МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ТРАНСПОРТА

Кафедра «ПРИКЛАДНАЯ МАТЕМАТИКА»

Г. Ю. МИШИН, Е. Л. САЗОНОВА
Т. Т. СНОПОК, Д. Н. ШЕВЧЕНКО

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Лабораторный практикум

Гомель 2001

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТРАНСПОРТА

Кафедра «ПРИКЛАДНАЯ МАТЕМАТИКА»

Г. Ю. МИШИН, Е. Л. САЗОНОВА
Т. Т. СНОПОК, Д. Н. ШЕВЧЕНКО

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Лабораторный практикум

Под редакцией канд. физ.-мат. наук доц. **В. С. Серёгиной**

Одобрен
методической комиссией
строительного факультета

Гомель 2001

УДК 519.22/.25(076.5)
М34

М34 **Математическая** статистика: Лабораторный практикум/ *Г. Ю. Мишин, Е. Л. Сазонова, Т. Т. Снопок, Д. Н. Шевченко*; Под ред. *В. С. Серёгиной*. – Гомель: БелГУТ, 2001. – 60 с.

Содержатся методические указания для выполнения лабораторных работ по курсу математической статистики. Приводятся инструкции по использованию необходимых процедур пакета обработки статистических данных Statgraphics.

Предназначен для студентов дневной формы обучения всех специальностей.

Рецензенты: зав. кафедрой автоматики и телемеханики БелГУТа, докт. техн. наук проф. **К. А. Бочков**;
зав. кафедрой математических проблем управления ГГУ им. Ф. Скорины, докт. техн. наук проф. **И. В. Максимей**.

Учебное издание

Геннадий Юрьевич МИШИН
Елена Леонидовна САЗОНОВА
Татьяна Тадеушевна СНОПОК
Дмитрий Николаевич ШЕВЧЕНКО

Математическая статистика
Лабораторный практикум

Редактор Т. М. Ризевская
Технический редактор В. М. Кучерова
Корректор М. П. Дежко
Компьютерный набор и вёрстка
кафедры «Прикладная математика» БелГУТа

Подписано в печать 2.08.2001.
Формат бумаги 60 × 84 $^{1}/_{16}$. Бумага газетная. Гарнитура Таймс.
Печать офсетная. Усл. печ. л. 3,49. Уч.-изд. л. 3,51. Тираж 600 экз.
Зак. № 1547. Изд. № 3552

Редакционно-издательский отдел БелГУТа,
246653, г. Гомель, ул. Кирова, 34. Лицензия ЛВ № 57 от 22.10.1997 г.

Типография БелГУТа,
246022, г. Гомель, ул. Кирова, 34. Лицензия ЛП № 360 от 26.07.1999 г.

Лабораторная работа № 1

ПЕРВИЧНАЯ ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ

Цель работы: ознакомиться с основными понятиями математической статистики и методикой проведения первичного исследования статистических данных.

Задание: выполнить первичную обработку полученных экспериментальных данных и сделать вывод о свойствах изучаемой случайной величины.

Основные теоретические сведения

1 Выборочный метод.

Допустим, результаты изучаемого вероятностного эксперимента могут быть описаны с помощью некоторой случайной величины X. Множество всех возможных значений, которые в результате опыта может принять данная величина в математической статистике принято называть **генеральной совокупностью** и обозначать Ω_X; другими словами, Ω_X – множество всех мыслимых наблюдений, которые могли бы быть зафиксированы при воспроизведении вероятностного эксперимента. Число элементов, образующих множество Ω_X, называется **объемом генеральной совокупности**.

Как правило, в большинстве встречающихся на практике задач, подвергнуть обследованию всю генеральную совокупность не представляется возможным (например, объем Ω_X может быть бесконечным, исследование может оказаться слишком дорогостоящим или приводить к разрушению объекта изучения). Поэтому для исследования свойств случайной величины X чаще всего используется выборочный метод. Сущность этого метода состоит в том, что из рассматриваемой генеральной совокупности случайным образом извлекается часть объектов, называемая **выборкой объема *n***, которая и подвергается детальному изучению. Затем, используя известные теоретико-вероятностные соотношения, по результатам выборочного обследования формулируются выводы о свойствах всей генеральной совокупности.

Можно сказать, что основное назначение математико-статистических методов именно в том и состоит, чтобы с их помощью на основании ограниченного числа выборочных данных получить как можно более полное представление об изучаемых случайных величинах.

Для того, чтобы по имеющейся выборке можно было сделать обоснованный вывод о свойствах всей генеральной совокупности, она должна быть репрезентативной (представительной), т. е. хорошо отображать свойства исследуемой генеральной совокупности. Один из разделов математической статистики посвящен теории получения репрезентативных выборок. В частности, доказано, что для получения представительной выборки выбор каждого элемента из генеральной совокупности должен осуществляться с сохранением принципа случайности.

2 Статистический закон распределения.

Итак, пусть для исследования свойств случайной величины X получена выборка объема n $\{x_1, x_2,\ldots, x_n\}$.

Последовательность наблюденных значений $x_1, x_2,\ldots, x_n$, записанных в порядке их появления, представляет собой исходный статистический материал и называется **простым статистическим рядом**.

Для компактного, удобного и наглядного представления имеющихся статистических данных необходимо произвести их первичную обработку.

Запишем все элементы выборки в порядке неубывания и обозначим члены такой последовательности $x_{(i)}$, $i = 1, 2, ..., n$:

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)},$$

$$x_{(1)} = \min\{x_1, x_2, ..., x_n\},$$

$$x_{(n)} = \max\{x_1, x_2, ..., x_n\}.$$

Каждый элемент $x_{(i)}$ называется ***порядковой статистикой*** (***вариантой***), а последовательность

$$x_{(1)}, x_{(2)}, ..., x_{(n)}$$

называется ***вариационным рядом***, соответствующим имеющейся выборке.

Если изучается *дискретная случайная величина*, число различных наблюденных значений которой не велико, то для каждого из отличающихся друг от друга значений (обозначим их $\widetilde{x}_i$) подсчитываются частоты m_i и относительные частоты (частости) m_i/n появления этих значений в выборке.

Результаты вычислений заносятся в таблицу 1.1, которая называется **сгруппированным статистическим рядом**.

Таблица 1.1 – **Сгруппированный статистический ряд**

$\tilde{x}_i$ – наблюденные значения	$\tilde{x}_1$	$\tilde{x}_2$	…	$\tilde{x}_k$	$k \le n$
m_i – частоты	m_1	m_2	…	m_k	$\sum_{i=1}^{k} m_i = n$
m_i / n – частости	m_1 / n	m_2 / n	…	m_k / n	$\sum_{i=1}^{k} m_i / n = 1$

Если изучается *непрерывная случайная величина* либо дискретная случайная величина, число различных значений которой достаточно велико, то интервал всех наблюденных значений разбивается на k разрядов длины h, и подсчитывается число вариант, попавших в каждый из разрядов. Результаты расчетов заносятся в таблицу 1.2, которая называется **интервальным статистическим рядом**.

Таблица 1.2 – **Интервальный статистический ряд**

Границы интервалов	$[C_1, C_2)$	$[C_2, C_3)$	…	$[C_k, C_{k+1})$	
Среднее значение интервала $\tilde{x}_i$	$\tilde{x}_1$	$\tilde{x}_2$	…	$\tilde{x}_k$	
Частоты m_i	m_1	m_2	…	m_k	$\sum_{i=1}^{k} m_i = n$
Частости m_i / n	m_1 / n	m_2 / n	…	m_k / n	$\sum_{i=1}^{k} m_i / n = 1$

Для определения границ частичных интервалов $[C_i, C_{i+1})$ можно воспользоваться следующей методикой:

1 Вычислить размах варьирования наблюденных значений: $R = x_{\max} - x_{\min}$, где $x_{\min}$ и $x_{\max}$, соответственно, минимальное и максимальное значения вариационного ряда.

2 Определить длину шага разбиения $h = \dfrac{R}{k}$, где k – число разрядов разбиения. Для примерной ориентации в выборе значения k можно воспользоваться формулой Стерджесса: $k \approx 1 + 3{,}322 \lg n$ ($5 \le k \le 15$), где n – объем выборки. (Для упрощения последующих расчетов полученное значение h может быть несколько округлено в бóльшую или меньшую сторону).

