СТАТИСТИКА

УЧЕБНОЕ ПОСОБИЕ
ДЛЯ ИНОСТРАННЫХ СТУДЕНТОВ

ПОДГОРНЫЙ А. З.

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,
МОЛОДЕЖИ И СПОРТА УКРАИНЫ

ОДЕССКИЙ НАЦИОНАЛЬНЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ

А.З. Подгорный
О.Г. Мылашко
С.М. Киршо
Н.М. Шилофост

СТАТИСТИКА

**УЧЕБНОЕ ПОСОБИЕ
ДЛЯ ИНОСТРАННЫХ СТУДЕНТОВ**

Одесса
2012

УДК 31(075.8)
ББК 60.6я73
П 44

Рекомендовано Министерством образования и науки, молодежи и спорта Украины в качестве пособия для студентов высших учебных заведений от 24.10.2012 №1/11 – 16544

Рецензенты:

Бараник З.П., доктор экономических наук, профессор (Киевский национальный экономический университет им. Вадима Гетьмана),
Сидорова А.В., доктор экономических наук, профессор (Донецкий национальный университет)

Подгорный А.З., Мылашко О.Г.,
Киршо С.М., Шилофост Н.М.
Статистика: Учебное пособие для иностранных студентов. – Одесса: Атлант, 2012. – 195 с.

В учебном пособии изложены теоретические основы статистики. Освещены вопросы теории статистического наблюдения, методы исследования вариационных рядов. Особое внимание уделено статистическим группировкам, методам статистического исследования зависимостей, анализу динамики и прогнозирования динамических рядов. Представлены типовые примеры с решениями и контрольные вопросы по изучаемому материалу.

Для иностранных студентов экономических специальностей, преподавателей, аспирантов.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В современном обществе статистика является одним из важнейших инструментов управления национальной экономикой. Она позволяет давать всестороннюю характеристику изучаемому явлению, отмечать успехи и недостатки, намечать пути и мероприятия по устранению нежелательных тенденций. Статистика служит основой для оценки реальной экономической ситуации на всех уровнях хозяйствования – от микроуровня (домохозяйство, предприятие) до макроуровня (экономика страны).

Дисциплина «Статистика» играет важнейшую роль в формировании высококвалифицированного экономиста, являясь одной из фундаментальных экономических наук. Ее изучают студенты всех экономических специальностей. Эта наука вооружает будущего специалиста знаниями методов и способов сбора, обработки и анализа экономической информации. После изучения курса «Статистика» студент способен решать следующие задачи: – разрабатывать программу статистических наблюдений; – осуществлять сводку и группировку массовых данных о процессах общественной жизни; – рассчитывать обобщающие характеристики структуры совокупностей; – измерять интенсивность динамики явлений; определять факторы, формирующие развитие общественных явлений и оценивать силу их влияния.

Структура учебного пособия учитывает требования методики преподавания курса «Статистика» иностранным студентам.

Для лучшего освоения материала и формирования навыков использования теоретических знаний на практике в каждой теме приведены решения типовых задач. Это дает возможность лучше понять суть расчетных формул, оценить конечные результаты с экономической и социальной точки зрения. После каждой темы предложены вопросы для самоконтроля, что делает пособие полезным в самостоятельной работе студентов.

РАЗДЕЛ 1. МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ СТАТИСТИКИ

1.1. ПРЕДМЕТ СТАТИСТИКИ

История развития человечества показала, что без статистических данных невозможно управление государством, отраслью экономики, предприятием. Статистические методы позволяют оценивать состояние явления, его структуру, интенсивность развития, разрабатывать стратегию роста как на микро – так и на макроуровне. Статистика применяется в разных сферах, областях экономики, но методы статистической работы, правила сбора, обработки и анализа информации общие.

Слово «статистика» латинского происхождения (от *status* – состояние, положение вещей). От корня этого слова образовалось итальянское слово *stato* – государство. Государственных деятелей, политиков, которые обладают знаниями об устройстве и состоянии дел в государстве, называли *statista*.

Возникновение статистики началось с зарождения хозяйственного учета. Для государственной деятельности, а именно: сбора налогов, набора войск – необходимо было определить численность населения, состав земель, поголовье скота, иметь данные о торговле. В Китае такие сведения начали собирать более 4 тысяч лет назад, в Древней Греции – 2,5–3 тысячи лет назад. Постепенно расширялся круг данных, необходимых для государственной деятельности. Возникли правила наблюдения и методы анализа показателей, что и составило суть статистической науки.

Первыми научными статистическими школами были: **немецкая описательная (государствоведение)** и **английская школа политических арифметиков**.

Видными представителями описательной школы были Г. Конринг (1606-1661), Г. Ахенваль (1719-1772), А. Бюшинг (1724-1793). Задачей статистики они считали описание «государственных достопримечательностей», а именно: территории государства, населения, религии, внешней политики. Именно Г. Ахенвалем, профессором Геттингенского университета, в 1743 г. был введен термин «статистика» для обозначения совокупности знаний, которые характеризуют государственное устройство. Вместе с тем представители описательной школы недооценивали математические свойства познания. Количественные оценки трактовались ими как частный случай общего описания.

Значительно ближе к современному пониманию статистики стала английская школа политических арифметиков. Ее основатели (У. Петти (1623-1687), Дж. Граунт (1620-1674)) ставили целью изучать общественные явления с помощью числовых характеристик. Их важнейшей заслугой явилось понимание необходимости массовых данных для получения устойчивых выводов и организации специального наблюдения.

Исследования представителей двух школ статистики создали предпосылки для обобщения накопленного опыта и создания научной теории статистики. Таким образом, **статистика сложилась из элементов политической арифметики и государствоведения**. Ее дальнейшее развитие характеризовалось совершенствованием методов и способов сбора и обработки данных, необходимых для анализа разнообразных массовых социально-экономических процессов и явлений.

Сегодня термин «статистика» используется в трех значениях:

1) цифровой материал, который характеризует какую-либо сторону общественной жизни;
2) отрасль практической деятельности, цель которой сбор, обработка, анализ, публикация массовых данных о различных общественных явлениях;
3) наука, которая имеет свой предмет и метод.

Предметом статистики является количественная сторона массовых социально-экономических явлений в неразрывной связи с качественной стороной в конкретных условиях места и времени.

Основным в определении предмета статистики является следующее:

- статистика изучает явления с **количественной стороны**. Количественную сторону явления определяют его размеры, уровни, объемы, темпы развития и т.д.;
- статистика исследует **качественную сторону** массовых явлений, которая и выражает их специфику, внутреннюю особенность, отличающую данное явление от других;

- статистика изучает не отдельные факты, а **массовые** явления, так как именно в них проявляются закономерности изменения количественных характеристик;
- все общественные явления протекают во времени и пространстве, и в отношении любого из них всегда можно установить, **когда** оно возникло и **где** оно развивалось. Следовательно, статистика изучает явления **в конкретных условиях места и времени**.

1.2. ЭТАПЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ И МЕТОДЫ СТАТИСТИКИ

Теоретической основой статистики как общественной науки являются философия, экономическая теория, которая включает политическую экономию, макро- и микроэкономику. На основе этих наук статистика выявляет количественные изменения общественных явлений, но использует свою, специфическую статистическую методологию.

Статистическая методология – это система приемов и методов, которые направлены на изучение количественных закономерностей, проявляющихся в структуре, динамике и взаимосвязях социально-экономических явлений.

Любое статистическое исследование осуществляют в **3 этапа**:

1. Статистическое наблюдение.
2. Сводка и группировка данных наблюдения.
3. Анализ полученных сводных материалов.

Прохождение каждого этапа исследования связано с использованием специальных методов статистики.

Первый этап статистического исследования проводят с помощью метода массового наблюдения. **Статистическое наблюдение** представляет собой планомерное, научно-организованное собирание сведений об изучаемых социально-экономических явлениях и процессах.

Важным методом, который используется в ходе проведения второго этапа исследования, является **группировка**. Группировка позволяет выделить однородные статистические совокупности. На данном этапе также используют методы наглядного представления информации, такие как **табличный** и **графический**, а также широко используется **балансовый** метод.

Статистический анализ – заключительная стадия статистического исследования. В процессе проведения анализа используют следующие статистические методы:

- **методы расчета обобщающих статистических показателей;**
- **методы анализа динамики;**
- **методы анализа вариации;**
- **индексный метод;**
- **методы изучения взаимосвязей между факторами.**

Изучению этапов статистического исследования и методов, которые используются на каждом из этапов, посвящены последующие разделы данного учебного пособия.

1.3. ОСНОВНЫЕ КАТЕГОРИИ СТАТИСТИКИ

Важнейшими категориями статистики являются:

- статистическая закономерность;
- статистическая совокупность;
- единица совокупности;
- признак;
- вариация;
- показатель;

Статистическая закономерность – это количественная закономерность изменения массовых явлений и процессов общественной жизни в пространстве и во времени в результате действия объективных законов.

Статистическая закономерность свойственна не отдельным единицам, а всей совокупности в целом и проявляется только при достаточно большом числе наблюдений, то есть в основе статистической закономерности лежит **закон больших чисел**. Объективной основой существования статистических закономерностей является переплетение, взаимосвязь основных и случайных причин. При достаточно большом числе событий влияние случайных причин взаимно уравновешивается, благодаря чему закономерности становятся видимыми.

Статистическая совокупность – это множество социально-экономических объектов или явлений общественной жизни, которые объединены общей качественной основой, но отличаются друг от друга отдельными признаками.

В это определение входят **три основные черты совокупности**:

– во-первых, это **множество** явлений;

– во-вторых, это множество явлений, которые объединены **общим качеством**;

– в-третьих, эти явления, объекты **отличаются по своим характеристикам**.

Например, статистической совокупностью являются предприятия гостиничного хозяйства Одесской области. Данное множество предприятий связывает то, что все предприятия, которые входят в совокупность, расположены в одном регионе. Все предприятия оказывают услуги по размещению, обладают многими другими общими характеристиками. В то же время, каждое предприятие отличается от другого набором и качеством услуг, которые оно предоставляет, ценами, количеством персонала и его заработной платой и т.д.

Статистическая совокупность состоит из единиц совокупности.

Единица совокупности – носитель признака и основа счета.

Под термином **признак** в статистике понимается свойство, характерная черта единиц совокупности.

В нашем примере отдельные предприятия являются единицами совокупности всех предприятий гостиничного хозяйства области. Каждое из них является носителем всех признаков, которые характеризуют совокупность. Примерами признаков, которые отличают предприятия одно от другого, являются, например, форма собственности, стоимость основного капитала, зарплата работников и т.д.

Объектом статистического изучения являются именно признаки.

Виды признаков:

1. **Статические и варьирующие (изменяющиеся)**.

Статические признаки характеризуются неизменными значениями у всех единиц совокупности.

Вариация – изменение, колеблемость признака у единиц совокупности. Признаки, значения которых меняются в пространстве или во времени, называются **варьирующими**. Значения признака у отдельных единиц совокупности называются **вариантами**. Например, вариантами признака *форма собственности гостиниц* являются частная и государственная формы; признака *число номеров в гостиницах* – 10 номеров, 20 номеров, 50 номеров и т.п.

2. **Атрибутивные и количественные**.

Атрибутивные – это признаки, варианты которых выражены словесно. Совокупность предприятий гостиничного хозяйства можно охарактеризовать такими атрибутивными признаками: форма собственности, место расположения (к какому району города или области относится гостиница), страна проживания иностранных посетителей и т.д.

Количественные – признаки, варианты которых выражены числами.

Количественные признаки делятся на дискретные и непрерывные.

Варианты **дискретных** признаков выражаются целочисленными значениями: число гостиниц, количество номеров, число обслуженных посетителей.

Варианты **непрерывных** признаков могут принимать любые значения: рентабельность предприятий, товарооборот ресторанов при гостиницах и т.д.

Кроме признаков, состояние совокупности, которая исследуется, характеризуют показатели. **Показатель** – обобщенная количественно-качественная характеристика социально-экономического явления или процесса в определенных условиях места и времени. Виды показателей будут нами далее рассмотрены в соответствующих темах.

Показатели и признаки в полной мере характеризуют статистическую совокупность, позволяют проводить ее исследование, что и является задачей статистической науки.

1.4. ОРГАНИЗАЦИЯ СТАТИСТИКИ В УКРАИНЕ

По своему назначению государственная система статистики должна быть одновременно **инструментом** фиксации ситуации, которая складывается в стране; эффективного контроля над изменениями в социально-экономической жизни общества; прогнозирования перспектив развития страны.

Главным органом государственной статистики Украины является **Государственная служба статистики Украины**.

Задачей службы является развитие статистической системы в стране по следующим направлениям:

- определение государственной политики в сфере статистики;
- координация статистической деятельности в стране;
- сбор, обработка, сохранение, защита, распространение и анализ статистической информации;
- обеспечение достоверности и объективности статистических данных;
- разработка, внедрение и совершенствование статистической методологии;
- разработка, внедрение и совершенствование системы статистических классификаций;
- разработка и ведение Единого государственного реестра предприятий и организаций.

Статистическая система состоит из трех уровней – центральный аппарат, региональные управления и местные отделения статистики.

Структуру центрального аппарата составляет 21 департамент, каждый из которых включает несколько отделений.

Второй уровень статистической системы представлен 28 региональными управлениями статистики и включает Главное межрегиональное управление статистики в г. Киеве, Главное управление статистики в Автономной республике Крым, 24 областные управления статистики, а также управление в г. Киеве и Севастополе. Региональные управления статистики отвечают за сбор, проверку и передачу данных в

Государственную службу статистики Украины, а также за распространение этих данных на территориальном уровне.

Третий уровень – местные органы государственной статистики – включает 587 отделений статистики районного и городского уровней.

Государственная служба статистики финансируется за счет средств госбюджета. Эти средства выделяются на проведение обследований, включенных в ежегодный план государственных статистических наблюдений.

Государственная служба статистики ежемесячно готовит доклад и статистический бюллетень о социально-экономическом развитии Украины, а также оперативную статистическую информацию.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Когда возник термин «статистика» и кем он был предложен?
2. Поясните сущность описательного направления в развитии статистики и назовите его представителей.
3. В чем сущность английской школы политических арифметиков?
4. Что обозначает термин «статистика» сегодня?
5. Сформулируйте предмет статистики и поясните основные положения данного определения.
6. Назовите этапы статистического исследования и методы, которые применяют на каждом из этапов.
7. Что такое статистическая закономерность и в чем выражается закон больших чисел?
8. Дайте определение статистической совокупности и приведите примеры.
9. Поясните, что такое единица совокупности и приведите примеры единиц различных статистических совокупностей.
10. Сформулируйте понятие признака и варианта в статистике.
11. Какие виды признаков вы знаете? Поясните различие между ними.
12. Как организована система государственной статистики в Украине?

РАЗДЕЛ 2. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

2.1. ПОНЯТИЕ О СТАТИСТИЧЕСКОМ НАБЛЮДЕНИИ

Первой стадией любого статистического исследования является статистическое наблюдение.

Статистическое наблюдение – планомерная, научно-организованная регистрация данных о массовых социально-экономических явлениях и процессах.

Статистическое наблюдение может быть первичным и вторичным.

Первичное статистическое наблюдение – это регистрация данных, которые поступают непосредственно от объекта, который эти данные продуцирует (например, текущий учет рождаемости).

Вторичное статистическое наблюдение – это сбор ранее собранных и обработанных данных (например, основой вторичного наблюдения могут быть данные статистических ежегодников).

Итак**, задачей** статистического наблюдения является получение данных, на основании которых можно исследовать процесс, который изучается. В связи с этим к статистическим данным предъявляются следующие **требования**:

- достоверность;
- полнота;
- своевременность;
- сопоставимость во времени и в пространстве;
- сопоставимость по единицам измерения;
- доступность.

Процесс подготовки наблюдения начинается с составления плана наблюдения.

План статистического наблюдения – совокупность программно-методологических и организационных вопросов.

2.2. ПРОГРАММНО-МЕТОДОЛОГИЧЕСКИЕ ВОПРОСЫ ПЛАНА СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

Программно-методологические вопросы определяют:

- цель исследования;
- объект наблюдения;
- единицу совокупности;
- источник информации (единицу наблюдения);
- программу наблюдения;
- инструментарий наблюдения.

Целью исследования является получение достоверной информации для выявления закономерностей развития явлений и процессов. Например, цель переписи населения страны - получение данных о численности и составе населения.

Объектом наблюдения является статистическая совокупность, в которой происходят массовые явления и процессы, которые исследуются. Например, объектом наблюдения может быть совокупность населения страны, совокупность учебных заведений, совокупность промышленных предприятий.

Объект наблюдения состоит из отдельных элементов – единиц совокупности.

Единица совокупности –носитель признаков, которые должны регистрироваться. Например, при переписи населения единицей совокупности является каждый человек.

Единица наблюдения – первичная единица, от которой получают информацию. При переписи населения единицей наблюдения может выступать домохозяйство или же отдельный его член.

Программа наблюдения – перечень (список) вопросов, на которые должны быть получены ответы. Так, в программе переписи населения содержатся вопросы о поле, возрасте, образовании, источниках средств существования и другие.

Программа статистического наблюдения оформляется в виде документа – статистического формуляра. Это необходимо, во-первых, для однообразия сведений, которые получают от каждой единицы, и, во-вторых, для дальнейшей обработки информации.

Статистический формуляр – это документ единого образца, в котором содержится программа и регистрируются ответы на вопросы наблюдения.

К формуляру разрабатывается **инструкция**, в которой определены порядок проведения наблюдения и заполнения форм.

Набор статистических формуляров и инструкций образует **инструментарий программы статистического наблюдения**.

2.3. ОРГАНИЗАЦИОННЫЕ ВОПРОСЫ ПЛАНА СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

Второй частью плана статистического наблюдения являются **организационные вопросы**, которые определяют:

- органы и персонал, которые проводят наблюдение;
- место наблюдения;
- материально-техническое обеспечение;
- систему контроля, которая обеспечивает точность результатов;
- время и период наблюдения.

Органами, которые обеспечивают проведение наблюдения, могут быть центральные и местные органы статистики, министерства и ведомства, учреждения, организации, предприятия.

Место наблюдения – это место, где должна производиться регистрация фактов, которые наблюдаются, и заполняются формуляры.

Время наблюдения – это время, когда происходят факты, которые регистрируются.

Если объектом наблюдения является **процесс**, то время наблюдения определяется **интервалом времени**. Если же объектом наблюдения является **состояние**, то выбирается, так называемый, **критический момент времени**. Например, число детей, которые родились в регионе, определяется за определенный интервал времени (например, за год), а численность населения – на **момент времени** (например, этот момент – «0» часов с 11 на 12 января).

Так как зарегистрировать численность населения одномоментно невозможно, то, кроме времени наблюдения, устанавливается период наблюдения.

Период наблюдения (субъективное время) – время, на протяжении которого осуществляется регистрация данных. Так, субъективное время переписи населения обычно 10 дней.

2.4. ФОРМЫ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

В отечественной статистике используют такие **3 организационные формы статистического наблюдения**:

- отчетность;
- специально организованное наблюдение;
- реестры.

Отчетность – форма статистического наблюдения, которая характеризуется тем, что каждый субъект деятельности регулярно предоставляет свои данные в государственные органы статистики в виде отчетов специально утвержденной формы.

Отчетность носит обязательный характер и базируется на документах первичного учета.

Статистическая отчетность бывает ежедневная, недельная, двухнедельная, месячная, квартальная и годовая.

Специально организованное наблюдение представляет собой регистрацию данных, которую организуют статистические органы:

- для изучения явлений, которые не отражены в отчетах;
- для более глубокого изучения отчетных данных, их проверки и уточнения.

К специально организованным наблюдениям относят переписи населения, материальных ресурсов, оборудования, многолетних насаждений, незавершенного строительства.

Реестры – это форма непрерывного статистического наблюдения за долговременными процессами в рамках конкретного времени. Реестр является инструментом, который постоянно следит за состоянием единиц наблюдения, каждая из которых характеризуется совокупностью показателей. Все показатели хранятся до тех пор, пока единица наблюдения находится в реестре и не закончила своего существования. Некоторые показатели остаются неизменными все время, пока единица наблюдения находится в реестре, другие могут меняться время от времени.

В практике статистики различают реестры населения и реестры предприятий.

Так, реестры предприятий включают данные о времени создания (регистрации) предприятия, название, адрес, телефон, вид экономической деятельности и другие.

2.5. ВИДЫ И СПОСОБЫ НАБЛЮДЕНИЯ

Статистическое наблюдение представляет собой процесс, который с точки зрения его организации может иметь различные виды и способы проведения.

Виды наблюдения по степени охвата единиц совокупности:

- сплошное;
- несплошное.

Сплошное наблюдение предусматривает регистрацию всех без исключения единиц совокупности. К сплошному наблюдению относится отчетность, большинство переписей. Недостатком данного вида наблюдения является высокая стоимость сбора и обработки всего массива информации и недостаточная оперативность, скорость получения данных.

Несплошное – наблюдение, при котором регистрируется лишь часть единиц совокупности, которую определенным образом отобрали. Такой вид наблюдения дает возможность получения информации в более краткие сроки и с меньшими расходами, чем при проведении сплошного наблюдения.

Основными видами организации несплошного наблюдения являются:

- выборочное наблюдение;
- наблюдение основного массива;
- монографическое наблюдение.

Выборочное – это наблюдение, когда регистрируется определенная часть единиц совокупности, которая выделяется методом случайного отбора. При правильной организации выборочное наблюдение дает достаточно точные результаты, которые можно распространить с определенной вероятностью на всю совокупность.

Наблюдение основного массива – регистрация данных наиболее важных единиц совокупности, которые по основному для наблюдения признаку имеют наибольший вес в совокупности. Например, наблюдение за ценами на больших рынках, обследование городов с наибольшим уровнем загрязнения окружающей среды.

Монографическое наблюдение – детальное обследование отдельных единиц совокупности, которые обладают какими-либо особенными характеристиками или представляют какое-либо новое явление. Целью такого наблюдения является выявление тенденций, которые уже существуют или только зарождаются в развитии данного процесса. Например, наблюдение за деятельностью отдельно взятой фермы с высокими показателями производства молока.

Виды наблюдения по времени регистрации фактов:

- непрерывное;
- прерывное.

Непрерывное (текущее) наблюдение осуществляется путем непрерывной регистрации фактов по мере их возникновения. Например, регистрация процессов производства и реализации продукции, поступления платежей.

Прерывное наблюдение проводится либо регулярно, через определенные промежутки времени (**периодическое** наблюдение), либо нерегулярно, однократно, по мере необходимости (**единовременное** наблюдение). Примером периодического наблюдения является перепись населения, к единовременным наблюдениям относятся различные маркетинговые обследования сегментов рынка.

Статистическая информация может быть получена различными **способами**, основными из которых являются:

- непосредственное наблюдение;
- документальный учет;
- опрос.

При **непосредственном наблюдении** учет фактов осуществляется лично регистратором путем подсчета, измерения, осмотра, оценки. Такой учет используется при регистрации товарных запасов, наличной денежной массы в кассах и т.д.

Документальный учет основан на использовании в качестве источника информации данных различных документов. Так, по документам первичного учета составляется статистическая отчетность об объемах материальных, трудовых и финансовых ресурсов.

Опрос – способ наблюдения, в соответствии с которым информацию получают со слов респондента (человека, который отвечает на вопросы), когда явления и процессы невозможно непосредственно наблюдать. Он характерен для различного вида социологических обследований, опросов общественного мнения. Различают следующие **виды опроса**:

- экспедиционный;
- саморегистрация;
- корреспондентский;
- анкетный.

Экспедиционный опрос – регистрация фактов специально подготовленными счетчиками с одновременной проверкой точности регистрации. Например, в Украине и во многих других странах перепись населения осуществляют с помощью экспедиционного опроса.

Саморегистрация – регистрация фактов самими респондентами после предварительного инструктажа, разъяснения. Данный вид опроса применяется, например, при обследовании бюджетов домохозяйств.

Корреспондентский опрос – регистрация фактов на местах их возникновения лицами, которые добровольно передают информацию в соответствующие инстанции. Корреспондентский опрос используется, например, для исследования рынка товаров и услуг отдельных регионов.

Анкетный опрос – регистрация мнений, соображений, намерений лицами, которые получили анкету. Данный вид опроса часто используется при социологических обследованиях.

2.6. ОШИБКИ НАБЛЮДЕНИЯ

Важнейшей задачей статистического наблюдения является обеспечение достоверности, правдивости первичной информации. Однако в процессе регистрации данных могут возникнуть некоторые неточности, ошибки наблюдения.

Ошибкой наблюдения называется расхождение между расчетным и реальным значением признака, который изучается.

Различают **две группы ошибок наблюдения**:

1) ошибки регистрации;
2) ошибки репрезентативности.

Ошибки регистрации могут возникнуть и при сплошном, и при несплошном наблюдении.

Причины ошибок регистрации:

- неправильное установление фактов в результате наблюдения;
- неверная запись данных в формуляре.

Виды ошибок регистрации:

- случайные;
- систематические.

Случайные ошибки регистрации могут возникнуть в результате описки (случайной ошибки при письме), оговорки (случайной ошибки в речи), неточности в подсчете. Они могут быть направлены и в сторону преуменьшения, и в сторону преувеличения показателя. Однако в результате действия закона больших чисел такие ошибки взаимно погашаются, уничтожаются.

Систематические ошибки регистрации возникают в результате определенных постоянных причин, поэтому имеют одинаковую тенденцию или к преуменьшению, или к преувеличению значения признака у единиц совокупности. Направленность систематической ошибки в одну сторону делает ошибочным значение признака, который регистрируется.

Систематические ошибки регистрации могут быть преднамеренными и непреднамеренными.

Преднамеренные ошибки возникают, если лицо, которое опрашивается или регистратор, знают реальное положение вещей, но специально дают неправильную информацию.

Непреднамеренные ошибки могут быть связаны либо с невнимательностью регистраторов, либо, например, с такими явлениями при переписи населения, как преуменьшение возраста женщинами («женское кокетство»), преувеличение возраста пожилыми людьми («старческое кокетство»), округление возраста.

Ошибки репрезентативности возникают только при несплошном наблюдении. Они связаны с тем, что отобранная и обследованная совокупность недостаточно точно воспроизводит генеральную совокупность. Так же, как и ошибки регистрации, **ошибки репрезентативности бывают двух видов**:

- случайные;
- систематические.

Систематические ошибки возникают из-за тенденциозного, необъективного отбора единиц, при котором нарушаются принципы выборки.

Основной задачей выборочного метода является измерение случайных ошибок, которые возникают при выполнении всех требований проведения выборки.

Выборочному наблюдению будет посвящен специальный раздел пособия.

Для недопущения ошибок наблюдения, их выявления и исправления, необходимо обеспечить:

- четкую разработку программно-методологического и организационного плана статистического наблюдения;
- качественное обучение персонала, который осуществляет наблюдение;
- арифметический и логический контроль результатов наблюдения.

Арифметический контроль заключается в арифметической проверке итоговых и расчетных показателей, то есть он основывается на количественных взаимосвязях между значениями разных показателей. Например, размер совокупных расходов домохозяйства можно проверить, если просуммировать его расходы по отдельным статьям.

Логический контроль основан на логических взаимосвязях между показателями. Например, в переписном листе в графе *возраст* указано 10 лет, а в графе *образование* указано высшее образование. Наличие данных записей одновременно свидетельствует, что одна из них не соответствует действительности. Логический контроль устанавливает лишь наличие ошибки, а не ее величину.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Сформулируйте сущность статистического наблюдения и охарактеризуйте его задачи.
2. Что представляет собой первичное и вторичное наблюдение?
3. Что включают в себя программно-методологические вопросы плана статистического наблюдения?
4. Что представляет собой объект статистического наблюдения? Поясните различие понятий *единица совокупности* и *единица наблюдения.*
5. Что включает программа статистического наблюдения?
6. Какой инструментарий необходим для проведения наблюдения?

7. Что определяют организационные вопросы плана статистического наблюдения?
8. Раскройте понятия *время наблюдения* и *период наблюдения*.
9. Какие организационные формы наблюдения используют в статистике?
10. Что представляет собой статистическая отчетность?
11. Приведите примеры специально организованных наблюдений.
12. Что такое *реестры*?
13. Какие существуют виды наблюдения по степени охвата единиц совокупности?
14. Назовите виды несплошного наблюдения и поясните содержание каждого из них.
15. Охарактеризуйте виды наблюдения по времени регистрации фактов.
16. Какими способами может быть получена статистическая информация?
17. Что представляет собой непосредственное наблюдение?
18. Назовите источники информации при документальном наблюдении.
19. Сформулируйте понятие опроса и охарактеризуйте его виды.
20. Что представляют собой ошибки наблюдения? Какие две группы ошибок наблюдения различают?
21. Назовите причины и виды ошибок регистрации.
22. С каким видом наблюдения связаны ошибки репрезентативности?
23. В результате каких причин могут возникать ошибки репрезентативности?
24. Что представляет собой арифметический и логический контроль результатов наблюдения?
25. Какие меры можно предложить для недопущения или смягчения ошибок наблюдения?

РАЗДЕЛ 3. СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

3.1. СОДЕРЖАНИЕ И ЗАДАЧИ СВОДКИ

В результате статистического наблюдения мы получаем информацию, которая характеризует отдельные единицы совокупности. Для характеристики совокупности в целом, то есть для определения ее типичных черт, закономерностей, основных тенденций, данные наблюдения необходимо обобщить, упорядочить, систематизировать. Эти задачи решаются с помощью сводки – второго этапа статистического исследования.

Сводка – комплекс последовательных действий по систематизации и обобщению первичных данных с целью выявления типичных черт и закономерностей, присущих явлению в целом.

В узком смысле слова сводка – это подсчет итогов.

В широком смысле слова сводка включает:

1. Разработку программы систематизации первичного статистического материала.
2. Обоснование системы показателей, характеризующих рассматриваемое явление.
3. Проектирование макетов таблиц и графиков.
4. Группировку, расчет групповых и общих итогов, обобщающих показателей.
5. Изложение материалов сводки в форме таблиц и графиков.

Для осуществления перечисленных этапов еще до начала проведения сводки составляется **программа статистической сводки**, содержащая перечень групп, их границы в соответствии с группировочными признаками, систему показателей, характеризующих совокупность, и методику их вычисления, систему макетов таблиц, в которых будут представлены итоги расчетов.

Наряду с программой разрабатывается **план проведения сводки**. В нем рассматриваются организационные вопросы ее проведения, такие, как последовательность и сроки выполнения отдельных этапов сводки, сроки предоставления общих результатов работы, определяются ответственные за осуществление сводки, предусматривается координация работы всех организаций, задействованных в ее проведении.

Различают **виды статистической сводки** по ряду признаков:

- по сложности построения;
- по организации проведения;
- по технике исполнения.

По сложности построения различают простую и сложную сводку.

Простая сводка предусматривает лишь подсчет итогов первичного материала без его систематизации. **Сложная** – включает комплекс всех процедур, связанных с проведением статистической сводки.

С точки зрения **организации проведения** различают централизованную и децентрализованную сводку.

Централизованная сводка предусматривает передачу всего первичного статистического материала в единый центр, в котором его обработка проводится по заранее разработанной программе. Централизованная сводка используется, например, при переписях населения.

При **децентрализованной сводке** обработка и обобщение первичного материала осуществляется от низшего звена статистических органов до высшего. Примером децентрализованной сводки является обработка отчетов предприятий.

По технике исполнения различают **ручную сводку** (все операции проводятся вручную) и **механизированную** (все операции исполняются и контролируются с помощью компьютеров).

3.2. СУТЬ СТАТИСТИЧЕСКОЙ ГРУППИРОВКИ

Основным элементом статистической сводки является группировка.

Статистическая группировка – это распределение единиц совокупности на группы по каким-либо существенным признакам.

Признак, по которому производится группировка, называется **группировочным**. Группировочные признаки бывают **количественными** и **атрибутивными**.

С помощью группировок решают **3 вида задач**:

- исследование структуры совокупности;
- выделение социально-экономических типов;
- анализ взаимовлияния между отдельными признаками совокупности.

В соответствии с решаемыми задачами существует **3 вида группировок**:

- структурные;
- типологические;
- аналитические.

Структурная группировка – распределение качественно однородной совокупности на группы по определенным признакам. Структурная группировка необходима в силу того, что в пределах однородной совокупности ее единицы различаются по значениям свойственных им признаков. Например, для изучения структуры промышленных предприятий изучают их распределение по объему выпускаемой продукции, прибыли, числу работников, их заработной плате. Предложенные в таблицах 3.1, 3.2 и 3.3 распределения работников одного предприятия по различным признакам также являются примерами структурных группировок. В основе первого распределения лежит атрибутивный признак – пол работников, в основе распределения, представленного в таблице 3.2, – количественный дискретный признак – тарифный разряд (уровень квалификации), в основе группировки по выработке (производство продукции в единицу времени) лежит соответствующий количественный непрерывный признак.

Таблица 3.1

Распределение работников предприятия по полу

Пол	Число рабочих	
	человек	% к итогу
Женский	180	60,0
Мужской	120	40,0
Итого	300	100,0

Таблица 3.2

Распределение работников предприятия по тарифному разряду

Тарифный разряд	Число рабочих	
	человек	% к итогу
1	30	10,0
2	100	33,3
3	80	26,7
4	40	13,3
5	35	11,7
6	15	5,0
Итого	300	100,0

Типологической называется группировка качественно неоднородной совокупности на однородные группы, классы, социально-экономические типы. Типологическая группировка применяется для изучения распределения населения по социальным группам, уровню бедности, предприятий – по формам собственности (пример данного распределения представлен в таблице 3.4).

Таблица 3.3

Распределение работников предприятия по средней дневной выработке

Группы рабочих по выработке, кг	Число рабочих	
	человек	% к итогу
Менее 600	70	23,3
600 - 1000	80	26,7
1000 - 1400	50	16,7
1400 - 1800	60	20,0
1800 и более	40	13,3
Итого	300	100,0

Таблица 3.4

Распределение предприятий отрасли по формам собственности

Форма собственности	Число предприятий	
	единиц	% к итогу
Частная	368	43,3
Коллективная	104	12,2
Государственная	255	30,0
Другие	123	14,5
Итого	850	100,0

Аналитическая группировка необходима для изучения наличия и направления связи между признаками. Эти признаки называют факторным и результативным. **Факторным** является признак, от которого зависит **результативный**. Для осуществления аналитической группировки необходимо разгруппировать единицы совокупности по факторному признаку, а затем для каждой группы рассчитать среднюю величину результативного. Аналитическая группировка дает наглядное представление о наличии прямой или обратной связи между признаками или же об ее отсутствии.

Пример аналитической группировки представлен в таблице 3.5. Данные таблицы однозначно свидетельствуют, что между факторным признаком, возрастом матери, и результативным, числом рожденных детей, существует взаимосвязь. Наибольшее число рожденных детей наблюдается у женщин в возрасте 20-25 лет, а затем, с увеличением возраста, число рождений уменьшается, то есть связь между признаками становится обратной.

Таблица 3.5

Зависимость рождаемости от возраста матери

Возраст матери, лет	Удельный вес женщин, %	Среднее число рожденных детей на 1000 женщин
До 20	16,1	33,5
20-25	14,1	82,5
25-30	15,9	60,4
30-35	16,6	30,0
35-40	15,6	10,2
40-45	13,3	2,4
Более 45	8,4	0,1
Итого (в среднем)	100,0	30,6

В зависимости от числа группировочных признаков, положенных в основу группировки, различают группировки:

- простые;
- комбинационные.

Группировка называется **простой**, если в ее основу положен один группировочный признак. Все вышеприведенные группировки являются простыми.

Если группировка проводится по нескольким признакам, она называется **комбинационной**.

При построении комбинационной группировки следует придерживаться определенных требований к ней. Эти **требования** касаются:

- **последовательности признаков**, по которым группировка осуществляется: группировку рекомендуется начинать с атрибутивного признака, причем наиболее значимого в данном исследовании;
- **числа признаков**, положенных в основу группировки: группировка, построенная более чем по трем признакам, сложна для восприятии и для анализа.

Пример комбинационной группировки дан в таблице 3.6.

Таблица 3.6

Распределение населения региона
по экономической активности и полу

(млн.)

Экономически активное население	Всего	в том числе	
		мужчины	женщины
Занятые	20,4	10,0	10,4
Безработные	2,3	1,1	1,2
Итого	22,7	11,1	11,6

3.3. ВОПРОСЫ МЕТОДОЛОГИИ ПОСТРОЕНИЯ СТАТИСТИЧЕСКИХ ГРУППИРОВОК

В процессе группировки ставятся и решаются такие **3 задачи**:

- выбор группировочного признака;
- определение числа групп;
- определение интервалов групп.

При выборе группировочных признаков надо учитывать их значение в дальнейшем анализе, а также условие места и времени. Например, при анализе уровня жизни населения одним из важнейших показателей является его материальное состояние. Для оценки материального благополучия можно произвести распределение населения по его доходам. Однако часто критерием, показателем благополучия является другой группировочный признак – расходы населения.

