

А. И. Богданов
Л. Н. Никитина
Е. А. Крайкина

СТАТИСТИКА

Министерство науки и высшего образования Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«Санкт-Петербургский государственный университет
промышленных технологий и дизайна»**

А. И. Богданов Л. Н. Никитина Е. А. Крайкина

СТАТИСТИКА

Утверждено Редакционно-издательским советом
университета в качестве учебного пособия

Санкт-Петербург
2023

УДК 519.22 (075.8)
ББК 60.6:22.172я73
Б73

Рецензенты:

кандидат экономических наук, доцент Санкт-Петербургского государственного
университета промышленных технологий и дизайна

Б. П. Касаткин;

кандидат социологических наук, доцент Петербургского государственного
университета путей сообщения Императора Александра I

Л. А. Кравченко

Богданов, А. И.

Б73 Статистика: учебное пособие / А. И. Богданов, Л. Н. Никитина, Е. А. Крайкина. – Санкт-Петербург: ФГБОУВО «СПбГУПТД», 2023. – 75 с.
ISBN 978-5-7937-2370-1

В учебном пособии рассмотрены статистические методы обработки экономической информации, включая и некоторые методы математической статистики, виды и способы построения группировок статистических данных, понятия математического ожидания, моды и медианы как характеристик закона распределения случайной величины, общие принципы оценки параметров законов распределения случайных величин и проверки статистических гипотез, особенности оценки математического ожидания и доли по выборочным данным при повторном и бесповторном отборе, вопросы определения выборочных статистических показателей по сгруппированным данным, особенности вычисления средних значений для относительных показателей, вопросы применения индексов в экономическом анализе. Теоретические положения проиллюстрированы многочисленными практическими примерами.

Предназначено для изучения бакалаврами экономических направлений дисциплины «Статистика».

УДК 519.22 (075.8)
ББК 60.6:22.172я73

ISBN 978-5-7937-2370-1

© ФГБОУВО «СПбГУПТД», 2023

© Богданов А. И., 2023

© Никитина Л. Н., 2023

© Крайкина Е. А., 2023

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
1. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ. СБОР, ГРУППИРОВКА И ОТОБРАЖЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ.....	5
1.1. Предмет, основные понятия и методология статистики.....	5
1.2. Организация и виды статистического наблюдения.....	5
1.3. Группировка статистических данных.....	6
1.4. Понятие о рядах распределения. Полигон и гистограмма.....	14
ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ.....	17
2. МЕТОДОЛОГИЯ СТАТИСТИКИ. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ.....	20
2.1. Понятие о случайных событиях и их вероятностях.....	20
2.2. Понятие о случайных величинах и их характеристиках.....	29
2.3. Основные задачи статистики.....	40
ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ.....	44
3. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ХАРАКТЕРИСТИК ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ (МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ, ДИСПЕРСИИ, МОДЫ, МЕДИАНЫ, ДОЛИ) ПО ВЫБОРКЕ.....	47
3.1. Свойства выборочной средней и частоты.....	47
3.2. Дисперсия оценок математического ожидания и доли объектов в генеральной совокупности при повторном и бесповторном отборах.....	48
3.3. Определение выборочных статистических показателей по сгруппированным данным.....	51
3.4. Абсолютные и относительные статистические показатели. Вычисление средних значений относительных показателей.....	54
ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ.....	58
4. ПРИМЕНЕНИЕ ИНДЕКСОВ В ЭКОНОМИЧЕСКОМ АНАЛИЗЕ.....	63
4.1. Индивидуальные и агрегатные индексы.....	63
4.2. Взаимосвязь агрегатных и индивидуальных индексов.....	64
4.3. Индексы постоянного состава и структурных сдвигов.....	65
ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ.....	66
5. ЗАВИСИМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ.....	67
5.1. Система случайных величин.....	67
5.2. Условные законы распределения.....	71
5.3. Зависимые и независимые случайные величины. Коэффициент корреляции.....	71
ЗАКЛЮЧЕНИЕ.....	74
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	75

ВВЕДЕНИЕ

Учебная дисциплина «Статистика» занимает одно из центральных мест в системе подготовки экономистов и менеджеров и ставит своей целью сформировать компетенции в области основных положений теории статистики, научного подхода при проведении прикладных статистических исследований, а также выработать практические навыки решения конкретных статистических задач в сфере экономики.

При составлении учебного пособия использован многолетний опыт чтения лекций и соответствующие материалы по дисциплине «Статистика» проф. Л. Н. Никитиной, а также некоторые разделы математической статистики, читаемой в прошлом проф. А. И. Богдановым. В подготовке практических заданий активное участие принимала канд. техн. наук Е. А. Крайкина.

В учебном пособии рассмотрены статистические методы обработки экономической информации, включая и некоторые методы математической статистики, виды и способы построения группировок статистических данных, понятия математического ожидания, моды и медианы как характеристик закона распределения случайной величины, общие принципы оценки параметров законов распределения случайных величин и проверки статистических гипотез, особенности оценки математического ожидания и доли по выборочным данным при повторном и бесповторном отборе, вопросы определения выборочных статистических показателей по сгруппированным данным, особенности вычисления средних значений для относительных показателей, вопросы применения индексов в экономическом анализе. Теоретические положения проиллюстрированы многочисленными практическими примерами.

В учебном пособии не нашла отражение тема «Временные ряды», так как этот вопрос подробно рассматривается в дальнейшем в рамках дисциплины «Эконометрика», и такое дублирование представляется нецелесообразным.

1. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ. СБОР, ГРУППИРОВКА И ОТОБРАЖЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ

1.1. Предмет, основные понятия и методология статистики

Предмет статистики – количественная сторона массовых явлений, носящих случайный характер, выявление их закономерностей.

Термин «статистика» употребляется в трех смыслах:

- область деятельности, связанная со сбором, обработкой, анализом и публикацией данных о массовых явлениях;
- сам цифровой материал;
- научная дисциплина.

Базовым понятием статистики является понятие «*статистической совокупности*» как некоторого множества «*элементов совокупности*» (предприятий, людей и т. д.).

Элемент совокупности может характеризоваться различными *признаками*, а статистическая совокупность – *статистическими показателями*.

Признак представляет собой качественную или количественную характеристику элемента совокупности (пол, возраст, стаж работы и т. д.). Статистический показатель – некая характеристика статистической совокупности (численность населения, средняя заработная плата и т. д.).

Методология статистики во многом опирается на результаты, полученные в математической статистике.

1.2. Организация и виды статистического наблюдения

Статистическое изучение тех или иных явлений требует наличия информации об этих явлениях, поэтому первый этап (начало статистического исследования) заключается в сборе необходимой информации.

Научно организованный сбор сведений, заключающийся в регистрации тех или иных признаков, относящихся к каждому элементу изучаемой совокупности, называют *статистическим наблюдением*.

Если при сборе первичных статистических данных допущена ошибка, это повлияет на достоверность как теоретических, так и практических выводов. Поэтому статистическое наблюдение должно быть тщательно продуманным и четко организованным.

Статистическое наблюдение должно носить массовый характер для того, чтобы получить правдивые статистические данные, характеризующие всю совокупность в целом.

Статистическое наблюдение может быть *сплошным* и *выборочным*. *Сплошное* наблюдение предполагает полный учет всех элементов изучаемой совокупности (например, перепись населения). При *выборочном* наблюдении характеристика всей совокупности дается по результатам исследования некоторой ее части (выборки), отобранной в случайном порядке (например, опросы

перед выборами). Сама же изучаемая статистическая совокупность (множество всех элементов) называется **генеральной совокупностью**.

Выборка должна быть **репрезентативной** (представительной), чтобы правильно отражать свойства генеральной совокупности (например, не должен быть опрос только мужчин или городских жителей).

Для однородной генеральной совокупности репрезентативность гарантируется случайностью отбора.

Отбор элементов из неоднородной совокупности предполагает разделение генеральной совокупности на однородные группы с последующей случайной выборкой внутри групп. Объекты разных групп включаются в выборку пропорционально их численности в генеральной совокупности.

Отбор может быть **повторным** (с возвращением выбранного элемента обратно в совокупность) или **бесповторным** (один элемент не может быть включен в выборку дважды).

1.3. Группировка статистических данных

В результате проведения статистического наблюдения получают данные о признаках каждого обследованного элемента статистической совокупности. Однако эти массивы данных собирают не для того, чтобы получить характеристики каждого отдельного элемента совокупности, а для того, чтобы получить характеристики статистической совокупности в целом.

Для этого необходимо обобщить и систематизировать сведения, полученные в ходе статистического наблюдения. В статистической практике этот этап статистического исследования называют группировкой.

Группировка представляет собой разбиение совокупности на группы, однородные по какому-либо признаку.

Существуют 4 типа группировок: типологическая, структурная, аналитическая и комбинационная.

Типологическая группировка служит для исследования распределения совокупности по какому-либо одному качественному признаку (цвет, тип упаковки товара и т. п.). Качественные признаки не предполагают цифрового выражения, например пол, семейное положение, форма собственности и т. д. Число выделенных групп строго соответствует числу градаций качественного признака. Так, например, группировка населения по полу допускает образование только двух групп – мужской и женской. Группировка промышленных предприятий по формам собственности предполагает образование столько же групп, сколько существует ее различных форм. В *табл. 1* приведен пример группировки музеев по классификации ЮНЕСКО, которая предусматривает выделение следующих групп: искусствоведческие, исторические и археологические, краеведческие, естественно-научные, научно-технические, комплексные, отраслевые, специализированные и др.

Т а б л и ц а 1. Группировка музеев РФ по видам

Вид музея	Число музеев	Доля музеев
Искусствоведческий	309	0,122
Исторический и археологический	532	0,210
Краеведческий	1294	0,510
Естественно-научный	39	0,015
Научно-технический	20	0,008
Комплексный	190	0,075
Отраслевой, специализированный и проч.	155	0,061
Итого	2539	1,000

Структурная группировка служит для исследования распределения совокупности по одному количественному признаку. Количественные признаки обычно имеют числовое выражение, например объем выпускаемой продукции, численность работников, возраст человека, доход семьи и т. д.

В зависимости от характера вариации признака различают дискретный и интервальный варианты.

Дискретный вариант характерен для признака, принимающего ограниченное число целых значений, например, оценки на экзамене (2, 3, 4, 5). При этом число выделенных групп совпадает с числом принимаемых значений признака (в нашем примере 4).

Пример такой группировки приведен в *табл. 2*.

Т а б л и ц а 2. Распределение студентов по экзаменационному баллу

Экзаменационный балл	Число студентов	Доля
5	16	0,32
4	23	0,46
3	7	0,14
2	4	0,08
Итого	50	1,00

Если признак может принимать вещественные значения или целые, но число возможных значений признака достаточно велико, то необходимо установить число групп и определить границы соответствующих интервалов.

Такая группировка основывается на группировочном признаке и интервале, который представляет собой промежуток между минимальным и максимальным значением признака в группе.

Интервалы могут быть равными и неравными, открытыми и закрытыми. Под закрытым интервалом понимают интервал, имеющий как нижнюю, так и верхнюю границу, например (100; 200). Под открытым интервалом понимают интервал, одна из границ которого равна ∞ (например, свыше 500).

При определении числа групп и величины интервалов следует иметь в виду, что число объектов в выделенных группах должно быть достаточным,

чтобы характеристики, рассчитанные для отдельных групп, были статистически достоверными.

Рекомендации по выбору числа групп k приведены в *табл. 3*.

Т а б л и ц а 3. Рекомендации по выбору числа групп k

Число наблюдений (n)	Число групп (k)
25–40	5–6
40–60	6–8
60–200	8–12
> 200	10–15

Число групп также можно определить по формуле $k = 1 + 3,32 \lg n$.

Наличие равных интервалов облегчает вычисление различных статистических характеристик, поэтому в дальнейшем будем считать их равными.

Тогда величина интервала вычисляется по формуле

$$\Delta = \frac{x_{\max} - x_{\min}}{k},$$

где $x_{\max}$ — максимальное значение признака в исследуемой совокупности;
 $x_{\min}$ — минимальное значение признака в исследуемой совокупности.

Затем определяются границы каждого интервала:

для первого интервала: от $x_{\min}$ до $x_{\min} + \Delta$;

для второго интервала: от $x_{\min} + \Delta$ до $x_{\min} + 2\Delta$;

.....

для k -го интервала: от $x_{\min} + (k-1)\Delta$ до $x_{\max}$.

Нетрудно видеть, что верхняя граница i -го интервала равна нижней границе $(i + 1)$ -го интервала.

Если в основе группировки лежит целочисленный признак, то нижняя граница $(i + 1)$ -го интервала граница равна верхней границе i -го интервала, увеличенной на 1. Например, группы предприятий по числу работников будут иметь вид: 1–100, 101–200, 201–300 и т. д.

Пример такой группировки приведен в *табл. 4*.

Т а б л и ц а 4. Распределение предприятий региона по числу работников

Численность работников	Число предприятий	Доля
1–100	12	0,15
101–200	18	0,22
201–300	25	0,31
301–400	14	0,18
401–500	11	0,14
Итого	80	1,00

Аналитическая группировка служит для выявления зависимости между признаками. При этом выделяют признак-фактор и признак-результат. Группировка осуществляется по признаку-фактору. В каждой группе рассчитывается

среднее значение признака-результата. Анализируя изменение средних значений признака-результата от группы к группе, можно сделать вывод о наличии или отсутствии взаимосвязи между признаками.

Различие групповых средних позволяет утверждать, что признаки взаимозависимы. Если изменение величины признака-фактора в определенном направлении вызывает изменение признака-результата в том же направлении, то говорят, что связь положительная, а в противном случае (в противоположном направлении) – отрицательная. Результаты аналитической группировки представляются в виде *табл. 5*.

Т а б л и ц а 5

Значение признака-фактора	Количество элементов совокупности в отдельной группе	Среднее значение признака-результата
От до От до		
Итого	Общее число элементов совокупности	

Проследить зависимость между признаками можно также на основе **комбинационной группировки**, которая осуществляется одновременно по двум признакам (*табл. 6*).

Т а б л и ц а 6

Группировка по признаку-фактору	Группировка по признаку-результату				
	От ... до ...	От ... до ...		От ... до ...	Всего
От ... до ...	n_{11}	n_{12}		n_{1m}	$\sum n_{1j}$
От ... до ...	n_{21}	n_{22}		n_{2m}	$\sum n_{2j}$
.....					
От ... до ...	n_{k1}	n_{k2}		n_{km}	$\sum n_{kj}$
Всего	$\sum n_{i1}$	$\sum n_{i2}$		$\sum n_{im}$	n

Здесь n_{ij} – количество случаев совместного появления i -й градации признака-фактора ($i = 1, \dots, k$) и j -й градации признака-результата ($j = 1, \dots, m$).

Если наибольшие числа каждой строки и каждого столбца располагаются вдоль "главной диагонали" таблицы, то можно сделать вывод, что связь положительная и близкая к линейной.

Если наибольшие числа располагаются вдоль другой диагонали таблицы, то можно сделать вывод, что связь отрицательная и близкая к линейной.

Если числа во всех клетках примерно одинаковые, то связи между признаками нет.

Пример 1. Собраны данные о численности работников на 30 предприятиях сферы торговли.

Номер предпр.	Число работников	Номер предпр.	Число работников	Номер предпр.	Число работников
1	12	11	11	21	11
2	10	12	12	22	13
3	14	13	12	23	14
4	14	14	14	24	12
5	13	15	12	25	13
6	12	16	11	26	12
7	12	17	13	27	12
8	12	18	12	28	14
9	15	19	13	29	12
10	13	20	14	30	14

Провести структурную группировку этих предприятий по признаку численности работников.

В данной ситуации признак принимает ограниченное число целых значений. Поэтому нет необходимости вводить интервалы, а просто подсчитать количество предприятий, где число работников равно 10, 11, 12, ..., 15.

Например, только одно предприятие имеет 10 работников, а 11 работников имеется на трех предприятиях.

Результаты таких расчетов приведены в *табл. 7*.

Т а б л и ц а 7

Число работников	Число предприятий
10	1
11	3
12	12
13	6
14	7
15	1
Всего	30

Пример 2. Имеются данные о средней недельной заработной плате на 30 предприятиях (тыс. руб.)

Номер предпр.	Зараб. плата	Номер предпр.	Зараб. плата	Номер предпр.	Зараб. плата
1	8,2	11	9,0	21	7,2
2	9,7	12	6,0	22	6,4
3	5,6	13	7,6	23	7,7
4	7,4	14	8,1	24	9,0
5	8,0	15	11,8	25	8,1
6	6,4	16	5,8	26	7,1
7	6,6	17	9,3	27	7,1
8	6,8	18	7,3	28	8,8
9	8,4	19	8,2	29	7,5
10	7,1	20	7,2	30	9,2

Заметим, что максимальная заработная плата ($x_{\max}$) равна 11,8 тыс. руб., а минимальная ($x_{\min}$) равна 5,6 тыс. руб. В соответствии с приведенными выше рекомендациями выберем число групп (k), равное 5.

