МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Московский государственный университет экономики,
статистики и информатики
Московский международный институт эконометрики,
информатики, финансов и права**

**В.Г. Минашкин
А.Б. Гусынин
Н.А. Садовникова
Р.А. Шмойлова**

Курс лекций по теории статистики

Москва, 2002

УДК 311
ББК 60.6
М 613

Минашкин В.Г. Гусынин А.Б. Садовникова Н.А. Шмойлова Р.А. Курс лекций по теории статистики. /Московский государственный университет экономики, статистики и информатики. - М., 2002. - 189с.

Пособие подготовлено коллективом преподавателей кафедры теории статистики и прогнозирования МЭСИ:
доц.Минашкин В.Г. - гл.3,5,6,10,11 введение, заключение;
доц.Гусынин А.Б. - гл.1,7;
доц.Садовникова Н.А. -гл.2,8;
проф.Шмойлова Р.А. - гл.4,9.
Под общей редакцией доц. Минашкина В. Г.

СОДЕРЖАНИЕ

Введение

Полная и достоверная статистическая информация является тем необходимым основанием, на котором базируется процесс управления экономикой. Принятие управленческих решений на всех уровнях - от общегосударственного или регионального и до уровня отдельной корпорации или частной фирмы - невозможно без должного статистического обеспечения. Именно статистические данные позволяют определить объемы валового внутреннего продукта и национального дохода, выявить основные тенденции развития отраслей экономики, оценить уровень инфляции, проанализировать состояние финансовых и товарных рынков, исследовать уровень жизни населения и другие социально-экономические явления и процессы.

Статистика - это наука, изучающая количественную сторону массовых явлений и процессов в неразрывной связи с их качественной стороной, количественное выражение закономерностей общественного развития в конкретных условиях места и времени.

Для получения статистической информации органы государственной и ведомственной статистики, а также коммерческие структуры проводят различного рода статистические исследования. Процесс статистического исследования включает три основные стадии: сбор данных, их сводка и группировка, анализ и расчет обобщающих показателей.

От того, как собран первичный статистический материал, как он обработан и сгруппирован в значительной степени зависят результаты и качество всей последующей работы. Недостаточная проработка программно-методологических и организационных аспектов статистического наблюдения, отсутствие логического и арифметического контроля собранных данных, несоблюдение принципов формирования групп в конечном итоге могут привести к абсолютно ошибочным выводам.

Не менее сложной, трудоемкой и ответственной является и заключительная, аналитическая стадия исследования. На этой стадии рассчитываются средние показатели и показатели распределения, анализируется структура совокупности, исследуется динамика и взаимосвязи между изучаемыми явлениями и процессами.

Используемые на всех стадиях исследования приемы и методы сбора, обработки и анализа данных являются предметом изучения **общей теории статистики**, которая является базовой отраслью статистической науки. Разработанная ею методология применяется в макроэкономической статистике, отраслевых статистиках (промышленности, сельского хозяйства, торговли и прочих), статистике населения, социальной статистике и в других статистических отраслях.

Данное учебное пособие предназначено для студентов, изучающих курсы «Элементарные методы анализа статистических данных», «Теория статистики» или «Общая теория статистики».

Глава 1. Статистическое наблюдение

1.1. Основные формы, виды и способы статистического наблюдения

Статистическое наблюдение представляет собой планомерный, научно организованный и, как правило, систематический сбор данных о явлениях и процессах общественной жизни путем регистрации заранее намеченных существенных признаков с целью получения в дальнейшем обобщающих характеристик этих явлений и процессов.

Например, при переписи населения специальные работники, привлеченные к ее проведению, по поручению органов статистики записывают для каждого жителя страны сведения о его поле, возрасте, семейном положении, образовании и др., а затем на основе этих сведений статистические органы определяют численность населения, его возрастную структуру, размещение по территории страны и многие другие показатели. Некоторые данные последних переписей населения и текущего учета приведены в табл. 1.1 и 1.2.

Статистическое наблюдение можно классифицировать по различным признакам. Один из вариантов такой классификации представлен на рис. 1.

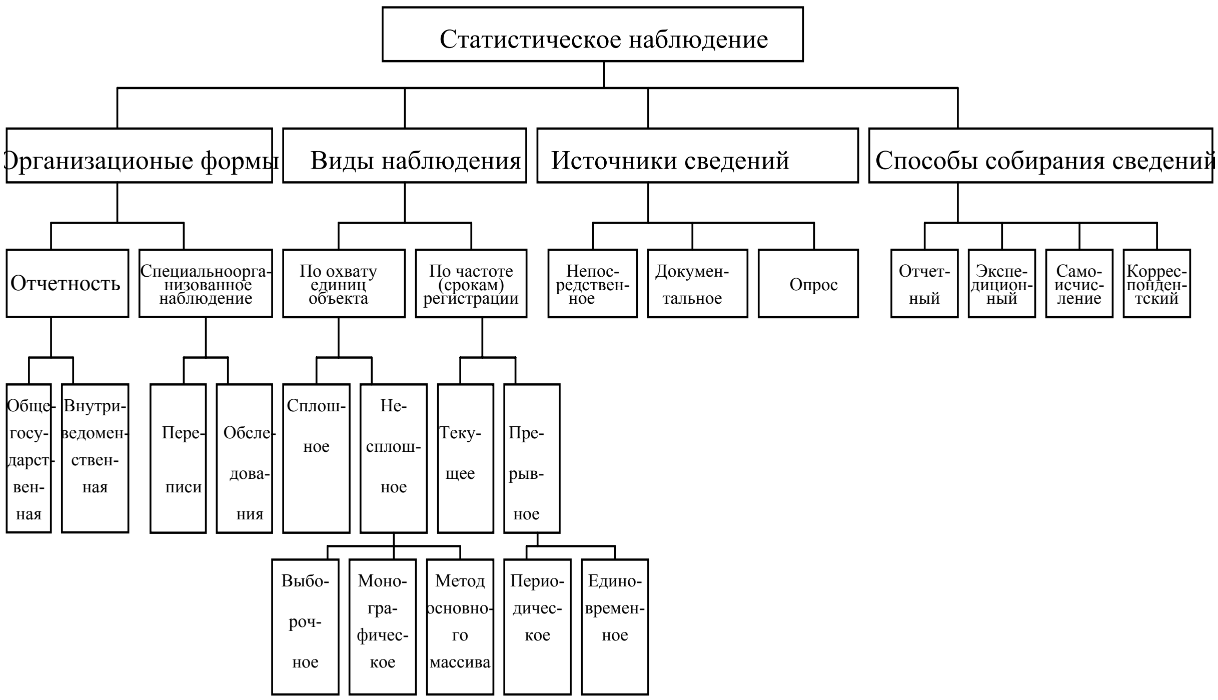


Рис.1. Классификация форм, видов и способов статистического наблюдения

Отчетностью называют такую организационную форму статистического наблюдения, при которой сведения поступают в статистические органы от предприятий, учреждений и организаций в виде обязательных отчетов об их деятельности.

В отчетах содержатся основные учетно-статистические данные о состоянии и деятельности предприятий, учреждений и организаций всех форм собственности эти данные необходимо для целей обобщения, контроля, анализа и прогнозирования, для оперативного руководства всеми субъектами, включенными в сферу действия отчетности.

Примерами такой формы статистического наблюдения могут служить уже упоминавшиеся переписи населения, а также бюджетные обследования домашних хозяйств, опросы общественного мнения и т.п.

Сплошным называется такое наблюдение, при котором обследованию подвергаются все без исключения единицы изучаемой совокупности (объекта наблюдения).

Примером такого наблюдения являются переписи, при которых по основной программе обследованию подлежит все без исключения население страны.

Таблица 1.1.

Численность населения Российской Федерации в 1970-1994 гг.[1]

Годы	Все население, млн. чел.	В том числе		В общей численности населения ,%	
		городское	сельское	городское	сельское
1970	130,1	81,0	49,1	62	38
1976	134,7	91,1	43,6	68	32
1979	137,6	95,4	42,2	69	31
1981	139,0	97,7	41,3	70	30
1986	143,8	104,1	39,7	72	28
1989	147,4	108,4	39,0	74	26
1990	148,0	109,2	38,8	74	26
1991	148,5	109,8	38,7	74	26
1992	148,7	109,7	39,0	74	26
1993	148,7	108,9	39,8	73	27
1994	148,4	108,5	39,9	73	27

[1] Российский статистический ежегодник. М., Госкомстат России, 1994, с.17

Таблица 1.2.

Распределение населения РФ по возрастным группам в 1970-1994гг.[1]

	1970	**1979**	**1989**	**1993**	**1994**
Все население	129941	137410	147022	148295	147997
в том числе в возрасте, лет: 0-4	9326	10523	12032	9759	8841
5-9	11975	9707	11360	12205	12079
10-14	13202	9512	10592	11103	11437
15-19	12290	12385	9968	10453	10613
20-24	9706	12995	9755	9710	9922
25-29	7102	11902	12557	10116	9657
30-34	11708	8016	12863	12818	12481
35-39	9327	8399	11684	12471	12721
40-44	10925	10485	7663	11178	11514
45-49	6698	9376	7955	6273	7359
50-54	5253	9716	9593	9201	7660
55-59	6874	5595	8399	8233	8983
60-64	5510	5065	8360	8387	7714
65-69	4181	5493	4510	6865	7326
70 и старше	5806	8200	9646	9559	9690

Специально организованное статистическое наблюдение представляет собой наблюдение, организуемое с какой-либо особой целью для получения данных, которые в силу тех или иных причин не собираются посредством отчетности, или для проверки, уточнения данных отчетности.

Несплошное - это такое наблюдение, при котором обследованию подвергаются не все единицы совокупности, а только часть их. В статистической практике применяется несколько видов несплошного наблюдения. Главными из них являются выборочное наблюдение, монографическое обследование и метод основного массива.

Выборочным называют наблюдение, основанное на принципе случайного отбора тех единиц изучаемой совокупности, которые должны быть подвергнуты наблюдению.

Выборочное наблюдение при правильной его организации и проведении дает достаточно достоверные данные для характеристики изучаемой совокупности в целом. Во многих случаях им вполне можно заменить сплошной учет. В условиях рыночной экономики сфера применения выборочного наблюдения постоянно расширяется.

[1] Российский статистический ежегодник. М., Госкомстат России, 1994, с.26

Монографическое обследование представляет собой детальное, глубокое изучение и описание отдельных, характерных в каком-либо отношении единиц совокупности.

Монографическое обследование проводится с целью выявления имеющихся или намечающихся тенденций в развитии явления, для выявления имеющихся резервов, изучения опыта отдельных субъектов рыночной экономики и т.п.

Метод основного массива заключается в том, что обследованию подвергаются наиболее крупные единицы, которые вместе взятые имеют преобладающий удельный вес в совокупности по основному для данного исследования признаку (признакам).

Единицы совокупности, обладающие незначительной величиной изучаемого признака, обследованию не подвергаются. Например, цены на продовольственных рынках могут регистрироваться лишь в крупных городах, где проживает большая часть населения России.

По частоте (срокам регистрации) наблюдение может быть непрерывным (текущим) и прерывным. Последнее, в свою очередь, подразделяется на периодическое и единовременное.

Текущим называют такое наблюдение, которое ведется непрерывно, и регистриция фактов производится по мере их свершения.
Пример такого наблюдения - регистрация актов гражданского состояния: рождений, смертей, браков, разводов.

Периодическое - это наблюдение, которое повторяется через определенные, равные промежутки времени. Таковым является, в частности, ежеквартальное представление финансовых отчетов в налоговые службы.

Единовременным называется такое наблюдение, которое проводится по мере необходимости, время от времени, без соблюдения строгой периодичности или вообще проводится один раз и больше не повторяется. Примером такого рода наблюдения может служить учет товарных остатков и денежной наличности на момент денежной реформы.

По источнику сведений различают непосредственное наблюдение, документальное наблюдение и опрос.

Непосредственным называют такое наблюдение, при котором сами регистраторы путем непосредственного замера, взвешивания или подсчета устанавливают факт и на этом основании производят запись в формуляре наблюдения. Таковы, например, инвентаризация имущества, снятие остатков товаров в магазине и др.

Документальное наблюдение предполагает запись ответов на вопросы формуляра на основании соответствующих документов. Примером такого наблюдения является сбор данных об успеваемости студентов вуза на основе зачетно - экзаменационных ведомостей.

Опрос - это наблюдение, при котором ответы на вопросы формуляра наблюдения записываются со слов опрашиваемого. Так проводятся, в частности, переписи населения. Основанием записи сведений о возрасте, семейном положении, образовании и т.д. служат ответы опрашиваемого.

В статистике применяются следующие способы собирания сведений: отчетный, экспедиционный, самоисчисление, анкетный, корреспондентский.

Сущность **отчетного** способа заключается, как уже отмечалось, в представлении предприятиями, учреждениями и организациями статистических отчетов о своей деятельности в строго обязательном порядке.

Экспедиционный способ наблюдения заключается в том, что специально привлеченные и обученные работники посещают каждую единицу наблюдения и сами заполняют формуляр наблюдения.

Этим способом собираются сведения при переписях населения.

При способе **самоисчисления (саморегистрации)** формуляры заполняют сами опрашиваемые.

Обязанность специально привлеченных для получения информации сотрудников состоит в раздаче формуляров опрашиваемым, инструктаже их, сборе заполненных формуляров и проверке правильности их заполнения. Способ саморегистрации применяется органами статистики, например, для изучения так называемой маятниковой миграции - передвижения населения от места жительства до места работы и обратно.

Анкетный способ - это сбор статистических данных с помощью специальных вопросников, рассылаемых определенному кругу лиц или публикуемых в периодической печати.

В современных условиях данный способ собирания сведений применяется очень широко, особенно в различных социологических обследованиях.

Сущность **корреспондентского** способа наблюдения заключается в том, что статистические органы договариваются с определенными лицами, которые берут на себя обязательство вести наблюдение за какими-либо явлениями, процессами и в установленные сроки сообщать результаты наблюдений статистическим органам.

Таким образом проводятся, в частности, экспертные оценки по конкретным вопросам социально - экономического развития страны.

1.2.Программно - методологические вопросы статистического наблюдения

При статистическом наблюдении необходимо прежде всего определить его объект и единицу.

Объектом статистического наблюдения называется та совокупность, о которой должны быть собраны нужные сведения.

Объектом наблюдения может быть, например, совокупность жителей страны, промышленных предприятий, крестьянских хозяйств, коммерческих банков, высших учебных заведений и т.п.

Единицей наблюдения называют тот составной элемент объекта наблюдения, который является носителем признаков, подлежащих регистрации.

Определение единицы наблюдения должно содержать указание ее важнейших отличительных признаков. Единицей наблюдения может быть человек, промышленное предприятие, фермерское хозяйство, коммерческий банк или вуз, в зависимости от того, какой объект подвергается наблюдению.

Единица наблюдения, как и объект в целом, обладают, как правило, множеством различных признаков. Все их учесть невозможно, а многие и не нужно. Поэтому при организации статистического наблюдения возникает вопрос о том, какие признаки следует регистрировать в процессе наблюдения.

Перечень признаков, регистрируемых в процессе наблюдения, называется **программой статистического наблюдения**.

К программе наблюдения предъявляется ряд требований, которым она должна отвечать при любом статистическом исследовании. Важнейшие из них - следующие:

1) программа должна содержать существенные признаки, по возможности непосредственно характеризующие изучаемое явление, его тип, основные черты, свойства;

2) в программу не следует включать второстепенные вопросы, так как они затрудняют работу по сбору информации, а в дальнейшем - по ее обработке и анализу;

3) разрабатывая программу, необходимо стремиться к полноте собираемых сведений;

4) в программу наблюдения должны включаться только такие вопросы, на которые действительно можно получить объективные и достаточно точные ответы;

5) в программу иногда следует включать вопросы контрольного характера, служащие целям проверки и уточнения собираемых сведений.

Одновременно с программой наблюдения должна составляться и программа разработки его материалов. Программа разработки конкретизирует задачи статистического наблюдения, она яснее показывает, какие именно данные следует собирать и в каком виде оформлять результаты их обработки, другими словами, она позволяет уточнить программу наблюдения.

Для записи ответов на вопросы программы конструируется формуляр наблюдения.

Формуляр наблюдения представляет собой особым образом разграфленный лист (листы) бумаги, в котором содержится перечень вопросов программы, свободные места для записи ответов на них, а также для записи шифров (кодов) ответов.

Обязательными элементами любого формуляра являются титульная и адресная части. **В титульной части** обычно содержится наименование статистического наблюдения, указывается наименование органа, проводящего наблюдение, кем и когда утвержден этот формуляр, иногда и номер, присвоенный ему в общей системе формуляров наблюдений, осуществляемых данным органом статистики. **В адресной части** предусматривается запись точного адреса единицы или совокупности единиц наблюдения и некоторые другие сведения о них (например, кому подчинено предприятие, сведения о котором записываются в формулярах наблюдения).

Во многих случаях в формулярах статистического наблюдения, кроме того, указывается, в какие сроки и в какие адреса должны высылаться заполненные формуляры, а также предусматриваются подписи лиц, ответственных за правильность содержащихся в них сведений (см. пример 1.2).

Обычно различают две системы статистического формуляра: индивидуальную (карточную) и списочную и в соответствии с этим выделяют бланк - карточку и бланк- список.

Бланком - карточкой называют формуляр, предназначенный для записи в нем ответов на вопросы программы только об одной единице наблюдения, а **бланком - списком** - о нескольких единицах наблюдения.

Приведенные в практических примерах формуляры наблюдения являются банками - карточками.

Формулировка вопросов бланка наблюдения имеет исключительно большое значение. Вопросы должны быть сформулированы как можно более кратко, ясно и определено, чтобы их понимание не вызывало затруднений и чтобы не возникала возможность разного их толкования (см. примеры).

Однако, какими бы ясными не казались вопросы формуляра, к нему обычно дается инструкция. **Инструкцией** называют совокупность разъяснений и указаний, главным образом по программе статистического наблюдения.

Инструкция может быть представлена в виде отдельного документа (часто - брошюры) или, как в приведенном практическом примере, изложена на формуляре наблюдения. Инструкцию следует писать кратко, просто, пояснения и указания должны быть ясными и четкими.

Иногда в формуляре после вопроса сразу же даются некоторые варианты возможных ответов на него. Перечень возможных ответов на поставленный вопрос называется **статистическим подсказом**. Если приводится исчерпывающий перечень возможных ответов на вопрос, подсказ называется **полным**, если же указываются только некоторые из возможных ответов - **неполным**. Пример формуляра с подсказом приведен ниже (анкета слушателя учебной фирмы).

При организации статистического наблюдения необходимо решить вопрос о времени данного наблюдения, включая выбор сезона, установления срока (периода) наблюдения, а в некоторых случаях и так называемого критического момента.

Период (срок) наблюдения - это время, в течение которого осуществляется регистрация единиц наблюдения по установленной программе.

Период наблюдения определяется многими факторами, важнейшими из которых являются следующие:

1) особенности объекта наблюдения (его размеры, состояние в то или иное время и т.д.);
2) объем и сложность программы наблюдения;
3) вид наблюдения по источнику сведений (см. раздел 1.1.).
4) наличие кадров, которые могут быть привлечены к проведению наблюдения, их количество и степень квалификации.

Критическим моментом статистического наблюдения (как правило, переписи) называется момент времени, по состоянию на который производится регистрация собираемых сведений.

В практической деятельности обычно стремятся к тому, чтобы процесс регистрации сведений был не слишком отдален от критического момента. Ведь чем дальше он будет отдален, тем больше произойдет изменений в объекте наблюдения и тем труднее будет восстановить состояние объекта в критический момент.

Срок наблюдения, как правило, обозначается указанием даты (иногда и часа) начала и окончания наблюдения. В некоторых случаях добавляется указание и числа дней, в течение которых оно должно быть проведено. Для некоторых статистических наблюдений устанавливается срок (день недели, число месяца или количество дней после критического момента), не позднее которого данные должны быть представлены по назначению. При этом чем короче период наблюдения, тем скорее должны быть представлены сведения по окончании этого периода.

Пример 1.1.

Анкета слушателя учебной фирмы

Анкета является конфиденциальным документом руководства отдела учебных фирм. Она предназначена для анализа учебного процесса на учебных фирмах и его совершенствования.
Благодарим Вас за участие в анкетировании.

Учебная группа ____

Просим Вас ответить на следующие вопросы:

1. Как Вы оцениваете продолжительность пребывания на учебной фирме:
- Долго
- Достаточно
- Мало

2. Считаете ли Вы целесообразной предложенную структуру прохождения практики на учебной фирме, включающую три этапа (регистрация фирмы, коммерческие операции, финансовый анализ):
   – Да – Нет

3. Оцените, пожалуйста, продолжительность каждого этапа практики:

1 этап	2 этап	3 этап
Долго	Долго	Долго
Достаточно	Достаточно	Достаточно
Мало	Мало	Мало

4. Ваши предложения по совершенствованию структуры и продолжительности практики на учебной фирме (расширить круг рассматриваемых вопросов, увеличить время на выполнение отдельных заданий, перераспределить время между этапами т.д.)

____
____
____

5. Оцените качество учебно - методических материалов:
- Хорошее
- Удовлетворительное
- Неудовлетворительное

Укажите свои замечания и предложения по доработке методических материалов ____
____

6. Оцените качество используемого программного обеспечения (формуляры на дискетах):
- Хорошее
- Удовлетворительное
- Неудовлетворительное

7. Укажите Ваши предложения по совершенствованию программного обеспечения

____

____

8. Оцените достаточность использования программных продуктов на учебной фирме:

– Достаточное
– Недостаточное

Ваши предложения______________________________________

____

9. Оцените организацию работы на учебной фирме:

– Хорошее
– Удовлетворительное

Ваши предложения______________________________________

____

10. Оцените уровень предварительной теоретической подготовки, необходимый для выполнения заданий на учебной фирме:

– Достаточный
– Недостаточный

1.3.Основные организационные вопросы статистического наблюдения

В целях успешного проведения наблюдения разрабатывается его организационный план.

Организационный план статистического наблюдения - это документ, в котором фиксируется решение важнейших вопросов подготовки и проведения статистического наблюдения с указанием конкретных сроков проведения намеченных мероприятий.

В организационном плане указываются:

1) объект наблюдения (дается его определение, описание, указываются отличительные признаки);
2) цели и задачи наблюдения;
3) органы наблюдения, осуществляющие подготовку и проведение наблюдения и несущие ответственность за эту работу;
4) время и сроки наблюдения;
5) подготовительные работы к наблюдению (в том числе порядок комплектования и обучения кадров, необходимых для проведения наблюдения);
6) порядок проведения наблюдения;
7) порядок приема и сдачи материалов наблюдения;

8) порядок получения и представления предварительных и окончательных итогов и др.

Организационные планы составляются разными звеньями системы статистических учреждений страны от высших до низших. В разных звенья объем и содержание планов несколько различаются. Так, в высших статистических органах главное внимание уделяется решению общих организационных вопросов, в оргпланах же низших звеньев на первое место выступает решение конкретных вопросов организации статистического наблюдения на местах.

Помимо постоянных органов, осуществляющих статистическое наблюдение, иногда, главным образом для проведения крупных обследований, создаются временные органы (бюро, управления, отделы, секторы и т.п.) соответствующих учреждений.

Место наблюдения - это место, где должна проводиться регистрация наблюдаемых фактов, где заполняются формуляры наблюдения.

Практический вопрос о месте наблюдения возникает лишь при специально организованном наблюдении, и только для объекта, единицы которого меняют или могут менять место своего пребывания (например, люди). Вопрос о месте наблюдения имеет существенное значение во многих социологических обследованиях. Если, например, необходимо получить некоторые сведения от студентов вуза путем их опроса, то необходимо установить, где этот опрос целесообразнее провести - в институте или в общежитии.

Иногда устанавливаются особые места для регистрации, и лица, обязанные сообщать соответствующие сведения, должны являться в эти места и сообщать их. В качестве примера таких мест наблюдения можно привести отделы записи актов гражданского состояния (ЗАГС). Хотя ЗАГСы и не статистические органы, именно через них органы статистики ведут наблюдение за естественным движением населения - рождаемостью, смертностью, брачностью и разводимостью.

1.4. Точность статистического наблюдения и ее контроль

Точностью статистического наблюдения называют степень соответствия значения какого-либо признака, найденного посредством статистического наблюдения, действительному его значению.

Точность характеризуется отношением и разностью данных наблюдения и действительных значений изучаемых величин. Расхождения между установленными статистическим наблюдением и действительными значениями изучаемых величин называются **ошибками наблюдения**. Они являются следствием неточностей при установлении и регистрации значений изучаемых признаков.

В зависимости от характера, степени влияния на окончательные результаты наблюдения, источников и причин возникновения неточностей различают несколько типов ошибок наблюдения.

Ошибки регистрации образуются вследствие неправильного установления фактов в процессе наблюдения, или ошибочной их записи, или того и другого вместе.

Случайными называют ошибки регистрации, которые возникают вследствие различных случайных причин.

Например, опрашиваемый мог оговориться, регистратор мог ослышаться или случайно переставить местами цифры, скажем, при записи возраста вместо 27 лет записать 72. При достаточно большом числе наблюдений благодаря действию закона больших чисел эти ошибки более или менее взаимно погашаются.

Систематические ошибки регистрации возникают под действием определенных причин. В каждом случае они действует в одном и том же направлении и приводят к серьезным искажениям общих результатов статистического наблюдения.

Примерами систематических ошибок регистрации при переписи населения могут служить случаи округления возраста населения, как правило, на цифрах, оканчивающихся на 5 и особенно - на 0. Систематические ошибки регистрации могут быть следствием сознательного, преднамеренного искажения фактов (например, в финансовой отчетности предприятий для сокрытия доходов от налогообложения).

Ошибки регистрации могут иметь место как при сплошном, так и при несплошном наблюдении.

Ошибки представительности (репрезентативности) свойственны только несплошному наблюдению. Отклонение величины изучаемого признака в отобранной для обследования части совокупности от его величины во всей совокупности называется **ошибкой представительности (репрезентативности).**

Случайные ошибки репрезентативности возникают в силу того, что совокупность отобранных на основе принципа случайности единиц наблюдения неполно воспроизводит совокупность в целом. Величина этой ошибки может быть оценена.

Систематические ошибки репрезентативности возникают в следствие нарушения принципа случайности отбора тех единиц изучаемой совокупности, которые должны быть подвергнуты наблюдению. Размеры этих ошибок обычно не поддаются количественному измерению.

По окончании наблюдения материалы, собранные в процессе его проведения, должны быть тщательно проверены. Проверка осуществляется с точки зрения а) полноты охвата объекта наблюдением и б) качества заполнения формуляров и других документов наблюдения. В последнем случае различают два вида контроля: логический и арифметический.

При контроле **полноты охвата объекта наблюдения** устанавливается, от всех ли единиц совокупности, подлежащих наблюдению, получены данные. Например, по истечении срока представления отчетности предприятиями города следует проверить, от всех ли подотчетных единиц наблюдения поступили необходимые данные. При специально организованном наблюдении контроль полноты материала возможен только в том случае, если заранее было известно, от кого или о ком (о чем), в крайнем случае - от какого числа единиц наблюдения должны были быть получены сведения . Если обнаружена неполнота охвата объекта наблюдением, дальнейшие действия зависят от того, представляется возможным восполнение пробелов или нет.

Логический контроль состоит в сопоставлении между собой ответов на вопросы формуляра наблюдения и выяснения их логической совместимости. При обнаружении логически несовместимых ответов пытаются путем дальнейших сопоставлений с ответами на другие вопросы или каким-либо иным путем установить, какой из ответов является неправильным.

Арифметический контроль состоит в проверке различных расчетов, результаты которых проведены в формуляре наблюдения, в частности, итогов, вычисления процентов, расчетов средних величин и т.п.

Пример 1.2

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

КОНФЕДЕНЦИАЛЬНОСТЬ ГАРАНТИРУЕТСЯ ПОЛУЧАТЕЛЕМ ИНФОРМАЦИИ

Код формы по ОКУД	Код отчитывающейся организации по ОКПО	Наименование отчитывающейся организации
0614426		Почтовый адрес

СВЕДЕНИЯ О РОЗНИЧНОЙ ТОРГОВЛЕ

за _______________ 199__г.

Форма №1 – торг

Представляют	Сроки представления
I. Юридические лица, их обособленные подразделения (предприятия розничной торговли, общественного питания, неторговые предприятия (организации), имеющие на своем балансе предприятия торговли и общественного питания) независимо от формы собственности (кроме отчитывающихся по формам: №№ МП и 1-ВЭС): статистическому органу по месту, установленному статистическим органом республики, края, области, автономного округа.	2 числа после отчетного периода
II. ОРСы (УРСы), продснабы статистическому органу по месту, установленному статистическим органом республики, края, области, автономного округа.	3 числа после отчетного периода
III. 1. Хозрасчетные предприятия и организации потребительской кооперации (кроме райпо) статистическому органу по месту, установленному статистическим органом республики, края, области, автономного округа; райпотребсоюзам, райпо и другим формированиям потребительской кооперации, в состав которых они входят.	2 числа после отчетного периода
2. Райпотребсоюзы, райпо и другие формирования потребительской кооперации статистическому органу по месту, установленному статистическим органом республики, края, области, автономного округа; потребсоюзам республик (в составе Российской Федерации), краев областей.	3 числа после отчетного периода
3. Потребсоюзы республик (в составе Российской Федерации), краев, областей, подразделения Центрсоюза статистическому органу республики в составе Российской Федерации, края, области по месту своего нахождения; Центросоюзу Российской Федерации.	4 числа после отчетного периода

0614426											
1	2	3	4	5	6	7	8	9	10	11	12
формы документа по ОКУД	отчитывающейся организации по ОКПО	отрасли по ОКОНХ	вида деятельности по ОКДП	территории по СОАТО	министерства (ведомства),	организационно-правовой формы по КОПФ	формы собственности по КФС				контрольной суммы (гр. 1-11)
Коды представляет отчитывающаяся организация											

Пояснения по заполнению сведений раздела I.

1 Сведения этого раздела представляют юридические лица, их обособленные подразделения независимо от форм собственности, для которых основным видом деятельности является розничная торговля или общественное питание.

2 Предприятия розничной торговли, осуществляющие, кроме основного вида деятельности, также оптовую торговлю, производственную деятельность, общественное питание и т.п., заполняют раздел I «Товарооборот и численность занятых» с учетом этих видов деятельности. При наличии сельскохозяйственной деятельности последняя не включается.

Обособленными подразделениями являются представительство и филиал юридического лица, которые расположены вне места его нахождения, наделены имуществом, выделены в самостоятельный баланс. Представительство и филиал действуют на основании доверенности и в соответствии с положением, утвержденным создавшим их юридическим лицом. Представительства и филиалы не являются юридическими лицами.

Сведения по статистике торговли обособленные подразделения представляют органам государственной статистики по месту своего нахождения.

3 Товарооборот - это общая сумма всех учетных продаж и услуг, а также выручка от комиссионных, прямых и транзитных операций (товар поступает от поставщика непосредственно к покупателю без промежуточного хранения на складе), комиссионных платежей и возмещений расходов, связанных с посреднической деятельностью (а не стоимость реализованных товаров).

В товарооборот не включается сконто (скидка в стоимости товара в случае уплаты за товар до наступления срока платежа), скидка с цены, годовые возмещения, бонус (дополнительная скидка, предоставляемая продавцом покупателю в соответствии с условиями сделки).

4. Численность занятых - все лица, которые заключили договор об оплате труда или обучении, включая работающих владельцев предприятия и безвозмездно помогающих членов семьи. Временно отсутствующие (напр. больные или находящиеся в отпуске) также включаются в этот показатель.

К занятым полное рабочее время относятся все лица, чье рабочее время соответствует рабочей неделе, принятой в данной местности, отрасли и заведении. Для занятых неполное рабочее время среднее время работы короче принятого в местности, отрасли и заведении.

5. По графе 2 объем товарооборота показывается за отчетный месяц, указываемый в графе 1.

6. Коды форм проявления розничной торговли и форм обслуживания покупателей проставляются по кодировке, сообщенной органами государственной статистики.

Код формы проявления розничной торговли	Код формы обслуживашия

I. Товарооборот и численность занятых

Отчетный месяц	Товарооборот включая НДС и акцизы млн. руб., код по СОЕИ - 0373	Численность занятых, включая владельцев, на конец отчетного месяца, человек, код по СОЕИ - 0792	
		занятые полное рабочее время	занятые неполное рабочее время
1	2	3	4

Возможные исправления за предыдущие месяцы
(Просьба указывать окончательный результат,
а не расхождения с сообщенными данными)

II. Розничный товарооборот и товарные запасы

№ п.п.	Наименование показателей	Розничный товарооборот	
		с начала отчетного года	за отчетный месяц
		млн. руб. код по СОЕИ-0373	
1	2	3	4
01	Всего (стр. 01=стр. 02+стр. 03)		
02	розничная торговля		
03	общественное питание		
04	Из строки 01 - продовольственные товары		
05	из них - алкогольные напитки		
06	Товарные запасы в предприятиях розничной торговли и общественного питания (на конец отчетного периода)	×	

Глава 2. Статистическая сводка и группировка

2.1.Задачи сводки и ее содержание

Важнейшим этапом исследования социально-экономических явлений и процессов является систематизация первичных данных и получение на этой основе сводной характеристики всего объекта при помощи обобщающих показателей, что достигается путем сводки и группировки первичного статистического материала

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

По глубине и точности обработки материала различают сводку простую и сложную.

Простая сводка - это операция подсчета общих итогов по совокупности единиц наблюдения.

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

Проведение сводки необходимо включает следующие этапы:

– выбор группировочного признака;
– определение порядка формирования групп;
– разработка системы статистических показателей для характеристики групп и объекта в целом;
– разработка макетов статистических таблиц для представления результатов сводки.

По форме обработки материала сводка бывает:

– централизованная, когда весь первичный материал поступает в одну организацию, подвергается в ней обработке от начала до конца;
– децентрализованная, когда отчеты предприятий сводятся статистическими органами субъектов РФ, а полученные итоги поступают в Госкомстат РФ и там определяются итоговые показатели в целом по народному хозяйству страны.

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

2.2.Виды статистических группировок

Группировкой называется расчленение единиц изучаемой совокупности на однородные группы по определенным существенным для них признакам. Группировки являются важнейшим статистическим методом обобщения статистических данных, основой для правильного исчисления статистических показателей.

С помощью метода группировок решаются следующие задачи:

- выделение социально-экономических типов явлений;
- изучение структуры явления и структурных сдвигов, происходящих в нем;

выявление связи и зависимости между явлениями.

В соответствии с этими задачами различают следующие виды группировок: типологические, структурные, аналитические.

Типологическая группировка - это расчленение разнородной совокупности на отдельные качественно однородные группы и выявление на этой основе экономических типов явлений. При построении группировки этого вида основное влияние должно быть уделено идентификации типов и выбору группировочного признака. Решение вопроса об основании группировки должно осуществляться на основе анализа сущности изучаемого явления.

Структурной называется группировка, которая предназначена для изучения состава однородной совокупности по какому-либо варьирующему признаку.

Группировка, выявляющая взаимосвязи между изучаемыми явлениями и их признаками, называется **аналитической группировкой.**

В статистике признаки делятся на факторные и результативные. Факторными называются признаки, под воздействием которых изменяются другие - результативные признаки. Взаимосвязь проявляется в том, что с возрастанием значения факторного признака систематически возрастает или убывает значение признака результативного.

