Министерство науки и высшего образования
Российской Федерации
Рубцовский индустриальный институт (филиал)
ФГБОУ ВО «Алтайский государственный технический
университет им. И.И. Ползунова»

О.А. Чиркова, Н.А. Ляпкина

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ
ЧАСТЬ 1

Учебное пособие для студентов всех форм обучения
экономических направлений подготовки

Рекомендовано Рубцовским индустриальным институтом (филиалом) ФГБОУ ВО «Алтайский государственный технический университет им. И.И. Ползунова» в качестве учебного пособия для студентов, обучающихся по направлению подготовки «Экономика»

Рубцовск 2021

ББК 60.6

Чиркова О.А., Ляпкина Н.А. Общая теория статистики: Часть 1: Учебное пособие по дисциплине «Статистика (теория статистики, социально – экономическая статистика)» для студентов всех форм обучения экономических направлений подготовки / Рубцовский индустриальный институт. – Рубцовск, 2021. – 84 с.

В методическом пособии изложены основные темы общей теории статистики. По каждой теме представлены: теоретический материал, вопросы для контроля теоретических знаний, примеры решения задач, задания для самостоятельной работы. Кроме того, приведена рекомендуемая литература.

Рассмотрено и одобрено на заседании НМС РИИ.
Протокол № 3 от 24.11.21.

Рецензент: к.э.н., доцент Э.С. Маршалов

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Цель преподавания дисциплины – сформировать у студентов экономических направлений подготовки комплекс знаний и умений в области общей теории статистики, основ социально-экономической статистики, ознакомить с технологиями и механизмом статистических расчетов и использования методов статистического анализа, а также сформировать навыки самостоятельного, творческого использования теоретических знаний в практической деятельности.

Для достижения цели преподавания дисциплины в процессе изучения курса необходимо решить следующие задачи: ознакомить студентов с базовыми понятиями общей теории статистики и социально-экономической статистики, методологическими и методическими основами проведения статистических расчетов и статистического анализа.

Для освоения дисциплины студенты должны обладать базовыми знаниями математики, экономической теории и других общественных наук в рамках курса средней школы.

Дисциплина «Статистика (теория статистики, социально-экономическая статистика)» формирует у студентов комплекс знаний, умений и навыков, необходимых для изучения дисциплин «Методы финансовых и коммерческих расчетов», «Математические модели в управлении», «Анализ финансово-хозяйственной деятельности», «Экономика предприятия» и т.д.

Тема 1. ПРЕДМЕТ, МЕТОД, ЗАДАЧИ СТАТИСТИКИ

1. Предмет статистики

В настоящее время термин «статистика» употребляется в трех значениях:

1. под *статистикой* понимают отрасль практической деятельности, которая имеет своей целью сбор, обработку, анализ и публикацию массовых данных о самых различных явлениях общественной жизни;

2. *статистикой* называют цифровой материал, служащий для характеристики какой-либо области общественных явлений или территориального распределения какого-то показателя;

3. *статистикой* называется отрасль знания, особая научная дисциплина.

Т.о., предмет статистики – размеры и количественные соотношения качественно определенных социально-экономических явлений, закономерности их связи и развития в конкретных условиях места и времени.

Свой предмет статистика изучает с помощью метода обобщающих показателей.

Теоретической основой статистики являются положения социально-экономической теории, которые рассматривают законы развития социально-экономического явления, выясняют их природу и значение в жизни общества.

2. Метод статистики

Для изучения предмета статистики разработаны и применяют специфические приемы, совокупность которых образует методологию статистики. Общей основой разработки и применения статистической методологии является диалектический метод познания, согласно которому общественные явления и процессы рассматриваются в развитии, взаимной связи и причинной обусловленности. При этом статистика опирается на такие диалектические категории, как количество и качество, необходимость и случайность, причинность и закономерность, единичное и массовое, индивидуальное и общее.

Статистические методы используются комплексно (системно). Это обусловлено сложностью процесса экономико-статистического исследования, состоящего из трех основных стадий:

а) сбор первичной информации. На этой стадии применяется метод массового статистического наблюдения;

б) статистическая сводка и обработка первичной информации. На данной стадии используют метод статистических группировок;

в) обобщение и интерпретация статистической информации. На этой стадии проводится анализ статистической информации на основе применения обобщающих статистических показателей.

При изучении статистической информации широкое применение имеют табличный и графический методы.

3. Категории статистики

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

Статистическая совокупность – совокупность социально-экономических объектов или явлений общественной жизни, объединенных некоей качественной основой, общей связью, но отличающихся друг от друга отдельными признаками. Пример: совокупность предприятий, фирм.

Совокупность называется однородной, если один или несколько существенных признаков ее объектов являются общими для всех единиц. Совокупность, в которую входят явления разного типа, считается разнородной.

Единица совокупности – первичный элемент статистической совокупности, являющийся носителем признаков, подлежащих регистрации и основой ведущегося при обследовании счета.

Признак – качественная особенность единицы совокупности.

Значение признака у отдельной единицы называют *вариантом*.

Статистический показатель – понятие, отражающее количественные характеристики соотношения признаков общественных явлений.

Система статистических показателей – совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями.

4. Задачи статистики

Основные задачи статистики:

1. всестороннее исследование происходящих в обществе глубоких преобразований экономических и социальных процессов на основе научно обоснованной системы показателей;
2. обобщение и прогнозирование тенденций народного хозяйства;
3. выявление имеющихся резервов эффективности общественного производства;
4. своевременное обеспечение надежной информацией законодательной власти, управленческих, исполнительных и хозяйственных органов, а также широкой общественности.

Вопросы для контроля теоретических знаний

1. От какого латинского слова происходит термин «статистика»? Что он означает?
2. Какие статистические работы проводились в древние и средние века?
3. К какому времени относится становление статистики как науки?
4. Почему статистика относится к общественным наукам? В чем ее отличие от других общественных наук?

5. Что такое закономерность? Статистическая закономерность и ее особенности.

6. Что такое совокупность, единица совокупности? Понятие вариации и признака.

7. В чем сущность и значение закона больших чисел для статистики?

8. Дайте определение предмета статистики.

9. Что является теоретической основой статистической науки?

10. Почему каждое статистическое исследование должно опираться на изучение всех относящихся к данному вопросу фактов?

11. Почему статистика изучает явления общественной жизни в движении, изменении и развитии?

12. Перечислите специфические методы, присущие статистическому исследованию.

13. Какие принципы и методы излагаются в общей теории статистики? Почему изучение статистической науки начинается с общей теории статистики?

14. Что определяет многообразие и сложность задач и функций статистики?

Тема 2. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

1. Понятие статистического наблюдения

Статистическая информация – первичный статистический материал о социально-экономических явлениях, формирующийся в процессе статистического наблюдения, который затем подвергается систематизации, сводке, анализу и обобщению.

Основные свойства статистической информации – массовость и стабильность.

Статистическим можно назвать только такое наблюдение, которое обеспечивает регистрацию устанавливаемых фактов в учетных документах для последующей обработки. Статистическое наблюдение должно быть обязательно массовым, систематическим, проводиться на научной основе по заранее разработанному плану и программе. В результате статистического наблюдения должна быть получена только объективная, сопоставимая и достаточно полная информация.

Процесс проведения статистического наблюдения включает следующие этапы: подготовка наблюдения; проведение массового сбора данных; подготовка данных к автоматизированной обработке; разработка предложений по совершенствованию статистического наблюдения.

2. Программно – методологические вопросы наблюдения

Любое статистическое исследование необходимо начинать с точной формулировки его цели и задач. После этого определяется объект и единица наблюдения, разрабатывается программа, выбираются вид и способ наблюдения.

Объект наблюдения – совокупность социально – экономических процессов и явлений, которые подлежат исследованию. Чтобы определить объект статистического наблюдения, необходимо установить границы изучаемой совокупности. Для этого следует указать важнейшие признаки, отличающие его от других сходных. Определяя объект наблюдения, необходимо точно указать единицу наблюдения.

Единица наблюдения – составная часть объекта наблюдения, которая служит основой счета и обладает признаками, подлежащими регистрации (при переписи населения единицами наблюдения являлся каждый отдельный человек).

Программа наблюдения – перечень вопросов, по которым собираются данные, либо перечень признаков и показателей, подлежащих регистрации.

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

Организационные вопросы статистического наблюдения включают в себя определение субъекта, места, времени, формы и способа наблюдения.

При организации статистического наблюдения должен быть решен вопрос о времени наблюдения. При этом устанавливается период, в течение которого будет проводиться наблюдение (срок наблюдения) и точно определяется время, к которому относятся регистрируемые сведения (объективное время наблюдения). Это может быть либо определенный момент, либо тот или иной период (сутки, месяц). Момент времени, к которому приурочены регистрируемые сведения, называются критическим моментом наблюдения.

3. Формы, виды и способы наблюдения

Формами статистического наблюдения являются отчетность и специально организованное наблюдение.

Отчетность – предусмотренная действующим законодательством форма организации статистического наблюдения за деятельностью предприятий, организаций, по которым органы государственной статистики получают информацию в виде установленных отчетных документов.

Специально организованное статистическое наблюдение – представляет собой сбор сведений посредством переписей, единовременных учетов и обследований.

Статистическое наблюдение подразделяется на виды:

1) по времени регистрации фактов:

1.1. непрерывное – отчетность, постоянная регистрация данных по мере их возникновения;

1.2. периодическое – регистрация данных по мере надобности;

2) по степени охвата единиц совокупности:

1.1. сплошное – регистрации подлежат все без исключения единицы изучаемой совокупности;

1.2. несплошное – обследованию подвергается часть единиц изучаемой совокупности, на основе которой можно получить обобщающую характеристику всей совокупности.

В свою очередь, *несплошное* наблюдение подразделяется на следующие виды:

1. наблюдение основного массива – обследованию подвергается та часть совокупности, у которой величина изучаемого признака является преобладающей во всем объеме;
2. монографическое наблюдение – применяется для детального изучения отдельных единиц совокупности;
3. выборочное наблюдение – отбор единиц совокупности осуществляется случайно.

Статистическое наблюдение проводится следующими способами:

- *непосредственное наблюдение* – работники органов статистики сами устанавливают факты и записывают их в формуляры;
- *документальное наблюдение* – данные берутся из документов;
- *опрос* – данные записываются со слов опрашиваемого;
- *анкетное наблюдение*, основано на принципе добровольного заполнения анкет (применяется, когда не требуется высокая точность данных).

4. Ошибки наблюдений

Всякое статистическое наблюдение ставит задачу получения таких данных, которые бы точнее отображали действительность. Точность и достоверность статистической информации – важная задача статистического наблюдения.

Под *точностью* информации понимают соответствие изучаемого показателя показателю, полученному при наблюдении. Отклонение между этими показателями называют *ошибками* наблюдения.

Выделяют ошибки регистрации и ошибки репрезентативности (представительности).

Ош*ибки регистрации* возникают вследствие неправильного установления фактов или неправильной их записи. Могут быть как при сплошном, так и несплошном наблюдении. Различают случайные и систематические ошибки регистрации.

Ошибки репрезентативности свойственны несплошному наблюдению. Возникают в результате нарушения правил отбора из изучаемой совокупности.

Для выявления и устранения ошибок могут применяться следующие виды контроля:

- *счетный* – проверка точности арифметических расчетов, применявшихся при составлении отчетов
- *логический* – проверка отчетов на вопросы программы наблюдения путем их логического осмысления.

Вопросы для теоретического контроля знаний

1. Что понимают под статистической информацией и ее распространением?
2. Для чего и кому нужна статистическая информация в современных условиях?
3. Каким образом осуществляется формирование информационной базы статистического исследования?
4. Дайте определение статистического наблюдения. В чем его сущность?
5. Охарактеризуйте этапы проведения статистического наблюдения.
6. Каковы цель наблюдения, объект, единица наблюдения, отчетная единица?
7. Дайте определение программы статистического наблюдения, перечислите требования, предъявляемые к программе статистического наблюдения.
8. Что такое статистический формуляр?
9. Назовите важнейшие организационные вопросы статистического наблюдения.
10. Перечислите основные формы, виды и способы статистического наблюдения.
11. Что понимается под точностью наблюдения?
12. Назовите основные ошибки наблюдения.
13. Какие виды контроля могут применяться для выявления ошибок наблюдения?

Задания для самостоятельной работы студентов

1. Составьте перечень наиболее существенных признаков следующих единицы статистического наблюдения:

а) фермерских хозяйств;
б) жилого дома (для жилищной переписи);
в) вуза.

2. Сформулируйте объект, единицу и цель наблюдения и разработайте программу:

а) обследования детских садов;
б) обследования автозаправочных станций.

3. С помощью логического контроля подвергните проверке следующие ответы на вопросы переписного листа переписи населения:

а) фамилия, имя, отчество – Иванова Ирина Петровна;
б) пол – мужской;
в) возраст – 5 лет;
г) состоит ли в браке в настоящее время – да;
д) национальность – русская;
е) родной язык – русский;
ж) образование – среднее специальное;

з) место работы – детский сад;
и) занятие по этому месту работы – мед. сестра.

Тема 3. СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКОГО МАТЕРИАЛА

1. Понятие статистической сводки

Статистическая сводка – научно организованная обработка материалов наблюдения. Она позволяет перейти к обобщающим показателям совокупности в целом и отдельных ее частей, осуществить анализ и прогнозирование изучаемых процессов.

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

Сложная сводка представляет собой комплекс операций, включающих группировку единиц наблюдения, подсчет итогов по каждой группе и по всему объекту и предоставление результатов группировки и сводки в виде статистических таблиц.

Статистические сводки различаются по ряду признаков: по сложности построения, месту проведения, способу обработки материалов статистического наблюдения.

Способ разработки статистической сводки может быть централизованным и децентрализованным.

По способу обработки материалов сводку делят на ручную и автоматизированную.

Проведению сводки предшествует разработка ее программы, которая состоит из следующих этапов:

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

2. Статистическая группировка

В результате сводки статистического материала отдельные единицы статистической совокупности объединяются в группы при помощи метода группировок.

Статистическая группировка – процесс образования однородных групп на основе расчленения статистической совокупности на части или объеденения изучаемых единиц в частные совокупности по определенному признаку (группировка промышленных предприятий по формам собственности).

Особым видом группировок является классификация. *Классификация* – устойчивая номенклатура классов и групп, образованных на основе сходства и различия единиц изучаемого объекта.

Группировка создает основу для последующей сводки и анализа данных.

Для решения этих задач применяют три вида группировок.

- Типологическая группировка.
- Структурная группировка.
- Аналитическоя группировка.

Всю совокупность признаков можно разделить на две группы: факторные и результативные.

Группировка, в которой группы образованы по одному признаку, назывется *простой. Сложной* называется группировка, в которой разделение совокупности на группы производится по двум и более признакам, взятым в сочетании (комбинации).

3. Этапы построения статистической группировки

Группировка статистических материалов осуществляется в несколько этапов.

1. Выбор группировочного признака.
2. Распределение единиц совокупности по группам.
3. Определение величины интервала.

Интервал – промежуток между двумя значениями количественного признака, в пределах которого все значения признака относятся к данной группе. Величина интервала зависит от количества групп (чем больше групп, тем меньше интервал). Интервалы бывают равными и неравными. Равные интервалы имеют одинаковые границы во всех группах.

Для определения количества групп группировки с равными интервалами может быть использована формула (*Стерджесса):*

$$n = 1+3.322\lg N, \tag{3.1}$$

где N – численность всей совокупности.

Для определения величины интервала при известном числе групп используется формула:

$$i = \frac{x_{\max} - x_{\min}}{n}, \tag{3.2}$$

где n – количество групп;

x_{min}, x_{max} – наибольшее и наименьшее значение признака.

4. Статистические ряды распределения

Статистические ряды распределения – упорядоченное расположение единиц изучаемой совокупности на группы по группировочному признаку.

Ряды распределения, образованные по качественному признаку, называются *атрибутивными*.

При группировке ряда по количественному признаку получают *вариационные ряды*. Вариационные ряды состоят из двух элементов.

Варианта – отдельное значение варьируемого признака, которое он принимает в ряду распределения.

Частота – численность отдельных вариант или каждой группы вариационного ряда.

Частоты, выраженные в долях единицы или в процентах к итогу, называются *частостями*. Сумма частот составляет объем ряда распределения и равна 1 или 100%.

В практике исследований часто встречается понятие накопленных частот, которое определяется путем суммирования к частоте данной группы показателей предшествующих групп.

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

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

5. Статистические графики

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

Основные элементы графика:

- графический образ
- поле графика
- пространственные ориентиры графика
- масштабные ориентиры
- масштабная шкала
- носитель шкалы
- экспликация
- абсцисса
- ордината

Классификация видов графиков:

Существует множество видов графических изображений. Их классификация основана на ряде признаков:

а) способ построения графического образа;

б) геометрические знаки, изображающие статистические показатели и отношения;

в) задачи, решаемые с помощью графического изображения.

Линейные	Плоскостные	Объемные
- статистические кривые	- столбиковые - полосовые - квадратные - фоновые - круговые - секторные - фигурные - точечные	- поверхности - распределения

Рис. 3.1. Классификация статистических графиков
по форме графического образа

Диаграммы	Статистические карты
- сравнения - динамики - структурные	- картограммы - картодиаграммы

Рис. 3.2. Классификация статистических графиков
по способу построения и задачам изображения

6. Статистические таблицы

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

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

Сказуемое статистической таблицы образует система показателей, которыми характеризуется объект изучения, т.е. подлежащее таблицы.

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

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

Основные приемы, определяющие технику формирования статистических таблиц, следующие:

1. Таблица должна быть компактной и содержать только те исходные данные, которые непосредственно отражают исследуемое явление в статике и динамике и необходимы для познания его сущности.

2. Заголовок таблицы и названия граф и строк должны быть четкими, лаконичными, представлять собой законченное целое, органично вписываться в содержание текста.

3. Информация, располагаемая в столбцах (графах) таблицы, завершается итоговой строкой.

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

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

6. Графы и строки полезно нумеровать. Графы, слева заполненные названием строк, принято обозначать заглавными буквами алфавита (А, В) и т.д., а все последующие графы – номерами в порядке возрастания.

7. Взаимосвязанные и взаимозависимые данные, характеризующие одну из сторон анализируемого явления, целесообразно располагать в соседних друг с другом графах.

8. Графы и строки должны содержать единицы измерения, соответствующие поставленным в подлежащем и сказуемом показателям. При этом используются общепринятые сокращения единиц измерения.

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

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

11. По возможности числа целесообразно округлять.

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

- Если данная позиция (на пересечении соответствующих графы и строки) вообще не подлежит заполнению, то ставится знак «Х»;
- Когда по какой-либо причине отсутствую сведения, то ставится многоточие «…», или «Нет свед.», или «Н. св.»;
- При отсутствии явления клетка заполняется тире («-«) или остается пустой;
- Для отображения очень маленьких чисел используют обозначения (0,0) или (0,00), предполагающие возможность наличия числа.

13. В случае необходимости дополнительной информации – разъяснений – к таблице могут даваться примечания.

Вопросы для теоретического контроля знаний

1. Назовите задачи сводки и ее содержание.
2. В чем суть метода группировки и его место в системе статистических методов?
3. Перечислите виды статистических группировок.

4. Укажите принципы построения статистических группировок и классификаций.
5. Дайте определение рядам распределения.
6. В чем заключается сравнимость статистических группировок?
7. Что представляют собой многомерные классификации?
8. Укажите место группировок и классификаций в практике статистики.
9. Дайте определение статистической таблицы.
10. Перечислите элементы статистической таблицы.
11. Назовите виды таблиц по способу разработки подлежащего.
12. Назовите виды таблиц по способу разработки сказуемого.
13. Определите основные правила построения таблиц.
14. В чем заключается чтение и анализ таблиц?
15. Дайте определение статистического графика.
16. Укажите основные элементы статистического графика.
17. Охарактеризуйте классификацию видов графиков.

