Министерство образования Республики Беларусь
Учреждение образования
«Брестский государственный технический университет»
Кафедра высшей математики

Практикум по статистике

для студентов экономических специальностей

ЧАСТЬ I

Брест 2016

УДК 311 (076.5)

Настоящая разработка содержит в краткой форме методические указания по курсу «Статистика». Рассмотрены статистические показатели, необходимые для решения практических заданий, приведены типовые примеры с подробными решениями, способствующие развитию профессиональных навыков, которые могут быть использованы для решения задач в социально-экономической статистике, статистике финансов, в системе национальных счетов. Также включены задания для самостоятельной работы.

Материалы данной разработки могут быть использованы на занятиях со студентами экономических специальностей всех форм обучения.

Издаётся в 2-х частях. Часть 1.

Составители: Журавель М.Г., ассистент
Золотухина Л.С., старший преподаватель
Копайцева Т.В., старший преподаватель
Кузьмина Е.В., старший преподаватель
Шамовская Г.В., ассистент

СТАТИСТИЧЕСКАЯ СВОДКА И ГРУППИРОВКА. СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ. ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ

Статистическая сводка *– систематизация единичных фактов, позволяющая перейти* к обобщающим показателям, относящимся ко всей изучаемой совокупности и ее частям, и осуществлять анализ и прогнозирование изучаемых явлений и процессов.

Группировка – это процесс образования однородных групп на основе разбиения статистической совокупности на части или объединение изучаемых единиц в частные совокупности по существенным для них признакам. *С помощью метода группировок решаются следующие задачи:*

- выделение социально-экономических типов явлений;
- изучение структуры явления и структурных сдвигов, происходящих в нем;
- выявление связи и зависимости между явлениями.

В соответствии с задачами группировки различают следующие ее виды: *типологическая, структурная, аналитическая.*

Типологическая группировка – это разбиение разнородной совокупности на отдельные качественно однородные группы и выявление на этой основе экономических типов явлений.

Структурная группировка предназначена для изучения состава однородной совокупности по какому-либо варьирующему признаку.

Группировка, выявляющая взаимосвязи между изучаемыми явлениями и их признаками, называется **аналитической группировкой**. Особенностью аналитической группировки является то, что единицы группируются по факторному признаку; каждая выделенная группа характеризуется средними значениями результативного признака. **Факторными** называются признаки, оказывающие влияние на изменение результативных. **Результативными** называются признаки, изменяющиеся под влиянием факторных.

Построение группировки начинается с определения состава группировочных признаков. **Группировочным признаком** называется признак, по которому проводится разбиение единиц совокупности на отдельные группы. Группировка выполняется по одному признаку (простая группировка) и по нескольким признакам (комбинированная группировка). Группировочные признаки делят на **атрибутивные** (профессия рабочих, социальная группа населения) и **количественные** (стаж работы, размер дохода). После определения основания группировки следует решить вопрос о количестве групп. Число групп зависит от задач исследования и вида показателя, положенного в основание группировки, численности совокупности, степени вариации признака. Если группировка строится по *атрибутивному* признаку, то групп, как правило, будет столько, сколько имеется градаций, видов состояний у этого признака. Если группировка проводится по *количественному* признаку, то тогда необходимо обратить особое внимание на число единиц исследуемого объекта и степень колеблемости группировочного признака.

Определение числа групп можно осуществить и математическим путем с использованием формулы Стерджесса:

$$n = 1 + 3{,}322 \cdot lgN,$$

где n – число групп; N – число единиц совокупности.

Когда определено число групп, то следует определить интервалы группировки.

Интервал – это значение варьирующего признака, лежащее в определенных границах. **Нижней границей** интервала называется наименьшее значение признака в интервале, а **верхней границей** – наибольшее значение признака в интервале.

Если вариация признака проявляется *в сравнительно узких границах и распределение носит равномерный характер*, то строят группировку с **равными интервалами**. Величина равного интервала определяется по следующей формуле:

$$h = \frac{x_{max} - x_{min}}{n},$$

где x_{max} и x_{min} – максимальное и минимальное значения признака в совокупности; n – число групп.

Если *размах вариации признака совокупности велик и значения признака варьируются неравномерно*, то необходимо использовать группировку **с неравными интервалами**. Величина интервалов, изменяющихся в арифметической прогрессии, определяется следующим образом: $h_{i+1} = h_i + a$; в геометрической прогрессии: $h_{i+1} = h_i \cdot q$, где a и q – константы.

Закрытыми называются интервалы, у которых имеются верхняя и нижняя границы. У **открытых** интервалов указана только одна граница: верхняя – у первого, нижняя – у последнего.

При определении границ интервалов статистических группировок необходимо исходить из того, что изменение количественного признака приводит к появлению нового качества. В этом случае граница интервала должна устанавливаться там, где происходит переход от одного качества в другое. Это достигается путем использования группировок со **специализированными** интервалами.

Результаты группировки и сводки материалов оформляются в виде **статистических таблиц**.

В статистической таблице выделяются два элемента: **подлежащее** (помещается в первой вертикальной или горизонтальной графе) – перечень единиц или групп, на которые подразделена вся масса единиц наблюдения, и **сказуемое** – цифры, при помощи которых характеризуются выделенные в подлежащем единицы или группы. Над таблицей помещается заголовок, отражающий в сжатой форме ее основное содержание, время, место, к которому относятся изложенные в таблице данные.

Таблица 1

Группировка малых и средних коммерческих банков одного из регионов по величине уставного капитала на 01.01.2008 г.

№ группы	Группы банков по величине уставного капитала, тыс. руб.	Число банков, ед.	Работающие активы, тыс. руб.	Капитал, тыс. руб.	Уставный капитал, тыс. руб.
1	2100 - 7350	18	504898	342889	71272
2	7350 - 12600	6	343932	204694	58227
3	12600 - 17850	3	174059	130680	48281
4	17850 - 23100	3	217842	128573	62238
Итого		30	1240731	806836	240018

В даннном примере группировочный признак – уставный капитал.

После определения группировочного признака и границ групп строится ряд распределения.

Статистический **ряд распределения** – это упорядоченное распределение единиц совокупности на группы по определенному варьирующему признаку. В зависимости от признака, положенного в основу образования ряда распределения, различают *атрибутивные* (построенные по качественному признаку) и *вариационные* (построенные по количественному признаку) ряды распределения.

Любой вариационный ряд состоит из двух элементов: вариант и частот. *Вариантами* считаются отдельные значения признака, которые он принимает в вариационном ряду.

Частоты – это численности отдельных вариант или каждой группы вариационного ряда. Распределение единиц совокупности по дискретному признаку характеризует дискретный вариационный ряд, по непрерывному признаку – интервальный вариационный ряд.

Графические методы в статистике являются способом наглядного изображения результатов статистической сводки и обработки массового материала.

Для анализа рядов распределения строят полигон, гистограмму, огиву и кумуляту распределения.

Для построения **полигона** в прямоугольной системе координат по оси абсцисс в одинаковом масштабе откладываются ранжированные значения варьирующего признака, а по оси ординат наносится шкала для выражения величины частот. Полученные на пересечении абсцисс и ординат точки соединяют прямыми линиями и получают ломаную линию, называемую полигоном частот.

При построении **гистограммы** на оси абсцисс откладываются величины интервалов, а частоты изображаются прямоугольниками, построенными на соответствующих интервалах. Высота столбиков должна быть пропорциональна частотам. Если середины верхних сторон прямоугольников соединить прямыми, то гистограмма может быть преобразована в полигон распределения.

При построении **кумуляты** интервального вариационного ряда по оси абсцисс откладываются варианты ряда, а по оси ординат – накопленные частоты, которые наносят на поле графика в виде перпендикуляров к оси абсцисс в верхних границах интервалов. Затем эти перпендикуляры соединяют и получают ломаную линию, т.е. кумуляту. Если при графическом изображении вариационного ряда в виде кумуляты оси поменять местами, то получим **огиву**.

Для графического изображения взаимосвязи между явлениями на оси абсцисс необходимо поместить значения признака-фактора, а на оси ординат – значения признака-результата.

Пример 1. Пусть имеются следующие данные о тарифных разрядах 50 рабочих одного из цехов завода:

3	5	6	3	2	4	3	5	5	6
4	3	2	3	4	5	4	2	4	6
5	3	4	5	4	3	3	6	2	3
4	6	3	4	4	5	4	5	3	4
2	6	3	4	5	3	4	4	5	4

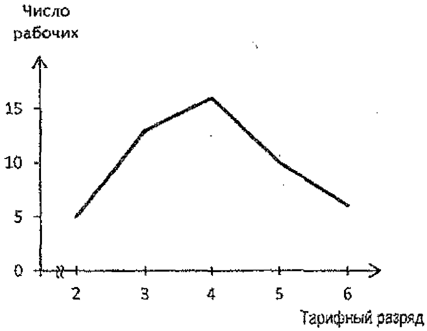
Построить вариационный ряд — распределение 50 рабочих по тарифному разряду. Изобразить ряд графически.

Решение. Выпишем все значения тарифного разряда (дискретный признак) в порядке возрастания и подсчитаем число рабочих в каждой группе:

Таблица 2

Тарифный разряд	Численность рабочих, чел.
2	5
3	13
4	16
5	10
6	6
Всего	50

Это дискретный вариационный ряд, у которого вариантами являются значения тарифного разряда, а частотами — число рабочих. Графическим изображением такого ряда является полигон.



Пример 2. Пусть имеются следующие данные о стоимости основных фондов предприятий, млн руб. (непрерывный признак):

9,4	8,0	6,3	10,0	15,0	8,2	7,3	9,2	5,8	8,7
5,2	13,2	8,1	7,5	11,8	14,6	8,5	7,8	10,5	6,0
5,1	6,8	8,3	7,7	7,9	9,0	10,1	8,0	12,0	14,0
8,2	9,8	13,5	12,4	5,5	7,9	9,2	10,8	12,1	12,4
12,9	12,6	6,7	9,7	8,3	10,8	15,0	7,0	13,0	9,5

Построить вариационный ряд — распределение заводов по стоимости основных фондов. Изобразить ряд графически.

Решение. Сначала решим вопрос о количестве групп, которые мы хотим выделить. Предположим, решено выделить пять групп заводов. Чтобы определить величину интервала в группе, найдем разность между максимальным и минимальным значениями признака и разделим ее на число выделяемых групп. В нашем примере

$$h = \frac{x_{max} - x_{min}}{n} = \frac{15 - 5,1}{5} \approx 2 \text{ (млн руб.)}$$

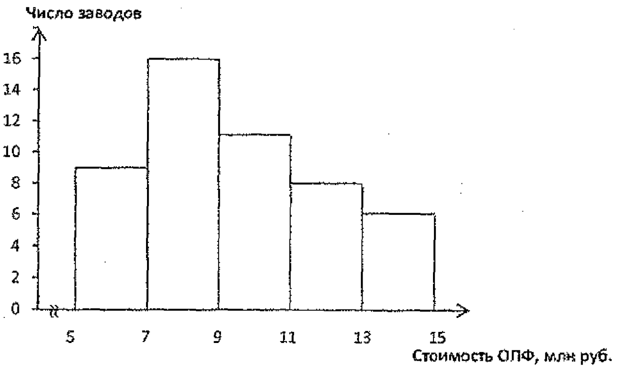
Выделим теперь группы с интервалом 2 млн руб. и подсчитаем число заводов в каждой группе (частоту):

Таблица 3

Стоимость основных фондов, млн руб.	Число заводов
5-7	9
7-9	16
9-11	11
11-13	8
13-15	6
Всего	50

Это интервальный вариационный ряд с равными интервалами. При такой записи непрерывного признака, когда одна и та же величина встречается дважды (как верхняя граница одного интервала и как нижняя граница другого интервала), единица, обладающая этим значением, обычно относится к той группе, где эта величина выступает в роли верхней границы. Так, в нашем примере завод со стоимостью основных фондов 9 млн руб. отнесен ко второй группе (а не к третьей).