Тогда

$$\Delta = \frac{x_{\max} - x_{\min}}{k} = \frac{11,8 - 5,6}{5} = 1,24.$$

Рассчитаем границы всех интервалов.

Для первого интервала: от 5,6 до $5,6 + 1,24 = 6,84$.

Для второго интервала: от 6,84 до $6,84 + 1,24 = 8,08$.

Для третьего интервала: от 8,08 до $8,08 + 1,24 = 9,32$.

Для четвертого интервала: от 9,32 до $9,32 + 1,24 = 10,56$.

Для пятого интервала: от 10,56 до $10,56 + 1,24 = 11,8$.

В первый интервал (от 5,6 до 6,84) попадают предприятия с номерами 3, 6, 7, 8, 12, 16, 22 (всего 7 предприятий). Прodelывая аналогичные расчеты для других интервалов, окончательно получим результаты, приведенные в *табл. 8*.

Т а б л и ц а 8

Заработная плата, тыс. руб.	Количество предприятий
От 5,6 до 6,84	7
От 6,84 до 8,08	11
От 8,08 до 9,32	10
От 9,32 до 10,56	1
От 10,56 до 11,8	1
Всего	30

Пример 3. В таблице приведены данные о численности работников и объемах выпуска продукции на 1 человека по 16 предприятиям.

Номер предпр.	Число работников	Объем выпуска, тыс. руб.	Номер предпр.	Число работников	Объем выпуска, тыс. руб.
1	456	33,4	9	478	35,4
2	375	31,8	10	783	42,8
3	484	38,0	11	365	36,0
4	456	33,4	12	1300	50,5
5	788	42,5	13	1998	53,4
6	7900	90,0	14	2320	62,0
7	9581	98,0	15	3650	64,0
8	3550	67,0	16	3450	71,0

Выполним аналитическую группировку, считая число работников признаком-фактором, а объем выпуска продукции – признаком-результатом.

Зададим число групп (k), равное 3. Заметим, что

$$x_{\max} = 9581, x_{\min} = 365.$$

Тогда

$$\Delta = \frac{9581 - 365}{3} = 3072.$$

Рассчитаем границы интервалов.

Для первого интервала: от 365 до $365 + 3072 = 3437$.

Для второго интервала: от 3437 до $3437 + 3072 = 6509$.

Для третьего интервала: от 6509 до $6509 + 3072 = 9581$.

Заметим, что в первый интервал попадут предприятия с номерами 1, 2, 3, 4, 5, 9, 10, 11, 12, 13, 14 (всего 11 предприятий).

Рассчитаем по этим предприятиям среднее значение признака-результата:

$$\bar{y}_1 = \frac{33,4 + 31,8 + 38,0 + 33,4 + 42,5 + 35,4 + 42,8 + 36,0 + 50,5 + 53,4 + 62}{11} = 41,7.$$

Во второй интервал попадут предприятия с номерами 8, 15, 16 (всего 3 предприятия). Среднее значение признака-результата для этих предприятий составит

$$\bar{y}_2 = \frac{67,0 + 64,0 + 71,0}{3} = 67,3.$$

В третий интервал попадут предприятия с номерами 6 и 7. Среднее значение признака-результата для этих предприятий составит

$$\bar{y}_3 = \frac{90,0 + 98,0}{2} = 94,0.$$

Полученные результаты расчетов сведены в таблицу аналитической группировки (табл. 9).

Т а б л и ц а 9

Число работников	Количество предприятий	Среднее значение признака-результата ($\bar{y}_i$)
365–3437	11	41,7
3437–6509	3	67,3
6509–9581	2	94,0
Всего	16	

Анализ полученных результатов показывает, что среднее значение объема выпуска продукции на 1 человека устойчиво растет от группы к группе, а следовательно, между количеством работников и объемом выпуска продукции на 1 человека имеется положительная связь.

Для проведения комбинационной группировки рассчитаем интервалы для признака – результата.

Зададим число групп (m), равное 3. Заметим, что

$$y_{\max} = 98,0, y_{\min} = 31,8.$$

Тогда

$$\Delta = \frac{98,0 - 31,8}{3} = 22,1.$$

Рассчитаем границы интервалов.

Для первого интервала: от 31,8 до $31,8 + 22,1 = 53,9$.

Для второго интервала: от 53,9 до $53,9 + 22,1 = 76,0$.

Для третьего интервала: от 76,0 до 98,0.

Результаты комбинационной группировки представлены ниже.

	31,8–53,9	53,9–76,0	76,0–98,0	Итого
365–3437	10	1	-	11
3437–6509	-	3	-	3
6509–9581	-	-	2	2
Итого	10	4	2	16

Из данных результатов видно, что наибольшие значения каждого столбца и каждой строки лежат на главной диагонали. Поэтому можно сделать вывод о положительной связи между численностью работников предприятия и объемом выпуска продукции на 1 работника.

1.4. Понятие о рядах распределения. Полигон и гистограмма

Рядами распределения называются числовые ряды, характеризующие структуру совокупности по некоторому признаку. Ряд распределения, полученный по качественному признаку, называется атрибутивным, а по количественному – вариационным. Вариационный ряд может быть дискретным (признак принимает ограниченное число возможных значений, например 2, 3, 4, 5) или интервальным (значения признака выражены вещественными числами или число возможных значений признака достаточно велико).

Характеристиками ряда являются:

x_i – *варианта* (отдельное возможное численное значение признака) ($i = 1, k$);

n_i – *частота* (численность отдельных групп);

n – общее число элементов совокупности;

q_i – *частость* (доля отдельных групп во всей совокупности).

При этом $q_i = \frac{n_i}{n}$.

Вариационный ряд оформляется в виде таблицы, где в первой графе указываются варианты (интервалы) значений признака, а в следующих – частота и частость.

Ряд распределения в целом характеризует структуру совокупности по данному признаку. Однако могут использоваться и кумулятивные ряды, т. е. ряды накопленных частот (частостей).

Накопленная частота (частость) – это число (доля) элементов совокупности, у которых значения признака не превышают данного.

Обозначим:

$F(x)$ – накопленная частота для данного значения x ;

$G(x)$ – накопленная частость для данного значения x .

Эти характеристики обладают следующими свойствами:

$$0 \leq F(x) \leq n; \quad 0 \leq G(x) \leq 1.$$

Рассмотрим интервал с номером $i : [x_i ; x_{i+1}]$.

Накопленная частота на конец i -го интервала определяется по формуле

$$F(x_{i+1}) = \sum_{l=1}^i n_l.$$

Вариационный ряд можно изобразить в виде графика.

Изображением дискретного ряда является **полигон**. При его построении по оси абсцисс откладываются варианты (x_i), а по оси ординат – частоты или частости – f_i . Затем точки с координатами ($x_i; f_i$) последовательно соединяются отрезками прямой, в результате чего получают ломаную линию. Графически распределение студентов по экзаменационному баллу (табл. 2) представлено на рис. 1.

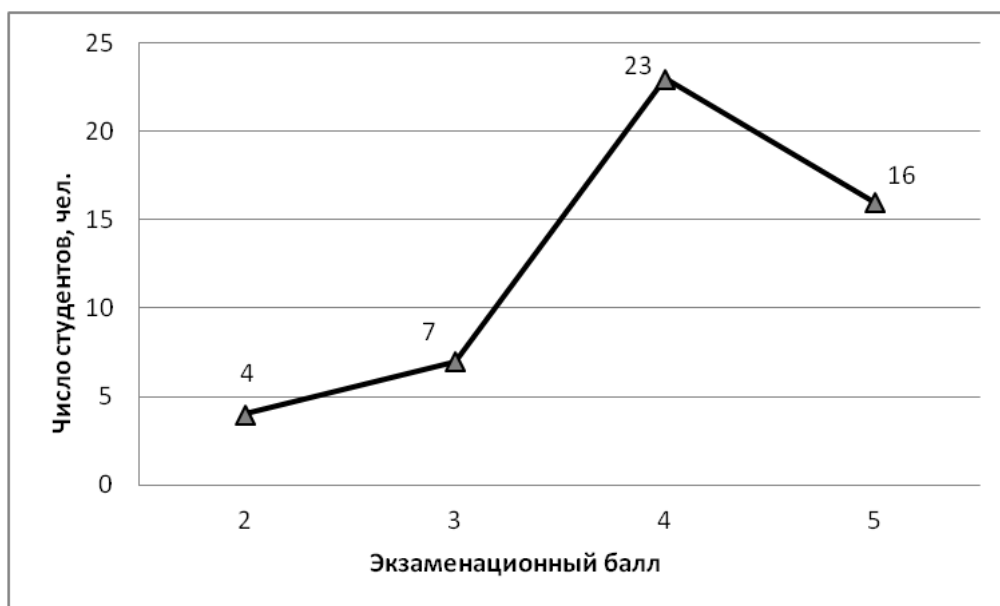


Рис. 1. Полигон распределения студентов по экзаменационному баллу

Изображением интервального ряда является *гистограмма*. При ее построении по оси абсцисс откладываются интервалы ряда. Над осью абсцисс строится прямоугольник, основанием которого является интервал, а высотой – значение частоты или частоты.

Гистограмма распределения предприятий по числу работников (см. табл. 4) представлена на рис. 2.

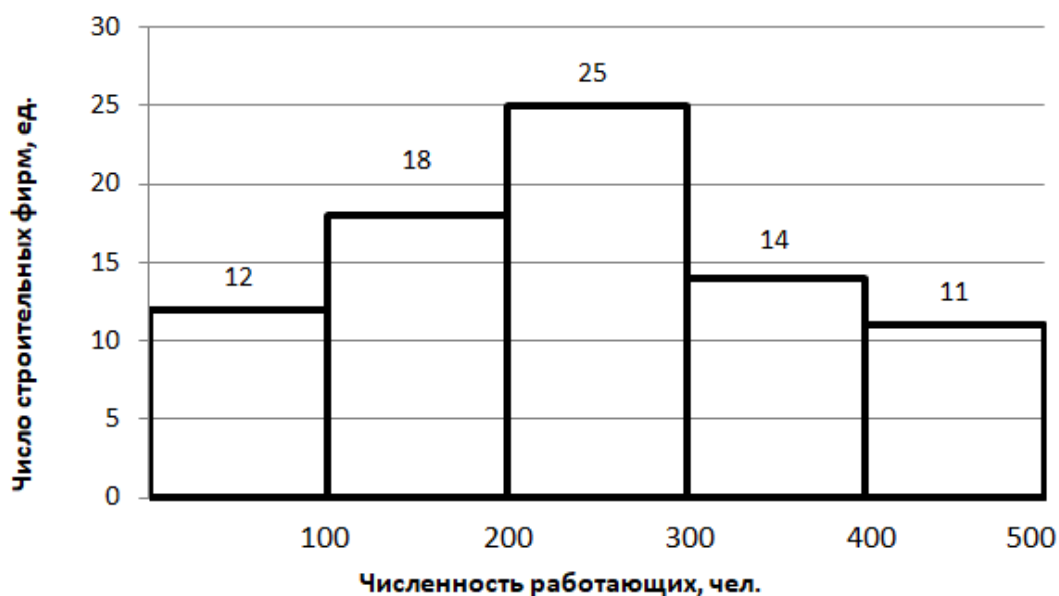


Рис. 2. Гистограмма распределения предприятий по числу работников

Изображением ряда накопленных частот является *кумулята*. Накопленные частоты откладываются по оси ординат для границ интервалов и соединяются отрезками прямых.

Пример 1. Распределение квартир дома по числу жителей приведено в таблице. Полигон и кумулята представлены на *рис. 3*.

Число живущих в квартире x_i	Число квартир (частота) n_i	Накопленная частота F_i
1	2	2
2	3	5
3	10	15
4	23	38
5	9	47
6	2	49
7	1	50
ВСЕГО	50	

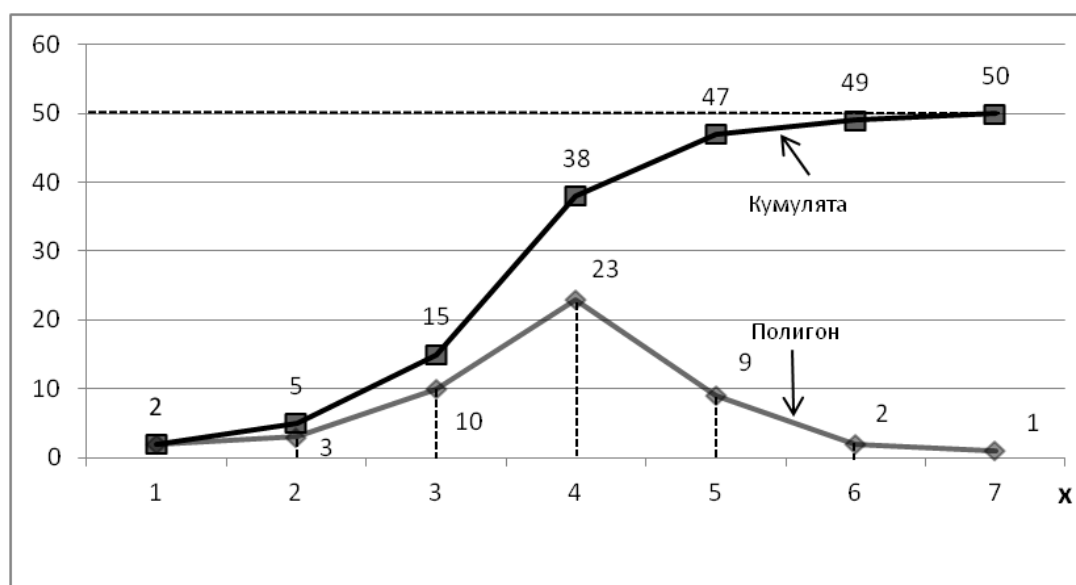


Рис. 3. Полигон и кумулята распределения квартир по числу жителей

Пример 2. Распределение банков по степени риска приведено в таблице. Гистограмма и кумулята представлены на *рис. 4*.

Степень риска, %	Доля банков (частость) q_i	Накопленная частость G_i
0–10	0,61	0,61
10–20	0,04	0,65
20–30	0,35	1,00
ВСЕГО	1,00	

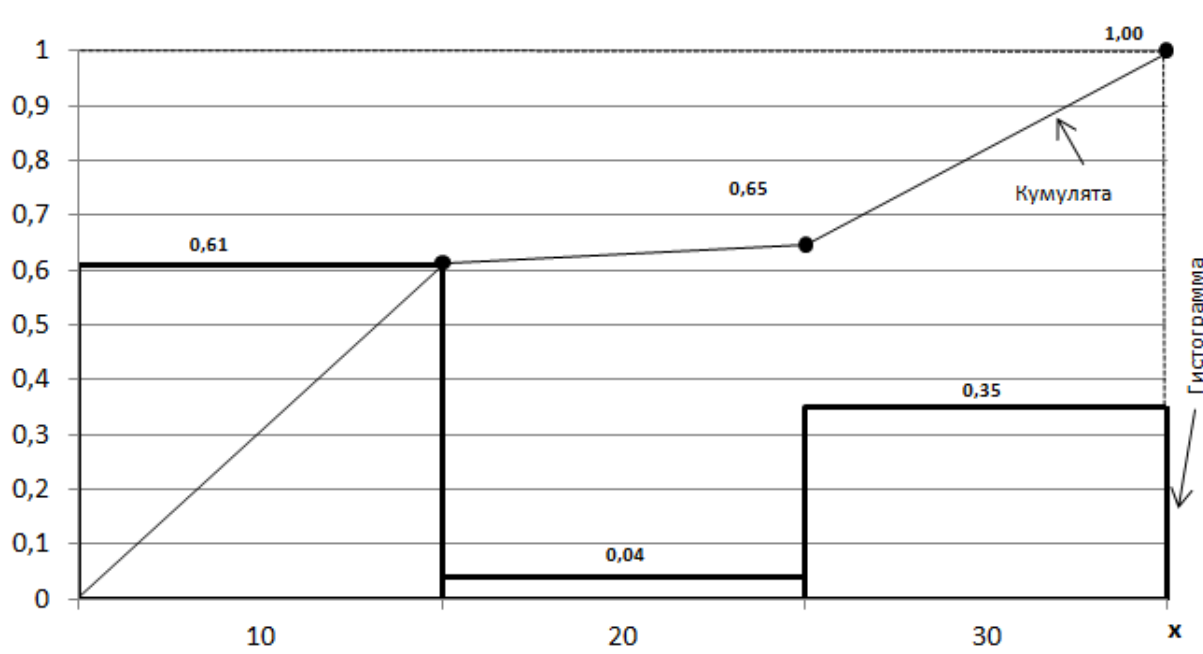


Рис. 4. Гистограмма и кумулята распределения банков по степени риска

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 1

В таблице представлены данные о численности работников на 30 предприятиях малого бизнеса. Проведите структурную группировку данных предприятий по признаку численности работников при числе групп, равном 5.

