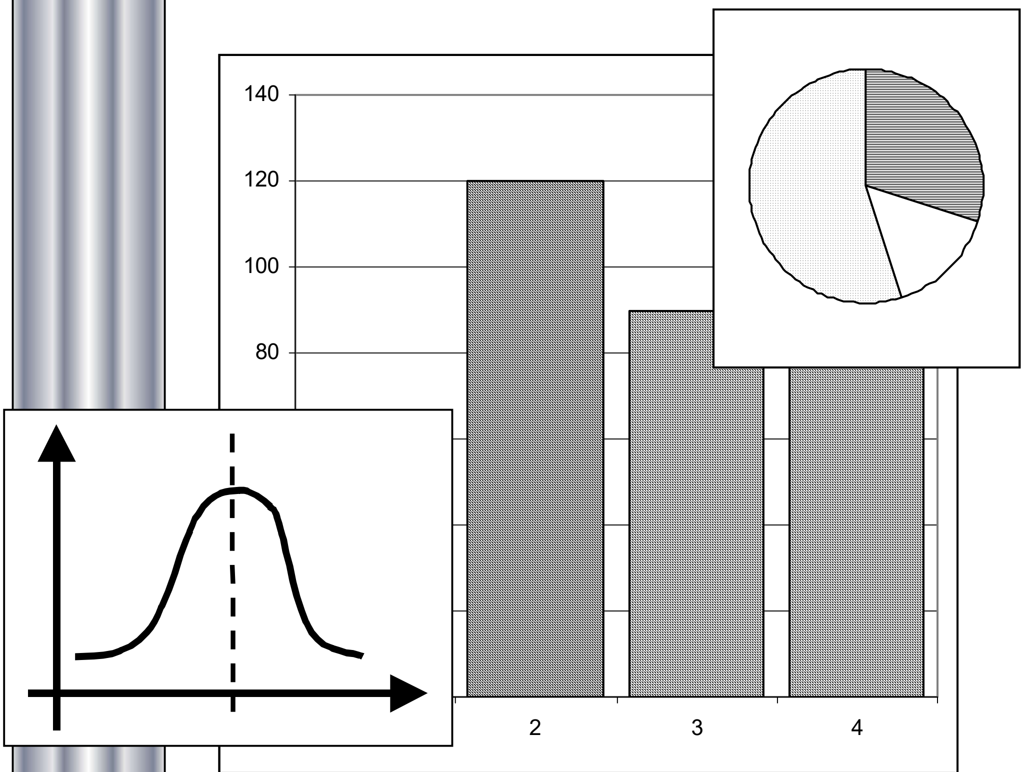
Кафедра 502 * Экономика промышленности *

Малютина О.А.
Аветисян А.С.

СТАТИСТИКА



Малютина О.А.
Аветисян А.С.

СТАТИСТИКА

Издание девятое стереотипное

Москва
ИВАКО Аналитик
2010

ББК 60.6

Малютина О.А., Аветисян А.С. Статистика. – М.: ИВАКО Аналитик, 2010. – 52 с.: ил.
ISBN 5-89547-066-1

Курс лекций, читаемый для студентов, обучающихся по специализации «Антикризисное управление предприятием» на кафедре 502 МАИ. Разработан в соответствии с Государственным образовательным стандартом высшего профессионального образования по специальности 0608 – «Экономика и управление на предприятии (по отраслям)».

Освещены этапы статистического исследования, сводка и группировка данных статистического наблюдения, анализ рядов распределения, использование корреляционного анализа, анализ рядов динамики, индексный метод анализа.

Для студентов, аспирантов, экономистов предприятий.

ISBN 5-89547-066-1

СТАТИСТИКА

Малютина Ольга Александровна, Аветисян Арташес Суренович

Издательство «ИВАКО Аналитик», 125993, г. Москва, А-80, ГСП, Волоколамское шоссе, д.4.
Тел. (499) 158-4764,
e-mail: economy@mai.ru
Тираж 200. Объем 3,1 п.л. 60x90/8.

1. ПРЕДМЕТ И МЕТОД СТАТИСТИКИ

1.1.Предмет, методы и основные категории статистики как науки

Термин «статистика» произошел от латинского слова “status” («статус»), что означает состояние, положение вещей. Первоначально статистика давала словесное описание «достопримечательностей» государства и лишь с XIXв. статистические сведения стали сообщать в количественной форме. Сейчас слово «статистика» имеет несколько значений. Прежде всего, оно относится к обозначению собранных статистических данных о различных явлениях общественной жизни. Под статистикой понимают также отрасль практической деятельности, задачами которой являются сбор, обработка и анализ статистических данных. И, наконец, под статистикой понимают специальную научную дисциплину, занимающуюся разработкой теоретических положений и методов, используемых статистической практикой.

Возникновение статистической науки связано с именами английского экономиста В. Петти (1623-1687) и Д. Граунта (1629-1674), статистические и демографические идеи которых были развиты их последователями – немецким пастором И. Зюсмильхом (1707-1767) и виднейшим бельгийским ученым XIXв. А. Кетле (1796-1874).

Как и всякая наука, статистика имеет свой предмет. Таковым является количественная сторона массовых общественных явлений в неразрывной связи с их качественной стороной, или их содержанием, а также количественное выражение закономерностей общественного развития в конкретных условиях места и времени.

Свой предмет статистика изучает при помощи определенных категорий, то есть понятий, которые отражают наиболее общие и существенные свойства, признаки, связи и отношения предметов и явлений объективного мира.

В статистике это следующие понятия:

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

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

Признак – общее свойство, характерная черта или иная особенность единиц совокупности, которые могут быть наблюдаемы или измерены. По характеру отображения свойств единиц изучаемой совокупности варьирующие признаки делятся на две основные группы:

признаки, имеющие непосредственное количественное выражение, например возраст, стаж работы, средний заработок и т.д. Они могут быть дискретными и непрерывными;

признаки, не имеющие непосредственного количественного выражения. В этом случае отдельные единицы совокупности различаются своим содержанием (например, профессии – характером труда: учитель, столяр, швея-мотористка и т.д.). Такие признаки обычно называют атрибутивными (в философии «атрибут» - неотъемлемое свойство предмета). В случае, когда имеются противоположные по значению варианты признака, говорят об альтернативном признаке (да, нет). Например, продукция может быть годной или бракованной (не годной).

Особенностью статистического исследования является то, что в нем изучаются только варьирующие признаки, то есть признаки, принимающие различные значения (для атрибутивных, альтернативных признаков) или имеющие различные количественные уровни у отдельных единиц совокупности.

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

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

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

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

Статистические данные – это конкретные численные значения статистических показателей. Они всегда определены не только качественно, но и количественно и зависят от конкретных условий места и времени.

Свой предмет статистика изучает при помощи своего, специфического метода. Метод статистики – это целая совокупность приемов, пользуясь которыми статистика исследует свой предмет. Она включает в себя три группы собственно методов: метод массовых статистических наблюдений, метод группировок, метод обобщающих показателей.

Статистическое наблюдение заключается в сборе первичного статистического материала, в научно организованной регистрации всех существенных фактов, относящихся к рассматриваемому объекту. Это первый этап статистического исследования.

Метод группировок дает возможность все собранные в результате массового статистического наблюдения факты подвергать систематизации и классификации. Это второй этап статистического исследования.

Метод обобщающих показателей позволяет характеризовать изучаемые явления и процессы при помощи статистических величин – абсолютных, относительных и средних. На этом этапе статистического исследования выявляются взаимосвязи и масштабы явлений, определяются закономерности их развития, даются прогнозные оценки.

Некоторые общие черты формирования обобщающих показателей устанавливаются посредством измерения их вариации. Показатели вариации дополняют средние величины, за которыми скрываются индивидуальные различия. Они характеризуют степень однородности статистической совокупности по данному признаку. Показатели вариации определяют границы вариации признака, а их соотношения выражает взаимосвязь признаков.

1.2. Принципы организации государственной статистики и структура статистической науки

Преобразования в российской экономике, связанные с рыночными реформами, обусловили необходимость адекватного изменения как организации государственной статистики, так и ее функций.

Госкомстат Российской Федерации является федеральным органом исполнительной власти. В своей деятельности он руководствуется Конституцией, законами, распоряжениями и постановлениями Президента и Правительства, а также статистическими программами, которые формируются с учетом предложений федеральных органов исполнительной и законодательной власти, органов государственной власти субъектов Российской Федерации, научных и других организаций и утверждаются Госкомстатом по согласованию с Правительством.

Методология статистических показателей, формы и методы сбора и обработки статистических данных, установленные Госкомстатом, являются официальными статистическими стандартами Российской Федерации.

Особую значимость в современных условиях приобретает деятельность административно-территориальных и местных органов государственной статистики в связи с развитием региональной и муниципальной статистики.

Таким образом, Госкомстат России вместе с территориальными органами представляет собой централизованную иерархическую систему, выполняющую важнейшую функцию – информационное обеспечение государственного и муниципального управления.

Чтобы управлять современным обществом, следует отслеживать состояние и взаимодействие друг с другом двух основных сфер – экономики и социальной системы.

Внутренние процессы, протекающие как в сфере экономики, так и в социальной сфере, различны по содержанию и могут классифицироваться в зависимости от макро- и микроуровня, вида деятельности, отрасли и т.д.

Эти обстоятельства предопределили выделение в составе единой статистической науки отдельных разделов и отраслей. Структура статистической науки представлена на рис. 1.1.

Методологические принципы и методы, применяющиеся при изучении различных сторон общественной жизни, излагаются в разделе «Общая теория статистики». Этот раздел статистической науки посвящен разработке общих методов и правил проведения статистического исследования: наблюдений, группировок, анализа и обобщения первичных данных, выявления тенденций и закономерностей в развитии изучаемых объектов, явлений и процессов.

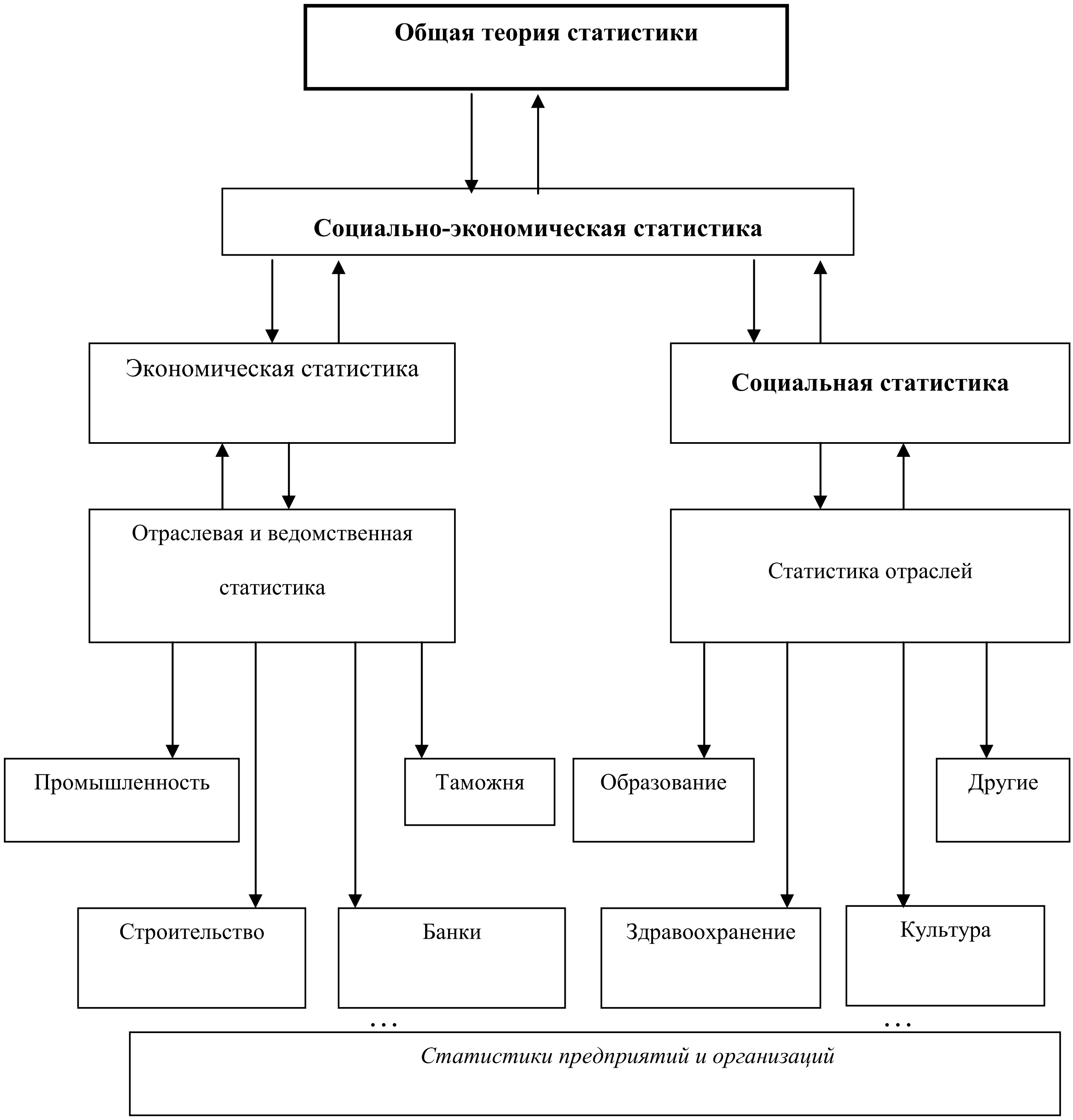


Рис. 1.1. Структура статистической науки

На разных уровнях управления и в различных отраслях и сферах деятельности существуют различные возможности и потребности в приемлемых методах анализа фактических данных и системах показателей, используемых для характеристики явлений общественной жизни.

Обобщающие показатели экономической статистики характеризуют состояние и тенденции развития национальной экономики, наличие, размещение и эффективность использования всех видов ресурсов (национального богатства, трудовых и финансовых ресурсов), состояние товарных рынков, позволяет проводить анализ инвестиций, инноваций и др.

Экономическая статистика раскрывает важнейшие пропорции и соотношения между производством, распределением, накоплением и потреблением.

Задачей социальной статистики является разработка и анализ системы показателей для характеристики качества и уровня жизни населения страны, отражающей демографическую ситуацию и различные аспекты социальных отношений.

Контрольные вопросы.

1. От какого латинского слова происходит термин «статистика»? Что он означает?
2. Перечислите основные значения слова «статистика».
3. К какому времени относится становление статистики как науки?
4. Дайте определение предмета статистики.

5. Что такое статистическая закономерность?
6. Что такое совокупность, единица совокупности?
7. Понятие вариации, признака, статистического показателя, статистических данных.
8. Как делятся варьирующие признаки?
9. Перечислите специфические методы, присущие статистическому исследованию.
10. Какова роль показателей вариации при использовании обобщающих статистических показателей?
11. Какие принципы положены в основу организации статистики в России?
12. Какова структура статистической науки?
13. Чем занимается экономическая статистика?
14. Задачи социальной статистики.

2. ЭТАПЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

2.1. Понятие статистической информации

Для исследования социально экономических явлений и процессов общественной жизни следует прежде всего собрать о них необходимые сведения - статистические данные. Под статистическими данными (информацией) понимают совокупность количественных характеристик социально-экономических явлений и процессов, полученных в результате статистического наблюдения, их обработки или соответствующих расчетов.

В общей системе экономической информации статистике принадлежит важная роль. Это объясняется тем, что действия предприятий и организаций определяются нуждами отраслей народного хозяйства и общества в целом. Следовательно, информация об этих действиях должна собираться и обобщаться для получения необходимых итоговых данных по всем уровням хозяйствования. А эти данные и составляют статистику.

Кроме того, статистика дает информацию для решения региональных задач, для предпринимательской деятельности - об уровне цен на товары, объемах реализации товаров и услуг, условиях кредитования, уровнях и темпах инфляции и т.д.

На основе статистической информации разрабатывается экономическая и социальная политика государства, оцениваются ее результаты, составляются экономические прогнозы, обеспечивается подготовка экономических соглашений между государствами.

Но на любом уровне и в любой сфере эффективность использования статистики во многом определяется качеством исходных данных, т.е. насколько достоверна, сравнима, своевременна и полно собрана статистическая информация. Поэтому сбор данных лежит в основе всего исследования.

Во всяком статистическом исследовании изучается конкретное массовое явление, имеющее определённую природу, качественное содержание, что необходимо знать прежде чем приступить к измерению количественных размеров явления.

Основой качественного анализа экономических объектов является экономика различных отраслей народного хозяйства. Лишь после того, как экономическое явление теоретически проанализировано, можно приступить непосредственно к статистическому исследованию.

Статистическое исследование включает следующие этапы:

1) статистическое наблюдение - сбор первичной информации об изучаемом явлении;
2) сводку и группировку данных статистического наблюдения, расчет обобщающих показателей;
3) анализ закономерностей и выводы по проведённому исследованию. Эти этапы неразрывно связаны друг с другом и составляют единое целое.

2.2. Статистическое наблюдение

Статистическое наблюдение - первый этап статистического исследования, представляющий собой планомерный, научно организованный по единой программе учет фактов о явлениях и процессах общественной жизни и сбор полученных на основе этого учета массовых первичных данных.

Любое статистическое наблюдение должно иметь программу и организационный план проведения. При этом решаются вопросы о том, каким способом, какими средствами и в какие сроки будет произведен учет данных, как будут организованы сбор и проверка полученного первичного материала. Научная организация статистического наблюдения необходима для получения объективных данных, так как последние являются фундаментом статистического исследования.

Сбор необходимой статистической информации осуществляется с помощью двух организационных форм статистического наблюдения: статистической отчетности и специально организованного наблюдения. Статистическая отчетность - это форма наблюдения, при которой статистические органы и вышестоящие учреждения в определенные сроки получают от предприятий, учреждений и организаций необходимые данные в виде законодательно установленных отчетных документов, скрепленных подписями лиц, ответственных за предоставление и достоверность сообщаемых сведений.

Статистическая отчетность отражает результаты деятельности предприятий, учреждений и организаций, носит периодический характер. При этом источником сведений являются первичные учетные документы. Действующая статистическая отчетность делится на типовую и специализированную. Состав показателей в типовой отчетности является единым для предприятий всех отраслей народного хозяйства. В специализированной отчетности состав показателей зависит от особенностей отраслей экономики.

Отчетность является основной формой статистического наблюдения.

Специально организованное статистическое наблюдение - это наблюдение, которое проводится для изучения вопросов, не охватываемых отчетностью, а также для проверки правильности и качества отчетности. Такое наблюдение осуществляется через определенные промежутки времени и подразделяется на переписи и специально организованные обследования.

Виды статистического наблюдения чаще всего классифицируются по следующим трем признакам:

1) охвату наблюдения единиц совокупности, подлежащих статистическому исследованию;

2) систематичности наблюдения;

3) источнику полученных сведений.

По степени охвата единиц изучаемой совокупности наблюдения делятся на сплошные, при которых регистрируются все без исключения единицы изучаемой совокупности, и несплошные, когда учету подлежит только часть единиц исследуемой совокупности.