Графическим изображением является гистограмма.



Мы рассмотрели примеры группировок по одному признаку. Однако в ряде случаев для решения поставленных задач такая группировка является недостаточной. В этих случаях переходят к группировке исследуемой совокупности по двум и более существенным признакам во взаимосвязи (сложной группировке).

Пример 3. Имеются следующие данные по заработной плате водителей за сентябрь.

Таблица 4

Табельный номер водителя	Класс водителя	Процент выполнения сменных заданий, %	Заработная плата за сентябрь, руб.
1	I	110,2	4100,3
2	II	102,0	3600,8
3	II	111,0	3970,7
4	I	107,9	4050,2
5	II	106,4	3740,5
6	I	109,0	3985,4
7	I	115,0	4300,8
8	II	112,2	4015,7
9	I	105,0	3790,2
10	II	107,4	3700,7
11	I	112,5	4280,2
12	I	108,6	4170,1

Требуется для выявления зависимости заработной платы от уровня квалификации и процента выполнения сменных заданий произвести аналитическую группировку. Интервалы группировки водителей по проценту выполнения норм выработки разработать самостоятельно. На основе выполненной группировки построить комбинационную таблицу. Сформулировать вывод.

Решение. Для решения задачи необходимо произвести группировку водителей по двум признакам-факторам: сначала – на группы по квалификации, затем внутри каждой группы на подгруппы по проценту выполнения сменного задания. По проценту выполнения сменного задания принимаются две подгруппы: водители, выполняющие норму от 100% до 110%; водители, выполняющие норму на 110% и выше.

Результаты группировки представлены во вспомогательной таблице 5.

На основе вспомогательной таблицы по каждой подгруппе определяют численность признака (общую сумму заработной платы), результаты оформляют в виде комбинационной таблицы 6.

Таблица 5

Группы водителей по квалификации	Водители II класса		Водители I класса	
Подгруппы водителей по проценту выполнения сменного задания	100 - 110	110 и выше	100 - 110	110 и выше
Табельный номер рабочего	2; 5; 10	3; 8	4; 6; 9; 12	1; 7; 11
Заработная плата за месяц, руб.	3600,8 3740,5 3700,7	3970,7 4015,7	4050,2 3985,4 3790,2 4170,1	4100,3 4300,8 4280,2

Таблица 6

Группы водителей по уровню квалификации	Подгруппы водителей по проценту выполнения сменного задания	Число водителей, чел.	Общая сумма заработной платы, руб.	Средняя заработная плата одного водителя, руб.	Изменение средней заработной платы по сравнению с низшей подгруппой, %
II класс	100 -110 110 и выше	3 2	11042,0 7986,4	3680,7 3993,2	100,0 108,5
Итого по группе		5	19028,4	3805,7	-
1	2	3	4	5	6
I класс	100 -110 110 и выше	4 3	15995,9 12681,3	3999,0 4227,1	108,6 114,8
Итого по группе		7	28677,2	4096,7	-
Всего		12	47705,6	3975,5	-

Вывод: из табл. 6 следует, что с ростом квалификации водителей и процента выполнения сменного задания увеличивается заработная плата. Так, заработная плата водителей I класса, выполняющих норму выработки на 110% и выше, на 14,8% превышает заработную плату водителей II класса, выполняющих норму от 100-110%.

Аудиторные задания

1. Выполнение установленных норм выработки рабочими бригады строителей-монтажников характеризуется следующими данными за июнь:

№	ФИО	Выполнение нормы, %	№	ФИО	Выполнение нормы, %
1	Спицын М.И.	101,3	13	Щербаков А.В.	110,3
2	Кузьмин К.С.	124,6	14	Башлыков Б.Р.	102,6
3	Федоренко А.П.	103,5	15	Тенисон В.Ф.	100,1
4	Симонович В.В.	130,1	16	Устинов Г.Д.	153,0
5	Потапенко Н.М.	113,9	17	Ильин М.С.	109,9
6	Глебов Л.Н.	89,6	18	Дронов Б.Г.	117,6
7	Фролов И.И.	104,7	19	Киселев Ф.А.	102,4
8	Бариков Б.Я.	120,0	20	Вельдин М.С.	105,6
9	Кузнецов А.И.	119,5	21	Сомов Е.Н.	99,7
10	Петров С.С.	108,6	22	Федин З.П.	107,4
11	Синицын Н.К.	97,8	23	Мохов Р.И.	106,1
12	Королев А.М.	107,2	24	Стогов К.М.	112,2

Постройте ряд распределения рабочих по степени выполнения ими установленных норм выработки, приняв следующие интервалы: 1) до 100%; 2) от 100 до 110%; 3) от 110 до 120%; 4) от 120% и выше. Решение представьте в форме статистической таблицы.

2. Ниже приведены данные о количестве членов семьи в 50 обследованных фермерских хозяйствах:

2	5	5	6	3	2	5	6	5	6
6	6	4	3	3	5	7	3	5	5
5	4	5	6	4	4	4	4	7	4
4	3	5	3	7	4	6	6	4	7
4	4	6	7	6	3	3	5	8	5

Построить вариационный ряд. Изобразить ряд графически.

3. Имеются следующие данные об урожайности озимой пшеницы в 40 обследованных хозяйствах:

27,1	18,2	16,3	22,0	24,3	24,8	33,0	27,3	25,3	21,2
28,5	15,1	19,5	28,1	25,1	26,7	28,4	29,6	23,9	22,6
23,7	18,0	31,0	19,8	26,0	23,5	20,2	25,1	21,5	25,8
22,8	27,0	20,4	24,0	29,5	22,9	19,9	27,0	23,1	23,8

Построить интервальный вариационный ряд с равными интервалами, выделив шесть групп хозяйств по величине урожайности. Изобразить ряд графически.

4. Известны следующие данные по основным показателям деятельности крупнейших банков одного из регионов (данные условные, в тыс. руб.):

№ п/п	Сумма активов	Собственный капитал	Привлеченные ресурсы	Балансовая прибыль	Объем вложений в государственные ценные бумаги	Ссудная задолженность
1	645,6	12,0	27,1	8,1	3,5	30,8
2	636,9	70,4	56,3	9,5	12,6	25,7
3	629,0	41,0	95,7	38,4	13,3	26,4
4	619,6	120,8	44,8	38,4	4,4	25,3
5	616,4	49,4	108,7	13,4	15,0	20,9
6	614,4	50,3	108,1	30,1	19,1	47,3
7	608,6	70,0	76,1	37,8	19,2	43,7
8	601,1	52,4	26,3	41,1	3,7	29,1
9	600,2	42,0	46,0	9,3	5,2	56,1
10	600,0	27,3	24,4	39,3	13,1	24,9
11	592,9	72,0	65,5	8,6	16,7	39,6
12	591,7	22,4	76,0	40,5	7,5	59,6
13	585,5	39,3	106,9	45,3	6,7	44,9
14	578,6	70,0	89,5	8,4	11,2	32,2
15	577,5	22,9	84,0	12,8	19,3	45,1
16	553,7	119,3	89,4	44,7	19,4	24,5
17	543,6	49,6	93,8	8,8	5,7	31,1
18	542,0	88,6	26,7	32,2	7,8	37,1
19	517,0	43,7	108,1	20,3	8,3	23,1
20	516,7	90,5	25,2	12,2	9,7	15,8

Постройте группировку коммерческих банков по величине собственного капитала, выделив не более пяти групп с равными интервалами. Рассчитайте по каждой группе сумму активов, собственный капитал, привлеченные ресурсы, балансовую прибыль. Результаты группировки представьте в табличной форме и сформулируйте выводы.

5. Постройте структурную группировку банков по величине балансовой прибыли, выделив четыре группы банков с открытыми интервалами для характеристики структуры совокупности коммерческих банков, перечисленных в задаче 4.

6. Постройте аналитическую группировку коммерческих банков, перечисленных в задаче 4, по величине балансовой прибыли, выделив четыре-пять групп. Рассчитайте по каждой группе два-три показателя, взаимосвязанных с балансовой прибылью. Результаты группировки изложите в табличной форме и сделайте выводы о взаимосвязи показателей.

7. Используя данные задачи 4, постройте группировку коммерческих банков в целях выявления взаимосвязи между показателями привлеченных ресурсов, объемом вложений в государственные ценные бумаги и ссудной задолженностью от результатов деятельности банков (показатель, выражающий результаты деятельности банков, определите самостоятельно).

8. По данным задачи 4 постройте группировку коммерческих банков по двум признакам: по величине балансовой прибыли и сумме активов. По каждой группе и подгруппе

определите число банков и два-три показателя, взаимосвязанных с группировочными. Результаты оформите в виде таблицы, сделайте выводы.

9. Известны следующие данные о результатах сдачи абитуриентами ЦТ по математике в 2015 г. (баллов):

48	12	18	15	84	44	14	62	13	56
16	16	77	9	52	52	16	54	42	87
17	20	56	35	23	18	82	49	48	88
18	36	19	91	10	61	36	53	49	15

Постройте: 1) ряд распределения абитуриентов по результатам сдачи ими ЦТ по математике, выделив четыре группы с равными интервалами; 2) ряд, делящий абитуриентов на набравших и не набравших проходной балл, учитывая, что проходной балл составил 15 баллов. Укажите, по какому группировочному признаку построен каждый из этих рядов распределения: атрибутивному или количественному.

Домашние задания

1. На основании следующих данных произведите группировку по среднесписочной численности работников, разделив всю совокупность магазинов на три группы. По каждой группе рассчитайте годовой товарооборот всего и в среднем на одного работника. Сведите данные в таблицу и сделайте соответствующие выводы.

№ магазина	Среднесписочная численность работников, чел.	Торговая площадь, $м^2$	Годовой товарооборот, млн руб.	№ магазина	Среднесписочная численность работников, чел.	Торговая площадь, $м^2$	Годовой товарооборот, млн руб.
1	21	186	1295	16	48	390	2660
2	68	579	2876	17	20	150	920
3	45	630	2411	18	30	175	1376
4	45	510	2460	19	42	620	1775
5	34	468	1900	20	47	350	2520
6	18	196	902	21	51	492	2200
7	53	420	2692	22	45	380	1990
8	41	486	1475	23	63	537	2560
9	48	441	2430	24	18	203	700
10	29	280	1032	25	57	370	2912
11	45	750	2343	26	60	550	2710
12	34	240	1810	27	19	250	820
13	40	458	2312	28	40	581	2405
14	32	190	1600	29	20	190	1306
15	32	240	1284	30	65	545	2601

2. На основании данных, представленных в задаче 1 произведите группировку по торговой площади, разделив совокупность на три группы. По каждой группе рассчитайте годовой товарооборот всего и в среднем на один магазин. Оформите результаты в виде таблицы. Сделайте соответствующие выводы.

3. На основании данных задачи 1 выявите зависимость годового товарооборота от среднесписочной численности работников и торговой площади, произведя комбинированную группировку, разделив совокупность магазинов на три группы по одному группировочному признаку и на две подгруппы по второму группировочному признаку. Оформите комбинационную таблицу. Сделайте соответствующие выводы.

4. Изобразите графически ряд распределения магазинов по признаку из задачи 1.

СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ. СИСТЕМА ПОКАЗАТЕЛЕЙ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ СТАТИСТИКЕ

Статистический показатель (СП) представляет собой количественную характеристику социально-экономических процессов в условиях качественной определенности. Все СП классифицируются на абсолютные, относительные и средние.

Абсолютные показатели отражают либо суммарное число единиц, либо суммарное свойство объекта. Данные показатели выражаются только именованными величинами.

Единицы измерения абсолютных показателей

Натуральные	Стоимостные	Трудовые
(кг, км, тонны, штуки, часы и др.)	(ден. единицы)	(общие затраты труда, трудоёмкость, человека – дни и др.)

