

Министерство образования и науки Российской Федерации
Федеральное агентство по образованию
Московский государственный институт электронной техники
(технический университет)

**В.В.Бардушкин, В.В.Лесин,
В.Н.Земсков, Н.Н.Мустафин**

**Лабораторный практикум по курсу
«Теория вероятностей и
математическая статистика»**

Утверждено редакционно-издательским советом института

Москва 2009

УДК 519.2

Рецензент докт. физ.-мат. наук, проф. *В.Б. Яковлев*

Бардушкин В.В., Лесин В.В., Земсков В.Н., Мустафин Н.Н.

Лабораторный практикум по курсу «Теория вероятностей и математическая статистика». – М.: МИЭТ, 2009. – 116 с.: ил.

Лабораторный практикум состоит из трёх разделов и двух приложений. Первый раздел содержит необходимые теоретические сведения по основным методам обработки статистических данных. Во втором разделе приведены общие сведения о системе программирования Matlab с пакетом прикладных программ по статистике «Statistics Toolbox», элементы программирования на языке Matlab, а также описаны средства пакета для изучения одномерных распределений случайных величин. Третий раздел содержит описание четырёх лабораторных работ. В приложениях приведены варианты индивидуальных заданий.

Предназначен для студентов 2-го курса технических факультетов МИЭТ.

© МИЭТ, 2009

Бардушкин Владимир Валентинович

Лесин Виктор Васильевич

Земсков Владимир Николаевич

Мустафин Наиль Нухович

Лабораторный практикум по курсу «Теория вероятностей и математическая статистика»

Редактор *Е.Г. Кузнецова*. Технический редактор *Л.Г. Лосякова*. Вёрстка авторов.

Подписано в печать с оригинал-макета 03.03.09. Формат 60x84 1/16. Печать офсетная. Бумага офсетная. Гарнитура Times New Roman. Усл. печ. л. 6,73.

Уч.-изд. л. 5,8. Тираж 1000 экз. Заказ 25.

Отпечатано в типографии ИПК МИЭТ.

124498, Москва, Зеленоград, проезд 4806, д. 5, МИЭТ.

1. Краткие теоретические сведения

1.1. Основные распределения, используемые в математической статистике

Равномерное распределение. Случайная величина непрерывного типа X распределена *равномерно* на отрезке $[a; b]$ (обозначение $X \sim \mathbf{R}(a; b)$), если плотность вероятности имеет вид:

$$f_X(x) = \begin{cases} 0, & x \notin [a; b], \\ \frac{1}{b-a}, & x \in [a; b]. \end{cases}$$

Графики плотности распределения, построенные в системе Matlab для нескольких случайных величин $\mathbf{R}(a; b)$, приведены на рис.1.1.

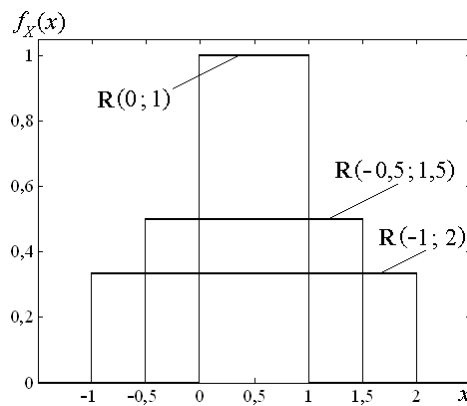


Рис.1.1. Графики плотности распределения $\mathbf{R}(a; b)$

Характеристическая функция распределения $\mathbf{R}(a; b)$ имеет вид $\varphi(t) = \frac{e^{ibt} - e^{iat}}{it(b-a)}$, $t \in \mathbf{R}$,

математическое ожидание $m_X = (a+b)/2$, дисперсия $D_X = (b-a)^2/12$. Равномерное распределение не имеет моды, а медиана совпадает с математическим ожиданием.

С помощью линейного преобразования $Y = (b-a)X + a$ от *стандартной равномерно распределённой величины* $X \sim \mathbf{R}(0; 1)$ можно перейти к случайной величине $Y \sim \mathbf{R}(a; b)$ с произвольными параметрами a и b .

Ошибка, возникающая в результате округления числа с точностью до 10^{-n} , удовлетворительно описывается равномерным на промежутке $[-0,5 \cdot 10^{-n}; 0,5 \cdot 10^{-n}]$ распределением.

Нормальное распределение. Случайная величина непрерывного типа X имеет *нормальное распределение*, или *распределение Гаусса*, с параметрами m и σ (обозначение $X \sim \mathbf{N}(m; \sigma)$), если плотность распределения имеет вид:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad m \in \mathbf{R}, \quad \sigma > 0.$$

Графики плотности распределения, построенные в системе Matlab для нескольких случайных величин $\mathbf{N}(m; \sigma)$, приведены на рис.1.2.

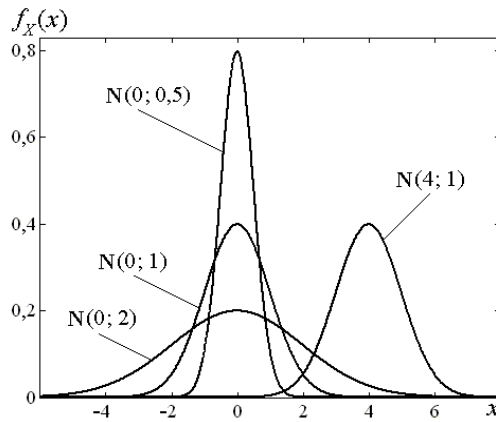


Рис.1.2. Графики плотности распределения $N(m; \sigma)$

Характеристическая функция распределения $N(m; \sigma)$ имеет вид $\varphi(t) = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right)$, $t \in \mathbf{R}$,

математическое ожидание $m_X = m$, дисперсия $D_X = \sigma^2$. Мода и медиана $X \sim N(m; \sigma)$ совпадают с математическим ожиданием.

Фундаментальная роль нормального распределения объясняется тем, что при некоторых (обычно выполняющихся) условиях закон распределения суммы случайных величин, имеющих произвольные законы распределения, приближается к нормальному закону с ростом числа суммируемых величин. Точная формулировка этого утверждения является содержанием *центральной предельной теоремы* [1].

С помощью линейного преобразования $Y = \sigma X + m$ от случайной величины со *стандартным нормальным распределением* $X \sim N(0; 1)$ можно перейти к случайной величине $Y \sim N(m; \sigma)$ с произвольными параметрами m и σ . Функция распределения стандартной нормальной величины

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

называется *функцией Лапласа*, значения которой табулируются (см., например, табл.П1 в [1]).

Нормально распределённая случайная величина с большой вероятностью принимает значения, близкие к своему математическому ожиданию, что выражается *правилом k сигм*:

$$P\{|X - m| < k\sigma\} = \begin{cases} 0,6827, & k = 1, \\ 0,9545, & k = 2, \\ 0,9973, & k = 3. \end{cases}$$

Обычно используется *правило трёх сигм*: значения нормальной случайной величины $X \sim N(m; \sigma)$ практически достоверно отклоняются от её математического ожидания m менее чем на 3σ .

Квантиль u_p нормального распределения является корнем уравнения $\Phi(u_p) = p$, т.е. $u_p = \Phi^{-1}(p)$.

Для обратной функции Лапласа $\Phi^{-1}(p)$ существуют специальные таблицы (см., например, табл.П1 в [1]). В силу симметричности нормального распределения

$$u_p = -u_{1-p}.$$

Вероятность попадания случайной величины $X \sim N(m; \sigma)$ в интервал $(a; b)$ вычисляется по формуле

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

Нормальное распределение *композиционно устойчиво*, т.е. сумма независимых нормально распределённых случайных величин также имеет нормальное распределение.

Гамма-распределение. Случайная величина непрерывного типа X имеет *гамма-распределение* с параметрами a и b (обозначение $X \sim \Gamma(a; b)$), если её плотность распределения имеет вид:

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}}, & x > 0, a > 0, b > 0, \\ 0, & x \leq 0, \end{cases}$$

где $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$ – гамма-функция, обладающая свойствами

$$\Gamma(x+1) = x\Gamma(x), \quad \Gamma(k+1) = k!, \quad \text{где } k \in \mathbf{N},$$

$$\Gamma(1) = \Gamma(2) = 1, \quad \Gamma(0,5) = \sqrt{\pi}.$$

Графики плотности распределения, построенные в системе Matlab для нескольких случайных величин $\Gamma(a; b)$, приведены на рис.1.3.

Характеристическая функция гамма-распределения имеет вид $\varphi(t) = (1 -ibt)^{-a}$, $t \in \mathbf{R}$, математическое ожидание $m_X = ab$, дисперсия $D_X = ab^2$.

Гамма-распределение *композиционно устойчиво* при фиксированном параметре b , т.е. сумма независимых случайных величин, имеющих гамма-распределение с параметром b , также имеет гамма-распределение с параметром b .

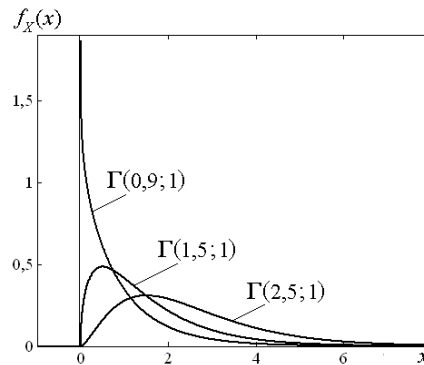


Рис.1.3. Графики плотности распределения $\Gamma(a; b)$

Экспоненциальное (показательное) распределение. Это распределение – частный случай гамма-распределения при $a = 1$ (обозначение $X \sim \mathbf{E}(\lambda)$, где $\lambda = 1/b$). Таким образом, функция плотности гамма-распределения имеет вид:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \lambda > 0, \\ 0, & x \leq 0. \end{cases}$$

Графики плотности распределения, построенные в системе Matlab для нескольких случайных величин $\mathbf{E}(\lambda)$, приведены на рис.1.4.

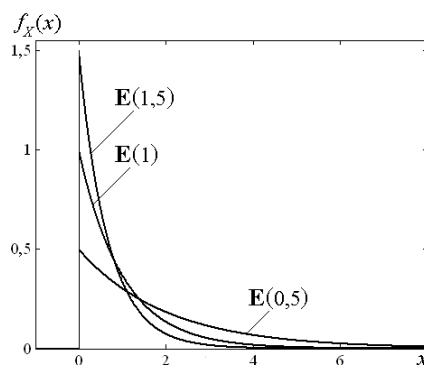


Рис.1.4. Графики плотности распределения $\mathbf{E}(\lambda)$

Характеристическая функция распределения $X \sim \mathbf{E}(\lambda)$ имеет вид $\varphi(t) = \lambda(\lambda - it)^{-1}$, $t \in \mathbf{R}$, математическое ожидание $m_X = 1/\lambda$, дисперсия $D_X = 1/\lambda^2$.

Распределение χ^2 («хи-квадрат») с k степенями свободы. Это распределение – частный случай гамма-распределения при $a = k/2$, $k \in \mathbf{N}$, $b = 2$ (обозначение $X \sim \chi^2(k)$). Таким образом, функция плотности распределения $\chi^2(k)$ имеет вид:

$$f_X(x) = \begin{cases} \frac{1}{2^{k/2} \Gamma(k/2)} e^{-\frac{x}{2}} x^{\frac{k}{2}-1}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Графики плотности распределения, построенные в системе Matlab для нескольких случайных величин $\chi^2(k)$, приведены на рис.1.5.

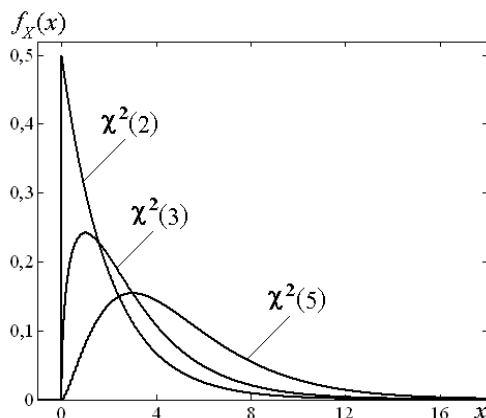


Рис.1.5. Графики плотности распределения $\chi^2(k)$

Характеристическая функция распределения $\chi^2(k)$ имеет вид $\varphi(t) = (1 - 2it)^{-k/2}$, $t \in \mathbf{R}$, математическое ожидание $m_X = k$, дисперсия $D_X = 2k$.

По закону $\chi^2(k)$ распределена сумма квадратов k независимых стандартных нормальных случайных величин, т.е. если случайные величины $X_i \sim \mathbf{N}(0; 1)$, $i = 1, 2, \dots, k$, то

$$Y = \sum_{i=1}^k X_i^2 \sim \chi^2(k).$$

Существуют таблицы квантилей распределения $\chi^2(k)$ при различных k (см., например, табл.П5 в [1]).

С ростом числа степеней свободы k распределение $\chi^2(k)$, согласно центральной предельной теореме, приближается к нормальному распределению, т.е. если $Y \sim \chi^2(k)$, то при больших k (на практике при $k \geq 30$) можно приближённо считать, что $Y \sim \mathbf{N}(k; \sqrt{2k})$. При этом квантили $\chi_p^2(k)$ распределения χ^2 могут быть приближённо найдены с помощью квантилей u_p стандартного нормального закона $\mathbf{N}(0; 1)$:

$$\chi_p^2(k) \approx u_p \sqrt{2k} + k.$$

Более точные значения квантилей $\chi_p^2(k)$ распределения χ^2 можно получить по формулам

$$\chi_p^2(k) \approx 0,5(\sqrt{2k-1} + u_p)^2$$

или

$$\chi_p^2(k) \approx k \left(1 - \frac{2}{9k} + u_p \sqrt{\frac{2}{9k}} \right)^3.$$

Один из критериев проверки согласия эмпирических данных с гипотетической функцией распределения $F_X(x)$ основан на изучении χ^2 -статистики Пирсона

$$\chi^2 = \sum_{j=1}^m \frac{(n_j - np_j)^2}{np_j},$$

где $p_j = F_X(x_j) - F_X(x_{j-1})$, $x_0 = -\infty < x_1 < \dots < x_m = +\infty$ — произвольное разбиение интервала $(-\infty; +\infty)$, n_j — число результатов измерений, попавших в промежуток $[x_{j-1}; x_j)$. χ^2 -статистики

Пирсона (в предположении истинности гипотезы о законе распределения) имеют в пределе при $n = \sum_{j=1}^m n_j \rightarrow \infty$ распределение $\chi^2(m-1)$ независимо от вида $F_X(x)$.

Распределение Стьюдента. Случайная величина непрерывного типа X имеет *распределение Стьюдента* с k степенями свободы (обозначение $X \sim \mathbf{St}(k)$), если её плотность распределения имеет вид:

$$f_X(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(k/2)\sqrt{\pi k}} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}, \quad x \in \mathbf{R}, \quad k \in \mathbf{N}.$$

Графики плотности распределения, построенные в системе Matlab для нескольких случайных величин $\mathbf{St}(k)$, приведены на рис.1.6.

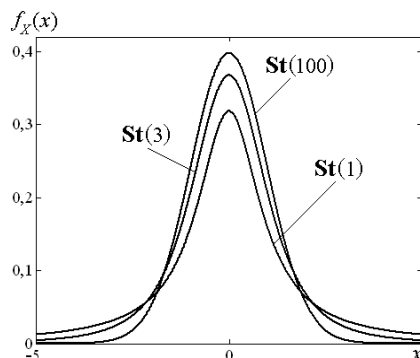


Рис.1.6. Графики плотности распределения $\mathbf{St}(k)$

Основные числовые характеристики распределения $\mathbf{St}(k)$:

- математическое ожидание $m_X = \begin{cases} 0, & k > 1, \\ \text{не существует,} & k = 1; \end{cases}$
- дисперсия $D_X = \begin{cases} k/(k-2), & k > 2, \\ \infty, & k \leq 2. \end{cases}$

Квантили $t_p(k)$ распределения $\mathbf{St}(k)$ находятся из таблиц (см., например, табл.Пб в [1]). В силу симметричности распределения

$$t_p(k) = -t_{1-p}(k).$$

При больших k распределение Стьюдента приближается к стандартному нормальному распределению. Поэтому при $k \geq 30$ можно использовать приближённое равенство $t_p(k) \approx u_p$, где u_p – квантиль распределения $\mathbf{N}(0; 1)$. Более точно квантили $t_p(k)$ можно находить по формуле

$$t_p(k) = u_p \left(\left(1 - \frac{1}{4k} \right)^2 - \frac{u_p^2}{2k} \right)^{-\frac{1}{2}}.$$

Распределение $\mathbf{St}(k)$ широко применяется в математической статистике. Это объясняется тем, что если случайные величины U и V независимы, причём $U \sim \mathbf{N}(0; 1)$ и $V \sim \chi^2(k)$, то отношение $\frac{U}{\sqrt{V/k}}$ имеет распределение Стьюдента с k степенями свободы:

$$\frac{U}{\sqrt{V/k}} \sim \mathbf{St}(k).$$

Распределение Фишера. Случайная величина непрерывного типа X имеет *распределение Фишера* с k_1 и k_2 степенями свободы (обозначение $X \sim \mathbf{F}(k_1; k_2)$), если её плотность распределения имеет вид:

$$f_X(x) = \begin{cases} \frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} \left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}} \frac{x^{\frac{k_1}{2}-1}}{\left(1+\frac{k_1}{k_2}x\right)^{\frac{k_1+k_2}{2}}}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

где $x \in \mathbf{R}$, $k_1, k_2 \in \mathbf{N}$.

Графики плотности распределения, построенные в системе Matlab для нескольких случайных величин $\mathbf{F}(k_1; k_2)$, приведены на рис.1.7.

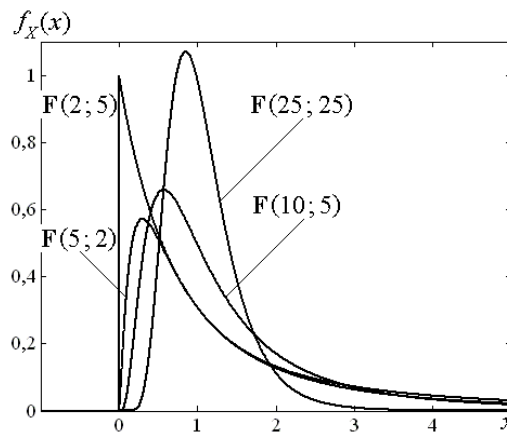


Рис.1.7. Графики плотности распределения $\mathbf{F}(k_1; k_2)$

Основные числовые характеристики распределения $\mathbf{F}(k_1; k_2)$:

- математическое ожидание $m_X = k_2/(k_2 - 2)$, $k_2 > 2$;
- дисперсия $D_X = \frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}$, $k_2 > 4$.

Квантили $F_p(k_1; k_2)$ распределения Фишера находятся из таблиц (см., например, табл.П7 в [1]), причём квантили порядков p и $1-p$ связаны соотношением

$$F_{1-p}(k_1; k_2) = \frac{1}{F_p(k_2; k_1)}.$$

При $k_1, k_2 \gg 1$ для вычисления квантилей $F_p(k_1; k_2)$ можно использовать приближённую формулу

$$F_p(k_1; k_2) \approx \frac{k_2}{k_2 - 2} \sqrt{\frac{2(k_1 + k_2 - 2)}{k_1(k_2 - 4)}} \cdot u_p + \frac{k_2}{k_2 - 2},$$

где u_p – квантиль порядка p распределения $\mathbf{N}(0; 1)$.

Распределение Фишера применяется в математической статистике в связи с тем, что если $X \sim \chi^2(k_1)$ и $Y \sim \chi^2(k_2)$ – независимые случайные величины, то

$$\frac{k_2 X}{k_1 Y} \sim \mathbf{F}(k_1; k_2).$$

1.2. Основные понятия математической статистики

Математическая статистика – это наука, изучающая случайные явления посредством обработки и анализа результатов наблюдений и измерений.

Случайную величину X будем называть *генеральной совокупностью* X . Исходным материалом для изучения свойств генеральной совокупности X являются статистические данные, т.е. значения X , полученные в результате повторения случайного опыта (измерения случайной величины X). Предполагается, что опыт может быть повторён сколько угодно раз в неизменных условиях. Это означает,

что распределение случайной величины X_i , $i = 1, 2, \dots$, заданной на множестве исходов i -го опыта, не зависит от i и совпадает с распределением генеральной совокупности X .

Набор n независимых в совокупности случайных величин (X_1, X_2, \dots, X_n) , где X_i соответствует i -му опыту, называется *случайной выборкой* из генеральной совокупности X . Число n называется *объёмом выборки*.

Совокупность чисел (x_1, x_2, \dots, x_n) , полученных в результате n -кратного повторения опыта по измерению генеральной совокупности X , называется *реализацией случайной выборки* или просто *выборкой объёмом n* .

В основе большинства результатов математической статистики лежит *выборочный метод*, состоящий в том, что свойства генеральной совокупности X устанавливаются путём изучения тех же свойств на случайной выборке.

Прежде чем перейти к детальному анализу статистических данных, обычно проводят их *предварительную обработку*. Иногда её результаты уже сами по себе дают ответы на многие вопросы, но в большинстве случаев они являются исходным материалом для дальнейшего анализа. Рассмотрим основные понятия, связанные с предварительной обработкой выборки.

Выборка (x_1, x_2, \dots, x_n) объёмом n из генеральной совокупности X , упорядоченная по неубыванию своих элементов, т.е. $x^{(1)} \leq x^{(2)} \leq \dots \leq x^{(n)}$, называется *вариационным рядом*:

$$x^{(1)}, x^{(2)}, \dots, x^{(n)}.$$

Разность между максимальным и минимальным элементами выборки $w = x^{(n)} - x^{(1)}$ называется *размахом выборки*.

Пусть выборка (x_1, x_2, \dots, x_n) содержит m различных чисел z_1, z_2, \dots, z_m , где $m \leq n$ и $z_1 < z_2 < \dots < z_m$, причём число z_j встречается в выборке n_j раз, $j = 1, 2, \dots, m$. Так бывает либо когда генеральная совокупность X – дискретная случайная величина, либо когда X – непрерывна, но её значения при измерении округляют. Число n_j называют *частотой* элемента выборки z_j , а отношение n_j/n – *относительной частотой* этого элемента.

Статистическим рядом выборки (x_1, x_2, \dots, x_n) называют таблицу, которая в первой строке содержит значения выборки z_j (напомним, что $z_1 < z_2 < \dots < z_m$), во второй строке – частоты n_j , а в третьей строке – относительные частоты n_j/n этих значений (табл.1.1). Статистические данные, представленные в виде статистического ряда, называют *группированными*.

Таблица 1.1

Статистический ряд выборки

Значения z_j	z_1	z_2	...	z_m
Частоты n_j	n_1	n_2	...	n_m
Относительные частоты n_j/n	n_1/n	n_2/n	...	n_m/n

Другой способ группировки, который используют обычно при больших объёмах выборки ($n \geq 50$), состоит в следующем. Отрезок $[x^{(1)}; x^{(n)}]$, содержащий выборку (x_1, x_2, \dots, x_n) , разбивают на m промежутков $\Delta_j = [d_{j-1}; d_j]$, как правило, одинаковой длины $\Delta = (x^{(n)} - x^{(1)})/m$. Далее подсчитывают частоты n_j попадания выборочных значений x_i в промежутки Δ_j и относительные частоты n_j/n . Получающийся в результате этого статистический ряд называют *интервальным статистическим рядом* (табл.1.2). При этом в качестве z_j рассматривают середины промежутков $\Delta_j = [d_{j-1}; d_j]$.

Таблица 1.2

Интервальный статистический ряд выборки

Промежутки Δ_j	$\Delta_1 = [d_0; d_1]$	$\Delta_2 = [d_1; d_2]$...	$\Delta_m = [d_{m-1}; d_m]$
Частоты n_j	n_1	n_2	...	n_m
Относительные частоты n_j/n	n_1/n	n_2/n	...	n_m/n

Может оказаться, что часть выборочных значений x_i совпадают с некоторыми границами d_j между соседними промежутками Δ_j и Δ_{j+1} . Тогда (исходя из определения функции распределения $F_X(x) = P\{X < x\}$) при подсчёте частот n_j и n_{j+1} для таких значений x_i считают, что $x_i \in \Delta_{j+1}$, т.е. каждое выборочное значение $x_i = d_j$ увеличивает частоту n_{j+1} на единицу и не влияет на частоту n_j . В частности, к первому интервалу группировки относят элементы выборки, попадающие на левую границу указанного промежутка. Кроме того, в последний интервал группировки включают также элементы выборки, попадающие на правую границу указанного промежутка.

Замечание. Многие специалисты определяют функцию распределения как $F_X(x) = P\{X \leq x\}$. Это относится и к разработчикам пакета Matlab. Поэтому система Matlab элементы выборки, попадающие на границу интервала, включает в соседний слева интервал группировки; кроме того, в первый интервал группировки Matlab включает также элементы выборки, попадающие на его левую границу.

Количество m промежутков, на которые разбивают отрезок $[x^{(1)}; x^{(n)}]$, выбирают в зависимости от объёма выборки n . Существуют различные способы выбора m . Например, ориентировочную оценку числа промежутков можно получить из соотношения $m = [1,86n^{0,33}]$, где $[a]$ – целая часть числа a . Часто для определения числа интервалов искусственной группировки пользуются формулой Старджеса [10, 13] $m \approx 1 + 3,322 \cdot \lg n$. Иногда m может быть задано природой исследуемого явления или условиями проведения эксперимента. Возможны ситуации, когда ширина каждого интервала может быть отличной от других (*неравномерная группировка*).

В некоторых источниках в качестве статистического ряда рассматривается таблица, содержащая либо только частоты n_j , либо только относительные частоты n_j/n .

Наряду с частотами n_j и относительными частотами n_j/n подсчитываются *накопленные частоты* $\sum_{i=1}^j n_i$ и *накопленные относительные частоты* $\sum_{i=1}^j n_i/n$, $j = 1, 2, \dots, m$. Полученные результаты сводятся в таблицу, называемую *таблицей частот группированной выборки*.

Для любого $x \in \mathbf{R}$ обозначим через $n(x)$ число значений выборки (x_1, x_2, \dots, x_n) , удовлетворяющих неравенству $x_i < x$. *Эмпирической функцией распределения* называется функция

$$\tilde{F}_X(x) = \frac{n(x)}{n}.$$

Очевидно, что $\tilde{F}_X(x) = 0$ при $x \leq z_1$ и $\tilde{F}_X(x) = 1$ при $x > z_m$. Кроме того, $\tilde{F}_X(x)$ – кусочно-постоянная, непрерывная слева функция с разрывами первого рода (скачками величины n_j) в точках $x = z_j$, $j = 1, 2, \dots, m$.

Гистограммой частот называют диаграмму, составленную из прямоугольников (столбиков) с основаниями $\Delta_j = [d_{j-1}; d_j]$ и высотами $h_j = \frac{n_j}{\Delta}$, $j = 1, 2, \dots, m$. Поскольку $\sum_{j=1}^m \Delta \cdot \frac{n_j}{\Delta} = \sum_{j=1}^m n_j = n$, площадь такой диаграммы равна объёму выборки n . Кроме того, площади прямоугольников равны частотам n_j попадания элементов выборки в промежутки $\Delta_j = [d_{j-1}; d_j]$ статистического ряда.

Замечание. В некоторых случаях при построении гистограммы частот высоты столбиков выбирают равными $h_j = n_j$. Это относится, в частности, к функции hist системы Matlab.

Если генеральная совокупность X является случайной величиной непрерывного типа, то *эмпирической плотностью распределения*, соответствующей выборке (x_1, x_2, \dots, x_n) , называется кусочно-постоянная функция

$$p_{n, \Delta}(x) = \begin{cases} \frac{n_j}{n\Delta}, & x \in [d_{j-1}; d_j], j = 1, 2, \dots, m, \\ 0, & x \notin [x^{(1)}; x^{(n)}]. \end{cases}$$

Её график называют *гистограммой относительных частот*.

При построении гистограмм подсчёт частот n_j попадания выборочных значений x_i в промежутки Δ_j (особенно это относится к последнему интервалу) надо осуществлять так же, как при составлении интервального статистического ряда.

Помимо гистограммы используют и другое графическое представление статистического ряда – *полигон частот*. В случае интервальной группировки это ломаная, звенья которой соединяют середины горизонтальных отрезков, ограничивающих гистограмму частот сверху. Аналогично (при интервальной группировке) определяют *полигон относительных частот*.

