ОСНОВЫ СТАТИСТИКИ

Конспект лекций

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования «Витебский государственный
технологический университет»

ОСНОВЫ СТАТИСТИКИ

Конспект лекций для слушателей ФПК экономических специальностей

Витебск
2013

УДК 31:658
ББК 60.6
О 75

Рецензенты :

главный экономист Главного статистического управления Витебской области Сысоева Т.А.;

декан ФПК и ПК, к.э.н., доцент Семенчукова И.Ю.

Рекомендовано к изданию редакционно-издательским советом УО «ВГТУ», протокол № 6 от 26 сентября 2013 г. .

О **Основы статистики** : конспект лекций / сост. : Т. В. Касаева. – Витебск : УО «ВГТУ», 2013. – 125 с.

ISBN 978-985-481-322-6

Пособие содержит методологию исчисления и методы статистического анализа показателей, используемых в оценке деятельности промышленных предприятий.

Пособие раскрывает все основные темы дисциплины в соответствии с учебной программой курса «Основы статистики».

Рекомендуется для слушателей ФПК экономических специальностей.

УДК 31:658
ББК 60.6

ISBN 978-985-481-322-6

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Основы статистики – одна из важнейших дисциплин в подготовке и переподготовке специалистов экономического профиля, так как статистическая грамотность – неотъемлемая составляющая экономического образования.

Специалисту любого звена и уровня управления необходимо понимать качественную природу и способы формирования количественных данных, характеризующих социально-экономические явления и процессы. Следовательно, он должен обладать знаниями и навыками в области статистики. Кроме того, он должен уметь использовать различные статистические методы анализа массовых явлений. Всему этому учит курс «Основы статистики».

В данном издании рассмотрены основные приемы и методы сбора, обработки и анализа статистических данных, позволяющие обобщить их и выявить особенности распределения единиц совокупностей по тем или иным признакам, определить средние значения отдельных показателей и взаимосвязи между ними, оценить результаты выборочных данных и т. д.

Основные теоретические положения в данном пособии сопровождаются конкретными примерами, что позволит слушателям успешно применять на практике полученные теоретические знания и навыки статистических методов исследования социально-экономических процессов.

Статистические данные в примерах в большинстве случаев носят условный характер.

Содержание пособия соответствует программе курса «Основы статистики» для слушателей переподготовки по экономическим специальностям.

1 ПРЕДМЕТ И МЕТОД СТАТИСТИКИ

1.1 Предмет статистики и её теоретические основы

Происхождение термина «Статистика» – связывают с латинскими «Status» – состояние, положение вещей (позже политическое состояние); «Stato» – государство.

В современном толковании термин «статистика» многозначен:

- во-первых, под статистикой понимают совокупность сведений, фактов о явлениях и процессах в жизни общества. Эти данные публикуются в специальных сборниках;
- во-вторых, статистика – это практическая деятельность людей по сбору, обработке и анализу массовых данных, относящихся к определенным сферам общественной жизни;
- в-третьих, под статистикой понимают науку, изучающую количественную сторону массовых общественных явлений и их закономерностей.

Развитие статистики как науки шло одновременно по двум направлениям:

Первое направление – «государствоведение» (описательная школа статистики) – появилось в Германии в XVII веке. Его основные представители: Герман Конринг, Готфрид Ахенваль, российские ученые Кирилов Иван Кириллович, Татищев Василий Никитич, Ломоносов Михаил Васильевич, Герман Карл Федорович и др. Представители «государствоведения» считали своей основной задачей описание государства, его достопримечательностей: территории, населения, климата, политического устройства, благосостояния граждан, вероисповедания и т. п. (изобретали системы описания государства).

Второе направление – «политическая арифметика» – развивалось одновременно с первым в Англии. Его основателями были Уильям Петти и Джон Граунт. Основная заслуга политических арифметиков:

- использование массовых данных для установления закономерностей;
- использование группировок, средних и относительных величин;
- исследование взаимосвязи явлений и т. д.

Несмотря на то, что государствоведение и политическая арифметика развивались каждая своим путем, предмет исследования у них был общий – общество (государство), а также происходящие в нем массовые явления и процессы.

Первым создателем теории статистики считался известный бельгийский ученый, математик по образованию Адольф Кетле (1796 – 1874), который дал определение ее как орудия социального познания.

Современная трактовка предмета статистики следует из определения статистики.

Статистика – общественная наука, изучающая количественную сторону качественно определенных массовых социально-экономических явлений и закономерностей их развития в конкретных условиях места и времени.

То есть предмет статистики – количественная сторона качественно определенных массовых социально-экономических явлений и закономерностей их развития в конкретных условиях места и времени. Это определение позволяет выделить следующие основные черты предмета статистики:

- статистика изучает массовые общественные явления (численность населения, объем промышленной продукции), то есть явления, состоящие из множества фактов, обладающих различными признаками;
- статистика изучает общественные социальные и экономические явления и дает количественное, цифровое освещение общественных явлений;
- статистика изучает количественную сторону общественных явлений в неразрывной связи с их качественной стороной (наблюдает процесс перехода количественных изменений в качественные);
- статистика изучает количественную сторону общественных явлений в конкретных условиях места и времени (численность населения по областям РБ в 2009 году; динамика занятости населения по секторам национальной экономики и т. д.), то есть характеризует явления и процессы в конкретных пространственных и временных границах;
- статистика изучает количественные связи между общественными явлениями с помощью специальной методологии (методов сбора, обработки и анализа данных).

Статистика считается многоотраслевой наукой, то есть включает в себя систему научных дисциплин: теория статистики; экономическая статистика (национальная экономика в целом); отраслевые статистики (промышленности, сельского хозяйства, транспорта т. д.); социальная статистика и её отрасли (статистика населения, образования, здравоохранения и т. д.).

Теория статистики разрабатывает общие методы исследования общественных явлений и категорийный аппарат, то есть является *методологической основой* всех отраслевых статистик.

Теоретической основой статистики как науки являются основные положения общественных наук, то есть социально-экономической теории. Они рассматривают законы развития социальных и экономических явлений, выясняют их природу, значение, формулируют категории и понятия.

Статистика, опираясь на знание положений экономической теории, анализирует формы проявления законов, оценивает размеры явлений, разрабатывает методы их изучения. Экономическая теория, опираясь на факты, представленные статистикой, формулирует закономерности, складывающиеся в развитии общества.

1.2 Основные категории и методы статистики

Основными категориями статистики являются:

1. Статистическая совокупность – это массовое явление в виде множества однокачественных единиц с отличающимися индивидуальными признаками.

Другими словами, статистическая совокупность – это совокупность социально-экономических объектов или явлений общественной жизни, объединенных качественной основой, но отличающихся друг от друга отдельными признаками.

2. Единица совокупности – первичный элемент статистической совокупности (при переписи населения – человек, при изучении успеваемости – студент). Она является носителем признака, подлежащего регистрации (то есть изучению).

3. Признак – это качественная особенность единицы совокупности. Другими словами, признак – это характерное свойство изучаемого явления.

Количественные признаки – выражаются числовыми характеристиками (возраст, средний балл успеваемости, объём выпускаемой продукции).

Атрибутивные (качественные) – не имеют количественного выражения, но отличаются смысловыми, содержательными понятиями (вид продукции, пол, профессия и т. д.). Частный случай атрибутивного признака – альтернативный. Альтернативный признак позволяет разделить исследуемую совокупность на 2 части с противоположными характеристиками (пол – мужчины и женщины; продукция – годная и брак и т. д.)

Прерывные (дискретные) могут выражаться только определёнными значениями без промежуточных между ними (в большинстве случаев целыми числами), например, год рождения, курс университета и т. д.

Непрерывные могут принимать любые значения в определённых границах (рост человека, процент выполнения норм выработки и т. д.).

Первичные признаки – могут быть измерены, взвешены и т. д. (оценка на экзамене у студента).

Вторичные (расчётные) – могут быть рассчитаны, но не могут быть измерены, взвешены и т. д. (средний балл успеваемости группы).

Моментные признаки характеризуют единицу совокупности на определённый момент времени (в экономике – на дату). Например, остатки сырья на складе на 01.09.2013.

Интервальные, характеризующие единицу совокупности (явление, процесс) за определённый период времени. Например, выпуск продукции за третий квартал 2013 года.

Существенные признаки неразрывно связаны с сущностью изучаемого явления, характеризуют его наиболее важные стороны. При изучении деятельности производственной организации – это например, объем продукции, стоимость основных средств, численность работников.

К числу *второстепенных* при этом могут быть отнесены адрес, название предприятия и т. д.

Многообразие признаков, характеризующих социально-экономические явления, приводит к необходимости их классификации. В наиболее общем виде эта классификация может быть представлена следующей схемой (рисунок 1.1).

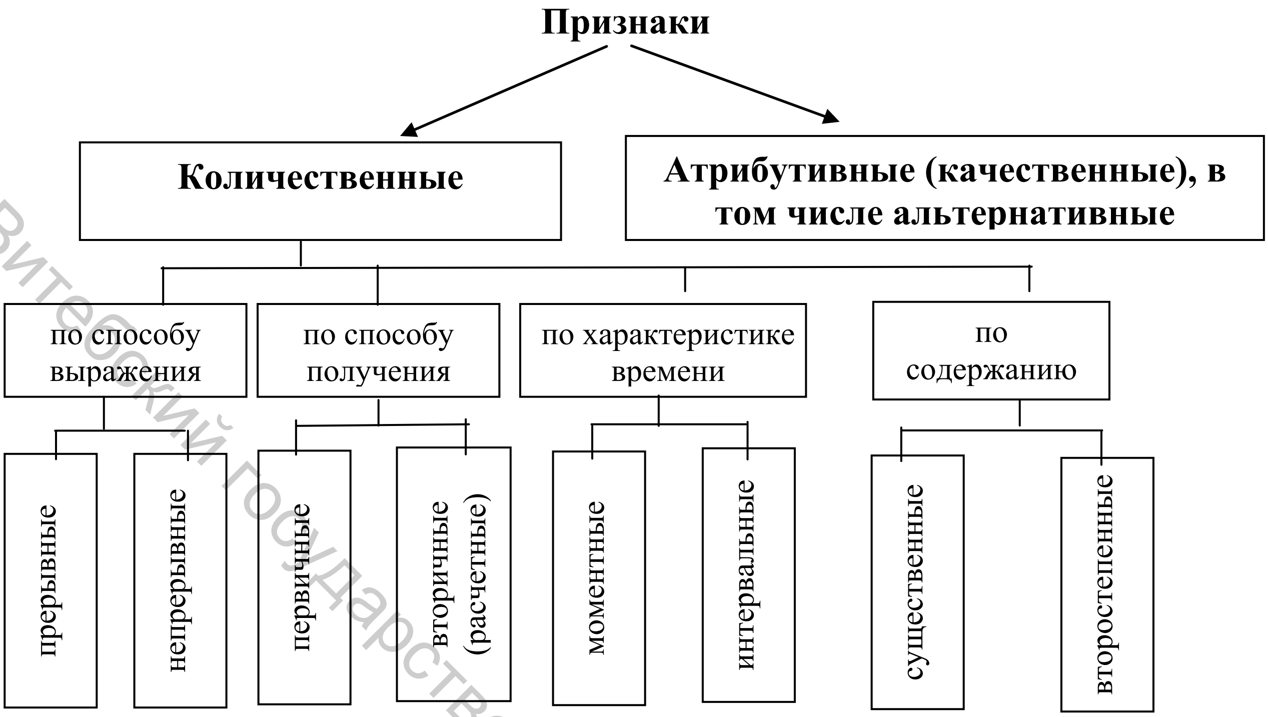

Рисунок 1.1 – Классификация признаков

4. Вариация – это изменение (колеблемость) признака при переходе от единой единицы наблюдения к другой.

Особенностью статистических исследований является тот факт, что изучаются только варьирующие признаки, то есть признаки, имеющие различные значения у отдельных единиц совокупности.

5. Статистический показатель – это количественная характеристика (размер) соотношения признаков общественных явлений. Статистические показатели могут быть (в зависимости от целевой функции):

- *учётно-оценочными (объёмными)*, которые характеризуют размеры изучаемых явлений (стоимость основных средств, численность населения и т. п.). Они, например, могут характеризовать уровень распространения в пространстве, уровни развития, достигнутые на определённый момент, и т. д.

- *аналитическими (расчётными)*, которые характеризуют особенности развития данного явления. Это могут быть: относительные величины, средние величины, показатели вариации, показатели динамики, показатели тесноты связи и др.

Основная задача статистики: определение содержания показателя и методологии его расчёта.

6. Система статистических показателей – это совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями. Система статистических показателей охватывает все стороны жизни общества (например, показатели СНС).

7. Статистическая закономерность – это закономерность, выявленная на основе массового наблюдения, то есть проявившаяся в большой массе явлений. Исследуя массу явлений, можно выявить и измерить закономерности.

Совокупность *методов*, используемых статистикой, должна рассматриваться со следующих позиций:

Статистика как наука использует общенаучные приёмы и методы (синтез, анализ, сравнение и т. д.).

Статистика как наука общественная опирается на диалектический метод познания, согласно которому все явления рассматриваются в развитии, во взаимной связи и причинной обусловленности.

При этом статистика использует такие категории диалектики, как количество и качество, необходимость и случайность, единичное и массовое, индивидуальное и общее и т. д.

Статистика как наука самостоятельная сформировала свою *статистическую методологию*.

Учитывая, что любое статистическое исследование состоит из 3-х этапов (стадий):

- статистическое наблюдение;
- статистическая сводка и обработка первичной информации;
- обобщение и анализ (интерпретация) статистической информации, – для каждого из этих этапов статистика использует свои специфические приёмы и методы соответственно:

– метод массового статистического наблюдения;
– метод сводки; методы группировок; графический метод; табличный метод.
– метод средних величин; методы оценки вариации; метод изучения динамики явлений; индексный метод; методы изучения взаимосвязей, оценки тесноты связей и т. д.

Вся совокупность этих приёмов и методов и образует статистическую методологию.

Так как выводы статистики основаны на большом числе единичных (случайных) явлений, она неизбежно связана с теорией вероятности, с математической статистикой. Так, например, статистические закономерности обнаруживаются благодаря действию закона больших чисел. В исследовании динамики и взаимосвязей в статистике широко используется метод наименьших квадратов и т. д.

2 СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

2.1 Понятие статистического наблюдения и его основные задачи

Статистическое наблюдение – это планомерный, научно организованный сбор данных о явлениях и процессах общественной жизни путём регистрации характеризующих их признаков.

Это определение включает две основные характеристики статистического наблюдения:

1) планомерность – это означает, что статистическое наблюдение заранее подготавливается и проводится по специально разработанному плану, который состоит из двух разделов:

1. Программно-методологические вопросы статистического наблюдения:
1 – определение цели статистического наблюдения;
2 – установление объекта наблюдения;
3 – выбор единицы наблюдения;
4 – разработка программы наблюдения;
5 – выбор системы формуляров.

2. Организационные вопросы статистического наблюдения:
1 – выбор времени статистического наблюдения;
2 – выбор места наблюдения;
3 – выбор формы наблюдения;
4 – выбор вида наблюдения;
5 – выбор способа наблюдения;
6 – определение органов, организующих и выполняющих наблюдение;

2) массовость статистического наблюдения, которая означает, что:
а) оно охватывает возможно большее число случаев данного явления;
б) в результате наблюдения мы хотим получить характеристики не единицы совокупности, а всей своей совокупности в целом.

Часто эти характеристики дополняются характеристикой систематичности, которая означает, что статистическое наблюдение проводится не от случая к случаю, а либо непрерывно, либо регулярно (через равные промежутки времени).

То есть статистическое наблюдение – планомерный, систематический, базирующийся на научной основе сбор данных о явлениях и процессах общественной жизни посредством регистрации их наиболее важных признаков в соответствии с программой наблюдения.

Основными задачами статистического наблюдения являются:

1) получение *достоверной* исходной информации;

2) обеспечение *полноты и сопоставимости* данных;

3) получение информации в возможно короткие сроки, то есть *своевременность*.

2.2 Формы, виды и способы статистического наблюдения

Проведение статистического наблюдения требует правильного выбора форм, видов и способов наблюдения, разновидности которых приведены на рисунке 2.1.

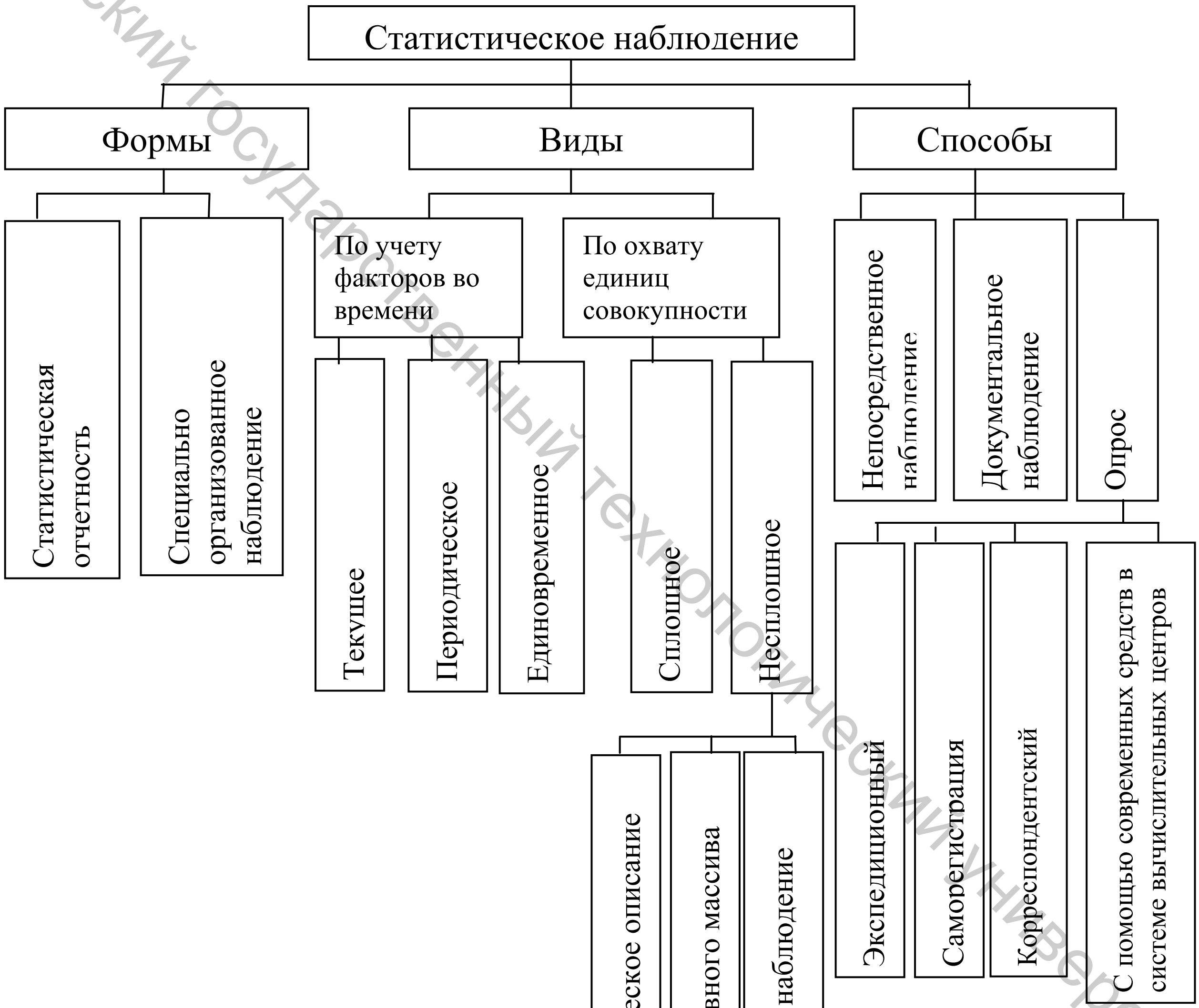

Рисунок 2.1 – Формы, виды и способы статистического наблюдения

Таким образом, статистическое наблюдение осуществляется в *двух формах*:

- путём предоставления отчётности;

- путём проведения специально организованных статистических обследований.

Статистическая отчётность – это основная форма статистического наблюдения. Она представляет собой совокупность статистических показателей, которые предоставляются всеми предприятиями, организациями, учреждениями в органы государственной статистики (и в свои вышестоящие организации) в строго определённые сроки и по строго установленным формам. Формы статистической отчётности разрабатываются и утверждаются Национальным статистическим комитетом.

Специально организованное статистическое обследование, как правило, применяется для тех объектов, которые не охвачены статистической отчётностью. Такие обследования проводятся специалистами (счётчиками), например, в виде:

- переписи;
- единовременного учёта;
- тематического статистического обследования.

Перепись – специально организованное статистическое обследование, характеризующее массовое явление или процесс на определённый момент (период) времени. Например, перепись населения, перепись учреждений и т. д.

Проведению переписей предшествует большая подготовительная работа: составление списков, разбивка административных районов на переписные участки, подготовка кадров и т. д.

Единовременный учёт – изучение численности и размещения изучаемого объекта (или его частей) по определённой территории на определённый момент (период) времени. Например, учёт установленного (работающего) оборудования, учёт численности студентов по факультетам.

Тематическое статистическое обследование, как правило, несёт выборочный характер и решает текущие задачи. Например, изучение семейных бюджетов.

Одним из основных вопросов организации статистического наблюдения является *выбор вида наблюдения*. Существует *классификация видов статистического наблюдения* по двум основным признакам:

1. *По полноте охвата единиц совокупности статистическое наблюдение* может быть:

- сплошное;
- несплошное.

Сплошное статистическое наблюдение имеет своей задачей обеспечение полного учёта единиц всей генеральной совокупности.

Несплошное статистическое наблюдение предполагает регистрацию части единиц генеральной совокупности и подразделяется на:

- монографическое описание;

- способ основного массива;
- выборочное наблюдение.

Монографическое описание (монографическое наблюдение) используется для подробного изучения единичных типичных объектов (или небольшого числа этих объектов). Основное правило – типичность исследуемого объекта.

Способ основного массива предполагает отбор наиболее крупных единиц наблюдения, в которых сосредоточена основная часть всех исследуемых фактов (например, обследуются только 15 % предприятий отрасли, которые производят 95 % продукции отрасли).

Выборочное наблюдение предполагает обследование отобранной в определённом порядке части единиц генеральной совокупности, а полученные характеристики распространяются на всю генеральную совокупность. Это наиболее распространённый вид несплошного статистического наблюдения, который широко используется в различных сферах: выборочный контроль качества продукции, выборочное обследование жилищных условий и т. д.

2. *По учёту фактов во времени* различают следующие виды статистического наблюдения: текущее, периодическое, единовременное.

Текущее (или непрерывное) статистическое наблюдение ведётся постоянно, непрерывно, по мере возникновения явлений (учёт рождаемости, учёт явок и неявок на работу и т. д.).

Периодическое статистическое наблюдение предполагает регистрацию исследуемых явлений через определённые, обычно одинаковые промежутки времени (остатки сырья на складе на 1 число каждого месяца).

Единовременное статистическое наблюдение проводится по мере надобности, без соблюдения строгой периодичности (время от времени). Например, перепись установленного оборудования и т. д.

С точки зрения *способа* регистрации фактов различают: непосредственное наблюдение, документальный способ, опрос.

При *непосредственном наблюдении* лица, проводящие статистическое наблюдение, получают необходимые сведения путём личного учёта единиц совокупности непосредственно на местах обследования: путём взвешивания, пересчёта, измерения, осмотра и т. д.

Документальный способ основан на использовании документов учётного характера. Проводится в виде систематических записей в первичных учётных документах, которые лежат в основе статистической отчётности.

Опрос основан на регистрации ответов, которые дают опрашиваемые лица. Этот способ получения статистической информации имеет следующие разновидности:

а) экспедиционный опрос;

б) саморегистрация;

в) корреспондентский опрос;

г) опрос с помощью современных компьютерных технологий (средств вычислительной техники).

Экспедиционный опрос состоит в том, что представитель статистических органов выезжает на место исследования и сам производит опрос и регистрирует ответы.

Саморегистрация состоит в том, что представитель статистических органов раздаёт бланки, инструктирует их заполнение и собирает заполненные бланки.

Корреспондентский опрос состоит в том, что статистическая организация рассылает бланки и инструкции по из заполнению. Получившие эти бланки лица заполняют их и высылают обратно в адрес статистической организации. Иногда создаётся постоянная сеть корреспондентов. Строго соблюдается признак добровольности.

В современных условиях компьютеризации есть возможность получения информации от корреспондентов по сети «Интернет».

2.3 План статистического наблюдения

С целью создания условий для получения объективно правильных материалов необходимо научно организовать статистическое наблюдение, для чего перед его проведением составляется план статистического наблюдения, структура которого приведена на рисунке 2.2.

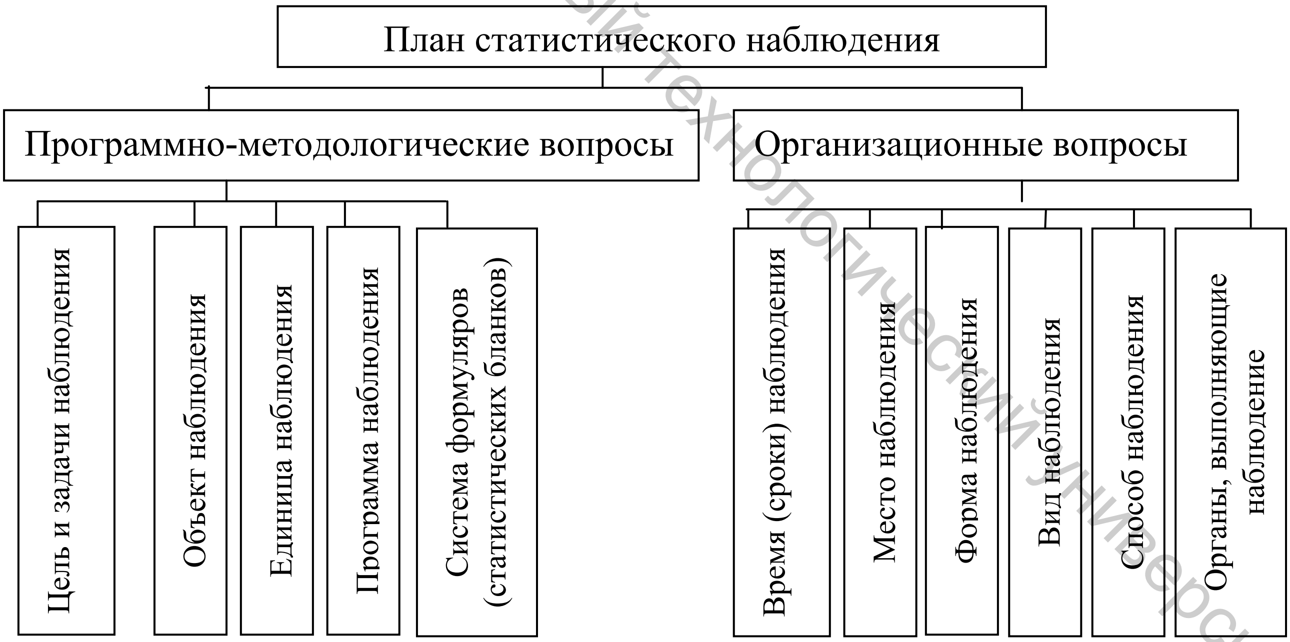

Рисунок 2.2 – Структура плана статистического наблюдения

Определить *цель* статистического наблюдения – значит дать чёткую формулировку задачи, стоящей перед статистическим исследованием.

Установление цели и задачи – исходный пункт при организации любого статистического наблюдения. Цель должна быть сформулирована ясно, чётко, а также развёрнуто, то есть с указанием задач, стоящих перед данным статистическим наблюдением.

Например, перепись населения.

Цель: определить численность и состав населения РБ и её регионов и выявить закономерности в их изменении.

Объект статистического наблюдения – это совокупность единиц изучаемого явления или процесса, которая подлежит статистическому наблюдению. Определить объект – не значит дать ссылку на изучаемое явление. Необходимо чётко определить состав и границы совокупности.

Определяя объект наблюдения, необходимо правильно указать единицу наблюдения.

Единица статистического наблюдения – это составной элемент изучаемой совокупности, который должен регистрироваться по определённым признакам в процессе статистического наблюдения.

При выборе единицы наблюдения необходимо чётко указать, каким единым существенным признаком должна обладать каждая единица, чтобы она могла попасть в изучаемую совокупность (например, при переписи – гражданство).

Необходимо различать понятия «единица наблюдения» и «отчётная единица».

Отчётная единица (учётная единица) – это источник сведений, та первичная ячейка, от которой должны поступать сведения о единицах наблюдения (это может быть предприятие, объединение, вуз и т. д.).

Программа статистического наблюдения – это перечень вопросов, на которые должны быть получены ответы по каждой единице наблюдения. Иными словами, это перечень признаков и показателей, подлежащих регистрации при проведении статистического наблюдения.

Требования к вопросам программы наблюдения были сформулированы в XIX веке бельгийским статистиком Адольфом Кетле:

1) программа статистического наблюдения должна включать только те вопросы, которые необходимы для решения поставленной цели;

2) в программу не следует включать вопросы, на которые возможно получить ответ неудовлетворительного качества;

3) в программу нельзя включать вопросы, которые могут расцениваться как вмешательство в личные вопросы опрашиваемых.

Современная теория статистики формулирует эти требования следующим образом:

1) формулировка вопросов должна быть чёткая, краткая и понятная;

2) формулировка вопроса должна быть такой, чтобы всеми быть понятной одинаково, чтобы можно было сравнивать ответы;

3) для цифровых объектов должны быть указаны единицы измерения.

К программе наблюдения прилагается *инструкция*, в которой даются пояснения, как следует понимать и отвечать на вопросы, даётся методика исчисления отдельных показателей.

Система формуляров – это специальные документы (бланки, карточки и т. д.), в которых регистрируются ответы на вопросы программы статистического наблюдения.

Существует 2 системы формуляров: списочная, индивидуальная.

При *списочной* системе в одном формуляре регистрируются ответы нескольких единиц наблюдения.

Её преимущества:

- проще автоматизировать (перенести информацию в ЭВМ);
- экономия бумаги;
- быстрая проверка результатов наблюдения.

При *индивидуальной* системе для каждой единицы наблюдения вводится свой формуляр.

Преимущество:

- можно включать большое количество признаков.

В отечественной государственной статистике наиболее распространена индивидуальная система формуляров.

Выбор *времени* проведения наблюдения – это период времени, к которому относятся полученные сведения. Например:

а) при переписи – зима 2009 – 2010 года;

б) показатели рентабельности в 2009 году (в I квартале 2009 года).

Срок проведения наблюдения – это время, в течение которого происходит заполнение формуляров: время начала и окончания сбора данных.

Критический момент (критическая дата) – это момент времени, по состоянию на который будет производиться регистрация явления. Например, перепись 1989 года (СССР): 12 часов ночи с 11 на 12 января (то есть переписи подлежали и те, кто умер после 12 часов ночи).

Выбор *места* проведения наблюдения имеет важное значение при изучении объектов, перемещающихся в пространстве (например, работа транспорта). Тогда речь идёт об установлении *пункта наблюдения*.