От выбора группировочного признака во многом зависит **определение числа групп.**

Если в основу группировки положен **атрибутивный** признак, то число групп зависит от числа разновидностей этого признака (таблицы 3.1, 3.6).

Если признак альтернативный, возможны лишь две группы, например, работники, имеющие высшее образование (первая группа), и не имеющие его (вторая группа).

В случае если атрибутивный признак имеет много разновидностей, имеет смысл выделить наиболее распространенные из них и группу «другие» (таблица 3.4).

В статистической практике часто используют классификации и номенклатуры.

Классификация – это распределение элементов совокупности на группы по определенным, в основном, атрибутивным признакам. Данное распределение рассматривается как статистический стандарт и утверждается международными и отечественными статистическими органами. Признаки, которые лежат в основе классификации, остаются неизменными на протяжении длительного периода времени.

Классификации, в которых каждому значению группировочного признака дано, присвоено буквенно-цифровое обозначение – **код**, называются **классификаторами**. Классификаторы способствуют упрощению техники сбора, обработки и передачи статистической информации.

В Украине в практику статистической работы внедрены около 20 национальных статистических классификаций, которые базируются на методологических принципах международных статистических классификаций.

Важнейшей национальной классификацией является Классификация видов экономической деятельности (ее новая редакция введена в действие 4 января 2006 г.).

Если в основу группировки положен **количественный дискретный признак** и он варьирует в небольшом диапазоне, то число групп зависит от числа значений данного признака (таблица 3.2).

Если же в основе группировки лежит непрерывный признак или дискретный, принимающий достаточно большое число значений, определяют число групп и интервалы группировки.

Для определения числа групп необходимо учитывать следующие **два условия построения группировки**:

- выделенные группы должны отличаться качественной однородностью;
- количество единиц в каждой группе должно быть достаточно велико.

Ориентировочно количество групп можно определить по формуле, предложенной американским статистиком Х. Стерджесом:

$$n = 1 + 3{,}332 \lg m, \qquad (3.1)$$

где n – число групп,

m – общее число единиц совокупности.

Предположим, что необходимо разгруппировать 150 предприятий отрасли по среднегодовой стоимости основного капитала. На основе формулы 3.1 определим число групп:

$$n = 1 + 3{,}332 \lg 150 = 8 \text{ групп}$$

Формулу Стерджеса можно использовать при условии, что распределение единиц совокупности по данному признаку приближается к нормальному закону распределения.

Интервал – промежуток между крайними значениями в группе.

Наименьшее значение признака в интервале называется его **нижней границей**, а наибольшее – **верхней**.

Если в основе группировки лежит **дискретный** признак, то нижняя граница i-го интервала равна верхней границе i-1 интервала, увеличенной на 1. Например, при формировании групп, характеризующих численность работников предприятия, интервалы могут быть следующими: 20 – 30, 31 – 40, 41 – 50 и т.д.

Если количественный признак **непрерывный**, то чаще всего верхняя граница каждого интервала совпадает с нижней границей последующего (таблица 3.5). В этом случае для граничных значений признака необходимо указывать, в какой интервал попадает соответствующая единица совокупности с помощью слов «включительно» или «исключительно».

Например, по данным таблицы 3.5: слово «включительно» будет означать, что женщины в возрасте 20, 25, 30 и т.д. лет попадают соответственно в 1-ый, 2-ой, 3-ий и т.д. интервалы; слово «исключительно» предполагает, что женщины данных возрастов учитываются соответственно во 2-м, 3-м, 4-м и т.д. интервалах.

Разница между верхней и нижней границей интервала представляет собой **ширину интервала**. Если верхняя граница интервала не совпадает с нижней границей последующего (для дискретных интервальных рядов), то ширина интервала определяется либо как разница между верхними границами смежных интервалов, либо как разница между их нижними границами.

Виды интервалов с точки зрения их размеров (ширины):

- равные;
- неравные.

Если применяются **равные** интервалы, то расчет их **ширины** (h) производится по формуле:

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

где X_{max} – максимальное значение группировочного признака,

X_{min} – минимальное значение признака,

Технически удобнее работать с равными интервалами. Однако природа социально-экономических явлений такова, что часто приходится применять неравные, прогрессивно, постоянно увеличивающиеся или уменьшающиеся интервалы. Это связано с тем, что абсолютное изменение группировочного признака на одну и ту же величину имеет далеко не одинаковое значение для групп с большим и малым значением признака.

Неравные интервалы могут определяться как **равнонаполненные**. При этом совокупность разделяется на группы равного объема с числом единиц:

$$m_i = \frac{m}{n}, \tag{3.3}$$

Данные ранжируются (устанавливаются в порядке возрастания), отсчитывается число единиц, составляющих первую группу, затем вторую и т.д. Границы интервалов будут соответствовать фактическим значениям признака в каждой группе.

Виды интервалов с точки зрения наличия границ:

- закрытые;
- открытые.

В **закрытых интервалах** указаны нижняя и верхняя границы. В **открытых интервалах** указана одна из границ. Открытые интервалы применяются только для крайних групп (таблица 3.5). В процессе анализа часто необходимо использовать данные о ширине открытых интервалов. Условно принимают, что величина открытого интервала равна величине интервала смежного, граничащего с ним.

Рассмотрим пример построения структурной и аналитической группировки по данным таблицы 3.7. Нам необходимо:

1. Распределить работников предприятия по стажу и построить структурную группировку.
2. Для выявления зависимости между стажем и выработкой построить аналитическую группировку.

Таблица 3.7

Данные о стаже работы и средней дневной выработке

50 работников предприятия

№ п/п	Стаж, полных лет	Выработка, грн	№ п/п	Стаж, полных лет	Выработка, грн
1	8	250	26	10	270
2	4	200	27	14	310
3	7	220	28	1	140
4	6	240	29	5	170
5	6	160	30	7	270
6	5	150	31	15	450
7	9	320	32	9	190
8	11	350	33	3	150
9	8	250	34	6	200
10	12	350	35	10	250
11	10	190	36	7	220
12	4	170	37	9	290
13	11	310	38	1	140
14	7	230	39	6	180
15	18	340	40	4	160
16	10	280	41	9	350
17	3	150	42	8	220
18	4	170	43	6	180
19	9	290	44	3	150
20	10	320	45	8	240
21	12	370	46	7	210
22	0	130	47	8	260
23	7	260	48	6	180
24	2	150	49	6	150
25	13	220	50	3	150

Отвечая на первый вопрос, заметим, что стаж работников является непрерывным признаком. Поэтому необходимо определить количество групп и ширину интервала. Воспользуемся формулой 3.1 и рассчитаем количество групп:

$$n = 1 + 3{,}332 \lg 50 = 6{,}7$$

Сформируем 6 групп с равными интервалами. Для этого определим ширину интервала на основе формулы 3.2:

$$h = \frac{18-0}{6} = 3 \text{ года}$$

Получим следующие группы работников по стажу: 0 – 3, 3 – 6, 6 – 9, 9 – 12, 12 – 15, 15 – 18. Рассчитаем число работников в каждой группе и занесем результаты расчетов в таблицу 3.8:

Таблица 3.8

Распределение работников предприятия по стажу

Группы по стажу, лет	Число работников	
	человек	% к итогу
0 – 3	4	8
3 – 6	10	20
6 – 9	18	36
9 – 12	12	24
12 – 15	4	8
15 – 18	2	4
Всего	50	100

Данные структурной группировки свидетельствуют о том, что больше всего работников в группе, стаж которых 6 – 9 лет, а меньше всего – в группе 15 – 18 лет.

Для построения аналитической группировки необходимо разгруппировать единицы совокупности по признаку-фактору и для каждой группы рассчитать среднюю величину результативного показателя. Очевидно, что в данной совокупности работников признаком-фактором является стаж, а зависимым –дневная выработка.

Группировка по стажу уже произведена. Для расчета средней выработки в каждой группе удобно предварительно построить рабочую таблицу (таблица 3.9). По ее данным рассчитаем среднюю выработку работника в каждой группе по стажу и заполним таблицу 3.10.

Таблица 3.9

Рабочая таблица

Группы по стажу, лет	№ работника (итого работников)	Выработка, грн.	Группы по стажу, лет	№ работника (итого работников)	Выработка, грн.
0 – 3 искл.	22	130		43	180
	24	150		45	240
	28	140		46	210
	38	140		47	260
Итого	**4**	**560**		48	180
3 – 6	2	200		40	150
	6	150	**Итого**	**18**	**3920**
	12	170	9 – 12	7	320
	17	150		8	350
	18	170		11	190
	29	170		13	310
	33	150		16	280
	40	160		19	290
	44	150		20	320
	50	150		26	270
Итого	**10**	**1620**		32	190
6 – 9	1	250		35	250
	3	220		37	290
	4	240		41	350
	5	160	**Итого**	**12**	**3410**
	9	250	12 – 15	10	350
	14	230		21	370
	23	260		25	220
	30	270		27	310
	34	200	**Итого**	**4**	**1250**
	36	220	15 – 18	15	340
	39	180		31	450
	42	220	**Итого**	**2**	**790**
-			**Всего**	**50**	**11550**

Данные аналитической группировки свидетельствуют о наличии прямой связи между стажем и выработкой: с ростом стажа средняя выработка 1 работника растет.

Таблица 3.10

Зависимость дневной выработки работников предприятия от стажа

Группы по стажу, лет	Число работников, чел.	Общая выработка, грн.	Средняя выработка 1 работника, грн.
0 – 3	4	560	140
3 – 6	10	1620	162
6 – 9	18	3920	218
9 – 12	12	3410	284
12 – 15	4	1250	313
15 – 18	2	790	395
Всего (в среднем)	50	11550	231

В практике статистической работы иногда возникает необходимость перегруппировать единицы совокупности и построить вторичную группировку.

Вторичная группировка – группировка, которая была произведена на основе осуществленного ранее распределения единиц совокупности на группы.

Задачами вторичной группировки являются:

- обеспечение сопоставимости двух и более группировок с разными интервалами;
- образование более укрупненных групп с целью выявления характера распределения;
- формирование на основе группировки, которая была построена по количественному признаку, качественно однородных групп.

Вторичная группировка может строиться путем непосредственного укрупнения групп и методом пропорционального дробления (разделения на части) групп.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Дайте определение сводки. Что обозначает этот термин в узком и в широком смысле слова?
2. Какие задачи ставятся перед сводкой как этапом статистического исследования?
3. Какие вопросы рассматриваются в Плане проведения сводки?
4. Назовите виды статистической сводки.
5. Дайте определение статистической группировки и группировочного признака.
6. Какие задачи решает группировка?
7. Поясните, какие виды группировок различают в зависимости от решаемых с их помощью задач.
8. Приведите примеры структурной и типологической группировки.
9. Объясните понятие признака-фактора и результативного признака и сформулируйте правило построения аналитической группировки.
10. Какие различают виды группировок в зависимости от числа группировочных признаков?
11. Приведите примеры простой и комбинационной группировки.
12. Какие методологические задачи решаются в процессе группировки?
13. Как определяют число групп при построении группировки?
14. Что представляют собой классификации и классификаторы?
15. Что такое интервал и как определяется его ширина?
16. Какие виды интервалов вы знаете?
17. Приведите формулу расчета ширины интервала при группировке с равными интервалами.
18. В какой ситуации используют неравные интервалы? Что такое равнонаполненные интервалы?
19. Какие задачи решает вторичная группировка и как она осуществляется?

РАЗДЕЛ 4. ТАБЛИЧНЫЙ И ГРАФИЧЕСКИЙ МЕТОДЫ ПРЕДСТАВЛЕНИЯ СТАТИСТИЧЕСКИХ ДАННЫХ

4.1. СУЩНОСТЬ ТАБЛИЧНОГО МЕТОДА

Статистическая таблица – форма рационального (целесообразного), наглядного (визуального) и компактного изложения статистических данных. Таблица используется также как средство анализа информации.

Таблица дает сводную, обобщенную количественную характеристику совокупности. В ней все одноименные показатели располагают в одной и той же горизонтальной строке или в одной и той же вертикальной графе.

Каждая графа в верхней ее части и каждая строка в левой части таблицы имеют **краткий заголовок (наименование)**.

Строки и графы таблицы без наименований и статистических показателей образуют сетку, которая называется **скелетом таблицы**.

Если скелет таблицы заполнить наименованиями строк и граф, то получим **макет таблицы** (таблицы 4.1 – 4.3).

Таблицу можно рассматривать как **статистическое предложение**. В ней обязательно есть **подлежащее** – та статистическая совокупность, о которой идет речь в таблице, и **сказуемое** – цифровые показатели, характеризующие подлежащее.

Подлежащее чаще всего размещается в левой части таблицы и содержит перечень, список строк. Сказуемое представлено перечнем различных признаков в соответствующих графах таблицы.

Виды таблиц в зависимости от вида подлежащего:

- простые;

- групповые;
- комбинационные.

Подлежащим **простой таблицы** является перечень элементов совокупности, территориальный (пространственный) ряд, хронологический (временной) ряд. Простые таблицы делятся на такие виды:

- перечневые (таблица 4.1);
- территориальные (таблица 4.2);
- хронологические (таблица 4.3).

Пример макета простой перечневой таблицы

Таблица 4.1

Задолженность предприятий региона перед бюджетом

№ предприятия	Сумма задолженности, тис. грн.
1	
2	
…	

Пример макета простой территориальной таблицы

Таблица 4.2

Объем инвестиций в регионы Украины

Регион	Объем инвестиций, млн. дол. США
Одесская область	
Днепропетровская область	
…	

Пример макета простой хронологической таблицы

Таблица 4.3

Динамика выпуска продукции предприятия

Год	Выпуск, тыс. грн.
2009	
2010	
…	

Групповые таблицы содержат в подлежащем группировку единиц совокупности по одному признаку. Примерами групповых таблиц являются таблицы 3.1 – 3.5, которые были рассмотрены в предыдущем разделе.

В подлежащем **комбинационных таблиц** группы единиц, которые были образованы по одному признаку, подразделяются на подгруппы по одному или нескольким другим признакам (таблица 3.6). Она дают возможность проследить зависимость показателей сказуемого от нескольких признаков, которые были положены в основу группировки.

4.2. ПРАВИЛА ПОСТРОЕНИЯ СТАТИСТИЧЕСКИХ ТАБЛИЦ

Оформление таблиц не может быть произвольным (свободным).

Существуют общие **правила построения таблиц**:

1. Слово «Таблица» пишется по правому краю листа. Если в тексте несколько таблиц, их следует нумеровать.
2. Таблица должна иметь название (заголовок), которое отражает ее содержание. Располагается заголовок по центру страницы. В заголовке указывается место и время, к которым относится статистическая информация (в случае фактических данных). Если единицы измерения для всей совокупности одинаковы, их выносят в общий заголовок.
3. Свои названия имеют строки и графы. Названия заголовков пишутся с большой буквы, подзаголовков – с маленькой. Слова в названиях желательно писать полностью и проставлять единицы измерения, если только они не общие для всей таблицы. Если в таблице много граф, их нумеруют. При этом первая графа (наименование подлежащего) обозначается буквой А, а графы сказуемого нумеруются цифрами. Нумерация граф дает также возможность показать способ расчета некоторых показателей (таблица 4.4).

4. Числа в таблице не должны быть многозначными и занимать много места. Каждая клетка содержит только одно число. В границах одного ряда или графы желательно округлять числа с одинаковой степенью точности. Если число знаков после запятой не одинаково, то числа, у которых знаков меньше, дополняют нулями.
5. Большинство таблиц содержат итоговые строки. В сложных таблицах следует различать «итого» и «всего» (таблица 3.9). «Итого» характеризует суммарный результат для части совокупности, а «всего» – для совокупности в целом.

Макет таблицы

Таблица 4.4

Показатели внешней торговли стран СНГ (млн. дол. США)

Регион	Экспорт	Импорт	Внешнеторговый оборот	Сальдо внешней торговли
А	1	2	3 = 1 + 2	4 = 1 - 2
Украина				
Белорусь				
…				

6. Каждая клетка таблицы должна быть заполнена. Если числа в некоторых клетках отсутствуют, то клетка заполняется следующим образом: в случае отсутствия сведений о величине какого-либо показателя, в клетке проставляются три точки (…); в случае отсутствия самого явления, проставляется тире (-); если клетка не подлежит заполнению, ставится знак х; если числовое значение показателя меньше, чем принятая в таблицы точность округления, проставляется 0,0.
7. В случае продолжения таблицы на другой странице, слово «Таблица» указывается один раз перед первой частью таблицы, над другой частью пишут «Продолжение таблицы» с указанием ее номера.

И в дополнение к сказанному. Для удобства прочтения таблицы она не должна быть большой. Лучше построить две маленькие наглядные таблицы, чем одну громоздкую.

4.3. СТАТИСТИЧЕСКИЕ ГРАФИКИ

Часто статистические таблицы дополняются графиками, когда ставится цель подчеркнуть какую-то особенность данных, провести их сравнение. Графики являются самой эффективной формой представления статистических данных с точки зрения их восприятия, понимания. Часто графики используются и без связи с таблицей. С помощью графиков достигается наглядность, понимание характеристики структуры, динамики, сравнения, взаимосвязи явлений.

График – это особый способ наглядного изображения и обобщения статистической информации с помощью точек, линий, фигур.

Целью построения графиков является:

- изучение структуры явления и структурных сдвигов (изменений);
- сравнение размеров различных явлений и их отдельных частей;
- изучение распределения единиц совокупности по определенным признакам;
- отображение интенсивности развития явлений во времени, их тенденций (направлений развития), сезонных явлений;
- отражение взаимосвязи между явлениями.

Каждый график состоит из графического образа и вспомогательных элементов.

Графический образ – это совокупность точек, линий и фигур, с помощью которых изображаются статистические данные.

Вспомогательными элементами графика являются:

1) **поле графика** – это пространство, в котором размещаются геометрические или другие графические знаки, образующие график;

2) **пространственные ориентиры** – это элементы графика, которые определяют порядок размещения графических знаков в его поле. Пространственные ориентиры задаются системой координатных сеток или контурных линий, которые делят это поле на части. В большинстве случаев в статистических графиках применяется система прямоугольных координат, но встречаются и круговые графики, построенные по принципу полярных координат;
3) **масштабные ориентиры**, придающие геометрическим знакам количественную определенность;
4) **экспликация графика**, состоящая из названия графика и объяснения смыслового значения каждого знака, который применяется на данном графике. Пояснительные тексты могут располагаться в пределах графического образа или рядом с ним, а также выноситься за его пределы.

Графики классифицируют по различным признакам:

- **по содержанию** (графики сравнения в пространстве, структуры, динамики, графики вариационных рядов, взаимосвязанных показателей);
- **по способу построения** (диаграммы, картодиаграммы, картограммы);
- **по характеру графического образа** (точечные, линейные, плоскостные, объемные); плоскостные, в свою очередь, делятся на столбиковые, квадратные, круговые, секторные, фигурные.

Рассмотрим **правила построения столбиковой диаграммы**, которая используется чаще всего для сравнения одноименных показателей, характеризующих различные объекты или территории.

Значение показателей, которые сравниваются, изображаются в виде прямоугольных столбиков, имеющих одинаковую ширину и расположенных на общей горизонтальной или вертикальной базовой линии. Высота (или длина) каждого столбика в определенном масштабе соответствует величине показателя, который изображается. Столбики могут располагаться вплотную (рядом друг с другом) либо на одинаковом расстоянии друг от друга. Примером такой диаграммы является рисунок 4.1.

Рис. 4.1. Среднемесячная заработная плата наемных работников в Украине в 2010 г. (грн.)

Основной формой **структурных** диаграмм являются секторные диаграммы. Секторная диаграмма – это графические изображения на площади круга, разделенные радиусами на отдельные сектора по количеству составных частей признака (рис. 4.2). Геометрическим параметром (геометрической величиной) в секторной диаграмме удельных весов (соотношение части совокупности и целого) служит величина угла между радиусами: 1 % принимается на диаграмме равным 3,6 градуса, а сумма всех углов, составляющая 360 градусов, приравнивается к 100 %. Возможность применения секторных диаграмм ограничена двумя обстоятельствами. Во-первых, они сохраняют свою выразительность при делении совокупности на небольшое число частей – не более 4 – 5, а, в-вторых, секторная диаграмма выглядит убедительно (доказательно) только при существенных различиях структур, которые сравниваются.

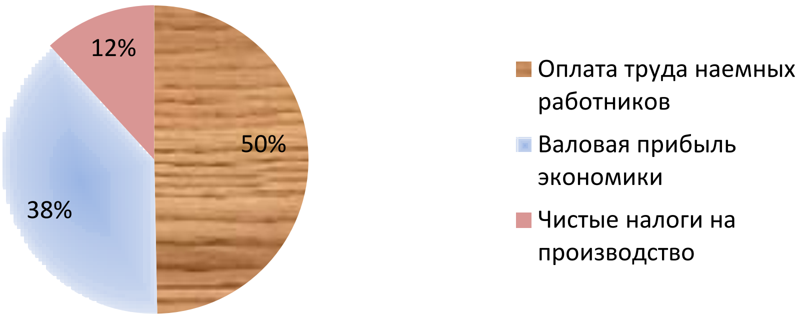

Рис. 4.2. Структура ВВП Украины в 2010 г.

Для изображения экономических явлений, протекающих во времени, применяют **динамические диаграммы** (диаграммы развития во времени), объектом отображения которых являются процессы (рис. 4.3). Геометрическими знаками-символами на таких диаграммах служат точки и последовательно соединяющие их прямые линии, складывающиеся в ломаные кривые. Конфигурация (общий вид) этих ломаных дает представление о процессе, который изображается. Ось абсцисс является в такой диаграмме осью времени, а ось ординат – осью значений, которые принимает с течением времени показатель, который изучается.

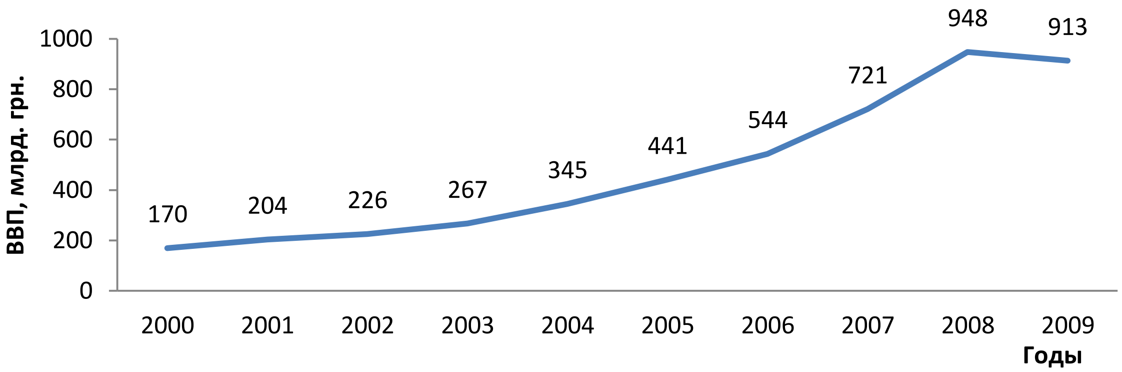

Рис. 4.3. Динамика ВВП Украины

Для сопоставления интенсивности динамики различных показателей ломаные линии можно нанести на одном графике.

Графики вариационных рядов будут рассмотрены в 5 разделе.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Сформулируйте определение и объясните значение статистической таблицы.
2. Поясните понятия «скелет» и «макет» таблицы.
3. Что является подлежащим и сказуемым таблицы?
4. Какие вы знаете виды таблиц в зависимости от построения подлежащего?
5. Какие подвиды простых таблиц существуют?
6. Постройте макеты простой, групповой и комбинационной таблицы.
7. Назовите правила оформления статистических таблиц.
8. Что представляет собой статистический график?
9. Что такое графический образ?
10. Поясните понятие «поле графика».
11. Что представляют собой пространственные и масштабные ориентиры графика?
12. Что включает в себя понятие «экспликация графика»?
13. По каким признакам классифицируют графики?
14. Для чего используется столбиковая диаграмма? Сформулируйте правила ее построения.
15. Что представляет собой секторная диаграмма? Для анализа каких процессов она используется?
16. Назовите правила построения и область применения динамической диаграммы.

РАЗДЕЛ 5. ОБОБЩАЮЩИЕ СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

5.1. ПОНЯТИЕ И ВИДЫ СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ

После сводки и группировки данных наблюдения приступают к третьему этапу статистического исследования – анализу данных. Анализ осуществляется на основе расчета статистических показателей.

Статистический показатель – обобщенная качественно-количественная характеристика социально-экономического явления или процесса в определенных условиях места и времени.

Данное определение указывает, во-первых, на **качественное содержание показателя**. Оно определяется сутью явления, местом и временем его существования и отображается в названии, например: прибыль банка в текущем месяце. Во-вторых, определение статистического показателя указывает на **количественную сторону явления**, то есть конкретную цифровую характеристику с указанием единиц измерения. Например, в 2010 году инвестиции в регион составили 1,2 млн. грн.

Итак, **каждый статистический показатель должен иметь**:

- название, отражающее суть явления;
- обозначение территории, на которой наблюдается явление;
- время регистрации явления;
- единицу измерения.

Остановимся на соотношении между признаком и показателем. Признак определяет качественное содержание показателя, его объективную основу (например, возраст). Признаки существуют вне зависимости от того, изучает их статистика или нет. Показатель создается наукой и служит инструментом познания (например: средний возраст работников предприятия; возраст, который наиболее часто встречается; возраст, меньше которого у половины работников).

В зависимости от способа определения показателей, признака времени и других свойств существуют различные виды статистических показателей.

По способу определения различают показатели:

- первичные;
- производные.

Первичные показатели определяются путем сводки и группировки данных и представляются в форме абсолютных величин (например: число семей с двумя детьми).

Производные показатели определяются на базе первичных и имеют форму средних или относительных величин (например: средняя себестоимость, соотношение уровней цен).

По признаку времени различают показатели:

- интервальные;
- моментные.

Интервальные показатели характеризуют результат развития явления за определенный период времени (например: выпуск продукции за год).

Моментные показатели характеризуют состояние явления на определенный момент времени (например: запасы продукции на конец года).

По характеру явления различают показатели:

- объемные (количественные, экстенсивные);
- качественные (интенсивные).

Объемные показатели характеризуют общий объем признака (например: общий выпуск продукции) или общее число единиц совокупности (например: число студентов).

Качественные показатели характеризуют объем признака в расчете на единицу совокупности (например: выпуск продукции в единицу времени, т. е. производительность труда).

По форме выражения различают показатели:

- абсолютные;
- относительные;
- средние.

5.2. ПОНЯТИЕ, ВИДЫ И ЕДИНИЦЫ ИЗМЕРЕНИЯ АБСОЛЮТНЫХ ВЕЛИЧИН

Абсолютная величина – это количественный показатель, который характеризует размеры, уровни, объемы массовых социально-экономических явлений в определенных условиях места и времени.

Виды абсолютных величин по степени охвата совокупности:

- индивидуальные;
- групповые;
- суммарные.

Индивидуальные абсолютные величины характеризуют отдельные единицы совокупности (например: заработная плата одного работника). Их получают на стадии статистического наблюдения.

Групповые абсолютные величины отображают размеры признака в отдельных частях совокупности (например: фонд заработной платы в отдельном цеху предприятия).

Суммарные абсолютные величины характеризуют объем статистической совокупности или общий объем признака в ней (например: фонд заработной платы в целом по предприятию).

Групповые и суммарные абсолютные величины получают в результате сводки и группировки данных наблюдения либо в результате подведения итогов, либо путем специальных расчетов (например, прогнозный объем выпуска).

Все абсолютные величины являются именованными числами, то есть они имеют единицы измерения (например: 10 л, 50 м, 120 грн.). В отличие от математического понятия «абсолютной величины» статистические абсолютные величины могут быть как положительными, так и отрицательными (например: в случае убытков, потерь и т.д.).

Виды единиц измерения абсолютных величин:

- натуральные;
- трудовые;
- стоимостные.

Натуральные единицы измерения характеризуют явления в свойственной им натуральной форме. Они могут выражаться в мерах длины: метры (м), километры (км); веса: килограммы (кг), тонны (т); объема: литры (л) и т.д. или же количеством единиц, числом событий (штуки, случаи, события).

В ряде случаев используют **комбинированные единицы измерения**, представляющие собой произведение двух величин. Так, например, производство электроэнергии измеряется в кВт.-час., грузооборот – в т-км, и т.д.

Если индивидуальные абсолютные величины характеризуют отдельные разновидности продукции, близкие по потребительным свойствам, используют **условно-натуральные единицы измерения**.

Например, мыловаренный завод произвел 2 тыс. т мыла с 40 % содержанием жирных кислот и 3 тыс. т с 80 % содержанием жирных кислот. Необходимо определить общий объем производства в пересчете на 40-процентное мыло.

Прежде всего определим коэффициент пересчета 80-% мыла в условное 40-%: $80 \div 40 = 2$. Тогда общий объем производства мыла в пересчете на 40-% составит:

$$2000 \cdot 1 + 3000 \cdot 2 = 8000 \text{ т.}$$

Трудовые единицы измерения характеризуют наличие и использование трудовых ресурсов или затраты рабочего времени на производство продукции. Они могут быть простыми и комбинированными. Например, число занятых на производстве (чел.) – простая единица измерения, а человеко-дни (чел.-дни), человеко-часы (чел.-час) – комбинированные.

Стоимостные (денежные) единицы измерения служат для определения общего размера различных явлений, несопоставимых между собой в натуральных единицах измерения. Они позволяют оценить общие результаты деятельности предприятий, отраслей, регионов, экономики в целом. В качестве стоимостного измерителя может выступать как национальная валюта, так и валюта других стан.

5.3. ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ. ИХ ФОРМЫ ВЫРАЖЕНИЯ И ВИДЫ

Абсолютные величины играют важную роль в системе обобщающих статистических показателей. Но достаточно полного представления о явлении, которое изучается, они дать не могут. С помощью абсолютных величин нельзя проанализировать структуру явления, соотношение между его отдельными частями, изучить изменение его во времени и пространстве.

Такую задачу могут решить относительные величины.

Относительная величина – это мера соотношения двух других величин. Логическая формула относительной величины следующая:

$$\text{Относительная величина} = \frac{\text{Сравниваемая величина(величина,которая сравнивается)}}{\text{Базисная величина (база сравнения)}}$$

В зависимости от базы сравнения различают следующие **формы выражения относительных величин**:

- коэффициент;
- процент (%);

- промилле (‰),

Относительная величина выражена в форме **коэффициента**, если база сравнения принимается за 1 единицу.

Относительная величина выражена в форме **процента**, если база сравнения принимается за 100 единиц.

Относительная величина выражена в форме **промилле**, если база сравнения принимается за 1000 единиц.

Реже, лишь в случаях, если сравниваемая величина значительно меньше базы сравнения, относительные величины выражают в продецимиллях, просантимиллях (база сравнения соответственно принимается за 10 000 и 100 000 единиц).

По содержанию, характеру задач, которые решаются с помощью относительных величин, различают следующие **виды относительных величин**:

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

Относительная величина планового задания (ОВПЗ) характеризует изменение уровня, которое было предусмотрено планом на последующий период, по сравнению с предыдущим и определяется соотношением уровня, который был запланирован ($У_{пл}$), и базисного ($У_0$):

$$ОВПЗ = \frac{У_{пл}}{У_0}. \qquad (5.1)$$

Относительная величина динамики (ОВД) или темп (скорость) роста характеризует направление и интенсивность изменения явления во времени, и определяется отношением текущего (отчетного) уровня ($У_1$) к базисному:

$$ОВД = \frac{у_1}{у_0}. \quad (5.2)$$

Относительная величина выполнения плана (ОВВП) характеризует степень выполнения плана и определяется соотношением текущего (фактического) уровня и уровня, который был запланирован на данный период:

$$ОВВП = \frac{у_1}{у_{пл}}. \quad (5.3)$$

ОВПЗ, ОВД и ОВВП выражаются в форме коэффициентов и процентов.

Пример. В 2009 году предприятие произвело продукции на 400 тыс. грн. На 2010 год было запланировано увеличить выпуск до 410 тыс. грн. Фактически выпуск 2010 года составил 415 тыс. грн. Определим относительные величины планового задания, динамики и выполнения плана.

В соответствии с формулой 5.1:

$$ОВПЗ = \frac{410}{400} = 1{,}025 \text{ или } 102,5\ \% \ (+2{,}5\ \%).$$

На основании формулы 5.2:

$$ОВД = \frac{415}{400} = 1{,}038 \text{ или } 103{,}8\ \% \ (+3{,}8\ \%).$$

На базе формулы 5.3:

$$ОВВП = \frac{415}{410} = 1{,}012 \text{ или } 101{,}2\ \% \ (+1{,}2\ \%).$$

Таким образом, на 2010 год было запланировано увеличить выпуск продукции в 1,025 раз или на 2,5 %. Фактически в 2010 году по сравнению с 2009 годом выпуск увеличился в 1,038 раз или на 3,8 %. Таким образом, план 2010 года был перевыполнен в 1,012 раз или на 1,2%.

Взаимосвязь между относительными величинами планового задания, динамики и выполнения плана:

$$ОВД = ОВПЗ \cdot ОВВП. \quad (5.4)$$

Действительно:

$$\frac{у_1}{у_0} = \frac{у_{пл}}{у_0} \cdot \frac{у_1}{у_{пл}}.$$

Пример. На отчетный год было запланировано снизить себестоимость на 15,6%, фактически в отчетном году по сравнению с базисным себестоимость удалось снизить на 12, 4 %. Определим относительную величину выполнения плана по снижению себестоимости.

На основании формулы 5.4 записываем, чему равна ОВВП, и подставляем значение ОВД:

100 – 12,4 = 87,6 % или 0,876

и ОВПЗ:

100 – 15,6 = 84,4 % или 0,844.

Получаем:

$$\text{ОВВП} = \frac{\text{ОВД}}{\text{ОВПЗ}} = \frac{0{,}876}{0{,}844} = 1{,}038 \text{ или } 103{,}8\ \% \ (+3{,}8\ \%).$$

То есть, в отчетном году план по снижению себестоимости был недовыполнен на 3,8 %.

Примечание. Такая интерпретация (объяснение) результата поясняется тем, что фактическая себестоимость оказалась на 3,8 % выше плановой. В нашей задаче планировалось снизить себестоимость как затратный показатель. Таким образом, план по снижению недовыполнен (выполнен не полностью).

Относительная величина структуры (ОВС) характеризует состав, структуру совокупности по тому или иному признаку и показывает долю, удельный вес части совокупности ($У_i$) в ее общем объеме ($\sum У_i$):

$$\text{ОВС} = \frac{У_i}{\sum У_i}. \qquad (5.5)$$

Относительная величина структуры может быть выражена в форме коэффициента, но чаще ее выражают в форме процента.

Количество относительных величин структуры в совокупности такое же, как число составных частей в ней. Сумма значений относительных величин совокупности, которые выражены в форме коэффициентов, равна 1, а сумма значений относительных величин совокупности, которые выражены в форме процентов, равна 100 %.

Относительная величина координации (ОВК) характеризует соотношение между отдельными частями одного целого и рассчитывается как отношение одной из его частей ($У_i^1$) к другой, которая принята за базу сравнения ($У_i^2$):

$$ОВК = \frac{У_i^1}{У_i^2}. \quad (5.6)$$

Относительная величина координации показывает, во сколько раз одна часть больше другой или сколько единиц одной части приходится на 10, 100, 1000 единиц второй.

Относительная величина сравнения (ОВСР) характеризует соотношение одноименных показателей, относящихся к различным объектам или территориям за один и тот же период времени. Базой сравнения может выступать любой объект или территория (А или Б). От выбора базы будет зависеть интерпретация результатов. Формула расчета ОВСР может быть записана:

$$ОВСР = \frac{У_А}{У_Б}. \quad (5.7)$$

Относительная величина сравнения выражается в форме коэффициента или процента.

Пример. По данным таблицы 5.1 о численности населения отдельных областей Украины определить:

1. Относительные величины структуры общей численности населения Днепропетровской области.
2. Относительную величину координации численности населения Донецкой области. За базу сравнения принимаем численность сельского населения.
3. Относительную величину сравнения общей численности населения Днепропетровской и Донецкой областей. За базу сравнения принимаем численность населения Днепропетровской области.

Таблица 5.1

Численность населения регионов Украины

на 1.05.2011 г. (млн. чел)

Область	Все население	в том числе	
		городское	сельское
Днепропетровская	3,3	2,8	0,5
Донецкая	4,4	4,0	0,4

Отвечая на первый вопрос, заметим, что признаком, по которому совокупность жителей Днепропетровской области разделена на две части, является их место проживания – город или село. Таким образом, необходимо рассчитать две относительные величины структуры. Воспользуемся формулой 5.5:

$$\text{Удельный вес горожан} = \frac{\text{Численность горожан}}{\text{Общая численность населения}} \cdot 100 = \frac{2{,}8}{3{,}3} \cdot 100 =$$

$$= 84{,}8\ \%,$$

$$\text{Удельный вес селян} = \frac{\text{Численность селян}}{\text{Общая численность населения}} \cdot 100 = \frac{0{,}5}{3{,}3} \cdot 100 = 15{,}2\ \%.$$

Проверим правильность расчётов: 84,8 + 15,2 = 100 %.

