

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Чувашский государственный университет имени И.Н. Ульянова»

В.А. Тобоев
С.И. Фролов

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Чебоксары
2023

УДК 519.21 (076.5)

ББК В172я73

Т50

Рецензенты:

А.Н. Кулагина – канд. физ.-мат. наук, доцент кафедры актуарной и финансовой математики факультета ПМФИТ Чувашского государственного университета имени И.Н. Ульянова;

В.П. Филиппов – канд. физ.-мат. наук, доцент кафедры экономики и информатики Чебоксарского кооперативного института (филиал) Российского университета кооперации

Тобоев В.А.

Т50 Теория вероятностей и математическая статистика: учеб. пособие / В.А. Тобоев, С.И. Фролов; Чуваш. гос. ун-т им. И.Н. Ульянова. – Чебоксары, 2023. – 104 с.

ISBN 978-5-7677-3631-7

Содержит краткие теоретические сведения, задания по вариантам, подробный порядок выполнения, а также контрольные вопросы и задачи для самостоятельного решения.

Для студентов I–II курсов очной и заочной форм обучения радиотехнического, машиностроительного, строительного факультетов и факультетов энергетики и электротехники, информатики и вычислительной техники.

Ответственный редактор доцент А.С. Сабилов

Утверждено Учебно-методическим советом университета

УДК 519.22 (076.5)

ББК В172я73

© Чувашский госуниверситет
им. И.Н. Ульянова, 2023

© Тобоев В.А., Фролов С.И., 2023

ISBN 978-5-7677-3631-7

ПРЕДИСЛОВИЕ

В настоящее время вероятностно-статистические методы широко применяются в радиотехнике, электронике, теории связи, теории автоматизированных систем управления, теории автоматического управления, вычислительной технике. Любая процедура управления техническим устройством протекает в заранее неизвестных, случайных условиях и неизбежно сопровождается случайными ошибками измерений, выполнения команд и т.д. Анализ работы такой системы требует учета большого количества случайных факторов. Выявление и изучение статистических закономерностей в случайных явлениях – главная цель теории вероятностей и математической статистики. Поэтому знакомство с методами теории вероятностей и математической статистики необходимо сегодня каждому грамотному специалисту.

Предлагаемое пособие предназначено для студентов технических специальностей, изучающих курс теории вероятностей и математической статистики, и включает три комплексных задания (одна контрольная работа, две лабораторные работы) по основам вероятностно-статистических методов. Лабораторные работы предполагают проведение полной статистической обработки полученных в результате случайного эксперимента данных – случайной выборки. По каждой из рассматриваемых тем приведены основные теоретические положения, примеры решения задач, варианты заданий, подробный порядок выполнения работ. К каждой работе даны контрольные вопросы и задачи для самостоятельного решения. Представленные задачи служат для усвоения материала основных разделов теории вероятностей и математической статистики на конкретных примерах, возникающих в практических приложениях. Численные расчеты по основам теории вероятностей можно выполнять с использованием калькулятора. В задачах математической статистики для проведения вычислений необходимо использовать наиболее распространенные и доступные программные продукты (системы компьютерной математики Mathcad, Matlab, Stadia, Scilab и другие, а также электронные таблицы Excel).

1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ.

ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ

1.1. Элементы комбинаторики

При решении комбинаторных задач применяются два универсальных правила.

Правило сложения. Пусть требуется выполнить одно из каких-либо k действий, взаимно исключающих друг друга. Если первое действие можно выполнить n_1 способами, второе действие – n_2 способами и так до k -го действия, которое можно выполнить n_k способами, то одно из этих k действий можно выполнить $n_1 + n_2 + n_3 + \dots + n_k$ способами.

Правило умножения. Если требуется выполнить одно за другим каких-либо k действий, которые можно выполнить соответственно $n_1, n_2, \dots, n_k$ способами, то все k действий вместе могут быть выполнены $n_1 \cdot n_2 \cdot \dots \cdot n_k$ способами.

Рассмотрим некоторое множество Q , состоящее из n различных элементов.

Множество, состоящее из k элементов ($1 \leq k \leq n$), называется упорядоченным, если каждому элементу этого множества ставится в соответствие число от 1 до k , причем различным элементам множества соответствуют разные числа.

Размещениями из n элементов по k называются упорядоченные подмножества множества Q , состоящие из k различных элементов и отличающиеся друг от друга составом элементов и порядком их расположения.

Число размещений из n элементов по k равно

$$A_n^k = n \cdot (n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}. \quad (1.1)$$

Перестановками из n элементов называются размещения из n элементов по n . Другими словами, перестановки – это упорядоченные подмножества множества Q , состоящие из всех элементов данного множества и отличающиеся друг от друга только порядком их расположения.

Число перестановок из n элементов равно

$$P_n = A_n^n = n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1. \quad (1.2)$$

Пустое множество можно упорядочить только одним способом

$$P_0 = 0! = 1. \quad (1.3)$$

Сочетаниями из n элементов по k называются упорядоченные подмножества множества Q , состоящие из k различных элементов и отличающиеся друг от друга только составом элементов.

Число сочетаний из n элементов по k равно

$$C_n^k = \frac{A_n^k}{P_k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{1 \cdot 2 \cdot \dots \cdot k}. \quad (1.4)$$

Если в перестановках из общего числа n элементов есть k различных элементов, при этом первый элемент повторяется n_1 раз, 2-й – n_2 раз, k -й – n_k раз, причем $n_1 + n_2 + \dots + n_k = n$, то такие перестановки называют перестановками с повторением.

Число перестановок с повторениями из n элементов определяется по формуле

$$P_n(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}. \quad (1.5)$$

Число сочетаний с повторениями из n элементов по m определяется по формуле

$$\tilde{C}_n^m = C_{n+m-1}^m = \frac{(n+m-1)!}{m!(n-1)!}, \quad (1.6)$$

а число размещений с повторениями из n элементов по m определяется по формуле

$$\tilde{A}_n^m = n^m. \quad (1.7)$$

1.2. Классическое определение вероятности

Случайное событие – это любой факт, который в результате испытания может произойти или не произойти.

Численная мера возможности наступления события A называется его вероятностью и обозначается $P(A)$.

Классическое определение вероятности события A :

$$P(A) = \frac{m}{n}, \quad (1.8)$$

где n – общее число равновозможных исходов испытания (элементарных событий); m – число равновозможных исходов, благоприятствующих событию A .

Из классического определения вероятности вытекают несколько свойств вероятности.

1. $0 \leq P(A) \leq 1$ для любого события A .

2. Вероятность достоверного события

$$P(\Omega) = \frac{n}{n} = 1.$$

3. Вероятность противоположного события

$$P(\bar{A}) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(A).$$

4. Вероятность невозможного события

$$P(\emptyset) = \frac{0}{n} = 0.$$

1.3. Геометрическое определение вероятности

Геометрическая вероятность – это обобщение классического определения вероятности на случай бесконечного числа равновозможных исходов.

Согласно геометрическому определению вероятности вероятность попадания точки в любую область A (событие A) пропорциональна ее мере

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}. \quad (1.9)$$

Здесь предполагается, что все исходы эксперимента изображаются точками некоторой области $\Omega \in R^n$ и попадание точки в любые области $A \subseteq \Omega$, имеющие конечную меру μ , равновозможно и не зависит от формы и расположения A внутри Ω .

В одномерном случае ($n = 1$) под мерой μ понимается длина подмножества на числовой прямой $\mathbb{R}$

$$P(A) = \frac{l(A)}{l(\Omega)}.$$

При $n = 2$ под мерой μ понимается площадь подмножества на плоскости $\mathbb{R}^2$

$$P(A) = \frac{S(A)}{S(\Omega)}.$$

При $n = 3$ под мерой μ понимается объем подмножества в пространстве $\mathbb{R}^3$

$$P(A) = \frac{V(A)}{V(\Omega)}.$$

Геометрическая вероятность имеет те же свойства, что и классическая вероятность.

1.4. Теоремы сложения и умножения вероятностей

Если события могут произойти одновременно, то они называются совместными событиями.

Два или несколько событий называются независимыми событиями, если шансы на наступление каждого из них не зависят от того, произошли какие-либо из остальных событий или нет.

Суммой событий $A_1, A_2, \dots, A_n$ называется событие $A_1 + A_2 + \dots + A_n$, состоящее в том, что произошло хотя бы одно из этих событий.

Произведением событий $A_1, A_2, \dots, A_n$ называется событие $A_1 \cdot A_2 \cdot \dots \cdot A_n$, состоящее в том, что произошли все эти события.

Событие, состоящее в наступлении события B при условии, что произошло событие A , обозначается B/A . Вероятность этого события $P(B/A)$ или $P_A(B)$ называется условной вероятностью события B .

Теорема сложения вероятностей. Пусть A и B совместные события. Тогда

$$P(A+B) = P(A) + P(B) - P(AB). \quad (1.10)$$

Если A и B несовместные события ($P(AB) = P(\emptyset) = 0$), то

$$P(A+B) = P(A) + P(B). \quad (1.11)$$

Теорема умножения вероятностей. Пусть A и B зависимые события. Тогда

$$P(AB) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B). \quad (1.12)$$

Если A и B события независимые, то

$$P(A/B) = P(A) \text{ и } P(B/A) = P(B)$$

и поэтому

$$P(AB) = P(A) \cdot P(B). \quad (1.13)$$

1.5. Формула полной вероятности, формула Байеса

Пусть событие A может произойти только с одним из несовместимых событий $H_1, H_2, \dots, H_n$, образующих полную группу $\left(\sum_{i=1}^n H_i = \Omega, H_i H_j = \emptyset, i \neq j \right)$.

Тогда событие A можно представить в следующем виде:

$$A = H_1 A + H_2 A + \dots + H_n A.$$

Формула полной вероятности имеет вид

$$P(A) = \sum_{k=1}^n P(H_k) \cdot P(A/H_k). \quad (1.14)$$

Пусть теперь событие A произошло и необходимо с учетом этого провести корректировку априорных (до проведения опыта) вероятностей $P(H_1), P(H_2), \dots, P(H_n)$. Для этого определяют апостериорные (после проведения опыта) вероятности $P(H_1/A), P(H_2/A), \dots, P(H_n/A)$.

Эти вероятности вычисляются по формуле Байеса

$$P(H_k/A) = \frac{P(H_k) \cdot P(A/H_k)}{P(A)}, \quad (k = \overline{1, n}). \quad (1.15)$$

1.6. Формула Бернулли. Локальная и интегральная теоремы Муавра – Лапласа. Формула Пуассона

Пусть производится n независимых испытаний в одинаковых условиях, в результате которых может произойти событие A с вероятностью $P(A) = p$ и не произойти с вероятностью $P(\bar{A}) = 1 - p = q$.

Вероятность того, что в n независимых испытаниях событие A появится ровно m раз, вычисляется по формуле Бернулли

$$P_n(m) = C_n^m p^m q^{n-m} \frac{n!}{m!(n-m)!} p^m q^{n-m}. \quad (1.16)$$

Следствие 1. Вероятность $P_n(n)$ появления события A n раз в n независимых испытаниях равна $P_n(n) = p^n$.

Следствие 2. Вероятность $P_n(0)$ не появления события ни разу в n независимых испытаниях равна $P_n(0) = q^n$.

Следствие 3. Вероятность появления события A хотя бы один раз в n независимых испытаниях равна $P_n(k \geq 1) = 1 - q^n$.

Таким образом, вероятность того, что в n независимых испытаниях событие наступит:

- 1) менее m раз: $P_n(0) + P_n(1) + \dots + P_n(m-1)$;
- 2) не более m раз: $P_n(0) + P_n(1) + \dots + P_n(m)$;
- 3) более m раз: $P_n(m+1) + P_n(m+2) + \dots + P_n(n)$;
- 4) не менее m раз: $P_n(m) + P_n(m+1) + \dots + P_n(n)$;
- 5) не менее m_1 раз и не более m_2 раз:
 $P_n(m_1) + P_n(m_1+1) + \dots + P_n(m_2)$.

В испытаниях Бернулли существует целое число m_0 , которому соответствует максимальное значение вероятности.

Наивероятнейшее число успехов m_0 удовлетворяет двойному неравенству

$$np - q \leq m_0 \leq np + p. \quad (1.17)$$

При больших значениях числа испытаний n и m пользоваться формулой Бернулли затруднительно. В этих случаях пользуются асимптотическими (приближенными) формулами вычисления вероятности $P_n(m)$.

Локальная предельная теорема Муавра – Лапласа. Если вероятность наступления события A при каждом отдельном испытании постоянна и равна p , то вероятность наступления события A ровно m раз в серии из n испытаний может быть представлена приближенной формулой (при условии $npq \geq 10$)

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi(x), \text{ где } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x = \frac{m - np}{\sqrt{npq}}, \quad q = 1 - p. \quad (1.18)$$

Функция Гаусса $\varphi(x)$ представлена в таблицах к учебникам по теории вероятностей. Эта функция может быть также вычислена в системе компьютерной математики (Mathcad, Matlab, Stadia и др.)

Интегральная теорема Муавра – Лапласа.

Вероятность того, что число m появления события A в схеме Бернулли находится в заданном промежутке $m_1 \leq m \leq m_2$ при большом числе испытаний n вычисляется по формуле

$$P_n(m_1 \leq m < m_2) \approx \frac{1}{\sqrt{2\pi}} \int_0^{x_2} e^{-\frac{x^2}{2}} dx - \frac{1}{\sqrt{2\pi}} \int_0^{x_1} e^{-\frac{x^2}{2}} dx = \Phi(x_2) - \Phi(x_1), \quad (1.19)$$

где $x_1 = \frac{m_1 - np}{\sqrt{npq}}$, $x_2 = \frac{m_2 - np}{\sqrt{npq}}$; $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ – функция

Лапласа.

Функция $\Phi(x)$ – нечетная. Значения функции Лапласа представлены в таблице (прил. 1).

Теорема Пуассона. Пусть в серии из n испытаний вероятность p появления события A в каждом испытании мала, а произведение $\lambda = np \leq 10$ ($n \geq 50$, $p < 0,1$).

Тогда вероятность того, что событие A появится ровно в m испытаниях вычисляется по приближенной формуле Пуассона

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}. \quad (1.20)$$

1.7. Случайные величины

Случайной называется величина, которая в зависимости от случая принимает те или иные значения с некоторой вероятностью. Случайная величина обозначается заглавными буквами латинского алфавита $X, Y, Z, \dots$, а принимаемые ими значения – соответствующими строчными буквами $x, y, z, \dots$

Случайные величины делятся на дискретные и непрерывные.

Дискретной называется случайная величина, которая может принимать лишь конечное или счетное число значений.

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

Закон распределения дискретной случайной величины может быть задан в виде таблицы (ряд распределения), графически (многоугольник или полигон распределения вероятностей) и аналитически.

Ряд распределения может быть задан в виде таблицы

x_i	x_1	x_2	x_3	$\dots$	x_n
p_i	p_1	p_2	p_3	$\dots$	p_n

Основными числовыми характеристиками дискретной случайной величины являются:

$$M(X) = m_x = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i; \quad (1.21)$$
$$D(X) = M[(X - m_x)^2] = \sum_{i=1}^n (x_i - m_x)^2 p_i = M(X^2) - m_x^2; \quad (1.22)$$
$$\sigma_x = \sqrt{D(X)}; \quad (1.23)$$

– медиана Me (геометрически – абсцисса точки, в которой площадь многоугольника распределения делится пополам)

$$P(X < Me) = P(X > Me) = \frac{1}{2}. \quad (1.24)$$

$$F(x) = P(X < x). \quad (1.25)$$
$$F(x) = \begin{cases} 0, & x \leq x_1; \\ p_1, & x_1 < x \leq x_2; \\ p_1 + p_2, & x_2 < x \leq x_3; \\ p_1 + p_2 + p_3, & x_3 < x \leq x_4; \\ \\ p_1 + p_2 + p_3 + ... + p_n, & x_{n-1} < x \leq x_n; \\ 1, & x > x_n. \end{cases} \quad (1.26)$$

Отметим, что значение функции распределения в точке медианы Me равно

$$F(Me) = \frac{1}{2}.$$

Непрерывной называется случайная величина, которая может принимать любые значения на всей числовой оси или на некотором ее промежутке.

Распределение непрерывной случайной величины однозначно определяется ее функцией распределения $F(x)$

$$P(x_1 < X < x_2) = F(x_2) - F(x_1). \quad (1.27)$$

Производная функции распределения

$$f(x) = F'(x) \quad (1.28)$$

называется плотностью распределения непрерывной случайной величины X .

Свойства плотности распределения.

1. Кривая распределения лежит в верхней полуплоскости

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

2. Площадь фигуры, ограниченной кривой распределения и осью абсцисс, равна единице (условие нормировки)

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

3. Вероятность попадания случайной величины X в промежутки от x_1 до x_2 равна площади фигуры, ограниченной кривой распределения, осью абсцисс и прямыми $x = x_1$ и $x = x_2$

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) dx.$$

4. Функция распределения $F(x) = \int_{-\infty}^x f(x) dx$.

Основными числовыми характеристиками непрерывной случайной величины являются:

– математическое ожидание (среднее значение)

$$M(X) = m_x = \int_{-\infty}^{\infty} x f(x) dx; \quad (1.29)$$

– дисперсия

$$D(X) = M \left[(X - m_x)^2 \right] = \int_{-\infty}^{\infty} (x - m_x)^2 f(x) dx; \quad (1.30)$$

– среднее квадратическое отклонение

$$\sigma_x = \sqrt{D(X)};$$

– мода Mo – такое значение случайной величины, при котором плотность распределения имеет максимум;

– медиана Me (абсцисса точки, в которой площадь под кривой распределения делится пополам).

1.8. Основные законы распределения непрерывных случайных величин

Равномерное распределение

Непрерывная случайная величина X имеет равномерное распределение на отрезке $[a, b]$, если её плотность распределения имеет вид:

$$f(x) = \begin{cases} 0, & x < a; \\ C, & a \leq x \leq b; \\ 0, & x > b. \end{cases} \quad (1.31)$$

Постоянная C определяется из условия равенства единице площади, ограниченной кривой распределения

$$\int_{-\infty}^{\infty} f(x) dx = \int_a^b C dx = Cx \Big|_a^b = C(b-a) = 1 \Rightarrow C = \frac{1}{b-a}.$$

Функция распределения на отрезке определяется формулой

$$F(x) = \int_{-\infty}^x f(x) dx = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x \leq b; \\ 1, & x > b. \end{cases}$$

Основные числовые характеристики

$$M(X) = \frac{a+b}{2}; \quad D(X) = \frac{(b-a)^2}{12}; \quad \sigma(X) = \frac{b-a}{2\sqrt{3}}.$$

Показательное (экспоненциальное) распределение

Непрерывная случайная величина X имеет показательное распределение с параметром λ , если её плотность распределения имеет вид:

$$f(x) = \begin{cases} 0, & x < 0; \\ \lambda e^{-\lambda x}, & x \geq 0, \end{cases} \quad (1.32)$$

где $\lambda > 0$ – параметр распределения.

Функция распределения случайной величины, распределенной по показательному закону, равна

$$F(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Основные числовые характеристики

$$M(X) = \frac{1}{\lambda}; \quad D(X) = \frac{1}{\lambda^2}; \quad \sigma(X) = \frac{1}{\lambda}.$$

Нормальное распределение

Непрерывная случайная величина X имеет нормальное распределение с параметрами m и $\sigma > 0$, если её плотность распределения имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (1.33)$$

Функция распределения нормальной случайной величины равна

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-m)^2}{2\sigma^2}} dt. \quad (1.34)$$

Математическое ожидание случайной величины, распределенной по нормальному закону, равно параметру m , а дисперсия равна σ^2

$$M(X) = m; \quad D(X) = \sigma^2.$$

Вероятность попадания случайной величины в заданный промежуток от x_1 до x_2 имеет следующий вид:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Последний интеграл не выражается через элементарные функции. После преобразования, запишем его в виде

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{t^2}{2}} dt = \Phi(\beta) - \Phi(\alpha), \quad (1.35)$$

где $\beta = \frac{x_2 - m}{\sigma}$, $\alpha = \frac{x_1 - m}{\sigma}$; $\Phi(x)$ – функция Лапласа, введенная ранее. Функция Лапласа – нечетная функция $\Phi(-x) = -\Phi(x)$.

Вероятность попадания в симметричный интервал дается формулой

$$P(|X - m| \leq \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

1.9. Двумерные случайные величины

Пусть (x, y) – пара действительных чисел, тогда множество

$$\{(X < x), (Y < y)\}$$

является множеством событий, а их вероятность задает значение двумерной функции распределения.

Функцией распределения системы двух случайных величин называется функция двух аргументов $F(x, y)$, равная вероятности совместного выполнения двух неравенств $X < x$, $Y < y$

$$F(x, y) = P\{X < x, Y < y\}. \quad (1.36)$$

Свойства функции распределения двух случайных величин.

1. Значения функции распределения лежат в интервале

$$0 \leq F(x, y) \leq 1.$$

2. Если оба аргумента стремятся к бесконечности, то

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

3. При стремлении одного или обоих аргументов к минус бесконечности

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

4. Если один из аргументов стремится к плюс бесконечности, то

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

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

Здесь $F_1(x)$ и $F_2(x)$ функции распределения одномерных случайных величин X и Y , составляющих двумерную случайную величину $\{X, Y\}$.

Законом распределения дискретной двумерной случайной величины называется множество ее значений и соответствующих вероятностей

$$p_{ij} = P\{(X = x_i), (Y = y_j)\}. \quad (1.37)$$

Закон распределения двумерной случайной величины с конечным числом значений представляется в виде двумерной матрицы распределения, в каждой ячейке которой содержатся вероятности произведения событий

$$A_{ij} = \{(X = x_i), (Y = y_j)\}; \quad i = 1, 2, 3, \dots, n, \quad j = 1, 2, 3, \dots, m.$$

События A_{ij} состоят в том, что случайная величина X принимает значение x_i , а Y принимает значение y_j , то есть они наступают с соответствующими вероятностями p_{ij} .

Если события A_{ij} несовместны и образуют полную группу событий, то

$$\sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1.$$

При этом распределения одномерных случайных величин определяются формулами

$$p_i = P(X = x_i) = \sum_{j=1}^m p_{ij}; \quad p_j = P(Y = y_j) = \sum_{i=1}^n p_{ij}. \quad (1.38)$$

Непрерывной называется двумерная случайная величина, если ее функция распределения $F(x, y)$ непрерывна по обоим аргументам.

Плотность и функция распределения двумерной случайной величины связаны соотношениями:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}, \quad F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy.$$

Двумерная плотность распределения неотрицательна и удовлетворяет условию нормировки:

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

Вероятность попадания случайной точки в некоторую область D в двумерном пространстве выражается интегралом:

$$P((X, Y) \in D) = \iint_D f(x, y) dx dy. \quad (1.39)$$

Одномерные плотности распределения составляющих двумерной случайной величины имеют вид

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

Тогда соответствующие функции распределения равны:

$$F_1(x) = F(x, \infty) = \int_{-\infty}^x f_1(x) dx; \quad F_2(y) = F(\infty, y) = \int_{-\infty}^y f_2(y) dy.$$

Если случайные величины X и Y зависимы между собой, то закон распределения $F(x, y)$ не может быть выражен через законы распределения составляющих. В этом случае вводится понятие условного закона распределения (распределение одной случайной величины при условии, что другая случайная величина приняла определенное значение).