Полигон частот используют также для представления статистических данных о генеральной совокупности X дискретного типа. В этом случае на оси абсцисс откладывают значения z_j , полученные в измерениях, и отмечают на координатной плоскости точки $(z_j; n_j)$, затем соседние точки соединяют отрезками прямой.

Результаты предварительной обработки выборки несут следующую информацию о законе распределения генеральной совокупности X .

1. Для любого $x \in \mathbf{R}$ последовательность случайных величин $\{\tilde{F}_X(x)\}$ при $n \rightarrow \infty$ сходится по вероятности к значению $F_X(x)$ функции распределения генеральной совокупности X в точке x (*теорема Гливенко*):

$$\tilde{F}_X(x) \xrightarrow{P} F_X(x), n \rightarrow \infty.$$

Таким образом, *эмпирическую функцию распределения $\tilde{F}_X(x)$ можно рассматривать как приближённое представление функции распределения $F_X(x)$ генеральной совокупности X* . С увеличением объёма выборки n точность и достоверность этого представления возрастают.

2. Для генеральной совокупности X дискретного типа относительные частоты n_j/n значений z_j (см. табл.1.1) при $n \rightarrow \infty$ сходятся по вероятности к вероятностям $p_j = P\{X = z_j\}$ этих значений:

$$\frac{n_j}{n} \xrightarrow{P} p_j = P\{X = z_j\}, n \rightarrow \infty.$$

Это означает, что для генеральной совокупности дискретного типа X статистический ряд (см. табл.1.1) даёт приближённое представление закона распределения (ряда распределения) X .

3. Для генеральной совокупности X непрерывного типа относительные частоты n_j/n попадания значений выборки в промежутки Δ_j (см. табл.1.2) при $n \rightarrow \infty$ сходятся по вероятности к вероятностям $p_j = P\{X \in \Delta_j\}$ попадания в эти промежутки значений генеральной совокупности X :

$$\frac{n_j}{n} \xrightarrow{P} p_j = P\{X \in \Delta_j\}, n \rightarrow \infty.$$

Таким образом, в случае генеральной совокупности X непрерывного типа статистический ряд (см. табл.1.2) приближённо представляет закон распределения X , так как позволяет приближённо найти вероятность попадания значений случайной величины X в любое множество, являющееся объединением каких-либо промежутков Δ_j .

4. В случае генеральной совокупности X непрерывного типа эмпирическая плотность распределения $p_{n,\Delta}(x)$ при $n \rightarrow \infty$ и $\Delta \rightarrow 0$ сходится по вероятности к функции плотности распределения $f_X(x)$ в любой точке $x \in \mathbf{R}$:

$$p_{n,\Delta}(x) \xrightarrow{P} f_X(x), n \rightarrow \infty, \Delta \rightarrow 0.$$

Это означает, что *гистограмма относительных частот $p_{n,\Delta}(x)$ даёт приближённое представление функции плотности распределения вероятности $f_X(x)$ генеральной совокупности X* .

Сказанное выше показывает, что предварительная обработка выборки, несмотря на её относительно небольшую трудоёмкость, даёт достаточно полное и наглядное представление о законе распределения генеральной совокупности X .

Пример 1.1. В результате наблюдений случайной величины X получена выборка объёмом $n = 50$: 9,04; 9,83; 12,70; 8,47; 8,21; 12,73; 10,05; 10,95; 12,42; 5,22; 9,30; 7,38; 6,39; 9,02; 5,17; 9,48; 8,84; 7,85; 8,37; 6,98; 10,48; 11,44; 7,63; 7,58; 6,04; 7,49; 3,00; 9,01; 10,22; 9,73; 9,18; 17,00; 10,99; 7,76; 11,87; 10,92; 7,23; 11,37; 7,64; 9,58; 9,46; 11,77; 8,76; 5,89; 5,39; 7,67; 8,93; 8,57; 8,74; 13,80.

Построить вариационный ряд и найти размах выборки w . Провести группировку (семь интервалов) и статистическую обработку результатов наблюдений, включающую построение (для группированной выборки):

- таблицы частот;
- полигона и гистограммы частот;
- гистограммы относительных частот;
- эмпирической функции распределения.

Решение. Вариационный ряд выборки имеет вид:

3,00; 5,17; 5,22 5,39; 5,89; 6,04; 6,39; 6,98; 7,23; 7,38; 7,49; 7,58; 7,63; 7,64; 7,67; 7,76; 7,85; 8,21; 8,37; 8,47; 8,57; 8,74; 8,76; 8,84; 8,93; 9,01; 9,02; 9,04; 9,18; 9,30; 9,46; 9,48; 9,58; 9,73; 9,83; 10,05; 10,22; 10,48; 10,92; 10,95; 10,99; 11,37; 11,44; 11,77; 11,87; 12,42; 12,70; 12,73; 13,80; 17,00.

Размах выборки определяется по соотношению $w = x^{(n)} - x^{(1)}$, тогда $w = 17,00 - 3,00 = 14$. Разобьём отрезок $[3; 17]$ на семь равных частей (разрядов) длины $\Delta = 2$. Результаты группировки данных представлены в табл.1.3.

Таблица 1.3

Таблица частот одномерной группированной выборки

Номер разряда i	Разряд	Середина разряда z_i	Частота n_i	Относительная частота $\frac{n_i}{n}$	Высота гистограммы $\frac{n_i}{nb}$	Накопленная частота $\sum_{j=1}^i n_j$	Относительная накопленная частота $\frac{1}{n} \sum_{j=1}^i n_j$
1	[3; 5)	4	1	0,02	0,01	1	0,02
2	[5; 7)	6	7	0,14	0,07	8	0,16
3	[7; 9)	8	17	0,34	0,17	25	0,50
4	[9; 11)	10	16	0,32	0,16	41	0,82
5	[11; 13)	12	7	0,14	0,07	48	0,96
6	[13; 15)	14	1	0,02	0,01	49	0,98
7	[15; 17]	16	1	0,02	0,01	50	1,00

По результатам группировки строим гистограмму и полигон частот (рис.1.8,*а*), а также гистограмму относительных частот (рис.1.8,*б*). Для наглядности на гистограмму относительных частот нанесён график теоретической функции плотности вероятности $f_X(x)$.

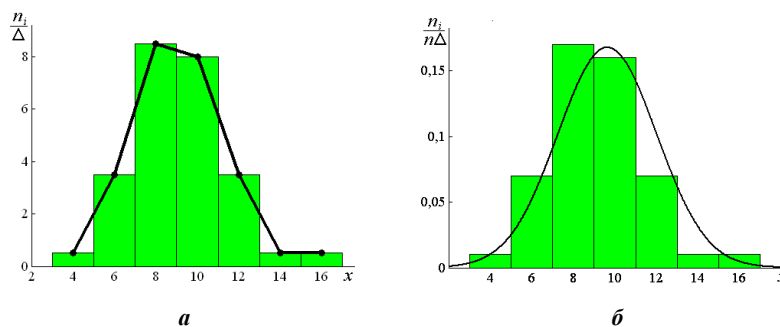


Рис.1.8. Графическое представление группированной выборки:

а – гистограмма и полигон частот;

б – гистограмма относительных частот с наложением графика теоретической функции плотности вероятности $f_X(x)$

По данным третьего и последнего столбцов табл.1.3 можно построить эмпирическую функцию распределения $\tilde{F}_X(x)$, которая имеет скачки в точках z_i , соответствующих серединам интервалов группировки. График $\tilde{F}_X(x)$ изображён на рис.1.9 (для наглядности на него нанесён график теоретической функции распределения $F_X(x)$).

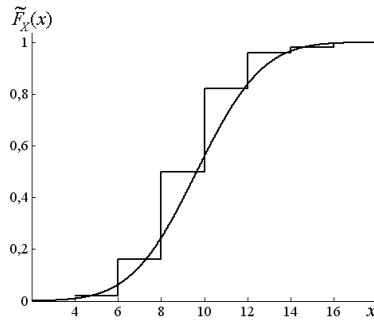


Рис.1.9. Эмпирическая функция распределения с наложением графика теоретической функции распределения $F_X(x)$

1.3. Числовые характеристики выборки

Пусть (x_1, x_2, \dots, x_n) – выборка объёмом n из генеральной совокупности с функцией распределения $F_X(x)$. Рассмотрим *выборочное распределение*, т.е. распределение дискретной случайной величины, принимающей значения x_1, x_2, \dots, x_n с вероятностями, равными $1/n$. Числовые характеристики этого выборочного распределения называются *выборочными (эмпирическими)*.

Замечание. Выборочные числовые характеристики являются характеристиками данной выборки, но не являются характеристиками распределения генеральной совокупности. Поэтому в дальнейшем изложении при обозначении эмпирических числовых характеристик будет использоваться знак « \sim ».

1.3.1. Выборочные числовые характеристики негруппированной выборки

Математическое ожидание случайной величины дискретного типа определяется по формуле $m_X = \sum_{i=1}^n x_i p_i$.

Так как для выборочного распределения $p_i = 1/n$, то *выборочное математическое ожидание (среднее)* определяется по соотношению

$$\tilde{m}_X = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Аналогично определяется *выборочная дисперсия*

$$\tilde{D}_X = \sum_{i=1}^n (x_i - \bar{x})^2 p_i = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right).$$

Выборочные начальные и центральные моменты порядка s ($s = 1, 2, \dots$) определяются по следующим формулам:

$$\tilde{\alpha}_s = \frac{1}{n} \sum_{i=1}^n x_i^s, \quad \tilde{\mu}_s = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^s.$$

Выборочной модой \tilde{d}_X унимодального (одновершинного) распределения называется элемент выборки, встречающийся с наибольшей частотой.

Выборочной медианой называется число \tilde{h}_X , которое делит *вариационный ряд* на две части, содержащие равное число элементов. Если объём выборки n – число нечётное ($n = 2l + 1$), то $\tilde{h}_X = x^{(l+1)}$; если $n = 2l$, то $\tilde{h}_X = \frac{1}{2} (x^{(l)} + x^{(l+1)})$.

1.3.2. Выборочные числовые характеристики группированной выборки

Пусть данные представлены в группированном виде, когда отрезок $[x^{(1)}; x^{(n)}]$, содержащий выборку (x_1, x_2, \dots, x_n) , разбит на m промежутков Δ_j , $j = 1, 2, \dots, m$. Обозначим через z_1, z_2, \dots, z_m середины интервалов группировки, а через n_j – частоты попадания выборочных значений в соответствующие промежутки Δ_j , $j = 1, 2, \dots, m$. Выборочное среднее определяется по соотношению

$$\bar{x} = \frac{1}{n} \sum_{j=1}^m z_j n_j.$$

Выборочная дисперсия равна

$$\tilde{D}_X = \frac{1}{n} \sum_{j=1}^m (z_j - \bar{x})^2 n_j = \frac{1}{n} \left(\sum_{j=1}^m z_j^2 n_j - n \bar{x}^2 \right).$$

Выборочные начальные и центральные моменты порядка s ($s = 1, 2, \dots$) определяются по следующим формулам:

$$\tilde{\alpha}_s = \frac{1}{n} \sum_{j=1}^m z_j^s n_j, \quad \tilde{\mu}_s = \frac{1}{n} \sum_{j=1}^m (z_j - \bar{x})^s n_j.$$

Выборочная мода вычисляется по формуле

$$\tilde{d}_X = a_d + \left(\frac{n_d - n_{d-1}}{2n_d - n_{d-1} - n_{d+1}} \right) \cdot \Delta,$$

где a_d – нижняя граница интервала, содержащего наибольшее число элементов выборки; n_d – число элементов выборки в этом интервале; n_{d-1} , n_{d+1} – количества элементов выборки в соседних интервалах; Δ – длина интервала группировки.

Оценка медианы по группированной выборке может быть вычислена по формуле

$$\tilde{h}_X = a_h + \left(\frac{\frac{n}{2} - (n_1 + n_2 + \dots + n_{h-1})}{n_h} \right) \cdot \Delta,$$

где a_h – нижняя граница интервала, которому принадлежит середина вариационного ряда; n_h – число элементов выборки в этом интервале; $n_1 + n_2 + \dots + n_{h-1}$ – число элементов выборки в разрядах, лежащих слева от интервала, содержащего середину вариационного ряда; Δ – длина интервала группировки.

1.4. Оценки параметров распределения

Пусть неизвестная функция распределения генеральной совокупности зависит от некоторого параметра θ . Нужно по наблюдениям оценить этот параметр. Для построения оценок используются *статистики*, т.е. функции $\tilde{\theta}_n = \tilde{\theta}_n(X_1, X_2, \dots, X_n)$ случайного выборочного вектора (X_1, X_2, \dots, X_n) . Статистика $\tilde{\theta}_n$ рассматривается как приближённое значение параметра θ . Например, такими статистиками могут быть *выборочное среднее*, *выборочная дисперсия* и т.п.

Замечание. Как правило, для оценки параметра θ можно использовать несколько статистик, получая при этом различные значения оценок.

Как измерить «близость» оценки $\tilde{\theta}_n$ к истинному значению θ или как определить качество оценки?

Качество оценки определяется не по одной конкретной выборке, а по всему мыслимому набору конкретных выборок, т.е. по случайному выборочному вектору (X_1, X_2, \dots, X_n) . Поэтому для установления качества полученных оценок следует во всех формулах раздела 1.3 заменить конкретные выборочные значения x_i на случайные величины X_i . Например, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $\tilde{D}_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ и т.п.

Качество точечной оценки $\tilde{\theta}_n$ параметра θ устанавливают, проверяя, выполняются ли следующие три свойства:

1) *несмещённость*, т.е. $M[\tilde{\theta}_n] = \theta$. Это свойство желательно, но не обязательно. Часто полученная оценка бывает смещённой, но её можно поправить так, что она станет несмещённой. Иногда оценка бывает смещённой, но *асимптотически несмещённой*, т.е. $\lim_{n \rightarrow \infty} M[\tilde{\theta}_n] = \theta$;

2) *состоятельность*, т.е. $\lim_{n \rightarrow \infty} P\{|\tilde{\theta}_n - \theta| < \varepsilon\} = 1$. Это свойство является обязательным, несостоятельные оценки не используются;

3) *эффективность*. Пусть $\tilde{\theta}_n^{(1)}$ и $\tilde{\theta}_n^{(2)}$ – две различные оценки параметра θ . Если $\tilde{\theta}_n^{(1)}$ и $\tilde{\theta}_n^{(2)}$ – несмещённые, то при $D[\tilde{\theta}_n^{(1)}] > D[\tilde{\theta}_n^{(2)}]$ считается, что оценка $\tilde{\theta}_n^{(2)}$ более эффективная, чем оценка $\tilde{\theta}_n^{(1)}$. Если $\tilde{\theta}_n^{(1)}$ и $\tilde{\theta}_n^{(2)}$ – смещённые, то сравнение оценок по эффективности осуществляется с помощью неравенства $\Delta(\tilde{\theta}_n^{(1)}) > \Delta(\tilde{\theta}_n^{(2)})$, где $\Delta(\tilde{\theta}_n) = \sqrt{M[\tilde{\theta}_n - \theta]^2 + D[\tilde{\theta}_n]}$ – средний квадрат отклонения оценки.

Например, выборочное среднее $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ является *несмещённой* и *состоятельной* оценкой для математического ожидания, а в случае, когда генеральная совокупность имеет нормальное распределение, \bar{X} является и наиболее *эффективной* из оценок.

Выборочная дисперсия $\tilde{D}_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ является *смещённой* оценкой дисперсии σ^2 генеральной совокупности с конечным математическим ожиданием m . Эта оценка занижает в среднем истинное значение дисперсии на величину σ^2/n . Очевидно, что это смещение «сходит на нет» при $n \rightarrow \infty$, т.е. оценка \tilde{D}_X – *асимптотически несмещённая*. На практике всегда лучше использовать в качестве *несмещённой* и *состоятельной* оценки дисперсии σ^2 «исправленную» *выборочную дисперсию*, которая имеет вид:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} \cdot \tilde{D}_X.$$

К сожалению, при оценке параметров не всегда оказывается возможным одновременное выполнение требований несмещённости, эффективности и состоятельности точечной оценки.

В ряде задач для параметра θ требуется не только найти подходящую точечную оценку $\tilde{\theta}_n$, но и указать, к каким ошибкам может привести замена параметра θ его оценкой $\tilde{\theta}_n$. Другими словами, требуется оценить точность и надёжность оценки. Такого рода задачи особенно актуальны при малом числе наблюдений, когда точечная оценка $\tilde{\theta}_n$ в значительной мере случайна и замена θ на $\tilde{\theta}_n$ может привести к серьёзным ошибкам. Для определения точности оценки $\tilde{\theta}_n$ в математической статистике пользуются доверительными интервалами, а для определения надёжности – доверительными вероятностями.

Доверительным интервалом для параметра θ называется интервал $(\theta_1; \theta_2)$, содержащий его истинное значение с заданной вероятностью $p = 1 - \alpha$. Таким образом,

$$P\{\theta_1 < \theta < \theta_2\} = 1 - \alpha.$$

Число $p = 1 - \alpha$ называется *доверительной вероятностью*, а значение α – *уровнем значимости*.

Чем меньше для выбранной доверительной вероятности длина интервала $|\theta_2 - \theta_1|$, тем точнее оценка неизвестного параметра θ , и наоборот, если этот интервал велик, то оценка, произведённая с его помощью, малоприменна для практики.

Замечание. Нижняя θ_1 и верхняя θ_2 границы доверительного интервала определяются по результатам наблюдений и, следовательно, являются случайными величинами. Поэтому говорят, что доверительный интервал «накрывает» оцениваемый параметр с вероятностью $p = 1 - \alpha$ (или в $100 \cdot (1 - \alpha)$ % случаев).

Выбор доверительной вероятности определяется конкретными условиями задачи. Обычно используются следующие значения $1 - \alpha$:

$$0,90; 0,95; 0,99.$$

Часто применяют *односторонние доверительные интервалы*, границы которых определяются из условий: $P\{\theta < \theta_2\} = 1 - \alpha$ (левосторонний), $P\{\theta_1 < \theta\} = 1 - \alpha$ (правосторонний).

Ниже приведены доверительные интервалы при оценке параметров по случайным выборкам из нормально распределённых генеральных совокупностей.

1. Доверительный интервал для математического ожидания m (дисперсия σ^2 неизвестна):

$$\bar{X} - \frac{S}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1) < m < \bar{X} + \frac{S}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1).$$

2. Доверительный интервал для дисперсии σ^2 (математическое ожидание m неизвестно):

$$\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}.$$

Пример 1.2. Используя выборку примера 1.1, вычислить для негруппированных и группированных (см. табл.1.3) данных эмпирические числовые характеристики: медиану, среднее, дисперсию (смещённую и несмещённую оценки), а также для группированных данных получить оценку моды. Считая, что выборка взята из нормально распределённой генеральной совокупности с неизвестными m и σ^2 , найти доверительные интервалы для математического ожидания и дисперсии при доверительной вероятности, равной 0,95.

Решение. Оценка медианы по негруппированным данным:

$$\tilde{h}_X = \frac{1}{2}(x^{(25)} + x^{(26)}) = \frac{8,93 + 9,01}{2} = 8,97.$$

Для получения оценок среднего и дисперсии по негруппированным данным воспользуемся средствами Matlab:

$$\bar{x} = \frac{1}{50} \sum_{i=1}^{50} x_i = 9,0708, \quad \tilde{D}_X = \frac{1}{50} \left(\sum_{i=1}^{50} x_i^2 - 50 \cdot \bar{x}^2 \right) \approx 5,9858,$$

$$s^2 = \frac{1}{49} \left(\sum_{i=1}^{50} x_i^2 - 50 \cdot \bar{x}^2 \right) \approx 6,1080.$$

Оценки медианы, среднего, дисперсии и моды по группированным данным:

$$\tilde{h}_X = a_h + \left(\frac{\frac{n}{2} - (n_1 + n_2 + \dots + n_{h-1})}{n_h} \right) \cdot \Delta = 7 + \frac{25 - (1 + 7)}{17} \cdot 2 = 9,$$

$$\bar{x} = \frac{1}{50} \cdot (4 \cdot 1 + 6 \cdot 7 + 8 \cdot 17 + 10 \cdot 16 + 12 \cdot 7 + 14 \cdot 1 + 16 \cdot 1) = 9,12,$$

$$\tilde{D}_X = \frac{1}{50} \cdot (4^2 \cdot 1 + 6^2 \cdot 7 + 8^2 \cdot 17 + 10^2 \cdot 16 + 12^2 \cdot 7 + 14^2 \cdot 1 + 16^2 \cdot 1 - 50 \cdot 9,12^2) = 5,1456,$$

$$s^2 = \frac{50}{50-1} \cdot \tilde{D}_X \approx 5,251,$$

$$\tilde{d}_X = a_d + \left(\frac{n_d - n_{d-1}}{2n_d - n_{d-1} - n_{d+1}} \right) \cdot \Delta = 7 + \frac{17-7}{2 \cdot 17 - 7 - 16} \cdot 2 \approx 8,818.$$

Доверительный интервал для дисперсии σ^2 найдём из условия

$$\bar{x} - \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1) < m < \bar{x} + \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1).$$

Задавшись доверительной вероятностью $1 - \alpha = 0,95$ с учётом $t_{0,975}(49) = 2,0096$, получим:

$8,369 < m < 9,773$ – для негруппированной выборки;

$8,469 < m < 9,771$ – для группированной выборки.

Доверительный интервал для математического ожидания m найдём из условия

$$\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} < \sigma^2 < \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}$$

Задавшись доверительной вероятностью $1 - \alpha = 0,95$ с учётом $\chi_{0,975}^2(49) = 70,2224$, $\chi_{0,025}^2(49) = 31,5549$, получим:

$$4,262 < \sigma^2 < 9,485 \text{ – для негруппированной выборки;}$$

$$3,664 < \sigma^2 < 8,154 \text{ – для группированной выборки.}$$

1.5. Проверка статистических гипотез

Во многих случаях результаты наблюдений используются для проверки гипотез относительно тех или иных свойств распределения генеральной совокупности. Пусть X – наблюдаемая дискретная или непрерывная случайная величина. *Статистической гипотезой H* называется предположение относительно параметров или вида распределения случайной величины X . Статистическая гипотеза H называется *простой*, если она однозначно определяет распределение случайной величины X ; в противном случае гипотеза H называется *сложной*. Например, гипотеза о том, что $X \sim N(0; 1)$ – простая, а гипотеза о том, что $X \sim N(m; 1)$, где $m \in [0; 3]$, – сложная.

Если распределение случайной величины X известно и по выборке наблюдений необходимо проверить предположения о значении параметров этого распределения, то такие гипотезы называются *параметрическими*. Если же по выборке наблюдений необходимо проверить гипотезу о том, что она извлечена из генеральной совокупности X с функцией распределения $F_X(x)$, то такие гипотезы называются гипотезами *о законе распределения*. В математической статистике проверяемую гипотезу называют нулевой гипотезой и обозначают H_0 . Наряду с гипотезой H_0 рассматривают одну из альтернативных (конкурирующих) гипотез H_1 . Например, если проверяется гипотеза о равенстве параметра θ некоторому заданному значению θ_0 , т.е. $H_0 : \theta = \theta_0$, то в качестве альтернативной гипотезы можно рассматривать одну из следующих гипотез:

$$H_1^{(1)} : \theta \neq \theta_0; H_1^{(2)} : \theta > \theta_0; H_1^{(3)} : \theta < \theta_0; H_1^{(4)} : \theta = \theta_1, \text{ где } \theta_1 \neq \theta_0.$$

Выбор альтернативной гипотезы определяется конкретными условиями задачи.

Практическое применение математической статистики состоит в проверке фактического соответствия реальных результатов экспериментов предполагаемой гипотезе H_0 . С этой целью выбирается правило проверки гипотезы, называемое *критерием*, позволяющее по результатам наблюдений принимать или отвергать данную гипотезу. Поскольку решение принимается на основе выборки наблюдений случайной величины X , то необходимо выбрать подходящую статистику Z критерия. При проверке простой параметрической гипотезы $H_0 : \theta = \theta_0$ в качестве статистики критерия обычно выбирают ту же статистику, что и для оценки параметра θ , т.е. $\tilde{\theta}$.

Проверка статистической гипотезы основывается на принципе, в соответствии с которым маловероятные события считаются невозможными, а события, имеющие большую вероятность, считаются достоверными. Этот принцип реализуется следующим образом. Перед анализом выборки фиксируется некоторая малая вероятность α , называемая *уровнем значимости*. Критерии, основанные на использовании заранее заданного уровня значимости, называются *критериями значимости*. Пусть W – множество значений статистики Z . Тогда W разбивается на два непересекающихся подмножества: U и V , причём $P\{Z \in V | H_0\} = \alpha$. Обозначим z_b выборочное значение статистики Z , вычисленное по выборке наблюдений. Критерий формулируется так: если $z_b \in U$, то считается, что гипотеза H_0 подтверждается эмпирическими данными; если же $z_b \in V$, то утверждается, что гипотеза H_0 не согласуется с результатами наблюдений, т.е. H_0 отвергается и принимается H_1 . Множество U называется *областью принятия гипотезы*, множество V – *критической областью*. Уровень значимости α определяет «размер» критической области V . Расположение V на множестве значений W статистики Z зависит от альтернативной гипотезы H_1 . На рис.1.10 показано расположение критической области V для различных альтернативных гипотез H_1 ($f(z | H_0)$ – плотность распределения статистики Z критерия при условии, что верна гипотеза H_0).

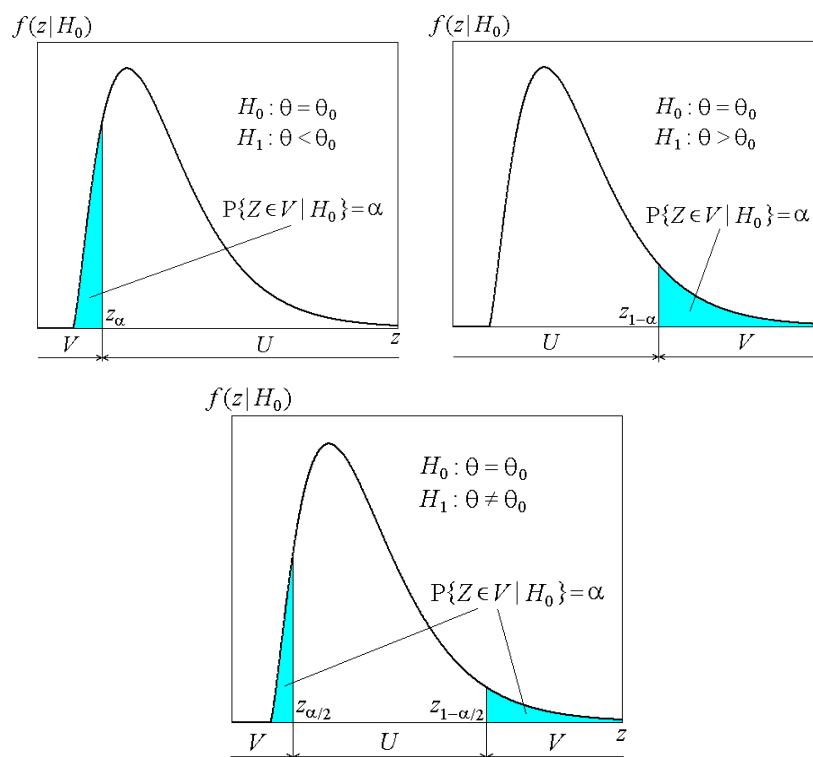


Рис.1.10. Расположение критической области V для различных альтернативных гипотез H_1

Применение правила проверки гипотезы сопряжено с ошибками двух родов: гипотеза отвергается, когда она верна (*ошибка первого рода*); гипотеза принимается, когда она неверна (*ошибка второго рода*). При выборе критерия проверки гипотезы желательно добиваться минимальных значений ошибок обоих родов. Однако в большинстве практически важных ситуаций невозможно построение критериев со сколь угодно малыми ошибками первого и второго родов. Правило проверки гипотез имеет статистический смысл, т.е. при многократном применении определённого критерия проценты числа неверных решений выражаются вероятностями ошибок первого и второго родов.

Согласование экспериментальных данных с предполагаемой гипотезой не означает, что невозможно согласование этих же данных с другой гипотезой. При применении статистических критериев на основании наблюдений нельзя доказать ту или иную гипотезу, можно лишь утверждать, что результаты наблюдений не противоречат принятой гипотезе. Таким образом, выводы, принимаемые на основании статистических данных, формулируются в следующем виде: экспериментальные данные согласуются с данной гипотезой (противоречат ей).