Таким образом, на 1.05.2011 г. удельный вес городских жителей в общей численности жителей Днепропетровской области Украины составлял 84,8 %, а сельских – 15,2 %.

Относительную величину координации численности населения Донецкой области определим по формуле 5.6:

$$\text{ОВК} = \frac{\text{Численность горожан}}{\text{Численность селян}} = \frac{4{,}0}{0{,}4} = 10.$$

Итак, на 1.05.2011 г. численность городского населения Донецкой области превышала численность сельского в 10 раз.

Примечание. Если относительная величина, выраженная в форме коэффициента, превышает 2, то ее пояснение, интерпретация в форме процентов не обязательна. Теоретически можно сказать, что число горожан превышает число сельских жителей на 900 % (1000 – 100), но данный вывод ассоциируется (связывается) с увеличением в 9, а не в 10 раз, что является неверным.

Относительная величина сравнения численности населения двух регионов рассчитывается по формуле 5.7:

$$\text{ОВСР} = \frac{4{,}4}{3{,}3} = 1{,}333 \text{ или } 133{,}3\ \% \ (+33{,}3\ \%).$$

Таким образом, на 1.05.2011 г. численность населения Донецкой области превышала численность населения Днепропетровской в 1,333 раз или на 33,3 %.

Вывод по данному показателю можно сформулировать и иначе. На 1.05.2011 г. на каждую 1000 жителей Днепропетровской области приходилось в среднем 1333 жителя Донецкой.

Относительная величина интенсивности (ОВИ) характеризует степень распространения или развития явления в определенной среде и представляет собой соотношение разноименных показателей, взятых за один период времени.

Тот факт, что **ОВИ является соотношением разноименных показателей**, отличает ее от других относительных величин.

Еще одно существенное отличие заключается в том, что данная относительная величина **имеет единицы измерения**. Они соответствуют единицам измерения тех величин, на основе которых ОВИ рассчитывается, и показывают, сколько единиц одной совокупности приходится на единицу другой.

Примерами относительных величин интенсивности являются: плотность населения (соотношение численности населения и площади территории проживания), показатели рождаемости, смертности, брачности и разводимости, представляющие собой соотношение соответствующих демографических характеристик к численности населения. К относительным величинам интенсивности относятся такие характеристики производственной деятельности, как производительность труда (соотношение объёма выпуска продукции и затрат времени), трудоемкость (обратная производительности труда величина), капиталоотдача (соотношение объёма выпуска продукции и среднегодовой стоимости капитала) и другие.

5.4. СРЕДНИЕ ВЕЛИЧИНЫ. ИХ ВИДЫ И МЕТОДЫ РАСЧЕТА

Средние величины играют в статистике очень важную роль. Они являются обобщающей характеристикой большого числа индивидуальных значений варьирующего признака.

Средняя величина – обобщающий статистический показатель, абстрактная характеристика, отображающая типичный уровень варьирующего признака в расчете на единицу совокупности в конкретных условиях места и времени.

Основным в определении средней величины является:

1. **Средняя величина – обобщающий показатель**, то есть средняя показывает то общее, что характеризует явление в целом. Отклонения значений признака, связанные с действием случайных причин, в средней величине взаимно погашаются (уничтожаются). В результате этого средние величины можно использовать для сравнения величин признака в разных совокупностях. Например, необходимо сравнить уровень заработной платы на двух предприятиях. Зарплаты отдельных работников (абсолютные величины) могут быть нетипичными для этих предприятий. Лишь средняя зарплата, приходящаяся на одного работника каждого предприятия, освобождена от влияния случайных причин и дает ответ на вопрос, где уровень зарплаты выше.
2. **Средняя величина – абстрактная характеристика совокупности**, так как она абстрагируется от значений (не учитывает значений) признака у отдельных единиц совокупности и может с ними не совпадать. Средняя величина может быть выражена в дробях даже для показателей, выраженных в штуках, событиях, случаях. Непременным

(обязательным) является лишь свойство средней величины быть больше наименьшего значения признака и меньше наибольшего.

3. **Средняя величина в статистике всегда имеет единицы измерения**. Они такие же, как единицы измерения признака, который исследуется. В этом смысле средняя величина в статистике отличается от математического понятия средней, где она абстрагируется от качественной стороны явления.

4. **Средняя величина рассчитывается только для варьирующего признака**, при отсутствии вариации ее применение бессмысленно.

Главными условиями использования средних величин являются:

- наличие качественно однородной совокупности;
- массовый характер данных совокупности, где действует закон больших чисел.

Признак, для которого определяется средняя величина, называется **осредняемым признаком**.

При использовании средних величин введем следующие **обозначения**:

- $\bar{x}$ – среднее значение признака;
- x_i или x – отдельные значения осредняемого признака (варианты);
- f_i или f – частоты повторений (веса);
- $F = xf$ – общий объем признака;
- n – общее число единиц совокупности.

В зависимости от характера осредняемого признака и характера связи между отдельными значениями признака и его общим объемом (сумма, произведение, степень, квадратный корень) различают следующие основные **виды средних величин**:

- средняя арифметическая;
- средняя гармоническая;
- средняя квадратическая;
- средняя геометрическая и другие.

Каждый из приведенных видов средней величины может выступать в двух **формах:**

- простой форме;
- взвешенной форме.

Простая форма используется при определении средней по первичным, несгруппированным данным.

Взвешенная форма используется, если данные сгруппированы.

В процессе выбора вида и формы средней величины необходимо исходить из экономического смысла осредняемого показателя. Часто осредняемый показатель может быть записан в виде следующего соотношения, **логической формулы**:

$$\bar{x} = \frac{\text{Общий объем признака}}{\text{Общее число единиц совокупности}}.$$

Приведем несколько примеров формул расчета осредняемых показателей.

$$\begin{matrix}\text{Средняя выработка продукции} \\ \text{в единицу времени}\end{matrix} = \frac{\text{Количество произведенной продукции}}{\text{Общие затраты времени}};$$

$$\begin{matrix}\text{Средняя заработная плата} \\ \text{одного работника}\end{matrix} = \frac{\text{Фонд заработной платы}}{\text{Количество работников}};$$

$$\begin{matrix}\text{Средняя цена} \\ \text{единицы продукции}\end{matrix} = \frac{\text{Стоимость всей продукции}}{\text{Общее количество продукции}};$$

$$\begin{matrix}\text{Средняя урожайность зерновых} \\ \text{с одного га площади}\end{matrix} = \frac{\text{Валовой (общий) сбор}}{\text{Общая площадь в га}};$$

$$\begin{matrix}\text{Средняя себестоимость} \\ \text{единицы продукции}\end{matrix} = \frac{\text{Общая себестоимость продукции}}{\text{Объем произведенной продукции}}.$$

Логическая формула часто является исходной базой расчета и критерием правильности выбора вида средней величины. В общем же случае средняя величина должна рассчитываться таким образом, чтобы при замене каждого варианта осредняемого показателя средней величиной, неизменным оставался бы некий итоговый показатель, связанный с осредняемым. Данный итоговый показатель называют **определяющим**.

Средняя арифметическая – наиболее распространенный вид средних величин. Она используется тогда, когда известны отдельные значения осредняемого признака и общее число единиц совокупности.

Средняя арифметическая простая имеет вид:

$$\overline{x} = \frac{\sum x}{n}. \tag{5.8}$$

Например, в первой группе 15 студентов, во второй – 18, в третьей – 24 студента. Среднее число студентов в группе рассчитывается с помощью средней арифметической простой, то есть, на основе формулы 5.8. запишем:

$$\overline{x} = \frac{15+18+24}{3} = 19 \text{ студентов.}$$

Средняя арифметическая взвешенная имеет вид:

$$\overline{x} = \frac{\sum xf}{\sum f}. \tag{5.9}$$

Например, в бригаде 3 рабочих зарабатывают по 2200 грн. в месяц, 2 рабочих – по 2500 грн. и 1 рабочий – 3100 грн. Определим на основе формулы 5.9 среднюю зарплату одного рабочего:

$$\overline{x} = \frac{2200 \cdot 3 + 2500 \cdot 2 + 3100 \cdot 1}{6} = 2450 \text{ грн.}$$

Определяющим показателем для средней арифметической является общий объем признака, то есть числитель формулы. Продемонстрируем, что он остается без изменения при замене отдельных значений заработной платы на среднюю величину:

$$\overline{x} = \frac{2450 \cdot 3 + 2450 \cdot 2 + 2450 \cdot 1}{6} = 2450 \text{ грн.}$$

Средняя гармоническая используется тогда, когда известны варианты осредняемого признака и общий объем признака.

Средняя гармоническая простая выражается формулой:

$$\overline{x} = \frac{n}{\sum \frac{1}{x}}. \tag{5.10}$$

Формула средней гармонической взвешенной имеет вид:

$$\overline{x} = \frac{\sum F}{\sum \frac{F}{x}}. \tag{5.11}$$

Например, на протяжении часа двое рабочих изготавливали одинаковые детали. Первый рабочий на одну деталь затратил полчаса времени, второй – четверть часа. Определим среднее время на изготовление одной детали.

На первый взгляд задача решается с помощью формулы средней арифметической простой:

$$\overline{x} = \frac{30+15}{2} = 22{,}5 \text{ мин.}$$

Данное значение было бы правильным, если бы каждый рабочий произвел по одной детали, а не работал в течение часа.

Правильно начать расчет с записи логической формулы:

$$\begin{matrix}\text{Среднее время} \\ \text{на изготовление 1 детали}\end{matrix} = \frac{\text{Общее время}}{\text{Общее число деталей}};$$

$$\overline{x} = \frac{1+1}{\frac{1}{0{,}5} + \frac{1}{0{,}25}} = \frac{1}{3} \text{ час. или 20 мин.}$$

Последнее количественное соотношение отвечает формуле средней гармонической простой.

Пример. На основе данных таблицы 5.2 определим среднюю по трем бригадам зарплату рабочего:

Таблица 5.2

Данные по трем бригадам предприятия

Бригада	Средняя зарплата рабочего, грн. x	Фонд зарплаты, грн. F	$\frac{F}{x}$, чел.
1	1900	57000	30
2	2050	36900	18
3	2120	21200	10
Итого	-	115100	58

Запишем логическую формулу расчета средней зарплаты:

$$\begin{matrix}\text{Средняя заработная плата} \\ \text{одного работника}\end{matrix} = \frac{\text{Фонд заработной платы}}{\text{Количество работников}}.$$

По условию задачи нам известны отдельные значения признака – уровни средней зарплаты в каждой бригаде и общий объем признака – фонд заработной платы. Необходимо использовать формулу средней гармонической взвешенной (формула 5.11). Знаменатель формулы, общее число работников, определим в таблице. В итоге:

$$\overline{x} = \frac{115100}{58} = 1984 \text{ грн.}$$

Таким образом, на предприятии средняя по трем бригадам заработная плата одного рабочего составляет 1984 грн.

Другие виды средних величин будут рассмотрены в последующих разделах.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Раскройте суть статистического показателя и поясните, какие атрибуты должен иметь каждый из показателей.
2. Поясните различие между признаком и показателем.
3. Что представляют собой интервальные и моментные показатели?
4. Поясните различие между объемными и качественными показателями.
5. Какие существуют формы выражения статистических показателей?
6. Что представляет собой абсолютная величина?
7. Назовите виды абсолютных величин по степени охвата единиц совокупности и поясните различие между ними.
8. Какие существуют виды единиц измерения абсолютных величин?
9. Что характеризуют натуральные единицы измерения? Приведите примеры данных единиц измерения.
10. представляют собой условно-натуральные единицы измерения?
11. Приведите примеры трудовых единиц измерения и опишите, что они характеризуют.
12. Для чего служат стоимостные единицы измерения?
13. Какие задачи можно решить с использованием относительных величин?

14. Сформулируйте определение и запишите логическую формулу относительной величины.
15. Назовите формы представления относительных величин?
16. Какой формой выражена относительная величина, если база сравнения принимается за единицу?
17. Что представляет собой форма относительной величины – промилле?
18. Перечислите виды относительных величин.
19. Что характеризуют и как рассчитываются относительные величины планового задания, динамики и выполнения плана?
20. Какая существует взаимосвязь между относительными величинами планового задания, динамики и выполнения плана?
21. Что характеризует и как рассчитывается относительная величина структуры?
22. Расскажите про относительную величину координации.
23. Что показывает относительная величина сравнения?
24. Какие особенности у относительной величины интенсивности?
25. Опишите роль средней величины в статистических исследованиях.
26. Дайте определение средней величины и поясните его.
27. Почему среднюю величину называют абстрактной характеристикой совокупности?
28. Назовите главные условия использования средних величин.
29. Назовите виды средних величин.
30. Какие существуют формы выражения средних величин? Когда какая форма применяется?
31. Приведите формулы расчета средней арифметической и средней гармонической и поясните, когда какая из них используется.

РАЗДЕЛ 6. АНАЛИЗ РЯДОВ РАСПРЕДЕЛЕНИЯ

6.1. СУЩНОСТЬ И ВИДЫ РЯДОВ РАСПРЕДЕЛЕНИЯ

Внутри совокупности, которая изучается, индивидуальные значения признака изменяются от одной единицы к другой. Чтобы изучить характер распределения единиц совокупности по определенному признаку, выявить закономерности такого распределения, строят ряды распределения.

Ряд распределения представляет собой распределение единиц совокупности на группы по какому-либо признаку с указанием либо числа единиц в каждой группе, либо числа единиц в процентах к итогу (к общему числу единиц).

Ряд распределения состоит из двух элементов: вариантов и частот (частостей).

Варианты (x) – это отдельные значения или разновидности признака.

Частоты (f) представляют собой численности единиц в каждой группе. Сумма всех частот ряда равна общей численности единиц изучаемой совокупности. Она называется **объемом совокупности** и обозначается n.

Частости (d) – это частоты, представленные относительными величинами в форме коэффициента (доли) или процента. Если частости выражены в форме коэффициента, то сумма долей равна единице; сумма всех частостей, выраженных в форме процента, равна 100 процентам.

Дополнительными характеристиками рядов распределения являются **кумулятивная (накопленная) частота** и **частость** (S). Они характеризуют объем совокупности со значениями вариантов, которые не превышают определенное значение признака.

В зависимости от статистической природы вариантов ряды распределения делятся на атрибутивные ряды и вариационные.

Атрибутивный ряд распределения – ряд, который построен по атрибутивному признаку (таблица 6.1). При построении атрибутивного ряда число групп соответствует числу разновидностей признака, положенного в его основу.

Таблица 6.1

Структура посевных площадей района

Культура	Зерновые (пшеница, рис…)	Технические (хлопок…)	Картофель	Итого
В % к общей площади	50	10	40	100

Вариационный ряд распределения – ряд, который строится по количественному признаку. Он может быть дискретным (таблица 6.2) или интервальным (таблица 6.3). Если интервалы на протяжении всего ряда сохраняют одну и ту же величину, имеем вариационный ряд с равными интервалами.

Таблица 6.2

Распределение работников предприятия по тарифному разряду (квалификации)

Разряд	1	2	3	4	5	6	Итого
Число рабочих	4	8	9	4	3	2	30
Накопленная частота	4	12	21	25	28	30	-

Для графического изображения **атрибутивных рядов распределения** используют следующие виды диаграмм:

- столбиковые;
- секторные;
- кольцевые.

Таблица 6.3

Распределение рабочих предприятия по средней дневной выработке

Группы рабочих по выработке, кг	Число рабочих, чел. *f*	Середина интервала, кг *x*	*xf*	*S*
Менее 600	70	400	28000	70
600 - 1000	80	800	64000	150
1000 - 1400	50	1200	60000	200
1400 - 1800	60	1600	96000	260
1800 и более	40	2000	80000	300
Итого	300	-	328000	-

На рисунке 6.1 изображена секторная диаграмма, построенная по данным таблицы 6.1.

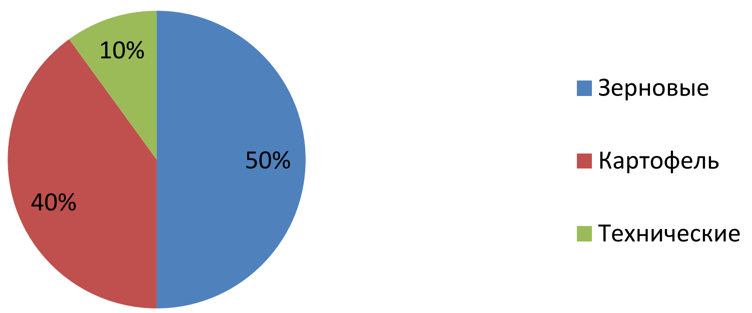

Рис. 6.1. Структура посевных площадей района

Для графического изображения дискретного вариационного ряда используют **полигон распределения**. Его изображают в прямоугольной системе координат, где на оси абсцисс откладывают значения вариантов *x*, а на оси ординат – частоты *f*. Точки пересечения соединяют прямыми линиями.

На рисунке 6.2 представлен полигон распределения по данным таблицы 6.2.

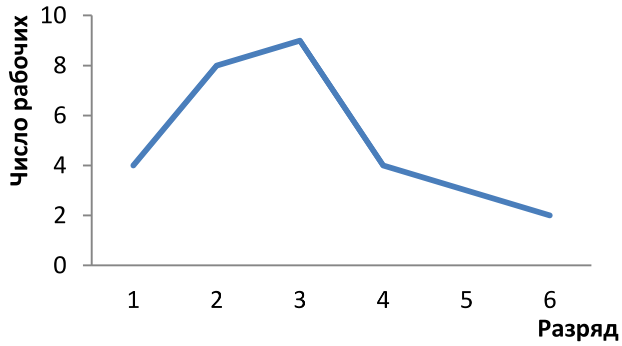

Рис.6.2. Полигон распределения рабочих предприятия по тарифному разряду

Графическое изображение интервального вариационного ряда представляют в виде **гистограммы**. Для построения гистограммы на оси абсцисс в соответствии с принятым масштабом откладывают границы интервалов. Эти интервалы служат основаниями прямоугольников, высота которых пропорциональна частотам или частостям. Гистограмма изображена на рисунке 6.3 и отражает распределение рабочих по дневной выработке (данные таблицы 6.3).

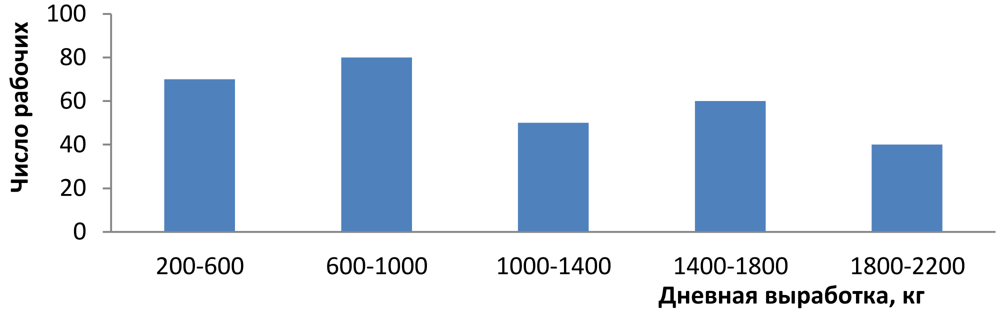

Рис.6.3 - Гистограмма распределения рабочих по дневной выработке

Для иллюстрации рядов распределения также используется **кумулятивная кривая (кумулята)**. Для ее построения на оси абсцисс откладываются значения дискретного признака, а на оси ординат – нарастающие частоты или частости, соответствующие этим значениям признака. На рисунке 6.4 изображена кумулята распределения рабочих по тарифному разряду (квалификации) (на основе данных таблицы 6.2).

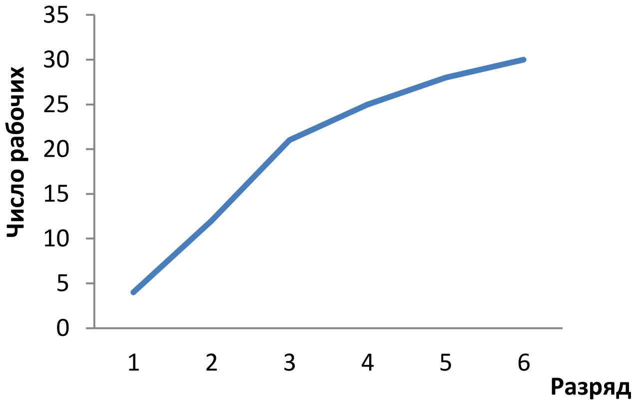

Рис. 6.4. Кумулята распределения рабочих
по тарифному разряду

В интервальных вариационных рядах с равными интервалами частоты и частости дают представление о степени их заполнения единицами совокупности. Если же интервалы не равны, напрямую сравнивать частоты или частости нельзя, поскольку их величины зависят не только от значений признака, которые определяют границы интервалов, но и от величины интервалов. Для таких рядов определяют плотность интервалов, или плотность распределения.

Плотность распределения показывает, сколько единиц совокупности или какой их процент приходится в среднем на единицу величины интервала. Она определяется делением частоты или частости соответствующего интервала на его ширину.

Для количественной оценки рядов распределения используют следующие группы статистических характеристик:

- характеристики центра распределения и порядковые статистики;
- характеристики вариации;
- характеристики формы распределения;
- характеристики дифференциации и концентрации (см. раздел 7).

6.2. ХАРАКТЕРИСТИКИ ЦЕНТРА РАСПРЕДЕЛЕНИЯ

К характеристикам центра распределения относятся средняя величина, мода, медиана.

Средняя величина характеризует типичный размер признака, дает обобщающую характеристику всего многообразия его индивидуальных значений.

По данным ряда распределения средняя величина определяется как арифметическая взвешенная.

Формула расчета средней величины на основе частот ряда распределения имеет следующий вид:

$$\bar{x} = \frac{\sum xf}{\sum f}. \quad (6.1)$$

Если вместо частот в процессе вычисления средней величины по данным ряда распределения используются частости, то формула расчета запишется:

$$\bar{x} = \sum xd. \quad (6.2)$$

При исследовании интервальных рядов предполагается равномерное распределение элементов совокупности в пределах каждого интервала. Поэтому в качестве варианта x используют середину интервала. При этом ширину открытого интервала условно считают равной ширине соседнего закрытого интервала. Данная ситуация представлена в таблице 6.3. После определения середины каждого из интервалов, в том числе открытых, производим расчет средней величины по формуле 6.1:

$$\bar{x} = \frac{328000}{300} = 1093{,}33 \text{ кг}$$

Таким образом, средняя по предприятию выработка 1 рабочего составляет 1093,33 кг.

Мода (*Mo*) – это наиболее распространенное значение признака.

В дискретном вариационном ряду мода – это вариант, который имеет наибольшую частоту (частость). В интервальном ряду по такому же принципу определяется модальный интервал, а конкретное модальное значение внутри интервала вычисляется по формуле:

$$Mo = x_{Mo} + h \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})}, \qquad (6.3)$$

где x_{Mo} – нижняя граница модального интервала;

h – ширина интервала;

$f_{Mo}, f_{Mo-1}, f_{Mo+1}$ – частоты (частости) модального, предмодального и послемодального интервалов соответственно.

Определим модальную выработку на предприятии по данным таблицы 6.3:

$$Mo = 600 + 400 \cdot \frac{80-70}{(80-70)+(80-50)} = 700 \text{ кг.}$$

Таким образом, на данном предприятии наиболее часто встречаются рабочие, средняя дневная выработка которых составляет 700 кг.

Поскольку мода не зависит от крайних значений признака, то ее целесообразно использовать в том случае, если ряд распределения имеет неопределенные границы.

Характеристикой центра распределения является также медиана (*Me*).

Медиана – значение признака, которое приходится на середину ранжированного ряда и делит его пополам – на две равные по объему части.

Таким образом, у одной половины единиц совокупности значение варьирующего признака меньше медианы, у другой – больше.

Если совокупность, которую мы исследуем, содержит нечетное число единиц, значение медианы имеет единица совокупности, расположенная в середине ранжированного ряда.

Если ранжированный ряд имеет четное число единиц, то ни одна из них не будет иметь значения медианы. Медиана в этом случае равна половине суммы двух значений признаков единиц совокупности, которые занимают среднее положение в ранжированном ряду.

Для вычисления медианы в рядах распределения используют кумулятивные частоты или частости, на основе которых определяется номер той единицы совокупности, которая имеет значение медианы.

В дискретном ряду медианой будет значение признака, первая кумулятивная частота (частость) которого равна или превышает половину объема совокупности. В интервальном ряду по этому принципу определяют медианный интервал, а значение медианы внутри интервала вычисляют по формуле:

$$Me = x_{Me} + h \cdot \frac{0{,}5 \sum f - S_{Me-1}}{f_{Me}}, \qquad (6.4)$$

где x_{Me}– нижняя граница медианного интервала;

$0{,}5 \sum f$ - половина объема совокупности;

S_{Me-1} – кумулятивная (накопленная) частота (частость) предмедианного интервала;

f_{Me} – частота (частость) медианного интервала.

Рассмотрим расчет медианы на основе данных о распределении рабочих по средней дневной выработке (таблица 6.3):

$$Me = 600 + 400 \cdot \frac{150-70}{80} = 1000 \text{ кг.}$$

Итак, на предприятии у половины робочих выработка менее 1000 кг, а у половины – более.

Таким образом, мода и медиана, в отличие от средней арифметической, определяются не на основе всех значений показателя у единиц совокупности, и их величина не зависит от крайних значений признака.

Во многом значения моды и медианы обусловлены величиной интервалов группировки.

Если распределение по форме приближается к нормальному, то между средней арифметической, модой и медианой существует следующее приближенное равенство:

$$Mo \approx \bar{x} - 3(\bar{x} - Me) \qquad (6.5)$$

или

$$\bar{x} \approx Mo - 3(\bar{x} - Me). \qquad (6.6)$$

Это соотношение играет важную роль для характеристики формы распределения, что будет рассмотрено ниже.

Определение моды и медианы, наряду со средней величиной, имеет особое практическое значение.

Например, в процессе планирования объема производства и реализации одежды, обуви ориентируются на наиболее распространенный размер, т.е. модальный.

Медиана имеет **минимальное** свойство, которое заключается в том, что сумма абсолютных отклонений вариантов от медианы – величина минимальная, т.е. меньше суммы абсолютных отклонений от любой другой величины. Это свойство медианы используется при проектировании размещения остановок городского транспорта, торговых и бытовых предприятий, так как только медиана дает возможность определить наименьшее расстояние до этих объектов.

6.3. КВАНТИЛИ РАСПРЕДЕЛЕНИЯ И ИХ РОЛЬ ПРИ ОПИСАНИИ ЗАКОНОМЕРНОСТЕЙ РАСПРЕДЕЛЕНИЯ

Для углубленного изучения структуры рядов распределения и выявления их характерных особенностей важное значение имеют так называемые порядковые статистики – квантили.

Квантиль – это значение признака у тех единиц совокупности, которые занимают определенное порядковое место в совокупности, упорядоченной по возрастанию (увеличению) или убыванию (уменьшению) значений исследуемого признака. С помощью квантилей можно найти значение признака, которое делит ряд на четыре части, пять частей, десять, сто. В соответствии с этим среди квантилей различают квартили, квинтили, децили, перцентили. К квантилям относится также и медиана, которая разбивает ряд на две равные части.

Квартили – это варианты, которые делят объем совокупности на четыре равные части. То есть квартилей насчитывается три.

Первый (нижний) квартиль (Q_1) характеризуется тем, что 25% единиц совокупности имеют значение признака меньше квартиля, а 75% единиц совокупности – больше. Формула расчета первого квартиля выглядит следующим образом:

$$Q_1 = x_{Q_1} + h \cdot \frac{0{,}25 \sum f - S_{Q_1 - 1}}{f_{Q_1}}, \qquad (6.7)$$

где x_{Q_1} – нижняя граница квартильного интервала;

S_{Q_1-1} – кумулятивная частота (частость) предквартильного интервала;

f_{Q_1} – частота (частость) квартильного интервала.

Третий (верхний) квартиль (Q_3) характеризуется тем, что 75% единиц совокупности имеют значение признака меньше квартиля, а 25% единиц – больше. Формула расчета запишется:

$$Q_3 = x_{Q_3} + h \cdot \frac{0{,}75 \sum f - S_{Q_3 - 1}}{f_{Q_3}}, \qquad (6.8)$$

где x_{Q_3} – нижняя граница квартильного интервала;

S_{Q_3-1} – кумулятивная частота (частость) предквартильного интервала;

f_{Q_3} – частота (частость) квартильного интервала.

Медиану можно рассматривать как второй квартиль.

Децили – это значения признака, которые делят совокупность на 10 равных частей, то есть каждая часть содержит 10% единиц совокупности. Децилей может быть девять, при этом пятым децилем является медиана.

Например, первый дециль характеризуется тем, что у 10% единиц совокупности значение признака меньше дециля, а у 90% – больше. Формула расчета первого дециля следующая:

$$D_1 = x_{D_1} + h \cdot \frac{0{,}1\sum f - S_{D_1-1}}{f_{D_1}}, \qquad (6.9)$$

где x_{D_1} – нижняя граница децильного интервала;

S_{D_1-1} – кумулятивная частота (частость) преддецильного интервала;

f_{D_1} – частота (частость) децильного интервала.

Результаты расчета первого дециля по данным таблицы 6.3 следующие:

$$D_1 = 200 + 400 \cdot \frac{30-0}{70} = 371{,}4 \text{ кг.}$$

Следовательно, на данном предприятии у 10 % рабочих средняя дневная выработка не превышает 371, 4 кг.

Квинтили - это значения признака, которые разбивают совокупность на пять равных частей, т. е. на части, каждая из которых содержит 20% единиц.

Например, первый квинтиль характеризуется тем, что у 20% единиц совокупности значение признака меньше квинтиля, а у 80% единиц – больше. Рассчитывают также второй, третий и четвертый квинтили. Формулы расчета квинтилей строят аналогично приведенным выше.

6.4. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ХАРАКТЕРИСТИКИ ВАРИАЦИИ

Обобщающей характеристикой ряда распределения является средняя величина. Средняя характеризует ряд в целом, но не показывает, как размещены вокруг нее отдельные значения, т. е. она не показывает характера вариации признака и степени его колебания (отклонения).

Вариация – изменение значения признака у отдельных единиц совокупности.

Для измерения вариации используют характеристики:

- абсолютные;
- относительные.

Абсолютные показатели вариации являются именованными величинами, т. е. они имеют такие же единицы измерения, как и признаки, которые изучаются. Рассмотрим их.

Размах вариации (R) характеризует диапазон вариации и рассчитывается как разность между максимальным и минимальным значениями признака:

$$R = x_{max} - x_{min}. \qquad (6.10)$$

В интервальном ряду распределения размах вариации определяют как разность между верхней границей последнего интервала и нижней границей первого или как разность между средними значениями в этих интервалах.

По данным таблицы 6.3 размах вариации составит:

R = 2000 – 400 = 1600 кг.

Таким образом, на данном предприятии максимальная дневная выработка больше минимальной выработки на 1600 кг.

Преимуществом размаха вариации как ее меры является простота расчета. Но данный показатель учитывает лишь крайние (граничные) значения признака и поэтому не дает представления о степени вариации внутри совокупности.

Другие абсолютные характеристики вариации учитывают все отклонения значений признака от средней величины.

Среднее линейное отклонение ($\bar{l}$) представляет собой среднюю арифметическую из абсолютных значений отклонений отдельных вариантов от их средней величины, т. е. оно показывает, на сколько единиц по модулю отклоняются в среднем отдельные значения признака от средней.

Как и средние величины, большинство показателей вариации имеют две формы: простую и взвешенную, которые используются соответственно для первичной, несгруппированной информации, и для вторичной, сгруппированной.

Запишем простую форму среднего линейного отклонения:

$$\bar{l} = \frac{\sum|x-\bar{x}|}{n} \qquad (6.11)$$

и взвешенную:

$$\bar{l} = \frac{\sum|x-\bar{x}|f}{\sum f}. \tag{6.12}$$

Рассчитаем среднее линейное отклонение по данным таблицы 6.4.

На основе формулы 6.12:

$$\bar{l} = \frac{144000}{300} = 480 \text{ кг.}$$

Это значит, что дневная выработка отдельных рабочих отклоняется от средней по предприятию выработки в среднем (по модулю) на 480 кг.

Среднее линейное отклонение объективно отражает абсолютные размеры вариации признака, который исследуется, на основе всех его значений. Но этот показатель вариации не учитывает знаков отклонений.

Чтобы устранить этот недостаток, отклонения значений признака от их средней величины возводят в квадрат и рассчитывают показатель вариации, который называют дисперсией.

Таблица 6.4

К расчету показателей вариации средней дневной выработки рабочих

Группы рабочих по выработке, кг	Число рабочих, чел. f	Середина интервала, кг x	$\|x-\bar{x}\|$	$\|x-\bar{x}\|f$	$(x-\bar{x})^2 f$
Менее 600	70	400	693,33	48530	33647310
600 - 1000	80	800	293,33	23470	6884460
1000 - 1400	50	1200	106,67	5330	568550
1400 - 1800	60	1600	506,67	30400	15402770
1800 и более	40	2000	906,67	36270	32884920
Итого	300	-	-	144000	89388010

Дисперсия (σ^2) представляет собой средний квадрат отклонений вариантов от их средней арифметической величины. Простая и взвешенная формы дисперсии имеют соответственно вид:

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n} \qquad (6.13)$$

и $$\sigma^2 = \frac{\sum(x-\bar{x})^2 f}{\sum f}. \qquad (6.14)$$

Дисперсию записывают без единиц измерения, поскольку они лишены экономического содержания.

На основе данных таблицы 6.4 и формулы 6.14 произведём расчет дисперсии:

$$\sigma^2 = \frac{89388010}{300} = 297960{,}03.$$

Чтобы получить показатель вариации, который имел бы такие же единицы измерения, как и признак, который исследуется, определяют среднее квадратическое отклонение как корень квадратный из дисперсии.

Среднее квадратическое отклонение (σ) характеризует меру абсолютного колебания признака относительно средней величины. Простая и взвешенная формы имеют вид:

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} \qquad (6.15)$$

и $$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2 f}{\sum f}}. \qquad (6.16)$$

Все рассмотренные показатели вариации характеризуют абсолютный размер вариации. Вследствие этого они не пригодны для сравнения степени вариации различных признаков или одного признака во времени. Для решения этой задачи используются относительные показатели вариации – коэффициенты вариации.

Коэффициент вариации (V) характеризует относительное колебание значений признака относительно средней величины и представляет собой выраженное в процентах отношение среднего линейного, среднего квадратического отклонения или размаха вариации к средней арифметической.

Рассчитывают следующие коэффициенты вариации:

линейный – $$V_l = \frac{\bar{l}}{\bar{x}} \cdot 100, \qquad (6.17)$$

квадратичный – $V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100,$ (6.18)

осцилляции – $V_R = \frac{R}{\bar{x}} \cdot 100.$ (6.19)

На основе формул 6.16 и 6.18 рассчитаем среднее квадратическое отклонение и квадратический коэффициент вариации выработки:

$$\sigma = \sqrt{297960{,}03} = 545{,}86 \text{ кг},$$

$$V_\sigma = \frac{545{,}86}{1093{,}33} \times 100 = 49{,}9\ \%.$$

Таким образом, на предприятии выработка отдельных рабочих отклоняется от средней выработки в среднем на 545,86 кг или на 49,9 %.

В соответствии с математическими свойствами величина среднего квадратического отклонения больше величины среднего линейного отклонения. Соответственно и квадратический коэффициент вариации больше, чем линейный.

Квадратический коэффициент вариации используют как критерий однородности совокупности по признакам, которые изучаются. Если квадратический коэффициент вариации меньше 33%, то совокупность считается однородной, а средняя величина - надежной и типичной.

В практике часто используется следующая формула дисперсии, которую получают путем несложных математических преобразований:

$$\sigma^2 = \overline{x^2} - \bar{x}^2, \quad (6.20)$$

где $\overline{x^2} = \frac{\sum x^2 f}{\sum f}$ – средний квадрат вариантов,

$\bar{x}^2 = (\frac{\sum xf}{\sum f})^2$ – квадрат средней величины.

Определим дисперсию вариации отпусков работников предприятия по данным таблицы 6.5.

На основании формулы 6.20:

$$\sigma^2 = \frac{33192}{100} - \left(\frac{1788}{100}\right)^2 = 12{,}23.$$

Среднеквадратическое отклонение составит (формула 6.16):

$$\sigma = \sqrt{12{,}23} = 3{,}5 \text{ дней},$$

а квадратический коэффициент вариации (формула 6.18):

$$V_\sigma = \frac{3{,}5}{17{,}88} \times 100 = 19{,}6\ \%.$$

Можно сформулировать вывод о том, что продолжительности отпусков отдельных работников фирмы отклоняются от средней продолжительности отпуска в среднем на 3,5 дня или на 19,6 %.