Несплошное наблюдение различают по виду: выборочное наблюдение, основного массива, монографическое, анкетное. Различие между этими видами наблюдений заключается в способе отбора единиц, подвергшихся наблюдению. Выборочным считается наблюдение, при котором отбор производится в случайном порядке, по специальным правилам, обеспечивающим объективность отбора. Наблюдение основного массива заключается в том, что из всей совокупности единиц выбирается та ее часть, которая имеет преобладающий удельный вес в совокупности по изучаемым признакам. Монографическое наблюдение характеризуется тем, что отбираются только отдельные, характерные в каком-либо отношении, единицы в совокупности (лучшие, типичные или худшие), подвергающиеся детальному статистическому описанию. При анкетном наблюдении данные получают путем рассылки определенному кругу лиц анкет, заполнение и возвращение которых основаны на добровольных началах.

По систематичности различают непрерывное или текущее, и прерывное наблюдения. Прерывное наблюдение может быть периодическим или единовременным. Непрерывным называется наблюдение, которое проводится постоянно; при этом данные подлежащие регистрации, фиксируются по мере их возникновения. Прерывное наблюдение проводится с перерывами, время от времени. Проводящееся регулярно, т. е. через равные промежутки времени, прерывное наблюдение называется периодическим. При отсутствии регулярности говорят о единовременном наблюдении.

По источнику получения сведений различают: наблюдение непосредственное, когда факты, подлежащие регистрации, устанавливаются лицами, проводящими наблюдение, путём замеров, подсчетов и т. п.; наблюдение документальное, при котором источником необходимых сведений являются различного рода документы; опрос, основанный на регистрации ответов со слов опрашиваемых.

В статистике применяются следующие виды опросов:

1) экспедиционный способ, при котором представители статистических органов опрашивают обследуемых лиц и сами регистрируют необходимые сведения и доставляют их в статистические органы;

2) саморегистрация - способ, при котором опрашиваемые дают нужные сведения, самостоятельно заполняя заранее полученные бланки регистрации;

3) корреспондентский способ, при котором сведения в органы статистики сообщает штаб добровольных корреспондентов;

4) анкетный способ, при котором разработанная анкета рассылается определенному кругу лиц и

после заполнения возвращается организациям, проводившим наблюдение.

2.3. Программно-методологические и организационные вопросы статистического наблюдения

Процесс статистического наблюдения включает три этапа. На первом этапе проводятся все подготовительные работы, на втором этапе собирается вся необходимая информация, на третьем проверяется достоверность собранных данных.

На первом этапе разрабатывается специальный метод наблюдения, содержащий программно-методологические и организационные вопросы. К программно-методологическим вопросам относятся определение цели наблюдения, определение объекта и единицы наблюдения, разработка программы наблюдения.

Цель статистического наблюдения устанавливается исходя из общих задач статистического исследования и отражается в документах, на основании которых проводится наблюдение. Такими документами могут быть приказы, распоряжения, договора, постановления правительства.

Для правильной организации статистического исследования важное значение имеет научно обоснованное определение объекта и единицы наблюдения.

Объектом наблюдения называют совокупность единиц изучаемого явления, которая подвергается статистическому исследованию. Эта совокупность имеет характерные особенности, отличающие её от других совокупностей. Определить объект наблюдения - значит установить важнейшие отличительные признаки, определяющие границы объекта по времени, территории и содержанию изучаемого явления.

Единицей наблюдения называется первичный элемент объекта наблюдения, который является носителем признаков, подлежащих регистрации. В каждом исследовании в зависимости от сложности объекта или задач исследования выделяется одна или несколько единиц наблюдения.

Программой наблюдения является перечень признаков единиц наблюдения, подлежащих регистрации. Программа наблюдения отражается в перечне вопросов, ответы на которые необходимо получить в процессе наблюдения. Вопросы и ответы фиксируются в основном документе статистического наблюдения - статистическом формуляре (переписном листе, анкете, бланке и т. п.).

На практике применяются два вида формуляра: списочный, в который заносятся результаты обследования нескольких единиц наблюдения, и индивидуальный, заполняемый на каждую единицу отдельно. Статистический формуляр должен быть удобен для заполнения, чтения и механизированной обработки данных. К статистическому формуляру прилагается инструкция по заполнению.

В план статистического наблюдения включаются организационные мероприятия, необходимые для успешного осуществления наблюдения. Указываются статистический орган, проводящий наблюдение, а также место, время и способ проведения наблюдения.

Организационный план включает также вопросы подготовки кадров, привлекаемых к наблюдению, и их инструктаж, работу по разъяснению целей и задач наблюдения и т.п.

При некоторых наблюдениях необходимо установить критический момент наблюдения, по состоянию на который регистрируют сведения, собираемые в процессе наблюдения.

После решения программно-методологических и организационных вопросов, а также сбора необходимой статистической информации следует провести контроль полноты и качества данных наблюдений.

Такой контроль обусловлен важнейшей задачей наблюдения - получением доброкачественных, достоверных данных. Как известно, в процессе наблюдения могут возникнуть погрешности, называемые ошибками наблюдения.

Все погрешности, появляющиеся при сплошном наблюдении, называются ошибками регистрации. Ошибки, возникающие в силу стечения случайных обстоятельств во время записи ответов на вопросы наблюдения (описки, оговорки и т.д.), называются случайными и влияют на точность сведений как в сторону их преувеличения, так и в сторону преуменьшения. В соответствии с законом больших чисел эти ошибки взаимно погашаются и не оказывают существенного влияния на точность результатов наблюдения.

Погрешности, возникающие под действием определённых постоянных причин (неточность измерительного инструмента, приписки в отчетности и т.д.), называются систематическими ошибками. Они вызывают одностороннее преувеличение или преуменьшение данных, что приводит к искажению результатов наблюдения.

При несплошном наблюдении могут возникать специфические погрешности, называемые ошибками репрезентативности. Они появляются в результате недостаточного охвата наблюдением единиц совокупности, приводящего к неточному воспроизведению всей совокупности этой частью единиц.

Для предупреждения или уменьшения этих погрешностей в организационном плане следует предусматривать специальные контрольные мероприятия по проверке полноты и качества данных. Контроль полноты - это проверка того, насколько полно охвачен объект наблюдением, т.е. о всех ли единицах наблюдения собраны данные. Контроль качества данных может быть логическим и арифметическим. Логический контроль заключается в сопоставлении ответов на различные вопросы и выявлении логически несовместимых ответов. Арифметический (счётный) контроль состоит в проверке правильности подведения итогов и проведения различных расчетов.

При контроле данных статистического наблюдения важно не только выявить ошибки, но и вскрыть причины их возникновения.

Контрольные вопросы

1. Что понимается под статистической информацией?
2. Поясните роль качественного анализа в статистическом исследовании.
3. Перечислите основные этапы статистического исследования.
4. Изложите содержание статистического наблюдения.
5. Каковы требования, предъявляемые к данным статистического наблюдения?
6. Какие вопросы разрабатываются при составлении плана проведения статистического наблюдения?
7. Какова сущность статистической отчетности?
8. Чем вызвана необходимость проведения специально организованного наблюдения?
9. Какими организационными признаками характеризуются виды статистического наблюдения?
10. Расскажите о выборе объекта и единицы наблюдения.
11. Каково содержание программы статистического наблюдения?
12. Какие ошибки могут возникать в процессе статистического наблюдения?
13. В чём заключается контроль данных наблюдений?

3. СВОДКА И ГРУППИРОВКА ДАННЫХ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

3.1. Понятие о сводке. Виды сводки

Полученные в результате статистического наблюдения данные о каждой единице наблюдения могут быть использованы для характеристики изучаемой совокупности в целом только после проведения статистической сводки.

Сводка - это процесс теоретического обобщения статистических данных, включающий сложный комплекс операций по систематизации и научной обработке этих данных.

Главным содержанием сводки является получение обобщающих статистических показателей, раскрывающих сущность явления и определяющих статистические закономерности. В результате осуществляется переход от первичных данных, характеризующих отдельные явления, к данным, характеризующим всю совокупность в целом - к итоговым обобщающим показателям.

В зависимости от сложности задач исследования различают три вида сводки. Простейшая сводка заключается в простом подведении итогов без их предварительной группировки. Сложная сводка предусматривает выделение однородных групп и подгрупп, подсчет групповых и общих итогов. Наиболее сложная сводка предусматривает выделение однородных групп и подгрупп, подсчет групповых и общих итогов, а также расчет обобщающих показателей для характеристики однородных групп и подгрупп.

Сводку статистических данных проводят по заранее составленной программе, содержание которой определяется задачами исследования. Программа предусматривает перечень объектов и показателей для изучения исследуемого явления.

Программа включает систему макетов таблиц, в которых в определённом порядке перечислены объекты, группы и подгруппы и необходимые для характеристики показатели.

По организации работ различают централизованную и децентрализованную сводки.

При централизованной материалы наблюдения сосредотачиваются в одном центральном статистическом органе (например, в Госкомстате РФ).

При децентрализованной сводке обобщение собранных сведений производится по единому плану на местах - в районных, городских, областных, краевых управлениях статистики, и вышестоящему статистическому органу для дальнейшего обобщения передаются сводные итоги.

По технике выполнения сводка может быть ручной и машинной. В современных условиях доминирующей является машинная сводка, при которой статистические материалы обрабатываются на различных ЭВМ. Поэтому одной из задач практической перестройки статистических органов является их полное техническое переоснащение современными ЭВМ.

3.2. Группировка статистических данных

Группировка является одним из наиболее эффективных методов обработки статистических данных. Это важнейший этап исследования массовых общественных явлений. Первостепенное значение группировки в экономическом анализе статистических данных проявляется в применении обобщающих показателей: средних, относительных и др. Лишь на основе правильной, рациональной группировки обобщающие показатели приобретают научную и практическую значимость.

Группировкой называется разделение единиц совокупности на однородные группы по существенным варьирующим признакам. Группировка может быть проведена только для целей данного исследования, для анализа отдельного вопроса. Позднее при изучении этого вопроса может быть принята другая группировка.

Варьирующие признаки, положенные в основу группировки, называются группировочными признаками. По характеру варьирования они разделяются на атрибутивные, варианты которых выражены словесно, и количественные, варианты которых выражены числом. Так, отрасль производства, профессии работающих - атрибутивные признаки предприятия, а стоимость основных фондов, численность работающих, стоимость продукции - количественные. Группировка единиц совокупности по одному признаку называется простой группировкой. Группы, образованные по одному признаку можно расчленять на подгруппы по другому признаку, которые в свою очередь можно расчленять на подгруппы по третьему признаку и т.д. Группировки по двум и более признакам, взятые в комбинации, называются комбинированными.

В отличии от группировки классификация - это устойчивое разграничение объектов. Классификация может быть либо постоянной, либо устанавливаемой на долгое время. Основой классификации может быть только атрибутивный признак.

На этапе перехода от одной формы хозяйствования к другой возникла потребность в новых классификациях, таких, как классификация организационно-правовых форм хозяйствующих субъектов.

Из всего многообразия задач, решаемых с помощью группировок, принято выделять три основные задачи:

- разделение всей совокупности на качественные однородные совокупности, или, иными словами, выделение социально - экономических типов; такие группировки называются типологическими;
- изучение состава совокупности по тем или иным признакам; такие группировки называются структурными;
- изучение взаимосвязанного изменения варьирующих признаков в пределах совокупности; такие группировки называются аналитическими.

Если в группировке выделяют типы, то есть образуются качественно однородные совокупности, то такие группировки являются типологическими независимо от того, дают ли они дополнительную характеристику составу совокупности и взаимосвязанному изменению варьирующих признаков или нет.

Структурные группировки обычно характеризуют состав и изменение состава совокупности по одному или нескольким признакам. К структурным группировкам близки ряды распределения единиц совокупности по варьирующим признакам. Однако задачи у них разные: структурные группировки показывают закономерности изменения структуры, а ряды распределения - закономерности в характере распределения.

С помощью аналитических или факторных группировок решается задача выявления причинно - следственных связей между явлениями и между признаками явлений.

Разделение группировок на виды в зависимости от решаемых с их помощью задач имеет важное значение, так как от вида группировок зависит решение основных вопросов метода группировок: выбор группировочного признака, правило образования групп, состав системы показателей с помощью которых характеризуются группы.

Для правильного теоретического обобщения статистических данных первостепенное значение имеют типологические группировки, позволяющие выделить и охарактеризовать социально - экономические типы явлений.

Примером типологической группировки может служить табл. №3.1 (пример условный).

В статистике классификации имеют широкое применение. Так, классификация промышленных предприятий по отраслям, хотя и не остается неизменной в силу происходящих в промышленности изменений, но все же устанавливается на ряд лет и служит основой статистики экономических явлений.

Таблица №3.1.

Группировка предприятий по формам собственности

Группы предприятий по формам собственности	Приватизировано предприятий	В % к итогу
1	2	3
Муниципальная	7957	48.9
Субъектов федерации	3843	23.7
Федеральная	4448	27.4
Итого	16248	100

При выделении качественно однородных типов посредством типологических группировок необходимо учитывать, что типы не являются неизменными. Во-первых, типы развиваются во времени и в пространстве, и, во-вторых, они выявляются и формируются в соответствии с поставленной задачей исследования и в зависимости от конкретных условий. Поэтому из одного и того же множества явлений могут быть выделены различные типы.

Процесс построения типологической группировки включает в себя следующее:

- предварительную наметку возможных типов, которые теоретически могут существовать, исходя из целевой задачи группировки на основе анализа изучаемого социально-экономического явления;
- выбор основного группировочного признака для выделения типов явления;
- выделение групп и подгрупп характеризующих отдельные типы явлений;
- непосредственная группировка, подсчет численности групп;
- характеристика выделенных групп и подгрупп с помощью систем статистических показателей.

Выбор группировочного признака - важнейший вопрос метода группировок, поскольку от правильности выбора зависят результаты исследования. Из всех возможных группировочных признаков необходимо выбрать тот, который разбивает всю изучаемую совокупность на качественные однородные группы. Число групп в типологической группировке зависит от числа социально экономических явлений.

В области экономических явлений выбор группировочного признака в каждом конкретном случае должен быть обоснован экономической теорией. Только на основе правильного экономического анализа законов развития явлений могут быть определены признаки, которые должны быть положены в основу типологических группировок. При этом важно понять сущность изучаемых явлений, а затем, руководствуясь задачами исследования, положить в основу группировки самые существенные признаки.

Для определения внутреннего строения явления (структуры явления) и изучения вариации внутри однотипной совокупности используются структурные группировки, которые могут быть построены как по атрибутивному, так и по количественному признаку. В первом случае число групп определяется числом вариантов атрибутивного признака, во втором - объемом статистической совокупности и задачами исследования.

Структурные группировки, взятые за период времени, характеризуют изменение структуры изучаемого явления (структурные сдвиги). Оценка структурных сдвигов производится сопоставлением структуры явления за ряд периодов. При этом устанавливаются тенденции в развитии явления.

Примеров структурной группировки является табл. №3.2.

Построение структурной группировки предполагает:

- выбор группировочного признака;
- выделение групп;

- непосредственную группировку, подсчет численности групп;
- определение удельного веса каждой группы.

При этом наиболее важное значение имеет определение удельного веса каждой группы, позволяющее оценить характеристику структуры явления, в нашем примере структуру предприятий по объему товарооборота.

Вопросы исследования взаимосвязи варьирующих признаков решаются с помощью аналитических (факторных) группировок.

Явления общественной жизни и их признаки тесно связаны между собой и зависят друг от друга. То же происходит и в экономических явлениях. Так, розничный товарооборот торговых предприятий зависит от ряда факторов, в том числе от общего объема товарооборота предприятия.

Таблица №3.2.

Группировка предприятий по величине уставного капитала.

№ п/п	Группы предприятий по величине уставного капитала (млн.руб.)	Число предприятий
1.	1215 – 2340	14
2.	2340 – 3465	4
3.	3465 – 4590	4
4.	4590 – 5715	2
5.	5715 – 6840	5
6.	6840 – 7965	1
ИТОГО:		30

Группируя торговые предприятия по общему объему товарооборота и исчисляя для каждой группы среднюю величину розничного товарооборота, можно статистически выразить зависимость между этими признаками.

Таблица №3.3.

Группировка коммерческих банков региона по сумме активов баланса.

№ п/п	Группы банков по сумме активов баланса, (млн. руб.) X	Количество банков, (единиц)	В среднем на один банк	
			Численность занятых, (человек) $\overline{У_1}$	Балансовая прибыль (млрд. руб.) $\overline{У_2}$
1.	до 20000	19	184	22,5
2.	20000 – 30000	8	313	31,6
3.	30000 – 40000	7	374	36,0
4.	40000 – 50000	9	468	69,2
5.	50000 - 60000	7	516	205,6
Всего:		50	$\overline{У_1}$=323	$\overline{У_2}$ = 60,0

Полученную аналитическую группировку можно представить в виде табл. №3.3. Аналитические группировки помогают раскрыть и изучить многообразные связи и зависимости между варьирующими признаками, отражающими различные свойства совокупностей. Особенность таких группировок заключается в том, что каждая группа, выделенная по существенному факторному признаку, характеризуется средними значениями результативного признака. Аналитические группировки позволяют анализировать влияние изменения группировочного признака на варьирование результативного признака и на этой основе измерять тесноту связи между вариациями этих признаков.

3.3. Статистические таблицы

Данные сводки и группировки материалов принято оформлять в виде таблиц. Это наиболее рациональная форма изложения статистических данных, так как таблица наглядно показывает связь между признаками изучаемого явления.

Статистической таблицей называется таблица, содержащая числовую характеристику совокупности по одному или нескольким признакам в их взаимной связи и раскрывающая содержание сводки.

Можно провести аналогию между статистической таблицей и грамматическим предложением. Статистическая таблица, языком цифр рассказывающая об изучаемом явлении, называется статистическим предложением. Подлежащее статистической таблицы - то, о чем говорится в таблице (обычно это перечень групп изучаемой совокупности).

Сказуемое - то, что говорится о подлежащем (обычно это показатели, характеризующие выделенные группы и всю совокупность в целом). Подлежащее получают в результате образования групп, сказуемое - в результате расчетов и подсчетов.

В зависимости от построения подлежащего, таблицы делятся на простые, групповые и комбинационные.

Простыми называются таблицы, которые содержат в подлежащем лишь перечень единиц наблюдения без статистической группировки. Простые таблицы делятся на перечневые, территориальные, хронологические.

Сопоставление однородных простых таблиц дает возможность установить некоторые тенденции в изменении изучаемых явлений.

Макет простой таблицы выглядит следующим образом:

Год	Численность РФ (млн. чел.)

Групповыми называются таблицы, подлежащее которых содержит перечень групп единиц наблюдения по одному варьирующему признаку, наиболее существенному для данного явления. Сказуемое групповых таблиц содержит показатели, характеризующие подлежащее. Групповые таблицы позволяют анализировать изучаемое явление, оценить его структуру. Макет групповой таблицы имеет вид:

Группы предприятий по уровню энерговооруженности, тыс. кВт-ч в год	Число предприятий
Итого	

Комбинационные таблицы содержат группировку единиц наблюдения по двум и большему числу варьирующих признаков. Группы, образованные по одному ведущему признаку, подразделяются на подгруппы по второму признаку. Если группировка ведется по трем варьирующим признакам одновременно, то подгруппы, образованные по второму признаку, подразделяются на подгруппы по третьему признаку. Таким образом, устанавливается зависимость между группами, образованными варьирующим признаком и показателями сказуемого.

Макет комбинационной таблицы имеет вид:

Группы предприятий по выработке продукции на одного работающего (тыс. шт.)	Группы предприятий по энерговооруженности (тыс. кВт-ч на одного работающего)	Итого
Итого		

Таблицы могут различаться в зависимости от разработки сказуемого.

