ТЕОРИЯ СТАТИСТИКИ

Под редакцией профессора Р.А.Шмойловой

Третье издание, переработанное

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

"Финансы и статистика"

УДК 311(075.8) ББК 60.6я73 Т33

АВТОРЫ: **Р.А. Шмойлова, канд. экон. наук, проф. — предисловие,**
параграфы 1.1, 1.2, 1.3, 1.4, 1.6,
главы 5, 10, алфавитно-предметный указатель;
Е.Б. Шувалова, канд. экон. наук, доц. — главы 2, 12,
параграфы 3.1, 3.2, 3.3, 3.4, 3.6, 3.7, 3.8;
Н.Ю. 1лубокова, канд. экон. наук, доц. — параграфы 1.7, 3.5;
А.Б. Гусынин, канд. экон. наук, доц. - параграфы 7.1, 7.2, 7.3
(совместно с проф. Р.А. Шмойловой), 7.4, 7.5, 7.6, 7.7, 8.1, 8.3, 8.4, 8.5;
В.Г. Минашкин, канд. экон. наук, доц. — главы 6,11, параграфы 7.8, 8.2;
А.А. Романов, д-р экон. наук, проф. — параграфы 1.5, 8.6;
Н.А. Садовникова, канд. экон. наук, доц. — главы 4, 9, 13,
приложения 1—13. В приложении 14 принимали участие все авторы.

РЕЦЕНЗЕНТЫ:
ГЛ. Громыко,
д-р экон. наук, проф. кафедры статистики Московского государственного университета им. М.В. Ломоносова;
О.Э. Башиня,
д-р экон. наук, проф., зав. кафедрой статистики Московского государственного университета коммерции

Теория статистики: Учебник/Под ред. проф. Р.А. ШмойлоТ33 вой. — 3-е изд., перераб. — М.: Финансы и статистика, 2001. — 560 с.: ил.
ISBN 5-279-01951-8

Излагаются общие вопросы теории статистики. Рассматриваются методы группировок, относительных и средних величин, показатели вариации, корреляционный и динамический анализ, статистическая взаимосвязь социально-экономических явлений, экономические индексы. Третье издание (2-е изд. — 1998 г.) дополнено тестами; в конце каждой главы введен краткий обзор формул и основных понятий по изученному материалу. Учебник снабжен математико-статистическими таблицами и дополнен новым приложением, содержащим краткий обзор основных формул и понятий теории статистики, алфавитно-предметным указателем.

Для преподавателей, студентов экономических вузов и бизнес-школ.

т **0702000000-011** ш_99
010(01)-2001

УДК 311(075.8)
ББК60.6я73

ISBN 5-279-01951-8

ПРЕДИСЛОВИЕ

Независимо от уровня и стадии экономического развития, характера политической системы, статистика на протяжении сотен лет своего существования всегда выступала как необходимый и эффективный инструмент государственного управления и одновременно как наука, исследующая количественную сторону массовых явлений. 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 и прикладных практических нужд различного рода структур, организаций и населения. Роль статистики в нашей жизни настолько значительна, что люди, часто не задумываясь и не осознавая, постоянно используют элементы статистической методологии в повседневной практике. 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 предпринимает определенные решения и действия. Таким образом, в каждом человеке генетически 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 в себе научно-исследовательскую компоненту статистического мышления, т. е. постичь множество специальных научных правил, методов и приемов количественного анализа разного рода информации. Что же необходимо для этого? Не очень много и не совсем мало серьезность и вдумчивость, неплохое знание математики, истории, экономической теории, основ предпринимательства и информатики. Чем шире кругозор и эрудиция в самых различных областях знаний, тем значительнее будут ваши успехи в изучении статистики.

Основными объектами приложения статистической теории и методологии выступают экономическая деятельность, население, условия жизни людей и управление. Ядром статистической системы знаний выступает **теория статистики,** обеспечивающая теоретическую и методологическую подготовку профессиональных статистиков, экономистов высшей квалификации, финансистов, менеджеров, коммерсантов и бухгалтеров, демографов и социологов, а также лиц других профессиональных интересов, самостоятельно изучающих статистику. Настоящий учебник представляет собой изложение курса теории статистики, ориентированное на развитие исследовательских и предпринимательских навыков, современное состояние российской экономики, развивающейся по пути реформ.

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

Результатом изучения курса теории статистики должны быть **знания** 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 (абсолютных и относительных, средних величин, индексов, показателей вариации, динамики, взаимосвязи), методов статистического моделирования и прогнозирования.

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

Структура учебника несколько отлична от традиционной и соответствует двум основным функциям статистической науки - описательной (дескриптивная — описательная статистика) и объясняющей (аналитическая статистика).

Первый раздел - **«Введение в теорию статистики»** - выделен с целью изучения:
сущности статистики как науки и сферы практической деятельности;
принципов и особенностей статистической теории и методологии;
отраслевой структуры статистики, ее основных понятий и категорий.

Во втором разделе — **«Описательная статистика»**
- излагаются:

методы сбора статистической информации (формы, виды и способы статистического наблюдения);
вопросы методологии и практики осуществления статистической сводки и группировки;
виды и методология построения статистических таблиц и графиков;
теория, методология и практика исчисления абсолютных, относительных и средних величин, их использования в анализе социально-экономических явлений.

В третьем разделе — **«Аналитическая статистика»**
- подробно рассматриваются:

методы анализа вариации и частотных распределений;
вопросы теории и практики выборочного наблюдения;
методы и показатели оценки взаимосвязей признаков:
методология статистического изучения динамики;
основные характеристики, виды и способы исчисления экономических индексов.

В отличие от прежних изданий курсов теории статистики в учебнике впервые введены самостоятельные главы, посвященные статистическому изучению структуры и структурных различий, а также общим вопросам анализа и обобщения статистических данных.

При подготовке настоящего учебника авторы испытывали некоторые трудности, обусловленные широтой современных проблем, возникающих в связи с переходом к рыночной экономике, необходимостью новых, принятых в международной практике подходов к статистической методологии, адаптации международных стандартов к российским условиям.

Учебник подготовлен авторским коллективом кафедры теории статистики и прогнозирования Московского государственного университета экономики, статистики и информатики и другими работающими в МЭ'СИ авторами под руководством кандидата экономических наук, профессора Р. А. Шмойловой.

РАЗДЕЛ

———~\!/——~~~~

ВВЕДЕНИЕ В ТЕОРИЮ СТАТИСТИКИ

ГЛАВА 1

СТАТИСТИКА КАК НАУКА

1.1

ПОНЯТИЕ СТАТИСТИКИ

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 Госкомстата РФ о выполнении планов федеральных программ развития национального хозяйства, публикуемые сведения о количестве родившихся и умерших в республике, данные о браках и разводах и т. д. К рубрике в газетах «Немного статистики» 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.

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

Итак, что же представляет собой термин «статистика»? Это слово многолико, многозначно и согласно одному из статистических терминов многомерно. В настоящее время насчитывается около тысячи определений статистики. Первое из них относится . к 1749 г.: на протяжении 250 лет оно уточнялось и дополнялось. Дать определение статистики как науки пытались философы, математики, экономисты, социологи, государственные деятели и, конечно, сами статистики. Сначала статистику определяли как «Staaten Kunde» - государствоведение (описание достопримеча-

Таблица 1.1

Численность постоянного населения РФ за 1959-1998 гг. (на начало года)

Год	Все население, млн человек	В том числе		В общей численности населения. %	
		городское	сельское	городское	сельское
1959	117,5	61,6	55,9	52	48
1970	129,9	80,6	49,3	62	38
1979	137,4	94,9	42,5	69	31
1989	147,0	107,9	39,1	73	27
1992	148,3	109,2	39,1	74	26
1993	148,3	108,5	39,8	73	27
1994	148,0	108,0	40,0	73	27
1995	147,9	107,9	40,0	73	27
1998	147,1	107,5	39,6	73	27

Едва ли не в каждом своем аспекте явления природы, а также человеческая и прочая деятельность поддаются сейчас измерению при помощи статистических показателей» *(Рейхман У. Дж.* Применение статистики. - М.: Статистика, 1969. - С. 11).

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

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

Многочисленные нелестные эпитеты, адресованные статистике, не затрагивают непосредственно ее как науку, рассматривающую совокупность методов, с помощью которых можно иссле-

тельностей государств). Оставаясь на протяжении многих лет «государствоведением», статистика постепенно отходила от описания достопримечательностей (их текстового изложения). Тем более что с развитием знаний вопросами государствоведения стали заниматься многие науки.

В настоящее время под термином «статистика» чаще всего понимают следующее.

Статистика - одна из **общественных наук,** имеющая целью сбор, упорядочивание, анализ и сопоставление числового представления фактов, относящихся к самым разнообразным массовым явлениям. Это вместе с тем **учение о системе показателей,** т. е. количественных характеристик, дающих всестороннее представление об общественных явлениях, о национальном хозяйстве в целом и отдельных его отраслях. Статистика - это **эффективное орудие, инструмент познания,**

используемый в естественных и общественных науках для установления тех специфических закономерностей, которые действуют в конкретных массовых явлениях, изучаемых данной наукой.

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

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

Часто слово «статистика» употребляется в качестве более короткого эквивалента для слов **«статистические методы».** Статистические методы можно охарактеризовать как методы, применяемые при сборе, представлении, анализе и интерпретации данных. В качестве примера можно упомянуть о методах, применяемых при сборе данных о совокупности студентов вузов, обработке этих данных, обобщении и представлении в виде различных итоговых абсолютных, относительных и средних показателей с помощью графиков, таблиц.

Применение статистических методов особенно важно там, где из больших массивов данных требуется выделить полезную для нас информацию.

Иногда слово «статистика» может **употребляться одновременно в нескольких значениях.** Известный английский статистик У. Дж. Рейхман (р. 1920) заметил: «Мы живем в век статистики.

Политические арифметики путем обобщения и анализа фактов стремились цифрами охарактеризовать состояние и развитие общества, показать закономерности развития общественных явлений, проявляющиеся в массовом материале. Цели и задачи, которые ставили перед собой ученые, близки к современному пониманию сущности статистики.

Идеи Д. Граунта, Э. Галлея и В. Петти имели своих последователей как на их родине, в Англии, так и в других европейских государствах. Наибольшее развитие эта школа получила в XVII и XVIII вв. в Англии, Голландии, Франции.

В дальнейшем понимание школы политических арифметиков в основном свелось к демографии и особенно к вопросам страхования жизни, финансовым расчетам.

В первой половине XIX в. возникло третье направление статистической науки — статистико-математическое. Особый вклад в развитие этого направления внес бельгийский статистик А. Кетле (1796 - 1874). Он назвал статистику «социальной физикой», т. е. наукой, изучающей законы общественной системы с помощью количественных методов. Важнейшей его заслугой 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.

Двумя другими учеными, внесшими значительный вклад в развитие статистики, являются два англичанина - Ф. Гальтон (1822 - 1911) и К. Пирсон (1857 - 1936). Гальтон, родственник Чарлза Дарвина, серьезно заинтересовался проблемой наследственности, к анализу которой он вскоре применил статистические методы. Кроме всего прочего им было разработано использование понятия перцентиля. Пирсон также провел много плодотворных исследрваний в статистике. И Гальтон, и Пирсон внесли значительный вклад в развитие теории корреляции.

Наиболее известным ученым XX в. в области статистики Запада является Р. Фишер (1890 - 1962). Фишер продуктивно работал с 1912 по 1962 г., и многие его исследования оказали существенное воздействие на современную статистику.

Представители статистико-математического направления считают основой статистики теорию вероятностей, составляющую одну из отраслей прикладной статистики.

В российской статистике не было четкого обособления школ и направлений, и тем не менее можно отметить русскую описательную школу, русскую школу политических арифметиков, ста-
университете начал читать новую учебную дисциплину, которую он назвал статистикой.

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. 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. Хотя авторы трудов называли их статистическими, с этим нельзя было согласиться, так как в них было словесное описание «достопримечательностей государства», т. е. в большей мере эти описания носили этнографический характер. Содержание, задачи, предмет изучения статистики в понимании ученых были далеки от современного взгляда на статистику как науку.

Гораздо ближе к современному пониманию статистики была английская **школа политических арифметиков,** которая 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 наметились два направления: демографическое с преобладанием вопросов страхования жизни у Д. Граунта и Э. Галлея и статистико-экономическое у В. Петти. Д. Граунт на основе обработки бюллетеней о естественном движении населения Лондона впервые открыл некоторые закономерности массовых общественных явлений и показал, как следует обрабатывать и анализировать массовый первичный материал. Он впервые попытался построить таблицу смертности для стационарного населения. Теоретическую разработку проблемы смертности, начатую Д. Граунтом, продолжил Э. Галлей. Знаменитый английский астроном высказал идею закона больших чисел и применил методы устранения случайных отклонений.

Друг и современник Д. Граунта В. Петти написал по статистике ряд научных работ, главные идеи которых состоят в стремлении конкретно оценить то или иное явление в условиях явной нехватки числовых данных. В. Петти является создателем экономической статистики.

Свою роль *в* истории статистики сыграли представители академической школы статистики, характерной особенностью которой было стремление заменить изучение государства изучением общества. Основоположниками этой школы явились 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 !954 г. этот вопрос обсуждался на научном совещании, которое еще раз подтвердило значение статистики как самостоятельной общественной науки. После совещания вышли в свет новые монографии, учебники по общей теории статистики. В это время значительный вклад в теорию индексного метода был внесен учеными С. М. Югенбергом, Г. И. Баклановым, Л. С. Казинцом, В. Е. Адамовым и др. Заслуживают серьезного внимания труды по изучению статистической связи Я. И. Лукомского.

Большим шагом вперед к развитию статистической науки послужило комплексное применение, наряду со статистическими, экономико-математических методов и широкое использование компьютерной техники в анализе социально-экономических явлений.

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

тистическую мысль революционеров-демократов русской социологической школы, различные направления в русской академической статистике.

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

Превращению статистики из науки описательной в науку теоретическую, формированию статистики как науки 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 большое значение имеют работы К. И. Арсеньева (1789 - 1856), в которых он утверждал, что статистика в состоянии дать адекватную характеристику жизни государства.

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

Большое влияние на развитие русской статистической мысли оказали русские демократы-революционеры: А. Н. Радищев (1749 - 1802), А. И. Герцен (1812 - 1870), Н. П. Огарев (1813 - 1874).

Эти выдающиеся деятели внесли определенный вклад в теорию и практику статистики. Ими разработаны программные вопросы экономической и судебной статистики, делались попытки определять средние величины, поставлен вопрос о социально-экономическом значении метода группировок.

(объектов, явлений, процессов) по каким-либо существенным признакам, но различающихся по каким-либо другим признакам. Например, из названных совокупностей множество сельскохозяйственных предприятий наряду с качественной определенностью (принадлежностью к разряду предприятий, причем определенной отрасли - сельского хозяйства) обладает различиями по размеру земельных угодий, численности работающих, голов скота, различной технологической оснащенностью и т. д.

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

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

Именно наличие вариации предопределяет необходимость статистики.

Вариация признака, к примеру, может быть отражена следующим статистическим распределением (табл. 1.2).

Данные таблицы показывают, что 50,2% безработных приходится на самый активный трудоспособный возраст 30 - 49 лет, т. е. средний возраст безработного - 34,4 года, и на людей, имеющих профессиональное и среднее общее образование -70%.

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

Заметим, что характер распределения и его параметры зависят от конкретных,--шповии яиетгга~РГЪремени.

1.3

ОСНОВНЫЕ ЧЕРТЫ ПРЕДМЕТА СТАТИСТИКИ И ЕГО ОПРЕДЕЛЕНИЕ

Как мы уже говорили, под статистикой понимают особую науку, которая имеет свой предмет и специфические методы исследования.

Рассмотрим основные черты и особенности предмета статистической науки.

Первой особенностью статистики как науки является то, что исследуются не отдельные факты, **а массовые социально-экономические явления и процессы,** выступающие как множества отдельных фактов, обладающих как индивидуальными, так и общими признаками.

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

Объект статистического исследования (в каждом конкретном случае) называют статистической совокупностью. **Статистическая совокупность - это множество единиц, обладающих массовостью, однородностью, определенной целостностью, взаимозависимостью состояний отдельных единиц и наличием вариации.** Например, в качестве особых объектов статистического исследования, т. е. статистических совокупностей, могут выступать множества: сельскохозяйственных предприятий, семей, браков, студентов, граждан какой-либо страны. Важно помнить, что статистическая совокупность состоит из реально существенных материальных объектов.

Каждый отдельно взятый элемент данного множества называется **единицей статистической совокупности.** Единицы статистической совокупности характеризуются общими свойствами, именуемыми в статистике признаками, т. е. под качественной однородностью совокупности понимается сходство единиц

Таблица 1.3

Число вынужденных переселенцев и беженцев на 1 января 1997 г.

	Человек
Всего в Российской Федерации	458 929
Из них ранее проживали на территории:	
Азербайджана	45 396
Армении	3 010
Беларуси	184
Грузии	50 029
Казахстана	82 914
Киргизии	30 470
Латвии	8 923
Литвы	1 241
Молдовы	9 722
России	61 046
Таджикистана	85 443
Туркмении	5 864
Узбекистана	65 740
Украины	3 193
Эстонии	5 318
Территория не указана	436

мени, т. е. **предметом статистики выступают размеры и количественные соотношения социально-экономических явлений, закономерности их связи и развития.** Например, статистика изучает экономические характеристики производства, распределения и потребления, уровень материального благосостояния населения, явления культурной жизни, численность населения земного шара, его распределение по континентам и странам и т. д.

Другим выражением количественной стороны общественной жизни *являются* **числовые соотношения** размеров общественных явлений. Например в 1997 г. на 1000 человек населения в РФ приходилось 6,3 брака и 3,8 развода. В 1996 г. в нашей стране в расчете на душу населения за год потреблено мяса и мясопродуктов 48 кг, яиц 173 шт., рыбы и рыбопродуктов 9,0 кг. Валовой внутренний продукт РФ *в 1997 г. по* сравнению с ! 996 г. составил 100,8%.

Таблица 1.2

Распределение численности безработных в РФ по полу, возрасту и образованию в 1997 г. (на конец октября, %)

	Всего	В том числе распределение по полу	
		мужчины	женщины
Безработные — всего,	100,0	100,0	100,0
в том числе в возрасте, лет:			
16-19	8,7	7,3	10,2
20-24	16,8	17,0	16,6
25-29	12,0	12,2	11,8
30-49	51,1	51,0	51,2
50 - 54	5,0	4,8	5,3 '
55-59	4,4	5,5	3,2
60 - 72	2,0	2,2	1,7
Средний возраст, лет	34,4	35,2	33,6
Безработные — всего,	100,0	100,0	100,0
в том числе имеют образование:			

высшее профессиональное	9,3	8,!	10,7
среднее профессиональное	27,7	23,0	33,0
среднее (полное) общее	43,2	45,8	40,3
основное общее	16,4	18,8	13,8
не имеют основного общего	3,3	4,3	2,2

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

Второй особенностью статистики как науки является то, что она изучает прежде всего количественную сторону общественных явлений и процессов в конкретных условиях места и вре-

Таблица 1.4

Распределение численности занятого населения по секторам экономики в РФ за 1985 - 1997 гг.

	Млн человек		% к итогу	
	1985	1997	1985	1997
Всего занято в экономике	74,9	65,4	100,0	100,0
Из них:				
на государственных и муниципальных предприятиях и в организациях	68,2	27,4	91,1	41,9
в частном секторе	0,7	23,7	0,9	36,2
в общественных организациях, фондах	-	0,4	-	0,6
на совместных предприятиях	-	0,6	-	0,9
на предприятиях и в организациях смешанной формы собственности	6,0	13,3	8,0	20,4

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

Такое свойство совокупности, как типичность, измеряется средними показателями.

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

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

При анализе структуры выявляются составные части социальноэкономических явлений. Эти составные части сопоставляются с явлением в целом и между собой. Производятся сравнение данной структуры с другими однотипными структурами, а также с задан-

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

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

Однако чем же тогда статистика отличается от математики?

Основная особенность статистики, по нашему мнению, состоит в том, что **она изучает количественную сторону качественно определенных массовых общественных явлений в данных условиях места и времени.** При этом **качественную определенность единичных явлений обычно определяют сопряженные науки.** Так, например, статистика изучает смертность как массовое явление, а факт и причину смерти устанавливает медицина.

Если для арифметики конкретное содержание количества не имеет значения (всегда 4-4=16), то для статистики цифры без качественного содержания, а также без обстоятельства места и времени лишены всякого смысла. Назовем цифру 8 664 тыс. - это численность населения Москвы (качественная определенность). Однако число жителей в городе постоянно меняется, следовательно, необходимо назвать и обстоятельство времени - 1996 г.

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

Данные табл. 1.4 характеризуют качественные изменения, которые произошли за рассматриваемый период в сфере экономики в России. За 1985 - 1996 гг. появились предприятия негосударственных форм собственности, увеличился удельный вес работников, занятых на этих предприятиях.

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

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

Таблица 1.5

Динамика обеспеченности населения РФ жильем (в среднем на одного жителя, м^2)

	1980	1985	1990	1996
Все население	13,4	14,9	16,4	18,3
В том числе проживающее в местности:				
городской	13,3	14,4	15,7	18,1,
сельской	13,7	16,2	18,2	18,8

ственных связей, чтобы воздействовать на общественные явления с целью их изменения в интересах общества.

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

Учитывая вышесказанное, сформулируем определение статистики как науки. **Статистика - общественная наука, которая изучает количественную сторону качественно определенных массовых социально-экономических явлений и процессов, их структуру и распределение, размещение в пространстве, движение во времени, выявляя действующие количественные зависимости, тенденции и закономерности, причем в конкретных условиях места и времени.**

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

- уровня и структуры массовых социально-экономических явлений;
- взаимосвязей массовых социально-экономических явлений и процессов;
- динамики массовых социально-экономических явлений. Таким образом, цель статистического исследования, как и

любого научного исследования, - раскрытие сущности массовых явлений и процессов, присущих им закономерностей. Отличительной особенностью этих закономерностей является то, что они относятся не к каждой отдельной единице совокупности, а ко всей массе единиц в целом. Общим принципом, лежащим в основе исследования статистических закономерностей, выступает так называемый закон больших чисел (ЗБЧ).

ной (плановой, нормативной и т. п.) и выявление причин отклонений. Подготавливаются предложения по оптимизации структуры. В процессе ее анализа используется метод группировок.

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

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

Таким образом, каждому общественному явлению свойственны изменения в пространстве и времени. Изменения в пространстве, т. е. **в статике,** выявляются посредством анализа структуры общественного явления, а изменения уровня и структуры явления исследуются во времени, т. е. **в**

динамике. Такова **четвертая особенность** статистики как науки. Анализ динамики включает: установление размера уровня общественных явлений на определенные моменты или промежутки времени и среднего уровня; выявление характера изменений за каждый промежуток времени и в целом; определение величины и темпов изменения; установление основной тенденции изменений, их закономерностей и составление статистического прогноза.

Рассмотрим для примера данные об обеспеченности населения Российской Федерации жильем (в среднем на одного жителя м^2 общей площади) (табл. 1.5).

Только статистика может так наглядно отразить изменение социально-экономического явления во времени.

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

1.4

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ СТАТИСТИКИ КАК НАУКИ

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

В предшествующем параграфе было дано общее представление об основных категориях статистической науки. Рассмотрим их подробнее.

Статистическая совокупность — это множество (масса) однокачественных (однородных) хотя бы по одному какому-либо признаку явлений, существование которых ограничено в пространстве и времени. Статистической совокупностью можно считать, к примеру, совокупность жителей России по состоянию на 1 января 1998 г., совокупность фермерских хозяйств Ростовской области в 1998 г., совокупность студентов III курса МГУ в 1998/99 учебном году и т. п. Однако статистическая совокупность (множество) совсем не обязательно представляет большую численность единиц, в принципе она может быть и очень маленькой; например, объем совокупности малой выборки может составлять иногда 8-10 единиц.

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

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

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

Существует понятие однородности статистической совокупности. Оно относительно и вовсе не означает полного соответствия всех единиц совокупности, а лишь подразумевает наличие для всех единиц совокупности основного свойства, качества, типичности. Одна и та же совокупность единиц, к примеру, может быть однородна по одному признаку и неоднородна по другому. Однородность единиц статистической совокупности формируется под воздействием определенных внутренних причин и условий. Одинаковые для всех единиц данной совокупности причины и условия существования создают то общее, что объединяет единицы совокупности, но эти же причины и условия формируют то, что отличает одну единицу совокупности от другой. В статистической совокупности эти отличия чаще имеют количественную природу. Количественные изменения значений признака при переходе от одной единицы совокупности к другой называются вариацией. **Вариация** возникает под воздействием случайных, прежде всего внешних, причин.

Социально-экономические явления, как правило, обладают большой вариацией. Например, вариация городов страны по численности населения складывается под влиянием большого числа факторов: исторических, этногеографических, экономических, социальных и множества других.

В большинстве теоретических и практических статистических исследований широко используются показатели вариации, которые показывают, как группируются значения признака вокруг средней величины совокупности. Показатели вариации выступают одновременно и мерами однородности совокупностей.

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

- атрибутивные (описательные, качественные);
- количественные.

Атрибутивные (качественные) признаки не поддаются прямому количественному (числовому) выражению.

Отличие количественных признаков от качественных состоит в том, что первые можно выразить итоговыми значениями, например общий объем добычи нефти в стране, выплавка стали; вторые - только числом единиц в совокупности (табл. 1.6).

Количественные признаки делятся на дискретные (прерывные) и непрерывные.

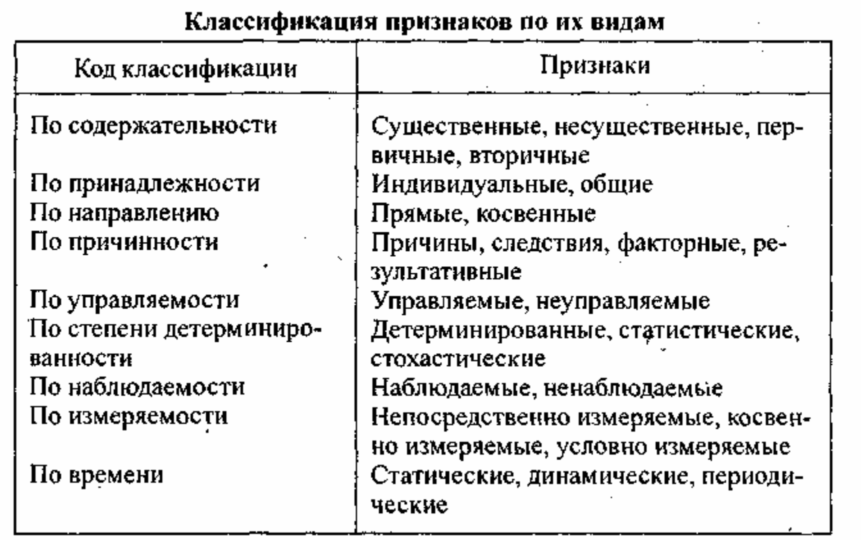
Таблица 1.6

Распределение театров в РФ по их направлению за 1980 — 1996 гг. (на конец года)

	1980	1996
Всего профессиональных театров	324	489
В том числе:		
оперы и балета	22	53
драмы, комедии и музыкальных	199	288
детских и юного зрителя	103	141

Все разнообразие признаков можно представить в табл. 1.7.

Таблица 1.7



Классификация признаков по их видам

Код классификации	Признаки
По содержательности	Существенные, несущественные, первичные, вторичные
По принадлежности	Индивидуальные, общие
По направлению	Прямые, косвенные
По причинности	Причины, следствия, факторные, результативные
По управляемости	Управляемые, неуправляемые
По степени детерминированности	Детерминированные, статистические, стохастические
По наблюдаемости	Наблюдаемые, ненаблюдаемые
По измеряемости	Непосредственно измеряемые, косвенно измеряемые, условно измеряемые
По времени	Статические, динамические, периодические

Важнейшей категорией статистики является статистическая закономерность. **Под закономерностью вообще принято называть повторяемость, последовательность и порядок изменений в явлениях.**

Статистическая же закономерность в статистике рассматривается как количественная закономерность изменения в пространстве и времени массовых явлений и процессов общественной жизни, состоящих из множества элементов (единиц совокупности). *Она* свойственна не отдельным единицам совокупности, а всей их массе, или совокупности в целом. В силу этого закономерность, присущая данному явлению (процессу), проявляется только при достаточно большом числе наблюдений и только в среднем. Таким образом, это закономерность усредненных параметров некоторого основного свойства (качества или типичности).

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

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

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

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

Статистическая закономерность - объективная количественная закономерность массового процесса. Она возникает в результате действия объективных законов, выражая каузальные отношения.

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

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

Так как статистическая закономерность обнаруживается в итоге массового статистического наблюдения, это обусловливает ее взаимосвязь с законом больших чисел.

Закон больших чисел в наиболее простой формулировке гласит, что количественные закономерности массовых явлений отчетливо проявляются лишь в достаточно большом их числе. Например, 104-106 мальчиков рождаются на 100 девочек, однако в отдельной семье и даже в небольшом населенном пункте это соотношение может быть совершенно иным.

Характеризуя роль закона больших чисел и его значение для статистики, А. Боярский отмечает, что в средней величине момент случайности оказывается снятым. Следовательно, закон больших чисел по форме говорит о том же, о чем гласит и статистика, - о результате, складывающемся из ряда (многих) элементов. Закон больших чисел позволяет определить число «необходимых» наблюдений за массовым процессом, а также и некоторые условные критерии для оценки «однородности» явлений. Он характеризует достоверность полученных результатов и необходим для понимания ошибок массового статистического наблюдения. Однако закон больших чисел не определяет собой постановку наблюдения и исследования в целом. Последняя диктуется необходимостью точного и всестороннего учета существенных факторов, влияющих на результат исследования.

Закон больших чисел выражает диалектику случайного и необходимого, единичного, особенного и всеобщего. В результате взаимопогашения случайных отклонений средние величины, исчисленные для величин одного и того же вида, становятся типичными, отражающими действие постоянных и существенных факторов в данных условиях места и времени.

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

При анализе общественных явлений весьма важно его вести таким образом, чтобы отразить и существенные стороны единичного, т. е. отдельного факта. О единичном в статистическом анализе нельзя забывать потому, что оно хотя и теряет свое значение в общей массе, однако продолжает существовать. Поэтому экономист часто убеждается, что как типичные выступают не только массовые, но и единичные, частные явления. Надо только уметь разобраться, имеем ли мы дело с отдельными изолированными фактами, не связанными с целым, или речь идет об-уникальных явлениях, за которыми кроются определенные социальные или экономические причины.

Вообще проникновению в сущность общественных явлений лучше всего способствует сочетание статистических обобщений с анализом отдельных связанных с ним фактов.

Итак, к помощи статистики прибегают тогда, когда исследователю необходимо раскрыть наличие статистических фактов, чтобы дать им соответствующее объяснение на основе причинн&х связей и отношений.

1.5

ОСОБЕННОСТИ СТАТИСТИЧЕСКОЙ МЕТОДОЛОГИИ. МЕТОД СТАТИСТИКИ

Методологической основой статистики является теория познания, определяющая научный подход к изучению явлений природы и общества. Это означает, что статистика анализирует социально-экономические явления и процессы не изолированно, а в их взаимодействии, взаимосвязи, не в состоянии покоя и неизменности, а в движении, в изменении и развитии.

Философия и логика заложили прочную основу для развития статистической теории. Научная разработка таких категорий, как количество и качество, необходимость и случайность, причинность, закономерность и другие, дает возможность философски осмыслить предмет, метод и задачи статистики как науки.

Знание диалектических законов и категорий позволяет статистике правильно понять и истолковать явления, подлежащие статистическому исследованию, выбрать надлежащий инструмент и методологически правильный подход к их изучению. Вместе с тем статистика как наука имеет свои специфические приемы изучения, зависящие от особенностей ее предмета. Совокупность этих приемов, с помощью которых статистика изучает свой предмет, образует статистическую методологию.

Под статистической методологией понимается система приемов, способов и методов, направленных на изучение количественных закономерностей, проявляющихся в структуре, динамике и взаимосвязях социально-экономических явлений.

Статистическое исследование состоит из трех основных стадий:

1) статистическое наблюдение;
2) первичная обработка, сводка и группировка результатов наблюдения;
3) анализ полученных сводных материалов.

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

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

Метод массовых наблюдений. Начальной стадией статистических исследований является **статистическое наблюдение** научно организованный сбор сведений об изучаемых социальноэкономических процессах или явлениях.

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

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

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

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

Для правильного выделения качественно однородных групп следует выбирать основные, наиболее существенные для данного явления или процесса признаки. В зависимости от числа и вида признаков, решаемых задач и исходных данных группировки подразделяются на: простые и комбинационные; по количественным и качественным признакам: типологические, структурные и аналитические; многомерные; первичные и вторичные. Одним из этапов процесса группировки является построение **рядов распределения,** т. е. группировка единиц наблюдения по величине или значению признака.

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

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

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

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

Выделяют следующие основные этапы анализа:

1) констатация фактов и их оценка;

2) установление характерных черт и причин явления;

3) сопоставление явления с другими, принятыми за базу сравнения - нормативными, плановыми и прочими явлениями;

4) формулирование гипотез, выводов и предположений;

5) статистическая проверка выдвинутых гипотез с помощью специальных статистических показателей.

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

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

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

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

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

1.6

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ КАК ОТРАСЛЬ СТАТИСТИЧЕСКОЙ НАУКИ

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

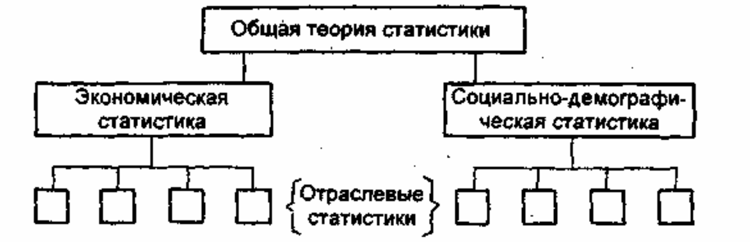


Рис. 1.1. Отрасли статистической науки

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

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

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

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

Таким образом, роль и место общей теории статистики особые. Она формирует общестатистическое мировоззрение экономики независимо от места его работы. Вот почему изучение статистики начинают с общих принципов и методов статистического исследования, т. е. с общей теории статистики.

На втором уровне выделены две большие обобщающие отрасли: экономическая и социально-демографическая статистики.

Экономическая статистика изучает явления и процессы в области экономики: структуру, пропорции, взаимосвязи отраслей и элементов общественного воспроизводства.

Социально-демографическая статистика изучает население, а также социальные (неэкономические) явления и процессы, которые характеризуют условия жизнедеятельности людей, их взаимоотношения в процессе труда и внепроизводственной деятельности. Ее основной целью является комплексное изучение различных сторон социальных условий и образа жизни людей.

На третьем уровне выделены **отрасли экономической и социально-демографической статистики.** В структуре экономической статистики выделяется макроэкономическая статистика, разрабатывающая методы комплексного изучения экономики страны, межотраслевые связи и др. Помимо этого в совокупность отраслей экономической статистики входят статистики промышленности, сельского хозяйства, торговли и других отраслей производственной сферы. В состав социально-демографической статистики входят статистики населения, уровня жизни, культуры, общественного мнения, политическая, моральная и другие отрасли. Каждая отраслевая статистика представляет собой науку о количественных изменениях, происходящих в соответствующих отраслях народного хозяйства. Задачей всех отраслевых статистик является разработка статистических показателей соответствующих отраслей.

Статистика развивается как единая наука, и развитие каждой отрасли содействует ее совершенствованию в целом.

Исходя из вышеизложенного отметим следующее: с точки зрения преподавания статистики в высшей школе она включает в себя целый ряд дисциплин: теория статистики, экономическая статистика, социальная статистика и целая серия отраслевых статистик: фирмы, ценообразования, рынка товаров и услуг, населения, демографическая, труда, здравоохранения, культуры и т. д.

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

1.7

ОСНОВНЫЕ ЗАДАЧИ И ПРИНЦИПЫ ОРГАНИЗАЦИИ ГОСУДАРСТВЕННОЙ СТАТИСТИКИ В РОССИЙСКОЙ ФЕДЕРАЦИИ

Задачи статистики. Одной из важнейших задач статистики является всестороннее освещение социально-экономического положения Российской Федерации, происходящих изменений, связанных с переходом к рыночным отношениям, т.е. получение полной и объективной количественной характеристики происходящих в стране преобразований.

Особое место в механизме управления экономикой принадлежит статистике, поскольку состав информации, **ее качество** и актуальность определяют **качественный** уровень принимаемых управленческих решений. Наличие систематической и полной информации о происходящих процессах и явлениях становится необходимым условием принятия эффективных управленческих решений на государственном и региональном уровнях. В этой связи обеспечение информационных запросов управленческих структур является весьма актуальной задачей статистики, способствующей успешному реформированию экономики.

С развитием рыночных отношений тесно связано реформирование российской статистики. В настоящее время выделяют два этапа проведения реформ в статистике. Начало глубокого реформирования статистики связано с Государственной программой перехода Российской Федерации на принятую в международной практике систему учета и статистики в соответствии с требованиями рыночной экономики и характеризует первый этап, который охватывает период с 1993 по 1996 г. включительно. Главной задачей первого этапа явилось содействие рыночным преобразованиям в стране, создание общеметодологических и организационных основ государственной статистики, соответствующих экономике переходного периода.

На первом этапе был определен состав показателей, адекватно отражающих процесс и результативность реформирования, а также внедрена система национальных счетов, наиболее полно отвечающая запросам рынка и соответствующая международной практике учета и статистики. В действующую систему показателей и учета были внесены изменения, касающиеся состава показателей рыночной экономики, методологии их оп- „деления. Начались внедрение цензового принципа организаии учета и переход от отраслевого метода сбора информации статистике предприятий. Формы государственной статистической отчетности были существенно пересмотрены и обновпен технический потенциал статистической системы. Главным итогом первого этапа реформ явились укрепление статистической системы, обеспечение объективности данных, активное осуществление перехода на принятую в мировой практике методологию статистики и учета.

Второй этап реформирования статистики начался в 1997 г. с разработкой Федеральной целевой программы «Реформирование статистики в 1997 - 2000 годах», основной целью которой является завершение реформирования статистики. Новый этап дальнейшего развития реформ связан с переходом на системный принцип реформирования статистики и адаптации ее к рыночным условиям. На современном этапе предусматривается комплексное, взаимосвязанное совершенствование всех элементов статистического наблюдения с учетом формирующегося рыночного спроса на информацию, новых требований к качеству информации различных потребителей. При достаточно полном удовлетворении информационных потребностей органов государственной власти остальные разрезы информации, ориентированные на спрос со стороны предпринимательского, научного и индивидуального секторов требуют серьезной проработки. В этой связи в качестве актуальных задач выступает совершенствование статистической информационной базы для максимального отражения общественных и социально-экономических явлений на основе развития системы статистических показателей, методологии их расчета и методов сбора статистической отчетности.

Крайне актуальными для второго этапа являются разработка методологии и организации получения информации о теневой экономике, осуществление взаимосвязи показателей макро- и микроуровней, разработка системы показателей, комплексно характеризующих экономику переходного периода, рыночные отношения, дальнейшее совершенствование организации статистического наблюдения, в том числе регистрового метода, классификаций, расширение сети мониторингов. Достаточно остро стоит проблема повышения оперативности сбора и обработки информации.

В соответствии с требованиями развития рыночной экономики встала задача компьютеризации статистики, являющаяся составной частью программы информатизации России, предусматривающей решение ряда задач по программной и тех* нической совместимости различных автоматизированных систем на федеральном и региональном уровнях. Предстоит создать информационно-телекоммуникационную статистику (ИТСС) на базе информационно-вычислительной сети, в основе которой лежит создание локальных вычислительных сетей (ЛВС), позволяющих перейти к новым информационным технологиям.

Успешному решению стоящих перед статистикой задач по информационному обеспечению общества и органов управления будет способствовать принятие Закона о статистической деятельности, который послужит правовой основой работы органов государственной статистики.

Организация государственной статистики в России. В основу организации статистической работы в России положены следующие принципы:

1) централизованное руководство статистикой;
2) единые организационное строение и методология;
3) неразрывная связь статистических органов с органами государственного управления.

В соответствии с Положением о Государственном комитете РФ по статистике, утвержденным постановлением Правительства РФ от 9 июля 1994 г., определено создание Государственного комитета Российской Федерации по статистике (Госкомстат России), который является федеральным органом исполнительной власти, осуществляющим руководство российской статистикой.

В соответствии с государственным устройством и административно-территориальными образованиями Российской Федерации создана единая система государственной статистики, которая проводит работу по единым плану и методологии. Методология статистических показателей, формы, методы сбора и обработки статистических данных, устанавливаемые Госкомстатом России, являются официальными статистическими стандартами Российской Федерации.

Система государственной статистики находится в ведении Правительства РФ и ему подотчетна, что обеспечивает неразрывную связь с органами государственного управления.

Система государственной статистики в России имеет иерархическую структуру, включающую федеральный, республиканский, краевой, областной, окружной, городской и районный уровни (рис. 1.2).

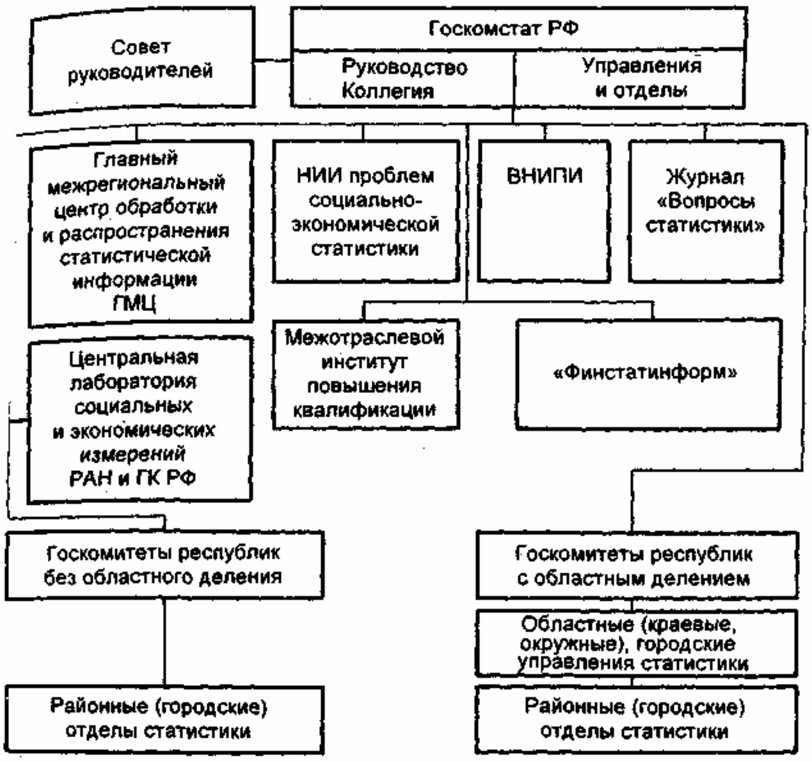


Рис. 1.2. Общая схема организации государственной статистики в РФ

Низовыми органами государственной статистики являются городские и районные управления государственной статистики. В областях, краях и республиках, а также в Москве и СанктПетербурге имеются статистические комитеты по статистике. Руководящим организационным и методологическим центром является Госкомстат РФ. Он осуществляет государственное управление всей находящейся в его ведении единой системой статистических органов, делом статистики, учета и отчетности во всех отраслях народного хозяйства, созданием и функционированием статистической информационной системы на основе единой наУЧНОЙ методологии. При Председателе Госкомстата России действует Совет руководителей органов государственной статистики, созданный для координации деятельности региональных органов государственной статистики.

В состав самого Госкомстата РФ входят такие управления, как аналитическое, информационных ресурсов и регистров, статистических стандартов и классификаций, организации статистического наблюдения, СНС и балансов, статистики финансов и платежного баланса, статистики цен, статистики товарных рынков ц торговли и др.

Структура госкомитетов и статистических управлений в основном повторяет структуру Госкомстата РФ, но подразделения низшего уровня являются более мелкими и с меньшим числом структурных подразделений.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 1

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

Аналитическая статистика - процедуры оценки характеристик совокупности по данным выборок.

Понятие статистики - общественная наука, имеющая целью сбор, упорядочивание, анализ и сопоставление данных, относящихся к самым разнообразным массовым явлениям.

Предмет статистики - количественная сторона качественно опре~ деленных массовых социально-экономических явлений и процессов, отображаемая посредством статистических показателей.

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

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

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

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

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

Статистический показатель - обобщающая количественная *ха*-

-теристика социально-экономических явлений в конкретных услоЙя* места и времени.

Система показателей - совокупность взаимосвязанных показате$_{е}$й которые отражают состояние и развитие массовых социально-экономических явлений с разных сторон.

Закономерность - повторяемость, последовательность и порядок изменений в явлениях.

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

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

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

Контрольные вопросы к главе 1

1. От какого латинского слова происходит термин «статистика»? Что он означает?

2. Какие статистические работы проводились в древние и средние века?

-?. К какому времени относится становление статистики как науки?

4. Почему статистика относится к общественным наукам? В чем ее отличие от других общественных неук?

3. Что такое закономерность? Статистическая закономерность и ее особенность.

6. Что такое совокупность, единица совокупности? Понятие вари°Чии и признака.

7. В чем сущность и значение закона больших чисел для статистики? # Дайте определение предмета статистики.

Я *Что является теоретической основой статистической науки? Ю. Почему каждое статистическое исследование должно опирать^{ся} на изучение всех относящихся к данному вопросу фактов?*

//. Почему статистика изучает явление общественной жизни $_{в}$ движении, изменении и развитии?

12. Перечислите специфические методы, присущие статистическому исследованию.

13. Какие принципы и методы излагаются в общей теории стати. с.тики? Почему изучение статистической науки начинается с оби/ру *теории статистики?*

14. Что определяет многообразие и слояс/юсть задач и функций статистики?

15. Какие принципы положены в основу организации статистики е *России?*

16. Какова организационная структура Госкомстата РФ?

РАЗДЕЛ

-\H^- ОПИСАТЕЛЬНАЯ СТАТИСТИКА

ГЛАВА 2

СБОР СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ (теория статистического наблюдения)

2.1

ПОНЯТИЕ О СТАТИСТИЧЕСКОМ НАБЛЮДЕНИИ, ЭТАПЫ ЕГО ПРОВЕДЕНИЯ

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

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

Процесс проведения статистического наблюдения включает следующие этапы:

- подготовка наблюдения;
- проведение массового сбора данных;
- подготовка данных к автоматизированной обработке;
- разработка предложений по совершенствованию статистического наблюдения.

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

Подготовка статистического наблюдения - процесс, включающий разные виды работ. Сначала необходимо решить методологические вопросы, важнейшими из которых являются определение цели и объекта наблюдения, состава признаков, подлежащих регистрации; разработка документов для сбора данных; выбор отчетной единицы и единицы, относительно которой будет проводиться наблюдение, а также методов и средств получения данных.

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

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

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

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

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

2.2

ПРОГРАММНО-МЕТОДОЛОГИЧЕСКИЕ ВОПРОСЫ СТАТИСТИЧЕСКОГО

НАБЛЮДЕНИЯ

Цель наблюдения. Статистические наблюдения чаще всего преследуют практическую цель - получение достоверной информации для выявления закономерностей развития явлений и процессов. Например, целью микропереписи населения России в 1994 г. было получение данных о численности, составе населения, условиях его проживания.

Ядача наблюдения предопределяет его программу и формы анизации. Неясно поставленная цель может привести к тому, °Р процессе наблюдения будут собраны ненужные данные или, ^{4} оборот, не будут получены сведения, необходимые для анализа.

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

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

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

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

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

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

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

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

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

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

К программе статистического наблюдения предъявляются следующие требования.

- Программа должна содержать существенные признаки, непосредственно характеризующие изучаемое явление, его тип, основные черты, свойства. Не следует включать в программу признаки, имеющие второстепенное значение по отношению к цели обследования или значения которых заведомо будут недостоверны или отсутствовать, например, в представлении такой информации, которая является предметом коммерческой тайны.
- Вопросы программы должны быть точными и недвусмысленными (иначе полученный ответ может содержать неверную информацию), а также легкими для понимания во избежание лишних трудностей при получении ответов.
- При разработке программы следует не только определить состав вопросов, но и их последовательность. Логичный порядок исследования вопросов (признаков) поможет получить достоверные сведения о явлениях и процессах.
- В программу целесообразно включать вопросы контрольного характера для проверки и уточнения собираемых данных.

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

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

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

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

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

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

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

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

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

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

Выбор времени наблюдения заключается в решении двух вопросов:

- установление критического момента (даты) или интервала времени;
- определение срока (периода) наблюдения.

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

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

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

2.3

ВАЖНЕЙШИЕ ОРГАНИЗАЦИОННЫЕ ВОПРОСЫ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

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

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

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

Размножение документации самого обследования, документации для проведения инструктажей и рассылка их республиканским, краевым, областным комитетам и управлениям статистики также относятся к организационным вопросам наблюдения.

В период подготовки большая роль отводится массоворазъяснительной работе: проведению лекций, бесед, организации выступлений в печати, по радио и телевидению о значении, целях и задачах предстоящего обследования.

Для согласования деятельности всех служб, занятых подготовкой и проведением наблюдения, целесообразно составить календарный план, представляющий собой перечень (наимечование) работ и сроки их исполнения отдельно для каждой °Рганизации, занятой в проведении обследования.

2.4

ОСНОВНЫЕ ОРГАНИЗАЦИОННЫЕ ФОРМЫ, ВИДЫ И СПОСОБЫ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

На этапе подготовки обследования нужно выяснить, как часто оно будет проводиться, будут ли обследоваться все единицы совокупности или только часть их, как получать информацию об объекте (путем интервью по телефону, по почте, простым наблюдением и т. п.). Другими словами, необходимо определить формы, способы и виды статистического наблюдения (табл. 2.1).

Таблица 2.1 **Формы, виды и способы статистического наблюдения**

Организационные формы статистического наблюдения	Виды статистического наблюдения		(~^ПГ\СТ\^1-*1 РТЯТ1ЛР \^ 1 lUkAJVJDI V I <л 1 rIL> тического наблюдения
	по времени регистрации фактов	по охвату единиц совокупности	
1 . Статистическая отчетность	1 . Текущее или непрерывное	1. Сплошное	1 . Непосредственное
2. Специально организованное наблюдение	2. Прерывное: а) периодическое; б) единовременное	2. Несплошное: а) выборочное; б) основного массива; в) монографическое	2. Документальное
3. Регистры			3. Опрос: а) экспедиционный; б) саморегистрации; в) корреспондентский; г) анкетный; д) явочный

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

- отчетность (предприятий, организаций, учреждений и т. п.);
- специально организованное статистическое наблюдение (пе-

писи, единовременные учеты, обследования сплошного и несплошного характера);

. регистры.

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

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

Для отчетности характерно то, что, во-первых, она утверждается органами государственной статистики. Предоставление информации по неутвержденным формам является нарушением отчетной дисциплины. Во-вторых, она имеет обязательный характер (т. е. все предприятия, учреждения, организации должны предоставлять ее в указанные сроки) и юридическую силу, потому что подписывается руководителем предприятия (учреждения, организации), а также документальную обоснованность, так как все данные базируются на документах первичного учета.

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

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

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

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

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

Характерными особенностями переписи являются: одновременность проведения ее на всей территории, которая должна быть охвачена обследованием; единство программы наблюдения; регистрация всех единиц наблюдения по состоянию на один и тот же критический момент времени. Программа наблюдения, приемы и способы получения данных по возможности должны ос^ таваться неизменными. Это позволяет обеспечить сопоставимость собираемой информации и получаемых в ходе разработки материалов переписи обобщающих показателей. Тогда можно не только определить численность и состав исследуемой совокупности, но и проанализировать ее количественное изменение в период между двумя обследованиями.

Из всех переписей наиболее известны переписи населения. Цель последних состоит в установлении численности и размещения населения по территории страны, характеристики его состава по полу, возрасту, занятиям и другим показателям. Первая всеобщая перепись населения России была проведена в 1897 г., а последняя - в 1989 г.

В период подготовки всеобщей переписи для уточнения и апробации программно-методических и организационных вопросов наблюдения проводят пробную перепись. Например, такая перепись была осуществлена в декабре 1986 г. Это обследование охватило не все, а только пять процентов населения страны. Запись сведений при переписи населения всегда проводится на основе его опроса (без требования предъявить какие-либо документы, подтверждающие правильность ответа).

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

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

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

Кроме переписей статистика проводит и другие специально организованные наблюдения, в частности бюджетные обследования, которые характеризуют структуру потребительских расходов и доходов семей.

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

Организация и ведение регистра невозможны без решения следующих вопросов.

* Когда заносить в регистр и исключать из него единицы совокупности?

* Какая информация должна храниться?

* Из каких источников следует брать данные?

* Как часто обновлять и дополнять информацию?

В практике статистики различают регистры населения и регистры предприятий.

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

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

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

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

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

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

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

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

Информационный фонд регистра содержит: во-первых, регистровый код субъекта; во-вторых, сведения об отраслевой, территориальной принадлежности субъекта, его подчиненности, виде собственности, организационной форме; в-третьих, справочные сведения (фамилии руководителей, адреса, номера телефонов, факсов и т. д., сведения об учредителях) и, наконец, в-четвертых, экономические показатели. Значения последних будут заноситься в регистр на основе бухгалтерской и статистической отчетностей, представляемых в региональные органы статистики.

Регистр содержит данные о следующих показателях: среднесписочная численность работников;

средства, направляемые на потребление; остаточная стоимость основных средств; балансовая прибыль (убыток); уставный фонд. Так как регистр ведется по отдельным территориям, региональные статистические службы могут расширять состав экономических показателей в случае необходимости.

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

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

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

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

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

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

Документальный способ наблюдения основан на использовании в качестве источника статистической информации различного рода документов, как правило, учетного характера. При надлежащем контроле за постановкой первичного учета и правильном заполнении статистических формуляров документальный способ дает наиболее точные результаты.

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

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

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

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

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

Анкетный способ предполагает сбор информации в виде анкет. Определенному кругу респондентов вручаются специальные вопросники (анкеты) либо лично, либо путем публикации в периодической печати. Заполнение этих вопросников носит добровольный характер и осуществляется, как правило, анонимно. Обычно обратно получают меньше анкет, чем рассылают. Этот способ сбора информации используется при несплошном наблюдении. Анкетный опрос применяется в обследованиях, где не требуется высо-

точность, а нужны приближенные, ориентировочные резуль^{К} ты1 например при изучении общественного мнения о работе городского транспорта, торговых предприятий и т. д.

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

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

Виды статистического наблюдения. Статистические наблюдения можно разбить на группы по следующим признакам:

- времени регистрации фактов;
- охвату единиц совокупности.

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

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

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

По охвату единиц совокупности статистическое наблюдение бывает сплошное и несплошное.

Задачей **сплошного** наблюдения является получение информации о всех единицах исследуемой совокупности.

До последнего времени российская система государственной статистики опиралась прежде всего на сплошное наблюдение. Однако такой вид наблюдения имеет серьезные недостатки: высокую стоимость получения и обработки всего объема информации; большие затраты трудовых ресурсов; недостаточную оперативность информации, так как для ее сбора и обработки необходимо много времени. И наконец, ни одно сплошное наблюдение, как правило, не обеспечивает полного охвата всех без исключения единиц совокупности. Большее или меньшее число единиц обязательно остается вне наблюдения как при проведении единовременных обследований, так и при такой форме наблюдения, как отчетность. Например, в настоящее время значительная часть предприятий негосударственного сектора не представляет необходимой информации органам государственной статистики, даже несмотря на принятый *Закон РФ «Об ответственности за нарушение порядка представления государственной статистической отчетности».*

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

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

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

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

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

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

Этот вид наблюдения применяется при проведении обследований доходов населения.

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

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

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

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

2.5

ТОЧНОСТЬ НАБЛЮДЕНИЯ

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

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

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

- обеспечить качественное обучение персонала, который будет проводить наблюдение;
- организовать специальные частичные или сплошные контрольные проверки правильности заполнения статистических формуляров;
- провести логический и арифметический контроль полученных данных после окончания сбора информации.

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

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

Ошибки регистрации бывают **случайные и систематические.**

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

систематические ошибки регистрации всегда имеют одина$_{\text{vfO}}$ тенденцию либо к увеличению, либо к уменьшению зна$^{\text{К}\circ}$ния показателей по каждой единице наблюдения, и поэтому ^{4} личина показателя по совокупности в целом будет включать в ^{B} б$_{\text{я}}$ накопленную ошибку. Примером статистической ошибки истрации $_{\text{п}}$р$_{\text{и}}$ проведении социологических опросов населения может служить округление возраста населения, как правило, на цифрах, оканчивающихся на 5 и 0. Многие опрашиваемые, например, вместо 48 - 49 и 51-52 лет говорят, что им 50 лет.

В отличие от ошибок регистрации ошибки репрезентативности характерны только для несплошного наблюдения. Они возникают потому, что отобранная и обследованная совокупность недостаточно точно воспроизводит (репрезентирует) всю исходную совокупность в целом.

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

Ошибки репрезентативности также бывают случайные и систематические.

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

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

После получения статистических формуляров следует прежде всего провести проверку полноты собранных данных, т. е. определить, все ли отчетные единицы заполнили статистические формуляры и значения всех ли показателей отражены в них. Следующим этапом контроля точности информации является арифметический контроль. Он основывается на использовании количественных связей между значениями различных показателей. Например, если среди собранных данных имеются сведения о численности промышленно-производственного персонала, выработке товарной продукции в среднем на одного работающего и стоимости товарной продукции, то произведение первых двух показателей должно дать значение третьего показателя. Если арифметический контроль покажет, что данная зависимость не выполняется, это будет свидетельствовать о недостоверности собранных данных.' Поэтому в программу статистического наблюдения целесообразно включать показатели, которые дают возможность провести арифметический контроль.

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

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

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 2

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

Цель наблюдения - получение достоверной информации для выявления закономерностей развития явлений to процессов.

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

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

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

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

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

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

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

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

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

регистровое наблюдение - форма непрерывного статистического блюдения за долговременными процессами, имеющими фиксированe начало, стадию развития и фиксированный конец. Н **Непосредственное наблюдение** - регистраторы путем непосредвенного замера, взвешивания, подсчета или проверки работы и так далее устанавливают факт, подлежащий регистрации, и на этом основании производят записи в формуляре наблюдения.

Документальный способ наблюдения - основан на использовании в качестве источника статистической информации различного рода документов, как правило, учетного характера.

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

Текущее наблюдение - наблюдение, когда изменения в отношении изучаемых явлений фиксируются по мере их наступления.

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

Сплошное наблюдение - получение информации о всех единицах исследуемой совокупности.

Несплошное наблюдение - обследованию подлежит лишь часть единиц изучаемой совокупности.

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

Ошибка наблюдения — расхождение между расчетным и действительным значением изучаемых величин.

ТЕСТ К ГЛАВЕ 2

1. Объект статистического наблюдения — это
а) единица наблюдения; "б) статистическая совокупность;
в) единица статистической совокупности;
г) отчетная единица.

2. Субъект, от которого поступают данные в ходе статистического наблюдения, называется:
а) единица наблюдения;
б) единица статистической совокупности; -в) отчетная единица.

3- Перечень признаков (или вопросов), подлежащих регистрации в процессе наблюдения, называется:
а) статистический формуляр;
б) профамма наблюдения;
в) инструментарий наблюдения.

4. Срок наблюдения — это
а) время, в течение которого происходит заполнение статистических формуляров;
б) конкретный день года, час дня, по состоянию на который должна быть проведена регистрация признаков

по каждой единице исследуемой совокупности.

5. Статистическая отчетность - это

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

6. Метод основного массива - это

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

7. Перепись населения России (1989 г) — это

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

8. Инвентаризация незавершенного производственного строительства 1980 г. - это

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

9. Метод моментных наблюдений - это разновидность:

а) сплошного наблюдения;
б) монографического обследования;
в) метода основного массива;
г) выборочного наблюдения.

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

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

ГЛАВА 3

СТАТИСТИЧЕСКАЯ СВОДКА И ГРУППИРОВКА

3.1

ЗАДАЧИ СВОДКИ И ЕЕ СОДЕРЖАНИЕ

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

Поэтому цель следующего этапа статистического исследования состоит в систематизации первичных данных и получении на этой основе сводной характеристики всего объекта при помощи обобщающих статистических показателей.

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

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

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

По глубине обработки материала сводка бывает простая и сложная.

Простой сводкой называется операция по подсчету общих Итогов по совокупности единиц наблюдения.

Сложная сводка представляет собой комплекс операц$_{и}$* включающих группировку единиц наблюдения, подсчет ито гов по каждой группе и по всему объекту и представленц результатов группировки и сводки в виде статистических таб лиц.

Проведению сводки предшествует разработка ее программы которая состоит из следующих этапов: выбор группировочных признаков; определение порядка формирования групп; разработка системы статистических показателей для характеристики групп и объекта в целом; разработка системы макетов статистических таблиц, в которых должны быть представлены результаты сводки.

По форме обработки материала сводка бывает децентрализованная и централизованная.

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

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

По технике выполнения статистическая сводка подразделяется на механизированную и ручную.

Механизированная сводка - это способ выполнения сводки статистических данных, при котором все операции осуществляются с помощью применения электронно-вычислительных машин. При **ручной сводке** все основные операции (подсчет групповых и общих итогов) осуществляются вручную. В настоящее время с появлением персональных компьютеров, созданием автоматизированных рабочих мест, разработкой статистических пакетов прикладных программ ручная сводка в обработке информации используется крайне редко.

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

3.2

МЕТОД ГРУППИРОВКИ И ЕГО МЕСТО В СИСТЕМЕ СТАТИСТИЧЕСКИХ МЕТОДОВ

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

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

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

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

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

С *помощью метода группировок решаются следующие задачи:*

′ выделения социально-экономических типов явлений;

• изучения структуры явления и структурных сдвигов, происходящих в нем;

* выявления связи и зависимости между явлениями.

3.3

ВИДЫ СТАТИСТИЧЕСКИХ ГРУППИРОВОК

Статистические группировки по задачам, решаемым с их помощью, делятся на: типологические, структурные и аналитические.

Типологическая группировка - это разделение исследуемой качественно разнородной совокупности на классы, социальноэкономические типы, однородные группы единиц в соответствии с правилами научной группировки. Примером типологической группировки является группировка промышленных предприятий по формам собственности (табл.3.1).

Табл ица 3.1

Группировка промышленных предприятий одного из регионов России по формам собственности в 1994 г.

№ п/п	Группы предприятий по формам собственности	Число предприятий	
		всего, единиц	% к итогу
1	Федеральная собственность	26326	93.6
2	Муниципальная собственность	89	0.3
3	Частная собственность	1 366	4,9
4	Смешанная собственность	331	1,2
	Всего	28 112	100

Согласно данным таблицы, подавляющее большинство предприятий находилось в федеральной собственности, менее 5% составляли предприятия с частной формой собственности и только 1,5% - с муниципальной и смешанной собственностью.

Типологические группировки широко применяются в исследовании социально-экономических явлений и процессов. Они позволяют проследить зарождение, развитие и отмирание различных типов явлений (табл. 3.2).

Исследуя данные табл. 3.2, отметим, что за три месяца число предприятий возросло в 4,4 раза, в том числе предприятий бытового обслуживания населения - почти в 6 раз. Наибольшее их количество приходится на розничную торговлю. За исследуемый период их численность возросла на 5329 единиц и составила более 52% общего числа предприятий.

Таблица 3.2

г уппировка предприятий службы сервиса региона России по назначению в 1994 г.

№ п/п	Группа предприятий по назначению	Число приватизированных предприятий, единиц	
		01.04.94г.	01.07.94 г.
1	Розничная торговля (палатки, магазины и др.)	1 194	6523
7	Общественное питание (столовые, рестораны и др.)	962	2 100 3
3	Бытовое обслуживание населения (ателье, мастерские и др.)	662	871
	Всего	2818	12494

Одна и та же совокупность может быть качественно однородной в одном статистическом исследовании и разнородной в другом. Так, совокупность промышленных предприятий является однородной в случае анализа показателей брака при производстве какой-либо продукции. Эта совокупность будет качественно неоднородной, если изучается налогообложение предприятий. Из всей совокупности единиц необходимо выделить, например, группу предприятий, на которых работают инвалиды, и группу малых предприятий, имеющих значительные льготы при уплате налогов.

При проведении типологической группировки основное внимание должно быть уделено идентификации типов социальноэкономических явлений. Она производится на базе глубокого теоретического анализа исследуемого явления.

Другой вид группировки - структурная. Структурной называется группировка, в которой происходит разделение однород^{ои} совокупности на группы, характеризующие ее структуру ^{По} какому-либо варьирующему признаку. С помощью таких Руппировок могут изучаться: состав населения по полу, возра^{Т}У> месту проживания; состав предприятий по численности занятых, стоимости основных фондов; структура депозитов *п* сроку их привлечения и т. д. Структурная группировка пр$_{е}$„ ставлена в табл. 3.3.

Таблица Зз

Группировка населения России по размеру среднедушевого дохода **в апреле 1994 г.**

No п/п	Группа населения по размеру среднедушевого денежного дохода, тыс. руб. в месяц	Численность населения	
		всего, млн человек	% к итогу
1	До 40	2,4	1,6
?	40-80	23,4	15,8
3	80 т 120	34,8	23,5
4	120- 160	29,4	19,8
<	160-200	20,7	13,9
6	200 - 240	13,5	9,1
7	240-280	8,7	' 5,9
8	280 и более	15,5	10,4
	Всего	148,4	100,0

Данные группировки показывают, что более 43% населения имело среднедушевой денежный доход от 80 до 160 тыс. руб. в месяц, доход до 40 тыс. руб. получало 1,6% населения.

В изменении структуры общественных явлений отражаются важнейшие закономерности их развития.

Группировка в табл. 3.4 показывает, что с 1959 по 1994 гг. численность населения, проживающего в городах, постоянно увеличивалась, а численность сельского - падала. Доля городского населения за указанный период увеличилась на 21 процентный пункт. В 1994 г. в городах проживало в 2,7 раза больше человек, чем в сельской местности.

Явления общественной жизни и отражающие их признаки тесно взаимосвязаны. Группировка, выявляющая взаимосвязи между изучаемыми явлениями и их признаками, называется **аналитической группировкой.**

Всю совокупность признаков можно разделить на две группы: факторные и результативные. **Факторными** называются признаки, под воздействием которых изменяются другие признаки - они и образуют группу **результативных** признаков. Взаимосвязь проявляется в том, что с возрастанием значения факторного признака систематически возрастает или убывает среднее значение признака результативного. Например, производительность труда зависит от технического уровня предприятия: чем он выше, тем при прочих равных условиях выше производительность труда занятых на предприятии. Поэтому, группируя промышленные предприятия, производящие одну и ту же продукцию, по техническому уровню производства (по уровню фондовооруженности труда) и исчисляя для каждой группы среднюю выработку товарной продукции на одного работающего, можно статистически отразить эту зависимость между факторами.

Таблица 3.4

Группировка населения России по месту проживания за 1959-1994гг.

.№ п/п	г --- --- ' Группа населения по месту жительств	Численность населения					
		1959		1979		1994	
		всего, млн человек	в%к итогу	всего, млн человек	в%к итогу	всего, млн человек	в%к . итогу
1	Городское	61,6	52	95,4	69	108,5	73
2	Сельское	55,9	48	42,2	31	39,9	27
	Всего	117,5	100	137,6	100	148,4	100

Особенности аналитической группировки следующие: во-первых, в основу группировки кладется факторный признак; во-вторых, каждая выделенная группа характеризуется средними значениями результативного признака. Аналитическая группировка приведена в табл. 3.5.

Данные таблицы характеризуют зависимость между суммой активов банка и численностью занятых, а также суммой балансовой прибыли. Чем больше сумма активов, тем больше прибыль ^{анка} и численность его сотрудников. У первой группы средняя численность занятых в 2,8 раза меньше, чем у пятой, а балансо$^{\text{Вая}}$ прибыль меньше в 9,1 раза.

Аналитические группировки позволяют изучить многообразие $^{\text{в}}$язей и зависимости между варьирующими признаками. Премущество метода аналитических группировок перед другими

Таблица 35

Группировка коммерческих банков России
по сумме активов баланса
(данные условные)

№ п/п	Группа банков по сумме активов баланса, млн руб.	Количеств банков, единиц	В среднем на один банк	
			численность занятых, человек	балансовая прибыль, млрд руб.
1	До 20 000	19	184	22,5
2	20 000 - 30 000	8	313	31,6
3	30 000 - 40 000	7	374	36,0
4'	40 000 - 50 000	9	468	69,2
5	50 000 и более	7	516	205,6
	Всего	50	323	60,0

методами анализа связи (например, корреляционно-регрессионным) состоит в том, что он не требует соблюдения каких-либо условий для своего применения, кроме одного - качественной однородности исследуемой совокупности.

Все рассмотренные в данном параграфе группировки объединяет то, что единицы объекта разделены на группы по какомуто одному признаку: форма собственности (табл. 3.1); назначение предприятий (табл. 3.2); размер среднедушевого дохода (табл. 3.3); место проживания (табл. 3.4) и сумма активов баланса банка (табл. 3.5).

Группировка, в которой группы образованы по одному признаку, называется **простой.** Для характеристики явления бывает недостаточно разбить совокупность на группы по какому-либо одному признаку. В этом случае строят сложные группировки.

Сложной называется группировка, в которой разделение совокупности на группы производится по двум и более признакам, взятым в сочетании (комбинации).

Сначала группы формируются по одному признаку, затем они делятся на подгруппы по другому признаку, которые, в свою очередь, подразделяются по третьему и т. д. Таким образом, ые группировки дают возможность изучать распределение $^{\text{олю>к}}$ и совокупности одновременно по нескольким признакам. $^{\text{С}\wedge\text{И}}$В табл. 3.6 приведена сложная группировка семей России по месту проживания и числу детей.

Таблица 3.6

Группировка семей России по месту проживания
и числу детей в 1989 г. (по материалам переписи населения)

п/п	Группа семей по месту проживания	В том числе подгруппа семей по числу детей	Число семей, тыс.
	Городское население	1 ребенок 2 детей 3« 4« 5 и более детей	9605 6936 971 153 76
Итого по группе			17741
2	Сельское население	1 ребенок 2 детей 3« 4« 5 и более детей	2328 2306 757 213 141

Итого по группе			5745
	Итого по подгруппам	1 ребенок 2 детей 3« 4« 5 и более детей	11 933 9242 I 728 366 217
Всего			23486

По первому признаку образованы две группы, а по второму "five"ять. Группировка показывает, что большинство семей и в гороДе, и в деревне имеет только одного ребенка, а в общей числен$^{\text{н°}}$сти семей они составляют почти 51% (11 933 : 23 486),

Число городских семей, в которых два ребенка, меньше почти в $^{'4}$ раза, чем семей с одним ребенком. В сельской местности чисенность таких семей различается незначительно - всего на 1%.

Меньше всего семей с численностью детей 5 и более челове Однако среди сельского населения они составляют 2,5%, гор$_{0\text{д}}$ского .- лишь 0,4%, а во всем населении - 0,9%.

При построении сложной группировки возникает вопрос $_{0}$ последовательности разбиения единиц объекта по признакам. Как правило, рекомендуется сначала производить группировку $_{\text{По}}$ атрибутивным признакам, значения которых имеют ярко выр$_{\text{а}}$. женные качественные различия.

С увеличением числа группировочных признаков в сложных группировках быстро растет количество групп. Группировка с большим числом групп становится ненаглядной. Поэтому на практике строят сложные группировки не более чем по трем признакам.

3.4

ПРИНЦИПЫ ПОСТРОЕНИЯ СТАТИСТИЧЕСКИХ ГРУППИРОВОК И КЛАССИФИКАЦИЙ

Приемы проведения статистических группировок весьма разнообразны. Это связано с разными задачами, которые в соответствии с целью исследования ставятся перед группировками.

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

В качестве основания группировки следует использовать существенные признаки. В каждом конкретном исследовании включение признака в состав группировочных должно быть теоретически обосновано. Только на базе теоретического анализа экономических законов развития исследуемого явления можно правильно определить состав признаков.

В основание группировки могут быть положены как количественные, так и качественные признаки. Первые имеют числовое выражение (объем торгов, курс доллара в рублях, возраст человека, денежный доход семьи и т. д.), а вторые отражают состояние единицы совокупности (пол человека, его национальность, семей-, ное положение, отраслевая принадлежность предприятия, его форма собственности и организационно-правовая форма и т. д.).

После определения основания группировки следует решить вопрос о количестве групп, на которые надо разбить исследуемую совокупность.

$_{\text{о}}$ групп зависит от задач исследования и вида признака, иного $_{\text{в}}$ основание группировки, численности совокупно$^{\text{ПоЛо}}$степени вариации признака.

$^{\text{стИ}}$с $_{\text{ини}}$цы анализируемого объекта могут быть разбиты по $_{\text{ом}}$у и тому же признаку на разное число групп. Например, $^{\text{од}}$ $_{\text{г}}$руппировке *населения по* возрасту с целью определения

удовых ресурсов страны все население в практической статитике делится на три группы: население моложе трудоспособно-

возраста, трудоспособное население и население старше трудоспособного возраста. Если же анализируется продолжительность жизни, то строится более детальная группировка и выделяются пятигодичные группы.

При построении группировки по качественному признаку групп, как правило, будет столько, сколько имеется градаций, видов, состояний у этого признака. Например, в случае проведения группировки населения по полу можно образовать только две группы: мужчины и женщины. Если проводится группировка производства товаров народного потребления по экономическим районам, то вся исследуемая совокупность *делится* на ! 1 групп: именно на столько экономических районов поделена территория страны.

В табл. 3.7 приведена группировка приватизированных российских предприятий по формам собственности, т, е. по качественному признаку.

Таблица 3.7

Группировка предприятий России по формам собственности, приватизированных в январе - августе 1994 г.

№ п/п k _ ____	Группа предприятий по формам собственности	Приватизировано предприятий, единиц		
		Всего	В том числе распределение по способу приватизации путем	
			продажи и выкупа	ционировани я
1	Муниципальная	7957 3	7 131	826
2.>	Субъектов Федерации	843	919	2924
j	Федеральная	4448	565	3883
	Всего	16248	8615	7633

Согласно данным табл. 3.7, вся исследуемая совокупно^ разбита на три группы: предприятия, находящиеся в федерад,^{ь} ной, муниципальной собственности и в собственности су&ь$_{ек}$ тов Федерации.

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

Отличительной чертой классификации является, *во-первых* то, что в основу ее кладется качественный признак. *Во-вгпа. рых,* классификации стандартны. Они устанавливаются органами государственной и международной статистики. Если в каждом конкретном исследовании строится своя группировка, то классификация едина для любого исследования независимо от того, проводят ли его органы государственной статистики или другие учреждения и ведомства (министерства, налоговые органы и т. п.). *В-третьих,* классификации устойчивы. Они остаются неизменными в течение длительного времени. Однако если появляются новые группы единиц, их классы, разряды, то в классификации вносятся соответствующие изменения и дополнения.

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

Если группировка проводится по количественному признаку, то необходимо обратить особое внимание на число единиц исследуемого объекта и степень колеблемости группировочного признака.

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

Часто группировка по количественному признаку имеет задачу отразить распределение единиц совокупности по этому при знаку. В данном случае количество групп зависит в первую очередь от степени колеблемости группировочного признака: чем больше его колеблемость, тем больше следует образовать групп. (Степень колеблемости признака измеряется показателями вариации, которые подробно рассмотрены в гл. 7.) Чем больше групп,

$_{чн}$ее будет воспроизведен характер исследуемого объекта. ^{TeVl} $_{о}$ слишком большое число групп затрудняет выявление

ономерностей при исследовании социально-экономических ^{33} $_{ен}$цй и процессов. Поэтому в каждом конкретном случае при ^{яв} $_{деле}$нии числа групп следует исходить не только из степени ° леблемости признака, но еще учитывать и особенности объек*$_{а}$ $_{и}$ цель исследования.

г]ри использовании электронно-вычислительных машин и „ерсональных компьютеров для обработки статистических данных группировка единиц объекта проводится с помощью стандартных процедур.

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

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

где п - число групп;

N - число единиц совокупности.

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

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

Другой способ определения числа групп основан на применении показателя среднего квадратического отклонения (а). Если величина интервала равна 0,5 сг, то совокупность разбивается на 12 групп, а когда величина интервала равна 2/3<з и с, то совокупность делится соответственно на 9 и 6 групп.

Если совокупность разбивается на 12 групп, то интервалы групп строятся следующим образом:

от	$\bar{x} - 3{,}0\sigma$	до	$\bar{x} - 2{,}5\sigma$
«	$\bar{x} - 2{,}5\sigma$	«	$\bar{x} - 2{,}0\sigma$
«	$\bar{x} - 2{,}0\sigma$	«	$\bar{x} - 1{,}5\sigma$
«	$\bar{x} - 1{,}5\sigma$	«	$\bar{x} - 1{,}0\sigma$
«	$\bar{x} - 1{,}0\sigma$	«	$\bar{x} - 0{,}5\sigma$
«	$\bar{x} - 0{,}5\sigma$	«	$\bar{x}$
«	$\bar{x}$	«	$\bar{x} + 0{,}5\sigma$
«	$\bar{x} + 0{,}5\sigma$	«	$\bar{x} + 1{,}0\sigma$
«	$\bar{x} + 1{,}0\sigma$	«	$\bar{x} + 1{,}5\sigma$
«	$\bar{x} + 1{,}5\sigma$	«	$\bar{x} + 2{,}0\sigma$
от	$\bar{x} + 2{,}0\sigma$	до	$\bar{x} + 2{,}5\sigma$
«	$\bar{x} + 2{,}5\sigma$	«	$\bar{x} + 3{,}0\sigma$,

где x - среднее значение признака по совокупности, которое опред$_{е}$

ляется по формуле

$$\bar{x} = \frac{\Sigma x_i}{n},$$

где x$_{(}$— i-е значение варьирующего признака;

а— среднее квадратическое отклонение; согласно формуле (7.6)

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}.$$

Когда число групп равно 6, получаются следующие интервалы групп:

от	$\bar{x} - 3\sigma$	до	$\bar{x} - 2\sigma$
«	$\bar{x} - 2\sigma$	«	$\bar{x} - \sigma$
«	$\bar{x} - \sigma$	«	$\bar{x}$
«	$\bar{x}$	«	$\bar{x} + \sigma$
«	$\bar{x} + \sigma$	«	$\bar{x} + 2\sigma$
«	$\bar{x} + 2\sigma$	«	$\bar{x} + 3\sigma$.

Эти методы не дают гарантии в том, что не будут сформированы «пустые» или малочисленные группы. «Пустыми» считаются группы, в которые не попала ни одна единица совокупности. Наличие таких интервалов свидетельствует, что группировка построена неправильно.

После определения числа групп следует определить интервалы группировки.

Интервал - это значения варьирующего признака, лежащие в определенных границах. Каждый интервал имеет свою величину, верхнюю и нижнюю границы или хотя бы одну из них. **Нижней границей** интервала называется наименьшее значение признака в интервале, а **верхней границей** - наибольшее значение признака в нем. **Величина интервала** (ее еще часто называют интервальной разностью) представляет собой разность между верхней и нижней границами интервала.

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

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

личина равного интервала определяется по следующей формЯе:

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

о = у - Х,,,:,,, т. е. размах вариации;
где К- ^пгах I»"" ' К •* ,
Y *Х* • - *максимальное* и минимальное значения признака в со-
А,,,ах' Л""» к
вокулности.

Прежде чем определять размах вариации, из совокупности оекомендуется исключить аномальные наблюдения.

Если максимальные или минимальные значения сильно отличаются от смежных с ними значений вариантов в упорядоченном ряду значений группировочного признака, для определения величины интервала следует использовать не максимальное и минимальное значения, а значения, несколько превышающие минимум и несколько меньшие, чем максимум.

Полученную по формуле (3.2) величину округляют. Она является шагом интервала.

Существуют следующие правила определения шага интервала.

Если величина интервала, рассчитанная по формуле (3.2), представляет собой величину, имеющую один знак до запятой (например, 0,66; 1,372; 5,8), то полученные значения целесообразно округлить до десятых и их использовать в качестве шага интервала. В приведенном выше примере шагом интервала будут соответственно значения 0,7; 1,4; 5,8.

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

В случае когда рассчитанная величина интервала представляет собой трехзначное, четырехзначное и так далее число, эту величину необходимо округлить до бдижайшего числа, кратного '00 или 50. Например, 248 следует округлить до 250.

Рассмотрим пример. Пусть требуется произвести группировку с равными интервалами предприятий по стоимости основных фондов, при этом максимальное значение признака равно 2040 млн руб., а минимальное его значение - 290 млн. руб. Совокупность включает 80 единиц. Согласно формуле (3.1) она должна °ыть разбита на 7 групп. Сначала следует *найти:*

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

Затем определим величину интервала:

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

После этого построим интервалы групп (табл. 3.8).

Таблица 3j

Варианты построения групп

№ группы	I вариант	II вариант
I	От 290 до 540	До 540
II.	540 790	540 - 790
III	790 1040	790 - 1 040
IV	1 040 1 290	1 040 - 1 290
V	1 290 1 540	1 290 - 1 540
VI	1 540 1 790	1 540 - 1 790
VII	1 790 2 040	1 790 и более

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

Особенностью первого варианта построения групп является то, что у всех групп имеются закрытые интервалы. Во втором варианте первая и последняя группы - это группы с открытыми интервалами.

Открытые - это те интервалы, у которых указана только одна граница: верхняя - у первого, нижняя - у последнего. Например, открытыми будут первый и последний интервалы в группировке населения по размеру среднедушевого дохода (табл. 3.3).

Закрытыми называются интервалы, у которых обозначены обе границы.

Ширина открытого интервала принимается равной ширине смежного с ним интервала.

При группировке по количественному признаку границы интервалов могут быть обозначены по-разному. Если основанием группировки служит непрерывный признак, то одно и то же значение признака выступает и верхней, и нижней границами у двух смежных интервалов. Таким образом, верхняя граница 1-го интервала равна нижней границе i + 1-го интервала. Примером такой группировки является приведенная ранее группировка предприятий по стоимости основных фондов.

г\ и таком обозначении границ может возникнуть вопрос, в ' рпуппу включать единицы объекта, значения признака у К пых совпадают с границами интервалов. Например, во вто- Ко ю или третью группу должно войти предприятие со стоимосю фондов 790 млн руб. Если нижняя граница формируется по ть цИПу «включительно», а верхняя - по принципу «исключиельно», то предприятие должно быть отнесено к третьей групв противном случае - ко второй. Для того чтобы правильно отнести к той или иной группе единицу объекта, у которой значение признака совпадает с границами интервалов, можно использовать открытые интервалы. Так, единицы объекта (табл. 3.3), у которых размер среднедушевого денежного дохода равен 40 тыс. руб., попали во вторую группу (так как верхняя граница первой группы построена по принципу «исключительно»), а 80 тыс. руб. - в третью группу и т. д.

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

Например, пусть совокупность состоит из 80 предприятий и ее надо разделить на группы по численности занятых. Минимальное и максимальное значения группировочного признака соответственно равны 290 и 2040 человек. В этом случае возможны следующие варианты построения групп (табл. 3.9).

Таблица 3.9

Варианты построения групп

№ группы	I вариант	11 вариант
I	290 - 540	До 541
II	541 -790	541 -790
III	791 - 1040	791 - 1040
IV	1041-1290	1041 - 1290
V	1291 -1540	1291 -1540
VI	1541 -1790	1541 - 1790
VII	1791 -2040	1 79 1 и более

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

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

$$h_{i+}, =H(+ a;' \qquad (3.3)$$

в геометрической прогрессии:

$$hi,, = \wedge \bullet q, \qquad (3.4)$$

где а-константа - число, которое будет положительным при прогрессивно возрастающих интервалах и отрицательным при прогрессивно убывающих интервалах;

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

Например, если необходимо построить группировку предприятий отрасли по показателю выручки от реализации продукции, который варьирует от 500 млн руб. до 4 000 млн руб., то строить группировку с равными интервалами нецелесообразно, пиюму что, как правило, совокупность предприятий любой отрасли промышленности, торговли включает большое число малых предприятий, имеющих небольшую выручку. С ростом выручки от реализации продукции значительно снижается число предприятий. Таким образом, распределение числа предприятий по величине выручки является неравномерным. Поэтому следует построить группировку с неравными интервалами (табл. 3.!0).

Таблица 3.10 Группировка с неравными

интервалами

№ группы	Интервал
I	500 - 800
II	800- 1300
HI	1300-2000
IV	2000-290<>
V	2900- 4001;

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

п и определении границ интервалов статистических группи-
к исходят из того, что изменение количественного признака
Р°в дит к появлению нового качества. В этом случае граница
ПР е(:)вала устанавливается там, где происходит переход от одно-
ин ячества к другому. Рамки границ зависят от условий места и
г емени. Например, группировка предприятий по числу занятых
в называет, что в промышленности и строительстве предприя-
со среднесписочной численностью работающих 7 5 - 1 0 0
человек относятся к группе малых предприятий, а в отраслях
непроизводственной сферы и в розничной торговле - к крупным.

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

При изучении социально-экономических явлений на макроуровне часто применяют группировки, интервалы которых не будут ни прогрессивно возрастающими, ни прогрессивно убывающими. Такие интервалы называются **произвольными.** Например, при обработке материалов переписи населения 1989 г. для группировки семей и одиночек по размеру жилой площади, приходящейся на одного человека (м^2), применялись следующие группы: до 5; 5 - 6; 7 - 8; 9 - 12; 13-14; 15 - 19; 20 и более. Произвольные интервалы часто используются при группировке рабочих по выработке продукции, предприятий по уровню рентабельности.

Группировка с произвольными интервалами может быть построена с помощью коэффициента вариации, определяемого по формуле (7.15), V = x/o • 100.

Построение группировки этим методом начинается с упорядочения единиц совокупности по возрастанию или убыванию фуппировочного признака. В полученном ряду значений признака первые его значения объединяются в группу до тех пор, пока исчисленный для этой группы коэффициент вариации не станет равен 33%. Это будет свидетельствовать об образовании первой Фуппы, которая исключится из исходной совокупности. Оставшаяся ее часть принимается за новую совокупность, для которой повторяется алгоритм образования новой группы. И так до техП°Р, пока все единицы совокупности не будут объединены в ГРУппы.

Особенностью данного способа проведения группировки яв ляется то, что заранее, до проведения группировки, исследоВа тель не знает ни количество групп, ни границы интервалов.

После определения группировочного признака и границ груПп строится ряд распределения.

3.5

РЯДЫ РАСПРЕДЕЛЕНИЯ И ГРУППИРОВКИ

Статистический ряд распределения - это упорядоченное распределение единиц совокупности на группы по определенному варьирующему признаку.

В зависимости от признака, положенного в основу образования ряда распределения, различают атрибутивные и вариационные ряды распределения.

Атрибутивными называют ряды распределения, построенные по качественным признакам. Ряд распределения принято оформлять в виде таблиц. Ниже приведем атрибутивный ряд распределения юридической помощи адвокатов гражданам. Представленный в табл. 3.11 ряд показывает, как общее число случаев юридической помощи адвокатов распределялось по видам и формам правовой помощи в 1994 г.

Таблица 3.11

Распределение видов юридической помощи,
оказанной адвокатами гражданам одного из регионов РФ в 1994 г. (цифры условные)

№ п/п	Вид юридической помощи, оказанной адвокатами	Число случаев юридической помощи	
		всего, тыс.	% к итогу
1	Устные советы	5 109	69,43
2	Составление документов	991	13,47
3	Поручения по ведению		
4	уголовных дел	1 021	13,87
	Поручения по ведению		
	гражданских дел	238	3,23
	Всего	7359	100,00

$_{ем}$ентами этого ряда распределения являются значения ат-

, $_{ивног}$о признака, представленного названиями видов праР$^{и\ у}$~ $_{п}$омощи, оказанной адвокатами, и числа случаев, относя^{В} чся к каждому виду и форме помощи. Наибольший удельный ^{Щ} (почти 79%) приходится на оказание юридической помощи в виде устных советов.

Атрибутивные ряды распределения характеризуют состав совокупности по тем или иным существенным признакам. Взятые за несколько периодов, эти данные позволят исследовать изменение структуры.

Вариационными называют ряды распределения, построенные по количественному признаку. Любой вариационный ряд состоит из двух элементов: вариантов и частот. **Вариантами** считаются отдельные значения признака, которые он принимает в вариационном ряду, т е. конкретное значение варьирующего признака. **Частоты** - это численности отдельных вариантов или каждой группы вариационного ряда, т. е. это числа, показывающие, как часто встречаются те или иные варианты в ряду распределения. Сумма всех частот определяет численность всей совокупности, ее объем. **Частостями** называются частоты, выраженные в долях единицы или в процентах к итогу. Соответственно сумма частостей равна 1 или 100%.

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

Как известно, вариация количественных признаков может быть дискретной (прерывной) или непрерывной.

В случае дискретной вариации величина количественного признака принимает только целые значения. Следовательно, **дискретный вариационный ряд** характеризует распределение единиц совокупности по дискретному признаку. Примером дискретного вариационного ряда является распределение семей по числу комнат в отдельных квартирах, приведенное в табл. 3.12.

В первой колонке таблицы представлены варианты дискретного вариационного ряда, во второй - помещены частоты вариационного ряда, а в третьей - показаны частости.

^ случае непрерывной вариации величина признака у единиц совокупности может принимать в определенных пределах лю^{оЫ}е значения, отличающиеся друг от друга на сколько угодно малую величину. Построение **интервальных вариационных** . РЯДОВ целесообразно прежде всего при непрерывной вариации

Таблица 3] т

Распределение семей по числу занимаемых комнат
в отдельных квартирах в 1989 г. в РФ
(по данным переписи населения)

№ п/п	Группы семей, проживающих в квартирах с числом комнат	Число семей	
		всего, тыс. ед.	% к итогу
1	1	4064	16,3
2	2	12399	49,7
3	3	7659^	30,7
4	4 и более	832	3,3
	Всего	24954	100,0

признака, а также если дискретная вариация проявляется в широких пределах, т. е. число вариантов дискретного признака достаточно велико. В табл. 3.3 представлен интервальный вариационный ряд.

Удобнее всего ряды распределения анализировать при помощи их графического изображения, позволяющего судить и о форме распределения. Наглядное представление о характере изменения частот вариационного ряда дают полигон и гистограмма.

Полигон используется при изображении дискретных вариационных рядов. Для его построения в прямоугольной системе координат по оси абсцисс в одинаковом масштабе откладываются ранжи-

Таблица 3.13

Распределение жилого фонда городского района по типу квартир (цифры условные)

№ п/п	Группы квартир по числу комнат	Число квартир, тыс. ед.
1	1	10
2 .	2	35
3	3	30
4	4	15
5	5	5
	Всего	95

иные значения варьирующего признака, а по оси ординат наноР° ' щкаладля выражения величины частот. Полученные на пере^{cl}^$_{t\ ии}$ абсцисс и ординат точки соединяются прямыми линиями, ^{oe} езультате этого получают ломаную линию, называемую полиго^{B} м частот. Иногда для замыкания полигона предлагается крайние " чки (слева и справа на ломаной линии) соединить с точками на $^{T}{}_{си}$ абсцисс. В этом случае получается многоугольник.

Например, изобразим графически распределение жилого фонпа $_{п}$о типу квартир, представленных в табл. 3.13 (рис. 3.!).

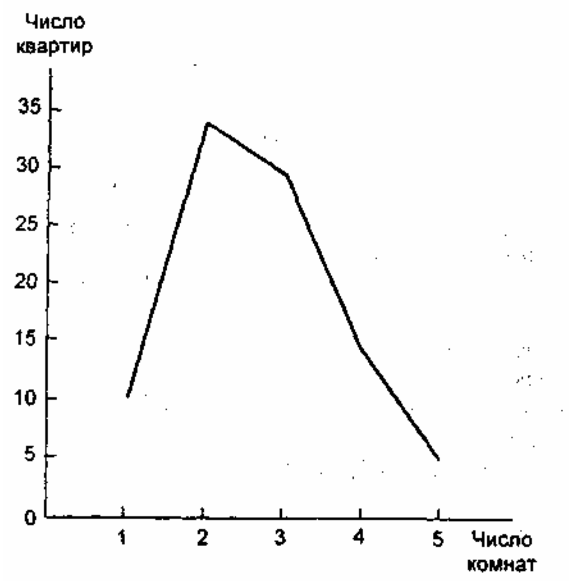


Рис. 3.1. Полигон распределения жилого фонда городского района по типу квартир

На оси ординат могут наноситься не только значения частот, но и частостей вариационного ряда.

Гистограмма применяется для изображения интервального вариационного ряда. При построении гистограммы на оси абсцисс откладываются величины интервалов, а частоты изображайся прямоугольниками, построенными на соответствующих интервалах. Высота столбиков в случае равных интервалов должна быть пропорциональна частотам. В результате мы получим гистограмму - график, на котором ряд распределения изображен в виде смежных друг с другом столбиков.

Изобразим графически интервальный ряд распределения, при. веденный в табл. 3.14 (рис. 3.2).

Таблица 3.14

Распределение семей по размеру жилой площади, приходящейся на одного человека (цифры условные)

№ п/п	Группы семей по размеру жилой площади, приходящейся на одного	Число семей с данным размером жилой	Накопленное число семей

	человека, м^2	площади	
1	3-5	10	10
2	5-7	20	30
3	7—9	30	70
4	9-11	40	100
5	11-13	50	115
	Всего	115	-

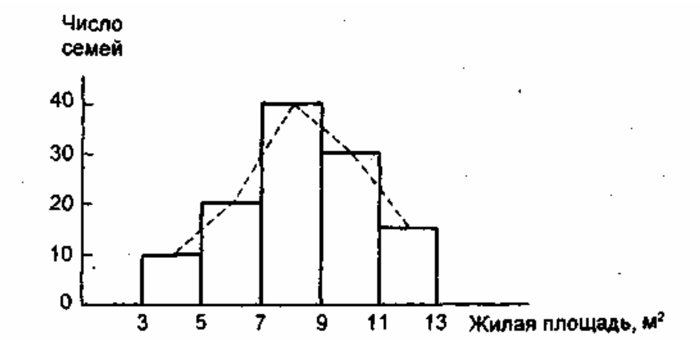


Рис. 3.2. Гистограмма распределения семей по размеру жилой площади, приходящейся на одного человека

Гистограмма может быть преобразована в полигон распределения, если найти середины сторон прямоугольников и затем эти точки соединить прямыми линиями. Полученный полигон распределения изображен на рис. 3.2 пунктирной линией.

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

валах. Это необходимо сделать для устранения влияния $^{\text{ИНТ}}$ чины интервала на распределение и получения возможнос$^{\text{ВеЛ}}$ павнивать частоты. **Плотность распределения -** это часто™ рассчитанная на единицу ширины интервала, т. е. сколько $^{\text{ТЗ'}}$ ниц в каждой группе приходится на единицу величины ин$^{\text{еД}}$ вала. Пример расчета плотности распределения представлен $_{\text{в}}$ табл. 3.15.

Таблица 3.15

Распределение предприятий по числу занятых (цифры условные)

№ п/п	Группа предприятий по числу занятых человек	Число предприятий	Ширина интервала, человек	Плотность распределения, единиц
	А	1	2	3 = 1 : 2
1	До 20	15	20	0,75
2	20-80	27	60	0,25
3	80-1 50	35	70	0,5
4	150-300	60	150	0,4
5	300 - 500	10	200	0,05
	Всего	147	-	-

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

При построении кумуляты интервального вариационного ряда °о оси абсцисс откладываются варианты ряда, а по оси ординат накопленные частоты, которые наносят на поле графика в виде перпендикуляров к оси абсцисс в верхних границах интервалов. $^{\text{атем}}$ эти перпендикуляры соединяют и получают ломаную ли$^{\text{ни}}$к>, т. $_{\text{е}}$. кумуляту.

Используя данные накопленного ряда (табл. 3.14), построим $^{\text{к}}$Умуляту распределения (рис. 3.3).

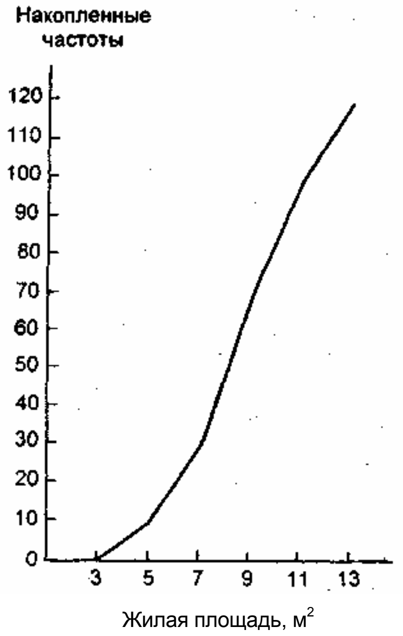


Рис. 3.3. Кумулята распределения семей по размеру жилой площади, приходящейся на одного человека

Изображение вариационного ряда в виде кумуляты особенно эффективно для вариационных рядов, частоты которых выражены в долях или процентах к сумме частот ряда, принятой соответственно за единицу или за 100%, т. е. частостями.

Если при графическом изображении вариационного ряда в виде кумуляты оси поменять местами, то мы получим **огиву.** На рис. 3.4 приведена огива, построенная на основе данных табл. 3.14.

С помощью кумулятивных кривых графически изображают процесс концентрации.

Широкое применение современных ЭВМ облегчает как построение рядов распределения, так и их графическое представление. Особо в этой связи следует отметить использование стандартизированных процедур определения величины интервала.

Ряд распределения представляет собой простейшую группировку, в которой каждая выделяемая группа характеризуется одним показателем - численностью единиц объекта, попавших в каждую группу. Построение рядов распределения является составной частью сводной обработки данных, при которой каждая группа единиц характеризуется многими показателями. Поэтому важным моментом в построении группировки является перечень тех показателей, которыми будет характеризоваться каждая группа.

Рис. 3.4. Огива распределения семей по размеру жилой площади, приходящейся на одного человека

Состав таких показателей формируется в соответствии с целями статистического исследования и задачами группировки. Для получения обобщенной, комплексной характеристики социально-экономического явления используют не отдельные показатели, а систему статистических показателей, которая предусматривает исчисление абсолютных, относительных и средних величин.

3.6

СРАВНИМОСТЬ СТАТИСТИЧЕСКИХ ГРУППИРОВОК

Группировки, построенные за один и тот же период времени, но для разных регионов или, наоборот, для одного региона, но за Д$^{\text{в}}$а разных периода времени, могут оказаться несопоставимыми $^{\text{из}}$-за различного числа выделенных групп или неодинаковости $^{\text{г}}$раниц интервалов. Для того чтобы привести такие группировки $^{\text{к}}$ сопоставимому виду (это позволяет провести их сравнитель-

$^{\text{ь'}}$и анализ), используется метод вторичной группировки. Суть Метода состоит в перегруппировке единиц объекта без обраще-

$^{\text{ия}}$ к первичным данным.

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

Применяют два способа образования новых групп. Первы\, наиболее простым и распространенным способом является объ$_{\text{е}}$^ **динение первоначальных интервалов.** Он используется в случае перехода от мелких к более крупным интервалам, а также когда границы новых и старых интервалов совпадают. Второй способ получил название **долевой перегруппировки** и состоит в образовании новых групп на основе закрепления за каждой группой определенной доли единиц совокупности. Этот способ употребляется, когда необходимо в ходе перегруппировки данных определить, какая часть (доля) единиц совокупности перейдет из старых групп в новые.

Рассмотрим первый способ проведения вторичной группировки.

Возьмем две группировки кредитов по сроку выдачи за ноябрь и декабрь (табл. 3.16 и 3.17).

Таблица 3.16

Группировка кредитов коммерческих банков по сроку выдачи, ноябрь 1995 г. (данные условные)

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

При анализе результатов двух группировок прежде всего их необходимо привести к сопоставимому виду, перегруппировав данные первой группировки. Для этого данные (табл. 3.16) первой и второй групп объединяют вместе, образовав одну группу краткосрочных кредитов. В эту группу включаются все кредиты, выданные в ноябре на срок от 1 до 6 месяцев. Данные третьей и четвертой групп полностью переносятся в табл. 3.18, в которой представлены результаты вторичной группировки кредитов коммерческих банков, выданных в ноябре и декабре.

При анализе результатов двух группировок прежде всего их необходимо привести к сопоставимому виду, перегруппировав данные первой группировки. Для этого данные (табл. 3.16) пер-

Таблица 3.17

пировка кредитов коммерческих банков по сроку выдачи, декабрь 1995 г. (данные условные)

- --- Xs п/п	--- ' -- ' Группы кредитов по сроку выдачи, месяцев	Число заключенных договоров, % от их общего количества	Сумма выданных кредитов, % от общей суммы
.———'	Краткосрочные (1-6)	86,54	97,91
2	Среднесрочные (6-12)	1,92	1,70
3	Долгосрочные (более 12)	11,54	0,39
__ ---	Итого	100,00	100,00

вой и второй групп объединяют вместе, образовав одну группу краткосрочных кредитов. В эту группу включаются все кредиты, выданные в ноябре на срок от 1 до 6 месяцев. Данные третьей и четвертой групп полностью переносятся в табл. 3.18, в которой представлены результаты вторичной группировки кредитов коммерческих банков, выданных в ноябре и декабре.

Табл и ца 3.18

Группировка кредитов коммерческих банков по сроку выдачи, ноябрь —декабрь 1995 г. (данные условные)

№ п/п	Группы кредитов по сроку выдачи, месяцев	Число заключенных договоров, % от их общего количества		Сумма выданных кредитов, % от общей суммы	
------- . ---		ноябрь	декабрь	ноябрь	декабрь
1	Краткосрочные (•1-6)	97.48	86.54'	91.73	97.91
2	Среднесрочные (6-12)	1.80	1.92	8.17	1.70
3 ~——------	Долгосрочные (более 12)	0,72	11,54	0,10	0,39
	Итого	100,00	100,00	100,00	100,00

Теперь можно сравнить итоги группировки ноябрьских и д$_е$, кабрьских кредитов. Доля заключенных договоров по краткое, рочным кредитам снизилась почти на 11 процентных пунктов доля среднесрочных кредитов осталась без изменения, а количе^ ство долгосрочных кредитов в анализируемом периоде значительно возросло. Несмотря на эти изменения, в декабре, так же как и ноябре, в структуре суммы выданных кредитов преобладающую долю занимали краткосрочные кредиты, затем следовали среднесрочные, а на последнем месте - долгосрочные кредиты.

Более сложным является способ долевой перегруппировки. Применение его рассмотрим на следующем *примере.* В табл. 3.19 приведены данные о распределении семей по размеру площади, приходящейся на одного человека по двум регионам.

Таблица 3.19

Группировка семей по размеру жилой площади, приходящейся на одного человека по двум-регионам в 1995 г.

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

Как ьидно из табл. 3.19, семьи первого региона разбиты на семь групп, а второго - на пять. Чтобы привести данные к сопоставимому виду, произведем перегруппировку семей второго регионаДля этого придется раздробить группы. Так как границы первой группы одинаковы у обеих группировок, то проведение каких-либо изменений нецелесообразно. Вторую группу (5-10) необходим^0

лить на три группы: семьи, в которых на одного человека Р^{а3}^$_{одитс}$я 5 и 6 м^2, должны образовать вторую группу; семьи, $^{П|7И}$ца человека приходится 7 и 8 м^2, - третью группу, а где 9 - 10 ^{Г}^$^{£}$ ледует включить в четвертую группу. Таким образом, вторую ^{М} ППУ в группировке семей второго региона следует разбить на и равные по величине интервала группы. При разбивке семей ^{т,7} $_{Г}$руппам полагают, что их распределение внутри группы 5-10 "авномерное. Тогда 1/3 семей группы 5-10 войдет в группу 5-6, \n

_ в группу 7 - 8, а оставшаяся часть должна быть включена в группу 9-12. Кроме того, в эту группу следует включить и часть семей из следующей третьей группы (11 - 15), т. е. семьи, в которых приходится 11 и 12м^2 жилой площади на одного человека. Поэтому 40% семей третьей группы надо включить в группу 9-12. Для составления группы 13 — 14 необходимо взять 40% семей группы 11 - 15. В группу 15-19 войдут оставшиеся 20% семей группы 11 - 15, т. е. семьи, в которых приходится на одного человека 15 м^2, и все семьи группы 16 - 19. Перегруппировка последней группы, как и первой, не нужна. Результаты перегруппировки представлены в табл. 3.20.

Таблица 3.20

Группировка семей по размеру жилой площади, приходящейся па одного человека по двум регионам в 1995 г.

№ группы	Группы семей по размеру жилой площади, приходящейся на одного человека, м^2	Доля семей в процентах к итогу	
		первый регион	второй регион
1	. До 5	3,6	6,2
2	5-6	11,4	1/3x46,3=15,43
o i	7-8	19,4	1/3x46,3=15,43
4	9-12	37,8	(46,3-2x15,43)+ +(0,4x28,5)=26,84
5	13- 14	11,1	0,4x28,5=11,4
6	15-19	13,0	0,2x28,5+10,8=16,5
7	20 и более	3,7	8,2
^-—•—	Итого	100	100

Из приведенной группировки следует, что размер жилой пло$^{Ди\ н}$а одного человека во втором регионе менее дифференциР°ван, чем в первом.

3.7

МЕТОД ГРУППИРОВОК И МНОГОМЕРНЫЕ КЛАССИФИКАЦИИ

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

Аналитические группировки, построенные по одному признаку, и сложные группировки позволяют установить связь и определить направление между результативными и 1 - 3-факторными признаками. Но часто этого бывает недостаточно, так как в действительности на изменение величины результативного признака оказывает влияние множество факторов, действующих в разных направлениях. Для исследования таких многофакторных связей используются многомерные группировки. Целью таких группировок является расчленение совокупности социально-экономических явлений на качественно однородные группы по большому числу признаков одновременно и определение на их основе связи и влияния факторных признаков на результативный. В основу построения многомерной группировки положен принцип перехода от величин, имеющих определенную размерность (рубли, тонны, гектары и т. д.), к безразмерным относительным ве-

Таблица 3.21

Матрица отношений

№ п/п	Результативный признак	Факторные признаки				
	Yi У	x: ч xJ	x: *2	X: _'l x.i		x: ч
1	Q,	P„	P,2	P,з		p·k
2	Qa	p	P>2	P$_2$з		n
1 j	Q,	P3,	P„	n»		n»
п	Q	p.	P,	p,		pk

ru I I

., Так абсолютные значения результативного признака личинам- > у, _ £у
нтгся отношениями Q, = 4г, где у = —, а абсолютные заме^{ня1}^у п
^{x},iения факторных признаков -
отношениями Ру=_1, где

ТХ..
_ _.._Л. В результате такой замены получается матрица отно^{X}J$^{=}$ п$_{ш}$ений(табл. 3.21).
Если связь между результативным и факторным (Xj) признакаобратная, то для каждой единицы объекта исследования определяется величина — .

ЕР На основе отношений Р^ исчисляется

показатель р$_{(}$. = —5L,

ЕС

$_{где}$ к _ число факторных признаков.

Этот показатель и будет основанием многомерной группировки, которая покажет взаимосвязь между множеством исследуемых факторных и одним результативным признаками.

На основе многомерной группировки можно построить уравнение регрессии, количественно отражающее степень связи между признаками.

3.8

ГРУППИРОВКИ И КЛАССИФИКАЦИИ В ПРАКТИКЕ СТАТИСТИКИ

Группировки как метод исследования широко используются ^{8} практике статистики.

"ри анализе формирования рыночных отношений большое значение имеет группировка предприятий по численности заня^{Т}bix в отраслях производственной и непроизводственной сфер с обязательным выделением группы малых предприятий, группировка предприятий по формам собственности, организационноправовым формам.

Ьольшое значение играет и группировка по экономическому начению продукции. Так, при анализе продукции промышчности используют группировку, позволяющую выделить из всего объема продукции производство средств производства производство предметов потребления.
"

В статистике розничной торговли используется группиров товаров на продовольственные и непродовольственные, в стати^{Э}тике сельского хозяйства - группировка продукции на проду$_{К}$ц$_{и}$^ растениеводства и животноводства, а внутри этих групп - делены продукции по ведущим производственным направлениям.

Среди группировок, применяемых отечественной статистикой особое место принадлежит группировке по отраслям народного хозяйства. Она используется при анализе валового внутреннего продукта, валового национального дохода, капитальных вложений и ввода в действие основных фондов, структуры ввоза и вывоза продукции.