Особенностями аналитической группировки являются то, что, во-первых, единицы группируются по факторному признаку и, во-вторых, каждая группа характеризуется средними величинами результативного признака.

Все рассмотренные группировки объединяет то, что единицы объекта разделены на группы по какому-либо признаку.

Группировка, в которой группы образованы по одному признаку называется **простой**.

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

Сначала группы формируются по одному признаку, затем группы делятся на подгруппы по другому признаку, а эти в свою очередь делятся по третьему и так далее. Таким образом, комбинационные группировки дают возможность изучить единицы совокупности одновременно по нескольким признакам.

При построении комбинационной группировки возникает вопрос о последовательности разбиения единиц объекта по признакам. Как правило, рекомендуется сначала производить группировку по атрибутивным признакам, значения которых имеют ярко выраженные качественные различия.

2.3. Принципы построения статистических группировок и классификаций

Построение группировки начинается с определения состава группировочных признаков.

Группировочным признаком называется признак, по которому проводится разбиение единиц совокупности на отдельные группы. От правильного выбора группировочного признака зависят выводы статистического исследования. В качестве основания группировки необходимо использовать существенные, теоретически обоснованные признаки.

В основание группировки могут быть положены как количественные, так и качественные признаки. Первые имеют числовое выражение (объем торгов, возраст человека, доход семьи и т. д.), а вторые отражают состояние единицы совокупности (пол, семейное положение, отраслевая принадлежность предприятия, его форма собственности и т. д.).

После того, как определено основание группировки следует решить вопрос о количестве групп, на которые надо разбить исследуемую совокупность.

Число групп зависит от задач исследования и вида показателя, положенного в основание группировки, объема совокупности, степени вариации признака. Например, группировка предприятий по формам собственности учитывает муниципальную, федеральную и собственность субъектов федерации.

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

При небольшом объеме совокупности не следует образовывать большого количества групп, так как группы будут включать недостаточное число единиц объекта. Поэтому показатели, рассчитанные для таких групп, не будут представительными и не позволят получить адекватную характеристику исследуемого явления.

Часто группировка по количественному признаку имеет задачу отразить распределение единиц совокупности по этому признаку. В этом случае количество групп зависит, в первую очередь, от степени колеблемости группировочного признака: чем больше его колеблемость, тем больше можно образовать групп. Чем больше групп, тем точнее будет воспроизведен характер исследуемого объекта. Однако, слишком большое число групп затрудняет выявление закономерностей при исследовании социально-экономических явлений и процессов. Поэтому в каждом конкретном случае при определении числа групп следует исходить не только из степени колеблемости признака, но и из особенностей объекта и цели исследования.

Определение числа групп можно осуществить и математическим путем с использованием формулы Стерджесса:

$$n = 1 + 3{,}322 \times \lg N, \qquad (2.1)$$

где n - число групп
N - число единиц совокупности.

Согласно этой формуле 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 если распределение единиц по признаку, положенному в основание группировки, близко к нормальному.

Другой способ определения числа групп основан на применении показателя среднего квадратического отклонения (σ). Если величина интервала равна 0.5σ, то совокупность разбивается на 12 групп, а когда величина интервала равна $2/3\sigma$ и σ, то совокупность делится, собственно, на 9 и 6 групп. Однако, при определении групп данными методами существует большая вероятность получения «пустых» или малочисленных групп.

Когда определено число групп, то 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 интервале. Величина интервала представляет собой разность между верхней и нижней границами.

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 равномерный характер, то строят группировку с **равными интервалами.**

Величина равного интервала определяется по следующей формуле:

$$h = \frac{R}{n} = \frac{x_{\max} - x_{\min}}{n} \qquad (2.2)$$

где x_{max}, x_{min} - максимальное и минимальное значения признака в совокупности;

n -число групп.

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

Полученную по формуле (2.2) величину округляют и она будет являться шагом интервала.

Существуют следующие правила определения шага интервала.

Если величина интервала, рассчитанная по формуле (2.2) представляет собой величину, которая имеет один знак до запятой (например: 0,88; 1,585; 4,8), то в этом случае полученные значения целесообразно округлить до десятых и их использовать в качестве шага интервала. В приведенном выше примере это будут соответственно значения: 0,9; 1,6; 4,7.

Если рассчитанная величина интервала имеет две значащие цифры до запятой и несколько после запятой (например 15,985), то это значение необходимо округлить до целого числа (до 16).

В случае, когда рассчитанная величина интервала представляет собой трехзначное, четырехзначное и так далее число, то эту величину следует округлить до ближайшего числа, кратного 100 или 50. Например, 557 следует округлить до 550 или до 600.

Если размах вариации признака в совокупности велик и значения признака варьируют неравномерно, то надо использовать группировку с неравными интервалами.

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

$$h_{i+1} = h_i + a,$$

а в геометрической прогрессии:

$$h_{i+1} = h_i \times q,$$

где а - константа: для прогрессивно возрастающих интервалов имеет знак «+«, и знак «-» - при прогрессивно - убывающих;

q - константа: больше «1» - для прогрессивно - возрастающих и меньше «1» - в другом случае.

Применение неравных интервалов обусловлено тем, что в первых группах небольшая разница в показателях имеет большое значение, а в последних группах эта разница не существенна.

Например, при построении группировки предприятий отрасли по показателю численности промышленно- производственного персонала, который варьирует от 200 человек до 2000 человек, нецелесообразно рассматривать равные интервалы, т. к. учитываются как малые, так и крупнейшие предприятия отрасли. Поэтому следует образовывать неравные интервалы: 200-500, 500-1100, 1100-2000, т. е. величина каждого последующего интервала больше предыдущего на 300 человек и увеличивается в арифметической прогрессии.

Интервалы группировок могут быть закрытыми и открытыми.

Закрытыми называются интервалы, у которых имеются верхняя и нижняя границы.

Открытые - это те интервалы, у которых указана только одна граница: верхняя - у первого, нижняя - у последнего. Например, группы коммерческих банков по числу работающих в них сотрудников (чел.): до 200, 200-300, 300-400, 400 и более.

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

Если основанием группировки служит непрерывный признак (например, группы строительных фирм по объему работ (млн. руб.): 1200-1400, 1400-1600, 1600-1800, 1800-2000), то одно и то же значение признака выступает и верхней и нижней границами двух смежных интервалов. В данном случае объем работ 1400 млн. руб. составляет верхнюю границу первого интервала и нижнюю границу второго, 1600 млн. руб. - соответственно второго и третьего и т. д.,т. е. верхняя граница i - го интервала равна нижней границе (i+1) - го интервала.

При таком обозначении границ может возникнуть вопрос, в какую группу включать единицы объекта, значения признака у которых совпадают с границами интервалов. Например, во вторую или третью группу должна войти строительная фирма с объемом работ 1600 млн. рублей? Если верхняя граница формируется по принципу «исключительно», то фирма должна быть отнесена к третьей группе, в противном случае - ко второй. Для того, чтобы правильно отнести к той или иной группе единицу объекта, значение признака которой совпадает с границами интервалов, можно использовать открытые интервалы (по нашему примеру группы строительных фирм по объему работ преобразуются в следующие: до 1400, 1400-1600, 1600-1800, 1800 и более). В данном случае, вопрос отнесения отдельных единиц совокупности, значения которых являются граничными, к той или иной группе решается на основе анализа последнего открытого интервала. Возможны два случая обозначения последнего открытого интервала: 1).1800 млн. руб. и более; 2). более 1800 млн. руб. В первом случае, строительные фирмы с объемом работ 1600 млн. руб. попадут в третью группу; во втором случае - во вторую группу.

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

При определении границ интервалов статистических группировок иногда исходят из того, что изменение количественного признака приводит к появлению нового качества. В этом случае граница интервала устанавливается там, где происходит переход от одного качества к другому.

Строя такую группировку, следует дифференцированно устанавливать границы интервалов для разных отраслей народного хозяйства. Это достигается путем использования группировок со специализированными интервалами. **Специализированные** - это такие интервалы, которые применяются 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** и, как правило, используются при группировке предприятий, например, по уровню рентабельности.

Пример.

Произведем анализ 30 самых надежных малых и средних банков РФ (на 1.04.96 г.) применяя метод группировок по следующим данным:

Номер банка	Капитал, млн. руб.	Рабочие активы, млн. руб.	Уставный фонд, млн. руб.
1	20780	11706	2351
2	19942	19850	17469
3	9273	2556	2626
4	59256	43587	2100
5	24654	29007	23100
6	47719	98468	18684
7	24236	25595	5265
8	7782	6154	2227
9	38290	79794	6799
10	10276	10099	3484
11	35662	30005	13594
12	20702	21165	8973
13	8153	16663	2245
14	10215	9115	9063
15	23459	31717	3572
16	55848	54435	7401
17	10344	21430	4266
18	16651	41119	5121
19	15762	29771	9998
20	6753	10857	2973
21	22421	53445	3415
22	13614	22625	4778
23	9870	11744	5029
24	24019	27333	6110
25	22969	70229	5961
26	75076	124204	17218
27	56200	90367	20454
28	60653	101714	10700
29	14813	18245	2950
30	41514	127732	12092

В качестве группировочного признака возьмем уставный фонд. Образуем четыре группы банков с равными интервалами. Величину интервала определим по формуле:

$$h = \frac{x_{max} - x_{min}}{n} = \frac{23100 - 2100}{4} = 5250$$

Обозначим границы групп:
2100-7350 - 1-я группа
7350-12600 - 2-я группа
12600-17850 - 3-я группа
17850-23100 - 4-я группа

После того как определен группировочный признак - уставный фонд, задано число групп - 4 и образованы сами группы, необходимо отобрать показатели, которые характеризуют группы и определить их величины по каждой группе. Показатели, характеризующие банки, разносятся по четырем указанным группам и подсчитываются групповые итоги. Результаты группировки заносятся в таблицу и определяются общие итоги по совокупности единиц наблюдения по каждому показателю.

Таблица 2.1

Группировка коммерческих банков по величине уставного фонда

Группы банков по величине уставного фонда, млн. руб.	**Число банков**	**Уставный фонд, млн. руб.**	**Работающие активы, млн. руб.**	**Капитал, млн. руб.**
А	1	2	3	4
2100-7350	18	71272	504898	342889
7350-12600	6	58227	343932	204694
12600-17850	3	48281	174059	130680
17850-23100	3	62238	217842	128573
Итого	30	240018	1240731	806836

Структурная группировка коммерческих банков на основе данных таблицы 2.1 будет иметь вид:

Таблица 2.2

Группировка коммерческих банков по величине уставного фонда (в %% к итогу).

Группы банков по величине ус-тавного фонда, млн. руб.	Число банков	Уставный фонд	Капитал	Работающие активы
2100-7350	60,0	40,7	42,5	29,7
7350-12600	20,0	27,7	25,4	24,3
12600-17850	10,0	14,0	16,2	20,1
17850-23100	10,0	17,6	15,9	25,9
Итого	100,0	100,0	100,0	100,0

Из таблицы 2.2 видно, что в основном преобладают малые банки - 60%, на долю которых приходится 42,5% всего капитала. Более конкретный анализ взаимосвязи показателей можно сделать на основе аналитической группировки.

Таблица 2.3

Группировка коммерческих банков по величине уставного фонда

Группы банков по величине уставного фонда, млн. руб.	Число банков	Капитал. млн.руб.		Работающие активы, млн. руб.	
		Всего	В среднем на один банк	Всего	В среднем на один банк
2100- 7350	18	342889	19049	504898	28050
7350-12600	6	204694	34116	343932	57322
12600-17850	3	130680	43560	174059	58020
17850-23100	3	128573	42858	217842	72614
Итого	30	806836	26895	1240731	41358

Величина капитала и работающие активы прямо зависят между собой и чем крупнее банк, тем эффективнее управление работающими активами.

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

Произведем группировку данных коммерческих банков по двум признакам: величине капитала и работающим активам.

Каждую группу и подгруппу охарактеризуем следующими показателями: число коммерческих банков, капитал, работающие активы.

Таблица 2.4

Группировка коммерческих банков по величине капитала и работающим активам.

Номер группы	Группы банков по величине капитала, млн. руб.	Подгруппы по величине работающих активов, млн. руб.	Число банков	Капитал, млн. руб.	Работающие активы, млн. руб.
1	**2**	**3**	**4**	**5**	**6**
1	до 24700	2600-33900 33900-65200 65200-96500 96500-12780	18 2 1 -	274577 39072 22969 -	325632 94564 70229 -
	Итого		21	336618	490425
2	24700-41600	2600-33900 33900-65200 65200-99100 99100-164300	1 1 - 1	35662 38290 - 41514	30005 79794 - 127732
	Итого		3	115466	237531
3	41600-58500	2600-33900 33900-65200 65200-99100 99100-164300	- 1 2 -	- 55848 103919 -	- 54435 188835 -
	Итого		3	159767	243270
4	58500-75400	2600-33900 33900-65200 65200-99100 99100-164300	- 1 - 2	- 59256 - 135729	- 43587 - 225918
	Итого		3	194985	269505
5	Всего по подгруппам	2600-33900 33900-65200 65200-99100 99100-164300	19 5 3 3	310239 192466 126888 177423	355637 272380 259064 353650
	Итого		30	806836	1240731

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

Отличительными чертами классификаций является то что: в основу их кладется качественный признак; они стандартны и устанавливаются органами государственной и международной статистики; они устойчивы, так как остаются неизменными в течение длительного периода времени.

Ряды распределения представляют собой простейшую группировку, в которой каждая выделенная группа характеризуется одним показателем.

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, то есть признакам, не имеющим числового выражения.

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

Вариационными рядами называют ряды распределения, построенные по количественному признаку. Любой вариационный ряд состоит из двух элементов: вариантов и частот. **Вариантами** называются отдельные значения признака, которые он принимает в вариационном ряду, то есть конкретное значение варьирующего признака. **Частотами** называются численности отдельных вариант или каждой группы вариационного ряда, то есть это числа, которые показывают, 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.

Дискретный вариационный ряд характеризует распределение единиц совокупности по дискретному признаку, принимающему только целые значения. Например, группы семей по числу детей (чел.): 1, 2, 3 и более.

В случае непрерывной вариации величина признака у единиц совокупности может принимать в определенных пределах любые значения, отличающиеся друг от друга на сколь угодно малую величину.

Построение **интервальных вариационных рядов** целесообразно прежде всего при непрерывной вариации признака, а также если дискретная вариация проявляется в широких пределах, то есть число вариантов дискретного признака достаточно велико.

Правила построения рядов распределения аналогичны правилам построения группировки.

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

Полигон используется при изображении дискретных вариационных рядов. Для его построения в прямоугольной системе координат по оси абсцисс в одинаковом масштабе откладываются ранжированные значения варьирующего признака, а по оси ординат наносится шкала для выражения величины частот. Полученные на пересечении абсцисс и ординат точки соединяются прямыми линиями, в результате чего получают ломаную линию, называемую полигоном частот. Иногда для замыкания полигона предлагается крайние точки (слева и справа на ломаной линии) соединить с точками на оси абсцисс, в результате чего получается многоугольник.

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

Гистограмма может быть преобразована в полигон распределения, если середины верхних сторон прямоугольников соединить прямыми.

При построении гистограммы распределения вариационного ряда с неравными интервалами по оси ординат наносят не частоты, а плотность распределения признака в соответствующих интервалах. Это необходимо сделать для устранения влияния величины интервала на распределение интервала и получения возможности сравнивать частоты. **Плотность распределения** - это частота, рассчитанная на единицу ширины интервала, то есть сколько единиц в каждой группе приходится на единицу величины интервала.

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

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

Если при графическом изображении вариационного ряда в виде кумуляты оси поменять местами, то получим **огиву.**

2.4.Сравнимость статистических группировок. Вторичная группировка

Группировки, построенные за один и тот же период времени, но для разных объектов или, наоборот, для одного объекта, но за два разных периода времени могут оказаться несопоставимыми из-за различного числа выделенных групп или неодинаковости границ интервалов.

Вторичная группировка, или перегруппировка сгруппированных данных применяется для: лучшей характеристики изучаемого явления (в случае, когда первоначальная группировка не позволяет четко выявить характер распределения единиц совокупности), либо для приведения к сопоставимому виду группировок с целью проведения сравнительного анализа.

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

Применяют два способа образования новых групп. Первым, наиболее простым и распространенным способом является изменение (чаще укрупнение) первоначальных интервалов. Второй способ получил название долевой перегруппировки и состоит в образовании новых групп на основе закрепления за каждой группой определенной доли единиц совокупности. Проиллюстрируем методику вторичной группировки на следующем примере.

Пример:

Распределение персонала строительной фирмы по уровню дохода (данные условные):

Группы работающих по уровню доходов, руб.	Число работающих
до 400	16
400 - 1000	20
1000 - 1800	44
1800 - 3000	74
3000 - 4000	37
4000 и более	9
Итого	200

Произведем перегруппировку данных, образовав новые группы с интервалами до 500, 500 - 1000, 1000 - 2000, 2000 - 3000, свыше 3000 руб.

В первую новую группу войдет полностью первая группа сотрудников и часть второй группы. Чтобы образовать группу до 500 руб., необходимо от интервала второй группы взять 100 руб. Величина интервала этой группы составляет 600 руб. Следовательно, необходимо взять от нее 1/6 (100:600) часть. Аналогичную же часть во вновь образуемую первую группу надо взять и от численности работающих, то есть $20 \times \frac{1}{6} = 3$ чел. Тогда в первой группе будет работающих: 16 + 3 = 19 чел.

Вторую новую группу образуют работающие второй группы за вычетом отнесенных к первой, то есть 20 - 3 = 17 чел. Во вновь образованную третью группу войдут все сотрудники третьей группы и часть сотрудников четвертой. Для определения этой части от интервала 1800 - 3000 (ширина интервала равна 1200) нужно добавить к предыдущему 200 (чтобы верхняя граница интервала была равна 2000 руб.). Следовательно, необходимо взять часть интервала, равную [200:1200 = 1:6]. В этой группе 74 человека, значит надо взять 74×(1:6) = 12 чел. В новую третью группу войдут 44 +12 = 56 чел. Во вновь образованную четвертую группу войдут 74 - 12 = 62 чел., оставшихся от прежней четвертой группы. Пятую вновь образованную группу составят работающие пятой и шестой прежних групп: 37 +9 = 46 чел.

В результате получим следующую новую группу:

Группы работающих по уровню доходов, руб.	**Число работающих**
до 500	19
500 - 1000	17
1000 - 2000	56
2000 - 3000	62
свыше 3000	46
Итого	200

2.5. Статистическая таблица и ее элементы

Результаты сводки и группировки материалов статистического наблюдения, как правило, представляются 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 так далее, могут носить табличную форму, но еще не являются статистическими таблицами.

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

Основные элементы статистической таблицы, составляющие как бы ее остов (основу), показаны на схеме 2.1.

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

Таким образом, внешне таблица представляет собой пересечение граф и строк, которые формируют остов таблицы.

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 макетом таблицы по центру и является внешним заголовком. Верхние заголовки характеризуют содержание граф (заголовки сказуемого), а боковые (заголовки подлежащего) - строк. Они являются внутренними заголовками.

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

Название таблицы
(общий заголовок)

<table>
<tr><td>Содержание строк</td><td colspan="6">Наименование граф (верхние заголовки)</td></tr>
<tr><td>А</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>...</td></tr>
<tr><td rowspan="3">Наименование строк (боковые заголовки)</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Итоговая строка</td><td></td><td></td><td></td><td></td><td></td><td>Итоговая графа</td></tr>
</table>

*) Примечания к таблице.

Схема 2.1. Остов (основа) статистической таблицы

Цифровой материал может быть представлен абсолютными (численность населения РФ), относительными (индексы цен на продовольственные товары) и средними (среднемесячный доход служащего коммерческого банка) величинами.

Таблицы могут сопровождаться примечанием, используемым с целью пояснения, в случае необходимости, заголовков, методики расчета некоторых показателей, источников информации и так далее.

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

Подлежащим статистической таблицы называется объект, который характеризуется цифрами. Это может быть одна или несколько совокупностей, отдельные единицы совокупности в порядке их перечня или сгруппированные по каким-либо признакам, территориальные единицы и так далее. Обычно подлежащее таблицы дается в левой части, в наименовании строк.

Сказуемое статистической таблицы образует система показателей, которыми характеризуется объект изучения, то есть подлежащее таблицы. Сказуемое формирует верхние заголовки и составляет содержание граф с логически последовательным расположением показателей слева направо.

Расположение подлежащего и сказуемого в отдельных случаях может меняться местами для более полного и лучшего способа прочтения и анализа исходной информации об исследуемой совокупности.

2.6. Виды таблиц по характеру подлежащего

В практике экономико-статистического анализа используются различные виды статистических таблиц.

В зависимости от структуры подлежащего, от группировки единиц в нем, различают статистические таблицы простые и сложные, а последние, в свою очередь, подразделяются на групповые и комбинационные.

Простой называется такая таблица, в подлежащем которой дается перечень каких-либо объектов или территориальных единиц.

Простые таблицы различают монографические и перечневые. Монографические таблицы характеризуют не всю совокупность единиц изучаемого объекта, а только одну какую-либо группу из нее, выделенную по определенному признаку (табл. 2.5).

Таблица 2.5.

Характеристика выпуска государственных краткосрочных облигаций в РФ в 1995г.

	Объем поданных заявок (шт.)	Объем выпуска (млн. руб.)		Доля ГКО, приобретенная сторонними инвесторами, %
		объявленный	реальный	
А	1	2	3	4
Государственные краткосрочные облигации	476354	295000	230569	34,7

Перестроив подлежащее таблицы 2.5, таким образом, чтобы были показаны ГКО по номерам, то есть, показав каждую единицу совокупности, получаем перечневую таблицу (см. табл. 2.6.).

Таким образом, простыми перечневыми таблицами называются таблицы, подлежащее которых содержит перечень единиц изучаемого объекта.

Таблица 2.6.

Характеристика выпусков государственных краткосрочных облигаций в РФ в 1995г.

Номер ГКО	Объем поданных заявок (шт.)	Объем выпуска (млн.руб.)		Доля ГКО, приобретенная сторонними инвесторами, %
		объявленный	реальный	
А	1	2	3	4
21003 RMFS7	40256	90000	37020	21,0
21004 RMFS5	164609	55000	49848	37,2
21005 RMFS2	271489	150000	143701	44,7
Всего	476354	295000	230569	34,5

Подлежащее простой таблицы может быть сформировано по видовому (например, табл 2.6); территориальному (например, численность населения по странам СНГ); временному и так далее принципам.

Простые таблицы не дают возможности выявить социально-экономические типы изучаемых явлений, их структуру, а также взаимосвязи и взаимозависимости между характеризующими их признаками.

Эти задачи более полно могут быть решены с помощью сложных - групповых 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 ряды распределения. Групповая таблица может быть более сложной, если в сказуемом дополнительно приводятся ряд показателей, характеризующих группы подлежащего. Такие таблицы часто используются в целях сопоставления обобщающих показателей по группам.

Таблица 2.7.

Распределение предприятий, выставивших акции на чековые аукционы РФ в 1995г. по величине уставного капитала

Группы предприятий по величине уставного капитала	Число предприятий	Количество акций (шт.)
1	2	3
1215-2340	14	7395
2340-3465	4	3402
3465-4590	4	4085
Итого	22	14882

Таблица 2.7 отражает количественное распределение предприятий, выставивших акции на чековых аукционах, по величине уставного капитала.

Таким образом, групповые таблицы позволяют выявить и охарактеризовать социально-экономические типы явлений, их структуру в зависимости только от одного признака.

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

Таблица 2.8.

Группировка предприятий, выставивших акции на чековые аукционы РФ в 1996г. по величине уставного капитала и числу занятых

Группы предприятий по величине уставного капитала (млн. руб.)	**Группы предприятий по числу занятых (чел.)**	**Число предприятий**	**Количество проданных акций (шт.)**
1	2	3	4
1235-2340	14-33 33-52 52-71	3 7 4	1206 4729 1390
Итого по группе	-	14	7325
2340-3465	14-33 33-52 52-71	3 - 1	2508 - 894
Итого по группе	-	4	3402
Итого по подгруппам	14-33 33-52 52-71	6 7 5	3714 4729 2284
Всего	-	18	10727

Подлежащим в таблице являются группы предприятий по величине уставного капитала и числу занятых.

Комбинационные таблицы позволяют характеризовать типические группы, выделенные по нескольким признакам и связь между ними. Последовательность разбиения единиц совокупности на однородные группы по признакам определяется либо важностью одного из них в их комбинации, либо порядком их изучения.

2.7. Виды таблиц по разработке сказуемого

В сказуемом статистической таблицы, как уже говорилось, приводятся показатели, которые являются характеристикой изучаемого объекта.

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 в 1996г.

Пред-приятия	Приобре-тено ак-ций (все-го)	в том числе			
		приватизи-рованные типа А	обыкно-венные	на льгот-ных ус-ловиях	по цене, опре-дел. Госкоми-муществом

После заполнения данного фрагмента таблицы получается подробная характеристика приватизированных предприятий по структуре их субъектов-владельцев. По каждому предприятию можно получить информацию о числе и ценовых условиях продажи акций.

Сложная разработка сказуемого предполагает деление признака, формирующего его, на подгруппы:

Пред-приятия	Приобре-тено ак-ций (все-го)	в том числе			
		На льготных усло-виях		По цене определенной Госкомимуществом	
		Привели-гирован-ные типа А	обык-новен-ные	Привели-гирован-ные типа А	обыкновенные

При этом получается более полная и подробная характеристика объекта.

Здесь оба признака сказуемого (ценовой и видовой) тесно связаны друг с другом. Можно проанализировать не только количество приобретенных акций по видам и условиям приобретения их сотрудниками приватизированных предприятий, но и определить число привилегированных и обыкновенных акций, приобретенных на разных ценовых условиях. То есть, при сложной разработке сказуемого явление или объект могут быть охарактеризованы различной комбинацией признаков, формирующих их.

Исследователь при построении статистических таблиц должен руководствоваться оптимальным соотношением показателей сказуемого.

2.8. Основные правила построения таблиц

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

Основными приемами, определяющими технику формирования статистических таблиц, являются следующие:

1. Таблица должна быть компактной и содержать только те данные, которые непосредственно отражают исследуемое явление в статике и динамике и необходимы для познания его сущности. Цифровой материал необходимо излагать таким образом, чтобы при анализе таблицы сущность явления раскрывалась чтением строк слева направо и сверху вниз;

2. Заголовок таблицы и названия граф и строк должны быть четкими, краткими, лаконичными, представлять собой законченное целое, органично вписывающееся в содержание текста. В названии таблицы должны найти отражение объект, признак, время и место совершения события. Например: «Курс доллара США на торгах ММВБ в 1996г.» Названия 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».

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

5. Графы и строки полезно нумеровать. Графы слева, заполненные названием строк, принято обозначать заглавными буквами алфавита (А), (В) и так далее, а все последующие графы - номерами в порядке возрастания.

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

7. Графы и строки должны содержать единицы измерения, соответствующие поставленным в подлежащем и сказуемом показателям. При этом используются общепринятые сокращения единиц измерения (чел., руб., кВт/ч и так далее).

8. Числа целесообразнее, по возможности, округлять. Округление чисел в пределах одной и той же графы или строки следует проводить с одинаковой степенью точности (до целого знака или до десятого и так далее).

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

9. Отсутствие данных об анализируемом социально-экономическом явлении может быть обусловлено различными причинами и это по-разному отмечается:

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

б) если по какой-либо причине отсутствуют сведения, то ставится многоточие «...» или «нет свед.»;

в) если отсутствует явление, то клетка заполняется тире (-). Для отображения очень малых чисел используют обозначения (0,0) или (0,00.

10. В случае необходимости дополнительной информации - разъяснений к таблице, могут даваться примечания.

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

2.9. Чтение и анализ таблицы

Анализу статистических таблиц предшествует этап ознакомления - чтения их.

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

Анализ предполагает реализацию двух его направлений - структурного и содержательного.

Структурный анализ предполагает анализ строения таблицы и характеристику представленных в ней:

- совокупности и единиц наблюдения, формирующих ее;
- признаков и их комбинации, формирующих подлежащее и сказуемое таблицы;
- признаков - количественные или атрибутивные;
- соотношение 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 соотношений и пропорций между группами явлений по одному и разным признакам; сравнительный анализ и формулировка выводов по отдельным группам и по всей совокупности в целом, установление закономерностей и определение резервов развития изучаемого объекта.

Прежде чем приступить к анализу числовой информации, необходимо проверить ее достоверность и научную обоснованность, источники ее получения. Должна быть произведена проверка данных: логическая (например, абсурдно, если численность работающих на фирме составила 106,7 чел.) и счетная - выборочный расчет отдельных значений признаков по группе, либо итоговых значений.

Анализ отдельных признаков и групп необходимо начинать с изучения абсолютных величин, затем - связанных с ними относительных величин.

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

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

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

Глава 3. Абсолютные и относительные статистические показатели

3.1. Классификация 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 финансовых средств, 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 и уровень его квалификации, производительность труда, фондовооруженность и другие.

В отличие от признака, статистический показатель получается расчетным путем. Это может быть простой подсчет единиц совокупности, суммирование их значений признака, сравнение двух или нескольких величин или более сложные расчеты.

Различают конкретный статистический показатель и показатель-категорию. **Конкретный статистический показатель** характеризует размер, величину изучаемого явления или процесса в данном месте и в данное время (под привязкой к месту понимается отношение показателя к какой-либо территории или объекту). Так, если мы называем конкретную величину стоимости промышленно-производственных фондов, то обязательно должны указать, к какому предприятию или отрасли и на какой момент 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 отличительный свойства конкретных статистических показателей одного и того же вида без указания места, времени и числового значения. Так, например, показатели розничного товарооборота предприятий торговли и общественного питания в г.Москве и г.С.-Петербурге в 1990г. и 1996г. 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: корпорацию, предприятие, цех, домохозяйство и т.п. Примером индивидуальных абсолютных показателей может служить численность промышленно-производственного персонала предприятия, объем реализованной продукции 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).

В отличие от индивидуальных **сводные показатели** характеризуют группу единиц, представляющую собой часть статистической совокупности или всю совокупность в целом. Эти показатели, в свою очередь, подразделяются на объемные и расчетные.

Объемные показатели получают путем сложения значений признака отдельных единиц совокупности. Полученная величина, называемая объемом признака, может выступать в качестве объемного абсолютного показателя (например, стоимость основных фондов предприятий отрасли), а может сравниваться с другой объемной абсолютной величиной (например, с численностью промышленно-производственного персонала этих предприятий) или объемом совокупности (в данном примере - с числом предприятий). В последних двух случаях получают объемный относительный и объемный средний показатели (в наших примерах - фондовооруженность и средняя стоимость основных фондов).

Расчетные показатели, вычисляемые по различным формулам, служат для решения отдельных статистических задач анализа - измерения вариации, характеристики структурных сдвигов, оценки взаимосвязи и т.д. Они также делятся на абсолютные, относительные и средние.

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 могут находить 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 Перми, соотношение численности детей дошкольного возраста и числа мест в детских дошкольных учреждениях и т.п.). Межобъектные показатели выражаются в форме относительных 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 только к какой-либо части территории или отдельному объекту.

В связи с тем, что классификация статистических показателей отличается многоплановостью, в дальнейшем изложении они сгруппированы по форме выражения. При этом в настоящей главе рассмотрены абсолютные и относительные показатели, которые являются формой выражения первичной информации и незаменимым инструментом элементарного анализа статистических данных. Показателям в форме средних величин будет уделено особое внимание в соответствующей главе.

3.2. Абсолютные показатели

Исходной, первичной формой выражения статистических показателей являются абсолютные величины. Статистические показатели в форме абсолютных величин характеризуют абсолютные размеры изучаемых статистикой процессов и явлений, а именно, их массу, площадь, объем, протяженность, отражают их временные характеристики, а также могут представлять объем совокупности, т.е. число составляющих ее единиц.

Индивидуальные абсолютные показатели , как правило, получают непосредственно в процессе статистического наблюдения как результат замера, взвешивания, подсчета и оценки интересующего количественного признака. В ряде случаев индивидуальные абсолютные показатели имеют разностный характер: разность между численностью зарегистрированных безработных в данном населенном пункте на конец и на начало года, разность между выручкой от реализации торгового предприятия и общей суммой затрат и т.п.

Сводные объемные показатели, характеризующие объем признака или объем совокупности 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, квадратные , кубические и простые метры, мили, километры, галлоны, литры, штуки и т.д. Например, производство электроэнергии в России в апреле 1996 г. составило 70,3 млрд. кВт·ч, в этом же месяце добыто 23,8 млн. т нефти и 51,7 млрд. куб. м газа.

В группу натуральных также входят условно-натуральные измерители, использумые в тех случаях, когда какой-либо продукт имеет несколько разновидностей и общий объем можно определить только исходя из общего для всех разновидностей потребительского свойства. Так, различные виды органического топлива переводятся в условное топливо с теплотой сгорания 29,3 МДж/кг (7000 ккал/кг), мыло разных сортов - в условное мыло с 40%-ным содержанием жирных кислот, консервы различного объема - в условные консервные банки объемом 353,4 куб.см и т.д.

Перевод в условные единицы измерения осуществляется на основе специальных коэффициентов, рассчитываемых как отношение потребительских свойств отдельных разновидностей продукта к эталонному значению. Так, например, 100 т торфа, теплота сгорания которого - 24 МДж/кг, будут эквивалентны 81,9 т условного топлива (100 * 24,0/29,3), а 100 т нефти при теплоте сгорания 45 МДж/кг будут оцениваться в 153,6 т условного топлива (100 * 45,0/29,3).

В отдельных случаях для характеристики какого-либо явления или процесса одной единицы измерения недостаточно, и используется произведение двух единиц. Примером этому могут служить такие показатели как грузооборот и пассажирооборот, оцениваемые соответственно в тонно-километрах и пассажиро-километрах, производство электроэнергии, измеряемое в киловатт-часах и т.д.

В условиях рыночной экономики наибольшее значение и применение имеют **стоимостные единицы измерения**, позволяющие получить денежную оценку социально-экономических явлений и процессов. Так, одним из важнейших стоимостных показателей в системе национальных счетов , характеризующим общий уровень развития экономики страны, является валовой внутренний продукт, который в России за 1 квартал 1996 года составил 508 трлн. рублей.

При анализе и сопоставлении стоимостных показателей необходимо иметь в виду, что в условиях высоких или относительно высоких темпов инфляции они становятся несопоставимыми. Так, сравнивать указанный выше ВВП России за 1 квартал 1996 года с его величиной за 1 квартал 1995 года вряд ли целесообразно, так как содержание рубля за этот период существенно 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.

3.3. Относительные показатели

Относительный показатель представляет собой результат деления одного абсолютного показателя на другой и выражает соотношение между количественными характеристиками социально-экономических процессов и явлений. Поэтому, по отношению к абсолютным показателям, относительные показатели или показатели в форме относительных величин являются производными, вторичными. Без относительных показателей невозможно измерить интенсивность развития изучаемого явления во времени, оценить уровень развития одного явления на фоне других взаимосвязанных с ним явлений, осуществить пространственно-территориальные сравнения, в том числе и на международном уровне.

При расчете относительного показателя абсолютный показатель, находящийся в числителе получаемого отношения, называется текущим или сравниваемым. Показатель же, с которым производится сравнение и который находится в знаменателе, называется основанием или базой сравнения. Таким образом, рассчитываемая относительная величина показывает, во сколько раз сравниваемый абсолютный показатель больше базисного, или какую 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 10000, то относительный показатель соответственно выражается в процентах (%), промилле (‰) и продецимилле (‱).

