

3
МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Амурский гуманитарно-педагогический государственный университет»
(ФГБОУ ВО «АмГПУ»)

Т. Б. Ершова

СТАТИСТИКА

Учебное пособие

В 2 частях

Часть 1

Комсомольск-на-Амуре
2015

Печатается по решению редакционно-издательского совета и кафедры экономики Амурского гуманитарно-педагогического государственного университета

УДК 83.03

ББК 60.6

Е 80

Рецензенты: Коваленко В. Н., к. э. н., доцент, заведующая кафедрой экономики ФГОУ ВО «АмГПУ»;
Жилин С. Г., к. т. н., доцент, к.т.н., доцент, старший научный сотрудник ИМиМ ДВО РАН

Ершова, Т. Б. Статистика: учебное пособие. В 2 ч. Ч.1 / Татьяна Борисовна Ершова. – 2 изд., перераб. и испр. – Комсомольск-на-Амуре: Изд-во АмГПУ, 2015.–73с.

В учебном пособии представлен теоретический материал по разделам курса «Статистика», посвященный изучению основополагающих приемов, методов и способов статистического исследования социально-экономических явлений и процессов. Даны примеры решения задач, упрощающие восприятие теоретического материала, тесты для самоконтроля, контрольные работы по вариантам.

Адресовано студентам, осваивающим данную дисциплину с использованием дистанционной образовательной технологии, а также всем, интересующимся вопросами статистики.

Вне плана

УДК 83.03

ББК 60.6

© Т. Б. Ершова, 2015.

© Издательство АмГПУ, 2015

Оглавление

Методические указания по использованию пособия.....	4
Методические рекомендации.....	5
1. Теоретические основы статистики как науки.....	7
1.1. Предмет статистики.....	7
1.2. Статистическая методология.....	8
1.3. Структура статистической науки.....	9
1.4. Статистическое наблюдение.....	10
1.5. Статистические таблицы и графики	13
Тесты для самоконтроля.....	15
2. Статистическая группировка и сводка материала.....	16
Тесты для самоконтроля.....	20
3. Статистические величины.....	21
3.1. Абсолютные величины.....	21
3.2. Относительные величины.....	22
3.3. Степенные средние величины	26
3.4. Структурные средние величины.....	29
Тесты для самоконтроля.....	35
4. Статистическое распределение.....	36
4.1. Понятие и измерение вариации признака.....	37
4.2. Виды дисперсий и правило их сложения.....	39
4.3. Способы построения вариационного ряда.....	41
4.4. Показатели центра распределения.....	43
4.5. Сопоставимость рядов распределения.....	44
4.6. Показатели формы распределения и графическое изображение вариационного ряда	46
Тесты для самоконтроля.....	53
5. Выборочное наблюдение	54
5.1. Виды выборочного наблюдения и формы отбора единиц в выборочную совокупность.....	54
5.2. Задачи, которые можно решать с помощью предельной ошибки выборки.....	55
5.3. Простая случайная и механическая выборки.....	56
Тесты для самоконтроля.....	62
Задания к контрольной работе №1.....	63
Вопросы для собеседования.....	70
Приложение 1. Удвоенная нормированная функция Лапласа ...	72
Библиографический список.....	73

Методические указания по использованию пособия

Термин «статистика» происходит от латинского слова status, что в средние века означало политическое состояние государства. В настоящее время под термином «статистика» чаще всего понимают следующее:

- статистика — одна из общественных наук, учение о системе количественных характеристик, дающих всестороннее представление об общественных явлениях, о национальном хозяйстве в целом и отдельных его отраслях;

- статистика — это также одна из форм практической деятельности, целью которой является сбор, обработка и анализ массовых данных о тех или иных явлениях;

- статистика — это эффективное орудие, инструмент познания, используемый в естественных и общественных науках для установления тех специфических закономерностей, которые действуют в конкретных массовых явлениях, изучаемых данной наукой;

- статистикой также называют цифровой материал, служащий для характеристики какой-либо области общественных явлений.

Статистика на протяжении сотен лет своего существования, независимо от политической системы и экономического развития общества, всегда выступала как необходимый и эффективный инструмент государственного управления и одновременно как наука, исследующая количественную сторону массовых явлений. Выполняя самые разнообразные функции сбора, систематизации и анализа сведений, характеризующих экономическое и социальное развитие общества, она всегда играла роль главного поставщика фактов для принятия управленческих решений, проведения научных исследований и характеристики разного рода структур, организаций и населения.

В современном обществе статистика является одним из важнейших инструментов управления национальной экономикой. Улучшение хозяйственного руководства неразрывно связано с возрастанием роли статистики и повышением научного уровня статистических исследований.

Основной целью преподавания дисциплины «Теория статистики» является формирование у студентов теоретических знаний и практических навыков проведения статистического исследования и анализа массовых социально-экономических явлений и процессов.

Задачи учебного курса:

- формирование представления о статистике как научной дисциплины, изучающей основные закономерности развития социально-экономических явлений;

- развитие у слушателей способности статистического мышления на основе изучения специальных научных правил, методов и приемов количественного анализа разного рода статистической информации;

- овладение знаниями общих основ статистики и навыками организации и проведения статистического исследования;
- изучение принципов и методов анализа социально-экономических явлений и процессов.

Место дисциплины в структуре ООП ВПО

Дисциплина относится к математическому и естественнонаучному циклу дисциплин учебного плана.

Изучение дисциплины «Теория статистики» базируется на знаниях, полученных в результате изучения дисциплин «Математический анализ», «Эконометрика», «Экономическая теория».

Компетенции обучающегося, формируемые в результате освоения дисциплины:

ОК-15 – владение методами количественного анализа и моделирования, теоретического и экспериментального исследования;

ОК-17 – владение основными методами, способами и средствами получения, хранения, переработки информации, навыками работы с компьютером как средством управления информацией.

В результате освоения дисциплины «Теория статистики» обучающийся должен:

- **знать:**
 - основные понятия и задачи статистики;
 - способы расчета и особенности применения абсолютных, относительных и средних величин;
 - виды, формы и способы статистического наблюдения;
 - виды и особенности группировки и сводки материала;
 - основные характеристики рядов распределения;
 - виды выборочного наблюдения и особенности их анализа;
 - методологию проведения корреляционно-регрессионного анализа парной линейной и множественной связей;
 - абсолютные, относительные и средние показатели динамики;
 - особенности индексного анализа;
- **уметь:**
 - рассчитывать абсолютные, относительные и средние показатели;
 - строить ряды распределения и анализировать их;
 - определять предельную и среднюю ошибки выборок, переносить результаты выборочного обследования на генеральную совокупность;
 - проводить анализ степени тесноты связи между признаками;
 - строить уравнение регрессии;
 - анализировать ряды динамики;
 - применять индексные методы анализа;
- **владеть:**
 - терминологией статистики;
 - основными методиками проведения статистического анализа.

На освоение дисциплины «Теория статистики» в учебном плане направления подготовки 080200.62 «Менеджмент» профиль «Управление малым бизнесом» отведено три семестра. Контроль качества усвоения материала проводится в два этапа: темы 1-5 – второй семестр, 7-10 – третий семестр. По окончании второго семестра студенты сдают зачет, вопросы к зачету представлены на странице 70. К зачету необходимо выполнить письменно контрольную работу № 1, которая находится в первой части учебного пособия. Третий семестр завершается сдачей экзамена. Вопросы на экзамен представлены на странице 55 второй части учебного пособия. Для допуска к экзамену необходимо представить выполненные контрольные работы № 2 и № 3, представленные во второй части учебного пособия.

Требования к выполнению контрольной работы:

- работа должна быть выполнена в сроки, установленные учебным планом;
- работа выполняется студентом на компьютере и сдается в распечатанном варианте на проверку;
- работа не должна содержать сокращений и обозначений, не принятых в статистике;
- задачи выполняются последовательно, условия заданий должны быть перенесены в сдаваемую контрольную работу;
- при решении задач записывается формула расчета, ее название и аргументация выбора данной формулы;
- ход решения задачи подробно описывается, расчеты должны быть проведены с точностью до 0,001, а проценты – до 0,01.
- завершающим этапом решения задач является построение вывода в текстовой форме.

Задания к контрольной работе составлены в шести вариантах. Номер варианта определяется по последней цифре номера зачетной книжки и начальной букве фамилии студента (см. табл. 1). Определенный номер варианта справедлив для всех контрольных работ, выполняемых по курсу «Теория статистики».

Таблица 1. Распределение вариантов контрольной работы

Начальная буква фамилии студента	Последняя цифра номера зачетной книжки									
	1	2	3	4	5	6	7	8	9	0
А, Б, В, Ф, Х	1	2	3	4	5	6	1	2	3	4
Г, Д, Е, Ц, Ч	2	3	4	5	6	1	2	3	4	5
Ж, З, И, К, У	3	4	5	6	1	2	3	4	5	6
Л, М, Ю, Я	4	5	6	1	2	3	4	5	6	1
Н, О, П, Ш, Щ	5	6	1	2	3	4	5	6	1	2
Р, С, Т, Э	6	1	2	3	4	5	6	1	2	3

1. Теоретические основы статистики как науки

1.1. Предмет статистики

Предмет исследования статистики – массовые социально-экономические явления и процессы. Статистика изучает количественную сторону этих явлений в неразрывной связи с их качественным содержанием в конкретных условиях места и времени.

Качественная сторона явления характеризуется сходством единиц по каким-либо качественным признакам (например, пол человека, принадлежность предприятия к определенной отрасли экономики).

Количественная сторона явления характеризуется числовыми значениями признаков каждой отдельной единицы явления, определяет размер явления и соотношение его частей (например, в населенном пункте 123 мужчины и 149 женщин, предприятий машиностроения в регионе 530).

При анализе массового явления необходимо учитывать единство качественной и количественной сторон. Если не определена качественная сторона явления, то невозможно определить и его количественную сторону (например, при анализе производительности труда необходимо знать само понятие производительности, его основные характеристики и факторы, влияющие на нее).

Массовое социально-экономическое явление – явления и процессы жизни общества, обладающие индивидуальными и общественными признаками (например, производство и потребление продукции, перевозка грузов, экспорт и импорт продукции и т.д.).

Объектом статистического исследования является статистическая совокупность (в каждом конкретном случае своя).

Статистическая совокупность – это множество однородных единиц, обладающих определенными признаками, объединяющими их (например, предприятия отдельной отрасли, население города, сберегательные банки и т.д.).

По форме внешнего выражения признаки делятся на:

- **атрибутивные** (описательные, качественные) признаки характеризуют качественную сторону изучаемого явления;
- **количественные** (объемные) признаки, характеризуют количественную сторону изучаемого явления;

Количественные признаки всегда имеют цифровую оценку и могут быть представлены дискретными, т.е. прерывными, неделимыми величинами (например, люди, тарифный разряд) или непрерывными, т. е. величинами, которые могут принимать не только целые, но и дробные значения (например, возраст, денежный доход, прибыль).

Единица статистической совокупности – отдельно взятый элемент совокупности (например, человек, отдельное предприятие). Единицы совокупности обладают индивидуальными особенностями и различиями, отличающими их друг от друга. В статистической

совокупности эти отличия чаще всего отражаются количественной стороной. Количественные изменения значений признака при переходе от одной единицы совокупности к другой называются **вариацией признака**.

Явления и процессы в статистике характеризуются статистическими показателями. **Статистический показатель** – это количественная оценка свойств изучаемого явления. С помощью них определяются размеры изучаемого явления, его особенности, закономерности развития и взаимосвязь с другими явлениями.

Виды статистических показателей:

1. *Учетно-оценочные* отражают объем и уровень изучаемого явления (например, число предприятий машиностроительного комплекса в регионе).

2. *Аналитические показатели* характеризуют особенности развития явления, распространенности его в пространстве, соотношение его частей, взаимодействие с другими явлениями (например, анализ себестоимости предприятий легкой промышленности в определенные периоды времени, анализ динамики объемов производства, сравнительная характеристика каждого предприятия по качеству продукции). К аналитическим показателям относятся средние величины, показатели вариации признака, относительные величины, показатели динамики, индексы, показатели выполнения плана и планового задания и т.д.

Задачи статистики как науки:

- изучение уровня и структуры массовых социально-экономических явлений и процессов;
- анализ взаимосвязей массовых социально-экономических явлений и процессов;
- изучение динамики развития массовых социально-экономических явлений и процессов.

Статистическая методология

Статистическая методология – это система принципов, приемов, способов и методов, направленных на изучение количественных закономерностей массовых явлений и процессов социально-экономической жизни общества.

Методы:

- 1) *теоретический (качественный) анализ явления* (необходимо знать сущность изучаемого явления или процесса, законы его развития, факторы, влияющие на него, взаимосвязь с другими явлениями и процессами);
- 2) *метод аналогии* (перенесение свойств одного явления или процесса на другое);
- 3) *метод гипотез* (научно обоснованное предположение о возможных причинно-следственных связях между явлениями);

4) *метод цифрового освещения явления* (применение обобщающих статистических показателей для анализа особенностей и динамики развития явления или процесса). Обобщающие показатели: абсолютные, относительные, средние и индексные показатели.

Последний метод состоит из трех основных этапов проведения статистического исследования:

1) массовое статистическое наблюдение (сбор первичной информации об отдельных единицах изучаемого явления), в результате формируется исходный материал (совокупность) для проведения статистического исследования;

2) первичная обработка, группировка и свodka материала (расчленение всей массы единиц совокупности на группы и подгруппы, расчет итогов по каждой группе и оформление результатов в виде статистических таблиц);

3) обработка статистических показателей, полученных при сводке материала и анализ результатов для построения выводов. Выводы, как правило, излагаются в текстовой форме и имеют две составные части:

- констатация факта наличия и направления связи между признаками;
- аргументация факта с помощью полученных цифровых результатов.

Таким образом, методы статистики основаны на соединении анализа и синтеза: сначала выделяются и отдельно изучаются части явления, определяются причины различий между ними, а затем дается характеристика явления в целом.

Структура статистической науки

Современная статистическая наука представлена тремя основными уровнями (см. рис. 1).

Первый уровень – общая теория статистики – является наукой о наиболее общих принципах, правилах и законах цифрового освещения социально-экономических явлениях. Важнейшие ее разделы: статистическое наблюдение, статистическая группировка и свodka материала, статистические показатели. Категории, показатели и методы общей теории статистики используются на всех уровнях статистического исследования и во всех отраслевых статистиках.

Второй уровень включает в себя экономическую и социально-демографическую статистики. *Экономическая статистика* изучает явления и процессы в области экономики: структуру, взаимосвязи отраслей экономики, анализирует показатели общественного производства. *Социально-демографическая* статистика изучает население, социальные явления и процессы: условия жизнедеятельности людей, их взаимоотношения, образ жизни людей, соотношение различных социальных групп.

Третий уровень занимают отрасли экономической и социально-демографической статистики



Рис. 1. Структура статистической науки

Статистическое наблюдение

Статистическое наблюдение – это планомерный, научно обоснованный сбор данных или сведений о социально-экономических явлениях и процессах.

Требования, предъявляемые к статистическим данным:

1) **достоверность** данных обеспечивается компетентностью исследователя, совершенством инструментария (бланков, инструкций, анкет и т.д), заинтересованностью и готовностью объекта наблюдения, полнотой данных;

2) **обоснованный отбор** единиц наблюдения в исследуемую статистическую совокупность (так как явления социально-экономической жизни общества многообразны, исследователь вынужден проводить сбор данных лишь по части совокупности);

3) **сопоставимость** обеспечивается едиными единицами измерения, сбором данных в определенных границах пространства и времени;

4) **своевременность** (устаревшие данные не отражают действительности).

Статистическое наблюдение могут проводить органы государственной статистики, научно-исследовательские институты, экономические службы банков, предприятий, фирм, бирж.

Элементы статистического наблюдения:

- **объект наблюдения** – исследуемая статистическая совокупность (население при переписи, предприятия, населенные пункты, банки и т.д.);

- **единица наблюдения** – отдельный элемент исследуемой совокупности, который обладает определенными признаками, подлежащими регистрации (человек, факт, предмет, процесс и т.д.). Единицы наблюдения обладают множеством признаков, у человека – это возраст, пол, социальное происхождение, семейное положение, доход;

- **статистический признак** – это конкретное свойство, качество, отличительная черта единицы наблюдения (например, пол работника, квалификация, профессия). Правила формирования признаков: отбор признаков производится с учетом целей исследования и возможностей их обработки, количество отобранных признаков должно быть не большим, признаки должны взаимно дополнять друг друга, отбор признаков должен производиться с учетом возможностей исследователя;

- **отчетная единица** – единица наблюдения, от которой поступают в установленном порядке данные по утвержденным формам (предприятия, организации, банки);

- **программа наблюдения** – перечень признаков (вопросов), подлежащих регистрации в процессе наблюдения. Требования к программе наблюдения: программа должна содержать признаки, которые непосредственно характеризуют свойства, основные черты изучаемого явления; вопросы программы должны быть точными и понятными; программа должна содержать определенную последовательность вопросов;

- **статистический формуляр** – документ единого образца, содержащий программу и результаты наблюдения;

- **момент или период наблюдения** – это время, по состоянию на которое регистрируются данные. **Критический момент наблюдения** – это конкретная дата (например, при переписи населения 1989 года критическим моментом наблюдения было 11 января 24.00 часа) **Период наблюдения** – это срок, в течение которого происходит регистрация данных, т.е. время необходимое для сбора данных.

Выделяют три основные **формы** статистического наблюдения:

1) **отчетность** – основная форма наблюдения, с помощью которой статистические органы в определенные сроки получают от предприятий, учреждений и организации отчетные данные;

2) **специально организованное наблюдение** (перепись) – проводится с целью получения сведений, отсутствующих в отчетности, или для проверки ее данных;

3) **регистрационное наблюдение** – это форма непрерывного статистического наблюдения за долговременными процессами.

Способы статистического наблюдения:

1) **непосредственное наблюдение** – наблюдение, при котором регистраторы непосредственно осуществляют необходимые замеры, взвешивание, подсчет и т.д.

2) **документальное наблюдение** – наблюдение, при котором источников статистической информации служат различного рода документы;

3) **опрос** – это способ наблюдения, при котором необходимые сведения получают со слов респондента (опрашиваемого, т.е. единицы наблюдения). Может проводиться в устной форме или с помощью специально разработанных опросных листов;

4) **корреспондентский способ** – сведения наблюдателям сообщают добровольные корреспонденты;

5) **анкетный способ** – сбор информации с помощью анкет;

6) **явочный способ** – представление сведений регистраторам в явочном порядке (например, при регистрации брака, рождений и т.д.).

Виды статистического наблюдения:

1) *по времени регистрации* выделяют:

- **непрерывное (текущее)** – систематизированная регистрация фактов по мере их возникновения (например, дорожно-транспортные происшествия, акты гражданского состояния: брак, рождение, развод);

- **периодическое** – сбор данных, отражающих изменение объекта, через определенные периоды времени (например, раз в 10 лет перепись населения);

- **единовременное** – регистрация состояний объекта по мере возникновения потребности в данных (например, инвентаризация незавершенного строительства 1990 г.).

2) *по охвату единиц совокупности* существуют:

- **сплошное** – охватывает при регистрации все без исключения единицы наблюдения;

- **несплошное** – при регистрации производится сбор данных лишь по части единиц совокупности:

- обследование **основного массива**, т.е. наблюдение за частью наиболее крупных единиц;

- **выборочное наблюдение** – наблюдение за определенной частью единиц, отобранных в исследуемую совокупность;

- **монографическое наблюдение** – подробное описание отдельных единиц наблюдения в совокупности

Точность статистического наблюдения – это степень соответствия полученных результатов в ходе наблюдения и их действительного значения.

Расхождение между расчетными и действительными значениями изучаемых величин называют **ошибкой наблюдения**.