Но часто под местом проведения наблюдения понимаются *территориальные границы*.

Выбор *формы* статистического наблюдения: статистическая отчётность и специально организованное статистическое обследование.

Выбор *вида* статистического наблюдения:

- сплошное или несплошное;
- текущее, периодическое или единовременное.

Выбор *способа проведения* статистического наблюдения: непосредственное наблюдение, документальный способ, способ опроса.

Органы, выполняющие наблюдение:

- Национальный статистический комитет;
- территориальные статистические организации;
- учётно-экономические службы организаций, предприятий, учреждений;
- специально подготовленные люди (счётчики) и т. д.

2.4 Ошибки статистического наблюдения и методы контроля его результатов

Ошибки статистического наблюдения принято разбивать на 2 группы: ошибки регистрации, ошибки репрезентативности.

Ошибки регистрации присущи всем видам статистического наблюдения и возникают в результате неправильного установления фактов или неправильной их записи.

В теории статистики их принято делить на:

- случайные и систематические;
- преднамеренные и непреднамеренные.

Случайные ошибки (они непреднамеренные) возникают в результате описок, оговорок, низкой квалификации наблюдателей и т. д. (например, цифра записана не в ту графу). Они с одинаковой вероятностью могут давать искажение как в большую, так и в меньшую сторону.

Систематические ошибки (они могут быть преднамеренными и непреднамеренными) более опасны, так как они в значительной степени влияют на итоговые показатели. Они возникают за счёт округлений, за счёт неточности измерительных приборов, а иногда за счёт тяги к круглым цифрам (возраст 30 могут выбрать и 29-летние и 31-летние).

Преднамеренные ошибки искажают сведения в одном направлении: либо увеличивают, либо уменьшают. Их принято относить к разряду систематических. Возникают они в силу сознательного стремления лиц, дающих сведения, исказить истину.

Ошибки репрезентативности присущи только выборочным наблюдениям и возникают вследствие того, что наблюдению подвергается только часть единиц совокупности, которая не может полностью точно отобразить всю генеральную совокупность.

Для проверки правильности полученных в результате статистического наблюдения сведений используется *логический и арифметический контроль.*

Логический контроль состоит в сопоставлении ответов на взаимосвязанные вопросы программы.

Например:

1) ФИО Иванов И.И.;

2) пол мужской;

3) дата рождения ***01.01.2010***;

4) образование ***высшее.***

Арифметический (счётный) контроль имеет своей целью проверку правильности вычислений. Он сводится к проверке общих и групповых итогов, их сопоставлению.

Например: Численность студентов, чел.:

- дневная форма обучения 3700;
- заочная форма обучения 3300;

Всего ***8000.***

3 СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

3.1 Понятие статистической сводки и группировки. Виды статистических группировок

3.2 Выбор группированных признаков. Построение статистических группировок

3.3 Приемы вторичной группировки

3.4 Статистические ряды распределения, их графическое изображение

3.5 Табличный и графический метод представления статистической информации

3.1 Понятие статистической сводки, ее содержание. Виды статистических группировок

В результате статистического наблюдения (первого этапа статистического исследования) получают сведения о признаках каждой обследованной единицы статистической совокупности.

Дальнейшая задача статистики заключается в том, чтобы систематизировать эти материалы, дать сводную характеристику изучаемой совокупности. Поэтому вторым этапом статистического исследования является сводка и группировка информации, полученной в результате статистического наблюдения.

Основной задачей этого этапа исследования является обобщение и анализ первичных статистических данных для получения полной и всесторонней характеристики совокупности в целом и отдельных ее частей и представление информации в наиболее удобной для пользователя форме.

Статистическая сводка – это научно организованная обработка материалов наблюдения с целью характеристики изучаемой совокупности обобщающими показателями.

По глубине обработки материала сводка может быть простой или сложной.

Простая (итоговая) сводка проводится без распределения полученных сведений на группы, а предполагает подведение общего итога по изучаемой совокупности.

Сложная сводка предполагает предварительное распределение совокупности на группы и подсчет итогов по группам и в целом.

По технике или способу выполнения сводка может быть: ручная и механизированная (автоматизированная), то есть с помощью ЭВМ.

По форме обработки материала или по организации сводки она может быть централизованная и децентрализованная.

Централизованная – предполагает, что весь первичный материал поступает в одну организацию, где и подвергается обработке от начала до конца (например, в Национальном статистическом комитете).

Децентрализованная – предполагает поэтапную обработку материалов (например, сначала в облстатуправлениях, а затем итоги по области – в Национальный статистический комитет).

Проведение сводки включает в себя 3 этапа:

1) предварительный контроль материалов (то есть проверка исходных данных);

2) группировка данных по заданным признакам, определение производных показателей;

3) оформление результатов сводки в виде статистических таблиц, удобных для восприятия информации.

Для того, чтобы была достигнута цель исследования, сводка научно организуется (обосновывается), то есть разрабатывается ее программа и план.

Разработка программы сводки состоит из следующих этапов:

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

Программу сводки дополняет план, который содержит указание о:

- последовательности выполнения сводки;
- сроках выполнения отдельных частей сводки;
- исполнителях;
- порядке изложения и представления результатов.

При проведении сводки статистического материала отдельные единицы изучаемой совокупности объединяются в группы при помощи метода группировок.

Статистическая группировка – это процесс образования однородных групп на основе расчленения статистической совокупности на части или объединения изучаемых единиц в частные совокупности по существенным для них признакам. Каждая из этих групп характеризуется системой статистических показателей.

Метод статистических группировок позволяет:

1) рассчитать сводные показатели по группам;
2) сравнивать и анализировать причины различий между группами;
3) изучать взаимосвязи между признаками;
4) создать основу для последующей сводки и анализа данных.

Все это определяет роль группировок как научной основы сводки.

С помощью статистических группировок статистика решает важные и многообразные задачи. Среди них можно выделить три основные группы задач:

1) выделение качественно однородных экономических групп или типов общественных явлений из разнородной совокупности;

2) определение структуры и структурных сдвигов в совокупности однородных единиц, расчленение совокупности по величине варьирующего признака;

3) выявление и изучение связи и взаимообусловленности между явлениями.

В зависимости от цели (задачи) исследования все группировки делятся соответственно на три вида:

1) ***типологическая*** – решает задачу выделения качественно однородных экономических групп или типов общественных явлений из разнородной совокупности;

2) ***структурная*** – решает задачу определения структуры и структурных сдвигов в совокупности однородных единиц, расчленения совокупности по величине варьирующего признака;

3) ***аналитическая*** – решает задачу выявления и изучения связи и взаимообусловленности между явлениями. В аналитической группировке непременно есть признак-фактор (факторный признак) и признак-результат (результативный, результатный признак). При этом признак-фактор оказывает влияние на признак-результат. Группировка производится по признаку-фактору, а далее по каждой группе рассчитываются значения признака-результата.

Классификация группировок на типологические, структурные и аналитические часто оказывается весьма относительной. То есть группировка может быть *универсальной*: одновременно выделять типы явлений, указывать структуру и вскрывать взаимосвязи признаков (например, таблица 3.1).

Таблица 3.1 – Группировка предприятий города по размерам и объему выпускаемой продукции (данные условные)

Размер предприятий	Процент продукции, выпускаемой предприятиями по годам		
	200…г.	200…г.	200…г.
Крупные	50	40	30
Средние	30	35	38
Малые	20	25	32
Всего	100	100	100

Эта группировка:

1) характеризует структуру и структурные сдвиги;

2) выделяет типы предприятий;

3) указывает, что с уменьшением доли крупного бизнеса одновременно растет доля среднего и мелкого.

Второй признак классификации группировок – число группировочных признаков, положенных в основу группировки. В зависимости от количества группировочных признаков все группировки делятся на:

- простые (строятся по одному признаку);
- сложные (строятся по нескольким признакам).

В свою очередь, сложные группировки могут быть:

а) комбинационные (комбинированные),

б) многомерные.

Комбинационные строятся путем разбивки каждой группы на подгруппы в соответствии с дополнительными признаками.

Многомерная группировка осуществляется не последовательно по нескольким признакам, а одновременно. В этой группировке – суть новых подходов, новых принципов группировки, отличных от традиционных. Современная практика экономического анализа потребовала таких группировок, и наука предложила решение задачи с использованием кластер-анализа [16].

Можно указать и еще одну классификацию группировок в зависимости от источников информации:

- первичная – производится на основе исходных данных статистического наблюдения;
- вторичная – производится на основе уже имеющейся первичной.

3.2 Выбор группировочных признаков. Построение статистических группировок

Признаки, по которым производится распределение единиц совокупности на группы, называются группировочными признаками, или основанием группировки.

Первым и наиболее сложным вопросом теории группировок является правильный выбор этих признаков. При отборе группировочных признаков руководствуются следующими правилами:

1) необходимо брать типичные, существенные признаки, в соответствии с целями исследования;
2) необходимо учитывать конкретные условия места и времени: уместные в одном случае признаки могут оказаться неуместными в другом случае;
3) при изучении сложных явлений группировку следует производить по нескольким признакам.

В случае, когда группировка производится по атрибутивному признаку, количество групп равно количеству вариантов признака (например, по форме обучения – 3: дневная, вечерняя и заочная).

По альтернативному признаку образуется 2 группы с противоположными характеристиками (например, продукция: годная и брак; семейное положение: одинокие и семейные и т. д.).

При составлении группировок по количественным признакам необходимо определить количество групп и величину (ширину) интервала.

Количество образуемых групп зависит:

1) от числа единиц наблюдения;
2) от степени вариации группировочного признака;
3) от задачи исследования и особенностей изучаемого явления.

Число групп может быть задано на основе опыта предыдущих обследований. Если же вопрос приходится решать самостоятельно, то можно использовать формулу американского ученого Стерджесса:

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

где K – число групп (всегда целое число);

n – число единиц наблюдения.

Тогда при n = 10

K = 1 + 3,322 lg 10 = 4,322 (принимается 4).

Соответственно, можно рассчитать:

n	K
15 – 24	5
25 – 44	6
45 – 89	7
90 – 179	8
180 – 359	9
360 – 719	10

Эта формула пригодна при условиях:

а) распределение единиц совокупности по данному признаку приближается к нормальному;

б) интервалы образуются равные.

Интервал – это количественное значение, отделяющее одну группу от другой, то есть он очерчивает количественные границы групп.

В зависимости от характера распределения единиц совокупности по определенному признаку интервалы могут быть: равные и неравные.

Равные интервалы образуются в тех случаях, когда вариация признака проявляется в сравнительно узких границах и распределение является практически равномерным. Ширина интервала в этом случае определяется по формуле

$$\boldsymbol{i}_x = \frac{\boldsymbol{X}_{masx} - \boldsymbol{X}_{\min}}{\boldsymbol{K}}, \qquad (3.2)$$

где $X_{\max}, X_{\min}$ – соответственно максимальное и минимальное значение признака,

либо по формуле Стерджесса:

$$\boldsymbol{i}_x = \frac{\boldsymbol{X}_{\max} - \boldsymbol{X}_{\min}}{1 + 3{,}322 \lg \boldsymbol{n}}. \qquad (3.3)$$

Например, 10 рабочих характеризуются следующими показателями выполнения норм выработки (НВ):

X, % выполнения норм выработки: 96, 109, 98, 102, 105, 104, 100, 106, 112 ,103.

Ширина интервала группировки для n = 10 и $к$ = 4 определяется по формуле

$$i_x = \frac{112 - 96}{4} = 4.$$

Тогда группировка принимает вид (таблица 3.2):

Таблица 3.2 – Группировка рабочих по степени выполнения норм выработки

% выполнения НВ	Число рабочих
96 –100	2
100 –104	3
104 –108	3
108 –112	2
	n = 10

Неравные интервалы (как правило, прогрессивно возрастающие или прогрессивно убывающие) образуются в тех случаях, когда группировочный признак изменяется неравномерно или в больших пределах.

Например, прибыль организаций города.

X, млн. руб. 520, 3800, 157, 1900, 37850 и т.д.

Интервалы: до 100
100 – 1000
1000 – 10000
10000 – 50000
50000 и более

Образуемые интервалы могут быть: - закрытыми;
- открытыми.

Закрытыми называют интервалы, у которых указаны обе границы.

Открытыми называют интервалы с одной границей: верхней – у первого интервала; нижней – у последнего.

Если закрытым интервалам часто присуще свойство неопределенности при включении в группу тех значений, которые являются границами, то в открытых интервалах эта неопределенность убирается с помощью терминов:

«до»; «выше» или «свыше»; «более» и «менее».

После того, как выбран группировочный признак, определено количество групп, составлена группировка, необходимо установить перечень показателей, которые будут характеризовать группу. Эти показатели определяют в зависимости от цели исследования и задачи группировки.

3.3 Приемы вторичной группировки

Специфическим видом группировок является так называемая вторичная группировка.

Вторичная группировка – операция образования новых групп на основании уже имеющейся группировки.

Вторичная группировка используется для решения следующих задач:

1) образование качественно однородных групп (типов) на основе группировок по количественному признаку;

2) приведение к единому, то есть сопоставимому, виду группировок с различными интервалами;

3) образование укрупненных групп, в которых более четко (более ясно) проявляется характер распределения.

Образование новых групп на основании уже имеющихся возможно двумя способами перегруппировки:

а) изменением (объединением) первоначальных интервалов (метод укрупнения интервалов);

б) долевой перегруппировкой (на основе пропорционального дробления групп).

Первый способ применяется при переходе от мелких к более крупным интервалам, когда границы новых и старых интервалов совпадают.

Например, имеется информация о заработной плате работников двух предприятий отрасли (таблицы 3.3 и 3.4).

Таблица 3.3 – Группировка работников предприятия № 1 по уровню заработной платы

№ гр.	Заработная плата, тыс. руб.	Число работников, % к итогу
1	1800 – 2400	3
2	2400 – 3000	8
3	3000 – 3600	14
4	3600 – 4500	25
5	4500 – 6000	30
6	6000 – 7500	14
7	7500 – 9000	6
		100

Таблица 3.4 – Группировка работников предприятия № 2 по уровню заработной платы

№ гр.	Заработная плата, тыс. руб.	Число работников, % к итогу
1	1200 – 2100	2
2	2100 – 2400	5
3	2400 – 3000	6
4	3000 – 3750	14
5	3750 – 4500	23
6	4500 – 5100	22
7	5100 – 6300	12
8	6300 – 7500	9
9	7500 – 9000	7
		100

Перегруппируем данные в интервалах:

до 3000
3000 – 4500
4500 – 7500
7500 – 9000.

Получаем следующую группировку:

Группы рабочих по уровню заработной платы, тыс. руб.	Удельный вес рабочих, % к итогу	
	предприятие № 1	предприятие № 2
до 3000	11 (3 + 8)	13 (2 + 5+ 6)
3000 – 4500	39 (14 + 25)	37 (14 + 23)
4500 – 7500	44 (30 + 14)	43 (22 + 12 + 9)
7500 – 9000	6	7
Итого	100	100

Второй способ – способ долевой перегруппировки – применяется в тех случаях, когда интервалы новой группировки имеют иные границы и необходимо распределить число единиц совокупности, приходящихся на интервал первичной группировки, между новыми интервалами.

Например, интервалы новой группировки:

1200 – 2700

2700 – 4200

4200 – 5700

5700 – 7200

7200 и более

Основа перегруппировки: предположение о равномерном распределении признака внутри группы.

Например, интервал первичной группировки по предприятию № 1 4500 – 6000 необходимо распределить между 3-ей и 4-ой группой новой группировки. Ширина новой группировки 1500. В четвертую группу войдут значения от 4500 до 5700, то есть

$30\frac{1200}{1500}$ = 24 (%) – 3-ья группа новой группировки.

Значения от 5700 до 6000 войдут в 4-ую группу.

$30\frac{300}{1500}$ = 6 (%) – 4-ая группа новой группировки.

Получается следующая группировка:

Группы рабочих по уровню заработной платы, тыс. руб.	Удельный вес рабочих групп, % к итогу	
	предприятие № 1	предприятие № 2
1. 1200 – 2700	$3 + 8\frac{300}{600} = 7$	$2 + 5 + 6\frac{100}{200} = 10$
2. 2700 – 4200	$8\frac{300}{600} + 14 + 25\frac{600}{900} = 35$	$6\frac{100}{200} + 14 + 23\frac{150}{250} = 31$
3. 4200 – 5700	$25\frac{300}{900} + 30\frac{1200}{1500} = 32$	$23\frac{100}{250} + 22 + 12\frac{200}{400} = 37$
4. 5700 – 7200	$30\frac{300}{1500} + 14\frac{1200}{1500} = 17$	$12\frac{200}{400} + 9\frac{300}{400} = 13$
5. 7200 и более	$14\frac{300}{1500} + 6 = 9$	$9\frac{100}{400} + 7 = 9$
Итого	100	100

3.4 Статистические ряды распределения и их графическое изображение

В результате группировки единиц совокупности по какому-либо признаку получают ряды распределения.

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

В более широком смысле ряд распределения – это первичная характеристика массовой статистической совокупности, в которой находят количественное выражение закономерности массовых явлений и процессов общественной жизни.

Ряды распределения дают возможность:

а) проследить закономерность распределения;

б) судить об однородности совокупности и границах ее вариации;

в) исчислить различные обобщающие показатели (среднюю, моду, дисперсию и т. д.).

Ряды распределения могут быть построены по различным признакам: по атрибутивному – атрибутивные ряды распределения; по количественному – вариационные ряды распределения. Ряды распределения (вариационные) состоят из 2-х элементов: вариант (x) и частот (f).

Вариантами (x) называются отдельные значения признака.

Частотами (f) называются величины, показывающие, сколько раз повторяется данная варианта.

Иногда частоты могут выражаться в относительных величинах: долях единицы или в процентах. Тогда их называют частости.

Сумму всех частот (Σf) называют объемом ряда распределения или численностью (объемом) совокупности и обозначают N (n).

$\Sigma f = 1$, если это частости, выраженные в долях единицы;

$\Sigma f = 100$, если это частости, выраженные в процентах;

$\Sigma f = N\ (n)$ – численность совокупности (N – генеральной, n – выборочной).

В зависимости от характера вариации признака вариационные ряды могут быть: а) дискретными, когда величина признака принимает целочисленные значения.

Например:

Количество детей в семье x	Число семей f
0	100
1	120
2	110
3	50
4	5
5	2
6	1

б) интервальными, когда величина признака представлена в виде интервалов (как правило, непрерывный признак, но не обязательно).

Например:

Балл успеваемости студента x	Число студентов f
4 – 6	15
6 – 8	20
8 – 10	35

Порядок построения дискретного ряда распределения следующий:

1) ряд данных ранжируют, то есть располагают в порядке возрастания или убывания;

2) считают число повторений каждого значения признака, то есть частоту.

Например, имеются сведения об оценках, полученных на экзамене по статистике студентами группы (24 чел.):

Поряд.№ студента в списке			01	02	03	04	05	06	07	08	09	10	11
Оценка, x			5	6	8	10	2	6	9	6	7	4	6
№	12	13	14	15	16	17	18	19	20	21	22	23	24
x	9	3	9	7	10	5	8	9	7	8	4	8	7

1. Ранжируем уровни ряда (в порядке возрастания):

x : 2, 3, 4, 4, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10.

2. Строим дискретный ряд распределения (так как признак – прерывный):

x	2	3	4	5	6	7	8	9	10	
f	1	1	2	2	4	4	4	4	2	Σf = 24

При построении интервальных рядов порядок действий следующий:

1) определяют число групп,

2) определяют ширину интервала,

3) считают число значений признака, попадающих в каждый интервал.

Например, в нашем примере:

1) $n = 24$, следовательно, по формуле Стерджесса число групп $K = 5$;

2) ширина интервала $\frac{X\max - X\min}{K} = \frac{10-2}{5} = 1{,}6$;

3) строим интервальный ряд (таблица 3.5):

Таблица 3.5 – Группировка студентов по успеваемости

Оценка на экзамене, x	Число студентов, f
2,0 – 3,6	2
3,6 – 5,2	4
5,2 – 6,8	4
6,8 – 8,4	8
8,4 – 10,0	6
Всего	$\Sigma f = 24$

Для графического изображения рядов распределения широко применяются линейные и плоскостные диаграммы.

Так, для изображения дискретных вариационных рядов используется полигон.

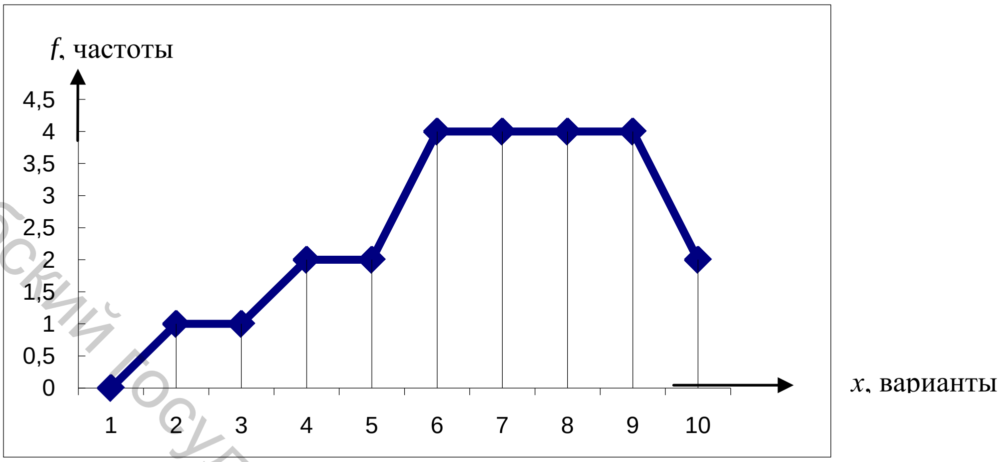

Рисунок 3.1 – Полигон распределения оценок, полученных на экзамене по статистике

При этом на оси абсцисс откладываются значения признака (x), на оси ординат – частоты (f). В местах их пересечения ставятся точки, которые затем соединяются ломаной линией (полигоном).

Для изображения интервальных рядов распределения обычно используют *гистограммы*. На оси абсцисс откладывают интервалы признака, на оси ординат – частоты и строят прямоугольники (основание прямоугольника – ширина интервала, высота – соответствующая частота).

В нашем примере:

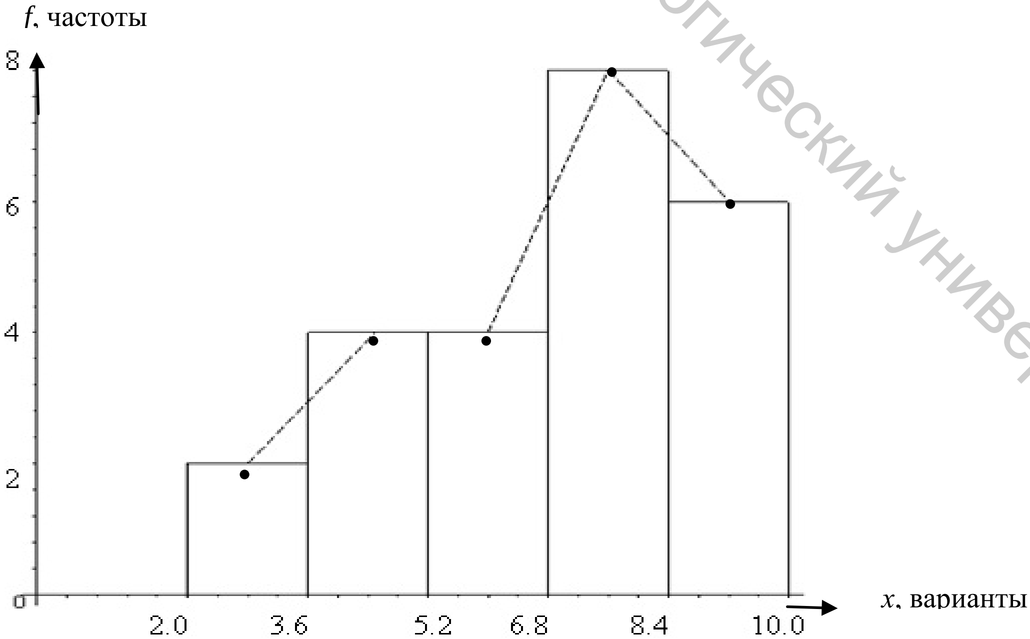

Рисунок 3.2 – Гистограмма распределения оценок, полученных на экзамене по статистике

Если интервалы в интервальном ряду *неравные,* то при построении гистограммы на ось ординат наносится *плотность распределения* – частота, рассчитанная на единицу ширины интервала.

Если на графике (рисунок 5) соединить ломаной линией центры интервалов, то получится полигон распределения для интервального ряда.

Ряды распределения могут также изображаться с помощью *кумуляты* (кумулятивной кривой или кривой сумм). На оси абсцисс – значения признака, на оси ординат – накопленные частоты. Накопленные частоты определяются последовательным суммированием по группам. Они показывают, сколько единиц совокупности имеют значения признака не большее, чем рассматриваемое значение.

Например, наш вариационный ряд (дискретный):

x	*f*	Накопленная частота
2	1	1
3	1	2 (1+1)
4	2	4 (1+1+2)
5	2	6 (1+1+2+2) и т. д.
6	4	10
7	4	14
8	4	18
9	4	22
10	2	24
	$\sum f = 24$	

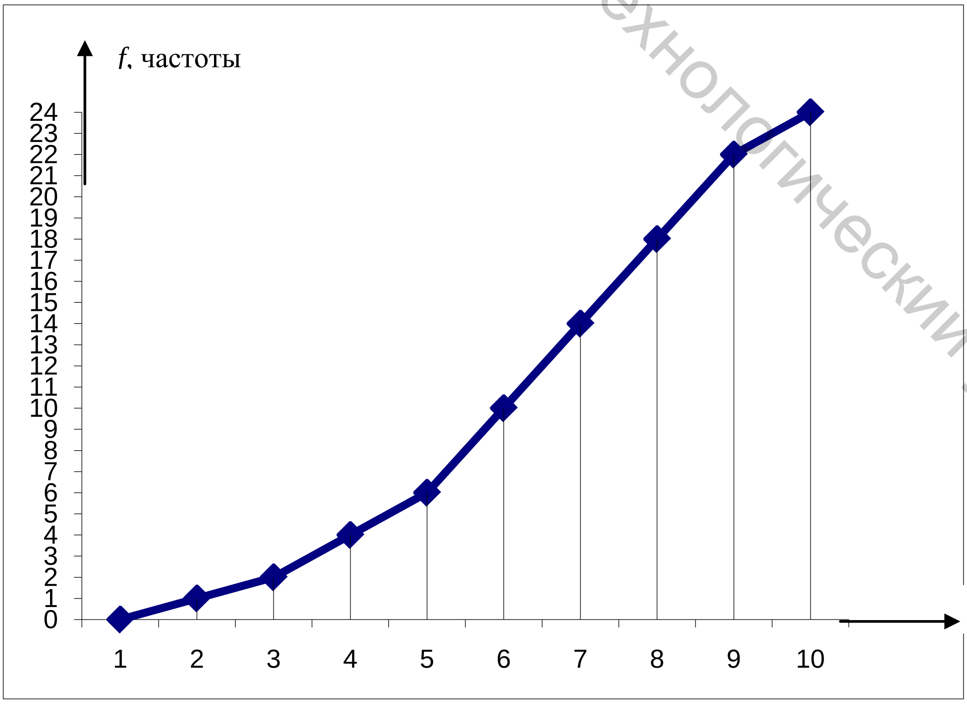

Рисунок 3.3 – Кумулята распределения оценок, полученных на экзамене по статистке

Помимо указанных способов графического изображения рядов распределения, некоторые авторы указывают на возможность использования для этих целей огивы. Однако существуют разные точки зрения. Так, в [15] отмечается, что огиву можно получить, если на оси ординат не частоты, а частости. А в [1] авторы указывают, что огива получается, если при графическом изображении кумуляты поменять местами оси.

3.5 Табличный и графический метод представления статистической информации

Результаты обработки статистических данных, как правило, оформляются в виде таблиц. Элементами статистической таблицы являются ее подлежащее (объекты изучения) и сказуемое (показатели, характеризующие эти объекты).

Незаполненная цифрами статистическая таблица называется макетом, то есть макет таблицы – это сетка, состоящая из горизонтальных строк и вертикальных колонок (граф), каждая из которых имеет название (если названия не указаны, то это скелет таблицы, ее остров). Клетки, образуемые на пересечении строк и граф заполняются статистическими данными.

По характеру подлежащего различают таблицы (рисунок 3.4):

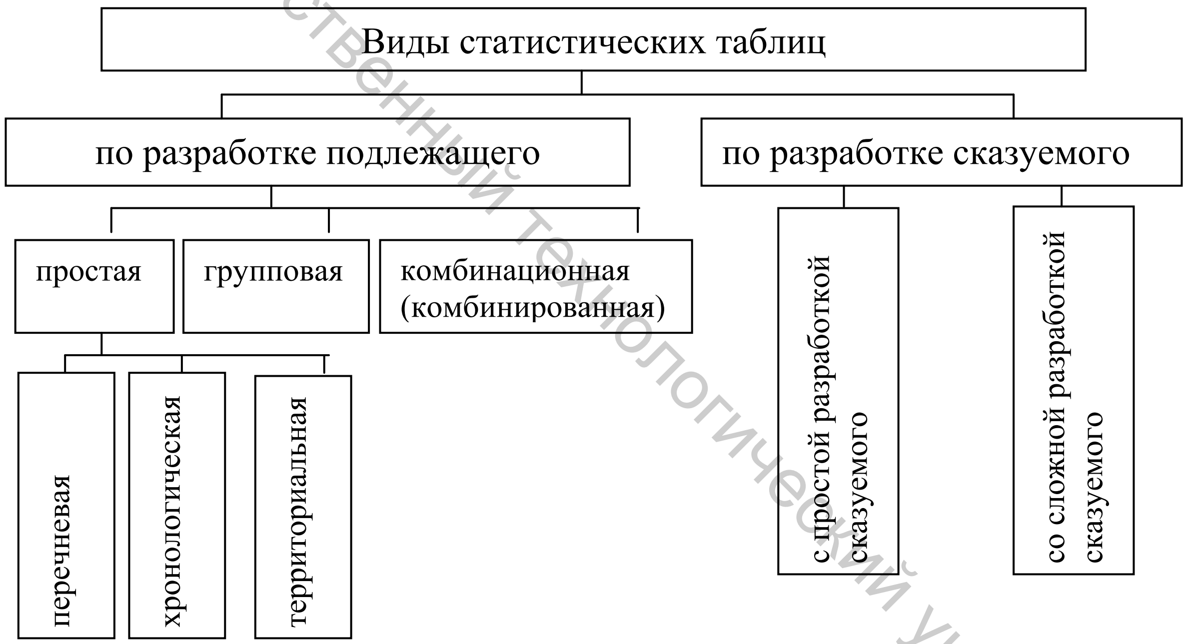

Рисунок 3.4 – Виды статистических таблиц

- *простые*, в которых подлежащее – это перечень отдельных единиц изучаемого объекта;
- *групповые*, в которых статистическая совокупность разбивается на отдельные группы по определенному (одному) признаку;
- *комбинационные*, в которых объект исследования разбивается на группы по нескольким признакам.