Кроме группировок в практической статистике широко применяются **классификации.** Существует много различных классификаций. В макроэкономической статистике применяют классификацию отраслей народного хозяйства, основных фондов; в статистике капитального строительства - классификацию капитальных вложений и строительных машин; в статистике труда классификацию профессий, а в статистике внешней торговли классификацию, называемую «Товарная номенклатура внешнеэкономической деятельности».

На этапе "перехода от одной формы хозяйствования к другой возникла потребность в новых классификаторах, таких, как классификатор форм собственности и организационно-правовых форм хозяйствующих субъектов.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 3

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

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

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

Структурная группировка - разделение однородной совокупности на группы, характеризующие ее структуру по какому-либо варьирУ^{10}' щему признаку.

Аналитическая группировка - группировка, выявляющая взаимО' .связи между изучаемыми явлениями и их признаками.

$_{\text{ров}}$**очный признак** - признак, по которому производится П*^{3}"^{1} $_{\text{един}}$иц совокупности на отдельные группы, раз&ие $_{\text{сн}}$**ф**$_{\text{ика}}$**ц**$_{\text{и}}$**я** - систематическое распределение явлений и объек-
лппеделенные группы, классы, разряды на основании их сходтов $^{\text{на}}$ " т,,а и различия.

и тервал - значения варьирующего признака, лежащие в опреде$_{\text{лен}}$ных границах.

Величина интервала - разность между верхней и нижней границами интервала.

Открытые интервалы - интервалы, у которых указана только одна

граница.

Закрытые интервалы — интервалы, у которых обозначены обе границы.

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

Атрибутивный ряд распределения - ряд, построенный по качественному признаку.

Вариационный ряд распределения - ряд, построенный по количественному признаку.

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

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

Дискретный вариационный ряд - распределение единиц совокупности по дискретному признаку.

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

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

ТЕСТ К ГЛАВЕ 3

1- Группировка, в которой происходит разбиение однородной совокупности на группы, называется:
а) типологической группировкой;
б) структурной группировкой;
$^{\text{в}}$) аналитической группировкой.

2. По технике выполнения статистическая сводка делится на:
а) простую и сложную;
°j Централизованную и децентрализованную;
$^{\text{в}}$) механизированную и ручную.

3- Основанием, группировки может быть:
а) качественный признак;
б) количественный признак;
$^{\text{в}}$) как качественный, так и количественный признак.

4. Особое внимание нужно обратить на число единиц иссл$_{\text{е}}$мого объекта, если основанием группировки выбран:
а) качественный признак;
б) количественный признак;
в) как качественный, так и количественный признак.

5. Наибольшее значение признака в интервале называется-
а) нижней границей;
б) верхней границей интервала.

6. Величина равного интервала определяется по формуле: а) h_{H}, = hj+a;
6)$h_{i+1} = h_i$-q;

в) h = —.
n

7. Если величина интервала равна 0,5ст, то совокупность разбивается на:
а) 6 групп;
б) 9 групп;
в) 12 групп.

8. При непрерывной вариации признака целесообразно построить:
а) дискретный вариационный ряд;
б) интервальный вариационный ряд;
в) ряд распределения.

9. Накопленные частоты используются при построении:
а) огивы;
б) гистограммы;
в) полигона.

10. Если две группировки несопоставимы из-за различного числа выделенных групп, то они могут быть приведены к сопоставимому виду:
а) с помощью метода вторичной группировки;
б) путем построения сложной группировки.

ГЛАВА 4

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

4.1

ПОНЯТИЕ О СТАТИСТИЧЕСКОЙ ТАБЛИЦЕ. ЭЛЕМЕНТЫ СТАТИСТИЧЕСКОЙ ТАБЛИЦЫ

Результаты сводки и группировки материалов статистического наблюдения, как правило, излагаются в виде таблиц.

Таблица является наиболее рациональной, наглядной и компактной формой представления статистического материала.

Однако не всякая таблица является статистической. Таблица умножения, опросный лист социологического обследования и т. д. могут носить табличную форму, но еще не являются статистическими таблицами.

Статистическую таблицу от других табличных форм отличает следующее:

- она должна содержать результаты подсчета эмпирических данных;
- она является итогом сводки первоначальной информации.

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

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

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

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

Статистическая таблица содержит три вида заголовков: общий, верхние и боковые. *Общий* заголовок отражает содержание всей таблицы (к какому месту и времени она относится), располагается над ее макетом *по* центру и является внешним заголовком, *Верхние* заголовки характеризуют содержание граф, а *боковые* строк. Они являются внутренними заголовками.

Название таблицы*
(общий заголовок)

Содержание строк	Наименование граф (верхние заголовки)					
А	1	2	3	4	5	~
Наименование строк (боковые заголовки)						
						" 1
Итоговая строка						— Итоговая графа

* Примечание к таблице.

Рис. 4.1. Остов (основа) статистической таблицы

Остов таблицы, заполненный заголовками, образует ее макет. Если на пересечении граф и строк записать цифры, то получается полная статистическая таблица.

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

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

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

Сказуемое статистической таблицы образует система показателей, 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 *менять*^ местами, что зависит от достижения каждым исследователем отдельности наиболее полного и лучшего способа прочтения анализа исходной информации об исследуемой совокупности.

4.2

ВИДЫ ТАБЛИЦ ПО ХАРАКТЕРУ ПОДЛЕЖАЩЕГО

ктике $_{\text{экон}}$омико-статистического анализа используются иные виды статистических таблиц, отличающихся различР$^{\text{азЛ}}$ $_{\text{стро}}$ением подлежащего и сказуемого, структурой и соот$^{\text{НЬ}}$шением признаков, формирующих их.

В зависимости от структуры подлежащего и группировки в единиц объекта различают статистические таблицы *простые* ^{Н} *сложные,* а последние, в свою очередь, подразделяются на *,рупповые* и *комбинационные.*

В простой таблице в подлежащем дается простой перечень каких-либо объектов или территориальных единиц, т. е. в подлежащем нет группировки единиц совокупности. Простые таблицы бывают *монографические* и *перечневые.* Монографические таблицы характеризуют не всю совокупность единиц изучаемого объекта, а только одну какую-либо группу из него, выделенную по определенному, заранее сформулированному признаку (табл. 4.1).

Табл и ца

Характеристика итогов торгов облигаций федерального займа на ММВБ 23.03.98 г.

(цифры условные)

4.1

1	Количество сделок, шт.	Объем сделок, тыс. руб.	Цена сделок, % к номиналу	Доходность к погашению, %
Облигации федерального займа	9.4	8662	47,50	29,70

Перестроив подлежащее табл. 4.1, чтобы были показаны облигации федерального займа по номерам выпуска, т. е. показав каждую единицу совокупности, получаем перечневую таблицу (табл. 4.2).

Гаким образом, простыми перечневыми таблицами называют$^{\text{Ся}}$ таблицы, подлежащее которых содержит перечень единиц $^{\text{Из}}$Учаемого объекта.

'фостые таблицы не дают возможность выявить социально°номические типы изучаемых явлений, их структуру, а также аимосвязи и взаимозависимости между характеризующими их "Ризнаками.

Таблиц$_{\text{а}}$

Характеристика итогов торгов облигаций федерального займа на ММВБ 23.03.98 г.

(цифры условные)

ер выпуска ОФЗ	Количество сделок, шт.	Объем сделок, тыс. руб.	Цена сделок, % к номиналу	Доходность к погашению, »/
26001	5	120	50,00	30,15
26002	1	48	47,49	29,80
26003	88	8494	45,02	29,13
Всего	94	8662	47,50	29,70~~~~

Эти задачи более полно решаются с помощью сложных: трупповых и особенно комбинационных таблиц.

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

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

Таблица 4.3

Распределение коммерческих банков РФ по величине капитала на 01.01.98 г.

(цифры условные)

№ ' п/п	Группы коммерческих банков по величине капитала, млн руб.	Число коммерческих банков	Удельный вес банков, % к итогу
1	6-20	16	53,3
2	20-34	6	20,0
3	34-48	8	26,7
Итого		30	100,0

43 отражает количественное распределение коммерческийнков РФ по величине капитала.

кИ^г им образом, групповые таблицы позволяют выявить и теризовать социально-экономические типы явлений, их ^^TVPY В зависимости только от одного признака. СТ^**мбинационными** называются статистические таблицы,

ежащее которых содержит группировку единиц совокупноПад новременно по двум и более признакам: каждая из групп, оТ троенная по одному признаку, разбивается, в свою очередь, "а подгруппы по какому-либо другому признаку и т. д.

Таблица 4.4

Распределение коммерческих банков РФ по величине капитала и работающих активов на 01.01.98 г.

(цифры условные)

№ п/п	Группы коммерческих банков по величине капитала, млн руб.	Подгруппы коммерческих банков по величине работающих активов, млн руб.	Число коммерческих банков
1	6-20	10-47 47-84 84-121	151
	Итого по группе		16
2 3	20-34	10-47 47-84 84-121	2 4
	Итого по группе		б
	34-48	10-47 47-84 84-121	1 1 6
	Итого по группе		8
	Итого по подгруппам	10-47 47 - 84 84-121	186 6
	Всего		30

Комбинационные таблицы позволяют характеризовать т ческие группы, выделенные по нескольким признакам, ц ИГ1|(между последними. Последовательность разбиения единщ8"51' вокупности на однородные группы по признакам определи Со~ либо важностью одного из них в их комбинации, либо поряпТся их изучения.

Например, подлежащим в табл. 4.4 являются группы комм ческих банков по величине капитала и величине работают''' активов. Из данных табл. 4.4 видно, что между величиной кап"* тала и величиной работающих активов имеется определенна неярко выраженная зависимость.

Таким образом, комбинационные таблицы позволяют глубже раскрыть сущность и закономерность изучаемых социально-экокомических явлений и процессов.

4.3

ВИДЫ ТАБЛИЦ ПО РАЗРАБОТКЕ СКАЗУЕМОГО

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

По структурному строению сказуемого различают статистические таблицы с простой и сложной его разработкой.

При **простой разработке сказуемого** показатель, определяющий его, не подразделяется на подгруппы, и итоговые значения получаются путем простого суммирования значений по каждому признаку отдельно независимо друг от друга. Примером простой разработки сказуемого может служить следующий фрагмент статистической таблицы.

Распределение акций среди работников приватизированных предприятий промышленности

Предприятия	Приоб-ретено акций, всего	В том числе распределение по видам акций			
		приватизированные типа А	обыкновенные	на льготных условиях	по цене, определенной Госкомимуществом

заполнения данного фрагмента таблицы получается ^°йная характеристика приватизированных предприятий по Г10ДР^0 $_{е\ их}$ субъектов-владельцев. По каждому предприятию стрУ^к $_{ол}$Учить информацию о числе и ценовых условиях пропажи акций.

г о*ная разработка сказуемого предполагает деление приака, формирующего его, на подгруппы.

Распределение акций среди работников приватизированных предприятий промышленности

Предприятия	Приобретено акций, всего	В том числе распределение			
		на льготных условиях		по цене, определенной Госкомимуществом	
		привилегированные типа А	обыкновенные	привилегированные типа А	обыкновенные

При сложной разработке сказуемого получается более полная и подробная характеристика объекта.

Однако сложная разработка сказуемого может привести к безмерному увеличению размерности статистических таблиц, что, в свою очередь, снижает их наглядность, чтение и анализ.

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

4.4

ОСНОВНЫЕ ПРАВИЛА ПОСТРОЕНИЯ ТАБЛИЦ

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

Основные приемы, определяющие технику формирования ста^{пи}стических таблиц, следующие.

'• Таблица должна быть компактной и содержать только те ^{с}*одные данные, которые непосредственно отражают исследу-
^{06} социально-экономическое явление в статике и динамике и
ооходимы для познания его сущности.

Следует избегать ненужной, второстепенной, бессодерж$_а$ной к данному объекту исследования информации. Циф^^{1} материал необходимо излагать таким образом, чтобы при $_{ан}$ ^{Boi}^ зе таблицы сущность явления раскрывалась чтением строк $_С$л направо и сверху вниз. ^{8}

2. Заголовок таблицы и названия граф и строк должны бк четкими, краткими, лаконичными, представлять собой законч ' ное целое, органично вписывающееся в содержание текста

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

Если название таблицы состоит из двух и более предложений точка ставится с целью отделения предложений друг от друга но не после последнего.

В заголовках граф допускаются точки только при необходимых сокращениях.

В заголовке таблицы должны найти отражение объект, признак, время и место совершения события. Но при этом следует помнить, что чем более краток и лаконичен текст заголовка таблицы, тем она яснее и доходчивее для чтения и анализа, естественно, если это осуществляется не в ущерб ее точности и познавательности. Заголовки таблицы, граф и строк пишутся полностью, без сокращений.

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

• строка «Итого» или «Всего» завершает статистическую таблицу;

• итоговая строка располагается первой строкой таблицы и соединяется с совокупностью ее слагаемых словами «В том числе».

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

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

5. Если названия отдельных граф повторяются между собой, содержат повторяющиеся термины или несут единую смысл0' вую нагрузку, то им необходимо присвоить общий объединяю' щий заголовок.

Данный прием используется и для подлежащего, и для сказУ' емого таблиц.

Г фы и строки полезно нумеровать. Графы, слева заполе названием строк, принято обозначать заглавными букваНеИялфавита (А), (В) и т. д., а все последующие графы - номеми в порядке возрастания.

1 Взаимосвязанные и взаимозависимые данные, характеризу-
' одну из сторон анализируемого явления (например, число
пприятий и удельный вес заводов (в % к итогу), абсолютный
ПР ост и темп роста и т. д.), целесообразно располагать в со-
сеРдних друг с другом графах.

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

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

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

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

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

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

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

а) если данная позиция (на пересечении соответствующих графы и строки) вообще не подлежит заполнению, то ставится знак «Х»;

б) когда по какой-либо причине отсутствуют сведения, то ставится многоточие «...» или «Нет свед.», или «Н. св.»;

в) при отсутствии явления клетка заполняется тире («-») и остается пустой.

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

13. В случае необходимости дополнительной информаци разъяснений - к таблице могут даваться примечания.

Соблюдение приведенных правил построения и оформлен статистических таблиц делает их основным средством npeflCTaHliления, обработки и обобщения статистической информации состоянии и развитии анализируемых социально-экономически0 явлений.

4.5

ЧТЕНИЕ И АНАЛИЗ ТАБЛИЦЫ

Анализу статистических таблиц предшествует этап ознакомления - их чтения.

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

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

Анализ таблицы как метод научного исследования .путем разбиения предмета изучения на части делится на структурный и содержательный.

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

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

Содержательный анализ предполагает изучение внутреннего содержания таблицы: анализ отдельных групп подлежащего по соответствующим признакам сказуемого; выявление соотношения и пропорций между группами явлений по одному и разным признакам; сравнительный анализ и формулировку выводе^{8} по отдельным группам и по всей совокупности в целом; уста-

закономерностей и определение резервов развития "°*^*ого* объема.

$^{\text{и}}$ ^ де чем приступать к анализу числовой информации, ^Р$^{\text{е}}$ $_{\text{М0}}$ проверить ее достоверность и научную обосноч$^{\text{е}}$ $_{\text{ь}}$ Исследователь должен убедиться в достоверности и $^{\text{ванН}}$ $_{\text{ости}}$ источника информации данных и критически оце$^{\text{наде}}$ $_{\text{х}}$ цифровые значения. Следует произвести логическую $^{\text{нит}}$етную проверки данных. **Логическая проверка** состоит в $^{\text{И}\circ}$можности определения конкретных признаков теми или $^{\text{В}\circ}$ыми числовыми значениями (например, абсурдно, если Келейность работающих на фирме составила 106,7 человека) **Счетная проверка** предполагает выборочный расчет отдельных значений признаков по группе либо итоговых значений строк или граф и т. д.

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

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

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

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

Комплексный анализ таблиц, содержащих ряды динамики, дает возможность исследователю выявлять и обобщать тенденции и закономерности в развитии совокупности единиц наблюдения. Анализ групповых и комбинационных таблиц позволяет охарактеризовать типы социально-экономических явлений, структуру совокупности, соотношения и пропорции между отдельными группами и единицами наблюдения; выявить характер и направление взаимосвязей и взаимозависимостей между различными, определенными логикой экономического анализа, сочетаниями признаков и зависимости признаков-следствия от признаков-причин.

Соблюдение правил и последовательности работы со статисическими таблицами помогает исследователю осуществить научно обоснованный экономико-статистический анализ объектов $^{\text{и}}$ процессов.

4.6

ТАБЛИЦЫ И МАТРИЦЫ

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

Матрицей называется прямоугольная таблица числовой ин формации, состоящая из m строк и п столбцов. Таким образом матрица имеет размерность m x п:

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

где Эу — элемент матрицы, стоящий на пересечении i-й строки и j-ro столбца.

Различают два вида матриц:

- прямоугольная (размерность m х п);
- квадратная. Если число строк строго равно числу столбцов (т = п), то матрица называется квадратной порядка п.

Квадратная матрица порядка п называется **диагональной** (Д), если все элементы, стоящие вне главной диагонали (d,, d_2..., d_n), равны нулю.

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

Если в диагональной матрице Д все dj = 1, то матрица называется **единичной,** при $\mathbf{d_s = 0}$ **- нулевой.**

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

Таблицы-матрицы широко применяются на практике, например в экономике в виде балансово-нормативных моделей, отражаюШ$^{\text{иХ}}$ соотношение результатов производства, нормативов производственных затрат и т. д. Успешно используют матрицы и в межотраслевом балансе, системе национального счетоводства и т. д.

4.7

ТАБЛИЦЫ СОПРЯЖЕННОСТИ

Т блицей сопряженности называется таблица, которая содерсводную числовую характеристику изучаемой совокупности $^{\text{жи}}$ пвум $^{\text{и}}$ более атрибутивным (качественным) признакам или $^{\text{П}}$ мбинации количественных и атрибутивных признаков. $^{\text{К}}$ Таблицы сопряженности получили наибольшее распространение при изучении социальных явлений и процессов: общественного мнения, уровня и образа жизни, общественно-политического строя и т. д.

Наиболее простым видом таблиц сопряженности является *таблица частот* 2x2.

Общая схема таблицы частот 2x2

	B,	B_2	Всего
A,	fn	f.2	fio
A_2	; f»,	fn	f_M
Всего	fo,	fe	foo

Построение данной таблицы исходит из предположения, что ответы респондентов или анализируемые атрибутивные признаки будут принимать только два значения: А, и A_2, B(и B_2. Внутреннее цифровое наполнение таблицы представляют частоты (fy), обладающие одновременно i-M (1 = 1,2) значением одного (Aj) и j-м (j = 1,2) значением (ВЛ другого качественного признака.

Итоговая графа и строка содержат информацию о количественном распределении совокупности соответственно по А и В а«риоутивньш признакам. (Подробнее пример расчета и анализа на основе таблиц сопряженности см. в п. 9.7.)

Для более полного описания и анализа явлений и процес-$^{\text{сов}}$, характеризующихся атрибутивными признаками, исполь-УЮтся *таблицы сопряженности большей размерности: i* х *j,*
$^{\text{Де ı =}}$ 1,2, ..., k - число вариантов значений (например, ответов
Респондентов и т. д.) одного признака (например, признака А);
'> 2, ..., п - число вариантов значений другого признака (В).

Общая схема таблицы сопряженности большей размерности

	B,	B_2		**B.**	-------- Всего
A,	f.i	f.i		f.i	~—~~, f,o
A_2	f2,	f22		fcj	———, f* ____

Ai	fil	fe		fij	fio
Всего	fo,	fo2		fo,	Too

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

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 4

Статистическая таблица — способ рационального изложения и обобщения данных о социально-экономических явлениях при помощи цифр, расположенных в определенном порядке.

Подлежащее статистической таблицы характеризует объект исследования. В нем дается перечень единиц совокупности либо групп исследуемого объекта по существенным признакам.

Сказуемое статистической таблицы - система показателей, которыми характеризуется *объек-f* изучения.

Простая таблица — таблица, в подлежащем которой дается простой перечень объектов или территориальных единиц.

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

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

Простая разработка сказуемого — показатели в сказуемом даны параллельно один другому, без разделения на подгруппы.

Сложная разработка сказуемого - показатели в сказуемом даны в комбинации друг с другом.

Матрица — прямоугольная таблица числовой информации, состоящая из m строк и п столбцов.

ца **сопряженности** - таблица, которая содержит сводную ^Я ю характеристику изучаемой совокупности по двум и более числГ ивным признакам или комбинации количественных и атрибутив-

*$У_К$ признаков.

ч ние вышеперечисленных основных понятий позволит в достаточ$_0$Й мере разобраться в теме курса.

ТЕСТ К ГЛАВЕ 4

1 Статистическая таблица представляет собой:

а) форму наиболее рационального изложения результатов статистического наблюдения;

б) сведения о чем-нибудь, расположенные по строкам и графам;

в) числовые характеристики, размещенные в колонках таблицы.

2. К статистической таблице можно отнести:

а) таблицу умножения;

б) опросный лист социологического обследования;

в) таблицу, характеризующую численность населения по полу и возрасту.

3. По характеру разработки подлежащего различают статистические таблицы:

а) простые;

б) перечневые;

в) комбинационные.

4. Монографические таблицы характеризуют:

а) совокупность единиц изучаемого объекта;

б) группу единиц совокупности по признаку;

в) каждую единицу совокупности.

5. Подлежащее групповых статистических таблиц содержит:

а) перечень единиц совокупности по признаку;

б) группировку единиц совокупности по одному признаку;

^{в}) группировку единиц совокупности по нескольким признакам.

6- По характеру разработки сказуемого различают статистические таблицы:

^{а}) монографические;

^{б}) перечневые;

^{в}) сложные.

'• Сказуемым статистической таблицы является:

^{а}) исследуемый объект;

ч) показатели, характеризующие исследуемый объект;

^{в}) сведения, расположенные в верхних заголовках таблицы.

8. Имеются следующие данные по основным показат*„ тельности некоторых крупнейших банков РФ на 01.01.98 г Ч1И % условные). '""чф^.

№ банка	Сумма активов, млн руб.	Собственный капитал, млн руб.	Привлечем Ресурсы, млнруб
1	480,6	37,3	Ю,1
2	485,3	39,5	16,3
^> i	490,2	38,3	8,4
Итого	1456,1	115,1	3AsT^~^

Подлежащим таблицы являются:
а) номер банка РФ;
б) сумма активов;
в) собственный капитал;
г) привлеченные ресурсы.

9. Представлен макет статистической таблицы, характеризующий зависимость заработной платы рабочих фирмы от выполнения ими норм выработки продукции.

Группа рабочих по уровню заработной платы, руб.	Группа рабочих по проценту выполнения норм выработки	Число рабочих	Фонд заработной платы, тыс. руб.
До 1500	До 100 100 и более		
Итого по группе			
1500-2500	До 100 1 00 и более		
Итого по группе			- 1
Итого по подгруппам	До 100 100 и более		
Всего			

„reov подлежащего макет таблицы: по характер \ гоупповои, 55ЙбинаШ«оннь.и;

в) пйТлоиИца сопряженности строится по:
^количественным признакам;
3 ^£SS^^^и атрибутивных признаков,

ГЛАВА 5

ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ

5.1

ПОНЯТИЕ О СТАТИСТИЧЕСКОМ ГРАФИКЕ. ЭЛЕМЕНТЫ СТАТИСТИЧЕСКОГО ГРАФИКА

Современную науку невозможно представить без применения графиков. Они стали средством научного обобщения.

Выразительность, доходчивость, лаконичность, универсальность, обозримость графических изображений сделали их незаменимыми в исследовательской работе и в международных сравнениях и сопоставлениях социально-экономических явлений.

Впервые о технике составления статистических графиков упоминается в работе английского экономиста У. Плейфейра «Коммерческий и политический атлас», опубликованной в 1786 г. и положившей начало развитию приемов графического изображения статистических данных.

Трактовка графического метода как особой знаковой системы - искусственного знакового языка - связана с развитием *семиотики,* науки о знаках и знаковых системах.

Знак в семиотике служит символическим выражением некоторых явлений, свойств или отношений.

Существующие в семиотике знаковые системы принято разделять на неязыковые и языковые.

Неязыковые знаковые системы дают представление о явлениях окружающего нас мира (например, шкала измерительного прибора, высота столбика ртути в термометре и т. д.).

Языковые знаковые системы выполняют сигнальные функции, а также задачи сопоставления совокупностей явлений и их анализа. Характерно, что в этих системах сочетание знаков приобретает смысл только тогда, когда их объединение производится по определенным правилам.

В языковых знаковых системах различают естественные и искусственные системы знаков, или языков.

С точки зрения семиотики человеческая речь, выраженная знаками-буквами, составляет **естественный язык.**

стве**нные языковые системы** используются в различ-
бластях жизни и техники. К ним относятся системы мате-
нь'* еских, химических знаков, алгоритмические языки, графи-

кИ а исключая естественного языка, искусственные, или символеские языки упрощают изложение специальных вопросов Лпределенной области знаний.

° Таким образом, **статистический график** - это чертеж, на ором статистические совокупности, характеризуемые опреде-
к ными показателями, описываются с помощью условных геометрических образов или знаков. Представление данных табли-
1Ы в виде графика производит более сильное впечатление, чем цифры, позволяет лучше осмыслить результаты статистического наблюдения, правильно их истолковать, значительно облегчает понимание статистического материала, делает его наглядным и доступным. Это, однако, вовсе не означает, что графики имеют лишь иллюстративное значение. Они дают новое знание о предмете исследования, являясь методом обобщения исходной информации.

Значение графического метода в анализе и обобщении данных велико. Графическое изображение прежде всего позволяет осуществить контроль достоверности статистических показателей, так как, представленные на графике, они более ярко показывают имеющиеся неточности, связанные либо с наличием ошибок наблюдения, либо с сущностью изучаемого явления. С помощью графического изображения возможны изучение закономерностей развития явления, установление существующих взаимосвязей. Простое сопоставление данных не всегда дает возможность уловить наличие причинных зависимостей, в то же время их графическое изображение способствует выявлению причинных связей, в особенности в случае установления первоначальных гипотез, подлежащих затем дальнейшей разработке- Графики также широко используются для изучения структуры явлений, их изменения во времени и размещения в пространстве. В них более выразительно проявляются сравниваемые характеристики и отчетливо видны основные тенденции развития и взаимосвязи, присущие изучаемому явлению или процессу.

При построении графического изображения следует соблюдать РЯД требований. Прежде всего график должен быть достаточно наглядным, так как весь смысл графического изображения как
етода анализа в том и состоит, чтобы наглядно изобразить ста-
Истические показатели. Кроме того, график должен быть выразиТельным, доходчивым и понятным. Для выполнения вышепере-
численных требований каждый график должен включать ряд новных элементов: графический образ; поле графика; прости ° ственные ориентиры; масштабные ориентиры; экспликацию г "" фика. '

Рассмотрим подробнее каждый из указанных элементов. Гп **фический образ (основа графика)** - это геометрические знак т. е. совокупность точек, линий, фигур, с помощью котори изображаются статистические показатели. Важно правильн выбрать графический образ, который должен соответствовать цели графика и способствовать наибольшей выразительности изображаемых статистических данных. Графическими являются лишь те образы, в которых свойства геометрических знаков фигура, размер линий, расположение частей - имеют существенное значение для выражения содержания изображаемых статистических величин, причем каждому изменению выражаемого содержания соответствует изменение графического образа.

Поле графика - это часть плоскости, где расположены графические образы. Поле графика имеет определенные размеры, которые зависят от его назначения.

Пространственные ориентиры графика задаются в виде системы координатных сеток. Система координат необходима для размещения геометрических знаков в поле графика. Наиболее распространенной является система прямоугольных координат (рис. 5.18). Для построения статистических графиков используется обычно только первый и изредка первый и четвертый квадраты.

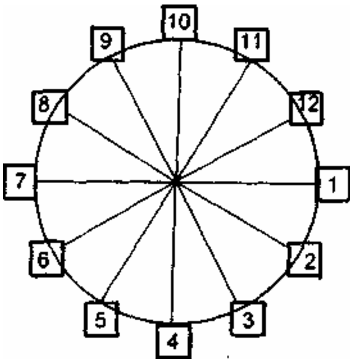


Рис. 5.1. Полярная система координат

В практике графического изображения применяются также полярные координаты. Они необходимы для наглядного изображения циклического движения во времени. В полярной системе координат (рис. 5.1) один из лучей, обычно правый горизонтальный, принимается за ось ординат, относительно которой определяется угол луча. Второй координатой считается ее расстояние от центра сетки, называемое **радиусом,** В радиальных графиках лучи обозначают момен ты времени, а окружности - величины изучаемого явления. На статистических картах пространствен$_{\text{ные}}$ ориентиры задаются контурной сеткой (контуры рек, береговая линей и океанов, границы государств) и определяют те тер$^{\text{Н1}}$^$^{\text{я}}$ м к которым относятся статистические величины. ^$^{\text{И}}$ *м* **сштабные ориентиры** статистического графика определя-маош1х^ттабом и системой масштабных шкал. Масштаб статикугся еского графика - это мера перевода числовой величины в графическую. .

Масштабной шкалой называется линия, отдельные точки опой могут быть прочитаны как определенные числа. Шкала $^{\text{к}}$ $_{\text{ет}}$ большое значение в графике и включает три элемента: инию (или носитель шкалы), определенное число помеченных точками $_{\text{т}}$очек, которые расположены на носителе шкалы в определенном порядке, цифровое обозначение чисел, соответствующих отдельным помеченным точкам. Как правило, цифровым обозначением снабжаются не все помеченные точки, а лишь некоторые из них, расположенные в определенном порядке. По правилам числовое значение необходимо помещать строго против соответствующих точек, а не между ними (рис. 5.2).

Рис. 5.2. Числовые интервалы

Носитель шкалы может представлять собой как прямую, так и кривую линии. Поэтому различают **шкалы прямолинейные** (например, миллиметровая линейка) и **криволинейные** - дуго$^{\text{В}}$bie и круговые (циферблат часов).

Графические и числовые интервалы бывают равными и неравными. Если на всем протяжении шкалы равным графическим интервалам соответствуют равные числовые, такая шкала назы$^{\text{Вается}}$ **равномерной. Когда** же равным числовым интервалам оответствуют неравные графические интервалы и наоборот,

кала называется **неравномерной.**

Масштабом равномерной шкалы называется **длина отрезка** ^ Рафический интервал), принятого за единицу и измеренного в

$^{\text{Ки}}$х-либо мерах. Чем меньше масштаб (рис. 5.3), тем гуще располагаются на шкале точки, имеющие одно и то же знамен Построить шкалу - это значит на заданном носителе щ$_{\text{Ка}}$ $^{\text{е}}$ разместить точки и обозначить их соответствующими числам^ согласно условиям задачи.
и

0 1
1 i i i i i i i i i i Масштаб 50 **мм**

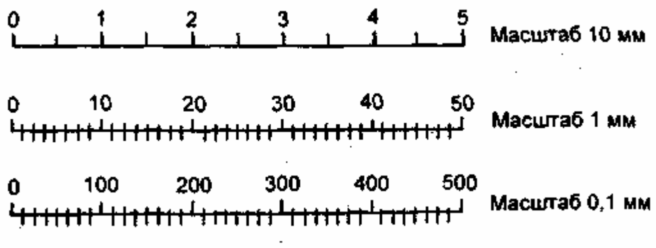
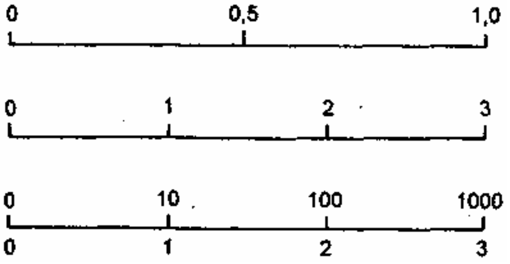


Рис. 5.3. Масштабы

Как правило, масштаб определяется примерной прикидкой возможной длины шкалы и ее пределов. Например, на поле в 20 клеток надо построить шкалу от 0 до 850. Так как 850 не делится удобно на 20, то округляем число 850 до ближайшего удобного числа, в данном случае 1000 (1000 : 20 = 50), т. е. в одной клетке 50, а в двух клетках 100; следовательно, масштаб - 100 в двух клетках.

Из неравномерных 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 10 Igl = 0; IglO = 1; Igl00 = 2 и т. д. (рис. 5.4).



Числа Логарифмы чисел

Рис. 5.4. Шкалы

ледний элемент графика - **экспликация.** Каждый график н иметь словесное описание его содержания. Оно включа- Дол>к gj, название графика, которое в краткой форме передает еТ В ^держание; подписи вдоль масштабных шкал и пояснения ^отдельным частям графика.

5.2

КЛАССИФИКАЦИЯ ВИДОВ ГРАФИКОВ

Существует множество видов графических изображений (рис. 5.5 и *5.6).* Их классификация основана на ряде признаков: а) способ построения графического образа; б) геометрические знаки, изображающие статистические показатели; в) задачи, решаемые с помощью графического изображения.

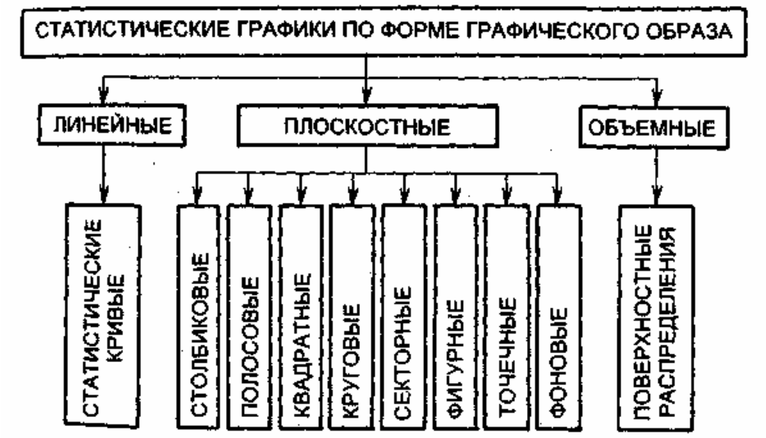


Рис. 5.5. Классификация статистических графиков по форме графического образа

По способу построения статистические графики делятся на **Диаграммы и статистические карты.**

Диаграммы - наиболее распространенный способ графических изображений. Это графики количественных отношений. Виды и способы их построения разнообразны. Диаграммы применяются для наглядного сопоставления в различных аспек- (пространственном, временном и др.) независимых друг от друга величин: территорий, населения и т. д. При этом сравнение исследуемых совокупностей производится по какому-либо существенному варьирующему признаку. Статистические карты - графики количественного распределения по поверхности. По своей основной цели они близко примыкают к диаграммам и специфичны лишь в том отношении, что представляют собой условные изображения статистических данных на контурной географической карте, т. е. показывают пространственное размещение или пространственную распространенность статистических данных. **Геометрические знаки,** как было сказано выше, - это либо точки, либо линии или плоскости, либо геометрические фигуры. В соответствии с этим различают графики точечные, линейные, плоскостные и пространственные (объемные)..

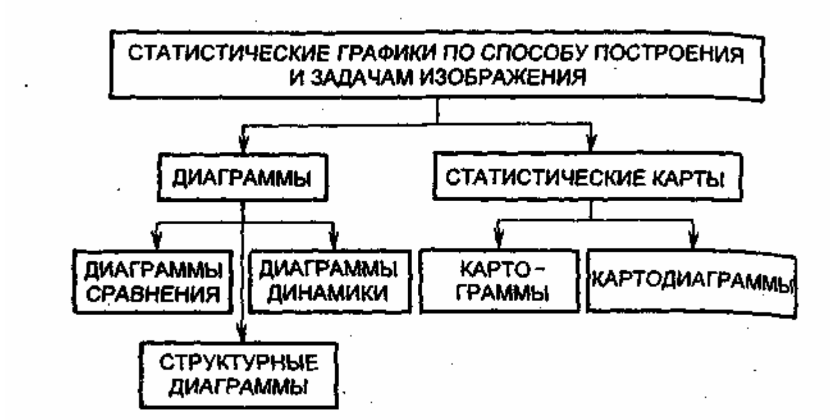


Рис. 5.6. Классификация статистических графиков по способу построения и задачам изображения

При построении точечных диаграмм в качестве графических образов применяются совокупности точек; при построении линейных - линии. Основной принцип построения всех плоскостных диаграмм сводится к тому, что статистические величины изображаются в виде геометрических фигур и, в свою очередь, подразделяются на столбиковые, полосовые, круговые, квадратные и фигурные.

Статистические карты по графическому образу делятся на картограммы и картодиаграммы.

В зависимости от круга решаемых задач выделяют диаграммы сравнения, структурные диаграммы и диаграммы динамики.

к tivf видом графиков являются диаграммы распределения O^{c}° пРеДставленных вариационным рядом.-Это гистограм-

веЛИ Ч п ОЛИ гоН, ОГИВа, к у м у лята,

5.3

ДИАГРАММЫ СРАВНЕНИЯ

Наиболее распространенными диаграммами сравнения являются столбиковые **диаграммы,** принцип построения которых остоит в изображении статистических показателей в виде поставленных по вертикали прямоугольников - столбиков. Каждый столбик изображает величину отдельного уровня исследуемого статистического ряда. Таким образом, сравнение статистических показателей возможно потому, что все сравниваемые показатели выражены в одной единице измерения.

При построении столбиковых диаграмм необходимо начертить систему прямоугольных координат, в которой располагаются столбики. На горизонтальной оси располагаются основания столбиков, величина основания определяется произвольно, но уста•навливается одинаковой для всех.

Шкала, определяющая масштаб столбиков по высоте, расположена по вертикальной оси. Ве'личина каждого столбика по вертикали соответствует размеру изображаемого на графике статистического показателя. Таким образом, у всех столбиков, составляющих диаграмму, переменной величиной является только одно измерение. Покажем построение столбиковой диаграммы по данным табл. 5.1, характеризующим вклады граждан в учреждения Сбербанка в 1998 г. (рис. 5.7).

Таблица 5.1

Вклады граждан в учреждения Сбербанка в 1998 г. (цифры условные)

Месяц 1 2 3 4 5 6 7 8 9 1 0 1 1 1

j Вклады, jJ^"Руб. 550 560 560 640 640

11001100110016301610161025

Эта группа графиков рассматривается в гл. 3.

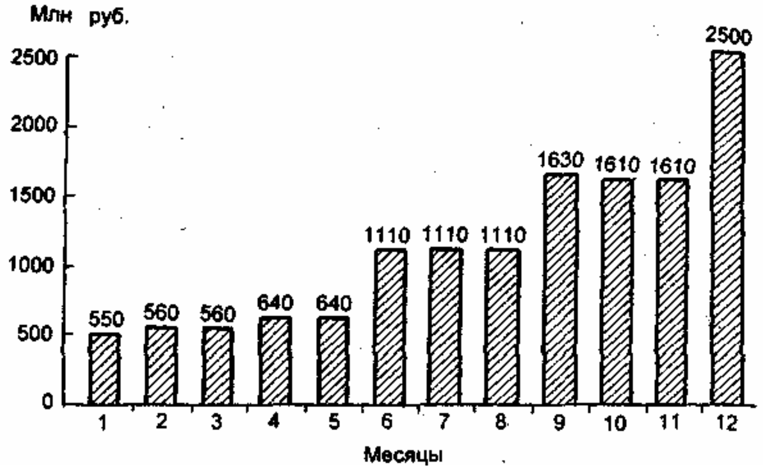


Рис. 5.7. Вклады граждан в учреждения Сбербанка в 1998 г.

В соответствии с изложенными выше правилами на горизонтальной оси размещаются основания двенадцати столбиков на одинаковом расстоянии друг от друга, в данном случае 0,5 см. Ширина столбиков принята 0,5 см. Масштаб на оси ординат 500 млн руб. - 1 см. Наглядность данной диаграммы достигается сравнением величины столбиков.

Размещение столбиков в поле графика может быть различным:

- на одинаковом расстоянии друг от друга (рис. 5.7);
- вплотную друг к другу (рис. 5.8);
- в частном наложении друг на друга (рис. 5.9).

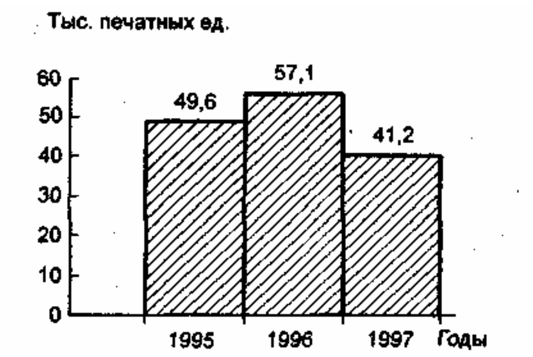


Рис. 5.8. Динамика выпуска книг и брошюр в одном из регионов России за 1995 - 1997 гг.

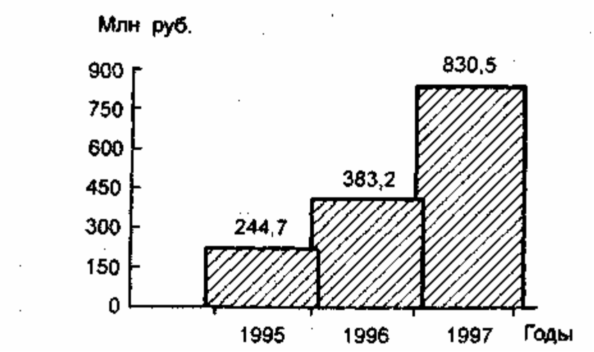


Рис. 5.9. Динамика денежных доходов населения в регионе за 1995 - 1997 гг.

Правила построения столбиковых диаграмм допускают одновременное расположение на одной горизонтальной оси изображений нескольких показателей. В этом случае столбики располагаются группами, для каждой из которых может быть принята разная размерность варьирующих признаков (рис. 5.10).

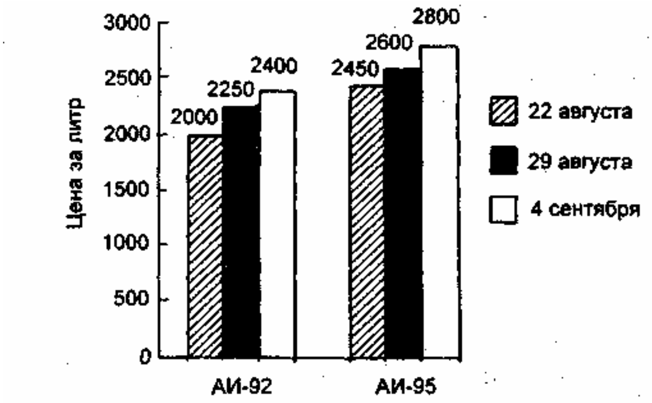


Рис. 5.10. Средние розничные цены на бензин в Москве в августе и сентябре 1997 г.

Разновидности 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.

Область применения столбиковых и полосовых диагп одинакова, так как идентичны правила их построения. Одном*^{1}*^{1}ность изображаемых статистических показателей и -их одном^ штабность для различных столбиков и полос требуют выпол^3' ния единственного положения: соблюдения соразмерности ($_{ст}$ ^{е}биков - по высоте, полос - по длине) и пропорционально ^{Л} изображаемым величинам. Для выполнения этого требован" необходимо: во-первых, чтобы шкала, по которой устанавлив" ется размер столбика (полосы), начиналась с нуля; во-вторых эта шкала должна быть непрерывной, т. е. охватывать все числ данного статистического ряда; разрыв шкалы и соответственно столбиков (полос) не допускается. Невыполнение указанных правил приводит к искаженному графическому представлению анализируемого статистического материала.

В качестве примера приведем полосовую диаграмму сравнения по данным табл. 5.2 (рис. 5.11).

Таблица 5.2 **Общий объем промышленного производства в некоторых** странах **СНГ в I квартале 1995 г. (в % к I кварталу 1994** *г.)* **(цифры условные)**

Страны СНГ	Общий объем промышленного производства
Казахстан	88,7
Беларусь	83,5
Россия	80,7
Кыргызстан	77,6
Таджикистан	71,8
Армения	41,6

Разновидностью столбиковых (ленточных) диаграмм являются **направленные диаграммы.** Они отличаются от обычных двусторонним расположением столбиков или полос и имеют начало отсчета по масштабу в середине. Обычно такие диаграммы применяются для изображения величин противоположного качественного значения. Сравнение между собой столбиков (полос), направленных в разные стороны, менее эффективно, чем расположенных рядом в одном направлении. Несмотря на это, анализ направленных диаграмм позволяет делать достаточно содержательные выводы, так как особое расположение придав графику яркое изображение. К группе двусторонних относятся диаграммы числовых отклонений. В них полосы направлены в обе стороны от вертикальной нулевой линии: вправо - для прироста; влево - для уменьшения. С помощью таких диаграмм удобно изображать отклонения от плана или некоторого уровня, принятого за базу сравнения. Важным достоинством рассматриваемых диаграмм является возможность видеть размах колебаний изучаемого статистического признака, что само по себе имеет большое значение для экономического анализа (рис. 5.12).

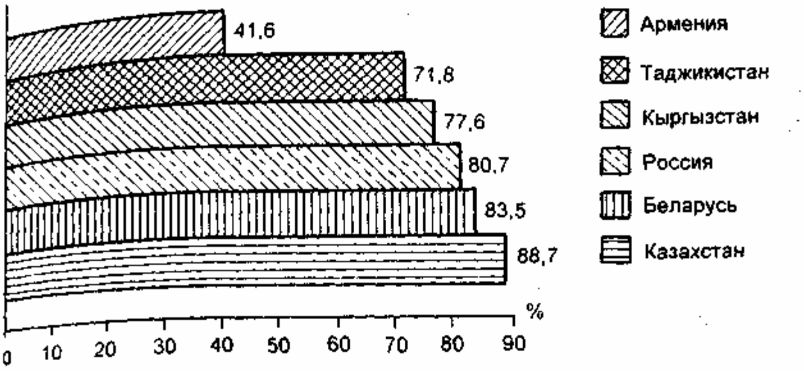


Рис. 5.11. Общий объем промышленного производства в странах СНГ в I квартале 1995 г. (в % к I кварталу 1994 г.)

Для простого сравнения независимых друг от друга показателей могут также использоваться диаграммы, принцип построения которых состоит в том, что сравниваемые величины изображаются в виде правильных геометрических фигур, которые строятся так, чтобы площади их относились между собой как количества, этими фигурами изображаемые. Иными словами, **эти Диаграммы выражают величину изображаемого явления размером своей площади.**

Для получения диаграмм рассматриваемого типа используют
Разнообразные геометрические фигуры - квадрат, круг, реже -
прямоугольник. Известно, что площадь квадрата равна квадра-
v £го стороны, а площадь круга определяется пропорциональ-
о квадрату его радиуса. Поэтому для построения диаграмм не-
ходимо сначала из сравниваемых величин извлечь квадрат-
'и корень. Затем на базе полученных результатов определить
°рону квадрата или радиус круга соответственно принятому

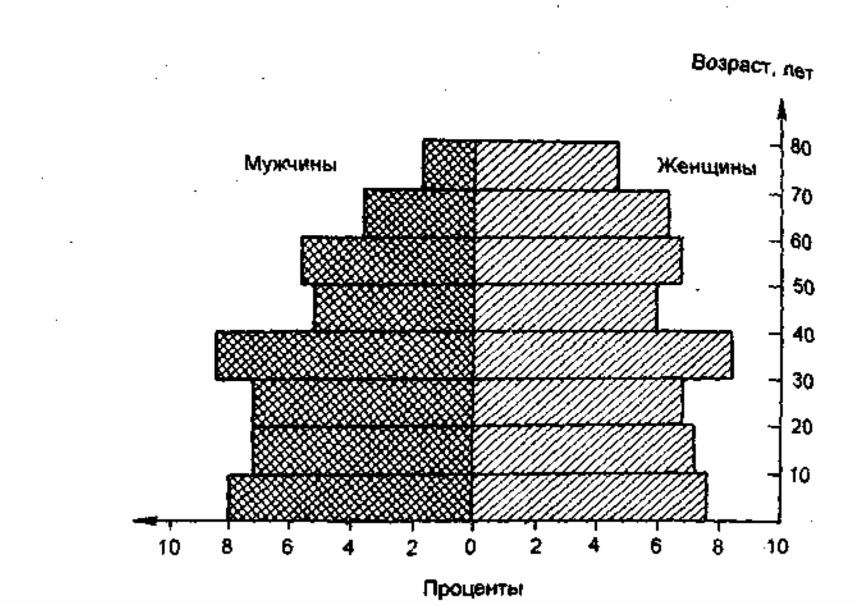


Рис. 5.12. Распределение населения одного из регионов России по полу и возрасту в 1998 г.

Например, если изобразить в виде квадрата поставки российского газа в ближайшее зарубежье, то сначала нужно извлечь квадратные корни из этих цифр (табл. *53).*

Таблица 5.3

Поставки российского газа в страны ближнего зарубежья, январь - август 1995 г.

Странд>1 ближнего зарубежья	Млн $м^3$
Украина ,	44460,1
Беларусь	10250,0
Литва •	2458,0

Это составит: для Украины - 210,9; Беларуси -101,2; Литвы 49,6. Затем установить масштаб и по этим данным построить квадраты. Для нашего примера примем 1 см равным 30 млн м •
сторона первого квадрата составит 7,03 см (210,9 : 30); вто-
Т^{о2}^ з 4 см; третьего - 3,65 см (рис. 5.13).
рого " '

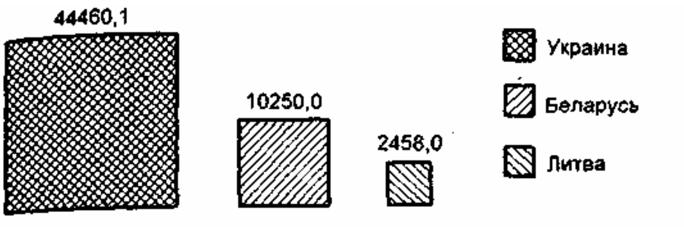


РИС 5.13. Поставки российского газа в страны ближнего зарубежья, январь-август 1995 г.

Для правильного построения диаграмм квадраты или круги необходимо расположить на одинаковом друг от друга расстоянии, 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 с различной заштриховкой или закраской. Такие диаграммы также позволяют сравнивать между собой ряд исследуемых величин. На рис. 5.14 показан такой вариант круговой диаграммы (цифры условные).

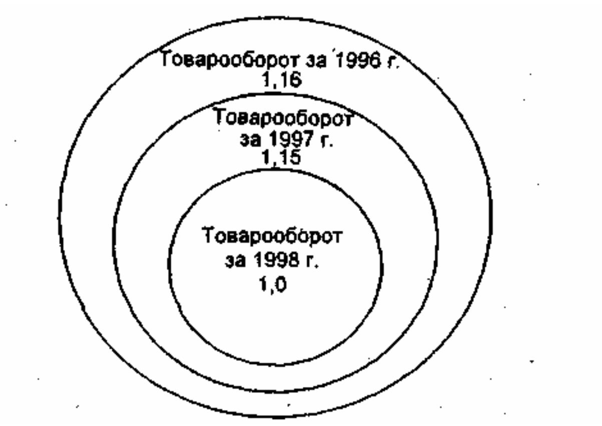


Рис. 5.14. Рост товарооборота вещевых, смешанных
и продовольственных рынков за 1996 - 1998 гг.
(товарооборот 1998 г. принят за единицу)

Наиболее выразительным и легко воспринимаемым явл i способ построения **диаграмм сравнения в виде фигур-**$_{\text{зна}}{}^{\text{е}\wedge\text{С!|}}$

В этом случае статистические совокупности изображаютс $^{\text{оВ'}}$ геометрическими фигурами, а символами или знаками, восп ^ изводящими в какой-то степени внешний образ статистическ данных. Достоинство такого способа графического изображ$_{\text{ен}}{}^{\text{Их}}$ заключается в высокой степени наглядности, в получении под ^ ного отображения, отражающего содержание сравниваемых вокупностей.

Важнейший признак любой диаграммы - масштаб, Поэто му, чтобы правильно построить фигурную диаграмму, необходимо определить **единицу счета.** В качестве последней принимается отдельная фигура (символ), которой условно присваивается конкретное численное значение. А исследуемая статистическая величина изображается отдельным количеством одинаковых по размеру фигур, последовательно располагающихся на рисунке. Однако в большинстве случаев не удается изобразить статистический показатель целым количеством фигур. Последнюю из них приходится делить на части, так как по масштабу один знак является слишком крупной единицей измерения. Обычно эта часть определяется на глаз. Сложность точного ее определения является недостатком фигурных диаграмм. Однако, если большая точность представления статистических данных не преследуется, то результаты получаются вполне удовлетворительными.

Рассмотрим построение фигурной диаграммы по следующим данным табл. 5.4

Таблица 5.4

Численность фермерских хозяйств в одном из регионов России за 1996 — 1998 гг. (данные условные)

Год	1996	1997	1998
Численность фермерских хозяйств	49	183	270

Примем условно за один знак 40 тыс. фермерских хозяйствТогда число хозяйств в России в 1996 г. в размере 49 тыс. буД$^{\text{еТ}}$ изображено в количестве 1,22 хозяйства, в 1997 г. •- 4,6 хозяйства и т. д. (рис. 5.15).

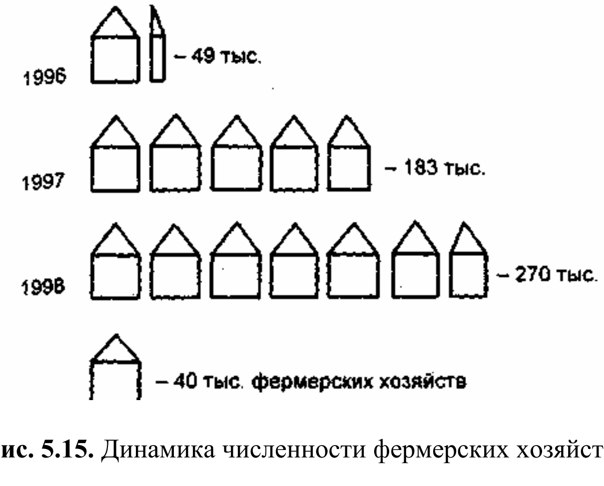


Рис. 5.15. Динамика численности фермерских хозяйств в одном из регионов России за 1996 - 1998 гг.

Как правило, фигурные диаграммы широко используются для популяризации статистических данных и рекламы.

5.4

СТРУКТУРНЫЕ ДИАГРАММЫ

Основное назначение структурных диаграмм заключается в графическом представлении состава статистических совокупностей, характеризующихся как соотношение различных частей каждой из совокупностей. Состав статистической совокупности графически может быть представлен с помощью как абсолютных, так и относительных показателей. В первом случае не только размеры отдельных частей, но и размер графика в целом' определяются статистическими величинами и изменяются в соответствии с изменениями последних. Во втором - размер всего графика не меняется (так как сумма всех частей любой совокупности составляет 100%), а меняются только размеры отдельных

о частей. Графическое изображение состава совокупности по

солютным и относительным показателям способствует проведению более глубокого анализа и позволяет проводить междуна-

^НЬ1е сопоставления и сравнения социально-экономических явлений.

В качестве графического образа для изображения стру^ совокупностей применяются прямоугольники - для постро столбиковых и полосовых диаграмм и круги - для построе*^8 Ия секторных диаграмм.

Покажем построение указанных выше диаграмм на конкп ных примерах.

Чтобы по приведенным данным табл. 5.5 построить диаграму отражающую структуру сравниваемых совокупностей по соотн^ шению в них отдельных видов часов, ряд абсолютных показателе" заменяется рядом относительных величин. В этом случае каждая и полос диаграммы будет иметь одинаковую длину, так как при переходе к относительным величинам погашаются различия в абсолютных размерах совокупностей. В то же время структурные различия проявляются значительно четче. Графическое изображение структуры с помощью столбиковых (полосовых) диаграмм позволяет изучить особенности многих изучаемых экономических явлений Так, приведенная на рис. 5.16 диаграмма характеризует увеличение доли наручных часов в общем производстве.

Таблица 5.5

Производство часов по видам в одном из регионов России за 1990 - 1998 гг. (цифры условные)

	1990г.		1998г.	
	млн шт.	%	млн шт.	%
Часы - всего	52,5	100,0	60,1	100,0
В том числе:				
наручные	24,4	46,5	31,6	52,6
настенные	9,3	17,7	10,5	17,5
будильники	18,8	35,8	18,0	29,9

Более распространенным способом графического изображения структуры статистических совокупностей является секторная диаграмма, которая считается основной формой диаграммы такого назначения. Это объясняется тем, что идея целого очень хорошо и наглядно выражается кругом, который представляет всю совокупность. Удельный вес каждой части COBOJ купности в секторной диаграмме характеризуется величине центрального угла (угол между радиусами круга). Сумма все углов круга, равная 360°, приравнивается к 100%, а следов^ тельно, 1% принимается равным 3,6°.

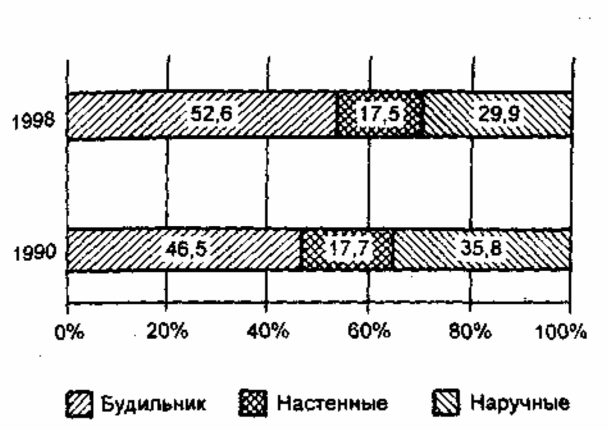


рис. 5.16. Динамика удельного веса производства часов по видам за 1990 - 1998 гг.

Приведем пример построения секторной диаграммы по даным табл. 5.6.

Таблица 5.6 Динамика доли негосударственного сектора экономики
в розничной торговле (в % к общему объему розничного товарооборота в России)

	1992г.	1993г.
Государственный сектор	78	49
Негосударственный сектор	2251В том числе предприятия: частной и смешанной форм	
собственности	1,8	31
потребительской кооперации	20	16
прочих форм собственности	0,2 t	4

Построение секторной диаграммы начинается с определения Центральных углов секторов. Для этого процентное выражение ^Дельных частей совокупностей умножают на 3,6°. Например, $^{\text{Для}}$Данных:
.J992 г.: 78-3,6° = 280,8°; 1,8 • 3,6° = 6,5°; 20 -3,6° = 72°; 0,2- 3,6° = о,7°.;
1993 г.: 49-3,6°= 176,4°; 31 • 3,6° = 111,6°; 16 - 3,6° = 57,6°; ^{3}'6° = 14,4°.

По найденным значениям углов круги делятся на соответ ющие секторы (рис. 5.17). $^{\text{СТб}}$)

Рис. 5.17, Динамика доли негосударственного сектора экономики в розничной торговле (в % к общему объему розничного товарооборота в России)

Применение секторных диаграмм позволяет не только графически изобразить структуру совокупности и ее изменение, но и показать динамику численности этой совокупности. Для этого строятся круги, пропорциональные объему изучаемого признака, а затем секторами выделяются его отдельные части.

Рассмотренные способы графического изображения структ/' ры совокупности имеют как достоинства, так и недостатки.

Так, секторная диаграмма сохраняет наглядность и выразительность лишь при небольшом числе частей совокупности, в против ном случае ее применение малоэффективно. Кроме того, нагляД ность секторной диаграммы снижается при незначительных изм нениях структуры изображаемых совокупностей: она выше, ее' имеются существенные различия сравниваемых структур. Преи^У ществом столбиковых (ленточных) структурных диаграмм по ср нению с секторными являются их большая емкость, возможно отразить более широкий объем полезной информации.

5.5

ДИАГРАММЫ ДИНАМИКИ

изображения и внесения суждений о развитии явления во

ени строятся диаграммы динамики. В"л наглядного изображения явлений в рядах динамики исполь- -я диаграммы: столбиковые, ленточные, квадратные, круго3,Ю линейные, радиальные и др. Выбор вида диаграммы зависит ВЬ" вном от особенностей исходных данных, цели исследоваВ я Например, если имеется ряд динамики с несколькими нерав"оотстоящими уровнями во времени (1913, 1940,1950, 1980, 1985, 1997 гг.), то часто для наглядности используют столбиковые, квадратные или круговые диаграммы. Они зрительно впечатляют, хорошо запоминаются, но не годны для изображения большого числа уровней, так как громоздки. Когда число уровней в ряду динамики велико, целесообразно применять линейные диаграммы, которые воспроизводят непрерывность процесса развития в виде непрерывной ломаной линии. Кроме того, линейные диаграммы удобно использовать, если целью исследования является изображение общей тенденции и характера развития явления; когда на одном графике необходимо изобразить несколько динамических рядов с целью их сравнения; если наиболее существенным является сопоставление темпов роста, а не уровней.

Для построения линейных графиков применяют систему прямоугольных координат. Обычно по оси абсцисс откладывается время (годы, месяцы и т. д.), а по оси ординат - размеры изображаемых явлений или процессов. На оси ординат наносят масштабы. Особое внимание следует обратить на их выбор, так как от этого зависит общий вид графика. Обеспечение равновесия, пропорциональности между осями координат необходимо в графике 8 связи с тем, что нарушение равновесия между осями координат Дает неправильное изображение развития явления. Если масштаб Для шкалы на оси абсцисс очень растянут по сравнению с масшаоом на оси ординат, то колебания в динамике явлений мало 'Делаются, и наоборот, преувеличение масштаба по оси ординат ^сравнению с масштабом на оси абсцисс дает резкие колебания. —ным периодам времени и размерам уровня должны соответ- овать равные отрезки масштабной шкалы.

^^ статистической практике чаще всего применяются графиОн К"е изображения с равномерными шкалами. По оси абсцисс

еРутся пропорционально числу периодов времени, а по оси ординат - пропорционально самим уровням. Масштабом мерной шкалы будет длина отрезка, принятого за един^цР3"1^

Рассмотрим построение линейной диаграммы на осно следующих данных (табл. 5.7). Bai4i,

Таблица 5-

Динамика валового сбора зерновых культур в регионе за 1985-1994 гг.

Год	1985	1986	1987	1988	1989	1990	1991	1992	1993 ^Г
Млн т	237,4	179,2	189,1	158,2	186,8	192,2	172,6	191,7210,!	2Ц3!

Изображение динамики валового сбора зерновых культур на координатной сетке с неразрывной шкалой значений, начинающихся от нуля, вряд ли целесообразно, так как 2/3 поля диаграммы остаются неиспользованными и ничего не дают для выразительности изображения. Поэтому в данных условиях рекомендуется строить шкалу без вертикального нуля, т. е. шкала значений разрывается недалеко от нулевой линии и на диаграмму попадает лишь часть всего возможного поля графика. Это не приводит к искажениям в изображении динамики явления, и процесс его изменения рисуется диаграммой более четко (рис. 5.18).

на одном линейном графике приводится несколько НеР$^{6\wedge}$ $_{отО}$рые дают сравнительную характеристику динамики . рцвы^{х}> $_{пок}$азателей или одного и того же показателя. азли^{чн} ^ графического изображения сразу нескольких покаДТявляется рис. 5Л9.

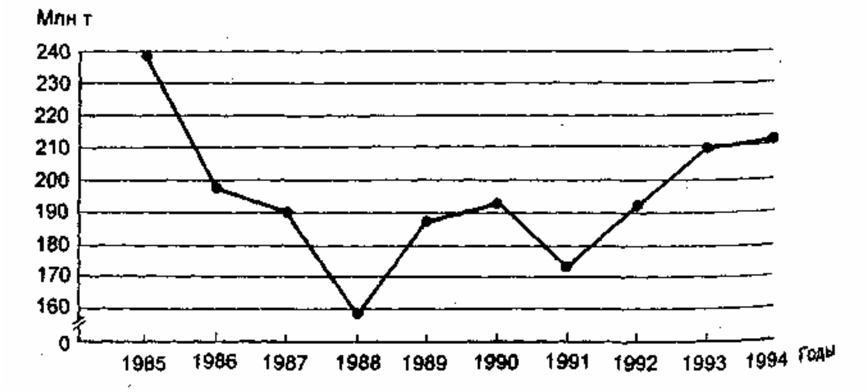


Рис. 5.18. Динамика валового сбора зерновых культур в регионе за 1985 - 1994 гг.

Рис. 5.19. Динамика производства чугуна и готового проката в регионе за 1985 - 1994 гг.

Однако на одном графике не следует помещать более трехчетырех кривых, так как большое их количество неизбежно осложняет чертеж и линейная диаграмма теряет наглядность.

В некоторых случаях нанесения на один график двух кривых дает возможность одновременно изобразить динамику третьего показателя, если он является разностью первых двух. Например, при изображении динамики рождаемости и смертности площадь между двумя кривыми показывает величину естественного прироста или естественной убыли населения.

Иногда необходимо сравнить на графике динамику двух показателей, имеющих различные единицы измерения. В таких случаях понадобится не одна, а две масштабные шкалы. Одну из ^{НИх} Р^{а}змещдют справа, другую - слева.

чцнако такое сравнение кривых не дает достаточно полной ^ртины динамики этих показателей, так как масштабы произволь^{Ь1}- Поэтому сравнение динамики уровня двух разнородных покачен следует осуществлять на основе использования одного сщ$_{та}$5$_{а\ после}$ преобразования абсолютных величин в относиьные. Примером такой линейной диаграммы является рис. 5.20.

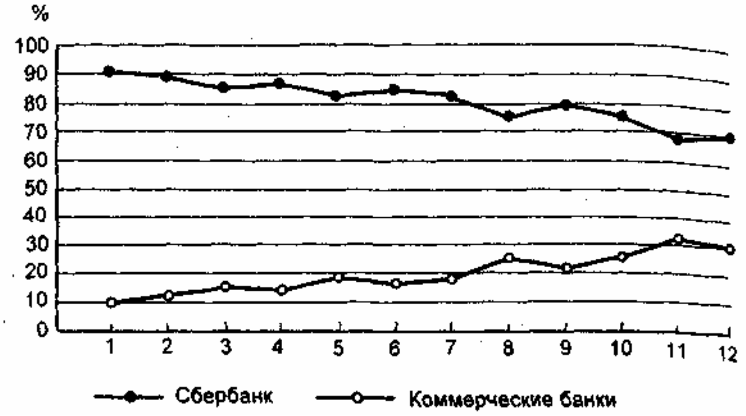


Рис. 5.20. Доли вкладов граждан в Сбербанк и коммерческие банки в одном из городов в 1997 г. (%)

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 протяжении исследуемого периода. Однако при изучении динамики важно знать относительные изменения исследуемых показателей по сравнению с достигнутым уровнем или темпы их изменения. Именно относительные изменения экономических показателей в динамике искажаются при их изображении на координатной диаграмме с равномерной вертикальной шкалой. Кроме того, в обычных координатах теряет всякую наглядность и даже становится невозможным изображение для рядов динамики с резко изменяющимися уровнями, которые обычно имеют место в динамических рядах за длительный период времени.

В этих случаях следует отказаться от равномерной шкалы и положить в основу графика полулогарифмическую систему. Основная идея полулогарифмической системы состоит в том, что в ней равным линейным отрезкам соответствуют равные значения лога*рифмов чисел. Такой подход имеет преимущество: возможность уменьшения размеров больших чисел через их логарифмические эквиваленты. Однако с масштабной шкалой в виде логарифме график малодоступен для понимания. Необходимо рядом с лог рифмами, обозначенными на масштабной шкале, проставить сам числа, характеризующие уровни изображаемого ряда динамик ., которые соответствуют указанным числам логарифмов. Такого рДО графики носят название графиков на полулогарифмической сет

тт лулогаРиФмическо,, сеткой называется сетка, в которой
иной оси нанесен линейный масштаб, а на другой - лога-
нзлмический. В данном случае логарифмический масштаб на-
ится на ось ординат, а на оси абсцисс располагают равномер-
Ноо шкалу для отсчета времени по принятым интервалам (го-
нум кварталам, месяцам, дням и пр.).

^Техника построения логарифмической шкалы следующая (рис- 5.21).

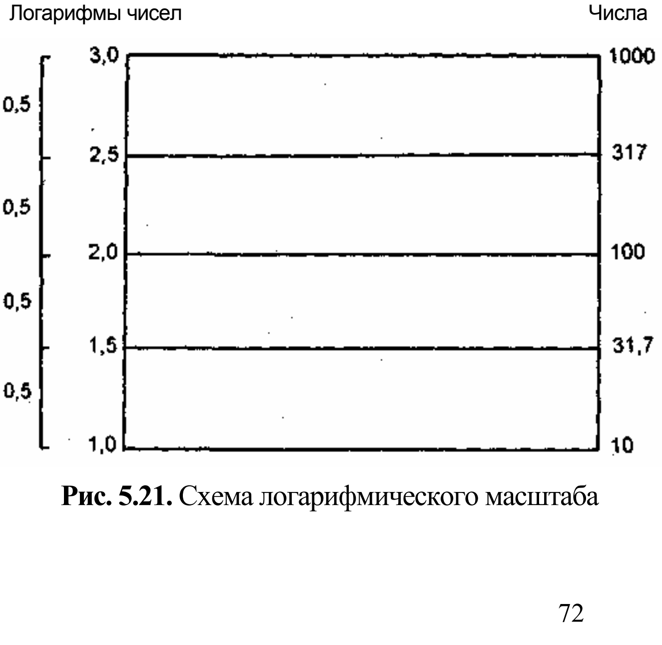


Рис. 5.21. Схема логарифмического масштаба

Необходимо найти логарифмы исходных чисел, начертить ординату и разделить ее на несколько равных частей. Затем нанести на ординату (или равную ей параллельную линию) отрезки, пропорциональные абсолютным приростам этих логарифмов. Далее записать соответствующие логарифмы чисел и их антилогарифмы, например (0,000; 0,3010; 0,4771; 0,6021; ...; 1,000, что Дает 1, 2, 3, 4, ..., 10). Полученные антилогарифмы окончательно Дают вид искомой шкалы на ординате.

Приведем *пример* логарифмического масштаба.

Допустим, что надо изобразить на графике динамику производства электроэнергии в регионе за 1965 - 1994 гг., за эти годы °но выросло в 9,1 раза. С этой целью находим логарифмы для каждого уровня ряда (табл. 5.8).

Определив минимальное и максимальное значение логариф-

^{в} производства электроэнергии, построим масштаб с таким рас^{е}том, чтобы все данные разместились на графике.

' , $^{\text{Табл}}$**ИЦа,,**

Динамика производства электроэнергии в регионе
за 1965 - 1994 гг. (млрд кВт- ч)

Год	Y,	LgY,	Год	Yi	ij^
1965	170	2,23	1985	1039	3 02 !
1970	292	2,46	1990	1294	з'ц 1
1975	507	2,70	1994	1544	з|19
1980	741	2,84			

Учитывая масштаб, находим соответствующие точки, которые соединим прямыми линиями, в результате получим график (рис. 5.22) с использованием логарифмического масштаба на оси ординат. Он называется диаграммой на полулогарифмической сетке. Полной логарифмической диаграммой он станет в том случае, если по оси абсцисс будет построен логарифмический масштаб. В рядах динамики это никогда не применяется, так как логарифмирование времени лишено всякого смысла.

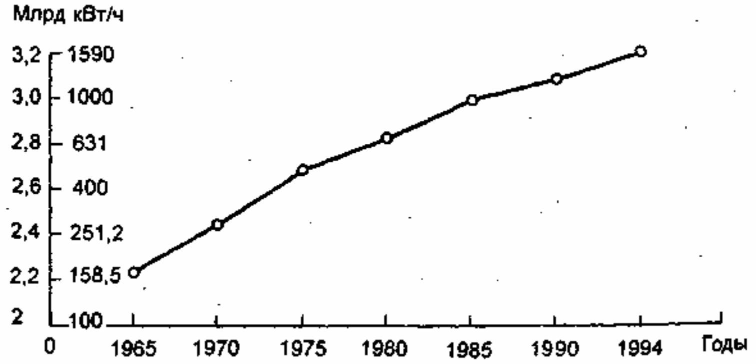


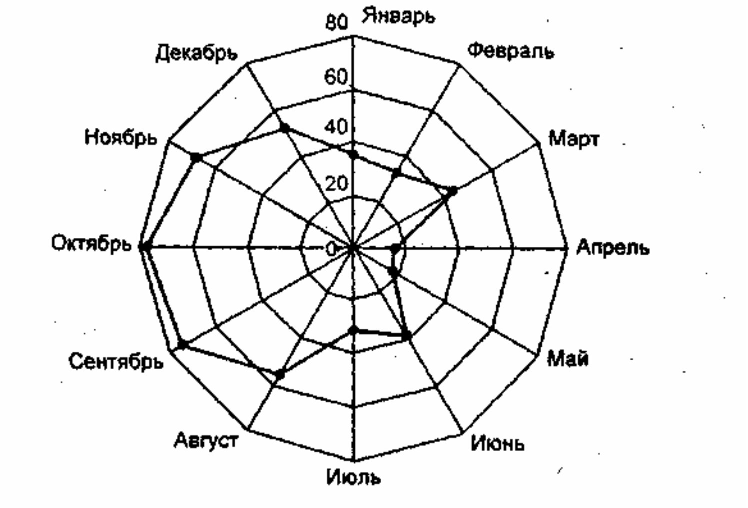
Рис. 5.22. Динамика производства электроэнергии в регионе за *1965* - 1994 гг.

Применяя логарифмический масштаб, можно без всяких вычислений характеризовать динамику уровня. Если кривая на логарифмическом масштабе несколько отклонена от прямой и становится вогнутой к оси абсцисс, значит, имеет место падение темпов; когда кривая в своем течении приближается к прямой - стабильность

если она отклоняется от прямой в сторону, выпуклую к оси ^virioB, у$_{\text{ЧаеМ}}$0е явление имеет тенденцию к росту с увеличива$^{\text{абСЦ}}$м$^{\text{С}}$ися темпами.

^гт амику изображают и **радиальные диаграммы,** строящиеся пных координатах. Радиальные диаграммы преследуют цель в $^{\text{ЛоЛ}}$$_{\text{я}}$ного изображения определенного ритмического движения во $^{\text{нагл}}$ $_{\text{и}}$ Чаще всего эти диаграммы применяются для иллюстра$^{\text{в}}$Р$^{\text{еМ}}$$_{\text{сезОНН}}$ых колебаний. Радиальные диаграммы разделяются на "$^{\text{И}}$ **кнутые и спиральные.** По технике построения радиальные ^{33}аграммы отличаются друг от друга в зависимости от того, что $^{\text{Д}}$ в $_{\text{ка}}$честве пункта отсчета - центр круга или окружность. $^{\text{В}}$ **Замкнутые диаграммы** отражают внутригодичный цикл дина(ики какого-либо одного года. **Спиральные диаграммы** показывав внутригодичный цикл динамики за ряд лет.

Построение замкнутых диаграмм сводится к следующему: вычерчивается круг, среднемесячный показатель приравнивается к радиусу этого круга. Затем весь круг делится на 12 радиусов, которые на графике приводятся в виде тонких линий. Каждый радиус обозначает месяц, причем расположение месяцев аналогично циферблату часов: январь - в том месте, где на часах 1, февраль - 2, и т. д. На каждом радиусе делается отметка в определенном месте

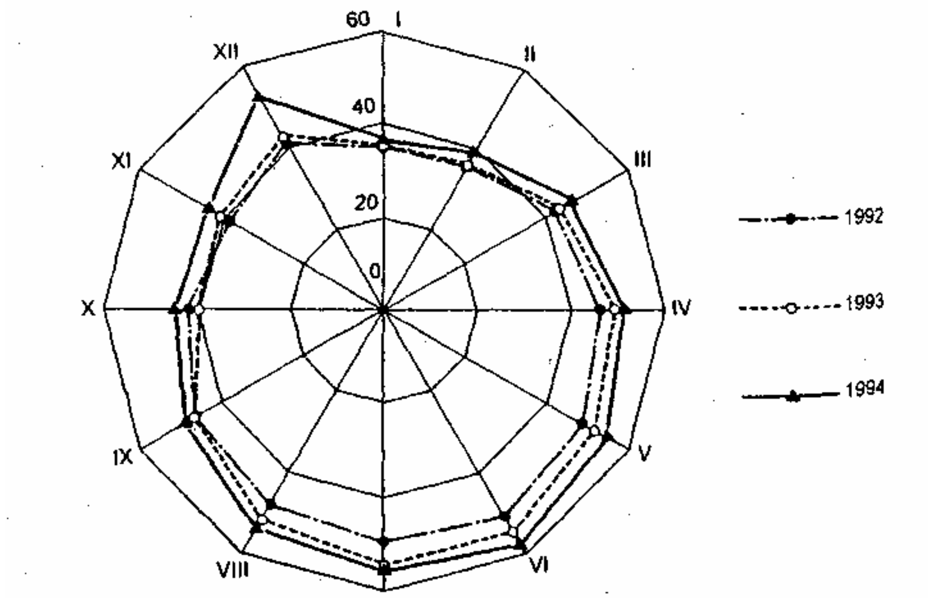


——« ---- Производство мяса, млн т

Рис. 5.23. Сезонные колебания производства мяса в одном из регионов России в 1997 г.

согласно масштабу исходя из данных за соответствующий м$_{ес}$ Если данные превышают среднемесячный уровень, отметка дзд$_{а}$$^{Я|*'}$ ся за пределами окружности на продолжении радиуса. Затем отм^ ки различных месяцев соединяются отрезками. В приведены примере (рис. 5.23) R = 44,8 тыс. т, длина радиуса- 3,0 см. Слег/^{1} вательно, 1 см = 44,8 : 3,0 ~ 15 тыс. т. Данная замкнутая диаграмм наглядно показывает, что производство мяса подвергнуто сезонным колебаниям. Минимум производства мяса приходится на апрель, май затем наблюдается медленное его повышение к августу, резкий подъем в сентябре, октябре и опять спад в декабре, январе. Если же в качестве базы для отсчета взять не центр круга, а окружность, то диаграммы называются спиральными.

Построение спиральных диаграмм отличается от замкнутых тем, что в них декабрь одного года соединяется не с январем данного же года, а с январем следующего года. Это дает возможность изобразить весь ряд динамики в виде спирали. Особенно наглядна такая диаграмма, когда наряду с сезонными изменениями происходит неуклонный рост из года в год (рис. 5.24).



Гоеди различных видов графиков особое место занимает кри-

Рис. 5.24. Продажа пива в розничной торговле в городе за 1992 - 1994 гг.

именуемая **моделью Лоренца, или кривой Лоренца. Дан-**
взя кривая дает возможность графически изобразить уровень
нз пентрации явления. Пример построения кривой Лоренца

описан в главе 11.

5.6

СТАТИСТИЧЕСКИЕ КАРТЫ

Статистические карты представляют собой вид графических изображений статистических данных на схематической географической карте, 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.

Картограмма - это схематическая географическая карта, на которой штриховкой различной густоты, точками или окраской определенной степени насыщенности показывается сравнительная интенсивность какого-либо показателя в пределах каждой единицы нанесенного на карту территориального деления (например, плотность населения по областям или республикам, распределение районов по урожайности зерновых культур и т. п.). Картограммы делятся на фоновые и точечные.

Картограмма фоновая - вид картограммы, на которой штриховкой различной густоты или окраской определенной степени насыщенности показывают интенсивность какого-либо показателя в пределах территориальной единицы.

Картограмма точечная - вид картограммы, где уровень выбранного явления изображается с помощью точек. Точка изображает одну единицу совокупности или некоторое их количество, показывая на географической карте плотность или частоту проявления определенного признака.

Фоновые картограммы, как правило, используются для изображения средних или относительных показателей, точечные -для °°ъемных (количественных) показателей (численность населения, "оголовье скота и т. д.).

Рассмотрим построение картограммы, используя данные табл. 5.9.

Таблица 5.9 Плотность населения восьми районов области (цифры условные)

№ района	1	2	3	4	5	6	7	8
Плотность населения на 1000 км^2, тыс. человек	3,0	4,0	11,0	14,0	17,0	13,0	11,0	з,о

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

Согласно данным табл. 5.9 все районы по плотности населения можно разбить на три группы: 1) районы, имеющие плотность населения до 4 тыс. человек; 2) от 4 до 12 тыс. человек; 3) от 12 до 17 тыс. человек. 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 (рис. 5.25). Другим примером фоновой картограммы является рис. 5.26.

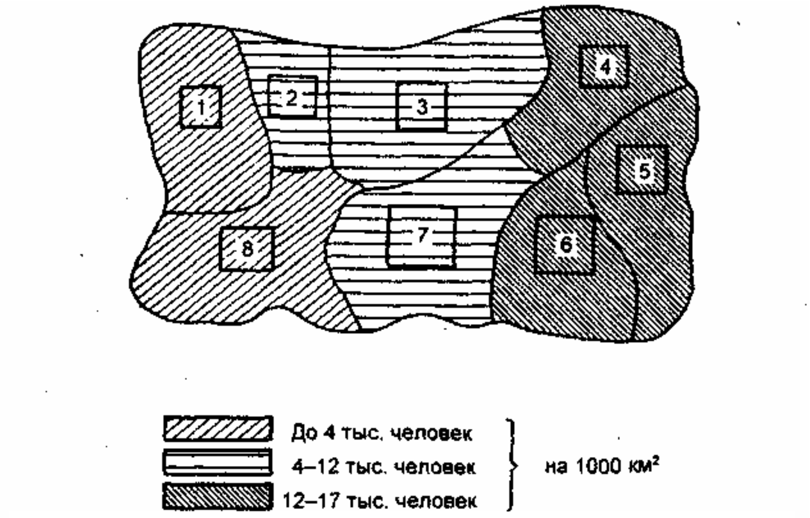


Рис. 5.25. Картограмма плотности населения восьми районов облает

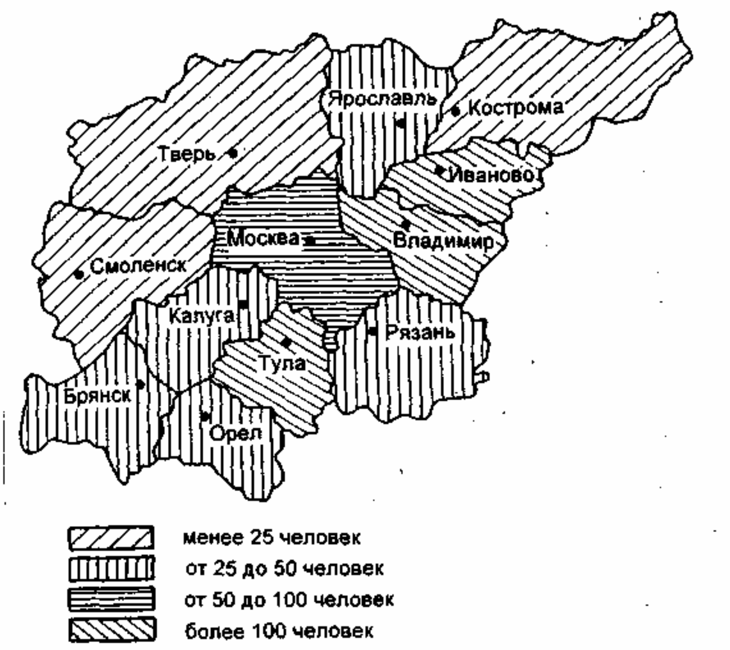


Рис. 5.26. Плотность населения в областях Центрального района России (человек на 1 м^2)

Вторую большую группу статистических карт составляют **картодиаграммы,** представляющие собой сочетание диаграмм с географической картой. В качестве изобразительных знаков в картодиаграммах используются диаграммные фигуры (столбики, квадраты, круги, фигуры, полосы), которые размещаются на контуре географической карты. Картодиаграммы дают возможность географически отразить более сложные статистико-географические построения, чем картограммы.

Среди картодиаграмм следует выделить картодиаграммы простого сравнения, графики пространственных перемещений, изолиний.

На картодиаграмме простого сравнения в отличие от обычной Диаграммы диаграммные фигуры, изображающие величины ис^едуемого показателя, расположены не в ряд, как на обычной ^^{и}5^{г}Рамме, а разносятся по всей карте в соответствии с тем Районом, областью или страной, которые они представляют.

- лементы простейшей картодиаграммы можно обнаружить на р ^{Лит}Ической карте, где города отличаются различными геомет-

^{Че}скими фигурами в зависимости от числа жителей.

В качестве примера картодиаграммы возьмем изображен валового сбора зерна Центрального района России (рис. 5,2тГ

Рис. *5.27.* Валовой сбор зерна Центрального района России (данные условные)

Изолинии (от греч. isos - равный, одинаковый, подобный) это линии равного значения какой-либо величины в ее распространении на поверхности, в частности на географической карте или графике. Изолиния отражает непрерывное изменение исследуемой величины в зависимости от двух других переменных и применяется при картографировании природных и социальноэкономических явлений. Изолинии используются для получения количественных характеристик исследуемых величин и для анализа корреляционных связей между ними.

Перечисленные виды графиков не являются исчерпывающими, но они наиболее часто употребляемы.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 5

Статистический график - чертеж, на котором статистические совокупности, характеризуемые определенными показателями, описываются с помощью условных геометрических образов или знаков.

Графический образ — совокупность точек, линий, фигур, с помощью которых изображаются статистические показатели.

П ле графика - часть плоскости, где расположены графические ° **пространственные ориентиры**

графика - система координатных

^{СеТ}**|\1асштабные ориентиры** — масштаб и система масштабных шкал.

Масштабная шкала - линия, отдельные точки которой могут быть

читаны как определенные числа (прямолинейная или криволинейная).

Носитель шкалы - прямая или **кривая линия.**

Экспликация — словесное описание содержания графика.

Координаты линейной диаграммы — оси *x* и *y* графика.

Абсцисса (ось x) - горизонтальная ось графика. На ней откладываются значения независимой переменной или времени, или значения признака.

Ордината (ось у) - вертикальная ось графика. На ней откладываются значения зависимой переменной или уровни ряда динамики, или частота повторения значений признака.

Диаграммы сравнения - столбиковые, ленточные, направленные, квадратные, круговые, фигур-знаков.

Структурные диаграммы - полосовые, столбиковые и секторные.

Диаграммы динамики —линейные, спиральные, радиальные, квадратные, круговые, ленточные, фигур-знаков, секторные.

Статистические карты - графическое изображение статистических данных на схематической географической карте, характеризующих уровень или степень распространения того или иного явления на определенной территории.

Картограмма - на схематическую географическую карту наносится штриховка различной частоты, точки или окраска определенной насыщенности, которая показывает сравнительную интенсивность какоголибо показателя в пределах каждой единицы нанесенного на карту территориального деления.

Картодиаграмма представляет собой сочетание диаграмм с географической картой.

ТЕСТ К ГЛАВЕ 5

1. Основными элементами статистического графика являются:
а) поле графика;
б) масштабные ориентиры;
^{в}) геометрические знаки;
^{г}) экспликация графика.

2. Какие виды диаграмм можно использовать по форме геометрического образа?
а) линейные;
б) плоскостные;
^{в}) объемные;
^{г}) статистические карты.

3. Каковы виды статистических графиков по способу построен
а) диаграммы; "•
б) статистические карты;
в) линейные;
г) плоскостные.

4. Каковы виды статистических графиков по задачам изображ ния социально-экономических явлений?
а) диаграммы сравнения;
б) диаграммы динамики;
в) диаграммы структуры;
г) картограммы;
д) картодиаграммы.

5. Выберите способ графического изображения данных о распределении научных работников по отраслям наук на конец года по региону:
а) картограмма;
б) картодиаграмма;
в) столбиковая;

г) секторная.

6. При изображении структуры и структурных сдвигов в совокупности явлений на графике применяются диаграммы:

а) полосовые;
б) квадратные;
в) секторные;
г) фигур-знаков.

7. При построении линейных диаграмм используются масштабные шкалы:

а) равномерные;
б) логарифмические; • в) радиальные.

8. При изображении на графике сезонных колебаний применяются диаграммы:

а) линейные;
б) радиальные;
в) спиральные;
г) столбиковые.

9. При изображении взаимосвязи между факторным и результативным признаками на графике применяются диаграммы:

а) столбиковые;
б) линейные;
в) фигур-знаков;
г) круговые.

10. При изображении социально-экономических явлений при помощи картограмм применяются их виды:

а) фоновые;
б) точечные;
в) знаков-символов.

ГЛАВА 6

СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

6.1

ПОНЯТИЕ, ФОРМЫ ВЫРАЖЕНИЯ И ВИДЫ СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Статистическое исследование независимо от его масштабов и целей всегда завершается расчетом и анализом различных по виду и форме выражения статистических показателей.

Статистический показатель представляет собой количественную характеристику социально-экономических явлений и процессов в условиях качественной определенности. Качественная определенность показателя заключается в том, что он непосредственно связан с внутренним содержанием изучаемого явления или процесса, его сущностью.

Как правило, изучаемые статистикой процессы и явления достаточно сложны, и их сущность не может быть отражена посредством одного отдельно взятого показателя. В таких случаях используется система статистических показателей.

Система статистических показателей - это совокупность взаимосвязанных показателей, имеющая одноуровневую или многоуровневую структуру и нацеленная на решение конкретной статистической задачи. Так, например, сущность промышленного предприятия заключается в производстве какой-либо продукции на базе эффективного взаимодействия средств производства и трудовых ресурсов. Следовательно, для полной экономической характеристики функционирования предприятия необходимо использовать систему, включающую прежде всего такие показатели, как прибыль, рентабельность, численность промышленно-производственного персонала, производительность труда, фондовооруженность и др.

В отличие от признака статистический показатель получается расчетным путем. Это могут быть простой подсчет единиц сово^{К}Упности, суммирование их значений признака, сравнение двух ^{и}ли нескольких величин или более сложные расчеты.

Различают конкретный статистический показатель и показатель-категорию.

Конкретный статистический показатель характеризует раз$_{м}$ величину изучаемого явления или процесса в данном месте ц*^{5}' данное время (под привязкой к месту понимается отношение пок ^{В} зателя к какой-либо территории или объекту). Так, если мы назь" ваем конкретную величину стоимости промышленно-производствен ных фондов, то обязательно должны указать, к какому предпр$_{ия}$ тию или отрасли и какому моменту времени она относится. Однако в теоретических работах и на этапе проектирования статистического наблюдения (при построении системы статистических показателей, обосновании методики их расчета) также оперируют и абстрактными показателями или показателями-категориями.

Показатель-категория отражает сущность, общие отличительные свойства конкретных статистических показателей одного и того же вида, без указания места, времени и числового значения. Например, показатели розничного товарооборота предприятий торговли и общественного питания в Москве и Санкт-Петербурге в 1990 и 1998 гг. отличаются местом, временем и конкретными числовыми значениями, но имеют одну и ту же сущность (продажа товаров через розничную торговую сеть и сеть предприятий общественного питания), которая отражена в показателе-категории «Розничный товарооборот предприятий торговли и общественного питания».

Все статистические показатели по охвату единиц совокупности разделяются на индивидуальные и сводные, а по форме выражения - на абсолютные, относительные и средние.

Индивидуальные показатели характеризуют отдельный объект или отдельную единицу совокупности - предприятие, фирму, банк, домохозяйство и т. п. Примером индивидуальных абсолютных показателей может служить численность промышленно-производственного персонала предприятия, оборот торговой фирмы, совокупный доход домохозяйства.

На основе соотнесения двух индивидуальных абсолютных показателей, характеризующих один и тот же объект или единицу, получают индивидуальный относительный показатель. В статистике рассчитываются и индивидуальные средние показатели, но только во временном измерении (среднегодовая численность промышленно-производственного персонала предприятия).

Сводные показатели в отличие от индивидуальных характеризуют группу единиц, представляющую собой часть статистической совокупности или всю совокупность в целом. Эти показатели, в свою очередь, подразделяются на объемные и расчетные^

Объемные показатели получают путем сложения значении признака отдельных единиц совокупности. Полученная величина, называемая объемом признака, может выступать в каче-

объемного абсолютного показателя (например, стоимость СТВ<овных фондов предприятий отрасли), а может сравниваться оС ЛУГОЙ объемной абсолютной величиной (например, с чисС ' ностью промышленно-производственного персонала этих ле пРИятий) или объемом совокупности (в данном примере "числом предприятии). В последних двух случаях получают Сбъемный относительный и объемный средний показатели (в аших примерах - фондовооруженность и средняя стоимость основных фондов).

Расчетные показатели, вычисляемые по различным формулам, служат для решения отдельных статистических задач анализа - измерения вариации, характеристики структурных сдвигов, оценки взаимосвязи и т. д. Они также делятся на абсолютные, относительные или средние. В эту группу входят индексы, коэффициенты тесноты связи, ошибки выборки и прочие показатели, подробно рассмотренные в соответствующих главах.

Охват единиц совокупности и форма выражения являются основными, но не единственными классификационными признаками статистических показателей. Важным классификационным признаком является также временной фактор. Социально-экономические процессы и явления находят свое отражение в статистических показателях либо по состоянию на определенный момент времени, как правило, на определенную дату, начало или конец месяца, года (численность населения, стоимость основных фондов, дебиторская задолженность), либо за определенный период — день, неделю, месяц, квартал, год (производство продукции, число заключенных браков, сумма страховых выплат). В первом случае показатели являются **моментными,** во втором — **интервальными.**

В зависимости от принадлежности к одному или двум объектам изучения различают **однообъектные и межобъектные показатели.** Если первые характеризуют только один объект, то вторые получают в результате сопоставления двух величин, относящихся к разным объектам (соотношение численности населения городов Екатеринбурга и Челябинска, соотношение численности детей дошкольного возраста и числа мест в детских Дошкольных учреждениях и т. п.). Межобъектные показатели Сражаются в форме относительных или средних величин.

С точки зрения пространственной определенности статистические показатели подразделяются на **общетерриториальные,** Характеризующие изучаемый объект или явление, в целом по Стране. **региональные и местные (локальные),** относящиеся к акой-либо части территории или отдельному объекту.

6.2

АБСОЛЮТНЫЕ ПОКАЗАТЕЛИ

Исходной, первичной формой выражения статистических пок зателей являются абсолютные величины. Статистические пока/ тели в форме абсолютных величин характеризуют абсолютны размеры изучаемых статистикой процессов и явлений: их масел площадь, объем, протяженность; отражают их временные харак теристики, а также могут представлять объем совокупности, т. е число составляющих ее единиц.

Индивидуальные абсолютные показатели, как правило, получают непосредственно в процессе статистического наблюдения как результат замера, взвешивания, подсчета и оценки интересующего количественного признака. В ряде случаев индивидуальные абсолютные показатели имеют разностный характер: разность между численностью зарегистрированных безработных в данном населенном пункте на конец и на начало года, разность между выручкой от реализации торгового предприятия и общей суммой затрат и т. п.

Сводные объемные показатели, характеризующие объем признака или объем совокупности как в целом по изучаемому объекту, так и по какой-либо его части, получают в результате сводки и группировки индивидуальных значений. " Абсолютные статистические показатели всегда являются именованными числами. В зависимости от социально-экономической сущности исследуемых явлений, их физических свойств они выражаются в натуральных, стоимостных или трудовых единицах измерения.

В международной практике используются такие **натуральные единицы измерения,** как тонны, килограммы,- унции, квадратные, кубические и простые метры, мили, километры, галлоны, литры, штуки и т. д. Например, производство электроэнергии в России в марте 1998 г. составило 79,0 млрд кВт-ч, в этом ж$^{£}$ месяце добыто 25,0 млн т нефти и 2,4 млрд м^{3} газа.

В группу натуральных также входят условно-натуральные измерители, используемые в тех случаях, когда какой-либо продукт имеет несколько разновидностей, и общий объем можно опреД$^{6'}$ лить только исходя из общего для всех разновидностей потребительского свойства. Так, различные виды органического топлива переводятся'в условное топливо с теплотой сгорания 29,3 МДж^ (7000 Ккал/кг); мыло разных сортов - в условное мыло с 40/°" ным содержанием жирных кислот; консервы различного объема в условные консервные банки объемом 353,4 см^{3} и т. Д-

$_{ев}$од в условные единицы измерения осуществляется на *в* специальных коэффициентов, рассчитываемых как отно$^{\circ CH\circ}$ & потребительских свойств отдельных разновидностей про$^{UJ£H}$ $_{к}$ эталонному значению. Так, например, 100 т торфа, теп*Р-У*^{К} $_{сго}$рания которого 24 МДж/кг, будут эквивалентны 81,9 т ровного *топлива* (100-24,0/29,3), а 100 т нефти при теплоте У^{с} $_{пания}$ 45 МДж/кг оцениваются в 153,6 т условного топлива Д. 45,0/29,3).

В отдельных *случаях* для характеристики какого-либо явлеили процесса одной единицы измерения недостаточно и используется произведение двух единиц. Например, показатели грузооборота и пассажирооборота, оцениваемые соответственно в $_{То}$нно-километрах и пассажиро-километрах, производство электроэнергии, измеряемое в киловатт-часах, и т. д.

В условиях рыночной экономики наибольшее значение и применение *имеют* **стоимостные единицы измерения,** дающие денежную оценку социально-экономическим явлениям и процессам. Так, одним из важнейших стоимостных показателей в системе национальных счетов, характеризующим общий уровень развития экономики страны, является валовой внутренний продукт, который в России в 1997 г. составил 2675 трлн руб.

При анализе и сопоставлении стоимостных показателей необходимо иметь в виду, что в условиях высоких темпов инфляции они становятся несопоставимыми. Так, сравнивать указанный выше ВВП России в 1997 г. с его величиной в 1996 г. вряд ли целесообразно, так как содержание рубля за этот период изменилось. Для того чтобы произвести подобные сравнения, там где это возможно, осуществляют пересчет в сопоставимые цены.

К трудовым единицам измерения, позволяющим учитывать как общие затраты труда на предприятии, так и трудоемкость отдельных операций технологического процесса, относятся человеко-дни и человеко-часы.

6.3

ОТНОСИТЕЛЬНЫЕ ПОКАЗАТЕЛИ

Относительный показатель представляет собой результат де^{е}ния одного абсолютного показателя на другой и выражает соотШение между количественными характеристиками социальнономических процессов и явлений. Поэтому по отношению к олютным показателям относительные показатели или показате-
ли в форме относительных величин являются производными (вт ричными). Без относительных показателей невозможно измени интенсивность развития изучаемого явления во времени, оценит" уровень развития одного явления на фоне других взаимосвязанны с ним явлений, осуществить пространственно-территориальны сравнения, в том числе и на международном уровне.

При расчете относительного показателя абсолютный показатель, находящийся в числителе получаемого отношения, называется **текущим** или **сравниваемым.** Показатель же, с которы$_{м}$ производится сравнение и который находится в знаменателе называется **основанием или базой сравнения.** Таким образом рассчитываемый относительный показатель указывает, во сколько раз сравниваемый абсолютный показатель больше базисного или какую он составляет от него долю, или сколько единиц первого приходится на 1, 100, 1000 и т. д. единиц второго.

Относительные показатели могут выражаться в коэффициентах, процентах, промилле, продецимилле или быть именованными числами. Если база сравнения принимается за 1, то относительный показатель выражается в коэффициентах, если база принимается за 100, 1000 или 10 000, то относительный показатель соответственно выражается в процентах ($^{\circ}/_{0}$), промилле (‰$_{0}$) и продецимилле ($^{\circ}/_{00}$).

Проценты, как правило, используются в тех случаях, когда сравниваемый абсолютный показатель превосходит базисный не более чем в 2 - 3 раза. Проценты же свыше 200 - 300 обычно заменяются кратным отношением, коэффициентом. Так, вместо 470% говорят, что сравниваемый показатель превосходит базисный в 4,7 раза.

Относительный показатель, полученный в результате соотнесения разноименных абсолютных показателей, в большинстве случаев должен быть именованным. Его наименование представляет собой сочетание наименований сравниваемого и базисного показателей (например, производство какой-либо продукции в соответствующих единицах измерения в расчете на душу населения).

Все используемые на практике *относительные статистические показатели можно подразделить на следующие виды:*

- динамики;
- плана;
- реализации плана;
- структуры;
- координации;
- интенсивности и уровня экономического развития;
- сравнения.

Относительный показатель динамики (ОГГД) представляет бой отношение уровня исследуемого процесса или явления за °° ный период времени (по состоянию на данный момент вре^^{а} ч $_{к}$ уровню этого же процесса или явления в прошлом:

$$ОПД = \frac{\text{Текущий показатель}}{\text{Предшествующий или базисный показатель}}.$$

Рассчитанная таким оОразом величина показывает, во сколько аз текущий уровень превышает предшествующий (базисный) или какую долю от последнего составляет. Если данный показатель выражен кратным отношением, он называется **коэффициентом роста,** при домножении этого коэффициента на 100% получают **темп роста.**

Например, если известно, что оборот торгов Московской межбанковской валютной биржи 25 марта 1998 г. составил 51,9 млн долл., а 24 марта - 43,2 млн долл., то относительный показатель

[519 1
динамики, или темп роста, будет равен: 120,1%=——-100% .
└4j,~ j

Все субъекты финансово-хозяйственной сферы,, начиная от небольших семейных предприятий и заканчивая крупными концернами, в той или иной степени осуществляют перспективное планирование своей деятельности, а также сравнивают реально . достигнутые результаты с ранее намеченными. Для этой цели используются **относительные показатели плана (ОПП) и реализации" плана (ОПРП):**

Предположим, оГюрот ком мер ческой фирмы в 1997 г. соста^{Вил} 2,0 млрд руб. Исходя из проведенного анализа складывающихся на рынке тенденций, руководство фирмы считает реаль$^{Чым в}$

$$ОПП = \frac{\text{Показатель, планируемый на (i+1) период}}{\text{Показатель, достигнутый в i-м периоде}};$$

$$ОПРП = \frac{\text{Показатель, достигнутый в (i+1) периоде}}{\text{Показатель, планируемый на (i+1) период}}$$

следующем году довести оборот до 2,8 млрд руб. В этом ^{л}Учае относительный показатель плана, представляющий собой ^{н}ошение планируемой величины к фактически достигнутой, ^ставит Н0% (2,8/2,0 • 100%). Предположим теперь, что фак^{е}ский оборот фирмы за 1998 г. составил 2,6 млрд руб. Тогда относительный показатель реализации плана, определяемый как отношение фактически достигнутой величины к ранее планированной, составит 92,9% (2,6/2,8-100%).

Между относительными показателями плана, реализации плана и динамики существует следующая взаимосвязь:

$$ОПП \cdot ОПРП = ОПД.$$

В нашем примере:

$$1{,}40 \cdot 0{,}929 = 1{,}3, \text{ или } ОПД = \frac{2{,}6}{2{,}0} \approx 1{,}3.$$

Основываясь на этой взаимосвязи по любым двум известным величинам, при необходимости всегда можно определить третью, неизвестную величину.

Относительный показатель структуры (ОПС) представляет собой соотношение структурных частей изучаемого объекта и их целого:

$$\text{ОПС} = \frac{\text{Показатель, характеризующий часть совокупности}}{\text{Показатель по всей совокупности в целом}}.$$

Относительный показатель структуры выражается в долях единицы или в процентах. Рассчитанные величины (dj), соответственно называемые долями или удельными весами, показывают, какой долей обладает или какой удельный вес имеет *1-я* часть в общем итоге.

Рассмотрим структуру внешнеторгового оборота РФ в 1997 г. (табл. 6.1).

Таб'лица 6.1 Структура внешнеторгового оборота РФ в 1997 г.

	Трлн руб.	% к итогу
А	1	2
Внешнеторговый оборот — всего	896,7	100,0
в том числе:		
экспорт	505,6	56,4
импорт	391,1	43,6

Рассчитанные в графе 2 табл. 6.1 проценты представляют собой относительные показатели структуры (в данном случае удельные веса). Они получены как отношения объемов экспорта и импорта к общему объему внешнеторгового оборота РФ. Сумма всех удельных весов всегда должна быть строго равна 100%*.

Относительные показатели координации (ОПК) характеризуют соотношение отдельных частей целого между собой:

$$\text{ОПК} = \frac{\text{Показатель, характеризующий i-ю часть совокупности}}{\text{Показатель, характеризующий часть совокупности, выбранную в качестве базы сравнения}}.$$

При этом в качестве базы сравнения выбирается та часть, которая имеет наибольший удельный вес или является приоритетной с экономической, социальной или какой-либо другой точки зрения. В результате получают, сколько единиц каждой структурной части приходится на 1 единицу (иногда на 100, 1000 и т. д. единиц) базисной структурной части. Так, на основе данных приведенной выше табл. 6.1 мы можем вычислить, что на каждый триллион рублей импорта приходилось 1,29 трлн руб. экспорта (505,6/391,1).

Относительный показатель интенсивности (ОПИ) характеризует степень распространения изучаемого процесса или явления в присущей ему среде:

$$\text{ОПИ} = \frac{\text{Показатель, характеризующий явление А}}{\text{Показатель, характеризующий среду распространения явления А}}.$$

Этот показатель исчисляется, когда абсолютная величина оказывается недостаточной для формулировки обоснованных выводов о масштабах явления, его размерах, насыщенности, плотности распространения. Как и в предшествующем случае, он может выражаться в процентах, промилле или быть именованной величиной. Например, для определения уровня рождаемости, измеряемого в $^0/_{00}$, рассчитывается число родившихся на ЮОО человек населения, для определения плотности населения рассчитывается число людей, приходящихся на 1 км^2 территории.

Расчет относительных показателей интенсивности в ряде случаев связан с проблемой выбора наиболее обоснованной, соответствующей данному процессу или явлению базы сравнения.

* Показатели, позволяющие оценить структурные сдвиги в пространстве и во времени, будут рассмотрены в соответствующей главе.

Разновидностью относительных показателей интенсивности являются **относительные показатели**

уровня экономического развития, характеризующие производство продукции в расчете на душу населения и играющие важную роль в оценке развития экономики государства. Так, зная лишь то, что валовой внутренний продукт России в 1997 г. составил 2675 трлн руб., мы не можем сказать, насколько это много, оценить, «почувствовать» эту величину. Для того чтобы на основе этой цифры сделать вывод об уровне развития экономики, необходимо сопоставить ее с численностью населения страны (146,9 млн человек)*. В результате размер ВВП на душу населения составит 18,2 млн руб. (2675 трлн руб./-146,9 млн человек). Сделав перерасчет на доллары, данный относительный показатель уже можно использовать для временных и территориальных сравнений (между странами).

По форме выражения относительные показатели интенсивности и уровня экономического развития близки средним показателям, что нередко приводит к их смешиванию или отождествлению. Разница же между ними заключается в том, что при расчете среднего показателя мы имеем дело с совокупностью единиц, каждая из которых является носителем осредняемого признака. Например, при расчете среднедушевого дохода осредняется масса индивидуальных доходов отдельных людей. При расчете же относительного показателя интенсивности каждая единица не является носителем признака (при определении плотности населения отсутствует какое-либо закрепление конкретной территории за конкретными людьми).

Относительный показатель сравнения (ОПСр) представляет собой соотношение одноименных абсолютных показателей, характеризующих разные объекты (предприятия, фирмы, районы, области, страны и т. п.):

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

Например, располагая данными па конец 1993 г. о размере инвестиционных фондов США (3583 млрд марок), Европы (2159 млрд марок) и Японии (785 млрд марок), можно на основе относительных показателей сравнения сделать вывод о том, что инвестиционные фонды США в 1,7 раза мощнее европейских (3583/ 2159) и в 4,6 раза больше японских (3583/785).

---

* При отличных от нуля темпах роста численности населения необходимо использовать среднегодовое значение, исчисляемое как полусумма показателей на начало и конец периода.

6.4

СУЩНОСТЬ И ЗНАЧЕНИЕ СРЕДНИХ ПОКАЗАТЕЛЕЙ

Наиболее распространенной формой статистических показателей, используемой в социально-экономических исследованиях, является **средняя величина, представляющая собой обобщенную количественную характеристику признака в статистической совокупности в конкретных условиях места и времени.** Показатель в форме средней величины выражает типичные черты и дает обобщенную характеристику однотипных явлений по одному из варьирующих признаков. Он отражает уровень этого признака, отнесенный к единице совокупности. Широкое применение средних объясняется тем, что они имеют ряд положительных свойств, делающих их незаменимыми в анализе явлений и процессов общественной жизни.

Проиллюстрируем значение средних показателей на следующем примере. Одной из задач органов государственной статистики является характеристика уровня жизни населения в целом и, в частности, уровня его доходов в разрезе различных социальных групп. Очевидно, что данный объект включает столь большое число единиц, что сравнение индивидуальных доходов каждой семьи рабочего, служащего, предпринимателя, студента и т. д. является абсолютно невозможным. Не представляет особого интереса и сравнение суммарных доходов отдельных социальных групп, так как эти группы существенно различаются по численности (например, численность рабочих и численность людей, занятых в сфере предпринимательства). В данном случае мы можем использовать лишь средние показатели, а именно среднюю величину доходов в расчете на одного человека или на одну семью по каждой социальной группе.