Выделяют два основных вида ошибок наблюдения:

1. Непреднамеренные:

- **случайные ошибки** связаны с неточностью измерительных приборов, невнимательностью регистратора, небрежностью заполнения документов и т.д. (например, опечатки, ошибки в формулах расчета по невнимательности и т.д.);

- **систематические ошибки** наблюдения могут быть связаны, например, с округлением полученных значений в большую сторону;

- **ошибки репрезентативности** (представительности) встречаются только при проведении несплошного наблюдения и связаны с отбором не тех единиц в исследуемую совокупность (например, для анализа успеваемости студентов группы были выбраны только отличники, следовательно, вывод по группе в целом будет сделан ошибочный);

- **случайные ошибки репрезентативности** возникают, если отобранные единицы не характеризуют совокупность в целом;

- **систематические ошибки репрезентативности** связаны с нарушением принципов отбора единиц в исследуемую совокупность.

2. Преднамеренные (злостные), как правило, встречаются при получении неточных данных от отчетной единицы, где значения величин специально завышены или занижены.

1.5. Статистические таблицы и графики

Статистическая таблица – это способ рационального изложения и обобщения данных о социально-экономических явлениях при помощи цифр, расположенных в определенном порядке.

Внешне таблица представляет собой пересечение граф и строк, которые формируют ее *остов* (см. рис.2). Каждое пересечение образует клетку таблицы.

Таблица содержит три вида заголовков: общий, верхние и боковые. *Общий* заголовок отражает содержание всей таблицы, располагается над ее макетом по центру и является внешним заголовком. *Верхние* заголовки характеризуют содержание граф, а *боковые* – строк. Они являются внутренними заголовками. Остов таблицы, заполненный заголовками, образует ее *макет*.

Основными элементами статистической таблицы являются подлежащее и сказуемое:

- **подлежащее** статистической таблицы характеризует объект исследования, это могут быть отдельные единицы совокупности (фирмы, объединения) в порядке их перечисления или сгруппированные по каким-либо признакам, обычно подлежащее таблицы размещается в левой части, в наименовании строк;

- **сказуемое** таблицы образует система показателей, которыми характеризуется объект исследования, т.е. подлежащее таблицы. Сказуемое формирует верхние заголовки и составляет содержание граф.

Название таблицы
(общий заголовок)

Содержание строк	Наименование граф (верхние заголовки)				
А	1	2	3	4	...
Наименование строк (боковые заголовки)					
Итоговая строка					Итоговая строка

Рис. 2. Остов (основа) статистической таблицы

В зависимости от структуры подлежащего и группировки в нем единиц совокупности различают простые и сложные таблицы.

Простая таблица – таблица, подлежащее которой состоит из простого перечня объектов или территориальных единиц, т.е. в подлежащем нет группировки единиц совокупности.

Простые таблицы не дают возможности выявить социально – экономические типы изучаемых явлений, их структуру, а также взаимосвязь между характеризующими их признаками.

Сложная таблица – таблица, которая содержит группировку единиц совокупности по одному и более признакам. Различают групповые и комбинационные сложные таблицы.

Групповыми называются таблицы, подлежащее которых содержит группировку единиц совокупности по одному признаку, а **комбинационными** – по двум и более признакам.

Статистический график – это рисунок, на котором с помощью геометрических образов, линий, символов или знаков описывается статистическая совокупность.

Основными элементами *графики* являются:

- **графический образ**, т.е. совокупность точек, линий, фигур, с помощью которых изображаются статистические показатели;

- **поле графика** – часть плоскости ограниченной осями, где расположены графические образы;

- **пространственные ориентиры** задаются в виде системы координатных сеток;

- **масштабные ориентиры** определяются масштабом и системой масштабных шкал;

- **экспликация графика** - словесное описание содержания графика, включающее в себя название графика, подписи вдоль масштабных шкал и пояснения к отдельным частям графика.

По способу построения статистические графики делятся на диаграммы и карты.

Диаграммы – наиболее распространенный способ графических изображений, основанный на применении геометрических фигур. Наиболее часто в статистических исследованиях используются линейные, столбиковые, полосовые, и секторные диаграммы.

Статистические карты – это графики количественного распределения по поверхности, которые показывают пространственное размещение или пространственную распространенность статистических данных.

Тесты для самоконтроля

1. Объект статистического наблюдения – это
 - а) единица наблюдения; б) статистическая совокупность;
 - в) единица статистической совокупности; г) отчетная единица.
2. Субъект, от которого поступают данные в ходе статистического наблюдения:
 - а) единица наблюдения; б) отчетная единица; в) единица совокупности.
3. Подлежащее групповых статистических таблиц содержит:
 - а) перечень единиц совокупности по признаку;
 - б) группировку единиц совокупности по одному признаку;
 - в) группировку единиц совокупности по нескольким признакам.
4. Сказуемым статистической таблицы является:
 - а) исследуемый объект; б) показатели, характеризующие исследуемый объект; в) сведения, расположенные в верхних заголовках таблицы.
5. Основными элементами статистического графика являются:
 - а) поле графика; б) формулы; в) фон графика; г) экспликация графика.
6. По характеру разработки подлежащего различают статистические таблицы: а) простые; б) перечневые; в) комбинационные; г) групповые.
7. Каковы виды статистических графиков по способу построения?
 - а) диаграммы; б) статистические карты; в) линейные; г) плоскостные.
8. Перепись населения России – это
 - а) единовременное, специально организованное, сплошное наблюдение;
 - б) периодическое, специально организованное, сплошное наблюдение;
 - в) единовременное, регистровое, сплошное наблюдение.
9. Экспликация – это: а) наименование осей графика; б) словесное описание содержания графика; в) система координатных сеток.
10. Статистическая отчетность – это: а) вид статистического наблюдения; б) способ наблюдения; в) форма статистического наблюдения.
11. Наблюдение, когда изменения в отношении изучаемых явлений фиксируются по мере их поступлений – это: а) единовременное обследование; б) регистровое наблюдение; в) текущее наблюдение.

Правильные ответы: 1. б; 2. б; 3. б; 4. б; 5. а, б; 6. а, г; 7. а, б; 8. а; 9. б; 10. в; 11. в.

2. Статистическая группировка и сводка материала

Группировка – это расчленение единиц генеральной совокупности на однородные в отношении изучаемых признаков группы и подгруппы.

В зависимости от целей проведения выделяют три основных вида группировки:

1) **типологическую** – разделение всей совокупности единиц на качественно однородные группы (типы) (например, группировка предприятий по форме собственности, виду деятельности и т.д.);

2) **структурную** - разделение всей совокупности единиц с целью анализа структуры явления (например, группировка населения по полу, возрасту, количеству детей в семье);

3) **аналитическую** - разделение всей совокупности единиц с целью изучения взаимосвязи между отдельными признаками явления (например, группировка банков по стоимости активов для определения влияния ее на прибыль банков)

При проведении группировки и сводки материала выделяют два вида признаков: факторный (группировочный) и результативный. В свою очередь факторный признак обозначается как x , а результативный – как y . Расчленение всей совокупности единиц на группы проводится по **факторному** признаку. В зависимости от числа группировочных признаков выделяют:

- **простую группировку** – разделение единиц совокупности на группы по одному (факторному) признаку, здесь результативный признак отсутствует (см. пример 1);

- **сложную группировку** - разделение единиц совокупности на группы по двум и более факторным признакам (x_1, x_2, x_3, \dots).

Пример 1. Известны результаты сдачи студентами экзамена: 3; 4; 3; 4; 2; 4; 4; 5; 3; 4; 5; 5; 3; 4; 3; 5; 4; 4; 3; 4; 3; 5; 5; 4; 3; 4; 5; 3; 4; 4.

Проведите простую группировку и сделайте вывод.

Решение

Факторный признак здесь - оценка, полученная на экзамене.

Таблица 1. Распределение студентов по оценке, полученной на экзамене

Оценка	2	3	4	5	Итого
Число студентов, чел.	1	9	13	7	30

Вывод: один студент неуспевающий, 7 человек получили «отлично», наибольшее количество студентов сдали экзамен на «хорошо».

Факторные признаки могут быть:

- **атрибутивными** - качественными, например, профессия рабочих, социальная принадлежность населения;

- *количественными* - объемными, например, возраст рабочих, число человек в семье, количество мужчин в цехе.

При непрерывном изменении значений признака-фактора применяют разделение его значений на интервалы. **Интервал** – это разность между максимальным и минимальным значениями признака в группе. В статистике применяют три вида интервалов: *равные, неравные* (постепенно увеличивающиеся) и *специализированные*.

В контрольной работе №1 вам необходимо изучить **аналитическую группировку** с выделением равных интервалов факторного признака.

Этапы проведения аналитической группировки

Этап 1. Выбор из двух признаков результативного и факторного. При этом надо учесть, что факторный признак каким-то образом влияет на результативный. Например, из двух признаков: прибыль и производительность труда – прибыль является результативным признаком (y), а на него влияет факторный признак – производительность труда (x). При проведении любого вида группировки результативный признак один, а факторных может быть несколько.

Этап 2. Определение **числа групп** (если не задано) проводится по формуле Стерджесса: $k = 1.0 + 3.322 \cdot \lg N$,
где N – число единиц совокупности.

Этап 3. Определение **шага интервала** по формуле: $i = \frac{X_{\max} - X_{\min}}{k}$,
где X_{\max} – максимальное значение факторного признака в совокупности;
 X_{\min} – минимальное значение факторного признака в совокупности.

Этап 4. Определение границ групп. Границами группы является интервал значений факторного признака в группе, который содержит нижнюю и верхнюю границы интервала.

Нижняя граница первого интервала соответствует минимальному значению признака-фактора, **верхняя** – нижней границе плюс шаг интервала (см. табл. 2).

Таблица 2. Определение значений верхней и нижней границ интервала

№ группы	Нижняя граница интервала	Верхняя граница интервала
1	X_{\min}	$X_{\min} + i = X_1$
2	X_1	$X_1 + i = X_2$
3	X_2	$X_2 + i = X_3$
...
k	X_{k-1}	$X_{k-1} + i = X_k$

Верхняя граница интервала последней группы должна быть больше или равна максимальному значению признака в совокупности: $X_k \geq X_{\max}$.

Этап 5. Составление **вспомогательной (рабочей) таблицы**, которая является основанием для проведения сводки и обработки материала.

Этап 6. На основе рабочей таблицы строится **сводная групповая таблица**. Обе таблицы должны быть оформлены статистически грамотно: иметь заглавие, наименование подлежащего, сказуемого, единицы измерения показателей, итоговые характеристики показателей и т.д.

При проведении сводки материала вписываются итоговые строки, полученные в рабочей таблице, и определяются средние значения факторного и результативного признаков по каждой группе (\bar{x}_i, \bar{y}_i).

Этап 7. На основе проведенной сводки материала строится вывод, в котором отражается степень зависимости результативного признака от факторного. При построении вывода необходимо анализировать рассчитанные средние величины результативного и факторного признаков.

Пример 2. Используя данные таблицы 3, проведите группировку и сводку материала, выделив 4 группы, и сделайте вывод.

Таблица 3. Данные о работе предприятий легкой промышленности

№	Число работников в смену, чел.	Прибыль, млн. руб.	№	Число работников в смену, чел.	Прибыль, млн. руб.
1	35	16,5	6	38	16,6
2	40	17,4	7	34	16,2
3	30	15,5	8	33	16,0
4	42	17,2	9	36	16,7
5	37	17,3	10	31	15,3

Решение

1. Результативным признаком (y) в данном случае является «Прибыль», а факторным (x) – «Число работников в смену».

2. Число групп задано и равно: $k = 4$.

3. Определение шага интервала: $i = \frac{X_{\max} - X_{\min}}{k} = \frac{42 - 30}{4} = 3$ (чел.).

Таблица 4. Определение значений верхней и нижней границ интервала

№ группы	Нижняя граница интервала	Верхняя граница интервала
1	30	33 (30+3)
2	33	36 (33+3)
3	36	39 (36+3)
4	39	42 (39+3)

4. Расчет границ групп представлен в таблицу 4.

Данную таблицу можно не указывать, а границы группы представить сразу в рабочей таблице.

5. Проведение группировки (см. табл. 5).

Таблица 5. **Рабочая таблица**

№ группы	Границы группы	№ предприятия в группе	Число работников в смену, чел.	Прибыль, млн. руб.
1	30-33	3	30	15,5
		8	33	16,0
		10	31	15,3
Итого:		3	94	46,8
2	33-36	1	35	16,5
		7	34	16,2
		9	36	16,7
Итого:		3	105	49,4
3	36-39	5	37	17,3
		6	38	16,6
Итого:		2	75	33,9
4	39-42	2	40	17,4
		4	42	17,2
Итого:		2	82	34,6
Всего:		10	356	164,7

Посмотрите внимательно на группировку: в первую группу вошли предприятия, где численность работников составила 30, 31, 32 и 33 человека. Следовательно, во вторую группу войдут предприятия, где число работников в смену больше 33 человек, так как значение признака – 33 – уже включено в первую группу. Таким образом, под нижней границей интервалов подразумевается значения: 33,1; 36,1; 39,1.

6. Сводка материала представлена в таблице 6.

Таблица 6. **Распределение предприятий легкой промышленности по числу работников в смену**

№ гр	Границы группы	Число предприятий в группе n_i	Число работников в смену, чел.		Прибыль, млн. руб.	
			Всего по группе x_i	Среднее число работников в группе, чел. \bar{x}_i	Всего по группе y_i	Средняя прибыль в группе, млн.руб. \bar{y}_i
1	30-33	3	94	31,3	46,8	15,60
2	33-36	3	105	35,0	49,4	16,47
3	36-39	2	75	37,5	33,9	16,95
4	39-42	2	82	41,0	34,6	17,30
Итого:		10	356	35,6	164,7	16,47

Среднее значение признака по группе определяется по формулам:

$\bar{x}_i = \frac{x_i}{n_i}$; $\bar{y}_i = \frac{y_i}{n_i}$, а общее среднее значение признака определяется делением

итоговых показателей: $\bar{x} = \frac{\sum x}{n}$; $\bar{y} = \frac{\sum y}{n}$.

7. Вывод: С ростом числа работников в смену увеличивается прибыль предприятия. Если среднее число работников составляет 35 чел., то средняя прибыль равна 16,47 млн. руб., а если среднее число работников возрастает до 41 чел., то и прибыль увеличивается до 17,3 млн. руб.

Тесты для самоконтроля

1. Основанием группировки может быть:
 - а) качественный признак; б) количественный признак;
 - в) как качественный, так и количественный признак.
2. Наибольшее значение признака в интервале называется:
 - а) нижней границей; б) средней величиной; в) верхней границей.
3. Группировка единиц совокупности проводится по: а) результативному признаку; б) количественному признаку; в) факторному признаку.
4. Группировка, в которой происходит разбиение однородной совокупности на группы, называется: а) типологической группировкой; б) структурной группировкой; в) аналитической группировкой.
5. Разделение единиц совокупности на группы по одному факторному признаку при отсутствии результативного признака называют:
 - а) простой группировкой; б) комбинированной группировкой;
 - в) сложной группировкой; г) аналитической группировкой.
6. С помощью формулы Стерджесса определяется:
 - а) шаг интервала; б) количество групп; в) число единиц в группе.
7. В рабочей таблице проводится непосредственно:
 - а) сводка материала; б) группировка единиц совокупности;
 - в) определение средних величин факторного признака по группам.
8. Разность между максимальным и минимальным значениями признака в группе – это: а) интервал значений факторного признака; б) интервал значений результативного признака; в) временной интервал значений.
9. При расчете верхней границы интервала:
 - а) от нижней границы отнимается среднее значение признака-фактора;
 - б) к нижней границе прибавляется шаг интервала;
 - в) к нижней границе прибавляется среднее значение признака-фактора.
10. Среднее значение признака в группе определяется: а) разницей между наибольшим и наименьшим значениями признака в группе; б) делением суммы значений признака по группе на число единиц в группе; в) делением наибольшего значения признака в группе на число единиц в группе.

Правильные ответы: 1. в; 2. в; 3. в; 4. а; 5. а; 6. б; 7. б; 8. а; 9. б; 10. б.

3. Статистические величины

В статистике любой анализ исследуемой совокупности проводится с использованием статистических величин. *Статистическая величина* – это обобщающий показатель, характеризующий изменение, динамику, объем совокупности. В статистических исследованиях применяются три вида статистических величин: абсолютные, относительные и средние.

3.1. Абсолютные величины

Абсолютные величины характеризуют численность и размер (объемность) социально – экономического явления или процесса в определенные периоды времени.

Они образуются путем суммирования значений признака и всегда имеют единицы измерения. Выбор единицы измерения обусловлен сущностью изучаемого явления и задачами, поставленными перед исследователем.

Виды абсолютных величин

1) величины, характеризующие объем изучаемого явления *на строго определенную дату* (например, численность рабочих на первое июня);

2) величины, характеризующие объем явления *за определенный период времени* (например, объем реализованной продукции предприятия за полугодие, квартал, месяц, год).

Единицы измерения абсолютных величин

1) *натуральные* – единицы измерения соответствуют потребительским свойствам продукта:

- *простые* (например, ткань – в метрах; молоко – в литрах; цветочные горшки – в штуках);
- *сложные* – образуются из двух и более простых единиц измерения (например, рабочее время – в человеко-часах, человеко-днях; потребление электроэнергии – в кВт/час, скорость – в км/ч);

2) *условно–натуральные* – применяются в том случае, когда надо суммировать продукцию, обладающую общностью основного потребительского свойства (например, если на предприятии используют разные виды топлива (мазут, уголь, нефть,...), то при определении общего объема потребления топлива все виды топлива переводят в тонны условного топлива с помощью коэффициентов калорийности, а потом суммируют полученные значения)

3) *стоимостные* – денежные единицы измерения (рубли, доллары, тысячи рублей). Очень широко используются в экономико– статистических исследованиях.

Пример 3. На производственные нужды предприятие использует несколько видов топлива (табл. 7). Средние коэффициенты калорийности перевода в

условное топливо следующие: мазут – 1,37 т; уголь – 0,9 т; газ – 1,2 тыс. м³; нефть – 1,3 т.

Определите общее потребление условного топлива.

Таблица 7. Расход топлива на производственные нужды за 1 квартал

Вид топлива	Единицы измерения	Расход
1. Нефть	т	210
2. Газ природный	тыс. м ³	450
3. Уголь	т	320
4. Мазут топочный	т	480

Решение

Суммируем расход топлива с учетом коэффициентов пересчета:

$$y = 210 \cdot 1.3 + 450 \cdot 1.2 + 320 \cdot 0.9 + 480 \cdot 1.37 = 1758.6 \text{ (усл. ед.)}.$$

Общее потребление условного топлива составляет 1758,6 усл. ед.

3.2. Относительные величины

Относительные величины - это результат сопоставления двух статистических показателей, которое дает цифровую меру их соотношения. Получает их путем деления сравниваемого показателя на показатель, принятый за базу сравнения.

Виды относительных величин:

1. Результат сопоставления **одноименных** статистических показателей

1) результат сопоставления с прошлым периодом:

- **относительные величины динамики** – это коэффициенты роста, которые характеризуют изменение явления во времени. Определяется коэффициент роста по формуле: $K_p = \frac{y_1}{y_0}$,

где y_0 - уровень явления в базисном периоде;

y_1 - уровень явления в текущем периоде.

Коэффициент роста не имеет единиц измерения и показывает, во сколько раз увеличился или уменьшился объем явления за определенный период времени.

Если коэффициент роста перевести в проценты, т.е. умножить на 100, то получим **темпы роста**. Определить темпы роста можно двумя способами: базисным и цепным. **Темпы роста с переменной базой сравнения (цепные)** получают при сравнении уровня явления каждого периода с уровнем предшествующего периода:

$$T_{p_1} = \frac{y_1}{y_0}; T_{p_2} = \frac{y_2}{y_1}; T_{p_3} = \frac{y_3}{y_2} \text{ и т.д.}$$

Темпы роста с постоянной базой сравнения (базисные) получают путем сопоставления уровня явления в каждом периоде с уровнем периода, принятого за базу сравнения: $T_{p_1} = \frac{y_1}{y_0}$; $T_{p_2} = \frac{y_2}{y_0}$; $T_{p_3} = \frac{y_3}{y_0}$; ...,

где y_0 - база сравнения (как правило, уровень первого периода принимается за базисный);

y_1 - уровень явления второго периода;

y_2 - уровень явления третьего периода;

y_3 - уровень явления четвертого периода.