Так, одним из важнейших стоимостных показателей в системе национальных счетов (СНС) является ВВП.

Относительный показатель представляет собой результат деления одного абсолютного показания на другой и выражает соотношение между количественными характеристиками социально-экономических процессов и явлений. Относительные показатели выражаются: в коэффициентах (долях единицы), процентах (%), промилле (‰) и т. д., а могут быть именованными величинами.

Относительный показатель динамики (ОПД) характеризует изменение уровня развития какого-либо явления во времени, т. е.

$$\text{ОПД} = \frac{\text{Текущий уровень}}{\text{Предшествующий или базисный уровень}}.$$

Относительные показатели плана (ОПП) и реализации плана (ОПРП) используются всеми субъектами финансово-хозяйственной деятельности, так как помогают осуществлять как оперативное, так и стратегическое планирование, а также сравнивать достигнутые результаты с ранее намеченными.

$$\text{ОПП} = \frac{\text{Уровень, планируемый на } (i+1) - \text{й период}}{\text{Уровень, достигнутый в } i - \text{м периоде}}.$$

$$\text{ОПРП} = \frac{\text{Уровень, достигнутый в } (i+1) - \text{м периоде}}{\text{Уровень, планируемый на } (i+1) - \text{й период}}.$$

Между относительными показателями плана, выполнением плана и динамики существует следующая взаимосвязь: $\text{ОПП} \cdot \text{ОПРП} = \text{ОПД}$.

Пример 1. Объем продаж компании Samsung в странах СНГ в первом полугодии 1996 г. составил 250 млн. долларов. В целом же за год компания планировала реализовать товаров на 600 млн. долларов. Вычислить относительный показатель плана на второе полугодие.

Решение. Относительный показатель плана на второе полугодие есть отношение планируемой величины на второе полугодие к фактически достигнутой величине в первом полугодии: $\frac{600-250}{250} \cdot 100\% = 140\%$.

Пример 2. Предприятие планировало увеличить выпуск продукции в 2015 г. по сравнению с 2014 г. на 18%. Фактически же объем продукции составил 112,3% от прошлогоднего уровня. Определить относительный показатель выполнения плана.

Решение. Относительный показатель выполнения плана в 2015 г. есть отношение фактического выпуска продукции к планируемому в 2015 г.:

$$\frac{1{,}123}{1+0{,}18}\cdot 100\% = \frac{1{,}123}{1{,}18}\cdot 100\% = 95{,}2\%,$$

т. е. плановое задание было недовыполнено на 4,8%.

Относительный показатель структуры (ОПС) характеризует доли, удельные веса составных элементов в общем итоге.

$$\text{ОПС} = \frac{\text{Показатель, характеризующий часть совокупности}}{\text{Показатель по всей совокупности в целом}}.$$

Относительный показатель интенсивности (ОПИ) характеризует степень распределения или развития данного явления в той или иной среде

$$\text{ОПИ} = \frac{\text{Показатель, характеризующий явление А}}{\text{Показатель, характеризующий среду распространения явления А}}.$$

ОПИ используется при изучении демографических процессов (показатели рождаемости, смертности, брачности и др.). Так, например, коэффициент рождаемости можно рассчитать по формуле:

$$\text{К}_{\text{рожд.}} = \frac{\text{Число родившихся за год}}{\text{Среднегодовая численность населения}}\cdot 1000\ ‰.$$

Разновидностью ОПИ являются относительные показатели уровня экономического развития, характеризующие производство продукции в расчете на душу населения.

Пример 3. Произведство ВВП в стране «Оранжевая речка» в 2015 г. в текущих ценах составило 26 781,1 млрд ден. ед. Среднегодовая численность населения в 2015 г. составила 142,8 млн чел. Вычислить размер ВВП на душу населения.

Решение. С учетом перевода миллиардов в миллионы и получим:

$$\frac{26\,781{,}1}{142{,}8}\cdot 100 = 187\,543\ (\text{ден. ед.})$$

В социально-экономической статистике часто используются следующие показатели: фондоотдача (Φ_o), фондоёмкость (Φ_e), фондовооруженность ($\Phi_в$).

$$\Phi_o = \frac{\text{Объем произведенной в данном периоде продукции}}{\text{Средняя за данный период стоимость ОПФ}}.$$

$$\Phi_e = \frac{\text{Средняя за данный период стоимость ОПФ}}{\text{Объем произведенной в данном периоде продукции}}.$$

$$\Phi_в = \frac{\text{Средняя за данный период стоимость ОПФ}}{\text{Среднесписочная численность работающих}}.$$

Фондоотдача показывает, сколько продукции произведено в данном периоде на 1 рубль стоимости основных производственных фондов (ОПФ). Фондоемкость (обратная величина) характеризует стоимость ОПФ, приходящуюся на 1 рубль произведенной продукции. Показатель фондовооруженности применяется для характеристики степени оснащенности труда работающих.

СРЕДНИЕ ВЕЛИЧИНЫ

Наиболее распространенной формой статистических показателей, используемой в экономических исследованиях, является средняя величина, представляющая собой обобщенную количественную характеристику признака в статистической совокупности в конкретных условиях места и времени.

На практике определить среднюю величину во многих случаях можно через **исходное соотношение средней** (ИСС) или ее логическую формулу:

$$\text{ИСС} = \frac{\text{Суммарное значение или объем осредняемого признака}}{\text{Число единиц или объем совокупности}}.$$

Заметим, что любой статистический показатель имеет единственное ИСС.

Для изучения и анализа социально-экономических явлений применяются различные виды степенных средних величин (средняя арифметическая, гармоническая и др.), а также структурные средние.

Средние могут рассчитываться в двух вариантах: взвешенные и невзвешенные (простые). Веса могут быть представлены как абсолютными величинами, так и относительными (в % или долях единиц).

Вид степенной средней	Формула расчета	
	простая	взвешенная
арифметическая	$\overline{x} = \frac{\sum x_i}{n}$	$\overline{x} = \frac{\sum x_i f_i}{\sum f_i}$
гармоническая	$\overline{x} = \frac{n}{\sum \frac{1}{x_i}}$	$\overline{x} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}}$
геометрическая	$\overline{x} = \sqrt[n]{\prod x_i}$	$\overline{x} = \sqrt[\sum f_i]{\prod (x_i)^{f_i}}$
квадратическая	$\overline{x} = \sqrt{\frac{\sum x_i^2}{n}}$	$\overline{x} = \sqrt{\frac{\sum x_i^2 f_i}{f_i}}$

В интервальном вариационном ряду для расчета средней арифметической взвешенной определяются и используются значения середины интервалов.

<u>**Пример 1.**</u> Рассмотрим методологию применения разных видов степенных средних на основе расчета среднего курса акций по двум эмитентам.

Данные о реализации акций на фондовой бирже

Эмитент	Дата торгов			
	10.11		15.11	
	Курс акций, $	Количество проданных акций, шт.	Курс акций, $	Объем продаж, $
А	25,4	400	27,3	1365,0
В	23,9	950	20,5	19450,2

Решение. По условию задачи необходимо рассчитать средний курс акций на каждую дату торгов.

Определим логическую формулу для расчета осредняемого признака.

$$ИСС(\text{средний курс акций}) = \frac{\text{Общий объем продаж, \$}}{\text{Общее количество проданных акций, шт.}}.$$

Обратим внимание на то, что единицей совокупности является акция, а не эмитент.

На 10.11 мы располагаем данными о курсе акций (признак X) и количеству проданных акций (признак У) по каждому эмитенту. Таким образом, известен знаменатель, но неизвестен числитель. Объем продаж можно найти, умножив курс акций на количество проданных акций ($x_i \cdot y_i$).

Аналитическая формула имеет вид:

$$\bar{x} = \frac{\sum x_i y_i}{\sum y_i}.$$

Тогда

$$\bar{x} = \frac{\sum x_i y_i}{\sum y_i} = \frac{25{,}4 \cdot 400 + 23{,}9 \cdot 950}{400 + 950} = 24{,}3\ (\$).$$

На вторую дату, 15.11, имеем данные о курсе акций (признак X) и общем объеме продаж (признак Z) по каждому эмитенту. В этом случае известен числитель, а неизвестен знаменатель, который можно получить делением объема продаж на курс акций $\left(\frac{z_i}{x_i}\right)$.

Аналитическая формула имеет вид:

$$\bar{x} = \frac{\sum z_i}{\sum \frac{z_i}{x_i}}.$$

Тогда

$$\bar{x} = \frac{\sum z_i}{\sum \frac{z_i}{x_i}} = \frac{1365{,}0 + 19450{,}20}{\frac{1365{,}0}{27{,}3} + \frac{19450{,}2}{20{,}5}} = 20{,}8\ (\$).$$

Таким образом, в первом случае применена формула средней арифметической взвешенной, а во втором – гармонической взвешенной.

Пример 2. Имеем следующие условные данные о распределении страховых организаций по размеру уставного капитала.

Порядок распределения страховых организаций по размеру уставного капитала

Группы страховых организаций по размеру уставного капитала, млн руб.	Количество страховых организаций, % к итогу
До 3	19,1
3-5	14,5
5-7	9,3
7-9	7,4
9-11	10,9
Свыше11	38,8
Итого	100

Определить средний размер уставного капитала по страховым организациям.

Решение.

Определим логическую формулу для расчета «среднего размера уставного капитала»:

$$\text{ИСС}\begin{pmatrix}\text{средний размер}\\ \text{уставного капитала}\end{pmatrix} = \frac{\text{Общий размер уставного капитала (млн руб.)}}{\text{Общее количество страховых организаций}}.$$

Обозначим значения осредняемого признака (размер уставного капитала) через x_i, а частоту повтора данного признака (количество страховых организаций) через f_i.

Роль «количества страховых организаций» в данном случае выполняет его доля в общем итоге, выраженная в процентах.

Группы страховых организаций по размеру уставного капитала, млн руб.	Середина интервала	Количество страховых организаций, % к итогу
$x_{i-1} - x_i$	x_i'	f_i
До 3	2	19,1
3-5	4	14,5
5-7	6	9,3
7-9	8	7,4
9-11	10	10,9
Свыше 11	12	38,8
Итого	–	100,0

Для расчета будем использовать формулу средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x_i f_i}{\sum y_i} = \frac{2 \cdot 19{,}1 + 4 \cdot 14{,}5 + 6 \cdot 9{,}3 + 8 \cdot 7{,}4 + 10 \cdot 10{,}9 + 12 \cdot 38{,}8}{100} = 7{,}9 \text{ (млн руб.)}$$

<u>**Пример 3.**</u> Имеются следующие данные о работе малых предприятий за текущий период:

Предприятие	Фактический объем реализации, тыс. руб.	Средний объем реализации на одного работника, тыс. руб.	Прибыль к объему реализации, %	Процент совместителей в общей численности работников, %
	x	y	z	d
1	190	38	19	66
2	160	40	20	70
3	200	50	26	60
4	192	32	20	75

Определить по малым предприятиям района средние значения:

1) реализованной продукции на одно предприятие;
2) производительности труда;
3) рентабельности продукции;
4) доли совместителей в общей численности работников. Указать вид рассчитанных средних величин и сделать выводы.

Решение.

Введем обозначения:

X – фактический объем реализации;

Y – средний объем реализации на одного работника;

Z – прибыль к объему реализации;

D – процент совместителей в общей численности работников.

Выбор вида средней величины необходимо начинать с построения логической формулы, исходя из качественного содержания усредняемого показателя.