Примеры решения задач

1. Требуется произвести группировку с равными интервалами по данным об уровне месячной заработной платы бюджетных работников, которая колеблется в пределах от 600 до 750 руб., и необходимо при этом выделить 5 групп. Величина интервала, руб.:

$$i = \frac{x_{\max} - x_{\min}}{n} = \frac{750 - 600}{5} = 30.$$

Прибавляя к минимальному значению признака (в данном случае 600 руб.) найденное значение интервала, получаем верхнюю границу первой группы: 600 + 30 = 630.

Прибавляя далее величину интервала к верхней границе первой группы, получаем верхнюю границу второй группы: 630 + 30 = 660 и т.д.

В результате получим следующие группы работников по размеру заработной платы, руб.:

600-630; 630-660; 660-690; 690-720; 720-750.

2. В таблицах 3.1, 3.2 определить подлежащее и сказуемое.

Таблица 3.1

Котировка облигаций государственного сберегательного займа на 05.03.2013 г.

Облигации государственного сберегательного займа	Объем покупки	Объем продаж
	482,70	469,65

Подлежащее: облигации государственного сберегательного займа.

Сказуемое: объем покупки, объем продаж.

Таблица 3.2

Распределение клиентов страховых компаний по категориям и страховым суммам в 1 квартале 2021 г.

Страховая компания	Всего клиентов, чел.	В том числе распределение клиентов по категориям и страховым суммам на одного застрахованного					
		Руководители коммерческих структур		Сотрудники предприятий, работающие в офисе		Охранники, милиционеры, инкассаторы	
		20 – 50 тыс. руб.	Свыше 50 тыс. руб.	20 – 50 тыс. руб.	Свыше 50 тыс. руб.	20 – 50 тыс. руб.	Свыше 50 тыс. руб.
1	444	195	180	13	12	23	21
2	390	150	180	12	15	15	18
3	595	210	300	26	10	21	28

Подлежащее: страховая компания, численность клиентов.

Сказуемое: категории застрахованных лиц и страховая сумма.

Задания для самостоятельной работы

1. Имеются следующие данные об успеваемости студентов группы по теории статистики в летнюю сессию: 5,4,4,4,3,2,5,3,4,4,4,3,2,5,2,5,5,2,3,3. Постройте:

а) ряд распределения студентов по баллам оценок, полученных в сессию;

б) ряд распределения студентов, по уровню успеваемости, выделив в нем две группы студентов: неуспевающие (2 балла), успевающие (3 балла).

По данным о численности населения за 2015-2020 гг. (млн. чел.) постройте столбиковые диаграммы.

2015	2016	2017	2018	2019	2020
117,5	130,1	137,6	147,4	148,0	148,0

2. Изобразить с помощью секторной диаграммы графические данные о структуре ВВП по видам первичных доходов РФ за 2019 и 2020 гг.: оплата наемных работников – 41,0%; валовая прибыль экономики и валовые смешанные доходы – 50,0; чистые налоги на производство и импорт – 9%; в 2020 г. соответственно – 33,0; 57,0; 10,0%.

Тема 4. СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

1. Виды и значение обобщающих статистических показателей

Статистические показатели имеют взаимосвязанные качественную и количественную стороны. Качественная сторона показателя отражается в его содержании безотносительно к конкретному размеру признака. Количественная сторона статистического показателя – его числовое значение.

Статистические показатели можно подразделить на группы по следующим признакам:

1) по сущности изучаемых явлений.

2) по степени агрегирования явлений.

3) в зависимости от характера изучаемых явлений.

Обобщающие показатели могут быть выражены абсолютными, относительными и средними величинами.

2. Абсолютные статистические величины

Абсолютными в статистике называют суммарные обобщающие показатели, характеризующие размеры (уровни, объемы) общественных явлений в конкретных условиях места и времени.

Различают два вида абсолютных величин: индивидуальные и суммарные.

Абсолютные статистические величины выражаются в натуральных, стоимостных и трудовых единицах измерения. Абсолютные статистические величины могут быть положительными (доходы) и отрицательными (убытки).

а) *натуральные единицы измерения* – соответствуют природным или потребительским свойствам предмета и выражаются в физических мерах веса, длины, объема (тонны, штуки, метры, литры).

Иногда одна натуральная единица измерения недостаточно характеризует изучаемое явление, поэтому применяют составные натуральные единицы измерения (киловатт – час, тонна – километр).

В статистике применяют условно-натуральные единицы измерения. Их получают, приводя различные натуральные, единицы к одной, принятой за основу (различные виды топлива пересчитываются в условное топливо).

б) *стоимостные единицы измерения* – выражаются в денежной форме.

в) *трудовые единицы измерения* – выражают затраты труда (человеко-день, человеко-час).

В практической деятельности, при отсутствии необходимой информации, абсолютные величины можно получить расчетным путем. Для этого используется *метод балансовой увязки показателя.*

$$Ок = Он + П + Р, \qquad (4.1)$$

где Ок – конечные остатки; Он – остатки начала; П – прибыль; Р – расход.

На основе абсолютных величин выводы делать нельзя.

3. Относительные величины

Относительные величины представляют собой частное от деления двух статистических величин и характеризуют количественное отношение между ними.

При расчете относительных величин следует иметь в виду, что в числителе всегда находится показатель, отражающий изучаемое явление, а в знаменателе – показатель, с которым проводится сравнение, принимаемый за базу сравнения или основание.

В зависимости от базы сравнения относительные величины могут быть выражены в:

- коэффициентах, база сравнения принимается за единицу и показывает, во сколько раз изучаемое явление отличается от базы сравнения;
- процентах, база сравнения принимается за 100;
- промилле, база сравнения принимается за 1000;
- продецимиле, база сравнения принимается за 10000000.

При расчете относительных величин необходимо обеспечить сопоставимость сравниваемых показателей.

В зависимости от задач относительные величины подразделяются:

- *относительный показатель динамики*

$$ОПД = \frac{текущий\ показатель}{предшествующий\ или\ базисный\ показатель} \quad (4.2)$$

- *относительный показатель планового задания*

$$ОППЗ = \frac{показатель,\ планируемый\ на\ (i+1)\ период}{показатель,\ достигнутый\ в\ i-м\ периоде} \quad (4.3)$$

- *относительный показатель выполнения плана*

$$ОПВП = \frac{показатель,\ достигнутый\ в\ (i+1)\ периоде}{показатель,\ планируемый\ на\ (i+1)\ период} \quad (4.4)$$

Между относительной величиной динамики, планового задания, выполнения плана существует взаимосвязь: показатель динамики представляет собой произведение планового задания на выполнение плана.

ОПД =ОППЗ *ОПВП) (4.5)

- *относительный показатель выполнения договорных обязательств*

$$ОПВД = \frac{фактический\ уровень}{уровень,\ предусмотренный\ договором} \quad (4.6)$$

- *относительный показатель структуры*

$$ОПС = \frac{показатель,\ характеризующий\ часть\ совокупности}{показатель\ по\ всей\ совокупности\ в\ целом} \quad (4.7)$$

- *относительный показатель интенсивности*

$$ОПИ = \frac{показатель,\ характеризующий\ явление\ A}{показатель,\ характеризующий\ среду\ рапространения\ явления\ A} \quad (4.8)$$

- *относительный показатель координации*

$$ОПК = \frac{показатель,\ характеризующий\ i-ю\ часть\ совокупности}{\begin{array}{c}показатель,\ характеризующий\ часть\ совокупности,\\ выбранную\ в\ качестве\ базы\ сравнения\end{array}} \quad (4.9)$$

- *относительный показатель сравнения*

$$ОПСр = \frac{показатель,\ характеризующий\ объект\ А}{показатель,\ характеризующий\ объект\ Б} \quad (4.10)$$

4. Средние величины

Средняя величина – это обобщающая характеристика однородной совокупности явлений по определенному признаку. Средняя величина выражает характерное, типичное значение признака у всех единиц. Средние величины используют для сравнения и выявления закономерностей.

Для того, чтобы рассчитать любую среднюю величину, необходимо соблюдать следующие требования:

1. средняя величина должна быть рассчитана только на основе массовых достоверных данных;

2. средняя величина будет объективна и типична, если она рассчитывается для качественно однородной совокупности;

3. числитель и знаменатель должны быть сравнимы.

Все средние величины объединяются в общей формуле средней степенной (при различной величине k):

$$\bar{x} = \sqrt[k]{\frac{\sum x_i^k f_i}{\sum f_i}}, \quad (4.11)$$

где x_i – отдельные значения изучаемого явления (вариант);

f_i – частота значений признака (вес i-го варианта);

n – объем совокупности.

Из формулы степенной средней, придавая К различные значения, можно вывести формулу различных средних величин. Выделяют следующие степенные средние величины:

- Средняя арифметическая простая

$$\bar{x}_{АРИФМ} = \frac{\sum x_i}{n} \quad (4.12)$$

- Средняя арифметическая взвешенная

$$\bar{x}_{АРИФМ} = \frac{\sum x_i f_i}{\sum f_i} \quad (4.13)$$

Средняя арифметическая обладает следующими свойствами:

1. Произведение средней на сумму частот всегда равно сумме произведений вариант на частоты.

2. Если от каждой варианты отнять (прибавить) какое-либо произвольное число, то новая средняя уменьшится (увеличится) на то же число.

3. Если каждую варианту умножить (разделить) на какое-то произвольное число, то средняя арифметическая увеличится (уменьшится) во столько раз.

4. Если все частоты (веса) разделить или умножить на какое-либо число, то средняя арифметическая от этого не изменится.

5. Сумма отклонений отдельных вариантов от средней арифметической всегда равняется нулю.

Перечисленные свойства могут быть использованы для того, чтобы облегчить технику исчисления средней арифметической.

- Средняя гармоническая простая

$$\bar{x}_{ГМ} = \frac{n}{\sum \frac{1}{x_i}} \tag{4.14}$$

- Средняя гармоническая взвешенная

$$\bar{x}_{ГМ} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} \tag{4.15}$$

$w_i = x_i f_i$ – произведение варианты на частоту.

- Средняя хронологическая

$$\bar{x}_{ХРОН} = \frac{\frac{x_1}{2} + x_2 + x_3 + \ldots + \frac{x_n}{2}}{n-1} \tag{4.16}$$

- Средняя геометрическая простая

$$\bar{x}_{ГЕОМ} = \sqrt{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} \tag{4.17}$$

- Средняя геометрическая взвешенная

$$\bar{x}_{ГЕОМ} = \sqrt[\sum f]{(x_1)^{f_1} \cdot (x_2)^{f_2} \cdot \ldots \cdot (x_n)^{f_n}} \tag{4.18}$$

- Средняя квадратическая простая

$$\bar{x}_{КВ} = \sqrt{\frac{\sum x_i^2}{n}} \tag{4.19}$$

- Средняя квадратическая взвешенная

$$\bar{x}_{КВ} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}} \tag{4.20}$$

- Средняя кубическая простая

$$\bar{x}_{КУВ} = \sqrt{\frac{\sum x_i^3}{n}}\,. \tag{4.21}$$

- Средняя кубическая взвешенная

$$\bar{x}_{КУВ} = \sqrt{\frac{\sum x_i^3 f_i}{\sum f_i}}\,. \tag{4.22}$$

Чем выше показатель степени k, тем больше значение средней величины (если индивидуальные значения признака варьируют). В итоге, можно построить следующее соотношение, которое называется правилом мажорантности средних:

$$\overline{X}_{ГМ} \le \overline{X}_{геом} \le \overline{X}_{арифм} \le \overline{X}_{КВ} \le \overline{X}_{куб} \qquad (4.23)$$

Вопросы для теоретического контроля знаний

1. Что представляют статистические показатели?
2. Назовите основные виды статистических показателей.
3. Что такое абсолютные показатели и каково их значение?
4. В каких единицах измерения выражаются абсолютные статистические величины?
5. Что называется относительными показателями?
6. Какие виды относительных показателей Вы знаете?
7. Дайте определение средних показателей.
8. Какие виды средних применяются в статистике?
9. Средняя арифметическая и ее свойства.
10. Понятие и способы исчисления средней арифметической и средней гармонической.

Примеры решения задач

1. Производство сахара-песка в РФ в январе – апреле 2021 г. характеризуется следующими данными (табл. 4.1).

Таблица 4.1

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

Рассчитаем относительные показатели динамики с переменной и постоянной базой сравнения:

Переменная база сравнения (цепные показатели)	Постоянная база сравнения (базисные показатели)
$\frac{138}{108} * 100\% = 127,8\%$	$\frac{138}{108} * 100\% = 127,8\%$
$\frac{131}{138} * 100\% = 94,9\%$	$\frac{131}{108} * 100\% = 121,3\%$
$\frac{206}{138} * 100\% = 157,3\%$	$\frac{206}{108} * 100\% = 190,7\%$

2. Предположим, оборот торговой фирмы в 2000 г. составил 2,0 млрд. руб. Исходя из проведенного анализа складывающихся на рынке тенденций руководство фирмы считает реальным в следующем году довести оборот до 2,8 млрд. руб. В этом случае относительный показатель плана, представляющий

собой отношение планируемой величины к фактически достигнутой, составит 140% (2,8 : 2,0 *100%). Предположим теперь, что фактический оборот фирмы за 2001 г. составил 2,6 млрд. руб. Тогда относительный показатель реализации плана, определяемый как отношение фактически достигнутой величины к ранее запланированной, составит 92,9% (2,6 : 2,8 *100%).

3. По данным таблицы 4.2 рассчитаем среднюю заработную плату в целом по трем предприятиям АО.

Таблица 4.2

Заработная плата предприятий АО

Предприятие	Численность промышленно-производственного персонала, чел.	Месячный фонд заработной платы, тыс. руб.	Средняя заработная плата, руб.
А	*1*	*2*	*3*
1	540	564,84	1046
2	275	332,75	1210
3	458	517,54	1130
Итого	1273	1415,13	?

Если мы располагаем только данными о средней заработной плате и численности работников (гр. 1 и 3), то общая средняя может быть рассчитана по формуле средней арифметической взвешенной:

$$\bar{x}_{АРИФМ} = \frac{\sum xifi}{\sum fi} = \frac{1046 \bullet 540 + 1210 \bullet 275 + 1130 \bullet 458}{540 + 275 + 458} = 1112\,руб.$$

Можно рассчитать среднюю заработную плату в целом по трем предприятиям по формуле средней гармонической взвешенной:

$$\bar{x}_{ГМ} = \frac{\sum wi}{\sum \frac{wi}{xi}} = \frac{564840 + 332750 + 517540}{\frac{564840}{1046} + \frac{332750}{1210} + \frac{517540}{1130}} = 1112\,руб.$$

Задания для самостоятельной работы студентов

1. По данным таблицы (4.3) рассчитать относительные показатели структуры.

Таблица 4.3

Показатели валового внутреннего продукта РФ в 1 квартале 2021 г.

Показатель	Объем	
	Трлн. руб.	% к итогу
ВВП – всего	508,0	
В том числе:		
производство товаров	185,4	
производство	277,9	
чистые налоги на продукты	44,7	

2. По данным таблицы 4.3 рассчитать относительный показатель координации.

3. На начало мая 2021 г. численность граждан, состоящих на учете в службе занятости, составляла 3 064 тыс. человек, а число заявленных предприятиями вакансий – 309 тыс. Рассчитать относительный показатель интенсивности.

4. На начало 2021 г. операции ГКО – ОФЗ проводили в Москве 108 официальных дилеров; в Новосибирске 16 и в Санкт – Петербурге 13. Вычислить относительный показатель сравнения.

5. Определить величину среднедушевого денежного дохода в целом по РФ.

Таблица 4.4

Распределение населения РФ в 1 квартале 2021 г. по уровню среднедушевых денежных доходов

Среднедушевой денежный доход в среднем за месяц, тыс. руб.	Численность населения, % к итогу
до 400	30,2
400-600	24,4
600-800	16,7
800-1000	10,5
1000-1200	6,5
1200-1600	6,7
1600-2000	2,7
2000 и выше	2,3
Итого	100

Тема 5. ПОКАЗАТЕЛИ ВАРИАЦИИ

1. Понятие вариации

Различие индивидуальных значений признака внутри изучаемой совокупности в статистике называется *вариацией признака.* Она возникает в результате того, что его индивидуальные значения складываются под совокупным влиянием разнообразных факторов (условий), которые по-разному сочетаются в каждом отдельном случае.

Под вариацией в статистике понимают такие количественные изменения величины исследуемого признака в пределах однородной совокупности, которые обусловлены перекрещивающимся влиянием действия различных факторов.

2. Абсолютные и средние показатели вариации

Для характеристики колеблемости признака используется ряд показателей.

Наиболее простой – *размах вариации,* определяемый как разность между наибольшим (x_{max}) и наименьшим (x_{min}) значениями вариантов:

$$R = x_{max} - x_{min}. \qquad (5.1)$$

Чтобы дать обобщающую характеристику распределению отклонений, исчисляют *среднее линейное отклонение,* которое учитывает различие всех единиц изучаемой совокупности.

Среднее линейное отклонение определяется как средняя арифметическая из отклонений индивидуальных значений от средней, без учета знака этих отклонений:

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n} \text{ (негруппированные данные).} \qquad (5.2)$$

$$\bar{d} = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i} \text{ (группированные данные).} \qquad (5.3)$$

Но этот показатель редко применяется на практике, т.к. во многих случаях не устанавливает степень рассеивания.

На практике меру вариации более объективно отражает *показатель дисперсии*, определяемый как средняя из отклонений, возведенных в квадрат.

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \text{ (негруппированные данные).} \qquad (5.4)$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} \text{ (сгруппированные данные).} \qquad (5.5)$$

Свойства дисперсии

1) Дисперсия постоянной величины равна нулю.

2) Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат.

3) Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин.

4) Дисперсия разности двух независимых случайных величин равна сумме дисперсий этих величин.

Среднее квадратическое отклонение определяется как корень из дисперсии.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}. \qquad (5.6)$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}. \qquad (5.7)$$

3. Показатели относительного рассеивания

Для характеристики меры колеблемости изучаемого признака исчисляются показатели колеблемости в относительных величинах. Они позволяют сравнивать характер рассеивания в различных пределах (при сравнении разноименных совокупностей). Расчет показателей меры относительного рассеивания осуществляют как отношение абсолютного показателя рассеивания к средней арифметической, умножаемое на 100%.

• *коэффициент осцилляции* отражает относительную колеблемость крайних значений признака вокруг средней.

$$V_R = \frac{R}{\bar{x}} \cdot 100\% . \quad (5.8)$$

• *линейный коэффициент вариации* характеризует долю усредненного значения абсолютных отклонений от средней величины.

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\% \textit{или} V_{\bar{d}} = \frac{\bar{d}}{M_e} \cdot 100\% . \quad (5.9)$$

• *коэффициент вариации* – относительный показатель, является наиболее распространенным показателем колеблемости, который используется для характеристики типичности средней. Он определяется как отношение среднего квадратического отклонения к средней величине.

$$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\% . \quad (5.10)$$

Если коэффициент вариации больше 33%, то это говорит о большой колеблемости признака изучаемой совокупности. Для того, чтобы найти показатели вариации в интервальном ряду, необходимо:

1. определить среднее значение в каждом интервале. Если интервал закрытый, то среднее значение определяется как сумма верхней и нижней границ, деленная на два. Если интервал открытый, то предполагается, что расстояние между границами данного интервала такое же, как и в соседнем интервале.
2. определяют общую среднюю для всей совокупности по формуле средней арифметической взвешенной, при этом вместо отдельных единиц признака берутся средние значения интервалов.
3. определяют показатели вариации, при этом при расчете средних показателей берётся отношение средних значений от общих средних.

4. Виды дисперсий

Различают три вида (показателя) колеблемости признака в совокупности: общую дисперсию, межгрупповую дисперсию и среднюю из внутригрупповых дисперсий.

Общая дисперсия характеризует изменение (вариацию) признака, который зависит от всех условий в данной совокупности. При ее расчете находят отклонения отдельных значений признака от общей средней, рассчитанной для всех совокупностей.

