Министерство сельского хозяйства Российской Федерации
ФГБОУ ВО «Красноярский государственный аграрный университет»

С. Г. Лукинова, В. И. Иванов, В. Д. Жданова

МАТЕМАТИКА. ВЕРОЯТНОСТЬ И СТАТИСТИКА

Рекомендовано Учебно-методическим советом федерального государственного бюджетного образовательного учреждения высшего образования «Красноярский государственный аграрный университет» для внутривузовского использования в качестве учебного пособия для студентов, обучающихся по направлениям подготовки: 06.03.01 «Биология», 19.03.02 «Продукты питания из растительного сырья», 19.03.03 «Продукты питания животного происхождения», 20.03.02 «Природообустройство и водопользование», 21.03.02 «Землеустройство и кадастры», 35.03.03 «Агрохимия и почвоведение», 35.03.04 «Агрономия», 35.03.06 «Агроинженерия», 35.03.07 «Технология производства и переработки сельскохозяйственной продукции», 35.03.10 «Ландшафты и архитектура», 36.03.02 «Зоотехния» очной и заочной форм обучения

Красноярск, 2026

УДК 519.2
ББК 22.17я73
Л 84

Рецензенты:
Е. В. Еремин, д-р физ.-мат. наук, доцент,
профессор кафедры ФТТиНТ СФУ
С. С. Шатохин, канд. техн. наук, доцент, доцент кафедры
«Лесной инжиниринг» СибГУ им. М.Ф. Решетнева

Л 84 **Лукинова, С. Г.**
Математика. Вероятность и статистика: учебное пособие / С. Г. Лукинова, В. И. Иванов, В. Д. Жданова; Красноярский государственный аграрный университет. – Красноярск, 2026. – 170 с.

Представлены основные теоретические сведения по теории вероятностей и математической статистике. В каждом разделе приводятся типовые задачи с подробными решениями, задания для контрольных работ, варианты индивидуальных заданий, вопросы и упражнения для самопроверки и текущей аттестации.

Предназначено для студентов, обучающихся по направлениям подготовки: 06.03.01 «Биология», 19.03.02 «Продукты питания из растительного сырья», 19.03.03 «Продукты питания животного происхождения», 20.03.02 «Природообустройство и водопользование», 21.03.02 «Землеустройство и кадастры», 35.03.03 «Агрохимия и почвоведение», 35.03.04 «Агрономия», 35.03.06 «Агроинженерия», 35.03.07 «Технология производства и переработки сельскохозяйственной продукции», 35.03.10 «Ландшафты и архитектура», 36.03.02 «Зоотехния» очной и заочной форм обучения.

УДК 519.2
ББК 22.17я73

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Данное учебное пособие является частью учебно-методического комплекса дисциплин «Математика», «Математика и математическая статистика», «Высшая математика», в которой представлены разделы: теория вероятностей, математическая статистика.

В пособии изложены основные теоретические сведения по теории вероятностей и математической статистике. Оно профессионально ориентировано, в каждой части приведено много примеров и задач, связанных с будущей профессиональной деятельностью студентов по указанным направлениям подготовки с подробными решениями. Также для лучшего усвоения теории в пособие включен практикум по рассмотренным разделам, который также носит профессионально ориентированный характер.

Учебное пособие составлено в полном соответствии с Федеральным государственным образовательным стандартом высшего образования для указанных направлений.

Цель данного учебного пособия состоит в приобретении студентами прочных теоретических знаний, твердых практических навыков при изучении представленных разделов, получении представления о прикладном потенциале рассмотренных тем и оказании помощи студентам, будущим специалистам в сельском хозяйстве в освоении математических дисциплин в части теории вероятностей и математической статистики.

Учебное пособие содержит индивидуальные практические задания, также включены контрольные вопросы для самопроверки и итоговой аттестации, а также тестовые задания для итогового тестирования.

Пособие рассчитано как для самостоятельной работы студентов, так и для обучения в аудитории.

Задачи итогового теста необходимо выполнить перед сдачей зачета или экзамена. Индивидуальные практические задания необходимо выполнить, чтобы получить допуск к экзамену или зачету. В пособии подробно рассмотрены решения типовых практических заданий.

Защита практических заданий осуществляется в письменной форме или в виде собеседования (по решению преподавателя).

1. ПРЕДМЕТ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Теория вероятностей – раздел математики, изучающий закономерности массовых случайных явлений. Под случайными явлениями понимаются явления с неопределенным исходом, происходящие при неоднократном воспроизведении определенного комплекса условий.

Очевидно, что в природе, технике и экономике нет явлений, в которых не присутствовали бы элементы случайности. Существует два подхода к изучению этих явлений. Один из них – классический, состоит в том, что выделяются основные факторы, определяющие данное явление, а влиянием множества остальных, второстепенных факторов, пренебрегают. Таким образом, выявляется основная закономерность, свойственная данному явлению, позволяющая однозначно предсказать результат по заданным условиям. Второй подход состоит в том, что учитываются не только основные, но и множество второстепенных факторов. Элемент неопределенности, свойственный случайным явлениям и обусловленный второстепенными факторами, требует специальных методов изучения. Разработкой таких методов, изучением специфических закономерностей, наблюдаемых в случайных явлениях, и занимается теория вероятностей.

2. СЛУЧАЙНЫЕ СОБЫТИЯ

2.1. Случайные события, пространство элементарных событий

Случайным событием называется событие, которое в результате испытания может произойти или не произойти.

Обозначают события буквами латинского алфавита: $A, B, C, \ldots$.

Например:

$A = \{\textit{При бросании игральной кости выпадает четное число очков}\};$

$B = \{\textit{Посеянное семя взошло}\};$

$C = \{\textit{При бросании игральной кости выпадает шесть очков}\}.$

С каждым испытанием связано понятие множества элементарных событий (элементарных исходов).

Элементарные события – это возможные, исключающие друг друга результаты испытания. Элементарное событие нельзя разложить на более простые события. Элементарные события, при которых интересующее нас событие наступает, называются ***благоприятствующими*** этому событию.

Множество всех элементарных событий, связанных с данным испытанием, называется ***пространством элементарных событий*** и обозначается **Ω**, а элементарное событие – точка (элемент) пространства **Ω** обозначается □, □$\in\Omega$.

Очевидно, что для любого случайного события A, $A\subseteq\Omega$.

Пример 1

Описать *пространство элементарных событий* при однократном бросании игральной кости, а также события:

$A=\{\textit{Прибросанииигральнойкостивыпадаетчетноечислоочков}\}$;

$D=\{\textit{Прибросанииигральнойкостивыпадаетнечетноечислоочков}\}$.

Решение

Пусть событие ω_i заключается в выпадении i очков. При однократном подбрасывании игральной кости может произойти шесть элементарных исходов:

$$\omega_1=\{1\},\quad \omega_2=\{2\},\quad \omega_3=\{3\},\quad \omega_4=\{4\},\quad \omega_5=\{5\},\quad \omega_6=\{6\}.$$

Пространство элементарных событий – это множество всех элементарных исходов:

$$\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}=\{1, 2, 3, 4, 5, 6\}.$$

Благоприятствующими для события A будет выпадение 2, 4 или 6 очков, поэтому

$$A=\{\omega_2,\omega_4,\omega_6\}=\{2, 4, 6\},$$

аналогично $D=\{\omega_1,\omega_3,\omega_5\}=\{1, 3, 5,\}$.

Введем следующие понятия:

- события называются **несовместными**, если они не могут произойти одновременно при одном и том же испытании, в противном случае события называют **совместными.**
- событие $\overline{A}$ называется **противоположным** для события A, если появление одного из них равносильно не появлению другого.

Например, если
$A = \{Прибросанииигральнойкостивыпадает\ четноечислоочков\}$, то
$\overline{A} = \{Прибросанииигральнойкостивыпадает\ нечетноечислоочков\}$.

- Событие A называется **достоверным**, если оно обязательно произойдет, обозначают Ω.
- Событие $A = \varnothing$ называется **невозможным**, если в результате испытания оно не может произойти.

2.2. Алгебра событий

Над событиями определены операции суммы, разности и произведения.

1. Суммой двух событий A и B называется событие C, состоящее в наступлении хотя бы одного из данных событий, обозначают

$$A + B = C.$$

Сумму событий $A + B$ можно рассматривать как множество, состоящее из всех элементарных событий, принадлежащих, по крайней мере, одному из событий A **или** B.

2. Разностью двух событий A и B называется событие D, которое состоит в том, что событие A произойдет, а событие B не произойдет, обозначают

$$A - B = D.$$

В частности противоположное событие $\overline{A} = \Omega - A$ состоит в том, что событие A не произойдет,

$$\overline{A} = \Omega - A.$$

3. Произведением событий называется событие E, состоящее в совместном наступлении этих событий, обозначают

$$A \cdot B = E.$$

Если A и B – совместные события, то их произведение $A \cdot B$ означает наступление и события A, и события B **одновременно**.

Операции над событиями можно геометрически проиллюстрировать диаграммами Эйлера – Венна (рис. 1) (Ω – пространство элементарных событий).

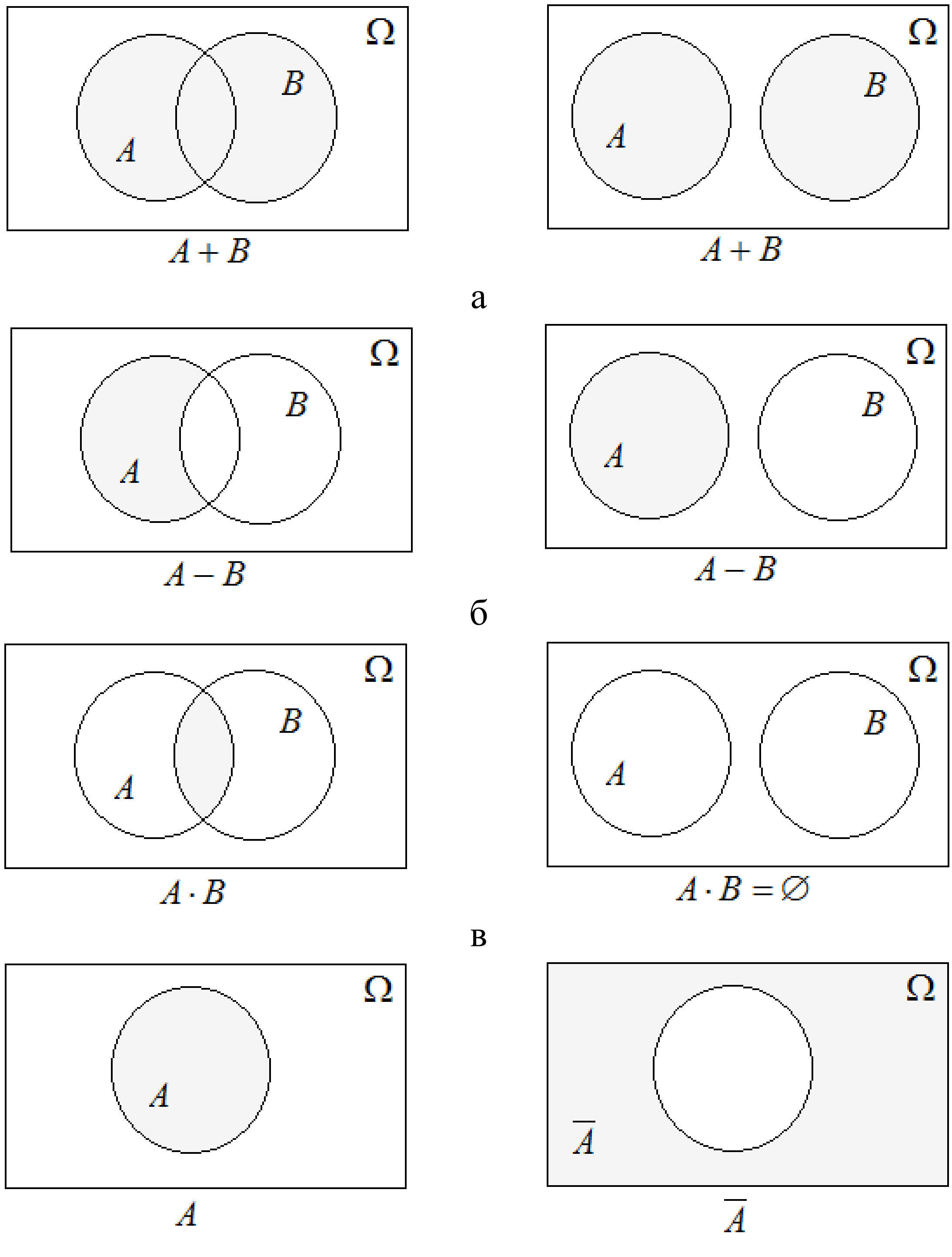


Рисунок 1 – Диаграммы Эйлера – Венна: а – сумма событий; б – разность событий; в – произведение; г – противоположное событие

***Свойства операций сложения и умножения событий*:**

1. $A+B=B+A$ (коммутативность сложения);
2. $A+B=B+A$ (ассоциативность сложения);
3. $A\cdot B=B\cdot A$ (коммутативность умножения);
4. $(A\cdot B)\cdot C=(A\cdot B)\cdot C$ (ассоциативность умножения);
5. $A\cdot(B+C)=A\cdot B+A\cdot C$ (дистрибутивность).

Пример 2

Описать *следующие события* $A+B$, $D-B$, $A\cdot B$, $\overline{A}\cdot B$, $C\cdot D$, если

$A=\{\textit{При бросании игральной кости выпадает четное число очков}\}$,

$B=\{\textit{При бросании игральной кости выпадает 5 очков}\}$,

$C=\{\textit{При бросании игральной кости выпадает 6 очков}\}$,

$D=\{\textit{При бросании игральной кости выпадает нечетное число очков}\}$.

Решение

$A+B=\{2,4,6,5\}$;

$D-B=\{1,3\}$;

$A\cdot B=\varnothing$;

$\overline{A}\cdot B=\{5\}$;

$C\cdot D=\varnothing$.

3. ПОНЯТИЕ ВЕРОЯТНОСТИ

3.1. Статистическое определение вероятности

Относительной частотой $W(A)$ случайного события A называется отношение числа опытов, в которых это событие появилось, к числу всех произведенных опытов.

$$W(A)=\frac{m}{n},$$

где m – число опытов, в которых событие A появилось; n – число всех произведенных опытов.

Очевидно*, что относительная частота обладает следующими свойствами:*

- $0\le W(A)\le 1$;
- $W(\varnothing)=0$;
- $W(\Omega)=1$.

Наблюдения позволили установить, что относительная частота обладает свойством *статистической устойчивости*: в различных сериях однородных испытаний (в каждом из которых может появиться или не появиться событие) относительная частота изменяется тем меньше, чем больше проведено испытаний, и принимает значения, достаточно близкие к некоторой постоянной. Эту постоянную, являющуюся объективной характеристикой явления, считают вероятностью данного события.

Статистической вероятностью случайного события A является относительная частота $W(A)$ этого события в различных сериях большого числа испытаний.

Статистическое определение вероятности применимо только в тех случаях, когда испытания могут быть воспроизведены неограниченное число раз при одном и том же комплексе условий.

3.2. Классическое определение вероятности

Вероятностью события A называется **число**, равное отношению числа m (благоприятствующих исходов событию A) к числу n (всех возможных исходов), обозначают так:

$$P(A)=\frac{m}{n}.$$

Пример 3

Найти вероятность выпадения четного числа очков при бросании игральной кости.

Решение

Множество всех элементарных исходов данного испытания состоит из шести элементов ($n=6$).

Благоприятствующими событию

$$A=\{Выпало четное число очков\}$$

будут элементарные события – выпадение 2, 4 и 6 очков, т. е. $m=3$.

Вероятность события $P(A)=\frac{m}{n}=\frac{3}{6}=\frac{1}{2}$.

3.3. Геометрическое определение вероятности

Классическое и статистическое определения вероятности предполагают *конечное число* возможных исходов испытания. На практике встречаются опыты, для которых число возможных исходов бесконечно.

Рассмотрим в качестве пространства элементарных событий Ω произвольную область (на прямой, плоскости или в пространстве), имеющую конечную меру (длину, площадь или объем).

Геометрической вероятностью события A называется отношение меры области A к мере всей области Ω:

$$P(A) = \frac{мера\,области\,A}{мера\,области\,\Omega}.$$

Если область A – отрезок в пространстве, то мера области А – длина этого отрезка, если область A – некоторая фигура на плоскости, то мера области A – площадь этой фигуры, если же область A – тело в пространстве R_3, мера области A – объем этого тела.

Пример 4

В квадрат со стороной, равной ***a***, наудачу брошена точка. Найти вероятность того, что точка попадет в круг, радиус которого равен r $(r \le a/2)$ и центр находится в точке пересечения диагоналей квадрата.

Решение

Площадь квадрата $S_{кв} = a^2$. Площадь круга $S_{кр} = \pi \cdot r^2$. Итак, искомая вероятность

$$\mathrm{P}(A) = \frac{S_{кр}}{S_{кв}} = \frac{\pi \cdot r^2}{a^2}.$$

4. СВОЙСТВА ВЕРОЯТНОСТИ

1) Для любого случайного события выполняется условие

$$0 \le P(A) \le 1.$$

2) Вероятность **достоверного** события равна единице

$$P(\Omega) = 1.$$

3) Вероятность **невозможного** события равна нулю

$$P(\Omega) = 1.$$

4) Теорема **сложения вероятностей**.

Вероятность суммы несовместных событий равна сумме вероятностей этих событий:

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

В случае n несовместных событий справедливо следующее:

$$P(A_1+A_2+\ldots+A_n)=P(A_1)+P(A_2)+\ldots+P(A_n).$$

5) Вероятность **противоположного** события

$$P(\overline{A})=1-P(A),$$

так как сумма вероятностей противоположных событий A и $\overline{A}$ равна единице:

$$P(A)+P(\overline{A})=1.$$

6) Теорема **умножения вероятностей**.

Вероятность произведения независимых событий равна произведению вероятностей этих событий:

$$P(A\cdot B)=P(A)\cdot P(B).$$

В случае n независимых событий справедливо следующее:

$$P(A_1\cdot A_2\cdot\ldots\cdot A_n)=P(A_1)\cdot P(A_2)\cdot\ldots\cdot P(A_n).$$

Упражнение

Показать самостоятельно, что если события совместные, то

$$P(A+B)=P(A)+P(B)-P(A\cdot B).$$

Пример 5

Два стрелка стреляют по мишени. Вероятность попадания для первого стрелка – 0,7, для второго – 0,8.

Найти вероятности следующих событий:

а) оба стрелка попадут в цель;

б) оба стрелка промахнутся;

в) только один стрелок попадет в цель;

г) хотя бы один попадет.

Решение

Рассмотрим следующие события:

$A=\{Попал\ первый\ стрелок\},$

$B=\{Попал\ второй\ стрелок\},$

$\overline{A}=\{Первый\ не\ попал\},$

$\overline{B}=\{Второй\ не\ попал\}.$

Тогда по условию задачи вероятности этих событий равны:

$$P(A)=0{,}7; \qquad P(\overline{A})-1-P(A)=0{,}3;$$
$$P(B)=0{,}8; \qquad P(\overline{B})-1-P(B)=0{,}2.$$

Используя операции над событиями, найдем искомые вероятности:

а) оба стрелка попадут в цель

$$P(A\cdot B)=P(A)\cdot P(B)=0{,}7\cdot 0{,}8=0{,}56;$$

б) оба стрелка промахнутся

$$P(\overline{A}\cdot\overline{B})=P(\overline{A})\cdot P(\overline{B})=0{,}3\cdot 0{,}2=0{,}06;$$

в) только один стрелок попадет в цель

$$P(\overline{A}\cdot B+\overline{B}\cdot A)=P(\overline{A})\cdot P(B)+P(\overline{B})\cdot P(A)=0{,}3\cdot 0{,}8+0{,}2\cdot 0{,}7=0{,}38;$$

г) хотя бы один попадет:

событие $\overline{A}\cdot\overline{B}$ («оба стрелка промахнулись») будет противоположным для события «хотя бы один попал в цель». Следовательно, искомая вероятность находится так:

$$1-P(\overline{A}\cdot\overline{B})=1-0{,}2\cdot 0{,}3=0{,}94.$$

5. ФОРМУЛЫ КОМБИНАТОРИКИ

Рассмотрим некоторое множество, состоящее из n различных элементов. Из элементов этого множества могут быть образованы различные подмножества (комбинации) из k элементов ($k\le n$).

На практике часто применяются следующие формулы комбинаторики:

- ***Сочетаниями*** из n различных элементов по k элементов называются такие комбинации из m элементов, которые отличаются друг от друга *только составом своих элементов*, обозначают C_n^k.

Число сочетаний C_n^k из n различных элементов по k элементов вычисляется по формуле

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Если число предметов в совокупности равно n и производится выбор k предметов без возвращения и без учета порядка, то количество всех элементарных исходов находится по формуле числа сочетаний.

Пример 6

Найти количество способов заполнения карточки спортлото «6 из 49».

Решение

В данной задаче эксперимент состоит в том, что случайным образом выбирается 6 чисел из 49 в карточке спортлото.

Пространство элементарных исходов здесь образуют различные кортежи из 6 чисел, которые выбирают без учета порядка из чисел:

1, 2, 3, …, 49.

Количество всех возможных способов (всех элементарных событий) находится по формуле числа сочетаний:

$$C_{49}^6 = \frac{49!}{6!\cdot(49-3)!} = \frac{49!}{6!\cdot 43!}.$$

- ***Перестановками*** из k различных элементов называются такие комбинации из этих элементов, которые отличаются друг от друга *только порядком своих элементов*, обозначают P_k.

Число перестановок из k элементов вычисляется по формуле

$$P_k = k!$$

Пример 7

Найти количество вариантов, чтобы рассадить шесть гостей за обеденным столом.

Решение

В данной задаче число различных вариантов будет равно перестановкам из шести элементов:

$$P_6 = 6! = 1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6 = 120.$$

- ***Размещениями*** из n различных элементов по k элементов называются такие комбинации из k элементов, которые отличаются друг от друга *либо составом элементов, либо порядком своих элементов*, обозначают A_n^k.

Число размещений из n различных элементов по m вычисляется по формуле

$$A_n^k = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-k+1)$$

$$\text{или } A_n^k = \frac{n!}{(n-k)!}.$$

Если выбор производится k предметов без возвращения и с учетом порядка из совокупности в n предметов, то количество всех элементарных исходов находится по формуле числа размещений.

Пример 8

В виварии имеются 10 подопытных животных. Найти количество способов выбора по одному животному для трех различных экспериментов.

Решение

Число различных вариантов выбора подопытных животных из вивария, содержащего 10 животных, по одному животному для трех экспериментов

$$A_{10}^3 = 10\cdot 9\cdot 8 = 720.$$

Если число предметов в совокупности равно n и производится выбор k предметов **с возвращением и учетом порядка**, то количество всех элементарных исходов будет равно n^k.

Пример 9

Сколько существует различных шестизначных телефонных номеров?

Решение.

Набор телефонного номера из шести цифр ($k=6$) можно рассматривать как результат выбора с возвращением и учетом порядка шести цифр из следующих десяти: 0, 1, …, 9 ($n=10$). Пространство элементарных исходов в этой задаче образует различные кортежи из шести цифр, например 210635, 201356, 112345, ….

Всего различных телефонных номеров будет

$$n^k = 10^6.$$

Формулы комбинаторики используют при непосредственном вычислении вероятностей.

Пример 10

Из пруда, в котором плавает 20 щук, выловили 8 щук, пометили их и пустили обратно в пруд. Во второй раз выловили 5 щук. Какова вероятность, что среди них окажутся только 3 помеченных щуки?

Решение

Вероятность события

$A = \{ \textit{Среди повторно выловленных щук окажутся 3 помеченных щуки} \}$

найдем по формуле классической вероятности: $P(A) = \frac{m}{n}$. Число всех возможных исходов n находится по формуле числа сочетаний:

$$n = C_{20}^{5} = \frac{20!}{5!\cdot 15!}.$$

Число исходов, благоприятствующих наступлению события A, равно числу исходов, где среди повторно выловленных пяти щук должно быть три помеченных щуки и две непомеченных. Поэтому число благоприятствующих исходов находится по формуле

$$m = C_8^3 \cdot C_{12}^2.$$

Тогда

$$P(A) = \frac{C_8^3 \cdot C_{12}^2}{C_{20}^5} = \frac{8!\cdot 12!\cdot 5!\cdot 15!}{3!\cdot 5!\cdot 2!\cdot 10!\cdot 20!},$$

выполнив преобразования, получим

$$P(A) = \frac{4\cdot 5\cdot 6\cdot 7\cdot 8\cdot 11\cdot 12}{1\cdot 2\cdot 16\cdot 17\cdot 18\cdot 19\cdot 20} = \frac{77}{323} = 0{,}239.$$

6. УСЛОВНАЯ ВЕРОЯТНОСТЬ. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ. ФОРМУЛЫ БАЙЕСА

Условной вероятностью события A при условии, что событие B произошло, называется вероятность

$$P_A(B) = \frac{P(A\cdot B)}{P(A)}.$$

Условная вероятность обозначается $P_A(B)$ или $P(B/A)$.

Аналогично определяется вероятность $P_B(A)$ или $P(A/B)$.

Теорема ***умножения*** *(для зависимых событий)*

Вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого:

$$P(AB) = P(A) \cdot P_A(B).$$

В случае если рассматривается n зависимых событий, теорема умножения имеет вид

$$P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) = P(A_1) \cdot P_{A_1}(A_2) \cdot P_{A_1 \cdot A_2}(A_3) \cdot \ldots \cdot P_{A_1 A_2 \ldots A_{n-1}}(A_n).$$

Пример 11

В стаде животных из 40 голов одной породы 15 животных не получили прививку. Наудачу последовательно, без возвращения отбираются два животных. Найти вероятность того, что оба выбранных животных не прививались.

Решение

Рассмотрим события

$A = \{\textit{Первое отобранное животное не прививалось}\}$,

$B = \{\textit{Второе отобранное животное не прививалось}\}$.

Очевидно, что вероятность $P(A) = \dfrac{15}{40}$.

Вероятность события B найдем из условия, что всех исходов осталось 39 и благоприятствующих 14, тогда

$$P(B) = \frac{14}{39}.$$

Поскольку была найдена вероятность события B, при условии, что событие A уже произошло, то правильнее будет записать эту условную вероятность как $P_A(B) = \dfrac{14}{39}$.

Итак, искомая вероятность того, что оба выбранных животных не прививались,

$$P(A \cdot B) = P(A) \cdot P_A(B) = \frac{15}{40} \cdot \frac{14}{39} = \frac{7}{52}.$$

Формула полной вероятности

События $H_1, H_2, H_3, ..., H_n$ образуют **полную группу событий**, если выполняются следующие условия:

1) $H_i \cdot H_j = \varnothing$, события попарно несовместны;

2) $\sum\limits_{i=1}^{n} H_i = \Omega$, их сумма равна достоверному событию.

Очевидно, что

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

Пусть требуется найти вероятность события A, которое может произойти после того, как произойдет одно из событий $H_1, H_2, H_3, ..., H_n$, образующих полную группу событий.

Искомая вероятность находится по следующей формуле, которая называется ***формулой полной вероятности*:**

$$P(A) = P(H_1)P_{H_1}(A) + P(H_2)P_{H_2}(A) + ... + P(H_n)P_{H_n}(A),$$

или

$$P(A) = \sum_{i=1}^{n} P(H_i)P_{H_i}(A),$$

где $\sum\limits_{i=1}^{n} P(H_i) = 1.$

Пример 11

На птицефабрике первым цехом производится 30 % яиц, вторым – 25 %, третьим – 45 %. Бракованных яиц с первого цеха поступает 2 %, со второго – 1 %, с третьего – 3 %. Какова вероятность того, что выбранное наудачу яйцо бракованное?

Решение

Обозначим интересующее нас событие:

$$A = \{Выбранное\ наудачу\ яйцо\ бракованное\}.$$

Поскольку выбранное яйцо могло поступить из трех цехов, введем события

$$H_i = \{Яйцо\ поступило\ из\ i-го\ цеха\}, \qquad i = 1, 2, 3.$$

Из условия задачи следует, что

$$P(H_1) = 0{,}30, \quad P(H_2) = 0{,}25, \quad P(H_3) = 0{,}45.$$

События H_1, H_2, H_3 попарно несовместны, сумма их вероятностей равна единице, следовательно, они образуют полную группу событий и можно воспользоваться формулой полной вероятности:

$$P(A) = P(H_1)P_{H_1}(A) + P(H_2)P_{H_2}(A) + P(H_3)P_{H_3}(A).$$

Из условия задачи $P_{H_1}(A)$ означает вероятность того, что наудачу выбранная деталь окажется бракованной и изготовленной на первом станке, $P_{H_1}(A) = 0{,}02$,

Аналогично

$$P_{H_2}(A) = 0{,}01, \quad P_{H_3}(A) = 0{,}03.$$

По формуле полной вероятности найдем требуемую вероятность события A:

$$P(A) = P(H_1)P_{H_1}(A) + P(H_2)P_{H_2}(A) + P(H_3)P_{H_3}(A) =$$
$$= 0{,}3 \cdot 0{,}02 + 0{,}25 \cdot 0{,}01 + 0{,}45 \cdot 0{,}03 = 0{,}022 \ .$$

Формулы Байеса

Формулы Байеса являются следствием формулы полной вероятности и применяются в случае, когда событие A, которое может произойти только после того, как произойдет одно из независимых событий (гипотез) $H_1, H_2, H_3, ..., H_n$, образующих полную группу событий, уже произошло и необходимо произвести количественную переоценку вероятностей гипотез $H_1, H_2, H_3, ..., H_n$, то есть найти условные вероятности гипотез $P_A(H_1), P_A(H_2), ..., P_A(H_n)$.

Эти вероятности находятся по следующим формулам, которые называются ***формулами*** **Байеса:**

$$P_A(H_i) = \frac{P(H_i) \cdot P_{H_i}(A)}{P(A)}, \qquad i = 1, 2, ..., n.,$$

где $P(A) = \sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)$ – формула полной вероятности.

Пример 12

При условии примера 11 найти вероятность того, что выбранное наудачу бракованное яйцо произведено в первом цехе.

Решение

По формуле Байеса $P_A(H_1)=\dfrac{P(H_1)\cdot P_{H_1}(A)}{P(A)}$ найдем искомую вероятность $P_A(H_1)=\dfrac{0{,}3\cdot 0{,}02}{0{,}022}=0{,}27.$

7. НЕЗАВИСИМЫЕ ИСПЫТАНИЯ

Схема независимых испытаний состоит в следующем: проводится n последовательных независимых экспериментов в одинаковых условиях, в каждом из которых рассматривается одно и то же событие A, которое может произойти или не произойти в результате эксперимента. Вероятность p наступления события A в каждом испытании одна и та же. Вероятность противоположного события $\overline{A}$ обозначим q, $q=1-p$.

7.1. Формула Бернулли

Вероятность того, что событие A наступит k раз в n независимых испытаниях, находится по ***формуле Бернулли:***

$$P_n(k)=C_n^k\cdot p^k\cdot q^{n-k}.$$

Число наступления события A называется ***наивероятнейшим***, если оно имеет наибольшую вероятность по сравнению с вероятностями наступления события A любое другое количество раз, обозначают k_0.

Наивероятнейшее число наступлений события A в n испытаниях заключено между числами $np-q$ и $np+q$:

$$np-q\le k_0\le np+p.$$

Пример 13

Вероятность всхожести одного семени растения равна 1/3. Посадили шесть семян.

а) Какова вероятность, что взойдут два семени?

б) Какова вероятность, что взойдут не менее двух семян?

в) Каково наивероятнейшее число взошедших семян?

Решение

Пусть $A = \{Посаженное\, семя\, взошло\}$. Тогда $p = P(A) = 1/3$, $q = P(\overline{A}) = 1 - p = 2/3$.

а) Число посаженных семян $n = 6$. Всхожесть каждого семени не зависит от всхожести любого другого. Тогда, полагая $k = 2$, по формуле Бернулли находим вероятность всхожести двух семян из шести:

$$P_6(2) = C_6^2 \cdot p^2 \cdot q^4 = \frac{6!}{2! \cdot 4!} \cdot \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^4 = 15 \cdot \frac{16}{3^6} = \frac{80}{243} = 0{,}329.$$

б) Событие «взойдет не менее двух семян» противоположно событию «одно семя взойдет или ни одного не взойдет», следовательно, искомая вероятность

$$P_6(k \geq 2) = 1 - P_6(0) - P_6(1) =$$

$$= 1 - C_6^0 \cdot p^0 \cdot q^6 - C_6^1 \cdot p^1 \cdot q^5 = 1 - \frac{6!}{0! \cdot 6!}\left(\frac{2}{3}\right)^6 - \frac{6!}{1! \cdot 5!} \cdot \frac{1}{3} \cdot \left(\frac{2}{3}\right)^5 =$$

$$= 1 - \left(\frac{2}{3}\right)^6 - 6 \cdot \left(\frac{2}{3}\right)^5 = 1 - \frac{64}{729} - \frac{192}{729} = \frac{473}{729} = 0{,}649.$$

в) Наивероятнейшее число взошедших семян

$$np - q \leq k_0 \leq np + p.$$

В нашем случае

$$6 \cdot \frac{1}{3} - \frac{2}{3} \leq k_0 \leq 6 \cdot \frac{1}{3} + \frac{1}{3} \text{ или } 1\frac{1}{3} \leq k_0 \leq 2\frac{2}{3}.$$

Следовательно, $k_0 = 2$.

Пример 14

Вероятность рождения бычка при отеле коровы примем равной 0,5. Найти вероятность того, что в группе из 10 стельных коров:

а) родятся не менее 2 и не более 4 бычков;

б) родится хотя бы один бычок.

Решение

а) Рассмотрим событие $A = \{У\, коровы\, родился\, бычок\}$, тогда $P(A) = \frac{1}{2} = p$, $q = 1 - p = \frac{1}{2}$, $n = 10$. Вероятность того, что у 10 коров родятся не менее 2 и не более 4 бычков, найдем с помощью формулы Бернулли:

$$P_{10}(2 \le k \le 4) = P_{10}(2) + P_{10}(3) + P_{10}(4) =$$

$$= C_{10}^{2} \cdot \left(\frac{1}{2}\right)^{2} \cdot \left(\frac{1}{2}\right)^{8} + C_{10}^{3} \cdot \left(\frac{1}{2}\right)^{3} \cdot \left(\frac{1}{2}\right)^{7} + C_{10}^{4} \cdot \left(\frac{1}{2}\right)^{4} \cdot \left(\frac{1}{2}\right)^{6} = \frac{375}{1024}.$$

б) Вероятность того, что у десяти коров родится хотя бы один бычок, находим по формуле

$$P_n(1 \le k \le n) = 1 - q^n, \text{ где } q = 1 - p.$$

$$P_{10}(1 \le k \le 10) = 1 - \left(\frac{1}{2}\right)^{10} = \frac{1023}{1024}.$$

Если число испытаний n велико ($n \ge 20$), то непосредственное вычисление вероятностей по формуле Бернулли технически сложно. Для этого существуют приближенные, так называемые асимптотические, формулы: формула Пуассона, локальная и интегральная формулы Муавра-Лапласа.

7.2. Формула Пуассона

Если вероятность p наступления события A в каждом испытании стремится к нулю при неограниченном увеличении числа испытаний, причем произведение $\lambda = np \le 10$, то вероятность того, что событие A появится k раз в n независимых испытаниях, находится *по формуле Пуассона*:

$$P_n(k) \approx \frac{\lambda^k \cdot e^{-\lambda}}{k!}.$$

Пример 15

Среди семян ржи 0,04 % семян сорняков. Какова вероятность при случайном отборе 5000 семян, обнаружить 5 семян сорняков?

Решение

По условию задачи $n = 5000, \quad k = 5, \quad p = 0{,}0004.$

Тогда $\lambda = n \cdot p = 5000 \cdot 0{,}0004 = 2 \le 10.$ Применив формулу Пуассона, получим:

$$P_{5000}(5)=\frac{2^5\cdot e^{-2}}{5!}=\frac{32\cdot 0{,}1352}{120}=0{,}036.$$

Пример 16

Для фермы закупили 1000 мешков комбикорма. Вероятность того, что степень измельчения комбикорма в одном мешке не соответствует зоотехническим требованиям, равна 0,001. Какова вероятность несоответствия зоотехническим требованиям измельчения купленного комбикорма:

а) в двух мешках;

б) не менее, чем в двух мешках.

Решение

Имеем $p=0{,}001$; $\lambda=n\cdot p=1000\cdot 0{,}001=1\le 10$. По формуле Пуассона:

$$\text{а) } P_{1000}(2)=\frac{1^2\cdot e^{-1}}{2!}=\frac{1}{2\cdot e}=0{,}184;$$

$$\text{б) } P_{1000}(k\ge 2)=1-P_{1000}(0)-P_{1000}(1)=1-\frac{1^0\cdot e^{-1}}{0!}-\frac{1^1\cdot e^{-1}}{1!}=1-\frac{2}{e}=0{,}26.$$

7.3. Локальная теорема Муавра – Лапласа

Если вероятность p наступления события A в каждом испытании одинакова и не близка к нулю, то вероятность того, что событие A наступит ровно k раз в n независимых испытаниях, приближенно равна

$$P_n(k)\approx\frac{1}{\sqrt{npq}}\cdot\varphi(x),$$

где $\varphi(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}$ – функция Гаусса, $x=\frac{k-np}{\sqrt{npq}}$.