В таблицах с простой разработкой сказуемого показатели располагаются параллельно друг другу и не зависят друг от друга:

Предприятие	Количество	М	Ж	Образование

	рабочих			среднее	высшее	Незаконченное высшее
Итого						

В таблицах с комбинированной разработкой сказуемого показатели располагаются в комбинации друг с другом и зависят друг от друга:

Предприятие	Количество рабочих	Образование					
		среднее		высшее		незаконченное высшее	
		М	Ж	М	Ж	М	Ж
Итого							

Правильное, статистически грамотное оформление таблицы обеспечивает максимальную компактность и рельефность содержащихся в ней данных, что значительно облегчает и упрощает анализ. Чтение и анализ таблицы начинается с заголовка, затем переходит к подлежащему и сказуемому. Особо важные и существенные выводы можно получить на основе системы статистических таблиц, содержащих разностороннюю характеристику исследуемого объекта.

Контрольные вопросы

1. Расскажите о содержании и видах сводки.
2. В чем заключается техника и организация сводки?
3. Дайте определение группировки статистических данных.
4. По каким признакам осуществляется группировка?
5. Дайте понятие классификаций.
6. Назовите основные задачи группировки и соответствующие этим задачам виды группировок (приведите примеры).
7. Расскажите о построении типологической группировки.
8. Какое значение имеет правильный выбор признаков группировки?
9. Расскажите о построении структурной группировки.
10. Поясните роль аналитических группировок.
11. Дайте определение статистической таблицы.
12. Назовите виды статистических таблиц по характеру подлежащего.
13. Каковы разновидности статистических таблиц по виду сказуемого.

4. АНАЛИЗ РЯДОВ РАСПРЕДЕЛЕНИЯ

4.1. Общее понятие о рядах распределения и порядок их построения

Статистическое исследование начинается с учета фактов о явлениях и процессах общественной жизни и сбора полученных на этой основе числовых первичных данных. Полученные таким образом материалы представляют собой неупорядоченную совокупность. Выявить какие-либо общие тенденции или закономерности по такой совокупности не представляется возможным. Поэтому собранный материал определенным образом систематизируется и уже после этого дается сводная характеристика всей совокупности. Результатом систематизации статистических данных могут быть ряды распределения. Рядом распределения в статистике называется упорядоченное распределение единиц совокупности на группы по какому-либо варьирующему признаку.

Признаки, по которым проводится группировка, могут быть атрибутивными (не имеющими количественной меры) и количественными.

Ряды, построенные по атрибутивному признаку, называются атрибутивными рядами распределения. Примером атрибутивного ряда распределения может служить распределение рабочих цеха по полу или распределение рабочих цеха по профессии (токарь, слесарь, фрезеровщик и т.д.). Атрибутивные ряды распределения показывают состав совокупности по тем или иным признакам.

Ряды распределения, построенные по количественному признаку, называются вариационными рядами.

В вариационном ряду различают:

- варианты x - отдельные значения группировочного признака;
- частоты m - числа, показывающие как часто встречаются варианты;

$\Sigma m = n$, где n - объем статистической совокупности;

- частости f - частоты, выраженные относительными значениями (процентами или долями единиц);

$$f = \frac{m}{n} \cdot 100\%$$

Так, если признак имеет прерывный (дискретный) характер (т.е. все его значения являются целыми числами, например, квалификация рабочего, число членов семьи и т.д.), то строят дискретный ряд распределения; если признак имеет непрерывный характер (т.е. может принимать сколь угодно мало отличающиеся друг от друга значения, например, размер заработной платы, процент выполнения плана и т.д.), то строят интервальный ряд распределения.

Примером дискретного ряда распределения могут быть данные табл. №4.1.

Как видно из таблицы, самым распространенным в цехе оказался третий тарифный разряд. Кроме того, в таблице №4.1 легко обнаруживается определенная закономерность в распределении частот. Сначала они растут до определенного значения признака, а затем, по мере роста признака частоты убывают, т.е. ряд распределения дает возможность установить характер распределения единиц совокупности по изучаемому признаку.

Таблица 4.1.

Распределение рабочих предприятия по тарифному разряду.

Варианты (X)	Абсолютная частота (m)	Относительная частота f (%)
1.	12	6,3
2.	48	25,3
3.	56	29,4
4.	37	19,4
5.	23	12,2
6.	14	7,4
Всего:	190	100%

Рассмотрим методику вычисления основных характеристик интервального ряда на примере распределения фирм города по расходам на рекламу.

Построения интервального ряда распределения в равных интервалах начинают с определения длины интервала группировки *(i)* по формуле:

$$i = \frac{r}{k},$$

где r - размах вариации признака;

k - количество интервалов.

Размах вариации признака определяется как разность между максимальным и минимальным значением признака.

$$r = Xmax - Xmin$$

Количество интервалов вычисляется по формуле Стреджесса, в зависимости от объема статистической совокупности *(n)*

$$k = 1 + 3{,}322\ lgn$$

При определении нижней границы первого интервала (a_1) рекомендуется отступить от минимального значения признака на половину длины интервала:

$a_1 = X_{min} - i/2$

Тогда верхняя граница первого интервала a_2 равна

$a_2 = a_1 + i$

Верхние границы устанавливаются таким образом, чтобы они совпадали с нижними границами последующих интервалов. Если значение признака совпадает с верхней границей предыдущего интервала и нижней границей последующего, признак можно отнести к любому из этих интервалов. При этом принцип отнесения значения признака к тому или иному интервалу должен оставаться неизменным для построения данного интервального ряда. Таким же образом строятся остальные интервалы до тех пор, пока максимальное значение признака не попадет в интервал.

Данные таблицы №4.2 показывают, что наиболее многочисленную группу составляют фирмы, расходы на рекламу которых лежат в пределах 15-20 т.р. Для групп выше и ниже этой характерно убывающее количество фирм.

Таблица 4.2

Распределение фирм города по расходам на рекламу.

Интервалы по X (тыс.руб.)	Абсолютная частота (m)	Относительная частота f (%)
5 – 10	46	4,6
10 – 15	123	12,3
15 – 20	525	52,5
20 – 25	228	22,8
25 – 30	35	3,5
30 – 35	28	2,8
35 - 40	12	1,2
40 - 45	3	0,3
Итого:	1000	100%

Для наглядности вариационный ряд можно представить в виде полигона, гистограммы, кумуляты и других графиков.

Полигон используется при графическом изображении дискретных вариационных рядов. Для его построения в прямоугольной системе координат по оси абсцисс в масштабе откладываются ранжированные значения варьирующего признака, а по оси ординат наносится шкала для выражения величины частот. Полученные на пересечении абсцисс и ординат точки соединяются отрезком прямой линии, в результате получают ломаную линию, называемую полигоном частот.

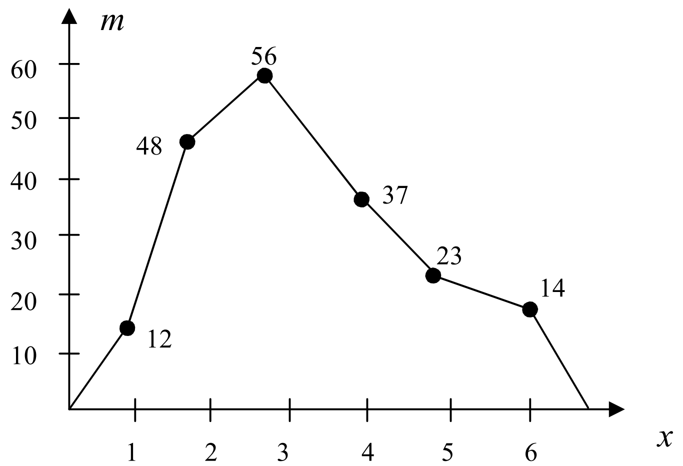


Рис. №4.1. Полигон распределения рабочих предприятия по тарифному разряду

Гистограмма применяется для графического изображения интервального вариационного ряда. При построении гистограммы на оси абсцисс откладываются величины интервалов, а частоты изображаются прямоугольниками, построенными на соответствующих интервалах. Высота столбиков в случае равных интервалов должна равняться частотам.

Если интервальный ряд построен в неравных интервалах, то при построении гистограммы по оси ординат наносят не частоты, а плотность распределения признака в соответствующих интервалах. Плотность распределения - это частота, рассчитанная на единицу длины интервала.

Интервальный ряд тоже можно представить в виде полигона. При этом предполагается, что в пределах интервалов частоты распределены равномерно. В этом случае их можно отнести к значению признака, которое находится в центре интервала *(Хц)* и определяется как полусумма нижней и верхней границ интервалов. Практически это означает, что полигон можно получить, соединив отрезками прямых точки с координатами *Хц* и *m.*

Ряды распределения можно также представить в виде кумуляты накопленных частот (кривой сумм интегральной функции распределения). Накопленные частоты получают суммированием частоты данного варианта и всех предыдущих. При построении кумуляты на оси абсцисс откладываются значения признака, а на оси ординат - накопленные частоты. Соединив полученные точки прямыми линиями, получим кумуляту распределения (рис. №4.3).

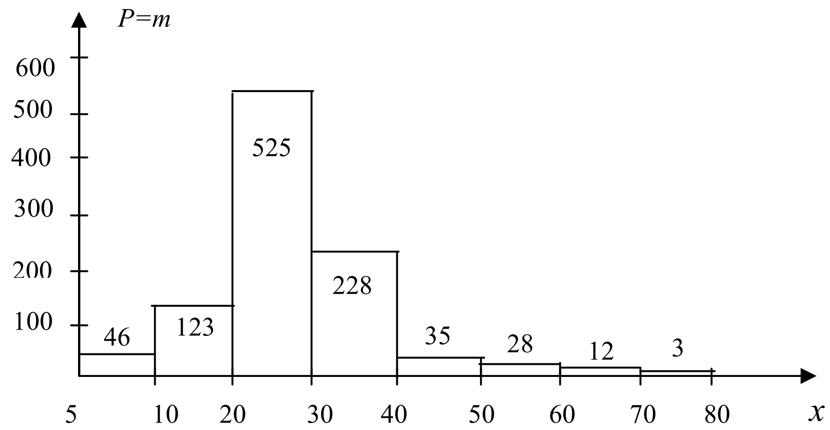


Рис. №4.2. Гистограмма распределения фирм города по расходам на рекламу

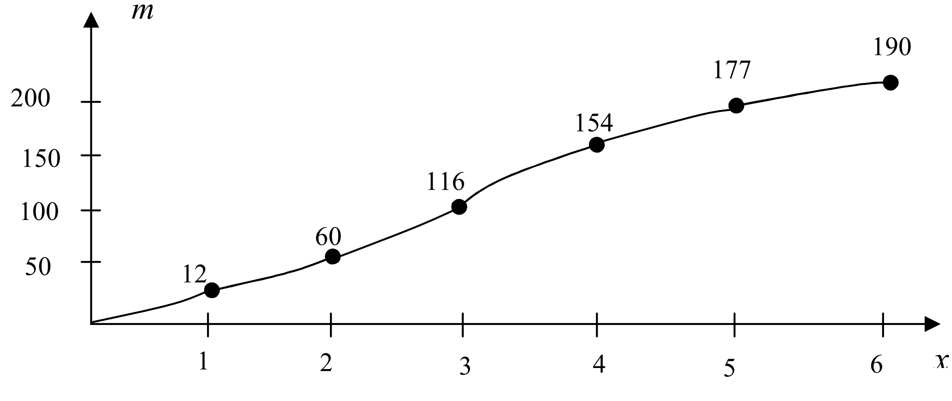


Рис. №4.3. Кумулята распределения рабочих предприятия по тарифному разряду

4.2. Статистические характеристики рядов распределения

Для выявления особенностей вариационных рядов рассчитывают ряд показателей - статистических характеристик ряда. Различают статистические характеристики положения и статистические характеристики рассеивания. В первую группу входят средние, мода и медиана, во вторую - показатели вариации.

Важнейшей характеристикой вариационного ряда является среднее, выражающее общий уровень, общую меру варьирующего признака. Средней называется показатель, дающий обобщенную характеристику варьирующего признака единиц однородной совокупности. Статистическая обработка методом средних заключается в замене индивидуальных значений варьирующего признака, т.е. в замене $X_1, X_2, ..., X_n$ некоторой уравненной величиной $\bar{x}$.

В процессе анализа вариационных рядов могут использоваться различные средние, вид которых (арифметическая, геометрическая, гармоническая и т.д.) в конкретном случае зависит от характера исследуемой совокупности и варьирующего признака, подлежащего усреднению.

В нашем примере средней (средним уровнем выполнения задания) является взвешенная средняя арифметическая:

$$\bar{x} = \frac{\Sigma Xц * m}{\Sigma m} = \frac{15160}{120} = 126,33\%$$

Для расчета средней использованы данные четырех столбцов табл. №4.3.

Таблица №4.3.

Данные для вычисления средней арифметической и дисперсии прямым способом

Выполнение задания (%)	Частота m	Центр интервала $X_ц$	$Xц*m$	$Xц- \bar{X}$	$/ Xц - \bar{X} / \bullet m$	$(Xц- \bar{X})^2$	$(Xц- \bar{X})^2 \bullet m$
1	2	3	4	5	6	7	8
90-100	4	95	380	-31,33	125,32	981,56	3926,24
100-110	13	105	1365	-21,33	277,29	454,96	4914,48
110-120	21	115	2415	11,33	237,93	128,36	2695,56
120-130	36	125	4500	1,33	47,88	1,76	63,36
130-140	28	135	3780	8,67	242,76	75,16	2104,48
140-150	11	145	1595	18,67	205,37	348,56	3834,16
150-160	3	155	465	28,67	86,01	821,96	2465,88
160-170	4	165	660	38,67	154,68	1495,36	5981,44
Итого	120	-	15160	-	1377,24		23151,82

Для упрощения расчетов и проверки правильности вычислений среднюю арифметическую можно исчислить более простым способом, который основан на использовании свойств средней арифметической и сводится к замене действительных значений варианта признака x к новым x'.

$$x' = \frac{Xц - C}{t}$$

В этой формуле С - новое начало отсчета; его следует выбирать равным значению центра интервала $(X_ц)$, которому соответствует наибольшая частота. В нашем случае С = 125, i = 10 (длина интервала). Для удобства расчетов данные сведём в табл. №4.4.

Таблица №4.4.

Данные для вычисления средней арифметической и дисперсии упрощенным способом.

$X_ц$	x'	m	$x' \cdot m$	$x'^2 \cdot m$
95	-3	4	-12	36
105	-2	13	-26	52
115	-1	21	-21	21
125	0	36	0	0
135	+1	28	+28	28
145	+2	11	+22	44
155	+3	3	+9	18
165	+4	4	+16	64
Итого		120	16	263

После выбора нового начала отсчета рассчитывают значения x' для каждого центра интервала:

$$x'_1 - \frac{(95-125)}{10} = -3; \qquad x'_2 - \frac{(105-125)}{10} = -2; \qquad \dots \qquad x'_8 - \frac{(165-125)}{10} = +4;$$

Далее находят произведение $x' \cdot m$ и заносят их в табл. №4.4.

Среднюю арифметическую для новых значений вариантов x' рассчитывают по формуле:

$$\bar{x} = \frac{\sum x' \cdot m}{\sum m} = 16/120 = 0.133$$

Действительное значение $\bar{x}$ можно получить следующим образом:

$$\bar{x} = \bar{x}' \cdot i + C = 0{,}133 \cdot 10 + 125 = 126.33\%,$$

что подтверждает правильность расчетов. Кроме средней арифметической находят структурные средние варьирующего признака - моду и медиану.

Модой в статистике называется наиболее часто встречающийся вариант ряда, т.е. значение признака, которому соответствует наибольшая частота.

При исчислении моды *Мо* для интервального вариационного ряда определяют модальный интервал (интервал с наибольшей частотой), а затем рассчитывают моду по формуле

$$M_o = x_o + i \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2}, \text{ где}$$

X_0 - начало модального интервала;
Δ_1 - разность частот модального и домодального интервалов;
Δ_2 - разность частот модального и замодального интервалов.

Для нашего примера мода находится в интервале 120-130%, поскольку именно на этот интервал приходится наибольшая частота (36), следовательно, *Хо* = 120.

Отсюда

$M_O = 120 + 10*(36 - 21)/((36 - 21) + (36 - 28)) = 126{,}52\%$

Это означает, что в данном цехе чаще всего рабочие выполняют задание на 126,52%.

Медиана - это значение признака, которым обладает элемент, находящийся в центре ранжированной (упорядоченной) совокупности, т.е. слева и справа от которого находится одинаковое число элементов совокупности. Ряд называется ранжированным, если его элементы расположены в порядке возрастания (убывания) значений признака.

Исчисление медианы *Ме* начинается с определения медианного интервала путем накапливания (суммирования) частот, начиная с 1-го интервала. Интервал, в котором сумма накопленных частот равна половине объема совокупности *(n/2)* или превышает ее, и будет медианным. В нашем случае *n* = 120, а *n/2* = 60. Значит, медиана находится в интервале 120 - 130, так как именно в этом интервале сумма накопленных частот (4+13+21+36= 74) превышает *n/2.*

Далее рассчитывается медиана:

$$M_e = x_e + i \cdot \frac{n/2 - \sum m_{e_{ПР}}}{m_e}, \text{ где}$$

$X_e = 120$ - начало медианного интервала;
$i = 10$ - длина интервала;
$\sum m_{e_{ПР}}$ =4+13+21=38 - сумма накопленных частот, предшествующих медианному интервалу;
$m_e = 36$ - частота медианного интервала.

Отсюда

$$M_e = 120 + 10 \cdot \frac{60-38}{36} = 126{,}11\%$$

Это означает, что в данном цехе одна половина рабочих (60 человек) выполняет задания ниже, чем на 126,11%, а другая - выше 126,11%.

Исчислив среднюю арифметическую для вариационного ряда, мы ничего не знаем о том, как отдельные значения изучаемого признака группируются вокруг нее, близко или далеко отстоят они от средней. Поэтому возникает задача измерения вариации (колеблемости) признака.

Вычислим характеристики колеблемости вариационного ряда.

Наиболее простой мерой колеблемости является размах вариации R = Х*max -Xmin,* который мы уже определили при исчислении длины интервала вариационного ряда *(R* = 164,7 - 98,3 == 66,4%). Поскольку каждое индивидуальное значение признака может отличаться от средней, мерой вариации может служить средняя из отклонений каждого варианта от средней, исчисленной для всего ряда.

Таким образом, распределение отклонений можно найти, исчислив отклонения всех вариантов от средней. Для того чтобы им дать обобщающую характеристику, необходимо вычислить среднюю из этих отклонений. Показателями, исчисленными по такому принципу, являются среднее линейное отклонение, дисперсия и среднее квадратическое отклонение.

Среднее линейное отклонение Q представляет собой среднюю из абсолютных отклонений отдельных вариантов от средней (все отклонения берутся с положительным знаком):

$$Q = \frac{\sum \left| x_ц - \bar{x} \right|}{\sum m}$$

Этот показатель можно рассчитать по данным табл. №4.3 (графа 6):

$Q = 1377.24/120 = 11.48\%$

Однако чаще отклонения от средней возводят в квадрат и из квадратов отклонений исчисляют среднюю, т.е. дисперсию или среднее квадратическое отклонение.