Таблица 6.5

Распределение работников фирмы по продолжительности отпуска

Продолжительность отпуска, дней x	Число работников в % к итогу f	xf	$x^2 f$
15	48	720	10800
18	30	540	9720
24	22	528	12672
Итого	100	1788	33192

Для **альтернативного признака**, вариация которого имеет два взаимоисключающих значения – «1» и «0», а распределение характеризуется двумя частостями:

p – доля единиц, обладающих признаком,

q – доля единиц, не обладающих признаком,

дисперсия рассчитывается как произведение частостей:

$$\sigma^2 = p \cdot q = p \cdot (1 - p). \qquad (6.21)$$

6.5. ХАРАКТЕРИСТИКИ ФОРМЫ РАСПРЕДЕЛЕНИЯ: КОЭФФИЦИЕНТЫ АСИММЕТРИИ И ЭКСЦЕССА

Анализ закономерностей распределения предусматривает оценку степени однородности совокупности. **Однородными** считаются такие совокупности, элементы которых имеют общие свойства и относятся к одному типу, классу. Как было отмечено выше, критерием однородности совокупности является величина квадратического коэффициента вариации. Об однородности также свидетельствует одновершинность кривых распределения, о неоднородности – многовершинность. Среди одновершинных распределений встречаются симметричные и асимметричные (скошенные), остро- и плосковершинные.

В симметричном распределении равноудалённые от центра значения признака имеют одинаковые частоты, а в **асимметричном** вершина распределения смещена, сдвинута. Если вершина смещена влево, имеем правостороннюю асимметрию, и наоборот.

В симметричном распределении характеристики центра – средняя, мода, медиана – имеют одинаковые значения, в асимметричном распределении между ними существуют расхождения (несовпадения).

При **правосторонней асимметрии** $\bar{x} > Me > Mo$.

В случае **левосторонней**, наоборот, $\bar{x} < Me < Mo$.

Чем больше асимметрия, тем больше отклонение $\bar{x} - Mo$.

Мерой асимметрии является **коэффициент асимметрии** (A):

$$A = \frac{\bar{x} - Mo}{\sigma}. \qquad (6.22)$$

Он характеризует направление и меру скошенности внутри распределения: при правосторонней асимметрии $A > 0$, при левосторонней – $A < 0$.

Другим свойством одновершинных распределений является степень сосредоточенности элементов совокупности вокруг центра распределения. Это свойство называют **эксцессом распределения**.

Комплексная оценка асимметрии и эксцесса осуществляется на основе центральных моментов распределения. **Центральный момент распределения** – это средняя арифметическая i-й степени отклонения индивидуальных значений признака от средней:

$$\mu_i = \frac{\sum (x - \bar{x})^i f}{\sum f}. \qquad (6.23)$$

Моментом второго порядка является дисперсия, которая характеризует вариацию.

Момент 3-го порядка характеризует асимметрию. В симметричном распределении $\mu_3 = 0$.

Для сравнения степени асимметрии различных распределений используется стандартизированный момент:

$$A_S = \frac{\mu_3}{\sigma^3} \ . \qquad (6.24)$$

Он зависит от крайних значений признака:

- при правосторонней асимметрии $A_S > 0$;
- при левосторонней $A_S < 0$.

Поэтому правосторонняя асимметрия называется **положительной**, а левосторонняя асимметрия – **отрицательной**.

Считается, что при $A_S < 0{,}25$ **асимметрия низкая**, если $A_S < 0{,}5$ – **средняя**, при $A_S > 0{,}5$ – **высокая**.

Для измерения эксцесса используется стандартизированный момент 4-го порядка:

$$E_i = \frac{\mu_4}{\sigma^4} \ . \qquad (6.25)$$

В **симметричном**, близком к нормальному, распределении $E_i = 3$, при **плосковершинном** $E_i < 3$, при **островершинном** $E_i > 3$.

Коэффициенты асимметрии и эксцесса – удобное средство для проверки близости распределения к нормальному. Условие нормального распределения играет важную роль при применении различных статистических методов.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Что представляет собой ряд распределения?
2. Назовите элементы ряда распределения и дайте им характеристику.
3. Приведите примеры атрибутивных рядов распределения.

4. Какие виды графиков используют для изображения атрибутивных рядов распределения?
5. Приведите примеры дискретных вариационных рядов. Какой вид графика используется для изображения дискретного ряда и каковы правила его построения?
6. Какие признаки лежат в основе интервальных вариационных рядов?
7. Назовите вид графика для отображения интервальных вариационных рядов. Сформулируйте правила его построения.
8. Что представляет собой кумулятивная кривая?
9. Дайте понятие плотности распределения.
10. Какие группы статистических характеристик используют для количественной оценки рядов распределения?
11. Как рассчитывается средняя величина в ряду распределения?
12. Сформулируйте понятие моды и запишите формулу ее расчета для интервального ряда распределения.
13. Что такое медиана и какова ее роль в экономических исследованиях?
14. Дайте понятие квантилей распределения и назовите известные вам квантили.
15. Расшифруйте понятие квартилей.
16. Сколько в ряду распределения можно рассчитать децилей?
17. Сформулируйте суть вариации и назовите абсолютные и относительные характеристики вариации.
18. Что представляет собой среднее линейное и среднее квадратическое отклонение и как их определяют при несгруппированных и сгруппированных данных?
19. Приведите формулы дисперсии по сгруппированным и несгруппированным данным.
20. Для чего предназначены относительные характеристики вариации и как они рассчитываются?
21. Как рассчитываются коэффициенты асимметрии и эксцесса?

РАЗДЕЛ 7. АНАЛИЗ КОНЦЕНТРАЦИИ, ДИФФЕРЕНЦИАЦИИ И ПОДОБИЯ СТРУКТУР

7.1. АНАЛИЗ НЕРАВНОМЕРНОСТИ РАСПРЕДЕЛЕНИЯ

Необходимо более детально (глубоко) проанализировать ряды распределения:

- с помощью оценки неравномерности распределения единиц совокупности по значению группировочного признака;
- путем исследования степени концентрации.

Для этого соответственно рассчитывают показатели дифференциации и концентрации.

Для **характеристики степени дифференциации** рассчитывают следующие показатели:

- коэффициент дифференциации;
- коэффициент децильной дифференциации;
- коэффициент квинтильной дифференциации.

Коэффициент дифференциации имеет вид:

$$K_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} \cdot 100, \quad (7.1)$$

где Q_1 – первый квартиль;

Q_3 – третий квартиль.

Децильные и квинтильные коэффициенты дифференциации широко используются для характеристики уровня материального обеспечения населения. **Коэффициент децильной дифференциации** представляет собой отношение девятого дециля к первому:

$$K_D = \frac{D_9}{D_1}. \quad (7.2)$$

Он характеризует отношение нижней границы доходов или расходов 10 % наиболее обеспеченного населения и верхней границы доходов или расходов 10 % наименее обеспеченного населения.

Коэффициент квинтильной дифференциации – это отношение четвертого квинтиля к первому:

$$K_q = \frac{q_4}{q_1}. \tag{7.3}$$

Квинтильный коэффициент характеризует отношение нижней границы доходов (расходов) 20 % наиболее обеспеченного населения и верхней границы доходов (расходов) 20 % наименее обеспеченного населения.

По данным таблицы 7.1, в которой представлено распределение населения Украины по уровню среднедушевых денежных доходов, рассчитаем коэффициент квинтильной дифференциации.

Таблица 7.1

Распределение населения Украины по уровню среднедушевых денежных доходов в 2009 году (проценты)

Группы населения по доходам, грн.	Численность в % к итогу f	Накопленные частоты, % S
До 480	9,5	9,5
480 – 840	36,1	45,6
840 – 1200	27,8	73,4
1200 – 1560	14,0	87,4
1560 – 1920	6,4	93,8
Более 1920	6,2	100,0
Итого	100,0	-

Квинтили делят совокупность на 5 равных частей. Первый квинтиль отсекает, отделяет 20 % совокупности снизу. Формула его расчета такая:

$$q_1 = x_{q_1} + h \cdot \frac{0{,}2\sum f - S_{q_1-1}}{f_{q_1}}.$$

Для определения первого квинтильного интервала строим графу накопленных частот и останавливаемся на том интервале, для которого накопленная частота впервые равна или превышает $0{,}2\sum f = 20$. Это интервал 480 – 840 грн. Рассчитаем первый квинтиль:

$$q_1 = 480 + 360 \cdot \frac{20-9{,}5}{36{,}1} = 584{,}7 \text{ грн.}$$

Таким образом, в Украине в 2009 году у 20 % населения среднедушевые денежные доходы не превышали 584,7 грн.

Четвертый квинтиль отсекает 20 % совокупности сверху. Он рассчитывается по формуле:

$$q_4 = x_{q_4} + h \cdot \frac{0{,}8\sum f - S_{q_4-1}}{f_{q_4}}.$$

Определяем четвертый квинтильный интервал. Он характеризуется тем, что накопленная частота, соответствующая данному интервалу, должна впервые быть равной или превышать $0{,}8\sum f$, то есть 80 %. Это интервал 1200 – 1560 грн. Производим расчет:

$$q_4 = 1200 + 360 \cdot \frac{80-73{,}4}{14} = 1369{,}7 \text{ грн.}$$

Данное значение показателя свидетельствует о том, что у 20 % населения Украины в 2009 году среднедушевые денежные доходы превышали 1369,7 грн.

Определяем коэффициент квинтильной дифференциации на основании формулы 7.3:

$$K_q = \frac{1369{,}7}{584{,}7} = 2{,}3.$$

Можно сформулировать вывод о том, что в Украине в 2009 году минимальные среднедушевые денежные доходы 20 % наиболее обеспеченного населения превышали максимальные среднедушевые денежные доходы 20 % наименее обеспеченного населения в 2,3 раза.

Перейдем к исследованию **степени концентрации**.

Вместе с распределением единиц совокупности по какому-либо признаку целесообразно изучение распределения по этим же группам другого показателя (таблица 7.2). Второе распределение может быть как равномерным (соответствовать распределению количества единиц совокупности), так и неравномерным.

Данное направление анализа актуально для характеристики равномерности распределения материальных и финансовых ресурсов по регионам; числа занятых – по отраслям; доходов и имущества – по отдельным группам населения.

Неравномерность распределения оценивают прежде всего с помощью коэффициентов концентрации и локализации.

Оценка концентрации основывается на отклонениях частостей двух распределений: первое распределение – это распределение по числу единиц совокупности d_j; второе – распределение по величине признака D_j.

Если распределение по величине признака в совокупности равномерное, то частости d_j=D_j; отклонения частостей свидетельствуют об определенной концентрации. Верхняя граница суммы отклонений $\sum\left|d_j - D_j\right| = 2$, а поэтому **коэффициент концентрации** определяется как полусумма модулей отклонений:

$$K = \frac{1}{2}\sum\left|d_j - D_j\right| \qquad (7.4)$$

Границы колебаний: $0 \leq K \leq 1$:

- при K = 0, констатируем равномерное распределение;
- при K = 1, фиксируем полную концентрацию;
- чем большая степень концентрации, тем больше значение K.

Представленные в таблице 7.2 данные о распределении предприятий региона по формам собственности и по количеству занятого населения свидетельствуют о неравномерности распределения.

На основании формулы 7.4 произведем расчет коэффициента концентрации:

$$K = \frac{0{,}7}{2} = 0{,}35.$$

Таблица 7.2

К расчету коэффициентов концентрации и локализации

Форма собственности	В % к итогу: число предприятий d_j	численность занятого населения D_j	Модуль отклонения частостей $\frac{1}{100}\cdot\left\|d_j - D_j\right\|$	$\frac{D_j}{d_j}$
Государственная	4,4	39,4	0,350	8,955
Коллективная	57,5	31,4	0,261	0,546
Частная	29,6	29,1	0,005	0,983
Другие	8,5	0,1	0,084	0,012
Итого	100,0	100,0	0,700	-

Величина данного показателя свидетельствует о среднем уровне концентрации занятых на предприятиях различных форм собственности (или о средней степени неравномерности распределения).

Таким образом, коэффициент концентрации является обобщающей характеристикой отклонения распределения от равномерного.

Для каждой j-й группы совокупности рассчитывают коэффициент локализации. **Коэффициент локализации** характеризует соотношение частостей двух распределений и часто используется для оценки территориального распределения:

$$L_j = \frac{D_j}{d_j}\cdot 100. \quad (7.5)$$

При равномерном распределении все значения $L_j = 1$. В случае концентрации значений признака в j-й группе $L_j > 1$ и наоборот.

Рассчитанные по данным таблицы 7.2 коэффициенты локализации свидетельствуют о концентрации занятого населения на предприятиях государственной формы собственности.

Для обеспечения сопоставимости структур одного объекта в динамике используют вторичные группировки. Для анализа изменений, например, в распределении населения по доходам (расходам), совокупность разбивают на равные по численности населения группы: на 10 групп по 10 % единиц в каждой или на 5 групп по 20 % единиц в каждой.

По такому принципу строят кривую концентрации Лоренца и на ее основе рассчитывают коэффициент концентрации доходов Джини, в котором d_j – доля населения j-й социальной группы в общей численности населения, D_j – доля доходов j-й социальной группы населения в общем объеме доходов.

Кривая концентрации Лоренца дает наглядное представление о характере распределения. Ее строят в прямоугольной системе координат, где на оси абсцисс и ординат наносят одинаковые масштабы шкал от 0 до 100. На оси ОХ откладывают значения кумулятивных долей (частостей), которые характеризуют распределение населения, а на оси ОУ – кумулятивные частости доходов.

В случае равномерного распределения доходов кумулятивные частости населения и доходов совпадают. **Линией равномерного распределения** является диагональ квадрата. Она означает полное отсутствие концентрации. Какие-либо отклонения от диагонали свидетельствуют о неравномерности распределения. Чем больше кривая Лоренца будет отклоняться от диагонали квадрата, тем более неравномерное распределение и выше концентрация доходов в отдельных группах населения. При полностью неравномерном распределении точкам 0, 20, 40, 60, 80, 100 по оси ОХ будут соответствовать 0, 0, 0, 0, 0, 100 по оси ОУ. В таком случае получаем **линию полного неравномерного распределения**.

На основе сравнения площадей фигур, первая из которых ограничена диагональю и кривой Лоренца, а другая – диагональю и линией полного неравномерного распределения, рассчитывают **коэффициент концентрации доходов Джини** (G). На практике для его определения используют следующую приближенную формулу:

$$G = \sum S_{d_j} S_{D_{j+1}} + \sum S_{d_{j+1}} S_{D_j}. \quad (7.6)$$

Коэффициент Джини принимает значения от 0 до 1; чем ближе к 1, тем больше степень концентрации и неравномерности распределения доходов между группами населения.

В таблице 7.3 приведены данные о распределении доходов населения между 20 %-ми группами.

Таблица 7.3

Распределение денежных доходов населения региона по 20%-м группам

20%-ые группы населения	Доля группы		S_{d_j}	S_{D_j}
	в общей численности населения d_j	в общих денежных доходах D_j		
Первая (с наименьшими доходами)	0,2	0,080	0,2	0,080
Вторая	0,2	0,133	0,4	0,213
Третья	0,2	0,181	0,6	0,394
Четвертая	0,2	0,245	0,8	0,639
Пятая (с наибольшими доходами)	0,2	0,361	1,0	1,000
Итого	1,0	1,000	x	x

На рисунке 7.1 представлена кривая концентрации Лоренца, построенная по данным таблицы 7.3.

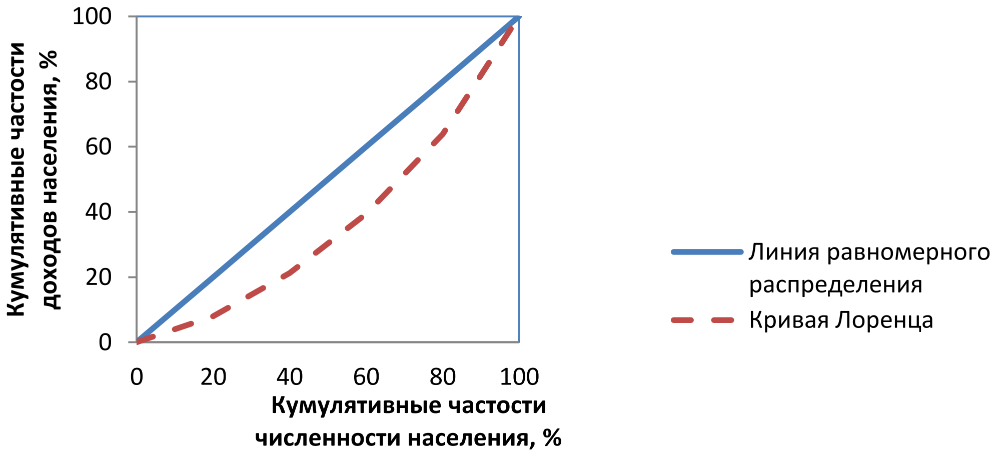

Рис. 7.1. Кривая Лоренца для характеристики распределения доходов населення

Произведем расчет коэффициента Джини по формуле 7.6:

$$G = (0{,}2 \cdot 0{,}213 + 0{,}4 \cdot 0{,}394 + 0{,}6 \cdot 0{,}639 + 0{,}8 \cdot 1{,}000) -$$

$$-(0{,}4 \cdot 0{,}08 + 0{,}6 \cdot 0{,}213 + 0{,}8 \cdot 0{,}394 + 1{,}0 \cdot 0{,}639) = 0{,}27.$$

Полученный результат свидетельствует о невысокой степени концентрации доходов в отдельных группах населения.

7.2. ОЦЕНКА ПОДОБИЯ СТРУКТУР РАЗНЫХ СОВОКУПНОСТЕЙ И ИНТЕНСИВНОСТИ СТРУКТУРНЫХ СДВИГОВ

Аналогично с коэффициентом концентрации (формула 7.4) рассчитывают коэффициент подобия структур двух совокупностей (P):

$$P = 1 - \frac{1}{2}\sum\left|d_j - d_k\right|. \qquad (7.7)$$

Если структуры одинаковые, то $P = 1$; если абсолютно противоположные, $P = 0$. Чем более схожие структуры, тем ближе значение P к 1.

В таблице 7.4 представлена структура потребительских расходов домохозяйств городской и сельской местности региона.

Рассчитаем по формуле 7.7 коэффициент подобия структур:

$$P = 1 - \frac{1}{2} \cdot 0{,}12 = 0{,}94.$$

Так как значение коэффициента близко к 1, то можно констатировать, что в целом структура потребительских расходов домохозяйств почти одинакова.

Таблица 7.4

Структура расходов домохозяйств региона

Потребительские расходы	Структура расходов в домохозяйствах местности, %		$\frac{1}{100} \cdot \lvert d_j - d_k \rvert$
	городской d_j	сельской d_k	
Продовольственные товары	68	67	0,01
Непродовольствен-ные товары	15	21	0,06
Оплата услуг	17	12	0,05
Итого	100	100	0,12

Структура любой совокупности динамична. Интенсивность структурных сдвигов оценивается с помощью среднего линейного ($\overline{l_d}$) или среднего квадратического (σ_d) отклонения частостей распределения:

$$\overline{l_d} = \frac{\sum\lvert d_{j_1} - d_{j_0}\rvert}{m}, \tag{7.8}$$

$$\sigma_d = \sqrt{\frac{\sum(d_{j_1} - d_{j_0})^2}{m}}, \tag{7.9}$$

где d_{j_0} и d_{j_1} – частости соответственно базисного и отчетного периодов;

m – число групп.

В данных формулах относительные величины структуры обычно выражаются в процентах, а не в долях. Проценты минус проценты, в отличие от их отношения, дают процентные пункты (п.п.).

В таблице 7.5 приведено распределение предприятий по отраслям в базисном и отчетном периодах.

Таблица 7.5

Распределение предприятий по отраслям экономики

Отрасли	Число предприятий в % к итогу		$(d_{j_1} - d_{j_0})^2$
	базисное d_{j_0}	отчетное d_{j_1}	
Добывающие	18	10	64
Обрабатывающие	17	15	4
Оказывающие услуги	65	75	100
Итого	100	100	168

Определим среднее квадратическое отклонение частостей распределения по формуле 7.9:

$$\sigma_d = \sqrt{\frac{168}{3}} = 7{,}5 \text{ п. п.}$$

Таким образом, в отчетном периоде по сравнению с базисным структура предприятий по отраслям экономики изменилась в среднем на 7,5 п.п.

7.3. ПРАВИЛО СЛОЖЕНИЯ ДИСПЕРСИЙ

В третьем разделе пособия нами формулировалась задача и рассматривались правила построения аналитической группировки.

Аналитическая группировка представляет собой распределение единиц совокупности по признаку-фактору x и характеристику групп средней величиной результативного признака y.

В такой совокупности дисперсия результативного признака y может быть разложена на: дисперсию в каждой группе (внутригрупповую) и дисперсию между группами (межгрупповую).

Общая дисперсия характеризует вариацию признака y около общей средней:

$$\sigma^2 = \frac{\sum(y-\bar{y})^2}{n}, \qquad (7.10)$$

где y - значения признака у отдельных единиц совокупности;

n – число единиц совокупности.

Межгрупповая дисперсия характеризует вариацию групповых средних ($\overline{y_j}$) около общей средней:

$$\delta^2 = \frac{\sum(\overline{y_j}-\bar{y})^2 f_j}{\sum f_j}, \qquad (7.11)$$

где f_j – частота j-й группы.

Внутригрупповая дисперсия рассчитывается отдельно для каждой j-й группы:

$$\sigma_j^2 = \frac{\sum(y-\overline{y_j})^2}{f_j}. \qquad (7.12)$$

Обобщающей характеристикой внутригрупповой вариации является **средняя из групповых дисперсий**:

$$\overline{\sigma^2} = \frac{\sum \sigma_j^2 f_j}{\sum f_j}. \qquad (7.13)$$

Взаимосвязь между тремя дисперсиями называют правилом сложения дисперсий:

$$\sigma^2 = \delta^2 + \overline{\sigma^2}\ . \qquad (7.14)$$

Правило сложения (разложения, декомпозиции) дисперсий звучит так: сумма межгрупповой и средней из групповых дисперсий равна общей дисперсии.

Рассмотрим вычисление дисперсий на следующем примере.

Количество деталей, которые изготавливают рабочие в разные смены, характеризуется следующими данными:

1-я смена: 26, 24, 23, 28, 25, 24;

2-я смена: 28, 30, 29, 33.

Для определения дисперсий условие удобно записать в форме таблицы:

Таблица 7.6

К расчету общей и групповой дисперсии

1-я смена			2-я смена		
№ рабочего	Число деталей, шт. y_1	y_1^2	№ рабочего	Число деталей, шт. y_2	y_2^2
1	26	676	1	28	784
2	24	576	2	30	900
3	23	529	3	29	841
4	28	784	4	33	1089
5	25	625	-	-	-
6	24	576	-	-	-
Итого	150	3766	Итого	120	3614

Определим групповые и общую средние по формуле средней арифметической простой (формула 5.8):

$$\overline{y_1} = \frac{150}{6} = 25 \text{ деталей;}$$

$$\overline{y_2} = \frac{120}{4} = 30 \text{ деталей;}$$

$$\bar{y} = \frac{150+120}{10} = 27 \text{ деталей.}$$

В таблице 7.7 приведена аналитическая группировка, характеризующая зависимость выработки от смены, в которой трудятся рабочие.

Таблица 7.7

Зависимость выработки от смены

Смена	Число рабочих f_j	Среднее число произведенных деталей 1 рабочим, шт. $\bar{y}_j$
1-я	6	25
2-я	4	30
Всего (в среднем)	10	$\bar{y}$ = 27

Рассчитаем групповые дисперсии по упрощенной формуле определения дисперсии (формула 6.20), которая для отдельных групп запишется:

$$\sigma_1^2 = \overline{y_1^2} - \overline{y_1}^2 = \frac{3766}{6} - 25^2 = 2{,}667;$$

$$\sigma_2^2 = \overline{y_2^2} - \overline{y_2}^2 = \frac{3614}{4} - 30^2 = 3{,}500.$$

Определим по формуле 6.20 общую дисперсию:

$$\sigma^2 = \frac{3766+3614}{10} - 27^2 = 9.$$

Средняя из групповых дисперсий (формула 7.13) будет равна:

$$\overline{\sigma^2} = \frac{2{,}667 \cdot 6 + 3{,}5 \cdot 4}{6+4} \approx 3.$$

Рассчитаем межгрупповую дисперсию (формула 7.11):

$$\delta^2 = \frac{(25-27)^2 \cdot 6 + (30-27)^2 \cdot 4}{6+4} = 6.$$

Осуществим проверку: сумма межгрупповой дисперсии и средней из групповых должна быть равной общей дисперсии:

6 + 3 = 9.

Итак, общая дисперсия для данного примера характеризует общую вариацию количества деталей, которые изготавливаются и в первую, и во вторую смены.

Внутригрупповые дисперсии отражают вариацию количества деталей, обусловленную всеми факторами, кроме фактора смены. Это могут быть различия в квалификации рабочих, в техническом состоянии станков, на которых рабочие изготавливают детали, проблемы с усталостью рабочих и т.д.

Межгрупповая дисперсия характеризует вариацию производства деталей, обусловленную только различием смен, в которых трудятся рабочие.

Отношение межгрупповой дисперсии к общей характеризует долю вариации результативного признака, которая обусловлена вариацией группировочного признака. Это отношение называется **эмпирическим индексом детерминации**:

$$\eta^2 = \frac{\delta^2}{\sigma^2}. \qquad (7.15)$$

В нашем примере:

$$\eta^2 = \frac{6}{9} = 0{,}667 \text{ или } 66{,}7\ \%.$$

Таким образом, 66,7 % общей вариации числа произведенных деталей обусловлены различием смен, в течение которых трудятся рабочие. Остальные 33,3 % общей дисперсии обусловлены вариацией всех прочих факторов.

Корень квадратный из эмпирического коэффициента детерминации называется **эмпирическим корреляционным отношением**:

$$\eta = \sqrt{\eta^2}\ . \qquad (7.16)$$

Данный показатель используется для оценки тесноты, близости связи между группировочным и результативным признаками.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Назовите показатели, используемые для характеристики степени дифференциации.
2. Опишите, что характеризует коэффициент децильной дифференциации.
3. Поясните смысл коэффициента квинтильной дифференциации.
4. Как рассчитывается и что характеризует коэффициент концентрации? Каковы границы его колебания?
5. Что представляет собой коэффициент локализации?
6. Поясните принцип построения кривой Лоренца. Какое явление описывается с помощью данной кривой?
7. Какие значения может принимать коэффициент концентрации доходов Джини и как он рассчитывается?
8. Что характеризует коэффициент подобия структур? Каковы границы его колебания?
9. С помощью каких показателей оценивают интенсивность структурных сдвигов?
10. Для какого вида группировок действует правило сложения дисперсий?
11. Что характеризуют и как рассчитываются групповые дисперсии?
12. Как определяется средняя из групповых дисперсий?
13. Как рассчитывается и что характеризует межгрупповая дисперсия?
14. Раскройте содержание общей дисперсии.
15. Сформулируйте правило сложения дисперсий.
16. Как рассчитывается и что характеризует эмпирический коэффициент детерминации и эмпирическое корреляционное отношение?

РАЗДЕЛ 8. ВЫБОРОЧНЫЙ МЕТОД

8.1. СУТЬ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ

При статистическом изучении явлений и процессов общественной жизни часто бывает нецелесообразным обследование всех единиц совокупности. Так, при обследовании бюджетов семей, определении средних расходов туристов, при проверке качества каких-либо изделий осуществление сплошного наблюдения затруднено или невозможно. В таких случаях используют несплошное наблюдение, разновидностью которого является выборочное наблюдение.

Выборочное наблюдение – это вид несплошного наблюдения, при котором обследуются не все элементы совокупности, которая изучается, а лишь определенным образом отобранная их часть.

Преимуществами выборочного обследования являются:

- обеспечение экономии материальных, трудовых, финансовых ресурсов и времени;
- возможность расширения программы наблюдения;
- при изучении определенного круга явлений выборочное наблюдение является единственно возможным, например, это касается проверки качества продукции.

Совокупность, из которой отбираются элементы для исследования, называют **генеральной** (общей).

Выборочная совокупность – это совокупность единиц, которые отобраны для исследования.

При проведении выборочного наблюдения необходимо учитывать следующее. Выборочная совокупность не точно отражает состав генеральной совокупности, а потому и выборочные оценки не совпадают с соответствующими характеристиками генеральной совокупности. Расхождения между ними называют **ошибками репрезентативности** (о них рассказывалось в разделе 2).

Итак, ошибки репрезентативности связаны с тем, что совокупность, которую отобрали и исследовали, недостаточно точно воспроизводит генеральную совокупность. **Ошибки репрезентативности бывают двух видов**:

- систематические;
- случайные.

Систематические ошибки возникают из-за тенденциозного, необъективного отбора единиц, при котором нарушаются принципы выборки.

Основной задачей выборочного метода является измерение случайных ошибок, возникающих при соблюдении всех требований проведения выборки.

Существуют следующие **требования к отбору**, **дающие объективную гарантию репрезентативности, представительности выборки**:

- выборка должна быть **случайной**, то есть каждая ее единица должна иметь одинаковую вероятность попасть в выборку;
- выборка должна быть из однородной совокупности, чтобы правильно репрезентовать генеральную совокупность.

При создании случайной выборки могут быть два подхода:

- отбор с помощью жеребьевки (лотереи);
- использование таблиц случайных чисел.

Способы формирования выборочной совокупности при жеребьевке:

- повторная выборка;
- бесповторная.

При **повторной выборке** единицы совокупности, которые попали в выборку, снова возвращаются в генеральную совокупность.

При **бесповторном отборе** единицы совокупности, которые попали в выборку, в генеральную совокупность уже не возвращаются.

По способу организации используют следующие виды выборки:

- простая случайная;
- механическая;
- типическая;
- серийная.

Самым распространенным способом отбора из генеральной совокупности является **простая случайная выборка**. При данном способе отбор единиц осуществляется из всей массы единиц генеральной совокупности без предварительного разделения ее на какие-либо группы и единицы отбора совпадают с единицами обследования. Важным условием репрезентативности случайного отбора является то, что каждой единице генеральной совокупности предоставляется одинаковая возможность попасть в выборочную совокупность. Это препятствует возникновению систематических ошибок отбора.

8.2. ВЫБОРОЧНЫЕ ОЦЕНКИ И ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ

Итак, в статистике принято строго различать параметры и свойства генеральной совокупности и их оценки по данным выборки.

Принята следующая система обозначений показателей в генеральной совокупности и оценочных показателей в выборочной (таблица 8.1).

Одной из задач выборочного наблюдения является определение ошибок выборки.

Ошибкой выборочного наблюдения является:

для средней – разница между выборочной и генеральной средней;

для доли – разница между выборочной и генеральной долей;

для дисперсии – отношение выборочной и генеральной дисперсии.

Таблица 8.1

Система обозначений показателей

Показатель	Генеральная совокупность	Выборочная совокупность
Объем совокупности (число единиц)	N	n
Среднее значение варьирующего признака	$\bar{x}$	$\tilde{x}$
Доля единиц, обладающих данным качеством	p	w
Общая дисперсия	σ_0^2	σ^2

По данным таблицы 8.2 рассчитаем средний балл студента, а также долю студентов, получивших неудовлетворительные оценки, по генеральной совокупности и по выборочной (20 % выборка).

Средний балл студента по данным ряда распределения рассчитывается по формуле средней арифметической взвешенной.

По генеральной совокупности средний балл одного студента:

$$\bar{x} = \frac{80\times2+300\times3+370\times4+250\times5}{1000} = 3{,}79 \text{ балла}$$

Средний балл студента по выборочной совокупности:

$$\tilde{x} = \frac{15\times2+55\times3+84\times4+46\times5}{200} = 3{,}81 \text{ балла}$$

Доля студентов, которые получили неудовлетворительные оценки в общей численности студентов:

по генеральной совокупности:

$$p = \frac{80}{1000} = 0{,}080$$

и по выборочной:

$$w = \frac{15}{200} = 0{,}075$$

Разность между показателями выборочной и генеральной совокупности покажет **случайную ошибку репрезентативности**:

$$\tilde{x} - \bar{x} = 3{,}81 - 3{,}79 = 0{,}02 \text{ балла}$$

$$w - p = 0{,}075 - 0{,}080 = -0{,}005$$

Таблица 8.2

Распределение студентов по успеваемости

Балл	Число студентов, чел.	
	Генеральная совокупность	Выборочная совокупность
2	80	15
3	300	55
4	370	84
5	250	46
Итого	1000	200

Таких выборок из генеральной совокупности может быть достаточно много. Каждая выборка дает свою ошибку. В связи с этим различают среднюю (стандартную) и предельную ошибки выборки.

Средняя ошибка выборки (μ) характеризует среднюю величину возможных расхождений выборочной и генеральной средней (или доли).

При **случайном повторном отборе средняя ошибка выборочной средней величины** определяется по формуле:

$$\mu = \sqrt{\frac{\sigma^2}{n}}, \tag{8.1}$$

где σ^2 - дисперсия показателя, который изучается, в генеральной совокупности (так как она, как правило, не известна, то фактически в формулу подставляется дисперсия выборочная, которая при большом числе наблюдений близка к генеральной).

При **бесповторном отборе** используется формула:

$$\mu = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}, \tag{8.2}$$

При бесповторном отборе подкоренное выражение умножается на величину $\left(1 - \frac{n}{N}\right)$, которая всегда меньше единицы, поэтому величина средней ошибки выборки при бесповторном отборе оказывается меньше, чем при повторном.

Средняя ошибка выборочной доли при случайном повторном отборе определяется так:

$$\mu = \sqrt{\frac{w(1-w)}{n}}. \qquad (8.3)$$

При бесповторном случайном отборе средняя ошибка выборочной доли рассчитывается по формуле:

$$\mu = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}\ . \qquad (8.4)$$

Предельная ошибка выборки (Δ) – это максимально возможная ошибка с принятой вероятностью. Она рассчитывается по формуле:

$$\Delta = t\mu,$$

где t– коэффициент доверия, определяющий размер ошибки в зависимости от принятого уровня вероятности P.

Уровень вероятности определяет величина коэффициента доверия и наоборот. Значения t даются в таблицах нормального распределения вероятностей. Чаще всего используют следующие сочетания:

t	P
1,0	0,683
1,5	0,866
2,0	0,954
2,5	0,988
3,0	0,997
3,5	0,999

Формула $\Delta = t\mu$ конкретизируется следующим образом (таблица 8.3).

Таблица 8.3

Формулы предельных ошибок средней и доли при простом случайном отборе

Параметры выборки	Повторная выборка	Бесповторная выборка
для средней	$\Delta_{\tilde{x}} = t\sqrt{\frac{\sigma^2}{n}}$ (8.5)	$\Delta_{\tilde{x}} = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$ (8.7)
для доли	$\Delta_w = t\sqrt{\frac{w(1-w)}{n}}$ (8.6)	$\Delta_w = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$ (8.8)

В практике выборочных обследований используют **два типа выборочных оценок**:

- точечные;
- интервальные.

Точечная оценка – это значение параметра по данным выборки: выборочная средняя и выборочная доля.

Интервальная оценка – это интервал значения параметра, рассчитанный по данным выборки для определенной вероятности, то есть доверительный интервал. Границы его определяются на основе точечных оценок и предельной ошибки выборки:

для средней $$\tilde{x} - t\mu \leq \bar{x} \leq \tilde{x} + t\mu, \quad (8.9)$$

для доли $$w - t\mu \leq p \leq w + t\mu. \quad (8.10)$$

В статистическом анализе часто бывает необходимо сравнить ошибки выборки различных признаков или же одного признака в разных совокупностях. Такие сравнения осуществляют с помощью **относительной ошибки,** которая показывает, на сколько процентов выборочная оценка может отклоняться от параметра генеральной совокупности.

Относительная стандартная ошибка средней – это коэффициент вариации выборочных средних:

$$V_\mu = \frac{\mu_x}{\tilde{x}} \cdot 100. \quad (8.11)$$

Ее размер можно определить также на основе коэффициента вариации признака V_x:

для повторной выборки

$$V_\mu = \frac{V_x}{\sqrt{n}} \cdot 100; \quad (8.12)$$

для бесповторной выборки

$$V_\mu = \frac{V_x}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} \cdot 100. \quad (8.13)$$

Относительную ошибку доли определяют по формуле:

$$V_\mu = \frac{\mu_w}{w} = \frac{\sqrt{\frac{w(1-w)}{n}}}{w} = \sqrt{\frac{1-w}{n \cdot w}}. \quad (8.14)$$

В анализе также используют предельную относительную ошибку, которая учитывает вероятность статистического вывода:

$$V_\Delta = tV_\mu. \quad (8.15)$$

В основе **простой случайной выборки** лежат понятия и категории, которые являются исходными при разработке других видов выборочного наблюдения.