Номер предприятия	Число работников	Номер предприятия	Число работников	Номер предприятия	Число работников
1	32	11	10	21	32
2	37	12	17	22	32
3	32	13	17	23	32
4	41	14	17	24	17
5	32	15	19	25	10
6	32	16	10	26	37
7	32	17	21	27	41
8	41	18	22	28	41
9	16	19	21	29	16
10	32	20	19	30	16

Задача 2

В таблице представлены данные о численности работников на 15 предприятиях. Проведите структурную группировку данных предприятий по признаку численности работников при числе групп, равном 3.

Номер предприятия	Число работников	Номер предприятия	Число работников	Номер предприятия	Число работников
1	17	6	20	11	29
2	17	7	20	12	18
3	19	8	19	13	18
4	20	9	26	14	17
5	27	10	25	15	19

Задача 3

В таблице приведены данные о численности работников и объемах выпуска продукции на 1 человека по 16 предприятиям.

Номер предприятия	Число работников	Объем выпуска, тыс. руб.	Номер предприятия	Число работников	Объем выпуска, тыс. руб.
1	10	33,4	9	11	35,4
2	18	41,8	10	10	32,8
3	15	38,0	11	19	39,0
4	15	38,4	12	21	45,1
5	21	44,5	13	14	38,8
6	13	32,4	14	15	35,1
7	17	40,3	15	13	36,4
8	22	45,8	16	15	40,7

Выполнить аналитическую группировку, считая число работников признаком-фактором, а объем выпуска продукции – признаком-результатом при числе групп, равном 3.

Задача 4

В таблице приведены данные о численности работников и себестоимости единицы продукции по 16 предприятиям.

Номер предприятия	Число работников	Себестоимость, руб.	Номер предприятия	Число работников	Себестоимость, руб.
1	50	200	9	47	220
2	60	140	10	40	250
3	51	210	11	45	240
4	52	180	12	38	260
5	48	220	13	35	270
6	45	235	14	40	250
7	45	240	15	36	265
8	40	250	16	34	300

Выполнить комбинационную группировку при числе групп по обоим признакам, равном 3.

Задача 5

Распределение квартир дома по числу жителей приведено в таблице. Найти накопленные частоты и занести в таблицу. Постройте полигон и кумуляту.

Число живущих в квартире x_i	Число квартир (частота) n_i	Накопленная частота F_i
1	5	
2	3	
3	20	
4	10	
5	2	
ВСЕГО		

Задача 6

Распределение банков по степени риска приведено в таблице. Найти накопленные частоты и занести в таблицу. Постройте гистограмму и кумуляту.

Степень риска, %	Доля банков (частость) q_i	Накопленная частота G_i
0–2	0,31	
2–4	0,34	
4–6	0,35	
ВСЕГО	1,00	

2. МЕТОДОЛОГИЯ СТАТИСТИКИ. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

2.1. Понятие о случайных событиях и их вероятностях

Есть события, которые при данных условиях происходят непременно, а есть такие, которые могут произойти или не произойти. Такие события называются *случайными*.

Отметим, что под событием обычно понимают исход какого-то опыта, эксперимента.

Говорят, что несколько событий образуют *полную группу*, если в результате опыта непременно должно произойти хотя бы одно из них.

Примеры:

- выпадение 1, 2, 3, ... 6 очков при бросании кубика;
- выпадение четного и нечетного количества очков;
- выпадение < 4 и выпадение > 2 очков. $\{1, 2, 3, 4, 5, 6\}$.

Рассмотрим пример. В результате бросания кубика возможны следующие исходы (события):

- A_1 – выпадение цифры 1;
- A_2 – выпадение цифры 2;
- A_3 – выпадение цифры 3;
- A_4 – выпадение цифры 4;
- A_5 – выпадение цифры 5;
- A_6 – выпадение цифры 6.

Эти события обладают следующими свойствами:

- а) одно из событий A_1 – A_6 обязательно произойдет (полная группа);
- б) события A_i и A_j ($i \neq j$) –взаимоисключающие, т. е. может произойти только одно из них.

События, обладающие такими свойствами, называются *элементарными* и образуют пространство событий

$$\Omega = \{ A_1, A_2, A_3, A_4, A_5, A_6 \}.$$

События могут быть и не элементарными.

Пример

- A – выпало четное число очков;
- B – выпало более трех очков.

A и B не исключают друг друга, например, при выпадении 4 или 6 происходят оба этих события. Может не произойти ни A , ни B (выпадение 1 или 3).

Событие может заключаться и в том, что получен какой-то результат измерения. Тогда в качестве события может рассматриваться точка на числовой оси, а пространство событий составляют все точки некоторого интервала.

Пример. Измерение веса студента. Событие-вес оказался равным 65 кг. Пространство событий $[50; 100]$ кг. Здесь два механизма появления случайности-случайность выбора и погрешность весов.

Любое подпространство пространства событий также является событием.

Примеры

A – выпадение четного числа очков при бросании куба – $\{A_2, A_4, A_6\}$;

B – выпадение числа, большего 2 – $\{A_3, A_4, A_5, A_6\}$;

C – получение результата измерения веса в диапазоне $[50; 60]$ кг.

Достоверное и невозможное событие

Пример. На трех карточках проставлены цифры 1, 2, 3. Эти карточки перетасовали и разложили в один ряд. Получилось трехзначное число α .

Пространство элементарных событий образуют следующие события:

$A_1 - \alpha = 123$;

$A_2 - \alpha = 132$;

$A_3 - \alpha = 213$;

$A_4 - \alpha = 231$;

$A_5 - \alpha = 312$;

$A_6 - \alpha = 321$.

Рассмотрим событие $\emptyset$ ($\alpha < 123$) и I ($\alpha \geq 123$). Событие $\emptyset$ в данном эксперименте не наступает никогда (невозможное событие), а событие I имеет место всегда (достоверное событие).

Соотношения между событиями

Пример

A – выпадение цифры 2;

B – выпадение четного числа очков.

Событие A является частью события B , так как событие B состоит из трех элементарных событий (выпадение 2, 4, 6).

Обозначение: $A \subset B$ (A влечет B).

Отметим, что из того, что событие A произошло следует, что произошло и событие B .

Геометрическая интерпретация приведена на *рис. 5*.

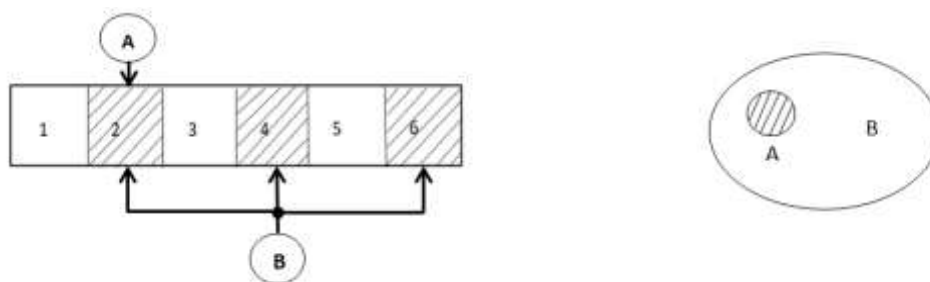


Рис. 5. Геометрическая интерпретация части события

Объединение и пересечение событий

Объединением событий A и B называется событие C , состоящее в наступлении хотя бы одного из событий A или B .

Обозначение: $C = A \cup B$.

Пример 1

Пусть A – выпадение цифры 1;

B – выпадение цифры 5.

Тогда $C = A \cup B$ заключается в выпадении одной из цифр $\{1, 5\}$.

Пример 2

Пусть A – выпадение одной из цифр $\{1, 2, 5\}$;

B – выпадение одной из цифр $\{1, 3, 4\}$.

Тогда $C = A \cup B$ заключается в выпадении одной из цифр $\{1, 2, 3, 4, 5\}$.

Объединением событий $A_1, A_2, \dots, A_n$ называется событие A , состоящее в появлении хотя бы одного из событий $A_1, A_2, \dots, A_n$.

$A = A_1 \cup A_2 \cup \dots \cup A_n$.

Пересечение событий $C = A \cap B$ есть событие, состоящее в одновременном наступлении обоих событий A и B .

Пример 1

Пусть A – выпадение одной из цифр $\{1, 3, 5\}$;

B – выпадение одной из цифр $\{2, 3, 5\}$.

Тогда пересечение событий A и B есть выпадение одной из цифр $\{3, 5\}$.

Пример 2

A – выбранное число кратно 2;

B – выбранное число кратно 3.

$C = A \cap B$ – выбранное число кратно 6.

Пересечением событий $A_1, A_2, \dots, A_n$ называется событие A , состоящее в одновременном наступлении всех $(A_1, A_2, \dots, A_n)$ событий.

$A = A_1 \cap A_2 \cap \dots \cap A_n$.

Два события A и B , пересечение которых – невозможное событие ($A \cap B = \emptyset$) называются несовместными (взаимоисключающими).

Элементарные события – несовместные.

Пример

Пусть A – выпадение цифры 2;

B – выпадение цифры 3.

$A \cap B = \emptyset$.

Если объединение событий A и B – достоверное событие, а их пересечение – невозможное событие, то события A и B называются противоположными.

Обозначение: $B = \bar{A}$, $A = \bar{B}$.

Событие $\bar{A}$ состоит из тех точек пространства событий, которые не составляют A .

Пример 1

A – выпадение герба;

$\bar{A}$ – выпадение решки.

Пример 2

A – выпадение четной цифры $\{2, 4, 6\}$;

$\bar{A}$ – выпадение нечетной цифры $\{1, 3, 5\}$.

Геометрическая интерпретация приведена на *рис. 6*.

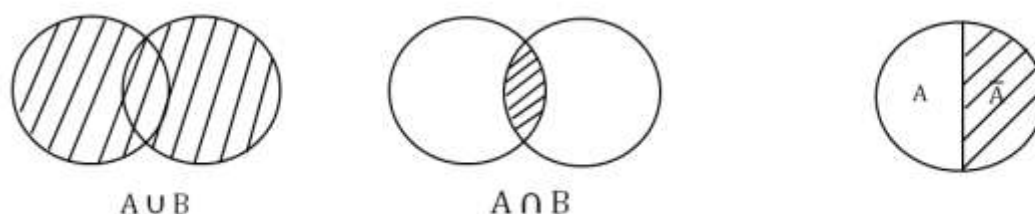


Рис. 6. Геометрическая интерпретация объединения, пересечения событий и противоположного события

Очевидно, что $A \cup \bar{A} = I$, $A \cap \bar{A} = \emptyset$.

Классическое понятие вероятности

Вероятность события представляет собой число, характеризующее степень возможности события. При этом достоверному событию приписывается вероятность, равная 1, а невозможному – 0.

В примере с кубиком могут выпасть 1, 2, 3, ..., 6 очков. Эти события элементарные. Здравый смысл подсказывает нам, что их можно считать равновероятными.

Событие A называется благоприятствующим некоторому событию B , если появление этого события влечет за собой событие B .

Например, событие появления числа 3 влечет событие появления нечетного числа.

Более возможно то событие, которому благоприятствуют большее число равновероятных элементарных событий данного пространства.

Пример

A: выпадение четного числа – {2, 4, 6};

B: выпадение числа, не превосходящего 2 – {1, 2}.

Количество всех элементарных равновозможных событий этого пространства – 6.

Вероятность случайного события *H* определяется по формуле

$$P(H) = \frac{m}{n},$$

где *m* – число равновозможных элементарных событий, которые благоприятствуют событию *H*;

n – число всех равновозможных элементарных событий пространства.

В нашем примере

$$P(A) = \frac{3}{6} = \frac{1}{2}; \quad P(B) = \frac{2}{6} = \frac{1}{3}.$$

Так как $0 \leq m \leq n$, то

$$0 \leq \frac{m}{n} \leq 1.$$

Отсюда вытекает свойство вероятностей

$$0 \leq P(H) \leq 1,$$

т. е. вероятность случайного события лежит в диапазоне от 0 до 1.

$$P(I) = \frac{n}{n} = 1 \text{ – вероятность достоверного события;}$$

$$P(\emptyset) = \frac{0}{n} = 0 \text{ – вероятность невозможного события.}$$

Замечание. Если $P(A) = 0$, то это не значит, что событие невозможное (пример – попадание в заданную точку).

Вероятностью случайного события *H* называется отношение числа равновозможных элементарных событий, благоприятствующих этому событию, к числу всех равновозможных элементарных событий пространства.

Возникает вопрос: А действительно ли равновозможны эти события? Что если кубик не идеальный? Как оценить эти вероятности тогда?

Поэтому далее рассмотрим статистическое определение вероятности.

Частота и вероятность событий

Произведем многократное бросание кубика (n раз) и подсчитаем число выпадений грани с цифрой 1 – событие A_1 (m раз).

Тогда $\frac{m}{n}$ – частота события A_1 в n испытаниях.

В теории вероятностей доказывается, что при увеличении n частота события стремится к некоторой постоянной величине, называемой вероятностью.

Вероятность объединения событий

Случай 1: события несовместны ($A_1 \cap A_2 = \emptyset$).

Пусть n – общее число шаров в урне;

n_a – число красных шаров;

n_b – число синих шаров;

$n_c = n_a + n_b$ – число цветных шаров.

Остальные шары – белые.

$$\frac{n_c}{n} = \frac{n_a + n_b}{n} = \frac{n_a}{n} + \frac{n_b}{n}.$$

Таким образом,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

Если случайные события несовместны, то вероятность их объединения равна сумме вероятностей этих событий.

Следствие 1. Если события $A_1, A_2, \dots, A_n$ образуют полную группу несовместных событий, то сумма их вероятностей равна 1.

$$\sum_{i=1}^n P(A_i) = 1.$$

Пример. В случае бросания куба события A_1 (выпадение цифры 1), A_2 (выпадение цифры 2), ..., A_6 (выпадение цифры 6) – несовместны.

$$P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6) = P(I) = 1 = P(A_1) + P(A_2) + \dots + P(A_6).$$

Так как $P(A_1) = P(A_2) = \dots P(A_6)$,

$$\text{то } P(A_1) = P(A_2) = P(A_3) = \dots P(A_6) = \frac{1}{6}.$$

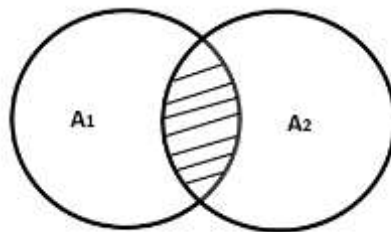
Следствие 2. Сумма вероятностей противоположных событий равна 1.

Пусть события A и $\bar{A}$ – противоположные ($A \cap \bar{A} = \emptyset$; $A \cup \bar{A} = I$).

$$P(A \cup \bar{A}) = P(I) = 1 = P(A) + P(\bar{A}),$$

$$\text{откуда } P(\bar{A}) = 1 - P(A).$$

Случай 2. События A_1 и A_2 – совместны ($A_1 \cap A_2 \neq \emptyset$).



Пусть n_1 – число попаданий в A_1 ;

n_2 – число попаданий в A_2 ;

n_{12} – число попаданий в A_1 и A_2 ;

n – общее число выстрелов.

Число попаданий в мишени ($A_1 \cup A_2$) равно $n_1 + n_2 - n_{12}$.

$$\text{Частота попаданий } \frac{n_1 + n_2 - n_{12}}{n} = \frac{n_1}{n} + \frac{n_2}{n} - \frac{n_{12}}{n}.$$

Переходя к пределу, получим

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

Вероятность объединения двух совместных событий равна сумме вероятностей этих событий минус вероятность их совместного осуществления.

Пример

Пусть A_1 – выпадение одной из цифр $\{1, 3, 5\}$;

A_2 – выпадение одной из цифр $\{2, 3, 6\}$;

$A_1 \cup A_2$ – выпадение одной из цифр $\{1, 2, 3, 5, 6\}$.

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2);$$

$$P(A_1) = \frac{3}{6} = \frac{1}{2}; P(A_2) = \frac{3}{6} = \frac{1}{2}; P(A_1 \cap A_2) = P(3) = \frac{1}{6};$$

$$P(A_1 \cup A_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}.$$

Условные вероятности. Вероятность пересечения событий

Понятие условной вероятности рассмотрим на конкретном примере.

Пусть в двух ящиках находится по 10 шаров, причем в 1-м ящике – 2 белых шара, во втором – 5.

Если мы осуществляем выбор из 1-го ящика, то вероятность выбора белого шара равна 0,2, а если из второго, то 0,5.

Таким образом, вероятность выбора белого шара зависит от другого события – выбора ящика.

Такая вероятность называется условной и обозначается $P(A_2/A_1)$ – вероятность события A_2 при условии, что произошло событие A_1 .

В нашем примере A_1 – выбор первого ящика, а A_2 – выбор белого шара.