Условное распределение случайной величины X при условии $Y = y_i$ задается формулой

$$p_j(x_i) = P(X/Y) = \frac{P(XY)}{P(Y)} = \frac{p_{ij}}{p_j}. \quad (1.40)$$

Аналогично, если произошло событие $X = x_i$, то распределение случайной величины Y вычисляется по формуле

$$p_i(y_j) = P(Y/X) = \frac{P(XY)}{P(X)} = \frac{p_{ij}}{p_i}. \quad (1.41)$$

В случае непрерывных двумерных величин условная плотность распределения вычисляется по формулам

$$f(x/y) = \frac{f(x, y)}{f_2(y)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx};$$

$$f(y/x) = \frac{f(x, y)}{f_1(x)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dy}.$$

Мерой взаимосвязи двух случайных величин X и Y может служить коэффициент ковариации:

$$\text{cov}(X, Y) = M[(X - m_x)(Y - m_y)] = M(XY) - m_x m_y. \quad (1.42)$$

На практике используется коэффициент корреляции, равный отношению коэффициента ковариации к произведению средних квадратических отклонений случайных величин X и Y :

$$r_{xy} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{M[(X - m_x) \cdot (Y - m_y)]}{\sigma_x \cdot \sigma_y} \quad (1.43)$$

Свойства коэффициента корреляции:

- 1) $-1 \leq r_{xy} \leq 1$;
- 2) для независимых случайных величин X и Y $r_{xy} = 0$. Если $r_{xy} \neq 0$, то X и Y являются зависимыми случайными величинами;
- 3) $|r_{xy}| = 1$ только в том случае, когда случайные величины X и Y связаны линейной зависимостью $Y = aX + b$, т.е. $r_{xy} = \text{sgn } a$.

Величины X и Y называются линейно коррелированными, если линии регрессии являются прямыми. Уравнения прямых имеют вид

$$y_x = m_y + r_{xy} \frac{\sigma_y}{\sigma_x} (x - m_x); \quad x_y = m_x + r_{xy} \frac{\sigma_x}{\sigma_y} (y - m_y). \quad (1.44)$$

Обе прямые проходят через точку, которую называют центром совместного распределения случайных величин X и Y .

Примеры решения задач

Пример 1. Сколькими способами можно расселить 8 студентов по трем комнатам: одноместной, трехместной, четырехместной?

Решение. Различные способы расселения по комнатам являются перестановками с повторениями. Воспользуемся формулой (1.5)

$$P_8(1, 3, 4) = \frac{8!}{1! \cdot 3! \cdot 4!} = 5 \cdot 7 \cdot 8 = 280.$$

Пример 2. В партии микросхем 95 стандартных и 5 бракованных. При контроле оказалось, что первые 50 микросхем стандартные. Найти вероятность того, что следующая микросхема будет стандартной.

Решение.

Обозначим событие: A – следующая микросхема после первого этапа проверки будет стандартной. При этом в

партии всего осталось $n = 95 + 5 - 50 = 50$ микросхем, из которых $m = 95 - 50 = 45$ стандартные.

Воспользуемся классическим определением вероятности (формула (1.8))

$$P(A) = \frac{m}{n} = \frac{45}{50} = 0,9.$$

Пример 3. Из разрезной азбуки выкладывается слово СТАТИСТИКА. Затем буквы этого слова перемешиваются и выкладываются в случайном порядке. Найти вероятность того, что снова получится слово СТАТИСТИКА.

Решение.

1-й способ. Событие A состоит в получении нужного слова СТАТИСТИКА. Общее число исходов $n = 10!$. Некоторые буквы в слове повторяются (С – два раза, Т – три раза, А – два раза, И – два раза), поэтому возможны перестановки, при которых слово не изменяется. Их число равно $m = 2!3!2!2! = 48$.

По формуле классической вероятности:

$$P(A) = \frac{48}{10!} = \frac{1}{75600}.$$

2-й способ. В слове СТАТИСТИКА две буквы С ($n_1 = 2$), три буквы Т ($n_2 = 3$), две буквы А ($n_3 = 2$), две буквы И ($n_4 = 2$) и одна буква К ($n_5 = 1$). Всего $n_1 + n_2 + n_3 + n_4 + n_5 = 10$ букв.

Общее число исходов в этом случае равно числу перестановок с повторениями и вычисляется по формуле (1.5)

$$P_{10}(2, 3, 2, 2, 1) = \frac{10!}{2! \cdot 3! \cdot 2! \cdot 2! \cdot 1!} = 75600.$$

По формуле классической вероятности:

$$P(A) = \frac{1}{P_{10}(2, 3, 2, 2, 1)} = \frac{1}{75600}.$$

Пример 4. Точку наудачу бросили на отрезок $[0;1,5]$. Какова вероятность ее попадания в отрезок $[0,3;0,9]$?

Решение.

Введем событие A – попадание случайной точки в отрезок $[0,3;0,9]$. Геометрически множество благоприятствующих исходов представляет собой отрезок $[0,3;0,9]$, а пространство

элементарных исходов – весь отрезок $[0;1,5]$. Длины этих отрезков равны $l(A) = 0,6$ и $l(\Omega) = 1,5$.

Согласно геометрическому определению вероятности

$$P(A) = \frac{l(A)}{l(\Omega)} = \frac{0,6}{1,5} = 0,4.$$

Пример 5. Студент может воспользоваться автобусом одного из двух маршрутов, чтобы добраться в университет. Автобусы первого маршрута следуют с интервалом в 15 минут, второго – с интервалом в 12 минут. Найти вероятность того, что на остановке студент будет ждать автобуса не более 7 минут.

Решение.

1-й способ. Введем обозначения:

A – событие, состоящее в том, что студент ждет автобуса меньше $t = 7$ минут;

t_1 и t_2 – время ожидания автобуса первого и второго маршрутов соответственно.

На оси OX отложим время прихода первого автобуса, а на оси OY – второго. Так как выполняются условия $0 \leq t_1 \leq 15$, $0 \leq t_2 \leq 12$, то пространство элементарных событий Ω совпадает с прямоугольником со сторонами 15 и 12 минут (рис. 1.1).

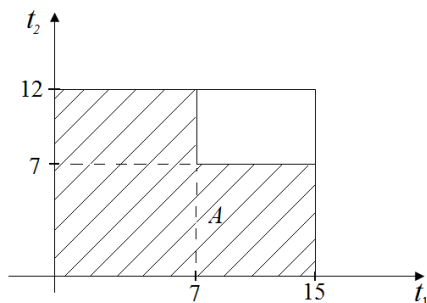


Рис. 1.1. Событие A и пространство элементарных событий Ω

Все благоприятствующие событию A исходы определяются двумя неравенствами:

$$0 \leq t_1 \leq t, 0 \leq t_2 \leq t,$$

что соответствует заштрихованной области.

Согласно геометрическому определению вероятности (формула (1.9))

$$P(A) = \frac{S(A)}{S(\Omega)} = \frac{12 \cdot 15 - (15 - 7)(12 - 7)}{12 \cdot 15} = \frac{7}{9}.$$

2-й способ. Введем обозначения:

A_1 – событие, состоящее в том, что студент ждет первого автобуса меньше $t = 7$ минут;

A_2 – событие, состоящее в том, что студент ждет второго автобуса меньше $t = 7$ минут.

Тогда событие A соответствует наступлению хотя бы одного из событий A_1 и A_2 :

$$A = A_1 + A_2.$$

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2) = \frac{7}{15} + \frac{7}{12} - \frac{7}{15} \cdot \frac{7}{12} = \frac{7}{9}.$$

Пример 6. Вероятность безотказной работы блока автоматики в течение заданного времени равна 0,8. Для повышения надежности устанавливается такой же резервный блок. Найти вероятность безотказной работы системы с резервным блоком.

Решение.

Обозначим события:

A – система безотказно работает в течение заданного времени;

A_1 – исходный блок безотказно работает в течение заданного времени;

$\bar{A}_1$ – исходный блок вышел из строя в течение заданного времени;

A_2 – резервный блок безотказно работает в течение заданного времени;

$\bar{A}_2$ – резервный блок вышел из строя в течение заданного времени.

$$A = A_1 \bar{A}_2 + \bar{A}_1 A_2 + A_1 A_2.$$

По формуле суммы несовместных событий (формула (1.11)):

$$\begin{aligned} p(A) &= p(A_1) p(\bar{A}_2) + p(\bar{A}_1) p(A_2) + p(A_1) p(A_2) = \\ &= 0,8 \cdot 0,2 + 0,2 \cdot 0,8 + 0,8 \cdot 0,8 = 0,96. \end{aligned}$$

Пример 7. В урне 5 белых и 2 черных шаров. Из урны вынимают сразу два шара. Найти вероятность того, что эти шары будут разных цветов.

Решение.

Введем события:

A – извлечен белый шар; B – извлечен черный шар; C – извлечены шары разных цветов. Для события C возможны два варианта: (белый, черный) и (черный, белый), т.е. C – есть сумма двух несовместных событий

$$C = AB + BA.$$

По теореме умножения вероятностей (формула (1.12)):

$$P(AB) = P(A) \cdot P(B/A) = \frac{5}{5+2} \cdot \frac{2}{5+2-1} = \frac{5}{21}.$$

$$P(BA) = P(B) \cdot P(A/B) = \frac{2}{5+2} \cdot \frac{5}{5+2-1} = \frac{5}{21}.$$

По формуле сложения несовместных событий:

$$p(C) = p(AB) + p(BA) = \frac{10}{21}.$$

Пример 8. В ящике перемешаны 5 черных и 6 белых шаров. Случайным образом вынимают 4 шара. Найти вероятность того, что среди них имеется:

- а) 2 белых шара;
- б) меньше чем 2 белых шара;
- в) хотя бы один белый шар.

Решение.

Испытанием будет случайное вынимание четырех шаров. Элементарными событиями являются всевозможные сочетания по 4 из 11 шаров. Их число равно

$$n = C_{11}^4 = \frac{8 \cdot 9 \cdot 10 \cdot 11}{1 \cdot 2 \cdot 3 \cdot 4} = 330.$$

A . Событие A_1 – среди вынутых шаров 2 белых. Значит, среди вынутых шаров 2 белых и 2 черных. Используя правило умножения, получаем

$$m = C_6^2 \cdot C_5^2 = \frac{5 \cdot 6}{1 \cdot 2} \cdot \frac{4 \cdot 5}{1 \cdot 2} = 15 \cdot 10 = 150,$$

$$P(A_1) = \frac{150}{330} = \frac{5}{11}.$$

Б. Событие A_2 – среди вынутых шаров меньше чем 2 белых шара. Это событие состоит из двух несовместных событий:

B_1 – среди вынутых шаров только один белый и 3 черных шара;

B_2 – среди вынутых шаров нет ни одного белого, все 4 шара черные:

$$A_2 = B_1 + B_2.$$

Так как события B_1 и B_2 несовместны, то:

$$P(A_2) = P(B_1) + P(B_2);$$

$$m_1 = C_6^1 \cdot C_5^3 = \frac{6 \cdot 4 \cdot 5}{1 \cdot 1 \cdot 2} = 60, \quad m_2 = C_5^4 = 5;$$

$$P(B_1) = \frac{60}{330} = \frac{5}{11}; \quad P(B_2) = \frac{5}{330}.$$

$$P(A_2) = \frac{60}{330} + \frac{5}{330} = \frac{65}{330} = \frac{13}{66}.$$

В. Событие A_3 – среди вынутых шаров хотя бы один белый шар.

Найдем вероятность противоположного события $\overline{A_3}$ – среди вынутых шаров нет ни одного белого

$$m = C_5^4 = 5, \quad P(\overline{A_3}) = \frac{5}{330} = \frac{1}{66}.$$

Тогда

$$P(A_3) = 1 - P(\overline{A_3}) = 1 - \frac{1}{66} = \frac{65}{66}.$$

Пример 9. Устройство состоит из трех независимых элементов, работающих в течение времени T безотказно соответственно с вероятностями 0,851, 0,751, и 0,701. Найти вероятность того, что за время T выйдет из строя:

- только один элемент;
- хотя бы один элемент.

Решение.

А. Обозначим события:

A_1 – за время T выходит из строя только один элемент;

B_1 – первый элемент выходит из строя;

B_2 – второй элемент выходит из строя;

B_3 – третий элемент выходит из строя;

$\overline{B_1}$ – первый элемент не выходит из строя;

$\overline{B_2}$ – второй элемент не выходит из строя;

$\overline{B_3}$ – третий элемент не выходит из строя.

$$A_1 = (B_1 \overline{B_2} \overline{B_3}) + (\overline{B_1} B_2 \overline{B_3}) + (\overline{B_1} \overline{B_2} B_3).$$

Учитывая независимость элементов устройств и несовместность событий B_i и $\overline{B_i}$, получаем следующую формулу:

$$P(A_1) = P(B_1) \cdot P(\overline{B_2}) \cdot P(\overline{B_3}) + P(\overline{B_1}) \cdot P(B_2) \cdot P(\overline{B_3}) + \\ + P(\overline{B_1}) \cdot P(\overline{B_2}) \cdot P(B_3).$$

По условию задачи $P(\overline{B_1}) = 0,851$, $P(\overline{B_2}) = 0,751$, $P(\overline{B_3}) = 0,701$, поэтому

$$P(B_1) = 0,149, P(B_2) = 0,249, P(B_3) = 0,299. \\ P(A_1) = 0,149 \cdot 0,751 \cdot 0,701 + 0,851 \cdot 0,249 \cdot 0,701 + \\ + 0,851 \cdot 0,751 \cdot 0,299 = 0,418.$$

Б. Обозначим события:

A_2 – за время T выходит из строя хотя бы один элемент;

$\overline{A_2}$ – за время T все элементы работают безотказно:

$$\overline{A_2} = \overline{B_1} \cdot \overline{B_2} \cdot \overline{B_3},$$

$$P(\overline{A_2}) = P(\overline{B_1}) \cdot P(\overline{B_2}) \cdot P(\overline{B_3}) = 0,851 \cdot 0,751 \cdot 0,701 = 0,448,$$

$$P(A_2) = 1 - P(\overline{A_2}) = 1 - 0,448 = 0,552.$$

Пример 10. Электрическая схема состоит из трех независимо работающих элементов. Вероятности отказов каждого из элементов равны $p_1 = 0,1$; $p_2 = 0,2$; $p_3 = 0,3$.

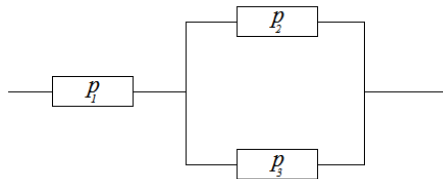


Рис. 1.2. Схема соединения элементов

Найти

а) вероятность отказа схемы;

б) схема отказала, какова вероятность отказа 1-го элемента?

3-го элемента?

Решение.

Обозначим события:

A – отказала схема;

A_k – отказал k -й элемент ($k = 1, 2, 3$).

а) событие $A = A_1 + A_2 A_3$, при этом A_1 и $A_2 A_3$ совместные события

$$p(A) = p(A_1 + A_2 A_3) = p(A_1) + p(A_2 A_3) - p(A_1 A_2 A_3);$$

$$p(A) = 0,1 + 0,06 - 0,006 = 0,154;$$

б) отказ 1-го элемента определим по формуле условной вероятности (формула 1.12) с учетом $A_1 \subset A$:

$$P(A_1/A) = \frac{P(A_1 A)}{P(A)} = \frac{P(A_1)}{P(A)} = \frac{0,2}{0,154} = \frac{1}{77};$$

$$\begin{aligned} P(A_3/A) &= \frac{P(A_3 A)}{P(A)} = \frac{P(A_1 A_3 + A_2 A_3)}{P(A)} = \frac{P(A_1 A_3) + P(A_2 A_3) - P(A_1 A_2 A_3)}{P(A)} = \\ &= \frac{0,03 + 0,06 - 0,006}{0,154} = \frac{3}{11}. \end{aligned}$$

В этом случае $A_3 \not\subset A$.

Пример 11. Однотипные приборы выпускается тремя заводами в количественном отношении $n_1 : n_2 : n_3$. Вероятности брака для этих заводов соответственно равны p_1, p_2 , и p_3 . Прибор, купленный лабораторией, оказался бракованным. Какова вероятность того, что это прибор изготовлен соответственно первым, вторым, третьим заводом?

Значения параметров вычислить по следующим формулам:

$$k = |14 - N|, n_1 = 5 + 2k, n_2 = 20 - k, n_3 = 25 - k;$$

$$p_1 = 0,01 + k/200; p_2 = 0,02 + k/320; p_3 = 0,03 + k/400.$$

Решение. Для $N = 30$ параметры равны:

$$k = 16, n_1 = 37, n_2 = 4, n_3 = 9, p_1 = 0,09, p_2 = 0,07, p_3 = 0,07.$$

Первым испытанием является выбор прибора, вторым проверка прибора. Обозначим события:

A – прибор оказался бракованным;

H_1 – прибор взят из продукции 1-го завода;

H_2 – прибор взят из продукции 2-го завода;

H_3 – прибор взят из продукции 3-го завода.

Вероятность события A вычисляем по формуле полной вероятности:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + P(H_3)P(A/H_3).$$

Условные вероятности заданы в условии задачи:

$$P(A/H_1) = 0,09, P(A/H_2) = 0,02, P(A/H_3) = 0,07.$$

Вероятности гипотез (прибор выпущен таким-то заводом):

$$n_1 + n_2 + n_3 = 50, n_1 : n_2 : n_3 = 37 : 4 : 9,$$

$$P(H_1) = \frac{37}{50}, P(H_2) = \frac{4}{50}, P(H_3) = \frac{9}{50}.$$

По формуле полной вероятности вычисляем вероятность события A

$$P(A) = \frac{37}{50} \cdot 0,09 + \frac{4}{50} \cdot 0,02 + \frac{9}{50} \cdot 0,07 = 0,085.$$

По формуле Байеса вычисляем условные вероятности событий (гипотез):

$$P(H_1/A) = \frac{0,74 \cdot 0,09}{0,085} = 0,784;$$

$$P(H_2/A) = \frac{0,08 \cdot 0,02}{0,085} = 0,066;$$

$$P(H_3/A) = \frac{0,18 \cdot 0,07}{0,085} = 0,15.$$

Пример 12. При передаче сообщения вероятность искажения одного знака равна 0,1. Найти вероятность того, что сообщение из 10 знаков: а) не будет искажено; б) содержит три искажения; в) содержит не менее трех искажений.

Решение.

Воспользуемся формулой Бернулли

$$P_n(m) = C_n^m p^m q^{n-m} \frac{n!}{m!(n-m)!} p^m q^{n-m}$$

при $n = 10$, $p = 0,1$ и $q = 0,9$.

а) $P_{10}(0) = C_{10}^0 0,1^0 \cdot 0,9^{10} = 0,349$;

б) $P_{10}(3) = C_{10}^3 0,1^3 \cdot 0,9^7 = 0,057$;

в) $P_{10}(k \geq 3) = 1 - P_{10}(0) - P_{10}(1) - P_{10}(2) = 1 - C_{10}^0 0,1^0 \cdot 0,9^{10} - C_{10}^1 0,1^1 \cdot 0,9^9 - C_{10}^2 0,1^2 \cdot 0,9^8 = 1 - 0,349 - 0,387 - 0,194 = 0,07$.

Пример 13. В каждом из 700 независимых испытаний событие A происходит с постоянной вероятностью 0,35. Найти вероятность того, что событие A происходит:

а) точно 270 раз;

б) меньше чем 270 и больше чем 230 раз;

в) больше чем 270 раз.

Решение.

Учитывая, что количество испытаний $n = 700$ довольно велико, можно использовать теорему Лапласа.

а) $n = 700$, $p = 0,35$, $m = 270$.

Используем локальную теорему Лапласа (формула 1.18)

$$\sqrt{npq} = \sqrt{700 \cdot 0,35 \cdot 0,65} = \sqrt{159,25} = 12,6;$$

$$x = \frac{270 - 700 \cdot 0,35}{12,6} = \frac{25}{12,6} = 1,98.$$

Значение функции Гаусса $\phi(1,98)$ или вычислим по формуле

$$\phi(1,98) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(1,98)^2}{2}} = 0,05618;$$

$$P_{700}(270) = 0,05618 / 12,6 = 0,00446.$$

б) $n = 700$, $p = 0,35$, $m_1 = 230$, $m_2 = 270$.

Вычислим

$$\sqrt{npq} = 12,6; \quad x_1 = \frac{230 - 700 \cdot 0,35}{12,6} = 1,19;$$

$$x_2 = \frac{270 - 700 \cdot 0,35}{12,6} = 1,98.$$

Воспользуемся формулой (1.19). Значения функции $\Phi(x)$ найдем из таблицы (прил. 1):

$$P_{700}(230 < m < 270) = \Phi(1,98) - \Phi(-1,19) = \Phi(1,98) + \Phi(1,19) = 0,4761 + 0,3830 = 0,8591.$$

в) $n = 700, p = 0,35, m_1 = 270, m_2 = 700$.

Поскольку

$$\sqrt{npq} = 12,6; x_1 = 1,98, x_2 = \frac{700 - 700 \cdot 0,35}{12,6} = 36,1;$$

$$P_{700}(270 < m < 700) = \Phi(36,1) - \Phi(1,98) = 0,5 - 0,4761 = 0,0239.$$

Пример 14. На телефонной станции неправильное соединение происходит с вероятностью $1/200$. Найти вероятность того, что среди 200 соединений произойдет:

- а) точно 1 неправильное соединение;
- б) меньше чем 3 неправильных соединения;
- в) больше чем 2 неправильных соединения.

Решение.

Здесь вероятность события мала, поэтому используем формулу Пуассона.

а) $n = 200, p = 1/200, m = 1$.

Так как $\lambda = np = 200 \cdot \frac{1}{200} = 1 < 10$, то вероятность того, что среди 200 соединений ровно одно неправильное соединение найдем по формуле Пуассона (1.20)

$$P_{200}(1) = \frac{1^1}{1!} e^{-1} = 0,3679.$$

б) $n = 200, p = 1/200, m < 3$.

Искомую вероятность запишем в виде

$$P_n(m < 3) = P_n(m = 0) + P_n(m = 1) + P_n(m = 2).$$

$$\begin{aligned} P_{200}(m < 3) &= P_{200}(0) + P_{200}(1) + P_{200}(2) = \\ &= 0,3679 + 0,3679 + 0,1839 = 0,9197. \end{aligned}$$

в) $n = 200, p = 1/200, m > 2$.

Искомую вероятность запишем в виде

$$P_n(m > 2) = 1 - P_n(m \leq 2) = 1 - P_n(m < 3).$$

$$P_n(m > 2) = 1 - P_n(m < 3) = 1 - 0,9197 = 0,0803.$$

Пример 15. Случайная величина X задана рядом распределения

x_i	3	5	7	11
p_i	0,14	0,20	0,49	?

Найти функцию распределения $F(x)$ случайной величины X и построить ее график. Вычислить для X ее среднее значение $M(X)$, дисперсию $D(X)$ и моду Mo .

Решение.