Для упрощения расчетов с использованием данной формулы составлена таблица значений функции Гаусса $\varphi(x)$ (приложение А). Пользуясь этой таблицей, необходимо помнить следующие свойства функции $\varphi(x)$:

- Функция $\varphi(x)$является четной, т. е. $\varphi(-x)=\varphi(x)$.

- Функция $\varphi(x)$ – монотонно убывающая при положительных значениях x, при $x \geq 4$ можно считать, что $\varphi(x) = 0$.

Пример 17

Найти вероятность того, что из 150 кукурузных стеблей с тремя початками вызревших будет ровно 70, если вероятность вызревания таких стеблей равна 0,4.

Решение

Пусть событие

$A = \{Кукурузный стебель с тремя початками вызрел\}$.

$p(A) = 0{,}4$, $q = 1 - p = 0{,}6$. Всего производится $n = 150$ независимых испытаний, в которых событие A должно произойти ровно $k = 70$ раз. Имеем

$$\sqrt{npq} = \sqrt{150 \cdot 0{,}4 \cdot 0{,}6} = 6,$$

$$x = \frac{k - np}{\sqrt{npq}} = \frac{70 - 150 \cdot 0{,}4}{6} = \frac{10}{6} \approx 1{,}67.$$

По таблице значений функции $\varphi(x)$ (приложение А) найдем $\varphi(1{,}67) = 0{,}0989$.

Тогда

$$P_{150}(70) = \frac{1}{6} \cdot \varphi(1{,}67) = \frac{1}{6} \cdot 0{,}0989 = 0{,}0165.$$

7.4. Интегральная теорема Муавра – Лапласа

Если вероятность p наступления события A в каждом испытании постоянна и отлична от 0 и 1, то вероятность того, что число k наступления события A в n независимых испытаниях будет заключено в пределах от k_1 до k_2 раз (включительно), при достаточно большом n находится по *интегральной формуле Муавра – Лапласа*:

$$P_n(k_1; k_2) = \Phi_0(x_2) - \Phi_0(x_1),$$

где $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ – нормированная функция Лапласа,

$$x_1=\frac{k_1-np}{\sqrt{npq}};\quad x_2=\frac{k_2-np}{\sqrt{npq}}.$$

Нормированная функция Лапласа $\Phi_0(x)$ табулирована (приложение В). Пользуясь таблицей, следует учитывать, что:

а) функция $\Phi_0(x)$ нечетная, т. е. $\Phi_0(-x)=-\Phi_0(x)$;

б) при $x\geq 5$, можно считать, что $\Phi_0(x)=0{,}5$.

Наряду с нормированной функцией Лапласа используют функцию

$$\Phi\,(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt,$$

называемую *функцией Лапласа*, для которой также имеются таблицы приближенных значений, которые можно найти в учебниках по теории вероятностей. При этом

$$P_n(k_1\leq k\leq k_2)=\Phi_0(x_2)-\Phi_0(x_1)=\Phi(x_2)-\Phi(x_1).$$

Для функций $\Phi(x)$ и $\Phi_0(x)$. справедливо равенство

$$\Phi(x)=0{,}5+\Phi_0(x).$$

Пример 18

Вероятность того, что живая масса поросенка при рождении отвечает стандарту, равна 0,96. Найти вероятность того, что среди 200 родившихся поросят с нестандартной массой окажется не более 10.

Решение

По условию задачи $n=200$, $p=0{,}04$ (вероятность рождения поросенка с нестандартной массой), $q=0{,}96$. Вероятность того, что среди 200 родившихся поросят с нестандартной массой окажется не более 10, вычислим с помощью интегральной формулы Муавра – Лапласа при $k_1=0$ и $k_2=10$. Находим, что

$$x_1=\frac{0-200\cdot 0{,}04}{\sqrt{200\cdot 0{,}04\cdot 0{,}96}}\approx -2{,}89;\quad x_2=\frac{10-200\cdot 0{,}04}{\sqrt{200\cdot 0{,}04\cdot 0{,}96}}\approx 0{,}72;$$

$$P_{200}(0\leq k\leq 10)=\Phi_0(0{,}72)-\Phi_0(-2{,}89)=0{,}26424+0{,}49807=0{,}76231.$$

Значения нормированной функции Лапласа $\Phi_0(x)$ найдены по таблице (приложение В).

Если вероятность p наступления события A в каждом испытании постоянна, отлична от 0 и 1, то при достаточно большом числе n независимых испытаний **вероятность отклонения относительной частоты** события $\frac{m}{n}$ от ее вероятности p на величину ε вычисляется по формуле

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=2\Phi_0\left(\varepsilon\cdot\sqrt{\frac{n}{pq}}\right).$$

8. ПОНЯТИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Случайной величиной называется величина, принимающая в результате опыта одно из своих возможных значений, причем заранее неизвестно, какое именно значение она примет. Обозначают случайные величины заглавными буквами латинского алфавита $X, Y, Z, \ldots$, а возможные значения случайной величины – соответствующими строчными буквами $x, y, z, \ldots$.

Событие, состоящее в том, что случайная величина X примет значение x, обозначают $X = x$. Вероятность этого события обозначают $P(X = x)$, аналогично $P(X < x)$ – вероятность того, что случайная величина X примет значение меньшее, чем x.

Случайные величины могут быть дискретными, непрерывными и смешанного типа. Возможные значения дискретной случайной величины составляют конечное или бесконечное, но счетное множество. Возможные значения непрерывной случайной величины составляют непрерывное множество.

8.1. Дискретная случайная величина, закон и функция распределения случайной величины

Случайная величина называется ***дискретной***, если она принимает отдельные изолированные значения. Множество ее значений конечное или бесконечное, но счетное.

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Для дискретной случайной величины закон распределения имеет вид

x_i	x_1	x_2	x_3	…	x_n
p_i	p_1	p_2	p_3	…	p_n

причем

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

Возможные значения случайной величины в таблице записываются в порядке возрастания.

Функцией распределения случайной величины X называется функция $F(x)$, значение которой для любого $x \in (-\infty; +\infty)$, равно вероятности того, что случайная величина X примет значение, меньшее x:

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

Из определения следуют следующие ***свойства функции распределения*** *(доказать самостоятельно)*:

1. Функция распределения случайной величины X есть неотрицательная функция, значения которой заключены между нулем и единицей:

$$0 \le F(x) \le 1.$$

2. Функция распределения есть неубывающая функция на всей числовой оси.

3. $\lim\limits_{x \to -\infty} F(x) = 0; \quad \lim\limits_{x \to +\infty} F(x) = 1.$

4. Вероятность попадания случайной величины в интервал $[x_1; x_2)$ равна приращению функции распределения на этом интервале:

$$P(x_1 \le X < x_2) = F(x_2) - F(x_1).$$

Пример 19

Допустим, что на данной ферме яйценоскость одной курицы-несушки оценивается в одно яйцо за два дня. Из популяции кур-несушек на ферме отбираются три птицы. Составить закон и функцию распределения случайной величины, равной числу снесенных яиц тремя курицами за день. Построить график функции распределения.

Решение

Пусть случайная величина X – число снесенных яиц тремя курицами за день. Случайная величина X может принимать значения 0,1, 2, 3, которые являются ее возможными значениями.

Найдем соответствующие этим значениям вероятности. По условию задачи, вероятность того, что курица снесет одно яйцо за день, равна $\frac{1}{2}$. По формуле Бернулли, имеем

$$P(X=0)=C_3^0 \cdot p^0 \cdot q^3 = \left(\frac{1}{2}\right)^3 = \frac{1}{8},$$

$$P(X=1)=C_3^1 \cdot p^1 \cdot q^2 = 3\cdot\left(\frac{1}{2}\right)\cdot\left(\frac{1}{2}\right)^2 = \frac{3}{8},$$

$$P(X=2)=C_3^2 \cdot p^2 \cdot q \; = 3\cdot\left(\frac{1}{2}\right)^2\cdot\left(\frac{1}{2}\right) \; = \frac{3}{8},$$

$$P(X=3)=C_3^3 \cdot p^3 \cdot q^0 = 1\cdot\left(\frac{1}{2}\right)^3\cdot\left(\frac{1}{2}\right)^0 = \frac{1}{8}.$$

Следовательно, закон распределения данной случайной величины имеет вид:

x_i	0	1	2	3
p_i	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Очевидно, что $\sum_{i=1}^{n} p_i = 1$.

Найдем аналитическое выражение для функции распределения данной случайной величины X, для этого будем задавать различные значения переменной x и находить вероятности того, что $(X < x)$, то есть значения функции

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

- если $x \le 0$, то $F(x)=0$;

- если $0 < x \le 1$, то $F(x) = \frac{1}{8}$;
- если $1 < x \le 2$, то $F(x) = P(X=0) + P(X=1) = \frac{1}{8} + \frac{3}{8} = 0{,}5$;
- если $2 < x \le 3$, то

$F(x) = P(X=0) + P(X=1) + P(X=2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$;

- если $x > 3$,

то $F(x) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = 1$.

Итак, функция $F(x)$ имеет следующий вид:

$$F(x) = \begin{cases} 0, & \textit{если} \quad x \le 0; \\ \frac{1}{8}, & \textit{если} \quad 0 < x \le 1; \\ \frac{1}{2}, & \textit{если} \quad 1 < x \le 2; \\ \frac{7}{8}, & \textit{если} \quad 2 < x \le 3; \\ 1, & \textit{если} \quad x > 3. \end{cases}$$

Графиком функции распределения дискретной случайной величины является ступенчатая линия, точки разрыва которой соответствуют значениям случайной величины (рис. 2).

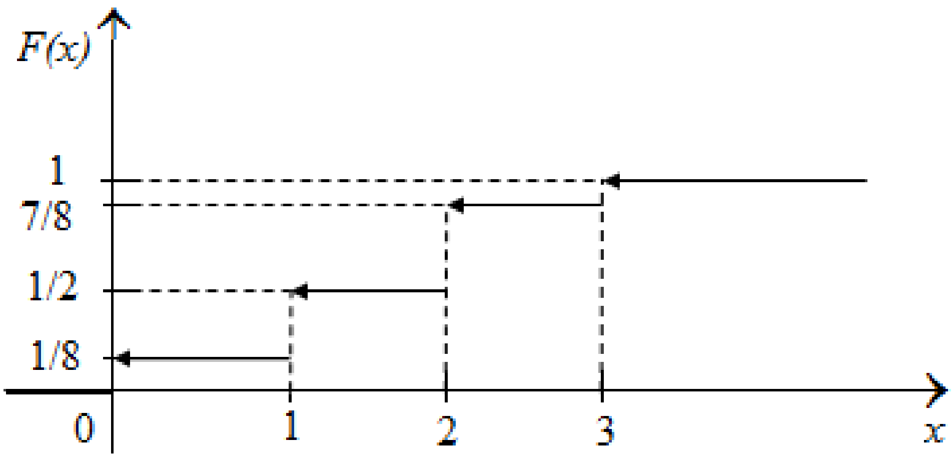


Рисунок 2 – График функции распределения дискретной случайной величины

8.2. Непрерывная случайная величина, функции распределения и плотности

Случайная величина X называется *непрерывной*, если ее функция распределения непрерывна в любой точке и дифференцируема всюду, кроме, быть может, отдельных точек.

Функция распределения непрерывной случайной величины имеет вид

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

где $f(x)$ – любая непрерывная, неотрицательная функция: $f(x) \geq 0$.

Плотностью вероятности (*плотностью распределения* или просто *плотностью*) непрерывной случайной величины X называется производная ее функции распределения, обозначается

$$f(x) = F'(x).$$

Про случайную величину X говорят, что она имеет распределение (распределена) с плотностью $f(x)$.

Плотность вероятности $f(x)$, как и функция распределения $F(x)$, является одной из форм закона распределения, но в отличие от функции распределения она существует только для непрерывных случайных величин.

Плотность вероятности иногда называют *дифференциальной функцией*, а функцию распределения непрерывной случайной величины – *интегральной функцией*.

Свойства функции плотности

1. Плотность вероятности – неотрицательная функция: $f(x) \geq 0$.
2. $f(x) = F'(x)$.
3. Несобственный интеграл от плотности вероятности непрерывной случайной величины равен единице:

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

4. Вероятность попадания непрерывной случайной величины в интервал $[a;b]$

$$P(a \le X \le b) = \int_a^b f(x)\,dx.$$

Замечание

Геометрически третье свойство плотности вероятности означает, что график функции плотности $f(x)$ – *кривая распределения* лежит не ниже оси абсцисс, а полная площадь области, ограниченной кривой распределения и осью абсцисс, равна единице.

Пример 20

Известна функция распределения некоторой случайной величины:

$$F(x) = \begin{cases} 0, & x \le 0; \\ \dfrac{x^2}{49}, & 0 < x \le 7; \\ 1, & x > 7. \end{cases}$$

Найти $f(x)$, $P(2 < x < 5)$.

Решение

Функцию плотности найдем из свойства $f(x) = F'(x)$, тогда

$$f(x) = F'(x) = \begin{cases} 0, & x \le 0; \\ \dfrac{2x}{49}, & 0 < x \le 7; \\ 0, & x > 7. \end{cases}$$

Вероятность того, что случайная величина x примет значения, принадлежащие интервалу $(2;5)$, найдем, используя четвертое свойство: $P(2 < x < 5) = F(5) - F(2) = \dfrac{5^2}{49} - \dfrac{2^2}{49} = \dfrac{21}{49} = \dfrac{3}{7}$.

9. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

К *числовым характеристикам* случайных величин относятся: *математическое ожидание, дисперсия, среднее квадратическое отклонение, медиана, мода и другие.*

9.1. Математическое ожидание случайной величины

Математическим ожиданием дискретной случайной величины X называют сумму произведений всех ее возможных значений на соответствующие им вероятности, обозначают

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

Математическим ожиданием непрерывной случайной величины X называют несобственный интеграл:

$$M(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx$$

Свойства математического ожидания

1. Математическое ожидание постоянной величины C равно самой постоянной:

$$M(C)=C.$$

2. Постоянный множитель можно выносить за знак математического ожидания:

$$M(C\cdot X)=C\cdot M(X).$$

3. Математическое ожидание алгебраической суммы конечного числа случайных величин равно сумме математических ожиданий этих величин:

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

4. Математическое ожидание произведения конечного числа независимых случайных величин равно произведению их математических ожиданий:

$$M(X\cdot Y)=M(X)\cdot M(Y).$$

9.2. Дисперсия случайной величины

Дисперсией случайной величины X называется математическое ожидание квадрата ее отклонения от математического ожидания, обозначают

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

Для вычисления дисперсии случайной величины часто на практике используют формулу

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

Действительно, используя свойства математического ожидания, докажем справедливость этого равенства:

$$D(X) = M[X - M(X)]^2 = M\left[X^2 - 2X \cdot M(X) + [M(X)]^2\right] =$$
$$= M(X^2) - 2M(X) \cdot M(X) + [M(X)]^2 == M(X^2) - [M(X)]^2.$$

На практике используют следующие формулы для вычисления дисперсии:

- дискретной случайной величины
$$D(X) = \sum_{i=1}^{n} x_i^2 \cdot p_i - \left(\sum_{i=1}^{n} x_i \cdot p_i\right)^2;$$
- непрерывной случайной величины
$$D(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx - (M(X))^2.$$

Свойства дисперсии

1. Из определения и свойств математического ожидания следует, что дисперсия любой случайной величины неотрицательна:

$$D(X) \geq 0.$$

2. Дисперсия постоянной величины равна нулю:

$$D(C) = 0.$$

3. Постоянный множитель можно выносить за знак дисперсии, возведя его в квадрат:

$$D(C \cdot X) = C^2 \cdot D(X).$$

4. Дисперсия суммы (разности) двух независимых случайных величин равна сумме или разности дисперсий этих величин:

$$D(X \pm Y) = D(X) \pm D(Y).$$

9.3. Среднее квадратическое отклонение случайной величины

Дисперсия имеет размерность квадрата случайной величины, что не всегда удобно. Поэтому в качестве показателя рассеяния используют также величину $\sqrt{D(X)}$.

Средним квадратическим отклонением $\sigma(x)$ случайной величины X называется арифметическое значение квадратного корня из ее дисперсии:

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

Размерность величин $M(X)$ и $\sigma(X)$ совпадает с размерностью самой случайной величины X.

Пример 21

Закон распределения случайной величины X задан таблицей

x_i	– 2	– 1	0	1	2
p_i	0,1	0,2	0,4	0,2	0,1

Найти числовые характеристики случайной величины X.

Решение

Сначала найдем математическое ожидание случайной величины:

$$M(X) = \sum_{i=1}^{5} x_i \cdot p_i = (-2) \cdot 0{,}1 + (-1) \cdot 0{,}2 + 0 \cdot 0{,}4 + 1 \cdot 0{,}2 + 2 \cdot 0{,}1 = 0,$$

а затем математическое ожидание квадрата случайной величины:

$$M(X^2) = \sum_{i=1}^{5} x^2{}_i \cdot p_i = (-2)^2 \cdot 0{,}1 + (-1)^2 \cdot 0{,}2 + 0^2 \cdot 0{,}4 + 1^2 \cdot 0{,}2 + 4 \cdot 0{,}1 = 1{,}2.$$

Итак, дисперсия

$$D(X) = M(X^2) - [M(X)]^2 = 1{,}2 - 0 = 1{,}2.$$

Среднее квадратическое отклонение:

$$\sigma(X)=\sqrt{D(X)}=\sqrt{1{,}2}=1{,}1.$$

Пример 22

Известна функция распределения некоторой случайной величины X:

$$F(x)=\begin{cases}0, & x\le 0;\\ \dfrac{x^2}{49}, & 0<x\le 7;\\ 1, & x>7.\end{cases}$$

Найти $f(x)$, $M(X)$, $D(X)$.

Решение

$$f(x)=F'(x)=\begin{cases}0, & x\le 0;\\ \dfrac{2x}{49}, & 0<x\le 7;\\ 0, & x>7.\end{cases}$$

$$M(X)=\int_{-\infty}^{+\infty}x\cdot f(x)\,dx=\int_0^7 x\cdot\frac{2x}{49}\,dx=\left.\frac{2x^3}{3\cdot 49}\right|_0^7=\frac{14}{3}=4\frac{2}{3}.$$

$$D(X)=\int_{-\infty}^{+\infty}x^2\cdot f(x)dx-M^2(X)=\int_0^7\frac{2x^3}{49}\,dx-\left(\frac{14}{3}\right)^2=2\frac{13}{18}.$$

9.4. Мода и медиана случайной величины

Модой $M_0(X)$ случайной величины X называется ее наиболее вероятное значение, т. е. то значение, для которого вероятность p_i или плотность вероятности $f(x)$ достигает максимума.

Если вероятность или плотность вероятности достигает максимума не в одной, а в нескольких точках, распределение называется **полимодальным.**

Медианой $M_e(X)$ непрерывной случайной величины X называется такое ее значение, для которого $P(X < M_e(X)) = P(X > M_e(X)) = 0{,}5.$

Геометрически прямая $x = M_e(X)$ делит площадь фигуры под кривой распределения на две равные части.

Пример 23

Найти моду, медиану и математическое ожидание случайной величины X, плотность вероятности которой задана функцией

$$f(x) = \begin{cases} 0, & x \le 0; \\ 4x^3, & 0 < x \le 1; \\ 0, & x > 1. \end{cases}$$

Решение

Очевидно, что плотность вероятности максимальна при $x = 1$, следовательно, мода $M_0(X) = 1$.

Медиану $M_e(X) = c$ найдем из условия $\int\limits_{-\infty}^{c} f(x)dx = \frac{1}{2}$.

$$\int\limits_{-\infty}^{c} f(x)dx = \int\limits_{-\infty}^{0} 0 \cdot dx + \int\limits_{0}^{c} 4x^3 dx = 4 \cdot \left.\frac{x^4}{4}\right|_0^c = c^4 = \frac{1}{2},\ c = \sqrt[4]{\frac{1}{2}} = 0{,}84,$$

таким образом, $M_e(X) = 0{,}84.$

Найдем математическое ожидание:

$$M(X) = \int\limits_0^1 x \cdot 4x^3 dx = \left.\frac{4}{5}x^5\right|_0^1 = \frac{4}{5} = 0{,}8.$$

9.5. Начальные и центральные моменты, асимметрия, эксцесс

Среди числовых характеристик случайной величины особое значение имеют *моменты* – начальные и центральные.

Начальным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени этой величины:

$$\nu_k = M(X^k).$$

Центральным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени отклонения случайной величины от ее математического ожидания:

$$\mu_k = M[X - M(X)]^k.$$

Очевидно,

при $k = 1$, $\nu_1 = M(X)$ – первый начальный момент случайной величины X есть ее математическое ожидание;

при $k = 2$, $\mu_2 = D(X)$ – второй центральный момент случайной величины X есть ее дисперсия.

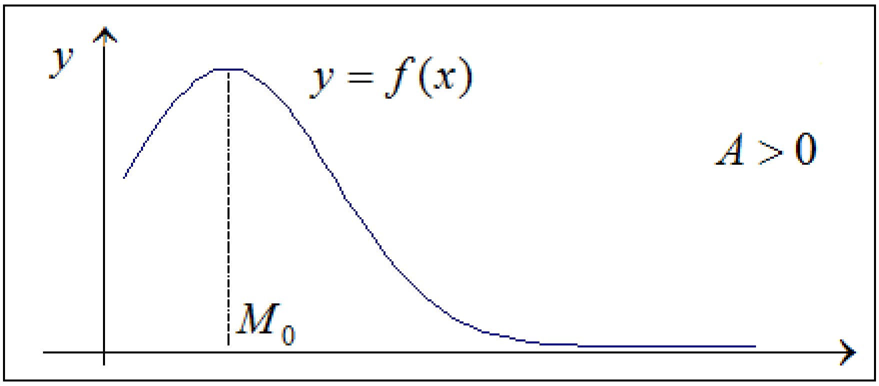
Третий центральный момент μ_3 служит для характеристики асимметрии (скошенности) распределения.

Коэффициентом асимметрии случайной величины X называется величина

$$A = \frac{\mu_3}{\sigma^3} = \frac{M[X - M(X)]^3}{\sigma^3}.$$

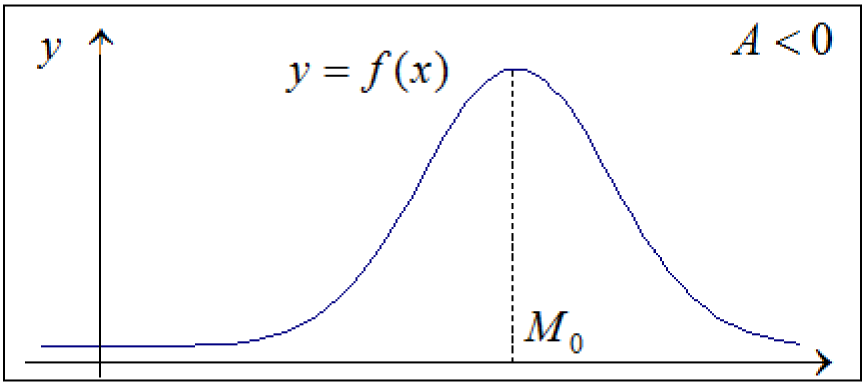
Имеют место следующие свойства коэффициента асимметрии:

- если распределение симметрично относительно математического ожидания, то коэффициент асимметрии $A = 0$;
- при $A > 0$ кривая функции плотности имеет положительную (правостороннюю) асимметрию (рис. 3, *а*);
- при $A < 0$ кривая функции плотности имеет отрицательную (левостороннюю) асимметрию (рис. 3, *б*).



а

Рисунок 3 – Кривая функции плотности



б

Окончание рис. 3

Четвертый центральный момент μ_4 служит для характеристики крутости (островершинности) распределения.

Коэффициентом эксцесса случайной величины называется число, равное

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{M\left[X - M(X)\right]^4}{\sigma^4} - 3.$$

10. СИСТЕМЫ ДВУХ СЛУЧАЙНЫХ ВЕЛИЧИН

10.1. Функция распределения двумерной случайной величины, ее свойства

Во многих практических задачах результат опыта описывается не одной случайной величиной, а некоторой системой случайных величин, которую называют также **многомерной (n-мерной) случайной величиной** или случайным вектором $\overline{X} = \left(X_1, X_2, \ldots, X_n\right)$.

Случайные величины $X_1, X_2, \ldots, X_n$, входящие в систему, могут быть как дискретными, так и непрерывными.

В частности $\overline{X} = \left(X_1, X_2\right)$ является **двумерной** случайной величиной.

Наиболее полным описанием многомерной случайной величины является закон ее распределения.

При конечном множестве возможных значений двумерной случайной величины такой закон может быть задан в виде таблицы, содержащей все возможные сочетания значений каждой из одномерных случайных величин и соответствующие им вероятности.

Если зафиксировать значение одного из аргументов, например $Y = y_j$, то полученное распределение случайной величины X называется ***условным распределением*** X при условии $Y = y_j$.

Вероятности $p_j(x_i)$ этого распределения называют ***условными вероятностями***, найденными при условии, что событие $Y = y_j$ произошло, обозначают $p_j(x_i) = \dfrac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \dfrac{p_{ij}}{p_j}$.

Функцией распределения двумерной случайной величины (X, Y) называется функция $F(x, y) = P(X < x, Y < y)$.

В случае дискретной двумерной случайной величины функция распределения определяется формулой

$$F(x, y) = \sum_i \sum_j p_{ij},$$

где суммирование вероятностей распространяется на все i, для которых $x_i < x$, и все j, для которых $y_j < y$.

Свойства функции распределения

1. Функция распределения $F(x, y)$ есть неотрицательная функция, для которой справедливо неравенство $0 \le F(x, y) \le 1$.

2. Функция распределения есть неубывающая функция по каждому из своих аргументов:

при $x_2 > x_1$, $F(x_2, y) \ge F(x_1, y)$,

при $y_2 > y_1$, $F(x\ , y_2) \ge F(x\ , y_1)$.

3. Если $x \to -\infty$ или $y \to -\infty$,

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

4. При $x, y \to \infty$, $F(x, y) = F(-\infty, +\infty) = 1$.

10.2. Числовые характеристики системы двух случайных величин

Счайные величины X и Y называются **независимыми,** если их функция распределения $F(x, y)$ представляется в виде произведения функций распределения этих случайных величин:

$$F(x, y) = F_1(x) \cdot F_2(x).$$

Зависимость между двумя случайными величинами называется ***вероятностной*** (*стохастической* или *статистической*), если каждому значению одной из них соответствует условное распределение другой.

Числовые характеристики ***непрерывной двумерной случайной величины*** (X,Y) ***определяются по формулам***

- математическое ожидание:

$$a_X = M(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\cdot f(x,y)\,dx\,dy,$$

$$a_Y = M(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\cdot f(x,y)\,dx\,dy,$$

- дисперсия

$$D(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-a_X)^2\cdot f(x,y)\,dx\,dy,$$

$$D(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y-a_Y)^2\cdot f(x,y)\,dx\,dy.$$

- **Условное математическое ожидание** случайной величины Y при $X=x$ $M_X(Y)$ есть функция от x, называемая ***функцией регрессии*** или просто *регрессией* Y по X; аналогично $M_Y(X)$ называется функцией регрессии или просто регрессией X по Y. Графики этих функций называются соответственно ***линиями регрессии.*** Степень зависимости между случайными величинами X и Y определяется такими числовыми характеристиками, как *ковариация* и *коэффициент корреляции*.

10.3. Коэффициент корреляции, его свойства

- ***Ковариацией*** (***корреляционным моментом***) K_{XY} случайных величин X и Y называется математическое ожидание произведения отклонений этих величин от своих математических ожиданий:

$$K_{XY} = M[(X-M(X))\cdot(Y-M(Y))]$$

или

$$K_{XY} = M[(X - a_X)\cdot(Y - a_Y)] = M(XY) - a_X \cdot a_Y.$$

Очевидно, что $K_{XY} = K_{YX}$.

Корреляционный момент характеризует как степень зависимости случайных величин, так и их рассеяние вокруг точки $(a_X; a_Y)$, кроме того это величина размерная, что затрудняет ее использования для оценки степени зависимости разных случайных величин.

- ***Коэффициентом корреляции*** двух случайных величин называется отношение корреляционного момента к произведению средних квадратических отклонений этих величин:

$$r_{XY} = \frac{K_{XY}}{\sigma(X)\cdot\sigma(Y)},$$

или

$$r_{XY} = \frac{M(XY) - a_X \cdot a_Y}{\sigma(X)\cdot\sigma(Y)}.$$

Из определения следует, что $r_{XY} = r_{YX} = r$. Очевидно также, что коэффициент корреляции r есть безразмерная величина.

Свойства коэффициента корреляции

1. Коэффициент корреляции принимает значения $-1 \le r \le 1$,

2. Если случайные величины **независимы**, то их коэффициент корреляции $r = 0$.

В этом случае случайные величины называются ***некоррелированными.***

3. Если коэффициент корреляции двух случайных величин $r = \pm 1$, то между этими случайными величинами существует **линейная функциональная зависимость** $y = ax + b$.

Пример 24

Закон распределения дискретной двумерной случайной величины задан таблицей:

x_i \ y_i	–2	0	2	3
2	0,10	0,25	0,30	0,15
3	0,10	0,05	0,00	0,05

Найти:

1) законы распределения одномерных случайных величин X и Y;

2) условные законы распределения случайной величины X при условии $Y=3$ и случайной величины Y при условии $X=2$;

3) найти ковариацию и коэффициент корреляции двумерной случайной величины *(X, Y).*

Решение

1) Случайная величина X может принимать значение $X=1$ с вероятностью $p_1 = 0{,}10+0{,}25+0{,}30+0{,}15 = 0{,}8$ и значение $X=2$ с вероятностью $p_2 = 0{,}10+0{,}05+0{,}05 = 0{,}2$. Поэтому закон распределения одномерной случайной величины X можно записать в виде

X	2	3
p	0,8	0,2

Аналогично, записывается закон распределения одномерной случайной величины Y:

Y	–2	0	2	3
p	0,2	0,3	0,3	0,2

2) Для того, чтобы записать условный закон распределения случайной величины X при условии, что $Y=3$, следует при расчете вероятностей каждую из вероятностей 0,15 и 0,05 разделить на их сумму: $\frac{0{,}15}{0{,}2}=0{,}75$; $\frac{0{,}05}{0{,}2}=0{,}25$, тогда

X	2	3
p	0,75	0,25

Аналогично, записывается условный закон распределения случайной величины Y при условии, что $X=2$:

Y	–2	0	2	3
p	0,125	0,3125	0,375	0,21875

3) В данной задаче в пункте 1) найдены законы распределения одномерных случайных величин X и Y:

X	2	3
p	0,8	0,2

Y	–2	0	2	3
p	0,2	0,3	0,3	0,2

Найдем математические ожидания и средние квадратические отклонения этих величин:

$$a_X = M(X) = 2\cdot 0{,}8 + 3\cdot 0{,}2 = 2{,}2;$$
$$a_Y = M(Y) = -2\cdot 0{,}2 + 0\cdot 0{,}3 + 2\cdot 0{,}3 + 3\cdot 0{,}2 = 0{,}8;$$

$$M(X^2) = 4\cdot 0{,}8 + 9\cdot 0{,}2 = 5;$$

$$D(X) = M(X^2) - [M(X)]^2 = 5 - 4{,}84 = 0{,}16;$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}16} = 0{,}4;$$

$$M(Y^2) = 4\cdot 0{,}2 + 0\cdot 0{,}3 + 4\cdot 0{,}3 + 9\cdot 0{,}2 = 3{,}8;$$

$$D(Y) = M(Y^2) - [M(Y)]^2 = 3{,}8 - 0{,}64 = 3{,}16;$$

$$\sigma(Y) = \sqrt{D(Y)} = \sqrt{3{,}16} = 1{,}78.$$

Найдем математическое ожидание двумерной случайной величины по исходной таблице, используя формулу

$$M(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i \cdot y_j \cdot p_{i\,j}.$$

$$M(XY) = 2\cdot(-2)\cdot 0{,}10 + 2\cdot 0\cdot 0{,}25 + 2\cdot 2\cdot 0{,}30 + 2\cdot 3\cdot 0{,}15 +$$
$$+\ 3\cdot(-2)\cdot 0{,}10 + 3\cdot 0\cdot 0{,}05 + 3\cdot 2\cdot 0 + 3\cdot 3\cdot 0{,}05 =$$
$$= -0{,}4 + 1{,}2 + 0{,}9 - 0{,}6 + 0{,}15 + 0{,}45 = 1{,}7.$$

Итак, ковариация и коэффициент корреляции заданной двумерной случайной величины

$$K_{XY} = M(XY) - a_X \cdot a_Y = 1{,}7 - 2{,}2\cdot 0{,}8 = -0{,}06.$$

$$r = \frac{K_{XY}}{\sigma(X)\cdot\sigma(Y)} = \frac{-0{,}06}{0{,}4\cdot 1{,}78} = -0{,}084.$$

11. ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ

11.1. Биномиальный закон распределения

Биномиальным называют закон распределения дискретной случайной величины X – числа появлений события в n независимых испытаниях, в каждом из которых вероятность наступления события одна и та же и равна p; вероятности возможных значений случайной величины $X = 0, 1, 2, \ldots, k, \ldots, n$ вычисляются по формуле Бернулли.

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}.$$

*Биномиальны*й закон распределения имеет вид

x_k	0	1	2	…	n
p_k	p_1	p_2	p_3	…	p_n

причем

$$P_k = P(X = k) = P_n(k),\ \sum_{k=0}^{n} p_k = 1.$$

Математическое ожидание случайной величины X, распределенной по биномиальному закону,

$$M(X) = np,$$

а ее дисперсия

$$D(X) = npq.$$

Биномиальный закон распределения широко используется в теории и практике статистического контроля качества продукции, при описании функционирования систем массового обслуживания и т. д.

Пример 25

Надежность работы одной ячейки доильной установки равна 0,8. Составить закон распределения случайной величины X (числа безотказно работающих ячеек доильной установки) во время дойки трех коров. Найти математическое ожидание $M(X)$, дисперсию $D(X)$, функцию распределения $F(x)$.

Решение

Из условия задачи следует, что $p=0{,}8;\quad q=0{,}2;\quad n=3$. Рассмотрим случайную величину X – числа безотказно работающих ячеек доильной установки, она принимает следующие значения: $x_0=0,\ x_1=1,\ x_2=2,\ x_3=3$. Таким образом, случайная величина X принимает 4 различных значения.

По формуле Бернулли найдем соответствующие им вероятности:

$$p_0=P(X=0)=P_3(0)=C_3^0\cdot 0{,}8^0\cdot 0{,}2^3=0{,}008;$$

$$p_1=P(X=1)=P_3(1)=C_3^1\cdot 0{,}8^1\cdot 0{,}2^2=0{,}096;$$

$$p_2=P(X=2)=P_3(2)=C_3^2\cdot 0{,}8^2\cdot 0{,}2^1=0{,}384;$$

$$p_3=P(X=3)=P_3(3)=C_3^3\cdot 0{,}8^3\cdot 0{,}2^0=0{,}512.$$

Биномиальный закон распределения данной случайной величины представляется следующей таблицей:

x_k	0	1	2	3
p_k	0,008	0,096	0,384	0,512

где $\sum\limits_{k=0}^{3} p_k=0{,}08+0{,}096+0{,}384+0{,}512=1.$

Функция распределения $F(x)=P(X<x)$ имеет вид

$$F(x)=\begin{cases}0, & x\le 0;\\ 0{,}008, & 0<x\le 1;\\ 0{,}104, & 1<x\le 2;\\ 0{,}488, & 2<x\le 3;\\ 1, & x>3.\end{cases}$$

Числовые характеристики:

$$M(X)=\sum_{k=0}^{3} p_k\cdot x_k=0\cdot 0{,}008+1\cdot 0{,}096+2\cdot 0{,}384+3\cdot 0{,}512=2{,}4$$

или $M(X)=np=3\cdot 0{,}8=2{,}4;$

$$D(X)=M(X^2)-[M(X)]^2=0\cdot 0{,}008+1\cdot 0{,}096+4\cdot 0{,}384+$$
$$+9\cdot 0{,}512-2{,}4^2=0{,}48$$

или $D(X)=npq=3\cdot 0{,}8\cdot 0{,}2=0{,}48.$

11.2. Закон распределения Пуассона

Закон распределения Пуассона – это закон распределения дискретной случайной величины X – числа появлений события в n независимых испытаниях, в каждом из которых вероятность наступления события одна и та же и равна p (число испытаний n велико, а вероятность p появления события в каждом испытании очень мала); вероятности возможных значений случайной величины $X=0, 1, 2, \ldots, k, \ldots, n$ вычисляются по приближенной формуле Пуассона:

$$P_n(k)\approx\frac{\lambda^k\cdot e^{-\lambda}}{k!},$$

где k – число появлений события в n независимых испытаниях;
$\lambda=np$ (среднее число появлений события в n испытаниях).

Математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона, совпадают и равны параметру этого закона: $M(X)=D(X)=\lambda$.

Пример 26

Вероятность того, что зерно пшеницы не прорастет, принимается равной 0,01. Найти математическое ожидание и дисперсию случайной величины X – число зерен, не давших всходов в партии из 1000 посаженных зерен.

