученной функции для экстраполяции найденных закономерностей в будущее. Применение кривых роста должно базироваться на предположении о сохранении выявленной тенденции в прогнозируемом периоде. Рассмотренные в данном подразделе различные статистические приемы и методы могут помочь исследователю при осуществлении сложного выбора подходящей кривой роста.

Пример 10.5. В табл. 10.15 представлены данные об остатках вкладов населения в банках за 15 мес. Остатки вкладов указаны на начало каждого месяца.

Необходимо рассчитать прогноз остатков вкладов населения в банках на начало 16-го месяца, исходя из предположения, что тенденция ряда может быть описана: 1) линейной моделью $y_t = a_0 + a_1 t$; 2) параболической моделью $y_t = a_0 + a_1 t + a_2 t^2$.

Решение. 1. Для расчета коэффициентов линейного тренда воспользуемся выражениями, полученными из системы нормальных уравнений после переноса начала координат в середину ряда (10.40). Коль скоро число уровней ряда динамики — нечетное ($n = 15$), то центральный уровень (восьмой) принимается за начало отсчета, ему соответствует $t = 0$. Вышестоящие уровни нумеруются с шагом –1, нижестоящие — с шагом +1 (гр. 3 табл. 10.16).

В табл. 10.16 представлены необходимые вспомогательные вычисления.

В соответствии с (10.40):

$$a_0 = \frac{\sum\limits_t y_t}{n} = \frac{495\,958}{15} = 33\,063{,}866; \quad a_1 = \frac{\sum\limits_t y_t t}{\sum\limits_t t^2} = \frac{899\,891}{280} = 3\,213{,}896.$$

Следовательно, уравнение линейного тренда имеет вид

$$\hat{y}_t = 33\,063{,}866 + 3\,213{,}896\, t.$$

Порядковый номер месяца	y_t	Порядковый номер месяца	y_t	Порядковый номер месяца	y_t
1	14 717	6	23 342	11	40 524
2	16 642	7	28 317	12	45 416
3	18 504	8	30 624	13	50 857
4	20 376	9	33 408	14	56 024
5	21 321	10	36 505	15	59 381

№	y_t	t	$y_t t$	t^2	$y_t t^2$	t^4
1	2	3	4	5	6	7
1	14 717	–7	–103 019	49	721 133	2 401
2	16 642	–6	–99 852	36	599 112	1 296
3	18 504	–5	–92 520	25	462 600	625
4	20 376	–4	–81 504	16	326 016	256
5	21 321	–3	–63 963	9	191 889	81
6	23 342	–2	–46 684	4	93 368	16
7	28 317	–1	–28 317	1	28 317	1
8	30 624	0	0	0	0	0
9	33 408	1	33 408	1	33 408	1
10	36 505	2	73 010	4	146 020	16
11	40 524	3	121 572	9	364 716	81
12	45 416	4	181 664	16	726 656	256
13	50 857	5	254 285	25	1 271 425	625
14	56 024	6	336 144	36	2 016 864	1 296
15	59 381	7	415 667	49	2 909 669	2 401
Σ	495 958	—	899 891	280	9 891 193	9 352

Согласно этой модели оценка среднего уровня ряда при $t = 0$ равна 33 063,9 млрд руб., а среднемесячный прирост остатков вкладов населения составляет 3 213,9 млрд руб.

Для прогнозирования на базе полученной модели на одну точку вперед необходимо в нее подставить соответствующее значение временнóго параметра, т. е. $t = 8$[1].

Прогноз равен: $\hat{y}_8 = 33\,063{,}866 + 3\,213{,}896 \cdot 8$; $\hat{y}_8 = 58\,775$ (млрд руб.).

2. Для расчета коэффициентов параболического тренда также воспользуемся выражениями, полученными из системы нормальных уравнений после переноса начала координат в середину ряда (10.41) (промежуточные вычисления представлены в табл. 10.16).

$$a_1 = \frac{899\,891}{280} \quad 3\,213{,}896; \quad a_2 = \frac{15 \cdot 9\,891\,193 - 280 \cdot 495\,958}{15 \cdot 9\,352 - (280)^2} = 153{,}517;$$

$$a_0 = 33\,063{,}866 - \frac{280}{15} \cdot 153{,}517 = 30\,198{,}16.$$

[1] Если бы оценки коэффициентов модели были получены без переноса начала координат в середину ряда, то следовало бы подставить в модель значение временнóго параметра $t = 16$.

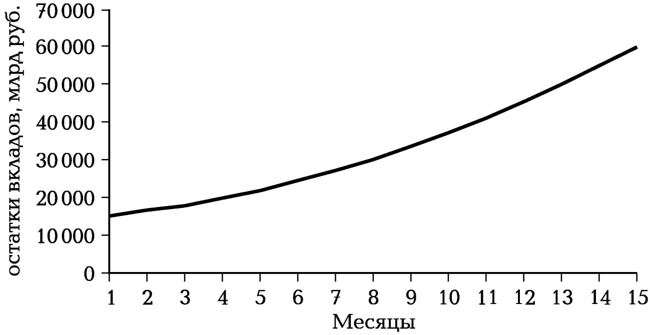


Рис. 10.6. Фактические и расчетные уровни ряда динамики, полученные по линейной и параболической модели:

—— фактические уровни; расчетные уровни (линейная модель); расчетные уровни (параболическая модель)

Следовательно, уравнение параболического тренда примет вид

$$\hat{y}_t = 30\,198{,}16 \quad 3\,213{,}896t + 153{,}517t^2.$$

Для определения прогноза показателя нужно подставить в полученную модель соответствующее значение временно́го параметра ($t = 8$).

Прогноз равен:

$$\hat{y}_8 = 30\,198{,}16 \quad 3\,213{,}896 \cdot 8 + 153{,}517 \cdot 8^2 = 65\,734 \text{ (млрд руб.)}.$$

На рис. 10.6 изображен график исходного временно́го ряда и выровненные значения уровней, полученные на основе двух трендовых моделей: линейной и параболической.

Графический анализ свидетельствует, что линейную модель нельзя признать адекватной. Полученный на ее основе прогноз будет сильно занижен. Значительно ближе к фактическим данным ложатся уровни, выровненные по параболической модели, хотя прогноз может быть несколько завышен. Дальнейшее исследование качества полученных моделей должно опираться на показатели, рассматриваемые в следующем подразделе.

Вопрос о возможности применения построенных моделей в целях анализа и прогнозирования явления может быть решен только после проверки адекватности, т. е. соответствия модели исследуемому процессу.

Проверка адекватности выбранных моделей реальному процессу (в частности, адекватности полученной кривой роста) строится на анализе остаточной компоненты. Остаточная компонента получается после выделения из исследуемого ряда систематической составляющей (тренда и периодической составляющей, если она присутствует во временном ряду).

Предположим, что исходный временной ряд описывает процесс, не подверженный сезонным колебаниям, т. е. примем гипотезу об аддитивной модели ряда, содержащей трендовую и случайную компоненты.

Тогда ряд остатков будет получен как отклонения фактических уровней временно́го ряда (y_t) от расчетных ($\hat{y}_t$):

$$e_t = y_t - \hat{y}_t. \tag{10.48}$$

При использовании кривых роста $\hat{y}_t$ вычисляют, подставляя в уравнения выбранных кривых соответствующие последовательные значения времени.

Принято считать, что модель адекватна описываемому процессу, если остаточная последовательность (ряд остатков) представляет собой случайную компоненту ряда, т. е. удовлетворяет определенным свойствам. Что это за свойства? 1. Случайность колебаний уровней остаточной последовательности. 2. Соответствие распределения остаточной компоненты нормальному закону с нулевым математическим ожиданием. 3. Независимость значений уровней ряда остатков между собой.

Рассмотрим подробнее последнее свойство. Если вид функции, описывающей систематическую составляющую, выбран неудачно, то последовательные значения ряда остатков могут не обладать свойствами независимости, ибо они могут коррелировать между собой. В этом случае говорят, что имеет место *автокорреляция* остатков.

Существует несколько приемов обнаружения автокорреляции. Наиболее распространен подход, опирающийся на критерий Дарбина — Уотсона. Критерий Дарбина — Уотсона связан с проверкой гипотезы об отсутствии автокорреляции первого порядка, т. е. автокорреляции между соседними остаточными членами ряда. Значение этого критерия определяется по формуле

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}. \tag{10.49}$$

Можно показать, что величина d приближенно равна

$$d \approx 2(1 - r_1), \tag{10.50}$$

где r_1 — коэффициент автокорреляции первого порядка (т. е. парный коэффициент корреляции между двумя рядами e_1, e_2, ..., e_{n-1} и e_2, e_3, ..., e_n).

Из последней формулы видно, что близость значения d к нулю означает наличие высокой положительной автокорреляции (коэффициент r_1 близок к единице); близость значения d к 4 означает наличие высокой отрицательной автокорреляции (коэффициент r_1 близок к –1). В случае отсутствия автокорреляции значение d будет близко к 2 (коэффициент r_1 не сильно отличается от нуля).

Для этого критерия найдены критические границы, позволяющие проводить проверку гипотезы об отсутствии автокорреляции. Авторами критерия границы определены для 1, 2,5 и 5 % уровней значимости. Фрагмент критических значений критерия Дарбина — Уотсона при 5%-м уровне значимости приведен в табл. 10.17.

В этой таблице d_1 и d_2 — соответственно нижняя и верхняя пороговые границы критерия Дарбина — Уотсона; k' — число объясняющих переменных в модели; n — длина временнóго ряда.

Применение на практике критерия Дарбина — Уотсона основано на сравнении величины d, рассчитанной по формуле (10.49), с теоретическими значениями d_1 и d_2, взятыми из таблицы. Отметим, что большинство программных пакетов статистической обработки данных осуществляет расчет именно этого критерия.