Пусть n_{A_1} – количество случаев выбора первого ящика;

n_{A_1, A_2} – количество случаев выбора белого шара из первого ящика ($A_1 \cap A_2$);

n – количество испытаний.

$$\frac{n_{A_1, A_2}}{n} = \frac{n_{A_1}}{n} \cdot \frac{n_{A_1, A_2}}{n_{A_1}}.$$

Переходя к пределу, получим

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 / A_1).$$

Аналогично

$$P(A_1 \cap A_2) = P(A_2) \cdot P(A_1 / A_2).$$

Вероятность пересечения двух событий равна произведению вероятности одного из этих событий на условную вероятность другого при условии, что первое произошло.

Из приведенных выше формул следует

$$P(A_2 / A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}; P(A_1 / A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)}.$$

Два события A_1 и A_2 называются стохастически независимыми, если

$$P(A_2 / A_1) = P(A_2);$$

$$P(A_1 / A_2) = P(A_1),$$

то есть появление одного из них не изменяет вероятности появления другого.

Отсюда для стохастически независимых событий

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2).$$

Вероятность пересечения двух независимых событий равна произведению вероятностей этих событий.

Это свойство может быть обобщено на случай произвольного числа независимых событий

$$P\left\{\bigcap_{i=1}^n A_i\right\} = \prod_{i=1}^n P(A_i).$$

Пример 1. Какова вероятность, что при трех бросках кубика выпадет 3 шестерки?

A_1 – выпадение шестерки в первом броске;

A_2 – выпадение шестерки во втором броске;

A_3 – выпадение шестерки в третьем броске.

Вероятность выпадения трех шестерок

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2) P(A_3) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}.$$

Пример 2. Пусть A_1 – событие, заключающееся в выборе дамы червей из колоды карт; A_2 – событие, заключающееся в выборе дамы из колоды карт.

$$\text{Тогда } P(A_1) = \frac{1}{36}; P(A_1 / A_2) = \frac{1}{4}.$$

Пример 3. Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,7, а для второго – 0,8.

Найти вероятность того, что при одном залпе в мишень попадает только один из стрелков.

Обозначим попадание в цель первым стрелком – событие A , вторым – событие B , промах первого стрелка – событие $\bar{A}$, промах второго – событие $\bar{B}$.

$$P(A) = 0,7; \quad P(\bar{A}) = 0,3; \quad P(B) = 0,8; \quad P(\bar{B}) = 0,2.$$

Событие C – попадание в мишень только одним из стрелков представляет собой объединение двух несовместных событий: C_1 – первый стрелок попадет, а второй промахнется ($C_1 = A \cap \bar{B}$) и C_2 – первый стрелок промахнется, а второй попадет ($C_2 = \bar{A} \cap B$).

Вероятность того, что первый стрелок попадет в мишень, а второй – нет равна

$$P(C_1) = P(A)P(\bar{B}) = 0,7 \cdot 0,2 = 0,14.$$

Вероятность того, что второй стрелок попадет в цель, а первый – нет равна

$$P(C_2) = P(\bar{A})P(B) = 0,3 \cdot 0,8 = 0,24.$$

Тогда вероятность попадания в цель только одним стрелком равна

$$P(C) = P(C_1) + P(C_2) = 0,14 + 0,24 = 0,38.$$

Пример 4. Чему равна вероятность того, что при бросании трех игральных костей 6 очков появится хотя бы на одной из костей?

Вероятность выпадения 6 очков при одном броске кости равна $\frac{1}{6}$. Вероятность того, что не выпадет 6 очков – $\frac{5}{6}$. Вероятность того, что при броске трех костей не выпадет ни разу 6 очков равна $P = \left(\frac{5}{6}\right)^3 = \frac{125}{216}$.

Тогда вероятность того, что хотя бы один раз выпадет 6 очков равна $1 - \frac{125}{216} = \frac{91}{216}$.

2.2. Понятие о случайных величинах и их характеристиках

Если сопоставить каждому возможному событию некоторое число (величину), то такое число называется случайной величиной, так как его значение будет случайным.

Более строго случайная величина X определяется как функция, заданная на множестве элементарных исходов (или в пространстве элементарных событий), т. е.

$$X = f(\omega),$$

где ω – элементарный исход (или элементарное событие), принадлежащее пространству Ω .

Случайная величина может быть дискретной, т. е. принимать только конечное или счетное число различных значений, и непрерывной, т. е. может принимать любые значения в определенном интервале числовой оси.

Примеры случайных величин:

X – число родившихся детей в течение суток в каком-то городе;

Y – число произведенных выстрелов до первого попадания;

Z – дальность полета снаряда.

В приведенных примерах X , Y – дискретные случайные величины, Z – непрерывная случайная величина.

Закон распределения случайной величины. Функция распределения вероятностей (ФРВ)

Хотя в каждом конкретном опыте нельзя точно предвидеть результат наблюдения, поведение случайных величин подчиняется определенным закономерностям и характеризуется законом распределения.

Для дискретной случайной величины закон распределения полностью определяется тем множеством значений, которые она может принимать, и их вероятностями.

Пример (для кубика)

x_i	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Ясно, что $\sum_{i=1}^n P(X = x_i) = 1$ (полная группа несовместных событий).

В теории вероятностей закон распределения выражается также функцией распределения вероятностей.

Функцией распределения вероятностей случайной величины X называется функция $F(x)$, выражающая для каждого x вероятность того, что случайная величина X примет значение, меньшее x .

$$F(x) = P(X < x),$$

где x пробегает все значения на числовой оси.

В примере с кубиком

$$P(X = 1) = P(X = 2) = \dots = P(X = 6) = \frac{1}{6}.$$

Для $-\infty < x < 1$

$$F(x) = 0,$$

$x = 1$

$$F(1) = P(X < 1) = 0,$$

$1 < x < 2$

$$F(x) = P(X < x) = P(X = 1) = \frac{1}{6},$$

$x = 2$

$$F(2) = P(X < 2) = P(X = 1) = \frac{1}{6},$$

$$2 < x \leq 3$$

$$F(x) = P(X < x) = P(X = 1) + P(X = 2) = \frac{2}{6}.$$

В общем виде

$$F(x) = \sum_{x_i < x} P(X = x_i).$$

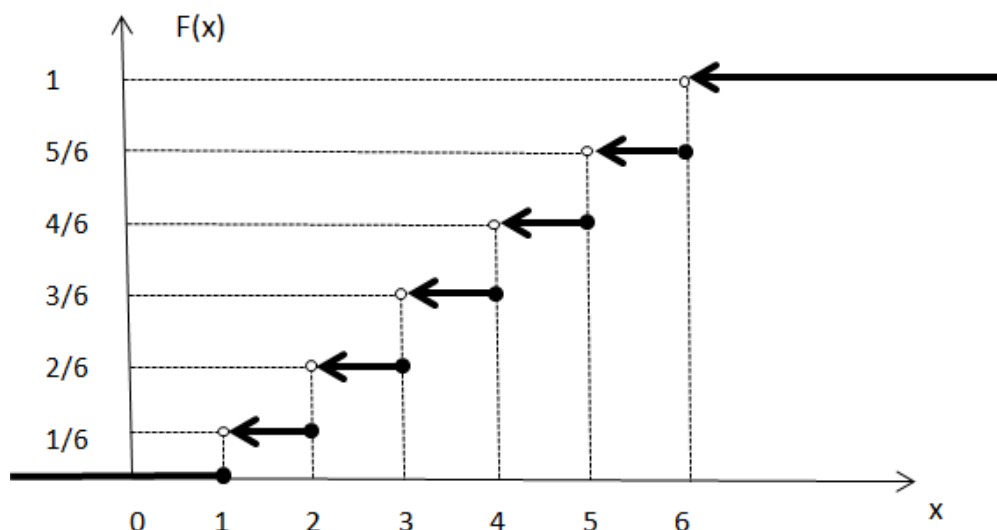


Рис. 7. Функция распределения вероятностей для кубика

Функция распределения вероятностей (ФРВ) используется и для непрерывных случайных величин. В этом случае ФРВ является непрерывной функцией.

ФРВ обладает следующими свойствами:

- 1) $0 \leq F(x) \leq 1$, так как это вероятность.
- 2) $F(x_2) \geq F(x_1)$ при $x_2 > x_1$, то есть функция неубывающая.
- 3) $F(-\infty) = 0$; $F(\infty) = 1$.

Вероятность попадания случайной величины X на участок $[a, b)$:

$$P(a \leq X < b) = F(b) - F(a).$$

Доказательство:

$$A: x < b$$

$$B: x < a$$

$$C: a \leq x < b$$

$$A = B \cup C$$

$$B \cap C = \emptyset$$

$$P(A) = P(B) + P(C)$$

$$\text{или } F(b) = F(a) + P(a \leq x < b).$$

$$\text{Отсюда } P(a \leq X < b) = F(b) - F(a).$$

Если случайная величина дискретная, ее закон распределения может также выражаться набором вероятностей.

Если случайная величина непрерывная, ее закон распределения может также выражаться плотностью распределения вероятностей.

Плотность распределения вероятностей

Рассмотрим производную от ФРВ:

$$\lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x) = f(x),$$

где $f(x)$ – плотность распределения вероятностей (ПРВ).

Известно, что

$$P(x \leq X < x + \Delta x) = F(x + \Delta x) - F(x).$$

Отсюда

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x}.$$

Физический смысл ПРВ – доля вероятности на единицу длины (плотность).

Плотностью распределения вероятностей $f(x)$ называется производная функции распределения вероятностей $F(x)$.

$F(x)$ – первообразная функции $f(x)$.

Поэтому

$$P(a \leq X < b) = F(b) - F(a) = \int_a^b f(x) dx.$$

Геометрически вероятность попадания случайной величины X на участок (a, b) равна площади под кривой ПРВ, опирающейся на этот участок (рис. 8).

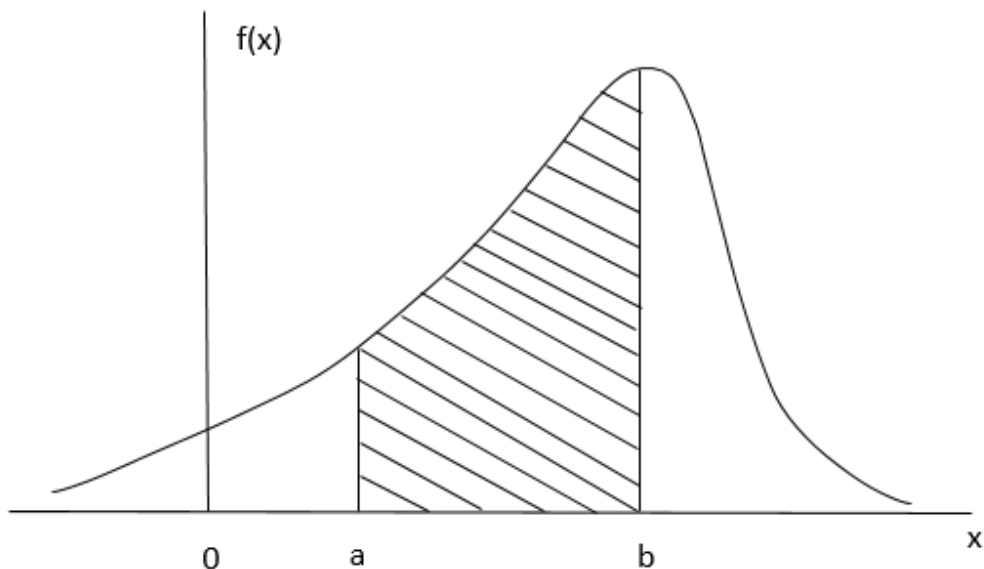


Рис. 8. Геометрическая интерпретация вероятности попадания случайной величины на участок (a, b)

При этом

$$F(x) = P(X < x) = P(-\infty < X < x) = \int_{-\infty}^x f(x)dx.$$

ПРВ обладает следующими свойствами:

1. Если все возможные значения X принадлежат промежутку $[a, b)$, то $a \leq X < b$ – достоверное событие и, следовательно:

$$\int_a^b f(x)dx = 1.$$

В частности

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

2. $f(x) \geq 0$, так как $F(x)$ – неубывающая функция.

Числовые характеристики закона распределения случайной величины

Закон распределения случайной величины полностью определяется функцией распределения вероятностей $F(x)$ или плотностью распределения вероятностей $f(x)$ (для дискретной случайной величины – набором вероятностей).

На практике представляют интерес такие параметры закона распределения, как математическое ожидание, дисперсия, среднее квадратичное отклонение.

Математическим ожиданием дискретной случайной величины называется сумма произведений всех возможных значений случайной величины на их вероятности:

$$m_x = M(X) = \sum_{i=1}^n x_i P(X = x_i).$$

Дисперсией (рассеиванием) случайной величины называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D(X) = M[X - M(X)]^2.$$

Дисперсия характеризует отклонение (разброс, вариацию) значений случайной величины относительно среднего значения.

Для дискретной случайной величины

$$D(X) = \sum_{i=1}^n (x_i - M(X))^2 P(X = x_i).$$

Математическим ожиданием непрерывной случайной величины X , возможные значения которой принадлежат отрезку $[a, b]$, называется определенный интеграл

$$M(X) = \int_a^b xf(x)dx.$$

Если возможные значения случайной величины рассматриваются на всей числовой оси, то математическое ожидание находится по формуле

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

Дисперсия непрерывной случайной величины вычисляется по формуле

$$D(X) = \int_{-\infty}^{\infty} [x - M(X)]^2 f(x)dx.$$

Для вычисления дисперсии удобнее пользоваться следующей формулой:

$$D(X) = M(X^2) - [M(X)]^2,$$

т. е. дисперсия равна разности между математическим ожиданием квадрата случайной величины X и квадратом ее математического ожидания.

Средним квадратичным отклонением случайной величины X называется квадратный корень из дисперсии

$$\sigma(X) = \sqrt{D(X)}.$$

Начальным моментом порядка s случайной величины X называется математическое ожидание величины X^s :

$$m_s[X] = M[X^s].$$

Для дискретной случайной величины

$$m_s[X] = \sum_{i=1}^n x_i^s P(X = x_i).$$

Для непрерывной случайной величины X начальным моментом s -го порядка называется интеграл:

$$m_s[X] = \int_{-\infty}^{\infty} x^s f(x) dx.$$

Начальный момент первого порядка равен математическому ожиданию.

Центральным моментом порядка s случайной величины X называется математическое ожидание величины $(X - m_x)^s$.

$$\mu_s[X] = M[(X - m_x)^s].$$

Для дискретной случайной величины

$$\mu_s[X] = \sum_{i=1}^n (x_i - m_x)^s P(X = x_i).$$

Для непрерывной случайной величины

$$\mu_s[X] = \int_{-\infty}^{\infty} (x - m_x)^s f(x) dx.$$

Центральный момент первого порядка всегда равен нулю, а центральный момент второго порядка равен дисперсии.

Модой M_0 дискретной случайной величины называется ее наиболее вероятное значение. Для непрерывной случайной величины мода – такое значение случайной величины, при которой плотность распределения вероятностей имеет максимум:

$$f(M_0) = \max f(x).$$

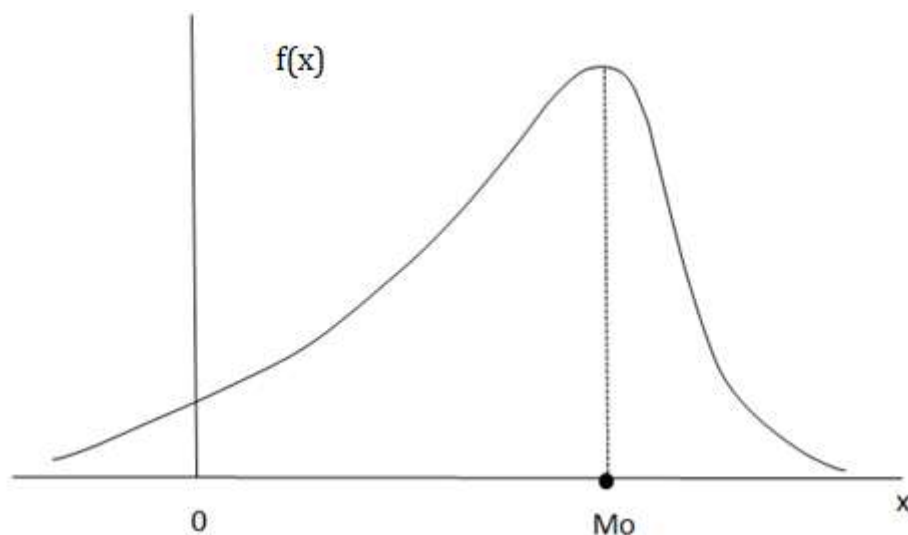


Рис. 9. Мода непрерывной случайной величины

Если многоугольник распределения для дискретной случайной величины или кривая распределения для непрерывной случайной величины имеет два или несколько максимумов, то такое распределение называется **двухмодальным** или **многомодальным**.