Важнейшее свойство средней величины заключается в том, что она отражает то общее, что присуще всем единицам исследуемой совокупности. Значения признака отдельных единиц совокупности могут колебаться в ту или иную сторону под влиянием множества» факторов, среди которых могут быть как основные, так и случайные. Например, доходы такой социальной группы, как студенты государственных вузов, в целом определяются действующим положением о начислении стипендии. В то же время доходы отдельно взятого студента могут быть и очень большими (предположим, вследствие занятия каким-либо бизнесом в свободное от учебы время или хорошо оплачиваемых сезонных работ), и совсем отсутствовать (например, при нахождении в академическом отпуске). Сущность средней в том и заключается, что в ней взаимопогашаются отклонения значений признака отдельных единиц совокупности, обусловленные действием случайных факторов, и учитываются изменения, вызванные действием факторов основных. Это позволяет средней отражать типичный уровень признака и абстрагироваться от индивидуальных особенностей, присущих отдельным единицам. Возможно, что ни один студент в границах исследуемой совокупности не имеет с точностью до рубля такого дохода, какой получен на основе расчета средней. Однако эта средняя отражает тот типичный уровень доходов, который характеризует студенчество как социальную группу.

Типичность средней непосредственным образом связана с однородностью статистической совокупности. Средняя величина только тогда будет отражать типичный уровень признака, когда она рассчитана по качественно однородной совокупности. Так, в приведенном примере, если мы рассчитаем средний уровень доходов служащих, то получим фиктивную среднюю. Это объясняется тем, что используемая для расчета средней совокупность, включающая служащих государственных, совместных, арендных, акционерных предприятий, а также органов государственного управления, сферы науки, культуры, образования и т. п., является крайне неоднородной. В этом и подобных случаях метод средних используется в сочетании с методом группировок: если совокупность неоднородна - общие средние должны быть заменены или дополнены групповыми средними, т. е. средними, рассчитанными по качественно однородным группам.

Теория средних достаточно подробно разработана в отечественных и зарубежных исследованиях. Среди ученых, внесших свой вклад в ее развитие, необходимо отметить И. Зюсмильха, А. Кетле, А. Боули, К. Джини, А. Я. Боярского, Т. В. Рябушкина, И. С. Пасхавера, В. Е. Овсиенко и др.

Сущность средней можно раскрыть через понятие ее **определяющего свойства,** сформулированное А. Я. Боярским и О. Кизини: средняя, являясь обобщающей характеристикой всей статистической совокупности, должна ориентироваться на определенную величину, связанную со всеми единицами этой совокупности. Эту величину можно представить в виде функции:

$$f(x_1, x_2, ..., x_n). \quad (6.1)$$

Так как данная величина в большинстве-случаев отражает реальную экономическую категорию, ее называют **определяющим показателем.**

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

$$f(x_1, x_2, ..., x_n) = f(\bar{x}, \bar{x}, ..., \bar{x}). \quad (6.2)$$

Исходя из данного равенства и определяется средняя. Определить среднюю во многих случаях можно через **исходное соотношение средней** (ИСС) или ее логическую формулу:

$$ИСС = \frac{\text{Суммарное значение или объем осредняемого признака}}{\text{Число единиц или объем совокупности}}.$$

Так, например, для расчета средней заработной платы работ' ников предприятия необходимо общий фонд заработной платы разделить на число работников:

$$ИСС = \frac{\text{Фонд заработной платы, тыс. руб.}}{\text{Число работников, человек}}.$$

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

Для каждого показателя, используемого в социально-экономическом анализе, можно составить только одно истинное исходное соотношение для расчета средней. Если, например, требуется рассчитать средний размер вклада в банке, то исходное соотношение будет следующим:

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

При необходимости определения средней процентной ставки по кредитам, выданным на один и тот же срок, потребуется следующее исходное соотношение:

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

Однако от того, в каком виде представлены исходные даннь для расчета средней, зависит, каким именно образом будет DC ^{C} лизовано ее исходное соотношение. *В каждом конкретном сл\ чае для реализации исходного соотношения потребуется одн из следующих форм средней величины:*

1) средняя арифметическая;
2) средняя гармоническая;
3) средняя геометрическая;
4) средняя квадратическая, кубическая и т. д. Перечисленные средние объединяются в общей формуле **средней степенной** (при различной величине k):

$$\bar{x} = \sqrt[k]{\frac{\Sigma x_i^k \cdot f_i}{\Sigma f_i}} ,$$

где x - средняя величина исследуемого явления;__

x, - i-й вариант осредняемого признака (i = $\overline{1, п}$); fj — вес i-ro варианта.

Помимо степенных средних в статистической практике также используются средние структурные, среди которых наиболее распространены мода и медиана.

При осреднении уровней динамических рядов применяются различные виды средней хронологической, которые будут рассмотрены в соответствующей главе.

6.5

СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ И ЕЕ СВОЙСТВА

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

Средняя арифметическая простая (невзвешенная). Эта форма средней используется в тех случаях, когда расчет осуществляется по несгруппированным данным.

Предположим, семь членов бригады имеют следующий стаж работы:

Табельный номер рабочего	1	2	3	4	5	6	7
Стаж работы (лет)	10	3	5	12	11	7	9

Для того чтобы определить средний стаж работы, необходимо воспользоваться следующим исходным соотношением:

$$ИСС = \frac{\text{Совокупный стаж работы всех рабочих, лет}}{\text{Число рабочих}} .$$

Используя приведенные в предыдущем параграфе условные обозначения, запишем формулу данной средней:

$$\bar{x} = \frac{x_1 + x_2 +...+ x_n}{n} = \frac{\Sigma x_i}{n}.$$

(6.3)

С учетом имеющихся данных получим:

$$\bar{x} = \frac{10+3+5+12+11+7+9}{7} = 8{,}1 \text{ года.}$$

В этом случае мы использовали формулу средней арифметической простой (невзвешенной).

Средняя арифметическая взвешенная. При расчете средних величин отдельные значения осредняемого признака могут повторяться, встречаться по нескольку раз. В подобных случаях расчет средней производится по сгруппированным данным или вариационным рядам, которые могут быть дискретными или интервальными.

Рассмотрим следующий пример из биржевой практики (табл. 6.2).

Таблица 6.2

Продажа акций АО «Дока-хлеб» на торгах фондовой секции ТМБ «Гермес»

Сделка	Количество проданных	курс продажи, руб. акций, шт.
1	500	1080
2,	300	1050
3	1 100	1145

Определим по данному дискретному вариационному ряду средний курс продажи акции, 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 результат следующий:

$$\bar{x} = \frac{1080 \cdot 500 + 1050 \cdot 300 + 1145 \cdot 1100}{500 + 300 + 1100} = \frac{2114500}{1900} = 1112{,}9 \text{ руб.}$$

Расчет среднего курса продажи произведен по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\Sigma x_i \cdot f_i}{\Sigma f_i}. \qquad (6.4)$$

В отдельных случаях веса могут быть представлены не абсолютными величинами, а относительными (в процентах или долях единицы). Так, в приведенном выше примере количество проданных в ходе каждой сделки акций соответственно составляет 26,3% (0,263), 15,8% (0,158) и 57,9% (0,579) от их общего числа. Тогда с учетом несложного преобразования формулы (6.4) получим:

$$\bar{x} = \Sigma\left(x_i \cdot \frac{f_i}{\Sigma f_i}\right), \qquad (6.5)$$

или $\bar{x} = 1080 \cdot 0{,}263 + 1050 \cdot 0{,}158 + 1145 \cdot 0{,}579 = 1112{,}9$ руб.

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

Можно ли по имеющимся данным определить среднюю заработную плату по предприятию в целом? Можно, но только в том случае, когда численность работников в 1 -м и 2-м цехах совпадает. Тогда средняя заработная плата по предприятию в целом составит 420 руб. (доказательство этого правила будет приведе-

Таблица 6.3 Заработная плата работников предприятия за май 1998 г.

Цех	Средняя заработная плата, руб.
1	430
2	410

но ниже). Однако в цехе 1 может быть занято, к примеру, 10 человек, а в цехе 2 - 100. Поэтому для расчета средней заработной платы потребуется средняя арифметическая взвешенная:

$$\bar{x} = \frac{430\cdot 10 + 410\cdot 100}{110} \approx 411,8 \text{ руб.}$$

Общий вывод заключается в следующем: использовать среднюю арифметическую невзвешенную можно только тогда, когда точно установлено отсутствие весов или их равенство.

При расчете средней по **интервальному вариационному ряду** для выполнения необходимых вычислений от интервалов переходят к их серединам. Рассмотрим следующий пример (табл. 6.4).

Таблица 6.4 **Распределение работников предприятия по возрасту**

Возраст, лет	Число работников, человек
До 25	7
25-30	13
30-40	38
40-50	42
50 - 60	16
60 и более	5
Итого	121

Для определения среднего возраста работника найдем середины возрастных интервалов. При этом величины открытых интервалов (первого и последнего) условно приравниваются к величинам интервалов, примыкающих к ним (второго и предпоследнего). Согласно вышеизложенному, середины интервалов будут следующими:

22,5 27,5 35,0 45,0 55,0 65,0.

Используя среднюю арифметическую взвешенную, определим средний возраст работника данного предприятия:

$$\bar{x} = \frac{22,5\cdot 7 + 27,5\cdot 13 + 35\cdot 38 + 45\cdot 42 + 55\cdot 16 + 65\cdot 5}{7 + 13 + 38 + 42 + 16 + 5} = 41 \text{ год.}$$

Свойства средней арифметической. Средняя арифметическая обладает некоторыми математическими свойствами, более полно раскрывающими ее сущность и в ряде случаев используемыми при ее расчетах. Рассмотрим эти свойства.

1. Произведение средней на сумму частот равно сумме произведений отдельных вариантов на соответствующие им частоты:

$$\bar{x}\Sigma f_i = \Sigma x_i \cdot f_i. \qquad (6.6)$$

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

$$1112{,}9 \cdot 1900 = 1080 \cdot 500 + 1050 \cdot 300 + 1145 \cdot 1100.$$

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

$$\Sigma(x_i - \bar{x}) \cdot f_i = 0.$$ (6.7) Для нашего примера:

$$(1080 - 1112{,}9) \cdot 500 + (1050 - 1112{,}9) \cdot 300 + (1145 - 1112{,}9) \cdot 1100 = 0.$$

Математическое доказательство данного свойства сводится к следующему:

$$\Sigma(x_i - \bar{x}) \cdot f_i = \Sigma x_i \cdot f_i - \Sigma\bar{x} \cdot f_i = \Sigma x_i \cdot f_i - \bar{x}\Sigma f_i = 0.$$

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

$$\begin{aligned}\Sigma(x_i - C)^2 \cdot f_i &= \Sigma(x_i - \bar{x} + \bar{x} - C)^2 \cdot f_i = \Sigma[(x_i - \bar{x}) + (\bar{x} - C)]^2 \cdot f_i = \\ &= \Sigma[(x_i - \bar{x})^2 + 2 \cdot (x_i - \bar{x}) \cdot (\bar{x} - C) + (\bar{x} - C)^2] \cdot f_i = \\ &= \Sigma(x_i - \bar{x})^2 \cdot f_i + 2 \cdot (\bar{x} - C)\Sigma(x_i - \bar{x}) \cdot f_i + \Sigma(\bar{x} - C)^2 \cdot f_i = \\ &= \Sigma(x_i - \bar{x})^2 \cdot f_i + 2 \cdot (\bar{x} - C) \cdot 0 + \Sigma(\bar{x} - C)^2 \cdot f_i. \qquad (6.8)\end{aligned}$$

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

$$\Sigma(\bar{x} - C)^2 \cdot f_i \quad \text{или} \quad (\bar{x} - C)^2 \cdot \Sigma f_i.$$

На использовании этого свойства базируется расчет центральных моментов, представляющих собой характеристики вариационного ряда при С = x*.

$$\mu_k = \frac{\Sigma(x_i - \bar{x})^k \cdot f_i}{\Sigma f_i},$$

где k определяет порядок момента (центральный момент второго порядка представляет собой дисперсию).

4. Если все осредняемые варианты уменьшить или увеличить на постоянное число А, то средняя арифметическая соответственно уменьшится или увеличится на ту же величину:

$$\frac{\Sigma(x_i \pm A) \cdot f_i}{\Sigma f_i} = \frac{\Sigma x_i \cdot f_i}{\Sigma f_i} \pm \frac{\Sigma A \cdot f_i}{\Sigma f_i} = \bar{x} \pm A. \qquad (6.9)$$

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

$$\bar{x} = \frac{1180 \cdot 500 + 1150 \cdot 300 + 1245 \cdot 1100}{1900} = 1212{,}9 \text{ руб.}$$

5. Если все варианты значений признака уменьшить или увеличить в А раз, то средняя также соответственно увеличится или уменьшится в А раз:

$$\frac{\Sigma(x_i / A) \cdot f_i}{\Sigma f_i} = \frac{(1/A)\Sigma x_i \cdot f_i}{\Sigma f_i} = \frac{1}{A}\bar{x}. \qquad (6.10)$$

* При С = 0 получают начальные моменты (начальный момент первого порядка — средняя арифметическая и т. д.).

Предположим, курс продажи в каждом случае возрастет в 1,5 раза. Тогда и средний курс увеличится на 50%.

$$\bar{x} = \frac{1080\cdot 1,5\cdot 500+1050\cdot 1,5\cdot 300+1145\cdot 1,5\cdot 1100}{1900} = 1112,9\cdot 1,5 = 1669,4 \text{ руб.}$$

6. Если все веса уменьшить или увеличить в А раз, то средняя арифметическая от этого не изменится:

$$\frac{\Sigma x_i\cdot (f_i/A)}{\Sigma (f_i/A)} = \frac{(1/A)\Sigma x_i\cdot f_i}{(1/A)\Sigma f_i} = \bar{x}. \qquad (6\text{-}11)$$

Так, в нашем примере удобнее было бы рассчитывать среднюю, предварительно поделив все веса на 100:

$$\bar{x} = \frac{1080\cdot 5+1050\cdot 3+1145\cdot 11}{19} = 1112,9 \text{ руб.}$$

Исходя из данного свойства можно заключить, что в случае равенства всех весов между собой расчеты по средней арифметической взвешенной и средней арифметической простой приведут к одному и тому же результату.

6.6

ДРУГИЕ ВИДЫ СРЕДНИХ

При расчете статистических показателей помимо средней арифметической могут использоваться и другие виды средних. Однако в каждом конкретном случае в зависимости от характера имеющихся данных существует только одно истинное среднее значение показателя, являющееся следствием реализации его исходного соотношения.

Средняя гармоническая взвешенная. Рассмотрим вариант, когда известен числитель исходного соотношения средней, но неизвестен его знаменатель (табл. 6.5).

Средняя урожайность любой сельскохозяйственной культуры по нескольким территориям, агрбфирмам, крестьянским хозяйствам и т. п. может быть определена только на основе следующего исходного соотношения:

Таблица 6.5

Валовой сбор и урожайность подсолнечника по Центрально-Черноземному району (в хозяйствах всех категорий)

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

$$\text{ИСС} = \frac{\text{Общий валовой сбор, тыс. ц}}{\text{Общая посевная площадь, тыс. га}}.$$

Общий валовой сбор мы получим простым суммированием валового сбора по областям. Данные же о посевной площади в таблице отсутствуют, но их можно получить, разделив валовой сбор по каждой области на урожайность. С учетом этого определим искомую среднюю, предварительно переведя для сопоставимости тонны в центнеры:

$$\bar{x} = \frac{970+2040+5+160+690}{\frac{970}{16{,}1}+\frac{2040}{9{,}5}+\frac{5}{4{,}8}+\frac{160}{10{,}9}+\frac{690}{7{,}0}} = \frac{3865}{389{,}3} = 9{,}9 \text{ ц/га}.$$

Таким образом, общая посевная площадь подсолнечника по Центрально-Черноземному району составила 389,3 тыс. га, а средняя урожайность - 9,9 ц с 1 га.

В данном случае расчет произведен по формуле средней гармонической взвешенной:

$$\bar{x} = \frac{\Sigma w_i}{\Sigma \frac{w_i}{x_i}}, \qquad (6.12)$$

где $w_i = x_i \cdot f_i$.

Средняя гармоническая невзвешенная. Эта форма средней используемая значительно реже, имеет следующий вид:

$$\bar{x} = \frac{n}{\Sigma \frac{1}{x_i}}.$$

(6-13)

Для иллюстрации области ее применения воспользуемся упрощенным условным примером. Предположим, в автохозяйстве эксплуатируются два электромобиля разных моделей, работающих на однотипных подзаряжаемых за ночь аккумуляторных батареях. Первый электромобиль расходует на 1 км пути 1,0 кВт-ч электроэнергии, второй - 0,6 кВт-ч. Каков средний расход электроэнергии на 1 пройденный километр?

На первый взгляд решение этой задачи заключается в осреднении индивидуальных значений потребления электроэнергии по двум электромобилям, т. е. (1,0 + 0,6) : 2 = 0,8 кВт-ч. Проверим обоснованность такого подхода на примере одного дня работы машин, в течение которого они расходуют один заряд аккумулятора, предположим, 60,0 кВт-ч (как будет показано ниже, конкретная цифра значения не имеет). За этот день первая машина пройдет 60 км (60,0/1,0), пробег второй составит 100 км (60,0/0,6), т. е. в сумме - 160 км. Если же заменить индивидуальные значения признака их предполагаемым средним значением, то общий пробег, выступающий в данном случае в качестве определяющего показателя, сократится до 150 км (60,0/0,8 + 60,0/0,8). Следовательно, полученная средняя рассчитана неверно.

Рассмотрим решение данной задачи через исходное соотношение средней. Для того чтобы определить средний расход энергии на 1 пройденный километр, необходимо общий расход энергии за какой-либо временной промежуток (день, неделю, месяц) поделить на сделанный за это время суммарный пробег:

$$\bar{x} = \frac{60{,}0+60{,}0}{\frac{60{,}0}{1{,}0}+\frac{60{,}0}{0{,}6}} = \frac{1+1}{\frac{1}{1}+\frac{1}{0{,}6}} = \frac{2}{1+1{,}667} = 0{,}75 \text{ кВт}\cdot\text{ч}.$$

При замене индивидуального значения признака их средней величиной общий пробег не изменится:

$$\frac{60{,}0}{0{,}75}+\frac{60{,}0}{0{,}75} = 160 \text{ км}.$$

Подведем итог: средняя гармоническая невзвешенная может использоваться вместо взвешенной в тех случаях, когда значения Wj для единиц совокупности равны (машины расходуют ежедневно одно и то же количество электроэнергии).

Средняя геометрическая. Еще одной формулой, по которой может осуществляться расчет среднего показателя, является средняя геометрическая:

- невзвешенная:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} = \sqrt[n]{\Pi x_i},$$

- взвешенная:

$$\bar{x} = \sqrt[\Sigma f]{(x_1)^{f_1} \cdot (x_2)^{f_2} \cdot (x_3)^{f_3} \cdot \ldots \cdot (x_n)^{f_n}} = \sqrt[\Sigma f]{\Pi (x_i)^{f_i}}. \quad (6. И)$$

Наиболее широкое применение этот вид средней получил в анализе динамики для определения среднего темпа роста (подробнее см. гл. 10).

Средняя квадратическая. В основе вычислений ряда свод. ных расчетных показателей лежит средняя квадратическая:

$$\bar{x} = \sqrt{\frac{\Sigma x_i^2}{n}} \text{ - невзвешенная;}$$

(6.15)

$$\bar{x} = \sqrt{\frac{\Sigma x_i^2 f_i}{\Sigma f_i}} \text{ - взвешенная.}$$

Наиболее широко этот вид средней используется при расчете показателей вариации (гл. 7).

В статистической практике также находят применение степенные средние 3-го и более высоких порядков.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 6

Статистический показатель — количественная характеристика социально-экономических явлений и процессов в условиях качественной определенности.

Система статистических показателей - совокупность взаимосвязанных показателей, имеющая одноуровневую или многоуровневую структуру и нацеленная на решение конкретной статистической задачи.

Абсолютный показатель — показатель в форме абсолютной величины, отражающий физические свойства, временные или стоимостные характеристики социально-экономических процессов и явлений.

Относительный показатель - показатель в форме относительной величины, получаемый как результат деления одного абсолютного показателя на другой и отражающий соотношение между количественными характеристиками изучаемых процессов и явлений.

Средний показатель - показатель в форме средней величины, представляющий собой обобщенную количественную характеристику признака в статистической совокупности в конкретных условиях места и времени.

Средняя величина является наиболее ценной, с аналитической точки зрения, и универсальной формой выражения статистических показателей. Наиболее распространенная средняя - средняя арифметическая обладает рядом математических свойств, которые могут быть использованы при ее расчете. В то же время при исчислении конкретной средней всегда целесообразно опираться на ее логическую формулу, представляющую собой отношение объема признака к объему совокупности. Для каждой средней существует только одно истинное исходное соотношение, для реализации которого, в зависимости от имеющихся данных, могут потребоваться различные формы средних. Однако во всех случаях, когда характер осредняемой величины подразумевает наличие весов, нельзя вместо взвешенных формул средних использовать их невзвешенные формулы.

ТЕСТ К ГЛАВЕ 6

1. К какому виду по степени охвата единиц совокупности относится показатель «Активы коммерческого банка»?

а) индивидуальный; •б) сводный.

2. К какому виду по временному фактору относится показатель «Число рекламаций на продукцию предприятия»?

а) моментный;

- б) интервальный.

3. Чтобы получить относительный показатель динамики с переменной базой сравнения для i-го периода, необходимо:

а) перемножить относительные показатели динамики с постоянной базой сравнения за i-й и (i— 1)-й периоды;

* б) разделить относительный показатель динамики с постоянной базой сравнения за i-й период на аналогичный показатель за период (i-1);

в) разделить относительный показатель динамики с постоянной базой сравнения за i-й период на аналогичный показатель за период (i+1).

4. Относительный показатель реализации предприятием плана производства продукции составил 103%, при этом объем производства по сравнению с предшествующим периодом вырос на 2%. Что предусматривалось планом?

а) снижение объема производства;

б) рост объема производства.

5. Сумма относительных показателей координации, рассчитанных по одной совокупности, должна быть:

-а) строго равной 100;

б) меньше 100 или равной 100;

в) меньше, больше или равной 100.

6. Объем совокупности — это:

а) сумма всех значений осредняемого признака по совокупности; - б) общее число единиц в совокупности.

' **7. В каких случаях взвешенные и иевзвешенные средние равны между собой?**

а) при отсутствии весов;

б) при равенстве весов;

в) при отсутствии или равенстве весов.

8. В каких случаях используется средняя гармоническая?

а) когда неизвестен числитель исходного соотношения;

б) когда неизвестен знаменатель исходного соотношения.

9. Если веса осредняемого показателя выражены в промилле, чему будет равен знаменатель при расчете средней арифметической?

а) 100;

б) 1000;

в) 10000.

10. Изменится ли средняя величина, если все веса уменьшить на некоторую постоянную величину?

а) изменится;

б) не изменится.

АНАЛИТИЧЕСКАЯ СТАТИСТИКА

ГЛАВА 7

ПОКАЗАТЕЛИ ВАРИАЦИИ И АНАЛИЗ ЧАСТОТНЫХ РАСПРЕДЕЛЕНИЙ

7.1

ПОНЯТИЕ ВАРИАЦИИ И ЕЕ ЗНАЧЕНИЕ

Рассматривая зарегистрированные в процессе статистического наблюдения величины того или иного признака у отдельных единиц совокупности, можно обнаружить между ними различия.

Колеблемость, многообразие, изменяемость величины признака у единиц совокупности называются **вариацией.**

Вариация порождается комплексом условий, действующих на совокупность и ее единицы. Например, вариация оценок на экзамене в вузе порождается, в частности, различными способностями студентов, временем, затрачиваемым ими на самостоятельную работу, различием социально-бытовых условий и т. д. Именно вариация и предопределяет необходимость статистики. Если бы все студенты получали одинаковые оценки или, например, семьи имели одинаковые доходы, то необходимость в статистическом исследовании отпала бы.

Исследование вариации в статистике имеет важное значение. Оно дает возможность оценить степень воздействия на данный признак других варьирующих признаков, установить, например, какие факторы и в какой степени влияют на смертность населения, финансовое положение предприятий, урожайность пшеницы и т. п. Определение вариации необходимо при организации выборочного

наблюдения, построении статистических моделей,
разработке материалов экспертных опросов и во многих других случаях.

Вариация существует в пространстве и во времени. Под **вариацией в пространстве** понимается колеблемость значений признака по отдельным территориям.

Объективно существует также **вариация во времени.** Под ней подразумевают изменение значений признака в различные периоды (или моменты) времени. Так, со временем изменяются средняя продолжительность жизни, срок службы товаров длительного пользования, мнения людей и т. д.

Наличие вариации в признаках изучаемых явлений ставит перед статистикой задачи ее исследования: определение меры вариации, ее измерение, нахождение соответствующих измерителей, показателей, характеризующих ее размеры, выявление их сущности и методов вычисления факторов, ее определяющих.

По степени вариации можно судить о многих сторонах процесса развития изучаемых явлений, в частности об однородности совокупности, устойчивости индивидуальных значений признака, типичности средней, о взаимосвязи между признаками одного и того же явления и признаками разных явлений. Статистические показатели, характеризующие вариацию, широко применяются в практической деятельности, например для оценки ритмичности работы промышленных предприятий, контроля за ходом других производственных процессов, устойчивости урожайности сельскохозяйственных культур тех или иных сортов или одного и того же сорта в определенных почвенно-климатических условиях. На основе показателей вариации в статистике разрабатываются другие показатели и методы изучения явлений и процессов общественной жизни - показатели тесноты связи между явлениями и их признаками, показатели оценки точности выборочного наблюдения.

Каким же образом статистика дает количественную оценку степени колеблемости признака в совокупности, измеряет вариацию?

В чисто математической части решения этой проблемы общая теория статистики опирается на математическую статистику, в которой излагается математическая сторона таких показателей вариации, как размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации. Однако эти показатели находят широкое применение и в социально-экономической статистике. Поэтому рассмотрим сущностную и логическую стороны этих показателей.

7.2

МЕРЫ ВАРИАЦИИ

Показатели вариации делятся на две группы: абсолютные и относительные. К **абсолютным** относятся размах вариации, среднее линейное отклонение, дисперсия и среднее квадратическое отклонение. Вторая группа показателей вычисляется как отношение абсолютных показателей вариации к средней арифметической (или медиане). **Относительными** показателями вариации являются коэффициенты осцилляции, вариации, относительное линейное отклонение и др.

Самым простым абсолютным показателем является **размах вариации (R).**

Размах показывает, насколько велико различие между единицами совокупности, имеющими самое маленькое и самое большое значение признака. Например, различие между максимальной и минимальной пенсией различных групп населения, заработной платой различных категорий работающих или нормами выработки у рабочих определенной специальности или квалификации; размах вариации урожайности в хозяйствах фермеров района, области.

Его рассчитывают как разность между наибольшим ($X_{тах}$) и наименьшим (X_{min}) значениями варьирующего признака, т. е.

$$R = X_{max} - X_{min}. \quad (7.1)$$

Знание подобного рода величин необходимо в практической и хозяйственной деятельности, а также в научных исследованиях.

Например, размах вариации применяется при контроле качества продукции для определения влияния систематически действующих причин на производственный процесс. Для этого отбирают через определенные промежутки времени несколько деталей и производят их измерение. Рассчитав по данным этих выборок показатели размаха вариации, на основе сопоставления результатов вычислений судят об устойчивости режима производственного процесса.

В учебной литературе по статистике обычно указывается, что размах имеет существенный недостаток. Его величина всецело зависит от крайних значений признака, и он не учитывает всех изменений варьирующего признака в пределах совокупности.

Этот упрек в адрес размаха вариации является не совсем верным. Какой же это недостаток, когда именно в этом заключается суть показателя.

Размах вариации для того и существует, чтобы измерять расстояние между крайними точками. Другое дело, что в изучении вариации нельзя ограничиться определением одного лишь ее размаха. Но это не исключает необходимости определения величины этого показателя, не умаляет его значения.

К действительным недостаткам размаха вариации можно отнести то обстоятельство, что очень низкое и очень высокое значения признака по сравнению с основной массой его значений в совокупности могут быть обусловлены какими-либо сугубо случайными обстоятельствами (т. е. эти значения являются аномальными в совокупности). В этих случаях размах вариации дает искаженную амплитуду колебания признака против, так сказать, нормальных ее размеров, так как в данную совокупность включены единицы другой совокупности с аналогичным признаком. Поэтому следует очистить совокупность от аномальных наблюдений, прежде чем определить величину размаха вариации. Например, нельзя вычислять размах вариации заработков работников какого-либо частного предприятия, если наряду с заработками наемных работников включен «заработок» его владельца.

Итак, размах вариации - важный показатель колеблемости признака, но не исчерпывающий его характеристику.

Для анализа вариации необходим и показатель, который отражает все колебания варьирующего признака, дающий обобщенную ее характеристику. Для многих варьирующих признаков возможно допущение, что при прочих равных условиях все единицы совокупности в соответствии с основными законами своего развития имели бы одинаковую и притом вполне определенную величину признака в данных условиях места и времени. Вполне логично в качестве такой величины условно принять **среднюю величину** из всех значений признака, поскольку в ней более или менее погашаются случайные отклонения от закономерного хода развития явления, и средняя тем самым отражает типичный размер признака у данной Однородной совокупности единиц. Но условия существования и развития отдельных единиц совокупности в определенной степени различны, что сказывается и на различии значений у них взятого нами признака. Средняя величина отражает эти средние условия.

Следовательно, средняя применяется в качестве своего рода центра тяжести, вокруг которого происходит колебание, рассеяние значений признака. При обобщении этих колебаний необходимо вновь прибегнуть к методу средних величин - найти среднюю величину этих отклонений.

Такая средняя называется **средним линейным отклонением (d).** Оно вычисляется как средняя арифметическая из абсолютных значений отклонений вариант x_i и х (взвешенная или простая в зависимости от исходных условий) по следующим формулам:

$$\bar{d} = \frac{\Sigma|x_i - \bar{x}|}{n}$$

(простая),

(7.2)

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

(взвешенная).

(7.3)

Поскольку сумма отклонений значений признака от средней величины равна нулю, приходится все отклонения брать по модулю, на что указывают прямые скобки в числителе формул.

Покажем расчет среднего линейного отклонения на следующем примере (табл. 7.1).

Таблица 7.1

Распределение промышленных фирм одного **из** регионов России по вооруженности работников промышленно-производственными основными фондами

Группы фирм по величине ППОФ на одного работника, тыс. руб. х	Число фирм, % к итогу, fi	Середи на интерва лов, х'	x'f	|x'-x|	|x'-x|fj
А	1	2	3'	4	5
До 1,0	7,8	0,5	3,90	6,16	48,048
1,1-2,0	12,2	1,5	18,30	5,16	62,952

2,1-3,0	14,9	2,5	37,25	4,16	61,984
3,1-5,0	23,3	4,0	93,20	2,66	61,978
5,1 -10,0	24,3	7,5	182,25	0,84	20,412.
10,1 -20,0	10,6	15,0	159,00	8,34	88,404
20,1 и более	6,9	25,0	172,50	8,34	126,546
Итого	100,0	-	666,40	-	470,324

1. Прежде всего находим середины интервалов (х') по исходным данным (гр. А) и записываем их в таблицу (гр. 2).

2. Определим произведения значений середин интервалов (х') на соответствующие им веса (f_f) (гр. 3). В итоге получаем 666,4. Рассчитаем среднюю величину по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\Sigma x_i \cdot f_i}{\Sigma f_i} = \frac{666{,}4}{100} = 6{,}664 \text{ тыс. руб.}$$

3. Для расчета среднего линейного отклонения находим абсолютные отклонения середины интервалов, принятых нами в качестве вариантов признака (х') от средней величины (х) (гр. 4).

4. Наконец, вычисляем произведения отклонений х' - х на их веса (fj) и подсчитываем сумму этих произведений. Эта сумма равна 470,324. Результаты записываем в гр. 5.

5. Делим эту сумму на сумму весов, чтобы получить искомую величину d:

$$\bar{d} = \frac{470{,}324}{100} = 4{,}70324 \text{ тыс. руб.}$$

Таково в среднем отклонение вариантов признака от их средней величины. Это отклонение по сравнению со средней величиной признака очень большое. Оно отличается от средней на 1,961 тыс. руб. Это свидетельствует о том, что данная совокупность в отношении нашего признака неоднородна, а средняя - нетипична. Действительно, средняя величина выведена из величин, резко различающихся (например, максимальное значение признака в 50 раз больше минимального - 25,0 против 0,5).

Таким образом, среднее линейное отклонение дает обобщенную характеристику степени колеблемости признака в совокупности. Однако при его исчислении приходится допускать некорректные с точки зрения математики действия, нарушать законы алгебры, что побудило математиков и статистиков искать иной способ оценки вариации для того, чтобы иметь дело только с положительными величинами. Самый простой выход - возвести все отклонения во вторую степень. Это столь простое решение привело в последующем к большим научным результатам. Оказалось, что обобщающие показатели вариации, найденные с использованием вторых степеней отклонений, обладают замечательными свойствами. Поэтому они получили широкое распространение в различных областях знаний, на их основе были разработаны новые методы исследования, а также новые показатели количественной характеристики большого класса явлений

Полученная мера вариации называется **дисперсией** (а^2), $_а$ корень квадратный из дисперсии - **средним квадратическим отклонением** (ст)*. Эти показатели являются общепринятыми мерами вариации и часто используются в статистических исследованиях, а также в технике, биологии и других отраслях знаний. Данные показатели нашли также свое широкое применение в международной практике учета и статистического анализа, в частности в системе национального счетоводства.

Дисперсия представляет собой средний квадрат отклонений индивидуальных значений признака от их средней величины и вычисляется по формулам простой и взвешенной дисперсий (в зависимости от исходных данных):

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$$ (простая дисперсия), (7.4)

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2 f_i}{\Sigma f_i}$$ (взвешенная дисперсия). (7.5)

Дисперсия есть средняя величина квадратов отклонений. В данном случае варианты признака выражены в первой степени, значит, и мера их вариации также должна быть взята в первой степени.

Для этого достаточно извлечь из дисперсии корень второй степени, получится среднее квадратическое отклонение (о~). Значит, среднее квадратическое отклонение равно корню квадратному из дисперсии:

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2 f_i}{\Sigma f_i}}.$$

или

(7.6)

(7.7)

Среднее квадратическое отклонение - это обобщающая характеристика размеров вариации признака в совокупности. Оно выражается в тех же единицах измерения, что и признак (в метрах, тоннах, рублях, процентах и т. д.).

* В зарубежной литературе этот показатель называется стандартным отклонением и применяется в различных стандартах.

Рассмотрим расчет дисперсии и среднего квадратического отклонения по данным табл. 7.2 о валовом сборе зерновых культур в фермерских хозяйствах.

Таблица 7.2 **Вычисление о^{г} и с по несгруппированным данным**

Хозяйство	Валовой сбор зерновых, ц X	л: х.	(x$_f$-x)^2
А	1	2	3
1	600	100	10000
2	520	20	400
3	400	- 100	10000
4	600	100	10000
5	500	0	0
6	380	- 120	14400
Итого	3000	0	44800

1. Определяем среднюю величину по исходным данным (гр. 1) по формуле средней арифметической простой:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{3000}{6} = 500 \text{ ц.}$$

2. Находим отклонения х(от х и записываем их в гр. 2. Возводим отклонения во вторую степень, отводим для них гр. 3. Определяем их сумму. Она равна 44 800.

3. Разделив ее на число единиц совокупности, получаем дисперсию:

$$\sigma^2 = \frac{44\,800}{6} = 7466{,}67.$$

4. Извлекая из дисперсии корень второй степени д/7466,67 = 86,4099 ц, получаем среднее квадратическое отклонение.

Степень вариации в данной совокупности невелика, так как средняя величина равна 500 ц. Это говорит об однородности рассматриваемой нами совокупности.

Рассмотрим вычисление дисперсии и среднего квадратического отклонения по данным о распределении сотрудников двух министерств по тарифному разряду (табл. 7.3).

Таблица 7.3

Расчет $а^2$ и а в вариационном ряду с разным распределением частот

Тариф разряд x_i	Число со трудник по- век f_i	x_i-x	$(x_j-x)^2$	$(Xi-x)^2f_i$	Тариф, разряд x_i	Число сотрудников, человек f_i	$x_i - x$	$(x_i-x)^2$	$(Xi-x)^2f_i$
12	1	-3	9	9	12	30	_^	9	270
13	5	_ *2*	4	20	13	20	-2	4	80
14	30	—1	1	30	14	10	_ j	1	10
15	60	0	0	0	15	50	0	0	0
16	30	1	1 '	30	16	- 10	1	1	10
17	5	*2*	4	20	17	20	2	4	80
18	1	*3*	9	9	18	30	3	9	270
^того	13	—	_	118	_	170	_	_	720

Министерство № 1:

$$\bar{x}_1 = 15;\ \sigma_1^2 = \frac{118}{132} = 0,89;\ \sigma_1 = \sqrt{0,89} = 0,94 \approx 1 \text{ разряд.}$$

Министерство № 2

$$\bar{x}_2 = 15;\ \sigma_2^2 = \frac{720}{170} = 4,24;\ \sigma_2 = \sqrt{4,24} = 2,05 \approx 2 \text{ разряд.}$$

Среднее квадратическое отклонение во втором случае более чем в два раза превышает среднее квадратическое отклонение в первом. Это свидетельствует о более высокой колеблемости тарифного разряда в министерстве № 2.

Расчет дисперсии может быть упрощен. В случае равных интервалов в вариационном ряду распределения используется способ отсчета от условного нуля (способ моментов). Для его понимания необходимо знать следующие **свойства дисперсии.**

Свойство 1. Дисперсия постоянной величины равна нулю.

Свойство 2. Уменьшение всех значений признака на одну и ту же величину А не меняет величины дисперсии:

$$\sigma^2_{(x-A)} = \sigma_x^2. \quad (7.8)$$

Значит, средний квадрат отклонений можно вычислить не по заданным значениям признака, а по отклонениям их от какого-то постоянного числа.

Свойство 3. Уменьшение всех значений признака в *k'* раз уменьшает дисперсию в k^2 раз, а среднее квадратическое отклонение - в *k* раз:

$$\sigma^2_{(x/k)} = \sigma_x^2 : k^2. \quad (7.9)$$

Значит, все значения признака можно разделить на какое-то постоянное число (скажем, на величину интервала ряда), исчислить среднее квадратическое отклонение, а затем умножить его на постоянное число:

$$\sigma_x = \sigma_{\frac{x}{k}} \cdot k. \quad (7.9\text{ a})$$

Свойство 4. Если исчислить средний квадрат отклонений от любой величины А, в той или иной степени отличающейся от средней арифметической (х), то он всегда будет больше, среднего квадрата отклонений, исчисленного от средней арифметической:

$$\sigma_A^2 > \sigma_X^2. \quad (7.10)$$

Средний квадрат отклонений при этом будет больше на вполне определенную величину - на квадрат разности средней и этой условно взятой величины, т. е. на $(x - A)^2$:

$$\sigma_A^2 = \sigma_x^2 + (\bar{x} - A)^2, \text{ или } \sigma_x^2 = \frac{\Sigma(x_i - A)^2 f_i}{\Sigma f_i} - (\bar{x} - A)^2. \quad (7.11)$$

Значит, дисперсия от средней всегда меньше дисперсий, исчисленных от любых других величин, т. е. она имеет **свойство минимальности.**

В случае когда А приравнивается к нулю и, следовательно, не вычисляются отклонения, формула принимает такой вид:

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2, \text{ или } \sigma_x^2 = \frac{\Sigma x_i^2 \cdot f_i}{\Sigma f_i} - \left(\frac{\Sigma x_i \cdot f_i}{\Sigma f_i}\right)^2$$

(7.12)

Значит, средний квадрат отклонений равен среднему квадрату значений признака минус квадрат среднего значения признака.

На приведенных выше математических свойствах дисперсии основываются способы, которые позволяют упростить ее вычисления; например, расчет дисперсии по способу моментов или способу отсчета от условного нуля применяется в вариационных рядах с равными интервалами. Расчет производится по формуле

(7.12а)

$$\sigma^2 = \frac{\Sigma\left(\frac{x_i - A}{k}\right)^2 f_i}{\Sigma f_i} \cdot k^2 - (\bar{x} - A)^2,$$

где k — ширина .интервала;

А- условный нуль, в качестве которого удобно использовать середину интервала, обладающего наибольшей частотой;

$$\frac{\Sigma\left(\frac{x_i - A}{k}\right)^2}{\Sigma f}$$ - так называемый момент второго порядка. (7.126)

Между средним линейным и средним квадратическим отклонениями существует следующее примерное соотношение: ст = 1,25d, если фактическое распределение близко к нормальному. Исчисление среднего квадратического отклонения для явно несимметричных распределений не^имеет смысла. По свойству мажорантности средних величин (гл. 6) среднее квадратическре отклонение (о) всегда больше среднего линейного отклонения (d).

Среднее квадратическое отклонение играет важную роль в анализе рядов распределения. В условиях нормального распределения существует следующая зависимость между величиной среднего квадратического отклонения и количеством наблюдений:

- в пределах $\bar{x} \pm 1\sigma$ располагается 0,683, или 68,3%, количества наблюдений;
- в пределах $\bar{x} \pm 2\sigma$ – 0,954, или 95,4%;
- в пределах $\bar{x} \pm 3\sigma$ – 0,997, или 99,7%, количества наблюдений,

В действительности на практике почти не встречаются отклонения, которые превышают ± За. Отклонение За может считаться максимально возможным. Это положение называют «**правилом трех**

сигм».

До сих пор говорилось о показателях вариации, выраженных в абсолютных величинах. Но для целей сравнения колеблемости различных признаков в одной и той же совокупности или же при сравнении колеблемости одного и того же признака в нескольких совокупностях представляют интерес показатели вариации, приведенные **в относительных величинах.** Базой для сравнения должна служить средняя арифметическая. Эти показатели вычисляются как отношение размаха вариации, среднего линейного отклонения или среднего квадратического отклонения к средней арифметической или медиане. Чаще всего они выражаются в процентах и определяют не только сравнительную оценку вариации, но и дают характеристику однородности совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 33% (для распределений, близких к нормальному). Различают следующие относительные показатели вариации (V):

Коэффициент осцилляции(V_R)**:**

$$V_R = \frac{R}{\bar{x}} \cdot 100\%. \qquad (7.13)$$

Линейный коэффициент вариации$(V_{\bar{d}})$

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\% \quad \text{или} \quad V_{\bar{d}} = \frac{\bar{d}}{M_e} \cdot 100\%. \qquad (7.14)$$

Коэффициент вариации(V_σ)**:**

$$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%. \qquad (7.15)$$

Наиболее часто в практических расчетах применяется показатель относительной вариации - коэффициент вариации.

Для примера, приведенного в табл. 7.3, коэффициенты вариации получились следующие:

$$V_1 = \frac{0{,}94}{15} \cdot 100 = 6{,}27\%;$$

$$V_2 = \frac{2{,}05}{15} \cdot 100 = 13{,}67\%.$$

Основываясь на коэффициенте вариации, можно сделать вывод, что по тарифному разряду рабочих совокупности министерства № 1 и министерства № 2 являются однородными. Однако вариация колеблемости тарифного разряда в министерстве № 2 несколько выше, чем вариация в министерстве № 1.

7.3

ВАРИАЦИЯ АЛЬТЕРНАТИВНОГО ПРИЗНАКА. ЭНТРОПИЯ РАСПРЕДЕЛЕНИЯ

Среди множества варьирующих признаков, изучаемых статистикой, существуют признаки, которыми обладают одни единицы совокупности и не обладают другие. Эти признаки называются альтернативными. Примером таких признаков являются: наличие бракованной продукции, ученая степень у преподавателя вуза, работа по полученной специальности и т. д. Вариация альтернативного признака количественно проявляется в значении нуля у единиц, которые этим признаком не обладают, или единицы у тех, которые данный признак имеют.

Пусть p - доля единиц в совокупности, обладающих данным признаком (p = m/n); q - доля единиц, не обладающих данным признаком, причем p + q = 1. Альтернативный признак принимает всего два значения —. О и 1 с весами соответственно q и p. Исчислим среднее значение альтернативного признака по формуле средней арифметической:

$$\bar{x} = \frac{1\cdot p + 0\cdot q}{p+q} = p. \qquad (7.16)$$

Дисперсия альтернативного признака определяется по формуле

$$\sigma_p^2 = \frac{(1-p)^2\cdot p + (0-p)^2\cdot q}{p+q} = \frac{q^2\cdot p + p^2\cdot q}{p+q} = pq. \qquad (7.17)$$

Таким образом, дисперсия альтернативного признака равна произведению доли на дополняющее эту долю до единицы число. Корень квадратный из этого показателя, т. е. $\sqrt{pq}$, соответствует среднему квадратическому отклонению альтернативного признака. Предельное значение дисперсии альтернативного признака равно 0,25 при p = 0,5.

Показатели вариации альтернативных признаков широко используются в статистике, в частности при проектировании выборочного наблюдения, обработке данных социологических обследований, статистическом контроле качества продукции, в ряде других случаев.

Исчислим дисперсию альтернативного признака по следующим данным: налоговой инспекцией одного из районов города проверено 86 коммерческих киосков, и в 37 обнаружены финансовые нарушения. Тогда

$$n = 86;\ m = 37;\ p = \frac{37}{86} = 0{,}43;\ q = 1 - 0{,}43 = 0{,}57.$$

Следовательно, дисперсия и среднее квадратическое отклонение доли коммерческих киосков, имеющих финансовые нарушения, во всей совокупности обследованных киосков равны:

$$\sigma^2_p = 0{,}43 \cdot 0{,}57 = 0{,}245;$$

$$\sigma_p = \sqrt{0{,}245} = 0{,}495.$$

Обобщенной характеристикой различий внутри ряда может служить **энтропия распределения.** Применительно к статистике энтропия - это мера неопределенности данных наблюдения, которая может иметь различные результаты. Энтропия зависит от числа градаций признака и от вероятности каждой из них. Она показывает, имеет ли закономерность в сосредоточении отдельных градаций наименьшее количество позиций или, напротив, заполненность распределения одинаковая. При этом сумма вероятностей всех возможных исходов равна единице. Энтропия измеряется в **битах.**

Показатель энтропии (H_x) представляет собой отрицательную сумму произведения вероятностей различных значений случайной величины (PJ) на логарифмы (при основании два) этих вероятностей, т. е.

$$H(x) = -\sum \log p_i. \qquad (7.18)$$

Если все варианты равновероятны, то энтропия максимальна. Если же все варианты, за исключением одного, равны нулю, то энтропия равна нулю.

Энтропия альтернативного признака (п = 2) при равновероятном распределении (p = 0,5) равна единице:

$$H(x) = -(0,5 \log 0,5 + 0,5 \log 0,5) = 1.$$

(7.19)

Расчет энтропии распределения можно показать на примере выпуска продукции различных сортов на одном из предприятий точного машиностроения (табл. 7.4).

Таблица 7.4 Вероятность различных сортов продукции

Сорт	1-й	2-й	3-й	Брак	Итого
Вероятность, p_i	0,90	0,04	0,05	0,01	1,00

Энтропия данного распределения равна:

$$H(x) = -(0,9 \log 0,9 + 0,04 \log 0,04 + 0,05 \log 0,05 + 0,01 \log 0,01) = 0,6051 \text{ бита.}$$

Энтропия сложной системы вычисляется следующим образом:

$$H(xy) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \cdot \log_2 p_{ij},$$

(7.20) где PJJ - вероятность любого возможного состояния сложной системы.

Показатель энтропии позволяет также измерять количество информации. Чем больше информации о случайном событии, тем определеннее его состояние. Чем больше вероятность случайного события P_a, тем меньше информации несет его осуществление. В случае $P_a = 1$

$$H(x) = -(1 \log 1 + 0 + \ldots + 0) = 0. \quad (7.21)$$

Следовательно, данное испытание не содержит никакой информации. Аналогично и при $P_a = 0$.

Энтропия распределения интерпретируется как мера рассредоточенное™ вариантов случайной переменной по ее возможным значениям или как мера неопределенности значения реализации. Неопределенность значений реализации случайной переменной предусматривает наличие некоторого наблюдателя, находящегося в том или ином отношении к источнику неопределенности. Очевидно, можно представить ситуацию, когда для двух наблюдений степени неопределенности результатов одного и того же наблюдения со случайными исходами существенно различаются. Например, результаты голосования при экспертных опросах для наблюдателя - участника голосования и наблюдателя, не участвующего в голосовании.

В связи с тем что верхнего предела энтропия распределения не имеет, целесообразно вычислять наряду с абсолютной и относительную величину неопределенности.

Относительная энтропия определяется как отношение ее фактической величины к максимальной, т. е.

$$H(x) : H\, max\, (x).$$

Это отношение изменяется от нуля до единицы и может быть интерпретировано. Чем меньше относительная энтропия, тем меньше неопределенность и выше однородность.

7.4

ВИДЫ ДИСПЕРСИЙ И ПРАВИЛО ИХ СЛОЖЕНИЯ

Наряду с изучением вариации признака по всей совокупности в целом часто бывает необходимо проследить количественные изменения признака по группам, на которые разделяется совокупность, а также и между группами. Такое изучение вариации достигается посредством вычисления и анализа различных видов дисперсии.

Выделяют **дисперсию общую, межгрупповую и внутригрупповую. Общая дисперсия** а^2 измеряет вариацию признака во всей совокупности под влиянием всех факторов, обусловивших эту вариацию:

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2 \cdot f_i}{\Sigma f_i}.$$

(7.22)

Межгрупповая дисперсия (5$_X^2$) характеризует систематическую вариацию, т. е. различия в величине изучаемого признака, возникающие под влиянием признака-фактора, положенного в основание группировки. Она рассчитывается по формуле

$$\delta_x^2 = \frac{\Sigma(\bar{x}_i - \bar{x})^2 \cdot n_i}{\Sigma n_i},$$

(7.23)

где X; и гц - соответственно групповые средние и численности по отдельным группам.

Внутригрупповая дисперсия (о$_($^{2}) отражает случайную вариацию, т. е. часть вариации, происходящую под влиянием неучтенных факторов и не зависящую от признака-фактора, положенного в основание группировки. Она исчисляется следующим образом:

$$\sigma_i^2 = \frac{\Sigma(x_i - \bar{x})^2 \cdot n_i}{\Sigma n_i}.$$

(7.24)

Средняя из внутригрупповых дисперсий (<т$_($):

$$\overline{\sigma_i^2} = \frac{\Sigma\sigma_i^2 \cdot n_i}{\Sigma n_i}$$ (7.25)

Существует закон, связывающий три вида дисперсии. Общая дисперсия равна сумме средней из внутригрупповых и межгрупповой дисперсий:

$$\sigma^2 = \overline{\sigma_i^2} + \delta_X^2.$$

(7.26)

Данное соотношение называют **правилом сложения дисперсий.** Согласно этому правилу, общая дисперсия, возникающая под действием всех факторов, равна сумме дисперсии, появляющейся под влиянием всех прочих факторов, и дисперсии, возникающей за счет группировочного признака.

Зная любые два вида дисперсий, можно определить или проверить правильность расчета третьего вида.

Правило сложения дисперсий широко применяется при исчислении показателей тесноты связи, в дисперсионном анализе, при оценке точности типической выборки и в ряде других случаев.

В статистическом анализе широко используется показатель, представляющий собой долю межгрупповой дисперсии в общей дисперсии. Он носит название **эмпирического коэффициента детерминации** *(г\\$^2$):*

$$\eta^2 = \frac{\delta_x^2}{\sigma^2}.$$ (7.27)

Этот коэффициент показывает долю (удельный вес) общей вариации изучаемого признака, обусловленную вариацией группировочного признака.

Корень квадратный из эмпирического коэффициента детерминации носит название **эмпирического корреляционного отношения** (г|):

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma^2}}.$$ (7.28)

Оно характеризует влияние признака, положенного в основание группировки, на вариацию результативного признака. Эмпирическое корреляционное отношение изменяется в пределах от О до 1. Если Г] = 0, то группировочный признак не оказывает влияния на результативный. Если Т) = 1, то результативный признак изменяется только в зависимости от признака, положенного в основание группировки, а влияние прочих факторных признаков равно нулю. Промежуточные значения оцениваются в зависимости от их близости к предельным значениям.

Покажем его практическое использование на следующем примере (табл. 7.5).

Таблица 7.5

Производительность труда двух групп рабочих одного из цехов НПО «Циклон»

Производительность труда рабочих								
прошедших техническое обучение (деталей за смену)					не прошедших техническое обучение (деталей за смену)			
84	93	95	- 101	102	62	68	[illegible]	8 !05

Рассчитаем общую и групповые средние и дисперсии:

$$\bar{x}_1 = \frac{\Sigma x_i}{n_i} = \frac{475}{5} = 95 \text{ дет.}; \quad \bar{x}_2 = \frac{\Sigma x_i}{n_2} = \frac{405}{5} = 81 \text{ дет.};$$

$$\bar{x} = \frac{475+405}{10} = 88 \text{ дет.};$$

$$\sigma_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1} = \frac{(84-95)^2+(93-95)^2+\ldots+(102-95)^2}{5} = 42;$$

$$\sigma_2^2 = \frac{\Sigma(x-\bar{x})^2}{n_2} = \frac{(62-81)^2+(68-81)^2+\ldots+(105-81)^2}{5} = 231,2;$$

$$\sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n} = \frac{(84-88)^2+(93-88)^2+\ldots+(105-88)^2}{10} = 185,6.$$

Исходные данные вычисления средней из внутригрупповых и межгрупповой дисперсии представлены в табл. 7.6.

(Таблица 7.6

Расчет а?и 5 по двум группам рабочих

Группы рабочих	Численность рабочих, человек	Средняя выработка деталей за смену одним рабочим	Дисперсия
Прошедшие техническое обучение	5.	95,	42,0
Не прошедшие техническое обучение	5	81	231,2
Все рабочие	10	88	185,6

Рассчитаем следующие показатели. Средняя из внутригрупповых дисперсий:

$$\sigma_i^2 = \frac{42,0\cdot 5+231,2\cdot 5}{10} = 136,6.$$

Межгрупповая дисперсия:

$$\delta^2 = \frac{(95-88)^2\cdot 5+(81-88)^2\cdot 5}{10} = 49{,}0.$$

Общая дисперсия:

$$\sigma^2 = 136{,}6 + 49{,}0 = 185{,}6.$$

Таким образом, эмпирическое корреляционное отношение:

$$\eta = \sqrt{(49{,}0/185{,}6)} = 0{,}264.$$

Фактор технического обучения объясняет в данном примере 26,4% вариации производительности труда рабочих, а неучтенные факторы - 73,6%.

Выше было рассмотрено правило сложения дисперсии для количественных признаков. Но наряду с вариацией количественных признаков может наблюдаться и вариация качественных признаков. Такое изучение вариации достигается, как и для долей количественных признаков, посредством вычисления и анализа следующих видов дисперсий.

Внутригрупповая дисперсия доли определяется по формуле

(7.28а)

$$\sigma_{p_i}^2 = p_i\cdot(1-p_i).$$

Средняя из рвнутригрупповых дисперсий рассчитывается так:

$$\overline{\sigma^2_{p_i}} = \overline{p_i\cdot(1-p_i)} = \frac{\Sigma p_i\cdot(1-p_i)\cdot n_i}{\Sigma n_i}. \quad (7.29)$$

Формула **межгрупповой дисперсии** имеет следующий вид:

$$(7.30) \quad \delta^2_{p_i} = \frac{\Sigma(p_i-\bar{p})^2\cdot n_i}{\Sigma n_i},$$

где п, - численность единиц в отдельных группах;

р - доля изучаемого признака во всей совокупности, которая оппелеляется по rhonMvne

$$(7.31) \quad \bar{p} = \frac{\Sigma p_i n_i}{\Sigma n_i}.$$

Общая дисперсия определяется по формуле

$$\delta^2_{\bar{p}} = \bar{p}\cdot(1-\bar{p}). \quad (7.32)$$

Три вида дисперсии связаны между собой следующим образом:

$$\sigma^2_p = \overline{\sigma^2_{p_i}} + \delta^2_{p_i}. \quad (7.33)$$

Это соотношение дисперсий называется **теоремой сложения дисперсии доли признака.**

Пример. Имеются следующие данные об удельном весе основных рабочих в трех цехах фирмы (табл. 7.7).

Таблица 7.7 Удельный вес основных рабочих фирмы

Цех	Удельный вес основных рабочих, % p_i	Численность всех рабочих, человек, n_i
1	80	100
2	75	200

3	90	150
Итого	-	450

1. Определим долю основных рабочих в целом по фирме (формула 7.31):

$$\bar{p} = \frac{0{,}80 \cdot 100 + 0{,}75 \cdot 200 + 0{,}90 \cdot 150}{450} = \frac{365}{450} = 0{,}81.$$

2. Общая дисперсия доли основных рабочих по всей фирме в целом будет равна (формула 7.32):

$$\sigma_p^2 = 0{,}81 \cdot (1 - 0{,}81) = 0{,}154.$$

3. Внутригрупповые дисперсии рассчитаем, применив формулу 7.28а:

$$\sigma_{p_1}^2 = 0{,}8 \cdot 0{,}2 = 0{,}16; \qquad \sigma_{p_2}^2 = 0{,}75 \cdot 0{,}25 = 0{,}19;$$

$$\sigma_{p_3}^2 = 0{,}9 \cdot 0{,}1 = 0{,}09.$$

4. Средняя из внутригрупповых дисперсий будет равна (формула 7.29):

$$\overline{\sigma_{pi}^2} = \frac{0{,}16 \cdot 100 + 0{,}19 \cdot 200 + 0{,}09 \cdot 150}{450} = \frac{675}{450} = 0{,}15.$$

5. Межгрупповую дисперсию определим по формуле 7.30:

$$\delta_p^2 = \frac{(0{,}8 - 0{,}81)^2 \cdot 100 + (0{,}75 - 0{,}81)^2 \cdot 200 + (0{,}9 - 0{,}81)^2 \cdot 150}{450} = \frac{365}{450} = 0{,}004.$$

Проверка вычислений показывает: 0,154 = 0,15 + 0,004.

7.5

ПОНЯТИЕ О ЗАКОНОМЕРНОСТЯ^РАСПРЕДЕЛЕНИЯ

В приведенных в предыдущих разделах главы вариационных рядах распределения (см. табл. 7.1, 7.3) можно заметить определенную зависимость между изменением значений варьирующего признака и частот. Частоты в этих рядах с увеличением значения варьирующего признака первоначально увеличиваются, а затем после достижения какой-то максимальной величины в середине ряда уменьшаются. Это свидетельствует о том, что частоты в вариационных рядах изменяются закономерно в связи с изменением варьирующего признака. Такие закономерности изменения частот в вариационных рядах называются **закономерностями распределения.**

Одна из важных целей статистического изучения вариационных рядов состоит в том, чтобы выявить закономерность распределения и определить ее характер. Как и статистические закономерности вообще, закономерности распределения наиболее отчетливо проявляются только при массовом наблюдении. Поэтому основной путь в выявлении закономерностей распределения состоит в построении вариационных рядов для достаточно больших по численности статистических совокупностей. Кроме того, большое значение для нахождения закономерностей распределения имеет правильное построение самого вариационного ряда.

Речь идет прежде всего о таком определении оптимального числа групп и размера интервала, при котором закономерность распределения видна более отчетливо.

Если сразу трудно определить оптимальное число групп, то первоначально разбивают совокупность на максимальное их число, а затем, укрупняя интервал и сокращая таким образом число групп, стремятся получить такое их число, при котором достаточно явно проявляется характер распределения.

Закономерности распределения выражают свойства явлений, общие условия, влияющие на формирование вариации признака. Так, например, распределение предприятий области по формам собственности отражает условия, определяющие развитие тех или иных форм на данном этапе. Эти условия выражаются, в частности, в степени разработанности законодательной базы, структуре экономики и ее состоянии в отдельных отраслях, социальном составе населения, уровне квалификации занятых и т. д. В меньшей степени для данного процесса характерно действие случайных причин, хотя полностью абстрагироваться от их проявления невозможно.

Когда мы говорим о характере, типе закономерностей распределения, то имеем в виду отражение в них общих условий, определяющих распределение. При этом следует учитывать, что речь идет о распределениях, отражающих однородные явления. Если мы смешаем вместе распределение предприятий по формам собственности в областях с различными социально-экономическими условиями, то сильно уменьшим действие общих факторов, определяющих его характер. В этом случае распределение может не проявиться отчетливо или проявиться как бимодальное (двухвершинное) или даже многовершинное.

7.6

ИЗУЧЕНИЕ ФОРМЫ РАСПРЕДЕЛЕНИЯ

Основная задача анализа вариационных рядов - выявление подлинной закономерности распределения путем исключения влияния второстепенных, случайных для данного распределения факторов — достигается увеличением объема исследуемой совокупности при одновременном уменьшении интервала ряда.

Из математической статистики известно, что если увеличить объем совокупности и уменьшить интервал группировки, изобразить эти данные графически, то полигон (гистограмма) распределения все более и более приближается к некоторой плавной линии, являющейся для него пределом и носящей название кривой распределения.

Таким образом, под **кривой распределения** понимается графическое изображение в виде непрерывной линии изменения частот в вариационном ряду, функционально связанного с изменением вариант.

Получение кривой распределения на основе полигона или гистограммы можно представить лишь для гипотетического случая, соответствующего бесконечно большому числу единиц совокупности и бесконечно малой ширине интервала ряда. Только при этих идеализированных условиях кривая распределения будет выражать функциональную связь между значениями варьирующего признака и соответствующими им частотами и представлять так называемое теоретическое- распределение.

Теоретической кривой распределения называется кривая, выражающая общую закономерность данного типа распределения в чистом виде, исключающего влияние случайных для закономерностей факторов. Но получение кривой распределения из эмпирических данных (полигон, гистограмма) возможно лишь для описанного идеального случая. Поэтому при проведении анализа вариационных рядов целесообразно свести эмпирическое распределение к одному из хорошо исследованных видов теоретического распределения, рассматриваемых математической статистикой. При этом теоретическое распределение играет роль некоторой идеализированной модели эмпирического распределения, а сам анализ вариационных рядов сводится к сопоставлению эмпирического и теоретического распределений и определению степени различия между ними.

В статистической практике встречаются разнообразные распределения. *Различают следующие разновидности кривых распределения:*

- одновершинные кривые: симметричные, умеренно асимметричные и крайне асимметричные;
- многовершинные кривые.

Для однородных совокупностей, как правило, характерны одновершинные распределения. Многовершинность' свидетельствует о неоднородности изучаемой совокупности. Появление двух и более вершин делает необходимой перегруппировку данных с целью выделения более однородных групп.

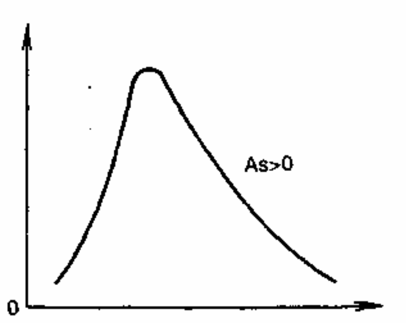
Выяснение общего характера распределения предполагает оценку его однородности, а также вычисление показателей асимУетрии и эксцесса. Для симметричных распределений частоты любых двух вариант, равноотстоящих в обе стороны от центра распределения, равны между собой. Рассчитанные для таких распределений средняя, мода и медиана также равны.

При сравнительном изучении асимметрии нескольких распределений с разными единицами измерения вычисляется относительный показатель асимметрии (A_s):

$$A_S = \frac{\bar{x} - Mo}{\sigma}, \quad \text{или} \quad A_S = \frac{\bar{x} - Me}{\sigma}.$$

(7.34)

Его величина может быть положительной и отрицательной. В первом случае речь идет о правосторонней асимметрии, а во втором - о левосторонней (рис. 7.1 и 7.2).



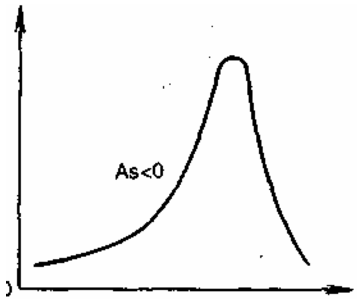


Рис. 7.1. Правосторонняя асимметрия

Рис. 7.2. Левосторонняя асимметрия

При правосторонней асимметрии Mo > Me > x. Наиболее широко (как показатель асимметрии) применяется отношение центрального момента третьего порядка* к среднему квадратическому отклонению данного ряда в кубе, т. е.

$$A_S = \frac{\mu_3}{\sigma^3}.$$

(7.35)

Применение данного показателя дает возможность определить не только величину асимметрии, но и проверить ее наличие в генеральной совокупности. Принято считать, что асимметрия выше 0,5 (независимо от знака) считается значительной; если она меньше 0,25, то незначительной.

Оценка существенности A_s производится на основе средней квадратической ошибки, коэффициента асимметрии (o_{As}), которая зависит от числа наблюдений (п) и рассчитывается по формуле

$$\sigma_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}.$$

(7.36)

В случае $\frac{|A_S|}{\sigma_{A_s}} > 3$ асимметрия существенна и распределение признака в генеральной совокупности несимметрично, в противном случае асимметрия несущественна и ее наличие может быть вызвано случайными обстоятельствами.

Покажем расчет коэффициента асимметрии на условном примере работы биржи по продаже акций (табл. 7.8).

Таблица 7.8

Расчет коэффициента асимметрии

Продано на сумму, млн руб. x	Количество продаж f	Середина интервала X'i	,,' x_i x_0	x'f	(*?f	(xff
			k k = 0,5 x_0= 1,75			
0,5- 1,0	46	0,75	~2	-92	184	-368
1,0-1,5	123	1,25	- 1	-123	123	- 123

* Центральными называются моменты распределения, при вычислении которых за исходную величину принимаются отклонения вариантов от средней арифметической данного ряда.

1,5-2,0	525	1,75	0	0	0	0
2,0 - 2,5	228	2,25	1	228	228	228
2,5-3,0	35	2,75	2	70	140	280
3,0-3,5	28	3,25	3	84	252	756
3,5-4,0	12	3,75	4	48	192	768
4,0-4,5	3	4,25	5	15	75	375
Итого	100	-	. -	230	1194	1916

201

Определим условные моменты т,, т$_2$ и т$_3$*, а также центральные моменты |1$_2$ и н,$_3$> необходимые для вычисления коэффициента асимметрии:

$$m_1 = \frac{230}{1000} = 0{,}23; \quad m_2 = \frac{1194}{1000} = 1{,}194; \quad m_3 = \frac{1916}{1000} = 1{,}916;$$

$$\mu_2 = m_2 - (m_1)^2 = 1{,}194 - (0{,}23)^2 = 1{,}1411;$$

'ц$_3$ = т$_3$ - т, • (2ц$_2$ + т$_2$) = 1,916 - 0,23 • (2,282 + 1,194) = 1,116. Коэффициент асимметрии для данного

ряда равен:

$$A_S = \frac{\mu_3}{\sigma^3} = \frac{1{,}116}{(1{,}068)^3} = 0{,}92.$$

Полученный результат свидетельствует о наличии значительной по величине и положительной по своему характеру асимметрии.

Для симметричных распределений может быть рассчитан показатель эксцесса (E$_k$). Наиболее точно он определяется по формуле с использованием центрального момента четвертого

$$\left(\mu_4 = \frac{\Sigma(x_i - \overline{x})^4}{n}\right):$$

порядка

$$E_k = \frac{\mu_4}{\sigma^4} - 3.$$

(7.37)

На рис. 7,3 и 7.4 представлены два распределения: островершинное (E$_k$ положительный) и плосковершинное (E$_k$ отрицательный). В нормальном распределении E$_k$ = 0.

Среднеквадратическая ошибка эксцесса (CT$_E$) рассчитывается по формуле

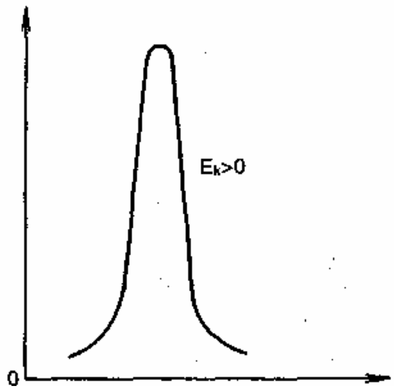
$$\sigma_{E_k} = \sqrt{\frac{24n \cdot (n-2) \cdot (n-3)}{(n-1)^2 \cdot (n+3) \cdot (n+5)}},$$

(7-38). Я

где п - число наблюдений.

---

* Условными называются моменты, вычисленные по отклонениям от произвольно взятой величины, т. е. от так называемого начала отсчета. На их использовании основан, в частности, один из способов расчета дисперсии (п. 7.2).



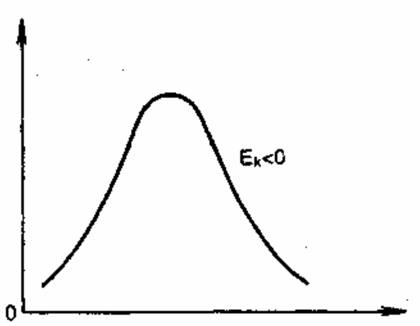


Рис. 7.3. Островершинное распределение
Рис. 7.4. Плосковершинное распределение

Для определения асимметрии и эксцесса можно пользоваться упрощенными формулами, предложенными Линдбергом:

$$A_S = P - 50, \quad (7.39)$$

где Р - удельный вес (в %) количества тех вариант, которые превосходят среднюю арифметическую, в общем количестве вариант данного ряда;

50- удельный вес (в %) вариант, превосходящих среднюю арифметическую ряда нормального распределения.

$$E_k = P - 38{,}29,$$

(7.40)

Р - доля (в %) количества вариант, лежащих в интервале, равном половине среднего квадратического отклонения (в ту или другую сторону от величины средней в общем количестве вариант данного ряда); ^__

38,29 - доля (в %) количества вариант, лежащих в интервале, равном половине среднего квадратического отклонения (в ту или другую сторону от величины средней), в общем количестве вариант ряда нормального распределения.

Необходимо отметить, что хотя показатели асимметрии и эксцесса характеризуют непосредственно лишь форму распределения признака в пределах изучаемой совокупности, однако их определение имеет не только описательное значение. Часто асимметрия и эксцесс дают определенные указания для даль-
где
нейшего исследования социально-экономических явлений. Например, появление значительного отрицательного эксцесса может указывать на качественную неоднородность исследуемой совокупности. Кроме того, эти показатели позволяют сделать вывод о возможности применения данного эмпирического распределения к типу кривых нормального распределения.

7.7

ТЕОРЕТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ В АНАЛИЗЕ ВАРИАЦИОННЫХ РЯДОВ

Многие явления, рассматриваемые каждое в отдельности, изолированно друг от друга, кажутся случайными. Однако если анализировать эти явления в совокупности с другими, аналогичными по своей сущности, то часто удается обнаружить закономерность, связанную с их возникновением. Например, мы не можем предсказать уровень дохода человека, если не располагаем о нем некоторой дополнительной информацией (о возрасте, профессиональной принадлежности, месте работы и т. д.). В то же время при рассмотрении группы людей закономерности формирования доходов проявляются более отчетливо. Так, во многих странах большинство населения имеет относительно низкий уровень дохода, некоторые - более высокий и только у незначительной части уровень дохода очень высокий. Именно существование подобных статистических закономерностей делает необходимым изучение индивидуальных, нередко на первый взгляд беспорядочно колеблющихся данных.

Если на практике часто встречается один и тот же тип распределения частот, целесообразно описать его с помощью математической формулы, которая может служить для сравнения и обобщения различных совокупностей аналогичных данных.

В статистике широко используются различные виды теоретических распределений - нормальное распределение, биномиальное распределение, распределение Пуассона и др. Каждое из теоретических распределений имеет специфику и свою область применения в различных отраслях знания.

Чаще всего в качестве теоретического распределения используется **нормальное распределение:**

$$y_t = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}t^2},$$

(7.41)

где у, - ордината кривой нормального распределения;

$t = \frac{x - \bar{x}}{\sigma}$ - стандартизованное отклонение;

е и л - математические постоянные;

х - варианты вариационного ряда;

х - их средняя величина;

о - среднее квадратическое отклонение.

Нормальное распределение полностью определяется двумя параметрами - средней арифметической (х) и средним квадратическим отклонением а. Подчиненность закону нормального распределения проявляется тем точнее, чем больше случайных величин действуют вместе. Если ни одна из случайно действующих причин по своему действию не окажется преобладающей над другими, то закон распределения очень близко подходит к нормальному.

Такая закономерность проявляется, например, в распределении отклонений в производственном процессе при нормальном уровне организации и технологии, в распределении населения определенного возраста по размеру обуви и т. д.

Часто возникают распределения, хотя и не отвечающие строго нормальному распределению, но имеющие с ним сходство. Такие сходные черты часто обусловлены тем, что крайние значения вариантов, близкие к X_{min} и $X_{тах}$, встречаются много реже, чем серединные.

Рассмотрим некоторые **свойства кривой нормального распределения:**

- f(t) - функция нормального распределения - четная, т. е. f(-t) = f(+t). Следовательно, изображающая ее кривая распределена симметрично относительно оси ординат, т. е. х = Мо = Ме;
- функция имеет бесконечно малые значения при t = ± °°. Это означает, что ветви кривой удалены в бесконечность и асимптотически приближаются к оси абсцисс. При этом чем'больше значения признака отклоняются от х, тем реже встречаются;
- функция имеет максимум при t = 0. Отсюда следует, что модального значения кривая достигает при t = 0 или при х = х.

Величина максимума составляет 1/-/2гс ;

- при t = ± 1 функция дает точки перегиба. Следовательно, при отклонении значений признака (х) от среднего значения (х) в положительном и отрицательном направлениях на одно стандартное (нормированное) отклонение (±ст от х) кривая дает переход от выпуклости к вогнутости;
- если случайная величина представляет сумму двух независимых случайных величин, следующих каждая нормальному закону, то она тоже следует нормальному закону;
- площадь между кривой и осью ot равна единице, как интеграл Пуассона.

Рассмотрим расчет значений частот теоретического ряда распределения на основании данных о содержании меди (в %) в 500 образцах сплава.

Поскольку нормальное распределение зависит от двух параметров: х и G, прежде всего определим соответствующие характеристики приведенного в табл. 7.9 распределения образцов содержания меди.

Таблица 7.9 **Расчет теоретических частот нормального распределения**

Содер-жание меди, %	Количе-ство образ-	Сере-дина интер-	x'-x	x'-x	f(t) ”—-|г-1- си || *	Теоретические частоты f =kSf f(t)
				О		''' О 0

x	f	x'			И	исчис-ленные	округ-ленные
1	2	3	4	5	6	7	8
56-58	5	57	-7,66	2,47	0,01888	6,09	6
58-60	29	59	-6,66	1,83	0,07477	24,12	24
60-62	63	61	-3,66	1,18	0,19886	64,15	64
62-64	116	63	- 1,66	0,54	0,34482	111,83	112
64-66	117	65	0,34	0,11	0,39654	127,92	128
66-68	102	67	2,34	0,75	0,30114	97,14	97
68-70	48	69	4,34	1,40	0,14973	48,30	48
70-72	14	71	6,34	2,04	0,04980	16,06	16
72-74	6	73 .	8,34	2,69	0,01071	3,55	4
Итого	500	-	-	-	-	499,06	499

x = 64,66%, a = 3,1% (расчет x и C произведен обычным способом и в табл. 7.9 не приводится). В графах 1 и 2 табл. 7.9 приведены фактические варианты и частоты.

Для расчета частот нормального распределения 500 образцов содержания меди со средней x = 64,66% и средним квадратическим отклонением a = 3,1% необходимо использовать формулу плотности вепоятности:

$$P_x = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{(x-\bar{x})^2}{2\sigma^2}}.$$

(7.42)

Чтобы прийти к частотам нормального распределения f_m, необходимо выразить их через P_x.

Для удобства вычислений вероятностей случайные величины нормируются, а затем используются заранее табулированные значения плотности функции распределения нормированной случайной величины. Первый множитель такой функции - величина постоянная для данного распределения. В нашем случае:

kZf 2-500 ^f ' x-x ,-
---- = ------- = 322,6, во втором множителе выражение ------- обо-
a 3,1 o
значим через t, тогда получим:

$$\frac{1}{\sqrt{2\pi}} \cdot e^{\frac{(x-\bar{x})^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}. \qquad (7.43)$$

Полученную функцию от t обозначим f(t):

$$f(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}.$$

(7.44)

В математической статистике существуют специальные таблицы для любых значений f(t).

Таким образом, $f_m = \frac{k\Sigma f}{\sigma} \cdot f(t)$ очень легко рассчитать, определив для каждого значения варианта x величину $t = \frac{x-\bar{x}}{\sigma}$ (гр.5) и найдя по таблицам соответствующие f(t) (гр. 6). Умножая f(t) на постоянный для всех частот множитель $\frac{k\Sigma f}{\sigma}$ получаем теоретические частоты нормального_распределения $т_т$ (гр.7).

Сравнивая полученные *f_m* (гр. 7) с фактическими частотами f (гр. 2), убеждаемся, что их расхождения невелики. На графике, представленном на рис. 7.5, видна довольно большая близость фактических частот распределения к теоретическим.

В то же время нельзя не отметить, что сопоставление графика эмпирических частот с теоретическими в целях определения соответствия эмпирического распределения нормально-

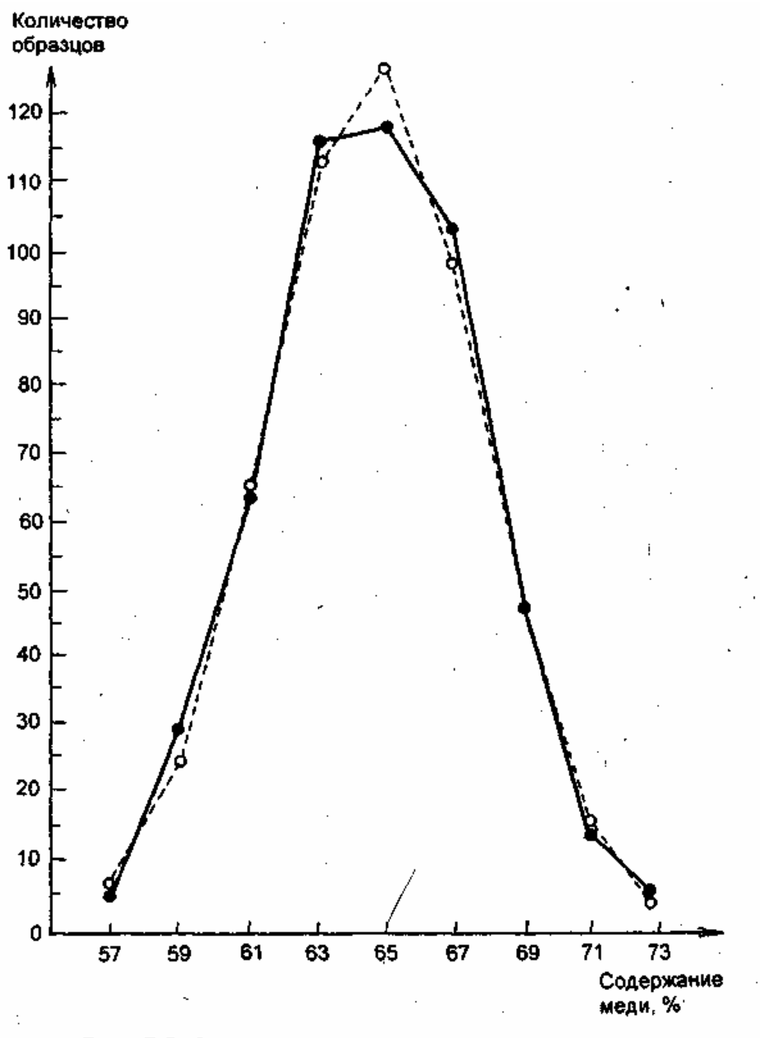


Рис. 7.5. Эмпирические и теоретические данные

му позволяет оценивать эти расхождения только субъективно. Объективная характеристика соответствия может быть получена с помощью особых статистических показателей - критериев согласия. Известны критерии согласия К. Пирсона (хиквадрат), В. И. Романовского, А. Н. Колмогорова и Б. С. Ястремского.

Критерий согласия Пирсона ($\%^2$) вычисляется по формуле

$$\chi^2 = \Sigma \frac{(f_э - f_т)^2}{f_т},$$

(7.45)

где f, и fj. - эмпирические и теоретические частоты соответственно.

С помощью величины $\%^2$ по специальным таблицам определяется вероятность $P(x^2)$- Входами в таблицу являются значения $\%^2$ и число степеней свободы у = п — 1. На основе Р выносится суждение о существенности или йесущественности расхождения между эмпирическим и теоретическим распределениями. При Р > 0,5 считается, что эмпирическое и теоретическое распределения близки, при Р е [0,2; 0,5] совпадение между ними удовлетворительное, в остальных случаях недостаточное.

Если число степеней свободы большое, то применяется соотношение, равное $д/2/^2$ — у2у —1. Расхождение между эмпирическим и теоретическим распределениями существенно при значениях этой разности, заметно превосходящих 2.

Критерий Романовского (С), также используемый для проверки близости эмпирического и теоретического распределений, определяется следующим образом:

$$C = \frac{\chi^2 - \gamma}{\sqrt{2\gamma}},$$

(7.46)

где %^{2} - критерий Пирсона, рассчитываемый по формуле (7.45);
у - число степеней свободы (при проверке гипотезы о нормальности распределения равно числу групп минус три).

При С < 3 различие несущественно, что позволяет считать эмпирическое распределение близким к нормальному.

$$L = \frac{\Sigma\frac{(f_э - f_т)^2}{Npq} - K}{\sqrt{2K + 4Q}},$$

Критерий Ястремского (L) может быть найден на основе следующего соотношения:

(7.47)

где N — объем совокупности; ;
pq - дисперсия альтернативного признака; К - число вариантов или групп;
Q - принимает значение 0,6 при числе вариантов или групп от 8 до 20.

Если L>3, то эмпирическое распределение соответствует теоретическому.

Критерий Колмогорова (А.) вычисляется по формуле

$$\lambda = \frac{D}{\sqrt{\Sigma f}},$$

(7.48)

где D - максимальное значение разности между накопленными эмпирическими и теоретическими частотами; £f - сумма эмпирических частот.

Необходимым условием использования этого критерия является достаточно большое число наблюдений (не меньше ста).

Рассмотрим применение критерия Колмогорова (табл. 7.10).

Как видно из табл. 7.10, максимальное значение разности между эмпирическими и теоретическими частотами составляет 7, т. е. D = 7.

Следовательно, величина критерия Колмогорова в нашем случае равна: д

$$\lambda \approx \frac{7}{\sqrt{500}} \approx 0{,}31.$$

По специальным таблицам вероятностей *P(k)* определяем, что \^, = 0,31 соответствует Р(х), близкой к 1,000. Это означает, что с вероятностью, близкой к 1, можно утверждать, что отклонения фактических частот от теоретических в нашем примере являются случайными. Следовательно, в основе фактического распределения образцов содержания меди лежит закон нормального распределения.

Рассмотрение других видов распределения, кроме нормального, не является предметом изучения в данной главе.

Таблица 7.10

Расчет критерия Колмогорова по данным о процентном содержании меди в 500 образцах

Содержание меди,	Частоты ряда распределения	Накопленные частоты	

% X	эмпириче-ское $f,$	теоретиче-ское f_T	эмпириче-ское $2f,.$	теоретиче-ское Zf_T	$\lvert 5I\ f\text{-}Zf_T \rvert$
56-58	5	6	5	6	1
58-60	29	24	34	30	4
60-62	63	64	97	94	3
62-64	116	112	213	206	7
64-66	117	128	330	334	4
66-68	102	97	432	431	1
68-70	48	48	480	479	1
70-72	14	16	494	495	1
72-74	6	4	500	499	1
Итого	500	499	-	-	-

7.8

СТРУКТУРНЫЕ ХАРАКТЕРИСТИКИ ВАРИАЦИОННОГО РЯДА РАСПРЕДЕЛЕНИЯ

Наряду с рассмотренными средними величинами в качестве статистических характеристик вариационных рядов распределения рассчитываются так называемые структурные средние - мода и медиана.

Мода (Мо) представляет собой значение изучаемого признака, повторяющееся с наибольшей частотой. Медианой (Me) называется значение признака, приходящееся на середину ранжированной (упорядоченной) совокупности.

Главное свойство медианы заключается в том, что сумма абсолютных отклонений значений признака от медианы меньше, чем от любой другой величины:

$$E|x_i - Me| = min. \qquad (7.49)$$

Рассмотрим определение моды и медианы по несгруппированным данным.

Предположим, рабочие бригады, состоящей из 9 человек, имеют следующие тарифные разряды: 4 3 4 5 3 3 6 2 6 .

Так как в данной бригаде больше всего рабочих 3-го разряда, этот тарифный разряд и будет модальным.

Для определения медианы необходимо провести ранжирование:

2 3 3 3 4 4 5 6 6 .

Центральным в этом ряду является рабочий 4-го разряда, следовательно, данный разряд и будет медианным. Если ранжированный ряд включает четное число единиц, то медиана определяется как средняя из двух центральных значений.

Если мода отражает типичный, наиболее распространенный вариант значения признака, то медиана практически выполняет функции средней для неоднородной, не подчиняющейся нормальному закону распределения совокупности. Проиллюстрируем ее познавательное значение следующим примером.

Допустим, нам необходимо дать характеристику среднего дохода группы людей, насчитывающей 100 человек, из которых 99 имеют доходы в интервале от 100 до 200 долл. в месяц, а месячные доходы последнего составляют 50 000 долл. (табл. 7.11).

Таблица 7.11 **Месячные доходы исследуемой группы людей**

№ п/п	1	2	3	4		50	51		99	100
Доход, долл.	100	104	104	107		162	164		200	50000

Если мы воспользуемся средней арифметической, то получим средний доход, равный примерно 600 - 700 долл., который не только в несколько раз меньше дохода 100-го человека, но и имеет мало общего с доходами остальной части группы. Медиана же, равная в данном случае 163 долл., позволит дать объективную характеристику уровня доходов 99% данной группы людей.

Рассмотрим определение моды и медианы по сгруппированным данным (рядам распределения).

Предположим, распределение рабочих уже не отдельной бригады, а всего предприятия в целом по тарифному разряду имеет следующий вид (табл. 7.12).

Таблица 7.12 **Распределение рабочих предприятия по тарифному разряду**

Тарифный разряд	Численность рабочих, человек
2	12
3	48
4	56
5	60
6	14
Всего	190

Определение моды по дискретному вариационному ряду не составляет большого труда - наибольшую частоту (60 человек) имеет 5-й тарифный разряд, следовательно, он и является модальным.

Для определения медианного значения признака по следующей формуле находят номер медианной единицы ряда (N_{Me}):

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

(7.50)

где п - 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-м и 96-м рабочими. Необходимо определить, к какой группе относятся рабочие с этими порядковыми номерами. Это можно сделать, рассчитав накопленные частоты. Очевидно, что рабочих с этими номерами нет в первой группе, где всего лишь 12 человек, нет их и во второй группе (12 + 48 = 60). 95-й и 96-й рабочие находятся в третьей группе (12 + 48 + 56 = 116), следовательно, медианным является 4-й тарифный разряд.

В отличие от дискретных вариационных рядо& определение моды и медианы по интервальным рядам требует проведения определенных расчетов на основе следующих формул:

$$Mo = x_0 + i \cdot \frac{(f_{Mo} - f_{Mo-1})}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})},$$

(7.51)

где x„ - нижняя граница модального интервала (**модальным** называется интервал, имеющий наибольшую частоту); i - величина модального интервала; f_{Mo} - частота модального интервала; $f_{Mo\cdot\_}$ | — частота интервала, предшествующего модальному; f_{Mo+1} - частота интервала, следующего за модальным.

$$Me = x_0 + i \cdot \frac{\frac{1}{2}\Sigma f_i - S_{Me-1}}{f_{Me}},$$

(7.52)

где x_0 — нижняя граница медианного интервала (медианным называется первый интервал, накопленная частота которого превышает половину общей суммы частот); i - величина медианного интервала;

S_{Me-1} - накопленная частота интервала, предшествующего медианному;

f_{Me} - частота медианного интервала.

Проиллюстрируем применение этих формул, используя данные табл. 7.13.

Таблица 7.13

Распределение населения РФ по уровню среднедушевых номинальных денежных доходов в I полугодии 1997 г.

Группы по уровню среднедушевого месячного дохода, тыс. руб.	Численность населения, млн чел.
До 400	29,6
400-600	30,6
600 - 800	25.1
800- 1000	18,4
1000-1200	12,8
1200-1400	9,4
1400-1600	5,6
1600-1800	4,1
1800-2000	3,3
Свыше 2000	8,6
Итого	147,5

Интервал с границами 400 - 600 в данном распределении будет модальным, так как он имеет наибольшую частоту. Используя формулу (7.51), определим моду:

$$Mo = 400+200\cdot\frac{30{,}6-29{,}6}{(30{,}6-29{,}6)+(30{,}6-25{,}1)} = 430{,}8 \text{ тыс. руб.}$$

Для установления медианного интервала необходимо определять накопленную частоту каждого последующего интервала до тех пор, пока она не превысит половины суммы накопленных частот (в нашем случае 73,75 млн чел.) (табл. 7.14).

Таблица 7.14

Определение медианного интервала

Интервал, тыс. руб.	Накопленная частота, млн чел.
До 400	29,6
400 - 600	60,2
600 - 800	85,3

Мы установили, что медианным является интервал с границами 600 - 800 тыс. руб. Определим теперь медиану:

$$Me = 600+200\cdot\frac{73{,}75-60{,}2}{25{,}1} = 708{,}0 \text{ тыс. руб.}$$

Таким образом, в качестве обобщенной характеристики значений определенного признака у единиц ранжированной совокупности могут быть использованы средняя арифметическая) мода и медиана. Каждая из них имеет свои особенности. \

Основной характеристикой центра распределения является средняя арифметическая, для которой характерно то, что вс$ отклонения от нее (положительные и отрицательные) в сумме равняются нулю; для медианы характерно, что сумма отклонений от нее по модулю является минимальной, а мода представляет собой значение признака, которое наиболее часто встреча-^ ется. Поэтому в зависимости от цели исследования распределения должна выбираться одна из упомянутых характеристик, либо же для сравнения - все три.

Соотношение моды, медианы и средней арифметической указывает на характер распределения признака в совокупности, позволяет оценить его асимметрию.

В симметричных распределениях все три характеристики со впадают. Чем больше расхождение между модой и средней арифметической, тем больше асимметричен ряд. Для умеренно асимметричных рядов разность между модой и средней примерно в три раза превышает разность между медианой и средней, т. е

$$|Mo - \bar{x}| = 3\ |Me - \bar{x}|. \qquad (7.53)$$

Моду и медиану в интервальном ряду можно определить графически. Мода определяется по гистограмме распределения. Для этого выбирается самый высокий прямоугольник, который является в данном случае модальным. Затем правую вершину модального прямоугольника соединяем с правым верхним углом предыдущего прямоугольника. А левую вершину модального прямоугольника - с левым верхним углом последующего прямоугольника. Далее из точки их пересечения опускают перпендикуляр на ось абсцисс.

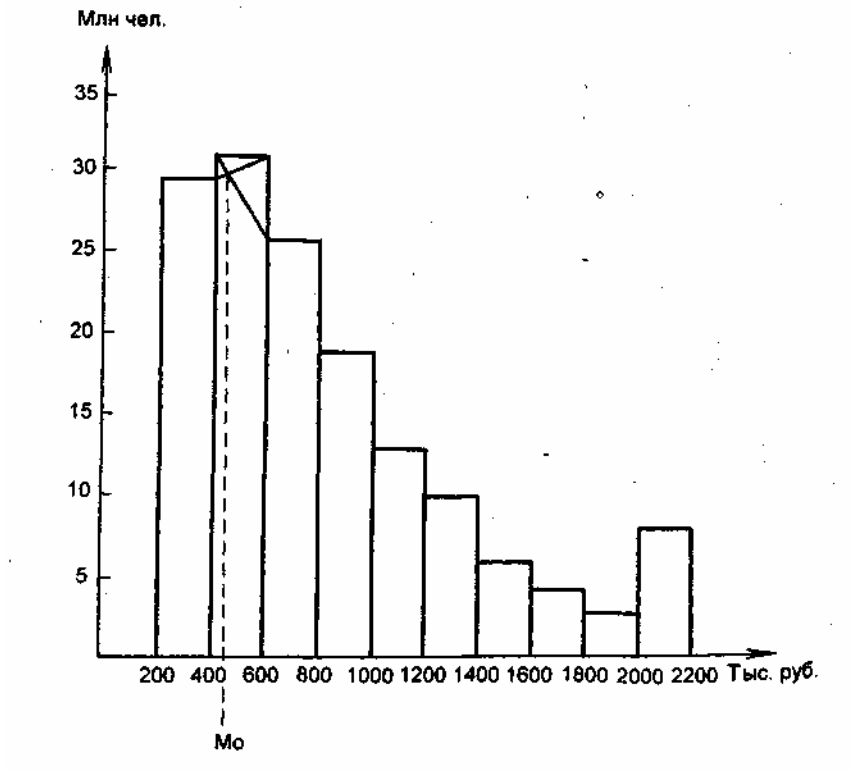


Рис. 7.6. Графическое определение моды по гистограмме

Абсцисса точки пересечения этих прямых и будет модой распределения (рис. 7.6). Медиана рассчитывается по кумуляте (рис. 7.7). Для ее определения из точки на шкале накопленных частот (частостей), соответствующей 50%, проводится прямая, параллельная оси абсцисс, до пересечения с кумулятой. Затем из точки пересечения указанной -прямой с кумулятой опускается перпендикуляр на ось абсцисс. Абсцисса точки пересечения является медианой.

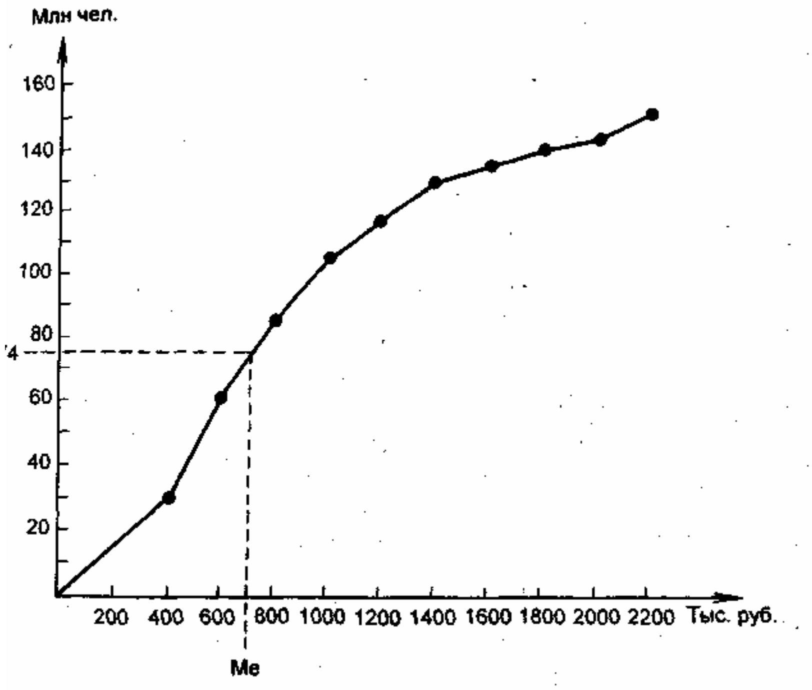


Рис. 7.7. Графическое определение медианы по кумуляте

Аналогично с нахождением медианы в вариационных рядах распределения можно отыскать значение признака у любой по порядку единицы ранжированного ряда. Так, например, можно найти значение признака у единиц, делящих ряд на чытыре равные части, на десять или сто частей. Эти величины называются «квартили», «децили» и «перцентили».

. **Квартили** представляют собой значения признака, делящие Ранжированную совокупность на четыре равновеликие части. Различают квартиль нижний (0|), отделяющий 1/4 часть совокупности с наименьшими значениями признака, и квартиль верхнний (Q_3), отсекающий 1/4 часть с наибольшими значениями признака. Это означает, что 25% единиц совокупности будут меньше по величине QJ; 25% единиц будут заключены между Q, и Q_2; 25% - между Q_2 и O_3 и остальные 25% превосходят Q_3. Средним квартилем Q_2 является медиана.

Для расчета квартилей по интервальному вариационному ряду используются формулы:

$$Q_1 = x_{Q_1} + i\frac{\frac{1}{4}\Sigma f - S_{Q_1-1}}{f_{Q_1}},$$

(7.54)

$$Q_3 = x_{Q_3} + i\frac{\frac{3}{4}\Sigma f - S_{Q_3-1}}{f_{Q_3}},$$

(7.55)

где x_{Q_1} 1

- нижняя граница интервала, содержащего нижний квартиль (интервал определяется по накопленной частоте, первой превышающей 25%);

x_{Q_3} — нижняя граница интервала, содержащего верхний квартиль (интервал определяется по накопленной частоте, первой превышающей 75%);

- величина интервала;

S_{Q_1-1} - накопленная частота интервала, предшествующего интервалу, содержащему нижний квартиль;

- то же для верхнего квартиля;

S_{Q_3-1} — частота интервала, содержащего нижний квартиль;

f_{Q_1} - то же для верхнего квартиля.

f_{Q_3}

Рассмотрим расчет нижнего и верхнего квартилей по данным табл. 7.13. Нижний квартиль находится в интервале 400 - 600, накопленная частота которого равна 60,2 млн чел. Верхний квартиль лежит в интервале 1000 - 1200 с накопленной частотой 116,5 млн чел. С учетом этого получим:

Кроме квартилей в вариационных рядах распределения могут определяться **децили** — варианты, делящие ранжированный ряд на десять равных частей. Первый дециль (c^) делит совокупность *в* соотношении 1/10 к 9/10, второй дециль (d-,) - в соотношении 2/10 к 8/10 и т. д.

$$Q_1 = 400 + 200 \cdot \frac{\frac{1}{4}147{,}5 - 29{,}6}{30{,}6} = 447{,}5 \text{ тыс. руб.}$$

$$Q_3 = 1000 + 200 \cdot \frac{\frac{3}{4}147{,}5 - 103{,}7}{12{,}8} = 1108{,}2 \text{ тыс. руб.}$$

Вычисляются они по той же схеме, что и медиана, и квартили:

$$d_1 = x_{d_1} + i\frac{\frac{1}{10}\Sigma f - S_{d_1-1}}{f_{d}},$$

(7.56)

$$d_2 = x_{d_2} + i\frac{\frac{2}{10}\Sigma f - S_{d_2-1}}{f_{d_2}}.$$

(7.57)

Значения признака, делящие ряд на сто частей, называются **перцентилями.** Поскольку эта характеристика применяется лишь при необходимости подробного изучения структуры вариационного ряда, приводить ее формулу и расчет не будем.

Соотношения медианы, квартилей, децилей и перцентилей представлены на рис. 7.8.

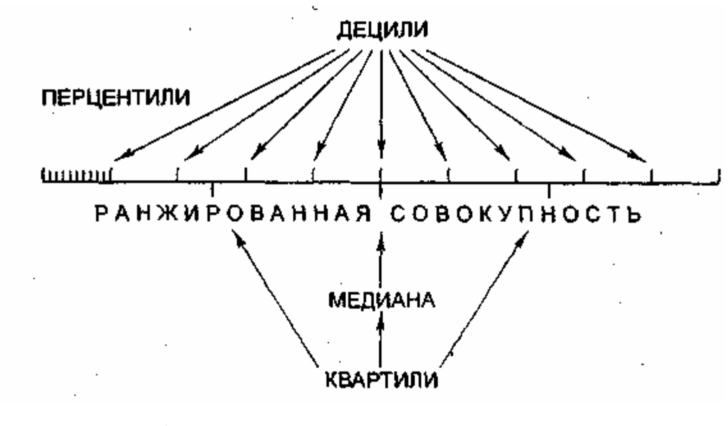


Рис. 7.8. Медиана, квартили, децили и перцентили

Использование в анализе вариационных рядов распределения рассмотренных выше характеристик позволяет более глубоко и детально охарактеризовать изучаемую совокупность.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 7

Вариация - колеблемость, многообразие, изменяемость величины признака у отдельных единиц совокупности.

К абсолютным показателям вариации относятся размах вариации, среднее линейное отклонение, среднее квадратическое отклонение, дисперсия и среднее квадратическое отклонение.

Относительные показатели вариации - это коэффициенты осцилляции, вариации, относительное линейное отклонение и др.

Размах вариации — разность между наибольшим и наименьшим значениями варьирующего признака.

Среднее линейное отклонение - средняя арифметическая из абсолютных значений отклонений вариант признака от их средней.

Дисперсия - средний квадрат отклонений индивидуальных значений признака от их средней величины.

Среднее квадратическое отклонение рассчитывается как корень квадратный из дисперсии. Среднее квадратическое отклонение, дисперсия и среднее линейное отклонение могут определяться по формулам простой и взвешенной (в зависимости от исходных данных).

Коэффициент осцилляции — процентное отношение размаха вариации к средней величине признака.

Линейный коэффициент вариации - процентное отношение среднего линейного отклонения к средней величине признака.

Коэффициент вариации - процентное отношение среднего квадратического отклонения к средней величине признака.

Энтропия - мера неопределенности данных наблюдения, которая может иметь различные результаты. Зависит от числа градаций признака и вероятности каждой из них.

Общая дисперсия измеряет вариацию признака во всей совокупности под влиянием всех факторов, обусловивших эту вариацию.

Межгрупповая дисперсия характеризует систематическую вариацию, т. е. различия в величине изучаемого признака, возникающие под действием признака-фактора, положенного в основание группировки.

Внутригрупповая дисперсия отражает случайную вариацию, т. е. часть вариации, происходящую под влиянием неучтенных факторов и не зависящую от признака-фактора.

Эмпирический коэффициент детерминации - доля межгрупповой дисперсии в общей дисперсии,

Эмпирическое корреляционное отношение - корень квадратный из эмпирического коэффициента детерминации.

Закономерности распределения - закономерности изменения частот в вариационных рядах.

Кривая распределения — графическое изображение « виде непрерывной линии изменения частот в вариационном ряду, функционально связанного с изменением вариант.

Теоретическая кривая распределения - кривая, выражающая общую закономерность данного типа распределения в чистом виде, исключающем влияние случайных факторов.

Критерии согласия - особые статистические показатели, характеризующие соответствие эмпирического и теоретического распределений. Известны критерии согласия К. Пирсона, В. И. Романовского, А. Н. Колмогорова, Б. С. Ястремского.

Мода и медиана - структурные средние. Мода - значение изучаемого признака, повторяющееся с наибольшей частотой. Медиана значение признака, приходящееся на середину ранжированной совокупности. Структурные средние могут быть определены по дискретным и интервальным рядам распределения.

Квартили — значения признака, делящие ранжированную совокупность на четыре равновеликие части.

Децили - варианты, делящие ранжированный ряд на десять равных частей.

Перцентили - значения признака, делящие ряд на сто частей.

ТЕСТ К ГЛАВЕ 7

1. Вариация - это:

а) изменение массовых явлений во времени;

б) изменение структуры статистической совокупности в пространстве;

,*.*в)* изменение значений признака во времени и в пространстве; г) изменение состава совокупности.

2. Какой из показателей вариации характеризует абсолютный размер колеблемости признака около средней величины?

а) коэффициент вариации; *б) дисперсия;

в) размах вариации;

г) среднее квадратическое отклонение.

3. Что характеризует коэффициент вариации?

а) диапазон вариации признака;

б) степень вариации признака;

в) тесноту связи между признаками;

г) пределы колеблемости признака.

4. Если все значения признака увеличить в 16 раз, то.дисперсия:

а) не изменится;

б) увеличится в 16 раз;

в) увеличится в 256 раз;

г) увеличится в 4 раза;

д) предсказать изменение дисперсии нельзя.

5.Чему"равна межгрупповая дисперсия, если отсутствуют различия между вариантами внутри групп?

а) единице;

б) нулю;

в) колеблется от нуля до единицы; в г) общей дисперсии;

д) средней из групповых дисперсий,

6. Коэффициент детерминации измеряет:

а) степень тесноты связи между исследуемыми явлениями;

б) вариацию, сложившуюся под влиянием .всех факторов;

о в) долю вариации признака-результата, сложившуюся под влиянием изучаемого (изучаемых) фактора (факторов);

г) вариацию, связанную с влиянием всех остальных факторов, кроме исследуемого (исследуемых).

7. При анализе данных о дальности перевозок грузов морским транспортом получен коэффициент асимметрии As = 0,732 **и показатель эксцесса Ek = 3,456. Это значит, что распределение:**

а) нормальное; ч

б) имеет правостороннюю асимметрию; *~)~^{7}

в) имеет левостороннюю асимметрию; ^**~

г) островершинное; t *К*-^°

д) плосковершинное. ъ\с^^

8. Проверяется соответствие эмпирического распределения нормальному. Статистическая совокупность из 245 **единиц разделена на 16 групп. Число степеней свободы для критерия** ^^{2} **равно:**

а) 244;

б) 242;

в) 16;

г) **15;** **' Д)** 13.

9. Критерий Колмогорова может быть рассчитан на основании:

а) индивидуальных данных; «б) частот;

в) частостей.

10. Средний размер реализованной коммерческой организацией спортивной обуви равен 39, мода - 39, медиана — 39. На основании этого можно сделать вывод, что распределение проданной спортивной обуви по размеру:

& а) симметричное;

б) приближенно симметричное;

в) с левосторонней асимметрией;

г) с правосторонней асимметрией;

д) данные не позволяют сделать вывод.

ГЛАВА 8

ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

8.1

ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ КАК ВАЖНЕЙШИЙ ИСТОЧНИК СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ

Понятие о выборочном наблюдении и его значение. Статистическая методология исследования массовых явлений различает, как известно, два способа наблюдения в зависимости от полноты охвата объекта: сплошное и несгаюшное. Разновидностью несплошного наблюдения является выборочное, которое в условиях развития рыночных отношений в России находит все более широкое применение. Переход статистики РФ на международные стандарты системы национального счетоводства требует более широкого применения выборки для получения и анализа показателей СНС не только в промышленности, но и в других секторах экономики.

Под выборочным наблюдением понимается такое несплошное наблюдение, при котором статистическому обследованию (наблюдению)' подвергаются единицы изучаемой совокупности, отобранные случайным способом. Выборочное наблюдение ставит перед собой задачу - по обследуемой части дать характеристику всей совокупности единиц при условии соблюдения всех правил и принципов проведения статистического наблюдения и научно организованной работы по отбору единиц.

К выборочному наблюдению статистика прибегает по различным причинам. На современном этапе появилось множество субъектов хозяйственной деятельности, которые характерны для рыночной экономики. Речь идет об акционерных обществах, малых, и совместных предприятиях, фермерских хозяйствах и т. д. Сплошное обследование этих статистических совокупностей, состоящих из десятков и сотен тысяч единиц, потребовало бы огромных материальных, финансовых и иных затрат. Использование же выборочного обследования позволяет значительно сэкономить силы и средства, что имеет немаловажное значение.

Наряду с экономией ресурсов одной из причин превращения выборочного наблюдения в важнейший источник статистической информации является возможность значительно ускорить получение необходимых данных. Ведь при обследовании, скажем, 10 % единиц совокупности будет затрачено гораздо меньше времени, а результаты могут быть представлены быстрее и будут более актуальными. Фактор времени важен для статистического исследования особенно в условиях изменяющейся социальноэкономической ситуации.

Роль выборочного обследования в получении статистических данных возрастает в силу возможности - когда это необходимо - расширения программы наблюдения. Так как исследованию подвергается сравнительно небольшая часть всей совокупности, можно более широко и детально изучить отдельные единицы и их группы.

Проведение статистического наблюдения вообще требует соответствующего кадрового обеспечения. Сплошное обследование занимает иногда слишком большое число людей для его организации проведения. Обращение же к опыту выборочного наблюдения приводит к тому, что необходимый штат сотрудников значительно уменьшается. Это позволяет привлекать более квалифицированных людей, снизить опасность появления субъективных ошибок, особенно при непосредственной регистрации фактов, и достичь поставленных целей с помощью меньшего количества более компетентных специалистов-статистиков.