Проценты, как правило, используются в тех случаях, когда сравниваемый абсолютный показатель превосходит базисный не более, чем в 2-3 раза, или базисный показатель превосходит сравниваемый не более, чем в 100 раз (например, 174% или 5%). Проценты свыше 200-300 обычно заменяются кратным отношением, коэффициентом. Так, вместо 470% говорят, что сравниваемый показатель превосходит базисный в 4,7 раза. Иногда относительный показатель может быть выражен в процентах и в том случае, когда знаменатель превосходит числитель более чем в 100 раз (например, 0,3% или 0,08%). Однако, это будет целесообразно, если предполагается сравнение с другими относительными показателями, превышающими 1%.

Если базисный показатель превышает сравниваемый более чем в 100 раз, но не более чем в 1000 раз, удобно использовать промилле , т.е. тысячную долю числа. Промилле наиболее часто применяются в статистике населения для характеристики уровня рождаемости, смертности, брачности и т.п.

В отдельных случаях базисная величина может приниматься за 10000 или 100000 .Так, в расчете на 10000 человек населения рассчитывается численность студентов высших учебных заведений, численность врачей всех специальностей, в расчете на 100000 человек населения приводится заболеваемость населения различными болезнями.

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:

1) динамики;
2) плана;
3) реализации плана;
4) структуры;
5) координации;
6) интенсивности и уровня экономического развития;
7) сравнения.

Относительный показатель динамики (ОПД) представляет собой отношение уровня исследуемого процесса или явления за данный период времени (по состоянию на данный момент времени) к уровню этого же процесса или явления в прошлом:

$$\text{ОПД} = \frac{\text{Текущий уровень}}{\text{Предшествующий или базовый уровень}}$$

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

Различают относительные показатели динамики с постоянной и переменной базой сравнения. Если сравнение осуществляется с одним и тем же базисным уровнем, например, первым годом рассматриваемого периода, получают относительные показатели динамики с постоянной базой (базисные). При расчете относительных показателей динамики с переменной базой (цепных) сравнение осуществляется с предшествующим уровнем, т.е. основание относительной величины последовательно меняется.

Для примера воспользуемся данными таблицы 3.1.

Таблица 3.1

Производство сахара-песка в РФ в январе-апреле 1996г.

Месяц	Январь	Февраль	Март	Апрель
Объем производства, тыс.т	108	138	131	206

Рассчитаем относительные показатели динамики с переменной и постоянной базой сравнения:

<u>переменная база сравнения</u> (цепные показатели)	<u>постоянная база сравнения</u> (базисные показатели)
$\frac{138}{108}\cdot 100\% = 127{,}8\%$	$\frac{138}{108}\cdot 100\% = 127{,}8\%$
$\frac{131}{138}\cdot 100\% = 94{,}9\%$	$\frac{131}{108}\cdot 100\% = 121{,}3\%$
$\frac{206}{131}\cdot 100\% = 157{,}3\%$	$\frac{206}{108}\cdot 100\% = 190{,}7\%$

Относительные показатели динамики с переменной и постоянной базой сравнения взаимосвязаны между собой следующим образом: произведение всех относительных показателей с переменной базой равно относительному показателю с постоянной базой за исследуемый период. Так, для рассчитанных показателей (предварительно переведя их из процентов в коэффициенты) получим:

$$1{,}278\cdot 0{,}949\cdot 1{,}573 = 1{,}907$$

Относительные показатели плана и реализации плана . Все субъекты финансово-хозяйственной деятельности, от небольших индивидуальных частных предприятий и до крупных корпораций, в той или иной степени осуществляют как оперативное, так и стратегическое планирование , а также сравнивают реально достигнутые результаты с ранее намеченными. Для этой цели используются относительные показатели плана (ОПП) и реализации плана (ОПРП):

$$\text{ОПП} = \frac{\text{Уровень, планируемый на (i + 1) период}}{\text{Уровень, достигнутый в i - м периоде}}$$

$$\text{ОПРП} = \frac{\text{Уровень, достигнутый в (i + 1) периоде}}{\text{Уровень, планируемый на (i + 1) период}}$$

Первый из этих показателей характеризует напряженность плана, т.е. во сколько раз намечаемый объем производства превысит достигнутый уровень или сколько процентов от этого уровня составит. Второй показатель отражает фактический объем производства в процентах или коэффициентах по сравнению с плановым уровнем.

Предположим, оборот торговой фирмы в 1995г. составил 2,0 млрд.руб. Исходя из проведенного анализа складывающихся на рынке тенденций руководство фирмы считает реальным в следующем году довести оборот до 2,8 млрд.руб. В этом случае относительный показатель плана, представляющий собой отношение планируемой величины к фактически достигнутой, составит 140% $(\frac{2,8}{2,0} \cdot 100\%)$. Предположим теперь, что фактический оборот фирмы за 1996г. составил 2,6 млрд. руб. Тогда относительный показатель реализации плана, определяемый как отношение фактически достигнутой величины к ранее запланированной, составит 92,9% $(\frac{2,6}{2,8} \cdot 100\%)$.

Между относительными показателями плана, реализации плана и динамики существует следующая взаимосвязь:

$$\text{ОПП} \cdot \text{ОПРП} = \text{ОПД}$$

В нашем примере:

$$1{,}40 \cdot 0{,}929 = 1{,}3 \quad \text{или} \quad \frac{2{,}6}{2{,}0} = 1{,}3$$

Основываясь на этой взаимосвязи по любым двум известным величинам при необходимости всегда можно определить третью неизвестную величину.

Относительный показатель структуры представляет собой соотношение структурных частей изучаемого объекта и их целого:

$$\text{ОПС} = \frac{\text{Показатель, характеризующий часть совокупности}}{\text{Показатель по всей совокупности в целом}}$$

Выражается относительный показатель структуры в долях единицы или в процентах. Рассчитанные величины, соответственно называемые долями или удельными весами, показывают, какой долей обладает или какой удельный вес имеет та или иная часть в общем итоге.

Рассмотрим структуру валового внутреннего продукта РФ в 1 квартале 1996г. (табл. 3.2.):

Таблица3.2

Структура валового внутреннего продукта РФ в 1 квартале 1996г.

	Объем	
	трлн.руб.	% к итогу
ВВП - всего	508,0	100
в том числе:		
- производство товаров	185,4	36,5
- производство услуг	277,9	54,7
- чистые налоги на продукты	44,7	8,8

Рассчитанные в последней графе данной таблицы проценты представляют собой относительные показатели структуры (в данном случае -

удельные веса). Сумма всех удельных весов всегда должна быть строго равна 100% или 1.

Относительный показатель координации представляет собой отношение одной части совокупности к другой части этой же совокупности:

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

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, во сколько раз данная часть больше базисной или сколько процентов от нее составляет, или сколько единиц данной структурной части приходится на 1 единицу (иногда - на 100, 1000 и т.д. единиц) базисной структурной части. Так, на основе данных приведенной выше таблицы 3.2 мы можем вычислить, что на каждый триллион рублей произведенных товаров приходится 1,50 трлн.руб произведенных услуг ($\frac{277,9}{185,4}$) и 0,24 трлн.руб чистых налогов на продукты ($\frac{44,7}{185,4}$).

Относительный показатель интенсивности характеризует степень распространения изучаемого процесса или явления и представляет собой отношение исследуемого показателя к размеру присущей ему среды:

$$\text{ОПИ} = \frac{\text{Показатель, характеризующий явление А}}{\text{Показатель, характеризующий среду распространиния явления А}}$$

Данный показатель получают сопоставлением уровней двух взаимосвязанных в своем развитии явлений. Поэтому, наиболее часто он представляет собой именованную величину, но может быть выражен и в процентах, промилле, продецимилле.

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

Так., по данным социальной статистики на начало мая 1996г. численность граждан, состоящих на учете в службе занятости, составляла 3064 тыс. человек, а число заявленных предприятиями вакансий - 309 тыс. Отсюда следует, что на каждых 100 незанятых приходилось 10 свободных мест $(\frac{309}{3064} \cdot 100)$.

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 как объемные показатели производства продукции по своей природе являются интервальными, а показатель численности населения - моментным, в расчетах используют среднюю за период численность населения (предположим, среднегодовую).

Например, рассматривая лишь абсолютный размер ВВП России в 1 квартале 1996 года (508 трлн.руб), трудно оценить или "почувствовать" эту величину. Для того, чтобы на основе данной цифры сделать вывод об уровне развития экономики, необходимо сопоставить ее со среднеквартальной численностью населения страны (148,1 млн.чел), которая в простейшем случае рассчитывается как полусумма численности населения на начало и на конец квартала. В результате размер ВВП на душу населения составит $3{,}43 \text{ млн.руб} (\frac{508000 \text{млрд.руб.}}{0{,}1481 \text{млрд.чел.}})$

Относительный показатель сравнения представляет собой соотношение одноименных абсолютных показателей, характеризующих разные объекты (предприятия, фирмы, районы, области, страны и т.п.):

$$\text{ОПС} = \frac{\text{Показатель, характеризующий объект А}}{\text{Показатель, характеризующий объект В}}$$

Для выражения данного показателя могут использоваться как коэффициенты, так и проценты.

Например, известно, что на начало 1996г. операции с ГКО-ОФЗ проводили в Москве 108 официальных дилеров, в Новосибирске - 16, и в Санкт-Петербурге - 13. Таким образом в Москве дилеров было в 6,8 раза больше, чем в Новосибирске $(\frac{108}{16})$ и в 8,3 раза больше, чем в Санкт-Петербурге $(\frac{108}{13})$. Соответственно, число дилеров Новосибирска и Санкт-Петербурга составляло 14,8% и 12,0% от их числа в Москве.

Глава 4. Графическое изображение статистических данных

4.1. Понятие о статистическом графике. Элементы статистического графика

Статистический график - это чертеж, на котором статистические совокупности, характеризуемые определенными показателями, описываются с помощью условных геометрических образов или знаков.

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 делают более очевидными имеющиеся неточности, связанные либо с наличием ошибок наблюдения, либо с сущностью изучаемого явления. 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 должен быть соблюден ряд требований. Прежде всего, графики должны быть достаточно наглядными, так как весь смысл графического изображения как метода анализа в том и состоит, чтобы наглядно изобразить статистические показатели. Кроме того, график должен быть выразительным, доходчивым и понятным. Чтобы все эти требования выполнялись, каждый график должен включать ряд основных элементов: графический образ; поле графика; пространственные ориентиры; масштабные ориентиры; экспликацию графика.

Рассмотрим подробнее каждый из указанных элементов.

Графический образ (основа графика) — это геометрические знаки, то есть совокупность точек, линий, фигур, с помощью которых изображаются статистические показатели. Важно правильно выбрать графический образ, который должен соответствовать цели графика и способствовать наибольшей выразительности изображаемых статистических данных.

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

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

В практике графического изображения применяются также полярные координаты. Они необходимы для наглядного изображения циклического движения во времени. В полярной системе координат (рис. 4.1) один из лучей, обычно правый горизонтальный, принимается за ось координат, относительно которой определяется угол луча. Второй координатой считается ее расстояние от центра сетки, называемое радиусом. На статистических картах пространственные ориентиры задаются контурной сеткой (контуры рек, береговая линия морей и океанов, границы государств) и определяют те территории, к которым относятся статистические величины.

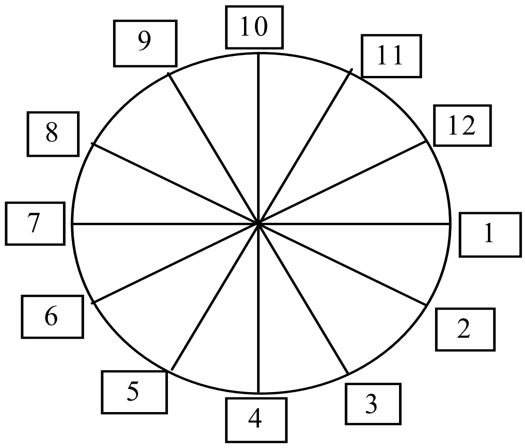


Рис. 4.1. Числовые интервалы

Масштабные ориентиры статистического графика определяются масштабом и системой масштабных шкал. Масштаб статистического графика - это мера перевода числовой величины в графическую. **Масштабной шкалой** называется линия, отдельные точки которой могут быть прочитаны как определенные числа. Шкала имеет большое значение в графике. В ней различают три элемента: линию (или носитель шкалы), определенное число помеченных черточками точек, которые расположены на носителе шкалы в определенном порядке, цифровое обозначение чисел, соответствующих отдельным помеченным точкам. Как правило, цифровым обозначением снабжаются не все помеченные точки, а лишь некоторые из них, расположенные в определенном порядке. По правилам числовое значение необходимо помещать строго против соответствующих точек, а не между ними (рис. 4.2).

Рис.4.2. Масштабная сетка

Носитель шкалы может представлять собой как прямую, так и кривую линию. В соответствии с этим различают шкалы прямолинейные (например миллиметровая линейка) и криволинейные - дуговые и круговые (циферблат часов).

Графические и числовые интервалы могут быть равными и неравными. Если на всем протяжении шкалы равным графическим интервалам соответствуют равные числовые, такая шкала называется **равномерной**. Если же равным числовым интервалам соответствуют неравные графические, и наоборот, - шкала называется **неравномерной**.

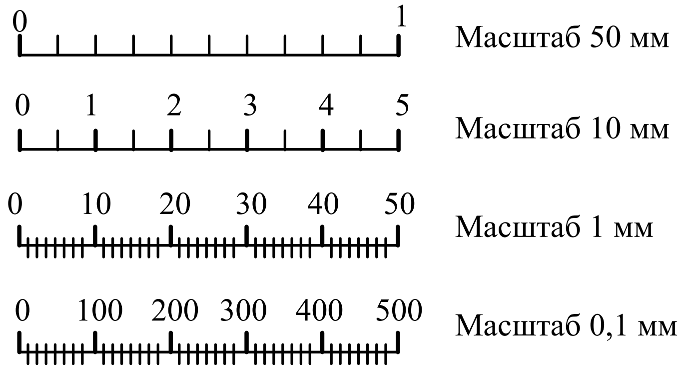


Рис. 4.3. Масштабы

Масштабом равномерной шкалы называется длина отрезка (графический интервал), принятого за единицу и измеренного в каких-либо мерах. Чем меньше масштаб (рис. 4.3), тем гуще располагаются на шкале точки, имеющие одно и то же значение. Построить шкалу - это значит на заданном носителе шкалы разместить точки и обозначить их соответствующими числами согласно условиям задачи. 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 $\lg 1 = 0$; $\lg 10 = 1$; $\lg 100 = 2$ и т.д. Для этих величин логарифмическая шкала может быть представлена так, как это сделано на рис. 4.4.

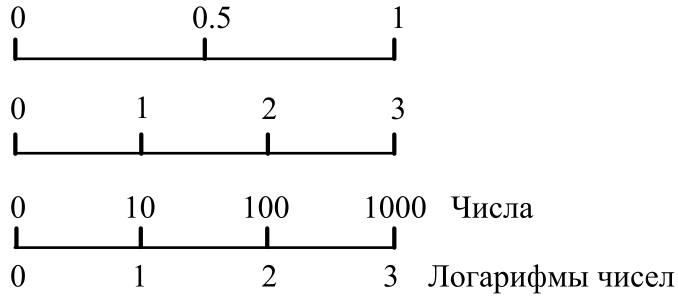


Рис.4.4. Шкалы

Последний элемент графика - **экспликация**. Каждый график должен иметь словесное описание его содержания. Оно включает в себя название графика, которое должно в краткой форме передавать его содержание, подписи вдоль масштабных шкал и пояснения к отдельным частям графика.

4.2. Классификация видов графиков

Существует множество графических изображений (рис. 4.6, 4.7). В основу их классификации может быть положен ряд признаков: а) способ построения графического образа; б) геометрические знаки, изображающие статистические показатели и отношения; в) задачи, решаемые с помощью графического изображения.

По способу построения статистические графики делятся на **диаграммы** и **статистические карты**. Диаграммы - наиболее распространенный способ графических изображений. Диаграммы применяются для наглядного сопоставления в различных аспектах (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. Они представляют собой условные изображения статистических данных на контурной географической карте, то есть показывают пространственное размещение и пространственную распространенность статистических данных.

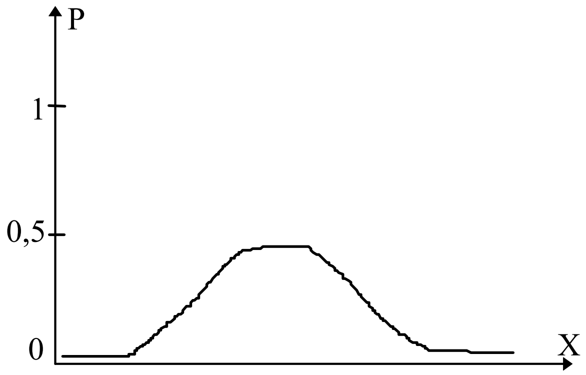
Геометрические знаки, как было сказано выше - это либо точки, либо линии или плоскости, либо геометрические тела. В соответствии с этим, различают графики точечные, линейные, плоскостные и пространственные (объемные) (рис. 4.5.).

При построении точечных диаграмм в качестве графических изображений применяются совокупности точек; при построении линейных - применяются 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, фигурные.

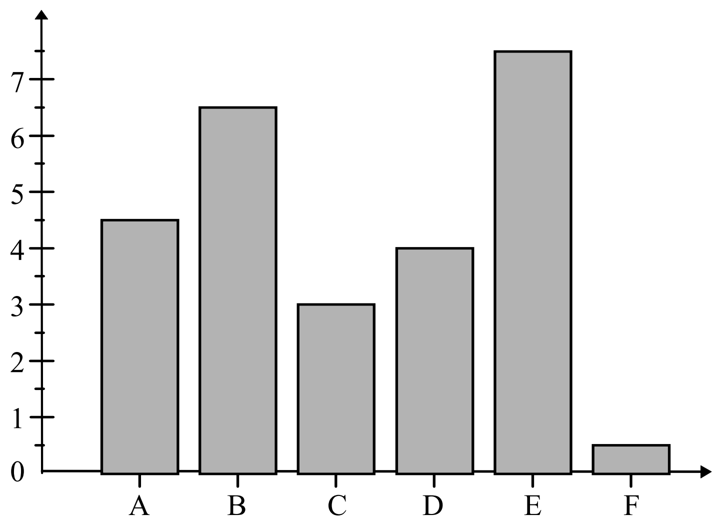
Статистические карты по графическому образу подразделяются на картограммы и картодиаграммы.

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

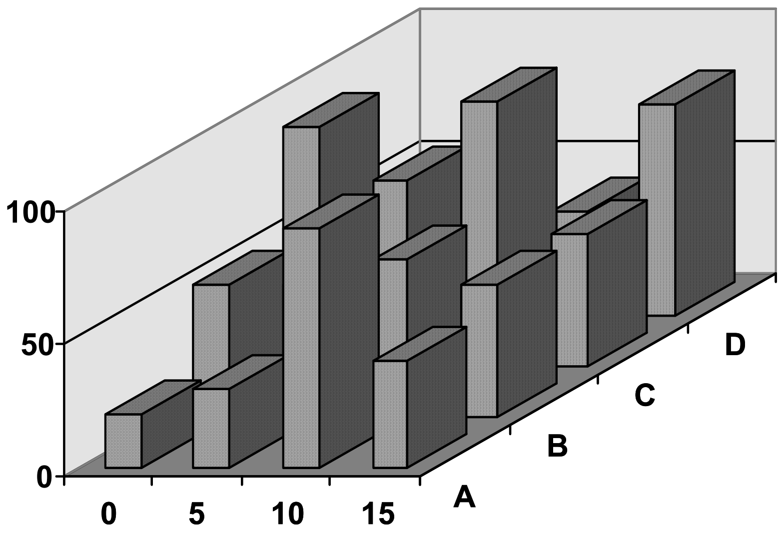
Особым видом графиков являются диаграммы распределения величин, представленных вариационным рядом. Это гистограмма, полигон, огива, кумулята.



а) линейные



б) плоскостные



в) объемные

Рис. 4.5. Виды диаграмм по форме геометрического образа

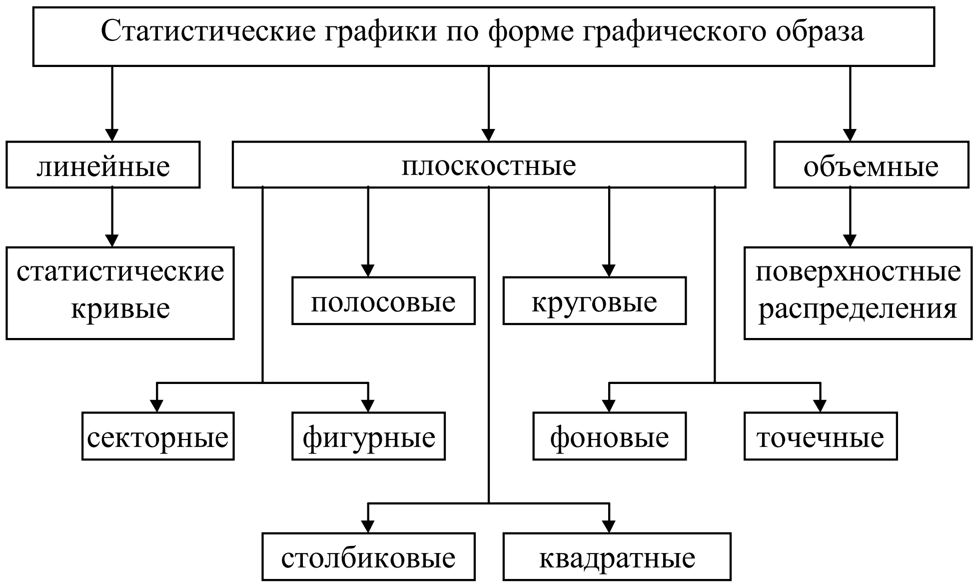


Рис. 4.6. Классификация статистических графиков по форме графического образа

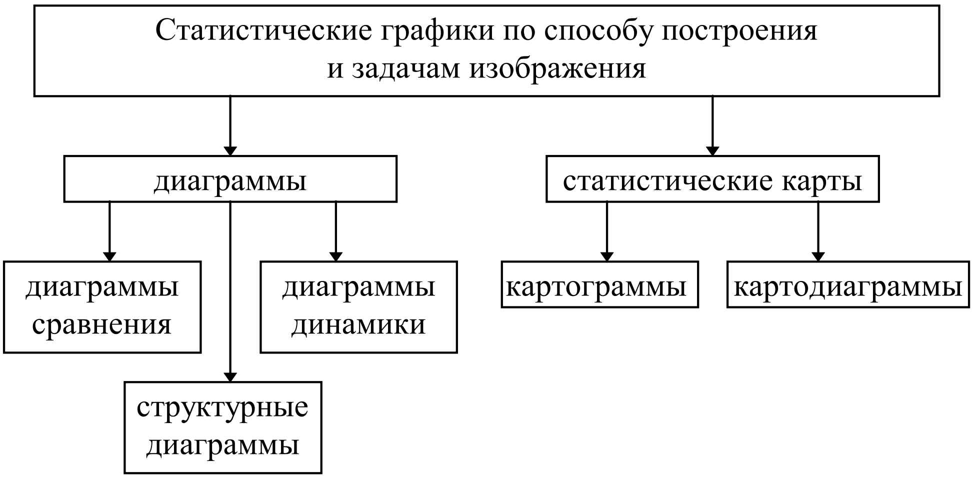


Рис. 4.7. Классификация статистических графиков по способу построения и задачам изображения

4.3. Диаграммы сравнения

Диаграммы сравнения применяются для графического отображения статистических данных с целью их наглядного сопоставления друг с другом в тех или иных разрезах.

Сравнительные диаграммы делятся на:
а) диаграммы простого сопоставления;
б) структурные диаграммы
в) изобразительные (фигур-знаков)

Диаграммы простого сопоставления дают наглядную сравнительную характеристику статистических совокупностей по какому-либо варьирующему признаку. При этом сопоставляемые совокупности и их части классифицируются по какому-либо атрибутивному или количественному признаку так, что отражаемый диаграммой статистический ряд представляет собой дискретный ряд цифр, на основе которого и строится график.

Диаграммы простого сопоставления между собой делятся на полосовые и столбиковые (см. рис. 4.8). Основной особенностью этих диаграмм является одномерность графического выражения величин варьирующего признака и их одномасштабность для различных столбцов или полос, характеризующих величину отражаемого признака в разных классификационных группах.

На **столбиковых** диаграммах статистические данные изображаются в виде вытянутых по вертикали прямоугольников (см. рис. 4.8а). Построение столбиковой требует применения вертикальной масштабной шкалы. Основания столбиков размещаются на горизонтальной линии, а высота столбиков устанавливается пропорционально изображаемым величинам. При построении столбиковых диаграмм необходимо выполнять следующие требования:

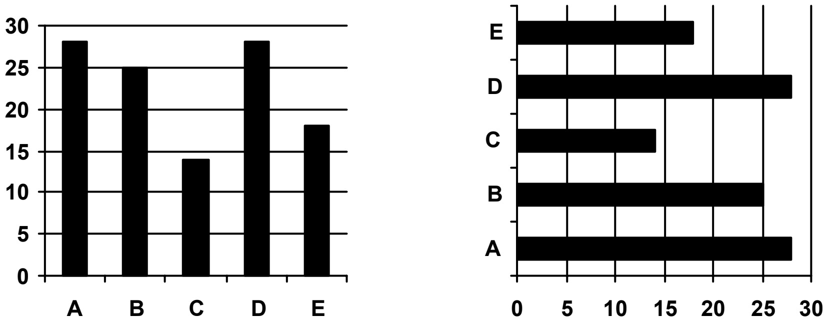
шкала, по которой устанавливается высота столбика должна начинать с нуля;

шкала должна быть непрерывной;

основания столбиков должны быть равны между собой;

наряду с разметкой шкалы соответствующими надписями следует снабжать сами столбцы.

Полосовые диаграммы (см. рис. 4.8б) состоят из прямоугольников, расположенных горизонтально. В этом случае масштабная шкала - горизонтальная ось. Принцип их построения тот же, что и в столбиковых.



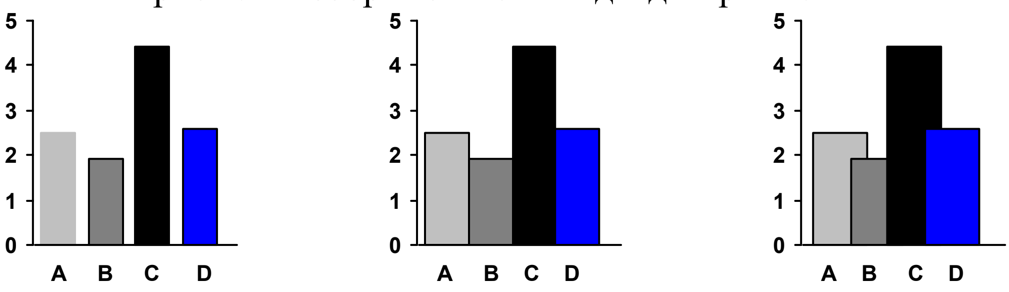
а) б)

Рис. 4.8. Пример построения столбиковой и полосовой диаграммы сравнения: а) столбиковая; б) полосовая.

На рисунке 4.8 изображены простейшие диаграммы сравнения. Сравниваются значения групп A, B, C, D, E. Вспомогательная сетка при построении диаграмм такого рода может быть опущена, она лишь помогает различать отклонения показателей разных групп друг от друга. Полосовые и столбиковые диаграммы являются однородными. Нетрудно заметить, что столбиковая диаграмма переходит в полосовую при повороте первой на 90 градусов. Выбор столбиковой или полосовой диаграммы в каждом конкретном случае равновозможен и обусловлен лишь эстетическими соображениями.

Размещение столбиков или полос в поле графика может быть различным: на одинаковом расстоянии друг от друга, вплотную друг к другу и в частичном наложении друг на друга.

На рис 4.8а изображены эти виды диаграмм.



а) б) в)

Рис. 8а. Пример расположения столбцов на диаграмме сравнения:

а) на одинаковом расстоянии;

б) вплотную;

в) с наплывом.

Более сложный вид принимают столбиковые и полосовые диаграммы при изображении на них статистических данных, показывающих какое-либо явление в нескольких разрезах. Такие диаграммы называются полосовыми или столбиковыми диаграммами с комбинированной группировкой показателей (см. рис. 4.9). Сходное назначение имеют диаграммы с подразделенными столбиками или полосами (см. рис. 4.10). Применение вышеуказанных диаграмм с группировкой по типам зависит от того, что наиболее важно в данных обстоятельствах подчеркнуть.

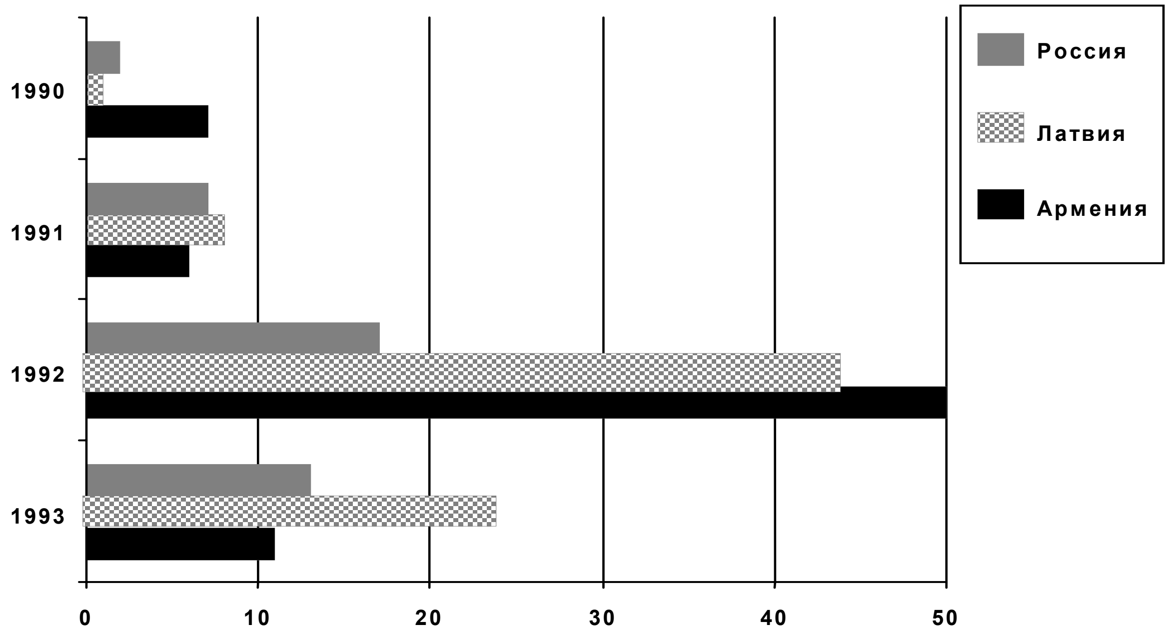


Рис.4.9. Уровень спада промышленного производства в трех странах бывшего СССР за период 1990-1993 гг. по сравнению с уровнем 1989 г. (базовый период 1989 г. = 100%).

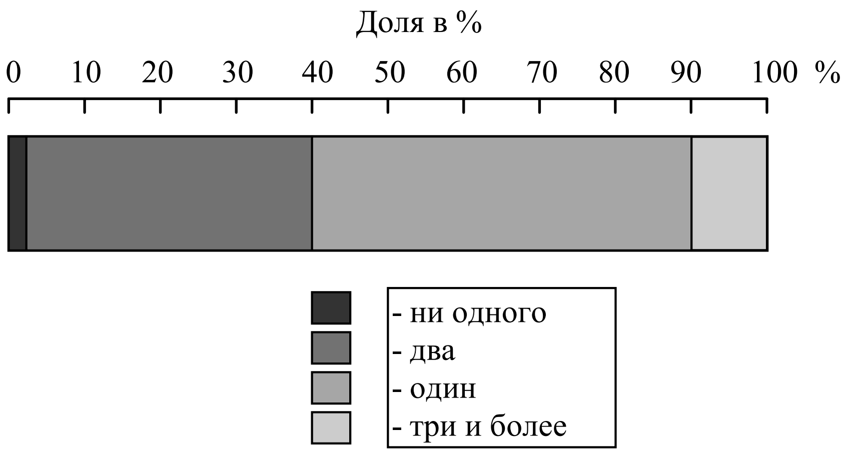


Рис. 4.10. Количество телевизоров в московской семье в 1993 г.

Для сопоставления изменяющихся во времени показателей, а также при сравнении величин, относящихся к одному и тому же периоду, могут использоваться **квадратные** и **круговые диаграммы**. В отличие от столбиковых или полосовых диаграмм они выражают величину изображаемого явления размером своей площади. Чтобы изобразить квадратную диаграмму, необходимо из сравниваемых статистических величин извлечь квадратные корни, а затем построить квадраты со сторонами, пропорциональными полученным результатам. Круговые диаграммы строятся аналогично. Разница состоит лишь в том, что на графике вычерчиваются круги, радиусы которых пропорциональны квадратному корню из изображаемых величин (рис.4. 11).

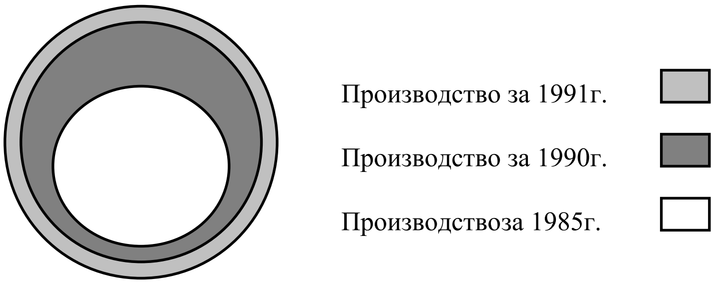


Рис. 4.11. Рост производства товаров народного потребления в г. Москве за 1985-1991 гг. (производство 1985 г. принято за единицу)

Показательные диаграммы прямого сопоставления статистических величин могут быть сделаны более выразительными, легче схватываемыми и запоминаемыми, если простые геометрические фигуры заменить символами, воспроизводящими в какой-то степени внешний образ отображаемых графиком статистических совокупностей или символизирующими их. **Изобразительные диаграммы** делятся на несколько типов.

Простейшей изобразительной диаграммой является такая, в которой в качестве графических знаков служат силуэтные изображения - символы сравниваемых статистических совокупностей, пропорциональные по своим размерам объемам этих совокупностей. Возражения против изобразительных диаграмм такого типа:

- отсутствие строгой соразмерности сравниваемых фигур;
- даже при точном соблюдении размерности величины отдельных знаков-символов отображаемым ими показателям диаграммы все равно оказываются маловыразительными;
- использование однородных фигур в расчете на их сравнение по одному условно-выбранному параметру.

Следующим типом изобразительных диаграмм являются диаграммы, в которых используются знаки-символы как масштабные знаки, как орудия счета. Диаграммы приятны для обозрения и легко запоминаются. В таких диаграммах часто приходится делить на части последний знак, так как по масштабу один знак является слишком крупной единицей измерения (это обычно делается на глаз).

Рассмотрим построение фигурной диаграммы по данным о числе фермерских хозяйств в одном из регионов России за 1994-1996 гг.:

Годы	1994	1995	1996
Тыс.	49	183	270

Примем условно за один знак 40 тыс. фермерских хозяйств. Тогда число хозяйств в 1994 году в размере 49 тыс. будет изображено в количестве 1,22 хозяйства, число в 1995 году - 4,6 хозяйства и так далее (рис. 4.12).