Вероятность p_4 найдем из условия

$$\sum_{i=1}^4 p_i = p_1 + p_2 + p_3 + p_4 = 1 \Rightarrow p_4 = 1 - 0,14 - 0,20 - 0,49 = 0,17.$$

Функцию распределения находим из ряда распределения по формуле (1.26):

$$F(x) = \begin{cases} 0, & x \leq 3, \\ 0,14, & 3 < x \leq 5, \\ 0,34, & 5 < x \leq 7, \\ 0,83, & 7 < x \leq 11, \\ 1, & x > 11. \end{cases}$$

Построим график функции распределения $F(x)$

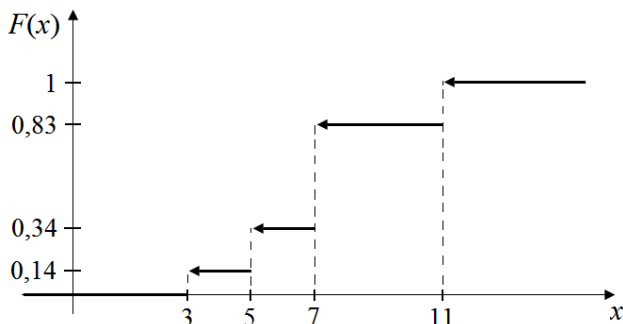


Рис. 1.3. Функция распределения $F(x)$

Среднее значение $M(X)$ вычислим по формуле (1.21):

$$M(X) = 3 \cdot 0,14 + 5 \cdot 0,2 + 7 \cdot 0,49 + 11 \cdot 0,17 = 6,72.$$

Для нахождения дисперсии воспользуемся формулой (1.22):

$$M(X^2) = 3^2 \cdot 0,14 + 5^2 \cdot 0,2 + 7^2 \cdot 0,49 + 11^2 \cdot 0,17 = 50,84;$$

$$D(X) = 50,84 - 6,72^2 = 5,6816.$$

Моду Mo найдем по максимальной вероятности: $Mo = 7$.

Пример 16. Дана функция

$$f(x) = \begin{cases} 0, & x < 0, \\ Cx, & 0 \leq x \leq 4, \\ 0, & x > 4. \end{cases}$$

При каком значении C эта функция является плотностью распределения случайной величины X ? Построить функцию распределения $F(x)$, а также графики функций $f(x)$ и $F(x)$. Вычислить для X ее среднее значение $M(X)$, дисперсию $D(X)$, моду Mo и медиану Me .

Решение.

Константу C найдем из условия нормировки

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^0 f(x) dx + \int_0^4 f(x) dx + \int_4^{\infty} f(x) dx = 1.$$

$$\int_{-\infty}^0 0 dx + \int_0^4 Cx dx + \int_4^{\infty} 0 dx = C \frac{x^2}{2} \Big|_0^4 = 8C = 1 \Rightarrow C = \frac{1}{8}.$$

Из свойства плотности распределения

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

При $x < 0$ плотность распределения $f(x) = 0$, поэтому $F(x) = 0$.

Рассмотрим промежуток $0 \leq x < 4$

$$F(x) = \int_0^x f(x) dx = \frac{1}{8} \int_0^x x dx = \frac{x^2}{16}.$$

Подставив в полученную формулу $x = 4$, получим, что $F(4) = 1$. Так как $F(x)$ является неубывающей функцией, то при $x > 4$ выполняется неравенство $F(x) \geq 1$, а из того, что $F(x)$ представляет собой вероятность, следует что $F(x) \leq 1$. Поэтому $F(x) = 1$ при $x > 4$.

Получаем следующее выражение для функции распределения $F(x)$:

$$F(x) = \begin{cases} 0, & x < 0, \\ x^2 / 16, & 0 \leq x \leq 4, \\ 1, & x > 4. \end{cases}$$

Построим графики функции плотности вероятности $f(x)$ и функции распределения $F(x)$.

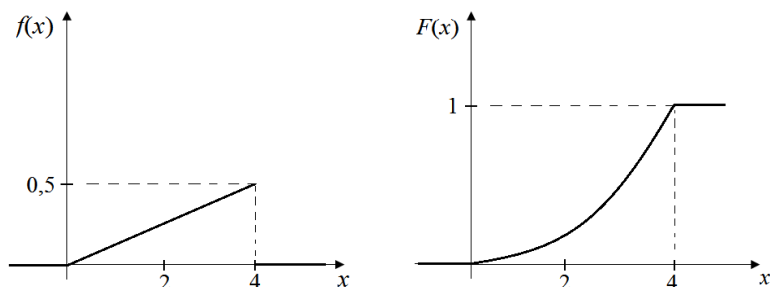


Рис. 1.4. Плотности вероятности $f(x)$ и функции распределения $F(x)$

Среднее значение X вычисляем по формуле (1.24)

$$M(X) = \int_0^4 x \cdot \frac{x}{8} dx = \frac{x^3}{24} \Big|_0^4 = \frac{8}{3}.$$

Дисперсию найдем по упрощенной формуле

$$D(X) = M(X^2) - m_x^2 = \int_0^4 x^2 \cdot \frac{x}{8} dx - \left(\frac{8}{3}\right)^2 = \frac{x^4}{32} \Big|_0^4 - \frac{64}{9} = 8 - \frac{64}{9} = \frac{8}{9}.$$

Плотность распределения достигает максимума в точке $x = 4$ и, значит, $Mo = 4$.

Медиану Me найдем из условия

$$F(Me) = \frac{1}{2} \Rightarrow \frac{x^2}{16} = \frac{1}{2} \Rightarrow x^2 = 8 \Rightarrow x = \pm 2\sqrt{2}; \quad Me = 2\sqrt{2}.$$

Пример 17. Время безотказной работы прибора подчинено показательному закону распределения с плотностью

$$f(x) = \begin{cases} 0,6e^{-0,6x} & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Найти среднее время безотказной работы и вероятность того, что за 5 часов не будет более одного отказа.

Решение.

Введем обозначения:

X – случайная величина, равная времени безотказной работы прибора;

Y – число отказов за время t .

Случайная величина X распределена по показательному закону с параметром $\lambda = 0,6$.

Среднее время безотказной работы равно математическому ожиданию случайной величины X

$$M(X) = \frac{1}{\lambda} = \frac{1}{0,6} = 1,67.$$

Дискретная случайная величина Y принимает целые значения $m = 0, 1, 2, \dots$ с вероятностями

$$P(Y = m) = \frac{a^m e^{-a}}{m!},$$

здесь параметр распределения $a = n \cdot \lambda = 5 \cdot 0,6 = 3$.

Отсюда получаем

$$P(Y \leq 1) = P(Y = 0) + P(Y = 1) = \frac{3^0 e^{-3}}{0!} + \frac{3^1 e^{-3}}{1!} = 0,199.$$

Пример 18. Дальномер имеет систематическую ошибку 0,1 м и среднюю квадратическую ошибку 0,4 м. Полагая, что ошибка измерения подчиняется нормальному закону распределения, найти вероятность того, что ошибка измерения не превысит 0,5 м.

Решение.

Так как по условию задачи $m = 0,1$, $\sigma = 0,4$, то искомая вероятность может быть вычислена по формуле (1.35)

$$\begin{aligned} P(|X| < 0,5) &= P(-0,5 < X < 0,5) = \Phi\left(\frac{0,5 - 0,1}{0,4}\right) - \Phi\left(\frac{-0,5 - 0,1}{0,4}\right) = \\ &= \Phi(1) - \Phi(-1,5) = \Phi(1) + \Phi(1,5) = 0,3413 + 0,4332 = 0,7745. \end{aligned}$$

Пример 19. Задана случайная величина $X \in N(0, 2)$. Найти вероятность того, что эта случайная величина принимает значение:

- а) в интервале $[-1, 2]$;
- б) меньше -1;
- в) больше 2;
- г) отличающееся от своего среднего значения по абсолютной величине не больше чем на 1.

Решение.

В первых трех случаях можно воспользоваться формулой (1.35)

$$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - m}{\sigma}\right) - \Phi\left(\frac{x_1 - m}{\sigma}\right).$$

а) $m = 0; \sigma = 2; x_1 = -1; x_2 = 2$

$$P(-1 \leq X \leq 2) = \Phi\left(\frac{2-0}{2}\right) - \Phi\left(\frac{-1-0}{2}\right) = \Phi(1) - \Phi(-0,5) = \\ = \Phi(1) + \Phi(0,5) = 0,3413 + 0,1915 = 0,5328.$$

б) $m = 0; \sigma = 2; x_1 = -\infty; x_2 = -1$

$$P(X < -1) = P(-\infty < X < -1) = \Phi\left(\frac{-1-0}{2}\right) - \Phi(-\infty) = \\ = \Phi(-0,5) + \Phi(\infty) = -0,1915 + 0,5 = 0,3085.$$

в) $m = 0; \sigma = 2; x_1 = 2; x_2 = \infty$

$$P(X \geq 2) = P(2 \leq X < \infty) = \Phi\left(\frac{\infty-0}{2}\right) - \Phi\left(\frac{2-0}{2}\right) = \\ = \Phi(\infty) - \Phi(1) = 0,5 - 0,3413 = 0,1587.$$

г) $m = 0; \sigma = 2; \varepsilon = 1$

Вероятность попадания в симметричный интервал найдем по формуле

$$P(|X - m| \leq \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right);$$

$$P(|X - 0| \leq 1) = 2\Phi\left(\frac{1}{2}\right) = 2 \cdot 0,1915 = 0,3830.$$

Пример 20. Закон распределения двумерной случайной величины (X, Y) имеет вид:

$X \backslash Y$	0	1	2
-1	$5/(N+10)$	$N/(5N+50)$	$N/(5N+50)$
0	$N/(5N+50)$	$2/(N+10)$	$N/(5N+50)$
1	$N/(5N+50)$	?	$2/(N+10)$

Заполнить свободную ячейку в таблице при заданном значении N (номер варианта), найти законы распределения составляющих двумерной случайной величины. Выяснить, зависимы случайные величины X и Y или нет? Найти $M(X)$, $M(Y)$, $D(X)$, $D(Y)$, $M(X+Y)$, $D(X+Y)$ и коэффициент корреляции r_{xy} .

Решение.

Для $N = 30$ вычислим соответствующие вероятности p_{ij} .

$X \backslash Y$	0	1	2
-1	$p_{11} = \frac{1}{8}$	$p_{12} = \frac{3}{20}$	$p_{13} = \frac{3}{20}$
0	$p_{21} = \frac{3}{20}$	$p_{22} = \frac{1}{20}$	$p_{23} = \frac{3}{20}$
1	$p_{31} = \frac{3}{20}$	$p_{32} = ?$	$p_{33} = \frac{1}{20}$

Вероятность p_{32} найдем из условия $\sum_{i=1}^3 \sum_{j=1}^3 p_{ij} = 1$:

$$p_{32} = 1 - p_{11} - p_{12} - p_{13} - p_{21} - p_{22} - p_{23} - p_{31} - p_{33} = \frac{1}{40}.$$

Распределения одномерных случайных величин X и Y определим по формулам (1.34).

Для случайной величины X вероятности ее значений $p_i = P(X = x_i)$ находятся суммированием вероятностей p_{ij} в i -й строке таблицы ($i = 1, 2, 3$)

x_i	-1	0	1
p_i	17/40	7/20	9/40

Для случайной величины Y вероятности ее значений $q_j = P(Y = y_j)$ находятся суммированием вероятностей p_{ij} в j -м столбце таблицы ($j = 1, 2, 3$)

y_j	0	1	2
q_j	17/40	9/40	7/20

Условием независимости случайных величин X и Y является равенство:

$$p_{ij} = p_i q_j, \text{ для всех } i, j = 1, 2, 3.$$

Так как в данном случае

$$p_{11} = P(X = -1, Y = 0) = \frac{1}{8}; \quad p_1 = P(X = -1) = \frac{17}{40}; \quad q_1 = P(Y = 0) = \frac{17}{40},$$

то $p_{11} \neq p_1 q_1$ и, следовательно, случайные величины X и Y зависимы.

Вычислим математические ожидания и дисперсии по формулам (1.21) и (1.22), используя одномерные законы

$$m_x = (-1) \cdot \frac{17}{40} + 0 \cdot \frac{7}{20} + 1 \cdot \frac{9}{40} = -\frac{8}{40}; \quad m_y = 0 \cdot \frac{17}{40} + 1 \cdot \frac{9}{40} + 2 \cdot \frac{7}{20} = \frac{37}{40}.$$

Для нахождения дисперсий формулу (1.19) представим в виде

$$D(X) = \sum_{i=1}^3 x_i^2 p_i - m_x^2 = (-1)^2 \cdot \frac{17}{40} + 0^2 \cdot \frac{7}{20} + 1^2 \cdot \frac{9}{40} - \left(-\frac{8}{40}\right)^2 = 0,61;$$

$$D(Y) = \sum_{j=1}^3 y_j^2 q_j - m_y^2 = 0^2 \cdot \frac{17}{40} + 1^2 \cdot \frac{9}{40} + 2^2 \cdot \frac{7}{20} - \left(\frac{37}{40}\right)^2 = 0,769.$$

$$\sigma_x = \sqrt{D(X)} = 0,781; \quad \sigma_y = \sqrt{D(Y)} = 0,877.$$

Ковариацию вычислим по формуле (1.42), предварительно записав в виде:

$$\begin{aligned} \text{cov}(X, Y) &= \sum_{i=1}^3 \sum_{j=1}^3 x_i y_j p_{ij} - m_x \cdot m_y = (-1) \cdot 0 \cdot \frac{1}{8} + (-1) \cdot 1 \cdot \frac{3}{20} + (-1) \cdot 2 \cdot \frac{3}{20} + \\ &+ 0 \cdot 1 \cdot \frac{1}{20} + 0 \cdot 2 \cdot \frac{3}{20} + 1 \cdot 0 \cdot \frac{3}{20} + 1 \cdot 1 \cdot \frac{1}{40} + 1 \cdot 2 \cdot \frac{1}{20} - \left(-\frac{8}{40}\right) \left(\frac{37}{40}\right) = -0,14. \end{aligned}$$

Коэффициент корреляции между X и Y

$$r_{xy} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{M[(X - m_x) \cdot (Y - m_y)]}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} = \frac{-0,14}{0,781 \cdot 0,877} = -0,204.$$

$$M(X + Y) = M(X) + M(Y) = -\frac{8}{40} + \frac{37}{40} = \frac{29}{40};$$

$$D(X + Y) = D(X) + D(Y) + 2 \text{cov}(X, Y) = 0,61 + 0,769 + 2(-0,14) = 1,1.$$

Пример 21. Двумерная случайная величина (X, Y) распределена равномерно в области D с плотностью

$$f(x, y) = \begin{cases} \frac{1}{S_D}, & (x, y) \in D, \\ 0, & (x, y) \notin D. \end{cases}$$

Найти $M(X)$, $M(Y)$, $D(X)$, $D(Y)$, стандартные отклонения и коэффициент корреляции r_{xy} . Найти уравнение линий регрессии

X на Y и Y на X и построить их. Область D определяется неравенствами: $\frac{x}{15} + \frac{y}{10} \leq 1; x \geq 0, y \geq 0$.

Решение. Площадь области S_D вычислим из основного свойства двумерной плотности

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

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

Интеграл равен площади треугольника с основанием 15 и высотой 10, т.е. $S_D = \frac{1}{2} \cdot 10 \cdot 15 = 75$.

Одномерные плотности распределения составляющих двумерной случайной величины имеют вид

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{75} \int_0^{10-2x/3} dy = \frac{10-2x/3}{75};$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx = \frac{1}{75} \int_0^{15-3y/2} dx = \frac{15-3y/2}{5}.$$

Кроме того, $f_1(x) = 0$ при $x \notin [0, 15]$ и $f_2(y) = 0$ при $y \notin [0, 10]$.

Так как $f(x, y) \neq f_1(x) \cdot f_2(y)$, то случайные величины X и Y зависимы.

Следовательно,

$$m_x = \frac{1}{75} \int_0^{15} (10 - \frac{2}{3}x)x dx = 5; \quad m_y = \frac{1}{75} \int_0^{10} y(15 - \frac{3}{2}y) dy = \frac{10}{3}.$$

$$D(X) = \frac{1}{75} \int_0^{15} x^2 (10 - \frac{2}{3}x) dx - 5^2 = \frac{75}{2} - 25 = \frac{25}{2}; \quad \sigma(X) = \frac{5}{\sqrt{2}}.$$

$$D(Y) = \frac{1}{75} \int_0^{10} y^2 (15 - \frac{3}{2}y) dy - \left(\frac{10}{3}\right)^2 = \frac{50}{3} - \frac{100}{9} = \frac{50}{9}, \quad \sigma(Y) = \frac{5\sqrt{2}}{3}.$$

Таким образом, центром рассеивания является точка $O(5; 10/3)$.

$$\text{cov}(X, Y) = \iint_D xy f(x, y) dx dy - m_x m_y = \frac{1}{75} \int_0^{15} dx \int_0^{\frac{30-2x}{3}} xy dy = \frac{25}{2} - 5 \cdot \frac{10}{3} = -\frac{25}{6}.$$

$$r_{xy} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = -\frac{25}{6} \bigg/ \left(\frac{5}{\sqrt{2}} \cdot \frac{5\sqrt{2}}{3} \right) = -0,5.$$

Линии регрессии найдем по формулам (1.40)

$$y - m_y = r_{xy} \frac{\sigma_y}{\sigma_x} (x - m_x), \quad y = \frac{10}{3} + (-0,5) \frac{5\sqrt{2} \cdot \sqrt{2}}{3 \cdot 5} (x - 5) = 5 - x/3,$$

$$x - m_x = r_{xy} \frac{\sigma_x}{\sigma_y} (y - m_y), \quad x = 5 + (-0,5) \frac{5 \cdot 3}{\sqrt{2} \cdot 5 \cdot \sqrt{2}} (y - \frac{10}{3}) = 7,5 - \frac{3}{4} y.$$

Линии регрессии приведены на рис. 1.5.

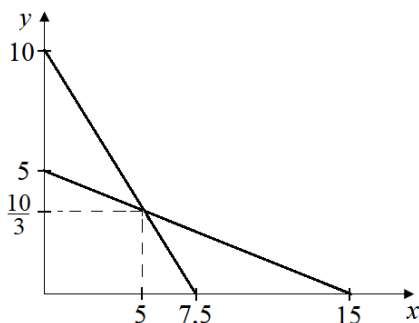


Рис. 1.5. Линии регрессии

Задачи для самостоятельного решения

1. Четыре студента сдают экзамен. Сколькими способами им могут быть выставлены положительные оценки?

Ответ: 81.

2. Среди перестановок цифр числа 12345 сколько есть таких, которые не заканчиваются пятеркой?

Ответ: 96.

3. Сколько сигналов можно подать с помощью шести флажков разного цвета?

Ответ: 1956.

4. Найти вероятность того, что среди шести карт, наудачу взятых из колоды в 36 карт, будет ровно два туза.

Ответ: 0,1.

5. Бросают 2 игральных кубика. Найти вероятности следующих событий: а) сумма выпавших очков равна шести; б) раз-

ность выпавших очков равна двум; в) сумма выпавших очков равна четырем, а разность – единице; г) сумма выпавших очков равна девяти, а произведение – двадцати; в) произведение числа очков не больше 10 .

Ответ: а) $5/36$; б) $2/9$; в) 0; г) $1/18$; д) $19/36$.

6. Какова вероятность того, что корни квадратного уравнения $x^2 + px + q = 0$ будут действительными числами, если коэффициенты p и q уравнения выбираются наудачу на отрезке $[0;1]$?

Ответ: $1/12$.

7. Две точки независимо друг от друга наудачу выбираются на отрезке $[0;1]$. Найти вероятность, что сумма координат точек меньше 1,5.

Ответ: $7/8$.

8. При включении зажигания двигатель начнет работу с вероятностью 0,9. Найти вероятность того, что для запуска двигателя потребуется включить зажигание не более трех раз.

Ответ: 0,999.

9. Вероятность хотя бы одного попадания в цель при четырех выстрелах равна 0,9984. Найти вероятность попадания в цель при одном выстреле.

Ответ: 0,8.

10. В одном ящике 3 белых и 5 черных шаров, в другом – 6 белых и 4 черных шара. Найти вероятность того, что хотя бы из одного ящика будет вынут белый шар, если из каждого ящика вынуто по одному шару.

Ответ: 0,75.

11. Определить надежность технической системы, изображенной на рис. 1.6, по известным вероятностям безотказной работы отдельных ее элементов: $p_1 = 0,8$ и $p_2 = 0,9$.

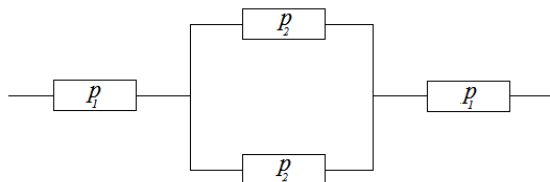


Рис. 1.6. Схема соединения элементов

Ответ: 0,63.

12. Два из четырех независимо работающих двигателя отказали. Найти вероятность того, что отказали первый и второй двигатели. Вероятности отказа двигателей равны соответственно 0,1; 0,2; 0,3; 0,4 .

Ответ: 0,0392.

13. По линии связи передано два сигнала типов A и B с вероятностями соответственно 0,8 и 0,2. В среднем принимается 60% сигналов типа A и 70% типа B . Найти вероятность того, что:

а) посланный сигнал будет принят; б) принятый сигнал – типа A .

Ответ: а) 0,62; б) 0,7742.

14. Батарея из трех орудий произвела залп, при этом два снаряда попали в цель. Найти вероятность того, что первое орудие дало промах, если вероятность попадания в цель для первого орудия 0,4; для второго – 0,3; для третьего – 0,5.

Ответ: 0,31.

15. При массовом производстве полупроводниковых диодов вероятность брака при формовке равна 0,1. Найти вероятность того, что из восьми диодов, проверяемых ОТК, бракованных будет:

а) два; б) не менее двух; в) не более двух.

Ответ: а) 0,1488; б) 0,1871; в) 0,9617.

16. Вероятность попадания в цель при одном выстреле равна 0,2. Найти наименее вероятное число попадания в цель: а) при 15 выстрелах; б) при 9 выстрелах; в) при 7 выстрелах.

Ответ: а) 3; б) 1 и 2; в) 1.

17. Станок состоит из 2000 независимо работающих узлов. Вероятность отказа одного узла в течение года равна 0,0005. Найти вероятность отказа в течение года двух узлов.

Ответ: 0,1838.

18. Вероятность появления событий в каждом из n независимых испытаний равна 0,25. Найти вероятность того, что событие наступит 50 раз в 243 испытаниях.

Ответ: 0,0167.

19. Среднее время работы электронного модуля равно 700 часов. Определить время безотказной работы модуля с надежностью 0,8.

Ответ: 156 ч.

20. Закон распределения случайной величины X задан рядом распределения

x_i	3	5	7	11
p_i	0,14	0,20	0,49	0,17

Для случайной величины найти математическое ожидание, дисперсию и моду.

Ответ: $M(X) = 6,72$; $D(X) = 5,68$; $Mo = 7$.

21. Стрелок производит три выстрела в цель. Вероятность попадания в цель при каждом выстреле равна 0,2. Для случайной величины X (число попадания в цель) найти математическое ожидание, дисперсию и моду.

Ответ: $M(X) = 0,6$; $D(X) = 0,48$; $Mo = 0$.

22. Случайная величина X подчинена закону Пуассона с математическим ожиданием, равным 3. Найти вероятность того, что случайная величина X примет значение, меньшее, чем ее математическое ожидание.

Ответ: 0,423.

23. Все значения равномерно распределенной случайной величины X лежат на отрезке $[2; 8]$. Найти вероятность попадания случайной величины X в промежуток $(3; 5)$.

Ответ: 0,3333.

24. Функция распределения случайной величины имеет вид X

$$F(x) = b + c \operatorname{arctg}(x/a), -\infty < x < \infty.$$

Чему должны быть равны a , b и c ? Найти плотность вероятности.

Ответ: $a > 0$, $b = 0,5$, $c = \frac{1}{\pi}$, $f(x) = \frac{1}{\pi(x^2 + a^2)}$.