При сравнении величины d с d_1 и d_2 возможны следующие варианты (при $d < 2$): 1) если $d < d_1$, то гипотеза об отсутствии автокорреляции отвергается; 2) если $d > d_2$, то гипотеза об отсутствии автокорреляции не отвергается; 3) если $d_1 \le d \le d_2$, то нет достаточных

N	$k' = 1$		$k' = 2$		$k' = 3$	
	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
20	1,20	1,41	1,10	1,54	1,00	1,68
25	1,29	1,45	1,21	1,55	1,12	1,66
30	1,35	1,49	1,28	1,57	1,21	1,65
35	1,40	1,52	1,34	1,58	1,28	1,65

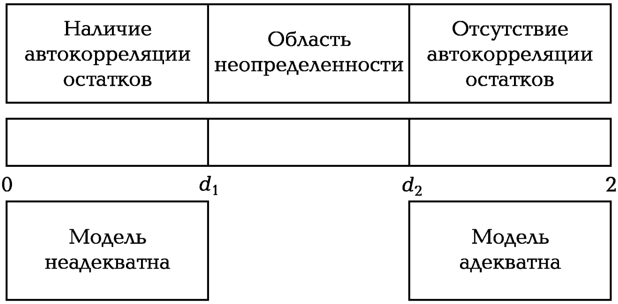


Рис. 10.7. Схема применения критерия Дарбина—Уотсона

оснований для принятия решений, ибо величина попадает в область «неопределенности».

Рассмотренные варианты относятся к случаю, когда в остатках проверяется отсутствие (наличие) положительной автокорреляции. Когда же расчетное значение d превышает 2, то проверяют существование отрицательной автокорреляции в остаточной последовательности. Для этого с критическими значениями d_1 и d_2 сравнивается не сам коэффициент d, а вспомогательная величина $4 - d$.

Таким образом, будем считать, что модель адекватна по этому критерию, если расчетное значение критерия (10.49) «не слишком отличается» от 2 (рис. 10.7).

Пример 10.6. Для прогнозирования численности промышленно-производственного персонала была выбрана модель $\hat{y}_t = a_0 \quad a_1 t$.

Оценка параметров модели осуществлялась по квартальным данным за период с 1 квартала 2008 г. по 3 квартал 2011 г.

Значение критерия Дарбина—Уотсона для ряда остатков: $d = 1{,}5$. Это значение указывает на то, что: а) модель адекватна реальному процессу по этому критерию; б) модель не адекватна реальному процессу по этому критерию; в) нет достаточных оснований для принятия решения об адекватности модели.

Решение. Из табл. 10.17 берем значения критических границ критерия Дарбина—Уотсона для $n = 15$ и $k' = 1$. Получаем $d_1 = 1{,}08$ и $d_2 = 1{,}36$.

Так как $d > d_2$ $(1{,}5 > 1{,}36)$, то гипотеза об отсутствии автокорреляции в остатках не отвергается, т.е. можно считать с доверительной вероятностью 0,95, что модель адекватна реальному процессу по данному критерию — вариант ответа а).

Важнейшими характеристиками качества модели, выбранной для прогнозирования, являются показатели ее точности. Они описывают величины случайных ошибок, полученных при использовании модели. Таким образом, чтобы судить о качестве выбранной модели, необходимо проанализировать систему показателей, характеризующих как адекватность модели, так и ее точность.

О точности можно судить по величине ошибки (погрешности) прогноза. Ошибка прогноза — величина, характеризующая расхождение между фактическим и прогнозным значением показателя. Абсолютная ошибка прогноза определяется по формуле

$$\Delta_t = \hat{y}_t \quad y_t, \tag{10.51}$$

где $\hat{y}_t$ — прогнозное значение показателя; y_t — фактическое значение.

Эта характеристика имеет ту же размерность, что и прогнозируемый показатель, и зависит от масштаба измерения уровней временно́го ряда.

На практике широко используется относительная ошибка прогноза, выраженная в процентах относительно фактического значения показателя:

$$\delta_t = \frac{\hat{y}_t - y_t}{y_t} \cdot 100. \tag{10.52}$$

Также используются средние ошибки по модулю (абсолютные и относительные):

$$\left|\overline{\Delta}\right| = \frac{\sum_{t=1}^{n}\left|\hat{y}_t - y_t\right|}{n}; \quad \left|\overline{\delta}\right| = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \cdot 100, \tag{10.53}$$

где n — число уровней временно́го ряда, для которых определялось прогнозное значение.

Из (10.51) и (10.52) видно, что если абсолютная и относительная ошибка больше 0, то это свидетельствует о «завышенной» прогнозной оценке, если — меньше 0, то прогноз был занижен. Очевидно, что все указанные характеристики могут быть вычислены лишь после того, как период упреждения уже закончился, и имеются фактические данные о прогнозируемом показателе или при рассмотрении показателя на ретроспективном участке. В последнем случае имеющаяся информация делится на две части: по первой — оцениваются параметры модели, а данные второй части рассматриваются в качестве фактических. Ошибки прогнозов, полученные ретроспективно (на втором участке), характеризуют точность применяемой модели.

На практике при проведении сравнительной оценки моделей может использоваться такая характеристика качества, как средняя квадратическая ошибка (s):

$$s = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}, \tag{10.54}$$

где n — длина временно́го ряда; y_t — фактическое значение уровня ряда; $\hat{y}_t$ — расчетное значение уровня, полученное по модели.

Часто знаменателем подкоренного выражения служит величина $(n - k)$, где k — число оцениваемых коэффициентов модели (10.48). Чем меньше значения всех рассмотренных характеристик, тем выше точность модели.

О точности модели нельзя судить лишь по значению одной ошибки прогноза. Например, если прогнозная оценка месячного уровня производства в июне совпала с фактическим значением, то это не является достаточным доказательством высокой точности модели. Надо учитывать, что единичный хороший прогноз может быть получен и по плохой модели, и наоборот.

О качестве применяемых моделей можно судить лишь по совокупности сопоставлений прогнозных значений с фактическими.

В заключение отметим, что не может быть чисто формальных подходов к выбору методов и моделей прогнозирования. Успешное применение статистических методов прогнозирования на практике возможно лишь при сочетании знаний в области самих методов с глубоким знанием объекта исследования, с содержательным анализом изучаемого явления.

Задача 10.1. Рассчитанное значение среднегодового запаса деталей на складе компании составило 1 050 ед. Это значение определялось на основе имевшихся данных о запасах деталей этого вида на складе на следующие даты: на 1 января — 1 035 ед., на 1 апреля — 1 053 ед., на 1 июля — 1 025 ед., на 1 сентября — 1 078 ед., на 1 января следующего года — ?.

Требуется восстановить пропущенное значение.

Решение. Среднегодовой запас деталей рассчитывался на основе ряда динамики, имеющего не равноотстоящие во времени уровни. Поэтому для решения целесообразно применить формулу (10.3), обозначив неизвестный уровень за x.

Среднегодовой запас деталей определялся следующим образом:

$$1\,050 = \frac{(1\,035 + 1\,053)3 + (1\,053 + 1\,025)3 + (1\,025 + 1\,078)2 + (1\,078 + x)4}{2(3 + 3 + 2 + 4)}.$$

В результате решения этого уравнения можно найти пропущенное значение:

$$4x = 1\,050 \quad 24 - 21\,016; \quad x = 1\,046 \text{ ед.}$$

Таким образом, восстановленное значение составило 1 046 деталей.

Задача 10.2. В табл. 10.18 представлена информация о динамике жилищного фонда в Московской области[1].

По данным табл. 10.18 требуется:

1) восстановить недостающие уровни временно́го ряда;

2) вычислить пропущенные значения показателей динамики;

3) рассчитать значения среднего темпа роста и среднего темпа прироста жилищного фонда Московской области.

Годы	Порядковый номер t	Жилищный фонд на конец года, млн м^2 (y_t)	Цепной абсолютный прирост, млн м^2	Цепной темп роста, %	Цепной темп прироста, %	Абсолютное значение 1 % прироста, млн м^2
2005	1		—	—	—	—
2006	2		7,2			1,646
2007	3			105,65		
2008	4				3,97	
2009	5	197,4				
2010	6		7,1			

Решение. 1. В 2006 г. абсолютное значение 1 % прироста жилищного фонда Московской области, рассчитанное как 1/100 от уровня предыдущего года, составило 1,646 млн м^2.

Согласно (10.21) уровень, соответствующий 2005 г., можно определить как

$$y_1 = 1{,}646 \quad 100 = 164{,}6 \text{ (млн м}^2\text{)}.$$

Тогда на основе (10.5), используя значение цепного абсолютного прироста для 2006 г., можно найти значение уровня 2006 г.:

[1] Задачи 10.2, 10.3 — по данным Росстата.

$$y_2 = 164{,}6 + 7{,}2 = 171{,}8 \text{ (млн м}^2\text{)}.$$

Пропущенное значение уровня для 2007 г. можно восстановить, опираясь на известное значение цепного темпа роста в соответствии с (10.11):

$$y_3 = 105{,}65 \cdot 171{,}8 / 100 = 181{,}5 \text{ (млн м}^2\text{)}.$$

Теперь с помощью представленного в табл. 10.18 значения цепного темпа прироста рассчитывается значение уровня, соответствующего 2008 г.:

$$y_4 = 181{,}5 \cdot (100 + 3{,}97) / 100 = 188{,}7 \text{ (млн м}^2\text{)}.$$

Пропущенное значение уровня для 2010 г. легко определяется на основе известного значения цепного абсолютного прироста для 2010 г.:

$$y_6 = 197{,}4 + 7{,}1 = 204{,}5 \text{ (млн м}^2\text{)}.$$

2. После восстановления недостающих уровней временно́го ряда можно рассчитать все пропущенные значения показателей динамики с помощью формул (10.5), (10.11), (10.17), (10.21). Полученные результаты расчетов представлены в табл. 10.19.

3. Средний темп роста можно определить по формуле

$$\overline{T} = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100\,\% = \sqrt[5]{\frac{204{,}5}{164{,}6}} \cdot 100\,\% = 104{,}44\,\%.$$

Таким образом, в среднем ежегодно жилищный фонд в Московской области составлял 104,44 % от уровня предыдущего года.