Следует также отметить, что на практике приходится сталкиваться со специфическими задачами изучения массовых процессов, которые решаются лишь с помощью методологии выборки. К таким задачам относится, например, исследование качества продукции, если она при этом уничтожается. На основе выборочного наблюдения изучается, например, качество электроламп, спичек, многих сплавов и т. д. Кроме того, в современных условиях развития внешнеэкономических связей России при наличии, в частности, большого числа импортируемых продуктов и непродовольственных товаров контроль их качества обеспечивается также путем выборочного исследования.

Наконец, важным фактором превращения выборочного наблюдения в важнейший источник статистической информации является возможность его использования в целях уточнения и для разработки данных сплошного обследования. Выборочная разработка данных сплошного наблюдения связана с потребностью представления оперативных предварительных итогов обследования. Кроме того, при обобщении данных сплошного учета невозможно вести сплошную разработку по всем сочетаниям рассматриваемых признаков. Она является сложной и дорогостоящей. В этих условиях выборочный метод позволяет получить необходимые сведения приемлемой точности, когда факторы времени и стоимости делают сплошную разработку нецелесообразной.

Характеристики выборочной и генеральной совокупности. Совокупность отобранных для обследования единиц в статистике принято называть **выборочной,** а совокупность единиц, из которых производится отбор, **- генеральной.**

Основные характеристики параметров генеральной и выборочной совокупностей обозначаются определенными символами (табл. 8.1).

Таблица 8.1

Символы основных характеристик параметров генеральной и выборочной совокупностей

№ п/п	Характеристика	Генеральная совокупность	Выборочная совокупность
1	Объем совокупности (численность единиц)	N	п
2	Численность единиц, обладающих обследуемым признаком	M	m
3	Доля единиц, обладающих обследуемым признаком	$P = \frac{M}{N}$	$W = \frac{m}{n}$
4	Средний размер признака	$\bar{x} = \frac{\Sigma x_i}{N}$	$\tilde{x} = \frac{\Sigma x_i}{n}$
5	Дисперсия количественного признака	$\sigma^2_{\bar{x}} = \frac{\Sigma(x_i - \tilde{x})^2}{n}$	$\sigma^2_{\tilde{x}} = \frac{\Sigma(x_i - \tilde{x})^2}{n}$
6	Дисперсия доли	$\sigma_p{}^2 = pq$	$\sigma_w{}^2 = W(1 - W)$

Средняя и предельная ошибки выборки. Теоремы Чебышева - Ляпунова. В процессе проведения выборочного наблюдения, как и вообще при анализе данных любого обследования, статистика выделяет два вида ошибок: регистрации и репрезентативности. Ошибки регистрации могут иметь случайный (непреднамеренный) или систематический (тенденциозный) характер. Их можно избежать при правильной организации и проведении наблюдения. Ошибки репрезентативности органически присущи выборочному наблюдению и возникают в силу того, что выборочная совокупность не полностью воспроизводит генеральную. Избежать ошибок репрезентативности нельзя, однако, пользуясь методами теории вероятностей, основанными на использовании предельных теорем закона больших чисел, эти ошибки можно свести к минимальным значениям, границы которых устанавливаются с достаточно большой точностью.

Ошибка выборочного наблюдения - это разность между величиной параметра в генеральной совокупности и его величиной, вычисленной по результатам выборочного наблюдения. Для среднего значения ошибка будет определяться так:

$$\Delta_{\tilde{x}} = |\bar{x} - \tilde{x}|, \quad (8.1)$$

где $\bar{x} = \frac{\Sigma x_i}{N}$;

$$\tilde{x} = \frac{\Sigma x_i}{n}.$$

Величина $А_г$ называется предельной ошибкой выборки.

Предельная ошибка выборки величина случайная. Исследованию закономерностей случайных ошибок выборки посвящены предельные теоремы закона больших чисел. Наиболее полно эти закономерности раскрыты в теоремах П. Л. Чебышева и А. М. Ляпунова.

Теорему П. Л. Чебышева применительно к рассматриваемому методу можно сформулировать следующим образом: при достаточно большом числе независимых наблюдений можно с вероятностью, близкой к единице (т. е. почти с достоверностью), утверждать, что отклонение выборочной средней от генеральной будет сколько угодно малым. В теореме П. Л. Чебышева доказано, что величина ошибки не должна превышать tu. В свою очередь, величина ц, выражающая среднее квадратическое отклонение выборочной средней от генеральной средней, зависит от колеблемости признака в генеральной совокупности <у и числа отобранных единиц п. Эта

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

зависимость выражается формулой

(8.2)

где |i зависит также и от способа производства выборки, о чем будет сказано ниже.

Величину $\sqrt{\frac{\sigma^2}{n}}$ называют **средней ошибкой выборки** и обозначают ц.

В этом выражении о^2 - генеральная дисперсия, а n - объем выборочной совокупности.

Рассмотрим, как влияет на величину средней ошибки число отбираемых единиц п. Логически нетрудно убедиться, что при отборе большого числа единиц расхождения между средними будут меньше, т. е. существует обратная связь между средней ошибкой выборки и числом отобранных единиц. При этом здесь образуется не только просто обратная математическая зависимость, а такая зависимость, которая показывает, что квадрат расхождения между средними обратно пропорционален числу отобранных единиц. '

Далее посмотрим, как влияет колеблемость признака в генеральной совокупности на величину ошибки. Нетрудно доказать, что увеличение колеблемости признака влечет за собой увеличение среднего квадратического отклонения, а следовательно, и ошибки. Если предположить, что все единицы будут иметь одинаковую величину признака, то среднее квадратическое отклонение станет равно нулю и ошибка выборки также исчезнет. Тогда, нет необходимости применять выборку. Однако следует иметь в виду, что величина колеблемости признака в генеральной совокупности нам'бывает неизвестна, поскольку неизвестны размеры единиц в ней. Мы можем рассчитать лишь колеблемость признака в выборочной совокупности.

Соотношение между дисперсиями генеральной и выборочной совокупности выражается формулой

$$\sigma^2_{\tilde{x}} = \sigma^2_{\tilde{x}} \cdot \frac{n}{n-1}. \quad (8.3)$$

Поскольку величина ----- при достаточно больших п близка
п -1
к 1, можно приближенно считать, что выборочная дисперсия равна генеральной дисперсии, т. е. $0^2_{ген}$ ~ $<3^2_{выб}$.

Следовательно, средняя ошибка выборки показывает, какие возможны отклонения характеристик выборочной совокупности от соответствующих характеристик генеральной совокупности. Однако о величине этой ошибки можно судить с определенной вероятностью. На величину вероятности указывает множитель t,

А. М. Ляпунов доказал, что распределение выборочных средних (а следовательно, и их отклонений от генеральной средней) при достаточно большом числе независимых наблюдений приближенно нормально при условии, что генеральная совокупность обладает конечной средней и ограниченной дисперсией.

Математически теорему Ляпунова можно записать так:

$$P\{|\bar{x}-\tilde{x}| \le \Delta_{\tilde{x}}\} \approx \frac{1}{\sqrt{2n}} \cdot \int_{-t}^{t} e^{-\frac{t^2}{2}} dt = F(t), \quad (8.4)$$

где $\Delta_{\tilde{x}} = \pm t\sqrt{\frac{\sigma^2}{n}}$; *к* = 3,14 (математическая постоянная);

АХ ~ **предельная ошибка выборки,** которая дает возможность выяснить, в каких пределах находится величина генеральной средней.

Значения этого интеграла для различных значений коэффициента доверия t вычислены и приводятся в специальных математических таблицах. В частности, при

$t = 1$ $F(t) = 0{,}683$; $t = 1{,}5$ $F(t) = 0{,}866$;
$t = 2$ $F(t) = 0{,}954$; $t = 2{,}5$ $F(t) = 0{,}988$;
$t = 3$ $F(t) = 0{,}997$; $t = 3{,}5$ $F(t) \doteq 0{,}999$.

Поскольку t указывает на вероятность расхождения |х - х|, т. е. на вероятность того, на какую величину генеральная средняя будет отличаться от выборочной средней, то это может быть прочитано так: с вероятностью 0,683 можно утверждать, что разность между выборочной и генеральной средними не превышает одной величины средней ошибки выборки. Другими словами, в 68,3% случаев ошибка репрезентативности не выйдет за пределы ± ц. С вероятностью 0,954 можно утверждать, что ошибка репрезентативности не превышает ± jj, (т. е. в 95% случаев). С вероятностью 0,997, т. е. довольно близкой к единице, можно ожидать, что разность между выборочной и генеральной средней не превзойдет'трехкратной средней ошибки выборки и т. д.

Логически связь здесь выглядит довольно ясно: чем больше пределы, в которых допускается возможная ошибка, тем с большей вероятностью судят о ее величине.

Для различных способов отбора предельная ошибка рассчитывается при проведении выборки по-разному.

Зная выборочную среднюю величину признака (х) и предельную ошибку выборки (Ах), можно определить границы (пределы), в которых заключена генеральная средняя:

$$\tilde{x} - \Delta_x \le \bar{x} \le \tilde{x} + \Delta_x \text{ или } \tilde{x} - \bar{x} = \pm \Delta_{\tilde{x}}. \quad (8.5)$$

Теорема Бернулли была доказана раньше теоремы Чебышева - Ляпунова, но является лишь частным случаем последней. Она рассматривает ошибку выборки для альтернативного признака, т. е. признака, у которого возможны только два исхода: наличие признака (1) и отсутствие его (0).

Теорема Бернулли утверждает, что при достаточно большом объеме выборки вероятность расхождения между долей признака в выборочной совокупности (w) и долей признака в генеральной совокупности (р) будет стремиться к единице.

В математических символах выражение теоремы Бернулли будет иметь вид:

$$P[|w - p| \le t\mu] \to 1, \quad (8.6)$$

т. е. с вероятностью, сколько угодно близкой к единице, можно утверждать, что при достаточно большом объеме выборки частость признака (выборочная доля) сколько угодно мало будет отличаться от доли признака (в генеральной совокупности).

Ввиду того что вероятность расхождения между частостью и долей следует закону нормального распределения, эту вероятность можно найти по функции F(t) в зависимости от задаваемой величины t

Из теоремы Бернулли следует, что величина расхождения между долей признака в выборочной совокупности (частостью) и долей этого признака в генеральной совокупности зависит, так же как и в расхождениях средних, от средней ошибки выборки.

Поскольку $\mu = \frac{\sigma}{\sqrt{n}}$, а среднее квадратическое отклонение в ге-

неральной совокупности для альтернативного признака равно

уРЯ' где q = 1 - р, то средняя ошибка выборки для альтернативного признака будет найдена по формуле

$$\mu = \sqrt{\frac{p \cdot q}{n}} \quad (8.7)$$

Однако доля признака в выборочной совокупности нам неизвестна, мы вынуждены заменить ее через долю того же признака в генеральной совокупности, т. е. принять w = р, а дисперсию альтернативного признака принять за w (I -w), тогда средняя ошибка выборки выразится формулой

$$\mu = \sqrt{\frac{w(1-w)}{n}}. \quad (8.8)$$

Предельная величина разности между частостью и долей называется **предельной ошибкой выборки.** О величине предельной ошибки можно судить с некоторой вероятностью, которая зависит от множителя т, поскольку A_w = tji.

Зная выборочную долю признака (w) и предельную ошибку выборки (A_{vv}), можно определить границы, в которых заключена генеральная доля (р):

$$w - \Delta_w \le p \le w + \Delta_w. \quad (8.9)$$

Уточнение формулы средней ошибки выборки. Если отбор единиц из генеральной совокупности произведен бесповторным способом, то в формулы средней ошибки выборки вносится поправка

$$\sqrt{1-\frac{n}{N}},$$

где п - объем выборочной совокупности; N - объем генеральной совокупности.

Результаты выборочного статистического исследования во многом зависят от уровня подготовки процесса наблюдения. **Под уровнем подготовки** в данном случае подразумевается соблюдение определенных правил и принципов проектирования выборочного обследования. Важнейшим элементом проектирования является составление организационного плана выборочного наблюдения. В *общем виде в организационный план включаются следующие вопросы:*

1. Постановка цели и задачи наблюдения.
2. Определение границ объекта исследования.
3. Отработка программы наблюдения (составление анкеты, опросного листа, формы отчета и т. д.) и разработка ее материалов.
4. Определение процедуры отбора, способа отбора и объема выборки.
5. Подготовка кадров для проведения наблюдения, размножение формуляров, инструктивных документов и др.
6. Расчет выборочных характеристик и определение ошибок выборки. >
7. Распространение выборочных данных на всю совокупность. Специфические вопросы организационного плана выборочного

статистического наблюдения будут рассмотрены ниже.

8.2

ОСНОВНЫЕ СПОСОБЫ ФОРМИРОВАНИЯ ВЫБОРОЧНОЙ СОВОКУПНОСТИ

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 целого ряда условий, а именно сущности исследуемого явления, объема совокупности, вариации и распределения наблюдаемых признаков, материальных и трудовых ресурсов, 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 отборе** - группы единиц, а **комбинированный отбор** предполагает сочетание группового и индивидуального отбора.

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 также приравнивается отбор 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 следующие выборки:*

- собственно-случайная;
- механическая;
- типическая;
- серийная;
- комбинированная.

Собственно-случайная выборка заключается в отборе единиц из генеральной совокупности наугад или 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, в списках или перечне отсутствуют пропуски, игнорирования отдельных единиц и т. п. Следует также установить четкие границы генеральной совокупности таким образом, чтобы включение или невключение в нее отдельных единиц не вызывало сомнений. Так, например, при обследовании 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 необходима подготовить достаточное количество жребиев — фишек, шаров, карточек, соответствующее объему 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 значности. Например, если генеральная совокупность включает 5000 единиц, потребуются четырехзначные столбцы, при этом числа больше 5000 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 для определения возможных границ генеральных характеристик рассчитываются средняя и предельная ошибки выборки.

Средняя ошибка повторной собственно-случайной выборки определяется по формуле

$$\mu = \frac{\sigma}{\sqrt{n}}. \qquad (8.10)$$

Предположим, в результате выборочного обследования жилищных условий жителей города, осуществленного на основе собственно-случайной повторной выборки, получен следующий ряд распределения (табл. 8.2).

Таблица 8.2

Результаты выборочного обследования жилищных условий жителей города

Общая (полезная) площадь жилищ, приходящаяся на 1 человека, м^2	До 5,0	5,0-10,0	10-15,0	15,0-20,0	20,0-25,0	25,0-30,0	30,0 и более
Число жителей	8	95	204	270	210	130	83

Для определения средней ошибки выборки нам необходимо прежде всего рассчитать выборочную среднюю величину и дисперсию изучаемого признака (табл. 8-.3).

Таблица 8.3

Расчет средней общей (полезной) площади жилищ, приходящейся на 1 человека, и дисперсии

Общая (полезная) площадь жилищ, приходящаяся на 1 человека, м^2	Число жителей f	Середина интервала x	xf	x^2f
До 5,0	8	2,5	20,0	50,0
5,0- 10,0	95	7,5	712,5	5343,75
10,0-15,0	204	12,5	2550,0	31875,0
15,0-20,0	270	17,5	4725,0	82687,5
20,0 - 25,0	210	22,5	4725,0	106312,5
25,0-30,0	130	27,5	3575,0	98312,5
30,0 и более	83	32,5	2697,5	87668,75
Итого	1000	-	19005,0	412250,0

$$\tilde{x} = \frac{19005,0}{1000} = 19,0;$$

$$\sigma^2 = \frac{412250}{1000} - 19,0^2 = 51,25;$$

$$\sigma = \sqrt{51,25} = 7,16.$$

Средняя ошибка выборки составит:

$$\mu_{\tilde{x}} = \frac{7,16}{\sqrt{1000}} = 0,23 \text{ м}^2.$$

Определим предельную ошибку выборки с вероятностью 0,954 (t = 2):

Установим границы генеральной средней:

$$\tilde{x} - \Delta_{\tilde{x}} \le \bar{x} \le \tilde{x} + \Delta_{\tilde{x}}$$

или с учетом полученных значений:

$$18,54 \le \bar{x} \le 19,46.$$

Таким образом, на основании проведенного выборочного обследования с вероятностью 0,954 можно заключить, что средний размер общей площади, приходящейся на 1 человека, в целом по городу лежит в пределах от 18,5 до 19,5 м^2.

При расчете средней ошибки собственно-случайной бесповторной выборки необходимо учитывать поправку на бесповторность отбора:

$$\mu = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}. \qquad (8.11)$$

Если предположить, что представленные в табл. 8.2 данные являются результатом 5%-ного бесповторного отбора (следовательно, генеральная совокупность включает 20 000 единиц), то средняя ошибка выборки будет несколько меньше:

$$\mu_{\tilde{x}} = \sqrt{\frac{51,25}{1000}\left(1 - \frac{1000}{20000}\right)} = 0,22 \text{ м}^2.$$

Соответственно уменьшится и предельная ошибка выборки, что вызовет сужение границ генеральной средней. Особенно ощутимо влияние поправки на бесповторность отбора при большом проценте выборки.

Воспользуемся еще раз данными табл. 8.2 для того, чтобы определить границы доли лиц, обеспеченность жильем которых составляет менее 10 м^2- Согласно результатам обследования, численность таких лиц составила 103 человека. Определим выборочную долю и дисперсию:

$$W = \frac{103}{1000} = 0,103;$$

$$\sigma^2_w = w\,(1 - w) = 0,103 \cdot 0,897 = 0,0924.$$

Рассчитаем среднюю ошибку выборки:

$$\mu_{\tilde{w}} = \sqrt{\frac{0,0924}{1000}\left(1 - \frac{1000}{20000}\right)} = 0,0094.$$

Предельная ошибка выборки с заданной вероятностью составит:

$$\Delta_w = 2 \cdot 0,0094 = 0,0188.$$

Определим границы генеральной доли: или

$$0,103 - 0,019 \le p \le 0,103 + 0,019$$

$$0{,}084 \le p \le 0{,}122.$$

Следовательно, с вероятностью 0,954 можно утверждать, что доля лиц, имеющих менее 10 м^2 на человека, в целом по городу находится в пределах от 8,4 до 12,2%.

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 механической выборки устанавливается пропорция отбора, которая определяется соотнесением объемов выборочной и генеральной совокупностей. Так, если из совокупности в 500 000 единиц предполагается получить 2%-ную выборку, т. е. отобрать 10 000 единиц, то пропорция отбора соста-

i _Г 1 1

вит 50 I 500000 -10000 Отбор единиц осуществляется в соответствии с установленной пропорцией через равные интервалы. Например, при пропорции 1 : 50 (2%-ная выборка) отбирается каждая 50-я единица, при пропорции 1 : 20 (5 %-ная выборка) каждая 20-я единица и т. д.

Генеральною совокупность при механическом отборелможно 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%ной 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%-ного 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 располагаются именно по 4 квартиры. Чтобы избежать систематической ошибки, в каждом новом подъезде счетчик менял начало отбора.

Для определения средней ошибки механической выборки" используется формула средней ошибки при собственно-случайном бесповторном отборе.

Типический отбор. Этот способ отбора 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:

$$n_i = n\frac{N_i}{N},$$

где NJ - объем i-й группы;

п(— объем выборки из i-й группы.

Средняя ошибка такой выборки находится по формулам:

$$\mu = \sqrt{\frac{\overline{\sigma}^2}{n}}$$ (повторный отбор), (8.12)

$$\mu = \sqrt{\frac{\overline{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)}$$

(бесповторный отбор), (8.13)

где о? - средняя из внутригрупповых дисперсий.

При выборке, пропорциональной дифференциации признака, число наблюдений по каждой группе рассчитывается по формуле

$$n_i = n\frac{\sigma_i N_i}{\Sigma\sigma_i N_i},$$

где Oj - среднее квадратическое отклонение признака в i-й группе.

Средняя ошибка такого отбора определяется следующим образом:

$$\mu = \frac{1}{N}\sqrt{\Sigma\frac{\sigma_i^2 N_i^2}{n_i}}$$

(повторный отбор), (8.15)

$$\mu = \frac{1}{N}\sqrt{\Sigma\frac{\sigma_i^2 N_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)} \text{ (бесповторный отбор).}$$

(8.16)

Отбор, пропорциональный дифференциации признака, дает лучшие результаты, однако на практике его применение затруднено вследствие трудности получения сведений о вариации до проведения выборочного наблюдения.

Рассмотрим оба варианта типической выборки на условном примере. Предположим, 10%-ный бесповторный типический отбор рабочих предприятия, пропорциональный размерам цехов, проведенный с целью оценки потерь из-за временной нетрудоспособности, привел к следующим результатам (табл. 8.4).

Таблица 8.4 **Результаты обследования рабочих предприятия**

Цех	Всего , рабочих, человек	Обследовано, человек	Число дней временной нетрудоспособности за год	
			средняя	дисперсия
I	1000	100	18	49
11	1400	140	12	25
HI	800	80	15	16

Рассчитаем среднюю из внутригрупповых дисперсий:

$$\bar{\sigma}^2 = \frac{\Sigma\sigma_i^2 n_i}{\Sigma n_i} = \frac{49\cdot 100+25\cdot 140+16\cdot 80}{100+140+80} = 30{,}25.$$

Определим среднюю и предельную ошибки выборки (с вероятностью 0,954):

$$\mu_{\tilde{x}} = \sqrt{\frac{30{,}25}{320}\left(1-\frac{320}{3200}\right)} = 0{,}29,$$

$$\Delta_{\tilde{x}} = 2\cdot 0{,}29 = 0{,}58.$$

Рассчитаем выборочную среднюю:

$$\tilde{x} = \frac{\Sigma x_i n_i}{\Sigma n_i} = \frac{18\cdot 100+12\cdot 140+15\cdot 80}{100+140+80} = 14{,}6 \text{ дня.}$$

С вероятностью 0,954 можно сделать вывод, что среднее число дней временной нетрудоспособности одного рабочего в целом по предприятию находится в пределах:

$$14{,}6 - 0{,}58 \le \bar{x} \le 14{,}6 + 0{,}58.$$

Воспользуемся полученными внутригрупповыми дисперсиями для проведения отбора, пропорционального дифференциации признака. Определим необходимый объем выборки по каждому цеху:

$$\Sigma\sigma_i N_i = \sqrt{49}\cdot 1000 + \sqrt{25}\cdot 1400 + \sqrt{16}\cdot 800 = 17200;$$

$$n_1 = 320\cdot\frac{\sqrt{49}\cdot 1000}{17200} = 130 \text{ человек;}$$

$$n_2 = 320\cdot\frac{\sqrt{25}\cdot 1400}{17200} = 130 \text{ человек;}$$

$$n_3 = 320\cdot\frac{\sqrt{16}\cdot 800}{17200} = 60 \text{ человек.}$$

С учетом полученных значений рассчитаем среднюю ошибку выборки:

$$\mu_{\bar{x}} = \frac{1}{3200}\sqrt{\frac{49\cdot 1000^2}{130}\left(1-\frac{130}{1000}\right)+\frac{25\cdot 1400^2}{130}\left(1-\frac{130}{1400}\right)+\frac{16\cdot 800^2}{60}\left(1-\frac{60}{800}\right)} = 0{,}28.$$

В данном случае средняя, а следовательно, и предельная ошибки будут несколько меньше, что отразится и на границах генеральной средней.' .

Серийный отбор. Данный способ отбора удобен в тех случаях, когда единицы совокупности объединены в небольшие группы или серии. В качестве таких серий могут рассматриваться упаковки с определенным количеством готовой продукции, партии товара, студенческие группы, бригады и другие объединения. Сущность серийной выборки заключается в собственнослучайном либо механическом отборе серий, внутри которых производится сплошное обследование единиц.

Поскольку внутри групп (серий) обследуются все без исключения единицы, средняя ошибка серийной выборки (при отборе равновеликих серий) зависит от величины только

$$\mu = \sqrt{\frac{\delta^2}{r}} \text{ (повторный отбор),}$$

межгрупповой (межсерийной) дисперсии и определяется по следующим формулам:

(8.17)

$$\mu = \sqrt{\frac{\delta^2}{r}\left(1-\frac{r}{R}\right)} \quad \text{(бесповторный отбор),} \quad (8.18)$$

где г - число отобранных серий; R - общее число серий.

Межгрупповую дисперсию вычисляют следующим образом:

$$\delta^2 = \frac{\Sigma(\tilde{x}_i - \tilde{x})^2}{r}, \quad (8.19)$$

где х, - средняя i-й серии;

х - общая средняя по всей выборочной совокупности.

Комбинированный отбор. В практике статистических обследований помимо рассмотренных выше способов отбора применяется и их комбинация. Так, например, можно комбинировать типическую и серийную выборки, когда серии отбираются в установленном порядке из нескольких типических групп. Возможна также комбинация серийного и собственно-случайного отборов, при которой 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.

В отличие от многоступенчатой **многофазная выборка** предполагает сохранение одной и той же единицы отбора на всех этапах его проведения, при этом отобранные на каждой стадии единицы подвергаются обследованию (на каждой последующей стадии отбора программа обследования расширяется).

Исходя из вышеизложенного приведем формулы предельной ошибки выборки для наиболее часто используемых на практике способов формирования выборочной совокупности (табл. 8.5).

Таблица 8.5

Предельная ошибка выборки для некоторых способов формирования выборочной совокупности

\ Метод Хртбора \ Зы- \ Зорка \	Повторный		Бесповторный	
	для сред-ней	для доли	для средней	для доли
1. Соб- ственносл учайная и механи- ческая	* a^2 1 n	jw(l-w) V п	tl$^{\circ2}$fl "1 1~1 NJ	|w(l-w)r $_n$] \ n [^{1} $_N$J
2. Типи- ческая (при пропорцио нальном объему групп отборе)	*t ^ n	/Wj(l-Wj) 1 п	tl^fl "1 *i n (*^{l} $_N$]	(w$_i$(l-w$_i$)f n^ 1 п i'-NJ
3. Серийная (довая)	* Гх2 M ~ m	ʌ/ʃ V г	сl^f^1__] V ' 1 ^{R}]	u 1 ^'-- 1 If rl RJ

8.3

ОПРЕДЕЛЕНИЕ НЕОБХОДИМОГО ОБЪЕМА ВЫБОРКИ

При проектировании выборочного наблюдения возникает вопрос о необходимой численности выборки. Эта численность может быть определена на базе допустимой ошибки при выборочном наблюдении исходя из вероятности, на основе которой можно гарантировать величину устанавливаемой ошибки, и наконец, на базе способа отбора. Рассмотрев вначале величину необходимой численности в общем виде, мы исследуем в дальнейшем особые условия, создающиеся в процессе ее вычисления при разных способах отбора.

Для определения необходимой численности выборки исследователь должен задать уровень точности выборочной совокупности с определенной вероятностью. В частности, необходимая численность случайной повторной выборки определяется по формуле

$$n = \frac{t^2 \cdot \sigma^2}{\Delta^2}, \quad (8.20)$$

которая вытекает из формулы предельной ошибки: $\Delta = t\sqrt{\frac{\sigma^2}{n}}$.

Эта формула показывает, что с увеличением предполагаемой ошибки выборки значительно уменьшается необходимый объем выборки. Так, увеличение допустимой ошибки выборки в 2 раза уменьшает необходимый ее объем в 4 раза. Необходимая численность выборки прямо пропорциональна дисперсии признака и величине t^2. Формула необходим эй численности выборки для разных способов отбора выводится из формулы предельной ошибки выборки.

Необходимая численность выборки рассчитывается по-разному для выборочного наблюдения, в котором устанавливается средний размер признака в совокупности, и для наблюдения, в котором определяется доля единиц, обладающих данным признаком, в силу различных методов вычисления меры колеблемости для варьирующего и альтернативного признаков.

На практике определение необходимого объема выборки часто составляет серьезную проблему. Она связана, в частности, с недостаточной разработанностью таких вопросов, как оценка вариации изучаемых признаков, обоснование численности выборки при изучении нескольких признаков, зависимость объема выборочной совокупности от программы разработки материалов наблюдения и др.

Трудности порождаются и тем, что кроме чисто статистических в определении необходимой численности выборочной совокупности большое значение принадлежит факторам организационного порядка, которые должны быть обязательно учтены. К ним относятся, например, обеспеченность обследования ресурсами, длительность обработки и срочность представления результатов. Согласование объема выборки с материальными, финансовыми, кадровыми ресурсами и тому подобное вызывает определенную сложность.

Одним из наиболее важных и в то же время сложных вопросов определения необходимого объема выборки в исследованиях является расчет показателя вариации изучаемого признака (а). При подготовке выборочного наблюдения у его организаторов часто отсутствуют необходимые для этих вычислений данные. Основой оценки степени колеблемости изучаемого признака служат, как правило, материалы предыдущих обследований. Обращение к ним при отсутствии какой-либо другой информации вполне оправданно. Однако следует иметь в виду, что использование данных прошлых обследований имеет смысл только тогда, когда за прошедший до нового обследования период в генеральной совокупности не произошло значительных изменений.

Во многих случаях более точное представление об изучаемой совокупности, в том числе о вариации интересующих исследователя признаков, может дать пробное обследование. По его данным возможно рассчитать среднее квадратическое отклонение и дисперсию для последующего обоснования необходимого объема выборки. Если же мера колеблемости признака неизвестна, то ее можно найти приближенно по величине предполагаемого размаха или среднего линейного отклонения по следующим формулам:

$$\sigma \approx \frac{R}{6} \quad (8.21)$$

$$\sigma = 1{,}25\bar{d}, \quad (8.22)$$

где о - среднее квадратическое отклонение;
R - размах вариации; . d - среднее линейное отклонение.

Важным условием практического использования этих формул является близость фактического распределения к нормальному. Исчисление среднего квадратического отклонения для явно несимметричных распределений не имеет смысла.

При статистическом исследовании социально-экономических явлений очень часто приходится

сталкиваться с качественными признаками, причем именно по ним нередко проводится расчет необходимого объема выборочной совокупности. Способ выра-. жения качественных признаков не позволяет рассчитать по ним
средние значения, поэтому оценка колеблемости производится, как правило, исходя из долей единиц, обладающих значениями этих признаков, т. е. **выборочных долей.** Выборочная доля также называется **частостью.**

Если расчет проводится по качественному альтернативному признаку и неизвестна его доля в генеральной совокупности (хотя бы приблизительно), рекомендуется принять ее равной 0,5, так как дисперсия доли достигает максимума: a^* = 0,25 при w = 0,5.

Преимущество такого приема заключается в том, что он позволяет определить численность выборочной совокупности, не располагая данными предыдущих обследований, и не проводить пробных обследований. Возможность экономии времени и ресурсов часто оказывается решающим фактором при обращении к данному методу.

Если же качественный признак, по которому определяется •необходимая численность выборочной совокупности, не является альтернативным, то использовать формулу a^ = w (1 - w) нельзя.

В ряде случаев приближенная оценка колеблемости может быть осуществлена с помощью превращения изучаемого признака в альтернативный. Например, все категории работников предприятия можно условно разделить в зависимости от при- ' надлежности работающих к рабочим и служащим. Однако при этом следует учитывать, что такое деление неизбежно приведет к потере некоторой части информации. Ведь существуют отдельные категории работников (МОП, охрана и др.), которые выделяются в самостоятельные группы. Поэтому применять описанный выше прием можно лишь при условии, что существует уверенность в незначительной доле неучтенных единиц во всей совокупности.

Приведем формулы необходимого объема выборки для наиболее часто используемых на практике способов формирования выборочной совокупности (табл. 8.6).

Рассмотрим несколько примеров расчета объема выборки при различных способах отбора.

Пример 1. В микрорайоне проживает 5000 семей. В порядке случайной бесповторной выборки предполагается определить средний размер семьи при условии, что ошибка выборочной средней не должна превышать 0,8 человека с вероятностью Р = 0,954 и при среднем квадратическом отклонении 3,0 человека (ошибка и среднее квадратическое отклонение определены на основе пробного обследования).

Таблица 8.6

Необходимый объем выборки для некоторых способов формирования выборочной совокупности

Вид выборочного наблюдения	Повторный отбор	Бесповторный отбор
Собственно- случайная выборка: а) при определении	n-$^{t?}$"4	t^2-a?.N
признака . б) при	4 t^2-W(1-W)	n—A^2s-N + t^2-o| t^2-W(1-W)-N
доли признака Механическая выборка Типическая выборка: а) при определении	" A'$_w$то же n - $^{t\ 2\ ,\ \wedge}$	$A^2{}_w$-N + t^2-W(l-W) Ч то же t^2.a-^{2}-N
признака	A|	A|-N + t^2-a|
б) при определении	t^2-W(1-W)	t^2-W(1-W)'N

доли признака Серийная выборка: а) при определении	" Ai r	$A^2{}_w$-N + t^2-W(l-W) t^2-
признака б) при определении	A| t^2-W_r(l- W_r)	A^2_s-R+t^2-8? t^2-
доли признака	г A^2_W	$A^2{}_w$-R + t^2-W_r(l-W_r)

При P = 0,954 t = 2 необходимая численность выборки

$$n = \frac{2^2 \cdot 3^2 \cdot 5000}{5000 \cdot 0{,}64 + 2^2 \cdot 3^2} = \frac{4 \cdot 9 \cdot 5000}{5000 \cdot 0{,}64 + 4 \cdot 9} = \frac{180000}{3236} \approx 56 \text{ семей.}$$

Пример 2. Для определения средней длины детали следует провести выборочное обследование методом случайного повторного отбора. Какое количество деталей надо отобрать, чтобы ошибка выборки не превышала 3 мм с вероятностью 0,997 при среднем квадратическом отклонении 6 мм?*

При t = 3 и P = 0,997 объем выборки рассчитывается следующим образом:

$$n = \frac{3^2 \cdot 6^2}{3^2} = 36 \text{ деталей.}$$

Пример 3. В фермерских хозяйствах области 10 000 коров. Из них в районе А - 5000, в районе Б - 3000, в районе В - 2000. 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 1600?

Рассчитаем необходимую численность типической выборки:

$$n = \frac{2^2 \cdot 1600 \cdot 10000}{5^2 \cdot 10000 + 2^2 \cdot 1600} = \frac{64000000}{250000 + 6400} = 249{,}6 \approx 250 \text{ коров.}$$

Необходимо отобрать 250 коров, из них

$$\text{в районе А: } n_1 = 250 \cdot \frac{5000}{10000} = 125 \text{ коров;}$$

$$\text{в районе Б: } n_2 = 250 \cdot \frac{3000}{10000} = 75 \text{ коров;}$$

$$\text{в районе В: } n_3 = 250 \cdot \frac{2000}{10000} = 50 \text{ коров.}$$

* Ошибка и среднее квадратиче.ское отклонение заданы исходя из технических нормативов.

Пример 4. 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 дисперсия серийной выборки равна 4. Определить необходимый объем выборки.

$$r = \frac{2^2 \cdot 4 \cdot 100}{100 \cdot 2^2 + 2^2 \cdot 4} = \frac{1600}{416} \approx 4 \text{ ящика.}$$

Подробное рассмотрение вопросов определения дисперсии для нахождения объема выборки не исключает использования в этих целях других показателей вариации. В последние годы были проведены исследования по разработке методики нахождения необходимой численности выборки при заданных значениях отдельных параметров, и в частности коэффициентов вариации. Методики, разработанные в рамках конкретных обследований и определенных способов формирования выборочной совокупности, требуют дальнейшего теоретического обоснования и практической проверки.

Ц^-ТлМАС. 8.4

ОЦЕНКА РЕЗУЛЬТАТОВ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ И РАСПРОСТРАНЕНИЕ ИХ НА ГЕНЕРАЛЬНУЮ СОВОКУПНОСТЬ

Заключительным этапом выборочного наблюдения является распространение его результатов на генеральную совокупность. Однако часто при статистическом изучении социально-экономических явлений этому процессу предшествует оценка результатов наблюдения с точки зрения самой возможности распространения.

Вывод о возможности распространения в значительной степени зависит от качества основы выборки, прежде всего от ее полноты. Под полнотой подразумевается наличие или представленность всех типов или групп данной генеральной совокупности в основе выборки. Неполнота основы может привести к нарушению представительности выборки и, как следствие, к неправильным выводам при анализе данных наблюдения.

Однако не следует обосновывать возможность распространения выборочных данных только анализом качества исходной информации для отбора. Более точной основой суждения о возможности распространения представляется расчет относительной ошибки:

$$\text{для средней: } \Delta_{\%} = \frac{\Delta_{\bar{x}}}{\bar{x}} \cdot 100\%, \qquad (8.23)$$

$$\text{для доли: } \Delta_{\%} = \frac{\Delta_{w}}{\bar{p}} \cdot 100\%, \qquad (8.24)$$

где Ло/о — относительная предельная ошибка выборки; А^ и $А_w$ — предельная ошибка для среднего значения или доли признака соответственно; х и р - генеральная средняя и доля соответственно.

Суждение о возможности распространения выборочных данных можно составить, если в формулах (8.23) и (8.24) заменить х и р соответствующими выборочными характеристиками. Необходимым условием при этом является соответствие плановой и фактической численности и структуры выборочной совокупности. При больших расхождениях использование этого приема может привести к ошибочным суждениям.

Если величина относительной ошибки не превышает заранее установленного для данного обследования предельного значения, то данные выборочного наблюдения являются представительными и могут быть распространены на генеральную совокупность. В противном случае следует попытаться восстановить исходные пропорции генеральной совокупности. Процесс восстановления пропорций выборки на основе исходной информации о таких пропорциях в генеральной совокупности принято называть **корректировкой*** выборки.

При обработке данных выборочного наблюдения целесообразно использовать два наиболее часто применяемых способа корректировки. Первый способ ориентирован на группу единиц, которые оказались недостаточно представлены в выборочной совокупности после наблюдения. Формуляры с данными об этих единицах, пригодные для обработки, следует сохранять в полном объеме. На основе сведений о количестве таких формуляров проводятся дополнительные расчеты. Их целью является определение числа хорошо представленных в фактической выборке формуляров остальных групп, часть которых необходимо исключить из обработки для сохранения пропорций генеральной совокупности. Данный способ корректировки, называемый **методом «отсечения»,** поясним на следующем условном примере.

С целью изучения общественного мнения из генеральной совокупности численностью 1000 человек было отобрано в порядке типической пропорциональной выборки 100 человек, принадлежащих к различным социальным группам: рабочие, служащие, студенты. При этом в генеральной совокупности было 50% рабочих, 35% служащих и 15% студентов, т. е. пропорция по группам населения составила примерно 3,3 : 2,3 : 1. Следовательно, для обеспечения представительности выборки по признаку социального положения требовалось бы получить данные о 50 рабочих, 35 служащих и 15 студентах. Однако по тем или иным причинам часть анкет не была получена, а другая часть была забракована. В результате пригодными для дальнейшей обработки оказались 40 анкет, заполненных рабочими, 30 - служащими и 10 - студентами. Таким образом, пропорции по различным группам в массиве для обработки составили 4:3:1, что свидетельствует о нарушении структуры генеральной совокупности.

Для проведения корректировки необходимо определить, анкеты какой социальной группы респондентов должны быть сохранены в процессе обработки полностью. Это можно сделать с помощью относительных величин, вычисляемых для каждой социальной категории как отношение числа пригодных для обработки анкет к общему количеству анкет по данной группе. Результат выражается в долях или процентах. Расчеты показывают, что наименьшая относительная величина получается по студентам (приблизительно 66,7%). Следовательно, формуляры, относящиеся к данной группе населения, необходимо сохранить полностью. Чтобы восстановить реальные пропорции генеральной совокупности, нужно вновь обратиться к ее структуре, выраженной соотношением 3,3 : 2,3 : 1.

Несложные вычисления показывают, что для сохранения представительности выборки в массиве анкет для дальнейшей обработки должны быть 33 анкеты, заполненные рабочими, 23 —. служащими и 10 - студентами. Таким образом, из дальнейшей •обработки следует «отсечь» по 7 анкет, относящихся к рабочим и служащим. Для «отсечения» возможно пользоваться процедурой случайной выборки.

После «отсечения» следует проверить, как исключение некоторого числа формуляров повлияло на обобщающие показатели фактической выборки. Для этого вначале следует найти средние по важнейшим показателям в совокупности пригодных для обработки формуляров, включая те, которые затем предполагается «отсечь». Затем те же средние рассчитать по совокупности формуляров, оставшихся после «отсечения», и сравнить полученные результаты. Для оценки различий средних можно воспользоваться принципами оценки точности выборки. Если расхождения между средними, рассчитанными до и после «отсечения», не превышают ± 5%, итоги корректировки считаются вполне удовлетворительными. В противном случае ее целесообразно повторить, исключив из обработки другие формуляры.

Для проверки степени влияния «отсечения» на обобщающие характеристики выборки используют и другой способ. Он заключается в сравнении средних, рассчитанных для различных вариантов «отсекаемых» формуляров, со средними, вычисленными до и после «отсечения».

Приведенный пример позволяет рассмотреть достоинства данного метода корректировки и его недостатки. Основным достоинством метода является то, что он дает возможность сохранить пропорции генеральной совокупности в массиве данных, на основе которого будут делаться обобщения. Это позволяет формулировать выводы на базе представительных данных.

В то же время корректировка способом «отсечения» имеет существенные недостатки. *Во-первых,* «отсечение» приводит к еще большему, если учитывать невозвращенные и забракованные формуляры, уменьшению объема выборки. *Во-вторых,* из обработки и анализа исключаются вполне пригодные для исследования формуляры. В нашем примере «отсекаются» 17,5% собранных формуляров по группе рабочих и примерно 23,3% по группе служащих. Сами по себе эти показатели весьма значительны. Однако и они могут возрасти, если для обработки окажутся пригодными не 10, а, скажем, 7 анкет студентов. Тогда из обработки потребуется исключить 17 из 40 анкет рабочих (42,5%) и 14 из 30 анкет служащих (около 46,7%). В таких случаях целесообразнее пользоваться другим способом корректировки - с помощью «взвешивания».

В отличие от первого способа корректировки **«взвешивание»** дает возможность сохранить в обрабатываемом массиве все или почти все полученные формуляры. Достигается это путем многократного использования при обработке части формуляров. При этом несколько раз используются, как правило, те формуляры,

число которых настолько мало, что вызывает необходимость исключения из дальнейшей обработки большого числа для исследования формуляров, относящихся к другим группам. Многократное применение формуляров проводится на основе специально рассчитанных для этой цели «весов».

Метод «взвешивания» наиболее удобно применять при обработке материалов выборочных обследований в случаях высокого процента невозвращенных или забракованных формуляров. Это характерно прежде всего для почтовых опросов.

---

* В отечественной литературе данный процесс называют также коррекцией, «ремонтом» выборки. В зарубежной литературе получил распространение термин «постстатификация».

Наиболее широко применяемые в социальных исследованиях методы корректировки выборки могут использоваться и при выборочных обследованиях промышленности, аграрного сектора и других отраслей экономики.

Собранные в результате выборочного наблюдения и при необходимости откорректированные данные распространяются на генеральную совокупность. Существуют два основных метода распространения - **прямой пересчет и способ коэффициентов.**

Сущность **способа прямого пересчета** заключается в умножении среднего значения признака, найденного в результате выборочного наблюдения, на объем генеральной совокупности." Практические расчеты при этом не вызывают серьезных затруднений. Например, на основании выборочного обследования 1000 молодых семей требуется оценить потребность в местах в детских яслях. С помощью метода прямого пересчета это можно сделать следующим образом. Известно, что ясли могут посещать дети в возрасте до трех лет. По материалам выборочного обследования следует вычислить среднее число детей этого возраста. Предположим, что оно составляет 1,3 человека. Умножив это число на численность генеральной совокупности, получим, что в детских яслях потребуется выделить 1300 мест.

Производя такие расчеты, мы считаем, что были обследованы все единицы, попавшие в выборочную совокупность. Однако в социальных исследованиях объемы фактической и запланированной выборки часто не совпадают, что всегда следует учитывать. Как правило, несоответствие фактической и запланированной выборки приводит, естественно, к неадекватному отражению в выборочных характеристиках, полученных по фактическим данным соответствующих характеристик генеральной совокупности.

Предположим, в нашем примере некоторое число семей по тем или иным причинам не было обследовано. Это привело к снижению объема фактической выборки по сравнению с запланированной. Среднее число детей, вычисленное по этой «неполной» выборке, составило не 1,3, а 1,2. Тогда прямой пересчет выборочной характеристики на объем генеральной совокупности даст результат 1 200 мест. Абсолютное отклонение от необходимого количества мест при условии охвата обследованием всей выборочной совокупности составит 100 мест, а относительное приблизительно 7,7%*. Если же объем генеральной совокупности был бы в 10 раз больше, т. е. 10 000 семей, то абсолютное отклонение также увеличится в 10 раз и при сохранении тех же различий среднего числа детей составит 1000 мест, относительное отклонение при этом не изменится. Таким образом, размер абсолютного отклонения находится в прямой зависимости от объема генеральной совокупности.

Данный пример показывает: недоучет обстоятельства, при котором на практике объемы фактической и запланированной выборок часто не совпадают, приводит к серьезным ошибкам при использовании распространенных на генеральную совокупность результатов таких исследований. Руководствуясь данными, рассчитанными на условном примере, пришлось бы принимать ошибочное решение о строительстве дополнительного числа детских учреждений, мест в которых не хватило бы на 100 (или на 1000 — в зависимости от объема генеральной совокупности) детей. Но могла возникнуть и обратная ситуация, когда вычисленное по «неполной» выборочной совокупности среднее число детей оказалось бы больше «истинного». В этом случае появились бы «лишние» места. Данный пример показывает, что результатами выборочного наблюдения необходимо пользоваться осторожно, особенно в случаях, когда их использование связано с большими материальными затратами.

В условиях существования большого числа факторов, влияющих на точность данных выборочного наблюдения, использование точечной оценки при распространении выборочных характеристик на генеральную совокупность в социальных исследованиях часто нецелесообразно. Во всех случаях, когда это возможно, правильнее пользоваться интервальной оценкой, позволяющей учесть размер предельной ошибки выборки, рассчитанной для средней или для доли признака. Так, если в нашем примере среднее число детей в возрасте до трех лет по выборочным данным составило 1,3 человека, а предельная ошибка - ± 0,1 человека, то требуемое, количество мест в детских учреждениях будет находиться в пределах от 1200 до 1400.

---

* Относительное отклонение рассчитано как отношение абсолютного отклонения к характеристике генеральной совокупности, вычисленной в процессе распространения, при условии совпадения фактической и плановой выборки.

Наряду со способом прямого пересчета при распространении данных выборочного наблюдения на генеральную совокупность применяется так называемый **способ коэффициентов.** Данный способ целесообразно использовать в случаях, когда выборочное наблюдение проводится с целью проверки и уточнения данных сплошного наблюдения, в частности численности учтенных единиц совокупности.

При этом следует использовать следующую формулу:

$$Y_1 = Y_0 \cdot \frac{y_1}{y_0}, \qquad (8.25)$$

где Y| - численность совокупности с поправкой на недоучет;
Y_0 - численность совокупности без этой поправки;
y_0 - численность совокупности в контрольных точках по первоначальным данным;
у, - численность совокупности в тех же точках по данным контрольных мероприятий.

Отметим, что цели исследования многих явлений могут быть достигнуты только путем сплошного наблюдения. Поэтому способ проверки результатов сплошного наблюдения на основе коэффициентов успешно применяется в социальной и экономической статистике. До сих пор возможности выборки при уточнении данных сплошного наблюдения используются недостаточно. В то же время в современных условиях данный способ может быть, например, инструментом в контроле за деятельностью коммерческих структур со стороны финансовых органов.

При уточнении данных сплошного наблюдения на основе контрольных выборочных мероприятий определяется так называемая **поправка на недоучет.** Метод ее расчета наиболее широко применяется в обследованиях относительно небольших совокупностей, когда их объем не превышает нескольких сотен или тысяч единиц.

Пример. При проведении учета коммерческих палаток в городе было зарегистрировано следующее их количество в районах: А - 2000; Б - 1500; В - 750. С целью проверки данных сплошного учета проведены контрольные обходы части обследованных районов. Их результаты содержатся в нижеприведенной табл. 8.7.

Рассчитанный по каждой категории работников коэффициент недоучета является основой уточнения при распространении данных, полученных при выборочных контрольных мероприятиях, на всю совокупность.

Таблица 8.7

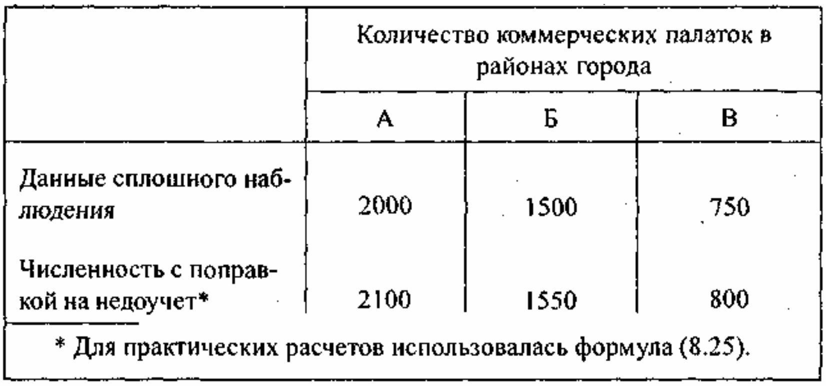
Количество коммерческих палаток в районах города до и после контрольных обходов

Район	Зарегистрировано при сплошном учете	Установлено при контрольном обходе	Коэффициент недоучета
А	400	420	1,050
Б	300	310	1,033
В	150	160	1,067

В нашем примере количество коммерческих палаток (по данным сплошного учета) следует умножить на рассчитанный для каждого района коэффициент недоучета. В итоге получим результаты, представленные в табл. 8.8.

Таблица 8.8

Уточненные данные учета коммерческих палаток в районах города



	Количество коммерческих палаток в районах города		
	А	Б	В
Данные сплошного наблюдения	2000	1500	750
Численность с поправкой на недоучет*	2100	1550	800

* Для практических расчетов использовалась формула (8.25).

Применение метода коэффициентов связано, как правило, с использованием выборочного наблюдения с целью проверки данных сплошного наблюдения. Однако это приводит к сознательному ограничению области применения данного метода. Метод коэффициентов можно использовать для проверки данных выборочного наблюдения, когда необходима очень высокая точность результатов и выборочная совокупность имеет большой объем - порядка нескольких тысяч или десятков тысяч единиц. В таких случаях списки единиц обследованной выборочной совокупности служат основой для отбора единиц в «контрольную» выборку, т. е. производится выборка из выборки. Контроль и, если это необходимо, уточнение данных основного обследования проводятся по методике, описанной выше на условном примере. Способ поправочных коэффициентов целесообразно использовать для распространения данных выборочного наблюдения в случаях, если его результаты значительно уступают в точности данным статистической отчетности или точность собранного статистического материала вызывает сомнение. При этом, естественно, данные способы уточнения результатов выборочного наблюдения могут быть использованы при наличии времени и средств для проведения контрольных мероприятий.

Завершая рассмотрение вопросов оценки точности и распространения данных выборочного наблюдения на генеральную совокупность, отметим: публикуя где бы то ни было статистические данные, рассчитанные на основе выборки, обязательно следует пояснять, что соответствующие показатели получены расчетным путем на базе материалов выборочного наблюдения. Одновременно было бы правильно указать пределы ошибок, которые могли быть допущены в процессе выборки, а также вероятность этих пределов.

8,5

МАЛАЯ ВЫБОРКА

В процессе оценки степени представительности данных выборочного наблюдения важное значение приобретает вопрос об объеме выборочной совокупности п. От него зависит не только величина пределов, которые с данной вероятностью не превзойдет ошибка выборки, но и способы определения этих пределов.

При большом числе единиц выборочной совокупности (п > 100) распределение случайных ошибок выборочной средней в соответствии с теоремой А. М. Ляпунова нормально или приближается к нормальному по мере увеличения числа наблюдений. Вероятность выхода ошибки за определенные пределы оценивается на основе таблиц интеграла Лапласа. Расчет

ошибки выборки базируется на величине генеральной дисперсии о~$^{2}_{ген}$, так как при больших п

коэффициент --, на который

для получения генеральной умножается выборочная дисперсия, большой роли не играет*.

Однако в практике статистического исследования в условиях рыночной экономики все чаще приходится сталкиваться с небольшими по объему так называемыми малыми выборками. Под **малой выборкой** понимается такое выборочное наблюдение, численность единиц которого не превышает 30. В настоящее время малая выборка используется более широко, чем раньше, прежде всего за счет статистического изучения деятельности малых и средних предприятий, коммерческих банков, фермерских хозяйств и т. д. Их количество в определенных случаях, особенно при региональных исследованиях, а также величина характеризующих их показателей (например, численность занятых) часто незначительны. Поэтому хотя общий принцип выборочного обследования (с увеличением объема выборки повышается точность выборочных данных) остается в силе, иногда приходится ограничиваться малым числом наблюдений. Наряду со статистическим изучением рыночных структур эта необходимость возникает при выборочной проверке качества продукции, в научно-исследовательской работе и в ряде других случаев.

Разработка теории малой выборки была начата английским статистиком В. С. Госсетом (печатавшимся под псевдонимом Стьюдент) в 1908 г. Он доказал, что оценка расхождения между средней малой выборки и генеральной средней имеет особый закон распределения.

При оценке результатов малой выборки величина генеральной дисперсии в расчетах не используется Для определения возможных пределов ошибки пользуются так называемым **критерием Стьюдента,** определяемым по формуле

$$t = \frac{\tilde{x} - \bar{x}}{\mu_{МВ}}, \qquad (8.26)$$

где- мера случайных колебаний выборочной средней в

$$\mu_{МВ} = \frac{\sigma}{\sqrt{n-1}}$$

малой выборке.

Величина а вычисляется на основе данных выборочного наблюдения. Она равна:

$$\sigma = \sqrt{\frac{\Sigma(x_i - \tilde{x})^2}{n}}.$$

(8.27)

Данная величина используется лишь для исследуемой совокупности, а не в качестве приближенной оценки <з в генеральной совокупности. При небольшой численности выборки распределение Стьюдента отличается от нормального: большие величины критерия имеют здесь большую вероятность, чем при нормальном распределении.

Предельная ошибка малой выборки (А$_{\text{мв}}$) в зависимости от средней ошибки (Ц$_{\text{М}}$в) представлена как

$$\Delta_{MB} = t \cdot \mu_{MB}.$$

(8.28)

Но в данном случае величина t иначе связана с вероятной оценкой, чем при большой выборке. Как указывалось ранее, английский ученый В. С. Госсет доказал, что при малой выборке действует особый закон распределения. Согласно распределению Стьюдента, вероятная оценка зависит как от величины t. так и от объема выборки в случае, если предельная ошибка не превысит t-кратную среднюю ошибку в малых выборках. Приведем выдержку из таблицы распределения Стьюдента (табл. 8.9).

Таблиц'а 8.9

Распределение вероятности в малых выборках в зависимости от коэффициента доверия t и объема выборки п*

^~-\^ п	4	5	6	7	8	9	10	15	20	со
0,5	348	356	362	366	368	370	372	376	378	383
1,0	608	626	636	644	650	654	656	666	670	683
1,5	770	792	806	816	832	828	832	846	850	865
2,0	860	884	908	908	914	920	924	936	940	954
2,5	933	946	955	959	963	966	968	975	978	988
3,0	942	960	970	976	980	938	984	992	992	977

* При п = °о в таблице даны вероятности нормального распределения. Для определения вероятности соответствующие табличные значения следует разделить на 1000.

Как видно из табл. 8.9, при увеличении п это распределение стремится к нормальному и при п = 20 уже мало от него отличается.

Покажем, как пользоваться таблицей распределения Стьюдента.

Пример. Предположим, что выборочное обследование 10 рабочих малого предприятия показало, что на выполнение одной из производственных операций рабочие затрачивали времени (мин.): 3,4; 4,7; 1,8; 3,9; 4,2; 3,9; 3,7; 3,2; 2,2; 3,9. Найдем выборочные средние затраты:

$$\tilde{x} = \frac{3,4+4,7+1,8+\ldots+2,2+3,9}{10} = 3,49 \text{ мин.}$$

Выборочная дисперсия

$$\sigma^2_{\tilde{x}} = \frac{(3,4-3,49)^2+(4,7-3,49)^2+\ldots+(3,9-3,49)^2}{10} = 0,713.$$

Отсюда средняя ошибка малой выборки

$$\mu_{MB} = \sqrt{\frac{0,713}{10-1}} = 0,28 \text{ мин.}$$

По табл. 8.9 находим, что для коэффициента доверия t = 2 и объема малой выборки п = 10 вероятность равна 0,924. Таким образом, с вероятностью 0,924 можно утверждать, что расхождение между выборкой и генеральной средней лежит в пределах от -2р. до + 2п_, т. е. разность х - х не превысит по абсолютной величине 0,56 (2 • 0,28). Следовательно, средние затраты времени во всей совокупности будут находиться в пределах от 2,93 до 4,05 мин. Вероятность того, что это предположение в действительности неверно и ошибка по случайным причинам будет по абсолютной величине больше, чем 0,56, равна: 1 - 0,924 = 0,076.

Таблица вероятностей СтьюДента часто приводится в иной форме, нежели в табл. 8.9. Считается, что в ряде случаев такая форма более удобна для практического использования (табл. 8.10).

Из табл. 8.10 следует, что для каждого числа степеней свободы k = п - 1 указана предельная величина t ($t_{\{)95}$ или $t_{()99}$), которая с данной вероятностью р не будет превышена в силу случайных колебаний результатов выборки. На основе указанной в табл. 8.10 величины t определяются **доверительные интервалы:** $x-A_{мв}$ и $x + A_{мв}$. Это область тех значений генеральной средней, выход за пределы которой имеет весьма малую вероятность, равную:

Табл и ца 8.10 **Некоторые значения t-распределения Стьюдента**

Число степеней свободы	*р			
	для одностороннего интервала		для двустороннего интервала	
	р = 0,95	р = 0,99	р = 0,95	р = 0,99
3	2,35	4,54	3,18	5,84
4	2,13	3,75	2,78	4,60
5	2,01	3,37	2,57	4,03
6	1,94	3,14	2,45	3,71
7	1,89	3,00	2,36	3,50
8	1,86	2,90	2,31	3,36
9	,83	2,82	2,26	3,35
10	,81	2,76	2,23	3,17
15	,75	2,60	2,13	2,95
20	,73	2,53	2,09	2,85
30	,70	2,46	2,04	2,75
60	,67	2,39	2,00	2,66
со	,64	2,33	1,96	2,58

Ч = 1 - Р-

В качестве доверительной вероятности при двусторонней проверке используют, как правило, р = 0,95 или р = 0,99, что не исключает, однако, выбора и других р, не приведенных в табл. 8.10.

Вероятности q случайного выхода оцениваемой средней величины за пределы доверительного интервала соответственно будут равны 0,05 и 0,01, т. е. весьма малы. Выбор между вероятностями 0,95 и 0,99 является до известной степени произвольным. Этот выбор во многом определяется содержанием тех задач, для решения которых применяется малая выборка.

В заключение отметим, что расчет ошибок в малой выборке мало отличается от аналогичных вычислений в большой выборке. Различие заключается в том, что при малой выборке вероятность нашего утверждения несколько меньше, чем при большой выборке (в частности, в приведенном ранее примере 0,924 и 0,954 соответственно). Однако все это не означает, что можно использовать малую выборку тогда, когда нужна большая выборка. Во многих случаях расхождения между найденными пределами могут достигать значительных размеров, что вряд ли удовлетворяет исследователей. Поэтому малую выборку следует применять в статистическом исследовании социально-экономических явлений с большой осторожностью, при соответствующем теоретическом и практическом обосновании.

Итак, выводы по результатам малой выборки имеют практическое значение лишь при условии, что распределение признака в генеральной совокупности является нормальным или асимптотически нормальным. Необходимо также принимать во внимание и то, что точность результатов выборки малого объема все же ниже, чем при большой выборке.

8.6

ОБЛАСТИ ПРИМЕНЕНИЯ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ В ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ ИССЛЕДОВАНИЯХ

Переход к рыночной экономике в значительной мере способствует расширению сферы использования выборочного наблюдения.

Проблемы применения конкретных видов выборочного наблюдения для решения тех или иных теоретических или прикладных задач решаются с учетом их специфики.

Выборочное наблюдение широко используется для: \) статистического оценивания и проверки гипотез; 2) решения производственных и управленческих задач; 3) отраслевых социально-экономических исследований; 4) разрешения задач в сфере предпринимательской деятельности.

Первая группа задач чаще всего связана с решением общетеоретических проблем, проведением исследований и экспериментов для получения информации о генеральной совокупности на основе выборочного наблюдения. Такие исследования могут решать два основных вида задач:

• поиск наилучших выборочных параметров (оценок) для отображения интересующих нас свойств генеральной совокупности (например, выбор в качестве лучшей оценки — средней, моды^ медианы или доли). Решения этих вопросов составляют суть теории статистического оценивания;

• выдвижение и формирование определенных гипотез о тех или иных свойствах генеральной совокупности и их последующая проверка с помощью результатов выборочного наблюдения. Изучением этих задач занимается теория проверки статистических гипотез.

Вторая группа задач (производственные и управленческие) связана с практическими интересами и приобретает все большее значение в области управления технологическими процессами, качеством продукции и работ. К основным этапам статистического управления качеством относятся: а) измерение параметров и создание системы показателей качества, контролируемых в производственном процессе; б) установление номинального (производственного, технологического) режима, отклонение от которого должно статистически оцениваться и иметь следствием принятие определенных решений; в) поиск оптимального режима, способов совершенствования процесса, альтернативных технологий на базе анализа производимых замеров; г) управление по номиналу и допускам. Оно чаще всего осуществляется на основе так называемых контрольных карт Шухарта или кумулятивных сумм, разрабатываемых на основе стандартов, специальных таблиц и номограмм.

Общая идея реализации перечисленных этапов состоит в проведении выборочных наблюдений на каждом этапе и анализе полученных результатов для принятия управленческих решений.

Выборочный приемочный и текущий контроль качества продукции подчас выступает единственно возможным, так как в ряде случаев подразумевает утрату потребительских свойств определенных видов продукции (например, при контроле продолжительности горения или устойчивости на перепады напряжения электрических лампочек).

Задачи отраслевых социально-экономических исследований, проводимых с использованием выборочного наблюдения, чаще всего решаются с помощью системы органов отраслевого управления и государственной статистики. В промышленности это, к примеру, изучение использования оборудования, рабочего времени, эффективности новых технологий; в сельскохозяйственном секторе - анализ продуктивности скота, урожайности, качества кормов; в торговле - выборочные исследования спроса на отдельные товары и степени его удовлетворения.

Большой опыт накоплен отечественной и зарубежной статистикой в области выборочных обследований населения. Выборочное наблюдение позволяет значительно расширить программы переписей населения, используется для получения предварительных итогов переписей, при контроле качества заполнения переписных листов и других контрольных мероприятиях. В отечественной практике широкое распространение получают национальные выборочные социально-демографические обследования (1985 и 1994 гг.), которые позволяют получить ряд важнейших социально-экономических характеристик в межпереписные периоды.

Широка область применения выборочного наблюдения в социальной статистике, в частности в изучении доходов, потребления материальных благ и услуг, жилищных условий и других характеристик уровня жизни населения. Главным источником информации об уровне жизни стали выборочные обследования бюджетов семей, позволяющие получить показатели занятости, размера доходов различных социальных групп населения, источников их формирования. Результаты бюджетных обследований дают возможность оценки дифференциации потребления продовольственных и непродовольственных товаров, одежды, обуви, мебели, предметов культурно-бытового назначения и других социальных характеристик (образовательного, профессионального статуса и др.).

Важнейшую социальную информацию дают выборочные обследования бюджета времени населения, являющегося зеркалом жизнедеятельности общества и представляющего многие параметры его образа и стиля жизни - культурный, образовательный, материальный уровень, характер использования рабочего и внерабочего времени, досуга и др. Все большую актуальность приобретают выборочные исследования перемещений, внутри- и межпоколенной мобильности, социальной стратификации, миграционных потоков, заболеваемости, контроля над рождаемостью и др. Расширяется область использования выборочных оп\росоВ, проводимых по самой широкой проблематике. К примеру, базы данных известных служб Гэллапа и Харриса позволяют дать обширную информацию о многих сторонах экономической, политической и социальной жизни США за весьма длительный (30 и более лет) период времени. Широко используются выборочные опросы, проводимые различными общественными и политическими организациями с целью анализа политической ситуации, настроений, популярности лидеров различных политических сил. Следует отметить, однако, что выборочные наблюдения в их классической форме здесь могут соседствовать с несобственно случайными методами формирования совокупностей - квотной выборкой (метод квот), стихийной выборкой (метод «первого встречного»), «снежным комом» и т. п.

Быстроразвивающейся в России областью использования выборочного наблюдения следует признать сферу **коммерции и бизнеса.** Развитие этой области обусловлено недостаточностью объемов и качества официальной информации (переписей предприятий, данных текущей торговой статистики и др.) для прогнозирования объемов производства и продаж, необходимой для предпринимателей. Это порождает потребность в систематическом получении дополнительных сведений о рынке, товарах и потребителях.

Для выяснения потребительских реакций на новые товары широко используется выборочное анкетирование, которое решает задачи анализа отношения потребителя к данному товару (внешний вид, калорийность, благоприятность отношения к данному товару со стороны определенных половозрастных групп, цена и т. д.). Широко используются выборочные наблюдения для исследования сегментации рынка, позиционирования товаров, изучения потребителей рекламной информации и в других областях предпринимательской деятельности.

Совершенствование теории и практики выборочного наблюдения, все более широкое применение различных сочетаний комбинированного, многоступенчатого отбора, современных компьютерных технологий информационной обработки в значительной мере расширяют области использования, скорость получения и качество результатов выборочного наблюдения.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИИ ГЛАВЫ 8

Выборочное наблюдение - несплошное наблюдение, при котором статистическому обследованию подвергаются единицы изучаемой совокупности, отобранные случайным способом.

Выборочная совокупность - совокупность отобранных для обследования единиц.

Генеральная совокупность — совокупность единиц, из которых производится отбор.

Ошибка выборочного наблюдения - разность между величиной параметра в генеральной совокупности и его величиной, вычисленной по результатам выборочного наблюдения.

Индивидуальный отбор предполагает формирование выборочной совокупности на основе отбора отдельных единиц генеральной совокупности.

Групповой отбор предполагает формирование выборочной совокупности на основе отбора групп единиц из генеральной совокупности.

Комбинированный отбор представляет собой сочетание индивидуального и группового отбора.

Виды выборки определяют конкретный механизм или процедуру отбора единиц из генеральной совокупности. В практике выборочных обследований наибольшее распространение получили следующие виды выборки: собственно-случайная, механическая, типическая, серийная, комбинированная.

Многоступенчатый отбор - 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 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.

Выборочная доля - удельный вес единиц, обладающих данным признаком в выборочной совокупности. Различия между выборочной долей и средним значением признака в выборке (выборочной средней) определяют особенности вычисления необходимого объема, ошибок выборки, доверительных интервалов и др.

Прямой пересчет и способ коэффициентов - методы распространения результатов выборочного наблюдения на генеральную совокупность.

Малая выборка - выборочное наблюдение, численность единиц которого не превышает 30. При малой выборке действует особый закон распределения. Величина вероятной ошибки зависит как от коэффициента доверия t, так и от объема выборки в случае, если предельная ошибка не превысит t-кратную среднюю ошибку в малых выборках.

ТЕСТ К ГЛАВЕ 8

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[1] 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 типический 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 на:

а) 36%;
б) 64%;
в) 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 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)-;
б) 1; 2;
в) 1; '
г) 2.

7. По данным 10%-ного 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 на:

а) 10%;
б) 19%;
в) 1%;
; г) предсказать результат невозможно.

11. По выборочным данным (2%-ный отбор), удельный вес неуспевающих студентов на IV курсе составил 10%, на III курсе — 15%. При одинаковой численности выборочной совокупности ошибка выборки больше:

а) на IV курсе; . б) на III курсе;
в) ошибки равны;
г) данные не позволяют сделать вывод.

ГЛАВА 9

СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

9.1

ПРИЧИННОСТЬ, РЕГРЕССИЯ, КОРРЕЛЯЦИЯ

Исследование объективно существующих связей между явлениями - важнейшая задача общей теории статистики. В процессе статистического исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие существенное влияние на вариацию изучаемых явлений и процессов. **Причинно-следственные отношения** - это связь явлений и процессов, когда изменение одного из них - причины - ведет к изменению другого - следствия.

Причина - это совокупность условий, обстоятельств, действие которых приводит к появлению следствия. Если между явлениями действительно существуют причинно-следственные отношения, то эти условия должны обязательно реализовываться вместе с действием причин. Причинные связи носят всеобщий и многообразный характер, и для обнаружения причинно-следственных связей необходимо отбирать отдельные явления и изучать их изолированно.

Особое значение при исследовании причинно-следственных связей имеет выявление временной последовательности: причина всегда должна предшествовать следствию, однако не каждое предшествующее событие следует считать причиной, а последующее следствием.

В реальной социально-экономической действительности причину и следствие необходимо рассматривать как смежные явления, появление которых обусловлено комплексом сопутствующих более простых причин и следствий. Между сложными группами причин и следствий возможны многозначные связи, когда за одной причиной будет следовать то одно, то другое действие или одно действие имеет несколько различных причин. Чтобы установить однозначную причинную связь между явлениями или предсказать возможные следствия конкретной причины, необходима полная абстракция от всех прочих явлений в исследуемой временной или пространственной среде. Теоретически такая абстракция воспроизводится. Приемы абстракции часто применяются при изучении взаимосвязей между двумя признаками (парной корреляции). Но чем сложнее изучаемые явления, тем труднее выявить причинно-следственные связи между ними. Взаимное переплетение различных внутренних и внешних факторов неизбежно приводит к некоторым ошибкам в определении причины и следствия.

Особенностью причинно-следственных связей в социальноэкономических явлениях является их транзитивность, т. е. причина X и следствие У связаны соотношением X—>X'—»X"—»У, а не непосредственно X—>У. Однако промежуточные факторы, как правило, при анализе опускаются.

Так, например, используя показатели международной методологии расчетов, фактором валовой прибыли (У) является валовое накопление основных и оборотных фондов (X), но при этом допускаются такие факторы, как валовой выпуск (X'), оплата труда (X") и т. д. Правильно вскрытые причинно-следственные связи позволяют установить силу воздействия отдельных факторов на результаты хозяйственной деятельности.

Социально-экономические явления представляют собой результат одновременного воздействия большого числа причин. Следовательно, при изучении этих явлений необходимо выявлять главные, основные причины, абстрагируясь от второстепенных.

В основе **первого этапа** статистического изучения связи лежит качественный анализ изучаемого явления, связанный с анализом природы социального или экономического явления методами экономической теории, социологии, конкретной экономики. **Второй этап** - построение модели связи. Он базируется на методах статистики: группировках, средних величинах, таблицах и т. д. **Третий, последний этап** - интерпретация результатов вновь связан с качественными особенностями изучаемого явления.

Статистика разработала множество методов изучения связей, выбор которых зависит от целей исследования и от поставленных задач. Связи между признаками и явлениями, ввиду их большого разнообразия, классифицируются по ряду оснований. Признаки по их значению для изучения взаимосвязи делятся на два класса. Признаки, обуславливающие изменения других, связанных с ними признаков, называются **факторными,** или просто **факторами.** Признаки, изменяющиеся под действием факторных признаков, являются **результативными.** Связи между явлениями и их признаками классифицируются по степени тесноты связи, направлению и аналитическому выражению.

В статистике различают функциональную связь и стохастическую зависимость. **Функциональной** называют такую связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака. Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы исследуемой совокупности.

Если причинная зависимость проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений, то такая зависимость называется **стохастической.** Частным случаем стохастической связи является **корреляционная** связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

По степени тесноты связи различают количественные критерии оценки тесноты связи (табл. 9.1).

Таблица 9.1 Количественные критерии оценки тесноты связи

Величина коэффициента корреляции					Характер связи
	До ±0,3	\ ±	0,3| ±0,5		Практически отсутствует Слабая
	±0,5 ±0,7	-	±0,7 ±1,0		Умеренная Сильная

По направлению выделяют связь **прямую и обратную. При** прямой связи с увеличением или уменьшением значений факторного признака происходит увеличение или уменьшение значений результативного. Так, например, рост производительности труда способствует увеличению уровня рентабельности производства. В случае обратной связи значения результативного признака изменяются под воздействием факторного, но в противоположном направлении по сравнению с изменением факторного признака. Так, с увеличением уровня фондоотдачи снижается себестоимость единицы производимой продукции.

По аналитическому выражению выделяют связи **прямолинейные** (или просто **линейные) и нелинейные.** Если статистическая связь между явлениями может быть приближенно выражена уравнением прямой линии, то ее называют **линейной** связью; если же она выражается уравнением какой-либо кривой линии (параболы, гиперболы, степенной, показательной, экспоненциальной и т. д.), то такую связь называют **нелинейной, или криволинейной.**

В статистике не всегда требуются количественные оценки связи, часто важно определить лишь ее направление и характер, выявить форму воздействия одних факторов на другие. Для выявления наличия связи, ее характера и направления в статистике используются методы: приведения параллельных данных; аналитических группировок; графический; корреляции.

Метод приведения параллельных данных основан на сопоставлении двух или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие связи и получить представление о ее характере. Сравним изменения двух величин X и У:

С увеличением величины X величина У также возрастает. Поэтому связь между ними прямая, и описать ее можно или уравнением прямой, или уравнением параболы второго порядка.

X	1	2	→j	4	5	6	' 7	8	9
У	5	6	9	10	14	17	15	20	23

Графически взаимосвязь двух признаков изображается с помощью поля корреляции. В системе координат на оси абсцисс откладываются значения факторного признака, а на оси ординат — результативного. Каждое пересечение линий, проводимых через эти оси, обозначается точкой. При отсутствии тесных связей имеет место беспорядочное расположение точек на графике. Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определенной линии, выражающей форму связи (рис.9.1).

Рис. 9.1. График корреляционного поля

Для социально-экономических явлений характерно, что наряду с существенными факторами, формирующими уровень результативного признака на него оказывают воздействие многие другие неучтенные и случайные факторы. Это свидетельствует о том, что взаимосвязи явлений, которые изучает статистика, носят корреляционный характер и аналитически выражаются функцией вида $y_x = f(x)$.

Корреляция - это статистическая зависимость между случайными величинами, не имеющими строго функционального характера, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

В статистике принято различать следующие варианты зависимостей.

\. Парная корреляция - связь между двумя признаками (результативным и факторным или двумя факторными).

2. Частная корреляция - зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков.

3. Множественная корреляция - зависимость результативного и двух или более факторных признаков, включенных в исследование.

Корреляционный анализ имеет своей задачей количественное определение тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

Теснота связи количественно выражается величиной коэффициентов корреляции. Коэффициенты корреляции, представляя количественную характеристику тесноты связи между признаками, дают возможность определять «полезность» факторных признаков при построении уравнений множественной регрессии. Величина коэффициента корреляции служит также оценкой соответствия уравнения регрессии выявленным причинно-следственным связям.

Первоначально исследования корреляции проводились в биологии, а позднее распространились и на другие области, в том числе на социально-экономическую. Одновременно с корреляцией начала использоваться и регрессия. Корреляция и регрессия тесно связаны между собой: первая оценивает силу (тесноту) статистической связи, вторая исследует ее форму. Та и другая служат для установления соотношения между явлениями, для определения наличия или отсутствия связи.

Корреляционно-регрессионный анализ как общее понятие включает в себя измерение тесноты, направления связи и установление аналитического выражения (формы) связи (регрессионный анализ).

Регрессионный анализ заключается в определении аналитического выражения связи, в котором изменение одной величины (называемой зависимой или результативным признаком) обусловлено влиянием одной или нескольких независимых величин (факторов), а множество всех прочих факторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения. Регрессия может быть однофакторной (парной) и многофакторной (множественной).

По форме зависимости различают:

а) линейную регрессию, которая выражается уравнением прямой (линейной функцией) вида: $У_x$ = a_0^ a,x;

б) нелинейную регрессию, которая выражается уравнениями вида:

парабола – $\overline{У}_x = a_0 + a_1x + a_2x^2$;

гипербола – $\overline{У}_x = a_0 + \frac{a_1}{x}$ и т. д.

По направлению связи различают:

а) прямую регрессию (положительную), возникающую при условии, если с увеличением или уменьшением независимой величины значения зависимой также соответственно увеличиваются или уменьшаются;

б) обратную (отрицательную) регрессию, появляющуюся при условии, что с увеличением или уменьшением независимой величины зависимая соответственно уменьшается или увеличивается.

Положительную и отрицательную регрессии можно легче понять, если использовать их графическое изображение (рис. 9.2, 9.3).

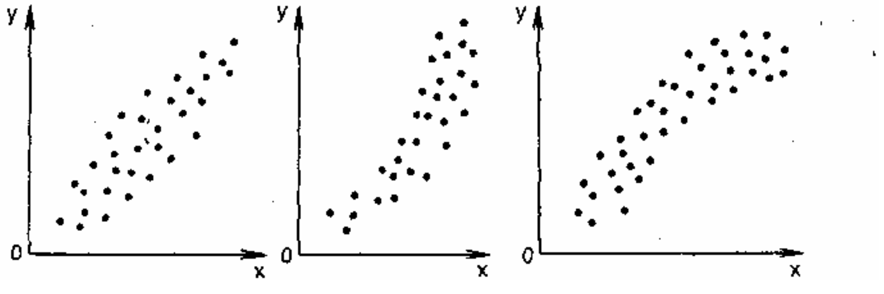


Рис. 9.2. Прямая (положительная) регрессия

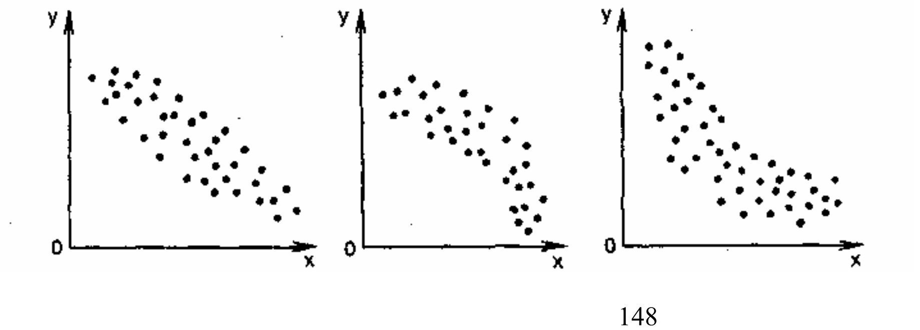


Рис. 9.3. Обратная (отрицательная) регрессия

Для простой (парной) регрессии в условиях, когда достаточно полно установлены причинно-следственные связи, приобретает практический смысл только последнее положение; при множественности причинных связей невозможно четко разграничить одни причинные явления от других.

9.2

ОСНОВНЫЕ ЗАДАЧИ И ПРЕДПОСЫЛКИ ПРИМЕНЕНИЯ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА

Все явления и процессы, характеризующие социально-экономическое развитие и составляющие единую систему национальных счетов, тесно взаимосвязаны и взаимозависимы между собой.

В статистике показатели, характеризующие эти явления, могут быть связаны либо корреляционной зависимостью, либо быть независимыми (табл. 9.1).

Корреляционная зависимость является частным случаем стохастической зависимости, при которой изменение значений факторных признаков (X|, x_2, ..., x_k) влечет за собой изменение среднего значения результативного признака.

Корреляционная зависимость исследуется с помощью методов корреляционного и регрессионного анализов.

Корреляционный анализ изучает взаимосвязи показателей и позволяет решить следующие задачи:

\. Оценка тесноты связи между показателями с помощью парных, частных и множественных коэффициентов корреляции (п. 9.6).

2. Оценка уравнения регрессии.

Основной предпосылкой применения корреляционного анализа является необходимость подчинения совокупности значений всех факторных (X|, x_2, ..., x_k) и результативного (У) признаков k-мерному нормальному закону распределения или близость к нему. Если объем исследуемой совокупности достаточно большой (п > 50), то нормальность распределения может быть подтверждена на основе расчета и анализа критериев Пирсона, Ястремского, Боярского, Колмогорова, чисел Вастергагада и т. д. Если п < 50, то закон распределения исходных данных определяется на базе построения и визуального анализа поля корреляции. При этом если в расположении точек имеет место линейная тенденция, то можно предположить, что совокупность исходных данных (У, x_h x_2, ..., x_k) подчиняется нормальному распределению.

Целью регрессионного анализа является оценка функциональной зависимости условного среднего значения результативного признака (У) от факторных (х,, x_2, ..., x_k).

Основной предпосылкой регрессионного анализа является то, что только результативный признак (У) подчиняется нормальному закону распределения, а факторные признаки х,, x_2, ... x_k могут иметь произвольный закон распределения. В анализе динамических рядов в качестве факторного признака выступает время t. При этом в регрессионном анализе заранее подразумевается наличие причинно-следственных связей между результативным (У) и факторными (х,, x_2, ..., x_k) признаками.

Уравнение регрессии, или статистическая модель связи социально-экономических явлений, выражаемая функцией

$$\bar{У}_x = f(x_1, x_2, ..., x_k),$$

является достаточно адекватным реальному моделируемому явлению или процессу в случае соблюдения следующих **требований их построения.**

1. Совокупность исследуемых исходных данных должна быть однородной и математически описываться непрерывными функциями.
2. Возможность описания моделируемого явления одним или> несколькими уравнениями причинно-следственных связей.
3. Все факторные признаки должны иметь количественное (цифровое) выражение.
4. Наличие достаточно большого объема исследуемой выборочной совокупности.
5. Причинно-следственные связи между явлениями и процессами следует описывать линейной или приводимой к линейной формами зависимости.
6. Отсутствие количественных ограничений на параметры модели связи.
7. Постоянство территориальной и временной структуры изучаемой совокупности.

Соблюдение данных требований позволяет исследователю построить статистическую модель связи, наилучшим образом аппроксимирующую моделируемые социально-экономические явления и процессы.

Теоретическая обоснованность моделей взаимосвязи, построенных на основе корреляционно-регрессионного анализа, обеспечивается соблюдением следующих **основных условий.**

1. Все признаки и их совместные распределения должны подчиняться нормальному закону распределения.
2. Дисперсия моделируемого признака (У) должна все время оставаться постоянной при изменении величины (У) и значений факторных признаков.
3. Отдельные наблюдения должны быть независимыми, т. е. результаты, полученные в i-м наблюдении, не должны быть связаны с предыдущими и содержать информацию о последующих наблюдениях,'а также влиять на них.

Отступление от выполнения этих условий и предпосылок приводит к тому, что модель регрессии будет неадекватно отражать реально существующие связи между анализируемыми признаками.

Одной из проблем построения уравнения регрессии является ее **размерность,** т. е. определение числа факторных признаков, включаемых в модель. Их число должно быть оптимальным.

Сокращение размерности за счет исключения второстепенных, несущественных факторов позволяет получить модель, быстрее и качественнее реализуемую. В то же время построение модели малой размерности может привести к тому, что она будет недостаточно полно описывать исследуемое явление или процесс в единой системе национального счетоводства.

Практика выработала определенный критерий, позволяющий установить оптимальное соотношение между числом факторных признаков, включаемых в модель, и объемом исследуемой совокупности. Согласно данному критерию число факторных признаков (k) должно быть в 5 — 6 раз меньше объема изучаемой совокупности.

Построение корреляционно-регрессионных моделей, какими бы сложными они ни были, само по себе не вскрывает полностью всех причинно-.следственных связей. Основой их адекватности является предварительный качественный анализ, основанный на учете специфики и особенностей сущности исследуемых социально-экономических явлений и процессов.

9.3

ПАРНАЯ РЕГРЕССИЯ НА ОСНОВЕ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ И МЕТОДА ГРУППИРОВОК

Парная регрессия характеризует связь между двумя признаками: результативным и факторным. Аналитическая связь между ними описывается уравнениями:

- **прямой** – $\bar{У}_x = a_0 + a_1 x$;
- **гиперболы** – $\bar{У}_x = a_0 + a_1/x$; (9-1)
- **параболы** – $\bar{У}_x = a_0 + a_1 x + a_2 x^2$ и т.д.

Определить тип уравнения можно, исследуя зависимость графически. Однако существуют более общие указания, позволяющие выявить уравнение связи, не прибегая к графическому изображению. Если результативный и факторный признаки возрастают одинаково, примерно в арифметической прогрессии, то это свидетельствует о том, что связь между ними линейная, а при обратной связи - гиперболическая. Если факторный признак увеличивается в арифметической прогрессии, а результативный - значительно быстрее, то используется параболическая), или степенная регрессия.

Оценка параметров уравнений регрессии (a_0, aj и a_2 - в уравнении параболы второго порядка) осуществляется методом наименьших квадратов, в основе которого лежит предположение о независимости наблюдений исследуемой совокупности.

Основной принцип **метода наименьших квадратов** рассмотрим на следующем примере: будем считать, что две величины (два показателя) X и У взаимосвязаны между собой, причем У находится в некоторой зависимости от X. Следовательно, У будет зависимой, а X - независимой величинами.

Сущность метода наименьших квадратов заключается в нахождении параметров модели (a_0, a,), при которых минимизируется сумма квадратов отклонений эмпирических (фактических) значений результативного признака от теоретических, полученных по выбранному уравнению регрессии:

$$S = \Sigma(У - \bar{У}_x)^2 \rightarrow \min.$$

Для прямой зависимости:

$$S = \Sigma(y - a_0 - a_1x)^2 \rightarrow \min.$$

Рассматривая S в качестве функции параметров a_0 и a, и проводя математические преобразования (дифференцирование), получаем:

$$\begin{cases} \dfrac{dS}{da_0} = \Sigma 2(a_0 + a_1x - y) = 0; \\ \dfrac{dS}{da_1} = \Sigma 2(a_0 + a_1x - y)x = 0. \end{cases}$$

Откуда система нормальных уравнений для нахождения параметров линейной парной регрессии методом наименьших квадратов имеет следующий вид:

$$\begin{cases} na_0 + a_1\Sigma x = \Sigma y; \\ a_0\Sigma x + a_1\Sigma x^2 = \Sigma xy, \end{cases} \qquad (9.2)$$

где п - объем исследуемой совокупности (число единиц наблюдений).

В уравнениях регрессии параметр a_0 показывает усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов; параметр *a_1* (а в уравнении параболы и a_2) - коэффициент регрессии показывает, насколько изменяется в среднем значение результативного признака при увеличении факторного на единицу собственного измерения.

Например, имеются данные, характеризующие деловую активность акционерных обществ закрытого типа (АОЗТ): прибыль (тыс. руб.) и затраты на 1 руб. произведенной продукции (коп.) (табл. 9.2).

Предположим наличие линейной зависимости между рассматриваемыми признаками.

Таблица 9.2

Расчет сумм для определения параметров парного линейного уравнения регрессии (данные условные)

№	Затраты на 1 руб. произведенно продукции. коп., Х	Прибыль, тыс. руб., У	X^2	ХУ	у*
1	77	1070	5929	82390	1 016
2	77	1001	5929	77077	1 016
3	81	789	5561	63 909	853
4	82	779	6724	63 878	812
5	89	606	7921	53934	527
6	96	221	9216	21 216	242
Итого	502	4466	42280	362 404	4466

Система нормальных уравнений для данного примера имеет вид:

$$\begin{cases} na_0 + a_1\Sigma x = \Sigma y; \\ a_0\Sigma x + a_1\Sigma x^2 = \Sigma xy. \end{cases}$$

$$\begin{cases} 6a_0 + 502a_1 = 4466; \\ 502a_0 + 42280a_1 = 362404. \end{cases}$$

Отсюда: a_0 = 4153,88; а, = - 40,75.

Следовательно, $У_x$ = 4153,88 - 40,75x.

На практике часто исследования проводятся по большому числу наблюдений. В этом случае исходные данные удобнее представлять в сводной корреляционной таблице. При этом анализу подвергаются сгруппированные данные и по факторному X, и по результативному У признакам, т. е. уравнение парной регрессии целесообразно строить на основе **сгруппированных данных.**

Если значения признаков X и У заданы в определенных интервалах (а - в), то для каждого интервала сначала определяют

середину интервалаа затем уже коррелируют зна-

$$\left(\frac{x'}{y'} = \frac{a+B}{2}\right),$$

чения x' и y' и строят уравнения регрессии между ними.

Например, определим зависимость между величиной капитала и величиной работающих активов коммерческих банков РФ на 01.01.98 г. (табл. 9.3).

Таблица 9.3

Распределение коммерческих банков РФ по величине капитала и величине работающих активов на 01.01.98 г. (данные условные)

Величина капитала, тыс. руб., У	Величина работающих активов, тыс. руб., X					Число коммерческих банков, f_y	\Tf __c	V"\ Tf ПУТ
		14-	70-126	126-182	182-238			
		42	98	154	210			
745-2684	1714,5	4	6	2	3	15	25717,5	2904363
2684-4624	3654,0	1	3	-	-	4	14616,0	1 227 744
4624-6564	5594,0	—	1	1	-	2	11 188,0	1409688
6564-8503	7533,5	1	1	2	-	4	30134,0	3 375 008
85,03-125842	67172,5	2	—	1	2	5	335862,5	44199505
Число предприятий f.	_	8	11	6	5	30	417518,	53 116308
xfx	-	336	1 078	924	1 050	3388		
X^2f_x	-	14112	105644	142296	220 500	482 552		

Предположим наличие линейной зависимости между рассматриваемыми признаками.

Система нормальных уравнений для определения коэффициентов уравнения регрессии примет вид:

$$\begin{cases} na_0 + a_1\Sigma xf_x = \Sigma yf_y; \\ a_0\Sigma xf_x + a_1\Sigma x^2 f_x = \Sigma xyf_y, \end{cases} \quad (9.3)$$

где *n* =30 - число анализируемых коммерческих банков;

f_x; f_v - число банков согласно распределению соответственно по факторному и результативному признакам;

$Уf_y$; Xf_x - значение результативного и факторного признаков по конкретной группе коммерческих банков.

Так, для первой группы:

$$Уf_y = 1\ 714,5 \cdot 15 = 25\ 717,5;$$

$$Xf_x = 42 \cdot 8 = 336;$$

$$XУf_y = 1\ 714,5 \cdot 4 \cdot 42 + 1\ 714,5 \cdot 6 \cdot 98 + 1\ 714,5 \cdot 2 \cdot 154 + 1\ 714,5 \cdot 3 \cdot 210 = 2\ 904\ 363;$$

$$X^2f_x = 42 \cdot 42 \cdot 8 = 14\ 112.$$

Таким образом, подставив в систему суммарные значения, получим:

$$\begin{cases} 30a_0 + 3\ 388a_1 = 417\ 518; \\ 3\ 388a_0 + 482\ 552a_1 = 53\ 116\ 308; \end{cases}$$

$$a_0 = 7177,6; \qquad a_1 = 59,7.$$

Отсюда: $\bar{У}_x = 7\ 177,6 + 59,7x.$

Если связь между признаками У и X криволинейная и описывается уравнением параболы второго порядка, то

$$\bar{У}_x = a_0 + a_1x + a_2x^2.$$

В данном случае задача сводится к определению неизвестных параметров: a_0, а,, a_2.

Значения величин X и У представлены двумя рядами данных:

$$\begin{matrix} У_1 & У_2 & У_3 & \dots & У_n; \\ X_1 & X_2 & X_3 & \dots & X_n. \end{matrix}$$

Если бы все значения, полученные по данным наблюдения, лежали строго на кривой, описываемой уравнением параболы, или для каждой из точек было бы справедливо равенство

$$У_i - a_0 - a_1x_{1i} - a_2x_{1i}^2 = 0,$$

то не существовало бы никаких проблем. Однако на практике имеем другое:

$$У_i - a_0 - a_1x_{1i} - a_2x_{1i}^2 = \Delta_i,$$

где А, - разность между данными наблюдения и данными, полученными по уравнению связи.

Эта разность как раз и появляется в силу наличия ошибок упрощения, поэтому возникает проблема нахождения таких коэффициентов уравнения (регрессии), при которых ошибка была бы минимальной. Можно минимизировать сумму абсолютных отклонений (ошибок), т. е.

$$S = \sum_{i=1}^{n} |\Delta_i| \to \min,$$

или минимизировать сумму кубических ошибок, и тогда получим метод наименьших кубов:

$S=\sum_{i=1}^{n}|\Delta_i^3| \rightarrow \min$ или, наконец, минимизировать наибольшую абсолютную

ошибку:

$$\min \max_i |\Delta_i|.$$

Однако наиболее оптимальным вариантом является оценка ошибки по методу наименьших квадратов:

$$S=\sum_{i=1}^{n}\Delta_i^2 \rightarrow \min.$$

Метод наименьших квадратов обладает тем замечательным свойством, что делает число нормальных уравнений равным числу неизвестных коэффициентов. Приведенное уравнение параболы второго порядка имеет три неизвестных коэффициента: *a^*, a,, a₂.
Следовательно, применяя метод наименьших квадратов, мы получим уравнение:

$$S=\sum_{i=1}^{n}\Delta_i^2=\sum_{i=1}^{n}\left(y-a_0-a_1x-a_2x^2\right)^2 \rightarrow \min.$$

Для нахождения значений неизвестных коэффициентов a(), a,, a₂, при которых функция f(a₀, a,, a₂) была бы минимальной, необходимо приравнять частные производные по этим величинам к нулю, т. е.

$$\begin{cases}\dfrac{dS}{da_0}=2\Sigma\left(y-a_0-a_1x-a_2x^2\right)=0;\\ \dfrac{dS}{da_1}=2\Sigma\left(y-a_0-a_1x-a_2x^2\right)\cdot x=0;\\ \dfrac{dS}{da_2}=2\Sigma\left(y-a_0-a_1x-a_2x^2\right)\cdot x^2=0.\end{cases}$$

Проделав простейшие преобразования, получим систему нормальных уравнений:

$$\begin{cases}na_0+a_1\Sigma x+a_2\Sigma x^2=\Sigma y;\\ a_0\Sigma x+a_1\Sigma x^2+a_2\Sigma x^3=\Sigma yx;\\ a_0\Sigma x^2+a_1\Sigma x^3+a_2\Sigma x^4=\Sigma yx^2.\end{cases}$$

(9.4)

Решив систему, найдем значения неизвестных коэффициентов a{), a,, a₂ и получим уравнение регрессии^. Вычислим по уравнению регрессии теоретические значения $У_x$ и сравним с данными наблюдения, т. е. рассчитаем так называемую остаточную сумму квадратов (табл. 9.4).

Остаточная сумма квадратов совпадает с минимальной возможной величиной по методу наименьших квадратов.

Оценка обратной зависимости между X и У, когда с увеличением (уменьшением) X уменьшается (увеличивается) значение результативного признака У, может быть осуществлена на основе уравнения гиперболы:

$$\overline{У}_x=a_0+\frac{a_1}{x}.$$

Таблица 9.4

Расчет остаточной суммы квадратов

Номер наблюдения	Значения по данным наблюдения	Значения по данным уравнения регрессии Ух	$А = У|\sim У_Х$	$А^2$;
1	YI	У|	А,	АГ
2	$У_2$	Уа	$А_2$	$А_2^2$
3	Уз	Уз	$А_3$	$Аз^2$
	•			
•			•	•
	У,	У,	А	$А^2$*
	•	•	•	
	.	•		
•				•
п	$У_п$	У,	А,	$А,^2$
* iX(i = l,2,...,n). i=l				

Систему нормальных уравнений для нахождения параметров гиперболы можно представить следующим образом:

$$\begin{cases} na_0 + a_1\Sigma\frac{1}{x} = \Sigma y; \\ a_0\Sigma\frac{1}{x} + a_1\Sigma\frac{1}{x^2} = \Sigma\frac{y}{x}. \end{cases}$$

(9.5)

Применение метода наименьших квадратов объясняется неизбежным наличием случайных ошибок в результатах опыта.

Статистические данные обладают ошибками упрощения, которые возникают как следствие:

- неполноты охвата, потому что часть единиц совокупности, полученных в результате наблюдения, не может быть использована в исследовании;
- неполноты факторов, определяющих то или иное социально-экономическое явление, в силу того, что ни в одно уравнение, или модель, нельзя включить бесконечное число аргументов (во всех случаях отбирается только часть воздействующих факторов, причем отбор носит чисто субъективный характер);
- характера выбранного уравнения связи. Как бы хорошо оно ни было обосновано, как бы теоретически адекватно ни описывало исследуемое явление, оно не может быть его точным аналогом.

Решение вопроса о возможности использования метода наименьших квадратов для изучения связей между социально-экономическими явлениями зависит от свойства оценок, получаемых с помощью этого метода.

Даже при сравнительно небольшом числе наблюдений применение метода наименьших квадратов позволяет получить достоверные оценки.

Метод наименьших квадратов также может быть использован в случаях проведения анализа косвенных наблюдений, являющихся функциями многих неизвестных.

9.4

МНОЖЕСТВЕННАЯ (МНОГОФАКТОРНАЯ) РЕГРЕССИЯ

Изучение связи между тремя и более связанными между собой признаками носит название **множественной (многофакторной) регрессии.** При исследовании зависимостей методами множественной регрессии задача формулируется так же, как и при использовании парной регрессии, т. е. требуется определить аналитическое выражение связи между результативным признаком (У) и факторными признаками $(x_1, x_2, x_3, \dots x_k)$, найти функ-
цию:

$$\bar{y}_{1,2,\dots,k} = f(x_1, x_2, \dots, x_k). \quad (9.6)$$

Построение моделей множественной регрессии включает несколько этапов:

- выбор формы связи (уравнения регрессии);
- отбор факторных признаков;
- обеспечение достаточного объема совокупности для получения несмещенных оценок.

Рассмотрим подробнее каждый из них.

Выбор формы связи затрудняется тем, что, используя математический аппарат, теоретически зависимость между признаками может быть выражена большим числом различных функций.

Выбор типа уравнения осложнен тем, что для любой формы зависимости выбирается целый ряд уравнений, которые в определенной степени будут описывать эти связи. Некоторые предпосылки для выбора определенного уравнения регрессии получают на основе анализа предшествующих аналогичных исследований или на базе анализа подобных работ в смежных отраслях знаний. Поскольку уравнение регрессии строится главным образом для объяснения и количественного выражения взаимосвязей, оно должно хорошо отражать сложившиеся между исследуемыми факторами фактические связи.

Наиболее приемлемым способом определения вида исходного уравнения регрессии является **метод перебора различных уравнений.**

Сущность данного метода заключается в том, что большое число уравнений (моделей) регрессии, отобранных для описания связей какого-либо социально-экономического явления или процесса, реализуется на ЭВМ с помощью специально разработанного алгоритма перебора с последующей статистической проверкой, главным образом, на основе t-критерия Стьюдента и F-критерия Фишера-Снедекора.

Способ перебора является достаточно трудоемким и связан с большим объемом вычислительных работ.

Практика построения многофакторных моделей взаимосвязи показывает, что все реально существующие зависимости между социально-экономическими явлениями можно описать, используя **пять типов моделей:**

1) линейная:

$$\bar{y}_{1,2,\dots,k} = a_0 + a_1x_1 + a_2x_2 + \dots a_kx_k;$$

2) степенная:

$$\bar{y}_{1,2,\dots,k} = a_0x_1^{a_1} \cdot x_2^{a_2}, \dots x_k^{a_k};$$

3) показательная:

$$\bar{y}_{1,2,\dots,k} = e^{a_0+a_1x_1+a_2x_2+\dots+a_kx_k};$$

4) параболическая:

$$\bar{y}_{1,2,\dots,k} = a_0 + a_1x_1^2 + a_2x_2^2 + \dots + a_kx_k^2;$$

5) гиперболическая:

$$\bar{y}_{1,2,\dots,k} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_k}{x_k}.$$

Основное значение имеют линейные модели в силу простоты и логичности их экономической интерпретации. Нелинейные формы зависимости приводятся к линейным путем линеаризации.

Важным этапом построения уже выбранного уравнения множественной регрессии является отбор и последующее включение факторных признаков.

Сложность формирования уравнения множественной регрессии заключается в том, что почти все факторные признаки находятся в зависимости один от другого.

Проблема размерности модели связи, т. е. определение оптимального числа факторных признаков, является одной из основных проблем построения множественного уравнения регрессии. С одной стороны, чем больше факторных признаков включено в уравнение, тем оно лучше описывает явление. Однако модель размерностью 100 и более факторных признаков сложно реализуема и требует больших затрат машинного времени. Сокращение размерности модели за счет исключения второстепенных, экономически и статистически несущественных факторов способствует простоте и качеству ее реализации. В то же время построение модели регрессии малой размерности может привести к тому, что такая модель будет недостаточно адекватна исследуемым явлениям и процессам.

Проблема отбора факторных признаков для построения моделей взаимосвязи может быть решена на основе эвристических или многомерных статистических методов анализа.

Метод экспертных оценок как эвристический метод анализа основных макроэкономических показателей, формирующих единую международную систему расчетов, основан на интуитивнологических предпосылках, содержательно-качественном анализе (подробнее данный метод рассмотрен в гл. 13). Анализ экспертной информации проводится на базе расчета и.анализа непараметрических показателей связи: ранговых коэффициентов корреляции Спирмена, Кендалла и конкордации (п. 9.8).

Наиболее приемлемым способом отбора факторных признаков является **шаговая регрессия** (шаговый регрессионный анализ). Сущность метода шаговой регрессии заключается в последовательном включении факторов в уравнение регрессии и последующей проверке их значимости. Факторы поочередно вводятся в уравнение так называемым "прямым методом". При проверке значимости введенного фактора определяется, насколько уменьшается сумма квадратов остатков и увеличивается величина множественного коэффициента корреляции (R). Одновременно используется и обратный метод, т. е. исключение факторов, ставших незначимыми на основе t-критерия Стьюдента. Фактор является незначимым, если его включение в уравнение регрессии только изменяет значение коэффициентов регрессии, не уменьшая суммы квадратов остатков и не увеличивая их значения. Если при включении в модель соответствующего факторного Признака величина множественного коэффициента корреляции увеличивается, а коэффициент регрессии не изменяется (или меняется несущественно), то данный признак существенен и его включение в уравнение регрессии необходимо.

Если же при включении в модель факторного признака коэффициенты регрессии меняют не только величину, но и знаки, а множественный коэффициент корреляции не возрастает, то данный факторный признак признается нецелесообразным для включения в модель связи.

Сложность и взаимное переплетение отдельных факторов, обуславливающих исследуемое экономическое явление (процесс), могут проявляться в так называемой **мультиколлинеарности.** Под мультиколлинеарностью понимается тесная зависимость между факторными признаками, включенными в модель.

Наличие мультикотинеарности между признаками приводит к:

- искажению величины параметров модели, которые имеют тенденцию к завышению;
- изменению смысла экономической интерпретации коэффицинтов регрессии;
- слабой обусловленности системы нормальных уравнений;
- осложнению процесса определения наиболее существенных факторных признаков.

В решении проблемы мультиколлинеарности можно выделить несколько этапов:

- установление наличия мультиколлинеарности;
- определение причин возникновения мультиколлинеарности;
- разработка мер по ее устранению.

Причинами возникновения мультиколлинеарности между признаками являются:

- изучаемые факторные признаки, характеризующие одну и ту же сторону явления или процесса. Например, показатели объема производимой продукции и среднегодовой стоимости основных фондов одновременно включать в модель не рекомендуется, так как они оба характеризуют размер предприятия;
- использование в качестве факторных признаков показателей, суммарное значение которых представляет собой постоянную величину;
- факторные признаки, являющиеся составными элементами друг друга;
- факторные признаки, по экономическому смыслу дублирующие друг друга.

Одним из индикаторов определения наличия мультиколлинеарности между признаками является превь/шение парным коэффициентом корреляции величины 0,8 ($г_x$,x,) и др.

Устранение мультиколлинеарности может реализовываться через исключение из корреляционной модели одного или нескольких линейно-связанных факторных признаков или преобразование исходных факторных признаков в новые, укрупненные факторы.

Вопрос о том, какой из факторов следует'отбросить, решается на основании качественного и логического анализов изучаемого явления.

Качество уравнения регрессии зависит от степени достоверности и надежности исходных данных и объема совокупности. Исследователь должен стремиться к увеличению числа наблюдений, так как большой объем наблюдений является одной из предпосылок построения адекватных статистических моделей.

Аналитическая форма выражения связи результативного признака и ряда факторных называется многофакторным (множественным) **уравнением регрессии, или моделью связи.**

Уравнение линейной множественной регрессии имеет вид:

$$\bar{y}_{1,2,3,\dots,k} = a_0 + a_1x_1 + a_2x_2 + \dots a_kx_k, \quad (9.7)$$

где $\bar{y}_{1,2,3,\dots k}$ - теоретические значения результативного признака, полученные в результате подстановки соответствующих значений факторных признаков в уравнение регрессии;

$x_1, x_2, \dots, x_k$ - факторные признаки;

$a_0, a_1, \dots, a_k$ - параметры модели (коэффициенты регрессии).

Параметры уравнения могут быть определены графическим методом, методом наименьших квадратов и т. д.

Наиболее приемлемым способом отбора факторных признаков является **шаговая регрессия** (шаговый регрессионный анализ). Сущность метода шаговой регрессии заключается в последовательном включении факторов в уравнение регрессии и последующей проверке их значимости. Факторы поочередно вводятся в уравнение так называемым "прямым методом". При проверке значимости введенного фактора определяется, насколько уменьшается сумма квадратов остатков и увеличивается величина множественного коэффициента корреляции (R). Одновременно используется и обратный метод, т. е. исключение факторов, ставших незначимыми на основе t-критерия Стьюдента. Фактор является незначимым, если его включение в уравнение регрессии только изменяет значение коэффициентов регрессии, не уменьшая суммы квадратов остатков и не увеличивая их значения. Если при включении в модель соответствующего факторного признака величина множественного коэффициента корреляции увеличивается, а коэффициент регрессии не изменяется (или меняется несущественно), то данный признак существенен и его включение в уравнение регрессии необходимо.

Если же при включении в модель факторного признака коэффициенты регрессии меняют не только величину, но и знаки, а множественный коэффициент корреляции не возрастает, то данный факторный признак признается нецелесообразным для включения в модель связи.

Сложность и взаимное переплетение отдельных факторов, обуславливающих исследуемое экономическое явление (процесс), могут проявляться в так называемой **мультиколлинеарности.** Под мультиколлинеарностью понимается тесная зависимость между факторными признаками, включенными в модель.

Наличие мультиколлинеарности между признаками приводит к:

- искажению величины параметров модели, которые имеют тенденцию к завышению;
- изменению смысла экономической интерпретации коэффицинтов регрессии;
- слабой обусловленности системы нормальных уравнений;
- осложнению процесса определения наиболее существенных факторных признаков.

В решении проблемы мультиколлинеарности можно выделить несколько этапов:

- установление наличия мультиколлинеарности;
- определение причин возникновения мультиколлинеарности;
- разработка мер по ее устранению.

Причинами возникновения мультиколлинеарности между признаками являются:

- изучаемые факторные признаки, характеризующие одну и ту же сторону явления или процесса. Например, показатели объема производимой продукции и среднегодовой стоимости основных фондов одновременно включать в модель не рекомендуется, так как они оба характеризуют размер предприятия;
- использование в качестве факторных признаков показателей, суммарное значение которых представляет собой постоянную величину;
- факторные признаки, являющиеся составными элементами друг друга;
- факторные признаки, по экономическому смыслу дублирующие друг друга.

Одним из индикаторов определения наличия мультиколлинеарности между признаками является превышение парным коэффициентом корреляции величины 0,8 (r_x,x,) и др.

Устранение мультиколлинеарности может реализовываться через исключение из корреляционной модели одного или нескольких линейно-связанных факторных признаков или преобразование исходных факторных признаков в новые, укрупненные факторы.

Вопрос о том, какой из факторов следует'отбросить, решается на основании качественного и логического анализов изучаемого явления.

Качество уравнения регрессии зависит от степени достоверности и надежности исходных данных и объема совокупности. Исследователь должен стремиться к увеличению числа наблюдений, так как большой объем наблюдений является одной из предпосылок построения адекватных статистических моделей.

Аналитическая форма выражения связи результативного признака и ряда факторных называется многофакторным (множественным) **уравнением регрессии; или моделью связи.**

Уравнение линейной множественной регрессии имеет вид:

$$\bar{У}_{1,2,3,\dots,k} = a_0 + a_1x_1 + a_2x_2 + \dots a_kx_k, \quad (9.7)$$

где $\bar{У}_{1,2,3\dots k}$ - теоретические значения результативного признака, полученные в результате подстановки соответствующих значений факторных признаков в уравнение регрессии;

$x_1, x_2, \dots, x_k$ - факторные признаки;

$a_0, a_1, \dots, a_k$ - параметры модели (коэффициенты регрессии).

Параметры уравнения могут быть определены графическим методом, методом наименьших квадратов и т. д.