- **относительные величины планового задания** определяется как отношение величины показателя по плану ($y_{пл}$) к его фактической величине в предшествующем периоде (y_0): $i_{пл} = \frac{y_{пл}}{y_0}$;

2) результат сопоставления *с планом*

- **относительные величины выполнения плана** - это отношение фактической величины показателя (y_1) к запланированной на тот же период его величине ($y_{пл}$): $i_{впл} = \frac{y_1}{y_{пл}}$.

Относительные величины планового задания, выполнения плана и динамики связаны между собой: $\frac{y_1}{y_0} = \frac{y_{пл}}{y_0} \times \frac{y_1}{y_{пл}} \Rightarrow i_1 = i_{пл} \cdot i_{в.пл}$;

3) результат сопоставления *части и целого или частей между собой*

- **относительные величины структуры** характеризуют долю отдельных частей в общем объеме совокупности и выражаются в долях единицы или в процентах. Они определяются по сгруппированным данным с помощью формулы: $\omega = \frac{n}{N}$,

где n - число единиц или объем признака по группе;

N - общее число единиц или объем признака по всей совокупности.

Если полученное значение относительной величины структуры умножить на 100, получим *удельный вес*;

- **относительные величины координации** отражает отношение численности двух частей единого целого, т.е. показывает, сколько единиц одной группы приходится в среднем на одну, на десять или на сто единиц другой группы изучаемой совокупности (например, сколько служащих приходится на сто рабочих);

4) результат сопоставления *в пространстве*:

- **относительные величины наглядности** отражает результаты сопоставления одноименных показателей, относящихся к одному и тому же периоду (или моменту) времени, но к разным объектам или территориям (например, сравнивается производительность труда на двух предприятиях).

2. Результат сопоставления **разноименных** показателей:

- **относительные величины интенсивности** являются именованными числами и показывают итог числителя на одну, на десять или на сто единиц знаменателя. К ним относятся:

- *показатели производства продукции*: $K = \frac{Q}{N}$,

где K – показатель производства продукции на душу населения;

Q - выпуск продукции в натуральном выражении за год;

N - среднегодовая численность населения;

- *показатели потребления продуктов питания и непродовольственных товаров на душу населения*;

- *показатели, отражающие обеспеченность населения материальными и культурными благами*;

- *показатели, характеризующие техническую оснащенность производства, рациональность расходования ресурсов*.

Пример 4. Используя данные таблицы 8, определите:

1) динамику фактического выпуска продукции; 2) величину планового задания; 3) величину выполнения плана в каждом периоде; 4) удельный вес продукции A в общем объеме выпуска.

Таблица 8. Данные об объемах выпуска предприятием за два года

Вид продукции	Объем выпуска, тыс. ед.			
	Первый год		Второй год	
	По плану	Фактически	По плану	Фактически
A	5,0	4,8	5,6	5,8
B	8,5	8,6	9,0	9,2

Решение

1. Определим темпы роста по каждому периоду:

$$T_p = \frac{y_1}{y_0} = \frac{5.8}{4.8} = 1.2084 \quad (1.2084 \cdot 100 = 120.84\%);$$

$$T_p = \frac{y_1}{y_0} = \frac{9.2}{8.6} = 1.0698 \quad (1.0698 \cdot 100 = 106.98\%).$$

Таким образом, объемы выпуска обеих видов продукции во втором году по сравнению с первым возросли: объем выпуска продукции A – на 20,84%, продукции B – на 6,98%.

2. При определении величины планового задания происходит сравнение запланированного показателя с фактическим, полученным в предшествующем периоде:

- продукция A : $i_{пл} = \frac{y_{пл}}{y_0} = \frac{5.6}{4.8} = 1.1667 \quad (1.1667 \cdot 100 = 116.67\%);$

- продукция B : $i_{пл} = \frac{y_{пл}}{y_0} = \frac{9.0}{8.6} = 1.0465$ ($1.0465 \cdot 100 = 104.65\%$).

Вывод: во втором году было запланировано увеличение объемов выпуска как продукции A (на 16,67%), так и продукции B (на 4,66%).

3. Величина выполнения плана продукции A :

- первый год: $i_{впл} = \frac{y_1}{y_{пл}} = \frac{4.8}{5.0} = 0.96$ ($0.96 \cdot 100 = 96.0\%$);

- второй год: $i_{впл} = \frac{y_1}{y_{пл}} = \frac{5.8}{5.6} = 1.0357$ ($1.0358 \cdot 100 = 103.57\%$).

Величина выполнения плана продукции B :

- первый год: $i_{впл} = \frac{y_1}{y_{пл}} = \frac{8.6}{8.5} = 1.0118$ ($1.0118 \cdot 100 = 101.18\%$);

- второй год: $i_{впл} = \frac{y_1}{y_{пл}} = \frac{9.2}{9.0} = 1.0223$ ($1.0223 \cdot 100 = 102.23\%$).

Как видно из расчетов, в первом году наблюдалось недовыполнение плана по продукции A на 4,0 %, во втором году план был перевыполнен на 3,58 %; по продукции B и в первом, и во втором году план был перевыполнен: в первом – на 1,18 %, во втором – на 2,23 %.

4. Для определения удельного веса продукции A в общем объеме выпуска используем форму относительной величины структуры.

Первый год: $\omega = \frac{n}{N} \cdot 100 = \frac{4.8}{4.8 + 8.6} \cdot 100 = 35.82$ (%);

второй год: $\omega = \frac{n}{N} \cdot 100 = \frac{5.8}{5.8 + 9.2} \cdot 100 = 38.67$ (%).

Вывод: во втором году удельный вес продукции A увеличился на 2,85% ($38,67 - 35,82$).

Пример 5. На транспортном предприятии работают 18 водителей, из них водители первой категории – 11 человек, второй категории – 7 человек.

Определите удельный вес водителей первой категории, а также удельный вес водителей первой категории, приходящихся на одного водителя второй категории.

Решение

Удельный вес водителей первой категории в общем числе составляет:

$$\omega = \frac{n}{N} \cdot 100 = \frac{11}{18} \cdot 100 = 61.12$$
 (%)

Величину координации можно определить по формуле:

$$k = \frac{n_1}{n_2} \cdot 100 = \frac{7}{11} \cdot 100 = 63.64$$
 (%)

таким образом, 63,64 % водителей первой категории приходится на одного водителя второй категории.

Пример 6. Используя данные таблицы 9, проведите сравнение себестоимости, объемов выпуска и общей себестоимости продукции на двух предприятиях.

Таблица 9. Данные о себестоимости и объемах выпуска однородной продукции на двух предприятиях

Предприятие	Себестоимость, тыс. руб.	Объем выпуска, ед.
1	32,8	54
2	29,6	80

Решение

Для сравнения данных, относящихся к разным единицам наблюдения, используют относительную величину координации.

Сравним данные первого предприятия с данными на втором предприятии:

$$1) \text{ себестоимость: } k_z = \frac{z_1}{z_2} = \frac{32.8}{29.6} = 1.108 \text{ (110.8\%)};$$

$$2) \text{ объемы выпуска: } k_q = \frac{q_1}{q_2} = \frac{54}{80} = 0.675 \text{ (0.675} \cdot 100 = 67.5\%);$$

$$3) \text{ общей себестоимости: } k_{zq} = \frac{z_1 q_1}{z_2 q_2} = \frac{32.8 \cdot 54}{29.6 \cdot 80} = \frac{1771.2}{2368.0} = 0.748 \text{ (74.8\%)}.$$

Таким образом, себестоимость единицы продукции на первом предприятии по сравнению со вторым больше на 10,8%; объемы выпуска на первом предприятии на 32,5% ниже, чем на втором; общая себестоимость первого предприятия меньше, чем второго на 25,2%.

3.3. Степенные средние величины

Средней величиной называется обобщающий показатель, характеризующий уровень варьирующего количественного признака на единицу совокупности в определенных условиях места и времени. Средняя величина всегда имеет ту же размерность, что и признак у отдельных единиц совокупности.

В экономических исследованиях и плановых расчетах применяют две категории средних: степенные и структурные. К **степенным** средним относятся: средняя арифметическая, средняя гармоническая, средняя квадратическая и средняя геометрическая.

Формулы средних величин могут быть получены на основе функции

$$\text{степенной средней: } \bar{x} = \sqrt[n]{\frac{\sum_{i=1}^n x_i^k f_i}{\sum_{i=1}^n f_i}},$$

где \bar{x} - средняя;

x_i - величина, для которой исчисляется средняя;

f_i - частота (повторяемость) индивидуальных значений признака;

k - степень средней;

n - число единиц в совокупности.

В зависимости от степени (k) получаются различные виды средних величин, их формулы представлены в таблице 10. Чем выше степень средней, тем большее значение средней получается в результате расчета:

$$\bar{X}_{\text{гарм.}} < \bar{X}_{\text{геом}} < \bar{X}_{\text{арифм}} < \bar{X}_{\text{квадрат}}$$

Таблица 10. Формулы расчета степенных средних величин

Значение k	Наименование средней	Формула средней	
		простая	взвешенная
-1	Гармоническая	$\bar{x} = \frac{n}{\sum \frac{1}{x}}$	$\bar{x} = \frac{\sum f}{\sum \frac{1}{x} f}$; $\bar{x} = \frac{\sum w}{\sum \frac{1}{x} w}$
0	Геометрическая	$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_n}$	$\bar{x} = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_n^{f_n}}$
1	Арифметическая	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum xf}{\sum f}$; $\bar{x} = \frac{\sum xw}{\sum w}$
2	Квадратическая	$\bar{x} = \sqrt{\frac{\sum x^2}{n}}$	$\bar{x} = \sqrt{\frac{\sum x^2 \cdot f}{\sum f}}$

Вопрос о выборе средней решается в каждом отдельном случае, исходя из задачи исследования и наличия исходной информации.

Средняя арифметическая и **средняя гармоническая** наиболее распространенные виды средних, их выбор определяется характером имеющейся в распоряжении исследователя информации. Используются они при расчетах общей средней из средних групповых, а также при выявлении взаимосвязи между признаками с помощью группировок.

Пример 7. Имеются данные о численности семьи работников организации (чел.): 2; 6; 3; 5; 6; 3; 3; 5; 4; 4; 2; 2; 4; 3; 5; 4; 3; 3; 3.

Определите среднее число человек в семье с помощью формулы средней арифметической: 1) простой; 2) взвешенной.

Решение

1. Так как исходные данные не сгруппированы, применяем среднюю арифметическую простую:

$$\bar{x} = \frac{\sum x}{n} = \frac{2+6+3+5+6+3+3+5+4+4+2+2+4+3+5+4+3+3+3}{19} = \frac{70}{19} = 3.7 \text{ (чел.)}.$$

2. Для применения средней арифметической взвешенной данные необходимо сгруппировать (см. табл. 11).

Таблица 11. Распределение работников по числу человек в семье

Число человек в семье, чел. (x)	2	3	4	5	6	Итого
Число работников, чел. (f)	3	7	4	3	2	19

Теперь, для определения среднего числа человек в семье используем среднюю арифметическую взвешенную:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{2 \cdot 3 + 3 \cdot 7 + 4 \cdot 4 + 5 \cdot 3 + 6 \cdot 2}{3 + 7 + 4 + 3 + 2} = \frac{70}{19} = 3.7 (\text{чел.}).$$

Следовательно, среднее число человек в семье работников организации составляет 3,7 человека.

Пример 8. Используя данные таблицы 12, определите среднюю успеваемость студентов факультета.

Таблица 12. Данные об успеваемости студентов факультета

Успеваемость, % (x)	60-70	70-80	80-90	90-100
Число студентов, чел. (f)	10	18	52	20

Решение

Данные сгруппированы, следовательно, для выполнения задания будем использовать среднюю арифметическую взвешенную.

Так как факторный признак представлен интервалами значений, то в числителе формулы вместо x применим x' - середину соответствующего интервала, которая определяется:

$$x'_1 = \frac{60+70}{2} = 65; \quad x'_2 = \frac{70+80}{2} = 75; \quad x'_3 = \frac{80+90}{2} = 85; \quad x'_4 = \frac{90+100}{2} = 95.$$

$$\bar{x} = \frac{\sum x'f}{\sum f} = \frac{65 \cdot 10 + 75 \cdot 18 + 85 \cdot 52 + 95 \cdot 20}{10 + 18 + 52 + 20} = \frac{8320}{100} = 83.2 (\%)$$

Таким образом, средняя успеваемость студентов факультета составляет 83,2%.

Пример 9. Определите среднюю месячную заработную плату работников предприятия, используя данные таблицы 13.

Таблица 13. Данные о заработной плате работников предприятия

№ цеха	Средняя месячная заработная плата, руб.	Фонд заработной платы, руб.
1	1790	368740
2	1730	349460
3	1500	429000

Решение

Средняя месячная заработная плата определяется делением фонда зарплаты на число рабочих предприятия.

Обозначим «Среднюю месячную заработную плату» - x , а «Фонд заработной платы» - w . Тогда число работников предприятия можно определить по формуле: $\sum \frac{w}{x} = \frac{w_1}{x_1} + \frac{w_2}{x_2} + \frac{w_3}{x_3}$.

Исходя из вышесказанного формула расчета примет вид:

$$\bar{x} = \frac{\sum w}{\sum \frac{w}{x}} = \frac{368740 + 349460 + 429000}{\frac{368740}{1790} + \frac{349460}{1730} + \frac{429000}{1500}} = \frac{1147200}{694} = 1653.03 \text{ (руб.)}.$$

Для расчета средней в данном случае использовалась формула средней гармонической взвешенной.

Средняя квадратическая применяется для расчета среднего квадратического отклонения, являющегося показателем вариации признака, а также в технике (например, при сооружении трубопроводов).

Средняя геометрическая (простая) используется при вычислении среднего коэффициента роста (среднего темпа роста) в рядах динамики.

Пример 10. Определите средний темп роста, используя данные таблицы 14.

Таблица 14. Данные о продажах телевизоров за три года

Год	2000	2001	2002
Объем продажи телевизоров, тыс. ед.	18,4	19,0	24,3

Решение

Для определения темпа роста от года к году используем относительную величину динамики:

$$T_p = \frac{y_1}{y_0} = \frac{19.0}{18.4} = 1.033; T_p = \frac{y_2}{y_1} = \frac{24.3}{19.0} = 1.279.$$

Средний темп роста рассчитывается по формуле средней геометрической, где степень корня определяется число перемножаемых под корнем чисел: $\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_n}$; $\bar{x} = \sqrt[3]{1.033 \cdot 1.279} = 1.149$ ($1.149 \cdot 100 = 114.9\%$).

Вывод: средний темп роста за три года составляет 114,9%

3.4. Структурные средние величины

К **структурным** средним величинам, применяемым в статистических исследованиях, относят моду и медиану. В отличие от степенных средних, структурные средние выступают как конкретные величины, совпадающие со значениями признака отдельных единиц совокупности.

Мода (M_o) – это наиболее часто встречающееся значение признака в статистическом ряду. При анализе рынка потребительских товаров модой может быть, например, наиболее распространенный размер обуви, одежды.

Для *графического определения* структурных средних используют три вида кривых: кумуляту, полигон распределения и гистограмму частот.

При определении моды используют два последних вида графиков.

Для построения *гистограммы* по оси абсцисс откладывают значения признака, а частоты повторения изображаются прямоугольниками (см. рис.3). Частота повторения обозначается буквой f и указывает, какое количество значений признака встречается в совокупности (в примере 10 частотой повторения является число рабочих определенного разряда: разряд 2 встречается 15 раз, 3 – 24 раза и т.д.).

С целью определения моды правую вершину модального прямоугольника (самого высокого) соединяют с правым верхним углом предыдущего прямоугольника, а левую – с левым верхним углом последующего прямоугольника. Абсцисса точки пересечения этих прямых и будет модой.

Полигон частот или **полигон распределения** – это замкнутая ломаная линия (см. рис. 4). По оси абсцисс откладывают значения признака, по оси ординат – частоты повторения. Полигон распределения всегда замыкается на оси абсцисс: на рисунке 4 полигон выходит с первого разряда, что свидетельствует о том, что рабочих первого разряда в цехе нет; замыкается полигон в седьмом разряде, что также свидетельствует об отсутствии его среди рабочих цеха.

Мода в данном случае находится как перпендикуляр, опущенный на ось абсцисс с самой высокой точки полигона.

Пример 11. Имеются данные о разрядах рабочих цеха, определите моду с помощью таблицы и графическим способом.

Таблица 15. Распределение рабочих цеха по уровню квалификации

Квалификационный разряд (x)	2	3	4	5	6
Число рабочих, чел. (f)	15	24	40	13	8

Решение

По таблице видно, что чаще всего встречается 4 квалификационный разряд: 40 рабочих имеют данный разряд, следовательно, $M_o = 4$.

На рисунках 3 и 4 представлено графическое определение моды.

Мода на рисунке 3 соответствует значению 3,4 разряда. Так как разряд – целая, неделимая, дискретная величина, то округление происходит всегда в большую сторону. Следовательно, $M_o \approx 4$.

По полигону частот (рис. 4) $Mo = 4$.

Медиана (M_e) – значение признака, которое располагается в середине ранжированного ряда и делит этот ряд на две равные по численности части.

Ранжированный ряд – это ряд статистических показателей, выстроенный в порядке возрастания или убывания значений признака.

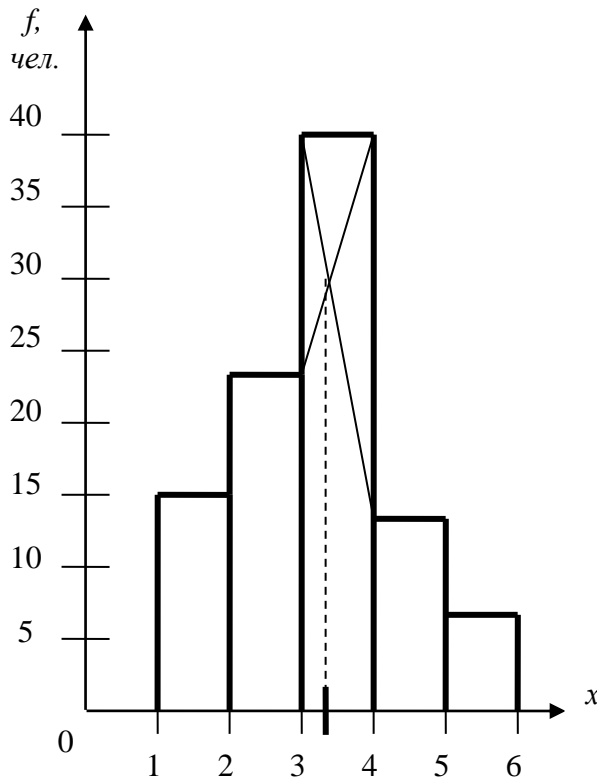


Рис. 3. Гистограмма

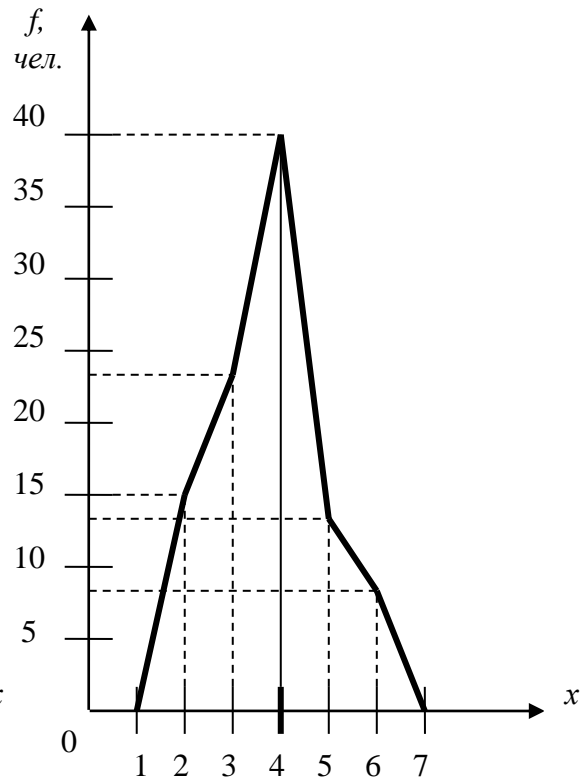


Рис. 4. Полигон частот

Сначала, для определения значения медианы, находится ее место в ранжированном ряду по формуле: $N_{Me} = \frac{n+1}{2}$,

где n – число единиц в статистическом ряду.