Годы	Порядковый номер t	Жилищный фонд на конец года, млн м^2	Цепной абсолютный прирост, млн м^2	Цепной темп роста, %	Цепной темп прироста, %	Абсолютное значение 1 % прироста, млн м^2
2005	1	164,6	—	—	—	—
2006	2	171,8	7,2	104,37	4,37	1,646
2007	3	181,5	9,7	105,65	5,65	1,718
2008	4	188,7	7,2	103,97	3,97	1,815
2009	5	197,4	8,7	104,61	4,61	1,887
2010	6	204,5	7,1	103,60	3,60	1,974

Средний темп прироста $\overline{K} = \overline{T} \quad 100 = 4{,}44\,\%$, т. е. в среднем ежегодно жилищный фонд в Московской области увеличивался на 4,44 %.

Задача 10.3. На основе данных табл. 10.20 требуется рассчитать средний темп роста жилищного фонда в России, сравнить полученное значение с аналогичным значением показателя для Московской области, вычисленным в предыдущей задаче.

Годы	2005	2006	2007	2008	2009	2010
Цепной темп роста, %	—	101,60	101,90	101,83	101,94	101,65

Решение. Согласно (10.13) средний темп роста можно рассчитать по формуле средней геометрической из цепных темпов роста:

$$\overline{T} = \sqrt[5]{1{,}016 \quad 1{,}019 \cdot 1{,}0183 \cdot 1{,}0194 \cdot 1{,}0165} = 1{,}0178, \text{ или } 101{,}78\,\%.$$

Таким образом, в анализируемый период в среднем ежегодное увеличение жилищного фонда в России составляло менее 2 % (1,78 %), что ниже соответствующего значения для жилищного фонда Московской области, равного 4,44 %. При проведении сравнительного анализа следует учитывать лидирующие позиции Московской области в сфере жилищного строительства по сравнению с другими регионами России.

Примечание. При использовании *Excel* можно упростить расчеты, применив для вычисления среднего темпа роста функцию *СРГЕОМ* (категория «Статистические»). При этом в качестве аргументов следует указать цепные темпы роста. Например, если цепные темпы роста находятся в ячейках A5:A9, то строка формул будет иметь вид

=СРГЕОМ(A5:A9)

Задача 10.4. В табл. 10.21 представлены квартальные данные о прибыли компании за четыре года. Для сглаживания колебаний требуется:

а) применить процедуру скользящих средних, приняв длину интервала сглаживания $g = 4$;

б) сравнить графически динамику исходного временно́го ряда и сглаженного.

Решение. Для вычисления значений четырехчленной скользящей средней воспользуемся формулой (10.29), тогда

$$\hat{y}_3 = \frac{0{,}5 \cdot 195{,}4 + 165{,}2 + 155{,}0 + 224{,}7 + 0{,}5 \cdot 223{,}3}{4} = 188{,}6;$$

$$\hat{y}_4 = \frac{0{,}5 \cdot 165{,}2 + 155{,}0 + 224{,}7 + 223{,}3 + 0{,}5 \cdot 183{,}6}{4} = 194{,}4 \text{ и т. д.}$$

№ года	Квартал	t	Прибыль, тыс. долл. (y_t)	Скользящая средняя ($g = 4$)
1	2	3	4	5
1	I	1	195,4	—
	II	2	165,2	—
	III	3	155,0	188,6
	IV	4	224,7	194,4
2	I	5	223,3	198,2
	II	6	183,6	202,3
	III	7	167,5	206,4
	IV	8	244,8	209,3
3	I	9	235,9	212,8
	II	10	194,2	217,2
	III	11	184,7	222,5
	IV	12	263,0	227,3
4	I	13	260,5	231,4
	II	14	208,1	236,3
	III	15	203,4	—
	IV	16	283,0	—

В гр. 5 табл. 10.21 представлены результаты расчетов четырехчленной скользящей средней, на рис. 10.8 представлена графическая иллюстрация результатов сглаживания.

В исходных данных видны значительные сезонные колебания прибыли компании в сочетании с монотонно возрастающим трендом. Сезонные «спады» характерны для второго и третьего кварталов. Ежегодные «всплески» присутствуют в значениях показателя для четвертого квартала, к ним достаточно близки уровни, соответствующие первому кварталу. Применение скользящей средней привело к сглаживанию случайных и сезонных колебаний, позволило наглядно показать характер тренда (см. рис. 10.8). Для последующего построения трендовой модели можно использовать кривые роста, например линейную модель.

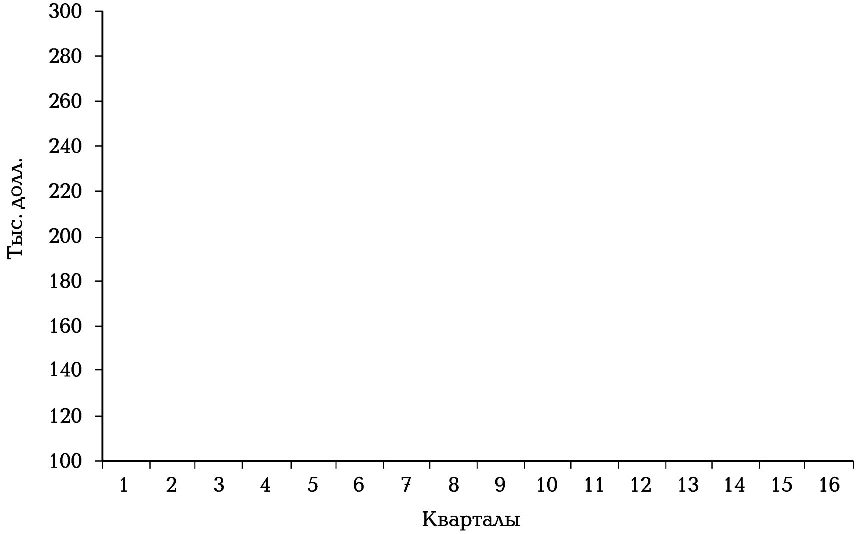


Рис. 10.8. Сглаживание временно́го ряда прибыли компании с помощью скользящей средней при $g = 4$:

фактические уровни; скользящая средняя при $g = 4$

Задача 10.5. В табл. 10.22 представлены ежемесячные данные об объемах продаж продукции фирмы. Требуется найти коэффициенты линейной трендовой модели $\hat{y}_t = a_0 \quad a_1 t$. С помощью полученной модели определить прогноз продаж продукции фирмы в следующем (16-м) месяце.

Решение. Для быстрого построения трендовых моделей целесообразно использовать имеющиеся возможности *Excel*. Для этого сначала на основе эмпирического временно́го ряда нужно построить диаграмму, выбрав, например, среди возможных тип диаграммы «*График*» (рис. 10.9). Щелкнув на диаграмме на одном из марке-

Порядковый номер месяца	Объем продаж	Порядковый номер месяца	Объем продаж	Порядковый номер месяца	Объем продаж
1	31,6	6	55,9	11	78,5
2	37,1	7	58,2	12	83,9
3	39,7	8	64,3	13	94,2
4	44,2	9	67,2	14	99,0
5	52,2	10	71,1	15	110,0

Рис. 10.9. Добавить линию тренда

ров данных правой кнопкой мыши, можно выделить сам временной ряд. При этом раскрывается контекстное меню, в котором следует выбрать команду «*Добавить линию тренда*».

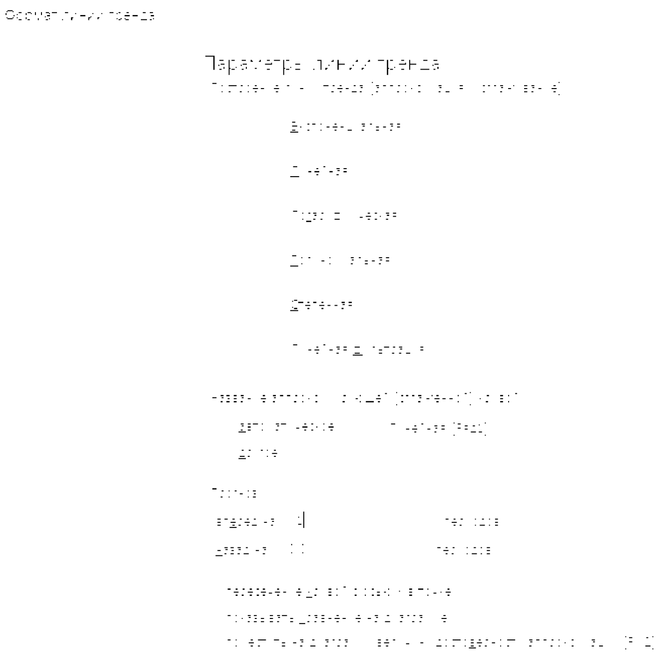


Рис. 10.10. Диалоговое окно «Формат линии тренда»

На экране появится диалоговое окно «*Формат линии тренда*» (рис. 10.10). В этом окне пользователь может выбрать требуемый тип модели и задать соответствующие параметры. На рис. 10.10 показан выбор линейной модели.

Также на диаграмме в качестве дополнительной информации можно представить уравнение модели и характеристику достоверности аппроксимации (коэффициент детерминации, R^2), установив соответствующие флажки, как показано на рис. 10.10.

На рис. 10.11 представлено уравнение линейной модели, полученное с помощью описанной схемы. Это уравнение ($\hat{y}_t = 23{,}820 \quad 5{,}248t$) в исходной системе координат (t = 1, 2, 3, ...), поэтому для расчета прогноза в модель следует подставить значение временно́го параметра $t = 16$.

Прогноз равен: $\hat{y}_{16} = 23{,}820 \quad 5{,}248 \cdot 16$; $\hat{y}_{16} = 107{,}8$ (тыс. ед.).

Следует отметить, что при использовании переноса начала координат в середину ряда и оценивании коэффициентов модели с помощью формул (10.40) будет получено уравнение линейного тренда: $\hat{y}_t = 65{,}800 \quad 5{,}248t$. В этом случае для расчета прогнозного значения следует подставить в модель $t = 8$.