Разработка сказуемого таблицы также может быть простой и сложной. При простой разработке показатели, характеризующие подлежащее таблицы, располагаются параллельно друг другу, а при сложной один признак комбинируется с другим.

Например, таблица 3.6 – это пример комбинационной таблицы с простой разработкой сказуемого.

При составлении таблиц соблюдают следующие требования:

а) обязательное наличие заголовка (объект изучения, место и время, иногда и единицы измерения);

б) оглавление граф и строк и, при необходимости, их нумерация;

в) наличие на пересечении строки и графы числа или заменяющего его знака:

- текстовые данные;

Таблица 3.6 – Группировка организаций обрабатывающей промышленности по видам экономической деятельности и формам собственности

Вид экономической деятельности	Форма собственности организации	Число организаций	Среднесписочная численность работающих	Стоимость основных средств	Объем произведенной продукции
Производство продуктов питания	государственная				
	частная				
	иностранная				
	всего				
Текстильное производство	государственная				
	частная				
	иностранная				
	всего				
…	…	…	…	…	…
Всего по обрабатывающей промышленности	государственная				
	частная				
	иностранная				
	всего				

- знак прочерк (-), то есть явление не имело место;

- знак многоточие (…), то есть сведения о явлении не могут быть получены;

- знак (0,0), то есть величина показателя незначительна;

- знак («Х») – на пересечении строки и столбца получается показатель, не имеющий смысла;

г) подведение итогов по графам и столбцам, кроме случаев, когда этот итог не имеет смысла;

д) соблюдение общепринятых правил оформления таблиц, в том числе их нумерация, перенос, использование общепринятых сокращений, наличие внешних границ (линий) и возможность отсутствия внутренних и т. д.

Для наглядного предоставления материалов сводки и группировки в статистике также широко применяются графики.

Элементами графика являются:

- графический образ, то есть основа графика, представляющая собой совокупность точек, линий, фигур, с помощью которых отображены данные;
- поле графика, то есть пространство, в котором размещен график;
- пространственные ориентиры (оси координат или контурные линии);
- масштабные ориентиры;
- экспликация графика, объясняющая название предмета, смысловое значение знака и т. д. (пояснительные тексты).

В статистике применяется большое разнообразие графиков и различные признаки их классификации. Например, по назначению, по способу построения, по характеру графического образа и т. д. (рисунок 3.5).

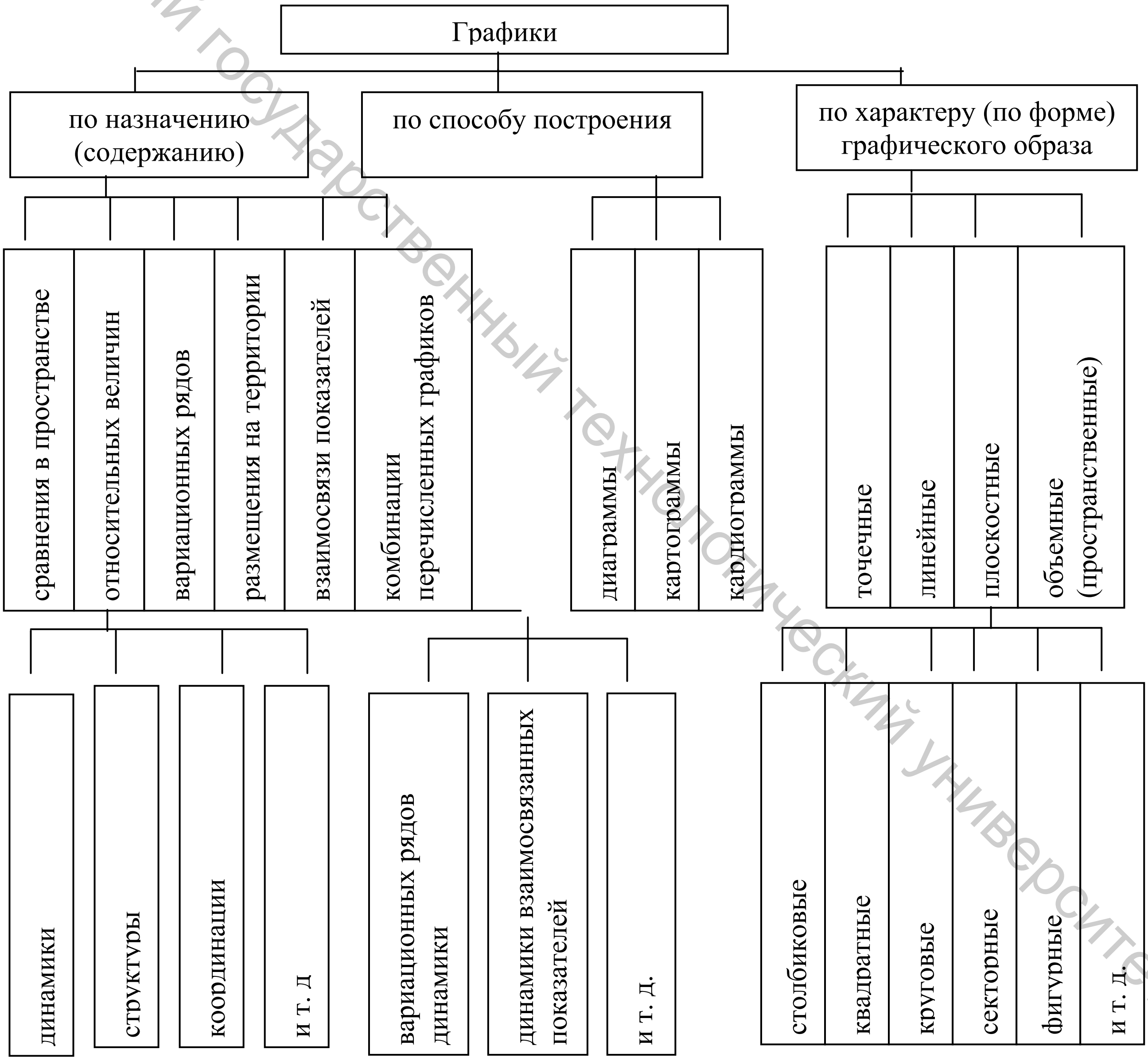

Рисунок 3.5 - Классификация графиков

Например, для сравнения одноименных показателей, характеризующих различные объекты или территории, целесообразно использовать столбиковую диаграмму, которая строится на общей горизонтальной или вертикальной базовой линии, отображаемые показатели – это высоты столбиков.

В тех случаях, когда сравниваемые объекты разнятся многократно и их невозможно сравнить с помощью столбиковой диаграммы (например, один столбик в 50 раз больше другого), используют квадратные или круговые диаграммы, у которых отображаемые показатели пропорциональны площадям квадратов или кругов.

Структура исследуемой совокупности чаще всего отображается с помощью секторной диаграммы, а иногда с помощью полосовой диаграммы удельных весов.

Для отражения динамики явлений и процессов чаще всего используют линейные диаграммы (линейные координатные диаграммы), у которых ось абсцисс – ось времени, а ось ординат – ось значений показателя.

Для изображения вариационных рядов применяют линейные и плоскостные диаграммы, в частности для сгруппированных данных – полигоны распределения, кумуляты, огивы и др.

Если необходимо сравнивать величины, представляющие собой произведение двух сомножителей (например, выпуск продукции представляется как произведение численности работников на их производительность труда), и показать роль каждого из них в формировании этой величины, строятся знаки Варзара.

4 СИСТЕМА СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ

4.1 Классификация статистических показателей
4.2 Абсолютные величины: их виды, способы получения и единицы измерения
4.3 Относительные величины: их виды и формы выражения

4.1 Классификация статистических показателей

В результате исследования статистической совокупности получают различные показатели, одни из которых характеризуют совокупность в целом, другие – ее части.

Статистический показатель – это количественная характеристика изучаемого объекта или его свойства.

Так, в результате сводки путем суммирования получают общий размер явления, например, общий фонд заработной платы. Если известно число отработанных человеко-дней и общий фонд заработной платы, путем деления второго на первое можно определить среднедневную заработную плату. Кроме этого, зная фонд заработной платы руководителей и специалистов и фонд заработной платы рабочих, можно определить долю (удельный вес) заработной платы рабочих в общем фонде заработной платы.

Все это примеры статистических показателей.

Совокупность взаимосвязанных статистических показателей, имеющая одноуровневую или многоуровневую структуру и нацеленная на решение конкретной задачи, образует систему статистических показателей.

Многообразие статистических показателей порождает необходимость их классификации по различным признакам.

Так, **по охвату единиц изучаемой совокупности** (или по степени агрегирования) их делят на:

а) индивидуальные – характеризующие один объект или одну единицу совокупности;

б) обобщающие (сводные) – характеризующие совокупность в целом или ее группы.

В свою очередь, обобщающие статистические показатели **по методологии исчисления** делятся на две большие группы:

1) экстенсивные (объемные) показатели, которые исчисляются по первичным признакам и характеризуют объем, массу общественных явлений. Они получаются как итог подсчета или суммирования статистических данных (то есть индивидуальных показателей или их составляющих частей) и показывают: численность единиц совокупности (например, численность работников) либо объем значений признака по совокупности (например, фонд заработной платы работников).

2) интенсивные показатели, которые исчисляются по вторичным признакам и рассчитываются на единицу совокупности. Это позволяет

устанавливать закономерности в развитии явлений. Интенсивные (качественные) показатели делятся на средние и относительные.

По форме выражения статистические показатели могут быть:

- абсолютными величинами,
- относительными величинами,
- средними величинами.

Для того, чтобы статистические показатели правильно отражали исследуемые процессы и явления, они должны соответствовать следующим требованиям:

1) теоретическая обоснованность – показатели должны выражать сущность изучаемого явления или процесса;

2) достоверность – их количественная оценка должна быть правильной, точной;

3) сопоставимость – они должны быть исчислены по единой методике, в одних пространственных границах и должна быть возможность сравнения их с плановыми, с другими периодами и т. д.

Кроме этого, большое значение имеет правильное название статистического показателя, его точная определенность. В названии статистического показателя должны быть отражены:

- его содержание (инвестиции, товарооборот и т. д.);
- его статистическая структура (среднее значение, процент к итогу, сумма и т. д.);
- позиция в классификации (если есть) (прямые иностранные инвестиции в РБ);
- единицы измерения;
- временные рамки (на начало года, за первый квартал 2013 г.);
- специальные уточнения (например, в ценах 2012 года и др.).

4.2 Абсолютные величины: их виды, способы получения и единицы измерения

Абсолютные статистические величины – это показатели, выражающие размер, объем и уровни общественных явлений.

Многообразие абсолютных величин требует их классификации по различным признакам:

1. По признаку **обобщения единиц совокупности** абсолютные величины могут быть:

– *индивидуальные* (полученные на основании статистического наблюдения);

- *итоговые* объемные (суммарные) (полученные в результате сводки и группировки)

2. По признаку **характеристики совокупности** их подразделяют на:

- показатели численности (численность экономически активного населения;

- показатели объема (фонд заработной платы, объем выпуска продукции).

3. По признаку **характеристики процесса развития** это могут быть:

- моментные показатели – характеризующие состояние явления на определенный момент времени (численность студентов на 01.02.2013);
- интервальные показатели – характеризующие результаты процессов за определенный период (выпуск продукции за год, прибыль за квартал и т. д.).

Все абсолютные величины являются именованными величинами, то есть выражаются в определенных единицах измерения. В качестве единиц измерения могут использоваться: натуральные единицы, условно-натуральные единицы, денежные (стоимостные), трудовые.

Учет в натуральных единицах принято называть натуральным учетом (кг, т, м, $м^3$, га и т. д.)

Натуральная единица измерения иногда может выражаться произведением двух натуральных измерителей: например, работа грузового транспорта – в тонно-километрах, электроэнергия – в киловатт-часах (кВт-час) и т. д.

Часто в статистических исследованиях объемов продукции применяют условно-натуральные показатели. Как правило, используются в случаях, если производится однородная, но не одинаковая продукция при анализе выполнения плана, динамики и т. д. Например, продукты химического производства, руды металлов соизмеряют по содержанию полезного вещества.

Наибольшее распространение при исследовании экономических явлений находят денежные (стоимостные) единицы измерения. Например, объем промышленной продукции, фонд заработной платы, прибыль, себестоимость и т. д. Денежный измеритель по праву называют универсальным.

Для измерения общих затрат труда на предприятии, трудоемкости отдельных операций технологического процесса и в других аналогичных случаях применяют трудовые единицы измерения: дни, часы, мин, сек., либо человеко-дни, человеко-часы.

4.3 Относительные величины, их виды и формы выражения

Чтобы провести полный анализ исследуемого явления, сделать правильные выводы о его развитии, недостаточно только абсолютных величин.

В ходе экономического анализа возникает необходимость в сравнении (сопоставлении).

Относительными величинами называются обобщающие показатели, характеризующие качественное соотношение двух сопоставляемых статистических величин.

При расчете относительных величин в числителе находится показатель, который сравнивается, а в знаменателе – база сравнения (основание).

В зависимости от того, к каким единицам приравнивается база сравнения, относительные величины имеют разную форму выражения:

а) если база сравнения принимается за 1, то относительная величина выражается как коэффициент;

б) если 100 – процент (%);

в) если 1000 – промилле (‰) (например, рождаемость на 1000 человек, смертность на 1000 человек, число кандидатов наук на 1000 человек);

г) если 10000 – продецимилле ($^{0}/_{000}$) (например, число врачей на 10000 жителей).

Иногда относительные величины могут быть именованными числами. Например, плотность населения – количество человек на 1 км^2.

По своему содержанию относительные величины подразделяются на виды:

- относительные величины динамики;
- относительные величины планового задания;
- относительные величины выполнения плана;
- относительные величины структуры;
- относительные величины координации;
- относительные величины интенсивности;
- относительные величины сравнения.

Относительные величины динамики (ОВД) характеризуют изменение уровня какого-либо явления во времени и называются темпами роста.

ОВД – отношение уровня признака в определенный период (момент) времени к уровню этого показателя в предшествующий период (момент).

Относительные величины планового задания (ОВПЗ) – это результат соотношения уровня, запланированного на предстоящий период, к фактическому уровню явления, сложившемуся в предшествующем периоде.

Относительные величины выполнения плана (относительные величины выполнения задания) (ОВВП) – это результат соотношения фактически достигнутого в данном периоде уровня к запланированному.

Относительные величины динамики, планового задания и выполнения плана связаны.

Относительные величины структуры (ОВС) характеризуют долю (удельный вес) отдельных частей совокупности в общем итоге.

Относительные величины координации (ОВК) характеризуют соотношение отдельных частей в изучаемой совокупности.

Относительные величины интенсивности (ОВИ) характеризуют степень распространенности или развития данного явления в той или иной среде.

Их определяют путем соотношения разноименных величин, находящихся в определенной взаимосвязи.

Например,

$$\frac{\textit{Число родившихся на данной территории}}{\textit{Среднегодовая численность населения данной территории}} * 1000 \text{ ‰}.$$

Разновидностью относительных величин интенсивности являются относительные показатели уровня экономического развития, например, ВВП на душу населения.

Относительные величины сравнения (ОВСр) характеризуют соотношение одноименных абсолютных величин, относящихся к одному и тому же периоду (моменту) времени, но к различным объектам либо территориям. Например, сравнение объемов промышленного производства Витебской и Могилевской области.

К общим принципам построения относительных величин можно отнести:

1) сопоставимость сравниваемых показателей (могут отличаться только одним атрибутом, например, временем, территорией);

2) взаимосвязь сравниваемых показателей;

3) необходимость учета границ их существования.

Например, абсолютные показатели численности работающих в организации характеризуется следующими данными:

Показатели	01.01.2012		01.01.2013	
	план	отчет	план	отчет
Всего работающих, чел.	500	495	510	515
В т. ч.				
- служащие	100	100	105	105
- рабочие	400	395	405	410

По данным этой таблицы может быть рассчитано множество относительных показателей. Приведем некоторые из них:

1. Относительная величина динамики:

$ОВД = \frac{515}{495} = 1,04$ (104 %).

Она показывает, что численность работающих организации 01.01.2013 г. в 1,04 раза больше чем на 01.01.2013.

2. Относительная величина планового значения:

$ОВПЗ = \frac{510}{495} = 1,03$ (103 %).

Это означает, что к 01.01.2013 планировалось увеличить численность работающих организации в 1,03 раза.

3. Относительная величина выполнения плана:

$ОВВП = \frac{515}{510} = 1,01$ (101 %).

План по численности работников перевыполнен на 101 %.

Проверим взаимосвязь:

$$ОВД = ОВПЗ \times ОВВП$$

$1,04 = 1,03 \times 1,01$

4. Относительная величина координации

$ОВК = \frac{400}{100} = 4$.

Следовательно по состоянию на 01.01.2013 в общей численности работающих численность рабочих составляла 4 %.

5 СРЕДНИЕ ВЕЛИЧИНЫ

5.1 Понятие средней величины и ее виды
5.2 Средняя арифметическая величина: ее расчет и свойств
5.3 Средняя гармоническая величина
5.4 Структурные средние: мода и медиана

5.1 Понятие средней величины и ее виды

Средняя величина является обобщающей характеристикой совокупности и представляет собой показатель, выражающий характерный, типичный, свойственный большинству признаков уровень.

По сути, средняя величина позволяет заменить множество индивидуальных значений признака, который варьирует у отдельных единиц наблюдения, какой-то одной определённой величиной.

Исходным соотношением средней является ее логическая формула:

$$\textit{Среднее значение признака в совокупности} = \frac{\textit{Сумма значений признака у всех единиц исследуемой совокупности}}{\textit{Число единиц (объем совокупности)}}$$

Определяющее свойство средней формируется следующим образом: сумма (произведение) индивидуальных значений признака равна сумме (произведению) средних значений признака.

Например, 10 рабочих одной и той же профессии с одним и тем же уровнем квалификации при выполнении однотипной работы имеют разные показатели заработной платы:

x_1 – 2700 тыс. руб.;	x_6 – 3180 тыс. руб.;
x_2 − 2820 тыс. руб.;	x_7 – 3060 тыс. руб.;
x_3 − 3060 тыс.руб.;	x_8 – 2700 тыс. руб.;
x_4 − 3180 тыс.руб.;	x_9 – 3240 тыс. руб.;
x_5 − 3240 тыс.руб.;	x_{10} – 2820 тыс. руб.

Тогда

$$\textit{средняя зар.плата } \bar{x} = \frac{2700 + 2820 + 3060 + 3180 + 3240 + 3180 + 3060 + 2700 + 3240 + 2820}{10} = 3000 \textit{ тыс.руб.}$$

Несмотря на то, что ни один рабочий не имел заработной платы, равной 3000 тыс. руб., результаты не искажены, так как

$$2700 + 2820 + 3060 + 3180 + 3240 + 3180 + 3060 + 2700 + 3240 + 2820 =$$
$$= 3000 * 10 = 30\,000$$

Вариация заработной платы (несмотря на целый ряд общих условий: одна операция, одинаковая квалификация, одинаковое оборудование и т. д.) объясняется действием множества случайных факторов: стаж работы, качество выполнения работы и т. д. Воздействие этих случайных факторов и погашается в средней величине. В этом выражается действие закона больших чисел: совокупное действие большого числа случайных факторов приводит, при некоторых весьма общих условиях, к результату, почти не зависящему от случая.

Средняя величина – это обобщённая количественная характеристика признака статистической совокупности в конкретных условиях места и времени.

Показатель в форме средней величины отражает типичные черты и даёт обобщающую характеристику однотипных явлений по какому-либо варьирующему признаку.

Для того, чтобы средняя величина была действительно типичной (типизирующей) величиной, необходимо соблюдать ряд требований:

1) при сборе и обработке информации необходимо обеспечить качественную однородность изучаемой совокупности.

2) необходимо обеспечить достаточный объём изучаемой совокупности, иначе не будет проявляться действие закона больших чисел, то есть взаимопогашения случайных факторов не произойдёт;

3) необходимо правильно выбрать вид средней величины.

Все средние величины делятся на два больших класса:

- степенные средние;
- структурные средние.

Из степенных средних в экономических исследованиях наибольшее распространение получили:

1) средняя арифметическая;
2) средняя гармоническая;
3) средняя геометрическая;
4) средняя квадратическая;
5) средняя кубическая и др.

К структурным средним относят моду и медиану.

Степенные средние в зависимости от представления исходных данных могут быть:

- простыми;
- взвешенными.

Простая средняя рассчитывается по несгруппированным данным, а взвешенная – по сгруппированным, то есть по дискретным или интервальным рядам, в которых указываются не только значения признака (x), но и частоты (повторяемости) – (f).

Доказано, что если рассчитать все виды средних для одних и тех же значений, то они будут неодинаковы: с увеличением показателя степени увеличивается и соответствующая средняя (правило мажорантности средних). Впервые правило мажорантности сформулировал русский статистик, профессор А.Я. Боярский.

$$\overline{X}_{\text{гарм.}} \le \overline{X}_{\text{геом.}} \le \overline{X}_{\text{арифм.}} \le \overline{X}_{\text{квадр.}} \le \overline{X}_{\text{куб.}}$$

В статистике чаще других используют средние арифметические и средние гармонические величины.

5.2 Средняя арифметическая величина: ее расчет и свойства

Наиболее распространённым видом средней величины является средняя арифметическая. Она может быть:

- простая;
- взвешенная.

Простая средняя арифметическая величина исчисляется в тех случаях, когда имеется несколько различных индивидуальных величин одного и того же вида. Тогда все они суммируются, и полученная сумма делится на их число.

Если обозначить эти индивидуальные значения x_1, x_2, x_3, x_4 … , а число индивидуальных значений (единиц наблюдения) – n, то средняя арифметическая простая будет равна:

$$\bar{x} = \frac{\sum x}{n}. \qquad (5.1)$$

Например, имеется информация о выработке за смену 10 рабочих бригады:

x, дет.: 32, 30, 34, 35, 34, 30, 33, 35, 34, 35.

$$\bar{x} = \frac{32+30+34+35+34+30+33+35+34+35}{10} = 33{,}2\ дет.$$

Однако в большинстве случаев исследователь имеет большую совокупность единиц, в которой уровни ряда x от случая к случаю повторяются. Тогда исходная информация представляется в виде дискретного ряда:

Таблица 5.1 – Информация о выработке рабочих, представленная в виде дискретного ряда

Выработка, дет.	Число рабочих, чел.	$x*f$
x	f	
30	2	60
32	1	32
33	1	33
34	3	102
35	3	105
	$\sum f = 10$	$\sum x*f = 332$

По сгруппированным данным рассчитывается средняя арифметическая взвешенная. Её формула:

$$\bar{x} = \frac{\sum x * f}{\sum f}. \qquad (5.2)$$

В нашем примере $\bar{x} = \frac{332}{10} = 33{,}2$ *дет.*

Частоты (f) в данном случае называют <u>весами</u>, поэтому средняя арифметическая <u>взвешенная</u>.

Аналогичным образом, по формуле средней арифметической взвешенной, рассчитывается средняя из <u>интервального ряда.</u>

Однако в данном случае

$$\bar{x} = \frac{\sum x'f}{\sum f}, \text{ где } x' - \text{середины или центры интервалов.} \qquad (5.3)$$

Например, необходимо рассчитать средний % выполнения норм выработки рабочими цеха.

Таблица 5.2 – Расчет среднего процента выполнения норм выработки

% выполнения норм выработки	Число рабочих, чел.	Центр (середина) интервала	$x'f$
x	f	x'	
100-105	10	102,5	1025,0
105-110	25	107,5	2687,5
110-115	20	112,5	2250,0
115-120	5	117,5	587,5
	$\sum f = 60$		$\sum x'f = 6550{,}0$

$$\bar{x} = \frac{\sum x'f}{\sum f} = \frac{6550}{60} = 109{,}2\ \%.$$

В случае, если имеются открытые интервалы, для определения центра первого интервала его ширину принимают равной ширине следующего за ним (то есть второго) интервала, а для определения центра последнего интервала приравнивают его ширину к ширине предшествующего (то есть предпоследнего).

Средняя арифметическая величина обладает рядом математических свойств, к основным из которых относятся следующие:

1. Произведение средней величины на сумму всех частот равно сумме произведений индивидуальных значений признака на соответствующие частоты (свойство вытекает из формулы $\bar{x} = \frac{\sum xf}{\sum f}$):

$$\bar{x} * \sum f = \sum xf. \qquad (5.4)$$

2. Сумма отклонений индивидуальных значений признака от средней величины равна 0:

$\sum\left(x-\overline{x}\right)=0$ – для несгруппированных данных;

$\sum\left(x-\overline{x}\right)*f=0$ – для сгруппированных данных.

3. Сумма квадратов отклонений индивидуальных значений признака от средней величины меньше суммы квадратов их отклонений от любой другой постоянной величины (x_0):

$\sum\left(x-\overline{x}\right)^2 < \sum\left(x-x_0\right)^2$ – для несгруппированных данных;

$\sum\left(x-\overline{x}\right)^2 * f < \sum\left(x-x_0\right)^2 * f$ – для сгруппированных данных.

4. Средняя арифметическая суммы варьирующих величин равна сумме средних арифметических этих величин:

если $x_i = y_i + z_i$, то $\overline{x}=\overline{y}+\overline{z}$.

5. Если все варианты ряда уменьшить (увеличить) в А раз, то средняя уменьшается (увеличивается) в А раз.

Это означает, что $\overline{x}=\frac{\sum xf}{\sum f}$ можно исчислять как $\overline{x}=\frac{\sum\frac{x}{A}}{\sum f}*A$ либо $\overline{x}=\frac{\sum xA*f}{\sum f}/A$.

6. Если все варианты ряда уменьшить (увеличить) на одно и то же число x_0, то и средняя величина уменьшиться (увеличиться) на x_0, то есть $\overline{x}=\frac{\sum xf}{\sum f}$ может быть рассчитана как $\overline{x}=\frac{\sum\left(x-x_0\right)*f}{\sum f}+x_0$ либо $\overline{x}=\frac{\sum\left(x+x_0\right)*f}{\sum f}-x_0$.

7. Если все частоты ряда разделить (умножить) на одно и то же число b, то средняя не изменится, то есть $\overline{x}=\frac{\sum xf}{\sum f}$ может быть рассчитана как $\overline{x}=\frac{\sum x*\frac{f}{b}}{\sum\frac{f}{b}}$

либо как $\overline{x}=\frac{\Sigma xf*b}{\Sigma f*b}$

Последние три свойства из перечисленных могут использоваться одновременно для упрощения расчетов, и тогда считается, что средняя рассчитывается по «способу моментов» или «методом отсчета от условного нуля». В данном случае важен факт правильного выбора А (чаще всего это величина интервала) и x_0 (чаще всего это середина какого-либо интервала).

Исчисление средней по «способу моментов» производится по формуле, вид которой меняется в зависимости от порядка применения свойств:

$$\bar{x} = \frac{\sum \frac{x - x_0}{A} * \frac{f}{b}}{\sum \frac{f}{b}} * A + x_0 \qquad (5.5)$$

либо
$$\bar{x} = \left[\frac{\sum \left(\frac{x}{A} - x_0 \right) * \frac{f}{b}}{\sum \frac{f}{b}} + x_0 \right] * A$$

и т. д.

Независимо от того, применяются либо не применяются свойства средней величины, результат расчета средней остается неизменным.

Например, необходимо определить среднюю численность работников организации, в т. ч. по «способу моментов».

Таблица 5.3 – Расчет средней численности работков

Численность работников, x	Число организаций, f	x'	$x'f$	$x-x_0$ ($x_0 = 1250$)	$\frac{x-x_0}{A}$ ($A = 500$)	$\frac{f}{b}$ ($b = 20$)	$\frac{x-x_0}{A} * \frac{f}{b}$
до 500	40	250	10000	-1000	-2	2	-4
500 - 1000	40	750	30000	-500	-1	2	-2
1000 – 1500	160	1250	200000	0	0	8	0
1500 – 2000	80	1750	140000	500	1	4	4
2000 – 2500	40	2250	90000	1000	2	2	4
2500 – 3000	20	2750	55000	1500	3	1	3
3000 и более	20	3250	65000	2000	4	1	4
	$\Sigma f = 400$		$\Sigma xf =$ 590000			$\Sigma \frac{f}{b} = 20$	$\Sigma \frac{x-x_0}{A} * \frac{f}{b} = 9$

Без использования свойств средней величины:

$$\bar{x} = \frac{\sum x' * f}{\sum f} = \frac{590000}{400} = 1475 \text{ чел.}$$

С использованием свойств средней величины:

$$\bar{x} = \frac{\sum \frac{x - x_0}{A} * \frac{f}{b}}{\sum \frac{f}{b}} * A + x_0 \quad . \qquad (5.6)$$

$$\bar{x} = \frac{9}{20} * 500 + 1250 = 1475 \text{ чел.}$$

5.3 Средняя гармоническая величина

Средняя гармоническая величина применяется в тех случаях, когда известны индивидуальные значения признака x и произведения $x*f$, но отсутствуют частоты f.

Например, необходимо определить среднюю заработную плату рабочих по предприятию, если известны: уровни средней заработной платы и фонды заработной платы:

Таблица 5.4 – Расчет средней заработной платы

№ цеха	Средняя заработная плата рабочих цеха, тыс. руб.	ФЗП рабочих, тыс. руб.	$f = \frac{W}{x}$
	x	$x*f = W$	
1	6000	300000	50
2	6240	468000	75
3	6600	462000	70
		$\sum W = 1230000$	$\sum \frac{W}{x} = 195$

В данном случае произведение $x*f$ обозначается W (может быть m, z и т. д.), и средняя рассчитывается по формуле

$$\bar{x} = \frac{\sum W}{\sum \frac{W}{x}}; \tag{5.7}$$

$$\bar{x} = \frac{1230000}{195} = 6307{,}70 \text{ тыс. руб.}$$

Средняя в такой форме называется средней гармонической взвешенной.

На практике к необходимости исчисления средней гармонической величины могут приводить, например, следующие случаи:

1) по имеющейся информации необходимо определить средний расход материала на одно изделие.

Таблица 5.5 – Порядок расчета среднего расхода материала

Вид продукции	Расход материала на единицу продукции	Расход материала на весь выпуск	Число изделий
	x	$W = x * f$	$f = \frac{W}{x}$
А			
Б			
…			

2) по имеющейся информации необходимо определить среднюю выработку одного рабочего по предприятию.

Таблица 5.6 – Порядок расчета средней выработки

№ цеха	Средняя выработка одного рабочего	Выпуск продукции цеха	Число рабочих
	x	$W = x*f$	$f = \frac{W}{x}$
1			
2			
…			

Средняя гармоническая взвешенная определяется по сгруппированным данным.

В тех же случаях, когда произведения $x*f$ одинаковы или равны единице, применяется <u>средняя гармоническая простая.</u>

$$\bar{x} = \frac{n}{\sum \frac{1}{x}} \tag{5.8}$$

5.4 Структурные средние: мода и медиана

Мода и медиана относятся к структурным средним и применяются для изучения внутреннего строения рядов распределения признака.

Мода (μ_0) − это наиболее часто встречающаяся величина признака в вариационном ряду.