Решение

Из условия задачи следует, что случайная величина X – число зерен, не давших всходов в этой партии, – распределена по закону Пуассона, она принимает значения 0,1,2,…,1000, а соответствующие им вероятности могут быть вычислены по формуле Пуассона ($n=1000$ достаточно велико, вероятность $p=0{,}01$ мала).

Определим параметр

$\lambda=np=1000\cdot 0{,}01=10$, тогда $M(X)=D(X)=\lambda=10.$

11.3. Равномерный закон распределения

Непрерывная случайная величина X распределена ***равномерно на отрезке*** $[a;b]$, если ее плотность вероятности имеет вид

$$f(x)=\begin{cases} 0, & при \quad x<a; \\ \dfrac{1}{b-a}, & при \quad a\le x\le b;. \\ 0, & при \quad x>b. \end{cases}$$

График функции плотности вероятности представлен на рисунке 4.

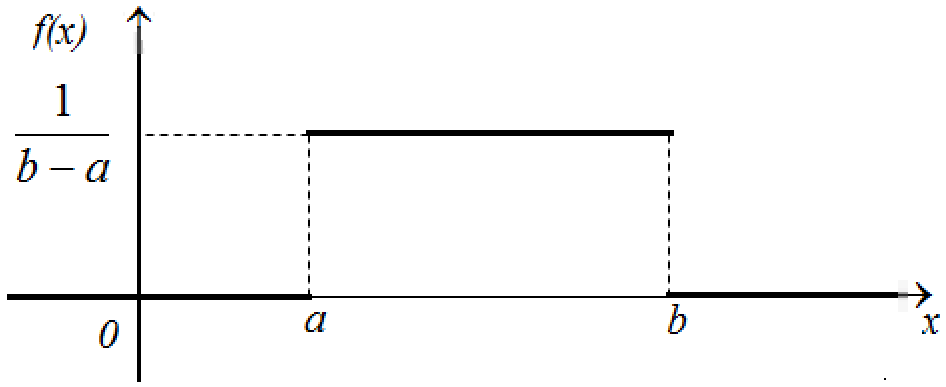


Рисунок 4 – График функции плотности вероятности

Упражнение

Показать самостоятельно, что

- математическое ожидание случайной величины X, распределенной равномерно на отрезке $[a;b]$ равно

$$M(X)=\frac{a+b}{2};$$

- дисперсия этой случайной величины равна $D(X)=\dfrac{(b-a)^2}{12}$.

Пример 27

На молочной ферме коровы ожидают своей очереди на доение в доильном зале. Время ожидания для каждой коровы составляет от 10 до 30 мин, причем каждое значение в этом диапазоне является равновероятным.

Найти:

а) плотность вероятности;

б) функцию распределения;

в) вероятность того, что случайно выбранная корова будет ждать менее 15 минут;

г) математическое ожидание и дисперсию.

Решение

а) Плотность вероятности равномерно распределенной случайной величины имеет вид

$$f(x)=\begin{cases} 0, & при \quad x<a; \\ \dfrac{1}{b-a}, & при \quad a\le x\le b; \\ 0, & при \quad x>b. \end{cases}$$

Интервал ожидания коровы на дойку составляет от 10 до 30 мин,

тогда $f(x)=\begin{cases} 0, & x<10; \\ \dfrac{1}{20}, & 10\le x\le 30; \\ 0, & x>30. \end{cases}$

б) Функцию распределения найдем, согласно определению

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=\begin{cases} 0, & x<10; \\ \dfrac{1}{20}\cdot x, & 10\le x\le 30; \\ 1, & x>30. \end{cases}$$

в) Вероятность того, что время ожидания не превзойдет 15 мин, найдем, используя свойства функции распределения:

$$P(10\le x\le 15)=F(15)-F(10)=0{,}05\cdot 15-0{,}05\cdot 10=0{,}05\cdot 5=0{,}25.$$

г) Найдем математическое ожидание и дисперсию:

$$M(X)=\frac{a+b}{2}=\begin{vmatrix}a=10\\b=30\end{vmatrix}=\frac{10+30}{2}=20,$$

$$D(X)=\frac{(b-a)^2}{12}=\frac{(30-10)^2}{12}=\frac{20^2}{12}=33{,}33.$$

11.4. Нормальный закон распределения

Нормальным распределением, или ***распределением Гаусса***, называется непрерывная случайная величина, плотность вероятности которой имеет вид

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

Числа a, σ $(\sigma>0)$ называются ***параметрами нормального распределения,*** обозначают нормальное распределение $N(a;\sigma)$.

График функции плотности ***нормальной,*** или ***Гауссовой кривой*** представлен на рисунке 5.

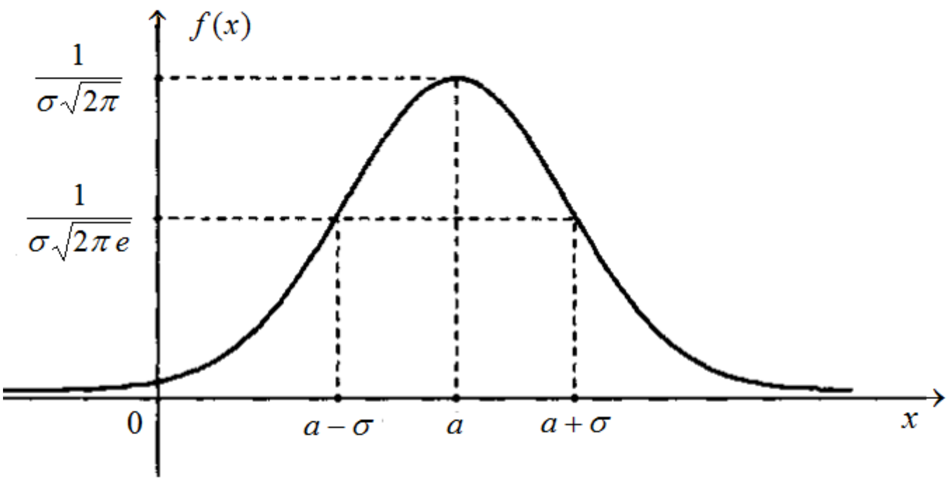


Рисунок 5 – График функции плотности нормальной, или Гауссовой кривой

Заметим, что график функции плотности нормального распределения симметричен относительно прямой $x=a$, имеет максимум в точке $x = a$, равный $\frac{1}{\sigma\sqrt{2\pi}}$, и две точки перегиба при $x = a \pm \sigma$.

Функцией распределения нормальной случайной величины является нормированная ***функция Лапласа***

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{t^2}{2}} dt,$$

значения которой представлены в таблице (приложение В).

Можно показать, что для нормальной случайной величины $X \sim N(a; \sigma)$:

$$M(X) = a, \; D(X) = \sigma^2(X).$$

Если $X \sim N(0; 1)$, то нормальное распределение называют **стандартным**.

Функцией распределения $X \sim N(0; 1)$ является функция Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Ранее было отмечено, что $\Phi(x) = 0{,}5 + \Phi_0(x)$.

Пример 28

Записать функцию плотности вероятности нормально распределенной случайной величины X, если $M(X) = 5, \; D(X) = 4$.

Решение

$a = M(X) = 5, \quad \sigma^2(X) = D(X) = 4, \quad \sigma(X) = 2,$ следовательно, функция плотности нормального распределения *N(5;2)* имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} = \frac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{(x-5)^2}{8}}.$$

Свойства случайной величины, распределенной по нормальному закону

1. Вероятность попадания случайной величины X, распределенной по нормальному закону, в интервал (α, β)

$$P(\alpha < x < \beta) = \Phi_0\left(\frac{\beta - a}{\sigma}\right) - \Phi_0\left(\frac{\alpha - a}{\sigma}\right).$$

2. Вероятность того, что отклонение нормально распределенной случайной величины X от ее математического ожидания a по абсолютной величине не превосходит заданного положительного числа δ,

$$P(|X - a| < \delta) = 2\Phi_0\left(\frac{\delta}{\sigma}\right).$$

3. ***Правило трех сигм.*** Вероятность того, что отклонение нормально распределенной случайной величины X от ее математического ожидания a по абсолютной величине не превосходит утроенного среднего квадратического отклонения, равна 0,9973:

$$P(|X - a| < 3\sigma) = 2\Phi_0(3) = 0{,}9973.$$

На практике это правило применяют так: если распределение изучаемой случайной величины неизвестно, но условие правила трех сигм выполняется, то есть основания предполагать, что изучаемая величина распределена нормально, в противном случае она подчиняется другому закону распределения.

Пример 29

Допустим, что среднее содержание гумуса в различных образцах почвы на поле составляет 3,5 % с дисперсией 0,64 %. Согласно правилу трех сигм, определить диапазон значений содержания гумуса, который будет наблюдаться в 99,7 % случаев (гумус – это органическое вещество почвы, характеризующее ее плодородность).

Решение

Пусть X – случайная величина содержания гумуса в образце почвы, которая распределена нормально. По условию имеем

$$a = M(X) = 3{,}5; \quad \sigma^2(X) = D(X) = 0{,}64; \quad \sigma(X) = \sqrt{D(X)} = 0{,}8.$$

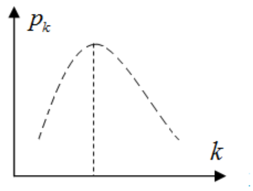
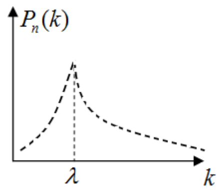
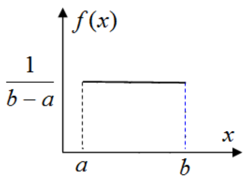
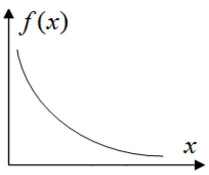
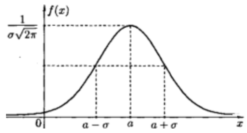
Поскольку случайная величина распределена нормально, диапазон значений процентного содержания гумуса в различных образцах почвы на поле, который будет наблюдаться в 99,7 % случаев, определяется интервалом

$$(a - 3\sigma;\, a + 3\sigma) = (3{,}5 - 3\cdot 0{,}8;\, 3{,}5 + 3\cdot 0{,}8) = (3{,}5 - 2{,}4;\, 3{,}5 + 2{,}4) = (1{,}1;\, 5{,}9).$$

11.5. Числовые характеристики законов распределения

Обобщим данные о законах распределения с помощью таблицы 1.

Таблица 1 – Числовые характеристики
некоторых законов распределения

Распределение	График	Числовые характеристики			
		$M(X)$	$D(X)$	A	E
Биномиальное $p_k = C_n^k p^k q^{n-k}$ $(k = 0, 1, \dots, n)$		np	npq	$\dfrac{1-2p}{\sqrt{npq}}$	$3+\dfrac{1-6pq}{npq}$
Пуассоновское $P_n(k) = \dfrac{\lambda^k}{k!} \cdot e^{-\lambda}$ $(\lambda > 0)$ $(k = 0, 1, \dots, n)$		$\lambda = np$	$\lambda = np$	$\dfrac{1}{\sqrt{\lambda}}$	$3+\dfrac{1}{\lambda}$
Равномерное $f(x) = \dfrac{1}{b-a}$, $x \in [a;b]$; $f(x) = 0,\ x \notin [a;b]$		$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	0	1,8
Показательное (экспоненциальное) $f(x) = \lambda \cdot e^{-\lambda x}$, $x \in [a;b]$		$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	2	9
Нормальное $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$		a	σ^2	0	3

12. ЗАКОН БОЛЬШИХ ЧИСЕЛ

Под законом больших чисел понимается общий принцип, согласно которому совокупное действие большого числа случайных факторов приводит к результату, почти не зависящему от случая. Другими словами, при большом числе случайных величин их средний результат перестает быть случайным и может быть предсказан с большой степенью определенности.

12.1. Неравенство Чебышева

Неравенство Маркова (*лемма Чебышева*).

Если случайная величина X принимает только неотрицательные значения и имеет математическое ожидание $M(X)$, то для любого положительного числа ε имеет место неравенство

$$P(X \geq \varepsilon) \leq \frac{M(X)}{\varepsilon}.$$

Это неравенство равносильно неравенству $P(X < \varepsilon) \geq 1 - \dfrac{M(X)}{\varepsilon}$.

Пример 30

Предположим, что агроном изучает урожайность определенного поля. Пусть X – случайная величина, представляющая урожайность в тоннах с гектара, зафиксированную случайным образом на нескольких участках поля. Допустим, известно, что средняя урожайность составляет 3 тонны с гектара. Агроному надо оценить вероятность того, что на каком-либо участке урожайность составит более 5 тонн с гектара.

Решение

По условию задачи $M(X) = 3,\ \ \varepsilon = 5$. Тогда по неравенству Маркова имеем

$$P(X \geq 5) \leq \frac{3}{5} = 0{,}6.$$

Это означает, что вероятность того, что урожайность на каком-либо участке будет больше 5 тонн с гектара, не превосходит 60 %.

Таким образом, используя неравенство Маркова, агроном может сделать вывод, что вероятность высокой урожайности достаточно низкая, поскольку средняя урожайность составляет всего 3 тонны с гектара.

Неравенство Чебышева. Если случайная величина X имеет математическое ожидание $M(X)$ и дисперсию $D(X)$, то для любого $\varepsilon > 0$ имеет место неравенство

$$P\left(\left|X - M(X)\right| > \varepsilon\right) \le \frac{D(X)}{\varepsilon^2}.$$

Неравенство Чебышева можно записать в другой форме:

$$P\left(\left|X - M(X)\right| \le \varepsilon\right) > 1 - \frac{D(X)}{\varepsilon^2}.$$

Неравенство Чебышева применимо для любых случайных величин. В первой форме оно устанавливает верхнюю границу, а во второй форме – нижнюю границу вероятности рассматриваемого события.

Запишем неравенство Чебышева в виде

$$P\left(\left|X - a\right| > \varepsilon\right) \le \frac{\sigma^2(X)}{\varepsilon^2},$$

где $M(X) = a$; $D(X) = \sigma^2(X)$.

Другими словами для любого $\varepsilon > 0$, если $\sigma^2(X)$ мало, то большие отклонения от a маловероятны.

Доказательство: пусть X непрерывная случайная величина, тогда

$$P(|X - a| > \varepsilon) = 1 - P(|X - a| \le \varepsilon = \int_{-\infty}^{+\infty} f(x)\,dx - \int_{a-\varepsilon}^{a+\varepsilon} f(x)\,dx =$$

$$= \int_{-\infty}^{a-\varepsilon} f(x)\,dx - \int_{a+\varepsilon}^{+\infty} f(x)\,dx = \int_{|x-a|>\varepsilon} f(x)\,dx \le \int_{|x-a|>\varepsilon} \frac{(x-a)^2}{\varepsilon^2} f(x)\,dx =$$

$$= \frac{1}{\varepsilon^2} \int_{|x-a|>\varepsilon} (x-a)^2 f(x)\,dx < \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (x-a)^2 f(x)\,dx = \frac{1}{\varepsilon^2} D(x).$$

Замечание. Другая форма записи неравенства Чебышева:

$$P\left(\left|X - a\right| \le \varepsilon\right) \le 1 - \frac{\sigma^2(X)}{\varepsilon^2}.$$

Пример 31

Оценить вероятность того, что отклонение любой случайной величины от ее математического ожидания будет по абсолютной величине не более трех средних квадратических отклонений (***правило трех сигм***).

Решение

По условию имеем $D(X)=\sigma^2(X)$, $\varepsilon=3\cdot\sigma(X)$, тогда согласно неравенству Чебышева

$$P(|X-M(X)|\leq 3\sigma(X))>1-\frac{\sigma^2(X)}{9\cdot\sigma^2(X)}=\frac{8}{9}=0{,}889,$$

т. е. не менее, чем 0,899.

Заметим, что для нормального распределения правило трех сигм выполняется с вероятностью 0,9973, для равномерного распределения – с вероятностью 1, для показательного – с вероятностью 0,9827. Таким образом, правило трех сигм с достаточно большой вероятностью его выполнения применимо для большинства случайных величин, встречающихся на практике.

12.2. Закон больших чисел в форме Чебышева

Теорема Чебышева

Если случайные величины $X_1, X_2,\ldots,X_n$ независимы, а их дисперсии ограничены постоянной C: $D(X)\leq C$, то для любого $\varepsilon>0$ среднее арифметическое большого числа случайных величин как угодно мало отличается (с вероятностью близкой к 1) от среднего арифметического их математических ожиданий:

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}M(X_k)\right|<\varepsilon\right)=1.$$

Докажем утверждение теоремы Чебышева, записанное в виде

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}a_k\right|>\varepsilon\right)\to 0,$$

где $M(X_k)=a_k$.

Доказательство

Обозначим $\frac{1}{n}\sum_{k=1}^{n} X_k = Y_n$; $\frac{1}{n}\sum_{k=1}^{n} a_k = M(Y_n)$, тогда, применяя неравенство Чебышева, получим

$$P\left(|Y_n - M(Y_n)| > \varepsilon\right) \le \frac{D(Y_n)}{\varepsilon^2} = \frac{D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)}{\varepsilon^2} = \frac{1}{\varepsilon^2}\cdot\frac{1}{n^2}\sum_{k=1}^{n} D(X_k) <$$

$$< \frac{1}{\varepsilon^2}\cdot\frac{1}{n^2}\cdot n\cdot C = \frac{C}{n\varepsilon^2} \to 0 \text{ при } n \to \infty.$$

Из ***теоремы Чебышева следует,*** что, если независимые случайные величины $X_1, X_2, \ldots, X_n$ имеют одинаковые математические ожидания, равные $M(X_i) = a$, и их дисперсии ограничены постоянной C: $D(X) \le C$, то для любого $\varepsilon > 0$ имеет место неравенство

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - a\right| \le \varepsilon\right) \ge 1 - \frac{C}{n\cdot\varepsilon^2}.$$

.

Последнее утверждение запишем в виде

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - a\right| > \varepsilon\right) \to 0.$$

Итак, с вероятностью равной единице, при $n \to \infty$ имеет место

$$\frac{1}{n}\sum_{k=1}^{n} X_k \approx a.$$

Пример 32

Для определения среднего содержания пектина в партии из 200 одинаковых ящиков с яблоками случайным образом было выбрано по одному яблоку из каждого ящика. Оценить вероятность того, что среднее содержание пектина в отобранных образцах 200 яблок отличается от среднего содержания пектина в яблоках во всей партии не более чем на 1 % (по абсолютной величине), если известно, что среднее квадратическое отклонение содержания пектина в каждом ящике меньше 2 %.

Пектин – это полисахарид (высокомолекулярный углевод), который естественным образом содержится в стенках клеток фруктов, ягод, овощей.

Решение

Пусть X_i– содержание пектина в яблоке, взятого из i-го ящика. По условию дисперсия $D(X_i) < 2^2 = 4$. Очевидно, что среднее содержание пектина в отобранных яблоках равна $(X_1 + X_2 + ... + X_n)/200$, а среднее содержание пектина в яблоках всей партии

$$(M(X_1) + M(X_2) + ... + M(X_{200}))/200,$$

тогда вероятность

$$P\left(\left|\frac{X_1 + X_2 + ... + X_n}{200} - \frac{M(X_1) + M(X_2) + ... + M(X_n)}{200}\right| \le 1\right) \ge 1 - \frac{4}{200 \cdot 1^2} = 0{,}98$$

Пример 32

Найти, сколько надо провести измерений данной случайной величины, чтобы с вероятностью не менее 0,95 гарантировать отклонение средней арифметической этих измерений от истинного значения величины не более чем на 1 (по абсолютной величине), если среднее квадратическое отклонение каждого из измерений не превосходит 5.

Решение

Пусть X_k– результат k-го измерения – истинное значение величины, т. е. $M(X_k) = a$ при любом k. Необходимо найти n, при котором $P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - a\right| \le 1\right) \ge 0{,}95$. Данное неравенство будет выполняться, если

$$1 - \frac{C}{n \cdot \varepsilon^2} = 1 - \frac{5^2}{n \cdot 1^2} \ge 0{,}95, \qquad \frac{25}{n} \le 0{,}05, \qquad n \ge \frac{25}{0{,}05} = 500.$$

Т. е. требуется провести не менее 500 измерений.

12.3. Центральная предельная теорема

Закон больших чисел устанавливает факт приближения средней большого числа случайных величин к определенным постоянным. Но этим не ограничиваются закономерности, возникающие в результате суммирования случайных величин. Оказывается, что при некоторых условиях совокупное действие случайных величин приводит к определенному, а именно – к нормальному закону распределения.

Центральная предельная теорема представляет собой группу теорем, посвященных установлению условий, при которых возникает нормальный закон распределения. Среди этих теорем важнейшее место принадлежит теореме Ляпунова.

Теорема Ляпунова. Если $X_1, X_2, ..., X_n$ – независимые случайные величины, у каждой из которых существует математическое ожидание $M(X_k) = a_k$, дисперсия $D(X_k) = \sigma_k{}^2$, абсолютный центральный момент третьего порядка $M\left(\left|X_k - a_k\right|^3\right)$, $\sum_{k=1}^{n} \sigma_k^2 = \beta_n^2$ и выполняется условие $\frac{1}{\beta_m^3} \sum_{k=1}^{n} M(x_k - a_k)^3 \to 0$ при $n \to \infty$, тогда

$$P\left(\frac{1}{\beta_n} \sum_{k=1}^{n} (x_k - a_k) < x\right) \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt,$$

где $\Phi(x)$ – функция Лапласа – функция стандартного нормального распределения $N(0;1)$.

Итак, закон распределения случайной величины $Y_n = X_1 + X_2 + ... + X_n$ при $n \to \infty$ неограниченно приближается к нормальному с математическим ожиданием $\sum_{i=1}^{n} a_k$ и дисперсией $\sum_{i=1}^{n} \sigma_k^2$.

Другими словами, большинство распределений, которые встречаются на практике, являются приближенно нормальными.

13. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Математическая статистика – раздел математики, изучающий методы сбора, систематизации и обработки результатов наблюдений (измерений) массовых случайных явлений с целью выявления закономерностей.

13.1. Выборка, дискретный и интервальный вариационные ряды

Совокупность всех возможных значений случайной величины X называется ***генеральной совокупностью (или популяцией)***.

Совокупность случайно отобранных из генеральной совокупности n значений (**вариант**) x_i называется ***выборкой***.

Число объектов в *выборке* n называют ***объемом выборки***.

Метод, основанный на том, что по данным выборки, выделенной из генеральной совокупности, делается заключение обо всей генеральной совокупности, называется ***выборочным методом***.

Выборка называется ***репрезентативной***, если ее значения отобраны случайно, т. е. все объекты генеральной совокупности имеют равную вероятность попасть в выборку.

Вариационным рядом называется таблица, состоящая из расположенных в порядке возрастания вариант x_i, соответствующих им частот n_i и (или) относительных частот w_i.

Число n_i, показывающее, сколько раз встречается варианта x_i, называется ***частотой***.

Относительной частотой (***частостью***) варианты x_i называется число $w_i = \frac{n_i}{n}$, где n – объем выборки.

Если значения вариант x_i таковы, что отличаются друг от друга не меньше, чем на некоторую конечную величину (например тарифные разряды отличаются друг от друга не меньше, чем на единицу), то вариационный ряд называется ***дискретным***.

Накопленной частотой n_x называется число, показывающее, сколько раз случайная величина X приняла значение, меньшее заданного значения x_i. Отношение накопленной частоты к общему числу наблюдений называют ***накопленной частостью*** и обозначают

$$w_x = \frac{n_x}{n}.$$

Если значения варианты x_i, которые может принимать случайная величина X, могут отличаться одно от другого на сколь угодно малую величину, либо число вариант x_i слишком велико, то данные группируют в интервалы.

Таблицу распределения частот (или частостей) по интервалам называют ***интервальным вариационным рядом***.

Для построения интервального вариационного ряда нужно определить оптимальный интервал h по формуле Стерджеса

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n},$$

где $x_{\max}$ и $x_{\min}$– максимальные и минимальные варианты; n – объем выборки. Число $k = 1 + 3{,}322 \cdot \lg n$ – число интервалов (за k берется ближайшее целое число). За начало первого интервала принимают

$$a_1 = x_{\min} - \frac{h}{2}.$$

Пример 33

Составить интервальный вариационный ряд частот n_i, относительных частот w_i и накопленных относительных частот $w_x = \frac{n_x}{n}$ некоторой величины X, результаты измерений которой приведены в таблице:

12	9	11	13	11	13	12	15	14	8
13	11	7	12	9	12	10	11	16	14
14	10	12	11	13	8	11	12	9	12
11	14	9	10	12	12	10	13	17	11
14	9	13	11	5	9	12	9	11	12
15	14	7	9	12	11	7	12	13	10
13	9	11	12	10	14	11	10	12	15
11	12	10	11	7	9	14	13	15	10
16	12	14	8	13	10	18	9	11	16
11	13	9	14	10	13	10	15	8	10

Решение

Статистическая обработка результатов эксперимента в случае выборки большого объема, в данной задаче $n = 100$, начинается с группировки выборочных значений, т. е. с разбиения наблюдаемых значений случайной величины на k частичных интервалов равной длины и подсчета частот попадания значений случайной величины в частичные интервалы.

Из данной таблицы найдем $x_{\max} = 18,\ x_{\min} = 5.$

Воспользуемся формулой Стерджеса:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \cdot \lg n} = \frac{18 - 5}{1 + 3{,}322 \cdot \lg 100} = \frac{21}{1 + 3{,}322 \cdot 2} \approx 1{,}86 \approx 2.$$

В качестве левой границы первого интервала возьмем величину, равную

$$a_1 = x_{\min} - \frac{h}{2} = 5 - \frac{2{,}0}{2} = 5 - 1 = 4.$$

Если a_1 – начало 1-го интервала, тогда

$$a_2 = a_1 + h = 4 + 2 = 6;$$
$$a_3 = a_2 + h = 6 + 2 = 8;$$
$$a_4 = a_3 + h = 8 + 2 = 10;$$
$$a_5 = a_4 + h = 10 + 2 = 12;$$
$$a_6 = a_5 + h = 12 + 2 = 14;$$
$$a_7 = a_6 + h = 14 + 2 = 16;$$
$$a_8 = a_7 + h = 16 + 2 = 18.$$

Заполним таблицу 2 с параметрами вариационных рядов.

Первый и четвертый столбцы таблицы составляют интервальный статистический ряд частот n_i.

Первый и пятый столбцы таблицы составляют интервальный статистический ряд относительных частот w_i.

Первый и шестой столбцы таблицы составляют интервальный статистический ряд накопленных относительных частот $w_x = \frac{n_x}{n}$.

Дискретный вариационный ряд частот задается *вторым и четвертым* столбцами.

Дискретный вариационный ряд относительных частот задается *вторым и пятым* столбцами.

Таблица 2 – Параметры интервального и дискретного вариационных рядов

Интервалы $(a_i; a_{i+1}]$	Середины интервалов $x_i = \dfrac{a_i + a_{i+1}}{2}$	Подсчет частот	Частоты n_i	Относит. частоты $w_i = \dfrac{n_i}{n}$	Накопленные относительные частоты $w_x = \dfrac{n_x}{n}$
(4; 6]	5		1	0,01	0,01
(6; 8]	7		8	0,08	0,09
(8; 10]	9		25	0,25	0,34
(10; 12]	11		34	0,34	0,68
(12; 14]	13		22	0,22	0,90
(14; 16]	15		8	0,08	0,98
(16; 18]	17		2	0,02	1,00

13.2. Графическое изображение вариационных рядов. Полигон, гистограмма, график эмпирической функции распределения

- ***Эмпирической функцией распределения*** называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$, т. е. $w_x(X \le x)$, и обозначают

$$F^*(x) = \frac{n_x}{n},$$

где n_x – накопленная частота, n – объем выборки.

Свойства эмпирической функции распределения

1. Значения функции принадлежат отрезку $[0; 1]$.
2. $F^*(x)$ – неубывающая функция.
3. Если x_1 – наименьшая варианта, то при $x \le x_1$

$$F^*(x) = 0;$$

4. Если x_k – наибольшая варианта, то при $x > x_k$

$$F^*(x) = 1.$$

Эмпирическая функция распределения $F^*(x)$ служит для оценки теоретической функции распределения $F(x)$ генеральной совокупности.

Имеет место равенство (теорема ***Гливенко***)***:***

$$\lim_{n\to\infty} P\left(\left| F(x) - F^*(x) \right| < \varepsilon\right) = 1,$$

где $\varepsilon > 0$ как угодно мало.

- ***Полигоном частот*** называют ломаную, отрезки которой соединяют точки с координатами $(x_1;n_1),(x_2;n_2),...,(x_n;n_k)$, где x_i – варианты выборки; n_i– соответствующие им частоты; аналогично определяется ***полигон относительных частот.***

- ***Гистограммой частот*** называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высоты равны отношению n_i/h (плотность частоты); аналогично определяется ***гистограмма относительных частот***.

Пример 34

Найти эмпирическую функцию распределения, построить полигон, гистограмму, график эмпирической функции распределения по интервальному вариационному ряду примера 33.

Решение

Используя данные последнего столбца таблицы 2, составим эмпирическую функцию распределения $F^*(x) = \dfrac{n_x}{n}$:

$$F^*(x) = \begin{cases} 0, & \text{если} \quad 0 < x \le 5; \\ 0{,}01, & \text{если} \quad 5 < x \le 7; \\ 0{,}09, & \text{если} \quad 7 < x \le 9; \\ 0{,}34, & \text{если} \quad 9 < x \le 11; \\ 0{,}68, & \text{если} \quad 11 < x \le 13; \\ 0{,}90, & \text{если} \quad 13 < x \le 15; \\ 0{,}98. & \text{если} \quad 15 < x \le 17; \\ 1, & \text{если} \quad x > 17. \end{cases}$$

Построим график эмпирической функции распределения (рис. 6).

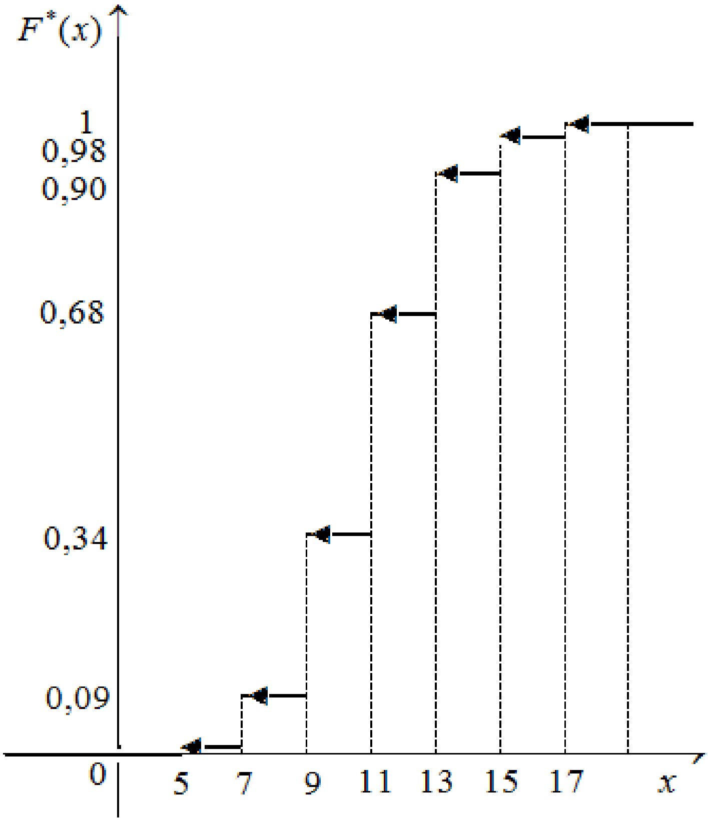


Рисунок 6 – График эмпирической функции распределения

Построим гистограмму интервального ряда (рис. 7) и полигон соответствующего дискретного ряда (столбцы два и четыре таблицы 2), соединяющего точки с координатами $(x_i; n_i)$ (рис. 8).

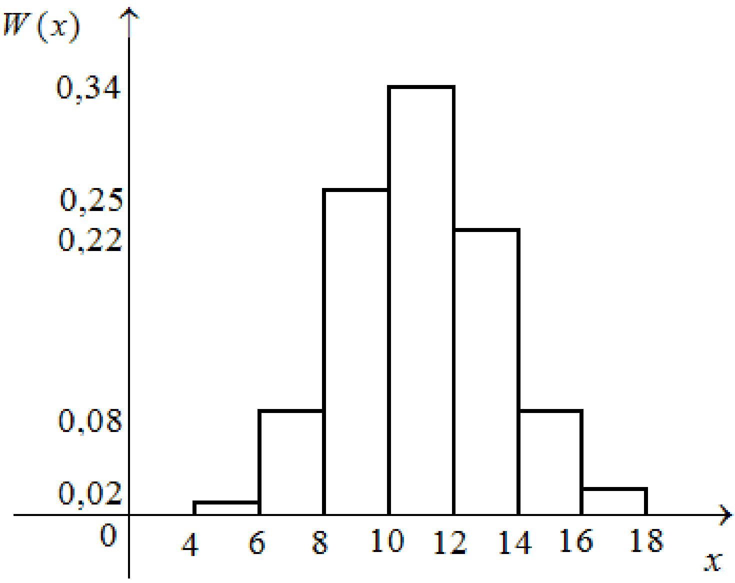


Рисунок 7 – Гистограмма интервального ряда

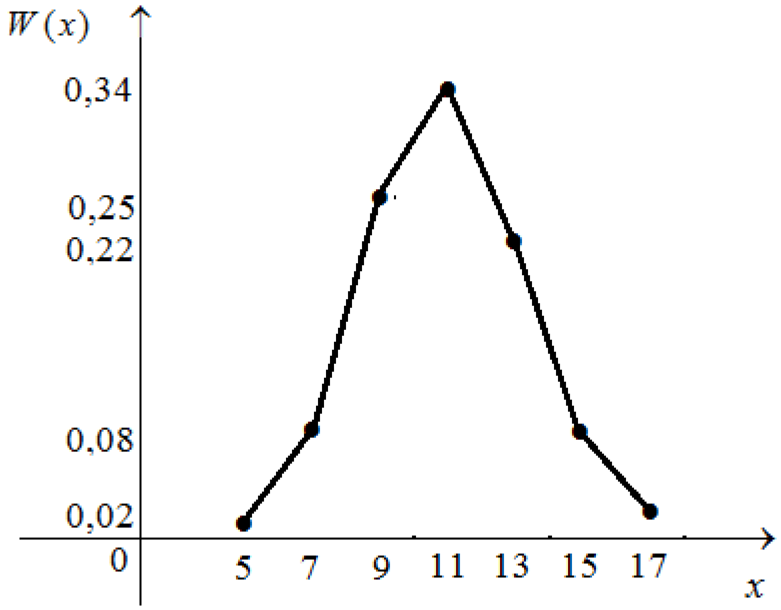


Рисунок 8 – Полигон дискретного ряда

14. ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

14.1. Выборочные числовые характеристики

Числовые характеристики случайной величины, вычисленные по выборке, называют выборочными или эмпирическими.

Рассмотрим статистическое распределение выборки объемом n:

x_i	x_1	x_2	x_3	…	x_k
n_i	n_1	n_2	n_3	…	n_k

Выборочной средней называется средняя арифметическая всех значений выборки:

$$\overline{X}_B = \frac{1}{n}\sum_{i=1}^{k} x_i n_i; \text{ или } \overline{X}_B = \sum_{i=1}^{k} x_i w_i;$$

где x_i– варианты; n_i– соответствующие частоты; n – объем выборки; $w_i = \dfrac{n_i}{n}$ – относительная частота.

Выборочной дисперсией D_B называют среднее арифметическое значение квадратов отклонений x_i от выборочной средней $\overline{X}_B$.

Для сгруппированных выборочных данных имеем формулу

$$D_B(X) = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{X}_B)^2 \cdot n_i.$$

Применяют также другую формулу для вычисления дисперсии:

$$D_B(X) = \frac{1}{n}\sum_{i=1}^{k} x_i^2 \cdot n_i - (\overline{X}_B)^2.$$

Выборочным средним квадратическим отклонением $\sigma_B(X)$ называется арифметическое значение квадратного корня из выборочной дисперсии.

$$\sigma_B(X) = \sqrt{D_B(X)}.$$

Исправленная выборочная дисперсия:

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{X}_B)^2 \cdot n_i.$$

Исправленное среднее квадратическое отклонение:

$$S_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{X}_B)^2 \cdot n_i}.$$

Коэффициент вариации – характеристика изменчивости, равная процентному отношению среднего квадратического отклонения $\sigma_B(X)$ к выборочной средней $\overline{X}_B$ (при $\overline{X}_B \neq 0$):

$$\tilde{v}V_B(X) = \frac{\sigma_B(X)}{\overline{X}_B} \cdot 100\%.$$

Если коэффициент вариации высок (например более 100 %), то это свидетельствует о неоднородности значений признака.

Медианой $M_e^*(X)$ называется середина вариационного ряда:

1) если $n = 2m$ – четное число дискретных наблюдений, то

$$M_e^*(X) = \frac{x_m + x_{m+1}}{2};$$

2) если $n = 2m+1$ – нечетное число дискретных наблюдений, то

$$M_e^*(X) = x_m.$$

Модой $M_o^*(X)$ называется такое значение x_i, которое наблюдается наибольшее число раз.