Все расчеты будем оформлять в таблице:

№ предприятия	Фактический объем реализации, тыс. руб. (x_i)	Средний объем реализации на одного работника, тыс. руб. (y_i)	Количество работающих чел., $\left(\frac{x_i}{y_i}\right)$	Прибыль к объему реализации, % (z_i)	Прибыль, тыс. руб., $\left(\frac{x_i \cdot z_i}{100}\right)$	Совместители, чел., $\frac{x_i}{y_i} \cdot \frac{d_i}{100}$	Процент совместителей в общей численности работников, % (d_i)
1	190	38	5	19	190·0,19 =36,1	3,3	66
2	160	40	4	20	160·0,20 =32,0	2,8	70
3	200	50	4	26	200·0,26=52,0	2,4	60
4	192	32	6	20	192·0,20 =38,4	4,5	75
$\sum$	742	–	19	–	158,5	13,0	–

$$1)\ \begin{matrix}\textit{Средний объем}\\ \textit{реализованной}\\ \textit{продукции}\end{matrix} = \frac{\textit{Совокупный объем реализованной продукции всеми предприятиями (тыс.руб.)}}{\textit{количество предприятий}}$$

Получаем, таким образом, среднюю арифметическую простую

$$\overline{x} = \frac{\sum x_i}{n} = \frac{742}{4} = 185{,}5\ (\text{тыс. руб.})$$

$$2)\ \begin{matrix}\text{Средняя}\\ \text{производительность}\\ \text{труда}\end{matrix} = \frac{\text{Совокупный объем реализованной продукции (тыс. руб.)}}{\text{Общее количество работающих (чел.)}}$$

Количество работающих найдем, зная, что:

$$\begin{matrix}\textit{Количество}\\ \textit{работающих}\end{matrix} = \frac{\textit{Объем реализованной продукции (тыс.руб.)}}{\textit{Средний объем реализованной продукции на одного работающего (тыс.руб.)}}$$

Получаем среднюю производительность труда по четырем предприятиям в виде средней гармонической взвешенной:

$$\overline{y} = \frac{\sum x_i}{\sum \frac{x_i}{y_i}} = \frac{742}{19} = 39{,}05\ (\text{тыс. руб.}),$$

т. е. по четырем предприятиям объем реализации продукции на одного работника предприятия составляет в среднем 39,05 тыс. руб.

$$3)\ \begin{matrix}\textit{Рентабельность}\\ \textit{продукции}\end{matrix} = \frac{\textit{Суммарная прибыль (тыс.руб.)}}{\textit{Совокупный объем реализованной продукции (тыс.руб.)}}$$

Среднюю рентабельность продукции в процентах найдем как среднюю арифметическую взвешенную:

$$\bar{z} = \frac{\sum \frac{x_i z_i}{100}}{\sum x_i} \cdot 100\% = \frac{\sum x_i z_i}{\sum x_i} = \frac{158{,}5}{742} \cdot 100\% = 21{,}4\ \%,$$

т. е. в среднем прибыль составляет 21,4% к объему реализации.

$$4)\ \begin{matrix}\text{Совместители}\\ \text{в общей численности}\\ \text{работников}\end{matrix} = \frac{\text{Общая численность совместителей (чел.)}}{\text{Общая численность работников (чел.)}}$$

$$\bar{d} = \frac{\sum \frac{x_i}{y_i} \cdot \frac{d_i}{100}}{\sum \frac{x_i}{y_i}} \cdot 100\%$$ – средняя арифметическая взвешенная.

$\bar{d} = \frac{13}{19} \cdot 100\% = 68{,}4\ \%$ – средний процент совместителей в общей численности работающих.

Т. е. по четырем предприятиям совместители составляют в среднем 68,4% от общей численности работающих.

Аудиторные задания

1. Имеются данные о финансовых результатах работы предприятий, входящих в состав коммерческой фирмы:

Предприятие	Прибыль, ден. ед.	Рентабельность,%
1	24	32
2	17	24
3	22	35

Определите среднюю рентабельность фирмы.

2. Данные о средней заработной плате по двум предприятиям следующие:

Предприятие	Январь		Февраль	
	Средняя заработная плата, руб.	Численность работников, чел.	Средняя заработная плата, руб.	Фонд оплаты труда, тыс. руб.
1	4900	450	5700	2565
2	5400	600	5800	3480

Определите среднюю заработную плату по двум предприятиям вместе: 1) за январь; 2) за февраль; 3) за оба эти месяца.

3. Имеются данные обследования трех семей:

Семья	Размер жилой площади, приходящейся на одного члена семьи, кв. м.	Число членов семьи, чел.	Процент жилой площади в общей площади,%
1	15,0	2	85
2	13,8	3	81
3	11,2	4	83

Определить по трем семьям средний процент жилой площади в общей площади на одного члена семьи.

4. По трем районам города имеются следующие данные (на конец год):

Район	Число отделений «ХХХ» банка	Среднее число вкладчиков в отделение, чел.	Размер вклада в отделении, руб.
1	4	1376	275
2	9	1559	293
3	5	1315	268

Определите средний размер вклада в «ХХХ» банке в целом по городу.

5. Товарооборот коммерческих киосков за отчетный период характеризуется следующими данными:

Киоск	Товарооборот, млн руб.	Товарооборот в расчете на одного работника, млн руб.	Средняя заработная плата на одного работника, тыс. руб.
1	30	7,1	600
2	28	5,2	595
3	32	6,8	610

Определите среднюю заработную плату на одного работника по всей совокупности объектов.

6. Определите среднюю урожайность и среднее значение затрат труда на 1 га посевной площади по совокупности объектов:

Колхозы	Валовой сбор, тыс. ц	Урожайность, ц/га	Затраты труда на 1 ц, чел/час
1	131	32,4	42,6
2	242	28,2	35,9
3	229	31,5	38,1

7. Имеются данные по предприятиям:

Предприятие	Объем произведенной продукции, млн руб.	Производительность труда, тыс. руб.	Доля пенсионеров в общей численности рабочих	Средняя заработная плата на одного работника, тыс. руб.
1	35,15	17,5	0,15	418
2	23,13	14,1	0,23	247
3	42,62	18,3	0,12	525

Определите среднее значение каждого признака по совокупности объектов, используя экономически обоснованные формулы расчета. Укажите вид и форму вычисленных средних.

Домашние задания

1. Имеем следующие условные данные о реализации акций на фондовой бирже:

Эмитент	Дата торгов					
	15.06		22.06		29.06	
	Объем продаж, тыс. руб.	Количество проданных акций, шт.	Курс акций, руб.	Объем продаж, тыс. руб.	Курс акций, руб.	Количество проданных акций, шт.
1	7,2	400	21,5	8,17	20,0	300
2	19,2	800	23,0	19,78	27,0	1000

Определите: 1) средний курс акций на бирже по каждой дате торгов (укажите вид используемых средних); 2) на сколько процентов изменялся курс акций на торгах от даты к дате и в среднем за рассматриваемый период.

2. Имеются следующие данные по трем магазинам о реализации товара «А»:

Номер магазина	Рентабельность, %	Объем реализации, ед.	Стоимость одной единицы товара, тыс. руб.
1	29	324	15
2	18	190	21
3	32	438	12

Определить среднюю рентабельность по совокупности.

3. Имеются следующие данные по трем заводам:

Завод	Стоимость ОПФ, млн ден. ед.	Из них активная часть, %	Стоимость ОПФ на одного рабочего, тыс. ден. ед.
1	32,9	85	9,4
2	23,8	68	11,9
3	24,0	79	9,0

Определить по трем заводам среднюю стоимость активной части ОПФ на одного рабочего.

4. Определите средний объем реализации на одного работника и средний процент продавцов в общей численности работников по совокупности объектов:

Магазины	Фактический объем реализации, млн ден. ед.	Средний объем реализации на одного работника, тыс. ден. ед.	Процент продавцов в общей численности работников,%
1	370	845	85
2	295	450	74
3	305	944	91

5. По результатам обследования сельхозпредприятий области получены следующие данные:

Группы сельхозпредприятий по среднегодовому надою молока от одной коровы, кг	Число предприятий, шт.	Среднегодовое поголовье коров (на 1 сельхозпредприятие), голов	Процент жирности в молоке,%
2000-2200	4	417	3,0
2200-2400	9	350	3,3
2400-2600	15	483	3,8

Определите средний надой молока на одну корову и среднюю жирность молока.

СТРУКТУРНЫЕ СРЕДНИЕ

Наиболее часто используемыми в экономической практике структурными средними являются мода и медиана.

Мода (M_o) – значение признака, повторяющееся с наибольшей частотой.

Медиана (M_e) – значение признака, приходящееся на середину ранжированной совокупности.

Для дискретных вариационных рядов модой будет значение варианта с наибольшей частотой. Если дискретный ряд распределения имеет нечетное число несгруппированных данных, то медианой будет значение признака, находящееся в середине ранжированного ряда; если же четное число членов - то медианой будет средняя арифметическая из двух значений признака, расположенных в середине ряда.

Пример 1. Предположим, что 11 торговых предприятий города реализуют товар А по следующим оптовым ценам (руб.): 51, 49, 50, 49, 52, 54, 55, 49, 50, 51, 54.

Определить моду и медиану оптовой цены товара А.

Решение.

Упорядочим данные:

49, 49, 49, 50, 50, 51, 51, 52, 54, 54, 55.

M_o=49 руб.

Центральной в этом ряду, т. е. шестой по счету, является цена 51 руб., значит, M_e=51 руб.

Пример 2. Известно распределение торговых предприятий города по уровню цен на товар А. Рассчитать структурные средние – моду и медиану.

Цена, x_i, руб.	52	53	54	56	59
Число предприятий, f_i	12	48	56	60	14
Накопленная частота, f_i'	12	60	116	176	190

Решение.

1. Мода M_o=56 руб. – наиболее часто встречающаяся цена товара А.
2. Для расчета медианы M_e определим

$$\frac{1}{2}\sum f_i = \frac{1}{2}(12+48+56+60+14)=95.$$

Подсчитаем накопленные частоты.

Определим накопленную частоту, впервые превышающую половину объема выборки.

116>95

Значит, M_e=54 руб., т. е. в среднем у половины предприятий города цена на товар А меньше 54 руб., у половины – больше 54 руб.

Пример 3. Известно распределение вкладчиков районного отделения сберегательного банка по размеру вкладов:

Группа вкладчиков по размеру вкладов, тыс. руб.	Число вкладчиков, тыс., чел.	*Расчетная графа* Накопленная частота
$x_{i-1}-x_i$	f_i	S_i
До 500	16	16
500 – 1000	29	45
1000 – 1500	13	58
1500 – 2000	25	83
2000 – 2500	14	97
2500 и более	9	106
Итого	106	–

Рассчитайте структурные средние – моду, медиану.

Решение.

Имеем интервальный вариационный ряд.

1. Мода для интервального ряда с равными интервалами рассчитывается по формуле:

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

где x_{Mo} – начало модального интервала,

h_{Mo} – длина модального интервала,

f_{M_o} – частота модального интервала,

f_{Mo-1} – частота домодального интервала,

f_{Mo+1} – частота послемодального интервала.

Модальный интервал определяем по наибольшей частоте $f_i = 29$, тогда имеем

$$\mathrm{M_o} = 500 + \frac{29-16}{(29-16)+(29-13)} \cdot 500 = \frac{500 \cdot 168}{307} = 724{,}1 \text{ (тыс. руб.)},$$

следовательно, наиболее часто встречающийся размер вклада 724,1 тыс. руб.

2. Медиана интервального распределения рассчитывается по формуле:

$$Me = x_{Me} + \frac{\frac{\sum f_i}{2} - S_{Me-1}}{f_{Me}} \cdot h_{Me},$$

где x_{Me} – начало медианного интервала,

h_{Me} – длина медианного интервала,

S_{Me-1} – накопленная частота домедианного интервала,

f_{Me} – частота медианного интервала.