Например, какой стаж работы встречается у рабочих наиболее часто

X, лет: 3, 12, 5, 7, 10, 3, 7, 1, 15, 7, 10. $\mu_0 = 7$ лет.

В дискретном ряду моду будет представлять то значение признака (та варианта), которое имеет наибольшую частоту.

Например, какие оценки на экзамене встречаются наиболее часто:

Оценка	x	2	3	4	5	6	7	8	9	10
							↑			
Число студентов	f	1	2	7	7	8	15	5	3	2
							f_{max}			

$\mu_o = 7.$

Для расчета моды в интервальном ряду вначале определяется модальный интервал, то есть интервал, имеющий наибольшую частоту. Затем рассчитывают моду по формуле

$$\mu_o = x_{\mu_0} + i_{\mu_0} \frac{f_{\mu_0} - f_{\mu_{0-1}}}{\left(f_{\mu_0} - f_{\mu_{0-1}}\right) + \left(f_{\mu_0} - f_{\mu_{0+1}}\right)}, \qquad (5.9)$$

где x_{μ_0} – начальная граница модального интервала,

i_{μ_o} – ширина модального интервала,

f_{μ_o} – частота модального интервала,

$f_{\mu_{o-1}}$ – частота интервала, предшествующего модальному,

$f_{\mu_{o+1}}$ – частота интервала, следующего за модальным.

Например, необходимо определить моду численности работающих в организации.

Таблица 5.7 – Группировка предприятий по числу работающих

Группы организаций по числу работающих	Количество организаций в группе	
x	*f*	
500 −1000	5	
1000 −1500	10	
1500 − 2000	15	– модальный интервал
2000 − 2500	14	
2500 − 3000	6	
Всего	$\Sigma f = 50$	

$$\mu_0 = 1500 + 500 \frac{15-10}{(15-10)+(15-14)} = 1916\textit{чел.}$$

Медиана (μ_e) – это величина варьирующего признака, которая находится в середине ранжированного ряда.

Например, имеется информация о стаже работы членов бригады, лет (*x*): 5, 2, 10, 15, 2, 5, 7, 8, 5.

Вначале ранжируем ряд:

x: 2, 2, 5, 5, 5, 7, 8, 10, 15

Затем определяем серединное значение: $\mu_e = 5$

То есть медиана делит ряд на 2 части, равные по численности. Половина значений меньше (либо равны) медианы, а вторая – больше (либо равны). Если ряд состоит из нечетного количества уровней (вариант), то порядковый номер медианы в ранжированном ряду:

$$№_{\mu_e} = \frac{n+1}{2}.$$

В нашем примере $\frac{9+1}{2} = 5 - ый$.

Если же ряд состоит из четного количества уровней, то медиана определяется как средняя арифметическая из варианты под $№ = \frac{n}{2}$ и варианты $№ = \frac{n}{2} + 1$.

Например, *x*, 2, 2, 4, 5, 5, 5, 7, 8, 10, 10, 15, 18

$\frac{n}{2} = 6$, $x_6 = 5$; $\frac{n}{2} + 1 = 7$, $x_7 = 7$.

Следовательно, в данном случае

$$\mu_e = \frac{5+7}{2} = 6 \ лет.$$

При определении медианы в дискретном ряду используют способ накопления частот. Частоты накапливают до тех пор, пока сумма накопленных частот ($S_{\mu e}$) не будет равна или больше половины суммы всех частот (Σf). Последняя накопленная частота и будет указывать то значение признака, которое является медианой.

Например, определить медиану заработной платы работников.

Таблица 5.8 – Определение медианы в дискретном ряду

ЗП, тыс. руб.	Число работников, чел.	$S_{\mu e}$
x	f	
4000	2	2
4800	6	8
6000	16	24
6800	12	
8000	4	
	$\Sigma f = 40$	

μ_e = 6000 тыс. руб.

В случае, если сумма накопленных частот составила ровно половину всех частот, медиана определяется как средняя из данного уровня и следующего за ним.

Таблица 5.9 – Определение медианы в дискретном ряду

ЗП, тыс. руб.	Число работников, чел.	$S_{\mu e}$
x	f	
4000	2	2
4800	6	8
6000	12	20
6800	16	
8000	4	
	$\Sigma f = 40$	

$$\mu_e = \frac{6000 + 6800}{2} = 6400\ тыс.руб.$$

Для определения медианы в интервальном ряду вначале с помощью суммы накопленных частот определяют медианный интервал, а затем рассчитывают медиану по формуле

$$\mu_e = x_{\mu_e} + i_{\mu_e} \frac{0.5\sum f - S_{\mu_{e-1}}}{f_{\mu_e}}, \qquad (5.10)$$

где $X_{\mu e}$ – начальная граница медианного интервала;

$i_{\mu e}$ – ширина медианного интервала;

$f_{\mu e}$ – частота медианного интервала;

$S_{\mu e\text{-}1}$ – сумма накопленных частот интервала, предшествующего медианному.

Например, определить медиану численности работников организации.

Таблица 5.10 – Исходные данные для расчета медианы численности работников

Группы организаций по числу работников	Количество организаций	$S_{\mu e}$	
x	*f*		
500 − 1000	5	5	
1000 − 1500	10	15	
1500 − 2000	15	30	медианный интервал
2000 − 2500	14	44	
2500 − 3000	6	50	
	$\Sigma f = 50$		

$$\mu_e = 1500 + 500\frac{0.5*50 - 15}{15} = 1833\,чел.$$

6 СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВАРИАЦИИ

6.1 Вариация признака и ее роль в статистических исследованиях
6.2 Показатели вариации
6.3 Виды дисперсии. Математические свойства дисперсии

6.1 Вариация признака и ее роль в статистических исследованиях

Вариацией признака называется его изменение при переходе от одной единицы наблюдения к другой.

С другой стороны, вариация – это то, что порождает необходимость статистики. Так например, лишено смысла исследование стоимости автобусного билета в городе Витебске, так как это явление не варьирующее.

После того, как исчислена средняя величина, возникает вопрос о её надёжности или её типичности. При этом необходимо учитывать, что типичность средней находится в обратной зависимости от вариации (колеблемости) уровней признака: чем больше вариация уровней исходной информации, тем меньше типичность (представительность, репрезентативность) средней величины. В случае слишком большой вариации уровней ряда можно получить фиктивную среднюю.

Рассмотрим два примера, в которых уровни средних величин равны, но эти средние имеют разные представительности.

Пример: имеется информация о заработной плате рабочих двух цехов.

Таблица 6.1 – Распределение рабочих цеха № 1 по уровню заработной платы

Заработная плата, млн. руб. x	Число рабочих, чел f	xf
3,0	10	30
4,2	10	420
4,6	10	46
5,8	10	58
	$\sum f = 40$	$\sum xf = 176$

Таблица 6.2 – Распределение рабочих цеха № 2 по уровню заработной платы

Заработная плата, млн. руб. x	Число рабочих, чел. f	xf
4,0	5	20
4,2	10	42
4,4	10	44
4,6	10	46
4,8	5	24
	$\sum f = 40$	$\sum xf = 176$

В обоих случаях средняя выработка рассчитывается по формуле средней арифметической взвешенной:

Цех № 1 $\bar{x} = \frac{\sum xf}{\sum f} = \frac{176}{40} = 4{,}4$ *млн. руб.*

Цех № 2 $\bar{x} = \frac{\sum xf}{\sum f} = \frac{176}{40} = 4{,}4$ *млн.руб.*

Однако, анализируя уровни исходной информации, замечаем:

Цех № 1 − вариация заработной платы от 3,0 до 5,8 млн. руб.

Цех № 2 − вариация выработки от 4,0 до 4,8 млн. руб.

Следовательно, $\bar{x} = 4{,}4$ млн. руб. наиболее типична для цеха № 2.

Для характеристики степени вариации признака (а следовательно, для оценки типичности средней величины) в статистике используют следующие показатели:

1) размах вариации (R);
2) среднее линейное отклонение (l);
3) дисперсия (σ^2);
4) среднее квадратическое отклонение (σ);
5) коэффициент осцилляции (K_r);
6) относительное линейное отклонение (K_l);
7) коэффициент вариации (V);

Первые четыре относятся к абсолютным показателям вариации, а последние три − к относительным.

Значение показателей вариации заключается в следующем:

1) они дополняют средние величины, за которыми скрываются индивидуальные различия отдельных единиц совокупности;
2) они характеризуют степень однородности статистической совокупности по изучаемому признаку;
3) они характеризуют границы вариации признака;
4) соотношение показателей вариации может быть использовано для характеристики взаимосвязи между признаками (см. 6.3).

5.2 Показатели вариации

Показатели вариации относятся к числу обобщающих показателей, они измеряют вариацию совокупности явлений.

Рассмотрим порядок их расчета на примере оценки вариации заработной платы рабочих цеха № 2.

Таблица 6.3 – Распределение рабочих цеха № 2 по заработной плате

Заработная плата, млн.руб. x	Число рабочих, чел. f	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert$ f	$(x-\bar{x})^2$	$(x-\bar{x})^2$ f	x^2	$x^2 f$
4,0	5	0,4	2	0,16	0,80	16,00	80,00
4,2	10	0,2	2	0,04	0,40	17,64	176,40
4,4	10	0	0	0	0	19,36	193,60
4,6	10	0,2	2	0,04	0,40	21,16	211,60
4,8	5	0,4	2	0,16	0,80	23,04	115,20
	$\sum f = 40$		$\sum\lvert x-\bar{x}\rvert f = 8$		$\sum(x-\bar{x})^2 f = 2,4$		$\sum x^2 f = 776,80$

6.2.1 Размах вариации – это разность между максимальным и минимальным значениями признака.

$$R = X_{\max} - X_{\min}. \qquad (6.1)$$

В нашем примере: R = 4,8 – 4,0 = 0,8 (млн.руб.)

Преимущество данного показателя: простота его исчисления.

Недостатки: он не учитывает внутреннюю колеблемость уровней ряда и часто зависит от случайности.

Область применения R поэтому ограничивается достаточно однородными совокупностями.

6.2.2 Среднее линейное отклонение – это средняя арифметическая величина, исчисленная из абсолютных отклонений индивидуальных значений признаков от средней величины.

Однако, учитывая нулевое свойство средней арифметической: сумма отклонений индивидуальных значений признака от средней равна 0. Поэтому при исчислении среднего линейного отклонения суммируются модули этих отклонений.

То есть формулы исчисления среднего линейного отклонения имеют вид:

$l = \dfrac{\sum\lvert x-\bar{x}\rvert}{n}$ для простого вариационного ряда (для несгруппированных данных) либо

$l = \dfrac{\sum\lvert x-\bar{x}\rvert * f}{\sum f}$ для дискретного ряда (для сгруппированных данных).

В нашем примере: l = 8 / 40 = 0,2 (млн. руб.)

Преимущество среднего линейного отклонения перед размахом в том, что оно учитывает внутреннюю вариацию уровней ряда.

Недостаток: необходимо абстрагирование от знака отклонения, следовательно – трудности в применении математических методов анализа вариации.

6.2.3 Дисперсия – наиболее распространенный в научной статистике показатель.

Дисперсия – это средняя из квадратов отклонений индивидуальных значений признака от его средней величины.

Она рассчитывается по формулам 6.2 и 6.3:

- для простого вариационного ряда

$$\sigma^2 = \frac{\sum\left(x-\bar{x}\right)^2}{n}; \tag{6.2}$$

- для дискретного ряда

$$\sigma^2 = \frac{\sum\left(x-\bar{x}\right)^2 * f}{\sum f}. \tag{6.3}$$

В нашем примере: $\sigma^2 = \frac{2{,}40}{40} = 0{,}06.$

Дисперсия может быть исчислена и другим способом: как разность между средней из квадратов индивидуальных значений признака и квадратом средней величины, то есть

$$\sigma^2 = \overline{x^2} - \bar{x}^2. \tag{6.4}$$

Тогда для простого вариационного ряда формула дисперсии принимает вид:

$$\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2, \tag{6.5}$$

а для дискретного

$$\sigma^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2. \tag{6.6}$$

В нашем примере:

$$\sigma_2^2 = \frac{776{,}80}{40} - \left(\frac{176}{40}\right)^2 = 0{,}06.$$

Результаты получаются одинаковые, независимо от применяемой формулы.

Недостаток этого показателя вариации – его размерность. Размерность дисперсии равна квадрату размерности изучаемого признака. Этот недостаток устраняется при переходе к среднему квадратическому отклонению.

6.2.4 Среднее квадратическое отклонение определяется как квадратный корень из среднего квадрата отклонений индивидуальных значений признака от их средней величины, то есть:

- для простого вариационного ряда

$$\sigma = \sqrt{\frac{\sum \left(x - \overline{x}\right)^2}{n}}\,; \qquad (6.7)$$

- для дискретного ряда

$$\sigma = \sqrt{\frac{\sum \left(x - \overline{x}\right)^2 * f}{\sum f}}\,. \qquad (6.8)$$

Следовательно, если дисперсия уже исчислена, среднее квадратическое отклонение рассчитывается извлечением квадратного корня из дисперсии:

$$\sigma = \sqrt{\sigma^2}\,. \qquad (6.9)$$

В наше примере:

$$\sigma_2 = \sqrt{0{,}06} = 0.2449\,(\text{млн.руб.})$$

Возведение $\left(x - \overline{x}\right)$ в квадрат, а затем извлечение квадратного корня позволяет избежать воздействия нулевого свойства средней арифметической.

В отличие от абсолютных показателей вариации, назначение относительных показателей – оценка вариации признака в % (либо в коэффициентах). Они рассчитываются как отношение абсолютных показателей вариации к средней арифметической величине признака.

Наиболее простыми и менее распространенными относительными показателями являются:

6.2.5 Коэффициент осцилляции:

$$K_R = \frac{R}{\overline{x}} * 100\,. \qquad (6.10)$$

В нашем примере:

$$K_R = 0{,}8 / 4{,}4 * 100 = 18{,}18\ \%.$$

6.2.6 Относительное линейное отклонение или линейный коэффициент вариации:

$$K_e = \frac{l}{\overline{x}} * 100\ . \qquad (6.11)$$

В нашем примере:

$$K_l = 0{,}2 / 4{,}4 * 100 = 4{,}54\ \%.$$

Самым распространенным относительным показателем вариации является коэффициент вариации:

$$V = \frac{\sigma}{\overline{x}} * 100\,. \qquad (6.12)$$

$$V = \frac{0{,}2449}{4{,}4} * 100 = 5{,}57\,\%$$

По величине коэффициента вариации можно судить о степени вариации признаков совокупности.

На практике коэффициент вариации находит широкое применение для сравнения вариации одного и того же признака в разных совокупностях, а также для сравнения вариации разных признаков одной и той же совокупности.

Кроме этого, коэффициент вариации используется в оценке ритмичности работы предприятия.

Совокупность считается достаточно однородной, если V ≤ 30 %.

6.3 Виды дисперсии. Математические свойства дисперсии

Вариация признака складывается под воздействием множества факторов, так как социально-экономические явления и процессы носят сложный характер. В исследованиях иногда возникает необходимость оценить не только общую вариацию признака, но и ту ее часть, которая обусловлена действием постоянных, стабильных, а не случайных факторов. В этих случаях рассчитывают три вида дисперсии:

- общую;
- межгрупповую;
- внутригрупповую.

Общая дисперсия характеризует общую вариацию признака под влиянием всех факторов (условий, причин). Она рассчитывается по формуле 6.13:

$$\sigma_0^2 = \frac{\sum (x-\bar{x})^2 * f}{\sum f}, \qquad (6.13)$$

где $\bar{x}$ − средняя по всей изучаемой совокупности.

Для определения влияния постоянного фактора на вариацию признака производят аналитическую группировку, в основании которой лежит данный фактор. Вариация, обусловленная фактором, положенным в основание группировки, оценивается с помощью межгрупповой дисперсии:

$$\delta_0^2 = \frac{\sum \left(\bar{x}_{гр} - \bar{x}\right)^2 * f_{гр}}{\sum f_{гр}}, \qquad (6.14)$$

где $\bar{x}_{гр}$ – средняя по отдельным группам;

$f_{гр}$ − численность отдельных групп.

Для определения влияния случайных факторов рассчитывают дисперсию внутри каждой группы, то есть внутригрупповую:

$$\sigma_{гр}^2 = \frac{\sum \left(x_{гр} - \bar{x}_{гр}\right) * f_{гр}}{\sum f_{гр}}, \qquad (6.15)$$

где $x_{гр}$ – индивидуальные значения признака в группе,

$f_{гр}$ – их частоты;

а затем <u>среднюю из внутригрупповых дисперсий:</u>

$$\overline{\sigma_{гр}^2} = \frac{\sum \sigma_{гр}^2 * f_{гр}}{\sum f_{гр}}. \qquad (6.16)$$

Доказано, что общая дисперсия равна сумме межгрупповой дисперсии и средней из внутригрупповых дисперсий:

$$\sigma_0^2 = \delta^2 + \overline{\sigma_{гр}^2}. \qquad (6.17)$$

Это равенство называется правилом сложения дисперсий.

Наряду с вариацией количественного признака часто возникает необходимость измерить вариацию альтернативного признака.

Если ввести обозначения:

p – доля единиц, обладающих данным признаком, интересующим исследователя;

q – доля единиц, не обладающих данным признаком.

Дисперсия альтернативного признака равна произведению доли единиц, обладающих данным признаком, на долю единиц, которые им не обладают, то есть

$$\sigma^2 = pq \text{ либо } \sigma^2 = p(1-p). \qquad (6.18)$$

Дисперсия облает рядом математических свойств, которые значительно упрощают её вычисление. К основным из них относятся следующие:

1. Если все значения признака увеличить или уменьшить в A раз, то дисперсия соответственно увеличится или уменьшится в A^2 раз.

2. Если все значения признака увеличить или уменьшить на какое-то постоянное число x_0, то дисперсия от этого не изменится.

3. Если все значения частот различить или умножить на какое-то число b, то дисперсия от этого не изменится.

Используя эти свойства одновременно, можно рассчитать дисперсию по «способу моментов». Если взять за основу исходную формулу дисперсии

$$\sigma^2 = \frac{\Sigma x^2 f}{\Sigma f} - \left(\frac{\Sigma xf}{\Sigma f}\right)^2, \qquad (6.19)$$

то формула дисперсии, исчисляемой по «способу моментов», будет иметь вид:

$$\sigma^2 = \left[\frac{\Sigma\left(\frac{x-x_0}{\mathrm{A}}\right)^2 * \frac{f}{b}}{\Sigma \frac{f}{b}} - \left(\frac{\Sigma \frac{x-x_0}{\mathrm{A}} * \frac{f}{b}}{\Sigma \frac{f}{b}}\right)^2\right] * \mathrm{A}^2. \qquad (6.20)$$

7 ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

7.1 Понятие выборочного наблюдения. Обобщающие характеристики генеральной и выборочной совокупности

Наибольшее распространение на практике получил такой вид несплошного наблюдения, как выборочное наблюдение.

Выборочным наблюдением называют такой способ несплошного наблюдения, при котором обследуется не вся совокупность, а лишь отобранная по определенным правилам ее часть, обеспечивающая получение данных для характеристики совокупности в целом.

Основная цель выборочного наблюдения: по характеристикам отобранной части единиц совокупности сделать вывод о характеристиках всей совокупности.

Многократное использование выборочного наблюдения и сопоставление его результатов с результатами сплошного наблюдения доказывают, что характеристики выборки достаточно точно воспроизводят характеристики генеральной совокупности.

Появление выборочного метода в статистике, как и сама статистика, было вызвано потребностями практики: проведение выборочных обмолотов зерна для определения сборов зерна, оценка качества земель и сенокосов, оценка качества товаров (вывозимых на мировой рынок) и т. п.

Русская наука имела заслуги не только в практическом использовании выборочного наблюдения, но и в его теоретическом обосновании. Основные положения теории выборочного наблюдения были разработаны такими выдающимися математиками, как П.Л. Чебышев, А.М. Ляпунов, А.А. Марков и др.

В настоящее время выборочное наблюдение находит широкое применение в различных сферах национальной экономики: в промышленности (контроль качества продукции, изучение использования рабочего времени, изучение использования оборудования и т. д.); в сельском хозяйстве (определение урожайности, определение потерь при уборке урожая и т. д.); в торговле (проверка качества поступивших товаров, изучение спроса населения и т. д.); в изучении потребления и благосостояния населения (обследование жилищных условий и т. д.) и т. п.

Если выборочное наблюдение обеспечивает получение характеристик, близких к характеристикам генеральной совокупности, то соблюдается условие репрезентативности, или представительности выборки. Для обеспечения репрезентативности выборки необходимо строго соблюдать основные правила отбора единиц из генеральной в выборочную совокупность:

1) строго объективный подход к отбору единиц: каждая единица совокупности должна иметь равную вероятность попадания в выборку (полностью должно отсутствовать субъективное мнение исследователя);

2) необходимо обеспечить достаточно большое количество единиц в выборке (действие закона больших чисел: взаимопогашение случайных нетипичных отклонений). Это правило демонстрирует основное противоречие, с которым сталкивается исследователь при определении количества единиц:

- с одной стороны, стремление уменьшить объем наблюдения;

- с другой стороны, необходимость обеспечить достаточно большой объем наблюдения;

3) необходимо учитывать степень вариации признака: чем меньше вариация, тем успешнее достигается требование репрезентативности.

Методологически выборочное наблюдение сложнее, чем сплошное, так как требует глубокой предварительной проработки программы, а в ряде случаев и организации пробного обследования (см. определение численности выборки).

Основные этапы выборочного наблюдения:

– определение цели, задач и составление программы наблюдения;

– формирование выборки;

– проведение статистического наблюдения, то есть сбор данных по разработанной программе;

– анализ полученных результатов и расчет основных характеристик выборочной совокупности;

– расчет ошибки выборки и распространение ее результатов на генеральную совокупность.

<u>Генеральной совокупностью</u> называется вся изучаемая совокупность единиц. Например, при изучении общественного мнения студентов УО «ВГТУ» – это все студенты УО «ВГТУ».

<u>Численность генеральной совокупности</u> обозначается в статистике N.

<u>Выборочная совокупность</u> – это та часть единиц генеральной совокупности, которая подвергается выборочному обследованию.

<u>Численность выборочной совокупности</u> обозначается n. Например, n при изучении общественного мнения – это число опрошенных студентов УО «ВГТУ».

Как уже отмечалось, задача выборочного наблюдения состоит в том, чтобы на основе изучения выборочной совокупности получить правильное представление о показателях генеральной совокупности.

В том случае, когда исследование проводится по <u>количественному признаку</u>, в качестве обобщающей характеристики совокупности применяется <u>среднее значение</u> (средняя величина, средний размер) <u>признака</u>.

Среднее значение варьирующего признака по всей совокупности называется <u>генеральной средней</u> и определяется как

$$\bar{x} = \frac{\sum xf}{\sum f} \text{ и при этом } \sum f = N. \tag{7.1}$$

Среднее значение признака у единиц, которые подверглись выборочному обследованию, называется выборочной средней и определяется по формуле

$$\tilde{x} = \frac{\sum xf_B}{\sum f_B}\text{, в свою очередь, } \sum f_B = n. \tag{7.2}$$

В данном случае задача выборочного наблюдения состоит в том, чтобы на основе выборочной средней $(\tilde{x})$ дать правильное представление о генеральной средней $(\bar{x})$.

Для характеристики совокупности по альтернативно варьирующему признаку в качестве обобщающего показателя используется доля (частость).

Доля определяется по генеральной совокупности и характеризует ту часть единиц генеральной совокупности, которая обладает признаком, интересующим исследователя. Обозначается латинской буквой p.

$$p = \frac{M}{N}, \tag{7.3}$$

где M – число единиц генеральной совокупности, обладающих интересующим исследователя признаком.

В свою очередь:

$q = 1 - p$ – это доля единиц, не обладающих данным признаком.

Выборочная доля, или доля в выборочной совокупности, называется частостью и определяется как отношение:

$$w = \frac{m}{n}, \tag{7.4}$$

где m – число единиц выборочной совокупности, обладающих интересующим исследователя признаком.

В случае альтернативного признака задача выборочного наблюдения состоит в том, чтобы на основе определения частости дать верное представление о доле в генеральной совокупности.

Кроме среднего размера признака и доли для характеристики генеральной и выборочной совокупности, могут быть использованы:

- генеральная и выборочная дисперсии;
- генеральное и выборочное среднее квадратическое отклонение;
- другие статистические характеристики.

Таблица 7.1 – Обобщающие характеристики генеральной и выборочной совокупности

Показатели	Совокупность	
	генеральная	выборочная
Объём (число единиц) совокупности	N	n
Среднее значение признака	$\bar{x}$	$\tilde{x}$
Число единиц, обладающих изучаемым признаком	M	m
Доля единиц, обладающих изучаемым признаком	p	w
Доля единиц, не обладающих изучаемым признаком	1-p (q)	1-w
Дисперсия	$\sigma_0^2 = \frac{\sum\left(x-\bar{x}\right)^2 * f}{\sum f}$	$\sigma_в^2 = \frac{\sum\left(x-\tilde{x}\right)^2 * f_в}{\sum f_в}$
Среднее квадратическое отклонение	$\sigma_0 = \sqrt{\frac{\sum\left(x-\bar{x}\right)^2 * f}{\sum f}}$	$\sigma_в = \sqrt{\frac{\sum\left(x-\tilde{x}\right)^2 * f_в}{\sum f_в}}$

Расхождения между генеральными и выборочными характеристиками изучаются на основе предельных теорем теории вероятности.

В статистике речь идёт о возможных ошибках выборки.

7.2 Отбор единиц из генеральной в выборочную совокупность

Для того, чтобы выборка была репрезентативной, необходима объективная система организации отбора единиц из генеральной совокупности.

Различают два вида выборочного наблюдения или отбора единиц:

- повторный отбор;
- бесповторный отбор.

Повторный отбор («схема возвратного шара») предполагает, что отобранная однажды единица возвращается обратно в генеральную совокупность и имеет равную с другими единицами возможность быть отобранной вновь. Объём генеральной совокупности во время отбора остаётся неизменным.

Бесповторный отбор («схема безвозвратного шара») подразумевает, что однажды отобранная единица не возвращается в генеральную совокупность. Поэтому численность генеральной совокупности всё время уменьшается, а следовательно, вероятность попадания оставшихся единиц в выборку всё время возрастает. Бесповторный отбор даёт более точные результаты и поэтому на практике находит более широкое применение, чем повторный.

При проведении выборочного наблюдения могут использоваться следующие способы отбора единиц из генеральной совокупности:

- индивидуальный отбор – в выборку отбираются отдельные единицы исследуемой совокупности;

- групповой отбор – в выборку попадают качественно однородные группы или серии изучаемых единиц;

- комбинированный отбор предполагает комбинацию (сочетание) индивидуального и группового отбора.

Для формирования выборочной совокупности возможны следующие методы отбора единиц:

- собственно-случайный (случайный);
- механический;
- типический;
- серийный;
- комбинированные наблюдения (многоступенчатый отбор, многофазный отбор);
- моментно-выборочное наблюдение (метод моментных наблюдений).

Случайный (собственно-случайный) отбор предполагает равную вероятность попадания единиц в выборочную совокупность. Он может проводиться при помощи жеребьёвки или таблицы случайных чисел.

Механический отбор заключается в отборе единиц из генеральной совокупности, производимом в каком-либо механическом порядке (например, в отборе каждой 5-й, каждой 10-й, 20-й и т. д. единицы). При этом для обеспечения репрезентативности выборки предполагается определённое расположение единиц генеральной совокупности (например, в алфавитном порядке либо в проранжированном по размеру признака виде и т. д.).

Промежуток, через который нужно отбирать единицы, или шаг отсчёта, зависит от пропорции отбора, исчисляемой делением численности генеральной совокупности на численность выборки:

$$N / n = 10.$$

Следовательно, например: 5-ая, 15-ая, 25-ая и т. д. – единицы.

Типический отбор имеет место при изучении неоднородных по исследуемым признакам совокупностей. Тогда перед производством выборки генеральная совокупность делится на группы, а затем механическим либо случайным методом отбираются единицы. Обычно из каждой группы берется определенный процент единиц (пропорциональный отбор), но иногда и одинаковое число единиц (непропорциональный отбор).

Типический отбор иногда называют расслоенным, или стратифицированным. Он даёт более точные результаты по сравнению с другими методами отбора.

Серийный отбор предполагает отбор из генеральной совокупности не отдельных единиц, а групп, которые принято называть сериями (гнёздами). Внутри отобранных серий проводится сплошное обследование всех единиц.

Например, при проведении 20 % статистического контроля качества отбираются пачки (транспортные партии) изделий по 10 ед. Сменный выпуск составляет 250 изделий. Тогда число серий в генеральной совокупности:

$$R = \frac{250}{10} = 25\,серий\,.$$

Число серий в выборочной совокупности можно определить двумя способами:

1) r = 25 серий*0,2 = 5 серий;

2) r = 250*0,2 / 10 = 5 серий.

Серийный отбор легче организовать и провести, однако он имеет существенный недостаток: не обеспечивается представительство каждой серии, а следовательно, не исключается влияние межсерийной вариации.

Все рассматриваемые методы формирования выборочной совокупности предполагают отбор единиц совокупности в выборку уже на первом этапе отбора. Такой отбор называется одноступенчатым.

На практике часто используется многоступенчатый отбор, при котором на первом этапе отбираются группы (серии), а на следующем − механическим или случайным способом из отобранных серий выбираются единицы наблюдения. Учитывая, что при построении многоступенчатой выборки происходит комбинация различных методов отбора, такую выборку иногда называют комбинированной.

К комбинированной выборке относят и многофазный отбор. Его особенностью является тот факт, что из числа единиц, отобранных на первом этапе, на следующих этапах отбирается всё меньше единиц, но расширяется программа наблюдения.

Основное отличие многофазного отбора от многоступенчатого состоит в том, что при многофазном отборе на каждом этапе единица отбора одна и та же, а при многоступенчатом − на каждом этапе своя единица отбора.

Особым методом выборочного наблюдения является метод моментных наблюдений (или моментно-выборочное наблюдение). Он используется достаточно широко в промышленности для изучения использования рабочего времени (исполнителей и оборудования).

Суть его заключается в том, что через определенные промежутки времени (в определенные моменты) происходит фиксация состояния изучаемой совокупности.

Например, по оборудованию может быть зафиксировано два состояния: а) работа; б) простой.

За смену по единице оборудования А зафиксировано 28 элементов «работа» и 4 элемента «простой».

Следовательно,

$$\frac{28}{28+4} = 0{,}875 \quad \textit{или} \quad *100 = 87{,}5\ \% - \textit{удельный вес элемента "работа"}.$$

$$\frac{4}{28+4} = 0{,}125 \quad \textit{или} \quad 12{,}5\ \% - \textit{удельный вес элемента "простой"}.$$

Так как в смене 8 часов, или 480 минут, то оборудование работало (480*0,875) 420 минут, или 7 часов, и простаивало (480*0,125) 60 минут, или 1 час.

Особенность такой выборки: выборка производится по времени. То есть генеральная совокупность – продолжительность смены, выборочная совокупность – совокупность отдельных моментных состояний объекта.