Итак, проверка параметрической статистической гипотезы при помощи критерия значимости может быть разбита на следующие этапы:

- 1) сформулировать проверяемую H_0 и альтернативную H_1 гипотезы;
- 2) установить уровень значимости α (вероятность ошибки первого рода);
- 3) выбрать статистику Z критерия для проверки гипотезы H_0 ;
- 4) определить распределение этой статистики при условии, что верна гипотеза H_0 ;
- 5) в зависимости от формулировки альтернативной гипотезы определить критическую область V ;
- 6) получить выборку наблюдений и вычислить значение z_B выбранной статистики Z критерия;
- 7) принять статистическое решение: если $z_B \in V$, то отклонить гипотезу как несогласующуюся с результатами наблюдений; если $z_B \in U$, то считать, что гипотеза H_0 не противоречит результатам наблюдений (выборке).

Проверка статистических гипотез может быть проведена на основе доверительных интервалов [3]. Для всех параметрических гипотез о нормальной генеральной совокупности область принятия гипотезы $H_0: \theta = \theta_0$ на уровне значимости α совпадает с доверительными интервалами для параметра θ при доверительной вероятности $1 - \alpha$. Двустороннему критерию значимости $H_1: \theta \neq \theta_0$ соответствует двусторонний доверительный интервал. Гипотеза H_0 согласуется с результатами наблюдений, если значение θ_0 накрывается соответствующим доверительным интервалом, в противном случае H_0 отклоняется.

Если проверяется гипотеза $H_0 : \theta_1 = \theta_2$, то рассматривается доверительный интервал для разности $\theta_1 - \theta_2$. Гипотеза H_0 согласуется с результатами наблюдений, если доверительный интервал для разности параметров $\theta_1 - \theta_2$ покрывает нулевое значение. Исключение составляет проверка гипотезы о равенстве дисперсий двух независимых нормальных генеральных совокупностей $H_0 : \sigma_1^2 = \sigma_2^2$. Поскольку доверительный интервал строится для отношения дисперсий, то в этом случае гипотеза H_0 согласуется с результатами наблюдений, если доверительный интервал покрывает значение, равное единице.

Пример 1.3. Для негруппированных данных примера 1.1 проверить с помощью построенных в примере 1.2 доверительных интервалов гипотезы $H_0^{(1)} : m_X = M_0$ и $H_0^{(2)} : D_X = A_0$, где $M_0 = \bar{x} + \sqrt{s}$, $A_0 = \sqrt{2} \cdot s^2$, и записать статистическое решение (принять или отклонить нулевую гипотезу).

Решение. Вычислим M_0 и A_0 :

$$M_0 \approx 9,0708 + \sqrt[4]{6,1080} \approx 10,643, \quad A_0 \approx 6,1080 \cdot \sqrt{2} \approx 8,638.$$

Поскольку вычисленное значение M_0 не покрывается доверительным интервалом $(8,369; 9,773)$, то гипотеза $H_0^{(1)} : m_X = M_0$ отклоняется.

Поскольку вычисленное значение A_0 покрывается доверительным интервалом $(4,262; 9,485)$, то гипотеза $H_0^{(2)} : D_X = A_0$ согласуется с результатами наблюдений (принимается).

1.5.1. Проверка гипотезы о равенстве дисперсий двух нормальных генеральных совокупностей при неизвестных средних

Пусть X_1 и X_2 – нормальные совокупности с дисперсиями σ_1^2 , σ_2^2 и неизвестными математическими ожиданиями m_1 , m_2 соответственно. Из этих совокупностей взяты две случайные независимые выборки объёмами n_1 и n_2 с параметрами \bar{x}_1 , s_1^2 и \bar{x}_2 , s_2^2 . На уровне значимости α требуется проверить нулевую гипотезу $H_0 : \sigma_1^2 = \sigma_2^2$ при альтернативной гипотезе $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Известно, что статистика $Z = \frac{S_1^2}{S_2^2}$ при $S_1^2 > S_2^2$ имеет в этом случае распределение Фишера с $n_1 - 1$ и

$n_2 - 1$ степенями свободы.

Критерий проверки гипотезы заключается в следующем:

– вычисляется выборочное значение статистики Z :

$$z_B = \frac{s_1^2}{s_2^2} \quad (\text{при } s_1^2 > s_2^2);$$

– если $F_{\frac{\alpha}{2}}(n_1 - 1; n_2 - 1) < z_B < F_{1 - \frac{\alpha}{2}}(n_1 - 1; n_2 - 1)$, то нет оснований отвергать гипотезу

$H_0 : \sigma_1^2 = \sigma_2^2$; иначе H_0 отклоняется.

Замечание. Если $S_1^2 < S_2^2$, то рассматривается статистика

$$Z = \frac{S_2^2}{S_1^2} \sim \mathbf{F}(n_2 - 1; n_1 - 1).$$

1.5.2. Проверка гипотезы о равенстве средних двух нормальных генеральных совокупностей при неизвестных дисперсиях

1. Пусть X_1 и X_2 – нормальные совокупности с равными, но неизвестными дисперсиями $\sigma_1^2 = \sigma_2^2 = \sigma^2$ и математическими ожиданиями m_1 и m_2 . Из этих совокупностей взяты две случайные независимые выборки объёмами n_1 и n_2 с параметрами \bar{x}_1 , s_1^2 и \bar{x}_2 , s_2^2 . На уровне значимости α требуется проверить нулевую гипотезу $H_0 : m_1 = m_2$ при альтернативной гипотезе $H_1 : m_1 \neq m_2$.

Известно, что статистика $Z = \frac{\bar{X}_1 - \bar{X}_2}{S \cdot \sqrt{1/n_1 + 1/n_2}}$, где $S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$, в этом случае имеет

распределение Стьюдента с $n_1 + n_2 - 2$ степенями свободы.

Критерий проверки гипотезы заключается в следующем:

– вычисляется выборочное значение статистики Z :

$$z_B = \frac{\bar{x}_1 - \bar{x}_2}{s \cdot \sqrt{1/n_1 + 1/n_2}}, \text{ где } s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}};$$

– если $|z_B| < t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2)$, то нет оснований отвергать гипотезу $H_0 : m_1 = m_2$; иначе H_0

отклоняется.

2. Пусть X_1 и X_2 – нормальные совокупности с неизвестными, но неравными дисперсиями $\sigma_1^2 \neq \sigma_2^2$ и математическими ожиданиями m_1 и m_2 . Из этих совокупностей взяты две случайные независимые выборки объёмами n_1 и n_2 с параметрами \bar{x}_1 , s_1^2 и \bar{x}_2 , s_2^2 . На уровне значимости α требуется проверить нулевую гипотезу $H_0 : m_1 = m_2$ при альтернативной гипотезе $H_1 : m_1 \neq m_2$.

Известно, что статистика $Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$ в этом случае имеет распределение Стьюдента с k

степенями свободы, где $k = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$.

Критерий проверки гипотезы заключается в следующем:

– вычисляется выборочное значение статистики Z :

$$z_B = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}};$$

– если $|z_B| < t_{1-\frac{\alpha}{2}}(k)$, где $k = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$, то нет оснований отвергать гипотезу

$H_0 : m_1 = m_2$; иначе H_0 отклоняется.

Замечание. Описанные алгоритмы проверки гипотез о равенстве дисперсий и средних справедливы и при отклонении распределений случайных величин X_1 и X_2 от нормального закона, но при условии, что n_1 и n_2 больше 30.

Пример 1.4. В банке в течение двух дней проводилось исследование времени обслуживания клиентов. Данные представлены в табл.1.4 (X_1 – время обслуживания клиентов в первый день, X_2 – во второй). Проверить гипотезы о равенстве дисперсий и средних при неизвестных σ_1^2 , σ_2^2 и m_1 , m_2 при уровне значимости $\alpha = 0,1$.

Таблица 1.4

Статистические данные времени обслуживания клиентов в банке

Номер интервала i	Время обслуживания, мин	Количество клиентов	
		1-й день	2-й день
1	[10; 12)	2	2
2	[12; 14)	4	4
3	[14; 16)	8	9
4	[16; 18)	12	13
5	[18; 20)	16	16
6	[20; 22)	10	8
7	[22; 24)	3	3

Решение. Очевидно, что $n_1 = n_2 = 55$. По статистическим данным получим оценки неизвестных параметров:

$$\bar{x}_1 = \frac{1}{55} \cdot (2 \cdot 11 + 4 \cdot 13 + 8 \cdot 15 + 12 \cdot 17 + 16 \cdot 19 + 10 \cdot 21 + 3 \cdot 23) \approx 17,84,$$

$$\bar{x}_2 = \frac{1}{55} \cdot (2 \cdot 11 + 4 \cdot 13 + 9 \cdot 15 + 13 \cdot 17 + 16 \cdot 19 + 8 \cdot 21 + 3 \cdot 23) \approx 17,65,$$

$$s_1^2 \approx \frac{1}{54} \cdot (2 \cdot 121 + 4 \cdot 169 + 8 \cdot 225 + 12 \cdot 289 + 16 \cdot 361 + 10 \cdot 441 + 3 \cdot 529 - 55 \cdot 17,84^2) \approx 8,55,$$

$$s_2^2 \approx \frac{1}{54} \cdot (2 \cdot 121 + 4 \cdot 169 + 9 \cdot 225 + 13 \cdot 289 + 16 \cdot 361 + 8 \cdot 441 + 3 \cdot 529 - 55 \cdot 17,65^2) \approx 8,30.$$

Проверим вначале гипотезу о равенстве дисперсий $H_0 : \sigma_1^2 = \sigma_2^2$ при альтернативной гипотезе $H_1 : \sigma_1^2 \neq \sigma_2^2$. Поскольку $s_1^2 > s_2^2$, то

$$z_B = \frac{s_1^2}{s_2^2} \approx \frac{8,55}{8,30} \approx 1,03.$$

Используя Matlab, найдём квантили распределения Фишера:

$$F_{0,05}(54; 54) = 0,6366, \quad F_{0,95}(54; 54) = 1,5709.$$

Так как $F_{0,05}(54; 54) < z_B < F_{0,95}(54; 54)$, то нет оснований отвергать гипотезу H_0 о равенстве дисперсий при уровне значимости $\alpha = 0,1$.

Для проверки гипотезы о равенстве средних $H_0 : m_1 = m_2$ при альтернативной гипотезе $H_1 : m_1 \neq m_2$ будем использовать критерий при равных неизвестных дисперсиях. Вычислим выборочное значение z_B статистики критерия:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \approx \sqrt{\frac{(55 - 1) \cdot 8,55 + (55 - 1) \cdot 8,30}{55 + 55 - 2}} \approx 2,90,$$

$$z_B = \frac{\bar{x}_1 - \bar{x}_2}{s \cdot \sqrt{1/n_1 + 1/n_2}} \approx \frac{17,84 - 17,65}{2,90 \cdot \sqrt{1/55 + 1/55}} \approx 0,34.$$

Используя Matlab, найдём квантиль распределения Стьюдента $t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2) = t_{0,95}(108) = 1,6591$.

Так как $|z_B| < t_{0,95}(108)$, то нет оснований отвергать гипотезу H_0 о равенстве средних при уровне значимости $\alpha = 0,1$.

Пример 1.5. При измерении производительности двух агрегатов получены результаты (в кг вещества за час работы), представленные в табл.1.5. Можно ли считать, что производительности агрегатов A и B одинаковы в предположении, что обе выборки получены из нормально распределённых генеральных совокупностей? Принять $\alpha = 0,1$.

Таблица 1.5

Статистические данные измерения производительности агрегатов

Номер замера	1	2	3	4	5
Агрегат A	14,1	10,1	14,7	13,7	14,0
Агрегат B	14,0	14,5	13,7	12,7	14,1

Решение. Очевидно, что $n_1 = n_2 = 5$. По статистическим данным получим оценки неизвестных параметров:

$$\bar{x}_1 = \frac{1}{5} \cdot (14,1 + 10,1 + 14,7 + 13,7 + 14,0) = 13,32,$$

$$\bar{x}_2 = \frac{1}{5} \cdot (14,0 + 14,5 + 13,7 + 12,7 + 14,1) = 13,80,$$

$$s_1^2 = \frac{1}{4} \cdot (14,1^2 + 10,1^2 + 14,7^2 + 13,7^2 + 14,0^2 - 5 \cdot 13,32^2) = 3,372,$$

$$s_2^2 = \frac{1}{4} \cdot (14,0^2 + 14,5^2 + 13,7^2 + 12,7^2 + 14,1^2 - 5 \cdot 13,80^2) = 0,46.$$

Проверим вначале гипотезу о равенстве дисперсий $H_0 : \sigma_1^2 = \sigma_2^2$ при альтернативной гипотезе $H_1 : \sigma_1^2 \neq \sigma_2^2$. Поскольку $s_1^2 > s_2^2$, то

$$z_B = \frac{s_1^2}{s_2^2} = \frac{3,372}{0,46} \approx 7,33.$$

Используя Matlab, найдём квантили распределения Фишера:

$$F_{0,05}(4; 4) = 0,1565, \quad F_{0,95}(4; 4) = 6,3882.$$

Так как $z_B > F_{0,95}(4; 4)$, то гипотеза H_0 о равенстве дисперсий отклоняется при уровне значимости $\alpha = 0,1$.

Для проверки гипотезы $H_0 : m_1 = m_2$ о равенстве средних при альтернативной гипотезе $H_1 : m_1 \neq m_2$ будем использовать критерий при неравных неизвестных дисперсиях. Вычислим выборочное значение z_B статистики критерия:

$$z_B = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{13,32 - 13,80}{\sqrt{3,372/5 + 0,46/5}} \approx -0,548.$$

Число степеней свободы в распределении Стьюдента получим по формуле

$$k = \frac{\left(\frac{s_1^2/n_1 + s_2^2/n_2}{s_1^2/n_1}\right)^2}{\frac{s_1^2/n_1}{n_1 - 1} + \frac{s_2^2/n_2}{n_2 - 1}} = \frac{(3,372/5 + 0,46/5)^2}{\frac{(3,372/5)^2}{4} + \frac{(0,46/5)^2}{4}} \approx 5.$$

Используя Matlab, найдём квантиль распределения Стьюдента $t_{1-\frac{\alpha}{2}}(k) = t_{0,95}(5) = 2,0150$. Так как

$|z_B| < t_{0,95}(5)$, то нет оснований отвергать гипотезу H_0 о равенстве средних при уровне значимости $\alpha = 0,1$.

1.5.3. Критерий согласия χ^2 (Пирсона)

Критерий χ^2 является наиболее распространённым непараметрическим критерием, используемым для проверки гипотез о виде распределения генеральной совокупности. Критерии, используемые для этих целей, называются обычно *критериями согласия*.

Итак, пусть требуется статистически проверить гипотезу H_0 о том, что данная выборка x_1, x_2, \dots, x_n извлечена из генеральной совокупности X с функцией распределения $F_X(x)$, которая точно известна. По выборке можно построить эмпирическое распределение $\tilde{F}_X(x)$ исследуемой случайной величины. Принцип сравнения эмпирического $\tilde{F}_X(x)$ и теоретического $F_X(x)$ распределений с помощью критерия согласия χ^2 состоит в следующем.

1. Пусть проверяемый закон распределения генеральной совокупности $F_X(x) = F_X(x; \theta_1, \theta_2, \dots, \theta_k, \delta_1, \delta_2, \dots, \delta_m)$ зависит от $k + m$ параметров $\theta_1, \theta_2, \dots, \theta_k, \delta_1, \delta_2, \dots, \delta_m$, причём параметры $\theta_1, \theta_2, \dots, \theta_k$ неизвестны. Тогда по выборке x_1, x_2, \dots, x_n (например, методом максимального правдоподобия [3]) для неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$ находят их оценки $\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_k$.

2. Если X – случайная величина непрерывного типа, то всё множество возможных значений X разбивают на r непересекающихся интервалов $\Delta_1, \Delta_2, \dots, \Delta_r$. Затем для каждого интервала Δ_i , $i = 1, 2, \dots, r$, подсчитывают n_i – число выборочных значений, фактически в него попавших.

Если X – случайная величина дискретного типа, то определяют частоты n_i , $i = 1, 2, \dots, r$, с которыми каждое значение (или группа значений) встречается в выборке.

Очевидно, что в обоих случаях $\sum_{i=1}^r n_i = n$, где n – объём выборки.

3. Если X – случайная величина непрерывного типа, то вычисляют p_i – вероятность того, что случайная величина X с законом распределения $\tilde{F}_X(x; \tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_k, \delta_1, \delta_2, \dots, \delta_m)$ попадает в интервал Δ_i , $i = 1, 2, \dots, r$.

Если X – случайная величина дискретного типа, то определяют вероятность p_i , $i = 1, 2, \dots, r$, с которой случайная величина X принимает каждое значение (или вероятность появления группы значений).

Очевидно, что в обоих случаях $\sum_{i=1}^r p_i = 1$.

4. Вычисляют выборочное значение статистики

$$\chi_{\text{в}}^2 = \sum_{i=1}^r \frac{(n_i - n'_i)^2}{n'_i},$$

где $n'_i = np_i$ – так называемые теоретические частоты.

Замечание. В литературе по математической статистике [4–6] показано, что при больших n статистика $\chi_{\text{в}}^2$ имеет распределение, близкое к распределению χ^2 с $r - k - 1$ степенями свободы (независимо от вида распределения X), где r – число интервалов, а k – число рассчитанных по выборке неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$.

5. Принимают статистическое решение: если $\chi_{\text{в}}^2 < \chi_{1-\alpha}^2(r - k - 1)$, где $\chi_{1-\alpha}^2(r - k - 1)$ – квантиль распределения χ^2 с $r - k - 1$ степенями свободы порядка $1 - \alpha$, то гипотеза H_0 о предполагаемом законе распределения не противоречит результатам наблюдений (выборке) при уровне значимости α ; в противном случае гипотеза H_0 отклоняется.

Замечания. 1. Необходимым условием применения критерия χ^2 является наличие в каждом из интервалов группировки по крайней мере пяти наблюдений. Поэтому все интервалы, для которых частоты $n_i < 5$, следует объединить с соседними интервалами. При этом соответствующие им частоты n_i и теоретические частоты $n'_i = np_i$ также надо сложить.

2. Если производилось объединение частот, то при определении числа степеней свободы по выражению $r - k - 1$ следует в качестве r принять число групп выборки, оставшихся после проведения указанной процедуры.

3. При проверке гипотезы о законе распределения контролируется лишь ошибка первого рода.

Пример 1.6 (Проверка гипотезы о нормальном распределении генеральной совокупности). Используя критерий χ^2 , требуется на уровне значимости $\alpha = 0,05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с эмпирическим распределением выборки объёмом 50, рассмотренной в примере 1.1.

Решение. Воспользуемся уже произведённой группировкой данных (см. табл.1.3). Для выполнения условия $n_j \geq 5$ объединим разряды с номерами 1, 2 и 5, 6, 7.

С помощью таблицы значений функции Лапласа вычислим вероятности p_i попадания случайной величины в каждый из интервалов выбранного разбиения, предположив, что она имеет нормальное распределение. В качестве параметров m и σ возьмём их оценки $\bar{x} = 9,071$ и $s = \sqrt{5,641} \approx 2,375$:

$$p_1 = P\{X < 7\} \approx \Phi\left(\frac{7 - 9,071}{2,375}\right) = \Phi(-0,87) = 1 - \Phi(0,87) = 0,1922,$$

$$p_2 = P\{7 \leq X < 9\} \approx \Phi\left(\frac{9 - 9,071}{2,375}\right) - \Phi\left(\frac{7 - 9,071}{2,375}\right) = 0,2958,$$

$$p_3 = P\{9 \leq X < 11\} \approx \Phi\left(\frac{11 - 9,071}{2,375}\right) - \Phi\left(\frac{9 - 9,071}{2,375}\right) = 0,3030,$$

$$p_4 = P\{X \geq 11\} = 1 - P\{X < 11\} \approx 1 - \Phi\left(\frac{11 - 9,071}{2,375}\right) = 0,2090.$$

Найдём теоретические частоты $n'_i = np_i = 50p_i$:

$$n'_1 \approx 9,61, \quad n'_2 \approx 14,79, \quad n'_3 \approx 15,15, \quad n'_4 \approx 10,45.$$

Составим расчётную табл.1.6.

Таблица 1.6

Данные для подсчёта статистики χ^2

Номер интервала j	Интервал	n_j	n'_j	$\frac{(n_j - n'_j)^2}{n'_j}$
1	$(-\infty; 7)$	8	9,61	0,270
2	$[7; 9)$	17	14,79	0,330
3	$[9; 11)$	16	15,15	0,048
4	$[11; +\infty)$	9	10,45	0,201

Вычислим значение статистики:

$$\chi^2_{\text{в}} = \sum_{j=1}^4 \frac{(n_j - n'_j)^2}{n'_j} = 0,270 + 0,330 + 0,048 + 0,201 = 0,849.$$

Используя Matlab, найдём квантиль $\chi^2_{1-\alpha}(4 - 2 - 1) = \chi^2_{0,95}(1) = 3,8415$.

Поскольку $\chi^2_{\text{в}} < \chi^2_{0,95}(1)$, то нет оснований отвергать гипотезу о нормальном распределении генеральной совокупности.

Пример 1.7 (Проверка гипотезы об экспоненциальном распределении генеральной совокупности). В результате испытаний 200 элементов на продолжительность работы получено эмпирическое распределение, приведённое в табл.1.7 (во втором столбце указаны интервалы времени в часах, в третьем столбце – частоты). Требуется на уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что время работы элементов распределено по экспоненциальному закону.

Таблица 1.7

Таблица частот группированной выборки

Номер разряда i	$[x_i; x_{i+1})$	n_i
1	$[0; 5)$	133
2	$[5; 10)$	45
3	$[10; 15)$	15
4	$[15; 20)$	4
5	$[20; 25)$	2
6	$[25; 30)$	1

Решение. Найдём среднее время работы всех элементов:

$$\bar{x} = \frac{1}{200} \cdot (133 \cdot 2,5 + 45 \cdot 7,5 + 15 \cdot 12,5 + 4 \cdot 17,5 + 2 \cdot 22,5 + 1 \cdot 27,5) = 5.$$

Тогда $\tilde{\lambda} = \frac{1}{\bar{x}} = 0,2$.

Таким образом, $\tilde{f}_X(x) = 0,2 \cdot e^{-0,2x}$ ($x > 0$).

Найдём вероятности p_i попадания случайной величины в каждый из интервалов:

$$p_i = P\{x_i \leq X < x_{i+1}\} = \int_{x_i}^{x_{i+1}} \tilde{f}_X(x) dx = e^{-\tilde{\lambda}x_i} - e^{-\tilde{\lambda}x_{i+1}}.$$

Тогда

$$p_1 \approx 0,6321, \quad p_2 \approx 0,2326, \quad p_3 \approx 0,0855, \\ p_4 \approx 0,0315, \quad p_5 \approx 0,0116, \quad p_6 \approx 0,0042.$$

Найдём теоретические частоты $n'_i = np_i = 200p_i$:

$$n'_1 \approx 126,42, n'_2 \approx 46,52, n'_3 \approx 17,10,$$

$$n'_4 \approx 6,30, n'_5 \approx 2,32, n'_6 \approx 0,84.$$

Объединив малочисленные частоты ($4 + 2 + 1 = 7$) и соответствующие им теоретические частоты ($6,30 + 2,32 + 0,84 = 9,46$), получим расчётную табл.1.8.

Таблица 1.8

Данные для подсчёта статистики χ^2

Номер интервала j	$[x_j; x_{j+1})$	n_j	$n_{\check{y}}$	$\frac{(n_j - n_{\check{y}})^2}{n_{\check{y}}}$
1	[0; 5)	133	126,42	0,3425
2	[5; 10)	45	46,52	0,0497
3	[10; 15)	15	17,10	0,2579
4	[15; 30)	7	9,46	0,6397

Вычислим значение статистики:

$$\chi^2_{\text{в}} = \sum_{j=1}^4 \frac{(n_j - n'_{j'})^2}{n'_{j'}} = 0,3425 + 0,0497 + 0,2579 + 0,6397 = 1,29.$$

Используя Matlab, найдём квантиль $\chi^2_{1-\alpha}(4-1-1) = \chi^2_{0,95}(2) = 5,9915$. Поскольку $\chi^2_{\text{в}} < \chi^2_{0,95}(2)$, то нет оснований отвергать гипотезу об экспоненциальном распределении генеральной совокупности.

Пример 1.8 (Проверка гипотезы о равномерном распределении генеральной совокупности). Произведено 200 испытаний, в результате каждого из которых некоторое событие появлялось в различные моменты времени. В итоге получено эмпирическое распределение, приведённое в табл.1.9 (во втором столбце указаны интервалы времени в минутах, в третьем столбце – частоты, т.е. числа появления наблюдаемого события в соответствующих интервалах). Требуется на уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что время появления событий распределено равномерно.

Таблица 1.9

Таблица частот группированной выборки

Номер разряда i	$[x_i; x_{i+1})$	n_i
1	[2; 4)	21
2	[4; 6)	16
3	[6; 8)	15
4	[8; 10)	26
5	[10; 12)	22
6	[12; 14)	14
7	[14; 16)	21
8	[16; 18)	22
9	[18; 20)	18
10	[20; 22)	25

Решение. Получим вначале оценки математического ожидания и дисперсии для данной группированной выборки:

$$\bar{x} = \frac{1}{200} \cdot (21 \cdot 3 + 16 \cdot 5 + 15 \cdot 7 + 26 \cdot 9 + 22 \cdot 11 + 14 \cdot 13 +$$

$$+ 21 \cdot 15 + 22 \cdot 17 + 18 \cdot 19 + 25 \cdot 21) = 12,31,$$

$$s^2 = \frac{1}{199} \cdot (21 \cdot 3^2 + 16 \cdot 5^2 + 15 \cdot 7^2 + 26 \cdot 9^2 + 22 \cdot 11^2 + 14 \cdot 13^2 +$$

$$+ 21 \cdot 15^2 + 22 \cdot 17^2 + 18 \cdot 19^2 + 25 \cdot 21^2 - 200 \cdot (12,31)^2) \approx 33,95.$$

Найдём теперь оценки \tilde{a} и \tilde{b} – концов интервала $(a; b)$. Для этого решим систему:

$$\begin{cases} \bar{x} = \frac{\tilde{a} + \tilde{b}}{2}, \\ s^2 = \frac{(\tilde{b} - \tilde{a})^2}{12}; \end{cases} \Leftrightarrow \begin{cases} \tilde{a} = \bar{x} - s\sqrt{3}, \\ \tilde{b} = \bar{x} + s\sqrt{3}. \end{cases}$$

Затем найдём плотность вероятности предполагаемого распределения:

$$\tilde{f}_X(x) = \begin{cases} \frac{1}{\tilde{b} - \tilde{a}}, & x \in (\tilde{a}; \tilde{b}), \\ 0, & x \notin (\tilde{a}; \tilde{b}). \end{cases}$$

В данном примере $\bar{x} = 12,31$, $s \approx 5,827$. Следовательно, $\tilde{a} \approx 2,217$, $\tilde{b} \approx 22,403$. Отсюда

$$\tilde{f}_X(x) = \begin{cases} \frac{1}{20,185}, & x \in (2,217; 22,403), \\ 0, & x \notin (2,217; 22,403). \end{cases}$$

Вычислим вероятности p_i попадания случайной величины в каждый из интервалов группировки:

$$p_1 = P\{\tilde{a} \leq X < x_2\} = \int_{\tilde{a}}^{x_2} \tilde{f}_X(x) dx \approx \frac{1}{20,185} \cdot \int_{2,217}^4 dx \approx 0,088,$$

$$p_2 = P\{x_2 \leq X < x_3\} = \int_{x_2}^{x_3} \tilde{f}_X(x) dx \approx \frac{1}{20,185} \cdot \int_4^6 dx \approx 0,099.$$

Длины интервалов группировки под номерами 3 – 9 равны длине второго интервала, поэтому

$$p_3 = \dots = p_9 = p_2 \approx 0,099.$$

Осталось вычислить p_{10} :

$$p_{10} = P\{x_{10} \leq X < \tilde{b}\} = \int_{x_{10}}^{\tilde{b}} \tilde{f}_X(x) dx \approx \frac{1}{20,185} \cdot \int_{20}^{22,403} dx \approx 0,120.$$

Найдём теоретические частоты $n'_i = np_i = 200p_i$:

$$n'_1 \approx 17,6, \quad n'_2 = n'_3 = \dots = n'_9 \approx 19,8, \quad n'_{10} \approx 24,0.$$

Составим расчётную табл.1.10.