25. Из урны, в которой находятся шесть белых и четыре черных шара, наудачу извлекают два шара без возвращения. Случайная величина X – число белых шаров, Y – число черных шаров в выборке. Определить: а) зависимы ли случайные величины X и Y ; б) вычислить вероятность $P\{X \geq Y\}$.

Ответ: а) зависимы, б) $\frac{13}{15}$.

26. Задана плотность распределения системы случайных величин X и Y

$$f(x, y) = \frac{1}{\pi^2(x^2 + y^2 + x^2 y^2 + 1)}.$$

Выяснить, являются ли независимыми случайные величины X и Y .

Ответ: являются.

27. Плотность распределения вероятностей случайного вектора (X, Y) имеет следующий вид:

$$f(x, y) = \begin{cases} C(x+y), & (x, y) \in D, \\ 0, & (x, y) \notin D. \end{cases}$$

Определить: а) коэффициент C ; б) вероятность попадания случайной точки (x, y) в область $D = \{(x, y) : x + y < 1\}$; в) центр рассеивания.

Ответ: а) 1; б) $\frac{1}{3}$; в) $\left(\frac{7}{12}, \frac{7}{12}\right)$.

28. Непрерывная случайная величина X распределена равномерно на отрезке $[-2, 4]$, а непрерывная случайная величина Y нормально с параметрами $N(-1, 2)$. Известно, что коэффициент корреляции $r_{xy} = 0,5$. Найти $M(XY)$.

Ответ: $\sqrt{3} - 1$.

29. На вход измерительного прибора поступает случайный вектор (X, Y) с характеристиками $m_x = -1$, $m_y = 1$, $\sigma_x = 2$, $\sigma_y = 3$, $r_{xy} = 0,5$. На выходе измеряется величина $Z = (2X - 3Y + 5)^2$. Определить математическое ожидание случайной величины Z .

Ответ: 61.

Задания контрольной работы

Основы теории вероятностей

Задание 1. Фамилия и имя студента записаны с помощью карточек. Карточки с буквами смешивают и вынимают по одной карточке без возврата. Найти вероятность того, что буквы вынимаются в порядке следования в фамилии и имени.

Задание 2. В урне содержится K черных и H белых шаров. Случайным образом вынимают M шаров. Найти вероятность того, что среди них имеется:

- а) P белых шаров;
 б) меньше чем P белых шаров;
 в) хотя бы один белый шар.

Таблица 1

Вариант	1	2	3	4	5	6	7	8	9	10	11	12	13	14
K	5	6	6	7	4	8	6	4	5	7	8	6	4	8
H	6	5	5	4	5	6	7	7	6	4	6	5	6	6
M	5	4	5	4	4	5	4	4	5	4	4	4	4	5
P	3	2	3	2	2	3	4	2	3	2	3	3	3	2

Вариант	15	16	17	18	19	20	21	22	23	24	25
K	5	7	5	6	5	6	6	6	8	6	5
H	6	4	7	5	7	7	8	5	6	7	7
M	5	5	4	5	5	5	5	5	5	4	4
P	4	3	3	2	4	3	4	4	3	3	2

Задание 3. Техническое устройство состоит из трех независимых узлов, вероятности безотказной работы которых равны соответственно p_1, p_2, p_3 . Найти вероятность того, что за время T выйдет из строя:

- а) только один элемент;
 б) хотя бы один элемент.

Значения параметров вычислить по формулам (N -номер варианта, соответствует номеру в списке группового журнала):

$$p = \frac{N}{100}; \quad p_1 = 0,8 - p; \quad p_2 = 0,9 - p; \quad p_3 = 1 - p.$$

Задание 4. Однотипные приборы выпускаются тремя заводами в количественном отношении $n_1 : n_2 : n_3$. Вероятности брака для этих заводов соответственно равны p_1, p_2 , и p_3 . Прибор, купленный лабораторией, оказался бракованным. Какова вероятность того, что этот прибор изготовлен соответственно первым, вторым, третьим заводом?

Значения параметров вычислить по следующим формулам:

$$k = |14 - N|, \quad n_1 = 5 + 2k, \quad n_2 = 20 - k, \quad n_3 = 25 - k;$$

$$p_1 = 0,01 + k/200; \quad p_2 = 0,02 + k/320; \quad p_3 = 0,03 + k/400.$$

Задание 5. Вероятность того, что событие A произойдет в течение часа, равна p . Оказалось, что в течение первых t минут событие A не произошло. Какова вероятность того, что событие произойдет в оставшееся время?

$$p = 0,9 - \frac{N}{100}; t = 10 + 2|15 - N|.$$

Задание 6. Вероятность наступления события A в каждом из n испытаний равна p . Найти вероятность того, что событие A произойдет:

- а) ровно m раз;
- б) меньше чем m и больше чем k раз;
- в) больше чем m раз.

Значения параметров вычислить по формулам

$$n = 550 + 4N; \quad m = 250 + 2N; \quad p = 0,3 + \frac{N}{100}; \quad k = m - 5N.$$

Задание 7. На телефонной станции неправильное соединение происходит с вероятностью p . Найти вероятность того, что среди n соединений имеет место:

- а) точно K неправильных соединений;
- б) меньше чем T неправильных соединений;
- в) больше чем P неправильных соединений.

Значения параметров n , p , K , T , и P вычислить по следующим формулам: $p = 1/(3N+200)$, $n = 2n + 200$,

$K = \text{остаток}(N/5) + 1$; $T = \text{остаток}(N/6) + 3$; $P = \text{остаток}(N/8) + 2$.

$G = \text{остаток}(N/5) + 1$; $L = \text{остаток}(N/6) + 3$; $M = \text{остаток}(N/8) + 2$.

Задание 8. Случайная величина X задана законом распределения

x_1	x_2	x_3	x_4
p_1	p_2	p_3	?

Найти функцию распределения случайной величины X и построить ее график. Вычислить математическое ожидание, дисперсию и моду.

Значения параметров вычислить по следующим формулам:

$$k = \text{остаток}(N/4) + 3; x_1 = |N - 15|; x_{i+1} = x_1 + ik, i = 1, 2, 3;$$

$$p_1 = \frac{1}{k+3}, \quad p_2 = \frac{1}{k+4}, \quad p_3 = \frac{1}{8-k}.$$

Задание 9. Случайная величина X задана функцией плотности вероятности

$$f(x) = \begin{cases} 0, & x \leq 0, \\ \frac{x}{2k^2}, & 0 < x \leq c, \\ 0, & x > c. \end{cases}$$

Найти функцию распределения $F(x)$ случайной величины X . Построить графики функций $f(x)$ и $F(x)$. Вычислить математическое ожидание, дисперсию, моду и медиану.

Значения параметров вычислить по следующим формулам:

$$k = 2 + \text{остаток}(N/8), \quad c = 2k.$$

Задание 10. Случайная величина X распределена по нормальному закону $N(m, \sigma^2)$. Найти вероятность того, что случайная величина принимает значение

- а) в интервале (a, b) ;
- б) меньше a ;
- в) больше b .

Значения параметров вычислить по следующим формулам

$$m = \text{остаток}(N/5) + 2; \quad \sigma = \text{остаток}(N/3) + 3; \quad a = m - \sigma; \quad b = m + 0,5\sigma.$$

Задание 11. Закон распределения двумерной случайной величины (X, Y) имеет вид:

$X \backslash Y$	0	1	2
-1	$5/(N+10)$	$N/(5N+50)$	$N/(5N+50)$
0	$N/(5N+50)$	$2/(N+10)$	$N/(5N+50)$
1	$N/(5N+50)$	?	$2/(N+10)$

Заполнить свободную ячейку в таблице, найти распределение каждой из случайных величин X и Y (одномерные законы распределения). Зависимы случайные величины X и Y или нет? Найти $M(X)$, $M(Y)$, $D(X)$, $D(Y)$, $M(X+Y)$, $D(X+Y)$ и коэффициент корреляции r_{xy} .

Задание 12. Двумерная случайная величина (X, Y) распределена равномерно в области D с плотностью:

$$f(x, y) = \begin{cases} \frac{1}{S_D}, & (x, y) \in D \\ 0, & (x, y) \notin D \end{cases}$$

Найти $M(X)$, $M(Y)$, $D(X)$, $D(Y)$, стандартные отклонения и коэффициент корреляции r_{xy} . Найти уравнение линий регрессии X на Y и Y на X и построить их. Область D задается неравенствами:

$$\frac{2x}{N} + \frac{y}{10} \leq 1; x \geq 0, y \geq 0.$$

Вопросы к защите контрольной работы

1. Что изучает комбинаторика? Сформулируйте правило умножения и правило сложения.

2. Поясните понятия: число перестановок, число размещений, число сочетаний. Запишите для них формулы вычислений.

3. Что называется экспериментом (опытом) и событием в теории вероятностей? Приведите примеры достоверного, невозможного и случайного события.

4. Какие события называются равновероятными?

5. Какое событие называется суммой нескольких событий?

6. Какое событие называется произведением нескольких событий?

7. Совместные и несовместные события. Приведите примеры.

8. Когда несколько событий образуют полную группу событий?

9. Что такое вероятность события?

10. В каких числовых пределах находятся значения вероятности любого события?

11. Каковы вероятности невозможного, достоверного противоположного событий?

12. Геометрическое определение вероятности.
13. В чем различие относительной частоты от вероятности?
14. Сформулируйте теорему сложения вероятностей для двух и трех совместных событий; для двух и трех несовместных событий.
15. Дайте определение условной вероятности события.
16. Какие события называются независимыми? Зависимыми?
17. Чему равны условные вероятности независимых событий?
18. Сформулируйте теорему умножения вероятностей для двух и трех зависимых событий; для двух и нескольких независимых событий.
19. Как проверить, являются ли события независимыми?
20. Запишите формулу полной вероятности и объясните смысл входящих в нее вероятностей.
21. Когда применяется формула Байеса?
22. Для решения каких задач используется формула Бернулли?
23. Запишите формулу для вычисления вероятности наступления события m раз при n повторных независимых испытаниях (формула Бернулли).
24. Каким образом можно найти наименее вероятное число наступлений события в повторных независимых испытаниях?
25. Сформулируйте локальную теорему Муавра – Лапласа.
26. Запишите формулу Пуассона. Когда она применяется?
27. Что называется случайной величиной? Приведите примеры случайных величин.
28. Какие случайные величины называются дискретными, а какие непрерывными?
29. Дайте определение функции распределения случайной величины. Изобразите примерный график функции распределения.
30. Сформулируйте основные свойства функции распределения.
31. Изобразите примерный график функции распределения непрерывной случайной величины. Дайте определение плотности распределения.
32. Что называется кривой распределения? Сформулируйте ее основные свойства.
33. Дайте определение математического ожидания случайной величины.
34. Сформулируйте основные свойства математического ожидания.

35. Что показывает дисперсия случайной величины? Сформулируйте основные свойства дисперсии.

36. В чем заключается основной недостаток дисперсии в качестве характеристики разброса (рассеяния) значений случайной величины?

37. Дайте определение среднего квадратического отклонения случайной величины.

38. Поясните понятия: начальный момент, центральный момент, коэффициент асимметрии, эксцесс.

39. Что называется законом распределения случайной величины?

40. Какая случайная величина называется распределенной по биномиальному закону?

41. Запишите выражения для числовых характеристик случайной величины, распределенной по биномиальному закону.

42. Какая случайная величина называется распределенной по показательному закону?

43. Чему равны числовые характеристики случайной величины, распределенной по биномиальному закону?

44. Какая случайная величина называется распределенной по равномерному закону?

45. Запишите выражения для числовых характеристик случайной величины, распределенной по равномерному закону.

46. Запишите выражение плотности распределения нормально распределенной случайной величины.

47. По какой формуле вычисляется вероятность попадания нормально распределенной случайной величины в заданный интервал?

48. Чему равна сумма всех элементов матрицы распределения двумерной случайной величины?

49. Вероятность какого события задает сумма элементов первого столбца матрицы распределения?

50. Вероятность какого события задает сумма элементов первой строки матрицы распределения?

51. Чему равен коэффициент корреляции для независимых случайных величин?

52. Следует ли из независимости случайных величин их некоррелированность, и наоборот?

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. 2. ПЕРВИЧНЫЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ. ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ

Пусть $x_1, x_2, x_3, \dots, x_n$ – выборка объема n из некоторой генеральной совокупности, имеющей функцию распределения $F(x)$. Различные элементы выборки x_i называются **вариантами**.

Ряд вариант $x_1^* \leq x_2^* \leq x_3^* \leq \dots \leq x_n^*$, расположенных в порядке возрастания их значений, называется **вариационным рядом** ($x_1^* = \min(x_n)$; $x_n^* = \max(x_n)$). Им пользуются, в основном, при малых значениях n .

Если n велико, то ряд преобразуют в группировки по отдельным значениям признака x (дискретная группировка) или по интервалам изменения признака (интервальная группировка), для чего разбивают диапазон изменения признака x на k равных интервалов.

Дискретный статистический ряд (дискретная группировка)

x_j^{**}	x_1^{**}	x_2^{**}	x_3^{**}	x_4^{**}	...	x_r^{**}
n_j	n_1	n_2	n_3	n_4	...	n_r

r – число различных вариантов выборки, n_j – число одинаковых значений x_j^{**} , $\sum_{j=1}^r n_j = n$.

Интервальный статистический ряд (интервальная группировка)

I_i	$[a_0, a_1)$	$[a_1, a_2)$	$[a_2, a_3)$	...	$[a_{k-1}, a_k)$
m_i	m_1	m_2	m_3	...	m_k

m_i – число выборочных значений, попадающих в интервал $[a_{i-1}, a_i)$, $\sum_{i=1}^k m_i = n$.

2.1. Построение интервального статистического ряда

Построение интервального (группированного) статистического ряда по заданной выборке $x_1, x_2, x_3, \dots, x_n$ объема n проводят в следующей последовательности.

На первом этапе все множество возможных значений признака X разбивается на k непересекающихся полуоткрытых интервалов

$$I_i = [a_{i-1}, a_i), \quad (2.1)$$

границы которых определяются формулами

$$a_0 = \min \{x_i\}, \quad a_i = a_0 + i \cdot h, \quad i = 1, 2, \dots, k. \quad (2.2)$$

Длина каждого интервала при выбранном числе интервалов, зависящем от объема выборки, ширина интервала равна

$$h = \frac{\max \{x_i\} - \min \{x_i\}}{k}. \quad (2.3)$$

Оптимальное число интервалов, на которые разбивается диапазон выборки, рекомендуется выбирать по одной из формул (все три формулы дают приблизительно одинаковый результат):

$$k = \begin{cases} k < 5 \lg n, & 6 \leq k \leq 20, \\ \left[\sqrt{n} \right], & 5 \leq k \leq 25, \\ \left[1 + 3,322 \lg n \right], & 5 \leq k \leq 25. \end{cases} \quad (2.4)$$

где $[Q]$ – целая часть числа Q .

Эмпирическую частоту попадания элементов выборки в интервал $I_i = [a_{i-1}, a_i]$, ($i = 1, 2, \dots, k$) обозначим m_j . При этом выполняется равенство

$$\sum_{i=1}^k m_i = n. \quad (2.5)$$

Все выборочные значения, попавшие в интервал $I_i, (i = \overline{1, k})$, принимаются равными его середине $\bar{a}_i$

$$\bar{a}_i = \frac{a_{i-1} + a_i}{2}, \quad i = 1, 2, 3, \dots, k, \quad (2.6)$$

а статистические оценки вероятностей (частота), с которыми выборочная случайная величина X принимает значения $\bar{a}_i$, равны

$$p_i^* = \frac{m_i}{n}, \quad (2.7)$$

где m_i – число выборочных значений (наблюдений), попавших в интервал $I_i = [a_{i-1}, a_i]$.

I_i	I_1	I_2	I_3	$\dots$	I_k
$\bar{a}_i$	$\bar{a}_1$	$\bar{a}_2$	$\bar{a}_3$	$\dots$	$\bar{a}_k$
p_i^*	p_1^*	p_2^*	p_3^*	$\dots$	p_k^*

2.2. Построение эмпирической функции и плотности распределения

Эмпирическую функцию распределения $F_n^*(x)$ определяют по формуле

$$F_n^*(x) = \sum_{\bar{a}_i < x} p_i^* \quad , \quad (2.8)$$

которую удобно записать в аналитическом виде через накопленные частоты

[illegible]

Функция $F_n^*(x)$ представляет собой кусочно-постоянную функцию со скачками в середине интервалов $\bar{a}_i$ (рис. 2.1).

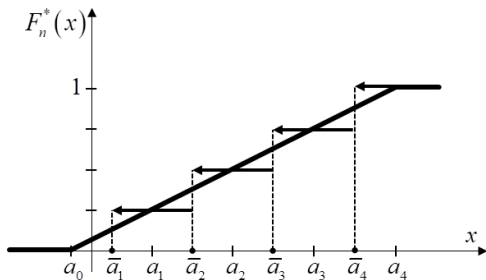


Рис. 2.1. Пример эмпирической функции распределения
при четырех интервалах статистического ряда

Эмпирическая плотность распределения $f_i^*(x)$ вычисляется по формуле

$$f_i^*(x) = \frac{p_i^*}{h}. \quad (2.10)$$

Аналитическое выражение для эмпирической плотности распределения на каждом интервале можно записать в виде

$$f_i^*(x) = \begin{cases} 0, & x \leq a_0 \\ f_1^*, & a_0 < x \leq a_1 \\ f_2^*, & a_1 < x \leq a_2 \\ \\ f_i^*, & a_{i-1} < x \leq a_i \\ \\ 0, & x > a_k \end{cases} \quad (2.11)$$

График эмпирической плотности распределения строится в виде кусочно-постоянной над интервалами $I_i = [a_{i-1}, a_i)$ линии и называется гистограммой (рис. 2.2).

Соединив точки гистограммы с абсциссами $\bar{a}_i$ при $i = 1, 2, 3, \dots, k$, можно построить полигон частот.

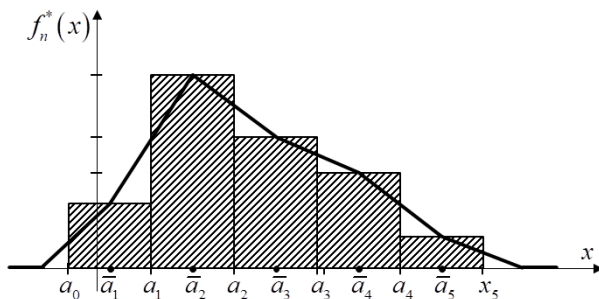


Рис. 2.2. Пример гистограммы и полигона частот интервального ряда

2.3. Точечные оценки параметров случайной величины

Точечными оценками параметров называют такие оценки, которые выражаются каким-то одним числом (точкой).

Поскольку точные значения параметров распределения признака X в генеральной совокупности в большинстве случаев

определить не представляется возможным, то эти параметры следует оценить, используя соответствующие параметры выборки.

Числовые характеристики выборки называются выборочными (эмпирическими) характеристиками.

Основные числовые характеристики:

– *среднее арифметическое (выборочное среднее)*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad (2.12a)$$

при интервальной группировке

$$\bar{x} = \frac{1}{n} \sum_{i=1}^K m_i \bar{a}_i, \quad (2.12б)$$

где m_i – частота; K – количество интервалов; $\bar{a}_i$ – середина i -го интервала;

– *выборочная (эмпирическая) дисперсия*

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right); \quad (2.13a)$$

при интервальной группировке

$$S^2 = \frac{1}{n-1} \sum_{i=1}^K m_i (\bar{a}_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^K m_i \bar{a}_i^2 - n \cdot \bar{x}^2 \right); \quad (2.13б)$$

– *выборочные начальные и центральные моменты порядка l*

$$\alpha_l = \frac{1}{n} \sum_{i=1}^n x_i^l, \quad \beta_l = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^l; \quad (2.14a)$$

при интервальной группировке

$$\alpha_l = \frac{1}{n} \sum_{i=1}^K m_i \bar{a}_i^l, \quad \beta_l = \frac{1}{n} \sum_{i=1}^K m_i (\bar{a}_i - \bar{x})^l; \quad (2.14б)$$

– *оценка коэффициента асимметрии A* (характеризует симметричность распределения относительно среднего $\bar{x}$)

$$A = \frac{\beta_3}{(\beta_2)^{\frac{3}{2}}}; \quad (2.15a)$$

при интервальной группировке

$$A = \frac{1}{n \cdot \bar{s}^3} \sum_{i=1}^K m_i (\bar{a}_i - \bar{x})^3, \quad (2.15б)$$

здесь $\bar{s} = \sqrt{S^2}$ – выборочное среднее квадратическое отклонение;

– оценка эксцесса E (меры островершинности распределения по сравнению с нормальным распределением)

$$E = \frac{\beta_4}{\beta_2^2} - 3; \quad (2.16a)$$

при интервальной группировке

$$E = \frac{1}{n \cdot S^4} \sum_{i=1}^K m_i (\bar{a}_i - \bar{x})^4 - 3. \quad (2.16b)$$

Если $E > 0$, то вершина более острая, а если $E < 0$, то более плоская, чем у нормального распределения. У нормального распределения $E = 0$.

Таким образом, в качестве точечной оценки параметров распределения признака X в генеральной совокупности могут приниматься соответствующие параметры выборки.

2.4. Проверка статистических гипотез

Статистическая гипотеза – это предположение относительно вида закона распределения или параметров распределения в генеральной совокупности некоторого признака X , являющегося случайной величиной.

По своему содержанию статистические гипотезы можно подразделить на несколько основных типов.

1. Гипотезы о виде закона распределения исследуемой случайной величины.
2. О числовых значениях параметров случайной величины.
3. Об общем виде модели, описывающей статистическую зависимость между признаками.
4. О принадлежности некоторого признака к тому или иному классу величин.

Естественно, как любая гипотеза, статистическая гипотеза нуждается в проверке. Проверяемая гипотеза называется основной или нулевой и обозначается H_0 .

Гипотеза H_0 является основной в том смысле, что было бы желательно убедиться в ее справедливости.

Если кроме нулевой выдвигаются и другие гипотезы, то они называются конкурирующими и обозначаются H_1 , H_2 и т.д.

Правило, по которому принимается решение принять или отклонить H_0 , называется статистическим критерием. Проверка

статистической гипотезы проводится путем рассмотрения некоторого параметра, определяемого по результатам изучения выборки. Этот параметр называется статистикой критерия.

Пусть по некоторым данным имеются основания выдвинуть предположения о законе распределения или о параметре закона распределения случайной величины (или генеральной совокупности, на множестве объектов которой определена эта случайная величина). Задача заключается в том, чтобы подтвердить или опровергнуть это предположение, используя выборочные (экспериментальные) данные.

В качестве примера проверки гипотезы о законе распределения генеральной рассмотрим два статистических критерия: χ^2 Пирсона и критерий Колмогорова.

Критерий χ^2 . Сумма квадратов n независимых одинаково распределенных случайных величин подчиняется закону, получившему название распределения χ^2 .

В качестве статистики критерия Пирсон предложил рассматривать величину

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - m'_i)^2}{m'_i}. \quad (2.17)$$

где m'_i – теоретическая частота попадания элементов выборки в интервал $I_i = [a_{i-1}, a_i)$, $i = \overline{1, k}$; m_i – эмпирическая частота попадания элементов выборки в тот же интервал.

Теоретическая частота попадания элементов выборки в интервал $I_i = [a_{i-1}, a_i)$ определяется для каждой области по формуле

$$m'_i = np_i, \quad (2.18)$$

где p_i – вероятность попадания признака X в соответствии с предполагаемым законом распределения в интервал $[a_{i-1}, a_i)$

$$p_i = P(a_{i-1} \leq X < a_i) = F(a_i) - F(a_{i-1}), \quad (2.19)$$

здесь $a_i, i = \overline{1, k}$ – границы интервалов статистического ряда; $F(x)$ – гипотетическая функция распределения.