Если число единиц в ряду четное, то медиану принимают равной средней арифметической величине из двух срединных значений.

Медиана применяется при статистическом контроле качества продукции и технологического процесса на промышленных предприятиях, при изучении распределения семей по величине дохода,...

Графически медиану определяют с помощью кумуляты.

Построение **кумулятивной кривой (кумуляты)**: на оси ординат откладывают накопленные частоты, на оси абсцисс – значения признака. Соединяя полученные точки плавной линией, которая, начиная с нуля, непрерывно поднимается над осью абсцисс до тех пор, пока не достигнет высоты, соответствующей общей сумме частот. По кумуляте определяется **медиана**. Для ее нахождения высоту наибольшей ординаты делят пополам.

Через полученную точку проводят прямую, параллельную оси абсцисс, до пересечения ее с кумулятой. Абсцисса точки пересечения является медианной величиной.

Пример 12. Дан ряд статистических показателей, отражающих возраст рабочих цеха (лет): 22; 36; 20; 30; 24; 36; 44; 20; 22; 28; 38; 26; 30.

Определите медиану.

Решение

Проранжируем заданный ряд в порядке роста значений признака:

20; 20; 22; 22; 24; 26; 28; 30; 30; 36; 36; 38; 44.

Теперь определим место медианы в ряду: $N_{Me} = \frac{n+1}{2} = \frac{13+1}{2} = 7$, т. е. седьмое по счету значение признака является медианой в данном ряду. Следовательно, медиана равна: $M_e = 28$ лет.

Для построения кумуляты необходимо сгруппировать данные и рассчитать накопленные частоты (см. табл. 16).

Таблица 16. Распределение рабочих цеха по возрасту.

Возраст (x)	20	22	24	26	28	30	36	38	44
Число рабочих, чел. (f)	2	2	1	1	1	2	2	1	1
Накопленная частота, чел. (S)	2	4	5	6	7	9	11	12	13

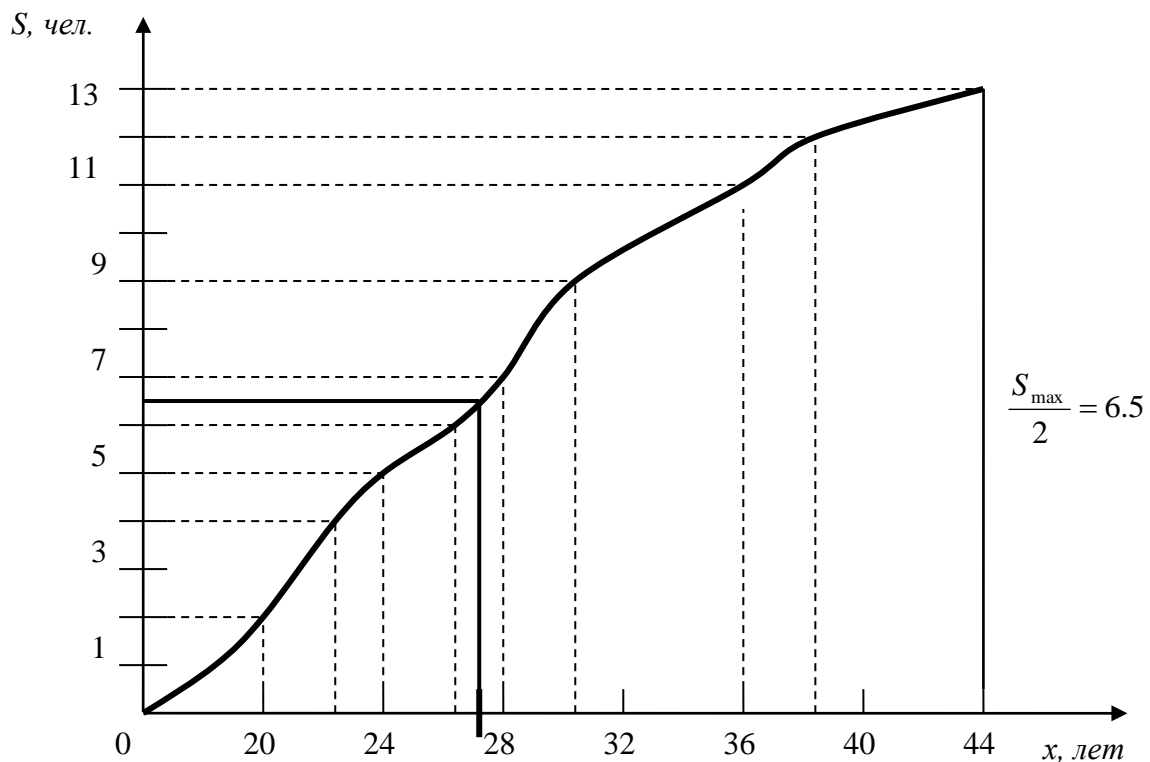


Рис. 5. Кумулята

Накопленная частота определяется как сумма частот текущего показателя с предшествующими, например, $S_{24} = 2 + 2 + 1 = 5$ (чел.).

Графическим способом полученное значение медианы соответствует:
 $Me = 27$ лет

Если факторный признак в группировке данных представлен интервалами значений, то определение моды и медианы имеет особенности.

При определении **медианы** сначала указывают *медианный интервал* – это первый интервал, в котором сумма накопленных частот превысит половину общего числа наблюдений.

Числовое значение медианы определяется по формуле:

$$Me = x_{Me} + i \cdot \frac{\frac{n+1}{2} - S_{(-1)}}{f_{Me}},$$

где x_{Me} – нижняя граница медианного интервала;

i – величина интервала;

$S_{(-1)}$ – накопленная частота интервала, предшествующего медианному;

f – частота медианного интервала.

При определении **моды** в интервальном ряду сначала определяется *модальный интервал*, т.е. тот интервал, который имеет наибольшую частоту повторения. Конкретное значение моды определяется по формуле:

$$Mo = x_{Mo} + i \cdot \frac{f_{Mo} - f_{(Mo-1)}}{[f_{Mo} - f_{(Mo-1)}] + [f_{Mo} - f_{(Mo+1)}]},$$

где x_{Mo} – нижняя граница модального интервала;

f_M – частота модального интервала;

$f_{(Mo-1)}$ – частота интервала, предшествующего модальному;

$f_{(Mo+1)}$ – частота интервала, следующего за модальным.

Пример 13. Используя данные таблицы 17, определите аналитическим и графическим способами моду и медиану.

Таблица 17. Распределение рабочих цеха по возрасту

Возраст, лет (x)	20-26	26-32	32-38	38-44	44-50
Число рабочих, чел. (f)	6	22	20	10	2
Накопленная частота, чел. (S)	6	28	48	58	60

Решение

Модальным интервалом является интервал [26-32], так как здесь наблюдается наибольшая частота повторения. Числовое значение определим по формуле:

$$Mo = x_{Mo} + i \cdot \frac{f_{Mo} - f_{(Mo-1)}}{[f_{Mo} - f_{(Mo-1)}] + [f_{Mo} - f_{(Mo+1)}]} = 26 + 6 \cdot \frac{22 - 6}{(22 - 6) + (22 - 20)} = 31.3 \text{ (лет)}.$$

Для определения медианного интервала, накопленную частоту разделим пополам: $\frac{S_{\max}}{2} = \frac{60}{2} = 30$. Полученное значение впервые встречается в строке накопленной частоты в третьем интервале: [32-38], следовательно, он является медианным.

Числовое значение медианы определим по формуле:

$$Me = x_{Me} + i \cdot \frac{\frac{n+1}{2} - S_{(-1)}}{f_{Me}} = 32 + 6 \cdot \frac{\frac{60+1}{2} - 28}{20} = 32.75 \text{ (лет)}.$$

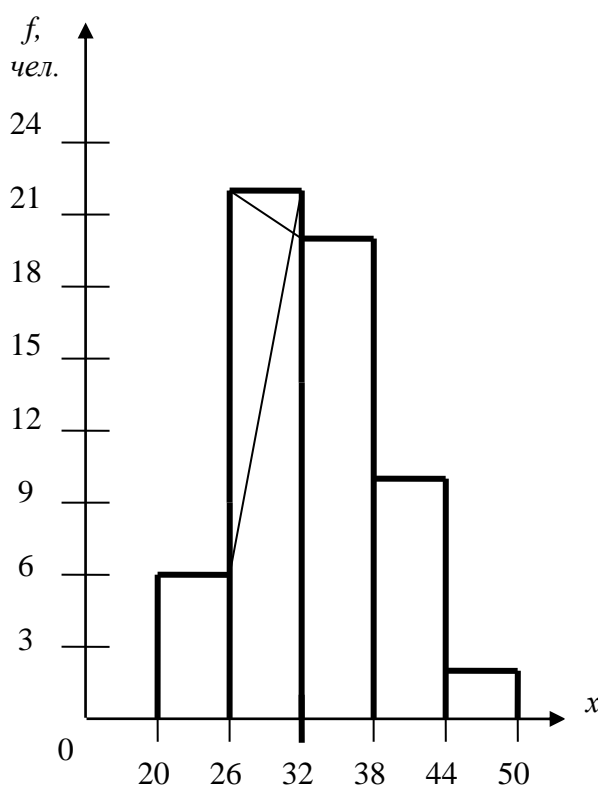


Рис. 6. Гистограмма

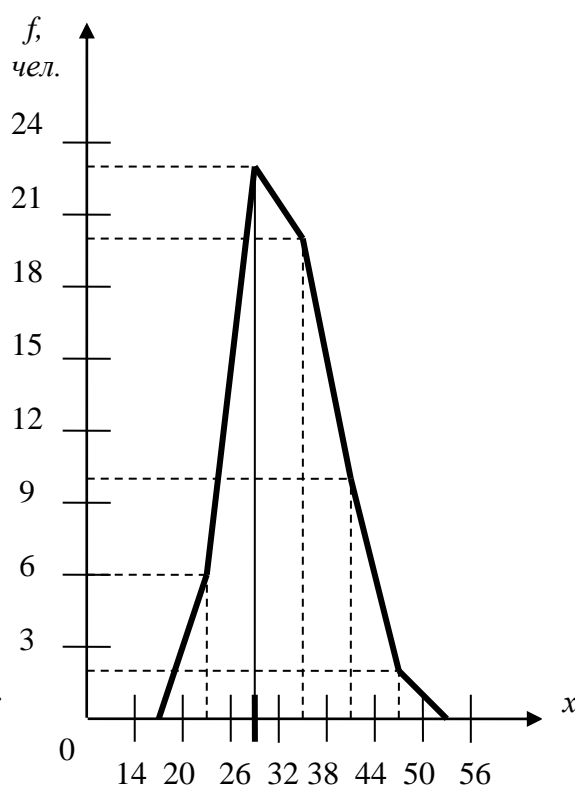


Рис. 7. Полигон частот

На рисунках 6 и 7 представлены гистограмма и полигон частот для определения моды, а на рисунке 8 – кумулята для определения медианы.

При построении для интервального ряда:

- **гистограммы** столбик выходит из нижней границы интервала и замыкается в верхней, т. е. охватывает весь интервал значений;
- **полигона частот** координатой точки является соответствующая частота повторения и *середина определенного интервала*;
- **кумуляты** координатой точки является соответствующая накопленная частота и *верхняя граница определенного интервала*.

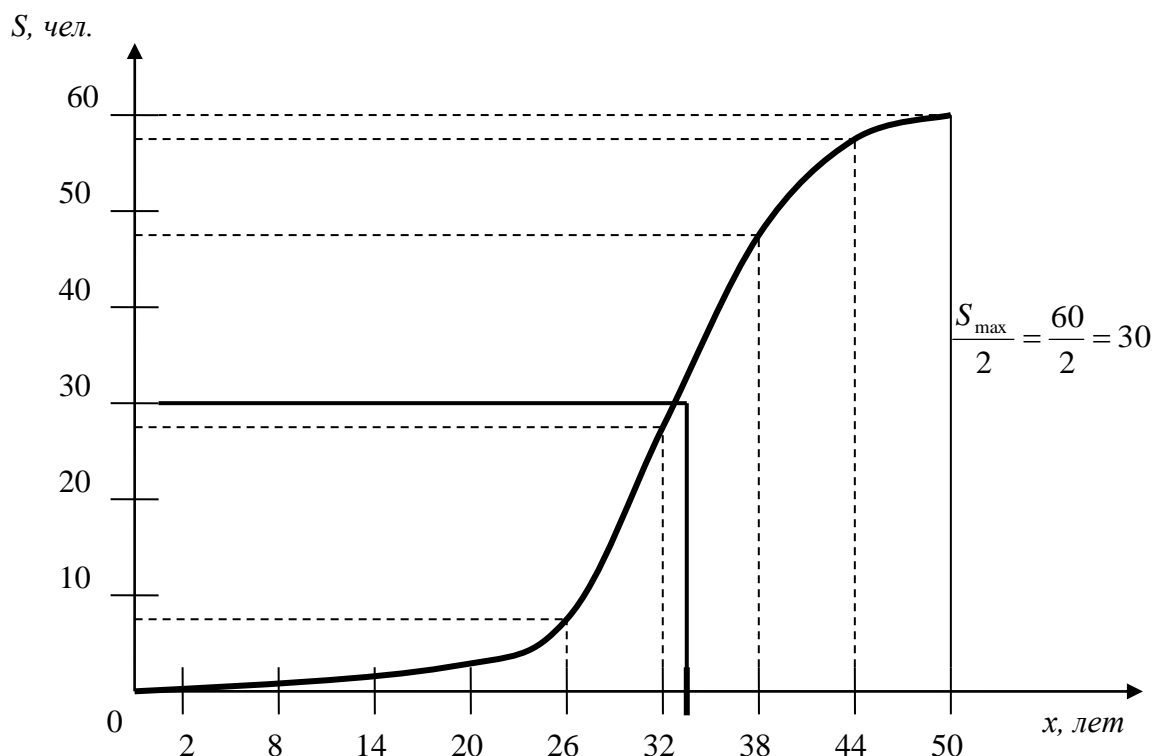


Рис. 8. Кумулята

При определении графическим способом получили значения:

- 1) моды близкое к 32 годам (рис. 6); 2) моды $M_o \approx 32$ г. (рис. 7);
- 3) медианы $M_e \approx 34$ г. (рис. 8).

Тесты для самоконтроля

1. Выберите, какие единицы измерения имеют абсолютные величины:
 - а) проценты; б) стоимостные; в) разы; г) не имеют единиц измерения.
2. Результат сопоставления двух статистических показателей, получаемый путем деления одного абсолютного показателя на другой и дающий цифровую меру их соотношения – это величина:
 - а) абсолютная; б) относительная; в) средняя.
3. Обобщающий показатель, характеризующий уровень варьирующего признака на единицу совокупности в определенных условиях места и времени – это величина:
 - а) абсолютная; б) относительная; в) средняя.
4. Показатель, полученный суммированием первичных данных – это величина: а) абсолютная; б) относительная; в) средняя.
5. Относительную величину динамики можно определить:
 - а) делением относительной величины выполнения плана на величину планового задания;
 - б) разницей между относительной величиной выполнения плана на величиной планового задания;

- в) произведением относительной величины выполнения плана на величину планового задания.
6. Доля единиц, обладающих определенным признаком, в общем числе единиц совокупности является относительной величиной:
а) координации; б) наглядности; в) структуры; г) интенсивности.
 7. При сгруппированных данных для определения среднего уровня ряда используются формулу средней:
а) простой; б) структурной; в) взвешенной.
 8. Наиболее часто встречающееся значение признака – это:
а) медиана; б) средняя; в) мода; г) величина структуры.
 9. Ряд, выстроенный в порядке возрастания или убывания значений признака, - это: а) атрибутивный ряд; б) ранжированный ряд;
в) ряд средних величин; г) дискретный ряд.
 10. При определении моды графическим способом строится:
а) кумулята; б) гистограмма; в) секторная диаграмма; г) полигон частот.
 11. При определении медианы графическим способом строится:
а) секторная диаграмма; б) гистограмма; в) кумулята; г) полигон частот.

Правильные ответы: 1. б; 2. б; 3. в; 4. а; 5. в; 6. в; 7. в; 8. в; 9. б; 10. б, г; 11. в.

4. Статистическое распределение

Сводная обработка данных статистического наблюдения предполагает построение рядов распределения, основной целью которого является выявление свойств и закономерностей развития исследуемой совокупности.

Выделяют *два вида рядов распределения*:

1) **атрибутивный ряд** – ряд распределения, построенный по качественным признакам (распределение населения по половому признаку, по национальности, образованию);

2) **вариационный ряд** - ряд распределения, построенный по количественным признакам (распределение населения по возрасту, по числу человек в семье, стажу работы).

Основными характеристиками вариационных рядов являются:

- показатели центра распределения;
- показатели степени вариации;
- показатели формы распределения.

Изучение вариации предполагает:

- построение вариационного ряда;
- его графическое изображение;
- определение основных характеристик распределения.

4.1. Понятие и измерение вариации признака

Различия индивидуальных значений признака у единиц совокупности называют **вариацией признака**. Вариация признака возникает из-за отличительных особенностей единиц совокупности, например, оценка, полученная студентом на экзамене, зависит от его способности воспринимать материал в ходе изучения, посещения занятий, способности самостоятельно изучать научную литературу и т.д.

При анализе вариации признака используются абсолютные и относительные показатели. К *абсолютным показателям* относятся:

1) **размах колебаний** (размах вариации): $R = X_{\max} - X_{\min}$,

где X_{\max} , X_{\min} - максимальное и минимальное значения факторного признака в совокупности, соответственно;

2) **среднее линейное отклонение** показывает, на сколько в среднем отличаются индивидуальные значения признака от его общей средней величины, определяется по формулам:

- для несгруппированных данных: $\bar{d} = \frac{\sum |x - \bar{x}|}{n}$;

- для сгруппированных данных: $\bar{d} = \frac{\sum |x - \bar{x}| f}{\sum f}$;

3) **среднее квадратическое отклонение** также показывает, на сколько в среднем отличаются индивидуальные значения признака от его среднего значения, определяется по формулам:

- для несгруппированных данных: $\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$;

- для сгруппированных данных: $\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}}$;

4) **дисперсия** – это средняя величина квадратов отклонений признака:

- для несгруппированных данных: $\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$;

- для сгруппированных данных: $\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}$.

Пример 14. Используя данные таблицы 18, проведите анализ вариации факторного признака с помощью абсолютных показателей.

Решение

Факторным признаком в данном случае является «Число работников в смену». Проанализируем его вариацию с помощью абсолютных показателей.

1. Размах вариации: $R = X_{\max} - X_{\min} = 42 - 30 = 12$ (чел.).