Среднее квадратическое отклонение (σ) определяется путем извлечения квадратного корня из дисперсии:

$$\sigma = \sqrt{\sigma^2}$$

представляющей собой средний квадрат отклонений вариантов от средней:

$$\sigma^2 = \frac{\sum (Xц - \bar{x})^2}{\sum m}$$

Вычислим дисперсию для нашего примера по данным табл. №4.3 (графа 8):

$$\sigma^2 = \frac{23151,82}{720} = 192,93\%$$

Тогда

$$\sigma = \sqrt{192.93} = 13,89\%$$

Абсолютные значения R и σ зависят не только от степени вариации признака, но и от абсолютных уровней вариантов и средней. Поэтому сравнивать эти характеристики вариационных рядов с разными уровнями нельзя. Для такого сравнения необходимо исчислить относительные показатели вариации.

Так, отношение размаха вариации к средней арифметической величине называется коэффициентом осцилляции:

R *80*

$$V_R = \frac{}{\bar{x}} \; 100\% = \frac{}{126,33} \cdot 100\,\% = 63,3\%$$

где R = $X_{max} - X_{min}$ = 170% - 90% = 80%

Если взять отношение среднего линейного отклонения к средней арифметической, получим линейный коэффициент вариации:

$$V_Q = \frac{Q}{\bar{x}} * 100\% = \frac{11.48}{126.33} * 100 = 9.08\%$$

Процентное отношение среднего квадратического отклонения к средней арифметической называется коэффициентом вариации:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{13.8}{126.33} \cdot 100 = 10.99\%$$

Коэффициент вариации можно использовать при сравнении вариации разных явлений. Например, с помощью *V* можно сравнивать колеблемости в производительности труда групп рабочих, занятых производством разных видов продукции.

Из всех рассмотренных показателей σ и *V* являются основными мерилами надежности $\bar{x}$. Чем меньше σ и *V,* тем однороднее совокупность, следовательно, тем типичнее средняя.

Простейшие свойства $\bar{x}$ и σ.

1) Если все частоты умножить или разделить на какое-либо постоянное число, то $\bar{x}$ и σ не изменяться.

$$\bar{x} = \frac{\sum x \cdot m}{\sum m} = \frac{\sum x \cdot m \cdot k}{\sum m \cdot k} = \bar{x}$$

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot m}{\sum m}} = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot m \cdot k}{\sum m \cdot k}} = \sigma$$

2) Если все варианты признака умножить или разделить на какое-либо постоянное число, то $\bar{x}$ умножится или разделится на это число, а σ умножится или разделится на модуль этого числа.

$$z = x \cdot k$$

$$\bar{z} = \frac{\sum z \cdot m}{\sum m} = \frac{\sum x \cdot k \cdot m}{\sum m} = k \cdot \frac{\sum x \cdot m}{\sum m} = \bar{x} \cdot k$$

$$\sigma_z = \sqrt{\frac{\sum (z - \bar{z})^2 \cdot m}{\sum m}} = \sqrt{\frac{\sum (x \cdot k - \bar{x} \cdot k)^2 \cdot m}{\sum m}} = \sqrt{\frac{\sum k^2 \cdot (x - \bar{x})^2 \cdot m}{\sum m}} = \sigma_x \cdot |k|$$

3) Если ко всем вариантам признака прибавить или вычесть какое-либо постоянное число, то $\bar{x}$ увеличится или уменьшится на это число, а σ останется без изменений.

$$z = x + k$$

$$\bar{z} = \frac{\sum z \cdot m}{\sum m} = \frac{\sum (x + k) \cdot m}{\sum m} = \frac{\sum x \cdot m}{\sum m} + \frac{k \cdot \sum m}{\sum m} = \bar{x} + k$$

$$\sigma_z = \sqrt{\frac{\sum (z - \bar{z})^2 \cdot m}{\sum m}} = \sqrt{\frac{\sum (x + k - \bar{x} - k)^2 \cdot m}{\sum m}} = \sigma_x$$

4) Алгебраическая сумма отклонений вариантов от средней равна нулю.

$$\sum (x - \bar{x}) = 0; \quad \bar{x} = \frac{\sum x}{n}; \quad \sum x = \bar{x} \cdot n; \quad \sum x - \sum \bar{x} = \bar{x} \cdot n - \bar{x} \cdot n = 0$$

5) Дисперсия равна среднему квадрату минус квадрат самой средней.

$$\sigma^2 = \frac{\sum x^2 \cdot m}{\sum m} - \overline{x}^2$$

$\overline{X}$ и σ^2 для объединенной статистической совокупности.

При анализе рядов распределения может решаться задача, когда, зная $\overline{x}$ и σ^2 для частных совокупностей, необходимо определить эти показатели для объединенной статистической совокупности. В этом случае используются два правила:

1) $\overline{x}$ для объединенной статистической совокупности равна средней арифметической из частных средних, взвешенных объемами частей.

$$\overline{x}_{об.} = \frac{\sum \overline{x}_i \cdot n_i}{\sum n_i}$$

2) Дисперсия объединенной статистической совокупности равна средней арифметической из частных дисперсий, плюс дисперсия самих частных средних относительно общей средней.

$$\sigma^2_{об.} = \overline{\sigma}_i^2 + \delta_i^2; \quad \overline{\sigma}_i^2 = \frac{\sum \sigma_i^2 \cdot n_i}{\sum n_i}; \quad \delta_i^2 = \frac{\sum (\overline{x}_1 - \overline{x}_{об.})^2 \cdot n_1 + \sum (\overline{x}_2 - \overline{x}_{об.})^2 \cdot n_2}{\sum n_i}$$

Упрощенный способ расчета $\overline{x}$ и σ.

Сущность этого способа заключается в том, что прежние значения x заменяются на x' по следующему соотношению:

$$x' = \frac{x - C}{i}, \text{ где}$$

С – это новое начало отсчета, в качестве которого используется вариант x или центр интервала $x_ц$, имеющие максимальную абсолютную частоту.

Порядок расчета:

1) Переход от прежних вариантов к новым по указанному соотношению.

2) Определение $\overline{x}'$ для новых вариантов: $\overline{x}' = \frac{\sum x' \cdot m}{\sum m}$.

3) Переход к действительному значению $\overline{x}$

4) Определение σ' для новых вариантов: $\sigma' = \sqrt{\frac{\sum \overline{x}^2 \cdot m}{\sum m} \cdot \overline{x}'^2}$

5) Переход к действительному значению σ: $\sigma = \sigma' \cdot i$.

4.3. Оценка формы кривой распределения

В приведенных рядах распределения рабочих по квалификации и уровню выполнения заданий можно заметить определенную связь в изменении частот и значений варьирующего признака. Такие изменения частот в вариационных радах называются закономерностями распределения. Закономерности выражают свойства явлений, условия, влияющие на образование вариации признака. Так, например, распределение рабочих по процентному уровню выполнения заданий и ее вариацию.

Эти условия выражаются в квалификационном и возрастном составе рабочих цеха, условиях труда и т.д. Однако, кроме этих общих условий, на фактическое выполнение заданий также влияют и случайные причины: болезнь отдельных рабочих, настроение, работа городского транспорта и т.д. Для обобщенной характеристики особенностей формы распределения используются кривые распределения. Кривая распределения выражает закономерность распределения единиц совокупности по значению варьирующего признака. Различают эмпирические и теоретические кривые распределения. Эмпирическая кривая распределения - это фактическая кривая распределения, полученная по данным наблюдения, в которой отражаются как общие, так и случайные условия, определяющие вариацию (полигон, гистограмма). Теоретическая кривая распределения - это кривая, выражающая функциональную связь между изменением варьирующего признака и изменением частот и характеризующая определенный тип распределения.

На практике часто возникает необходимость оценки степени близости эмпирического распределения одному из теоретических законов распределения. При обработке эмпирических данных чаще всего используется нормальный закон распределения случайной величины, описываемый уравнением

$$P_t = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}t^2}$$

где P_t - ордината кривой нормального распределения (частости);

t - нормальное отклонение, равное $t = \dfrac{x - \bar{x}}{\sigma}$

π =3,1415

e=2,7182

Таким образом, кривую нормального распределения можно построить по двум параметрам: средней арифметической и среднему квадратическому отклонению.

Кривая нормального распределения имеет ряд особенностей. Она симметрична. Если из точки на оси абсцисс, соответствующей значению $\bar{x}$, восстановить перпендикуляр, то по обе стороны от него образуются две подобные равномерно убывающие ветви, ассимптотически приближающиеся к оси абсцисс. Поэтому средняя арифметическая, мода и медиана кривой нормального распределения совпадают. Кривая имеет три особые точки: точку максимума (соответствует значению = $\bar{x}$) и две точки перегиба (при $x = \bar{x} \pm \sigma$).

Алгоритм построения теоретического нормального распределения по фактическим данным нашего вариационного ряда предусматривает:

1. Расчет значений нормативного отклонения по формуле

$$t = \frac{x_ц - \bar{x}}{\sigma}$$

Таблица №4.5.

Данные для расчета теоретических частот и критерия согласия

$X_ц$	m	t	P_t	$\tilde{m}$	$\Sigma_i m$	$\Sigma_i \tilde{m}$	$\left\lvert \sum_i m - \sum_i \tilde{m} \right\rvert$
1	2	3	4	5	6	7	8
95	4	-2,256	0,0310	2,59	4	2,59	1,41
105	13	-1,536	0,1219	10,54	17	13,13	3,87
115	21	-0,816	0,2850	24,62	38	37,75	0,25
125	36	-0,096	0,3970	34,29	74	72,04	1,96
135	28	0,624	0,3292	28,44	102	100,48	1,52
145	11	1,344	0,1626	14,05	113	114,53	1,53
155	3	2,064	0,0478	4,14	117	118,67	1,67
165	4	2,784	0,0084	0,73	120	119,40	0,60
Итого	120	-	-	119,40	-	-	-

Для этого используются данные табл. №4.3 (графа 5),а также найденные ранее значения $\bar{x}$ == 126,33% и σ = 13,89%. Тогда

t_1=-31.33/13.89=-2.256 t_2=-21.33/13.89=-1.536 . . .

t_8=38.67/13.89=2.784

Полученные значения t заносятся в табл. №4.5.

2. Нахождение по специальной таблице значений функции P_t, соответствующих рассчитанным значениям $t1$, $t2 \ldots t8$, а также значениям $t = 0$ (точка максимума) и $t = \pm 1$ (точки перегиба).

3. Вычисление теоретических частот по формуле

$$\widetilde{m} = \frac{n \cdot i}{\sigma} \cdot P_t$$

Значения приведены в табл. №4.5. Для нашего примера:

$\widetilde{m}_1$ = 120*10*0.0310/13.89=2.59

$\widetilde{m}_2$ =120*10*0.1219/13.89=10.54

$\widetilde{m}_8$ =120*10*0.0084/13.89=0.73

4. Построение теоретической кривой распределения по значениям ординат (теоретические частоты).

Для этого из центров интервалов, а также из точек $x = \bar{x}$ и $x = \bar{x} \pm \sigma$ на оси абсцисс восстанавливаются значения теоретических частот. Полученные точки соединяются плавной кривой, которая и является теоретической кривой нормального распределения.

Для оценки близости эмпирического распределения к теоретическому (в данном случае к нормальному) используются специальные показатели - критерии согласия.

Рассмотрим применительно к нашему примеру критерий согласия Колмогорова (критерий лямбда):

$$\lambda = \frac{|D|}{\sqrt{n}}$$

где

$|D|$ - максимальная разность по модулю между накопленными эмпирическими и теоретическими частотами ($\Sigma_i m$ и $\Sigma_i \widetilde{m}$, графы 6,7 табл. №4.5).

По специальным таблицам вероятностей для критерия согласия лямбда определяем соответствующую вероятность $P(\lambda)$, по значению которой судим о близости эмпирического и теоретического распределений.

Для нашего примера по табл. №4.5 (графа 8) находим максимальное значение $\left|\sum_i m - \sum_i \widetilde{m}\right|$ и получаем $\lambda = 3.87 / \sqrt{120} = 0.35$

Этому значению критерия лямбда соответствует значение $P(\lambda) = 0{,}999$ (найденное по таблице вероятностей для критерия согласия), т.е. с вероятностью 0,999 можно считать, что отклонения между эмпирическими и теоретическими частотами (распределениями) случайны, и наше распределение рабочих по процентному уровню выполнения заданий соответствует нормальному распределению.

При сравнении фактического распределения с теоретическим важно констатировать не только согласие этих распределений, но и характер их расхождения.

В симметричном распределении средняя арифметическая равна моде и медиане. При отсутствии этого равенства распределение является асимметричным. Для оценки асимметрии пользуются коэффициентом A, который представляет собой отношение разности между средней и модой к среднему квадратичному отклонению. Если $\bar{x} > M_o$, то коэффициент асимметрии положительный, что характеризует правостороннюю асимметрию. Если $\bar{x} < M_o$, то коэффициент отрицательный, что характеризует левостороннюю асимметрию. Для нашего примера с распределением рабочих по процентному уровню выполнения заданий

$$A = \frac{\bar{x} - M_o}{\sigma} = \frac{126.33 - 126.52}{13.89} = -0.013$$

Коэффициент асимметрии A показывает левостороннюю асимметрию.

Для характеристики крутизны ряда распределения пользуются эксцессом $Ек$ или нормированным моментом четвертого порядка

$$E_K = \frac{M_4}{\sigma_4} - 3$$

где M_4 - центральный момент четвертого порядка.

$$M_4 = \frac{\sum m\,(x - \bar{x})^4}{\sum m}$$

Если $E_K > 0$, имеем островершинное (высоковершинное) распределение, а при $E_K < 0$ - плосковершинное (низковершинное) распределение.

Контрольные вопросы.

1. Что представляет собой ряд распределения?
2. Чем отличаются вариационные ряды от атрибутивных? Приведите примеры.
3. В какой последовательности строят интервальные ряды?
4. Для чего применяются полигоны, гистограммы, кумуляты?
5. В чем сущность средней арифметической, моды и медианы? Каков порядок их исчисления?
6. Что представляет собой вариация признака?
7. Какие показатели характеризуют вариацию признака и как они исчисляются?
8. Что выражают кривые распределения? Какие кривые называются эмпирическими, а какие теоретическими?
9. Как строится теоретическая кривая распределения?
10. Что характеризует критерии согласия?
11. Как определяются коэффициенты асимметрии и эксцесса и что они характеризуют?

5. ИСПОЛЬЗОВАНИЕ КОРРЕЛЯЦИОННОГО АНАЛИЗА В ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

5.1. Статистическое изучение связи экономических явлений

Изучение взаимосвязи - одна из важнейших задач общей теории статистики. Исследование объективно существующих связей между явлениями имеет большое значение в экономическом анализе. Познание любого явления будет неполным, если не исследованы его связи с другими явлениями.

В процессе статистического исследования вскрываются причинно-следственные отношения между явлениями, что позволяет выявить факторы (признаки), оказывающие основное влияние на вариацию явлений. Особенности статистического изучения связи между явлениями заключаются в том, что имеется возможность не только выявить наличие и направление связи, но и дать количественную оценку этой связи, определить ее аналитическое выражение. В результате устанавливается, насколько изменение одного признака меняет вариацию другого.

Прежде чем приступить к изучению связи между явлениями необходимо разделить признаки по характеру их роли во взаимосвязи. Признак, определяющий изменение другого, связанного с ним признака, называется факторным или признаком-аргументом.

Признак, изменяющийся под действием факторного признака, называется результативным или признаком-функцией. Например, объем произведенной продукции - это признак-аргумент, балансовая прибыль - признак-функция.

Связи между явлениями ввиду их большого разнообразия классифицируются по ряду признаков. По степени тесноты различают полную или функциональную связь и связь неполную - корреляционную или статистическую. При функциональной связи каждому значению факторного признака (признаков) соответствует строго определенное значение результативного признака. Такие связи обычно выражаются формулами и имеют место в точных науках. При корреляционной связи одному и тому же значению факторного признака могут соответствовать несколько значений результативного признака. Так, при одном и том же объеме произведенной продукции на предприятиях уровень балансовой прибыли может быть различным.

В зависимости от направления действия как функциональные, так и корреляционные связи могут быть прямыми и обратными. При прямой связи направление изменения факторного признака совпадает с направлением изменения результативного признака, т.е. с увеличением признака-аргумента увеличивается и признак-функция и наоборот, с уменьшением признака-аргумента уменьшается и признак-функция. В противном случае между рассматриваемыми признаками существуют обратные связи. Например, чем выше квалификация рабочего, тем выше уровень производительности труда, прямая связь, но тем ниже себестоимость продукции - обратная связь.

По аналитическому выражению (форме) связи могут быть прямолинейными (линейными) и нелинейными. При прямолинейной связи с возрастанием факторного признака происходит непрерывное возрастание (или убывание) результативного признака, что математически выражается уравнением прямой линии. При нелинейной связи с возрастанием факторного признака возрастание (или убывание) результативного признака происходит неравномерно или направление его изменения меняется на обратное, что математически выражается уравнением кривой (параболы, гиперболы и т.п.).

5.2. Корреляционный анализ связей

Для выявления связи между явлениями, ее характера, направления в статистике используются различные методы: методы приведения параллельных рядов, балансовый метод, метод аналитических группировок, графический метод. Однако ни один из перечисленных методов не устанавливает единой меры связи. Такую меру дает корреляционный анализ.

Корреляционный анализ заключается в построении и анализе статистической модели в виде уравнения регрессии (уравнения корреляционной связи), приближенно выражающей зависимость результативного признака от одного или нескольких признаков-факторов.

Корреляционному исследованию предшествует качественный анализ изучаемого явления, связанный при изучении экономических явлений с методами конкретной экономики. Такой предварительный теоретический анализ должен доказать, что между признаком-аргументом и признаком-функцией имеется причинная связь. Необходимо также установить, прямая эта связь или обратная, прямолинейная или более сложная.

Следующим предварительным этапом, предшествующим непосредственно исследованию, является систематизация статистической информации по изучаемым признакам. Удобной формой изложения данных о взаимосвязанных признаках является корреляционная таблица, представляющая собой комбинационную статистическую таблицу. Для примера приведем корреляционную таблицу, характеризующую связь между объемом произведенной продукции x и балансовой прибылью y:

В подлежащем таблицы расположены: по горизонтали - группы предприятий по факторному признаку - объему произведенной продукции, по вертикали - группы предприятий по результативному признаку - балансовой прибыли.

Сказуемым таблицы является численность групп, т.е. частоты повторений вариантов признака.

Таблица №5.1.

Балансовая прибыль (млн. руб.)y	Объем произведенной продукции (млн. руб.) x					Итого, l
	6,0-7,5	7,5-9,0	9,0-10,5	10,5-12	12-13,5	
8,4-10,0	-	-	1	2	4	7
6,8-8,4	-	-	1	2	3	6
5,2-6,8	1	1	-	1	1	4
3,6-5,2	1	2	1	-	-	4
2-3,6	3	1	-	-	-	4
Итого, *h*	5	4	3	5	8	25

Заглавная и итоговая строки образуют ряд распределения по признаку-аргументу, а заглавный и итоговый столбцы - по признаку-функции; при этом каждый внутренний столбец вместе с заглавным образуют частный ряд распределения функции.

По мере увеличения значений аргумента x частные ряды распределения функции y смещаются в сторону больших значений последней, обнаруживая тем самым зависимость между исследуемыми признаками. Это корреляционная зависимость. Ей можно дать следующее определение: зависимость называется корреляционной, если каждому значению аргумента соответствует частный ряд распределения функции, причем все эти ряды закономерно смещаются относительно друг друга по мере изменения аргумента.