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} . \quad (5.11)$$

Межгрупповая дисперсия отражает изменение (вариацию) изучаемого признака, которая возникает под влиянием признака – фактора, положенного в основу группировки. Она характеризует колеблемость групповых (частных) средних около общей средней.

$$\delta_x^2 = \frac{\sum(\bar{x}_i - \bar{x})n_i}{\sum n_i}. \qquad (5.12)$$

где $\bar{x}_i$ и n_i – соответственно групповые средние и численности по отдельным группам.

Внутригрупповая (остаточная) дисперсия характеризует случайную вариацию в каждой отдельной группе. Эта вариация возникает под влиянием других, неучитываемых факторов и не зависит от условия (признака – фактора), положенного в основу группировки.

$$\delta_i^2 = \frac{\sum(x_i - \bar{x})^2 n_i}{\sum n_i}. \qquad (5.13)$$

Средняя из внутригрупповых дисперсий:

$$\bar{\delta}_i^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i}. \qquad (5.14)$$

Эти три показателя находятся в определенной зависимости между собой: показатель общей дисперсии может быть получен как сумма межгрупповой и остаточной.

$$\sigma^2 = \bar{\sigma}_i^2 + \delta_x^2. \qquad (5.15)$$

Это правило имеет большое практическое значение. С помощью него можно определить зависимость от фактора, положенного в основу группировки. Зависимость находится при помощи коэффициента детерминации.

$$\eta^2 = \frac{\delta_x^2}{\sigma^2}. \qquad (5.16)$$

Этот коэффициент показывает долю общей вариации изучаемого признака, обусловленную вариацией группировочного признака.

Корень квадратный из эмпирического коэффициента детерминации носит название эмпирическое корреляционное отношение:

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma^2}}. \qquad (5.17)$$

Оно характеризует влияние признака, положенного в основание группировки, на вариацию результативного признака.

5. Структурные характеристики вариационного ряда распределения

Для характеристики величины варьирующего признака пользуются структурными величинами. Они бывают двух видов:

- Мода – наиболее часто встречающееся значение ряда (варианты). Мода применяется, например, при определении размера обуви, одежды, пользующейся наибольшим спросом у покупателей.

Для дискретных рядов мода – это вариант, имеющий наибольшую частоту.

При расчете моды для интервального ряда необходимо вначале определить модальный интервал, т.е. интервал, который имеет наибольшую частоту, а затем значение модального признака. В этом случае моду рассчитывают по следующей формуле:

$$M_o = x_{Mo} + i_{Mo} \frac{(f_{Mo} - f_{Mo-1})}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})}, \qquad (5.18)$$

где $x_{мо}$ – нижняя граница модального интервала;

$i_{мо}$ – величина модального интервала;

$f_{мо}$ – частота, соответствующая модальному интервалу;

$f_{мо-1}$ – частота, предшествующая модальному интервалу;

$f_{мо+1}$ – частота интервала, следующего за модальным.

Мода определяет непосредственно размер признака, свойственный, хотя и значительной части, но все же не всей совокупности. Мода по своему обобщающему значению менее точна по сравнению со средней арифметической, характеризующей совокупность в целом с учетом всех без исключения элементов совокупности.

- Медиана – значение элемента, который больше или равен и одновременно меньше или равен половине остальных элементов ряда распределения. Медиана делит ряд на две равные части.

Медиана – значение признака у средней единицы ранжированного ряда. Ранжированный ряд – ряд, расположенный в порядке возрастания или убывания единиц.

Для ранжированного ряда с нечетным числом единиц медианой будет являться варианта, расположенная в центре ряда. Для ранжированного ряда с четным числом единиц медиана определяется как среднее арифметическое из двух смежных вариант, находящихся в центре ряда.

В интервальных рядах для определения медианы необходимо:

1. расположить значение признака по ранжиру;
2. для ранжированного ряда определить сумму частот;
3. найти медианный интервал. Он будет находиться там, где полусумма накопленных частот больше или равна сумме частот.

Значение медианы находится по формуле:

$$M_e = x_{Me} + i_{Me} \frac{\frac{1}{2}\sum f_i - S_{Me-1}}{f_{Me}}, \qquad (5.19)$$

где $x_{ме}$ – нижняя граница медианного интервала (медианным называется первый интервал, накопленная частота которого превышает половину общей суммы частот);

$i_{ме}$ – величина медианного интервала;

f /2 – полусумма частот ряда;

$S_{ме-1}$ – сумма накопленных частот, предшествующих медианному интервалу;

$f_{ме}$ – частота медианного интервала.

Медиана не зависит ни от амплитуды колебаний ряда, ни от распределения частот в пределах двух равных частей ряда, поэтому ее применение позволяет получить более точные результаты, чем при использовании других форм средних.

Вопросы для теоретического контроля знаний

1. Что представляет собой вариация признака, от чего зависят ее размеры?
2. Что такое размах вариации? По какой формуле он исчисляется? В чем его недостаток?
3. Что представляет собой среднее линейное отклонение? Его формулы, недостатки.
4. Какие показатели называются дисперсией и средним квадратическим отклонением? По каким формулам они вычисляются?
5. Что представляет собой вариация альтернативного признака? Чему она равна?
6. Какие виды дисперсий Вы знаете?
7. Каковы основные свойства дисперсий?
8. Как вычисляются показатели вариации в интервальном ряду?
9. Как осуществляется изучение формы распределения?
10. Теоретические распределения в анализе вариационных рядов.
11. Структурные характеристики вариационного ряда распределения.
12. Что представляют собой мода и медиана?
13. Как вычисляется мода в дискретных и интервальных рядах?
14. Как вычисляется медиана в дискретных и интервальных рядах?
15. Что представляет собой правило сложения дисперсий? В чем его практическое значение?
16. Дайте определение эмпирического коэффициента детерминации. В чем его смысл?
17. Дайте определение эмпирического корреляционного отношения. В чем его смысл?

Примеры решения задач

1. На основе данных (дискретный ряд) табл. 5.1 рассчитаем среднее линейное отклонение для дискретного ряда распределения.

Таблица 5.1

Распределение учителей средних школ района по стажу работы

Стаж работы, лет x_i	Число учителей в % к итогу, f_i	$x_i f_1$	$x_i - \bar{x}$	$\lvert xi - \bar{x}\rvert$	$\lvert xi - \bar{x}\rvert fi$
8	14	112	-2	2	28
9	20	180	-1	1	20
10	30	300	0	0	0
11	24	264	1	1	24
12	12	144	2	2	24
Итого	100	1000	0	-	96

Размах вариации стажа равен: $R = x_{max} - x_{min} = 12 – 8 = 4$ года.

Результаты вспомогательных расчетов даны в графах 3 – 5 табл. 5.1.

Средний стаж работы определяем по формуле средней арифметической взвешенной: $\bar{x} = 1000:100 = 10$ лет.

Отклонения индивидуальных значений стажа от средней с учетом и без учета знака содержатся в графах 4 и 5, а произведения отклонений по модулю на соответствующие частоты – в гр. 6.

Среднее линейное отклонение стажа работы учителей средних школ района:

$$\bar{d} = \frac{\sum \lvert xi - \bar{x}\rvert fi}{\sum fi} = \frac{96}{100} = 0{,}96 \ \text{года}.$$

2. Рассчитаем дисперсию и среднее квадратическое отклонение для следующего ряда распределения (табл. 5.2).

Таблица 5.2

Распределение магазинов города по товарообороту во 2 квартале 2021 г.

Группы магазинов по величине товарооборота, тыс. руб.	Число магазинов, f_i	Середина интервала, тыс. руб., x_i	$x_i f_i$	$xi - \bar{x}$	$(xi - \bar{x})^2$	$(xi - \bar{x})^2 fi$
40 – 50	2	45	90	49,2	2420,64	4841,28
50 – 60	4	55	220	-39,2	1536,64	6146,56
60 – 70	70	65	455	-29,2	852,64	5968,48
70 – 80	10	75	750	-19,2	368,64	3686,40
80 – 90	15	85	1275	-9,2	84,64	1269,60
90 – 100	20	95	1900	0,8	0,64	12,80
100 – 110	22	105	2310	10,8	116,64	2566,08
110 – 120	11	115	1265	20,64	432,64	4759,04
120 – 130	6	125	750	30,8	948,64	5691,84
130 – 140	3	135	405	40,8	1164,64	4993,92
Итого	100	0	9420	-	-	399936

При расчете вариации по интервальным рядам распределения необходимо сначала определить середины интервалов, а затем ввести дальнейшие расчеты, рассматривая ряд середин интервалов как дискретный ряд распределения.

Результаты вспомогательных расчетов для определения дисперсии и среднего и квадратического отклонения содержатся в графах 2 – 6 табл. 5.2.

Средний размер товарооборота определяется по средней арифметической взвешенной и составляет: $\bar{x} = \frac{9420}{100} = 94{,}2$ *тыс.руб*.

Дисперсия товарооборота:

$$\sigma^2 = \frac{\sum (xi - \bar{x})^2 fi}{\sum fi} = \frac{39936}{100} = 399.36.$$

Среднее квадратическое отклонение товарооборота определяется как корень квадратный из дисперсии:

$$\sigma = \sqrt{\frac{\sum (xi - \bar{x})^2 fi}{\sum fi}} = \sqrt{399.36 =} 19.98 \cong 20 \text{ тыс.руб.}$$

3. Определим групповые дисперсии, среднюю из групповых дисперсий, межгрупповую дисперсию, общую дисперсию по данным табл. 5.3 (аналитическая группировка).

Таблица 5.3

Производительность труда двух бригад рабочих токарей

1 бригада				2 бригада			
№	Изготовлено деталей за час, шт.	$xi - \bar{x}_1$	$(xi - \bar{x}_1)^2$	№	Изготовлено деталей за час, шт.	$xi - \bar{x}_2$	$(xi - \bar{x}_2)^2$
1	13	-2	4	7	18	-3	9
2	14	-1	1	8	19	-2	4
3	15	0	0	9	22	1	1
4	17	2	4	10	20	-1	1
5	16	1	1	11	24	3	9
6	15	0	0	12	23	2	4
90		-	10	126		-	24

Для расчета групповых дисперсий вычислим среднее по каждой группе:

$$\bar{x}_1 = \frac{90}{6} = 15 \text{ шт.}; \quad \bar{x}_2 = \frac{126}{6} = 21 \text{ шт.}$$

Промежуточные расчеты дисперсий по группам представлены в табл. 5.3. Подставив полученные значения в формулу, получим:

$$\sigma_1^2 = \frac{\sum (xi - \bar{x})^2}{n} = \frac{10}{6} = 1.666 \approx 1.67;$$

$$\sigma_2^2 = \frac{\sum (xi - \bar{x})^2}{n} = \frac{28}{6} \approx 4.67.$$

Средняя из групповых дисперсий:

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2 fi}{\sum fi} = \frac{1.67 \cdot 6 + 4.67 \cdot 6}{12} = \frac{10+28}{12} = \frac{38}{12} \cong 3.17 .$$

Затем рассчитаем межгрупповую дисперсию. Для этого предварительно определим общую среднюю как среднюю взвешенную из групповых средних:

$$\bar{x} = \frac{\sum \bar{x}_i fi}{\sum fi} = \frac{16 \cdot 6 + 21 \cdot 6}{12} = \frac{90+126}{12} = 18 \ \text{шт.}$$

Теперь определим межгрупповую дисперсию:

$$\delta^2 = \frac{\sum (xi - \bar{x})^2 fi}{\sum fi} = \frac{(15-18)^2 \cdot 6 + (21-18)^2 \cdot 6}{12} = \frac{9 \cdot 6 + 9 \cdot 6}{12} = \frac{108}{12} = 9 .$$

Задания для самостоятельной работы

1. Рассчитать показатели вариации по данным о распределении фирм по стоимости основных фондов (табл. 5.4).

Таблица 5.4

Расчет коэффициента асимметрии

Группы фирм по стоимости основных фондов, млн. руб.	Количество фирм
0,5-1,0	20
1,0-1,5	40
1,5-2,0	25
2,0-2,5	20
Итого	105

2. Рассчитать моду и медиану по данным таблицы 5.5.

Таблица 5.5

Распределение обуви, проданной коммерческой фирмой в январе 2021 г.

Размер	34	37	38	40	41	43	44 и более	Итого
Кол-во проданных пар, % к итогу	23	19	20	15	14	7	2	100

3. Рассчитать моду и медиану для интервального ряда распределения по данным таблицы 5.6.

Таблица 5.6

Распределение семей города по размеру среднедушевого дохода в январе 2021 г.

Группы семей по размеру дохода, руб.	Число семей
До 500	600
500 – 600	700
600 – 700	1 700
700 – 800	2 500
800 – 900	2 200
900 – 1000	1 500
Свыше 1000	800
Итого	10 000

Тема 6. ВЫБОРОЧНЫЙ МЕТОД В СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

1. Понятие выборочного наблюдения

Выборочное наблюдение – способ несплошного наблюдения, при котором обобщающие показатели изучаемой совокупности устанавливаются по некоторой ее части на основе положений случайного отбора. Цель выборочного наблюдения – нахождение средних характеристик всей совокупности по ее выборочной части. При этом подлежащая изучению статистическая совокупность, из которой производится отбор части единиц, называется *генеральной совокупностью*. Отобранная из генеральной совокупности некоторая часть единиц называется *выборочной совокупностью* или *выборкой*.

В генеральной совокупности доля единиц, обладающих изучаемым признаком, называется генеральной долей (Р), а средняя величина признака называется генеральной средней ($\bar{x}$).

В выборочной совокупности доля изучаемого признака называется выборочной долей (W), а средняя величина называется выборочной средней ($\tilde{x}$).

Основные характеристики генеральной и выборочной совокупностей обозначаются определенными символами (табл. 6.1).

Таблица 6.1

Символы основных характеристик параметров генеральной и выборочной совокупностей

Характеристика	Генеральная совокупность	Выборочная совокупность
Объем совокупности (численность единиц)	N	n
Численность единиц, обладающих обследуемым признаком	M	m
Доля единиц, обладающих обследуемым признаком	P = M/N	W = m/n
Средний размер признака	$\bar{x} = \frac{\sum x_i}{N}$	$\tilde{x} = \frac{\sum x_i}{n}$
Дисперсия количественного признака	$\sigma_{\bar{x}}^2 = \frac{\sum (x_i - \bar{x})^2}{N}$ $\sigma_{\bar{x}}^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$	$\sigma_{\tilde{x}}^2 = \frac{\sum (x_i - \tilde{x})^2}{n}$ $\sigma_{\tilde{x}}^2 = \frac{\sum (x_i - \tilde{x})^2 f_i}{\sum f_i}$
Дисперсия доли	$\sigma_p^2 = pq$	$\sigma_w^2 = W(1-W)$

2. Виды выборочного наблюдения

Существуют следующие виды выборок по форме организации:

а) *повторный отбор*, когда отобранная единица возвращается обратно в генеральную совокупность и для нее сохраняется возможность быть отобранной еще раз.

б) *бесповторный отбор,* когда каждая отобранная единица может быть обследована только один раз.

По способу организации различают пять видов выборок:

- *Собственно случайный отбор*

Состоит в том, что выборочная совокупность образуется в результате случайного отбора (жеребьевки) единиц из генеральной совокупности. При этом каждой единице предоставляется равная возможность попадания в выборку. Может быть повторной и бесповторной. К недостаткам этого вида отбора можно отнести: 1. нужно знать заранее размер генеральной совокупности, что не всегда возможно; 2. сложность его применения при очень больших совокупностях.

- *Механический отбор*

Это выборка, при которой генеральная совокупность разбивается на равные части, и из каждой части в выборку отбирают лишь одну единицу, через определенный интервал. При этом единицы генеральной совокупности должны быть распределены по любому признаку, который не будет изучаться в выборочном наблюдении. Промежуток, через который единицы попадают в выборку, зависит от размеров выборочной и генеральной совокупности. Механическая выборка является единственно возможной, когда совокупность заранее неизвестна, а формируется постепенно. По способу проведения механический отбор всегда бесповторен.

- *Типический отбор*

Это вид выборки, при котором генеральная совокупность разбивается на группы по какому-либо признаку, который будет изучаться в выборочной совокупности. Затем проводится случайный или механический отбор единиц из каждой совокупности. Типическая выборка применяется при изучении сложных статистических явлений.

- *Серийный (гнездовой) отбор*

Выборка, при которой отбору подлежат не отдельные единицы совокупности, а группы или серии, в которых единицы определенным образом связаны между собой. Отбор серии осуществляется случайно или механически, а внутри каждой серии осуществляется сплошное обследование.

- *Комбинированный отбор*

Это вид выборки, при котором генеральная совокупность разбивается на группы, затем проводится отбор групп, а из отобранных групп проводится отбор единиц – многоступенчатая выборка. При этом на отдельных стадиях могут применяться разные виды выборок.

Комбинированный отбор может быть двух видов:

Многоступенчатый – отбор проходит по ступеням, последовательным стадиям, причем на каждой ступени единицы отбора будут всегда разные.

Многофазный – программа обследования на различных фазах отбора меняется, а единицы отбора одни и те же.

Отличие комбинированной выборки от типической состоит в том, что при типической выборке выборочная совокупность формируется их всех групп, а при комбинированной осуществляется отбор самих групп, и поэтому не все они попадают в выборочную совокупность.

3. Ошибки выборочного наблюдения

В связи с тем, что статистическая совокупность состоит из единиц с варьирующим признаком, состав выборочной совокупности может отличаться от состава генеральной совокупности. Это значит, что обобщающие показатели выборки не совпадают со значениями этих показателей в генеральной совокупности. Расхождения между показателями выборочной и генеральной совокупностей называются *ошибкой выборки*.

При определении точности выборочных показателей различают среднюю и предельную ошибки выборки.

3.1. Средняя ошибка выборки зависит от степени вариации значений признака внутри совокупности и от размеров выборочной совокупности.

▪ *Средняя ошибка собственно-случайной выборки*

$$\mu = \frac{\sigma}{\sqrt{n}} \text{ (повторный отбор);} \quad (6.1)$$

$$\mu = \sqrt{\frac{\sigma^2}{n}(1-\frac{n}{N})} \text{ (бесповторный отбор).} \quad (6.2)$$

Для определения *средней ошибки механической выборки* используется формула средней ошибки выборки при случайном бесповторном отборе.

▪ *Средняя ошибка типической выборки*

$$\mu \cong \sqrt{\frac{\bar{\sigma}^2}{\mathrm{n}}} \text{ (повторный отбор),} \quad (6.3)$$

где $\bar{\sigma}^2$ – средняя из внутригрупповых дисперсий;

n – численность выборочной совокупности.

$$\mu \cong \sqrt{\frac{\bar{\sigma}^2}{n}(1-\frac{n}{N})} \text{ (бесповторный отбор),} \quad (6.4)$$

где N – численность генеральной совокупности.

▪ *Средняя ошибка серийного отбора*

$$\mu = \sqrt{\frac{\delta^2}{r}} \text{ (повторный отбор),} \quad (6.5)$$

$$\mu = \sqrt{\frac{\delta^2}{r}(1-\frac{r}{R})} \text{ (бесповторный отбор),} \quad (6.6)$$

где r – число отобранных серий;

R – общее число серий.

При этом $\delta^2 = \frac{\sum(\tilde{x}_i - \tilde{x})^2}{r}$, (6.7)

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

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

▪ *Средняя ошибка выборки для альтернативного признака*

$$\mu = \sqrt{\frac{w(1-w)}{n}}. \quad (6.8)$$

3.2. Предельная ошибка выборки

Полученные показатели генеральной совокупности посредством выше указанных формул можно гарантировать лишь с определенной долей вероятности. В математической статистике доказано, что пределы значений генеральной совокупности отличаются от показателей выборочной совокупности на величину средней ошибки выборки лишь с вероятностью 0,683. Это значит, что в 683 случаях из 1000 значения генеральной совокупности будут находиться в установленных пределах, а в остальных 317 они могут выйти за эти пределы. Но вероятность можно повысить, если расширить пределы отклонений. Для этого средняя ошибка выборки увеличивается в t раз. Коэффициент t называется коэффициентом доверия. Он определяется в зависимости от того, с какой доверительной вероятностью надо гарантировать результаты выборочного обследования.