7.3 Ошибки выборочного наблюдения

Информация, получаемая в результате выборочного наблюдения, может иметь расхождение с реальной действительностью. Так как речь идет о варьирующих признаках и обследованию подвергается не вся совокупность, а только ее часть, можно с уверенностью утверждать, что статистические показатели, рассчитанные по выборке, не будут абсолютно совпадать с показателями генеральной совокупности.

Так, средняя величина признака в генеральной совокупности имеет всегда одно и то же значение. В то же время средняя, рассчитанная по выборке, будет колебаться по мере того, как будут меняться единицы, отобранные в выборку. То же можно утверждать и о доле и частости.

Следовательно, речь должна идти о том, чтобы:

- во-первых, максимально приблизить показатели выборки к показателям генеральной совокупности;
- во-вторых, знать возможные пределы их отклонений;
- в-третьих, знать условия, от которых зависит величина этих отклонений.

Расхождения между характеристиками выборочной совокупности и характеристиками генеральной совокупности носят название ошибки выборочного наблюдения.

Различают ошибки выборки и ошибки регистрации.

Ошибки выборки называют ошибками репрезентативности. Возникают они вследствие естественного расхождения характеристики выборочной и генеральной совокупности, носят случайный характер и с равной вероятностью могут либо увеличивать, либо уменьшать характеристики генеральной совокупности.

Различают ошибки выборки:

- средние (стандартные);
- предельные.

Средними ошибки называются потому, что они будут разные в зависимости от того, какие единицы попали в выборку, то есть речь идет о средней величине из возможных ошибок.

Средняя ошибка выборки (μ) зависит от: а) объема (численности) выборочной совокупности (чем больше n, тем меньше μ); б) степени вариации изучаемого признака (чем больше σ^2, тем больше μ); в) схемы отбора единиц из генеральной в выборочную совокупность.

Степень вариации признака в данном случае оценивается дисперсией σ^2. При проведении выборочного наблюдения обычно генеральная дисперсия неизвестна. Представляется возможным расчет лишь выборочной дисперсии.

Доказано, что при достаточно больших n, когда величина $\frac{n}{n-1}$ близка к 1, выборочная дисперсия приближенно равна генеральной:

$$\sigma_0^2 \approx \sigma_B^2 \quad при \quad \frac{n}{n-1} \approx 1.$$

При случайном повторном отборе величина средней ошибки рассчитывается по формуле

$$\mu = \sqrt{\frac{\sigma_0^2}{n}} \qquad (7.5)$$

Учитывая, что $\sigma_0^2 \approx \sigma_B^2$:

$$\mu = \sqrt{\frac{\sigma^2}{n}}\ , \qquad (7.6)$$

где σ^2 – выборочная дисперсия.

Тогда: а) для средней величины

$$\mu_x = \sqrt{\frac{\sigma_x^2}{n}} = \frac{\sqrt{\sum (x-\tilde{x})^2}}{n}\ , \qquad (7.7)$$

б) для доли

$$\mu_w = \sqrt{\frac{\sigma_w^2}{n}} = \sqrt{\frac{w(1-w)}{n}}\ . \qquad (7.8)$$

При случайном бесповторном отборе численность единиц генеральной совокупности уменьшается в процессе отбора. Следовательно, и вероятность ошибки уменьшается. Потому при исчислении средних ошибок для бесповторного отбора в формулы μ_x и μ_w вводится дополнительный множитель $(1-\frac{n}{N})$ ($\frac{n}{N}$ – доля отобранных единиц из генеральной совокупности).

Величина средней ошибки в этом случае определяется по формулам:

а) для средней величины

$$\mu_x = \sqrt{\frac{\sigma_x^2}{n} * \left(1-\frac{n}{N}\right)}; \qquad (7.9)$$

б) для доли

$$\mu_w = \sqrt{\frac{w(1-w)}{n} * \left(1-\frac{n}{N}\right)}. \qquad (7.10)$$

Так как $n < N$, выражение $\left(1-\frac{n}{N}\right) \langle 1$, ошибки при бесповторном отборе будут меньше, чем при повторном.

При механическом отборе средняя ошибка определяется по формуле случайного бесповторного отбора.

В случае типического отбора в качестве показателя вариации выступает средняя из внутригрупповых дисперсий (см. раздел 6.3). Поэтому средняя ошибка выборки при типическом повторном отборе:

а) для средней величины

$$\mu_x = \sqrt{\frac{\overline{\sigma^2_{гр}}}{n}}\ ; \qquad \overline{\sigma^2_{гр}} = \frac{\sum \sigma^2_{гр} * f_{гр}}{\sum f_{гр}}, \tag{7.11}$$

где $f_{гр}$ – число единиц в изучаемой группе;

б) для доли

$$\mu_w = \sqrt{\frac{\overline{w(1-w)}}{n}}$$
$$\overline{w(1-w)} = \frac{\sum \left[w_{гр}\left(1-w_{гр}\right)\right] * f_{гр}}{\sum f_{гр}}. \tag{7.12}$$

В случае бесповторного отбора также добавляется множитель $\left(1-\frac{n}{N}\right)$, то есть при <u>типическом бесповторном отборе</u> средняя ошибка рассчитывается:

а) для средней величины

$$\mu_x = \sqrt{\frac{\overline{\sigma_{гр}}^2}{n}\,(1-\frac{n}{N})}\ ; \tag{7.13}$$

б) для доли

$$\mu_w = \sqrt{\frac{\overline{W(1-W)}}{n}\,(1-\frac{n}{N})}\ . \tag{7.14}$$

При <u>серийном отборе</u> оценка вариации признака производится по межсерийной дисперсии δ^2 (см. раздел 6.3), а численность выборки характеризуется числом отобранных серий – *r*.

Тогда для <u>повторного серийного отбора</u> средняя ошибка определяется как

$$\mu = \sqrt{\frac{\delta_в{}^2}{r}}\ , \tag{7.15}$$

а для бесповторного

$$\mu = \sqrt{\frac{\delta_в{}^2}{r}\,(1-\frac{r}{R})}\ . \tag{7.16}$$

Зная среднюю величину (средний размер) признака в выборке и среднюю ошибку, можно записать пределы (границы) генеральной средней:

$$\tilde{x} - \mu_x < \bar{x} < \tilde{x} + \mu_x . \quad (7.17)$$

Аналогично для доли:

$$w - \mu_w < p < w + \mu_w \ . \quad (7.18)$$

Вместе с тем, утверждать, что генеральная средняя или доля не выйдут за указанные пределы, можно только с определённой степенью вероятности − 0,683. Это означает, что если в генеральной совокупности 1000 единиц, то 683 из них будут находиться в указанных пределах, а 317 могут выходить за эти пределы. Следовательно, оценка генеральной совокупности по $\bar{x} = \tilde{x} \pm \mu_x$ или $p = w \pm \mu_w$ является довольно приблизительной, грубой.

На практике чаще всего требуется получение более точного результата. Для того, чтобы повысить вероятность гарантии пределов характеристик выборки, прибегают к исчислению не средних, а предельных ошибок. Известные математики П.Л. Чебышев и А.М. Ляпунов предложили для повышения вероятности невыхода значений генеральной совокупности за пределы характеристик выборки удвоить или утроить среднюю ошибки. То есть μ_x – средняя ошибка, то предельная ошибка:

$$\Delta_x = t * \mu_x , \quad (7.19)$$

где t – коэффициент доверия.

t = 1,2,3 (чаще всего, хотя могут быть другие промежуточные значения).

Аналогично для доли:

$$\Delta_w = t * \mu_w . \quad (7.20)$$

Формулы определения предельных ошибок зависят от способа отбора единиц в выборку.

Таблица 7.2 – Предельная ошибка выборки для различных способов отбора

Метод отбора	***Вид отбора***			
	Повторный		Бесповторный	
	Для средней	Для доли	Для средней	Для доли
Случайный и механический	$t * \sqrt{\frac{\sigma^2}{n}}$	$t * \sqrt{\frac{w(1-w)}{n}}$	$t * \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$	$t * \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)}$
Типический	$t * \sqrt{\frac{\overline{\sigma_i^2}}{n}}$	$t * \sqrt{\frac{\overline{w_i(1-w)}}{n}}$	$t * \sqrt{\frac{\overline{\sigma_i^2}}{n}\left(1 - \frac{n}{N}\right)}$	$t * \sqrt{\frac{\overline{w_i(1-w_i)}}{n}\left(1 - \frac{n}{N}\right)}$
Серийный	$t * \sqrt{\frac{\delta_x^2}{r}}$	$t * \sqrt{\frac{\delta_w^2}{r}}$	$t * \sqrt{\frac{\delta_x^2}{r}\left(1 - \frac{r}{R}\right)}$	$t * \sqrt{\frac{\delta_w^2}{r}\left(1 - \frac{r}{R}\right)}$

Условные обозначения, принятые в формулах:

t – коэффициент доверия;

σ^2 – дисперсия признака в выборочной совокупности;

δ^2 – межсерийная дисперсия;

r – число отобранных серий;

R – число серий в генеральной совокупности;

n – число отобранных единиц;

N – число единиц в генеральной совокупности.

Коэффициент доверия определяет вероятность, с которой можно утверждать, что максимальная ошибка выборки не превысит величины $t*\mu$.

При этом для t = 1 эта вероятность − 0,683; для t = 2 вероятность − 0,954; а для t = 3 вероятность − 0,997.

То есть с вероятностью 0,997 можно утверждать, что

$$\tilde{x} - 3\mu_x < \bar{x} < \tilde{x} + 3\mu_x. \quad (7.21)$$

После того, как рассчитаны показатели выборки, они распространяются на характеристики генеральной совокупности с помощью предельных ошибок:

$$\tilde{x} - \Delta_x < \bar{x} < \tilde{x} + \Delta_x \text{ или } w - \Delta_w < p < w + \Delta_w. \quad (7.22)$$

Предельные ошибки являются абсолютными величинами. Но на их основе могут быть рассчитаны и предельные относительные ошибки:

$$\Delta_X(\%) = \frac{\Delta_X}{\tilde{x}} * 100 \text{ или } \Delta_w(\%) = \frac{\Delta_w}{w} * 100. \quad (7.23)$$

К примеру, необходимо определить с вероятностью 0,954, в каких пределах находится удельный вес нестандартной продукции в партии изделий в 1000 единиц, если в отобранных случайным повторным методом 40 единицах оказалось 8 нестандартных.

1. Определяем частость:

$$w = \frac{m}{n} = \frac{8}{40} = 0,2.$$

2. Определяем дисперсию:

w – частость нестандартной продукции, 1-w − частость стандартной продукции.

$$\sigma_w^{\ 2} = 0,2 * 0,8 = 0,16.$$

3. Определяем среднюю ошибку:

$$\mu_w = \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0,16}{40}} = 0,063.$$

4. Определяем предельную ошибку, учитывая что, вероятность 0,954, следовательно, t = 2:

$$\Delta_w = t * \mu_w = 2 * 0{,}063 = 0{,}126.$$

5. С вероятностью 0,954 можно утверждать, что доля нестандартной продукции в исследуемой партии находится в пределах:

$$0{,}2 - 0{,}126 < p < 0{,}2 + 0{,}126$$

$$0{,}074 < p < 0{,}326,$$

от 7,4 % до 32,6 %

7.4 Определение численности выборки

Так как величина ошибок выборочного наблюдения зависит от объёма выборки, то на стадии организации выборочного наблюдения необходимо решить вопрос о том, каким должен быть объём выборки, чтобы была обеспечена требуемая точность результатов.

Формулы для определения необходимой численности выборки выводятся из формул предельных ошибок.

Так, для случайного повторного отбора:

$$\Delta = t * \mu = t * \sqrt{\frac{\sigma^2}{n}}, \text{ следовательно, } \Delta^2 = t^2 * \frac{\sigma^2}{n}.$$

Отсюда получаем $$n = \frac{t^2 \sigma^2}{\Delta^2}. \quad (7.24)$$

В случае количественного признака (то есть для средней величины):

$$n = \frac{t^2 \sigma_x^2}{\Delta_x^2}. \quad (7.25)$$

В случае качественного признака (для доли):

$$n = \frac{t^2 \sigma_w^2}{\Delta_w^2}. \quad (7.26)$$

В случае, когда t = 1, формулы принимают вид:

$$n = \frac{\sigma^2}{\mu^2}. \quad (7.27)$$

Для случайного бесповторного отбора численность выборки составит соответственно:

- для количественного признака (для средней величины):

$$n = \frac{t^2 \sigma_x^2 N}{\Delta_x^2 N + t^2 \sigma_x^2}; \quad (7.28)$$

- для качественного признака (для доли):

$$n = \frac{t^2 \sigma_w^2 N}{\Delta_w^2 N + t^2 \sigma_w^2} \qquad \sigma_w^2 = w(1 - w). \quad (7.29)$$

Аналогичным образом определяют и объем выборок, формируемых по другим схемам.

Математическая статистика разработала следующие формулы для определения необходимой численности выборки.

Таблица 7.3 – Необходимый объем выборки для различных способов формирования выборочной совокупности

Метод отбора	***Вид отбора***			
	Повторный		Бесповторный	
	Для средней	Для доли	Для средней	Для доли
Случайный и механический	$n = \frac{t^2\sigma^2}{\Delta_x^2}$	$n = \frac{t^2 w(1-w)}{\Delta_w^2}$	$n = \frac{t^2\sigma^2 N}{\Delta_x^2 N + t^2\sigma^2}$	$n = \frac{t^2 w(1-w)N}{\Delta_w^2 N + t^2 w(1-w)}$
Типический	$n = \frac{t^2\overline{\sigma}^2}{\Delta_x^2}$	$n = \frac{t^2\overline{w(1-w)}}{\Delta_w^2}$	$n = \frac{t^2\overline{\sigma}^2 N}{\Delta_x^2 N + t^2\overline{\sigma}^2}$	$n = \frac{t^2\overline{w(1-w)}N}{\Delta_w^2 N + t^2\overline{w(1-w)}}$
Серийный	$r = \frac{t^2\delta^2}{\Delta_x^2}$	$r = \frac{t^2 w_R(1-w_R)}{\Delta_x^2}$	$r = \frac{t^2\delta^2 R}{\Delta_x^2 R + t^2\delta^2}$	$r = \frac{t^2 w_R(1-w_R)R}{\Delta_w^2 R + t^2 w_R(1-w_R)}$

Следовательно, задаваясь величиной ошибки в абсолютном выражении, исследователь может определить численность единиц, которые нужно отобрать в выборку. Проблемным остается определение σ^2. Как правило, это происходит с помощью пробного обследования (обычно небольшого объема единиц).

7.5 Малая выборка и сфера ее применения

Несмотря на то, что уменьшение объема выборки сопровождается ростом величины стандартной ошибки, на практике часто приходится ограничиваться малым числом наблюдений. Чаще всего эта необходимость возникает при проверке качества продукции, связанной с ее уничтожением (так называемый разрушающий контроль качества продукции). Например, при проверке ткани на разрыв.

В этих случаях ограничиваются малыми выборками.

Под малой выборкой понимается такое выборочное наблюдение, численность единиц которого не превышает 30 единиц.

Первые работы в теории малых выборок были проведены Стьюдентом и продолжены Фишером. Доказано, что и при малых выборках характеристики выборочной совокупности могут быть перенесены на генеральную совокупность. Однако расчет средней и предельной ошибок в случае малой выборки имеет свои особенности.

В математической статистике доказано, что соотношение между выборочной и генеральной дисперсией

$$\sigma_0^2 = \sigma_B^2\left(\frac{n}{n-1}\right). \tag{7.30}$$

В случае достаточно большого объема выборки сомножителем $\frac{n}{n-1}$ пренебрегают, так как $\frac{n}{n-1} \approx 1$. Однако в случае малой выборки этого делать нельзя. Тогда формула средней ошибки малой выборки будет иметь вид:

$$\mu_{М.В.} = \sqrt{\frac{\sigma^2_{М.В.}}{n-1}}. \tag{7.31}$$

Предельная ошибка малой выборки

$$\Delta_{М.В.} = t\mu_{М.В.}. \tag{7.33}$$

Однако величина *t* иначе связана с вероятностной оценкой, чем при обычной выборке. В данном случае вероятностная оценка зависит не только от *t*, но и от *n*.

Таблица 7.4 – Распределение вероятности ошибок в зависимости от коэффициента доверия и численности малых выборок (фрагмент)

n / *t*	5	…	10	…	20	…	Обычная выборка
…							
1	0,626		0,657		0,670		0,683
…							
2	0,884		0,924		0,940		0,954
…							
3	0,960		0,984		0,992		0,997

Как и в случае обычной выборки, на заключительном этапе определяются доверительные интервалы, в которых может находиться генеральная средняя

$$\tilde{x} - \Delta_{М.В.} < \bar{x} < \tilde{x} + \Delta_{М.В.} \tag{7.34}$$

или генеральная доля

$$w - \Delta_{М.В.} < \rho < w + \Delta_{М.В.}. \tag{7.35}$$

8 СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ДИНАМИКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

8.1 Ряды динамики, их виды и правила построения

Ряд динамики – это ряд последовательно расположенных в хронологическом порядке показателей, которые характеризуют развитие явлений во времени.

Каждый ряд динамики (или динамический ряд, или временной ряд) состоит из двух элементов:

t – показатели времени;

y – соответствующие им уровни развития экономического явления. Их называют уровнями ряда динамики.

Исследование рядов динамики имеет большое значение для процесса познания, так как они дают возможность выявить закономерность в изменении во времени того или иного общественного явления.

С точки зрения характеристики развития явления во времени ряды динамики делятся на:

- моментные;
- интервальные.

Моментные ряды динамики отображают состояние изучаемых явлений на определённые моменты времени (в большинстве случаев− на дату (таблица 8.1)).

Таблица 8.1 – Динамика численности работников (данные условные)

Дата	*t*	01.09.2010	01.09.2011	01.09.2012	01.09.2013
Списочная численность студентов заочного факультета, чел.	*y*	1020	1184	1050	1010

Интервальные ряды динамики отображают состояние (развитие) изучаемых явлений за отдельные интервалы времени (год, квартал, месяц, декаду и т. п. (таблица 8.2).

Таблица 8.2 – Динамика выпуска продукции, млн. руб. (данные условные)

Квартал	t	I	II	III	IV
Выпуск продукции, млн. руб.	y	2800	3780	2820	3900

В зависимости от приводимых в них обобщающих показателей ряды динамики можно подразделить на:

- ряды динамики абсолютных величин;
- ряды динамики относительных величин.
- ряды динамики средних величин.

Основополагающими в этой системе являются ряды динамики абсолютных величин, так как на их основе составляются ряды динамики относительных и средних величин. В свою очередь, ряды динамики относительных и средних величин позволяют охарактеризовать качественные сдвиги в развитии явлений.

Важнейшей проблемой при построении рядов динамики является проблема сопоставимости уровней ряда. Для обеспечения этой сопоставимости при построении рядов динамики необходимо соблюдать следующие правила:

1) все показатели рядов динамики должны быть исчислены в одних и тех же единицах измерения;

2) все показатели рядов динамики должны быть исчислены по единой методологии;

3) все показатели рядов динамики должны быть исчислены в одних и тех же территориальных границах;

4) все показатели рядов динамики должны быть исчислены относительно одного и того же круга объектов (или объекта);

5) если уровни рядов динамики имеют стоимостную оценку, то цены должны быть сопоставимыми, то есть едиными.

8.2 Показатели динамики

8.2.1 Аналитические показатели ряда динамики

Аналитические показатели динамики получают в результате сопоставления уровней рядов динамики. Они могут быть определены цепным и базисным способами. При цепном способе каждый уровень ряда динамики сопоставляется с предыдущим, а при базисном способе каждый уровень сопоставляется с одним и тем же уровнем, принятым за базу сравнения (как правило, первым).

К числу важнейших аналитических показателей относят:

- абсолютный прирост;
- темп роста;
- темп прироста;
- вес (абсолютное значение) 1 % прироста;

иногда к ним добавляют:

- ускорение;
- коэффициент опережения;

- темп наращивания.

Абсолютный прирост показывает, на сколько (в единицах измерения уровней ряда) уровень одного периода больше или меньше уровня, принятого за базу сравнения. В зависимости от базы сравнения абсолютные приросты могут быть:

- цепные;
- базисные.

Цепной абсолютный прирост:

$$\Delta y_{ц} = y_i - y_{i-1} \ . \tag{8.1}$$

Базисный абсолютный прирост:

$$\Delta y_{б} = y_i - y_0 \ . \tag{8.2}$$

Если значения $\Delta y_{ц}$ постоянны, то уровни ряда изменяются равномерно. Если же $\Delta y_{ц}$ увеличиваются или уменьшаются, это означает, что развитие явления ускоряется или замедляется. Тогда имеет смысл рассчитывать показатель ускорения:

$$\Delta_{\Delta} = \Delta y_i - \Delta y_{i-1} \, . \tag{8.3}$$

Темп роста показывает, во сколько раз данный уровень ряда динамики больше (меньше) принятого за базу сравнения. Он может выражаться в виде коэффициента либо в процентах.

Цепной темп роста:

$$T_{рц} = \frac{y_i}{y_{i-1}} (*100\%) \, . \tag{8.4}$$

Базисный темп роста:

$$T_{рб} = \frac{y_i}{y_0} (*100\%) \, . \tag{8.5}$$

Между цепными и базисными темпами роста существует связь, которая позволяет при необходимости переходить от цепных к базисным и наоборот:

- произведение цепных темпов роста (коэффициентов) равно базисному;
- отношение базисного темпа роста (коэффициента) *i*-ого периода к базисному темпу роста (коэффициенту) (*i*–1)-ого периода равно цепному темпу роста *i*-го периода.

Иногда на практике приходится сравнивать темпы роста разных показателей, относящихся к одной и той же совокупности, либо одного и того же показателя, но рассчитанного в разных совокупностях. В таких случаях производится сопоставление темпов роста двух показателей, а результат этого сопоставления называют коэффициент опережения.

Темп прироста показывает, на сколько процентов данный уровень РД больше либо меньше принятого за базу сравнения. Темп прироста – это отношение абсолютного прироста к сравниваемому уровню (к базе сравнения):

- цепной темп прироста:

$$T_{пр\ ц} = \frac{\Delta y_ц}{y_{i-1}} \ (*100\%) \ \frac{y_1 - y_0}{y_0}; \frac{y_2 - y_1}{y_1}; \frac{y_3 - y_2}{y_2} \quad \text{и т. д.} \qquad (8.6)$$

- базисный темп прироста:

$$T_{пр\ б} = \frac{\Delta y_б}{y_0} \ (*100\%) \ \frac{y_1 - y_0}{y_0}; \frac{y_2 - y_0}{y_0}; \frac{y_3 - y_0}{y_0} \quad \text{и т. д.} \qquad (8.7)$$

На практике темп прироста чаще всего рассчитывают исходя из взаимосвязи между показателями темпов роста и темпов прироста:

$$\begin{aligned} &Тпр = Тр - 1 \ (коэффициент) \\ &либо \\ &Тпр = Тр - 100 \ (\%) \end{aligned} \qquad (8.8)$$

Абсолютное значение одного процента прироста (или вес 1% прироста) показывает, насколько весом каждый % прироста, какое содержание или какая абсолютная величина за ним скрывается:

$$Абс.зн.1\% = \frac{\Delta y}{Тпр\,(\%)} \ . \qquad (8.9)$$

Абсолютное значение 1 % прироста имеет смысл лишь для цепных показателей.

Темп наращивания исчисляется как отношение цепных абсолютных приростов к уровню, принятому за базисный:

$$Тн = \frac{\Delta y_ц}{y_0} \qquad \frac{y_1 - y_0}{y_0} \quad \frac{y_2 - y_1}{y_0} \quad \text{и т. д.} \qquad (8.10)$$

Как и ускорение, представляет интерес в том случае, если абсолютные приросты от одного периода к другому возрастают.

Таблица 8.3 – Пример расчета аналитических показателей динамики

Наименование показателей	Ед.изм.	Уровни показателей по годам			
		2009	2010	2011	2012
Прибыль	млрд. руб.	380	390	420	500
Аналитические показатели динамики:					
а) абсолютный прирост: - цепной - базисный	млрд. руб.	- -	10 10	30 40	80 120
б) темп роста: - цепной	коэф-т %	- -	1,026 102,6	1,077 107,7	1,190 119,0
- базисный	коэф-т %	- -	1,026 102,6	1,105 110,5	1,316 131,6
в) темп прироста: - цепной	коэф-т %	- -	0,026 2,6	0,077 7,7	0,190 19
- базисный	коэф-т %	- -	0,026 2,6	0,105 10,5	0,316 31,6
г) абсолютное значение одного процента прироста:	млрд. руб.	-	3,8	3,9	4,2
д) темп наращивания	коэф-т	-	0,026	0,079	0,210
	%	-	2,6	7,9	21,0
е) ускорение	млрд. руб.	-	-	20	50

8.2.2 Средние показатели ряда динамики

Для сравнения изменений того или иного показателя в разные периоды, в разных странах и т. п. необходимы обобщающие показатели в виде средних величин. Такими обобщающими характеристиками в рядах динамики являются:

- средний уровень ряда динамики;
- средний абсолютный прирост;
- средний темп роста,
- средний темп прироста.

Средний уровень ряда динамики рассчитывается неодинаково для различных видов рядов динамики. Кроме того, в исчислении средних величин по рядам динамики большое значение играет равенство (либо неравенство) промежутков времени между соседними уровнями.

Таблица 8.4 – Формулы определения среднего уровня ряда динамики

Виды рядов динамики		Промежутки времени между соседними уровнями	
		равные	неравные
Ряды динамики абсолютных и средних величин	интервальные	Средняя арифметическая простая $\bar{y} = \frac{\sum y_i}{n}$, n – число уровней	Средняя арифметическая взвешенная $\bar{y} = \frac{\sum y_i * t_i}{\sum t_i}$, t_i – интервал между соседними уровнями
	моментные	Средняя хронологическая простая $\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}$	Средняя хронологическая взвешенная $\bar{y} = \frac{\sum \frac{y_i + y_{i+1}}{2} * t_i}{\sum t_i}$

В нашем примере (таблица 8.3):

$$\bar{y} = \frac{380 + 390 + 420 + 500}{4} = 422{,}5 \text{ (млрд. руб.)}.$$

<u>Средний абсолютный прирост</u> рассчитывается как средняя арифметическая простая из абсолютных приростов (цепных):

$$\Delta\bar{y} = \frac{\sum \Delta y_ц}{n-1}, \tag{8.11}$$

где n – число уровней ряда динамики;

$n-1$ – число абсолютных приростов, которые могут быть получены по n уровням.

Либо учитывая накопление абсолютного прироста:

$$\sum \Delta y_ц = (y_1 - y_0) + (y_2 - y_1) + \ldots + (y_n - y_{n-1}) = y_n - y_0, \tag{8.12}$$

$$\Delta\bar{y} = \frac{y_n - y_0}{n-1}. \tag{8.13}$$

В нашем примере (таблица 8.3): $\Delta\bar{y} = \frac{10 + 30 + 80}{3} = 40$ *млрд. руб.*

Важную роль в анализе рядов динамики играет <u>средний темп роста</u>. Наиболее часто он рассчитывается как средняя геометрическая из цепных темпов роста:

$$\overline{T}p = \sqrt[n-1]{Tp_1 * Tp_2 * ... * Tp_{n-1}} \ . \qquad (8.14)$$

Используя выражения $Tp_i = \dfrac{y_i}{y_{i-1}}$, можно получить другую формулу:

$$\overline{T}p = \sqrt[n-1]{\frac{y_1}{y_0} * \frac{y_2}{y_1} * \frac{y_3}{y_2} * ... * \frac{y_n}{y_{n-1}}}, \text{ т.е. } \overline{T}p = \sqrt[n-1]{\frac{y_n}{y_0}} \ . \qquad (8.15)$$

В нашем примере (таблица 8.3): $\overline{T}p = \sqrt[3]{1{,}026 * 1{,}077 * 1{,}1190} = 1{,}095.$

Средний темп прироста определяется на основе взаимосвязи между темпами роста и прироста.

Если данные о средних темпах роста выражены в виде коэффициента:

$$\overline{T}np = \overline{T}p - 1 \ . \qquad (8.16)$$

1,095 – 1 = 0,095,

а если данные приводятся в процентах, то:

$$\overline{T}np = \overline{T}p - 100 \, . \qquad (8.17)$$

109,5 – 100 = 9,5 (%).

По средним показателям в нашем примере можно сделать следующие выводы:

а) размер среднегодовой прибыли за исследуемый период составляет 422,5 млрд. руб.;

б) из года в год прибыль увеличивается в 1,095 раза;

в) за каждый год прибыль возрастает в среднем на 40 млрд. руб. или на 9,5 %.

8.3 Приведение рядов динамики к единому основанию

Чаще всего этот метод на практике применяется при сравнительном анализе тенденций развития взаимосвязанных явлений:

- заработная плата и производительность труда по предприятию, отрасли и т. д.;
- фондовооружённость и производительность труда;
- выручка от реализации и прибыль от реализации и т. д.

Разные размерности и исходные уровни таких показателей не позволяют определить, уровень какого из них изменяется быстрее.

Например, необходимо сравнить динамику заработной платы и производительности труда в организации.

Таблица 8.5 – Динамика заработной платы и производительности труда рабочих

Квартал	I	II	III	IV
Среднемесячная заработная плата 1 рабочего, тыс. руб.	7380	7500	7800	8200
Среднемесячная производительность труда 1 рабочего, шт.	1540	1560	1560	1575

Проведение сравнительного анализа значительно облегчается, если рассматриваемые ряды динамики приведены к единому основанию, то есть их уровни выражены в процентах к начальному уровню (возможно, к среднему или другому характерному уровню).

Приняв в качестве базы сравнения (то есть за 100 %) начальный уровень ряда динамики, исходные ряды преобразуем в ряды базисных темпов роста:

Таблица 8.6 – Базисные темпы роста заработной платы и производительности труда рабочих

Квартал	I	II	III	IV
Темп роста заработной платы, %	100,0	101,6	105,7	111,1
Темп роста производительности труда, %	100,0	101,3	101,3	102,3

То есть заработная плата растёт быстрее, чем производительность труда. Для сравнения может быть рассчитан коэффициент опережения:

$$K_{оп.} = \frac{111{,}1}{102{,}3} = 1{,}086\,.$$

Этот же приём, то есть приведение рядов к единому основанию, может быть использован при сравнении динамики одного и того же показателя, но в разных совокупностях.

Особое внимание при использовании метода приведения рядов динамики к единому основанию уделяется выбору базы сравнения (или основанию), так как она оказывает весьма существенное влияние на результаты исследования.

Для наглядности часто строится график.