Более наглядное представление о существующей связи между исследуемыми признаками дает графическое изображение статистических характеристик, полученных в результате сводки и обработки исходной информации. Так, выявленную с помощью корреляционной таблицы связь между объемом произведенной продукции x и балансовой прибылью y наглядно показывает график, на оси абсцисс которого отложены значения признака x, а на оси ординат - значения признака y. В этой системе координат по соответствующим значениям аргумента и функции строится сетка, в клетках которой точками наносятся частоты повторений вариантов признаков. Такой точечный график в целом имеет форму квадрата (рис. №5.1).

По характеру расположения точек можно судить о направлении и силе связи. Если точки беспорядочно разбросаны по всему полю, это свидетельствует об отсутствии зависимости между двумя признаками. Если точки эллипсом концентрируются вокруг оси, идущей от нижнего левого угла в верхний правый, как это имеет место в нашем примере (см. рис. № 5.1), то имеется прямая зависимость между варьирующими признаками. Если точки будут концентрироваться вокруг оси, идущей от верхнего левого угла в нижний правый, то имеется обратная зависимость.

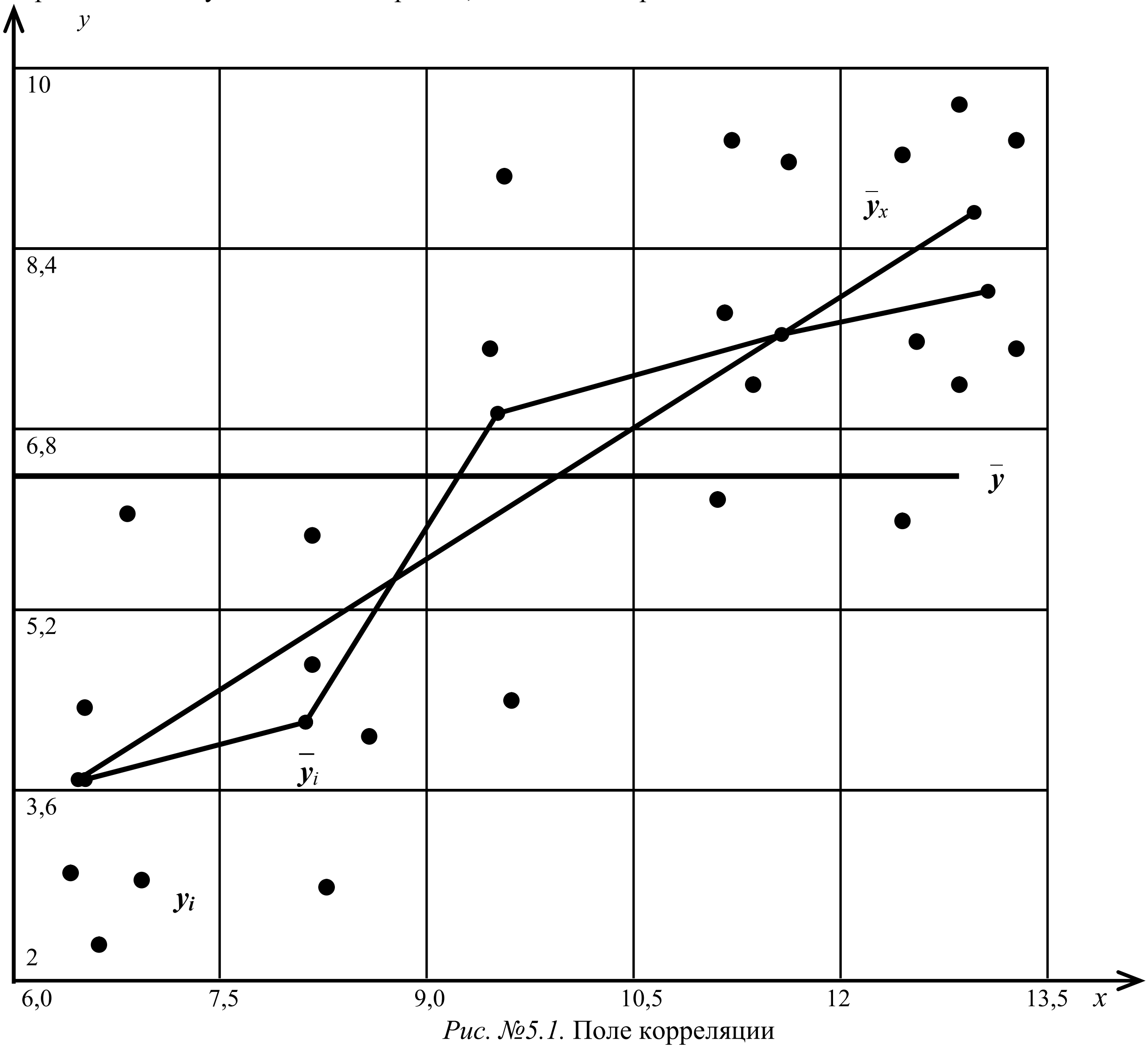


Рис. №5.1. Поле корреляции

5.3. Исследование формы связи

Начальным этапом непосредственно корреляционного анализа является установление формы связи между анализируемыми признаками. Это имеет решающее значение для оценки взаимосвязи, поскольку дальнейшие самые тщательные расчеты могут быть обесценены, если форма связи определена неверно. Форма связи устанавливается в результате решения первой задачи теории корреляции. Эта задача заключается в том, чтобы на основе наблюдения над большим количеством данных выяснить, как изменилась бы функция при изменении одного из ее аргументов, если бы другие аргументы не менялись. При анализе особенно экономических явлений следует иметь в виду, что все другие (неучтенные) факторы в действительности меняются и маскируют этим интересующую нас зависимость.

Корреляционная зависимость выражается в закономерном смещении частных рядов распределения функции по мере роста аргумента. Это смещение может быть быстрым, медленным, может происходить в сторону увеличения функции и в сторону уменьшения ее значений. В зависимости от этого частные ряды распределения "ползут" вверх или вниз. Таким образом, при изучении корреляционной зависимости необходимо ответить на вопрос: как быстро и в каком направлении смещаются частные ряды распределения функции?

Для этого следует дать точную оценку положения частных рядов распределения функции на оси ординат, что можно сделать с помощью метода средних путем определения групповых средних для каждого частного ряда распределения функции. Результаты расчетов сводятся в таблицу, где вместо интервалов группировки используются их центры.

Таблица №5.2.

$У'$	$X_ц$ / $У_ц$	6,75	8,25	9,75	11,25	12,75	Итого, U
+1	9,2	-	-	1	2	4	7
0	7,6	-	-	1	2	3	6
-1	6	1	1	-	1	1	4
-2	4,4	1	2	1	-	-	4
-3	2,8	3 '	1	-	-	-	4
	итого, hi	5	4	3	5	8	25
	$\sum_i my'$	-12	-8	-1	1	3	-
	$\overline{У_i'}$	-2,4	-2	-0,33	0,2	0,37	-
	$\overline{У}_i$	3,76	4,4	7,07	7,92	8,2	-

Полученные значения $\overline{y_i}$ наносятся на поле корреляции в виде точек (см. рис. №5.1), абсциссами которых являются середины интервалов признака. Последовательные значения $\overline{y_i}$ соединяются отрезками прямой. Полученная ломаная линия называется эмпирической линией регрессии и показывает, как быстро смещаются частные ряды распределения функции y по мере роста аргумента x или как в среднем изменяется функция по мере роста аргумента.

Эмпирическую линию регрессии всегда следует рассматривать на фоне индивидуальных точек, которые она усредняет. С помощью вычисления групповых средних нейтрализуется влияние прочих случайных факторов и выявляется то общее, что свойственно каждой частной совокупности.

Линия регрессии отмечена зигзагами, которые носят случайный характер и вызваны влиянием прочих, неучтенных факторов. Зигзаги сильнее на тех участках линии, где наблюдений мало и наоборот, там, где средняя рассчитана на достаточно большом числе наблюдений, они незначительны. По мере увеличения числа наблюдений зигзаги эмпирической линии регрессии сглаживаются, и она все более приближается к закону распределения, лежащему в ее основе.

Предельная теоретическая линия регрессии дает истинную оценку формы связи. Она показывает, как изменялась бы функция под влиянием одного из аргументов, если бы другие аргументы не менялись, т.е. решается первая задача теории корреляции, поэтому отыскание предельной теоретической линии регрессии и есть решение этой задачи.

Прямой путь отыскания предельной теоретической линии регрессии - это неограниченное увеличение числа наблюдений при одновременном, не столь быстром, уменьшении длины интервала аргумента до тех пор, пока закономерность, лежащая в основе зависимости y от x, не выявится с удовлетворяющей нас точностью. Поскольку на практике мы располагаем ограниченной информацией, то используем косвенные приемы обнаружения закономерностей, а именно выравнивание эмпирической линии регрессии.

Универсальный способ, с помощью которого можно было бы подобрать ту или иную кривую для выравнивания эмпирической линии регрессии, не существует, и поэтому часто используют несколько путей в сочетании друг с другом.

Эмпирический путь. О типе теоретической линии в этом случае судят по внешнему виду эмпирического графика. Этот путь основан на законе больших чисел и может дать хорошие результаты, если число наблюдений достаточно велико. Среди нескольких типов кривых предпочтение отдают тому, который зависит от меньшего числа параметров.

Очевидно, при малом числе наблюдений трудно определить сколько-нибудь сложный тип зависимости, так как особенности последней вуалируются случайными колебаниями. В этом случае желательно увеличить число наблюдений.

Теоретический путь.

1. Опыт предыдущих исследований. Если подобного рода зависимости изучались ранее и описаны в литературе, то учитывается опыт этих исследований. Этот путь неприменим для новых исследований, в которых данная зависимость изучается впервые.

2. Эксперимент. Этот путь можно применить там, где проводились аналогичные исследования и ставилась задача проверки и конкретизации зависимости в новых условиях или для иного материала. Цель эксперимента - искусственно сократить случайные зигзаги эмпирической линии регрессии путем закрепления или сокращения изменчивости ряда факторов, вызывающих колеблемость функции. Недостатком является ограниченная возможность экспериментирования во многих областях, особенно в экономических.

Логический анализ. В той или иной форме логический качественный анализ изучаемого явления должен сопутствовать любому другому способу определения типа зависимости. Значение и степень достоверности выводов логического анализа обусловлены уровнем научных знаний и характером имеющейся информации.

В нашем примере такой анализ показывает, что с ростом объема продукции балансовая прибыль должна возрастать. При этом, если обратиться к эмпирической линии связи, можно предположить, что наиболее удобно ее выравнивать по прямой. Видимо, в основе этой зависимости в данных условиях лежит прямолинейная связь. Уравнение прямой в общем виде $\overline{y_x} = A_o + A_1 \cdot x$. Найти теоретическое уравнение связи в данном случае значит определить параметры прямой. Их находят способом наименьших квадратов, который дает следующую систему нормальных уравнений для нахождения параметров прямой:

$$\begin{cases} \sum y = a_0 n + a_1 \sum x \\ \sum yx = a_0 \sum x + a_1 \sum x^2 \end{cases}$$

В нашем примере данные сгруппированы, т.е. все суммы берутся во взвешенном виде, а именно: суммы *по x* берутся с частотой h, *по y - с* частотой l, а суммы произведений *yx* - с частотой *m.* В этом случае система нормальных уравнений имеет вид:

$$\begin{cases} \sum ly = a_0 n + a_1 \sum hx \\ \sum (x \sum my) = a_0 \sum hx + a_1 \sum hx^2 \end{cases}$$

Для упрощения расчетов вводится новая переменная y' которая рассчитывается по формуле:

$$y' = \frac{y_ц + C_y}{i_y}$$

где $y_ц$ - центры интервалов функции;

C_y - новое начало отсчета; им является одно из значений центра интервала функции, в котором ожидается появление средней арифметической (в нашем примере - 7,6 тыс. шт.);

i_y - длина интервала функции.

Затем определяются частные средние в новых значениях перемен-

$$\overline{y'_i} = \frac{\sum_i my'}{\sum_i m}$$

ных:

Где *m* - частоты повторений вариантов признаков.

Действительные значения средних для каждого частного ряда распределения функции определяются по формуле

$$\overline{y_i} = \overline{y_i}' i + C_y$$

Таблица №5.3.

	x'	-4	-3	-2	-1	0	Итого l_i	$l_i \cdot У'$	$l_i \cdot y'^2$
$У'$	$X_ц$ / $У_ц$	6,75	8,25	9,75	11,25	12,75			
1	9,2	-	-	1	2	4	7	7	7
о	7.6	-	-	1	2	3	6	0	0
-1	6	1	1	-	1	1	4	-4	4
-2	4,4	1	' 2	1	-	-	4	. -8	16
-3	2,8	3	1	-	-	-	4	-12	36
Итого h_i		5	4	3	5	8	25	-17	63
$\sum_i$ my′		-12	-8	-1	1	3	-17	-	-
X’$\sum_i$ my′		48	24	2	-1	0	73	-	-
hx′		-20	-12	-6	-5	0	-43	-	-
hx'^2		80	36	12	5	0	133	-	-

Для вычисления необходимых при решении системы уравнений сумм можно применить расчетную таблицу № 5.3, в которой для упрощения расчетов используются новые значения переменных

$$y' = \frac{y_ц + C_y}{i_y}, \; x' = \frac{x_ц + C_x}{i_x}$$

С учетом новых значений переменных система нормальных уравнений преобразуется к виду:

$$\begin{cases} \sum ly' = a'_0 n + a'_1 \sum hx \\ \sum (x' \sum my') = a'_0 \sum hx' + a'_1 \sum hx'^2 \end{cases}$$

Подставляя в полученную систему уравнений значение *n = 25* (численность совокупности) и суммы из приведенной таблицы, решим систему уравнений относительно a'_o и a'_1:

$$\begin{cases} -17 = a'_0 25 + a'_1 (-43) \\ 73 = a'_0 43 + a'_1 133 \end{cases}$$

Умножая первое уравнение на -1,72 (-43 : 25), получаем *29,24 = а '$_o$• (-43) + а '$_1$ • 73,96.*
Вычитая из второго уравнения вновь полученное, находим

$59{,}04 * a'_1 = 43{,}76.$

Отсюда

$a'_1 = 43.76/59.04 = 0{,}74$

Подставляя значение a'_1 в первое уравнение, получаем

$-17 = a'_o 25 + 0{,}74\ (-43)$

откуда $25 \cdot a'_o = 14{,}28;\ a_o = 0{,}59$. От параметров a'_o и a'_1 переходим к параметрам a_o и a_1, используя следующие формулы перехода:

$a_o = \mathbf{C}_y + a'_o * i_y - a'_1 * C_x * i_y / i_x$

$a_o = 7{,}6 + 0{,}59 \cdot 1{,}6 - 0{,}74*12{,}75*1{,}6/1{,}5 = -1{,}4:$

$$a_1 = a_1' \frac{i_y}{i_x} = 0.74 \frac{1.6}{1.5} = 0.78$$

Следовательно, теоретическое уравнение связи примет вид

$$\overline{y}'_x = -1.4 + 0.78 * x$$

Подставляя в это уравнение значение $X_ц$, получим при $X_ц = 6{,}75$

$$\overline{y}'_x = -1.4 + 0.78 * 6.75 = 3.865$$

при $X_ц = 12{,}75$ имеем

$$\overline{y}'_x = -1.4 + 0.78 * 12.75 = 8.545$$

Полученные значения ординат наносим на поле корреляции (см. рис. №5.1.), восстанавливая перпендикуляры из центров соответствующих интервалов и отмечая полученные значения точками. Проведенная через эти точки линия, выравнивающая эмпирическую линию, и будет теоретической линией регрессии, а ее поиск, построение, анализ и практическое применение в статистике называют регрессионным анализом.

В случае высокой корреляционной зависимости между *у* и *х* параметр a_1 при ***х***, называемый коэффициентом регрессии, приобретает большое практическое значение. Он показывает, в какой мере увеличивается у с ростом *х*.

В нашем примере прирост на 1 млн. р. произведенной продукции в среднем увеличивает балансовую прибыль на 0,78 млн. руб. Следует лишь иметь в виду, что такие выводы мы вправе делать только для нашей совокупности предприятий при данных условиях их работы. Если данная совокупность и данные условия достаточно типичны, то коэффициент регрессии может быть использован для планирования и нормирования.

Количественную зависимость изменения теоретического значения *у* от изменения *х,* которую выражает коэффициент регрессии a_1, часто удобнее выразить в относительных величинах. Для этого исчисляется коэффициент эластичности Э, показывающий, на сколько процентов увеличивается *у* при увеличении *х* на 1%. Коэффициент эластичности рассчитывается по формуле:

$$Э = a_1 \frac{\overline{x}}{\overline{y}}$$

Так, в нашем примере коэффициент эластичности при *х(с чертой)*=9,75 и *у(счертой) = 6* составит

Э = 0,78 • 9,75 /6 = 1,26.

Следовательно, на 1% прироста объема произведенной продукции балансовая прибыль возрастает на 1,26%.

5.4 Измерение тесноты связи

Следующая задача корреляционного анализа - установление тесноты связи между исследуемыми признаками. Она решается путем определения степени влияния признака-аргумента на признак-функцию или путем определения искажающего влияния прочих факторов на интересующую нас зависимость.

Схематически это можно представить, например, следующим образом (рис. №5.2 и 5.3).

Как видно из рис. №5.2 и 5.3, линии регрессии на полях корреляции расположены одинаково. Однако, точки поля на рис. №5.3, расположенные по эллипсу, теснее примыкают к линии регрессии, чем точки на рис. №5.2, расположенные по окружности. Рассеяние точек поля относительно линии регрессии объясняется влиянием прочих неучтенных факторов. Чем сильнее это влияние, тем больше отклонение точек поля от линии регрессии. При незначительном влиянии прочих факторов точки поля корреляции приближены к линии регрессии. Так, связь, показанная на рис. №5.3, является более тесной.

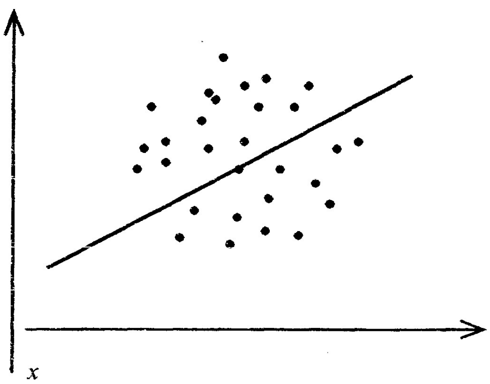


Рис. 5.2.

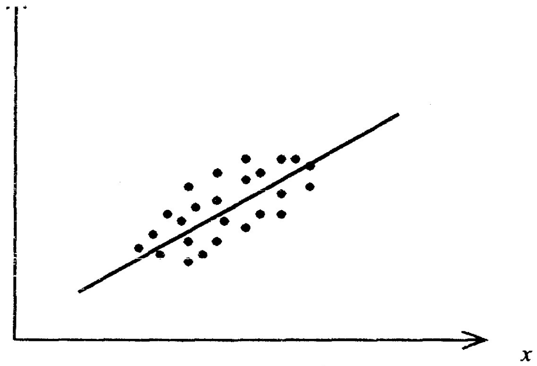


Рис. 5.3.

Если теснота связи велика, то, зная аргумент, по уравнению регрессии можно точно определить значение функции, т.е., воздействуя на аргумент, можно управлять функцией. Если теснота связи мала, то влияние на y обнаруживается лишь в среднем, т.е. управлять функцией нельзя.