При вычислении коэффициентов линейной модели также можно было воспользоваться функцией *ЛИНЕЙН* (категория «Статистические»).

Для этого сначала требуется выделить две ячейки в строке для значений оцениваемых коэффициентов модели (на рис. 10.12 это ячейки A19 и B19).

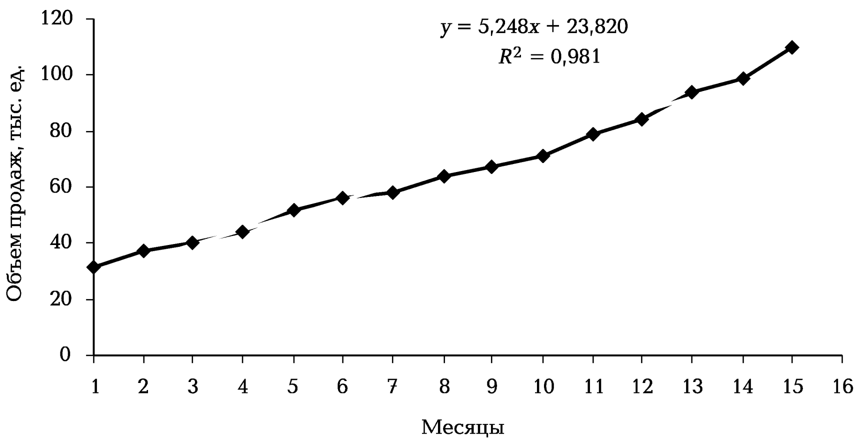
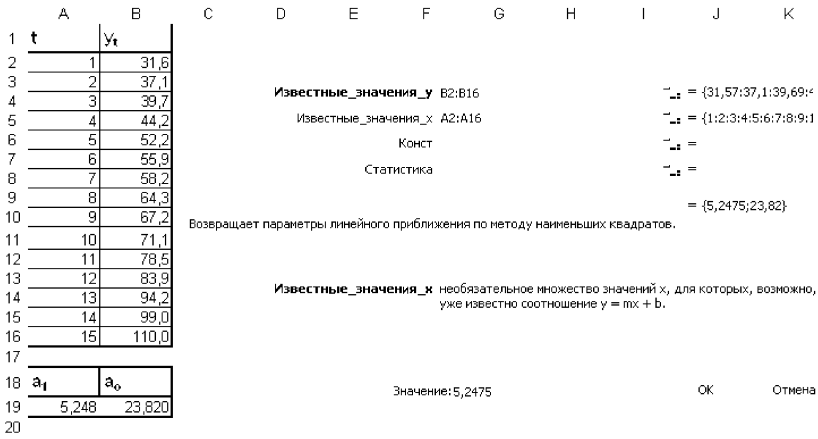


Рис. 10.11. Фактические уровни временно́го ряда и расчетные значения, полученные по линейной модели



t	y_t
1	31,6
2	37,1
3	39,7
4	44,2
5	52,2
6	55,9
7	58,2
8	64,3
9	67,2
10	71,1
11	78,5
12	83,9
13	94,2
14	99,0
15	110,0

a_1	a_0
5,248	23,820

Рис. 10.12. Применение функции *ЛИНЕЙН* для оценивания коэффициентов линейной трендовой модели

После этого следует выбрать функцию *ЛИНЕЙН*, а затем указать в первой строке диапазон ячеек, содержащих значения уровней временнóго ряда (на рис 10.12 — диапазон B2 : B16), во второй строке — диапазон ячеек, содержащих значения временнóго параметра $t = 1, 2, 3, ...15$ (на рис 10.12 — диапазон A2 : A16).

Одновременное нажатие трех клавиш на клавиатуре *Ctrl*, *Shift*, *Enter* позволит получить искомые коэффициенты модели. В ячейке B19, как показано на рис. 10.12, получено значение коэффициента a_0, в ячейке A19 представлено значение коэффициента a_1. В соответствии с полученной моделью среднемесячный прирост объемов продаж составляет примерно 5,248 тыс. ед.

Для анализа качества построенной модели целесообразно найти оценки характеристик точности, рассмотренных в подразд. 10.6, а также проверить отсутствие автокорреляции в остатках с помощью критерия Дарбина — Уотсона.

1. Представление уровней временнóго ряда в виде $y_t = u_t + s_t + \varepsilon_t$, где u_t — тренд; s_t — сезонная компонента; ε_t — случайная компонента, соответствует:
   а) мультипликативной модели;
   б) аддитивной модели;
   в) модели смешанного типа.

2. При сглаживании временно́го ряда с помощью 7-членной скользящей средней теряются:
   а) первые и последние 3 уровня;
   б) первые и последние 7 уровней;
   в) только первые 3 уровня;
   г) только первые 7 уровней.

3. Для описания процессов «с насыщением» используются следующие виды кривых роста:
   а) полином первого порядка (линейная модель);
   б) полином второго порядка (параболическая модель);
   в) показательная или экспоненциальная кривая;
   г) модифицированная экспонента.

4. Тенденция изменения численности промышленно-производственного персонала предприятия за 10 лет описывается показательной функцией: $\hat{y}_t = 431 \;\; 1{,}092^t$.
   Из этой модели следует, что среднегодовой темп роста численности составил:
   а) 109,2 %;
   б) 431 %;
   в) 92 %;
   г) 9,2 %.

5. Критерий Дарбина — Уотсона служит для:
   а) проверки наличия тенденции;
   б) проверки гипотезы о нормальном характере распределения ряда остатков;
   в) обнаружения автокорреляции в остатках.

6. Временной ряд может быть сглажен с помощью 3-, 5-, 7- или 9-членной скользящей средней. Более гладкий ряд, менее подверженный случайным колебаниям, будет получен:
   а) при сглаживании по 3-членной скользящей средней;
   б) при сглаживании по 5-членной скользящей средней;
   в) при сглаживании по 7-членной скользящей средней;
   г) при сглаживании по 9-членной скользящей средней.

7. На основе годовых данных об изменении урожайности картофеля в регионе были оценены коэффициенты линейного тренда: $\hat{y}_t = 172{,}2 \quad 4{,}418t$.
   В соответствии с этой моделью среднегодовой прирост урожайности составляет:
   а) 4,418 (ц/га);
   б) 172,2 (ц/га);
   в) (172,2 + 4,418) (ц/га);
   г) 4,418 %.

8. Темп роста вычисляется как:
   а) отношение уровней ряда;
   б) разность уровней ряда;

в) произведение уровней ряда;
г) разность темпа прироста и 100 %.

9. В таблицах приведены примеры рядов динамики.

Ряд динамики № 1. Объем продаж рекламного времени радиостанцией за 6 нед:

Показатель	Текущий номер недели					
	1	2	3	4	5	6
Проданное рекламное время, мин	125	922	125	238	264	82

Ряд динамики № 2. Среднемесячная заработная плата рабочих предприятия (руб.):

Показатель	Месяц					
	январь	февраль	март	апрель	май	июнь
Средняя заработная плата рабочих, руб.	9 570	10 900	11 950	11 200	13 100	16 000

Ряд динамики № 3. Цены акций промышленной компании на момент открытия торгов (долл.):

Показатель	Дата					
	21.11.11	22.11.11	23.11.11	24.11.11	25.11.11	28.11.11
Цены акций, долл.	280	291	287	289	294	286

К интервальным рядам динамики следует отнести:
а) ряд динамики № 1;
б) ряд динамики № 2;
в) ряд динамики № 3;
г) пример интервального ряда динамики отсутствует.

10. Для описания периодических колебаний, имеющих 3-месячный период, используется:
а) сезонная компонента;
б) случайная компонента;
в) трендовая компонента;
г) циклическая компонента.

11. Если значения цепных абсолютных приростов примерно одинаковы, то для вычисления прогнозного значения показателя в следующей точке корректно использовать:
а) средний абсолютный прирост;
б) средний темп роста;
в) средний темп прироста.

12. Данные об изменении урожайности озимой пшеницы (y_t) за 10 лет представлены в таблице (ц/га):

t	1	2	3	4	5	6	7	8	9	10
y_t	16,3	21,2	18,1	8,7	16,3	17,3	20,9	15,4	19,7	21,7

Проведите сглаживание временно́го ряда, используя 3-летнюю скользящую среднюю. В ответе укажите расчетные значения второго и девятого уровней ряда. Точность ответа — один знак после запятой. **Ответ:** 18,5; 18,9.

13. По данным задания 12 проведите сглаживание по 5-членной взвешенной скользящей средней, используя табл. 10.12. В ответе укажите расчетное значение третьего уровня ряда. Точность ответа — один знак после запятой. **Ответ:** 16,2.

14. В таблице представлены данные об объеме производства продукции (млн руб.) в течение 6 кварталов:

t	1	2	3	4	5	6
y_t	11,18	12,23	13,28	14,31	15,36	16,40

Рассчитайте средний абсолютный прирост. Точность ответа — три знака после запятой. **Ответ:** 1,044 (млн руб.).

15. По данным задания 14 рассчитайте прогноз производства в 8 квартале с помощью среднего абсолютного прироста. Точность ответа — три знака после запятой. **Ответ:** 18,488 (млн руб.).

16. Ежеквартальная динамика процентной ставки банка в течение 5 кварталов представлена в таблице:

t	1	2	3	4	5
y_t (%)	7,3	8,0	8,8	9,7	10,7

Вычислите для приведенных данных средний темп роста и средний темп прироста. Точность ответа — два знака после запятой. **Ответ:** 110,03 %; 10,03 %.

17. С помощью среднего темпа роста рассчитайте прогноз процентной ставки банка в 6 и 7 кварталах (по данным задания 16). Точность ответа — один знак после запятой. **Ответ:** 11,8 %; 13,0 %.

18. Динамика прибыли производственного объединения представлена в таблице:

t	1	2	3	4	5	6	7
y_t (млн руб.)	70	76	83	88	94	101	106

Рассчитайте коэффициенты линейного тренда $\hat{y}_t = a_0 \quad a_1 t$. В ответе укажите значение коэффициента a_1. Точность ответа — четыре знака после запятой. **Ответ:** 6,0357.