Методом наименьших квадратов (см. п. 9.3) минимизируется выражение:

$$S=\sum_{i=1}^{n}\left(У-a_0-a_1x_1-a_2x_2-\dots-a_kx_k\right)^2=\min.$$

$$\frac{ds}{da_0}=0; \quad \frac{ds}{da_1}=0; \quad \frac{ds}{da_2}=0; \quad \dots; \quad \frac{ds}{da_k}=0.$$

Например, по параметру a,:

$$\frac{ds}{da_1}=\Sigma 2\left(У-a_0-a_1x_1-a_2x_2-\dots-a_kx_k\right)\cdot\left(-x_1\right)=0.$$

Делая соответствующие преобразования по всем значениям параметров a(, получаем:

$$-2\Sigma Уx_1 + 2a_0\Sigma x_1 + 2a_1\Sigma x_1^2 + 2a_2\Sigma x_2x_1 + \dots + 2a_k\Sigma x_kx_1,$$

отсюда:

$$a_0\Sigma x_1 + a_1\Sigma x_1^2 + a_2\Sigma x_2x_1 + \dots + a_k\Sigma x_kx_1 = \Sigma Уx.$$

В результате таких преобразований система нормальных уравнений с k неизвестными (по числу параметров a() имеет вид:

$$\begin{cases} na_0+a_1\Sigma x_1+a_2\Sigma x_2+\dots+a_i\Sigma x_i+\dots+a_k\Sigma x_k=\Sigma У; \\ a_0\Sigma x_1+a_1\Sigma x_1^2+a_2\Sigma x_2x_1+\dots+a_i\Sigma x_ix_1+\dots+a_k\Sigma x_kx_1=\Sigma У x_1; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0\Sigma x_k+a_1\Sigma x_1x_k+a_2\Sigma x_2x_k+\dots+a_i\Sigma x_ix_k+\dots+a_k\Sigma x_k^2=\Sigma У x_k. \end{cases}$$

(9.8)

Одним из способов построения множественных уравнений регрессии является построение **модели связи в стандартизованном масштабе.**

Оценка влияния каждого факторного признака, включенного в уравнение регрессии, на результативный признак может быть значительно затруднена, если факторные признаки различны по своей сущности и имеют различные единицы измерения. В этих случаях для более точной оценки влияния факторных признаков используют множественные модели регрессии в стандартизованном масштабе. Модель регрессии в стандартизованном масштабе предполагает, что все значения исследуемых признаков переводятся в стандарты по формуле

$$(9.9)\quad t_{x_i}=\frac{x_i-\bar{x}_i}{\sigma_i},$$

где Xj - значение признака в натуральном масштабе.

Уравнение множественной регрессии в стандартизованном масштабе следующее:

$$\bar{t}_{1,2,\dots k} = \beta_1 t_1 + \beta_2 t_2' + \dots + \beta_k t_k, \quad (9.10)$$

где $t_1, t_2, t_3, \dots, t_k$ — стандартизованные значения признаков X|, $x_2, \dots, x_k$;

$\bar{t}_{1,2,\dots k}$ - среднее значение стандартизованной переменной соответствующего результативного признака, полученного по уравнению регрессии;

$\beta_1, \beta_2, \dots, \beta_k$ - стандартизованные коэффициенты регрессии.

Параметры многофакторной модели регрессии в стандартизованном масштабе определяются методом наименьших квадратов, рассмотренным выше.

Представим систему нормальных уравнений:

$$\begin{cases} \Sigma tt_1 = \beta_1 \Sigma t_1^2 + \beta_2 \Sigma t_1 t_2 + \ldots + \beta_k \Sigma t_1 t_k; \\ \Sigma tt_2 = \beta_1 \Sigma t_1 t_2 + \beta_2 \Sigma t_2^2 + \ldots + \beta_k \Sigma t_2 t_k; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \Sigma tt_k = \beta_1 \Sigma t_1 t_k + \beta_2 \Sigma t_2 t_k + \ldots + \beta_k \Sigma t_k^2, \end{cases}$$

где t - значение результативного признака в стандартизованном масштабе.

Коэффициенты (3|5 Р2> ••• Pk дают возможность провести сравнительную оценку силы влияния изменения каждого факторного признака на изменение результативного (моделируемого) признака.

От уравнения в стандартизованном масштабе легко можно перейти к уравнению в натуральном масштабе. Коэффициенты а| получают из соотношения

$$a_i = \beta_i \frac{\sigma_i}{\sigma_y} \cdot (9.11)$$

а свободный член а$_п$ — из выражения

$$a_0 = \bar{y} - a_1 \bar{x}_1 - a_2 \bar{x}_2 - \dots - a_k \bar{x}_k.$$

По следующим данным о прибыли (У), затратах на 1 руб. произведенной продукции (Х|) и стоимости основных фондов (x_2) необходимо определить зависимость между признаками (табл. 9.5).

Таблица 9.5

Расчетная таблица для определения параметров уравнения регрессии (данные условные)

№ п/п	Затраты на 1 руб. произведенной продукции, коп. $X_($	Стоимость основных фондов млн X_2	Прибыль тыс. руб. У	X^2,	X, X_2	УХ,	$Y^2_п$	$УX_2$	ух
1	77	5,9	1070	5929	454,3	82390	34,81	6313,0	1012,8
2	77	5,9	1001	5929	454,3	77077	34,81	5905,9	1012,8
3	81	4,9	789	6561	396,3	63909	24,01	3866,1	854,7
4	82	4,3	779	6724	352,6	63878	18,49	3349,7	817,8
5	89	3,9	606	7921	347,1	53934	15,21	2363,4	530,8
6	96	4,3	221	9216	412,8	21216	18,49	950,3	237,1
Итого	502	29,2	4466	42280	2418,0	362404	145,82	22748,4	4466,0

Система нормальных уравнений имеет вид:

$$\begin{cases} na_0 + a_1\Sigma x_1 + a_2\Sigma x_2 = \Sigma У; \\ a_0\Sigma x_1 + a_1\Sigma x_1^2 + a_2\Sigma x_1 x_2 = \Sigma x_1 У; \\ a_0\Sigma x_2 + a_1\Sigma x_1 x_2 + a_2\Sigma x_2^2 = \Sigma x_2 У. \end{cases}$$

$$\begin{cases} 6a_0 + 502a_1 + 29{,}2a_2 = 4466; \\ 502a_0 + 42280a_1 + 2418a_2 = 362404; \\ 29{,}2a_0 + 2418a_1 + 145{,}82a_2 = 22748{,}4. \end{cases}$$

Таким образом

$$\bar{У}_x = 4247{,}79 - 41{,}43x_1 - 7{,}60x_2.$$

9.5

ОЦЕНКА СУЩЕСТВЕННОСТИ СВЯЗИ. ПРИНЯТИЕ РЕШЕНИЙ НА ОСНОВЕ УРАВНЕНИЯ РЕГРЕССИИ

Проверка **адекватности** моделей, построенных на основе уравнений регрессии, начинается с проверки значимости каждого коэффициента регрессии.

Значимость коэффициентов регрессии осуществляется с помощью t-критерия Стьюдента:

$$(9.12)\ t_p = \frac{|a_i|}{\sqrt{\sigma^2_{a_i}}}$$

где $O^2_а$, - дисперсия коэффициента регрессии.

Параметр модели признается статистически значимым, если

$$t_p > t_{kp}\ (\alpha;\ n = n - k - 1),$$

где а - уровень значимости критерия проверки гипотезы о равенстве нулю параметров, измеряющих связь, т. е. статистическая существенность связи утверждается при отклонении нулевой гипотезы об отсутствии связи;

n = n - k - 1 - число степеней свободы, которое характеризует число свободно варьирующих элементов совокупности.

Наиболее сложным в этом выражении является определение дисперсии, которая может быть рассчитана двояким способом.

Наиболее простой способ, выработанный методикой экспериментирования, заключается в том, что величина дисперсии коэффициента регрессии может быть приближенно определена по выражению:

$$\sigma^2_{a_i} = \frac{\sigma^2_y}{k}, \quad (9.13)$$

где O^2_y - дисперсия результативного признака;

k - число факторных признаков в уравнении.

Более точную оценку величины дисперсии можно получить по формуле

$$\sigma_{a_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2}}{\sigma_{x_i} \cdot \sqrt{n} \cdot \sqrt{1 - R_i}},$$

(9.14)

где RJ — величина множественного коэффициента корреляции по фактору Xj с остальными факторами.

Проверка адекватности всей модели осуществляется с помощью расчета F-критерия и величины средней ошибки аппроксимации £.

Если $F_p > F_a$ при а = 0,05 или *а* = 0,01, то H_0 - гипотеза о несоответствии заложенных в уравнении регрессии связей реально существующим отвергается. Величина F_a определяется по специальным таблицам на основании величины а = 0,05 или а = 0,01 и числа степеней свободы: v, = k - 1, v_2 = n - k, где n - число наблюдений, k - число факторных признаков в уравнении.

Значение средней ошибки аппроксимации не должно превышать 12 - 15%.

$$\bar{\varepsilon} = \frac{1}{n}\Sigma\frac{\left|y - \bar{y}_{1,2,\dots,k}\right|}{y}\cdot 100.$$

Наиболее сложным этапом, завершающим регрессионный анализ, является интерпретация уравнения, т. е. перевод его с языка статистики и математики на язык экономики.

Интерпретация моделей регрессии осуществляется методами той отрасли знаний, к которой относятся исследуемые явления. Но всякая интерпретация начинается со статистической оценки уравнения регрессии в целом и оценки значимости входящих в модель факторных признаков, т. е. с выяснения, как они влияют на величину результативного признака. Чем больше величина коэффициента регрессии, тем значительнее влияние данного признака на моделируемый. Особое значение при этом имеет знак перед коэффициентом регрессии. Знаки коэффициентов регрессии говорят о характере влияния на результативный признак. Если факторный признак имеет знак плюс, то с увеличением данного фактора результативный признак возрастает; если факторный признак со знаком минус, то с его увеличением результативный признак уменьшается. Интерпретация этих знаков полностью определяется социально-экономическим содержанием моделируемого (результативного) признака. Если его величина изменяется в сторону увеличения, то плюсовые знаки факторных признаков имеют положительное влияние. При изменении результативного признака в сторону снижения положительное значение имеют минусовые знаки факторных признаков. Если экономическая теория подсказывает, что факторный признак должен иметь положительное значение, а он со знаком минус, то необходимо проверить расчеты параметров уравнения регрессии. Такое явление чаще всего бывает в силу допущенных ошибок при решении. Однако следует иметь в виду, что при анализе совокупного влияния факторов, при наличии взаимосвязей между ними характер их влияния может меняться. Для того чтобы быть уверенным, что факторный признак изменил знак влияния, необходима тщательная проверка решения данной модели, так как часто знаки могут меняться в силу допущенных ошибок при сборе или обработке информации.

При анализе адекватности уравнения регрессии исследуемому процессу возможны следующие варианты.

1. Построенная модель на основе ее проверки по F-критерию Фишера в целом адекватна, и все коэффициенты регрессии значимы. Такая модель может быть использована для принятия решений и осуществления прогнозов.
2. Модель по F-критерию Фишера адекватна, но часть коэффициентов регрессии незначима. В этом случае модель пригодна для принятия некоторых решений, но не для производства прогнозов.
3. Модель по F-критерию Фишера адекватна, но все коэффициенты регрессии незначимы. В этом случае модель полностью считается неадекватной. На ее основе не принимаются решения И не осуществляются прогнозы.

С целью расширения возможностей экономического анализа используются **частные коэффициенты эластичности,** определяемые по формуле

$$Э_{x_i} = a_i\cdot\frac{\bar{x}_i}{\bar{y}}, \qquad (9.15)$$

где Xj - среднее значение соответствующего факторного признака; у - среднее значение результативного признака; ^ - коэффициент регрессии при соответствующем факторном признаке. . •

Коэффициент эластичности показывает, на сколько процентов в среднем изменится значение результативного признака при изменении факторного признака на 1%.

Рассчитаем коэффициент эластичности ($Э_x$,) по исходным данным зависимости между прибылью АОЗТ (у), затратами на 1 руб. произведенной продукции (х,) и стоимостью основных фондов (x_2), приведенным в табл. 9.5.

$$a_1 = -41{,}43; \quad a_2 = -7{,}6;$$

$$\bar{y} = \frac{4466}{6} = 744{,}3;$$

$$\overline{x}_1 = \frac{502}{6} = 83,67; \qquad \overline{x}_2 = \frac{29,2}{6} = 4,9;$$

$$Э_{x_1} = a_1 \frac{\overline{x}_1}{\overline{y}} = -41,43 \cdot \frac{83,67}{744,3} = -4,7;$$

$$Э_{x_2} = a_2 \frac{\overline{x}_2}{\overline{y}} = -7,6 \cdot \frac{4,9}{744,3} = -0,05.$$

Это значит, что при увеличении затрат на 1 руб. произведенной продукции на 1% прибыль АОЗТ снизится на 4,7%, а при увеличении стоимости основных фондов на 1% прибыль снизится на 0,05%.

Частный коэффициент детерминации:

$$d_{x_i} = r_{yx_i} \cdot \beta_{x_i}, \tag{9.16}$$

где v_{yxj} - парный коэффициент корреляции между результативным и i-м факторным признаками;

(3х| - соответствующий коэффициент уравнения множественной регрессии в стандартизованном масштабе.

* Частный коэффициент детерминации показывает, на сколько процентов вариация результативного признака объясняется вариацией 1-го признака, входящего в множественное уравнение регрессии, i

По данным, приведенным в табл. 9.5, рассчитаем частный коэффициент детерминации для фактора х, - затраты на ! руб. произведенной продукции:

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\sigma_y \sigma_{x_1}};$$

$$d_{x_1} = r_{yx_1} \cdot \beta_{x_1}.$$

d_{X2} - частный коэффициент детерминации для фактора x_2 стоимости основных фондов:

Наиболее приемлемым способом отбора факторных признаков является **шаговая регрессия** (шаговый

$$d_{x_2} = r_{yx_2} \cdot \beta_{x_2};$$

$$\overline{yx} \quad \overline{yx_2} = \frac{22\,748,4}{6} = 3791,4; \ 0,7;$$

$$\overline{y} \quad \overline{x_2} = \frac{29,2}{6} = 4,9; \quad 4,3;$$

$$n \qquad 6$$

$$\overline{x_1} = \frac{\Sigma x_1}{n} = \frac{502}{6} = 83,67;$$

$$\sigma_y^2 = \overline{y^2} - \left(\overline{y}\right)^2 = \frac{3\,792\,340}{6} - (744,3)^2 = 78\,074,2;$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{78\,074,2} = 279,4;$$

$$\sigma_{x_1}^2 = \overline{x_1^2} - \left(\overline{x_1}\right)^2 = \frac{42\,280}{6} - (83,67)^2 = 46;$$

$$\sigma_{x_1} = \sqrt{\sigma_{x_1}^2} = \sqrt{46} = 6,78;$$

$$r_{yx_1} = \frac{60\,400,7 - 744,3 \cdot 83,67}{279,4 \cdot 6,78} = -0,98;$$

$$\beta_{x_1} = a_1 \cdot \frac{\sigma_{x_1}}{\sigma_y} = -41,43 \frac{6,78}{279,4} = -1,006;$$

$$d_{x_1} = -0,98 \, . \, (-1,006) = 0,99.$$

регрессионный анализ). Сущность метода шаговой регрессии заключается в последовательном включении факторов в уравнение регрессии и последующей проверке их значимости. Факторы поочередно вводятся в уравнение так называемым "прямым методом". При проверке значимости введенного фактора определяется, насколько уменьшается сумма квадратов остатков и увеличивается величина множественного коэффициента корреляции (R). Одновременно используется и обратный метод, т. е. исключение факторов, ставших незначимыми на основе t-критерия Стьюдента. Фактор является незначимым, если его включение в уравнение регрессии только изменяет значение коэффициентов регрессии, не уменьшая суммы квадратов остатков и не увеличивая их значения. Если при включении в модель соответствующего факторного признака величина множественного коэффициента корреляции увеличивается, а коэффициент регрессии не изменяется (или меняется несущественно), то данный признак существенен и его включение в уравнение регрессии необходимо.

Если же при включении в модель факторного признака коэффициенты регрессии меняют не только величину, но и знаки, а множественный коэффициент корреляции не возрастает, то данный факторный признак признается нецелесообразным для включения в модель связи.

Сложность и взаимное переплетение отдельных, факторов, обуславливающих исследуемое экономическое явление (процесс), могут проявляться в так называемой **мультиколлинеарности.** Под мультиколлинеарностью понимается тесная зависимость между факторными признаками, включенными в модель.

Наличие мультиколлинеарности между признаками приводит к:

- искажению величины параметров модели, которые имеют тенденцию к завышению;
- изменению смысла экономической интерпретации коэффицинтов регрессии;
- слабой обусловленности системы нормальных уравнений;
- осложнению процесса определения наиболее существенных факторных, признаков.

В решении проблемы мультиколлинеарности можно выделить несколько этапов:

- установление наличия мультиколлинеарности;
- определение причин возникновения мультиколлинеарности;
- разработка мер по ее устранению.

Причинами возникновения мультиколлинеарности между признаками являются:

- изучаемые факторные признаки, характеризующие одну и ту же сторону явления или процесса. Например, показатели объема производимой продукции и среднегодовой стоимости основных фондов одновременно включать в модель не рекомендуется, так как они оба характеризуют размер предприятия;
- использование в качестве факторных признаков показателей, суммарное значение которых представляет собой постоянную величину;
- факторные признаки, являющиеся составными элементами друг друга;
- факторные признаки, по экономическому смыслу дублирующие друг друга.

Одним из индикаторов определения наличия мультиколлинеарности между признаками является превышение парным коэффициентом корреляции величины 0,8 ($r_{X|}x,$) и др.

Устранение мультиколлинеарности может реализовываться через исключение из корреляционной модели одного или нескольких линейно-связанных факторных признаков или преобразование исходных факторных признаков в новые, укрупненные факторы.

Вопрос о том, какой из факторов следует'отбросить, решается на основании качественного и логического анализов изучаемого явления.

Качество уравнения регрессии зависит от степени достоверности и надежности исходных данных и объема совокупности. Исследователь должен стремиться к увеличению числа наблюдений, так как большой объем наблюдений является одной из предпосылок построения адекватных статистических моделей.

Аналитическая форма выражения связи результативного признака и ряда факторных называется многофакторным (множественным) **уравнением регрессии, или моделью связи.**

Уравнение линейной множественной регрессии имеет вид:

$$\tilde{y}_{1,2,3,...,k} = a_0 + a_1x_1 + a_2x_2 + ... a_kx_k, \quad (9.7)$$

где $\tilde{y}_{1,2,3,...k}$ - теоретические значения результативного признака, полученные в результате подстановки соответствующих значений факторных признаков в уравнение регрессии;

$x_1, x_2, ..., x_k$ - факторные признаки;

$a_0, a_1, ..., a_k$ - параметры модели (коэффициенты регрессии).

Параметры уравнения могут быть определены графическим методом, методом наименьших квадратов и т. д.

Методом наименьших квадратов (см. п. 9.3) минимизируется выражение:

$$S=\sum_{i=1}^{n}\left(У-a_0-a_1x_1-a_2x_2-\ldots-a_kx_k\right)^2=\min.$$

$$\frac{ds}{da_0}=0;\quad \frac{ds}{da_1}=0;\quad \frac{ds}{da_2}=0;\quad \ldots;\quad \frac{ds}{da_k}=0.$$

Например, по параметру а,:

$$\frac{ds}{da_1}=\Sigma 2\left(У-a_0-a_1x_1-a_2x_2-\ldots-a_kx_k\right)\cdot\left(-x_1\right)=0.$$

Делая соответствующие преобразования по всем значениям параметров 3j, получаем:

$$-2\Sigma Уx_1+2a_0\Sigma x_1+2a_1\Sigma x_1^2+2a_2\Sigma x_2x_1+\ldots+2a_k\Sigma x_kx_1,$$

отсюда:

$$a_0\Sigma x_1+a_1\Sigma x_1^2+a_2\Sigma x_2x_1+\ldots+a_k\Sigma x_kx_1=\Sigma Уx.$$

В результате таких преобразований система нормальных уравнений с k неизвестными (по числу параметров а;) имеет вид:

$$\begin{cases} na_0+a_1\Sigma x_1+a_2\Sigma x_2+\ldots+a_j\Sigma x_j+\ldots+a_k\Sigma x_k=\Sigma У; \\ a_0\Sigma x_1+a_1\Sigma x_1^2+a_2\Sigma x_2x_1+\ldots+a_j\Sigma x_jx_1+\ldots+a_k\Sigma x_kx_1=\Sigma У\,x_1; \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_0\Sigma x_k+a_1\Sigma x_1x_k+a_2\Sigma x_2x_k+\ldots+a_j\Sigma x_jx_k+\ldots+a_k\Sigma x_k^2=\Sigma У\,x_k. \end{cases}$$

(9.8)

Одним из способов построения множественных уравнений регрессии *является* построение **модели связи в стандартизованном масштабе.**

Оценка влияния каждого факторного признака, включенного в уравнение регрессии, на результативный признак может быть значительно затруднена, если факторные признаки различны по своей сущности и имеют различные единицы измерения. В этих случаях для более точной оценки влияния факторных признаков используют множественные модели регрессии в стандартизованном масштабе. Модель регрессии в стандартизованном масштабе предполагает, что все значения исследуемых признаков переводятся в стандарты по формуле

$$t_{x_i}=\frac{x_i-\bar{x}_i}{\sigma_i}, \qquad (9.9)$$

где Xj - значение признака в натуральном масштабе.

Уравнение множественной регрессии в стандартизованном масштабе следующее:

$$\bar{t}_{1,2,\ldots k}=\beta_1t_1+\beta_2t_2+\ldots+\beta_kt_k, \qquad (9.10)$$

где $t_1, t_2, t_{t_{1,2,\ldots k}}$ - стандартизованные значения признаков x,, x_2, •••, x_k;
- среднее значение стандартизованной переменной соответствующего результативного признака, полученного по уравнению регрессии;

$\beta_1, \beta_2, \ldots, \beta_k$ - стандартизованные коэффициенты регрессии.

Параметры многофакторной модели регрессии в стандартизованном масштабе определяются методом наименьших квадратов, рассмотренным выше.

Представим систему нормальных уравнений:

$$\begin{cases} \Sigma tt_1 = \beta_1 \Sigma t_1^2 + \beta_2 \Sigma t_1 t_2 + \ldots + \beta_k \Sigma t_1 t_k; \\ \Sigma tt_2 = \beta_1 \Sigma t_1 t_2 + \beta_2 \Sigma t_2^2 + \ldots + \beta_k \Sigma t_2 t_k; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \Sigma tt_k = \beta_1 \Sigma t_1 t_k + \beta_2 \Sigma t_2 t_k + \ldots + \beta_k \Sigma t_k^2, \end{cases}$$

где t - значение результативного признака в стандартизованном масштабе.

Коэффициенты Р,, (3$_2$,... Р$_k$ дают возможность провести сравнительную оценку силы влияния изменения каждого факторного признака на изменение результативного (моделируемого) признака.

От уравнения в стандартизованном масштабе легко можно перейти к уравнению в натуральном масштабе. Коэффициенты а, получают из соотношения

$$a_i = \beta_i \frac{\sigma_i}{\sigma_y}, \qquad (9.11)$$

а свободный член а$_0$ - из выражения

$$a_0 = \bar{У} - a_1\bar{x}_1 - a_2\bar{x}_2 - \ldots - a_k\bar{x}_k.$$

По следующим данным о прибыли (У), затратах на 1 руб. произведенной продукции (х,) и стоимости основных фондов (х$_2$) необходимо определить зависимость между признаками (табл. 9.5).

Таблица 9.5

Расчетная таблица для определения параметров уравнения регрессии (данные условные)

№ т1/п	Затраты на 1 произведенной продукции, коп. Х,	Стоимость основных фондов млн X_2	Прибыль тыс. руб. У	X^2.	Х,X_2	УХ,	Х\	УX_2	У*
1	77	5,9	1070	5929	454,3	82390	34,81	6313,0	1012,8
2	77	5,9	1001	5929	454,3	77077	34,81	5905,9	1012,8
3	81	4,9	789	6561	396,3	63909	24,01	3866,1	854,7
4	82	4,3	779	6724	352,6	63878	18,49	3349,7	817,8
5	89	3,9	606	7921	347,1	53934	15,21	2363,4	530,8
6	96	4,3	221	9216	412,8	21216	18,49	950,3	237,1
Итого	502	29,2	4466	42280	2418,0	362404	145,82	22748,4	4466,0

Система нормальных уравнений имеет вид:

$$\begin{cases} na_0 + a_1\Sigma x_1 + a_2\Sigma x_2 = \Sigma У; \\ a_0\Sigma x_1 + a_1\Sigma x_1^2 + a_2\Sigma x_1 x_2 = \Sigma x_1 У; \\ a_0\Sigma x_2 + a_1\Sigma x_1 x_2 + a_2\Sigma x_2^2 = \Sigma x_2 У. \end{cases}$$

$$\begin{cases} 6a_0 + 502a_1 + 29{,}2a_2 = 4466; \\ 502a_0 + 42\,280a_1 + 2418a_2 = 362404; \\ 29{,}2a_0 + 2418a_1 + 145{,}82a_2 = 22\,748{,}4. \end{cases}$$

Таким образом

$$\bar{У}_x = 4247{,}79 - 41{,}43x_1 - 7{,}60x_2.$$

9.5

ОЦЕНКА СУЩЕСТВЕННОСТИ СВЯЗИ.

ПРИНЯТИЕ РЕШЕНИЙ НА ОСНОВЕ УРАВНЕНИЯ РЕГРЕССИИ

Проверка **адекватности** моделей, построенных на основе уравнений регрессии, начинается с проверки значимости каждого коэффициента регрессии.

Значимость коэффициентов регрессии осуществляется с помощью t-критерия Стьюдента:

$$t_p = \frac{|a_i|}{\sqrt{\sigma^2_{a_i}}}, \quad (9.12)$$

где o"$^2_{ai}$ - дисперсия коэффициента регрессии.

Параметр модели признается статистически значимым, если

$$t_p > t_{kp}\ (\alpha;\ n = n - k - 1),$$

где а - уровень значимости критерия проверки гипотезы о равенстве нулю параметров, измеряющих связь, т. е. статистическая существенность связи утверждается при отклонении нулевой гипотезы об отсутствии связи;

n = n — k — 1 - число степеней свободы, которое характеризует число свободно варьирующих элементов совокупности.

Наиболее сложным в этом выражении является определение дисперсии, которая может быть рассчитана двояким способом.

Наиболее простой способ, выработанный методикой экспериментирования, заключается в том, что величина дисперсии коэффициента регрессии может быть приближенно определена по выражению:

$$\sigma^2_{a_i} = \frac{\sigma^2_y}{k}, \quad (9.13)$$

где O"^2_y - дисперсия результативного признака;

k - число факторных признаков в уравнении.

Более **точную** оценку величины дисперсии можно получить по формуле

$$\sigma_{a_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2}}{\sigma_{x_i} \cdot \sqrt{n} \cdot \sqrt{1 - R_i}},$$

(9.14)

где RJ - величина множественного коэффициента корреляции по фактору X| с остальными факторами.

Проверка адекватности всей модели осуществляется с помощью расчета F-критерия и величины средней ошибки аппроксимации £.

Если Fp > F_a при а = 0,05 или а = 0,01, то H_0 - гипотеза о несоответствии заложенных в уравнении регрессии связей реально существующим отвергается. Величина F_a определяется по специальным таблицам на основании величины а = 0,05 или а = 0,01 и числа степеней свободы: v, = к - 1, v_2 = n - k, где п — число наблюдений, k — число факторных признаков в уравнении.

Значение средней ошибки аппроксимации не должно превышать 12 - 15%.

$$\bar{\varepsilon} = \frac{1}{n} \Sigma \frac{|y - \bar{y}_{1,2,\ldots,k}|}{y} \cdot 100.$$

Наиболее сложным этапом, завершающим регрессионный анализ, является интерпретация уравнения, т. е. перевод его с языка статистики и математики на язык экономики.

Интерпретация моделей регрессии осуществляется методами той отрасли знаний, к которой относятся исследуемые явления. Но всякая интерпретация начинается со статистической оценки уравнения регрессии в целом и оценки значимости входящих в модель факторных признаков, т. е. с выяснения, как они влияют на величину результативного признака. Чем больше величина коэффициента регрессии, тем значительнее влияние данного признака на моделируемый. Особое значение при этом имеет знак перед коэффициентом регрессии. Знаки коэффициентов регрессии говорят о характере влияния на результативный признак. Если факторный признак имеет знак плюс, то с увеличением данного фактора результативный признак возрастает; если факторный признак со знаком минус, то с его увеличением результативный признак уменьшается. Интерпретация этих знаков полностью определяется социально-экономическим содержанием моделируемого (результативного) признака. Если его величина изменяется в сторону увеличения, то плюсовые знаки факторных признаков имеют положительное влияние. При изменении результативного признака в сторону снижения положительное значение имеют минусовые знаки факторных признаков. Если экономическая теория подсказывает, что факторный признак должен иметь положительное значение, а он со знаком минус, то необходимо проверить расчеты параметров уравнения регрессии. Такое явление чаще всего бывает в силу допущенных ошибок при решении. Однако следует иметь в виду, что при анализе совокупного влияния факторов, при наличии взаимосвязей между ними характер их влияния может меняться. Для того чтобы быть уверенным, что факторный признак изменил знак влияния, необходима тщательная проверка решения данной модели, так как часто знаки могут меняться в силу допущенных ошибок при сборе или обработке информации.

При анализе адекватности уравнения регрессии исследуемому процессу возможны следующие варианты.

1. Построенная модель на основе ее проверки по F-критерию Фишера в целом адекватна, и все коэффициенты регрессии значимы. Такая модель может быть использована для принятия решений и осуществления прогнозов.
2. Модель по F-критерию Фишера адекватна, но часть коэффициентов регрессии незначима. В этом случае модель пригодна для принятия некоторых решений, но не для производства прогнозов.
3. Модель по F-критерию Фишера адекватна, но все коэффициенты регрессии незначимы. В этом случае модель полностью считается неадекватной. На ее основе не принимаются решения и не осуществляются прогнозы.

С целью расширения возможностей экономического анализа используются **частные коэффициенты эластичности,** определяемые по формуле

$$Э_{x_i} = a_i \cdot \frac{\bar{x}_i}{\bar{y}}, \qquad (9.15)$$

где x^ - среднее значение соответствующего факторного признака; у - среднее значение результативного признака; aj - коэффициент регрессии при соответствующем факторном признаке.

Коэффициент эластичности показывает, на сколько процен в среднем изменится значение результативного признака о& изменении факторного признака на 1%. ^и

Рассчитаем коэффициент эластичности ($Э_{X|}$) по исходнк данным зависимости между прибылью АОЗТ (у), затратами М 1 руб. произведенной продукции (х,) и стоимостью основных фондов (x_2), приведенным в табл. 9.5.

$$a_1 = -41,43; \quad a_2 = -7,6;$$

$$\bar{y} = \frac{4466}{6} = 744,3;$$

$$\bar{x}_1 = \frac{502}{6} = 83,67; \quad \bar{x}_2 = \frac{29,2}{6} = 4,9;$$

$$Э_{x_1} = a_i \frac{\bar{x}_1}{\bar{y}} = -41,43 \cdot \frac{83,67}{744,3} = -4,7;$$

$$Э_{x_2} = a_2 \frac{\overline{x_2}}{\overline{y}} = -7,6 \cdot \frac{4,9}{744,3} = -0,05.$$

Это значит, что при увеличении затрат на 1 руб. произведенной продукции на 1% прибыль АОЗТ снизится на 4,7%, а при увеличении стоимости основных фондов на 1% прибыль снизится на 0,05%.

Частный коэффициент детерминации:

$$d_{x_i} = r_{yx_i} \cdot \beta_{x_i}, \tag{9.16}$$

где r_{yx}, - парный коэффициент корреляции между результативным и i-М факторным признаками;

р\j - соответствующий коэффициент уравнения множественной регрессии в стандартизованном масштабе.

•V Частный коэффициент детерминации показывает, на сколько процентов вариация результативного признака объясняется вариацией i-го признака, входящего в множественное уравнение регрессии.v

По данным, приведенным в табл. 9.5, рассчитаем частный коэффициент детерминации для фактора x, - затраты на 1 РУ&произведенной продукции:

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\sigma_y \sigma_{x_1}};$$

$$d_{x_1} = r_{yx_1} \cdot \beta_{x_1},$$

$$\sigma^2_{x_2} = \frac{145,82}{6} - (4,9)^2 = 0,29;$$

$$yx_1\ \overline{yx_2} = \frac{22\,748,4}{6} = 3791,4; 0,7;$$

$$\overline{y} = \overline{x_2} = \frac{29,2}{6} = 4,9; \quad 3;$$

$$\overline{x_1} = \frac{\Sigma x_1}{n} = \frac{502}{6} = 83,67;$$

$$\sigma^2_y = \overline{y^2} - \left(\overline{y}\right)^2 = \frac{3\,792\,340}{6} - (744,3)^2 = 78\,074,2;$$

$$\sigma_y = \sqrt{\sigma^2_y} = \sqrt{78\,074,2} = 279,4;$$

$$\sigma^2_{x_1} = \overline{x_1^2} - \left(\overline{x_1}\right)^2 = \frac{42\,280}{6} - (83,67)^2 = 46;$$

$$\sigma_{x_1} = \sqrt{\sigma^2_{x_1}} = \sqrt{46} = 6,78;$$

$$r_{yx_1} = \frac{60\,400,7 - 744,3 \cdot 83,67}{279,4 \cdot 6,78} = -0,98;$$

$$\beta_{x_1} = a_1 \cdot \frac{\sigma_{x_1}}{\sigma_y} = -41,43 \frac{6,78}{279,4} = -1,006;$$

$$d_{x_1} = -0,98 \cdot (-1,006) = 0,99.$$

dx, - частный коэффициент детерминации для фактора x_2 стоимости основных фондов:

$$\sigma_{x_2} = 0{,}54;$$

$$r_{yx_2} = \frac{3791{,}4 - 744{,}3 \cdot 4{,}9}{279{,}4 \cdot 0{,}54} = -0{,}96;$$

$$\beta_{x_2} = 7{,}6 \frac{0{,}54}{279{,}4} = 0{,}01;$$

$$d_{x_2} = 0{,}96 \cdot 0{,}01 = 0{,}01.$$

Это свидетельствует о том, что на 99% вариация прибыли АОЗТ объясняется изменением затрат на 1 руб. произведенной продукции.

Множественный коэффициент детерминации (R^2), представляющий собой множественный коэффициент корреляции в квадрате, характеризует, какая доля вариации результативного признака обусловлена изменением факторных признаков, входящих в многофакторную регрессионную модель.

Для более точной оценки влияния каждого факторного признака на моделируемый используют **Q-коэффициент,** определяемый по формуле

$$Q_{x_1} = Э_{x_1} \cdot V_{x_1}, \tag{9.17}$$

где V_x - коэффициент вариации соответствующего факторного признака. Q_x - для фактора затрат на 1 руб.

произведенной продукции:

$$Q_{x_1} = Э_{x_1} \cdot V_{x_1};$$

$$Э_{x_1} = 4{,}7;$$

$$V_{x_1} = \frac{\sigma_{x_1}}{\bar{x}_1} \cdot 100\% = \frac{6{,}79}{83{,}67} \cdot 100\% = 8{,}1\%;$$

$$Q_{x_1} = -\,4{,}7 \cdot 0{,}081 = -\,0{,}38.$$

Q_x ~ для фактора x_2 - стоимости основных фондов:

$$Q_{x_2} = Э_{x_2} \cdot V_{x_2};$$

$$Э_{x_2} = -\,0{,}05;$$

$$V_{x_2} = \frac{\sigma_{x_2}}{\bar{x}_2} \cdot 100\% = \frac{0{,}54}{4{,}9} \cdot 100\% = 11{,}0\%;$$

$$Q_{x_2} = -0{,}05 \cdot 11{,}0 : 100 = -0{,}006.$$

В целом, оценивая положительно значение уравнений регрессии как адекватных моделей связи, необходимо отметить их отрицательные свойства. Хорошую аппроксимацию эти модели имеют только для тех значений результативного признака, которые находятся в середине ранжированного ряда индивидуальных значений признака. Ошибка аппроксимации для данных значений не превышает 1 - 2%. На концах же исходного ряда величина ошибки аппроксимации может достигать до 50%. На основе уравнений регрессии невозможно получить оптимальное значение моделируемого показателя. Модели на основе уравнений регрессии обладают слабыми экстраполяционными свойствами, так как не отражают тенденции развития социально-экономических явлений и процессов и годны для построения лишь краткосрочных прогнозов, носящих вероятностный характер.

Наиболее полная экономическая интерпретация моделей регрессии позволяет *выявить* резервы развития и повышения деловой активности субъектов экономики.

9.6

СОБСТВЕННО-КОРРЕЛЯЦИОННЫЕ ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ ИЗУЧЕНИЯ СВЯЗИ. ОЦЕНКА СУЩЕСТВЕННОСТИ КОРРЕЛЯЦИИ

Измерение тесноты и направления связи является важной задачей изучения и количественного измерения взаимосвязи социально-экономических явлений. Оценка тесноты связи между признаками предполагает определение меры соответствия вариации результативного признака от одного (при изучении парных зависимостей) или нескольких (множественных) факторов.

Линейный коэффициент корреляции был впервые введен в начале 90-х гг. Пирсоном, Эджвортом и Велдоном и характеризует тесноту и направление связи между двумя коррелируемыми признаками в случае наличия между ними линейной зависимости.

В теории разработаны и на практике применяются различные модификации формул расчета данного коэффициента:

$$r = \frac{\overline{(x-\bar{x})\cdot(y-\bar{y})}}{\sigma_x \cdot \sigma_y}, \qquad (9.18)$$

Используя математические свойства средней, получаем:

$$r = \frac{\overline{(x-\bar{x})\cdot(y-\bar{y})}}{\sigma_x \cdot \sigma_y} = \frac{\overline{xy} - \overline{\bar{x}y} - \overline{x\bar{y}} + \overline{xy}}{\sigma_x \cdot \sigma_y} = \frac{\overline{xy} - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y}}{\sigma_x \cdot \sigma_y} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \cdot \sigma_y}.$$

(9.19)

Преобразования данной формулы позволяют получить следующую формулу линейного коэффициента корреляции:

$$r = \frac{\Sigma(x-\bar{x})\cdot(y-\bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \text{ или } r = \frac{\Sigma(x-\bar{x})\cdot(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2 \cdot \Sigma(y-\bar{y})^2}}, \qquad (9.20)$$

где п -^ число наблюдений.

Производя расчет по итоговым значениям исходных переменных, линейный коэффициент корреляции можно вычислить по формуле

$$r = \frac{n \cdot \Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2]\cdot[n\Sigma y^2 - (\Sigma y)^2]}}$$

или

$$r = \frac{\Sigma xy - \Sigma x \cdot \frac{\Sigma y}{n}}{\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]\cdot\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}}.$$

(9.21)(9.22)

Коэффициент корреляции может быть выражен через дисперсии слагаемых:

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \cdot \sigma_y}. \qquad (9.23)$$

Формулы (9.21), (9.22), (9.23) применяются при изучении совокупностей малого объема (п < 20 -*• 30).

Между линейным коэффициентом корреляции и коэффициентом регрессии существует определенная зависимость, выражаемая формулой

$$r = a_i \frac{\sigma_{x_i}}{\sigma_y}, \qquad (9.24)$$

где aj - коэффициент регрессии в уравнении связи;

o_x - среднее квадратическое отклонение соответствующего, статистически существенного, факторного признака.

Линейный коэффициент корреляции имеет большое значение при исследовании социально-экономических явлений и процессов, распределение которых близко к нормальному. Легко доказывается, что условие г = 0 является необходимым и достаточным для того, чтобы величины X и У были независимы. При этом условии коэффициенты регрессии а^, a_{xy} также обращаются в нуль, а прямые регрессии У по X и X по У оказываются взаимно перпендикулярными (параллельными: одна оси абсцисс, а вторая оси ординат). *

Если же г = 1, то это означает, что все точки (X, У) находятся на прямой и зависимость между X и У является функциональной. Прямые регрессии в этом случае совпадают. Указанное положение распространяется также на случай нормального распределения трех и более величин.

Линейный коэффициент корреляции изменяется в пределах от -1 до 1: -1 < г < 1. Знаки коэффициентов регрессии и корреляции совпадают. При этом интерпретацию выходных значений коэффициента корреляции можно представить в табл. 9.6.

Значимость линейного коэффициента корреляции проверяется на основе t-критерия Стьюдента. При этом выдвигается и проверяется гипотеза (H_0) о равенстве коэффициента корреляции нулю [H_0 : г = 0]. При проверке этой гипотезы используется t-статистика:

$$t_p = \sqrt{\frac{r^2}{1-r^2}\cdot(n-2)} = \frac{|r|}{\sqrt{1-r^2}}\sqrt{n-2}.$$

(9.25)

Таблица 9.6 **Оценка линейного коэффициента корреляции**

Значение линейного коэффициента связи	Характер связи	Интерпретация связи
r = 0	Отсутствует	-
0<r< 1	Прямая	С увеличением X увеличивается У
-1 <г<0	Обратная	С увеличением X уменьшается У, и наоборот
r=1	Фукциональная	Каждому значению факторного признака строго соответствует одно значение результативного признака

При выполнении H_0 t-статистика имеет распределение Стьюдента с входными параметрами: {a, k = n — 2).

Если расчетное значение $t_p > t_k$ (табличное), то гипотеза H_0 отвергается, что свидетельствует о значимости линейного коэффициента корреляции, а следовательно, и о статистической существенности зависимости между X и У.

Данный критерий оценки значимости применяется для совокупностей n < 50.

При большом числе наблюдений (п > 100) используется следующая формула t-статистики:

$$t_p = \frac{|r|}{\sqrt{1-r^2}}\cdot\sqrt{n}.$$

(9.26)

Для статистически значимого линейного коэффициента корреляции можно построить интервальные оценки с помощью z-распределения Фишера:

$$z = \frac{1}{2} \cdot \ln \frac{1+r}{1-r}.$$

Первоначально определяется интервальная оценка для z по выражению

$$Z \in \left[Z' \pm t_Y \cdot \sqrt{\frac{1}{n-3}} \right],$$

(9.27)

где t_v - табулированные значения для нормального распределения, зависимые от у = 1 - а (а - уровень вероятности); Z' - табличные значения, Z = f{r) - распределения. Функция Z' нечетная, т. е. Z' = f(-r) = - f(r).

Пример. На основе выборочных данных о деловой активности однотипных коммерческих структур оценить тесноту связи между прибылью (тыс. руб.) (У) и затратами на 1 руб. произведенной продукции (X) (табл.9.7).

Таблица 9.7

Расчетная таблица для определения коэффициента корреляции

№ п/п	У	X	ух	$У^2$	X^2
1	221	96	21 216	48841	9216
2	1 070	77	82390	1 144900	5929
3	1 001	77	*non*	I 002 000	5929
4 '	606	89	53934	367 236	7921
5	779	82	63878	606 841	6724
6	789	81	63909	622 520	6561
Сумма	4466	502	362 404	3 792 338	42 280
Средняя	744,33	83,67	60400,67	632056,33	7046,67

1. Используя формулу (9.19), получаем:

$$r = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_{..} \cdot \sigma_{..}};$$

$$r = \frac{6 \cdot 362404 - 4466 \cdot 502}{\sqrt{[6 \cdot 42\,280 - (502)^2] \cdot [6 \cdot 3\,792\,338 - (4466)^2]}} =$$

$$= \frac{2\,174\,424 - 2\,241\,932}{\sqrt{(253\,680 - 252\,004) \cdot (22\,754\,028 - 19\,945\,156)}} = \frac{-67\,508}{\sqrt{1676 \cdot 2\,808\,872}} =$$

$$= \frac{-67\,508}{68\,612{,}46} = -0{,}98.$$

Результат тот же.
По формуле (9.22);

$$r = \frac{362\,404 - 4\,466 \cdot 502/6}{\sqrt{[42\,280 - (502)^2/6] \cdot [3\,792\,338 - (4466)^2/6]}} =$$

$$= \frac{362\,404 - 373\,655{,}3}{\sqrt{(42\,280 - 42\,000{,}67) \cdot (3\,792\,338 - 3\,324\,192{,}7)}} =$$

$$= \frac{-11\,251{,}3}{\sqrt{279{,}33 \cdot 468\,145{,}3}} = \frac{-11\,251{,}3}{11\,435{,}9} = -0{,}98.$$

2. По формуле (9.21) значение коэффициента корреляции со- • ставило:

Таким образом, результат по всем формулам одинаков и свидетельствует о сильной обратной зависимости между изучаемыми признаками.

Проверка значимости коэффициента корреляции:

$$t_p = \frac{|r|}{\sqrt{1-r^2}} \cdot \sqrt{n-2} = \frac{0{,}98}{\sqrt{1-(-0{,}98)^2}} \cdot \sqrt{6-2} = 14{,}036.$$

Гипотеза H_0 отвергается при уровне значимости а = 0,05 и числе степеней свободы k = 6 - 2 = 4, так как $|t_p| > t_{kp} = 2{,}776$, что свидетельствует о значимости данного коэффициента корреляции:

Доверительные интервалы линейного коэффициента корреляции между прибылью и затратами на 1 руб. выпускаемой продукции получились: а = 0,05; у $^=$ 1 - ot = 1 - 0,05 = 0,95. Тогда t, = .1,96 -для нормального закона распределения (приложение 1) |г| = |-0,98| =• 0,98; Z' = 2,2976

(таблица z-распределения Фишера, приложение 6):

$$Z' - t_\gamma \cdot \sqrt{\frac{1}{n-3}} \le Z \le Z' + t_\gamma \cdot \sqrt{\frac{1}{n-3}};$$

$$2{,}2976 - 1{,}96 \cdot \sqrt{\frac{1}{6-3}} \le Z \le 2{,}2976 + 1{,}96 \cdot \sqrt{\frac{1}{6-3}};$$

$$1{,}1659 \le Z \le 3{,}4292.$$

По таблице z-распределения Фишера (приложение 6) определяем г (0,83 < г < 0,988).

В случае наличия линейной и нелинейной зависимости между двумя признаками для измерения тесноты связи применяют так называемое **корреляционное отношение.** Различают эмпирическое и теоретическое корреляционное отношение.

Эмпирическое корреляционное отношение рассчитывается по данным группировки, когда 8 характеризует отклонения групповых средних результативного показателя от общей средней:

$$\eta = \sqrt{\frac{\sigma^2 - \bar{\sigma}^2}{\sigma^2}} = \sqrt{1 - \frac{\bar{\sigma}^2}{\sigma^2}} = \sqrt{\frac{\delta^2}{\sigma^2}},$$

(9.28)

где Т) - корреляционное отношение; 5^2 - общая дисперсия;

O^2 - средняя из частных (групповых) дисперсий; 5^2 - межгрупповая дисперсия (дисперсия групповых средних).

Все эти дисперсии являются дисперсиями результативного признака. **Теоретическое корреляционное отношение** определяется по формуле

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}},$$

(9.29)

где 5^2 .- дисперсия выравненных значений результативною признака,

т. е. рассчитанных по уравнению регрессии;

$$\delta = \sqrt{\frac{\Sigma(\bar{y}_x - \bar{y})^2}{n}} = \sqrt{\delta^2_{y_x}};$$

O^2 - дисперсия 'эмпирических (фактических) значений результативного признака.

Тогда

$$\sigma = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}} = \sqrt{\sigma_y^2}.$$

$$\eta = \sqrt{\frac{\Sigma(\overline{y_x}-\bar{y})^2}{\Sigma(y-\bar{y})^2}} \qquad (9.30)$$

объясняется влиянием факторного признака.

В основе расчета корреляционного отношения лежит правило сложения дисперсий (гл. 7), т. е.

$$\sigma^2 = \sigma^2 + \overline{\sigma_i^2}, \qquad (9.31)$$

где $\overline{\sigma_i^2}$ - при изучении степени коррелированности факторов отражает вариацию результативного признака (У) под влиянием всех не учтенных при анализе факторов, т. е. носит остаточный характер:

$$\overline{\sigma_i^2} \approx \sigma_{ост}^2 = \frac{\Sigma(y-\overline{y_x})^2}{n}.$$

Отсюда формула корреляционного отношения принимает вид:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} \approx \sqrt{\frac{\sigma^2-\sigma_{ост}^2}{\sigma^2}} = \sqrt{1-\frac{\sigma_{ост}^2}{\sigma^2}}. \qquad (9.32)$$

Корреляционное отношение изменяется в пределах от 0 до 1 (О < Г| < 1), и анализ степени тесноты связи полностью соответствует линейному коэффициенту корреляции (табл. 9.6).

Теоретическое корреляционное отношение также может вычисляться по формуле

$$\eta = \sqrt{1-\frac{\Sigma(y-\overline{y_x})^2}{\Sigma(y-\bar{y})^2}}.$$

Корреляционное отношение является более универсальным показателем тесноты связи по сравнению с линейным коэффициентом корреляции.

Для измерения тесноты связи при множественной корреляционной зависимости, т. е. при исследовании трех и более признаков одновременно, вычисляются множественный или совокупный и частные коэффициенты корреляции.

Множественный коэффициент корреляции рассчитывается при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Множественный коэффициент корреляции вычисляется по формуле

$$R_{y/x_1x_2..} = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{1-\frac{\sigma_{ост}^2}{\sigma^2}},$$

где 5^2 - дисперсия теоретических значений результативного признака, рассчитанная по уравнению множественной регрессии; $O_{(;\text{ст}}$ - остаточная дисперсия; $а^2$ - общая дисперсия результативного признака.

В случае оценки связи между результативными (У) и двумя факторными признаками (х,) и ($х_2$) множественный коэффициент корреляции можно определить по формуле

$$R_{y/x_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1}\cdot r_{yx_2}\cdot r_{x_1x_2}}{1-r^2_{x_1x_2}}}, \qquad (9.33)$$

где г - парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции можно рассчитать, используя парные коэффициенты г^ и коэффициенты регрессии в стандартизованном масштабе (Pj):

$$R_{x_1,x_2,\ldots,x_k} = \sqrt{\beta_1\cdot r_{yx_1} + \beta_2\cdot r_{yx_2} + \ldots + \beta_k\cdot r_{yx_k}},$$

гдеr_{yx_i} - парные коэффициенты;

β_i — коэффициенты в стандартизованном масштабе.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: 0 < R < 1.

Приближение R к единице свидетельствует о сильной зависимости между признаками.

При небольшом числе наблюдений величина коэффициента множественной корреляции, как правило, завышается.

Чтобы оценить общую вариацию результативного (моделируемого) признака в зависимости от факторных признаков, величина коэффициента множественной корреляции корректируется на основании следующего выражения:

$$\hat{R}_{y/x_1,x_2,\ldots,x_k} = \sqrt{1-(1-R^2)\frac{n-1}{n-k-1}}, \qquad (9.34)$$

л

•где R- скорректированное значение; п - число наблюдений; k — число факторных признаков.

Корректировка R не производится при условии, если

$$\frac{n-k}{k} \geq 20.$$

Проверка значимости коэффициента множественной корреляции осуществляется на основе F-критерия Фишера-Снедекора:

$$F_p = \frac{\frac{1}{2}\cdot R^2_{y/x_1x_2}}{\frac{1}{n-3}\cdot\left(1-R^2_{y/x_1x_2}\right)}. \qquad (9.35)$$

Гипотеза H_0о незначимости коэффициента множественной корреляции$(H_0 : R = 0)$отвергается, если$F_p > F_{kp}\ (\alpha;\ v_1 = 2;$

$\nu_2 = n - 3$).

Оценка доверительных границ R производится следующим образом: величина R приравнивается к гиперболическому тангенсу величины Z, т. е. R = thr, где

$$Z = \frac{1}{2} \cdot \ln \frac{1+R}{1-R}.$$

Плотность распределения Z является почти нормальной со средним значением

$$\bar{Z} = \frac{1}{2} \cdot \ln \frac{1+R}{1-R} + \frac{R}{2(N-1)} \qquad (9.36)$$

и дисперсией

$$\sigma_z^2 = \frac{1}{N-3}.$$

Следовательно,

$$P\{-t\sigma_z < Z - Z_0 < t\sigma_z\} = \phi(t),$$

отсюда:

R;

$$Z - t\sigma_R < Z_0 < Z + t\sigma$$

$$Z_1 = Z - t\sigma_Z; \qquad (9.37)$$

$$Z_2 = Z + t\sigma_Z.$$

По таблицам Z-преобразования Фишера находят R, и R_2, т. е. $R_1 < R < R_Z$ - верхняя и нижняя границы значений R. На основе данных табл. 9.5 рассчитаем коэффициент множественной корреляции и его ошибку:

$$r_{yx_1} = \frac{\overline{yx_1} - \bar{y}\bar{x}_1}{\sigma_y \cdot \sigma_{x_1}} = -0{,}98;$$

$$r_{yx_2} = \frac{\overline{yx_2} - \bar{y}\bar{x}_2}{\sigma_y \cdot \sigma_{x_2}} = 0{,}78;$$

$$r_{x_1x_2} = \frac{\overline{x_1x_2} - \bar{x}_1\bar{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}} = -0{,}86.$$

Матрица линейных коэффициентов корреляции имеет вид:

$$\begin{bmatrix} 1 & -0{,}98 & 0{,}78 \\ & 1 & -0{,}86 \\ & & 1 \end{bmatrix}.$$

Множественный коэффициент корреляции составит:

$$R_{y/x_1x_2} = \sqrt{\frac{-0{,}98^2 + 0{,}78^2 - 2(-0{,}98)\cdot 0{,}78\cdot(-0{,}86)}{1-(-0{,}86)^2}} = 0{,}99.$$

Проверка условия $(n-k)/k = (6-2)/2 = 2 \le 20$ v_. ._, . ч , подтверждает возможность проведения корректировки данного коэффициента:

$$\tilde{R}_{y/x_1x_2} = \sqrt{1-\left(1-R^2_{y/x_1x_2}\right)\frac{n-1}{n-k-1}} = \sqrt{1-\left(1-(0{,}99)^2\right)\frac{6-1}{6-3}} = 0{,}98.$$

Проверка значимости коэффициента множественной корреляции показала:

$$F_p = \frac{1/2\cdot 0{,}98^2}{1/3\cdot(1-0{,}98^2)} = \frac{0{,}4802}{0{,}0132} = 36{,}3.$$

Гипотеза о незначимости коэффициента корреляции отвергается, так как

кр $F_{kp} = 9{,}55$ $(\alpha = 0{,}05;\ v_1 = 2;\ v_2 = n-3 = 3)$, $[F_p = 36{,}35] > [F_{kp} = 9{,}55]$.

Определим доверительные границы, в которых находится R в генеральной совокупности. При этом доверительную вероятность примем равной $\gamma = 0{,}95$ $(\alpha = 0{,}05;\ \gamma = 1-\alpha)$, $t\gamma = 1{,}96$ (приложение 1 «Нормальный закон распределения»).

$$R = 0{,}98;\ z = 2{,}2976;\ \sigma_R = 1/(6-1) = 0{,}2;$$

$$-1{,}96\cdot 0{,}2 < Z - Z_0 \le 1{,}96\cdot 0{,}2;$$

$$-0{,}392 \le Z - Z_0 \le 0{,}392;$$

$$Z_1 = 2{,}2976 - 0{,}392 = 1{,}9056;$$

$$Z_2 = 2{,}2976 + 0{,}392 = 2{,}6896;$$

$$1{,}9056 \le Z \le 2{,}6896.$$

Отсюда по таблице Z-преобразования Фишера (приложение 6) получаем: 0,96 < R < 0,941.

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками х, и x_2 при фиксированном значении других (k - 2) факторных признаков, т. е. когда влияние x_3 исключается и оценивается связь между х, и x_2 в «чистом виде».

Коэффициент, в котором исключается влияние только одного факторного признака, называется коэффициентом частной корреляции первого порядка. В общем виде коэффициент корреляции первого порядка выражается так:

$$r_{1,\,2,\,3,\,4,\ldots,\,k} = \frac{r_{1,\,2,\,3,\,4,\ldots,\,k-1} - r_{1,\,k,\,3,\ldots,\,k-1} \cdot r_{2,\,k,\,3,\ldots,\,k-1}}{\sqrt{(1 - r^2_{1,\,k,\,3,\ldots,\,k-1}) \cdot (1 - r^2_{2,\,k,\,3,\ldots,\,k-1})}}.$$

В случае зависимости У от двух факторных признаков х, и x_9 коэффициент частной корреляции следующий:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1 - r^2_{x_2y}) \cdot (1 - r^2_{x_1x_2})}};$$

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{x_1y}) \cdot (1 - r^2_{x_1x_2})}}, \qquad (9.38)$$

где г - парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака х,, во втором —x_1. Значения парного и частного коэффициентов корреляции отличаются друг от друга, так как парный коэффициент характеризует связь между двумя признаками без учета влияния других признаков, а частный - учитывает наличие и влияние других факторов.

Проверка значимости и расчет доверительных интервалов для частных коэффициентов корреляции аналогичны, как для парных коэффициентов, с тем лишь отличием, что число степеней свободы *v* определяется как

$$n = n - k,$$

где k - порядок коэффициента частной корреляции.

На основании приведенных выше данных о зависимости трех факторов деятельности предприятий вычислим частные коэффициенты корреляции первого порядка:

$$r_{yx_1/x_2} = \frac{-0{,}99 - 0{,}78 \cdot (-0{,}86)}{\sqrt{(1 - (0{,}78)^2)} \cdot \sqrt{(1 - (-0{,}86)^2)}} = -0{,}999;$$

$$r_{yx_2/x_1} = \frac{0{,}78-(-0{,}99)\cdot(-0{,}86)}{\sqrt{(1-(-0{,}99)^2)}\cdot\sqrt{(1-(0{,}78)^2)}} = -0{,}992;$$

$$r_{x_1x_2/y} = \frac{-0{,}86-(-0{,}99)\cdot(0{,}78)}{\sqrt{(1-(-0{,}99)^2)}\cdot\sqrt{(1-(0{,}78)^2)}} = -0{,}994.$$

Проверка значимости частных коэффициентов корреляции показала:

$$t_{p(yx_1/x_2)} = \frac{r_{yx_1/x_2}}{\sqrt{1-r^2_{yx_1/x_2}}}\cdot\sqrt{n-3} = \frac{-0{,}999\cdot\sqrt{3}}{\sqrt{1-(-0{,}999)^2}} = -38{,}7;$$

$$t_{p(yx_2/x_1)} = \frac{r_{yx_2/x_1}}{\sqrt{1-r^2_{yx_2/x_1}}}\cdot\sqrt{n-3} = \frac{-0{,}994\cdot\sqrt{3}}{\sqrt{1-(-0{,}994)^2}} = -15{,}74;$$

$$t_{p(x_1x_2/y)} = \frac{r_{x_1x_2/y}}{\sqrt{1-r^2_{x_1x_2/y}}}\cdot\sqrt{n-3} = \frac{-0{,}992\cdot\sqrt{3}}{\sqrt{1-(-0{,}992)^2}} = -13{,}61;$$

$$\left\{\left|t_{p(yx_1/x_2)}\right|;\ \left|t_{p(yx_2/x_1)}\right|;\ \left|t_{p(x_1x_2/y)}\right| > t_{kp}\ \ (\alpha = 0{,}05, \nu = n-3 = 3) = 3{,}181\right\}$$

Следовательно, все приведенные коэффициенты корреляции значимы.

Для значимых частных коэффициентов корреляции доверительные интервалы составят:

$$Z' - t_\gamma \cdot \sqrt{\frac{1}{n-4}} \le Z \le Z' + t_\gamma \cdot \sqrt{\frac{1}{n-4}};$$

$$|r_{yx_1/x_2}| = 0{,}999;$$

$$Z = 3{,}8002;$$

$$3{,}8002 - 1{,}96 \cdot \sqrt{1/2} \le Z \le 3{,}8002 + 1{,}96 \cdot \sqrt{1/2};$$

$$2{,}414 \le Z \le 5{,}186.$$

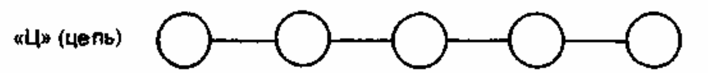
Используя таблицу Z-преобразований Фишера, получаем коэффициент корреляции в доверительных интервалах 0,98 < г < 1 и т. д.

9.7

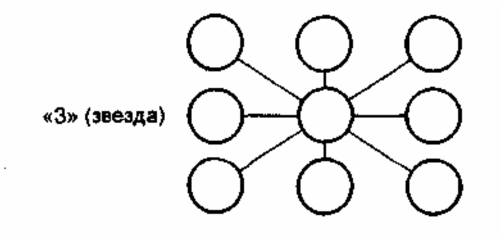
МЕТОДЫ ИЗУЧЕНИЯ СВЯЗИ СОЦИАЛЬНЫХ ЯВЛЕНИЙ

Важной задачей статистики является разработка методики статистической оценки социальных явлений, которая осложняется тем, что многие социальные явления не имеют количественной оценки.

Как правило, анализ социальных явлений, их связей и зависимостей должен начинаться с построения графиков связей. В настоящее время используются следующие графики, характеризующие связь социальных явлений:

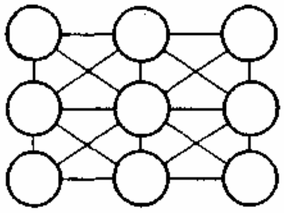


С помощью данного графика изображаются связи между социальными признаками, которые одинаково существенны и значимы.



изображает зависимость социальных явлений, которые тяготеют к одному наиболее значимому. Исключение данного признака нарушает взаимосвязи между оставшимися признаками.

«С» (сетка)



В этом случае выделяется несколько значимых признаков, которые тесно зависимы друг от друга.

Для количественной характеристики многомерных (многофакторных) связей социальных явлений используется **метод корреляционных плеяд,** основанный на расчете коэффициентов связи, которые носят общее название информативных коэффициентов. Метод корреляционных плеяд позволяет сгруппировать взаимосвязанные признаки в так называемые плеяды. Плеяды выделяются на основе матрицы информационных коэффициентов корреляции, коэффициентов сопряженности и т. д. Алгоритм построения корреляционных плеяд базируется на выделении максимальных значений информационных коэффициентов в исходной матрице значений. На базе отдельно выделенных плеяд строится «дерево плеяд», особенность которого в том, что внутриплеядные связи между факторными признаками тесные, а межплеядные - слабые.

Информационной основой для такого анализа служат данные различных социологических обследований на базе анкетирования.

Вычисление информационных коэффициентов служит основой для дальнейшего углубленного анализа связей между социальными явлениями. В настоящее время такой углубленный статистический анализ проводится по линии разработки корреляционных плеяд с дальнейшим переходом к факторному анализу или методу главных компонент.

Количественная оценка связей социальных явлений осуществляется на базе расчета и анализа целого ряда коэффициентов.

При наличии соотношения между вариацией качественных признаков говорят об их ассоциации, взаимосвязанности. Для оценки связи в этом случае используют ряд показателей.

Коэффициент ассоциации и контингенции. Для определения тесноты связи двух качественных признаков, каждый из которых состоит только из двух групп, применяются коэффициенты ассоциации и контингенции. При исследовании связи числовой материал располагают в виде таблиц сопряженности, например табл. 9.8. Для вычисления строится таблица, которая показывает связь между двумя явлениями, каждое из которых должно быть альтернативным, т. е. состоящим из двух качественно отличных друг от друга значений признака (например, хороший, плохой).

Коэффициенты определяются по формулам:

$$\text{ассоциации } K_a = \frac{ad - bc}{ad + bc};$$

Таблица 9.1

Таблица для вычисления коэффициентов ассоциации и контингенции

a	b	a + b
c	d	c + d
a + c	b + d	a + b + c + d

$$\text{контингенции } K_k = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}}. \qquad (9.40)$$

Коэффициент контингенции всегда меньше коэффициента ассоциации. Связь считается подтвержденной, если $K_a > 0{,}5$ или $K_k > 0{,}3$.

Пример. Исследовалась связь между успеваемостью студентов-заочников одного из вузов и работой их по специальности. Результаты обследования характеризуются следующими данными (табл. 9.9).

Таблица 9.9

Зависимость успеваемости студентов-заочников от работы их по специальности (цифры условные)

Студенты-заочники	Число студентов	Из них	
		получившие положительные оценки	получившие неудовлетвори-тельные оценки
Работающие по специальности	200	180	20
Не работающие по специальности	200	140	60
Итого	400	320	80

$$K_a = \frac{180 \cdot 60 - 140 \cdot 20}{180 \cdot 60 + 140 \cdot 20} = \frac{10\,800 - 2800}{10\,800 + 2800} = \frac{8000}{13\,600} = 0{,}6;$$

$$K_k = \frac{180 \cdot 60 - 140 \cdot 20}{\sqrt{(180+20)\cdot(20+60)\cdot(60+140)\cdot(140+180)}} =$$

$$= \frac{10\,800 - 2800}{\sqrt{200 \cdot 80 \cdot 200 \cdot 320}} = \frac{8000}{32\,000} = 0{,}3.$$

Таким образом, связь между успеваемостью студентов-заочников и работой их по специальности существенная.

Когда каждый из качественных признаков состоит более чем из двух групп, то для определения тесноты связи возможно применение **коэффициента взаимной сопряженности Пирсона Чупрова** (табл. 9.10).

Таблица 9.10

Вспомогательная таблица для расчета коэффициента взаимной сопряженности

7Г^—^ у	I	II	III	Всего
I			$п_{xy}$	$п_x$
•II				$п_x$
III				$п_x$
Итого	$п_y$	$п_y$	$п_y$	п

Этот коэффициент вычисляется по следующей формуле:

$$K_п = \sqrt{\frac{\varphi^2}{1+\varphi^2}}; \quad K_ч = \sqrt{\frac{\varphi^2}{\sqrt{(K_1-1)(K_2-1)}}}, \qquad (9.41)$$

где φ^2 - показатель взаимной сопряженности;

φ - определяется как сумма отношений квадратов частот каждой клетки таблицы к произведению итоговых частот соответствующего столбца и строки. Вычитая из этой суммы 1, получим величину $ф^2$.

$$\varphi^2 = \sum \frac{n_{xy}^2}{n_x n_y} - 1;$$

- чисто значений (групп) первого признака;

K_1

K_2 - число значений (групп) второго признака.

Чем ближе величины $K_ч$ и $K_п$ к 1, тем связь теснее.

$$1+\varphi^2 = \sum \frac{\sum \frac{n_{xy}^2}{n_x}}{n_y} = \sum \frac{\sum \frac{n_{xy}^2}{n_y}}{n_x}.$$

Пример. С помощью коэффициента взаимной сопряженности исследовать связь между себестоимостью продукции и произврдительностью труда (табл. 9.11).

Таблица 9.11

Зависимость между производительностью труда и себестоимостью продукции (данные условные)

Себестоимость	Производительность труда I	I высокая	средняя	низкая
Низкая	19	12	9	40
Средняя	*1*	18	15	40
Высокая	4	10	26	40
Итого	30	40	50	120

$$1+\varphi^2=\frac{\frac{19^2}{30}+\frac{12^2}{40}+\frac{9^2}{50}}{40}+\frac{\frac{7^2}{30}+\frac{18^2}{40}+\frac{15^2}{50}}{40}+$$

$$+\frac{\frac{4^2}{30}+\frac{10^2}{40}+\frac{26^2}{50}}{40}=0{,}431+0{,}356+0{,}414=1{,}201;$$

$$1+\varphi^2=1{,}201;$$

$$\varphi^2=0{,}201;$$

$$K_п=\sqrt{\frac{0{,}201}{1{,}201}}=\sqrt{0{,}167}=0{,}41;$$

$$K_ч=\sqrt{\frac{0{,}201}{\sqrt{2\cdot 2}}}=0{,}32.\ \text{Связь средняя.}$$

В статистике существуют модификации **коэффициента Чупрова,** например, **через расчет x^2 - критерия Пирсона.** Коэффициент взаимной сопряженности ($K_ч$) вычисляется по формуле

$$K_ч=\sqrt{\frac{\chi^2}{n+\chi^2}},\qquad (9.42)$$

$$\text{где}\ \chi^2=n\left\{\sum_{xy}n_{xy}^2/n_x\cdot n_y-1\right\}$$

- наиболее распространенный критерий согласия, используемый для проверки статистической гипотезы о виде распределения. Коэффициент Чупрова изменяется в пределах $0 < K_ч < 1$. :

t

По данным табл. 9.11 получили следующие результаты: ;

$$\chi^2=120\left\{\frac{19^2}{30\cdot 40}+\frac{12^2}{40\cdot 40}+\frac{9^2}{50\cdot 40}+\frac{7^2}{30\cdot 40}+\frac{18^2}{40\cdot 40}+\right.$$

$$\left.+\frac{15^2}{50\cdot 40}+\frac{4^2}{30\cdot 40}+\frac{10^2}{40\cdot 40}+\frac{26^2}{50\cdot 40}-1\right\}=$$

$$=120\cdot\{1{,}1998-1\}=23{,}976;$$

$$C=\sqrt{\frac{23{,}976}{120+23{,}976}}=0{,}41.$$

Другой модификацией **коэффициента взаимной сопряженности Чупрова** является:

$$K_ч = \sqrt{\frac{\chi^2}{n \cdot \sqrt{(k_1 - 1) \cdot (k_2 - 1)}}}, \quad (9.43)$$

гдеK_1 - число строк в таблице; K_2 - число граф в таблице; п - число наблюдений.

Вычислим величину $K_ч$ для приведенного примера:

$$K_ч = \sqrt{\frac{23{,}976}{120\sqrt{(3-1)\cdot(3-1)}}} = \sqrt{0{,}0999} \approx 0{,}32.$$ **Связь средняя.**

Особое значение для оценки связи имеет **биссериальный коэффициент корреляции,** который дает возможность оценить связь между качественным альтернативным и количественным варьирующим признаками. Данный коэффициент вычисляется по формуле

$$r = \frac{\left|\bar{y}_2 - \bar{y}_1\right|}{\sigma_y} \cdot \frac{pq}{Z}, \quad (9.44)$$

где $\bar{y}_2$ и $\bar{y}_1$ - средние в группах;

σ_y - среднее квадратическое отклонение фактических значений признака от среднего уровня; р - доля первой группы; q - доля второй группы;

' z- табулированные (табличные) значения z-распределения в зависимости от р.

Пример. Уровень доходов сотрудников одной коммерческой структуры характеризуется следующими данными (табл. 9.12).

Таблица 9.12

Зависимость уровня доходов сотрудников коммерческой структуры от уровня их образования (цифры условные)

	Уровень доходов, руб. ------------ !				Всего человек
	200-300	300-400	400-500	^{5}°£~ 600	
	250	350	450	550	
Окончившие вузы	5	7	6	4	22
Не обучавшиеся в вузах	9	4	2	1	16
Итого	14	1!	8	5	38

$$\overline{y}_1 = \frac{250 \cdot 5 + 350 \cdot 7 + 450 \cdot 6 + 550 \cdot 4}{22} = \frac{8600}{22} = 390{,}9;$$

$$\overline{y}_2 = \frac{250 \cdot 9 + 350 \cdot 4 + 450 \cdot 2 + 550 \cdot 1}{16} = \frac{5100}{16} = 318{,}8;$$

$$\overline{y}_{общ} = \frac{250 \cdot 14 + 350 \cdot 11 + 450 \cdot 8 + 550 \cdot 5}{38} = \frac{13700}{38} = 360{,}5;$$

$$\sigma_y = 104{,}7;$$

$$p = \frac{22}{38} = 0{,}58; \quad q = 0{,}42;$$

$$Z_{табл} = 0{,}3975;$$

$$p \cdot \frac{q}{Z} = 0{,}58 \cdot \frac{0{,}42}{0{,}3975} = 0{,}61;$$

$$r = \frac{|318{,}8 - 390{,}9|}{104{,}7} \cdot 0{,}61 = 0{,}42.$$

Величина биссериального коэффициента корреляции также подтверждает умеренную тесноту связи между изучаемыми признаками.

9.8

НЕПАРАМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ СВЯЗИ. РАНГОВЫЕ КОЭФФИЦИЕНТЫ СВЯЗИ

В анализе социально-экономических явлений часто приходится прибегать к различным условным оценкам, например рангам а взаимосвязь между отдельными признаками измерять с помощью непараметрических коэффициентов связи. Данные коэффициенты исчисляются при условии, что исследуемые признаки подчиняются различным законам распределения.

Ранжирование - это процедура упорядочения объектов изучения, которая выполняется на основе предпочтения.

Ранг - это порядковый номер значений признака, расположенных в порядке возрастания или убывания их величин. Если значения признака имеют одинаковую количественную оценку, то ранг всех этих значений принимается равным средней арифметической от соответствующих номеров мест, которые определяют. Данные ранги называются **связными.**

Пример. Проранжировать предприятия автомобильной про-* мышленности одного из регионов по величине балансовой прибыли (табл. 9.13).

Таблица 9.13

Балансовая прибыль предприятий автомобильной промышленности одного из регионов в 1998 г.

(цифры условные)

№ предприятия	Балансовая прибыль, млн руб.	Ранжирование (ранги)
1	Ю	6,5
2	12	4
3	10	6,5
4	12	4
5	12	4
6	15	2
7	17	1

Наиболее предпочтительному предприятию, величина балансовой прибыли которого наибольшая, присваивается ранг «1»; затем в порядке уменьшения величины балансовой прибыли были проранжированы все рассматриваемые предприятия автомобильной промышленности.

Принцип нумерации значений исследуемых признаков является основой непараметрических методов изучения взаимосвязи между социально-экономическими явлениями и процессами.

Среди непараметрических методов оценки тесноты связи наибольшее значение имеют ранговые коэффициенты Спирмена (р) и Кендалла (т). Эти коэффициенты могут быть использованы для определения тесноты связей как между количественными, так и между качественными признаками при условии, если их значения упорядочить или проранжировать по степени убывания или возрастания признака.

Коэффициент корреляции рангов (коэффициент **Спирмена)** рассчитывается по формуле (для случая, когда нет связных рангов):

$$\rho_{x/y} = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)}, \quad (9.45)$$

где d,^{2} - квадрат разности рангов;

п - число наблюдений (число пар рангов).

Коэффициент Спирмена принимает любые значения в интервале [-1; 1]. Значимость коэффициента корреляции рангов Спирмена проверяется на основе t-критерия Стьюдента. Расчетное значение критерия определяется по формуле

$$t_p = \rho_{x/y} \cdot \sqrt{\frac{n-2}{1-\rho_{x/y}^2}}. \quad (9.46)$$

Значение коэффициента корреляции считается статистически существенным, если $t_p > t_{kp}$ (ос; k = n - 2).

Пример. По данным группы предприятий одной из отраслей промышленности (табл. 9.14) определить с помощью коэффициента Спирмена зависимость между величиной балансовой прибыли (тыс. руб) и объемом реализованной продукции (млн руб.).

Таблица 9.14 Расчет коэффициента Спирмена

(данные условные)

пред- приятия	Объем реали- зован- ной продук- Ц""" руб.	Ба_ лан· совая при· быль> тыс. у X	Ранжирование				Сравне- ние рангов		Разность рангов d = R - R,	dj2
			X	R,	.У	R,	R_x	R,	' "	^ млн $р_{уб}$
1	1,8	20	1,3	1	20	1	2	1	1	1
2	2,3	75	1,8	2	42	2	3	3	О	О
3	8,6	42	2,3	3	75	3	10	2	8	64

4	1,3	80	3,5	4	80	4	1	4	-3	9
5	3,5	107	3,7	5	107	5	4	5	-1	1
6	3,8	125	3,8	6	125	6	6	6	O	O
7	4,5	140	4,5	7	140	7	7	7	O	O
8	5,8	175	5,8	8	175	8	8	8	O	O
9	3,7	200	6,5	9	200	9	5	9	-4	16
10	6,5	210	8,6	10	210	10	9	10	~Г	1
Итого	-	-	_	_	_	_	-	_	-	92

*

$$\rho_{x/y} = 1 - \frac{6 \cdot 92}{10 \cdot 99} = 1 - \frac{552}{990} = 0{,}44.$$

Связь, близкая к умеренной.

Если совокупность значений по исследуемому признаку содержит связные ранги, то коэффициент корреляции Спирмена вычисляется по формуле

$$\rho_{x/y} = \frac{\frac{1}{6} \cdot \left(n^3 - n\right) - \sum_{i=1}^{n} d_i^2 - T_x - T_y}{\sqrt{\left[\frac{1}{6} \cdot \left(n^3 - n\right) - 2T_x\right] \cdot \left[\frac{1}{6} \cdot \left(n^3 - n\right) - 2T_y\right]}}, \qquad (9.47)$$

где $T_{x/y} = \frac{1}{12} \cdot \sum_{i=1}^{k} \left(t_j^3 - t_j\right);$

t_j - число одинаковых рангов в }-м ряду.

На практике, если величины T_x и T_y несущественно отличаются относительно значения $\left[\frac{1}{6} \cdot \left(n^3 - n\right)\right]$, пользуются формулой

$$\rho_{x/y} = 1 - \frac{\sum_{i=1}^{n} d_i^2}{\frac{1}{6} \cdot \left(n^3 - n\right) - \left(T_x + T_y\right)}. \qquad (9.48)$$

Пример. По данным итогов торгов на биржевом рынке с 03.03.98 г. - 09.03.98 г. определить зависимость средней цены сделки от номинальной стоимости акции с помощью коэффициента Спирмена (табл. 9.15).

$$T_x = 1/12 \cdot [(5^3 - 5) + (3^3 - 3)] = 12;$$
$$T_y = 1/12 \cdot (2^3 - 2) = 0{,}5.$$

Применив формулу (9.47), получим:

$$\rho_{x/y} = \frac{1/6 \cdot (100-10) - 23{,}5 - 12 - 0{,}5}{\sqrt{[1/6 \cdot (1000-10) - 2 \cdot 12] \cdot [1/6 \cdot (1000-10) - 2 \cdot 0{,}5]}} =$$

$$= \frac{129}{\sqrt{141 \cdot 164}} = 0{,}85.$$

Таблица 9.15

Расчет коэффициента Спирмена

№	Номи-	Средняя цена	Ранжирование				Сравнение рангов		Разность рангов	d^2
та	руб X	руб. У	X	*	У	"у	RX	"у	d_1 — X P * у	
1	1,0	2,0	1,0	i	2,0	1	3	1	2	4
2	1,0	6,0	1,0	3	4,0	2,5	3	5	-2	4
3	1,0	4,0	1,0	3	4,0	2,5	3	2,5	0,5	0,25
4	1,0	4,0	1,0	3	5,7	4	3	2,5	0,5	0,25
5	2,5	7,8	1,0	3	6,0	5	6	6	0	0
6	10,	16,0	2,5	6	7,8	6	9	8	1	1
7	10,0	10,8	5,0	7	10,8	7	9	7	2	4
8	5,0	20,0	10,0	9	16,0	8	7	10	°	9
9	10,	16,4	10,0	9	16,4	9	9	9	0	0
10	1,0	5,7	10,0	9	20,0	10	3	4	_ j	1
Итого	-	-	-	-	-	-	-	-	-	23,5

Опыт проведения статистического анализа показывает, что формула 9.47 несущественно учитывает связные ранги при расчетах по формуле 9.45. Так, используя данные примера табл. 9.15, получаем, что

$$\rho_{x/y} = 1 - \frac{6 \cdot 23{,}5}{10 \cdot 99} = 0{,}85.$$

Но при этом вычисления по формуле (9.47) более громоздки. Поэтому на практике используют формулу расчета коэффициента Спирмена (9.45) как для случая отсутствия, так и наличия связных рангов.

Результаты расчетов по формуле (9.48) подтвердили предыдущие:

$$\rho_{x/y} = 1 - \frac{23{,}5}{1/6 \cdot (1000-10) - (12 + 0{,}5)} = 0{,}85,$$

Ранговый коэффициент корреляции Кендалла (т) может также использоваться для измерения взаимосвязи между качественными и количественными признаками, характеризующими однородные объекты, ранжированные по одному принципу. Расчет рангового коэффициента Кендалла осуществляется по формуле

$$\tau = \frac{2S}{n(n-1)}, \qquad (9.49)$$

где п — число наблюдений;

S - сумма разностей между числом последовательностей и числом инверсий по второму признаку.

Расчет данного коэффициента выполняется в следующей последовательности:

1) значения X ранжируются в порядке возрастания или убывания;

2) значения У располагаются в порядке, соответствующем значениям X;

3) для каждого ранга У определяется число следующих за ним значений рангов, превышающих его величину. Суммируя таким образом числа, определяется величина Р как мера соответствия последовательностей рангов по X и У и учитывается со знаком (+);

4) для каждого ранга У определяется число следующих за ним рангов, меньших его величины. Суммарная величина обозначается через Q и фиксируется со знаком (-);

5) определяется сумма баллов по всем членам ряда. В приведенном примере (табл. 9.14):

$$P = 6 + 8 + 6 + 5 + 1 + 3 + 2 + 1 + 0 = 32;$$

$$Q = (-3) + 0 + (-1) + (-1) + (-4) + (-1) + (-1) + (-1) + (-1) = -13.$$ Таким образом,

$$\tau = \frac{2(32-13)}{10(10-1)} = 0{,}42,$$

что свидетельствует о сильной связи между рассматриваемыми признаками.
что свидетельствует о наличии близкой к умеренной связи между рассматриваемыми признаками.

Если в изучаемой совокупности есть связные ранги, то расчеты необходимо проводить по следующей формуле:

$$\tau = \frac{S}{\sqrt{[n(n-1)/2 - V_x]\cdot[n(n-1)/2 - V_y]}}, \qquad (9.50)$$

По данным табл. 9.15 рассмотрим расчет коэффициента кор. реляции рангов Кендалла

где $V_{x/y} = 1/2\sum^{k} t_i \cdot (t_i - 1)$.

$$P = 9 + 5 + 6 + 6 + 5 + 4 + 0 + 1 + 1 = 37;$$

$$Q = 0 + (-3) + 0 + 0 + 0 + 0 + (-3) + (-1) + 0 = -7;$$

$$S = 37 + (-7) = 30;$$

$$V_x = 1/2 \cdot [5 \cdot (5-1) + 3 \cdot (3-1)] = 13;$$

$$V_y = 1/2 \cdot [2 \cdot (2-1)] = 1;$$

для случая наличия связных рангов:

что свидетельствует о существенной связи между номинальной стоимостью акций и средней ценой сделки на биржах. Как пра^ вило, коэффициент Кендалла меньше коэффициента Спирмена. При достаточно большом объеме совокупности значения данных коэффициентов имеют следующую

$$\tau = \frac{30}{\sqrt{[10\cdot(10-1)/2-13]\cdot[10\cdot(10-1)/2-1]}} = \frac{30}{\sqrt{(45-13)\cdot(45-1)}} = 0{,}8,$$

зависимость:

$$\tau = \frac{2}{3}\rho_{x/y}.$$

Связь между признаками можно признать статистически значимой, если значения коэффициентов ранговой корреляции Спирмена и Кендалла больше 0,5.

Для определения тесноты связи между произвольным числом ранжированных признаков применяется **множественный коэффициент ранговой корреляции (коэффициент конкордации)** *(W),* который вычисляется по формуле

(9.51) $$W = \frac{12S}{m^2 \cdot (n^3 - n)},$$

где m - количество факторов; n - число наблюдений;
S - отклонение суммы квадратов рангов от средней квадратов рангов.

Пример. Определить тесноту связи между уставным капиталом, числом выставленных акций и числом занятых на предприятиях, выставивших акции на чековые аукционы в 1998 г. (табл. 9.16).

Таблица 9.16

Расчет коэффициента конкордации (данные условные)

№ пред прия	Уставный капитал, тыс. руб.	Число выставленных акций У	Число занятых на предприя Z	R*	R>	R_7	Сумма стро	Квадраты сумм
1	2954	856	119	9	7	1	17	289
2	1 605	930	125	1	9	2	12	144
3	4 102	1563	132	10	10	3	23	529
4	2350	682	141	6	5	4	15	225
5	2625	616	150	7	3	5	15	225
6	1 795	495	165	4	2	6	12	144
7	2813	815	178	8	6	7	21	441
8	1 751	858	181	3	8	8	19	361
9	1 700	467	201	2	1	9	12	144
10	2264	661	204	5	4	10	19	361
	-	-	-	-	-	-	165	2863

$$S = 2863 - \frac{(165)^2}{10} = 2863 - 2722{,}5 = 140{,}5;$$

$$W = \frac{12S}{m^2(n^3 - n)} = \frac{12 \cdot 140{,}5}{(91000 - 10)} = 0{,}19.$$

Значимость коэффициента конкордации проверяется на основе 5С^2 - критерия Пирсона:

$$\chi^2 = \frac{12S}{m \cdot n(n-1)}. \quad (9.52)$$

Для нашего примера:

$$\chi^2_p = \frac{12 \cdot 140{,}5}{3 \cdot 10(10-1)} = 6{,}24.$$

Расчетное значение %$^2_р$ = 6,24 больше %$^2_{кр}$ = 16,919 (а = 0,05, v = n — 1 =9), что подтверждает незначимость коэффициента конкордации и свидетельствует о слабой связи между рассматриваемыми признаками.

В случае наличия связных рангов коэффициент конкордации определяется по формуле

$$W = \frac{S}{1/12m^2 \cdot (n^3 - n) - m \cdot \sum_{j=1}^{m} T_j}, \quad (9.53)$$

где $T_j = 1/2\sum_{j=1}^{m}(t_j^3 - t_j);$

t_j - количество связных рангов по отдельным показателям. Проверка значимости осуществляется по формуле -

$$\chi^2 = \frac{S}{1/12m \cdot n(n-1) - \frac{1}{(n-1)} \cdot \sum_{j=1}^{m} T_j}. \qquad (9.54)$$

Коэффициент конкордации принимает любые значения в интервале [-1; 1].

Пример. По данным предприятий хлопчатобумажной промышленности определить зависимость прибыли от реализации (тыс. руб.), от среднегодовой стоимости основных производственных фондов (млн руб.) и объема валовой продукции (млн руб.) (табл. 9.17).

Таблица 9.17 Расчет коэффициента конкордации (данные условные)

№ предприятия	Прибыль от реализации, тыс. руб.	Среднегодовая стоимость основных производственных фондов, млн руб.	Объем валовой продукции, млн руб.	R_x	R_y	R_z	Сумма строк	Квадрат сумм
	x	y	z					
1	40	1,7	0,27	1,5	1	1,5	4	16
2	75	3,2	0,55	3	5	4	12	144
3	82	2,9	0,97	4,5	3,5	5,5	13,5	182,25
4	40	1,8	0,27	1,5	2	1,5	5	25
5	106	11,8	0,98	6	6,5	7	19,5	380,25
6	82	2,9	0,35	4,5	3,5	3	-11	121
7	109	11,8	0,97	7	6,5	5,5	19	361
Итого	-		-	28	28	28	84	1229,5

Объем Среднего При- вал о- довая быль № вой стоимость от про- основных пред- реали- Сумма дук- произвол- К, Ry R, прия- зации, ции, ственных тия тыс. млн фондов, руб. млн руб. л У Z
1 40 1,7 0,27 1,5
1 1,5 4 16
2 75 3,2 0,55
3 5 4 12 144
3 82 2,9 0,97 4,5
3,5 5,5 13,5 182,25
4 40 1,8 0,27 1,5
2 1,5 5 25
5 106 11,8 0,98 6
6,5 7 19,5
380,25

$$T_x = \frac{1}{12}\left[(2^3-2)+(2^3-2)\right] = 1;$$

$$T_y = \frac{1}{12}\left[(2^3-2)+(2^3-2)\right] = 1;$$

$$T_z = \frac{1}{12}\left[(2^3-2)+(2^3-2)\right] = 1;$$

$$\sum T_j = T_x + T_y + T_z = 1+1+1 = 3.$$

$$W = \frac{S}{1/12m^2(n^3-n)-m\sum T_j} = \frac{221,5}{1/12\cdot 3^2(7^3-7)-3\cdot 3} = \frac{221,5}{243,0} = 0,91.$$

Расчетное значение ^-критерия Пирсона для проверки значимости коэффициента конкордации по данным нашего примера составил:

$$\chi_p^2 = \frac{221,5}{1/12\cdot 3\cdot 7\cdot (7-1)-\frac{1}{7-1}\cdot 3} = 22,15.$$

Расчетное значение % $_{\text{р}}$ = 22,15 больше % $_{\text{кр}}$ = 12,592, (а = 0,05, v = п —1 =6), что подтверждает значимость коэффициента конкордации и свидетельствует о сильной связи между рассматриваемыми признаками.

Преимуществом ранговых коэффициентов корреляции Спирмена, Кендалла и конкордации является то, что с их помощью можно измерять и оценивать связи как между количественными, так и между атрибутивными признаками, которые поддаются ранжированию.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 9

Причинно-следственные отношения — связь явлений и процессов, когда изменение одного из них - причины - ведет к изменению другого - следствия. Социально-экономические явления — это результат одновременного воздействия большого числа причин.

Признак- основная отличительная черта, особенность изучаемого явления или процесса.

Результативный признак - признак, изменяющийся под действием факторных признаков.

Факторный признак - признак, оказывающий влияние на изменение результативного.

Функциональная связь - связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака.

Стохастическая связь - связь, которая проявляется не в каждом отдельном случае, а в общем, среднем или большом числе наблюдений.

Корреляционная связь - изменение среднего значения результативного признака, которое обусловливается изменением факторных признаков.

Прямая связь - с увеличением или уменьшением значений факторного признака увеличивается или уменьшается значение результативного.

Обратная связь - с увеличением или уменьшением значений факторного признака уменьшается или увеличивается значение результативного.

Линейная связь — статистическая связь между явлениями, выраженная уравнением прямой линии.

Нелинейная связь - статистическая связь между социально-экономическими явлениями, аналитически выраженная уравнением кривой линии (параболы, гиперболы и т. д.).

Корреляция - статистическая зависимость между случайными величинами, которая не имеет строго функционального характера и при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

Регрессионный анализ - аналитическое выражение связи, в котором изменение одной величины - результативного признака - обусловлено влиянием одной или нескольких независимых величин (факторов), а множество всех прочих факторов, также оказывающих влияние на зависимую величину принимается за постоянные и средние значения.

Парная регрессия - аналитическое выражение связи двух признаков.

Множественная регрессия - модель связи трех и более признаков.

Коэффициент регрессии a_t показывает, насколько в среднем изменяется значение результативного признака при изменении факторного на единицу собственного измерения.

Мультиколлинеарность - наличие тесной зависимости между факторными признаками.

Коэффициент эластичности показывает, на сколько процентов в среднем изменится значение результативного признака при изменении факторного признака на 1%.

Коэффициент детерминации показывает, на сколько процентов вариация результативного признака объясняется вариацией i-го признака (частный) или всех вошедших в модель факторных признаков (множественный).

Линейный коэффициент корреляции определяет тесноту и направление связи между двумя коррелируемыми признаками.

Корреляционное отношение показывает связь между двумя признаками.

Множественный коэффициент корреляции отражает связь между результативным и несколькими факторными признаками.

Частный коэффициент корреляции показывает степень тесноты связи между двумя признаками при фиксированном значении остальных факторных признаков.

Экономическая интерпретация модели - основные выводы и заключения на основе расчета и анализа частных коэффициентов эластичности, частных и множественного коэффициентов детерминации, Q-коэффициента.

Коэффициенты ассоциации и контингенции - определение тесноты связи двух качественных признаков, каждый из которых состоит только из двух групп.

Коэффициент взаимной сопряженности Пирсона - Чупрова определение тесноты связи двух качественных признаков, каждый из которых состоит более чем из двух групп.

Биссериальный коэффициент корреляции - оценивание связи между качественным альтернативным и количественным варьирующим признаками.

Ранг - порядковый номер значения признака, расположенного в порядке возрастания или убывания величин.

Ранжирование - процедура упорядочения объектов изучения, которая выполняется на основе предпочтения значений признака в порядке возрастания или убывания.

Коэффициенты корреляции Спирмена и Кендалла - определение тесноты связи между двумя количественными или качественными признаками после предварительного ранжирования их по возрастанию или убыванию.

Коэффициент конкордации - определение тесноты связи между произвольным числом ранжированных признаков.

ТЕСТ К ГЛАВЕ 9

1. По направлению связи бывают:
а) умеренные;
б) прямые;
в) прямолинейные.

2. По аналитическому выражению связи различаются:
а) обратные;
б) тесные;
в) криволинейные.

3. Функциональной является связь:
а) между двумя признаками;
б) при которой определенному значению факторного признака соответствует несколько значений результативного признака;
в) при которой определенному значению факторного признака соответствует одно значение результативного признака.

4. Аналитическое выражение связи определяется с помощью методов анализа:
а) корреляционного;
б) регрессионного;
в) группировок.

5. Анализ тесноты и направления связей двух признаков осуществляется на основе:
а) парного коэффициента корреляции;
б) частного коэффициента корреляции;
в) множественного коэффициента корреляции.

6. Мультиколлинеарность - это связь между:
а) признаками;
б) уровнями;
в) явлениями.

7. Оценка значимости параметров модели регрессии осуществляется на основе:
а) коэффициента корреляции;
б) средней ошибки аппроксимации;
в) t-критерия Стьюдента.

8. Оценка значимости уравнения регрессии осуществляется на основе:

, а) коэффициента детерминации;

б) средней квадратической ошибки;

в) F-критерия Фишера.

9. Оценка связей социальных явлений производится на основе:

а) коэффициента ассоциации;

б) коэффициента контингенции; ,. в) коэффициента эластичности.

10. Коэффициент корреляции рангов Спирмена можно применять для оценки тесноты связи между:

а) количественными признаками;

•. б) качественными признаками, значения которых могут быть упорядочены;

в) любыми качественными признаками.

ГЛАВА 10

СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ДИНАМИКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

10.1

ПОНЯТИЕ И КЛАССИФИКАЦИЯ РЯДОВ ДИНАМИКИ

Процесс развития, движения социально-экономических явлений во времени в статистике принято называть динамикой. Для отображения динамики строят **ряды динамики** (хронологические, временные), которые представляют собой ряды изменяющихся во времени значений статистического показателя, расположенных в хронологическом порядке. В нем процесс экономического развития изображается в виде совокупности перерывов непрерывного, позволяющих детально проанализировать особенности развития при помощи характеристик, отражающих изменение параметров экономической системы во времени.

Составными элементами ряда динамики являются показатели уровней ряда и периоды времени (годы, кварталы^ месяцы, сутки) или моменты (даты) времени.

Уровни ряда обычно обозначаются через «у», моменты или периоды времени, к которым относятся уровни, - через «t».

Существуют различные виды рядов динамики. Их можно классифицировать по следующим признакам.

1. В зависимости от способа выражения уровней ряды динамики подразделяются на ряды абсолютных, относительных и средних величин.

Примером рядов динамики указанных выше видов являются данные табл. 10.1.

В табл. 10.1 рядом динамики абсолютных величин являются данные первой строки; рядом средних величин - второй строки; рядом относительных величин - третьей строки.

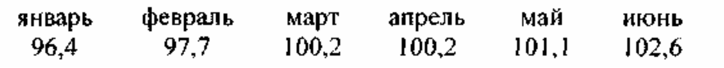
2. В зависимости от того, как выражают уровни ряда состояние явления на определенные моменты времени (на начало месяца, квартала, года и т. п.) или его величину за определенные интервалы времени (например, за сутки, месяц, год и т. п.), различают соответственно моментные и интервальные ряды динамики.

Таблица 10.1

Число квартир, построенных предприятиями и организациями всех форм собственности и их средний размер в РФ

	1980	1985	—'.......	1995	1997
I . Число квартир, тыс.	1190	1151	682	602	426
2. Средний размер квартир, м^2 общей площади	49,9	54,4	60,8	68,2	76,5
3. Удельный вес жилой площади в общей площади квартир, %	62,7	60,7	60,0	60,1	62,1

Примером моментного ряда может служить ряд динамики, показывающий число вкладов населения в учреждениях Сбербанка РФ (на первое число месяца 1997 г., в трлн руб.):



январь	февраль	март	апрель	май	июнь
96,4	97,7	100,2	100,2	101,1	102,6

Уровни этого ряда - обобщающие итоги статистики вкладов населения по состоянию на определенную дату (на начало каждого месяца).

Примером интервального ряда динамики являются данные, приведенные в табл. 10.1.

Из различного характера интервальных и моментных рядов динамики вытекают некоторые особенности уровней соответствующих рядов.

Уровни интервального ряда динамики абсолютных величин характеризуют собой суммарный итог какого-либо явления за определенный отрезок времени. Они зависят от продолжительности этого периода времени, и поэтому их можно суммировать как не содержащие повторного счета.

Отдельные же уровни моментного ряда динамики абсолютных величин содержат элементы повторного счета, например, число вкладов населения, учитываемых за январь, существует и в настоящее время, являясь единицами совокупности и в июне. Все это делает бессмысленным суммирование уровней моментных рядов динамики.

3. В зависимости от расстояния между уровнями ряды динамики подразделяются на ряды динамики с равноотстоящими уровнями и неравноотстоящими уровнями во времени. Ряды динамики следующих друг за другом периодов или следующих через определенные промежутки дат называются равноотстоящими (см. пример о числе вкладов в Сбербанк РФ за январь - июнь 1997 г.). Если же в рядах даются прерывающиеся периоды или неравномерные промежутки между датами, то ряды называются неравноотстоящими (см. пример в табл. 10.1).

4. В зависимости от наличия основной тенденции изучаемого процесса ряды динамики подразделяются на стационарные и нестационарные.

Если математическое ожидание значения признака и дисперсия (основные характеристики случайного процесса) постоянны, не зависят от времени, то процесс считается стационарным и ряды динамики также называются стационарными. Экономические процессы во времени обычно не являются стационарными, так как содержат основную тенденцию развития, но их можно преобразовать в стационарные путем исключения тенденций.

10.2

СОПОСТАВИМОСТЬ УРОВНЕЙ И СМЫКАНИЕ РЯДОВ ДИНАМИКИ

Важнейшим условием правильного построения ряда динамики является сопоставимость всех входящих в него уровней. Данное условие решается либо в процессе сбора и обработки данных, либо путем их пересчета.

Проблема сопоставимости данных особенно остро стоит в рядах динамики, потому что они охватывают значительные периоды времени, за которые могли произойти изменения, приводящие к несопоставимости статистических'данных. Рассмотрим основные причины несопоставимости уровней ряда динамики.

Несопоставимость уровней ряда может возникнуть вследствие изменения единиц измерения или единиц счета. Нельзя, например, сравнивать и анализировать цифры о производстве тканей, если за одни годы цифры даны в погонных метрах, а за другие в квадратных метрах.

На сопоставимость уровней ряда динамики непосредственно влияет **методология учета** или **расчета показателей.** Например, если в одни годы среднюю урожайность считали с засеянной площади, а в другие - с убранной, то такие уровни будут несопоставимы.

Условием сопоставимости уровней ряда динамики является **периодизация динамики.** В процессе развития во времени прежде всего происходят количественные изменения явлений, а затем на определенных ступенях совершаются качественные скачки, приводящие к изменению закономерности явления. Поэтому научный подход к изучению рядов динамики заключается в том, чтобы ряды, охватывающие большие периоды времени, разбивать на такие, которые бы объединяли лишь однокачественные периоды развития совокупности, характеризующейся одной закономерностью развития.

Процесс выделения однородных этапов развития рядов динамики носит название **периодизации динамики.** Вопрос о том, какие этапы развития прошло то или иное явление за определенный исторический отрезок времени, решается теорией той науки, к области которой относится изучаемая совокупность явления.

Необходимость формировать ряды динамики по строго однородным периодам или этапам не означает отрицания возможности построения и изучения рядов динамики, охватывающих длительные исторические отрезки времени, включающие различные этапы развития явления. Нужно помнить, что само понятие однородности периодов весьма относительно, оно зависит от уровня абстракции, принятой в исследовании. Например, весь советский период развития России, несомненно, является особым однородным периодом, кардинально отличающимся от предыдущего развития нашей страны. Внутри советского периода, в свою очередь, можно выделить более короткие, однородные в определенном отношении интервалы времени - довоенные годы, годы Великой Отечественной войны, послевоенные годы восстановления народного хозяйства и т. д. Если ряды динамики не готовятся непосредственно для анализа, а играют чисто информационную роль, они могут быть не периодизированы.

Важно также, чтобы *в* ряду динамики **интервалы** или **моменты,** по которым определены уровни, имели **одинаковый экономический смысл.** Скажем, при изучении роста поголовья скота бессмысленно сравнивать цифры поголовья по состоянию на 1 октября с 1 января, так как первая цифра включает не только скот, оставшийся на зимовку, но и предназначенный к убою, а вторая цифра включает только скот, оставленный на зимовку.

Условием сравнимости уровней интервального ряда является наличие равных интервалов, по которым даны уровни. Совершенно очевидно, что нельзя сравнивать квартальную продукцию с годовой.

Уровни ряда динамики могут оказаться несопоставимыми **по кругу охватываемых объектов** вследствие перехода ряда объектов из одного подчинения в другое.

Несопоставимость уровней ряда может возникнуть вследствие изменений **территориальных границ** областей, районов и т. д. При этом, говоря об изменении территории, к которой относятся уровни ряда за разное время, следует иметь в виду, что вопрос о сопоставимости или несопоставимости при изменении территории решается по-разному в зависимости от цели исследования. Если, например, ставится задача показать изменение численности населения или объема промышленного производства в связи с изменением административно-территориальных границ области или района, то не только можно, но и нужно сопоставлять данные в фактических границах этой области или района. Если же изучаются показатели темпов естественного прироста населения или темпов развития промышленности, то, очевидно, сравниваемые показатели должны относиться к одним и тем же территориальным границам. Аналогичные проблемы возникают в отдельных городах и даже государстве в целом, если их административно-территориальные границы меняются.

Следовательно, прежде чем анализировать динамический ряд, надо, исходя из цели исследования, убедиться в сопоставимости уровней ряда и при отсутствии последней добиваться ее, пользуясь дополнительными расчетами.

Для того чтобы привести уровни ряда динамики к сопоставимому виду, иногда приходится прибегать к приему, который называется **«смыкание рядов динамики».** Под смыканием понимают объединение в один ряд (более длинный) двух или нескольких рядов динамики, уровни которых исчислены по разной методологии или разным территориальным границам. Для осуществления смыкания необходимо, чтобы для одного из периодов (переходного) имелись данные, исчисленные по разной методологии (или в разных границах). Предположим, по одному из промышленных объединений имеются данные о произведенной продукции, методика получения которых в течение рассматриваемого периода претерпела некоторые изменения (табл. 10.2). Для анализа динамики объема продукции за 1991 - 1998 гг. необходимо сомкнуть (объединить) исследуемые два ряда в один. А чтобы уровни нового ряда были сопоставимы, следует пересчитать данные 1991 — 1994 гг. по новой методике. Для этого на основе данных об объеме продукции за 1994 г. в новой и старой методике находим соотношение между ними: 22,8 : 21,2 = 1,1. Умножая на полученный коэффициент данные за 1991 - 1993 гг., приводим их таким образом в сопоставимый вид с

Таблица 10.2 **Динамика объема продукции** (цифры условные)

	1991	1992	1993	1994	1995	1996	1997	1998
Объем продукции, млн руб.:								
по старой методике	19,1	19,7	20,0	21,2	-	-	—	-
по новой методике	-	-	-	22,8	23,6	24,5	26,2	28,1
Сомкнутый (сопоставимый) ряд абсолютных величин, млн руб.	21,0	21,7	22,0	22,8	23,6	24,5	26,2	28,1
Сопоставимый ряд относительных величин, в % к 1994 г.	90,1	92,9	94,3	100,0	103,5	107,5	114,9	123,2

последующими уровнями. Сомкнутый (сопоставимый) ряд динамики показан в предпоследней строке табл. 10.2.

Другой способ смыкания рядов динамики заключается в том, что уровни года, в котором произошли изменения (в нашем примере уровни 1994 г.), как до изменений, так и после изменений (в старой и новой методике, т. е. 21,2 и 22,8) принимаются за 100%, а остальные пересчитываются в процентах по отношению к этим уровням соответственно (в старых ценах - по отношению к 21,2, в новых ценах - к 22,8). В результате получаем сомкнутый ряд динамики, который показан в последней строке табл. 10.2.

Та же проблема приведения к сопоставимому виду возникает и при параллельном анализе развития во времени экономических показателей отдельных стран, административных и территориальных районов. Это, во-первых, вопрос о сопоставимости цен сравниваемых стран, во-вторых, о сопоставимости методики расчета сравниваемых показателей. В таких случаях ряды динамики приводятся к **одному основанию,** т. е. к одному и тому же периоду или моменту времени, уровень которого принимается за базу сравнения, а все остальные уровни выражаются в виде коэффициентов или в процентах по отношению к нему.

Например, имеются данные о производстве цемента в двух странах, млн т (табл. 10.3). Различные значения абсолютных уровней приведенных рядов динамики затрудняют выявление особенностей производства цемента в странах А и Б. Поэтому приведем абсолютные уровни рядов динамики **к общему основанию,** приняв за постоянную базу сравнения уровни 1993 г.; получим следующие данные (в % к 1993 г., табл. 10.4).

Таблица 10.3 **Производство цемента за 1993 - 1997 гг.**

	1993	1994	1995	1996	1997
Страна А	45,5	72,4	95,2	122,0	128,0
Страна Б	56,1	65,1	66,5	65,0	67,0

Таблица 10.4 **Производство цемента за 1993 - 1997 гг.**

	1993	1994	1995	1996	1997
Страна А	100,0	159,1	209,2	268,1	281,3
Страна Б	100,0	116,0	118,5	115,9	119,4

В относительных величинах, выраженных в базисных темпах роста по каждой стране, несопоставимость уровней рядов динамики нивелируется. Различный характер развития выступает более наглядно.

Из данных табл. 10.4 видно, что производство цемента в стране А непрерывно и быстро возрастает, значительно превосходя темпы роста в стране Б.

10.3

ПОКАЗАТЕЛИ ИЗМЕНЕНИЯ УРОВНЕЙ РЯДА ДИНАМИКИ

Анализ скорости и интенсивности развития явления во времени осуществляется с помощью статистических показателей, которые получаются в результате сравнения уровней между собой. К таким показателям относятся: абсолютный прирост, темп роста и прироста, абсолютное значение одного процента прироста. При этом принято сравниваемый уровень называть отчетным, а уровень, с которым производят сравнение, - базисным.

Абсолютный прирост (Ay) характеризует размер увеличения (или уменьшения) уровня ряда за определенный промежуток времени. Он равен разности двух сравниваемых уровней и выражает абсолютную скорость роста:

$$\Delta_{yi} = Y_i - Y_{i-k}, \qquad (Ю.1)$$

где $i = 1, 2, 3, \ldots, n$.

Если k = 1, то уровень $у_{м}$ является предыдущим для данного ряда, а абсолютные приросты изменения уровня будут цепными. Если же k постоянно для данного ряда, то абсолютные приросты будут базисными.

Интенсивность изменения уровня оценивается отношением отчетного уровня к базисному, которое всегда представляет собой положительное число.

Показатель интенсивности изменения уровня ряда в зависимости от того, выражается ли он в виде коэффициента или в процентах, принято называть **коэффициентом роста или темпом роста.** Иными словами, коэффициент роста и темп роста представляют собой две формы выражения интенсивности изменения уровня. Однако необходимо отметить, что не нужно пользоваться одновременно двумя формами, которые по существу идентичны. Разница между ними заключается только в единице измерения.

Коэффициент роста показывает, во сколько раз данный уровень ряда больше базисного уровня (если этот коэффициент больше единицы) или какую часть базисного уровня составляет уровень текущего периода за некоторый промежуток времени (если он меньше единицы). В качестве базисного уровня в зависимости от цели исследования может приниматься какой-то постоянный для всех уровень (часто начальный уровень ряда) либо для каждого последующего предшествующий ему:

$$T_{p_{i/1}} = \frac{Y_i}{Y_1}\cdot 100 \quad \text{или} \quad T_{p_{i/i-1}} = \frac{Y_i}{Y_{i-1}}\cdot 100. \qquad (10.2)$$

В первом случае говорят о базисных темпах роста, во втором о цепных темпах роста.

Наряду с темпом роста можно рассчитать показатель **темпа прироста,** характеризующий относительную скорость изменения уровня ряда в единицу времени. Темп прироста показывает, на какую долю (или процент) уровень данного периода или момента времени больше (или меньше) базисного уровня.

Темп прироста есть отношение абсолютного прироста к уровню ряда, принятого за базу:

Если темп роста всегда положительное число, то темп прироста может быть положительным,

$$T_{пр_{i}} = \frac{\Delta_{i/i-1}}{Y_{i-1}} = \frac{Y_i - Y_{i-1}}{Y_{i-1}}\cdot 100 = \left(K_{p_{i/i-1}} - 1\right)\cdot 100 = T_{pi/i-1} - 100. \quad (10.3)$$

отрицательным и равным нулю.