Медианой M_e случайной величины X называется такое ее значение, относительно которого равновероятно получение большего или меньшего значения случайной величины.

$$P(X < M_e) = P(X > M_e).$$

Геометрически медиана – абсцисса точки, в которой площадь, ограниченная кривой распределения, делится пополам.

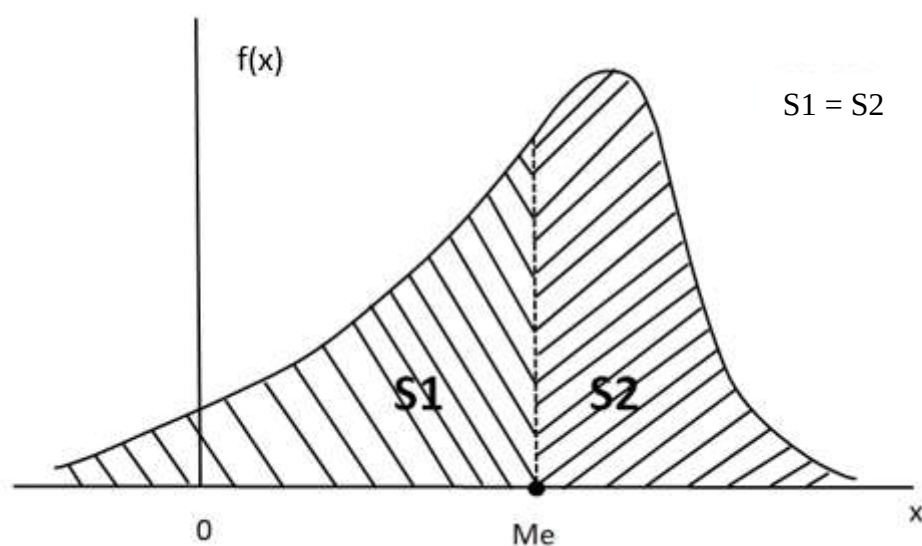


Рис. 10. Медиана непрерывной случайной величины

Свойства математического ожидания

1. Математическое ожидание постоянной величины равно самой постоянной.

$$M(C) = C.$$

2. Постоянный множитель можно выносить за знак математического ожидания.

$$M(CX) = CM(X).$$

3. Математическое ожидание произведения двух независимых случайных величин (вопрос независимости случайных величин будет рассмотрен в главе 5) равно произведению их математических ожиданий:

$$M(XY) = M(X)M(Y).$$

Это свойство справедливо для произвольного числа случайных величин.

4. Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий слагаемых:

$$M(X + Y) = M(X) + M(Y).$$

Это свойство также справедливо для произвольного числа случайных величин.

Свойства дисперсии

1. Дисперсия постоянной величины равна нулю:

$$D(C) = 0.$$

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

$$D(CX) = C^2 D(X).$$

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

$$D(X + Y) = D(X) + D(Y).$$

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

$$D(X - Y) = D(X) + D(Y).$$

Справедливость этого равенства вытекает из свойства 2.

Нормальный закон распределения

Нормальный закон распределения занимает центральное место в теории вероятностей. Это обусловлено тем, что этот закон проявляется во всех случаях, когда случайная величина является результатом действия большого числа различных факторов.

Плотность распределения вероятностей нормального закона распределения имеет вид

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Можно показать, что параметры m и σ , входящие в плотность распределения вероятностей, являются соответственно математическим ожиданием и средним квадратичным отклонением случайной величины X .

Кривая плотности распределения вероятностей нормального закона имеет симметричный вид. Максимальная ордината кривой, равная $\frac{1}{\sigma\sqrt{2\pi}}$, соответствует точке $x = m$. По мере удаления от точки m плотность распределения вероятностей падает и при $x \rightarrow \pm\infty$ асимптотически приближается к оси абсцисс.

Кривая симметрична относительно $x = m$. Математическое ожидание показывает положение максимума. Параметр σ характеризует форму кривой. Наибольшая ордината кривой обратно пропорциональна σ . При увеличении σ максимальная ордината уменьшается, а так как площадь под кривой плотности распределения вероятностей должна оставаться равной 1, то кривая становится более плоской, растягиваясь вдоль оси абсцисс. При уменьшении σ кривая плотности распределения вероятностей вытягивается вверх, одновременно сжимаясь с боков и становится более крутой.

Построим график функции плотности распределения вероятностей (рис. 11).

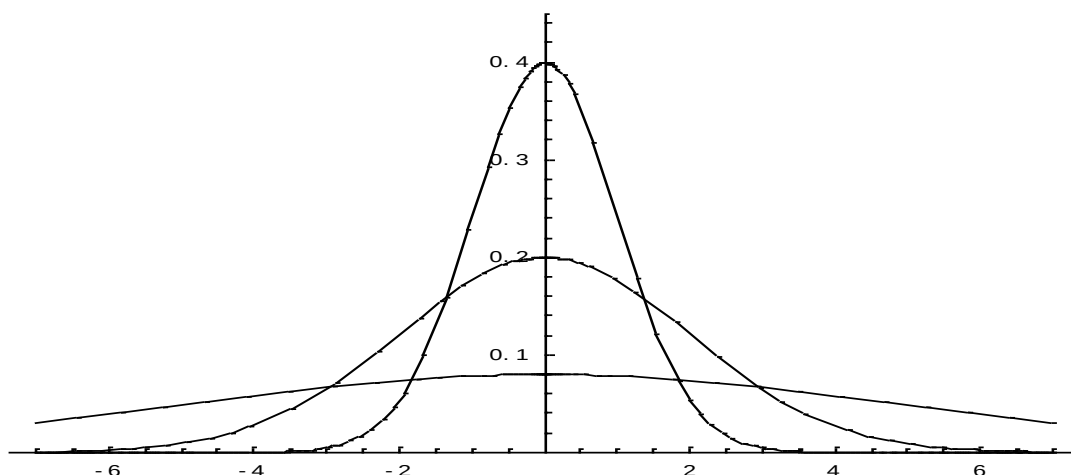


Рис. 11. Графики плотности распределения вероятностей нормального закона распределения при различных параметрах

Построены графики при $m = 0$ и трех возможных значениях среднего квадратичного отклонения $\sigma = 1$, $\sigma = 2$ и $\sigma = 7$. Как видно, при увеличении значения среднего квадратичного отклонения график становится более пологим, а максимальное значение уменьшается.

Если $m > 0$, то график сместится в положительном направлении, а если $m < 0$ – в отрицательном.

Законы распределения, связанные с нормальным

Распределение χ^2

Распределением χ^2 с n степенями свободы называется распределение суммы квадратов n независимых стандартных нормальных величин

$$\chi^2(n) = \sum_{i=1}^n Z_i^2$$

$$Z_i \sim N(0,1).$$

Распределение Стьюдента

Закону распределения Стьюдента с n степенями свободы удовлетворяет отношение

$$t(n) = \frac{Z}{\sqrt{\chi^2(n)/n}},$$

где Z – стандартная нормальная величина;

$\chi^2(n)$ – независимая от Z случайная величина, имеющая χ^2 -распределение с n степенями свободы.

При $n \rightarrow \infty$ распределение Стюдента приближается к нормальному. Практически уже при $n > 30$ распределение $t(n)$ можно считать нормальным.

Распределение Фишера

F -распределением с m, n степенями свободы называется распределение F -отношения:

$$F(m, n) = \frac{\chi^2(m)/m}{\chi^2(n)/n}.$$

Как видим, F -отношение – положительная случайная величина.

Сходимость по вероятности

Последовательность случайных величин $X_1, X_2, \dots, X_n, \dots$ сходится по вероятности к случайной величине X , если для любого $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$$

или

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$$

Сходимость по вероятности символически записывается как $X_n \Rightarrow X$.

2.3. Основные задачи статистики

Математическая статистика

Предположим, что задано некоторое множество, содержащее конечное число элементов, и мы выбираем наугад какой-нибудь элемент этого множества. При этом регистрируем значение какой-то характеристики (признака) этого элемента – x_i и затем возвращаем элемент обратно в множество. Заданное множество элементов будем называть генеральной совокупностью, а группу выбранных элементов – выборкой.

При этом нас интересуют не сами выбранные объекты, а только их характеристики – $x_1, x_2, \dots, x_n$. Последовательность n наблюдений $x_1, x_2, \dots, x_n$ будет рассматриваться как выборка из совокупности значений X , а генеральная совокупность объектов заменяется генеральной совокупностью значений их признаков.

Число объектов в генеральной совокупности и в выборке называют их объемами. Генеральная совокупность может быть как конечной, так и бесконечной.

В статистике выводы о генеральной совокупности делаются на основании изучения некоторой ее части (выборки). При этом рассматриваются две основные задачи:

– оценка параметров генеральной совокупности;

– проверка гипотез относительно каких-то параметров или свойств генеральной совокупности.

Оценка параметров

Предположим, что X – случайная величина с законом распределения $F(x, \theta)$, где θ – параметр закона распределения.

Функцию результатов наблюдения, с помощью которой судят о значении параметра θ , называют оценкой параметра θ . Эту оценку обозначают $\hat{\theta}$:

$$\hat{\theta} = f(x_1, x_2, \dots, x_n).$$

Она представляет собой случайную величину со своим законом распределения.

Если генеральная совокупность имеет нормальный закон распределения, нужно знать m и σ для полного определения $f(x)$.

Так как мы имеем выборку, то можем найти лишь выборочное среднее – $\bar{x}$ и выборочное среднее квадратичное отклонение – s , которые представляют собой оценки для m и σ , являясь при этом случайными величинами:

$$\hat{m} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$
$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}.$$

В качестве оценки математического ожидания выступает линейная функция от выборочных наблюдений с равными коэффициентами $1/n$, а в качестве оценки дисперсии – квадратичная функция от выборочных значений.

Как измерить близость оценки $\hat{\theta}$ к истинному значению θ или как определить качество оценки? Качество оценки устанавливают, проверяя, выполняются ли следующие три свойства: состоятельность, несмещенность и эффективность.

Оценка $\hat{\theta}$ называется *состоятельной*, если она сходится по вероятности к истинному значению

$$\hat{\theta} \Rightarrow \theta.$$

В случае использования состоятельных оценок оправдывается увеличение объема выборки, так как при этом становятся маловероятными значительные ошибки при оценивании.

Свойство состоятельности является обязательным, несостоятельные оценки не используются.

Оценка $\hat{\theta}$ называется **несмещенной**, если ее математическое ожидание равно истинному значению

$$M(\hat{\theta}) = \theta.$$

Если это равенство не выполняется, то оценка $\hat{\theta}$ будет в среднем либо завышать, либо занижать значение θ . Таким образом, требование несмещенности гарантирует отсутствие систематических ошибок при оценивании.

Свойство несмещенности желательно, но не обязательно. Часто полученная оценка бывает смещенной, но ее можно поправить так, чтобы она стала несмещенной.

Оценка $\hat{\theta}^*$ называется **эффективной** в определенном классе оценок Ω , если она самая точная среди оценок этого класса, т. е. имеет минимальную дисперсию:

$$D(\hat{\theta}^*) = \min_{\hat{\theta} \in \Omega} D(\hat{\theta}).$$

Проверка статистических гипотез

Под гипотезой понимают всякое высказывание о генеральной совокупности, проверяемое по выборке (гипотезы о законе распределения, гипотезы о параметрах).

Основную проверяемую гипотезу называют нулевой и обозначают H_0 . Наряду с выдвинутой гипотезой рассматривают и противоречащую ей гипотезу H_1 . Гипотеза H_1 называется альтернативной.

Проверка гипотезы осуществляется на основании выборки из n наблюдений $x_1, x_2, \dots, x_n$.

Для проверки нулевой гипотезы используют специально подобранную случайную величину (статистический критерий). В зависимости от вида распределения критерий обозначают через z (если он распределен по нормальному закону), t (если он распределен по закону Стьюдента), F (если он распределен по закону Фишера) и т. д. Поэтому для обозначения критерия в общем случае будем использовать букву K :

$$K = f(x_1, x_2, \dots, x_n).$$

После выбора критерия множество всех его возможных значений разбивают на два подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается (**критическая область** – S), а другое содержит те значения критерия, при которых гипотеза принимается (**область принятия гипотезы** – $\bar{S}$).

Основной принцип проверки статистических гипотез можно сформулировать так: если наблюдаемое значение критерия принадлежит критической об-

ласти, то нулевую гипотезу отвергают в пользу альтернативной гипотезы; если наблюдаемое значение критерия принадлежит области принятия гипотезы, то нулевую гипотезу принимают.

При этом может иметь место ошибка двух родов:

1) **ошибка первого рода** – будет принята гипотеза H_1 , тогда как на самом деле верной является гипотеза H_0 . Ее вероятность равна

$$\alpha = P(K \in S / H_0);$$

2) **ошибка второго рода** – будет принята гипотеза H_0 , тогда как на самом деле верной является гипотеза H_1 . Ее вероятность равна

$$\beta = P(K \in \bar{S} / H_1).$$

Величина α называется уровнем значимости (на практике $\alpha = 0,05$; $0,01$). Обычно по ней и определяется критическая область (рис. 12).

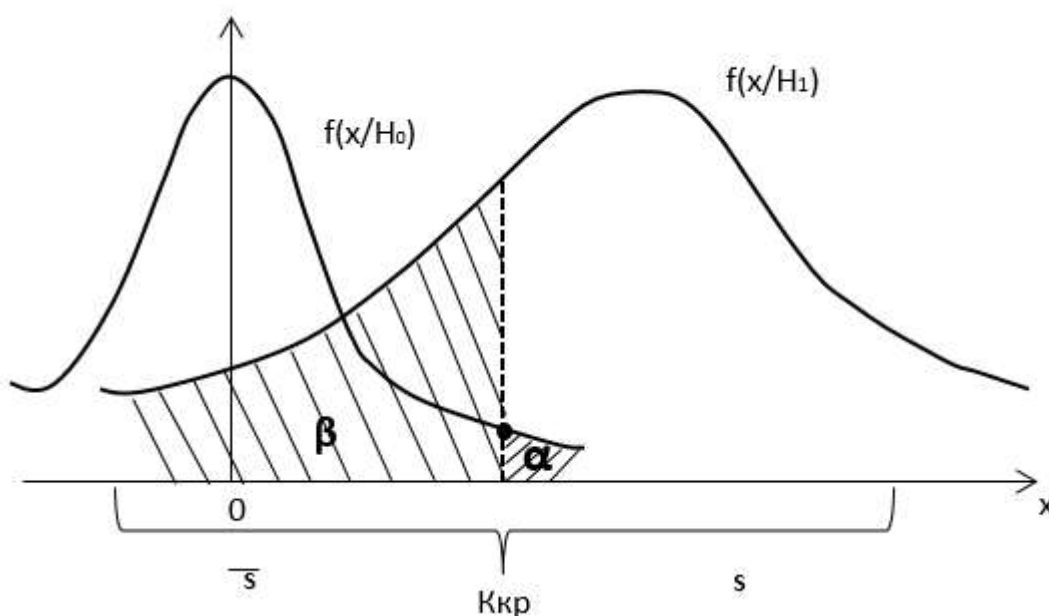


Рис. 12. Критическая область и ошибки первого и второго рода

Величина

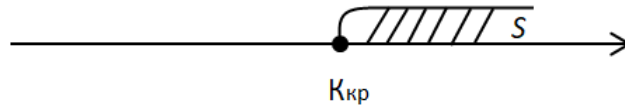
$1 - \beta = P(K \in S / H_1)$ называется мощностью критерия.

При фиксированном объеме выборки уменьшение вероятности ошибки 1-го рода α неизбежно приводит к увеличению вероятности ошибки 2-го рода β , и наоборот. Лишь при увеличении объема выборки возможно одновременное уменьшение вероятностей α и β .

Поскольку критерий K – одномерная случайная величина, все ее возможные значения принадлежат некоторому интервалу, поэтому критическую область и область принятия гипотезы можно разделить точками. **Критическими точками** $K_{кр}$ называют точки, отделяющие критическую область от области

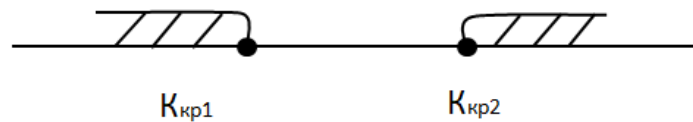
принятия гипотезы. Различают одностороннюю (правостороннюю и левостороннюю) и двустороннюю критические области.

Правосторонней называют критическую область, определяемую неравенством $K > K_{кр}$. При этом $P(K > K_{кр}/H_0) = \alpha$.



Левосторонней называют критическую область, определяемую неравенством $K < K_{кр}$. При этом $P(K < K_{кр}/H_0) = \alpha$.