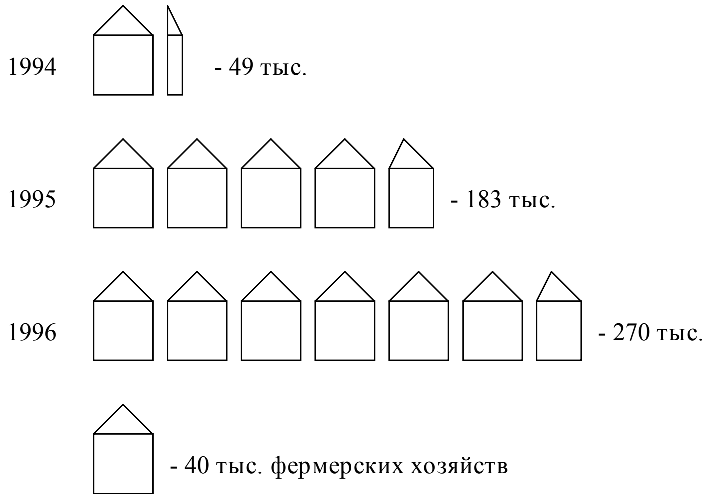


Рис. 4.12. Динамика фермерских хозяйств в одном из регионов России за 1994-1996 гг.

4.4. Диаграммы структуры

Вторую большую группу показательных графиков составляют структурные диаграммы. Это такие диаграммы, в которых отдельные статистические совокупности сопоставляются по их структуре, характеризующейся соотношением разных параметров совокупности или ее отдельных частей.

Простейшим видом структурных статистических диаграмм являются **диаграммы удельных весов**, отражающие структуры сравниваемых совокупностей по процентному соотношению в них отдельных частей, выделяемых по тому или иному количественному или атрибутивному признаку (рис. 4.10). Эти диаграммы получены путем преобразования простой полосовой диаграммы с подразделенными полосами. Полосовые диаграммы удельных весов могут вскрыть экономические существенные особенности многих изучаемых экономических явлений.

Другой широко распространенный метод графического изображения структур статистических совокупностей по соотношению удельных весов заключается в составлении структурных круговых или секторных диаграмм (рис. 4.13). **Секторные диаграммы** удобно строить следующим образом: вся величина явления принимается за сто процентов, рассчитываются доли отдельных частей в процентах. Круг разбивается на секторы пропорционально частям изображаемого целого. Таким образом, на 1% приходится 3,6 градуса. Для получения центральных углов секторов, изображающих доли частей целого, необходимо их процентное выражение умножить на 3,6 градуса. Секторные диаграммы позволяют не только разделить целое на части, но и сгруппировать отдельные части, давая как бы комбинированную группировку долей по двум признакам (см. рис. 4.13б).

Рассмотрим построения секторной диаграммы по данным представленным в таблице 4.1.

Таблица 4.1.

Количество телевизоров в московской семье в 1993 г.

Количество телевизоров	ни одного	один	два	три и более
Доля группы к итогу (%)	2	50	39	9

Построение секторной диаграммы начинается с определения центральных углов секторов. Для этого процентное выражение отдельных частей совокупности умножим на 3,6 градуса, т.е. $2 \times 3{,}6 = 7{,}2^\circ$; $50 \times 3{,}6 = 180^\circ$; $39 \times 3{,}6 = 140{,}4^\circ$; $9 \times 3{,}6 = 32{,}4^\circ$. По найденным значениям углов круг делится на соответствующие сектора (рис. 4.13а).

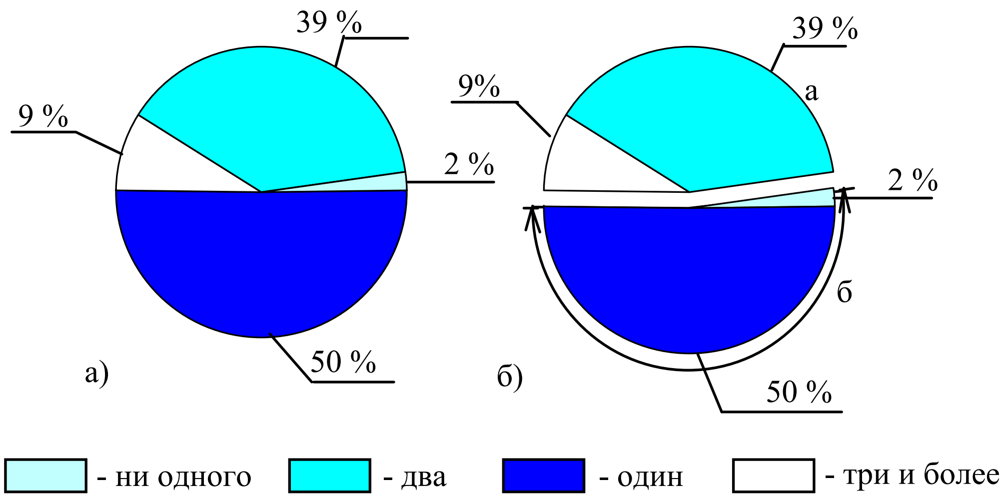


Рис. 4.13. Количество телевизоров в московской семье в 1993г.

На графике представлены два варианта структурной секторной диаграммы:

а) простая;

б) с группировкой долей;

Вариант б) помимо общего деления, показывает две специфические группы семей:

а) семьи, имеющие два телевизора и больше;

б) семьи, имеющие меньше двух телевизоров.

Такой тип диаграммы бывает удобен для выделения отдельных, наиболее типичных групп совокупности. Так в данном случае - это группа семей имеющих менее двух телевизоров.

Каждая доля (сектор, группа секторов), выделенная из круга, строится на биссектрисе общего угла доли, т.е. центр дуги этой доли принадлежит биссектрисе и находится на заданном расстоянии от общего центра диаграммы.

При большом числе долей, группировка дает хорошие результаты, позволяя лучше различать по своему весу нужные элементы совокупности.

Секторные диаграммы выглядят убедительно при существенных различиях сравниваемых структур, а при небольших различиях она может быть недостаточно выразительна. Значительным преимуществом полосовых структурных диаграмм по сравнению с секторными является их большая емкость, возможность отразить на небольшом пространстве большой объем полезной информации.

Для одновременного изображения трех величин, связанных между собой таким образом, что одна величина является произведением двух других, применяются диаграммы, называющиеся "**знаком Варзара**" (рис. 4.14). "Знак Варзара" представляет собой прямоугольник, у которого один сомножитель принят за основание , другой за высоту, а вся площадь равна произведению.

Оба показателя откладываются на шкалах (каждый на своей), третий (результат) изображается в виде прямоугольника в поле графика.

На рис. 4.14 средний размер вклада, умноженный на их число, дает общую сумму вкладов, что и отображается в виде площади (данные в центре прямоугольников, млрд. руб.).

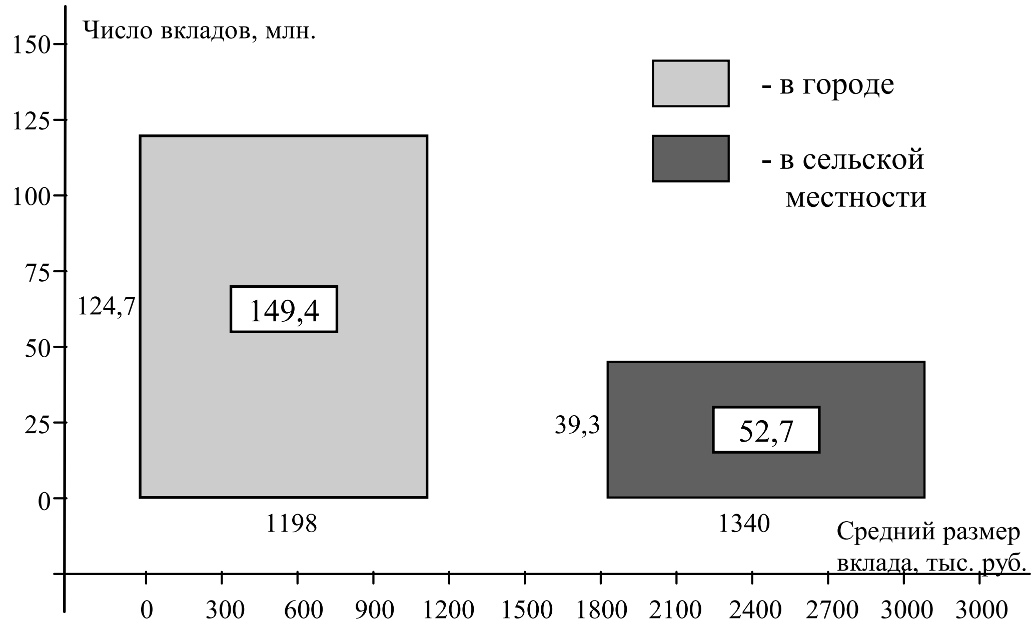


Рис. 4.14. Данные о вкладах населения в сберегательные банки региона в 1994 г.

4.5. Диаграммы динамики

Для изображения и внесения суждений о развитии явления во времени строятся диаграммы динамики. В рядах динамики используются для наглядного изображения явлений многие диаграммы: столбиковые, ленточные, квадратные, круговые, линейные, радиальные и другие. Выбор вида диаграмм зависит в основном от особенностей исходных данных, от цели исследования. Например, если имеется ряд динамики с несколькими неравноотстоящими уровнями во времени (1913, 1940, 1950, 1980,1985, 1995), 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, и если число уровней в ряду динамики велико, то целесообразно применять **линейные диаграммы**, которые воспроизводят непрерывность процесса развития в виде непрерывной ломаной линии. Кроме того, линейные диаграммы удобно использовать: когда целью исследования является изображение общей тенденции и характера развития явления; когда на одном графике необходимо изобразить несколько динамических рядов с целью их сравнения; когда наиболее существенным является сопоставление темпов роста, а не уровней.

Для построения линейных диаграмм используют систему прямоугольных координат. Обычно по оси абсцисс откладывается время (годы, месяцы и т.д.), а по оси ординат - размеры отображаемых 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 диаграмме, так как нарушение равновесия дает неправильное изображение развития явления. Если масштаб для шкалы на оси абсцисс очень растянут по сравнению с масштабом на оси ординат, то колебания в динамике явлений мало выделяются, и наоборот, преувеличение масштаба по оси ординат по сравнению с масштабом на оси абсцисс дает резкие колебания. Если в ряду динамики данные за некоторые годы отсутствуют, это должно быть уточнено при построении графика. Равным периодам времени и размерам уровня должны соответствовать равные отрезки масштабной шкалы.

Рассмотрим построение линейной диаграммы на основании данных таблицы 4.2.:

Таблица 4.2.

Динамика валового сбора зерновых культур в регионе за 1986 -1995 гг.

Годы	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
млн. тонн	237,4	179,2	189,1	158,2	186,8	192,2	172,6	191,7	210,1	211,3

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 остается неиспользованным и ничего не дает для выразительности изображения. Поэтому в данных условиях рекомендуется строить шкалу без вертикального нуля, то есть шкала значений разрывается недалеко от нулевой линии и на диаграмму попадает лишь часть возможного поля графика. Это не приводит к искажениям в изображении динамики явления и процесс его изменения рисуется диаграммой более четко (рис. 4.15).

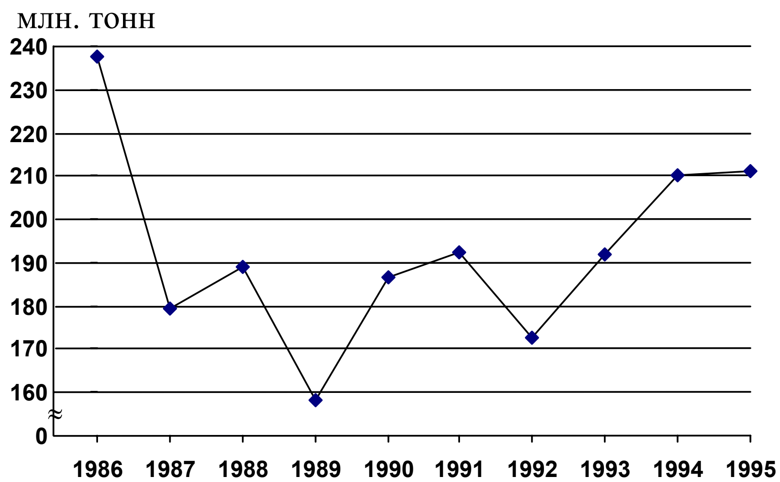


Рис. 4.15. Динамика валового сбора зерновых культур в регионе за 1986-1995 гг.

Нередко на одном линейном графике приводится несколько кривых, которые дают сравнительную характеристику динамики различных показателей или одного и того же показателя в разных странах. Примером графического изображения сразу нескольких показателей может служить рис. 4.16.

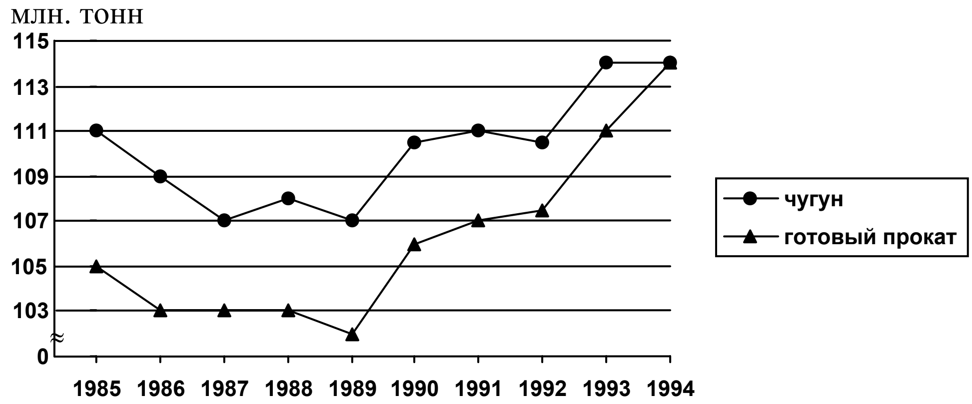


Рис. 4.16. Динамика производства чугуна и готового проката в регионе за 1985-1994 гг.

Линейные диаграммы с равномерной шкалой имеют один недостаток, снижающий их познавательную ценность. Этот недостаток заключается в том, что 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 изображении их на координатной диаграмме с равномерной вертикальной шкалой. Кроме того, в обычных координатах теряет всякую наглядность и даже становится невозможным изображение рядов динамики с резко изменяющимися уровнями, которые обычно имеют место в динамических рядах за длительный период времени.

В этих случаях следует отказаться от равномерной шкалы и положить в основу графика полулогарифмическую систему.

Полулогарифмической сеткой называется сетка, в которой на одной оси нанесен линейный масштаб, а на другой логарифмический. В данном случае логарифмический масштаб наносится на ось ординат, а на оси абсцисс располагают равномерную шкалу для отсчета времени по принятым интервалам (годам, кварталам, месяцам, дням и прочее). Техника построения логарифмической шкалы следующая (рис. 4.17): необходимо найти логарифмы исходных чисел; начертить ординату и разделить 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. Логарифмический масштаб лучше понять на примере.

Допустим нам надо изобразить на графике динамику производства электроэнергии в регионе за 1966 - 1995 гг., за эти годы оно возросло в 9,1 раза. С этой целью находим логарифмы для каждого уровня ряда (см. таблицу 4.3.)

Таблица 4.3.

Динамика производства электроэнергии в регионе за 1966 - 1995 гг. (млрд. кВт. ч.)

Годы	**Y_1**	**LgY_1**
1966	170	2,23
1970	292	2,46
1975	507	2,70
1980	741	2,84
1985	1039	3,02
1990	1294	3,11
1995	1544	3,19

Найдя минимальное и максимальное значения логарифмов производства электроэнергии, строим масштаб с таким расчетом, чтобы все данные разместились на графике. В соответствии с масштабом находим соответствующие точки, которые соединим прямыми линиями. В результате получим график (см. рис. 4.18) с использованием логарифмического масштаба на оси ординат. Поэтому он называется диаграммой на полулогарифмической сетке. Полной логарифмической диаграммой он будет 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. 4.17. Схема логарифмического масштаба

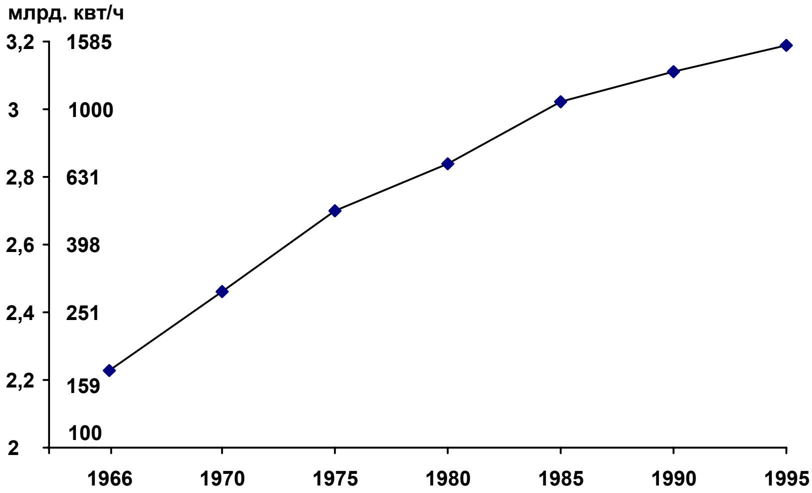
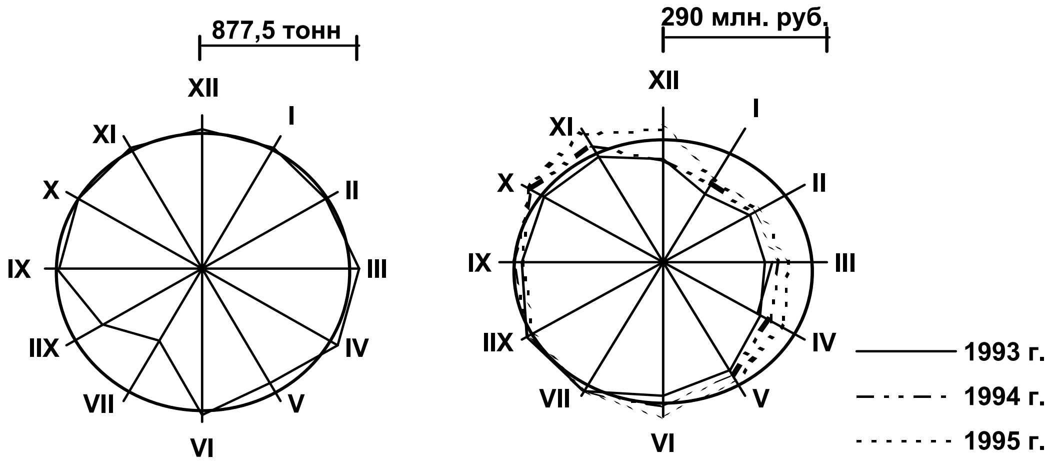


Рис. 4.18. Динамика производства электроэнергии в регионе за 1966 - 1995 гг.

К диаграммам динамики относятся и **радиальные диаграммы**, построенные в полярных координатах и предназначенные для отражения процессов, ритмически повторяющихся во времени (см. рис. 4.19). Чаще всего эти диаграммы применяются для иллюстрации сезонных колебаний, и в этом отношении они имеют преимущество перед статистическими кривыми. Радиальные диаграммы делятся на два вида: замкнутые и спиральные. Эти два вида диаграмм отличаются друг от друга по технике построения, все завит от того, что взято в качестве базы отсчета - центр круга или окружность.

Замкнутые диаграммы (см. рис. 4.19а) отражают весь внутригодичный цикл динамики какого либо одного года. Их построение сводится к следующему: вычерчивается круг, среднемесячный показатель приравнивается к радиусу этого круга, затем весь круг делится на двенадцать равных секторов, посредством проведения радиусов, которые изображаются в виде тонких линий. Каждый радиус изображает месяц, причем расположение месяцев аналогично циферблату часов. На каждом радиусе делается отметка в определенном месте, согласно масштабу, исходя из данных на соответствующий месяц. Если данные превышают среднегодовой уровень, то отметка делается вне окружности на продолжении радиуса. Затем отметки различных месяцев соединяются отрезками.

Если в качестве базы отсчета берется окружность, такого рода диаграммы называются спиральными (см. рис. 4.19б). Спиральные диаграммы отличаются 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 динамический ряд в виде одной кривой. Особенно наглядна такая диаграмма тогда, когда наряду с сезонным ритмом ряд обнаруживает неуклонный рост из года в год.



а) б)

Рис. 4.19.а. Колебания месячной продуктивности одной из кондитерских фабрик (в тоннах).

Рис. 4.19.б. Динамика объема работ строительного треста по месяцам (1993 - 1995 гг.) по сметной стоимости (млн. руб.).

Для отображения зависимости одного показателя от другого строится **диаграмма взаимосвязи** (см. рис. 4.20). Один показатель принимается за X, а другой за Y (т.е. функцию от X). Строится прямоугольная система координат с масштабами для показателей, в которой вычерчивается график.

Рис. 4.20. Зависимость уровня затрат на реализацию продукции от стоимости основных производственных фондов.

Построенный выше график показывает, что с увеличением стоимости основных производственных фондов происходит увеличение затрат на реализацию продукции. Согласно данным графика, можно утверждать, что с увеличением числа исследуемых предприятий зависимость двух показателей будет определяться линейной связью.

Диаграмма взаимосвязи имеет огромное применение на практике, так как множество различных величин связаны между собой той или иной формой прямой или обратной связи. Она может использоваться также для отображения различных циклических процессов (например инфляционной спирали), взаимонакладывающихся явлений и т.п.

4.6. Статистические карты.

Карты статистические представляют собой вид графических изображений статистических данных на схематичной 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 различной 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 различной степени насыщенности показывают интенсивность какого-либо показателя в пределах территориальной единицы.

Картограмма точечная - вид картограммы, где уровень какого-либо явления изображается с помощью точек. Точка изображает одну единицу совокупности или некоторое их количество, чтобы показать на географической карте плотность или частоту появления определенного признака.

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.

Глава 5. Средние показатели.

5.1. Сущность средних показателей

Наиболее распространенной формой статистических показателей, используемой в 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 обобщающую 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 незаменимым инструментом анализа явлений и процессов в экономике.

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

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, если мы рассчитаем средний курс по акциям всех предприятий, реализуемых в данный день на данной бирже, то получим фиктивную среднюю. Это будет объясняться тем, что используемая для расчета совокупность является крайне неоднородной. В этом и подобных случаях метод средних используется в сочетании с методом группировок: если совокупность неоднородна - общие средние должны быть заменены или дополнены групповыми средними, т.е. средними, рассчитанными по качественно однородным группам.

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

$$f\,(x_1, x_2, ..., x_n) \tag{5.1.}$$

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

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

$$f\,(x_1, x_2, ..., x_n)=f(\bar{x}, \bar{x}, \ldots, \bar{x}) \tag{5.2.}$$

Исходя из данного равенства и определяется средняя. На практике определить среднюю во многих случаях можно через **исходное соотношение средней** (ИСС) или ее логическую формулу:

$$ИСС = \frac{\text{Суммарное зна чение или объем осредняемого признака}}{\text{Число единиц или объем совокупности}}$$

Так, например, для расчета средней заработной платы работников предприятия необходимо общий фонд заработной платы разделить на число работников:

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

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

Для каждого показателя, используемого в экономическом анализе, можно составить только одно истинное исходное соотношение для расчета средней. Если, например, требуется рассчитать средний размер вклада в банке, то исходное соотношение будет следующим:

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

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

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

Однако от того, в каком виде представлены исходные данные для расчета средней, зависит, каким именно образом будет реализовано ее исходное соотношение. В каждом конкретном случае для реализации исходного соотношения потребуется одна из следующих форм средней величины:

- средняя арифметическая,
- средняя гармоническая,
- средняя геометрическая,
- средняя квадратическая, кубическая и т.д.

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

$$\bar{x} = \sqrt[k]{\frac{\sum x_i^k f_i}{\sum f_i}} \qquad i=\overline{1,n}$$

где x_i - i-ый вариант осредняемого признака ($i=\overline{1,n}$)

f_i - вес i-го варианта.

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

5.2. 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, пять торговых центров фирмы имеют следующий объем товарооборота за месяц:

Торговый центр	А	Б	В	Г	Д
Товарооборот (млн.руб.)	130	142	125	164	127

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

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

Используя приведенные в предыдущем параграфе условные обозначения, запишем формулу данной средней:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n} \qquad (5.3.)$$

С учетом имеющихся данных получим:

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

В данном случае мы использовали формулу средней арифметической простой (невзвешенной).

Средняя арифметическая взвешенная. При расчете средних величин отдельные значения осредняемого признака могут повторяться, встречаться по несколько раз. В подобных случаях расчет средней производится по сгруппированным данным или вариационным рядам, которые могут быть дискретными или интервальными.

Рассмотрим следующий пример:

Таблица 5.1.

Продажа акций АО “Дока-хлеб” на торгах фондовой секции ТМБ “Гермес” 11-17 мая 1994 г.

Сделка	Количество проданных акций, шт.	Курс продажи, руб.
1	500	1080
2	300	1050
3	1100	1145

Определим по данному дискретному вариационному ряду средний курс продажи 1 акции, что можно сделать только используя следующее исходное соотношение:

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

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

$$\overline{X} = \frac{1080 \times 500 + 1050 \times 300 + 1145 \times 1100}{500 + 300 + 1100} = \frac{2114500}{1900} = 1112{,}9 \text{ руб.}$$

Расчет среднего курса продажи произведен по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \qquad (5.4.)$$

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 (5.4.) получим:

$$\bar{x} = \sum (x_i \frac{f_i}{\Sigma f_i}) \qquad (5.5.)$$

или

$$\overline{X} = 1080 \times 0{,}263 + 1050 \times 0{,}158 + 1145 \times 0{,}579 = 1112{,}9 \text{ руб.}$$

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

Таблица 5.2.

Средние цены оптовых рынков на товар А

Оптовый рынок	Средняя цена (руб./шт.)
1	43
2	41

Можно ли по имеющимся данным определить среднюю цену данного товара по двум рынкам, вместе взятым? Можно, но только в том случае, когда объемы реализации этого товара на двух рынках совпадают. Тогда средняя цена реализации составит 42 руб. (доказательство этого правила будет приведено ниже.). Однако на первом рынке может быть реализовано, к примеру, 100 единиц товара, а на втором - 1000 единиц. Тогда для расчета средней цены потребуется уже средняя арифметическая взвешенная:

$$\overline{X} = \frac{43 \times 100 + 41 \times 1000}{100 + 1000} = 41{,}2\text{руб.}$$

Общий вывод заключается в следующем: использовать среднюю арифметическую невзвешенную можно только тогда, когда точно установлено отсутствие весов или их равенство.

При расчете средней по **интервальному вариационному ряду** для выполнения необходимых вычислений от интервалов переходят к их серединам. Рассмотрим следующий пример:

Таблица 5.3.

Распределение менеджеров корпорации по возрасту

Возраст (лет)	Число менеджеров (чел.)
до 25	7
25 - 30	13
30 - 40	38
40 - 50	42
50 - 60	16
60 и более	5
Итого:	121

Для определения среднего возраста управленческого персонала найдем середины возрастных интервалов. При этом величины открытых интервалов (первого и последнего) условно приравниваются к величинам интервалов, примыкающих к ним (второго и предпоследнего). С учетом этого середины интервалов будут следующими:

22, 5 27,5 35,0 45,0 55,0 65,0

Используя среднюю арифметическую взвешенную, определим средний возраст менеджера данной корпорации:

$$\overline{X} = \frac{22,5\times 7 + 27,5\times 13 + 35\times 38 + 45\times 42 + 55\times 16 + 65\times 5}{7+13+38+42+16+5} = 41 \text{ год}$$

Свойства средней арифметической. Средняя арифметическая обладает некоторыми математическими свойствами, более полно раскрывающими ее сущность и в ряде случаев используемыми при ее расчете. Рассмотрим эти свойства:

1. Произведение средней на сумму частот равно сумме произведений отдельных вариантов на соответствующие им частоты:

$$\overline{x}\sum f_i = \sum x_i f_i \qquad (5.6.)$$

Действительно, если мы обратимся к приведенному выше примеру расчета среднего курса продажи акций (табл. 5.1.), то получим следующее равенство (за счет округления среднего курса правая и левая части равенства в данном случае будут незначительно отличаться):
1112,9×1900=1080×500+1050×300+1145×1100

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

$$\sum (x_i - \overline{x}) f_i = 0 \qquad (5.7.)$$

Для нашего примера:

(1080-1112,9)×500+(1050-1112,9)×300+(1145-1112,9)×1100=0

Математическое доказательство данного свойства сводится к следующему:

$$\sum (x_i - \overline{x}) f_i = \sum x_i f_i - \sum \overline{x} f_i = \sum x_i f_i - \overline{x}\sum f_i = 0$$

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

$$\begin{aligned}&\sum (x_i - C)^2 f_i = \sum (x_i - \overline{x} + \overline{x} - C)^2 f_i = \sum \left[(x_i - \overline{x}) + (\overline{x} - C)\right]^2 f_i = \\ &= \sum \left[(x_i - \overline{x})^2 + 2(x_i - \overline{x})(\overline{x} - C) + (\overline{x} - C)^2\right] f_i = \sum (x_i - \overline{x})^2 f_i + \\ &+ 2(\overline{x} - C)\sum (x_i - \overline{x}) f_i + \sum (\overline{x} - C)^2 f_i = \sum (x_i - \overline{x})^2 f_i + \\ &+ 2(\overline{x} - C)\cdot 0 + \sum (\overline{x} - C)^2 f_i\end{aligned} \qquad (5.8.)$$

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

$$\sum(\bar{x}-C)^2 f_i \text{ или } (\bar{x}-C)^2\sum f_i$$

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

$$\mu_k=\frac{\sum(x_i-\bar{x})^k f_i}{\sum f_i},$$

где к 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:

$$\frac{\sum(x_i\pm A)f_i}{\sum f_i}=\frac{\sum x_i f_i}{\sum f_i}\pm\frac{\sum Af_i}{\sum f_i}=\bar{x}\pm A \qquad (5.9.)$$

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

$$\bar{x}=\frac{1180\times500+1150\times300+1245\times1100}{1900}=1212{,}9 \text{ руб.}$$

5. Если все варианты значений признака уменьшить или увеличить в А раз, то средняя также соответственно увеличится или уменьшится в А раз:

$$\frac{\sum\frac{x_i}{A}f_i}{\sum f_i}=\frac{\frac{1}{A}\sum x_i f_i}{\sum f_i}=\frac{1}{A}\bar{x} \qquad (5.10.)$$

Предположим, курс продажи в каждом случае возрастет в 1,5 раза. Тогда и средний курс также увеличится на 50%:

$$\bar{x}=\frac{1080\times1{,}5\times500+1050\times1{,}5\times300+1145\times1{,}5\times1100}{1900}=1112{,}9\times1{,}5=1669{,}4 \text{ руб.}$$

6. Если все веса уменьшить или увеличить в А раз, то средняя арифметическая от этого не изменится:

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

$$\frac{\sum x_i \frac{f_i}{A}}{\sum \frac{f_i}{A}} = \frac{\frac{1}{A}\sum x_i f_i}{\frac{1}{A}\sum f_i} = \bar{x} \qquad (5.11.)$$

Так, в нашем примере удобнее было бы рассчитывать среднюю, предварительно поделив все веса на 100:

$$\bar{x} = \frac{1080 \times 5 + 1050 \times 3 + 1145 \times 11}{19} = 1112{,}9 \text{ руб.}$$

Исходя из данного свойства, можно заключить, что если все веса равны между собой, то расчеты по средней арифметической взвешенной и средней арифметической простой приведут к одному и тому же результату.

5.3. 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. Данная форма используется, когда известен числитель исходного соотношения средней, но неизвестен его знаменатель. Рассмотрим расчет средней урожайности, являющейся одним из основных показателей эффективности производства в агробизнесе:

Таблица 5.4.

Валовой сбор и урожайность подсолнечника по Центрально-Черноземному району (в хозяйствах всех категорий)

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

Средняя урожайность любой сельскохозяйственной культуры в среднем по нескольким территориям, агрофирмам, фермерским хозяйствам и т.п. может быть определена только на основе следующего исходного соотношения:

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

Общий валовой сбор мы получим простым суммированием валового сбора по областям. Данные же о посевной площади 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 составляла 389,3 тыс.га, а средняя урожайность - 9,9 ц с одного гектара.

В данном случае расчет произведен по формуле средней гармонической взвешенной:

$$\bar{x} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}}, \text{ где } w_i = x_i f_i \qquad (5.12.)$$

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

Средняя гармоническая невзвешенная. Эта форма средней, используемая значительно реже, имеет следующий вид:

$$\bar{x} = \frac{n}{\sum \frac{1}{x_i}} \qquad (5.13.)$$

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

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

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

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

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

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

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

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

Средняя геометрическая. Еще одной формулой, по которой может осуществляться расчет среднего показателя, является средняя геометрическая:

$$\bar{x}=\sqrt[к]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_к}=\sqrt[к]{\Pi x_i} \qquad \text{- невзвешенная}$$

(5.14.)

$$\bar{x}=\sqrt[\Sigma m]{x_1^{m_1} \cdot x_2^{m_2} \cdot x_3^{m_3} \cdot \ldots \cdot x_к^{m_к}}=\sqrt[\Sigma m]{\Pi x_i^{m_i}} \qquad \text{- взвешенная}$$

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

Средняя квадратическая. В основе вычислений ряда сводных расчетных показателей лежит средняя квадратическая:

$$\bar{x}=\sqrt{\frac{\sum x_i^2}{n}} \qquad \text{- невзвешенная}$$

(5.15.)

$$\bar{x} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}}$$ - взвешенная

Наиболее широко этот вид средней используется при расчете показателей вариации.

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

5.4. Структурные средние

Наиболее часто используемыми в экономической практике структурными средними являются мода и медиана. **Мода** представляет собой значение изучаемого признака, повторяющееся с наибольшей частотой. **Медианой** 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:

$$\sum |x_i - M_e| = \min$$

Рассмотрим определение моды и медианы по **несгруппированным данным**.

Предположим, что 9 торговых фирм города реализуют товар А по следующим оптовым ценам (тыс.руб.).

4,4 4,3 4,4 4,5 4,3 4,3 4,6 4,2 4,6

Так как чаще всего встречается цена 4,3 тыс.руб., то она и будет модальной.

Для определения медианы необходимо провести ранжирование:

4,2 4,3 4,3 4,3 4,4 4,4 4,5 4,6 4,6

Центральной в этом ряду является цена 4,4 тыс.руб., следовательно, данная цена и будет медианой. Если ранжированный ряд 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 нормальном закону распределения совокупности. Она также используется в тех случаях, когда средняя не позволяет объективно оценить исследуемую совокупность вследствие сильного влияния максимальных и минимальных значений.

Проиллюстрируем познавательное значение медианы следующим примером.

Допустим, нам необходимо дать характеристику среднего дохода группы людей, насчитывающей 100 человек, из которых 99 имеют доходы в интервале от 100 до 1000 долл. в месяц, а месячные доходы последнего составляют 50000 долл.:

№ 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 совокупности людей.

Рассмотрим определение моды и медианы по сгруппированным **данным** (рядам распределения).