Таблица 1.10

Данные для подсчёта статистики χ^2

Номер интервала i	n_i	n'_i	$\frac{(n_i - n'_i)^2}{n'_i}$
1	21	17,6	0,657
2	16	19,8	0,729
3	15	19,8	1,164
4	26	19,8	1,941
5	22	19,8	0,244
6	14	19,8	1,699
7	21	19,8	0,073
8	22	19,8	0,244
9	18	19,8	0,164
10	25	24,0	0,042

Вычислим значение статистики:

$$\begin{aligned} \chi^2_{\text{в}} &= \sum_{i=1}^{10} \frac{(n_i - n'_i)^2}{n'_i} = 0,657 + 0,729 + 1,164 + 1,941 + 0,244 + \\ &+ 1,699 + 0,073 + 0,244 + 0,164 + 0,042 = 6,957. \end{aligned}$$

Используя Matlab, найдём квантиль $\chi^2_{1-\alpha}(10-2-1) = \chi^2_{0,95}(7) = 14,0671$. Поскольку $\chi^2_{\text{в}} < \chi^2_{0,95}(7)$, то нет оснований отвергать гипотезу о равномерном распределении генеральной совокупности.

1.6. Статистическое описание и выборочные характеристики двумерного случайного вектора

Пусть $(x_i; y_i)$, $i = 1, 2, \dots, n$, – выборка объёмом n из наблюдений двумерного случайного вектора $(X; Y)$. Предварительное представление о двумерной генеральной совокупности можно получить, изображая элементы выборки точками на плоскости с выбранной декартовой прямоугольной системой координат. Это представление выборки называется *диаграммой рассеивания* (*корреляционным полем*).

Распределением двумерной выборки называется распределение двумерного дискретного случайного вектора, принимающего значения $(x_i; y_i)$, $i = 1, 2, \dots, n$, с вероятностями, равными $1/n$. *Выборочные числовые характеристики* вычисляются как соответствующие числовые характеристики двумерного случайного вектора дискретного типа.

Двумерную выборку *большого объёма* представляют в виде *корреляционной таблицы*. С этой целью группируют реализации величин X и Y по интервалам длины Δ_x и Δ_y , а в клетки таблицы записывают число пар исходной выборки (т.е. частоты) для каждой комбинации интервалов. Эту процедуру можно также выполнить непосредственно по диаграмме рассеивания, нанося на неё сетку горизонтальных и вертикальных прямых, взятых с постоянными шагами Δ_x и Δ_y . Наблюдения, которые попадают на верхнюю и правую границы рассматриваемого прямоугольника, относятся соответственно к соседним верхнему и правому прямоугольникам. В дальнейших вычислениях эмпирических числовых характеристик используются середины интервалов и соответствующие частоты.

Обозначим середины интервалов группировки через x'_i , $i = 1, 2, \dots, k$, и y'_j , $j = 1, 2, \dots, l$, а соответствующие им частоты – через n_{ij} (очевидно, что $\sum_{i=1}^k \sum_{j=1}^l n_{ij} = n$). Положим также, что $\sum_{j=1}^l n_{ij} = n_{i\bullet}$,

$\sum_{i=1}^k n_{ij} = n_{\bullet j}$. Тогда для *группированных* данных *выборочные средние*, «исправленные» *дисперсии* и *ковариация* определяются по соотношениям

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^k x'_i n_{i\bullet}, \quad \bar{y} = \frac{1}{n} \sum_{j=1}^l y'_j n_{\bullet j}; \\ s_X^2 &= \frac{1}{n-1} \sum_{i=1}^k (x'_i - \bar{x})^2 n_{i\bullet} = \frac{1}{n-1} \left(\sum_{i=1}^k x'^2_{i\bullet} n_{i\bullet} - n\bar{x}^2 \right), \\ s_Y^2 &= \frac{1}{n-1} \sum_{j=1}^l (y'_j - \bar{y})^2 n_{\bullet j} = \frac{1}{n-1} \left(\sum_{j=1}^l y'^2_{\bullet j} n_{\bullet j} - n\bar{y}^2 \right), \\ \tilde{K}_{X,Y} &= \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^l (x'_i - \bar{x})(y'_j - \bar{y}) n_{ij} = \frac{1}{n-1} \left(\sum_{i=1}^k \sum_{j=1}^l n_{ij} x'_i y'_j - n\bar{x}\bar{y} \right). \end{aligned}$$

Для *негруппированных* данных аналогичные формулы упрощаются и приобретают вид:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{m=1}^n x_m, \quad \bar{y} = \frac{1}{n} \sum_{m=1}^n y_m; \\ s_X^2 &= \frac{1}{n-1} \sum_{m=1}^n (x_m - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{m=1}^n x_m^2 - n\bar{x}^2 \right), \\ s_Y^2 &= \frac{1}{n-1} \sum_{m=1}^n (y_m - \bar{y})^2 = \frac{1}{n-1} \left(\sum_{m=1}^n y_m^2 - n\bar{y}^2 \right), \end{aligned}$$

$$\tilde{K}_{X,Y} = \frac{1}{n-1} \sum_{m=1}^n (x_m - \bar{x})(y_m - \bar{y}) = \frac{1}{n-1} \left(\sum_{m=1}^n x_m y_m - n\bar{x}\bar{y} \right).$$

Выборочный коэффициент корреляции (как для группированных, так и для негруппированных данных) определяется по формуле

$$\tilde{\rho}_{X,Y} = \frac{\tilde{K}_{X,Y}}{s_X s_Y},$$

где $s_X = \sqrt{s_X^2}$, $s_Y = \sqrt{s_Y^2}$. Выборочный коэффициент корреляции $\tilde{\rho}_{X,Y}$ обладает всеми свойствами, которыми обладает теоретико-вероятностный коэффициент корреляции $\rho_{X,Y}$. В частности, для любой выборки $|\tilde{\rho}_{X,Y}| \leq 1$. При этом, чем ближе $\tilde{\rho}_{X,Y}$ к 1 (или к -1), тем сильнее выражена линейная зависимость между X и Y .

Пример 1.9. Дана двумерная выборка объёмом $n = 50$ (табл.1.11).

Таблица 1.11

Выборочные данные двумерного случайного вектора

№ п/п	X	Y	№ п/п	X	Y
1	10,30	5,94	26	6,77	3,50
2	2,38	1,33	27	6,38	2,68
3	3,80	3,71	28	2,59	-1,05
4	-0,70	-0,49	29	12,18	6,01
5	5,60	3,03	30	6,89	4,77
6	4,47	0,81	31	1,13	4,92
7	3,25	3,72	32	5,70	5,93
8	4,85	2,21	33	14,09	9,08
9	-0,23	-1,90	34	8,84	5,97
10	7,55	-0,15	35	7,42	6,19
11	2,57	0,99	36	7,31	1,74
12	1,76	5,03	37	6,92	-1,10
13	5,41	-1,13	38	10,65	7,81
14	9,84	5,81	39	1,87	1,79
15	0,36	-3,38	40	7,91	4,75
16	4,25	3,58	41	4,18	1,22
17	11,73	3,14	42	6,69	3,87
18	13,01	0,66	43	3,30	5,64
19	3,28	1,70	44	3,12	-1,54
20	7,68	3,88	45	7,24	5,29
21	12,64	5,29	46	5,70	1,77
22	6,87	3,06	47	6,53	3,05
23	7,39	6,62	48	4,71	3,46
24	-0,87	-4,56	49	7,21	2,12
25	0,96	2,30	50	9,96	5,35

Выполнить предварительную обработку результатов наблюдений, включающую:

- 1) построение диаграммы рассеивания (корреляционного поля);
- 2) группировку данных и построение корреляционной таблицы;
- 3) оценку числовых характеристик для негруппированных и группированных данных.

Решение. 1. Диаграмма рассеивания с нанесённой на неё сеткой горизонтальных и вертикальных прямых ($\Delta_x = \Delta_y = 3$) представлена на рис.1.11.

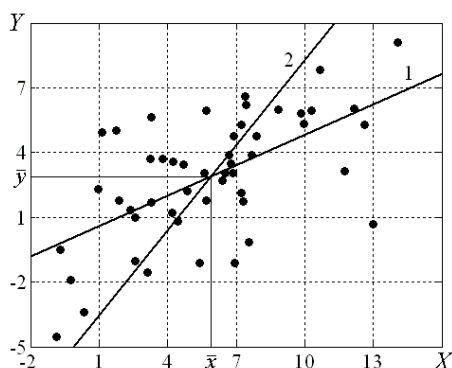


Рис.1.11. Диаграмма рассеивания с нанесённой на неё сеткой для группировки данных и графики уравнений выборочных линейных регрессий:
1 – Y на x ; 2 – X на y

2. Используя графическую группировку данных (с помощью нанесённой на рис.1.11 сетки), построим таблицу частот (табл.1.12).

Таблица 1.12

Таблица частот группированной двумерной выборки

$Y \backslash X$	-3,5	-0,5	2,5	5,5	8,5	$n_{i \cdot}$
-0,5	2	2	1	0	0	5
2,5	0	3	5	3	0	11
5,5	0	3	11	2	0	16
8,5	0	1	3	7	0	11
11,5	0	0	1	3	1	5
14,5	0	1	0	0	1	2
$n_{\cdot j}$	2	10	21	15	2	50

В первом столбце и в первой строке табл.1.12 указаны середины интервалов группировки по X и по Y соответственно.

3. Проведём вычисление выборочных числовых характеристик:

– для негруппированных данных:

$$\bar{x} = 5,8688, \bar{y} = 2,8884, s_X^2 = 13,4696, s_Y^2 = 8,3570, \tilde{K}_{X,Y} = 6,3564, \tilde{\rho}_{X,Y} = 0,5991;$$

– для группированных данных:

$$\bar{x} = 5,86, \bar{y} = 2,80, s_X^2 = 14,5616, s_Y^2 = 7,4388, \tilde{K}_{X,Y} = 5,5837, \tilde{\rho}_{X,Y} = 0,5365.$$

1.7. Элементы корреляционного и регрессионного анализа данных

1.7.1. Корреляционный анализ

Корреляционный анализ – это совокупность методов обнаружения так называемой корреляционной зависимости между случайными величинами.

Для двух случайных величин X и Y корреляционный анализ включает следующие этапы:

– построение корреляционного поля (диаграммы рассеивания) и составление корреляционной таблицы;

– вычисление выборочного коэффициента корреляции;

– проверку статистической гипотезы о значимости корреляционной связи.

Корреляционное поле и *корреляционная таблица* являются исходными данными при корреляционном анализе. По характеру расположения точек на диаграмме рассеивания можно составить предварительное представление о форме зависимости случайных величин (например, о том, что одна из них в среднем возрастает или убывает с возрастанием другой) (см. пример 1.9, рис.1.11).

Корреляционная таблица даёт возможность выдвинуть предположение о наличии или отсутствии корреляционной связи, а также выяснить её направление (см. пример 1.9, табл.1.12). Если частоты в корреляционной таблице расположены по диагонали из левого верхнего угла в правый нижний угол, то можно предположить наличие прямой корреляционной зависимости между признаками. Если же частоты расположены по диагонали из правого верхнего угла в левый нижний, то можно предположить наличие

обратной связи между признаками. Необходимо подчеркнуть, что при рассмотрении корреляционной таблицы важно установить расположение основной части частот. Возможны варианты, когда все клетки корреляционной таблицы окажутся заполненными. Однако это обстоятельство ещё не означает, что корреляционная связь между признаками отсутствует.

Проверка гипотезы о наличии корреляции осуществляется следующим образом. Основная гипотеза – отсутствие линейной статистической связи ($H_0: \rho_{X,Y} = 0$); альтернативной гипотезой может выступать любая из трёх возможных ($H_1^{(1)}: \rho_{X,Y} < 0$, $H_1^{(2)}: \rho_{X,Y} \neq 0$ или $H_1^{(3)}: \rho_{X,Y} > 0$).

В тех случаях, когда справедливо предположение о нормальном распределении двумерного генерального вектора $(X; Y)$, подходящей статистикой Z для проверки основной гипотезы об отсутствии линейной статистической связи (по двумерной выборке объёмом n) является

$$Z = \frac{\tilde{\rho}_{X,Y} \cdot \sqrt{n-2}}{\sqrt{1-\tilde{\rho}_{X,Y}^2}} \sim \text{St}(n-2).$$

Критерий проверки гипотезы $H_0: \rho_{X,Y} = 0$ при альтернативной гипотезе $H_1^{(2)}: \rho_{X,Y} \neq 0$ заключается в следующем:

- вычисляется выборочное значение z_B статистики Z ;
- если $|z_B| < t_{1-\frac{\alpha}{2}}(n-2)$, то нет оснований отвергать гипотезу H_0 ; иначе H_0 отклоняется с

ошибкой первого рода α в пользу гипотезы $H_1^{(2)}$ (корреляция значима).

Замечание. Статистика Стьюдента сильно зависит от объёма выборки. Для получения надёжных выводов о наличии корреляции при использовании статистики Z желательно иметь более 100 наблюдений.

Пример 1.10. Для негруппированных данных примера 1.9 проверить гипотезу $H_0: \rho_{X,Y} = 0$ об отсутствии линейной статистической связи между компонентами X и Y при двусторонней альтернативе $H_1^{(2)}: \rho_{X,Y} \neq 0$ и $\alpha = 0,05$.

Решение. Для негруппированных данных примера 1.9 выборочная оценка коэффициента корреляции $\tilde{\rho}_{X,Y}$ равна 0,5991. Используя Matlab, найдём квантиль распределения Стьюдента $t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(48) = 2,0106$. Тогда выборочное значение статистики Z равно

$$z_B = \frac{0,5991 \cdot \sqrt{48}}{\sqrt{1-0,5991^2}} \approx 5,184.$$

Так как $|z_B| > t_{0,975}(48)$, то гипотеза H_0 отклоняется в пользу гипотезы $H_1^{(2)}$. Корреляция значима.

Менее чувствительной к объёму n выборки из генеральной совокупности, имеющей двумерное нормальное распределение, является статистика U для проверки более общей гипотезы $H_0: \rho_{X,Y} = \rho_0$ против любой из трёх альтернатив $H_1^{(1)}: \rho_{X,Y} < \rho_0$, $H_1^{(2)}: \rho_{X,Y} \neq \rho_0$ или $H_1^{(3)}: \rho_{X,Y} > \rho_0$:

$$U = \frac{\text{arth } \tilde{\rho}_{X,Y} - \text{arth } \rho_0}{1/\sqrt{n-3}},$$

где $\text{arth } x = \frac{1}{2} \ln \frac{1+x}{1-x}$ ($|x| < 1$), а ρ_0 – истинное (но неизвестное) значение коэффициента корреляции двумерного генерального вектора $(X; Y)$.

Фишером было доказано, что если гипотеза H_0 верна, то случайная величина U при $n \geq 30$ имеет распределение, близкое к нормальному $\mathbf{N}(0; 1)$:

$$U \sim \mathbf{N}(0; 1).$$

Критерий проверки гипотезы $H_0: \rho_{X,Y} = \rho_0$ при альтернативной гипотезе $H_1^{(2)}: \rho_{X,Y} \neq \rho_0$ заключается в следующем:

- вычисляется выборочное значение u_B статистики U ;

– если $|u_B| < u_{1-\frac{\alpha}{2}}$, то нет оснований отвергать гипотезу H_0 ; иначе H_0 отклоняется с ошибкой первого

рода α в пользу гипотезы $H_1^{(2)}$.

Пример 1.11. Пользуясь статистикой Фишера U , для негруппированных данных примера 1.9 проверить гипотезу $H_0: \rho_{X,Y} = 0$ об отсутствии линейной статистической связи между компонентами X и Y при двусторонней альтернативе $H_1^{(2)}: \rho_{X,Y} \neq 0$ и $\alpha = 0,05$.

Решение. Для негруппированных данных примера 1.9 выборочная оценка коэффициента корреляции $\tilde{\rho}_{X,Y}$ равна 0,5991. Используя Matlab, найдём квантиль нормального распределения $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,9600$. Тогда выборочное значение статистики U равно (учитывая $\rho_0 = 0$)

$$u_B = \frac{\sqrt{47}}{2} \cdot \ln \frac{1+0,5991}{1-0,5991} \approx 4,7423.$$

Так как $|u_B| > u_{0,975}$, то гипотеза H_0 отклоняется в пользу гипотезы $H_1^{(2)}$. Корреляция значима.

Как уже отмечалось, при малом объёме выборки точечными оценками числовых характеристик пользоваться некорректно. Необходима интервальная оценка.

Доверительный интервал для коэффициента корреляции:

$$\text{th} \left(\frac{1}{2} \ln \frac{1+\tilde{\rho}}{1-\tilde{\rho}} - \frac{u_{1-\alpha/2}}{\sqrt{n-3}} - \frac{\tilde{\rho}}{2(n-1)} \right) < \rho < \text{th} \left(\frac{1}{2} \ln \frac{1+\tilde{\rho}}{1-\tilde{\rho}} + \frac{u_{1-\alpha/2}}{\sqrt{n-3}} - \frac{\tilde{\rho}}{2(n-1)} \right),$$

где $\text{th } x$ – гиперболический тангенс аргумента x ; ρ и $\tilde{\rho}$ – истинное значение и выборочная оценка коэффициента корреляции соответственно; u_p – квантиль нормального распределения порядка p ; α – уровень значимости.

Пример 1.12. Для негруппированных данных примера 1.9 получить интервальную оценку истинного значения коэффициента корреляции $\rho_{X,Y}$ при уровне значимости $\alpha = 0,05$.

Решение. Для негруппированных данных примера 1.9 выборочная оценка коэффициента корреляции $\tilde{\rho}_{X,Y}$ равна 0,5991. Тогда, используя Matlab, найдём:

$$u_{1-\alpha/2} = u_{0,975} = 1,9600,$$

$$\frac{1}{2} \ln \frac{1+\tilde{\rho}}{1-\tilde{\rho}} - \frac{u_{1-\alpha/2}}{\sqrt{n-3}} - \frac{\tilde{\rho}}{2(n-1)} \approx 0,3997, \quad \text{th}(0,3997) \approx 0,3797,$$

$$\frac{1}{2} \ln \frac{1+\tilde{\rho}}{1-\tilde{\rho}} + \frac{u_{1-\alpha/2}}{\sqrt{n-3}} - \frac{\tilde{\rho}}{2(n-1)} \approx 0,9715, \quad \text{th}(0,9715) \approx 0,7494.$$

Следовательно, $0,3797 < \rho_{X,Y} < 0,7494$.

1.7.2. Регрессионный анализ

Зависимость между случайными величинами X и Y называется *стохастической*, если с изменением одной из них (например, X) меняется закон распределения другой (Y). В качестве примеров такой зависимости приведём зависимость веса человека (Y) от его роста (X), предела прочности стали (Y) от её твёрдости (X) и т.д.

Напомним, что в теории вероятностей стохастическую зависимость Y от X описывают условным математическим ожиданием

$$y(x) = M[Y | X = x],$$

которое, как видно из записи, является функцией от независимой переменной x , имеющей смысл возможного значения случайной величины X . Уравнение $y = y(x)$ называется *уравнением регрессии Y на x* . Переменная x называется *регрессионной* переменной или *регрессором*. График функции $y = y(x)$ называется *линией* или *кривой регрессии*.

Кривые регрессии обладают следующим свойством: среди всех действительных функций $\varphi(x)$ минимум $M[(Y - \varphi(x))^2]$ достигается для функции $y(x) = M[Y | X = x]$. Иными словами, регрессия Y на x даёт наилучшее (в среднеквадратическом смысле) предсказание величины Y по заданному значению $X = x$. На

практике это свойство используется для прогноза Y по X , если непосредственно наблюдаемой величиной является лишь компонента X случайного вектора $(X; Y)$. Например, если X – диаметр сосны, то в качестве прогнозируемого значения Y (высота сосны) берётся условное математическое ожидание $y(x)$.

Наиболее простым является случай, когда регрессия Y на x линейна:

$$y(x) = a_0 + a_1 x.$$

Напомним, что если $(X; Y)$ – случайный вектор, распределённый по двумерному нормальному закону, то коэффициенты a_0 и a_1 определяются равенствами

$$a_0 = m_Y - \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} m_X, \quad a_1 = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}.$$

Уравнением регрессии в этом случае является прямая линия

$$y(x) = m_Y + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \cdot (x - m_X),$$

проходящая через центр рассеивания $(m_X; m_Y)$ с угловым коэффициентом $\eta_{X,Y} = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}$,

называемым коэффициентом регрессии Y на x .

В реальных экспериментах, связанных со статистической обработкой опытных данных, условный закон распределения случайной величины Y при $X = x$ обычно заранее неизвестен. В таком случае речь может идти лишь о каком-либо приближении к теоретической кривой регрессии, построенном на основе выборочных данных. Другими словами, задача заключается в подборе подходящей функциональной зависимости, наилучшим образом (в некотором статистическом смысле) приближающей стохастическую зависимость.

Во многих случаях можно считать, что «независимая» переменная X находится под контролем экспериментатора и может быть измерена с любой заданной точностью, в то время как измеряемые значения Y как функции от X (выборочные значения y_i при фиксированных значениях x_i) определяются с ошибкой (содержат шум измерения). Если вид функциональной зависимости зафиксирован, то статистическую модель регрессии можно записать следующим образом:

$$y_i = y(x_i) = \varphi(x_i; a_0, a_1, \dots, a_m) + \varepsilon(x_i),$$

где a_0, a_1, \dots, a_m – набор неизвестных параметров, определяющих функциональную зависимость (параметры регрессии); $\varepsilon(x_i) = \varepsilon_i$ – случайные величины, складывающиеся при каждом фиксированном значении x_i из шума измерения и ошибки модели. При исследовании качества построенной модели важно уметь разделять эти ошибки.

Следует иметь в виду, что наличие шума измерения делает невозможной задачу интерполяции, т.е. график искомой зависимости не должен проходить через все выборочные точки, а должен проходить таким образом, чтобы «сгладить» шум. Поскольку уровень шума определяется дисперсией $D[\varepsilon]$, то задача состоит в подборе параметров a_0, a_1, \dots, a_m , минимизирующих $D[\varepsilon]$. В действительности минимизируется не сама дисперсия (она неизвестна), а её выборочная оценка, которая, как будет показано ниже, пропорциональна сумме квадратов отклонений (по оси Oy) кривой регрессии от соответствующих выборочных значений y_i , т.е. пропорциональна величине

$$\Psi = \sum_{k=1}^n [y_k - \varphi(x_k; \tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_m)]^2.$$

Указанный критерий минимизации суммы квадратов отклонений носит название *метода наименьших квадратов* (МНК), а полученные в результате решения этой задачи оценки $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_m$ параметров называются МНК-оценками.

Как правило, в модели регрессии

$$y_i = \varphi(x_i; a_0, a_1, \dots, a_m) + \varepsilon_i$$

случайные величины ε_i , $i = 1, 2, \dots, n$, независимы в совокупности и одинаково распределены по закону $\mathbf{N}(0; \sigma)$ (физически условие $M[\varepsilon_i] = 0$ означает, что систематические ошибки измерений отсутствуют, а условие $D[\varepsilon_i] = \sigma^2$, $i = 1, 2, \dots, n$, означает, что измерения проводятся с одинаковой точностью).

Регрессионный анализ проводится в *три этапа*.

На *первом этапе* по характеру корреляционного поля выдвигают гипотезу о виде функциональной зависимости $\varphi(x_k; a_0, a_1, \dots, a_m)$. Довольно часто используют следующее представление функции φ :

$$\varphi(x_k; a_0, a_1, \dots, a_m) = a_0 + \sum_{k=1}^m a_k \varphi_k(x),$$

где $\varphi_k(x)$ – известные *координатные функции*. Такая модель регрессии называется *линейной по параметрам*. В частном случае, когда $\varphi_k(x) = x^k$, модель называется *полиномиальной*.

На *втором этапе* по имеющимся выборочным данным осуществляют подгонку модели, т.е. находят МНК-оценки неизвестных параметров регрессии a_0, a_1, \dots, a_m .

На *третьем этапе* анализируют качество построения модели: проверяются так называемые корректность и адекватность модели. Этот этап осуществляется средствами проверки статистических гипотез.

Пример 1.13 (*Построение прямой регрессии Y на x*). Пусть получена выборка $(x_k; y_k)$, $k = 1, 2, \dots, n$, из двумерного распределения $(X; Y)$. Корреляционный анализ показал, что корреляционная связь Y от X значима на некотором уровне α . Выдвигается гипотеза о том, что уравнение прямой регрессии

$$y(x) = ax + b + \varepsilon(x)$$

должно хорошо аппроксимировать стохастическую зависимость Y от X . Найти МНК-оценки параметров a и b .

Решение. Пусть задан план эксперимента, т.е. совокупность точек x_1, x_2, \dots, x_n . Выбор этих точек – отдельная задача, решаемая в рамках теории оптимального планирования эксперимента, и здесь не обсуждается.

Искомые оценки являются решениями следующей задачи минимизации:

$$\psi(a, b) = \sum_{k=1}^n \varepsilon_k^2 = \sum_{k=1}^n [y_k - ax_k - b]^2 \rightarrow \min_{a, b}.$$

Применим классический метод поиска безусловного экстремума дифференцируемой функции $\psi(a, b)$. Запишем необходимые условия экстремума:

$$\frac{\partial \psi}{\partial a} = 0, \quad \frac{\partial \psi}{\partial b} = 0.$$

Получим следующую систему линейных алгебраических уравнений для неизвестных значений a и b :

$$\begin{cases} a \sum_{k=1}^n x_k^2 + b \sum_{k=1}^n x_k = \sum_{k=1}^n x_k y_k, \\ a \sum_{k=1}^n x_k + bn = \sum_{k=1}^n y_k. \end{cases}$$

Разделив обе части каждого уравнения системы на n и введя обычные обозначения для выборочных характеристик случайного вектора $(X; Y)$, приведём данную систему к виду

$$\begin{cases} a \tilde{\alpha}_{2,0} + b \bar{x} = \tilde{\alpha}_{1,1}, \\ a \bar{x} + b = \bar{y}, \end{cases}$$

где $\tilde{\alpha}_{r,s} = \frac{1}{n} \sum_{k=1}^n x_k^r y_k^s$ – начальный выборочный момент порядка $r+s$; \bar{x} и \bar{y} – средние значения соответствующих переменных.

Решение $(\tilde{a}; \tilde{b})$ последней системы, как нетрудно убедиться, имеет вид:

$$\tilde{a} = \frac{\tilde{K}_{X,Y}}{s_X^2} = \tilde{\rho}_{X,Y} \cdot \frac{s_Y}{s_X}, \quad \tilde{b} = \bar{y} - \bar{x} \cdot \tilde{\rho}_{X,Y} \cdot \frac{s_Y}{s_X},$$

где $\tilde{K}_{X,Y}$, s_X , s_Y , $\tilde{\rho}_{X,Y}$ – выборочные числовые характеристики, определённые в разделе 1.6.

Тогда уравнение линейной регрессии, полученное методом наименьших квадратов, приобретает вид:

$$y(x) = \bar{y} + \tilde{\rho}_{X,Y} \frac{s_Y}{s_X} (x - \bar{x}).$$

Замечание. Полученное уравнение аналогично теоретическому уравнению регрессии $y(x) = m_Y + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \cdot (x - m_X)$, если заменить все входящие в него вероятностные моменты соответствующими выборочными оценками.

Аналогично рассматривается регрессия X на y . Оценки параметров a' и b' линейной регрессии $x(y) = a'y + b' + \varepsilon'(y)$ вычисляются по формулам

$$\tilde{a}' = \frac{\tilde{K}_{X,Y}}{s_Y^2} = \tilde{\rho}_{X,Y} \cdot \frac{s_X}{s_Y}, \quad \tilde{b}' = \bar{x} - \bar{y} \cdot \tilde{\rho}_{X,Y} \cdot \frac{s_X}{s_Y}.$$

Тогда уравнение приобретает вид:

$$x(y) = \bar{x} + \tilde{\rho}_{X,Y} \frac{s_X}{s_Y} (y - \bar{y}).$$

Для контроля правильности расчётов можно использовать соотношение $\sqrt{\tilde{a} \cdot \tilde{a}'} = |\tilde{\rho}_{X,Y}|$.

Прямые $y = \tilde{a}x + \tilde{b}$ и $x = \tilde{a}'y + \tilde{b}'$ пересекаются в точке с координатами $(\bar{x}; \bar{y})$, причём угол между ними уменьшается при увеличении коэффициента корреляции между случайными величинами X и Y . При $|\rho_{X,Y}| = 1$ обе прямые совпадают. Следует подчеркнуть, что $y = \tilde{a}x + \tilde{b}$ и $x = \tilde{a}'y + \tilde{b}'$ – различные прямые. Первая прямая получается в результате решения задачи о минимизации суммы квадратов отклонений по вертикали, а вторая – при решении задачи о минимизации суммы квадратов отклонений по горизонтали.