3 Определить границы интервалов разбиения: $C_1 = x_{\min} - h/2$, $C_2 = C_1 + h$, $C_3 = C_2 + h$, и т. д. Процесс разбиения продолжается до тех пор, пока максимальный элемент выборки не попадет в интервал. Среднее значение каждого частичного интервала можно определить как среднее арифметическое его границ.

Элементы выборки, попавшие на границы разрядов разбиения, могут быть приписаны к какому-то одному из этих интервалов (например, к правому, как это сделано в таблице 1.2), либо частоты этих значений могут быть

разделены поровну между двумя соседними интервалами.

3 Графическое изображение статистического закона распределения.

Для графического представления сгруппированного статистического ряда обычно используется ***столбцовая диаграмма*** (рисунок 1.1), которая представляет собой последовательность вертикальных отрезков длины m_i/n, отложенных от оси абсцисс в точках с координатами $\tilde{x}_i$.

Для графического изображения интервального статистического ряда чаще всего используется *гистограмма относительных частот* (рисунок 1.2). При построении гистограммы на оси абсцисс необходимо отложить границы интервалов наблюденных значений $[C_i, C_{i+1})$ ($i = \overline{1, k}$) и на каждом из этих интервалов, как на основании, построить прямоугольники, площади которых равны m_i/n. Высоты прямоугольников равны $m_i/(nh_i)$.

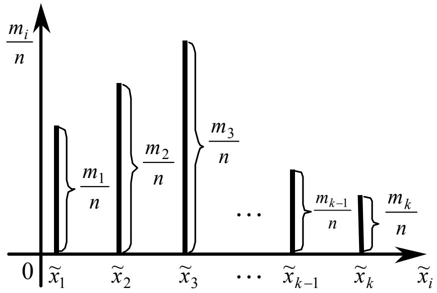


Рисунок 1.1 – Столбцовая диаграмма

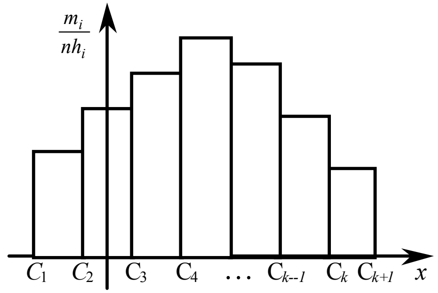


Рисунок 1.2 – Гистограмма относительных частот

4 Эмпирическая функция распределения.

Эмпирической функцией распределения $\hat{F}(x)$ называется функция, которая каждому значению x ставит в соответствие относительную частоту события $\{X < x\}$, т. е. $\hat{F}(x) = \dfrac{n_x}{n}$, где n – объём исследуемой выборки, n_x – число вариант выборки, меньших данного фиксированного значения x.

Для определения эмпирической функции распределения по данным сгруппированного статистического ряда можно использовать формулу

$$\hat{F}(x) = \sum_{\tilde{x}_i < x} \frac{m_i}{n}.$$

Важнейшее свойство эмпирической функции распределения состоит в том, что при увеличении объема выборки n, значение этой функции в каждой точке приближается к значению функции распределения $F(x)$ в указанной точке, т. е. эмпирическая функция распределения является экспериментальным аналогом (оценкой) неизвестной исследователю функции распределения.

5 Оценки числовых характеристик.

Важнейшим этапом обработки статистических данных является вычисление оценок числовых характеристик исследуемой случайной величины. Полученные оценки позволяют в числовой форме описать характерные черты статистического распределения и являются базой для построения математической модели изучаемого случайного явления.

Любая величина $\tilde{\Theta}$, определяемая как функция выборочных значений $\tilde{\Theta} = U(x_1, x_2, ..., x_n)$, называется ***выборочной статистикой*** или просто ***статистикой***. Статистика $\tilde{\Theta}$, используемая в качестве приближённого значения неизвестного параметра Θ, называется ***статистической оценкой параметра*** Θ.

Существует два вида оценок параметров: точечные и интервальные. ***Точечные оценки*** определяют точку $\tilde{\Theta}$, являющуюся некоторым приближением оцениваемого параметра Θ. ***Интервальная оценка*** – это интервал $(\tilde{\Theta}_1, \tilde{\Theta}_2)$, который с заданной вероятностью содержит неизвестное значение параметра Θ.

Ниже приведены формулы для вычисления точечных оценок случайной величины на основании имеющихся выборочных данных.

1 В качестве оценки ***математического ожидания*** используется среднее арифметическое $\bar{x}$ наблюденных значений. Эта статистика называется **выборочным средним**.

$$\tilde{M}[X] = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

2 Для оценивания по выборочным данным моды распределения, используется то значение сгруппированного статистического ряда $\tilde{x}_{\text{mod}}$, которому соответствует наибольшее значение частоты. По интервальному статистическому ряду определяется модальный интервал, в который попало наибольшее число элементов выборки, и в качестве точечной оценки моды может использоваться среднее значение этого интервала.

3 Для определения выборочного значения медианы используется вариационный ряд. В качестве оценки медианы $\tilde{x}_{\text{med}}$ принимают средний (т. е. $\frac{1}{2}(n+1)$-й) член этого ряда, если значение n – нечётно и среднее арифметическое между двумя средними (т. е. между $\frac{1}{2}n$-м и $\left(\frac{1}{2}n+1\right)$-м) членами этого ряда, если n – чётно.

4 В качестве оценки дисперсии используется статистика

$$\tilde{D}[X] = \tilde{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

5 Оценка среднего квадратического отклонения

$$\tilde{\sigma}[X] = \sqrt{\tilde{D}[X]} .$$

6 Оценка коэффициента вариации

$$\tilde{V} = \frac{\tilde{\sigma}}{\bar{x}} .$$

7 Оценка коэффициента асимметрии

$$\tilde{A}[X] = \tilde{\beta}_1 = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\tilde{\sigma}^3} .$$

8 Оценка коэффициента эксцесса

$$\tilde{E}x[X] = \tilde{\beta}_2 = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\tilde{\sigma}^4} - 3 .$$

Пример 1.1. При изучении характеристик работы цеха, в результате наблюдений получена выборка значений случайной величины *X*, обозначающей число отказов оборудования в течение рабочей смены: 3, 1, 0, 1, 3, 6, 2, 2, 1, 1, 4, 4, 1, 4, 2, 0, 1, 5, 3, 2, 2, 1, 3, 1, 1, 4, 2, 1, 1, 3, 1, 2, 1, 0, 2, 1, 2, 0, 1, 1, 1, 4, 4, 2, 1, 4, 1, 2, 3, 2.

Задание: исследовать полученные экспериментально данные с целью изучения свойств случайной величины *X*.

Решение. 1 Построим вариационный ряд, соответствующий полученной выборке: 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 6.

2 Поскольку изучаемая случайная величина является дискретной, ее статистический закон распределения запишем в виде сгруппированного статистического ряда:

$\tilde{x}_i$	0	1	2	3	4	5	6
m_i	4	19	12	6	7	1	1
m_i / n	0,08	0,38	0,24	0,12	0,14	0,02	0,02

Столбцовая диаграмма, построенная по данным сгруппированного статистического ряда, изображена на рисунке 1.3.

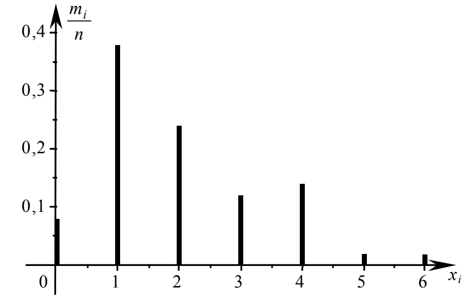


Рисунок 1.3 – Столбцовая диаграмма (пример 1.1)

3 Построим эмпирическую функцию распределения исследуемой случайной величины:

$$\tilde{F}(x)=\begin{cases} 0 & \text{при} \quad x\le 0;\\ 0{,}08 & \text{при} \quad 0<x\le 1;\\ 0{,}46 & \text{при} \quad 1<x\le 2;\\ 0{,}70 & \text{при} \quad 2<x\le 3;\\ 0{,}82 & \text{при} \quad 3<x\le 4;\\ 0{,}96 & \text{при} \quad 4<x\le 5;\\ 0{,}98 & \text{при} \quad 5<x\le 6;\\ 1 & \text{при} \quad x>6. \end{cases}$$

График функции $\tilde{F}(x)$ изображён на рисунке 1.4.