19. По данным задания 18 на основе линейной модели определите прогноз прибыли производственного объединения в точке $t = 8$ (время упреждения прогноза $i = 1$). Точность ответа — один знак после запятой. **Ответ:** 112,4 (млн руб.).

20. Ежеквартальная динамика объемов реализованной продукции объединения представлена в таблице:

t	1	2	3	4	5	6	7	8	9
y_t (млн руб.)	95	100	108	113	124	135	149	160	169

Оцените коэффициенты параболического тренда $\hat{y}_t = a_0 + a_1 t + a_2 t^2$. В ответе укажите значение коэффициента a_2. Точность ответа — четыре знака после запятой. **Ответ:** 0,4978.

21. На основе параболической модели, полученной в задании 20, определите прогноз реализованной продукции объединения в точке $t = 10$ (время упреждения прогноза $i = 1$). Точность ответа — один знак после запятой. **Ответ:** 185,6 (млн руб.).

22. Для описания динамики численности промышленно-производственного персонала предприятия выбрана модель вида $\hat{y}_t = a_0 + a_1 t$. Длина временно́го ряда $n = 20$. Значение критерия Дарбина — Уотсона для ряда остатков $d = 2{,}7$. Значение d указывает на то, что: а) модель неадекватна реальному процессу по данному критерию; б) модель адекватна реальному процессу по данному критерию; в) нет достаточных оснований для принятия решения об адекватности модели. Уровень значимости 0,05. **Ответ:** (в).

23. Анализ динамики числа собственных автомобилей на 1 000 чел. населения региона показал: в 2005 г. значение показателя составило 123 ед.; в 2006 г. значение цепного темпа роста составило 109,35 %; в 2007 г. значение цепного темпа прироста 11,9 %; в 2008 г. значение цепного абсолютного прироста 19,6 ед. Восстановите значения уровней исследуемого временного ряда в 2006 — 2008 гг.

24. Рассчитанное значение среднегодового запаса готовой продукции предприятия составило 960 ед. Это значение определялось на основе имевшихся данных о запасах готовой продукции на складе предприятия на следующие даты (ед.): на 1 января — 925, на 1 марта — ?,

на 1 июня — 945, на 1 сентября — 960, на 1 января следующего года — 975. Восстановите пропущенное значение. **Ответ:** 980 ед.

25. В таблице представлены цепные темпы роста (%) инвестиций в основной капитал предприятия (в сопоставимых ценах). На основе имеющихся данных рассчитайте средний годовой темп роста и средний годовой темп прироста инвестиций в основной капитал предприятия в исследуемый период.

Показатель	Порядковый номер года				
	1	2	3	4	5
Цепной темп роста, %	—	102,55	102,70	102,30	101,90

Ответ: 102,36 %; 2,36 %.

26. Динамика импорта оборудования в компании представлена в таблице:

Показатель	Порядковый номер года				
	1	2	3	4	5
Импорт оборудования, тыс. долл.	581	595	610	622	635

Рассчитайте цепные, базисные и средние:

а) абсолютные приросты;

б) темпы роста;

в) темпы прироста.

Базисный уровень — первый уровень временно́го ряда.

27. Анализ динамики добычи естественного газа в России за пять лет показал, что средний годовой абсолютный прирост был равен 15 млрд $м^3$, во второй исследуемый год абсолютное значение 1 % прироста составило 5,81 млрд $м^3$. Определите средний годовой темп роста и средний годовой темп прироста добычи естественного газа в России в исследуемом периоде.
Ответ: 102,49 %; 2,49 %.

28. Анализ динамики экспорта продукции компании в 2007—2010 гг. показал, что в 2008 г. экспорт продукции увеличился на 20 %, или на 100 тыс. долл. по сравнению с предыдущим годом, при этом в 2009—2010 гг. экспорт продукции возрастал на 5 % ежегодно. Определите:

а) значения экспорта продукции компании в 2007—2010 гг.;

б) цепной абсолютный прирост экспорта продукции компании в 2010 г.;

в) базисный темп роста экспорта продукции компании в 2010 г., приняв за базисный — уровень 2007 г.

29. В таблице представлена динамика числа собственных легковых автомобилей на 1 000 чел. населения города:

Порядковый номер года	Число собственных легковых автомобилей на 1 000 чел. населения, значение за год, ед.	Порядковый номер года	Число собственных легковых автомобилей на 1 000 чел. населения, значение за год, ед.
1	131	7	156,4
2	134,9	8	164,6
3	137,7	9	176,7
4	142,8	10	190,5
5	146,7	11	209,1
6	152,8		

Оцените коэффициенты параболического тренда $\hat{y}_t = a_0 \quad a_1 t + a_2 t^2$. В ответе укажите значение коэффициента a_2. Точность ответа — три знака после запятой.
Ответ: 0,742.

30. На основе параболической модели, полученной в задании 29, определите прогноз числа собственных легковых автомобилей на 1 000 человек населения города для следующего года (время упреждения прогноза $i = 1$).

31. В таблице представлены квартальные данные об объемах продаж продукции фирмы:

Порядковый номер квартала	Объем продаж, тыс. ед.	Порядковый номер квартала	Объем продаж, тыс. ед.	Порядковый номер квартала	Объем продаж, тыс. ед.
1	846	6	932	11	1058
2	878	7	953	12	1063
3	891	8	996	13	1095
4	905	9	1005	14	1125
5	926	10	1032	15	1155

Требуется:

а) найти коэффициенты двух трендовых моделей: линейной ($\hat{y}_t = a_0 + a_1 t$) и параболической ($\hat{y}_t = a_0 + a_1 t + a_2 t^2$);

б) провести сравнительный анализ характеристик точности полученных моделей, опираясь на значения средних абсолютных и относительных ошибок по модулю (см. формулу (10.53));
в) с помощью критерия Дарбина—Уотсона (см. формулу (10.49)) проверить гипотезу об отсутствии автокорреляции первого порядка в остатках;
г) рассчитать прогноз объема продаж в следующем (16-м) квартале на основе выбранной модели.

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

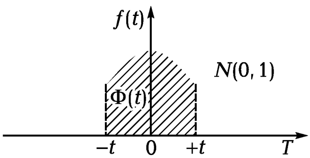
Методические указания к использованию таблиц

Таблица П.1

В табл. П.1 протабулирована функция (интеграл вероятностей) Лапласа:

$$\Phi(t) = \int_{-t}^{t} f(x)dx \quad \frac{2}{\sqrt{2\pi}}\int_{0}^{t} e^{-\frac{x^2}{2}}dx,$$

где $f(t)$ — плотность нормированной (стандартной) нормально распределенной случайной величины $T \in N(0,1)$.



Функция $\Phi(t)$ обладает следующими свойствами:

1) $\Phi(t)$ — нечетная функция: $\Phi(-t) = -\Phi(t)$;

2) $\Phi(t)$ — монотонно возрастающая функция: при $t \to +\infty$ $\Phi(t) \to 1$; $\Phi(+\infty) = 1$.

Вероятность попадания случайной величины T в интервал от t_1 до t_2 вычисляется по формуле

$$P(t_1 < T < t_2) = \frac{1}{2}[\Phi(t_2) \quad \Phi(t_1)]. -$$

Пример П.1. Пусть $T \in N(0,1)$. Требуется определить $P(-1{,}36 < T < 2{,}15)$.

На пересечении строки, соответствующей t = 1,3, и столбца, соответствующего 6 сотым долям t, находим в середине табл. П.1 значение функции Лапласа. Это $\Phi(t) = 0{,}8262$. Таким образом, с учетом нечетности функции Лапласа, $t_1 = -1{,}36$ соответствует $\Phi(t_1) = \Phi(-1{,}36) = -\Phi(1{,}36) = -0{,}8262$.

Аналогично $t_2 = 2{,}15$ соответствует $\Phi(t_2) = \Phi(2{,}15) = 0{,}9684$.

$$P(-1{,}36 < T < 2{,}15) = \frac{1}{2}\left[\Phi(2{,}15) - \Phi(-1{,}36)\right] = \frac{1}{2}[0{,}9684 \quad 0{,}8262] = +0{,}8973.$$

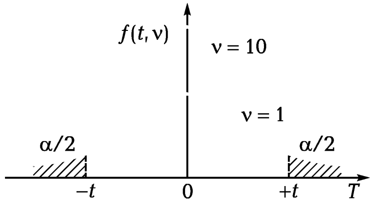
Пример П.2. Дано значение функции Лапласа $\Phi(t)$ = 0,98. Требуется найти t.

В середине табл. П.1 находятся значения функции Лапласа $\Phi(t)$. Находим ближайшее к 0,98 значение, это $\Phi(t) = 0{,}9802$. Значит, соответствующее значение t будет равно 2,33.

Таблица П.2

В табл. П.2 протабулирована вероятность выхода за пределы интервала от $-t$ до $+t$ случайной величины, имеющей распределение Стьюдента (t-распределение) с числом степеней свободы ν:

$$\alpha = St(t;\nu) \quad P(|T| > t);$$



$f(t;\nu)$ — плотность распределения Стьюдента с числом степеней свободы ν.

Функция $St(t)$ обладает следующими свойствами:

1) $St(-t) = 2 - St(t)$;

2) $St(\infty) = 0$; $St(-\infty) = 2$; $St(0) = 1$.

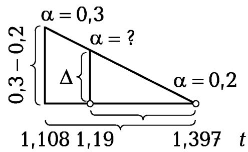
Вероятность попадания случайной величины T в интервал от t_1 до t_2 вычисляется по формуле

$$P(t_1 < T < t_2) = \frac{1}{2}\left[St(t_1) \quad St(t_2)\right].-$$

Пример П.3. Для заданной вероятности $\alpha = 0{,}05$ и числа степеней свободы $\nu = 12$ найти соответствующее значение t.

На пересечении строки табл. П.2, соответствующей $\nu = 12$, и столбца, в котором $\alpha = 0{,}05$, находим значение $t = 2{,}179$.

Пример П.4. Определить, какой вероятности α соответствует найденное значение $t = 1{,}19$ при числе степеней свободы $\nu = 8$.