Мода – это значение x_i, повторяющееся с наибольшей частотой, поэтому моду используют в случаях, когда нужно ответить на вопрос, какая продукция имеет наибольший спрос, каковы преобладающие в данный момент уровни производительности труда, себестоимости продукции и т. д.

Замечание. Медиана и мода предпочтительнее средней арифметической в тех случаях, когда крайние варианты по сравнению с остальными оказались чрезмерно большими или малыми. Медиана и мода не изменяются при изменении крайних членов вариационного ряда, т. е. обладают определенной устойчивостью к изменению признака.

Под ***выборочным центральным моментом k-го порядка*** понимается средняя взвешенная k-й степени отклонения x_i от выборочной средней $\overline{X}_B$:

$$\mu_k^*(X) = \frac{1}{n}\sum_i (x_i - \overline{X}_B)^k \cdot n_i.$$

Выборочным коэффициентом асимметрии $A_B(X)$ называется отношение эмпирического центрального момента третьего порядка $\mu_3^*(X)$ к кубу эмпирического среднего квадратического отклонения $\sigma_B^3(X)$:

$$A_B(X) = \frac{\mu_3^*(X)}{\sigma_B^3(X)} = \frac{\sum_i (x_i - \overline{X}_B)^3 \cdot n_i}{n \cdot \sigma_B^3(X)}.$$

Если $A_B(X) = 0$, то распределение признака имеет симметричную форму; если преобладают варианты $x_i < \overline{X}_B$, то $A_B(X) < 0$ и говорят, что имеет место левосторонняя асимметрия, так как левая ветвь полигона длиннее правой. Если же преобладают варианты $x_i > \overline{X}_B$, то $A_B(X) > 0$ и имеет место правосторонняя асимметрия, так как более длинной является правая ветвь полигона.

Выборочным эксцессом, или коэффициентом крутости полигона $E_B(X)$ называется уменьшенное на 3 единицы отношение эмпирического центрального момента четвертого порядка $\mu_4^*(X)$ к четвертой степени среднего квадратического отклонения:

$$E_B(X)=\frac{\mu_4^*(X)}{\sigma_B^4(X)}-3=\frac{\sum_i (x_i-\overline{X}_B)^4\cdot n_i}{n\cdot\sigma_B^4(X)}-3.$$

За стандартное значение эксцесса принимают $E_B(X)=0$ – эксцесс нормальной кривой, если $E_B(X)<0$ – «плосковершинные кривые», если $E_B(X)>0$ – «островершинные кривые» распределения по сравнению с нормальной кривой (рис. 9).

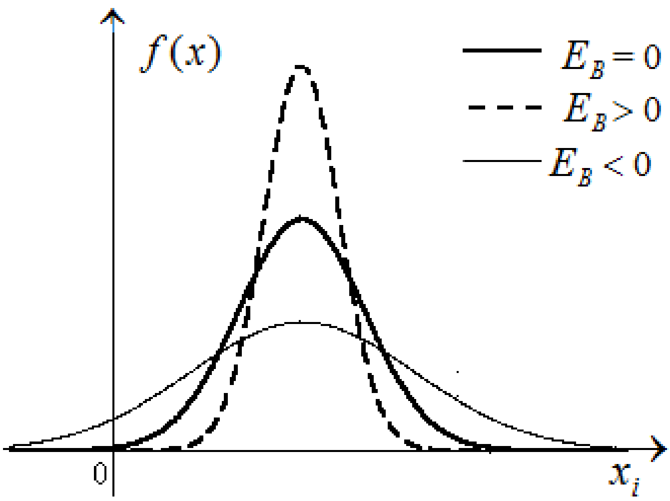


Рисунок 9 – Кривые распределения

Пример 35

Найти точечные оценки $\overline{X}_B$, $D_B(X), \sigma_B(X)$, $V_B(X)$, $A_B(X)$, $E_B(X)$ по выборке, заданной в примере 33.

Решение

Используя данные таблицы 2, вычисляем выборочное среднее по формуле

$$\overline{X}_B=\frac{1}{n}\sum_{i=1}^{k}x_i\cdot n_i\,.$$

$$\overline{X}_B=\frac{1}{100}\sum_{i=1}^{7}x_i\cdot n_i=\frac{1}{100}\cdot(5\cdot1+7\cdot8+9\cdot25+11\cdot34+13\cdot22+$$

$$+15\cdot8+17\cdot2)=\frac{1100}{100}=11.$$

Для вычисления других числовых характеристик выборки составим таблицу 3 (в первых двух столбцах приведены сгруппированные исходные данные – второй и четвертый столбцы таблицы 2 примера 33).

Контроль: при заполнении таблицы нужно помнить, что сумма элементов четвертого столбца должна быть равна нулю.

Таблица 3 – Таблица для расчета числовых характеристик выборки

Середины интервалов x_i	Частоты n_i	$x_i-\bar{x}_B$	$(x_i-\bar{x}_B)\cdot n_i$	$(x_i-\bar{x}_B)^2\cdot n_i$	$(x_i-\bar{x}_B)^3\cdot n_i$	$(x_i-\bar{x}_B)^4\cdot n_i$
5	1	– 6	– 6	36	– 216	1296
7	8	– 4	– 32	128	– 512	2048
9	25	– 2	– 50	100	– 200	400
11	34	0	0	0	0	0
13	22	2	44	88	176	352
15	8	4	32	128	512	2048
17	2	6	12	72	432	2592
$\sum$	100	–	0	552	192	8736

Выборочная дисперсия:

$$D_B(X)=\frac{1}{100}\sum_{i=1}^{8}(x_i-\overline{X}_B)^2\cdot n_i=\frac{1}{100}\cdot(5^2\cdot 1+7^2\cdot 8+9^2\cdot 25+11^2\cdot 34+$$

$$+13^2\cdot 22+15^2\cdot 8+17^2\cdot 2)=\frac{552}{100}=5{,}52.$$

Выборочное среднее квадратическое отклонение:

$$\sigma_B(X)=\sqrt{D_B(X)}=\sqrt{5{,}52}\approx 2{,}35.$$

$\sigma_B(X)$ показывает разброс выборочных значений x_i относительно выборочного среднего $\overline{X}_B=11$.

Коэффициент вариации:

$$V_B(X)=\frac{\sigma_B(X)}{\overline{X}_B}\cdot 100\%=\frac{2,35}{11}\cdot 100\%=21,4\%.$$

Выборочные коэффициент асимметрии и эксцесс:

$$A_B(X)=\frac{\mu_3^*(X)}{\sigma_B^3(X)}=\frac{\sum\limits_i (x_i-\overline{X}_B)^3\cdot n_i}{n\cdot\sigma_B^3(X)}=\frac{192}{100\cdot 2,35^3}=\frac{192}{1297,79}\approx 0,1479;$$

$$E_B(X)=\frac{\mu_4^*(X)}{\sigma_B^4(X)}-3=\frac{\sum\limits_i (x_i-\overline{X}_B)^4\cdot n_i}{n\cdot\sigma_B^4(X)}-3=\frac{8736}{100\cdot 2,35^4}-3=$$
$$=\frac{8736}{100\cdot 30,49800625}-3\approx -0,13555.$$

$A_B(X)\neq 0$ говорит о несимметричности полигона (гистограммы) относительно выборочного среднего $\overline{X}_B$.

Положительный знак выборочного коэффициента асимметрии $A_B(X)$ свидетельствует о правосторонней асимметрии данного распределения.

Отрицательность выборочного коэффициента эксцесса показывает, что полигон менее крут, чем нормальная кривая (кривая Гаусса).

14.2. Основные свойства точечных оценок: несмещенность, эффективность, состоятельность

По выборке x_1, x_2, ..., x_n из генеральной совокупности значений случайной величины X находятся выборочные числовые характеристики случайной величины – ***точечные оценки***, обозначим их θ^*; неизвестный параметр теоретического распределения обозначим θ.

Состав выборки случаен, тогда каждое возможное значение x_i можно считать частным значением случайной величины X, меняющейся от выборки к выборке. Следовательно, для каждой выборки из одной и той же генеральной совокупности получим различные оценки параметра: θ_1^*, θ_2^*, ..., θ_k^*. Таким образом, статистическая оценка θ^* является случайной величиной и поэтому определяется своим законом распределения и числовыми характеристиками $M(\theta^*)$, $D(\theta^*)$, $\sigma(\theta^*)$.

Для выбора наилучшей оценки θ_1^*, θ_2^*, ..., θ_k^* каждому θ_i^* ставится в соответствие вероятность, характеризующая степень достоверности (точности, надежности) оценки.

К основным свойствам точечных оценок относятся следующие:

- Оценку θ^* параметра θ называют ***несмещенной***, если ее математическое ожидание равно оцениваемому параметру $M(\theta^*)=\theta$.

Если $M(\theta^*)>\theta$, то оценка θ^* завышена, если $M(\theta^*)<\theta$, то оценка θ^* занижена.

- ***Эффективной*** оценкой называют такую оценку θ^*, которая имеет наименьшую дисперсию среди всех возможных несмещенных оценок, вычисленных по выборкам одинакового объема n.

Пусть θ^* и θ^{**} – две несмещенные оценки параметра θ:

$$M(\theta^*)=\theta \text{ и } M(\theta^{**})=\theta.$$

Если $D(\theta^*)<D(\theta^{**})$, тогда θ^* – эффективная оценка.

Оценка θ^* параметра θ называется ***состоятельной***, если она подчиняется закону больших чисел. При достаточно большом числе независимых наблюдений n можно утверждать с вероятностью, близкой к единице ($p>1-\alpha$, где α мало), что разность $\left|\theta^*-\theta\right|$ меньше сколь угодно малого положительного числа ε

$$P(\left|\theta^*-\theta\right|<\varepsilon)>1-\alpha.$$

(см. неравенство Чебышева).

- ***Устойчивой*** называется ***оценка***, которая не меняется от выборки к выборке

$$\theta_1^*=\theta_2^*=...=\theta_k^*.$$

На практике при оценке параметра θ не всегда удается удовлетворить одновременно требованиям несмещенности, эффективности, состоятельности и устойчивости. Рассмотрим, обладают ли этими свойствами основные точечные оценки: математическое ожидание и дисперсия.

1. **Выборочная средняя** $\overline{X}_B = \frac{1}{n}\sum_{i=1}^{n} x_i$ является несмещенной, состоятельной и эффективной оценкой математического ожидания случайной величины X, т. е. справедливы следующие утверждения:

$$M(\overline{X}_B) = M(X);$$

$$P(\left|\overline{X}_B - M(X)\right| < \varepsilon) > 1-\alpha;$$

$D(\overline{X}_B)$ – минимальна и равна $D(\overline{X}_B) = \frac{1}{n}D(X)$ (дисперсия выборочного среднего $\overline{X}_B$ в n раз меньше дисперсии самой величины X).

2. **Выборочная дисперсия**, вычисленная по n независимым наблюдениям над случайной величиной X, является смещенной оценкой дисперсии $D(X)$:

$$D_B(X) = \frac{1}{n}\sum_i (x_i - \overline{X}_B)^2 \cdot n_i,\ M(D_B) < D(X),\ \text{оценка занижена.}$$

Несмещенной оценкой генеральной дисперсии является **исправленная выборочная дисперсия**

$$S_X^2 = \frac{1}{n-1}\sum_i (x_i - \overline{X}_B)^2 \cdot n_i.$$

Выборочная дисперсия $D_B(X)$ и исправная выборочная дисперсия $S_B^2(X)$, вычисленные по n независимым наблюдениям, являются **состоятельными** оценками дисперсии $D(X)$.

Если случайная величина X распределена по *нормальному закону* и известно математическое ожидание $a = M(X)$, то несмещенной, состоятельной и эффективной оценкой дисперсии $D(X)$ является исправленная выборочная дисперсия.

14.3. Основные точечные оценки

Для удобства точечные оценки приведем в виде таблицы 4.

Таблица 4 – Основные точечные оценки параметров распределения

Название	Обозначения	Форма задания выборки	
		$x_1, x_2, \ldots, x_n$	$\begin{array}{|c|c|c|c|} \hline x_1 & x_2 & \ldots & x_k \\ \hline n_1 & n_2 & \ldots & n_k \\ \hline \end{array}$ $\sum_{i=1}^{k} n_i = n$
Выборочное среднее	$\bar{X}_B$	$\bar{X}_B = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\bar{X}_B = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$
Выборочная дисперсия	$D_B(X)$	$D_B(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X}_B)^2$	$D_B(X) = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{X}_B)^2$
Исправленная выборочная дисперсия	S_X^2	$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X}_B)^2$	$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{X}_B)^2$
Выборочное среднее квадра-тическое отклонение	$\sigma_B(X)$	$\sigma_B(X) = \sqrt{D_B(X)}$	$\sigma_B(X) = \sqrt{D_B(X)}$
Эмпирическая функция распределения	$F^*(x)$	$F^*(x) = \frac{n_x}{n}$	$F^*(x) = \frac{n_x}{n}$
Выборочный коэффициент асимметрии	A_B	$A_B = \frac{1}{nS_X^2} \cdot$ $\cdot \sum_{i=1}^{n}(x_i - \bar{X}_B)^3$	$A_B = \frac{1}{nS_X^2} \cdot$ $\cdot \sum_{i=1}^{k}(x_i - \bar{X}_B)^3 \cdot n_i$
Выборочный коэффициент эксцесса	E_B	$E_B = \frac{1}{nS_X^4} \cdot$ $\cdot \sum_{i=1}^{n}(x_i - \bar{X}_B)^4 - 3$	$E_B = \frac{1}{nS_X^4} \cdot$ $\cdot \sum_{i=1}^{k}(x_i - \bar{X}_B)^3 \cdot n_i - 3$
Выборочный коэффициент корреляции*	r_B	$r_B = \frac{1}{S_X \cdot S_Y} \cdot$ $\cdot \sum_{i=1}^{n}(x_i - \bar{X}_B)(y_i - \bar{Y}_B)$	$r_B = \frac{1}{S_X S_Y}(\sum_{i=1}^{k}(x_i y_j n_{ij} - \bar{X}_B$

* Эта точечная оценка подробнее будет рассмотрена в п. 17.

15. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ

15.1. Понятие доверительного интервала. Доверительная вероятность.

При малом числе наблюдений n, когда точечная оценка θ^* малонадежна, применяют интервальное оценивание.

- ***Ошибкой измерения*** называется разность $\theta - \theta^*$ между результатом измерения θ^* (точечной оценкой) и истинным значением измеряемой величины θ.

- ***Доверительной вероятностью (надежностью)*** оценки θ называют вероятность γ, такую, что

$$P(|\theta - \theta^*| < \delta) = \gamma$$

или

$$P(\theta^* - \delta < \theta < \theta^* + \delta) = \gamma,$$

где число $\delta > 0$ называется ***точностью оценки***.

- ***Доверительным*** называется ***интервал***

$$I_\gamma = (\theta^* - \delta, \theta^* + \delta),$$

который покрывает неизвестное значение параметра θ с заданной доверительной вероятностью γ.

Доверительная вероятность γ обычно задается равной 0,95; 0,99; 0,999.

- Вероятность $\alpha = 1 - \gamma$ называется ***уровнем значимости*** ($\alpha = 0{,}05;\ 0{,}01;\ 0{,}001$).

Статистический смысл доверительной вероятности $\gamma = 0{,}95$ ($\alpha = 0{,}05$) состоит в том, что если рассмотрено большое количество выборок, то для 95 % из них доверительный интервал покрывает оцениваемый параметр θ и лишь 5 % интервалов его не содержат.

Итак, доверительный интервал – это случайный интервал, в пределах которого с вероятностью γ находится оцениваемый параметр, поэтому его называют ***интервальной оценкой*** неизвестного параметра. Величина доверительного интервала существенно зависит от объема выборки n (уменьшается с ростом n) и значения доверительной вероятности γ (увеличивается с приближением γ к единице).

15.2. Распределение Стъюдента

Пусть случайная величина X подчиняется *нормальному закону распределения* $N(a, \sigma)$, с математическим ожиданием a и дисперсией σ^2, причем каждое из независимых наблюдений $X_1, X_2, \ldots, X_n$ имеет те же числовые характеристики, что и X:

$$M(X_i) = a;\ D(X_i) = \sigma^2,\ (i = 1, 2, \ldots, n),$$

тогда выборочная средняя $\bar{X}_B = \frac{1}{n}\sum_i X_i$ также подчиняется нормальному закону распределения $N(a, \frac{\sigma^2}{n})$ с параметрами:

$$M(\bar{X}_B) = M(X) = a\ ,\ D(\bar{X}_B) = \frac{\sigma^2}{n},\ \sigma(\bar{X}_B) = \frac{\sigma}{\sqrt{n}}.$$

- Случайная величина $Z = \dfrac{\bar{X}_B - a}{\sigma(\bar{X}_B)}$ называется ***нормированным отклонением средней арифметической***.

Случайная величина Z распределена по нормальному закону $N(0;1)$, ее функция плотности вероятности имеет вид

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\quad,$$

значения этой функции приведены в Приложении А, а значения функции распределения – интегральной функции Лапласа

$$\Phi_0(z) = \frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{z^2}{2}}dz$$

приведены в Приложении В.

В практических задачах дисперсия генеральной совокупности $D(X)$ почти всегда оказывается неизвестной, поэтому параметр σ заменяют его точечной оценкой – исправленным средним квадратическим отклонением S_X, где

$$S_X^2 = \frac{1}{n-1}\sum_i (x_i - \overline{X}_B)^2 \cdot n_i$$

несмещенная, состоятельная оценка дисперсии.

- Случайную величину $t_\nu = \dfrac{\overline{X}_B - a}{S_X / \sqrt{n}}$ называют ***распределением Стьюдента***, или t_ν – ***распределением с числом степеней свободы*** $\nu = n-1$.

Случайная величина t_ν применяется при решении разнообразных вопросов статистического анализа при малых объемах выборок. При $n > 50$ распределение Стъюдента мало отличается от нормального.

При $n \to \infty$ случайная величина t_ν имеет распределение, близкое к распределению случайной величины Z, т. е. к нормальному закону $N(0;1)$ с параметрами $a = 0$, $\sigma = 1$ (рис. 10).

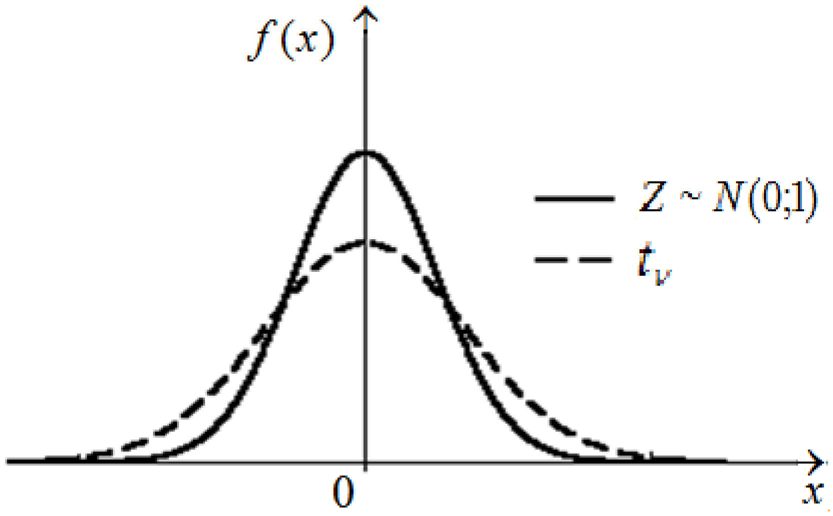


Рисунок 10 – Распределение случайной величины t_ν при $n \to \infty$

15.3. Доверительный интервал для математического ожидания нормально распределенной случайной величины (σ^2 – известна)

Пусть случайная величина X распределена нормально $N(a;\sigma)$, дисперсия генеральной совокупности σ^2 **известна**.

Построим доверительный интервал для неизвестного математического ожидания $M(X) = a$.

Наилучшей несмещенной, состоятельной и эффективной оценкой в этом случае является выборочная средняя

$$\overline{X}_B = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Из теории вероятностей известно, что вероятность попадания нормально распределенной случайной величины X в интервал симметричный математическому ожиданию a

$$P(|X - a| < \delta) = 2\Phi_0\left(\frac{\delta}{\sigma}\right).$$

Оценкой для a является $\overline{X}_B$, известно, что

$$D(\overline{X}_B) = \frac{\sigma^2}{n}, \text{ тогда } \sigma(\overline{X}_B) = \frac{\sigma}{\sqrt{n}}.$$

Следовательно,

$$P\left(|\overline{X}_B - a| < \delta\right) = 2\Phi_0\left(\frac{\delta\sqrt{n}}{\sigma}\right) = 2\Phi_0(t),$$

где $\dfrac{\delta\sqrt{n}}{\sigma} = t$ или $\delta = \dfrac{t\sigma}{\sqrt{n}}$,

тогда

$$P(\overline{X}_B - \frac{t\sigma}{\sqrt{n}} < a < \overline{X}_B + \frac{t\sigma}{\sqrt{n}}) = 2\Phi_0(t) = \gamma .$$

Итак, доверительный интервал

$$I_\gamma = \left(\overline{X}_B - \frac{t\sigma}{\sqrt{n}}, \overline{X}_B + \frac{t\sigma}{\sqrt{n}}\right)$$

покрывает неизвестное значение параметра a с заданной доверительной вероятностью γ. Точность оценки $\delta = \dfrac{t\sigma}{\sqrt{n}}$, где t находится из равенства

$$2\Phi_0(t)=\gamma.$$

Значения функции $\Phi_0(t)$ приведены в Приложении В.

Пример 36

Случайная величина X распределена нормально $N(a;9)$. По заданной доверительной вероятности $\gamma=0{,}95$ найти доверительный интервал для оценки математического ожидания a генеральной совокупности X, если объем выборки равен 36 и выборочное среднее равно 4.

Решение

Построим доверительный интервал для неизвестного математического ожидания $M(X)=a$.

Решение

По условию задачи дисперсия генеральной совокупности $\sigma^2=9$, $n=36$, $\gamma=0{,}95$, $\overline{X}_B=4$.

Так как $\gamma=2\Phi_0(t)=0{,}95$, то по таблице значений функции $\Phi_0(t)$ (Приложение В) найдем значение аргумента $t=1{,}96$.

Точность оценки $\delta=\dfrac{t\sigma}{\sqrt{n}}=\dfrac{1{,}96\cdot 3}{6}=0{,}98.$

Доверительный интервал для оценки математического ожидания

$$I_\gamma=(\overline{X}_B-\frac{t\sigma(X)}{\sqrt{n}},\overline{X}_B+\frac{t\sigma(X)}{\sqrt{n}})=(4-0{,}98;4+0{,}98)=(3{,}02;\,4{,}98).$$

15.4. Доверительный интервал для математического ожидания нормально распределенной случайной величины (σ^2 – неизвестна)

Пусть случайная величина X распределена нормально, дисперсия генеральной совокупности σ^2 **неизвестна**. Построим доверительный интервал для $M(X)=a$ – неизвестного математического ожидания.

По заданной выборке найдем точечные оценки:

$$\overline{X}_B=\frac{1}{n}\sum_i x_i \text{ и } S_X^2=\frac{1}{n-1}\sum_i (x_i-\overline{X}_B)^2\cdot n_i.$$

Построим распределение Стьюдента

$$t_\nu = \frac{\overline{X}_B - a}{S_X / \sqrt{n}}$$

с *числом степеней свободы* $\nu = n - 1$.

Рассмотрим неравенство $\dfrac{\left|\overline{X}_B - a\right|}{S_X / \sqrt{n}} \le t_\nu$, выполним преобразования

$$-\frac{t_\nu \cdot S_X}{\sqrt{n}} < \overline{X}_B - a < \frac{t_\nu \cdot S_X}{\sqrt{n}}.$$

Итак, получили доверительный интервал для математического ожидания a нормально распределенной случайной величины X при неизвестном среднем квадратическом отклонении

$$\overline{X}_B - \frac{t_\nu \cdot S_X}{\sqrt{n}} < a < \overline{X}_B + \frac{t_\nu \cdot S_X}{\sqrt{n}}.$$

Числовые значения статистики t_ν зависят не только от доверительной вероятности $\gamma = 1 - \alpha$, но и от объема выборки n. Как отмечалось выше, при $n \to \infty$ распределение Стьюдента приближается к нормальному.

Пример 37

Распределение длины колоса ячменя задается таблицей:

Длина колоса ячменя (см) x_i	7,5	8,5	9,5	10,5	11,5	12,5	13,5
Число колосьев n_i	4	10	14	12	5	4	1

Найти точечную и интервальную оценки математического ожидания a случайной величины X – длина колоса ячменя с надежностью $\gamma = 0{,}95$.

Решение

Объем выборки в данном случае:

$$n = \sum_i n_i = 4 + 10 + 14 + 12 + 5 + 4 + 1 = 50.$$

Полагая, что изменчивость колоса ячменя описывается нормальным законом, найдем точечную оценку математического ожидания случайной величины X.

$$\overline{X}_B = \frac{1}{n} \cdot \sum_i x_i n_i =$$

$$= \frac{1}{50}(7{,}5 \cdot 4 + 8{,}5 \cdot 10 + 9{,}5 \cdot 14 + 10{,}5 \cdot 12 + 11{,}5 \cdot 5 + 12{,}5 \cdot 4 + 13{,}5 \cdot 1) =$$

$$= \frac{495}{50} = 9{,}9.$$

Значение $\overline{X}_B$ показывает, что в среднем длина колоса у 50 колосьев равна 9,9 см.

Исправленная выборочная дисперсия:

$$S_X^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_i x_i^2 \cdot n_i - \overline{X}_B^2\right) =$$

$$= \frac{50}{49}(\frac{1}{50}(7{,}5^2 \cdot 4 + 8{,}5^2 \cdot 10 + 9{,}5^2 \cdot 14 + 10{,}5^2 \cdot 12 + 11{,}5^2 \cdot 5 + 12{,}5^2 \cdot 4 +$$

$$+ 13{,}5^2 \cdot 1) - 9{,}9^2) = 2{,}08.$$

Несмещенной оценкой среднего квадратического отклонения является исправленное выборочное среднее квадратическое отклонение $S_X = \sqrt{S_X^2} = \sqrt{2{,}08} = 1{,}44.$

Доверительный интервал для математического ожидания при неизвестном среднем квадратическом отклонении найдем по формуле

$$\overline{X}_B - \frac{t_\nu \cdot S_X}{\sqrt{n}} < a < \overline{X}_B + \frac{t_\nu \cdot S_X}{\sqrt{n}},$$

где t_ν найдем из таблицы Стьюдента (приложение С) для уровня значимости $\alpha = 1 - 0{,}95 = 0{,}05$ и числа степеней свободы $k = n - 1 = 50 - 1 = 49$, итак $t_{49;0,05} = 2{,}26$. По условию задачи

$$\overline{X}_B = 9{,}9; \quad t_{49;\,0,05} = 2{,}26; \quad S_X = 1{,}44; \quad n = 50.$$

Итак, искомый доверительный интервал для математического ожидания

$$9{,}5 < a < 10{,}3.$$

Следовательно, если будет произведено достаточно большое число выборок (данных о длине колоса ячменя) одного и того же объема $n = 50$, то в 95 % выборок среднее значение длины колоса будет находиться в пределах интервала (9,5; 10,3).

15.5. Распределение χ^2 Пирсона, доверительного интервала для дисперсии

Величина $\chi^2 = \dfrac{nS_X^2}{\sigma^2}$, называется ***случайной величиной с распределением*** $\boldsymbol{\chi^2}$ (хи-квадрат) и $\nu = n-1$ степенями свободы.

Составим соотношение

$$P\left(\chi_1^2 < \frac{n\cdot S_X^2}{\sigma^2} < \chi_2^2\right) = 1-\alpha,$$

которое показывает, что случайная величина $\chi^2 = \dfrac{nS_X^2}{\sigma^2}$ покрывается доверительным интервалом $I_\gamma = \left(\chi_1^2\ ,\chi_2^2\right)$ с заданной вероятностью $\gamma = 1-\alpha$.

Значение χ_1^2 находят из таблицы (приложение D) для $\nu = n-1$ степеней свободы и уровня значимости $1-\alpha/2$.

Значение χ_2^2 находят из таблицы (приложение D) для $\nu = n-1$ степеней свободы и уровня значимости $\alpha/2$.

График плотности распределения χ^2 представлен на рисунке 11.

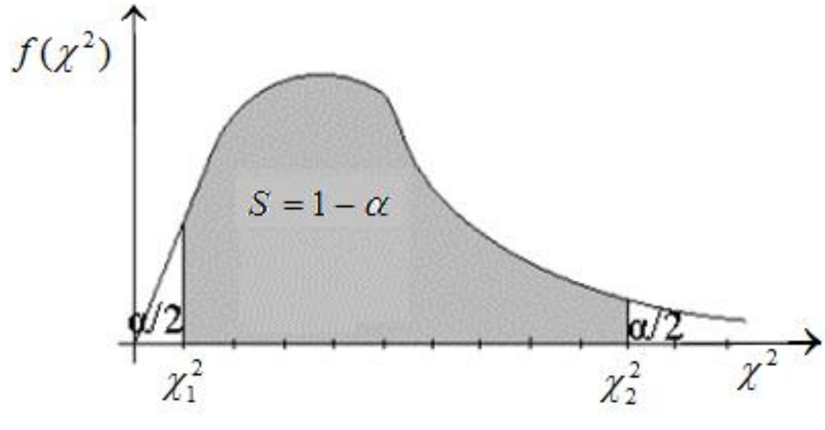


Рисунок 11 – График плотности распределения $\square^2$

Заштрихованная площадь криволинейной трапеции под графиком плотности $f(\chi^2)$ равна вероятности $\gamma=1-\alpha$ попадания случайной величины χ^2 в интервал $I_\gamma=\left(\chi_1^2\ ,\chi_2^2\right)$.

Из двойного неравенства

$$\left(\chi_1^2\ <\frac{n\cdot S_X^2}{\sigma^2}<\chi_2^2\right)$$

найдем доверительный интервал для дисперсии σ^2 нормально распределенной случайной величины X:

$$\frac{nS_X^2}{\chi_2^2}<\sigma^2<\frac{nS_X^2}{\chi_1^2},$$

Пример 38

Случайная величина X распределена по нормальному закону. По выборке объема $n=20$ найдена исправленная выборочная дисперсия $S_X^2\ =10$. Найти доверительный интервал (интервальную оценку) для дисперсии σ^2 с доверительной вероятностью $\gamma=0{,}9$.

Решение

Найдем интервальную оценку дисперсии σ^2. Так как $\alpha=1-\gamma=1-0{,}9=0{,}1$, то и 1 – $\frac{\alpha}{2}$ = 0,95. По таблице (Приложение D) находим χ_1^2 для вероятности $P=1-\frac{\alpha}{2}=0{,}95$ и $k=n-1=19$ степеней свободы, получим $\chi_1^2=10{,}1$; аналогично $\chi_2^2=30{,}1$ для $P=\frac{\alpha}{2}=0{,}05$ и $k=n-1=19$. Тогда доверительный интервал для дисперсии σ^2 $\frac{20\cdot 10}{30{,}1}<\sigma^2<\frac{20\cdot 10}{10{,}1}$ или $6{,}645<\sigma^2<19{,}802$.

Следовательно, для 90 % выборок объема $n=20$ дисперсия покрывается интервалом (6,645; 19,802).

16. СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ

16.1. Понятие статистической гипотезы, постановка задачи проверки гипотез

Под ***статистической гипотезой*** понимают всякое высказывание о генеральной совокупности, проверяемое по выборке.

Например:

- гипотезы о виде неизвестного закона распределения генеральной совокупности (***непараметрические гипотезы)***;
- гипотезы о равенстве числовых значений параметров распределения (***параметрические гипотезы)***.

Гипотеза называется ***основной*** (***нулевой***), если для закона распределения случайной величины X с одним параметром $f(x,\theta)$ необходимо проверить гипотезу, что неизвестный параметр θ равен заданной величине a, обозначается

$$H_0:\theta = a.$$

Гипотеза $H_1:\theta \neq a$ называется ***конкурирующей*** (***альтернативной***).

Также альтернативными гипотезами являются следующие:

$$H_1:\theta > a;\quad H_1:\theta < a.$$

При проверке гипотез возникают следующие ошибки:

- ***Ошибка первого рода*** состоит в том, что основная гипотеза H_0 отвергается, в то время как она верна.

Вероятность ошибки первого рода обозначают α (доля ложных заключений) и называют ***уровнем значимости***. Чем меньше α, тем меньше вероятность отклонить по результатам выборки проверяемую гипотезу, если она верна; но увеличивается риск принять гипотезу, когда она неверна.

- ***Ошибка второго рода*** состоит в том, что H_0 принимается, но в действительности она неверна.

Вероятность ошибки второго рода условились обозначать β, ее величина зависит от H_1.

В задачах контроля качества продукции вероятность α представляет «риск поставщика», связанный с забраковкой (по отрицательному результату выборочного контроля) всей партии, а вероятность β – «риск потребителя», связанный с принятием (по положительному анализу выборки) партии, не удовлетворяющей стандарту.

Статистическим критерием K называют правило, по которому определяется мера расхождения результатов выборочного наблюдения с выдвинутой гипотезой H_0.

Вероятность $(1-\beta)$ принятия проверяемой гипотезы H_0 называется ***мощностью критерия***.

При определении меры расхождения результатов выборочного наблюдения с выдвинутой гипотезой H_0 используют следующие случайные величины, которые называют ***статистиками***:

- $\mathrm{Z}=\dfrac{\overline{X}_B-a}{\sigma_B(X)/\sqrt{n}}$ – случайная величина, распределенная по

нормальному закону $N(0;1)$, с параметрами $a=0$ и $\sigma=1$ (нормированное нормальное распределение). Значения ее функции распределения (функция Лапласа) приведены в таблице (Приложение В).

- $t_\nu=\dfrac{\overline{X_B-a}}{S_X/\sqrt{n}}$ – случайная величина, распределенная по закону

Стьюдента с $\nu=n-1$ степенями свободы (Приложение С).

- $\chi^2=\dfrac{nS_X^2}{\sigma^2}$, – случайная величина «хи-квадрат», распределение Пирсона (Приложение D).
- $F=\dfrac{S_1^2}{S_2^2}$, где S_1^2 и S_2^2 – исправленные выборочные дисперсии

двух выборок.

В таблице (Приложение Е) приведены значения F критерия Фишера.

Для критерия K введены понятия критической области и области допустимых значений.

Критической областью W критерия называется совокупность таких значений случайной величины K, при которых основная гипотеза H_0 отклоняется.

Областью принятия гипотезы H_0 (областью допустимых значений) называют совокупность таких значений случайной величины K, при которых гипотеза H_0 не отклоняется.

Точки, отделяющие критическую область W от области принятия гипотезы, называют ***критическими точками*** $k_{кр}$.

Критическая область, определяемая неравенством $K > k_{кр}$, называется ***правосторонней***.

Левосторонней называют критическую область, определяемую неравенством $K < k_{кр}$.

Двусторонней называют критическую область, определяемую неравенствами $K < k_1$, $K > k_2$ $(k_1 < k_2)$. В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется неравенством: $|K| > k_{кр}$ $(k_{кр} > 0)$.

Критическую область W следует выбирать так, чтобы вероятность попадания в нее критерия K была минимальной и равной заданной величине α, если верна нулевая гипотеза H_0, и максимальной в противном случае.

Другими словами, критическая область должна быть такой, чтобы при заданном уровне значимости α мощность критерия $(1-\beta)$ была максимальной.

Вид критической области зависит от вида альтернативной гипотезы H_1. Например, если для нулевой гипотезы $H_0 : \theta = a$ рассматривают альтернативную гипотезу $H_1 : \theta \neq a$, то критическая область будет двусторонней симметричной, для которой

$$P(K < k_1) + P(K > k_2) = \alpha/2 + \alpha/2 = \alpha.$$

Тогда вероятность правильных решений $P(k_1 < K < k_2) = 1-\alpha$.

Для альтернативной гипотезы $H_1 : \theta \neq a$ правосторонняя критическая область определяется из уравнения $P(K > k_{кр}) = \alpha$, а для альтернативной гипотезы $H_1 : \theta < a$ левосторонняя критическая область определяется из уравнения $P(K < k_{кр}) = \alpha$.

16.2. Проверка гипотез о законе распределения. Критерий согласия χ^2 Пирсона

Критерием согласия называется критерий проверки гипотезы H_0 о предполагаемом законе неизвестного распределения.

Пусть X – исследуемая случайная величина. Требуется проверить гипотезу H_0 о том, что X подчиняется нормальному закону распределения:

$$H_0 : X \in N(a, \sigma^2).$$

Для проверки данной гипотезы используют различные критерии согласия: Пирсона (χ^2), Колмогорова, Смирнова и др.

Рассмотрим критерий Пирсона (χ^2).

Пусть задана выборка объема n:

Выборочные значения x_i	x_1	x_2	...	x_k
Эмпирические частоты n_i	n_1	n_2	...	n_k

Алгоритм критерия χ^2 Пирсона.