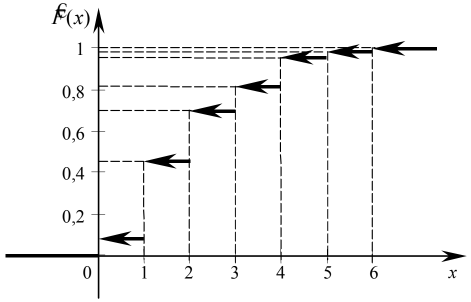


Рисунок 1.4 – График эмпирической функции распределения (пример 1.1)

4 Вычислим точечные оценки числовых характеристик случайной величины X.

Оценка математического ожидания

$$\tilde{M}[X]=\bar{x}=\frac{1}{n}\sum_{i=1}^{7}\tilde{x}_i m_i=$$

$$=\frac{1}{50}(0\cdot 4+1\cdot 19+2\cdot 12+3\cdot 6+4\cdot 7+5\cdot 1+6\cdot 1)=2\,[\text{отказа}].$$

Оценка моды данной случайной величины равна 1 ($\tilde{x}_{\text{mod}}=1$), так как этому значению соответствует наибольшее значение частоты.

Оценку медианы определим на основании вариационного ряда

$$\tilde{x}_{\text{med}}=\frac{x_{(25)}+x_{(26)}}{2}=\frac{2+2}{2}=2\,[\text{отказа}].$$

Оценка дисперсии

$$\tilde{D}[X]=\tilde{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{7}(\tilde{x}_i-\bar{x})^2 m_i=\frac{1}{49}((0-2)^2\cdot 4+(1-2)^2\cdot 19+(2-2)^2\cdot 12+$$

$$+(3-2)^2\cdot 6+(4-2)^2\cdot 7+(5-2)^2\cdot 1+(6-2)^2\cdot 1)=1{,}9184\,[\text{отказа}^2].$$

Оценка среднего квадратического отклонения

$$\tilde{\sigma}[X]=\sqrt{\tilde{D}[X]}=1{,}385\,[\text{отказа}].$$

Оценка коэффициента вариации

$$\tilde{V}=\frac{\tilde{\sigma}}{\bar{x}}=\frac{1{,}385}{2}=0{,}6925.$$

Выводы. Итак, мы выполнили первичную обработку выборки значений дискретной случайной величины, характеризующей число отказов оборудования в течение рабочей смены. Объём выборки $n=50$. Минимальное число отказов оборудования в течение рабочей смены равно 0, максимальное – 6. Среднее значение числа отказов равно 2; наиболее вероятное – 1, а средневероятное – 2. Среднеквадратическое отклонение числа отказов оборудования от среднего значения равно 1,385. Коэффициент вариации – 0,6925.

Пример 1.2. Исследуется работа технического устройства, которое может безотказно функционировать до тех пор, пока регистрируемые параметры его работы не превысят допустимые значения. При выходе параметров работы устройства из области допустимых значений производится переналадка этого устройства, после чего оно снова продолжает свою работу. При проведении экспериментов фиксировались значения случайной величины X, характеризующей время безотказной работы устройства между двумя последовательными переналадками.

В приведённых экспериментальных данных значения времени безотказной работы выражены в часах:

6,421; 5,034; 0,599; 10,687; 26,294; 7,852; 14,040; 8,933; 4,062; 1,573;
5,455; 2,810; 15,658; 3,692; 1,825; 17,760; 8,030; 3,218; 2,872; 8,247;
0,417; 1,995; 0,611; 12,059; 0,665; 21,434; 22,102; 10,709; 2,283; 5,649;
6,773; 30,034; 3,702; 12,834; 2,723; 5,255; 12,595; 3,533; 34,540; 9,238;
1,673; 24,919; 15,511; 7,154; 1,816; 8,401; 22,239; 2,902; 2,083; 7,176.

Задание: произвести первичную обработку полученных опытных данных с целью изучения свойств случайной величины X.

Р е ш е н и е. 1 Построим вариационный ряд:

0,417; 0,599; 0,61; 0,665; 1,573; 1,673; 1,816; 1,825; 1,995; 2,083;
2,283; 2,723; 2,810; 2,872; 2,902; 3,218; 3,533; 3,692; 3,703; 4,062;
4,176; 5,034; 5,255; 5,455; 5,649; 6,421; 6,773; 7,154; 7,852; 8,030;
8,247; 8,401; 8,933; 9,238; 10,687; 10,709; 12,059; 12,595; 12,834; 14,040;
15,511; 15,658; 17,760; 28,434; 22,102; 22,239; 24,919; 26,294; 30,034; 34,540.

Для записи статистического закона распределения данной непрерывной случайной величины построим интервальный статистический ряд.

Определим длину интервала

$$h=\frac{R}{k}=\frac{x_{\max}-x_{\min}}{1+3{,}322\lg n}=\frac{34{,}54-0{,}417}{1+3{,}322\lg 50}=5{,}14\,.$$

Определив границы интервалов разбиения ($C_1=x_{\min}-h/2=0{,}417-2{,}57=$ $=-2{,}153$, $C_2=C_1+h=-2{,}153+5{,}14=2{,}987$, $C_3=C_2+h=2{,}987+5{,}14=8{,}127$ и т. д.), построим интервальный статистический ряд:

$[C_i, C_{i+1})$	[–2,153; 2,987)	[2,987; 8,127)	[8,127; 13,267)	[13,267; 18,407)	[18,407; 23,547)	[23,547; 28,687)	[28,687; 33,827)	[33,827; 38,967)
$\tilde{x}_i$	0,417	5,557	10,697	15,837	20,977	26,117	31,257	36,397
m_i	15	15	9	4	3	2	1	1
m_i / n	0,3	0,3	0,18	0,08	0,06	0,04	0,02	0,02

(Для контроля убедимся, что $\sum_{i=1}^{k} m_i = n = 50$, $\sum_{i=1}^{k} m_i / n = 1$).

Графическое изображение интервального статистического ряда приведено на рисунке 1.5.

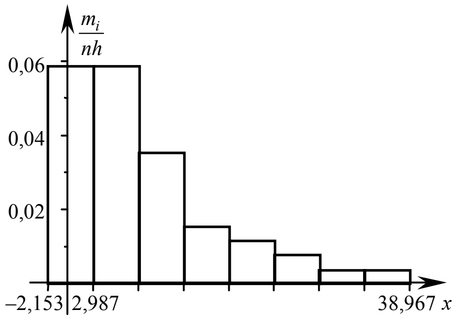


Рисунок 1.5 – Гистограмма относительных частот (пример 1.2)

2 Для приближённого построения эмпирической функции распределения воспользуемся соотношением $\hat{F}[x]=\sum_{\tilde{x}_i<x}\frac{m_i}{n}$ *:

$$\hat{F}(x)=\begin{cases} 0 & \text{при} \quad x\le 0{,}417;\\ 0{,}3 & \text{при} \quad 0{,}147<x\le 5{,}557;\\ 0{,}6 & \text{при} \quad 5{,}557<x\le 10{,}697;\\ 0{,}78 & \text{при} \quad 10{,}697<x\le 15{,}837;\\ 0{,}86 & \text{при} \quad 15{,}837<x\le 20{,}977;\\ 0{,}92 & \text{при} \quad 20{,}977<x\le 26{,}117;\\ 0{,}96 & \text{при} \quad 26{,}117<x\le 31{,}257;\\ 0{,}98 & \text{при} \quad 31{,}257<x\le 36{,}397;\\ 1 & \text{при} \quad x>36{,}397. \end{cases}$$

* При использовании этой формулы принимаем допущение о том, что исследуемая случайная величина принимает только значения, соответствующие серединам интервалов $[C_i; C_{i+1})$ с частотами, равными m_i ($i=1, 2, \ldots, k$).

График эмпирической функции распределения приведён на рисунке 1.6, *а*.

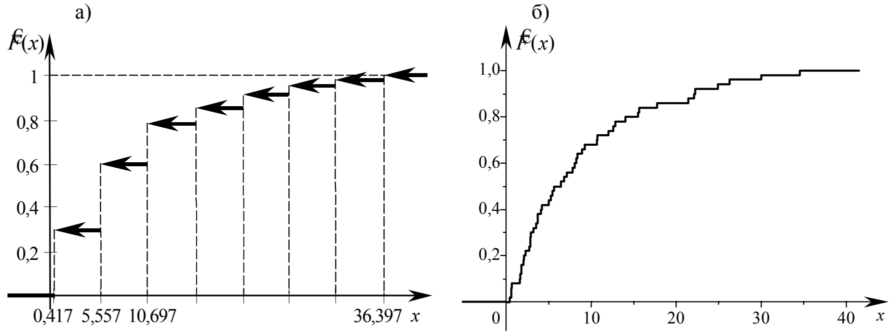


Рисунок 1.6 – График эмпирической функции распределения (пример 1.2)

На рисунке 1.6, *б* приведён график эмпирической функции распределения данной случайной величины, построенный при использовании всего набора имеющихся экспериментальных данных.