Установлено, что при t = 1 вероятность равна 0,683 (68,3%); при t = 2 вероятность равна 0,954 (95,4%); при t = 3, вероятность равна 0,997 (99,7%).

Средняя ошибка выборки, умноженная на коэффициент доверия, называется *предельной ошибкой выборки.*

$$\Delta_{\tilde{x}} = t\mu_{\tilde{x}}. \qquad (6.9)$$

Зная выборочную среднюю величину признака ($\tilde{x}$) и предельную ошибку выборки ($\Delta_{\tilde{x}}$), можно определить границы (пределы), в которых заключена генеральная средняя:

$$\tilde{x} - \Delta_{\tilde{x}} \le \bar{x} \ge \tilde{x} + \Delta_{\tilde{x}}, \quad \text{или} \quad \tilde{x} - \bar{x} = \pm\Delta_{\tilde{x}}. \qquad (6.10)$$

Зная выборочную долю признака (w) и предельную ошибку выборки (Δ_w), можно определить границы, в которых заключена генеральная доля (p):

$$w - \Delta_w \le p \ge w + \Delta_w. \qquad (6.11)$$

4. Определение необходимого объема выборки

При проектировании выборочного наблюдения возникает вопрос о необходимой численности выборки. Эта численность может быть определена на базе допустимой ошибки при выборочном наблюдении исходя из вероятности, на основе которой можно гарантировать величину устанавливаемой ошибки, и, наконец, на базе способа отбора.

Приведем формулы необходимого объема выборки для наиболее часто используемых на практике способов формирования выборочной совокупности (таблица 6.2).

Таблица 6.2

Необходимый объем выборки для некоторых способов формирования выборочной совокупности

Вид выборочного наблюдения	*Повторный отбор*	*Бесповторный отбор*
Случайная выборка		
а) при определении среднего размера признака	$n = \frac{t^2\sigma^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}}$	$n = \frac{t^2\sigma^2_{\tilde{x}}N}{\Delta^2_{\tilde{x}}N + t^2\sigma^2_{\tilde{x}}}$
б) при определении доли признака	$n = \frac{t^2W(1-W)}{\Delta^2_w}$	$n = \frac{t^2W(1-W)N}{\Delta^2_w N + t^2W(1-W)}$
Механическая выборка		
а) при определении среднего размера признака	-	$n = \frac{t^2\sigma^2_{\tilde{x}}N}{\Delta^2_{\tilde{x}}N + t^2\sigma^2_{\tilde{x}}}$

б) при определении доли признака	-	$n = \dfrac{t^2 W(1-W)N}{\Delta^2_w N + t^2 W(1-W)}$
Типическая выборка		
а) при определении среднего размера признака	$n = \dfrac{t^2 \bar{\sigma}^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}}$	$n = \dfrac{t^2 \bar{\sigma}^2_{\tilde{x}} N}{\Delta^2_{\tilde{x}} N + t^2 \bar{\sigma}^2_{\tilde{x}}}$
б) при определении доли признака	$n = \dfrac{t^2 \overline{W(1-W)}}{\Delta^2_w}$	$n = \dfrac{t^2 \overline{W(1-W)} N}{\Delta^2_w N + t^2 \overline{W(1-W)}}$
Серийная выборка		
а) при определении среднего размера признака	$r = \dfrac{t^2 \delta^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}}$	$r = \dfrac{t^2 \delta^2_{\tilde{x}} R}{\Delta^2_{\tilde{x}} R + t^2 \delta^2_{\tilde{x}}}$
б) при определении доли признака	$r = \dfrac{t^2 W_r (1-W_r)}{\Delta^2_w}$	$r = \dfrac{t^2 W_r (1-W_r) R}{\Delta^2_w R + t^2 W_r (1-W_r)}$

5. Способы распространения результатов выборочного наблюдения на генеральную совокупность

На практике используют два способа распространения выборочного наблюдения на генеральную совокупность:

1. способ прямого пересчета применяется тогда, когда с помощью выборки рассчитываются средние объемные показатели. Например, при определении общего количества молока, полученного от коров в личных подсобных хозяйствах. Для этого на основе выборочного обследования бюджетов крестьян устанавливают средний надой на одну корову. По данным переписи скота рассчитывают поголовье коров в личных подсобных хозяйствах. Перемножая эти данные, определяют общий объем молока.

2. способ поправочных коэффициентов применяется для уточнения и проверки данных сплошного наблюдения. Его используют в том случае, когда можно сопоставить данные выборочного и сплошного наблюдения. При этом рассчитывают поправочный коэффициент, по которому проводят корректировку итогов сплошного наблюдения. Например, в результате сплошной переписи скота установлено, что в личных подсобных хозяйствах области имелось 100 000 коров, в т.ч. в районе А 1000 коров. Выборочное обследование этого района показало, что фактическое число коров составило 1015, т.е. при сплошной переписи не были учтены 15 коров. Поправочный коэффициент будет равен: 0,015 (15/1000). Тогда по области с учетом поправочного коэффициента число коров составит: (0,015*100 000): 1000+ 100000 = 1500 + 100000 = 101 500 голов.

Вопросы для теоретического контроля знаний

1. Какое наблюдение называется выборочным?
2. В чем преимущества выборочного наблюдения перед сплошным?
3. Какие вопросы необходимо решить для проведения выборочного наблюдения?
4. Почему при выборочном наблюдении неизбежны ошибки и как они классифицируются?
5. Какие условия правильного отбора единиц совокупности при выборочном наблюдении?
6. Как производятся собственно – случайный, механический, типический и серийный отборы?
7. В чем различие повторной и бесповторной выборки?
8. Что представляет собой средняя ошибка выборки (доли и средней)?
9. По каким расчетным формулам находят средние ошибки выборки (для средней и доли) при повторном и бесповторном отборах?
10. Что характеризует предельная ошибка выборки и по каким формулам она исчисляется (для средней и доли)?
11. Что показывает коэффициент доверия?
12. По каким формулам определяется необходимая численность выборки, обеспечивающая с определенной вероятностью заданную точность наблюдения?
13. Раскройте сущность понятия «малая выборка».
14. Какими способами осуществляется распространение результатов выборочного наблюдения на всю совокупность?

Примеры решения задач

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

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

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

Вначале определим на основе имеющегося распределения семей выборочную среднюю и дисперсию:

$$7400 : 5000 = 1,5 \text{ чел.}; \qquad 7650 : 5000 = 1,53.$$

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

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

$$\Delta_X = t\sqrt{\frac{\delta_X^{\ 2}}{2}(1-\frac{n}{N})} = 2\sqrt{\frac{1,53}{500}\left(1-\frac{5000}{250000}\right)} = 0,035\,.$$

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

Задания для самостоятельной работы

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

2. С целью определения средних затрат времени при поездках на работу населением города планируется выборочное наблюдение на основе случайного повторного отбора. Сколько людей должно быть обследовано, чтобы с вероятностью 0,954 ошибка выборки не превышала 1 мин. при среднем квадратическом отклонении 15 мин.?

3. Из партии импортируемой продукции на посту Московской региональной таможни было взято в порядке случайной повторной выборки 20 проб продукта А. В результате проверки установлена средняя влажность продукта А в выборке, которая оказалась равной 6% при среднем квадратическом отклонении 1%. С вероятностью 0,683 определите пределы средней влажности продукта во всей партии импортируемой продукции.

Тема 7. ИНДЕКСНЫЙ МЕТОД В СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

1. Понятие экономических индексов

Индекс – относительный показатель, который выражает соотношение величин какого – либо явления во времени, пространстве или сравнение фактических данных с любым эталоном (план, норматив, прогноз).

По степени охвата явления индексы бывают:

- Индивидуальные, служат для характеристики изменения отдельных элементов сложного явления;
- Сводные (общие), служат для измерения динамики сложного явления, составные части которого непосредственно не соизмеримы;
- Групповые (субиндексы), охватывают не все элементы сложного явления, а только их часть.

По базе сравнения все индексы можно разделить на:

- Динамические, отражают изменения явления во времени (бывают цепные и базисные);
- Территориальные, применяются для межрегиональных сравнений.

По виду весов индексы бывают:

- С постоянными весами;
- С переменными весами.

В зависимости от формы построения различают:

- Индексы агрегатные;
- Индексы средние.

По характеру объема исследования общие индексы подразделяются (в основе деления лежит вид индексируемой величины):

- Индексы количественных показателей; к ним относятся: объем продукции в натуральном выражении, численность работников, величина материальных затрат, величина основных фондов.
- Индексы качественных показателей; к ним относятся: цена, себестоимость, производительность труда, материалоотдача, фондоотдача.

По объекту исследования индексы бывают:

- Индексы производительности труда;
- Индексы себестоимости;
- Индексы физического объема продукции;
- Индексы стоимости продукции и т.д.

По составу явления можно выделить две группы индексов:

- Постоянного состава;
- Переменного состава.

По периоду исчисления индексы подразделяются на:

- Годовые;
- Квартальные;
- Месячные;
- Недельные.

Буквой «i» – обозначаются индивидуальные индексы, «I» – общие индексы. Знак внизу справа означает период: «0» – базисный, «1» – отчетный. Помимо этого используются определенные символы для обозначения индексируемых показателей:

q – количество товара в натуральном выражении;

p – цена единицы продукции;

z – себестоимость единицы продукции;

t – затраты времени на производство единицы продукции;

w – выработка продукции в стоимостном выражении на одного рабочего или в единицу времени;

v – выработка продукции в натуральном выражении на одного рабочего или в единицу времени;

T – общие затраты времени (tq) или численность рабочих;

pq – стоимость продукции или товарооборот;

zq – издержки производства.

2. Индивидуальные и общие индексы

Индивидуальные индексы получают в результате сравнения однотоварных явлений. Представляют собой относительные величины динамики, выполнения плана и их расчет не требует знания специальных правил.

- *Индивидуальный индекс физического объема продукции:*

$$i_q = \frac{q_1}{q_0}. \qquad (7.1)$$

Этот индекс показывает, во сколько раз возрос (уменьшился) выпуск какого-либо одного товара в отчетном периоде по сравнению с базисным или сколько процентов составляет рост (снижение) выпуска товара. В знаменателе может быть не только количество продукции, произведенной за какой – то предыдущий период, но и плановое значение ($q_{пл}$), нормативное ($q_н$) или эталонное значение, принятое за базу сравнения ($q_э$). Тогда формула примет следующий вид:

$$i_q = \frac{q_1}{q_{пл}};\ i_q = \frac{q_1}{q_н};\ i_q = \frac{q_1}{q_э}. \qquad (7.2\text{-}7.4)$$

Индексы других показателей строятся аналогично.

В экономических расчетах чаще всего используются общие индексы, которые характеризуют изменение совокупности в целом. В индексной теории сложились две концепции, которые по-разному интерпретируют общие индексы.

Согласно *синтетической концепции* особенность общих индексов состоит в том, что они выражают относительное изменение сложных явлений, отдельные части которых непосредственно несоизмеримы. Таким образом, *синтетические свойства* индексов состоят в том, что посредством индексного метода производится соединение (агрегирование) в целое разнородных единиц статистической совокупности.

В аналитической теории индексы трактуются как показатели, необходимые для измерения влияния изменения составных частей сложного явления на изменение уровня этого явления. Таким образом, *аналитические свойства* индексов состоят в том, что посредством индексного метода определяется влияние факторов на изменение изучаемого показателя.

3. Агрегатная форма общего индекса

Агрегатный индекс – сложный относительный показатель, который характеризует среднее изменение социально – экономического явления, состоящего из несоизмеримых элементов.

Числитель и знаменатель агрегатного индекса представляет собой сумму произведений двух величин, одна из которых меняется (индексируемая величина), а другая остается неизменной в числителе и знаменателе (вес индекса).

Индексируемая величина – признак, изменение которого изучается.

Вес индекса – величина, служащая для целей соизмерения индексируемых величин.

Методика построения агрегатного индекса предусматривает решение трех вопросов:

1. какая величина будет индексируемой;
2. по какому составу разнородных элементов явления необходимо исчислить индекс;
3. что будет служить весом при расчете индекса.

При выборе веса индекса принято руководствоваться правилом:

- Если строится индекс количественного показателя, то веса берутся за базисный период;
- При построении индекса качественного показателя используются веса отчетного периода.

Формулы для расчета общих индексов

Наименование индекса	Формула расчета индекса	Что показывает индекс
Индекс стоимости продукции (товарооборота)	$I_{pq}=\frac{\sum p_1q_1}{\sum p_0q_0}$ (7.5)	Во сколько раз изменилась стоимость продукции отчетного периода по сравнению с базисным
Индекс физического объема продукции	$I_q=\frac{\sum q_1p_0}{\sum q_0p_0}$ (7.6)	Во сколько раз изменилась стоимость продукции в результате изменения объема ее производства
Индекс цен	$I_p=\frac{\sum p_1q_1}{\sum p_0q_1}$ (7.7)	Во сколько раз изменилась стоимость продукции в результате изменения ее цен
$I_{pq}=I_p\times I_q$ (7.8)		

Индексы других показателей строятся аналогично.

4. Средние индексы

Средние индексы – индексы, вычисленные как средняя величина из индивидуальных индексов.

Выделяют следующие виды средних индексов:

• *средний арифметический индекс*, характеризует изменение количественных показателей.

Применяется в тех случаях, когда имеются данные об общей величине изучаемого явления в базисном периоде и изменения в процентах количественного показателя в отчетном периоде по сравнению с базисным.

Средний арифметический индекс физического объема продукции

$$I_q = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0}. \tag{7.5}$$

Индекс характеризует изменение количества продукции в отчетном периоде, по сравнению с базисным, в относительных величинах. Разница между числителем и знаменателем этого индекса показывает абсолютное изменение объема производства за счет изменения в количестве продукции.

Средний арифметический индекс производительности труда:

$$I_t = \frac{\sum i_t t_1 q_1}{\sum t_1 q} = \frac{\sum i_t T_1}{\sum T_1}; \tag{7.6}$$

• *средний гармонический индекс,* используется для характеристики качественных показателей.

Он рассчитывается в том случае, когда имеются данные об общей величине изучаемого явления за отчетный период и изменения в процентах качественного показателя в отчетном периоде по сравнению с базисным.

Средний гармонический индекс цен:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}. \tag{7.7}$$

Индекс характеризует относительное изменение цены в отчетном периоде по сравнению с базисным. Разница между его числителем и знаменателем показывает абсолютное изменение объема производства за счет изменения цен.

Средний гармонический индекс себестоимости:

$$I_z = \frac{\sum z_1 q_1}{\sum \frac{z_1 q_1}{i_z}}. \tag{7.8}$$

5. Выбор базы и весов индексов

Выбор базы сравнения и весов индексов – это два важнейших методологических вопроса построения систем индексов.

Система индексов – ряд последовательно построенных индексов.

В зависимости от базы сравнения системы индексов бывают:

• базисные – это ряд последовательно вычисленных индексов одного и того же явления с постоянной базой сравнения, т.е. в знаменателе всех индексов находится индексируемая величина базисного периода.

• цепные – это ряд индексов одного и того же явления, вычисленных с меняющейся от индекса к индексу базой сравнения.

Базисные индексы дают более наглядную характеристику общей тенденции развития исследуемого явления, а цепные – четче отражают последовательность изменения уровней во времени.

Примеры построения системы индивидуальных индексов:

Название индекса	*Система индексов*	
	базисные	*цепные*
Индекс стоимости	$\frac{p_1q_1}{p_0q_0};\frac{p_2q_2}{p_0q_0};...;\frac{p_nq_n}{p_0q_0}$	$\frac{p_1q_1}{p_0q_0};\frac{p_2q_2}{p_1q_1};...;\frac{p_nq_n}{p_{n-1}q_{n-1}}$
Индекс физического объема	$\frac{q_1}{q_0};\frac{q_2}{q_0};...;\frac{q_n}{q_0}$	$\frac{q_1}{q_0};\frac{q_2}{q_1};...;\frac{q_n}{q_{n-1}}$
Индекс цен	$\frac{p_1}{p_0};\frac{p_2}{p_0};...;\frac{p_n}{p_0}$	$\frac{p_1}{p_0};\frac{p_2}{p_1};...;\frac{p_n}{p_{n-1}}$

Системы базисных и цепных индексов могут быть построены для агрегатных индексов.

а) цепные индексы стоимости: $\frac{\sum p_1q_1}{\sum p_0q_0};\frac{\sum p_2q_2}{\sum p_1q_1};...;\frac{\sum p_nq_n}{\sum p_{n-1}q_{n-1}};$

б) базисные индексы стоимости: $\frac{\sum p_1q_1}{\sum p_0q_0};\frac{\sum p_2q_2}{\sum p_0q_0};...;\frac{\sum p_nq_n}{\sum p_0q_0}.$

При построении систем индексов можно использовать постоянные и переменные веса.

Системой индексов с постоянными весами называется система сводных индексов одного и того же явления, вычисленных с весами, не меняющимися при переходе от одного индекса к другому. Постоянные веса позволяют исключить влияние структуры на величину индекса.

- Система базисных индексов физического объема с постоянными весами

$$\frac{\sum q_1p_0}{\sum q_0p_0};\frac{\sum q_2p_0}{\sum q_0p_0};...;\frac{\sum q_np_0}{\sum q_0p_0}.$$

- Система цепных индексов физического объема с постоянными весами

$$\frac{\sum q_1p_0}{\sum q_0p_0};\frac{\sum q_2p_0}{\sum q_1p_0};...;\frac{\sum q_np_0}{\sum q_{n-1}p_0}.$$

Система индексов с переменными весами представляет собой систему сводных индексов одного и того же явления, вычисленных с весами, последовательно меняющимися от одного индекса к другому.

- Система базисных индексов цен с переменными весами

$$\frac{\sum p_1q_1}{\sum p_0q_1};\frac{\sum p_2q_2}{\sum p_0q_2};...;\frac{\sum p_nq_n}{\sum p_0q_n}.$$

- Система цепных индексов цен с переменными весами

$$\frac{\sum p_1q_1}{\sum p_0q_1};\frac{\sum p_2q_2}{\sum p_1q_2};...;\frac{\sum p_nq_n}{\sum p_{n-1}q_n}.$$

Системы агрегатных индексов обладают теми же свойствами, что и системы индивидуальных индексов.

6. Индексы постоянного, переменного составов и структурных сдвигов

При изучении динамики социально – экономических явлений приходится производить индексные сопоставления более чем за два периода. В тех случаях, когда необходимо установить изменение качественного показателя по одному и тому же признаку, но относящемуся к разным объектам, применяются индексы постоянного, переменного состава и структурных сдвигов.

- *индекс переменного состава* – индекс, выражающий соотношение средних уровней изучаемого явления, относящихся к разным периодам времени.

Индекс переменного состава себестоимости продукции одного и того же вида:

$$I_{nc}=\frac{\bar{z}_1}{z_0}=\frac{\sum z_1q_1}{\sum q_1}\div\frac{\sum z_0q_0}{\sum q_0}. \tag{7.9}$$

Отражает изменение не только индексируемой величины (себестоимости), но и структуры совокупности (весов).

- *индекс постоянного (фиксированного) состава* – индекс, исчисленный с весами, зафиксированными на уровне одного какого – либо периода, и показывающий изменение только индексируемой величины.

$$I_{фс}=\frac{\sum z_1q_1}{\sum q_1}\div\frac{\sum z_0q_1}{\sum q_1}=\frac{\sum z_1q_1}{\sum z_0q_1}. \tag{7.10}$$

- *индекс структурных сдвигов* – индекс, характеризующий влияние изменения структуры изучаемого явления на динамику среднего уровня этого явления.

$$I_{c.c.}=\frac{\sum z_0q_1}{\sum q_1}\div\frac{\sum z_0q_0}{\sum q_0}=\frac{\sum z_0q_1}{\sum z_0q_1}\div\frac{\sum q_1}{\sum q_0}. \tag{7.11}$$

$$I_{nc}=I_{фс}\times I_{c.c.}. \tag{7.12}$$

Аналогично строятся системы индексов для других показателей.