Так, в нашем примере наблюдаются значительные колебания объемов выпускаемой продукции на разных предприятиях. Эти колебания объясняются различием предприятий по многим факторам (уровню организации труда, оснащенности производства, уровню использования вычислительной техники и т.д.)

Действие какого-либо фактора на функцию проявляется в изменчивости функции под влиянием изменчивости данного фактора. Поэтому показатель колеблемости (дисперсия) является мерой оценки влияния на функцию изменчивости различных факторов.

Чтобы выяснить степень влияния на функцию интересующего нас аргумента, нужно разложить дисперсию функции на две части, из которых одна будет вызвана влиянием данного аргумента ***x***, другая - влиянием прочих, действующих независимо от x, факторов. Это можно сделать исходя из теоремы о дисперсии объединенной статистической совокупности, аналитическое выражение кото-

$$\sigma^2{}_y = \overline{\sigma^2}_i + \overline{\delta^2}_i$$

рой имеет вид

где σ^2_y - общая дисперсия функции;

$\overline{\sigma_i^2}$ - внутригрупповая дисперсия функции, вызванная изменчивостью прочих факторов;

$\overline{\delta_i^2}$ - межгрупповая дисперсия функции, вызванная изменчивостью аргумента x. Поясним эти составляющие на нашем примере. Прежде всего проведем на нашем графике (см. рис. №5.1) горизонтальную линию, выражающую среднее значение $\overline{y}$ ($\overline{y} = \sum ym / \sum m = 162{,}8/25 = 6{,}51$).

Таким образом, на графике три линии:

- эмпирическая линия $\overline{y_i}$, построенная по фактическим данным;

- прямая наклонная линия $\overline{y}_x$ - теоретическое значение *y* при абстрагировании от влияния всех факторов, кроме фактора *x;*
- прямая горизонтальная линия $\overline{y}$ - среднее значение *y,* в которой исключено влияние всех без исключения факторов.

Поэтому $\overline{y}$ - постоянная средняя, а $\overline{y_i}$ и $\overline{y}_x$ - переменные средние, функционально изменяющиеся под влиянием изменения х. Отличие переменных средних от постоянной свидетельствует о силе влияния фактора х. Несовпадение линии $\overline{y}_x$ с линией у говорит о том, что связь между х и у неполная, не функциональная. Значит, чтобы измерить тесноту связи, т.е. определить, насколько она близка к функциональной связи, нужно знать значения составляющих теоремы разложения дисперсии. Это, прежде всего, внутригрупповая дисперсия, измеряющая отклонения точек поля корреляции у от эмпирической линии регрессии $\overline{y_i}$ и характеризующая остаточную вариацию, обусловленную прочими факторами:

$$\overline{\sigma}_э{}^2 = \frac{\sum (y_i - \overline{y}_i)^2 \cdot h_i}{\sum h_i}, \text{ где}$$

$\overline{\sigma}_э{}^2$ - средний квадрат отклонений точек поля корреляции от эмпирической линии регрессии. Также необходимо знать межгрупповую дисперсию, измеряющую отклонения эмпирической линии регрессии от линии общей средней и характеризующую вариацию, обусловленную фактором *x:*

$$\overline{\delta^2}_э = \frac{\sum (\overline{y}_i - \overline{y})^2 h_i}{\sum h_i}$$

где

$\overline{\delta}_э{}^2$ - средний квадрат отклонений эмпирической линии регрессии от линии общей средней. Общая дисперсия функции, таким образом, измеряет отклонения точек поля корреляции от линии общей средней и характеризует вариацию, обусловленную как фактором х, так и прочими факторами:

$$\overline{\sigma}_y{}^2 = \frac{\sum (y - \overline{y})^2 \cdot l_i}{\sum l_i}, \text{ где}$$

$\overline{\sigma}_y{}^2$ - средний квадрат отклонения точек поля корреляции от линии общей средней. В этом случае теорема разложения дисперсии запишется так:

$$\sigma_y{}^2 = \overline{\sigma_э{}^2} + \overline{\delta_э{}^2}$$

Используя упрощенный способ расчета и данные таблицы №5.1, определим межгрупповую дисперсию $\overline{\delta}_э{}'^2$ и общую дисперсию функции $\overline{\sigma}_y{}^2$ по следующей расчетной таблице и формулам:

Таблица №5.4.

$X_ц$	h	$\overline{y_i'}$	$\overline{y_i'}^2$	$\overline{y_i'}^2 \cdot h_i$
67,5	5	-2,4	5,76	28,8
8,25	4	-2	4	16
9,75	3	-0,33	0,1	0,3
11,25	5	0,2	0,04	0,20
12,75	8	0,37	1,137	1,095
Итого	25	-	-	46,39

$$\overline{y'} = \frac{\sum l_i \cdot y'}{\sum l_i} = \frac{-17}{25} = -0.64$$

$$\overline{\delta'_э{}^2} = \frac{\sum \overline{y'_i}^2 \cdot h_i}{\sum h_i} - \overline{y'}^2 = \frac{46.39}{25} - (-0.64)^2 = 1.395$$

$$\sigma'_y{}^2 = \frac{\sum l_i \cdot y'^2}{\sum l_i} - \overline{y'}^2 = \frac{63}{25} - (-0.64)^2 = 2.06$$

Рассчитаем теперь коэффициент детерминации $\eta^2_э$ и эмпирическое корреляционное отношение

$$\eta^2{}_э = \frac{\overline{\delta^{2}}'_э}{\delta^{2}{}'_y} = 1.395 / 2.06 = 0.677$$

$\eta_э$:

Коэффициент детерминации показывает, какая часть полной колеблемости функции *y* обусловлена изменчивостью аргумента.

Так, в нашем примере объем произведенной продукции на 67,7% определяет вариацию балан-

$$\eta_э = \sqrt{\frac{\delta'^2{}_э}{\sigma'^2{}_y}} = \sqrt{0.677} = 0.82$$

совой прибыли.

Эмпирическое корреляционное отношение является одним из показателей, характеризующих тесноту корреляционной зависимости, т.е. степень ее приближения к функциональной связи.

Эмпирическое корреляционное отношение изменяется от 0 до 1. Если связь отсутствует, то $\eta_э = 0$. В этом случае общая дисперсия функции объясняется влиянием только внутригрупповой дисперсии, а межгрупповой вариации не существует. Следовательно, признак-аргумент никак не влияет на образование общей вариации. Если связь функциональная, то $\eta_э = 1$. В этом случае общая дисперсия функции объясняется влиянием только межгрупповой вариации, т.е. признак-аргумент целиком определяет дисперсию признака-функции. Чем больше значение $\eta_э$ приближается к 1, тем сильнее связь между *x* и *y,* тем ближе зависимость к функциональной. Так, в нашем примере значение $\eta_э$= 0,82 показывает, что связь между объемом продукции и балансовой прибылью достаточно высокая.

В формуле $\sigma_y{}^2 = \overline{\sigma_э{}^2} + \overline{\delta_э{}^2}$ при ограниченном числе наблюдений допускается систематическая ошибка. Она возникает по той причине, что *y* (при ограниченном числе наблюдений) изменяется не только под влиянием *x,* но и под влиянием прочих факторов, не зависящих от *x.* Случайная колеблемость у, вызванная влиянием этих прочих факторов, преувеличивает $\overline{\delta_э{}^2}$ и одновременно преуменьшает $\overline{\sigma_э{}^2}$, так как в последней составляющей недоучтено влияние прочих факторов.

Для устранения математической ошибки при расчете показателей $\overline{\sigma_э{}^2}$ и $\overline{\delta_э{}^2}$ эмпирическую линию регрессии заменяют теоретической, т.е. вместо y_i используют y_x. Теорема разложения дисперсии в этом случае запишется следующим образом:

$$\sigma_y{}^2 = \overline{\sigma_T{}^2} + \overline{\delta_T{}^2}, \text{ где}$$

$\overline{\sigma_T{}^2}$ - средний квадрат отклонений точек поля корреляции от теоретической линии регрессии (измеряет влияние прочих факторов);

$$\overline{\sigma_T{}^2} = \frac{\sum \sigma_T{}^2 \cdot h_i}{\sum h_i} = \frac{\sum (y - \overline{y_x})^2 \cdot h_i}{\sum h_i}$$

$\overline{\delta_T^{\ 2}}$ - средний квадрат отклонений теоретической линии регрессии от линии общей средней (измеряет влияние фактора x).

$$\overline{\delta_T^{\ 2}} = \frac{\sum \delta_T^{\ 2} \cdot h_i}{\sum h_i} = \frac{\sum (\overline{y_x} - \overline{y})^2 \cdot h_i}{\sum h_i}$$

Вместо показателя $\eta_э$ можно использовать показатель η_T, называемый теоретическим корреляционным отношением y по x.

$$\eta_T = \sqrt{\frac{\overline{\delta_T^{\ 2}}}{\sigma_y^{\ 2}}}$$

С помощью показателя η_T в чистом виде оценивается степень приближения зависимости к функциональной.

Теоретическое корреляционное отношение также лежит в пределах от 0 до 1. Рассмотрим несколько случаев функциональной зависимости между y и x.

1. $\eta_T = 1$ - строгая функциональная зависимость между у и x (рис. №5.4):

$\sigma_y^{\ 2} = \overline{\delta_T^{\ 2}}$, $\overline{\sigma_T^{\ 2}} = 0$. Изменчивость у формируется за счет фактора х.

2. $\eta_T = 0$ - отсутствие корреляционной зависимости (рис. №5.5):

$\sigma_y^{\ 2} = \overline{\sigma_T^{\ 2}}$, $\overline{\delta_T^{\ 2}} = 0$. Изменчивость у формируется за счет прочих факторов.

3. $0 < \eta_T < 1$ - более или менее тесная корреляционная зависимость (рис. №5.6):

$\delta^2_T \neq 0$, $\sigma^2_T \neq 0$. Изменчивость y формируется как за счет x, так и за счет прочих факторов.

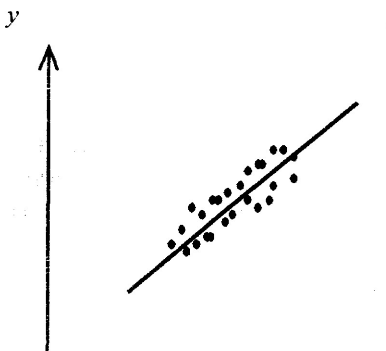


Рис. 5.4.

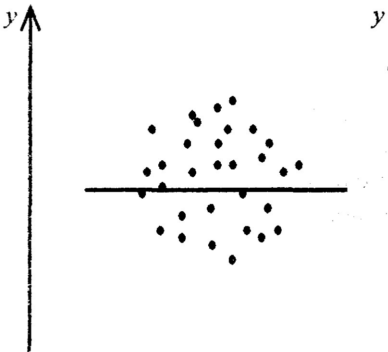


Рис. 5.5.

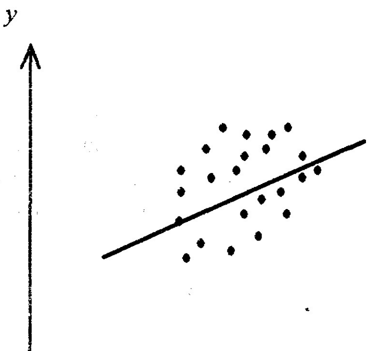


Рис. 5.6.

Рассчитаем значение теоретического корреляционного отношения для нашего примера, используя расчетную таблицу и полученное ранее значение σ'^2_y:

Таблица №5.5.

x'	*Хц*	*h*	y'_x	$\overline{y'_x} - \overline{y'}$	$(\overline{y'_x} - \overline{y'})^2$	$(\overline{y'_x} - \overline{y'})^2 * h_I$
-4	67.5	5	-2.45	-1.77	3.13	15.6
-3	8,25	4	-1,69	-1,01	1,02	4,08

-2	9,75	3	-0,93	-0,25	0,0625	0,18
-1	11,25	5	-0,17	0,51	0,26	1,30
0	12,75	8	0,59	1,27	1,61	12,8
Итого	-	25	-	-	-	33,96

$$\overline{\delta}^2{}_T = \frac{\sum(\overline{y}_x - \overline{y}')^2 h_i}{\sum h_i} = 33.96/25 = 1.36$$

$$\eta_T = \sqrt{\frac{\overline{\delta_T}^2}{\sigma_y^2}} = \sqrt{\frac{1.36}{2.06}} = \sqrt{0.66} = 0.81$$

Для качественной оценки тесноты связи на основе показателей корреляционных отношений можно воспользоваться таблицей, предложенной американским ученым Чэддоком:

Таблица №4.6.

η_T	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
Сила связи	Слабая	Умеренная	Заметная	Высокая	Очень высокая

Полученное значение $\eta_T = 0,81$ подтверждает высокую связь между объемом произведенной продукции и балансовой прибылью.

Как эмпирическое, так и теоретическое корреляционные отношения измеряют лишь тесноту связи, не показывая ее направление.

Коэффициент корреляции r*, с* помощью которого оценивается теснота связи только в ее линейной форме, показывает также и направление этой связи. Значения r изменяются от - 1 до + 1. Если коэффициент корреляции имеет знак плюс, связь прямая, если минус - связь обратная.

Другим достоинством коэффициента корреляции является его независимость от выбора единиц измерения признаков, в то время как значение коэффициента регрессии a_1 зависит от размерности признаков. Такую независимость коэффициенту корреляции придает стандартизированная система единиц σ_x, σ_y в которой он рассчитывается:

$r = a_1 * \sigma_x / \sigma_y$

С помощью коэффициента корреляции можно ответить на вопрос: на сколько стандартизированных единиц σ_y изменится функция *y,* если аргумент *x* изменится на одну стандартизированную единицу σ_x?

Для расчета r в нашем примере воспользуемся той его модификацией, которая позволяет сделать расчет по имеющимся данным:

$$r = \frac{n\sum(\overline{x}' * \sum my') - \sum hx' * \sum ly'}{\sqrt{n * \sum hx'^2 - (\sum hx')^2} * \sqrt{n\sum ly'^2 - (\sum ly')^2}} = \frac{25 * 73 - (-43) * (-17)}{\sqrt{25 * 133 - 1849} * \sqrt{25 * 63 - 289}} = 0.8$$

Близость коэффициента корреляции к единице ($r = 0,8$) также характеризует близость нашей зависимости к функциональной, а положительный знак указывает на то, что с ростом объема произведенной продукции балансовая прибыль в среднем возрастает.

6. АНАЛИЗ РЯДОВ ДИНАМИКИ

6.1. Основные понятия рядов динамики

Изучение изменений экономических явлений во времени является при рыночной системе хозяйствования одной из важнейших задач статистики. Эта задача решается при помощи составления и анализа рядов динамики важнейших экономических показателей. Результаты такого анализа способствуют последующему уточнению форм и методов производственной и хозяйственной деятельности.

Ряд динамики (или временной ряд) представляет собой ряд значений строго определённого статистического показателя, отражающий закономерность развития изучаемого явления во времени.

Числовые явления того или иного статистического показателя, составляющие динамический ряд, называют уровнями ряда.

Каждый временной ряд состоит из двух граф. В первой графе приводятся моменты или периоды времени, во второй - уровни ряда. При графическом изображении динамического ряда на оси абсцисс строится шкала времени, а на оси ординат - шкала уровней ряда.

Статистические показатели, приводимые в динамическом ряду могут быть абсолютными, средними или относительными величинами.

Динамический ряд, уровни которого характеризуют состояние явления на определённую дату, называется моментным динамическим рядом. Примером могут служить данные за ряд лет о численности населения страны на 1 января каждого года. Интервальным называется такой динамический ряд, уровни которого выражают размеры явления за определённый промежуток времени. Например, данные о производстве продукции предприятия за ряд лет. Особенностью интервальных рядов является возможность суммирования рядом стоящих уровней, получая новые уровни большей длительности. Данные моментных рядов суммировать нельзя, т.к. единицы, из которых слагаются уровни, последовательно повторяются в различных уровнях ряда.

Важнейшим требованием построения динамических рядов является сопоставимость статистических данных, входящих в состав ряда динамики. Только при соблюдении этого требования может быть обеспечена правильность тех выводов, которые являются результатом анализа ряда.

Соблюдение принципа сопоставимости ряда означает, что полезно лишь такое сопоставление, которое учитывает существо изучаемого явления и цель, для которой производится сравнение.

В качестве важнейших причин несопоставимости уровней ряда можно выделить следующие:
- изменение методологии учета или расчета показателей;
- изменение цен (для стоимостных показателей);
- различная продолжительность периодов, к которым относятся уровни;
- различные единицы измерения уровней;
- изменение территории, к которой отнесены те или иные показатели.

Могут быть и другие причины несопоставимости экономических показателей. Это требует, прежде чем анализировать динамический ряд с целью выявления закономерностей изменения, убедиться, исходя из цели исследования, в сопоставимости уровней ряда. Если последняя отсутствует, добиться её дополнительными расчетами, когда это возможно.

Важной задачей при изучении временных рядов экономических показателей является характеристика изменения уровней ряда от периода к периоду, от даты к дате. В результате сравнения уровней получается следующий набор абсолютных и относительных показателей: абсолютный прирост, темп роста, темп прироста, абсолютное значение одного процента прироста.

В результате сравнения уровней получается система абсолютных и относительных показателей динамики. При этом возможно два варианта сопоставления:

1. Каждый уровень динамического ряда сравнивается с предшествующим ему уровнем, т.е. база сравнений является переменной. Такие показатели называются цепными.

2. Каждый уровень динамического ряда сравнивается с одним и тем же уровнем, принятым за базу сравнения. Обычно это начальный уровень ряда. Такие показатели с постоянной базой сравнения называются базисными.

Цепные показатели динамики характеризуют интенсивность изменения уровня от периода к периоду или от даты к дате. Базисные показатели динамики характеризуют окончательный результат всех изменений в уровнях ряда за период от момента начала развития явления до данного момента.

Для того, чтобы дать характеристику общего изменения явления за рассматриваемый промежуток времени, используют средние показатели (средний уровень ряда, средний абсолютный прирост, средний темп роста, средний темп прироста). Кроме того, к числу задач анализа динамических рядов экономических показателей относятся также:

- выявление закономерностей динамики ряда в целом;
- краткосрочное прогнозирование тенденций развития экономического явления, представленного динамическим рядом;
- оценка степени взаимосвязи в изменениях уровней динамических рядов.

Решение перечисленных выше задач анализа динамических рядов экономических показателей позволит точно оценить тенденцию и интенсивность развития явления, что является в условиях рынка важным элементом принятия управленческих решении.

6.2. Показатели динамических рядов

Абсолютный прирост уровня ряда показывает, на сколько в абсолютном выражении уровень отчётного периода больше (меньше) уровня базисного периода. Может быть рассчитан с постоянной и переменной базой сравнения.

Цепной абсолютный прирост:

$$\Delta_ц = y_i - y_{i-1}$$

Базисный абсолютный прирост:

$$\Delta_Б = y_i - y_Б,\ где$$

y_i - любой уровень ряда, кроме 1-го (называется отчётным),
y_{i-1} - уровень ряда, предшествующего отчётному,
$y_Б$ - уровень периода, принятого за базу (называется базисным).