1. Разбить выборку на k интервалов $\Delta_1, \Delta_2, ..., \Delta_k$ с шагом $h = const.$

2. Вычислить теоретические вероятности p_i по формуле

$$p_i = P(x_1 < X < x_{i+1}) = \Phi_0\left(\frac{x_{i+1} - \overline{X}_B}{S_X}\right) - \Phi_0\left(\frac{x_i - \overline{X}_B}{S_X}\right).$$

Теоретическое число (теоретическая частота) значений случайной величины X, попавших в интервал Δ_i, равно $n \cdot p_i$. При этом сумма теоретических частот всех значений случайной величины X, попавших в соответствующий интервал Δ_i, равна объему выборки: $\sum_{\mathrm{i}=1}^{k} n_i \cdot p_i = n,$ а сумма теоретических вероятностей $\sum_{\mathrm{i}=1}^{k} p_{\mathrm{i}} = 1.$

3. Подсчитать эмпирические частоты n_i – количество наблюдаемых значений случайной величины, попавших в каждый из интервалов выборки, при этом сумма этих частот равна объему выборки: $\sum_{\mathrm{i}=1}^{k} n_i = n.$

4. Вычислить выборочное значение величины

$$\chi_B^2 = \sum_{i-1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i},$$

имеющей закон распределения Пирсона или χ^2 – «хи-квадрат» с $\nu = k - r - 1$ степенями свободы, где r – число параметров распределения; k – число интервалов выборки.

5. Сравнить выборочное значение χ_B^2 с критическим значением $\chi_{кр}^2$, найденным по таблице распределения χ^2 (Приложение D), для уровня значимости α (обычно $\alpha = 0{,}05$) и для $\nu = k - r - 1$ степеней свободы.

Если $\chi_B^2 \le \chi_{кр}^2$, то считается, что гипотеза H_0 не противоречит эмпирическим данным, а различие между эмпирическими и теоретическими частотами незначительно: $n \cdot p_i \approx n_i$.

Если $\chi_B^2 > \chi_{кр}^2$, то говорят, что гипотеза H_0 не согласуется с выборочными данными, и она отвергается.

Замечание 1. Статистика χ^2 имеет хи-квадрат распределение лишь при достаточно большом объеме выборки n, поэтому необходимо проверить, чтобы в каждом интервале Δ_i было, по меньшей мере, 5–7 наблюдений. Если количество наблюдений в отдельном интервале мало, то соседние интервалы объединяют.

Замечание 2. Для предварительной проверки гипотезы о нормальном распределении генеральной совокупности можно использовать следующие оценки, основанные на свойствах нормального распределения:

✓ Практически все отклонения $|x_i - \overline{X}_B|$ выборочных значений x_i от среднего $\overline{X}_B$ должны быть меньше утроенного выборочного исправленного среднего квадратического отклонения S_X (с вероятностью 0,99):

$$|x_i - \overline{X}_B| < 3S_X \text{ (правило трех «сигм»).}$$

✓ Примерно $2/3$ отклонений (68,3 %) выборочных значений x_i от $\overline{X}_B$ должны удовлетворять неравенству $|x_i - \overline{X}_B| < S_X$.

✓ Выборочные коэффициенты асимметрии и эксцесса должны быть близки к нулю: $A_B \approx 0, \quad E_B \approx 0.$

Пример 39

По выборке, представленной в примере 33, требуется проверить, что случайная величина X подчиняется нормальному закону распределения при уровне значимости $\alpha = 0{,}05.$

Решение

Гипотезу о том, что генеральная совокупность, из которой извлечена выборка, распределена по нормальному закону, обозначим

$$H_0 : X \in N(a, \sigma).$$

Проверим справедливость выдвинутой гипотезы с помощью критерия согласия χ^2 Пирсона.

В качестве критерия проверки нулевой гипотезы принимается случайная величина

$$\chi_B^2 = \sum_{i-1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}.$$

Вероятности p_i рассчитаем с помощью нормированной функции Лапласа $\Phi_0(x)$ (Приложение В):

$$p_i = P(x_1 < X < x_{i+1}) = \Phi_0\left(\frac{x_{i+1} - \overline{X}_B}{\sigma_B(X)}\right) - \Phi_0\left(\frac{x_i - \overline{X}_B}{\sigma_B(X)}\right).$$

По заданной выборке были найдены точечные оценки (Пример 34): выборочное среднее $\overline{X}_B = 11,$ выборочное исправленное среднее квадратическое отклонение $\sigma_B(X) = 2{,}35.$

Тогда $p_1 = P(-\infty < x \le 6) = \Phi_0\left(\frac{6-11}{2{,}35}\right) - \Phi_0(-\infty) =$

$$= \Phi_0(-2{,}13) - \Phi_0(-\infty) = -\Phi_0(2{,}13) + \Phi_0(+\infty) = -0{,}4834 + 0{,}5 = 0{,}0166$$

$$p_2 = P(6 < x \le 8) = \Phi_0\left(\frac{8-11}{2{,}35}\right) - \Phi_0\left(\frac{6-11}{2{,}35}\right) =$$

$$= \Phi_0(-1{,}28) + \Phi_0(-2{,}13) = -\Phi_0(1{,}28) + \Phi_0(2{,}13) = -0{,}3997 + 0{,}4834 = 0{,}0837;$$

$$p_3 = P(8 < x \le 10) = \Phi_0\left(\frac{10-11}{2{,}35}\right) - \Phi_0\left(\frac{8-11}{2{,}35}\right) =$$

$$=\Phi_0(-0{,}43)+\Phi_0(-1{,}28)=-\Phi_0(0{,}43)+\Phi_0(1{,}28)=-0{,}1664+0{,}3997=0{,}2333;$$

$$p_4=P(10<x\le 12)=\Phi_0\left(\frac{12-11}{2{,}35}\right)-\Phi_0\left(\frac{10-11}{2{,}35}\right)=$$

$$=\Phi_0(0{,}43)-\Phi_0(-0{,}43)=\Phi_0(0{,}43)+\Phi_0(0{,}43)=0{,}1664+0{,}1664=0{,}3328;$$

$$p_5=P(12<x\le 14)=\Phi_0\left(\frac{14-11}{2{,}35}\right)-\Phi_0\left(\frac{12-11}{2{,}35}\right)=$$

$$=\Phi_0(1{,}28)-\Phi_0(0{,}43)=0{,}3997-0{,}1664=0{,}2333;$$

$$p_6=P(14<x\le 16)=\Phi_0\left(\frac{16-11}{2{,}35}\right)-\Phi_0\left(\frac{14-11}{2{,}35}\right)=$$

$$=\Phi_0(2{,}13)-\Phi_0(1{,}28)=0{,}4834-0{,}3997=0{,}0837;$$

$$p_7=P(16<x<+\infty)=\Phi_0(+\infty)-\Phi_0\left(\frac{18-11}{2{,}35}\right)=$$

$$=\Phi_0(+\infty)-\Phi_0(1{,}45)=0{,}5-0{,}4834=0{,}0166.$$

Полученные результаты запишем в таблицу 5.

Таблица 5 – Расчетная таблица для вычисления χ^2_B

Интервалы $(x_i;\, x_{i+1}]$	Частоты эмпирические n_i	Вероятности p_i	Теоретические частоты $n\cdot p_i$	$\frac{(n_i-n\cdot p_i)^2}{n\cdot p_i}$
(-∞; 6]	1	0,0166	1,66	0,2624
(6; 8]	8	0,0838	8,38	0,0164
(8; 10]	25	0,2333	23,33	0,1195
(10; 12]	34	0,3328	33,28	0,0156
(12; 14]	22	0,2333	23,33	0,0758
(14; 16]	8	0,0838	8,38	0,0164
(16; + ∞)	2	0,0166	1,66	0,0696
$\sum$	100	1.00	100	$\chi^2_B=5{,}5757$

Значение $\chi^2_B=5{,}5757$.

Количество интервалов $k=7$, нормальное распределение имеет два параметра (математическое ожидание a и среднее квадратическое отклонение σ), поэтому $r=2$, тогда число степеней свободы равно $\nu=k-r-1=7-2-1=4$.

В таблице критических точек распределения χ^2_B (Приложение D) по уровню значимости $\alpha=0{,}05$ и числу степеней свободы $\nu=4$, найдем критическое значение $\chi^2_{кр} = (0{,}05; 4) = 9{,}488$, очевидно, что $5{,}5755 < 9{,}488$.

Так как $\chi^2_B \le \chi^2_{кр}$, то нет оснований для отклонения нулевой гипотезы при заданном уровне значимости $\alpha = 0{,}05$.

17. ОСНОВЫ ТЕОРИИ КОРРЕЛЯЦИИ

17.1. Понятие корреляционной зависимости. Корреляционная таблица

Зависимость между случайными величинами, состоящая в том, что каждому значению одной величины X ставится в соответствие какая-либо числовая характеристика соответствующего распределения другой величины Y, называется ***статистической корреляцией***, или просто ***корреляционной зависимостью***.

Таблица 6, в которой результаты $(x_i; y_j)$ испытания записаны в порядке возрастания x_i и y_j с указанием частот n_{ij}, называется ***корреляционной таблицей***.

Таблица 6 – Корреляционная таблица

$Y \backslash X$	x_1	x_2	…	x_k	n_{y_j}
y_1	n_{11}	n_{12}	…	n_{1k}	$\sum n_{1j} = n_{y_1}$
y_2	n_{21}	n_{22}	…	n_{2k}	$\sum n_{2j} = n_{y_2}$
…	…	…	…	…	…
y_m	n_{m1}	n_{m2}	…	n_{mk}	$\sum n_{mj} = n_{y_m}$
n_{x_i}	$\sum n_{i1} = n_{x_1}$	$\sum n_{i2} = n_{x_2}$	…	$\sum n_{ik} = n_{x_k}$	$\sum n_{xy} = n$

В таблице 6 числа n_{ij} – частоты, показывающие, сколько раз повторяется пара $(x_i; y_j)$.

Задача корреляционного анализа состоит в том, чтобы с помощью статистических методов получить надежные заключения о связи между изучаемыми случайными величинами. Зависимость между случайными величинами может быть ***линейной*** или ***нелинейной***; ***тесноту связи*** характеризуется выборочным коэффициентом корреляции.

17.2. Выборочный коэффициент корреляции, его свойства

Пусть X и Y – случайные величины.

Корреляционный момент, коэффициент корреляции для случайных величин были введены в разделе 15.3.

Рассмотрим точечные оценки этих числовых характеристик случайных величин, вычисленных по выборке.

- ***Выборочным корреляционным моментом*** *случайных величин* X и Y называется число

$$K_B = \frac{1}{n}\sum_i (x_i - \overline{X}_B)(y_i - \overline{Y}_B) \cdot \text{или } K_B = \frac{\sum_i x_i y_i}{n} - \overline{X}_B \overline{Y}_B \quad .$$

- ***Выборочным коэффициентом корреляции*** *случайных величин* X и Y называется число

$$r_B = \frac{K_B}{S_X S_Y} = \frac{\sum_i (x_i - \overline{X}_B)(y_i - \overline{Y}_B)}{nS_X S_Y}$$

или

$$r_B = \frac{\sum_i (x_i - \overline{X}_B)(y_i - \overline{Y}_B)}{\sqrt{\sum_i (x_i - \overline{X}_B)^2} \cdot \sqrt{\sum_i (y_i - \overline{Y}_B)^2}}.$$

Если результаты $(x_i; y_j)$ испытания записаны в корреляционной таблице, то выборочный коэффициент корреляции удобно вычислять по следующей формуле:

$$r_B = \frac{K_B}{S_X S_Y} = \frac{\sum\limits_i \sum\limits_j x_i y_j n_{ij} - \frac{\sum\limits_i x_i n_{x_i} \cdot \sum\limits_j y_j n_{y_j}}{n}}{n S_X S_Y}.$$

Свойства выборочного коэффициента корреляции

1. Коэффициент корреляции принимает значения на отрезке $[-1; 1]$:

$$|r_B| \le 1.$$

2. В случае ***линейной*** функциональной зависимости между X и Y коэффициент корреляции равен по модулю единице:

$$r_B = \pm 1.$$

3. Если $r_B > 0$, то связь между X и Y положительна и случайные величины одновременно возрастают (или одновременно убывают). Если $r_B < 0$, то при увеличении X убывает Y.

4. Если X и Y независимы, то их коэффициент корреляции равен нулю: $r_B = 0$, $(K_B = 0.)$

Обратное, вообще говоря, неверно. Если $r_B = 0{,}5$, то X и Y либо не коррелированны, либо они зависимы, но связь ***нелинейная***.

При наличии нелинейной связи для описания корреляции используют другие показатели.

Графическая интерпретация коэффициента корреляции между признаками X и Y приведена на рисунке 12.

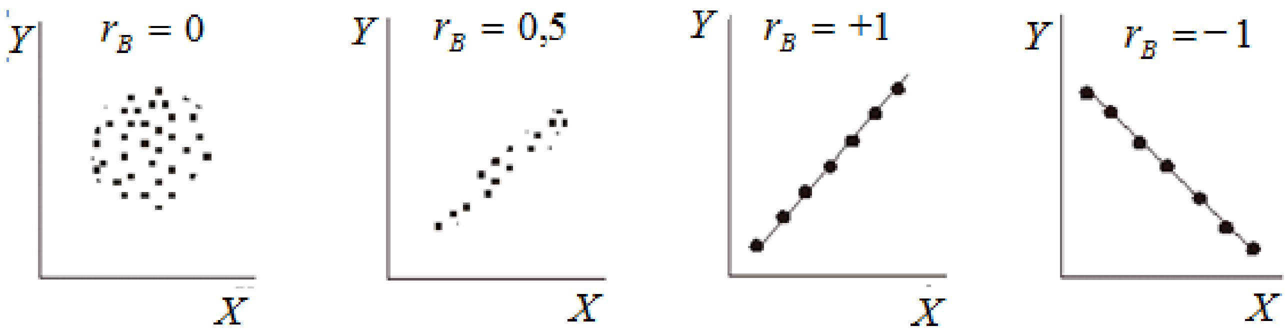


Рисунок 12 – Интерпретация коэффициента корреляции между признаками X и Y

Пример 40

По заданным выборкам двух случайных величин X – количество осадков (мм) за июнь-август и Y – урожайность картофеля (ц/га) построить корреляционную таблицу, найти коэффициент корреляции.

№	X	Y	№	X	Y	№	X	Y	№	X	Y	№	X	Y
1	113	93	**11**	129	72	**21**	112	55	**31**	146	181	**41**	127	68
2	115	95	**12**	171	94	**22**	165	119	**32**	80	36	**42**	149	88
3	117	96	**13**	170	110	**23**	152	78	**33**	112	61	**43**	145	86
4	113	94	**14**	153	92	**24**	138	77	**34**	119	54	**44**	140	84
5	120	98	**15**	181	119	**25**	80	34	**35**	160	93	**45**	144	84
6	128	103	**16**	188	120	**26**	131	68	**36**	129	67	**46**	155	95
7	117	79	**17**	123	66	**27**	151	92	**37**	104	44	**47**	161	102
8	120	80	**18**	113	66	**28**	137	174	**38**	159	93	**48**	155	95
9	131	87	**19**	212	153	**29**	86	38	**39**	136	72	**49**	137	76
10	136	78	**20**	116	48	**30**	120	68	**40**	116	67	**50**	164	100

Решение

Составим корреляционную таблицу. Для этого сгруппируем все значения выборки случайной величины X в следующие интервалы с шагом $h = 20$ (подобно как это показано в примере 33):

$$(70;90],\ (90;110],\ (110;130],\ (130;150],$$
$$(150;170],\ (170;190],\ (190;210],\ (210;230].$$

Аналогично, для случайной величины Y выберем следующие интервалы ($h = 22$):

$$(23;45],\ (45;67],\ (67;89],\ (89;111],$$
$$(111;133],\ (133;155],\ (155;177],\ (177;199].$$

Подсчитываем количество пар исходной выборки $(x_i; y_j)$, попадающих в прямоугольники, образованные границами интервалов, результаты запишем в таблицу 7.

В корреляционной таблице 8 вместо интервалов для случайной величины X и случайной величины Y записываем середины интервалов и соответствующие частоты n_x, n_y из таблицы 7.

Таблица 7 – Таблица для частот n_{xy} пар значений $(x_i; y_j)$

Интер-валы для Y	Интервалы для X							
	(70; 90]	90; 110]	(110; 130]	(130; 150]	(150; 170]	(170; 190]	(190; 210]	(210; 230]
(23; 45]								
(45; 67]								
(67; 89]								
(89; 111]								
(111; 133]								
(133; 155]								
(155; 177]								
(177; 199]								

Таблица 8 – Корреляционная таблица эмпирического распределения двумерной случайной величины (X, Y)

Y \ X	80	100	120	140	160	180	200	220	n_y
34	**3**	**1**							4
56			**8**						8
78			**5**	**10**	**1**				16
100			**6**		**9**	**1**			16
122					**1**	**2**			3
144								**1**	1
166				**1**					1
188				**1**					1
n_x	3	1	19	12	11	3	0	1	50

Если вычислить точечные оценки, то получится следующее:

$$\overline{X}_B = 136{,}4; \quad S_X = 26{,}71.$$

Аналогично можно подсчитать, что

$$\overline{Y}_B = 85{,}92; \quad S_Y = 30{,}72.$$

Далее вычислим выборочный корреляционный момент:

$$K_B = \frac{\sum_i x_i y_i n_{ij}}{n} - \overline{X}_B \cdot \overline{Y}_B.$$

$$K_B = \frac{1}{50}(80 \cdot 34 \cdot 3 + 100 \cdot 34 \cdot 1 + 120 \cdot 56 \cdot 8 + 120 \cdot 78 \cdot 5 +$$

$$+120 \cdot 100 \cdot 6 + 140 \cdot 78 \cdot 10 + 140 \cdot 166 \cdot 1 + 140 \cdot 188 \cdot 1 + 160 \cdot 78 \cdot 1 +$$

$$+160 \cdot 100 \cdot 9 + 160 \cdot 122 \cdot 1 + 180 \cdot 100 \cdot 1 + 180 \cdot 122 \cdot 2 + 220 \cdot 144 \cdot 1) -$$

$$-136{,}40 \cdot 85{,}92 = 12249{,}6 - 11719{,}49 = 530{,}11.$$

Выборочный коэффициент корреляции

$$r_B = \frac{K_B}{S_X \cdot S_Y}$$

$$r_B = \frac{530{,}11}{26{,}71 \cdot 30{,}72} = \frac{530{,}11}{820{,}5312} \approx 0{,}65.$$

Положительный знак выборочного коэффициента корреляции r_B показывает, что с увеличением значений случайной величины X эмпирические значения случайной величины Y в среднем возрастают.

18. ОСНОВЫ РЕГРЕССИОННОГО АНАЛИЗА

18.1. Задачи и этапы регрессионного анализа

З***адачами регрессионного анализа*** являются

- Нахождение уравнений регрессии (*функциональной* зависимости между случайными величинами X и Y);
- вычисление коэффициентов, входящих в уравнения;
- прогноз значений зависимой переменной.

При нахождении эмпирического уравнения регрессии $\bar{y}_x = \varphi(x)$ зависимая переменная y есть величина случайная, а независимая переменная x – величина неслучайная; и наоборот, при нахождении уравнения $\bar{x}_y = \psi(y)$ зависимая переменная x есть величина случайная, а независимая переменная y – величина неслучайная.

Обозначение $\bar{y}_x$ читается как «игрек на икс», обозначение $\bar{x}_y$ читается как «икс на игрек».

Основные этапы регрессионного анализа

На практике регрессионный анализ случайных величин X и Y состоит из трех этапов.

I. *На первом этапе* выдвигают гипотезу о виде искомой зависимости $\bar{y}_x = \varphi(x)$ (или $\bar{x}_y = \psi(y)$).

На практике встречаются следующие зависимости между случайными величинами X и Y:

– линейная $y = ax + b$;

– дробно-линейная $y = \dfrac{1}{a + b\,x}$;

– гиперболическая $y = a + \dfrac{b}{x}$;

– степенная $y = a \cdot x^b$;

– показательная $y = a \cdot b^x$;

II. *На втором этапе* по имеющимся экспериментальным данным находят оценки неизвестных параметров уравнения регрессии.

III. *На третьем этапе* проверяют справедливость принятой гипотезы относительно результатов измерений (находят оценку значимости параметров уравнения регрессии).

18.2. Метод наименьших квадратов

С помощью этого метода находят оценки неизвестных параметров уравнения регрессии по имеющимся экспериментальным данным.

Найдем с помощью метода наименьших квадратов уравнение **линейной регрессии.**

Пусть известны выборочные значения (x_i, y_i) двух случайных величин X и Y.

Уравнение линейной регрессии имеет вид $y = ax + b$, где a, b – неизвестные параметры (или коэффициенты) линейной функции.

Требуется найти a, b ***методом наименьших квадратов*** по имеющимся экспериментальным данным.

По методу наименьших квадратов параметры a, b выбирают так, чтобы сумма квадратов отклонений

$$\sum_{i=1}^{n}(y_i-(ax_i+b))^2$$

была минимальной.

Для нахождения минимума данной функции двух переменных приравниваем к нулю частные производные по этим переменным a, b.

$$\begin{cases} 2\sum_{i=1}^{n}(y_i-ax_i-b)=0, \\ 2\sum_{i=1}^{n}x_i\cdot(y_i-ax_i-b)=0. \end{cases}$$

После простых алгебраических преобразований получаем систему линейных уравнений для определения параметров a, b:

$$\begin{cases} a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_i\,y_i, \\ a\sum_{i=1}^{n}x_i+bn=\sum_{i=1}^{n}y_i. \end{cases}$$

Решая полученную систему, найдем значения a, b.

Можно показать, что ***угловой коэффициент*** прямой линии регрессии $\bar{y}_x$ равен

$$a=r_B\cdot\frac{S_Y}{S_X}.$$

Таким образом, если зависимость между X и Y линейная, то угловой коэффициент линии регрессии выражается через коэффициент корреляции.

Итак, функциональная зависимость значений величины Y от значений величины X выражается с помощью уравнения линейной регрессии:

$$\bar{y}_x=\bar{Y}_B+r_B\cdot\frac{S_Y}{S_X}\cdot(x-\bar{X}_B).$$

Аналогично, функциональная зависимость значений величины X от значений величины Y выражается с помощью уравнения линейной регрессии, имеющего вид

$$\bar{x}_y = \bar{X}_B + r_B \cdot \frac{S_X}{S_Y}(y - \bar{Y}_B),$$

отсюда угловой коэффициент прямой регрессии $\bar{x}_y$

$$a_1 = r_B \cdot \frac{S_X}{S_Y}.$$

Точка пересечения прямых линий регрессий $\bar{y}_x = ax + b$ и $\bar{x}_y = a_1 y + b_1$ называется ***центром рассеивания*** системы случайных величин X и Y.

Пример 41

По условию примера 40 провести регрессионный анализ, найти выборочные уравнения линейной регрессии.

Решение

Построим корреляционное поле, для этого изобразим заданные выборочные значения (x_i, y_i) в виде точек в декартовой системе координат (рис. 13).

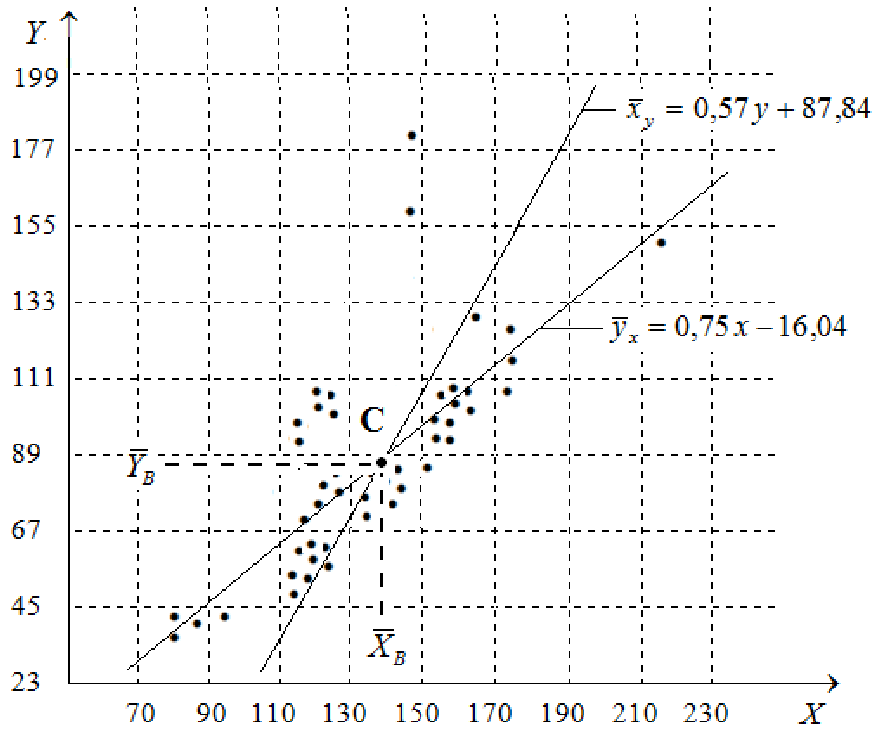


Рисунок 13 – Корреляционное поле в декартовой системе координат

По виду корреляционного поля (точки на плоскости) видно, что между X и Y имеется заметная прямолинейная регрессионная зависимость.

Найдем выборочное уравнение регрессии «игрек на икс»:

$$\bar{y}_x = \overline{Y}_B + r_B \cdot \frac{S_Y}{S_X} \cdot (x - \overline{X}_B).$$

По условию задачи

$$\bar{y}_x = 85{,}92 + 0{,}65 \cdot \frac{30{,}72}{26{,}71}(x - 136{,}4),$$

или после вычисления и округления параметров:

$$\bar{y}_x = 0{,}75x - 16{,}04.$$

Аналогично, находим выборочное уравнение регрессии «икс на игрек»:

$$\bar{x}_y = \overline{X}_B + r_B \cdot \frac{S_X}{S_Y}(y - \overline{Y}_B)$$

В нашем случае

$$\bar{x}_y = 136{,}4 + 0{,}65 \cdot \frac{26{,}71}{30{,}72}(y - 85{,}92),$$

$$\bar{x}_y = 0{,}57y + 87{,}84.$$

Графики полученных линий регрессий – прямые – построим на том же рисунке, где изображено корреляционное поле (см. рис. 13).

Центром рассеивания системы случайных величин X и Y является точка C – точка пересечения прямых линий регрессий (см. рис. 13).

18.3. Нелинейная регрессия

Если функция регрессии не является линейной, то с помощью замены переменной можно перейти к линейной функции.

Рассмотрим некоторые нелинейные функции:

а) **дробно-линейная** $y = \dfrac{1}{ax + b}$;

сделаем замену $y_1 = \dfrac{1}{y}$,

получим линейную функцию $y_1 = ax + b.$

b) **гиперболическая** $y = a + \dfrac{b}{x}$;

сделаем замену $y_1 = x \cdot y,$

получим линейную функцию $y_1 = ax + b.$

c) **степенная** $y = a \cdot x^b$;

сделаем замену $y_1 = \lg y, \quad b = \lg a,$

получим линейную функцию $y_1 = ax + b.$

ПРАКТИКУМ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Занятие 1

Тема: Пространство элементарных событий, операции над событиями. Классическое определение вероятности.

1.1. Основные знания и умения

Необходимо:

знать

✓ определение пространства элементарных событий, случайного события, операций над событиями (суммы, произведения, разности событий);

✓ определение вероятности события: аксиоматическое и классическое;

уметь

✓ пользоваться основными формулами комбинаторики;

✓ описывать пространство элементарных событий в конкретной задаче (в случайном эксперименте);

✓ находить вероятность события, используя классическое определение;

✓ находить вероятность противоположного события, суммы и произведения событий.

1.2. Применение знаний при решении типовых примеров и задач

Задача 1. Описать пространство элементарных событий и найти количество всех возможных элементарных событий в следующих задачах:

а) набор шестизначного телефонного номера;

б) изучение совместного влияния на агрохимические показатели почвы любой тройки удобрений из пяти имеющихся, если порядок внесения удобрений несущественен;

в) в виварии имеются 10 подопытных животных. Для трех различных экспериментов требуется отобрать по одному животному.

Решение

а) Набор телефонного номера из шести цифр можно рассматривать как результат выбора с возвращением и с учетом порядка шести цифр из следующих десяти: 0, 1, …, 9.

Пространство элементарных исходов образует различные кортежи из шести цифр, например 210635, 201356, 112345, … . Всего различных телефонных номеров будет 10^6.

Если число предметов в совокупности равно n и производится выбор k предметов с возвращением и учетом порядка, то количество всех элементарных исходов будет равно n^k.

б) В данной задаче эксперимент состоит в том, что случайным образом выбирается 3 удобрения из 5 имеющихся.

Будем считать, что удобрения пронумерованы, тогда пространство элементарных исходов здесь образуют различные кортежи из 3 чисел, которые выбирают без учета порядка из чисел: 1, 2, 3, 4, 5. Количество всех возможных элементарных событий находится по формуле для числа сочетаний: $C_5^3 = \dfrac{5!}{3!\cdot 2!}$.

Если число предметов в совокупности равно n и производится выбор k предметов без возвращения и без учета порядка, то количество всех элементарных исходов $C_n^k = \dfrac{n!}{k!(n-k)!}$.

в) Число различных вариантов выбора подопытных животных из вивария, содержащего 10 животных, будет равно числу размещений из 10 элементов по 3 элемента:

$$A_{10}^3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = \frac{7!\cdot 8\cdot 9\cdot 10}{7!} = 8\cdot 9\cdot 10 = 720.$$

Задача 2. Найти вероятность получения потомка с доминантным признаком B при скрещивании двух гетерозиготных особей Bb.

Хромосома – это особая структура внутри ядра клетки организма, которая хранит и передает генетическую информацию.

Гетерозиготные особи – это организмы, у которых в определенных участках хромосом представлены разные гены. Например, если один родитель передал признак B, а другой b, то у потомка будет гетерозиготное состояние Bb.

Гомозиготные особи – это организмы, у которых в определенных участках парных хромосом унаследованы одинаковые гены. Например, если оба родителя передали признак B или b, то у потомка возможны гомозиготные состояния BB или bb.

Решение

Рассмотрим событие

$$A = \{\text{Получение потомка с доминантным признаком } B\}.$$

Воспользуемся формулой классической вероятности:

$P(A) = \dfrac{m}{n}$, где m – число исходов, благоприятствующих событию A, n – число всех возможных исходов.

При скрещивании двух гетерозиготных особей Bb возможны четыре элементарных исхода (появление особей типа BB, Bb, bB, bb). Тогда получению потомка с доминантным признаком B благоприятствует три исхода (BB, Bb, bB). Таким образом, по условию задачи $n = 4$, $m = 3$. Тогда

$$P(A) = \frac{3}{4} = 0{,}75.$$

Задача 3. Два стрелка стреляют по мишени. Вероятность попадания для первого стрелка – 0,7, для второго – 0,8. Найти вероятность следующих событий:

а) оба стрелка попадут в цель;

б) оба стрелка промахнутся;

в) только один стрелок попадет в цель;

г) хотя бы один попадет.

Решение

Рассмотрим события

$$A = \{\text{Попал первый стрелок}\},\ B = \{\text{Попал второй стрелок}\},$$
$$\overline{A} = \{\text{Первый не попал}\},\ \overline{B} = \{\text{Второй не попал}\}.$$

Тогда

$$P(A) = 0{,}7; \qquad P(\overline{A}) - 1 - P(A) = 0{,}3;$$
$$P(B) = 0{,}8; \qquad P(\overline{B}) - 1 - P(B) = 0{,}2.$$

а) $P(A \cdot B) = P(A) \cdot P(B) = 0{,}7 \cdot 0{,}8 = 0{,}56;$

б) $P(\overline{A} \cdot \overline{B}) = P(\overline{A}) \cdot P(\overline{B}) = 0{,}3 \cdot 0{,}2 = 0{,}06;$

в) $P(\overline{A} \cdot B + \overline{B} \cdot A) = P(\overline{A}) \cdot P(B) + P(\overline{B}) \cdot P(A) = 0{,}3 \cdot 0{,}8 + 0{,}2 \cdot 0{,}7 = 0{,}38;$

г) событие $\overline{A}\cdot\overline{B}$ («оба стрелка промахнулись») будет противоположным для события «хотя бы один попал в цель». Следовательно, исходная вероятность находится так: $1-P(\overline{A}\cdot\overline{B})=1-0{,}2\cdot 0{,}3=0{,}94$.

Задача 4. В группе 20 студентов, из них 12 юношей и 8 девушек. На конференцию выбрали делегацию из пяти человек. Какова вероятность того, что в составе делегации 3 девушки?

Решение

Вероятность искомого события

$$A=\{В\ составе делегации\ 3\ девушки\}$$

находится по формуле классической вероятности: $P(A)=\dfrac{m}{n}$.

Число всех возможных исходов N находится по формуле числа сочетаний: $n=C_{20}^{5}=\dfrac{20!}{5!\cdot 15!}$.

Число исходов, благоприятствующих наступлению события A, что в составе делегации из пяти студентов должно быть 3 девушки и 2 юноши, находится по формуле $m=C_{8}^{3}\cdot C_{12}^{2}$.

Тогда

$$P(A)=\frac{C_8^3\cdot C_{12}^2}{C_{20}^5}=\frac{8!\cdot 12!\cdot 5!\cdot 15!}{3!\cdot 5!\cdot 2!\cdot 10!\cdot 20!};$$

$$P(A)=\frac{4\cdot 5\cdot 6\cdot 7\cdot 8\cdot 11\cdot 12}{1\cdot 2\cdot 16\cdot 17\cdot 18\cdot 19\cdot 20}=\frac{77}{323}=0{,}239.$$

1.3. Самостоятельное применение знаний, умений, навыков

Задача 1. Описать пространство элементарных событий в следующих задачах:

а) выпадение очков при бросании игральной кости (кубика);

б) выпадение очков при бросании двух игральных костей.

Задача 2. Найти вероятность получения гомозиготного потомка с доминантным признаком B при скрещивании двух гетерозиготных особей Bb.

Задача 3. Найти вероятность того, что из шести отмеченных чисел в карточке спортлото 4 числа будут выигрышными.

Задача 4. На рассаду вырастили 7 корней красных и 5 корней желтых помидоров, наугад высаживают 7 корней. Найти вероятность того, что при этом 3 корня будут с красными помидорами, а 4 – с желтыми.

Задача 5. Вы купили 3 лотерейных билета, вероятность выигрыша по одному билету равна 0,1. Какова вероятность выиграть:

а) по всем билетам;

б) ни по одному билету;

в) хотя бы по одному билету?

Занятие 2

Тема: Условная вероятность. Формула полной вероятности. Схема и формула Бернулли.

2.1. Основные знания и умения

Необходимо:

знать

- ✓ определение условной вероятности;
- ✓ понятие полной группы событий;
- ✓ формулу полной вероятности;
- ✓ вероятную схему Бернулли;
- ✓ формулу Бернулли;
- ✓ формулу для наивероятнейшего числа;

уметь

- ✓ находить условную вероятность;
- ✓ описать полную группу событий;
- ✓ пользоваться формулой полной вероятности;
- ✓ описать схему Бернулли;
- ✓ применять формулу Бернулли;
- ✓ находить наивероятнейшее число.

2.2. Применение знаний при решении типовых примеров и задач

Задача 1. В каждой из двух урн находится по 4 белых и 6 черных шара. Из первой урны наудачу переложили один шар во вторую, после этого из второй урны наудачу извлекли шар. Какова вероятность того, что этот шар белый?

Решение

Рассмотрим следующие события:

$A = \{Шар, извлеченный\ из\ второй\ урны\ белый\}$;

$H_1 = \{Переложили\ белый\ шар\ из\ первой\ урны\ во\ вторую\}$;

$H_2 = \{Переложили\ черный\ шар\ из\ первой\ урны\ во\ вторую\}$;

H_1 и H_2 образуют полную группу событий.

Тогда по формуле полной вероятности имеем

$$P(A) = P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2).$$

Условные вероятности событий H_1 и H_2 находим, используя классическое определение вероятности:

$$P(H_1) = \frac{m}{n} = \frac{4}{10}; \qquad P(H_2) = \frac{6}{10};$$

$$P(A/H_1) = \frac{5}{11}; \qquad P(A/H_2) = \frac{4}{11}.$$

Итак, $\quad P(A) = \frac{4}{10} \cdot \frac{5}{11} + \frac{6}{10} \cdot \frac{4}{11} = \frac{44}{110} = 0{,}4.$

Задача 2. Найти вероятность того, что при 4 подбрасываниях монеты орел выпадет:

а) 3 раза;

б) не менее 3 раз;

в) более 3 раз.

Решение

Вероятность выпадения орла при одном бросании монеты равна $p = 0{,}5;$ тогда $q = 1 - p = 0{,}5$.

По формуле Бернулли $P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}$, при $n = 4, k = 3$ имеем

а) $P_4(3) = C_4^3 \cdot p^3 \cdot q^{4-3} = \frac{4!}{3! \cdot 1!} \cdot (0{,}5)^3 \cdot 0{,}5 = \frac{4}{16} = \frac{1}{4};$

б) $P_4(3) + P_4(4) = \frac{1}{4} + C_4^4 \cdot (0{,}5)^4 \cdot (0{,}5)^0 = \frac{5}{16};$

в) $P_4(4) = \frac{1}{16}.$

Задача 3. Примем, что вероятность появления петухов из яиц составляет около 49 %. Курица высиживает 10 яиц. Найти вероятность того, что событие A – «из яйца появился петушок» наступит:

а) не менее 2 и не более 4 раз;

б) хотя бы 1 раз.