Предположим, распределение торговых предприятий города по уровню розничных цен на товар А имеет следующий вид :

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

Определение моды по дискретному вариационному ряду не состав

ляет большого труда - наибольшую частоту (60 предп.) имеет цена 55 руб., следовательно она и является модальной.

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

$$N_{me} = \frac{n+1}{2} \qquad (5.16)$$

где n - объем совокупности.

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

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

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

$$M_o = x_o + i \times \frac{(f_{M_o} - f_{M_o-1})}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})} \quad (5.17)$$

где Xo - нижняя граница модального интервала (модальным называется интервал, имеющий наибольшую частоту);

i - величина модального интервала;

f_{Mo} - частота модального интервала;

f_{Mo-1} - частота интервала, предшествующего модальному;

f_{Mo+1} - частота интервала, следующего за модальным.

и

$$M_e = x_0 + i \times \frac{\frac{1}{2}\sum f_i - s_{M_{e-1}}}{f_{M_e}} \quad (5.18)$$

где Xo - 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 таблицы 5.5.

Информация, подобная представленной в этой таблице, необходима для получения четкого представления о покупательной способности населения страны или региона, для оценки эластичности спроса и, в конечном итоге, для выбора того или иного метода ценообразования и обоснования окончательной цены на товар.

Таблица 5.5.

Распределение населения РФ по уровню среднедушевого денежного дохода в январе-августе 1995 г.

Среднедушевой денежный доход (в среднем за месяц), тыс.руб.	Удельный вес населения, %
100 и менее	2,4
100 - 200	15,4
200 - 300	20,1
300 - 400	17,2
400 - 500	12,8
500 - 600	9,2
600 - 700	6,5
700 - 800	4,5
800 - 900	3,2
900 - 1000	2,3
свыше 1000	6,4
Всего	100,0

Интервал с границами 200 - 300 в данном распределении будет модальным, так как он имеет наибольшую частоту. Использую формулу (5.17), определим моду :

$$M_0 = 200 + 100 \times \frac{20{,}1 - 15{,}4}{(20{,}1 - 15{,}4) + (20{,}1 - 17{,}2)} = 262 \text{ тыс.руб.}$$

Для определения медианного интервала необходимо определять накопленную частоту каждого последующего интервала до тех пор, пока она не превысит 1/2 суммы накопленных частот (в нашем случае - 50%) :

Интервал	Накопленная частота, %
100 и менее	2,4
100 - 200	17,8
200 - 300	37,9
300 - 400	55,1

Мы определили, что медианным является интервал с границами 300 - 400. Определим медиану :

$$M_e = 300 + 100 \times \frac{50,0 - 37,9}{17,2} = 370 \text{ тыс.руб.}$$

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

На основе полученных в последнем примере значений структурных средних можно заключить, что наиболее распространенным, типичным является среднедушевой доход порядка 260 тыс.руб. в месяц. В то же время, более половины населения располагает доходом свыше 370 тыс.руб. при среднем уровне 435 тыс.руб. (средняя арифметическая взвешенная). Из соотношения этих показателей следует вывод о правосторонней асимметрии распределения населения по уровню среднедушевых денежных доходов, что позволяет предполагать о достаточной емкости рынка дорогих товаров повышенного качества и товаров престижной группы.

Глава 6. Анализ вариации.

6.1. Основные показатели вариации.

Информация о средних уровнях исследуемых показателей обычно бывает недостаточной для глубокого анализа изучаемого процесса или явления. Необходимо учитывать и разброс или вариацию значений отдельных единиц, которая является важной характеристикой изучаемой совокупности. В наибольшей степени вариации подвержены курсы акций, объемы спроса и предложения, процентные ставки в разные периоды и в разных местах.

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

Для иллюстрации расчетов этих показателей воспользуемся следующими данными:

Таблица 6.1

Итоги торгов на валютных биржах России 21 января 1999г. (спецсессия)

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

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

$$R=Xmax-Xmin=22{,}83-22{,}40=0{,}43 \text{ руб.}$$

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

$$\sigma^2 = \frac{\sum\left(x_i - \overline{x}\right)^2}{n} \quad \text{невзвешенная формула} \qquad (6.1)$$

$$\sigma^2 = \frac{\sum\left(x_i - \overline{x}\right)^2 f_i}{\sum f_i} \quad \text{взвешенная формула} \qquad (6.2)$$

По данным нашего примера определим средневзвешенный курс доллара по итогам всех торгов и рассчитаем дисперсию:

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

$$\sigma^2 = \frac{(22{,}73 - 22{,}71)^2 \times 158{,}0 + (22{,}63 - 22{,}71)^2 \times 10{,}0 + ... + (22{,}56 - 22{,}71)^2 \times 0{,}7}{158{,}0 + 10{,}0 + ... + 0{,}7} = 0{,}004$$

Дисперсию в отдельных случаях удобнее рассчитывать по другой формуле, представляющей собой алгебраическое преобразование выражений (5.19.) и (5.20.):

$$\sigma^2 = \overline{x^2} - \left(\overline{x}\right)^2 \qquad (6.3)$$

$$\text{где } \overline{x^2} = \frac{\sum x_i^2}{n} \quad \text{или} \quad \overline{x^2} = \frac{\sum x_i^2 f_i}{\sum f_i} \qquad (6.4)$$

Наиболее удобным и широко распространенным на практике показателем является **среднее квадратическое отклонение**. Оно определяется как квадратный корень из дисперсии и имеет ту же размеренность, что и изучаемый признак:

$$\sigma = \sqrt{\frac{\sum\left(x_i - \overline{x}\right)^2}{n}} \quad \text{невзвешенная формула} \qquad (6.5)$$

$$\sigma = \sqrt{\frac{\sum\left(x_i - \overline{x}\right)^2 f_i}{\sum f_i}} \quad \text{взвешенная формула} \qquad (6.6)$$

В нашем случае получим:

$$\sigma = \sqrt{0{,}004} = 0{,}06\,руб.$$

Рассмотренная величина показывает, что курсы доллара на биржах отклонялись от средневзвешенного курса в среднем на 17,4 руб.

Рассмотренные показатели позволяют получить абсолютное значение вариации, т.е. оценивают ее в единицах измерения исследуемого признака. В отличие от них, **коэффициент вариации** измеряет колеблемость в относительном выражении, относительно среднего уровня, что во многих случаях является предпочтительнее:

$$V = \frac{\sigma}{\overline{x}} \times 100\% \qquad (6.7)$$

Определим значение этого показателя по нашим данным:

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

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

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

6.2. Использование показателей вариации в анализе взаимосвязей.

Показатели вариации могут быть использованы не только в анализе колеблемости или изменчивости изучаемого признака, но и для оценки степени воздействия одного признака на вариацию другого признака, т.е. в анализе взаимосвязей между показателями.

При проведении такого анализа исходная совокупность должна представлять собой множество единиц, каждая из которых характеризуется двумя признаками - факторным и результативным. Факторным называется признак, оказывающий влияние на взаимосвязанный с ним признак. В свою очередь, этот второй признак, подверженный влиянию, называется результативным.

Для выявления взаимосвязи исходная совокупность делится на две или более групп во факторному признаку. Выводы о степени взаимосвязи базируются на анализе вариации результативного признака. При этом применяется правило сложения дисперсий:

$$\sigma_0^2 = \overline{\sigma}^2 + \delta^2 \qquad (6.8)$$

где σ_0^2 - общая дисперсия;

$\overline{\sigma}^2$ - средняя из внутригрупповых дисперсий;

δ^2 - межгрупповая дисперсия.

Межгрупповая дисперсия отражает ту часть вариации результативного признака, которая обусловлена воздействием признака факторного. Это воздействие проявляется в отклонении групповых средних от общей средней:

$$\delta^2 = \frac{\sum(\overline{x}_i - \overline{x}_0)^2 n_i}{\sum n_i}, \qquad (6.9)$$

где $\overline{x}_i$ - среднее значение результативного признака по i-ой группе;

$\overline{x}_0$ - общая средняя по совокупности в целом;

n_i - объем (численность) i-ой группы.

Если факторный признак, по которому производилась группировка, не оказывает никакого влияния не признак результативный, то групповые средние будут равны между собой и совпадут с общей средней. В этом случае межгрупповая дисперсия будет равна нулю.

Средняя из внутригрупповых дисперсий отражает ту часть вариации результативного признака, которая обусловлена действием всех прочих неучтенных факторов, кроме фактора, по которому осуществлялась группировка:

$$\overline{\sigma}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i}, \qquad (6.10)$$

где σ_i^2 -дисперсия результативного признака в i-ой группе;

n_i -объем (численность) i-ой группы.

Теснота связи между факторным и результативным признаком оценивается на основе эмпирического корреляционного отношения:

$$\eta_э = \sqrt{\frac{\delta^2}{\sigma_0^2}} \qquad (6.11)$$

Данный показатель может принимать значения от 0 до 1. Чем ближе к 1 будет его величина, тем сильнее взаимосвязь между рассматриваемыми признаками.

На следующем условном примере исследуем зависимость между собственными и привлеченными средствами коммерческих банков региона:

Таблица №6.2

Банк	Собственные средства, млн.руб.	Привлеченные средства, млн. руб.
1.	70	300
2.	90	400
3.	140	530
4.	110	470
5.	75	255
6.	150	650
7.	90	320
8.	60	240
9.	95	355
10.	115	405

Если взаимосвязь между рассматриваемыми показателями существует, то она обусловлена влиянием объема собственных средств на объем привлеченных средств. Поэтому объем собственных средств выступает в данном примере в качестве факторного признака (X), а объем привлеченных средств в качестве результативного признака (Y).

Произведем группировку банков, выделив две группы по величине собственных средств, например, группу "да 100 млн.руб." и группу "100 млн. руб. и более". Результаты такой группировки представлены в следующей таблице:

Таблица №6.3

№ группы	Собственные средства, млн. руб.	Привлеченные средства, млн. руб.
1.	До 100	300 400 255 320 240 355
2.	100 и более	530 470 650 405

Расчет эмпирического корреляционного отношения включает несколько этапов:

1) рассчитываем групповые средние:

$$\bar{x}_i = \frac{\sum x_{ij}}{n_i},$$

где i- номер группы;
j- номер единицы в группе.

$$\bar{x}_1 = \frac{300 + 400 + ... + 355}{6} = 311{,}7 \text{млн.руб};$$

$$\bar{x}_2 = \frac{530 + ... + 405}{4} = 513{,}8 \text{млн.руб};$$

В данном примере при расчете групповых средних мы использовали невзвешенные формулы. Однако, при повторении вариантов для расчета

необходимо использовать средние взвешенные.
2) рассчитываем общую среднюю:

$$\bar{x}_0 = \frac{\sum \bar{x}_i n_i}{\sum n_i} = \frac{311{,}7 \cdot 6 + 513{,}8 \cdot 4}{10} = 392{,}5 \text{млн.руб}$$

Данную среднюю также можно было получить как отношение суммы всех единиц исходной совокупности (без учета деления на группы) к объему всей совокупности, т.е. к общему числу единиц.

3) рассчитываем внутригрупповые дисперсии:

$$\sigma_i^2 = \frac{\sum \left(x_{ij} - \bar{x}_i\right)^2}{n_i};$$

$$\sigma_1^2 = \frac{(300 - 311{,}7)^2 + (400 - 311{,}7)^2 + ... + (355 - 311{,}7)^2}{6} = 3039;$$

$$\sigma_2^2 = \frac{(530 - 513{,}8)^2 + ... + (405 - 513{,}8)^2}{4} = 8142.$$

Если бы варианты имели веса, то для расчета внутригрупповых дисперсий также требовались бы взвешенные формулы.
4) вычисляем среднюю из внутригрупповых дисперсий:

$$\overline{\sigma}^2 = \frac{3039 \cdot 6 + 8142 \cdot 4}{10} = 5080$$

5) определяем межгрупповую дисперсию:

$$\delta^2 = \frac{(311{,}7 - 392{,}5)^2 \cdot 6 + (513{,}8 - 392{,}5)^2 \cdot 4}{10} = 9803$$

находим общую дисперсию по правилу сложения:

$$\sigma_0^2 = 5080 + 9803 = 14883.$$

На этом этапе возможна проверка правильности выполненных ранее расчетов. Если возвратиться к исходной совокупности и не раздета ее на группы рассчитать дисперсию признака “у”, то она должна совпасть с общей дисперсией, полученной по правилу сложения.
рассчитываем эмпирическое корреляционное отношение:

$$\eta_э = \sqrt{\frac{9803}{14883}} = 0{,}81.$$

Полученная величина свидетельствует о том, что фактор, положенный в основание группировки (собственные средства), существенно влияет на размер привлеченных банками средств.

Глава 7. Выборочное наблюдение

7.1. Выборочное наблюдение как важнейший источник статистической информации

Статистическая методология исследования массовых явлений различает, как известно, два способа наблюдения в зависимости от полноты охвата объекта: сплошное и несплошное. Разновидностью несплошного наблюдения является выборочное, которое в условиях 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 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 социально-экономической ситуации.

Роль выборочного исследования в получении статистических данных возрастает в силу возможности, - когда это необходимо - расширения программы наблюдения. Так как исследованию подвергается сравнительно небольшая часть всей совокупности, можно с помощью многофазной выборки 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 и проведения. Обращение 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 таможенный и иной контроль обеспечивается также на основе выборки.

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.

Совокупность отобранных для обследования единиц в статистике принято называть **выборочной**, а совокупность единиц, из которых производится отбор, - **генеральной**.

Основные характеристики параметров генеральной и выборочной совокупности обозначаются определенными символами (табл. 7.1).

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 разработки ее материалов.
4. Определение процедуры отбора, способа отбора и объема выборки.
5. Подготовка кадров для проведения наблюдения, тиражирование 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.

Таблица 7.1.

Символы основных характеристик параметров генеральной и выборочной совокупностей

№ п/п	Характеристики	Генеральная совокупность	Выборочная Совокупность
1	Объем совокупности (численность единиц)	N	n
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	$\overline{x} = \frac{\sum x_i}{N}$	$\widetilde{x} = \frac{\sum x_i}{n}$
5	Дисперсия количественного признака	$\sigma^2_{\overline{x}} = \frac{\sum (x_i - \overline{x})^2}{N}$	$\sigma^2_{\widetilde{x}} = \frac{\sum (x_i - \widetilde{x})^2}{n}$
6	Дисперсия доли	$\sigma^2_p = pq$	$\sigma^2_w = w(1-w)$

7.2.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 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. 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.

Способ отбора определяет конкретный механизм или процедуру выборки единиц из генеральной совокупности. В практике выборочных обследований наибольшее распространение получили следующие виды 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 необходимо 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 количества 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 ПК и представляют собой абсолютно произвольные столбцы цифр. В соответствии с объектом генеральной совокупности выбирается любой столбец с числами необходимой значимости. Например, если генеральная совокупность включает 5000 единиц, потребуется 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.

Эти два вида ошибок связаны следующим соотношением:

$$\Delta = \mathbf{t}\mu,$$

где Δ - предельная ошибка выборки;

μ - средняя ошибка выборки;

t - коэффициент доверия, определяемый в зависимости от уровня вероятности **p**.

Ниже приведены некоторые значения t.

Таблица 7.2.

Вероятность, p_i	0,683	0,866	0,954	0,988	0,997	0,999
Значение t	1,0	1,5	2,0	2,5	3,0	3,5

Величина средней ошибки выборки рассчитывается дифференцированно в зависимости от способа отбора и процедуры выборки. Так, при случайном повторном отборе средняя ошибка определяется по формуле:

$$\mu = \frac{\sigma}{\sqrt{n}},$$

а при бесповторном:

$$\mu = \sqrt{\frac{\sigma^2}{n}(1 - \frac{n}{N})},$$

где σ^2 - выборочная (или генеральная)[1] дисперсии;

σ - выборочное (или генеральное) среднее квадратическое отклонение;

n - объем выборочной совокупности;

N - объем генеральной совокупности.

Расчет средней и предельной ошибок выборки позволяет определить возможные пределы, в которых будут находиться характеристики генеральной совокупности. Например, для выборочной средней такие пределы устанавливаются на основе следующих соотношений:

$$\tilde{x} - \Delta_{\tilde{x}} \leq \tilde{x} \leq \tilde{x} + \Delta_{\tilde{x}},$$

где $\bar{x}$ и $\tilde{x}$ - генеральная и выборочная средняя соответственно;

$\Delta_{\tilde{x}}$ - предельная ошибка выборочной средней.

Покажем практическое применение рассмотренной выше методики на следующих примерах.

Пример 1. При проверке веса импортируемого груза на таможне методом случайной повторной выборки было отобрано 200 изделий. В результате был установлен средний вес изделия 30 г. при среднем квадратическом отклонении 4 г. С вероятностью 0,997 определите пределы, в которых находится средний вес изделия в генеральной совокупности.

Решение. Рассчитаем сначала предельную ошибку выборки. Так как при p = 0,997 t = 3, она равна:

$$\Delta_{\tilde{x}} = t \cdot \frac{\sigma_{\tilde{x}}}{\sqrt{n}} = 3 \cdot \frac{4}{\sqrt{200}} = 0{,}84.$$

Определим пределы генеральной средней:

$$30 - 0{,}84 \le \overline{X} \le 30 + 0{,}84$$

или

$$29{,}16 \le \overline{X} \le 30{,}84\,.$$

Следовательно, с вероятностью 0,997 можно утверждать, что средний вес изделий в генеральной совокупности находится в пределах от 29,16 г. до 30,84 г.

Пример 2. В городе проживает 250 тыс. семей. Для определения среднего числа детей в семье была организована 2%-ная случайная бесповторная выборка семей. По ее результатам было получено следующее распределение семей по числу детей:

Число детей в семье	0	1	2	3	4	5
Количество семей	1000	2000	1200	400	200	200

С вероятностью 0,954 определите пределы, в которых будет находиться среднее число детей в генеральной совокупности.

Решение. Вначале на основе имеющегося распределения семей определим выборочные среднюю и дисперсию:

Число детей в семье, x_i	Количество семей, f_i	$x_i f_i$	$x_i - \tilde{x}$	$(x_i - \tilde{x})^2$	$(x_i - \tilde{x})^2 f_i$
0	1000	0	-1,5	2,25	**2250**
1	2000	2000	-0,5	0,25	500
2	1200	2400	0,5	0,25	300
3	400	1200	1,5	2,25	900
4	200	800	2,5	6,25	1250
5	200	1000	3,5	12,25	2450
Итого	5000	7400	-	-	7650

$$\tilde{x} = \frac{7400}{5000} \approx 1{,}5\,(\text{чел.});.\ \sigma_x^2 = \frac{7650}{5000} \approx 1{,}53\,.$$

Вычислим теперь предельную ошибку выборки (с учетом того, что при p = 0,954 t = 2).

$$\Delta_{\tilde{x}} = t\sqrt{\frac{\sigma_{\tilde{x}}^2}{n}\left(1 - \frac{n}{N}\right)} = 2\sqrt{\frac{1{,}53}{5000}\left(1 - \frac{5000}{250000}\right)} \approx 0{,}035\,.$$

Следовательно, пределы генеральной средней:

$$\overline{x} = \tilde{x} \pm \Delta_{\tilde{x}} = 1{,}5 \pm 0{,}035\,.$$

Таким образом, с вероятностью 0,954 можно утверждать, что среднее число детей в семьях города практически не отличается от 1,5, т.е. в среднем на каждые две семьи приходится три ребенка.

Наряду с определением ошибок выборки и пределов для генеральной средней эти же показатели могут быть определены для доли признака. В этом случае особенности расчета связаны с определением дисперсии доли, которая вычисляется так:

$$\sigma_w^2 = w\,(1 - w),$$

где $w = \frac{m}{n}$ - доля единиц, обладающих данным признаком в выборочной совокупности, определяемая как отношение количества соответствующих единиц к объему выборки.

Тогда, например, при собственно-случайном повторном отборе для определения предельной ошибки выборки используется следующая формула:

$$\Delta_w = t\sqrt{\frac{\sigma_w^2}{n}} = t\sqrt{\frac{w(1-w)}{n}}.$$

Соответственно, при бесповторном отборе:

$$\Delta_w = t\sqrt{\frac{\sigma_w^2}{n}\left(1-\frac{n}{N}\right)} = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}.$$

Пределы доли признака в генеральной совокупности p выглядят следующим образом:

$$w - \Delta_w \le p \le w + \Delta_w.$$

Рассмотрим пример.

Пример 3. С целью определения средней фактической продолжительности рабочего дня в государственном учреждении с численностью служащих 480 человек, в январе 1998 г. было проведена 25%-ная случайная бесповторная выборка. По результатам наблюдения оказалось, что у 10% обследованных потери времени достигали более 45 мин. в день. С вероятностью 0,683 установите пределы, в которых находится генеральная доля служащих с потерями рабочего времени более 45 мин. в день.

Решение. Определим объем выборочной совокупности:

$$n = 480 \cdot 0{,}25 = 120 \text{ чел.}$$

Выборочная доля w равна по условию 10%.

Учитывая, что при p = 0,683 t = 1, вычислим предельную ошибку выборочной доли:

$$\Delta_w = t\sqrt{\frac{w(1-w)}{n}(1-\frac{n}{N})} = 1 \cdot \sqrt{\frac{0{,}1(1-0{,}1)}{120}(1-\frac{120}{480})} = 0{,}0237 \approx 0{,}024 \text{ или } 2{,}4\%.$$

Пределы доли признака в генеральной совокупности:

$$10 - 2{,}4 \le p \le 10 + 2{,}4$$

или

$$7{,}6 \le p \le 12{,}4\,.$$

Таким образом, с вероятностью 0,683 можно утверждать, что доля работников учреждения с потерями рабочего времени более 45 мин. в день находится в пределах от 7,6% до 12,4%.

Механическая выборка применяется в случаях, когда генеральная совокупность каким-либо образом упорядочена, т.е. имеется определенная последовательность в расположении единиц (табельные номера работников, списки избирателей, телефонные номера респондентов, номера домов и квартир и т.п.).

Для проведения 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 составит $\frac{1}{50} = \left(\frac{1}{500\,000 : 10\,000}\right)$. 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.

Генеральную совокупность при механическом отборе можно ранжировать или упорядочить по величине изучаемого или коррелирующего с ним признака, что позволит повысить репрезентативность выборки. Однако в этом случае возрастает опасность систематической ошибки, связанной с занижением значений изучаемого признака (если из каждого интервала регистрируется первое значение) или с его завышением (если из каждого интервала регистрируется последнее значение). Поэтому целесообразно отбор начинать с середины первого интервала, например, при 5%-ной выборке отобрать 10-ю, 30-ю, 50-ю, 70-ю и с таким же интервалом последующие единицы.

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 обследовании населения такими группами могут быть, например, районы, социальные, возрастные или образовательные группы, при обследовании предприятий – отрасль или 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 N_i - объем i-ой группы;

n_i - объем выборки из i-ой группы.

Средняя ошибка такой выборки находится по формулам:

$$\mu = \sqrt{\frac{\overline{\sigma^2}}{n}} \text{ (повторный отбор)}$$

$$\mu = \sqrt{\frac{\overline{\sigma^2}}{n}\left(1 - \frac{n}{N}\right)} \text{ (бесповторный отбор)}$$

где $\overline{\sigma}_i^2$ - средняя из внутригрупповых дисперсий.

При выборке, пропорциональной дифференциации признака, число наблюдений по каждой группе рассчитывается по формуле:

$$n_i = n\frac{\sigma_i N_i}{\sum \sigma_i N_i},$$

где σ_i - среднее квадратическое отклонение признака в i-ой группе.
Средняя ошибка такого отбора определяется следующим образом:

$$\mu = \frac{1}{N}\sqrt{\sum \frac{\sigma_i^2 N_i^2}{n_i}} \text{ (повторный отбор)}$$

$$\mu = \frac{1}{N}\sqrt{\sum \frac{\sigma_i^2 N_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)} \text{ (бесповторный отбор).}$$

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% бесповторный типический отбор рабочих предприятия, пропорциональный размерам цехов, проведенный с целью оценки потерь из-за временной нетрудоспособности, привел к следующим результатам (табл. 7.3.).

Таблица 7.3.

Результаты обследования рабочих предприятия

Цех	Всего рабочих, человек	Обследовано, человек	Число дней временной нетрудоспособности за год	
			средняя	дисперсия
I	1000	100	18	49
II	1400	140	12	25
III	800	80	15	16

Рассчитаем среднюю из внутригрупповых дисперсий:

$$\overline{\sigma}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i} = \frac{49 \cdot 100 + 25 \cdot 140 + 16 \cdot 80}{100 + 140 + 80} = 30{,}25\,.$$

Определим среднюю и предельную ошибки выборки (с вероятностью 0,954):

$$\mu = \sqrt{\frac{30{,}25}{320}(1 - \frac{320}{3200})} = 0{,}29\,;$$

$$\Delta_x = 2 \cdot 0{,}29 = 0{,}58\,.$$

Рассчитаем выборочную среднюю:

$$\widetilde{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{18 \cdot 100 + 12 \cdot 140 + 15 \cdot 80}{100 + 140 + 80} = 14{,}6 \text{ дня.}$$

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:

$$\sum \sigma_i N_i = \sqrt{49} \cdot 1000 + \sqrt{25} \cdot 1400 + \sqrt{16} \cdot 800 = 17200;$$

$$n_1 = 320 \cdot \frac{\sqrt{49} \cdot 1000}{17200} = 130 \text{ человек;}$$

$$n_2 = 320 \cdot \frac{\sqrt{25} \cdot 1400}{17200} = 130 \text{ человек;}$$

$$n_3 = 320 \cdot \frac{\sqrt{16} \cdot 800}{17200} = 60 \text{ человек;}$$

С учетом полученных значений рассчитаем среднюю ошибку выборки:

$$\mu = \frac{1}{3200}\sqrt{\frac{49 \cdot 1000^2}{130}(1 - \frac{130}{1000}) + \frac{25 \cdot 1400^2}{130}(1 - \frac{130}{1400}) + \frac{16 \cdot 800^2}{60}(1 - \frac{60}{80})} = 0{,}28$$

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 собственно-случайном или 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 межгрупповой (межсерийной) дисперсии и определяется по следующим формулам:

$$\mu=\sqrt{\frac{\delta^2}{r}} \text{ (повторный отбор),}$$

$$\mu=\sqrt{\frac{\delta^2}{r}\left(1-\frac{r}{R}\right)} \text{ (бесповторный отбор),}$$

где r - число отобранных серий;

R - общее число серий.

Межгрупповую дисперсию вычисляют следующим образом:

$$\delta^2=\frac{\sum\left(\tilde{x}_i-\tilde{x}\right)^2}{r},$$

где $\tilde{x}_i$ - средняя i-й серии;

$\tilde{x}$ - общая средняя по всей выборочной совокупности.

Пример 4. В области, состоящей из 20 районов, проводилось выборочное обследование урожайности на основе отбора серий (районов). Выборочные средние по районам составили соответственно 14,5 ц/га; 16 ц/га; 15,5 ц/га; 15 ц/га и 14 ц/га. С вероятностью 0,954 определите пределы урожайности во всей области.

Решение. Рассчитаем общую среднюю:

$$\tilde{x}=\frac{14{,}5+16+...+14}{5}=15 \text{ ц/га.}$$

Межгрупповая (межсерийная) дисперсия равна:

$$\delta^2=\frac{(14{,}5-15)^2+(16-15)^2+...+(14-15)^2}{5}=0{,}5$$

Определим теперь предельную ошибку серийной бесповторной выборки (t = 2 при p = 0,954):

$$\Delta=2\sqrt{\frac{0{,}5}{5}(1-\frac{5}{20})}\approx 1{,}7\,.$$

Следовательно, урожайность в области будет с вероятностью 0,954 находиться в пределах:

$$15-1{,}7\le\bar{x}\le 15+1{,}7$$

или

$$13{,}3 \text{ ц/га} \le\bar{x}\le 16{,}7 \text{ ц/га}$$

7.3. 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.

Формулы необходимого объема выборки для различных способов формирования выборочной совокупности могут быть выведены из соответствующих соотношений, используемых при расчете предельных ошибок выборки. Приведем наиболее часто применяемые на практике выражения необходимого объема выборки:

- собственно-случайная и механическая выборка:

$$n=\frac{t^2\sigma^2}{\Delta^2} \text{ (повторный отбор);}$$

$$n=\frac{t^2\sigma^2 N}{\Delta^2 N+t^2\sigma^2} \text{ (бесповторный отбор);}$$

- типическая выборка:

$$n=\frac{t^2\overline{\sigma_i^2}}{\Delta^2} \text{ (повторный отбор);}$$

$$n=\frac{t^2\overline{\sigma_i^2} N}{\Delta^2 N+t^2\overline{\sigma_i^2}} \text{ (бесповторный отбор);}$$

- серийная выборка:

$$r=\frac{t^2\delta^2}{\Delta^2} \text{ (повторный отбор);}$$

$$r=\frac{t^2\delta^2 R}{\Delta^2 R+t^2\delta^2} \text{ (бесповторный отбор).}$$

При этом в зависимости от целей исследования дисперсии и ошибки выборки могут быть рассчитаны для средней величины или доли признака.

Рассмотрим примеры определения необходимого объема выборки при различных способах формирования выборочной совокупности.

Пример 5. В 100 туристических агентствах города предполагается провести обследование среднемесячного количества реализованных путевок методом механического отбора. Какова должна быть численность выборки, чтобы с вероятностью 0,683 ошибка не превышала 3 путевок, если по данным пробного обследования дисперсия составляет 225.

Решение. Рассчитаем необходимый объем выборки:

$$n = \frac{1^2 \cdot 225 \cdot 100}{3^2 \cdot 100 + 1^2 \cdot 225} = \frac{22500}{1125} = 20 \text{ агентств.}$$

Пример 6. С целью определения доли сотрудников коммерческих банков области в возрасте старше 40 лет предполагается организовать типическую выборку пропорциональную численности сотрудников мужского и женского пола с механическим отбором внутри групп. Общее число сотрудников банков составляет 12 тыс. чел., в том числе 7 тыс. мужчин и 5 тыс. женщин.

На основании предыдущих обследований известно, что средняя из внутригрупповых дисперсий составляет 1600. Определите необходимый объем выборки при вероятности 0,997 и ошибке 5%.

Решение. Рассчитаем общую численность типической выборки:

$$n = \frac{3^2 \cdot 1600 \cdot 12000}{5^2 \cdot 12000 + 3^2 \cdot 1600} = 550 \text{ чел.}$$

Вычислим теперь объем отдельных типических групп:

$$n_1 = \frac{550 \cdot 7000}{12000} = 319 \text{ чел.}$$

$$n_2 = \frac{550 \cdot 5000}{12000} = 231 \text{ чел.}$$

Таким образом, необходимый объем выборочной совокупности сотрудников банков составляет 550 чел., в т.ч. 319 мужчин и 231 женщина.

Пример 7. В акционерном обществе 200 бригад рабочих. Планируется проведение выборочного обследования с целью определения удельного веса рабочих, имеющих профессиональные заболевания. Известно, что межсерийная дисперсия доли равна 225. С вероятностью 0,954 рассчитайте необходимое количество бригад для обследования рабочих, если ошибка выборки не должна превышать 5%.

Решение. Необходимое количество бригад рассчитаем на основе формулы объема серийной бесповторной выборки:

$$n = \frac{2^2 \cdot 225 \cdot 200}{200 \cdot 5^2 + 2^2 \cdot 225} = 30 \text{ бригад.}$$

7.4. 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:

для средней $$\Delta_{\%}=\frac{\Delta_x}{\bar{x}}\cdot 100\%,$$

для доли $$\Delta_{\%}=\frac{\Delta_w}{\bar{p}}\cdot 100\%,$$

где $\Delta_{\%}$ - относительная предельная ошибка выборки;

Δ_x и Δ_w - предельная ошибка для среднего значения или доли признака соответственно;

$\bar{x}$ и $\bar{p}$ - генеральная средняя и доля соответственно.

Суждение о возможности распространения выборочных данных можно составить, если в формулах заменить $\bar{x}$ и $\bar{p}$ 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.

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 в расчете на 1 семью. Предположим, что оно составляет 0,3 человека. Умножив это число на численность генеральной совокупности, получим, что в детских яслях потребуется выделить 300 мест.

В условиях существования большого числа факторов, влияющих на точность данных выборочного наблюдения, использование точечной оценки при распространении выборочных характеристик на генеральную совокупность в статистических 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 число детей в возрасте до трех лет по выборочным данным составило 0,3 человека, а предельная ошибка - $\pm 0{,}1$ человека, то требуемое количество мест в детских учреждениях будет находиться в пределах от 200 до 400.

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:

$$Y_1 = Y_0 \frac{y_1}{y_0},$$

где Y_1 - численность совокупности с поправкой на недоучет;

Y_0 - численность совокупности без этой поправки;

y_0 - численность совокупности в контрольных точках по первоначальным данным;

y_1 - численность совокупности в тех же точках по данным контроль ных мероприятий.

До сих пор возможности выборки при уточнении данных сплошного наблюдения используются недостаточно. В то же время в современных условиях данный способ может быть, например, одним из инструментов контроля деятельности 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 8. 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. 7.4.

Таблица 7.4.

Количество коммерческих палаток в районах города до и после контрольных обходов

Районы	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

Рассчитанный по каждой категории работников коэффициент недоучета является основой уточнения имеющихся данных.

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. 7.5.

Таблица 7.5.

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

7.5. Малая выборка

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.

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:

$$t = \frac{\tilde{x} - \bar{x}}{\mu_{M.B}},$$

где $\mu_{M.B} = \frac{\sigma}{\sqrt{n-1}}$ - мера случайных колебаний выборочной средней в малой выборке.

Приведем выдержку из таблицы распределения Стьюдента.

Таблица 7.6.

Распределение вероятности в малых выборках в зависимости от коэффициента доверия t и объема выборки n*

t \ n	4	5	6	7	8	9	10	15	20	∞
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	970	980	938	984	992	992	997

* При n = ∞ в таблице даны вероятности нормального распределения. Для определения вероятности соответствующие табличные значения следует разделить на 1000.

Как видно из таблицы, при увеличении n это распределение стремится к нормальному и при n = 20 уже мало от него отличается.

Покажем, как пользоваться таблицей распределения Стьюдента.

Пример 9. 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; 4,2; 3,9; 3,7; 3,2; 2,2; 3,9. Найдем выборочные средние затраты:

$$\tilde{x} = \frac{3{,}4 + 4{,}7 + 1{,}8 + ... + 2{,}2{,}+3{,}9}{10} = 3{,}49 \text{ мин.}$$

Выборочная дисперсия:

$$\sigma_x^2 = \frac{(3{,}4 - 3{,}49)^2 + (4{,}7 - 3{,}49)^2 + ... + (3{,}9 - 3{,}49)^2}{10} = 0{,}713.$$

Отсюда средняя ошибка малой выборки равна:

$$\mu_{M.B} = \sqrt{\frac{0{,}713}{10-1}} = 0{,}28 \text{ мин.}$$

По табл. 7.6. находим, что для коэффициента доверия t = 2 и объема малой выборки n =10 вероятность равна 0,924. Таким образом, с вероятностью 0,924 можно утверждать, что расхождение между выборкой и генеральной средними лежит в пределах от -2μ до +2μ, т.е. разность $\tilde{x} - \bar{x}$ не превысит по абсолютной величине 0,56 (2×0,28). Следовательно, средние затраты времени во всей совокупности будут находится 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.

Глава 8. Статистическое изучение взаимосвязи социально-экономических явлений

8.1.Причинность, регрессия, корреляция

Исследование объективно существующих связей между явлениями - важнейшая задача теории статистики. В процессе статистического исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие основное влияние на вариацию изучаемых явлений и процессов. Причинно-следственные отношения - это такая связь явлений и процессов, когда изменение одного из них - причины ведет к изменению другого - следствия.