Вычислим точечные оценки числовых характеристик случайной величины, обозначающей время безотказной работы оборудования между двумя последовательными переналадками.

$$\hat{M}[X]=\bar{x}=\frac{1}{n}\sum_{i=1}^{50}x_i=8{,}9017 \ \text{[ч]};$$

$$\hat{D}[X]=\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{50}(x_i-\bar{x})^2=69{,}5796 \ [\text{ч}^2];$$

$$\hat{\sigma}[X]=\hat{\sigma}=8{,}34 \ \text{[ч]};$$

$$\hat{V}=\frac{\hat{\sigma}}{\bar{x}}\approx 0{,}937;$$

$$\hat{A}[X]=\hat{\beta}_1=\frac{1}{(n-1)\hat{\sigma}^3}\sum_{i=1}^{50}(x_i-\bar{x})^3=1{,}3626\,;$$

$$\hat{E}_x[X]=\hat{\beta}_2=\frac{1}{(n-1)\hat{\sigma}^4}\sum_{i=1}^{50}(x_i-\bar{x})^4-3=1{,}3559\,.$$

В качестве оценки моды можно принять среднее значение между двумя модальными интервалами [–2,153; 2,987) и [2,987; 8,127): $\hat{x}_{\text{mod}}=2{,}987$ ч.

В качестве оценки медианы примем среднее значение между 25-м и 26-м элементами вариационного ряда: $\hat{x}_{\text{med}} = (5{,}649 + 6{,}421)/2 = 6{,}035$ ч.

Выводы. В результате исследования выборки значений непрерывной случайной величины, характеризующей время безотказной работы устройства между двумя последовательными переналадками, получили следующие результаты, ч: минимальное время безотказного функционирования – 0,417, максимальное – 34,54, среднее значение времени безотказного функционирования устройства – 8,901, наиболее вероятное значение – 2,987, средневероятное – 6,035. Среднеквадратическое отклонение времени безотказного функционирования устройства от среднего значения – 8,34.

Порядок выполнения работы

1 Изучить теоретические сведения.

2 Получить у преподавателя выборку значений исследуемой случайной величины.

3 Произвести вручную первичную обработку статистических данных:
- построить вариационный ряд;
- построить статистический закон распределения и его графическое изображение;
- вычислить эмпирическую функцию распределения и построить график этой функции;
- вычислить точечные оценки числовых характеристик изучаемой случайной величины.

4 Произвести первичную обработку полученной выборки с помощью ЭВМ:
- записать выборку на диск (приложение А, пп. 1, 2);
- вычислить оценки числовых характеристик (приложение А, п. 3);
- построить графическое изображение статистического закона распределения (приложение А, п. 7).

5 Сравнить результаты, полученные при ручном расчёте и расчёте на ЭВМ.

6 Сделать обоснованный вывод о свойствах изучаемой случайной величины.

Лабораторная работа № 2

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ПАРАМЕТРИЧЕСКИХ ГИПОТЕЗ

Цель работы: изучить основные понятия теории статистической проверки гипотез и проверки параметрических гипотез.

Задание: по заданным выборкам провести статистическую проверку:
- гипотез о равенстве математического ожидания некоторому фиксированному значению;
- гипотезы о равенстве математических ожиданий двух генеральных совокупностей.

Основные теоретические сведения

1 Основные понятия.

На практике часто решаются задачи сравнения технологических процессов по различным параметрам: производительности, экономичности, точности достижения определенной технической характеристики и т. д. На языке математической статистики такие задачи формулируются как задачи сравнения или статистической проверки гипотез относительно параметров распределения. Рассмотрим, например, случайную величину X – время пребывания вагона на сортировочной станции. Пусть с. в. X имеет функцию распределения $F(x, \theta)$, где θ – некоторый параметр распределения. Предположим, что с целью уменьшения времени простоя вагонов на станции введены усовершенствования в технологический процесс обработки составов. Если мы возьмем одну выборку наблюденных значений с. в. X, полученную до внесения изменений в технологический процесс, а другую – после изменений, и вычислим две точечные оценки $\hat{\theta}_1$ и $\hat{\theta}_2$ параметра θ распределения $F(x, \theta)$, то они, очевидно, не будут равны между собой. Спрашивается, связано ли это с изменениями, внесенными в технологический процесс, либо разница в значениях оценок обусловлена случайными факторами?

Статистическая гипотеза называется ***непараметрической***, если в ней сформулировано предположение о виде закона распределения исследуемой случайной величины.

Статистическая гипотеза называется ***параметрической***, если в ней сформулированы предположения относительно значений параметров распределения.

Суждения относительно истинности (ложности) статистических гипотез формулируются на основании выборки объема n. Наряду с выдвинутой гипотезой будем рассматривать одну или несколько альтернативных (конкурирующих) гипотез.

Нулевой гипотезой называют выдвинутую гипотезу и обозначают H_0. Обычно нулевые гипотезы утверждают, что различия между сравниваемыми величинами (параметрами или функциями распределения) отсутствуют, а наблюдаемые отклонения обусловлены случайными колебаниями выборки.

Альтернативной называется гипотеза, конкурирующая с нулевой гипотезой в том смысле, что если отвергается нулевая гипотеза, то принимается альтернативная. Ее обозначают H_a.

Пусть для решения задачи сравнения из генеральной совокупности извлечена выборка $\{x_1, x_2, ..., x_n\}$ объема n. Предположим, экспериментатор по каким-либо соображениям выдвинул гипотезу о виде распределения исследуемой случайной величины X. Пусть по выборке $\{x_1, x_2, ..., x_n\}$ выдвинута, например, гипотеза о нормальном распределении генеральной совокупности с математическим ожиданием a и среднеквадратическим отклонением σ.

Приведем примеры нулевых и альтернативных параметрических гипотез (таблица 2.1).

Таблица 2.1 – **Примеры нулевых и альтернативных параметрических гипотез**

Нулевые	Альтернативные
$H_0: \ a = a_0$	$H_a: \ a \neq a_0$, или $a < a_0$, или $a > a_0$;
$H_0: \ \sigma = \sigma_0$	$H_a: \ \sigma \neq \sigma_0$, или $\sigma < \sigma_0$, или $\sigma > \sigma_0$;

Статистическим критерием называется случайная величина R, с помощью которой принимается решение о принятии либо отклонении нулевой гипотезы. Для проверки статистических гипотез по критериям значимости необходимо знать условный закон распределения построенной случайной величины в предположении выполнения нулевой гипотезы.

При проверке статистических гипотез по выборочным данным всегда существует возможность принятия ложного решения. Это объясняется тем, что объем выборки конечен, и поэтому нельзя точно определить ни вид функции распределения, ни значения параметров.

Ошибкой первого рода называется ошибка отклонения верной нулевой гипотезы H_0.

Уровнем значимости α ***статистического критерия*** называется вероятность совершения ошибки первого рода.

Ошибкой второго рода называется принятие ложной нулевой гипотезы H_0. Вероятность совершения ошибки второго рода принято обозначать буквой β.

Мощностью M статистического критерия R называется вероятность несовершения ошибки второго рода, т. е. $M = 1 - \beta$.

Будем рассматривать только один вид статистических критериев – ***статистические критерии значимости***. Это значит, что будет заранее фиксироваться вероятность совершения ошибки первого рода (***уровень значимости*** α). На фиксированном уровне значимости мы можем принять только одно из двух решений: «отклонить проверяемую нулевую гипотезу» или «результаты выборки не дают основания для отклонения нулевой гипотезы».

Зададим ***уровень значимости*** α – вероятность настолько малую, что событие с такой вероятностью будет практически невозможным. Обычно выбирают значения $\alpha = 0{,}05$; $\alpha = 0{,}01$. Например, отклонение нулевой гипотезы при уровне значимости $\alpha = 0{,}05$ означает, что мы либо не ошибаемся, т. е. нулевая гипотеза H_0 действительно ложна, либо мы совершаем ошибку первого рода, отклоняя верную гипотезу H_0, причем последнее происходит в среднем в пяти из ста случаев применения данного статистического критерия.