В частности, если основная гипотеза H_0 состоит в том, что случайная величина X распределена по нормальному закону, то

$$p_i = \Phi\left(\frac{a_i - \bar{x}}{\bar{s}}\right) - \Phi\left(\frac{a_{i-1} - \bar{x}}{\bar{s}}\right), \quad (2.20)$$

где $\bar{x}$ – среднее выборочное; $\bar{s}$ – исправленное выборочное среднее квадратическое отклонение; $\Phi(x)$ – функция Лапласа, которая определяется формулой (прил. 1)

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

В случае если гипотеза H_0 состоит в том, что случайная величина X распределена по показательному закону

$$p_i = e^{-\lambda a_i} - e^{-\lambda a_{i-1}}, \quad (2.21)$$

где параметр $\lambda = \frac{1}{\bar{x}}$; $\bar{x}$ – среднее выборочное.

В предположение о равномерном распределении случайной величины X на отрезке $[a_0, a_k]$, вероятности p_i определяются по формуле

$$p_i = \frac{1}{a_k - a_0} (a_i - a_{i-1}), \quad i = 1, 2, \dots, k. \quad (2.22)$$

Число независимых слагаемых, входящих в сумму квадратов, называется числом степеней свободы и обозначается v

$$v = k - r - l,$$

где k – число рассматриваемых интервалов; r – число параметров в принятом распределении; l – число независимых линейных ограничивающих связей.

В рассматриваемом случае $l = 1$ из-за одной имеющейся связи: равенство суммы частот объему выборки.

При проверке гипотезы о нормальном распределении признака X с заранее неизвестными параметрами a и σ величина r равна 2. Тогда выражение для v при исследовании нормально распределенной величины имеет вид

$$v = k - 3.$$

При проверке гипотезы о равномерном распределении проводится оценка двух параметров a и b – концы интервала, в котором наблюдались возможные значения X , поэтому число степеней свободы принимается равным $v = k - 3$.

Для показательного распределения оценивается один параметр (λ), поэтому эмпирические и теоретические частоты сравнивают, приняв число степеней свободы равным $v = k - 2$.

По значениям v и α (уровень значимости) находят значение $\chi_{\text{кр}}^2$ по таблицам (прил. 3). Если $\chi^2 < \chi_{\text{кр}}^2$, то гипотеза H_0 принимается, а если $\chi^2 \geq \chi_{\text{кр}}^2$, то эта гипотеза отклоняется.

Замечание. Критерий χ^2 Пирсона следует применять, если каждая частота m_i в рассматриваемой области имеет значение не менее $\alpha \cdot n$, как правило, $m_i \geq 5$. Если $m_j < 5$ при некотором j , то j -й интервал объединяется соседним, при этом эмпирические и теоретические частоты суммируются.

Критерий Колмогорова. Согласно критерию Колмогорова вычисляется статистика

$$\xi_n = \sqrt{n} \max_x |F_n^*(x) - F(x)| \quad (2.24)$$

в граничных точках интервального ряда, т.е. в точках a_i . Здесь в качестве количественной меры рассматривается максимальное значение абсолютной величины разности между эмпирической функцией распределения $F_n^*(x)$ и соответствующей теоретической функцией распределения $F(x)$ (рис. 2.3). При вычислении статистики ξ_n значения эмпирической функции распределения $F_n^*(x)$ определяют по формуле (2.9)

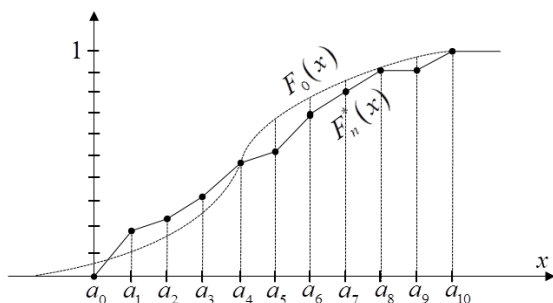


Рис. 2.3. Эмпирическая функция распределения в граничных точках интервального ряда и гипотетическая функция $F_0(x)$

Значения гипотетической функции распределения $F(x)$ в точке a_i вычисляются в зависимости от выдвинутой гипотезы

$$H_0 : F(x) = F_0(x) :$$

– если выдвинута гипотеза о нормальном распределении, то

$$F_0(a_i) = 0.5 + \Phi\left(\frac{a_i - \bar{x}}{\bar{s}}\right), \quad (2.25)$$

где $\bar{x}$ – среднее выборочное; $\bar{s}$ – исправленное выборочное среднее квадратическое отклонение; $\Phi(x)$ – функция Лапласа, значения которой берутся из таблицы (прил. 1);

– если выдвинута гипотеза о показательном распределении, то

$$F_0(a_i) = \begin{cases} 0, & a_i < a_0 \\ 1 - e^{-\lambda a_i}, & a_i \geq a_0 \end{cases}, \quad (2.26)$$

где параметр $\lambda = \frac{1}{\bar{x}}$; $\bar{x}$ – среднее выборочное;

– если выдвинута гипотеза о равномерном распределении, то

$$F_0(a_i) = \frac{a_i - a_0}{a_k - a_0}, \quad a_0 \leq a_i \leq a_k, \quad (2.27)$$

где a_i – границы интервалов интервального ряда.

Выбирая наибольшее значение $|F_n^*(a_i) - F_0(a_i)|$, вычисляется статистика $\xi_n = \sqrt{n} \max_x |F_n^*(a_i) - F_0(a_i)|$. По заданному значению уровня значимости q определяется λ_q такое, что $p(\xi_n \geq \lambda_q) = q$ из таблицы распределения Колмогорова. Если $\xi_n \geq \lambda_q$, то основная гипотеза отвергается. Если $\xi_n < \lambda_q$, то гипотеза принимается, т.е. генеральное распределение считается совпадающим с гипотетическим.

ЗАДАНИЯ ЛАБОРАТОРНОЙ РАБОТЫ № 1

Первичный анализ экспериментальных данных

По статистическим данным, полученным в результате проведения опыта, требуется:

1. Произвести интервальную группировку, построить статистическое распределение относительных частот и изобразить его графически.

2. Найти эмпирическую функцию распределения и построить ее график.

3. Вычислить выборочную среднюю, выборочную дисперсию, выборочное среднее квадратическое отклонение, моду, коэффициенты асимметрии и эксцесса.

4. Выдвинуть гипотезу о законе генерального распределения и пользуясь критериями χ^2 и Колмогорова на уровне значимости 0,01, установить случайно или значимо расхождение между формой распределения выборки и генеральной совокупности.

Порядок выполнения задания

1. Построить интервальный статистический ряд. Упорядочить выборку, т.е. составить вариационный ряд, разбить весь диапазон выборки на $k = \lceil \sqrt{n} \rceil$ частичных интервалов одинаковой длины $h = (x_{\max} - x_{\min})/k$, вычислить статистические оценки (частоты), подсчитав сколько значений признака попадает в каждый частичный интервал (значения, совпадающие с граничными, следует отнести к левому интервалу).

2. Вычислить эмпирическую функцию распределения (формула 2.9) и построить ее график.

Вычислить эмпирические плотности распределения (формула 2.10), построить гистограмму и полигон.

3. Получить точечные статистические оценки параметров распределения (выборочное среднее, выборочная дисперсия, выборочная асимметрия, выборочный эксцесс) по формулам:

$$\begin{aligned}\bar{x} &= \sum_{i=1}^k \bar{a}_i p_i^*; \\ s^2 &= \sum_{i=1}^k \bar{a}_i^2 p_i^* - \bar{x}^2, \quad \bar{s}^2 = \frac{n}{n-1} s^2, \quad \bar{s} = \sqrt{\bar{s}^2}; \\ A^* &= \frac{1}{\bar{s}^3} \sum_{i=1}^k (\bar{a}_i - \bar{x})^3 p_i^*; \\ E^* &= \frac{1}{\bar{s}^4} \sum_{i=1}^k (\bar{a}_i - \bar{x})^4 p_i^* - 3.\end{aligned}$$

Здесь эмпирические частоты p_i^* вычисляются по формуле (2.7).

4. Выдвинуть гипотезу о законе генерального распределения и построить теоретическую кривую. Для этого в точках, яв-

ляющихся серединами интервалов, вычислить значения плотностей гипотетического распределения (нормального, показательного или равномерного).

5. Проверить истинность выдвинутой гипотезы по критерию Пирсона по следующей схеме:

а) вычислить теоретические частоты $m'_i = np_i$ (формула 2.18) и найти

$$\chi^2_{\text{расч}} = \sum_{i=1}^k \frac{(m_i - m'_i)^2}{m'_i} = \sum_{i=1}^k \frac{m_i^2}{m'_i} - n;$$

б) по заданному уровню значимости α и числу степеней свободы ν из таблицы критических точек распределения (прил. 3) найти $\chi^2_{\text{кр}}$ ($\nu = k - 3$ для равномерного и нормального распределения, $\nu = k - 2$ для показательного распределения, k – число групп интервалов);

в) если $\chi^2_{\text{расч}} < \chi^2_{\text{кр}}$, то нет оснований отвергать гипотезу; Если не выполняется данное условие ($\chi^2_{\text{расч}} \geq \chi^2_{\text{кр}}$) – то гипотеза отклоняется.

6. Проверить истинность гипотезы по критерию Колмогорова. Для этого вычислить статистику ξ_n (формула 2.24) в граничных точках интервального ряда, т.е. в точках a_i и сравнить с табличным значением λ_q , найденным по значению уровня значимости q из таблицы распределения Колмогорова (прил. 5). Если $\xi_n < \lambda_q$, то основная гипотеза принимается, если $\xi_n \geq \lambda_q$, то отвергается.

Пример решения варианта задания

Исходные данные

26,7; 26,9; 26,8; 26,1; 26,5; 26,2; 26,5; 26,4; 26,4; 26,3; 26,2; 25,8; 26,5; 27,0; 25,7; 28,6; 28,2; 28,7; 25,4; 28,9; 26,4; 27,9; 27,6; 26,8; 26,3; 26,6; 26,9; 27,0; 27,0; 26,7; 27,2; 27,1; 27,2; 27,3; 27,2; 27,5; 27,5; 27,6; 27,6; 27,8; 27,7; 27,7; 27,4; 27,6; 24,8; 27,6; 24,8; 25,7; 26,7; 24,7.

1. Построение интервального статистического ряда

1. Составим вариационный ряд

24,7; 24,8; 24,8; 25,4; 25,7; 25,7; 25,8; 26,1; 26,2; 26,2; 26,3; 26,3; 26,4; 26,4; 26,4; 26,5; 26,5; 26,5; 26,6; 26,7; 26,7; 26,7; 26,8; 26,8; 26,9; 26,9; 27,0; 27,0; 27,0; 27,1; 27,2; 27,2; 27,2; 27,3; 27,4; 27,5; 27,5; 27,6; 27,6; 27,6; 27,6; 27,6; 27,7; 27,7; 27,8; 27,9; 28,2; 28,6; 28,7; 28,9.

2. Объем выборки $n = 50$, $x_{\min} = 24,7$, $x_{\max} = 28,9$.

3. Диапазон выборки $x \in [24,7; 28,9]$, то есть размах выборки $R = 28,9 - 24,7 = 4,2$.

4. Количество интервалов $k = [\sqrt{n}] = [\sqrt{50}] = [7,071] = 7$.

5. Длина интервала $d = \frac{R}{k} = 0,6$.

6. Границы интервалов статистического ряда и сами интервалы

$$I_1 = [24,7; 25,3), I_2 = [25,3; 25,9), I_3 = [25,9; 26,5), I_4 = [26,5; 27,1),$$

$$I_5 = [27,1; 27,7), I_6 = [27,7; 28,3), I_7 = [28,3; 28,9).$$

Число выборочных значений, попавших в каждый интервал:

$$m_1 = 3, m_2 = 4, m_3 = 8, m_4 = 14, m_5 = 13, m_6 = 5, m_7 = 3.$$

7. Частоты интервального ряда

$$p_1^* = \frac{m_1}{n} = \frac{3}{50} = 0,06; p_2^* = \frac{4}{50} = 0,08; p_3^* = \frac{8}{50} = 0,16; p_4^* = \frac{14}{50} = 0,28;$$

$$p_5^* = \frac{13}{50} = 0,26; p_6^* = \frac{5}{50} = 0,10; p_7^* = \frac{3}{50} = 0,06.$$

8. Середины интервалов

$$\bar{a}_1 = \frac{24,7 + 25,3}{2} = 25,0; \bar{a}_2 = 25,6; \bar{a}_3 = 26,2; \bar{a}_4 = 26,8;$$

$$\bar{a}_5 = 27,4; \bar{a}_6 = 28,0; \bar{a}_7 = 28,6.$$

2. Построение эмпирической функции распределения

1. Накопленные частоты

$$z_1 = p_1^* = 0,06; z_2 = p_1^* + p_2^* = 0,14; z_3 = p_1^* + p_2^* + p_3^* = 0,30;$$

$$z_4 = p_1^* + p_2^* + p_3^* + p_4^* = 0,58; z_5 = p_1^* + p_2^* + p_3^* + p_4^* + p_5^* = 0,84;$$

$$z_6 = p_1^* + p_2^* + p_3^* + p_4^* + p_5^* + p_6^* = 0,94;$$

$$z_7 = p_1^* + p_2^* + p_3^* + p_4^* + p_5^* + p_6^* + p_7^* = 1.$$

2. Дискретный статистический ряд, в котором указаны середины интервалов $\bar{a}_i$, частоты p_i^* и накопленные частоты z_i .

$\bar{a}_i$	25,0	25,6	26,2	26,8	27,4	28,0	28,6
p_i^*	0,06	0,08	0,16	0,28	0,26	0,10	0,06
z_i	0,06	0,14	0,30	0,58	0,84	0,94	1

3. Эмпирическая функция распределения $F_n^*(x)$ (рис. 2.4):

$$F_n^*(x) = \begin{cases} 0, & x \leq 25,0; \\ 0,06, & 25,0 < x \leq 25,6; \\ 0,14, & 25,6 < x \leq 26,2; \\ 0,30, & 26,2 < x \leq 26,8; \\ 0,58, & 26,8 < x \leq 27,4; \\ 0,84, & 27,4 < x \leq 28,0; \\ 0,94, & 28,0 < x \leq 28,6; \\ 1, & x > 28,6. \end{cases}$$

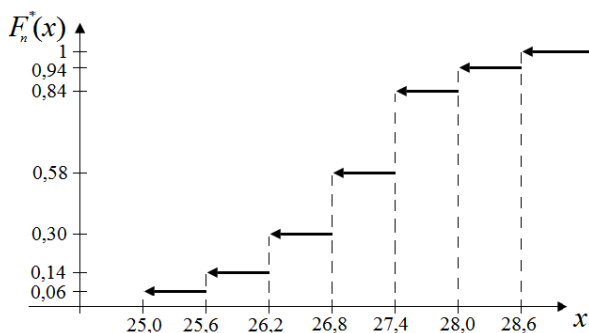


Рис. 2.4. Эмпирическая функция распределения

3. Построение эмпирической плотности распределения

1. Эмпирическая плотность (рис. 2.5) вычисляется по формуле (2.10):

$$f_1^*(x) = \frac{p_1^*}{d} = \frac{0,06}{0,6} = 0,1; \quad f_2^*(x) = 0,133; \quad f_3^*(x) = 0,267; \quad f_4^*(x) = 0,467; \\ f_5^*(x) = 0,433; \quad f_6^*(x) = 0,167; \quad f_7^*(x) = 0,1.$$

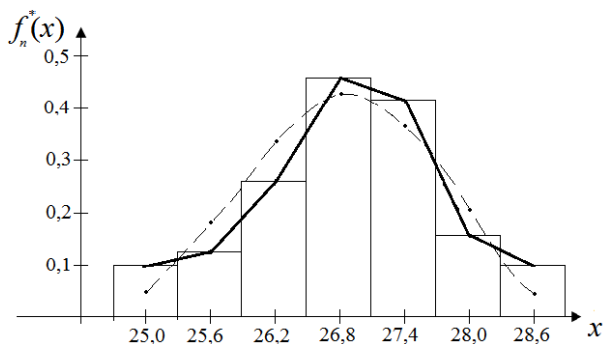


Рис. 2.5. Эмпирическая плотность распределения, полигон частот и нормальная кривая (пунктирная линия)

Аналитическое выражение для эмпирической плотности распределения на каждом интервале принимает вид

$$f_i^*(x) = \begin{cases} 0, & x \leq 24,7; \\ 0,1, & 24,7 < x \leq 25,3; \\ 0,133, & 25,3 < x \leq 25,9; \\ 0,267, & 25,9 < x \leq 26,5; \\ 0,467, & 26,5 < x \leq 27,1; \\ 0,433, & 27,1 < x \leq 27,7; \\ 0,167, & 27,7 < x \leq 28,3; \\ 0,1, & 28,3 < x \leq 28,9; \\ 0, & x > 28,9. \end{cases}$$

2. Ряд распределения эмпирической плотности примет вид

$\bar{a}_i$	25,0	25,6	26,2	26,8	27,4	28,0	28,6
$f_i^*(x)$	0,1	0,133	0,267	0,467	0,433	0,167	0,1

4. Получение точечных статистических оценок

Выборочное среднее вычислим по формуле

$$\bar{x} = \sum_{i=1}^k \bar{a}_i p_i^* = 26,88.$$

Найдем выборочную дисперсию

$$s^2 = \sum_{i=1}^k \bar{a}_i^2 p_i^* - \bar{x}^2 = 0,79.$$

Исправленная выборочная дисперсия и исправленное выборочное среднее квадратическое отклонение вычислим по формулам

$$\bar{s}^2 = \frac{n}{n-1} s^2 = 0,81$$

$$\bar{s} = \sqrt{\bar{s}^2} = 0,90.$$

Выборочная медиана равна $\bar{a}_4 = 26,8$, так как накопленная частота $\frac{25,3-2,88}{0,9}$, то есть интервал I_4 – медианный. Этому же интервалу соответствует наибольшая частота, поэтому выборочная мода равна $Mo = \bar{a}_4 = 26,8$.

Выборочную асимметрию вычислим по формуле

$$A^* = \frac{1}{\bar{s}^3} \sum_{i=1}^k (\bar{a}_i - \bar{x})^3 p_i^* = -0,20.$$

Выборочный эксцесс вычислим по формуле

$$E^* = \frac{1}{\bar{s}^4} \sum_{i=1}^k (\bar{a}_i - \bar{x})^4 p_i^* - 3 = -0,40.$$

По найденным значениям $a = \bar{x} = 26,88$ и $\sigma = \bar{s} = 0,90$ на рис. 2.5 построена теоретическая кривая нормального распределения в точках $\bar{a}_i$ по формуле

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

5. Проверка гипотезы о нормальном распределении

Проверка истинности по критерию χ^2

По формуле (2.20) вычислим теоретические вероятности

$$\begin{aligned} p_1 &= \Phi\left(\frac{25,3-26,88}{0,90}\right) - \Phi\left(\frac{24,7-26,88}{0,90}\right) = \Phi(-1,75) - \Phi(-2,42) = \\ &= -\Phi(1,75) + \Phi(2,42) = -0,4599 + 0,4922 = 0,0323; \end{aligned}$$

$$p_2 = -\Phi(1,09) + \Phi(1,75) = -0,3621 + 0,4599 = 0,0978;$$

$$p_3 = \Phi(-0,42) - \Phi(-1,09) = -0,1628 + 0,3621 = 0,1993;$$

$$p_4 = \Phi(0,24) - \Phi(-0,42) = 0,0948 + 0,1628 = 0,2576;$$

$$p_5 = \Phi(0,91) - \Phi(0,24) = 0,3186 - 0,0871 = 0,2315;$$

$$p_6 = \Phi(1,58) - \Phi(0,91) = 0,4429 - 0,3186 = 0,1243;$$

$$p_7 = \Phi(2,24) - \Phi(1,58) = 0,4875 - 0,4429 = 0,0446.$$

Теоретические частоты соответственно равны

$$m'_1 = np_1 = 1,62; \quad m'_2 = 4,89; \quad m'_3 = 9,97; \quad m'_4 = 12,88;$$

$$m'_5 = 11,58; \quad m'_6 = 6,22; \quad m'_7 = 2,23.$$

Объединим малочисленные частоты и соответствующие им теоретические частоты

m_i	7	8	14	13	8
m'_i	6,51	9,97	12,88	11,58	8,45

По полученным данным вычислим наблюдаемое значение критерия Пирсона

$$\chi^2_{\text{набл}} = \sum_{i=1}^k \frac{(m_i - m'_i)^2}{m'_i} = 0,76.$$

По таблице критических точек распределения $\chi^2_{\text{кр}}$, по уровню значимости $\alpha = 0,01$ и числу степеней свободы $\nu = 5 - 3 = 2$ найдем критическую точку правосторонней области $\chi^2_{\text{кр}} = 9,2$.

Так как $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$ – нет оснований отвергнуть гипотезу о нормальном распределении.

Замечание. Если гипотеза о нормальном законе генеральной совокупности отвергается, то используя критерий Пирсона, нужно проверить гипотезу о показательном или равномерном распределении. В этих случаях необходимо пересчитать теоретические вероятности по формулам (2.21) и (2.22).

Для того чтобы проверить гипотезу о показательном законе распределения, следует:

1) принять в качестве оценки параметра λ величину, обратную выборочной средней

$$\lambda = \frac{1}{\bar{x}};$$

2) найти теоретические вероятности попадания случайной величины X в интервал $[a_{i-1}, a_i)$ по формуле (2.21)

$$p_i = e^{-\lambda a_i} - e^{-\lambda a_{i-1}};$$

3) вычислить теоретические частоты

$$m'_i = np_i.$$

Сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, приняв число степеней свободы $\nu = k - 2$.

Для того чтобы проверить гипотезу о равномерном законе распределения, т.е. по закону

$$f(x) = \begin{cases} 0, & x < a; \\ \frac{1}{b-a}, & a \leq x \leq b; \\ 0, & x > b, \end{cases}$$

следует:

1) оценить параметры a и b по формулам

$$a^* = \bar{x} - \sqrt{3} \cdot \bar{s}; \quad b^* = \bar{x} + \sqrt{3} \cdot \bar{s};$$

2) найти плотность вероятности предполагаемого распределения

$$f(x) = \frac{1}{b^* - a^*};$$

3) найти теоретические частоты по формуле (2.22)

$$m'_i = np_i = n \frac{1}{b^* - a^*} (a_i - a_{i-1}), i = 1, 2, \dots, k;$$

4) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, приняв число степеней свободы $\nu = k - 3$.