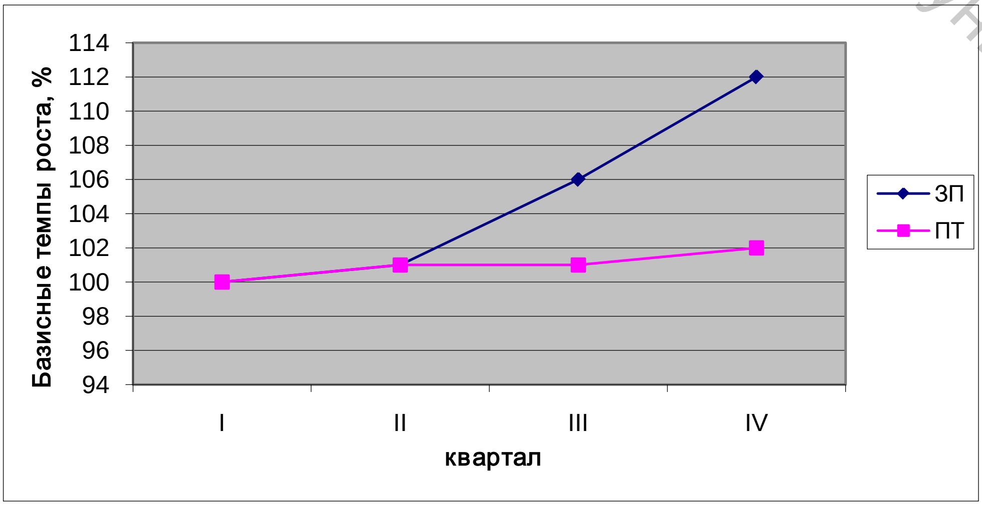

Рисунок 8.1 – Темпы роста заработной платы и производительности труда рабочих

8.4 Метод аналитического выравнивания рядов динамики

Этот метод изучения закономерностей в рядах динамики нашёл наиболее широкое применение на практике, так как он имеет существенное преимущество: он позволяет приближенно выразить определённым математическим законом развитие явления, то есть получить математическое описание этого развития в виде функции:

$$y = f(t). \tag{8.18}$$

По экономическому смыслу и возможности последующей интерпретации математических расчётов рекомендуется выделять <u>4 типа развития явления во времени:</u>

1) <u>равномерное развитие</u>, то есть в арифметической прогрессии (с постоянным абсолютным приростом):

$$y_t = a_0 + a_1 t \ \text{– прямая,} \tag{8.19}$$

где y_t – теоретические уровни ряда динамики;

a_0 – начальный уровень явления;

a_1 – абсолютное изменение явления за единицу времени (скорость ряда динамики).

При этом $a_1 < 0$ – тенденция уменьшения, а $a_1 > 0$ – тенденция роста.

2) <u>равноускоренное (или равнозамедленное) развитие</u>, то есть движение с постоянным во времени ускорением (замедлением):

$$y_t = a_0 + a_1 t + a_2 t^2 \ \text{– парабола,} \tag{8.20}$$

где a_2 – величина постоянного изменения скорости в единицу времени (то есть величина ускорения или замедления).

Если $a_2 < 0$ – замедление, а $a_2 > 0$ – ускорение.

3) <u>развитие с переменным ускорением (замедлением):</u>

$$y_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \ \text{– кубическая парабола,} \tag{8.21}$$

где a_3 характеризует эффект возрастания ускорения ($a_3 > 0$) или его замедление, ($a_3 < 0$) во времени.

4) <u>развитие по экспоненциальному закону</u> с постоянным темпом роста, то есть в геометрической прогрессии:

$$y_t = a_0 * T_p^t, \quad (8.22)$$

где T_p – темп роста (снижения) в единицу времени.

Выбор типа развития осуществляется по минимальной величине средней квадратической ошибки аппроксимации:

$$\sigma_y = \sqrt{\frac{\sum (y_t - y)^2}{n}} \quad . \quad (8.23)$$

Параметры уравнений определяются методом наименьших квадратов. Имеются стандартные программы для расчётов на ЭВМ. При ручном счёте для облегчения расчётов вводятся такие обозначения периодов или моментов времени, чтобы $\sum t = 0$, следовательно, $\sum t^3 = 0; \sum t^5 = 0$.

Рассмотрим получение уравнения тренда на примере прямой:

$$y_t = a_0 + a_1 t. \quad (8.24)$$

Нахождение параметров a_0 и a_1 основано на использовании известного в математике метода наименьших квадратов, согласно которому расчёт a_0 и a_1 сводится к решению системы уравнений:

$$\begin{cases} a_0 n + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum yt \end{cases}. \quad (8.25)$$

Для упрощения расчётов параметру t придаются такие значения, чтобы $\sum t = 0$ (способ «отсчёта от условного 0»).

Для ряда динамики, состоящего из нечетного количества уровней:

Квартал	I	II	III	IV
t	-3	-1	1	3

Для ряда динамики, состоящего из четного количества уровней:

Месяц	01	02	03	04	05	06
t	-5	-3	-1	1	3	5

При $\sum t = 0$ система уравнений принимает вид:

$$\begin{cases} a_0 n = \sum y \\ a_1 \sum t^2 = \sum yt \end{cases} . \quad (8.26)$$

Отсюда легко определить:

$$a_0 = \frac{\sum y}{n} \text{ и } a_1 = \frac{\sum yt}{\sum t^2} \quad . \qquad (8.27, 8.28)$$

Однако, учитывая тот факт, что «условный 0» – то середина ряда динамики, и анализируя формулу расчёта a_0, приходим к выводу, что a_0 – средний (серединный) уровень ряда динамики.

Например, по данным таблицы 8.7 определить уравнение тренда, характеризующее тенденцию изменения текучести кадров за 5 лет.

Таблица 8.7 – Расчет параметров уравнения тренда

Годы	Прибыль, млрд. руб.		t	$y*t$	t^2	$y_t = 8{,}6 - 1{,}3t$	
2008 2009 2010 2011 2012	Эмпирические уровни	12 9 8 8 6	-2 -1 0 1 2	-24 -9 0 8 12	4 1 0 1 4	Теоретические уровни	11,2 9,9 8,6 7,3 6,0
	$\Sigma y = 43$		$\Sigma t = 0$	$\Sigma yt = 56$	$\Sigma t^2 = 10$	$\Sigma y_t = 43$	

Уравнение прямой $y_t = a_0 + a_1 t$

При этом: $$a_0 = \frac{\sum y}{n} = \frac{43}{5} = 8{,}6,$$

$$a_1 = \frac{\sum yt}{\sum t^2} = \frac{-13}{10} = -1{,}3.$$

Тогда уравнение тренда принимает вид:

$$y_t = 8{,}6 - 1{,}3\,t.$$

По уравнению определяем теоретические уровни и анализируем расхождение $\sum y$ и $\sum y_t$, оно не должно превышать 1 %.

Экономическая интерпретация тренда: за анализируемый период среднегодовой размер текучести кадров составил 8,6 %. Наблюдается тенденция снижения текучести в среднем за год на 1,3 п.п.

Несколько по иному дается интерпретация тренда, рассчитанного для ряда динамики, состоящего из четного количества уровней.

Например, по данным таблицы 8.8 определить уравнение тренда для характеристики динамики прибыли за 4 года.

Таблица 8.8 – Расчет параметров уравнения тренда

Годы	Прибыль, млрд. руб.	t	$y*t$	t^2	$y_t = 422{,}5 + 19{,}5t$
2009	380	-3	-1140	9	364
2010	390	-1	-390	1	403
2011	420	1	420	1	442
2012	500	3	1500	9	481
	$\Sigma y = 1690$		$\Sigma yt = 390$	$\Sigma t^2 = 20$	$\Sigma y_t = 1690$

$$a_0 = \frac{\sum y}{n} = \frac{1690}{4} = -422{,}5$$

$$a_1 = \frac{\sum yt}{\sum t^2} = \frac{390}{20} = 19{,}5$$

y_t = 422,5 + 19,5*t – уравнение тренда.

Экономическая интерпретация тренда: среднегодовой размер прибыли составляет 422,5 млрд. руб., однако наблюдается снижение уровня прибыли со среднегодовым уменьшением 39 млрд. руб. (19,5 * 2).

8.5 Сезонные колебания в рядах динамики и методы их измерения

Сезонными колебаниями называются внутригодичные постоянно повторяющиеся изменения уровней изучаемых явлений.

То есть необходимость исследования сезонности возникает при анализе рядов внутригодовой динамики, характеризующих развитие явлений по месяцам.

Внутригодовые колебания явлений наблюдаются во многих сферах деятельности:

- сельское хозяйство,
- пассажирский транспорт,
- торговля,
- бытовое обслуживание населения и т. д.

Существуют различные методы исследования сезонных колебаний в рядах динамики для тех случаев, когда информация приводится за несколько лет (например, аналитическое выравнивание по ряду Фурье).

Наиболее простой метод – построение сезонной волны.

Например, имеются сведения о количестве пассажиров маршрутных такси, следующих рейсом «Витебск – Могилев» за 3 года (таблица 8.9):

Таблица 8.9 – Динамика пассажирских перевозок маршрутными такси, следующими рейсом «Витебск − Могилев»

Месяц	Количество пассажиров, чел.			В среднем за три года, $\overline{y_i}$	Индекс сезонности, $i_{сез.} = \frac{\overline{y_i}}{\overline{y}} * 100$, %
	2010 год	2011 год	2012 год		
01	743	732	772	749	89,3
02	784	828	751	788	93,9
03	793	834	765	797	95,0
04	809	835	844	829	98,8
05	811	854	836	834	99,4
06	1029	1084	1100	1071	127,7
07	1010	924	1008	981	116,9
08	833	840	876	848	101,1
09	857	859	789	835	99,5

10	813	750	826	806	96,1
11	767	782	804	784	93,5
12	731	738	768	744	88,7
В среднем за год	834	838	844	$\bar{y} = 839$	

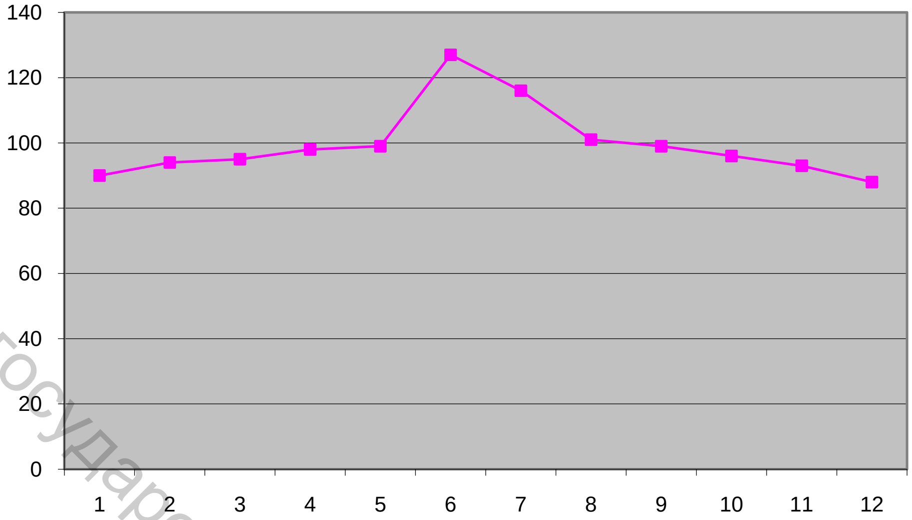

Рисунок 8.2 – Сезонная волна, характеризующая динамику количества пассажиров

Для оценки сезонных колебаний могут использоваться следующие показатели:

1) размах сезонных колебаний:

$$R_{сез} = y_{\max} - y_{\min};$$

2) коэффициент сезонных колебаний:

$$K_{сез} = \frac{y_{\max}}{y_{\min}};$$

3) индексы сезонных колебаний (они могут рассчитываться различными способами):

$$i_{сее} = \frac{y}{y_{\min}},$$

$$либо\ \ i_{сез} = \frac{y}{y_{\max}},$$

$$либо\ i_{сез} = \frac{y}{\bar{y}},$$

где $\bar{y}$ – средняя за 12 месяцев, или скользящая средняя.

4) среднее линейное отклонение:

$$\bar{d} = \frac{\sum(y-\bar{y})}{12};$$

5) среднее квадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum(y-\bar{y})^2}{12}}$$

или

$$\sigma = \sqrt{\frac{\sum i_{сез}^2}{12}};$$

6) дисперсия:

$$\sigma^2 = \frac{\sum(y-\bar{y})^2}{12}$$

или

$$\sigma^2 = \frac{\sum(i_{сез}-100)^2}{12};$$

7) коэффициент вариации:

$$V = \frac{\sigma}{\bar{y}}*100.$$

8.6 Экстраполяция и интерполяция в рядах динамики

Экстраполяция – это распространение выявленных при анализе рядов динамики закономерностей развития явления в будущее. В результате продолжения выявленных закономерностей в будущее определяются значения уровней явления на перспективу. Но могут определяться и прошлые значения – ретроспективная экстраполяция.

Экстраполируемый уровень ряда динамики может рассчитываться:

а) при помощи среднего абсолютного прироста:

$$y_{n+t} = y_n + \overline{\Delta y}*t.$$

где y_n – последний известный уровень ряда динамики;

t – количество периодов времени от уровня y_n до экстраполируемого уровня.

б) при помощи среднего темпа роста:

$$y_{n+t} = y_n * \overline{Tp}^t;$$

в) по уравнению тренда, подставив значение t для экстраполируемого уровня исходя из обозначений этого параметра при построении уравнения тренда для исходного ряда динамики.

Таблица 8.10 – Расчет параметров уравнения тренда

	2009	2010	2011	2012	
t	-3	-1	1	3	
y	69	71	73	77	$\sum y = 290$
t^2	9	1	1	9	$\sum t^2 = 20$
yt	-207	-71	73	231	$\sum yt = 26$

$$a_0 = \frac{290}{4} = 72{,}5,$$

$$a_1 = \frac{26}{20} = 1{,}3.$$

Уравнение тренда $y_{t\,=}$ 72,5+1,3t.

Тогда уровень 2014 г. будет иметь параметр t:

Годы	2011	2012	2013	2014
t	1	3	5	7

,

то есть $t_{2014} = 7$, а следовательно:
$Y_{2014\,=}$ 72,5 + 1,3 * 7 = 81,6.

Ретроспективная экстраполяция (например, необходимо определить уровень 2007 года) осуществляется аналогичным образом: вначале устанавливается значение параметра t:

Годы	2007	2008	2009	2010
t	-7	-5	-3	-1

,

а затем рассчитывается экстраполируемый уровень:

$Y_{2009\,=}$ 72,5 - 1,3 * 7 = 63,4.

Интерполяция – это нахождение недостающих промежуточных уровней внутри ряда динамики. Она также производится:

а) с помощью средних абсолютных приростов:

$$y_i = y_1 + \Delta\overline{y} * (i-1),$$

где y_1 – начальный уровень ряда динамики;

i – порядковый номер искомого уровня.

Например, отсутствует уровень 2011 г., то есть:

Годы	2009	2010	2011	2012
t	69	71	…	77

$$\Delta\overline{y} = \frac{77-69}{3} = 2{,}667$$

$У_{2011}$ = 69 + 2,667*(3-1) = 74,334 млн.руб.

б) с помощью средних темпов роста:

$$y_i = y_1 * \overline{T}_p^{\,(i-1)},$$

$$\overline{T}_p = \sqrt[3]{\frac{77}{69}} = 1{,}037$$

$У_{2011}$ = 69*1,037$^{(3\text{-}1)}$ = 74,200 млрд. руб.

в) по уравнению тренда:

Годы	2009	2010	2011	2012
	69	71	…	77
t	1	2	3	4

$$a_0 + a_1 t = yt$$

В данном случае параметры уравнения тренда определяют, исходя из следующих рассуждений. Уравнение $y_t = a_0 + a_1 t$ приравнивают к двум известным уровням ряда (например, первому и последнему).

Тогда уровень 2009 может быть представлен как $y_{2005} = a_0 + a_1 \cdot 1 = = 69$, а уровень 2012 года $y_{2008} = a_0 + a_1 \cdot 4 = 77$. В результате получаем систему уравнений:

$$\begin{cases} a_0 + a_1 \cdot 1 = 69 \\ a_0 + a_1 \cdot 4 = 77 \end{cases}.$$

Решая ее, получаем значения a_1 = 2,667, a_0 = 66,333.

Следовательно: yt = 66,333 + 2,667t,

y_{2011} = 66,333 + 3 · 2,667 = 74,334.

9 ИНДЕКСНЫЙ МЕТОД В СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

9.1 Понятие индекса. Классификация индексов

<u>В статистике индекс</u> – это относительная величина, полученная в результате соотношения двух уровней одного явления.

Показатель, изменения которого характеризуется индексом, называется индексируемой величиной.

В статистике при использовании индексного метода применяется своя терминология и символика. Так, каждая индексируемая величина имеет свое обозначение:

q – физический объем продукции, товара (работ, услуг) (количество в натуральном выражении): л, м, м^2, кг, тонны, шт., пары и т. д.;

p – цена единицы продукции, товара;

$p*q = s$ – стоимость данного вида продукции, товара (объем в стоимостном выражении);

$S = \Sigma qp$ – стоимость всей продукции предприятия, товаров магазина и т. д. (то есть это могут быть показатели товарной продукции, реализованной продукции, товарооборота и т. д .);

z − себестоимость единицы продукции;

$Z = \Sigma\ qz$ – издержки, то есть себестоимость всей продукции по совокупности (цеху, предприятию и т. п.).

Обозначение самих индексов:

i – индивидуальный индекс, то есть индекс, который характеризует изменение признака у отдельного элемента изучаемой совокупности.

Так, индекс физического объема определенного продукта (товара)

$$i_q = \frac{q_1}{q_0} \qquad (9.1)$$

При построении индексов для обозначения базового значения индексируемой величины используется подстрочный знак «0», а для обозначения отчетного – «1».

Вместе с тем, при исследовании экономических явлений наряду с индивидуальными индексами, которые характеризуют изменение отдельных элементов изучаемой совокупности, широко используются сводные относительные величины для характеристики изменения совокупности (продукции, товаров и т. д.) в целом. Для этих целей рассчитывают общие индексы, которые обозначают *I*.

Например, индекс стоимости продукции:

$$I_s = \frac{\sum q_1 p_1}{\sum q_0 p_0}, \tag{9.2}$$

Эти индексы позволяют оценить изменение индексируемой величины в целом по сложной совокупности, отдельные элементы которой несопоставимы (то есть несоизмеримы в физических единицах). Например, товарооборот магазина: молоко (л) + мясо (кг) + сигареты (шт) + и т. д.

Множество индексов, разработанных статистикой, классифицируются по различным признакам.

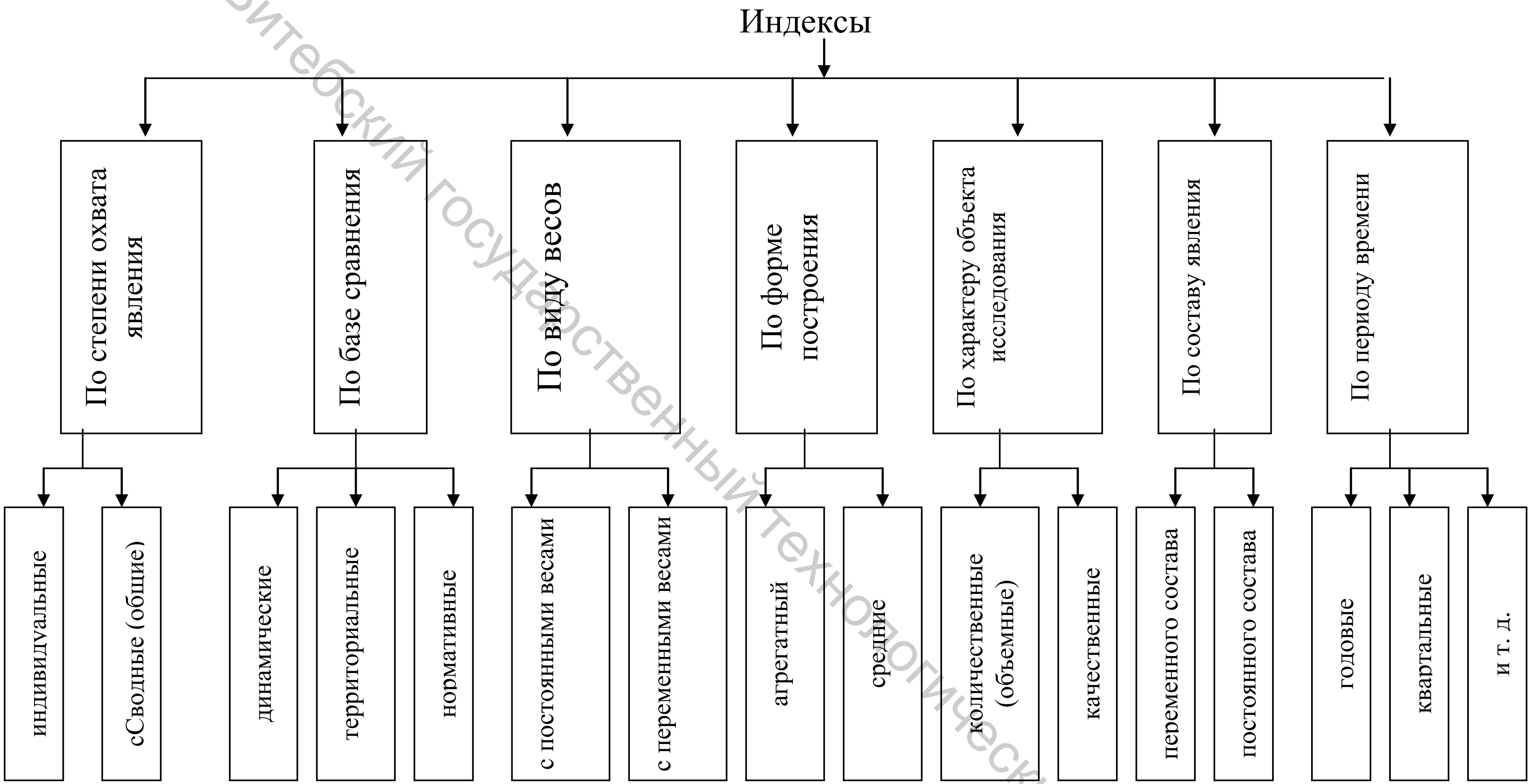

Рисунок 9.1 – Классификация индексов по основным признакам

9.2 Агрегатные индексы и их взаимосвязь

Агрегатные индексы – это индексы, в которых числитель и знаменатель представляют собой сумму произведений показателей (то есть агрегат). Как правило, перемножаются два показателя: один из них – индексируемая величина, а второй – ее вес.

Агрегатная форма индекса является основной, наиболее распространенной формой индекса в экономических исследованиях.

Агрегатный способ построения индексов позволяет с помощью определенных соизмерителей получить итоговое (суммарное) значение несопоставимых единиц сложной совокупности и в последующем сопоставить эти суммы в отчетном и базисном периодах (см. пример, таблица 9.1).

Наиболее типичные представители агрегатных индексов – индексы физического объема и цен, а также индекс стоимости. В тех случаях, когда имеются данные о производстве (о продаже) различных несоизмеримых в физических единицах продуктов в одной и той же совокупности за 2 периода и необходимо охарактеризовать изменение всего объема продуктов, они с помощью определенных соизмерителей (цены, себестоимости, трудоёмкости и т. п.) выражаются в одинаковых единицах измерения: например, *q*p, q*z, q*t,* и т. д. Это дает возможность соизмерить объемы продуктов в отчетном и базисном периодах:

$$I_s = I_{pq} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0} \quad \textit{агрегатный индекс стоимости,} \qquad (9.3)$$

который показывает относительное изменение стоимости продуктов как за счет изменения цен, так и за счет изменения объема продуктов.

Абсолютное изменение стоимости продуктов может быть оценено как

$$\Delta S = \Sigma q_1 p_1 - \Sigma q_0 p_0 . \qquad (9.4)$$

Индекс стоимости показывает, что изменение индексируемой величины обусловлено воздействием двух факторов: физического объема и цен. Для того, чтобы выявить влияние каждого из них, рассчитываются так называемые факторные индексы, которые по форме построения также являются агрегатными.

Агрегатный индекс физического объема получают путем сопоставления стоимости продукции отчетного и базисного периодов, исчисленной в неизменных ценах: ценах базисного периода:

$$I_q = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}, \qquad (9.5)$$

где $\Sigma q_1 p_0$ – стоимость продукции отчетного периода в базисных ценах;

$\Sigma q_0 p_0$ – стоимость продукции базисного периода.

Вычитая из числителя знаменатель, можно определить, на сколько в абсолютном выражении изменилась стоимость продукции за счет изменения физического объема:

$$\Delta Sq = \Sigma q_1 p_0 - \Sigma q_0 p_0. \tag{9.6}$$

По аналогии строится и агрегатный индекс цен. Только в данном случае индексируемая величина – *p*, а веса – физические объемы, то есть q_1.

$$I_p = \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0}, \tag{9.7}$$

где $\Sigma q_1 p_1$ – стоимость продукции отчетного периода;

$\Sigma q_1 p_0$ – стоимость продукции отчетного периода в базисных ценах.

Изменение стоимости продукции за счет изменения цен в абсолютном выражении:

$$\Delta Sp = \Sigma q_1 p_1 - \Sigma q_1 p_0. \tag{9.8}$$

Например, необходимо проанализировать товарооборот магазина

Таблица 9.1 – Динамика товарооборота магазина за май-июнь отчетного года

Вид товара	Ед. измер.	Кол-во проданных товаров		Цена ед. товара, тыс. руб.		Товарооборот		
		май q_0	июнь q_1	май p_0	июнь p_1	q_0p_0	q_1p_1	q_1p_0
А	л	500	600	1,200	1,250	600	750	720
Б	кг	400	440	9,000	9,000	3400	3960	3740
В	шт	1200	1100	1,150	1,150	1320	1265	1210
						$S_0 = \Sigma q_0 p_0 = 5320$	$S_1 = \Sigma q_1 p_1 = 5975$	$\Sigma q_1 p_0 = 5670$

Общее изменение товарооборота:

$$I_s = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0} = \frac{5975}{5320} = 1{,}1231 \ (+12{,}31\ \%)$$

или в абсолютном выражении:

$$\Delta S = 5975 - 5320 = 655 \text{ тыс. р.}$$

В том числе

1) за счет изменения количества проданных товаров:

$$I_q = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{5670}{5320} = 1{,}0658 \ (+6{,}58\%)$$

или в абсолютном выражении:

$$\Delta S_q = 5670 - 5320 = 350 \text{ тыс. р.}$$

2) за счет изменения цен:

$$I_p = \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} = \frac{5975}{5670} = 1{,}0538,$$

что в абсолютном выражении составит:

$$\Delta S_q = 5975 - 5670 = 305 \text{ тыс. руб.}$$

Вывод. Товарооборот в июне увеличился по сравнению с предыдущим месяцем на 12,31 %, или на 655 тыс. р. Это увеличение вызвано воздействием двух факторов: за счет роста физического объема проданных товаров – на 6,58 % (350 тыс. р.) и за счет роста цен – на 5,38 % (305 тыс. руб.).

Взаимосвязь индексов $I_s = I_q * I_p$ носит название сопряженности индексов.

Аналогично, агрегатный индекс общих издержек:

$$Iz = \frac{\Sigma q_1 z_1}{\Sigma q_0 z_0}, \tag{9.9}$$

в том числе изменение издержек за счет физического объема:

$$I_q = \frac{\Sigma q_1 z_0}{\Sigma q_0 z_0}; \tag{9.10}$$

изменение издержек за счет себестоимости единицы продукции:

$$I_z = \frac{\Sigma q_1 z_1}{\Sigma q_1 z_0}. \tag{9.11}$$

Тогда $I_z = I_q * I_z$ – сопряженные индексы.

Свойство сопряженности используется:

а) для проверки правильности расчетов;

б) для нахождения неизвестного индекса по остальным известным, с которыми он связан как сопряженный. Например:

$$I_q = \frac{I_s}{I_p} \quad \text{или} \quad I_z = \frac{I_Z}{I_q}.$$

9.3 Средние арифметические и средние гармонические индексы

Общие индексы могут быть вычислены не только как агрегатные, но и как средние из индивидуальных индексов. Для исчисления агрегатных индексов необходима информация об индексируемых величинах и весах в отчетном и базисном периодах. На практике часто приводится информация, в которой вместо этих данных приводятся данные об индивидуальных индексах. В этих случаях формой построения индекса становится средняя величина, причем в виде взвешенной, а рассчитанные индексы называют средними арифметическими или средними гармоническими.

Рассмотрим порядок их построения и условия применения.

1. <u>Средний арифметический индекс физического объема.</u>

Индекс физического объема в форме агрегатного:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \tag{9.12}$$

требует наличия информации о q_1, q_0, p_0.

Если отсутствует информация о q_1, но известны индивидуальные индексы физического объема i_q, то построение среднего индекса основывается на рассуждениях:

$$i_q = \frac{q_1}{q_0} \Rightarrow q_1 = i_q \cdot q_0, \text{ следовательно } q_1 p_0 = i_q q_0 p_0 ,$$

а индекс физического объема:

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} . \tag{9.13}$$

Проведём аналогию: $i_q = x; \; q_0 p_0 = f$.

$$I_q = \bar{x} = \frac{\sum xf}{\sum f} . \tag{9.14}$$

Следовательно, это средневзвешенная арифметическая величина из индивидуальных индексов физического объёма, в которой в качестве частот (весов) используются стоимости базисного периода.

Например, по имеющейся информации необходимо определить, как изменился выпуск (физический объём) продукции в целом по предприятию.

Таблица 9.2 – Расчет среднеарифметического индекса физического объема

Изделие	Индивидуальный индекс объёма продукции	Стоимость продукции предыдущего периода	$i_q q_0 p_0$
	i_q	$q_0 p_0$	
А	1,027	1200	1232
Б	1,065	800	852
В	1,112	500	556
		$\sum q_0 p_0 = 2500$	$\sum i_q q_0 p_0 = 2640$

$$I_q = \frac{\sum i_q * q_0 p_0}{\sum q_0 p_0} = \frac{2640}{2500} = 1{,}056 \ (+5{,}6\ \%).$$

2. Средний гармонический индекс физического объёма.

Рассуждаем аналогично:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \text{ есть } q_1,\ p_0, \text{ отсутствует } q_0, \text{ но есть и } i_q. \quad (9.15)$$

$$\text{Тогда } i_q = \frac{q_1}{q_0} \Rightarrow q_0 = \frac{q_1}{i_q} \Rightarrow q_0 p_0 = \frac{q_1 p_0}{i_q} \Rightarrow I_q = \frac{\sum q_1 p_0}{\sum \frac{q_1 p_0}{i_q}}.$$

Если обозначить: $i_q = x$, $q_1 p_0 = W$, то

$$I_q = \bar{x} = \frac{\sum W}{\sum \frac{W}{x}}, \text{ то есть это средняя гармоническая величина.}$$

Таблица 9.3 – Расчет среднегармонического индекса физического объема

Изделие	Индивидуальный индекс объёма продукции	Стоимость продукции отчетного периода в базисных ценах	$\frac{q_1 p_0}{i_q}$
	$i_q = q_1 / q_0$	$q_1 p_0$	
А	1,027	1232	1200
Б	1,065	852	800
В	1,112	556	500
		$\sum q_1 p_0 = 2640$	$\sum \frac{q_1 p_0}{i_q} = 2500$

$$I_q = \frac{\sum q_1 p_0}{\sum \frac{q_1 p_0}{i_q}} = \frac{2640}{2500} = 1{,}056.$$

Аналогичным образом выводятся формулы средних индексов цен.