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, связанный с анализом природы социального или экономического явления методами экономической теории, социологии, конкретной экономики. Второй этап - построение модели связи, базируется на методах статистики: группировках, средних величинах, таблицах и так далее. Третий, последний этап - интерпретация результатов, вновь связан с качественными особенностями изучаемого явления. Статистика разработала множество методов изучения связей. Выбор метода изучения связи зависит от цели исследования, от поставленной задачи.

Признаки по их значению для изучения взаимосвязи делятся на два класса. Признаки, обуславливающие изменения других, связанных с ними признаков, называют **факторными**, или просто факторами. Признаки, изменяющиеся под действием факторных признаков, называют **результативными**.

В статистике различают функциональную и стохастическую зависимости. **Функциональной** называют такую связь, при которой определенному значению факторного признака соответствует одно значение результативного.

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.

Связи между явлениями и их признаками классифицируются по степени тесноты, направлению и аналитическому выражению.

По степени тесноты связи различают:

Таблица 8.1

Количественные критерии оценки тесноты связи

Величина коэффициента корреляции	Характер связи
До ±0,3	практически отсутствует
±0,3 - ±0,5	слабая
±0,5 - ±0,7	умеренная
±0,7 - ±1,0	сильная

По направлению выделяют связь **прямую** и **обратную.** Прямая - это такая связь, при которой с увеличением или с 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 приблизительно выражена уравнением прямой линии, то ее называют **линейной** связью вида:

$$\bar{y}_x = a_0 + a_1 x \qquad (8.1)$$

Если же связь может быть выражена уравнением какой-либо кривой линии, например:

параболы - $\bar{y}_x = a_0 + a_1 x + a_2 x^2$ (8.2)

гиперболы - $\bar{y}_x = a_0 + a_1 \frac{1}{x}$; и т.д., то такую связь называют нелинейной или криволинейной. 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 изменение двух величин:

X	1	2	3	4	5	6	7	8	9
Y	5	9	6	10	14	17	15	20	23

Мы видим, что с увеличением величины X величина Y также возрастает. Можно сделать предположение, что связь между ними прямая и что ее можно описать или уравнением прямой или уравнением параболы второго порядка.

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, обозначаются 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.

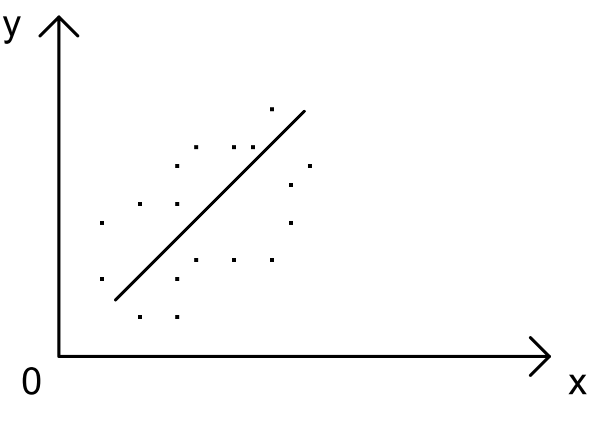


Рис. 8.1. График корреляционного поля.

Корреляция - это статистическая зависимость между случайными величинами, не имеющая строго функционального характера, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

В статистике принято различать следующие варианты зависимостей:

1. 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, которые давая количественную характеристику тесноты связи между признаками, позволяют определять «полезность» факторных признаков при построении уравнения множественной регрессии.

Регрессия тесно связана с корреляцией: первая оценивает силу (тесноту) статистической связи, вторая исследует ее форму.

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).

Одной из проблем построения уравнений регрессии является их размерность, то есть 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.

При построении моделей регрессии должны соблюдаться следующие требования:

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 должны описываться линейной или приводимой к линейной форме зависимостью.
6. Отсутствие количественных ограничений на параметры модели связи.

7. Постоянство территориальной и временной структуры изучаемой совокупности.

Соблюдение данных требований позволяет построить модель, наилучшим образом описывающую реальные явления и процессы.

8.2.Парная регрессия на основе метода наименьших квадратов и метода группировок

Парная регрессия характеризует связь между двумя признаками: результативным и факторным. Аналитически связь между ними описывается уравнениями:

прямой $\bar{y}_x = a_0 + a_1 x$

гиперболы $\bar{y}_x = a_0 + a_1 \frac{1}{x}$ (8.3)

параболы $\bar{y}_x = a_0 + a_1 x + a_2 x^2$

и так далее.

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, то это свидетельствует о том, что связь между ними линейная, а при обратной связи - гиперболическая. Если результативный признак увеличивается в арифметической прогрессии, а факторный значительно быстрее, то используется параболическая или степенная регрессия.

Оценка параметров уравнений регрессии (a_0, a_1, и a_2 - в уравнении параболы второго порядка) осуществляется методом наименьших квадратов, в основе которого лежит предположение о независимости наблюдений исследуемой совокупности и нахождении параметров модели (a_0, a_1), при которых минимизируется сумма квадратов отклонений эмпирических (фактических) значений результативного признака от теоретических, полученных по выбранному уравнению регрессии:

$$S = \sum (y - \bar{y}_x)^2 \to \min$$

Система нормальных уравнений для нахождения параметров линейной парной регрессии методом наименьших квадратов имеет следующий вид:

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \end{cases} \qquad (8.4)$$

где n - объем исследуемой совокупности (число единиц наблюдения).

В уравнениях регрессии параметр a_0 показывает усредненное влияние на результативный признак неучтенных в уравнении факторных признаков; коэффициент регрессии a_1 показывает, на сколько изменяется в среднем значение результативного признака при увеличении факторного на единицу собственного измерения.

Например, имеются данные, характеризующие деловую активность закрытого акционерного общества (ЗАО): прибыль (млн. руб.) и затраты на 1 руб. произведенной продукции.

Таблица 8.2

Зависимость между прибылью ЗАО и затратами на 1 руб. произведенной продукции

№ п/п	Прибыль (млн. руб.) (y)	Затраты на 1 руб. произведенной продукции (коп.) (x)	x^2	xy	$\overline{y}_x$
1	221	96	9216	21216	193
2	1070	77	5929	82390	1044
3	1001	77	5929	77077	1044
4	606	89	7921	53934	507
5	779	82	6724	63878	813
6	789	81	6561	63909	865
Итого	4466	502	42280	362404	4466

Предположим наличие линейной зависимости между рассматриваемыми признаками.

Система нормальных уравнений для данного примера имеет вид:

$$\begin{cases} 6a_0 + 502a_1 = 4466 \\ 502a_0 + 42280a_1 = 362404 \end{cases}$$

Отсюда: $a_0 = 4494{,}06$; $a_1 = -44{,}8$

Следовательно, $\overline{y}_x = 4494{,}06 - 44{,}8x$.

На практике исследования часто проводятся по большому числу наблюдений. В этом случае исходные данные удобнее представлять в сводной групповой таблице. При этом анализу подвергаются сгруппированные данные и по факторному (x) и по результативному (y) признакам, то есть уравнения парной регрессии целесообразно строить на основе сгруппированных данных.

Если значения x и y заданы в определенных интервалах (a, b), то для каждого интервала сначала необходимо определить середину ($x'/y' = (a+b)/2$), а затем уже коррелировать значения x' и y' и строить уравнения регрессии между ними.

Например, определим зависимость между величиной уставного капитала и числом занятых на предприятиях, выставивших акции на чековые аукционы в 1996 г. в одном из регионов, который характеризуется следующими данными:

Таблица 8.3

Распределение предприятий, выставивших акции на чековые аукционы в 1996 г., по величине уставного капитала и числу занятых в одном из регионов

Уставной капитал (млн.руб.) (y)	Число занятых (чел.) (x)					Число предприятий, f_y	yf_y	xyf_y
		14-70	70-126	126-182	182-238			
	x_i / y_i	42	98	154	210			
745-2684	1714,5	4	6	2	3	15	2517,5	2904363
2684-4624	3654,0	1	3	-	-	4	14616,0	1227744
4624-6564	5594,0	-	1	1	-	2	11188,0	1409688
6564-8503	7533,5	1	1	2	-	4	30134,0	3374995
8503-125842	67172,5	2	-	1	2	5	335862,5	44199505
Число предприятий, f_x	-	8	11	6	5	30	417518,0	53533813
xf_x	-	336	1078	924	1050	3388		
x^2f_x		14112	105644	142296	220500	482552		

Предположим наличие линейной зависимости между рассматриваемыми признаками.

Система нормальных уравнений для определения коэффициентов уравнения регрессии примет вид:

$$\begin{cases} na_0 + a_1\sum xf_x = \sum yf_y \\ a_0\sum xf_x + a_1\sum x^2f_x = \sum xyf_y \end{cases}$$

где n=30 - число анализируемых предприятий;

f_x/f_y - число предприятий, согласно распределению, соответственно по факторному и результативному признакам;

yf_y / xf_x - значения результативного и факторного признака по конкретной группе предприятий.

Так для первой группы:

yf_y=1714,5×15=25717,5; xf_x=42×8=336

xyf_y=1714,5×4×42+1714,5×6×98+1714,5×2×154+1714,5×3×210= =2904363

x^2f_x=42×42×8=14112

Таким образом, подставив в систему суммарные значения, получим:

$$\begin{cases} 30a_0 + 3388a_1 = 417518 \\ 3388a_0 + 482552a_1 = 53533813 \end{cases}$$

a_0=6640; a_1=64

Отсюда $\bar{y}_x$ =6640+64x.

8.3. Множественная (многофакторная) регрессия

Изучение связи между тремя и более связанными между собой признаками носит название **множественной (многофакторной)** регрессии:

$$\bar{y}_{1,2,\ldots,k} = f(x_1, x_2, \ldots, x_k)$$

Построение моделей множественной регрессии включает несколько этапов:

1. Выбор формы связи (уравнения регрессии);
2. Отбор факторных признаков;
3. Обеспечение достаточного объема совокупности.

Выбор типа уравнения затрудняется тем, что для любой формы зависимости можно выбрать целый ряд уравнений, которые в определенной степени будут описывать эти связи. Основное значение имеют линейные модели в силу простоты и логичности их экономической интерпретации.

Важным этапом построения уже выбранного уравнения множественной регрессии является отбор и последующее включение факторных признаков.

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 интуитивно-логических или многомерных статистических методов анализа.

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, на сколько уменьшается сумма квадратов остатков и увеличивается величина множественного коэффициента корреляции (R^2). Одновременно используется и обратный метод, то есть исключение факторов, ставших незначимыми.

Фактор является незначимым, если его включение в уравнение регрессии только изменяет значения 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 коэффициента регрессии не изменяется (или меняется несущественно), то данный признак существенен и его включение в уравнение регрессии необходимо. В противном случае, фактор нецелесообразно включать в модель регрессии.

При построении модели регрессии возможна проблема мультиколлинеарности, под которой понимается тесная зависимость между факторными признаками, включенными в модель $(r_{x_{ij}} > 0{,}8)$.

Наличие мультиколлинеарности между признаками приводит к:

- искажению величины параметров модели, которые имеют тенденцию к завышению, чем осложняется процесс определения наиболее существенных факторных признаков;
- изменению смысла экономической интерпретации коэффициентов регрессии.

В качестве причин возникновения мультиколлинеарности между признаками, можно выделить следующие:

- изучаемые факторные признаки являются характеристикой одной и той же стороны 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;
- факторные признаки являются составляющими элементами друг друга;
- факторные признаки по экономическому смыслу дублируют друг друга.

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 анализа 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 связи результативного признака от ряда факторных выражается и называется многофакторным (множественным) уравнением регрессии или моделью связи.

Линейное уравнение множественной регрессии имеет вид:

$$\bar{y}_{1,2,\ldots,k} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_k x_k$$

где $\bar{y}_{1,2,3,\ldots,k}$ - теоретические значения результативного признака, полученные в результате подстановки соответствующих значений факторных признаков в уравнение регрессии;

$x_1, x_2, \ldots, x_k$ - факторные признаки;

$a_1, a_2, \ldots, a_k$ - параметры модели (коэффициенты регрессии).

Параметры уравнения могут быть определены графическим методом, методом наименьших квадратов и так далее.

Пример. По следующим данным о прибыли (y), затратах на 1 руб. произведенной продукции (x_1) и стоимости основных фондов (x_2) определим зависимость между признаками.

Таблица 8.4

Расчетная таблица для определения параметров уравнения регрессии

Прибыль (млн. руб.) (y)	Затраты на 1 руб. произведенной продукции (коп.) (x_1)	Стоимость основных фондов (млрд. руб.) (x_2)	x_1^2	$x_1 x_2$	$x_1 y$	x_2^2	$x_2 y$
221	96	4,3	9216	412,8	21216	18,49	950,3
1070	77	5,9	5929	454,3	82390	34,81	6313,0
1001	77	5,9	5929	454,3	77070	34,81	5905,9
606	89	3,9	7921	347,1	53934	15,21	2363,4
779	82	4,3	6724	352,6	63878	18,49	3349,7
789	81	4,9	6561	396,9	63909	24,01	3866,1
4466	502	29,2	42280	2418	362404	145,82	22748,4

Система нормальных линейных уравнений имеет вид:

$$\begin{cases} na_0 + a_1\sum x_1 + a_2\sum x_2 = \sum y \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 = \sum x_1y \\ a_0\sum x_2 + a_1\sum x_1x_2 + a_2\sum x_2^2 = \sum x_2y \end{cases}$$

$$\begin{cases} 6a_0 + 502a_1 + 29{,}2a_2 = 4466 \\ 502a_0 + 42280a_1 + 2418a_2 = 362404 \\ 29{,}2a_0 + 2418a_1 + 145{,}82a_2 = 22748{,}4 \end{cases}$$

Таким образом:

$$\bar{y}_{x_1,x_2} = 4078{,}9 - 40{,}02x_1 + 2{,}87x_2 .$$

8.4. Собственно-корреляционные параметрические методы изучения связи.

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) факторных.

Линейный коэффициент корреляции 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 формулы расчета данного коэффициента:

$$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x \cdot \sigma_y} \tag{8.5}$$

Производя расчет по итоговым значениям исходных переменных, линейный коэффициент корреляции можно вычислить по формуле:

$$r = \frac{n\sum xy - \sum x\sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\cdot\left[n\sum y^2 - \left(\sum y\right)^2\right]}} \tag{8.6}$$

Между линейным коэффициентом корреляции и коэффициентом регрессии существует определенная зависимость, выражаемая формулой:

$$r = a_i \frac{\sigma_{x_i}}{\sigma_y} \tag{8.7}$$

где a_i - коэффициент регрессии в уравнении связи;

σ_{x_i} - среднеквадратическое отклонение соответствующего, статистически существенного, факторного признака.

Линейный коэффициент корреляции изменяется в пределах от -1 до 1: $-1 \le r \le 1$. Знаки коэффициентов регрессии и корреляции совпадают. При этом интерпретацию выходных значений коэффициента корреляции можно представить в следующей таблице:

Таблица 8.5

Оценка линейного коэффициента корреляции

Значение линейного коэффициента связи	Характер связи	Интерпретация связи
r = 0	отсутствует	-
0<r<1	прямая	с увеличением *x* увеличивается *y*
-1<*r*<0	обратная	с увеличением *x* уменьшается *y* и наоборот
r=1	функциональная	каждому значению факторного признака строго соответствует одно значение результативного признака

Пример: На основе выборочных данных о деловой активности однотипных коммерческих структур оценить тесноту связи между прибылью y (млн. руб.) и затратами на 1 руб. произведенной продукции *x* (коп.).

Таблица 8.6

Расчетная таблица для определения коэффициента корреляции

№ п/п	y	x	yx	y^2	x^2
1	221	96	21216	48841	9216
2	1070	77	82390	1144900	5929
3	1001	77	77077	1002000	5929
4	606	89	53934	367236	7921
5	779	82	63878	606841	6724
6	789	81	63909	622520	6561
Сумма	4466	502	362404	3792338	42280
Средняя	744,33	83,67	60400,67	632056,33	7046,67

1. Используя формулу (8.5) получаем:

$$\sigma_y^2 = \overline{y}^2 - (\overline{y})^2 = 632056{,}3 - (744{,}3)^2 = 78029{,}3$$

$$\sigma_x^2 = \overline{x}^2 - (\overline{x})^2 = 7046{,}67 - (83{,}67)^2 = 46$$

$$r = \frac{60400{,}67 - 744{,}33 \cdot 83{,}67}{\sqrt{78029{,}3 \cdot 46}} = -0{,}98$$

2. По формуле (8.6) значение коэффициента корреляции составило:

$$r = \frac{6 \cdot 362404 - 4466 \cdot 502}{\sqrt{\left[6 \cdot 42280 - (502)^2\right] \cdot \left[6 \cdot 3792338 - (4466)^2\right]}} =$$

$$= \frac{2174424 - 2241932}{\sqrt{(253680 - 252004) \cdot (22754028 - 19945156)}} =$$

$$= \frac{-67508}{\sqrt{1676 \cdot 2808872}} = \frac{-67508}{68612{,}46} = -0{,}98$$

Таким образом, результат по всем формулам одинаков и свидетельствует о сильной обратной зависимости между изучаемыми признаками.

В случае наличия линейной и нелинейной зависимости между двумя признаками для измерения тесноты связи применяют так называемое **корреляционное отношение.** Различают эмпирическое и теоретическое корреляционное отношение.

Эмпирическое корреляционное отношение рассчитывается по данным группировки, когда δ^2 характеризует отклонения групповых средних результативного показателя от общей средней:

$$\eta = \sqrt{\frac{\sigma^2 - \overline{\sigma}^2}{\sigma^2}} = \sqrt{1 - \frac{\overline{\sigma}^2}{\sigma^2}} = \sqrt{\frac{\delta^2}{\sigma^2}} \qquad (8.8)$$

где η - корреляционное отношение;
σ^2 - общая дисперсия;
$\overline{\sigma}^2$ - средняя из частных (групповых) дисперсий;
δ^2 - межгрупповая дисперсия (дисперсия групповых средних).

Все эти дисперсии есть дисперсии результативного признака.

Теоретическое корреляционное отношение определяется по формуле:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{1 - \frac{\sigma^2_{ост}}{\sigma^2}} \quad (8.9)$$

где δ^2 - дисперсия выравненных значений результативного признака, то есть рассчитанных по уравнению регрессии;

σ^2 - дисперсия эмпирических (фактических) значений результативного признака.

Корреляционное отношение изменяется в пределах от 0 до 1 $(0 \le \eta \le 1)$ и анализ степени тесноты связи полностью соответствует линейному коэффициенту корреляции (таблица 8.1).

Для измерения тесноты связи при множественной корреляционной зависимости, то есть при исследовании трех и более признаков одновременно, вычисляется множественный и частные коэффициенты корреляции.

Множественный коэффициент корреляции вычисляется при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Множественный коэффициент корреляции для двух факторных признаков вычисляется по формуле:

$$R_{y/x_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}} \quad (8.10)$$

где r_{yx_i} - парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \le R \le 1$.

Приближение R к единице свидетельствует о сильной зависимости между признаками.

На основе данных таблицы 8.4 рассчитаем коэффициент множественной корреляции и его ошибку:

$$r_{yx_1} = \frac{\overline{yx}_1 - \bar{y} \cdot \bar{x}_1}{\sigma_y \cdot \sigma_{x_1}} = -0{,}98; \quad r_{yx_2} = \frac{\overline{yx}_2 - \bar{y} \cdot \bar{x}_2}{\sigma_y \cdot \sigma_{x_2}} = 0{,}78; \quad r_{x_1x_2} = -0{,}86.$$

Множественный коэффициент корреляции составит:

$$R_{y/x_1x_2} = \sqrt{\frac{-0{,}98^2 + 0{,}78^2 - 2(-0{,}98) \cdot 0{,}78 \cdot (-0{,}86)}{1 - (-0{,}86)^2}} = 0{,}99$$

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками x_1 и x_2 при фиксированном значении других $(k-2)$ факторных признаков, то есть когда влияние x_3 исключается, то есть оценивается связь между x_1 и x_2 в «чистом виде».

В случае зависимости y от двух факторных признаков x_1 и x_2 коэффициенты частной корреляции имеют вид:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{\left(1-r^2_{x_2y}\right)\cdot\left(1-r^2_{x_1x_2}\right)}}$$

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{\left(1-r^2_{x_1y}\right)\cdot\left(1-r^2_{x_1x_2}\right)}} \qquad (8.11)$$

где r - парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака x_2, во втором - x_1.

На основании приведенных выше данных о зависимости трех факторов деятельности предприятий вычислим частные коэффициенты корреляции (см. табл. 8.4):

$$r_{yx_1/x_2} = \frac{-0{,}99 - 0{,}78\cdot(-0{,}86)}{\sqrt{\left(1-(0{,}78)^2\right)}\cdot\sqrt{\left(1-(-0{,}86)^2\right)}} = -0{,}999$$

$$r_{yx_2/x_1} = -0{,}992;\; r_{x_1x_2/y} = -0{,}994.$$

8.5. Принятие решений на основе уравнений регрессии

Интерпретация моделей регрессии осуществляется методами той отрасли знаний, к которой относится исследуемое явление. Но всякая интерпретация начинается со статистической оценки уравнения регрессии в целом и оценки значимости входящих в модель факторных признаков.

Чем больше величина коэффициента регрессии, тем значительнее влияние данного признака на моделируемый.

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 имеет знак минус, то с его увеличением результативный признак уменьшается.

Если экономическая теория подсказывает, что факторный признак должен иметь положительное значение, а он имеет знак минус, то необходимо проверить расчеты параметров уравнения регрессии. Такое явление чаще всего бывает в силу допущенных ошибок при решении. Однако следует иметь ввиду, что когда рассматривается совокупное влияние факторов, то в силу наличия взаимосвязей между ними характер их влияния может меняться.

С целью расширения возможностей экономического анализа, используются **частные коэффициенты эластичности**, определяемые по формуле:

$$Э_{x_i} = a_1 \cdot \frac{\overline{x}_i}{\overline{y}} \qquad (8.12)$$

где $\overline{x}_i$ - среднее значение соответствующего факторного признака;

$\overline{y}$ - среднее значение результативного признака;

a_1 - 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 ($Э_{x_i}$) по исходным данным о зависимости между прибылью ЗАО (y), затратами на 1 руб. произведенной продукции (x_1) и стоимостью основных фондов (млн. руб.), приведенным в таблице 8.4.

$a_1 = -40{,}02$; $a_2 = 2{,}87$.

$\overline{y} = \frac{4466}{6} = 744{,}3$; $\overline{x}_1 = \frac{502}{6} = 83{,}7$; $\overline{x}_2 = \frac{29{,}2}{6} = 4{,}9$.

$Э_{x_1} = a_1 \cdot \frac{\overline{x}_1}{\overline{y}} = -40{,}02 \cdot \frac{83{,}7}{744{,}3} = -4{,}5$; $Э_{x_2} = a_2 \frac{\overline{x}_2}{\overline{y}} = 0{,}02$.

Это значит, что при увеличении затрат на 1 рубль произведенной продукции на 1%, прибыль ЗАО снизится на 4,5%, а при увеличении стоимости основных фондов на 1%, прибыль увеличится на 0,02%.

Частный коэффициент детерминации:

$$d_{x_i} = r_{yx_i} \cdot \beta_{x_i} \qquad (8.13)$$

где r_{yx_i} - парный коэффициент корреляции между результативным и i- ым факторным признаком;

β_{x_i} - соответствующий стандартизованный коэффициент уравнения множественной регрессии: $\beta_{x_1} = a_1 \cdot \frac{\sigma_{x_i}}{\sigma_y}$ (8.14)

Частный коэффициент детерминации показывает на сколько процентов вариация результативного признака объясняется вариацией i- го признака, входящего в множественное уравнение регрессии.

По данным, приведенным в таблице 8.4 рассчитаем частный коэффициент детерминации для фактора x_2 - затраты на 1 руб. произведенной продукции:

$$d_{x_1} = r_{yx_1}\beta_{x_1};\ r_{yx_1} = \frac{\overline{yx_1} - \bar{y}\cdot\bar{x}}{\sigma_y \cdot \sigma_{x_1}};\ \beta_{x_1} = a_1\frac{\sigma_{x_1}}{\sigma_y}$$

$$\overline{yx_1} = \frac{\sum yx_1}{n} = \frac{362404}{6} = 60400{,}7;\ \bar{y} = \frac{\sum y}{n} = \frac{4466}{6} = 744{,}3;$$

$$\bar{x}_1 = \frac{\sum x_1}{n} = \frac{502}{6} = 83{,}7;$$

$$\sigma_y^2 = \overline{y^2} - \left(\bar{y}\right)^2 = \frac{3792340}{6} - \left(744{,}3\right)^2 = 78074{,}2$$

$$\sigma_x^2 = \bar{x}_1^2 - (\bar{x}_1)^2 = \frac{42280}{6} - \left(83{,}7\right)^2 = 46{,}6$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{78074{,}2} = 279{,}4;\ \sigma_x = \sqrt{46{,}6} = 6{,}8$$

$$r_{yx_1} = \frac{60400{,}7 - 744{,}3\cdot 83{,}7}{279{,}4\cdot 6{,}8} = -0{,}99$$

$$\beta_{x_1} = -40{,}02\cdot\frac{6{,}8}{279{,}4} = -0{,}97;\ d_{x_1} = -0{,}99\cdot\left(-0{,}97\right) = 0{,}96.$$

- частный коэффициент детерминации для фактора x_2 - стоимость основных фондов:

$$d_{x_2} = r_{yx_2}\cdot\beta_{x_2} = 0{,}006.$$

Из расчетов следует, что вариация прибыли ЗАО на 96% объясняется изменением затрат на 1 руб. произведенной продукции.

Наиболее полная экономическая интерпретация моделей регрессии позволяет выявить резервы развития и повышения деловой активности субъектов экономики.

8.6. Методы изучения связи качественных признаков

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.

Для их вычисления строится таблица, которая показывает связь между двумя явлениями, каждое из которых должно быть альтернативным, то есть состоящим из двух качественно отличных друг от друга значений признака (например, хороший, плохой).

Таблица 8.7

Таблица для вычисления коэффициентов ассоциации и контингенции

a	b	a+b
c	d	c+d
a+c	b+d	a+b+c+d

Коэффициенты вычисляются по формулам:

ассоциации: $$K_a = \frac{ad - bc}{ad + bc} \qquad (8.15)$$

контингенции: $$K_k = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}} \qquad (8.16)$$

Коэффициент контингенции всегда меньше коэффициента ассоциации. Связь считается подтвержденной, если $K_a \geq 0{,}5$ или $K_k \geq 0{,}3$.

Пример. Исследуем связь между участием в забастовках рабочих и уровнем их образования. Результаты обследования характеризуются следующими данными:

Таблица 8.8

Зависимость участия рабочих в забастовках от образовательного уровня

Группы рабочих	Число рабочих	Из них	
		участвующих в забастовке	не участвующих в забастовке
Имеют среднее образование	100	78	22
Не имеют среднего образования	100	32	68
Итого	200	110	90

$$K_a = \frac{78 \cdot 68 - 32 \cdot 22}{78 \cdot 68 + 32 \cdot 22} = \frac{4600}{6608} = 0{,}766$$

$$K_k = \frac{78 \cdot 68 - 32 \cdot 22}{\sqrt{(78+22)\cdot(22+68)\cdot(78+32)\cdot(32+68)}} = \frac{5304 - 704}{\sqrt{99000000}} = 0{,}46$$

Таким образом, связь между участием в забастовках и их образовательным уровнем имеет место, но не столь существенна.

Когда каждый из качественных признаков состоит более чем из двух групп, то для определения тесноты связи возможно применение **коэффициента взаимной сопряженности Пирсона-Чупрова.** Этот коэффициент вычисляется по следующей формуле:

$$K_п = \sqrt{\frac{\varphi^2}{1+\varphi^2}}; \qquad K = \sqrt{\frac{\varphi^2}{\sqrt{(K_1-1)\cdot(K_2-1)}}} \qquad (8.17)$$

где φ^2 - показатель взаимной сопряженности;

φ - определяется как сумма отношений квадратов частот каждой клетки таблицы к произведению итоговых частот, соответствующего столбца и строки. Вычитая из этой суммы «1», получим величину φ^2:

$$\varphi^2 = \sum \frac{n_{xy}^2}{n_x n_y} - 1;$$

K_1 - число значений (групп) первого признака;

K_2 - число значений (групп) второго признака.

Чем ближе величина $K_п$ и $K_ч$ к 1, тем теснее связь.

Таблица 8.9

Вспомогательная таблица для расчета коэффициента взаимной сопряженности

y \ x	I	II	III	Всего
I			n_{xy}	n_x
II				n_x
III				n_x
Итого	n_y	n_y	n_y	n

$$1+\varphi^2 = \sum \frac{\sum \frac{n_{xy}^2}{n_x}}{n_y} = \sum \frac{\sum \frac{n_{xy}^2}{n_y}}{n_x}$$

Пример. С помощью коэффициента взаимной сопряженности исследуем связь между себестоимостью продукции и накладными расходами на реализацию.

Таблица 8.10

Зависимость между себестоимостью продукции и накладными расходами на реализацию

Накладные расходы	Себестоимость			Итого
	Низкая	Средняя	Высокая	
Низкие	19	12	9	40
Средние	7	18	15	40
Высокие	4	10	26	40
Итого	30	40	50	120

$$1+\varphi^2=\frac{\frac{19^2}{30}+\frac{12^2}{40}+\frac{9^2}{50}}{40}+\frac{\frac{7^2}{30}+\frac{18^2}{40}+\frac{15^2}{50}}{40}+\frac{\frac{4^2}{30}+\frac{10^2}{40}+\frac{26^2}{50}}{40}=$$

$$=0{,}431+0{,}356+0{,}414=1{,}183$$

$1+\varphi^2=1{,}183$; $\varphi^2=0{,}183$

$K_n=\sqrt{\frac{0{,}183}{1{,}183}}=\sqrt{0{,}155}=0{,}39$; $K_ч=\sqrt{\frac{0{,}183}{\sqrt{2\cdot 2}}}=0{,}21$.

Связь средняя.

Особое значение для оценки связи имеет **биссериальный коэффициент корреляции,** который дает возможность оценить связь между качественным альтернативным и качественным варьирующим признаками. Данный коэффициент вычисляется по формуле:

$$r=\frac{|\overline{y_2}-\overline{y_1}|}{\sigma_y}\cdot\frac{pq}{Z} \tag{8.18}$$

где $\overline{y_2}$ и $\overline{y_1}$ - средние в группах;

σ_y - среднее квадратическое отклонение фактических значений признака от среднего уровня;

p - доля первой группы;

q - доля второй группы;

Z - табулированные (табличные) значения Z-распределения в зависимости от p.

Пример. Уровень дохода сотрудников одной коммерческой структуры характеризуется следующими данными:

Таблица 8.11

Зависимость уровня доходов сотрудников коммерческой структуры от уровня их образования

Уровень образования	Уровень доходов, (руб.)				Всего (чел.)
	200-300	300-400	400-500	500-600	
	250	350	450	550	
Имеют высшее образование	5	7	6	4	22
Не имеют высшего образования	9	4	2	1	16
Итого	14	11	8	5	38

$$\overline{y_1} = \frac{250 \cdot 5 + 350 \cdot 7 + 450 \cdot 6 + 550 \cdot 4}{22} = \frac{8600}{22} = 390{,}9$$

$$\overline{y_1} = \frac{250 \cdot 9 + 350 \cdot 4 + 450 \cdot 2 + 550 \cdot 1}{16} = \frac{5100}{16} = 318{,}8$$

$$\bar{y}_{общ} = \frac{250 \cdot 14 + 350 \cdot 11 + 450 \cdot 8 + 550 \cdot 5}{38} = \frac{13700}{38} = 360{,}5$$

$$\sigma = 104{,}7\,;\ Z_{табл} = 0{,}3975$$

$p = \frac{22}{38} = 0{,}58$; $q = 0{,}42$; $p \cdot \frac{q}{Z} = 0{,}58 \cdot \frac{0{,}42}{0{,}3975} = 0{,}61$

$r = \frac{|318{,}8 - 390{,}9|}{104{,}7} \cdot 0{,}61 = 0{,}42$

Величина биссериального коэффициента корреляции также подтверждает умеренную тесноту связи между изучаемыми признаками.

8.7. Ранговые коэффициенты связи

В анализе социально-экономических явлений часто приходится прибегать к различным условным оценкам с помощью рангов, а взаимосвязь между отдельными признаками измерять с помощью непараметрических коэффициентов связи.

Ранжирование - это процедура упорядочения объектов изучения, которая выполняется на основе предпочтения.

Ранг - это порядковый номер значений признака, расположенных в порядке возрастания или убывания их величин. Если значения признака имеют одинаковую количественную оценку, то ранг всех этих значений принимается равным средней арифметической из соответствующих номеров мест, которые определяют. Данные ранги называются связными.

Среди непараметрических методов оценки тесноты связи наибольшее значение имеют ранговые коэффициенты Спирмена(ρ) и Кендалла (τ). Эти коэффициенты могут быть использованы для определения тесноты связи как между количественными, так и между качественными признаками.

Коэффициент корреляции рангов (коэффициент **Спирмена**) рассчитывается по формуле:

$$\rho_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \qquad (8.19)$$

где d_i^2 - квадраты разности рангов;

n - число наблюдений (число пар рангов).

Коэффициент Спирмена принимает любые значения в интервале $[-1;\ 1]$.

Пример. По данным группам предприятий, выставившим акции на чековые аукционы в 1996 г. определить с помощью коэффициента Спирмена зависимость между величиной уставного капитала и количеством выставленных акций.

Таблица 8.12

Расчет коэффициента Спирмена

№ предприятия	Уставной капитал (млн. руб.) (X)	Число выставленных акций (Y)	Ранги		Разность рангов $d_i = R_x - R_y$	d_i^2
			R_x	R_y		
1	2	3	8	9	10	11
1	2954	856	9	7	2	4
2	1605	930	1	9	-8	64
3	4102	1563	10	10	0	0
4	2350	682	6	5	1	1
5	2625	616	7	3	4	16
6	1795	495	4	2	2	4
7	2813	815	8	6	2	4
8	1751	858	3	8	-5	25
9	1700	467	2	1	1	1
10	2264	661	5	4	1	1
$\sum$						120

$$\rho_{x/y} = 1 - \frac{6 \cdot 120}{10 \cdot 99} = 1 - \frac{720}{990} = 0{,}3 \text{ (связь слабая).}$$

Ранговый коэффициент корреляции Кендалла (τ) также может использоваться для измерения взаимосвязи между качественными и количественными признаками, характеризующими однородные объекты и ранжированные по одному принципу. Расчет рангового коэффициента Кендалла осуществляется по формуле:

$$\tau = \frac{2S}{n(n-1)} \qquad (8.20)$$

где n - 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 Y располагаются в порядке, соответствующем значениям X;
3. Для каждого ранга Y определяется число следующих за ним значений рангов, превышающих его величину. Суммируя таким образом числа определяется величина P, как мера соответствия последовательностей рангов по X и Y и учитывается со знаком (+);
4. Для каждого ранга Y определяется число следующих за ним значений рангов, меньших его величины. Суммарная величина обозначается через Q и фиксируется со знаком (-);
5. Определяется сумма баллов по всем членам ряда.