7. Индексы пространственно – территориального сопоставления

В статистической практике часто возникает потребность в сопоставлении уровней экономических явлений в пространстве: по странам, экономическим районам и т.д., т.е. в исчислении территориальных индексов. При построении территориальных индексов приходится решать вопрос, какие веса использовать при их исчислении.

Если стоит задача сравнить цены двух регионов (А и В), то можно построить два индекса:

$$I_{A/B} = \frac{\sum p_A q_A}{\sum p_B q_A} \qquad (7.13)$$

(индекс, в котором в качестве базы сравнения применяются данные по региону А);

$$I_{B/A} = \frac{\sum p_B q_B}{\sum p_A q_B} \qquad (7.14)$$

(индекс, в котором в качестве базы сравнения применяются данные по региону В).

В теории и практике статистики предлагаются различные методы построения территориальных индексов, в т.ч. метод стандартных весов. Суть его в том. Что значения индексируемой величины взвешивается не по весам какого – то одного региона, а по весам области, экономического района, республики, в котором находятся сравниваемые регионы:

$$I_p = \frac{\sum p_A (q_A + q_B)}{\sum p_B (q_A + q_B)}. \qquad (7.15)$$

Вопросы для теоретического контроля знаний

1. Что называется индексом в статистике?
2. Какие задачи решают при помощи индексов?
3. Что характеризуют индивидуальные индексы? Приведите примеры.
4. В чем сущность общих индексов?
5. Как исчисляется агрегатный индекс стоимости продукции (товарооборота в фактических ценах) и что он характеризует?
6. Как исчисляется агрегатный индекс физического объема продукции (товарооборота) и что он характеризует? Напишите формулу.
7. Когда возникает необходимость преобразования индекса физического объема в средний арифметический и средний гармонический; каким образом происходят такие преобразования? Покажите на примерах.
8. Как исчисляют агрегатные индексы цен (Паше и Ласпейреса), себестоимости, производительности труда и что они показывают? Напишите их формулы.

9. Когда возникает необходимость преобразования агрегатного индекса цен в средний арифметический и средний гармонический; каким образом происходят такие преобразования? Покажите на примерах.

10. Что называется индексом переменного состава, как он исчисляется и что он характеризует? Напишите его формулу.

11. Какой индекс называется индексом постоянного состава, как он исчисляется и что он характеризует? Напишите его формулу.

12. Что характеризует индекс структурных сдвигов и как он исчисляется?

13. Какая взаимосвязь существует между индексами постоянного, переменного состава и структурных сдвигов?

14. Как строятся базисные и цепные индексы и какая между ними существует взаимосвязь?

15. Что представляют собой индексы с постоянными и переменными весами?

16. В чем выражается взаимосвязь индексов цен, физического объема и товарооборота, как практически она используется?

17. Какая система взаимосвязанных индексов используется при анализе себестоимости, физического объема и затрат в производстве?

18. Что представляют собой территориальные индексы?

Примеры решения задач

1. Если цена товара А в текущем периоде составляла 30 руб., а в базисном году 25 руб., то индивидуальный индекс цены I_p = 30: 25 = 1,2, или 120,0%.

$$i_p = \frac{p_1}{p_0}.$$

В данном примере цена товара А возросла по сравнению с базисным уровнем в 1,2 раза, или на 20%.

2. Имеются следующие данные о реализации плодово-ягодной продукции в области (табл. 7.1).

Таблица 7.1

Реализация плодово-ягодной продукции в области

Наименование товара	Июль		Август		Расчетные графы, руб.		
	Цена за 1 кг, руб., p_0	Продано т, q_0	Цена за 1 кг, руб., p_1	Продано т, q_0	$p_0 q_0$	$p_1 q_1$	$p_0 q_1$
Черешня	12	18	12	15	216	180	180
Персики	11	22	10	27	242	270	297
Виноград	9	20	7	24	180	168	216
Итого	*	*	*	*	638	618	693

1. Индекс товарооборота:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{618}{638} = 0.969,\ \textit{или}\ 96{,}9\%.$$

Мы получили, что товарооборот в целом по данной товарной группе в текущем периоде по сравнению с базисным уменьшился на 3,1% (100-96,9%).

2. Общий индекс цен:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{618}{693} = 0.892,\ \textit{или}\ 89{,}2\%.$$

Цены в августе по сравнению с июлем в среднем снизились на 10,8%.

Разность числителя и знаменателя будет показывать величину экономии или перерасхода покупателей от изменения цен:

$$E = \sum p_1 q_1 - \sum p_0 q_1 = 618 - 693 = -75\ \textit{тыс.руб.}$$

3. Индекс физического объема реализации:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{693}{638} = 1.086,\ \textit{или}\ 108{,}6\%.$$

Т.о., физический объем реализации увеличился на 8,6%.

Используя взаимосвязь индексов, проверим правильность вычислений:

$$I_{pq} = I_p\ I_q = 0{,}892 * 1{,}086 = 0{,}969 \text{ или } 96{,}9\%.$$

4. Индексы постоянного, переменного состава и структурных сдвигов.

Проведем анализ изменения цен реализации товара А в двух регионах (таблица 7.2).

Таблица 7.2

Регион	Июнь		Июль		Расчетные графы, руб.		
	Цена, руб., p_o	Продано, шт.,q_o	Цена, руб., p_1	Продано, шт.,q_1	p_oq_o	p_1q_1	p_oq_1
1	12	10 000	13	18 000	120 000	234 000	216 000
2	17	20 000	19	9 000	340 000	171 000	153 000
Итого	*	30 000	*	27 000	460 000	405 000	369 000

1. Индекс переменного состава:

$$I_p^{\textit{ПС}} = \frac{\sum p_1 q_1}{\sum q_1} \div \frac{\sum p_0 q_0}{\sum q_0} = \frac{405000}{27000} \div \frac{460000}{30000} = 15.00 \div 15.33 = 0.0978.$$ Из табл. 7.2. видно, что цена в каждом регионе в июле по сравнению с июнем возросла. В целом же средняя цена снизилась на 2,2% (97,8 – 100). Несоответствие объясняется влиянием изменения структуры реализации товаров по регионам: в июне по более высокой цене продавали товара вдвое больше, в июле же ситуация изменилась.

2. Индекс структурных сдвигов:

$$I\textit{сст} = \frac{\sum p_0 q_2}{\sum q_1} \div \frac{\sum p_0 q_0}{\sum q_0} = \frac{369000}{27000} \div \frac{460000}{30000} = 0.891.$$

Вывод: за счет структурных сдвигов цены снизились.

3. Рассчитанный индекс постоянного состава равен 1,098, или 109,8%. Отсюда следует, что, если бы структура реализации товара по регионам не изменилась, средняя цена возросла бы на 9,8%.

4. Рассчитать средний арифметический индекс по следующим данным:

Товар	Реализация в базисном периоде, руб. $q_0 p_0$	Изменение физического объема реализации в текущем периоде по сравнению с базисным, % $i_q \cdot 100\% - 100\%$	Расчетные графы	
			i_q	$i_q \cdot q_0 p_0$
Мандарины	46000	-6,4	0,936	43056
Грейпфруты	27000	-8,2	0,918	24786
Апельсины	51000	+1,3	1,013	51663
Итого	124000	-	-	119505

Решение:

$$I_p = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{119505}{124000} = 0.964, \ \textit{или} \ 96,4\%.$$

Физический объем реализации данных товаров в среднем снизился на 3,6%.

5. Рассчитать территориальный индекс, если известны данные о ценах и объемах реализации товаров по двум регионам:

Товар	Регион А		Регион Б		Расчетные графы		
	Цена, руб., p_a	Реализовано, т., q_a	Цена, руб., p_b	Реализовано, т., q_b	$Q = q_a + q_b$	$p_a Q$	$p_b Q$
1	11,0	30	12,0	35	65	715,0	780,0
2	8,5	45	9,0	50	95	807,5	855,0
3	17,0	15	16,0	90	105	1785,0	1680,0
Итого	*	*	*	*	*	3307,5	3315,0

$$I_{pb/a} = \frac{\sum p_b Q}{\sum p_a Q} = \frac{3315.0}{3307.5} = 1{,}002, \ \textit{или} \ 100,2\%.$$

Цены в регионе В на 0,2% превышают цены в регионе А.

Задания для самостоятельной работы

1. Имеются данные о ценах на уголь и объемах его производства в РФ во 2 квартале 2021 г.:

Месяц	Цена за 1 т, тыс. руб.	Произведено, млн. т
Апрель	120	23,2
Май	121	20,2
Июнь	116	18,7

При условии 100-% реализации угля в каждом месяце определите цепные и базисные индексы цен, физического объема реализации и товарооборота. Проверьте взаимосвязь цепных и базисных индексов.

2. Известны следующие данные о реализации фруктов предприятиями розничной торговли:

Товар	Цена за 1 кг, руб.		Товарооборот, тыс. руб.	
	июль	август	июль	август
Яблоки	8	6	143,5	167,1
Груши	11	10	38,9	45,0

Рассчитайте сводные индексы: а) товарооборота; б) цен; в) физического объема реализации. Определите абсолютную величину экономии покупателей от снижения цен.

3. Строительно-производственная деятельность двух ДСК характеризуется следующими данными:

ДСК	Построено жилья, тыс. $м^2$		Себестоимость 1 $м^2$, млн. руб.	
	2011	2012	2011	2012
ДСК – 1	53	68	1,5	1,7
ДСК – 2	179	127	1,7	1,9

Рассчитать: индексы себестоимости переменного и постоянного составов, а также структурных сдвигов. Объясните результаты расчетов.

Тема 8. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ И ПРОЦЕССОВ

1. Измерение связи

Все явления общественной жизни взаимосвязаны и взаимообусловлены. Задача статистики состоит в том, чтобы выявить и измерить связи и зависимости между изучаемыми явлениями.

Взаимосвязанные признаки подразделяются на *факторные* (под их воздействием изменяются другие, зависящие от них признаки) и *результативные.*

Связи по степени тесноты могут быть *функциональными* (при которых определенному значению факторного признака соответствует строго определенное значение результативного признака; для выявления такой связи достаточно одного наблюдения), *статистическими* (когда одному и тому же значению факторного признака могут соответствовать несколько значений результативного признака; эти связи проявляются в массе случаев и при том – в среднем). Функциональные связи иначе называются полными, а статистические – неполными или корреляционными.

Корреляционная зависимость проявляется только в средних величинах и выражает числовое соотношение между ними в виде тенденции к возрастанию или убыванию одной переменной величины при возрастании и убывании другой.

Корреляционная связь является свободной, неполной и неточной связью. Например, себестоимость величины продукции зависит от уровня производительности труда: чем выше производительность труда, тем ниже себестоимость. Но себестоимость зависит также и от ряда других факторов: стоимости сырья и материалов, топлива, электроэнергии, их расхода на единицу продукции, цеховых общезаводских расходов и т.д. Поэтому нельзя утверждать, что при повышении производительности труда, допустим, на 10% себестоимость снизится также на 10%. Может случиться, что, несмотря на рост производительности труда, себестоимость не только не снизится, но даже несколько повысится, если на нее окажут более сильное влияние действующие в обратном направлении другие факторы.

По направлению различают *прямую* и *обратную* связь. Если с увеличением аргумента x функция y также увеличивается без всяких единичных исключений, то такая связь называется полной прямой связью. Если с увеличением аргумента х функция у уменьшается без всяких единичных исключений, то такая связь называется полной обратной. Кроме того, в виде исключений, которые, однако, не нарушают общей тенденции, встречается частичная связь – прямая или обратная. Когда признаки варьируют независимо друг от друга, говорят о полном отсутствии связи.

По аналитическому выражению корреляционная связь может быть *прямолинейной* и *криволинейной.* Прямолинейной называется связь, когда величина явления изменяется приблизительно равномерно в соответствии с изменением величины влияющего фактора. Математически прямолинейная связь может быть выражена уравнением прямой:

$$V=a_0 +a_1x, \qquad (8.1)$$

которое называется *линейным уравнением регрессии.*

Если происходит неравномерное изменение явления в связи с изменением величины влияющего фактора, то такая связь называется криволинейной. Математически криволинейная зависимость может быть выражена уравнением криволинейной связи. В экономическом анализе для ее выражения часто *пользуются уравнением параболы второго порядка:*

$$V = a_0 + a_1 x + a_2x^2. \qquad (8.2)$$

Уравнение криволинейной связи может быть выражено и в виде *дробной функции:*

$$V = \frac{a_0}{x} + a_1; \qquad (8.3)$$

показательной функции:

$$V = a_0 a_1^x. \qquad (8.4)$$

Однако корреляционные связи могут быть выражены лишь приблизительно, в то время как функциональные связи имеют точное аналитическое выражение.

2. Основные приемы изучения взаимосвязей

Для изучения, измерения и количественного выражения взаимосвязей между явлениями статистикой применяются различные методы, такие как: метод сопоставления параллельных рядов, балансовый, графический, методы аналитических группировок, дисперсионного и корреляционного анализа.

Метод параллельных рядов заключается в том, что полученные в результате сводки и обработки материалы располагают в виде параллельных рядов и сопоставляют их между собой для установления характера и тесноты связи.

Балансовый метод состоит в том, что данные взаимосвязанных показателей изображаются в виде таблицы и располагаются таким образом, чтобы итоги между отдельными ее частями были равны, т.е. чтобы был баланс. Балансовый метод используется для характеристики взаимосвязи между производством и распределением продуктов, денежными доходами и расходами населения и т.д. Почти все внутренние и внешние хозяйственные связи выражаются в виде балансов.

Метод аналитических группировок. Сущность метода аналитических группировок состоит в том, что единицы статистической совокупности группируются, как правило, по факторному признаку, и для каждой группы рассчитывается средняя или относительная величина по результативному признаку. Затем изменения средних или относительных значений результативного признака сопоставляются с изменениями факторного признака для выявления характера связи между ними.

Дисперсионный анализ дает прежде всего возможность определить значение систематической и случайной вариаций в общей вариации, а также установить роль интересующего нас фактора в изменении результативного признака.

Для характеристики тесноты корреляционной связи между признаками в аналитических группировках межгрупповую дисперсию сопоставляют с общей. Это сопоставление называется *корреляционным отношением* и обозначается:

$$\eta^2 = \frac{\delta^2}{\sigma^2}. \qquad (8.5)$$

Оно характеризует долю вариации результативного признака, вызванной действием факторного признака, положенного в основание группировки. Корреляционное отношение по своему абсолютному значению колеблется в пределах от 0 до 1. Чем ближе корреляционное отношение к 1, тем большее влияние оказывает факторный признак на результативный. Если же факторный признак не влияет на результативный, то вариация, обусловленная им, будет равна нулю ($\delta^2 = 0$) и корреляционное отношение также будет равно нулю ($\eta^2 = 0$), что говорит о полном отсутствии связи. И наоборот, если результативный признак изменяется только под воздействием одного факторного признака, то вариация, обусловленная этим признаком, будет равна общей вариации ($\eta^2 = \eta^2$) и корреляционное отношение будет равно единице ($\eta^2 = 1$), что говорит о существовании полной связи.

Дисперсионный анализ позволяет не только определить роль случайной и систематической вариаций в общей вариации, но и оценить достоверность вариации, обнаруженной методом аналитических группировок. Определение достоверности вариации дает возможность с заданной степенью вероятности установить, вызвана ли межгрупповая вариация признаком, положенным в основание группировки, или она является результатом действия случайных причин. Для оценки существенности корреляционного отношения пользуются критическими значениями корреляционного отношения η^2 при разных уровнях вероятности или значимости α.

Уровень значимости – это достаточно малое значение вероятности, отвечающее событиям, которые в данных условиях исследования будут считаться практически невозможными. Появление такого события считается указанием на неправильность начального предположения. Чаще всего пользуются уровнями $\alpha = 0{,}05$ или $\alpha = 0{,}01$. Критические значения корреляционного отношения содержатся в специальных таблицах.

В этих таблицах распределение η^2 при случайных выборках зависит от числа степеней свободы факторной и случайной дисперсии. Число степеней свободы факторной дисперсии равно:

$$K_1=m-1, \tag{8.6}$$

где m – число групп, а для случайной дисперсии:

$$K_1=n-m, \tag{8.7}$$

где n – число вариант; m – число групп.

При проверке существенности связи чаще пользуются критерием Фишера, потому что при больших числах степеней свободы его табличные значения мало изменяются в отличие от корреляционного отношения, которое требует более громоздких таблиц. *Критерий Фишера* представляет собой отношение межгрупповой дисперсии к средней из внутригрупповых дисперсий, исчисленных с учетом числа степеней свободы:

$$F=\frac{\delta^2}{\sigma^2}*\frac{n-m}{m-1}. \tag{8.8}$$

Для этих отношений составляются таблицы, по которым можно определить, какая величина F при данном числе степеней свободы по факторной вариации (K_1) и остаточной вариации (K_2) дает основание утверждать с определенной вероятностью (например, 0,95; 0,99), что положенный в основание группировки признак является существенным, или не дает такого основания, и, следовательно, группировочный признак является несущественным.

Зная корреляционное отношение, можно определить критерий Фишера по следующей формуле:

$$F = \frac{\eta^2}{1-\eta^2} * \frac{n-m}{m-1}. \tag{8.9}$$

Подобный дисперсионный анализ может проводиться при группировке по одному факторному признаку или при комбинационной группировке по двум и более факторам. В таком случае необходима оценка достоверности влияния не только каждого положенного в основание группировки фактора в отдельности, но и результата их взаимодействия. Последний определяется как разность между эффектом совместного влияния двух группировочных признаков и суммой эффектов влияния каждого из этих факторных признаков, взятых в отдельности. Это осложняет расчеты суммы квадратов отклонений и числа степеней свободы вариации. Но сам принцип дисперсионного анализа, основанный на сопоставлении факторной дисперсии со случайной для оценки достоверности результатов статистической группировки, остается применим независимо от числа признаков группировки.

3. Корреляционный анализ

Корреляционный анализ решает две основные задачи.

Первая задача заключается в определении формы связи, т.е. в установлении математической формы, в которой выражается данная связь. Это очень важно, так как от правильного выбора формы связи зависит конечный результат изучения взаимосвязи между признаками.

Вторая задача состоит в измерении тесноты, т.е. меры связи между признаками, с целью установить степень влияния данного фактора на результат. Она решается математически путем определения параметров корреляционного уравнения.

Затем проводятся оценка и анализ полученных результатов при помощи специальных показателей корреляционного метода (коэффициентов детерминации, линейной и множественной корреляции и т.д.), а также проверка существенности связи между изучаемыми признаками.

Определяющая роль в выборе формы связи между явлениями принадлежит теоретическому анализу. Так, например, чем больше размер основного капитала предприятия (факторный признак), тем больше при прочих равных условиях оно выпускает продукции (результативный признак). С ростом факторного признака здесь, как правило, равномерно растет и результативный, поэтому зависимость между ними может быть выражена уравнением прямой $\boldsymbol{Y = a_0 + a_1 x}$, которое называется *линейным уравнением регрессии.*

Параметр a_1 называется *коэффициентом регрессии* и показывает, насколько в среднем отклоняется величина результативного признака y при отклонении величины факторного признаках на одну единицу. При х = 0 $a_0 = Y$. Увеличение количества внесенных удобрений приводит, при прочих равных условиях, к росту урожайности, но чрезмерное внесение их без изменения других элементов к дальнейшему повышению урожайности не приводит, а наоборот, снижает ее. Такая зависимость может быть выражена уравнением параболы $\boldsymbol{Y = a_0 + a_1 x + a_2 x^2}$.

Параметр $\boldsymbol{a_2}$ характеризует степень ускорения или замедления кривизны параболы, и при $\mathbf{a_2} > 0$ парабола имеет минимум, а при $a_2 < 0$ – максимум. Параметр $\mathbf{a_1}$ характеризует угол наклона кривой, а параметр $\mathbf{a_0}$ – начало кривой.

Однако с помощью теоретического анализа не всегда удается установить форму связи. В таких случаях приходится только предполагать наличие определенной формы связи. Проверить эти предположения можно при помощи графического анализа, который используется для выбора формы связи между явлениями, хотя графический метод изучения связи применяется и самостоятельно.