Пусть для проверки некоторой нулевой гипотезы H_0 относительно параметров распределения служит статистический критерий R, для которого известна плотность $f(x \mid H_0)$ распределения вероятностей при условии, что нулевая гипотеза верна. Найдем $R_{1-\alpha/2}$ и $R_{\alpha/2}$ – критические точки (квантили) распределения R из условий

$$P\{R < R_{1-\alpha/2}\} = \alpha/2; \quad P\{R > R_{\alpha/2}\} = \alpha/2.$$

Область $(R_{1-\alpha/2}; R_{\alpha/2})$ называется областью допустимых значений случайной величины R. Решающее правило состоит в следующем: если вычисленное по выборке значение статистики R попадает в критическую область, то нулевая гипотеза отклоняется, в противном случае результаты выборки не дают основания для отклонения нулевой гипотезы. Следует отметить, что чем меньше уровень значимости, тем шире область допустимых значений выборочной статистики. При выборе $\alpha = 0$ область допустимых значений будет представлять всю числовую прямую, и проверка гипотезы в таком случае лишена смысла.

Критическая область, изображенная на рисунке 2.1, называется двусторонней критической областью. Иногда используются односторонние критические области, если экспериментатор убежден, что $R > R_0$ или $R < R_0$ (рисунок 2.2).

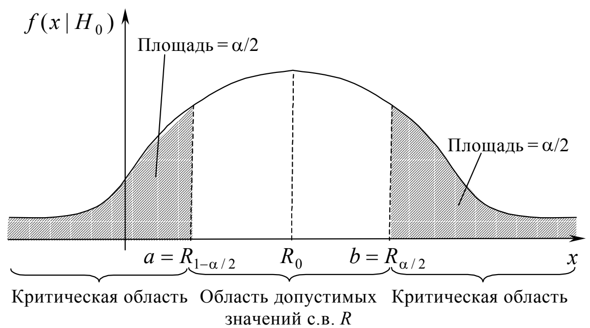


Рисунок 2.1 – Двусторонняя критическая область статистического критерия

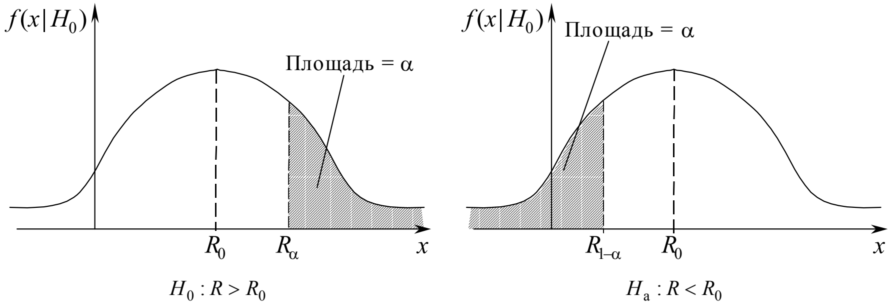


Рисунок 2.2 – Односторонние критические области проверки статистического критерия

2 Проверка гипотез о математическом ожидании случайной величины, имеющей нормальное распределение.

Пусть случайная величина X имеет нормальное распределение с параметрами a и σ. Требуется на основании выборки $\{x_1, x_2, ..., x_n\}$ проверить гипотезу о том, что математическое ожидание с. в. равно некоторому предполагаемому значению a_0:

$$H_0: a = a_0.$$

Можно рассмотреть два случая.

1 Известно значение среднеквадратического отклонения σ.

Для проверки гипотезы используется критерий значимости

$$u = \frac{\bar{x} - a_0}{\sigma/\sqrt{n}}.$$

Доказано, что если гипотеза H_0 верна, то с. в. u имеет нормальное распределение с параметрами (0; 1), т. е. $u \sim N(0; 1)$.

В таблице 2.2 приведены условия отклонения нулевой гипотезы в зависимости от вида альтернативной гипотезы, где u_α и $u_{\alpha/2}$ – квантили стандартного нормального распределения для вероятностей α и α/2 соответственно. Значения квантилей стандартного нормального распределения находят по таблице значений функции Лапласа (приложение В). Некоторые из них приведены в таблице 2.3. Отметим, что в силу симметричности стандартного нормального распределения $u_{1-\alpha} = -u_\alpha$.

Таблица 2.2 – **Условия отклонения гипотезы о равенстве математического ожидания заданному значению при известном среднеквадратическом отклонении**

Альтернативная гипотеза H_a	Условие отклонения нулевой гипотезы в пользу альтернативной
$a \neq a_0$	$u \le -u_{\alpha/2}$ либо $u \ge u_{\alpha/2}$
$a > a_0$	$u \ge u_\alpha$
$a < a_0$	$u \le -u_\alpha$

Таблица 2.3 – **Значение квантилей стандартного нормального распределения**

α	u_α	α	u_α	α	u_α	α	u_α
0,005	2,5758	0,025	1,9600	0,950	–1,6449	0,990	–2,3263
0,010	2,3263	0,050	1,6449	0,975	–1,9600	0,995	–2,5758

2 Значение среднеквадратического отклонения σ неизвестно.

Для проверки гипотезы используется следующий критерий значимости:

$$t = \frac{\bar{x} - a_0}{\tilde{S}/\sqrt{n-1}}.$$

Известно, что если гипотеза H_0 верна, то с. в. t имеет распределение Стьюдента с $\nu = n - 1$ степенями свободы, т. е. $t \sim T(\nu)$.

В таблице 2.4 приведены условия отклонения нулевой гипотезы в зависимости от вида альтернативной гипотезы.

Таблица 2.4 – **Условия отклонения гипотезы о равенстве математического ожидания заданному значению при неизвестном среднеквадратическом отклонении**

Альтернативная гипотеза H_a	Условие отклонения нулевой гипотезы в пользу альтернативной
$a \neq a_0$	$t \le -t_{\alpha/2;\, n-1}$ либо $t \ge t_{\alpha/2;\, n-1}$
$a > a_0$	$t \ge t_{\alpha;\, n-1}$
$a < a_0$	$t \le -t_{\alpha;\, n-1}$

Пример 2.1. В графике движения поездов на участке Гомель – Жлобин на основании тяговых расчетов установлено время следования поездов по участку $a_0 = 121$ мин. Известно, что случайная величина – время следования поездов на перегоне – имеет нормальное распределение со среднеквадратическим отклонением $\sigma = 5{,}1$ мин. Инженер-графист отдела перевозок провел анализ графика исполненного движения поездов в течение 6 суток и установил затраты времени для 40 поездов. В результате расчетов выборочное среднее времени следования поезда на участке Гомель – Жлобин составило 115 мин.

Можно ли утверждать, что нормы времени на следование поездов на указанном участке, принятые на основании тяговых расчетов, завышены, и поэтому следует откорректировать времена хода по перегонам участка в нормативном графике?

Из условия следует, что необходимо проверить нулевую гипотезу H_0: $a = 121$ мин против альтернативной гипотезы H_a: $a < 121$ мин. Так как из условия известно, что среднеквадратическое отклонение исследуемой случайной величины $\sigma = 5{,}1$ мин, то для проверки нулевой гипотезы воспользуемся критерием u. Вид альтернативной гипотезы указывает на то, что следует воспользоваться критерием с левосторонней критической областью. Примем $\alpha = 0{,}05$. Вычислим значение критерия значимости

$$u = \frac{\bar{x} - a_0}{\sigma/\sqrt{n}} = \frac{115 - 121}{5{,}1/\sqrt{40}} \approx -7{,}44 .$$

В таблице 2.3 квантилей стандартного нормального распределения по уровню значимости $\alpha = 0{,}05$ находим критическую точку $-u_\alpha = -u_{0{,}05} = -1{,}64$. Поскольку $u = -7{,}4 \le -u_\alpha = -1{,}64$, то на заданном уровне значимости нулевая гипотеза отклоняется. Это означает, что с вероятностью ошибки, меньшей 0,05, можно утверждать, что тяговыми расчетами установлены завышенные нормы времени на следование поездов на указанном участке и поэтому следует откорректировать нормативный график движения поездов.