Двусторонней называют критическую область, определяемую неравенствами $K < K_{кр1}$ и $K > K_{кр2}$. При этом $P(K > K_{кр2}/H_0) = P(K < K_{кр1}/H_0) = \alpha/2$.



ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 1

Для сигнализации об аварии установлены два независимо работающих сигнализатора. Вероятность того, что при аварии сигнализатор сработает, равна 0,95 для первого сигнализатора и 0,9 для второго. Найти вероятность того, что при аварии сработает только один сигнализатор.

Задача 2

В читальном зале имеется шесть учебников по теории вероятностей, из которых три в переплете. Библиотекарь наудачу взял два учебника. Найти вероятность того, что оба учебника окажутся в переплете.

Задача 3

Студент знает 20 из 25 вопросов программы. Найти вероятность того, что студент знает предложенные ему экзаменатором три вопроса.

Задача 4

На стеллаже библиотеки в случайном порядке расставлены 15 учебников, причем 5 из них в переплете. Библиотекарь берет наудачу 3 учебника. Найти вероятность того, что хотя бы один из взятых учебников окажется в переплете (событие A).

Задача 5

В ящике 10 деталей, из которых четыре окрашены. Сборщик наудачу взял три детали. Найти вероятность того, что хотя бы одна из взятых деталей окрашена.

Задача 6

В электрическую цепь последовательно включены три элемента, работающие независимо один от другого. Вероятности отказов первого, второго и третьего элементов соответственно равны: $p_1 = 0,1$; $p_2 = 0,15$; $p_3 = 0,2$. Найти вероятность того, что тока в цепи не будет.

Задача 7

Устройство состоит из трех элементов, работающих независимо. Вероятности безотказной работы (за время t) первого, второго и третьего элементов соответственно равны 0,6; 0,7; 0,8. Найти вероятности того, что за время t безотказно будут работать: а) только один элемент; б) только два элемента; в) все три элемента.

Задача 8

Вероятность хотя бы одного попадания стрелком в мишень при трех выстрелах равна 0,875. Найти вероятность попадания при одном выстреле.

Задача 9

Найти математическое ожидание дискретной случайной величины X , заданной законом распределения:

$$\begin{array}{l} \text{а) } X = 4 \quad 6 \quad 10 \\ p = 0,2 \quad 0,3 \quad 0,5 \end{array}$$

$$\begin{array}{l} \text{б) } X = 0,21 \quad 0,54 \quad 0,61 \\ p = 0,1 \quad 0,5 \quad 0,4 \end{array}$$

Задача 10

После ответа студента на вопросы экзаменационного билета экзаменатор задает студенту дополнительные вопросы. Преподаватель прекращает задавать дополнительные вопросы, как только студент обнаруживает незнание заданного вопроса. Вероятность того, что студент ответит на любой заданный дополнительный вопрос, равна 0,9. Требуется составить закон распределения случайной дискретной величины X – числа дополнительных вопросов, которые задаст преподаватель студенту.

Задача 11

Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leq -1, \\ \frac{3}{4}x + \frac{3}{4} & \text{при } -1 < x \leq 1/3, \\ 1 & \text{при } x > 1/3. \end{cases}$$

Найти вероятность того, что в результате испытания величина X примет значение, заключенное в интервале $(0, 1/3)$.

Задача 12

Дана функция распределения непрерывной случайной величины X

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \sin x & \text{при } 0 < x \leq \pi/2, \\ 1 & \text{при } x > \pi/2. \end{cases}$$

Найти плотность распределения $f(x)$.

Задача 13

Непрерывная случайная величина X в интервале $(0, \infty)$ задана плотностью распределения вероятностей $f(x) = ae^{-ax}$ ($a > 0$); вне этого интервала $f(x) = 0$. Найти вероятность того, что X примет значение, принадлежащее интервалу $(1, 2)$.

Задача 14

Задана плотность распределения непрерывной случайной величины X :

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \cos x & \text{при } 0 < x \leq \pi/2, \\ 0 & \text{при } x > \pi/2. \end{cases}$$

Найти функцию распределения $F(x)$.

Задача 15

Случайная величина X задана плотностью распределения $f(x) = 2x$ в интервале $(0, 1)$; вне этого интервала $f(x) = 0$. Найти математическое ожидание и дисперсию величины X .

Задача 16

Случайная величина X задана плотностью распределения $f(x) = 1/2x$ в интервале $(0, 2)$; вне этого интервала $f(x) = 0$. Найти математическое ожидание и дисперсию величины X .

3. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ХАРАКТЕРИСТИК ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ (МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ, ДИСПЕРСИИ, МОДЫ, МЕДИАНЫ, ДОЛИ) ПО ВЫБОРКЕ

3.1. Свойства выборочной средней и частоты

Свойства выборочной средней как оценки математического ожидания

Рассмотрим выборочную среднюю как оценку математического ожидания.

Для случайной выборки оценка математического ожидания имеет вид

$$\hat{m} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Несмещенность устанавливаем прямой проверкой:

$$M(\hat{m}) = M\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n M(X_i) = \frac{1}{n} \sum_{i=1}^n m = m.$$

Можно доказать, что $\hat{m}$ эффективна в классе линейных несмещенных оценок, а также является состоятельной.

Свойства частоты как оценки вероятности

Частота m/n сходится по вероятности к вероятности события

$$\frac{m}{n} \Rightarrow p.$$

Данная теорема имеет исключительно важное значение: для оценки вероятности случайного события A надо провести эксперимент из n испытаний и подсчитать частоту (долю положительных исходов). Эта частота и будет хорошей оценкой вероятности события $P(A) = p$ при больших n .

Это относится и к оценке доли объектов в генеральной совокупности, обладающих каким либо свойством, по выборочной доле

$$\frac{m}{n}.$$

3.2. Дисперсия оценок математического ожидания и доли объектов в генеральной совокупности при повторном и бесповторном отборах

Часто по выборке определяется среднее значение какого-либо признака — $\bar{x}$ (выборочное среднее) или выборочная доля элементов, обладающих каким-либо качественным признаком (w).

Разница между этими показателями в выборке и генеральной совокупности называется ошибкой оценки характеристик генеральной совокупности по выборке.

Выборочное среднее и выборочная доля являются случайными величинами, которые могут принимать различные значения в зависимости от того, какие элементы совокупности попали в выборку.

Следовательно, ошибки оценки этих характеристик генеральной совокупности по выборке также являются случайными величинами.

Среднее квадратичное отклонение оценок характеристик генеральной совокупности по выборке

1. Для оценки среднего значения генеральной совокупности по выборке

$$\sigma^2(\bar{x}) = \sigma^2\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2(X_i) = \frac{1}{n^2} n\sigma_x^2 = \frac{\sigma_x^2}{n}$$

или

$$\sigma(\bar{x}) = \sqrt{\frac{\sigma_x^2}{n}},$$

где σ_x — среднее квадратичное отклонение для самих данных;
 n — количество элементов в выборке.

2. Рассмотрим теперь оценку доли в генеральной совокупности по выборке.

Выборочная доля (частота события) определяется по формуле

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

где m — число случаев выпадения события A в n испытаниях.

Отметим, что

$$m = \sum_{i=1}^n X_i,$$

где $X_i = 0$, если случайное событие A не происходит и $X_i = 1$, если случайное событие A происходит в i -м эксперименте.

При этом

$$P(A) = p; \quad P(\bar{A}) = 1 - p = q.$$

Тогда

$$M(X_i) = 1 \cdot p + 0 \cdot (1 - p) = p;$$

$$D(X_i) = (1 - p)^2 p + (0 - p)^2 (1 - p) = (1 - p)p(1 - p + p) = p(1 - p).$$

Далее

$$D(w) = D\left(\frac{m}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{p(1 - p)}{n}.$$

Таким образом,

$$\sigma(w) = \sqrt{\frac{p(1 - p)}{n}},$$

где p – доля объектов, обладающих признаком, в генеральной совокупности.

В приведенных формулах σ_x и p являются характеристиками генеральной совокупности, которые при выборочном наблюдении неизвестны. Поэтому их заменяют аналогичными характеристиками выборочной совокупности – s_x и w .

Тогда

$$\hat{\sigma}(\bar{x}) = \sqrt{\frac{s_x^2}{n}},$$

$$\hat{\sigma}(w) = \sqrt{\frac{w(1 - w)}{n}}.$$

При бесповторном отборе подкоренное выражение умножается на величину

$$1 - \frac{n}{N},$$

где N – объем генеральной совокупности.

При $n \ll N$ выборку можно считать повторной.

Предельные ошибки оценок характеристик генеральной совокупности

Для решения практических задач необходимо знать не только среднюю квадратичную, но и предельную ошибку с гарантирующим ее уровнем довери-

тельной вероятности. Формулы для определения предельной ошибки Δ приведены в *табл. 10*.

Т а б л и ц а 10. Определение предельной ошибки

Метод отбора	Для средней	Для доли
Повторный	$\Delta = z_{\alpha} \sqrt{\frac{s_x^2}{n}}$	$\Delta = z_{\alpha} \sqrt{\frac{w(1-w)}{n}}$
Бесповторный	$\Delta = z_{\alpha} \sqrt{\frac{s_x^2}{n} \left(1 - \frac{n}{N}\right)}$	$\Delta = z_{\alpha} \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$

Величина z_{α} зависит от требуемого уровня доверительной вероятности и определяется по таблицам функции Лапласа.

Например, при $P = 0,95$ $z_{\alpha} = 1,96$, а при $P = 0,997$ $z_{\alpha} = 3$.

Определение численности выборки

Разрабатывая программу выборочного наблюдения, задают величину допустимой ошибки Δ и доверительную вероятность P . Неизвестным является тот минимальный объем выборки n , который должен обеспечить заданную точность.

Формулы для определения численности выборки приведены в *табл. 11*.

Т а б л и ц а 11. Определение численности выборки

Метод отбора	Для средней	Для доли
Повторный	$n = \frac{z_{\alpha}^2 s_x^2}{\Delta^2}$	$n = \frac{z_{\alpha}^2 w(1-w)}{\Delta^2}$
Бесповторный	$n = \frac{z_{\alpha}^2 s_x^2 N}{N\Delta^2 + z_{\alpha}^2 s_x^2}$	$n = \frac{z_{\alpha}^2 w(1-w)N}{N\Delta^2 + z_{\alpha}^2 w(1-w)}$

Величины s_x , w оцениваются по выборке меньшего размера. Часто в качестве w выбирается 0,5 (по наихудшему случаю).

Пример

Рассмотрим пример оценки доли полезного использования рабочего времени на предприятии (генеральной совокупности). Допустим, что предварительных данных об использовании рабочего времени нет. Допустимую ошибку установим в размере 0,05, а уровень значимости 0,05 (доверительная вероятность равна 0,95). Тогда необходимое число наблюдений составит

$$n = \frac{1,96^2(1-0,5)0,5}{0,05^2} \approx 400.$$

Допустим, что по данным 400 наблюдений получено, что время полезной работы наблюдается в 90 % наблюдений. Используя приведенную выше формулу, можно определить среднюю квадратичную ошибку оценки доли полезного времени:

$$\hat{\sigma}(w) = \sqrt{\frac{0,9(1-0,9)}{400}} = 0,015.$$

Предельная ошибка выборочной доли составляет 0,03. Поэтому с вероятностью 0,95 можно утверждать, что истинное значение доли полезно используемого рабочего времени находится в диапазоне (0,87 – 0,93), т. е. составляет от 87 до 93 %.

3.3. Определение выборочных статистических показателей по сгруппированным данным

Характеристики центра распределения

1. Средняя арифметическая

– для не сгруппированных данных $\bar{x} = \frac{\sum_{i=1}^n x_i}{n},$

– для сгруппированных данных $\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{n},$

где x_i – варианта или середина интервала i -й группы;
 n_i – частота i -й группы;
 k – количество групп.

2. Медиана $Me(x)$

Медиана представляет собой такое значение признака, которое делит объем совокупности пополам в том смысле, что число элементов совокупности со значениями признака, меньшими медианы, равно числу элементов совокупности со значениями признака, большими медианы.

Численное значение медианы можно определить по ряду накопленных частот. Накопленная частота для медианы равна половине объема совокупности:

$$F(Me(x)) = n/2.$$

Для интервального ряда сначала определяется интервал, в котором будет находиться медиана. Само же значение $Me(x)$ может быть приближенно определено с помощью интерполяции

$$Me(x) = x_0 + \Delta \frac{\frac{n}{2} - F(x_0)}{n_0},$$

где x_0 – начало интервала, содержащего медиану;
 Δ – величина интервала, содержащего медиану;
 $F(x_0)$ – накопленная частота на начало интервала, содержащего медиану;
 n – объем совокупности;
 n_0 – частота интервала, в котором расположена медиана.

3. Мода $Mo(X)$ – наиболее часто встречающееся значение признака в совокупности.

Для дискретного ряда – это то значение признака, которому соответствует наибольшая частота распределения.

Для интервального ряда вначале определяется интервал, содержащий моду (с наибольшей частотой). Затем приближенно вычисляется значение моды по формуле

$$Mo(X) = x_0 + \Delta \frac{n_0 - n_{-1}}{(n_0 - n_{-1}) + (n_0 - n_1)},$$

где x_0 – начало интервала, содержащего моду;
 Δ – величина интервала;
 n_0 – частота интервала, в котором расположена мода;
 n_{-1} – частота интервала, предшествующего модальному;
 n_1 – частота интервала, следующего за модальным.

Показатели вариации

1. **Выборочная дисперсия (s_x^2)** – это среднее значение квадратов отклонений индивидуальных значений признака от средней величины:

– для не сгруппированных данных

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n},$$

– для сгруппированных данных

$$s_x^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{n}.$$

Если ряд интервальный, то в качестве x_i берется середина i -го интервала. Более удобны следующие формулы вычислений:

$$s_x^2 = \frac{\sum_{i=1}^n x_i^2}{n} - (\bar{x})^2$$

(для не сгруппированных данных);

$$s_x^2 = \frac{\sum_{i=1}^k x_i^2 n_i}{n} - (\bar{x})^2$$

(для сгруппированных данных).

2. *Среднее квадратичное отклонение* (s_x) представляет собой квадратный корень из дисперсии:

$$s_x = \sqrt{s_x^2}.$$

3. *Коэффициент вариации* характеризует относительную величину варируемости признака в данной совокупности по отношению к среднему значению

$$cv = \frac{s_x}{\bar{x}} 100 \%.$$

Пример

Имеются сгруппированные данные по зарплате

Зарплата, тыс. руб.	Середина интервала x_i	Частоты n_i	Накопленные частоты F_i
8,6–9,4	9,0	2	2
9,4–10,2	9,8	6	8
10,2–11,0	10,6	15	23
11,0–11,8	11,4	23	46
11,8–12,6	12,2	25	71
12,6–13,4	13,0	17	88
13,4–14,2	13,8	7	95
14,2–15,0	14,6	5	100
Итого		100	

Найдем медиану. В данном случае $\frac{n}{2} = 50$. Эта величина больше 46, но меньше 71, следовательно, медиана находится в интервале (11,8–12,6). Рассчитаем ее значение:

$$Me(x) = x_0 + \Delta \frac{\frac{n}{2} - F(x_0)}{n_0} = 11,8 + 0,8 \frac{50 - 46}{25} = 11,93.$$

Найдем моду по этим данным. Мода находится в том же интервале, так как максимальная частота (25) приходится на этот интервал.

$$Mo(X) = x_0 + \Delta \frac{n_o - n_{-1}}{(n_o - n_{-1}) + (n_0 - n_1)} = 11,8 + 0,8 \frac{25 - 23}{(25 - 23) + (25 - 17)} = 11,96.$$

Средняя арифметическая

$$\bar{x} = \frac{9,0 \cdot 2 + 9,8 \cdot 6 + 10,6 \cdot 15 + 11,4 \cdot 23 + 12,2 \cdot 25 + 13,0 \cdot 17 + 13,8 \cdot 7 + 14,6 \cdot 5}{100} = 11,94.$$

Выборочная дисперсия

$$s_x^2 = \frac{\sum_{i=1}^k x_i^2 n_i}{n} - (\bar{x})^2 = \frac{9,0^2 \cdot 2 + 9,8^2 \cdot 6 + \dots + 14,6^2 \cdot 5}{100} - 11,94^2 = 1,5.$$

Среднее квадратичное отклонение

$$s_x = \sqrt{1,5} = 1,225.$$

Коэффициент вариации

$$cv = \frac{1,225}{11,94} 100\% = 10,3 \ %.$$

3.4. Абсолютные и относительные статистические показатели.

Вычисление средних значений относительных показателей

Под абсолютными показателями в статистике понимают исходные показатели статистического наблюдения (объем продукции, количество населения и т. д.). Они могут быть как моментными (на определенный момент времени), так и интервальными (за определенный период). Любая абсолютная величина (показатель) имеет присущую ей единицу измерения (штуки, килограммы, метры и т. д.). Часто в качестве абсолютных показателей используют стоимостные показатели (в рублях).