Во многих случаях графическое представление данных показывает, что интересующая зависимость не может быть описана линейной регрессией. Очень часто достаточно точным для описания фактической взаимозависимости является уравнение второго порядка

$$y = \tilde{a}x^2 + \tilde{b}x + \tilde{c}.$$

Когда вид нелинейной зависимости между двумя переменными известен, можно с помощью преобразований одной или обеих переменных получить линейную зависимость между ними. Например, если предположить, что описываемая зависимость представляет собой функцию вида

$$y = a \cdot b^x,$$

то следует применить логарифмирование $\ln y = \ln a + x \ln b$. Для оценки коэффициентов линейной регрессии $\ln a$ и $\ln b$ можно использовать метод наименьших квадратов (см., например, [1]).

Пример 1.14. Для негруппированных и группированных данных примера 1.9 составить уравнения линейной регрессии Y на x и X на y . Для негруппированной двумерной выборки построить прямые регрессии Y на x и X на y .

Решение. Рассмотрим вначале случай негруппированных данных.

В примере 1.12 построен доверительный интервал для коэффициента корреляции $\rho_{X,Y}$ при $\alpha = 0,05$:

$$0,3797 < \rho_{X,Y} < 0,7494.$$

Этот интервал не содержит нуля, т.е. с доверительной вероятностью $1 - \alpha = 0,95$ существует корреляция между X и Y и имеет смысл построение уравнений регрессии.

Подставим найденные в примере 1.9 значения \bar{x} , \bar{y} , s_X , s_Y , $\tilde{\rho}_{X,Y}$ в уравнения линейной регрессии Y на x и X на y :

$$y(x) = \bar{y} + \tilde{\rho}_{X,Y} \frac{s_Y}{s_X} (x - \bar{x}), \quad x(y) = \bar{x} + \tilde{\rho}_{X,Y} \frac{s_X}{s_Y} (y - \bar{y}).$$

Получим

$$y = 0,1189 + 0,4719x, \quad x = 3,672 + 0,7606y.$$

На рис.1.11 построены корреляционное поле, линии регрессии Y на x (прямая 1) и X на y (прямая 2). Прямые регрессии пересекаются в точке с координатами $(\bar{x}; \bar{y}) = (5,8688; 2,8884)$.

Проверка: $\sqrt{\tilde{a} \cdot \tilde{a}'} = \sqrt{0,4719 \cdot 0,7606} \approx 0,5991 = |\tilde{\rho}_{X,Y}|$.

Рассмотрим теперь случай группированных данных.

Подставим найденные в примере 1.9 значения \bar{x} , \bar{y} , s_X , s_Y , $\tilde{\rho}_{X,Y}$ в уравнения линейной регрессии Y на x и X на y . Получим:

$$y = 0,5529 + 0,3835x, \quad x = 3,7583 + 0,7506y.$$

$$\text{Проверка: } \sqrt{\tilde{a} \cdot \tilde{a}'} = \sqrt{0,3835 \cdot 0,7506} \approx 0,5365 = |\tilde{\rho}_{X,Y}|.$$

Качество аппроксимации результатов наблюдений $(x_i; y_i)$, $i = 1, 2, \dots, n$, выборочной регрессией $y = \tilde{a}x + \tilde{b}$ (ограничимся далее регрессией Y на x) определяется величиной остаточной дисперсии s^2 , вычисляемой по формуле

$$s^2 = \frac{Q_e}{n-2}.$$

Величина Q_e называется *остаточной суммой квадратов* и определяется выражением

$$Q_e = \sum_{i=1}^n (y_i - \tilde{y}_i)^2,$$

где $y_i - \tilde{y}_i$, $i = 1, 2, \dots, n$, – разности между наблюдаемыми значениями переменной Y при $x = x_i$ и расчётными значениями $\tilde{y}_i = \tilde{a}x_i + \tilde{b}$, называемые *остатками*.

В практических вычислениях остаточную сумму квадратов получают из тождества

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \tilde{y}_i)^2,$$

которое записывается в виде

$$Q_y = Q_R + Q_e,$$

где

$$Q_y = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2,$$

$$Q_R = \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2 = \tilde{a} \cdot (n-1) \tilde{K}_{X,Y} = \tilde{a}^2 \cdot (n-1) s_X^2 = \frac{(n-1) \tilde{K}_{X,Y}^2}{s_X^2}.$$

Величина Q_R называется *суммой квадратов, обусловленной регрессией*.

Полезной характеристикой линейной регрессии является *коэффициент детерминации* R^2 , вычисляемый по формуле

$$R^2 = \frac{Q_R}{Q_y} = 1 - \frac{Q_e}{Q_y}.$$

Коэффициент детерминации R^2 равен той доле разброса результатов наблюдений $(x_i; y_i)$, $i = 1, 2, \dots, n$, относительно горизонтальной прямой $y = \bar{y}$, которая объясняется выборочной регрессией $y = \tilde{a}x + \tilde{b}$. Величина $R = +\sqrt{R^2}$ является оценкой коэффициента корреляции $\tilde{\rho}_{Y,\tilde{Y}}$ между результатами наблюдений y_i и вычисленными значениями \tilde{y}_i , предсказываемыми регрессией $y = \tilde{a}x + \tilde{b}$, т.е. $R = \tilde{\rho}_{Y,\tilde{Y}}$. В случае линейной регрессии Y на x (одной независимой переменной x) между коэффициентом R и выборочным коэффициентом корреляции $\tilde{\rho}_{X,Y}$ имеется следующее соотношение:

$$\tilde{\rho}_{X,Y} = (\text{знак } \tilde{a})R.$$

Линейная регрессионная модель $y(x) = ax + b + \varepsilon(x)$ называется *незначимой*, если параметр $a = 0$. Для проверки гипотезы $H_0: a = 0$ можно использовать доверительный интервал для параметра a :

$$\tilde{a} - t_{1-\frac{\alpha}{2}}(n-2) \cdot \sqrt{\frac{S^2}{(n-1) \cdot S_X^2}} < a < \tilde{a} + t_{1-\frac{\alpha}{2}}(n-2) \cdot \sqrt{\frac{S^2}{(n-1) \cdot S_X^2}},$$

где α – уровень значимости; $t_p(n-2)$ – квантиль распределения Стьюдента с $n-2$ степенями свободы порядка p ; S^2 – остаточная дисперсия; S_X^2 – выборочная «исправленная» дисперсия компоненты X .

Часто для проверки гипотезы $H_0: a = 0$ используется статистика

$$F = \frac{(n-2) \cdot Q_R}{Q_e}.$$

Если гипотеза $H_0 : a = 0$ верна, то статистика F имеет распределение Фишера с 1 и $n-2$ степенями свободы:

$$F \sim \mathbf{F}(1; n-2).$$

Критерий проверки гипотезы H_0 при альтернативной гипотезе $H_1 : a \neq 0$ заключается в следующем:

- вычисляется выборочное значение f_b статистики F ;
- если $F_{\frac{\alpha}{2}}(1; n-2) < f_b < F_{1-\frac{\alpha}{2}}(1; n-2)$, то нет оснований отвергать гипотезу H_0 ; иначе H_0

отклоняется с ошибкой первого рода α в пользу гипотезы H_1 (здесь $F_p(1; n-2)$ – квантиль распределения Фишера с 1 и $n-2$ степенями свободы).

В случае, когда гипотеза $H_0 : a = 0$ отклоняется, говорят, что регрессионная модель $y(x) = ax + b + \varepsilon(x)$ статистически значима.

В линейной регрессионной модели $y(x) = ax + b + \varepsilon(x)$ полезными также являются:

- доверительный интервал для параметра b :

$$\tilde{b} - t_{1-\frac{\alpha}{2}}(n-2) \cdot \sqrt{\frac{S^2 \cdot \sum x_i^2}{n(n-1) \cdot S_X^2}} < b < \tilde{b} + t_{1-\frac{\alpha}{2}}(n-2) \cdot \sqrt{\frac{S^2 \cdot \sum x_i^2}{n(n-1) \cdot S_X^2}};$$

- доверительный интервал для среднего значения Y_0 , соответствующий заданному значению $x = x_0$ (границы интервала имеют вид):

$$\tilde{y}_0 \pm t_{1-\frac{\alpha}{2}}(n-2) \cdot \sqrt{S^2 \cdot \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1) \cdot S_X^2} \right)};$$

- доверительный интервал для дисперсии ошибок наблюдений σ^2 :

$$\frac{(n-2) \cdot S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-2)} < \sigma^2 < \frac{(n-2) \cdot S^2}{\chi_{\frac{\alpha}{2}}^2(n-2)},$$

где $\chi_p^2(n-2)$ – квантиль распределения χ^2 с $n-2$ степенями свободы.

Пример 1.15. Для негруппированных данных примера 1.9 по найденным оценкам параметров линейной регрессии Y на x :

- получить оценку s^2 дисперсии ошибок наблюдений σ^2 ;
- найти коэффициент детерминации R^2 ;
- построить доверительные интервалы для параметров регрессии a и b , дисперсии ошибок наблюдений σ^2 и среднего значения Y при $x = x_0$.

Проверить значимость линейной регрессии Y на x . Принять $\alpha = 0,05$.

Решение. Для негруппированных данных примера 1.9 были получены следующие оценки числовых характеристик и коэффициентов регрессии: $\bar{x} = 5,8688$, $\bar{y} = 2,8884$, $s_X^2 = 13,4696$, $s_Y^2 = 8,3570$, $\tilde{K}_{X,Y} = 6,3564$, $\tilde{\rho}_{X,Y} = 0,5991$, $\tilde{a} = 0,4719$, $\tilde{b} = 0,1189$.

Используя Matlab, найдём сумму квадратов, обусловленную регрессией $Q_R = 146,9833$, а также $Q_y = 409,4935$ и остаточную сумму квадратов $Q_e = 262,5102$.

Тогда оценка дисперсии ошибок наблюдений равна

$$s^2 = \frac{Q_e}{n-2} = 5,4690.$$

Коэффициент детерминации равен $R^2 = 0,3589$.

Поскольку $\tilde{\rho}_{X,Y} = (\text{знак } \tilde{a})R$, то сделаем проверку правильности расчётов: $0,5991 = +\sqrt{0,3589}$ (верно).

Полученный результат для коэффициента детерминации означает, что уравнение регрессии $y = 0,1189 + 0,4719x$ на 35,89% объясняет общий разброс результатов наблюдений относительно горизонтальной прямой $y = 2,8884$.

С помощью Matlab найдём квантили распределений Стьюдента и χ^2 :

$$t_{0,975}(48) = 2,0106, \chi_{0,975}^2(48) = 69,0226, \chi_{0,025}^2(48) = 30,7545.$$

Тогда можно построить доверительные интервалы для параметров линейной регрессии, дисперсии ошибок наблюдений

$$0,2889 < a < 0,6549, \quad -1,1444 < b < 1,3822, \quad 3,8033 < \sigma^2 < 8,5357,$$

а также найти границы доверительных интервалов для среднего значения Y при $x = x_0$:

$$\tilde{y}_0 \pm 2,0106 \cdot \sqrt{5,4690 \cdot \left(\frac{1}{50} + \frac{(x_0 - 5,8688)^2}{660,0104} \right)}.$$

На рис.1.12 пунктирными линиями показаны границы доверительных интервалов для среднего значения переменной Y при всех $x \in (-2; 16)$.

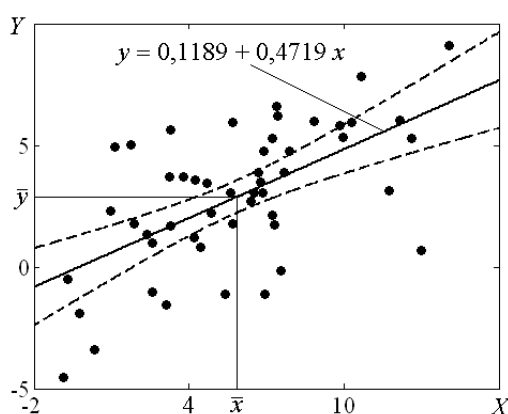


Рис.1.12. Линейная регрессия Y на x : графическое изображение границ доверительных интервалов для среднего значения переменной Y при всех $x \in (-2; 16)$ и $\alpha = 0,05$

Проверим значимость линейной регрессии Y на x .

Гипотеза $H_0: a = 0$ отклоняется на уровне значимости $\alpha = 0,05$, так как доверительный интервал $0,2889 < a < 0,6549$ не покрывает нуль с доверительной вероятностью 0,95.

Этот же результат можно получить, используя для проверки гипотезы $H_0: a = 0$ статистику $F = (n - 2) \cdot Q_R / Q_e$.

С помощью Matlab найдём квантили распределения Фишера:

$$F_{0,025}(1; 48) = 0,0010, \quad F_{0,975}(1; 48) = 5,3541.$$

Выборочное значение статистики F равно

$$f_B = 26,8759.$$

Поскольку $f_B \notin (0,0010; 5,3541)$, то гипотеза $H_0: a = 0$ отклоняется на уровне значимости $\alpha = 0,05$. Таким образом, линейная регрессия Y на x статистически значима.

Замечание. Из статистической значимости не следует, что регрессионная модель $y(x) = ax + b + \varepsilon(x)$ хорошо согласуется с результатами наблюдений, т.е. адекватна им.

Проверку адекватности модели можно проводить следующим образом. Пусть при различных значениях x_1, x_2, \dots, x_m переменной X получены повторные наблюдения переменной Y , причём при $x = x_i$

проведено n_i наблюдений Y : $y_{i1}, y_{i2}, \dots, y_{in_i}$. Очевидно, что $\sum_{i=1}^m n_i = n$, где n – объём всей выборки. Для

проверки адекватности модели используется статистика G , имеющая распределение Фишера с $m - 2$ и $n - m$ степенями свободы:

$$G = \frac{Q_n / (m - 2)}{Q_p / (n - m)} \sim \mathbf{F}(m - 2; n - m),$$

где $Q_p = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$, $Q_n = \sum_{i=1}^m n_i (\bar{y}_i - \tilde{y}_i)^2$, $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$, \tilde{y}_i – расчётные значения в линейной регрессионной модели, $i = 1, 2, \dots, m$.

Если выборочное значение статистики G удовлетворяет условию

$$g_b < F_{1-\alpha}(m-2; n-m)$$

(здесь $F_p(m-2; n-m)$ – квантиль распределения Фишера с $m-2$ и $n-m$ степенями свободы), то гипотеза об адекватности линейной регрессии результатам наблюдений принимается.

В этом случае остаточную дисперсию s^2 можно использовать в качестве оценки дисперсии ошибок наблюдений σ^2 , найти доверительные интервалы для параметров линейной регрессии и проверить гипотезы о параметрах. В противном случае нужно попытаться использовать другую модель, например, параболическую регрессию.

Пример 1.16. Для данных примера 1.9, сгруппированных только по X , проверить адекватность линейной регрессии Y на x . Принять $\alpha = 0,05$.

Решение. Для проверки адекватности воспользуемся корреляционной таблицей (см. табл.1.12). Будем считать, что середины интервалов группировки x_i , $i = 1, \dots, 6$, являются значениями компоненты X . Тогда число m повторных наблюдений равно 6. Запишем результаты этих наблюдений в виде табл.1.13.

Таблица 1.13

Повторные наблюдения, сгруппированные по X

x_i	-0,5	2,5	5,5	8,5	11,5	14,5
			-1,13			
			-1,10			
		-1,54	0,81	-0,15		
		-1,05	1,22	1,74		
		0,99	1,77	2,12		
	-4,56	1,33	2,21	3,88	3,14	
	-3,38	1,70	2,68	4,75	5,29	
	-1,90	1,79	3,03	5,29	5,94	0,66
	-0,49	3,71	3,05	5,35	6,01	9,08
	2,30	3,72	3,06	5,81	7,81	
		4,92	3,46	5,97		
		5,03	3,50	6,19		
		5,64	3,58	6,62		
			3,87			
			4,77			
			5,93			
n_i	5	11	16	11	5	2
\bar{y}_i	-1,606	2,385	2,544	4,325	5,638	4,87

Для удобства расчётов в последней строке таблицы приведены средние значения \bar{y}_i , $i = 1, \dots, 6$.

С помощью Matlab получим уравнение выборочной линейной регрессии Y на x для данных, сгруппированных по X :

$$y = 0,2851 + 0,4442x,$$

а также вычислим:

$$Q_p = \sum_{i=1}^6 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = 235,4766,$$

$$Q_n = \sum_{i=1}^6 n_i (\bar{y}_i - \tilde{y}_i)^2 = 33,1952,$$

где $\tilde{y}_i = 0,2851 + 0,4442 \cdot x_i$, $i = 1, \dots, 6$.

Выборочное значение статистики $G = \frac{Q_n/(m-2)}{Q_p/(n-m)}$ равно

$$g_{\text{в}} \approx \frac{33,1952/(6-2)}{235,4766/(50-6)} \approx 1,5507 .$$

Так как квантиль распределения Фишера, вычисленный с помощью Matlab, равен

$$F_{0,95}(4; 44) = 2,5837 ,$$

то $g_{\text{в}} < F_{0,95}(4; 44)$, а значит, линейная регрессия Y на x для данных примера 1.9, сгруппированных по X , адекватна результатам наблюдений.

2. Некоторые сведения о системе Matlab

Matlab (**Matrix Laboratory** – матричная лаборатория) – универсальная интегрированная система, предлагаемая её разработчиками как язык программирования высокого уровня для технических вычислений.

Matlab запускается нажатием левой клавиши мыши на ярлыке Matlab в рабочем меню операционной системы Windows. После этого появляется командное окно системы Matlab, и система готова к проведению вычислений в командном режиме. Сеанс работы с Matlab называется сессией. Сессия является текущим документом, отражающим работу пользователя с системой Matlab. В ней имеются строки ввода, вывода и сообщений об ошибках. После запуска программы Matlab в окне команд появляется символ `>>`, правее которого располагается рабочее поле программы для ввода команды. Набранная команда вводится клавишей «Enter». Команды, вводимые в процессе работы программы, сохраняются в памяти и могут быть вызваны в рабочее поле клавишами «↑» (вызов предыдущей команды) и «↓» (вызов следующей команды).

Команду, помещённую в рабочее поле, можно редактировать с использованием следующих клавиш: «←» – перемещение курсора на одну позицию влево, «→» – перемещение курсора на одну позицию вправо, «Delete» – удаление текущего символа, «Backspace» (или «←») – удаление символа слева от курсора. Исправленную команду можно вводить клавишей «Enter» с любой позиции курсора.

Удалить из памяти все величины, с которыми проводились вычисления, можно командой `clear`; `clear x` уничтожает в рабочем пространстве определение переменной `x`; `clear a,b,c` уничтожает в рабочем пространстве определения переменных списка. Уничтоженная (стёртая в рабочем пространстве) переменная становится неопределённой. Использовать эти переменные нельзя, такие попытки сопровождаются выдачей сообщений об ошибке (строка сообщений об ошибках начинается символами `???`). По мере задания одних переменных и уничтожения других рабочая область перестаёт быть непрерывной и содержит «дыры» и всякий «мусор». Во избежание непроизводительных потерь памяти при работе с объёмными данными следует использовать команду `pack`, осуществляющую дефрагментацию рабочей области. Команда `clc` очищает экран и размещает курсор в левом верхнем углу пустого экрана.

Справочная информация о функциях, командах и символах, применяемых в системе Matlab, может быть получена с помощью команды `help`. Например, с помощью команды `help if` на экран будет выдана справка об особенностях использования оператора «если».

Очень много сведений можно почерпнуть из демонстрационной программы пакета Matlab.

2.1. Ввод матриц и формирование упорядоченных числовых последовательностей

Система Matlab ориентирована на работу с матричными переменными. По умолчанию предполагается, что каждая заданная переменная – это матрица. Даже обычные константы и переменные рассматриваются в Matlab как матрицы размером 1×1 . Обозначив матрицу `A`, `B`, `A1`, `AB`, `a` и т.д. (можно использовать как заглавные, так и строчные буквы), можно ввести её следующим образом:

$$A=[1\ 2\ 3;4\ 5\ 6;7\ 8\ 9]$$

или

$$A = [1\ 2\ 3 \\ 4\ 5\ 6 \\ 7\ 8\ 9]$$

В первом случае элементы матрицы находятся внутри скобок «[» и «]», а строки разделяются символом «;», во втором случае элементы матрицы находятся внутри скобок «[» и «]», а ввод каждой строки заканчивается нажатием клавиши «Enter». Элементы каждой строки вводятся через пробел. Набрав, в частности, `b=[14\ 9\ -12]`, получим вектор-строку. Введя `c=[14;9;-12]`, получим вектор-столбец.

Замечание. Если в конце команды ставится символ «;», то команда выполняется, но результат её выполнения не выводится на экран.

Введённая матрица сохраняется в памяти до выхода из программы Matlab или до переопределения матрицы. Для вывода матрицы на экран достаточно ввести символ, её обозначающий.

Матрицу размерности $m \times n$ можно рассматривать как двумерный массив. Значение любого элемента матрицы можно изменять. Так, для рассмотренной ранее матрицы `A` ввод команды `A(2,3) = 11` приведёт к замене значения элемента $a_{23} = 6$ на 11.

Для формирования упорядоченных числовых последовательностей в Matlab применяется оператор двоеточие:

Начальное значение : Шаг : Конечное значение

Данная конструкция порождает последовательность (массив) чисел, которая начинается с начального значения, идёт с заданным шагом и завершается конечным значением. Если шаг не задан, то он принимает значения 1 или -1 .

Пример 2.1. `>> i=1:6`

```
i=
1 2 3 4 5 6
>> x=0:0.5:3
x=
0 0.5000 1.0000 1.5000 2.0000 2.5000 3.0000
>> x=3:-0.5:0
x=
3.0000 2.5000 2.0000 1.5000 1.0000 0.5000 0
```

2.2. Операции над массивами

Если A и B – матрицы, c – скалярная величина, то команды $D=A+B$, $D=A-B$, $D=A*B$ и $D=c*A$ приводят к вычислению матрицы D , являющейся соответственно суммой, разностью, произведением матриц A и B , а также произведением матрицы A на число c .

Матрицу B , транспонированную к матрице A , получают командой $B=A'$. Например, имея вектор-строку $b=[14\ 9\ -12]$, можно найти квадрат длины (нормы) вектора при помощи команды $r=b*b'$, а длину (норму) вектора – командой $\text{sqrt}(b*b')$, где функция sqrt означает извлечение квадратного корня. Отметим, что длину вектора b можно найти и непосредственно командой $\text{norm}(b)$.

Используются также следующие символы: «/» (слэш) или «\» (обратный слэш) для деления матриц. Например, $A \setminus B$ означает левое деление матрицы B на матрицу A , соответствующее математическому произведению $A^{-1}B$, где A^{-1} – обратная матрица. Запись A/B означает правое деление матрицы A на матрицу B , что соответствует произведению AB^{-1} .

Использование деления вместо умножения на обратную матрицу имеет преимущества по точности и времени выполнения, так как применяет метод исключения Гаусса без явного обращения матриц.

Возможны поэлементные операции над матрицами одинакового размера с помощью использования точки перед символами умножения или деления. В частности, команда $R=A.*B$ приводит к вычислению матрицы R , каждый элемент которой представляет собой произведение элементов матриц A и B с

одинаковыми индексами строки и столбца. Например, если $A = \begin{pmatrix} -1 & 2 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 3 \\ -2 & 1 \end{pmatrix}$, то

$A.*B = \begin{pmatrix} -2 & 6 \\ -6 & 4 \end{pmatrix}$. Аналогично вычисляется $R=A./B$, т.е. каждый элемент матрицы R определяется равенством $r(i, j) = a(i, j)/b(i, j)$.

Возведение в степень каждого элемента матрицы A выполняется командой «.^». Например, команда $R=A.^2$ осуществляет возведение в квадрат каждого элемента матрицы A .

Символ «:» в применении к матрицам имеет следующий смысл: $A(:, j)$ означает j -й столбец матрицы A , а $A(i, :)$ – i -ю строку этой матрицы. Например, команда $A(2, :) = a$ поместит на место второй строки матрицы A вектор-строку a . (*Внимание!* Количества элементов в векторе a и строке матрицы A должны совпадать!)

В применении к индексам запись $j:k$ означает, что будут выбраны только объекты (в качестве объектов могут выступать, например, компоненты вектора) с индексами $j, j+1, \dots, k$. Так, если имеется вектор $b=[7\ 8\ 5\ 4\ 6\ 9\ 1]$, то командой $r = b(2:5)$ получим вектор $r=[8\ 5\ 4\ 6]$.

Удалить i -ю строку матрицы A размера $m \times n$ можно командой $A(i, :) = []$. После выполнения этой команды получается матрица размера $(m-1) \times n$. Команда $A(:, j) = []$ удаляет j -й столбец, и матрица размера $m \times n$ становится матрицей размера $m \times (n-1)$.

Если A – матрица размера $m \times n$, b – вектор-столбец размера $m \times 1$, то с помощью команды $B=[A\ b]$ можно добавить к матрице A справа столбец b , получится матрица B размера $m \times (n+1)$.

Matlab допускает максимум 4096 символов в строке. Если для выражения не хватает одной строки или нет желания заходить в невидимую область окна, то выражение можно перенести на новую строку с помощью многоточия ... (три идущие подряд точки).

Пример 2.2. `>> x=2;`

```
>> y=2;
>> r=sqrt(x^2+ ...
y^2)
r=
2.8284
```

2.3. Создание, редактирование и запуск m-файла

В Matlab имеется возможность написать программу на языке операторов и функций системы, сохранить её на диске в виде файла с расширением .m (называемым m-файлом) с целью последующего многократного выполнения. Для этого после запуска Matlab необходимо выбрать на панели инструментов (нажатием левой клавиши мыши) «File», затем «New», далее «M-File». После этого появляется окно «Editor» редактора-отладчика системы Matlab, в котором необходимо набрать нужные команды программы и сохранить полученный файл на диске (с помощью нажатия на значок «Save» панели окна «Editor»). Удобно также пользоваться нажатием на значок «Run» панели окна «Editor», в этом случае происходит одновременный запуск m-файла и его сохранение на диске.

Важно отметить, что в названии m-файла возможно использование только латинских букв и арабских цифр (например, Sidorov_1). Для выполнения m-файла необходимо набрать в окне команд правее символа >> его имя (без расширения .m) и нажать клавишу «Enter».

2.4. Некоторые функции и операторы системы Matlab

В данном разделе в алфавитном порядке приведены основные функции и операторы системы Matlab, которые могут пригодиться при выполнении студентами лабораторных работ.

% Знак процента является концом интерпретируемой строки, содержащей исполняемые выражения, любой последующий текст игнорируется. Используется для создания текстовых комментариев внутри m-файлов.

^ Оператор возведения в степень. Если задано $Y=X^p$, то это означает X в степени p (где p принадлежит к скалярному типу). Поэлементное возведение в степень достигается с помощью операции $Y=X.^p$.

= Оператор двойного знака равенства. Используется как оператор сравнения на равенство в операциях отношения, когда нельзя применять знак =, поскольку он отведён под указание операции присваивания. См. пример в описании оператора if.

< Оператор отношения «меньше чем».

<= Оператор отношения «меньше или равно».

> Оператор отношения «больше чем».

>= Оператор отношения «больше или равно».

~= Оператор отношения «не равно».

Функция abs. Если x – число, то $abs(x)$ вычисляет абсолютное значение (модуль) этого числа. Если X – матрица, то $abs(X)$ создает матрицу такого же размера, элементы которой имеют значения, равные абсолютным значениям соответствующих элементов исходной матрицы.

Функция acos. Используется в виде $acos(X)$, где X – матрица или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения арккосинуса соответствующих элементов матрицы или вектора X .

Функция asin. Используется в виде $asin(X)$, где X – матрица или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения арксинуса соответствующих элементов матрицы или вектора X .

Функция atan. Используется в виде $atan(X)$, где X – матрица или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения арктангенса соответствующих элементов матрицы или вектора X .

Оператор axis. Отменяет режим автомасштабирования графиков и с помощью своих параметров обеспечивает принудительное задание масштабов по осям x и y .

Оператор $axis(V)$, где V – вектор из четырёх компонент, устанавливает заданные пользователем масштабы по осям x и y . Например, последовательность команд:

```
V=[1 9 -12 4];
axis(V)
```

возвращает вектор [1 9 -12 4], компоненты которого (по оси x : от 1 до 9; по оси y : от -12 до 4) определяют масштабы для последнего графического построения.

Оператор bar. Применяется для построения столбцовых диаграмм и гистограмм. См. также оператор hist.