2. Так как данные не сгруппированы, то для расчета среднего линейного отклонения применим простую формулу (см. табл. 19):

$$\bar{d} = \frac{\sum |x - \bar{x}|}{n} = \frac{30}{10} = 3.0 \text{ (чел.)}.$$

Таблица 18. Данные о работе предприятий легкой промышленности

№	Число работников в смену, чел.	Прибыль, млн. руб.	№	Число работников в смену, чел.	Прибыль, млн. руб.
1	35	16,2	6	38	16,6
2	40	17,4	7	34	16,2
3	30	15,5	8	33	16,0
4	42	17,2	9	36	16,7
5	37	17,3	10	31	15,3

3. Среднее квадратическое отклонение (см. табл. 19):

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{\frac{130.4}{10}} = 3.6 \text{ (чел.)}.$$

4. Дисперсия (см. табл. 19): $\sigma^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{130.4}{10} = 13.04 \text{ (чел.)}.$

Таблица 19. Вычисление абсолютных показателей вариации

№ предпр.	Число работников в смену, чел. (x)	Прибыль, млн. руб. (y)	$ x - \bar{x} $	$(x - \bar{x})^2$
1	35	16,2	0,6	0,36
2	40	17,4	4,4	19,36
3	30	15,5	5,6	31,36
4	42	17,2	6,4	40,96
5	37	17,3	1,4	1,96
6	38	16,6	2,4	5,76
7	34	16,2	1,6	2,56
8	33	16,0	2,6	6,76
9	36	16,7	0,4	0,16
10	31	15,3	4,6	21,16
Итого	356	164,7	30,0	130,4

Общее среднее значение факторного признака: $\bar{x} = \frac{356}{10} = 35.6 \text{ (чел.)}.$

Относительные показатели вариации определяются как отношение абсолютных показателей вариации к средней арифметической или медиане:

1) коэффициент осцилляции: $K_R = \frac{R}{\bar{x}} \cdot 100;$

2) относительное линейное отклонение: $K_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100;$

3) **коэффициент вариации**: $V_x = \frac{\sigma}{\bar{x}} \cdot 100$.

Наиболее часто применяется *коэффициент вариации*, с помощью которого не только характеризуется вариация признака, но и проводится проверка совокупности на однородность. Совокупность считается однородной, если $V_x \leq 33\%$.

Пример 15. Используя данные примера 13, а также рассчитанные абсолютные показатели, проведите проверку совокупности на однородность.

Решение

Рассчитаем коэффициент вариации: $V_x = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{3.6}{35.6} \cdot 100 = 10.12 \%$.

Анализируемая совокупность является однородной, так как $V_x = 10.12 < 33\%$.

4.2. Виды дисперсий и правило их сложения

Выделяют три показателя дисперсии:

1) **общая дисперсия** характеризует вариацию признака, сформированную под влиянием всех факторов, определяющих уровень признака у единиц совокупности:

$$\sigma_0^2 = \frac{\sum (x - \bar{x}_0)^2 f}{\sum f},$$

где x_0 - общая средняя арифметическая для всей изучаемой совокупности;

2) **межгрупповая дисперсия** (дисперсия групповых средних) отражает различия в значениях изучаемого признака, которые возникают под влиянием факторного признака:

$$\delta_2 = \frac{\sum (\bar{x}_i - \bar{x}_0)^2 n_i}{\sum n_i},$$

где \bar{x}_i - средняя по определенной группе;

n_i - число единиц в определенной группе.

3) **средняя внутригрупповая дисперсия** характеризует случайную вариацию, возникающую под влиянием неучтенных факторов:

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i},$$

где σ_i^2 - дисперсия по отдельной группе, определяется по формуле:

$$\sigma_i^2 = \frac{\sum (x - \bar{x}_i)^2 f}{\sum f}.$$

Правило сложения дисперсий: величина общей дисперсии равна сумме межгрупповой дисперсии и средней внутригрупповой дисперсии:

$$\sigma_0^2 = \delta^2 + \bar{\sigma}^2.$$

Пример 16. Используя данные таблицы 20, проверьте правило сложения дисперсий.

Таблица 20. Данные о заработной плате десяти рабочих за май

Профессия	Число рабочих	Месячная заработная плата каждого рабочего, руб.
Токари	4	1252; 1548; 1600; 1400
Слесари	6	1450; 1380; 1260; 1700; 1250; 1372

Решение

Общая средняя величина факторного признака равна

$$\bar{x}_0 = \frac{\sum x}{n}; \quad \bar{x}_0 = \frac{1252+1548+\dots+1372}{10} = \frac{142120}{10} = 14212 \text{ (руб.)}$$

Общая дисперсия определим по формуле, используя результаты расчетов таблицы 21: $\sigma^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{215097,6}{10} = 21509,76 \text{ (руб.)}$.

Таблица 21. Вспомогательная таблица

№	Месячная заработная плата каждого рабочего, руб.	$x - \bar{x}$	$(x - \bar{x})^2$
1	1252	-169,2	28628,64
2	1548	126,8	16078,24
3	1600	178,8	31969,44
4	1400	-21,2	449,44
5	1450	28,8	829,44
6	1380	-41,2	1697,44
7	1260	-161,2	25985,44
8	1700	278,8	77729,44
9	1250	-171,2	29309,44
10	1372	-49,2	2420,64
Σ	14212,0	-	215097,6

Для определения средней внутригрупповой дисперсии необходимо рассчитать дисперсию по каждой группе. Воспользуемся составленной рабочей таблицей 22.

Среднее значение признака в первой группе: $\bar{x}_1 = \frac{5800}{4} = 1450,0 \text{ (руб.)}$;

во второй группе - $\bar{x}_2 = \frac{8412}{6} = 1402,0 \text{ (руб.)}$.

Используя результаты расчета (табл.22, последняя колонка), определим внутригрупповые дисперсии:

$$\sigma_1^2 = \frac{\sum (x - \bar{x}_1)^2 \cdot f}{\sum f} = \frac{73808}{4} = 184520 \text{ (руб.)};$$

$$\sigma_2^2 = \frac{\sum (x - \bar{x}_2)^2 \cdot f}{\sum f} = \frac{135760}{6} = 22626.67 \text{ (руб.)}.$$

Подставим полученные значения дисперсий каждой группы в формулу средней внутригрупповой дисперсии:

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i} = \frac{184520 \cdot 4 + 22626.67 \cdot 6}{4 + 6} = \frac{209568}{10} = 20956.8 \text{ (руб.)}.$$

Таблица 22. Вспомогательная таблица расчета групповых дисперсий

Месячная заработная плата рабочего, руб. x	Число рабочих, чел. f	$x - \bar{x}_1$	$(x - \bar{x}_1)^2 \cdot f$
Группа 1 - Токари			
1252	1	-198	39204
1548	1	98	9604
1600	1	150	22500
1400	1	-50	2500
Итого: 5800	4	-	73808
Группа 2 - Слесари			
1450	1	48	2304
1380	1	-22	484
1260	1	-142	20164
1700	1	298	88804
1250	1	-152	23104
1372	1	-30	900
Итого: 8412	6	-	135760

Межгрупповая дисперсия равна:

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x}_0)^2 n_i}{\sum n_i} = \frac{(1450 - 1421.2)^2 \cdot 4 + (1402 - 1421.2)^2 \cdot 6}{10} = 552.96 \text{ (руб.)}.$$

Теперь проверим правило сложения дисперсий – общая дисперсия равна сумме средней внутригрупповой и межгрупповой дисперсий:

$$\sigma_0^2 = \delta^2 + \bar{\sigma}^2 = 552.96 + 20956.8 = 21509.76 \text{ (руб.)}.$$

Таким образом, правило сложения дисперсий выполняется.

4.3. Способы построения вариационного ряда

Вариационный ряд – это статистический ряд, представленный в виде групповой таблицы, построенной по количественному признаку. В сказуемом данной таблицы отражается число единиц в каждой группе.

Таблица 23. Представление вариационного ряда

Признак-фактор	x_1	x_2	x_3	$\dots x_n$
Частота повторения	f_1	f_2	f_3	$\dots f_n$
Частость	$w_1 = \frac{f_1}{\sum f}$	$w_2 = \frac{f_2}{\sum f}$	$w_3 = \frac{f_3}{\sum f}$	$w_i = \frac{f_i}{\sum f}$
Накопленная частота	$S_1 = f_1$	$S_2 = f_1 + f_2$	$S_3 = S_2 + f_3$	$S_n = S_{n-1} + f_n$

Как правило, ряд распределения представляют в таблице, состоящей из четырех строк (см. табл. 23):

- в *первой строке* таблицы указываются конкретные значения каждого индивидуального **значения признака-фактора** (x_i);
- во *второй строке* отражается численность единиц с определенным значением признака, т.е. **частота повторения** (f_i);
- в *третьей строке* определяются **частости** - частоты, выраженные в относительных единицах (долях или процентах):

$$w_i = \frac{f_i}{\sum f_i},$$

где f_i – частоты ряда;

$\sum f_i$ - общая сумма частот: $\sum f_i = f_1 + f_2 + f_3 + \dots + f_n$;

- в *четвертой строке* – определяются **накопленные частоты** (S_i) путем последовательного прибавления к частоте первого интервала частот последующих интервалов.

Способ построения вариационного ряда зависит от характера изменения изучаемого признака, он может быть построен в форме дискретного ряда или в форме интервального ряда распределения.

Вариационный ряд представляется в форме **дискретного ряда**, если:

- факторный признак представлен дискретными величинами (неделимыми, целыми);
- число значений изучаемого признака небольшое.

Пример 17. Имеются данные о квалификационном разряде рабочих цеха: 3; 5; 4; 4; 4; 5; 3; 3; 2; 5; 3; 4; 2; 4; 3; 4; 5; 4; 4; 2; 3; 4; 4; 3; 4; 4; 5; 3; 4.

Постройте дискретный ряд распределения.

Решение

Для построения дискретного ряда распределения необходимо в первой строке таблицы указать встречающиеся разряды по мере их возрастания. Во второй строке – подсчитать количество каждого разряда. В третьей строке – определить частость, в последней строке таблицы – накопленную частоту (табл. 24).

Таблица 24. Распределение рабочих цеха по уровню квалификации.

Разряд (x)	2	3	4	5	Итого
Количество рабочих, чел. (f)	3	8	13	5	29
Частость (w)	$\frac{3}{29} = 0.1$	$\frac{8}{29} = 0.28$	$\frac{13}{29} = 0.45$	$\frac{5}{29} = 0.17$	1,0
Накопленная частота, чел. (S)	3	11=3+8	24=11+13	29=24+5	-

Пример 18. Известны данные о возрасте рабочих предприятия: 46; 20; 40; 35; 33; 30; 28; 36; 38; 38; 32; 24; 25; 33; 32; 40; 39; 39; 34; 28; 33; 40; 35; 26; 22; 34; 35 20; 26; 49.

Представьте интервальный ряд распределения, выделив 3 группы.

Решение

Для признака, имеющего непрерывное изменение, строится *интервальный вариационный ряд*, состоящий, так же как и дискретный ряд, из четырех строк. При его построении в первой строке отдельные значения признака-фактора указываются в виде интервалов, во второй строке – число единиц, входящих в интервал. Интервалы используются, как правило, равные и закрытые.

Величина шага интервала определяется по формуле: $i = \frac{R}{k}$,

где R – размах колебаний признака: $R = X_{\max} - X_{\min}$; k – число групп.

Число групп задано - 3. Величину шага интервала округляют до целого числа. Определим шаг интервала: $i = \frac{50-20}{3} = 10$ (лет). Далее определяется верхняя и нижняя границы интервалов (табл. 1).

Ряд распределения представим в таблице 25.

Таблица 25. Распределение рабочих по возрасту

Возраст, лет (x)	20-30	30-40	40-50	Итого
Количество рабочих, чел. (f)	10	18	2	30
Частость (w)	0,33	0,6	0,07	1,0
Накопленная частота, чел. (S)	10	28	30	-

4.4. Показатели центра распределения

К показателям центра распределения относятся: средняя арифметическая, мода и медиана.

Для дискретного и интервального рядов распределения *средняя арифметическая* определяется по формулам:

$$\bar{x} = \frac{\sum x \cdot f}{\sum f}, \quad \bar{x}' = \frac{\sum x' \cdot f}{\sum f},$$

где x – варианты значений признака;

x' – середина соответствующего интервала значения признака

f – частота повторения данного варианта.

Пример расчета среднего уровня для дискретного ряда распределения приведен в примере 7 (стр. 27), для интервального – в примере 8 (стр. 28).

Медиана (Me) соответствует значению признака, стоящему в середине ранжированного ряда.

Мода (Mo) – наиболее часто встречающееся значение признака.

Подробно структурные средние величины рассмотрены в параграфе 3,4; пример расчета медианы для дискретного ряда приведен в примере 11 (стр. 30), для интервального ряда – в примере 12 (стр. 32). Расчет моды для интервального ряда представлен в примере 12 (стр. 32), для дискретного ряда – в примере 10 (стр. 29).

4.5. Сопоставимость рядов распределения

Для проведения сравнительного анализа двух и более рядов распределения, в которых значения факторного признака представлены неравными интервалами или интервалами значений, имеющими разрывы, необходимо на первом этапе привести ряды к сопоставимому виду. Сопоставимость рядов распределения характеризуется равными интервалами значений факторного признака и отсутствием разрывов в цепи данных (например, представлены интервалы данных о возрасте: 20-24; 24-30; 34-38, здесь, во-первых, не равные интервалы, во-вторых – есть разрыв: отсутствуют данные о возрасте 30-34 года).

Для приведения ряда к сопоставимому виду необходимо рассчитать абсолютную или относительную плотности распределения. Если частота повторения представлена абсолютными величинами (число рабочих, число студентов), то рассчитывается абсолютная плотность распределения.

Абсолютная плотность распределения – это количество рабочих (студентов и т.д., т.е. величина частоты повторения) приходящихся на единицу шага интервала. Определяется она по формуле: $p = \frac{f}{i}$,

где f – частота повторения данного интервала в совокупности;

i – размер (шаг) интервала.

Если частота повторения представлена относительными величинами (удельный вес банков, доля рабочих), то рассчитывается относительная плотность распределения.

Относительная плотность распределения: $p' = \frac{w}{i}$,

где w – частость отдельной группы.

Пример 19. Приведите ряды к сопоставимому виду, используя данные таблицы 26.

Таблица 26. Успеваемость студентов двух групп

Группа 1		Группа 2	
Успеваемость, %	Число студентов, чел.	Успеваемость, %	Число студентов, чел.
60-70	4	50-60	2
70-90	10	60-80	10
90-100	6	80-90	8
Итого	20	90-100	4
		Итого	24

Решение

Заданные ряды распределения не сопоставимы, так как интервалы успеваемости в первой группе отличаются от интервалов второй группы.

Частота повторения представлена абсолютными показателями – числом студентов, следовательно, будем рассчитывать абсолютную плотность распределения.

Перед расчетом плотности, составим итоговую таблицу, где представим факторный признак - «успеваемость» - равными интервалами для обеих групп студентов (табл. 27).

В примере лучше взять шаг интервала, равный 10 %, так как минимальное значение факторного признака равно 50 % (табл. 26), а максимальное – 100 %. Шаг интервала при приведении рядов к сопоставимому виду выбираются исследователем самостоятельно. Таким образом, в таблице 26 на данный момент заполнено лишь подлежащее, т.е. первая графа.

Таблица 27. Распределение студентов по уровню успеваемости

Успеваемость, %	Группа 1	Группа 2
50-60	0	2
60-70	4	5
70-80	5	5
80-90	5	8
90-100	6	4
Итого:	20	24

Теперь таблицу 27 надо заполнить, для этого смотрим на данные *первой группы*:

- интервал [50-60] - нет студентов, т.е. 0 человек (вносим в таблицу);
- интервал [60-70] – 4 человека (вносим в таблицу);

- интервал [70-90] – надо рассчитывать;
- интервал [90-100] – 6 человек (вносим в таблицу).

Аналогично заполняем известные значения в *группе 2*:

- интервал [50-60] – 2 человека (вносим в таблицу);
- интервал [60-80] – шаг интервала 20%, т.е. надо рассчитать плотность;
- интервал [80-90] – 8 человек (вносим в таблицу);
- интервал [90-100] – 4 человека (вносим в таблицу).

Необходимо рассчитать абсолютную плотность распределения в интервале [70-90] в первой группе: $P = \frac{f}{i} = \frac{10}{20} = 0.5$.

Нам в итоговой таблице (табл. 26) неизвестны два интервала: [70-80] и [80-90], оба они входят в интервал [70-90], плотность которого мы рассчитали. Теперь выразим из формулы частоту повторения, получим: $f = p \cdot i$, следовательно, частота повторения:

- в интервале [70-80] равна: $f = 0.5 \cdot 10 = 5$ (чел.) (учтите, что теперь шаг интервала равен 10 %);
- в интервале [80-90] равна: $f = 0.5 \cdot 10 = 5$ (чел.).

Результаты занесем в таблицу 27.

Аналогично рассчитаем недостающие частоты в группе 2.

Абсолютная плотность распределения в интервале [60-80] (табл. 26) равна: $P = \frac{f}{i} = \frac{10}{20} = 0.5$.

Данный интервал можно разбить на два необходимых нам интервала: [60-80]=[60-70]+[70-80]. Рассчитаем частоту повторения в каждом из этих интервалов: 1) в интервале [60-70] - $f = 0.5 \cdot 10 = 5$ (чел.);

2) в интервале [70-80] - $f = 0.5 \cdot 10 = 5$ (чел.).

Занесем полученные значения частот в таблицу 26.

Теперь данные по двум группам сопоставимы, можно проводить анализ успеваемости в них.

4.6. Показатели формы распределения и графическое изображение вариационного ряда

Первым этапом изучения вариационного ряда является его графическое изображение. Дискретный и интервальный вариационные ряды чаще всего изображаются в виде полигона распределения частот. Графики строятся в прямоугольной системе координат.

Выделяют два основных вида кривых распределения:

1) **многовершинные кривые** (рис. 9) свидетельствуют о ненормальности распределения, следовательно, анализ таких рядов продолжается только после перегруппировки данных;

2) **одновершинные кривые** свидетельствуют о нормальности распределения.

Эталоном ряда распределения является кривая нормального распределения, представленная на рисунке 10.

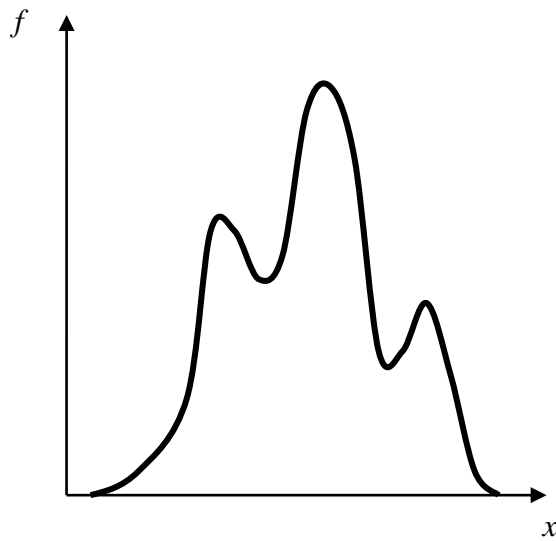


Рис. 9. Многовершинная кривая распределения

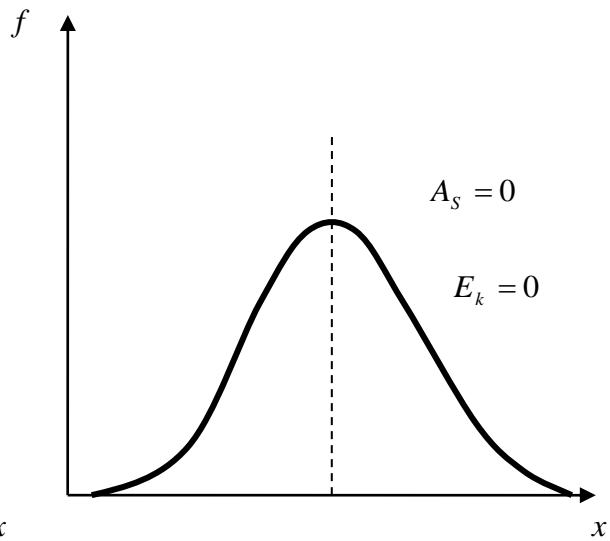


Рис. 10. Кривая нормального распределения

Среди одновершинных кривых распределения выделяют ассиметричные и симметричные кривые.