Применение методов корреляционного анализа дает возможность выражать связь между признаками аналитически – в виде уравнения – и придавать ей количественное выражение. Рассмотрим применение приемов корреляционного анализа на конкретном примере.

Допустим, что между стоимостью основного капитала и выпуском продукции существует прямолинейная связь, которая выражается *уравнением прямой* $\boldsymbol{Y = a_0 + a_1 x}$. Необходимо найти параметры $\boldsymbol{a_0}$ и $\boldsymbol{a_1}$, что позволит определить теоретические значения Y для разных значений x. Причем $\boldsymbol{a_0}$ и $\boldsymbol{a_1}$ должны быть такими, чтобы было достигнуто максимальное приближение к первоначальным (эмпирическим) значениям теоретических значений Y. Эта задача решается при помощи способа наименьших квадратов, основное условие которого сводится к определению параметров $\boldsymbol{a_0}$ и $\boldsymbol{a_1}$ таким образом, чтобы $\sum (y_i - Y)^2 = \min$.

Математически доказано, что условие минимума обеспечивается, если параметры $\boldsymbol{a_0}$ и $\boldsymbol{a_1}$ определяются при помощи *системы* двух *нормальных уравнений,* отвечающих требованию метода наименьших квадратов:

$$\begin{cases} \sum y = na_0 + a_1 \sum x, \\ \sum xy = a_0 \sum x + a_1 \sum x^2. \end{cases} \qquad (8.10)$$

Первое уравнение есть сумма всех первоначальных уравнений. Второе получается умножением обеих частей уравнения прямой на один и тот же множитель. Математически доказано, что условие $\sum (y_i - Y)^2 = \min$

соблюдается, если в качестве такого множителя принять значение факторного признака, т.е. если уравнение прямой умножить на х.

Кроме рассмотренных функций связи в экономическом анализе часто применяются степенная, показательная и гиперболическая функции.

Степенная функция имеет вид $Y = a_0 \; x^{a1}$. Параметр a_1 степенного уравнения называется показателем эластичности и указывает, на сколько процентов изменится y при возрастании x на 1%. При $x = 1 \quad a_0 = Y$.

Для определения параметров степенной функции вначале ее приводят к линейному виду путем логарифмирования: $\lg y = \lg a_0 + a_1 \lg x_1$, а затем строят систему нормальных уравнений:

$$\begin{cases} \sum \lg y = n \lg a_0 + a_1 \sum \lg x; \\ \sum \lg y \lg x = \lg a_0 \sum \lg x + a_1 \sum \lg x^2. \end{cases} \qquad (8.11)$$

Решив систему двух нормальных уравнений, находят логарифмы параметров логарифмической функции a_0 и a_1, а затем и сами параметры a_0 и a_1. При помощи степенной функции определяют, например, зависимость между фондом заработной платы и выпуском продукции, затратами труда и выпуском продукции и т.д.

Если факторный признак х растет в арифметической прогрессии, а результативный y – в геометрической, то такая зависимость выражается *показательной функцией* $Y = a_0 * a_1{}^2$. Для определения параметров показательной функции ее также вначале приводят к линейному виду путем логарифмирования: $\lg y = \lg a_0 + x \lg a_1$, а затем строят систему нормальных уравнений:

$$\begin{cases} \sum \lg y = n \lg a_0 + \lg a_1 \sum x; \\ \sum x \lg y = \lg a_0 \sum x + \lg a_1 \sum x^2. \end{cases} \qquad (8.12)$$

Вычислив соответствующие данные и решив систему двух нормальных уравнений, находят параметры показательной функции a_o *и* a_1..

В ряде случаев обратная связь между факторным и результативным признаками может быть выражена *уравнением гиперболы:*

$$Y = a_0 + \frac{a_1}{x}. \qquad (8.13)$$

И здесь задача заключается в нахождении параметров a_0 и a_t при помощи системы двух нормальных уравнений:

$$\begin{cases} \sum y = na_0 + a_1 \sum \frac{1}{x} \\ \sum y \frac{1}{x} = a_0 \sum \frac{1}{x} + a_1 \sum \frac{1}{x^2}. \end{cases} \qquad (8.14)$$

При помощи гиперболической функции изучают, например, связь между выпуском продукции и себестоимостью, уровнем издержек обращения (в % к товарообороту) и товарооборотом в торговле, сроками уборки и урожайностью и т.д.

Таким образом, применение различных функций в качестве уравнения связи сводится к определению параметров уравнения по способу наименьших квадратов при помощи системы нормальных уравнений.

В малых совокупностях значение коэффициента регрессии подвержено случайным колебаниям. Поэтому возникает необходимость в определении достоверности коэффициента регрессии. Достоверность коэффициента регрессии определяется так же, как и в выборочном наблюдении, т.е. устанавливаются средняя и предельная ошибки для выборочной средней и доли. *Средняя ошибка коэффициента регрессии* определяется по формуле:

$$\mu_B = \sqrt{\frac{\sigma_0^2}{\sigma^2(n-2)}}, \qquad (8.15)$$

где σ_0^2 – случайная дисперсия; σ^2 – общая дисперсия, n – число коррелируемых пар.

4. Измерение тесноты связи

Чтобы измерить тесноту прямолинейной связи между двумя признаками, пользуются *парным коэффициентом корреляции,* который обозначается r.

Так как при корреляционной связи имеют дело не с приращением функции в связи с изменением аргумента, а с сопряженной вариацией результативных и факторных признаков, то определение тесноты связи, по существу, сводится к изучению этой сопряженности, т.е. того, в какой мере отклонение от среднего уровня одного признака сопряжено с отклонением другого. Это значит, что при наличии *полной прямой связи* все значения *(x – X)* и (y – *Y)* должны иметь одинаковые знаки, при *полной обратной* – разные, при частичной связи знаки в преобладающем числе случаев будут совпадать, а при отсутствии связи – совпадать примерно в равном числе случаев.

Для оценки существенности коэффициента корреляции пользуются специально разработанной таблицей критических значений r.

Коэффициент корреляции r_{xy}. применяется только в тех случаях, когда между явлениями существует прямолинейная связь.

$$r_{xy} = \frac{r\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x - \left(\sum x\right)\cdot\left[n\sum y - \left(\sum y\right)\right]\right]}}. \qquad (8.16)$$

Если же связь криволинейная, то пользуются *индексом корреляции,* который рассчитывается по формуле:

$$R = \sqrt{1 - \frac{\sum (y - Y)^2}{\sum (y - \bar{y})^2}}, \qquad (8.17)$$

где y – первоначальные значения; $\bar{y}$ – среднее значение; Y – теоретические (выравненные) значения переменной величины.

Показатель остаточной, случайной дисперсии определяется по формуле:

$$\sigma_0^2 = \frac{\sum (y - Y)^2}{n}. \quad (8.18)$$

Она характеризует размер отклонений эмпирических значений результативного признака у от теоретических *Y*, т.е. случайную вариацию.

Общая дисперсия

$$\sigma^2 = \frac{\sum [y - \bar{y}]^2}{n}, \quad (8.19)$$

характеризует размер отклонений эмпирических значений результативного признака y от $\bar{y}$, т.е. общую вариацию. Отношение случайной дисперсии к общей характеризует долю случайной вариации в общей вариации, а

$$1 - \frac{\sum (y - Y)^2}{\sum (y - \bar{y})^2}, \quad (8.20)$$

есть не что иное, как *доля факторной вариации*

$$\sigma_y^2 = \frac{\sum (Y - \bar{y})^2}{n}, \quad (8.21)$$

в общей, потому что по правилу сложения дисперсий общая дисперсия равна сумме факторной и случайной дисперсий:

$$\sigma = \sigma_Y^2 + \sigma_0^2. \quad (8.22)$$

Подставим в формулу индекса корреляции соответствующие обозначения случайной, общей и факторной дисперсий и получим:

$$\boldsymbol{R} = \sqrt{1 - \frac{\sum (y - Y)^2}{\sum (y - \bar{y})^2}} = \sqrt{1 - \frac{\sigma_0^2}{\sigma^2}} = \sqrt{\frac{\sigma_Y^2}{\sigma^2}}. \quad (8.23)$$

Таким образом, индекс корреляции характеризует долю факторной вариации в общей:

$$R = \sqrt{\frac{\sigma_Y^2}{\sigma^2}}, \quad (8.24)$$

однако с той лишь разницей, что вместо групповых средних берутся теоретические значения Y.

Индекс корреляции по своему абсолютному значению колеблется в пределах от 0 до 1. При функциональной зависимости случайная вариация $\sum(y-Y)^2=0$, индекс корреляции равен 1. При отсутствии связи R = 0, потому что Y = *y*.

Коэффициент корреляции является мерой тесноты связи только для линейной формы связи, а индекс корреляции – и для линейной, и для криволинейной. При прямолинейной связи коэффициент корреляции по своей абсолютной величине равен индексу корреляции:

$$|r| = R\,. \tag{8.25}$$

Если индекс корреляции возвести в квадрат, то получим *коэффициент детерминации*:

$$R^2 = \frac{\sigma_y^2}{\sigma^2}\,. \tag{8.26}$$

Он характеризует роль факторной вариации в общей вариации и по построению аналогичен корреляционному отношению η^2. Как и корреляционное отношение, коэффициент детерминации R^2 может быть исчислен при помощи дисперсионного анализа, так как дисперсионный анализ позволяет расчленить общую дисперсию на факторную и случайную. Однако при дисперсионном анализе для разложения дисперсии пользуются методом группировок, а при корреляционном анализе – корреляционными уравнениями.

Коэффициент детерминации является наиболее конкретным показателем, так как он отвечает на вопрос о том, какая доля в общем результате зависит от фактора, положенного в основание группировки.

При прямолинейной парной связи факторную дисперсию можно определить без вычисления теоретических значений Y по следующей формуле:

$$\sigma^2 = \frac{1}{n}\left(a_0 \sum y + a_1 \sum xy\right) - y^{-2}\,. \tag{8.27}$$

5. Множественная корреляция

До сих пор мы рассматривали корреляционные связи между двумя признаками: результативным (y) и факторным *(x)*. Например, выпуск продукции зависит не только от размера основного капитала, но и от уровня квалификации рабочих, состояния оборудования, обеспеченности и качества сырья и материалов, организации труда и т.д. В связи с этим возникает необходимость в изучении, измерении связи между результативным признаком, двумя и более факторными. Этим занимается множественная корреляция.

Множественная корреляция решает три задачи. Она определяет:

1) форму связи;

2) тесноту связи;

3) влияние отдельных факторов на общий результат.

Определение формы связи сводится обычно к отысканию уравнения связи у с факторами *x,z,w,...y*. Так, линейное уравнение зависимости результативного признака от двух факторных определяется по формуле

$$Y_{xz} = a_0 + a_1 x + a_2 z\,. \tag{8.28}$$

Для определения параметров a_0, a_1 и a_2, по способу наименьших квадратов необходимо решить следующую систему трех нормальных уравнений:

$$\begin{cases} \sum y = a_0 n + a_1 \sum x + a_2 \sum z; \\ \sum yx = a_0 \sum x + a_1 \sum x^2 + a_2 \sum xy; \\ \sum yx = a_0 \sum z + a_1 \sum xz + a_2 \sum z^2. \end{cases} \quad (8.29)$$

При определении тесноты связи для множественной зависимости пользуются коэффициентом множественной (совокупной) корреляции, предварительно исчислив коэффициенты парной корреляции. Так, при изучении связи между результативным признаком y и двумя факторными признаками – x и z нужно предварительно определить тесноту связи между y и x, между y и z, т.е. вычислить коэффициенты парной корреляции, а затем для определения тесноты связи результативного признака от двух факторных исчислить *коэффициент множественной корреляции* по следующей формуле:

$$R_{y(x.z)} = \sqrt{\frac{r_{xy}^2 + r_{zy}^2 - 2r_{xy} * r_{zy} * r_{xz}}{1 - r_{xz}^2}}, \quad (8.30)$$

где r_{xy}, r_{zy}, r_{xz} – парные коэффициенты корреляции.

Коэффициент множественной корреляции колеблется в пределах от 0 до 1. Чем он ближе к 1, тем в большей мере учтены факторы, определяющие конечный результат.

Если коэффициент множественной корреляции возвести в квадрат, то получим совокупный коэффициент детерминации, который характеризует долю вариации результативного признака y под воздействием всех изучаемых факторных признаков.

Совокупный коэффициент детерминации, как и при парной корреляции, можно исчислить по следующей формуле:

$$R^2 = \frac{\sigma_Y^2}{\sigma_y^{\ 2}}, \quad (8.31)$$

где σ_Y^2 – дисперсия факторных признаков, σ_y^2 – дисперсия результативного признака. Однако вычисление теоретических значений Y при множественной корреляции и сложно, и громоздко. Поэтому *факторную дисперсию* η_Y^2 исчисляют по следующей формуле:

$$\sigma^2 = \frac{1}{n}\left(a_0 \sum y + a_1 \sum xy + a_2 \sum zy\right) - y^{-2}. \quad (8.32)$$

Проверка существенности связи при множественной корреляции, по сути, ничем не отличается от проверки при парной корреляции.

Поскольку факторные признаки действуют не изолированно, а во взаимосвязи, то может возникнуть задача определения тесноты связи между результативным признаком и одним из факторных при постоянных значениях прочих факторов. Она решается при помощи *частных коэффициентов корреляции.* Например, при линейной связи частный коэффициент корреляции между x и y при постоянном z рассчитывается по следующей формуле:

$$r_{xy(z)} = \frac{r_{xy} - r_{zy} r_{xz}}{\sqrt{\left(1 - r_{zy}^2\right)\left(1 - r_{xz}^2\right)}}. \qquad (8.33)$$

В настоящее время на практике широкое распространение получил многофакторный корреляционный анализ.

6. Методы измерения тесноты связи

Измерение тесноты связи при помощи дисперсионного и корреляционного анализа связано с определенными сложностями и требует громоздких вычислений. Для ориентировочной оценки тесноты связи пользуются приближенными показателями, не требующими сложных, трудоемких расчетов. К ним относятся: коэффициент корреляции знаков Фехнера, коэффициент корреляции рангов, коэффициент ассоциации и коэффициент взаимной сопряженности.

Коэффициент корреляции знаков основан на сопоставлении знаков отклонений от средней и подсчете числа случаев совпадения и несовпадения знаков, а не на сопоставлении попарно размеров отклонений индивидуальных значений факторного и результативного признаков от средней (х – х) и (у – у):

$$i = \frac{u - v}{u + v}, \qquad (8.34)$$

где u – число пар с одинаковыми знаками отклонений x и у от х и у; v – число пар с разными знаками отклонений х и у от х и у. Коэффициент корреляции знаков колеблется в пределах от -1 до +1. Чем ближе коэффициент к 1, тем теснее связь. Если u > v, то i > 0, так как число согласованных знаков больше, чем несогласованных, и связь прямая. При u < v имеем i < 0, потому что число несогласованных знаков больше, чем согласованных, и связь обратная. Если u = v, то $I=0$, и связи нет.

Коэффициент кореляции рангов исчисляется не по первичным данным, а по рангам (порядковым номерам), которые присваиваются всем значениям изучаемых признаков, расположенным и порядке их возрастания. Если значения признака совпадают, то определяется средний ранг путем деления суммы рангов на число значений. Коэффициент корреляции рангов определяется по формуле:

$$p = 1 - \frac{s\sum d^2}{n\left(n^2 - 1\right)}, \qquad (8.35)$$

где d^2 – квадрат разности рангов для каждой единицы, d = x – у; n – число рангов; s – средний ранг.

Коэффициент корреляции рангов также колеблется в пределах от -1 до +1. Если ранги по обоим признакам совпадают, то $\eta d^2 = 0$, значит, p = 1 и, следовательно, связь полная прямая. Если p = -1, связь полная обратная, при р = 0 связь между признаками отсутствует.

Коэффициент ассоциации применяется для установления меры связи между двумя качественными альтернативными признаками. Для его вычисления строится комбинационная четырех клеточная таблица, которая выражает связь между двумя альтернативными явлениями.

Коэффициент ассоциации рассчитывается по формуле:

$$A = \frac{ad - bc}{\sqrt{(a+d)(c+d)(a+c)(b+d)}}. \qquad (8.36)$$

Коэффициент ассоциации также изменяется от -1 до +1. Чем A ближе к единице, тем сильнее связаны между собой изучаемые признаки. При $ad > bc$ связь прямая, а при $ad < bc$ связь обратная, при $ad = bc$ $A = 0$ связь отсутствует.

В тех случаях, когда требуется установить связь между качественными признаками, каждый из которых состоит из трех и более групп, применяется *коэффициент взаимной сопряженности*.

Различия между условным и безусловным распределением свидетельствуют о влиянии факторного признака на распределение совокупности по результативному признаку, т.е. о наличии связи между факторным и результативным признаками, а чем больше эти различия, тем в большей мере признаки связаны между собой, тем теснее связь между ними.

Коэффициент взаимной сопряженности определяется по следующей формуле:

$$C = \sqrt{\frac{x^2}{n(m_1-1)(m_2-1)}}, \qquad (8.37)$$

где n – число единиц совокупности; m_1 и m_2 – число групп по первому и второму признакам; χ^2 – *показатель абсолютной квадратической сопряженности Пирсона,* характеризующий близость условных распределений к безусловным, который, как и критерий χ^2, исчисляется по формуле:

$$\chi^2 = \sum f \sum \frac{(\omega_{ij} - \omega_j)^2}{\omega_j}, \qquad (8.38)$$

где ω_{ij}. – частости условного распределения в i-й строке; ω_j – частости безусловного распределения; j – номер столбца.

Если признаки независимы, то $w_{ij}.. = \omega_j.$, откуда $\chi^2 = O$ и, значит, C = 0. Если же связь функциональная, то коэффициент взаимной сопряженности будет равен единице.

Вопросы для теоретического контроля знаний

1. Определите понятие «статистическая связь».
2. Какие вы знаете формы и методы изучения статистической связи?
3. Укажите основные задачи корреляционно – регрессионного анализа.
4. Каковы определенные значения корреляционного отношения?
5. Что такое множественная корреляция?

6. В чем состоит различие между корреляционной и функциональной связью?

7. Какие основные проблемы решает исследователь при изучении корреляционной зависимости?

8. Какие показатели являются мерой тесноты связи между двумя признаками?

9. Как оценить существенность линейного коэффициента корреляции?

10. Как определяется парная регрессия на основе метода наименьших квадратов и метода группировок?

11. Что такое множественная регрессия?

12. Какое значение имеет расчет индекса корреляции?

13. Какие решения можно принимать на основе уравнения регрессии?

14. Укажите собственно – корреляционные параметрические методы изучения связи. Оценка существенности корреляции.

15. Раскройте суть методов изучения связи социальных явлений.

16. Какие непараметрические показатели связи вы знаете?

17. Назовите ранговые коэффициенты связи.

18. Для чего рассчитываются частные коэффициенты корреляции?

19. Как подходить к отбору факторов для включения их в уравнение множественной регрессии?

Примеры решения задач

1. По данным о сумме активов и кредитных вложений коммерческих банков необходимо определить направление и тесноту связи между признаками

Таблица 8.1

Распределение кредитных вложений и сумм активов коммерческих банков

Банк	Кредитные вложения, млрд. руб., х	Сумма активов, млрд. руб., у	x^2	ху	y_x
1	311	518	96721	161098	1140,6
2	658	1194	432964	785652	1502,5
3	783	2941	613089	2302803	1632,9
4	1142	1865	1304164	2129830	2007,3
5	1319	1997	1739761	2634043	2191,9
6	1962	3066	3849444	6015492	2862,4
7	2496	3176	6230016	7927296	3419,4
Итого	8671	14757	14266159	21956214	14757,0

Сопоставив полученные данные ряда данных х и у, можно наблюдать наличие прямой зависимости между признаками, когда увеличение кредитных вложений увеличивает сумму активов коммерческих банков. Исходя из этого можно сделать вывод, что связь между признаками прямая и ее можно описать уравнением прямой.