Определим медианный интервал: $\frac{1}{2}\sum f_i = 53$, следовательно, это третий интервал (1000,1500). Тогда

$$\mathrm{M_e} = 1000 + \frac{53-45}{13} \cdot 500 = 1307{,}7 \text{ (тыс. руб.)},$$

Таким образом, половина вкладчиков имеют вклады менее 1307,7 тыс. руб., остальные – более 1307,7 тыс. руб.

Аудиторные задания

1. Результаты экзамена по теории статистики в одной из студенческих групп представлены в таблице:

Экзаменационные оценки	Отлично (10-8)	Хорошо (7-6)	Удовлетворительно (5-4)	Неудовлетворительно (3-1)	Итого
Число оценок	6	15	4	2	27

Найдите модальный и медианные баллы успеваемости студентов.

2. При изучении качества семян пшеницы было получено следующее распределение семян по проценту всхожести:

Процент всхожести	70	75	80	85	90	92	95	Свыше 95	Итого
Число проб, % к итогу	2	4	7	29	46	8	3	1	100

Рассчитайте моду и медиану.

3. С целью исследования качества деталей на предприятии проверена партия из 100 деталей. Результаты представлены в следующей таблице:

Вес деталей, г	До 50	50-60	60-70	70-80	80-90	90-100	100-110	Более 110	Итого
Число деталей, шт.	2	4	12	18	21	24	11	8	100

Определите моду, медиану, квартили.

4. По нижеследующим данным вычислите моду, медиану и квартили.

Группы порций торфа по влажности, %	Число проб	Группы порций торфа по влажности, %	Число проб
20 – 22	18	26 – 28	20
22 – 24	26	28 – 30	12
24 – 26	34	30 – 32	6
Итого	–	–	116

Домашние задания

1. Вычислите моду и медиану количественного состава семей города на основании следующего их распределения по числу совместно проживающих членов семьи:

Число членов семьи, чел.	2	3	4	5	6	7	Итого
Число семей, % к итогу	15	34	25	16	8	2	100

2. По нижеследующим данным моду, медиану, квартили.

Группы магазинов по размеру товарооборота, тыс. руб.	Число магазинов	Группы магазинов по размеру товарооборота, тыс. руб.	Число магазинов
До 200	12	500 – 600	15
200 – 300	14	600 – 700	7
300 – 400	18	700 – 800	6
400 – 500	23	Свыше 800	4
Итого	–	–	100

3. Данные о размере семьи работников отдела (число человек в семье) следующие:

3, 4,5, 2,3,6,4,2,5,3,4,2,1,2,3,4,4,2,2,1,4,5,2,3,1,2,1,2,5,4,4,2.

Рассчитайте среднее число человек в семье и структурные средние.

4. Имеем следующие условные данные по объему инвестиций в основной капитал предприятий области в 2015, млн руб.:

24,7; 71,4; 85,6; 36,8; 35,1; 94,0; 20,9; 37,3; 46,7; 25,8; 42,0; 73,8; 27,6;
50,6; 33,2; 60,9; 18,5; 38,2; 34,0; 36,7; 30,6; 55,3; 40,9; 22,3; 60,0; 41,5.

Рассчитать моду и медиану объема инвестиций.

ПОКАЗАТЕЛИ ВАРИАЦИИ

Многообразие, изменяемость величины признака у единиц совокупности называется **вариацией**. Показатели вариации делятся на две группы: абсолютные и относительные. К **абсолютным** относятся размах вариации, среднее линейное отклонение, дисперсия и среднее квадратическое отклонение.

Размах вариации (R) является наиболее простым измерителем вариации признака:

$$R = x_{max} - x_{min},$$

где x_{max} - наибольшее значение варьирующего признака; x_{min} - наименьшее значение признака.

Среднее линейное отклонение $(\bar{d})$ представляет собой среднюю величину из отклонений вариантов признака от их средней. Его можно рассчитать по формуле средней арифметической, как невзвешенной, так и взвешенной, в зависимости от отсутствия или наличия частот в ряду распределения:

$$\overline{d} = \frac{\sum |x_i - \overline{x}|}{n} \text{ – невзвешенное среднее линейное отклонение;}$$

$$\overline{d} = \frac{\sum |x_i - \overline{x}| f_i}{\sum f_i} \text{ – взвешенное среднее линейное отклонение.}$$

Здесь n – объем совокупности; x_i – i-й вариант признака; f_i – вес i-го варианта; $\overline{x}$ – средняя.

<u>**Пример 1.**</u> На основе данных следующей таблицы рассчитаем среднее линейное отклонение для дискретного ряда распределения.

Распределение учителей средних школ района по стажу работы

Стаж работы, лет, x_i	Число учителей в % к итогу, f_i	Дополнительные расчеты			
		$x_i \cdot f_i$	$x_i - \overline{x}$	$\lvert x_i - \overline{x}\rvert$	$\lvert x_i - \overline{x}\rvert f_i$
1	2	3	4	5	6
8	14	112	-2	2	28
9	20	180	-1	1	20
10	30	300	0	0	0
11	24	264	1	1	24
12	12	144	2	2	24
Итого	100	1000	0	–	96

Решение.

Размах вариации стажа равен: R= 12 – 8 = 4 года.

Средний стаж работы определяем по формуле средней арифметической взвешенной:

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{1000}{100} = 10 \text{ (лет).}$$

Среднее линейное отклонение стажа работы учителей средних школ района:

$$\overline{d} = \frac{\sum |x_i - \overline{x}| f_i}{\sum f_i} = \frac{96}{100} = 0{,}96 \text{ (года).}$$

Дисперсия представляет собой средний квадрат отклонений индивидуальных значений признака от их средней величины и вычисляется по формулам простой невзвешенной и взвешенной:

$$\sigma^2 = \frac{\sum (x_i - \overline{x})^2}{n} \text{ – невзвешенная;}$$

$$\sigma^2 = \frac{\sum (x_i - \overline{x})^2 f_i}{\sum f_i} \text{ – взвешенная.}$$

При вычислении дисперсии часто пользуются формулами вида:

$$\sigma^2 = \overline{x^2} - (\overline{x})^2 \text{ или } \sigma^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - \left(\frac{\sum x_i f_i}{\sum f_i}\right)^2.$$

Среднее квадратическое отклонение представляет собой корень второй степени из среднего квадрата отклонений отдельных значений признака от их средней:

$$\sigma = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n}} \text{ – невзвешенное;}$$

$$\sigma = \sqrt{\frac{\sum(x_i - \overline{x})^2 f_i}{\sum f_i}} \text{ – взвешенное.}$$

Среднее квадратическое отклонение – величина именованная, имеет размерность осредняемого признака.

Пример 2. Рассчитаем дисперсию и среднее квадратическое отклонение для ряда распределения магазинов города по товарообороту во II квартале 2008 г. (см. табл.).

Решение. При расчете показателей вариации по интервальным рядам распределения сначала определяем середины интервалов и ведем дальнейшие расчеты, рассматривая ряд середин интервалов как дискретный ряд распределения. Результаты вспомогательных расчетов содержатся в графах 2-6.

Распределение магазинов города по товарообороту во II квартале 2008г.

Группы магазинов по величине товарооборота, тыс. руб.	Число магазинов, f_i	Дополнительные расчеты				
		Середина интервала, тыс. руб., x_i'	$x_i' \cdot f_i$	$x_i' - \overline{x}$	$(x_i' - \overline{x})^2$	$(x_i' - \overline{x})^2 f_i$
А	1	2	3	4	5	6
40–50	2	45	90	-49,2	2420,64	4841,28
50–60	4	55	220	-39,2	1536,64	6146,56
60 –70	7	65	455	-29,2	852,64	5968,48
70 – 80	10	75	750	-19,2	368,64	3686,40
80–90	15	85	1275	-9,2	84,64	1269,60
90 – 100	20	95	1900	0,8	0,64	12,80
100– 110	22	105	2310	10,8	116,64	2566,08
110– 120	11	115	1265	20,64	432,64	4759,04
120 – 130	6	125	750	30,8	948,64	5691,84
130 – 140	3	135	405	40,8	1664,64	4993,92
Итого	100	0	9420	–	–	39936,00

Средний размер товарооборота определяется по средней арифметической взвешенной и составляет:

$$\overline{x} = \frac{\sum x_i' f_i}{\sum f_i} = \frac{9420}{100} = 94{,}2 \text{ (тыс. руб.)}.$$

Дисперсия товарооборота:

$$\sigma^2 = \frac{\sum(x_i' - \overline{x})^2 f_i}{\sum f_i} = \frac{39936}{100} = 399{,}36 \text{ (тыс. руб.)}.$$

Среднее квадратическое отклонение товарооборота:

$$\sigma = \sqrt{399{,}36} \approx 20 \text{ (тыс. руб.)}.$$

Относительные показатели вариации вычисляются для целей сравнения колеблемости различных признаков в одной и той же совокупности или же при сравнении колеблемости одного и того же признака в нескольких совокупностях. Базой для сравнения служит средняя арифметическая. Чаще всего они выражаются в процентах и характеризуют не только сравнительную оценку вариации, но и дают характеристику однородности совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 33% (для распределений, близких к нормальному). Различают следующие относительные показатели вариации (V):

Коэффициент осцилляции	**Линейный коэффициент вариации**	**Коэффициент вариации**
$V_R = \frac{R}{\bar{x}} \cdot 100\%$	$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\%$	$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%$

Наиболее часто в практических расчетах из этих трех показателей применяется коэффициент вариации.

Рассмотрим примеры определения этих показателей.

По данным примера 1, коэффициент осцилляции и линейный коэффициент вариации соответственно равны:

$$V_R = \frac{4}{10} \cdot 100\% = 40{,}0\% \text{ и } V_{\bar{d}} = \frac{0{,}96}{10} \cdot 100\% = 9{,}6\%.$$

То есть разница между крайними значениями на 40% превышает средний стаж работы учителей и доля усредненного значения абсолютных отклонений стажа от его средней величины составляет 9,6%.

Коэффициент вариации вычислим на основе ряда распределения, представленного в таблице примера 2:

$$V_\sigma = \frac{19{,}98}{94{,}2} \cdot 100\% = 21{,}2\% < 33\%.$$

Значение коэффициента вариации свидетельствует о том, что совокупность достаточно однородна.

Наряду с вариацией количественных признаков часто необходимо определить вариацию качественных или альтернативных признаков.

Альтернативным называется признак, который принимает только два значения: наступление или не наступление события. Условно считается, что альтернативный признак принимает значение, равное 1, если событие наступило, т. е. обусловливаемая единица обладает данным признаком, и равно 0, если событие не наступило.

Удельный вес значений признака, для которых событие наступило, обозначим p, а для которых не наступило – q. Тогда p+q=1.

Средняя арифметическая равна:

$$\bar{x} = \frac{1 \cdot p + 0 \cdot q}{p + q} = p.$$

Дисперсия альтернативного признака:

$$\sigma^2 = pq = p(1 - p).$$

Пример 3. В партии готовой продукции имеется 5% нестандартной. Определим дисперсию стандартной продукции.

Решение. В партии готовой продукции доля стандартной продукции составила 95%, нестандартной 5%. Следовательно, средняя доля стандартной продукции $\bar{x}=p=0{,}95$, а дисперсия $\sigma^2=0{,}95\cdot0{,}05=0{,}0475$.

Аудиторные задания

1. Распределение студентов одного из факультетов по возрасту характеризуется следующими данными:

Возраст студентов, лет	17	18	19	20	21	22	23	24	Всего
Число студентов, чел.	20	80	90	110	130	170	90	60	750

Вычислите: а) размах вариации; б) среднее линейное отклонение; в) дисперсию; г) среднее квадратическое отклонение; д) относительные показатели вариации возраста студентов.