3. Средний арифметический индекс цен.

$$I_p = \frac{\sum i_p \cdot q_1 p_0}{\sum q_1 p_0}\ . \tag{9.16}$$

4. Средний гармонический индекс цен.

$$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0}. \tag{9.17}$$

На практике по исходной информации важно правильно определить вид индекса, который необходимо исчислять для характеристики изменения показателя.

Например: по данным таблицы 9.4 необходимо определить, как в среднем изменились цены в отчётном периоде по сравнению с предыдущим и какова экономия (или перерасход) денежных средств у населения.

Таблица 9.4 – Расчет среднего индекса цен

Вид товара	Товарооборот отчётного периода, млн. руб.	Индивидуальные индексы цен по группам товаров	$\frac{q_1 p_1}{i_p}$
	q_1p_1	$i_p = \frac{p_1}{p_0}$	
А	2500	1,12	2232
Б	3800	1,05	3619
В	4200	0,98	4286
Г	3800	1,15	3304
Д	1600	1,08	1481
	$\sum q_1 p_1 = 15900$		$\sum \frac{q_1 p_1}{i_p} = 14922$

Введя условные обозначения, видим, что есть информация о q_1p_1 и i_p. Следовательно, можно воспользоваться средним гармоническим индексом цен:

$$I_p = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_p}}; \tag{9.18}$$

то есть $I_p = \frac{15900}{14922} = 1{,}0655$ (+6,55 %)

Перерасход денежных средств у населения:

15900 - 14992 = 978 млн. руб.

По форме средних индексов цен строятся известные во всём мире индексы ценных бумаг: индекс Доу-Джонса (Dow Jones Industrial Average Index); индекс Стэндарда и Пура (Standart fnd Poor’s 500 Stock Index).

9.4 Индексы с постоянной и переменной базой сравнения, с постоянными и переменными весами

При построении динамических индексов, если известны данные за несколько периодов, может быть построен ряд (система индексов).

Ряд индексов, каждый из которых рассчитан по отношению к предыдущему периоду, называют цепными индексами, а ряд индексов, рассчитанных по отношению к одному (постоянному) периоду, называют базисными индексами.

Так, цепные индивидуальные индексы цен имеют вид:

$$i_{p_{1/0}} = \frac{p_1}{p_0};\ i_{p_{2/1}} = \frac{p_2}{p_1};\ i_{p_{3/2}} = \frac{p_3}{p_2}\ \dots\ i_{p_{n/n-1}} = \frac{p_n}{p_{n-1}}, \quad (9.19)$$

а цепные индивидуальные индексы физического объёма:

$$i_{q\,1/0} = \frac{q_1}{q_0};\ i_{q\,2/1} = \frac{q_2}{q_1};\ i_{q\,3/2} = \frac{q_3}{q_2} \dots\ i_{q\,n/n-1} = \frac{q_n}{q_{n-1}}. \quad (9.20)$$

В свою очередь, базисные индивидуальные индексы цен:

$$i_{p\,1/0} = \frac{p_1}{p_0};\quad i_{p\,2/0} = \frac{p_2}{p_0};\quad i_{p\,3/0} = \frac{p_3}{p_0}\ \dots\ i_{p\,n/0} = \frac{p_n}{p_0}; \quad (9.21)$$

и базисные индивидуальные индексы физического объёма

$$i_{q\,1/0} \frac{q_1}{q_0};\quad i_{q\,2/0} = \frac{q_2}{q_0};\quad i_{q\,3/0} = \frac{q_3}{q_0}\ \dots\ i_{q\,n/0} = \frac{q_n}{q_0}. \quad (9.22)$$

Между цепными и базисными индексами существует определённая взаимосвязь, которая позволяет переходить от одних индексов к другим:

- произведение цепных индексов даёт базисный индекс соответствующего периода;
- отношение базисного индекса данного периода к базисному индексу предыдущего периода даёт цепной индекс данного периода.

Например, $i_{p_{3/0}} = i_{p_{1/0}} * i_{p_{2/1}} * i_{p_{3/2}}$ или $\dfrac{i_{p_{4/0}}}{i_{p_{3/0}}} = i_{p_{4/3}}$. (9.23)

Таким образом, цепные индексы имеют всё время изменяющуюся базу сравнения и поэтому называются индексами с переменной базой сравнения. В то же время базисные индексы имеют постоянную базу сравнения и, соответственно, называются индексами с постоянной базой сравнения.

Цепные и базисные индексы могут быть построены и для общих индексов. Однако в данном случае дополнительно решается вопрос о весах, так как они также могут быть постоянными и переменными.

Допустим, что имеется информация о выпуске (q) и ценах (p) нескольких видов продукции (А, Б и т. д.) за несколько периодов (например, за 4 квартала).

Таблица 9.5− Макет таблицы для расчета цепных и базисных индексов с переменными и постоянными весами

Вид продукции	I квартал		II квартал		III квартал		IV квартал	
	выпуск, шт.	цена, тыс. руб.	выпуск, шт.	цена, тыс. руб	выпуск, шт.	цена, тыс. руб.	выпуск, шт.	цена, тыс. руб.
	q_0	p_0	q_1	p_1	q_2	p_2	q_3	p_3
А								
Б								
…								

В данном случае при расчёте цепных индексов физического объёма по агрегатной формуле продукцию всех периодов можно оценить в одних и тех же ценах, например в p_0. Тогда такие цепные индексы будут иметь следующий вид:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad I_{q_{2/1}} = \frac{\sum q_2 p_0}{\sum q_1 p_0}; \quad I_{q_{3/2}} = \frac{\sum q_3 p_0}{\sum q_2 p_0}. \qquad (9.24)$$

Эти индексы будут называться индексами с постоянными весами, т.к. все они имеют один и тот же соизмеритель – p_0.

Однако при исчислении цепных индексов физического объема можно поступить и иначе: принимать в качестве весов цены предыдущего периода. В этом случае цепные индексы физического объема будут иметь такой вид:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad I_{q_{2/1}} = \frac{\sum q_2 p_1}{\sum q_1 p_1}; \quad I_{q_{3/2}} = \frac{\sum q_3 p_2}{\sum q_2 p_2}. \qquad (9.25)$$

Так как эти индексы построены по разным соизмерителям (p_0, p_1, p_2), их принято называть индексами с переменными весами.

По аналогии с агрегатными индексами физического объема могут быть построены цепные агрегатные индексы цен:

- с постоянными весами

$$I_{p\,1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1};\; I_{p\,2/1} = \frac{\sum p_2 q_1}{\sum p_1 q_1};\; I_{p\,3/2} = \frac{\sum p_3 q_1}{\sum p_2 q_1}, \tag{9.26}$$

- с переменными весами

$$I_{p\,1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1};\; I_{p\,2/1} = \frac{\sum p_2 q_2}{\sum p_1 q_2};\; I_{p\,3/2} = \frac{\sum p_3 q_3}{\sum p_2 q_3} \quad \text{и т. д.} \tag{9.27}$$

В свою очередь, базисные агрегатные индексы цен:
- с постоянными весами

$$I_{p\,1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1};\; I_{p\,2/0} = \frac{\sum p_2 q_1}{\sum p_0 q_1};\; I_{p\,3/0} = \frac{\sum p_3 q_1}{\sum p_0 q_1}, \tag{9.28}$$

- с переменными весами

$$I_{p\,1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1};\; I_{p\,2/0} = \frac{\sum p_2 q_2}{\sum p_0 q_2};\; I_{p\,3/0} = \frac{\sum p_3 q_3}{\sum p_0 q_3}. \tag{9.29}$$

Соответственно, базисные агрегатные индексы физического объема:
- с постоянными весами

$$I_{q\,1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0};\; I_{q\,2/0} = \frac{\sum q_2 p_0}{\sum q_0 p_0};\; I_{q\,3/0} = \frac{\sum q_3 p_o}{\sum q_0 p_0}, \tag{9.30}$$

- с переменными весами

$$I_{q\,1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0};\; I_{q\,2/0} = \frac{\sum q_2 p_1}{\sum q_0 p_1};\; I_{q\,3/0} = \frac{\sum q_3 p_2}{\sum q_0 p_2}. \tag{9.31}$$

В случае агрегатных индексов взаимосвязи между цепными и базисными индексами (сформулированные для индивидуальных индексов) действуют только для индексов с постоянными весами.

На цепные и базисные агрегатные индексы с переменными весами эти взаимосвязи не распространяются.

9.5 Индексный метод анализа динамики среднего уровня (индексы переменного состава, постоянного состава, структурных сдвигов)

При исследовании различных экономических процессов и явлений используются не только объемные показатели: товарная продукция, реализованная продукция, товарооборот, издержки и т. п.; но и целый ряд средних показателей: средняя заработная плата 1 работника, средняя выработка 1 рабочего (работника), средняя себестоимость единицы продукции, средняя цена товара и т. д.

Во всех этих случаях необходимо учитывать, как влияет на динамику показателя изменение структуры изучаемой совокупности. Например, средняя заработная плата на предприятии может вырасти не только за счет того, что она выросла у отдельных работников, но и за счет того, что увеличился удельный вес (доля) высокооплачиваемых работников.

При исследовании динамики таких (средних) показателей важно определить, в какой мере изменение показателя вызвано структурными сдвигами, а в какой непосредственно изменением показателя у отдельных единиц совокупности. Эта задача на практике решается с помощью индексного метода.

Если ввести следующие обозначения:

x – (индивидуальные значения) групповые средние (средняя заработная плата по категориям работающих, средняя производительность труда по цехам, средний балл успеваемости по группам и т. д.);

f – численность единиц совокупности в группе, (численность единиц с данным уровнем),

то $\bar{x}$ – средний уровень по совокупности (средняя заработная плата по предприятию, средняя производительность труда по предприятию, средний балл успеваемости по факультету или университету и т. д.), очевидно, что $\bar{x} = \dfrac{\sum xf}{\sum f}$.

Тогда изменение среднего уровня показателя по совокупности в целом может быть оценено с помощью следующего индекса:

$$I_{\bar{x}} = \frac{\overline{x_1}}{\overline{x_0}} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}. \qquad (9.32)$$

Такой индекс получил в статистике название <u>индекса переменного состава</u>. Он отражает влияние на динамику показателя <u>двух факторов</u>: изменения индексируемой величины (x) и изменения структуры совокупности.

Так, индекс средних цен может быть построен следующим образом:

$$I_{\bar{p}\ п.с.} = \frac{\sum p_1 q_1}{\sum q_1} \div \frac{\sum p_0 q_0}{\sum q_0}, \quad т.е.\ I_{\bar{p}} = \frac{\overline{p_1}}{\overline{p_0}}. \qquad (9.33)$$

Для обособленной оценки влияния на изменение средней величины каждого из указанных ранее факторов строятся следующие индексы:

а) для оценки влияния <u>изменения индексируемой величины</u>:

$$I_{\bar{x}} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1} \quad или \quad I_{\bar{x}} = \frac{\sum x_1 f_1}{\sum x_0 f_1} \qquad (9.34)$$

Такой индекс носит название индекса постоянного (фиксированного) состава, т. к. состав изучаемой совокупности фиксируется на одном уровне – уровне отчетного периода.

Например, индекс постоянного состава средней цены строится по формуле

$$I_{\bar{p}_{ф.с.}} = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_1}{\sum q_1} \quad или \quad I_{\bar{p}} = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \qquad (9.35)$$

б) для оценки влияния изменения структуры изучаемой совокупности, то есть структурных сдвигов:

$$I_{\bar{x}} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} \quad или \quad I_{\bar{x}} = \frac{\sum x_0 f_1}{\sum x_0 f_0} : \frac{\sum f_1}{\sum f_0} \ . \qquad (9.36)$$

Этот индекс принято называть <u>индексом структурных сдвигов</u>.

Так, индекс структурных сдвигов для средней цены определяется по формуле

$$I_{\bar{p}_{с.с.}} = \frac{\sum p_0 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0} \quad или \quad I_{\bar{p}} = \frac{\sum p_0 q_1}{\sum p_0 q_0} : \frac{\sum q_1}{\sum q_0}. \qquad (9.37)$$

Можно легко доказать, что индексы переменного, постоянного состава и структурных сдвигов находятся во взаимосвязи:

$$I_{\bar{x}п.с.} = I_{\bar{x}ф.с.} * I_{\bar{x}с.с.} \qquad (9.38)$$

$$\frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} = \left(\frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1} \right) * \left(\frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_o}{\sum f_0} \right) = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}. \qquad (9.39)$$

Как и в случае агрегатных индексов, эта взаимосвязь используется:

1) для контроля за правильностью выполнения расчетов;

2) для нахождения значения неизвестного индекса по значениям известных.

Например: по данным таблицы 9.6 необходимо определить, как изменилась средняя цена единицы продукции и под воздействием каких факторов сложилось это изменение (цифры условные).

Таблица 9.6 – Расчет индексов средней цены

Вид продукции	Количество продукции, тыс. ед.		Цена единицы продукции, тыс.р.		p_0q_0	p_1q_1	p_0q_1
	базисный период, q_0	отчетный период, q_1	базисный период, p_0	отчетный период, P_1			
А Б	40 40	60 20	100 200	125 250	4000 8000	7500 5000	6000 4000
	$\Sigma q_0 = 80$	$\Sigma q_1 = 80$			$\Sigma p_0q_0 =$ $= 12000$	$\Sigma p_1q_1 =$ $= 12500$	$\Sigma p_0q_1 =$ $= 10000$

Изменение средней цены составило:

$$I_{\bar{p}\, п.с.} = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0} = \frac{12500}{400} : \frac{12000}{400} = 31{,}25 : 30 = 1{,}0416\ (+4{,}16\,\%)$$

или в абсолютном выражении

$\Delta\bar{p} = 31{,}25 - 30 = 1{,}25\, млн.\, руб.$

В том числе за счет непосредственного изменения цен на продукцию:

$$I_{\bar{p}ф.с.} = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_1}{\sum q_1} = \frac{12500}{400} : \frac{10000}{400} = 31{,}25 : 25 = 1{,}2500\,(+25{,}00\,\%)$$

в абсолютном выражении $\Delta\bar{p}_p = 31{,}25 - 25 = 6{,}25\, млн.\ руб.$;

и за счет структурных сдвигов:

$$I_{\bar{p}\, с.с.} = \frac{\sum p_0 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0} = 25 : 30 = 0{,}8333\,(-16{,}67\,\%),$$

в абсолютном выражении $\Delta\bar{p}_{с.с} = 25 - 35 = -5\, млн.\ руб.$

Вывод. Средняя цена единицы продукции в отчетном периоде по сравнению с базисным возросла на 4,16 % (на 1,35 тыс. руб.). Этот рост был вызван непосредственно ростом цен на продукцию на 25 %. Одновременно структурные сдвиги в составе продукции: увеличение доли дешевой и уменьшение доли дорогой продукции привели к снижению средней цены на 16,67 %.

10 СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ СВЯЗИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

10.1 Понятие корреляционной связи

Все явления общественной жизни находятся во взаимосвязи и взаимообусловленности. Одной из важнейших задач статистики является установление и измерение связи и зависимости между явлениями.

Так как формы и виды этих взаимосвязей весьма разнообразны, существуют их различные классификации.

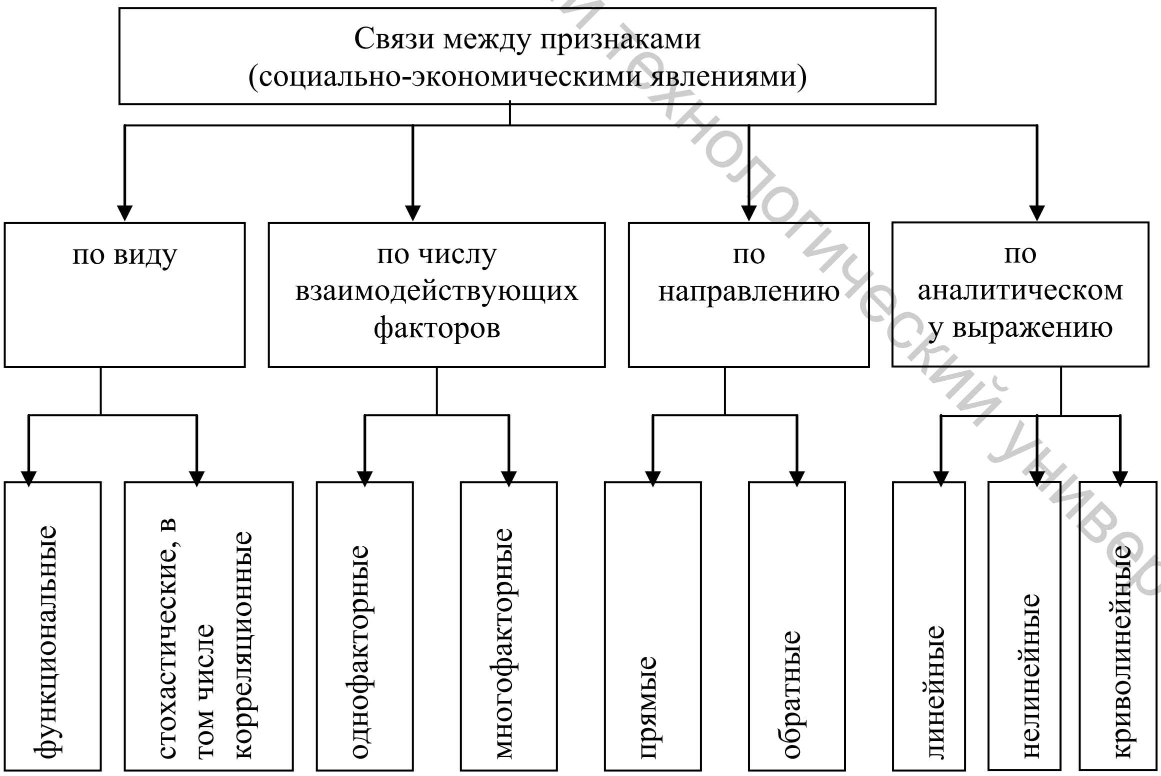

Рисунок 10.1 – Формы и виды взаимосвязей

Объектом статистического изучения являются факторные или причинно-следственные связи, которые по степени тесноты связи (по степени детерминизма) связи могут быть двух видов:

- функциональные;
- стохастические (статистические).

Функциональная связь предполагает, что определённому значению признака-фактора соответствует одно, строго определённое значение признака-результата, например:

$$S = \pi r^2$$
$$y = 3{,}14x^2$$

$x = 1$	$y = 3{,}14$
$x = 2$	$y = 12{,}56$
$x = 3$	$y = 28{,}26$
$x = \ldots$	$y = \ldots$
$x = 10$	$y = 314{,}00$ и т. д.

Такие связи принято называть жёсткими или полными. Они присутствуют в физике, математике, астрономии и других точных науках.

Статистические – характеризуются тем, что в данном случае связь наблюдается не в каждом конкретном случае, а в среднем, при большом количестве наблюдений.

Частным случаем статистической связи является корреляционная связь.

Общая задача статистического изучения взаимосвязей может быть сформулирована следующим образом: по результатам *n* измерений исследуемых факторов *x* и *y*

x_1	x_2	x_3	x_4	…	x_n
y_1	y_2	y_3	y_4	…	y_n

необходимо получить функцию, которая позволила бы по заданным значениям факторных переменных (x) восстанавливать (прогнозировать) значения результирующих переменных (y), то есть функцию y = f(x).

Однако задачи исследования взаимосвязей могут быть конкретизированы и носить более частный характер в зависимости от цели исследования:

1) выявление наличия или отсутствия корреляционной связи между изучаемыми признаками;

2) измерение тесноты связи между признаками;

3) определение математической модели для описания зависимости между признаками: признаком-результатом и одним либо несколькими признаками-факторами.

Для решения каждой из этих задач теория статистики разработала свои приёмы и методы:

- 1-я задача может быть решена с помощью так называемых элементарных методов изучения взаимосвязей: графического, балансового, метода аналитических группировок, метода сравнения параллельных рядов;
- 2-я задача может быть решена с помощью корреляционного анализа, дисперсионного анализа;

- 3-я задача требует построения функции $y = f(x)$, то есть проведения регрессионного анализа.

10.2 Методы выявления взаимосвязей в статистике

Неотъемлемым элементом любого метода изучения взаимосвязей является предшествующий теоретический анализ, в результате которого будет выявлено, что связь между изучаемыми признаками возможна.

Например, с ростом заработной платы возможен рост производительности труда.

Но: рост успеваемости студентов экономического факультета УО «ВГТУ» и рост производства продукции РУПП «Витязь» – ложная корреляция!

10.2.1 Метод сравнения параллельных рядов

Метод сравнения параллельных рядов позволяет установить направление связи между изучаемыми признаками. Для этого единицы изучаемой совокупности располагаются в порядке возрастания (убывания) признака-фактора. Параллельно записываются соответствующие им значения признака-результата. На основании логического анализа делается вывод о наличии и характере взаимосвязи между признаками:

а) если с возрастанием уровней признака-фактора наблюдается рост уровней признака-результата, то между ними существует прямая связь; то есть: связь есть, связь прямая;

б) если с возрастанием уровней признака-фактора наблюдается снижение уровней признака-результата, связь есть, и она обратная;

в) если с возрастанием уровней признака-фактора наблюдается хаотичное распределение значений признака-результата, то связь отсутствует.

Например: имеются сведения об уровне фондоотдачи и удельном весе активной части в общей стоимости основных средств (ОС) по 10 организациям. Необходимо установить, существует ли взаимосвязь между этими показателями и каков её характер (или каково её направление).

Произведём ранжирование уровней *x* (удельного веса активной части ОС) и расположим элементы изучаемой совокупности в порядке возрастания признака-фактора (таблица 10.2, графы 1 и 2).

Таблица 10.1 – Показатели использования и структуры основных средств по организациям (данные условные)

№ организации	1	2	3	4	5	6	7	8	9	10
Фондоотдача, руб.	3,7	3,4	2,9	2,8	3,0	2,5	3,7	2,9	2,8	3,2
Удельный вес активной части ОС, %	65	63	56	53	58	48	61	58	48	61

Анализируя полученные параллельные ряды, можем сделать вывод о наличии прямой связи (то есть связь есть, связь прямая) между фондоотдачей (у) и удельным весом активной части ОС (x), хотя она проявляется не в каждом конкретном случае, а в целом, в среднем.

10.2.2 Метод группировок

При сравнении индивидуальных значений признаков (метод сравнения параллельных рядов) не всегда отчетливо просматривается корреляционная зависимость. Более наглядное представление об этой зависимости можно получить, если сравнивать не индивидуальные значения признака, а групповые средние значения признака-фактора и признака-результата. Такой прием изучения взаимосвязей получил название метода аналитических группировок. Чтобы выявить наличие и направление связи между признаками, производится группировка единиц изучаемой совокупности по признаку-фактору (x), и по каждой группе рассчитывается среднее значение признака-результата (y).

Далее, как и в методе сравнения параллельных рядов, сравнивается направление изменения признака-фактора и признака-результата. Если они совпадают – связь есть, связь прямая; если они не совпадают – связь есть, связь обратная.

Для построения группировки в нашем примере принимается количество интервалов $к = 4$ (так как $n = 10$, исходя из формулы Стерджеса), следовательно, ширина интервала группировки составит:

$$i_x = \frac{x_{\max} - x_{\min}}{k} = \frac{65 - 48}{4} = 4{,}25 .$$

Таблица 10.2 – Группировка предприятий отрасли по удельному весу активной части основных средств и фондоотдачи

Группы предприятий по удельному весу активной части ОС, x	Число предприятий в группе, $f_{гр}$	Суммарное значение признака-результата в группе $\Sigma y_{гр} * f_{гр}$	Среднее значение признака-результата по группе $\bar{y}_{гр} = \frac{\sum y_{гр} * f_{гр}}{\sum f_{гр}}$	$\bar{y}_{гр} - \bar{y}$	$(\bar{y}_{гр} - \bar{y})^2$	$(\bar{y}_{гр} - \bar{y})^2 * f_{гр}$
1	2	3	4	5	6	7
48,00-52,25	2	2,5+2,8=5,3	2,65	-0,44	0,1936	0,3872
52,25-56,50	2	2,8+2,9=5,7	2,85	-0,24	0,0576	0,1152
56,50-60,75	2	2,9+3,0=5,9	2,95	-0,14	0,0196	0,0392
60,75-65,00	4	3,7+3,2+3,4+3,7=14,0	3,50	0,41	0,1681	0,6724
Сумма	10	30,9				1,2140

Таблица 10.3 – Расчетная таблица

x	y	$x-\bar{y}$	$(y-\bar{y})^2$	x^2	xy	y_x = -0,279+ +0,059x	y^2	$(y-y_x)^2$	Знаки отклонений		Ранги		Разность рангов, d	d^2
									x от $\bar{x}$	y от $\bar{y}$	x	y		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
48	2,5	-0,59	0,3481	2304	120,0	2,553	6,25	0,0028	-	-	1,5	1	0,5	0,25
48	2,8	-0,29	0,0841	2304	134,4	2,553	7,84	0,0610	-	-	1,5	2,5	-1,0	1,00
53	2,8	-0,29	0,0841	2809	148,4	2,848	7,84	0,0023	-	-	3	2,5	0,5	0,25
56	2,9	-0,19	0,0361	3136	162,4	3,025	8,41	0,0156	-	-	4	4,5	-0,5	0,25
58	2,9	-0,19	0,0361	3364	168,2	3,143	8,41	0,0590	+	-	5,5	4,5	1,0	1,00
58	3,0	-0,09	0,0081	3364	174,0	3,143	9,00	0,0204	+	-	5,5	6	-0,5	0,25
61	3,7	0,61	0,3721	3721	225,7	3,320	13,69	0,1444	+	+	7,5	9,5	-2,0	4,00
61	3,2	0,11	0,0121	3721	195,2	3,320	10,24	0,0144	+	+	7,5	7	0,5	0,25
63	3,4	0,31	0,0961	3969	214,2	3,438	11,56	0,0014	+	+	9	8	1,0	1,00
65	3,7	0,61	0,3721	4225	240,5	3,556	13,69	0,0207	+	+	10	9,5	0,5	0,25
$\sum x = 571$	$\sum y = 30,9$		$\sum (y-\bar{y})^2 = 1,4490$	$\sum x^2 = 32917$	$\sum xy = 1783,0$	$\sum y_x = 30,899$	$\sum y^2 = 96,93$	$\sum (y-y_x)^2 = 0,3420$						$\sum d^2 = 8,50$

Так как направления изменения признаков (графа 1 и графа 4 таблицы 10.3) совпадают, принимается вывод о наличии прямой связи между фондоотдачей и удельным весом активной части ОС.

Метод аналитической группировки является единственно приемлемым в случае большого числа единиц совокупности, например не 10, а 1000. Это является его существенным преимуществом

10.2.3 Графический метод (метод корреляционного поля)

Графический метод часто называют методом корреляционного поля. Сущность его заключается в следующем: на график, у которого одна ось x – признак-фактор, а другая ось y – признак-результат, наносятся точки, отображающие исходную информацию (удобнее в ранжированном виде, по таблице 10.3), и соединяются ломаной линией. Далее по расположению этих точек на графике делается вывод о наличии, направлении и, частично, о тесноте связи:

а) если точки на графике концентрируются около некоторой прямой, направленной из левого нижнего в правый верхний угол, то принимается вывод о наличии прямой связи (связь есть, связь прямая);

б) если точки концентрируются около прямой, направленной из левого верхнего в правый нижний угол, связь есть и она обратная;

в) если точка концентрируется в виде дуги около некоторой кривой (например, параболы) принимается вывод о наличии криволинейной связи;

г) если на корреляционном поле наблюдается хаотичный разброс точек, принимается вывод об отсутствии взаимосвязи исследуемых признаков.

Примерный вывод о тесноте связи делается на основании разброса точек на корреляционном поле. Чем ближе они концентрируются вокруг некоторой прямой или кривой, то есть чем меньше их рассеяние, тем теснее корреляционная связь.

В нашем примере (рисунок 10.2) точки на корреляционном поле концентрируются около прямой, направленной из левого нижнего в правый верхний угол, что позволяет сделать вывод о наличии прямой зависимости между фондоотдачей и удельным весом активной части в общей стоимости основных средств. Более того, точки сконцентрированы достаточно близко к некоторой прямой.

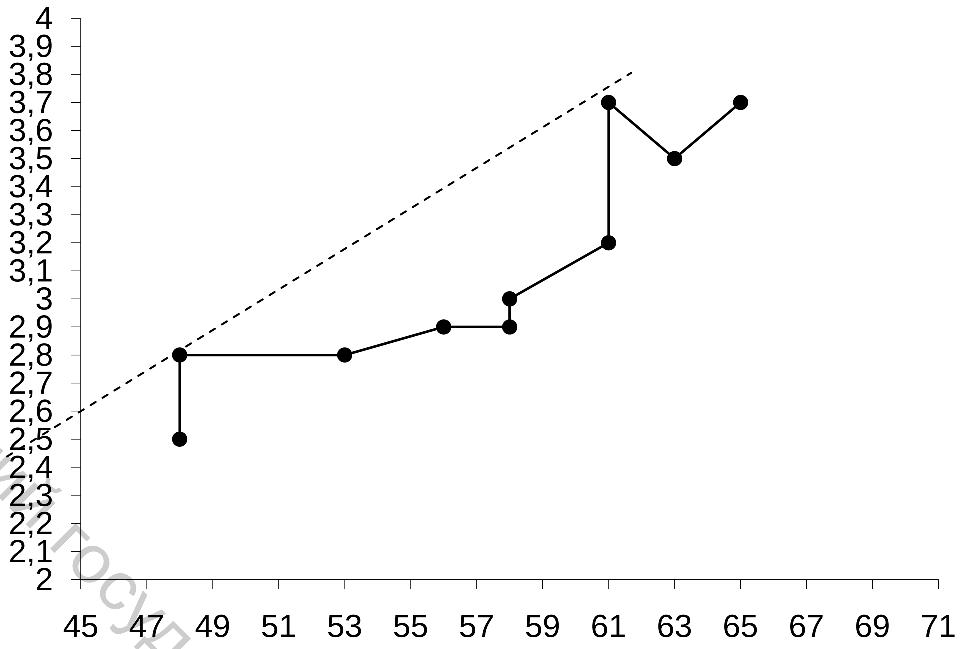

Рисунок 10.2 – Корреляционное поле зависимости фондоотдачи (y) от удельного веса активной части основных средств (x)

Вывод: связь есть, связь прямая и достаточно тесная.

10.2.4 Табличный (балансовый) метод

Этот метод имеет и целый ряд других названий: табличный метод, метод корреляционной таблицы, метод корреляционной решетки.

Для построения такой таблицы (она имеет форму шахматной таблицы), группируются уровни x и y исходя из следующих правил:

- интервалы устанавливаются равные, то есть ширина интервала определяется по формуле:

для признака-фактора $i_x = \frac{x_{\max} - x_{\min}}{k}$,

для признака-результата $i_y = \frac{y_{\max} - y_{\min}}{k}$;

- количество групп (k) одинаковое для x и для y;
- количество интервалов не следует делать большим, так как таблица теряет наглядность (хотя строгих правил нет).

В нашем примере примем $k = 4$, тогда

$$i_x = \frac{x_{\max} - x_{\min}}{k} = \frac{65 - 48}{4} = 4{,}25,$$

$$i_y = \frac{y_{\max} - y_{\min}}{k} = \frac{3{,}7 - 2{,}5}{4} = 0{,}3.$$

После этого строится макет корреляционной таблицы по строкам – признак-фактор, по столбцам – признак-результат.