Решение:

а) Условия задачи соответствуют схеме испытаний Бернулли, поэтому вероятность успеха $p = P(A) = 0{,}49$, а вероятность неудачи $q = P(\overline{A}) = 1 - p = 1 - 0{,}49 = 0{,}51$, количество испытаний $n = 10$.

Вероятность того, что в n испытаниях событие наступит от k_1 до k_2 раз, находим по формуле

$$P_n(k_1 < k < k_2) = \sum_{k=k_1}^{k_2} C_n^k \cdot P^k \cdot q^{n-k}.$$

Тогда

$$P_{10}(2 < k < 4) = P_{10}(2) + P_{10}(3) + P_{10}(4) =$$
$$= C_{10}^2 \cdot (0{,}49)^2 \cdot (0{,}51)^8 + C_{10}^3 \cdot (0{,}49)^3 \cdot (0{,}51)^7 + C_{10}^4 \cdot (0{,}49)^4 \cdot (0{,}51)^6 = 0{,}389.$$

б) Вероятность того, что в n испытаниях событие наступит хотя бы один раз, находим по формуле

$$P_n(1 < k < n) = 1 - q^n.$$

В нашем случае

$$P_{10}(1 < k < 10) = 1 - (0{,}51)^{10} = 0{,}999.$$

Задача 4. Сколько надо посеять семян, всхожесть которых 70 %, чтобы наивероятнейшее число взошедших семян было равно 60?

Решение

По условию задачи $p = 0{,}7$ – вероятность того, что посеянное семя взойдет, а вероятность $q = 1 - p = 1 - 0{,}7 = 0{,}3$ – вероятность того, что семя не даст всхода, $k_0 = 60$ – наивероятнейшее число взошедших семян. Требуется найти n – число посеянных семян. Наивероятнейшее число k_0 находится из двойного неравенства

$$np - q \le k_0 \le np + p.$$

Подставив известные величины в неравенство, получим

$$n \cdot 0{,}7 - 0{,}3 \le 60 \le n \cdot 0{,}7 + 0{,}7.$$

Переходим к системе неравенств:

$$\begin{cases} n\cdot 0{,}7-0{,}3\le 60, \\ n\cdot 0{,}7+0{,}7\ge 60. \end{cases} \Rightarrow \begin{cases} n\cdot 0{,}7\le 60{,}3, \\ n\cdot 0{,}7+0{,}7\ge 59{,}3. \end{cases} \Rightarrow \begin{cases} n\le 86{,}1, \\ n\ge 84{,}7. \end{cases} \Rightarrow 84{,}7\le n\le 86{,}1.$$

Поскольку число семян n является целым числом, то $n=85$ или $n=86$.

2.3. Самостоятельное применение знаний, умений, навыков

Задача 1. В студенческом стройотряде две группы студентов, в первой группе 6 юношей и 4 девушки, во второй – 5 юношей и 5 девушек. Наудачу из стройотряда выбрали одну из групп, а из нее одного человека для поездки в штаб студенческих отрядов. Какова вероятность того, что выбран юноша?

Задача 2. У молочного завода четыре поставщика молока. От первого поставщика поступает 20 % молока, от второго – 40 %, от третьего – 30 %, от четвертого – 10 %. Отклонения от качества и безопасности в пробах выявляются в 3 % молока от первого поставщика, в 1 % – от второго, в 2 % – от третьего, в 5 % – от четвертого. Какова вероятность того, что в выбранной наудачу пробе найдут отклонения от качества и безопасности?

Задача 3. На сельскохозяйственном предприятии имеется 8 комбайнов. Вероятность того, что каждый комбайн выйдет на уборку урожая, равна 0,8. Какова вероятность нормальной работы предприятия, если для этого необходимо иметь в поле на уборке как минимум 6 машин?

Задача 4. Всхожесть гороха 80 %. Найти вероятность того, что среди 10 замоченных горошин не взойдет:

а) 3 горошины;

б) не более 3 горошин.

Задача 5. В саду растут 12 яблонь. Вероятность того, что яблоня даст урожай в конкретном году, равна 0,75. Каково наивероятнейшее число яблонь, которые дадут урожай в этом году.

Задача 6. Известно, что всхожесть пшеницы составляет 90 %. Сколько необходимо взять зерен, чтобы взошло 270 растений?

Занятие 3

Тема: Приближенные формулы для схемы испытаний Бернулли.

3.1. Основные знания и умения

Необходимо:

знать

- ✓ формулу Пуассона;
- ✓ функцию Гаусса;
- ✓ локальную теорему Муавра-Лапласа;
- ✓ нормированную функцию Лапласа;
- ✓ функцию Лапласа;
- ✓ интегральную теорему Муавра-Лапласа;
- ✓ знать условия применения формул.

уметь

- ✓ пользоваться формулой Пуассона;
- ✓ применять локальную формулу Муавра-Лапласа;
- ✓ вычислять вероятность с помощью интегральной формулы Муавра – Лапласа.

3.2. Применение знаний при решении типовых примеров и задач

Задача 1. Вероятность того, что зерно пшеницы не прорастет, принимается равным 0,05. Какова вероятность того, что из 800 посеянных семян не дадут всходов: а) не более 5 семян; б) более 5 семян.

Решение

Пусть событие A заключается в том, что зерно пшеницы не прорастет. По условию задачи событие A является редким, так как число испытаний велико ($n = 800$), а вероятность успеха мала ($p = P(A) = 0{,}005$), кроме того $\lambda = n \cdot p = 800 \cdot 0{,}005 = 4 \le 10$, поэтому возникает основание применить формулу Пуассона:

$$P_n(k) \approx \frac{\lambda^k \cdot e^{-\lambda}}{k!}.$$

а)

$$P_{800}(k \le 5) = \sum_{k=0}^{5} \frac{\lambda^k \cdot e^{-2}}{k!} = P_{800}(0) + P_{800}(1) + P_{800}(2) + P_{800}(3) + P_{800}(4) + P_{800}(5) =$$

$$= \frac{4^0 \cdot e^{-4}}{0!} + \frac{4^1 \cdot e^{-4}}{1!} + \frac{4^2 \cdot e^{-4}}{2!} + \frac{4^3 \cdot e^{-4}}{3!} + \frac{4^4 \cdot e^{-4}}{4!} + \frac{4^5 \cdot e^{-4}}{5!} \approx 0{,}7852.$$

б) $P_{800}(k > 5) = 1 - P_{800}(k \le 5) = 1 - 0{,}7852 = 0{,}2148.$

Задача 2. Лежкость картофеля в данных условиях равна 90 %. Найти вероятность того, что из 300 кг картофеля, заложенного на хранение на зиму, до весны сохранится: а) 260 кг; б) от 260 до 280 кг.

Решение

Пусть событие

$A = \{\textit{Заложенныйна хранениекилограммкартофелясохранился}\}.$

$p(A) = 0{,}9,\ q = 1 - p = 0{,}1$. Всего производится $n = 300$ независимых испытаний.

а) Это случай в схеме испытаний Бернулли, когда число испытаний n велико, а вероятность p не близка к нулю, поэтому применим локальную формулу Муавра – Лапласа. Событие A должно произойти ровно $k = 260$ раз. Имеем

$$\sqrt{npq} = \sqrt{300 \cdot 0{,}9 \cdot 0{,}1} = \sqrt{27} = 5{,}196;$$

$$x = \frac{k - np}{\sqrt{npq}} = \frac{260 - 300 \cdot 0{,}9}{5{,}196} = -\frac{10}{5{,}196} \approx -1{,}92.$$

По таблице значений функции $\varphi(x)$ (Приложение А) найдем

$$\varphi(-1{,}92) = \varphi(1{,}92) = 0{,}0632.$$

Тогда

$$P_{300}(260) = \frac{1}{5{,}196} \cdot \varphi(-1{,}92) \approx \frac{1}{5{,}196} \cdot 0{,}0632 = 0{,}012.$$

б) Поскольку число опытов велико, а количество исходов k рассматривается в заданном промежутке $k_1 < k < k_2$, вычислим вероятность по интегральной формуле Муавра – Лапласа.

По условию задачи: $k_1 = 260, k_2 = 280,$ тогда

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{280 - 300 \cdot 0{,}9}{5{,}196} = \frac{10}{5{,}196} \approx 1{,}92;$$

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{260 - 300 \cdot 0{,}9}{5{,}196} = -\frac{10}{5{,}196} \approx -1{,}92;$$

Искомую вероятность найдем по формуле

$$P_n(k_1 \le k \le k_2) = \Phi_0(x_2) - \Phi_0(x_1).$$

По таблице значений нормированной функции Лапласа $\Phi_0(x)$ (Приложение В) находим, что $\Phi_0(1{,}92) = 0{,}47257.$ Поскольку функция $\Phi_0(x)$ нечетная, то $\Phi_0(-1{,}92) = -\Phi_0(1{,}92) = -0{,}47257.$ Тогда
$P_{300}(260 \le k \le 280) = \Phi_0(1{,}92) - \Phi_0(-1{,}92) = 2\Phi_0(1{,}92) = 2 \cdot 0{,}47257 \approx 0{,}95.$

3.3. Самостоятельное применение знаний, умений, навыков

Задача 1. Предприятие по изготовлению вин отправило по договору поставки 1500 бутылок вина «Кагор». Вероятность того, что в пути бутылка может разбиться, равна 0,002. Найти вероятность того, что в пути будет разбито не более 3 бутылок.

Задача 2. Найти вероятность одновременной остановки 30 доильных аппаратов из 100 работающих на животноводческом комплексе, если вероятность остановки для каждого доильного аппарата равна 0,2.

Задача 3. Вероятность, что продукт получится высшего качества, равна 0,6. Найти вероятность, что из 400 изделий число изделий высшего качества составит от 194 до 208.

Задача 4. Вероятность появления события в каждом из 900 независимых испытаний равна 0,5. Найти вероятность того, что относительная частота появления события отклоняется от его вероятности не более чем на 0,02.

Занятие 4

Тема: Случайные величины, их числовые характеристики. Основные вероятностные распределения.

4.1. Основные знания и умения

Необходимо:

знать

✓ определение случайной величины (дискретной и непрерывной);

✓ определение функций распределения в плотности случайной величины;

✓ определение числовых характеристик: математического ожидания, дисперсии, среднего квадратического отклонения для дискретных и непрерывных случайных величин;

✓ основные вероятностные распределения;

уметь

✓ находить функцию распределения, плотность вероятности;

✓ находить числовые характеристики случайных величин, вероятность попадания значения случайной величины в заданный интервал.

4.2. Применение знаний при решении типовых примеров и задач

Задача 1. Производится 3 независимых опыта, в каждом из которых вероятность появления события A одинакова и равна 0,4 ($P(A)=0{,}4$). Рассматривается случайная величина X – число появления события A в трех опытах. Построить ряд распределения случайной величины X. Найти ее математическое ожидание $M(X)$, дисперсию $D(X)$, функцию распределения $F(X)$.

Решение

Закон распределения данной случайной величины представляется следующей таблицей:

x_i	0	1	2	3
p_i	0,216	0,432	0,288	0,064

где $p_i = P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}$; $k = i-1$; $\quad i = 1, \ldots, n+1$.

Вычисляем

$$p_1 = P_3(0) = C_3^0 \cdot 0{,}4^0 \cdot 0{,}6^3 = 0{,}216;$$

$$p_2 = P_3(1) = C_3^1 \cdot 0{,}4^1 \cdot 0{,}6^2 = 0{,}432.$$

Функция распределения $F(x) = P(X < x)$ имеет вид

$$F(x) = \begin{cases} 0, & x \le 0; \\ 0{,}216, & 0 < x \le 1; \\ 0{,}648, & 1 < x \le 2; \\ 0{,}936, & 2 < x \le 3; \\ 1, & x > 3. \end{cases}$$

Числовые характеристики:

$$M(X) = \sum_{i=1}^{4} p_i \cdot x_i = 0 \cdot 0{,}216 + 1 \cdot 0{,}432 + 2 \cdot 0{,}288 + 3 \cdot 0{,}064 = 1{,}2;$$

$$D(X) = M(X^2) - [M(X)]^2 =$$
$$= 0 \cdot 0{,}216 + 1 \cdot 0{,}432 + 4 \cdot 0{,}288 + 9 \cdot 0{,}064 - 1{,}2^2 = 0{,}72.$$

Задача 2. Известна функция распределения некоторой случайной величины:

$$F(x) = \begin{cases} 0, & x \le 0; \\ \dfrac{x^2}{49}, & 0 < x \le 7; \\ 1, & x > 7. \end{cases}$$

Найти $f(x)$, $M(X)$, $D(X)$, $(2 < x < 5)$.

Решение

$$f(x)=F'(x)=\begin{cases}0, & x\le 0;\\ \dfrac{2x}{49}, & 0<x\le 7;\\ 0, & x>7.\end{cases}$$

$$M(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx=\int_0^7 x\cdot\frac{2x}{49}dx=\frac{2x^3}{3\cdot 49}\Big|_0^7=\frac{14}{3}=4\frac{2}{3};$$

$$D(X)=\int_{-\infty}^{+\infty} x^2\cdot f(x)dx-M(X)^2=\int_0^7\frac{2x^3}{49}dx-(\frac{14}{3})^2=2\frac{13}{18}.$$

Вероятность того, что случайная величина X примет значения принадлежащие интервалу $(2;5)$, находим по формуле

$$P(2<x<5)=F(5)-F(2)=\frac{5^2}{49}-\frac{2^2}{49}=\frac{21}{49}=\frac{3}{7}.$$

Задача 3. Автобусы идут с интервалом 10 мин. Предполагая, что время ожидания автобуса на остановке имеет равномерное распределение, найти:

а) плотность вероятности;

б) функцию распределения;

в) вероятность того, что время ожидания не превзойдет 4 мин.

Решение

а) Плотность вероятности равномерно распределенной случайной величины имеет вид

$$f(x)=\begin{cases}0, & x\le a;\\ \dfrac{1}{b-a}, & a<x\le b;\\ 0, & x>b.\end{cases}$$

Так как интервал движения автобуса равен 10 мин, то

$$\text{а) } f(x)=\begin{cases}0, & x\le 0;\\ \dfrac{1}{10}, & 0<x\le 10;\\ 0, & x>10.\end{cases}$$

$$\text{б) } F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases}0, & x\le 0;\\ 0{,}1x, & 0<x\le 10;\\ 1, & x>10.\end{cases}$$

$$\text{в) } P(x\le 4)=F(4)-F(0)=0{,}1\cdot 4-0=0{,}4.$$

Задача 4. Размер диаметра втулок, изготовленных заводом, можно считать случайной величиной, распределенной по нормальному закону с параметрами $a=2{,}5,\ \sigma=0{,}01$.

Найти выражение для плотности вероятности и функции распределения.

Решение

Плотность вероятности нормального распределения имеет вид

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

а функция распределения $F(x)=\dfrac{1}{\sigma\cdot\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}dt.$

Следовательно, в нашем случае

$$f(x)=\frac{1}{0{,}01\cdot\sqrt{2\pi}}=e^{-\frac{(x-2{,}5)^2}{0{,}0002}};\qquad F(x)=\frac{1}{0{,}01\cdot\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-2{,}5)^2}{0{,}0002}}dt.$$

Используя нормированную функцию Лапласа

$$\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt,$$

значения которой приводятся в таблице (приложение В), запишем

$$F(x)=0{,}5+\Phi_0\left(\frac{x-a}{\sigma}\right).$$

При $a=2{,}5$ и $\sigma=0{,}01$ имеем

$$F(x)=0{,}5+\Phi_0\left(\frac{x-2{,}5}{0{,}01}\right)=0{,}5+\Phi_0(100x-250).$$

3.1. Самостоятельное применение знаний, умений, навыков

Задача 1. В некоторой партии зерна $1/5$ часть зерен невсхожие. Наугад выбираются 4 зерна. Найти закон распределения случайной величины X, равной количеству выбранных всхожих зерен.

Найти:

а) функцию распределения, построить ее график;

б) вероятность $P(1\le x\le 2)$;

в) $M(X)$, $D(X)$.

Задача 2. Случайная величина X задана функцией распределения:

$$F(x)=\begin{cases}0, & x\le 1;\\ x-1, & 1<x\le 2;\\ 1, & x>2.\end{cases}$$

Найти:

а) функцию плотности $f(x)$, построить ее график;

б) $M(X)$, $D(X)$.

Задача 3. Случайная величина X имеет плотность вероятности (закон Коши) $f(x)=\dfrac{c}{x^2+1}$.

Найти:

а) постоянную c (применить свойство функции плотности $\int\limits_{-\infty}^{+\infty} f(x)dx=1$);

б) функцию распределения $F(X)$;

в) вероятность $P(-1<x<1)$.

Задача 4. Даны все возможные значения дискретной случайной величины X: $X_1=1$; $X_2=2$; $X_3=3$.

Известны $M(X)=2{,}3$ и $M(X^2)=5{,}9$.

Найти закон распределения X.

ПРАКТИКУМ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Тема: Статистическая обработка экспериментальных данных

Даны результаты 100 наблюдений над некоторыми случайными величинами X и Y.

Требуется для каждой случайной величины X и Y:

1. Построить интервальный и дискретный статистический ряды распределения частот и относительных частот.
2. Построить гистограмму и полигон относительных частот.
3. Найти эмпирическую функцию распределения и построить ее график.
4. Вычислить числовые характеристики выборки: выборочную среднюю, выборочную дисперсию, выборочное среднее квадратическое отклонение, выборочные коэффициенты асимметрии и эксцесса.
5. Определить закон распределения наблюдаемой случайной величины, исходя из механизма ее образования, по виду гистограммы и полигона относительных частот и по значениям выборочных коэффициентов асимметрии и эксцесса.
6. Найти точечные оценки параметров нормального закона распределения, предполагая, что наблюдаемая случайная величина распределена по нормальному закону, и записать функцию плотности распределения вероятностей.
7. Проверить с помощью критерия согласия Пирсона гипотезу о том, что выборка извлечена из генеральной совокупности с предполагаемым нормальным законом распределения.
8. В случае принятия гипотезы найти интервальные оценки параметров нормального закона распределения (доверительную вероятность принять равной $\gamma = 0{,}95$).
9. Провести корреляционный анализ случайной **двумерной величины** (X, Y):

9.1. Составить корреляционную таблицу.

9.2. Найти выборочный коэффициент корреляции.

9.3. Проверить значимость выборочного коэффициента корреляции r_B при $\lambda = 0{,}05$ ($H_0 : p = 0$), при альтернативной гипотезе H_λ: $p \neq 0$.

10. Провести регрессионный анализ:

10.1. Построить корреляционное поле и по характеру расположения точек на нем подобрать общий вид функции регрессии.

10.2. Найти эмпирические функции регрессии Y на X, X на Y и построить их графики.

Выборочные значения случайных величин X и Y приведены в следующей таблице:

№	X	Y	**№**	X	Y	**№**	X	Y	**№**	X	Y	**№**	X	Y
1	17	34	**21**	16	38	**41**	13	37	**61**	11	39	**81**	15	37
2	9	42	**22**	20	41	**42**	15	39	**62**	7	43	**82**	7	44
3	12	35	**23**	15	33	**43**	3	47	**63**	14	37	**83**	17	32
4	11	37	**24**	9	37	**44**	14	38	**64**	13	34	**84**	15	36
5	11	40	**25**	20	38	**45**	15	36	**65**	11	36	**85**	13	37
6	14	37	**26**	11	40	**46**	11	39	**66**	20	30	**86**	17	34
7	21	32	**27**	11	39	**47**	13	35	**67**	12	42	**87**	12	39
8	18	34	**28**	18	33	**48**	14	37	**68**	8	44	**88**	9	41
9	14	37	**29**	9	41	**49**	12	38	**69**	12	37	**89**	17	36
10	6	46	**30**	10	31	**50**	16	37	**70**	20	29	**90**	15	37
11	16	40	**31**	16	35	**51**	14	36	**71**	14	40	**91**	8	43
12	15	38	**32**	13	40	**52**	14	35	**72**	15	34	**92**	13	38
13	13	39	**33**	18	33	**53**	19	28	**73**	11	41	**93**	12	40
14	13	36	**34**	14	37	**54**	15	35	**74**	10	39	**94**	16	36
15	18	32	**35**	13	39	**55**	11	41	**75**	9	38	**95**	19	31
16	13	37	**36**	17	33	**56**	5	47	**76**	10	42	**96**	10	41
17	11	39	**37**	18	36	**57**	13	38	**77**	8	41	**97**	18	34
18	5	33	**38**	12	39	**58**	8	47	**78**	18	34	**98**	12	40
19	21	30	**39**	9	42	**59**	14	39	**79**	7	41	**99**	10	41
20	14	36	**40**	13	38	**60**	16	36	**80**	7	43	**100**	13	37

Проведем статистическую обработку выборки для случайной величины X, согласно заданиям 1–8.

1. Статистическая обработка результатов эксперимента в случае выборки большого объема $(n \geq 50)$ начинается с группировки выборочных значений, т. е. с разбиения наблюдаемых значений *СВ* на k частичных интервалов равной длины и подсчета частот попадания значений случайной величины в частичные интервалы.

Сделаем группировку наблюдаемых значений. Оптимальную длину интервала определим по формуле Стерджеса:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \cdot \lg n},$$

где $x_{\max}, x_{\min}$ – соответственно максимальное и минимальное выборочные значения случайной величины X; n – объем выборки. Если h окажется дробным, то за величину интервала нужно взять либо ближайшее целое число, либо ближайшую несложную дробь.

Для случайной величины X $n = 100$, $x_{\max} = 21$, $x_{\min} = 3$. Следовательно,

$$h=\frac{21-3}{1+3{,}322\cdot\lg 100}=\frac{18}{1+3{,}322\cdot 2}\approx 2{,}4.$$

Вычислим количество интервалов:

$$k=1+3{,}322\cdot\lg 100=1+3{,}322\cdot 2\approx 8.$$

В качестве левой границы первого интервала возьмем величину

$$a_1=x_{\min}-\frac{h}{2}=3-\frac{2{,}4}{2}=3-1{,}2=1{,}8.$$

Если a_1 – начало 1-го интервала, тогда

$$a_2=a_1+h=1{,}8+2{,}4=4{,}2;$$

$$a_3=a_2+h=4{,}2+2{,}4=6{,}6 \text{ и т. д.}$$

Составим рабочую таблицу.

Таблица 9 – Параметры интервального и дискретного статистических рядов

Интервалы $(a_i;a_{i+1}]$	Середины интервалов $\frac{a_i+a_{i+1}}{2}$	Подсчет частот	Частоты n_i	Относит. частоты $w_i=\frac{n_i}{n}$	Накопленные относительные частоты $w_x=\frac{n_x}{n}$
(1.8; 4.2]	3		1	0,01	0,01
(4.2; 6.6]	5,4		3	0,03	0,04
(6.6; 9]	7,8		14	0,14	0,18
(9; 11.4]	10,2		15	0,15	0,33
(11.4; 13.8]	12,6		21	0,21	0,54
(13.8; 16.2]	15		26	0,26	0,8
(16.2; 18.6]	17,4		12	0,12	0,92
(18.6; 21]	19,8		8	0,08	1,00

2. Первый и пятый столбцы таблицы составляют интервальный статистический ряд относительных частот, графическое изображение которого – *гистограмма* относительных частот (рис. 14).

Дискретный статистический ряд относительных частот задается вторым и пятым столбцами, графическое изображение которого – *полигон* относительных частот – изображен на рисунке 14 ломаной линией.

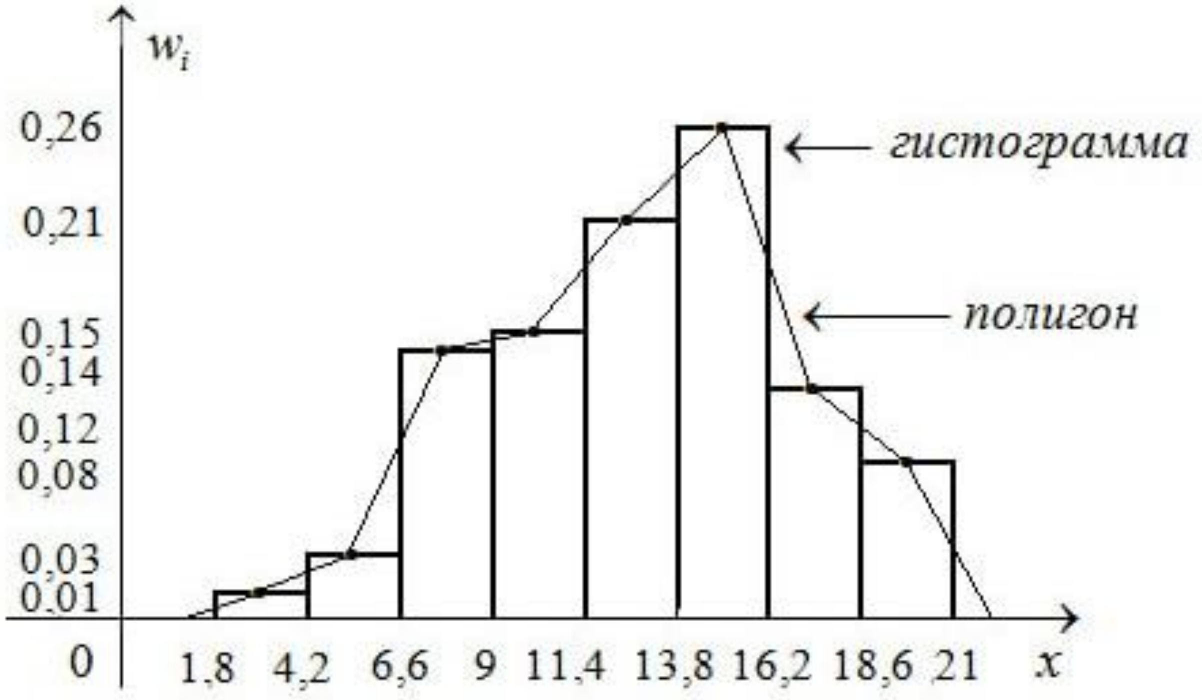


Рисунок 14 – Гистограмма и полигон относительных частот

3. Эмпирическая функция распределения $F^*(x)$ выборки служит для оценки функции распределения $F(x)$ генеральной совокупности. Функция $F^*(x)$ определяет для каждого значения x относительную частоту события $X < x$: $F^*(x) = \frac{n_x}{n}$, где n_x – число выборочных значений, меньших x; n – объем выборки. Шестой столбец таблицы содержит накопленные частоты, т. е. значения эмпирической функции распределения $F^*(x)$, они относятся к верхней границе частотного интервала.

Эмпирическая функция распределения $F^*(x)$ имеет следующий вид (ее график изображен на рисунке 15):

$$F^*(x)=\begin{cases} 0, & \text{если} \quad x \le 1{,}8; \\ 0{,}01, & \text{если} \quad 1{,}8 < x \le 4{,}2; \\ 0{,}04, & \text{если} \quad 4{,}2 < x \le 6{,}6; \\ 0{,}18, & \text{если} \quad 6{,}6 < x \le 9; \\ 0{,}33, & \text{если} \quad 9 < x \le 11{,}4; \\ 0{,}54, & \text{если} \quad 11{,}4 < x \le 13{,}8; \\ 0{,}80, & \text{если} \quad 13{,}8 < x \le 16{,}2; \\ 0{,}92, & \text{если} \quad 16{,}2 < x \le 18{,}6; \\ 1, & \text{если} \quad x > 18{,}6. \end{cases}$$

График эмпирической функции распределения $F^*(x)$ изображен на рисунке 15.

4. Вычисляем выборочное среднее по формуле

$$\overline{X}_B = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i :$$

$$\overline{X}_B = \frac{1}{100}\sum_{i=1}^{8} x_i \cdot n_i = \frac{1}{100}\cdot(3\cdot 1 + 5{,}4\cdot 3 + 7{,}8\cdot 14 + 10{,}2\cdot 15 + 12{,}6\cdot 21 +$$
$$+15\cdot 26 + 17{,}4\cdot 12 + 19{,}8\cdot 8) = \frac{1303{,}2}{100} = 13{,}032.$$

Выборочное среднее $\overline{X}_B$ дает усредненное значение длины колосьев ячменя для данной выборки.

Для вычисления числовых характеристик выборки D_B, σ_B, V_B, A_B, E_B удобно использовать таблицу 10.

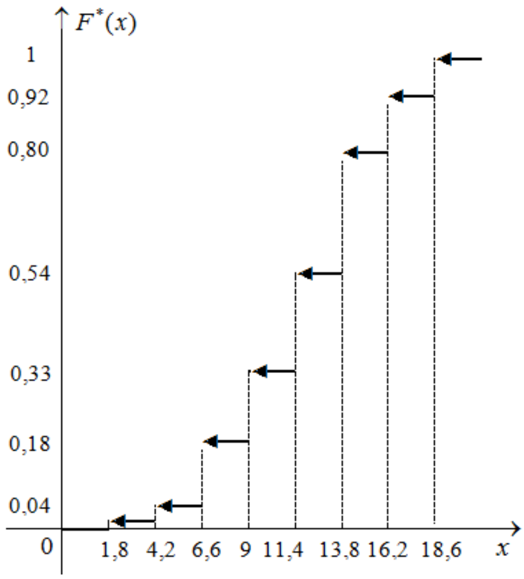


Рисунок 15 – График эмпирической функции распределения F(x)*

Таблица 10 – Таблица для расчета числовых характеристик выборки

Середины интервалов x_i	Частоты n_i	$x_i-\bar{x}_B$	$(x_i-\bar{x}_B)\cdot n_i$	$(x_i-\bar{x}_B)^2\cdot n_i$	$(x_i-\bar{x}_B)^3\cdot n_i$	$(x_i-\bar{x}_B)^4\cdot n_i$
3	1	–10,032	–10,032	100,641	–10009,63	10128,41
5,4	3	–7,632	–22,896	174,742	–1333,63	10178,26
7,8	14	–5,232	–73,248	383,234	–2005,08	10490,58
10,2	15	–2,232	–42,48	120,303	–340,70	964,86
12,6	21	–0,432	–9,072	3,919	–1,69	0,73
15	26	1,968	51,168	100,699	198,16	389,98
17,4	12	4,368	52,416	228,953	1000,07	4368,30
19,8	8	6,768	54,144	366,446	2480,11	16785,38
$\sum$	100	–	0	1478,94	–1012,39	53306,5

При заполнении таблицы нужно помнить, что сумма элементов четвертого столбца должна быть равна нулю.

Выборочную дисперсию для сгруппированных данных вычисляют по формуле

$$D_B(X)=\frac{1}{100}\sum_{i=1}^{8}(x_i-\bar{X}_B)^2\cdot n_i=\frac{1}{100}\cdot(3^2\cdot 1+5{,}4^2\cdot 3+7{,}8^2\cdot 14+10{,}2^2\cdot 15+$$

$$+12{,}6^2\cdot 21+15^2\cdot 26+17{,}4^2\cdot 12+19{,}8^2\cdot 8)=\frac{1478{,}94}{100}=14{,}789.$$

Выборочное среднее квадратическое отклонение находят по формуле $\sigma_B(X)=\sqrt{D_B(X)}$.

Итак, $\sigma_B(X)=\sqrt{D_B(X)}=\sqrt{14{,}789}\approx 3{,}84$(см) показывает разброс выборочных значений x_i относительно выборочного среднего $\bar{X}_B=13{,}032$.

Выборочные коэффициент асимметрии и эксцесс:

$$A_B=\frac{\mu_3^*}{\sigma_B^3}=\frac{\sum_i(x_i-\bar{X}_B)^3\cdot n_i}{n\cdot\sigma_B^3}=\frac{-1012{,}39}{100\cdot 3{,}84^3}=\frac{-1012{,}39}{100\cdot 56{,}7897}\approx -0{,}1782;$$

$$E_B=\frac{\mu_4^*}{\sigma_B^4}-3=\frac{\sum_i(x_i-\bar{X}_B)^4\cdot n_i}{n\cdot\sigma_B^4}-3=\frac{53306{,}5}{100\cdot 3{,}84^4}-3=$$

$$=\frac{53306{,}5}{100\cdot 218{,}7145}-3\approx -0{,}5628.$$

$A_B \neq 0$ говорит о несимметричности полигона (гистограммы) относительно выборочного среднего $\overline{X}_B$.

Отрицательный знак выборочного коэффициента асимметрии A_B свидетельствует о левосторонней асимметрии данного распределения.

Отрицательность выборочного коэффициента эксцесса показывает, что полигон менее крут, чем нормальная кривая.

5. Мы предварительно предполагаем, что СВ X распределена нормально по совокупности следующих признаков.

Вид полигона и гистограммы относительных частот напоминает нормальную кривую (кривую Гаусса).

Выборочные коэффициенты асимметрии $A_B = -0{,}1782$ и эксцесса $E_B = -0{,}5628$ отличаются от значений асимметрии и эксцесса для нормального распределения (которые равны нулю) не более, чем на их утроенные средние квадратические ошибки:

$$|A_B| = |-0{,}1782| < 0{,}7161 = 3 \cdot S_{A_B},$$

$$|E_B| = |-0{,}3628| < 1{,}3917 = 3 \cdot S_{E_B},$$

где $S_{A_B} = \sqrt{\dfrac{6(n-1)}{(n+1)(n+3)}} = \sqrt{\dfrac{6 \cdot 99}{101 \cdot 103}} \approx 0{,}2387,$

$$S_{E_B} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}} = \sqrt{\frac{24 \cdot 100 \cdot 98 \cdot 97}{99^2 \cdot 103 \cdot 105}} \approx 0{,}4639.$$

Итак, по совокупности указанных признаков можно предположить, что распределение случайной величины X является нормальным.

6. Функция плотности нормального распределения имеет вид

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

В качестве неизвестных параметров a и σ возьмем их точечные оценки $\overline{X}_B = 13{,}03$ и $\sigma_B(X) = 3{,}84$ соответственно. Тогда дифференциальная $f(x)$ и интегральная функции $F(x)$ предполагаемого нормального закона распределения примут вид

$$f(x)=\frac{1}{3{,}84\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-13{,}02)^2}{2\cdot 3{,}84^2}};\quad F(x)=\frac{1}{3{,}84\cdot\sqrt{2\pi}}\cdot\int\limits_{-\infty}^{x}e^{-\frac{(x-13{,}02)^2}{2\cdot 3{,}84^2}}dx.$$

7. Гипотезу о том, что генеральная совокупность, из которой извлечена выборка, распределена по предлагаемому нормальному закону, назовем нулевой $H_0: X\in N(a;\sigma)$. Проверим ее с помощью критерия согласия χ^2 Пирсона.

Согласно критерию Пирсона, сравниваются эмпирические n_i (наблюдаемые) и теоретические $n\cdot p_i$ (вычисленные в предложении нормального распределения) частоты. В качестве критерия проверки нулевой гипотезы принимается случайная величина

$$\chi^2_{набл}=\sum_{1}^{k}\frac{(n_i-n\cdot p_i)^2}{n\cdot p_i}.$$

По таблице критических точек распределения χ^2 по заданному уровню значимости α и числу степеней свободы $\nu=k-r-l$ ($k-$ число интервалов, $r-$ число параметров предполагаемого распределения СВ X) находится критическое значение $\chi^2_{кр}(\alpha;\nu)$ (Приложение D).

Если $\chi^2_{\text{набл}}\le\chi^2_{\text{кр}}$ (a,ν), то считается, что данный критерий не дает оснований для отклонений гипотезы при данном уровне значимости $\alpha=0{,}05$. В противном случае считается, что гипотеза не согласуется с экспериментальными данными и ее надо отвергнуть.

Если проверяется гипотеза о нормальном распределении, то вероятности p_i рассчитываются с помощью нормированной функции Лапласа $\Phi_0(x)$:

$$p_i=P(x_1<X<x_{i+1})=\Phi_0\left(\frac{x_{i+1}-\overline{X}_B}{\sigma_B(X)}\right)-\Phi_0\left(\frac{x_i-\overline{X}_B}{\sigma_B(X)}\right),$$

где выборочное среднее $\overline{X}_B=13{,}03$; выборочное среднее квадратическое отклонение $\sigma_B(X)=3{,}84$ (Приложение В).