2. Определите среднюю длину пробега автофургона торгово-посреднической фирмы и вычислите все показатели вариации, если известны:

Пробег за один рейс, км	30-50	50-70	70-90	90-110	110-130	130-150	Всего
Число рейсов за квартал, шт.	20	25	14	18	9	6	92

3. Имеется следующий ряд распределения телеграмм, принятых отделением связи, по числу слов:

Количество слов в телеграмме	12	13	14	15	16	17	18	Итого
Число телеграмм, шт.	18	22	34	26	20	13	7	140

Рассчитайте абсолютные и относительные показатели вариации.

4. Акционерные общества области по среднесписочной численности работающих на 1 января 2013 г. распределялись следующим образом:

Группы АО по численности работающих, чел.	Количество АО, шт.
до 400	11
400 – 600	23
600 – 800	36
800 – 1000	42
1000 – 1200	28
1200 – 1400	17
1400 – 1600	9
1600 – 1800	4
Итого	170

Рассчитайте: а) среднее линейное отклонение; б) дисперсию; в) среднее квадратическое отклонение; г) коэффициент вариации. Сделайте выводы.

5. Средняя величина признака в совокупности равна 19, а средний квадрат индивидуальных значений этого признака – 397. Определите коэффициент вариации.

6. Дисперсия признака равна 9, средний квадрат индивидуальных его значений – 130. Чему равна средняя?

7. Средняя величина в совокупности равна 16, среднее квадратическое отклонение – 8. Определите средний квадрат индивидуальных значений этого признака.

8. В отчетном периоде распределение работников по продолжительности отпусков характеризовалось показателями:

Продолжительность отпуска, дней	Численность работников, % к итогу	
	Рабочие	ИТР
18	12	–
20	20	4
22	25	6
23	18	16
24	15	58
26	10	13
30	–	3
Итого	100	100

Определите для каждой группы работников все абсолютные и относительные показатели вариации. Сравните полученные значения по группам работников и сделайте вывод.

9. Если дисперсия равна 20 000 единицам, а коэффициент вариации – 30%, то каков будет средний квадрат отклонений индивидуальных значений признака от величины, равной 250 единицам?

10. Доля неуспевающих студентов на II курсе составляет 10%. Определите дисперсию доли неуспевающих студентов.

11. Определите дисперсию и среднее квадратическое отклонение альтернативного признака, если общая численность новорожденных составила 184, в том числе 104 мальчика.

12. В студенческой группе по списку числится 27 человек, в том числе 18 девушек. Вычислите среднюю долю и дисперсию девушек в группе.

13. В классе из 30 учеников 4 являются неуспевающими. Определите дисперсию и среднее квадратическое отклонение альтернативного признака.

14. В результате контроля качества при приемке 500 готовых изделий 50 оказались бракованными. Определите дисперсию и среднее квадратическое отклонение альтернативного признака.

15. Урожай яблок с одной яблони характеризуется следующими данными:
$\bar{x} = 150$ кг, $\overline{x^2}$=22600. Определите среднеквадратическое отклонение.

16. Средний тарифный разряд рабочих завода составил 3,8 при $\overline{\sigma^2}$=0,903; средняя заработная плата 1200 тыс. руб. при σ^2 =64000. Если возможно, сравните вариацию заработной платы с вариацией тарифного разряда и обоснуйте ответ.

Домашние задания

1. По результатам летней сессии 2014 г. знания студентов по статистике всех форм обучения были оценены следующим образом:

Балл	1	2	3	4	5	6	7	8	9	10	Всего
Число студентов, чел.	27	28	30	35	38	39	37	27	15	9	285

Определить абсолютные и относительные показатели данной вариации: размах вариации, среднее линейное отклонение, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Сделайте выводы.

2. По данным статистического наблюдения получено статистическое распределение.

Выполнение норм выработки, %	Число работающих, чел.
до 100	18
100-110	66
110-120	144
120-130	53
130-140	10
140 и более	9

Для характеристики вариации определите:

1) среднее линейное отклонение;
2) среднее квадратическое отклонение;
3) коэффициент вариации.

Сделайте выводы.

ПРАВИЛО СЛОЖЕНИЯ ДИСПЕРСИЙ

Наряду с изучением вариации признака по всей совокупности в целом часто бывает необходимо проследить количественные изменения признака по группам, на которые разделяется совокупность, а также и между группами. Такое изучение вариации достигается посредством вычисления и анализа различных видов дисперсии. Если данные представлены в виде аналитической группировки, то можно вычислить дисперсию **общую, межгрупповую и внутригрупповую.**

Общая дисперсия измеряет вариацию признака во всей совокупности под влиянием всех факторов, обусловливающих эту вариацию:

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

Межгрупповая дисперсия характеризует систематическую вариацию, т. е. различия в величине изучаемого признака, возникающие под влиянием признака-фактора, положенного в основание группировки. Она рассчитывается по формуле:

$$\delta_x^2 = \frac{\sum(\overline{x_i} - \overline{x})^2 n_i}{\sum n_i},$$

где $\overline{x_i}$ и n_i - соответственно средние и численности по отдельным группам.

Внутригрупповая дисперсия отражает случайную вариацию, т. е. часть вариации, происходящую под влиянием неучтенных факторов и не зависящую от признака-фактора, положенного в основание группировки. Она исчисляется следующим образом:

$$\sigma_i^2 = \frac{\sum\left(x_j - \overline{x_i}\right)^2 \cdot f_j}{n_i}.$$

Средняя из внутригрупповых дисперсий:

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 \cdot n_i}{\sum n_i}.$$

Существует закон, связывающий три вида дисперсий. Общая дисперсия равна сумме средней из внутригрупповых и межгрупповых дисперсий:

$$\sigma^2 = \overline{\sigma_i^2} + \delta_x^2.$$

Данное соотношение называют **правилом сложения дисперсий.** Согласно этому правилу общая дисперсия, возникающая под влиянием всех факторов, равна сумме дисперсий, возникающих под влиянием всех прочих факторов, и дисперсии, возникающей за счет группировочного признака.

Зная любые два вида дисперсий, можно определить или проверить правильность расчета третьего вида.

Пример 1. Определим групповые дисперсии, среднюю из групповых дисперсий, межгрупповую дисперсию, общую дисперсию по данным следующей таблицы:

Производительность труда двух бригад рабочих-токарей

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

Решение. Для расчета групповых дисперсий вычислим среднюю по каждой группе:

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

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

$$\sigma_1^2 = \frac{\sum (x_j - \bar{x}_1)^2}{n_1} = \frac{10}{6} = 1{,}666 \approx 1{,}67; \; \sigma_2^2 = \frac{\sum (x_j - \bar{x}_2)^2}{n_2} = \frac{28}{6} \approx 4{,}67.$$

Средняя из групповых дисперсий

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 \cdot n_i}{\sum n_i} = \frac{1{,}67 \cdot 6 + 4{,}67 \cdot 6}{12} = \frac{10 + 28}{12} = \frac{38}{12} \approx 3{,}17.$$

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

$$\bar{x} = \frac{\sum \bar{x}_i f_i}{\sum f_i} = \frac{15 \cdot 6 + 21 \cdot 6}{12} = \frac{90 + 126}{12} = \frac{216}{12} = 18 \text{ (шт./час)}.$$

Теперь определим межгрупповую дисперсию:

$$\delta_x^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 \cdot n_i}{\sum n_i} = \frac{(15-18)^2 \cdot 6 + (21-18)^2 \cdot 6}{12} = \frac{9 \cdot 6 + 9 \cdot 6}{12} = 9.$$

Таким образом, общая дисперсия по правилу сложения дисперсий

$$\sigma^2 = \overline{\sigma_i^2} + \delta_x^2 = 3{,}17 + 9 = 12{,}17.$$

Проверим полученный результат, исчислив общую дисперсию обычным способом:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{(13-18)^2 + (14-18)^2 + ... + (23-18)^2}{12} = \frac{146}{12} \approx 12,17.$$

На основании правила сложения дисперсий можно определить показатели тесноты связи между группировочным (факторным) и результативным признаками. **Эмпирическое корреляционное отношение** (η) измеряет, какую часть общей колеблемости результативного признака вызывает изучаемый фактор. Соответственно оно рассчитывается как отношение факторной дисперсии к общей дисперсии результативного признака:

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma^2}}$$ **– эмпирическое корреляционное отношение.**

Этот показатель принимает значения в интервале[0,1]: чем ближе к 1, тем теснее связь, и наоборот. Величина корреляционного отношения будет равна нулю, когда нет колеблемости в величине средних по выделенным группам. В тех случаях, когда внутригрупповая дисперсия близка к нулю, т. е. практически вся вариация результативного признака обусловлена действием фактора X, величина корреляционного отношения близка к 1.

Для нашего примера эмпирическое корреляционное отношение $\eta = \sqrt{\frac{9}{12,17}} \approx 0,86$.

Величина 0,86 характеризует существенную связь между группировочным и результативным признаками.

Коэффициент детерминации рассчитывается как возведенное в квадрат эмпирическое корреляционное отношение

$$\eta^2 = \frac{\delta_x^2}{\sigma^2}$$ **– коэффициент детерминации.**

Он характеризует долю общей колеблемости результативного признака, вызванную действием факторного признака, положенного в основу группировки.

Наряду с вариацией индивидуальных значений признака вокруг средней может наблюдаться и вариация индивидуальных долей признака вокруг средней доли. Такое изучение вариации достигается посредством вычисления и анализа следующих видов дисперсий.

Внутригрупповая дисперсия доли определяется по формуле

$$\sigma_{p_i}^2 = p_i \cdot (1 - p_i).$$

Средняя из внутригрупповых дисперсий

$$\overline{\sigma_{p_i}^2} = \overline{p_i \cdot (1 - p_i)} = \frac{\sum p_i \cdot (1 - p_i) \cdot n_i}{\sum n_i}.$$

Формула **межгрупповой дисперсии** имеет вид:

$$\delta_{p_i}^2 = \frac{\sum (p_i - \bar{p})^2 \cdot n_i}{\sum n_i},$$

где n_i – численность единиц в отдельных группах; $\bar{p}$ – доля изучаемого признака во всей совокупности, которая определяется по формуле

$$\overline{p} = \frac{\sum p_i \cdot n_i}{\sum n_i}.$$

Общая дисперсия имеет вид:

$$\sigma_p^2 = \overline{p} \cdot (1 - \overline{p}).$$

Три вида дисперсии связаны между собой следующим образом:

$$\sigma_p^2 = \overline{\sigma_{p_i}^2} + \delta_{p_i}^2.$$

Данное соотношение дисперсий называется теоремой сложения дисперсии доли признака. Эта теорема широко используется в изучении колеблемости качественных признаков.

Пример 2. Определим групповые дисперсии, среднюю из групповых, межгрупповую и общую дисперсии по следующим данным:

Численность и удельный вес одной из категорий птиц фермерских хозяйств района

Хозяйство	Удельный вес кур, %, p_i	Всего кур, шт., n_i
1	90	50
2	95	20
3	80	30
Итого	265	100

Решение. Определим долю кур в целом по трем хозяйствам:

$$\overline{p} = \frac{0,90 \cdot 50 + 0,95 \cdot 20 + 0,80 \cdot 30}{100} = \frac{88}{100} = 0,88.$$

Общая дисперсия доли птиц:

$$\sigma_p^2 = 0,88 \cdot (1 - 0,88) = 0,1056.$$

Внутригрупповые дисперсии:

$$\sigma_{p_1}^2 = 0,90 \cdot (1 - 0,90) = 0,09,$$

$$\sigma_{p_2}^2 = 0,95 \cdot (1 - 0,95) = 0,0475, \ \sigma_{p_3}^2 = 0,80 \cdot (1 - 0,80) = 0,16.$$

Средняя из внутригрупповых дисперсий:

$$\overline{\sigma_{p_i}^2} = \frac{0,09 \cdot 50 + 0,0475 \cdot 20 + 0,16 \cdot 30}{100} = \frac{10,25}{100} = 0,1025.$$

Межгрупповая дисперсия:

$$\delta_{p_i}^2 = \frac{(0,9 - 0,88)^2 \cdot 50 + (0,95 - 0,88)^2 \cdot 20 + (0,8 - 0,88)^2 \cdot 30}{100} = \frac{0,31}{100} = 0,0031.$$

Используя правило сложения дисперсий, получаем: 0,1025+0,0031=0,1056. Пример решен правильно.