Проверка истинности по критерию Колмогорова

Согласно критерию Колмогорова вычислим статистику

$$\xi_n = \sqrt{n} \max_x |F_n^*(x) - F_0(x)|$$

в граничных точках

$$a_0 = 24,7; a_1 = 25,3; a_2 = 25,9; a_3 = 26,5;$$

$$a_4 = 27,1; a_5 = 27,7; a_6 = 28,3; a_7 = 28,9.$$

Эмпирическая функция распределения была найдена

$$F_n^*(x) = \begin{cases} 0, & x \leq 25,0; \\ 0,06, & 25,0 < x \leq 25,6; \\ 0,14, & 25,6 < x \leq 26,2; \\ 0,30, & 26,2 < x \leq 26,8; \\ 0,59, & 26,8 < x \leq 27,4; \\ 0,84, & 27,4 < x \leq 28,0; \\ 0,94, & 28,0 < x \leq 28,6; \\ 1, & x > 28,6. \end{cases}$$

$$F_n^*(24,7) = 0; F_n^*(25,3) = 0,06; F_n^*(25,9) = 0,14; F_n^*(26,5) = 0,30;$$

$$F_n^*(27,1) = 0,59; F_n^*(27,7) = 0,84; F_n^*(28,3) = 0,94; F_n^*(28,9) = 1.$$

Значения теоретической функции $F_0(x)$ найдем по формуле (2.25)

$$F_0(a_i) = 0,5 + \Phi\left(\frac{a_i - \bar{x}}{\bar{s}}\right)$$

с учетом вычисленных значений $\bar{x} = 26,88$ и $\bar{s} = 0,90$.

$$F_0(24,7) = 0,5 + \Phi\left(\frac{24,7 - 26,88}{0,9}\right) =$$

$$= 0,5 + \Phi(-2,42) = 0,5 - 0,4922 = 0,008;$$

$$F_0(25,3) = 0,040; F_0(25,9) = 0,139;$$

$$F_0(26,5) = 0,337; F_0(27,1) = 0,595;$$

$$F_0(27,7) = 0,819; F_0(28,3) = 0,943; F_0(28,9) = 0,987.$$

Модуль разностей эмпирической и теоретической функций принимает максимальное значение в точке $a_3 = 26,5$

$$|F_n^*(26,5) - F_0(26,5)| = 0,037.$$

Тогда значение статистики Колмогорова:

$$\xi_n = \sqrt{n} \max_x |F_n^*(x) - F_0(x)| = \sqrt{50} \cdot 0,037 = 0,262.$$

Из таблицы распределения Колмогорова (прил. 5) при заданном уровне значимости $q = 0,01$ находим $\lambda_q = 1,627$.

Так как $\xi_n = 0,262 < \lambda_q = 1,627$, основная гипотеза H_0 принимается, то есть генеральное распределение считается нормальным.

Таким образом, данные наблюдений согласуются гипотезой о нормальном распределении по обоим критериям.

Замечание. Если гипотеза о нормальном законе генеральной совокупности отвергается, то используя критерий Колмогорова, нужно проверить гипотезу о показательном или равномерном распределении. В этих случаях необходимо переписать теоретические функции распределения по формулам (2.26) и (2.27).

Вопросы к защите лабораторной работы

1. Какие задачи решаются методами математической статистики?

2. Что называется генеральной совокупностью, генеральной средней и генеральной дисперсией?

3. Что называется выборкой, какие требования предъявляются к ней?

4. Какие существуют формулы для определения числа интервалов в статистическом ряду?

5. Что называется вариационным рядом?

6. Что называется статистическим рядом распределения?

7. Как определяются частоты в интервальном ряду? Чему должна быть равна их сумма и почему?

8. Как составляется дискретный ряд распределения, и на основе какого допущения интервальный ряд заменяется приближенным дискретным?

9. Что называются накопленными частотами и каков их смысл?

10. Что называется эмпирической функцией распределения? Как выглядит ее график?

11. Что называется эмпирической плотностью распределения?

12. Чему должна быть равна площадь под гистограммой и почему?
13. Каким соотношением связаны эмпирическая функция распределения и эмпирическая плотность распределения?
14. Что называется точечной статистической оценкой?
15. Что является статистической оценкой математического ожидания?
16. Что является статистической оценкой дисперсии?
17. Что является статистической оценкой среднего квадратического отклонения?
18. Как вычисляются выборочные эксцесс и асимметрия?
19. Какие выводы можно сделать о распределении генеральной совокупности по величине эксцесса и асимметрии?
20. Что называется статистической гипотезой?
21. Что такое основная (нулевая) статистическая гипотеза и альтернативная?
22. Что называется статистическим критерием?
23. Как определяется критерий согласия?
24. Что называется критической областью?
25. Каким образом с помощью критерия согласия принимается или отвергается основная статистическая гипотеза?
26. Что принимается в математической статистике за уровень значимости и мощность критерия проверки гипотез?
27. Какие ошибки называются ошибками 1-го рода? 2-го рода?
28. Какая статистика выбирается для проверки истинности выдвинутой гипотезы в критерии Пирсона?
29. Чем определяется критическая область в критерии Пирсона?
30. Какая статистика выбирается для проверки истинности выдвинутой гипотезы в критерии Колмогорова?
31. Чем определяется критическая область?
32. Что называется доверительной вероятностью?
33. Как определяется доверительный интервал для параметра статистического распределения?
34. Что называется надежностью и точностью интервальной оценки параметра распределения?

Варианты заданий

Вариант № 1

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

Вариант № 2

9,5	9,4	10,2	9,7	10,4	10,4	10,3	9,3	9,8	9,7
9,7	9,7	10,0	9,8	10,1	9,8	10,1	12,3	13,1	14,2
13,7	15,3	14,9	16,2	17,4	17,8	17,5	16,4	15,8	15,6
14,1	13,7	13,5	12,5	12,2	11,9	11,6	10,9	10,4	9,9
9,8	10,7	11,3	11,9	13,5	14,7	15,3	15,6	15,1	14,8

Вариант № 3

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

Вариант № 4

10,3	10,8	11,8	17,2	18,6	18,4	14,2	14,1	19,6	23,4
10,4	10,8	12,3	16,7	18,9	17,7	13,9	14,5	20,8	23,2
10,7	10,8	12,6	16,5	18,7	17,2	13,6	14,9	21,3	22,9
10,8	10,8	12,8	17,4	18,3	16,8	13,3	15,7	21,7	22,7
10,9	10,8	13,3	16,8	18,1	16,6	13,4	16,4	21,8	22,4

Вариант № 5

9,6	9,6	9,8	9,5	9,9	9,4	9,5	10,3	11,8	11,9
13,3	13,7	15,4	15,7	16,4	18,1	17,5	17,6	17,1	13,3
16,8	16,4	15,7	16,0	17,6	18,4	18,3	18,3	17,4	16,6
17,5	17,6	17,1	16,6	17,2	16,4	16,1	13,7	13,5	13,7
13,5	11,7	10,8	9,7	9,3	9,4	9,5	9,3	9,6	9,7

Вариант № 6

-3,9	-3,8	-3,9	-3,9	-4,1	-3,9	-4,0	-4,2	-4,4	-4,9
-5,1	-5,0	-5,3	-5,6	-5,6	-5,6	-6,2	-6,7	-7,0	-7,2
-7,8	-7,9	-8,3	-8,4	-9,1	-9,6	-9,6	-9,8	-9,8	-9,9
-9,2	-9,8	-9,6	-9,3	-9,2	-9,0	-8,6	-9,1	-9,2	-8,7
-9,2	-9,6	-9,3	-9,6	-9,4	-9,4	-9,4	-8,6	-8,3	-7,6

Вариант № 7

27,7	26,1	27,2	23,7	24,2	24,2	24,8	25,0	25,3	26,1
26,4	26,6	26,7	27,1	27,1	27,2	26,9	24,7	26,2	27,4
28,8	24,9	25,3	25,4	25,4	25,7	26,3	28,1	28,8	29,1
29,5	29,5	29,3	28,4	29,1	28,6	27,9	27,7	27,3	26,7
26,9	28,8	28,8	28,3	28,1	28,1	27,5	26,8	26,4	26,1

Вариант № 8

18,2	14,9	14,1	13,4	18,1	12,2	11,6	10,8	10,6	12,9
17,8	14,7	14,1	13,5	18,1	12	11,5	10,8	10,3	13,4
17,3	14,8	13,9	13,6	18,3	11,9	11,3	10,7	10,2	13,7
16,8	14,7	13,8	13,8	19,1	11,7	11,2	10,7	10,1	13,5
16,3	14,4	13,8	13,9	18,5	11,3	11,2	10,8	10,2	13,2

Вариант № 9

10,6	10,9	15,2	17,3	18,2	16,1	13,2	17,1	22,9	21,4
10,3	10,9	15,2	17,5	18,3	15,8	14,2	17,7	22,8	19,9
10,6	11,1	16,2	17,3	18,4	15,6	14,3	18,2	23,4	19,6
10,7	11,4	16,9	18	18,7	14,8	14,2	18,6	23,1	18,8
10,8	11,8	17,2	18,6	18,4	14,2	14,1	19,6	23,4	18,3

Вариант № 10

14,7	13,7	18,8	23,2	19,7	17	15,7	15,7	18,9	15,9
14,3	13,4	19,8	23,6	19,2	16,4	15,7	16,5	18,2	16
14,2	13,1	20,5	22,6	18,3	16,3	15,3	17,1	17,8	14,4
13,9	13,2	21,6	22,4	18,2	16,1	15,1	17,4	17,4	13,8
14	13,3	22,3	22,5	17,8	16,2	14,8	18	17,1	13,6

Вариант № 11

1,2	1,4	1,9	2,2	1,3	1,2	0,8	-1,3	-2,3	-1,7
1,3	1,4	2,2	2,1	1,3	1,2	0,8	-1,2	-2,2	-1,9
1,4	1,5	2,1	2,1	1,3	1,2	0,6	-1,7	-1,8	-1,8
1,5	1,5	2,1	2,1	1,2	1,1	0,6	-1,8	-1,9	-1,7
1,7	1,6	2,1	2	1,2	1,1	0,4	-2	-1,9	-1,9

Вариант № 12

-2,4	-8,9	-3,5	-3,9	-3,3	-3,7	-2,2	-3,1	-3,9	-4,8
-5,6	-6,1	-3,4	-3,9	-6,3	-2,6	-4,9	-3,2	-3,9	-4,8
-6,5	-5,2	-3,4	-3,9	-6	-2,6	-1,8	-3,1	-3,9	-4,8
-8	-3,9	-4,3	-3,8	-5,4	-2,6	0,1	-3,1	-3,8	-6,1
-7,9	-4,4	-4,2	-3,8	-4	-2,6	-0,4	-3,1	-3,8	-6,1

Вариант № 13

12,6	13,1	13,2	12,9	12,9	12,9	13,1	12,6	12,4	14,2
12,6	12,9	13,4	12,8	12,8	12,6	12,9	12,6	12,2	14,2
12,8	12,9	13,1	13,4	13,1	12,8	12,9	12,6	12,7	14,2
12,8	13,1	13,1	13,3	13,2	12,7	12,8	12,6	13,2	13,9
12,8	13,1	13,1	12,5	12,9	12,9	12,8	12,8	13,1	13,8

Вариант № 14

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

Вариант № 15

-8,4	-2,9	-5,2	-9,3	-6,8	-6,1	-9,9	-7,7	-6,3	-4,3
-7,5	-2,4	-6,1	-9,6	-3,8	-3,3	-8,7	-4,8	-3,3	-5,3
-6,5	-2,1	-6,3	-9,9	-4,9	-5,2	-9,6	-3,9	-1,8	-3,2
-5,9	-1,9	-7	-6,2	-7,5	-4,8	-5,8	-6,6	-1,9	-5,7
-5	-1,7	-7,7	-8,9	-8,7	-8,7	-6,7	-8,1	-2,2	-8,8

Вариант № 16

16	13,1	10,3	9,8	7,1	6,7	7,6	7,4	6,6	5,6
15,8	12,7	10,3	9,8	7,0	6,7	7,7	7,1	6,5	5,3
15,8	11,9	10,2	9,6	6,5	6,8	7,6	6,9	6,6	5,2
15,7	11,8	10,2	9,0	6,6	7,2	7,6	7,1	6,3	5,3
15,7	11,6	10,0	8,9	5,8	7,2	7,6	6,7	6,6	5,3

Вариант № 17

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

Вариант № 18

9,9	10,6	11,0	11,3	11,1	9,7	11,1	11,6	11,4	10,6
9,9	10,6	11,1	11,3	11,5	9,9	11,1	11,6	11,6	9,9
10	10,8	11,3	11,2	10,2	10,3	11,3	11,4	11,4	10
10,3	10,9	11,7	11,5	11,1	10,3	11,3	11,4	11,1	10,2
10	10,8	11,1	11,3	11,1	10,5	11,4	11,4	10,4	9,9

Вариант № 19

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

Вариант № 20

18,3	14,6	13,9	13,8	23	22,7	16,8	13,6	12,2	16,7
16,8	14,8	13,8	14,2	20,4	23,1	16,3	13,6	12,3	16,9
16,6	14,9	13,4	15,2	23,2	22,2	16	13,4	12,9	20,6
16	14,6	13,3	14,3	23,5	21,2	15,3	12,9	13,9	21,2
16,2	14,4	13,2	14,3	20,4	19,7	14,8	12,9	15,5	21,2

Вариант № 21

-2,3	0,4	0,7	2,3	4,1	3,2	1,8	-0,2	1,2	3,2
-2,2	0,1	0,8	3,7	5,1	3	-0,1	0,5	1,1	3,6
-1,2	0,2	0,9	4,6	3,7	2,8	-0,4	0,9	1,4	3,7
-1,7	-0,2	0,8	5,4	2,7	2,6	-0,8	1,2	1,4	4,5
-1,2	0	1,1	5,7	2,6	2,2	-0,3	1,3	1,8	3,8

Вариант № 22

-1,2	0,3	1,3	2,5	5,8	4,2	3,9	1,2	2,7	2,8
-1,1	0,9	1,8	2,4	6,1	4,8	3,6	1,7	2,6	3
-1	1,6	1,2	2,7	6,1	4,4	3,4	1,1	2,7	3,4
-1,2	1,4	1,6	3,2	5,8	4,4	3,7	0,8	2,6	3,9
-1	1,6	2,1	3,3	5,1	4,3	3,3	0,5	2,6	3,7

Вариант № 23

-6,6	-9,3	-9,1	-8,1	-8,3	-8,2	-9,8	-9,3	-7,6	-6,6
-7,2	-8,7	-8,7	-8,1	-9,2	-9,4	-9,3	-9,4	-8,2	-6,4
-6,6	-9,2	-9	-8,1	-8,8	-5,9	-9,2	-9,9	-8,4	-6,2
-7,3	-9,2	-8,3	-8,1	-8,8	-8,1	-9,8	-7,6	-8,1	-6,7
-7,8	-8,8	-8,7	-8,5	-8,4	-6,1	-9,5	-8,1	-7,7	-6,3

Вариант № 24

10,8	4,8	8,8	9,5	18,1	20,3	18,7	17,7	11,6	17,1
10,1	4,6	7,9	9,6	18,2	20,0	17,8	17,2	13,3	18,3
9,4	4,3	8,9	12,1	19,6	19,9	18,5	16,3	14,4	19,2
8,2	4,3	9,3	13,4	19,4	19,5	18,1	15,6	14,6	19,3
7,9	3,2	9,7	13,9	19,7	19,7	17,8	13,6	14,6	17,8

Вариант № 25

8,9	4,8	5,6	1,6	8,9	16,1	16,1	10,5	1,6	4
5,6	4,8	4,8	1,6	9,7	17,7	16,1	11,3	1,6	3,2
6,4	4,8	3,2	1,6	9,7	16,9	12,9	9,7	2,4	4,8
6,4	4,8	4	3,2	13,7	16,9	15,3	7,2	3,2	5,6
5,6	4,8	4	4	12,1	16,9	13,7	5,6	3,2	5,6
5,6	3,2	4	4	12,9	16,9	12,9	3,2	3,2	4,8

Вариант № 26

12,8	12,8	13,1	13,1	12,5	12,9	12,9	12,8	12,8	13,1
12,5	12,8	13,1	12,5	12,8	12,7	12,9	12,7	13,6	12,4
12,6	12,7	13,1	11,9	12,8	12,9	12,8	12,7	16,7	13,0
12,5	13,1	13,0	12,1	12,9	12,7	12,8	12,7	16,0	13,4
12,7	13,0	13,2	12,1	12,9	12,8	13,2	12,6	12,8	14,1

Вариант № 27

-5,6	-7,9	-4,4	-4,2	-3,8	-4	-2,6	-0,4	-3,1	-3,8
-5,7	-8,1	-4,2	-4,2	-3,8	-3,9	-9,2	-1,6	-3,2	-4,9
0,2	-7,9	-3,9	-4,2	-3,7	-3,9	-7,8	-1,7	-3,1	-4,9
-1,3	-7,9	-3,8	-4,1	-3,7	-3,8	-7,8	-2,8	-3,9	-4,9
-2,3	-7,1	-3,6	-4,3	-3,7	-3,8	-3,8	-3,0	-3,9	-4,9

Вариант № 28

0,8	1,7	1,6	2,1	2	1,2	1,1	0,4	-2	-1,9
1,1	1,7	1,7	2,2	1,9	1,2	0,9	-0,1	-2	-1,9
1,2	1,6	1,7	2,1	1,8	1,2	0,9	-0,2	-2,2	-1,8
1,1	1,3	1,8	2,1	1,6	1,2	0,8	-0,6	-2,4	-1,8
1,1	1,3	1,8	2,1	1,4	1,2	0,8	-0,8	-2,2	-1,7

Вариант № 29

14,5	14	22,3	22,1	17,5	15,9	14,7	18,7	16,7	13,3
14,4	15,1	22,3	21,6	17,4	15,9	14,7	18,3	16,6	13
14,4	15,8	22,3	21,2	16,9	15,8	14,9	18,9	16,3	12,8
14	17,3	22,7	20,2	17,2	15,8	15,3	18,7	16,1	12,6
13,7	18,8	23,2	19,7	17	15,7	15,7	18,9	15,9	12,4

Вариант № 30

26,7	26,9	26,8	26,8	26,5	26,6	26,5	26,4	26,4	26,9
25,6	25,3	25,5	25,7	25,7	28,6	28,2	29,2	28,4	29,5
28,4	27,9	27,6	27,1	26,8	26,6	26,9	27,0	27,0	27,1
27,2	27,1	27,2	27,3	27,2	27,5	27,5	27,6	27,6	27,8
27,7	27,7	27,4	27,6	27,8	28,6	28,1	26,1	24,3	24,8

3. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ. ОСНОВНЫЕ ПОЛОЖЕНИЯ. ДВУМЕРНАЯ МОДЕЛЬ

На практике часто генеральная совокупность исследуется не по одному, а по нескольким признакам или переменным, являющимся случайными величинами. В случае двух переменных X и Y требуется установить наличие и степень связи между ними.

Статистические связи между признаками (переменными) можно изучать методами дисперсионного, корреляционного и регрессионного анализа. Методами дисперсионного анализа устанавливается наличие влияния заданного фактора на изучаемый процесс. Корреляционный анализ позволяет оценить силу такой связи, а методами регрессионного анализа можно выбрать конкретную математическую модель и оценить ее адекватность.

Корреляционный анализ применяют, когда данные наблюдений или эксперимента считаются случайными и выбранными из нормальной совокупности.

Поскольку исследовать всю генеральную совокупность по этим признакам в большинстве случаев не представляется возможным, то проводится случайная выборка.

В результате изучения выборки исследователь получает пары чисел (точки):

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n),$$

где n – объём выборки.

Если нанести эти точки на координатную плоскость, то можно получить диаграмму рассеяния или корреляционное поле (рис. 2.6).

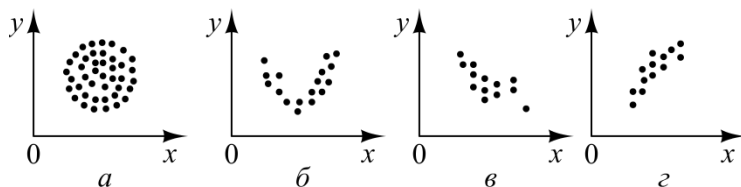


Рис. 2.6. Некоторые диаграммы рассеяния точек выборки

Парная регрессия – это уравнение связи двух переменных (признаков) $X = \{x_1, x_2, x_3, \dots, x_n\}$ и $Y = \{y_1, y_2, y_3, \dots, y_n\}$

$$y = \varphi(x).$$

В случае линейной регрессии:

$y_x = ax + b$ – регрессия Y на X ;

$x_y = cy + d$ – регрессия X на Y .

Построение уравнений регрессии сводится к оценке их параметров. Для оценки параметров регрессий используют метод наименьших квадратов (МНК). Метод позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических y_x минимальна.

Применяя к исходным данным метод наименьших квадратов, для нахождения параметров a и b получается систему уравнений

$$\begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i + b \cdot n = \sum_{i=1}^n y_i \end{cases} \quad (2.28)$$

здесь

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad \overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i.$$

Параметры c и d находятся из системы уравнений

$$\begin{cases} c \sum_{i=1}^n y_i^2 + d \sum_{i=1}^n y_i = \sum_{i=1}^n x_i y_i \\ c \sum_{i=1}^n y_i + d \cdot n = \sum_{i=1}^n x_i. \end{cases} \quad (2.29)$$

Линейное уравнение регрессии обычно записывается в принятой математической статистике форме

$$y_x - \bar{y} = a \cdot (x - \bar{x}) \quad (2.30)$$

или

$$y_x - \bar{y} = r \frac{S_y}{S_x} (x - \bar{x}).$$

Аналогично, уравнение регрессии X на Y имеет вид

$$x_y - \bar{x} = r \frac{S_x}{S_y} (y - \bar{y}). \quad (2.31)$$

Алгоритм нахождения выборочного коэффициента корреляции r следующий:

а) вычисляются выборочные средние квадратические отклонения переменных X и Y

$$S_x = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2}, \quad S_y = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - (\bar{y})^2}; \quad (2.32)$$

б) вычисляется выборочный корреляционный момент (выборочная ковариация)

$$K_{XY} = \overline{xy} - \bar{x} \cdot \bar{y}; \quad (2.33)$$

в) вычисляется выборочный коэффициент корреляции

$$r = \frac{K_{XY}}{S_x \cdot S_y}. \quad (2.34)$$

1. Коэффициент корреляции принимает значения на отрезке $[-1; 1]$, т.е.

$$-1 \leq r \leq 1.$$

В зависимости от того, насколько $|r|$ приближается к единице, различают слабую, умеренную и сильную связь: чем ближе $|r|$ к единице, тем теснее связь между признаками X и Y .

2. Если переменные X и Y умножить на одно и то же число, то коэффициент корреляции не изменится.

3. Если $r = \pm 1$, то корреляционная связь между X и Y представляет собой линейную зависимость.

Выборочный коэффициент корреляции r является оценкой генерального коэффициента корреляции ρ , тем более точной, чем больше объем выборки n .

Положительное значение коэффициента корреляции говорит о положительной связи между x и y , когда с ростом одной из переменных другая тоже растет. Отрицательное значение коэффициента корреляции означает, что с ростом одной из переменных другая убывает, с убыванием одной из переменных другая растет.

Так как коэффициент корреляции r вычисляется по данным выборки, то в разных выборках значение этой величины – есть величина случайная. Если $r \neq 0$, возникает вопрос: объясняется ли это действительно линейной связью между X и Y или вызвано случайными факторами?

Для выяснения значимости коэффициента корреляции рассматриваются две гипотезы: нулевая $H_0: \rho = 0$ – генеральный линейный коэффициент корреляции равен нулю и альтернативная гипотеза $H_1: \rho \neq 0$ – существует линейная корреляционная связь между признаками X и Y .

В качестве критерия проверки гипотезы $H_0: \rho = 0$ принимают случайную величину, распределенную по закону Стьюдента с количеством степеней свободы $k = n - 2$:

$$T_{\text{набл}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

Гипотеза $H_0: \rho = 0$ отвергается, если

$$|T_{\text{набл}}| > t_{\text{кр}}(k),$$

где $t_{\text{кр}}(k)$ – табличное значение распределения Стьюдента, определенное на уровне значимости α при числе свободы $k = n - 2$.