Пример 2.2. Технологией развоза местного груза на участке Минск – Молодечно предусмотрена норма времени стоянки сборного поезда на станции Уша для выполнения операций прицепки-отцепки вагонов в 45 мин. В отдел перевозок поступил доклад начальника станции Уша, в котором утверждается, что в связи с увеличением числа местных вагонов для станции фактическое время стоянки превысило нормативное. Начальник станции предлагает пересмотреть норму времени стоянки сборного поезда в сторону увеличения. Инженер-технолог отдела перевозок проанализировал отчетные данные времени стоянки 30 сборных поездов на станции Уша за последний месяц. Как на основании собранных статистических данных дать аргументированный ответ на предложение начальника станции?

Пусть по выборке установлено, что выборочное среднее $\hat{a} = \bar{x} = 47$ мин и выборочное среднеквадратическое отклонение $\hat{\sigma} = 6{,}7$ мин. Зададим уровень значимости $\alpha = 0{,}01$.

Из условия следует, что необходимо проверить нулевую гипотезу H_0: $a = 45$ мин против альтернативной гипотезы H_a: $a > 45$ мин. Так как точное значение среднеквадратического отклонения неизвестно и альтернативная гипотеза H_a: $a > a_0$, воспользуемся t-критерием с правосторонней критической областью. Вычислим значение t-критерия

$$t = \frac{\bar{x} - a_0}{\hat{\sigma}/\sqrt{n-1}} = \frac{47 - 45}{6{,}7/\sqrt{30-1}} \approx 1{,}608.$$

В таблице квантилей распределения Стьюдента (приложение Д) по уровню значимости $\alpha = 0{,}01$ и числу степеней свободы $\nu = n - 1 = 29$ находим $t_{\alpha;\, n-1} = t_{0{,}01;\, 29} = 2{,}462$. Так как $t = 1{,}608 < t_{0{,}01;\, 29} = 2{,}462$, то имеющиеся данные не дают оснований для отклонения нулевой гипотезы, т. е. нет оснований пересматривать в сторону увеличения норму времени стоянки сборного поезда на ст. Уша.

3 Проверка гипотез равенства двух случайных величин, имеющих нормальное распределение.

Пусть исследуются две случайные величины X и Y, причем обычно предполагается, что обе они имеют нормальное распределение: $X \sim N(a_1, \sigma_1)$, $Y \sim N(a_2, \sigma_2)$.

В качестве нулевой гипотезы рассмотрим гипотезу о равенстве математических ожиданий исследуемых величин H_0: $a_1 = a_2$. В таблице 2.5 приведены решающие правила для проверки подобного рода гипотез:

Таблица 2.5 – **Решающие правила для проверки гипотез о равенстве математических ожиданий двух случайных величин**

Дисперсия	Критерий	Распределение критерия	Альтернативная гипотеза	Условие отклонения нулевой гипотезы
σ_1 и σ_2 известны	$u = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	Стандартное нормальное $u \sim N(0;1)$	$H_a: \; a_1 \ne a_2$	$u \le -u_{\alpha/2}$ или $u \ge u_{\alpha/2}$
			$H_a: \; a_1 < a_2$	$u \le -u_\alpha$
			$H_a: \; a_1 > a_2$	$u \ge u_\alpha$
σ_1 и σ_2 неизвестны, $\sigma_1 = \sigma_2$	$t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{n_1\hat{\sigma}_1^2 + n_2\hat{\sigma}_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	Стьюдента с $\nu = n_1 + n_2 - 2$ степенями свободы $t \sim T(\nu)$	$H_a: \; a_1 \ne a_2$	$t \le -t_{\alpha/2;\, n_1+n_2-2}$ или $t \ge t_{\alpha/2;\, n_1+n_2-2}$
			$H_a: \; a_1 < a_2$	$t \le -t_{\alpha;\, n_1+n_2-2}$
			$H_a: \; a_1 > a_2$	$t \ge t_{\alpha;\, n_1+n_2-2}$

Пример 2.3. На станции Могилёв-2 произведена реконструкция сортировочной горки: механизирована операция торможения вагонов при скатывании их с горки на пути сортировочного парка, установлены две централизованные тормозные позиции. Необходимо оценить эксплуатационную эффективность новой технологии по сравнению с прежней, когда торможение производилось вручную регулировщиками скорости движения вагонов. С этой целью собраны данные о дополнительных затратах времени за сутки на осаживание вагонов в сортировочном парке. Сбор данных производился в течение 60 дней до реконструкции и 60 дней после реконструкции.

Предположим, что дополнительные затраты времени за сутки на осаживание вагонов в сортировочном парке по старой и новой технологиям соответственно являются нормально распределенными случайными величинами $X \sim N(a_1, \sigma_1)$, $Y \sim N(a_2, \sigma_2)$, причем a_1 и a_2 неизвестны, σ_1 и σ_2 неизвестны, но предполагается $\sigma_1 = \sigma_2$.

Пусть также по имеющимся выборкам получены точечные оценки математических ожиданий и среднеквадратических отклонений этих случайных величин $\bar{x} = 170$ мин, $\hat{\sigma}_1 = 17{,}3$ мин, $\bar{y} = 160$ мин, $\hat{\sigma}_2 = 16{,}5$ мин. ($\bar{x}$ и $\bar{y}$ – точечные оценки a_1 и a_2 соответственно).

Необходимо проверить нулевую гипотезу о равенстве математических ожиданий H_0: $a_1 = a_2$ против альтернативной H_a: $a_1 > a_2$. Найдем значение статистики:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{n_1\hat{\sigma}_1^2 + n_2\hat{\sigma}_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{170 - 160}{\sqrt{\dfrac{60 \cdot 17{,}3^2 + 60 \cdot 16{,}7^2}{60 + 60 - 2}\left(\dfrac{1}{60} + \dfrac{1}{60}\right)}} \approx 3{,}194.$$

По таблицам квантилей распределения Стьюдента (см. приложение *Д*) для числа степеней свободы $\nu = n_1 + n_2 - 2 = 60 + 60 - 2 = 118$ находим $t_{0,05;\,118} \approx$ $\approx 1{,}658$. Так как $t = 3{,}194 > t_{0,05;\,118} = 1{,}658$, нулевая гипотеза отклоняется на заданном уровне значимости, т. е. собранные данные позволяют утверждать, что внедрение новой технологии привело к уменьшению затрат времени за сутки на осаживание вагонов в сортировочном парке.

Порядок выполнения работы

1 Изучить теоретические сведения.

2 Записать на диск две выборки случайных величин, которые требуется исследовать (приложение А, пп. 1, 2).

3 По одной из выборок и заданному значению математического ожидания провести проверку гипотезы о равенстве математического ожидания заданному значению против предложенного вида альтернативной гипотезы:
- вручную рассчитать значение выборочной статистики и сравнить его с критическим значением;
- с помощь процедуры «One-Sample Analysis» пакета Statgrafics (приложение А, п. 5)

4 По двум выборкам провести проверку гипотезы о равенстве математических ожиданий двух нормально распределенных случайных величин против предложенного вида альтернативной гипотезы:
- вручную рассчитать значение выборочной статистики и сравнить его с критическим значением;
- с помощь процедуры «Two-Sample Analysis» пакета Statgrafics (приложение А, п. 6).

5 Сделать выводы.

Лабораторная работа № 3

ПОДБОР ЗАКОНА РАСПРЕДЕЛЕНИЯ ОДНОМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Цель работы: изучить методику применения критерия χ^2 Пирсона для проверки гипотезы о виде закона распределения случайной величины.

Задание: с помощью критерия χ^2 проверить согласование выдвинутой гипотезы о виде закона распределения исследуемой случайной величины с имеющимися выборочными данными.

Основные теоретические сведения

1 Введение.

При проведении статистического исследования экспериментальных данных часто возникает задача подбора закона распределения изучаемой случайной величины *X*.

Таблица 3.1 – **Основные сведения о наиболее часто встречающихся на практике законах**

Название закона распределения	Возможные значения	Пара-метры	Статистическая оценка параметров	Числовые характеристики		
				$M[X]$	$D[X]$	$\sigma[X]$
Биномиальный	$X = 0, 1, 2, ..., n$	p, n	$\hat{p} = \hat{M}[X]/\hat{n}$	np	npq	$\sqrt{npq}$
Пуассона	$X = 0, 1, 2, ..., m, ...$	a	$\hat{a} = \hat{M}[X]$	a	a	$\sqrt{a}$
Геометрический	$X = 0, 1, 2, ..., m, ...$	p	$\hat{p} = 1/(\hat{M}[X] + 1)$	$\dfrac{1}{p} - 1$	$\dfrac{q}{p^2}$	$\sqrt{\dfrac{q}{p^2}}$

Предположим, что согласно методике, описанной в разделе 1 (лабораторная работа № 1), произведена первичная обработка имеющихся статистических данных.