Ассиметричные кривые распределения в свою очередь делятся на два вида: кривые с правосторонней асимметрией (рис. 11) и кривые с левосторонней асимметрией (рис. 12).

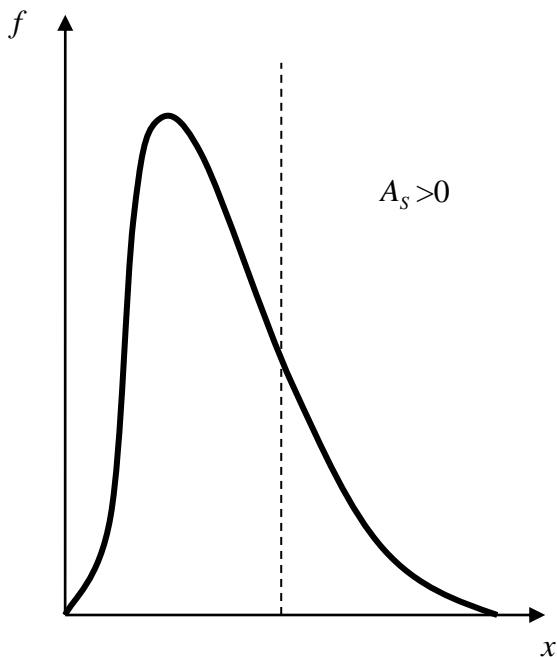


Рис. 11. Правосторонняя асимметрия

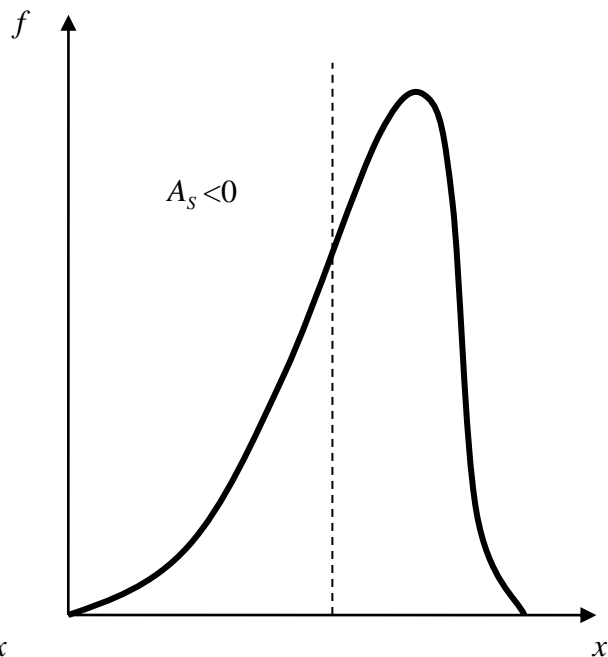


Рис. 12. Левосторонняя асимметрия

Способы определения показателя асимметрии и оценка его существенности

1. Анализ степени асимметрии проводится на основе определения **относительного показателя асимметрии**, предложенного английским

статистиком К. Пирсоном: $A_s = \frac{\bar{x} - Mo}{\sigma_x}$ или $A_s = \frac{\bar{x} - Me}{\sigma_x}$,

где \bar{x} - среднее значение признака;

Mo - модальное значение признака;

Me – медианное значение признака;

σ_x - среднее квадратическое отклонение факторного признака.

Показатель асимметрии может принимать как положительные, так и отрицательные значения. Принято считать, что асимметрия считается значительной, если $|A_s| > 0,5$. Незначительной асимметрия признается, если $|A_s| < 0,25$. Если же значение показателя асимметрии находится в интервале $0,25 \leq |A_s| \leq 0,5$, то асимметрия считается умеренной.

2. Другой способ расчета **показателя асимметрии** предложил шведский математик **Линдберг**: $A_s = \Pi - 50$,

где Π – удельный вес (процент) значений признака, превышающих величину средней арифметической;

50 – константа, которая характеризует процент вариантов, превосходящих среднюю арифметическую ряда нормального распределения.

3. Наиболее часто применяется **показатель асимметрии**, определяемый через **центральный момент третьего порядка**: $A_s = \frac{\mu_3}{\sigma^3}$,

где μ_3 - центральный момент третьего порядка, определяется по формуле:

$$\mu_3 = \frac{\sum (x - \bar{x})^3 \cdot f}{\sum f}.$$

Среднее квадратическое отклонение определяется по формулам:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}; \quad \sigma = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}}.$$

Оценка существенности показателя асимметрии производится на основе **средней квадратической ошибки**: $\sigma_{A_s} = \sqrt{\frac{6 \cdot (n-1)}{(n+1) \cdot (n+3)}}$,

где n – число наблюдений.

Асимметрия является **существенной** и распределение признака-фактора в генеральной совокупности несимметрично, если $\frac{|A_s|}{\sigma_{A_s}} > 3$, если же

дробь меньше 3, то асимметрия является *несущественной*, ее наличие объясняется влиянием случайных величин.

Симметричные кривые распределения встречаются двух видов: островершинные (рис. 13) и плосковершинные (рис. 14).

При анализе симметричных рядов распределения определяется **показатель островершинности (эксцесса)**: $E_k = \frac{\mu_4}{\sigma^4} - 3$,

где μ_4 - центральный момент четвертого порядка: $\mu_4 = \frac{(x - \bar{x})^4 \cdot f}{\sum f}$.

Положительное значение показателя эксцесса свидетельствует об островершинности кривой распределения, отрицательное значение – о плосковершинности кривой.

Для приближенного определения показателя эксцесса используют **формулу Линдберга**: $E_k = \Pi - 38.29$,

где Π – удельный вес (процент) количества вариантов, лежащих в интервале, равном половине среднего квадратического отклонения (в обе стороны от величины средней);

38,29 – процент количества вариантов, лежащих в интервале, равном половине среднего квадратического отклонения, в общем количестве вариантов ряда нормального распределения.

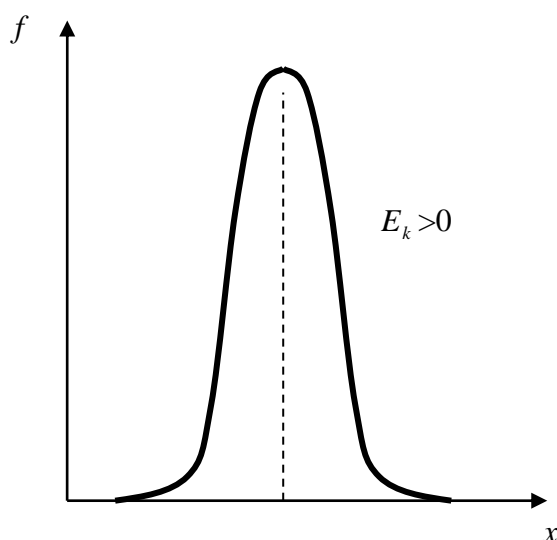


Рис. 13. Островершинное распределение

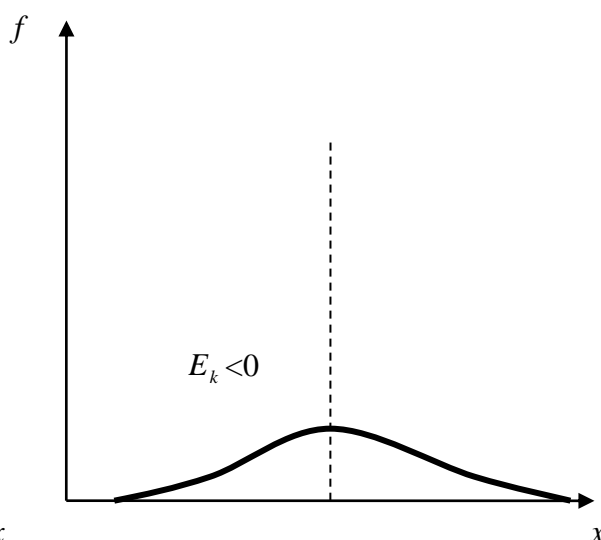


Рис. 14. Плосковершинное распределение

Распределение может считаться нормальным, если показатель эксцесса не превышает **среднеквадратической ошибки эксцесса**, которая определяется по формуле:

$$\sigma_{E_k} = \sqrt{\frac{24 \cdot n \cdot (n-2) \cdot (n-3)}{(n-1)^2 \cdot (n+3) \cdot (n+5)}}.$$

Пример 20. Используя построенный вариационный ряд в примере 17 (табл.24), определите показатели формы распределения, графически представьте ряд распределения.

Таблица 28. Распределение рабочих цеха по уровню квалификации.

Разряд (x)	2	3	4	5	Итого
Количество рабочих, чел. (f)	3	8	13	5	29
Частость (w)	0,1	0,28	0,45	0,17	1,0
Накопленная частота, чел. (S)	3	11	24	29	-

Решение

Определим показатель асимметрии: $A_s = \frac{\mu_3}{\sigma^3}$, $\mu_3 = \frac{\sum (x - \bar{x})^3 \cdot f}{\sum f}$.

Для расчета необходимо найти среднее значение признака:

$$\bar{x} = \frac{\sum x \cdot f}{\sum f} = \frac{2 \cdot 3 + 3 \cdot 8 + 4 \cdot 13 + 5 \cdot 5}{3 + 8 + 13 + 5} = \frac{107}{29} = 3.7.$$

Центральный момент третьего порядка составляет:

$$\mu_3 = \frac{(2 - 3.7)^3 \cdot 3 + (3 - 3.7)^3 \cdot 8 + (4 - 3.7)^3 \cdot 13 + (5 - 3.7)^3 \cdot 5}{3 + 8 + 13 + 5} = \frac{-6.06}{29} = -0.21.$$

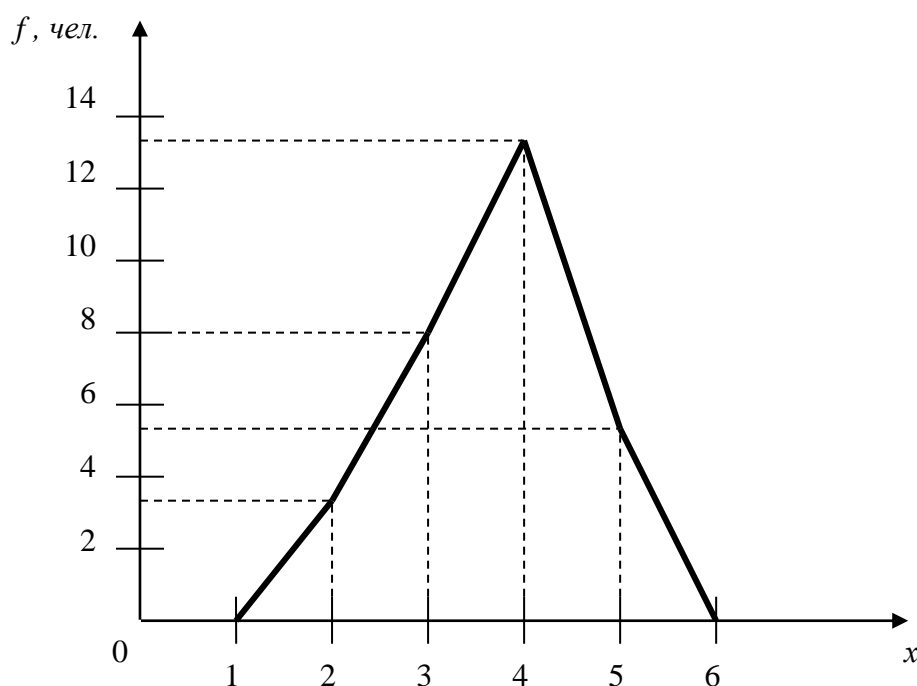


Рис. 15. Линия распределения рабочих цеха по разряду

При определении показателя асимметрии используется среднее квадратическое отклонение, определим его:

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

$$\sigma = \sqrt{\frac{(2-3.7)^2 \cdot 3 + (3-3.7)^2 \cdot 8 + (4-3.7)^2 \cdot 13 + (5-3.7)^2 \cdot 5}{3+8+13+5}} = \sqrt{\frac{22.21}{29}} = 0.88.$$

Значение показателя асимметрии равно: $A_s = \frac{\mu_3}{\sigma^3} = \frac{-0.21}{0.68} = -0.3$.

Полученное значение показателя асимметрии свидетельствует об умеренной левосторонней асимметрии.

Рассчитаем показатель эксцесса по формулам:

$$E_k = \frac{\mu_4}{\sigma^4} - 3, \quad \mu_4 = \frac{\sum (x - \bar{x})^4 \cdot f}{\sum f}.$$

Центральный момент четвертого порядка равен:

$$\mu_4 = \frac{(2-3.7)^4 \cdot 3 + (3-3.7)^4 \cdot 8 + (4-3.7)^4 \cdot 13 + (5-3.7)^4 \cdot 5}{3+8+13+5} = \frac{41.36}{29} = 1.43.$$

$$E_k = \frac{\mu_4}{\sigma^4} - 3 = \frac{1.43}{(0.88)^4} - 3 = 2.39 - 3 = -0.61.$$

Полученная величина показателя эксцесса свидетельствует о плосковершинном распределении.

Графически заданный вариационный ряд представлен на рисунке 15.

Пример 21. Используя данные примера 18 (табл. 25), определите показатели формы распределения, графически представьте заданный ряд.

Решение

Для определения показателей формы распределения, составим вспомогательную таблицу (табл. 29).

Таблица 29. Распределение рабочих по возрасту

Данные		Рассчитанные показатели			
Возраст, лет (x)	Количество рабочих, чел. (f)	$(x' - \bar{x})$	$(x' - \bar{x})^2 \cdot f$	$(x' - \bar{x})^3 \cdot f$	$(x' - \bar{x})^4 \cdot f$
20-30	10	-7,3	532,9	-3890,2	28398,2
30-40	18	2,7	131,2	354,3	956,6
40-50	2	12,7	322,6	4096,8	52028,9
Итого	30	-	986,7	560,9	81383,7

На первом этапе определим средний возраст рабочих:

$$\bar{x} = \frac{\sum x' \cdot f}{\sum f} = \frac{25 \cdot 10 + 35 \cdot 18 + 45 \cdot 2}{30} = 32.3 \text{ (лет)}.$$

Производим расчет элементов формул и заносим их в таблицу 27, учитывая, что x' - середина соответствующего интервала.

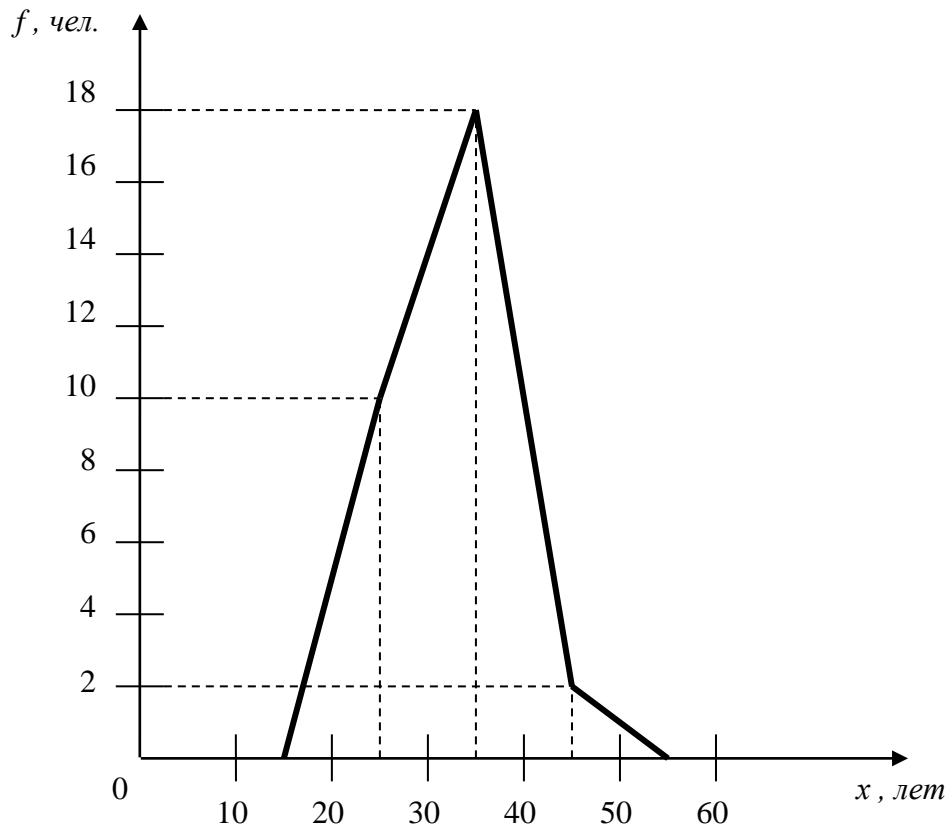


Рис. 16. Линия распределения рабочих по возрасту

Центральный момент третьего порядка составляет:

$$\mu_3 = \frac{\sum (x - \bar{x})^3 \cdot f}{\sum f} = \frac{560.9}{30} = 18.7;$$

среднее квадратическое отклонение равно: $\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}} = \sqrt{\frac{986.7}{30}} = 5.7;$

рассчитаем величину показателя асимметрии: $A_s = \frac{\mu_3}{\sigma^3} = \frac{18.7}{(5.7)^3} = \frac{18.7}{185.2} = 0.1.$

Полученная величина показателя асимметрии свидетельствует о незначительной правосторонней асимметрии ($A_s > 0$ и $A_s < 0.25$).

Так как асимметрия незначительна, то целесообразно рассчитать показатель эксцесса: $\mu_4 = \frac{\sum (x - \bar{x})^4 \cdot f}{\sum f} = \frac{81383.7}{30} = 2712.8;$

$$E_k = \frac{\mu_4}{\sigma^4} - 3 = \frac{2712.8}{(5.7)^4} - 3 = \frac{2712.8}{1055.6} - 3 = 2.57 - 3 = -0.43.$$

Полученная величина показателя эксцесса свидетельствует о плосковершинном распределении ($E_k = -0.43 < 0$).

Графически заданный вариационный ряд представлен на рисунке 16.

При значительной или умеренной асимметрии показатель эксцесса не определяется, так как о симметричности кривой распределения не может

быть и речи, но при выполнении контрольной работы желательно рассчитать оба показателя и сделать вывод по каждому из них.

Тесты для самоконтроля

1. Вариация – это:
 - а) изменение массовых явлений во времени;
 - б) изменение структуры статистической совокупности в пространстве;
 - в) изменение значений признака во времени и в пространстве;
 - г) изменение состава совокупности.
2. Размах вариации – это:
 - а) доля межгрупповой дисперсии в общей дисперсии;
 - б) разность между наибольшим и наименьшим значениями варьирующего признака;
 - в) графическое изображение линии изменения частот.
3. С помощью коэффициента вариации характеризуется:
 - а) размах вариации; б) однородность совокупности;
 - в) интенсивность вариации; г) нормальность распределения.
4. Графическое представление вариационного ряда осуществляется с помощью: а) кумуляты; б) секторной диаграммы; в) полигона частот.
5. Если значения факторного признака представлены дискретными величинами, то строится ряд распределения:
 - а) интервальный; б) дискретный; в) атрибутивный.
6. Для построения интервального ряда распределения необходимо определить: а) число групп; б) коэффициент вариации; в) среднее квадратическое отклонение; г) шаг интервала.
7. Частоты, выраженные в относительных единицах - это:
 - а) накопленная частота; б) частость; в) частота повторения.
8. К показателям центра распределения относятся:
 - а) медиана; б) дисперсия; в) коэффициент осцилляции;
 - г) среднее значение признака; д) мода.
9. Для приведения рядов распределения к сопоставимому виду рассчитывают: а) среднее квадратическое отклонение; б) относительное линейное отклонение; в) плотность распределения; г) дисперсию.
10. При анализе симметрии ряда распределения определяют:
 - а) показатель эксцесса; б) частость; в) показатель асимметрии;
 - г) среднюю квадратическую ошибку показателя асимметрии.
11. Оценка существенности показателя асимметрии определяется с помощью: а) среднего квадратического отклонения; б) средней квадратической ошибки эксцесса; в) средней квадратической ошибки A_s .