Удобно использовать в форме `bar(X,Y,width,'color')`, здесь X – вектор, компоненты которого не должны повторяться, Y – вектор, имеющий такую же размерность, что и вектор X , `width` – ширина столбцов (по умолчанию ширина равна 0,8), `color` – цвет столбцов. Например, команда `bar(X,Y,0.6,'g')` выдаёт графическое изображение значений компонент $y(i)$ вектора Y в виде столбцов зелёного цвета ширины 0,6, середины оснований которых находятся в точках $x(i)$, где $x(i)$ – компоненты вектора X . Для более подробного ознакомления с возможностями оператора `bar` можно воспользоваться оператором `help`.

Оператор `break`. Используется для принудительного прерывания выполнения цикла. Цикл прекращается досрочно, как только `break` встречается в теле цикла. Обычно `break` используется совместно с логическими условиями, управляющими ходом работы программы.

Оператор `contour`. Используется для построения линий уровня для функции двух переменных $z = f(x, y)$. Команда `contour(x,y,z,n)` строит n линий уровня функции $z = f(x, y)$ по данным матрицы z с указанием спецификаций для массивов x и y . Массивы x и y можно получить с помощью команды `meshgrid`. Например, последовательность команд

```
[x,y]=meshgrid(-2:0.1:2,-2:0.1:2);
z=x.^2+y.^2;
contour(x,y,z,5);
grid;
```

позволяет построить график пяти линий уровня функции $z = x^2 + y^2$ с нанесением на рисунок сетки, при этом аргументы x и y изменяются от -2 до 2 с шагом $0,1$.

Оператор `corrcoef`. Команда `R=corrcoef(A)` возвращает корреляционную матрицу R (симметрическую матрицу, в которой на главной диагонали стоят единицы, а вне главной диагонали – выборочные коэффициенты корреляции столбцов матрицы A). Если X и Y – вектор-столбцы одинакового размера, то команда `corrcoef(X,Y)` возвращает корреляционную матрицу 2×2 .

Функция `cos`. Используется в виде `cos(X)`, где X – матрица или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения косинуса соответствующих элементов матрицы или вектора X .

Оператор `cov`. Команда `K=cov(A)` возвращает ковариационную матрицу R (симметрическую матрицу, в которой на главной диагонали стоят несмещённые оценки дисперсий каждого из столбцов матрицы A , а вне главной диагонали – несмещённые оценки ковариации столбцов матрицы A). Если X и Y – вектор-столбцы одинакового размера, то команда `cov(X,Y)` возвращает ковариационную матрицу 2×2 .

Оператор `det`. Если A – квадратная матрица, то команда `det(A)` возвращает определитель матрицы A .

Оператор `diag`. Используется для формирования диагональных матриц.

Если b – n -мерный вектор-столбец или вектор-строка, то команда `R=diag(b)` создаёт диагональную матрицу R размера $n \times n$ с компонентами вектора b на главной диагонали.

Если X – квадратная матрица n -го порядка, то команда `a=diag(X)` даёт вектор a , компонентами которого являются элементы главной диагонали матрицы X . Команда `A=diag(diag(X))` создаёт диагональную матрицу A .

Оператор `disp`. Используется в виде `disp(X)`. Отображает на экране монитора матрицу X без вывода её имени. Если вектор X содержит текст (в апострофах), то на экран выводится этот текст. Например, `disp('Привет')` выводит на экран сообщение: Привет.

Оператор `eig`. Возвращает собственные значения и собственные векторы матрицы X .

Команда `a=eig(X)` даёт вектор a , содержащий собственные значения квадратной матрицы X .

Выражение `[V,D]=eig(X)` задаёт диагональную матрицу D , содержащую собственные значения матрицы X , и матрицу V , столбцы которой образованы соответствующими собственными векторами матрицы X , причём $X*V=V*D$.

Оператор `else`. Оператор условного выражения, используемый вместе с оператором `if`. См. `if` и `elseif`.

Оператор `elseif`. Оператор условного выражения, используемый вместе с оператором `if`. См. `if` и `else`.

Оператор `end`. Указывает на конец действия операторов `for`, `while` и `if`. Без `end` операторы `for`, `while` и `if` будут распространять своё поле действия на последующие операторы и функции. Каждый оператор `end` связан с предыдущим оператором `for`, `while` или `if` и закрывает его поле действия.

Функция `exp`. Используется в виде `exp(X)`, где X – матрица или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения числа e в степени x соответствующих элементов матрицы или вектора X .

Оператор `eye`. Команда `R=eye(n)` создаёт единичную матрицу R размера $n \times n$.

Оператор `for`. Задаёт цикл в формате:

```
for <переменная>=<выражение>;
... Операторы ...
end;
```

Пример 2.3 (Задание двойного цикла).

```
for I=1:N;
    for J=1:N;
```

```

        A(I,J)=1/(I+J-1);
    end;
end;

```

Оператор grid. Создаёт в поле вывода графиков координатную сетку из штрихпунктирных линий. Его рекомендуется использовать при одновременном выводе ряда графиков после построения последнего графика. См. axis, hold и plot.

Оператор hist. Используется для получения и отображения гистограммы (см. оператор bar), а также для получения данных для гистограммы. Для получения данных для гистограммы оператор hist записывается в следующем виде:

$N = \text{hist}(X, M)$, возвращает вектор чисел попаданий вектора X в M интервалов группировки одинаковой длины (если X – матрица, то hist работает со столбцами);

$N = \text{hist}(X, Y)$, где Y – вектор, возвращает количество элементов вектора X , попавших в интервалы с центрами, заданными вектором Y . Число интервалов в этом случае равно числу элементов вектора Y ;

$[N, Y] = \text{hist}(X, M)$, возвращает числа попаданий N вектора X в M интервалов группировки и данные о центрах интервалов Y (аналог статистического ряда).

Для построения гистограммы используется команда hist(X,M). Отображает числа попаданий значений элементов вектора X в M интервалов группировки в виде столбцовой диаграммы, т.е. строит прямоугольники высотой $h_j = n_j$, где n_j – число элементов, попавших в j -й интервал, $j = 1, 2, \dots, M$.

Оператор hold. Обеспечивает наложение одного графика на другой, что позволяет отображать на экране одновременно ряд кривых.

Оператор hold on обеспечивает построения с помощью оператора plot на фоне уже существующего изображения, используя при этом установленные первоначально масштабы по осям x и y графика (см. оператор axis). Оператор hold off выключает этот режим.

Пример 2.4 (Построение семейства кривых).

```

V=[0 5 0 1];
axis(V);                % Установка масштабов по осям.
plot(5,1)              % Построение точки (5; 1) и осей.
grid;
hold on;               % Установка режима наложения.
for i=1:10;
    x=5:-0.1:0;        % Построение десяти графиков
    plot(x,1-exp(-x./(i/2)), 'b'); % функций с наложением и
end;                   % пунктирной сеткой.
hold off;              % Отмена режима наложения.

```

Заметим, что программа должна начинаться с установки масштаба изображения с помощью оператора axis и построения точки с координатами (5; 1). Последнее необходимо для вывода координатных осей с оцифровкой. При построении последующих кривых они располагаются в созданной координатной системе и масштаб графика уже не меняется.

Оператор input. Обеспечивает ввод данных с клавиатуры. Команда $r = \text{input}(\text{'Сколько Вам лет?'})$ выводит на экран данный вопрос и ожидает ввода с клавиатуры любого Matlab-выражения, которое вычисляется. После ввода выражения и нажатия клавиши «Enter» результат возвращается в r . Команда $r = \text{input}(\text{'Введите текст','s'})$ даёт приглашение в текстовой строке и ожидает ввода символьной строки. Напечатанный текст не вычисляется; символы просто возвращаются как Matlab-строка.

```

Пример 2.5. r=input('Введите угол в радианах:')
Введите угол в радианах:  2*pi
r=6.2832
r=input('Введите угол в радианах:','s')
Введите угол в радианах:  2*pi
r=2*pi

```

Оператор inv. Если A – квадратная матрица, то команда inv(A) возвращает обратную матрицу для матрицы A .

Оператор if. Задаёт цикл с условными выражениями. Простая форма:

```

if <Условие>
... Инструкции ...
end;

```

В условиях используются следующие логические операторы: $=$, $<$, $>$, $<=$, $>=$, \sim . Двойные символы не имеют между собой пробелов.

Более полная конструкция цикла имеет вид:

```

if <Условие>
... Инструкции 1 ...

```

```

else
... Инструкции 2 ...
end;

```

В этом варианте выполняется Инструкция 1, если выполняется <Условие>, или Инструкция 2 в противном случае.

Возможно и применение оператора elseif.

Пример 2.6. if I = J
A(I, J)=2;
elseif abs(I - J) = 1
A(I, J) = -1;
else
A(I, J) = 0;
end;

Функция log. Используется в виде log(X), где X – матриц или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения натурального логарифма соответствующих элементов матрицы или вектора X.

Оператор max. Используется в виде max(X). Для векторов: возвращает наибольший элемент вектора X. Для матриц: возвращает вектор-строку, состоящую из максимальных элементов каждого столбца матрицы X.

Оператор mean. Используется в виде mean(X). Возвращает среднее арифметическое значений компонент вектора X.

Оператор median. Используется в виде median(X). Возвращает медиану вектора X.

Оператор mode. Используется в виде mode(X). Возвращает моду вектора X.

Оператор mesh. Служит для построения сетчатой поверхности, представляющей трёхмерную фигуру, задаваемую функцией двух переменных $z = f(x, y)$. См. также plot3.

Команда mesh(X,Y,Z) вычерчивает различные строки, полученные из столбцов двумерных массивов X, Y, Z одинаковой размерности. Цвет линий сетчатого графика определяется высотой поверхности. Массивы X и Y можно получить с помощью команды meshgrid. Например, последовательность команд

```

[X,Y]=meshgrid(-2:0.1:2,-2:0.1:2);
Z=X.^2+Y.^2;
mesh(X,Y,Z);

```

позволяет построить график функции $z = x^2 + y^2$, при этом аргументы x и y изменяются от -2 до 2 с шагом 0,1.

Функция meshgrid. Трёхмерные поверхности обычно описываются функцией двух переменных $z = f(x, y)$. Специфика построения трёхмерных графиков требует не просто задания ряда значений x и y, т.е. векторов x и y, а определения двумерных массивов X и Y. Для создания таких массивов служит функция meshgrid.

Команда [X,Y]=meshgrid(x,y) преобразует область, заданную векторами x и y, в двумерные массивы X и Y, которые могут быть использованы для вычисления значений функций двух переменных и построения трёхмерных графиков. Эта функция формирует массивы X и Y таким образом, что строки выходного массива X являются копиями вектора x, а столбцы выходного массива Y – копиями вектора y.

Оператор min. Используется в виде min(X). Для векторов: возвращает наименьший элемент вектора X. Для матриц: возвращает вектор-строку, состоящую из минимальных элементов каждого столбца матрицы X.

Оператор norm. Команда norm(b) вычисляет длину (норму) вектора b.

Оператор ones. Команда R=ones(n) создаёт матрицу R размера $n \times n$, состоящую из единиц. Команда R=ones(m,n) создаёт матрицу единиц размера $m \times n$. Команда R=ones(A) образует матрицу R единиц такого же размера, как и матрица A.

Оператор pause. В форме pause прекращает ход вычислительного процесса и ожидает нажатия любой клавиши для продолжения. В форме pause(n) создаёт паузу в n секунд.

Константа pi. Системная константа pi=3.1415926535897...

Оператор plot. Используется для построения графиков по данным вектора или матрицы. Команда plot(X,Y) строит по точкам (x; y) график функции $y(x)$, где x и y – элементы векторов X и Y. Перед применением команды необходимо задать векторы X и Y независимой переменной x и значений функции y. Эти векторы должны быть одинаковой размерности. Например, последовательность команд

```

x = -2:0.1:3;      % Вычисляет значения аргумента с шагом 0,1.
y = sin(x);       % Вычисляет синусы для компонент вектора x.
plot(x,y)

```

выводит на экран график синуса на промежутке [-2; 3].

Команда `plot(X1,Y1,X2,Y2)` даёт другой способ создания линий в поле вывода, при котором строятся графики двух функций: $y_1(x_1)$ и $y_2(x_2)$. Так можно задать построение и большего числа функций.

Команда `plot(X1,Y1,'g-',X2,Y2,'r+')` используется для построения штрихпунктирной линии зелёного цвета для первой кривой и символа + красного цвета для построения узловых точек для второй кривой.

Таким образом, специальные указания в апострофах позволяют задавать дополнительные указания о цвете и типе узловых точек или линий. Цвет линии или узловой точки определяется символами: **y** (жёлтый), **m** (фиолетовый), **c** (голубой), **r** (красный), **g** (зелёный), **b** (синий), **w** (белый), **k** (чёрный). Тип узловой точки определяется символами: **.** (точка), **o** (окружность), **x** (крестик), **+** (плюс), ***** (звёздочка), **s** (квадрат), **d** (ромб), **< ^ >** (треугольники различной направленности), **p** (пятиугольник), **h** (шестиугольник). Тип линии определяется символами: **-** (непрерывная), **:** (короткие штрихи), **-.** (штрихпунктир), **--** (длинные штрихи).

Оператор `plot(Y)` даёт построение графика по данным вектора **Y** относительно его индексов.

Оператор plot3. Оператор `plot3(...)` является трёхмерным аналогом оператора `plot(...)`. Служит для построения поверхности, задаваемой функцией двух переменных $z = f(x, y)$. См. также оператор `mesh`.

Команда `plot3(X,Y,Z)` вычерчивает различные строки, полученные из столбцов двумерных массивов **X**, **Y**, **Z** одинаковой размерности. Массивы **X** и **Y** можно получить с помощью команды `meshgrid`. Например, последовательность команд

```
[X,Y]=meshgrid(-2:0.1:2,-2:0.1:2);
Z=X.^2+Y.^2;
plot3(X,Y,Z);
grid;
```

позволяет построить график функции $z = x^2 + y^2$, при этом аргументы x и y изменяются от -2 до 2 с шагом $0,1$.

Оператор prod. Если **X** – вектор, то команда `prod(X)` вычисляет произведение его компонент. Если **X** – матрица, то `prod(X)` – вектор-строка, компонентами которой являются произведения элементов каждого столбца матрицы **X**. Например, для матрицы $X = \begin{pmatrix} -1 & 2 \\ 3 & 4 \end{pmatrix}$ `prod(X) = (-3; 8)`.

Оператор rand. Используется для генерации случайных чисел и матриц. Команда `rand(n)` даёт $n \times n$ -матрицу, заполненную случайными числами. Команда `rand(m,n)` даёт $m \times n$ -матрицу, заполненную случайными числами. Команда `rand` без аргументов даёт одно случайное число. Случайные числа распределены равномерно в интервале $(0; 1)$.

Оператор rank. Команда `rank(A)` возвращает число – ранг матрицы **A**.

Функция sign. Для каждого элемента x матрицы **X** команда `sign(x)` возвращает 1, если значение аргумента больше нуля; 0, если его значение равно нулю; (-1) , если значение элемента меньше нуля.

Функция sin. Используется в виде `sin(X)`, где **X** – матрица или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения синуса соответствующих элементов матрицы или вектора **X**.

Оператор sort. В форме `sort(X)` сортирует вектор **X** по неубыванию. В случае, когда **X** – матрица, оператор `sort(X)` сортирует каждый столбец **X** в неубывающем порядке.

Команда `[Y,INDEX]=sort(X)` наряду с отсортированным вектором **Y** возвращает вектор индексов **INDEX**. С помощью этого массива можно восстановить структуру исходного вектора **X**.

Функция sqrt. Используется в виде `sqrt(X)`, где **X** – матрица или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения корня квадратного из соответствующих элементов матрицы или вектора **X**.

Оператор stairs. В форме `stairs(X)` строит лестничный (ступенчатый) график по значениям элементов вектора **X**.

Функция `stairs(Z,SF)` строит лестничный график по значениям элементов вектора **SF** в точках скачков, определённых в **Z**. Значения **Z** должны располагаться в возрастающем порядке. Используется для получения и графического отображения эмпирической функции распределения. В этом случае **Z** – вектор различных значений элементов для негруппированной выборки или вектор середин интервалов для группированной выборки, а **SF** – вектор накопленных относительных частот выборки.

Оператор std. В форме `std(X)` или `std(X,0)` возвращает несмещённую оценку $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ стандартного отклонения вектора **X**. В форме `std(X,1)` возвращает смещённую оценку $\tilde{\sigma}_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ стандартного отклонения вектора **X**.

Оператор sum. Если X – вектор, то команда $\text{sum}(X)$ вычисляет сумму его компонент. Если X – матрица, то $\text{sum}(X)$ – вектор-строка, компонентами которой являются суммы элементов каждого столбца матрицы X . Например, для матрицы $X = \begin{pmatrix} -1 & 2 \\ 3 & 4 \end{pmatrix}$ $\text{sum}(X) = (2; 6)$.

Функция tan. Используется в виде $\text{tan}(X)$, где X – матрица или вектор. Возвращает аналогичные по размерам матрицу или вектор, элементы которых имеют значения тангенса соответствующих элементов матрицы или вектора X .

Оператор tril. Команда $\text{tril}(A)$ для квадратной матрицы A возвращает нижнюю треугольную матрицу.

Оператор var. В форме $\text{var}(X)$ или $\text{var}(X,0)$ возвращает несмещённую оценку дисперсии вектора X . В форме $\text{var}(X,1)$ возвращает смещённую оценку дисперсии вектора X .

Оператор while. Задаёт цикл. Используется следующим образом:

```
while <Условие>
... Инструкции ...
end;
```

Цикл типа `while` выполняется до тех пор, пока выполняется `<Условие>`. Для прекращения выполнения цикла можно использовать оператор `break`.

Пример 2.7. $X = -2:0.5:3$; $N = 1$;

```
while X(N)<0;
    X(N) = -X(N);
    N = N + 1;
end;
```

Оператор zeros. Команда $R = \text{zeros}(n)$ создаёт матрицу R размера $n \times n$, состоящую из нулей. Команда $R = \text{zeros}(m,n)$ создаёт матрицу нулей размера $m \times n$. Команда $R = \text{zeros}(A)$ образует матрицу R нулей такого же размера, как и матрица A .

2.5. Средства Matlab для изучения одномерных распределений

В пакете **stats** статистических программ Matlab (каталог\Matlab\toolbox\stats) имеются программы расчёта плотностей вероятности и функций распределения для многих известных распределений. Имена функций для расчёта плотностей вероятности оканчиваются буквами **pdf** (probability density function), а для расчёта функций распределения – буквами **cdf** (cumulative distribution function).

2.5.1. Функции Matlab для расчёта плотностей вероятности

$y = \text{unifpdf}(x,a,b)$ – расчёт значения плотности вероятности в точке x для равномерного на промежутке $(a; b)$ распределения.

$y = \text{normpdf}(x,m,\text{sigma})$ – расчёт значения плотности вероятности в точке x для нормального распределения, где m – математическое ожидание, sigma – среднее квадратическое отклонение.

$y = \text{gampdf}(x,a,b)$ – расчёт значения плотности вероятности в точке x для гамма-распределения с параметрами a, b .

$y = \text{expdf}(x,\text{lambda})$ – расчёт значения плотности вероятности в точке x для экспоненциального распределения с параметром lambda , равным *математическому ожиданию* (!) случайной величины.

$y = \text{chi2pdf}(x,k)$ – расчёт значения плотности вероятности в точке x для распределения χ^2 с k степенями свободы.

$y = \text{tpdf}(x,k)$ – расчёт значения плотности вероятности в точке x для распределения Стьюдента с k степенями свободы.

$y = \text{fpdf}(x,k1,k2)$ – расчёт значения плотности вероятности в точке x для распределения Фишера с $k1, k2$ степенями свободы.

2.5.2. Функции Matlab для расчёта функций распределения

$y = \text{unifcdf}(x,a,b)$ – расчёт значения функции распределения в точке x для равномерного на промежутке $(a; b)$ распределения.

$y = \text{normcdf}(x,m,\text{sigma})$ – расчёт значения функции распределения в точке x для нормального распределения, где m – математическое ожидание, sigma – среднее квадратическое отклонение.

$y=\text{gamcdf}(x,a,b)$ – расчёт значения функции распределения в точке x для гамма-распределения с параметрами a, b .

$y=\text{expcdf}(x,\lambda)$ – расчёт значения функции распределения в точке x для экспоненциального распределения с параметром λ , равным *математическому ожиданию* (!) случайной величины.

$y=\text{chi2cdf}(x,k)$ – расчёт значения функции распределения в точке x для распределения χ^2 с k степенями свободы.

$y=\text{tcdf}(x,k)$ – расчёт значения функции распределения в точке x для распределения Стьюдента с k степенями свободы.

$y=\text{fcdf}(x,k1,k2)$ – расчёт значения функции распределения в точке x для распределения Фишера с $k1, k2$ степенями свободы.

Замечание. Для расчёта значений гамма-функции в точке x в Matlab имеется функция $y=\text{gamma}(x)$.

2.6. Алгоритмы моделирования случайных величин

Случайные числа с различными законами распределения обычно моделируются с помощью преобразований одного или нескольких независимых значений *базовой случайной величины*. Базовая случайная величина α – это случайная величина с распределением $\mathbf{R}(0; 1)$ (равномерным распределением в интервале $(0; 1)$). Независимые случайные величины с распределением $\mathbf{R}(0; 1)$ обозначаются символами $\alpha_1, \alpha_2, \dots$. В любой системе программирования имеется стандартная программа моделирования базовой случайной величины (см. раздел 2.4, оператор `rand`).

Рассмотрим алгоритмы моделирования случайных величин, имеющих законы распределения, описанные в разделе 1 пособия.

Нормальное распределение $\mathbf{N}(m; \sigma)$

<p><i>Алгоритм моделирования 1:</i></p> <ol style="list-style-type: none"> 1) зарезервировать константу $c = 2\pi$; 2) $r = \sqrt{-2 \ln \alpha_1}$; 3) $\varphi = c\alpha_2$; 4) $X_1 = r \cos \varphi, X_2 = r \sin \varphi$; 5) $X_1 = m + \sigma X_1, X_2 = m + \sigma X_2$. 	<p><i>Алгоритм моделирования 2:</i></p> <ol style="list-style-type: none"> 1) $\xi_1 = 2\alpha_1 - 1, \xi_2 = 2\alpha_2 - 1$; 2) $s = \xi_1^2 + \xi_2^2$; 3) если $s \geq 1$, вернуться к п. 1; 4) $r = \sqrt{-(2 \ln s)/s}$; 5) $X_1 = \xi_1 r, X_2 = \xi_2 r$; 6) $X_1 = m + \sigma X_1, X_2 = m + \sigma X_2$.
--	---

В качестве случайного числа можно взять любое из чисел X_1 и X_2 .

Равномерное распределение $\mathbf{R}(a; b)$ ($a < b$)

Алгоритм моделирования:
 $X = a + (b - a)\alpha$.

Экспоненциальное распределение $\mathbf{E}(\lambda)$ ($\lambda > 0$)

Алгоритм моделирования:
 $X = -\lambda^{-1} \ln \alpha$.

Распределение χ^2 с k степенями свободы $\chi^2(k)$

Алгоритм моделирования:

$$X = \sum_{i=1}^k u_i^2,$$

где $u_i \sim \mathbf{N}(0; 1)$ – независимые случайные величины.

Распределение Стьюдента с k степенями свободы $\mathbf{St}(k)$

Алгоритм моделирования:

$$X = \frac{u}{\sqrt{v/k}},$$

где $u \sim \mathbf{N}(0; 1), v \sim \chi^2(k)$ – независимые случайные величины.

Распределение Фишера с k_1, k_2 степенями свободы $F(k_1; k_2)$

Алгоритм моделирования:

$$X = \frac{v/k_1}{w/k_2},$$

где $v \sim \chi^2(k_1)$, $w \sim \chi^2(k_2)$ – независимые случайные величины.

Замечание 1. Описанные алгоритмы позволяют получить только одно случайное число. Для получения массивов объёмом 300 используйте цикл `for` (см. раздел 2.4).

Замечание 2. В Matlab (см. раздел 2.4) имеется системная константа π : `pi=3.1415926535897...`

Натуральный логарифм числа x можно получить с помощью оператора `log`. Синус (или косинус) числа x можно получить с помощью операторов `sin` (или `cos`).

2.7. Средства Matlab для моделирования случайных величин

В пакете `stats` статистических программ Matlab (каталог\Matlab\toolbox\stats) имеются программы моделирования случайных чисел из многих известных законов распределения. Имена указанных функций оканчиваются буквами `rnd`.

`y=unifrnd(a,b)` – возвращает случайное число из равномерного на промежутке $(a; b)$ распределения.

`y=normrnd(m,sigma)` – возвращает случайное число из нормального распределения, где m – математическое ожидание, σ – среднее квадратическое отклонение.

`y=gamrnd(a,b)` – возвращает случайное число из гамма-распределения с параметрами a, b .

`y=expnrnd(lambda)` – возвращает случайное число из экспоненциального распределения с параметром λ , равным математическому ожиданию случайной величины.

`y=chi2rnd(k)` – возвращает случайное число из распределения χ^2 с k степенями свободы.

`y=trnd(k)` – возвращает случайное число из распределения Стьюдента с k степенями свободы.

`y=frnd(k1,k2)` – возвращает случайное число из распределения Фишера с k_1, k_2 степенями свободы.

Замечание. В Matlab имеется программа для моделирования многомерных случайных чисел с нормальным распределением.

`r=mvnrnd(m,sigma,N)` – возвращает матрицу случайных чисел, выбранных из многомерного нормального распределения с вектором средних m и ковариационной матрицей σ . Параметр N является количеством строк в r (количеством многомерных случайных чисел).

2.8. Средства Matlab для получения выборочных числовых характеристик и интервальных оценок

В пакете `stats` статистических программ Matlab (каталог\Matlab\toolbox\stats) имеются программы расчёта выборочных числовых характеристик, а также получения интервальных оценок параметров распределений. Имена указанных функций оканчиваются буквами `fit`.

`normfit(X)` – возвращает оценку математического ожидания для нормально распределённой генеральной совокупности, полученную по выборке X .

`[m,sigma,m_int,sigma_int]=normfit(X,alpha)` – возвращает несмещённые оценки и $100 \cdot (1 - \alpha)$ -процентные доверительные интервалы для параметров нормально распределённой генеральной совокупности (математическое ожидание и среднее квадратическое отклонение) по выборке X .

`expfit(X)` – возвращает несмещённую оценку математического ожидания экспоненциального распределения, полученную по выборке X методом максимального правдоподобия.

`[m,m_int]=expfit(X,alpha)` – возвращает максимально правдоподобную оценку и $100 \cdot (1 - \alpha)$ -процентный доверительный интервал для математического ожидания экспоненциального распределения.

`gamfit(X)` – возвращает оценки параметров гамма-распределения, полученные по выборке X методом максимального правдоподобия.

`[par,par_int]=gamfit(X,alpha)` – возвращает максимально правдоподобные оценки и $100 \cdot (1 - \alpha)$ -процентные доверительные интервалы для параметров гамма-распределения.

`unifit(X)` – возвращает оценку параметра a равномерного распределения на $(a; b)$, полученную по выборке X методом максимального правдоподобия.

`[a,b,a_int,b_int]=unifit(X,alpha)` – возвращает максимально правдоподобные оценки и $100 \cdot (1 - \alpha)$ -процентные доверительные интервалы для параметров равномерного распределения.

Замечание. По умолчанию необязательный параметр `alpha=0,05`, что соответствует 95-процентным доверительным интервалам.

2.9. Средства Matlab для нахождения квантилей Распределений

В пакете **stats** статистических программ Matlab (каталог\Matlab\toolbox\stats) имеются программы расчёта квантилей для многих известных распределений. Имена функций для расчёта квантилей оканчиваются буквами **inv**.

x=norminv(p,m,sigma) – возвращает значение квантили порядка p для нормального распределения с математическим ожиданием m и средним квадратическим отклонением σ .

x=chi2inv(p,k) – возвращает значение квантили порядка p для распределения χ^2 с k степенями свободы.

x=tinv(p,k) – возвращает значение квантили порядка p для распределения Стьюдента с k степенями свободы.

x=finv(p,k1,k2) – возвращает значение квантили порядка p для распределения Фишера с k_1, k_2 степенями свободы.

3. Описание лабораторных работ

Лабораторная работа № 1. Статистическая обработка одномерной выборки

Цель работы: 1) приобретение практических навыков по статистической обработке результатов наблюдений одномерной случайной величины; 2) изучение возможностей системы Matlab по статистической обработке результатов наблюдений одномерной случайной величины.