Гипотеза о виде закона распределения изучаемой случайной величины обычно выдвигается на основании вида графического изображения статистического закона распределения, сведений о механизме формирования значений этой величины и значений оценок числовых характеристик.

В таблицах 3.1 и 3.2 приведены сведения о наиболее часто используемых при решении практических задач законах распределения дискретных и непрерывных случайных величин. В графе «Примечание» этих таблиц приведены примеры случайных величин, подчиняющихся указанным законам, либо описан механизм формирования значений этих величин.

Если изучается непрерывная случайная величина, то вид полученной гистограммы относительных частот обычно значительно облегчает задачу выдвижения гипотезы H_0. Например, по виду гистограммы, изображённой на рисунке 3.1, *а*, логично выдвинуть гипотезу о равномерном законе распределения исследуемой случайной величины. Вид гистограмм, приведённых на рисунках 3.1, *б* и 3.1, *в*, напоминает соответственно кривые экспоненциального и нормального законов распределения.

распределения дискретных случайных величин

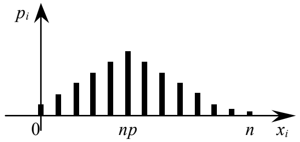
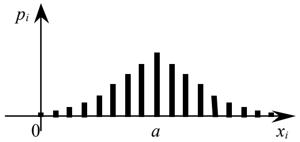
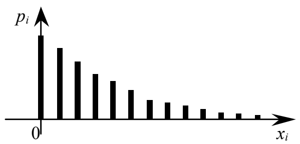
Вероятности возможных значений, столбцовая диаграмма	Примечание
$P(X=k)=C_n^k p^k q^{n-k}$ 	Случайная величина X характеризует число появлений события A в серии из n независимых испытаний, в каждом из которых это событие может осуществиться с вероятностью p.
$P(X=k)=\frac{a^k}{k!}e^{-a}$ 	**Пример:** число событий простейшего потока, характеризующегося интенсивностью a, произошедших в течение единицы времени.
$P(X=k)=q^k p$ 	Случайная величина X характеризует число независимых испытаний, произведённых до первого появления события A, которое в каждом из этих испытаний может произойти с вероятностью p (при этом испытание, в котором появляется событие A, не учитывается).

Таблица 3.2 – **Основные сведения о наиболее часто встречающихся на практике законах**

Закон распределения	Возможные значения	Параметры	Статистическая оценка параметров	Числовые характеристики					Вероятность попадания значений с. в. в отрезок $[\alpha;\beta]$
				$M[X]$	$D[X]$	$\sigma[X]$	$A[X]$	$Ex[X]$	
Равномерный	$X\in[a;b]$	a b	$\tilde{a}=x_{\min}-\frac{x_{\max}-x_{\min}}{n-1}$, $\tilde{b}=x_{\max}+\frac{x_{\max}-x_{\min}}{n-1}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{b-a}{2\sqrt{3}}$	0	–1,2	$P(\alpha\le X\le\beta)=\frac{\beta-\alpha}{b-a}$
Экспоненциальный (показательный)	$X\in[0;\infty)$	λ	$\tilde{\lambda}=1/\tilde{M}[X]$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{\lambda}$	2	6	$P(\alpha\le X\le\beta)=e^{-\lambda\alpha}-e^{-\lambda\beta}$
Нормальный	$X\in R$	m σ	$\tilde{m}=\tilde{M}[X]$ $\tilde{\sigma}=\tilde{\sigma}[X]$	m	σ^2	σ	0	0	$P(\alpha\le X\le\beta)=\Phi\left(\frac{\beta-m}{\sigma}\right)-\Phi\left(\frac{\alpha-m}{\sigma}\right)$; $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

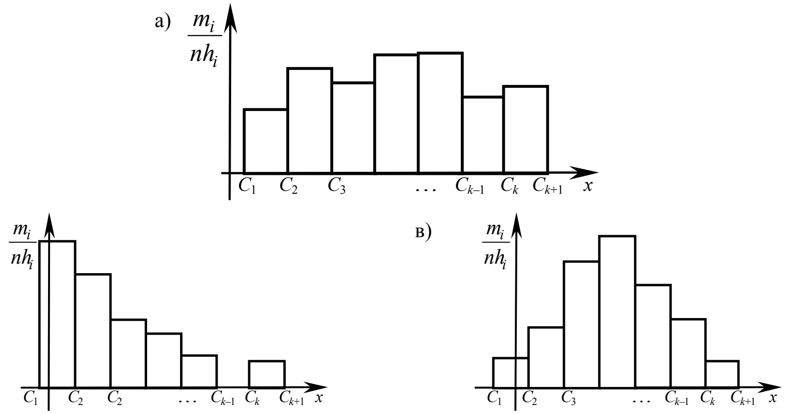
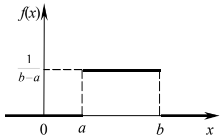
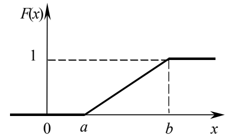
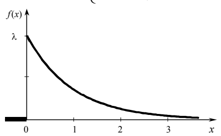
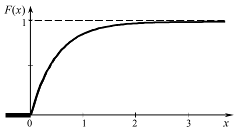
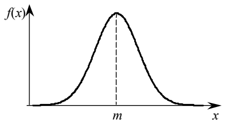
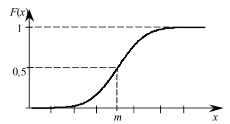


Рисунок 3.1 – Примеры гистограмм относительных частот

распределения непрерывных случайных величин

Функция плотности распределения вероятностей	Функция распределения	Примечание
$f(x)=\begin{cases}\dfrac{1}{b-a}, & x\in[a;b];\\ 0, & x\notin[a;b]\end{cases}$ 	$F(x)=\begin{cases}0, & x\le a;\\ \dfrac{x-a}{b-a}, & a<x\le b;\\ 1 & x>b\end{cases}$ 	Если все возможные значения непрерывной случайной величины принадлежат отрезку [a; b], и все значения, попадающие на этот отрезок равновозможны, то данная случайная величина X распределена по равномерному закону. **Пример:** величина погрешности при округлении данных.
$f(x)=\begin{cases}0, & x<0;\\ \lambda e^{-\lambda x}, & x\ge 0\end{cases}$ 	$F(x)=\begin{cases}0, & x\le 0;\\ 1-e^{-\lambda x}, & x>0\end{cases}$ 	**Примеры:** – промежуток времени между моментами наступления двух последовательных событий простейшего потока; – разнообразные временные характеристики функционирования технических устройств (время безотказной работы оборудования, время простоя в ожидании ремонта и т. д.)
$f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$ 	$F(x)=\dfrac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{(t-m)^2}{2\sigma^2}}dt$ 	Если случайная величина X представляет собой сумму большого числа независимых (или слабо зависимых) случайных величин, сопоставимых по уровню своего влияния на суммарный результат, то эта величина имеет распределение, близкое к нормальному. **Пример:** реальные значения параметров изготовленного изделия.

Большое внимание при выдвижении гипотез уделяется и вычисленным значениям числовых характеристик (примеры 3.1, 3.2).

2 Применение критерия Пирсона χ^2 для проверки гипотезы о виде закона распределения случайной величины.

Одним из наиболее широко используемых на практике критериев согласия является критерий χ^2 Пирсона. Он может применяться для проверки гипотез о распределении исследуемой случайной величины по любому из известных законов распределения (как дискретных, так и непрерывных случайных величин).

Для применения этого критерия необходимо представление эмпирического распределения в виде интервального или сгруппированного статистического ряда. Обозначим, как и ранее, число используемых разрядов разбиения k, а частоту i-го разряда – m_i, $(i=1, 2, \ldots, k)$.

Критерий χ^2 основан на сопоставлении эмпирических (m_i) и теоретических (np_i), т. е. вычисленных в предположении справедливости проверяемой гипотезы, частот попадания значений исследуемой случайной величины в рассматриваемые частичные разряды. В качестве меры расхождения эмпирического и теоретического распределений используется статистика

$$\chi^2=\sum_{i=1}^{k}\frac{(m_i-np_i)^2}{np_i},$$

которая при $n\to\infty$ независимо от вида предполагаемого распределения стремится к распределению χ^2 с $\nu=k-r-1$ степенями свободы (здесь r – число параметров теоретического распределения, оцениваемых по выборке).