Точность построенной модели линейной регрессии определяется средней ошибкой аппроксимации, которая представляет собой среднее относительное отклонение расчетных значений от исходных данных

$$\Delta = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_x}{y_i} \right|.$$

ЗАДАНИЯ ЛАБОРАТОРНОЙ РАБОТЫ № 2

Корреляционный и регрессионный анализ

По данным, полученным в результате выборочного наблюдения (вторая строка – X ; третья строка таблицы – Y), требуется:

1. С помощью метода наименьших квадратов получить уравнения линейной регрессии Y на X и X на Y .

2. Оценить тесноту связи. Получить уравнения выборочных регрессий Y на X и X на Y в терминах корреляционного анализа.
3. Построить корреляционное поле и графики регрессий.
4. Проверить значимость уравнения регрессии Y на X по критерию Фишера – Снедекера при уровне значимости $\alpha = 0,05$.
5. Найти интервальную оценку коэффициентов a и b с доверительной вероятностью $\gamma = 0,95$.
6. Проверить значимость выборочного коэффициента корреляции.

Порядок выполнения работы

1. Для получения уравнения линейной регрессии $y = ax + b$ необходимо решить систему уравнений (2.28) относительно параметров a и b .

При отыскании уравнения регрессии X на Y параметры c и d вычисляются из системы уравнений (2.29).

2. Теснота связи оценивается с помощью выборочного коэффициента корреляции

$$r = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \cdot \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}}.$$

Получить уравнения регрессии Y на X и X на Y с учетом коэффициента корреляции.

3. Построить корреляционное поле и уравнения линейной регрессии Y на X и X на Y .

4. Для проверки значимости уравнения регрессии вычислить значение критерия Фишера $F_{\text{расч}} = s_R^2 / s^2$, где

$$s_R^2 = \frac{\sum_{i=1}^n (y_{x,i} - \bar{y})^2}{l-1} = \frac{\sum_{i=1}^n [(ax_i + b) - \bar{y}]^2}{l-1},$$

$$s^2 = \frac{\sum_{i=1}^n (y_i - y_{x,i})^2}{n-2} = \frac{\sum_{i=1}^n [y_i - (ax_i + b)]^2}{n-2}$$

и сравнить с табличным значением $F_{кр}$ для данного уровня значимости $\alpha = 0,05$ при $k_1 = l - 1$ и $k_2 = n - l$ (прил. 3).

Если $F_{расч} < F_{кр}$, то уравнение регрессии не является значимым с уровнем значимости α . А если $F_{расч} \geq F_{кр}$, то подтверждается значимость уравнения регрессии.

5. Доверительные интервалы коэффициентов a и b с доверительной вероятностью γ найти из неравенств

$$a - t_{\gamma} s_a < a_{ист} < a + t_{\gamma} s_a$$

$$b - t_{\gamma} s_b < b_{ист} < b + t_{\gamma} s_b,$$

где t_{γ} – параметр распределения Стьюдента с $k = n - 2$ степенями свободы (прил. 2); s_a и s_b – средние квадратические отклонения коэффициентов a и b , вычисляемые соответственно по формулам

$$s_a = \sqrt{\frac{\sum_{i=1}^n [y_i - (ax_i + b)]^2}{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}}, \quad s_b = \sqrt{\frac{\sum_{i=1}^n [y_i - (ax_i + b)]^2}{(n-2)n}}.$$

Замечание. Если в границы доверительного интервала попадает ноль, то есть нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается равным нулю, так как он не может одновременно принимать и положительное, и отрицательное значения.

6. Проверить гипотезу о значимости коэффициента корреляции при уровне значимости α .

В качестве критерия проверки гипотезы $H_0: \rho = 0$ принять случайную величину, распределенную по закону Стьюдента:

$$T_{набл} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

В таблице критических распределений Стьюдента (прил. 2) найти критическое значение $t_{кр}(k)$. Пользуясь таблицей, следует вместо n подставлять вычисленное значение $k = n - 2$ и вероятность $\gamma = 1 - \alpha$.

При $|T_{набл}| > t_{кр}(k)$ нулевая гипотеза $H_0: \rho = 0$ отвергается в пользу гипотезы $H_1: \rho \neq 0$.

Пример решения варианта задания

Исходные данные:

	1	2	3	4	5	6	7	8	9	10
x_i	3,1	3,2	2,7	2,2	2,4	2,9	3,2	2,6	2,8	3,3
y_i	6,3	5,4	5,8	5,7	4,5	5,9	4,9	5,4	4,9	5,8

11	12	13	14	15	16	17	18	19	20
3,4	3,5	3,7	3,9	4,4	4,2	4,5	4,7	4,5	4,8
5,6	6,3	7,2	7,4	7,6	8,5	7,2	8,3	8,9	7,7

21	22	23	24	25	26	27	28	29	30
4,6	4,8	4,9	4,7	4,8	4,3	4,5	3,5	3,2	2,8
7,6	9,4	8,0	8,5	7,9	6,7	7,2	7,0	5,5	5,8

1. Получение уравнений линейной регрессии методом наименьших квадратов.

Уравнение линейной регрессии Y на X имеет вид $y = ax + b$.

Для нахождения параметров a, b решим систему уравнений

$$\begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i; \\ a \sum_{i=1}^n x_i + b \cdot n = \sum_{i=1}^n y_i. \end{cases}$$

Разделив каждое уравнение системы на n , по заданной выборке вычислим суммы:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 3,737; \quad \overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2 = 14,6543; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 6,763;$$

$$\overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i = 26,212; \quad \overline{y^2} = \frac{1}{n} \sum_{i=1}^n y_i^2 = 47,405.$$

Система уравнений принимает вид

$$\begin{cases} 14,654a + 3,737b = 26,212; \\ 3,737a + b = 6,763. \end{cases}$$

Решив систему, получим: $a = 1,36$, $b = 1,68$.

Уравнение регрессии имеет вид

$$y = 1,36x + 1,68.$$

Параметры c и d уравнения линейной регрессии X на Y вычислим из системы уравнений (формула 2.29)

$$\begin{cases} c \sum_{i=1}^n y_i^2 + d \sum_{i=1}^n y_i = \sum_{i=1}^n x_i y_i; \\ c \sum_{i=1}^n y_i + d \cdot n = \sum_{i=1}^n x_i. \end{cases}$$

Решив систему, получим: $c = 0,56$, $b = -0,08$.

Уравнение регрессии X на Y имеет вид

$$x = 0,56y - 0,08.$$

2. Оценка выборочного коэффициента корреляции.

Оценим тесноту связи с помощью выборочного коэффициента корреляции, для этого вычислим выборочные средние квадратические отклонения переменных X и Y

$$S_x = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2} = 0,83, \quad S_y = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - (\bar{y})^2} = 1,29,$$

вычислим выборочный корреляционный момент (выборочная ковариация)

$$K_{XY} = \overline{xy} - \bar{x} \cdot \bar{y} = 26,212 - 3,737 \cdot 6,763 = 0,939.$$

Тогда выборочный коэффициент корреляции

$$r = \frac{K_{XY}}{S_X \cdot S_Y} = \frac{0,939}{0,83 \cdot 1,29} = 0,88.$$

Уравнения выборочных регрессий Y на X и X на Y в терминах корреляционного анализа примут вид

$$y_x = \bar{y} + r \frac{S_y}{S_x} (x - \bar{x}) = ax + b = 1,36x + 1,68,$$

$$x_y = \bar{x} + r \frac{S_x}{S_y} (y - \bar{y}) = 0,56y - 0,084.$$

Эти прямые пересекаются в точке $(m_x; m_y) = (3,73; 6,76)$.

3. Корреляционное поле и графики регрессий.

Построим корреляционное поле и уравнения линейной регрессии Y на X и X на Y (рис. 2.7).

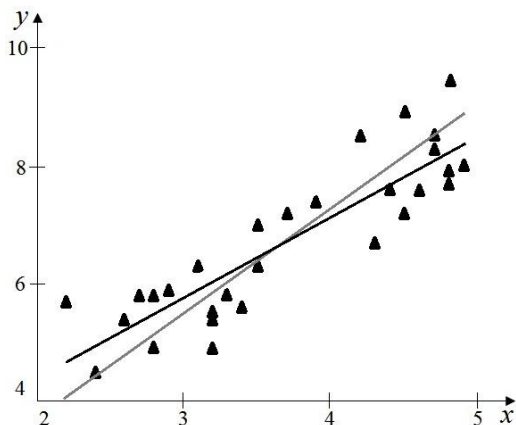


Рис. 2.7. Корреляционное поле и линии регрессии

4. Проверка значимости уравнения регрессии $y = ax + b$.

Для проверки значимости уравнения регрессии вычисляем значение критерия Фишера:

$$s_R^2 = \frac{\sum_{i=1}^n [(ax_i + b) - \bar{y}]^2}{2 - 1} = 38,272,$$

$$s^2 = \frac{\sum_{i=1}^n [y_i - (ax_i + b)]^2}{n - 2} = 0,414,$$

$$F_{\text{расч}} = s_R^2 / s^2 = 92,399.$$

Согласно критерию Фишера – Снедекера, уравнение регрессии $y = ax + b$ является значимым при уровне значимости α , если наблюдаемое значение статистики $F_{\text{расч}} > F_{\text{кр}}$.

Табличное значение $F_{\text{кр}} = 4,2$ (прил. 3) для уровня значимости $\alpha = 0,05$ при $k_1 = l - 1 = 2 - 1 = 1$ и $k_2 = n - l = 30 - 2 = 28$ (l – число оцениваемых параметров уравнения регрессии).

Так как $F > F_{\text{кр}}$ уравнение регрессии является значимым.

5. Доверительные интервалы для коэффициентов a и b .

Среднеквадратические отклонения s_a и s_b коэффициентов a и b , определяются по формулам

$$s_a = \sqrt{\frac{\sum_{i=1}^n [y_i - (ax_i + b)]^2}{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}} = 0,141,$$

$$s_b = \sqrt{\frac{\sum_{i=1}^n [y_i - (ax_i + b)]^2}{(n-2)n}} = 0,118.$$

При $\gamma = 0,95$, $k = n - 2 = 28$, $t_\gamma = 2,04$ (прил. 2).

Находим доверительный интервал коэффициента a

$$a - t_\gamma \cdot s_a = 1,36 - 2,04 \cdot 0,141 = 1,07;$$

$$a + t_\gamma \cdot s_a = 1,36 + 2,04 \cdot 0,141 = 1,65;$$

$$1,07 \leq a_{\text{ист}} \leq 1,65.$$

Аналогично определим доверительный интервал для b

$$b - t_\gamma \cdot s_b = 1,68 - 2,04 \cdot 0,118 = 1,44;$$

$$b + t_\gamma \cdot s_b = 1,68 + 2,04 \cdot 0,118 = 1,92;$$

$$1,44 \leq b_{\text{ист}} \leq 1,92.$$

6. Проверка значимости коэффициента корреляции.

Проверим гипотезу о значимости выборочного коэффициента корреляции $r = 0,88$ (при $\alpha = 0,05$).

Для проверки гипотезы $H_0 : \rho = 0$ вычислим статистический критерий

$$T_{\text{набл}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,88\sqrt{30-2}}{\sqrt{1-0,88^2}} = 9,80.$$

В таблице критических распределений Стьюдента при $\gamma = 1 - \alpha = 0,95$ и $k = n - 2 = 28$ (прил. 2) либо с помощью соответствующих функций Mathcad, Excel находим критическое значение

$$t_{\text{кр}}(28) = 2,05.$$

Так как $|T_{\text{набл}}| > t_{\text{кр}}(28)$, то нулевую гипотезу $H_0 : \rho = 0$ отвергаем в пользу гипотезы $H_1 : \rho \neq 0$.

Таким образом, выборочное значение коэффициента корреляции $r = 0,88$ оказалось статистически значимым, т.е. генеральный коэффициент корреляции значимо отличается от нуля ($\rho \neq 0$).

Вопросы к защите лабораторной работы № 2

1. Что является основной задачей корреляционного анализа?
2. В чем разница между функциональной, стохастической и корреляционной зависимостями?
3. Что означает коррелированность случайных величин X и Y ?
4. Объясните смысл выборочного коэффициента корреляции и его значимость.
5. В каком случае коэффициент корреляции равен по модулю единице?
6. Как проверить значимость коэффициента парной корреляции?
7. Как строится ковариационная матрица?
8. В чем заключается задача регрессионного анализа?
9. В чем состоит значение уравнения регрессии?
10. Объясните связь между уравнением линейной регрессии и методом наименьших квадратов
11. Что характеризуют параметры уравнения регрессии?
12. Какая существует связь между линейным коэффициентом корреляции и коэффициентом регрессии?
13. Перечислите основные предпосылки регрессионного анализа.
14. Сформулируйте понятие интервальной оценки и правило проверки значимости уравнения регрессии.

Варианты заданий

Вариант № 1

1	2	3	4	5	6	7	8	9	10
28,3	28,4	28,5	30,5	28,5	28,6	30,6	28,8	28,8	29,3
26,1	27,2	26,3	26,2	27,3	26,5	27,7	26,9	26,9	26,0

11	12	13	14	15	16	17	18	19	20
31,8	31,1	31,4	30,1	31,2	31,9	37,3	35,8	34,3	32,3
27,4	26,8	27,0	27,6	28,6	28,9	33,5	32,2	31,0	31,2

21	22	23	24	25	26	27	28	29	30
30,9	30,4	30,2	29,9	29,4	29,0	28,9	29,7	29,3	29,5
28,5	28,1	27,9	27,8	27,5	27,0	27,3	26,7	26,6	26,4

Вариант № 2

1	2	3	4	5	6	7	8	9	10
9,6	11,9	10,2	11,2	9,6	9,7	13,3	10,7	9,5	12,2
-2,6	-2,2	-2,4	-1,7	-2,1	-2,5	-2,3	-2,8	-2,9	-2,7

11	12	13	14	15	16	17	18	19	20
10,3	11,5	13,2	12,4	10,9	13,7	12,5	10,6	13,8	14,1
-1,9	-2,0	-3,0	-3,3	-3,2	-3,7	-3,9	-3,5	-3,1	-3,8

21	22	23	24	25	26	27	28	29	30
9,2	8,8	8,5	9,2	7,8	13,4	13,9	14,3	14,5	13,0
-4,3	-4,1	-4,5	-4,7	-4,9	-3,4	-3,6	-4,2	-4,6	-1,8

Вариант № 3

1	2	3	4	5	6	7	8	9	10
12,6	12,5	12,9	13,1	13,7	13,5	13,7	12,8	13,1	12,9
24,0	23,8	23,6	23,4	23,3	23,2	23,1	22,8	22,6	22,3

11	12	13	14	15	16	17	18	19	20
13,1	12,9	12,9	12,7	12,5	12,7	12,8	12,7	12,7	12,9
22,1	22,0	21,9	21,8	21,8	21,7	21,7	21,7	21,7	21,6

21	22	23	24	25	26	27	28	29	30
12,9	12,9	13,0	12,9	13,0	13,0	12,9	13	12,9	12,8
21,6	21,6	21,6	21,6	21,5	21,5	21,5	21,5	21,4	21,4

Вариант № 4

1	2	3	4	5	6	7	8	9	10
0,9	0,9	1,8	0,9	1,3	0,4	1,8	2,2	1,8	1,8
9,4	9,4	9,4	9,4	9,4	9,6	9,6	9,6	9,7	9,7

11	12	13	14	15	16	17	18	19	20
1,8	2,2	2,2	2,7	2,2	1,8	2,2	1,8	1,8	1,8
9,7	9,9	9,9	10,1	10,1	10,3	10,3	10,3	10,5	10,5

21	22	23	24	25	26	27	28	29	30
1,8	1,8	2,2	1,8	1,3	0,9	1,3	0,4	0,9	0,9
10,5	10,5	10,5	10,5	10,7	10,7	10,7	10,7	10,7	10,9

Вариант № 5

1	2	3	4	5	6	7	8	9	10
4,0	4,2	4,4	4,5	4,8	5,1	4,7	5,1	5,5	5,5
2,2	1,8	4,0	4,0	4,9	3,6	4,5	4,9	5,4	4,5

11	12	13	14	15	16	17	18	19	20
5,5	5,6	6,0	6,1	6,2	6,4	6,9	7,0	7,2	7,1
4,5	5,4	4,9	5,4	6,3	6,7	6,7	7,2	6,7	5,8

21	22	23	24	25	26	27	28	29	30
7,1	7,6	8,0	8,2	8,3	8,6	8,7	8,8	8,7	8,7
5,8	5,4	5,8	4,0	3,6	3,1	2,7	2,2	2,7	2,7

Вариант № 6

1	2	3	4	5	6	7	8	9	10
10,6	10,9	11,1	11,0	10,9	11,1	11,3	11,4	11,4	11,5
10,7	10,9	11,2	11,2	11,0	11,6	11,4	11,5	11,5	11,6

11	12	13	14	15	16	17	18	19	20
11,5	16,4	11,9	12,1	12,1	10,8	10,6	10,4	10,1	10,2
11,6	11,6	11,4	15,1	11,2	11,1	10,2	10,6	10,4	10,2

21	22	23	24	25	26	27	28	29	30
10,2	10,1	10,2	10,4	10,7	11,1	11,4	11,3	12,1	12,2
10,2	10,2	10,3	10,4	10,7	11,1	11,4	11,4	12,1	12,3

Вариант № 7

1	2	3	4	5	6	7	8	9	10
3,2	3,2	2,4	2,4	2,4	2,4	3,2	2,4	2,4	3,2
6,3	5,4	5,8	5,4	4,5	5,4	4,9	5,4	4,9	5,8

11	12	13	14	15	16	17	18	19	20
3,2	3,2	4,0	4,0	4,0	4,8	4,8	4,8	4,0	4,8
5,8	6,3	7,2	7,2	7,6	8,5	7,2	8,9	8,9	7,6

21	22	23	24	25	26	27	28	29	30
4,8	4,8	4,8	4,8	4,8	4	4	3,2	3,2	2,4
7,6	9,4	8	8,5	7,6	6,7	7,2	7,2	5,8	5,8

Вариант № 8

1	2	3	4	5	6	7	8	9	10
7,3	7,1	7,4	7,4	7,5	7,3	7,2	7,1	6,8	6,9
18,1	18,2	17,9	16,9	16,7	17,2	17,7	17,7	17,8	17,8

11	12	13	14	15	16	17	18	19	20
6,9	6,8	6,8	6,8	6,8	6,7	6,6	6,6	6,6	6,2
17,8	18,1	17,9	17,9	17,9	17,8	17,8	17,8	17,8	17,8

21	22	23	24	25	26	27	28	29	30
6,2	6,1	6,1	6	6	6	5,9	5,9	5,8	5,5
17,8	17,8	17,8	17,8	17,8	17,8	17,8	17,8	17,7	17,7

Вариант № 9

1	2	3	4	5	6	7	8	9	10
8,1	8	7,8	7,6	7,7	7,7	7	7,3	7,1	6,6
16,8	16,8	16,7	16,7	16,6	16,6	16,6	16,6	16,6	16,4

11	12	13	14	15	16	17	18	19	20
6,6	6,2	6,3	6,3	6,1	5,9	5,9	5,9	6,2	6,3
16,4	16,4	16,3	16,3	16,3	16,3	16,3	16,4	16,6	16,7

21	22	23	24	25	26	27	28	29	30
6,3	7,6	6,2	7	6,6	6,9	8	6,2	6,9	6,9
16,7	17,8	19,1	20,5	20,7	20,9	21,2	20,9	21,2	20,6

Вариант № 10

1	2	3	4	5	6	7	8	9	10
-6,8	-8,1	-6,9	-5,7	-7	-7	-6,7	-6,6	-5,4	-5,5
28,2	28,2	28,2	28,2	28,2	28,2	28,2	28,2	28,2	28,3

11	12	13	14	15	16	17	18	19	20
-5,5	-7	-7,1	-7,1	-7,3	-7,2	-6,1	-6	-7,2	-7,3
28,3	28,3	28,3	28,3	28,3	28,4	28,4	28,4	28,5	28,5

21	22	23	24	25	26	27	28	29	30
-7,3	-7,3	-5,8	-5,5	-5,8	-6,2	-5,9	-6	-5,9	-6,1
28,5	28,5	28,5	28,6	28,6	28,6	28,7	28,7	28,7	28,7

Вариант № 11

1	2	3	4	5	6	7	8	9	10
15,1	14,9	14,6	14,2	14	14	14	14,2	14,4	14,9
10,3	10,3	10,4	10,2	10,2	10,2	10,2	10,4	10,6	10,7

11	12	13	14	15	16	17	18	19	20
14,9	15,7	15,9	16,2	16,8	16,8	17,1	18,5	19,6	19,7
10,7	11,2	11,3	11,6	12,1	12,1	12,4	13,1	12,1	13

21	22	23	24	25	26	27	28	29	30
19,7	20,3	21,1	22,4	23,6	24,7	25,1	25,6	25,6	25,2
13	13	13,5	14,5	14,3	13,3	13	12,4	11,7	11,7

Вариант № 12

1	2	3	4	5	6	7	8	9	10
1,7	2,2	2,1	2,3	1,9	1,8	1	0,6	0,3	0,2
8,5	8,2	8,6	8,8	9,3	8,8	9	8,9	8,8	8,9

11	12	13	14	15	16	17	18	19	20
3,9	3,6	2,5	2,2	2,1	1,9	1,8	1,4	1	0,6
12,6	12,6	12,9	12,7	12,7	12,5	12,6	12,5	12,4	12,3

21	22	23	24	25	26	27	28	29	30
0,8	0,9	2,2	1,6	1,8	1,9	2,6	1,1	0,1	0
12,2	12,2	12,7	12,7	12,5	12,3	12,3	12,3	12,2	12,7

Вариант № 13

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

11	12	13	14	15	16	17	18	19	20
12,7	11,8	12,6	12,6	12,9	12,7	12,7	12,5	12,6	12,5
2,5	3,7	3,9	3,6	2,5	2,2	2,1	1,9	1,8	1,4

21	22	23	24	25	26	27	28	29	30
12,5	5	5,2	5,5	4,8	5,1	5,5	5,6	5,5	5,3
1,4	1,1	1,1	1	1,2	1,2	2,4	1,9	1,4	1,4

Вариант № 14

1	2	3	4	5	6	7	8	9	10
0,4	0,4	0,4	0,4	0,4	0,9	1,3	1,8	1,3	0,9
0	0,3	0,7	0,8	0,8	1,1	1,9	1,9	2,3	2,7

11	12	13	14	15	16	17	18	19	20
0,9	0,9	0,9	1,3	1,8	2,7	2,2	2,7	2,2	1,8
2,7	2,5	1,9	1,8	2,1	1,3	1,1	0,9	0,7	0,7

21	22	23	24	25	26	27	28	29	30
1,8	1,8	1,3	1,8	0,9	1,3	1,3	0,9	0,9	0,9
2,3	2,7	2,9	3	3,1	3,1	2,9	2,7	2,4	2

Вариант № 15

1	2	3	4	5	6	7	8	9	10
18,6	19,6	20,8	21,3	21,7	21,8	22,9	22,8	23,4	23,1
18,2	20,9	21,7	22,7	22,6	23,1	22,9	21,9	20,5	18,4

11	12	13	14	15	16	17	18	19	20
14,9	15,3	15,6	16,5	17,1	17,4	18	18,7	18,3	18,9
3	5,5	6,6	8,2	7,5	8,5	9	13,3	6,5	8,3

21	22	23	24	25	26	27	28	29	30
17,5	17,9	18	17,7	18,2	18,3	18,3	18,8	18,5	18,8
17,9	19,4	20,7	22	21,9	20	21	20,9	18,2	15,4