Под относительными показателями в статистике понимают показатели, характеризующие соотношение двух абсолютных показателей (ВНП на душу населения, производительность труда, себестоимость продукции и т. д.).

Различают относительные величины структуры, координации, динамики, сравнения и интенсивности.

Относительные величины структуры показывают долю каждой группы в общей численности совокупности. Их получают путем деления численности каждой группы на численность всей совокупности.

Относительные величины координации получают как соотношение между частями одной совокупности. Например, это может быть отношение числа мужчин к числу женщин.

Относительные величины динамики – это результат сопоставления уровней одного и того же показателя в разные моменты или периоды времени. Например, сопоставляя объем добычи нефти в России в 2019 и 2018 гг., получим относительную величину динамики.

Относительные величины сравнения получают в результате сопоставления двух одноименных показателей, относящихся к разным совокупностям. Например, при сравнении величины основных фондов двух разных регионов.

Относительные величины интенсивности получают, сопоставляя разноименные признаки одной совокупности. Например, коэффициент рождаемости равен отношению числа родившихся детей к числу жителей, а себестоимость продукции равна отношению полных затрат к объему выпуска продукции.

Для расчета **средних значений относительных величин** используются формулы различных взвешенных средних в зависимости от экономического смысла показателей. В статистике используются различные виды средних величин.

Наиболее часто применяются следующие средние величины:

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

Все указанные средние величины можно рассчитать по общей формуле **степенной средней**

$$\bar{x} = \sqrt[m]{\frac{\sum_{i=1}^n x_i^m}{n}}.$$

Если данные сгруппированы, то

$$\bar{x} = \sqrt[m]{\frac{\sum_{i=1}^k x_i^m n_i}{n}}.$$

Последние две формулы позволяют получить различные виды средних при разных значениях m (табл. 12).

Т а б л и ц а 12. **Формулы для расчета простой и взвешенной средней**

Вид степенной средней	m	Формула расчета простой средней	Формула расчета взвешенной средней
Гармоническая	-1	$\bar{x} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$	$\bar{x} = \frac{\sum_{i=1}^n V_i}{\sum_{i=1}^n \frac{V_i}{x_i}}$
Геометрическая	$\rightarrow 0$	$\bar{x} = \sqrt[n]{x_1 x_2 \dots x_n}$	$\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i^{V_i}}$
Арифметическая	1	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^n V_i x_i}{\sum_{i=1}^n V_i}$
Квадратическая	2	$\bar{x} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$	$\bar{x} = \sqrt{\frac{\sum_{i=1}^n V_i x_i^2}{\sum_{i=1}^n V_i}}$

Средняя арифметическая, средняя гармоническая, средняя геометрическая и средняя квадратическая, рассчитанные для одних и тех же исходных данных, отличаются друг от друга. При этом всегда выполняется следующее соотношение:

$$\bar{x}_{\text{гармоническая}} < \bar{x}_{\text{геометрическая}} < \bar{x}_{\text{арифметическая}} < \bar{x}_{\text{квадратическая}}.$$

Приведем несколько примеров использования средних взвешенных зависимостей.

Пример 1. Найдем средний коэффициент выполнения плана по предприятиям отрасли.

Пусть N_i – план i -го предприятия;

α_i – относительный показатель выполнения плана (в долях);

n – число предприятий отрасли.

Тогда фактический объем выпуска продукции составит

$$\sum_{i=1}^n N_i \alpha_i.$$

Плановый объем выпуска продукции по отрасли

$$\sum_{i=1}^n N_i.$$

Средний показатель выполнения плана по отрасли

$$\bar{a} = \frac{\sum_{i=1}^n N_i \alpha_i}{\sum_{i=1}^n N_i}.$$

Этот показатель представляет собой средневзвешенное арифметическое показателей α_i с весами, соответствующими плану производства – N_i .

Пример 2. Найдем среднюю скорость движения автомобиля, если он проехал расстояние S_1 со скоростью v_1 , а затем расстояние S_2 со скоростью v_2 . Для нахождения средней скорости надо разделить суммарное расстояние $S_1 + S_2$ на суммарное время, затраченное на этот путь. Суммарное время в пути будет равно

$$\frac{S_1}{v_1} + \frac{S_2}{v_2}.$$

Таким образом, средняя скорость составит

$$\bar{v} = \frac{S_1 + S_2}{S_1/v_1 + S_2/v_2}.$$

В общем случае при наличии n участков с различной скоростью

$$\bar{v} = \frac{\sum_{i=1}^n S_i}{\sum_{i=1}^n \frac{S_i}{v_i}}.$$

Нетрудно видеть, что средняя скорость представляет собой средневзвешенное гармоническое из скоростей на отдельных участках пути с весами, равными длине участков пути.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 1

В итоге пяти измерений длины стержня одним прибором (без систематических ошибок) получены следующие результаты (в мм): 92; 94; 103; 105; 106. Найти: а) выборочную среднюю длину стержня; б) выборочную дисперсию ошибок прибора.

Задача 2

В итоге четырех измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: 8; 9; 11; 12. Найти: а) выборочную среднюю результатов измерений; б) выборочную дисперсию ошибок прибора.

Задача 3

Результаты моментного наблюдения за поведением покупателей в магазине самообслуживания приведены ниже.

Код действия покупателя	1	2	3	4	5	6	7
Количество покупателей	215	105	200	46	100	473	15

- 1 – ищут нужный отдел;
- 2 – подходят к прилавку;
- 3 – изучают ассортимент товаров и их цены;
- 4 – выбирают необходимый товар;
- 5 – переносят товар к кассе;
- 6 – оплачивают товар;
- 7 – выходят из магазина.

Найти выборочную долю покупателей, которые в момент обследования ищут нужный отдел, и предельную ошибку для оценки доли в генеральной совокупности с доверительной вероятностью $P = 0,95$.

Задача 4

Имеются сгруппированные данные по заработной плате. Рассчитайте моду и медиану.

Зарплата, тыс. руб.	Середина интервала x_i	Частоты n_i	Накопленные частоты F_i
10–20	15	2	
20–30	25	6	
30–40	35	15	
40–50	45	23	
50–60	55	25	
60–70	65	17	
70–80	75	7	
80–90	85	5	
Итого			

Задача 5

Имеются сгруппированные данные по заработной плате. Рассчитайте среднее значение, моду и медиану.

Зарплата, тыс. руб.	Середина интервала x_i	Частоты n_i	Накопленные частоты F_i
10–20	15	41	
20–30	25	40	
30–40	35	34	
40–50	45	17	
50–60	55	11	
60–70	65	43	
70–80	75	32	
80–90	85	27	
Итого			

Задача 6

Имеются сгруппированные данные по заработной плате. Рассчитайте показатели вариации (дисперсию, среднее квадратичное отклонение, коэффициент вариации).

Зарплата, тыс. руб.	Середина интервала x_i	Число предприятий n_i
10–20	15	20
20–30	25	29
30–40	35	18
40–50	45	26
50–60	55	18
60–70	65	16
70–80	75	15
80–90	85	32
Итого		

Задача 7

В таблице представлены исходные данные по объемам реализации товара в сгруппированном виде, рассчитайте выборочное среднее и выборочную дисперсию.

Объем реализации товара, млн руб.	Число предприятий n_i
10–15	4
15–20	5
20–25	9
25–30	2

Задача 8

В таблице представлены исходные данные по объемам реализации товара в сгруппированном виде, рассчитайте выборочную дисперсию и среднее квадратичное отклонение.

Объем реализации товара, млн руб.	Число предприятий n_i
10–20	10
20–30	9
30–40	17
40–50	20

Задача 9

В таблице представлены исходные данные по объемам реализации товара в сгруппированном виде, рассчитайте коэффициент вариации.

Объем реализации товара, млн руб.	Число предприятий F_i
50–57	10
57–64	9
64–71	11
71–78	10

Задача 10

В таблице приведены данные о продажах одежды в одном из магазинов города за 1 квартал прошедшего года. Определите долю продаж каждого товара (относительный показатель структуры).

Товар	Число проданных товаров, тыс. шт.	Доля
Джинсы	126	
Кофты	215	
Брюки	210	
Пиджаки	432	
Лосины	211	
Итого		

Задача 11

По региону имеются следующие данные об объеме продаж одежды. Определить долю продаж отдельных товаров (относительный показатель структуры) в 2021 и 2022 гг.

Товар	Продажи			
	2021 г.		2022 г.	
	Штуки	Доля	Штуки	Доля
Джинсы	32		35	
Кофты	19		21	
Брюки	32		35	
Итого				

Задача 12

По региону имеются следующие данные об объеме продаж ткани. Определить долю продаж (относительный показатель структуры) в 2021 и 2022 гг.

Товар	Продажи			
	2021 г.		2022 г.	
	м ²	Доля	м ²	Доля
Трикотин	71		69	
Туаль	207		241	
Шифон	68		52	
Итого				

Задача 13

Определить среднюю заработную плату по всем предприятиям с использованием формулы средней гармонической.

Номер предприятия	Месячный фонд заработной платы, тыс. руб.	Средняя заработная плата, тыс. руб.
1	6000	50,0
2	6240	48,0
3	5500	50,0

Задача 14

Даны значения темпов роста прибыли предприятия за первые пять месяцев года: 100 %; 102,9 %; 102,4 %; 99,1 %; 100,8 %. Определить средний темп роста с использованием формулы средней геометрической.

Задача 15

Провели опрос о желаемом размере заработной платы у пяти сотрудников офиса. По результатам опроса выяснили, что желаемый размер заработной платы составляет соответственно для каждого сотрудника: 70 000, 80 000, 100 000, 90 000, 80 000 руб. Определить среднюю желаемую заработную плату с использованием формулы средней арифметической.

Задача 16

Имеются 4 цеха промышленного предприятия в виде квадрата со сторонами: 100, 200, 300, 200 м. Определить среднюю площадь цеха данного предприятия с использованием формулы средней квадратической.

Задача 17

Имеются следующие данные по фермерским хозяйствам области.

Группы хозяйств по себестоимости 1 т сахарной свеклы, тыс. руб.	Число хозяйств	Валовой сбор в среднем на 1 хозяйство, т
20–22	30	110
22–24	60	90
24–26	120	120
26–28	20	130

Определите среднюю себестоимость 1 т свеклы в целом по фермерским хозяйствам области.

Задача 18

Производственная деятельность предприятий корпорации за месяц характеризуется следующими данными.

Предприятие	Общие затраты на производство, тыс. руб.	Затраты на единицу произведенной продукции, тыс. руб.
1	2000	7,5
2	8000	7,1
3	4000	7,3
4	3500	7,8

Определите средние затраты на единицу произведенной продукции в целом по корпорации.

Задача 19

По трем районам города имеются следующие данные (на конец года).

Район	Число отделений Сбербанка	Среднее число вкладов в отделении	Средний размер вклада, тыс. руб.
1	5	1300	278
2	10	1600	325
3	5	1200	256

Определите средний размер вклада в Сбербанке в целом по городу.

Задача 20

Качество продукции предприятия характеризуется следующими данными (за месяц).

Вид продукции	Процент брака	Стоимость бракованной продукции, руб.
A	1,0	20 000
B	0,9	35 000
C	2,0	10 000

Определите средний процент брака в целом по предприятию.

4. ПРИМЕНЕНИЕ ИНДЕКСОВ В ЭКОНОМИЧЕСКОМ АНАЛИЗЕ

4.1. Индивидуальные и агрегатные индексы

Индивидуальные индексы

В статистике под индексом понимают относительную величину, характеризующую результат сравнения двух уровней одноименных показателей.

Каждый индекс включает два вида данных: данные текущего (отчетного) периода и данные базисного периода, служащие базой для сравнения.

Различают индивидуальные и общие (агрегатные) индексы.

Индивидуальные индексы характеризуют изменение отдельных элементов статистической совокупности

$$i_x = \frac{x_1}{x_0},$$

где x_1 – текущий уровень величины;

x_0 – базисный уровень величины.

Различают индексы абсолютных показателей (объема продукции, товарооборота и т. д.) и относительных показателей (цен, себестоимости и т. д.).

Общие (агрегатные) индексы

Общие (агрегатные) индексы строятся с учетом изменения всех элементов статистической совокупности.

А) Агрегатный индекс товарооборота

$$I_{pq} = \frac{\sum_{j=1}^n p_{1j} q_{1j}}{\sum_{j=1}^n p_{0j} q_{0j}},$$

где n – количество товаров;

p_{1j}, p_{0j} – цена j -го товара в текущем и базовом периодах соответственно;

q_{1j}, q_{0j} – количество j -го товара в текущем и базовом периодах соответственно.

Б) Агрегатный индекс цен

Так как совокупность состоит из элементов, непосредственно не поддающихся суммированию, то агрегатный индекс включает набор значений цен и соответствующих им весовых коэффициентов

$$I_p = \frac{\sum_{j=1}^n p_{1j} W_j}{\sum_{j=1}^n p_{0j} W_j}.$$

При этом возникает вопрос что использовать в качестве весов W_j . В экономическом анализе приняты два варианта.

В первом варианте в качестве весов принимается физический объем производства текущего периода (индекс Пааше)

$$I_p = \frac{\sum_{j=1}^n p_{1j} q_{1j}}{\sum_{j=1}^n p_{0j} q_{1j}}.$$

Во втором варианте в качестве весов принимается физический объем производства базисного периода (индекс Ласпейреса)

$$I_p = \frac{\sum_{j=1}^n p_{1j} q_{0j}}{\sum_{j=1}^n p_{0j} q_{0j}}.$$

Обычно на практике используется индекс Пааше.

В) Индекс физического объема

Индекс физического объема использует в качестве весов цены базисного периода

$$I_p = \frac{\sum_{j=1}^n p_{0j} q_{1j}}{\sum_{j=1}^n p_{0j} q_{0j}}.$$

Тогда индекс товарооборота будет равен

$$I_{pq} = \frac{\sum_{j=1}^n p_{1j} q_{1j}}{\sum_{j=1}^n p_{0j} q_{0j}} = \frac{\sum_{j=1}^n p_{1j} q_{1j}}{\sum_{j=1}^n p_{0j} q_{1j}} \cdot \frac{\sum_{j=1}^n p_{0j} q_{1j}}{\sum_{j=1}^n p_{0j} q_{0j}} = I_p \cdot I_q.$$

4.2. Взаимосвязь агрегатных и индивидуальных индексов

Агрегатный индекс связан с индивидуальными индексами. При этом агрегатный индекс является некоторой средней из индивидуальных индексов с соответствующими весами.

Предположим, что известны индивидуальные индексы цен $i_{p1}, i_{p2}, \dots, i_{pn}$. Тогда

$$I_p = \frac{\sum_{j=1}^n p_{1j} q_{1j}}{\sum_{j=1}^n p_{0j} q_{1j}} = \frac{\sum_{j=1}^n p_{1j} q_{1j}}{\sum_{j=1}^n \frac{p_{1j}}{i_{pj}} q_{1j}}.$$

Данная формула представляет собой средневзвешенное гармоническое из индивидуальных индексов цен с весами $p_{1j}q_{1j}$.

Предположим, что известны индивидуальные индексы физического объема $i_{q1}, i_{q2}, \dots, i_{qn}$.

Тогда

$$I_q = \frac{\sum_{j=1}^n p_{0j}q_{1j}}{\sum_{j=1}^n p_{0j}q_{0j}} = \frac{\sum_{j=1}^n p_{0j}q_{0j}i_{qj}}{\sum_{j=1}^n p_{0j}q_{0j}}.$$

Данная формула представляет собой средневзвешенное арифметическое из индивидуальных индексов физических объемов с весами $p_{0j}q_{0j}$.

4.3. Индексы постоянного состава и структурных сдвигов

Агрегатный индекс средних величин

Рассмотрим агрегатный индекс средних величин на примере индекса средней цены

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\sum_{j=1}^n p_{1j}q_{1j}}{\sum_{j=1}^n q_{1j}} / \frac{\sum_{j=1}^n p_{0j}q_{0j}}{\sum_{j=1}^n q_{0j}}.$$

Отметим, что на величину индекса средней цены влияет как непосредственное изменение цен, так и изменение структуры потребляемой продукции. Поэтому индекс средней величины называют индексом переменного состава.

Индекс постоянного состава (индекс Пааше) имеет вид

$$I_p = \frac{\sum_{j=1}^n p_{1j}q_{1j}}{\sum_{j=1}^n p_{0j}q_{1j}}.$$

Индекс структурных сдвигов имеет вид

$$I_{c/c} = \frac{\sum_{j=1}^n p_{0j}q_{1j}}{\sum_{j=1}^n q_{1j}} / \frac{\sum_{j=1}^n p_{0j}q_{0j}}{\sum_{j=1}^n q_{0j}}.$$

При этом выполняется следующее соотношение

$$I_{\bar{p}} = I_p I_{c/c}.$$

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 1

Данные о количестве проданных товаров и ценах в базисном и текущем периодах приведены в таблице. Найти индекс цен Пааше и Ласпейреса.