Тогда

$$p_1=P(-\infty<x\le 4{,}2)=\Phi_0\left(\frac{4{,}2-13{,}03}{3{,}84}\right)-\Phi_0(-\infty)=$$

$$=\Phi_0(-2{,}3)-\Phi_0(-\infty)=-\Phi_0(2{,}3)+\Phi_0(+\infty)=-0{,}4893+0{,}5=0{,}0107;$$

$$p_2 = P(4,2 < x \le 6,6) = \Phi_0\left(\frac{6,6-13,03}{3,84}\right) - \Phi_0\left(\frac{4,2-13,03}{3,84}\right) =$$
$$= \Phi_0(-1,67) - \Phi_0(-2,3) = -\Phi_0(1,67) + \Phi_0(2,3) = -0,4525 + 0,4893 = 0,0368;$$

$$p_3 = P(6,6 < x \le 9) = \Phi_0\left(\frac{9-13,03}{3,84}\right) - \Phi_0\left(\frac{6,6-13,03}{3,84}\right) =$$
$$= \Phi_0(-1,05) + \Phi_0(1,67) = -\Phi_0(1,05) + \Phi_0(1,67) = -0,3531 + 0,4525 = 0,0994;$$

$$p_4 = P(9 < x \le 11,4) = \Phi_0\left(\frac{11,4-13,03}{3,84}\right) - \Phi_0\left(\frac{9-13,03}{3,84}\right) =$$
$$= \Phi_0(-0,42) + \Phi_0(1,05) = -\Phi_0(0,42) + \Phi_0(1,05) = -0,1228 + 0,3531 = 0,2303;$$

$$p_5 = P(11,4 < x \le 13,8) = \Phi_0\left(\frac{13,8-13,03}{3,84}\right) - \Phi_0\left(\frac{11,4-13,03}{3,84}\right) =$$
$$= \Phi_0(0,20) - \Phi_0(-0,42) = \Phi_0(0,20) + \Phi_0(0,42) = 0,0793 + 0,1228 = 0,2021;$$

$$p_6 = P(13,8 < x \le 16,2) = \Phi_0\left(\frac{16,2-13,03}{3,84}\right) - \Phi_0\left(\frac{13,8-13,03}{3,84}\right) =$$
$$= \Phi_0(0,83) - \Phi_0(0,20) = 0,2967 - 0,0793 = 0,2174;$$

$$p_7 = P(16,2 < x \le 18,6) = \Phi_0\left(\frac{18,6-13,03}{3,84}\right) - \Phi_0\left(\frac{16,2-13,03}{3,84}\right) =$$
$$= \Phi_0(1,45) - \Phi_0(0,83) = 0,4265 - 0,2967 = 0,1298;$$

$$p_8 = P(18,6 < x \le +\infty) = \Phi_0(+\infty) - \Phi_0\left(\frac{18,6-13,03}{3,84}\right) =$$
$$= \Phi_0(+\infty) - \Phi_0(1,45) = 0,5 - 0,4265 = 0,0735.$$

Полученные результаты запишем в таблицу 11.

Таблица 11 – ***Расчетная таблица для вычисления*** χ_B^2

Интервалы $(x_i; x_{i+1}]$	Частоты эмпирические n_i	Вероятности p_i	Теоретические частоты $n \cdot p_i$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
(–∞; 4,2]	1	0,0107	1,07	0,0046
(4,2; 6,6]	3	0,0368	3,68	0,1257
(6,6; 9]	14	0,0994	9,94	1,6583
(9; 11,4]	15	0,2303	23,03	2,7999
(11,4; 13,8]	21	0,2021	20,21	0,0309
(13,8; 16,2]	26	0,2174	21,74	0,8348
(16,2; 18,6]	12	0,1298	12,98	0,0740
(18,6; + ∞)	8	0,0735	7,35	0,0575
$\sum$	100	1.00	100	$\chi_B^2 = 5{,}5855$

Значение $\chi_B^2 = 5{,}5855$.

Количество интервалов $k = 8$, нормальное распределение имеет два параметра (математическое ожидание α и среднее квадратическое отклонение σ), поэтому $r = 2$, тогда число степеней свободы

$$\nu = k - r - 1 = 8 - 2 - 1 = 5.$$

В таблице критических точек распределения χ_B^2 (Приложение D) по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $\nu = 5$ найдем критическое значение $\chi_{кр}^2(0{,}05; 5) = 11{,}070$, очевидно, что $5{,}5855 < 11{,}070$. Так как $\chi_B^2 \le \chi_{кр}^2$, то нет оснований для отклонения нулевой гипотезы при заданном уровне значимости $\alpha = 0{,}05$.

Построим график эмпирической функции плотности нормального распределения $f^*(x)$ на фоне полигона относительных частот (рис. 16). Для этого из середины частичных интервалов восстановим перпендикуляры высотой, равной вероятностям p_i попадания случайной величины X в соответствующий частичный интервал. Полученные точки соединим плавной кривой.

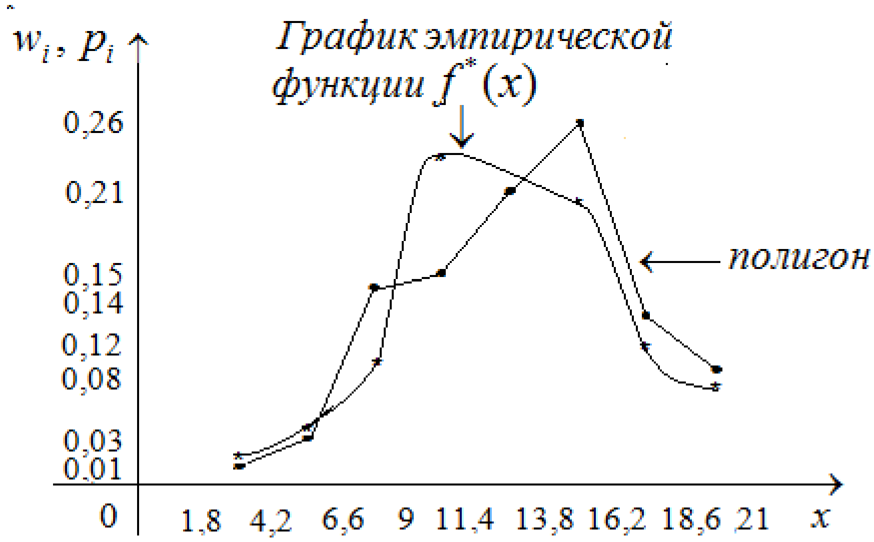


Рисунок 16 – График эмпирической функции плотности нормального распределения f(x) на фоне полигона относительных частот*

Сравнение полигона относительных частот и графика эмпирической функции плотности нормального распределения $f^*(x)$ (см. рис. 16) показывает, что построенная нормальная кривая достаточно близка к полигону, т. е. удовлетворительно сглаживает полигон.

8. Найдем интервальные оценки параметров нормального закона распределения. Для нахождения доверительного интервала, покрывающего математическое ожидание СВ X, найдем по таблицам квантилей распределения Стьюдента (Приложение С) по заданной доверительной вероятности $1-\alpha=\gamma=0{,}95$ и числу степеней свободы $\nu=n-1=100-1=99$ число $t_\gamma=1{,}984$.

С учетом того, что объем выборки большой ($n\geq 30$), вычислим предельную погрешность интервального оценивания по формуле

$$\varepsilon=t_\gamma\frac{\sigma_B(X)}{\sqrt{n}}=1{,}984\cdot\frac{3{,}84}{\sqrt{100}}\approx 0{,}762.$$

Запишем искомый доверительный интервал для математического ожидания a:

$$\overline{X}_B-\varepsilon<a<\overline{X}_B+\varepsilon,$$

$$13{,}032-0{,}762<a<13{,}032+0{,}762,$$

$$12{,}270<a<13{,}794.$$

Если будет произведено достаточно большое число выборок объема n случайной величины X из одной и той же генеральной совокупности, то в 95 % выборок доверительный интервал (12,270; 13,794) покроет математическое ожидание a; и только в 5 % выборок математическое ожидание может выйти за границы доверительного интервала.

Для нахождения доверительного интервала, покрывающего неизвестное среднее квадратическое отклонение σ с заданной вероятностью $1-\alpha=\gamma=0{,}95$, найдем по $\gamma=0{,}95$ и числу степеней свободы $\nu=n-1=100-1=99$ два числа: $\gamma_1=0{,}878$ и $\gamma_2=1{,}161$ (Приложение Е).

Искомый доверительный интервал равен следующему:

$$\gamma_1\cdot\sigma_B(X)<\sigma<\gamma_2\cdot\sigma_B(X),$$

$$0{,}878\cdot 3{,}84<\sigma<1{,}161\cdot 3{,}84,$$

$$3{,}372<\sigma<4{,}458.$$

Если будет произведено достаточно большое число выборок объема n СВ X из одной и той же генеральной совокупности, то в 95 % выборок доверительный интервал $(3{,}372;\,4{,}458)$ покроет среднее квадратическое отклонение σ, и только в 5 % среднее квадратическое отклонение σ может выйти за границы доверительного интервала.

9. Проведем *корреляционный и регрессионный анализ* случайной *двумерной величины* (X,Y) по заданной выборке.

9.1. Составим корреляционную таблицу. Как известно, для СВ X выбраны следующие интервалы: $(1{,}8;\,4{,}2]$, $(4{,}2;\,6{,}6]$, $(6{,}6;\,9]$, $(9;\,11{,}4]$, $(11{,}4;\,13{,}8]$, $(13{,}8;\,16{,}2]$, $(16{,}2;\,18{,}6]$, $(18{,}6;\,21]$.

Аналогично, для случайной величины Y выберем следующие интервалы ($h=2{,}4$): $(26{,}8;\,29{,}2]$, $(29{,}2;\,31{,}6]$, $(31{,}6;\,34]$, $(34;\,36{,}4]$, $(36{,}4;\,38{,}8]$, $(38{,}8;\,41{,}2]$, $(41{,}2;\,43{,}6]$, $(43{,}6;\,46]$, $(46;\,48{,}4]$.

Подсчитываем количество пар исходной выборки, попадающих в прямоугольники, образованные границами интервалов. Для этого принадлежность пары $(x_i;y_j)$ к определенному прямоугольнику отмечаем внутри этого прямоугольника в таблице с помощью одного из элементов «конвертика».

В результате получим таблицу 12.

Таблица 12 – Таблица для частот n_{xy} пар значений $(x_i; y_j)$

Интер-валы для Y	Интервалы для X							
	(1,8; 4,2]	(4,2; 6,6]	(6,6; 9]	(9; 11,4]	(11,4; 13,8]	(13,8; 16,2]	(16,2; 18,6]	(18,6; 21]
(26,8; 29,2]								• •
(29,2; 31,6]				•				• • •
(31,6; 34]		•			•	• •		•
(34; 36,4]				•	• • •		• •	
(36,4; 38,8]			• •	•				•
(38,8; 41,2]			• • • •			• • • •		•
(41,2; 43,6]				•	•			
(43,6; 46]		•	• •					
(46; 48,4]	•	•	•					

В окончательной корреляционной таблице 13 вместо интервалов для случайной величины X и случайной величины Y записываем середины интервалов и соответствующие частоты n_x, n_y.

Таблица 13 – Корреляционная таблица для случайной величины (X, Y)

Y \ X	3	5,4	7,8	10,2	12,6	15	17,4	19,8	n_y
28								**2**	2
30,4				**1**				**3**	4
32,8		**1**			**1**	**2**	**10**	**1**	15
35,2				**1**	**3**	**9**	**2**		15
37,6			**2**	**1**	**9**	**11**		**1**	24
40			**4**	**11**	**7**	**4**			27
42,4			**5**	**1**	**1**				7
44,8		**1**	**2**						3
47,2	**1**	**1**	**1**						3
n_x	1	3	14	15	21	26	12	8	100

9.2. Зная, что $\overline{X}_B = 13{,}032$; $\sigma_B(X) = 3{,}84$, аналогично, можно подсчитать, что $\overline{Y}_B = 37{,}528$; $\sigma_B(Y) = 4{,}00$.

Далее вычислим сначала выборочный корреляционный момент:

$$K_B = \frac{\sum_i x_i y_i n_{ij}}{n} - \overline{X}_B \cdot \overline{Y}_B.$$

$$K_B = \frac{1}{100}(28\cdot 19{,}8\cdot 2 + 30{,}4\cdot 10{,}2\cdot 1 + 30{,}4\cdot 19{,}8\cdot 3 + 32{,}8\cdot 5{,}4\cdot 1 +$$
$$+32{,}8\cdot 12{,}6\cdot 1 + 32{,}8\cdot 15\cdot 2 + 32{,}8\cdot 17{,}4\cdot 10 + 32{,}8\cdot 19{,}8\cdot 1 + 35{,}2\cdot 10{,}2\cdot 1 +$$
$$+35{,}2\cdot 12{,}6\cdot 3 + 35{,}2\cdot 15\cdot 9 + 35{,}2\cdot 17{,}4\cdot 2 + 37{,}6\cdot 7{,}8\cdot 2 + 37{,}6\cdot 10{,}2\cdot 1 +$$
$$+37{,}6\cdot 12{,}6\cdot 9 + 37{,}6\cdot 15\cdot 11 + 37{,}6\cdot 19{,}8\cdot 1 + 40\cdot 7{,}8\cdot 4 + 40\cdot 10{,}2\cdot 11 +$$
$$+40\cdot 12{,}6\cdot 7 + 40\cdot 15\cdot 4 + 40\cdot 19{,}8\cdot 1 + 42{,}4\cdot 7{,}8\cdot 5 + 42{,}4\cdot 10{,}2\cdot 1 +$$
$$+42{,}4\cdot 12{,}6\cdot 1 + 44{,}8\cdot 5{,}4\cdot 1 + 44{,}8\cdot 7{,}8\cdot 2 + 47{,}2\cdot 3\cdot 1 + 47{,}2\cdot 5{,}4\cdot 1 +$$
$$+47{,}2\cdot 7{,}8\cdot 1) - 13{,}032\cdot 37{,}528 = \frac{47786{,}4}{100} - 489{,}064896 = 11{,}200896.$$

Выборочный коэффициент корреляции:

$$r_B = \frac{K_B}{\sigma_B(X)\cdot\sigma_B(Y)},$$

$$r_B = -\frac{11{,}200896}{3{,}84\cdot 4{,}00} = -\frac{11{,}16}{15{,}36} = -0{,}729225 \approx -0{,}73.$$

Отрицательный знак выборочного коэффициента корреляции r_B показывает, что с увеличением значений случайной величины X эмпирические значения случайной величины Y в среднем убывают.

9.3. Проверим значимость полученного выборочного коэффициента корреляции, т. е. проверим нулевую гипотезу о том, что коэффициент корреляции r равен нулю ($H_0 : r = 0$) при альтернативной гипотезе $H_0 : r \neq 0$.

Вычислим статистику:

$$t_{набл} = \frac{r_B\cdot\sqrt{n-2}}{\sqrt{1-r_B^2}} = \frac{-0{,}73\cdot\sqrt{98}}{\sqrt{1-(-0{,}73)^2}} \approx -10{,}57.$$

Принятие гипотезы H_0 при уровне значимости $\alpha = 0{,}05$ означает, что выборочный коэффициент корреляции $r_в$ отличается от нуля с ошибкой 5 %.

Найдем по таблицам квантилей распределение Стьюдента по наиболее употребляемому уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $\nu = n - 2$ квантиль $t(1-\alpha; n-2) = t(0{,}95; 98) = 1{,}984$ (Приложение С).

Так как $|t_{набл}| = 10{,}57 > 1{,}984$, то нулевая гипотеза отвергается и коэффициент корреляции можно считать существенным, а связь между случайными величинами достоверной, т. е. выборочный коэффициент корреляции значимо отличается от нуля. Это означает, что между СВ X и СВ Y существует корреляционная зависимость.

10. Проведем регрессионный анализ.

10.1. Построим корреляционное поле. Изобразим результаты измерений $(x_i; y_j)$ в виде точек в декартовой системе координат (рис. 17).

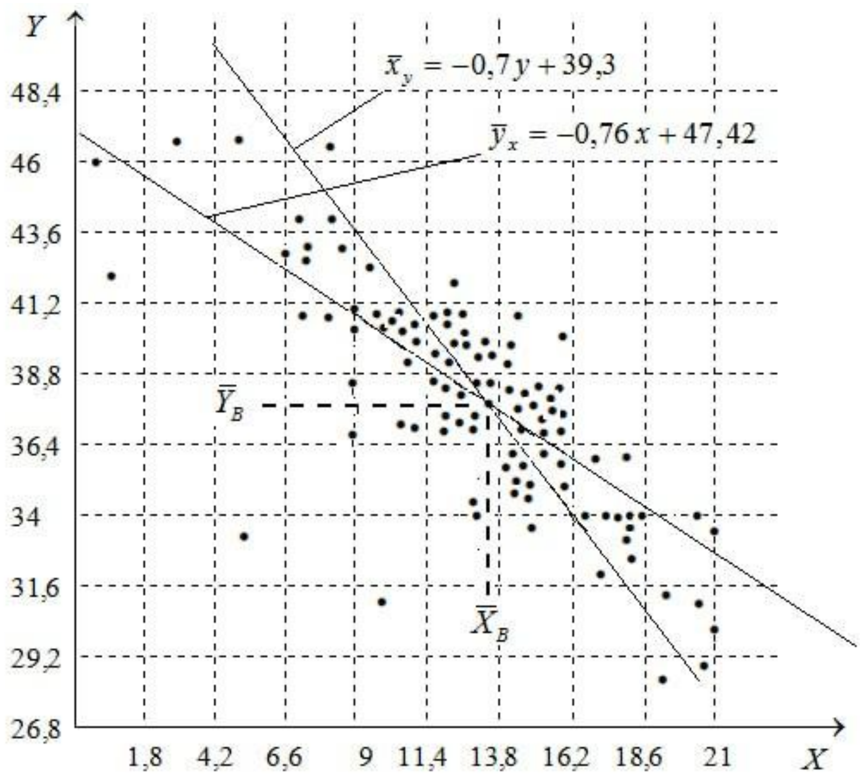


Рисунок 17 – Корреляционное поле

По виду корреляционного поля видно, что между X и Y имеется прямолинейная регрессионная зависимость.

10.2. Найдем выборочное уравнение регрессии «игрек на икс»:

$$\bar{y}_x = \bar{Y}_B + r_B \cdot \frac{\sigma_B(Y)}{\sigma_B(X)} \cdot (x - \bar{X}_B).$$

По условию задачи

$$\bar{y}_x = 37{,}528 - 0{,}729 \cdot \frac{4{,}00}{3{,}84}(x - 13{,}032),$$

$$\bar{y}_x = -0{,}759375x + 47{,}424175$$

или после округления параметров

$$\bar{y}_x = -0{,}76x + 47{,}42.$$

Аналогично находим выборочное уравнение регрессии «икс на игрек»:

$$\bar{x}_y = \overline{X}_B + r_B \cdot \frac{\sigma_B(X)}{\sigma_B(Y)} \cdot (y - \overline{Y}_B).$$

В нашем случае

$$\bar{x}_y = 13{,}032 - 0{,}729 \cdot \frac{3{,}84}{4{,}00}(y - 37{,}528),$$

$$\bar{x}_y = -0{,}69984y + 39{,}29559552$$

или

$$\bar{x}_y = -0{,}7y + 39{,}3.$$

Контроль вычислений:

$$(-0{,}759375) \cdot (-0{,}69984) = 0{,}531441 = r_B^2.$$

ВОПРОСЫ И УПРАЖНЕНИЯ ДЛЯ САМОПРОВЕРКИ

Теория вероятностей

1. Упростить выражение $\overline{\overline{A}\cap\overline{B}}$.

2. Пусть A и B – два события. Найти выражение для множества тех исходов, при которых происходит только одно из событий A и B.

3. Совместны ли события A и $\overline{A\cup B}$?

4. Каким множеством достаточно дополнить пару A, B несовместных событий, чтобы получилась полная группа событий?

5. При каком условии имеет место теорема сложения:

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

6. Известно, что наступление события A влечет наступление события B. Поставить знак неравенства между $P(A)$ и $P(B)$.

7. Всегда ли верно, что $P(A/B)=P(A)\cdot P(B)$?

8. Следует ли из независимости событий в совокупности их попарная независимость?

9. Следует ли из попарной независимости событий их независимость в совокупности?

10. Известно, что A и B – независимые события. Верно ли, что события $\overline{A}$ и $\overline{B}$ также независимы?

11. Написать формулу вероятности того, что произойдет событие A, при условии, что произошло событие B.

12. Пусть события A и B являются независимыми. Чему равна $P(A/B)$?

13. Написать формулу полной вероятности.

14. Написать формулу Байеса.

15. Пусть проведено n независимых испытаний Бернулли, вероятность успеха в единичном испытании равна p. Чему равна вероятность того, что произошло ровно k успехов?

16. Что такое функция распределения случайной величины X?

17. Могут ли две различные случайные величины иметь одинаковые функции распределения?

18. Пусть функция распределения есть

$$F(x)=\begin{cases} 0, если & x\le 0. \\ x/2, если & 0<x<1. \\ 1, если & x>1. \end{cases}$$

Найти функцию плотности, построить графики функций.

19. Пусть $f(x)$ есть плотность распределения случайной величины X. Чему равна $F(x)$ – функция распределения случайной величины X?

20. Пусть $f(x)$, $-\infty<x<+\infty$, есть плотность распределения случайной величины X. Чему равна $P(a<X<b)$?

21. Пусть X – дискретная случайная величина. $P(X=x_k)=p_k$, $k=1,2,\ldots$. Чему равно $M(X)$?

22. Пусть X – абсолютно непрерывная случайная величина, $f(x)$ – плотность ее распределения. Чему равно $M(X)$?

23. Может ли случайная величина не иметь математического ожидания?

24. Может ли случайная величина не иметь дисперсии?

25. Пусть существует $M(X)=a<\infty$. Чему равно $M(2X+3)$?

26. Что такое дисперсия случайной величины?

27. Пусть существует $D(X)=\sigma^2<+\infty$. Чему равна $D(3X+2)$?

28. Всегда ли дисперсия суммы двух случайных величин равна сумме их дисперсий (известно, что дисперсии существуют)?

29. Что такое ковариация случайных величин X_1 и X_2?

30. Пусть $f(x)=\dfrac{1}{\sqrt{8\pi}}e^{-\frac{(x-1)^2}{8}}$ – плотность распределения случайной величины X. Найти численные значения математического ожидания и дисперсии.

31. Чему равно математическое ожидание случайной величины, равномерно распределенной в интервале $(-1;1)$?

32. Записать неравенство Чебышева.

33. К чему сходится при $n\to\infty$ последовательность случайных величин $Y_n=\dfrac{1}{n}\sum_{k=1}^{n}X_к$ в законе больших чисел?

Математическая статистика

1. Что является объектом изучения математической статистики?

2. Что называется генеральной и выборочной совокупностями?

3. Поясните термин «репрезентативная выборка». Каким образом обеспечивается такая выборка?

4. Как из непрерывного статистического ряда получить дискретный ряд, соответствующий данному непрерывному ряду?

5. Как выбирается оптимальная длина n интервала и начало первого интервала?

7. Как называются графики, применяющиеся для изображения вариационных рядов: а) дискретных, б) интервальных?

7. В чем сходство и различие между понятиями «закон распределения случайных величин» в теории вероятностей и «вариационный ряд» в математической статистике?

8. Что называется эмпирической функцией распределения $F^*(x)$?

9. Перечислите основные свойства эмпирической функции распределения.

10. Назовите числовые характеристики выборки.

11. Напишите формулы для выборочной средней $\overline{X}_B$, исправленной выборочной дисперсией S^2, исправленного выборочного среднего квадратического отклонения S, медианы M_e, моды M_0.

12. Перечислите свойства выборочной средней.

13. Перечислите свойства выборочной дисперсии.

14. Какая оценка параметра называется точечной?

15. Перечислите требования, предъявляемые к точечным оценкам неизвестных параметров распределения.

16. Какие числовые характеристики выборки оценивают $M(X)$?

17. Какие числовые характеристики выборки оценивают $D(X)$?

18. Что является несмещенной оценкой генеральной дисперсии. Запишите формулу.

19. Пусть θ_1^* и θ_2^* – несмещенные оценки параметра θ, причем $D(\theta_1^*) < D(\theta_2^*)$; $M(\theta_1^*) < M(\theta_2^*) = \theta$. Какая из оценок является эффективной?

20. Пусть θ^* – точечная оценка параметра θ. Что характеризует число $\varepsilon = \left|\theta - \theta^*\right|$?

21. Какие теоретические моменты приравниваются соответствующим эмпирическим моментам для нахождения параметров нормального закона распределения?

23. Что называется доверительной вероятностью, доверительным интервалом (интервальной оценкой) параметра распределения?

24. Каков статистический смысл доверительной вероятности? Что показывает доверительная вероятность $\gamma = 0{,}99$?

25. Какая случайная величина Z называется нормированным отклонением средней арифметической $\overline{X}_B$ от математического ожидания? Приведите формулу Z и укажите ее закон распределения.

26. По таблице Стьюдента найдите значение t_γ для уровня значимости $\alpha = 0{,}01$; $\alpha = 0{,}05$; и числа степеней свободы $k = 40$.

27. Какая случайная величина называется нормированным отклонением случайной величины X от математического ожидания?

28. Приведите формулы случайной величины χ^2.

29. Как изменяется длина доверительного интервала при увеличении (уменьшении): а) объема выборки n; б) доверительной вероятности γ?

30. Приведите формулы доверительных интервалов для математического ожидания при известном и неизвестном значении дисперсии.

31. Приведите формулу для нахождения доверительного интервала для дисперсии нормально распределенной случайной величины X при неизвестном математическом ожидании.

32. Приведите формулу для нахождения доверительного интервала для среднего квадратического отклонения нормально распределенной случайной величины X при неизвестном математическом ожидании.

33. Дайте понятие статистической гипотезы.

34. Какие статистические гипотезы называют нулевой и альтернативной?

35. Поясните смысл ошибок первого и второго рода, возникающих при проверке гипотез.

36. Дайте определение статистического критерия. Приведите примеры статистических критериев.

37. Какая совокупность критерия K называется критической областью и областью допустимых значений?

38. Сформулируйте общий принцип проверки гипотез о параметрах распределения с помощью статистических критериев.

39. На основании каких признаков производится предварительный выбор теоретического закона распределения по экспериментальным данным?

40. Сформулируйте алгоритм применения критерия согласия χ^2 Пирсона.

41. В чем состоит различие между функциональной и статистической зависимостью двух случайных величин X и Y? Приведите примеры.

42. В чем состоит основная задача корреляционного анализа?

43. Сформулируйте и запишите математически свойства коэффициента корреляции

44. Приведите графическую интерпретацию значений выборочного коэффициента корреляции для случаев: а) $r_B = 1$; б) $r_B = -1$; в) $r_B = 0$; г) $r_B > 0$; д) $r_B < 0$.

45. Как строится корреляционное поле? На основании каких признаков можно судить об общем виде эмпирической функции регрессии? Приведите пример.

46. Опишите метод наименьших квадратов для линейной регрессии.

ПРАКТИЧЕСКАЯ РАБОТА № 1

Студент выполняет задачи по варианту, номер которого совпадает с последней цифрой его номера зачетной книжки.

Вариант	Номер задачи					
1	1	11	21	31	41	51
2	2	12	22	32	42	52
3	3	13	23	33	43	53
4	4	14	24	34	44	54
5	5	15	25	35	45	55
6	6	16	26	36	46	56
7	7	17	27	37	47	57
8	8	18	28	38	48	58
9	9	19	29	39	49	59
0	10	20	30	40	50	60

Задание 1

1. Из пяти букв разрезной азбуки составлено слово «КНИГА». Ребенок, не умеющий читать, рассыпал буквы и затем собрал их в произвольном порядке. Найти вероятность того, что у него снова получится слово «КНИГА».
2. Имеется два ящика изделий, причем в одном ящике все изделия доброкачественны, а во втором только половина. Изделие, взятое наудачу из выбранного ящика, оказалось доброкачественным. Найти, на сколько отличаются вероятности того, что изделие принадлежит первому и второму ящику, если количество изделий в ящиках одинаково.
3. Из контейнера, содержащего одинаковое число деталей четырех предприятий, взяли на проверку одну деталь. Какова вероятность обнаружения бракованной продукции, если продукция двух предприятий содержит по ¾ бракованных деталей, а вся продукция остальных предприятий доброкачественна?
4. Агрегат имеет четыре двигателя и способен функционировать, если работают, по крайней мере, два из них. Вероятность выйти из строя первому двигателю – 0,01; второму – 0,02; третьему – 0,03 и четвертому – 0,04. Какова вероятность выйти из строя агрегату?
5. В двух ящиках содержится по 20 деталей, из которых в первом ящике – 16, а во втором – 10 стандартных. Из первого ящика извлекается и перекладывается во второй 1 деталь. Определить вероят-

ность того, что наудачу извлеченная после этого деталь из этого ящика будет стандартной.

6. Из 20 выбранных деталей 5 изготовлено на станке № 1, 10 изготовлено на станке № 2, а остальные – на станке № 3. Вероятность изготовления стандартной детали на станке № 1 равно 0,96, на станке № 2 – 0,98. Найти вероятность изготовления стандартной детали на третьем станке, если вероятность при случайном отборе получить стандартную деталь из указанных 20 равна 0,97.

7. Из 20 стрелков 7 попадает в цель с вероятностью 0,6; 8 – с вероятностью 0,5 и 5 – с вероятностью 0,7. Наудачу выбранный стрелок произвел выстрел, поразив цель. К какой из групп вероятней всего принадлежал этот стрелок?

8. Три партии деталей содержат соответственно по 1/2, 2/3 и 1/2 бракованных. Из каждой партии взято по одной детали, причем обнаружено две бракованных. Определить вероятность того, что доброкачественная деталь принадлежит третьей партии.

9. Из партии в 4 детали наудачу взята одна, оказавшаяся доброкачественной. Количество доброкачественных деталей равновозможно любое. Какое предположение о количестве бракованных деталей наиболее вероятно и какова его вероятность?

10. Вероятность изготовления на автоматическом станке бракованной детали равна 0,1. Какова вероятность того, что из 4 деталей бракованных окажется не более 2?

Задание 2

11. При сборке в среднем 2 % механизмов оказываются с дефектами. Контролер проверяет взятые наудачу 6 механизмов. Определить вероятность того, что среди них с дефектами окажется не более 1 механизма.

12. При установившемся технологическом процессе автомат производит 0,75 числа деталей первого сорта и 0,25 – второго. Установить, что является более вероятным – получить 3 первосортных детали среди 5 наудачу отобранных или 4 первосортных среди 6 наудачу отобранных.

13. Сколько раз надо подбрасывать монету, чтобы вероятность появления герба три раза равнялась 0,25?

14. Вероятность выигрыша по билету равна 0,2. Сколько нужно приобрести билетов, чтобы наивероятнейшее число выигрышей билетов равнялось 15?

15. Среди изделий, произведенных на станке-автомате, в среднем бывает 905 изделий первого сорта. Какова вероятность того, что среди 5 наудачу выбранных изделий будет не менее 4 первого сорта?

16. Для прядения смешаны поровну белый и окрашенный хлопок. Какова вероятность среди 8 случайно отобранных волокон смеси обнаружить менее 4 окрашенных?

17. Что вероятнее, выиграть у равносильного противника не менее 3 партий из 4 или не менее 5 из 8?

18. Среди изделий, изготовленных вручную, бывает в среднем 4 % брака. Какова вероятность того, что среди взятых наудачу пяти изделий будет 40 % бракованных?

19. Каждый из четырех станков в течение 5 часов работы останавливается несколько раз и всего в сумме стоит 1 час, причем остановка его в любой момент времени равновероятна. Найти вероятность того, что в данный момент времени будут работать два станка.

20. При данном технологическом процессе 80 % всей продукции оказывается продукцией высшего сорта. Определить наивероятнейшее число изделий высшего сорта в партии из 200 изделий и его вероятность.

Задание 3

21. В результате проверки качества приготовленного для посева зерна было установлено, что 80 % зерен всхожие. Определить вероятность того, что среди отобранных и высаженных 100 зерен прорастет не менее 70.

22. Вероятность того, что пара обуви, взятая наудачу из изготовленной партии, окажется первого сорта, равна 0,7. Определить вероятность того, что среди 2100 пар, поступающих на контроль, число пар первосортной обуви окажется не менее 1000 и не более 1500?

23. Наиболее вероятная частота проявления события при независимых испытаниях равна 50, а дисперсия – 25. Оценить абсолютную величину отклонения частности появления события от вероятности появления с вероятностью 0,9973.

24. Вероятность рождения мальчиков $p = 0{,}515$. Какова вероятность, что из 1000 рождающихся детей мальчиков будет 520?

25. В ОТК с конвейера для проверки поступила партия изделий в количестве 600 штук. Какова вероятность того, что среди этих изделий имеется 500 изделий 1-го сорта, если известно, что с конвейера в среднем поступает 85 % продукции 1-го сорта?

26. Книга в 500 страниц содержит 50 опечаток. Оценить вероятность того, что на случайно выбранной странице не менее 3 опечаток.

27. Вероятность изготовления бракованного изделия равна 0,001. Вычислить вероятность того, что контролер, проверяющий партию в 200 изделий, обнаружит число бракованных более 2.

28. Вероятность изготовления консервной банки с недостаточной герметизацией равна 0,02. Среди скольких банок, отобранных случайным образом, можно с вероятностью 0,9 ожидать отсутствие бракованных?

29. Всхожесть семян некоторого растения составляет 70 %. Какова вероятность того, что из 100 посеянных семян взойдет не менее 80?

30. Всхожесть семян некоторого растения составляет 85 %. Какова вероятность того, что из 100 посеянных семян взойдет не менее 70?

Задание 4

В задачах № 31–40 непрерывная случайная величина X задана дифференциальной функцией:

$$f(x)=\begin{cases} 0, & \text{при} \quad x\le 0; \\ ax^2, & \text{при} \quad 0<x\le b; \\ 0, & \text{при} \quad x>b. \end{cases}$$

Требуется найти:

а) значение параметра a;

б) интегральную функцию $F(x)$;

в) математическое ожидание и дисперсию X;

г) вероятность того, что X примет значение, заключенное в интервале $(0;1/3)$;

д) построить графики функций $f(x)$ и $F(x)$, если значение параметра b следующие:

31. $1/2$;
32. $2/3$;
33. $3/2$;
34. $4/5$;
35. $5/6$;

36. 6/7;
37. 7/8;
38. 8/9;
39. 9/10;
40. 10/11.

Задание 5

В задачах № 41–50 найти математическое ожидание и дисперсию случайной величины $Y = \alpha X + \beta$ для заданного ряда распределения случайной величины X и значений параметров α и β.

41.

x_i	1	2	3	4
p_i	0,1	0,2	0,3	0,4

; $\alpha = -2, \quad \beta = 3.$

42.

x_i	5	6	7	8
p_i	0,1	0,2	0,3	0,4

; $\alpha = 3, \quad \beta = -4.$

43.

x_i	9	10	11	12
p_i	0,1	0,4	0,4	0,1

; $\alpha = -0{,}2, \quad \beta = 0{,}3.$

44.

x_i	13	14	15	16
p_i	0,4	0,2	0,2	0,2

; $\alpha = -0{,}3, \quad \beta = 0{,}4.$

45.

x_i	2	3	4	5
p_i	0,2	0,2	0,4	0,2

; $\alpha = -0{,}4, \quad \beta = 0{,}5.$

46.

x_i	3	4	5	6
p_i	0,3	0,4	0,2	0,1

; $\alpha = -4, \quad \beta = 2.$

47.

x_i	4	5	6	7
p_i	0,1	0,3	0,4	0,2

; $\alpha = 4, \quad \beta = -3.$

48.

x_i	1	2	3	4
p_i	0,1	0,2	0,3	0,4

; $\alpha = -0{,}5, \quad \beta = 2.$

49.

x_i	14	15	16	17
p_i	0,5	0,3	0,1	0,1

; $\alpha = -0{,}1, \quad \beta = 4.$

50.

x_i	4	5	6	7
p_i	0,6	0,2	0,1	0,1

; $\alpha = 3, \quad \beta = -0{,}2.$

Задание 6

По данным задач № 51–60 необходимо:

1) начертить графики: полигон, гистограмму и кумуляту;

2) вычислить среднюю арифметическую, дисперсию, среднее квадратическое отклонение;

3) рассчитать и построить на графиках гистограммы, теоретические нормальные кривые $f(x)$ и $F(x)$;

4) определить вероятность $P(x_1 < x < x_2)$;

5) произвести оценку степени близости теоретического распределения эмпирическому ряду с помощью критерия согласия Пирсона.

51. Распределение затрат на 100 руб. продукции по предприятиям хлопчатобумажной промышленности.

Интервал	96,3–97,3	97,3–98,3	98,3–99,3	99,3–100,3	100,3–101,3	101,3–102,3	102,3–103,3	103,3–104,3	104,3–105,3
Кол-во пред-приятий	3	3	12	12	24	18	17	4	2

$$P(98{,}4 < x < 101{,}2) = ?$$

52. Распределение объема товарной продукции на 1 кв. м производственной площади (млн руб.).

Интервал	0–0,2	0,2–0,4	0,4–0,6	0,6–0,8	0,8–1,0	1,0–1,2	1,2–1,4
Кол-во предприятий	7	11	26	24	17	10	5

$$P(0{,}35 < x < 1{,}03) = ?$$

53. Распределение объема основных фондов (млрд руб.) предприятий трикотажной промышленности.

Интервал	2,0–2,14	2,14–2,28	2,28–2,42	2,42–2,56	2,56–2,70	2,70–2,84	2,84–2,98	2,98–3,12
Кол-во предприятий	5	12	18	40	15	5	3	2

$$P(2{,}29 < x < 2{,}97) = ?$$

54. Распределение оплаты труда на одном из малых предприятий за месяц.

Интервал	260–274	274–288	288–302	302–316	316–330	330–344	344–358	358–372
Кол-во работ	1	2	9	25	28	18	12	5

$$P(275 < x < 342) = ?$$

55. Распределение производственных площадей (тыс. м^2) предприятий текстильной промышленности.