Группы предприятий по y / Группы предприятий по x	2,5 - 2,8	2,8 - 3,1	3,1 - 3,4	3,4 - 3,7	Всего
48,00 – 52,25	•	•			2
52,25 – 56,50		••			2
56,50 – 60,75		••			2
60,75 – 65,00			•	...	4
Всего	1	5	1	3	10

Заполнение построенной таблицы производится методом точек или черточек: на пересечении соответствующей строки (x) и столбца (y) в любом месте квадрата (прямоугольника) ставится точка либо черточка. Иногда ставится число, показывающее общее количество единиц совокупности, которое попало в данный прямоугольник (в левом верхнем квадрате должна быть 1, а в правом нижнем – 3).

На последнем этапе производится анализ расположения единиц совокупности по группам, то есть в таблице:

а) если точки впиваются в эллипс, направленный из левого верхнего в правый нижний угол, связь есть, и она прямая;

б) если точки вписываются в эллипс, направленный из левого нижнего в правый верхний угол, связь есть, и она обратная;

в) если точки концентрируются около некоторой дуги, делается предположение о наличии криволинейной связи;

г) при хаотичном разбросе данных принимается вывод об отсутствии связи между исследуемыми признаками.

В нашем примере точки в корреляционной таблице вписываются в эллипс, направленный из левого верхнего в правый нижний угол, следовательно, между фондоотдачей и удельным весом активной части основных средств существует прямая связь.

10.3 Дисперсионный анализ и его использование в оценке тесноты связи

Дисперсионный анализ применяют:

1) для оценки тесноты связи между признаками в аналитических группировках (в балансовом методе также представляется такая возможность);

2) для определения роли исследуемого признака-фактора в изменении (вариации) признака-результата.

Для характеристики тесноты связи между признаками рассчитывают эмпирическое корреляционное отношение:

$$\eta = \sqrt{\frac{\delta y^2}{\sigma_y^{\ 2}}}, \qquad (10.1)$$

где δy^2 – межгрупповая дисперсия признака-результата;

$\sigma_y^{\ 2}$ – общая дисперсия признака-результата (см. «Виды дисперсий»).

В нашем примере общую дисперсию можно вычислить по формуле (по таблице 10.2)

$$\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n}\ . \qquad (10.2)$$

Учитывая, что $\bar{y} = \frac{\sum y}{n} = \frac{30{,}9}{10} = 3{,}09$ и проведя необходимые расчеты в таблице 10.2, получим:

$$\sigma_y^2 = \frac{1{,}4490}{10} = 0{,}1449$$

Межгрупповая дисперсия рассчитывается по формуле (по таблице 10.3)

$$\delta^2 = \frac{\sum (\bar{y}_{гр} - \bar{y})^2 f_{гр}}{\sum f_{гр}}\ , \qquad (10.3)$$

где $\bar{y}_{гр}$ – среднее значение признака-результата по группе;

$f_{гр}$ – число единиц совокупности в группе.

$$\delta^2 = \frac{1{,}2140}{10} = 0{,}1214.$$

Тогда эмпирическое корреляционное отношение составит:

$$\eta = \sqrt{\frac{0{,}1214}{0{,}1449}} = 0{,}915\,.$$

Чем ближе η к 1, тем более тесная связь. В нашем примере она весьма тесная, так как η достаточно близко к 1.

Для того, чтобы определить, какая доля вариации признака-результата вызвана действием признака-фактора, положенного в основание группировки, используют <u>коэффициент детерминации:</u>

$$\eta^2 = \frac{\delta_y^2}{\sigma_y^2}\,. \qquad (10.4)$$

В нашем примере $\eta^2 = 0{,}838$. Следовательно, вариация фондоотдачи по организациям на 83,8 % вызвана изменением доли активной части ОС в их общей стоимости.

10.4 Корреляционно-регрессионный анализ

Корреляционно-регрессионный анализ решает две важные, неразрывные и дополняющие друг друга задачи:

1) определение формы связи между признаками *x* и *y*, то есть установление математической модели или аналитического выражения этой связи;

2) измерение тесноты, то есть меры связи между *x* и *y*.

1-я задача решается с помощью регрессионного анализа, 2-я – с помощью корреляционного анализа. Последовательность их решения может быть различной: вначале регрессионный анализ, а затем корреляционный либо наоборот.

Регрессионный анализ начинается с выбора формы связи между признаками *x* и *y*. Определяющая роль в этом выборе отводится теоретическому анализу (например, рост текучести кадров будет вызывать падение уровня производительности труда; рост заработной платы будет сопровождаться ростом производительности труда и т. д.).

В зависимости от характера изменения признака-результата под влиянием изменения признака-фактора теоретическая форма связи может принимать различные виды уравнений:

– прямой $y_z = a_0 + a_1 x$;

– параболы $y_x = a_0 + a_1 x + a_2 x^2$;

– гиперболы $y_x = a_0 + a_1 \dfrac{1}{x}$;

– показательной функции $y_z = a_0 + a_1^{\,x}$;

– и др.

Выбор формы связи всегда является несколько условным, так как статистическая зависимость только приближается к функциональной, а исследователь осуществляет поиск функциональной связи. Для выбора формы связи могут быть использованы такие элементарные методы изучения взаимосвязей, как графический или балансовый.

Теоретическая линия связи, с помощью которой описывается исследуемая статистическая связь, называется <u>уравнением регрессии,</u> выбор, построение и анализ этого уравнения – <u>регрессионным анализом</u>.

Рассмотрим на примере линейной зависимости:

$$y_x = a_0 + a_1 x .$$

После того, как определён выбор типа функции, необходимо решить уравнение регрессии, то есть найти параметры этого уравнения a_0 и a_1.

Независимо от формы связи параметры a_0 и a_1 уравнения регрессии определяются с помощью метода наименьших квадратов.

Система нормальных уравнений метода наименьших квадратов для линейного уравнения имеет вид:

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

Для нахождения $\sum x, \sum y, \sum x^2, \sum xy$ в нашем примере используем таблицу 10.2. В результате получаем:

$$\begin{cases} 10a_0 + 571a_1 = 30{,}9 \\ 571a_0 + 32917a_1 = 1783 \end{cases}.$$

В результате решения этой системы получаем значения: a_0 =0,279, a_1 = = 0,059.

Для нашего примера уравнение регрессии принимает вид:

$$y_x = -0{,}279 + 0{,}059x.$$

Подставляя значения x в уравнение регрессии, определяем теоретические уровни признака-результата (таблица 10.2), а затем рассчитываем ошибку (или расхождение), которая не должна превышать 1 %:

$$E = \frac{\sum y_x - \sum y}{\sum y} * 100. \qquad (10.5)$$

В нашем примере:

$$E = \frac{30{,}899 - 30{,}9}{30{,}9} * 100 = 0{,}003\ \% \langle 1\ \%.$$

Следовательно, форма связи выбрана правильно.

Анализ (экономическая интерпретация) уравнения регрессии основан на параметре a_1, который называют <u>коэффициентом регрессии</u>. Он показывает, на сколько в абсолютном выражении изменится признак-результат при изменении признака-фактора на единицу.

В нашем примере: увеличение удельного веса активной части в общей стоимости основных средств на 1 процентный пункт вызывает рост фондоотдачи на 0,059 рублей.

Для более удобного восприятия результатов регрессионного анализа целесообразно рассчитывать <u>коэффициент эластичности</u>. Он выражает зависимость y от x в %-ах и определяется по формуле

$$Э = a_1 \frac{\bar{x}}{\bar{y}}. \qquad (10.6)$$

В нашем примере: $\bar{x} = 57{,}1$; $\bar{y} = 3{,}09$;

$$Э = 0{,}059 \frac{57{,}1}{3{,}09} = 1{,}090.$$

Это означает, что при увеличении удельного веса активной части основных средств на 1 % фондоотдача возрастает на 1,09 %.

Если уравнение регрессии y_x = -0,279 + 0,059x нанести на график (корреляционное поле) и провести на нём ещё одну линию $y = \bar{y}$ (рисунок 10.2), то на графике получится три линии, расположение которых имеет своё объяснение:

- большой угол наклона (y_x) теоретической линии связи (2) к $(\bar{y} = 3{,}09)$ горизонтальной линии (3) свидетельствует о наличии тесной связи между x и y;

- несовпадение теоретической линии (2) связи (y_x) и эмпирической (1) (ломаной линии) объясняется действием на признак-результат не только фактора x, но и других факторов.

Значение уравнения регрессии на практике: предполагая, что признак-фактор примет определённое значение, можно составить прогноз признака-результата.

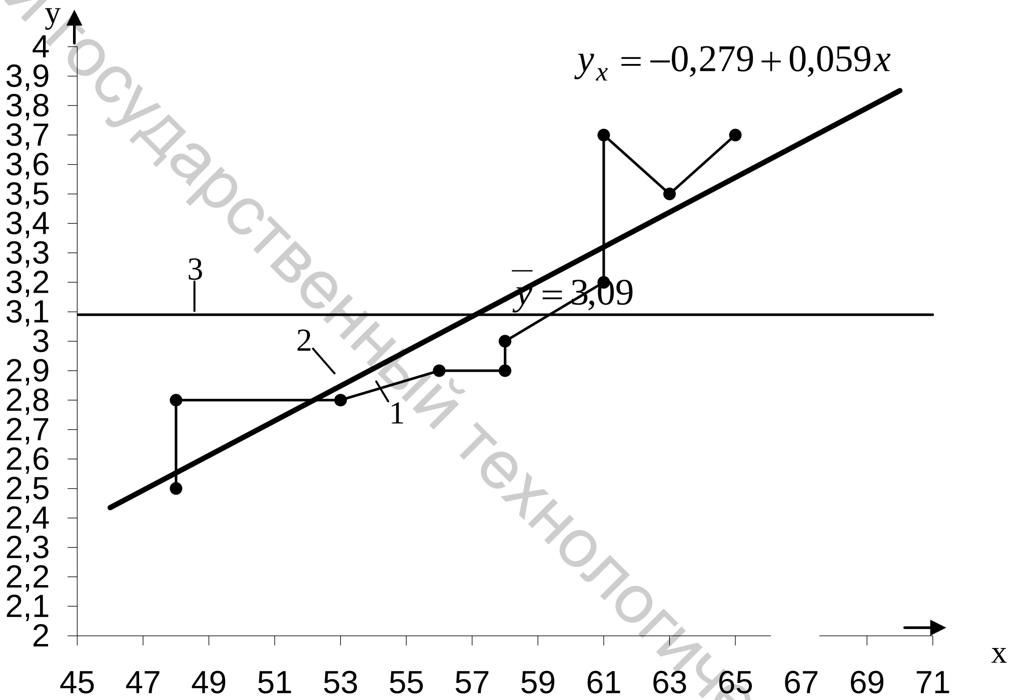

Рисунок 10.2 – Корреляционное поле зависимости фондоотдачи (y) от удельного веса активной части основных средств (x)

<u>Корреляционный анализ</u> предполагает оценку тесноты связи между признаками x и y.

В случае линейной зависимости для оценки степени тесноты этой связи используется <u>линейный коэффициент корреляции</u> (он нашёл наибольшее распространение на практике).

В теории статистики существует множество формул для определения линейного коэффициента корреляции. Исходным положением является следующее: линейный коэффициент корреляции представляет собой среднюю величину из произведений нормированных отклонений для x и y:

$$r = \frac{\sum\left(\frac{x-\bar{x}}{\sigma_x}\right)*\left(\frac{y-\bar{y}}{\sigma_y}\right)}{n} \quad . \qquad (10.7)$$

Другой вид формулы получается в том случае, если σ_x и σ_y как постоянные величины выносятся за знак суммы:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{n*\sigma_x*\sigma_y} \quad . \qquad (10.8)$$

Путём математических преобразований можно данную формулу привести к виду:

$$r = \frac{\overline{xy}-\bar{x}*\bar{y}}{\sigma_x*\sigma_y}\text{, при этом:} \qquad (10.9)$$

$$\sigma_x = \sqrt{\overline{x^2}-(\bar{x})^2}\,, \qquad (10.10)$$

$$\sigma_y = \sqrt{\overline{y^2}-(\bar{y})^2}\,. \qquad (10.11)$$

Рассчитаем необходимые составляющие последней формулы для нашего примера (по таблице 10.2):

$\overline{xy} = \frac{1783}{10} = 178{,}3$; $\bar{x} = \frac{571}{10} = 57{,}1$; $\bar{y} = \frac{30{,}9}{10} = 3{,}09$.

Из раздела 10.3 $\sigma_y^2 = 0{,}1449$, следовательно $\sigma_y = \sqrt{0{,}1449} = 0{,}381$.

Все дальнейшие необходимые расчеты проведем в таблице 10.2. По данным графы 5:

$$\overline{x^2} = \frac{\sum x^2}{n} = \frac{32917}{10} = 3291{,}7\,;$$

а квадрат средней величины признака-фактора: $(\bar{x})^2 = 57{,}1^2 = 3260{,}4$.

Следовательно: $\sigma_x = \sqrt{3291{,}7 - 3260{,}4} = 5{,}595$.

По данным графы 8 таблицы 10.2: $\overline{y^2} = \frac{\sum y^2}{n} = \frac{96{,}93}{10} = 9{,}693$;

а квадрат среднего значения признака-результата: $(\bar{y})^2 = 3{,}09^2 = 9{,}548$.

Тогда $\sigma_y = \sqrt{9{,}693 - 9{,}548} = 0{,}381$.

Подставляя полученные значения в формулу 10.9, получаем

$$r = \frac{178{,}3 - 57{,}1*3{,}09}{5{,}595*0{,}381} = 0{,}87\,.$$

Иногда линейный коэффициент корреляции удобно рассчитывать по итоговым суммам:

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}} \qquad (10.12)$$

В нашем примере (по данным таблицы 10.2):

$$r = \frac{1783 - \frac{571*30{,}9}{10}}{\sqrt{\left(32917 - \frac{571^2}{10}\right)\left(96{,}93 - \frac{30{,}9^2}{10}\right)}} = 0{,}87 .$$

Достаточно часто линейный коэффициент корреляции может быть рассчитан и по более простой формуле:

$$r = a_1 \frac{\sigma_x}{\sigma_y} . \qquad (10.13)$$

В примере: $r = 0{,}059 \frac{5{,}595}{0{,}381} = 0{,}87$.

Линейный коэффициент корреляции может быть рассчитан и по другим производным от указанных формул, однако методика его исчисления на результат не влияет.

Коэффициент корреляции может принимать значения от -1 до +1. При этом положительное значение коэффициента указывает на наличие прямой связи, а отрицательное – обратной.

В оценке тесноты связи обычно руководствуются следующими соотношениями:

$[r]$	связь
$< 0{,}3$	слабая
0,3:0,5	умеренная
0,5:0,7	заметная
$> 0{,}7$	высокая (тесная).

В нашем примере r = 0,87, следовательно, между признаками существует прямая тесная связь.

Учитывая, что r рассчитывается по выборке, он, как и любой выборочный показатель, подвержен случайным ошибкам. Оценка значимости линейного коэффициента корреляции производится по критерию Стьюдента:

$$t_{расч} = \frac{|r|}{\sigma_r}, \qquad (10.14)$$

где σ_r – средняя квадратическая ошибка r.

При небольшом n (n < 30) средняя ошибка:

$$\sigma_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}} \; . \qquad (10.15)$$

Тогда расчетное значение t-критерия определяется по формуле

$$t_{расч} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (10.16)$$

и сравнивается с табличным.

Условие $t_{расч} \geq t_{табл}$ должно выполняться.

В нашем примере:

$$t_{расч} = \frac{0{,}87\sqrt{8}}{\sqrt{1-0{,}87^2}} = 4{,}99$$

При n = 10 $t_{табл}$ = 3,35.

Следовательно, $t_{расч}$ > $t_{табл}$, а это означает, что полученное значение коэффициента корреляции достоверно.

Линейный коэффициент корреляции служит показателем тесноты связи в линейных зависимостях. Однако универсальным показателем тесноты связи считается теоретическое корреляционное отношение. Оно представляет собой относительную величину сравнения среднего квадратического отклонения теоретических уровней признака результата от $\bar{y}$ и среднего квадратического отклонения эмпирических уровней признака результата от $\bar{y}$:

$$\eta_{теор.} = \sqrt{\underbrace{\frac{\sum(y_x-\bar{y})^2}{n}}_{\substack{\text{факторная}\\ \text{дисперсия}}} \Big/ \underbrace{\frac{\sum(y-\bar{y})^2}{n}}_{\substack{\text{общая}\\ \text{дисперсия}}}} \; . \qquad (10.17)$$

Эта формула может быть преобразована следующим образом:

$$\eta_{теор.} = \sqrt{\frac{\sum(y_x-\bar{y})^2}{\sum(y-\bar{y})^2}} \, . \qquad (10.18)$$

Если учесть, что дисперсия эмпирического ряда характеризует общую вариацию признака-результата за счёт всех факторов (включая и фактор x), а дисперсия теоретического ряда характеризует только ту часть вариации, которая обусловлена действием фактора x, то отношение второй дисперсии к первой показывает, какую долю в общей дисперсии занимает дисперсия, вызванная фактором *x*.

Это отношение получило название «теоретический коэффициент детерминации»:

$$\eta^2_{теор.} = \frac{\sum (y_x - \bar{y})^2}{\sum (y - \bar{y})^2}. \qquad (10.19)$$

Если учесть, что остаточная дисперсия (то есть дисперсия, вызванная действием других, неучтённых факторов) может быть рассчитана по формуле

$$\sigma^2_{ост.} = \frac{\sum (y - y_x)^2}{n} \qquad (10.20)$$

по правилу сложения дисперсий:

$\sigma^2_{общая}$ = $\sigma^2_{факторная}$ (т.е. вызванная фактором х) + $\sigma^2_{остаточная}$ (вызванная другими факторами).

Следовательно:

$$\sigma^2_{факторная} = \sigma^2_{общая} - \sigma^2_{остаточная}. \qquad (10.21)$$

Тогда используемое в формуле теоретического корреляционного отношения выражение:

$$\frac{\sigma^2_{факт.}}{\sigma^2_{общ.}} = \frac{\sigma^2_{общ.} - \sigma^2_{ост.}}{\sigma^2_{общ.}} = 1 - \frac{\sigma^2_{ост.}}{\sigma^2_{общ.}}. \qquad (10.22)$$

Рассчитанное в таком виде корреляционное отношение обычно называют <u>индексом корреляции</u>:

$$R = \sqrt{1 - \frac{\sum (y - y_x)^2}{\sum (y - \bar{y})^2}} \text{ } или \text{ } R = \sqrt{1 - \frac{\sigma^2_{ост.}}{\sigma^2_{общ.}}}. \qquad (10.23)$$

<u>Индекс корреляции</u> применяется для оценки тесноты связи линейной и нелинейной, парной и множественной.

Индекс корреляции может находиться в пределах от 0 до 1:

R = 1 – связь функциональная $\sigma^2_{ост.} = 0$,

R = 0 – связь отсутствует $\sigma^2_{ост.} = \sigma^2_{общая}$.

Интерпретация индекса корреляции обычно производится аналогично коэффициенту корреляции.

Рассчитаем индекс корреляции для нашего примера (по данным таблицы 10.2.):

$$\sum(y-\bar{y})^2 = 1{,}4490,$$
$$\sum(y-y_x)^2 = 0{,}3420,$$

$$R=\sqrt{1-\frac{0{,}3420}{1{,}4490}}=0{,}87.$$

10.5 Непараметрические методы оценки тесноты связи

Оценка тесноты связи с помощью дисперсионного и корреляционного анализа достаточно сложна и громоздка. Для измерения тесноты связи между исследуемыми признаками могут быть использованы менее точные, но более простые методы.

Методы дисперсионного и корреляционного анализа основаны на вычислении параметров распределения (средних величин, дисперсий) и поэтому их называют параметрическими методами оценки тесноты связи.

В свою очередь, методы измерения тесноты связи между признаками, которые не предусматривают использование количественных значений признаков (а следовательно, и параметров распределения), принято называть непараметрическими. К ним относят такие показатели, как:

- коэффициент корреляции знаков;
- коэффициент корреляции рангов;
- коэффициент ассоциации
- и другие.

Коэффициент корреляции знаков (коэффициент Фехнера) предполагает установление знаков отклонений каждого значения *x* от $\bar{x}$ и каждого значения *y* от $\bar{y}$. Затем определяется число единиц изучаемой совокупности, у которых эти знаки совпадают – *C*, и число единиц, у которых они не совпадают – *H*. Коэффициент Фехнера определяется по формуле

$$K_{Фех.}=\frac{C-H}{C+H}. \qquad (10.24)$$

Очевидно, что $C + H = n$ (число единиц наблюдения).

В нашем примере (таблица 10.2, графы 10, 11):
$C = 8$.
$H = 2$.
$K_{Фех.}=\frac{8-2}{8+2}=0{,}6$.
Коэффициент Фехнера меняется в диапазоне $-1 < K_ф < 1$,
$K_ф = 0{,}6$ означает, что связь прямая и достаточно тесная.

Коэффициент корреляции рангов. Известны два коэффициента: Спирмена и Кендэлла, но наиболее распространенным является коэффициент Спирмена.

Он определяется по рангам. Ранг – это порядковый номер, присваиваемый каждому индивидуальному значению *x* и *y* в ранжированных рядах. Если несколько значений одинаковых, то их ранги определяются делением приходящейся на них суммы мест на число значений признака (таблица 10.2, графы 12, 13). После того, как определены по каждой единице совокупности ранги *x* и *y*, определяют их разность *d* (для каждой единицы) (таблица 10.2, графа 14). Коэффициент ранговой корреляции определяется по формуле

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}. \qquad (10.25)$$

В нашем примере:

$$\rho = 1 - \frac{6 * 8{,}5}{10(100 - 1)} = 0{,}948.$$

Следовательно, связь прямая, тесная.

Основное преимущество коэффициента ранговой корреляции: он может быть использован там, где нет возможности измерить признак, но можно его проранжировать, например, оттенки цветов.

<u>Коэффициент ассоциации</u> применяется для оценки тесноты связи между альтернативными признаками. В этих случаях строится 4-клеточная таблица, в которой отражена связь между двумя альтернативными признаками.

Например, необходимо установить, имеется ли зависимость между семейным положением работников и обеспеченностью их жильем, если из 65 семейных обеспечены отдельными квартирами 55 человек, а из 40 одиноких квартиры имеют 25 человек.

Таблица 10.4 – Группировка работников по семейному положению и обеспеченности жильем

Обеспеченность жильём / Семейное положение	Имеют отдельную квартиру	Не имеют отдельной квартиры
Семейные	(*a*) 55	(*b*) 10
Одинокие	(*c*) 25	(*d*) 15

Каждая клетка этой таблицы имеет условное обозначение: *a, b, c, d.*

Коэффициент ассоциации определяют по формуле

$$K_a = \frac{ad - bc}{ad + bc}. \qquad (10.26)$$

В нашем примере:

$$K_a = \frac{55*15 - 10*25}{55*15 + 10*25} = \frac{575}{1075} = 0{,}535.$$

Следовательно: связь есть, связь прямая.

На практике известны случаи, когда один из квадратов такой таблицы может оказаться пустым (= 0), тогда K_a = 1. В этих случаях прибегают к исчислению коэффициента контингенции:

$$K_{конт.} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} . \qquad (10.27)$$

В нашем примере:

$$K_{конт.} = \frac{55*15 - 10*25}{\sqrt{(55+10)(25+15)(55+25)(10+15)}} = \frac{575}{\sqrt{65*40*80*25}} = \frac{575}{2280} = 0{,}252.$$

На практике значение $K_{конт.}$ всегда меньше значения K_a.

Некоторые авторы [18, с. 85] приводят только один коэффициент для оценки тесноты связи альтернативных признаков, называя его коэффициентом ассоциации:

$$K_{ас.} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} . \qquad (10.28)$$

10.6 Понятие множественной корреляции

Уровень социально-экономических явлений складывается под воздействием целого ряда факторов, часто взаимодействующих между собой. Попытка определить совокупное влияние нескольких признаков-факторов на признак-результат получила название множественной корреляции.

Если в случае парной корреляции необходимо получить уравнение регрессии $y = f(x)$, то при исследовании множественной корреляции это уравнение имеет вид:

$$y = f(x_1, x_2, x_3 \ldots x_m),$$

где m – число исследуемых признаков-факторов.

Наиболее сложный момент – это отбор признаков-факторов. Он производится в несколько этапов:

1. На основе теоретического (логического) анализа выбираются факторы, которые могут влиять на признак-результат и интересуют исследователя.

2. Производят проверку наличия достаточной связи между признаком-результатом и каждым признаком-фактором с помощью линейного коэффициента корреляции.

3. Производят проверку факторов на коллинеарность, то есть на наличие тесноты связи между двумя факторами или мультиколлинеарность – наличие тесноты связи между несколькими факторами из числа взаимосвязанных

факторов в модели оставляют те, которые в наибольшей степени интересуют исследователя.

4. Строятся различные варианты многофакторных моделей, и если они противоречат теории или логике, производится поочерёдное исключение факторов и выбирается наиболее приемлемая модель.

Экономическая интерпретация полученного уравнения регрессии производится по коэффициентам a_1, a_2, a_3…, которые указывают размер изменения признака-результата при изменении *j*-го фактора на единицу.

Оценка тесноты связи – по коэффициенту множественной корреляции *R*. Например, в случае двухфакторной модели формула его расчета имеет вид:

$$R_{y(x_1,x_2)} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} * r_{yx_2} * r_{x_1x_2}}{1 - r_{x_1x_2}^2}},$$

Как правило, расчёты выполняются на ЭВМ по стандартным программам. Задача исследователя состоит в следующем:

а) правильное определение факторов;

б) набор необходимого количества данных;

в) умение выбрать модель, наиболее полно решающую задачи исследования и имеющую теоретическое (логическое) объяснение.

Многофакторная модель показывает, изменение какого фактора даёт наиболее ощутимое изменение результата, а следовательно, на что должны быть направлены усилия менеджеров при управлении признаком-результатом.

ЛИТЕРАТУРА

1. Батуева, А. Д. Статистика : учеб. пособие / А. Д. Батуева, Е. П. Петецкая, М. А. Кокарев. – Москва : Издательство «Экзамен», 2008. – 255 с.

2. Доннели, Р. А. Статистика / Р. А Доннели; пер. с англ. Н. А. Ворониной. – Москва : АСТ : Астрель, 2007. – 367 с.

3. Елисеева, И. И. Общая теория статистики : учебник / под ред. И. И. Елисеевой, И. И. Юзбашева. – Москва : Финансы и статистика, 1995. – 368 с.

4. Ефимова, М. Р. Практикум по общей теории статистики : учеб. пособие / М. Р. Ефимова. – Москва : Финансы и статистика, 2000. – 280 с.

5. Касаева, Т. В. Статистика предприятия : курс лекций / Т. В. Касаева. – Витебск : УО «ВГТУ», 2007. – 151 с.

6. Республика Беларусь. Законы. О государственной статистике : Закон Республики Беларусь, 28 ноября 2004 г. // Национальный реестр правовых актов Республики Беларусь. – 2004. – № 192.

7. Республика Беларусь. Об утверждении инструкции по анализу и контролю за финансовым состоянием и платежеспособностью субъектов предпринимательской деятельности : Постановление Министерства Финансов Республики Беларусь, Министерства Экономики Республики Беларусь, Министерства Статистики и анализа Республики Беларусь : принято 14.05.2004г. № 81/128/65 // Нац. реестр правовых актов РБ. – 2004. – № 90.

8. Ряузов, Н. Н. Общая теория статистики : учебник / Н. Н. Ряузов.− 4-е изд., перераб. и доп. – Москва : Финансы и статистика, 1984. – 343 с.

9. Общая теория статистики. Статистическая методология в изучении коммерческой деятельности : учебник / А. И. Харламов [и др.] ; под ред. А. А. Спирина. – Москва : Финансы и статистика, 1994. – 296 с.

10. Общая теория статистики : учебник / А. Я. Боярский [и др.] ; под ред. А. М. Гольдберга. – Москва : Финансы и статистика, 1985. – 367 с.

11. Переяслова, И. Г. Основы статистики : учеб. пособие / И. Г. Переяслова, Е. Б. Колбачев. – Ростов-на-Дону : Феникс, 1999. – 320 с.

12. Годин, А. М. Статистика : учебник / А. М. Годин. – Москва : Дашков и К, 2005. – 472 с.

13. Переяслова, И. Г. Статистика : учеб. пособие / И. Г. Переяслова, Е. Б. Колбачев, О. Г. Переяслова. – Ростов-на-Дону : Феникс, 2005. – 282 с.

14. Колесникова, И. И. Статистика : учеб. пособие / И. И. Колесникова, Г. В. Круглякова. – Москва : Новое знание, 2006. – 208 с.

15. Статистика : учеб. пособие / Л. П. Харченко [и др.] ; Новосибирская государственная академия экономики и управления. – Москва : Инфра, 2005. – 384 с.

16. Статистика. Показатели и методы анализа : справ. пособие / Н. Н. Бондаренко [и др.] ; под ред. М. М. Новикова. – Минск : Современная школа, 2005. – 628 с.

17. Статистика промышленности : учебник / В. Е. Адамов [и др.] ; под ред. В. Е. Адамова. – Москва : Финансы и статистика, 1987. – 320 с.

18. Сиденко, А. В. Статистика : учебник / А. В. Сиденко, Г. Ю. Попов, В. М. Матвеева. – Москва : Издательство «Дело и Сервис», 2000. – 464 с.

19. Статистика : национальные счета, показатели и методы анализа : справочное пособие / Н. П. Дащинская [и др.] ; под ред. И. Е. Теслюка. – Минск, 1995. – 340 с.

20. Теория статистики : учебник / под ред. проф. Г. Л. Громыко. – Москва : ИНФРА-М, 2000. – 414 с.

21. Чичкан, Л. Г. Статистика в промышленности : учебно-методическое пособие для студентов вузов / Л. Г. Чичкан. – Минск : НО ООО «БИП-С», 2002. – 110 с.

22. Шимко, П. Д. Статистика / П. Д. Шимко, М. П. Власов. – Ростов-на-Дону : Феникс, 2003. – 448 с.

23. Экономика и статистика фирм / под ред. проф. С. Д. Ильенковой. – Москва : Финансы и статистика, 2002. – 256 с.

Учебное издание

Основы статистики

Конспект лекций

Составитель:
Касаева Тамара Васильевна

Редактор *А. А. Кахро*
Технический редактор *Т. В. Лапехо*
Корректор *Т. А. Осипова*
Компьютерная верстка *Н. С. Васильева*

Подписано к печати _______. Формат 60x84 1/16. Бумага офсетная № 1. Гарнитура «Таймс». Усл. печ. л. ____. Уч.-изд. л. _____. Тираж _____ экз. Заказ № ____.

Учреждение образования «Витебский государственный технологический университет» 210035, г. Витебск, Московский пр., 72.

Отпечатано на ризографе учреждения образования ”Витебский государственный технологический университет”.
Лицензия № 02330/0494384 от 16 марта 2009 г.