В качестве постоянной базы сравнения обычно принимается начальный уровень ряда. Абсолютный прирост, рассчитанный с переменной базой, называется скоростью роста. Абсолютный прирост может быть положительным, отрицательным, близким к нулю. Однако по величине абсолютного прироста нельзя судить об интенсивности динамики явления, т.к. он зависит от величины уровней (чем они больше, тем выше абсолютный прирост). Поэтому наряду с абсолютными показателями динамики вычисляются ещё относительные показатели роста и прироста уровней ряда.

Темп роста показывает, во сколько раз уровень отчетного периода больше (меньше) уровня базисного периода. Может быть рассчитан с переменной и постоянной базой сравнения. Цепной темп роста:

$$Тр_ц = y_i / y_{i-1}$$

Базисный темп роста:

$$Тр_Б = y_i / y_0$$

Темп роста характеризует изменение явления от срока к сроку, даёт непрерывную линию развития.

Темп прироста показывает на сколько процентов уровень отчетного периода больше (меньше) уровня базисного периода. Представляет собой отношение абсолютного прироста к предыдущему или базисному уровню.

Цепной темп прироста:

$$Тпр_ц = \Delta_ц * 100\% / y_{i-1}$$

Базисный темп прироста:

$$Тпр_Б = \Delta_Б * 100\% / y_0$$

или:

$$Тпр = \mathrm{Тр} - 100\%$$

Темп прироста может быть величиной положительной и отрицательной. Абсолютное значение 1 % прироста определяется как результат деления абсолютного прироста на соответствующий темп прироста, выраженный в %:

$A = \varDelta_ц / Тпр_ц$

Абсолютное значение 1% прироста показывает содержание 1% прироста, его весомость. При снижении *Тр* и *Тпр* абсолютное содержание 1% прироста может расти.

Не менее важное значение при анализе экономических показателей имеют средние годовые абсолютные приросты и средние темпы роста показателей.

Средний уровень динамического ряда рассчитывается различно для интервального и моментного рядов.

Для интервального ряда с равными интервалами времени:

$$\overline{y} = \frac{\sum y}{n}$$

Для интервального ряда с неравными интервалами времени:

$$\overline{y} = \frac{\sum yt}{\sum t}$$

Для моментного ряда:

$$\overline{y} = \frac{1/2 \cdot y_1 + y_2 + ... + y_{n-1} + 1/2 \cdot y_n}{n-1}$$

Средний абсолютный прирост определяется как средняя арифметическая из отдельных приростов, рассчитанных с переменной базой.

$$\overline{\Delta} = \frac{\sum \Delta y}{n-1}, \text{ или } \overline{\Delta} = \frac{y_n - y_o}{n-1}$$

Средний темп роста используется для характеристики средней относительной скорости равномерного изменения изучаемого явления.

$$\overline{T_P} = \sqrt[n-1]{Tp_1 \cdot Tp_2 \cdot \ldots \cdot Tp_{n-1}}$$

Средний темп прироста показывает, на сколько процентов увеличивался уровень ряда по сравнению с предыдущим в среднем за единицу времени, и характеризует среднюю интенсивность роста уровней ряда, т.е. среднюю относительную скорость увеличения уровней ряда.

В соответствии с теорией средних величин, расчет всех указанных динамических средних можно производить лишь для такого периода времени который характеризуется одинаковыми условиями

$$\overline{T_{Пр}} = \overline{T_p} * 100\% - 100\%$$

развития явления. Кроме того, общая средняя может быть дополнена средними показателями за отдельные промежутки этого периода.

6.3. Выравнивание динамических рядов и особенности статистического прогнозирования

Динамические ряды, уровни которых на протяжения длительного периода времени не изменяются, встречаются довольно редко. Наоборот, чаще уровни ряда со временем меняются, колеблются, но эта колеблемость для различных явлений неодинакова и может быть вызвана различными причинами. Колебания уровней могут вызваться действием каких-либо определяющих факторов, способствующих повышению или снижению показателей, влиянием сезонности, и, наконец, случайными причинами.

Таким образом, перед статистикой встает задача выявления основной закономерности развития явлений в отдельные отрезки времени, т.е. следует выявить общую тенденцию в изменении уровней ряда, освобождённую от действия случайных различных факторов. Для этой цели используют как элементарные приёмы обработки динамических рядов, так и более сложные с применением математической статистики.

К элементарным способам обработки ряда с целью выявления закономерностей изменения его уровней относятся: способ переменной средней (или способ укрупнённых интервалов); способ скользящей средней.

Обработка динамических рядов методом скользящей средней также, как и использование метода укрупненных интервалов рекомендуется, главным образом, для предварительного анализа исходных данных. Более совершенным способом обработки динамических рядов в целях выявления основной закономерности развития, с последующим использованием её для прогнозирования, является аналитическое выравнивание ряда. Суть этого метода состоит в том, что уровни ряда рассматриваются как функции времени, и задача выравнивания сводится к определению вида функции, отысканию её параметров по эмпирическим данным и расчету теоретических уровней по найденному уравнению.

Чаще всего при выравнивании используются следующие простейшие формулы, выражающие тенденции развития (тренд):

- прямая вида: $y_t = a_0 + a_1 \bullet t$, где y_t - теоретический уровень, t - время, a_0 , a_1 - параметры прямой;
- показательная функция: $y_t = a_0 * a_1^t$;
- парабола 2-го порядка: $y_t = a_0 + a_1 \bullet t + a_2 \bullet t^2$,
- гипербола: $y_t = a_0 + a_1 /t$.

При выборе типа функции чаще всего используют графическое изображение имеющихся эмпирических данных (линейную диаграмму).

На основе графического представления ряда динамики не всегда удаётся произвести однозначный выбор формы уравнения. Поэтому при наличии большого колебания в значениях уровней целесообразно воспользоваться графическим изображением сглаженных с помощью скользящих средних уровней. Однако всегда следует помнить, что основанием для выбора вида кривой должен служить, прежде всего, анализ сущности данного общественного явления.

При выборе вида кривой можно использовать также метод конечных разностей, который основан на свойствах различных кривых, применяемых при выравнивании. Обязательным условием применения этого метода является равенство интервалов между уровнями в динамическом ряду. Метод заключается в следующем:

1. Если общая тенденция выражается линейным уравнением $y_t = a_0 + a_1 \bullet t$, то первые разности (абсолютные приросты) постоянны: $\Delta'_i = y_i - y_{i-1}$, вторые разности нулевые: $\Delta''_i = y'_i - y'_{i-1}$. Здесь следует иметь в виду, что равенство Δ' постоянной величине и Δ'' нулю при вычислении их по эмпирическим данным рассматривается не для каждого отдельного случая, а в целом по всей совокупности уровней:

при t=0 $y_0 = a_0$; $\Delta'_1 = y_1 - y_0 = a_1$; $\Delta''_1 = \Delta'_2 - \Delta'_1 = 0$

при t=1 $y_1 = a_0 + a_1$; $\Delta'_2 = y_2 - y_1 = a_1$; $\Delta''_2 = \Delta'_3 - \Delta'_2 = 0$

при t=2 $y_2 = a_0 + 2a_1$; $\Delta'_3 = y_3 - y_2 = a_1$; $\Delta''_3 = \Delta'_4 - \Delta'_3 = 0$

и т.д.

2. Если общая тенденция будет выражена параболой 2-го порядка $y_t = a_0 + a_1 \bullet t + a_2 \bullet t^2$, то вторые разности постоянные, третьи разности нулевые:

при t=0 $y_0 = a_0$.

при t=1 $y_1 = a_0 + a_1 + a_2$; $\Delta'_1 = a_1 + a_2$; $\Delta''_1 = 2a_2$; $\Delta'''_1 = 0$

при t=2 $y_2 = a_0 + 2a_1 + 4a_2$; $\Delta'_2 = a_1 + 3a_2$; $\Delta''_2 = 2a_2$; $\Delta'''_2 = 0$

и т.д.

При выборе формы уравнения следует исходить и из объема имеющейся информации. Чем больше параметров содержит уравнение тренда, тем больше должно быть наблюдений при одной и той же степени надежности оценивания. Проиллюстрируем выравнивание динамического ряда по прямой, используя данные табл. №6.1.

Таблица №6.1.

Годы	Объём продукции производства, тыс. шт.	Скорость роста, тыс. шт.	t	t^2	y_t	$\overline{y_t}$	$y_i - \overline{y_t}$	$(y_i - \overline{y_t})^2$
1	2	3	4	5	6	7	8	9
1980	30,4	-	-4	16	-121.6	30.4	0	0
1981	31,8	1,4	-3	9.	-95.4	31.8	0	0
1982	33,1	1,3	-2	4	-66.2	33.2	-0.1	0.01
1983	34,7	1,6	-1	1	-34.7	34.6	+0.1	0.01
1984	36.0	1,3	0	0	0	36.0	0	0
1985	37,5	1,5	1	1	37.5	37.4	+0.1	0.01
1986	38,7	1,2	2	4	77.4	38.8	-0.1	0.01
1987	40,3	1,6	3	9	120.9	40.2	+0.1	0.01
1988	41,6	1,3	4	16	166.4	41.7	-0.1	0.01
Итого	324.1			60	84.3	324.1		0.06

Выравнивание по прямой $y_t = a_o + a_1 \cdot t$ даёт эффект в тех случаях, когда абсолютные приросты более или менее постоянны, т.е. когда уровни ряда изменяются приблизительно в арифметической прогрессии. Таким образом, рассчитанные в графе 3 скорости роста подтверждают правильность выбора уравнения прямой ($y_t = a_o + a_1 \cdot t$) с целью выявления основной закономерности изменения объема продукции по годам.

Параметры уравнения a_o *и* a_1 находятся путем решения системы нормальных уравнений, полученных по способу наименьших квадратов:

$$\begin{cases} \sum_{1}^{n} y_i = a_0 n + a_1 \sum_{1}^{n} t_i \\ \sum_{1}^{n} y_i t_i = a_0 \sum_{1}^{n} t_i + a_1 \sum_{1}^{n} t_i^2 \end{cases}$$

Вычислительный процесс при нахождении параметров искомой прямой может быть значительно упрощен (при сохранении полной идентичности конечных результатов), если отсчет времени производится так, чтобы сумма показателей времени изучаемого ряда динамики была равна нулю.

При нечётном числе уровней ряда серединная точка (год, месяц) принимается за 0; тогда предшествующие периоды обозначаются соответственно через -1, -2, -3 и пр., а следующие за средним периоды соответственно+1,+2,+3 и пр.

При четном числе уровней ряда два серединных момента времени принимаются за -1 и +1, а все остальные соответственно обозначаются через два интервала, а именно: -3, -5 и т.д., или +3, +5 и т.д. При этих условиях

$$\sum_{1}^{n} t_i = 0$$

и система нормальных уравнений будет иметь вид:

$$\begin{cases} \sum_{1}^{n} y_i = a_0 n \\ \sum_{1}^{n} y_i t_i = a_1 \sum_{1}^{n} t_i^2 \end{cases}$$

откуда:

$$a_0 = \frac{\sum_{1}^{n} y_i}{n}, \; a_1 = \frac{\sum_{1}^{n} y_i \cdot t_i}{\sum_{1}^{n} t_i^2}$$

Необходимые расчеты произведем в табл. №6.1. Используя результаты итоговой строки табл. №6.1 находим:

a_0=324.1/9=36.01 a_1=84.3/60=1.41

По исчисленным параметрам составляем уравнение прямой ряда динамики объема производства продукции:

$$\overline{y_t} = 36.01 + 1.41t$$

Параметры a_0 и a_1 имеют конкретное экономическое содержание: так, параметр a_0, равный 36,01 тыс. шт., представляет средний уровень данного ряда динамики, а параметр a_1, равный 1,41 тыс. шт. - средний абсолютный прирост.

По найденному уравнению рассчитаем уровни выровненного ряда динамики:

для 1980г. $\overline{y_{t=-4}}$ *= 36,01 + 1,41 • (-4) = 30,4 тыс.* шт.

для 1981г. $\overline{y_{t=-3}}$ *= 36,01 + 1,41 • (-3) = 31,8* тыс. шт.

и т.д. (см. табл. №6.1 графу 7).

Найденное аналитическое уравнение используется для краткосрочных прогнозов (экстраполяции уровней), которые можно получить путём продления в будущее тенденции, наблюдавшейся в прошлом. Экстраполируя при *t = 5,* находим уровень 1989г.:

$\overline{y_{t=5}}$ *= 36,01 + 1,41 • 5 = 43,1* тыс. шт.

Фактический объем производства (для наших данных) в 1989г. составил 56,3 тыс. шт. Как видно, есть расхождение между фактическим и экстраполируемым уровнем. Здесь следует заметить, что любому прогнозированию в виде экстраполяции ряда должно предшествовать тщательное изучение длительных рядов динамики, которое позволило бы определить тенденцию изменения, поскольку в действительности тенденция развития, в свою очередь, может изменяться, то данные, получаемые путем экстраполяции ряда надо рассматривать как вероятностные, как своего рода оценки. Исходя из этого, при составлении прогнозов используют не точечную, а интервальную оценку, определяя доверительные интервалы прогнозов. Для этого необходимо рассчитать среднюю квадратическую ошибку уравнения ($\overline{S}y_t$), характеризующую колеблемость фактических уровней вокруг выровненных.

Величина $\overline{S}y_t$ определяется по формуле:

$$\overline{S}\, y_t = \sqrt{\frac{\sum (y_i - \overline{y}_t)^2}{n - m}}$$

где y_i и $\overline{y_t}$ - соответственно фактические и выравненные значения уровней динамического ряда;

n — число уровней ряда;

m - количество параметров в уравнении тренда. Для нашего примера среднее квадратическое отклонение от тренда будет равно:

$$\overline{Sy_t} = \sqrt{\frac{0.06}{9-2}} = 0.093 тыс.шт$$

Расчеты $(y_i - \overline{y_t})$ и $(y_i - \overline{y_t})^2$ приведены в графах 8 и 9 табл. №6.1 Величина доверительного интервала определяется в общем виде таким образом:

$\overline{y_t} = \pm t_ф \cdot \overline{Sy_t}$, где $t_ф$ - значение *t* критерия Стьюдента.

Вероятность отклонений выборочной средней от генеральной в малых выборках определяется при помощи интеграла Стьюдента:

$$P\left[|t| > t_ф\right] = 2\left[1 - S(t_ф)\right]$$

тогда, задавшись определенной вероятностью, например, P = 0,99, и зная, что средняя квадратическая ошибка уравнения составила 0,093, находим $t_ф$:

$$2^*\text{-}\ S(t_ф)\text{-}\ 1=0{,}99;\ 2^*\ S(t_ф)=1{,}99;\ S(t_ф)=0{,}995$$

По таблице *S(t)* в распределении Стьюдента находим $t_ф=3,7$.

Тогда:

$$y_{прогноз} = \overline{y_t} \pm t_ф * \overline{Sy_t} = 43.1 \pm 3.7 * 0.93 = 43.1 \pm 0.34$$

$$42{,}76 \leq y_{прогноз} \leq 43{,}44$$

Теперь видно, что и наш экстраполируемый уровень, и фактический уровень, охватываются одним и тем же доверительным интервалом.

Следует отметить, что данный момент экстраполяции динамического ряда только тогда может дать оценки, близкие к действительности, если в эмпирическом ряду случайные колебания, выражающиеся в разности (y_i - $\overline{y_t}$) и измеряемые средним квадратическим отклонением, будут небольшими и если между случайными отклонениями отсутствует автокорреляция (зависимость между последовательными уровнями в рядах динамики).

При изучении развития явления во времени часто возникает необходимость оценивать степень взаимосвязи в изменениях уровней двух каких-то рядов различного содержания, но связанных между собой. Например, динамика энерговооруженности и фондовооруженности влияет на изменение уровня производительности. В таких случаях используется корреляционный анализ.

Однако, применение корреляционного анализа в рядах динамики связано с некоторыми трудностями. Во-первых, в динамических рядах имеет место наличие связи между смежными уровнями, т.е. каждый последующий уровень зависим от предыдущего, что затушевывает корреляционную связь. Такое влияние смежных уровней носит название автокорреляция.

Другим обстоятельством является наличие лага, т.е. смещение во времени изменения одного явления по сравнению с другим. Для устранения этого необходимо сдвинуть уровни одного ряда относительно другого на нужный промежуток времени.

При применении корреляционного анализа в динамических рядах решаются следующие задачи:

1) Измерение связи между изменением двух параллельных рядов разного содержания, но так или иначе связанных друг с другом.

2) Измерение связи последовательных уровней одного и того же динамического ряда.

В первом случае исчисляются коэффициенты корреляции и регрессии, во втором - автокорреляции и авторегрессии. В первом случае коэффициент корреляции исчисляется по непосредственным данным рядов динамики, когда эмпирические уровни одного ряда принимаются за аргумент, другого - за функцию. Во втором случае коэффициент автокорреляции рассчитывается по отклонениям фактических уровней от тренда, представляющего основную тенденцию развития ряда.

Ряды динамики, у которых зависимость одних уровней от предыдущих может быть выражена определенным уравнением регрессии, называются авторегрессионными. Поэтому при наличии больших случайных колебаний во временных рядах рекомендуется при прогнозировании основываться на экстраполяции авторегрессивной функции уровней ряда.

Сущность данного метода заключается в нахождении авторегрессивного уравнения динамического ряда. Для этого прежде всего необходимо выявить основную тенденцию динамического ряда, после чего можно получить случайные отклонения эмпирических (фактических) значений от теоретических:

$$\Delta y_i = y_i - \overline{y_t}$$

Так как динамический ряд можно считать случайным, то для прогноза можно воспользоваться моделью марковского случайного процесса 1-го порядка. Она будет иметь следующий вид:

где $r_{i,\ i-1}$ - коэффициент автокорреляции 1 -го порядка. Коэффициент автокорреляции 1-го по-

$$\Delta y_i = r_{i,i-1} * \Delta y_{i-1}$$

рядка есть коэффициент корреляции между соседними уровнями. Он определяется по формуле:

$$r_{i,i-1} = \frac{\sum_{1}^{n} (\Delta y_i * \Delta y_{i-1})}{(n-1)\sigma^2{}_{\Delta y}}$$

где n - число уровней ряда;

$\sigma_{\Delta y}$ - среднее квадратическое отклонение;

Δy_i - значение ряда Δy.

Если значение коэффициента корреляции по модулю больше 0,5, то мы можем с уверенностью прогнозировать Δy_i, на один шаг. Спрогнозировав величину Δy_i, можем по ней спрогнозировать Δy_{i+1} при помощи этой же модели. Однако этот прогноз будет уже менее точен, т.к. $|r_{i,i-1}| < 1$ и, следовательно функция $y_i(\Delta y_i - 1)$ будет иметь затухающий экспоненциальный или синусоидальный вид. Получив значение Δy_i, и вычислив значение y_t ,можем получить окончательный прогноз:

$$y_i = y_i + \Delta y_i$$

В том случае, когда модель 1-го порядка не удовлетворяет нас (т.к. прогнозировать необходимо не на один шаг, а на несколько шагов) мы можем воспользоваться авторегрессивной моделью *i-го* порядка:

$$\Delta y_i = a_1 \Delta y_{i-1} + a_2 \Delta y_{i-2} + \ldots + a_m \Delta y_{i-m},$$

где Δy_i - прогнозируемое решение;

$\Delta y_{i-1} \ldots \Delta y_{i-m}$ - значение ряда Δy ;

$a_1, a_2 \ldots a_m$ - набор параметров авторегрессионного уравнения.