Механической называется такая выборка, при которой генеральная совокупность объемом N единиц, расположенных в определенном порядке (алфавитном, по территории и т.д.), делят на n равных частей и из каждой части обследуют одну единицу.

Ошибки выборки при механическом отборе единиц определяют по формулам простой случайной выборки.

При прочих равных условиях:

$$\begin{matrix}\text{ошибка при} \\ \text{случайном бес}- \\ \text{повторном отборе}\end{matrix} < \begin{matrix}\text{ошибка при} \\ \text{механическом} \\ \text{отборе}\end{matrix} < \begin{matrix}\text{ошибка при} \\ \text{случайном} \\ \text{повторном отборе}\end{matrix}$$

Типический отбор – это способ формирования выборки с учетом структуры генеральной совокупности. Он предусматривает ее предварительную структуризацию и независимый отбор элементов в каждой составляющей.

При бесповторном отборе внутри типов (групп, объектов, территорий):

средняя ошибка выборки для средней:

$$\mu_x = \sqrt{\frac{\bar{\sigma}_\text{в}^2}{n}\left(1 - \frac{n}{N}\right)}, \quad (8.16)$$

где $\bar{\sigma}_\text{в}^2$- средняя из групповых дисперсий.

средняя ошибка выборки для доли:

$$\mu_w = \sqrt{\overline{\frac{w(1-w)}{n}}\left(1 - \frac{n}{N}\right)}, \quad (8.17)$$

где $\overline{w(1-w)}$ - средняя из долей групп.

При прочих равных условиях ошибка выборки при типическом отборе меньше, чем при случайном и механическом.

Суть **серийной выборки** заключается в том, что отбираются отдельные серии (группы, гнезда) единиц генеральной совокупности. На практике чаще всего встречается отбор с равными сериями. В отобранных методом случайного бесповторного или механического отбора сериях проводят сплошное наблюдение всех единиц, которые туда вошли.

При равных сериях **средняя ошибка бесповторной выборки** запишется:

$$\mu = \sqrt{\frac{\delta^2}{r}\left(1-\frac{r}{R}\right)}, \tag{8.18}$$

где r – количество отобранных серий,

R – общее количество серий в генеральной совокупности,

δ^2 – межсерийная дисперсия.

8.3. ОСНОВНЫЕ ТИПЫ ЗАДАЧ НА ВЫБОРКУ

На основе свойств нормального распределения **для одной конкретной выборки можно определить:**

- ошибки репрезентативности – среднюю и предельную с принятой вероятностью;
- вероятность того, что ошибка выборки не превысит допустимый уровень;
- объем выборки, который обеспечит необходимую точность результатов с принятой вероятностью.

Пример. Из партии лампочек (генеральная совокупность) в 1000 шт. отобрано способом случайной бесповторной выборки 100 шт. Средняя продолжительность горения лампочек по отобранной части составляет 1200 ч., а среднее квадратическое отклонение – 200 ч. С вероятностью P = 0,997 $(t=3)$

определим границы среднего значения продолжительности горения лампочек по всей совокупности.

Для этого вычислим предельную ошибку средней по формуле 8.7:

$$\Delta_{\tilde{x}} = 3 \cdot \sqrt{\frac{200^2}{100}\left(1 - \frac{100}{1000}\right)} = 57 \text{ ч.}$$

Пределы средней продолжительности горения лампочек по всей совокупности составят (формула 8.9):

$$1200 - 57 \leq \bar{x} \leq 1200 + 57;$$

$$1143 \text{ ч.} \leq \bar{x} \leq 1256 \text{ ч.}$$

Таким образом, с вероятностью 99,7 % можно утверждать, что средняя продолжительность горения лапочек по всей партии будет заключена в пределах от 1143 ч до 1257 ч.

В дополнение к данному примеру известно, что из отобранных лампочек 90 шт. удовлетворяли стандарту. С вероятностью P = 0,954 ($t = 2$) определим границы доли лампочек, удовлетворяющих стандарту по всей партии лампочек.

Доля лампочек по выборке, удовлетворяющих стандарту, составила 90 % ($w = \frac{90}{100} \cdot 100$).

Вычислим предельную ошибку доли по формуле 8.8:

$$\Delta_w = 2 \cdot \sqrt{\frac{0{,}9(1-0{,}9)}{100}\left(1 - \frac{100}{1000}\right)} = 0{,}057 \text{ или } 5{,}7\ \%.$$

Доля лампочек, удовлетворяющих стандарту по всей партии, будет находиться в следующих пределах (формула 8.10):

$$80 - 5{,}7 \leq p \leq 80 + 5{,}7;$$

$$73{,}4\ \% \leq p\ 85{,}7\ \%.$$

Итак, с вероятностью 95,4 % можно утверждать, что доля лампочек, удовлетворяющих стандарту во всей партии, будет заключена в пределах от 74,3 до 85,7 %.

Продемонстрируем пример задачи на определение вероятности того, что ошибка выборки не превысит допустимый уровень.

Пример. На основе выборочного обследования 600 рабочих (n = 600, применялся способ случайного повторного отбора) одной из отраслей экономики установлено, что удельный вес женщин составил 40 % ($w = 0{,}4$).

Рассчитаем вероятность, с которой можно будет утверждать, что при определении доли женщин, занятых в этой отрасли, допущена ошибка (Δ), не превышающая 5% (0,05).

Для определения вероятности допуска ошибки из формулы 8.6 определяем t:

$$t = \frac{\Delta}{\sqrt{\frac{w(1-w)}{n}}} = \frac{0{,}05}{\sqrt{\frac{0{,}4 \cdot 0{,}6}{600}}} = 2{,}5.$$

Для $t = 2{,}5$ вероятность P = 0,988, то есть с вероятностью 98,8 % можно утверждать, что при определении доли женщин в общем числе рабочих допущена ошибка не более 5 %.

Таблица 8.4

Формулы определения численности выборки

Способ отбора	Виды выборки			
	для средней		для доли	
	повторная выборка	бесповторная выборка	повторная выборка	бесповторная выборка
Простая случайная выборка	$\frac{t^2\sigma^2}{\Delta_{\tilde{x}}^2}$	$\frac{Nt^2\sigma^2}{N\Delta_{\tilde{x}}^2 + t^2\sigma^2}$	$\frac{t^2 w(1-w)}{\Delta_w^2}$	$\frac{Nt^2 w(1-w)}{N\Delta_w^2 + t^2 w(1-w)}$
Типическая выборка	$\frac{t^2\overline{\sigma^2}}{\Delta_{\tilde{x}}^2}$	$\frac{Nt^2\overline{\sigma^2}}{N\Delta_{\tilde{x}}^2 + t^2\overline{\sigma^2}}$	$\frac{t^2\overline{w(1-w)}}{\Delta_w^2}$	$\frac{Nt^2\overline{w(1-w)}}{N\Delta_w^2 + t^2\overline{w(1-w)}}$
Серийная выборка	$\frac{t^2{\sigma^2}_r}{\Delta_{\tilde{x}(r)}^2}$	$\frac{Rt^2{\sigma^2}_r}{R\Delta_{\tilde{x}(r)}^2 + t^2{\sigma^2}_r}$	-	-

Одной из основных задач выборочного метода является **определение численности выборки *n***. Пусть требуется определить, сколько рабочих на предприятии необходимо обследовать в порядке случайной повторной выборки для определения их средней выработки, чтобы с вероятностью 95,4 % можно было бы гарантировать ошибку не более 5 единиц продукции. Предполагаемое среднее квадратическое отклонение 20 единиц.

Из формулы 8.5 находим n:

$$n = \frac{t^2\sigma^2}{\Delta^2} = \frac{4 \cdot 20^2}{5^2} = 64 \text{ чел.}$$

Формулы определения необходимой численности выборки при различных видах и способах отбора приведены в таблице 8.4.

8.4. ОСОБЕННОСТИ МАЛОЙ ВЫБОРКИ

Малой принято считать выборку, объем которой не превышает 30 единиц. Такие выборки используются при оценке деятельности коммерческих банков, малых предприятий, во время обследования качества продукции.

Репрезентативность выборки в значительной мере зависит от ее объема. При достаточно большом объеме распределение случайных ошибок выборки приближено к нормальному закону распределения. В силу этого по таблице значений функции $F(t)$ можно определить вероятность того, что случайная ошибка не выйдет за определенные пределы. При этом допускается условие, что дисперсия выборочной совокупности приблизительно равна дисперсии генеральной совокупности.

Особенность малой выборки состоит в том, что ее **случайные ошибки не подчиняются закону нормального распределения**.

Закон распределения ошибок малой выборки впервые разработал английский статистик В. Госсет (1908 г.), который подписывал свои работы именем Стьюдента. Распределение, которое он открыл, получило название ***t*-распределение Стьюдента**. Отношение выборочной средней к генеральной Стьюдент выразил в виде формулы, которая называется **отношение Стьюдента**:

$$t = \frac{\tilde{x} - \bar{x}}{\mu_{\text{м.в.}}}, \qquad (8.19)$$

где $\mu_{\text{м.в.}}$ – стандартная ошибка малой выборки, которая рассчитывается по формуле:

$$\mu_{\text{м.в.}} = \frac{\sigma}{\sqrt{n-1}}, \tag{8.20}$$

Предельная ошибка малой выборки определяется по формуле:

$$\Delta = t\mu_{\text{м.в.}}, \tag{8.21}$$

где t – отношение Стьюдента.

Вероятность того, что ошибка выборки будет не больше заданного значения $|\tilde{x} - \bar{x}| \leq t\mu_{\text{м.в.}}$, представляет собой функцию *S(t, n)*, приведенную в таблицах Стьюдента в литературе по математической статистике.

Вероятность того, что генеральная средняя находится в определенных границах, определяется по формуле:

$$P[\tilde{x} - t\mu_{\text{м.в.}} < \bar{x} < \tilde{x} + t\mu_{\text{м.в.}}] = 2S(t,n) - 1. \tag{8.22}$$

Из таблиц Стьюдента следует, что при увеличении объема выборки распределение Стьюдента приближается к нормальному закону и при n = 20 оно мало отличается от нормального распределения.

Пример. Средняя продолжительность горения, установленная путем испытания 10 случайно отобранных лампочек, оказалась равной 1280 ч. при среднем квадратическом отклонении 18 ч.

Определим, с какой вероятностью можно будет утверждать, что допущенная при этом предельная ошибка выборки не превысит 12 ч.

Так как $n < 30$, имеем дело с малой выборкой. Определяем среднюю ошибку малой выборки по формуле 8.20:

$$\mu_{\text{м.в.}} = \frac{18}{\sqrt{9}} = 6 \text{ ч.}$$

Из формулы 8.21 находим:

$$t = \frac{\Delta}{\mu_{\text{м.в.}}} = \frac{12}{6} = 2.$$

Так как при малой выборке вероятность той или иной ошибки выборки подчиняется (соответствует) распределению Стьюдента, то вероятность того, что генеральная средняя находится в определенных границах, определяем по формуле 8.22. В математических таблицах распределения Стьюдента находим для заданных t и n значение $S(t, n)$, а затем рассчитываем $2S(t, n) - 1$.

В нашем примере для $n = 10 - 1$ и $t = 2$ получаем $S(t, n)$ = 0,962.

Вероятность допуска ошибки не более 12 ч. равняется $2 \cdot 0{,}962 - 1 = 0{,}924$ или 92,4 %.

8.5. СПОСОБЫ РАСПРОСТРАНЕНИЯ ДАННЫХ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ НА ГЕНЕРАЛЬНУЮ СОВОКУПНОСТЬ

Конечной целью выборочного наблюдения является характеристика генеральной совокупности на основании данных, полученных по выборочной совокупности.

Применяются **два способа распространения данных выборочного наблюдения на генеральную совокупность**:

- способ прямого пересчета;
- способ поправочных коэффициентов.

Способ прямого пересчета заключается в том, что выборочная средняя и доля умножаются на численность генеральной совокупности и получаются соответствующие объемные показатели, характеризующие размеры явления по всей совокупности. Так, например, в туризме для определения объема поступлений от иностранного туризма средние расходы одного иностранного туриста (полученные по данным выборочного наблюдения) умножают на число иностранных туристов.

Способ поправочных коэффициентов используется при проведении контрольных выборочных наблюдений для проверки и уточнения данных сплошного наблюдения. Так, если выборочное наблюдение показало, что недоучет (неполный учет) величины явления, которое исследуется, составил 1 %, то данный поправочный коэффициент (1,01) распространяют на результат, полученный при сплошном наблюдении, путем увеличения его в 1,01 раза или на 1 %.

Когда мы рассматриваем способы распространения выборочных данных на генеральную совокупность, необходимо иметь в виду, что статистические сведения в этом случае получены расчетным путем. В этой связи нужно указывать пределы ошибок, которые допущены в процессе выборки, и вероятность этих пределов.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Что представляет собой выборочное наблюдение?
2. Перечислите преимущества выборки.
3. Что представляют собой генеральная и выборочная совокупность?
4. Поясните суть ошибок репрезентативности.
5. Какие существуют способы формирования выборочной совокупности при жеребьевке?
6. Перечислите виды выборки по способу ее организации.
7. Как осуществляется простая случайная выборка?
8. Что характеризует средняя ошибка выборки?
9. Как рассчитывается при случайном повторном отборе средняя ошибка выборочной средней величины?
10. Приведите формулу расчета стандартной ошибки выборочной средней при случайном бесповторном отборе.
11. При каком отборе, повторном или бесповторном, ошибка наблюдения больше?
12. Как определяется средняя ошибка выборочной доли при случайном повторном и бесповторном отборе?
13. Что представляет собой предельная ошибка выборки?
14. Приведите формулы расчета предельных ошибок средней и доли при простом случайном отборе.
15. Что представляет собой точечный и интервальный тип выборочных оценок?

16. Как определяются относительные стандартные ошибки средней и доли?
17. Поясните суть механической выборки.
18. Какие формулы расчета ошибок используют при механическом отборе?
19. При каком способе отбора, простом случайном или механическом, ошибка наблюдения больше?
20. Что представляет собой типическая выборка?
21. Как рассчитываются средние ошибки средней величины и доли при типическом отборе?
22. В чем суть серийной выборки и какие формулы расчета ошибок в ней используются?
23. Сформулируйте основные типы задач на выборку.
24. Поясните формулы определения численности выборки при различных способах отбора.
25. Что представляет собой малая выборка?
26. Назовите, в чем состоит особенность малой выборки.
27. Что представляет собой t-распределение Стьюдента?
28. Как рассчитывается отношение Стьюдента?
29. Запишите формулу расчета предельной ошибки малой выборки.
30. Как при применении малой выборке определяется вероятность того, что генеральная средняя находится в определенных границах?

РАЗДЕЛ 9. СТАТИСТИЧЕСКИЕ МЕТОДЫ ИЗМЕРЕНИЯ ВЗАИМОСВЯЗЕЙ

9.1. ВИДЫ ВЗАИМОСВЯЗЕЙ ЯВЛЕНИЙ И ЗАДАЧИ ИХ СТАТИСТИЧЕСКОГО ИЗУЧЕНИЯ

Все явления и процессы общественной жизни взаимосвязаны. Ни одно явление не может быть полностью познано, изучено, если его исследовать изолированно, отдельно от других процессов. Лишь при изучении причинно-следственных связей между явлениями можно дать им объективную оценку и прогнозировать их развитие.

Причинно-следственная связь – это связь явлений и процессов, когда изменение одного из них – причины – ведет к изменению другого – следствия. Признаки, которые характеризуют причины, называются **факторными** (независимыми) и обозначаются x, а те, что характеризуют следствие, называются **результативными** (зависимыми) и обозначаются y.

Различают **два основных вида связей между признаками**:

- функциональные;
- стохастические.

При **функциональной зависимости** каждому значению факторного признака строго соответствует одно определенное значение результативного. Такие взаимосвязи часто наблюдаются в математических, физических науках. Например, функциональной является связь между площадью треугольника, его основанием и высотой.

Между социально-экономическими явлениями функциональные связи встречаются реже. Это балансовая связь объемных показателей, зависимость средних величин от системы весов, взаимосвязь между индексируемыми показателями (см. Раздел 11). Часто одним из факторов при функциональной связи выступает средний уровень качественного показателя. Например, функциональной является связь между стоимостью изделий, их средней ценой и количеством; между валовым, общим сбором, средней урожайностью и посевной площадью.

Характерной особенностью функциональных связей является то, что нам известен полный перечень, список факторов, определяющих величину результативного признака, а также механизм этого влияния, выраженный определенным уравнением.

В большинстве случаев при изучении социально-экономических явлений мы встречаемся с тем, что связь между двумя или более признаками затушевывается (становится неясной, нечеткой), усложняется в результате действия других причин (факторов).

При **стохастической связи** при изменении факторного признака изменяются условные распределения результативного признака.

Частным случаем стохастической связи является **корреляционная связь**, при которой с изменением факторного признака изменяется средняя величина условных распределений результативного признака. Классическими примерами корреляционной связи являются связи между стажем работы и выработкой или стажем работы и заработной платой и т.д.

Особенностью корреляционной связи является то, что она проявляется не в отдельном случае, а при значительном числе единиц совокупности, которая обследуется. Поэтому корреляционные связи изучаются по так называемым **эмпирическим данным**. Эмпирические данные получают на основе статистического наблюдения массовых общественных явлений и процессов. Именно в таких данных отображается совокупное действие всех причин на явление, которое обследуется.

В зависимости от направления различают связи:

- прямые;
- обратные.

При **прямой связи** направления изменения факторного и результативного признаков совпадают. Например, прямая связь существует между стажем работы (факторный признак) и заработной платой (результативный признак), так как с увеличением стажа средняя заработная плата растет.

Если направления изменения факторного и результативного признаков не совпадают, говорят об **обратной связи** между ними. Примером обратной связи является связь между количеством часов пропущенных занятий (факторный признак) и оценкой на экзамене (результативный признак), так как с ростом пропусков средний балл падает.

В зависимости от числа признаков, между которыми измеряются связи, различают:

- однофакторную связь (парную корреляцию);
- множественную связь (множественную корреляцию).

При **однофакторной связи** исследуется зависимость результативного признака от какого-либо одного факторного. При этом влияние других факторов не учитывается, от него абстрагируются.

При **множественной связи** анализируется влияние на результативный признак нескольких факторных признаков.

По степени тесноты связи между признаками различают связь:

- слабую;
- умеренную;
- заметную;
- сильную.

Теснота связи выражается конкретными величинами в соответствии со смыслом статистического показателя.

Основной характеристикой корреляционной связи является линия регрессии.

Линия регрессии – это функция, которая связывает средние значения результативного признака со значениями факторного признака.

Способы задания линии регрессии:

- аналитический;
- табличный;
- графический.

При аналитическом способе задания линии регрессии различают **две аналитические формы корреляционных связей**:

- линейную;
- нелинейную (криволинейную).

При **линейной связи** зависимость между факторными и результативными признаками может быть выражена в виде уравнения прямой.

При **нелинейной связи** зависимость между факторными и результативными признаками может быть выражена в виде уравнения кривой (гиперболы, параболы, показательной функции и т.п.).

Задачами статистического изучения корреляционных связей являются:

- выявление наличия связи между признаками;
- определение характера связи, его направления и формы;
- установление аналитического выражения связи в виде математической функции;
- установление количественных оценок тесноты связи;
- проверка значимости статистических характеристик взаимосвязи.

Для решения этих задач рассмотрим **основные методы изучения корреляционных связей**:

- метод аналитических группировок и дисперсионный анализ;
- регрессионно-корреляционный анализ.

9.2. МЕТОД АНАЛИТИЧЕСКИХ ГРУППИРОВОК И ДИСПЕРСИОННЫЙ АНАЛИЗ

Задачи аналитических группировок были сформулированы в разделе 3 данного пособия, а в разделе 7 изложены основные понятия дисперсионного анализа.

В данном параграфе вернемся к обсуждению метода с точки зрения оценки им корреляционных связей.

Характеристикой корреляционной связи является **линия регрессии**. В модели аналитической группировки – это эмпирическая линия регрессии, которая образуется из групповых средних значений результативного признака $\bar{y}_j$ для каждого значения (интервала) x_j.

Напомним, что для построения аналитической группировки необходимо распределить единицы совокупности по факторному признаку *x*, а затем для каждой *j*-той группы рассчитать среднюю величину результативного признака $\bar{y}_j$.

На основе данных статистического наблюдения (таблица 3.7) в третьем разделе пособия нами была построена аналитическая группировка, характеризующая зависимость дневной выработки работников предприятия от стажа. Представим ее в таблице 9.1.

Таблица 9.1

Зависимость дневной выработки работников предприятия от стажа

Группы по стажу, лет x_j	Число работников, чел. f_j	Средняя выработка 1 работника, грн. $\bar{y}_j$	$\bar{y}_j - \bar{y}$	$(\bar{y}_j - \bar{y})^2 f_j$
0 – 3	4	140	-91	33124
3 – 6	10	162	-69	47610
6 – 9	18	218	-13	3042
9 – 12	12	284	53	33708
12 – 15	4	313	82	26896
15 – 18	2	395	164	53792
Всего (в среднем)	50	$\bar{y}$ =231	-	198172

Средние уровни дневной выработки представлены в третьей графе таблицы. Рост групповых средних от группы к группе свидетельствует о наличии корреляционной связи между стажем и выработкой.

На основе данных аналитической группировки определяют **эффект (силу) влияния факторного признака на результативный** (Δ_{yx}), который показывает, на сколько единиц в среднем изменяется $\bar{y}_j$ с изменением $\bar{x}_j$ на одну свою единицу:

$$\Delta_{yx}=\frac{\bar{y}_j-\bar{y}_{j-1}}{\bar{x}_j-\bar{x}_{j-1}} \tag{9.1}$$

В случае прямой связи $\Delta_{yx} > 0$, обратной - $\Delta_{yx} < 0$.

Так, при переходе от первой группы ко второй:

$$\Delta_{yx}=\frac{162-140}{4{,}5-1{,}5}=7{,}33\ \text{грн},$$

то есть, с ростом стажа на 1 год выработка увеличивается в среднем на 7 грн. 33 коп.

Силу связи можно определить для любых групп. В случае линейной связи между признаками определяется **средняя сила связи**:

$$\bar{\Delta}_{yx}=\frac{\bar{y}_n-\bar{y}_1}{\bar{x}_n-\bar{x}_1}, \tag{9.2}$$

где $\bar{y}_n$ и $\bar{y}_1$ – средние значения результативного признака в последней и первой группах,

$\bar{x}_n$ и $\bar{x}_1$ – средние значения факторного признака в последней и первой группах.

По данным аналитической группировки, приведенной в таблице 9.1, произведём расчет средней силы связи (связь между признаками линейная, так как с ростом стажа выработка практически равномерно растет):

$$\bar{\Delta}_{yx}=\frac{395-140}{16{,}5-1{,}5}=17\ \text{грн}.$$

Величина данного показателя свидетельствует о том, что в целом по совокупности работников с ростом стажа на один год их дневная выработка в среднем увеличивается на 17 грн.

Таким образом, **аналитическая группировка как метод выявления корреляционной связи дает возможность:**

- выявить наличие (либо отсутствие) связи между признаками;
- выявить направление связи;
- представить эмпирическую линию регрессии.

Для количественной оценки тесноты связи и проверки ее значимости используют **дисперсионный анализ**. Его основы были освещены в разделе 7.

Оценка тесноты связи основывается на правиле сложения дисперсий **(**$\sigma^2=\delta^2+\overline{\sigma^2}$).

В модели аналитической группировки мерой тесноты связи является отношение межгрупповой дисперсии к общей, которое называется **эмпирическим коэффициентом детерминации** (**η^2**) (формула 7.15). Данный показатель колеблется от 0 до 1 либо от 0 до 100 %. При отсутствии связи $\eta^2 = 0$, а при функциональной связи - $\eta^2 = 1$. Чем больше η^2 приближается к 1, тем связь теснее.

По данным таблицы 3.7 на основе упрощенной формулы расчета дисперсии (формула 6.20) была определена общая дисперсия выработки σ^2. Она составила 5465.

В таблице 9.1 представлены промежуточные расчеты для определения межгрупповой дисперсии. На основе формулы 7.11 рассчитаем ее для нашей совокупности:

$$\delta^2 = \frac{198172}{50} = 3963.$$

Эмпирический коэффициент детерминации составит:

$$\eta^2 = \frac{3963}{5465} = 0{,}726 \text{ или } 72{,}6\ \%.$$

Таким образом, вариация выработки на 72,6 % объясняется вариацией стажа рабочих и на 27,4 % - влиянием других факторов.

Вместе с тем, тесная связь, которую демонстрирует коэффициент детерминации, может возникнуть случайно. Необходимо проверить его значимость.

Проверка значимости связи – это сравнение фактического значения η^2с его критическим значением $\eta^2_{1-\alpha}(k_{1,}k_{2,})$ для определенного уровня значимости α и числа степеней свободы $k_1 = m - 1$ и $k_2 = n - m$, где m – число групп, n – объем совокупности. Если $\eta^2 > \eta^2_{1-\alpha}(k_{1,}k_{2,})$, то связь признается существенной.

Критические значения эмпирического коэффициента детерминации приведены в таблицах. Проведем анализ для $\alpha = 0{,}05$.

В нашем примере

$$k_1 = 6 - 1 = 5,$$

$$k_2 = 50 - 6 = 44.$$

$$\eta^2_{0,95}(5, 44) = 0{,}234.$$

Так как $\eta^2 = 0{,}726 > 0{,}234$, то связь признается значимой с вероятностью 0,95 (95 %).

Необходимо иметь в виду следующее. В программе Excel в процессе определения дисперсий используются не средние квадраты отклонений, а суммы квадратов отклонений, которые принято называть **девиатами** (SS). Тогда правило сложения дисперсий можно записать так:

$$\mathrm{SS}_y = \mathrm{SS}_Y + \mathrm{SS}_\varepsilon, \tag{9.3}$$

где $\mathrm{SS}_y,\ \mathrm{SS}_Y,\ \mathrm{SS}_\varepsilon$ – соответственно общая, межгрупповая и остаточная девиаты.

Итак, в программе Excel **общая девиата** определяется как сумма квадратов отклонений фактических значений результативного признака (y) от общей средней величины $(\bar{y})$:

$$\mathrm{SS}_y = \sum(y - \bar{y})^2. \tag{9.4}$$

Межгрупповая девиата определяется как сумма квадратов отклонений групповых средних $(\bar{y}_i)$ от общей средней величины результативного признака, взвешенных по соответствующим частотам:

$$\mathrm{SS}_Y = \sum(\bar{y}_i - \bar{y})^2 f_i\ . \tag{9.5}$$

Остаточная девиата определяется как разность между общей девиатой и межгрупповой ($SS_\varepsilon = \mathrm{SS}_y - \mathrm{SS}_Y$).

Тогда эмпирический коэффициент детерминации можно рассчитать соотношением соответствующих девиат:

$$\eta^2 = \frac{\mathrm{SS}_Y}{\mathrm{SS}_y}\ . \tag{9.6}$$

На основе встроенного блока «Однофакторный дисперсионный анализ» редактора Excel рассчитываются девиаты и осуществляется проверка значимости связи между факторным и результативным признаками (таблица 9.2)

Таблица 9.2

Результаты дисперсионного анализа выработки

Дисперсионный анализ						
Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	198162,222	5	39632,44	23,22385356	2,4295E-11	2,42704012
Внутри групп	75087,7778	44	1706,54			
Итого	273250	49				

Числа степеней свободы ($k_1 = 5$ и $k_2 = 44$) указываются в графе df.

Графа MS представляет собой скорректированные на число степеней свободы квадраты отклонений:

$$MS_Y = \frac{SS_Y}{k_1} = \frac{198162{,}222}{5} = 39632{,}44,$$

$$MS_\varepsilon = \frac{SS_\varepsilon}{k_2} = \frac{75087{,}7778}{44} = 1706{,}54.$$

Фактическое значение F-критерия рассчитывается:

$$F = \frac{MS_Y}{MS_\varepsilon} = \frac{39632{,}44}{1706{,}54} = 23{,}22385356.$$

Сравниваем фактическое значение F-критерия с критическим ($F_{\text{кр.}}$) (таблица 9.2). Убеждаемся, что $F > F_{\text{кр.}}$, то есть расхождения между групповыми средними не являются случайными. Это подтверждает, что связь между признаками существенна, значима.

9.3. РЕГРЕССИОННО-КОРРЕЛЯЦИОННЫЙ АНАЛИЗ (РКА)

Выше отмечалось, что характеристикой корреляционной связи является линия регрессии. В аналитической группировке – это эмпирическая линия регрессии, которая образуется из групповых средних значений результативного признака.

В модели регрессионного анализа характеристикой корреляционной связи является теоретическая линия регрессии.

Теоретическая линия регрессии описывается функцией $Y = f(x)$, которая называется **уравнением регрессии**.

На основе уравнения регрессии определяют средние значения результативного признака для каждого значения факторного признака x. Их обозначают Y и называют **теоретическими значениями** в отличие от фактических значений y, которые получают благодаря статистическому наблюдению и называют эмпирическими.

На первом этапе РКА решаются два вопроса:

1) выбор факторного и результативного признака;

2) выбор вида уравнения регрессии.

Решение первой проблемы основывается на теоретическом анализе сути изучаемого явления. В зависимости от задачи исследования необходимо выбрать показатель, характеризующий результативный признак. Например, при изучении туристической привлекательности региона в качестве результативного можно выбрать количество иностранных посетителей, число обслуженных туристов предприятиями гостиничного хозяйства, поступления в бюджет региона от туризма и т.п.

При выборе факторных признаков руководствуются следующим. Необходимо ограничиться изучением влияния лишь основных, существенных факторов. При этом они не должны дублировать, повторять друг друга. Например, нет смысла одновременно рассматривать число гостиниц и количество номеров в них в качестве факторов, влияющих на туристическую привлекательность. Обязательно должны быть исключены факторы, которые функционально связаны с результативным.

Важным требованием, обеспечивающим надежность РКА, является наличие достаточного объема совокупности, которая исследуется. Считается, что **число единиц совокупности должно быть в 5-6 раз больше числа факторов**.

Совокупность, которая исследуется, должна быть **качественно однородной**, что способствует точности полученных результатов анализа. Необходима также проверка **количественной однородности**, критерием которой служит коэффициент вариации, не превышающий 33 %.

Кроме того, необходимо учитывать, что факторные и результативный признаки должны подчиняться **нормальному распределению.**

Пример. Пусть целью использования РКА является исследование влияния математических способностей на успеваемость по статистике.

Теоретический анализ сути явления, которое изучается, показал, что в качестве результативного признака может быть балл на экзамене по статистике у группы студентов, а в качестве факторного целесообразно рассматривать экзаменационную оценку (балл) по математике у этих же студентов (таблица 9.3).

Таблица 9.3

Экзаменационные баллы студентов по математике и статистике

№ студента	1	2	3	4	5	6	7	8	9	10
Балл по математике x	50	85	54	68	70	95	53	55	73	51
Балл по статистике y	55	91	56	73	75	90	50	56	67	50

Современный подход к осуществлению всех этапов РКА с учетом использования редактора Excel системно описан профессором Янковым А.Г.

После выбора результативного и факторного признаков необходимо по результативному признаку проверить количественную однородность совокупности студентов. После того, как мы рассчитали с помощью редактора Excel среднее значение и среднее квадратическое отклонение баллов по статистике ($\bar{y} = 66{,}3$ балла, $\sigma_y = 14{,}7$ балла), определяем коэффициент вариации:

$$V_\sigma = \frac{\sigma_y}{\bar{y}} \cdot 100 = \frac{14{,}7}{66{,}3} \cdot 100 = 22{,}2\ \%.$$

Величина коэффициента вариации ($V_\sigma < 33\ \%$) свидетельствует о количественной однородности совокупности.

При выборе вида уравнения регрессии необходимо использовать:

- теоретическое обоснование;
- графический метод;
- аналитические группировки;
- способ перебора, пересмотра функций.

Теоретическое обоснование включает как логический анализ возможной взаимосвязи между признаками, так и изучение характера изменений результативного и факторного признаков. Если с изменением x признак y изменяется более-менее равномерно, выбирают линейную функцию, если же изменение взаимосвязанных признаков происходит неравномерно (с ускорением, замедлением или со сменой направления связи), применяют нелинейные функции.

Выбору типа уравнения регрессии помогает **графический метод.** Близость общего вида эмпирической линии регрессии к общему виду кривой соответствующей математической функции и определяет выбор того или иного типа уравнения регрессии.

Предварительное построение **аналитической группировки** дает возможность по эмпирической линии регрессии, которая образуется из групповых средних значений результативного признака, судить о типе уравнения регрессии.

Сегодня широко используется так называемый **способ перебора функций**. Он предполагает построение уравнений регрессии разных видов и из них на основе специальных математико-статистических критериев выбор лучшего. Выбранное уравнение регрессии характеризуется прежде всего наивысшим показателем тесноты связи между признаками, которые изучаются.

Итак, **в зависимости от характера связи используют:**

линейное уравнение:

$$Y = a_0 + a_1 x, \tag{9.7}$$

нелинейные уравнения:

- **параболы**:

$$Y = a_0 + a_1 x + a_2 x^2, \tag{9.8}$$

- **гиперболы**:

$$Y = a_0 + \frac{a_1}{x}, \tag{9.9}$$

- **показательной функции**:

$$Y = a_0 a_1{}^x, \tag{9.10}$$

- **степенной функции**:

$$Y = a_0 x^{a_1}. \tag{9.11}$$

Параметры для всех уравнений связи чаще всего определяют из так называемой системы нормальных уравнений, отвечающей требованию «метода наименьших квадратов» (МНК). Это требование можно записать как $\sum(y - Y)^2 \to min,$ то есть требуется определить, при каких значениях параметров a_0 и a_1 сумма квадратов отклонений фактических значений результативного признака от теоретических будет минимальной. Когда мы найдем частные производные указанной суммы по a_0 и a_1 и приравняем их к нулю, то можем легко записать систему уравнений, решение которой и дает параметры уравнения регрессии.

Система нормальных уравнений при линейной зависимости имеет вид:

$$\begin{cases} \sum y = na_0 + a_1 \sum x \\ \sum xy = a_0 \sum x + a_1 \sum x^2 \end{cases}. \tag{9.12}$$

В социально-экономических исследованиях **наибольшее распространение получила именно линейная функция**, так как ее параметры можно четко интерпретировать. **Параметр a_1 (коэффициент регрессии)** – величина именованная, имеет размерность результативного признака и рассматривается как эффект влияния x на y, то есть параметр показывает, как в среднем изменяется результативный признак при изменении факторного на единицу. **Параметр a_0** – свободный член уравнения регрессии. Он характеризует среднее значение y при $x = 0$.

В случае если факторный признак не может быть равен нулю, экономическую интерпретацию параметру не дают.

Предварительный анализ, в том числе графический, однозначно свидетельствует о наличии линейной связи между успеваемостью по статистике и баллами по математике.

Параметры уравнения регрессии можно определить на основе системы нормальных уравнений.

Современный подход к изучению корреляционных связей позволяет получить значение необходимых коэффициентов на основе встроенного блока РКА редактора Excel без проведения трудоемких расчетов (таблица 9.4).

Таблица 9.4

Результаты РКА успеваемости по статистике

ВЫВОД ИТОГОВ						
Регрессионная статистика						
Множественный R	0,958848945					
R-квадрат	0,919391299					
Нормированный R-ква	0,909315212					
Стандартная ошибка	4,66965322					
Наблюдения	10					
Дисперсионный анализ						
	df	*SS*	*MS*	*F*	*Значимость F*	
Регрессия	1	1989,65471	1989,65471	91,2448695	1,19369E-05	
Остаток	8	174,4452896	21,8056612			
Итого	9	2164,1				
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	3,854747067	6,701956801	0,5751674	0,580989875	-11,59999303	19,30948716
Переменная X 1	0,954820381	0,099957976	9,552218041	1,19369E-05	0,724316875	1,185323887

Значения коэффициентов регрессии расположены в нижней левой части таблицы 9.3 в столбике *Коэффициенты.* В ряду *У-пересечение* находится значение a_0, а в ряду *Переменная X1* - a_1. Таким образом, получаем следующую регрессионную модель:

$$Y = 3{,}854747067 + 0{,}954820381x,$$

или, округлив значения параметров:

$$Y = 3{,}85 + 0{,}95x$$

Знак коэффициента a_1 соответствует предполагаемому направлению связи между признаками.

Несданный экзамен по математике, оцененный в 0 баллов ($x = 0$), теоретически может соответствовать $Y = 3{,}85$ баллам по статистике (то есть также несданному экзамену).

Далее следует проверка значимости отдельных коэффициентов парной регрессии. Для этого сначала рассчитывают стандартные ошибки коэффициентов регрессии a_0 и a_1:

$$\mu_{a_0} = \frac{\sigma_\varepsilon}{\sqrt{n-2}}, \qquad (9.15)$$

$$\mu_{a_1} = \frac{\sigma_\varepsilon}{\sigma_x\sqrt{n-2}}, \qquad (9.16)$$

где ${\sigma^2}_\varepsilon$- остаточная дисперсия.