В строке табл. П.2, соответствующей заданному числу степеней свободы $\nu = 8$, находим ближайшее значение к $t = 1{,}19$ — это $t = 1{,}108$, что соответствует вероятности $\alpha = 0{,}3$.

Более точно значение вероятности α можно найти методом линейной интерполяции:

$$\frac{\Delta}{0{,}3 - 0{,}2} = \frac{1{,}397 - 1{,}19}{1{,}397 - 1{,}108} \;\Rightarrow\; \Delta \approx 0{,}07 \;\Rightarrow\; \alpha = 0{,}2 + \Delta \approx 0{,}27.$$

Пример П.5. При $\nu = 10$ определить $P(-1{,}36 < T < 2{,}15)$.

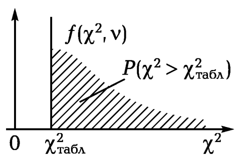
$$P(-1{,}36 < T < 2{,}15) = \frac{1}{2}\left[St(-1{,}36) - St(2{,}15)\right] =$$

$$= \frac{1}{2}[2 \quad St(1{,}36) - St(2{,}15)] \cong \frac{1}{2}[2 - St(1{,}372) - St(2{,}228)] = \frac{1}{2}[2 \quad 0{,}2 - 0{,}05] = 0{,}875.$$

Чтобы не прибегать к интерполяции, в строке, соответствующей $\nu = 10$, находим ближайшие к заданным значения.

Таблица П.3

В табл. П.3 протабулирована вероятность того, что наблюдаемое значение случайной величины χ^2, имеющей распределение Пирсона (хи-квадрат распределение) с числом степеней свободы ν, превысит табличное значение $\chi^2_{\text{табл}}$.



На рисунке представлен график функции $f(\chi^2, \nu)$ — плотности χ^2-распределения с числом степеней свободы ν.

Вероятность попадания случайной величины χ^2 в интервал от χ_1^2 до χ_2^2 вычисляется по формуле

$$P(\chi_1^2 < \chi^2 < \chi_2^2) = P(\chi^2 \quad \chi_1^2) \Rightarrow P(\chi^2 > \chi_2^2) = P(\chi_1^2) \quad P(\chi_2^2).$$

Функция $P(\chi^2)$ обладает следующими свойствами:

$$P(0) = P(\chi^2 > 0) = 1; \; P(+\infty) = P(\chi^2 \to +\infty) = 0.$$

Пример П.6. При числе степеней свободы $\nu = 8$ и вероятностях, равных $P(\chi_1^2) = 0{,}975$ и $P(\chi_2^2) = 0{,}025$, найти соответствующие значения χ_1^2 и χ_2^2.

На пересечении строки табл. П.3, соответствующей $\nu = 8$, и столбца, в котором $P(\chi^2_{\text{табл}}) = 0{,}975$, находим значение $\chi_1^2 = 2{,}180$. Аналогично на пересечении строки с $\nu = 8$, и столбца, в котором $P(\chi^2_{\text{табл}}) = 0{,}025$, находим значение $\chi_2^2 = 17{,}535$.

Пример П.7. При числе степеней свободы $\nu = 12$ и значениях $\chi_1^2 = 3{,}156$ и $\chi_2^2 = 20{,}48$ найти соответствующие им вероятности $P(\chi_1^2)$ и $P(\chi_2^2)$.

В строке табл. П.3, соответствующей заданному числу степеней свободы $\nu = 12$, находим ближайшее значение к $\chi_1^2 = 3{,}156$ — это $\chi_1^2 = 3{,}074$, что соответствует вероятности $P(\chi_1^2) = 0{,}995$. Аналогично в строке с $\nu = 12$ находим ближайшее значение к $\chi_2^2 = 20{,}48$ — это $\chi_2^2 = 21{,}026$, что соответствует вероятности $P(\chi_2^2) = 0{,}05$.

Более точно значение вероятностей $P(\chi^2)$ можно найти методом линейной интерполяции, приведенным в описании табл. П.2.

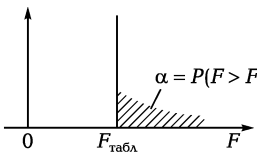
Пример П.8. Определить при числе степеней свободы $\nu = 10$ вероятность попадания случайной величины χ^2 в интервал от 2,5 до 19,0.

$$P(2{,}5 < \chi^2 < 19{,}0) = P(2{,}5) \quad P(19{,}0) \cong P(2{,}558) - P(18{,}307) = 0{,}99 \quad 0{,}05 = 0{,}94.$$

Чтобы не прибегать к интерполяции, в строке таблицы, соответствующей $\nu = 10$, берем значения, ближайшие к заданным.

Таблица П.4

В табл. П.4 для случайной величины F, имеющей закон распределения Фишера — Снедекора (F-распределение) с числами степеней свободы числителя ν_1 и знаменателя ν_2, протабулированы три значения, соответствующие трем вероятностям (уровням значимости): $\alpha = P(F > F_{\text{табл}}) = 0{,}05$ (верхнее значение); $\alpha = 0{,}01$ (среднее значение) и $\alpha = 0{,}001$ (нижнее значение).



Статистика F строится таким образом, чтобы наблюдаемое значение было не меньше единицы (числитель должен быть больше знаменателя).

Пример П.9. Найти значение F-распределения, соответствующее уровню значимости $\alpha = 0{,}01$ и числам степеней свободы числителя $\nu_1 = 5$ и знаменателя $\nu_2 = 7$.

На пересечении столбца табл. П.4, соответствующего числу степеней свободы числителя $\nu_1 = 5$, и строки, соответствующей числу степеней свободы знаменателя $\nu_2 = 7$, находим три значения F-распределения и выбираем среднее из них, соответствующее $\alpha = 0{,}01$. Таким образом, $F_{\text{табл}} = 7{,}46$.

Таблица П.5

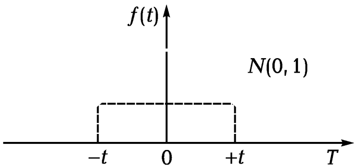
В табл. П.5 протабулирована функция плотности вероятностей (функция Гаусса) $f(t)$ нормированной (стандартной) нормально распределенной случайной величины $T \in N(0,1)$

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

Функция $f(t)$ обладает следующими свойствами:

1) $f(t)$ — четная функция: $f(-t) = f(t)$. Функция $f(t)$ симметрична относительно оси ординат;

2) $f(t)$ — монотонно убывающая функция: при $t \to \pm\infty$ $f(t) \to 0$; $f(+\infty) = 0$.



Пример П.10. Пусть $T \in N(0,1)$ и $t = 1{,}53$. Требуется определить $f(t)$.

На пересечении строки, соответствующей $t = 1{,}5$ и столбца, соответствующего 3 сотым долям t, находим в середине табл. П.5 значение функции Гаусса $f(t) = 0{,}1238$.

Пример П.11. Пусть $T \in N(0,1)$ и $f(t) = 0{,}31$. Требуется определить t.

В середине табл. П.5 находим значение функции Гаусса, ближайшее к заданному 0,31. Это $f(t) = 0{,}3101$. Теперь смотрим, какому t это соответствует, т.е. на пересечении каких строки и столбца находится найденное значение $f(t)$. Таким образом, получаем $t = 0{,}71$.

Таблица П.6

В табл. П.6 Фишера — Иейтса приведены критические значения коэффициента корреляции $r_{\text{кр}}(\alpha;\nu)$ для проверки значимости генеральных парных и частных коэффициентов корреляции по выборочным.

В таблице приведены значения для четырех уровней значимости: $\alpha = 0{,}05$; $\alpha = 0{,}02$; $\alpha = 0{,}01$ и $\alpha = 0{,}001$.

Число степеней свободы ν зависит от объема выборки n и вида коэффициента корреляции: $\nu = n - 2$ в случае парной корреляции и $\nu = n - l - 2$, где l — число исключенных величин, в случае частной корреляции.

Пример П.12. Найти критические значения коэффициента корреляции $r_{\text{кр}}(\alpha;\nu)$ для проверки значимости парных и частных коэффициентов корреляции в случае трехмерной модели (X, Y, Z) на уровне значимости $\alpha = 0{,}02$, если число наблюдений, по которым рассчитаны выборочные коэффициенты корреляции, равно $n = 23$.

При проверке значимости *парных* коэффициентов корреляции ρ_{xy}, ρ_{yz}, ρ_{xz} число степеней свободы $\nu = n - 2 = 21$ значение $r_{\text{кр}}(\alpha = 0{,}02;\ \nu = 21)$ находится на пересечении строки табл. П.6, соответствующей $\nu = 21$, и столбца со значением $\alpha = 0{,}02$ в верхней части табл. П.6 (для двусторонних границ). Значение $\nu = 21$ отсутствует, поэтому методом линейной интерполяции находим недостающее значение $r_{кр}$ для $\nu = 21$ между $r_{кр} = 0{,}492$ (для $\nu = 20$) и $r_{кр} = 0{,}445$ (для $\nu = 25$). Повышение степени свободы на 1 соответствует понижению $r_{кр}$ на $(0{,}492 - 0{,}445)/5 = 0{,}0094$; следовательно, $r_{кр}(\nu = 21) = 0{,}492 - 0{,}0094 = 0{,}4826$. Таким образом, для парных коэффициентов корреляции $r_{кр} = 0{,}4826$.

При проверке значимости *частных* коэффициентов корреляции $\rho_{xy/z}$, $\rho_{yz/x}$, $\rho_{xz/y}$ число степеней свободы $\nu = n - l - 2 = 23 - 1 - 2 = 20$ ($l = 1$ в случае трехмерной модели, фиксируется одна переменная). $r_{кр}(\alpha = 0{,}02;\ \nu = 20)$ находится на пересечении строки таблицы, в которой $\nu = 20$, и столбца со значением $\alpha = 0{,}02$ в верхней части таблицы (для двусторонних границ). Таким образом, для частных коэффициентов корреляции $r_{кр} = 0{,}492$.