$a_1, a_2 \ldots a_m$ определяется следующим образом: строится матрица коэффициентов корреляции:

$$Д = \begin{Vmatrix} 1 r_1; r_2 \ldots\ldots\ldots\ldots\ldots\ldots r_m \\ r_1; 1 r_1; r_2 \ldots\ldots\ldots\ldots\ldots r_{m-1} \\ r_2; r_1; 1 r_1; r_2 \ldots\ldots\ldots\ldots r_{m-2} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_m; r_{m-1} \ldots\ldots\ldots\ldots\ldots\ldots r_1 1 \end{Vmatrix}$$

где r_{ko} - автокорреляционный коэффициент между значениями Δy_i и Δy_i, взятым с шагом k (т.е. Δy_{i-k})

Автокорреляционный, коэффициент вычисляется по следующей формуле:

$$r_k = \frac{\sum_{1}^{n-k} (\Delta y_i * \Delta y_{i-k})}{(n-k-1)\sigma^2{}_{\Delta y}}$$

Из матрицы получим значение коэффициента a_k:

$$a_k = -r_k \frac{Д^{-1}{}_k}{Д^{-1}{}_1}$$

где $Д^{1}{}_{к}$ - алгебраическое дополнение элемента r-ого, стоящего в 1-й строке;

$Д^{1}{}_{1}$ - алгебраическое дополнение первого элемента первой строки матрицы. Получив все значения *а.к* и подставив их в уравнение, мы можем получить *m* значений динамического ряда Δy_i с необходимой точностью. Вычислив по формуле тренда *m* значений y_t, получим прогноз на *m* шагов:

$$y_i = \overline{y}_t + \Delta y_i$$

Рассмотрим еще один метод прогнозирования уровней в рядах динамики. Это метод нахождения регрессионного уравнения связи двух динамических рядов. Учитывая, что между изменениями нескольких показателей существует зависимость, можно экстраполировать один ряд динамики на основе имеющихся сведений об изменении второго ряда, связанного с ним. Так, например, определив зависимость между изменением объема капитальных вложений и объемом выпускаемой продукции в той или иной отрасли, можно экстраполировать данные о производстве продукции на основе данных о намечаемых капиталовложениях.

Рассмотрим данную методику прогнозирования в общем виде. Допустим, у нас имеются два динамических ряда x и *y,* которые зависят друг от друга. Причем, эта связь прямая и сдвинута во времени, т.е. рост значения X_i вызовет рост ряда *y* только в *i+1* или *i+2* периоде. Практически это означает, что если в 1988г. значение x_i увеличивалось, то значение y_i , должно соответственно увеличиваться в 1989г.

Этот сдвиг во времени необходимо устанавливать в зависимости от характера связи этих двух динамических рядов. Установив характер этой связи, мы должны соответствующим образом расположить эти два ряда.

Так, например, при сдвиге на один шаг значение первого уровня динамического ряда будет соответствовать значению второго уровня ряда. Естественно, при этом отрезки времени, за которые даются уровни, должны быть одинаковые и в том, и в другом ряду (это может быть год, месяц, декада и пр.). Теперь можно приступить к нахождению уравнения связи.

Допустим, связь между признаками *x и y* у нас выразится уравнением прямой:

$$\overline{y}_x = a_0 + a_1 x$$

где a_o, a_1 - параметры регрессионного уравнения.

Параметры a_o и a_1 для искомой прямой находятся путем решения системы нормальных уравнений, полученных по способу наименьших квадратов:

$$\begin{cases} \sum\limits_1^n y_i = a_0 n + a_1 \sum\limits_1^n x_i \\ \sum\limits_1^n y_i x_i = a_0 \sum\limits_1^n x_i + a_1 \sum\limits_1^n x_i^2 \end{cases}$$

где *n* - количество уровней ряда.

В случае обратной зависимости соответствующих значений рядов *x* и *y* необходимо выбрать следующий вид регрессионного уравнения связи:

$$\overline{y}_i = a_0 + \frac{a_1}{x}$$

Решая систему нормальных уравнений:

$$\begin{cases} \sum\limits_1^n y_i = a_0 n + a_1 \sum\limits_1^n \frac{1}{x_i} \\ \sum\limits_1^n \frac{y_i}{x_i} = a_0 \sum\limits_1^n \frac{1}{x_i} + a_1 \sum\limits_1^n \frac{1}{x_i^2} \end{cases}$$

получим параметры a_o и a_1 регрессионного уравнения обратной связи рядов *x* и *y* с соответствующим сдвигом.

При выборе функции надо опираться на характер взаимосвязи рядов х и *y,* для чего необходим теоретический анализ обоих признаков, а также на вид кривой *y* от *x*.

7. ИНДЕКСНЫЙ МЕТОД АНАЛИЗА

Важное значение в экономическом анализе имеет индексный метод. Полученные на основе этого метода показатели используются для изучения развития народного хозяйства в целом и его отраслей, анализа результатов производственно-хозяйственной деятельности предприятий, исследования роли отдельных факторов в формировании экономических показателей, выявления влияния структурных сдвигов. Индексы используются в международных сопоставлениях экономических показателей.

Индекс представляет собой относительную величину, получаемую в результате сравнения сложных совокупностей и отдельных их единиц по одному признаку. Результат индексного отношения может выражаться в виде коэффициента или в процентах.

Под сложной понимается такая статистическая совокупность, отдельные единицы которой непосредственно не подлежат суммированию. Например, промышленные предприятия выпускают, как правило, различные виды продукции, разнородные по своей потребительской стоимости. Поэтому получить общий объем продукции предприятия нельзя суммированием количества различных видов в натуральном выражении. Не подлежат непосредственному суммированию и данные о производстве и реализации производственных товаров.

Основой индексного метода при определении изменений в производстве и обращении продукции является переход от натурально-вещественной формы выражения к денежным измерителям. Именно посредством денежного выражения устраняется несравнимость товаров как потребительских стоимостей и достигается единство.

Для удобства в теории статистики разработана определенная символика. Каждая величина, изменение которой нас интересует (индексируемая величина), имеет свое обозначение. Так, количество единиц данного вида продукции (физический объем) обозначается *q,* цена единицы продукции -*p* и т.д. По степени охвата единиц совокупности различают индивидуальные и сводные индексы. Индивидуальные индексы характеризуют изменение только одной единицы совокупности. Например, изменение выпуска определенной марки самолета. Сводные индексы выражают общие результаты изменения всех единиц, образующих совокупность. Если индекс исчисляется для части единиц совокупности его называют групповым индексом. Индивидуальные индексы обозначают *i,* а сводные - I.

Таблица №7.1.

Виды продукции	Количество продукции, шт.		Оптовая цена за единицу, тыс.руб.		Расчетные показатели				
	Базисный период, q_o	Текущий период q_1	Базисный период P_o	Текущий период P_1	i_q	i_p	Стоимость, тыс.руб. Базисный период, q_o*P_o	Текущий период, q_1*P_1	Условная стоимость q_1*P_o
1	2	3	4	5	6	7	8	9	10
А	60	75	8,0	10,0	1,250	1,250	480,0	750,0	600,0
Б	100	100	6,5	6,5	1,000	1,000	650,0	650,0	650,0
В	150	130	4,0	3,8	0,867	0,950	600,0	494,0	520,0
Итого:							1730,0	1894,0	1770,0

Для характеристики периода, в котором берется индексируемый показатель, используются подстрочные знаки. Так, "O" обозначает базисный период, а "1" - текущий. Например, q_1 и q_0 - физический объем продукции в текущем и базисном периодах. Подстрочные знаки используются также для характеристики индексируемой величины. Например, i_q - индивидуальный индекс физического объема продукции, I_p -сводный индекс цен.

Рассмотрим порядок вычисления индивидуальных и сводных индексов на условном примере. Имеются данные (табл. №7.1) о выпуске продукции машиностроительным предприятием и ценах на них за два периода.

Индивидуальные индексы физического объема продукции:

$i_q = q_1 / q_0$ *(7.1.)*

где q_1 ,q_0- количество продукции соответственно в текущем и базисном периодах.

Для нашего примера эти индексы составят по продукции А, Б и В:

$$i^A{}_q = \frac{q^A{}_1}{q^A{}_0} = 75/60 = 1.250 \text{или} 125\%$$

$$i^Б{}_q = \frac{q^Б{}_1}{q^Б{}_0} = 100/100 = 1.000 \text{или} 100\%$$

$$i^В{}_q = \frac{q^В{}_1}{q^В{}_0} = 130/150 = 0.867 \text{или} 86.7\%$$

Вычисленные индивидуальные индексы показывают, что выпуск продукции А в текущем периоде возрос в 1,25 раза (рост на 25%), продукции Б не изменился, а продукции В - снизился на 13,3%. Индивидуальные индексы цен

$i_p = p_1 / p_0$ *(7.2)*

где p_1 *и* p_0 оптовые цены за единицу продукции в текущем и базисном периодах.

Результаты расчетов этих индексов (гр.7, табл. №7.1) показывают, что в текущем периоде цена на продукцию А увеличилась на 25%, на продукцию Б не изменилась, а на продукцию В снизилась на 5%.

При построении сводного индекса стоимости продукции необходимо сопоставить стоимость всей выпущенной предприятием продукции в текущем и базисном периодах. Для этого необходимо вычислить сумму произведений физического объема каждого вида продукции на цены.

Стоимость продукции базисного периода

$q^A{}_0 p^A{}_0 + q^Б{}_0 p^Б{}_0 + q^В{}_0 p^В{}_0 = \Sigma q_0 p_0$ *(7. 3)*

Стоимость продукции текущего периода

$q^A{}_1 p^A{}_1 + q^Б{}_1 p^Б{}_1 + q^В{}_1 p^В{}_1 = \Sigma q_1 p_1$ *(7.4)*

Сводный индекс стоимости продукции (итоги гр.9:гр.8, табл. №7.1)

$I_{qp} = \Sigma q_1 p_1 / \Sigma q_0 p_0 = 1894/1730 = 1.095$ *или 109,5%* *(7.5)*

т.е. по данной номенклатуре продукции (А, Б, В) в целом прирост стоимости в текущем периоде составил 9,5%.

При сравнении числителя и знаменателя формулы (7.5) в разности определяется показатель абсолютного прироста (уменьшения) стоимости *(Δ qp)* за счет действия двух факторов - изменения физического объема *(Δ q)* и цен на продукцию *(Δ p)* в текущем периоде по сравнению с базисным.

$\Delta qp = \Sigma q_1 p_1 - \Sigma q_0 p_0 = 1894{,}0 - 1730{,}0 = 164{,}0$ *тыс. руб.* *(7.6)*

Стоимость продукции в текущем периоде увеличилась на 164,0 тыс. руб., что соответствует относительному ее росту на 9,5%. Для определения количественного влияния физического объема продукции на изменение стоимости необходимо элиминировать (устранить) влияние цен. Это достигается путем построения сводного индекса физического объема продукции (итоги гр.10:гр.8, табл. №7.1)

$I_q = \Sigma q_1 p_0 / \Sigma q_0 p_0 = 1770/1730 = 1{,}023$ *или 102,3%* *(7.7)*

Где $\Sigma q_1 p_0$ - стоимость продукции в текущем периоде по ценам базисного периода (условная стоимость).

Разность между числителем и знаменателем формулы (7.7) характеризует абсолютный прирост (уменьшение) стоимости продукции в текущем периоде, обусловленный изменениями физического объема *(Δq).*

$\Delta q = \Sigma q_1p_0 - \Sigma q_0p_1 = 1770{,}0 - 1730{,}0 = 40{,}0$ *тыс. руб.* *(7.8)*

т.е. в результате роста физического объема продукции в среднем на 2,3% стоимость в сопоставимых ценах возросла на 40,0 тыс. руб.

Для определения количественного влияния цен на стоимость продукции рассчитывают сводный индекс цен (итоги гр.9:гр.10 табл. №7.1).

$I_p = \Sigma q_1p_1/\Sigma q_1p_0$=1894/1770= 1,070 *или 107,0%* *(7.9)*

Абсолютный прирост (уменьшение) стоимости за счет изменения цен в текущем периоде (Δp).

$\Delta p = \Sigma q_1p_1 - \Sigma q_1p_0 == 1894{,}0 - 1770{,}0 = 124{,}0$ *тыс.руб.* *(7.10)*

т. е. цены в текущем периоде в среднем возросли на 7%, что привело к абсолютному росту стоимости на 124,0 тыс. руб.

Индексы стоимости, физического объема продукции и цен взаимосвязаны:

$I_{qp} = I_q * I_p$ *(7.11)*

Значение формулы (7.11) состоит в том, что на ее основе можно определить влияние отдельных факторов на изменение стоимости продукции. Так, если в текущем периоде стоимость возросла по сравнению с базисным периодом на 14% (I_{qp} = 1,14), а цены на продукцию снижены в среднем на 4% (I_p = 0.96), то не имея никакой дополнительной информации можно рассчитать, что стоимость продукции в сопоставимых ценах увеличится на 18,7 % (1,14: 0.96 =1,187 или 118,7 %).

Индекс цен (7.9) и физического объема продукции (7.7) представляют собой агрегатную форму сводных индексов.

Часто в практических расчетах для определения сводных обобщающих показателей изменения цен (физического объема продукции) используется средняя гармоническая форма сводного индекса.

Средний гармонический индекс получается путем преобразования агрегатного индекса. Суть этого преобразования заключается в том, что на основе формулы (7.2) в значение Σp_0q_1 (7.9) вместо p_0 подставляется p_1: $i_p = p_0$

$\Sigma p_0q_1 = \Sigma(p_1/i_p)*q_1 = \Sigma(p_1 * q_1/i_p)$ *(7.12)*

тогда сводный индекс цен в среднегармонической форме:

$$I_p = \frac{\sum p_1q_1}{\sum \frac{p_1q_1}{i_p}} \qquad (7.13)$$

Другой формой средних индексов является средний арифметический индекс. Для получения среднего арифметического индекса необходимо в значение Σp_1q_1 (7.9) подставить значение $p_0 * I_p = p_1$ (7.2):

$I_p = (\Sigma(i_p\, p_0q_1))/\Sigma p_0q_1$ *(7.14)*

Итак, любой сводный индекс в агрегатной форме может быть преобразован в средний арифметический и средний гармонический индексы.

Изучаемые в статистике показатели находятся между собой в определенной связи. Так, себестоимость единицы продукции зависит от количества произведенной продукции, от структурных сдвигов (изменение удельного веса в общем количестве от периода к периоду) и ряда других факторов. Для изучения взаимного влияния статистических показателей используется система взаимосвязанных индексов. Эта система позволяет выявить влияние факторов на динамику показателя. В качестве примера рассмотрим связь между изменением объема товарооборота продукта *N,* реализуемого через сеть магазинов, количеством продаж этого товара и уровнем их цен, данные по которым приведены в таблице №7.2.

Таблица №7.2

Магазин	Базисный период		Текущий период		Расчетные графы		
	Цена 1 кг, руб. P	Количество, кг q_o	Цена 1 кг, руб. P_1	Количество, кг q_1	i_p	Удельный вес реализации, %	
						Базисный период	Текущий период
1	1	3	4	5	6	7	8
1	40	100	39	400	0,98	20,0	40,0
2	32	200	31	250	0,97	40,0	30,0
3	38	200	37	350	0,97	40,0	30,0
Итого	Х	500	Х	1000	Х	100,0	100,0

Если анализировать изменение уровней цен (индивидуальных) в каждом магазине, то можно отметить, что они в текущем периоде по сравнению с базисным снизились соответственно на 2%, 3% и 3% (гр. 6 таблицы №7.2.).

В данном случае анализ производился без учета объемов реализации. Чтобы дать обобщающую характеристику процесса, необходимо соизмерять средний уровень цен по всей совокупности магазинов от периода к периоду. Для этого рассчитывают сводный индекс переменного состава:

$$\overline{I}_p = \frac{\overline{p}_1}{\overline{p}_0} = \frac{\sum q_1 p_0}{\sum q_1} : \frac{\sum q_0 p_0}{\sum q_0} \qquad (7.15)$$

где $\overline{p}_1$ и $\overline{p}_0$ - средняя цена продукта соответственно в текущем и базисном периодах.

Применительно к данным таблицы №7.2:

$$\overline{p}_1 = \frac{39*400 + 31*250 + 37*350}{400 + 250 + 350} = \frac{36300}{1000} = 36.3\,руб$$

$$\overline{p}_0 = \frac{40*100 + 32*200 + 38*200}{100 + 200 + 200} = \frac{18000}{500} = 36\,руб$$

Следовательно, $\overline{I_p}$ = 36,3 : 36,0 = 1,008 (100,8%), т.е. в среднем (по трем магазинам) цена продукта в текущем периоде возросла на 0,8%. Население при покупке каждого килограмма данного продукта переплачивало по 0,3 руб.(36,3 - 36,0).

Такое изменение цен произошло под влиянием как изменения индивидуальных цен, так и за счет влияния структурных сдвигов.

Для определения влияния изменения индивидуальных цен рассчитывают сводный индекс постоянного состава (I*). При этом элиминируют влияние структурных сдвигов, путем закрепления количества реализованной продукции на уровне текущего периода:

$$I^* = \frac{\sum q_1 p_1}{\sum q_1} : \frac{\sum q_1 p_0}{\sum q_1} = \frac{\sum q_1 p_1}{\sum q_1 p_0} \qquad (7.16)$$

В формуле (7.16)

$\overline{p_o}'$ - расчетная средняя цена реализации 1 кг в текущем периоде по ценам базисного периода, и для наших данных:

$$p_0' = \frac{40 \cdot 400 + 32 \cdot 250 + 38 \cdot 350}{400 + 250 + 350} = \frac{37300}{1000} = 37{,}3 \text{ руб.}$$

Следовательно, I* = 36,3 : 37,3 = 0,973 (97,3%), т.е. в отчетном периоде цены по трем магазинам в среднем снизились на 2,7% (97,3 - 100 = 2,7%). В абсолютном выражении это дало населению фактическую экономию при приобретении каждого килограмма данного продукта 1,0 руб.: (36,3 -37,3 = -1,0).

Для определения влияния структурных сдвигов на величину товарооборота необходимо элиминировать влияние индивидуальных цен путем их закрепления на уровне базисного периода.

Сводный индекс структурных сдвигов *($I_{стр}$)* будет иметь вид:

$$I_{стр} = \frac{\sum q_1 \cdot p_o}{\sum q_1} \div \frac{\sum q_o \cdot p_o}{\sum q_o} \quad (7.17)$$

Для нашего примера $I_{стр}$ = 37,3 : 36,0 = 1,036 (103,6%), т.е. структурные сдвиги привели к повышению цен в среднем на 3,6% и в абсолютном выражении это вызвало переплату населением на каждом килограмме приобретенной продукции 1,3 руб. (37,3 - 36,0).

Проведенный анализ показывает, что рост в текущем периоде средней цены продукта на 0,8% вызван, с одной стороны, снижением на 2,7% цен в отдельных магазинах и, с другой стороны, ростом на 3,6% в результате структурных сдвигов в объеме реализации.

В абсолютном выражении рост в текущем периоде средней цены 1 кг продукта на 0,3 руб. вызван увеличением на 1,3 руб. за счет структурных сдвигов и снижением в среднем на 1,0 руб. цен в отдельных магазинах (0,3 == 1,3 - 1,0)

Рассчитанные выше индексы взаимосвязаны:

$$\overline{I_p} = I^* \cdot I_{стр} \quad (7.18)$$

Для практических расчетов система (7.18) удобна тем, что с ее помощью по любым двум известным индексам можно рассчитать третий неизвестный индекс.

СОДЕРЖАНИЕ