По результатам РКА (таблица 9.3) стандартные ошибки для нашего примера составили $\mu_{a_0} = 6{,}701956801$ и $\mu_{a_1} = 0{,}099957976$.

Если справедлива нулевая гипотеза H_0: $a_j = 0$ против альтернативной H_a: $a_j \neq 0$ ($j = 0$, 1) отношение соответствующего коэффициента регрессии к его стандартной ошибке

$$t_j = \frac{a_j}{\mu_j} \qquad (9.17)$$

подчиняется t-распределению Стьюдента с уровнем значимости α и k = n – 2 степенями свободы. В таблице 9.3 приведены необходимые значения: $t_0 = 0{,}5751674$; $t_1 = 9{,}552218041$; $p_0 = 0{,}580989875$; $p_1 = 1{,}19369E-05$.

Мы видим, что $p_1 = 1{,}19369E-05 < 0{,}05$, то есть нулевая гипотеза H_0: $a_1 = 0$ отклоняется. Можно утверждать, что коэффициент a_1 существенно отличается от нуля. Данный коэффициент является статистически значимым и надежным.

Что касается a_0, то $p_0 = 0{,}580989875 > 0{,}05$. Таким образом, нулевая гипотеза H_0: $a_0 = 0$ не отклоняется. Это означает, что необходимо использовать модель $Y = a_1x$, график которой проходит через начало координат.

Итак, уравнение регрессии необходимо пересчитать без свободного члена, который не является статистически значимым.

Для пересчета уравнения регрессии в диалоговом окне блока «Регрессия» необходимо активизировать флажок *Константа-ноль*. В случае если незначимым является коэффициент при факторном признаке, следует пересмотреть набор признаков в уравнении регрессии.

В результате пересчета уравнения регрессии получены новые результаты РКА, отраженные в таблице 9.5.

Таблица 9.5

Результаты РКА успеваемости по статистике (П вариант)

Регрессионная статистика						
Множественный R	0,998028683					
R-квадрат	0,996061252					
Нормированный R-кв	0,884950141					
Стандартная ошибка	4,49269759					
Наблюдения	10					
Дисперсионный анализ						
	df	*SS*	*MS*	*F*	*Значимость F*	
Регрессия	1	45939,34102	45939,34102	2275,990201	4,12145E-11	
Остаток	9	181,6589847	20,18433164			
Итого	10	46121				
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Переменная X 1	1,010900031	0,021189613	47,70733907	3,91374E-12	0,962965797	1,058834265

Коэффициент a_1 =1,01. Он является значимым, так как $p_1 = 3{,}91374E - 05 < 0{,}05$. Уравнение регрессии будет иметь вид:

$$Y = 1{,}01 \cdot x.$$

Можно сформулировать вывод, что с увеличением на 10 баллов оценки по математике, успеваемость по статистике улучшается в среднем на 10,1 балла.

Характеристикой относительного изменения результативного признака за счет факторного является коэффициент эластичности (Э). Он показывает, на сколько процентов в среднем изменяется результативный признак со сменой факторного на 1 %. Рассчитывается данный показатель по формуле:

$$Э = a_1 \frac{\bar{x}}{\bar{y}}. \qquad (9.18)$$

Для нашего примера $Э = 0{,}996$. Таким образом, с ростом балла по математике на 1 % балл по статистике увеличивается на 0,996 %.

Следующей задачей изучения корреляционной зависимости является **измерение тесноты связи**.

Теснота линейной связи измеряется с помощью **линейного коэффициента корреляции** r:

$$r = a_1 \frac{\sigma_x}{\sigma_y}, \qquad (9.19)$$

где σ_x и σ_y- соответственно среднее квадратическое отклонение в ряду x и в ряду y.

Линейный коэффициент может быть рассчитан и по другим формулам, тождественным предыдущей формуле, например:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}, \qquad (9.20)$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \qquad (9.21)$$

Линейный коэффициент корреляции может принимать по модулю значения от 0 до 1 (знак «+» при прямой зависимости, и «–» - при обратной).

Для качественной характеристики тесноты связи можно использовать так называемую шкалу Чеддока:

Оценка тесноты связи	Слабая	Умеренная	Заметная (средняя)	Значительная (сильная)	Очень сильная
Значение r	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99

Мерой тесноты как линейной, так и нелинейной связи является **коэффициент детерминации R^2**. Методика его расчета основывается на правиле разложения дисперсии.

Вариация результативного признака оценивается прежде всего с помощью общей дисперсии. **Общая дисперсия** представляет собой средний квадрат отклонений фактических значений результативного признака от общей средней:

$$\sigma^2 = \frac{\sum (y - \bar{y})^2}{n}. \qquad (9.22)$$

Вариация результативного признака, связанная с изменением исследуемого фактора, измеряется факторной дисперсией. **Факторная дисперсия** ($\sigma_Y{}^2$) представляет собой средний квадрат отклонений теоретических значений результативного признака от общей средней:

$$\sigma_Y{}^2 = \frac{\sum (Y - \bar{y})^2}{n}. \qquad (9.23)$$

Вариация, связанная с влиянием всех других факторов, кроме исследуемого нами x, измеряется остаточной дисперсией. **Остаточная дисперсия** представляет собой средний квадрат отклонений фактических значений результативного признака от теоретических:

$$\sigma_\varepsilon{}^2 = \frac{\sum(y-Y)^2}{n}. \quad (9.24)$$

Сумма факторной и остаточной дисперсии равна общей дисперсии.

В программе Excel при осуществлении РКА, как и при дисперсионном анализе, вместо дисперсии определяются **девиаты**. Подход к их расчету аналогичный.

Общая девиата (SS) определяется:

$$\mathrm{SS} = \sum(y-\bar{y})^2. \quad (9.25)$$

Остаточная девиата рассчитывается:

$$\mathrm{SS}_\varepsilon = \sum(y-Y)^2. \quad (9.26)$$

Факторная девиата определяется как разность между общей девиатой и остаточной:

$$\mathrm{SS}_Y = \mathrm{SS} - \mathrm{SS}_\varepsilon. \quad (9.27)$$

Тогда коэффициент детерминации R^2 можно рассчитать соотношением факторной и общей дисперсий (девиат) с помощью следующих формул:

$$R^2 = \frac{\sigma_Y{}^2}{\sigma^2}, \quad (9.28)$$

$$R^2 = \frac{\mathrm{SS}_Y}{\mathrm{SS}}, \quad (9.29)$$

$$R^2 = 1 - \frac{\sigma_\varepsilon{}^2}{\sigma^2}, \quad (9.30)$$

$$R^2 = 1 - \frac{\mathrm{SS}_\varepsilon}{\mathrm{SS}}. \quad (9.31)$$

Коэффициент детерминации показывает, какой удельный вес в общей дисперсии результативного признака занимает дисперсия, вызванная вариацией фактора x.

Если $\sigma_Y{}^2 = \sigma^2$ и соответственно $R^2 = 1$, это означает полную зависимость y от x. Если $\sigma_Y{}^2 = 0$ и соответственно $R^2 = 0$, это означает, что вариация x никак не влияет на вариацию y.

Корень квадратный из коэффициента детерминации называется **индексом корреляции** (коэффициентом корреляции) R.

При линейной форме связи абсолютная величина линейного коэффициента корреляции равна индексу корреляции, то есть $R = |r|$.

Итак, показателем тесноты корреляционной связи для любой модели служит коэффициент множественной корреляции R . Для парной линейной модели $R = |r|$.

В таблице 9.4 R – первый показатель в разделе *Регрессионная статистика*. Его значение, а значит и значение коэффициента парной корреляции, – 0,998028683 свидетельствует об очень тесной корреляционной связи между баллами по математике и статистике.

Абсолютной мерой точности полученной модели является **средняя квадратическая (стандартная) ошибка регрессии.** Для нашей модели она рассчитывается автоматически, и в таблице 9.5 ее величина отражена в строке *Стандартная ошибка* (4,49269759). Проф. А.Г.Янковой дает следующую рекомендацию по толкованию, объяснению данной величины: для одних и тех же исходных данных меньшая стандартная ошибка отвечает более точной модели.

Относительной характеристикой точности регрессионной модели является коэффициент детерминации R^2. Для парной линейной модели $R^2 = r^2$. Он показывает долю вариации результативного признака, которая объясняется вариацией факторного. В таблице 9.4 $R^2 = 0{,}996061252.$

В разделе *Регрессионная статистика* отражена также величина *Нормированный коэффициент детерминации* ($R_н{}^2$). Он рассчитывается для малых выборок, и его значение всегда ниже R^2. Данный показатель учитывает соотношение числа наблюдений и число коэффициентов уравнения регрессии. Для нашего примера $R_н{}^2 = 0{,}884950141,$ то есть даже с учетом особенностей малой выборки модель является точной. 88,5 % вариации баллов по статистике можно объяснить изменением баллов по математике.

Проверка значимости связи в РКА осуществляется с помощью тех же самых критериев, что и в дисперсионном. Для определения числа степеней свободы сохраняются аналогичные формулы: $k_1 = m - 1$ и $k_2 = n - m$, только в данном случае m - число параметров в уравнении регрессии, а не число групп.

Фактическое значение F-критерия рассчитывается:

$$F = \frac{MS_Y}{MS_\varepsilon} = \frac{45939{,}34102}{20{,}18433164} = 2275{,}990201.$$

Современный подход к использованию данного критерия предполагает использование встроенного блока КРА (результаты в таблице 9.4). В столбике *Значимость F* приводятся расчетные значения F–критерия и его p-значимость (4,12145E-11). В соответствии со схемой проверки гипотез последняя величина сравнивается с уровнем значимости $\alpha = 0{,}05$. Так как 4,12145E-11 $< 0{,}05$, нулевая гипотеза H_0: $r^2 = 0$ отклоняется. Можно констатировать, что модель статистически значима. Этот вывод совпадает с выводом значимости коэффициента парной корреляции.

Далее рассчитываются верхние и нижние доверительные интервалы коэффициентов регрессии (в нашем примере только коэффициента a_1):

$$a_1 \pm t_{\alpha,k} \cdot \mu_{a_1}. \quad (9.18)$$

В таблице 9.4 в столбиках *Нижние 95 %* , *Верхние 95 %* приводятся значения 95 %-х доверительных интервалов коэффициентов регрессии:

$$0{,}962965797 \leq a_1 \leq 1{,}058834265.$$

Таким образом, с вероятностью 95 % можно утверждать, что истинные значения коэффициента регрессии a_1 находятся в указанных выше интервалах.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Что представляет собой причинно-следственная связь между явлениями?
2. Поясните особенности корреляционной связи.

3. Какие существуют виды связи в зависимости от ее направления?
4. Какие различают виды связей в зависимости от числа признаков?
5. Что представляет собой линия регрессии?
6. Какие существуют способы задания линии регрессии?
7. Какие существуют виды нелинейной связи?
8. Сформулируйте задачи статистического изучения корреляционных связей.
9. Что является линией регрессии в модели аналитической группировки?
10. Как определяют эффект (силу) влияния факторного признака на результативный?
11. Как рассчитывается и что характеризует эмпирический коэффициент детерминации?
12. Как осуществляется проверка значимости связи при дисперсионном анализе?
13. Какая линия является характеристикой корреляционной связи в модели регрессионного анализа?
14. Как решаются проблемы выбора результативного и факторных признаков?
15. Каков критерий достаточного объема исследуемой совокупности?
16. Величина какого показателя свидетельствует о количественной однородности совокупности?
17. Как рассчитывают параметры уравнений регрессии?
18. Что характеризуют параметры линейной функции?
19. Как измеряется теснота связи при изучении корреляционной зависимости?
20. Какие значения может принимать линейный коэффициент корреляции?
21. В чем заключается анализ точности парной линейной модели? Что является абсолютной мерой точности?
22. Какой показатель является относительной характеристикой точности регрессионной модели?

РАЗДЕЛ 10. АНАЛИЗ ИНТЕНСИВНОСТИ ДИНАМИКИ И ТЕНДЕНЦИЙ РАЗВИТИЯ

10.1 ПОНЯТИЯ, ЭЛЕМЕНТЫ И ВИДЫ РЯДОВ ДИНАМИКИ

Под динамикой в экономике, статистике понимают изменение явления во времени. Изучают динамику на основе построения рядов динамики (динамических рядов, временных рядов).

Ряд динамики – это ряд чисел, расположенных в хронологической последовательности и характеризующий изменение явления во времени.

Элементами ряда динамики являются:

- периоды либо моменты времени (t_i);
- уровни ряда динамики, относящиеся к тому либо иному периоду (моменту) времени (y_i).

В зависимости от сути социально-экономического явления, которое исследуется, и от того, что описывают уровни (моменты или периоды времени), различают следующие **виды рядов динамики**:

- моментные;
- периодические (интервальные),

Уровни моментных рядов динамики характеризуют состояния явления на определенные моменты времени. В качестве примера моментных рядов динамики можно привести ряды динамики запаса топлива в регионе на начало квартала, численности населения области на конец года, запаса продукции предприятия на первое число каждого месяца, основных средств на конец года (таблица 10.1) и т.п.

Уровни интервальных рядов динамики характеризуют результаты развития явления за определенные периоды, интервалы времени. Ряды динамики выпуска продукции, числа родившихся, товарооборота за определенные периоды времени, валового внутреннего продукта (таблица 10.1) являются примерами интервальных динамических рядов.

В зависимости от формы выражения уровней ряда динамики различают такие **виды рядов динамики**:

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

В таблице 10.1 приведены интервальный ряд динамики абсолютных величин (динамика ВВП), интервальный ряд динамики средних величин (динамика среднемесячной зарплаты), моментный ряд динамики абсолютных величин (динамика основных средств на конец года), моментный ряд динамики относительных величин (динамика уровня безработицы на конец года).

Таблица 10.1

Динамика основных социально-экономических показателей развития Украины

Год	ВВП, млрд. грн.	Основные средства на конец года, трлн. грн.	Уровень безработицы на конец года, %	Среднемесячная заработная плата, грн.
2005	441,5	1,3	3,1	806
2006	544,2	1,6	2,7	1041
2007	720,7	2,0	2,3	1351
2008	948,1	3,2	3,0	1806
2009	914,7	3,9	1,9	1906

Остановимся на некоторых **особенностях интервальных и моментных рядов динамики**.

Уровни интервальных рядов абсолютных величин можно суммировать и делить. Например, интервальный ряд динамики квартальных уровней можно путем суммирования преобразовать в интервальный ряд динамики годовых уровней.

В то же время уровни интервальных рядов динамики относительных и средних величин суммировать не имеет смысла, так как уровни таких рядов являются производными и определяются делением абсолютных величин.

Следующей особенностью интервальных рядов является тот факт, что **величина уровней в них зависит от длины периода**. Например, число детей, родившихся за год больше, чем за месяц, квартал и т.д.

Говоря об **особенностях моментных рядов динамики**, заметим следующее.

Во-первых, **уровни моментного ряда нельзя суммировать**, так как одни и те же единицы исследуемого явления входят в состав нескольких его уровней. Например, часть лиц, принимавших участие в переписи населения, обязательно будут принимать участие в последующей переписи.

Во-вторых, **величина уровней моментного ряда не зависит от длины периода между отдельными датами**. Характеризует ли ряд динамики численность населения на конец каждого года или на конец каждого месяца, это не влияет на размер показателя численности населения.

Важной проблемой построения ряда динамики является сопоставимость уровней в нем. Уровни должны быть сопоставимы:

- по единицам измерения;
- по времени, к которому они относятся;
- по территории;
- по охвату объектов наблюдения;
- по использованию единой методологии их расчета.

10.2. СРЕДНИЙ УРОВЕНЬ РЯДА ДИНАМИКИ

Для обобщающей характеристики ряда динамики рассчитывают средние уровни динамических рядов. Подход к расчету среднего уровня зависит от вида ряда динамики.

Средний уровень интервального ряда динамики рассчитывается по формуле средней арифметической простой:

$$\bar{y} = \frac{\sum y_i}{n}, \qquad (10.1)$$

где y_i– отдельные уровни ряда,

n – число уровней ряда (продолжительность периода).

Пусть требуется по данным таблицы 10.1 рассчитать среднегодовой уровень ВВП за 2007-2009 годы.

На основе формулы 10.1 рассчитаем:

$$\bar{y}_{07-09}^{\text{год.}} = \frac{720{,}7+948{,}1+914{,}7}{3} = 871{,}2 \text{ млрд. грн.}$$

Таким образом, в период 2007-2009 гг. среднегодовой уровень ВВП Украины составлял 871,2 млрд. грн.

Формулы расчета **среднего уровня моментного ряда динамики** зависят от характера исходной информации.

1. При наличии **полной, исчерпывающей информации** обо всех изменениях уровня внутри периода применяют формулу средней арифметической взвешенной:

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i}, \qquad (10.2)$$

где t_i – продолжительность времени, на протяжении которого уровень остается неизменным.

Например, известно, что 1 марта текущего года на расчетном счете предприятия было 230 тыс. грн. 10 марта на счет поступило еще 20 тыс. грн., 17 марта было списано 80 тыс. грн., 23 марта списано еще 50 тыс. грн., 28 марта поступило на счет 60 тыс. грн. Других изменений в марте не было. Считая, что изменения происходили с утра, определим средний остаток средств на текущем счете предприятия.

Итак, известны данные на начало марта и все изменения внутри месяца. Это говорит о том, что необходимо для расчета среднего уровня использовать формулу 10.2. Запишем уровни ряда в таблице 10.2, там же произведем и необходимые расчеты.

Таблица 10.2

К расчету среднего уровня моментного ряда динамики

Дата	y_i, тыс. грн.	Период	t_i, дни	$y_i t_i$
1.03	230	1.03-9.03	9	2070
10.03	250	10.03-16.03	7	1750
17.03	170	17.03-22.03	6	1020
23.03	120	23.03-27.03	5	600
28.03	180	28.03-31.03	4	720
Итого	-	-	31	6160

$$\bar{y} = \frac{6160}{31} = 198{,}7 \text{ тис. грн.}$$

Таким образом, в марте средние остатки средств на расчетном счете предприятия составляли 198,7 тис. грн.

2. Если **информация** об изменениях уровня моментного ряда динамики **не полная**, рассматривают несколько возможных ситуаций.

2.1. Известны данные на начало и на конец периода. В таком случае для расчета среднего уровня моментного ряда динамики используют формулу средней арифметической простой:

$$\bar{y} = \frac{y_н + y_к}{2}, \quad (10.3)$$

где $y_н$ и $y_к$ – уровни соответственно на начало и на конец периода.

Определим по данным таблицы 10.1 среднюю стоимость основных средств в Украине за 2009 год. Так как известны данные на начало 2009 г. и данные на конец 2009 г. (для моментных показателей уровни на конец предыдущего и начало последующего периодов совпадают), воспользуемся формулой 10.3 и рассчитаем:

$$\bar{y} = \frac{3,2+3,9}{2} = 3,55 \text{ трлн. грн.}$$

Таким образом, в 2009 году в Украине средняя стоимость основных средств составляла 3,55 трлн. грн.

2.2. Известны данные на начало и на конец периода, а также некоторые промежуточные данные, причем интервалы времени между известными уровнями не равны. В данной ситуации используется формула средней арифметической взвешенной модифицированной:

$$\bar{y} = \frac{\sum \bar{y}_i t_i}{\sum t_i}, \tag{10.4}$$

где $\bar{y}_i$- средний уровень в каждом промежутке времени t_i.

Предположим, что известны запасы топлива на предприятии на 1.01 – 300 т, на 1.03 – 320 т и на 1.07 – 270 т. Необходимо определить средние запасы топлива за 1-е полугодие.

Итак, известны данные на начало 1-го полугодия – 300 т, на конец – 270 т, и есть промежуточное значение запасов на 1.03. Между датами промежутки не равны. Используем для расчета среднего уровня формулу 10.4:

$$\bar{y} = \frac{\frac{(300+320)}{2}\cdot 2 + \frac{(320+270)}{2}\cdot 4}{6} = 300 \text{ т.}$$

Можно сформулировать вывод, что в 1-м полугодии средние запасы топлива на предприятии составляли 300 т.

2.3. Известны данные на начало и на конец периода, а также некоторые промежуточные данные, причем интервалы времени между известными уровнями равны. В данной ситуации используется формула средней хронологической простой (строго говоря, она представляет собой частный случай формулы 10.4):

$$\bar{y} = \frac{\frac{1}{2}\cdot y_1 + y_2 + \ldots + y_{n-1} + \frac{1}{2}\cdot y_n}{n-1}, \tag{10.5}$$

где n – число уровней,

$n - 1$ – число промежутков.

Определим по данным таблицы 10.1 среднюю стоимость основных средств в Украине за 2006-2009 гг.

Итак, известны данные на начало периода – 1,3 трлн. грн., на конец – 3,9 трлн. грн., известны также промежуточные данные на конец 2006, 2007 и 2008 гг. Данные условия соответствуют ситуации, которую мы рассматриваем. Производим расчет средней по формуле 10.5:

$$\bar{y} = \frac{\frac{1}{2}\cdot 1{,}3 + 1{,}6 + 2{,}0 + 3{,}2 + \frac{1}{2}\cdot 3{,}9}{5-1} = 2{,}35 \text{ трлн. грн.}$$

Таким образом, в Украине за период 2006-2009 гг. средняя стоимость основных средств составляла 2,35 трлн. грн.

10.3. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ХАРАКТЕРИСТИКИ ИНТЕНСИВНОСТИ ДИНАМИКИ

Расчет показателей динамики базируется на сравнении уровней ряда. При этом база сравнения может быть постоянной либо переменной.

При **постоянной базе** каждый уровень сравнивается с начальным или с тем, который считается исходным в развитии явления, базисным уровнем. При таком сравнении показатели динамики называются **базисными**.

Если же сравнение осуществляется с предыдущим уровнем (**переменная база сравнения**), такие показатели называются **цепными**.

К основным показателям анализа динамики относят:

- абсолютный прирост;
- темп роста;
- темп прироста;
- абсолютное значение 1 % прироста.

Абсолютный прирост (**снижение**) (Δ_t) характеризует абсолютную скорость роста (снижения) и показывает, на сколько единиц увеличился (уменьшился) уровень за период времени:

базисный – $\Delta_t = y_i - y_{i-t}$, (10.6)

$$\text{цепной} \quad - \quad \Delta_{\substack{t \\ t=1}} = y_i - y_{i-1,} \qquad (10.7)$$

где y_i — сравниваемый уровень,

y_{i-t} — базисный уровень,

y_{i-1} – предыдущий уровень,

t – длина периода.

Если абсолютная скорость роста положительна, имеем абсолютный прирост, если отрицательна – абсолютное снижение.

По данным таблицы 10.3 определим цепные абсолютные приросты:

$$\Delta_{08} = y_{08} - y_{07} = 61{,}4 - 55{,}7 = 5{,}7 \text{ млн. грн.}$$

$$\Delta_{09} = y_{09} - y_{08} = 62{,}1 - 61{,}4 = 0{,}7 \text{ млн. грн.}$$

$$\Delta_{10} = y_{10} - y_{09} = 55{,}3 - 62{,}1 = -6{,}8 \text{ млн. грн.}$$

Таким образом, в 2008 году по сравнению с 2007 годом выпуск продукции увеличился на 5,7 млн. грн., в 2009 году по сравнению с 2008 годом – на 0,7 млн. грн., а в 2010 году по сравнению с предыдущим годом выпуск уменьшился на 6,8 млн. грн.

Таблица 10.3

Динамика выпуска продукции

Год	Выпуск, млн.грн. y_i	Абсолютный прирост, млн. грн. Δ_t		Темп роста T_{p_t}		Темп прироста, % $T_{\text{пр}_t}$		Абсолютное содержание 1% прироста, млн. грн. A_t
		цепной	базисный	цепной	базисный	цепной	базисный	
2007	55,7	…	0,0	…	1,000	…	0,0	…
2008	61,4	5,7	5,7	1,102	1,102	10,2	10,2	0,557
2009	62,1	0,7	6,4	1,011	1,115	1,1	11,5	0,614
2010	55,3	-6,8	-0,4	0,890	0,993	-11,0	-0,7	0,621

Определим теперь базисные абсолютные приросты, приняв за базисный уровень выпуск начального, 2007 года. Для 2008 года цепной и базисный приросты совпадут, а для следующих периодов запишем формулы и расчеты:

$$\Delta_{\substack{08-09\\t=2}} = y_{09} - y_{07} = 62{,}1 - 55{,}7 = 6{,}4 \text{ млн. грн.}$$

$$\Delta_{\substack{08-10\\t=3}} = y_{10} - y_{07} = 55{,}3 - 55{,}7 = -0{,}4 \text{ млн. грн.}$$

То есть, за период 2008-2009 гг. или в 2009 г. по сравнению с 2007 г. выпуск продукции увеличился на 6,4 млн. грн.

За период 2008-2010 гг. или в 2010 г. по сравнению с 2007 г. выпуск снизился на 0,4 млн. грн.

Между цепными и базисными абсолютными приростами существует следующая взаимосвязь: сумма последовательных цепных абсолютных приростов равна приросту за весь период, то есть соответствующему базисному абсолютному приросту. Для нашего примера:

$$\Delta_{\substack{08\\t=1}} + \Delta_{\substack{09\\t=1}} = y_{08} - y_{07} + y_{09} - y_{08} = y_{09} - y_{07} = \Delta_{\substack{08-09\\t=2}}.$$

То есть базисные абсолютные приросты можно рассчитать следующим образом:

$$\Delta_{\substack{08-09\\t=2}} = 5{,}7 + 0{,}7 = 6{,}4 \text{ млн. грн.}$$

$$\Delta_{\substack{08-10\\t=3}} = 5{,}7 + 0{,}7 - 6{,}8 = -0{,}4 \text{ млн. грн.}$$

Темп роста (T_{p_t}) характеризует относительную скорость роста и показывает, во сколько раз увеличился уровень за период времени или какую часть базисного уровня составляет:

$$\text{базисный} \quad - \quad T_{p_t} = \frac{y_i}{y_{i-t}}, \qquad (10.8)$$

$$\text{цепной} \quad - \quad T_{\substack{p_t\\t=1}} = \frac{y_i}{y_{i-1,}} \qquad (10.9)$$

Темп роста является относительной величиной динамики и, как большинство относительных величин, имеет форму коэффициента и процента.

В соответствующих графах таблицы 10.3 рассчитаны цепные и базисные темпы роста. Приведем формулы и расчеты цепного и базисного показателей для строки 2010 г.:

$$T_{p_{10}} = \frac{y_{10}}{y_{09}} = \frac{55{,}3}{62{,}1} = 0{,}890 \text{ или } 89{,}0\ \%,$$

$$Tp_{\substack{08-10\\t=3}} = \frac{y_{10}}{y_{07}} = \frac{55{,}3}{55{,}7} = 0{,}993 \text{ или } 99{,}3\ \%.$$

Так как темпы роста за указанные периода оказались меньше единицы, выводы не формулируются.

Между цепными и базисными темпами роста существует следующая взаимосвязь: произведение последовательных цепных темпов роста равно темпу роста за весь период, то есть соответствующему базисному темпу роста. Для нашего примера:

$$T_{p\,\substack{08\\t=1}} \cdot T_{p\,\substack{09\\t=1}} \cdot T_{p\,\substack{10\\t=1}} = \frac{y_{08}}{y_{07}} \cdot \frac{y_{09}}{y_{08}} \cdot \frac{y_{10}}{y_{09}} = \frac{y_{10}}{y_{07}} = Tp_{\substack{08-10\\t=3}}$$

Таким образом, базисные темпы роста можно рассчитать следующим образом:

$$Tp_{\substack{08-10\\t=3}} = 1{,}102 \cdot 1{,}011 \cdot 0{,}890 = 0{,}993.$$

Темп прироста ($T_{\text{пр}_t}$) характеризует относительную скорость прироста:

$$\text{базисный - } T_{\text{пр}_t} = \frac{\Delta_t}{y_{i-t}} = \frac{y_i - y_{i-t}}{y_{i-t}} = T_{p_t} - 1, \quad (10.10)$$

$$\text{цепной - } T_{\text{пр}\,\substack{t\\t=1}} = \frac{\Delta_{\substack{t\\t=1}}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} = T_{p\,\substack{t\\t=1}} - 1. \quad (10.11)$$

Выраженный в %, темп прироста характеризует, на сколько % увеличился или уменьшился уровень за период времени:

$$\text{базисный - } T_{\text{пр}_t}(\%) = T_{p_t}(\%) - 100, \quad (10.12)$$

$$\text{цепной - } T_{\text{пр}\,\substack{t\\t=1}}(\%) = T_{p\,\substack{t\\t=1}}(\%) - 100. \quad (10.13)$$

Если относительная скорость прироста положительная, имеем темп прироста, если отрицательная, имеем темп снижения.

В таблице 10.3 рассчитаны цепные и базисные темпы прироста. Приведем формулы и расчеты цепного и базисного показателей для строки 2010 г.:

$$T_{\text{пр}_{10}}(\%) = T_{p_{10}}(\%) - 100 = 0{,}890 \cdot 100 - 100 = -11\ \%,$$

$$T_{\text{пр}_{08-10}}(\%) = T_{p_{08-10}}(\%) - 100 = 0{,}993 \cdot 100 - 100 = -0{,}7\ \%.$$

То есть в 2010 г. по сравнению с 2009 г. выпуск продукции снизился на 11 %. За период 2008-2010 гг. или в 2010 г. по сравнению с 2007 г. выпуск продукции снизился на 0,7 %.

Абсолютное значение 1 % прироста (A_t) показывает, сколько единиц содержит в себе каждый % прироста:

$$A_t = \frac{\Delta_t}{T_{\text{пр}_t}(\%)} = \Delta_t \div \left(\frac{\Delta_t}{y_{i-t}} \cdot 100\right) = \frac{y_{i-t}}{100}. \qquad (10.14)$$

Результаты расчета данного показателя приведены в таблице 10.3 (на цепной основе). Продемонстрируем формулу и расчет A_t для 2010 г.:

$$A_{10} = \frac{y_{09}}{100} = \frac{62{,}1}{100} = 0{,}621 \text{ млн. грн.}$$

Таким образом, в 2010 г. по сравнению с 2009 г. каждому % снижения выпуска продукции соответствовала 621 тыс. грн.

Приведенные выше формулы расчета основных показателей динамики применяют для интервального динамического ряда. Заметим, что сравниваемым уровнем для интервального ряда является последний уровень периода, а базисным – уровень, предшествующий периоду, за который делается расчет.

В моментном ряду динамики в качестве сравниваемого уровня берется уровень на конец периода, а в качестве базисного – уровень на начало периода.

Запишем формулы и определим основные показатели динамики для моментного ряда – ряда динамики стоимости основных средств в Украине (по данным таблицы 10.1).

Рассчитаем абсолютные приросты, темпы роста и прироста: 1) для 2010 года; 2) для периода 2008-2010 гг.

1) $\Delta_{10} = y_{\text{к.}10} - y_{\text{к.}09} = 3{,}9 - 3{,}2 = 0{,}7 \text{ трлн. грн.}$,

$$T_{p_{10}} = \frac{y_{\text{к.}10}}{y_{\text{к.}09}} = \frac{3{,}9}{3{,}2} = 1{,}219 \text{ или } 121{,}9 \text{ \%},$$

$$T_{\text{пр}_{10}}(\%) = T_{p_{10}}(\%) - 100 = 121{,}9 - 100 = 21{,}9 \text{ \%}.$$

Итак, за 2010 год стоимость основных средств в Украине возросла на 0,7 трлн. грн. или в 1,219 раза или на 21,9 %.

2) $\Delta_{08-10}= y_{\text{к}.10} - y_{\text{к}.07} = 3{,}9 - 1{,}3 = 2{,}6$ трлн. грн.,

$T_{p_{08-10}} = \frac{y_{\text{к}.10}}{y_{\text{к}.07}} = \frac{3{,}9}{1{,}3} = 1{,}300$ или $130{,}0$ %,

$T_{\text{пр}_{08-10}}(\%) = T_{p_{08-10}}(\%) - 100 = 130{,}0 - 100 = 30{,}0$ %.

Таким образом, за период 2008-2010 гг. стоимость основных средств в Украине возросла на 2,6 трлн. грн. или в 1,3 раза или на 30 %.

10.4. СРЕДНИЕ ПОКАЗАТЕЛИ ДИНАМИКИ

Со временем изменяются не только уровни, но и показатели динамики – абсолютные приросты, темпы роста и прироста. **Для обобщенной характеристики развития используют соответствующие средние показатели динамики**:

- средний абсолютный прирост;
- средний темп роста;
- средний темп прироста.

Средний абсолютный прирост характеризует среднюю абсолютную скорость роста и показывает, на сколько единиц увеличивался или уменьшался уровень по сравнению с базисным уровнем в среднем за единицу времени (в среднем ежегодно, ежеквартально, ежемесячно):

$$\overline{\Delta_t}= \frac{y_i - y_{i-t}}{t}, \tag{10.15}$$

или

$$\overline{\Delta_t}= \frac{\Delta_t}{t}, \tag{10.16}$$

или

$$\overline{\Delta_t}= \frac{\sum \Delta_{\text{ц}}}{t} \tag{10.17}$$

где $\sum \Delta_{\text{ц}}$ – сумма последовательных цепных абсолютных приростов.

Рассчитаем среднегодовой абсолютный прирост выпуска продукции за 2008-2009 гг. по данным таблицы 10.3:

На основе формулы 10.15:

$$\bar{\Delta}_{\substack{08-09\\t=2}}=\frac{y_{09}-y_{07}}{2}=\frac{62,1-55,7}{2}=3,2\text{ млн. грн.}$$

По формуле 10.16:

$$\bar{\Delta}_{\substack{08-09\\t=2}}=\frac{\Delta_{08-09}}{2}=\frac{6,4}{2}=3,2\text{ млн. грн.}$$

На основе формулы 10.17:

$$\bar{\Delta}_{\substack{08-09\\t=2}}=\frac{\Delta_{08}+\Delta_{09}}{2}=\frac{5,7+0,7}{2}=3,2\text{ млн. грн.}$$

Средний темп роста характеризует среднюю относительную скорость роста. Выраженный в форме коэффициента, показывает, во сколько раз увеличивался уровень в среднем за единицу времени (или какую часть базисного уровня составлял):

$$\bar{T}_{p_t}=\sqrt[t]{\frac{y_i}{y_{i-t}}}, \tag{10.18}$$

или

$$\bar{T}_{p_t}=\sqrt[t]{T_{p_t}}, \tag{10.19}$$

или

$$\bar{T}_{p_t}=\sqrt[t]{\Pi T_{p_ц}}, \tag{10.20}$$

где $\Pi T_{p_ц}$- произведение последовательных цепных темпов роста.

Формула 10.20 представляет собой формулу средней геометрической простой.

Рассчитаем среднегодовой темп роста выпуска продукции за 2008-2009 гг. по данным таблицы 10.20.

На основе формулы 10.18:

$$\bar{T}_{p\substack{08-09\\t=2}}=\sqrt[2]{\frac{y_{09}}{y_{07}}}=1,056\text{ или }105,6\text{ \%}.$$

По формуле 10.19:

$$\bar{T}_{p\substack{08-09\\t=2}}=\sqrt[2]{T_{p_{08-09}}}=1,056\text{ или }105,6\text{ \%}.$$

На основе формулы 10.20:

$$\bar{T}_{\underset{t=2}{p08-09}} = \sqrt[2]{T_{p_{08}} \cdot T_{p_{09}}} = 1{,}056 \text{ или } 105{,}6\ \%.$$

Средний темп прироста характеризует среднюю относительную скорость прироста. Выраженный в форме %, показывает, на сколько % увеличивался или уменьшался уровень в среднем за единицу времени:

$$\bar{T}_{\text{пр}_t}(\%) = \bar{T}_{p_t}(\%) - 100. \tag{10.21}$$

$$\bar{T}_{\underset{t=2}{\text{пр}08-09}}(\%) = \bar{T}_{\underset{t=2}{p08-09}}(\%) - 100 = 105{,}6 - 100 = 5{,}6\ \%.$$

Таким образом, за период 2008-2009 гг. выпуск продукции в среднем ежегодно увеличивался на 3,2 млн. грн. или в 1,056 раза или на 5,6 %.

10.5. ОЦЕНКА УСКОРЕНИЯ (ЗАМЕДЛЕНИЯ) РАЗВИТИЯ. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ДИНАМИЧЕСКИХ РЯДОВ

Если скорость роста в пределах изучаемого периода неодинакова, измеряют ускорение или замедление динамики путем сравнения одноимённых характеристик скорости роста.

Причем, если сравниваемые периоды одинаковой продолжительности, то можно пользоваться как элементарными показателями динамики, так и средними характеристиками скорости роста и прироста. Если же периоды не одинаковы, то сравнивают только средние показатели динамики.

На базе абсолютных приростов, в том числе средних, оценивают абсолютное и относительное ускорение.