Определим параметры уравнения прямой на основе метода наименьших квадратов, решив систему нормальных уравнений:

$$\begin{cases} \sum y = na_0 + a_1 \sum x, \\ \sum xy = a_0 \sum x + a_1 \sum x^2. \end{cases}$$

Получаем:

$7a_0 + 8671\ a_1 = 14757$;

$8671a_0 + 14266159\ a_1\ 21956214$.

$a_0 = 816{,}2878$; $a_1 = 1{,}0429$.

Аналитическая связь описывается уравнением:

$\bar{y}_x = a_0 + a_1 x$.

Исходя из полученных результатов получаем:

$\bar{y}_x = 816{,}2878 + 1{,}0429x$.

Следовательно, с увеличением кредитных вложений на 1 млрд. руб. сумма активов возрастет на 1,0429 млрд. руб.

2. По данным задачи 2 оценить тесноту связи между стоимостью активов (у) и кредитными вложениями (x_1), используя различные модификации коэффициента корреляции. Проверить его значимость.

Решение:

$$r_{yx1} = \frac{\overline{yx_1} - \bar{y} \cdot \bar{x}_1}{\sigma_{x1}\sigma_y};$$

$$\overline{yx_1} = \frac{\overline{yx_1}}{n} = \frac{21956214}{7} = 3136602;$$

$$\bar{y} = \frac{\sum y}{n} = \frac{14757}{7} = 2108, \bar{x}_1 = \frac{\sum x_1}{n} = \frac{8671}{7} = 1238.7;$$

$$\sigma_{x1} = \sqrt{\bar{x}_1^2 - (\bar{x}_1)^2} = \sqrt{\frac{\sum x_1^2}{n} - \left(\frac{\sum x_1}{n}\right)^2} = \sqrt{\frac{14266159}{7} - (123.7)^2} = 709.7;$$

$$\sigma_y = \sqrt{\bar{y}^2 - (\bar{y})^2} = \sqrt{\frac{\sum y^2}{n} - (\bar{y})^2} = \sqrt{\frac{37297007}{7} - (2108.0)^2} = 949.6;$$

$$r_{yx1} = \frac{3136602 - 2108.0 \cdot 1238.7}{709.7 \cdot 949.6} = 0.78.$$

Связь прямая сильная.

Проверим значимость r_{yx1} :

$$t_p = \sqrt{\frac{r^2}{1 - r_{yx1}}} \cdot (n-1) = \sqrt{\frac{0.78^2}{1-0.78}}(7-2) = 3.72;$$

$$\alpha = 0.05, \nu = n - 1 = 6, t_{kp} = 2.447.$$

Так как $t_p = 3.72 \geq t_{kr} = 2.447$, *то коэффициент корреляции значим.*

Эмпирическое корреляционное отношение:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{\frac{86{,}6}{124{,}8}} = 0{,}83.$$

Связь сильная.

Задания для самостоятельной работы студентов

1. Имеются следующие данные:

№ серии	Средневзвешенная цена, x	Объем продаж, y
001	84,42	79,5
002	82,46	279,7
003	80,13	71,4
004	63,42	242,8
005	76,17	76,3
006	75,13	74,7
007	74,84	210,7

008	73,03	75,1
009	73,41	75,5
010	71,34	335,3

Составьте линейное уравнение регрессии. Вычислите параметры и рассчитайте линейный коэффициент корреляции и корреляционное отношение. Сравните их, сделайте выводы.

2. Установите направление и характер связи между четырьмя факторами по 15 банкам Японии, применив метод приведения параллельных данных:

№ банка	Суммарный актив, млрд. долл.	Объем вложений акционеров, млрд. долл.	Чистый доход, млрд. долл.	Депозиты, млрд. долл.
1	507,2	19,5	352,9	448,1
2	506,6	19,8	187,1	451,9
3	487,8	21,1	375,2	447,9
4	496,0	18,6	287,9	444,3
5	493,6	19,6	444,0	443,2
6	458,9	11,7	462,4	411,7
7	429,3	10,5	459,5	328,6
8	386,9	13,6	511,3	314,7
9	311,5	10,8	328,6	259,4
10	302,2	10,9	350,0	187,7
11	262,0	10,3	298,7	238,5
12	242,4	10,6	529,3	269,4
13	231,9	8,5	320,0	284,0
14	214,3	6,7	502,0	172,3
15	208,4	8,3	194,9	166,4

3. Составьте линейное уравнение регрессии зависимости чистого дохода от величины суммарных активов 15 крупнейших банков Японии. Определите параметры уравнения a_0 и a_1. Проанализируйте полученные параметры.

№ банка	Суммарный актив, млрд. долл.	Объем вложений акционеров, млрд. долл.	Чистый доход, млрд. долл.	Депозиты, млрд. долл.
1	507,2	19,5	352,9	448,1
2	506,6	19,8	187,1	451,9
3	487,8	21,1	375,2	447,9
4	496,0	18,6	287,9	444,3

5	493,6	19,6	444,0	443,2
6	458,9	11,7	462,4	411,7
7	429,3	10,5	459,5	328,6
8	386,9	13,6	511,3	314,7
9	311,5	10,8	328,6	259,4
10	302,2	10,9	350,0	187,7
11	262,0	10,3	298,7	238,5
12	242,4	10,6	529,3	269,4
13	231,9	8,5	320,0	284,0
14	214,3	6,7	502,0	172,3
15	208,4	8,3	194,9	166,4

Тема 9. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ДИНАМИКИ

1. Статистические ряды динамики

Основная цель статистического изучения динамики явлений состоит в выявлении и измерении закономерностей их развития во времени. Это достигается посредством построения и анализа статистических рядов динамики.

Рядами динамики называются статистические данные, отображающие развитие изучаемого явления во времени.

В каждом ряду динамики имеются два основных элемента:

1. показатель времени t;
2. соответствующие им уровни развития изучаемого явления у.

В качестве показаний времени в рядах динамики выступают либо определенные даты времени (моменты), либо отдельные периоды (годы, кварталы, месяцы, сутки).

Уровни рядов динамики отображают количественную оценку (меру) развития во времени изучаемого явления. Они могут выражаться абсолютными, относительными и средними величинами.

В зависимости от характера изучаемого явления уровни рядов динамики могут относиться или к определенным датам (моментам), или к отдельным периодам. В соответствии с этим ряды динамики подразделяются на моментные и интервальные.

Моментные ряды динамики отображают состояние изучаемых явлений на определенные даты (моменты) времени. Особенностью моментного ряда динамики является то, что в его уровни могут входить одни и те же единицы изучаемой совокупности. Поэтому при суммировании уровней моментного ряда динамики может возникнуть повторный счет. Посредством моментных рядов динамики изучают товарные запасы, состояние кадров, количество оборудования и т.д.

Интервальные ряды динамики отображают итоги развития (функционирования) изучаемых явлений за отдельные периоды (интервалы) времени.

Особенностью интервального ряда динамики является то, что каждый его уровень складывается из данных за более короткие интервалы (субпериоды) времени. Посредством интервальных рядов динамики изучают изменение во времени поступления и реализации товаров, суммы издержек обращения и т.д.

Статистическое отображение развития изучаемого явления во времени может быть представлено *рядами динамики с нарастающим итогом*. При составлении таких рядов производится последовательное суммирование смежных уровней. Этим достигается суммарное обобщение результата развития изучаемого показателя с начала отчетного периода (месяца, квартала, года).

С помощью рядов динамики изучение закономерностей развития социально – экономических явлений осуществляется в следующих направлениях:

- характеристика уровней развития изучаемых явлений;
- измерение динамики изучаемых явлений посредством системы статистических показателей;
- выявление и количественная оценка основных тенденций развития (тренда);
- изучение периодических колебаний;
- экстраполяция и прогнозирование.

2. Сопоставимость в рядах динамики

Основным условием для получения правильных выводов при анализе рядов динамики является сопоставимость его элементов. Показатели ряда динамики, подлежащие сопоставлению, должны быть однородны по экономическому содержанию. Основные условия, обеспечивающие сопоставимость в рядах динамики:

- единые единицы измерения (цены берутся за определенный период);
- единство территории;
- единство методологии сбора, учета и обобщения информации.

Для того чтобы привести уровни ряда динамики к сопоставимому виду, иногда приходится прибегать к приему, который называется «смыкание рядов динамики». Под смыканием понимают объединение в один ряд (более длинный) двух или несколько рядов динамики, уровни которых исчислены по разной методологии или территориальным границам. Для осуществления смыкания необходимо, чтобы для одного из периодов (переходного) имелись данные, исчисленные по разной методологии (или в разных границах).

При анализе рядов динамики можно решить следующие основные задачи:

1. определить интенсивность изменений в уровнях ряда из периода в период или от даты к дате;
2. определить средние показатели за рассматриваемый период;
3. выявить закономерности изменения явления во времени и в целом за рассматриваемый период;

4. выявить факторы, вызывающие изменение изучаемого явления во времени;

5. прогнозировать развитие явления на будущее.

3. Статистические показатели динамики

В основе расчета показателей рядов динамики лежит сравнение его уровней. В зависимости от применяемого способа сопоставления показатели динамики могут вычисляться на постоянной и переменной базах сравнения.

Для расчета показателей динамики на постоянной базе сравнения каждый уровень ряда сравнивается с одним и тем же базисным уровнем. Исчисляемые при этом показатели называются *базисными*. Для расчета показателей динамики на переменной базе каждый последующий уровень ряда сравнивается с предыдущим. Вычисленные таким образом показатели динамики называются *цепными*.

- *Абсолютные показатели динамики*

Важнейшим статистическим показателем динамики является абсолютный прирост, который определяется в разносном сопоставлении двух уровней ряда динамики в единицах измерения исходной информации.

Базисный абсолютный прирост исчисляется как разность между сравниваемым уровнем (y_i) и уровнем, принятым за постоянную базу сравнения (начальным) (y_{oi}):

$$\Delta y_{бi} = y_i - y_{oi}\,. \qquad (9.1)$$

Цепной абсолютный прирост – разность между сравниваемым уровнем (y_i) и уровнем, который ему предшествует (y_{i-1}):

$$\Delta y_{цi} = y_i - y_{i-1}. \qquad (9.2)$$

Абсолютный прирост может иметь и отрицательный знак, показывающий, насколько уровень изучаемого периода ниже базисного.

Между базисными и цепными абсолютными приростами имеется связь: сумма цепных абсолютных приростов равна базисному абсолютному приросту последнего периода ряда динамики.

$$\Delta y_{бп} = \sum \Delta y_ц\,. \qquad (9.3)$$

- *Относительные показатели динамики*

Распространенным статистическим показателем динамики является *темп роста*. Он характеризует отношение двух уровней ряда и может выражаться в виде коэффициента или в процентах.

Базисные темпы роста исчисляются делением сравниваемого уровня y_i на уровень, принятый за постоянную базу сравнения y_{oi}.

$$Тр_{бi} = y_i : y_{oi}. \qquad (9.4)$$

Цепные темпы роста исчисляются делением сравниваемого уровня y_i на предыдущий уровень y_{i-1}.

$$Тр_{цi} = y_i : y_{i-1}. \qquad (9.5)$$

Если темп роста больше единицы (или 100%), то это показывает на увеличение изучаемого уровня по сравнению с базисным. Темп роста, равный единице (или 100%), показывает, что уровень изучаемого периода не изменился по сравнению с базисным. Темп роста меньше единицы (или 100%) показывает на уменьшение уровня изучаемого периода по сравнению с базисным. Темп роста всегда имеет положительный знак.

Произведение последовательных цепных темпов роста равно базисному темпу роста, а частное от деления последующего базисного темпа роста на предыдущий равно соответствующему цепному темпу роста.

Темпы прироста характеризуют абсолютный прирост в относительных величинах. Исчисленный в процентах темп прироста показывает, на сколько процентов изменился сравниваемый уровень с уровнем, принятым за базу сравнения.

Базисный темп прироста вычисляется делением сравниваемого базисного абсолютного прироста $у_{бi}$ на уровень, принятый за постоянную базу сравнения $у_{oi}$.

$Тп_{бi} = \Delta у_{бi} : у_{oi}.$ (9.6)

Цепной темп прироста – это отношение сравниваемого цепного абсолютного прироста $у_{цi}$ к предыдущему уровню $у_{i-1}$.

$Тп_{цi} = \Delta у_{цi} : у_{i-1}.$ (9.7)

Взаимосвязь темпа роста и темпа прироста:

$Тп_i (\%) = Тр_i (\%) - 100$ (в процентах); (9.8)

$Тп_i = Тр_i - 1$ (в коэффициентах). (9.9)

Темп наращивания характеризует наращивание во времени экономического потенциала и определяется как отношение цепного абсолютного прироста $у_{цi}$ к начальному уровню $у_{oi}$, выражается в процентах. Может быть как положительным, так и отрицательным.

$Тн_i = \Delta у_{цi} : у_{oi}.$ (9.10)

- *Средние показатели динамики*

Средний уровень ряда динамики характеризует типическую величину абсолютных уровней ряда.

В интервальных рядах динамики средний уровень определяется делением суммы уровней на их число.

$$\bar{y} = \frac{\sum y_i}{n} = \frac{y_1 + y_2 + ... + y_n}{n}, \quad (9.11)$$

где $у_{2,} у_n$ – уровни ряда;

n – число рядов.

В моментном ряду динамики с равностоящими датами времени средний уровень определяется по формуле:

$$\bar{y} = \frac{\frac{1}{2} y_1 + y_2 + ... + \frac{1}{2} y_n}{n-1}. \quad (9.12)$$

В моментном ряду динамики с неравностоящими датами средний уровень определяется по формуле:

$$\bar{y} = \frac{\sum t_i y_i}{\sum t_i} = \frac{t_1 y_1 + t_2 y_2 + \ldots + t_n y_n}{t_1 + t_2 + \ldots + t_n}, \tag{9.13}$$

где y_i – уровни ряда динамики, сохранившиеся без изменения в течение промежутка времени t_i.

Средний абсолютный прирост представляет собой обобщенную характеристику индивидуальных абсолютных приростов ряда динамики и определяется как частное от деления суммы цепных абсолютных приростов на их число.

$$\Delta\bar{y} = \sum \Delta y_{цi} \div n. \tag{9.14}$$

Средний темп роста – обобщающая характеристика индивидуальных темпов роста ряда динамики.

$$\overline{Tp} = \sqrt[n]{Tp_1 \cdot Tp_2 \cdot \ldots \cdot Tp_n}, \tag{9.15}$$

где Tp_1, Tp_2, Tp_n – индивидуальные (цепные) темпы роста (в коэффициентах);

n – число индивидуальных темпов роста.

$$\overline{Tp}_y = \sqrt[n-1]{y_n \div y_1}. \tag{9.16}$$

Средний темп прироста можно определить на основе взаимосвязи между темпами роста и прироста.

$$\overline{\text{Тп}} = \overline{\text{Тр}} - 1. \tag{9.17}$$

4. Изучение основной тенденции развития

Тренд – долговременная компонента ряда динамики, она характеризует основную тенденцию его развития, при этом остальные компоненты рассматриваются только как мешающие процедуре его определения. При наличии ряда наблюдаемых значений для различных моментов времени следует найти подходящую трендовую кривую, которая сгладила бы остальные колебания.

После установления тенденции в ряду динамики производится ее описание с помощью методов сглаживания. Методы сглаживания разделяются на две основные группы:

1. сглаживание или механическое выравнивание отдельных членов ряда динамики с использованием фактических значений соседних уровней,
2. выравнивание с применением кривой, проведенной между конкретными уровнями, т.о., чтобы она отображала тенденцию, присущую ряду, и одновременно освободила его от незначительных колебаний.

Рассмотрим некоторые из них.

- Метод усреднения по левой и правой половине

Разделяют ряд на две части, находят для каждой из них среднее арифметическое значение и проводят через полученные точки линию тренда на графике.

- Метод укрупнения интервалов

Если рассматривать уровни экономических показателей за короткие промежутки времени, то в силу влияния различных факторов, действующих в разных направлениях, в рядах динамики наблюдаются снижение и повышение этих уровней. Это мешает видеть основную тенденцию развития изучаемого явления. Поэтому для наглядного представления тренда применяется метод укрупнения интервалов, основанный на укрупнении периодов времени, к которым относятся уровни ряда. Например, ряд ежесуточного выпуска продукции заменяется рядом месячного выпуска продукции и т.д.

- Метод простой скользящей средней

Сглаживание ряда динамики с помощью скользящей средней заключается в том, что вычисляется средний уровень из определенного числа первых по порядку уровней ряда, затем – средний уровень такого же числа уровней, начиная со второго, далее – начиная с третьего и т.д. Т.о., при расчетах среднего уровня как бы «скользят» по ряду динамики от его начала к концу, каждый раз отбрасывая один уровень в начале и добавляя один следующий. Каждое звено скользящей средней – это средний уровень за соответствующий период, который относится к середине выбранного периода. Применяется, если графическое изображение ряда динамики напоминает прямую линию.

- Метод центрирования

Определение скользящей средней по четному числу членов ряда динамики несколько сложнее, т.к. средняя может быть отнесена только к середине между двумя датами, находящимися в середине интервала сглаживания. Чтобы ликвидировать такой сдвиг, применяют способ центрирования. Центрирование заключается в нахождении средней из двух смежных скользящих средних для отнесения полученного уровня к определенной дате. При центрировании необходимо находить скользящие суммы, скользящие средние нецентрированные по этим суммам и средние из двух смежных нецентрированных скользящих средних.

- Метод взвешенной скользящей средней

Отличается от простой скользящей средней тем, что уровни, входящие в интервал усреднения, суммируются с различными весами.

- Выбор уравнения тренда, отображающего развитие социально – экономических явлений во времени.

Для отображения основной тенденции развития явлений во времени применяются полиномы разной степени.

- полином первой степени $\bar{y}_t = a_0 + a_1 t$; (9.18)

- полином второй степени $\bar{y}_t = a_0 + a_1 t + a_2 t^2$; (9.19)

- полином третьей степени $\bar{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$; (9.20)

- полином n-й степени $\bar{y}_t = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n$, (9.21)

где $a_0, a_1, a_2, \ldots, a_n$ – параметры полиномов,

n – условное обозначение времени.

Суть способа: нахождение такой прямой или кривой, ординаты точек которой были бы максимально близки к уравнениям исследуемого динамического ряда.

Закономерно изменяющийся уровень изучаемого общественного явления ($\hat{y}_t$) рассчитывается как функция времени (тренд):

$$\hat{y}_t = f(t). \tag{9.22}$$

Параметры аналитического уравнения выбранной линии находят, используя способ наименьших квадратов. Предполагается, что сумма квадратов отклонений фактических уровней (y) от выровненных ($\hat{y}_t$), т.е. расположенных на искомой линии, должна быть минимальной:

$$\sum (y - \hat{y}_t) \to \min . \tag{9.23}$$

Рассмотрим технику выравнивания ряда динамики по уравнению тренда прямой:

$$\hat{y}_t = a_0 + a_1 t, \tag{9.24}$$

где t – условное обозначение времени;

a_0; a_1 – параметры искомой прямой.

Выравнивание по уравнению тренда прямой применяется в тех случаях, когда характер движения изучаемого явления ближе всего к прямолинейному. Для этого типа динамики характерны постоянные цепные абсолютные приросты: $\Delta y_{ц} \cong const$.

Параметры a_0 ,a_1, удовлетворяющие методу наименьших квадратов, находятся путем решения следующей системы нормальных уравнений:

$$\begin{cases} \sum y = na_0 + a_1 \sum t, \\ \sum yt = a_0 \sum t + a_1 \sum t^2. \end{cases} \tag{9.25}$$

где y – фактические уровни ряда динамики;

n – число уровней ряда;

t – нумерация фактора времени.