Правильные ответы: 1. в; 2. б; 3. б; 4. в; 5. б; 6. а, г; 7. б; 8. а, д; 9. в; 10. а; 11. в.

5. Выборочное наблюдение

Одним из наиболее эффективным способом несплошного наблюдения является *выборочное наблюдение*.

Преимущества:

- экономия труда;
- быстрота проведения;
- экономия средств на получение и обработку информации.

Недостатком выборочного наблюдения является наличие ошибки. Здесь встречаются все виды ошибок сплошного наблюдения, а также возможно появление *ошибки репрезентативности*, которые свойственны только несплошным наблюдениям. Они возникают вследствие представления в выборочную совокупность единиц наблюдения, не отражающих действительность (например, если взять из группы студентов только отличников, то будет сделан вывод об очень высоком уровне знаний всей группы).

5.1. Виды выборочного наблюдения и формы отбора единиц в выборочную совокупность

Все *виды выборочного наблюдения* классифицируются по двум основным признакам:

- 1) *по степени охвата единиц* выделяют большие и малые выборки;
- 2) *по способу формирования* выделяют:

- простую случайную выборку;
- механическую выборку;
- типическую (расслоенную, районированную) выборку;
- серийную выборку;
- ступенчатую выборку;
- комбинированную выборку;
- многофазную выборку;
- метод моментных наблюдений.

Наиболее часто используются простая случайная и механическая виды выборок. Именно эти виды рассмотрены в данном пособии, в контрольную работу № 1 также вошли задачи только по этим видам выборок.

Формы отбора единиц в выборочную совокупность:

- *повторный отбор единиц* (единица, один раз включенная в выборку из общей массы единиц генеральной совокупности, может быть в нее включена дважды, трижды и т.д.);
- *бесповторный отбор единиц* (единица, один раз включенная в выборку, уже не может быть отобрана снова).

Общепринятые обозначения:

- N* - объем генеральной совокупности;
- n* - объем выборочной совокупности;

- \bar{x} - среднее значение признака в генеральной совокупности;
 \tilde{x} - среднее значение признака в выборочной совокупности;
 p - доля единиц, обладающих определенным признаком в генеральной совокупности;
 ω - доля единиц, обладающих определенным признаком в выборочной совокупности;
 σ^2 - дисперсия признака в генеральной совокупности;
 S^2 - дисперсия признака в выборочной совокупности;
 σ - среднее квадратическое отклонение признака в генеральной совокупности;
 S - среднее квадратическое отклонение в выборочной совокупности.

5.2. Задачи, которые можно решать с помощью предельной ошибки выборки

При проведении анализа выборочной совокупности всегда определяется **предельная ошибка выборки**, которая показывает величину отклонения выборочной средней (доли) от генеральной, вероятность превышения которой вследствие случайных причин очень мала.

При любом способе проведения выборочного наблюдения **предельная ошибка** определяется по формуле:

$$\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}} \text{ или } \Delta_p = t \cdot \mu_p,$$

где $\Delta_{\bar{x}}$ - предельная ошибка генеральной средней;

Δ_p - предельная ошибка генеральной доли;

t - коэффициент кратности;

$\mu_{\bar{x}}$ - средняя ошибка генеральной средней;

μ_p - средняя ошибка генеральной доли.

Средние ошибки ($\mu_{\bar{x}}, \mu_p$) для каждого вида выборочного наблюдения определяются по своим формулам (см. далее).

Коэффициент кратности (t) определяется с помощью таблицы значений «Удвоенной нормированной функции Лапласа» (Приложение 1) на основе доверительной вероятности.

Доверительная вероятность определяет точность расчетов:

$$P = 1 - \alpha,$$

например, если $P = 0.985$, то заданная точность расчетов составляет 98,5 %, а допустимая ошибка равна $\alpha = 1 - P = 0.015$ (1,5 %).

С помощью предельной ошибки выборки решаются **три типа задач**.

1. *Определение пределов генеральных характеристик* с заданной доверительной вероятностью:

- доверительные интервалы для генеральной средней:

$$\bar{x} = \tilde{x} \pm \Delta_{\bar{x}};$$

$$\tilde{x} - \Delta_{\bar{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\bar{x}};$$

- доверительные интервалы для генеральной доли:

$$p = \omega \pm \Delta_p;$$

$$\omega - \Delta_p \leq p \leq \omega + \Delta_p.$$

2. *Определение доверительной вероятности* того, что генеральная характеристика отличается от выборочной не более чем на определенную заданную величину предельной ошибки:

- доверительная вероятность является функцией от t , следовательно, определив коэффициент кратности: $t = \frac{\Lambda_{\bar{x}}}{\mu_{\bar{x}}}$, можно найти значение

доверительной вероятности по таблице значений «Удвоенной нормированной функции Лапласа».

3. *Определение необходимой численности выборки*, при заданном значении предельной ошибки выборки. Формулы определения численности выборки представлены в таблице 30.

Таблица 30. **Определение необходимой численности выборки**

Численность выборки	Способ отбора единиц	
	повторный	бесповторный
1. Для средней	$n = \frac{t^2 \cdot S^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 \cdot N \cdot S^2}{\Delta_{\bar{x}}^2 \cdot N + t^2 \cdot S^2}$
2. Для доли	$n = \frac{t^2 \cdot \omega(1-\omega)}{\Delta_p^2}$	$n = \frac{t^2 \cdot N \cdot \omega(1-\omega)}{\Delta_p^2 \cdot N + t^2 \cdot \omega(1-\omega)}$

5.3. Простая случайная и механическая выборки

При проведении *простой случайной выборки* отбор единиц в выборочную совокупность производится непосредственно из всей массы единиц генеральной совокупности в форме случайного (повторного или бесповторного) отбора. Формулы определения ошибок выборки представлены в таблице 31.

Отбор единиц из всей массы единиц генеральной совокупности в *механическую выборку* производится через равные промежутки из определенного расположения единиц в генеральной совокупности: по алфавиту, в пространстве, во времени. Например, в выборку включается каждая пятая единица генеральной совокупности. Таким образом, отбор единиц в механическую выборку осуществляется только в форме *бесповторного* отбора.

При формировании механической выборки необходимо:

- *определить шаг отчета* (γ) - расстояние между отбираемыми единицами: $\gamma = \frac{N}{n}$;

- *начало отсчета*, т.е. единицу, с которой начинается отбор единиц в выборку.

Таблица 31. Формулы расчета средней ошибки простой случайной выборки

Средняя ошибка	Способ отбора единиц	
	повторный	бесповторный
- для средней	$\mu_{\bar{x}} = \sqrt{\frac{S^2}{n}}$	$\mu_{\bar{x}} = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$
- для доли	$\mu_p = \sqrt{\frac{\omega(1-\omega)}{n}}$	$\mu_p = \sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}$

Определение ошибок механической выборки производится по формулам простой случайной выборки при бесповторном отборе (табл. 31).

Пример 22. При анализе возраста рабочих на предприятии была проведена 25%-ная простая случайная бесповторная выборка.

Таблица 32. Результаты выборочного обследования

Возраст, лет	20-26	26-32	32-38	38-44
Число рабочих, чел.	20	48	20	12

Определите:

1) доверительные интервалы среднего возраста рабочих для предприятия в целом, гарантируя результат с вероятностью 0,978;

2) доверительные интервалы доли рабочих старше 38 лет по предприятию в целом, гарантируя результат с вероятностью 0,984;

3) необходимую численность выборки при определении среднего возраста рабочих предприятия так, чтобы предельная ошибка не превышала 0,9 лет (вероятность – 0,968);

4) необходимую численность выборки при определении доли рабочих старше 38 лет по предприятию в целом так, чтобы предельная ошибка не превышала 4 % (вероятность – 0,954).

Решение

1. Доверительный интервал среднего возраста рабочих для предприятия в целом определяется выражением: $\tilde{x} - \Delta_{\bar{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\bar{x}}$.

Средний возраст рабочих по выборке составляет:

$$\tilde{x} = \frac{\sum x' \cdot f}{\sum f}; \quad \tilde{x} = \frac{23 \cdot 20 + 29 \cdot 48 + 35 \cdot 20 + 41 \cdot 12}{20 + 48 + 20 + 12} = \frac{3044}{100} = 30.44 \text{ (лет)}.$$

Предельная ошибка для генеральной средней равна: $\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}}$.

Для ее определения необходимо найти среднюю ошибку генеральной средней (табл. 31): $\mu_{\bar{x}} = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$,

где S^2 - дисперсия выборочной совокупности, для интервального ряда определяется по формуле: $S^2 = \frac{\sum (x' - \tilde{x})^2 \cdot f}{\sum f}$;

$$S^2 = \frac{(23 - 30.44)^2 \cdot 20 + (29 - 30.44)^2 \cdot 48 + (35 - 30.44)^2 \cdot 20 + (41 - 30.44)^2 \cdot 12}{20 + 48 + 20 + 12} = 29.61$$

Следовательно, дисперсия составляет 29,61 лет.

Теперь определим среднюю ошибку, учитывая, что $\frac{n}{N} = 0.25$, так как произведена по условию задачи 25%-ная выборка, т.е. доля единиц выборки в общем числе единиц генеральной совокупности равна 0,25:

$$\mu_{\bar{x}} = \sqrt{\frac{29.6}{100} (1 - 0.25)} = 0.47 \text{ (лет)}.$$

По таблице «Удвоенной нормированной функции Лапласа» (Приложение 1) определим t : в первом задании $P = 0,978$, внутри таблицы ищем цифры после запятой: 9780, затем выходим по сроке на значение t (2,2), поднявшись вверх от цифры 9780 получим 9 – это сотые.

Следовательно, $t = 2.29$.

Тогда предельная ошибка равна: $\Delta_{\bar{x}} = 2.29 \cdot 0.47 = 1.076$.

Составим доверительный интервал среднего возраста рабочих:

$$30.44 - 1.08 \leq \bar{x} \leq 30.44 + 1.08 \\ 29.36 \leq \bar{x} \leq 31.52.$$

Вывод: средний возраст рабочих предприятия находится в интервале от 29,36 до 31,52 года.

2. Доверительные интервалы доли рабочих в возрасте старше 38 лет на предприятии в целом, определяются с помощью выражения:

$$\omega - \Delta_p \leq p \leq \omega + \Delta_p.$$

Доля рабочих в возрасте старше 38 лет по выборке составляет:

$$\omega = \frac{n_i}{n}; \quad \omega = \frac{12}{100} = 0.12.$$

Для определения предельной ошибки генеральной доли: $\Delta_p = t \cdot \mu_p$, необходимо найти

- 1) коэффициент кратности при $P = 0.984$ (Приложение 1): $t = 2.41$;
- 2) среднюю ошибку доли для бесповторной случайной выборки:

$$\mu_p = \sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}; \quad \mu_p = \sqrt{\frac{0.12 \cdot (1 - 0.12)}{100} (1 - 0.25)} = 0.0281.$$

Тогда $\Delta_p = 2.41 \cdot 0.0281 = 0.0677$, следовательно, интервал равен:

$$0.12 - 0.0677 \leq p \leq 0.12 + 0.0677;$$

$$0.0523 \leq p \leq 0.1877;$$

$$5,23\% \leq p \leq 18,77\% .$$

Вывод: удельный вес рабочих в возрасте старше 38 лет на предприятии в целом составляет от 5,23% до 18,77 %.

3. Необходимая численность выборки для генеральной средней при проведении случайной бесповторной выборки определяется по формуле:

$$n = \frac{t^2 \cdot N \cdot S^2}{\Delta_{\bar{x}}^2 \cdot N + t^2 \cdot S^2} .$$

При доверительной вероятности 0,968 - $t = 2.15$ (Приложение 1), дисперсия выборочной совокупности определена в первом задании: $S^2 = 29.6$ лет, предельная ошибка задана в третьем задании: $\Delta_{\bar{x}} = 0.9$ лет, численность единиц генеральной совокупности можно рассчитать с помощью математической пропорции:

$$\frac{25\% - 100\text{чел}}{100\% - N} ; N = \frac{100}{25} \cdot 100 = 400 \text{ (чел.)}.$$

$$\text{Таким образом } n = \frac{(2.15)^2 \cdot 400 \cdot 29.6}{(0.9)^2 \cdot 400 + (2.15)^2 \cdot 29.6} = 118.8 \approx 119.0 \text{ (чел.)}.$$

Предельная ошибка выборки не будет превышать 0,9 человек, если численность выборочной совокупности составит 119 человек.

4. Необходимая численность выборки для генеральной доли случайной бесповторной выборки определяется по формуле:

$$n = \frac{t^2 \cdot N \cdot \omega(1-\omega)}{\Delta_p^2 \cdot N + t^2 \cdot \omega(1-\omega)} .$$

В последнем задании задана $\Delta_p = 0.04$; во втором задании была рассчитана доля рабочих старше 38 лет по выборке: $\omega = 0.12$, общая численность рабочих предприятия составляет 400 человек; коэффициент кратности при $P = 0.954$ равен: $t = 2.0$ (Приложение 1).

Тогда необходимая численность выборки составляет:

$$n = \frac{(2)^2 \cdot 400 \cdot 0.12 \cdot (1-0.12)}{(0.04)^2 \cdot 400 + (2)^2 \cdot 0.12 \cdot (1-0.12)} = 159.5 \approx 160.0 \text{ (чел.)}.$$

Вывод: чтобы предельная ошибка доли рабочих старше 38 лет не превышала 4%, в выборочную совокупность необходимо включить 160 человек.

Пример 23. При проведении 15%-ной простой случайной повторной выборки был проанализирован доход рабочих предприятия (табл. 33).

Определите:

1) доверительные интервалы среднего уровня заработной платы рабочих для предприятия в целом (доверительная вероятность 0,976);

2) необходимую численность выборки при определении средней заработной платы рабочих предприятия, чтобы предельная ошибка не превышала 135 рублей (вероятность – 0,946);

3) долю рабочих предприятия, доход которых не превышает 6900 рублей, гарантируя результат с вероятностью 0,964;

4) необходимую численность выборки при определении доли рабочих предприятия с доходом менее 6900 рублей, чтобы предельная ошибка не превышала 7 % (вероятность – 0,954).

Таблица 33. Распределение рабочих предприятия по уровню дохода

Средняя заработная плата за месяц, руб.	6300-6900	6900-7500	7500-8100	8100-8700
Число рабочих, чел.	18	30	26	6

Решение

1. Доверительный интервал среднего уровня заработной платы рабочих для предприятия в целом определяется: $\bar{x} - \Delta_{\bar{x}} \leq \bar{x} \leq \bar{x} + \Delta_{\bar{x}}$.

Средняя месячная заработная плата рабочих по выборке составляет:

$$\bar{x} = \frac{\sum x' \cdot f}{\sum f}; \quad \bar{x} = \frac{6600 \cdot 18 + 7200 \cdot 30 + 7800 \cdot 26 + 8400 \cdot 6}{18 + 30 + 26 + 6} = \frac{588000}{80} = 7350,0 \text{ (руб.)}.$$

Предельная ошибка для генеральной средней равна: $\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}}$.

Определим среднюю ошибку генеральной средней, учитывая, что проведен повторный отбор (табл. 31):

$$\mu_{\bar{x}} = \sqrt{\frac{S^2}{n}}; \quad S^2 = \frac{\sum (x' - \bar{x})^2 \cdot f}{\sum f};$$

$$S^2 = \frac{(6600 - 7350)^2 \cdot 18 + (7200 - 7350)^2 \cdot 30 + (7800 - 7350)^2 \cdot 26 + (8400 - 7350)^2 \cdot 6}{18 + 30 + 26 + 6} = 2835000$$

Подставим полученное значение дисперсии: $S^2 = 2835000$ руб.:

$$\mu_{\bar{x}} = \sqrt{\frac{2835000}{80}} = 59,53 \text{ (руб.)}.$$

По таблице «Удвоенной нормированной функции Лапласа» определим коэффициент кратности: $t = 2,25$ при $P = 0,976$.

Тогда предельная ошибка равна: $\Delta_{\bar{x}} = 59,53 \cdot 2,25 = 133,9$ (руб.).

Составим доверительный интервал средней заработной платы за месяц рабочих предприятия: $7350 - 133,9 \leq \bar{x} \leq 7350 + 133,9$
 $7216,1 \leq \bar{x} \leq 7483,9$.

Вывод: средний уровень месячной заработной платы рабочих предприятия находится в интервале от 7216,1 до 7483,9 рубля.

2. Необходимая численность выборки для генеральной средней при проведении случайной повторной выборки определяется по формуле:

$$n = \frac{t^2 \cdot S^2}{\Delta_{\bar{x}}^2}.$$

Известные данные: $S^2 = 2835000$ руб.; $t = 1.93$ (Приложение 1); $\Delta_{\bar{x}} = 135.0$ руб., подставим их в формулу, получим:

$$n = \frac{(1.93)^2 \cdot 2835000}{(135)^2} = 57.9 \approx 58.0 (\text{чел.}).$$

Предельная ошибка выборки не будет превышать 135 рублей, если численность выборочной совокупности составит 58 человек.

3. Доверительный интервал доли рабочих с доходом, не превышающем 6900 рублей, на предприятии в целом составляет:

$$\omega - \Delta_p \leq p \leq \omega + \Delta_p.$$

Доля рабочих, уровень заработной платы меньше 6900 рублей по выборке составляет: $\omega = \frac{n_i}{n}$; $\omega = \frac{18}{80} = 0.225$.

Для определения предельной ошибки генеральной доли: $\Delta_p = t \cdot \mu_p$, необходимо найти

- 1) коэффициент кратности при $P = 0.964$ (Приложение 1): $t = 2.1$;
- 2) среднюю ошибку доли для повторной случайной выборки:

$$\mu_p = \sqrt{\frac{\omega(1-\omega)}{n}}; \mu_p = \sqrt{\frac{0.225 \cdot (1-0.225)}{80}} = 0.0469.$$

Тогда $\Delta_p = 2.1 \cdot 0.0469 = 0.0985$, следовательно, интервал равен:

$$0.225 - 0.0985 \leq p \leq 0.225 + 0.0985;$$

$$0.1265 \leq p \leq 0.3235;$$

$$12.65\% \leq p \leq 32.35\%.$$

Вывод: удельный вес рабочих с доходом менее 6900 на предприятии в целом находится в интервале от 12,65% до 32,35 %.

4. Необходимая численность выборки для генеральной доли случайной повторной выборки определяется по формуле: $n = \frac{t^2 \cdot \omega(1-\omega)}{\Delta_p^2}$

Нам известны: $\Delta_p = 0.07$; $\omega = 0.225$; $t = 2.0$ при $P = 0.954$. Тогда необходимая численность выборки составляет:

$$n = \frac{(2)^2 \cdot 0.225 \cdot (1-0.225)}{(0.07)^2} = 142.35 \approx 143.0 (\text{чел.}).$$

Вывод: чтобы предельная ошибка доли рабочих с доходом менее 6900 рублей не превышала 7%, в выборочную совокупность необходимо включить 143 человека.

Пример 23. Определите доверительную вероятность для повторного и бесповторного отбора, если известно, что при проведении 20%-ной простой случайной выборки получены следующие результаты:

$$n = 100; S^2 = 3.0; \Delta_{\bar{x}} = 0.3.$$

Решение

Так как задана предельная ошибка средней, следовательно, анализируется генеральная средняя величина.

Доверительная вероятность является функцией от t , т.е. рассчитав коэффициент кратности, найдем значение доверительной вероятности. Выразим из формулы предельной ошибки коэффициент кратности:

$$\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}} \Rightarrow t = \frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}}.$$

$$\text{Для повторного отбора: } \mu_{\bar{x}} = \sqrt{\frac{S^2}{n}}; \mu_{\bar{x}} = \sqrt{\frac{3}{100}} = 0.17; t = \frac{0.3}{0.17} = 1.77.$$

Теперь, используя «Удвоенную нормированную функцию Лапласа», найдем значение доверительной вероятности: $P = 0.9233$.