Аудиторные задания

1. Имеются данные о распределении рабочих двух участков, использующих старую и новую технологию, по уровню производительности труда.

Показатель	Производительность труда, млн руб.	Число рабочих, чел.
Работающие по старой технологии	120, 135, 145, 150, 170,175	6
Работающие по новой технологии	270, 250, 180, 210, 160, 220, 230	7

Определите все виды дисперсий. Сделайте проверку. Вычислите коэффициент детерминации и эмпирическое корреляционное отношение.

Сделайте выводы.

2. *По данным таблицы о распределении пряжи по крепости нити вычислите все виды дисперсий. Определите общую дисперсию по правилу сложения дисперсий.*

I группа пряжи (менее крепкая)		II группа пряжи (более крепкая)	
Крепость нити, г	Число проб	Крепость нити, г	Число проб
120–130	2	200–210	25
130–140	6	210–220	28
140–150	8	220–230	16
150–160	15	230–240	10
160–170	25	240–250	8
170–180	29	250–260	7
180–190	35	260–270	5
190–200	30		

3. *Имеются данные о распределении семей сотрудников финансовой корпорации по количеству детей:*

Число детей в семье, чел.	Число семей сотрудников по подразделениям		
	первое	второе	третье
0	4	7	5
1	6	10	13
2	3	3	3
3	2	1	

Вычислите: а) внутригрупповые дисперсии; б) среднюю из внутригрупповых дисперсий; в) межгрупповую дисперсию; г) общую дисперсию. Проверьте правильность произведения расчетов с помощью правила сложения дисперсий.

4. Распределение основных фондов по малым предприятиям отрасли характеризуется следующими данными:

Группы предприятий по стоимости основных фондов, млн руб.	Число предприятий	Основные фонды в среднем на предприятии, млн руб.	Групповые дисперсии
12–27	18	18	1,14
27–42	40	32	1,09
42–57	26	48	1,69
57–72	12	69	1,84

Рассчитайте коэффициент детерминации и эмпирическое корреляционное отношение. Сделайте выводы.

5. Распределение стоимости продукции, предназначенной для экспортных поставок, по цехам предприятия представлено следующими данными:

Цех	Стоимость всей произведенной продукции, тыс. руб.	В том числе стоимость экспортной продукции, тыс. руб.
1	340	110
2	290	140
3	180	180
Итого	810	430

Вычислите:

1) среднюю из внутригрупповых, межгрупповую и общую дисперсии доли экспортной продукции;
2) коэффициент детерминации и эмпирическое корреляционное отношение.

Сделайте выводы.

6. Средняя часовая выработка 10 рабочих бригады составила 18 деталей. При этом средняя часовая выработка 4 рабочих составила 15 деталей, а остальных 6 рабочих – 20 деталей. Среднеквадратическое отклонение часовой выработки равно 3,2 детали. Определите:

1) среднюю из внутригрупповых дисперсий;
2) коэффициент детерминации.

7. Имеются следующие данные:

Магазины	Число магазинов	Средний дневной оборот, млн руб.	Дисперсия дневного оборота
Специализированные	20	30	14
Неспециализированные	8	20	18

Определите общую дисперсию дневного оборота в магазинах торга и эмпирическое корреляционное отношение. Сделайте выводы.

8. Среднеквадратическое отклонение дневной выработки первой группы ткачей (40 человек) составило 15 м, второй группы (60 человек) – 10 м. Общая дисперсия равна 400. Определите межгрупповую дисперсию.

Домашние задания

1. Количество изготавливаемых за смену одинаковых деталей рабочими 4-го и 5-го разрядов характеризуется следующими данными:

Порядковый номер рабочего	Количество изготавливаемых деталей, шт.	
	рабочими 4-го разряда	рабочими 5-го разряда
1	7	9
2	8	9
3	9	11
4	10	10
5	–	12
6	–	14

Определите:

1) общую среднюю и групповые средние величины выработки деталей для рабочих 4-го, 5-го разрядов;

2) общую и внутригрупповые дисперсии выработки деталей;
3) среднюю из внутригрупповых дисперсий;
4) межгрупповую дисперсию.

Проверьте правило сложения дисперсий. Объясните смысл рассчитанных величин. Рассчитайте коэффициент детерминации и эмпирическое корреляционное отношение. Сделайте выводы.

2. Имеются следующие данные по одному предприятию:

Группы рабочих по организации труда	Дневная выработка, шт.	Число рабочих, чел.
Индивидуальная	100	2
	110	4
	115	6
Бригадная	120	5
	127	12
	135	3

Определите:

1) общую и групповые средние величины дневной выработки деталей в группах по видам организации труда;
2) общую и внутригрупповые дисперсии выработки деталей;
3) среднюю из внутригрупповых дисперсий;
4) межгрупповую дисперсию.

Проверьте правило сложения дисперсий. Объясните смысл рассчитанных величин. Рассчитайте коэффициент детерминации и эмпирическое корреляционное отношение. Сделайте выводы.

3. Имеются следующие данные, характеризующие фермерские хозяйства региона:

Группы хозяйств по стоимости удобрений на 1 га зерновых, тыс. руб.	Число хозяйств	Средняя урожайность, ц/га	Дисперсия урожайности в группе
До 1	6	27	6,25
1–2	10	30	3,61
2 и более	7	34	8,41

Определите коэффициент детерминации и эмпирическое корреляционное отношение при условии, что посевные площади под зерновыми культурами во всех хозяйствах одинаковы. Сделайте выводы.

ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

Выборочное наблюдение – широко применяемый вид несплошного наблюдения. При проведении выборочного наблюдения обследуются не все единицы изучаемого объекта (не все единицы **генеральной совокупности**), а лишь некоторая, так или иначе отобранная в случайном порядке часть этих единиц. Однако наблюдение организуется таким образом, что эта часть отобранных единиц в уменьшенном масштабе представляет (репрезентирует) всю совокупность. Часть единиц генеральной совокупности, подлежащей непосредственному наблюдению, называют **выборочной совокупностью**.

По результатам выборочного наблюдения можно оценить искомые параметры генеральной совокупности. Между характеристиками выборочной совокупности и искомыми характеристиками генеральной совокупности существует расхождение, которое называют **ошибкой выборочного наблюдения**. Расчет ошибок позволяет решить одну из главных проблем организации выборочного наблюдения – оценить репрезентативность (представительность) выборочной совокупности. Различают среднюю и предельную ошибки выборки. Эти два вида ошибок связаны следующим соотношением:

$$\Delta = t\mu,$$

где Δ – предельная ошибка выборки; μ – средняя ошибка выборки; t – коэффициент доверия, определяемый в зависимости от уровня вероятности.

Величина средней ошибки выборки рассчитывается в зависимости от способа отбора и процедуры выборки.

По способу организации различают следующие виды выборки:

1) собственно-случайная;
2) механическая (например, каждый 10, 20 и т. д.);
3) *типическая, когда генеральная совокупность разбита на группы и в каждой* группе обследуются по нескольку объектов));
4) серийная, когда случайным образом отбираются целые серии.

Наиболее простой способ формирования выборочной совокупности – собственно случайный отбор. При **простой случайной выборке** отбор производится из всей генеральной совокупности без предварительного расчленения ее на какие-либо группы, и единица отбора совпадает с единицей наблюдения.

Собственно случайный отбор может быть **повторным и бесповторным**. При повторном отборе каждая единица, отобранная в случайном порядке из генеральной совокупности, *после проведения наблюдения возвращается в эту совокупность и может быть* вновь подвергнута обследованию. При бесповторном отборе обследованные единицы в генеральную совокупность не возвращаются и не могут быть обследованы повторно.

Теоретические основы выборочного метода, первоначально разработанные для собственно случайного отбора, используют и для определения ошибок выборки при других способах наблюдения. Так, **средняя ошибка выборки**

при случайном повторном отборе: $\mu = \dfrac{\sigma}{\sqrt{n}}$;

при случайном бесповторном отборе: $\mu = \sqrt{\dfrac{\sigma^2}{n}\left(1 - \dfrac{n}{N}\right)}$,

где σ^2 – выборочная (или генеральная) дисперсия; σ – выборочное (или генеральное) среднее квадратическое отклонение; n – объем выборочной совокупности; N – объем генеральной совокупности.

Расчет средней и предельной ошибок выборки позволяет определить возможные пределы, в которых будут находиться характеристики генеральной совокупности. Так, **пределы для выборочной средней** устанавливаются на основе соотношений:

$$\tilde{x}-\Delta_{\tilde{x}} \le \bar{x} \le \tilde{x}+\Delta_{\tilde{x}},$$

где $\bar{x}$ и $\tilde{x}$ – генеральная и выборочная средние соответственно; $\Delta_{\tilde{x}}$ – предельная ошибка выборочной средней.

Пример 1. При проверке веса импортируемого груза на таможне методом случайной повторной выборки было отобрано 200 изделий. В результате был установлен средний вес изделия 30 г при среднем квадратическом отклонении 4 г. С вероятностью 0,997 определите пределы, в которых находится средний вес изделий в генеральной совокупности.

Решение. Рассчитаем сначала предельную ошибку выборки. Так, при доверительной вероятности $\gamma=0{,}997$ из равенства $\Phi(t)=\dfrac{\gamma}{2}=\dfrac{0{,}997}{2}=0{,}4985$ по таблице значений функции $\Phi(t)$ найдем значение t = 3. Тогда получим предельную ошибку:

$$\Delta_{\tilde{x}}=t\cdot\frac{\sigma_{\tilde{x}}}{\sqrt{n}}==3\cdot\frac{4}{\sqrt{200}}\approx 0{,}84.$$

Определим пределы генеральной средней:

$$30-0{,}84\le\bar{x}\le 30+0{,}84 \text{ или } 29{,}16\le\bar{x}\le 30{,}84.$$

Следовательно, с вероятностью 0,997 можно утверждать, что средний вес изделий в генеральной совокупности находится в пределах от 29,16 до 30,84 г.

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

Число детей в семье, чел.	0	1	2	3	4	5
Количество семей, шт.	1000	2000	12000	400	200	200

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

Решение. Вначале на основе имеющегося распределения семей определим выборочные среднюю и дисперсию:

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

$$\tilde{x}=\frac{7400}{5000}\approx 1{,}5\,(\textit{ребенка}),\ \sigma^2_{\tilde{x}}=\frac{7650}{5000}=1{,}53.$$

Вычислим теперь предельную ошибку выборки (с учетом того, что γ=0,954 $\Rightarrow t$=2):

$$\Delta_{\tilde{x}} = t \cdot \sqrt{\frac{\sigma_{\tilde{x}}^2}{n}\left(1-\frac{n}{N}\right)} = 2 \cdot \sqrt{\frac{1{,}53}{5000}\left(1-\frac{5000}{250000}\right)} \approx 0{,}035.$$

Следовательно, пределы генеральной средней: $\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}} = 1{,}53 \pm 0{,}035$.

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

Наряду с определением ошибок выборки и пределов для генеральной средней эти же показатели могут быть определены *для доли признака*. В этом случае особенности расчета связаны с определением **дисперсии доли**, которая вычисляется так:

$$\sigma_w^2 = w(1-w),$$

где $w = \frac{m}{n}$ – доля единиц, обладающих данным признаком в выборочной совокупности, определяемая как отношение количества соответствующих единиц к объему выборки.