Эта система уравнений значительно упрощается, если значения t подобрать, чтобы их сумма = 0. Тогда получается следующая система уравнений:

$$\begin{cases} \sum y = na_0 \\ \sum yt = a_1 \sum t^2 \end{cases}, \tag{9.26}$$

решая которую получаем:

$$a_0 = \frac{\sum y}{n}; \quad a_1 = \frac{\sum yt}{\sum t^2}.$$

Если уровней нечетное число, то отсчет ведется от середины, принятой за 0:

Год	2016	2017	2018	2019	2020
t	-2	-1	0	1	2

Если уровней нечетное число, то условное обозначение времени t принимает следующий вид:

Год	2015	2016	2017	2018	2019	2020
t	-5	-3	-1	1	3	5

a_0 – среднее значение динамического ряда;

a_1 – ежегодный абсолютный прирост выравненного уровня, обусловленный изменением фактора времени t.

5. Статистическое изучение сезонных колебаний

Сезонные колебания характеризуются специальными показателями – индексами сезонности. Совокупность этих показателей отражает сезонную волну. Индексами сезонности являются процентные отношения фактических внутригодовых уровней к постоянной или переменной средней.

Если ряд динамики не содержит ярко выраженной тенденции в развитии, то индексы сезонности вычисляются по эмпирическим данным без их предварительного выравнивания.

Для каждого месяца рассчитывается средняя величина уровня ($\bar{y}_i$), затем из них вычисляется среднемесячный уровень для всего ряда ($\bar{y}$), и в заключение определяется процентное отношение средних для каждого месяца к общему среднемесячному уровню ряда.

$$I_s = \frac{\bar{y}_i}{\bar{y}} \cdot 100\% . \qquad (9.27)$$

Если ряд динамики содержит определенную тенденцию в развитии, то прежде чем вычислить сезонную волну, фактические данные должны быть обработаны так, чтобы была выявлена общая тенденция. Обычно для этого прибегают к аналитическому выравниванию ряда динамики.

При использовании способа аналитического выравнивания ход вычислений индексов сезонности следующий:

- По соответствующему полиному вычисляются для каждого месяца (квартала) выравненные уровни на момент времени (t),
- Определяются отношения фактических месячных (квартальных) данных (y_i) к соответствующим выравненным данным ($\bar{y}_t$) в процентах:

$$I_i = (y_i \div \bar{y}_t) \cdot 100 . \qquad (9.28)$$

- Находятся средние арифметические из процентных соотношений, рассчитанных по одноименным периодам в процентах

$$I_i = (I_1 + I_2 + I_3 + \ldots + I_n) : n, \qquad (9.29)$$

где n – число одноименных периодов.

В общем виде формулу расчета индекса сезонности можно записать так:

$$I_s = \left[\sum \frac{y_i}{\bar{y}_t} \right] \div n . \qquad (9.30)$$

6. Элементы прогнозирования и интерполяции

Применение прогнозирования предполагает, что закономерность развития, действующая в прошлом (внутри ряда динамики), сохранится и в прогнозируемом будущем, т.е. прогноз основан на экстраполяции. Экстраполяция, проводимая в будущее, называется перспективной, и в прошлое – ретроспективной.

Теоретической основой распространения тенденции на будущее является инерционность, которая позволяет выявить сложившиеся взаимосвязи как между уровнями динамического ряда, так и между группой связанных рядов динамики. На основе рядов динамики получаются весьма надежные прогнозы, если уровни ряда динамики сопоставимы и получены на основе единой методологии.

Экстраполяцию в общем виде можно представить формулой:

$$\hat{y}_{i+T} = f(y_i, T, a_j), \tag{9.31}$$

где $\hat{y}_{i+T}$ – прогнозируемый уровень;

y_i – текущий уровень прогнозируемого ряда;

T – период упреждения;

a_j – параметр уравнения тренда.

В зависимости от тог, какие принципы и исходные данные положены в основу прогноза, выделяют следующие элементарные методы экстраполяции: среднего абсолютного прироста, среднего темпа роста и экстраполяцию на основе выравнивания рядов по какой – либо аналитической формуле.

- Прогнозирование по среднему абсолютному приросту

Для нахождения интересующего нас аналитического выражения тенденции на любую дату t необходимо определить средний абсолютный прирост и последовательно прибавить его к последнему уровню ряда столько раз, на сколько периодов экстраполируется ряд.

$$\hat{y}_{i+t} = y_i + \overline{\Delta} \cdot t, \tag{9.32}$$

где $\hat{y}_{i+t}$ – экстаполируемый уровень: (i+t) – номер этого уровня (года),

i – номер последнего уровня (года) исследуемого периода, за который рассчитан $\overline{\Delta}$,

t – срок прогноза период упреждения,

$\overline{\Delta}$ – средний абсолютный прирост.

Однако следует иметь в виду, что использование среднего абсолютного прироста для прогноза возможно только при следующем условии:

$$\sigma_{ocm}^{2} \le p^2;$$

$$p^2 = \frac{1}{2} \cdot \frac{\sum \Delta_i^2}{n}, \quad \sigma_{ocm}^{2} = \frac{\sum (y_i - \bar{y}_{\overline{\Delta}})^2}{n}. \tag{9.33}$$

- Прогнозирование по среднему темпу прироста

Осуществляется в случае, когда есть основание считать, что общая тенденция ряда характеризуется показательной (экспоненциальной) кривой. Для нахождения тенденции необходимо определить средний коэффициент роста, возведенный в степень, соответствующую периоду экстраполяции, т.е. по формуле:

$$\hat{y}_{i+t} = y_i \cdot \bar{k}_p^{\,t}, \tag{9.34}$$

где y_i – последний уровень ряда динамики,

t – срок прогноза,

$\bar{k}_p$ – средний коэффициент роста.

- Прогнозирование на основе аналитического выравнивания тренда

При этом для выхода за границы исследуемого периода достаточно продолжить значения независимой переменной времени (t).

При таком подходе к прогнозированию предполагается, что размер уровня, характеризующего явление, формируется под воздействием множества факторов, причем не представляется возможным выделить отдельно их влияние. В связи с этим ход развития связывается не с какими–либо конкретными факторами, а с течением времени, т.е. $y = f(t)$. При анализе рядов динамики иногда приходится прибегать к определению некоторых неизвестных уровней внутри данного ряда динамики, т.е. *интерполяции.*

Интерполяция может проводиться на основе абсолютного прироста, среднего темпа роста и с помощью аналитического выравнивания. Она также основана на том или ином предположении тенденции изменения уровней, характер этого прогноза несколько иной: здесь уже не приходится предполагать, что тенденция, характерная для прошлого, сохранится и в будущем.

Вопросы для теоретического контроля знаний

1. Для чего нужно изучать динамику явлений?
2. Дайте определение ряда динамики. Из каких элементов он состоит и каков их смысл?
3. Какие существуют виды рядов динамики?
4. Какие динамические ряды называются моментными и почему их уровни нельзя суммировать?
5. Какие ряды статистических величин называются интервальными? Почему их уровни можно суммировать? Приведите примеры.
6. Каковы причины возникновения несопоставимости динамических рядов?
7. Какие приемы применяются для преобразования несопоставимых рядов динамики в сопоставимые?
8. Как исчисляется средняя для интервального ряда? Приведите примеры.
9. Как исчисляется средняя для моментного ряда? Приведите примеры.
10. Что характеризуют показатели абсолютного прироста и как они исчисляются?
11. Что представляет собой темп роста? Как он исчисляется?

12. Какая существует взаимосвязь между последовательными цепными коэффициентами роста и базисным коэффициентом роста за соответствующий период? Каково практическое применение этой взаимосвязи?

13. Чему равен средний абсолютный прирост?

14. По какой формуле исчисляется средний темп роста?

15. Как исчисляется средний темп прироста?

16. Какими наиболее распространенными статистическими методами осуществляется изучение тренда в рядах динамики?

17. В чем сущность метода укрупнения интервалов и для чего он применяется?

18. Как производится сглаживание рядов динамики способом скользящей (подвижной) средней? В чем достоинства и недостатки этого метода?

19. В чем сущность метода аналитического выравнивания динамических рядов?

20. Как определяется тип уравнения тенденции динамики?

21. Охарактеризуйте технику выравнивания ряда динамики по прямой.

22. Что представляют собой сезонные колебания, в чем практическое значение их изучения?

23. Как исчисляются индексы сезонности?

24. Что такое экстраполяция и интерполяция рядов динамики?

Примеры решения задач

1. Требуется провести анализ динамики продажи мясных консервов за 2008 -2012 гг. Для удобства и наглядности исходные и рассчитанные показатели изложены в табличной форме (табл. 9.1).

Таблица 9.1

Динамика продажи мясных консервов в одном из регионов за 2016-2020 гг. и расчет аналитических показателей динамики (данные условные)

Год	Консервы мясные, млн. усл. банок	Абсолютный прирост (снижение), млн. усл. банок		Темпы роста, %		Темпы прироста, %		Абсолютное значение 1% прироста, млн. усл. банок
		с предыдущим годом	с 2016 г.	с предыдущим годом	с 2016 г.	с предыдущим годом	с 2016 г.	
1	2	3	4	5	6	7	8	9
2016	891	-	-	-	100,0	-	0,0	-
2017	806	-85	-85	90,5	90,5	-9,5	-9,5	8,91
2018	1595	+789	+704	197,9	179,0	97,9	79,0	8,06
2019	1637	+42	+746	102,63	183,7	2,63	83,7	15,95
2020	1651	+14	+760	100,85	185,3	0,85	85,3	-16,37
Итого	6580	+780	-	-	-	-	-	-

Абсолютный прирост: $\Delta_ц = y_1 - y_{i-1}, \Delta_б = y_i - y_0$.

В нашем примере абсолютное уменьшение продажи консервов в 2017 г. по сравнению с 2016 г. составило: 806 – 891 = - 85 млн. усл. банок (табл. 9.1, гр. 2), а по сравнению с базисным 2016 г. продажа консервов возросла в 2020 г. на 760 млн. усл. банок (гр. 3).

Темп роста: $Tp = \frac{y_i}{y_{i-1}} \cdot 100, \quad Tp = \frac{y_i}{y_0} \cdot 100$.

Для 2020 г. *темп роста* по сравнению с 2016 г. составил $\left(\frac{1651}{891}\right)$*100 = 185,3% (табл. 9.1, гр. 5).

Темп прироста: $Tnp = \frac{\Delta}{y_{i-1}} \cdot 100, \quad Tnp = \frac{\Delta}{y_0} \cdot 100$.

Показывает, например, на сколько процентов продажа консервов в 2020 г. возросла по сравнению с 2016 г.: $\left(\frac{760}{891}\right)100 = 85{,}3\% = 185{,}3 - 100 = 85{,}3\%$ (гр. 6,7).

Показатель абсолютного значения одного процента прироста: $|\%| = \frac{\Delta}{Tnp}$ *или* $0{,}01 y_{i-1}$.

Для 2020 г. абсолютное значение 1% прироста равно: 0,01*16,37 = 16,37 или 14: 0,0855 = 16,37 млн. банок.

Среднегодовой абсолютный прирост ($\overline{\Delta} = \frac{\sum \Delta ц}{n}, \overline{\Delta} = \frac{y_n - y_0}{n-1}$) продажи мясных консервов за 2016-2020 гг. равен $\Delta = \frac{760}{4} = 190$ *или* $\Delta = \frac{1651 - 891}{4} = 190$ *млн.усл.банок*.

Среднегодовой темп роста ($\overline{Tp} = \sqrt[m]{K_1 \cdot K_2 \cdot K_3 \cdot \ldots \cdot K_n}$, *или* $\overline{Tp} = \sqrt[n-1]{\frac{y_n}{y_0}}$) продажи мясных консервов за 2016-2020 гг. рассчитаем двумя способами:

Тр = $4\sqrt{0{,}905 * 1{,}979 * 1{,}026 * 1{,}009} = 4\sqrt{}1{,}853$ = 1,167 или 116,7%;

Тр = $4\sqrt{\frac{1651}{891}}$ = $4\sqrt{1{,}853}$ =1,167 или 116,7%.

Среднегодовой темп прироста ($\overline{T}np = \overline{T}p - 100$) получим, вычтя из среднего темпа роста 100%.

Тпр = 116,7% - 100% = 16,7%.

2. Выявить сезонные колебания по следующим данным о числе разводов по месяцам за 2018-2020 гг.

Месяц	Число расторгнутых браков				Индекс сезонности
	2018	2019	2020	в среднем за три года	
Январь	195	158	144	165,7	122,4
Февраль	164	141	136	147,0	108,6
Март	153	153	146	150,7	111,3
Апрель	136	140	132	136,0	100,4
Май	136	136	136	136,0	100,4
Июнь	123	129	125	125,7	92,8
Июль	126	128	124	126,0	93,1
Август	121	122	119	120,7	89,1
Сентябрь	118	118	118	118,0	87,2
Октябрь	126	130	128	128,0	94,5
Ноябрь	129	131	135	131,7	97,3
Декабрь	138	141	139	139,3	102,9
Средний уровень ряда	138,7	135,6	131,8	135,4	100,00

1. Произведем осреднение уровней одноименных периодов:

$$\text{январь} = \frac{у_{янв.\ 10} + у_{янв.\ 11} + у_{янв.\ 12}}{3} = \frac{195+158+144}{3} = 165{,}7;$$

$$\text{декабрь} = \frac{у_{дек.10} + у_{дек.\ 11} + у_{дек.\ 12}}{3} = \frac{138+141+139}{3} = 139{,}3.$$

2. По месячным средним уровням определим средний общий уровень

$$\bar{y} = \frac{\sum \bar{y}_i}{n} = \frac{1624{,}8}{12} = 135{,}4.$$

3. По месяцам определим индексы сезонности

$$Is = \frac{\sum y_i}{\bar{y}} \cdot 100\% .$$

$I_{янв}$. = (165,7 : 135,4) 8 100% = 122,4%.

$I_{дек.}$ = (147,0 : 135,4) * 100% = 108,6%.

3. На основе данных о производстве стиральных машин произвести сглаживание методом трехчленной средней.

Месяц	Стиральные машины, тыс. шт.	Трехчленные скользящие суммы	Трехчленные скользящие средние
1	155	-	-
2	163	-	161,7
3	167	485	153,7
4	131	461	152,0
5	158	456	145,3
6	147	436	145,0

7	130	435	140,7
8	145	422	134,3
9	128	403	137,7
10	140	413	142,3
11	159	427	153,00
12	160	459	155,3

Взяв данные за первые три месяца, исчисляем трехчленные суммы, а затем средние:

$$\bar{y}_1 = \frac{155+163+167}{3} = 161{,}7.$$

$$\bar{y}_2 = \frac{163+167+131}{3} = 153{,}7 \quad \text{и т.д.}$$

4. С помощью метода аналитического выравнивания по прямой выразите основную тенденцию по следующим данным о производстве молока:

Год	млн. т., y	t	t^2	ty	$\bar{y}_i$	$y - \bar{y}_i$	$(y - \bar{y}_I)^2$
2008	13,3	-2	4	-26,6	13,02	0,28	0,08
2009	13,5	-1	1	-13,5	13,94	-0,44	0,19
2010	14,8	0	0	0	14,86	-0,0	0,00
2011	16,1	1	1	16,1	15,78	-0,32	0,10
2012	16,6	2	4	33,2	16,70	-0,1	0,01
Итого	74,3	-	10	9,2	74,30	-	0,38

Для выравнивания ряда динамики по прямой используем уравнение: $\bar{y}_t = a_0 + a_1 t$.

Расчеты необходимых значений представлены в таблице.

a_0 = 74,3 : 5 = 14,86; a_1 = 9,2 : 10 = 0,92.

В результате получаем следующее уравнение основной тенденции производства молока в регионе за 2008-2012 гг. *:* $\bar{y}_t = 14.86 + 0.92t.$

Подставляя в уравнение принятые обозначения t, вычислим выровненные уровни ряда динамики*:*

2008: $\bar{y}_1$ = 14,86 +0,92 (-2) = 13,02;

2009: $\bar{y}_2$ = 14,86 + 0,92 (-1) = 13,94 и т.д.

Задания для самостоятельной работы

1. Имеются следующие данные о производстве продуктов животноводства:

Год	Мясо, млн. т	Молоко, млн. т.
2011	10,0	72,6
2012	12,3	83,6
2013	15,0	90,8
2014	13,6	89,7

2015	14,7	94,9
2016	15,3	94,5
2017	15,5	93,3
2018	15,1	90,9
2019	15,2	88,9
2020	15,3	90,1

По каждому виду продукции рассчитайте среднегодовые абсолютные приросты, среднегодовые темпы роста и прироста.

2. По следующим данным о производстве молока в РФ за 2015-2020 гг., млн. т:

2015	2016	2017	2018	2019	2020
13,3	13,5	14,8	16,1	16,6	16,4

определите среднегодовые абсолютные приросты, среднегодовые темпы роста, среднегодовые темпы прироста.

3. Используя взаимосвязь показателей динамики, определите уровни ряда динамики и недостающие в таблице цепные показатели динамики по следующим данным о производстве продукции предприятиями (в сопоставимых ценах):

Год	Производство продукции, млн. руб.	По сравнению с предыдущим годом		
		абсолютный прирост, млн. руб.	темп роста, %	темп прироста, %
2015	92,5			
2016		4,8		
2017			104,0	
2018				5,8
2019				7,4
2020	7,0			

4. Имеются следующие данные о внутригодовой динамике поставки шерстяных тканей в розничную сеть области по кварталам за 2018-2020 гг., млн. руб.:

Квартал	2018	2019	2020
1	171,9	160,0	172,1
2	132,8	113,1	176,8
3	144,4	124,2	139,1
4	154,7	155,8	141,2

Для анализа внутригодовой динамики продажи шерстяных тканей: а) определите индексы сезонности; б) изобразите графически сезонную волну развития изучаемого явления по месяцам года. Сделайте выводы.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Основная литература

1. Гусаров, В.М. Статистика: [текст]: Учеб. пособие / В.М. Гусаров. – 2-е изд., испр. и доп.. – М.: ЮНИТИ, 2008. – 479 с.
2. Статистика: Учебник: [текст] / Ред. И.И. Елисеева. – М.: Проспект, 2013. – 448 с.

ЭБС «Лань»

3. Годин, А. М. Статистика : учебник / А. М. Годин. — 11-е изд., перераб. и испр. — Москва : Дашков и К, 2017. — 412 с. — ISBN 978-5-394-02183-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/93468 (дата обращения: 29.10.2021). — Режим доступа: для авториз. пользователей.
4. Горлач, Б. А. Теория вероятностей и математическая статистика. Практикум для студентов технических и экономических специальностей вузов : учебное пособие для вузов / Б. А. Горлач, С. В. Подклетнова. — Санкт-Петербург : Лань, 2021. — 116 с. — ISBN 978-5-8114-6736-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/162372 (дата обращения: 29.10.2021). — Режим доступа: для авториз. пользователей.

Программное обеспечение и Интернет-ресурсы

1. Научно-образовательный портал «Экономика и управление на предприятиях»: http://www.eup.ru
2. Портал «Гуманитарное образование»: http://www.humanities.edu.ru/
3. Федеральный портал «Российское образование»: http://www.edu.ru/
4. Электронная энциклопедия «Википедия»: http://ru.wikipedia.org/wiki/
5. Официальный сайт журнала «Справочник экономиста»: http://www.profiz.ru/se
6. Официальный сайт компании «Консультант Плюс»: http://www.consultant.ru/
7. Информационно-правовой портал «Гарант»: http://www.garant.ru/
8. Официальный сайт Федеральной службы государственной статистики: http://www.gks.ru/
9. Сайт территориального органа Федеральной службы государственной статистики по Алтайскому краю: http://ak.gks.ru/default.aspx.

Чиркова Ольга Александровна
Ляпкина Наталья Александровна

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ
Часть 1

Учебное пособие по дисциплине «Статистика (теория статистики, социально – экономическая статистика)» для студентов всех форм обучения экономических направлений подготовки

Подписано к печати 25.11.21. Формат 60х84 /16.
Усл. печ. л. 5,25. Тираж 60 экз. Заказ 2117122. Рег. №71.

Отпечатано в ИТО Рубцовского индустриального института
658207, Рубцовск, ул. Тракторная, 2/6.