В приведенном примере (таблица 8.12)

$P = 1 + 8 + 1 + 6 + 4 + 3 + 3 + 2 + 1 = 29$

$Q = (-8) + 0 + (-6) + 0 + (-1) + (-1) + 0 + 0 + 0 = -16$

Таким образом:

$$\tau = \frac{2 \cdot (29 - 16)}{10 \cdot (10 - 1)} = 0{,}28$$

что свидетельствует о практическом отсутствии связи между рассматриваемыми признаками.

Как правило, коэффициент Кендалла меньше коэффициента Спирмена. При достаточно большом объеме совокупности значения данных коэффициентов имеют следующую зависимость:

$$\tau = \frac{2}{3}\rho_{x/y}$$

Связь между признаками признается статистически значимой, если значения коэффициентов ранговой корреляции Спирмена и Кендалла больше 0,5.

Для определения тесноты связи между произвольным числом ранжированных признаков применяется множественный коэффициент ранговой корреляции (коэффициент конкордации) W , который вычисляется по формуле:

$$W = \frac{12S}{m^2 \cdot (n^3 - n)} \qquad (8.21)$$

где m - количество факторов

n - число наблюдений

S - отклонение суммы квадратов рангов от средней квадратов рангов.

Пример. Одновременно с проведенными выше расчетами определялась теснота связи между уставным капиталом, числом выставленных акций и числом занятых на этих предприятиях.

Таблица 8.13

Расчет коэффициента конкордации

№ предприятия	Уставной капитал (млн. руб.) (X)	Число выставленных акций (Y)	Число занятых на предприятиях (Z)	R_x	R_y	R_Z	Сумма строк	Квадраты сумм
1	2954	856	119	9	7	1	17	289
2	1605	930	125	1	9	2	12	144
3	4102	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	1795	495	165	4	2	6	12	144
7	2813	815	178	8	6	7	21	441
8	1751	858	181	3	8	8	19	361
9	1700	467	201	2	1	9	12	144
10	2264	661	204	5	4	10	19	361
$\sum$							165	2863

$$S = 2863 - \frac{(165)^2}{10} = 2863 - 2722{,}5 = 140{,}5$$

$$W = \frac{12S}{m^2(n^3 - n)} = \frac{12 \cdot 140{,}5}{9(1000 - 10)} = 0{,}19\,,$$

что свидетельствует о слабой связи между рассматриваемыми признаками.

Ранговые коэффициенты Спирмена, Кендалла и конкордации имеют то преимущество, что с помощью их можно измерять и оценивать связи как между количественными так и между атрибутивными признаками, которые поддаются ранжированию.

Глава 9. Статистическое изучение динамики социально-экономических явлений

9.1. Понятие и классификации 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.

Составными элементами ряда динамики являются показатели уровней ряда и показатели времени (годы, кварталы, месяцы, сутки) или моменты (даты) времени.

Уровни ряда обычно обозначаются через «y», моменты или периоды времени, к которым относятся - 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 таблицы 9.1:

Таблица 9.1

Число построенных квартир предприятиями и организациями всех форм собственности и их средний размер

	1980	1985	1992	1993	1994
Число квартир, тыс.	1190	1151	682	682	630
Их средний размер квартиры, $м^2$ общей площади	49,9	54,4	60,8	61,3	61,9
Удельный вес жилой площади в общей площади квартир, процентов	62,7	60,7	60,0	60,1	60,1

В таблице 9.1 рядом динамики абсолютных величин являются данные первой строки; рядом средних величин - второй строки; рядом относительных величин - третьей строки.

2) В зависимости от того 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.

Примером моментного ряда может служить ряд динамики, показывающий число вкладов населения в учреждениях сберегательного банка РФ (на конец года, млн.):

1990 г.	1991 г.	1992 г.	1993 г.	1994 г.
124,9	141,0	203,7	210,9	234,2

Уровни этого ряда - обобщающие итоги статистики вкладов населения по состоянию на определенную дату (конец каждого года).

Интервальные ряды динамики содержат данные о производстве продукции по месяцам или по годам, о товарообороте, о числе родившихся за период и т. п.

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, так как, например, часть вкладов населения, учтенных в 1990 г., существуют и в настоящее время, являясь единицами совокупности и в 1994 г. Все это делает бессмысленным суммирование уровней моментных рядов динамики.

3) В зависимости от расстояния между уровнями, ряды динамики подразделяются на ряды **с равноотстоящими уровнями и неравноотстоящими уровнями во времени**. Ряды динамики следующих друг за другом периодов или следующих через определенные промежутки дат называется равноотстоящими (см. пример о числе вкладов в сберегательные банки РФ за 1990-1994 гг.). Если же в рядах даются прерывающиеся периоды или неравномерные промежутки между датами, то ряды называются неравноотстоящими (см. пример в таблице 9.1).

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, т.к. содержат основную тенденцию развития, но их можно преобразовать в стационарные путем исключения тенденций.

9.2.Сопоставимость уровней и смыкание рядов динамики

Важнейшим условием правильного построения ряда динамики являются 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 могут охватывать значительные периоды времени, за которые могли произойти изменения, приводящие к несопостовимости статистических рядов. Рассмотрим основные причины несопоставимости уровней ряда динамики.

Несопоставимость уровней ряда может возникнуть вследствие изменения **единиц измерения и единиц счета**. Нельзя сравнивать и анализировать цифры о производстве тканей, если за одни годы оно дано в погонных метрах, а за другие - в квадратных метрах.

На сопоставимость уровней ряда динамики непосредственно влияет **методология учета или расчета показателей**. Например, если в они 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 закономерностей явления. Поэтому научный подход к изучению рядов динамики заключаются в том, чтобы ряды, охватывающие большие периоды времени, расчленять 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 явлений.

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.

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 так далее.

Следовательно, прежде чем анализировать динамический ряд, надо, исходя из цели исследования, убедится в сопоставимости уровней ряда и, если последняя отсутствует, добиться ее дополнительными расчетами. Для того, чтобы привести уровни ряда динамики к сопоставимому виду, иногда приходится прибегать к приему, который носит название **смыкания рядов динамики**. Под смыканием понимают объединение в один ряд (более длинный) двух или нескольких рядов динамики, уровни которых исчислены по разной методологии или в разных территориальных границах. 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 следующие данные о произведенной продукции, методика получения которых в течение рассматриваемого периода претерпела некоторые изменения.

Таблица 9.2

Динамика объема продукции

Годы	1988	1989	1990	1991	1992	1993	1994	1995
Объем продукции (млн. руб.) по старой методике по новой методике	 19,1 -	 19,7 -	 20,0 -	 21,2 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 1991 г.	90,1	92,9	94,3	100,0	103,5	107,5	114,9	123,2

Чтобы проанализировать динамику объема продукции за 1988-1995 гг., необходимо сомкнуть (объединить) приведенные выше два ряда в один. А чтобы уровни нового ряда были сопоставимы, необходимо пересчитать данные 1988-1991 гг. по новой методике. Для этого на основе данных об объеме продукции за 1991 г. в новой и старой методике находим соотношение между ними: 22,8 : 21,2=1,1. Умножая на полученный коэффициент данные за 1988-1991 гг. приводим их таким образом в сопоставимый вид с последующими уровнями. Сомкнутый (сопоставимый) ряд динамики показан в предпоследней строке таблицы 9.2.

Другой способ смыкания рядов динамики заключается в том, что уровни года, в котором произошли изменения (в нашем примере - уровни 1991 г.), как до изменений, так и после изменений (для нашего примера 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 нашем примере 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 таблицы 9.2.

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, вопрос о сопоставимости методики расчета сравниваемых показателей. В таких случаях ряды динамики приводятся к одному основанию, то есть 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 следующие данные (табл. 9.3.):

Таблица 9.3

Производство цемента, млн. т

Год	1991	1992	1993	1994	1995
Страна А	45,5	72,4	95,2	122,0	128,0
Страна Б	56,1	65,1	66,5	65,0	67,0

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 1991 г., получим следующие данные (табл. 9.4.):

Таблица 9.4

Темпы роста производства цемента ,
в % к 1991г.

Год	1991	1992	1993	1994	1995
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.

9.3. 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 происходит сравнение - базисным.

Абсолютный прирост (Δ_y) характеризует размер увеличения (или уменьшения) уровня ряда за определенный промежуток времени. Он равен разности двух сравниваемых уровней и выражает абсолютную скорость роста:

$$\Delta_i = y_i - y_{i-k} \quad (i=1,2,3,...,n) \qquad (9.1)$$

Если k=1, то уровень y_{i-1} является предыдущим для данного уровня, а абсолютные приросты изменения уровня будут цепными. Если же k постоянны 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 ненужно 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:

$$T_{р_б} = \frac{y_i}{y_1}100 \text{ или } T_{р_ц} = \frac{y_i}{y_{i-1}}100 \qquad (9.2)$$

В первом случае говорят о базисных темпах роста, во втором - о цепных темпах роста.

Наряду с темпом роста можно рассчитать показатель **темпа прироста**, характеризующий относительную скорость изменения уровня ряда в единицу времени. Темп прироста показывает, на какую долю (или процент) уровень данного периода или момента времени больше (или меньше) базисного уровня.

Темп прироста есть отношение абсолютного прироста к уровню ряда, принятого за базу:

$$T_{пр_ц} = \frac{\Delta_{i/i-1}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} \times 100 = (Кр_{i/i-1} - 1) \times 100 = Тр_{i/i-1} - 100 \quad (9.3)$$

Если темп роста всегда положительное число, то темп прироста может быть положительным, отрицательным и равным нулю.

В статистической практике часто вместо расчета и анализа темпов роста и прироста рассматривают **абсолютное значение одного процента прироста**. Оно представляет собой одну сотую часть базисного уровня и в то же время - отношение абсолютного прироста к соответствующему темпу прироста:

$$|\%| = \frac{\Delta_{i/i-1}}{Тпр_{i/i-1} \times \%} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}}100} = \frac{y_{i-1}}{100} = 0{,}01 \times y_{i-1} \qquad (9.4)$$

где |%| - обозначение абсолютного значения одного процента прироста.

Для иллюстрации расчетов рассмотренных статистических показателей приведем следующий ряд динамики в таблице 9.5.

Средний уровень ряда динамики ($\bar{y}$) рассчитывается по средней хронологической. **Средней хронологической** называется средняя, исчисленная из значений, изменяющихся во времени. Такие средние обобщают хронологическую вариацию. В хронологической средней отражается совокупность тех условий, в которых развивалось изучаемое явление в данном промежутке времени.

Таблица 9.5

Динамика производства газа в регионе за 1991-1995 гг.
(цифры условные)

Годы	Производство	Абсолютный прирост (млн. $м^3$)		Темп роста, в %		Темп прироста, в %		Абсолютное значение одного процента прироста, млн. $м^3$
	млн. $м^3$	по сравнению с предыдущим годом	по сравнению с 1991 г.	по сравнению с предыдущим годом	по сравнению с 1991г.	по сравнению с предыдущим годом	по сравнению с 1991 г.	
А	1	2	3	4	5	6	7	8
1991	289	-	-	-	100,0	-	-	-
1992	321	32	32	111,1	111,1	11,1	11,1	2,9
1993	346	25	57	107,8	119,7	7,8	19,7	3,2
1994	372	26	83	107,5	128,7	7,5	28,7	3,4
1995	407	35	118	109,4	140,8	9,4	40,8	3,7
Итого	1735	118	-	-	-	-	-	-

Методы расчета среднего уровня интервального и моментного рядов динамики различны. Для интервальных равноотстоящих рядов средний уровень находится по формуле средней арифметической простой и для неравноотстоящих рядов по средней арифметической взвешенной:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad (9.5.)$$

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i} \qquad (9.6.)$$

где y_i - уровень ряда динамики;
n - число уровней;
t_i - длительность интервала времени между уровнями.

Так, в таблице 9.5. приведен интервальный ряд динамики с равноотстоящими уровнями. По этим данным можно рассчитать среднегодовой уровень производства газа за 1991-1995 гг. Он будет равен 347 млн.м^3, то есть ($\bar{y}$=1735/5).

Средний уровень моментного ряда динамики так исчислить нельзя, так как отдельные уровни содержат элементы повторного счета. Средний уровень моментного равноотстоящего ряда динамики находится по формуле средней хронологической:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \frac{y_3 + y_4}{2} + \ldots + \frac{y_n + y_{n-1}}{2}}{n-1} =$$

$$= \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1} \quad \text{или} \quad \bar{y} = \frac{\frac{y_1 + y_n}{2} + \sum_{i=2}^{n-1} y_i}{n-1} \qquad (9.7)$$

Средний уровень моментных рядов динамики с неравноотстоящими уравнениями определяются по формуле средней хронологической взвешенной:

$$\bar{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \ldots + (y_{n-1} + y_n)t_{n-1}}{2(t_1 + t_2 + \ldots + t_{n-1})} =$$

$$\frac{\sum_{i=1}^{n-1} (y_i + y_{i+1})t_i}{2\sum_{i=1}^{n-1} t_i} \qquad (9.8)$$

где y_i, y_n - уровни рядов динамики;
t_i - длительность интервала времени между уровнями.

Покажем расчет среднего уровня моментного ряда динамики с равноотстоящими уровнями.

Например, если известны товарные остатки магазина на 1-ое число каждого месяца (тыс. руб.):

1/I	1/II	1/III	1/IV
18	14	16	20

то среднемесячный товарный остаток за 1 квартал по формуле 9.7. составит

$$\bar{y} = \frac{18/2 + 14 + 16 + 20/2}{3} = \frac{49}{3} = 16{,}3 \text{ тыс. руб.}$$

Другой пример. Известна списочная численность рабочих организации на некоторые даты 1994 г. (чел):

1/I	1/III	1/VI	1/IX	1/I-1995
1200	1100	1250	1500	1350

Среднегодовая численность работников за 1994 г. по формуле 9.8 составит:

$$\bar{y} = \frac{(1200+1100)2 + (1100+1250)3 + (1250+1500)3 + (1500+1350)4}{2(2+3+3+4)} =$$

$$\frac{31300}{24} \approx \text{чел}$$

Обобщающим показателем скорости изменения явления во времени является **средний абсолютный прирост** за весь период, ограничивающий ряд динамики. Для его определения можно воспользоваться формулой средней арифметической простой:

$$\bar{\Delta} = \frac{\sum_{n-1}^{n-1} \Delta_{i/i-1}}{n-1} \tag{9.9.}$$

или

$$\bar{\Delta} = \frac{y_n - y_1}{n-1} \tag{9.10.}$$

Так, для условтй нашего примера (см. таблицу 9.2.) средний абсолютный прирост равен 29,5млн.м^3 [(407-289)/4].

Свободной обощающей 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 нужно определять 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 в геометрической ($a, aq, aq^2,...,aq^n$), которая характеризуется постоянным отношением, называемым знаменателем прогрессии (q). Вопрос, следовательно, состоит в том, чтобы найти этот знаменатель. Знаменатель геометрической прогрессии (q) определяется делением последующего уровня прогрессии на его предыдущий. при делении n-го уровня на первый, получаем:

$$\frac{aq^{n-1}}{a} = q^{n-1}$$

отсюда следует:

$$q = \sqrt[n-1]{\frac{aq^{n-1}}{a}} = \sqrt[n-1]{\frac{b_n}{b_1}} \qquad (9.11.)$$

где $b_1=a$ - первый член прогрессии.

Зная q, мы точно можем определить какую тенденцию развития явления имеет неометрическая последовательность, которая применяется тогда, когда определяющий показатель является не суммой значений, а их произведением. Следовательно, во всех тех случаях, где варианты связаны между собой не знаком сложения, а знаком произведения, можно вычислить среднюю геометрическую. Обычно средний темп роста вычисляется по формуле средней неометрической из цепных коэффициентов роста:

$$\overline{\text{T}}\text{p}\_\_ = \sqrt[m]{k_{2/1}k_{3/2}...k_{n/n-1}} = \sqrt[m]{\Pi K_{P_{i/i-1}}} \qquad (9.12.)$$

Например, средний темп роста производства газа за 1991-1995 гг. (см. таблицу 9.5) равен:

$$\overline{Tp_y} = \sqrt[4]{1{,}111 \cdot 1{,}078 \cdot 1{,}075 \cdot 1{,}094} = \sqrt[4]{1{,}41} = 1{,}089 \text{ или } 108{,}9\%$$

Поскольку всякий темп роста является отношением уровней ряда динамики, так, что $Tp_{2/1} = \frac{y_2}{y_1}100$; $\text{Tp}_{3/2} = \frac{y_3}{y_2}100...$ в формуле средней геометрической темпы роста заменяются соответствующим отношением уравнений. Заменив темпы роста выражающими их отношениями и учитывая, что эти величины перемножаются, найдем подкоренное выражение как:

$$\frac{y_2}{y_1} \times \frac{y_3}{y_2} \times \frac{y_4}{y_3} \times \ldots \times \frac{y_n}{y_{n-1}} = \frac{y_n}{y_1}$$

Следовательно, средний темп роста может быть выражен формулой:

$$\overline{Tp_y} = \sqrt[n-1]{\frac{y_n}{y_1}} \qquad (9.13)$$

Продолжим наш пример (см. таблицу 9.5). Средний темп роста производства газа за 1991-1995 гг. будет равен:

$$\overline{T}p_y = \sqrt[4]{\frac{407}{289}} = \sqrt[4]{1{,}41} = 1{,}089 \quad \text{или} \quad 108{,}9\%$$

Когда приходится вести расчет средних темпов роста по периодам различной продолжительности (разноотстоящие ряды динамики), то пользуются средними геометрическими, взвешенными по продолжительности периодов. Формула средней геометрической взвешенной будет иметь вид:

$$\overline{Tp_y} = \sqrt[\Sigma t]{(k_{2/1})^{t_1} \ (k_{3/2})^{t_2} \ldots..(k_{n/n-1})^{t_{n-1}}} \qquad (9.14)$$

где t - интервал времени, в течении которого сохраняется данный темп роста;
Σ - 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%:

$$\overline{\text{T}}\text{пр}_\text{y} = \overline{\text{T}}\text{р} - 100 \qquad (9.15)$$

9.4. Методы анализа основной тенденции (тренда) в рядах динамики

Важной задачей статистики при анализе рядов динамики является определение основной тенденции развития, присущей тому или иному ряду динамики. Например, за колебаниями урожайности какой-либо сельскохозяйственной культуры в отдельные годы тенденция роста (уменьшения) урожайности может не просматриваться непосредственно, и поэтому должна быть выявлена статистическими методами.

Методы анализа основной тенденции в рядах динамики 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 методы каждой группы.

Метод укрупнения интервалов. Если рассматривать уровни экономических показателей за короткие промежутки времени, то в силу влияния различных факторов, действующих в разных направлениях, в рядах динамики наблюдается снижение и повышение этих уровней. Это мешает видеть основную тенденцию развития изучаемого явления. В этом случае для наглядного представления тренда применяется метод укрупнения интервалов, который основан 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 и уровень вначале и добавляя один следующий. Отсюда название - **скользящая средняя**.

Каждое звено скользящей средней - это средней уровень за соответствующий период, который относится к **середине выбранного периода,** если число уровней ряда динамики нечетное.

Нахождение скользящей средней по четному числу членов рядов динамики несколько сложнее, так как средняя может быть отнесена только к середине между двумя датами, находящимися в середине интервала сглаживания. Например, средняя, найденная для четырех членов, относится к середине между вторым и третьим, третьим и четвертым уровнями и так далее. Чтобы ликвидировать такой сдвиг, применяют так называемый **способ центрирования**. **Центрирование** заключается в нахождении средней из двух смежных скользящих средних для отнесения полученного уровня к определенной дате. При центрировании необходимо находить скользящие суммы, скользящие средние нецентрированные по этим суммам и средние из двух смежных нецентрированных скользящих средних.

Покажем расчет 5-летней и 4-летней скользящей средней на примере данных таб. 9.6.

Таблица 9.6

Сглаживание урожайности зерновых культур в хозяйстве за 1980-1995 гг. методом скользящей средней

Годы	Центнеров с га	Скользящие пятилетние суммы	Пятилетние скользящие средние	Скользящие четырехлетние суммы	Четырехлетние скользящие средние (нецентрированные)	Четырехлетние скользящие средние (центрированные)
А	1	2	3	4	5	6
1980	9,5	-	-	-	-	-
1981	13,7	-	-	-	-	-
					12,3	
1982	12,1	-	12,5	-		12,8
					13,2	
1983	14,0	-	13,7	49,3		13,5
					13,7	
1984	13,2	63,5	14,1	53,0		14,1
					14,6	
1985	15,6	68,6	14,4	54,9		14,6
					14,6	
1986	15,4	70,3	15,2	58,2		15,1
					15,7	
1987	14,0	72,2	15,6	58,2		15,6
					15,6	
1988	17,6	75,8	14,7	62,6		15,0
					14,5	
1989	15,4	78,0	15,1	62,4		14,9
					15,3	
1990	10,9	73,5	15,3	57,9		15,0
					14,7	
1991	17,5	75,4	15,5	61,4		15,1
					15,5	
1992	15,0	76,4	15,2	58,8		15,8
					16,3	
1993	18,5	77,3	16,0	61,9		15,97
					15,65	
1994	14,2	76,1	-	65,2		-
					-	
1995	14,9	80,1	-	62,6		-

Недостаток метода простой скользящей средней состоит в том, что сглаженный ряд динамики сокращается ввиду невозможности получить сглаженные уровни для начала и конца ряда. Этот недостаток устраняется применением метода аналитического выравнивания для анализа основной тенденции.

Аналитическое выравнивание предполагает представление уровней данного ряда динамики в виде функции времени - y=f(t).

Для отображения основной тенденции развития явлений во времени применяются различные функции: полиномы степени, экспоненты, логистические кривые и другие виды.

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-ой степени $\bar{y}_t = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n$

Здесь a_0; a_1; a_2; ... a_n - параметры полиномов, t - условное обозначение времени. В статистической практике параметры полиномов невысокой степени иногда имеют конкретную интерпретацию характеристик динамического ряда. Так, параметр a_0 трактуется как характеристика средних условий ряда динамики, параметры a_1, a_2, a_3 - как изменения 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.

После выбора вида уравнения необходимо определить параметры уравнения. Самый распространенный способ определения параметров уравнения - это **метод наименьших квадратов**.

Суть данного метода изложена в главе 8.

Согласно этому методу, для нахождения параметров полинома р-й степени необходимо решить систему так называемых нормальных уравнений:

$$\begin{cases} na_0 + a_1\sum t + a_2\sum t^2 + ... + a_p\sum t^p = \sum y \\ a_0\sum t + a_1\sum t^2 + a_2\sum t^3 + ... + a_p\sum t^{p+1} = \sum yt \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0\sum t^p + a_1\sum t^{p+1} + a_2\sum t^{p+2} + ... + a_p\sum t^{2p} = \sum yt^p \end{cases} \qquad (9.16)$$

где n - число членов в ряду динамики: t=1,2,...,n

Система 9.16, состоящая из «р» уравнений, содержит в качестве известных величин $\sum y,\ \sum yt,...,\sum yt^p$, то есть суммы наблюдаемых значений уровней динамического ряда, умноженные на показатели времени в степени 0,1,2,...,р и неизвестных величин a_j. Решение этой системы относительно a_0, a_1,...,a_p и дает искомые значения параметров.

Системы для расчета параметров полиномов невысоких степеней намного проще. Обозначим последовательные параметры полиномов как a_0, a_1, a_2. Тогда системы нормальных уравнений для оценивания параметров прямой $\bar{y}_t = a_0 + a_1 t$ примет вид:

$$\begin{cases} na_0 + a_1\sum t = \sum y \\ a_0\sum t + a_1\sum t^2 = \sum yt \end{cases} \qquad (9.17)$$

для параболы второго порядка ($y_t=a_0+a_1t+a_2t^2$):

$$\begin{cases} na_0 + a_1\sum t + a_2\sum t^2 = \sum y \\ a_0\sum t + a_1\sum t^2 + a_2\sum t^3 = \sum yt \\ a_0\sum t^2 + a_1\sum t^3 + a_2\sum t^4 = \sum yt^2 \end{cases} \qquad (9.18)$$

Решение системы (9.17) относительно искомых параметров a_0 и a_1 дает:

$$a_0 = \frac{\sum y\sum t^2 - \sum t\sum yt}{n\sum t^2 - \left(\sum t\right)^2}$$

$$a_1 = \frac{n\sum yt - \sum y\sum t}{n\sum t^2 - \left(\sum t\right)^2}$$

В статической практике применяется упрощенный расчет параметров уравнений, который заключается в переносе начала координат в середину ряда динамики. В этом случае упрощаются сами нормальные уравнения, кроме того уменьшаются абсолютные значения величин, участвующих в расчете. В самом деле, если до переноса начала координат t было равно 1,2,3,...,n, то после переноса t=...-4,-3,-2,-1,0,1,2,3,4,..., если число членов ряда нечетное. Если же число членов ряда четное, то t=...,-5,-3,-1,1,3,5,... Следовательно, Σt и все Σt^p у которых «р» – нечетное число, равны 0. Таким образом, все члены уравнений, содержащие Σt с такими степенями могут быть исключены. Системы нормальных уравнений теперь упрощаются для прямой:

$$\begin{cases} a_0 n = \sum y \\ a_1 \sum t^2 = \sum ty \end{cases} \tag{9.19}$$

для параболы второго порядка:

$$\begin{cases} a_0 n + a_2 \sum t^2 = \sum y \\ a_1 \sum t^2 = \sum ty \\ a_0 \sum t^2 + a_2 \sum t^4 = \sum t^2 y \end{cases} \tag{9.20}$$

Решая системы (9.19), (9.20) относительно неизвестных параметров, получим величины параметров соответствующих полиномов.

Параметр a_1 выражает начальную скорость роста, а коэффициент a_2 - постоянную скорость изменения прироста.

При сглаживании ряда динамики по показательной кривой ($y_t = a_0 a_1^t$) для определения параметров применяется метод наименьших квадратов к логарифмам исходных данных. Так, для нахождения параметров показательной функции необходимо решить следующую систему уравнений:

$$\begin{cases} \sum \lg y = n \lg a_0 + \lg a_1 \sum t \\ \sum t \lg y = \lg a_0 \sum t + \lg a_1 \sum t^2 \end{cases} \tag{9.21}$$

Если Σt=0, то параметры уравнения $\lg a_0$ и $\lg a_1$ находим по формулам: $\lg a_0 = \frac{\sum \lg y}{n}$; $\lg a_1 = \frac{\sum t \lg y}{\sum t^2}$.

Рассмотрим следующий пример. Необходимо определить основную тенденцию ряда динамики урожайности зерновых культур в хозяйстве за 1981-1995 гг. по следующим данным (см. табл. 9.7).

Начнем определение тенденции с самого простого полинома-уравнение прямой (9.19). Решая систему нормальных уравнений, получим искомые параметры: a_0=14,8; a_1=0,17, а само уравнение запишется следующим образом $\bar{y}_t = 14{,}8 + 0{,}17t$, что выражает тенденцию динамики урожайности зерновых культур в 1981-1995 гг., т.е. в течение исследуемого периода урожайность зерновых культур в хозяйстве увеличивалась в среднем на 0,17 ц. с га в год.

Таблица 9.7

Динамика урожайности зерновых культур в хозяйстве
и определение параметров уравнения методом наименьших квадратов

Годы	Урожайность ц. с га (у)	t	t^2	$y_i t$	y_t
1981	13,7	-7	49	-95,5	13,6
1982	12,1	-6	36	-72,6	13,8
1983	14,0	-5	25	-70,0	13,9
1984	13,2	-4	16	-52,8	14,1
1985	15,6	-3	9	-46,8	14,3
1986	15,4	-2	4	-30,8	14,5
1987	14,0	-1	1	-14,0	14,6
1988	17,6	0	0	0	14,8
1989	15,4	1	1	15,4	15,0
1990	10,9	2	4	21,8	15,1
1991	17,5	3	9	52,5	15,3
1992	15,0	4	16	60,0	15,5
1993	18,5	5	25	92,5	15,7
1994	14,2	6	36	85,2	15,8
1995	14,9	7	49	104,3	16,0
Итого	222,0	0	280	48,8	222,0

9.5. Методы выявления сезонной компоненты

При рассмотрении квартальных или месячных данных многих социально-экономических явлений часто обнаруживаются определенные, постоянно повторяющиеся колебания, которые существенно не изменяются за длительный период времени. Они являются результатом влияния природно-климатических условий, общих экономических факторов, а также ряда многочисленных разнообразных факторов, которые частично являются регулируемыми. В статистике периодические колебания, которые имеют определенный и постоянный период, равный годовому промежутку, носят название «сезонных колебаний» или «сезонных волн», а динамический ряд в этом случае называют **тренд-сезонным**, или просто **сезонным рядом динамики**.

Сезонные колебания характеризуются специальным 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 (обычно не менее трех) берутся для того, чтобы выявить устойчивую сезонную волну, на которой не отражались бы случайные условия одного года.

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 ($\bar{y}_i$), затем из них рассчитывается среднемесячный уровень для всего ряда ($\bar{y}$) и в заключение определяется процентное отношение средних для каждого месяца к общему среднемесячному уровню ряда, то есть:

$$I_S = \frac{y_i}{\bar{y}} 100\% \qquad (9.22)$$

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:

- по соответствующему полиному вычисляются для каждого месяца (квартала) выровненные уровни на момент времени (t);

- вычисляются отношения фактических месячных (квартальных) данных (y_i) к соответствующим выровненным данным ($\bar{y}_t$) в процентах

$$\left[I = (y_i : \bar{y}_t) \times 100\right];$$

- находятся средние арифметические из процентных соотношений, рассчитанных по одноименным периодам в процентах $I_i=(I_1+I_2+I_3+...+I_n):n$, где n - число одноименных периодов.

В общем виде формулу расчета индекса сезонности данным способом можно записать так:

$$I_S = \left[\frac{\sum y_i}{\bar{y}_t}\right] : n \qquad (9.23)$$

Расчет заканчивается проверкой правильности вычислений индексов, так как средний индекс сезонности для всех месяцев (кварталов) должен быть 100 процентов, то сумма полученных индексов по месячным данным равна 1200, а сумма по четырем кварталам - 400.

9.6. Элементы прогнозирования и интерполяции

Исследования динамики социально-экономических явлений, выявление и характеристика основной тенденции развития дают основание для прогнозирования - определения будущих размеров уровня экономического явления.

Важное место в системе методов прогнозирования занимают статистические методы. Применение прогнозирования предполагает, что закономерность развития, действующая в прошлом (внутри ряда динамики), сохранится и в прогнозируемом будущем, то есть прогноз основан на **экстраполяции**. Экстраполяция, проводимая в будущее, называется **перспективой** и в прошлое - **ретроспективой**. Обычно, говоря об экстраполяции рядов динамики, подразумевает чаще всего перспективную экстраполяцию.

В зависимости от того, какие принципы и какие исходные данные положены в основу прогноза, можно выделить следующие элементарные методы экстраполяции: **среднего абсолютного прироста**, **среднего темпа роста** и экстраполяция 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, то есть 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, то есть экстраполяцию можно сделать по следующей формуле:

$$\hat{y}_{i+t} = y_i + \overline{\Delta}t \qquad (9.24)$$

где $\hat{y}_{i+t}$ - экстраполируемой уровень, (i+t) - номер этого уровня (года);

i - номер последнего уровня (года) исследуемого периода, за который рассчитан $\overline{\Delta}$

t - срок прогноза (период упреждения);

$\overline{\Delta}$ - средний абсолютный прирост.

Прогнозирование по среднему темпу роста можно осуществлять в случае, когда есть основание считать, что общая тенденция ряда характеризуется показательной (экспоненциальной) кривой. Для нахождения тенденции в этом случае необходимо определить средний коэффициент роста, возведенный в степень, соответствующую периоду экстраполяции, то есть по формуле:

$$\hat{y}_{i+1} = y_i \cdot \overline{K}_p^t \qquad (9.25)$$

где y_i - последний уровень ряда динамики;
t - срок прогноза;
$\overline{K}_p$ - средний коэффициент роста.

Если же ряду динамики свойственна иная закономерность, то данные, полученные при экстраполяции на основе среднего темпа роста, будут отличаться от данных, полученных другими способами экстраполяции.

Рассмотренные способы экстраполяции тренда, будучи простейшими, в то же время являются и самыми приближенными.

Поэтому наиболее распространенным методом прогнозирования является **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, характеризирующего 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, то есть y=f(t).

Поэтому целесообразно определение доверительных интервалов прогноза.

Величина доверительного интервала определяется следующим образом:

$$\overline{y}_{t+1} \pm t_\alpha \sigma_t \qquad (9.26)$$

где σ - средняя квадратическая ошибка тренда;
$\overline{y}_{t+1}$ - расчетное значение уровня;
t_α - доверительная величина.

Вместо t_α - критерия удобно использовать коэффициент ($К^*$).

Например, необходимо провести прогноз на 1996-1999 гг. по данным таблицы (9.7) об урожайности зерновых культур в хозяйстве.

Для экстраполяции используем уравнение тренда, полученное по прямой: $\bar{y}_t = 14,8 + 0,17t$. Подставив соответствующее значение t в наше уравнение, получим точечные прогнозы на 1996-1999 гг. (графа 2 таблицы 9.8). Для построения интервальных прогнозов рассчитаем среднеквадратическую ошибку тренда (σ_t=1,797) и используем значения $К^*$.

Результаты прогноза представлены в таблице (9.8):

Таблица 9.8

Прогнозные значения урожайности зерновых культур в хозяйстве на 1996-1999 гг.

Годы	t	$\bar{y}_{t+1}$	К* [1]	σ_tК*	$\hat{y}_{i+T} \pm \sigma \cdot K^*$
А	1	2	3	4	5
1996	8	16,2	2,0153	3,6	12,6 - 19,8
1997	9	16,3	2,0621	3,7	12,6 - 20,0
1998	10	16,5	2,1131	3,8	12,7 - 20,3
1999	11	16,7	2,1680	3,9	12,8 - 20,6

При анализе рядов динамики иногда приходится прибегать к определению некоторых неизвестных уровней внутри данного ряда динамики, то есть к **интерполяции**.

Как и экстраполяция, интерполяция может производится на основе среднего абсолютного прироста, среднего темпа роста, а также с помощью аналитического выравнивания. При интерполяции предполагается, что ни выявленная тенденция, ни ее характер не претерпели существенных изменений в том промежутке времени, уровень (уровни) которого нам неизвестны.

[4] Значения К* взяты из книги Е.М. Четыркина «Статистические методы прогнозирования». М., 1975 г., стр. 183

Глава 10. Статистический анализ структуры

10.1. Понятие структуры и основные направления ее исследования

Изучаемые статистикой процессы и явления в сфере промышленного или сельскохозяйственного производства, финансов, коммерции, демографии, в социальной и политической областях, как правило, характеризуются внутренней структурой, которая с течением времени может изменяться. Динамика структуры вызывает изменение внутреннего содержания исследуемых объектов и их экономической интерпретации, приводит к изменению установившихся причинно-следственных связей. Именно поэтому изучение структуры и структурных сдвигов занимает важное место в экономико-статистическом анализе.

В статистике под **структурой** понимают совокупность элементов социально-экономических явлений, обладающих определенной устойчивостью внутригрупповых связей при сохранении основных свойств, характеризующих эту совокупность как целое. В качестве примеров можно привести структуру населения региона по возрасту или уровню доходов, структуру предприятий отрасли по численности промышленно-производственного персонала или стоимости основных фондов и другие.

Классификация структур прежде всего предполагает 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 (например, структура парка транспортных средств). **Интервальные** структуры характеризуют строение социально-экономических явлений за определенные периоды времени - дни, недели, месяцы, кварталы, годы (например, структура экспорта и импорта).