Содержание и порядок выполнения работы

1. Получение допуска к работе. Необходимо переписать данные своего варианта N (см. приложение 1, выборка объёмом 50) и выполнить дома вручную в отдельной тетради следующие задания (надо **обязательно учесть** приведённое ниже замечание, касающееся особенностей пакета Matlab, и **выполнить** расчёты так, чтобы не было расхождений между результатами, полученными вручную и на компьютере с помощью Matlab):

- построить вариационный и статистический ряды, найти размах выборки;
- построить таблицу частот группированной выборки, сделав семь интервалов **одинаковой** длины;
- провести статистическую обработку результатов наблюдений, включающую построение (для интервально группированной выборки) полигона и гистограммы частот, эмпирической функции распределения;
- найти оценки математического ожидания, дисперсии (смещённую и несмещённую), медианы и моды для исходной (негруппированной) и интервально группированной выборок.

Замечание. Matlab в силу отличий в определении функции распределения элементы выборки, попадающие на границу интервала, включает в соседний слева интервал группировки; кроме того, в первый интервал группировки Matlab включает элементы выборки, попадающие на его левую границу.

2. Изучение возможностей системы Matlab. Требуется научиться создавать, редактировать и запускать m-файл, проводить первичную статистическую обработку одномерной выборки при помощи пакета Matlab. Для этого необходимо (см. раздел 2):

- ввести в созданном m-файле данные своего варианта в виде вектор-строки (или вектор-столбца);
- дополнить m-файл операторами построения вариационного и статистического рядов (см. sort);
- дополнить m-файл операторами построения гистограмм и полигонов частот для интервально группированной выборки (см. bar, hist, plot), совместить в одном графическом окне графики соответствующих полигонов и гистограмм (см. axis, hold);
- дополнить m-файл операторами построения графиков эмпирической функции распределения для исходной (негруппированной) и интервально группированной выборок (см. stairs);
- дополнить m-файл операторами вывода на экран числовых характеристик для исходной (негруппированной) и интервально группированной выборок: размах, мода, медиана, выборочное среднее, дисперсия (см. min, max, mode, median, mean, std, var);
- запустить m-файл (не забывайте при этом об операторе pause), убедиться в совпадении результатов, полученных дома вручную и в классе на компьютере, отчитаться перед преподавателем за проделанную работу.

Требования к отчёту

Отчёт, представленный в отдельной тетради, должен содержать:

- 1) все требуемые таблицы, графики (гистограмму, полигон и эмпирическую функцию распределения);
- 2) результаты расчётов числовых характеристик выборок.

Контрольные вопросы

1. Дайте определения:

- вариационного ряда;
- статистического ряда;
- гистограммы;
- полигона;
- эмпирической функции распределения;
- выборочной числовой характеристики (среднее, дисперсия, мода, медиана).

2. Каковы требования, предъявляемые к точечным оценкам?

Лабораторная работа № 2. Моделирование случайной величины с заданным законом распределения

Цель работы: 1) ознакомление с графическими возможностями системы Matlab по изучению основных законов распределений одномерных случайных величин, визуализированное исследование зависимости графиков функций распределения и функций плотности вероятности от параметров распределений; 2) изучение возможностей системы Matlab по моделированию одномерных случайных величин.

Содержание и порядок выполнения работы

1. Получение допуска к работе. Необходимо выполнить дома следующие задания:

– для каждого из следующих трёх видов распределений (параметры распределений взять из п. 3 данной лабораторной работы):

нормального $\mathbf{N}(m; \sigma)$ ($\sigma > 0$),

равномерного $\mathbf{R}(a; b)$ ($a < b$),

экспоненциального $\mathbf{E}(\lambda)$ ($\lambda > 0$)

написать программу вывода в графическом окне по два графика функций плотности вероятности, причём один из этих графиков получить по собственной программе, написанной для расчёта значений функции плотности вероятности по формулам раздела 1.1, другой – с использованием функций системы Matlab (см. раздел 2.5);

– для каждого из следующих трёх видов распределений (параметры распределений взять из п. 3 данной лабораторной работы):

нормального $\mathbf{N}(m; \sigma)$ ($\sigma > 0$),

равномерного $\mathbf{R}(a; b)$ ($a < b$),

экспоненциального $\mathbf{E}(\lambda)$ ($\lambda > 0$)

написать программу вывода в графическом окне графиков функций распределения с использованием функций системы Matlab (см. раздел 2.5).

2. В компьютерном классе для каждого из следующих трёх видов распределений:

χ^2 с k степенями свободы,

Стьюдента с k степенями свободы,

Фишера с k_1, k_2 степенями свободы

вывести в графическом окне по два графика функций плотности вероятности, причём один из графиков получить по собственной программе, написанной для расчёта значений функции плотности вероятности по формулам раздела 1.1, другой – с использованием функций системы Matlab (см. раздел 2.5). По указанию преподавателя исследовать зависимость графиков от параметров распределений.

3. Написать программу, осуществляющую моделирование массивов объёмом 300 для законов распределения: нормального, равномерного, экспоненциального (см. раздел 2.6).

Параметры моделируемых массивов:

– для нормального распределения $\mathbf{N}(m; \sigma)$ ($\sigma > 0$):

$$m = \ll \text{номер варианта} \gg, \quad \sigma = \ll \sqrt{\text{номер варианта}} \gg;$$

– для равномерного распределения $\mathbf{R}(a; b)$ ($a < b$):

$$a = \ll \text{номер варианта} \gg, \quad b = 2a;$$

– для экспоненциального распределения $\mathbf{E}(\lambda)$ ($\lambda > 0$):

$$\lambda = \begin{cases} N/2, & N < 10, \\ N/5, & 10 \leq N \leq 20, \text{ где } N = \ll \text{номер варианта} \gg. \\ N/10, & 20 < N, \end{cases}$$

Для каждого из этих массивов вывести в графическом окне график гистограммы относительных частот группированной выборки (шесть – восемь интервалов группировки). Продемонстрировать полученные результаты преподавателю.

Замечание. Необходимо запомнить, как выглядят гистограммы выборок, полученных из генеральных совокупностей, имеющих нормальное, равномерное и экспоненциальное распределения. Это потребуется при выполнении лабораторной работы № 3.

Требования к отчёту

Отчёт, представленный на отдельных листах формата А4, должен содержать:

1) распечатку программы (с комментариями) по расчёту значений функций распределения, плотности вероятности и по моделированию массивов для нормального, равномерного и экспоненциального законов распределения;

2) графики функций распределения и плотности вероятности, а также графики гистограмм моделируемых массивов.

Контрольные вопросы

Дайте определения:

- закона распределения случайной величины;
- функции распределения;
- функции плотности вероятности;
- случайной величины непрерывного типа, имеющей распределение: равномерное, нормальное, гамма, экспоненциальное, χ^2 , Стьюдента, Фишера (необходимо уметь объяснять, от скольких параметров зависит и каков их смысл).

Лабораторная работа № 3. Проверка гипотезы о законе распределения генеральной совокупности

Цель работы: 1) изучение методов проверки параметрических гипотез с помощью доверительных интервалов и гипотез о законе распределения с помощью критерия χ^2 ; 2) получение навыков проверки параметрических гипотез с помощью доверительных интервалов в системе Matlab.

Содержание и порядок выполнения работы

1. Получение допуска к работе. Необходимо выполнить дома вручную следующие задания, оформив их в отдельной тетради (см. приложение 1, выборка объёмом 50 из лабораторной работы № 1, а также раздел 1.4):

– в предположении, что массив получен из нормально распределённой генеральной совокупности с неизвестными m и σ^2 , построить доверительные интервалы для математического ожидания и дисперсии при доверительной вероятности, равной 0,95 (рассмотреть случаи негруппированной и группированной выборок);

– проверить (с помощью построенных доверительных интервалов) гипотезы $H_0^{(1)} : m_X = M_0$ и $H_0^{(2)} : D_X = A_0$, где $M_0 = \bar{x} + 0,5s$, $A_0 = 2s^2$, и записать статистическое решение (принять или отклонить нулевую гипотезу);

– вычислить статистику χ^2 для проверки гипотезы о **нормальном** распределении генеральной совокупности (число интервалов группировки равно семи, уровень значимости $\alpha = 0,05$) и записать статистическое решение (принять или отклонить нулевую гипотезу).

2. Используя средства Matlab, для массива объёмом 50 из лабораторной работы № 1 (в предположении, что массив получен из нормально распределённой генеральной совокупности) осуществить вывод на экран границ доверительных интервалов и выборочных числовых характеристик (см. раздел 2.8).

3. Проанализировать три выборки объёмом 200 из генеральных совокупностей, имеющих экспоненциальное, нормальное и равномерное распределения (по одной выборке каждого из трёх указанных видов распределений).

Для этого в рабочем поле команд после символа `>>` ввести с клавиатуры команду `forlab3` (являющуюся специально написанной программой для данной лабораторной работы) и нажать клавишу «Enter». После появления приглашения

Введите номер Вашего варианта N, затем нажмите Enter: N=

надо набрать с клавиатуры номер своего варианта и нажать клавишу «Enter». На экране появятся три столбца, каждый из которых – это выборка объёмом 200. Необходимо скопировать данные этих трёх массивов в виде матрицы размером 200×3 в специально созданный m-файл.

Далее для каждой из этих выборок необходимо:

- найти выборочное среднее и несмещённую оценку дисперсии (эти числовые характеристики должны войти в отчёт по данной лабораторной работе);
- построить гистограммы относительных частот (сделать семь интервалов группировки); распечатать гистограммы (для помещения их в отчёт по данной лабораторной работе);
- получить таблицы частот, при этом можно воспользоваться оператором $[N, Y]=\text{hist}(X, M)$ (эти таблицы должны войти в отчёт по данной лабораторной работе).

Проанализировав данные по всем трём выборкам, выдвинуть гипотезы о законах распределения генеральных совокупностей, представленных этими массивами.

Требования к отчёту

Отчёт, представленный в отдельной тетради, должен содержать:

- 1) вычисления, касающиеся выборки объёмом 50 из лабораторной работы № 1, которые служили допуском к выполнению данной лабораторной работы;
- 2) таблицы частот, графики гистограмм, а также результаты расчётов по проверке гипотез о законах распределения по критерию согласия χ^2 для массивов, полученных с помощью программы forlab3.

Контрольные вопросы

1. Дайте определения:
 - доверительного интервала;
 - доверительной вероятности и уровня значимости;
 - квантилей используемых распределений.
2. Опишите принцип проверки статистических гипотез (простые и сложные гипотезы, параметрические или гипотезы о законе распределения; нулевая и альтернативная гипотезы; уровень значимости, критическая область; ошибки первого и второго родов; связь с доверительными интервалами).
3. Перечислите основные положения, связанные с использованием критерия согласия χ^2 для проверки гипотезы о законе распределения.

Лабораторная работа № 4. Анализ данных в линейной регрессионной модели

Цель работы: 1) приобретение практических навыков по статистической обработке результатов наблюдения двумерного случайного вектора, а также оценки зависимости между случайными величинами; 2) изучение и освоение некоторых процедур регрессионного анализа системы Matlab.

Содержание и порядок выполнения работы

1. Получение допуска к работе. Необходимо переписать данные своего варианта N (см. приложение 2, парная выборка $(x_i; y_i)$ объёмом 50 двумерного нормально распределённого случайного вектора) и выполнить дома вручную в отдельной тетради следующие задания (см. разделы 1.5 – 1.7):

– осуществить предварительную обработку результатов наблюдений, включающую построение диаграммы рассеивания; группировку данных и построение корреляционной таблицы; оценку числовых характеристик для негруппированных и группированных данных;

– для негруппированных данных проверить гипотезу $H_0: \rho_{X,Y} = 0$ об отсутствии линейной статистической связи между компонентами X и Y при альтернативной гипотезе $H_1: \rho_{X,Y} \neq 0$ (уровень значимости $\alpha = 0,05$);

– для негруппированных данных получить интервальную оценку для истинного значения коэффициента корреляции $\rho_{X,Y}$ при уровне значимости $\alpha = 0,05$;

– для негруппированных и группированных данных составить уравнения линейной регрессии Y на x и X на y ;

– для негруппированных данных нанести графики выборочных регрессионных прямых на диаграмму рассеивания;

– для негруппированных данных по найденным оценкам параметров линейной регрессии Y на x получить оценку s^2 для дисперсии ошибок наблюдений σ^2 , найти коэффициент детерминации R^2 , построить доверительные интервалы для параметров регрессии a и b , дисперсии ошибок наблюдений σ^2 и среднего значения Y при $x = x_0$;

– для негруппированных данных проверить значимость линейной регрессии Y на x (уровень значимости $\alpha = 0,05$);

– для данных, сгруппированных только по X , проверить адекватность линейной регрессии Y на x (уровень значимости $\alpha = 0,05$);

– для негруппированных данных проверить гипотезу $H_0: m_X = m_Y$ при альтернативной гипотезе $H_1: m_X \neq m_Y$ (уровень значимости $\alpha = 0,05$).

2. В компьютерном классе для защиты выполненной дома работы необходимо (см. раздел 2):

– ввести свой двумерный массив в специально созданный m-файл в виде матрицы A (два столбца, по столбцам – компоненты X и Y);

– сравнить результаты расчётов, проведённых дома вручную для получения допуска к данной лабораторной работе, с компьютерными оценками в системе Matlab (см. cov, corrcoef);

– построить диаграмму рассеивания и графики выборочных регрессионных прямых, используя режим наложения (см. hold, plot);

– в присутствии преподавателя проверить свои расчёты с помощью процедур обработки данных в графическом окне.

Замечание 1. В системе Matlab аппроксимацию и графическую визуализацию можно осуществлять прямо в окне редактора графики Property Editor. Для этого в позиции Tools графического окна имеются команды Basic Fitting (открывает окно с различными видами аппроксимации и регрессии) и Data Statistics (открывает окно с результатами простейшей статистической обработки данных). При этом выбор вида аппроксимации, показ уравнения аппроксимации (Show equations) и вывод тех или иных числовых параметров выборочного вектора осуществляются с помощью установки «птичек» в нужных полях. Тогда можно сразу наблюдать соответствующие автоматические перестроения на диаграмме рассеивания.

Замечание 2. Для детального ознакомления с возможностями Matlab по обработке данных методами регрессионного анализа можно воспользоваться командой help и изучить синтаксис оператора regress.

Требования к отчёту

Отчёт, представленный в отдельной тетради, должен содержать:

- 1) вычисления, касающиеся парной выборки $(x_i; y_i)$ объёмом 50 двумерного нормально распределённого случайного вектора, которые служили допуском к выполнению данной лабораторной работы (*должны быть приведены не только окончательные результаты, но и промежуточные выкладки*);
- 2) диаграмму рассеивания и прямые регрессии.

Контрольные вопросы

1. Дайте определения:
 - распределения двумерной выборки;
 - выборочных числовых характеристик (приведите формулы для их вычисления).
2. Объясните принципы:
 - получения оценок коэффициентов выборочной регрессии;
 - проверки значимости и адекватности линейной регрессии;
 - проверки статистических гипотез о равенстве дисперсий и средних при неизвестных дисперсиях и математических ожиданиях для нормально распределённых генеральных совокупностей.

Литература

1. Сборник задач по математике для вузов: В 4-х ч. – 3-е изд., перераб. и доп. / *Под ред. А.В. Ефимова и А.С. Поспелова*. – М.: Физматлит, 2003. – Ч. 4.
2. *Земсков В.Н.* Лекции по теории вероятностей и математической статистике. – М.: МИЭТ, 2002. – 152 с.
3. *Земсков В.Н.* Методы статистического анализа экспериментальных данных с использованием ЭВМ. – М.: МИЭТ, 1988. – 124 с.
4. *Вуколов Э.А.* Статистические методы обработки экспериментов и их реализация на ЭВМ. – М.: МИЭТ, 1984. – 110 с.
5. *Вуколов Э.А., Лесин В.В., Лисовец Ю.П., Ревякин А.М.* Лабораторный практикум по математической статистике. – М.: МИЭТ, 1986. – 90 с.
6. *Вуколов Э.А.* Регрессионный анализ. – М.: МИЭТ, 2000. – 51 с.
7. *Вуколов Э.А.* Основы статистического анализа: Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL. – М.: ФОРУМ: ИНФРА-М, 2004. – 464 с.
8. *Крамер Г.* Математические методы статистики. – М.: Мир, 1975. – 648 с.
9. *Боровков А.А.* Математическая статистика. – М.: Наука, 1984. – 472 с.
10. *Никитина Н.Ш.* Математическая статистика для экономистов. – 2-е изд., перераб. и доп. – М.: ИНФРА-М; Новосибирск: изд-во НГТУ, 2001. – 170 с.
11. *Муха В.С., Служанова Т.В.* Статистические методы обработки данных. – Минск: БГУИР, 2004. – 98 с.
12. *Мхитарян В.С., Трошин Л.И., Астафьева Е.В., Миронкина Ю.Н.* Математическая статистика (для бизнесменов и менеджеров). – М.: МГУЭСИ, 2004. – 214 с.
13. *Ефимова М.Р., Петрова Е.В., Румянцев В.Н.* Общая теория статистики. – 2-е изд., испр. и доп. – М.: ИНФРА-М, 2002. – 416 с.
14. *Колемаев В.А., Калинина В.Н.* Теория вероятностей и математическая статистика. – М.: ИНФРА-М, 2001. – 302 с.
15. *Кибзун А.И., Горяинова Е.Р., Наумов А.В.* Теория вероятностей и математическая статистика: Базовый курс с примерами и задачами. – 2-е изд., испр. и доп. – М.: Физматлит, 2005. – 232 с.

Приложение 1

Варианты заданий к лабораторной работе № 1

№ п/п	N=1	N=2	N=3	N=4	N=5	N=6	N=7	N=8	N=9	N=10
1	5	3	7	12	4	7	6	2	10	9
2	2	8	6	10	9	9	7	7	2	5
3	7	14	5	9	6	9	0	10	2	7
4	5	4	7	14	7	13	13	7	2	4
5	6	1	3	1	10	6	6	7	3	2
6	2	4	6	7	6	4	5	8	1	7
7	2	7	3	4	8	7	8	4	1	4
8	6	7	8	1	3	11	12	7	4	7
9	5	11	5	7	8	12	2	8	6	13
10	4	10	5	14	7	13	11	6	2	4
11	3	6	4	7	1	9	12	6	5	6
12	6	8	3	5	3	9	10	11	2	8
13	3	6	5	3	10	8	9	7	3	4
14	2	8	5	13	10	12	8	6	6	7
15	2	6	5	5	6	12	16	5	3	7
16	4	1	5	6	8	13	11	12	4	8
17	2	8	5	4	8	13	13	8	4	1
18	4	10	6	1	9	8	8	10	1	3
19	6	14	5	3	6	5	11	2	3	3
20	5	9	5	13	8	10	10	12	2	2
21	1	6	3	5	7	13	8	7	6	7
22	7	13	8	3	7	5	9	11	3	10
23	4	1	1	2	8	5	5	5	6	9
24	5	0	5	13	7	8	11	12	4	9
25	3	13	8	9	10	5	5	7	7	7
26	3	2	6	13	7	12	10	2	5	5
27	3	5	8	11	5	9	6	4	3	5
28	2	14	5	7	9	2	6	7	4	6
29	4	3	7	11	8	7	7	13	5	7
30	6	12	6	5	4	8	3	12	3	11
31	1	1	6	11	8	3	12	4	7	6
32	2	12	9	6	11	5	9	11	1	3
33	3	6	4	4	4	4	9	5	2	8
34	3	11	3	13	10	5	7	6	1	10
35	3	7	4	4	9	11	13	5	6	7
36	5	3	2	2	12	4	8	8	4	0
37	1	4	9	2	12	6	5	6	6	-1
38	4	11	7	8	4	11	7	5	2	6
39	4	8	5	7	4	4	12	6	2	5
40	4	3	7	4	5	9	5	11	5	5
41	11	12	6	6	6	8	6	5	2	4
42	8	4	4	13	7	6	5	6	4	-1
43	7	5	4	8	5	7	10	6	2	6
44	3	9	6	10	4	12	4	10	6	5
45	0	9	6	10	7	10	10	4	3	4
46	6	10	5	7	6	6	6	4	8	2
47	1	10	6	10	6	13	5	6	2	6
48	5	8	8	14	4	5	6	3	6	6
49	4	8	7	3	8	6	13	8	1	7
50	3	0	8	7	5	5	7	5	6	6

Продолжение

№ п/п	N=11	N=12	N=13	N=14	N=15	N=16	N=17	N=18	N=19	N=20
1	16	2	8	3	1	11	11	6	1	1
2	9	9	5	1	4	4	7	16	2	-1
3	4	6	5	1	5	2	14	8	9	2
4	13	10	10	1	9	10	4	7	5	3
5	4	3	8	0	3	7	9	12	1	0
6	18	2	1	4	0	8	8	8	4	8
7	5	9	3	1	4	9	11	9	8	4
8	8	2	6	3	4	5	7	12	3	4
9	17	5	5	5	5	4	2	11	2	3
10	5	7	7	6	0	10	11	9	0	6
11	8	4	11	1	7	2	8	6	11	3
12	11	0	4	2	4	9	4	7	5	3
13	4	7	3	4	10	11	10	9	8	3
14	17	1	5	2	3	7	11	6	2	4
15	17	0	5	0	4	8	2	10	3	3
16	12	8	7	6	2	5	11	14	8	2
17	8	6	5	2	4	11	9	8	2	1
18	5	6	6	3	3	11	6	4	9	7
19	12	5	3	5	4	14	3	7	5	2
20	17	5	4	2	2	2	14	9	8	1
21	4	7	5	5	5	10	7	13	10	1
22	7	4	8	4	8	7	13	8	9	7
23	10	6	8	3	5	10	6	12	10	4
24	15	8	9	1	5	6	5	6	1	1
25	16	8	4	3	9	8	7	8	10	1
26	7	10	4	5	7	7	5	10	2	8
27	14	1	6	5	6	13	2	10	1	-2
28	9	2	6	2	4	2	7	7	6	6
29	15	1	2	1	5	8	12	9	5	2

30	8	4	5	5	0	7	10	8	1	1
31	11	2	8	6	4	6	4	7	0	5
32	12	0	5	2	1	10	2	5	7	4
33	10	1	7	4	5	8	9	12	7	6
34	5	5	4	0	6	6	9	10	9	2
35	8	1	4	4	7	4	10	9	0	0
36	11	6	6	3	4	9	2	14	6	4
37	14	7	2	6	3	11	12	6	1	3
38	12	4	7	3	8	6	13	15	10	3
39	11	3	5	3	5	8	9	8	3	9
40	17	3	4	6	11	3	4	8	8	-1
41	18	1	4	5	6	5	4	7	6	2
42	15	9	8	4	4	10	7	10	9	6
43	6	3	5	3	2	9	8	16	1	5
44	4	9	3	1	4	9	12	6	7	5
45	11	1	2	6	4	6	4	10	9	10
46	14	10	5	1	5	3	9	13	11	8
47	6	8	2	2	4	8	10	4	10	3
48	8	6	8	3	6	10	5	16	6	2
49	15	2	5	3	2	11	3	10	10	3
50	7	5	7	2	4	4	13	8	10	5
№ п/п	N=21	N=22	N=23	N=24	N=25	N=26	N=27	N=28	N=29	N=30
1	2	6	8	0	10	3	2	1	2	7
2	3	1	-1	2	7	8	1	3	1	5
3	6	0	5	5	-1	8	6	3	4	8
4	0	-2	6	1	-1	4	11	3	6	10
5	4	3	6	4	9	5	4	0	5	7
6	4	4	10	1	2	5	8	4	4	6
7	1	5	2	7	11	10	7	6	6	9
8	4	6	3	1	-1	11	2	0	5	9
9	2	2	-2	14	7	8	3	0	9	9
10	4	7	8	20	0	6	0	5	4	7
11	7	5	5	1	2	9	2	0	5	10
12	4	2	0	4	2	7	5	1	9	9
13	6	-1	6	0	-1	6	5	3	-2	7
14	5	5	-4	6	2	8	5	6	1	10
15	2	5	2	1	6	7	14	1	4	10
16	5	-1	8	7	5	6	3	4	0	8
17	3	3	5	1	11	2	14	8	8	12
18	2	6	5	2	11	3	1	5	7	4
19	0	3	6	3	6	3	10	2	3	3
20	2	6	9	9	1	6	11	1	5	8
21	5	3	7	3	8	4	5	1	-2	5
22	2	0	-1	2	10	4	1	6	0	7
23	0	3	1	1	3	6	6	3	-1	6
24	4	3	5	5	5	7	3	13	3	6
25	4	8	11	0	5	6	13	3	7	11
26	0	1	9	0	3	7	2	1	-1	7
27	4	0	-1	1	0	7	14	2	4	2
28	5	7	3	0	0	6	0	2	5	4
29	3	1	9	14	10	5	5	3	-2	9
30	2	-2	7	9	11	12	12	0	-2	9
31	2	3	-4	18	6	4	5	2	-1	15
32	3	-2	5	3	7	9	18	2	2	12
33	3	6	6	7	6	9	0	0	3	9
34	2	1	11	5	0	10	4	1	5	6
35	3	7	-2	4	9	3	1	1	4	8
36	4	2	10	0	-1	10	2	0	3	9
37	4	6	-5	4	9	9	10	3	3	12
38	2	4	9	4	11	6	12	0	0	7
39	3	8	-1	12	0	7	3	0	0	11
40	4	4	0	4	0	4	4	1	1	5
41	3	4	2	6	8	1	2	1	-1	8
42	2	-1	4	0	6	3	2	2	1	6
43	2	1	5	1	0	8	2	0	-1	9
44	0	3	3	2	10	6	1	1	-2	10
45	6	0	4	1	10	2	1	12	7	5
46	3	2	8	0	8	3	12	6	8	6
47	3	0	5	3	8	6	4	9	8	4
48	4	6	4	10	7	7	0	10	5	7
49	2	0	5	2	4	7	3	1	6	10
50	5	5	5	4	8	4	3	5	2	7