Таблица П.7

В табл. П.7 приведены значения прямого и обратного Z-преобразования Фишера, соответствующие гиперболическим арктангенсам и гиперболическим тангенсам заданных чисел:

$$Z_r = \frac{1}{2}\ln\frac{1+r}{1-r} = \operatorname{arcth}(r); \quad r = \operatorname{th}(Z) = \frac{e^{2Z} - 1}{e^{2Z} + 1}.$$

Эти преобразования используются в статистике при расчете интервальных оценок генеральных коэффициентов корреляции. Для значимых выборочных парных и частных коэффициентов корреляции можно построить с заданной надежностью γ доверительный интервал $\rho_{\min} \leqslant \rho \leqslant \rho_{\max}$ с помощью Z-преобразования Фишера:

$$Z_r - \frac{t_\gamma}{\sqrt{n-l-3}} \le MZ \le Z_r + \frac{t_\gamma}{\sqrt{n-l-3}}; \quad Z^{-1}(Z_{\min}) \le \rho \le Z^{-1}(Z_{\max}).$$

Пример П.13. Для выборки из 15 наблюдений двумерной нормальной совокупности (X, Y) получено, что выборочный коэффициент корреляции равен $r = -0{,}85$. С надежностью $\gamma = 0{,}95$ построить доверительный интервал для генерального коэффициента корреляции $\rho_{\min} \leqslant \rho \leqslant \rho_{\max}$.

1. По найденному выборочному коэффициенту корреляции $r = 0{,}85$ с помощью табл. П.7 находим соответствующее значение Z_r на пересечении строки, в которой $r = 0{,}8$, и столбца, соответствующего 5 сотым долям r. Получаем $Z_r = -1{,}2562$ ($Z(-r) = -Z(r)$, так как $Z(r)$ — функция нечетная).

2. Вычисляем $\Delta Z = \dfrac{t_\gamma}{\sqrt{n-3}} = \dfrac{1{,}96}{\sqrt{15\text{-}3}}$ 0,5658 (значение $t_\gamma = 1{,}96$, соответствующее заданной надежности $\gamma = 0{,}95$, находим по табл. П.1 функции Лапласа для $\Phi(t_\gamma) = \gamma = 0{,}95$).

Затем рассчитываем $Z_{\min} = Z_r - \Delta Z = -1{,}2562 - 0{,}5658 = -1{,}822$; $Z_{\max} = Z_r + \Delta Z = -1{,}2562 + 0{,}5658 = -0{,}6904$.

3. Соответствующие значения $\rho_{\min}$ и $\rho_{\max}$ находятся с помощью обратного преобразования Фишера следующим образом. В середине табл. П.6 находим ближайшие к рассчитанным $Z_{\min}$ и $Z_{\max}$ значения и определяем, каким коэффициентам корреляции они соответствуют.

$Z_{min} = -1,822$; в середине табл. П.7 находим ближайшее значение, это $Z = 1,8318$. Это соответствует коэффициенту корреляции, равному 0,95 (смотрим, на пересечении каких строки и столбца находится найденное значение Z). Следовательно, нижняя граница генерального коэффициента корреляции $\rho_{min} = -0,95$.

Аналогично для $Z_{max} = -0,6904$ в середине табл. П.7 находим ближайшее значение, это $Z = 0,6932$. Это соответствует коэффициенту корреляции, равному 0,60. Следовательно, верхняя граница генерального коэффициента корреляции $\rho_{max} = -0,60$. (Более точно значения ρ_{min} и ρ_{max} можно получить, рассчитав по формуле гиперболического тангенса, приведенной выше.) Таким образом, $P(-0,95 \leqslant \rho \leqslant -0,60) = \gamma = 0,95$.

Таблица П.8

В табл. П.8 протабулированы значения функции Пуассона вероятности того, что в n повторных независимых испытаниях событие A наступит ровно m раз. Формула Пуассона используется при вероятности наступления события p, близкой к нулю, и параметре $\lambda = np$, находящемуся в пределах $0,1 \leqslant \lambda \leqslant 10$:

$$P_{n,m} = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Пример П.14. Событие A происходит в каждом из испытаний с вероятностью $p = 0,005$. Найти вероятность, что при проведении 1 000 испытаний событие A наступит 3 раза (т. е. $n = 1\,000$; $m = 3$; $p = 0,005$; $P_{1000}(X = 3) = ?$).

$\lambda = np = 1\,000 \cdot 0,005 = 5$. Искомую вероятность находим в табл. П.8 на пересечении строки, соответствующей $m = 3$, и столбца, в котором $\lambda = 5$. Таким образом, $P_{1000}(X = 3) = 0,1404$.

Таблица П.9

В табл. П.9 приведены значения G-распределения, которые используются в статистике при проверке гипотезы об однородности ряда дисперсий k генеральных совокупностей с помощью критерия Кохрана. Значения G-распределения приводятся в таблице для двух уровней значимости. Первое значение соответствует уровню значимости $\alpha = 0,05$, а второе — $\alpha = 0,01$.

Кроме того, значение G зависит от числа степеней свободы $\nu = n - 1$ (где n — объем независимых выборок из каждой совокупности) и количества рассматриваемых совокупностей k, для которых проверяется равенство генеральных дисперсий.

Пример П.15. Для проверки гипотезы о равенстве генеральных дисперсий четырех совокупностей осуществили независимые выборки по 10 наблюдений в каждой. Критерий Кохрана проверяется на уровне значимости $\alpha = 0,05$. Найти соответствующее значение G-распределения $G_{кр}$.

На пересечении строки табл. П.9, соответствующей $k = 4$, и столбца, в котором число степеней свободы $\nu = n - 1 = 10 - 1 = 9$, находятся два значения G-распределения. Берем верхнее из них, соответствующее заданному уровню значимости $\alpha = 0,05$. Таким образом, $G_{кр} = 0,502$.

Целые и десятичные 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	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1663	0,1741	0,1819	0,1897	0,1974	0,2051	0,2128	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2960	0,3035
0,4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3688	0,3759
0,5	0,3829	0,3899	0,3969	0,4039	0,4108	0,4177	0,4245	0,4313	0,4381	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,9	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6679	0,6729	0,6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7994	0,8029
1,3	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8638
1,5	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882

1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,8	0,9281	0,9297	0,9312	0,9327	0,9342	0,9357	0,9371	0,9385	0,9399	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9841	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928
2,7	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9951	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,5	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997
3,6	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	0,9998
3,7	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,8	0,999900	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	0,999900	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
4,0	0,999936	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
4,5	0,999994	—	—	—	—	—	—	—	—	—
5,0	0,99999994	—	—	—	—	—	—	—	—	—

ν	Вероятность $\alpha = St(t) = P(\lvert T \rvert > t_{табл})$												
	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	4,043	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,953	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	0,397	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	0,858	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

ν	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	6,737	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	19,820	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	27,336
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

ν	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

ν_1 \ ν_2	1	2	3	4	5	6	8	12	24	∞	*t*
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	253,3	12,71
	4 052	4 999	5 403	5 625	5 764	5 859	5 981	6 106	6 234	6 366	63,66
	406 523	500 016	536 700	562 527	576 449	585 953	598 149	610 598	623 432	636 535	636,2
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50	4,30
	98,49	99,01	00,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50	9,92
	998,46	999,00	999,20	999,20	999,20	999,20	999,40	999,60	999,40	999,40	31,00
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53	3,18
	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12	5,84
	67,47	148,51	141,10	137,10	134,60	132,90	130,60	128,30	125,90	123,50	12,94
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63	2,78
	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46	4,60
	74,13	61,24	56,18	53,43	51,71	50,52	49,00	47,41	45,77	44,05	8,61
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36	2,57
	16,26	13,27	12,06	11,39	10,97	10,67	10,27	9,89	9,47	9,02	4,03
	47,04	36,61	33,20	31,09	20,75	28,83	27,64	26,42	25,14	23,78	6,86
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67	2,45
	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88	3,71
	35,51	26,99	23,70	21,90	20,81	20,03	19,03	17,99	16,89	15,75	5,96

7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23	2,36
	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65	3,50
	29,22	21,69	18,77	17,19	16,21	15,52	14,63	13,71	12,73	11,70	5,40
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,99	2,31
	11,26	8,65	7,59	7,10	6,63	6,37	6,03	5,67	5,28	4,86	3,36
	25,42	18,49	15,83	14,39	13,49	12,86	12,04	11,19	10,30	9,35	5,04
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71	2,26
	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31	3,25
	22,86	16,39	13,90	12,56	11,71	11,13	10,37	9,57	8,72	7,81	4,78
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54	2,23
	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91	3,17
	21,04	14,91	12,55	11,28	10,48	9,92	9,20	8,45	7,64	6,77	4,59
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40	2,20
	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60	3,11
	19,69	13,81	11,56	10,35	9,58	9,05	8,35	7,62	6,85	6,00	4,49
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30	2,18
	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36	3,06
	18,64	12,98	10,81	9,63	8,89	8,38	7,71	7,00	6,25	5,42	4,32
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21	2,16
	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16	3,01
	17,81	12,31	10,21	9,07	8,35	7,86	7,21	6,52	5,78	4,97	4,12

Продолжение табл. П.4

ν_1 \ ν_2	1	2	3	4	5	6	8	12	24	∞	t
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13	2,14
	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00	2,98
	17,14	11,78	9,73	8,62	7,92	7,44	6,80	6,13	5,41	4,60	4,14
15	4,45	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07	2,13
	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87	2,95
	16,59	11,34	9,34	8,25	7,57	7,09	6,47	5,81	5,10	4,31	4,07
16	4,41	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01	2,12
	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75	2,92
	16,12	10,97	9,01	7,94	7,27	6,80	6,20	5,55	4,85	4,06	4,02
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96	2,11
	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65	2,90
	15,72	10,66	8,73	7,68	7,02	6,56	5,96	5,32	4,63	3,85	3,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92	2,10
	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,01	2,57	2,88
	15,38	10,39	8,49	7,46	6,81	6,35	5,76	5,13	4,45	3,67	3,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88	2,09
	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49	2,86
	15,08	10,16	8,28	7,26	6,61	6,18	5,59	4,97	4,29	3,52	3,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84	2,09
	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42	2,84
	14,82	9,95	8,10	7,10	6,46	6,02	5,44	4,82	4,15	3,38	3,85