Абсолютное ускорение (δ) – это разница между абсолютными приростами за два последовательных периода:

$$\delta = \Delta_2 - \Delta_1, \tag{10.22}$$

где 1 и 2 – номера сравниваемых периодов.

Ускорение характеризуется положительной величиной, а замедление – отрицательной.

Относительное ускорение (коэффициент ускорения абсолютной скорости роста δ') – это соотношение абсолютных приростов двух последовательных периодов:

$$\delta' = \frac{\Delta_2}{\Delta_1}. \tag{10.23}$$

При этом сравнивать есть смысл лишь абсолютные скорости с одинаковыми знаками.

Если абсолютный прирост сравниваемого периода меньше базисного (по модулю), в числителе записывают базисный абсолютный прирост, в знаменателе – сравниваемый. В этом случае полученный показатель будет характеризовать замедление.

Пример. За период 2006-2007 гг. средняя номинальная заработная плата в Украине в среднем ежегодно увеличивалась на 272,5 грн., а за 2008-2009 гг. – этот же показатель в среднем ежегодно возрастал на 277,5 грн. (цифры получены по данным таблицы 10.1).

Определим абсолютное и относительное ускорение среднегодовой заработной платы за период 2008-2009 гг. по сравнению с периодом 2006-2007 гг.

На основе формул 10.22 и 10.23 запишем:

$$\delta_{\bar{\Delta}} = \bar{\Delta}_{08-09} - \bar{\Delta}_{06-07} = 277{,}5 - 272{,}5 = 5 \text{ грн.},$$

$$\delta'_{\bar{\Delta}} = \frac{\bar{\Delta}_{08-09}}{\bar{\Delta}_{06-07}} = \frac{277{,}5}{272{,}5} = 1{,}018 \text{ или } 101{,}8\ \% \ (+1{,}8\ \%).$$

Таким образом, в Украине за период 2008-2009 гг. по сравнению с периодом 2006-2007 гг. среднегодовая абсолютная скорость роста номинальной заработной платы увеличилась на 5 грн. или в 1,018 раза или на 1,8 %.

На основе темпов прироста, имеющих одинаковые знаки, рассчитывают **коэффициенты ускорения (замедления) относительных скоростей развития** ($K_{T_{\text{пр}}}^{\text{уск.}}$). При этом делителем является больший по модулю темп прироста.

Приведем формулы коэффициентов ускорения (замедления) для средних темпов прироста:

$$K_{T_{\text{пр}}}^{\text{уск.}} = \bar{T}_{\text{пр}2} \div \bar{T}_{\text{пр}1}, \text{если } \left|\bar{T}_{\text{пр}2}\right| > \left|\bar{T}_{\text{пр}1}\right|, \tag{10.24}$$

$$K_{T_{\text{пр}}}^{\text{зам.}} = \bar{T}_{\text{пр}1} \div \bar{T}_{\text{пр}2}, \text{если } \left|\bar{T}_{\text{пр}1}\right| > \left|\bar{T}_{\text{пр}2}\right|. \tag{10.25}$$

Сравним относительные скорости прироста среднемесячной заработной платы в Украине в 2009 г. по сравнению с 2008 г. (по данным таблицы 10.1).

Прежде всего, рассчитаем темпы прироста за 2008 г. и за 2009 г. по формуле 10.13:

$$T_{\text{пр}_{08}}(\%) = T_{p_{08}}(\%) - 100 = \frac{y_{08}}{y_{07}} \cdot 100 - 100 = \frac{1806}{1351} \cdot 100 - 100 = 33{,}7\ \%,$$

$$T_{\text{пр}_{09}}(\%) = T_{p_{09}}(\%) - 100 = \frac{y_{09}}{y_{08}} \cdot 100 - 100 = \frac{1906}{1806} \cdot 100 - 100 = 5{,}5\ \%.$$

Так как темп прироста сравниваемого периода меньше темпа прироста базисного, рассчитываем коэффициент замедления по формуле 10.25:

$$K_{T_{\text{пр}}}^{\text{зам.}} = \frac{T_{\text{пр}_{08}}}{T_{\text{пр}_{09}}} = \frac{33{,}7}{5{,}5} = 6{,}127$$

Таким образом, в Украине в 2009 г. по сравнению с 2008 г. относительная скорость прироста среднемесячной заработной платы уменьшилась более чем в 6 раз.

Для двух рядов динамики с одноименными показателями (для одного периода) рассчитывают **коэффициент опережения:**

$$K_t^{\text{оп.}} = \frac{T_t^{A}}{T_t^{\text{Б}}}, \tag{10.26}$$

где T - темпы роста или прироста (с одинаковыми знаками) одноименного показателя,

А и Б – наименование региона или объекта.

Делителем является больший по значению показатель динамики.

Например, за текущий период на первом предприятии производительность труда повысилась на 5 %, а на втором – на 6,7 %. Определим на основании формулы 10.26 коэффициент опережения относительной скорости прироста:

$$K_t^{\text{оп.}} = \frac{T_{\text{пр}t}^2}{T_{\text{пр}t}^1} = \frac{6{,}7}{5{,}0} = 1{,}340 \text{ или } 134{,}0\ \% \ (+34{,}0\ \%).$$

Итак, в текущем периоде относительная скорость прироста производительности труда на втором предприятии оказалась выше, чем на первом в 1,340 раза или на 34,0 %.

Если в рядах динамики с разноименными признаками один из них может рассматриваться как факторный (x), а другой – как результативный (y), то сравнивают темпы прироста результативного и факторного признаков. Такое соотношение называется **коэффициентом эластичности**:

$$Э_t = \frac{T_{\text{пр}t}^y}{T_{\text{пр}t}^x}. \qquad (10.27)$$

Коэффициент эластичности показывает, на сколько процентов изменяется результативный показатель при изменении факторного на 1 %.

Темпы прироста в формуле 10.27 могут быть с разными знаками.

Например, в текущем периоде магазин повысил цены на бытовую технику на 10 %, объем продаж за этот период уменьшился на 20 %. Определим коэффициент эластичности спроса на бытовую технику в зависимости от цены на нее, используем формулу 10.27:

$$Э_t = \frac{-20}{10} = -2\ \%.$$

Итак, в текущем периоде с увеличением цен на бытовую технику на 1 % спрос на нее снизился на 2 %.

10.6. ВЫЯВЛЕНИЕ И ИЗМЕРЕНИЕ ТЕНДЕНЦИЙ РАЗВИТИЯ

Уровни ряда динамики со временем меняются. Данные изменения вызваны различными причинами. **Динамика ряда включает три компонента**:

- тенденцию (долговременные изменения);
- кратковременные систематические изменения, например, сезонные;
- несистематические случайные изменения.

Тенденция – общее направление развития, определенная закономерность изменения уровней. Одним рядам динамики присуща тенденция к росту, другим – к снижению, третьим – к стабилизации.

Для выявления и изучения основной тенденции развития применяют разные методы. Их использование предваряет и дополняет графический метод, прежде всего, построение линейной диаграммы по фактическим данным ряда динамики. На основе диаграммы можно сделать выводы о наличии тенденции и ее типе.

Если цепные показатели динамики от года к году практически не изменяются, то для количественной характеристики основной тенденции можно использовать средний абсолютный прирост и средний темп прироста (формулы 10.15, 10.21).

Но если тенденция недостаточно отчетлива, цепные показатели динамики сильно колеблются, формулы 10.15 и 10.21 исказят основную характеристику развития явления. Это связано с тем, что формулы учитывают только крайние значения, а они могут быть нетипичными для данного ряда динамики.

Необходимы методы, которые учитывали бы и промежуточные значения во временном ряду.

Статистика предлагает следующие **методы характеристики общей тенденции развития**:

- сглаживание уровней ряда;
- аналитическое выравнивание.

Метод сглаживания уровней ряда заключается в преобразовании и замене первичного ряда динамики другим, уровни которого относятся к большим по продолжительности периодам времени. То есть ряд динамики, например, месячных уровней, путем их суммирования преобразуют в ряд динамики квартальных, годовых и т.д. уровней. При этом отклонения в уровнях, обусловленные действием случайных причин, взаимно погашаются, сглаживаются. Проявляется основная тенденция изменения уровней.

Существуют **два способа сглаживания**, в соответствии с которыми образуется новый ряд динамики:

- ступенчатых средних;
- скользящих средних.

Рассмотрим использование данных способов на примере анализа тенденции динамики числа иностранных туристов в регионе (таблица 10.4).

Для построения **ряда ступенчатых средних** перейдем от ежегодных уровней к среднегодовым уровням за каждые последовательные три года. Рассчитаем их по формуле средней арифметической простой и запишем в таблице 10.4 на уровне центрального года каждого трехлетнего периода. Продемонстрируем расчет первой ступенчатой средней:

$$\bar{y}_{02-04} = \frac{y_{02}+y_{03}+y_{04}}{3} = \frac{10{,}0+10{,}7+12{,}0}{3} = 10{,}9 \text{ тыс.}$$

Таблица 10.4

Динамика числа иностранных туристов в регионе

Год	y_i, тыс.	3-летняя ступенчатая средняя, тыс.	3-летняя скользящая средняя, тыс.	t	t^2	ty	Y_t	$(y_i - Y_t)^2$
2002	10,0	-	-	-4	16	-40,0	9,30	0,49
2003	10,7	10,9	10,9	-3	9	-32,1	10,41	0,08
2004	12,0	-	11,0	-2	4	-24,0	11,52	0,23
2005	10,3	-	11,8	-1	1	-10,3	12,63	5,43
2006	12,9	13,2	13,2	0	0	0,0	13,74	0,71
2007	16,3	-	14,9	1	1	16,3	14,85	2,10
2008	15,6	-	16,6	2	4	31,2	15,96	0,13
2009	17,8	17,1	17,1	3	9	53,4	17,07	0,53
2010	18,0	-	-	4	16	72,0	18,18	0,03
Итого	123,6	-	-	0	60	66,5	123,66	9,73

Новый ряд динамики ступенчатых средних свидетельствует о тенденции к росту числа иностранных туристов в регионе, так как $10{,}9 < 13{,}2 < 17{,}1$.

Явным недостатком данного способа сглаживания является значительное сокращение числа уровней.

Использование **способа скользящих средних** не ведет к резкому сокращению уровней. Способ заключается в следующем. Рассчитывают среднюю величину из определенного числа первых уровней, затем – из такого же числа уровней, но, начиная со второго, далее – с третьего и т.д., то есть происходит скольжение (движение) по ряду динамики: начальный уровень периода отбрасывается и добавляется следующий за ним.

Для того, чтобы среднюю можно было отнести к конкретному периоду времени (году, например), целесообразно формировать интервалы с нечетным числом уровней. Если это невозможно (при изучении сезонности, например), осуществляют процедуру центрирования. Данная процедура будет рассмотрена в следующем параграфе.

Итак, по данным таблицы 10.4 рассчитаем уровни и построим новый ряд динамики трехлетних скользящих средних. Для периода 2002-2004 расчет и значение среднегодового числа туристов совпадут с продемонстрированными нами выше. Сдвинемся (опустимся) на один год вниз по ряду динамики и определим среднюю величину для 2003-2005 гг.:

$$\bar{y}_{03-05} = \frac{y_{03}+y_{04}+y_{05}}{3} = \frac{10{,}7+12{,}0+10{,}3}{3} = 11{,}0 \text{ тыс.}$$

Итак, перед нами новый ряд динамики трехлетних скользящих средних. Он сглаживает случайные колебания отдельных лет и однозначно свидетельствует о тенденции к увеличению числа иностранных туристов в регионе. Данную тенденцию демонстрирует графическое изображение фактических и сглаженных уровней (рис 10.1).

Наиболее эффективным методом выявления основной тенденции развития является **аналитическое выравнивание**. При использовании этого метода фактические уровни ряда (y_i) заменяют на теоретические уровни, рассчитанные на основе определенной функции времени $Y_t = f(t)$, которую называют **уравнением тренда**. Выбор функции производится на основе характера закономерностей динамики изучаемого явления.

На практике преимущество отдается функциям, параметры которых имеют четкий экономический смысл и измеряют абсолютную или относительную скорость развития.

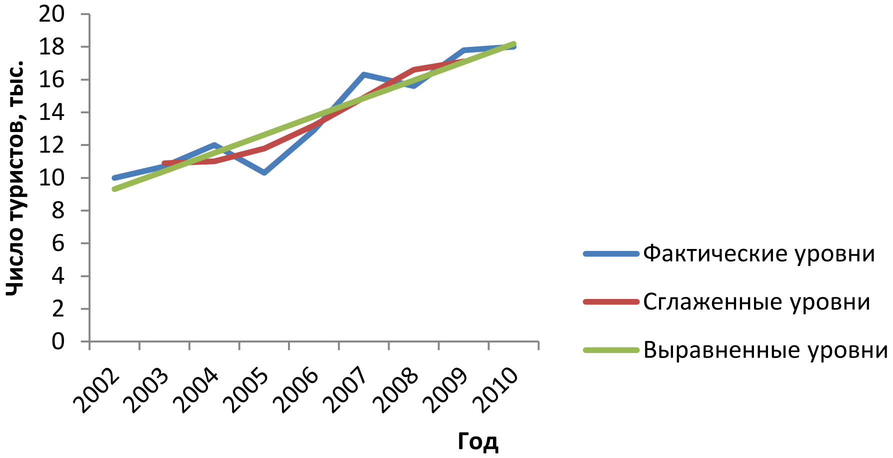

Рис. 10.1 - Динамика числа иностранных туристов

Если цепные абсолютные приросты относительно стабильны, выравнивание ряда осуществляется на основе линейной функции:

$$Y_t = a_0 + a_1 t, \tag{10.28}$$

где t – порядковый номер периода,

a_0– уровень ряда при $t = 0$,

a_1 – характеристика средней абсолютной скорости развития.

Если относительно стабильными являются цепные темпы прироста, выравнивание ряда производят на основе экспоненциальной функции:

$$Y_t = a_0 {a_1}^t, \tag{10.29}$$

где a_1– характеристика средней относительной скорости развития.

Параметры трендовых уравнений определяют методом наименьших квадратов. В соответствии с ним для линейной функции получают следующую систему нормальных уравнений:

$$\begin{cases} \sum y = na_0 + a_1 \sum t \\ \sum ty = a_0 \sum t + a_1 \sum t^2 , \end{cases} \tag{10.30}$$

где n – число уровней ряда динамики.

В рядах динамики техника расчета параметров уравнения может быть упрощена. Для этой цели показателям времени t придают такие значения, чтобы их сумма была равна нулю, то есть $\sum t = 0$. При этом уравнения системы примут следующий вид:

$$\begin{cases} \sum y = na_0 \\ \sum ty = a_1 \sum t^2 \end{cases}, \tag{10.31}$$

откуда:

$$a_0 = \frac{\sum y}{n}, \tag{10.32}$$

$$a_1 = \frac{\sum ty}{\sum t^2}. \tag{10.33}$$

По данным о динамике числа иностранных туристов, отраженной в таблице 10.4, произведем расчет параметров:

$$a_0 = \frac{123{,}6}{9} = 13{,}74$$

$$a_1 = \frac{66{,}5}{60} = 1{,}11.$$

В результате получаем следующее уравнение основной тенденции числа иностранных туристов в регионе:

$$Y_t = 13{,}74 + 1{,}11t.$$

Таким образом, среднегодовое число иностранных туристов составило 13,74 тыс. В среднем ежегодно численность иностранных туристов увеличивается на 1110 человек.

Подставляя в уравнение принятые обозначения t, вычислим выравненные уровни ряда динамики:

$$Y_{02} = 13{,}74 + 1{,}11 \cdot (-4) = 9{,}3,$$

$$Y_{03} = 13{,}74 + 1{,}11 \cdot (-3) = 10{,}41$$

и т.д. (см. значение Y_t в таблице 10.4).

По окончании расчета основной тенденции целесообразно представить выравненные уровни графически (рис. 10.1).

Вернемся к выбору формы тренда. Для такого выбора целесообразно использовать сумму квадратов отклонений фактических уровней ряда динамики от теоретических уровней, скорректированную на число степеней свободы:

$$S_\varepsilon^2 = \frac{\sum(y_i - Y_t)^2}{n-m}, \qquad (10.34)$$

где m- число параметров тренда. Выбирается тот тренд, для которого S_ε^2 меньше. Также в качестве критерия может выступать коэффициент детерминации, представляющий собой отношение факторной дисперсии к остаточной. Выбирается тот тип функции, для которого коэффициент больше.

Можно воспользоваться также встроенным блоком КРА редактора Excel.

Результаты такого анализа свидетельствуют, что трендовая модель $Y_t = 13{,}74 + 1{,}11t$ достаточно точно описывает динамику числа иностранных туристов в регионе. Значение $R^2 = 0{,}939850707$ говорит о том, что 94 % вариации числа туристов объясняется линейным трендом. Коэффициент a_1 является статистически значимым и надежным.

При условии, что комплекс причин, которые формируют тенденцию, в ближайшее время не изменится, можно продолжить тенденцию за границы ряда динамики, то есть осуществить **экстраполяцию тренда** и определить Y_{t+L}, где L – период упреждения.

Ожидаемое число иностранных туристов в 2011 году составит 19,29 тыс. чел.:

$$Y_{11} = 13{,}74 + 1{,}11 \cdot 5 = 19{,}29.$$

Значение 19,29 тыс. чел. – это **точечный**, маловероятный прогноз.

Так как статистический прогноз имеет вероятностный характер, прогнозные значения обычно определяют в виде доверительных интервальных значений, то есть определяют **интервальный прогноз**.

Интервальный прогноз рассчитывается с определенной вероятностью по формуле:

$$Y_{t+L} \pm t_{1-\alpha} S_p, \qquad (10.35)$$

где $t_{1-\alpha}$ – коэффициент доверия для принятого уровня вероятности,

S_p – стандартная ошибка прогноза.

Ошибка прогноза рассчитывается по формуле:

$$S_p = S_\varepsilon \sqrt{\frac{n+1}{n} + \frac{3(n+2L-1)^2}{n^3-n}}, \qquad (10.36)$$

где S_ε – величина средней квадратической ошибки, определяемая как $\sqrt{S_\varepsilon^2}$.

По формуле 10.34:

$$S_\varepsilon^2 = \frac{9{,}73}{9-2} = 1{,}39.$$

Следовательно

$$S_\varepsilon = \sqrt{1{,}39} = 1{,}18.$$

Определим коэффициент:

$$\sqrt{\frac{n+1}{n} + \frac{3(n+2L-1)^2}{n^3-n}} = \sqrt{\frac{9+1}{9} + \frac{3(9+2-1)^2}{9^3-9}} = 1{,}23.$$

Критическое значение t - критерия для $\alpha = 0{,}05$ и числа степеней свободы $n-2 = 9 - 2 = 7$ составляет 2,36.

Поэтому

$$t_{1-\alpha} S_p = 2{,}36 \cdot 1{,}18 \cdot 1{,}23 = 3{,}43.$$

Таким образом, доверительные интервалы прогнозного значения

$$19{,}29 \pm 3{,}43.$$

Итак, с вероятностью 95 % можно прогнозировать, что в 2011 г. численность иностранных туристов в регионе будет колебаться от 15,86 тыс. до 22,72 тыс. чел.

На практике анализ основной тенденции развития, прогнозирование проводят с помощью компьютерных программ.

10.7. ВЫЯВЛЕНИЕ И ИЗМЕРЕНИЕ СЕЗОННЫХ КОЛЕБАНИЙ

Сезонные колебания – это более или менее стойкие (постоянные) внутригодовые колебания уровней социально-экономических явлений. Они обусловлены влиянием одних и тех же факторов климатического и социального порядка.

Сезонные колебания проявляются прежде всего в таких сферах деятельности, как сельское хозяйство и обработка сельскохозяйственной продукции, транспорт, туризм. К сезонным явлениям относятся потребление электроэнергии, газа, спрос на определенные виды продукции.

Сезонные колебания усложняют работу предприятий, вызывая неравномерность использования их материально-технической базы, трудовых ресурсов, ведут к другим негативным, отрицательным последствиям экономического, социального, организационно-технического характера.

В связи с этим сезонные колебания нуждаются в регулировании, а значит – статистическом изучении.

Характер сезонных колебаний описывается сезонной волной, которую образуют индексы сезонности.

Индексы сезонности (I_i) – это процентное отношение фактических уровней за одноименные месяцы (кварталы) к теоретическим.

Формула расчета индексов сезонности уточняется в зависимости от исходной информации.

Если имеется информация только за один год, используется следующая формула:

$$I_i = \frac{y_i}{\bar{y}} \cdot 100, \tag{10.37}$$

где $\bar{y}$ – среднемесячный (среднеквартальный) уровень за данный год.

Так как явление сезонности присуще такой развивающейся сфере, как туризм, в данном разделе будут приведены примеры, связанные с этим видом деятельности.

В таблице 10.5 представлены рассчитанные индексы сезонности по данным об уровнях туристопотока в регионе. Для этого по формуле средней арифметической простой был рассчитан среднемесячный уровень в текущем году, а затем по формуле 10.37 для каждого месяца определен индекс сезонности.

Таблица 10.5

Динамика туристопотока (количества туристов)
в регионе в текущем периоде

Месяц	y_i, тыс. чел.	I_i, %
Январь	30	54,5
Февраль	28	50,9
Март	27	49,1
Апрель	32	58,2
Май	40	72,7
Июнь	64	116,4
Июль	95	172,4
Август	110	200,0
Сентябрь	90	163,6
Октябрь	65	118,2
Ноябрь	44	80,0
Декабрь	35	63,6
Итого	660	-

Данные таблицы свидетельствуют, что пик, самая высокая точка сезонной волны, приходилась на август ($I_{\text{авг.}} = 200\ \%$) , и лишь 49,1 % от среднемесячного уровня приходился на март – месяц, когда туристов минимальное количество.

Результаты расчетов индексов сезонности являются более надежными, если в качестве исходной информации используются месячные (квартальные) уровни не за один год, а за ряд лет. Это объясняется тем, что в каждом конкретном году одновременно происходят сезонные колебания и чисто случайные.

Если известны данные за несколько лет, тогда знаменателем формулы индекса сезонности выступает либо средний уровень, либо сглаженный или выравненный уровень.

Средние месячные (квартальные) уровни используются, если ряд динамики, кроме сезонных колебаний, не выявляет определенной основной тенденции развития.

Если же ряду динамики свойственна тенденция к росту или снижению, то знаменателем в формуле индекса сезонности должны стать сглаженные или выравненные уровни.

Ниже рассмотрена ситуация, при которой ряд динамики, кроме сезонных (и случайных) колебаний, не выявляет определенной основной тенденции (таблица 10.6).

Таблица 10.6

Динамика туристопотока в регионе

Квартал	y_{ij}, тыс. чел				$\sum y_i$	$\bar{y}_i$	I_i, %	$(I_i - 100)^2$
	2007	2008	2009	2010				
1	341,1	333,4	348,7	216,5	1239,7	309,9	68,5	992,25
2	496,9	569,4	390,4	175,8	1632,5	408,1	90,2	96,04
3	1013,7	1051,3	658,4	371,5	3094,9	773,7	171,0	5041,00
4	364,0	432,3	318,2	157,4	1271,9	318,0	70,3	882,09
Итого (в средн)	2215,7	2386,4	1715,7	921,2	7239,0	$\bar{y}_{ij}$= 452,4	400,0	7011,38

В этом случае используется формула:

$$I_i = \frac{\bar{y}_i}{\bar{y}_{ij}} \cdot 100, \qquad (10.38)$$

где $\bar{y}_i$- средние уровни за одноименные месяцы (кварталы),

$\bar{y}_{ij}$- общая средняя (месячная или квартальная),

i- обозначение (номер) месяца (квартала),

j- обозначение (номер) года.

Промежуточные расчеты отражены в таблице. Полученные индексы сезонности отражают внутригодовые колебания, очищенные от случайных колебаний каждого года. Пик сезонной волны приходится на третий квартал, а в первом квартале количество туристов наименьшее.

Для наглядности восприятия изобразим сезонную волну графически.

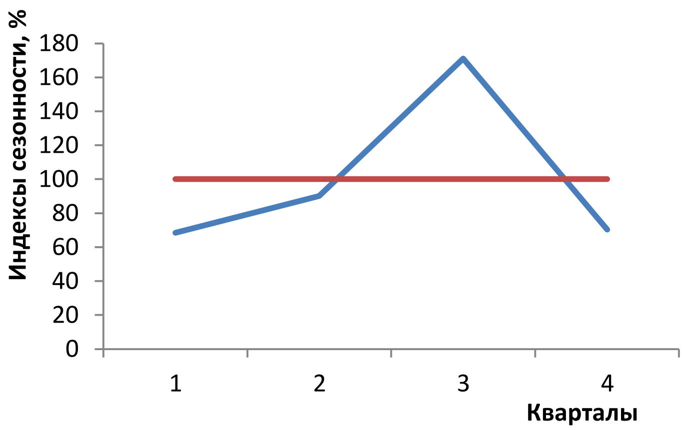

Рис. 10.2 - Сезонная волна туристопотока

Если на протяжении периода кроме сезонных колебаний наблюдается достаточно выраженная тенденция к росту или уменьшению уровней, расчет индексов сезонности усложняется. Их определению предшествует процедура сглаживания или выравнивания уровней ряда динамики известными нам методами. Сглаженные или выравненные уровни будут знаменателями индексов сезонности.

Рассмотрим **этапы расчета индексов сезонности с помощью скользящей средней**.

1. Определяются четырехквартальные (12–месячные) скользящие средние, так как именно 4 квартала (12 месяцев) охватывают год и всю сезонную волну. Из-за того, что четырехквартальные (12–месячные) показатели рассчитывают на основе парного числа уровней, скользящую среднюю нельзя отнести к конкретному кварталу (месяцу). Ее записывают в промежутке между кварталами (месяцами), для которых эта средняя определялась.
2. Осуществляют процедуру **центрирования**. Она заключается в определении **центрированных скользящих средних** – это средние величины из двух соседних нецентрированных скользящих средних величин. Центрированные средние уже можно отнести к конкретным кварталам (месяцам).
3. Для тех уровней, для которых рассчитаны центрированные скользящие средние, определяют индексы сезонности по формуле:

$$I_i = \frac{y_i}{\bar{y}_{\text{скольз.}}} \cdot 100, \qquad (10.39)$$

где $\bar{y}_{\text{скольз.}}$ – центрированные скользящие средние.

4. Для каждого квартала (месяца) находят средние индексы сезонности по формуле средней арифметической простой.
5. Осуществляют арифметический контроль: для квартальных уровней сумма индексов сезонности должна быть равна 400, для месячных – 1200.

По данным таблицы 10.7 определим индексы сезонности, предварительно осуществим необходимые промежуточные расчеты.

Во-первых, убедимся, что представленный в таблице 10.7 ряд динамики характеризуется тенденцией к росту. От года к году года растут и одноименные квартальные уровни, и суммарные за год. В таблице представлены расчеты нецентрированных и центрированных скользящих средних, рассчитаны индексы сезонности.

Таблица 10.7

Динамика туристопотока и расчет индексов сезонности с помощью скользящей средней

Год и квартал	y_i, тыс. чел.	4–квартальная скользящая средняя (нецентрированная)	Двухчленная скользящая средняя (центрированная)	I_i, %
2008 г. 1	24,4	-	-	-
2	52,6	-	-	-
		42,9	-	-
3	60,4	-	43,9	137,6
		44,9		
4	34,0	-	45,4	74,9
		45,8		
2009 г. 1	32,7	-	46,7	70,0
		47,6		
2	56,2	-	47,9	117,3
		48,1		
3	67,3	-	48,8	137,9
		49,4		
4	36,2	-	50,6	71,5
		51,7		
2010 г. 1	37,8	-	52,4	72,1
		53,1	-	-
2	65,3	-	53,4	122,3
		53,7	-	-
3	73,1	-	-	-
				-
4	38,4	-	-	-
Итого	578,4	-	-	-

Для одноименных индексов определим средние индексы сезонности, которые и описывают сезонную волну:

$$\bar{I}_1 = \frac{70 + 72{,}1}{2} = 71{,}1\ \%,$$

$$\bar{I}_2 = \frac{117{,}3 + 122{,}3}{2} = 119{,}8\ \%,$$

$$\bar{I}_3 = \frac{137{,}6 + 137{,}9}{2} = 137{,}8\ \%,$$

$$\bar{I}_4 = \frac{74{,}9 + 71{,}5}{2} = 73{,}2\ \%.$$

Арифметический контроль подтверждает правильность расчётов (небольшая погрешность в результате округлений).

Рассмотрим **этапы расчета индексов сезонности с помощью аналитического выравнивания**.

1. Для каждого квартала (месяца) рассчитывают теоретические уровни на основе аналитического выравнивания. В таблице 10.8 произведены промежуточные расчеты, на основе которых определим (по формулам 10.32 и 10.33) параметры линейной модели:

$$a_0 = \frac{578{,}4}{12} = 48{,}2,$$

$$a_1 = \frac{213{,}4}{143} = 1{,}492,$$

и построим следующее уравнение основной тенденции динамики туристопотока:

$$Y_t = 48{,}2 + 1{,}492t.$$

2. Определяют индексы сезонности как отношение фактических и теоретических уровней ряда динамики по формуле:

$$I_i = \frac{y_i}{Y_t} \cdot 100. \qquad (10.40)$$

3. Для одноименных кварталов (месяцев) рассчитывают средние индексы сезонности по формуле средней арифметической простой:

$$\bar{I}_1 = \frac{61 + 71{,}1 + 72{,}8}{3} = 68{,}3\ \%,$$

$$\bar{I}_2 = \frac{126{,}7 + 118{,}3 + 122{,}3}{3} = 122{,}4\ \%,$$

$$\bar{I}_3 = \frac{140{,}5 + 137{,}6 + 133{,}2}{3} = 137{,}1\ \%,$$

$$\bar{I}_4 = \frac{76{,}4 + 71{,}8 + 68{,}1}{3} = 72{,}1\ \%.$$

4. Осуществляют арифметический контроль:

$$68{,}3 + 122{,}4 + 137{,}1 + 72{,}1 \approx 400.$$

5. Производят корректировку тренда на сезонность (таблица 10.8) по формуле:

$$Y_t^{\text{сез.}} = Y_t I_i. \qquad (10.41)$$

6. Осуществляют прогнозирование на каждый квартал следующего года с учетом основной тенденции и сезонных колебаний. Продемонстрируем расчет и результаты точечных прогнозов на каждый квартал 2011 года:

$$Y_{2011}^{1\text{кв}} = (48{,}2 + 1{,}492 \cdot 6{,}5) \cdot 0{,}683 = 39{,}5 \text{ тыс. туродней}$$

$$Y_{2011}^{2\text{кв}} = (48{,}2 + 1{,}492 \cdot 7{,}5) \cdot 1{,}224 = 72{,}7 \text{ тыс. туродней}$$

$$Y_{2011}^{3\text{кв}} = (48{,}2 + 1{,}492 \cdot 8{,}5) \cdot 1{,}371 = 83{,}5 \text{ тыс. туродней}$$

$$Y_{2011}^{4\text{кв}} = (48{,}2 + 1{,}492 \cdot 9{,}5) \cdot 0{,}721 = 45{,}0 \text{ тыс. туродней}$$

Таблица 10.8

К расчету индексов сезонности
с помощью аналитического выравнивания

Год и квартал	y_i, тыс. чел.	t	ty	t^2	Y_t	I_i, %	$Y_t^{\text{сез.}}$
2008 г. 1	24,4	-5,5	-134,20	30,25	40,0	61,0	27,3
2	52,6	-4,5	-236,70	20,25	41,5	126,7	50,8
3	60,4	-3,5	-211,40	12,25	43,0	140,5	59,0
4	34,0	-2,5	-85,00	6,25	44,5	76,4	32,1
2009 г. 1	32,7	-1,5	-49,05	2,25	46,0	71,1	31,4
2	56,2	-0,5	-21,10	0,25	47,5	118,3	58,1
3	67,3	0,5	33,65	0,25	48,9	137,6	67,0
4	36,2	1,5	54,30	2,25	50,4	71,8	36,3
2010 г. 1	37,8	2,5	94,20	6,25	51,9	72,8	35,4
2	65,3	3,5	228,55	12,25	53,4	122,3	65,4
3	73,1	4,5	328,95	20,25	54,9	133,2	75,3
4	38,4	5,5	211,20	30,25	56,4	68,1	40,7
Итого	578,4	0,0	213,40	143,00	578,4	х	578,8

Кроме индексов сезонности соответствующее явление характеризуют другие показатели сезонности.

Самой простой характеристикой сезонной волны является **амплитуда (широта) сезонных колебаний:**

$$R = I_{max} - I_{min}, \tag{10.42}$$

где I_{max}, I_{min} – соответственно максимальное и минимальное значение индексов сезонности.

По данным таблицы 10.6 амплитуда сезонных колебаний составит:

$$R = 171{,}0 - 68{,}5 = 102{,}5\ \%.$$

Следующей характеристикой сезонных колебаний является **абсолютный размах сезонных колебаний** (Δ_i). Он рассчитывается для каждого квартала (месяца) и показывает, на сколько единиц уровень каждого квартала (месяца) больше или меньше среднего. По данным таблицы 10.6, то есть в условиях, когда сезонность изучается по данным нескольких лет, формула абсолютного размаха сезонных колебаний запишется:

$$\Delta_i = \bar{y}_i - \bar{y}_{ij}. \tag{10.43}$$

Относительный размах сезонных колебаний (Δ'_i) также рассчитывается для каждого квартала (месяца) и показывает, на сколько процентов уровень в отдельные кварталы (месяцы) больше или меньше среднего:

$$\Delta'_i = I_i - 100. \tag{10.44}$$

Результаты расчётов абсолютного и относительного размаха сезонных колебаний по данным таблицы 10.6 отражены в таблице 10.9.

Таблица 10.9

Показатели абсолютного и относительного размаха сезонных колебаний туристопотока

Показатель	1 квартал	2 квартал	3 квартал	4 квартал	Итого
Δ_i, тыс. чел.	-142,5	-44,3	321,3	-134,5	0,0
Δ'_i, %	-31,5	-9,5	71,0	-29,7	0,0

Выводы по результатам расчетов следующие.

В первом квартале объем туристопотока оказался меньше среднеквартального на 142,5 тыс. чел. или на 31,5 %.

Во втором квартале также объем туристопотока меньше среднеквартального: в абсолютном выражении – на 44,3 тыс. чел., а в относительном – на 9,5 %.

В третьем квартале наблюдается пик сезонной волны туристопотока – его уровень превышает среднеквартальный на 321,3 тыс. чел. или на 71 %.

В четвертом квартале вновь уровень туристопотока оказался ниже среднего на 134,5 тыс. чел. или на 29,7 %.

Основным показателем силы сезонных колебаний в ряду динамики является **среднеквадратическое отклонение индексов сезонности** (σ):

$$\sigma = \sqrt{\frac{\sum(I_i - 100)^2}{4(12)}} \qquad (10.45)$$

Этот показатель следует применять при сравнении силы сезонных колебаний в пространстве или во времени.

По данным таблицы 10 с помощью формулы 10.45 рассчитаем среднее квадратическое отклонение индексов сезонности:

$$\sigma = \sqrt{\frac{7011{,}38}{4}} = 41{,}9\ \%.$$

Таким образом, индексы сезонности в отдельные кварталы отклоняются от 100 % в среднем на 41,9 %. Если величина данного показателя уменьшается в текущем периоде по сравнению с базисным, то это является свидетельством эффективности мер по борьбе с сезонностью.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Дайте определение ряда динамики.
2. Из каких обязательных элементов состоит ряд динамики?
3. Назовите виды динамических рядов.
4. Что характеризуют уровни интервального ряда динамики и каковы особенности данного вида временных рядов?
5. Опишите суть и особенности моментных рядов динамики.
6. Сформулируйте основные принципы построения динамических рядов.
7. Как рассчитывается средний уровень интервального ряда динамики?
8. Какие формулы и в каких ситуациях используют при определении среднего уровня моментных рядов динамики?
9. Назовите основные характеристики интенсивности динамики.
10. Что представляют собой цепные и базисные показатели динамики?
11. Как рассчитывается и что характеризует абсолютный прирост?
12. Приведите формулы расчеты цепного и базисного темпа роста.
13. Что представляет собой темп прироста и каковы формулы его расчета?
14. По какой формуле рассчитывается абсолютное значение 1 % прироста?
15. Назовите средние показатели динамики.
16. Что представляет собой средний абсолютный прирост?
17. Приведите формулы расчета средней абсолютной скорости роста.
18. Какие формулы можно использовать для определения среднего темпа роста?
19. Что характеризует показатель средней относительной скорости прироста?
20. Как рассчитываются абсолютное и относительное ускорение абсолютной скорости роста?