Интервал	1,03–1,37	1,37–1,71	1,71–2,05	2,05–2,39	2,39–2,73	2,73–3,07	3,07–3,41	3,41–3,75	3,75–4,09
Кол-во предприятий	2	12	15	17	23	12	14	3	2

$$P(2{,}31 < x < 2{,}99) = ?$$

56. Распределение средних удоев молока в фермерском хозяйстве (литров) от одной коровы за день.

Интервал	7,5–10,5	10,5–13,5	13,5–16,5	16,5–19,5	19,5–22,5	22,5–25,5	25,5–28,5	28,5–31,5	31,5–34,5
Кол-во коров	2	6	10	17	33	11	9	7	5

$$P(15{,}4 < x < 28{,}4) = ?$$

57. Распределение средней урожайности (ц/га) в фермерских хозяйствах области.

Интервал	9,8–13,2	13,2–16,6	16,6–20,0	20,0–23,4	23,4–26,8	26,8–30,2	30,2–33,6	33,6–37,0
Кол-во хозяйств	2	4	8	30	29	14	10	3

$$P(23{,}4 < x < 32{,}9) = ?$$

58. Распределение декадной выручки от реализации (млн руб.) в коммерческих торговых палатках микрорайона.

Интервал	5–10	10–15	15–20	20–25	25–30	30–35	35–40
Кол-во палаток	3	10	19	30	20	14	4

$$P(23{,}4 < x < 32{,}9) = ?$$

59. Распределение индекса цен по группе продовольственных товаров (%).

Интервал	98,4–99,6	99,6–100,8	100,8–102,0	102,0–103,2	103,2–104,4	104,4–105,6	105,6–106,8	106,8–108,0	108,0-109,2
Кол-во детей	3	7	18	19	20	17	13	2	1

$$P(x > 100) = ?$$

60. Распределение длины резьбы на муфте вентилей (мм).

Интервал	51,48–51,72	51,72–51,96	51,96–52,20	52,20–52,44	52,44–52,68	52,68–52,92	52,92–53,16	53,16–53,40
Кол-во детей	3	4	18	19	288	21	6	1

$$P(51{,}9 < x < 52{,}4) = ?$$

ПРАКТИЧЕСКАЯ РАБОТА № 2

Даны результаты 100 наблюдений над некоторыми случайными величинами X и Y.

Требуется для каждой случайной величины X и Y:

1. Построить интервальный и дискретный статистический ряды распределения частот и относительных частот.
2. Построить гистограмму и полигон относительных частот.
3. Найти эмпирическую функцию распределения и построить ее график.
4. Вычислить числовые характеристики выборки: выборочную среднюю, выборочную дисперсию, выборочное среднее квадратическое отклонение, выборочные коэффициенты асимметрии и эксцесса.
5. Определить закон распределения наблюдаемой случайной величины, исходя из механизма ее образования, по виду гистограммы и полигона относительных частот и по значениям выборочных коэффициентов асимметрии и эксцесса.
6. Найти точечные оценки параметров нормального закона распределения, предполагая, что наблюдаемая случайная величина распределена по нормальному закону, и записать функцию плотности распределения вероятностей.
7. Проверить с помощью критерия согласия Пирсона гипотезу о том, что выборка извлечена из генеральной совокупности с предполагаемым нормальным законом распределения.
8. В случае принятия гипотезы найти интервальные оценки параметров нормального закона распределения (доверительную вероятность принять равной $\gamma = 0{,}95$)
9. Провести корреляционный анализ случайной **двумерной величины** (X, Y):

9.1. Составить корреляционную таблицу.

9.2. Найти выборочный коэффициент корреляции.

9.3. Проверить значимость выборочного коэффициента корреляции r_B при $\lambda = 0{,}05$ $(H_0 : p = 0)$, при альтернативной гипотезе H_λ: $p \neq 0$.

10. Провести регрессионный анализ:

10.1. Построить корреляционное поле и по характеру расположения точек на нем подобрать общий вид функции регрессии.

10.2. Найти эмпирические функции регрессии Y на X, X на Y и построить их графики.

ВАРИАНТ № 1

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	17	34	**21**	16	38	**41**	13	37	**61**	11	39	**81**	15	37
2	9	42	**22**	20	41	**42**	15	39	**62**	7	43	**82**	7	44
3	12	35	**23**	15	33	**43**	3	47	**63**	14	37	**83**	17	32
4	11	37	**24**	9	37	**44**	14	38	**64**	13	34	**84**	15	36
5	11	40	**25**	20	38	**45**	15	36	**65**	11	36	**85**	13	37
6	14	37	**26**	11	40	**46**	11	39	**66**	20	30	**86**	17	34
7	21	32	**27**	11	39	**47**	13	35	**67**	12	42	**87**	12	39
8	18	34	**28**	18	33	**48**	14	37	**68**	8	44	**88**	9	41
9	14	37	**29**	9	41	**49**	12	38	**69**	12	37	**89**	17	36
10	6	46	**30**	10	31	**50**	16	37	**70**	20	29	**90**	15	37
11	16	40	**31**	16	35	**51**	14	36	**71**	14	40	**91**	8	43
12	15	38	**32**	13	40	**52**	14	35	**72**	15	34	**92**	13	38
13	13	39	**33**	18	33	**53**	19	28	**73**	11	41	**93**	12	40
14	13	36	**34**	14	37	**54**	15	35	**74**	10	39	**94**	16	36
15	18	32	**35**	13	39	**55**	11	41	**75**	9	38	**95**	19	31
16	13	37	**36**	17	33	**56**	5	47	**76**	10	42	**96**	10	41
17	11	39	**37**	18	36	**57**	13	38	**77**	8	41	**97**	18	34
18	5	33	**38**	12	39	**58**	8	47	**78**	18	34	**98**	12	40
19	21	30	**39**	9	42	**59**	14	39	**79**	7	41	**99**	10	41
20	14	36	**40**	13	38	**60**	16	36	**80**	7	43	**100**	13	37

ВАРИАНТ № 2

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	15	37	**21**	11	39	**41**	13	37	**61**	16	38	**81**	17	34
2	7	44	**22**	7	43	**42**	15	39	**62**	20	41	**82**	9	42
3	17	32	**23**	14	37	**43**	3	47	**63**	15	33	**83**	12	35
4	15	36	**24**	13	34	**44**	14	38	**64**	9	37	**84**	11	37
5	13	37	**25**	11	36	**45**	15	36	**65**	20	38	**85**	11	40
6	17	34	**26**	20	30	**46**	11	39	**66**	11	40	**86**	14	37
7	12	39	**27**	12	42	**47**	13	35	**67**	11	39	**87**	21	32
8	9	41	**28**	8	44	**48**	14	37	**68**	18	33	**88**	18	34
9	17	36	**29**	12	37	**49**	12	38	**69**	9	41	**89**	14	37
10	15	37	**30**	20	29	**50**	16	37	**70**	10	31	**90**	6	46
11	8	43	**31**	14	40	**51**	14	36	**71**	16	35	**91**	16	40
12	13	38	**32**	15	34	**52**	14	35	**72**	13	40	**92**	15	38
13	12	40	**33**	11	41	**53**	19	28	**73**	18	33	**93**	13	39
14	16	36	**34**	10	39	**54**	15	35	**74**	14	37	**94**	13	36
15	19	31	**35**	9	38	**55**	11	41	**75**	13	39	**95**	18	32
16	10	41	**36**	10	42	**56**	5	47	**76**	17	33	**96**	13	37
17	18	34	**37**	8	41	**57**	13	38	**77**	18	36	**97**	11	39
18	12	40	**38**	18	34	**58**	8	47	**78**	12	39	**98**	5	33
19	10	41	**39**	7	41	**59**	14	39	**79**	9	42	**99**	21	30
20	13	37	**40**	7	43	**60**	16	36	**80**	13	38	**100**	14	36

ВАРИАНТ № 3

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	16	40	**21**	16	35	**41**	14	36	**61**	14	40	**81**	8	43
2	15	38	**22**	13	40	**42**	14	35	**62**	15	34	**82**	13	38
3	13	39	**23**	18	33	**43**	19	28	**63**	11	41	**83**	12	40
4	13	36	**24**	14	37	**44**	15	35	**64**	10	39	**84**	16	36
5	18	32	**25**	13	39	**45**	11	41	**65**	9	38	**85**	19	31
6	13	37	**26**	17	33	**46**	5	47	**66**	10	42	**86**	10	41
7	11	39	**27**	18	36	**47**	13	38	**67**	8	41	**87**	18	34
8	5	33	**28**	12	39	**48**	8	47	**68**	18	34	**88**	12	40
9	21	30	**29**	9	42	**49**	14	39	**69**	7	41	**89**	10	41
10	14	36	**30**	13	38	**50**	16	36	**70**	7	43	**90**	13	37
11	17	34	**31**	16	38	**51**	13	37	**71**	11	39	**91**	15	37
12	9	42	**32**	20	41	**52**	15	39	**72**	7	43	**92**	7	44
13	12	35	**33**	15	33	**53**	3	47	**73**	14	37	**93**	17	32
14	11	37	**34**	9	37	**54**	14	38	**74**	13	34	**94**	15	36
15	11	40	**35**	20	38	**55**	15	36	**75**	11	36	**95**	13	37
16	14	37	**36**	11	40	**56**	11	39	**76**	20	30	**96**	17	34
17	21	32	**37**	11	39	**57**	13	35	**77**	12	42	**97**	12	39
18	18	34	**38**	18	33	**58**	14	37	**78**	8	44	**98**	9	41
19	14	37	**39**	9	41	**59**	12	38	**79**	12	37	**99**	17	36
20	6	46	**40**	10	31	**60**	16	37	**80**	20	29	**100**	15	37

ВАРИАНТ № 4

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	13	37	**21**	17	33	**41**	5	47	**61**	10	42	**81**	10	41
2	11	39	**22**	18	36	**42**	13	38	**62**	8	41	**82**	18	34
3	5	33	**23**	12	39	**43**	8	47	**63**	18	34	**83**	12	40
4	21	30	**24**	9	42	**44**	14	39	**64**	7	41	**84**	10	41
5	14	36	**25**	13	38	**45**	16	36	**65**	7	43	**85**	13	37
6	16	40	**26**	16	35	**46**	14	36	**66**	14	40	**86**	8	43
7	15	38	**27**	13	40	**47**	14	35	**67**	15	34	**87**	13	38
8	13	39	**28**	18	33	**48**	19	28	**68**	11	41	**88**	12	40
9	13	36	**29**	14	37	**49**	15	35	**69**	10	39	**89**	16	36
10	18	32	**30**	13	39	**50**	11	41	**70**	9	38	**90**	19	31
11	17	34	**31**	16	38	**51**	13	37	**71**	11	39	**91**	15	37
12	9	42	**32**	20	41	**52**	15	39	**72**	7	43	**92**	7	44
13	12	35	**33**	15	33	**53**	3	47	**73**	14	37	**93**	17	32
14	11	37	**34**	9	37	**54**	14	38	**74**	13	34	**94**	15	36
15	11	40	**35**	20	38	**55**	15	36	**75**	11	36	**95**	13	37
16	14	37	**36**	11	40	**56**	11	39	**76**	20	30	**96**	17	34
17	21	32	**37**	11	39	**57**	13	35	**77**	12	42	**97**	12	39
18	18	34	**38**	18	33	**58**	14	37	**78**	8	44	**98**	9	41
19	14	37	**39**	9	41	**59**	12	38	**79**	12	37	**99**	17	36
20	6	46	**40**	10	31	**60**	16	37	**80**	20	29	**100**	15	37

ВАРИАНТ № 5

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	14	36	**21**	16	34	**41**	6	48	**61**	9	41	**81**	11	42
2	10	38	**22**	17	36	**42**	13	37	**62**	7	42	**82**	20	34
3	5	33	**23**	12	39	**43**	8	47	**63**	18	34	**83**	12	40
4	21	30	**24**	9	42	**44**	14	39	**64**	7	41	**84**	10	41
5	14	36	**25**	13	38	**45**	16	36	**65**	7	43	**85**	13	37
6	16	40	**26**	16	35	**46**	14	36	**66**	14	40	**86**	8	43
7	15	38	**27**	13	40	**47**	14	35	**67**	15	34	**87**	13	38
8	13	39	**28**	18	33	**48**	19	28	**68**	11	41	**88**	12	40
9	13	36	**29**	14	37	**49**	15	35	**69**	10	39	**89**	16	36
10	18	32	**30**	13	39	**50**	11	41	**70**	9	38	**90**	19	31
11	17	34	**31**	16	38	**51**	13	37	**71**	11	39	**91**	15	37
12	9	42	**32**	20	41	**52**	15	39	**72**	7	43	**92**	7	44
13	12	35	**33**	15	33	**53**	3	47	**73**	14	37	**93**	17	32
14	11	37	**34**	9	37	**54**	14	38	**74**	13	34	**94**	15	36
15	11	40	**35**	20	38	**55**	15	36	**75**	11	36	**95**	13	37
16	14	37	**36**	11	40	**56**	11	39	**76**	20	30	**96**	17	34
17	21	32	**37**	11	39	**57**	13	35	**77**	12	42	**97**	12	39
18	18	34	**38**	18	33	**58**	14	37	**78**	8	44	**98**	9	41
19	14	37	**39**	9	41	**59**	12	38	**79**	12	37	**99**	17	36
20	6	46	**40**	10	31	**60**	16	37	**80**	20	29	**100**	15	37

ВАРИАНТ № 6

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	15	35	**21**	16	35	**41**	7	48	**61**	10	41	**81**	12	41
2	9	37	**22**	18	36	**42**	13	38	**62**	8	43	**82**	21	33
3	5	33	**23**	12	39	**43**	8	47	**63**	18	34	**83**	12	40
4	21	30	**24**	9	42	**44**	14	39	**64**	7	41	**84**	10	41
5	14	36	**25**	13	38	**45**	16	36	**65**	7	43	**85**	13	37
6	16	40	**26**	16	35	**46**	14	36	**66**	14	40	**86**	8	43
7	15	38	**27**	13	40	**47**	14	35	**67**	15	34	**87**	13	38
8	13	39	**28**	18	33	**48**	19	28	**68**	11	41	**88**	12	40
9	13	36	**29**	14	37	**49**	15	35	**69**	10	39	**89**	16	36
10	18	32	**30**	13	39	**50**	11	41	**70**	9	38	**90**	19	31
11	17	34	**31**	16	38	**51**	13	37	**71**	11	39	**91**	15	37
12	9	42	**32**	20	41	**52**	15	39	**72**	7	43	**92**	7	44
13	12	35	**33**	15	33	**53**	3	47	**73**	14	37	**93**	17	32
14	11	37	**34**	9	37	**54**	14	38	**74**	13	34	**94**	15	36
15	11	40	**35**	20	38	**55**	15	36	**75**	11	36	**95**	13	37
16	14	37	**36**	11	40	**56**	11	39	**76**	20	30	**96**	17	34
17	21	32	**37**	11	39	**57**	13	35	**77**	12	42	**97**	12	39
18	18	34	**38**	18	33	**58**	14	37	**78**	8	44	**98**	9	41
19	14	38	**39**	10	41	**59**	16	38	**79**	12	37	**99**	19	38
20	7	45	**40**	10	30	**60**	12	37	**80**	21	28	**100**	14	36

ВАРИАНТ № 7

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	21	39	**21**	17	31	**41**	6	45	**61**	11	38	**81**	13	42
2	8	36	**22**	15	37	**42**	23	35	**62**	8	41	**82**	20	32
3	5	33	**23**	12	39	**43**	8	47	**63**	18	34	**83**	12	40
4	21	30	**24**	9	42	**44**	14	39	**64**	7	41	**84**	10	41
5	14	36	**25**	13	38	**45**	16	36	**65**	7	43	**85**	13	37
6	16	40	**26**	16	35	**46**	14	36	**66**	14	40	**86**	8	43
7	15	38	**27**	13	40	**47**	14	35	**67**	15	34	**87**	13	38
8	13	39	**28**	18	33	**48**	19	28	**68**	11	41	**88**	12	40
9	13	36	**29**	14	37	**49**	15	35	**69**	10	39	**89**	16	36
10	18	32	**30**	13	39	**50**	11	41	**70**	9	38	**90**	19	31
11	17	34	**31**	16	38	**51**	13	37	**71**	11	39	**91**	15	37
12	9	42	**32**	20	41	**52**	15	39	**72**	7	43	**92**	7	44
13	12	35	**33**	15	33	**53**	3	47	**73**	14	37	**93**	17	32
14	11	37	**34**	9	37	**54**	14	38	**74**	13	34	**94**	15	36
15	11	40	**35**	20	38	**55**	15	36	**75**	11	36	**95**	13	37
16	14	37	**36**	11	40	**56**	11	39	**76**	20	30	**96**	17	34
17	21	32	**37**	11	39	**57**	13	35	**77**	12	42	**97**	12	39
18	18	34	**38**	18	33	**58**	14	37	**78**	8	44	**98**	9	41
19	14	38	**39**	10	41	**59**	16	38	**79**	12	37	**99**	19	38
20	5	45	**40**	11	32	**60**	13	35	**80**	22	29	**100**	12	34

ВАРИАНТ № 8

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	20	37	**21**	17	31	**41**	6	45	**61**	11	38	**81**	13	42
2	9	35	**22**	14	37	**42**	23	32	**62**	8	43	**82**	22	32
3	5	33	**23**	12	39	**43**	8	47	**63**	18	34	**83**	12	40
4	21	30	**24**	9	42	**44**	14	39	**64**	7	41	**84**	10	41
5	14	36	**25**	13	38	**45**	16	36	**65**	7	43	**85**	13	37
6	16	40	**26**	16	35	**46**	14	36	**66**	14	40	**86**	8	43
7	15	38	**27**	13	40	**47**	14	35	**67**	15	34	**87**	13	38
8	13	39	**28**	18	33	**48**	19	28	**68**	11	41	**88**	12	40
9	13	36	**29**	14	37	**49**	15	35	**69**	10	39	**89**	16	36
10	18	32	**30**	13	39	**50**	18	41	**70**	9	38	**90**	19	31
11	17	34	**31**	16	38	**51**	13	37	**71**	11	39	**91**	15	37
12	9	42	**32**	20	41	**52**	15	39	**72**	7	43	**92**	7	44
13	12	35	**33**	15	33	**53**	3	47	**73**	14	37	**93**	17	32
14	11	37	**34**	9	37	**54**	14	38	**74**	13	34	**94**	15	36
15	11	40	**35**	20	38	**55**	15	36	**75**	11	36	**95**	13	37
16	14	37	**36**	11	40	**56**	11	39	**76**	20	30	**96**	17	34
17	21	32	**37**	11	39	**57**	13	35	**77**	12	42	**97**	12	39
18	18	34	**38**	18	33	**58**	14	37	**78**	8	44	**98**	9	41
19	13	37	**39**	10	43	**59**	16	39	**79**	13	38	**99**	20	36
20	7	42	**40**	14	33	**60**	17	36	**80**	25	31	**100**	15	34

ВАРИАНТ № 9

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	5	32	**21**	12	38	**41**	10	25	**61**	18	33	**81**	15	43
2	19	35	**22**	12	35	**42**	9	42	**62**	19	35	**82**	17	28
3	7	33	**23**	11	38	**43**	8	47	**63**	18	34	**83**	12	40
4	21	30	**24**	9	42	**44**	14	39	**64**	7	41	**84**	10	41
5	14	36	**25**	13	38	**45**	16	36	**65**	7	43	**85**	13	37
6	16	40	**26**	16	35	**46**	14	36	**66**	14	40	**86**	8	43
7	15	38	**27**	13	40	**47**	14	35	**67**	15	34	**87**	13	38
8	13	39	**28**	18	33	**48**	19	28	**68**	11	41	**88**	12	40
9	13	36	**29**	14	37	**49**	15	35	**69**	10	39	**89**	16	36
10	18	32	**30**	13	39	**50**	11	41	**70**	9	38	**90**	19	31
11	17	34	**31**	16	38	**51**	13	37	**71**	11	39	**91**	15	37
12	9	42	**32**	20	41	**52**	15	39	**72**	7	43	**92**	7	44
13	12	35	**33**	15	33	**53**	3	47	**73**	14	37	**93**	17	32
14	11	37	**34**	9	37	**54**	14	38	**74**	13	34	**94**	15	36
15	11	40	**35**	20	38	**55**	15	36	**75**	11	36	**95**	13	37
16	14	37	**36**	11	40	**56**	11	39	**76**	20	30	**96**	17	34
17	21	32	**37**	11	39	**57**	13	35	**77**	12	42	**97**	12	39
18	18	34	**38**	18	33	**58**	14	37	**78**	8	44	**98**	9	41
19	13	37	**39**	10	43	**59**	16	39	**79**	13	38	**99**	20	36
20	21	37	**40**	14	33	**60**	14	37	**80**	27	31	**100**	16	35

ВАРИАНТ № 10

№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*	№	*X*	*Y*
1	10	34	**21**	13	37	**41**	11	26	**61**	17	37	**81**	14	45
2	18	34	**22**	11	34	**42**	10	12	**62**	18	33	**82**	15	29
3	7	33	**23**	11	38	**43**	8	47	**63**	18	34	**83**	12	40
4	21	30	**24**	9	42	**44**	14	39	**64**	7	41	**84**	10	41
5	14	36	**25**	13	38	**45**	16	36	**65**	7	43	**85**	13	37
6	16	40	**26**	16	35	**46**	14	36	**66**	14	40	**86**	8	43
7	15	38	**27**	13	40	**47**	14	35	**67**	15	34	**87**	13	38
8	13	39	**28**	18	33	**48**	19	28	**68**	11	41	**88**	12	40
9	13	36	**29**	14	37	**49**	15	35	**69**	10	39	**89**	16	36
10	18	32	**30**	13	39	**50**	11	41	**70**	9	38	**90**	19	31
11	17	34	**31**	16	38	**51**	13	37	**71**	11	39	**91**	15	37
12	9	42	**32**	20	41	**52**	15	39	**72**	7	43	**92**	7	44
13	12	35	**33**	15	33	**53**	3	47	**73**	14	37	**93**	17	32
14	11	37	**34**	9	37	**54**	14	38	**74**	13	34	**94**	15	36
15	11	40	**35**	20	38	**55**	15	36	**75**	11	36	**95**	13	37
16	14	37	**36**	11	40	**56**	11	39	**76**	20	30	**96**	17	34
17	21	32	**37**	11	39	**57**	13	35	**77**	12	42	**97**	12	39
18	18	34	**38**	18	33	**58**	14	37	**78**	8	44	**98**	9	41
19	14	37	**39**	11	43	**59**	10	39	**79**	13	36	**99**	19	36
20	20	39	**40**	12	33	**60**	14	41	**80**	27	37	**100**	8	35

ИТОГОВЫЙ ТЕСТ

1. Распределение с плотностью вероятностей

$$f(x) = \frac{1}{b-a}, \quad a \le x \le b:$$

1) показательное;
2) равномерное;
3) нормальное;
4) биномиальное;
5) Пуассона.

2. Несмещенной оценкой называют статистическую оценку, математическое ожидание которой равно:

1) оцениваемому параметру генеральной совокупности;
2) единице;
3) нулю;
4) средней арифметической;
5) среднему квадратическому отклонению.

3. Между величинами X и Y, заданных таблицей

X	1	2	3	4	5
Y	2,1	2,9	4,1	5,2	6,1

существует связь:

1) прямая;
2) обратная;
3) средняя;
4) слабая.

4. Основным методом оценки параметров уравнений в регрессионном анализе является:

1) метод наименьших квадратов;
2) метод локальной аппроксимации;
3) метод минимизации суммы модулей;
4) метод наименьших кубов.

5. Коэффициент корреляции между величинами X и Y находится в границах:

1) $-1 \le r \le 1$;
2) $r \ge 0$;
3) $-1 < r < 1$;
4) $0 \le r \le 1$.

6. Предельная ошибка выборки при случайном повторном отборе определяется соотношением:

1) $t \cdot \dfrac{\sigma}{\sqrt{n}}$;
2) $t \cdot \dfrac{\sigma}{n}$;
3) $t \cdot \dfrac{\sigma^2}{\sqrt{n}}$;
4) $t \cdot \dfrac{\sqrt{n}}{\sigma}$.

7. Коэффициент вариации V определяется по формуле:

1) $\dfrac{\sigma}{\bar{X}_B} \cdot 100\%$;
2) $\dfrac{\sqrt{A_B}}{\bar{X}_B} \cdot 100\%$;
3) $\dfrac{\sigma^2}{\bar{X}_B} \cdot 100\%$;
4) $\dfrac{\bar{X}_B}{\sigma} \cdot 100\%$.

8. Функция распределения $F(x)$ может принимать значения на промежутке:

1) $[0;1]$;
2) $[0;+\infty)$;

3) $(0;1)$;
4) $(-\infty;+\infty)$.

9. Математическое ожидание непрерывной случайной величины определяется по формуле:

1) $M(x) = x \cdot \int\limits_{-\infty}^{+\infty} f(x)dx;$

2) $M(x) = f(x) \cdot \int\limits_{-\infty}^{+\infty} x\,dx;$

3) $M(x) = \int\limits_{-\infty}^{+\infty} f(x)dx \int\limits_{-\infty}^{+\infty} x\,dx;$

4) $M(x) = \int\limits_{-\infty}^{+\infty} f(x)dx.$

10. В опыте бросания двух игральных костей событие A_1 – «хотя бы на одной кости появится три очка» и событие A_2 – « на каждой кости появится четное число очков» являются:

1) Равновозможными;
2) совместными;
3) несовместными;
4) единственно возможными.

11. Из пяти карточек с буквами Е, А, Т, И, К последовательно, по одной вынимают три карточки и располагают в ряд. Вероятность того, что в результате опыта получится слово «КИТ» равна ……. .

12. В урне 5 белых и 4 черных шара. Наугад выбираются два шара. Вероятность того, что будет выбрано два черных шара, равна числу ……. .

13. Вероятность попадания в мишень при одном выстреле для данного стрелка равна 0,7. Для вычисления вероятности того, что при 200 выстрелах мишень будет поражена 160 раз, следует использовать:

1) формулу полной вероятности;
2) формулу Пуассона;
3) локальную формулу Муавра-Лапласа;
4) формулу Байеса.

14. Посажено 5 семян растения. Вероятность того, что семя не прорастет, равна 0,2. Тогда наивероятнейшим числом невзошедших семян будет:
1) 2;
2) 1;
3) 4;
4) 3.

15. Вероятность попадания случайной величины X в интервал $(0;1)$ заданной функцией распределения

$$F(x)=\begin{cases} 0, если & x\le 1 \\ x/2, если & 0<x<2 \\ 1, если & x>2 \end{cases}$$ равна ….. .

16. Среднее квадратическое отклонение $\sigma(x)$ случайной величины X есть:

1) $\sqrt{M(X)}$;
2) $\sqrt{D(X)}$;
3) $D^2(X)$;
4) $D(X)-M(X)$.

17. Пусть X и Y – независимые дискретные случайные величины, причем $DX=2$, $DY=9$. Тогда дисперсия $D(5X-3Y+2)$ равна:
1) 133;
2) 39;
3) 37;
4) 131.

18. При выборке объемом $n = 10$ найдена выборочная дисперсия $D_B = 3{,}6$. Тогда исправленное среднее квадратическое отклонение равно:

1) 2;
2) 4;
3) 3,24;
4) 8.

19. Статистическое распределение выборки объемом $n = 50$ имеет вид

x_i	1	2	3	4
n_i	5	14	n_3	8

Тогда n_3 равен:

1) 7;
2) 23;
3) 50;
4) 24.

20. Если основная гипотеза имеет вид $H_0 : a = 7$, то конкурирующей гипотезой может являться:

1) $H_0 : a \leq 7$;
2) $H_0 : a > 7$;
3) $H_0 : a \geq 7$;
4) $H_0 : a \neq 7$.

ЗАКЛЮЧЕНИЕ

Учебное пособие предназначено для оказания помощи студентам в освоении разделов «Теория вероятностей» и «Математическая статистика», для подготовки к практическим занятиям, для выполнения индивидуальных заданий. В пособии изложены основные теоретические сведения, рассматриваются алгоритмы практических расчетов по статистике. В каждом разделе приводятся типовые задачи с подробными решениями.

Следует подчеркнуть важность применения вероятностных и статистических методов в агрономии, зоотехнии и других смежных областях. Владение основами теории вероятностей и статистики позволяет специалистам в аграрном секторе принимать обоснованные решения, опираться на количественные данные и тем самым повышать эффективность производства и оптимизировать процессы управления. В условиях изменения климата, демографии населения и ограниченных ресурсов, аграрный сектор сталкивается с новыми вызовами, требующими комплексного подхода к анализу и прогнозированию. Методы математической статистики позволяют не только анализировать текущее состояние, но и выделять тенденции, оценивая вероятные риски и возможности. Это позволяет эффективно планировать агрономические мероприятия, управлять ресурсами и повышать продуктивность сельского хозяйства.

Представленное учебное пособие станет для студентов полезным инструментом в изучении и применении теории вероятностей и статистики в их будущей профессиональной деятельности, будущих исследованиях и практике в области сельского хозяйства.

БИБЛИОГРАФИЯ

1. Данко, П. Е. Высшая математика в упражнениях и задачах. Ч. 2: учебное пособие для вузов / П. Е. Данко, А. Г. Попов, Т. Я. Кожевников. – Москва : Астрель, 2011. – 416 с.

2. Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учебник для вузов / В. Е. Гмурман. – 11-е изд., перераб. и доп. – Москва : Юрайт, 2024. – 406 с.

3. Письменный, Д. Т. Конспект лекций по теории вероятностей и математической статистике / Д. Т. Письменный. – Москва : Айрис-пресс, 2024. – 256 с.

4. Зайцев, И. А. Высшая математика : учебник для вузов / И. А. Зайцев. – 4-е изд., стереотип. – Москва : Дрофа, 2005. – 398 с.

5. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. – 5-е изд., перераб. и доп. – Москва : Юрайт, 2024. – 538 с.

6. Соколов, Г. А. Математическая статистика : учебник для вузов. – Москва : Экзамен, 2004. – 432 с.

ПРИЛОЖЕНИЯ

Приложение А

Значения функции $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	**0,3989**	**0,3989**	**0,3989**	**0,3988**	**0,3986**	**0,3984**	**0,3982**	**0,3980**	**0,3977**	**0,3973**
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3525	3503	3485	3467	3448	3429	3411	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2987	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2631	2613	2589	2565	2541	2516	2492	2468	2444
1,0	**0,2420**	**0,2396**	**0,2371**	**0,2347**	**0,2323**	**0,2299**	**0,2275**	**0,2251**	**0,2227**	**0,2203**
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1734	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0762	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	**0,0540**	**0,0529**	**0,0519**	**0,0508**	**0,0498**	**0,0488**	**0,0478**	**0,0468**	**0,0459**	**0,0449**
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	**0,0044**	**0,0043**	**0,0042**	**0,0040**	**0,0039**	**0,0038**	**0,0037**	**0,0036**	**0,0035**	**0,0034**
3,1	0033	0032	0031	0030	0029	0028	0027	0036	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001
4,0	**0,0001**	**0,0001**	**0,0001**	**0,0001**	**0,0001**	**0,0001**	**0,0001**	**0,0001**	**0,0001**	**0,0001**

Приложение В

Значения функции $\Phi_0(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	**0,0000**	**0,0040**	**0,0080**	**0,0120**	**0,0160**	**0,0199**	**0,0239**	**0,0279**	**0,0319**	**0,0359**
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0754
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1228	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2504	2088	2123	2157	2190	2224
0,6	2258	2291	2324	2357	2389	2422	2454	2486	2518	2549
0,7	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2996	3023	3051	3079	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	**0,3413**	**0,3438**	**0,3461**	**0,3485**	**0,3508**	**0,3531**	**0,3554**	**0,3577**	**0,3599**	**0,3621**
1,1	3413	3438	3461	3486	3508	3531	3554	3577	3599	3621
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4130	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
1,6	4452	4483	4474	4485	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4615	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4700	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	**0,4773**	**0,4778**	**0,4783**	**0,4788**	**0,4793**	**0,4798**	**0,4803**	**0,4808**	**0,4812**	**0,4817**
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4865	4867	4871	4875	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	4963	4955	4956	4957	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4980	4980	4981
2,9	4981	4982	4983	4983	4984	4984	4985	4985	4986	4986

x	Десятые доли x									
	0	1	2	3	4	5	6	7	8	9
3,0	49865	49903	49931	49952	49966	49977	49984	49989	49993	49995
4,0	499968					499997				
5,0	49999997									

Приложение C

Значения функции $t_\gamma = t(\gamma; n)$

n	γ			n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999		0,95	0,99	0,999
5	2,78	4,60	8,61	**15**	2,15	2,98	4,14	**45**	2,016	2,692	3,527
6	2,57	4,03	6,86	**16**	2,13	2,95	4,07	**50**	2,009	2,679	3,502
7	2,45	3,71	5,96	**17**	2,12	2,92	4,02	**60**	2,001	2,662	3,464
8	2,37	3,50	5,41	**18**	2,11	2,90	3,97	**70**	1,996	2,649	3,439
9	2,31	3,36	5,04	**19**	2,10	2,68	3,92	**80**	1,001	2,640	3,416
10	2,26	3,25	4,78	**20**	2,093	2,861	3,883	**90**	1,987	2,633	3,403
11	2,23	3,17	4,59	**25**	2,064	2,797	3,745	**100**	1,984	2,627	3,392
12	2,20	3,11	4,44	**30**	2,045	2,756	3,659	**120**	1,980	2,617	3,374
13	2,18	3,06	4,32	**35**	2,032	2,720	3,600	**140**	1,960	2,576	3,291
14	2,16	3,01	4,22	**40**	2,023	2,708	3,558	∞	1,960	2,576	3,291

Приложение D

Критические точки распределения $\chi^2(\alpha; \nu)$

(ν – число степеней свободы)

$\nu \backslash \alpha$	**0,10**	**0,05**	**0,02**	**0,01**	$\nu \backslash \alpha$	**0,10**	**0,05**	**0,02**	**0,01**
1	2,706	3,841	5,412	6,635	**16**	13,542	26,296	29,633	32,000
2	4,605	5,991	7,824	9,210	**17**	24,769	27,587	30,995	33,409
3	6,251	7,815	9,837	11,345	**18**	25,989	28,869	32,346	34,805
4	7,779	9,488	11,668	13,277	**19**	27,204	30,144	33,687	36,191
5	9,236	11,070	13,388	15,086	**20**	28,412	31,410	35,020	37,566
6	10,645	12,592	15,033	16,812	**21**	29,615	32,671	36,343	38,932
7	12,017	14,067	16,622	18,475	**22**	30,813	32,924	37,659	40,289
8	13,362	15,507	18,168	20,090	**23**	32,07	35,172	38,868	41,638
9	14,684	16,919	19,679	21,666	**24**	33,196	36,415	40,270	42,980
10	15,987	18,307	21,161	23,209	**25**	34,382	37,652	41,566	44,314
11	17,275	19,675	22,618	24,618	**26**	35,563	38,885	42,856	45,642
12	18,549	21,026	24,054	26,217	**27**	36,741	40,113	44,140	46,963
13	19,812	22,362	25,472	26,688	**28**	37,916	41,337	45,419	48,278
14	21,064	23,685	26,683	29,141	**29**	39,087	42,557	46,693	49,588
15	23,307	24,996	28,259	30,578	**30**	40,256	43,773	47,962	50,892

Приложение Е

Нижние γ_1 и верхние γ_2 границы доверительного интервала

$$\gamma_1 \cdot S < \sigma < \gamma_2 \cdot S$$

γ	**0,95**		γ	**0,95**		γ	**0,95**	
$\nu = n - 1$	γ_1	γ_2	$\nu = n - 1$	γ_1	γ_2	$\nu = n - 1$	γ_1	γ_2
1	0,446	31,19	**16**	0,745	1,522	**40**	0,821	1,279
2	0,521	6,28	**17**	0,750	1,499	**50**	0,837	1,243
3	0,566	3,73	**18**	0,756	1,479	**60**	0,849	1,217
4	0,599	2,87	**19**	0,760	1,460	**70**	0,858	1,197
5	0,624	2,45	**20**	0,765	1,444	**80**	0,866	1,183
6	0,644	2,202	**21**	0,769	1,429	**90**	0,873	1,171
7	0,661	2,215	**22**	0,773	1,416	**100**	0,878	1,161
8	0,675	1,916	**23**	0,777	1,402	**200**	0,912	1,090
9	0,688	1,826	**24**	0,781	1,391			
10	0,699	1,755	**25**	0,784	1,380			
11	0,708	1,698	**26**	0,788	1,371			
12	0,717	1,651	**27**	0,791	1,361			
13	0,725	1,611	**28**	0,794	1,352			
14	0,732	1,577	**29**	0,796	1,344			
15	0,739	1,548	**30**	0,799	1,337			

МАТЕМАТИКА.
ВЕРОЯТНОСТЬ И СТАТИСТИКА

Учебное пособие

ЛУКИНОВА Светлана Георгиевна
ИВАНОВ Владимир Иванович
ЖДАНОВА Вера Даниловна

Редактор Е.С. Хакимова

Подписано в печать 29.04.2026. Формат 60 × 84/16. Бумага тип. № 1.
Офсетная печать. Объем п.л. 11,0. Тираж 100 экз. Заказ № 472
Отпечатано в типографии Красноярского государственного аграрного университета
660017, Красноярск, ул. Ленина, 117