Вариант № 16

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

11	12	13	14	15	16	17	18	19	20
10,2	9,5	8,4	9,1	10,1	9,7	9,7	8,7	7,4	9,2
13,1	11,7	10,2	10,7	9,2	6,9	4,3	2,7	1	0,3

21	22	23	24	25	26	27	28	29	30
18,8	18,8	17,7	16,9	17,4	16,1	15,6	15,3	14,4	13,1
0,2	0,8	1,9	4,7	7,2	9,3	10,9	12,1	12,9	13

Вариант № 17

1	2	3	4	5	6	7	8	9	10
-6,9	-5,5	-5,5	-5,5	-5,6	-6	-6	-5,9	-5,8	-5,8
29,3	29,3	29,3	29,3	29,2	28,6	28,6	28,7	28,8	28,9

11	12	13	14	15	16	17	18	19	20
-8,5	-8,7	-13	-7,7	-7,2	-6,9	-6,8	-6,8	-6,8	-6,9
24,6	27,1	25,4	28,2	28,9	29,3	29,4	29,4	29,4	29,3

21	22	23	24	25	26	27	28	29	30
-6,8	-6,5	-9,3	-9,4	-9,8	-8	-9,9	-9,6	-9,7	-9,5
29,4	29,8	20	17,4	19,9	27,8	25,3	25,9	27	24,6

Вариант № 18

1	2	3	4	5	6	7	8	9	10
9	9	8	11	12	9	6	7	6	7
29,4	29,8	20	17,4	19,9	27,8	25,3	25,9	27	24,6

11	12	13	14	15	16	17	18	19	20
-3,1	-3,7	-3,2	-3,4	-4,1	-4,3	-4,3	-4,5	-4,5	-5,2
25,5	26	26,1	26,2	26,3	26,3	26,4	26,5	26,5	26,6

21	22	23	24	25	26	27	28	29	30
5,2	4,3	4,1	4,6	4,4	5,1	6,8	7	5,4	6,7
19,3	19,2	19,2	19,1	19,1	19,2	19,4	19,6	19,6	19,8

Вариант № 19

1	2	3	4	5	6	7	8	9	10
6,6	5,8	5,5	6	5,6	5,7	6,5	6,2	6,3	7,1
53	53	53	52	52	52	51	51	51	50

11	12	13	14	15	16	17	18	19	20
7,1	7,1	7,4	7,9	7,8	8,2	8,6	8,1	9,2	8,7
50	49	46	47	48	48	49	49	49	47

21	22	23	24	25	26	27	28	29	30
8,7	9,1	7,8	7,8	8,1	7,8	7,4	7,3	7,1	6,6
47	49	50	51	51	51	52	52	53	53

Вариант № 20

1	2	3	4	5	6	7	8	9	10
6,8	6,6	6,6	6,7	6,8	6,8	6,8	7	7	7,1
10,1	9,8	10,1	9,8	9,8	9,8	9,8	9,8	9,8	9,8

11	12	13	14	15	16	17	18	19	20
7,1	7,1	7	7	6,8	6,8	6,8	6,7	6,7	6,7
9,8	9,8	9,8	9,8	9,8	9,8	9,8	9,7	9,7	9,7

21	22	23	24	25	26	27	28	29	30
6,7	6,7	6,7	6,7	6,6	6,6	6,5	6,5	6,5	6,6
9,7	9,7	9,7	9,7	9,7	9,7	9,7	9,6	9,6	9,6

Вариант № 21

1	2	3	4	5	6	7	8	9	10
7,3	7,3	7,3	7,3	7,2	7,2	7,1	7,1	7,1	7,1
5,9	5,9	5,9	5,3	5,4	5,9	5,3	4,8	4,7	5,3

11	12	13	14	15	16	17	18	19	20
7,1	7,1	7,2	7,1	7,1	7,1	6,9	6,9	6,9	6,8
5,3	5,3	5,3	4,8	5,3	5,3	4,6	4,2	4,1	4,4

21	22	23	24	25	26	27	28	29	30
6,8	6,7	6,6	6,6	6,4	6,3	6,4	6,4	6,4	6,3
4,4	3,8	3,8	3,7	3,4	3,3	2,9	3,4	3,4	3,2

Вариант № 22

1	2	3	4	5	6	7	8	9	10
6,9	6,8	6,7	6,6	6,6	6,4	6,4	6,5	6,5	6,4
5,8	5,2	5,2	5,1	4,8	4,8	4,5	4,9	4,9	4,7

11	12	13	14	15	16	17	18	19	20
6,4	6,3	6,3	6,2	6,3	6,4	6,3	6,3	6,1	6,1
4,7	4,7	4,9	4,4	5,1	5,6	5	4,1	4,2	4,3

21	22	23	24	25	26	27	28	29	30
6,1	6,2	6,3	7,2	7,4	8,4	8,6	8,7	8,3	7,7
4,3	4,4	5	6,1	6,4	7,8	7,4	7,1	6,3	6,2

Вариант № 23

1	2	3	4	5	6	7	8	9	10
5,9	6,5	6,1	6,3	7,1	7,5	7,9	8,3	7,7	7,3
8,5	8,6	8,4	8,2	8,3	8,4	8,1	8,3	7,9	7,8

11	12	13	14	15	16	17	18	19	20
7,2	6,4	6,3	5,6	5,4	5,9	5,7	5,8	5,5	5,2
7,6	7,7	7,6	7,6	7,5	7,4	7,3	7,4	7,1	7,3

21	22	23	24	25	26	27	28	29	30
5,2	4,9	4,7	4,4	4,1	3,9	3,8	3,7	3,6	3,9
7,3	7,6	7,3	7,1	7,3	7,2	6,8	6,9	6,7	6,7

Вариант № 24

1	2	3	4	5	6	7	8	9	10
5,2	4,9	4,7	4,4	4,1	3,9	3,8	3,7	3,6	3,4
23,1	23,1	23,1	23,1	23,1	22,9	22,9	22,7	22,7	22,7

11	12	13	14	15	16	17	18	19	20
3,4	3,3	3,2	2,9	2,7	2,6	2,8	2,8	2,6	2,3
22,7	22,7	22,7	22,7	22,7	22,7	22,7	22,7	22,7	22,7

21	22	23	24	25	26	27	28	29	30
2,3	2,8	3,1	4,5	5,9	6,6	7,1	8,1	8,4	10,1
22,7	22,8	23,3	24,2	24,8	23,4	22,3	20,9	20,9	21,2

Вариант № 25

1	2	3	4	5	6	7	8	9	10
-10,6	-11	-11,2	-11,2	-10,3	-9,9	-10	-9,4	-9,5	-9,7
-4,3	-10,2	-11,4	-12,4	-12,3	-12,9	-12	-12,8	-8,5	-8

11	12	13	14	15	16	17	18	19	20
-9,7	-9,5	-9,5	-9,9	-10	-9,4	-9,4	-9,5	-9,7	-10,6
-8	-5,6	-3,7	-1,8	-1,3	-4,4	-10,7	-14,9	-16,3	-17,1

21	22	23	24	25	26	27	28	29	30
-10,6	-10,3	-10,0	-10,5	-10,4	-10,4	-9,8	-9,9	-10,5	-10,1
-17,1	-15,4	-15,9	-16,3	-16,6	-17,1	-13,0	-11,4	-10,4	-9,9

Вариант № 26

1	2	3	4	5	6	7	8	9	10
5,9	6,1	6,1	6,3	7,1	7,1	7,9	8,3	7,7	7
8,5	8,6	8,6	8,2	8,3	8,4	8,4	8	7,9	7,8

11	12	13	14	15	16	17	18	19	20
7	6,4	6	5,6	5,4	5,4	5,7	5,8	5,5	5,2
7,8	7,7	7,6	7,6	7,5	7,4	7,4	7,4	7,4	7,3

21	22	23	24	25	26	27	28	29	30
5,2	4,9	4,7	4,4	4,1	3,9	3,8	3,7	3,6	3,4
7,3	7,3	7,3	7,3	7,3	7,2	6,8	6,7	6,7	6,7

Вариант № 27

1	2	3	4	5	6	7	8	9	10
2,9	4,1	4,1	4,3	5,1	5,1	5,9	6,3	5,4	5,0
5,5	5,6	5,6	5,2	5,3	5,4	5,4	5,1	4,9	4,8

11	12	13	14	15	16	17	18	19	20
5,2	4,4	4,3	3,6	3,4	3,4	3,7	3,8	3,5	3,2
4,6	4,7	4,6	4,7	4,5	4,4	4,4	4,4	4,4	4,3

21	22	23	24	25	26	27	28	29	30
3,2	2,9	2,7	2,4	2,1	1,9	1,8	1,7	1,6	1,9
4,3	4,3	4,5	4,3	4,3	4,2	3,8	3,7	3,7	3,7

Вариант № 28

1	2	3	4	5	6	7	8	9	10
2,4	2,1	2,4	2,6	2,4	2,9	2,3	2,8	2,3	2,9
1,0	1,3	1,7	1,8	1,8	2,1	1,9	2,9	2,3	2,7

11	12	13	14	15	16	17	18	19	20
2,4	2,9	2,4	3,3	3,8	3,7	3,2	3,7	3,2	3,8
2,7	2,5	2,9	2,8	2,1	2,3	1,1	1,9	2,7	1,7

21	22	23	24	25	26	27	28	29	30
3,8	2,8	3,3	2,8	2,9	2,3	2,3	2,1	2,7	2,9
2,3	2,7	2,9	3,2	3,7	3,1	1,9	2,7	2,4	2,9

Вариант № 29

1	2	3	4	5	6	7	8	9	10
1,4	1,2	1,2	1,1	1,4	1,3	1,9	1,8	2,3	1,9
0,3	0,4	0,7	0,6	0,8	1,1	2,0	1,9	2,3	2,7

11	12	13	14	15	16	17	18	19	20
1,9	1,8	1,9	2,3	1,8	2,7	3,2	3,2	3,2	2,8
2,7	2,5	1,9	1,8	2,1	1,4	1,7	0,9	1,7	1,4

21	22	23	24	25	26	27	28	29	30
----	----	----	----	----	----	----	----	----	----

2,8	2,8	2,3	1,8	1,9	2,3	2,3	1,9	1,5	1,9
2,3	2,7	2,9	3,1	3,1	3,3	2,9	2,7	2,4	2,5

Вариант № 30

1	2	3	4	5	6	7	8	9	10
0,6	0,7	0,6	0,5	0,4	0,8	1,3	1,5	1,1	0,9
0,4	0,2	0,5	0,8	0,7	1,2	1,7	1,9	2,1	2,4

11	12	13	14	15	16	17	18	19	20
1,3	0,8	1,1	1,6	1,8	2,4	2,8	2,6	2,2	1,8
2,5	2,1	1,7	1,8	2,1	1,9	1,5	1,2	0,9	1,7

21	22	23	24	25	26	27	28	29	30
1,8	1,9	1,6	1,8	1,2	1,3	1,5	1,1	0,9	0,8
2,3	2,7	2,9	2,5	3,1	3,1	2,9	2,4	2,3	1,8

СПИСОК ЛИТЕРАТУРЫ

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М.: Юрайт, 2019.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Юрайт, 2019.
3. Вентцель Е.С. Теория вероятностей. М.: Высшая школа, 1999.
4. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения: учеб. пособие для вузов. М.: Академия, 2003.
5. Вентцель Е.С., Овчаров Л.А. Задачи и упражнения по теории вероятностей: учеб. пособие для втузов. М.: Академия, 2005.
6. Кремер Н.Ш. Теория вероятностей и математическая статистика. М.: Юнити-Дана, 2018.
7. Письменный Д.Т. Конспект лекций по теории вероятностей и математической статистике. М.: Айрис-пресс, 2004.
8. Тобоев В.А., Каргузова Т.В., Толстов М.С. Прикладной статистический анализ для инженеров. Чебоксары. 2012.

ПРИЛОЖЕНИЯ

1. Таблица значений функции Лапласа $\Phi(x)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,27	0,1064	0,54	0,2054	0,81	0,2910
0,01	0,0040	0,28	0,1103	0,55	0,2088	0,82	0,2939
0,02	0,0080	0,29	0,1141	0,56	0,2123	0,83	0,2967
0,03	0,0120	0,30	0,1179	0,57	0,2157	0,84	0,2995
0,04	0,0160	0,31	0,1217	0,58	0,2190	0,85	0,3023
0,05	0,0199	0,32	0,1255	0,59	0,2224	0,86	0,3051
0,06	0,0239	0,33	0,1293	0,6	0,2257	0,87	0,3078
0,07	0,0279	0,34	0,1331	0,61	0,2291	0,88	0,3106
0,08	0,0319	0,35	0,1368	0,62	0,2324	0,89	0,3133
0,09	0,0359	0,36	0,1406	0,63	0,2357	0,9	0,3159
0,10	0,0398	0,37	0,1443	0,64	0,2389	0,91	0,3186
0,11	0,0438	0,38	0,1480	0,65	0,2422	0,92	0,3212
0,12	0,0478	0,39	0,1517	0,66	0,2454	0,93	0,3238
0,13	0,0517	0,4	0,1554	0,67	0,2486	0,94	0,3264
0,14	0,0557	0,41	0,1591	0,68	0,2517	0,95	0,3289
0,15	0,0596	0,42	0,1628	0,69	0,2549	0,96	0,3315
0,16	0,0636	0,43	0,1664	0,7	0,2580	0,97	0,3340
0,17	0,0675	0,44	0,1700	0,71	0,2611	0,98	0,3365
0,18	0,0714	0,45	0,1736	0,72	0,2642	0,99	0,3389
0,19	0,0753	0,46	0,1772	0,73	0,2673	1	0,3413
0,20	0,0793	0,47	0,1808	0,74	0,2704	1,01	0,3438
0,21	0,0832	0,48	0,1844	0,75	0,2734	1,02	0,3461
0,22	0,0871	0,49	0,1879	0,76	0,2764	1,03	0,3485
0,23	0,0910	0,5	0,1915	0,77	0,2794	1,04	0,3508
0,24	0,0948	0,51	0,1950	0,78	0,2823	1,05	0,3531
0,25	0,0987	0,52	0,1985	0,79	0,2852	1,06	0,3554
0,26	0,1026	0,53	0,2019	0,8	0,2881	1,07	0,3577
1,08	0,3599	1,43	0,4236	1,78	0,4625	2,26	0,4881
1,09	0,3621	1,44	0,4251	1,79	0,4633	2,28	0,4887
1,1	0,3643	1,45	0,4265	1,8	0,4641	2,3	0,4893

1,11	0,3665	1,46	0,4279	1,81	0,4649	2,32	0,4898
1,12	0,3686	1,47	0,4292	1,82	0,4656	2,34	0,4904
1,13	0,3708	1,48	0,4306	1,83	0,4664	2,36	0,4909
1,14	0,3729	1,49	0,4319	1,84	0,4671	2,38	0,4913
1,15	0,3749	1,5	0,4332	1,85	0,4678	2,4	0,4918
1,16	0,3770	1,51	0,4345	1,86	0,4686	2,42	0,4922
1,17	0,3790	1,52	0,4357	1,87	0,4693	2,44	0,4927
1,18	0,3810	1,53	0,4370	1,88	0,4699	2,46	0,4931
1,19	0,3830	1,54	0,4382	1,89	0,4706	2,48	0,4934
1,2	0,3849	1,55	0,4394	1,9	0,4713	2,5	0,4938
1,21	0,3869	1,56	0,4406	1,91	0,4719	2,52	0,4941
1,22	0,3888	1,57	0,4418	1,92	0,4726	2,54	0,4945
1,23	0,3907	1,58	0,4429	1,93	0,4732	2,56	0,4948
1,24	0,3925	1,59	0,4441	1,94	0,4738	2,58	0,4951
1,25	0,3944	1,6	0,4452	1,95	0,4744	2,6	0,4953
1,26	0,3962	1,61	0,4463	1,96	0,4750	2,62	0,4956
1,27	0,3980	1,62	0,4474	1,97	0,4756	2,64	0,4959
1,28	0,3997	1,63	0,4484	1,98	0,4761	2,66	0,4961
1,29	0,4015	1,64	0,4495	1,99	0,4767	2,68	0,4963
1,3	0,4032	1,65	0,4505	2	0,4772	2,7	0,4965
1,31	0,4049	1,66	0,4515	2,02	0,4783	2,72	0,4967
,32	0,4066	1,67	0,4525	2,04	0,4793	2,74	0,4969
1,33	0,4082	1,68	0,4535	2,06	0,4803	2,76	0,4971
1,34	0,4099	1,69	0,4545	2,08	0,4812	2,78	0,4973
1,35	0,4115	1,7	0,4554	2,1	0,4821	2,8	0,4974
1,36	0,4131	1,71	0,4564	2,12	0,4830	2,82	0,4976
1,37	0,4147	1,72	0,4573	2,14	0,4838	2,84	0,4977
1,38	0,4162	1,73	0,4582	2,16	0,4846	2,86	0,4979
1,39	0,4177	1,74	0,4591	2,18	0,4854	2,88	0,4980
1,4	0,4192	1,75	0,4599	2,2	0,4861	2,9	0,4981
1,41	0,4207	1,76	0,4608	2,22	0,4868	2,92	0,4982
1,42	0,4222	1,77	0,4616	2,24	0,4875	2,94	0,4984
2,96	0,4985	3,2	0,49931	3,8	0,49993	5	0,5
2,98	0,4986	3,4	0,49966	4	0,49997		
3	0,49865	3,6	0,49984	4,5	0,5		

2. Таблица значений $t_\gamma = t(\gamma, n)$

$n \backslash \gamma$	0,95	0,99	0,999
4	2,78	4,60	8,61
5	2,57	4,03	6,86
6	2,45	3,71	5,96
7	2,37	3,50	5,41
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,06	4,32
13	2,16	3,01	4,22
14	2,15	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,02
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,093	2,861	3,883
20	2,085	2,845	3, 845
25	2,064	2,797	3,745
30	2,045	2,756	3,659
35	2,032	2,720	3,600
40	2,023	2,708	3,558
45	2,016	2,692	3,527
50	2,009	2,679	3,502
60	2,001	2,662	3,464
70	1,996	2,649	3,439
80	1,991	2,640	3,418
90	1,987	2,633	3,403
100	1,984	2,627	3,392
120	1,980	2,617	3,374
∞	1,960	2,576	3,291

3. Критические точки распределения χ^2 – Пирсона

Число степеней свободы ν	Уровень значимости α		
	0,01	0,025	0,05
1	6,6	5,0	3,8
2	9,2	7,4	6,0
3	11,3	9,4	7,8
4	13,3	11,1	9,5
5	15,1	12,8	11,1
6	16,8	14,4	12,6
7	18,5	16,0	14,1
8	20,1	17,5	15,5
9	21,7	19,0	16,9
10	23,2	20,5	18,3
11	24,7	21,9	19,7
12	26,2	23,3	21,0
13	27,7	24,7	22,4
14	29,1	26,1	23,7
15	30,6	27,5	25,0
16	32,0	28,8	26,3
17	33,4	30,2	27,6
18	34,8	31,5	28,9
19	36,2	32,9	30,1
20	37,6	34,2	31,4
21	38,9	35,5	32,7
22	40,3	36,8	33,9
23	41,6	38,1	35,2
24	43,0	39,4	36,4
25	44,3	40,6	37,7
26	45,6	41,9	38,9
27	47,0	43,2	40,1
28	48,3	44,5	41,3
29	49,6	45,7	42,6
30	50,9	47,0	43,8

4. Критические точки распределения F Фишера – Снедекора

Уровень значимости $\alpha = 0,05$ (k_1 – число степеней свободы большей дисперсии, k_2 – число степеней свободы меньшей дисперсии)

$k_2 \backslash k_1$	1	2	3	4	5	6
1	161	200	216	225	230	234
2	18,5	19,00	19,16	19,25	19,30	19,33
3	10,13	9,55	9,28	9,12	9,01	8,94
4	7,71	6,94	6,59	6,39	6,26	6,16
5	6,61	5,79	5,41	5,19	5,05	4,95
6	5,99	5,14	4,76	4,53	4,39	4,28
7	5,59	4,74	4,35	4,12	3,97	3,87
8	5,32	4,46	4,07	3,84	3,69	3,58
9	5,12	4,26	3,86	3,63	3,48	3,37
10	4,96	4,10	3,71	3,48	3,33	3,22
11	4,84	3,98	3,59	3,36	3,20	3,09
12	4,75	3,88	3,49	3,26	3,11	3,00
13	4,67	3,80	3,41	3,18	3,02	2,92
14	4,60	3,74	3,34	3,11	2,96	2,85
15	4,54	3,68	3,29	3,06	2,90	2,79
16	4,49	3,63	3,24	3,01	2,85	2,74
17	4,45	3,59	3,20	2,96	2,81	2,70
18	4,41	3,55	3,16	2,93	2,77	3,66
19	4,38	3,52	3,13	2,90	2,74	2,63
20	4,35	3,49	3,10	2,87	2,71	2,60
21	4,32	3,47	3,07	2,84	2,68	2,57
22	4,30	3,44	3,05	2,82	2,66	2,55
23	4,28	3,42	3,03	2,80	2,64	2,53

24	4,26	3,40	3,01	2,78	2,62	2,51
25	4,24	3,88	2,99	2,76	2,60	2,49
26	4,22	3,37	2,98	2,74	2,59	2,47
27	4,21	3,35	2,96	2,73	2,57	2,46
28	4,20	3,34	2,95	2,71	2,56	2,44
29	4,18	3,33	2,93	2,70	2,54	2,43
30	4,17	3,32	2,92	2,69	2,53	2,42
40	4,08	3,23	2,84	2,61	2,45	2,34
60	4,00	3,15	2,76	2,53	2,37	2,25
120	3,92	3,07	2,68	2,45	2,29	2,17
∞	3,84	2,99	2,60	2,37	2,21	2,09

**5. Значение λ_q распределения Колмогорова
для вероятности $q = P\{\xi_n \geq \lambda_q\}$**

q	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
λ_q	0,828	0,895	0,975	1,073	1,224	1,358	1,520	1,627	1,950

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ.....	3
1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ. ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ	4
Примеры решения задач.....	18
Задачи для самостоятельного решения	37
Задания контрольной работы.....	41
Вопросы к защите контрольной работы	45
2. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. ПЕРВИЧНЫЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ. ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ	48
ЗАДАНИЯ ЛАБОРАТОРНОЙ РАБОТЫ № 1. Первичный анализ экспериментальных данных.....	57
Пример решения варианта задания	59
Вопросы к защите лабораторной работы.....	67
3. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ. ОСНОВНЫЕ ПОЛОЖЕНИЯ. ДВУМЕРНАЯ МОДЕЛЬ	75
ЗАДАНИЯ ЛАБОРАТОРНОЙ РАБОТЫ № 2. Корреляционный и регрессионный анализ	78
Пример решения варианта задания	81
Вопросы к защите лабораторной работы № 2	85
СПИСОК ЛИТЕРАТУРЫ	97
ПРИЛОЖЕНИЯ	98

Учебное издание

ТОБОЕВ Вячеслав Андреевич
ФРОЛОВ Сергей Иванович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Редактор *А.Н. Антонова*
Компьютерная верстка и правка *Е.В. Ивановой*

Согласно Закону № 436-ФЗ от 29 декабря 2010 года
данная продукция не подлежит маркировке

Подписано в печать 28.04.2023. Формат 60×84/16.
Бумага газетная. Печать офсетная. Гарнитура Times.
Усл. печ. л. 6,04. Уч.-изд. 5,98. Тираж 100 экз. Заказ № 568.

Чувашский государственный университет имени И.Н. Ульянова
Типография университета
428015 Чебоксары, Московский просп., 15