21. Какой показатель используют для сравнения относительной скорости прироста за два последовательных периода?
22. Что характеризует и как определяется коэффициент опережения?
23. Что показывает коэффициент эластичности?
24. Что представляет собой основная тенденция динамики?
25. Какие существуют методы анализа основной тенденции?
26. Какие существуют способы сглаживания уровней?
27. Опишите метод скользящей средней.
28. Опишите метод аналитического выравнивания?
29. Что представляет собой уравнение тренда?
30. При каких условиях для выравнивания уровней ряда динамики используется линейная функция?
31. На основании какой функции производят выравнивание ряда динамики при относительно стабильных цепных темпах прироста?
32. С помощью какого метода определяют параметры трендовых уравнений?
33. Что представляет собой экстраполяция тренда и как осуществляется точечный и интервальный прогноз?
34. Что представляют собой сезонные колебания и индексы сезонности?
35. Как рассчитываются индексы сезонности для одногодичных данных?
36. Приведите формулу расчёта индексов сезонности при отсутствии явно выраженной основной тенденции в ряду динамики за ряд лет.
37. Какой подход необходим к оценке сезонности при наличии четко выраженной основной тенденции?
38. Как рассчитываются индексы сезонности с помощью скользящей средней и с помощью аналитического выравнивания?
39. Что представляет собой среднеквадратическое отклонение индексов сезонности?

РАЗДЕЛ 11. ИНДЕКСНЫЙ МЕТОД

11.1. СУЩНОСТЬ И ФУНКЦИИ ИНДЕКСОВ В АНАЛИЗЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

Индекс - это относительная величина, которая характеризует изменение уровня явления во времени, в пространстве и по сравнению с планом.

Индексами можно считать относительные величины динамики, сравнения, планового задания и выполнения плана.

Индекс имеет две формы выражения – коэффициент и процент.

Величина, изменение которой характеризует индекс, называется **индексируемой величиной**.

Индексы выполняют две функции:

- синтетическую;
- аналитическую.

Синтетическая функция заключается в расчете относительного изменения индексируемого показателя.

Аналитическая функция состоит в определении влияния факторов на изменение результативного признака.

По степени охвата единиц совокупности различают следующие виды индексов:

- индивидуальные;
- сводные (общие).

Индивидуальный индекс (i) характеризует соотношение уровней одного элемента совокупности.

Сводный индекс (I) характеризует соотношение уровней нескольких элементов или совокупности в целом.

Индексируемые величины могут быть качественными (интенсивными) и объемными (количественными, экстенсивными).

Качественные показатели характеризуют объем признака в расчете на единицу совокупности (выпуск продукции в единицу времени, цена, себестоимость единицы продукции).

Объемные показатели характеризуют общий объем признака или общее число единиц совокупности (выпуск продукции, общая себестоимость, затраты времени и т.п.).

В статистике используется следующая **система обозначений индексируемых величин**:

Q – количество (объем) произведенной продукции или количество проданного товара данного вида в натуральном выражении (физический объем товарооборота: т, м, л, шт. и т.д.);

T – общие затраты рабочего времени на производство продукции данного вида (чел.-час, чел.-день) или среднесписочное число занятых (чел.);

z – себестоимость единицы продукции (грн.);

p – цена единицы продукции (грн.)

M – общий расход сырья, материала, топлива на производство продукции данного вида (кг, л, м);

На основе данных показателей могут быть получены следующие:

$q = \frac{Q}{T}$ – производство продукции данного вида в единицу времени (выработка продукции на одного рабочего), то есть производительность труда (т, м, л, шт.);

$t = \frac{T}{Q}$ – затраты рабочего времени на производство единицы продукции данного вида, то есть трудоемкость единицы продукции (чел.-час, чел.-день, чел.);

zQ – общая себестоимость продукции данного вида, то есть денежные затраты на ее производство;

pQ – общая стоимость произведенной продукции данного вида либо общая стоимость проданного товара данного вида, то есть товарооборот (грн.);

$m = \frac{M}{Q}$ – расход сырья, материла или топлива данного вида на производство единицы продукции, то есть удельный расход сырья, материала или топлива (кг, л, м).

В индексах динамики предыдущее значение величины, которое принимается за базу сравнения, обозначают «0», а отчетное (текущее, фактическое) обозначают «1».

Построим **индивидуальные индексы** динамики (далее индивидуальные индексы) некоторых показателей:

индивидуальный индекс цен

$$i_p = \frac{p_1}{p_0}; \tag{11.1}$$

индивидуальный индекс количества проданных товаров

$$i_Q = \frac{Q_1}{Q_0}; \tag{11.2}$$

индивидуальный индекс товарооборота

$$i_{pQ} = \frac{p_1 Q_1}{p_0 Q_0}. \tag{11.3}$$

Так как индивидуальные индексы – относительные величины, **между индексами сохранены все существующие между относительными величинами взаимосвязи**:

произведение индексов планового задания и выполнения плана равно индексу динамики (темпу роста):

$$\frac{z_{\text{пл.}}}{z_0} \cdot \frac{z_1}{z_{\text{пл.}}} = \frac{z_1}{z_0}, \tag{11.4}$$

произведение цепных индексов равно соответствующему базисному индексу:

$$\frac{z_1}{z_0} \cdot \frac{z_2}{z_1} \cdot \frac{z_3}{z_2} = \frac{z_3}{z_0}, \tag{11.5}$$

если произведение двух или нескольких показателей представляет собой новый показатель, имеющий реальный экономический смысл, то произведение индексов-сомножителей равно индексу их произведения, то есть индексу нового показателя:

$$\frac{z_1}{z_0} \cdot \frac{Q_1}{Q_0} = \frac{z_1 Q_1}{z_0 Q_0}. \qquad (11.6)$$

Например, в текущем периоде объем производства увеличился на 26 %, а производительность труда – на 25 %. Определим относительное изменение общих затрат времени и трудоемкости единицы продукции.

Так как $T = \frac{Q}{q}$, то $i_T = \frac{i_Q}{i_q} = \frac{1{,}26}{1{,}25} = 1{,}008$ или $100{,}8\ \%$ $(+0{,}8\ \%)$.

Из формулы $t = \frac{1}{q}$ следует, что $i_t = \frac{1}{i_q} = \frac{1}{1{,}25} = 0{,}800$ или $80{,}0\%$ $(-20\ \%)$.

Таким образом, в текущем периоде по сравнению с базисным общие затраты времени увеличились в 1,008 раза или на 0,8 %, а трудоемкость продукции снизилась на 20 %.

Построим **сводные индексы** некоторых показателей:

сводный индекс товарооборота

$$I_{pQ} = \frac{\sum p_1 Q_1}{\sum p_0 Q_0}, \qquad (11.7)$$

сводный индекс общей себестоимости

$$I_{zQ} = \frac{\sum z_1 Q_1}{\sum z_0 Q_0}, \qquad (11.8)$$

сводный индекс общих затрат рабочего времени

$$I_T = \frac{\sum T_1}{\sum T_0}. \qquad (11.9)$$

сводный индекс количества однородной продукции

$$I_Q = \frac{\sum Q_1}{\sum Q_0}. \qquad (11.10)$$

По данным таблицы 11.1 определим индивидуальные индексы физического объема продаж, цен и стоимости проданных товаров А и Б.

По формуле 11.2 рассчитаем индивидуальные индексы физического объема проданных товаров:

для А – $i_Q = \frac{82}{80} = 1{,}025$ или $102{,}5\ \%\ (+2{,}5\ \%)$,

для Б – $i_Q = \frac{124}{130} = 0{,}954$ или $95{,}4\ \%\ (-4{,}6\ \%)$.

Таким образом, в текущем периоде по сравнению с базисным физический объем реализации товара А увеличился в 1,025 раза или на 2,5 %, а товара Б – снизился на 4,6 %.

Определим индивидуальные индексы цен по формуле 11.1:

для А – $i_p = \frac{3{,}1}{2{,}5} = 1{,}240$ или $124{,}0\ \%\ (+24{,}0\ \%)$,

для Б – $i_p = \frac{6}{7{,}4} = 0{,}811$ или $81{,}1\ \%\ (-18{,}9\ \%)$.

Итак, в текущем периоде по сравнению с базисным цена тонны товара А возросла в 1,240 раза или на 24,0 %, а цена единицы товара Б снизилась на 18,9 %.

По формуле 11.3 рассчитаем индивидуальные индексы товарооборота, определив предварительно p_0Q_0 и p_1Q_1 (см. таблицу 11.1):

для А – $i_{pQ} = \frac{254{,}2}{200} = 1{,}271$ или $127{,}1\ \%\ (+27{,}1\ \%)$,

для Б – $i_{pQ} = \frac{744{,}0}{962} = 0{,}773$ или $77{,}3\ \%\ (-22{,}7\ \%)$.

То есть, в текущем периоде по сравнению с базисным товарооборот по товару А увеличился в 1,271 раза или на 27,1 %, а по товару Б – снизился на 22,7 %.

Таблица 11.1

Объем продаж и цены товаров

Товар	Физический объем реализации		Цена, тыс. грн.		p_0Q_0, тыс. грн.	p_1Q_1, тыс. грн.	p_0Q_1, тыс. грн.
	базисный Q_0	текущий Q_1	базисная p_0	текущая p_1			
А (т)	80	82	2,5	3,1	200,0	254,2	205,0
Б (шт.)	130	124	7,4	6,0	962,0	744,0	917,6
Итого	-	-	-	-	1162,0	998,2	1122,6

Проверим взаимосвязи между исчисленными индексами.

Так как $p \cdot Q = pQ, \Rightarrow i_p \cdot i_Q = i_{pQ}.$

Действительно:

для А — $1{,}240 \cdot 1{,}025 = 1{,}271$

для Б — $0{,}811 \cdot 0{,}954 \approx 0{,}773$

Рассчитаем сводный индекс товарооборота по формуле 11.7:

$$I_{pQ} = \frac{998{,}2}{1162} = 0{,}859 \text{ или } 85{,}9\ \% \ (-14{,}1\ \%).$$

Можно сформулировать вывод, что в текущем периоде по сравнению с базисным общий по двум товарам товарооборот уменьшился на 14,1 %.

Предпосылкой расчета всех индивидуальных индексов, а также сводных индексов, построенных выше, является сопоставимость измерения числителя и знаменателя.

При агрегировании (соединении) совокупности элементов, которые имеют различные единицы измерения, а также для построения сводных индексов качественных показателей используют специальные методологические принципы построения сводных индексов.

11.2. АГРЕГАТНАЯ ФОРМА – ОСНОВНАЯ ФОРМА СВОДНЫХ ИНДЕКСОВ

Социально-экономические явления могут быть соизмеримыми, то есть иметь общую меру, и несоизмеримыми.

Для построения сводных индексов физического объема разнородной продукции необходимо первоначально физические объемы привести к общей мере с помощью коэффициентов соизмерения. Единство различных видов продукции состоит в том, что они являются продуктами труда, а значит, имеют определенную себестоимость, трудоемкость и цену. Именно эти качественные показатели выступают как соизмерители, веса. В сводных индексах при этом используют так называемые **агрегаты**: $\sum Qp$ или $\sum Qz$, или $\sum Qt$.

При построении сводных индексов качественных показателей также возникает проблема: суммировать качественные показатели не имеет смысла. Поэтому при построении этих индексов используют веса. Весами для качественных показателей выступают объемные показатели.

Так как и в индексах объемных, и в индексах качественных показателей используется процесс агрегирования, форма индексов получила название агрегатной.

Правила построения агрегатной формы индексов:

- индексируемая величина рассматривается не изолированно, а во взаимосвязи с другой величиной, называемой **весами**;
- веса фиксируются в числителе и в знаменателе на одном уровне.

В статистике используют **две равноправные индексные системы**:

Ласпейреса:

$$I_p = \frac{\sum p_1 Q_0}{\sum p_0 Q_0}, \tag{11.11}$$

$$I_Q = \frac{\sum Q_1 p_0}{\sum Q_0 p_0}, \tag{11.12}$$

Пааше:

$$I_p = \frac{\sum p_1 Q_1}{\sum p_0 Q_1}, \tag{11.13}$$

$$I_Q = \frac{\sum Q_1 p_1}{\sum Q_0 p_1}. \tag{11.14}$$

Вышеописанные индексы увязывают, объединяют в **систему взаимосвязанных индексов**: $I_p \cdot I_Q = I_{pQ}$.

Данная связь между индексами обеспечивается, если индексы-сомножители имеют веса разных периодов. Во многих статистических исследованиях, например, в статистике внешней торговли, весами в индексах качественных показателей служат объемные факторы, зафиксированные на уровне отчетного периода (формула 11.13), а в индексах объемных показателей соизмерителями являются качественные факторы базисного уровня (формула 11.12). В этом случае система взаимосвязанных индексов имеет вид:

$$\frac{\sum p_1Q_1}{\sum p_0Q_1}\cdot\frac{\sum Q_1p_0}{\sum Q_0p_0}=\frac{\sum p_1Q_1}{\sum p_0Q_0} \qquad (11.15)$$

Заметим при этом, что агрегаты $\sum p_0Q_0$ и $\sum p_1Q_1$ являются фактическими величинами, которые характеризуют товарооборот соответственно базисного и отчетного периодов, а агрегат $\sum p_0Q_1$ – это условная величина, которая показывает, каким был бы товарооборот в отчетном периоде, если бы цены остались на уровне базисного периода.

Данная система взвешивания обеспечивает возможность решения следующей задачи: **расчета суммы экономии (дополнительных расходов) в результате изменения качественного показателя**.

Разница между фактической и условной величиной сводного индекса качественного показателя является характеристикой суммы экономии или дополнительных затрат за счет этого качественного показателя ($\text{Э}\,(\text{д. р.})_p$).

Например, на основе сводного индекса цен (формула 11.3):

$$\text{Э}\,(\text{д. р.})_p=\sum p_1Q_1-\sum p_0Q_1. \qquad (11.16)$$

По данным таблицы 11.1 рассчитаем сводные индексы физического объема реализации и цен.

На основе формулы 11.12 запишем:

$$I_Q=\frac{1122{,}6}{1162{,}0}=0{,}966 \text{ или } 96{,}6\ \% \ (-3{,}4\ \%).$$

Таким образом, в текущем периоде по сравнению с базисным в целом по двум товарам физический объем их реализации снизился на 3,4 %.

По формуле 11.13 определим сводный индекс цен:

$$I_p=\frac{998{,}2}{1122{,}6}=0{,}889 \text{ или } 88{,}9\ \% \ (-10{,}1\ \%).$$

Рассчитаем экономию (дополнительные расходы) покупателей от изменения цен, воспользовавшись формулой 11.16:

$$\text{Э}\,(\text{д. р.})_p=998{,}2-1122{,}6=-124{,}4 \text{ тыс. грн.}$$

Итак, в текущем периоде по сравнению с базисным в целом по двум товарам цены на них снизились на 10,1 %, вследствие чего покупатели сэкономили 124,4 тыс. грн.

Проверим взаимосвязь между индексами, описанную формулой 11.15:

$$0{,}889 \cdot 0{,}966 = 0{,}859\ .$$

Подход к построению агрегатной формы индексов других качественных и объемных показателей аналогичен.

11.3. СРЕДНЕВЗВЕШЕННЫЕ ИНДЕКСЫ

Агрегатная форма – это основная форма сводных индексов. Но на практике при построении сводных индексов, например, потребительских цен, физического объема ВВП, сталкиваются с проблемой отсутствия информации о ценах или физическом объеме всех единиц совокупности. Это связано с тем, что сплошное наблюдение – трудоемкий и дорогостоящий процесс. Использование выборочного наблюдения приводит к необходимости использования другой формы сводного индекса – **средневзвешенной**. При этом **основной формой является агрегатная**, так как:

- в агрегатной форме индекса непосредственно ясен смысл числителя, знаменателя и индекса в целом;
- средняя форма индекса получается путем преобразования агрегатной.

Преобразование агрегатного индекса в средний осуществляется следующим образом: либо в числителе либо в знаменателе агрегатного индекса индексируемую величину заменяют ее выражением через индивидуальный индекс. Если замену осуществляют в числителе, получают среднюю арифметическую взвешенную форму сводного индекса, если в знаменателе – среднюю гармоническую взвешенную форму индекса. Замену в агрегатном индексе осуществляют в той его части, где находится условная величина.

Преобразуем агрегатную форму сводного индекса себестоимости единицы продукции в среднюю. Замену произведем в знаменателе, так как именно там находится условная величина. Индексируемый показатель в знаменателе представлен своим базисным уровнем. Выражаем его через индивидуальный индекс: $z_0 = \frac{z_1}{i_z}$. Таким образом, получаем:

$$I_z = \frac{\sum z_1 Q_1}{\sum z_0 Q_1} = \frac{\sum z_1 Q_1}{\sum \frac{z_1 Q_1}{i_z}} . \quad (11.17)$$

Формула 11.17 – это формула среднего гармонического взвешенного индекса себестоимости единицы продукции.

Осуществим соответствующее преобразование сводного индекса физического объема производства. Условная величина находится в числителе индекса, индексируемой величиной является Q. Заменяем Q_1 его выражением через индивидуальный индекс: $Q_1 = i_Q Q_0$. Получаем следующую формулу среднего арифметического взвешенного индекса физического объема:

$$I_Q = \frac{\sum Q_1 z_0}{\sum Q_0 z_0} = \frac{\sum i_Q Q_0 z_0}{\sum Q_0 z_0}. \quad (11.18)$$

По данным таблицы 11.2 рассчитаем сводный индекс цен. Используем среднюю форму индекса, которую получим путем преобразования агрегатной.

Таблица 11.2

Производство товаров народного потребления в регионе в 2010 г.

Товары	Стоимость в фактических ценах, млн. грн. p_1Q_1	Изменение цен в 2010г., %	i_p	$\frac{p_1Q_1}{i_p}$, млн. грн.
Продовольственные	25,7	+ 19	1,19	23,6
Непродовольственные	17,2	- 2	0,98	17,6
Итого	42,9	-	-	41,2

Промежуточные расчетные данные занесем в таблицу и получим:

$$I_p = \frac{\sum p_1 Q_1}{\sum p_0 Q_1} = \frac{\sum p_1 Q_1}{\sum \frac{p_1 Q_1}{i_p}} = \frac{42,9}{41,2} = 1,041 \text{ или } 104,1\ \% \ (+4,1\ ).$$

Итак, в 2010 г. по сравнению с 2009 г. в целом по товарам народного потребления цены на них увеличились в 1,041 раза или на 4,1 %.

11.4. ИНДЕКСНЫЙ МЕТОД АНАЛИЗА ВЛИЯНИЯ ФАКТОРОВ

Если результативный признак представляет собой произведение двух и более других признаков, они могут рассматриваться как факторы, от величины которых функционально зависит результативный показатель.

Ставятся и решаются такие две задачи:

- рассчитать абсолютный прирост результативного признака за счет факторов-сомножителей, то есть определить **факторные абсолютные приросты**;
- рассчитать относительный прирост результативного признака за счет факторов-сомножителей, то есть определить **факторные относительные приросты.**

Роль каждого отдельного фактора в относительном или абсолютном изменении агрегата определяется в рамках системы взаимосвязанных индексов (формула 11.18).

Итак, **абсолютный прирост результативного показателя за счет каждого из факторов-сомножителей** рассчитывается как разница между числителем и знаменателем соответствующего индекса.

Например, **абсолютный прирост товарооборота в целом по нескольким товарам (общий прирост)**:

$$\Delta\sum pQ = \sum p_1Q_1 - \sum p_0Q_0 \qquad (11.19)$$

раскладывается на две составляющие:

за счет изменения цен (факторный прирост):

$$\Delta\sum pQ_p = \sum p_1Q_1 - \sum p_0Q_1, \qquad (11.20)$$

за счет изменения физического объема продаж (факторный прирост):

$$\Delta\sum pQ_Q = \sum p_0Q_1 - \sum p_0Q_0. \qquad (11.21)$$

При этом соблюдается равенство:

$$\Delta\sum pQ = \Delta\sum pQ_p + \Delta\sum pQ_Q\ . \qquad (11.22)$$

Рассчитаем абсолютные приросты, общий и факторные, по данным таблицы 11.1.

На основе формулы 11.19 запишем:

$$\Delta \sum pQ = 998{,}2 - 1162{,}0 = -163{,}8 \text{ тыс. грн.}$$

На основе формул 11.20 и 11.21:

$$\Delta \sum pQ_p = 998{,}2 - 1122{,}6 = -124{,}4 \text{ тыс. грн.}$$

$$\Delta \sum pQ_Q = 1122{,}6 - 1162{,}0 = -39{,}4 \text{ тыс. грн.}$$

Осуществим арифметический контроль: $-124{,}4 - 39{,}4 = -163{,}8$.

Таким образом, в текущем периоде по сравнению с базисным в целом по двум товарам товарооборот снизился на 163,8 тыс. грн. В том числе в результате изменения цен на товары товарооборот уменьшился на 124,4 тыс. грн., а за счет изменения физического объема продаж – на 39,4 тыс. грн.

Формулы расчета общего и факторных абсолютных приростов для одного товара будут иметь вид:

$$\Delta pQ = p_1 Q_1 - p_0 Q_0 \qquad (11.23)$$

$$\Delta pQ_p = p_1 Q_1 - p_0 Q_1 = (p_1 - p_0) \cdot Q_1, \qquad (11.24)$$

$$\Delta \sum pQ_Q = \sum p_0 Q_1 - \sum p_0 Q_0 = (Q_1 - Q_0) \cdot p_0. \qquad (11.25)$$

Итак, **абсолютный прирост результативного показателя за счет качественного фактора** равен произведению абсолютного прироста самого качественного фактора и объемного фактора, зафиксированного на отчетном уровне.

Абсолютный прирост результативного показателя за счет объемного фактора равен произведению абсолютного прироста самого объемного фактора и качественного фактора, зафиксированного на базисном уровне.

Произведем соответствующие расчеты (данные таблицы 11.1) по формулам 11.23−11.25 для товара А:

$$\Delta pQ = 254{,}2 - 200{,}0 = 54{,}2 \text{ тыс. грн.},$$

$$\Delta pQ_p = (3{,}1 - 2{,}5) \cdot 82 = 49{,}2 \text{ тыс. грн.},$$

$$\Delta \sum pQ_Q = (82 - 80) \cdot 2{,}5 = 5 \text{ тыс. грн.}$$

Осуществим арифметический контроль: $49{,}2 + 5 = 54{,}2$.

Таким образом, в текущем периоде по сравнению с базисным товарооборот по товару А увеличился на 53,2 тыс. грн. В том числе за счет роста цены товара А товарооборот возрос на 49,2 тыс. грн., а в результате увеличения физического объема продаж – на 5 тыс. грн.

Относительный прирост результативного показателя и его факторные относительные приросты рассчитываются путем деления соответствующих абсолютных приростов (общего и факторных) на базисный уровень результативного показателя.

Для группы товаров формулы будут иметь вид:

относительный прирост товарооборота в целом по нескольким товарам:

$$\Delta' \sum pQ = \frac{\sum p_1Q_1 - \sum p_0Q_0}{\sum p_0Q_0} = I_{pQ} - 1, \qquad (11.26)$$

относительный прирост товарооборота за счет изменения цен:

$$\Delta' \sum pQ_p = \frac{\sum p_1Q_1 - \sum p_0Q_1}{\sum p_0Q_0} = I_{pQ} - I_Q \ , \qquad (11.27)$$

относительный прирост товарооборота за счет изменения физического объема продаж:

$$\Delta' \sum pQ_Q = \frac{\sum p_0Q_1 - \sum p_0Q_0}{\sum p_0Q_0} = \ I_Q - 1. \qquad (11.28)$$

При этом соблюдается равенство:

$$\Delta' \sum pQ = \Delta' \sum pQ_p + \ \Delta' \sum pQ_Q. \qquad (11.29)$$

Таким образом, мы видим, что **относительный прирост результативного показателя за счет изменения качественного фактора** равен разности индексов результативного и объемного показателей.

Относительный прирост результативного показателя за счет изменения объемного фактора равен относительному приросту самого объемного фактора.

По данным таблицы 11.1 определим относительные приросты общего по двум товарам товарооборота, воспользовавшись формулами 11.26−11.28:

$$\Delta' \sum pQ = 0{,}859 - 1 = -0{,}141 \text{ или} - 14{,}1\ \%,$$

$$\Delta' \sum pQ_p = 0{,}859 - 0{,}966 = -0{,}107 \text{ или} - 10{,}7\ \%,$$

$$\Delta' \sum pQ_Q = 0{,}966 - 1 = -0{,}034 \text{ или} - 3{,}4\ \%.$$

Произведем арифметический контроль: $-10{,}7 - 3{,}4 = -14{,}1$.

Итак, в текущем периоде по сравнению с базисным в целом по товарам А и Б товарооборот снизился на 14,1 %. В том числе в результате изменения цен на товары товарооборот уменьшился на 10,7 %, а за счет изменения количества проданных товаров – на 3,4 %.

Относительный прирост результативного показателя и его факторные относительные приросты для одного товара можно рассчитать по следующим формулам:

$$\Delta' pQ = \frac{p_1Q_1 - p_0Q_0}{p_0Q_0} = i_{pQ} - 1, \tag{11.30}$$

$$\Delta' pQ_p = \frac{p_1Q_1 - p_0Q_1}{p_0Q_0} = i_{pQ} - i_Q, \tag{11.31}$$

$$\Delta' pQ_Q = \frac{p_0Q_1 - p_0Q_0}{p_0Q_0} = i_Q - 1. \tag{11.32}$$

Используем данные формулы и рассчитаем относительные приросты – общий и факторные – для товара А (таблица11.1):

$$\Delta' pQ = 1{,}271 - 1 = 0{,}271 \text{ или } 27{,}1\ \%,$$

$$\Delta' pQ_p = 1{,}271 - 1{,}025 = 0{,}246 \text{ или } 24{,}6\ \%,$$

$$\Delta' pQ_Q = 1{,}025 - 1 = 0{,}025 \text{ или } 2{,}5\ \%.$$

Сделаем проверку: $24{,}6 + 2{,}5 = 27{,}1$.

Таким образом, в текущем периоде по сравнению с базисным товарооборот по товару А увеличился на 27,1 %, в том числе: за счет роста цены на данный товар товарооборот возрос на 24,6 %, а в результате увеличения физического объема продаж – на 2,5 %.

11.5. АНАЛИЗ ДИНАМИКИ СРЕДНЕГО УРОВНЯ КАЧЕСТВЕННОГО ПОКАЗАТЕЛЯ

Для нескольких предприятий, выпускающих или реализующих однородную продукцию, можно определить среднюю по предприятиям себестоимость единицы продукции, среднюю трудоемкость единицы продукции, среднюю цену. Расчет средней величины производится по формуле арифметической взвешенной. Для перечисленных показателей формулы запишутся:

$$\bar{z} = \frac{\sum zQ}{\sum Q}, \tag{11.33}$$

$$\bar{t} = \frac{\sum tQ}{\sum Q}, \tag{11.34}$$

$$\bar{p} = \frac{\sum pQ}{\sum Q}. \tag{11.35}$$

Если в приведенных формулах используются не частоты, а частости, то формула средней себестоимости единицы продукции, например, будет иметь вид:

$$\bar{z} = \sum z d_Q, \tag{11.36}$$

где $d_Q = \frac{Q}{\sum Q}$ - доля отдельных предприятий в общем выпуске продукции.

На основе формулы 11.36 можно сделать вывод, что **факторами относительного изменения средней себестоимости единицы продукции являются**:

- себестоимость единицы продукции на каждом предприятии (z);
- доля предприятий в общем выпуске продукции или, другими словами, структурные сдвиги выпуска продукции ($\frac{Q}{\sum Q}$).

Задачей индексного анализа динамики среднего уровня качественного показателя является определение относительного изменения среднего уровня: общее и под влиянием изменения факторов. Данная задача решается с помощью построения **системы индексов**:

- индекса переменного состава;
- индекса постоянного (фиксированного) состава;
- индекса структурных сдвигов.

Индекс переменного состава ($I^{\text{п.с.}}$) характеризует общее относительное изменение среднего уровня качественного показателя под влиянием всех факторов. Запишем формулу данного индекса для анализа динамики средней цены:

$$I_{\bar{p}}^{\text{п.с}} = \bar{p}_1 \div \bar{p}_0 = \frac{\sum p_1 Q_1}{\sum Q_1} \div \frac{\sum p_0 Q_0}{\sum Q_0}. \qquad (11.37)$$

В таблице 11.3 представлены данные об изменении цены и количества реализованного товара А в двух регионах. Определим среднюю по двум регионам цену товара А в июне и в июле (промежуточные расчеты в таблице 11.3) и рассчитаем по формуле 11.37 индекс средней цены переменного состава:

$$I_{\bar{p}}^{\text{п.с}} = \frac{405}{27} \div \frac{460}{30} = 15{,}00 \div 15{,}33 = 0{,}978 \text{ или } 97{,}8\ \% \ (-2{,}2\ \%).$$

То есть в июле по сравнению с июнем средняя по двум регионам цена единицы товара А снизилась на 2,2 %.

Таблица 11.3

К расчету индексов динамики
среднего уровня качественного показателя

Регион	Июнь		Июль		p_0Q_0, тыс. грн.	p_1Q_1, тыс. грн.	p_0Q_1, тыс. грн.
	цена, грн. p_0	количество проданного товара А, тыс. шт. Q_0	цена, грн. p_1	количество проданного товара А, тыс. шт. Q_1			
1	12	10	13	18	120	234	216
2	17	20	19	9	340	171	153
Итого	-	30	-	27	460	405	369

Обратим внимание, что и в первом, и во втором регионе цена товара А в июле по сравнению с июнем увеличилась, и одновременно с этим средняя цена уменьшилась. Очевидно, что снижению средней цены способствовал второй фактор, фактор структуры. Мы видим, что в июле выросла доля региона, в котором цена низкая, а доля региона, где цена значительно выше, упала.

Для количественной оценки влияния факторов на изменение средней цены строят индексы фиксированного состава и структурных сдвигов. В основе их построения лежит допущение, что первым изменяется фактор структуры, а затем – качественный фактор. Для правильной записи формул удобно пользоваться следующей схемой:

$$\bar{p}_0 = \frac{\sum p_0 Q_o}{\sum Q_o} \xrightarrow{\frac{Q}{\sum Q}} \overline{p'} = \frac{\sum p_0 Q_1}{\sum Q_1} \xrightarrow{p} \bar{p}_1 = \frac{\sum p_1 Q_1}{\sum Q_1}$$

Результатом сравнения средней цены отчетного периода и базисного, как мы видим, является индекс переменного состава (формула 11.37).

Результатом сравнения средней цены отчетного периода и условной средней цены ($\overline{p'}$) является индекс постоянного (фиксированного) состава ($I_{\bar{p}}^{\text{ф.с}}$):

$$I_{\bar{p}}^{\text{ф.с}} = \bar{p}_1 \div \overline{p'} = \frac{\sum p_1 Q_1}{\sum Q_1} \div \frac{\sum p_0 Q_1}{\sum Q_1}. \tag{11.38}$$

Данный индекс характеризует относительное изменение среднего уровня качественного показателя только под влиянием изменения самого качественного фактора.

Результатом сравнения условной средней цены и средней цены базисного периода является индекс структурных сдвигов ($I_{\bar{p}}^{\text{с.с}}$):

$$I_{\bar{p}}^{\text{с.с}} = \overline{p'} \div \bar{p}_0 = \frac{\sum p_0 Q_1}{\sum Q_1} \div \frac{\sum p_0 Q_0}{\sum Q_0}. \tag{11.39}$$

Индекс структурных сдвигов оценивает относительное изменение среднего уровня качественного показателя за счет изменения структуры совокупности.

Между индексами переменного состава, фиксированного состава и индексом структурных сдвигов существует следующая взаимосвязь:

$$I_{\bar{p}}^{\text{п.с}} = I_{\bar{p}}^{\text{ф.с}} \cdot I_{\bar{p}}^{\text{с.с}}.$$

Рассчитаем индексы фиксированного состава и структурных сдвигов по формулам 11.38 и 11.39 для данных таблицы 11.3:

$$I_{\bar{p}}^{\text{ф.с}} = 15 \div \frac{369}{27} = 15 \div 13{,}67 = 1{,}097 \text{ или } 109{,}7\ \% \ (+9{,}7).$$

Итак, в текущем периоде по сравнению с базисным средняя по двум регионам цена товара А увеличилась в 1,097 раза или на 9,7 % только за счет изменения уровней цен по регионам.

$$I_{\bar{p}}^{\text{с.с}} = 13{,}67 \div 15{,}33 = 0{,}892 \text{ или } 89{,}2\ \% \ (-10{,}8\ \%).$$

Как мы и предположили, изменение структуры продаж снизило среднюю цену товара А. В текущем периоде по сравнению с базисным в результате изменения структуры продаж средняя по двум регионам цена товара А снизилась на 10,8 %.

Подход к анализу динамики среднего уровня других качественных показателей аналогичен.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

1. Дайте определение статистического индекса.
2. Какие функции выполняет индекс?
3. Назовите виды индексов по степени охвата единиц совокупности.
4. Что представляет собой индексируемая величина и какие ее виды вы знаете?
5. Какая система обозначений принята в статистике система обозначений при использовании индексного метода?
6. Какие взаимосвязи существуют между индексами?
7. Что значит соизмеримые и несоизмеримые явления?
8. Какая проблема возникает при построении сводных индексов физического объема для несопоставимой продукции?
9. Сформулируйте правила построения агрегатной формы индексов.
10. В чем отличие систем индексов Ласпейреса и Пааше?

11. Какая система взвешивания обеспечивает возможность оценки суммы экономии (дополнительных расходов) под влиянием изменения качественного показателя?
12. Поясните, почему агрегатная форма сводных индексов является основной?
13. Сформулируйте принципы предобразования агрегатной формы сводного индекса в среднюю форму.
14. Приведите примеры средней арифметической взвешенной формы сводного индекса.
15. Запишите примеры средней гармонической взвешенной формы сводного индекса.
16. Как рассчитывается абсолютный прирост результативного показателя под влиянием факторов-сомножителей для одного и для нескольких видов продукции?
17. Приведите формулы расчета относительного прироста результативного показателя за счет изменения факторов-сомножителей для одного и для нескольких видов продукции.
18. Какая система индексов используется для анализа динамики среднего уровня качественного показателя?
19. Что характеризует индекс переменного состава?
20. Запишите формулу и поясните суть индекса фиксированного состава.
21. Поясните содержание индекса структурных сдвигов.
22. Какая существует взаимосвязь между индексами переменного состава, фиксированного состава и индексом структурных сдвигов?

СПИСОК РЕКОМЕНДОВАННОЙ ЛИТЕРАТУРЫ

1. Кравець О.С. Статистика: навч. посіб. / О.С. Кравець – Одеса: Пальміра, 2008. – 387 с.
2. Лугінін О.Є. Статистика: підручник / О.Є. Лугінін, С.В. Білоусова – К. Центр навчальної літератури, 2005. – 580 с.
3. Матковський С.О. Теорія статистики: навч. посіб. / С.О. Матковський, О.Р. Марець – К.: Знання, 2010. – 534 с.
4. Методологія статистичного забезпечення розвитку регіону: монографія / [А.З. Підгорний, К.В. Вітковська, О.Г. Милашко та ін.]; під ред. А.З. Підгорного – Одеса: Атлант, 2012. – 303 с.
5. Моторин Р.М. Статистика для економістів: навч. посіб. / О.М. Моторин, Е.В. Чекотовський – К.: Знання, 2009. – 430 с. + ком пакт-диск.
6. Підгорний А.З. Теорія статистики: навчальний посібник / А.З.Підгорний – Одеса: ОДЕУ, 2001. – 150 с.
7. Общая теория статистики: учебник / [А.Я. Боярский, Л.Л. Викторова, А.М. Гольдберг и др.]; под ред. А.М. Гольдберга, В.С. Козлова. – М.: Финансы и статистика, 1985. – 367 с.
8. Статистика: підручник / [С.С. Герасименко, А.В. Головач, А.М. Эрина та ін.]. – К.: КНЕУ, 2002. – 389 с.
9. Статистика: підручник / [А.В.Головач, А.М.Єріна, О.В.Козирєв та ін.]. – К.: Вища шк., 1993. – 422 с.
10. Підгорний А. З. Міжнародна статистика : навчальний посібник / А. З. Підгорний, О. Г. Милашко, О. П. Русєва. – Одеса: ОНЕУ, 2012. – 162 с.
11. Уманець Т.В. Статистика: навч. посіб. / Т.В. Уманець, Ю.Б. Пігарєв – К.: Вікар, 2003. – 623 с.
12. Янковой О.Г. Моделювання парних зв’язків в економіці: навч. посіб. / О.Г. Янковой – Одеса: Оптимум, 2001. – 198 с.

А.З. Подгорный
О.Г. Мылашко
С.М. Киршо
Н.М. Шилофост

СТАТИСТИКА

Учебное пособие
для иностранных студентов

Підписано до друку 12.12.2012. Формат 60*84/16. Папір офсетний.
Гарнітура Times New Roman. Ум. друк. арк.. 11,27
Тираж 300 прим.

Друкарня «Атлант» ВОІ СОІУ
65029, м. Одеса, Ольгієвський узвіз, 8
Свідоцтво ДК №3564 від 31.08.2009
Тел.: 728-45-71
e-mail: ev_atlant@mail.ru