В статистической практике часто вместо расчета и анализа темпов роста и прироста рассматривают **абсолютное значение одного процента прироста.** Оно представляет собой одну сотую часть базисного уровня и в то же время - отношение абсолютного прироста к соответствующему темпу прироста:

$$|\%| = \frac{\Delta_{i/i-1}}{T_{пр_{i/i-1}}\cdot\%} = \frac{Y_i - Y_{i-1}}{\dfrac{Y_i - Y_{i-1}}{Y_{i-1}}\cdot 100} = \frac{Y_{i-1}}{100} = 0{,}01\cdot Y_{i-1}, \qquad (10.4)$$

где |%| — обозначение абсолютного значения 1% прироста.

Абсолютное значение одного процента прироста служит косвенной мерой базисного уровня и вместе с темпом прироста позволяет рассчитать абсолютный прирост уровня за рассматриваемый период.

Абсолютным ускорением в статистике .называется разность между последующим и предыдущим абсолютными приростами (A' = Ay; - Ay; _ ,). Ускорение показывает, насколько данная скорость больше (меньше) предыдущей.

Таким образом, абсолютное ускорение есть скорость изменения скорости. Оно может быть положительным и отрицательным числом.

Относительным ускорением называется отношение абсолютного ускорения к абсолютному приросту, принятому за базу (A'/Ayj), т. е. относительное ускорение есть темп прироста абсолютного прироста. Оно вычисляется лишь в том случае, если абсолютный прирост, принятый за базу сравнения, число положительное. Например, для ряда 30, 33, 35, 39, 44 абсолютные приросты составят 3, 2, 4, 5; абсолютные ускорения -1,2, 1; относительные ускорения $(-V_3)$-100 = -33,3%; V_2-100 = 100%; V_4-100 = 25%.

Для иллюстрации расчетов рассмотренных статистических показателей приведем следующий ряд динамики в табл. 10.5.

Таблица 10.5

Динамика производства газа в регионе за 1994- 1998 гг. (цифры условные)

Год	Млн $м^3$	Абсолютный прирост, мпн $м^3$		Темп роста, %		Темп прирос-та, %		Абсо-лютное зна чени одного процент прироста. млн $м^3$
		по сравнению с предыдущим годом	по сравнению с 1994 г.	по сравнению с предыдущим годом	по сравнению с 1994г.	по сравнению с предыдущим годом	сравнению с 1994г	
1994	289	—	—	—	100,0	—	—	—
1995	321	32	32	111,1	111,1	11,1	11,1	2,9
1996	346	25	57	107,8	119,7	7,8	19,7	3,2
1997	372	26	83	107,5	128,7	7,5	28,7	3,4
1998	407	35	118	109,4	140,8	9,4	40,8	3,7
Итого	1735	118	-	-	-	-	-	-

Средний уровень ряда динамики (у) рассчитывается по средней хронологической^ **Средней хронологической** называется средняя, исчисленная из значений, изменяющихся во времени. Такие средние обобщают хронологическую вариацию. В хронологической средней отражается совокупность тех условий, в которых развивалось изучаемое явление в данном промежутке времени.

[Методы расчета среднего уровня интервального и моментного рядов динамики различны/Для интервальных рядов с равноотстоящими урЪвнями средний уровень находится по формуле

средней арифметическо^Гпростой, а для неравноотстоящих уровней - по средней арифметической взвешенной:у *t*

Покажем расчет среднего уровня моментного ряда динамики. *Например,* если известны товарные остатки магазина на 1-е число каждого месяца (тыс. руб.) на:

(10.5) $$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n};$$

(10.6) $$\bar{y} = \frac{\sum_{i=1}^{n} y_i \cdot t_i}{\Sigma t_i},$$

где y_i - уровень ряда динамики; п - число уровней; t_i - длительность интервала времени между уровнями.

Так, в табл. 10.5 приведен интервальный ряд динамики с равноотстоящими уровнями. По этим данным можно рассчитать среднегодовой уровень производства газа за 1994 - 1998 гг. Он будет равен 347 млн $м^3$, т. е. у = 1735/5.

Средний уровень моментного ряда динамики так исчислять нельзя, так как отдельные уровни содержат элементы повторного счета. Средний уровень моментного равноотстоящего ряда динамики находится по формуле средней хронологической:

1.I	1.II	1.III	1.IV
18	14	16	20

то среднемесячный товарный остаток за 1 квартал по формуле (10.7) составит:

$$\bar{y} = \frac{18/2+14+16+20/2}{3} = \frac{49}{3} = 16{,}3 \text{ тыс. руб.}$$

Этот же показатель можно получить иначе. При вычислении среднего уровня моментного ряда условно предполагается непрерывное, равномерное изменение уровня в промежутках между двумя датами. Основываясь на этом предположении, определим средние товарные остатки магазина на каждый месяц как полусумму остатков на начало и конец месяца. Средние товарные остатки за месяц:

$$\text{январь} = \frac{18+14}{2} = 16 \text{ тыс. руб.}$$

$$\bar{y} \approx \frac{\frac{y_1+y_2}{2} + \frac{y_2+y_3}{2} + \frac{y_3+y_4}{2} + \ldots + \frac{y_{n-1}+y_n}{2}}{n-1} =$$

$$\approx \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots y_{n-1} + \frac{y_n}{2}}{n-1} \quad \text{или} \quad \bar{y} = \frac{\frac{y_1+y_n}{2} + \sum_{i=2}^{n-1} y_i}{n-1}. \qquad (10.7)$$

Средний уровень моментных рядов динамики с неравноотстоящими уровнями определяется по формуле средней хронологической взвешенной:

$$\text{февраль} = \frac{14+16}{2} = 15 \text{ тыс. руб.}$$

$$\text{март} = \frac{16+20}{2} = 18 \text{ тыс. руб.}$$

Среднемесячный товарный остаток магазина за I квартал в данном случае определяется как простая средняя арифметическая:

$$\bar{y} = \frac{16+15+18}{3} = 16{,}3 \text{ тыс. руб}$$

$$\bar{y} \approx \frac{(y_1+y_2)\cdot t_1 + (y_2+y_3)\cdot t_2 + \ldots + (y_{n-1}+y_n)\cdot t_{n-1}}{2\cdot(t_1+t_2+t_3\ldots+t_{n-1})} = \frac{\Sigma(y_i+y_{i+1})\cdot t_i}{2\cdot\Sigma t_i}, \qquad (10.8)$$

где y_i, y_n - уровни рядов динамики;

t_i - длительность интервала времени между уровнями.

Другой пример. Известна списочная численность рабочих организации на некоторые даты 1998 г. (челрвек) на:

1.I	1.III	1.VI	1.IX	1.I 1999 г.
1200	1100	1250	1500	1350

Среднегодовая численность работников за 1997 г. по формуле (10.8) составит:

$$\bar{y} = \frac{(1200+1100)2 + (1100+1250)3 + (1250+1500)3 + (1500+1350)4}{2(2+3+3+4)} =$$

$$= \frac{31300}{24} \approx 1304 \text{ человека}$$

Обобщающим показателем скорости изменения_явления во времени является **средний абсолютный прирост** (А).

Этот показатель дает возможность установить, насколько в среднем за единицу времени должен увеличиваться уровень ряда (в абсолютном выражении), чтобы, отправляясь от начального уровня за данное число периодов (например, лет), достигнуть конечного уровня. Определяющим свойством интересующего нас показателя среднего абсолютного прироста при такой постановке задачи является общий абсолютный прирост за весь период, ограничивающий ряд динамики. Для его определения воспользуемся формулой средней арифметической простой:

$$\overline{\Delta y} = \frac{\sum_{i=1}^{n-1} \Delta_{i/i-1}}{n-1} \qquad (10.9)$$

или

Ау = ^f п-1

Так, для условий нашего примера (см. табл. 10.2) средний абсолютный прирост равен 29,5 млн м^3 [(407 - 289)74].

Возможен и другой способ расчета среднего абсолютного прироста исходя из кумулятивных данных:

Сводной обобщающей характеристикой интенсивности изменения уровней ряда динамики служит **средний темп роста,** показывающий, во сколько раз в среднем за единицу времени изменился уровень динамического ряда.

Необходимость исчисления среднего темпа роста возникает вследствие того, что темпы роста из года в год колеблются. Кроме того, средний темп роста часто следует определить в тех случаях, когда имеются данные об уровне в начале какого-либо периода и в конце его, а промежуточные данные отсутствуют. Такого рода средний темп роста можно исчислить, если положить в основу расчетов рост не в арифметической прогрессии, которая характеризуется постоянной разностью, а геометрической (a, aq, aq , ... aq'^{1}), характеризующейся постоянным отношением, называемым знаменателем прогрессии (q). Следовательно, вопрос состоит в том, чтобы найти этот знаменатель. Знаменатель геометрической прогрессии (q) определяется делением последующего уровня прогрессии на его предыдущий. При делении п-го уровня на первый получаем:

$$\frac{aq^{n-1}}{a} = q^{n-1},$$

отсюда следует:

$$(10.11)\; q = \sqrt[n-1]{\frac{aq^{n-1}}{a}} = \sqrt[n-1]{\frac{B_n}{B_1}}$$

где В| = а - первый член прогрессии.

Зная q, мы точно можем определить, какую тенденцию развития явления имеет геометрическая последовательность. Формула (10.11) является средней геометрической и применяется в случае, когда определяющий показатель является не суммой значений, а их произведением. Следовательно, если варианты связаны между собой не знаком сложения, а знаком произведения, нужно вычислить среднюю геометрическую. Обычно средний темп роста вычисляется по формуле средней геометрической из цепных коэффициентов роста:

$$(10.10) \quad \overline{\Delta y} = \frac{2\left(\sum_{i=1}^{n} y_i - ny_1\right)}{n(n+1)}.$$

Обе формулы применяются в зависимости от цели исследования.

$$\overline{T}_{p_y} = \sqrt[m]{K_{2/1} \cdot K_{3/2} \ldots K_{n/n-1}} = \sqrt[m]{\Pi K_{pi/i-1}} \cdot \qquad (10.12)$$

Например, средний темп роста производства газа за 1994 1998 гг. (табл. 10.5)

$$\overline{T}_{p_y} = \sqrt[4]{1,111 \cdot 1,078 \cdot 1,075 \cdot 1,094} = \sqrt[4]{1,41} = 1,089, \text{ или } 1$$

Поскольку всякий темп роста является отношением уровней ряда динамики, так что Т 7/, =—=-100;
У|

_ У2 ,пп. m ',,3/9 = — 100..., в формуле
У2

средней геометрической темпы роста заменяются соответствующим отношением уровней. Заменив темпы роста выражающими их отношениями и учитывая, что эти величины перемножаются, найдем подкоренное выражение:

$$\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \frac{y_4}{y_3} \ldots \frac{y_n}{y_{n-1}} = \frac{y_n}{y_1}.$$

Следовательно, средний темп роста может быть выражен формулой

$$\bar{T}_{p_y} = \sqrt[n-1]{\frac{y_n}{y_1}} \quad (10.13)$$

Продолжим расчеты (см. табл. 10.5). Средний темп роста производства газа за 1994 - 1998 гг. по данной формуле будет равен:

$$\bar{T}_{p_y} = \sqrt[4]{\frac{407}{289}} = \sqrt[4]{1,41} = 1,089, \text{ или } 108,9\%.$$

При расчете средних темпов роста по периодам различной продолжительности (разноотстоящие ряды динамики) пользуются средними геометрическими взвешенными по продолжительности периодов. Формула средней геометрической взвешенной будет иметь вид:

$$\bar{T}_{p_y} = \sqrt[\Sigma t]{(K_{2/1})^{t_1} \cdot (K_{3/2})^{t_2} \cdot \ldots \cdot (K_{n/n-1})^{t_n}}, \quad (10.14)$$

где t - интервал времени, в течение которого сохраняется данный темп роста; Zt - сумма отрезков времени периода.

Средний темп прироста не может быть определен непосредственно на основании последовательных темпов прироста или показателей среднего абсолютного прироста. Для его вычисления необходимо вначале найти средний темп роста, а затем уменьшить его на единицу, или 100%.

$$\bar{T}_{пр_y} = \bar{T}_p - 100. \quad (10.15)$$

Для проведения глубокого анализа динамики социально-экономических явлений следует параллельно использовать показатели скорости и интенсивности изменения уровней. Анализ, основанный на использовании какого-либо одного из этих показателей, неизбежно будет иметь односторонний характер.

10.4

КОМПОНЕНТЫ РЯДА ДИНАМИКИ

Ряд динамики может быть подвержен влиянию факторов эволюционного и осциллятивного характера, а также находиться под влиянием факторов разного воздействия.

Влияния эволюционного характера - это изменения, определяющие некое общее направление развития, как бы многолетнюю эволюцию, которая пробивает себе дорогу через другие систематические и случайные колебания. Такие изменения динамического ряда называются **тенденцией развития,** или **трендом.**

Влияния осциллятивного характера - это **циклические (конъюнктурные) и сезонные колебания.** Циклические (или периодические) состоят в том, что значение изучаемого признака в течение какого-то времени возрастает, достигает определенного максимума, затем понижается, достигает определенного минимума, вновь возрастает до прежнего значения и т. д. Иначе циклические колебания можно схематически представить в виде синусоиды у = sint. Циклические колебания в экономических процессах примерно соответствуют так называемым циклам конъюнктуры. Сезонные колебания - это колебания, периодически повторяющиеся в некоторое определенное время каждого года, дня месяца или часа дня. Эти изменения отчетливо наблюдаются на графиках многих рядов динамики, содержащих данные за период не менее одного года.

Наконец, рассмотрим **нерегулярные колебания,** которые для социально-экономических явлений можно разделить на две группы: а) спорадически наступающие изменения, вызванные, например, войной или экологической катастрофой; б) случайные колебания, являющиеся результатом действия большого количества относительно слабых второстепенных факторов.

Следовательно, первоначальные значения ряда динамики подвергаются самым разнообразным воздействиям. Выделим его четыре основные компоненты: основную тенденцию (тренд) (Т), *m*

циклическую или конъюнктурную (K), сезонную (S), случайные колебания (E). Если ряд динамики разбить на различные компоненты, то он представляется в следующем виде:

$$y = f\,(T, K, S, E).$$

В зависимости от взаимосвязи их между собой может быть построена **аддитивная** или **мультипликативная модель** ряда динамики.

Аддитивная модель ряда динамики $y = T + K + S + E$ характеризуется главным образом тем, что характер циклических и сезонных флюктуации (колебаний) остается постоянным (рис. 10.1).

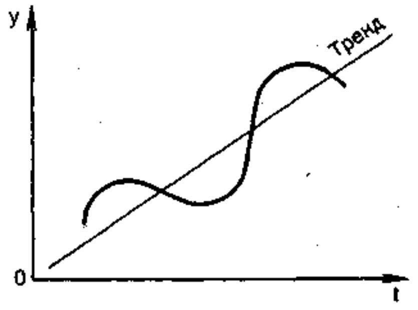


Рис. 10.1. Сочетание различных составляющих ряда динамики при аддитивной связи

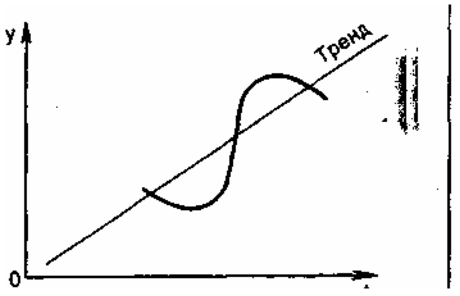


Рис. 10.2. Сочетание различных составляющих ряда динамики при мультипликативной связи

Мультипликативная модель ряда динамики $y = T \cdot K \cdot S + E$. В этой модели характер циклических и сезонных флюктуации остается постоянным только по отношению к тренду (рис. 10.2).

10.5

ВИДЫ ТРЕНДОВОЙ КОМПОНЕНТЫ И ПРОВЕРКА ГИПОТЕЗЫ О СУЩЕСТВОВАНИИ ТЕНДЕНЦИИ

Тренд - это долговременная компонента ряда динамики. Она = характеризует основную тенденцию его развития, при этом остальные компоненты рассматриваются только как мешающие процедуре его определения. При наличии ряда наблюдаемых

значений для различных моментов времени следует найти подходящую трендовую кривую, которая сгладила бы остальные колебания.

В *социально-экономических рядах динамики можно наблюдать тенденции трех видов:*

- среднего уровня;
- дисперсии;
- автокорреляции.

Тенденция среднего уровня аналитически выражается с помощью математической функции, вокруг которой варьируют фактические уровни исследуемого явления. В таком случае значения тренда в отдельные моменты времени будут являться математическими ожиданиями ряда динамики. Часто тенденция среднего уровня называется детерминированной составляющей исследуемого явления, и соответствующий ряд динамики выражается следующим уравнением:

$$y_t = f_t + e_t. \tag{10.16}$$

Тенденция дисперсии представляет собой тенденцию изменения отклонений между эмпирическими уровнями и детерминированной компонентой ряда.

Тенденцией автокорреляции является тенденция изменения связи между отдельными уровнями ряда динамики. Графически это изменение не прослеживается.

Однако прежде чем перейти к выделению тренда, следует проверить гипотезу о том, существует ли он вообще. Отсутствие основной тенденции (тренда) означает неизменность среднего уровня ряда во времени.

В настоящее время известно около десятка критериев для проверки наличия тренда, различающихся как по мощности, так и по сложности математического аппарата. Рассмотрим два из них: метод, основанный на проверке разности средних двух разных частей одного и того же ряда, и метод Фостера-Стюарта.

Проверка существенности разности средних. Ряд динамики разбивается на две равные или почти равные части. Проверяется гипотеза о существовании разности средних: H_0 : у, = y_2.

Воспользуемся методом проверки, разработанным для малых выборок, так как число членов анализируемого ряда, как правило, довольно незначительно. За основу проверки берется toe критерий Стьюдента. При t > tcx гипотеза об отсутствии тренда отвергается, при t < ta гипотеза (H_0) принимается. Здесь t расчетное значение, найденное для анализируемых данных,

tec - табличное значение этого критерия при уровне вероятности ошибки, равном а. В случае равенства или при несущественном различии дисперсий двух исследуемых совокупностей (a^2, = $O^2{}_2$) исчисляется отношение средних с помощью выражения:

(10.17)

$$= \frac{\bar{y}_1 - \bar{y}_2}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

где $\bar{y}_1$ и $\bar{y}_2$ - средние для первой и второй половины ряда динамики;

n_1 и n_2 - число наблюдений в этих частях ряда;

σ - среднее квадратическое отклонение разности средних.

Значение tec берется с числом степеней свободы, равным $n_1 + n_2 - 2$. Необходимое значение а можно определить на основе средней взвешенной величины дисперсий отдельных совокуп-

чгчг"г^й'

$$\sigma = \sqrt{\frac{(n_1 - 1)^2 \cdot \sigma_1^2 + (n_2 - 1)^2 \cdot \sigma_2^2}{n_1 + n_2 - 2}}. \qquad (10.18)$$

поворота тенденции оказывается близкой к середине ряда, поэтому средние двух отрезков будут близки, а проверка может не показать наличие тенденции.

$S = \Sigma S_i; \qquad d =$

где $S_i = U_i + e_i;$

$d_i = U_i - e_j.$

Метод Фостера - Стюарта кроме определения наличия тенденции явления позволяет обнаружить тренд дисперсии уровней ряда динамики, что важно знать при анализе и прогнозировании экономических явлений.

Расчет состоит из следующих этапов.

\. Сравнивается каждый уровень ряда со всеми предыдущими, при этом

если $y_i > y_{i-1}; y_{i-2} \dots, y_1$, то $U_i = 1; e_i = 0;$

при $y_i < y_{i-1}; y_{i-2} \dots, y_1$, то $U_1 = 0; e_j = 1.$

2. Вычисляются значения величин S и d:

(10,20)

При оценивании дисперсий для первой и второй частей ряда динамики $а,^2$ и $о\sim_2{}^2$ возьмем число степеней свободы, равное

$n_1 - 1$ и $n_2 - 1$, соответственно:

$$\sigma_i^2 = \frac{\sum (y_i - \bar{y}_i)^2}{n - 1}, \quad i = 1, 2, \dots, n.$$

Проверка гипотезы о равенстве дисперсий реализуется с помощью F-критерия, который основан на сравнении расчетного отношения с табличным.

$$F = \frac{\sigma_1^2}{\sigma_2^2}, \qquad (10.19)$$

где $\sigma_1{}^2 > \sigma_2{}^2$.

Если расчетное значение F меньше, чем -табличное, при заданном уровне вероятности, то можно принять гипотезу о равенстве дисперсий. Если же F больше, чем табличное значение, то гипотеза о равенстве дисперсий отклоняется и формула (10.17) для испытания разности средних не может быть применена.

Следует заметить, что данный метод дает вполне приемлемые результаты лишь в случае рядов с монотонной тенденцией. Когда же ряд динамики меняет общее направление развития, то точка

Анализируя формулу (10.20), нетрудно заметить, что величина S может принимать значения 0 < S < n - 1, причем S = 0, когда все уровни ряда равны между собой, и S = п - 1, когда ряд динамики монотонно убывает или возрастает. Показатель S характеризует тенденцию изменения дисперсии ряда динамики.

Показатель d имеет нижний предел, равный - (п - 1), и верхний - (п - 1). В первом случае ряд является монотонно убывающим, во втором - монотонно возрастающим. Кроме того, показатель d может быть равен нулю:

- если все уровни ряда равны между собой, тогда ZU; = Ze_r (Данное условие выполняется для ряда, который в первой половине является монотонно убывающим, а во второй - монотонно возрастающим);
- если уровни подъема и спада чередуются, причем каждое следующее значение уровня подъема (спада) должно быть больше (меньше) всех последующих.

Перечисленные случаи, при которых показатель d = 0, представляют лишь теоретический интерес, и вероятность их использования при проведении практических расчетов крайне незначительна.

Из формулы (10.20) определяем
приложения 13 при п = 20, ц = 5,1
Подставляя полученные значения в фо

$$t_s = \frac{7-5,196}{l,677} = l,076; \quad t_d =$$

Показатель d характеризует изменение тенденций в среднем.

Оба показателя, S и d, асимптотически нормальны и имеют независимые распределения.

3. Проверяется с использованием t-критерия Стьюдента гипотеза о том, можно ли считать случайными разности S - ц и d - 0:

$$t_s = \frac{S-\mu}{\sigma_1}; \quad t_d = \frac{d-0}{\sigma_2}, \tag{10.21}$$

где ц - среднее значение величины S, определенное для ряда, в котором уровни расположены случайным образом; о, - стандартная ошибка величины S; O_2 - стандартная ошибка величины d.

Значения величин μ, σ_1 и σ_2 табулированы и приведены в таблице приложения 13.

4. Сравниваются расчетные значения t_s и t_d с табличным при заданном уровне значимости. Если t_s < $t_{ra\&]}$ и t_d < $1_{,,абл}$, то гипотеза об отсутствии тренда в средней и дисперсии подтверждается.

Рассмотрим в качестве примера определение наличия тенденции в ряду динамики производства реализованной продукции на производственном объединении за 1977 - 1997 гг. (табл. 10.6).

Таблица 10.6

Реализованная продукция производственного объединения. Определение Ut и &

Год	Млн руб. Y,	И,	6|	Год	Млн руб. YJ	И,	е(
1977	63,5	0	0	1987	63,0	0	0
1978	62,1	0	1	1989	59,9	0	1
1979	61,6	0	1	1990	62,0	0	0
1980	61,3	0	1	1991	63,4	0	0
1981	61,5	0	0	1992	64,5	0	0
1982	61,3	0	0	1993	58,0	0	1
1983	62,4	0	0	1994	54,5	0	1
1984	65,5	1	0	1995	56,0	0	0
1985	64,8	0	0	1996	55,2	0	0 *
1986	64,3	0	0	1997	56,1	0	0

Ближайшее табличное значение $1_{,\cdot абл}$ для двустороннего критерия при уровне значимости 0,10 равно $1_{габл}$ = 1,725, т. е. $t_{\cdot raf)J1}$ > t_s, *табл $^{<\ 1}$d- Следовательно, гипотеза об отсутствии тенденции в дисперсии показателя реализованной продукции подтвердилась, а в средней - отвергнута.

10.6

МЕТОДЫ АНАЛИЗА ОСНОВНОЙ ТЕНДЕНЦИИ (ТРЕНДА) В РЯДАХ ДИНАМИКИ

После того как установлено наличие тенденции в ряду динамики, производится ее описание с помощью методов сглаживания. *Методы сглаживания разделяются на две основные группы:*

1) сглаживание или механическое выравнивание отдельных членов ряда динамики с использованием фактических значений соседних уровней;

2) выравнивание с применением кривой, проведенной между конкретными уровнями таким образом, чтобы она отображала тенденцию, присущую ряду, и одновременно освободила его от незначительных колебаний.

Рассмотрим каждый из них.

Метод усреднения по левой и правой половине. Разделяют ряд динамики на две части, находят для каждой из них среднее арифметическое значение и проводят через полученные точки линию тренда на графике.

Метод укрупнения интервалов. Если рассматривать уровни экономических показателей за короткие промежутки времени, то в силу влияния различных факторов, действующих в разных направлениях, в рядах динамики наблюдаются снижение и повышение этих уровней. Это мешает видеть основную тенденцию развития изучаемого явления. Поэтому для наглядного представления тренда применяется метод укрупнения интервалов, основанный на укрупнении периодов времени, к которым относятся уровни ряда. Например, ряд ежесуточного выпуска продукции заменяется рядом месячного выпуска продукции и т. д.

Метод простой скользящей средней. Сглаживание ряда динамики с помощью скользящей средней заключается в том, что вычисляется средний уровень из определенного числа первых по порядку уровней ряда, затем - средний уровень из такого же числа уровней, начиная со второго, далее - начиная с третьего и т. д. Таким образом, при расчетах среднего уровня как бы «скользят» по ряду динамики от его начала к концу, каждый раз отбрасывая один уровень в начале и добавляя один следующий. Отсюда название **- скользящая средняя.**

Каждое звено скользящей средней - это средний уровень за соответствующий период, который относится к **середине выбранного периода.**

Для каждого конкретного ряда динамики ($y_,$, y_2, ..., $y_п$) алгоритм расчета скользящей средней следующий.

1. Определить интервал сглаживания, т. е. число входящих в него уровней m (m < n), используя правило: если необходимо сгладить мелкие, беспорядочные колебания, то интервал сглаживания берут по возможности большим, и, наоборот, интервал сглаживания уменьшают, когда нужно сохранить более мелкие волны и освободиться от периодически повторяющихся колебаний, возникающих, например, из-за автокорреляций уровней.

2. Вычислить среднее значение уровней, образующих интервал сглаживания, которое одновременно является сглаживающим значением уровня, находящегося в центре интервала сглаживания, при условии, что m - нечетное число, по одной из формул:

$$y_t = \frac{\sum\limits_{i=t-p}^{t+p} y}{m} \quad \text{или} \quad \bar{y}_t = \bar{y}_{t-1} + \frac{y_{t+p} - y_{t-p-1}}{m}, \qquad (10.22)$$

где

y_i - фактическое значение 1-го уровня;

m - число уровней, входящих в интервал сглаживания $(m = 2p + 1)$;

y_t - текущий уровень ряда динамики;

i - порядковый номер уровня в интервале сглаживания; p - при нечетном m равно: $p = (m - 1) / 2$.

Определение скользящей средней по четному числу членов ряда динамики несколько сложнее, так как средняя может бывь отнесена только к середине между двумя датами, находящимися в середине интервала сглаживания.

Если число членов скользящей средней обозначить через 2т, то серединным будет уровень, относящийся к m + 1/2 члену ряда, т. е. имеет место сдвиг периода, к которому относится уровень. Например, средняя, найденная для четырех членов, относится к середине между вторым и третьим периодами, следующая средняя - к середине между третьим и четвертым, и т. д. Чтобы ликвидировать такой сдвиг, применяют так называемый **способ центрирования. Центрирование** заключается в нахождении средней из двух смежных скользящих средних для отнесения полученного уровня к определенной дате. При центрировании необходимо находить скользящие суммы, скользящие средние нецентрироранные по этим суммам и средние из двух смежных нецентрированных скользящих средних.

3. Сдвинуть интервал сглаживания на одну точку вправо, потом вычислить по формуле (10.22) сглаженное значение для t + 1 члена, снова произвести сдвиг и т. д. В результате последовательного применения приведенной итеративной процедуры получится n - (m - 1) новых сглаженных уровней.

Первые и последние p членов ряда с помощью данного алгоритма сгладить нельзя, так как их значения теряются.

Покажем расчет 5-летней и 4-летней скользящих средних на примере данных табл. 10.7. Как видим, скользящая средняя дает более или менее плавное изменение уровней (рис. 10.3).

Таблица 10.7

Сглаживание урожайности зерновых культур в хозяйстве

Год	Центнеров с 1 га	Скользящие пятилетние суммы	Пятилетние скользящие средние	Скользящие ----- рехлетние суммы	Четырехлетние скользящие средние (нецентрированные)	Четырехлетние скользящие средние (центрированные)	Взвешенные пятилетние скользящие средние
А	1	2	3	4	5	6	7
1982	9,5	_	_	-		• -	-
1983	13.7		-	—	12.3		
1984	12,1	_	12,5	-	13,2	12,8	12,8
1985	14,0	_	13,7	49,3	13,7	13,5	13,4
1986	13,2	63,5	14,1	53,0	14,6	14,1	14,1
1987	15,6	68,6	14,4	54,9	14,6	14,6	14,7
1988	15,4	70,3	15,2	58,2	15,7	15,1	15,1
1989	14,0	72,2	15,6	58,2	15,6	15,6	15,4
1990	17,6	75,8	14,7	62,6	14,5	15,0	15,6
1991	15,4	78,0	15,1	62,4	15,3	14,9	14,9
1992	10,9	73,5	15,3	57,9	14,7	15,0	14,3
1993	17,5	75,4	15,5	61,4	15,5	15,1	15,2
1994	15,0	76,4	15,2	58,8	16,3	15,8	16,2
1995	18,5	77,3	16,0	61,9	15,65	15,97	16,3
1996	14,2	76,1	-	65,2	•	—	~~

Рис. 10.3. Динамика урожайности зерновых культур в хозяйстве за 1982 - 1997 гг.

1997	14,9	80,1	—	62,6	-		

Метод простой скользящей средней вполне приемлем, если графическое изображение ряда динамики напоминает прямую линию. В этом случае не искажается динамика исследуемого явления. Однако когда тренд выравниваемого ряда имеет изгибы и к тому же желательно сохранить мелкие волны, использовать для сглаживания ряда метод простой скользящей средней нецелесообразно, так как простая скользящая средняя может привести к значительным искажениям исследуемого процесса. В таких случаях более надежным является использование взвешенной скользящей средней.

Метод взвешенной скользящей средней. Взвешенная скользящая средняя отличается от простой скользящей средней тем, что уровни, входящие в интервал усреднения, суммируются с различными весами. Это связано с тем, что аппроксимация сглаживаемого ряда динамики в пределах интервала сглаживания осуществляется с использованием уровней, рассчитанных по полиному y~j = a_0 + a, • i + a_2 • i^2 + ... (здесь i - порядковый номер уровня в интервале сглаживания). Полином первого порядка yj = a_0 + a, • i есть уравнение прямой, следовательно, метод простой скользящей средней является частным случаем метода взвешенной скользящей средней. Коэффициенты полиномов находятся по способу наименьших квадратов (гл. 9).

' На первом этапе сглаживания определяются интервал сглаживания и порядок аппроксимирующего полинома - параболы. Считается, что при использовании полиномов высоких степеней и при меньших размерах интервалов сглаживание ряда динамики будет более «гибким».

Центральная ордината параболы принимается за сглаженное значение соответствующего фактическим данным уровня. Поскольку отсчет времени в пределах интервала сглаживания производится от его середины, т. е. (t = i) i =..., -2, -1,0, 1,2, ..., то сглаженное значение уровня равно параметру a_0 подобранной параболы и является соответствующей скользящей средней. Поэтому для сглаживания нет необходимости прибегать к процедуре подбора системы парабол, так как величину a_0 можно получить как взвешенную среднюю из m уровней.

Например, если в интервал сглаживания входят пять последовательных уровней ряда со сдвигом во времени на один шаг, а выравнивание проводится по полиному второго порядка, то коэффициенты полинома находятся из условия

$$\sum_{t=1}^{a}\left(y-a_0-a_1\cdot i-a_2\cdot i^2\right)\to \min.$$

Учитывая, что для нечетных $k\Sigma i^k = 0$, приходим к системе:

$$\begin{cases} \Sigma y-5a_0-a_2\Sigma i^2=0; \\ \Sigma yi-a_1\Sigma i^2=0; \\ \Sigma yi^2-a_0\Sigma i^2-a_2\Sigma i^4=0. \end{cases}$$

Для определения a_0 необходимо найти значения $\sum_{-p}^{p} i^2$ и $\sum_{-p}^{p} i^4$.

$-$

$-p$

$p_{>2}$Так как интервал сглаживания равен m = 5, то Z,' $^{=}$Ю и

$$\sum_{-p}^{p} i^4 = 34.$$

Нормальное уравнение, определяющие a_0 и a_2, в этом случае записываются так:

$$\begin{cases} 5a_0+10a_2=\sum_{-p}^{p} yi; \\ 10a_0+34a_2=\sum_{-p}^{p} y\cdot i^2. \end{cases}$$

Решение этой системы относительно a_0 может быть представ лено следующим образом:

$$a_0 = \frac{34\Sigma y - 10\Sigma y^2}{5 \cdot 34 - 10^2} = \frac{17\Sigma y - 5\Sigma yi^2}{35} =$$

(10.23,

$$= \frac{1}{35}(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{1+2})$$

Аналогичным путем получим выражения и для других интервалов сглаживания по параболе второго и третьего порядка. Так, например, для

$$\begin{cases} m = 7 \ \bar{y}_t = \frac{1}{21}(-2y_{t-3} + 3y_{t-2} + 6y_{t-1} + 7y_t + 6y_{t+1} + 3y_{t+2} - 2y_{t+3}) \\ m = 9 \ \bar{y}_t = \frac{1}{231}(-21y_{t-4} + 14y_{1-3} + 39y_{t-2} + 5y_{t-1} + 59y_t + 54y_{t+1} + \\ \quad 39y_{t+2} + 14y_{1+3} - 21y_{1+4}) \end{cases}$$

Согласно приведенным формулам, веса симметричны относительно центрального уровня (y_t) и их сумма с учетом общего множителя, вынесенного за скобки, равна единице.

По данным рассмотренного выше примера с урожайностью зерновых получим следующие значения взвешенных скользящих средних для m = 5 (табл. 10.7 гр. 7). Пятичленная скользящая средняя показывает, что на протяжении периода с 1982 по 1997 гг. наблюдался рост урожайности зерновых культур.

Выбор уравнения тренда, отображающего развитие социально-экономических явлений во времени. Для отображения основной тенденции развития явлений во времени применяются полиномы разной степени, экспоненты, логистические кривые и другие функции.

Полиномы имеют следующий вид:

- полином первой степени $\bar{y}_t = a_0 + a_1 t$;
- полином второй степени $\bar{y}_t = a_0 + a_1 t + a_2 t^2$;
- полином третьей степени $\bar{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$;
- полином n-й степени $\bar{y}_t = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n$,

где a_0, a,, a_2, ... $a_п$ - параметры полиномов;
t - условное обозначение времени.

В статистической практике параметры полиномов невысокой степени иногда имеют конкретную интерпретацию характеристик динамического ряда. Так, параметр a_0 трактуется как характеристика средних условий ряда динамики, параметры a,, a_2, a_3 - изменение ускорения.

В статистике выработано правило выбора степени полинома модели развития, основанное на определении величин конечных разностей уровней динамических рядов. Согласно этому правилу, полином первой степени (прямая) применяется как модель такого ряда динамики, у которого первые разности (абсолютные приросты) постоянны, полиномы второй степени - для отражения ряда динамики с постоянными вторыми разностями (ускорениями), полиномы третьей степени - с постоянными третьими разностями и т.д.

Для полиноминальных моделей характерно отсутствие прямой связи между абсолютными приростами и приростами уровней рядов динамики.

Предполагаемой функцией, отражающей процесс роста явления, может быть и экспонента $y_t = a_0 a|'$ или $y_t = a_0 \cdot (a,)^ь 1^{1+ь} 2'$. Экспоненты характеризуют прирост, зависящий от величины основания функции.

Отдельные уравнения выражают различные типы динамики. Монотонное возрастание или убывание процесса характеризуют функции: 1) линейная; 2) параболическая; 3) степенная; 4) экспоненциальная простая (показательная) и производная от нее логарифмическая линейная; 5) сложная экспоненциальная и производная от нее логарифмическая парабола; 6) гиперболическая (главным образом убывающих процессов); 7) комбинация их видов.

Для моделирования динамических рядов, проявляющих быстрое развитие в начале ряда и затухающее его развитие к концу, т. е. тех, которые характеризуются стремлением к некоторой предельной величине, применяются **логистические функции.**

Логистическую функцию часто записывают в следующем виде:

$$\bar{y}_t = \frac{K}{1 + C^{-a_0 t}} \quad \text{или} \quad \bar{y}_t = \frac{K}{1 + 10^{a_0 + a_1 t}},$$

менение во времени первых разностей кривой Гомперца асимметрично, а у логистической кривой их изменение во времени имеет симметричный вид, напоминающий нормальное распределение. Для выбора уравнения можно воспользоваться формулой стандартной ошибки

$$S=\sqrt{\frac{\Sigma(y_t-\bar{y}_t)^2}{n-p-1}}, \quad (10.24)$$

где р - число параметров уравнения,

где С - основание натурального логарифма.

Логистическая кривая симметрична относительно точки перегиба и при *t* = — *оо* стремится к нулю, а при t = + °° стремится к некоторой постоянной величине, к которой кривая асимптотически приближается. Если найти вторую производную от y_t по t и приравнять ее к нулю, то для логистической кривой, выражаемой через местоположение точки перегиба кривой, $t = \lg a_1 : a_0$;

$\bar{y}_t = n : 2$.

Тип процессов, характеризующийся наличием экстремальных значений, описывается **кривой Гомперца,** имеющей следующее уравнение:

$$\bar{y}_t = K\cdot(a_0)^{a_1 t}.$$

Возможны четыре варианта этой кривой. Для экономистов наибольшее значение имеет кривая, у которой Iga$_0$ < 0 и а, < 1. Развитие уровня такой кривой имеет следующие этапы. Если коэффициент а, меньше единицы при отрицательном значении IgaQ, то на первом этапе прирост кривой незначителен. Он медленно увеличивается по мере роста t, но на следующем этапе прирост увеличивается быстрее, а затем, после точки перегиба, он начинает уменьшаться, и на подходе к линии асимптоты прирост кривой опять незначителен.

Прологарифмировав функцию Гомперца ;

$$\lg\bar{y}_t = \lg k + \lg a_0 \cdot a_1^{t},$$

получим модифицированную экспоненту. Вводя в модифицированной экспоненте величину, обратную y_t получим логистическую кривую. Следовательно, логистическая кривая имеет сходство с кривой Гомперца. Различие между ними состоит в том, что из-

или применить критерий наименьшей суммы квадратов отклонения эмпирических уровней от теоретических

$$S=\sum_{t=1}^{n}(y_t-\bar{y}_t)^2 \to \min.$$

Из множества возможных уравнений тренда можно выбрать то уравнение, которому соответствует минимальное значение, т. е. критерий наименьших квадратов отклонений, либо использовать формулу средней ошибки аппроксимации:

$$\varepsilon_t = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t-\bar{y}_t|}{y_t}\cdot 100. \quad (10.25)$$

Все эти характеристики имеют один и тот же смысл: показывают, как близко аналитическая функция выравнивания огибает все значения исходного ряда. Поэтому, проводя сравнительную оценку моделей тренда, можно использовать лишь одну из перечисленных характеристик. Результаты такой оценки, полученные на основе прочих характеристик, как правило, совпадают. Наиболее часто в качестве меры точности аппроксимации выбирают остаточную дисперсию или остаточное среднее квадратическое отклонение.

Расчет параметров полиномов различными методами. После того как выяснен характер кривой развития, необходимо определить ее параметры. Элементарный метод определения параметров уравнения тренда, описанного полиномом или экспонентой, состоит в решении системы уравнений **по известным уровням ряда динамики.** Если, например, дан ряд динамики, представленный табл. 10.8, то, приняв условные обозначения времени через t и предположение взять две точки - конечный и начальный уровни, можно построить уравнение прямой по этим двум точкам.

Таблица 10.8 **Динамика производства готовой продукции на фирме**

	1992	1993	1994	1995	1996	1997
Готовая продукция фирмы, тыс. руб.	18	21	26	22	25	28
t	1	2	1 j	4	5	6

Уравнение прямой имеет вид: $\bar{y}_t = a_0 + a_1t$. Для 1992 г. его уровень составил: $a_0 + a_1 \cdot 1 = 18$; для 1997 г.: $a_0 + 6 \cdot a_1 = 28$.

Решая эти уравнения как систему уравнений, получим: $10 = 5a_1$; $a_1 = 2$; $a_0 = 28 - 6a_1 = 28 - 12 = 16$. Следовательно, приближенная модель динамики готовой продукции выражается уравнением $\bar{y}_t = 16 + 2t$. Здесь параметр a_1 соответствует абсолютному приросту.

Можно предположить и развитие по параболе второго порядка: $\bar{y}_t = a_0 + a_1t + a_2t^2$, но тогда следует взять три точки, например 1992, 1995, 1997 гг., т. е. уровни при t = 1, t = 4, t = 6.

Составим систему трех уравнений с тремя неизвестными:

$$\begin{cases} a_0 + 1a_1 + 1a_2 = 18; \\ a_0 + 4a_1 + 16a_2 = 22; \\ a_0 + 6a_1 + 36a_2 = 28. \end{cases}$$

Решая эту систему, получим: $a_0 = 18$, $a_1 = 0{,}3$ и $a_2 = 0{,}3$, а само уравнение применительно к нашему примеру выразится $\bar{y}_t = 18 + 0{,}3t + 0{,}3t^2$, что в приближенной форме определит модель динамики данного явления.

Отрицательным моментом в таком моделировании тренда служат "разные числовые выражения параметров в различных точках их определения.

Другим способом определения параметров уравнения является **метод средних значений (линейных отклонений),** заключающийся в следующем: ряд расчленяется на две примерно равные части и вводится требование, чтобы сумма выравненных значений в каждой части совпала с суммой фактических значений, т. е. чтобы сумма отклонений фактических данных от выравненных равнялась нулю.

В случае выравнивания по прямой линии $y_t = a_0 + a$,t, где a_0 и a, - параметры, получаем:

(10.26) $$\Sigma(y - a_0 - a_1t) = 0,$$

откуда

$$na_0 + a_1\Sigma t = \Sigma y.$$

Если применить это требование к каждой из двух частей ряда, то, вычисляя для каждой части динамического ряда Zt и 2y, получим два уравнения с двумя неизвестными. В результате решения этой системы уравнений находим параметры а() и а,, т. е. начальный уровень и скорость ряда. При этом значение t = 1, 2, 3, 4,..., п.

Разобьем приведенный в табл. 10.7 ряд динамики урожайности зерновых культур на два периода:

1 период - 1982 - 1989 гг.;

И период- 1990 - 1997 гг.,

тогда: £,y = 107,5; E_2y = 124,0; £,t = 45; Z_2t = 91.

Для определения параметров a_0 и *a;* решим систему:

$$\begin{cases} 8a_0 + 45a_1 = 107{,}5; \end{cases}$$

Вычтем из второго уравнения первое. В результате получим:

$$a_1 = 0{,}359; \quad a_0 = 11{,}42.$$

Искомое уравнение будет иметь следующий вид: .

$$\bar{y}_t = 11{,}42 + 0{,}359t.$$

Метод средних значений прост и требует минимального количества вычислений. Его недостаток заключается в том, что при произвольном расчленении ряда на две части мы будем получать разные результаты. Метод средних значений, как и выравнивание ряда динамики с помощью среднего прироста и темпа роста, может применяться для ориентировочных расчетов.

Выравнивание ряда динамики с помощью метода конечных разностей. Этот метод заключается в следующем.

Пусть ряд динамики у, описывается полиномом р-й степени. Для полинома р-й степени вычислим первые разности:

Л/" = Ун. - У„

вторые разности:

Рассмотрим следующий пример (табл. 10.9).

$$\Delta_t^{(2)} = \Delta^{(1)}_{t+1} - \Delta_t^{(1)}$$

и т. д.

Общая формула р-й разности:

$$\Delta_t^{(p)} = y_p + py_{p-1} + \frac{p(p-1)}{2!} \cdot y_{p-2} - \frac{p(p-1)(p-2)}{3!} \cdot y_{p-3} + \ldots + (-1)^p y_0 \cdot (1027)$$

Любой членy_i $(i = 0, 1, 2, 3, ..., n)$ряда динамики можно выразить через начальный уровень ряда y_0 и конечные разности:

$$y_1 = y_0 + \Delta_0^{(1)}; \quad y_2 = y_0 + \Delta_0^{(1)} + \Delta_1^{(1)},$$

но$\Delta_1^{(1)}$ $= y_2 = y_0 + 2\Delta_0^{(1)} + \Delta_0^{(2)}$

и т. д. Отсюда получаем:

$$\bar{y}_i = y_0 + t\Delta_0^{(1)} + \frac{t(t-1)}{2!}\Delta_0^{(2)} + \frac{t(t-1)(t-2)}{3!}\Delta_0^{(3)} + \ldots + \Delta_0^{(t)}. \quad (10.28)$$

Если первые разности не равны, но варьируют с незначительными отклонениями друг от друга, а средняя арифметическая вторых разностей настолько мала, что ею можно пренебречь, то первые разности можно считать практически равными.

Окончательная формула для расчета уровней ряда динамики при равных или почти равных первых разностях будет:

$$\bar{y}_t = \bar{y} + \bar{\Delta}^{(1)} \cdot t^1. \quad (10.29)$$

Если, анализируя вторые разности, мы придем к выводу, что они практически равны, то, вычисляя коэффициенты параболы второго порядка, получаем тренд ряда динамики:

10.9

Таблица

Сглаживание удельного веса прибытия воздушных судов, выполненных без опоздания по сравнению с расписанием одной из авиакомпаний за 1987 — 1997 гг. методом конечных разностей

Год	Удельный вес прибытий, У. %	Условное обозначение време-ни t	Разности		Выравненные значения У,
			ДО)	Д(2)	
1987	91,6	-5	–	–	91,6
1988	91,5	-4	-0,1	-	91,4
1989	91,3	-3	~0,2	-0,1	91,3
1990	91,1	-2	~0,2	0	91,1
1991	91,0	-1	-0,1	0,1	91,0
1992	90,8	0	-0,2	-0,1	90,8
1993	90,6	1	-0,2	0	90,6

1994	90,4	2	-0,2	0	90,4
1995	90,2	3	-0,2	0	90,2
1996	90,0	4	-0,2	0	90,1
1997	89,9	5	-0,1	0,1	89,9
Итого	998,4	0	- 1,7	0,0	998,4
Средние значения У	90,76	—	-0,17	—	—

$$\bar{y}_t = \bar{y} - \frac{n^2-1}{24}\bar{\Delta}^{(2)} + \bar{\Delta}^{(1)} \cdot t^1 + \frac{\bar{\Delta}^{(2)}}{2} \cdot t^2,$$

где $\bar{y}_t$ - выравненное значение ряда динамики;
$\bar{y}$ - средний уровень ряда динамики;
$\bar{\Delta}^{(1)}$ - средняя арифметическая первых разностей;
$\bar{\Delta}^{(2)}$ - средняя арифметическая вторых разностей; п - число уровней;
t - независимая переменная (время).

(10.30)

Способом конечных разностей получили уравнение, выражающее тенденцию динамики данного в ряде процесса:

$$\bar{y}_t = 90{,}76 - 0{,}17t.$$

Метод наименьших квадратов при расчете параметров полиномов. Этот метод при моделировании рядов динамики можно рассматривать как некоторый прием получения оценки

В экономике часто применяется функция вида:

детерминированной компоненты f(t), которая характеризует тренд изучаемого явления.

$$\bar{y}_t = a_0 + \sum_{i=1}^{p} a_i \cdot t^i. \qquad (10.31)$$

Отсюда ад, а,, а$_2$, ..., а в формулах находятся методом наименьших квадратов, сущность которого нам известна из регрессионного анализа.

Согласно этому методу, для нахождения параметров полинома р-й степени необходимо решить систему так называемых нормальных уравнений:

$$\begin{cases} na_0 + a_1\Sigma t + a_2\Sigma t^2 + \ldots + a_p\Sigma t^p = \Sigma y; \\ a_0\Sigma t + a_1\Sigma t^2 + a_2\Sigma t^3 + \ldots + a_p\Sigma t^{p-1} = \Sigma yt; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0\Sigma t^p + a_1\Sigma t^{p+1} + a_2\Sigma t^{p+2} + \ldots + a_p\Sigma t^{2p} = \Sigma yt^p, \end{cases} \qquad (10.32)$$

где п- число членов в ряду динамики;
1 = 1 , 2 , ...

Система (10.32), состоящая из «р» уравнений, содержит в качестве известных величин Еу, Eyt, ..., Hyt^p, т. е. суммы наблюдаемых значений уровней динамического ряда, умноженные на показатели времени в степени 0, 1,2, ..., р, и «р» неизвестных величин а$_;$. Решение этой системы относительно а$_0$, а,, ..., а$_p$ и дает искомые значения параметров.

Системы для расчета параметров полиномов невысоких степеней намного проще. Обозначим последовательные параметры полиномов как а$_0$, а,, а$_2$. Тогда системы нормальных уравнений для оценивания параметров прямой (у$_($ = а$_0$ + а$_;$1) примут вид:

Решение системы (10.33) относительно искомых параметров а$_0$ и а,:

$$\begin{cases} a_0 = \dfrac{\Sigma y \Sigma t^2 - \Sigma t \Sigma yt}{n\Sigma t^2 - (\Sigma t)^2}; \\ a_1 = \dfrac{n\Sigma yt - \Sigma y \Sigma t}{n\Sigma t^2 - (\Sigma t)^2}. \end{cases}$$

Аналогичным путем можно было бы подойти и к системе (10.34). Однако такой путь расчета параметров достаточно трудоемок, если он не выполняется на ЭВМ. Поэтому перейдем к упрощенным приемам расчета параметров, применение которых дает ощутимую экономию труда без какой-либо потери точности результатов.

Упрощенный расчет параметров уравнений заключается в переносе начала координат в середину ряда динамики. В этом случае упрощаются сами нормальные уравнения, кроме того, уменьшаются абсолютные значения величин, участвующих в расчете. В самом деле, если до переноса начала координат t было равно 1, 2, 3, ..., п, то после переноса t = ... -4, -3, -2, -1,0, 1, 2, 3, 4, ..., если число членов ряда нечетное. Когда же число ряда четное, то t = ..., -5, -3, -1, 1 ,3 ,5 , ... Следовательно, Zt и все Zt^p, у которых «р» - нечетное число, равны 0. Таким образом, все члены уравнений, содержащие Xt с такими степенями, могут быть исключены. Системы нормальных уравнений теперь упрощаются для прямой:

$$\begin{cases} a_0 n = \Sigma y; \\ a_1 \Sigma t^2 = \Sigma ty; \end{cases} \qquad (10.35)$$

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

$(\bar{y}_t = a_0 + a_1 t + a_2 t^2)$:

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

(10.33)

(10.34)

для параболы второго порядка:

$$\begin{cases} a_0 n + a_2 \Sigma t^2 = \Sigma y; \\ a_1 \Sigma t^2 = \Sigma ty; \\ a_0 \Sigma t^2 + a_2 \Sigma t^4 = \Sigma t^2 y. \end{cases} \qquad (10.36)$$

Решая системы (10.35), (10.36) относительно неизвестных параметров, получим величины параметров соответствующих полиномов.

Параметр а, выражает начальную скорость роста, а коэффициент a_2 - постоянную скорость изменения прироста. Если уровень явления растет с ускорением, то величина этого ускорения в среднем за изучаемый период равна $2a_2$ единицам.

При сглаживании ряда динамики по экспоненте ($y_t = a_0 e^{a|(}$) для определения параметров применяется метод наименьших квадратов к логарифмам исходных данных. Так, для нахождения параметров экспоненты необходимо решить следующую систему уравнений:

$$\begin{cases} \Sigma \lg y = n \lg a_0 + \lg a_1 \Sigma t; \\ \Sigma t \lg y = \Sigma t \lg a_0 + \lg a_1 \Sigma t^2. \end{cases} \qquad (10.37)$$

Если $\Sigma t = 0$, то параметры уравнения $\lg a_0$ и $\lg a_1$ находим по формулам:

$$\lg a_0 = \frac{\Sigma \lg y}{n}; \quad \lg a_1 = \frac{\Sigma t \lg y}{\Sigma t^2}.$$

Экспонента отражает постоянный относительный рост, равный $e^{a'}$ единицам. Рассмотрим следующий *пример.* Необходимо определить основную тенденцию ряда динамики урожайности зерновых культур в хозяйстве за 1983 - 1997 гг. по данным табл. 10.10.

Начнем определение тенденции с самого простого полинома - уравнения прямой (10.35). Решая систему нормальных уравнений, получим искомые параметры: a_0 = 14,8; а, = 0,17. А само уравнение представим следующим образом: y_t - 14,8 + 0,17t, которое показывает тенденцию динамики урожайности зерновых культур в 1983 - 1997 гг., т. е. в течение исследуемого периода урожайность зерновых культур в хозяйстве увеличилась в среднем на 0,17 ц/га в год.

Во многих случаях моделирование рядов динамики с помощью полиномов или экспоненциальной функции не дает удовлетворительных результатов, так как в рядах динамики содержатся заметные периодические колебания вокруг общей тенденции или наблюдается автокорреляция не в самих уровнях, а в их отклонениях от полученных по определенным аналитическим формулам теоретических значений. В таких случаях следует использовать гармонический анализ.

Целью данного анализа являются выявление и измерение периодических колебаний в рядах динамики и автокорреляции в остатках ряда.

10.10

Таблица

Динамика урожайности зерновых культур в хозяйстве
и определение параметров уравнения
методом наименьших квадратов

Год	Урожай-ность, ц с 1 га, У|	t	t^2	y'	y
1983	13,7	—7	49	-95,9	13,6
1984	' 12,1	-6	36	-72,6	13,8
1985	14,0	-5	25	-70,0	13,9
1986	13,2	-4	16	-52,8	14,1
1987	15,6	[illegible]	9	-46,8	14,3
1988	15,4	-2	4	-30,8	14,5
1989	14,0	-1	1	- 14,0	14,6
1990	17,6	0	0	0'	14,8
1991	15,4	1	1	15,4	15,0
1992	10,9	2	4	21,8	15,1
1993	17,5	[illegible]	9	52,5	15,3
1994	15,0	4	16	60,0	15,5
1995	18,5	5	25	92,5	.15,7
1996	14,2	6	36	85,2	15,8
1997	14,9	7	49	104,3	16,0
Итого	222,0	0	280	48,8	222,0

Функцию, заданную в каждой точке изучаемого интервала времени, можно представить бесконечным рядом синусоидальных и косинусоидальных функций. Нахождение конечной суммы уровней с использованием функций косинусов и синусов времени называется **гармоническим анализом.**

Другими словами, гармонический анализ представляет собой операцию по выражению заданной периодической функции в виде ряда Фурье по гармоникам разных порядков. Каждый член ряда представляет собой слагаемое постоянной величины с функциями косинусов и синусов определенного периода.

В простейшем случае динамика явлений, обладающих периодичностью, может быть аппроксимирована синусоидой:

$$y = A\sin(\alpha t + \beta),$$

где t - время;

A - полуамплитуда колебания, т. е. наибольшее и наименьшее отклонения от оси t;

$t = \pi/\alpha$ - период (длина волны) колебательного движения;

β - начальная фаза колебания.

При t = 0 получаем $y_{t=0} = A\sin\beta$.

Аппроксимация динамики экономических явлений рядом Фурье состоит в выборе таких гармонических колебаний, наложение которых друг на друга (сумма) отражало бы периодические колебания фактических уровней динамического ряда. С помощью ряда Фурье можно представить динамику явлений в виде некоторой функции времени, в которой слагаемые расположены по убыванию периодов:

$$\bar{y}_t = a_0 + \sum_{t=0}^{m}(a_k \cos kt + b_k \sin kt). \quad (10.38)$$

В этом уравнении величина k определяет гармонику ряда Фурье и может быть взята целым числом (чаще всего от 1 до 4). Параметры уравнения рассчитываются методом наименьших квадратов.

Найдя частные производные этой функции и приравняв их к нулю, получим систему нормальных уравнений, из которой вычислим параметры:

1 2 2
a_0 =— 2y; a_k = — Sycoskt; b_k = —Sysinkt. n n n

Последовательные значения t обычно определяются от 0 с увеличением (приростом), равным 2я/п, где n - число уровней ряда динамики.

Для изучения специфического периодического явления - сезонности берется n = 12, по числу месяцев в году.

Тогда ряд динамики годового производства можно записать так:

Ряд Фурье с двумя гармоникам

$$\bar{y}_t = a_0 + a_1 \cos t + b_1 \sin t$$

где $a_2 = \frac{\Sigma y \cos 2t}{6}$; $b_2 = \frac{\Sigma y \sin 2t}{6}$.

$$0;\ \frac{\pi}{6};\ \frac{\pi}{3};\ \frac{\pi}{2};\ \frac{2\pi}{3};\ \frac{5\pi}{6};\ \pi;\ \frac{7\pi}{6};\ \frac{4\pi}{3};\ \frac{3\pi}{2};\ \frac{5\pi}{3};\ \frac{11\pi}{6}.$$

$$y_0;\ y_1;\ y_2;\ y_3;\ y_4;\ y_5;\ y_6;\ y_7;\ y_8;\ y_9;\ y_{10};\ y_{11}.$$

Для определенных в каждом конкретном случае t находят значения синусов и косинусов разных гармоник, которые для удобства располагают в табл. 10.11.

Таблица 10.11

Коэффициент гармонического анализа месячных наблюдений для расчета параметров аи и bk

t	cost	cos2t	cos3t	cos4t	sint	sin2t	sin3t	sin4t
0	1	1	1	1	0	0	0	0
л/6	0,866	0,5	0	-0,5	0,5	0,866	1	0,866
л/3	0,5	-0,5	- 1	-0,5	0,866	0,866	0	- 0,866
л/2	0 •	-1	0	1	1	0	-1	0
2л/3	-0,5	-0,5	1	-0,5	0,866	- 0,866	0	0,866
5л/6	- 0,866	0,5	0	-0,5	0,5	-0,866	1	- 0,866
л	- 1	1	-1	1	0	0	0	0
7л/6	- 0,866	0,5	0	-0,5	-0,5	0,866	-1	0,866
4л/3	-0,5	-0,5	1	-0,5	- 0,866	0,866	0	- 0,866
3л/2	0	- 1	0	1	- 1	0	1	0
5л/3	0,5	-0,5	-1	-0,5	- 0,866	- 0,866	0	0,866
Ил/6	0,866	0,5	0	-0,5	-0,5	- 0,866	-1	-0,866

Полагая гармоники k соответственно равными 1, 2, 3 и т. д., находим все значения coskt и sinkt. Тогда, например, первая гармоника ряда Фурье примет вид:

$$\bar{y}_t = a_0 + a_1\cos t + b_1\sin t,$$

здесь:

$$a_0 = \frac{\Sigma y}{12}; \quad a_1 = \frac{\Sigma y \cos t}{6}; \quad b_1 = \frac{\Sigma y \sin t}{6}. \qquad (10.39)$$

Исчисление параметров ряда Фурье может производиться и другими способами, а также путем использования различных шаблонов.

Пример. Полагая наличие периодичности, проведем гармонический анализ динамики отклонений от линейной тенденции данных об урожайности ярового ячменя в одном из хозяйств на 1986 - 1997 гг. (Uj) (табл. 10.12), Произведем расчеты первой гармоники (для значений синусов и косинусов используем табл. 10.11).

Таблица 10.12 **Отклонения от линейной тенденции данных об урожайности ярового ячменя и расчет параметров ai и bi в модели сезонной волны**

Год	Ц = У,-	t	cost	sint	a, cost	b, sint	Ц
1986	2,1	0	2,1	0	0,81	0	0,81
1987	-2,0	я/6	- 1,732	-1,0	0,701	-0,212	0,49
1988	0,5	я/3	0,25	0,433	0,405	-0,366	0,04
1989	-0,1	я/2	0	0,1	0	- 0,423	- 0,423
1990	2,8	2тс/3	-1,4	2,425	- 0,405	- 0,366	- 0,797
1991	-2,5	5тс/6	2,165	- 1,25	-0,701	-0,212	-0,913
1992	-3,1	я	3,1	0	-0,81	0	-0,81
1993	-2,0	77С/6	1,732	1,0	-0,701	0,121	- 0,489
1994	3,4	471/3	- 1,7	- 2,944	- 0,405	0,366	-0,04
1995	-0,6	3гс/2	0	0,6	0	0,423	0,423
1996	2,6 Л // –1,1	5ТС/3"	1,3-	- 2,252	0,405	0,366	0,771
1997		11 я/6	0,953	0,55	0,701	0,212	0,912
Итого	0,0	-	4,862	-2,538	-	-	0,0

Отсюда можно определить параметры:

$$a_0 = 0; \; a_1 = 0,81; \; b_1 = -0,423.$$

~~(Следовательно, первая гармоника описывается уравнением

$$_1\bar{U}_t = 0{,}81\cos t - 0{,}423\sin t.$$

Аналогично рассчитываются гармоники второго и высших порядков, и значения их последовательно присоединяются к значениям первой гармоники. Запишем уравнения искомых отклонений с учетом второй и третьей гармоник.

Для второй гармоники:

$$_2\bar{U}_t = 0{,}81\cos t - 0{,}423\sin t - 1{,}46\cos 2t - 0{,}27\sin 2t.$$

Для третьей гармоники:

$$_3\bar{U}_t = 0{,}81\cos t - 0{,}42\sin t - 1{,}46\cos 2t - 0{,}27\sin 2t + 1{,}38\cos 3t - 0{,}32\sin 3t.$$

Подставив в уравнение конкретные значения cost, sint, cos2t, sin2t, cos3t, sinSt, получим выравненные уровни отклонений урожайности ярового ячменя за 1986-1997 гг. Затем, рассчитав остаточные дисперсии ($o^{2}_{ост} = £(y_r y_t)^{2}:n$) для трех гармоник, можно сделать вывод, какая гармоника ряда Фурье наиболее близка к фактическим уровням ряда.

10.7

МЕТОДЫ ВЫЯВЛЕНИЯ ПЕРИОДИЧЕСКОЙ КОМПОНЕНТЫ. МОДЕЛИ СЕЗОННЫХ КОЛЕБАНИЙ

Для проверки предположения о существенности периодической компоненты ряда динамики целесообразно использовать такие критерии случайности, которые имеют наибольшую мощность относительно альтернативной гипотезы о цикличности ряда. Наиболее простым для применения и зрительно понятным является критерий «пиков» и «ям». В основе этого критерия лежит подсчет числа экстремальных точек ряда р, который осуществляется следующим образом:

$$\hat{p} \approx \sum_{t=2}^{n-1} p_t,$$

$$где\ \ p_t = \begin{cases} 1,\ если\ y_{t-1} < y_t > y_{t+1} \\ или\ y_{t-1} > y_t < y_{t+1} \\ 0\ в\ противном\ случае \end{cases}$$

$t = 1 + n;$

n – число наблюдений в ряду динамики.

Для случайного ряда математическое ожидание числа экстремальных точек

$$\bar{p} = \frac{2(n-2)}{3}.$$

Проверка гипотезы сводится к сравнению р с расчетным значением р. Если эти значения близки, то можно отказаться от дальнейшей проверки и признать ряд случайным. Если же р и р значительно отличаются друг от друга, то производится дальнейшая проверка гипотезы, основанная на подсчете фаз различной длины.

Фазой называется интервал между двумя соседними уровнями, для которых р, = 1. Для определения **длины фазы** е достаточно просто найти разности индексов двух соседних экстремальных точек. Затем подсчитывается число фаз N„ N_2, N_3 длин е, = 1, $e_2 = 2$, $e_3 = 3$. Теоретическое значение числа фаз длины е для случайного ряда следующее:

$$\hat{N}_e = \frac{2(n-e-2)(e^2-3e+1)}{(e+3)!}.$$

Естественная процедура проверки случайности сводится к сравнению наблюдаемых значений N„ N_2, N_3 с теоретическим значением N_e. Однако при небольшом числе наблюдений п критерий %^{2} здесь непосредственно использовать нельзя, так как в этом случае длины фаз е, не являются независимыми. Доказано, что при разбиении длины фазы на три группы: е = 1, е = 2, е =3 (две степени свободы) -

статистика /^{2} может быть использована в обычной форме (V = 2,5) при %^{2} = 6,3. Расчетные значения /^{2} в случае трех групп длин фазы определяются по формуле

$$\chi^2=\sum_{e=1}^{3}\frac{(N_e-\hat{N}_e)}{N_e}. \quad (10.42)$$

Если %^{2} > 6,3, то колебания исходного ряда нельзя считать чисто случайными и ряд содержит периодическую составляющую. Этот критерий весьма чувствителен к периодическим колебаниям и имеет практически нулевую эффективность относительно альтернативы наличия тренда, поэтому он может применяться непосредственно к исходному ряду динамики в отличие от других критериев, которые требуют, чтобы из ряда динамики предварительно была выделена систематическая составляющая.

После того как установлена периодическая составляющая, производится ее анализ.

При рассмотрении квартальных или месячных данных многих социально-экономических явлений часто обнаруживаются определенные, постоянно повторяющиеся колебания, которые существенно не изменяются за длительный период времени. Они являются результатом влияния природно-климатических условий, общих экономических факторов, а также ряда многочисленных разнообразных факторов, которые частично являются регулируемыми. В статистике периодические колебания, которые имеют определенный и постоянный период, равный годовому промежутку, носят название «сезонные колебания», или «сезонные волны», а динамический ряд в этом случае называют **тренд-сезонным,** или просто **сезонным рядом динамики.**

Сезонные колебания характеризуются специальными показателями, которые называются **индексами сезонности** (I_s). Совокупность этих показателей отражает сезонную волну. Индексами сезонности являются процентные отношения фактических внутригодовых уровней к постоянной или переменной средней.

Для выявления сезонных колебаний обычно берут данные за несколько лет, распределенные по месяцам. Данные за несколько лет (не менее трех) используют для того, чтобы выявить устойчивую сезонную волну, на которой не отражались бы случайные условия одного года.

Для вычисления индексов сезонности применяются различные методы (табл. 10.15). Если ряд динамики не содержит ярко выраженной тенденции в развитии, то индексы сезонности вычисляются непосредственно по эмпирическим данным без их предварительного выравнивания.

Для каждого месяца рассчитывается средняя величина уровня, например за три года (у;), затем из них вычисляется среднемесячный уровень для всего ряда (у), и в заключение определяется процентное отношение средних для каждого месяца к общему среднемесячному уровню ряда, т. е.

$$(10.43) \quad I_S=\frac{\bar{y}_i}{\bar{y}}100\%.$$

Пример расчета индексов сезонности покажем на месячных данных о внутригодовой динамике числа браков, расторгнутых населением города за 1996-1998 гг., представленных в табл. 10.13.

Таблица 10.13

Динамика браков, расторгнутых населением города, и расчет индексов сезонности

Месяц	Число расторгнутых браков				Индекс сезонности (у,: у) 100,%
	1996	1997	1998	в среднем за три года $\bar{}$ У:	
А	1	*2*	3	4	5
Январь	195	158	144	165,7	122,4
Февраль	164	141	136	147,0	108,6
Март	153	153	146	150,7	111,3
Апрель	136	140	132	136,0	100,4
Май	136	136	136	136,0	100,4
Июнь	123	129	125	125,7	92,8
Июль	126	128	124	126,0	93,1
Август	121	122	119	120,7	89,1
Сентябрь	118	118	118	118,0	87,2
Октябрь	126	130	128	128,0	94,5

Ноябрь Декабрь	129 138	131 141	135 139	131,7 139,3	97,3 102,9
Средний уровень ряда У	138,77	135,6	131,8	у= 135,4	100,0

По данным табл. 10.13 вычислим усредненные значения уровней по одноименным периодам способом средней арифметической простой:

$$\text{январь: } \bar{y}_1 = \frac{195+158+144}{3} = \frac{497}{3} = 165,7;$$

$$\text{февраль: } \bar{y}_2 = \frac{164+141+136}{3} = \frac{441}{3} = 147,0 \text{ и т. д. (гр. 4 табл. 10.13).}$$

Затем по вычисленным помесячным средним уровням (у;) определим общий средний уровень (у):

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{1624,8}{12} = 135,4$$

или

$$\bar{y} = \frac{\Sigma(\bar{y}_i)}{m} = \frac{406,1}{3} = 135,4.$$

где m - число лет; £(У|) ~ сумма среднегодовых уровней ряда динамики.

Далее рассчитаем по месяцам года индексы сезонности:

$$\text{январь: } I_{S_1} = \frac{165,7}{135,4}\cdot 100 = 122,4\%;$$

$$\text{февраль: } I_{S_2} = \frac{147,0}{135,4}\cdot 100 = 108,6\% \text{ и т. д. (гр. 5 табл. 10.13).}$$

Совокупность исчисленных индексов сезонности характеризует сезонную волну развития числа браков, расторгнутых населением города, во внутригодовой динамике. Для наглядного получения представления о сезонной волне желательно изобразить полученные данные в виде линейной диаграммы.

Если же ряд динамики содержит определенную тенденцию в развитии, то, прежде чем вычислить сезонную волну, фактические данные должны быть обработаны так, чтобы была выявлена общая тенденция. Обычно для этого прибегают к аналитическому выравниванию ряда динамики.

При использовании способа аналитического выравнивания ход вычислений индексов сезонности следующий (табл. 10.14):

- по соответствующему полиному вычисляются для каждого месяца (квартала) выравненные уровни на момент времени (t) (гр. 2);
- определяются отношения фактических месячных (квартальных) данных (у;) к соответствующим выравненным данным (у,) в процентах (гр. 3);

$$I_i = (y_i : \bar{y}_t) \cdot 100;$$

- находятся средние арифметические из процентных соотношений, рассчитанных по одноименным периодам в процентах (гр. 4):

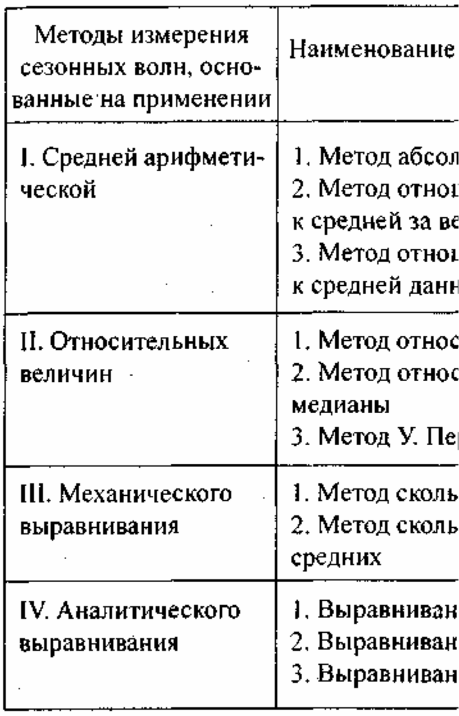
$$I_i = (I_1+I_2+I_3+...+I_n) : n,$$

где п - число одноименных периодов.

Таблица 10.14

Динамика поквартальной продажи безалкогольных напитков в одной из республик за 1995 - 1997 гг. и расчет сезонной волны

Годи квартал	Млн дкл У	Теорети-ческие уровни у,= 88,3 + 0,1 3t	Индекс сезонности по каждому кварталу года (У. :уО- 1QO	Индекс сезонности по одноименным кварталам Е(у, :у,)- 100 п
А	1	2	3	4
1995				
I	58,4	86,2	67,8	67,6
II	125,6	86,6	145,0	140,4
III	108,1	87,0	124,3	121,7
IV	60,8	87,3	69,6	70,3
1996				
I	57,7	87,7	6.5,8	67,6
II	115,4	88,1	131,0	140,4
III	103,9	88,5	117,4	121,7
IV	60,6	88,9	68,2	70,3,
1997				
I	61,8	89,2	69,3	67,6
II	130,2	89,7	145,2	140,4
III	111,0	90,0	123,3	121,7
IV	66,1	90,4	73,1	70,3
Итого	1059,6	1059,6	1 200	1 200



Методы измерения сезонных волн, основанные на применении	Наименование
I. Средней арифметической	1. Метод абсол 2. Метод отно к средней за ве 3. Метод отно к средней данн
II. Относительных величин	1. Метод относ 2. Метод относ медианы 3. Метод У. Пе
III. Механического выравнивания	1. Метод сколь 2. Метод сколь средних
IV. Аналитического выравнивания	1. Выравниван 2. Выравниван 3. Выравниван

В общем виде формулу расчета индекса сезонности данным способом можно записать так:
по месячным данным равна 1200, а сумма по четырем кварталам - 400.

В результате проведенных расчетов в табл. 10.14 получили ряд индексов (гр. 4), характеризующих сезонную волну продажи безалкогольных напитков (в процентах к среднегодовой продаже, принятой за 100%) по кварталам.

В табл. 10.15 дается классификация наиболее распространенных методов измерения сезонных волн.

Таблица 10.15 **Классификация методов измерения сезонных волн**

$$I_S = \left[\Sigma \frac{y_i}{\bar{y}_t}\right] : n. \quad (10.44)$$

Расчет заканчивается проверкой правильности вычислений индексов. Так как средний индекс сезонности для всех месяцев (кварталов) должен быть 100%, то сумма полученных индексов

Подобно сезонной компоненте ряда динамики **циклическая** компонента также представляет собой волнообразные движения (на графике), но она более продолжительна и менее предсказуема, чем сезонные колебания. Сущность классического метода устранения циклической компоненты ряда динамики заключается в исключении (или в усреднении) основной тенденции и сезонной компоненты из ряда динамики, так как при этом остается циклическая и, как правило, **нерегулярная** компонента. Поскольку эти компоненты составляют то, что остается после подобных расчетов, этот метод называется **остаточным.**

Экономисты уделяют большое внимание анализу деловых циклов и их причинам, но мы не ставим своей целью рассмотрение многочисленных теорий этого анализа.

10.8

РЕГРЕССИОННЫЙ АНАЛИЗ СВЯЗНЫХ ДИНАМИЧЕСКИХ РЯДОВ

Многомерные временные ряды, показывающие зависимость результативного признака от одного или нескольких факторных, называют **связными рядами динамики.** Применение методов наименьших квадратов для обработки рядов динамики не требует выдвижения никаких предположений о законах распределения исходных данных. Однако при использовании метода наименьших квадратов для обработки связных рядов следует учитывать наличие автокорреляции (авторегрессии), которая не учитывалась при обработке одномерных рядов динамики, поскольку ее наличие способствовало более плотному и четкому выявлению тенденции развития рассматриваемого социально-экономического явления во времени.

В значительной части рядов динамики экономических процессов между уровнями, особенно близко расположенными, существует взаимосвязь. Ее удобно представить в виде корреляционной зависимости между рядами у,, y_2, *у-$,...,* $y_п$ и этим же рядом, сдвинутым относительно первоначального положения на h моментов времени $y_{,+h}$, y_{2+h}, y_{3+h}, ..., y_{n+h}. Временное смещение L называется **сдвигом,** а само явление взаимосвязи - **автокорреляцией.**

Автокорреляционная зависимость особенно существенна между последующими и предшествующими уровнями ряда динамики. Поскольку классические методы математической статистики применимы лишь в случае независимости отдельных членов ряда между собой, то при анализе нескольких взаимосвязанных рядов динамики важно установить наличие и степень их автокорреляции.

Различаются два вида автокорреляции:

1) автокорреляция в наблюдениях за одной или более переменными;

2) автокорреляция ошибок или автокорреляция в отклонениях от тренда.

Наличие последней приводит к искажению величин средних квадратических ошибок коэффициентов регрессии, что затрудняет построение доверительных интервалов для коэффициентов регрессии, а также проверку их значимости.

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - (\bar{y}_t)^2}{\sigma^2_{y_t}} \quad \text{или} \quad r_a$$

Автокорреляцию измеряют при помощи **нециклического коэффициента автокорреляции,** который может рассчитываться не только между соседними уровнями, т. е. сдвинутыми на один период, но и между сдвинутыми на любое число единиц времени (L). Этот сдвиг, именуемый **временным лагом,** определяет и порядок коэффициентов автокорреляции: первого порядка (при L = 1), второго порядка (при L = 2) и т.д. Однако наибольший интерес для исследования представляет вычисление нециклического коэффициента (первого порядка), так как наиболее сильные искажения результатов анализа возникают при корреляции между исходными уровнями ряда (у$_{t|}$) и теми же уровнями, сдвинутыми на одну единицу времени, т. е. у$_{t\_(}$ (или у$_{t+)}$).

Тогда формулу коэффициента автокорреляции можно записать следующим образом:

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - \bar{y}_t \cdot \bar{y}_{t+1}}{\sigma_{y_t} \cdot \sigma_{y_{t+1}}}, \qquad (10.45)$$

где О"$_{yt}$, О"$_{yi+1}$- среднее квадратическое отклонение рядов у, и у$_{(н}$ соответственно.

Если значение последнего уровня (у$_п$) ряда мало отличается от первого (у,), то сдвинутый ряд не укорачивается, его можно условно дополнить, принимая у$_п$ = у,. Тогда у, = у,+, и а$_{у|}$ = а.,$_ж$, поскольку рассчитываются они для одного и того же ряда. При такой замене, т. е. если у$_($ = у$_{t+1}$ и а ~ о$_{уж}$, формула коэффициента автокорреляции примет вид:

$$r_a = \frac{\sum\limits_{t=1}^{n-1} y_t \cdot y_{t+1}}{\sum\limits_{t=1}^{n} y_t^2}. \qquad (10.47)$$

Если ряд динамики состоит из уровней, среднее значение которых равно нулю (у = 0), то выражение (Ш.46) значительно упрощается:

Для суждения о наличии или отсутствии автокорреляции в исследуемом ряду фактическое значение коэффициентов автокорреляции сопоставляется с табличным (критическим) для 5%-ного или 1%-ного уровня значимости (вероятности допустить ошибку при принятии нулевой гипотезы о независимости уровней ряда).

Одна из специальных таблиц, в которой определена критическая область проверяемой гипотезы (об отсутствии автокорреляции), составленная Р. Андерсеном в 1942 г., приведена в приложении 12.

Если фактическое значение коэффициента автокорреляции меньше табличного, то гипотеза об отсутствии автокорреляции в ряду может быть принята. Когда же фактическое значение больше табличного, можно сделать вывод о наличии автокорреляции в ряду динамики.

Покажем расчет коэффициента автокорреляции первого порядка на примере данных об объеме собственной продукции общественного питания в одной из областей за 1988 - 1997 гг. Исходные данные и расчет необходимых величин для формулы (10.46) представлены в табл. 10.16.

По итоговым данным табл. 10.16 находим

1) $\overline{y_t y_{t+1}} = \frac{49,20}{10} = 4,92;$

2) $\bar{y} = \frac{21,7}{10} = 2,17;$

3) $\overline{y^2} = 4,71;$

4) $\overline{y_t^2} = \frac{51,47}{10} = 5,15;$

5) $\sigma^2_{y_t} = \overline{y^2_t} - \bar{y}^2 = 5,15$ -4,71 = 0,44 . и их значения подставляем в формулу

коэффициента автокорре-
псгпыы'

$$r_a = \frac{4,92 - 4,71}{0,44} = 0,48.$$

где $e_t = y_t - \bar{y}_t$.

Таблица 10.16

Динамика объема собственной продукции общественного питания в одной из областей за 1988 - 1997 гг. и расчет коэффициента автокорреляции

Год	Собственная продукция, млн pvб. У,	Собственная продукция со сдвигом в 1 год $у_{,+|}$	Расчетные величины	
			$у.у_{,+|}$	$У^2$,
1988	1,3	1,4	1,82	1,69
1989	1,4	1,5	2,10	1,96
1990	1,5	1,7	2,55	2,25
1991	1,7	2,1	3,57	2,89
1992	2,1	2,2	4,62	4,41
1993	2,2	2,5	5,50	4,84
1994	2,5	2,7	6,75	6,25
1995	2,7	3,0	8,10	7,29
1996	3,0	3,3	9,90	9,00
1997	3,3	1,3	4,29	10,89
Итого	21,7	21,7	49,20	51,47

Приведем сопоставление полученного коэффициента автокорреляции с табличной величиной при численности п = 10. При уровне значимости Р = 0,05 (5%-ный уровень) величина r_a может только в пяти случаях из ста превышать 0,36 (см. таблицу в приложении 12). Следовательно, фактический коэффициент автокорреляции, равный 0,48, превысил табличное значение, что говорит о наличии автокорреляции в ряду динамики собственной продукции.

Для уменьшения автокорреляции применяют различные методы. Почти все они преследуют цель исключения основной тенденции (тренда) из первоначальных данных.

Наиболее распространенным примером выявления наличия автокорреляции'Ъ-отклонениях от тренда или от регрессионной модели является использование **критерия Дарбина - Уотсона,/** который рассчитывается по формуле *

$$d = \frac{\sum_{t=1}^{n}(e_{t+1} - e_t)^2}{\sum_{1}^{n} e_t^2}, \qquad (10.48)$$

Теоретическое основание применения этого критерия обусловлено тем, что в динамических рядах как сами наблюдения, так и отклонения от них распределяются в хронологическом порядке.

При условии, что отклонения уровней от тенденции (так называемые остатки) случайны, значения D, лежащие в интервале О - 4, всегда будут находиться ближе к 2. Если автокорреляция положительная, то D < 2; отрицательная - 2< = D < = 4. Следовательно, оценки, получаемые по критерию, являются не точечными, а интервальными. Их значения для трех уровней значимости (а = 0,01, ос = 0,025 и а = 0,05) с учетом числа наблюдений даны в специальных таблицах.

Рассчитаем критерий Дарбина - Уотсона по данным табл. 10.17.

Таблица 10.17

Динамика среднегодовой стоимости основных промышленно-производственных фондов объединения за 1982 - 1997 гг.

Год	Млн руб. У	t	t^2	yt	y_t	e_t	e_{t+1}	e^2,	$e_{t+1}-e_t$	$(e_{t+1}-e_t)^2$
1982	47	– 15	225	- 705	47,6	-0,6	-0,4	0,36	-0,2	0,04
1983	51	-13	169	-	51,4	-0,4	-0,1	0,16	0,3	0,09
1984	55	-11	121	-	55,1	-0,1	0,1	0,01	0,2	0,04
1985	59	-9	81	-	58,9	0,1	-0,6	0,01	-0,7	0,49
1986	62	-7	49	-	62,6	-0,6	-0,4	0,36	0,2	0,04
1987	66	-5	25	-	66,4	-0,4	-0,2	0,16	0,2	0,04
1988	70	-3	9	-	70,2	-0,2	1,1	0,04	1,3	1,69
1989	75	-1	)	-75	73,9	1,1	1,3	1,21	0,2	0,04
1990	79	1	1	79	77,7	1,3	0,6	1,69	-0,7	0,49
1991	82	3	9	246	81,4	0,6	0,8	0,36	0,2	0,04
1992	86	5	25	430	85,2	0,8	0	0,64	-0,8	0,64
1993	89	7	49	623	89,0	0	0,7	0	0,7	0,49

1994	92	9	81	828	82,7	-0,7	-0,5	0,49	0,2	0,04
1995	96	11	121	1 056	96,5	-0,5	-0,2	0,25	0,3	0,09
1996	100	13	169	1 300	100,2	-0,7	-0,1	0,04	0,3	0,09
1997	103	15	225	1 545	104,0	-1,0		1,0	-	-
	1 212	-	1 360	255 4	-	-	-	6,78	-	4,35

$$\bar{y}_t = a_0 + a_1 t;$$

$$\begin{cases} na_0 = \Sigma y; \\ a_1 \Sigma t^2 = \Sigma ty; \end{cases}$$

$$\begin{cases} 16a_0 = 1212; \\ 1360a_1 = 2554; \end{cases}$$

$$\begin{cases} a_0 = 75{,}8; \\ a_1 = 1{,}88. \end{cases}$$

Следовательно, $y_t = 75{,}8 + 1{,}88t$.

Величина критерия Дарбина - Уотсона D = 0,64 (4,35 : 6,78), т. е. D < 2, и это подтверждает наличие в исходном динамическом ряду положительной автокорреляции.

Существует ряд способов исключения или уменьшения автокорреляции (авторегрессии) в рядах динамики: а) метод включения времени в качестве дополнительного фактора; б) метод последовательных разностей; в) метод авторегрессионных преобразований.

Рассмотрим эти способы исключения автокорреляции (авторегрессии).

В соответствии с теоремой, доказанной Фришем и Воу, время вводится в систему связных динамических рядов в явной форме в качестве дополнительного фактора. Уровни исходных динамических рядов могут быть представлены показателями в любой форме, в том числе логарифмической, а время всегда вводится в линейной форме. Считается, что введение фактора времени исключает основную тенденцию развития всех явлений, представленных исследуемыми рядами динамики. Доказано, что введение времени аналогично использованию отклонения фактических данных от трендов.

Применение метода наименьших квадратов к обработке многомерных временных рядов не отличается от методологии применения его к обычным статистическим рядам. В рассматриваемом случае минимизируется следующее выражение:

$$S = \Sigma[y_i - f(x_1 x_2, \ldots, x_n, t)]^2 \rightarrow \min.$$

Рассмотрим *пример.* По условным данным о реализованной продукции и накладных расходах на реализацию найдем линейное уравнение связи.и рассчитаем неизвестные параметры (табл. 10.18).

Таблица 10.18 Расчет параметров уравнения регрессии

ются не сами уровни исходных рядов y_t, y_{t+1}, ..., y_{t+n}, и x_t, x_{t+},, x_{t+n}, а последовательные разности между ними:

Год	Реали-зованная про-дукция, млн X	Наклад-ные расходы млн. руб. У	t	xy	x^2	ty	t^2	tx	ух.
1989	9	27	1	243	81	27	1	9	30,5
1990	13	36	2	468	169	72	4	26	32,4
1991	17	29	3	493	289	87	9	51	34,2
1992	22	41	4	902	484	164	16	88	38,7
1993	29	54	5	1 566	841	270	25	145	48,3
1994	36	71	6	2566	1 296	426	36	216	58,0
1995	44	50	7	2200	1 936	350	49	308,	70,3

1996	51	81	8	4 131	2601	648	64	408	79,9
1997	60	98	9	5880	3600	882	81	540	94,8
	281	487	45	18439	11 297	2926	285	791	487,1

$$\begin{cases} na_0 + a_1\Sigma x + a_2\Sigma t = \Sigma y; \\ a_0\Sigma x + a_1\Sigma x^2 + a_2\Sigma xt = \Sigma yx; \\ a_0\Sigma t + a_1\Sigma xt + a_2\Sigma t^2 = \Sigma yt; \end{cases}$$

$$\begin{cases} 9a_0 + 281a_1 + 43a_2 = 487; \\ 281a_0 + 11287a_1 + 1791a_2 = 18439; \\ 45a_0 + 1791a_1 + 285a_2 = 2926; \end{cases}$$

$$\begin{cases} a_0 = 15,63; \\ a_1 = 2,61; \\ a_2 = -8,60; \end{cases}$$

$$\bar{y}_{xt} = 15,63 + 2,61x - 8,6t.$$

При исключении автокорреляции методом последовательных разностей обработке методом наименьших квадратов подверга-

$$\Delta y_1 = y_t - y_{t-1}; \qquad \Delta x_t = x_t - x_{t-1};$$

$$\Delta y_2 = y_{t-1} - y_{t-2}; \qquad \Delta x_2 = x_{t-1} - x_{t-2};$$

$$\ldots\ldots\ldots\ldots \qquad \ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots \qquad \ldots\ldots\ldots\ldots$$

$$\Delta y_k = y_{t-k} - y_{t-k-1}; \qquad \Delta x_k = x_{t-k} - x_{t-k-1}.$$

При использовании этого метода исходят из предположения, что все разности между уровнями динамических рядов, начиная с первой, будут содержать только случайную компоненту. Причем первые разности содержат случайную компоненту в линейной форме, вторые - описываемую параболой второго порядка, третьи - показательной функцией.

Метод авторегрессионных преобразований заключается в том, что определяют уравнение связи между отклонениями от тенденций двух связных рядов динамики:

$$y_1 - \bar{y}_{t1}; \qquad x_1 - \bar{x}_{t1};$$

$$y_2 - \bar{y}_{t2}; \qquad x_2 - \bar{x}_{t2};$$

$$\ldots\ldots \qquad \ldots\ldots$$

$$\ldots\ldots \qquad \ldots\ldots$$

$$y_n = \bar{y}_{tn}; \qquad x_n - \bar{x}_{tn}.$$

В этом случае также получают уравнения регрессии, не искаженные влиянием автокорреляции.

Введение времени в качестве дополнительной переменной является наиболее действенным способом обработки связных рядов динамики. Во всяком случае при линейной связи между исследуемыми рядами этот способ более точен, чем использование последовательных разностей или отклонений от трендов.

При обработке методом наименьших квадратов последовательных разностей или отклонений от трендов исследователь имеет дело с чисто случайными величинами, взаимосвязь между которыми является часто весьма сомнительной, так как исключение в обоих случаях тенденций нарушило бы существование между явлениями причинно-следственной связи.

10.9

КОРРЕЛЯЦИЯ РЯДОВ ДИНАМИКИ

При изучении развития явления во времени часто возникает необходимость оценить степень взаимосвязи в изменениях уровней двух или более рядов динамики различного содержания, но связанных между собой. Эта задача решается методами коррелирования: 1) уровней ряда динамики; 2) отклонений фактических уровней от тренда; 3) последовательных разностей, т. е. путем исчисления парного коэффициента корреляции.

Коррелирование уровней ряда динамики правильно показывает тесноту связи между рядами динамики лишь в том случае, если в каждом из них отсутствует автокорреляция.

В этом случае величину коэффициента корреляции находят по формуле

$$r = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x \cdot \sigma_y}, \qquad (10.49)$$

где x_i - уровни факторного ряда динамики;

y_i - уровни результативного ряда динамики.

Следовательно, прежде чем коррелировать ряды динамики (по уровням), необходимо проверить каждый из рядов на наличие или отсутствие в них автокорреляции (при помощи коэффициента автокорреляции, описанного в предыдущем параграфе). В случае наличия автокорреляции между уровнями ряда последняя должна быть устранена.

Рассмотрим способы ее исключения в рядах динамики.

Коррелирование отклонений от выравненных уровней (тренда). Этот способ состоит в том, что коррелируют не сами уровни, а отклонения фактических уровней от выравненных, отражающих тренд, т. е. коррелируют остаточные величины. Для этого каждый ряд динамики выравнивают по определенной, характерной для него аналитической формуле, затем из эмпирических уровней вычитают выравненные (т. е. находят $d_x = x_t - x_t$; $d_v = y_t - y_t$) и определяют тесноту связи между рассчитанными отклонениями ($d._x$ и d) по формуле

$$r_{d_x d_y} = \frac{\Sigma d_x \cdot d_y}{\sqrt{\Sigma d_x^2 \cdot \Sigma d_y^2}}. \qquad (10.50)$$

Коррелирование последовательных разностей. Исключить влияние автокорреляции можно путем вычитания из каждого уровня предшествующего ему, т. е. находя разности уровней (У| - У|_|). Алгебраически легко показать, что при переходе от уровней к их разностям исключается влияние общей тенденции на колеблемость. При этом при изменении уровней по прямой можно коррелировать первые разности, при изменении по параболе п-го порядка - n-е разности. Формула коэффициента разностей, используемая для измерения тесноты связи между исследуемыми рядами, имеет вид:

$$r_{\Delta_x \Delta_y} = \frac{\Sigma \Delta_x \cdot \Delta_y}{\sqrt{\Sigma \Delta_x^2 \cdot \Sigma \Delta_y^2}}. \qquad (10.51)$$

Коэффициент корреляции, рассчитанный для измерения тесноты зависимости изменения уровней двух рядов, является своего рода средним, обобщающим показателем. Однако для длительного периода эта зависимость не является постоянной, она может меняться во времени. Поэтому чтобы судить о том, в какие периоды зависимость между изменениями уровней двух рядов слабая или сильная, рекомендуется рассчитывать серию скользящих коэффициентов корреляции для определенного интервала времени.

10.10

ЭЛЕМЕНТЫ ПРОГНОЗИРОВАНИЯ И ИНТЕРПОЛЯЦИИ

Исследование динамики социально-экономических явлений, выявление и характеристика основной тенденции развития и моделей взаимосвязи дают основание для прогнозирования определения будущих размеров уровня экономического явления.

Особенно актуальными становятся вопросы прогнозирования в условиях перехода на международную методологию учета и анализа социально-экономических явлений.

Важное место в системе методов прогнозирования занимают статистические методы. Применение прогнозирования предполагает, что закономерность развития, действующая в прошлом (внутри ряда динамики), сохранится и в прогнозируемом будущем, т. е. прогноз основан на **экстраполяции.** Экстраполяция, проводимая в будущее, называется **перспективной** и в прошлое **ретроспективной.** Обычно, говоря об экстраполяции рядов динамики, подразумевают чаще всего перспективную экстраполяцию.

Теоретической основой распространения тенденции на будущее является известное свойство социально-экономических явлений, называемое **инерционностью.** Именно инерционность

где $\rho^2 = \frac{1}{2} \cdot \frac{\Sigma \Delta_i^2}{n}$; $\sigma^2_{ост} = \frac{\Sigma (y_i - \bar{y}_{\Delta})^2}{n}$.

позволяет выявить сложившиеся взаимосвязи как между уровнями динамического ряда, так и между группой связных рядов динамики. На основе рядов динамики получаются весьма надежные прогнозы, если уровни ряда динамики сопоставимы и получены на основе единой методологии.

Применение экстраполяции в прогнозировании базируется на следующих предпосылках:

- развитие исследуемого явления в целом следует описывать плавной кривой;
- общая тенденция развития явления в прошлом и настоящем не должна претерпевать серьезных изменений в будущем.

Поэтому надежность и точность прогноза зависят От того, насколько близкими к действительности окажутся эти предположения, а также как точно удалось охарактеризовать выявленную в прошлом закономерность. Экстраполяцию следует рассматривать как начальную стадию построения окончательных прогнозов. Механическое использование экстраполяции может стать причиной погрешности и неправильных выводов. Всегда следует учитывать все необходимые условия, предпосылки и гипотезы, связывая их с тщательным содержательным экономико-теоретическим анализом.

Разумеется, чем шире раздвигаются временные рамки прогнозирования, тем очевиднее становится недостаточность простого экстраполяционного метода (изменения тенденций, неизвестны точки поворота кривых, влияние новых факторов и т. д.). В этом случае динамичность экономических явлений и процессов вступает в противоречие с инерционностью их развития. Так как анализируемые экономические ряды динамики нередко относительно короткие, то временной горизонт экстраполяции не может быть бесконечным. Поэтому, чем короче **срок экстраполяции (период упреждения),** тем более надежные и точные результаты (при прочих равных условиях) дает прогноз. За короткий период не успевают сильно измениться условия развития явления и характер его динамики.

Экстраполяцию в общем виде можно представить формулой

$$\hat{y}_{i+T} = f(y_i, T, a_j), \quad (10.52)$$

где $\hat{y}_{i+T}$ - прогнозируемый уровень;

y_i - текущий уровень прогнозируемого ряда; Т - период упреждения:

a_j - параметр уравнения тренда.

В зависимости от того, какие принципы и исходные данные положены в основу прогноза, выделяют следующие элементарные методы экстраполяции: **среднего абсолютного прироста, среднего темпа роста** и экстраполяцию на основе **выравнивания рядов по какой-либо аналитической формуле.**

Прогнозирование по **среднему абсолютному приросту** может быть выполнено в том случае, если есть уверенность считать общую тенденцию линейной, т. е. метод основан на предположении о равномерном изменении уровня (под равномерностью понимается стабильность абсолютных приростов).

Для нахождения интересующего нас аналитического выражения тенденции на любую дату t необходимо определить средний абсолютный прирост и последовательно прибавить его к последнему уровню ряда столько раз, на сколько периодов экстраполируется ряд, т. е. экстраполяцию можно сделать по следующей формуле:

$$\hat{y}_{i+t} = y_i + \bar{\Delta} \cdot t, \quad (10.53)$$

где

$\hat{y}_{i+t}$ - экстраполируемый уровень: (i +t) - номер этого уровня (года);

i - номер последнего уровня (года) исследуемого периода, за который рассчитан А; t - срок прогноза (период упреждения);

$\bar{\Delta}$ - средний абсолютный прирост.

Однако следует иметь в виду, что использование среднего абсолютного прироста для прогноза возможно только при следующем условии:

$$\sigma^2_{ост} \leq \rho^2, \quad (10.54)$$

Например, по данным об удельном весе прибытия воздушных судов, выполненных без опоздания по сравнению с расписанием за 1987-1997 гг. (см. табл. 10.9) экстраполируем ряд на 1998— 1999 гг. Средний абсолютный прирост равен (-0,17%), остаточ-? ная дисперсия: $С^2_{ОСТ} = 0{,}003$; $р^2 = 0{,}02$. Следовательно, $СУ^2_{ОСТ} < р^2$, основное условие выполняется, можно делать прогноз:

$$\hat{y}_{98} = 89{,}9 + (-0{,}17) = 89{,}7\%;$$

$$\hat{y}_{99} = 89{,}7 + (-0{,}17) = 89{,}5\%.$$

Прогнозирование по **среднему темпу роста** осуществляется в случае, когда есть основание считать, что общая тенденция ряда характеризуется показательной (экспоненциальной) кривой. Для нахождения тенденции необходимо определить средний коэффициент роста, возведенный в степень, соответствующую периоду экстраполяции, т. е. по формуле

$$\hat{y}_{i+l} = y_i \cdot \bar{k}_p^{\,t}, \qquad (10.55)$$

где

y_i — последний уровень ряда динамики; t - срок прогноза; $\bar{k}_p$ - средний коэффициент роста.

Если же ряду динамики свойственна иная закономерность, то данные, полученные при экстраполяции на основе среднего темпа роста, будут отличаться от данных, рассчитанных другими способами экстраполяции.

Рассмотренные способы экстраполяции тренда, будучи простейшими, в то же время являются и самыми приближенными.

Поэтому наиболее распространенным методом прогнозирования считают **аналитическое выражение тренда.** При этом для выхода за границы исследуемого периода достаточно продолжить значения независимой переменной времени (t).

При таком подходе к прогнозированию предполагается, что размер уровня, характеризующего явление, формируется под воздействием множества факторов, причем не представляется возможным выделить отдельно их влияние. В связи с этим ход развития связывается не с какими-либо конкретными факторами, а с течением времени, т. е. $y = f(t)$.

Экстраполяция дает возможность получить точечное значение прогноза. Точное совпадение фактических данных и прогностических точечных оценок, полученных путем экстраполяции кривых, характеризующих тенденцию, имеет малую вероятность. *Возникновение таких отклонений объясняется следующими причинами.*

1. Выбранная для прогнозирования кривая не является единственно возможной для описания тенденции. Можно подобрать такую кривую, которая дает более точные результаты.

2. Построение прогноза осуществляется на основании ограниченного числа исходных данных. Кроме того, каждый исходный уровень обладает еще случайной компонентой. Поэтому и кривая, по которой осуществляется экстраполяция, будет содержать случайную компоненту.

3. Тенденция характеризует лишь движение среднего уровня ряда динамики, поэтому отдельные наблюдения от него отклоняются. Если такие отклонения наблюдались в прошлом, то они будут наблюдаться и в будущем.

Любой статистический прогноз носит приближенный характер. Поэтому целесообразно определение доверительных интервалов прогноза.

Величина доверительного интервала определяется следующим образом:

$$\hat{y}_t \pm t_\alpha \cdot \sigma_{\bar{y}_t},$$

где$\sigma_{\bar{y}_t}$ - средняя квадратическая ошибка тренда;

$\hat{y}_t$ - расчетное значение уровня;

t_α - доверительная величина. \

Вместо $1_а$-критерия Е. М. Четыркин предлагает брать коэффициент (k*).

Например, необходимо провести прогноз на 1998-2001 гг. по данным табл. 10.7 об урожайности зерновых культур в хозяйстве.

Для экстраполяции используем уравнение тренда, полученное по прямой: $y_t = 14{,}8 + 0{,}17t$. Подставив соответствующие значения t в наше уравнение, получим точечные прогнозы на 19982001 гг. (табл. 10.19 гр. 2). Для построения интервальных прогнозов рассчитаем среднюю квадратическую ошибку тренда (а$_t$ = 1,797) и значения k*. Результаты прогноза представлены в табл. 10.19.

При анализе рядов динамики иногда приходится прибегать к определению некоторых неизвестных уровней внутри данного ряда динамики, т. е. к **интерполяции.**

Как и экстраполяция, интерполяция может производиться на основе среднего абсолютного прироста, среднего темпа роста и с помощью аналитического выравнивания. Она также основана

Таблица 10.19

Прогнозные значения урожайности зерновых культур в хозяйстве на 1998 - 2001 гг.

Год	t	y_t	k*[1]	V^{k*}	$yi_{+},\pm oV^{k*}$
А	1	2	3	4	5
1998	8	16,2	2,0153	3,6	12,6-19,8
1999	9	16,3	2,0621	3,7	12,6-20,0
2000	10	16,5	2,1131	3,8	12,7-20,3
2001	11	16,7	2,1680	3,9	12,8-20,6

[1] Значения k* взяты из: *Четыркин Е. М.* Статистические методы прогнозирования. -М.: Статистика, 1975. -С. 183.

на том или ином предположении о тенденции изменения уровней, но характер этого прогноза несколько иной: здесь уже не приходится предполагать, что тенденция, характерная для прошлого, сохранится и в будущем.

При интерполяции считается, что ни выявленная тенденция, ни ее характер не претерпели существенных изменений в том промежутке времени, уровень (уровни) которого нам не известен. Такое предположение обычно является более обоснованным, чем предположение о будущей тенденции.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 10

Ряд динамики - ряд числовых значений определенного статистического показателя в последовательные моменты или периоды времени.

Интервальный ряд динамики — ряд числовых значений определенного статистического показателя, характеризующего размеры изучаемого явления за определенные промежутки (периоды, интервалы) времени.

Моментный ряд динамики - ряд числовых значений определенного статистического показателя, характеризующего размеры изучаемого явления на определенные даты, моменты времени.

Уровень ряда динамики - абсолютная (относительная, средняя) величина каждого члена динамического ряда.

Хронологическая средняя - средняя, исчисленная из уровней динамического ряда.

Средняя хронологическая интервального ряда исчисляется по формуле средней арифметической, причем при равных интервалах применяется средняя арифметическая простая, а при неравных - средняя арифметическая взвешенная.

Средняя хронологическая моментного ряда исчисляется как сумма всех уровней ряда, поделенного на число членов ряда без одного, причем первый и последний члены ряда берутся в половинном размере.

Смыкание рядов динамики - один из методов приведения несопоставимых рядов к сопоставимым путем прямого пересчета уровней с помощью специальных коэффициентов или относительных величин.

Абсолютный прирост измеряет абсолютную скорость роста (или снижения) уровня ряда за единицу времени (месяц, квартал, год и т. д.). Он показывает, на сколько единиц увеличился или уменьшился уровень по сравнению с базисным, т. е. за тот или иной промежуток времени.

Темп роста - относительный показатель, характеризующий интенсивность процесса роста (или снижения). Он показывает, сколько процентов составляет уровень данного периода по сравнению с базисным или предыдущим уровнем, т. е. характеризует относительную скорость изменения уровня ряда в единицу времени.

Темп прироста-относительный показатель, характеризующий величину прироста (снижения).

Абсолютный размер 1% прироста - абсолютный показатель, который показывает, какое содержание имеется в 1% прироста, сколько весом 1%.

Абсолютный прирост скорости (замедления) или ускорения абсолютный показатель, который показывает, на сколько данная скорость больше (меньше) предыдущей.

Коэффициент опережения (замедления) - относительный показатель, характеризующий сравнение динамических рядов, относящихся к двум пространственным объектам (странам, республикам и т. д.).

Средний абсолютный прирост - показатель, характеризующий среднюю абсолютную скорость роста (или снижения) уровня за отдельные периоды времени. Он показывает, на сколько единиц увеличился (или уменьшился) уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежемесячно и т. д.).

Средний темп роста - относительный показатель, выраженный в форме коэффициента и показывающий, во сколько раз увеличился уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежеквартально и т. п.).

Средний темп прироста - относительный показатель, выраженный в процентах и показывающий, на сколько увеличился (или уменьшился) уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежемесячно и т. п.).

Основная тенденция (тренд) - достаточно плавное и устойчивое изменение уровня явления во времени, более или менее свободное от случайных колебаний. Основную тенденцию можно представить либо аналитически - в виде уравнения (модели) тренда, либо графически.

Механическое сглаживание - метод нахождения плавных уровней'ряда динамики путем использования скользящих средних. Различают метод невзвешенных и взвешенных скользящих средних.

Аналитическое выравнивание динамического ряда проводится при помощи математической формулы, отражающей общую тенденцию ряда.

Сезонная компонента ряда динамики - внутригодичные колебания, имеющие более или менее регулярный характер. Их мерой обычно является индекс-сезонности.

Ряд Фурье дает возможность выделить периодические (сезонные) колебания, свойственные динамике многих экономических явлений.

Автокорреляция - корреляционная зависимость между последовательными (т. е. соседними) значениями уровней динамического ряда у, и $у_2$; У2 и y_3 и т. д.

Авторегрессия - регрессия, учитывающая влияние предыдущих уровней ряда на последующие.

Лаг - промежуток времени отставания, одного явления от другого, связанного с ним.

Интерполяция - приближенный расчет уровней, лежащих внутри ряда динамики, но почему-либо неизвестных.

Экстраполяция — нахождение уровней за пределами изучаемого ряда, т. е. продление ряда на основе выявленной закономерности изменения уровней в изучаемый отрезок времени.

ТЕСТ К ГЛАВЕ 10

1. **Ряд динамики, характеризует:**

а) структуру совокупности по какому-либо признаку;
б) изменение характеристики совокупности в пространстве;
в) изменение характеристики совокупности во времени.

2. **Уровень ряда динамики - это:**

а) определенное значение варьирующего признака в совокупности;
б) величина показателя на определенную дату или момент времени;
в) величина показателя за определенный период времени.

3. **Средний уровень интервального ряда динамики определяется как:**

*а) средняя арифметическая;
б) средняя гармоническая;
в) средняя хронологическая.

4. **Средний уровень моментного ряда исчисляется как средняя арифметическая взвешенная при:**

а) равноотстоящих уровнях между датами; «-б) неравноотстоящих уровнях между датами.

5. **Средний уровень моментного ряда исчисляется как средняя хронологическая при:**

а) равноотстоящих уровнях между датами; -б) неравноотстоящих уровнях между датами.

6. **Если сравниваются смежные уровни ряда динамики, показатели называются:**

-а) ценными;

б) базисными.

7. Если все уровни ряда динамики сравниваются с одним и тем же уровнем, показатели называются:

а) ценными;
• б) базисными.

8. Абсолютный прирост исчисляется как:

а) отношение уровней;
• б) разность уровней ряда.

9. Темп роста исчисляется как:

» а) отношение уровней ряда;
б) разность уровней ряда.

10. Основная тенденция представляет собой изменение ряда динамики:

а) равномерно повторяющееся через определенные промежутки времени внутри ряда;
-б) определяющее какое-то общее направление развития.

11. Сезонные колебания представляют собой изменения ряда динамики, равномерно повторяющиеся:

- а) через определенные промежутки времени с годичным интервалом;
б) внутри года.

12. Для выявления основной тенденции развития используются:

,*&)* метод укрупнения интервалов;
б) метод скользящей средней;
в) метод аналитического выравнивания;
г) ряд Фурье.

13. При сравнении динамики взаимосвязанных показателей применяются приемы:

а) приведения рядов динамики к одному основанию;
б) смыкания динамических рядов.

14. С целью приведения несопоставимых уровней ряда динамики к сопоставимому виду применяются приемы:

а) приведения рядов динамики к одному основанию; - б) смыкания динамических рядов.