Статистика имеет дело как с фактическими, реально существующими структурами, так и со структурами перспективными, прогнозными, оптимальными и стандартизованными. Последние представляют собой какие-либо 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 или более регионам рассчитывают стандартизованные коэффициенты на основе некоторой стандартизованной структуры, в качестве которой может использоваться возрастная структура населения в целом по стране.

Основные направления статистического изучения структуры включают:

а) характеристику структурных сдвигов отдельных частей совокупности за два и более периодов;

б) обобщающую характеристику структурных сдвигов в целом по совокупности;

в) оценку степени концентрации и централизации.

Рассмотрим последовательно эти три направления исследования.

10.2. Частные показатели структурных сдвигов

Анализ структуры и ее изменений базируется на относительных показателях структуры - долях или удельных весах, представляющих собой соотношения размеров частей и целого. При этом как частные, так и обобщающие показатели структурных сдвигов могут отражать либо «абсолютное» изменение структуры в процентных пунктах или долях единицы (кавычки показывают, что данные показатели являются абсолютными по методологии расчета, но не по единицам измерения), либо ее относительное изменение в процентах или коэффициентах.

«Абсолютный» прирост удельного веса i-ой части совокупности показывает, на сколько процентных пунктов возросла или уменьшилась данная структурная часть в j-ый период по сравнению с (j-1) периодом:[1]

$$\Delta_{d_i} = d_{i\,j} - d_{i\,j-1}, \qquad (10.1)$$

где d_{ij} - удельный вес (доля) i-ой части совокупности в j-ый период;

d_{ij-1} - удельный вес (доля) i-ой части совокупности в (j-1)-ый период.

Знак прироста показывает направление изменения удельного веса данной структуры части («+» - увеличение, «-» - уменьшение), а его значение - конкретную величину этого изменения.

Темп роста удельного веса представляет собой отношение удельного веса i-ой части в j-ый период времени к удельному весу той же части в предшествующий период:

$$Tp_{d_i} = \frac{d_{i\,j}}{d_{i\,j-1}} \cdot 100 \qquad (10.2)$$

Темпы роста удельного веса выражаются в процентах и всегда являются положительными величинами. Однако, если в совокупности имели место какие-либо структурные изменения, часть темпов роста будет больше 100%, а часть - меньше.

[1] Здесь и далее при исследовании моментных структур под периодами будут подразумеваться моменты времени.

Рассчитаем частные показатели структурных сдвигов по данным о распределении коммерческих банков по размеру объявленного уставного фонда (табл. 10.1.):

Таблица 10.1

Группы коммерческих банков по размеру объявленного уставного фонда (млрд.руб.)	Число банков		Удельный вес, в % к итогу		Годовой прирост удельного веса, проц. пунктов Δ_{d_i}	Годовой темп роста удельного веса, % Tp_{d_i}
	1.01.95	1.01.96	1.01.95 d_{i0}	1.01.96 d_{i1}		
А	1	2	3	4		6(гр.4:гр.3)* 100
до 1	1656	1175	65,8	45,6	-20,2	69,3
1 - 5	697	892	27,7	34,6	6,9	124,9
5 - 20	134	418	5,3	16,2	10,9	305,7
20 и более	30	93	1,2	3,6	2,4	300,0
Итого	2517	2578	100,0	100,0	0	Х

Как следует из данных таблицы 10.1, наиболее существенно в «абсолютном» выражении изменился удельный вес банков с уставным фондом до 1 млрд. руб. - снизился на 20,2 процентного пункта. В относительном выражении наиболее сильно (в 3 раза) выросла доля банков с уставным фондом свыше 5 млрд. руб.

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, т.е. в расчете средних показателей структурных сдвигов.

Средний «абсолютный» прирост удельного веса i-ой структурной части показывает, на сколько процентных пунктов в среднем за какой-либо период (день, неделю, месяц, год и т.п.) изменяется данная структурная часть:

$$\overline{\Delta}_{d_i} = \frac{d_{in} - d_{i1}}{n-1}, \qquad (10.3)$$

где n - число осредняемых периодов.

Сумма средних «абсолютных» приростов удельных весов всех k структурных частей совокупности, также как и сумма их приростов за один временной интервал, должна быть равна нулю.

Средний темп роста удельного веса характеризует среднее относительное изменение удельного веса i-ой структурной части за n периодов, и рассчитывается по формуле средней геометрической:

$$\overline{T}p_{d_i} = \sqrt[n-1]{Tp_{d\,i1} \cdot Tp_{d\,i2} \cdot Tp_{d\,i3} \cdot \ldots \cdot Tp_{d\,in-1}} \qquad (10.4)$$

Подкоренное выражение этой формулы представляет собой последовательное произведение цепных темпов роста удельного веса за все временные интервалы. После проведения несложных алгебраических преобразований данная формула примет следующий вид:

$$\overline{T}p_{d_i} = \sqrt[n-1]{\frac{d_{in}}{d_{i1}}} \cdot 100 \qquad (10.5)$$

Для иллюстрации этих формул воспользуемся приведенным выше примером (таблица 10.1). Рассчитаем средний месячный прирост (в данном случае - снижение) удельного веса банков 1-ой группы:

$\overline{\Delta}_{d_1} = \dfrac{45{,}6 - 65{,}8}{12 - 1} = -1{,}8$ проц. пункта.

По этой же группе определим средний месячный темп роста удельного веса:

$$\overline{T}p_{d_1} = \sqrt[11]{\frac{45{,}6}{65{,}8}} \cdot 100 = 96{,}7\%$$

Мы получили, что удельный вес банков данной группы в среднем ежемесячно снижался на 1,8 процентного пункта или на 3,3% (96,7%-100%).

При анализе структуры исследуемого объекта или явления за ряд периодов также можно определить **средний удельный вес** каждой i-ой части за весь рассматриваемый временной интервал. Однако для его расчета одних лишь относительных данных об удельных весах структурных частей недостаточно, необходимо располагать еще и информацией о размерах этих частей в абсолютном выражении. Используя эти данные, средний удельный вес любой i-ой структурной части можно определить по формуле:

$$\overline{d_i} = \frac{\sum_{j=1}^{n} x_{ij}}{\sum_{j=1}^{n} \sum_{i=1}^{k} x_{ij}} \cdot 100\ , \qquad (10.6)$$

где x_{ij} - величина i-ой структурной части в j- период времени в абсолютном выражении.

Проиллюстрируем эту формулу следующим примером. По данным первичного рынка государственных краткосрочных облигаций (ГКО) и облигаций федерального займа (ОФЗ) в третьем квартале 1995г. определим средний удельный вес ценных бумаг каждого вида в общем объеме выручки от их реализации (табл. 10.2.):

Таблица 10.2

Вид ценных бумаг	Объем выручки от продажи			
	Июль	Август	Сентябрь	Итого
ГКО, трлн.руб. (x_{1j})	5,5	8,1	11,0	24,6
в % к итогу (d_{1j})	80,9	98,9	99,1	...
ОФЗ, трлн. руб. (x_{2j})	1,3	0,09	0,1	1,49
в % к итогу (d_{2j})	19,1	1,1	0,9	...
Всего, трлн.руб.	6,8	8,19	11,1	26,09

Определим средний удельный вес выручки от продажи ГКО в общем объеме выручки от реализации государственных ценных бумаг:

$$\bar{d} = \frac{24,6}{26,09} \cdot 100 = 94,3\% .$$

Рассчитаем средний удельный вес выручки от продажи ОФЗ:

$$\bar{d} = \frac{1,49}{26,09} \cdot 100 = 5,7\% .$$

Итак, в августе-сентябре 1995г. на долю ГКО в среднем ежемесячно приходилось 94,3% общего объема выручки от реализации государственных ценных бумаг, на долю ОФЗ - только 5,7%.

10.3. Обобщающие показатели структурных сдвигов

В отдельных случаях исследователю необходимо в целом оценить структурные изменения в изучаемом социально-экономическом явлении за определенный временной интервал, которые характеризуют подвижность, или наоборот, стабильность, устойчивость данной структуры. Как правило, это требуется для сравнения динамики одной и той же структуры в различные периоды или нескольких структур, относящихся к разным объектам. Во втором случае число структурных частей у разных объектов необязательно должно совпадать.

Среди применяемых для этой цели обобщающих показателей наиболее распространен **линейный коэффициент «абсолютных» структурных сдвигов**, представляющий собой сумму приростов удельных весов, взятых по модулю, деленную на число структурных частей:

$$\overline{\Delta}_{d_1-d_0}=\frac{\sum_{i=1}^{k}\left|d_{ij}-d_{ij-1}\right|}{k} \qquad (10.7)$$

Этот показатель отражает то среднее изменение удельного веса (в процентных пунктах), которое имело место за рассматриваемый временной интервал в целом по всем структурным частям совокупности.

Для решения данной задачи также применяют **квадратический коэффициент «абсолютных» структурных сдвигов**, который рассчитывается по формуле:

$$\sigma_{d_1-d_0}=\sqrt{\frac{\sum_{i=1}^{k}(d_{ij}-d_{ij-1})^2}{k}} \qquad (10.8)$$

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**:

$$\sigma_{\frac{d_1}{d_0}}=\sqrt{\sum_{i=1}^{k}\frac{(d_{ij}-d_{ij-1})^2}{d_{ij-1}}\cdot 100} \qquad (10.9)$$

Данный показатель отражает тот средний относительный прирост удельного веса (в процентах), который наблюдался за рассматриваемый период.

По данным таблицы 10.3 (см. в конце раздела) рассчитаем обобщающие показатели структурных сдвигов.

Для расчета линейного коэффициента «абсолютных» структурных сдвигов за первый период (с октября по ноябрь) и за второй период (с ноября по декабрь) соответственно воспользуемся данными итогов гр.4 и гр.7 таблицы 10.3 :

$\overline{\Delta}^{I}_{d_1-d_0}=\frac{7,4}{5}=1,5$ проц. пункта

$\overline{\Delta}^{II}_{d_1-d_0}=\frac{11,4}{5}=2,3$ проц. пункта

Итак, с октября по ноябрь 1995г. удельный вес отдельных направлений использования доходов населения изменился в среднем на 1,5 процентного пункта. С ноября по декабрь «абсолютные» структурные сдвиги заметно увеличились. Этот вывод подтверждается квадратическими коэффициентами «абсолютных» структурных сдвигов (необходимые промежуточные расчеты выполнены в гр.5 и гр.8 таблицы 10.3):

$$\sigma^{I}_{d_1-d_0} = \sqrt{\frac{17,66}{5}} = 1,9 \text{ проц. пункта,}$$

$$\sigma^{II}_{d_1-d_0} = \sqrt{\frac{31,34}{5}} = 2,5 \text{ проц. пункта.}$$

Далее определим величину квадратических коэффициентов относительных структурных сдвигов, воспользовавшись итоговыми данными гр.6 и гр.9:

$$\sigma^{I}_{\frac{d_1}{d_0}} = \sqrt{11,86 \cdot 100} = 34,4\%$$

$$\sigma^{II}_{\frac{d_1}{d_0}} = \sqrt{5,76 \cdot 100} = 24,0\%$$

Расчеты показывают, что если в ноябре удельный вес каждой статьи расходов в среднем изменился более чем на треть своей величины, то в декабре - менее чем на четверть.

Для сводной оценки структурных изменений в исследуемой совокупности в целом за рассматриваемый временной интервал, охватывающий несколько недель, месяцев, кварталов или лет, наиболее удобным является **линейный коэффициент «абсолютных» структурных сдвигов за n периодов**:

$$\overline{\Delta}^{(n)}_{d_1-d_0} = \frac{\sum_{i=1}^{k} \left| d_{i\,n} - d_{i\,1} \right|}{k(n-1)} \qquad (10.10)$$

Используя итоговые данные гр.10 таблицы 10.3 получим:

$$\overline{\Delta}^{(n)}_{d_1-d_0} = \frac{17,6}{5 \cdot 2} = 1,8 \text{ проц.пункта}$$

Таким образом, за рассматриваемый период среднемесячное изменение по всем направлениям использования доходов составило 1,8 процентного пункта.

Необходимо отметить, что последний показатель 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.

10.4. 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 этого необходимо иметь частотное распределение единиц исследуемой совокупности и взаимосвязанное с ним частотное распределение изучаемого признака. Для удобства вычислений и повышения аналитичности данных единицы совокупности, как правило, разбиваются на равные группы - 10 групп по 10% единиц в каждой, 5 групп по 20% единиц и так далее.

Наиболее известным показателем концентрации является **коэффициент Джини**, обычно используемый как мера дифференциации или социального расслоения:

$$G = 1 - 2\sum_{i=1}^{k} d_{xi}\, d_{yi}^{H} + \sum_{i=1}^{k} d_{xi}\, d_{yi}, \qquad (10.11)$$

где d_{xi} - доля i-ой группы в общем объеме совокупности;

d_{yi} - доля i-ой группы в общем объеме признака;

d_{yi}^{H} - накопленная доля i-ой группы в общем объеме признака

Если доли выражены в процентах, данную формулу можно преобразовать:

для 10%-го распределения -

$$G = 110 - 0{,}2\sum_{i=1}^{k} d_{yi}^{H} \qquad (10.12)$$

для 20%-го распределения -

$$G = 120 - 0{,}4\sum_{i=1}^{k} d_{yi}^{H} \qquad (10.13)$$

Чем ближе к 1 (100%) значение данного признака, тем выше уровень концентрации; при нуле мы имеем равномерное распределение признака по всем единицам совокупности.

Оценка степени концентрации также может быть получена на основе **коэффициента Лоренца**:

$$L = \frac{\sum_{i=1}^{k} \left| d_{xi} - d_{yi} \right|}{2} \qquad (10.14)$$

При использовании данного коэффициента можно оперировать как долями единицы, так и процентами. Коэффициент Лоренца изменяется в тех же границах, что и коэффициент Джини.

Определим степень концентрации доходов населения по данным таблицы 10.4.:

Таблица 10.4

Распределение доходов населения России в январе - сентябре 1995г.

20%-ные группы населения	Объем денежных доходов		d_{xi}	$d_{xi}\ d_{yi}$	d_{yi}^{H}	$d_{xi}\ d_{yi}^{H}$	$\left\lvert d_{xi} - d_{yi} \right\rvert$
	% к итогу	d_{yi}					
А	1	2	3	4	5	6	7
Первая (с наименьшими доходами)	5,5	0,055	0,2	0,0110	0,055	0,0110	0,145
Вторая	10,2	0,102	0,2	0,0204	0,157	0,0314	0,098
Третья	15,2	0,152	0,2	0,0304	0,309	0,0618	0,048
Четвертая	22,4	0,224	0,2	0,0448	0,533	0,1066	0,024
Пятая (с наивысшими доходами)	46,7	0,467	0,2	0,0934	1,000	0,2000	0,267
Итого	100,0	1,0	1,0	0,2000	Х	0,4108	0,582

Для расчета коэффициента Джини воспользуемся итоговыми данными граф 4 и 6 таблицы 10.4 :

$G = 1 - 2 \cdot 0{,}4108 + 0{,}2 = 0{,}378$ или 37,8%.

Такой же результат мы получим, выполнив расчеты в процентах:
G = 120 - 0,4 (5,5 + 15,7 + 30,9 + 53,3 + 100,0) = 37,8%.

Второй способ расчета проще, однако, исходная формула незаменима в тех случаях, когда имеются неравные группы по объему совокупности (в нашем примере - по численности населения).

Используя данные графы 7 таблицы 10.4 определим коэффициент Лоренца:

$$L = \frac{0,582}{2} = 0,291 \text{ или } 29,1\%.$$

Оба коэффициента указывают на относительно высокую степень концентрации доходов населения.

Если под концентрацией понимается степень неравномерности распределения изучаемого признака, не связанная ни с 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** имеет следующий вид:

$$I_Z = \sum_{i=1}^{k} \left(\frac{m_i}{M} \right)^2 , \qquad (10.15)$$

где m_i - значение признака i-ой единицы совокупности;

M - объем признака всей совокупности.

Максимальное значение, равное 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 пример. Предположим, выпуск продукции А сконцентрирован на 5 предприятиях, расположенных в трех районах области (табл. 10.5.):

Таблица 10.5

Район	Число предприятий	Объем производства, млн.руб.		Доля одного предприятия в общем объеме продукции, (гр.3: Итог гр.2)
		всего	в среднем на 1 предприятие (гр.2:гр.1)	
А	1	2	3	4
А	1	5374	5374	0,584
Б	1	1225	1225	0,133
В	3	2610	870	0,094
Итого	5	9209	Х	Х

Вычислим показатель централизации производства данного вида продукции:

$$I_z = 0{,}584^2 + 0{,}133^2 + 3 \cdot 0{,}094^2 = 0{,}39$$

Рассчитанная величина свидетельствует о высокой степени централизации. Отметим, что аналитическая ценность показателей концентрации и централизации повышается при проведении сравнений во временном или территориальном аспектах.

Глава 11. Индексы

11.1. Общие понятия об индексах

«Индекс» в переводе с латинского - указатель или показатель. В статистике индексом называют показатель относительного изменения данного уровня исследуемого явления по сравнению с другим его уровнем, принятым за базу сравнения. В качестве такой базы может быть использован или уровень за какой-либо прошлый период времени (динамический индекс), или уровень того же явления по другой территории (территориальный индекс). В статистической практике динамические индексы получили большее распространение.

Индексы являются незаменимым инструментом исследования в тех случаях, когда необходимо сравнить во времени или пространстве две совокупности, элементы которых непосредственно суммировать нельзя. Предположим, нам требуется оценить рост заработной платы работников предприятия в текущем периоде по сравнению с базисным. Такая совокупность является однородной ,и поэтому вполне правомерно суммировать заработную плату работников в каждом периоде, рассчитать средние значения и сравнить их , поделив одну среднюю на другую. Рассмотрим теперь другой случай: нам необходимо оценить рост розничных цен. Здесь уже будет неправомерно складывать цены на разнородные товары, которые могут даже измеряться в различных единицах, а также рассчитывать какие-либо средние показатели. В подобных случаях и применяются индексы.

В целом, индексный метод направлен на решение следующих задач:

1) характеристика общего изменения уровня сложного социально-экономического явления;

2) анализ влияния каждого из факторов на изменение индексируемой величины путем элиминирования воздействия прочих факторов;

3) анализ влияния структурных сдвигов на изменение индексируемой величины.

В дальнейшем изложении будут использоваться следующие общепринятые обозначения:

i - индивидуальный индекс;

I - сводный индекс;

p - цена;

q - количество;

z - себестоимость;

r - урожайность;

s - посевная площадь;

1 - текущий период;

0 - базисный период.

Простейшим показателем, используемым в индексном анализе, является **индивидуальный индекс**, который характеризует изменение во времени экономических величин, относящихся к одному объекту:

$$i_p = \frac{p_1}{p_0} \text{ - } \quad \text{индекс цены,}$$

где p_1 - цена товара в текущем периоде;

p_0 - цена товара в базисном периоде;

$$i_q = \frac{q_1}{q_0} \text{ - индекс физического объема реализации;}$$

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0} \text{ - индекс товарооборота.}$$

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

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

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

11.2. Агрегатные индексы

В тех случаях, когда исследуются не единичные объекты, а состоящие из нескольких элементов совокупности, используются **сводные индексы**. Исходной формой сводного индекса является агрегатная.

При расчете агрегатного индекса для разнородной совокупности находят такой общий показатель, в котором можно объединить все ее элементы. Вернемся к рассмотрению задачи с розничными ценами. Как уже отмечалось, цены различных товаров складывать неправомерно, но суммировать товарооборот по этим товарам вполне допустимо. В текущем периоде такой товарооборот по n товарам составит:

$$p_1^1 q_1^1 + p_1^2 q_1^2 + p_1^3 q_1^3 + \ldots + p_1^n q_1^n = \sum p_1 q_1$$

Аналогично получим для базисного периода:

$$p_0^1 q_0^1 + p_0^2 q_0^2 + p_0^3 q_0^3 + \ldots + p_0^n q_0^n = \sum p_0 q_0$$

Если мы сравним товарооборот в текущем периоде с его величиной в базисном периоде, то получим **сводный индекс товарооборота**:

$$\mathrm{I}_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} \tag{11.1}$$

Для иллюстрации этого и последующих индексов воспользуемся следующими условными данными (табл. 11.1.):

Таблица 11.1

Цены и объем реализации трех товаров

Товар	Июль		Август	
	цена, руб.	продано, тыс.шт.	цена, руб.	продано, тыс.шт.
А	18	20	15	28
Б	50	11	40	13
В	40	12	35	12

Рассчитаем индекс товарооборота:

$$\mathrm{I}_{pq} = \frac{15\cdot 28 + 40\cdot 13 + 35\cdot 12}{18\cdot 20 + 50\cdot 11 + 40\cdot 12} = 0{,}978$$

Таким образом, товарооборот в целом по данной товарной группе в текущем периоде по сравнению с базисным уменьшился на 2,2% /100% - 97,8%/. Отметим, что размер товарной группы при расчете этого и последующих индексов значения не имеет.

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

$$\mathrm{I}_p = \frac{p_1^1 q_1^1 + p_1^2 q_1^2 + \ldots + p_1^n q_1^n}{p_0^1 q_1^1 + p_0^2 q_1^2 + \ldots + p_0^n q_1^n} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \tag{11.2}$$

$$\mathrm{I}_p = \frac{15\cdot 28 + 40\cdot 13 + 35\cdot 12}{18\cdot 28 + 50\cdot 13 + 40\cdot 12} = 0{,}832$$

По данной товарной группе цены в августе по сравнению с июлем снизились на 16,8%. При построении данного индекса цена выступает в качестве индексируемой величины, а количество проданного товара - в качестве веса.

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

Числитель и знаменатель сводного индекса цен также можно интерпретировать и по-другому. Числитель представляет собой сумму денег, фактически уплаченных покупателями за товары в текущем периоде. Знаменатель же показывает, какую сумму покупатели заплатили бы за те же товары, если бы цены не изменились. Разность числителя и знаменателя будет отражать величину экономии (если знак «-») или перерасхода («+») покупателей от изменения цен:

$$\mathrm{E} = \sum p_1 q_1 - \sum p_0 q_1 = 1360 - 1634 = -274 \text{ руб.}$$

Третьим индексом в данной индексной системе является **сводный индекс физического объема реализации**. Он характеризует изменение количества проданных товаров не в денежных, а в физических единицах измерения. Весами в данном случае выступают цены, которые фиксируются на базисном уровне:

$$\mathrm{I}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \qquad (11.3)$$

$$\mathrm{I}_q = \frac{28 \cdot 18 + 13 \cdot 50 + 12 \cdot 40}{20 \cdot 18 + 11 \cdot 50 + 12 \cdot 40} = 1{,}176$$

Физический объем реализации (товарооборота) увеличился на 17,6%.

Между рассчитанными индексами существует следующая взаимосвязь:

$$\mathrm{I}_p \cdot \mathrm{I}_q = \mathrm{I}_{pq}$$
$$0{,}832 \cdot 1{,}176 = 0{,}978$$

Мы рассмотрели применение индексного метода в анализе товарооборота и цен. Однако, эта же индексная система может использоваться для анализа результатов производственной деятельности предприятий, выпускающих разнородную продукцию. Тогда приведенные выше индексы соответственно называются:

I_{pq} - индекс стоимости продукции;
I_p - индекс оптовых цен;
I_q - индекс физического объема продукции.

Взаимосвязь между этими индексами остается прежней:

$$I_p \cdot I_q = I_{pq}.$$

Еще одна область применения индексов - анализ затрат на производство и себестоимости.

Индивидуальный индекс себестоимости характеризует изменение себестоимости отдельного вида продукции в текущем периоде по сравнению с базисным.

Для определения общего изменения уровня себестоимости нескольких видов продукции, выпускаемых предприятием, рассчитывается **сводный индекс себестоимости**. При этом себестоимость взвешивается по объему производства отдельных видов продукции:

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1} \qquad (11.4)$$

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

$$E = \sum z_1 q_1 - \sum z_0 q_1$$

Индекс физического объема продукции, взвешенный по себестоимости, имеет следующий вид:

$$I_q = \frac{\sum q_1 z_0}{\sum q_0 z_0} \qquad (11.5)$$

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

$$I_{zq} = \frac{\sum z_1 q_1}{\sum z_0 q_0} \qquad (11.6)$$

Все три индекса взаимосвязаны между собой:

$$I_z \cdot I_q = I_{zq}$$

Индексный метод также находит применение в статистике сельского хозяйства: **индекс валового сбора сельскохозяйственных культур** (Irs) может быть получен через **индекс урожайности** (Ir) и **индекс посевных площадей** (Is).

$$I_r \cdot I_s = I_{rs}$$

или (11.7).

$$\frac{\sum r_1 s_1}{\sum r_0 s_1} \cdot \frac{\sum S_1}{\sum S_0} = \frac{\sum r_1 s_1}{\sum r_0 s_0}$$

11.3. Сводные индексы в средней арифметической и средней гармонической формах

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

Предположим , мы располагаем данными о стоимости проданной продукции в текущем периоде и индивидуальными индексами цен, полученными, например, в результате выборочного наблюдения. Тогда при расчете сводного индекса цен можно использовать следующую замену:

$$p_0 q_1 = \frac{1}{i_p} p_1 q_1$$

В целом же сводный индекс цен в данном случае будет выражен в форме средней гармонической:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{1}{i_p} p_1 q_1} \qquad (11.8)$$

Рассмотрим следующий условный пример (табл. 11.2.):

Таблица 11.2

Данные о реализации и ценах по товарной группе

Т о в а р	Реализация в текущем периоде, руб.	Изменение цен в текущем периоде по сравнению с базисным,%
А	23000	+4,0
Б	21000	+2,3
В	29000	-0,8

Данные последней графы таблицы отражают изменение индивидуальных индексов цен, которые по товарам А, Б и В соответственно равны 1,040, 1,023 и 0,992.

С учетом этого получим:

$$\mathrm{I}_p = \frac{23000 + 21000 + 29000}{\dfrac{23000}{1{,}040} + \dfrac{21000}{1{,}023} + \dfrac{29000}{0{,}992}} = 1{,}016.$$

Цены по данной товарной группе в среднем возросли на 1,6%.

При расчете сводного индекса физического объема товарооборота можно использовать среднеарифметическую форму. При этом производится замена:

$$q_1 p_0 = i_q q_0 p_0$$

Тогда индекс имеет вид:

$$\mathrm{I}_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} \qquad (11.9)$$

Для иллюстрации этой формы расчета воспользуемся следующим примером (табл. 11.3.):

Таблица 11.3

Данные о реализации трех товаров в натуральном и стоимостном выражении

Т о в а р	Реализация в базисном периоде,руб.	Изменение физического объема реализации в текущем периоде по сравнению с базисным,%
А	46000	-6,4
Б	27000	-8,2
В	51000	+1,3

Индивидуальные индексы физического объема будут равны 0,936; 0,918; 1,013. С учетом этого рассчитаем среднеарифметический индекс:

$$\mathrm{I}_q = \frac{0{,}936 \cdot 46000 + 0{,}918 \cdot 27000 + 1{,}013 \cdot 51000}{46000 + 27000 + 51000} = 0{,}964.$$

Физический объем реализации данных товаров в среднем снизился на 3,6%.

11.4. Системы индексов

Индексы могут использоваться для анализа динамики социально-экономических явлений за ряд последовательных периодов. В этом случае для достижения сопоставимости они должны рассчитываться по единой схеме. Такая схема расчета индексов за несколько временных периодов называется системой индексов.

В зависимости от информационной базы и целей исследователя индексная система может строится в четырех вариантах. Рассмотрим их на примере сводного индекса цен, рассчитываемого за n периодов:

А. Цепные индексы цен с переменными весами:

$$\mathrm{I}_{p\,1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad \mathrm{I}_{p\,2/1} = \frac{\sum p_2 q_2}{\sum p_1 q_2}; \quad \mathrm{I}_{p\,3/2} = \frac{\sum p_3 q_3}{\sum p_2 q_3}; \; \ldots\ldots \quad \mathrm{I}_{p\,n/n-1} = \frac{\sum p_n q_n}{\sum p_{n-1} q_n};$$

Б.Цепные индексы цен с постоянными весами:

$$\mathrm{I}_{p\,1/0} = \frac{\sum p_1 q_0}{\sum p_0 q_0}; \quad \mathrm{I}_{p\,2/1} = \frac{\sum p_2 q_0}{\sum p_1 q_0}; \quad \mathrm{I}_{p\,3/2} = \frac{\sum p_3 q_0}{\sum p_2 q_0}; \; \ldots\ldots \quad \mathrm{I}_{p\,n/n-1} = \frac{\sum p_n q_0}{\sum p_{n-1} q_0};$$

В. Базисные индексы цен с переменными весами:

$$\mathrm{I}_{p\,1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad \mathrm{I}_{p\,2/0} = \frac{\sum p_2 q_2}{\sum p_0 q_2}; \quad \mathrm{I}_{p\,3/0} = \frac{\sum p_3 q_3}{\sum p_0 q_3}; \quad \ldots\ldots \quad \mathrm{I}_{p\,n/0} = \frac{\sum p_n q_n}{\sum p_0 q_n};$$

Г. Базисные индексы цен с постоянными весами:

$$\mathrm{I}_{p\,1/0} = \frac{\sum p_1 q_0}{\sum p_0 q_0}; \quad \mathrm{I}_{p\,2/0} = \frac{\sum p_2 q_0}{\sum p_0 q_0}; \quad \mathrm{I}_{p\,3/0} = \frac{\sum p_3 q_0}{\sum p_0 q_0}; \quad \ldots\ldots \; \mathrm{I}_{p\,n/0} = \frac{\sum p_n q_0}{\sum p_0 q_0};$$

Индексы системы "Б" по своей природе мультипликативны, т.е. последовательное произведение этих индексов приводит к сводному индексу цен за весь рассматриваемый период (система "Г").

11.5. Индексы постоянного и переменного состава

В предыдущих задачах рассматривались индексы, рассчитываемые по нескольким товарам или видам продукции, реализуемым или производимым в одном месте. Рассмотрим теперь случай, когда один товар или вид продукции реализуется или производится в нескольких местах (табл. 11.4.):

Таблица 11.4

Реализация товара А в двух регионах

Регион	Июнь		Июль	
	цена, руб.	продано, шт.	цена, руб.	продано,шт
1	12	10000	13	18000
2	17	20000	19	9000

Так как в данном случае реализуется один и тот же товар, вполне правомерно рассчитать его среднюю цену за июнь и за июль. Сравнением полученных средних значений получают **индекс цен переменного состава**:

$$\mathrm{I}_p^{nc} = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0} =$$
$$= \frac{13 \cdot 18000 + 19 \cdot 9000}{18000 + 9000} : \frac{12 \cdot 10000 + 17 \cdot 20000}{10000 + 20000} = \qquad (11.10)$$
$$= 15{,}00 : 15{,}33 = 0{,}978.$$

Из таблицы видно, что цена в каждом регионе в июле по сравнению с июнем возросла. В целом же, средняя цена снизилась на 2,2%. Такое несоответствие объясняется влиянием изменения структуры реализации товаров по регионам: в июне по более высокой цене продали товара вдвое больше, в июле ситуация принципиально изменилась (в данном условном примере для наглядности числа подобраны таким образом, чтобы это различие в структуре продаж было очевидным). Оценить воздействие этого фактора можно с помощью **индекса структурных сдвигов**:

$$
\begin{aligned}
\mathrm{I}^{cmp} &= \frac{\sum p_0 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0} = \\
&= \frac{12 \cdot 18000 + 17 \cdot 9000}{27000} : \frac{12 \cdot 10000 + 17 \cdot 20000}{30000} = 0{,}891
\end{aligned}
\qquad (11.11)
$$

Первая формула в этом индексе позволяет ответить на вопрос, какой была бы средняя цена в июле, если бы цены в каждом регионе сохранились на прежнем июньском уровне. Вторая часть формулы отражает фактическую среднюю цену июня. В целом по полученному значению индекса мы можем сделать вывод, что за счет структурных сдвигов цены снизились на 10,9%.

Последним в данной системе является рассмотренный выше **индекс цен фиксированного состава**, который не учитывает влияние структуры:

$$
\mathrm{I}_p^{\phi c} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = 1{,}098. \qquad (11.12).
$$

Итак, если бы структура реализации товара А по регионам не изменилась, средняя цена возросла бы на 9,8%. Однако, влияние на среднюю цену первого фактора оказалось сильнее, что отражается в следующей взаимосвязи:

$$
\mathrm{I}_{\mathrm{p}}^{\mathrm{фс}} \cdot \mathrm{I}^{-\mathrm{стр}} = \mathrm{I}_{\mathrm{p}}^{\mathrm{пс}}
$$

$$
1{,}098 \cdot 0{,}891 = 0{,}978.
$$

Аналогично строятся индексы структурных сдвигов, переменного и фиксированного состава для анализа изменения себестоимости, урожайности и пр.

Заключение

Мы рассмотрели приемы сбора, обработки и анализа статистических данных, которые являются методологическим базисом любой статистической работы. В то же время, необходимо отметить, что статистическое наблюдение не является обязательным этапом статистического исследования. Во многих случаях экономист-аналитик имеет дело с материалом, полученным из баз данных, бюллетеней информационных агентств, статистических сборников и других источников. Тогда работа должна начинаться с проверки полноты и качества данных, их группировки, а при отсутствии необходимости в этих этапах - с расчета индивидуальных и обобщающих показателей.

Рассмотренные приемы и методы с успехом могут использоваться не только в практике статистического анализа. Статистическая методология исследования в настоящее время заняла прочные позиции во многих областях знания. Статистические формулы находят применение в макро- и микроэкономике, оценке бизнеса и недвижимости, финансовом анализе, техническом анализе товарных и финансовых рынков.

Более того, подвергающийся статистической обработке материал не обязательно должен относиться к экономической области. В большинстве случаев, описанные приемы и показатели будут работоспособны и эффективны при обобщении и анализе технической, биологической, медицинской, демографической и социологической информации.

Рассматриваемые в пособии методы в большинстве случаев иллюстрированы практическими примерами. Подобные вычисления при небольших объемах совокупности или коротких динамических рядах не очень трудоемки. При работе же с большими массивами статистической информации необходимо использовать прикладное программное обеспечение, существенно ускоряющее и упрощающее все расчеты. Среди наиболее распространенных современных программных продуктов следует отметить пакеты Мезозавр, ОЛИМП, САНИ, Эвриста, STATISTICA, STATGRAPHICS и SPSS.

Рекомендуемая литература

1. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под. ред. чл.-корр. РАН И.И.Елисеевой. – 4-е изд., перераб. и доп. - М.: Финансы и статистика, 1998. - 480с.: ил.
2. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. - М.: ИНФРА-М, 1996. - 416с.
3. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: Учебник / Под ред. А.А.Спирина, О.Э.Башиной. - М.: Финансы и статистика, 1994. - 296с.: ил.
4. Практикум по теории статистики: Учеб. пособие /Под.ред. проф. Р.А. Шмойловой. – М.: Финансы и статистика, 1999. – 416с.: ил.
5. Статистический словарь / Гл. ред. М.А.Королев. - М.: Финансы и статистика, 1989. - 623с.: ил.
6. Теория статистики: Учебник / Под ред. проф. Шмойловой Р.А. –2-е изд., доп. и перераб. - М.: Финансы и статистика, 1998. -576с.:ил.
7. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере / Под ред. В.Э.Фигурнова. - М.: ИНФРА-М, Финансы и статистика, 1998. - 528с., ил.