Товар	Продано, м ²		Цена за 1 м ² , д. е.	
	Базисный период	Текущий период	Базисный период	Текущий период
Хлопок	4	5	1176	1207
Лён	160	180	1233	1244
Атлас	3	4	1656	1698

Задача 2

Данные о количестве проданных товаров и ценах в базисном и текущем периодах приведены в таблице.

Товар	Продано, кг		Цена за 1 кг, д. е.	
	Базисный период	Текущий период	Базисный период	Текущий период
Мясо	4000	5000	550	700
Картофель	160 000	180 000	30	40
Творог	300	400	60	85

На основании вышеприведенных данных вычислить:

- 1) индивидуальные индексы цен и физического объема;
- 2) агрегатные индексы товарооборота, цен и физического объема.

Задача 3

Данные о реализации фруктов и овощей на рынках города приведены в таблице.

Вид фрукта	Товарооборот, д. е.		Процент изменения количества проданных фруктов
	Базисный период	Отчетный период	
Слива	1400	1480	-8
Груша	750	1040	+4
Яблоко	1300	1560	Без изменения

На основании приведенных данных вычислите:

- 1) индивидуальные индексы физического объема;
- 2) агрегатный индекс физического объема;
- 3) агрегатный индекс товарооборота;
- 4) на основании агрегатных индексов товарооборота и физического объема найти агрегатный индекс цен.

Задача 4

Данные о реализации товаров в универсаме приведены в таблице.

Товар	Товарооборот, д. е.		Процент изменения цен проданных товаров
	Базисный период	Отчетный период	
Мясо	8450	9903	+10,3
Масло	5930	6191	+3,8
Овощи	2367	2596	-1,5

На основании приведенных данных вычислите:

- 1) индивидуальные индексы цен;
- 2) агрегатный индекс цен;
- 3) агрегатный индекс товарооборота;
- 4) на основании агрегатных индексов товарооборота и цен найти агрегатный индекс физического объема.

5. ЗАВИСИМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

5.1. Система случайных величин

Рассмотренные выше случайные величины были одномерными, т. е. определялись одним числом, однако существуют также случайные величины, которые определяются двумя, тремя и т. д. числами. Такие случайные величины называются двумерными, трехмерными и т. д.

В зависимости от типа входящих в систему случайных величин, системы могут быть дискретными, непрерывными или смешанными, если в систему входят различные типы случайных величин.

Более подробно рассмотрим системы двух случайных величин.

Функцией распределения вероятностей системы двух случайных величин (X, Y) называется вероятность совместного выполнения двух неравенств $X < x, Y < y$.

$$F(x, y) = P(X < x \cup Y < y).$$

Отметим следующие свойства функции распределения вероятностей системы двух случайных величин:

1. $0 \leq F(x, y) \leq 1$, так как эта функция равна вероятности.
2. Если один из аргументов стремится к плюс бесконечности, то функция распределения вероятностей системы стремится к функции распределения вероятностей одной случайной величины, соответствующей другому аргументу:

$$\lim_{y \rightarrow \infty} F(x, y) = F_1(x);$$

$$\lim_{x \rightarrow \infty} F(x, y) = F_2(y).$$

3. Событие $(X < \infty \cap Y < \infty)$ является достоверным, поэтому если оба аргумента стремятся к бесконечности, то функция распределения вероятностей системы стремится к единице:

$$F(\infty, \infty) = 1.$$

4. События $(X < -\infty), (Y < -\infty)$ являются невозможными, поэтому при стремлении одного или обоих аргументов к минус бесконечности функция распределения стремится к нулю:

$$F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0.$$

5. Функция распределения вероятностей является неубывающей функцией по каждому аргументу.

С помощью функции распределения системы можно найти вероятность того, что случайная точка (X, Y) попадет в прямоугольную область (рис. 13).

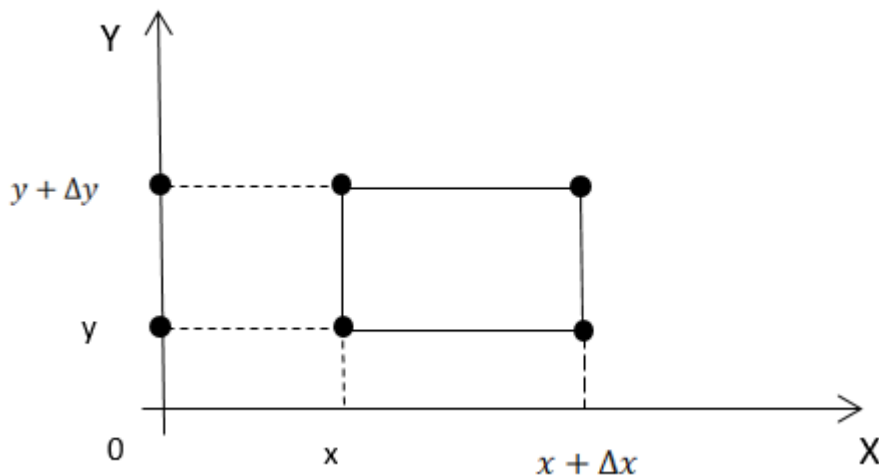


Рис. 13. Прямоугольная область попадания случайной величины

Введем в рассмотрение три случайных события:

$$A = (x \leq X < x + \Delta x) \cap (y \leq Y < y + \Delta y),$$

$$B = (X \leq x) \cap (Y < y + \Delta y),$$

$$C = (X < x + \Delta x) \cap (Y < y).$$

Тогда объединение этих событий

$$A \cup B \cup C = (X < x + \Delta x) \cap (Y < y + \Delta y).$$

События A и $B \cup C$ несовместны, поэтому

$$P(A \cup B \cup C) = P(A) + P(B \cup C) = P(A) + P(B) + P(C) - P(B \cap C),$$

но событие $B \cap C = (X < x) \cap (Y < y)$.

Таким образом, получаем

$$P(X < x + \Delta x \cap Y < y + \Delta y) = P(x \leq X < x + \Delta x \cap y \leq Y < y + \Delta y) +$$

$$+ P(X < x \cap Y < y + \Delta y) + P(X < x + \Delta x \cap Y < y) - P(X < x \cap Y < y)$$

или $F(x + \Delta x, y + \Delta y) = P(x \leq X < x + \Delta x \cap y \leq Y < y + \Delta y) +$

$$+ F(x, y + \Delta y) + F(x + \Delta x, y) - F(x, y).$$

Отсюда следует, что

$$P(x \leq X < x + \Delta x \cap y \leq Y < y + \Delta y) = F(x + \Delta x, y + \Delta y) -$$

$$- F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y). \quad (1)$$

Выведем формулу для плотности распределения вероятностей системы $f(x, y)$

$$f(x, y) = \lim_{\Delta x \Delta y} \frac{P(x \leq X < x + \Delta x \cap y \leq Y < y + \Delta y)}{\Delta x \Delta y}. \quad (2)$$

Используя формулу (2) и правило последовательного перехода к пределу, получим

$$f(x, y) = \lim_{\Delta x \Delta y} \frac{F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y)}{\Delta x \Delta y} =$$

$$= \lim_{\Delta y} \frac{1}{\Delta y} \left(\lim_{\Delta x} \frac{F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y)}{\Delta x} - \right.$$

$$\left. - \lim_{\Delta x} \frac{F(x + \Delta x, y) - F(x, y)}{\Delta x} \right) = \lim_{\Delta y} \frac{\frac{\partial F(x, y + \Delta y)}{\partial x} - \frac{\partial F(x, y)}{\partial x}}{\Delta y} =$$

$$= \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Плотностью распределения вероятностей системы (X, Y) называется вторая смешанная частная производная от функции распределения вероятностей:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Если известна плотность распределения вероятностей, то функция распределения вероятностей может быть найдена по формуле

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(x, y) dx dy.$$

Рассмотрим свойства плотности распределения вероятностей системы.

1. Плотность распределения вероятностей системы есть функция неотрицательная:

$$f(x, y) \geq 0.$$

Это ясно из того, что плотность распределения есть предел отношения двух неотрицательных величин – вероятности попадания в прямоугольник и площади прямоугольника. Следовательно, отрицательной быть не может.

2. Двойной интеграл с бесконечными пределами от плотности распределения вероятностей системы равен единице:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$$

По известной плотности совместного распределения можно найти плотности распределения каждой из составляющих двумерной случайной величины:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy; \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx;$$

а также функции распределения вероятностей

$$F_1(x) = F(x, \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(x, y) dy dx; \quad F_2(y) = F(\infty, y) = \int_{-\infty}^y \int_{-\infty}^{\infty} f(x, y) dx dy.$$

5.2. Условные законы распределения

Как было показано выше, зная совместный закон распределения, можно легко найти законы распределения каждой случайной величины, входящей в систему.

Однако на практике чаще стоит обратная задача – по известным законам распределения случайных величин найти их совместный закон распределения.

Для того чтобы исчерпывающим образом охарактеризовать систему, недостаточно знать распределение каждой из величин, входящих в систему, нужно еще знать зависимость между величинами, входящими в систему. Эта зависимость может быть охарактеризована с помощью так называемых условных законов распределения.

Распределение одной случайной величины, входящей в систему, найденное при условии, что другая случайная величина приняла определенное значение, называется *условным законом распределения*.

Условный закон распределения можно задавать как функцией распределения вероятностей, так и плотностью распределения вероятностей. Условная функция распределения вероятностей обозначается $F(x/y)$, а условная плотность распределения вероятностей $f(x/y)$. Так как системы непрерывных величин имеют основное практическое значение, ограничимся рассмотрением условных законов, заданных плотностью распределения вероятностей.

Условная плотность распределения вероятностей обладает всеми свойствами плотности распределения одной случайной величины.

Имеет место следующее соотношение, аналогичное соотношению для вероятности пересечения случайных событий

$$f(x, y) = f_1(x) f_2(y / x).$$

5.3. Зависимые и независимые случайные величины.

Коэффициент корреляции

Случайные величины называются независимыми, если закон распределения одной из них не зависит от того, какое значение принимает другая случайная величина.

Условные распределения независимых случайных величин равны их безусловным распределениям, т. е. $f_2(y / x) = f_2(y)$.

Для того чтобы случайные величины X и Y были независимы, необходимо и достаточно, чтобы плотность распределения вероятностей системы (X, Y) была равна произведению плотностей распределения вероятностей составляющих переменных:

$$f(x, y) = f_1(x) f_2(y).$$

Рассмотрим начальные и центральные моменты системы (X, Y) .

Начальным моментом порядка k, s системы (X, Y) называется математическое ожидание произведения X^k на Y^s :

$$m_{k,s} = M[X^k Y^s].$$

Центральным моментом порядка k, s системы (X, Y) называется математическое ожидание произведения k -й и s -й степеней соответствующих центрированных величин:

$$\mu_{k,s} = M[(X - m_x)^k (Y - m_y)^s].$$

Особую роль играет второй смешанный центральный момент, который называется ковариацией:

$$\text{cov}(X, Y) = M[(X - M(X))(Y - M(Y))].$$

Для непрерывных случайных величин:

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)(y - m_y) f(x, y) dx dy. \quad (3)$$

Ковариация служит для того, чтобы охарактеризовать связь между случайными величинами. Если случайные величины независимы, то их ковариация равна нулю.

Доказательство. Пусть X и Y – независимые непрерывные случайные величины с плотностью распределения вероятностей $f(x, y)$. Для независимых случайных величин

$$f(x, y) = f_1(x) f_2(y). \quad (4)$$

Подставляя выражение (4) в формулу (3), видим, что интеграл превращается в произведение двух интегралов:

$$\begin{aligned} \text{cov}(X, Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)(y - m_y) f(x, y) dx dy = \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} (x - m_x)(y - m_y) f_1(x) f_2(y) dx \right] dy = \\ &= \int_{-\infty}^{\infty} (y - m_y) f_2(y) \left[\int_{-\infty}^{\infty} (x - m_x) f_1(x) dx \right] dy = \\ &= \int_{-\infty}^{\infty} (x - m_x) f_1(x) dx \int_{-\infty}^{\infty} (y - m_y) f_2(y) dy. \end{aligned}$$

Интеграл $\int_{-\infty}^{\infty} (x - m_x) f_1(x) dx$ представляет собой не что иное, как первый центральный момент величины X , и, следовательно, равен нулю. По той

же причине равен нулю и второй сомножитель, следовательно, для независимых случайных величин $cov(X, Y) = 0$.

Таким образом, если ковариация двух случайных величин отлична от нуля, это есть признак наличия зависимости между ними.

Из формулы (3) видно, что ковариация характеризует не только зависимость величин, но и их рассеивание. Действительно, если, например, одна из величин (X, Y) весьма мало отклоняется от своего математического ожидания (почти не случайна), то ковариация будет мала, какой бы тесной зависимостью ни были связаны случайные величины (X, Y) . Поэтому для характеристики связи между величинами (X, Y) в чистом виде переходят от ковариации к безразмерной характеристике – коэффициенту корреляции.

Коэффициентом корреляции r случайных величин X и Y называется отношение ковариации к произведению средних квадратичных отклонений этих величин:

$$r = \frac{cov(X, Y)}{\sigma_x \sigma_y}.$$

Коэффициент корреляции является безразмерной величиной. Коэффициент корреляции независимых случайных величин равен нулю.

Если случайные величины независимы, то они и не коррелированы, но из некоррелированности нельзя сделать вывод о их независимости.

ЗАКЛЮЧЕНИЕ

В настоящем учебном пособии в традиционные разделы по дисциплине «Статистика» органично встроены элементы математической статистики, что существенно повышает его научный уровень.

Это особенно важно для дальнейшего изучения студентами направлений «Экономика» и «Менеджмент» таких дисциплин, как «Эконометрика» и «Методы принятия управленческих решений».

Раздел «Временные ряды» не включен в настоящее пособие, чтобы исключить дублирование, так как в дальнейшем эта тема будет подробно изучена в курсе «Эконометрика».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. **Статистика:** учебник для вузов / И. И. Елисеева и др. – 5-е изд. – Москва: Юрайт, 2021. – 572 с.
2. **Сазонов, А. Л.** Системы случайных величин / А. Л. Сазонов. – Санкт-Петербург: СПГУТД, 1998. – 51 с.
3. **Сазонов, А. Л.** Анализ систем числовых и нечисловых случайных переменных: учебное пособие / А. Л. Сазонов, Н. Н. Рожков. – Санкт-Петербург: ФГБОУВПО «СПГУТД», 2014. – 177 с.
4. **Годин, А. М.** Статистика: учебник для бакалавров / А. М. Годин. – Москва: Дашков и К, 2020. – 412 с.
5. **Балдин, К. В.** Теория вероятностей и математическая статистика: учебник / К. В. Балдин, В. Н. Башлыков, А. В. Рукосуев. – Москва: Дашков и К, 2020. – 472 с.
6. **Теория вероятностей и математическая статистика** / И. Л. Макарова и др. – Сочи: Сочинский государственный университет, 2020. – 130 с.
7. **Монсик, В. Б.** Вероятность и статистика: учебное пособие / В. Б. Монсик, А. А. Скрынников. – Москва: Лаборатория знаний, 2020. – 382 с.
8. **Статистика:** учебное пособие / под ред. В. Н. Селина, Е. Л. Шпаковской. – 3-е изд. – Москва: КНОРУС, 2014. – 504 с.
9. **Вентцель, Е. С.** Теория вероятностей / Е. С. Вентцель. – 4-е изд. – Москва: Наука, 1969. – 576 с.
10. **Колемаев, В. А.** Теория вероятностей и математическая статистика / В. А. Колемаев, В. Н. Калинина. – Москва: ИНФРА-М, 1999. – 302 с.
11. **Елисеева, И. И.** Общая теория статистики / И. И. Елисеева, М. М. Юзбашев. – 5-е изд. – Москва: финансы и статистика, 2004. – 656 с.
12. **Ляховецкий, А. М.** Статистика: учебное пособие / А. М. Ляховецкий, Е. В. Кремянская, Н. В. Климова; под ред. В. И. Негаева. – Москва: КНОРУС, 2018. – 362 с.

Учебное издание

**Богданов Александр Иванович
Никитина Людмила Николаевна
Крайкина Елизавета Андреевна**

СТАТИСТИКА

Издательский редактор Т. Н. Козлова

Учебное электронное издание сетевого распространения

Системные требования:
электронное устройство с программным обеспечением
для воспроизведения файлов формата PDF

Режим доступа: http://publish.sutd.ru/tp_get_file.php?id=2023176, по паролю.
– Загл. с экрана.

Дата подписания к использованию 12.09.2023 г. Рег. № 176/23

ФГБОУВО «СПбГУПТД»
Юридический и почтовый адрес:
191186, Санкт-Петербург, ул. Большая Морская, 18.
<http://sutd.ru/>