15. Индексы сезонности можно рассчитать как отношение фактического уровня за тот или иной месяц к:

а) среднемесячному уровню за год;
б) выравненному уровню за тот же месяц;
-в) среднемесячному выравненному уровню за год.

16. Можно ли изучить взаимосвязи социально-экономических явлений по данным рядов динамики?

а) да;
б) нет.

17. Влияет ли автокорреляция на результаты измерения связи?

в) да;
г) нет.

ГЛАВА 11

СТАТИСТИЧЕСКИЙ АНАЛИЗ СТРУКТУРЫ

11.1

ПОНЯТИЕ И ВИДЫ СТРУКТУРЫ СОЦИАЛЬНОЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

• Структура (от лат. structure - строение, расположение, порядок) в наиболее общем смысле этого слова представляет собой совокупность устойчивых внутренних связей объекта, обеспечивающих его целостность и тождественность самому себе, т. е. сохранение основных свойств при различных внешних и внутренних изменениях. В статистике под структурой понимают совокупность элементов социально-экономических явлений, обладающих определенной устойчивостью внутригрупповых связей, при сохранении основных свойств, характеризующих эту совокупность как целое. Статистическая структура - это распределение различных частей в пределах общего для них качества, распределение составляющих совокупность единиц по количественному или качественному признаку. Подавляющая масса изучаемых статистикой сложных объектов, процессов или явлений в сфере промышленного или сельскохозяйственного производства, финансов, коммерции, демографии, в социальной и других областях может быть исследована с точки зрения их внутренней структуры по тому или иному признаку.

Статистический анализ структуры непосредственно связан с группировкой данных. Если основанием структуры выступает качественный признак, то процесс группировки, как правило, не вызывает затруднений. Группировка по количественному признаку обычно сложнее, так как требует обоснованного установления границ перехода одного качества в другое. Анализ структуры совокупности одновременно по нескольким количественным признакам обычно проводится на основе методов многомерной классификации.

Статистические приемы и методы анализа позволяют проводить исследование конкретных социально-экономических структур в определенных условиях места и времени, заключающееся прежде всего в их точном количественном измерении и соизмерении, выявлении пропорций и закономерностей. Структура сложного социально-экономического явления всегда обладает той или иной степенью подвижности, имеет свойство меняться с течением времени как в количественном, так и в качественном отношении. Поэтому большое практическое значение имеют изучение структуры в динамике, оценка структурных сдвигов, выявление и характеристика основных тенденций ее развития. Углубленный анализ структуры, требующий применения корреляционно-регрессионного анализа и индексного метода, предполагает также изучение факторов, воздействующих на структуру, и оценку влияния структуры на взаимосвязанные с ней результативные показатели.

Классификация структур предполагает в первую очередь их разделение на два основных вида по временному фактору. **Моментные** структуры характеризуют строение социально-экономических явлений по состоянию на определенные моменты времени и отображаются посредством моментных относительных показателей (как правило, на начало или на конец периода). Моментными являются структуры: населения по полу, возрасту, уровню образования; основных фондов по отраслям народного хозяйства и по формам собственности; парка транспортных средств и т. п. **Интервальные** структуры характеризуют строение социально-экономических явлений за определенные периоды времени (дни, недели, месяцы, кварталы, годы). Они отображаются интервальными относительными показателями. Например, структуры экспорта и импорта, товарооборота, материальных затрат.

Статистика имеет дело как с **фактическими,** реально существующими структурами, так и со структурами **перспективными, прогнозными, оптимальными и стандартизованными.** Последние представляют собой какие-либо гипотетические (условные) или фактические структуры, принятые в качестве эталонных для расчета и сравнения стандартизованных показателей. Например, для сравнения уровней рождаемости, смертности, заболеваемости и т. п. по двум и более регионам рассчитывают стандартизованные коэффициенты на основе некоторой стандартизованной структуры, в качестве которой может использоваться возрастная структура населения в целом по стране.

11.2

ПОКАЗАТЕЛИ СТРУКТУРЫ И СТРУКТУРНЫХ СДВИГОВ

Различают две формы выражения относительных показателей структуры (гл. 6): долю и удельный вес, который представляет собой долю, выраженную в процентах. В дальнейшем изложении удельный вес (долю) i-й части совокупности (i = 1, k) в j-й период времени или по состоянию на j-й момент времени обозначим как djj, при этом все расчеты будем производить над величинами, выраженными в процентах.

Рассмотрим показатели, характеризующие изменения структуры, или структурные сдвиги. Отметим прежде всего, что термин «структурные сдвиги» применим лишь при исследовании структурных различий во времени. При территориальных же сравнениях структуры, а также при сравнении фактической структуры со стандартизованной более корректным является термин «структурные различия». Рассматриваемые в данной главе методы динамических сравнений большей частью применимы и для сравнений территориальных.

•Для статистической оценки структурных сдвигов за два или более периодов используются **две группы показателей:** показатели, основывающиеся на разностях между удельными весами одноименных частей совокупности; показатели, базирующиеся на отношениях удельных весов

одноименных частей
совокупности.

Исходным в первой группе является показатель «абсолютного» **прироста удельного веса** i-й части совокупности (dj), показывающий, на какую величину в долях единицы или процентах возросла или уменьшилась данная структурная часть в j-й период по сравнению с (j - 1) периодом*. Этот показатель рассчитывается по следующей формуле:

$$\Delta d_i = d_{ij} - d_{ij-1}. \tag{11.1}$$

Знак прироста показывает направление изменения удельного веса данной структурной части («+» - увеличение, «-» - уменьшение), а его величина - конкретное значение этого изменения в процентных пунктах.

Так как сумма удельных весов всех частей совокупности в любой момент времени всегда равна строго 100%, то при какихлибо изменениях в структуре одна часть приростов удельных весов всегда будет иметь положительный знак, а другая - отрицательный. Сумма же всех k приростов для совокупности в целом всегда равна нулю:

$$\sum_{i=1}^{k} \Delta d_i = \sum_{i=1}^{k} (d_{ij} - d_{ij-1}) = \sum_{i=1}^{k} d_{ij} - \sum_{i=1}^{k} d_{ij-1} = 100\% - 100\%.$$

В качестве примера рассмотрим динамику структуры внешнеторгового оборота РФ в 1997 г. (табл. 11.1).

Таблица 11.1 **Структура внешнеторгового оборота РФ в 1997 г. (%)**

Составляющие внешнеторгового оборота	Квартал				IV квартал по сравнению с III кварталом	
	I	II	III	IV	прирост удельного веса, проц. пунктов (гр.4- -гр.3)	рост удельного веса (гр. 4 : : гр. 3)
А	1	2	3	4	5	6
Экспорт	59,5	55,6	54,3	56,4	2,1	1,039
Импорт	40,5	44,4	45,7	43,6	-2,1	0,954
Всего	100	100	100	100	0	X

Для характеристики структурных сдвигов в IV квартале по сравнению с III кварталом по формуле (11.1) рассчитаны «абсолютные» приросты удельного веса (гр.5): экспорт возрос на 2,1 процентного пункта и соответственно на эту же величину снизился удельный вес импорта.

Для второй из выделенных выше групп показателей структурных сдвигов исходным является показатель **темп роста удельного веса ($Тр_{di}$),** представляющий собой отношение удельного веса i-й части в j-й период времени к удельному весу этой же части в предшествующий период:

---

* Здесь и далее при исследовании моментных структур под периодами будут подразумеваться моменты времени; термин «абсолютный» при исследовании приростов удельных весов принято брать в кавычки, так как абсолютными эти показатели являются по методологии расчета, но не по единицам измерения.

(11.2)

$$Tp_{di} = \frac{d_{ij}}{d_{ij-1}}.$$

Темпы роста удельного веса всегда являются положительными величинами. Однако если в совокупности имели место какиелибо структурные изменения, часть темпов роста будет больше единицы, а часть - меньше. В то же время их среднее значение, взвешенное по базисным удельным весам, всегда строго равно единице:

$$\overline{Tp}_d = \frac{\sum_{i=1}^{k} Tp_{di}\cdot d_{ij-1}}{\sum_{i=1}^{k} d_{ij-1}} = \frac{\sum_{i=1}^{k} d_{ij}}{\sum_{i=1}^{k} d_{ij-1}} = \frac{100\%}{100\%}.$$

Рассчитанные темпы роста удельных весов структурных частей внешнеторгового оборота приведены в графе 6 табл. 11.1. Проверим правильность их вычисления расчетом средневзвешенного значения:

$$\overline{Tp}_d = \frac{1{,}039\cdot 0{,}543+0{,}954\cdot 0{,}457}{0{,}543+0{,}457} \approx 1.$$

Мы рассмотрели показатели структурных сдвигов за один интервал между Двумя периодами. Если же изучаемая структура представлена данными за три и более периодов, появляется необходимость в динамическом осреднении приведенных выше показателей.

Средний «абсолютный» прирост удельного веса (Adj) i-й структурной части за п периодов определяется по формуле
дуальных приростов удельного веса экспорта суммой средних приростов, общий прирост останется без изменений:

$$\overline{\Delta}d_i = \frac{d_{in}-d_{i1}}{n-1},$$

$$-3{,}9 - 1{,}3 + 2{,}1 = -1{,}03 - 1{,}03 - 1{,}03 = -3{,}1.$$

(11.3)

где i - обозначает структурную часть$(i = 1, k)$; j - определяет временной интервал$(j = 1, n)$.

При отсутствии в расчетах ошибок и соблюдении достаточной точности вычислений сумма средних «абсолютных» приростов удельных весов всех *к* структурных частей совокупности, так же как и сумма их приростов за один временной интервал, должна быть равна нулю:

$$\sum_{i=1}^{k} \overline{\Delta}d_i = \sum_{i=1}^{k} \frac{d_{in}-d_{i1}}{n-1} = \frac{1}{n-1}\left(\sum_{i=1}^{k} d_{in} - \sum_{i=1}^{k} d_{i1}\right) = \frac{1}{n-1}(100\%-100\%).$$

Анализируя структуру внешнеторгового оборота за все четыре квартала 1997 г., мы можем определить не только квартальные и общие за весь период приросты удельного веса каждой структурной части, но и на основе формулы (11.3) - средний квартальный прирост, представляющий собой общий прирост, деленный на число интервалов между рассматриваемыми периодами (между четырьмя кварталами три интервала). Например, для экспорта получим:

$$\overline{\Delta}d_j = \frac{56{,}4-59{,}5}{3} = -1{,}03 \text{ проц. пункта.}$$

Относительным показателем, характеризующим изменение удельного веса i-й структурной части за п периодов, является **средний темп роста удельного веса.** При расчете этого показателя используется формула средней геометрической:

$$\overline{Tp}_{d_i} = \sqrt[n-1]{Tp_{d_{i1}}\cdot Tp_{d_{i2}}\cdot Tp_{d_{i3}}\cdot\ldots\cdot Tp_{d_{in-1}}}.$$

(11.4)

Подкоренное выражение этой формулы представляет собой последовательное произведение индивидуальных темпов роста удельного веса за все временные интервалы. После проведения несложных алгебраических преобразований формула (11.4) примет следующий вид:

$$\bar{T}p_{d_i} = \sqrt[n-1]{\frac{d_{in}}{d_{i1}}}. \qquad (11.5)$$

Используя эту формулу, по итогам года определим средний квартальный темп роста удельного веса экспорта

$$\bar{T}p_d = \sqrt[3]{\frac{56,4}{59,5}} = \sqrt[3]{0,948} = 0,982$$

и импорта

$$\bar{T}p_d = \sqrt[3]{\frac{43,6}{40,5}} = \sqrt[3]{1,0}$$

Средний «абсолютный» прирост удельного веса импорта составит:

$$\bar{\Delta}d_2 = \frac{43,6-40,5}{3} = 1,03 \text{ проц. пункта.}$$

Таким образом, удельный вес импорта ежеквартально увеличивался в среднем на 1 процентный пункт, а экспорт на эту же величину снижался.

Определяющим показателем среднего «абсолютного» прироста удельного веса является общий прирост удельного веса каждой i-й части. Например, если мы заменим сумму всех индиви-

Итак, импорт ежеквартально увеличивался в среднем в 1,03 раза, а экспорт снижался на 0,02 своей величины (1-0,98).

Определяющим показателем при расчете среднего темпа роста удельного веса i-й структурной части является общий темп роста-ее удельного веса за весь рассматриваемый период. Это означает, что, заменив произведение индивидуальных темпов роста произведением средних значений, общий прирост удельного веса данной части должен остаться без изменений. Про-

иллюстрируем это, используя темпы роста удельного веса импорта:

$$Tp_{d_{21}} = \frac{44,4}{40,5} = 1,096;$$

$$Tp_{d_{22}} = \frac{45,7}{44,4} = 1,029;$$

$$Tp_{d_{23}} = \frac{43,6}{45,7} = 0,954;$$

$$1,096 \cdot 1,029 \cdot 0,954 = 1,025 \cdot 1,025 \cdot 1,025.$$

При анализе структуры исследуемого объекта или явления за ряд периодов так же можно определить средний, или общий (в данном случае одно и то же), удельный вес каждой i-й части за весь рассматриваемый временной интервал. Однако для этого одних лишь относительных данных об удельных весах структурных частей недостаточно, необходимо располагать исходными данными о размерах этих частей в абсолютном выражении (табл. 11.2).

Таблица 112 **Исходные данные для определения удельных весов**

T

Используя эти данные, **средний удельный вес** любой i-й структурной части можно определить по формуле

$$\bar{d}_j = \frac{\sum_{j=1}^{n} x_{ij}}{\sum_{j=1}^{n}\sum_{i=1}^{k} x_{ij}} \cdot 100\%. \qquad (11.6)$$

Рассмотрим табл. 11.3.

Таблица 11.3 **Производство продукции промышленным предприятием**

Вид продукции	Квартал							
	I		II		III		IV	
	млн руб.	%	млн руб.	%	млн руб.	%	млн руб.	%
А	80	20	800	80	100	20	1200	80
В	320	80	200	20	400	80	300	20
Итого	400	100	1000	100	500	100	1500	100

\ Пе, \ риод \ СтрукХ турная \ часть \	1	2	3		n
1 2 3 k	xП X2l x3, xki	x|2 x22 x32 xk2	x» x23 x33 xk3		x!n . x2„ x3n xkn
Итого	k 5>П i=l	k I×i2 i=l	k Z×i3 i=!		k I× in i=l

Располагая лишь относительными показателями структуры, можно предположить, что в среднем за год средние удельные веса объемов продукции двух видов были равны:

$$\bar{d}_A = \frac{20+80+20+80}{4} = 50\%;$$

$$\bar{d}_B = \frac{80+20+80+20}{4} = 50\%.$$

Однако при этом мы не учли, что 1% в каждом квартале соответствует различная в абсолютном выражении величина. Рассчитаем истинные средние удельные веса по обоим видам продукции, используя формулу (11.6):

$$\overline{d}_A = \frac{80+800+100+1200}{400+1000+500+1500}\cdot 100\% = 64{,}1\%;$$

$$\overline{d}_B = \frac{320+200+400+300}{400+1000+500+1500}\cdot 100\% = 35{,}9\%.$$

Полученные значения существенно отличаются от предполагаемых.

Если в абсолютном выражении известен только общий объем признака за каждый период, а также удельные веса структурных частей, для расчета среднего удельного веса используют среднюю арифметическую взвешенную.

Среди предлагаемых для этих целей обобщающих показателей наиболее легко интерпретируется **линейный коэффициент «абсолютных» структурных сдвигов** (A$_{drdo}$), представляющий собой сумму приростов удельных весов, взятых без учета знака, деленную на число структурных частей:

$$\overline{\Delta}_{d_1-d_0} = \frac{\sum\limits_{i=1}^{k}\left|d_{ij}-d_{ij-1}\right|}{k}. \qquad (11.8)$$

$$\overline{d}_i = \frac{\sum\limits_{j=1}^{n}\left(d_{ij}\cdot\sum\limits_{i=1}^{k}x_{ij}\right)}{\sum\limits_{j=1}^{n}\cdot\sum\limits_{i=1}^{k}x_{ij}}\cdot 100\%. \qquad (11.7)$$

На основе этой формулы для нашего примера мы получим те же самые значения средних удельных весов:

Этот показатель отражает то среднее изменение удельного веса (в процентных пунктах), которое имело место за рассматриваемый временной интервал.

Для решения данной задачи применяется также **квадратический коэффициент «абсолютных» структурных сдвигов** (a$_d$ _d), который резче реагирует на происходящие в совокупности структурные изменения:

$$\overline{d}_A = \frac{20\cdot 400+80\cdot 100+20\cdot 500+80\cdot 1500}{400+1000+500+1500} = 64{,}1\%;$$

$$\overline{d}_B = \frac{80\cdot 400+20\cdot 1000+80\cdot 500+20\cdot 1500}{400+1000+500+1500} = 35{,}9\%.$$

$$\sigma_{d_1-d_0} = \sqrt{\frac{\sum\limits_{i=1}^{k}\left(d_{ij}\right.}{}}$$

11.3

СВОДНАЯ ОЦЕНКА СТРУКТУРНЫХ ИЗМЕНЕНИЙ ВО ВРЕМЕНИ И ПРОСТРАНСТВЕ

В предшествующем параграфе были рассмотрены показатели, позволяющие измерить те количественные изменения, которым подвергалась каждая отдельно взятая часть изучаемой совокупности. В то же время перед исследователем в ряде случаев встает задача в целом оценить структурные изменения в изучаемом социально-экономическом явлении, имеющие место за определенный временной интервал и характеризующие подвижность или, наоборот, стабильность, устойчивость данной структуры. Как правило, это необходимо для сравнения динамики одной и той же структуры в различные периоды или нескольких структур, относящихся к разным объектам. Во втором случае число структурных частей у разных объектов необязательно должно совпадать.

(11.9)

Линейный и квадратический коэффициенты «абсолютных» структурных сдвигов позволяют получить сводную оценку скорости изменения удельных весов отдельных частей совокупности. Для сводной характеристики интенсивности изменения удельных весов используется **квадратический коэффициент относительных структурных сдвигов**

$$\left(\sigma_{\frac{d_1}{d_0}}\right)$$

$$\sigma_{\frac{d_1}{d_0}} = \sqrt{\sum_{i=1}^{k} \frac{(d_{ij} - d_{ij-1})^2}{d_{ij-1}} \cdot 100.} \qquad (11.10)$$

Данный показатель отражает тот средний относительный прирост удельного веса (в процентах), который наблюдался за рассматриваемый период.

Рассмотрим применение формул (11.8) - (11.10) на примере структуры потребительских расходов населения региона (табл. 11.4).

Таблица 11.4

Структура потребительских расходов населения условного региона

Статья расходов	Удельный вес, % к итогу			Расчетные графы						
	1996	1997	1998	$\lvert d_{i2} - d_{i1}\rvert$	$(d_{i2} - d_{i1})^2$	$\frac{(d_{i2} - d_{i1})^2}{d_{i1}}$	$\lvert d_{i3} - d_{i2}\rvert$	$(d_{i3} - d_{i2})^2$	$\frac{(d_{i3} - d_{i2})^2}{d_{i2}}$	$\lvert d_{i3} - d_{i1}\rvert$
А	1	2	3	4	5	6	7	8	9	10
Продовольственные товары	33,7	43,9	45,2	10,2	104,04	3,09	1,3	1,69	0,04	11,5
Непродовольственные товары	54,2	45,3	42,0	8,9	79,21	1,46	3,3	10,89	0,24	12,2
Алкоголь	3,4	4,4	4,0	1,0	1,00	0,29	0,4	0,16	0,04	0,6
Услуги	8,7	6,4	8,8	2,3	5,29	0,61	2,4	5,76	0,90	0,1
Все расходы	100,0	100,0	100,0	22,4	189,54	5,45	7,4	18,50	1,22	24,4

Для расчета линейного коэффициента «абсолютных» структурных сдвигов за первый (с 1996 по 1997 г.) и за второй (с 1997 по 1998 г.) периоды воспользуемся итогами гр. 4 и 7 табл. 11.4.

$$\overline{\Delta}^{1}_{d_1 - d_0} = \frac{22,4}{4} = 5,6 \text{ проц. пункта;}$$

$$\overline{\Delta}^{II}_{d_1-d_0} = \frac{7,4}{4} = 1,9 \text{ проц. пункта.}$$

Итак, с 1996 по 1997 г. удельный вес отдельных статей потребительских расходов населения изменился в среднем на 5,6 процентного пункта. В течение же следующего года «абсолютные» структурные сдвиги уменьшились и структура потребительских расходов начала стабилизироваться. Эти выводы подтверждаются квадратическими коэффициентами «абсолютных» структурных сдвигов (необходимые промежуточные расчеты выполнены в гр. 5 и 8 табл. 11.4).

$$\sigma^{I}_{d_1-d_0} = \sqrt{\frac{189,54}{4}}6,9 \text{ проц. пункта;}$$

$$\sigma^{II}_{d_1-d_0} = \sqrt{\frac{18,50}{4}} = 2,2 \text{ проц. пункта.}$$

Теперь определим величину квадратических коэффициентов относительных структурных сдвигов, воспользовавшись итогами гр. 6 и 9 табл. 11.4:

$$\sigma^{I}_{\frac{d_1}{d_0}} = \sqrt{5,45 \cdot 100} = 23,3\%;$$

$$\sigma^{II}_{\frac{d_1}{d_0}} = \sqrt{1,22 \cdot 100} = 11,0\%.$$

Расчеты показывают, что если за первый год удельный вес каждой статьи расходов в среднем изменился почти на четверть своей величины, то за следующий год - только на 1/9.

Для сводной оценки структурных изменений в исследуемой совокупности в целом за рассматриваемый временной интервал, охватывающий несколько дней, месяцев, кварталов или лет, наиболее удобным является **линейный коэффициент «абсолютных» структурных сдвигов за п периодов** $\left(\overline{\Delta}^{(n)}_{d_1-d_0}\right)$:

$$\overline{\Delta}^{(n)}_{d_1-d_0} = \frac{\sum_{i=1}^{k}|d_{in} - d_{i1}|}{k(n-1)}. \quad \text{(П.Н)}$$

Используя итог гр. 10 табл. 11.4, получим:

$$\overline{\Delta}^{(n)}_{d_1-d_0} = \frac{24,4}{4 \cdot 2} = 3,1 \text{ проц. пункта.}$$

Таким образом, за рассматриваемый период среднее годовое изменение по всем статьям расходов составило 3,1 процентного пункта.

Необходимо отметить, что показатель (11.11) может использоваться как для сравнения динамики двух и более структур, так и для анализа динамики одной и той же структуры за разные по продолжительности периоды времени.

11.4

СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ КОНЦЕНТРАЦИИ И ЦЕНТРАЛИЗАЦИИ

Одна из задач статистического анализа структуры заключается в определении степени концентрации изучаемого признака по единицам совокупности или в оценке неравномерности его распределения. Такая неравномерность может иметь место в распределении доходов по группам населения, жилой площади по группам семей, прибыли по группам предприятий и т. д. При исследовании неравномерности распределения изучаемого признака по территории понятие «концентрация» обычно заменяется понятием «локализация».

Оценка степени концентрации наиболее часто осуществляется по **кривой концентрации (Лоренца)** и рассчитываемым на ее основе характеристикам. Для построения кривой концентрации необходимо иметь частотное распределение единиц исследуемой совокупности и взаимосвязанное с ним частотное распределение изучаемого признака. При этом для удобства вычислений и повышения аналитичности данных единицы совокупности обычно разбиваются на равные группы: 10 групп - по 10% единиц в каждой, 5 групп - по 20% единиц и т. д.

1

Проведем сравнительный анализ концентрации доходов населения условного региона в 1997 и 1998 гг. Для этого используем приведенную в табл. 11.5 группировку, в которой все население разбито на 10 равных групп таким образом, что 1-я группа объединяет 10% населения с наименьшими доходами, 2-я группа - следующие по уровню доходов 10% населения и так далее до последней, 10-й группы, объединяющей 10% населения с наибольшими доходами.

Таблица 11.5 **Распределение доходов населения региона, %**

Группы населения, ранжированные по уровню среднедушевого дохода (по 1 0% от общей ности населения)	1997		1998	
	Удельный вес в совокупном доходе	Накопленн ая частота	Удельный вес в совокупном доходе	Накопленн ая частота
1	4,3	4,3	3,2	3,2
2	6,1	10,4	4,8	8,0
3	7,1	17,5	6,1	14,1
4	8,1	25,6	7,2	21,3
5	9,1	34,7	8,4	29,7
6	10,1	44,8	9,7	39.4
7	11,2	56,0	11,3	50,7
8	12,6	68,6	13,2	63,9
9	14,3	82,9	15,8	79,7
10	17,1	100,0	20,3	100,0
Итого	100,0	444,8	100,0	410,0

Предварительный анализ этих данных позволяет заключить, что концентрация доходов, или дифференциация населения по уровню доходов, за рассматриваемый период усилилась. Так, если в 1997 г. 10% беднейших слоев населения располагали 4,3% совокупного дохода, а 10% наиболее обеспеченных людей - 17,1 % дохода, то в 1998 г. эти показатели составили соответственно 3,2 и 20,3%. Если в 1997 г. доходы менее обеспеченной половины населения составляли 34,7% общей величины, то в 1998 г. эта доля снизилась до 29,7%.

Кривая Лоренца строится в прямоугольной системе координат. На оси абсцисс откладываются накопленные частоты объема совокупности, а на оси ординат - накопленные частоты объема признака. Полученная при соединении точек кривая линия и будет характеризовать степень концентрации.

Если распределение является строго равномерным, то первые 10% единиц обладают 10% объема признака, первые 20% соответственно 20% объема признака и т. д. Такое распределение отображается прямой, проходящей из нижнего левого угла графика к верхнему правому углу, и называется **линией равномерного распределения.** Чем сильнее концентрация изучаемого признака, тем заметнее кривая Лоренца отклоняется от линии равномерного распределения, и, наоборот, чем слабее концентрация, тем ближе будет кривая к прямой.

Построенные по данным табл. 11.5 кривые концентрации доходов населения представлены на рис. 11.1.

Следует отметить, что кривая концентрации может сколько угодно близко приближаться к линии равномерного распределения, но никогда не пересекает ее. В нашем примере это объясняется следующим образом. Так как население ранжировано по уровню индивидуальных доходов от минимального значения к максимальному, то суммарный доход первых 10% населения не может превышать 10% общего совокупного дохода (иначе какая-либо другая 10%-ная группа с более высокими доходами должна была бы иметь менее 10% общего совокупного дохода, что невозможно). Суммарный доход первых 20% населения по той же причине не может превышать 20% общего совокупного дохода и т. д.

Степень концентрации определяется площадью фигуры А (см. рис. 11.1), ограниченной линией равномерного распределения и кривой концентрации. Чем больше площадь А и чем меньше площадь В, тем степень концентрации выше.

На сравнении площади А с площадью треугольника, расположенного ниже линии равномерного распределения, основан **коэффициент Джини (G):**

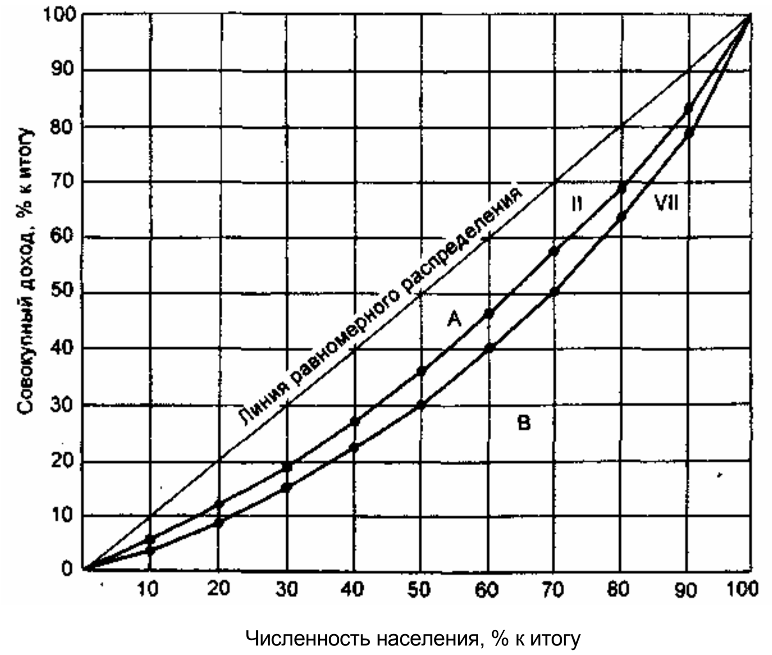


Рис. 11.1. Кривая концентрации доходов населения региона в 1997 и 1998 гг.

(11.12) $$G = 1 - 2\sum_{i=1}^{k} d_{xi} d_{yi}^{H} + \sum_{i=1}^{k} d_{xi} d_{yi},$$

где d_{xi} - доля i-й группы в общем объеме совокупности (в нашем примере в численности населения);

d_{yi} - доля i-й группы в общем объеме признака (в нашем примере в доходах);

d^{H}_{yi} - накопленная доля i-й группы в общем объеме признака.

Если исследуемая совокупность разделена на 10 равновеликих групп и частоты выражены в процентах, то данный коэффициент принимает следующий вид:

(11.13) $$G = 110 - 0{,}2\sum_{i=1}^{k} d_{yi}^{H}.$$

Используя итоги гр. 2 и 4 табл. 11.5 для наших распределений, получим:

$$G_{97} = 110 - 0{,}2 \cdot 444{,}8 = 21{,}0\%;$$
$$G_{98} = 110 - 0{,}2 \cdot 410{,}0 = 28{,}0\%.$$

Итак, если в 1997 г. концентрация доходов населения составляла 21%, то за год она возросла на 7 процентных пунктов.

В последние годы Госкомстат РФ при характеристике степени дифференциации доходов использует 20%-ные группы населения. Например, расслоение населения страны по доходам, сложившееся к началу 1998 г., отражают данные табл. 11.6.

Таблица 11.6

Распределение общего объема денежных доходов населения РФ в январе 1998 г.

20%-ные группы населения	Удельный вес в совокупном доходе, %	Накопленная частота, %
Первая (с наименьшими доходами)	6,2	6,2
Вторая	10,6	16,8
Третья	15,1	31,9
Четвертая	21,4	53,3
Пятая (с наивысшими доходами)	46,7	100,0
Итого	100,0	X

Для оценки уровня дифференциации в этом случае удобнее использовать следующее преобразование коэффициента Джини:

$$G = 120 - 0{,}4\sum_{i=1}^{k} d_{yi}^{H}. \qquad (11.14)$$

Используя данные последней графы табл. 11.6, получим:

$$G = 120 - 0{,}4(6{,}2+16{,}8+31{,}9+53{,}3+100{,}0) = 36{,}7\%.$$

Для построения кривой концентрации и расчета показателей концентрации необязательно иметь группы с равной численностью единиц. Рассмотрим решение этой задачи на основе структурной группировки с различными объемами групп (табл. 11.7).

Построенная по этим данным кривая концентрации будет иметь следующий вид (рис. 11.2).

Из графика видно, что в данном случае кривая Лоренца вогнута значительно сильнее, чем в предыдущем примере, и, следовательно, концентрация изучаемого признака выше. Это подтверждает и коэффициент Джини, рассчитанный по формуле (11.12):

$$G = 1 - 2 \cdot 0{,}2537 + 0{,}1685 = 0{,}6611\ (66{,}11\%).$$

Таблица 11.7

Численность занятых на приватизированных промышленных предприятиях

(на конец января 1994 г.)

Группы предприятий по численности занятых человек	Число предприятий			Численность занятых			dxidyi	d,A^{H}
	в абсолютных единицах	доля в общем итоге d_{xi}	накопленная доля $d_{,,}^{H}$	человек	доля в общем итоге $d_{y,}$	накопленная доля V		
1 -499	4941	0,632	0,632	0,99	0,125	0,125	0,0790	0,0790
500 - 999	1173	0,15	0,782	0,84	0,10	0,231	0,0159	0,0346

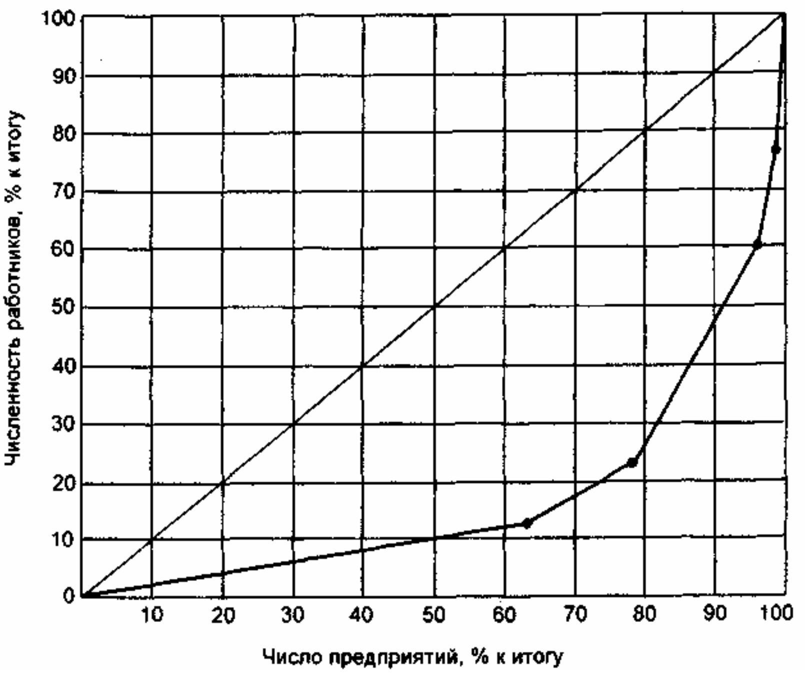
1000-4999	1408	0,180	0,962	2,92	0,36	0,600	0,0664	0,1080
5000 - 9999	202	0,026	0,988	1,36	0,17	0,772	0,0045	0,0201
10 000 и бол ее	94	0,012	1,000	1,81	0,228	1,000	0,0027	0,0120
Всего	7818	1,000	X	7,92	1,000	X	0,1685	0,2537

Получить количественную оценку степени концентрации можно и на основе сравнения по группам долей или удельных весов объема совокупности($c1_{x1}$) с долями или удельными весами объема признака (d_{yj}). По данным табл. 11.7 рассчитаем сумму абсолютных разностей удельных весов:

$$|63,2 - 12,5| + |15,0 - 10,6| + |18,0 - 36,9| + |2,6 - 17,2| + \\ + |1,2 - 22,8| = 110,2\%.$$

При равномерном распределении удельные веса должны совпадать. Так, при равномерном или близком к равномерному распределении работников по предприятиям на первых 63,2% предприятий было бы занято примерно 63,2% работников, на следующих 15,0% предприятий - примерно 15,0% и т. д.

При максимально возможном неравномерном распределении на предприятиях, входящих в первые



четыре группы, было бы занято 0 рабочих, а на предприятиях последней, 5-й группы -

Рассчитанное значение данного коэффициента, как и полученное ранее значение коэффициента Джини, свидетельствует о высокой степени концентрации работников на приватизированных предприятиях.

Если под **концентрацией** понимается степень неравномерности распределения изучаемого признака, не связанная ни с общим объемом совокупности, ни с численностью отдельных групп, то **централизация** означает сосредоточение объема признака у отдельных единиц (объема продукции данного вида на отдельных предприятиях, капитала в отдельных банках и т. п.). Обобщающий показатель централизации (I_z) имеет следующий вид:

$$I_Z = \sum_{i=1}^{n} \left(\frac{x_i}{\sum_{i=1}^{n} x_i} \right)^2, \quad (11.16)$$

где x_i- значение признака i-й единицы совокупности;

$\sum_{i=1}^{n} x_i$ - объем признака всей совокупности, где п - объем совокупности (число входящих в нее единиц).

Рис. 11.2. Кривая концентрации работников на приватизированных предприятиях все 100%. Тогда суммарная величина абсолютных разностей была бы следующей:

$$|63{,}2 - 0| + |15{,}0 - 0| + |18{,}0 - 0| + |2{,}6 - 0| + |1{,}2 - 100{,}0| = 197{,}6\%.$$

Теоретически возможное максимальное значение этой суммы при неравномерном распределении равно 200%. На соотношении фактической суммы с максимально возможной основан **коэффициент Лоренца** (L), также характеризующий степень концентрации:

$$[11.15]\quad L = \frac{\sum_{i=1}^{k} |d_{xi} - d_{yi}|}{200} \cdot 100\% = \frac{\sum_{i=1}^{k} |d_{xi} - d_{yi}|}{2}.$$

Максимального значения, равного 1, данный коэффициент достигает лишь в'том случае, когда совокупность состоит только из одной единицы, обладающей всем объемом признака. Минимальное значение коэффициента приближается к нулю, но никогда его не достигает.

Рассмотрим следующий условный пример, приведенный в табл. 11.8.

Используя формулы (11.15) и (11.16), определим степень концентрации и централизации производства данного вида продукции в районе А:

$$L_A = \frac{|25 - 25| + |25 - 25| + |50 - 50|}{} = 0;$$

$$L = \frac{110{,}2}{2} = 55{,}1\%.$$

$$I_{Z_A} = (0{,}25)^2 + (0{,}25)^2 + (0{,}25)^2 \cdot 2 = 0{,}25.$$

Для нашего примера получим:

В этом районе мы имеем равномерное распределение производимой продукции в полном соответствии с числом предприятий различных типов, поэтому степень концентрации нулевая. В то же время (так как все производство сосредоточено всего на

Таблица 11.8

Группировка предприятий отрасли по форме собственности

Тип предприятия	Район А						Район Б					
	Число предприятий		Объем продукции				Число предприятий		Объем продукции			
	единиц	%	млн руб.	%	млн руб. на 1 предприятие (гр. 3 : гр. 1)	доля в общем объеме (гр. 5 : Σгр. 3)	единиц	%	млн руб.	%	млн руб. на 1 предприятие (гр. 9 : гр. 7)	доля в общем объеме (гр. 11 : Σгр. 9)
А	1	2	3	4	5	6	7	8	9	10	11	12
ИЧП	1	25	200	25	200	0,25	5	25	1000	25	200	0,05
ООО	1	25	200	25	200	0,25	5	25	2000	50	400	0,10
АО	2	50	400	50	200	0,25	10	50	1000	25	100	0,025
Итого	4	100	800	100	X	X	20	100	4000	100	X	X

четырех предприятиях) имеет место сравнительно высокая централизация. Выполним аналогичные расчеты по району Б:

$$L_Б = \frac{|25-25|+|25-50|+|50-25|}{2} = 25\%;$$

$$I_{Z_Б} = (0{,}05)^2 \cdot 5 + (0{,}10)^2 \cdot 5 + (0{,}025)^2 \cdot 10 = 0{,}07.$$

В данном районе наблюдается концентрация производства в обществах с ограниченной ответственностью (ООО), вследствие этого соответствующий коэффициент увеличился до 25%. Но общее число предприятий в районе Б в 5 раз больше, что обусловливает небольшую степень централизации.

Отметим, что оценка степени централизации представляет практический интерес только при очень небольших объемах совокупности.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 11

Изучаемые статистикой сложные социально-экономические объекты и явления в большинстве своем характеризуются внутренней структурой, имеющей ту или иную степень подвижности и воздействующей на состояние и развитие объекта или явления в целом.

Структура - совокупность элементов социально-экономических явлений, обладающих определенной устойчивостью внутригрупповых связей, при сохранении основных свойств, характеризующих эту совокупность как цзлое.

Моментная структура - структура, характеризующая строение социально-экономических явлений по состоянию на определенные моменты времени (на определенную дату, начало или конец периода).

Интервальная структура - структура, характеризующая строение социально-экономических явлений за определенные периоды времени (дни, недели, месяцы, кварталы, годы).

Для статистического анализа динамики структуры используются две группы показателей: показатели, базирующиеся на приростах удельных весов одноименных частей совокупности, и показатели, основу которых составляют темпы роста удельных весов. Каждая из названных групп объединяет индивидуальные показатели, рассчитываемые для отдельных структурных частей, и обобщающие показатели, рассчитываемые в целом по совокупности. Как те, так и другие могут быть исчислены за один период или как средние за несколько периодов.

Важной задачей статистического изучения структуры является определение степени концентрации и централизации признака.

Концентрация - неравномерность распределения изучаемого признака внутри совокупности, не связанная с общим ее объемом.

Централизация - сосредоточение объема признака у отдельных единиц или неравномерность его распределения с учетом объема совокупности.

Необходимо уяснить различие этих понятий. Так, при нулевой концентрации вполне возможна сильная централизация и, наоборот, на фоне слабой централизации допустима высокая концентрация.

ТЕСТ К ГЛАВЕ 11

1. К какому виду по временному фактору относится половозрастная структура населения?

а) моментная;
б) интервальная.

2. Может ли темп роста удельного веса быть отрицательной величиной?

а) не может;
б) может в случае снижения удельного веса.

3. Средний "абсолютный" прирост удельного веса определяется по формуле:

а) средней арифметической;
б) средней геометрической.

4. Должна ли сумма средних темпов роста всех структурных частей исследуемой совокупности быть строго равной 100%?

а) должна;
б) не должна.

5. Должна ли сумма средних удельных весов всех структурных частей исследуемой совокупности быть строго равной 100%?

а) должна;
б) не должна.

6. Могут ли линейные и квадратические обобщающие коэффициенты, рассчитанные для сравнения структурных изменений в двух совокупностях, привести к противоположным выводам?

а) могут;
б) не могут.

7. Может ли кривая концентрации частично совпадать с лин!|ей равномерного распределения?

а) может;
б) не может.

8. Коэффициент Лоренца позволяет получить сводную оценку структурных сдвигов:

а) в процентных пунктах;
б) в процентах к удельному весу наибольшей структурной части;
в) в процентах к предельной величине структурных сдвигов.

9. Может ли высокая концентрация сопровождаться сильной централизацией?

а) может;
б) не может.

10. В каких границах может находиться коэффициент централизации в том случае, если все производство сосредоточено только на двух лредприятиях?

а) **от 0 до** 0,5;

б) от 0,5 до 1,0.

ГЛАВА 12

ЭКОНОМИЧЕСКИЕ ИНДЕКСЫ

12.1

ПОНЯТИЕ ЭКОНОМИЧЕСКИХ ИНДЕКСОВ. КЛАССИФИКАЦИЯ ИНДЕКСОВ

Индексы относятся к важнейшим обобщающим показателям. Слово «индекс» имеет несколько значений: показатель, указатель, опись, реестр. Оно используется как понятие в математике, экономике, в метеорологии и других науках.

В статистике под индексом понимается относительный показатель, который выражает соотношение величин какого-либо явления во времени, в пространстве или сравнение фактических данных с любым эталоном (план, прогноз, норматив и т. д.).

В международной практике индексы принято обозначать символами i и I (начальная буква латинского слова index). Буквой «i» обозначаются индивидуальные (частные) индексы, буквой «I» общие индексы. Знак внизу справа означает период: 0 - базисный; 1 - отчетный. Помимо этого используются определенные символы для обозначения индексируемых показателей:

q - количество (объем) какого-либо товара в натуральном выражении;

p - цена единицы товара;

z - себестоимость единицы продукции;

t - затраты времени на производство единицы продукции;

w - выработка продукции в стоимостном выражении на одного рабочего или в единицу времени;

v - выработка продукции в натуральном выражении на одного рабочего или в единицу времени;

T - общие затраты времени (tq) или численность рабочих;

pq - стоимость продукции или товарооборот;

zq - издержки производства.

Все экономические индексы можно классифицировать по следующим признакам:

- степень охвата явления;
- база сравнения;
- вид весов (соизмерителя);
- форма построения;
- характер объекта исследования;
- объект исследования;
- состав явления;
- период исчисления.

По степени охвата явления индексы бывают индивидуальные и сзодные. **Индивидуальные индексы** служат для характеристики изменения отдельных элементов сложного явления. Их примером могут быть изменения объема производства отдельных видов продукции (телевизоров, электроэнергии и т. д.), а также цен на акции какого-либо предприятия. Для измерения динамики сложного явления, составные части которого непосредственно несоизмеримы (изменения физического объема продукции, включающей разноименные товары, индекса цен акций предприятий региона и т. п.), рассчитывают **сводные,** или **общие, индексы.**

Если индексы охватывают не все элементы сложного явления, а только часть их, то такие индексы называются **групповыми или субиндексами,** например индексы физического объема продукции по отдельным отраслям промышленности, индексы цен по группам продовольственных и непродовольственных товаров. Групповые индексы отражают закономерности в развитии отдельных частей изучаемых явлений. В таких индексах проявляется их связь с методом группировок.

По базе сравнения все индексы можно разделить на две группы: **динамические и территориальные.** Первая группа индексов отражает изменение явления во времени. Например, индекс цен на продукцию в 1996 г. по сравнению с предыдущим годом; индекс стоимости потребительской корзины в августе по сравнению с июлем 1997 г.

При исчислении динамических индексов происходит сравнение значения показателя в отчетный период со значением этого же показателя за предыдущий период, который называют базисным. Однако в качестве последнего могут быть использованы и прогнозные, и плановые показатели.

Динамические индексы бывают базисные и цепные.

Вторая группа индексов **(территориальные)** применяется для межрегиональных сравнений. Большое значение эти индексы имеют в международной статистике при сопоставлении показателей социально-экономического развития различных стран. Например, индекс цен на фототовары в Италии по сравнению с Германией, индекс стоимости потребительской корзины в Москве по сравнению с Санкт-Петербургом.

По виду весов индексы бывают с постоянными и переменными весами.

В зависимости от формы построения различаются индексы агрегатные и средние. Последние делятся на арифметические и гармонические. Агрегатная форма общих индексов является основной формой экономических индексов. Средние индексы производные, они получаются в результате преобразования агрегатных индексов.

По характеру объема исследования общие индексы подразделяются на индексы количественных (объемных) и качественных показателей. В основе такого деления индексов лежит вид индексируемой величины. К первой группе индексов относятся, например, индексы объема продаж долларов США на Московской межбанковской валютной бирже, а ко второй - индекс курса немецкой марки.

По объекту исследования индексы бывают: производительности труда, себестоимости, физического объема продукции, стоимости продукции и т. д.

По составу явления можно выделить две группы индексов: постоянного (фиксированного) состава и переменного состава. Деление индексов на эти две группы используется для анализа динамики средних показателей.

По периоду исчисления индексы подразделяются на годовые, квартальные, месячные, недельные.

С помощью экономических индексов решаются следующие •задачи:

- измерение динамики социально-экономического явления за два и более периодов времени;
- измерение динамики среднего экономического показателя;
- измерение соотношения показателей по разным регионам;
- определение степени влияния изменений значений одних показателей на динамику других;
- пересчет значения макроэкономических показателей из фактических цен в сопоставимые.

Каждая из этих задач решается с помощью различных индексов.

12.2

ИНДИВИДУАЛЬНЫЕ И ОБЩИЕ ИНДЕКСЫ

Индивидуальные индексы получают в результате сравнений однотоварных явлений. Например, индекс цен на подсолнечное масло определяется как отношение цены на этот товар в текущем периоде к цене базисного периода.

Индивидуальные индексы представляют собой относительные величины динамики, выполнения плана, сравнения, и их расчет не требует знания специальных правил.

В зависимости от экономического назначения индивидуальные индексы бывают: физического объема продукции, себестоимости, цен, трудоемкости и т. д.

Индекс физического объема продукции i рассчитывается по формуле

$$i_q = \frac{q_1}{q_0}. \quad (12.1)$$

Этот индекс показывает, во сколько раз возрос (уменьшился) выпуск какого-либо одного товара в отчетном периоде по сравнению с базисным, или сколько процентов составляет рост (снижение) выпуска товара. Если из значения индекса, выраженного в процентах, вычесть 100%, то полученная величина покажет, на сколько процентов возрос (уменьшился) выпуск продукции. В знаменателе может быть не только количество продукции, произведенной за какой-то предыдущий период, но и плановое значение (q_{nil}), нормативное (q_H) или эталонное значение, принятое за базу сравнения ($ц_э$). Тогда формула (12.1) примет соответственно следующий вид:

(12.2) (12.3) (12.4)

$$i_q = \frac{q_1}{[illegible]};$$

$$i_p = \frac{p_1}{p_0}$$

$$i_q = \frac{}{q_э}.$$

(12.5)

характеризует изменение цены одного определенного товара в текущем

Индексы других показателей строятся аналогично.

Индивидуальный индекс цен

периоде по сравнению с базисным.

Индивидуальный индекс себестоимости единицы продукции

(12.6) $i_z = \frac{z_1}{z_0}$,

показывает изменение себестоимости единицы продукции в текущем периоде по сравнению с базисным.

Производительность труда может быть измерена количеством продукции, производимой в единицу времени (v), или затратами рабочего времени на производство единицы продукции (t). Поэтому можно построить:

- индекс количества продукции, произведенной в единицу времени:

W

ставляет рост (снижение) стоимости товара, и определяется по Лопмуле

(12.11) $i_{pq} = \frac{p_1q_1}{p_0q_0}$.

Индивидуальный индекс численности рабочих можно рассчитать следующим образом:

$i_v = \frac{v_1}{v_0} = \frac{q_1}{T_1}:\frac{q_0}{T_0}$; $i_T = \frac{T_1}{T_0} = \frac{t_1q_1}{t_0q_0}$.

(12.7)

(12.12)

- индекс производительности труда по трудовым затратам:

$i_t = \frac{t_0}{t_1}$. (12.8)

Так как между количеством продукции, произведенной в единицу времени, и затратами рабочего времени на производство единицы продукции существует обратно пропорциональная зависимость, т. е.

(12.9) $t = \frac{1}{v}$,

то индекс (12.8) получается в результате деления величины показателя в базисном периоде на величину в текущем периоде.

Для характеристики производительности труда часто используется индивидуальный индекс выработки продукции в стоимостном выражении на одного рабочего:

(12.10) $i_w = \frac{w_1}{w_0} = \frac{q_1p}{T_1}:\frac{q_0p}{T_0}$,

где р — сопоставимые цены.

Индивидуальные индексы (12.7 и 12.10) показывают, во сколько раз производительность труда в базисном периоде выше (ниже), чем в отчетном.

Индекс, исчисленный по формуле (12.8), показывает, во сколько раз производительность труда в базисном периоде выше (ниже), чем в отчетном.

Индивидуальный индекс стоимости продукции отражает, во сколько раз изменилась стоимость какого-либо товара в текущем периоде по сравнению с базисным, или сколько процентов со-

Он показывает, во сколько раз изменилась численность рабочих в текущем периоде по сравнению с базисным, или сколько процентов составляет рост (снижение) численности рабочих.

По данным о цене, количестве и стоимости проданных товаров (табл. 12.1) рассчитаем индивидуальные индексы. Результаты расчетов (представлены в гр. 7 - 9 этой таблицы) показывают, что больше всего возросли цены на кофе - почти на 6%, а индекс стоимости товаров и физического объема продукции самый высокий по чаю - соответственно 5,2 и 5 раз.

В экономических расчетах чаще всего используются общие индексы, которые характеризуют изменение совокупности в целом. Построение этих индексов и является содержанием индексной методологии. В индексной теории сложились две концепции: *синтетическая* и *аналитическая.* Они по-разному интерпретируют общие индексы.

Согласно **синтетической концепции** особенность общих индексов состоит в том, что они выражают относительное изменение сложных (разнотоварных) явлений, отдельные части или элементы которых непосредственно несоизмеримы, и поэтому индексы - показатели синтетические. Например, промышленные предприятия производят несколько видов продукции, имеющей различное назначение. Следовательно, путем суммирования количества произведенных товаров различных видов нельзя получить показатель физического объема продукции. Методология построения общих индексов предусматривает прежде всего приведение разнотоварных явлений к соизмеримому виду.

В аналитической теории индексы трактуются как показатели, необходимые для измерения влияния изменения составных частей, компонентов, факторов сложного явления на изменение уровня этого явления. Например, изменение общей величины

Таблица 12.1

Цена и количество проданных товаров

Товар	Единица измерения	Цена, руб.		Количество проданных товаров		Стоимость проданной продукции, тыс. руб.		Индивидуальный индекс, %			Стоимость продукции, проданной в мае, в ценах апреля, тыс. руб. p_0q_1	$i_q \cdot p_0q_0$	$\frac{p_1q_1}{i_p}$	Стоимость продукции, проданной в апреле, в ценах мая, тыс. руб. p_1q_0
		апрель p_0	май p_1	апрель q_0	май q_1	апрель p_0q_0	май p_1q_1	цен $i_p = \frac{p_1}{p_0}$	физического объема продукции $i_q = \frac{q_1}{q_0}$	стоимость $i = \frac{p_1q_1}{p_0q_0}$				
А	Б	1	2	3	4	5	6	7=2:1	8=4:3	9=6:5	10=1·4	11=8·5	12=6:7	13=2·3
Чай	Пачка	1 638	1 704	1 000	5 000	1 638	8 520	104,03	500	520	8 190	8 190	8 190	1 704
Кофе	Банка	6 925	7 340	2 000	2 500	13 850	18 350	105,99	125	132	17312,5	17312,5	17312.5	14 680
Сыр	Кг	5 040	5 240	400	500	2 016	2 620	103,97	125	130	2 520	2 520	2 520	2 096
Всего		–	–	–	–	17 504	29 490	–	–	–	28022,5	28022,5	28022,5	18 480

товарооборота в текущем периоде по сравнению с базисным связано как с изменением физического объема продаж товаров, так и с изменением цен по каждому виду товаров. Поэтому индексная методология предусматривает определение влияния каждого из факторов путем элиминирования влияния других факторов на уровень изучаемого явления.

Таким образом, общие индексы являются синтетическими и аналитическими показателями.

Общие индексы строят для количественных (объемных) и качественных показателей. В зависимости от цели исследования и наличия исходных данных используют различные форму построения общих индексов: агрегатную или средневзвешенную.

12.3

АГРЕГАТНЫЙ ИНДЕКС КАК ИСХОДНАЯ ФОРМА ИНДЕКСА

Агрегатный индекс - сложный относительный показатель, который характеризует среднее изменение социально-экономического явления, состоящего из несоизмеримых элементов.

Латинское слово «агрегат» (aggregatus) означает «складываемый, суммируемый». Особенность этой формы индекса состоит в том, что в агрегатной форме непосредственно сравниваются две суммы одноименных показателей. В настоящее время это наиболее распространенная форма индексов, используемая в практической статистике многих стран мира.

Числитель и знаменатель агрегатного индекса представляют собой сумму произведений двух величин, одна из которых меняется (индексируемая величина), а другая остается неизменной в числителе и знаменателе (вес индекса).

Индексируемой величиной называется признак, изменение которого изучается (цена товаров, курс акций, затраты рабочего времени на производство продукции, количество проданных товаров и т. д.). **Вес индекса** - это величина, служащая для целей соизмерения индексируемых величин.

За каждым экономическим индексом стоят определенные экономические категории. Экономическое содержание индекса предопределяет методику его расчета.

Методика построения агрегатного индекса предусматривает решение трех вопросов:

1) какая величина будет индексируемой;
2) по какому составу разнородных элементов явления необходимо исчислить индекс;
3) что будет служить весом при расчете индекса.

При выборе веса индекса принято руководствоваться следующим правилом: если строится индекс количественного показателя, то веса берутся за базисный период, при построении индекса качественного показателя используются веса отчетного периода.

Построим три индекса: стоимости продукции, физического объема продукции и цен.

Стоимость продукции - это произведение количества продукции в натуральном выражении (q) на ее цену (p).

Индекс стоимости продукции, или **товарооборота** (I_{pq}), представляет собой отношение стоимости продукции текущего периода (Zp,q|) к стоимости продукции в базисном периоде (Zp_0q_0) и определяется по формуле

$$(12.13)\quad I_{pq} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}.$$

Такой индекс показывает, во сколько раз возросла (уменьшилась) стоимость продукции (товарооборота) отчетного периода по сравнению с базисным, или сколько процентов составляет рост (снижение) стоимости продукции. Если из значения индекса стоимости (12.13) вычесть 100% (I - 100), то разность покажет, на сколько процентов возросла (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным. Разность числителя и знаменателя (*I.p$_l$q$_l$* - ^p_0Яо) показывает, на сколько рублей увеличилась (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным. Аналогично строятся индексы для показателей, которые являются произведением двух сомножителей: издержек производства (произведение себестоимости единицы продукции на количество продукции); затрат времени на производство всей продукции (произведение затрат времени на производство единицы продукции на количество выработанной продукции).

Рассчитаем индекс стоимости (товарооборота) продукции по данным табл. 12.1.

$$I_{pq} = \frac{29\,490}{17\,504} = 1{,}685, \text{ или } 168{,}5\%.$$

Следовательно, стоимость продукции (товарооборота) в мае по сравнению с апрелем возросла почти в 1,7 раза (рост составил 168,5%). Стоимость продукции увеличилась на 168,5% - 100% =

68,5%, или на 11 986 тыс. руб. (29 940-17 504). Значение индекса стоимости продукции (товарооборота) зависит от двух факторов: изменения количества продукции и цен, что обусловливает возможность и необходимость построения еще двух индексов: физического объема продукции и цен.

Индекс физического объема продукции - это индекс количественного показателя. В этом индексе индексируемой величиной будет количество продукции в натуральном выражении, а весом - цена. Только умножив несоизмеримые между собой количества разнородной продукции на их цены, можно перейти к стоимостям продукции, которые будут уже величинами соизмеримыми. Так как индекс физического объема - индекс количественного показателя, то весами будут цены базисного периода. Тогда формула индекса примет следующий вид:

$$I_q = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}, \quad (12.14)$$

где в числителе дроби - условная стоимость произведенных в текущем периоде товаров в ценах базисного периода, а в знаменателе - фактическая стоимость товаров, произведенных в базисном периоде. Если объектом исследования является отдельное предприятие, то индекс определяется по совокупности произведенных товаров; когда объект исследования - отрасль промышленности, индекс рассчитывается по совокупности всех товаров, произведенных в отрасли, или отдельным их группам в зависимости от цели анализа. Если же объектом исследования является какой-либо регион, то индекс рассчитывается по товарам, произведенным предприятиями региона.

Индекс физического объема продукции (12.14) показывает, во сколько раз возросла (уменьшилась) стоимость продукции из-за роста (снижения) объема ее производства или сколько процентов составляет рост (снижение) стоимости продукции в результате изменения физического объема ее производства. Если из значения индекса физического объема продукции (12.14) вычесть 100% (l_q- 100), то разность покажет, на сколько процентов возросла (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным из-за роста (снижения) объема ее производства. Разность числителя и знаменателя (lqiP$_0$ - ^ЯоРо) показывает, на сколько рублей изменилась стоимость продукции в результате роста (уменьшения) ее объема. Изменение цен на продукцию в текущем периоде по сравнению с базисным не влияет на величину индекса. Для расчета индекса воспользуемся данными табл. 12.1.

$$I_q = \frac{28\,022{,}5}{17\,504} = 1{,}6009, \text{ или } 160{,}09\%.$$

Следовательно, стоимость продукции в мае по сравнению с апрелем увеличилась в 1,6 раза (или рост стоимости составил 160%). Вычитая из числителя индекса знаменатель, получим:

$$28\,022{,}5 - 17\,504 = 10\,518{,}5 \text{ тыс. руб.},$$

т. е. за счет увеличения объема продукции на 60% (160% - 100%) ее стоимость в абсолютном выражении увеличилась на 10518,5 тыс. руб.

При построении агрегатного индекса цен, который в условиях рыночной экономики является наиболее широко распространенным показателем инфляции, исходят из тех же предпосылок, что и при построении индекса физического объема продукции.

Индекс цен - это индекс качественного показателя. Индексируемой величиной будет цена товара, так как этот индекс характеризует изменение цен. Весом будет выступать количество произведенных товаров. Умножив цену товара на его количество, получаем величину, которую можно суммировать и которая представляет собой показатель, соизмеримый с другими подобными ему величинами.

Индекс цен определяется по следующей формуле:

$$I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}, \quad (12.15)$$

где в числителе дроби - фактическая стоимость продукции текущего периода, а в знаменателе - условная стоимость тех же товаров в ценах базисного периода.

Индекс показывает, во сколько раз возросла (уменьшилась) стоимость продукции из-за изменения цен, или сколько процентов составляет рост (снижение) стоимости продукции в результате изменения цен. Если из значения индекса (12.15) вычесть 100% (1_p - 100%), то разность покажет, на сколько процентов возросла (уменьшилась) стоимость продукции из-за изменения цен, а разность числителя и знаменателя (Zpi4i - Spo4i) ~ ^{на} сколько рублей изменилась стоимость продукции в результате роста (снижения) цен. Изменение количества произведенной продукции в текущем периоде по сравнению с базисным не влияет на величину индекса.

Определим индекс цен по данным табл. 12.1:

$$I_p = \frac{29\,490}{28\,022,5} = 1,0523, \text{ или } 105,23\%.$$

Таким образом, в среднем по трем товарам цены выросли в 1,0523 раза (или рост цен составил 105,23%). В результате увеличения цен на 5,23% (105,12% - 100%) покупатели заплатили на 1 467,5 тыс. руб. больше в мае, чем в апреле (29 490 - 28 022,5 = 1 467,5).

Как уже отмечалось выше, стоимость продукции можно представить как произведение количества товара на его цену. Точно такая же связь существует и между индексами стоимости, физического объема и цен, т. е.

(12.16)

$$I_{pq} = I_p \cdot I_q$$

или

(12.17) $$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \cdot \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}.$$

Разность числителя и знаменателя каждого индекса-сомножителя выражает размер изменения общей абсолютной величины под слиянием изменения одного фактора. Алгебраическая сумма этих разностей равна разности числителя и знаменателя индекса стоимости продукции:

$$(\Sigma q_1 p_0 - \Sigma q_0 p_0) + (\Sigma q_1 p_1 - \Sigma q_1 p_0) = \Sigma q_1 p_1 - \Sigma q_0 p_0. \quad (12.18)$$

Равенства (12.16 - 12.18) выполняются в том случае, если при исчислении индекса объемного показателя веса были зафиксированы на уровне базисного периода, а при расчете индекса качественного показателя - на уровне отчетного периода.

Для нашего примера (табл. 12.1) индекс стоимости, исчисленный по формуле (12.16), равен:

$$1,6009 \cdot 1,0523 = 1,685, \text{ или } 168,5\%,$$

а алгебраическая сумма этих разностей (формула 12.18) равна:

$$10\,518,5 + 1\,467,5 = 11\,986 \text{ тыс. руб.}$$

Формулы для расчета общих индексов других показателей приведены в табл. 12.2.

Таблица 12.2

Основные формулы исчисления сводных, или общих, индексов

Наименование индекса	Формула расчета индекса	Что показывает индекс	Что показывает значение индекса, уменьшенное на 100%, т. е. I – 100	Что показывает разность числителя и знаменателя
Индекс физического объема продукции	$I_q = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}$	Во сколько раз изменилась стоимость продукции в результате изменения объема ее производства, или сколько процентов составил рост (снижение) стоимости продукции из-за изменения ее физического объема	На сколько процентов изменилась стоимость продукции в результате изменения объема ее производства	На сколько рублей изменилась стоимость продукции в результате роста (уменьшения) объема ее производства
Индекс цен	$I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$	Во сколько раз изменилась стоимость продукции в результате изменения цен, или сколько процентов составил рост (снижение) стоимости продукции из-за изменения цен	На сколько процентов изменилась стоимость продукции в результате изменения цен	На сколько рублей изменилась стоимость продукции в результате роста (уменьшения) цен
Индекс издержек производства	$I_{zq} = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_0}$	Во сколько раз возросли (уменьшились) издержки производства продукции, или сколько процентов составил рост (снижение) издержек производства продукции в текущем периоде по сравнению с базисным	На сколько процентов возросли (уменьшились) издержки производства продукции в текущем периоде по сравнению с базисным	На сколько рублей увеличились (уменьшились) издержки производства продукции в текущем периоде по сравнению с базисным
Индекс физического объема продукции	$I_q = \frac{\Sigma q_1 t_0}{\Sigma q_0 t_0}$	Во сколько раз изменились затраты времени на производство продукции в результате изменения объема ее производства, или сколько процентов составил рост (снижение) затрат времени на производство продукции из-за изменения физического объема ее производства	На сколько процентов изменились затраты времени на производство продукции в результате изменения объема ее производства	На сколько человеко-часов возросли (уменьшились) затраты времени на производство продукции в результате роста (уменьшения) объема производства продукции

12.4

СРЕДНИЕ ИНДЕКСЫ

Помимо агрегатных индексов в статистике применяется другая их форма - средневзвешенные индексы. К их исчислению прибегают тогда, когда имеющаяся в распоряжении информация не позволяет рассчитать общий агрегатный индекс. Так, если отсутствуют данные о ценах, но имеется информация о стоимости продукции в текущем периоде и известны индивидуальные индексы цен по каждому товару, то нельзя определить общий индекс цен как агрегатный, но возможно исчислить его как средний из индивидуальных. Точно так же, если не известны количества произведенных отдельных видов продукций, но известны индивидуальные индексы и стоимость продукции базисного периода, можно определить общий индекс физического объема продукции как средневзвешенную величину.

Средний индекс - это индекс, вычисленный как средняя величина из индивидуальных индексов. Агрегатный индекс является основной формой общего индекса, поэтому средний индекс должен быть тождествен агрегатному индексу. При исчислении средних индексов используются две формы средних: арифметическая и гармоническая.

Средний арифметический индекс тождествен агрегатному индексу, если весами индивидуальных индексов будут слагаемые знаменателя агрегатного индекса. Только в этом случае величина индекса, рассчитанного по формуле средней арифметической, будет равна агрегатному индексу.

Средний арифметический индекс физического объема продукции вычисляется по формуле

$$I_q = \frac{\Sigma i_q p_0 q_0}{\Sigma p_0 q_0}, \qquad (12.19)$$

Так как i_t x t, = t_0, то формула этого индекса может быть преобразована в агрегатный индекс трудоемкости продукции. Весами являются общие затраты времени на производство продукции в текущем периоде.

В статистике широко известен и другой средний арифметический индекс, который используется при анализе.производительности труда. Он носит название **индекса Струмилина** и определяется следующим образом:

$$I_v = \frac{\Sigma\left(\frac{q_1}{T_1}:\frac{q_0}{T_0}\right)\cdot T_1}{\Sigma T_1}.$$

Индекс производительности труда по трудовым затратам	$I_t = \frac{[illegible]}{\Sigma t_1 q_1}$	[illegible]ько раз увеличилась (уменьшилась) производительность труда, или сколько процентов составило снижение (рост) производительности труда в текущем периоде по сравнению с базисным	На сколько процентов изменилась производительность труда в текущем периоде по сравнению с базисным	Абсолютный размер экономии (перерасхода) затрат живого труда в связи с ростом (уменьшением) его производительности
Индекс затрат времени на производство продукции	$I_{tq} = \frac{\Sigma t_1 q_1}{\Sigma t_0 q_0}$	Во сколько раз изменились затраты времени на производство продукции, или сколько процентов составил рост (снижение) затрат времени на производство продукции в текущем периоде по сравнению с базисным	На сколько процентов изменились затраты времени на производство продукции в текущем периоде по сравнению с базисным	На сколько человеко-часов увеличились (уменьшились) затраты на производство продукции в текущем периоде по сравнению с базисным

(12.21)

Индекс показывает, во сколько раз возросла (уменьшилась) производительность труда, или сколько процентов составил рост (снижение) производительности труда в среднем по всем единицам исследуемой совокупности.

Средние арифметические индексы чаще всего применяются на практике для расчета сводных индексов количественных показателей, а из качественных показателей - для исчисления двух приведенных выше индексов (12.20 - 12.21).

Индексы других качественных показателей (цен, себестоимости и т. д.) определяются по формуле средней гармонической взвешенной величины.

Средний гармонический индекс тождествен агрегатному, если индивидуальные индексы будут взвешены с помощью слагаемых числителя агрегатного индекса. Например, индекс себестоимости можно исчислить так:

$$I_z = \frac{\sum z_1 q_1}{\sum \frac{z_1 q_1}{i_z}}, \quad (12.22)$$

а индекс цен:

Так как i_q х q_0 = q,, то формула этого индекса легко преобразуется в формулу (12.14). Весами в формуле (12.19) является стоимость продукции базисного периода.

Средний арифметический индекс производительности труда определяется следующим образом:

$$(12.20) \quad I_t = \frac{\sum i_t t_1 q_1}{\sum t_1 q_1} = \frac{\sum i_t T_1}{\sum T_1}.$$

$$(12.23) \quad I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}.$$

Таким образом, весами при определении среднего гармонического индекса себестоимости являются издержки производства текущего периода, а индекса цен - стоимость продукции этого периода.

Рассчитаем средние индексы цен и физического объема продукции по данным табл. 12.1 (гр. 11-12):

$$I_q = \frac{28022,5}{17504} = 1,6009, \text{ или } 160,09\%;$$

$$I_p = \frac{29490}{28022,5} = 1,0523, \text{ или } 105,23\%;$$

Этот же результат получился при расчете агрегатных индексов. Средние индексы широко используются для анализа рынка ценных бумаг. Наиболее известными являются индексы ДоуДжонса, Стэндарда и Пура.

Индекс Доу-Джонса (Dow Jones Industrial Average Index) определяется как средний арифметический индекс значений курсов акций, котирующихся на Нью-Йоркской фондовой бирже. Один сводный и три групповых индекса рассчитываются каждые полчаса, и ежедневно публикуется их значение на момент, закрытия биржи. Групповые индексы определяются по ценам акций 30 промышленных, 20 транспортных и 15 компаний сферы услуг. Общий индекс рассчитывается по всем 65 компаниям. Их перечень был составлен в 1928 г. В качестве базисного выбран 1920 г. Первоначальная методика исчисления индекса была разработана основателем и редактором крупнейшей в США газеты «Уолл-стрит джорнел» Чарлзом Доу.

Индекс Стэндарда и Пура (Standard and Poor's 500 Stock Index) - индекс, рассчитываемый по курсам акций 500 крупнейших компаний Нью-Йоркской фондовой биржи как средний взвешенный показатель, учитывающий общее число выпущенных компанией акций. В число компаний, акции которых включены в индекс, входят 400 промышленных корпораций, 40 - финансовых, 20 - транспортных и 40 - сферы услуг.

12.5

ВЫБОР БАЗЫ И ВЕСОВ ИНДЕКСОВ

Выбор базы сравнения и весов индексов - это два важнейших методологических вопроса построения систем индексов. Система используется при изучении динамики социально-экономических явлений за некоторый интервал времени, включающий более двух периодов времени.

Системой индексов называется ряд последовательно построенных индексов. Такие системы характеризуют изменения, происходящие в изучаемом явлениу в течение исследуемого периода времени.

В зависимости от базы сравнения системы индексов бывают базисными и цепными.

Система базисных индексов - это ряд последовательно вычисленных индексов одного и того же явления с постоянной базой сравнения, т. е. в знаменателе всех индексов находится индексируемая величина базисного периода.

Система цепных индексов - это ряд индексов одного и того же явления, вычисленных с меняющейся от индекса к индексу базой сравнения.

В экономико-статистических исследованиях выбор системы индексов (базисные или цепные) проводится в зависимости от цели анализа. Базисные индексы дают более наглядную характеристику общей тенденции развития исследуемого явления, а цепные - четче отражают последовательность изменения уровней во времени.

Системы цепных и базисных индексов могут быть построены для индивидуальных и общих индексов. Системы индивидуальных индексов стоимости продукции, физического объема продукции и цен (табл. 12.3) просты по построению. Аналогично им строятся системы индивидуальных индексов и для других показателей.

Таблица 12.3

Системы индивидуальных индексов

Название ального индекса	Система индексов	
	базисных	цепных
Индекс	P|Ч1 . P2Ч2 . p,,q,, PoЯo PoЧo PoЧo	Pl4l . P2Ч2 . . PпЧп PoЧo Pl4i Pn-l4n-l
Индекс го объема	Ль 42 . Чп Чo Чo Чo	Ч| . 42 . . Чп Чo 4i 4,,-i
Индекс	Pi . P2 . Pп Po Po Po	Pi . P2 . . P,, Po Pi P,,-i

Между цепными и базисными индексами существуют различные виды связи.

- Если известны цепные индексы, то путем их последовательного перемножения можно получить базисные индексы. Например,

$$\frac{p_1}{p_0}\cdot\frac{p_2}{p_1}=\frac{p_2}{p_0} \qquad (12.24)$$

или

$$\frac{q_1}{q_0}\cdot\frac{q_2}{q_1}\cdot\frac{q_3}{q_2}=\frac{q_3}{q_0}. \qquad (12.25)$$

- Зная последовательные значения базисных индексов, легко рассчитать на их основе цепные индексы:

$$\frac{p_2}{p_1}=\frac{p_2}{p_0}:\frac{p_1}{p_0} \qquad (12.26)$$

или

$$\frac{q_3}{q_2}=\frac{q_3}{q_0}:\frac{q_2}{q_0}. \qquad (12.27)$$

Системы базисных и цепных индексов могут быть построены для агрегатных индексов.

Система индексов стоимости имеет следующий вид: • цепные индексы:

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}; \quad \frac{\Sigma p_2 q_2}{\Sigma p_1 q_1}; \quad \ldots; \quad \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_{n-1}};$$

базисные индексы:

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}; \quad \frac{\Sigma p_2 q_2}{\Sigma p_0 q_0}; \quad \ldots; \quad \frac{\Sigma p_n q_n}{\Sigma p_0 q_0}.$$

Формирование системы индексов, например, цен или физического объема отличается от уже рассмотренных в этом параграфе систем индексов. Это связано с тем, что при построение систем этих индексов можно использовать постоянные и переменные веса.

Системой индексов с постоянными весами называется система сводных индексов одного и того же явления, вычисленных с весами, не меняющимися при переходе от одного индекса к другому. Постоянные веса позволяют исключить влияние изменения структуры на величину индекса.

Например, система базисных индексов физического объема продукции с постоянными весами (p_0) имеет следующий вид:

$$\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}; \quad \frac{\Sigma q_2 p_0}{\Sigma q_0 p_0}; \quad \ldots; \quad \frac{\Sigma q_n p_0}{\Sigma q_0 p_0},$$

а систему цепных индексов с теми же постоянными весами можно представить так:

$$\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}; \quad \frac{\Sigma q_2 p_0}{\Sigma q_1 p_0}; \quad \ldots; \quad \frac{\Sigma q_n p_0}{\Sigma q_{n-1} p_0}.$$

Система индексов с переменными весами представляет собой систему сводных индексов одного и того же явления, вычисленных с весами, последовательно меняющимися от одного индекса к другому. Переменные веса - это веса отчетного периода. Например, система базисных индексов цен с переменными весами следующая:

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}; \quad \frac{\Sigma p_2 q_2}{\Sigma p_0 q_2}; \quad \ldots; \quad \frac{\Sigma p_n q_n}{\Sigma p_0 q_n}.$$

Элементами этой системы являются индексы-дефляторы, которые необходимы для пересчета стоимостных показателей системы национальных счетов в сопоставимые цены.

Система цепных индексов цен с переменными весами выглядит так:

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}; \quad \frac{\Sigma p_2 q_2}{\Sigma p_1 q_2}; \quad \ldots; \quad \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_n}.$$

Отдельные индексы этой системы используются для пересчета стоимостных показателей отчетного периода в цены предыдущего периода.

Системы общих индексов других показателей строятся аналогично.

Системы агрегатных индексов обладают теми же свойствами, что и системы индивидуальных индексов, т. е., зная базисные индексы, можно рассчитать цепные; при наличии цепных индексов легко получить соответствующие им базисные. Например,

$$(12.28) \quad \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_2 q_2}{\Sigma p_1 q_1} = \frac{\Sigma p_2 q_2}{\Sigma p_0 q_0}$$

или

(12.29)

$$\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \cdot \frac{\Sigma q_2 p_0}{\Sigma q_1 p_0} \cdot \frac{\Sigma q_3 p_0}{\Sigma q_2 p_0} = \frac{\Sigma q_3 p_0}{\Sigma q_0 p_0},$$

или

(12.30)

$$\frac{\Sigma p_2 q_2}{\Sigma p_0 q_0} : \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{\Sigma p_2 q_2}{\Sigma p_1 q_1},$$

или

(12.31)

$$\frac{\Sigma p_n q_n}{\Sigma p_0 q_n} : \frac{\Sigma p_{n-1} q_n}{\Sigma p_0 q_n} = \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_n}.$$

12.6

ИНДЕКСЫ СТРУКТУРНЫХ СДВИГОВ

При изучении динамики качественных показателей приходится определять изменение средней величины индексируемого показателя, которое обусловлено взаимодействием двух факторов - изменением значения индексируемого показателя у отдельных групп единиц и изменением структуры явления. Под **изменением структуры явления** понимается изменение доли отдельных групп единиц совокупности в общей их численности. Так, средняя заработная плата на предприятии может вырасти в результате роста оплаты труда работников или увеличения доли высокооплачиваемых сотрудников. Снижение трудоемкости производства единицы продукции по совокупности предприятий отрасли может быть обусловлено повышением производительности труда на предприятиях или концентрацией производства продукции на заводах с низкой трудоемкостью. Так как на изменение среднего значения показателя оказывают воздействие два фактора, возникает задача определить степень влияния каждого из факторов на общую динамику средней.

Эта задача решается с помощью индексного метода, т. е. путем построения системы взаимосвязанных индексов, в которую включаются три индекса: переменного состава, постоянного состава и структурных сдвигов.

Индексом переменного состава называется индекс, выражающий соотношение средних уровней изучаемого явления, относящихся к разным периодам времени. Например, индекс переменного состава себестоимости продукции одного и того же вида рассчитывается по формуле

$$(12.32)\quad I_{пс} = \frac{\bar{z}_1}{\bar{z}_0} = \frac{\Sigma z_1 q_1}{\Sigma q_1} : \frac{\Sigma z_0 q_0}{\Sigma q_0},$$

где $1_{ПС}$ - индекс переменного состава.

Индекс переменного состава отражает изменение не только индексируемой величины (в данном случае себестоимости), но и структуры совокупности (весов).

Индекс постоянного (фиксированного) состава - это индекс, исчисленный с весами, зафиксированными на уровне одного какого-либо периода, и показывающий изменение только индексируемой величины. Индекс фиксированного состава определяется как агрегатный индекс. Так, индекс фиксированного состава себестоимости продукции рассчитывают по формуле

$$(12.33)\quad I_{фс} = \frac{\Sigma z_1 q_1}{\Sigma q_1} : \frac{\Sigma z_0 q_1}{\Sigma q_1} = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1},$$

где $1_{фс}$ - индекс фиксированного состава.

Под индексом структурных сдвигов понимают индекс, характеризующий влияние изменения структуры изучаемого явления на динамику среднего уровня этого явления. Индекс определяется по формуле (при изучении изменения среднего уровня себестоимости)

$$I_{сс} = \frac{\Sigma z_0 q_1}{\Sigma q_1} : \frac{\Sigma z_0 q_0}{\Sigma q_0} = \frac{\Sigma z_0 q_1}{\Sigma z_0 q_0} : \frac{\Sigma q_1}{\Sigma q_0}, \qquad (12.34)$$

где $1_{ОС}$ - индекс структурных сдвигов.

Система взаимосвязанных индексов при анализе динамики средней себестоимости имеет следующий вид:

$$\underset{\text{Индекс переменного состава}}{I_{пс}} = \underset{\text{Индекс фиксированного состава}}{I_{фс}} \cdot \underset{\text{Индекс структурных сдвигов}}{I_{сс}} \qquad (12.35)$$

Рассмотрим применение такой системы на конкретном **примере.** Пусть имеются данные о себестоимости единицы продукции на трех предприятиях в текущем и базисном периодах (табл. 12.4).

Таблица 12.4

Количество произведенной продукции и себестоимость единицы продукции одного вида по трем предприятиям отрасли

Номер приятия	Произведено продукции				Себестоимость продукции, тыс. руб.		Индивидуальные себестомости, %	Издержки млн руб.		
	всего единиц		% к итогу							
	базисный период Чо	текущий период 4i	базисный период	текущий период	базисный период z«	текущий период ^{z}i	; i	базисный период ЗДо	текущий период Z1fl1	^{2}оЧ1
А	1	2	3	4	5	6	7=6:5	8=5-1	9=6-2	10=5-2
1	1680	1500	70	50	20	20,3	101,5	33,6	30,45	30
2	480	600	20	20	18	18,4	102,2	8,64	11,04	10,8
j	240	900	10	30	15	15,5	103,3	3,6	13,95	13,5
Всег	2400	3000	100	100	19,1	18,48	_	45,8	55,4	54,3

$$\bar{z}_1 = \frac{55\,440}{3000} = 18{,}48 \text{ тыс. руб.}$$

Тогда

$$J_{пс} = 18{,}48 : 19{,}10 = 0{,}9675, \text{ или } 96{,}75\%.$$

Следовательно, средняя себестоимость по трем предприятиям снизилась в текущем периоде по сравнению с базисным на 3,25%, хотя на каждом из них в отдельности она возросла. Это - результат того, что исчисленный индекс учитывает влияние еще и структурного фактора.

Определим индекс себестоимости фиксированного состава.

$$I_{фс} = \frac{55{,}44}{54{,}3} = 1{,}021, \text{ или } 102{,}1\%.$$

Таким образом, себестоимость в текущем периоде по сравнению с базисным возросла в среднем на 2,1%.

Вычислим влияние изменения структуры на динамику средней себестоимости:

$$I_{сс} = \frac{54{,}3}{45.84} : \frac{3000}{2400} = 0{,}9476, \text{ или } 94{,}76\%.$$

Изменение доли предприятий в общем объеме произведенной продукции привело к снижению себестоимости на 5,24%.

Аналогично строятся системы индексов для других показателей. Так, для показателя производительности труда можно построить систему индексов, в которой:

$$I_{пс} = \bar{t}_0 : \bar{t}_1 = \frac{\Sigma T_0}{\Sigma q_0} :$$

$$I_{фс} = \frac{\Sigma t_0 q_1}{\Sigma t_1 q_1};$$

В текущем периоде по сравнению с базисным себестоимость производства продукции возросла на каждом предприятии (гр. 5-6); изменилась структура производства: уменьшилась доля первого предприятия в общем выпуске продукции, возросла доля третьего, а доля второго не изменилась (гр. 3-4).

Рассчитаем индекс переменного состава. Для. этого сначала определим среднюю себестоимость единицы продукции в текущем и базисном периодах:

(12.36)

(12.37)

(12.38) $$I_{сс} = \left[\frac{\Sigma T_0}{\Sigma q_0} : \frac{\Sigma T_1}{\Sigma q_1}\right] : \frac{\Sigma t_0 q_1}{\Sigma t_1 q_1} = \frac{\Sigma t_0 q_0}{\Sigma t_0 q_1} : \frac{\Sigma q_0}{\Sigma q_1}.$$

$$\bar{z}_0 = \frac{45\ 840}{2400} = 19{,}10 \text{ тыс. руб.};$$

12.7

ИНДЕКСЫ ПРОСТРАНСТВЕННОТЕРРИТОРИАЛЬНОГО СОПОСТАВЛЕНИЯ

В статистической практике часто возникает потребность в сопоставлении уровней экономического явления в пространстве: по странам, экономическим районам, областям, т. е. в исчислении территориальных индексов. При построении территориальных индексов приходится решать вопрос, какие веса использовались при их исчислении. Например, если стоит задача сравнить цены двух регионов (А и Б), то можно построить два индекса:

Эти формулы могут дать совершенно различное представление о соотношении уровней явления. Например, при расчете по формуле (12.39) значение признака будет ниже в регионе А, а по формуле (12.40) - в регионе Б.

Рассмотрим *пример* (табл. 12.5).

Рассчитав индексы по двум формулам, получаем:

$$I_{А/Б} = \frac{\Sigma p_А q_А}{\Sigma p_Б q_А} = \frac{28900}{23100} = 1{,}2510, \text{ или } 125{,}1\%$$

и

$$I_{Б/А} = \frac{\Sigma p_Б q_Б}{\Sigma p_А q_Б} = \frac{14250}{14200} = 1{,}0035, \text{ или } 100{,}35\%$$

где

(12.39) $$I_{А/Б} = \frac{\Sigma p_А q_А}{\Sigma p_Б q_А}$$

(12.40) $$I_{Б/А} = \frac{\Sigma p_Б q_Б}{\Sigma p_А q_Б},$$

'А/Б ~ индекс, в котором в качестве базы сравнения применяются данные по региону А; 'Б/А ~ индекс, используемый в качестве базы сравнения данных по региону Б.

Цена на продукты питания и количество

Индексы показывают, что при сравнении региона А с регионом Б цены выше в регионе А на 25,1%, а при сравнении региона Б с регионом А оказывается, что цены несколько выше в регионе Б. Таким образом, расчет индексов не позволяет определить, в каком регионе выше цены. Причина заключается в резком различии структуры продаж в отдельных регионах.

В теории и практике статистики предлагаются различные методы построения территориальных индексов, в том числе **метод стандартных весов.** Этот метод заключается в том, что значения индексируемой величины взвешиваются не по весам какого-то одного региона, а по весам области, экономического района, республики, в которых находятся сравниваемые регионы.

Таблица 12.5

проданной продукции по двум регионам

Наименование продуктов	Единица измерения	Регион А		Регион Б		Товарооборот		$Р_БЯ_Л$	$Р_ДЯ_Б$	Количество продукции, проданной в регионах А и Б ЯА+ЯБ	РА (ЯА+ЯБ)	РБ (ЯА+ЯБ)
		цена, тыс. руб. РА	количество проданной продукции Яд	цена, тыс. руб. РБ	количество проданной продукции $Я_Б$	регион А $Р_АЯ_А$	регион Б $Р_БЯ_Б$					
А	Б	1	2	3	4	5=1-2	6=3-4	7=3-2	8=1-4	9=2+4	10=1-9	11=3-9
Масло	кг	22	1200	16,5	300	26400	4950	19800	6600	1500	33 000	24750
Хлеб	кг	2	500	3	300	1 000	900	1 500	600	800	1 600	2400
Яйца	десяток	5	300	6	1400	1 500	8400	1 800	7000	1700	8 500	10200
Всего	-	-	-	-	-	28900	14250	23 100	14200	-	43 100	37350

В нашем примере в качестве весов можно использовать количество продукции, проданной в регионах А и Б, т. е.

(12.41) $$I_p = \frac{\Sigma p_А (q_А + q_Б)}{\Sigma p_Б (q_А + q_Б)}.$$

Определим значение индекса по данным табл. 12.5:

$$I_p = \frac{43100}{37350} = 1{,}154, \text{ или } 115{,}4\%.$$

Итак, цены в регионе А выше, чем цены в регионе Б, в среднем на 15,4%.

12.8

ВАЖНЕЙШИЕ ЭКОНОМИЧЕСКИЕ ИНДЕКСЫ И ИХ ВЗАИМОСВЯЗИ

Между важнейшими индексами существуют взаимосвязи, позволяющие на основе одних индексов получить другие. Зная, например, значение цепных индексов за какой-либо период времени, можно рассчитать базисные индексы. И наоборот, если известны базисные индексы, то путем деления одного из них на другой можно получить цепные индексы.

Существующие взаимосвязи между важнейшими индексами позволяют выявить влияние различных факторов на изменение изучаемого явления, например связь между индексом стоимости продукции, физического объема продукции и цен (п. 12.7). Другие индексы также связаны между собой. Так, индекс издержек производства - это произведение индекса себестоимости продукции и индекса физического объема продукции:

(12.42)

$$I_{zq} = I_z \cdot I_q$$

или

Отсюда если себестоимость увеличилась на 10%, а количество продукции снизилось на 8%, то индекс издержек на производство будет равен:

$$1{,}10 \cdot 0{,}92 = 1{,}012, \text{ или } 101{,}2\%.$$

Индекс затрат времени на производство продукции может быть получен в результате умножения индекса физического объема продукции и величины, обратной величине индекса трудоемкости, т. е. индекс производительности труда

(12.44) $$I_{tq} = I_q \cdot \frac{1}{I_t}$$

или

$$\frac{\Sigma t_1 q_1}{\Sigma t_0 q_0} = \frac{\Sigma q_1 t_0}{\Sigma q_0 t_0} : \frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}.$$

(12.45)

При увеличении физического объема продукции в текущем периоде на 15% по сравнению с базисным производительность понизилась на 18%, поэтому индекс затрат времени на производство продукции будет равен:

$$1{,}15 : 0{,}82 = 1{,}402, \text{ или } 140{,}2\%.$$

Существует важная взаимосвязь между индексами физического объема продукции и индексами производительности труда.

Индекс производительности труда рассчитывается на основе следующей формулы:

(12.46) $$I_w = \frac{\Sigma q_1 p_0}{\Sigma T_1} : \frac{\Sigma q_0 p_0}{\Sigma T_0},$$

т. е. представляет собой отношение средней выработки продукции (в сопоставимых ценах) в единицу времени (или на одного занятого) в текущем и базисном периодах.

Индекс физического объема продукции равен произведению индекса производительности труда на индекс затрат рабочего времени (или численности занятых):

$$\frac{\Sigma z_1 q_1}{\Sigma z_0 q_0} = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} \cdot \frac{\Sigma q_1 z_0}{\Sigma q_0 z_0}.$$

$$\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{\Sigma T_1}{\Sigma T_0} \cdot \left(\frac{\Sigma q_1 p_0}{\Sigma T_1} : \frac{\Sigma q_0 p_0}{\Sigma T_0} \right)$$

(12.43)

(12.47)

Таким образом, если численность рабочих возросла на 12%, а производительность труда — на 7%, то индекс физического объема продукции будет равен:

$$1{,}12 \cdot 1{,}07 = 1{,}20, \text{ или } 120\%.$$

Взаимосвязь между отдельными индексами может быть использована для выявления влияния отдельных факторов, оказывающих воздействие на изучаемое явление.

Рассмотрим *пример.* Пусть имеются следующие данные (табл. 12.6).

Таблица 12.6

Динамика отгруженной продукции, численности рабочих
и производительности труда на предприятии
в мае 1995 г. (данные условные)

Для того чтобы определить влияние каждого из факторов, введем следующие обозначения. Обозначим численность рабочих через «а», а другой фактор - выработку продукции на одного рабочего - через «Ь». Тогда стоимость отгруженной продукции - это а • Ь, а индекс стоимости продукции - это произведение индекса численности занятых на индекс производительности труда:

$$\frac{a_1 b_1}{a_0 b_0} = \frac{a_1}{a_0} \cdot \frac{b_1}{b_0}. \qquad (12.48)$$

Чтобы выявить влияние каждого фактора на общее изменение стоимости отгруженной продукции, необходимо определить динамику одного из факторов, оставив значение другого неизменным. Тогда индекс (12.48) можно представить в двух вариантах, во-первых,

Наименование показателя	Единиц измерения	Значение показателя в периоде		Абсолютное изменение	Относительная величина динамики, %
		базисном	текущем		
А	Б	1	2	3=2-1	4=2:1-100
Стоимость отгруженной продукции	млн	450	610	+ 160	136
Численность рабочих	человек	700	-730	+ 30	104,3
Выработка продукции на одного рабочего	тыс. руб	642,3	835,6	+ 193,3	130,1

Из таблицы следует, что численность рабочих увеличилась на 4,3% (730 : 700 = 1,043, или 104,3%), производительность труда - на 30,1% (835,6 : 642,3 = 1,301, или 130,1%), а отгруженная продукция - на 36% (610 : 450= 1,36, или 1,043 • 1,301 = 1,36, т. е. 136%).

Объем отгруженной продукции в текущем периоде по сравнению с базисным возрос на 160 млн руб. Это изменение является результатом роста численности рабочих и выработки продукции.

$$\frac{a_1b_1}{a_0b_0}=\frac{a_1b_0}{a_0b_0}\cdot\frac{b_1a_1}{b_0a_1}, \qquad (12.49)$$

во-вторых,

$$\frac{a_1b_1}{a_0b_0}=\frac{a_1b_1}{a_0b_1}\cdot\frac{b_1a_0}{b_0a_0}. \qquad (12.50)$$

Каждый из индексов-сомножителей (12.49 - 12.50) отражает влияние одного фактора на изменение стоимости отгруженной продукции при неизменном (базисном или отчетном) уровне другого фактора. Только в первом случае изменение выработки продукции (Ь) определяется при численности занятых в текущем периоде, а во втором случае - на уровне базисного периода. И наоборот, изменение численности занятых (а) в первом случае определяется при базисном значении выработки продукции, а во втором случае - при текущем его значении.

Рассчитаем влияние каждого из факторов по формуле (12.49):

$$\frac{610}{450}=\frac{730\cdot 642{,}3}{700\cdot 642{,}3}\cdot\frac{835{,}6\cdot 730}{642{,}3\cdot 730}=\frac{469}{450}\cdot\frac{610}{469}.$$

Следовательно,

1. Общий прирост продукции равен: 160 млн руб.$[a_1b_1 - a_0b_0 =$ $= 610 - 450]$.

2. За счет увеличения численности рабочих стоимость продукции возросла на 19 млн руб. $[a_1b_0 - a_0b_0 = 469 - 450]$.

3. За счет повышения производительности труда стоимость продукции увеличилась на 141 млн руб. [a,Ь| - a,Ь$_0$ = 610-469].

Теперь определим влияние каждого из факторов по формуле (12.50):

$$\frac{610}{450}=\frac{730\cdot 835{,}6}{700\cdot 835{,}6}\cdot\frac{700\cdot 835{,}6}{700\cdot 642{,}3}=\frac{610}{585}\cdot\frac{585}{450}.$$

Таким образом:

1. Общий прирост продукции равен 160 млн руб. *[a$_1$Ъ$_1$ - a\$Ъ$_0$* = = 610 - 450].

2. За счет увеличения численности рабочих стоимость продукции возросла на 25 млн руб.$[a_1b_1 - a_0b_1 = 610 - 585]$.

3. За счет повышения производительности труда стоимость продукции увеличилась на 135 млн руб. [a$_0$Ь| — a$_0$Ь$_0$ = 610 - 450].

Итак, мы получили различные результаты. Разницу в значениях отклонений, равную в нашем примере 6 млн руб. (141 - 135, или 25 - 19), называют изменением стоимости отгруженной продукции за счет совместного изменения обоих факторов, которая присоединяется то к одному, то к другому фактору. В экономических расчетах чаще применяется первый способ анализа влияния факторов (12.49), так как в этом случае индекс качественного признака определяется при неизменном значении количественного признака, зафиксированном на уровне текущего периода.

мы национальных счетов (совокупного общественного продукта, национального дохода, капитальных вложений и т. д.) из фактически действовавших (текущих) цен в сопоставимые.

Таким образом, *индексы цен необходимы для решения двух задач:*

- отражения динамики инфляционных процессов в народном хозяйстве страны; .
- пересчета важнейших стоимостных показателей СНС из фактических цен в сопоставимые при изучении динамики социально-экономических явлений.

Для реализации этих различных по содержанию задач служат два типа индексов:

- собственно индекс цен;
- индекс-дефлятор.

Рассмотрим подробнее первый индекс.

Первая формула для расчета индекса цен была сформулирована в 1738 г. французским экономистом Дюто, предложившим вычислять обобщенный" показатель изменения цен как отношение сумм цен на отдельные виды товаров в отчетном периоде к сумме цен на те же товары в базисном периоде. Эта формула имеет следующий вид:

$$I_p = \frac{\Sigma p_1}{\Sigma p_0}. \quad (12.51)$$

12.9

СВОЙСТВА ИНДЕКСОВ ЛАСПЕЙРЕСА И ПААШЕ

В рыночном хозяйстве особое место среди индексов качественных показателей отводится индексам цен. Основным назначением **индекса цен** является оценка динамики цен на товары производственного и непроизводственного потребления. Помимо этого индекс цен выполняет роль общего измерителя инфляции при макроэкономических исследованиях; используется при корректировке законодательно устанавливаемого минимального размера оплаты труда, установлении ставок налогов.

Индексы цен нужны при разработке технико-экономических обоснований и проектов строительства новых предприятий. Без них нельзя обойтись при пересчете основных показателей систе-

В 1764 г. итальянец Карли предложил определять общий индекс цен как простую среднюю арифметическую величину из индивидуальных индексов цен:

$$I_p = \frac{\Sigma \frac{p_1}{p_0}}{n} = \frac{\Sigma i_p}{n}. \quad (12.52)$$

И только в конце XIX в. были построены две формулы индекса цен, которые используются в качестве основных современной отечественной и зарубежной статистикой.

Автором первой формулы (12.15) является немецкий статистик Г. Пааше. Немецкий ученый Э. Ласпейрес предложил определять индекс цен следующим образом:

$$I_p = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}. \quad (12.53)$$

Индексируемой величиной обоих индексов являются цены. Весами же в индексе цен Пааше выступает количество продукции текущего периода, а в индексе цен Ласпейреса — количество продукции базисного периода.

Найдем значение индекса цен Ласпейреса по данным табл. 12.1. Он равен:

$$I_p = \frac{18480}{17504} = 1{,}056, \text{ или } 105{,}6\%.$$

Следовательно, в среднем по всем товарам цены возросли на 5,6%. В результате роста цен стоимость товаров базисного периода увеличилась на 976 тыс. руб. (18 480 - 17 504).

Значения индексов цен Пааше и Ласпейреса не совпадают. Отличие значений объясняется тем, что индексы имеют различное экономическое содержание.

Индекс цен, исчисленный по формуле Пааше, дает ответ на вопрос, насколько товары в текущем периоде стали дороже (дешевле), чем в базисном. Индекс цен Ласпейреса показывает, во сколько бы раз товары базисного периода подорожали (подешевели) из-за изменения цен на них в отчетный период.

Согласно практике индекс цен, рассчитанный по формуле Пааше, имеет тенденцию некоторого занижения, а по формуле Ласпейреса - завышения темпов инфляции.

До начала 90-х гг. XX в. отечественная статистика отдавала предпочтение индексу цен Пааше. Сложность его расчета заключается в том, что взвешивание по весам отчетного периода требует ежегодного (ежеквартального,'ежемесячного) сбора и обработки значительных объемов информации для формирования системы весов. А эта работа связана с большими затратами времени, материальных и трудовых ресурсов. Поэтому начиная с 1991 г. отечественные органы государственной статистики определяют индексы цен по формуле Ласпейреса, которой отдается предпочтение и в зарубежной статистике: Англии, Германии, США и др. При исчислении индекса цен по формуле Ласпейреса веса фиксируются на уровне базисного периода и остаются неизменными в течение некоторого промежутка времени, отсюда целью расчета индекса является измерение динамики стоимости базисного (неизменного) объема продукции.

Следует отметить, что индекс цен всегда имеет определенную степень условности. Это связано прежде всего с тем, что при его расчете учитываются изменения цен не по всей совокупности продукции, а по отдельным товарам-представителям, которые составляют так называемую товарную корзину. По мере отдаления от базисного года эта товарная корзина по видам, количеству и качеству вошедших в нее товаров-представителей все менее соответствует структуре и составу объема продукции текущего года. Поэтому состав товарной корзины, а следовательно, и система весов должны периодически пересматриваться. Только тогда они отражают современную структуру объема продукции. Особенно это важно в период резкого изменения экономических условий в народном хозяйстве страны.

При расчете индекса цен по формуле Ласпейреса необходимо решить три вопроса:

1) выбор базисного года для постоянных весов;
2) определение срока использования весовых коэффициентов без их пересмотра;
3) увязка индекса, рассчитанного по новым весам (после их пересмотра), с ранее существующими динамическими рядами индексов цен.

Например, при выборе базисного года немецкие статистики ориентируются на следующие критерии.

1. Базисный год должен находиться в середине длительной фазы подъема (снижения) экономического развитии.
2. Динамика цен в базисном году не должна быть ниже, чем в соседние с ним годы, но не должна быть и стабильной.
3. Год должен быть сравнительно «нормальным» для сельскохозяйственного производства, т. е. не выделяться по погодным условиям среди других лет.

В странах Европейского союза принято пересматривать весовые коэффициенты по истечении пятилетнего срока. В последнее время в связи с существенными колебаниями цен, что ведет к изменению структуры объема продукции, во многих странах системы весов пересматривают ежегодно, т. е. используют подвижную систему весов. Эта система остается неизменной в течение года, и по ней ежемесячно или ежеквартально рассчитываются индексы цен. Когда наступает новый календарный год, веса корректируют. Именно этот подход используется в настоящее время при исчислении индексов цен органами российской государственной статистики.

Увязка индекса, рассчитанного по новым, измененным весам (после их пересмотра), осуществляется с помощью процедуры смыкания динамических рядов (гл. 10).

Одним из важнейших показателей статистики цен, широко используемым в экономической и социальной политике государства, является индекс потребительских цен (ИПЦ). Он применяется для пересмотра правительственных социальных программ, служит основой для повышения минимального размера заработной платы, отражает реальную покупательную способность денег, которыми различные слои населения располагают для удовлетворения своих материальных, культурных и духовных потребностей. *Методология расчета этого показателя включает:*

1) отбор товаров (услуг)-представителей и торговых предприятий, по которым производится регистрация цен. Для вычисления ежемесячного ИПЦ отбор товаров (услуг)-представителей производится в соответствии с Общероссийским классификатором экономической деятельности, продукции, услуг и вновь разработанным классификатором на платные услуги населению.

От него несколько отличается набор товаров (услуг)-представителей, используемый для исчисления еженедельного индекса цен, целью которого является наиболее полное отражение изменения цен в условиях нестабильности потребительского рынка;

2) формирование структуры весов по отдельным группам товаров и услуг для расчета сводного индекса потребительских цен.

Для этого используются данные о структуре потребительских расходов населения.

Методология исчисления ИПЦ предполагает расчет индекса для отдельных регионов, товарных групп и услуг, отдельных групп населения е различным уровнем доходов, а также федерального индекса цен.

Если подходить к классификации индексов с чисто математических (формальных) позиций, то *все индексы* (не только индексы цен) *можно разделить на две группы:*

- индексы, при исчислении которых использовались веса базисного периода (формула Ласпейреса);
- индексы, рассчитанные по весам отчетного периода (формула Пааше).

В табл. 12.7 приведены варианты определения агрегатных индексов физического объема и цен.

Эти индексы, а также индекс стоимости (12.13) находятся в определенных соотношениях, которые порой бывает полезно применять в индексных расчетах.

Рассмотрим более подробно **свойства индексов Ласпейреса и Пааше.** Для удобства изложения введем следующие обозначения:

$1_р^п$, $I_q^n{}_л$ - индекс цен, физического объема с текущими весами 1^>ждекс Пааше);

$1_р^л$, $1ч^Л$ - индекс цен, физического объема с базисными весами (индекс Ласпейреса).

Таблица 12.7

Индекс Ласпейреса и Пааше

Наименование индекса	Формула индекса	
	Ласпейреса (индекс с базисными весами)	Пааше (индекс с отчетными весами)
Индекс физического объема	£qiPo £ЧоРо	SqiPi £qoPi
Индекс цен	\ £P]ЧО ^PoЯо	£P|Ч| £p0q>

Свойство 1:

$$I_p^П = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} = \frac{I_{pq}}{I_q^Л} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0} : \frac{\Sigma q_1p_0}{\Sigma q_0p_0}, \tag{12.54}$$

т. е. индекс цен в формуле Пааше равен отношению индекса стоимости продукции к индексу физического объема в формуле Ласпейреса. *Свойство 2:*

$$I^Л{}_p \cdot I^П{}_q = I^П{}_p \cdot I^Л{}_q = I_{pq} \tag{12.55}$$

или

$$\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \cdot \frac{\Sigma q_1p_1}{\Sigma q_0p_1} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \cdot \frac{\Sigma q_1p_0}{\Sigma q_0p_0} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}. \tag{12.56}$$

Данное свойство позволяет сократить объем вычислительной работы. Действительно, для определения индекса цен и физического объема необходимо иметь две величины условной стоимости: Zpiq₀ и ХроЧ|. Если рассчитать величину базисного объема товаров в текущих ценах (Piq₀), то можно исчислить сначала индекс цен по формуле Ласпейреса ($1^л{}_р$), а затем, разделив этот индекс на индекс стоимости (1™), получить индекс физического объема по формуле Пааше ($1^п{}_р$).

Свойство 3:

$$I_p^Л = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} = \frac{\Sigma p_1q_0 \cdot (p_0/p_0)}{\Sigma p_0q_0} = \frac{\Sigma p_1q_1 \cdot (p_1/p_0)}{\Sigma p_0q_0} \approx \frac{\Sigma p_0q_0 \cdot (i_p)}{\Sigma p_0q_0}. \tag{12.57}$$

Если имеются индивидуальные индексы цен, то индекс цен по формуле Ласпейреса может быть исчислен как средняя арифметическая величина, где в качестве весов используется стоимость продукции базисного периода (p₀Чо)- Именно этот способ определения индекса цен наиболее часто используется на практике зарубежными статистиками.

Свойство 4:

$$I_q^Л = \frac{\Sigma q_1p_0}{\Sigma q_0p_0} = \frac{\Sigma q_1p_0(q_0/q_0)}{\Sigma q_0p_0} = \frac{\Sigma q_0p_0(q_1/q_0)}{\Sigma q_0p_0} = \frac{\Sigma q_0p_0(i_q)}{\Sigma q_0p_0}. \tag{12.58}$$

Индекс физического объема по формуле Ласпейреса - это средняя арифметическая величина из индивидуальных индексов объема (i), взвешенных по стоимости базисного периода (p_0Чо)-

Свойство 5:

$$I_p^П = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{\Sigma p_1 q_1 (p_0/p_0)}{\Sigma p_0 q_1} = \frac{\Sigma p_0 q_1 (p_1/p_0)}{\Sigma p_0 q_1} = \frac{\Sigma p_0 q_1 (i_p)}{\Sigma p_0 q_1}. \qquad (12.59)$$

Так как коэффициенты вариации всегда положительны, а величина коэффициента корреляции между изменениями цен и физического объема на товарном рынке обычно отрицательна, то значение индекса по формуле Пааше всегда меньше значения индекса по формуле Ласпейреса.

12.10

ИДЕАЛЬНЫЙ ИНДЕКС ФИШЕРА

Индекс цен американского экономиста И. Фишера представляет собой среднюю геометрическую из произведения двух агрегатных индексов цен Ласпейреса и Пааше:

$$(12.62) \quad I_p = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}.$$

В данном случае в качестве весов используется условная стоимость - ($p_0$qi).

Свойство 6:

$$I_q^П = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} = \frac{\Sigma q_1 p_1 (q_0/q_0)}{\Sigma q_0 p_1} = \frac{\Sigma q_0 p_1 (q_1/q_0)}{\Sigma q_0 p_1} = \frac{\Sigma q_0 p_1 (i_q)}{\Sigma q_0 p_1}. \qquad (12.60)$$

Весами служит стоимость продукции базисного периода, исчисленная в ценах отчетного периода (P|$Я_0$)-

При взвешивании индекса по величине стоимости базисного периода (p_0q_0) возникает постоянная погрешность, причиной которой является тот факт, что цена входит как сомножитель в веса и между изменениями цен весов существует корреляция:

$$I_p^П : I_p^Л = 1 + r_{ip,iq} \cdot V_{ip} \cdot V_{iq}, \qquad (12.61)$$

где $r_{ip,\ iq}$ - коэффициент корреляции между индивидуальными индексами физического объема и цен на отдельные виды продукции;

V_{in} - коэффициент вариации индивидуальных индексов физического объема продукции;

V_{iD} - коэффициент вариации индивидуальных индексов цен.

Формула, предложенная Фишером, может быть использована и для определения индекса физического объема:

$$(12.63) \quad I_q = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \cdot \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}.$$

Геометрическая форма индексов имеет принципиальный недостаток: она лишена конкретного экономического содержания. Так, в отличие от агрегатного индекса Ласпейреса или Пааше разность м*ежду числителем и знаменателем не покажет никакой реальной экономии (или потерь) из-за изменения цен или физического объема продукции.

И Фишер назвал эту формулу расчета индекса идеальной формулой. Идеальность формулы заключается прежде всего в TIM, что индекс является обратимым во времени, т. е. при перестановке базисного и отчетного периодов полученный «обратный» индекс - это обратная величина величины первоначального индекса. Этому условию отвечает любой индивидуальный индекс. Например, индекс цен равен:

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

тогда обратный индекс цен определяется следующим образом:

$$\frac{1}{i_p} = \frac{p_0}{p_1}.$$

Если перемножить эти два индекса, то получится 1:

$$\frac{p_1}{p_0} \cdot \frac{p_0}{p_1} = 1.$$

Этому условию удовлетворяет идеальный индекс цен Фишера:

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \cdot \sqrt{\frac{\Sigma p_0 q_0}{\Sigma p_1 q_0} \cdot \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1}} = 1. \quad (12.64)$$

Индекс Фишера в силу сложности расчета и трудности экономической интерпретации на практике используется довольно редко. Чаще всего он применяется при исчислении индексов цен за длительный период времени для сглаживания тенденций в структуре и составе объема продукции, в которых происходят значительные изменения.