21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,82	2,08
	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36	2,83
	14,62	9,77	7,94	6,95	6,32	5,88	5,31	4,70	4,03	3,26	3,82
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78	2,07
	7,94	5,72	4,82	4,31	3,99	3,75	3,45	3,12	2,75	2,30	2,82
	14,38	9,61	7,80	6,81	6,19	5,76	5,19	4,58	3,92	3,15	3,79
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76	2,07
	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26	2,81
	14,19	9,46	7,67	6,70	6,08	5,56	5,09	4,48	3,82	3,05	3,77
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73	2,06
	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21	2,80
	14,03	9,34	7,55	6,59	5,98	5,55	4,99	4,39	3,74	2,97	3,75
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71	2,06
	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17	2,79
	13,88	9,22	7,45	6,49	5,89	5,46	4,91	4,31	3,66	2,89	3,72
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69	2,06
	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13	2,78
	13,74	9,12	7,36	6,41	5,80	5,38	4,83	4,24	3,59	2,82	3,71
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67	2,05
	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10	2,77
	13,61	9,02	7,27	6,33	5,73	5,31	4,76	4,17	3,52	2,76	3,69

ν_1 \ ν_2	1	2	3	4	5	6	8	12	24	∞	*t*
28	4,19	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65	2,05
	7,64	5,54	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06	2,76
	13,50	8,93	7,18	6,25	5,66	5,24	4,69	4,11	3,46	2,70	3,67
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64	2,05
	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03	2,76
	13,39	8,85	7,12	6,19	5,59	5,18	4,65	4,05	3,41	2,64	3,66
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62	2,04
	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01	2,75
	13,29	8,77	7,05	6,12	5,53	5,12	4,58	4,00	3,36	2,59	3,64
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39	2,00
	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60	2,66
	11,97	7,76	6,17	5,31	4,76	4,37	3,87	3,31	2,76	1,90	3,36
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,03	1,96
	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,04	2,58
	10,83	6,91	5,42	4,62	4,10	3,74	3,27	2,74	2,13	1,05	3,29

[1] 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; ν_1 — число степеней свободы числителя; ν_2 — знаменателя.

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	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3525	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3256	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2631	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,3012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127

Окончание табл. П.5

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0762	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0568	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025

3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

[1]

ν	Двусторонние границы			
	0,05	0,02	0,01	0,001
1	0,997	1,000	1,000	1,000
2	0,950	0,980	0,990	0,999
3	0,878	0,934	0,959	0,991
4	0,811	0,882	0,917	0,974
5	0,754	0,833	0,875	0,951
6	0,707	0,789	0,834	0,925
7	0,666	0,750	0,798	0,898
8	0,632	0,715	0,765	0,872
9	0,602	0,685	0,735	0,847
10	0,576	0,658	0,708	0,823
11	0,553	0,634	0,684	0,801
12	0,532	0,612	0,661	0,780
13	0,514	0,592	0,641	0,760
14	0,497	0,574	0,623	0,742
15	0,482	0,558	0,606	0,725
16	0,468	0,543	0,590	0,708
17	0,456	0,529	0,575	0,693
18	0,444	0,516	0,561	0,679
19	0,433	0,503	0,549	0,665
20	0,423	0,492	0,537	0,652
25	0,381	0,445	0,487	0,597
30	0,349	0,409	0,449	0,554
35	0,325	0,381	0,418	0,519
40	0,304	0,358	0,393	0,490
45	0,288	0,338	0,372	0,465
50	0,273	0,322	0,354	0,443
60	0,250	0,295	0,325	0,408
70	0,232	0,274	0,302	0,380
80	0,217	0,257	0,283	0,338
90	0,205	0,242	0,267	0,338
100	0,195	0,230	0,254	0,321
	0,025	0,01	0,005	0,0005
ν	Односторонние границы			

[1] Значения $r_{кр}$, найденные для уровня значимости α и чисел степеней свободы $\nu = n - 2$ в случае парной корреляции и $\nu = n - l - 2$, где l — число исключенных величин, в случае частной корреляции.

[1]

Десятые доли r	Сотые (для последней строки тысячные) доли 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
0,1	0,1003	0,1104	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
0,2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
0,3	0,3095	0,3205	0,3316	0,3428	0,3541	0,3654	0,3767	0,3884	0,4001	0,4118
0,4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
0,5	0,5493	0,5627	0,5764	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
0,6	0,6932	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
0,7	0,8673	0,8872	0,9077	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
0,8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	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

[1] Значения $Z = \frac{1}{2}\ln\frac{1+r}{1-r}$.

[1]

m \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1637	0,2223	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1216	0,1438	0,1547
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4	0,0000	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5		0,0000	0,0000	0,0001	0,0002	0,0003	0,0007	0,0012	0,0020
6				0,0000	0,0000	0,0000	0,0001	0,0002	0,0003
7							0,0000	0,0000	0,0000
8									
9									
10									
11									
12					0,0000				
13									
14									
15									
16									
17									
18									
19									

[1] $P_{n,m}$ — вероятности того, что в n испытаниях событие A наступит m раз при вероятности наступления события p, близкой к нулю, и параметре $\lambda = np$.

1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0
0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0001
0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
0,0613	0,1805	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
0,0153	0,0902	0,1681	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337	0,0189
0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378
0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631
0,0001	0,0034	0,0216	0,0595	0,1045	0,1377	0,1490	0,1396	0,1171	0,0901
0,0000	0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318	0,1126
	0,0002	0,0027	0,0132	0,0363	0,0689	0,1014	0,1241	0,1318	0,1251
	0,0000	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186	0,1251
		0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970	0,1137
		0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728	0,0948
		0,0000	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504	0,0729
			0,0001	0,0005	0,0022	0,0071	0,0169	0,0324	0,0521
			0,0000	0,0002	0,0009	0,0033	0,0090	0,0194	0,0347
				0,0000	0,0003	0,0015	0,0045	0,0109	0,0217
					0,0001	0,0006	0,0021	0,0058	0,0128
					0,0000	0,0002	0,0009	0,0029	0,0071
						0,0001	0,0004	0,0014	0,0037

[1]

k \ $\nu = n - 1$	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	0,838	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	0,254	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	0,15
	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

[1] Пяти- и однопроцентные пределы для отношения G наибольшей выборочной дисперсии к сумме k выборочных дисперсий, полученных из k независимых выборок объемом n. Первое значение соответствует уровню значимости $\alpha = 0{,}05$, а второе — $\alpha = 0{,}01$.

1. *Ефимова М. Р.* Общая теория статистики / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев — М. : ИНФРА-М, 2011.

2. *Ильенкова С. Д.* Экономика фирмы и микростатистика / С. Д. Ильенкова, Н. Д. Ильенкова, Н. В. Тихомирова, С. А. Орехов. — М. : Финансы и статистика, 2007.

3. Курс социально-экономической статистики / под ред. М. Г. Назарова. — М. : Финстатинформ, 2010.

4. Международная статистика / под ред. Б. И. Башкатова, А. Е. Суринова — М. : Юрайт, 2010.

5. *Мхитарян В. С.* Теория вероятностей и математическая статистика / В. С. Мхитарян, Е. В. Астафьева, Ю. Н. Миронкина, Л. И. Трошин. — М. : МФПА, 2011.

6. *Салин В. Н.* Статистика / В. Н. Салин, Э. Ю. Чурилова, Е. П. Шпаковская. — М. : Кнорус, 2012.

7. *Сигел Э.* Практическая бизнес-статистика : пер. с англ. / Эндрю Ф. Сигел. — М. : ИД «Вильямс», 2002.

8. *Симонова М. Д.* Система национальных счетов. Счет производства / М. Д. Симонова. — М. : Статистика России, 2007.

9. Статистика / под ред. И. И. Елисеевой. — М. : Юрайт, 2011.

10. Статистика / под ред. Л. И. Ниворожкиной. — М. : ИТД «Дашков и К°», 2009.

11. Статистика / под ред. С. А. Орехова. — М. : Эксмо, 2010.

12. Теория статистики / под ред. проф. Г. Л. Громыко — М. : ИНФРА-М, 2010.

13. Теория статистики / под ред. проф. Р. А. Шмойловой — М. : Финансы и статистика, 2009.

14. *Шмойлова Р. А.* Практикум по теории статистики / Р. А. Шмойлова, В. Г. Минашкин, Н. А. Садовникова ; под ред. проф. Р. А. Шмойловой. — М. : Финансы и статистика, 2009.

Учебное издание

Мхитарян Владимир Сергеевич,
Дуброва Татьяна Абрамовна,
Минашкин Виталий Григорьевич,
Шмойлова Римма Александровна,
Садовникова Наталья Алексеевна

Статистика

Учебник

Редактор *Т. В. Сергачева*
Технический редактор *О. Н. Крайнова*
Компьютерная верстка: *Д. В. Федотов*
Корректоры *Л. В. Гаврилина*, *И. И. Феоктистова*

Изд. № 112102680. Подписано в печать 05.03.2013. Формат 60 × 90/16.
Гарнитура «Балтика». Бумага офсетная № 1. Печать офсетная. Усл. печ. л. 19,0.
Тираж 3 000 экз. Заказ №

Издательский центр «Академия». www.academia-moscow.ru
129085, Москва, пр-т Мира, 101В, стр. 1.
Тел./факс: (495)648-0507, 616-0029.
Санитарно-эпидемиологическое заключение № РОСС RU. AE51. Н 16067 от 06.03.2012.

Отпечатано с электронных носителей издательства.
ОАО «Тверской полиграфический комбинат», 170024, г. Тверь, пр-т Ленина, 5.
Телефон: (4822) 44-52-03, 44-50-34. Телефон/факс: (4822) 44-42-15
Home page — www.tverpk.ru. Электронная почта (E-mail) — sales@tverpk.ru