Приложение 2

Варианты заданий к лабораторной работе № 4

№ п/п	N=1		N=2		N=3		N=4		N=5	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	0,75	3,72	2,22	2,40	3,99	1,24	12,81	1,20	0,96	8,37
2	4,62	7,84	6,24	-2,93	3,72	0,44	-2,78	18,95	3,03	22,56
3	4,01	4,09	-2,11	23,66	0,71	3,48	8,61	6,76	6,07	32,18
4	4,90	7,43	2,70	-3,96	4,08	1,34	5,02	10,29	5,01	11,60
5	1,84	2,03	0,62	15,34	3,50	0,16	8,02	6,14	6,30	21,63
6	7,21	7,65	5,68	-11,84	0,92	2,96	10,24	4,40	4,66	22,30
7	3,81	3,31	1,45	-1,86	0,61	2,88	5,20	14,50	6,05	15,46
8	7,60	7,74	3,70	-2,27	2,71	2,00	-1,49	16,55	5,15	21,54
9	7,59	8,63	-2,46	13,13	1,87	1,63	18,11	-5,81	10,45	24,72
10	5,56	7,02	3,54	1,19	1,66	2,36	22,93	-9,89	-1,58	6,66
11	3,78	5,22	0,41	8,06	2,88	1,81	9,42	6,11	2,69	6,56
12	7,94	8,46	6,64	-21,67	2,41	2,07	3,67	9,18	0,61	11,92
13	9,69	10,09	1,70	-14,24	3,37	0,90	15,72	-5,89	4,41	21,89
14	6,50	6,54	-1,63	13,60	0,52	3,24	13,11	-4,28	9,84	28,98
15	6,96	8,97	0,33	-0,04	1,77	2,12	1,51	12,68	4,27	15,35
16	6,56	7,40	3,86	-1,04	0,94	3,03	4,55	12,36	2,55	6,05
17	2,29	3,97	1,06	10,73	-0,15	3,62	11,12	7,10	2,17	13,24
18	1,55	4,09	5,73	-16,08	0,63	3,82	15,40	1,20	11,28	28,59
19	3,15	3,95	4,91	-3,68	2,75	0,56	4,57	5,50	2,87	10,68
20	6,29	4,48	1,81	1,35	0,94	2,36	-0,04	16,32	7,81	12,04
21	7,40	7,04	2,24	-3,67	1,76	2,59	1,64	7,34	4,80	29,13
22	3,75	4,14	4,06	1,68	3,14	1,66	-5,78	18,28	-0,77	11,45
23	7,22	7,93	7,46	-18,96	1,41	3,42	0,92	6,67	0,20	6,61
24	5,22	6,13	-3,33	22,12	4,05	1,06	2,26	16,06	6,28	23,97
25	3,44	3,56	8,26	-10,87	2,72	1,88	5,84	8,96	4,66	14,96
26	5,81	6,65	-1,07	-1,23	3,00	1,18	-2,53	19,24	2,48	17,07
27	4,15	5,24	1,56	16,07	2,22	3,22	5,09	7,40	11,25	29,38
28	5,38	4,85	6,55	-6,87	2,95	0,79	10,89	3,32	7,02	30,78
29	6,96	7,92	3,28	1,40	1,28	3,18	7,47	1,21	7,07	21,34
30	5,57	6,90	-0,21	2,79	3,41	2,27	-3,60	11,38	4,76	21,44
31	8,75	9,81	-1,12	8,06	0,90	3,10	-7,44	16,31	10,68	29,12
32	9,80	10,12	2,09	-7,14	1,55	2,52	4,99	15,22	0,90	17,31
33	4,24	4,51	-2,38	6,80	2,04	3,07	12,78	1,69	2,92	5,47
34	10,14	9,61	-4,83	16,03	2,53	1,22	4,97	6,76	7,26	23,93
35	2,02	2,60	0,06	8,77	3,11	1,41	-1,02	12,00	8,23	30,34
36	11,28	10,91	1,39	11,05	3,68	0,86	18,74	-9,98	4,76	7,82
37	3,85	4,01	-1,19	16,32	1,30	3,57	10,21	8,00	-1,01	5,17
38	1,84	4,60	0,19	2,98	0,85	3,27	8,65	8,27	10,66	14,87
39	7,47	9,01	3,78	-16,24	2,35	1,93	8,43	6,38	-0,04	5,26
40	7,30	8,43	-3,18	16,74	3,76	1,15	11,32	3,47	7,61	24,42
41	1,03	2,15	4,32	-10,62	0,20	3,39	8,75	9,43	8,23	30,34
42	-0,73	0,34	2,96	-11,78	3,40	1,63	13,17	4,36	2,69	12,63
43	9,44	11,88	1,07	5,02	0,08	4,26	7,72	11,13	3,63	19,67
44	3,48	2,53	1,31	4,97	2,36	2,08	3,08	5,92	2,09	4,43
45	2,44	2,38	1,13	6,89	0,45	3,65	9,30	1,15	3,29	23,02
46	5,77	9,18	-1,56	9,35	2,54	2,23	9,97	7,85	7,53	15,25
47	2,60	3,97	4,31	-9,02	1,00	3,32	10,56	6,82	3,12	11,32
48	5,36	4,37	7,64	-17,94	3,44	1,36	16,84	0,53	1,01	10,24
49	-0,21	2,07	6,35	-13,92	-0,05	3,61	-1,39	14,93	5,92	28,54
50	8,68	11,95	4,35	-5,97	2,43	2,34	6,45	7,57	1,62	5,32
№ п/п	N=6		N=7		N=8		N=9		N=10	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	12,03	12,01	17,83	1,40	0,12	-0,08	-2,11	-6,06	9,77	41,77
2	10,89	9,66	1,21	30,14	3,96	3,80	9,28	1,27	16,26	32,44
3	12,51	13,61	18,92	13,01	0,16	1,39	5,06	9,80	11,26	72,52
4	12,66	6,12	9,68	19,09	0,19	1,15	12,19	10,91	5,43	53,50
5	17,57	19,55	-1,06	59,62	0,46	1,41	3,87	5,98	4,42	65,10
6	5,92	5,25	12,07	25,07	1,24	1,70	4,40	6,14	13,17	26,70
7	8,90	6,93	13,67	15,96	2,84	4,73	5,31	6,19	22,26	27,90
8	12,44	13,12	11,95	4,07	0,49	0,90	4,56	1,35	19,39	17,87
9	8,22	9,43	14,41	4,32	2,05	3,39	8,46	5,47	14,59	29,00
10	7,41	9,83	6,41	28,17	0,38	1,14	5,74	5,05	3,52	74,50
11	2,08	4,77	10,09	23,20	2,77	4,22	4,97	1,02	8,57	85,41
12	3,20	6,66	8,45	26,03	1,10	1,57	7,73	4,30	1,63	85,02
13	3,32	2,34	12,61	12,91	1,85	1,55	12,60	19,29	7,39	57,85
14	2,33	0,07	3,75	56,37	2,57	3,28	-1,05	-0,02	14,85	10,28
15	3,35	1,34	8,47	42,23	0,15	1,27	9,89	5,82	13,99	49,89
16	7,16	5,76	2,86	61,61	3,58	4,68	10,18	10,82	8,32	68,50
17	12,90	15,44	4,66	46,25	3,86	4,06	6,34	2,63	7,45	60,67
18	4,98	-0,06	6,62	22,13	0,78	1,43	5,21	0,61	6,68	79,03
19	2,05	0,79	16,93	-16,53	0,62	2,12	3,57	3,33	2,00	96,97
20	13,12	12,74	8,19	18,58	0,87	2,15	8,05	12,66	10,53	1,88
21	0,67	3,18	9,38	28,24	2,22	1,98	9,55	10,86	14,58	43,49
22	6,87	3,35	18,86	-7,18	3,56	4,40	2,07	5,50	5,92	62,54
23	7,37	6,42	6,02	31,23	1,27	2,49	6,88	3,00	9,53	66,54
24	7,36	7,11	14,36	5,45	1,81	2,03	4,50	3,45	2,30	69,44
25	4,04	3,93	0,76	59,16	1,81	3,73	-2,20	-2,45	1,37	70,24
26	3,35	4,23	13,67	18,21	2,36	3,47	6,09	2,04	11,60	38,22
27	6,54	5,68	5,78	31,61	2,85	3,33	5,94	1,46	12,78	32,38
28	3,90	1,52	9,37	26,60	2,85	3,50	5,71	10,23	2,40	55,84
29	7,01	5,25	4,65	50,04	1,02	1,73	12,72	13,56	19,25	-14,38

30	15,62	14,21	13,66	7,05	0,46	1,17	4,81	3,03	11,01	50,37
31	4,74	8,47	1,74	19,02	0,43	2,10	10,44	14,46	14,33	33,88
32	9,87	10,24	4,20	47,00	0,46	0,94	12,76	14,49	16,35	33,91
33	0,43	0,69	10,45	-8,76	1,41	2,29	8,99	5,83	10,10	55,49
34	3,44	4,08	2,96	42,79	2,91	3,94	4,83	10,99	8,55	78,06
35	4,73	5,77	15,88	1,60	1,53	1,80	2,58	4,54	15,27	55,17
36	7,89	12,55	6,53	27,59	2,82	3,73	10,75	17,46	5,54	74,08
37	9,71	9,62	5,18	28,39	1,43	1,99	-1,32	4,04	18,29	12,04
38	16,30	19,31	5,77	38,13	0,06	1,01	-0,44	-2,38	0,99	91,36
39	4,28	4,90	0,33	60,50	0,66	1,64	13,09	8,72	10,58	29,80
40	7,21	9,00	0,94	52,23	3,46	3,49	6,85	7,36	18,19	10,54
41	8,44	9,44	4,09	41,63	2,23	2,43	2,92	2,40	17,07	12,66
42	12,57	16,75	14,42	5,41	0,14	2,03	0,47	-2,11	12,40	36,90
43	2,31	-0,64	11,86	23,36	0,67	1,78	6,40	9,19	5,50	71,39
44	13,38	9,39	13,52	14,63	0,68	1,58	-0,25	6,51	6,63	43,50
45	15,35	15,04	8,94	3,91	0,39	0,70	4,76	3,38	4,87	49,35
46	1,21	3,92	7,94	29,36	3,02	3,55	2,81	0,23	5,86	23,08
47	11,82	11,20	20,24	-6,33	0,97	2,29	5,97	5,78	19,88	2,17
48	13,39	11,55	14,75	1,37	3,05	3,66	6,97	10,42	12,41	40,29
49	11,71	15,10	9,42	21,20	2,16	3,30	-2,64	-5,26	13,09	39,93
50	14,72	11,49	16,18	-4,52	3,79	4,49	8,82	4,82	17,66	-3,04
№ п/п	N=11		N=12		N=13		N=14		N=15	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	6,09	6,71	14,72	11,01	3,67	4,03	8,51	6,55	-0,15	11,37
2	16,59	14,51	13,58	9,15	0,43	1,14	1,55	13,14	2,63	44,11
3	15,50	15,90	7,34	9,95	5,85	5,75	10,18	6,00	2,33	53,38
4	11,98	12,84	9,76	5,53	7,75	7,41	12,28	2,99	0,36	12,76
5	13,85	18,37	11,96	8,04	7,29	8,40	3,55	8,38	2,80	35,45
6	0,95	0,87	5,56	-0,60	11,16	11,16	3,92	15,31	2,33	14,45
7	6,31	6,83	9,15	4,40	9,44	10,21	15,99	1,29	1,17	21,97
8	0,93	-0,91	9,39	1,86	3,95	3,30	6,31	7,92	2,08	14,67
9	18,27	20,94	14,14	8,27	7,01	7,34	6,68	13,08	0,40	19,04
10	5,60	4,71	7,03	1,58	6,70	8,10	14,59	3,04	0,09	24,79
11	17,59	17,62	7,38	3,08	2,97	2,40	2,28	9,83	-0,25	15,02
12	8,97	10,71	7,28	2,57	9,37	10,04	5,70	9,53	0,28	18,16
13	8,38	10,55	14,96	8,66	6,44	6,66	12,26	4,98	3,87	49,45
14	8,34	5,48	15,26	11,05	5,06	5,47	11,91	7,13	0,03	25,05
15	20,17	19,20	7,85	7,81	4,19	4,26	11,51	3,39	1,67	30,95
16	12,21	11,72	15,02	9,87	6,01	4,02	4,03	13,05	-0,78	10,70
17	4,53	9,83	12,97	6,66	7,02	7,34	11,34	1,16	-2,68	9,81
18	11,75	11,91	5,37	3,58	3,92	6,09	3,21	12,44	0,81	14,65
19	19,47	18,42	12,32	5,85	8,08	9,37	7,41	10,09	2,10	47,39
20	13,24	15,45	13,13	5,99	4,69	4,73	0,23	15,33	4,21	49,97
21	0,44	-1,88	5,95	0,75	7,93	7,64	8,34	7,83	0,36	16,80
22	8,56	5,13	4,95	-0,85	8,57	9,37	6,37	6,09	4,62	52,71
23	12,15	13,07	8,21	4,79	2,88	2,92	14,65	7,90	3,54	40,97
24	5,98	1,51	7,04	5,57	1,37	1,24	4,13	16,35	5,18	46,51
25	17,09	14,34	8,78	5,68	8,70	9,27	0,52	12,64	1,02	29,88
26	5,03	6,26	17,98	10,77	8,84	10,85	9,75	6,71	0,59	30,58
27	4,78	6,86	8,24	4,98	3,74	1,22	7,57	8,16	1,96	20,91
28	5,78	3,64	2,90	0,62	3,77	4,97	10,05	7,49	-0,19	10,09
29	6,20	5,50	13,73	12,01	4,10	3,40	4,68	12,53	0,53	18,95
30	17,19	20,35	9,88	5,49	5,92	7,56	8,88	8,78	1,13	16,43
31	9,93	12,11	6,16	0,87	5,52	3,11	5,99	8,16	3,65	50,45
32	10,44	12,04	10,31	5,58	4,11	4,34	14,68	-1,94	0,81	13,33
33	12,61	11,76	6,01	2,47	7,51	8,08	14,64	3,11	1,15	7,19
34	13,62	16,23	14,43	8,42	5,80	5,79	16,06	1,54	2,15	25,28
35	12,20	13,87	13,85	7,53	2,31	2,57	6,72	9,35	1,27	40,88
36	17,79	18,96	6,20	2,03	5,39	4,08	14,40	6,36	3,51	47,42
37	9,85	5,69	7,95	1,30	11,61	10,75	8,15	10,20	1,49	32,17
38	16,41	15,47	3,82	1,39	8,11	7,63	9,77	3,16	-0,43	16,13
39	18,11	14,60	9,43	2,56	9,28	10,22	8,12	11,17	1,22	22,49
40	11,73	10,06	7,89	4,55	3,40	3,20	15,45	3,54	0,89	16,75
41	17,14	15,25	12,56	9,35	0,21	2,55	14,00	6,63	-0,50	7,77
42	4,09	6,20	14,91	13,69	5,73	6,40	9,65	6,25	1,93	27,23
43	9,74	9,54	9,01	6,45	3,53	1,79	6,50	10,84	2,96	36,99
44	7,79	9,86	17,05	11,05	0,23	0,18	5,12	9,45	0,84	11,60
45	8,20	8,42	14,48	10,79	4,51	4,23	14,93	4,87	2,18	42,99
46	10,29	10,21	9,30	6,44	5,38	6,00	1,71	15,36	2,07	52,08
47	1,39	-5,16	13,25	10,54	5,73	5,97	4,78	9,39	1,89	40,02
48	12,46	10,01	7,10	3,89	10,30	10,47	14,38	1,32	2,17	42,42
49	0,64	6,19	11,08	6,91	9,40	9,02	11,43	5,70	0,92	19,52
50	-0,55	-0,68	13,65	6,88	0,39	0,81	10,24	8,05	1,38	17,63

№ п/п	N=16		N=17		N=18		N=19		N=20	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	11,04	-6,66	12,16	87,37	7,41	7,77	-1,60	1,90	4,19	9,19
2	10,72	-4,97	11,12	89,11	8,08	7,99	3,80	4,72	3,04	11,94
3	18,27	-10,12	10,92	54,96	9,49	10,24	4,30	3,75	4,60	8,09
4	23,16	-18,05	11,70	71,49	8,50	6,71	-4,63	0,66	9,83	10,33
5	7,13	1,72	12,45	84,26	10,87	12,09	-3,14	2,17	8,66	7,15
6	1,47	8,10	9,52	69,89	6,86	6,80	2,04	6,26	1,30	12,34
7	16,92	-13,55	10,27	64,67	7,88	7,49	-0,92	2,64	4,22	16,35
8	21,38	-15,14	11,26	78,22	8,97	9,58	-1,49	4,19	5,11	7,70
9	19,88	-13,64	9,19	52,11	8,94	9,59	6,78	5,94	9,85	5,64
10	12,18	-3,63	9,65	64,17	7,73	8,75	-1,56	0,66	8,80	4,52
11	22,55	-14,72	7,53	50,46	7,00	7,93	-2,36	0,95	12,17	4,52
12	7,02	3,56	7,28	60,32	8,13	9,35	1,35	3,00	11,25	2,06
13	9,92	-0,35	7,81	50,08	7,47	7,22	-0,07	3,71	5,73	7,41
14	25,24	-21,30	7,13	42,69	7,73	7,03	-0,42	1,19	4,05	10,51
15	12,56	-5,22	7,35	44,06	8,14	7,55	-1,87	1,92	5,41	9,97
16	14,05	-11,05	9,24	65,96	8,13	7,87	-0,93	1,21	1,28	14,68
17	0,91	3,32	11,23	68,66	9,33	10,57	2,01	0,70	1,67	9,67
18	16,19	-11,43	8,19	71,55	8,10	6,55	-1,09	2,76	11,99	3,31
19	17,41	-14,88	7,85	60,26	6,53	6,14	0,56	2,25	7,66	5,93
20	20,25	-12,82	11,95	82,61	8,46	8,74	0,63	4,68	5,17	9,87
21	17,24	-12,19	6,54	45,90	7,40	8,23	0,85	3,07	3,26	11,52
22	8,45	0,38	10,03	71,04	6,81	5,83	2,57	5,67	12,58	2,88
23	26,12	-14,08	9,75	78,61	7,55	7,45	1,64	5,08	8,34	3,57
24	13,44	-8,34	10,42	50,39	6,60	6,72	7,22	9,16	5,79	4,39
25	20,81	-15,43	8,38	69,92	7,17	7,23	-1,97	1,23	3,42	9,71
26	9,66	1,31	7,89	51,92	7,38	7,75	1,23	2,24	4,44	9,13
27	12,47	-6,45	9,00	62,00	8,04	7,94	3,26	5,77	11,31	4,58
28	19,00	-10,06	8,29	65,72	7,19	6,49	0,61	2,82	7,57	3,14
29	15,63	-8,79	10,53	76,45	7,70	8,00	2,46	3,77	1,62	14,61
30	4,96	1,22	12,13	66,25	9,96	9,98	4,85	6,01	5,71	6,48
31	5,94	5,56	8,56	45,82	7,40	8,77	3,55	5,31	11,06	6,78
32	9,08	-2,10	9,88	72,43	9,13	9,54	2,57	4,74	10,35	2,15
33	16,96	-9,92	7,21	43,09	6,29	6,36	-0,19	4,23	2,46	9,66
34	15,85	-11,18	7,51	38,40	7,97	8,27	2,04	5,05	1,02	11,19
35	1,55	6,18	8,17	53,81	7,94	8,41	0,76	1,96	5,77	7,77
36	25,64	-17,77	9,84	63,47	7,79	9,57	0,03	3,50	8,63	4,05
37	10,33	-0,88	9,41	54,03	9,67	9,93	-0,11	2,24	6,91	4,76
38	6,86	2,46	12,21	71,29	10,32	11,83	2,95	3,30	3,56	8,54
39	6,30	-1,22	7,89	57,97	8,02	8,34	1,06	3,67	9,47	2,22
40	1,73	6,57	9,37	47,04	7,98	8,78	1,25	3,71	6,16	3,72
41	13,98	-5,04	9,66	77,79	8,43	9,01	1,14	2,94	8,26	3,57
42	16,56	-9,50	11,26	81,53	9,06	10,84	0,43	3,86	6,70	14,32
43	6,59	2,87	7,66	51,47	6,97	6,03	-1,26	3,19	4,95	10,64
44	5,69	-0,34	11,96	79,54	8,64	7,72	1,65	4,47	3,37	10,73
45	1,60	0,93	12,03	72,03	9,91	10,28	3,32	4,05	1,53	10,13
46	13,94	-6,76	6,58	38,08	7,73	8,64	-1,21	0,81	9,54	4,95
47	9,06	-0,81	11,11	77,99	8,75	8,90	3,99	4,26	3,11	5,38
48	1,24	1,65	12,18	89,22	8,32	8,12	-0,41	2,12	5,09	5,79
49	18,27	-7,74	11,38	87,79	8,28	9,77	1,85	4,03	11,08	3,87
50	25,08	-14,25	12,04	79,10	9,45	8,83	0,90	4,28	8,74	-2,23

№ п/п	N=21		N=22		N=23		N=24		N=25	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	10,47	12,04	9,39	20,27	1,52	-7,99	8,43	5,65	14,33	6,62
2	5,69	5,97	10,27	28,41	1,89	-16,79	9,13	5,13	10,06	12,36
3	9,24	9,53	11,77	25,82	1,98	-6,10	-0,48	1,30	9,13	8,29
4	3,36	5,48	11,45	23,01	-0,27	8,76	2,00	2,10	6,99	17,26
5	7,08	6,45	8,94	20,43	0,80	4,02	8,43	6,30	8,74	12,70
6	0,01	-0,23	5,05	18,14	0,24	0,62	-0,84	-0,20	10,40	21,52
7	8,39	5,93	10,36	28,38	-0,20	4,04	-0,23	1,36	13,99	8,65
8	13,00	9,74	15,86	31,57	0,81	-5,00	3,39	4,06	20,25	-4,48
9	7,90	6,48	10,26	23,80	0,53	1,93	6,38	2,57	11,11	11,42
10	12,47	12,73	14,12	31,98	0,76	-3,19	2,45	-1,53	5,63	17,37
11	2,21	1,19	14,37	38,48	1,00	-0,80	7,29	4,26	15,60	3,36
12	1,89	-0,55	11,79	27,51	1,67	-7,67	1,70	2,45	11,39	10,55
13	10,32	12,04	10,16	29,30	0,40	-0,76	6,79	4,71	11,78	6,83
14	12,79	11,26	11,58	25,79	0,33	0,21	-1,33	0,06	8,59	12,49
15	7,17	7,01	8,11	18,40	2,98	-27,58	2,14	2,36	8,11	14,57
16	9,77	9,57	9,00	25,34	0,78	-1,54	-0,30	-1,24	10,58	15,18
17	14,41	13,67	11,64	26,05	-0,41	14,21	4,36	7,75	5,91	18,18
18	-2,87	-2,10	12,93	27,78	1,51	-9,78	10,93	13,88	15,50	-5,08
19	8,49	7,84	14,73	30,12	1,35	-10,45	8,24	9,43	15,75	6,39
20	-0,46	-3,45	8,99	20,24	2,06	-19,42	4,91	4,60	8,47	15,83
21	4,81	2,86	12,19	26,12	0,15	-4,95	5,03	8,58	16,48	4,49
22	4,55	4,90	14,82	36,88	0,78	-1,63	6,03	4,14	8,48	15,96
23	16,73	14,13	15,52	31,49	1,70	-15,39	7,43	1,83	12,68	11,42
24	0,17	-0,46	8,60	20,70	1,36	-7,25	-1,55	-0,33	14,12	12,11
25	14,50	12,51	8,84	20,54	1,28	-2,48	4,49	2,99	9,82	16,11
26	14,28	11,88	9,91	13,89	1,34	-8,17	7,42	8,82	9,62	15,22
27	15,85	13,70	12,64	31,30	1,75	-25,28	2,60	-0,11	14,97	-0,59
28	11,21	10,27	10,58	26,81	0,32	7,12	9,38	7,56	11,75	5,29
29	8,08	8,23	13,34	28,46	0,50	-12,53	8,00	8,57	5,77	17,97
30	7,52	7,16	7,40	9,94	0,62	-3,32	7,20	7,78	5,36	11,74
31	4,29	6,12	13,59	27,27	1,32	-15,89	1,53	-0,33	5,43	13,14
32	11,50	10,83	12,19	29,04	1,44	-1,27	7,23	7,05	10,39	13,36
33	4,98	2,15	11,93	29,14	0,67	-1,58	4,52	5,25	6,40	14,68

34	16,96	16,64	13,88	29,05	1,93	-13,80	2,18	7,08	12,37	12,88
35	15,78	15,60	11,45	29,38	0,59	9,51	2,75	-1,58	10,64	5,00
36	7,54	7,95	11,84	25,75	1,62	-5,83	7,60	1,92	2,07	20,85
37	11,89	12,55	7,31	17,62	1,53	-3,99	2,19	2,50	8,66	5,81
38	5,93	8,11	9,32	17,43	1,12	-5,10	-4,35	-3,17	14,37	0,86
39	7,14	6,01	8,24	24,34	0,51	4,50	1,20	1,51	7,02	15,15
40	12,94	13,15	7,55	18,51	1,12	-4,61	-0,26	-2,17	12,12	8,32
41	11,81	11,55	10,67	18,89	0,91	3,96	2,33	0,68	7,66	8,82
42	3,39	2,28	9,58	13,79	1,17	-10,26	0,78	-4,54	11,77	8,06
43	12,93	16,52	10,14	24,06	0,53	12,94	10,04	13,09	12,15	6,52
44	12,36	11,65	12,78	28,92	3,35	-27,55	-2,78	-0,68	10,44	6,97
45	5,49	2,85	11,23	26,45	0,83	0,94	1,76	2,76	12,32	9,68
46	2,37	5,51	10,78	23,23	2,18	-18,13	11,31	10,63	16,28	-1,24
47	9,90	10,29	11,29	25,60	1,16	-1,14	3,00	6,14	9,00	13,26
48	8,62	6,44	9,98	22,34	1,74	-11,40	9,02	6,45	9,92	2,98
49	7,15	8,13	9,32	21,78	0,49	7,94	1,14	3,69	6,62	16,50
50	4,95	7,13	11,64	29,06	0,39	2,08	3,45	7,35	7,83	14,55
№ п/п	N=26		N=27		N=28		N=29		N=30	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	12,67	7,10	0,05	4,89	2,62	2,55	5,88	2,11	5,59	4,66
2	13,52	6,82	0,76	13,16	3,14	3,04	4,47	4,23	5,12	4,56
3	16,75	2,25	0,89	17,21	2,59	2,49	2,75	7,06	5,78	4,88
4	13,27	5,32	0,29	5,08	2,65	2,26	4,16	5,08	5,29	4,10
5	11,16	9,86	0,25	5,12	3,22	3,47	5,15	3,16	5,67	4,74
6	18,09	1,07	0,93	21,16	2,82	2,64	2,92	7,76	4,26	4,07
7	14,72	5,74	0,13	6,56	2,05	2,01	7,12	1,72	5,56	5,22
8	12,77	5,31	0,94	22,49	2,48	2,46	3,77	6,02	5,17	4,49
9	14,93	4,78	0,70	13,64	2,28	2,40	5,57	3,38	4,71	4,26
10	12,22	9,27	0,85	16,37	2,07	2,11	3,43	3,71	5,72	5,34
11	16,44	1,78	0,21	9,57	3,08	3,03	3,36	4,13	5,59	5,11
12	13,33	6,94	0,46	10,56	2,72	2,36	4,37	4,85	5,67	5,21
13	14,20	5,33	0,08	5,01	2,04	2,17	4,86	2,08	5,22	5,53
14	10,60	9,25	0,85	18,93	2,64	2,88	4,53	5,42	6,11	4,59
15	17,08	2,20	0,56	10,57	2,44	2,08	4,00	5,50	5,59	5,17
16	14,59	1,25	0,32	10,77	2,79	2,40	3,83	6,02	6,01	5,53
17	8,18	11,53	0,37	8,85	2,38	2,29	5,49	3,62	5,63	5,33
18	14,74	8,05	0,87	17,77	2,33	2,27	3,64	6,88	6,35	5,48
19	17,56	3,74	0,37	11,53	2,80	2,55	5,84	2,23	5,70	5,48
20	13,94	5,31	0,07	4,30	2,84	2,64	3,66	6,00	5,69	5,19
21	16,42	2,26	0,20	8,66	3,08	2,99	6,98	3,76	5,28	5,09
22	17,60	2,90	0,05	7,71	2,36	2,04	4,58	4,46	4,89	4,79
23	10,13	9,10	0,57	10,26	2,91	2,97	5,01	4,95	5,53	5,50
24	6,86	12,08	0,12	7,31	2,78	2,70	2,25	5,79	5,12	4,58
25	18,84	1,27	0,52	11,71	2,15	2,12	4,22	4,72	5,83	5,78
26	14,55	7,96	0,12	4,07	2,34	2,59	2,62	4,22	6,01	5,50
27	12,75	2,21	0,77	14,46	2,13	1,91	6,11	1,39	5,56	4,43
28	11,11	10,06	0,38	9,55	2,03	2,28	4,95	3,74	5,51	4,90
29	7,71	10,28	0,82	13,54	2,23	2,38	3,66	2,99	5,25	4,52
30	13,05	8,84	0,05	1,70	1,88	1,79	3,78	2,84	5,93	4,65
31	8,72	6,43	0,60	10,95	2,91	2,89	4,24	4,61	5,95	5,18
32	11,19	8,36	0,95	17,33	2,54	2,75	6,50	2,08	4,89	4,16
33	12,35	7,76	0,29	5,31	1,93	1,94	4,31	2,77	5,93	5,63
34	15,90	3,25	0,89	19,08	2,91	2,92	4,45	5,58	6,00	5,53
35	14,05	5,55	0,10	5,12	2,67	2,67	5,98	2,45	5,84	5,61
36	8,88	8,10	0,07	1,72	2,79	2,70	4,18	6,02	6,56	6,01
37	17,35	0,40	0,23	8,27	2,93	3,11	4,32	9,36	5,36	4,76
38	12,92	5,45	0,93	20,11	2,22	2,35	4,12	5,75	5,26	4,69
39	19,41	1,32	0,06	5,93	2,28	2,56	4,13	6,66	6,14	5,64
40	8,38	10,44	0,26	8,76	2,60	2,44	5,60	2,36	6,26	6,10
41	10,25	12,83	1,00	17,94	2,40	2,44	3,90	6,55	5,53	5,30
42	9,80	10,48	0,21	8,10	2,77	2,83	4,06	4,07	5,02	3,97
43	13,06	3,21	0,50	13,24	2,91	2,86	3,78	5,81	6,14	5,57
44	10,37	8,72	0,29	9,14	2,52	2,70	2,98	7,27	5,29	4,67
45	11,07	7,62	0,67	15,26	3,32	3,34	5,52	3,15	6,16	5,77
46	5,97	14,24	0,96	18,47	2,21	1,92	5,51	3,32	4,78	4,19
47	10,29	9,27	0,77	15,23	2,41	2,25	2,67	7,76	5,60	5,50
48	13,17	6,27	0,67	14,22	2,74	2,51	4,88	6,57	5,56	4,93
49	12,70	5,83	0,13	4,74	2,50	2,50	4,17	6,21	6,30	5,08
50	12,96	6,91	0,10	4,35	2,56	2,38	7,34	1,07	5,37	5,39