Индекс-дефлятор рассчитывается как отношение фактической стоимости продукции отчетного периода к стоимости объема продукции, структура которого аналогична структуре отчетного года, но определенного в ценах базисного года. В основе расчета индекса-дефлятора лежит формула Пааше - агрегатная формула индекса с текущими весами. Индекс-дефлятор для ВВП в 1992 г. определяется по формуле:

$$(12.65) \quad I_d = \frac{\Sigma p_{1992} \cdot q_{1992}}{\Sigma p_0 \cdot q_{1992}},$$

где

I_d - индекс-дефлятор;

4i992 ~ объем продукции в 1992 г.; Pi992' Po ~ цены фактически действовавшие в 1992 г. и базисном году соответственно.

Реальный ВВП за 1992 г. определяется по формуле:

$$(12.66) \quad R_{1992} = Q_{1992} : I_d,$$

гдеQ_{1992}- номинальный ВВП.

12.11

ИНДЕКСЫ-ДЕФЛЯТОРЫ

Пересчет важнейших стоимостных показателей системы национальных счетов (национальный доход, валовой национальный продукт и т. д.) из фактических цен в сопоставимые осуществляется с помощью индекса-дефлятора. **Дефлятор --** это коэффициент, переводящий значение стоимостного показателя за отчетны^ период в стоимостные гдеq_{1993}

p_{1993}

измерители базисного. Например, индекс-дефлятор валового внутреннего продукта (ВВП) представляет собой индекс цен, применяемый для корректировки номинального объема ВВП с учетом инфляции и получения на этой основе реального его объема*.

* Согласно экономической теории, с помощью номинального ВВП измеряется объем продукции текущего года (q|) в текущих ценах (p,). С помощью реального ВВП измеряется объем продукции текущего года (Я]) в ценах, которые сложились в базисном году (p_0).

Формулу (12.65) можно представить и в следующем виде:

$$I_d = \frac{Q_{1992}}{R_{1992}}. \quad (12.67)$$

Индекс-дефлятор для 1993 г. может быть исчислен так:

$$I_d = \frac{\Sigma p_{1993} \cdot q_{1993}}{\Sigma p_0 \cdot q_{1993}}, \quad (12.68)$$

объем продукции в 1993 г.;

цены, фактически действовавшие в 1993 г.

Сравним формулы (12.65) и (12.68). Легко заметить, что в них используются различные веса (Я|992 и Ч1993)- Поэтому важной особенностью индекса-дефлятора является то, что он не может быть использован для сравнительной оценки динамики цен за два периода, в данном случае за 1992 и 1993 гг. Индексы-дефляторы дают представление только об отношении стоимости продукции в текущем периоде к ее стоимости в базисном периоде. •

При этом не учитывается отличие состава и структуры продукции в базисный период по сравнению с отчетным.

Таким образом, индекс-дефлятор - это самостоятельный показатель.

В статистической практике индексы-дефляторы определяются не только в целом по народному хозяйству; они исчисляются по отдельным регионам, различным товарным группам, каналам реализации потребительских благ, отраслям экономики и т. д.

Индекс структурных сдвигов - индекс, характеризующий влияние изменения структуры изучаемого явления на динамику среднего уровня этого явления.

Территориальные индексы - индексы, которые отражают изменение явления во времени.

Индекс-дефлятор - отношение фактической стоимости продукции отчетного периода к стоимости объема продукции, структура которого аналогична структуре отчетного года, но определенного в ценах базисного года.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИИ ГЛАВЫ 12

Индекс - относительный показатель, который выражает соотношение величин какого-либо явления во времени, в пространстве или сравнение фактических данных с любым эталоном (план, прогноз, норматив и т. д.).

Индивидуальные индексы - относительные показатели, которые отражают результат сравнения однотоварных явлений.

Сводный, или общий, индекс — показатель, измеряющий динамику сложного явления, составные части которого непосредственно несоизмеримы.

Агрегатный индекс - сложный относительный показатель, который характеризует среднее изменение социально-экономического явления, состоящего из несоизмеримых элементов,

Индексируемая величина - признак, изменения которого изучаются.

Вес индекса - величина, служащая для целей соизмерения индексируемых величин.

сто.^Рпч

Средний индекс - индекс, вычисленный как средняя величина из индивидуальных индексов.

Система индексов - ряд последовательно построенных индексов.

Система базисных индексов - ряд последовательно вычисленных индексов одного и того же явления с постоянной базой сравнения.

Система цепных индексов - ряд индексов одного и того же явления, вычисленных с меняющейся от индекса к индексу базой сравнения.

Система индексов с постоянными весами - система сводных индексов одного и того же явления, вычисленных с весами, не меняющимися при переходе от одного индекса к другому.

Система индексов с переменными весами - система сводных индексов одного итого же явления, вычисленных с весами, последовательно меняющимися от одного индекса к другому.

Индекс переменного состава - индекс, выражающий отношение средних уровней изучаемого явления, относящихся к разным периодам времени.

Индекс постоянного (фиксированного) состава - индекс, исчисленный с весами, зафиксированными на уровне одного какого-либо периода, и показывающий изменение только индексируемой величины.

ТЕСТ К ГЛАВЕ 12 1. Индекс стоимости продукции исчисляется по формуле:

£Р|Я| . _ Spiqi . Iq,p,

а) $\frac{}{\Sigma p_0 q_1}$, б) $\frac{}{\Sigma p_0 q_0}$, в) $\frac{}{\Sigma q_0 p_1}$.

а) индекс переменного состава;
б) индекс постоянного состава;
в) индекс структурных сдвигов.

2. $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$; $\frac{\Sigma p_2 q_2}{\Sigma p_1 q_1}$; ...; $\frac{\Sigma p_n q_n}{-}$

а) цепная;
б) базисная.

3. $\frac{\Sigma z_0 q_1}{\Sigma z_0 q_0} : \frac{\Sigma q_1}{\Sigma q_0} -$

4. Индекс цен Ласпейреса определяется по формуле:

а) $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$; б) $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$; в) $\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$.

5. Индекс количества продукции, произведенной в единицу времени, рассчитывается по формуле:

а) $\frac{q_1}{T_1} : \frac{q_0}{T_0}$; б) $\frac{t_0}{t_1}$; в) $\frac{q_1 p}{T_1} : \frac{q_0 p}{T_0}$; г) $\frac{t_1 q_1}{t_0 q_0}$.

6. Индекс Струмилина рассчитывается:

а) как средний арифметический индекс;
б) как средний гармонический индекс;
в) как средний геометрический индекс.

7. Система базисных индексов физического объема продукции с постоянными весами имеет следующий вид:

ГЛАВА 13

ОБЩИЕ ВОПРОСЫ АНАЛИЗА И ОБОБЩЕНИЯ СТАТИСТИЧЕСКИХ ДАННЫХ

а) $\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}$; $\frac{\Sigma q_2 p_0}{\Sigma q_1 p_0}$; ...; $\frac{\Sigma q_n p_0}{\Sigma q_{n-1} p_0}$;

б) $\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}$; $\frac{\Sigma q_2 p_0}{\Sigma q_0 p_0}$; ...; $\frac{\Sigma q_n p_0}{\Sigma q_0 p_0}$;

в) $\frac{\Sigma q_1 p_1}{\Sigma q_0 p_0}$; $\frac{\Sigma q_2 p_2}{\Sigma q_0 p_0}$; ...; $\frac{\Sigma q_n p_n}{\Sigma q_0 p_0}$.

8. Если себестоимость увеличилась на 14%, а количество продукции снизилось на 6%, то индекс издержек производства будет равен:

а) 107;
б) 120;
в) 121.

9. Индекс-дефлятор - это индекс:

а) из системы цепных индексов цен с переменными весами;
б) из системы цепных индексов с постоянными весами;
в) из системы базисных индексов с переменными весами;
г) из системы базисных индексов с постоянными весами.

10.Если индекс переменного состава равен 118%, а индекс структурных сдвигов 107%, то индекс фиксированного состава равен:

а) ПО; **6)111; в)** 115.

13.1

ПОНЯТИЕ И ОСНОВНЫЕ ПРИНЦИПЫ ЭКОНОМИКО-СТАТИСТИЧЕСКОГО АНАЛИЗА

Анализ и обобщение статистических данных - заключительный этап статистического исследования, конечной целью которого является получение теоретических выводов и практических заключений о тенденциях и закономерностях изучаемых социально-экономических явлений и процессов.

Анализ - это метод научного исследования объекта путем рассмотрения его отдельных сторон и составных частей.

Экономико-статистический анализ - это разработка методики, основанной на широком применении традиционных статистических и математико-статистических методов, с целью контроля адекватного отражения исследуемых явлений и процессов.

Задачами статистического анализа являются: определение и оценка специфики и особенностей изучаемых явлений и процессов, изучение их структуры, взаимосвязей и закономерностей их развития.

В качестве этапов статистического анализа выделяются:

1) формулировка цели анализа;

2) критическая оценка данных;

3) сравнительная оценка и обеспечение сопоставимости данных;

4) формирование обобщающих показателей;

5) фиксация и обоснование существенных свойств, особенностей, сходств и различий, связей и закономерностей изучаемых явлений и процессов;

6) формулировка заключений, выводов и практических предложений о резервах и перспективах развития изучаемого явления.

Методы анализа должны меняться в зависимости от характера изучаемых процессов, их специфики, особенностей и форм проявления.

Статистический анализ данных проводится в неразрывной связи теоретического, качественного анализа сущности исследуемых явлений и соответствующего количественного инструментария изучения их структуры, связей и динамики.

Экономико-статистический анализ должен проводиться при строгом соблюдении следующих принципов, которые учитывают экономическую и статистическую их градацию.

Экономическими принципами являются:

- соответствие экономическим законам и положениям теории расширенного воспроизводства;
- адекватное отражение сущности экономической политики современного этапа общественно-экономического развития;
- ориентация на конечные экономические результаты;
- учет специфики изучаемого объекта, отрасли и т. д.;
- согласование интересов субъектов различных иерархических уровней как подразделений единого народнохозяйственного механизма.

К статистическим принципам относятся:

- четко определенная цель экономико-статистического исследования;
- согласованность систем по горизонтали и вертикали;
- сопоставимость во времени и пространстве;
- логическая взаимосвязь между показателями, характеризующими объект или явление;
- комплексность и полнота отображения объекта исследования в статистических показателях;
- максимальная степень аналитичности.

Соблюдение данных принципов наряду с предпосылками применения методологии статистического анализа позволяет осуществить научно обоснованное экономико-статистическое исследование субъектов экономики в соответствии с принятой международной методологией учета и статистики.

13.2

АПРИОРНЫЙ АНАЛИЗ И ЕГО РОЛЬ В ИССЛЕДОВАНИИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

Оценка эффективности и деловой активности субъектов экономического процесса и состояния социальной инфраструктуры общества во многом зависит от качества статистического анализа эмпирического материала, а также от того, насколько точно будут выявлены и научно обоснованы закономерности и тенденции развития. Основные трудности, связанные с применением количественных математико-статистических методов, заключаются в том, что они достаточно нейтральны к исследуемым социально-экономическим процессам. Поэтому важным этапом проведения статистического исследования на информационной базе, характеризующей реальные социально-экономические явления, является критическая оценка исходных данных с точки зрения их достоверности и научной обоснованности.

Под критической оценкой статистического материала следует понимать полноту, качество и достоверность его соответствия целям и задачам исследования.

Надежность выводов и заключений по анализу статистических данных обеспечивается минимизацией в исходной информации пробелов, неточностей, несопоставимости, неопределенности и т. д.

Во время информационного бума, которым характеризуется со- , временный этап общественного развития, необходимо больше внимания уделять критической оценке и априорному анализу исходной статистической информации. Развитие новых организационно-правовых форм во всех сферах общественной жизни, наличие коммерческой тайны и так далее увеличивают вероятность получения преднамеренно искаженных фактов, затушевывающих результаты производственно-хозяйственной деятельности фирм, банков и других структур.

Методы априорного анализа включают:

- выявление экономически обоснованных и существенных причинно-следственных связей между признаками и явлениями;
- оценку однородности исследуемой совокупности;
- анализ характера распределения совокупности по изучаемым признакам.

Понятия, используемые при проведении анализа статистическими методами, должны быть точно определены.

Например, при изучении производительности труда в строительстве необходимо установить, будет ли выработка рассчитываться на одного работника или одного рабочего занятого в строительном производстве. Необходимо также четко определить, к какому моменту или периоду времени относится исследуемое явление или процесс.

Одной из основополагающих предпосылок проведения научно обоснованного статистического анализа, адекватно отражающего причинно-следственные связи и зависимости, тенденции развития реальных явлений и процессов в статике и динамике, является однородность статистической совокупности (гл. 3).

Если данное условие выполняется, то значение у/^{1}' является скорректированным, не аномальным значением, занимает место УцпВ ряду и анализу подвергается $у_{i(2)}$.

Если условие не выполняется, то рекомендуется рассчитать $у_{l(2}$) и проверить на аномальность. Процесс корректировки носит итерационный характер.

При исследовании динамики наибольшее распространение получил метод Ирвина, основанный на определении А,$_{s}$ - статистики. При его использовании выявление аномальных наблюдений производится по схеме

$$\lambda_i = \frac{|y_i - y_{i-1}|}{\sigma_y}. \quad (13.6)$$

Если расчетное значение превысит уровень критического (с заданным уровнем точности и числом наблюдений) (табл. 13.1), то оно признается аномальным.

Таблица 13.1 Табулированные значения А.

Число наблюдений	Ч	
	0,95	0,99
2	2,8	3,7
o J	2,2	2,9 •
10	1,5	2,0
20	1,3	1,8
30	1,3	1,7
50	1,1	1,6
100	1,0	1,5

Схема реализации данного метода аналогична предыдущей с той лишь разницей, что у заменяется на $у_{и}$ (предыдущее значение ряда).

Способ, основанный на расчете q-статистики, применим для относительно стационарных рядов, так как при использовании для анализа динамических рядов, имеющих ярко выраженную тенденцию, он приводит к ошибкам. Более корректным является

применение статистики, в которой определяются отклонения от теоретических значений, полученных по уравнению тренда:

$$(13.7)\quad q_i = \frac{|y_t - \bar{y}_t|}{\sigma_y}.$$

В общем виде, систематизируя методику анализа аномальных наблюдений, описанную в статистической литературе, градацию статистических методов выявления аномальности в исходных данных можно представить следующей схемой (рис. 13.1).

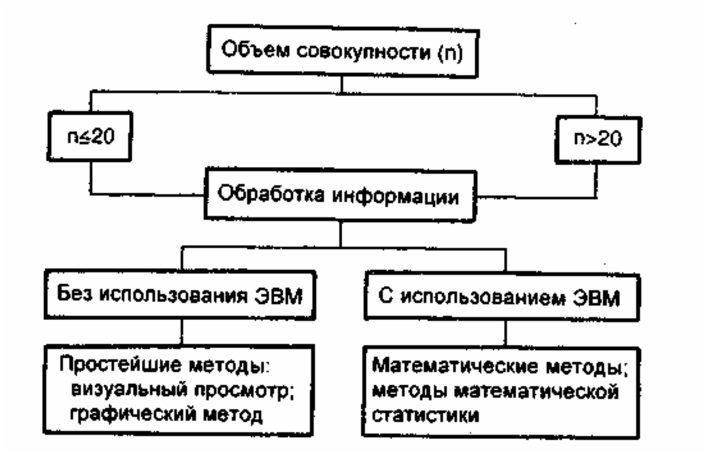


Рис. 13.1. Методы выявления аномальных наблюдений

Нецелесообразность исключения аномальных наблюдений из изучаемой совокупности реализуется широким использованием метода группировок.

Важной задачей статистических исследований на этапе априорного анализа является выделение однородных групп (даже аномальных). В данном случае в анализе эффективно применять сложные комбинационные группировки с развернутым сказуемым (гл. 3).

Всесторонний качественный, анализ исходных данных является залогом проведения научно-обоснованного, логически выверенного экономико-статистического исследования социально-экономических явлений и процессов.

13.3

КОМПЛЕКСНОЕ ПРИМЕНЕНИЕ МАТЕМАТИКО-СТАТИСТИЧЕСКИХ МЕТОДОВ АНАЛИЗА ДАННЫХ

Наряду с традиционными статистическими методами анализа данных, подробно описанными в главах 3-12 данного учебника, при исследовании реальных социально-экономических явлений и процессов широко используются математико-статистические методы исходя из отечественной и зарубежной методологии.

Комплексность применения математико-статистических методов предполагает наиболее полное раскрытие сущности, закономерностей и тенденций развития конкретных явлений и процессов с целью более адекватного отражения их свойств и особенностей, резервов и перспектив развития и путей совершенствования.

Усложнение структуры социально-экономических явлений предполагает использование ряда методов классификации и выделения однородных групп, в основе построения которых лежат меры близости или метрики. Сущность заключается в том, что распределение исследуемых объектов или явлений в совокупности должно подчиняться нормальному закону распределения, с тем, чтобы получить модели, которые действительно будут отражать качественно однородные группы.

Наибольшее распространение в практике анализа экономических явлений и процессов получили: кластерный анализ, метод главных компонент, факторный анализ

Пусть имеется *n* объектов, каждый из которых характеризуется набором ^-признаков.

Требуется разбить эту совокупность на однородные группы. Полученные в результате разбиения группы называются **кластерами,** а метод их нахождения - **кластерным анализом.**

Наиболее трудным считается определение однородности объектов, которые задаются введением расстояния между объектами

Xj и Xj (р (Xj, Xj)).

Объекты будут однородными в случае р (x_{J5} Xj) < $р_{пор}$, где $р_{пор}$ - заданное пороговое значение.

Выбор расстояния (р) является основным моментом исследования, от которого зависят окончательные варианты разбиения. Наиболее распространенными считаются процедуры «ближайшего соседа», основанные на близости объектов по совокупности рассматриваемых признаков, и «дальнего соседа».

В задачах кластерного анализа часто используют **Евклидово и Хемингово расстояния.**

Евклидово расстояние:

$$\rho_E(x_i, x_j) = \sqrt{\left(x_i^{1,2,\dots,k} - x_j^{1,2,\dots,k}\right)^2};$$

сравнивается близость двух объектов по большому числу признаков.

Хемингово расстояние:

$$\rho_E(x_i, x_j) = \sum_{i,j}^{1,2,\dots,k} \left| x_i^{1,2,\dots,k} - x_j^{1,2,\dots,k} \right|;$$

используется как мера различия объектов, задаваемых атрибутивными признаками.

Выбор метрики-расстояния определяется структурой признакового пространства и целью классификации.

При использовании процедур кластерного анализа расчленение объектов совокупности на качественно однородные группы производится одновременно по большому числу признаков, но при соблюдении условия, что ни один признак не выделяется по своей значимости так, что группировка на его основе является главной. Особенностью кластерного анализа является то, что различия между единицами, входящими в выделенную группу, незначительны, а различия между группами существенны.

Пример. Провести классификацию шести объектов, каждый из которых характеризуется двумя признаками (табл. 13.2).

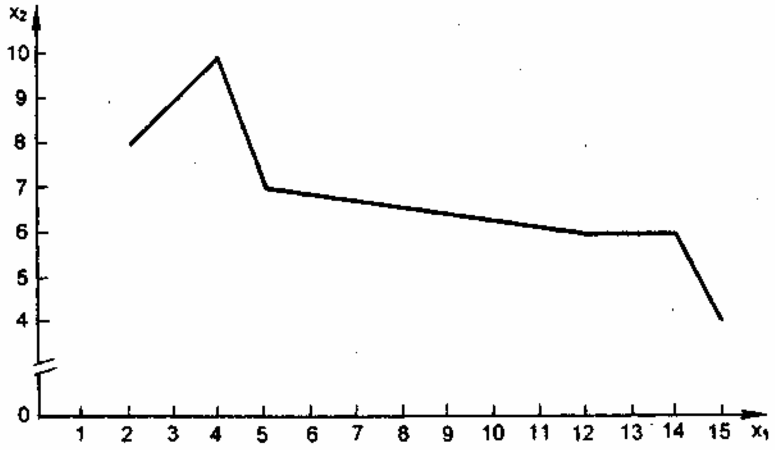
Таблица 13.2

№ п/п	1	2	3	4	5	6
^{x}i	2	4	*5*	12	14	15
X_2	8	10	*1*	6	6	4

где X, - объем выпускаемой продукции, млн руб.;

x_2 - среднегодовая стоимость основных промышленно-производ-ственных фондов, млн руб.

Представим эти данные графически (рис. 13.2)



13.2. Зависимость между объемом выпускаемой продукции и среднегодовой стоимостью основных промышленно-производственных фондов

$$\rho(x_i x_j) = \sqrt{\sum_{l=1}^{k} \left(x_{il} - x_{jl}\right)^2},$$

где l – признаки;
k – количество признаков.

$$\rho_{11} = 0;$$

$$\rho_{12} = \sqrt{(2-4)^2 + (8-10)^2} = \sqrt{8} \approx 2{,}83.$$

Расчеты последующих $\rho(x_i,x_j)$ аналогичны:

$\rho_{13} = 3{,}16$; $\rho_{14} = 10{,}19$; $\rho_{15} = 12{,}17$; $\rho_{16} = 13{,}60$;
$\rho_{23} = 3{,}16$; $\rho_{24} = 8{,}94$; $\rho_{25} = 10{,}77$; $\rho_{26} = 12{,}53$;
$\rho_{34} = 7{,}07$; $\rho_{35} = 9{,}06$; $\rho_{36} = 10{,}44$; $\rho_{45} = 2{,}00$;
$\rho_{46} = 3{,}61$; $\rho_{56} = 2{,}24$.

1. Принцип «ближайшег

$$R_1 = \begin{array}{r|ll} & 1 & 2 \\ 1 & 0 & 2{,}83 \\ 2 & & 0 \\ 3 & & \\ 4 & & \\ 5 & & \\ 6 & & \end{array}$$

$\rho_{min} = \rho_{45} = 2$ $(S_1;\ S_2;\ S_3$

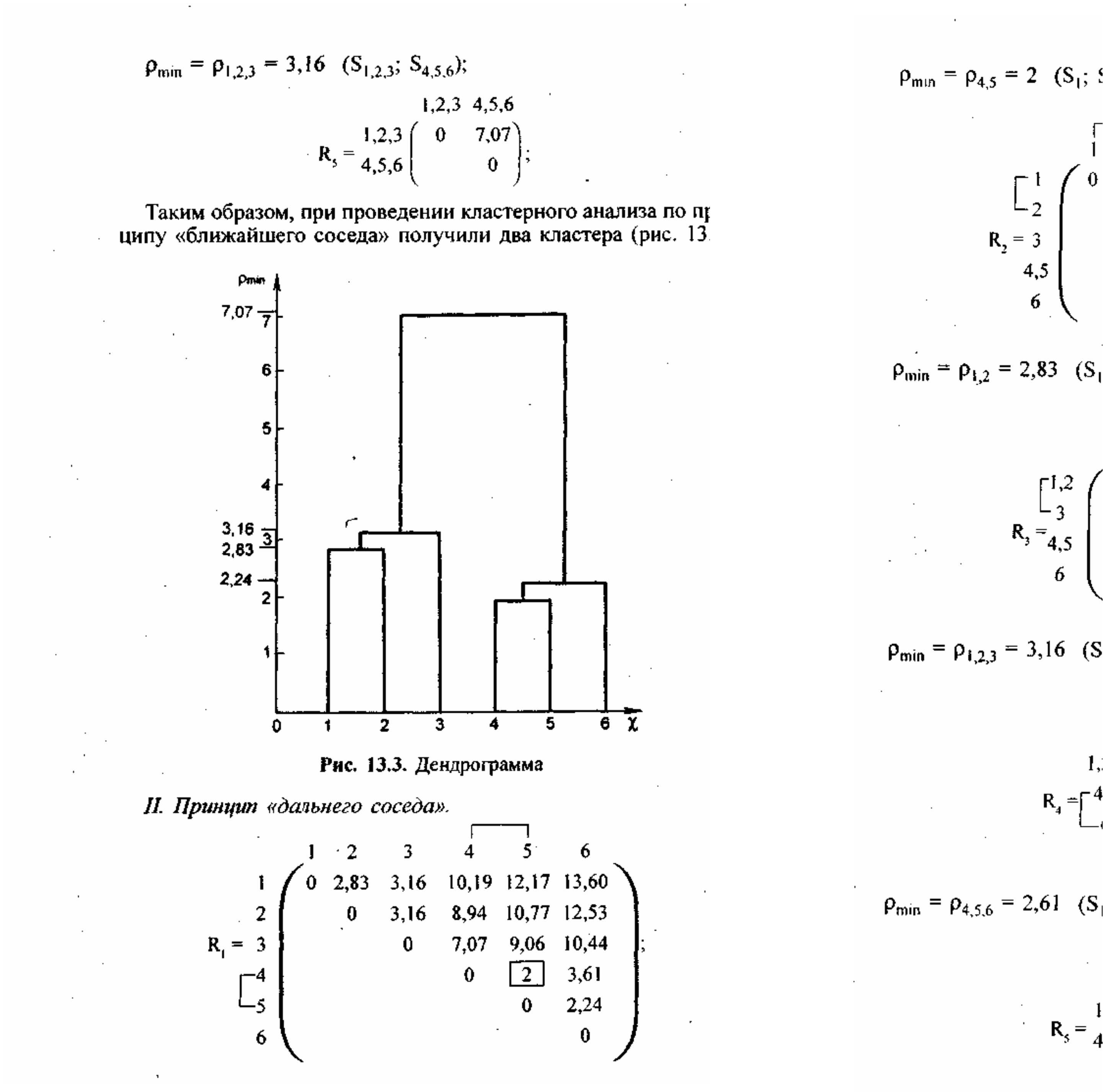
$$R_2 = \begin{array}{r|ll} & 1 & 2 \\ 1 & 0 & 2{,}83 \\ 2 & & 0 \\ 3 & & \\ 4{,}5 & & \\ 6 & & \end{array}$$

$\rho_{min} = \rho_{4,5,6} = 2{,}24$ $(S_1;\ S_2$

$$R_3 = \begin{array}{r|ll} & 1 & \\ 1 & 0 & \boxed{2} \\ 2 & & \\ 3 & & \\ 4{,}5{,}6 & & \end{array}$$

$\rho_{min} = \rho_{12} = 2{,}83$ $(S_{1,2};\ S_3;$

$$R_4 = \begin{array}{r|l} & 1, \\ 1{,}2 & 0 \\ 3 & \\ 4{,}5{,}6 & \end{array}$$



$$\rho_{\min} = \rho_{1,2,3} = 3{,}16 \quad (S_{1,2,3};\ S_{4,5,6});$$

$$R_5 = \begin{array}{c|cc} & 1,2,3 & 4,5,6 \\ \hline 1,2,3 & 0 & 7{,}07 \\ 4,5,6 & & 0 \end{array};$$

Таким образом, при проведении кластерного анализа по пр
ципу «ближайшего соседа» получили два кластера (рис. 13

Рис. 13.3. Дендрограмма

II. Принцип «дальнего соседа».

$$R_1 = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 2{,}83 & 3{,}16 & 10{,}19 & 12{,}17 & 13{,}60 \\ 2 & & 0 & 3{,}16 & 8{,}94 & 10{,}77 & 12{,}53 \\ 3 & & & 0 & 7{,}07 & 9{,}06 & 10{,}44 \\ 4 & & & & 0 & \boxed{2} & 3{,}61 \\ 5 & & & & & 0 & 2{,}24 \\ 6 & & & & & & 0 \end{array};$$

$\rho_{\min} = \rho_{4,5} = 2 \quad (S_1;$

$R_2 =$ 1, 2, 3, 4,5, 6; 0

$\rho_{\min} = \rho_{1,2} = 2{,}83 \quad (S$

$R_3 =$ 1,2, 3, 4,5, 6

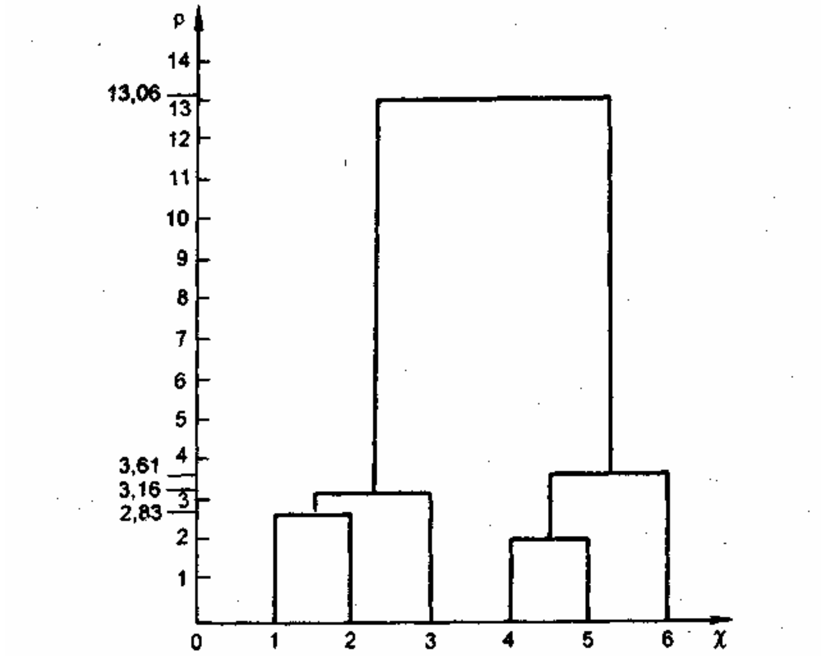
$\rho_{\min} = \rho_{1,2,3} = 3{,}16 \quad (S$

$R_4 =$ 4

$\rho_{\min} = \rho_{4,5,6} = 2{,}61 \quad (S$

$R_5 =$ 1, 4

При проведении кластерного анализа по принципу «дальнего соседа» получили два кластера (рис. 13.4).



13.4. Дендрограмма

Модели на основе результатов кластерного анализа позволяют исследовать однородные по основным экономико-техническим характеристикам и параметрам деятельности социально-экономические объекты и процессы а также степень их деловой активности.

По мере углубления анализа деятельности экономических структур - объектов и процессов - в рассмотрение включается все большее число признаков. При этом требуется обозримость. Закономерность распыляется на большое множество связей. Поэтому целесообразно осуществлять классификацию по нескольким обобщающим признакам, полученным с помощью **метода главных компонент или факторного анализа.**

Предпосылками, обусловливающими возможность применения этих методов, являются:

1) наличие сильно коррелированных признаков, приводящих к дублированию информации;
2) слабая информативность ряда факторных признаков;
3) возможность и целесообразность агрегирования несколь ких факторных признаков.

Итак, сокращение размерности методом главных компонент и факторного анализа предполагает переход от описания результативного признака меньшим числом наиболее информативных (с точки зрения их влияния на результативный), факторных признаков.

Метод главных компонент, как правило, рассматриваемый в качестве средства снижения размерности, также используется для проведения классификаций. Математической моделью на которой основывается метод главных компонент, является следующая:

Сущность метода заключается в выделении линейных комбинаций исходных факторных

$$y_j = V_{1j}F_1 + V_{2j}F_2 + \ldots + V_{pj}F_p = \sum_{p=1}^{n} V_{pj}F_p, \quad j = 1,2,\ldots,p,$$

где- весовой коэффициент общего результативного фактора

$V_{j1,2,...p}$

$'1.2....P$ - общий фактор (главная компонента). й;

признаков, имеющих максимально возможную дисперсию. При этом первая главная компонента обладает максимальной дисперсией и является нормированной линейной комбинацией всех возможных исходных признаков, а вторая - учитывает максимальное значение оставшейся дисперсии и корреляционно не связана с первой компонентой.

В целом, если переходить к оценке степени и направления связей между анализируемыми признаками на основе совместного использования методов регрессионного анализа из главных компонент, уравнение главных компонент включает меньшее их число, чем количество факторных признаков, так как из дальнейшего рассмотрения должны быть исключены главные компоненты, вклад которых в общую дисперсию несуществен и составляет менее 10%.

Комплексное использование корреляционного анализа и метода главных компонент выражается в расчете парных коэффициентов корреляции между исходными (включенными в исследование) признаками и соответствующими главными компонентами. На основе общей оценки значения парного коэффициента корреляции (по абсолютной величине) и доли вклада (в %-ном выражении) каждой компоненты в общую вариацию признака осуществляется отбор наиболее статистически существенных главных компонент.

К достоинствам метода главных компонент относятся следующие:

- в матрице значений компоненты размещаются в убывающей последовательности собственных значений, что способствует классификации признаков;
- число компонент соответствует количеству исходных фак торных признаков;
- главные компоненты не коррелированы между собой, что существенно при построении регрессионных моделей;
- главные компоненты полностью обусловливают вариацию исходных факторных признаков.

Факторный анализ заключается в переходе от исходной информации к обобщенным факторам, являющимися результатом их первоначальной агрегации и линейной комбинации.

Основная модель факторного анализа линейная и имеет вид:

$$y_j = a_{1j}F_1 + a_{2j}F_2 + ... + a_{jp}F_p + d_jV_j,$$

где- ($F_1, F_2,..., F_p$ факторы, обусловливающие систематическую вариацию и корреляционную связь между ними;
a_{ij} - факторные нагрузки;
V_j - характерные факторы, учитывающие вариацию, не объяснимую общими факторами.

Факторные нагрузки оценивают степень тесноты связи между исходными признаками $x_,$, x_2,...x_k и обобщенными факторами FJ. Связь считается существенной, если парный коэффициент корреляции больше или равен по абсолютному значению 0,5.

В практической деятельности вклад общего фактора в общую дисперсию составляет не менее 80-90%.

Совокупное использование методов факторного и регрессионного анализов невозможно без учета специфики и различий между ними. При применении регрессионного анализа в модель не могут быть включены все переменные, влияющие на результативный признак, что ведет к некоторой потере информации. На основе реализации факторного анализа в модель включаются общие факторы, являющиеся реальным отображением ряда экономически связанных между собой исходных переменных, воспроизводящих и объясняющих их свойства.

Переход от большого числа факторных признаков к обобщенным факторам или главным компонентам не служит существенной потерей информации. В отдельных случаях общие факторы могут отражать свойства исходных факторных признаков, непосредственно статистически не измеряемых и влияния на результат не оказывающих.

Модели регрессии на обобщенные факторы и главные компоненты заведомо не содержат коллинеарно связанные признаки.

Однако на практике целесообразнее строить модели регрессии на исходные переменные, так как использование общих факторов и главных компонент сильно осложняет экономическую интерпретацию параметров модели.

Комплексность методологии статистического анализа в оценке статистических и динамических совокупностей эмпирических данных осуществляется в следующей последовательности, которая может быть рассмотрена как задание для самостоятельной научно-исследовательской работы студентов.

А. Методика **комплексного анализа статической информации и выявление причинно-следственных связей.**

/. Априорный анализ исходных статистических данных. 1. Обобщение исходных данных:

а) построение интервальных вариационных рядов по каждому из рассматриваемых показателей на базе предварительного определения целесообразного количества групп;

б) графическое изображение полученных в п. 1 рядов распределения в виде гистограммы, полигона, кумуляты и огивы.

2. Оценка однородности совокупности на основе:

а) метода группировок;

б) показателей вариации;

в) анализа аномальных наблюдений на основе А,- и q-статистик.

3. Оценка характера распределения совокупности исходных данных:

а) вычисление и анализ средней, моды, медианы, показателей вари ации. ,

Вывод о характере распределения: близки ли к нормальному распределению случайных величин эмпирические распределения, полученные в виде вариационных рядов. С этой целью могут быть использованы различные модификации соотношений средних величин и показателей вариации;

б) проверка данных на базе одного из критериев: Пирсона, Ястремского, Романовского, Колмогорова.

//. Моделирование связи социально-экономических явлений. \. Отбор факторных признаков:

а) метод экспертных оценок и ранговые коэффициенты корреляции как инструмент анализа экспертной информации;

б) графический метод как способ наглядного отображения зависи мости результативного с каждым из факторных признаков;

в) метод корреляционного анализа в оценке характера парных и множественных зависимостей между исходными признаками. Расчет парных, частных и множественных коэффициентов корреляции. Иссле дование связей на мультиколлинеарность.

2. Построение модели связи и оценка ее существенности:

а) определение параметров модели методом наименьших квадратов;

б) построение уравнения связи методом пошагового регрессионного анализа;

в) проверка адекватности регрессионной модели исследуемому со циально-экономическому явлению:

- проверка значимости коэффициентов регрессии при факторных признаках, вошедших в модель, на основе t-критерия Стьюдента;
- проверка значимости уравнения регрессии на основе F-критерия Фишера - Снедекора;
- расчет и анализ средней ошибки аппроксимации (е);
- анализ средней квадратической ошибки ($a_{ош}$) и остаточной дис персии ($a^2_{ост}$).

3. Интерпретация модели связи (уравнения регрессии). С этой це лью расчет и анализ:

а) р-коэффициентов, построение модели связи в стандартизованном масштабе;

б) частных коэффициентов эластичности ($Э_x$.);

в) частных и множественного коэффициентов детерминации;

г) Д,у-коэффицентов;

д) Qj-коэффицентов.

Б. Методика комплексного анализа и прогнозирования динамической информации.

/. Анализ и прогнозирование тенденции.

1. Оценка аномальных наблюдений на основе Я,- и q-статистик.
2. Расчет аналитических (Д, T_p и $T_{пр}$) и средних показателей рядов динамики и на их основе анализ тенденций и закономерностей разви тия социально-экономических явлений.
3. Определение наличия тенденции средних и дисперсии в рядах динамики на базе методов:

- Фостера - Стюарта;
- сравнения средних уровней ряда динамики.

4. Определение наличия тенденции автокорреляции (для связных рядов динамики) на основе:

а) первого циклического коэффициента;

б) критерия Неймана;

в) критерия Дарбина - Уотсона.

5.Выявление основной тенденции динамического ряда методами:

а) укрупнения интервалов;

б) усреднения по левой и правой половине;

в) простой и взвешенной скользящей средней;

г) аналитического выравнивания и определения параметров на базе методов:

- конечных разностей;
- средних значений (линейных отклонений);
- наименьших квадратов.

6. Оценка адекватности выбранного уравнения тренда (в п. 5, г) на основе:

а) минимизации сумм квадратов отклонений эмпирических данных от теоретических (расчетных);

б) средней квадратической ошибки;

в) средней ошибки аппроксимации.

7. Корреляция рядов динамики. Модели, исключающие автокорреля цию, методами:

а) последовательных или конечных разностей;

б) отклонения эмпирических значений признаков от выравненных по тренду;

б) Фриша-Воу.

8. Прогнозирование динамики на основе простейших методов:

- среднего уровня ряда;
- среднего абсолютного прироста;
- среднего темпа роста;
- линейного тренда.

//. Выявление периодической компоненты. Модели сезонных колебаний:

а) графический анализ исходных данных;

б) выявление тенденции средней и дисперсии;

в) проверка ряда динамики на наличие сезонной компоненты на основе критерия «пиков» и «ям» и др.;

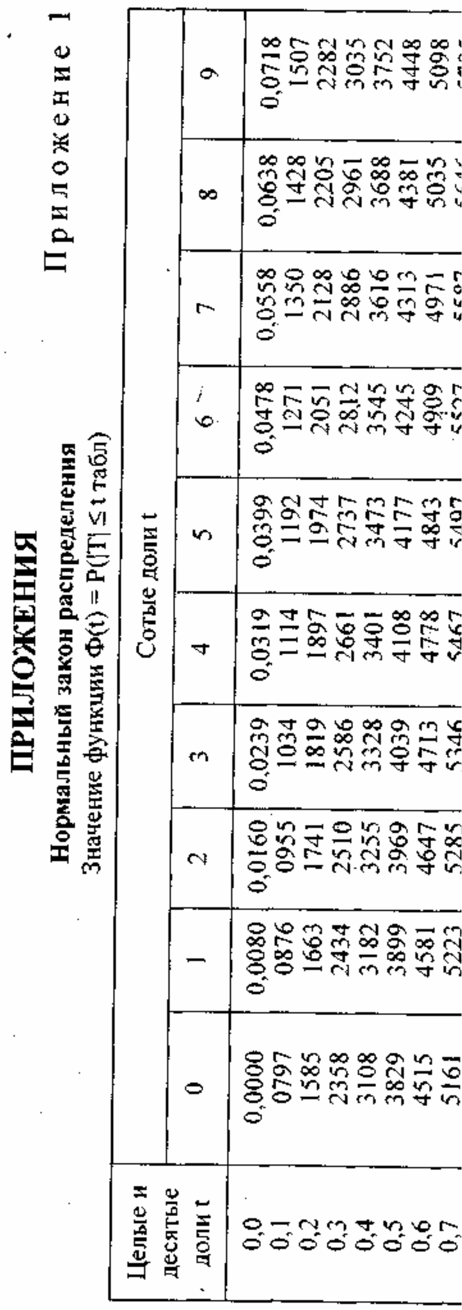
г) расчет параметров уравнения тренда и определение теоретичес ких уровней ряда динамики по тренду;

д) определение абсолютных и относительных отклонений фактичес ких уровней от выравненных по тренду. Графический метод в анализе амплитуды отклонений эмпирических и теоретических значений уров ней ряда динамики;

е) проверка абсолютных и относительных фактических уровней на наличие автокорреляции;

ж) построение модели сезонной волны по отклонениям фактических данных от тренда методами гармонического анализа. Определение гармоники Фурье, наилучшим образом отражающей периодичность изменения уровней ряда динамики на основе:

- минимизации суммы квадратов отклонений эмпирических данных от выравненных по гармонике;



П р и л о ж е н и е 1

ПРИЛОЖЕНИЯ

Нормальный закон распределения

Значение функции Ф(t) = P(|T| ≤ t табл)

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0718
0,1	0797	0876	0955	1034	1114	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3752
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4909	4971	5035	5098
0,7	5161	5223	5285	5346	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

- средней квадратической ошибки.

Изучение социально-экономических явлений и процессов на основе комплексной методики анализа, обобщения и прогнозирования на базе широкого применения традиционных статистических и математико-статистических методов позволит глубоко и досконально исследовать при-чинно-следственные связи и закономерности и показать природу изучаемого явления или процесса.

КРАТКИЙ ОБЗОР ОСНОВНЫХ ПОНЯТИЙ ГЛАВЫ 13

Заключительным этапом проведения статистического исследования является анализ статистических данных с целью получения теоретических выводов и практических заключений о тенденциях и закономерностях изучаемых социально-экономических явлений и процессов. Рассмотрим основные понятия данной главы учебника.

Анализ - метод научного исследования объекта путем рассмотрения его отдельных сторон и составных частей.

Экономико-статистический анализ - разработка методики, основанной на широком применении традиционных статистических и мате-матико-статистических методов, с целью контроля адекватного отражения исследуемых явлений и процессов.

Критическая оценка исходных данных - полнота, качество и достоверность соответствия эмпирического материала целям и задачам исследования.

Аномальные наблюдения - наблюдения, характерные для нестабильных явлений и процессов. Выявляются и корректируются с помощью методов Ирвина и q-статистики.

КОНТРОЛЬНЫЕ ВОПРОСЫ К ГЛАВЕ 13

1. *Дайте определение анализа.*
2. *Основные принципы экономико-статистического анализа.*
3. *Определение информации.*
4. *Виды информации и их краткая характеристика.*
5. *Требования, предъявляемые к статистической информационной базе.*
6. *Неполнота статистической информации и причины ее возникно вения.*
7. *Методика выявления аномальных наблюдений.*
8. *Однородность совокупности и пути ее определения.*
9. *Сущность и задачи кластерного анализа.*
10. *Сущность и задачи метода главных компонент. П. Сущность и задачи факторного анализа.*

Продолжение

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9785	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	9915	9917	9920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,5	0,999994	–	–	–	–	–	–	–	–	–
5,0	0,99999994	–	–	–	–	–	–	–	–	–

Приложение 2

Распределение Стьюдента (t-распределение)

υ	Вероятность $\alpha = S_t(t) = P(\lvert T\rvert > t_{табл})$												
	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	4,032	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959

Продолжение

υ	Вероятность $\alpha = S_t(t) = P(\lvert T\rvert > t_{табл})$												
	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
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,868	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,402	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Приложение 3

Распределение Пирсона (χ^2-распределение)

Значения $\chi^2_{табл}$ для вероятностей $P(\chi^2 > \chi^2_{табл})$

υ	Вероятность										
	0,999	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
1	0,05157	0,04393	0,03157	0,03628	0,03982	0,00393	0,0158	0,0642	0,102	0,148	0,455
2	0,00200	0,0100	0,0201	0,0404	0,0506	0,103	0,211	0,446	0,575	0,713	1,386
3	0,0243	0,0717	0,115	0,185	0,216	0,352	0,584	1,005	1,213	1,424	2,366
4	0,0908	0,207	0,297	0,429	0,484	0,711	1,064	1,649	1,923	2,195	3,357

Продолжение

υ	Вероятность										
	0,999	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
21	6,447	8,034	8,897	9,915	10,283	11,591	13,240	15,445	16,344	17,182	20,337
22	6,983	8,643	9,542	10,600	10,982	12,338	14,041	16,314	17,240	18,101	21,337
23	7,529	9,260	10,196	11,293	11,688	13,091	14,848	17,187	18,137	19,021	22,337
24	8,035	9,886	10,856	11,992	12,401	13,848	15,659	18,062	19,037	19,943	23,337
25	8,649	10,520	11,524	12,697	13,120	14,611	16,173	18,940	19,939	20,887	24,337
26	9,222	11,160	12,198	13,409	13,844	15,379	17,292	19,830	20,843	21,792	25,336
27	9,803	11,808	12,879	14,125	14,573	16,151	18,114	20,703	21,749	22,719	26,136
28	10,391	12,461	13,565	14,847	15,308	16,928	18,937	21,588	22,657	23,617	27,386
29	10,986	13,121	14,256	15,574	16,047	17,708	19,768	22,475	23,567	24,577	28,336
30	11,588	13,787	14,953	16,306	16,791	18,493	20,599	23,364	24,478	25,508	29,336

Продолжение

υ	Вероятность									
	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
1	1,074	1,323	1,642	2,706	3,841	5,024	5,412	6,635	7,879	10,827
2	2,408	2,773	3,219	4,605	5,991	7,378	7,824	9,210	10,597	13,815
3	3,665	4,108	4,642	6,251	7,815	9,348	9,837	11.345	12,838	16,268
4	4,878	5,385	5,989	7,779	9,488	11,143	11,668	13,277	14,860	18.465
5	6,064	6,626	7,289	9,236	11,070	12,839	13,388	15,086	16,750	20,517
6	7,231	7,841	8,558	10,645	12,592	14,449	15,033	16,812	18,548	22,457
7	8,383	9,037	9,803	12,017	14,067	16,013	16,622	18,475	20,278	24,322
8	9,524	10,219	11,030	13,362	15,507	17,535	18,168	20,090	21,955	26,125

Продолжение

υ	Вероятность									
	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
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

Приложение 4

Распределение Фишера – Снедекора (F-распределение)

Значения $F_{табл}$, удовлетворяющие условию $P(F > F_{табл})$. Первое значение соответствует вероятности 0,05; второе – вероятности 0,01 и третье – вероятности 0,001; υ_1 – число степеней свободы числителя; υ_2 – знаменателя

υ_2 \ υ_1	1	2	3	4	5	6	8	12	24	∞	t
1	161,4 4052 406523	199,5 4999 500016	215,7 5403 536700	224,6 5625 562527	230,2 5764 576449	234,0 5859 585953	238,9 5981 598149	243,9 6106 610598	249,0 6234 623432	253,3 6366 636535	12,71 63,66 636,2
2	18,51 98,49 998,46	19,00 99,01 999,00	19,16 99,17 999,20	19,25 99,25 999,20	19,30 99,30 999,20	19,33 99,33 999,20	19,37 99,36 999,40	19,41 99,42 999,60	19,45 99,46 999,40	19,50 99,50 999,40	4,30 9,92 31,00

Продолжение

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞	t
6	5,99 13,74 35,51	5,14 10,92 26,99	4,76 9,78 23,70	4,53 9,15 21,90	4,39 8,75 20,81	4,28 8,47 20,03	4,15 8,10 19,03	4,00 7,72 17,99	3,84 7,31 16,89	3,67 6,88 15,75	2,45 3,71 5,96
7	5,59 12,25 29,22	4,74 9,55 21,69	4,35 8,45 18,77	4,12 7,85 17,19	3,97 7,46 16,21	3,87 7,19 15,52	3,73 6,84 14,63	3,57 6,47 13,71	3,41 6,07 12,73	3,23 5,65 11,70	2,36 3,50 5,40
8	5,32 11,26 25,42	4,46 8,65 18,49	4,07 7,59 15,83	3,84 7,10 14,39	3,69 6,63 13,49	3,58 6,37 12,86	3,44 6,03 12,04	3,28 5,67 11,19	3,12 5,28 10,30	2,99 4,86 9,35	2,31 3,36 5,04
9	5,12 10,56 22,86	4,26 8,02 16,39	3,86 6,99 13,90	3,63 6,42 12,56	3,48 6,06 11,71	3,37 5,80 11,13	3,23 5,47 10,37	3,07 5,11 9,57	2,90 4,73 8,72	2,71 4,31 7,81	2,26 3,25 4,78
10	4,96 10,04 21,04	4,10 7,56 14,91	3,71 6,55 12,55	3,48 5,99 11,28	3,33 5,64 10,48	3,22 5,39 9,92	3,07 5,06 9,20	2,91 4,71 8,45	2,74 4,33 7,64	2,54 3,91 6,77	2,23 3,17 4,59
11	4,84 9,65 19,69	3,98 7,20 13,81	3,59 6,22 11,56	3,36 5,67 10,35	3,20 5,32 9,58	3,09 5,07 9,05	2,95 4,74 8,35	2,79 4,40 7,62	2,61 4,02 6,85	2,40 3,60 6,00	2,20 3,11 4,49
12	4,75 9,33 18,64	3,88 6,93 12,98	3,49 5,95 10,81	3,26 5,41 9,63	3,11 5,06 8,89	3,00 4,82 8,38	2,85 4,50 7,71	2,69 4,16 7,00	2,50 3,78 6,25	2,30 3,36 5,42	2,18 3,06 4,32

Продолжение

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞	t
13	4,67 9,07 17,81	3,80 6,70 12,31	3,41 5,74 10,21	3,18 5,20 9,07	3,02 4,86 8,35	2,92 4,62 7,86	2,77 4,30 7,21	2,60 3,96 6,52	2,42 3,59 5,78	2,21 3,16 4,97	2,16 3,01 4,12
14	4,60 8,86 17,14	3,74 6,51 11,78	3,34 5,56 9,73	3,11 5,03 8,62	2,96 4,69 7,92	2,85 4,46 7,44	2,70 4,14 6,80	2,53 3,80 6,13	2,35 3,43 5,41	2,13 3,00 4,60	2,14 2,98 4,14

Продолжение

υ_2 \ υ_1	1	2	3	4	5	6	8	12	24	∞	t
20	4,35 8,10 14,82	3,49 5,85 9,95	3,10 4,94 8,10	2,87 4,43 7,10	2,71 4,10 6,46	2,60 3,87 6,02	2,45 3,56 5,44	2,28 3,23 4,82	2,08 2,86 4,15	1,84 2,42 3,38	2.09 2,84 3,85
21	4,32 8,02 14,62	3,47 5,78 9,77	3,07 4,87 7,94	2,84 4,37 6,95	2,68 4,04 6,32	2,57 3,81 5,88	2,42 3,51 5,31	2,25 3,17 4,70	2,05 2,80 4,03	1,82 2,36 3,26	2,08 2,83 3,82
22	4,30 7,94 14,38	3,44 5,72 9,61	3,05 4,82 7,80	2,82 4,31 6,87	2,66 3,99 6,19	2,55 3,75 5,76	2,40 3,45 5,19	2,23 3,12 4,58	2,03 2,75 3,92	1,78 2,30 3,15	2,07 2,82 3,79
23	4,28 7,88 14,19	3,42 5,66 9,46	3,03 4,76 7,67	2,80 4,26 6,70	2,64 3,94 6,08	2,53 3,71 5,56	2,38 3,41 5,09	2,20 3,07 4,48	2,00 2,70 3,82	1,76 2,26 3,05	2,07 2,81 3,77
24	4,26 7,82 14,03	3,40 5,61 9,34	3,01 4,72 7,55	2,78 4,22 6,59	2,62 3,90 5,98	2,51 3,67 5,55	2,36 3,36 4,99	2,18 3,03 4,39	1,98 2,66 3,84	1,73 2,21 2,97	2,06 2,80 3,75
25	4,24 7,77 13,88	3,38 5,57 9,22	2,99 4,68 7,45	2,76 4,18 6,49	2,60 3,86 5,89	2,49 3,63 5,46	2,34 3,32 4,91	2,16 2,99 4,31	1,96 2,62 3,66	1,71 2,17 2,89	2,06 2,79 3,72
26	4,22 7,72 13,74	3,37 5,53 9,12	2,98 4,64 7,36	2,74 4,14 6,41	2,59 3,82 5,80	2,47 3,59 5,38	2,32 3,29 4,83	2,15 2,96 4,24	1,95 2,58 3,59	1,69 2,13 2,82	2,06 2,78 3,71

Продолжение

υ_2 \ υ_1	1	2	3	4	5	6	8	12	24	∞	t
27	4,21 7,68 13,61	3,35 5,49 9,02	2,96 4,60 7,27	2,73 4,11 6,33	2,57 3,78 5,73	2,46 3,56 5,31	2,30 3,26 4,76	2,13 2,93 4,17	1,93 2,55 3,52	1,67 2,10 2,76	2,05 2,77 3,69
28	4,19 7,64	3,34 5,45	2,95 4,57	2,71 4,07	2,56 3,75	2,44 3,53	2,29 3,23	2,12 2,90	1,91 2,52	1,65 2,06	2,05 2,76

Приложение 5

Таблица Фишера – Иейтса

Значения $r_{кр}$, найденные для уровня значимости α и чисел степеней свободы $\upsilon = n - 2$ в случае парной корреляции и $\upsilon = n - l - 2$, где l – число исключенных величин в случае частной корреляции

υ	Двусторонние границы				υ	Двусторонние границы			
	0,05	0,02	0,01	0,001		0,05	0,02	0,01	0,001
1	0,997	1,000	1,000	1,000	16	0,468	0,543	0,590	0,708
2	0,950	0,980	0,990	0,999	17	0,456	0,529	0,575	0,693
3	0,878	0,934	0,959	0,991	18	0,444	0,516	0,561	0,679
4	0,811	0,882	0,917	0,974	19	0,433	0,503	0,549	0,665
5	0,754	0,833	0,875	0,951	20	0,423	0,492	0,537	0,652
6	0,707	0,789	0,834	0,925	25	0,381	0,445	0,487	0,597
7	0,666	0,750	0,798	0,898	30	0,349	0,409	0,449	0,554
8	0,632	0,715	0,765	0,872	35	0,325	0,381	0,418	0,519
9	0,602	0,685	0,735	0,847	40	0,304	0,358	0,393	0,490
10	0,576	0,658	0,708	0,823	45	0,288	0,338	0,372	0,465
11	0,553	0,634	0,684	0,801	50	0,273	0,322	0,354	0,443
12	0,532	0,612	0,661	0,780	60	0,250	0,295	0,325	0,408
13	0,514	0,592	0,641	0,760	70	0,232	0,274	0,302	0,380
14	0,497	0,574	0,623	0,742	80	0,217	0,257	0,283	0,338
15	0,482	0,558	0,606	0,725	90	0,205	0,242	0,267	0,338
					100	0,195	0,230	0,254	0,321
υ	0,025	0,01	0,005	0,0005	υ	0,025	0,01	0,005	0,0005
	Односторонние границы					Односторонние границы			

Приложение 6

Таблица Z-преобразования Фишера

$$Z = \frac{1}{2}\{\ln(1+r) - \ln(1-r)\}$$

r	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0101	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
1	0,1003	0,1104	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923

Приложение 7

Критерий А. Н. Колмогорова.
Точные и асимптотические границы для верхней грани модуля разности истинной и эмпирической функции распределения

n	Уровень значимости 0,05			Уровень значимости 0,01		
	точная граница	асимптотическая граница	отношение	точная граница	асимптотическая граница	отношение
5	0,5633	0,6074	1,078	0,6685	0,7279	1,089
10	0,4087	0,4295	1,051	0,4864	0,5147	1,058
15	0,3375	0,3507	1,039	0,4042	0,4202	1,040
20	0,2939	0,3037	1,033	0,3524	0,3639	1,033
25	0,2639	0,2716	1,029	0,3165	0,3255	1,028
30	0,2417	0,2480	1,026	0,2898	0,2972	1,025
40	0,2101	0,2147	1,022	0,2521	0,2574	1,021
50	0,1884	0,1921	1,019	0,2260	0,2302	1,018
60	0,1723	0,1753	1,018	0,2067	0,2101	1,016
70	0,1597	0,1623	1,016	0,1917	0,1945	1,015
80	0,1496	0,1518	1,015	0,1795	0,1820	1,014
90	0,1412	0,1432	1,014			
100	0,1340	0,1358	1,013			

При n > 100 следует принять асимптотические границы

$$\bar{\varepsilon}_{0,05} = \frac{1,36}{\sqrt{n}} \quad \text{и} \quad \bar{\varepsilon}_{0,01} = \frac{1,63}{\sqrt{n}},$$

для которых истинные коэффициенты доверия несколько больше заданных величин 0,95 и 0,99 соответственно.

Приложение 8

Значения плотности $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ вероятности для нормированного нормального закона распределения. $f(-t) = f(t)$

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825

Продолжение

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,1	0,0001338	–	–	–	–	–	–	–	–	–
4,5	0,000160	–	–	–	–	–	–	–	–	–
5,0	0,0000015	–	–	–	–	–	–	–	–	–

Приложение 9

Пяти- и однопроцентные пределы для отношения G наибольшей выборочной дисперсии к сумме I выборочных дисперсий, полученных из I независимых выборок объема n

Первое значение соответствует уровню значимости $\alpha = 0,05$, а второе – $\alpha = 0,01$

I \ n–1	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,998	0,975	0,939	0,906	0,877	0,853	0,833	0,816	0,801	0,788	0,734	0,660	0,518	0,500

Приложение 10

Квантили распределения выборочных характеристик эксцесса $E_к$ и асимметрии A_c

Объем выборки n	Критические значения коэффициента					
	эксцесса $E_к$ при $1 - \alpha$				асимметрии A_c при $1 - \alpha$	
	0,99	0,95	0,05	0,01	0,95	0,99
50	4,92	4,01	2,13	1,95	0,533	0,787
100	40	3,77	35	2,18	389	567
150	14	66	45	30	321	464
200	3,98	57	51	37	280	403
250	87	51	55	42	251	360
300	79	47	59	46	230	329
350	72	44	62	50	213	305
400	67	41	64	52	200	285
450	63	39	66	55	188	269
500	60	37	67	57	179	255
550	57	35	69	58	171	243
600	54	34	70	60	163	233
650	52	33	71	61	157	224
700	50	31	72	62	151	215
750	48	30	73	64	146	208
800	46	29	74	65	142	202
850	45	28	74	66	138	196
900	43	28	75	66	134	190
950	42	27	76	67	130	185
1000	41	26	76	68	127	180
1200	37	24	78	71	116	165
1400	34	22	80	72	107	152
1600	32	21	81	74	100	142
1800	30	20	82	76	095	134
2000	28	18	83	77	090	127
2500	25	16	85	79	080	114
3000	22	15	86	81	073	104
3500	21	14	87	82	068	096
4000	19	13	88	83	064	090
4500	18	12	88	84	060	085
5000	17	12	89	85	057	081

Критические значен
и коэффи

а) урове

k_2 \ k_1	1	2	
3	0,771	865	9
4	658	776	8
5	569	699	7
6	500	632	7
7	444	575	6
8	399	527	6
9	362	488	5
10	332	451	5
11	306	420	4
12	283	394	4
14	247	345	4
16	219	312	3
18	197	283	3
20	179	259	3
22	164	238	2
24	151	221	2
26	140	206	2
28	130	193	2
30	122	182	2
32	115	171	2
34	108	162	2
36	102	153	1
38	097	146	1
40	093	139	1
50	075	113	1
60	063	195	1
80	047	072	0
100	038	058	0
120	032	049	0
200	019	030	0
400	010	015	0

Приложение 12

Таблица 5%-ного и 1%-ного уровней вероятности коэффициентов корреляции (r_a)

Размер выборки	Положительные значения r_a		Отрицательные значения r_a	
	5%-ный уровень	1%-ный уровень	5%-ный уровень	1%-ный уровень
5	0,253	0,297	– 0,753	– 0,798
6	0,354	0,447	– 0,708	– 0,863
7	0,370	0,510	– 0,674	– 0,799
8	0,371	0,531	– 0,625	– 0,764
9	0,366	0,533	– 0,593	– 0,737
10	0,360	0,525	– 0,564	– 0,705
11	0,353	0,515	– 0,539	– 0,679
12	0,348	0,505	– 0,516	– 0,655
13	0,341	0,495	– 0,497	– 0,634
14	0,335	0,485	– 0,479	– 0,615
15	0,328	0,475	– 0,462	– 0,597
20	0,299	0,432	– 0,399	– 0,524
25	0,276	0,398	– 0,356	– 0,473
30	0,257	0,370	– 0,324	– 0,433
35	0,242	0,347	– 0,300	– 0,401
40	0,229	0,329	– 0,279	– 0,376
45	0,218	0,313	– 0,262	–0,256
50	0,208	0,301	– 0,248	– 0,339

Приложение 13

Значения средней μ и стандартных ошибок σ_1 σ_2 для n от 10 до 50

n	μ	σ_1	σ_2
10	3,858	1,288	1,964
15	4,636	1,521	2,153
20	5,195	1,677	2,279
25	5,632	1,791	2,373
30	5,990	1,882	2,447
35	6,294	1,956	2,509
40	6,557	2,019	2,561
45	6,790	2,072	2,606
50	6,998	2,121	2,645

Приложение 14 ОСНОВНЫЕ ФОРМУЛЫ СТАТИСТИКИ

Формула Стерджесса (для определения оптимального числа групп):

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

где п - число групп;

N - число единиц совокупности.

Величина равного интервала:

$$h = \frac{x_{max} - x_{min}}{n} = \frac{R}{n},$$

где $R = x_{max} - x_{min}$,

т. е. размах вариации;

x_{max} - наибольшее значение варьирующего признака;
x_{min} - наименьшее значение варьирующего признака.

Величина неравных интервалов:

а) изменяющихся в арифметической прогрессии:

$$h_{i+1} = h_i + a,$$

б) изменяющихся в геометрической прогрессии:

$$h_{i+1} = h_i \cdot q,$$

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

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

СРЕДНИЕ ВЕЛИЧИНЫ

Общая формула степенной средней:

- простая (невзвешенная V

$$\bar{x} = \sqrt[k]{\frac{\Sigma x_i^k}{n}},$$

где x_i - i-й вариант осредняемого признака,
п - число вариантов;
$\bar{x}$ - средняя величина признака.

- взвешенная

$$\bar{x} = \sqrt[k]{\frac{\Sigma x_i^k f_i}{\Sigma f_i}}.$$

где f - частота (статистический вес) 1-го варианта; k - порядок средней.

При k = 2 получается средняя квадратическая$(\bar{x}_q)$; при к = 1 -средняя арифметическая (x); при к = 0 - средняя геометрическая $(\bar{x}_g)$; при к = -1 - средняя гармоническая$(\bar{x}_h)$

Правило мажорантности :$\bar{x}_q > \bar{x} > \bar{x}_g > \bar{x}_h$.

Средняя арифметическая:

- простая (невзвешенная)

Σx

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i}.$$

- взвешенная

Средняя гармоническая:

- простая (невзвешенная)

Средняя геометрическая:

- простая (невзвешенная)

$$\bar{x} = \sqrt[n]{\Pi x_i};$$

- взвешенная

$$\bar{x} = \sqrt[\Sigma f]{\Pi (x_i)^{f_i}}$$

СТРУКТУРНЫЕ (НЕПАРАМЕТРИЧЕСКИЕ) СРЕДНИЕ

Мода интервального ряда распределения:

$$Mo = x_o + i \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})},$$

где x_0 - нижняя граница модального интервала;
i - величина модального интервала; f_{Mo} - частота модального интервала; f_{Mo}_, - частота интервала, предшествующего модальному; *м<>+1 ~ частота интервала, следующего за модальным.

» взвешенная

$$\bar{x} = \frac{n}{\Sigma \frac{1}{x_i}};$$

где Wj - сложный вес; $w_(= x^$.

$$\bar{x} = \frac{\Sigma w_i}{\Sigma \frac{w_i}{x_i}},$$

Средняя квадратическая:

- простая (невзвешенная)

Медиана интервального ряда распределения:

$$Me = x_o + i \cdot \frac{\frac{1}{2}\Sigma f - S_{Me-1}}{f_{Me}},$$

где x_0 - нижняя граница медианного интервала; $\sum f$ — сумма частот;
S_{Me-1} - накопленная частота интервала, предшествующего медианно му;
f_{Me} - частота медианного интервала.

$$\bar{x} = \sqrt{\frac{\Sigma x_i^2}{n}}$$

$$\bar{x} = \sqrt{\frac{\Sigma x_i^2 f_i}{\Sigma f_i}}$$

$$Q_1 = x_{Q1} + i \cdot \frac{\frac{1}{4}\Sigma f -}{f_Q}$$

взвешенная

Квартили

Первый

Третий

$$Q_3 = x_{Q_3} + i \cdot \frac{\frac{3}{4}\Sigma f - S_{Q_3-1}}{f_{Q_3}},$$

где x_{Q_1} - нижняя граница интервала, содержащего нижний квартиль;
x_{Q_3} - нижняя граница интервала, содержащего верхний квартиль; S_{Q_1-1}
- накопленная частота интервала, предшествующего интервалу, содержащему нижний квартиль;
S_{Q_3-1} - то же, для верхнего квартиля;
- частота интервала, содержащего нижний квартиль;
f_{Q_1}
f_{Q_3} - то же, для верхнего квартиля.

Децили

Первый

Второй

$$d_2 = x_{d_2} + i \cdot \frac{\frac{2}{10}\Sigma f - S_{d_2-1}}{f_{d_2}},$$

$$d_1 = x_{d_1} + i \cdot \frac{10}{f_{d_1}}.$$

где x_{d_1} - нижняя граница интервала, содержащего первый дециль;
x_{d_2} - то же, для второго дециля;
i - величина интервала, содержащего первый (второй) де циль;

S_{d_1-1} и S_{d_2-1} - накопленные частоты интервалов, предшествующих интервалу, содержащему первый (второй) дециль;

f_{d_1} и f_{d_2} - частоты интервалов, содержащих первый (второй) дециль.

Остальные децили вычисляют по аналогичной схеме.

ПОКАЗАТЕЛИ ВАРИАЦИИ Размах вариации:

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

Среднее линейное отклонение:

- невзвешенное

$$\bar{d} = \frac{\Sigma|x_i - \bar{x}|}{n};$$

- взвешенное

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

где $|x_i - \bar{x}|$- абсолютное значение отклонений.

Дисперсияσ^2- средний квадрат отклонений

Дисперсия вариационного признака:

- невзвешенная

$$\Sigma(x_i - \bar{x})^2$$

[1] взвешенная

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2 f_i}{\Sigma f_i}.$$

Среднее квадратическое отклонение вариационного признака:

- невзвешенное

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}};$$

> взвешенное

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2 f_i}{\Sigma f_i}}.$$

Дисперсия альтернативного признака:

$$\sigma^2_p = pq,$$

где p - доля единиц в совокупности, обладающих данным признаком; q - доля единиц, не обладающих данным признаком.

Среднее квадратическое отклонение альтернативного признака:

$$\sigma_p = \sqrt{pq}.$$

Коэффициент осцилляции:

Линейный коэффициент вариации:

$$V_R = \frac{R}{\bar{x}} \cdot 100\%.$$

$$V_{\bar d} = \frac{d}{\bar x} \cdot 100\%.$$

Коэффициент вариации:

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

Показатель энтропии:

$$H(x) = -\sum p_i \cdot \log p_i,$$

гдеp_i- вероятности различных значений случайной величины.

Правило сложения дисперсий для средней величины признака:

$$\sigma^2 = \bar\sigma_i^2 + \delta_x^2,$$

где σ^2 — общая дисперсия;
$\bar\sigma_i^2$ - средняя из внутрифупповых дисперсий;
δ_x^2 - межгрупповая дисперсия.

Средняя из внутригрупповых дисперсий:

$$\bar\sigma_i^2 = \frac{\Sigma \sigma_i^2 n_i}{\Sigma n_i}$$

Эмпирическое корреляционное отношение:

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma^2}}.$$

Правило сложения дисперсий для доли признака:

$$\sigma_{\bar p}^2 = \overline{\sigma_{p_i}^2} + \delta_{p_i}^2,$$

где $\sigma_{\bar p}^2$— общая дисперсия доли;
$\bar\sigma_{p_i}^2$ - средняя из внутригрупповых дисперсий доли;
$\delta_{p_i}^2$ - межгрупповая дисперсия доли.

Общая дисперсия доли:

$$\sigma_{\bar p}^2 = \bar p(1-\bar p),$$

где$\bar p$ - доля изучаемого признака во всей совокупности, определяемая по формуле:

$\bar p = \frac{\Sigma p_i n_i}{\Sigma n}$ (p_i- групповые доли). **Средняя из внутригрупповых дисперсий**

доли:

$$\overline{\sigma_{p_i}^2} = \overline{p_i(1-p_i)} = \frac{\Sigma p_i(1-p_i)n_i}{\Sigma n_i}.$$

где σ_i^2- групповые дисперсии;n_i-число единице группах. **Межгрупповая дисперсия:**

$$\delta_x^2 = \frac{\Sigma(\bar x_i - \bar x)^2 n_i}{\Sigma n_i},$$

где $\bar{x}_i$ - групповые средние; $\bar{x}$— общая средняя.

Эмпирический коэффициент детерминации:

$$\eta^2 = \frac{\delta_x^2}{\sigma^2}.$$

Межгрупповая дисперсия доли:

$$\delta_{p_i}^2 = \frac{\Sigma(p_i - \bar{p})^2 n_i}{\Sigma n_i}.$$

Коэффициент асимметрии:

$$A_s = \frac{\mu_3}{\sigma^3},$$

где μ_3 - центральный момент третьего порядка, определяемый по фор-

$$\mu_3 = \frac{\Sigma(x_i - \bar{x})^3 f_i}{\Sigma f_i}.$$

Эксцесс:

$$E_k = \frac{\mu_4}{\sigma^4} - 3,$$

где μ_4 - центральный момент четвертого порядка, определяемый по фор муле:

$$\mu_4 = \frac{\Sigma(x_i - \bar{x})^4 f_i}{\Sigma f_i}.$$

где y_t — ордината кривой нормального распределения;

$t = \frac{x_i - \bar{x}}{\sigma}$ - стандартизированная (нормированная) величина; е и *n* - математические постоянные.

Нормальное распределение:

$$y_t = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2},$$

КРИТЕРИИ СОГЛАСИЯ Пирсона:

$$\chi^2 = \Sigma\frac{(f_\text{э} - f_T)^2}{f_T},$$

где $f_\text{э}$ - эмпирические частоты; f_T - теоретические частоты.

Романовского:

$$C = \frac{\chi^2 - \gamma}{\sqrt{2\gamma}},$$

где у- число степеней свободы (при проверке гипотезы о нормальности распределения) равно числу групп минус три.

Ястремского:

$$L = \frac{\Sigma\frac{(f_\text{э} - f_T H)^2}{N \cdot pq} - K}{\sqrt{2K + 4Q}},$$

где N - объем совокупности;
pq - дисперсия альтернативного признака;
K - число вариантов или групп;
Q - принимает значение 0,6 при числе вариантов или групп от 8 до 20.

Колмогорова:

$$\lambda = \frac{D}{\sqrt{\Sigma f}},$$

где D - максимальное значение разности между накопленными эмпирическими и теоретическими частотами; £f - сумма эмпирических частот.

ОШИБКА РЕПРЕЗЕНТАТИВНОСТИ (большая выборка)

Предельная ошибка выборки

$$\Delta = t\mu,$$

где t - коэффициент доверия, вычисляемый по таблицам в зависимости от вероятности; μ - средняя ошибка выборки.

Соотношение между генеральной и выборочной дисперси-

$$\sigma_{\bar{x}}^2 = \sigma_{\tilde{x}}^2 \cdot \frac{n}{n-1},$$

где сг^ - генеральная дисперсия; $\sigma_{\tilde{x}}^2$ - выборочная дисперсия; п - численность выборки.

Средняя ошибка собственно-случайной выборки:

- повторный выбор

$$\mu = \sqrt{\frac{\sigma_{\tilde{x}}^2}{n}};$$

[1] бесповторный отбор

$$\mu = \sqrt{\frac{\sigma_{\tilde{x}}^2}{n}\left(1 - \frac{n}{N}\right)},$$

где N - численность генеральной совокупности. **Средняя ошибка механической выборки:**

$$\mu = \sqrt{\frac{\sigma_{\tilde{x}}^2}{n}\left(1 - \frac{n}{N}\right)}.$$

Средняя ошибка типической выборки

При отборе, пропорциональном объему типических групп: • повторный отбор

$$\mu = \sqrt{\frac{\overline{\sigma_i^2}}{n}},$$

где$\overline{\sigma_i^2}$- средняя из внутригрупповых дисперсий; • бесповторный отбор

$$\mu = \sqrt{\frac{\overline{\sigma_i^2}}{n}\left(1 - \frac{n}{N}\right)}.$$

При отборе, пропорциональном дифференциации признака: • повторный отбор

$$\mu = \frac{1}{N}\sqrt{\sum \frac{\sigma_i^2 N_i^2}{n_i}},$$

гдеσ_i^2— дисперсия признака в i-й группе;
N_i - численность генеральной совокупности в i-й группе; п; - численность выборки в i-й группе;

- бесповторный отбор

$$\mu = \frac{1}{N}\sqrt{\sum \frac{\sigma_i^2 N_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)}.$$

Средняя ошибка серийной выборки:

- повторный отбор

$$\mu = \sqrt{\frac{\delta^2}{r}},$$

где δ^2 - межгрупповая (межсерийная) дисперсия; г - число отобранных серий;

- бесповторный отбор

Средняя ошибка малой выборки:

$$\mu_{MB} = \sqrt{\frac{\sigma^2_{MB}}{n}}$$

где σ^2_{MB} - выборочная дисперсия (в малой выборке);

п - число отобранных в малую выборку единиц. При этом

$$\sigma^2_{MB} = \frac{\Sigma(x - \tilde{x})^2}{n - 1}.$$

НЕОБХОДИМЫЙ ОБЪЕМ ВЫБОРКИ ПРИ ОПРЕДЕЛЕНИИ СРЕДНЕГО ЗНАЧЕНИЯ ПРИЗНАКА

Необходимый объем собственно-случайной и механической выборки:

- повторная

$$n = \frac{t^2\sigma^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}};$$

> бесповторная

$$n = \frac{t^2\sigma^2_{\tilde{x}}N}{\Delta^2_{\tilde{x}}N + t^2\sigma^2_{\tilde{x}}}.$$

Необходимый объем типической выборки

При отборе, пропорциональном объему типических групп: • повторная

$$n = \frac{t^2\overline{\sigma}^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}};$$

^{1} бесповторная

$$n = \frac{t^2\overline{\sigma}^2_{\tilde{x}}N}{\Delta^2_{\tilde{x}}N + t^2\overline{\sigma}^2_{\tilde{x}}}.$$

$$\mu = \sqrt{\frac{\delta^2}{r}(i - \frac{r}{R})},$$

где R - число серий в генеральной совокупности.

При отборе, пропорциональном дифференциации признака:

$$n_i = \frac{nN_i\sigma_i}{\Sigma N_i\sigma_i}.$$

Необходимый объем серийной выборки:

- повторная

$$r = \frac{t^2\delta^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}};$$

бесповторная

$$T_{пр} = \frac{\Delta_{i/i-1}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100 = (k$$

$$r = \frac{t^2\delta_{\tilde{x}}^2 R}{\Delta_{\tilde{x}}^2 R + t^2\delta_{\tilde{x}}^2}.$$

Относительный прирост (T - темп прироста): • базисный

$$T_{пр} = \frac{\Delta_{i/1}}{y_1} = (k_{p_{i/1}} - 1)\cdot 100 = T_{p_{i/1}} - 100;$$

- цепной

Абсолютное значение 1% пр

$$|\%| = \frac{\Delta_{i/i-1}}{T_{пр_{i/i-1}} \cdot \%} = \frac{y_i -}{\frac{y_i - y_{i-}}{y_{i-1}}}$$

ПОКАЗАТЕЛИ АНАЛИЗА РЯДА ДИНАМИКИ

Темп роста:

- базисный
  цепной

$$T_p = \frac{y_i}{y_{i-1}} \cdot 100,$$

где y_i - порядковый уровень ряда динамики; y_1 - базисный уровень ряда динамики.

Абсолютный прирост:

базисный

$$\Delta_y = y_i - y_1.$$

Средний абсолютный прирост:

- По базисному абсолютному приросту:

$$\bar{\Delta}_y = \frac{\Delta_{i/1}}{n-1} = \frac{y_n - y_1}{n-1},$$

где $y_п$ - конечный уровень ряда динамики; п - число уровней ряда динамики.

- по цепным абсолютным приростам:

$$\bar{\Delta}_y = \frac{\sum_{i=1}^{n-1}\Delta_{i/i-1}}{n-1}.$$

СРЕДНИЙ УРОВЕНЬ РЯДА ДИНАМИКИ: (У- **средняя хронологическая) В моментном ряду динамики:**

- с равноотстоящими уровнями

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \ldots + y_{n-1} + \frac{1}{2}y_n}{n-1} = \frac{\frac{y_1 + y_n}{2} + \sum_{i=2}^{n-1} y_i}{n-1};$$

- с неравноотстоящими уровнями

$$\overline{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \ldots + (y_{n-1} + y_n)t_{n-1}}{2(t_1 + t_2 + t_3 + \ldots + t_{n-1})} = \frac{\Sigma(y_i + y_{i+1})t}{2\sum\limits_{i=1}^{n-1} t},$$

где y_i, y_n - уровни ряда динамики;

t_i - длительность интервала времени между уровнями.

В интервальном ряду динамики:

- с равноотстоящими уровнями

$$\overline{y} = \frac{\sum\limits_{i=1}^{n} y_i}{n};$$

525

- **с неравноотстоящими** уровнями

$$\overline{y} = \frac{\sum\limits_{i=1}^{n} y_i t_i}{\Sigma t_i}.$$

Средний темп роста ($\overline{T}_p$ - средняя геометрическая в рядах динамики).

По цепным коэффициентам (темпам) роста:

$$\overline{T}_p = \sqrt[m]{k_{p2/1} \cdot k_{p3/2} \cdots k_{n/n-1}} = \sqrt[m]{\Pi k_{pi/i-1}},$$

где m - число темпов роста.

По абсолютным уровням ряда динамики:

$$\overline{T}_p = \sqrt[n-1]{\frac{y_n}{y_1}}.$$

По базисным темпам (коэффициентам) роста:

$$\overline{T}_p = \sqrt[n-1]{T_{p_{n/1}}}.$$

Для рядов динамики с неравноотстоящими уровнями:

$$\overline{T}_p = \sqrt[\Sigma t]{(k_{p_{2/1}})^{t_1} \cdot (k_{p_{3/2}})^{t_2} \cdots (k_{n/n-1})^{t_n}},$$

где t - интервал времени, в течение которого сохраняется данный темп роста; Σt - сумма отрезков времени периода.

Средний темп прироста:

$$\bar{T}_{пр} = \bar{T}_p - 100.$$

ПРОВЕРКА ГИПОТЕЗЫ О СУЩЕСТВОВАНИИ ОСНОВНОЙ ТЕНДЕНЦИИ В РЯДАХ ДИНАМИКИ

Метод проверки существенности разности средних:

- **t-критерий:**

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}};$$

- **среднее** квадратическое отклонение разности **средних:**

$$\sigma = \sqrt{\frac{(n_1 - 1)^2 \cdot \sigma_1^2 + (n_2 - 1)^2 \cdot \sigma_2^2}{n_1 + n_2 - 2}};$$

- **дисперсии** для первой и второй части ряда динамики:

$$\sigma_i^2 = \frac{\Sigma(y_i - \bar{y}_i)^2}{n - 1}, \quad i = 1,\ 2,\ \ldots,\ n.$$

где $\bar{y}_1$ и $\bar{y}_2$ - средние для первой и второй половины ряда динамики (У;);
n_1 и n_2 - число наблюдений в этих частях ряда;
y_i - фактические уровни ряда динамики; п - число уровней ряда динамики.

Метод Фостера-Стюарта:

t-критерий

$$t_s = \frac{S - \mu}{\sigma_1}; \quad t_d = \frac{d - 0}{\sigma_2},$$

где

$$S = \sum_{i=2}^{n} S_i; \quad d = \sum_{i=2}^{n} d_i,$$

где

$$S_i = U_i + e_i;\ d_i = U_i - e_i,$$

если $y_i > y_{i-1};\ y_{i-2}, \ldots, y_1$, то $U_i = 1,\ e = 0$; если $y_i < y_{i-1},\ y_{i-2}, \ldots, y_1$, то $U_i = 0,\ e = 1$; μ - среднее значение величины S, определенное для ряда, в котором уровни расположены случайным образом;
σ_1 - стандартная ошибка величины S; σ_2 - стандартная ошибка величины d.

Метод наименьших квадратов при расчете параметров полиномов.

- по прямой

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

где a_0 и a_1 - параметры уравнения;
t - показатели времени.

3

При $\Sigma t = 0$ параметры a_0 и a_1 определяются:

$$a_0 = \frac{\Sigma y}{n}; \quad a_1 = \frac{\Sigma yt}{\Sigma t^2};$$

- по параболе

$$y_t = a_0 + a_1 t + a_2 t_2.$$

При $\Sigma t = 0$ параметры a_0, a_1 и a_2 определяются:

$$\begin{cases} na_0 + a_2 \Sigma t^2 = \Sigma y \\ a_1 \Sigma t^2 = \Sigma ty \\ a_0 \Sigma t^2 + a_2 \Sigma t^4 = \Sigma t^2 y \end{cases}$$

- по показательной кривой

$$\bar{y}_t = a_0 a^t{}_1.$$

При $\Sigma t = 0$ параметры a_0 и a_1 определяются

$$n \lg a_0 = \Sigma \lg y$$

$$\lg a_1 \Sigma t^2 = \Sigma t \lg y;$$

Ряд Фурье:

$$\bar{y}_t = a_0 + \sum_{t=0}^{m} (a_k \cos kt + b_k \sin kt);$$

$$a_0 = \frac{1}{2} \Sigma y; \quad a_k = \frac{2}{n} \Sigma y \cos kt; \quad b_k = \frac{2}{n} \Sigma y \sin t,$$

. где t - время, последовательные значение которого определяются от с увеличением (приростом), равным *2п/п; п* - число уровней ряда динамики.

Первая гармоника ряда Фурье:

$$\bar{y}_t = a_0 + a_1 \cos t + b_1 \sin t;$$

$$a_0 = \frac{\Sigma y}{12}; \; a_1 = \frac{\Sigma y \cos t}{6}; \; b_1 = \frac{\Sigma y \sin t}{6}.$$

Показате I_S **и «сезонной волны»** (- индекс сезонности) • в стабильных рядах динамики

$$I_s = \frac{\bar{y}_i}{\bar{y}} \cdot 100\%,$$

где $\bar{y}_i$ - осредненные фактические данные по одноименным периодам; $\bar{y}$ - общая (постоянная) средняя;

- в рядах динамики с тенденцией роста (уменьшения)

$$I_s = \frac{\left[\Sigma \frac{y_i}{\bar{y}_t} \cdot 100\right]}{n},$$

где $\bar{y}_t$ - выравненные уровни ряда динамики по уравнению тренда (переменная средняя); п - число одноименных периодов.

СВЯЗНЫЙ АНАЛИЗ РЯДОВ ДИНАМИКИ Коэффициент автокорреляции:

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - \bar{y}_t \cdot \bar{y}_{t+1}}{\sigma_{y_t} \cdot \sigma_{y_{t+1}}} \cdot 100,$$

где $\sigma_{y_t}, \sigma_{y_{t+1}}$ - средние квадратические отклонения рядов у, и у,+| соответственно.

Коэффициент корреляции отклонений уровней ряда динамики от тренда:

$$r_{d_x d_y} = \frac{\Sigma d_x \cdot d_y}{\sqrt{\Sigma d_x^2 \cdot \Sigma d_y^2}},$$

где $d_x = x_i - \bar{x}_t$; $d_y = y_i - y_t$.

Коэффициент корреляции последовательных разностей уровней ряда динамики:

$$r_{\Delta_x \Delta_y} = \frac{\Sigma \Delta_x \cdot \Delta_y}{\sqrt{\Sigma \Delta_x^1 \cdot \Sigma \Delta_y^2}},$$

где $\Delta_x = x_i - x_{i-1}$, $\Delta_y = y_i - y_{i-1}$.

Экстраполяция тренда:

- по среднему абсолютному приросту

$$\hat{y}_{i+1} = \bar{y}_i + \bar{\Delta},$$

где $\hat{y}_{i+1}$ - экстраполируемый уровень:
(i+t) — номер этого уровня (года);
i - номер последнего уровня (года) исследуемого периода, за который рассчитан $\bar{\Delta}$; $\bar{\Delta}$ - средний абсолютный прирост;

- по среднему темпу роста

$$\hat{y}_{i+t} = y_i \cdot \bar{k}_p^t,$$

где y_i - последний уровень ряда динамики; t - срок прогноза;
$\bar{k}_p$ - средний коэффициент роста;

- по аналитическому выражению тренда

$$\hat{y}_t \pm t_\alpha \cdot \sigma_{\hat{y}_t},$$

где $\hat{y}_t$ - расчетное значение уровня ряда по уравнению тренда;
$\sigma_{\bar{y}}$ - средняя квадратическая ошибка тренда;
t_α - доверительная величина.

СТАТИСТИЧЕСКИЙ АНАЛИЗ СТРУКТУРЫ «Абсолютный» прирост удельного веса:

$$\Delta d_i = d_{ij} - d_{ij-1},$$

где d_{ij}—удельный вес i-й структурной части в j-й период времени. **Темп роста удельного веса:**

$$T_{p_{d_i}} = \frac{d_{ij}}{d_{ij-1}}.$$

Средний «абсолютный» прирост удельного веса:

$$\bar{\Delta}d_i = \frac{d_{in} - d_{i1}}{n-1},$$

где п - число периодов, за которые исследуются структурные сдвиги.

Средний темп роста удельного веса:

$$\bar{T}_{p_{d_i}} = \sqrt[n-1]{\frac{d_{in}}{d_{i1}}}.$$

Средний удельный вес:

$$\bar{d}_i = \frac{\sum_{j=1}^{n} x_{ij}}{\sum_{j=1}^{n}\sum_{i=1}^{k} x_{ij}} \cdot 100\%,$$

где x_{ij} - абсолютная величина i-й структурной части в j-й период времени;
k - общее число структурных частей в исследуемой совокупности.

Линейный коэффициент «абсолютных» структурных сдвигов:

$$\bar{\Delta}_{d_1-d_0} = \frac{\sum_{i=1}^{k}\left|d_{ij} - d_{ij-1}\right|}{k}.$$

Квадратический коэффициент «абсолютных» структурных сдвигов:

Квадратический коэффициент «относительных» структурных сдвигов:

$$\sigma_{\frac{d_1}{d_0}} = \sqrt{\sum_{i=1}^{k} \frac{(d_{ij} - d_{ij-1})^2}{d_{ij-1}} \cdot 100}.$$

Линейный коэффициент «абсолютных» структурных сдвигов за n периодов:

$$\overline{\Delta}^{(n)}_{d_1 - d_0} = \frac{\sum_{i=1}^{k} |d_{in} - d_{i1}|}{k(n-1)}.$$

Коэффициент Джини:

$$G = 1 - 2\sum_{i=1}^{k} d_{xi} d_{yi}^{H} + \sum_{i=1}^{k} d_{xi} d_{yi},$$

где d_{xi} - доля i-й группы в объеме совокупности;

d_{yi} - доля i-й группы в объеме признака;

d^{H}_{yi} - накопленная доля i-й группы в объеме признака.

Коэффициент Лоренца:

$$L = \frac{\sum_{i=1}^{k} |d_{xi} - d_{yi}|}{2},$$

где d_{xi} - удельный вес i-й группы в объеме совокупности;

d_{yi} - удельный вес i-й группы в объеме признака.

Показатель централизации:

$$I_z = \sum_{i=1}^{n} \left(\frac{x_i}{\sum_{i=1}^{n} x_i} \right)^2,$$

где x_i - значение признака i-й единицы совокупности.

ИНДЕКСЫ

Индивидуальные индексы:

- физического объема продукции

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

где q - количество продукции;

- цен

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

где p - цена единицы товара (продукта);

- себестоимости

$$i_z = \frac{z_1}{z_0},$$

где z - себестоимость единицы изделия (продукта);

- производительности труда

$$i_l = \frac{t_0}{t_1},$$

где t - затраты времени на изготовление единицы изделия (продукта);

- стоимости продукции

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0}.$$

Общие индексы в агрегатной форме:

- стоимости продукции или товарооборота

$$I_{pq} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0};$$

- физического объема

$$I_q = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0};$$

- цен

$$I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1};$$

- себестоимости продукции

$$I_z = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1};$$

- издержек производства

$$I_{zq} = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_0};$$

- производительности труда

$$I_t = \frac{\Sigma t_0 q_1}{\Sigma t_1 q_1};$$

- затрат времени на производство продукции

$$I_{tq} = \frac{\Sigma t_1 q_1}{\Sigma t_0 q_0}.$$

ПРЕОБРАЗОВАНИЕ ФОРМЫ АГРЕГАТНОГО ИНДЕКСА

Средняя арифметическая форма:

- индекс физического объема продукции

$$I_q = \frac{\Sigma i_q p_0 q_0}{\Sigma p_0 q_0}, \quad \text{где} \quad i_q = \frac{q_1}{q_0};$$

- индекс производительности труда

$$I_t = \frac{\Sigma i_t t_1 q_1}{\Sigma t_1 q_1} = \frac{\Sigma i_t T_1}{\Sigma T_1}, \quad \text{где} \quad i_t = \frac{t_0}{t_1};$$

- индекс Струмилина

$$I_v = \frac{\Sigma\left(\frac{q_1}{T_1} : \frac{q_0}{T_0}\right) \cdot T_1}{\Sigma T_1}.$$

Средняя гармоническая форма:

- индекс себестоимости

$$I_z = \frac{\Sigma z_1 q_1}{\Sigma \frac{z_1 q_1}{i_z}}, \quad \text{где} \quad i_z = \frac{z_1}{z_0};$$

- индекс цен

$$I_p = \frac{\Sigma p_1 q_1}{\Sigma \frac{p_1 q_1}{i_p}}, \quad \text{где} \quad i_p = \frac{p_1}{p_0}.$$

Взаимосвязи индексов:

$$I_{pq} = I_p \cdot I_q = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \cdot \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0};$$

$$I_{zd} = I_z \cdot I_q = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} \cdot \frac{\Sigma q_1 z_0}{\Sigma q_0 z_0};$$

$$I_{td} = I_q : I_t = \frac{\Sigma q_1 t_0}{\Sigma q_0 t_0} : \frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}.$$

Системы агрегатных индексов

индексы стоимости или товарооборота

а) цепные индексы

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}; \quad \frac{\Sigma p_2 q_2}{\Sigma p_1 q_1}; \quad \ldots; \quad \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_{n-1}};$$

б) базисные индексы

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}; \quad \frac{\Sigma p_2 q_2}{\Sigma p_0 q_0}; \quad \ldots; \quad \frac{\Sigma p_n q_n}{\Sigma p_0 q_0}.$$

Базисные индексы физического объема продукции с постоянными весами (p_0):

$$\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}; \quad \frac{\Sigma q_2 p_0}{\Sigma q_0 p_0}; \quad \ldots; \quad \frac{\Sigma q_n p_0}{\Sigma q_0 p_0}.$$

Цепные индексы физического объема продукции с постоянными весами (p_0):

$$\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}; \quad \frac{\Sigma q_2 p_0}{\Sigma q_1 p_0}; \quad \ldots; \quad \frac{\Sigma q_n p_0}{\Sigma q_{n-1} p_0}.$$

Базисные индексы цен с переменными весами:

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1};\quad \frac{\Sigma p_2 q_2}{\Sigma p_0 q_2};\quad \ldots;\quad \frac{\Sigma p_n q_n}{\Sigma p_0 q_n}.$$

Общая формула индекса переменного состава:

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\Sigma x_1 w_1}{\Sigma w_1} : \frac{\Sigma x_0 w_0}{\Sigma w_0},$$

где x - индексируемые величины; w - веса индекса.

Общая формула индекса постоянного состава:

$$I_x = \frac{\Sigma x_1 w_1}{\Sigma w_1} : \frac{\Sigma x_0 w_1}{\Sigma w_1}.$$

Общая формула индекса изменения структуры (структурных сдвигов):

$$I_{стр} = \frac{\Sigma x_0 w_1}{\Sigma w_1} : \frac{\Sigma x_0 w_0}{\Sigma w_0}.$$

«Идеальная» формула Фишера:

$$I_p = \sqrt{\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \cdot \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}}.$$

$$I_p = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}.$$

Индекс цен Ласпейреса: Индекс цен Пааше:

$$I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}.$$

ИЗМЕРЕНИЕ СВЯЗИ

Однофакторные модели связи (x)

- Прямолинейная

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

где a_0, a_1 — коэффициенты регрессии, определяются:

$$\begin{cases} na_0 + a_1 \Sigma x = \Sigma y; \\ a_0 \Sigma x + a_1 \Sigma x^2 = \Sigma xy, \end{cases}$$

или по сгруппированным данным:

$$\begin{cases} na_0 + a_1\Sigma x f_x = \Sigma y f_y; \\ a_0 \Sigma x f_x + a_1 \Sigma x^2 f_x = \Sigma x y f_y, \end{cases}$$

где f_x, f_y - число единиц совокупности согласно распределению соответственно по факторному и результативному признакам.

- По параболе второго порядка

$$\bar{y}_x = a_0 + a_1 x + a_2 x^2,$$

где a_0, a_1 и a_2 определяются:

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

- По гиперболе

$$\bar{y}_x = a_0 + a_1 \frac{1}{x},$$

где a_0, a_1 определяются:

$$\begin{cases} na_0 + a_1 \Sigma \frac{1}{x} = \Sigma y; \\ a_0 \Sigma \frac{1}{x} + a_1 \Sigma \frac{1}{x^2} = \Sigma \frac{1}{x} y. \end{cases}$$

Многофакторные модели связи

$(x_1, x_2, \ldots, x_k)$

- прямолинейная

$$\bar{y}_{1,2,\ldots,k} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_k x_k;$$

- степенная

$$\bar{y}_{1,2,\ldots,k} = a_0 x_1^{a_1} x_2^{a_2} \ldots x_k^{a_k};$$

- показательная

$$\bar{y}_{1,2,\ldots,k} = e^{a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_k x_k};$$

- парабола

$$\bar{y}_{1,2,\ldots,k} = a_0 + a_1x_1^2 + a_2x_2^2 + \ldots + a_kx_k^2;$$

- гипербола

$$\bar{y}_{1,2,\ldots,k} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \ldots + \frac{a_k}{x_k};$$

- в стандартизованном масштабе

$$\bar{t}_{1,2,\ldots k} = \beta_1t_1 + \beta_2t_2 + \ldots + \beta_kt_k,$$

где- стандартизованные значения признаковх, $_2$ ^определяе-
$t_{1,2,\ldots,k}$
мые по формуле:

$$t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_i},$$

^{1}12 k ~ среднее значение стандартизованной переменной соответствующего результативного признака, полученного по уравнению регрессии;

Pi 2 k ~ стандартизованные коэффициенты регрессии.

Оценка существенности связи:

t-критерий Стьюдента

t = , t_{KD}:(a;v = n-k-l),

где a - заданный уровень значимости; v - число степеней свободы; п - объем совокупности; k — число факторных признаков; a_a - дисперсия коэффициента регрессии, определяется по формуле

0> -i ^{a}*~k'

где a^2 - дисперсия результативного признака.

АНАЛИТИЧЕСКИЕ ПОКАЗАТЕЛИ Коэффициент эластичности:

где x$_{;}$ - среднее значение соответствующего факторного признака; у - среднее значение результативного признака; a$_{;}$ - коэффициент регрессии при соответствующем факторном признаке.

Частный коэффициент детерминации:

где r$_{yx}$ - парный коэффициент корреляции между результативным и i-м факторным признаками;

P$_x$ - соответствующий коэффициент уравнения множественной регрессии в стандартизованном масштабе.

Q-коэффициент:

Проверка значимости уравнения регрессии:

- F - критерий Фишера

k+1

тт ^2

n - k - 1

где V_x - коэффициент вариации соответствующего факторного признака.

ПАРАМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ СВЯЗИ Линейный коэффициент корреляции:

г= xy-x-y

У|2 k ~ теоретические значения результативного признака, полученные по уравнению регрессии; п - объем совокупности; k - число факторных признаков в модели;

Теоретическое корреляционное отношение:

a^2-(

$$\sqrt{1-\frac{\sigma^2_{ост}}{\sigma^2}}$$

средняя ошибка аппроксимации

где a^2 - общая дисперсия результативного признака; a^2 ост - остаточная дисперсия, рассчитанная по формуле:

с
т
о
с
т
=

Множественный коэффициент корреляции:

$$9\quad R_{y/x_1x_2...x_k}=\sqrt{1-\frac{\sigma^2_{ост}}{\sigma^2}},$$

где $\sigma^2_{ост}$- остаточная дисперсия;

σ^2 - общая дисперсия результативного признака.

Совокупный (общий) коэффициент корреляции:

$$R_{y/x_1x_2}=\sqrt{\frac{r^2_{yx_1}+r^2_{yx_2}-2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1-r^2_{x_1x_2}}},$$

где r_{xy} - линейные коэффициенты парной корреляции, или

$$R_{y/x_1x_2...x_k}=\sqrt{\beta_1 r_{yx_1}+\beta_2 r_{yx_2}+...+\beta_k r_{yx_k}},$$

где r_{xy_i} - парные коэффициенты корреляции;

β_{x_i} - коэффициенты регрессии в стандартизованном масштабе.

НЕПАРАМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ СВЯЗИ

Коэффициент корреляции рангов Спирмена:

- нет связных рангов:

$$\rho_{x/y} = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)},$$

где d_i - разность рангов вариантов факторного и результативного признаков; п - число наблюдений (число пар рангов).

Связные ранги (упрощенная формула):

$$\rho_{x/y} = 1 - \frac{\sum_{i=1}^{n} d_i^2}{\frac{1}{6}(n^3 - n) - (T_x + T_y)},$$

где $T_{x,y} = \frac{1}{12}\sum_{j=1}^{k}(t_j^3 - t_j)$; t_j- число одинаковых рангов в j-м ряду.

Коэффициент корреляции рангов Кендалла:

- нет связных рангов:

$$\tau = \frac{2S}{n(n-1)},$$

где S - сумма разностей между числом последовательностей и числом инверсий по второму признаку; п - число наблюдений;

- есть связные ранги:

$$\tau = \frac{S}{\sqrt{\left[\frac{n(n-1)}{2} - V_x\right] \cdot \left[\frac{n(n-1)}{2} - V_y\right]}},$$

где $V_{x/y} = \frac{1}{2}\sum_{i=1}^{k} t_j(t_j - 1)$; t_j - число одинаковых рангов в j-м ряду.

Коэффициент конкордации:

- нет связных рангов:

$$W = \frac{12S}{m^2(n^3 - n)},$$

где т - количество факторов; п - число наблюдений;
S - отклонение суммы квадратов рангов от средней квадратов рангов;

- есть связные ранги:

$$W = \frac{S}{\frac{1}{12}m^2(n^3 - n) - m\sum_{j=1}^{m} T_j},$$

где t_j - количество связных рангов по отдельным показателям.

$$T_j = \frac{1}{12}\sum_{i=1}^{m}(t_j^3 - t_j);$$

Коэффициент ассоциации:

$$K_a = \frac{ad - bc}{ad + bc},$$

Частные коэффициенты корреляции:

где a, Ь, c, d - градации признаков.

Коэффициент контингенции:

$$K_k = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}.$$

Коэффициент взаимной сопряженности Пирсона:

$$K_n = \sqrt{\frac{\varphi^2}{1+\varphi^2}},$$

где φ^2 - показатель взаимной сопряженности; φ

- сумма отношений квадратов частот каждой клетки таблицы к произведению итоговых частот соответствующего столбца и строки. Вычитая из этой суммы 1, получаем величину:

$$\varphi^2 = \Sigma\frac{n_{xy}^2}{n_x \cdot n_y} - 1.$$

Коэффициент взаимной сопряженности Чупрова:

$$K_ч = \sqrt{\frac{\varphi^2}{\sqrt{(k_1 - 1)(k_2 - 1)}}},$$

где k_1 - число значений (групп) первого признака;

k_2 - число значений (групп) второго признака.

Биссериальный коэффициент корреляции:

$$r = \frac{|\bar{y}_2 - \bar{y}_1|}{\sigma_y} \cdot \frac{pg}{z},$$

где $\bar{y}_1, \bar{y}_2$ - средние в группах;

σ_y - среднее квадратическое отклонение фактических значений

признака от среднего уровня; p- доля первой группы; g - доля второй группы; z - табличные значения распределения в зависимости от p.

$$r_{1,2,\ldots,k} = \frac{r_{1,2,\ldots,k-1} - r_{1,k,3,\ldots,k-1} \cdot r_{2,k,3,\ldots,k-1}}{\sqrt{(1 - r^2_{1,k,3,\ldots,k-1})(1 - r^2_{2,k,3,\ldots,k-1})}};$$

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1 - r^2_{x_2y})(1 - r^2_{x_1x_2})}};$$

$$r_{yx_2/x_1} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1 - r^2_{x_2y})(1 - r^2_{x_1x_2})}},$$

где r - парные коэффициенты корреляции между указанными в индексе переменными. Выявление аномальных наблюдений

- Ня пснпне (т-стятигтик-ы

$$g_t = \frac{|y_i - \bar{y}|}{\sigma_y},$$

где y_i - уровни ряда динамики; y - средний уровень ряда;

σ_y - среднее квадратическое отклонение значений ряда от их среднего уровня.

- Методом Ирвина

$$\lambda_i = \frac{|y_i - y_{i-1}|}{\sigma_y},$$

АЛФАВИТНО-ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

где y_{i-1} - значение предыдущего уровня ряда динамики.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. *Адамов В. Е.* Факторный индексный анализ (методология и пробле мы). - М: Статистика, 1977.
2. *Айвазян С. А., Енюков И. С., МешалкинЛ. Д.* Прикладная статистика. Исследование зависимостей. — М.: Финансы и статистика, 1985.
3. *Алан Р.* Экономические индексы. - М.: Статистика, 1980.
4. *Андерсен Т.* Статистический анализ временных рядов. -М.: Мир, 1976. • 5. *Бешенее С. Л., Гурвич Ф. Г.* Математико-статистические методы экс пертных оценок. - М.: Статистика, 1980.
6. *Вайну Я. Я.* Корреляция рядов динамики. - М.: Статистика, 1977.
7. *Герчук Я. П.* Графические методы в статистике. - М.: Статистика, 1968.
8. *Громыко Г. Л.* Статистика- М.: Издательство МГУ, 1981.
9. *Джессен Р.* Методы статистических обследований/Под ред. Е. М. Четыркина; пер. с англ. Ю. П. Лукашина, Я. Ш. Паппэ. - М.: Финан сы и статистика, 1985.
10. *Джини К.* Средние величины. - М.: Статистика, 1970.
11. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ: В 2-х кн. Кн. 1. - М.: Финансы и статистика, 1986.
12. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ: В 2-х кн. Кн. 2. - М.: Финансы и статистика, 1987.
13. *Дюран Б., Оделл П.* Кластерный анализ. - М.: Статистика, 1977.
14. *Езекиэл Н., Фокс К. А.* Методы анализа корреляций и регрессий. М.: Статистика, 1966.
15. *Елисеева И. И., Юзбашев М. М.* Общая теория статистики: Учебник. -М.:ИНФРА-М, 1998.
16. *Ефимова М. Р., Петрова Е. В., Румянцев В. Н.* Общая теория статис тики: Учебник. - М.: ИНФРА-М, 1998.
17. *Иващенко Г. А., Кильдишев Г. С., Шмойлова Р. А.* Статистическое изучение основной тенденции и взаимосвязи в рядах динамики. Томск: Издательство Томского университета, 1985.
18. *Йейтс Ф.* Выборочный метод в переписях и обследованиях /Под ред. А. Г. Волкова: пер. с англ. И. М. Сонина. - М.: Статистика, 1976.
19. *КазинецЛ. С.* Темпы роста и структурные сдвиги в экономике. -М.: Экономика, 1981.
20. *Ковалевский Г. В.* Индексный метод в экономике. - М.: Финансы и статистика, 1989.
21. *Королев Ю. Г., Рабинович П. М., Шмойлова Р. А.* Статистическое мо делирование и прогнозирование: Учебное пособие. - М.: МЭСИ, 1985.
22. *Лившиц Ф. Д.* Статистические таблицы. — М.: Госстатиздат, 1958.
23. *Миллс Ф.* Статистические методы. - М.: Госстатиздат, 1958.
24. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности /Под ред. О. Э. Башиной, А. А. Спири на. - М.: Финансы и статистика, 1999.
25. *Пасхавер И. С.* Закон больших чисел и статистическая закономер ность. - М.: Статистика, 1972.
26. *Перегудов В. Н.* Теоретические вопросы индексного анализа. - М.: Госстатиздат, 1960.

27. Практикум по теории статистики /Под ред. проф. Р. А. Шмойловой М.: Финансы и статистика, 1998.
28. Статистический словарь /Под ред. М. А. Королева. - М.: Финансы и статистика, 1989.
29. *Суслов И. П.* Общая теория статистики: Учебник. - М.: Статистика, 1970.
30. Теория статистики: Учебно-практическое пособие для системы дистан ционного образования /Под ред. В. Г. Минашкина. - М.: МЭСИ, 1998.
31. *Тинтнер Г.* Введение в эконометрию. - М.: Статистика, 1965.
32. *ТорвейР.* Индексы потребительских цен: Методология и руководство. - М.: Финансы и статистика, 1993.
33. *Фестер Э., Ренц Б.* Методы корреляционного и регрессионного анализа. - М.: Финансы и статистика, 1983.
34. *Четыркин Е. М.* Статистические методы прогнозирования. - М.: Статистика, 1975.
35. *Шварц Г.* Выборочный метод /Под ред. И. Г. Виницкого, В. М. Ива нова: пер. с нем. Я. Ш. Паппэ. - М.: Статистика 1978.
36. *Юзбашев М. М., Манелля А. И.* Статистический анализ тенденций и колеблемости. - М.: Финансы и статистика, 1983
37. *Silver M. S.* Business statistics. - 2-nd edition/ - McGraw-Hill, 1997.
38. *Плошко В. Г., Елисеева И. И.* История статистики. М.: Финансы и статистика, 1990.
39. Популярный экономико-статистический словарь-справочник/Под ред. И. И. Елисеевой. -М.: Финансы и статистика, 1993.

ОГЛАВЛЕНИЕ

Учебник

Шмойлова Римма Александровна
Шувалова Елена Борисовна Глубокова Наде^кда Юрьевна и др.

ТЕОРИЯ СТАТИСТИКИ

Зав. редакцией *Л. А. Табакова*
Редактор *Е. В. Стадниченко*
Художественный редактор *Ю. И. Артюхов*
Технический редактор *И. В. Завгородняя*
Корректоры *Т. М. Колпакова, Г. В. Хлопцева*
Оформление художника *О. В. Толмачева*

ИБ № 3679 Лицензия ЛР № 010156 от 29.01.97

Подписано в печать 25.10.2000. Формат 60x88/16 Гарнитура «Тайме». Печать офсетная
Усл. п. л. 34,3. Уч.-изд. л. 32,31 Доп. тираж 5000 экз. Заказ 3448. «С» OH

Издательство «Финансы и статистика»
101000, Москва, ул. Покровка, 7 Телефон (095) 925-35-02; факс (095) 925-09-57
E-mail: mail@finstat.ru, Aff/>://www.fmstat.m

Великолукская городская типография
Комитета по средствам массовой информации и связям
с общественностью администрации Псковской области,
182100, г. Великие Луки, ул. Полиграфистов, 78/12
Тел./факс: (811-53) 3-62-95
E-mail: VTL@MART.RU