Тогда, например, при собственно-случайном повторном и бесповторном отборах для определения предельной ошибки выборки используются следующие формулы соответственно:

$$\Delta_w = t \cdot \sqrt{\frac{\sigma_w^2}{n}} = t \cdot \sqrt{\frac{w(1-w)}{n}} \text{ и } \Delta_w = t \cdot \sqrt{\frac{\sigma_w^2}{n}\left(1-\frac{n}{N}\right)} = t \cdot \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}.$$

Следовательно, **пределы доли признака** в генеральной совокупности p:

$$w - \Delta_w \le p \le w + \Delta_w.$$

Пример 3. С целью определения средней фактической продолжительности рабочего дня в государственном учреждении с численностью служащих 480 человек в июне 2012 г. была проведена 25%-я механическая выборка. По результатам наблюдения оказалось, что у 10% обследованных потери времени достигали более 45 мин. в день. С вероятностью 0,683 *установите пределы, в которых находится генеральная доля служащих с потерями рабочего времени более 45 мин. в день.*

Решение. Определим объем выборочной совокупности: $n = 480 \cdot 0{,}25 = 120$ человек. Выборочная доля w равна по условию 10%. Учитывая, что показатели точности механической и собственно-случайной бесповторной выборки определяются одинаково, а также то, что при $\gamma = 0{,}683 \Rightarrow t = 1$, вычислим предельную ошибку выборочной доли:

$$\Delta_w = t \cdot \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} = 1 \cdot \sqrt{\frac{0{,}1(1-0{,}1)}{120}\left(1-\frac{120}{480}\right)} = 0{,}0237 \approx 0{,}024 \text{ или } 2{,}4\%.$$

Пределы доли признака в генеральной совокупности:

$$10 - 2{,}4 \le p \le 10 + 2{,}4 \text{ или } 7{,}6 \le p \le 12{,}4.$$

Таким образом, с вероятностью 0,683 можно утверждать, что доля работников учреждения с потерями рабочего времени более 45 мин. в день находится в пределах от 7,6 до 12,4%.

Ошибки и пределы генеральных характеристик при других способах формирования выборочной совокупности определяются на основе соответствующих формул, отражающих особенности этих видов выборки.

При применении выборочного наблюдения одной из основных задач является **определение объема выборки**, необходимого для получения требуемой точности результатов с заданной вероятностью. Объем выборки рассчитывается на стадии проектирования выборочного обследования. Для собственно-случайной и механической выборки необходимый объем выборки при повторном и бесповторном отборах находится соответственно по формулам:

$$n=\frac{t^2\sigma^2}{\Delta^2} \text{ и } n=\frac{t^2\sigma^2 N}{\Delta^2 N+t^2\sigma^2}.$$

Пример 4. В 100 туристических агентствах города предполагается провести обследование среднемесячного количества реализованных путевок методом механического отбора. Какова должна быть численность выборки, чтобы с вероятностью 0,683 ошибка не превышала 3 путевок, если по данным пробного обследования дисперсия составляет 225?

Решение. Рассчитаем необходимый объем выборки:

$$n=\frac{1^2\cdot 225\cdot 100}{3^2\cdot 100+1^2\cdot 225}=\frac{22500}{1125}=20 \text{ агентств.}$$

Наиболее часто употребляемые уровни доверительной вероятности и соответствующие значения t для выборок достаточно большого объема ($n\geq 30$):

t	1,00	1,96	2,00	2,58	3,00
$F(t)$	0,683	0,950	0,954	0,990	0,997

Аудиторные задания

1. Что произойдет с величиной предельной ошибки выборки, если вероятность, гарантирующую результат: а) увеличить с 0,954 до 0,997; б) уменьшить с 0,954 до 0,683: в) увеличить с 0,683 до 0,954; г) уменьшить с 0,997 до 0,954; д) увеличить с 0,683 до 0,997?

2. Определите, как изменится средняя ошибка случайной выборки, если необходимую численность выборочной совокупности: а) уменьшить в 2,5 раза; на 40%; б) увеличить в 1,5 раза; на 20%. Как нужно применить необходимую численность выборки, чтобы средняя ошибка уменьшилась в 2 раза; на 50%; на 30%?

3. Какой должна быть необходимая численность выборки при механическом отборе, чтобы установить генеральную долю с ошибкой не более 2%, если дисперсия доли неизвестна, а отбор производится из совокупности, включающей: а) 1000 единиц; б) 10 000 единиц; в) 100 000 единиц? Вероятность, гарантирующая результаты выборочного наблюдения, равна 0,954.

4. Каким должен быть объем случайной бесповторной выборки из генеральной совокупности численностью 10000 единиц при среднем квадратическом отклонении не более 20, предельной ошибке, не превышающей 5%, и вероятности 0,997?

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

6. Из партии в 1 млн шт. мелкокалиберных патронов путем случайного отбора взято для определения дальнобойности боя 1000 шт. Результаты испытаний представлены в следующей таблице:

Дальность боя, м	25	30	35	40	45	50	Итого
Число патронов, шт.	120	180	280	170	140	110	1000

С вероятностью 0,954 определите среднюю дальность боя по выборке, ошибку выборки и возможные пределы средней дальности боя для всей партии патронов.

7. В порядке механической выборки обследован возраст 100 студентов вуза из общего числа 2000 человек. Результаты обработки материалов наблюдения приведены в таблице:

Возраст, лет	17	18	19	20	21	22	23
Число студентов, чел.	11	13	18	23	17	10	8

Установите: а) средний возраст студентов вуза по выборке; б) величину ошибки при определении возраста студентов на основе выборки; в) вероятные пределы колебания возраста для всех студентов при вероятности 0,997.

8. Для изучения заработной платы работников завода строительных материалов из общей численности 5400 человек были отобраны в порядке собственно случайной бесповторной выборки 600 человек и зафиксированы следующие результаты:

Месячная зарплата, руб.	Число рабочих, чел.
от 15000 до 25000	20
от 25000 до 35000	79
от 35000 до 45000	130
от 45000 до 55000	150
от 55000 до 65000	110
от 65000 до 75000	100
от 75000 до 85000	11

Определите:

1) среднюю выборочную заработную плату;
2) среднюю ошибку выборочной средней;
3) предельную ошибку выборочной средней с вероятностью 0,96 и укажите интервал, в котором находится генеральная средняя;
4) долю рабочих в выборке с месячной зарплатой до 30000 руб.; свыше 72000 руб. и укажите интервалы, в которых находятся эти генеральные доли.

9. В процессе технического контроля из партии готовой продукции методом случайного бесповторного отбора было проверено 70 изделий, из которых 4 оказались бракованными. Можно ли с вероятностью 0,954 утверждать, что доля бракованных изделий во всей партии не превышает 7%, если процент отбора равен 10?

10. С целью определения среднего размера вклада в отделениях Сбербанка города предполагается провести механическую выборку лицевых счетов из общего числа 67800. По данным предыдущего обследования установлено среднее квадратическое отклонение размера вклада, равное 140 тыс. руб. С вероятностью 0,997 определите необходимый объем выборочной совокупности при условии, что ошибка выборки не превысит 10 тыс. руб.

11. Определите, сколько персональных компьютеров следует подвергнуть обследованию в порядке случайной бесповторной выборки, чтобы с вероятностью 0,954 предельная ошибка (в процентах к среднему сроку службы компьютера) не превышала 3%.

Коэффициент вариации среднего срока службы компьютеров по данным предыдущих обследований составляет 15%, а вся партия состоит из 1250 компьютеров.

12. В процессе подготовки выборочного обследования качества импортируемых кондитерских изделий была проведена пробная проверка 8 ящиков для сбора данных о вариации их веса. Результаты проверки представлены в следующей таблице:

№ ящика	1	2	3	4	5	6	7	8
Средний вес коробки в ящике, г	540	520	550	500	510	530	560	520

Сколько ящиков с кондитерскими изделиями необходимо отобрать для проверки качества в порядке бесповторного отбора, чтобы с вероятностью 0,997 ошибка выборки не превышала 20 г, если генеральная совокупность включает 1000 равных по величине серий?

Домашние задания

1. Для определения зольности угля месторождения в порядке случайной выборки взято 400 проб. В результате исследования установлена средняя зольность угля в выработке 16% при среднем квадратическом отклонении 4%. С вероятностью 0,997 определите пределы, в которых находится средняя зольность угля.

2. Для определения среднего размера вклада определенной категории вкладчиков в банках города, где число вкладчиков 5 000, необходимо провести выборку лицевых счетов методом механического отбора. Предварительно установлено, что среднее квадратическое отклонение размера вкладов составляет 120 у. е. Определите необходимую численность выборки при условии, что с вероятностью 0,954 ошибка выборки не превысит 10 у.е.

3. На заводе с числом рабочих 1000 человек было проведено 2%-е выборочное обследование возраста рабочих методом случайного бесповторного отбора. В результате обследования получены следующие данные:

Возраст рабочих, лет	До 30	30-40	40-50	50-60	Свыше 60
Число рабочих, чел.	8	22	10	6	4

С вероятностью 0,997 определите пределы, в которых находится средний возраст рабочих завода.

4. Научно-исследовательским институтом для изучения общественного мнения населения области о проведении определенных мероприятий в порядке случайного повторного отбора было опрошено 600 человек. Из числа опрошенных 360 человек одобрили мероприятия. С вероятностью 0,997 определите пределы, в которых находится доля лиц, одобривших мероприятия.

5. В городе N с числом семей 10 тыс. предполагается методом случайного бесповторного отбора определить долю семей с детьми ясельного возраста. Какова должна быть численность выборки, чтобы с вероятностью 0,954 ошибка выборки не превышала 0,04, если на основе предыдущих обследований известно, что дисперсия равна 0,24?

ЛИТЕРАТУРА

1. Елисеева, И.И. Общая теория статистики: учебник / И.И. Елисеева, М.М. Юзбашев; под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2005.
2. Ефимова, М.Р. Общая теория статистики: учебник / М.Р. Ефимова, Е.В. Петрова, В.Н. Румянцев – М.: ИНФРА – М, 2005.
3. Неганова, Л.М., Шевелева Ю.Г., Замедлина Е.А. Экзамен по статистике: Учебное пособие для вузов – М.: Приор-издат, 2006. – 144 с.
4. Практикум по общей теории статистики. Для студентов экономических специальностей. – Брест: Издательство БрГТУ, 2010. – Ч. 1.
5. Практикум по общей теории статистики: учебно-методическое пособие / Под ред. М.Г. Назарова. – М.: КНОРУС,2008. – 184 с.
6. Руденко, В.И. Статистика: Пособие студентам для подготовки к экзаменам / В.И. Руденко. – 6-е изд., перераб. и доп. – М.: Издательско-торговая корпорация «Дашков и К^0», 2010. – 188 с.
7. Социальная статистика: Учебник / Под ред. И.И. Елисеевой. – 2-е изд., доп. – М., 1999.
8. Социальная статистика: Учебник / Под ред. М.Г. Назарова. – М., 1988.
9. Социально-экономическая статистика / Под ред. Г.Л. Громыко. – М., 1997.
10. Теория статистики / Под ред. проф. Г.Л. Громыко – М.: ИНФРА – М., 2005.

СОДЕРЖАНИЕ

Учебное издание

Составители:
Журавель Мария Григорьевна
Золотухина Лада Станиславовна
Копайцева Татьяна Владимировна
Кузьмина Елена Викторовна
Шамовская Галина Владимировна

Практикум по статистике

для студентов экономических специальностей

ЧАСТЬ I

Ответственный за выпуск: Кузьмина Е.В.
Редактор: Боровикова Е.А.
Компьютерная верстка: Кузьмина Е.В.
Корректор: Никитчик Е.В.

Подписано к печати 6.09.2016 г. Формат 60x84 1/16. Бумага «Снегурочка». Усл. п.л. 2,56. Усл. изд. л. 2,75. Заказ 884. Тираж 120 экз. Отпечатано на ризографе учреждения образования «Брестский государственный технический университет». 224017, г. Брест, ул. Московская, 267.