Для бесповторного отбора средняя ошибка средней определяется по формуле: $\mu_{\bar{x}} = \sqrt{\frac{S^2}{n} (1 - \frac{n}{N})}$; $\mu_{\bar{x}} = \sqrt{\frac{3.0}{100} (1 - 0.2)} = 0.155$.

Тогда коэффициент кратности равен: $t = \frac{0.3}{0.155} = 1.94$, следовательно, доверительная вероятность соответствует величине: $P = 0.9476$.

Тесты для самоконтроля

1. К достоинствам выборочного наблюдения относятся:
 - а) наличие ошибки наблюдения; б) экономия времени на получение и обработку информации; в) простота проведения.
2. Если единица отобрана в выборочную совокупность из генеральной не один раз, а несколько, то осуществлен отбор:
 - а) бесповторный; б) повторный; в) случайный; г) механический.
3. Точность расчетов отражает: а) среднее значение признака в выборке; б) доверительная вероятность; в) коэффициент кратности.
4. Коэффициент кратности определяется по значению доверительной вероятности с помощью таблицы:
 - а) «Распределение Стьюдента (t распределение)»;
 - б) «Удвоенной нормированной функции Лапласа»;
 - в) «Распределение Пирсона (χ^2 - распределение)».
5. На основе известного значения предельной ошибки можно определить:
 - а) структурные средние величины; б) доверительную вероятность;
 - в) доверительные интервалы генеральных характеристик; г) дисперсию.
6. Механическая выборка проводится в форме:
 - а) повторного отбора единиц; б) случайного отбора единиц;
 - в) бесповторного отбора единиц.

Правильные ответы: 1. б, в; 2. б; 3. в; 4. б; 5. б, в; 6. в.

Задания к контрольной работе № 1

Вариант 1

Задача 1. Используя данные таблицы 34, проведите группировку и сводку материала, выделив 5 групп. За факторный признак примите «Среднесписочное число работников». Сформулируйте выводы.

Таблица 34. Данные о деятельности продовольственных магазинов за отчетный период

№	Товаро-оборот, млн. руб.	Среднесписочное число работников, чел.	№	Товаро-оборот, млн. руб.	Среднесписочное число работников, чел.
1	88	30	11	100	29
2	130	35	12	109	28
3	113	31	13	57	21
4	99	27	14	100	27
5	101	30	15	101	30
6	100	34	16	115	31
7	102	30	17	170	40
8	130	33	18	50	20
9	125	36	19	137	29
10	147	32	20	75	23

Задача 2. Определите среднюю заработную плату за январь по каждой профессии в отдельности и по предприятию в целом.

Таблица 35. Данные о заработной плате рабочих за январь

Профессия	Число рабочих	Заработная плата рабочего за январь, руб.
Токари	4	1152; 1512; 1600; 1400
Слесари	6	1450; 1380; 1260; 1700; 1250; 1372

Задача 3. Определите моду и медиану графическими способами.

Таблица 36. Данные о разрядах рабочих цеха

Разряд	2	3	4	5	6	7
Число рабочих, чел.	2	7	10	8	3	2

Задача 4. Известны данные о возрастном составе рабочих цеха (лет): 18; 39; 28; 29; 26; 38; 34; 22; 28; 30; 22; 23; 35; 33; 27; 24; 30; 32; 28; 25; 29; 26; 31; 24; 29; 27; 32; 25; 29; 29.

Для анализа распределения рабочих цеха по возрасту:

- 1) постройте интервальный ряд распределения, выделив семь групп;
- 2) определите показатели формы распределения.

Задача 5. Проведено выборочное обследование жилищных условий жителей города, осуществленного на основе простой случайной повторной

выборки (табл. 37). С вероятностью 0,964 определите доверительные интервалы средней полезной площади для жителей города в целом.

Таблица 37. Распределение жителей города по жилищным условиям

Общая полезная площадь, приходящаяся на 1 человека, м ²	10-15	15-20	20-25	25-30	30-35
Число жителей	204	270	210	130	83

Вариант 2

Задача 1. По данным таблицы 38 проведите аналитическую группировку и сводку материала, выделив 5 групп. За факторный признак примите «Число сотрудников». Сделайте выводы.

Таблица 38. Данные по двадцати коммерческим банкам региона за год

№ банка	Число сотрудников, чел.	Прибыль, млн. руб.	№ банка	Число сотрудников, чел.	Прибыль, млрд. руб.
1	102	47,6	11	303	95,1
2	300	92,8	12	50	31,5
3	220	79,5	13	70	46,5
4	27	21,0	14	177	75,9
5	374	99,3	15	265	93,4
6	43	32,3	16	14	14,9
7	192	68,7	17	162	69,9
8	32	26,8	18	121	54,0
9	20	20,2	19	136	60,0
10	210	81,1	20	157	66,5

Задача 2. Используя данные таблицы 39, определите среднюю месячную заработную плату рабочих каждой специальности в отдельности, а также среднюю заработную плату рабочих цеха в целом.

Таблица 39. Данные о заработной плате рабочих цеха за август

Профессия	Количество рабочих	Заработная плата рабочего, руб.
Токари	5	2650; 2600; 3000; 3200; 3600
Слесари	3	1550; 1630; 1590

Задача 3. Известны результаты сдачи студентами экзамена: «отлично» получили 10 человек; «хорошо» - 16; «удовлетворительно» - 8. Определите долю студентов, сдавших экзамен на «отлично» в их общем числе, а также долю троечников в числе хорошистов и отличников.

Задача 4. Известны данные о часовой интенсивности движения автомобилей на автомагистрали (авт/ч): 140; 99; 80; 50; 110; 130; 90; 210; 220; 140; 96; 261; 218; 48; 282; 340; 36; 312; 92; 60; 68; 152; 30; 80; 120; 130; 86; 131; 102; 190. Задание:

- 1) постройте интервальный ряд распределения, выделив 5 групп;

2) вычислите среднее линейное отклонение, дисперсию и среднее квадратическое отклонение.

Задача 5. На заводе с целью изучения производительности труда рабочих было проведено 20%-ное выборочное обследование 100 рабочих методом случайного бесповторного отбора.

Таблица 40. Данные о затратах времени на обработку одной детали

Время обработки одной детали, мин.	18-20	20-22	22-24	24-26	26-28	28-30
Число рабочих	2	8	24	50	12	4

Используя данные таблицы 40, определите:

1) доверительные интервалы среднего времени обработки одной детали рабочих завода в целом, гарантируя результат с вероятностью 0,997;

2) доверительные интервалы доли рабочих, затрачивающих на обработку одной детали 26 минут и более, в общей числе рабочих завода, гарантируя результат с вероятностью 0,958.

Вариант 3

Задача 1. Используя данные таблицы 41, проведите группировку и сводку материала, выделив 5 групп. За факторный признак примите «Число работников в смену». Сделайте вывод.

Таблица 41. Данные о работе предприятий легкой промышленности

№ п/п	Число работников в смену, чел.	Прибыль, млн. руб.	№ п/п	Число работников в смену, чел.	Прибыль, млн. руб.
1	35	16,2	11	31	15,3
2	54	33,9	12	57	35,0
3	30	15,5	13	40	20,4
4	42	21,2	14	61	40,6
5	69	46,4	15	43	27,3
6	74	52,1	16	47	26,6
7	34	16,2	17	65	44,0
8	33	16,0	18	36	16,7
9	55	34,2	19	80	57,0
10	44	26,8	20	48	25,8

Задача 2. Проведите простую группировку данных о полученных студентами оценках на экзамене: 4; 5; 4; 3; 3; 4; 4; 5; 3; 4; 4; 4; 3; 5; 4; 5; 4; 4; 4; 3; 2; 5; 4; 5; 4. Определите средний балл на экзамене.

Задача 3. Определите долю рабочих в возрасте старше 32 лет.

Таблица 42. Данные о возрасте рабочих цеха

Возраст, лет	20 - 24	24 - 28	28 - 32	32 - 36	36 - 40
Число рабочих, чел.	8	20	12	7	3

Задача 4. Имеются данные о возрастном составе рабочих цеха (лет): 22; 40; 18; 28; 36; 19; 22; 25; 22; 26; 40; 18; 32; 33; 45; 22; 35; 20; 44; 38; 34; 35; 33; 22; 25; 18; 35; 36; 39; 50; 21; 22; 27; 19.

Для анализа распределения рабочих цеха по возрасту необходимо:

- 1) построить интервальный ряд распределения;
- 2) дать графическое изображение ряда (полигон распределения);
- 3) определить показатели центра распределения.

Задача 5. Проведена 10%-ая простая случайная выборка (повторный отбор). Из 50 рабочих, попавших в выборку, 8 человек не выполняли норму выработки, а 5 человек – перевыполняли ее. Определите:

- 1) долю рабочих предприятия не выполняющих норму выработки, гарантируя результат с вероятностью 0,966;
- 2) необходимую численность выборки при определении доли рабочих предприятия, не выполняющих норму выработки, чтобы предельная ошибка не превышала 0,08 (вероятность – 0,968);

Вариант 4

Задача 1. По результатам работы 20 предприятий (табл. 43) проведите группировку и сводку материала, выделите 5 групп.

Таблица 43. Результаты работы предприятий

№	Производительность труда, ед./ч	Прибыль, млн. руб.	№	Производительность труда, ед./ч	Прибыль, млн. руб.
1	25	76,5	11	28	88,9
2	10	21,8	12	27	85,0
3	17	42,6	13	15	31,2
4	20	60,0	14	19	51,8
5	30	90,6	15	15	34,8
6	27	84,3	16	27	88,8
7	14	30,5	17	24	74,6
8	19	50,4	18	14	22,3
9	15	35,0	19	16	42,0
10	28	89,0	20	18	48,5

Задача 2. Определите средний возраст рабочих цеха.

Таблица 44. Распределение рабочих цеха по возрасту

Группы рабочих по возрасту, лет	18-22	22-26	26-30	30-34	34-38
Число рабочих, чел.	5	3	7	4	2

Задача 3. Используя данные таблицы 45, определите величину планового задания на оба вида продукции, величину выполнения плана в каждом периоде, а также удельный вес продукта B в общем объеме выпуска.

Таблица 45. Данные об объемах выпуска продукции предприятия

Вид продукции	Объем выпуска, тыс. ед.			
	Первый год		Второй год	
	По плану	Фактически	По плану	Фактически
А	60,5	58,4	65,6	66,0
В	85,0	86,0	79,4	79,2

Задача 4. Используя данные таблицы 46, проведите проверку совокупности на однородность. Определите среднее квадратическое отклонение, дисперсию и среднее линейное отклонение.

Таблица 46. Данные о работе предприятий легкой промышленности

№ п/п	Число работников в смену, чел.	№ п/п	Число работников в смену, чел.
1	35	6	38
2	40	7	34
3	30	8	33
4	42	9	36
5	37	10	31

Задача 5. Было проведено 25%-е выборочное обследование в форме простой случайной повторной выборки (табл. 47). Определите:

- 1) доверительные интервалы средних затрат времени на обработку одной детали для предприятия в целом, гарантируя вероятность 0,986;
- 2) необходимую численность выборки при определении средних затрат времени на обработку одной детали, чтобы предельная ошибка средней не превышала 0,4 минуты (вероятность– 0,956);

Таблица 47. Данные о затратах времени обработки детали

Затраты времени на обработку одной детали, мин.	1,5–2,5	2,5–3,5	3,5–4,5	4,5–5,5
Количество станков, ед.	7	10	15	8

Вариант 5

Задача 1. По результатам работы двадцати предприятий (табл. 48) проведите группировку и сводку материала, выделив 5 групп.

Таблица 48. Результаты работы предприятий

№	Средний стаж работы рабочих, лет	Прибыль, млн. руб.	№	Средний стаж работы рабочих, лет	Прибыль, млн. руб.
1	15,4	76,5	11	18,6	88,9
2	7,0	21,8	12	16,7	85,0
3	10,0	42,6	13	8,4	31,2
4	13,2	60,0	14	12,0	51,8
5	22,0	90,6	15	8,2	34,8
6	18,5	84,3	16	21,2	88,8
7	8,6	30,5	17	14,0	74,6
8	12,3	50,4	18	7,5	22,3
9	9,3	35,0	19	10,5	42,0
10	22,0	89,0	20	11,0	48,5

Задача 2. Определите среднемесячный темп роста объема грузовых перевозок, используя формулу средней геометрической.

Таблица 49. Перевозка грузов по автотранспортному предприятию

Показатель	Январь	Февраль	Март
Перевезено грузов, тыс.т	45,0	46,7	50,1

Задача 3. Используя данные таблицы 50, определите

- 1) долю женщин и долю мужчин в общем числе сотрудников;
- 2) долю женщин с высшим образованием в общем числе женщин;
- 3) долю мужчин с высшим образованием в общем числе мужчин;
- 4) общую долю сотрудников с незаконченным высшим образованием в общем числе сотрудников организации.

Таблица 50. Данные о составе сотрудников организации

Сотрудники	Образование	
	Незаконченное высшее	Высшее
Женщины	8	32
Мужчины	13	57

Задача 4. Даны тарифные разряды тридцати рабочих цеха: 2; 4; 3; 2; 6; 4; 4; 2; 3; 5; 5; 4; 4; 5; 4; 2; 3; 4; 4; 5; 2; 3; 4; 6; 5; 4; 2; 4; 3; 4; 3; 4; 6; 3; 4; 4; 3; 2; 4. Постройте дискретный ряд распределения, представить ряд графически, определите показатели формы распределения, а также определите их значимость.

Задача 5. При проведении простой случайной 25%-й выборки методом бесповторного отбора, были получены следующие данные: $n = 150$; $\bar{x} = 7,0$; $S = 3,0$.

Какова вероятность того, что предельная ошибка генеральной средней не превысит $\Delta_{\bar{x}} = 0,6$?

Вариант 6

Задача 1. По результатам работы двадцати предприятий (табл. 51) проведите группировку и сводку материала, выделив 5 групп.

Таблица 51. **Результаты работы предприятий**

№	Коэффициент износа оборудования, %	Прибыль, млн. руб.	№	Коэффициент износа оборудования, %	Прибыль, млн. руб.
1	13,6	76,5	11	10,8	88,9
2	47,9	21,8	12	12,0	85,0
3	30,4	42,6	13	36,4	31,2
4	19,2	60,0	14	23,5	51,8
5	10,0	90,6	15	32,3	34,8
6	12,4	84,3	16	11,0	88,8
7	35,6	30,5	17	14,3	74,6
8	24,4	50,4	18	50,0	22,3
9	38,5	35,0	19	30,2	42,0
10	11,5	89,0	20	27,4	48,5

Задача 2. Имеются следующие данные о размере семьи работников цеха (число человек в семье): 3; 4; 5; 2; 3; 6; 4; 2; 5; 3; 4; 2; 7; 3; 2; 3; 6; 8; 6; 7; 3; 4; 5; 4; 3; 3; 4. Определите среднее число человек в семье.

Задача 3. В населенном пункте проживает 20 тысяч семей, из них 5 тысяч пользуется услугами междугороднего пункта связи, 12 тысяч – подключены к сети сотовой связи, 16 тысяч – имеют стационарные телефонные аппараты.

Определите долю каждой категории семей в общем числе семей населенного пункта.

Задача 4. Используя данные таблицы 52, определите абсолютные показатели вариации и проведите проверку совокупности на однородность.

Таблица 52. **Распределение рабочих цеха по возрасту**

Возраст рабочих, лет.	16-24	24-32	32-40	40-48	48-56
Количество рабочих	20	30	25	19	6

Задача 5. Было проведено 20%-е выборочное обследование в форме простой случайной повторной выборки (табл. 53).

Таблица 53. **Данные о затратах времени обработки детали**

Затраты времени на обработку одной детали, мин.	1,5-2,5	2,5-3,5	3,5-4,5	4,5-5,5
Количество станков, ед.	7	10	15	8

Определите:

- 1) долю станков предприятия, с затратами времени на обработку одной детали более 4,5 минут, гарантируя результат с вероятностью 0,975;
- 2) необходимую численность выборки при определении доли станков предприятия, с затратами времени на обработку одной детали более 4,5 минут, чтобы предельная ошибка не превышала 12 % (вероятность – 0,98).

Вопросов к зачету

1. Предмет и метод статистики
2. Статистическое наблюдение: понятие, требования к данным, элементы
3. Формы, способы и виды статистического наблюдения
4. Ошибки наблюдения
5. Статистические графики
6. Статистические таблицы
7. Группировка и сводка материала. Аналитическая и механическая (простая) группировки
8. Виды статистических величин. Абсолютные величины
9. Относительные величины
10. Средние степенные величины
11. Средние структурные величины: графические и аналитические способы определения для дискретного и интервального рядов распределения
12. Выборочное наблюдение: основные понятия, достоинства и недостатки, ошибки
13. Задачи, которые можно решать с помощью предельной ошибки выборки
14. Простая случайная и механическая выборки
15. Серийная и типическая выборки
16. Вариация признака: понятие, абсолютные и относительные показатели
17. Виды рядов распределения. Способы построения вариационного ряда
18. Проверка совокупности на однородность и нормальность распределения
19. Кривые рядов распределения
20. Показатели центра распределения

Удвоенная нормированная функция Лапласа

t	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3683	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5385	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6176	6528	6579	6629	6680	6729	6778
1,0	6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7339	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9616	9625	9634
2,1	9643	9651	9660	9668	9674	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9922	9924	9926	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	9973	9981	9986	9990	9993	9995	9997	9998	9999	9999

* Значения ординат увеличены в 10000 раз.

Библиографический список

1. **Акулич М.В.** Статистика в таблицах, формулах и схемах: учеб. пособие / М.В. Акулич. – СПб: Питер, 2009. – 205 с.
2. **Баскакова О.В.** Экономика предприятия (организации): учебник для бакалавров / О.В. Баскакова, Л.Ф. Сейко. – М.: Дашков и Ко, 2012. – 370 с.
3. **Кузнецов Е.И.** Статистика: учебник / Е.И. Кузнецов, В.М. Гусаров. – М.: ЮНИТИ-ДАНА, 2011. – 479 с.
4. **Любушкин Н.П.** Экономика организации: учебник / Н.П. Любушкин. – М.: КНОРУС, 2010. – 304 с.
5. **Мелкумов Я.С.** Социально-экономическая статистика: учебник / Я.С. Мелкумов. – М.: Изд-во Инфра-М, 2011. – 236 с.
6. **Минашкин В.Г.** Теория статистики / В.Г. Минашкин, Н.А. Садовникова, Р.А. Шмойлова, Л.Г. Моисейкина, Е.С. Дарда; Под ред Минашкин В.Г. - М.: Евразийский открытый институт, 2011. – 400 с.
7. **Назаров М.Г.** Практикум по общей теории статистики: учеб. пособие / М.Г. Назаров. – Изд-во Кнорус, 2008. – 184 с.
8. **Образцова О.И.** Статистика предприятий и бизнес-статистика: учеб. пособие / О.И. Образцова. – М.: Издательский дом Высшей школы экономики, 2011. – 698 с.
9. **Практикум по социально-экономической статистике:** учеб. пособие / под ред. М.Г.Назарова. – М.: КНОРУС, 2009. – 368 с.
10. **Шмойлова Р.А.** Теория статистики: учебник / Р.А. Шмойлова, В.Г. Минашкин, Н.А. Садовникова, Е.Б. Шувалова. – 5-е изд. – М.: Финансы и статистика, 2011. – 656 с.
11. **Экономика промышленного предприятия:** учебник / под ред. Е.Л. Кантора, Г.А. Маховиковой. – Ростов н/Д: Феникс; М.: ИКЦ «МарТ», 2009. – 864 с.
12. **Экономическая статистика:** учебник / Под ред. проф. Ю.Н. Иванова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2007. – 736 с.