При вычислении значений p_i для нормального закона распределения можно воспользоваться таблицей значений функции плотности стандартного нормального распределения (приложение Б).

Легко заметить, что при незначительных отклонениях значений m_i от np_i значение критерия χ^2 будет близким к нулю. И наоборот, большое значение критерия χ^2 свидетельствует о существенном отклонении экспериментально полученного распределения от предполагаемого. Поэтому критическая область в данном случае задается условием $P(\chi^2>\chi^2_{\alpha,\nu})=\alpha$.

Критическое значение $\chi^2_{\alpha,\nu}$ определяется по таблице квантилей распределения χ^2 в зависимости от уровня значимости α и числа степеней свободы ν.

Необходимыми условиями применения критерия χ^2 являются достаточно большой объем выборки ($n\ge 50$) и отсутствие в теоретическом распределении разрядов с небольшим ($np_i<5$) числом наблюдений. Для обеспечения последнего условия интервалы, для которых $np_i<5$, необходимо объединить с соседними.

3 Алгоритм применения критерия χ^2 для проверки гипотезы о виде закона распределения исследуемой случайной величины.

1 Выборочные данные представляются в виде интервального или сгруппированного статистического ряда.

2 Выбирается уровень значимости α.

3 Формулируется гипотеза о виде закона распределения исследуемой случайной величины.

4 Вычисляются вероятности p_i попадания значений случайной величины X в рассматриваемые разряды разбиения: $p_i=P(C_i\le X<C_{i+1})=$ $=F(C_{i+1})-F(C_i)$, $(i=1, 2, \ldots, k)$, где $F(x)$ –гипотетическая функция распределения случайной величины X.

Замечание. Если изучается непрерывная случайная величина, то при вычислении значений p_i необходимо изменить границы первого и последнего частичных интервалов разбиения таким образом, чтобы учесть все возможные значения, которые может принять случайная величина предполагаемого класса. В зависимости от конкретного вида проверяемой гипотезы границы частичных интервалов необходимо изменить следующим образом:

Вид закона распределения	Первый интервал разбиения	Последний интервал разбиения
Равномерный	$[\hat{a}, C_2)$	$[C_k, \hat{b})$
Экспоненциальный	$[0, C_2)$	$[C_k, \infty)$
Нормальный	$(-\infty, C_2)$	$[C_k, \infty)$

5 Определяются значения теоретических частот np_i ($i = 1, 2,\ldots, k$). При необходимости для обеспечения условия $np_i \geq 5$, объединяются несколько соседних разрядов разбиения.

6 Вычисляется наблюдаемое значение критерия χ^2:

$$\chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}.$$

7 По таблицам квантилей распределения χ^2 определяется критическое значение $\chi^2_{\alpha,\nu}$, соответствующее заданному уровню значимости α и числу степеней свободы $\nu = k - r - 1$.

Если расчётное значение критерия попадает в критическую область, т. е. $\chi^2 > \chi^2_{\alpha,\nu}$, то проверяемая гипотеза отвергается (при этом вероятность отклонения верной гипотезы равна α).

В случаях, когда наблюденное значение χ^2 не превышает критического $\chi^2 \leq \chi^2_{\alpha,\nu}$, считают, что выдвинутая гипотеза не противоречит опытным данным. Подчеркнем, что полученный результат свидетельствует лишь о приемлемом согласовании проверяемой гипотезы с имеющимися выборочными данными и, в общем случае, не является доказательством истинности этой гипотезы.

Пример 3.1. На основании выборочных данных, приведённых в примере 1.1, подобрать закон распределения случайной величины X, характеризующей число отказов оборудования, произошедших в течение рабочей смены. Уровень значимости α принять равным 0,05.

Решение. После проведения первичной обработки полученных данных (см. пример 1.1), опираясь на сведения о механизме формирования значений исследуемой случайной величины (поток отказов оборудования обычно обладает свойствами простейшего потока), учитывая вид построенной столбцовой диаграммы и значения оценок числовых характеристик

($\hat{M}[X] = 2 \approx \hat{D}[X] = 1{,}92$), выдвигаем гипотезу о том, что изучаемая случайная величина подчиняется закону распределения Пуассона.

Для проверки этой гипотезы с помощью критерия χ^2 выполним следующие действия.

Вычислим оценку параметра a распределения Пуассона: $\hat{a} = \hat{M}[X] = 2$.

Вычислим вероятности наблюденных значений изучаемой случайной величины:

$$p_i = P(X = i) = \frac{a^i}{i!} e^{-a} \quad (i = 0, 1, 2, \ldots, 6);$$

$$P(X = 0) = \frac{2^0}{0!} e^{-2} = e^{-2} = 0{,}1353; \quad P(X = 1) = \frac{2^1}{1!} e^{-2} = 2e^{-2} = 0{,}27067;$$

$$P(X = 2) = \frac{2^2}{2!} e^{-2} = 2e^{-2} = 0{,}27067; \quad P(X = 3) = \frac{2^3}{3!} e^{-2} = \frac{4}{3} e^{-2} = 0{,}1804;$$

$$P(X = 4) = \frac{2^4}{4!} e^{-2} = \frac{2}{3} e^{-2} = 0{,}0902; \quad P(X = 5) = \frac{2^5}{5!} e^{-2} = \frac{4}{15} e^{-2} = 0{,}036;$$

$$P(X = 6) = \frac{2^6}{6!} e^{-2} = \frac{4}{45} e^{-2} = 0{,}012;$$

$$\sum_{i=0}^{6} p_i = \sum_{i=0}^{6} P(X = i) = 0{,}9948.$$

Вычислим значения теоретических частот и заполним расчётную таблицу:

$\tilde{x}_i$	0	1	2	3	4	5	6
m_i	4	19	12	6	7	1	1
p_i	0,1353	0,27067	0,27067	0,1804	0,0902	0,036	0,012
np_i	6,765	13,5335	13,5335	9,02	4,51	1,8	0,6

Учитывая, что теоретические частоты наблюденных значений, находящихся в трёх последних столбцах таблицы, не превышают пяти единиц, при вычислении значения критерия χ^2 эти три разряда объединим в один:

$$\chi^2 = \sum_{i=1}^{5} \frac{(m_i - np_i)^2}{np_i} = \frac{(4 - 6{,}765)^2}{6{,}765} + \frac{(19 - 13{,}5335)^2}{13{,}5335} + \frac{(12 - 13{,}5335)^2}{13{,}5335} +$$

$$+\frac{(6-9{,}02)^2}{9{,}02}+\frac{(9-6{,}91)^2}{6{,}91}=5{,}155.$$

По таблицам квантилей распределение χ^2 определим критическое значение $\chi^2_{\alpha,\nu}$, соответствующее $\alpha=0{,}05$, $\nu=\mathrm{k}-\mathrm{r}-1=5-1-1=3$: $\chi^2_{0{,}05;3}=7{,}815$.

Поскольку выборочное значение критерия меньше критического $\chi^2\le\chi^2_{\alpha,\nu}$, можно сделать вывод о том, что экспериментальные данные не дают оснований для отклонения проверяемой гипотезы.

Пример 3.2. На основании опытных данных, приведённых в примере 1.2 подобрать закон распределения непрерывной случайной величины *X*, характеризующей время безотказной работы оборудования между двумя последовательными переналадками. Уровень значимости α принять равным 0,05.

Решение. Используя результаты первичной обработки выборочных данных (вид полученной гистограммы и значения оценок числовых характеристик: $\hat{M}[X]=8{,}9\approx\hat{\sigma}[X]=8{,}34$), а также учитывая сведения о физическом смысле полученных значений, выдвигаем гипотезу о том, что случайная величина *X* распределена по экспоненциальному закону.

Проверим согласование сформулированной гипотезы с экспериментальными данными с помощью критерия χ^2.

Вычислим оценку параметра экспоненциального закона распределения

$$\hat{\lambda}=\frac{1}{\hat{M}[X]}=\frac{1}{8{,}901}\approx 0{,}112.$$

При вычислении вероятностей $p_i=P(C_i\le X<C_{i+1})$ изменим границы первого и последнего интервалов разбиения в соответствии с замечанием, приведенным на с. 26.

$$p_i=P(C_i\le X<C_{i+1})=e^{-\lambda C_i}-e^{-\lambda C_{i+1}}\ (i=1,2,\ldots,8);$$

$$p_1=P(0\le X<2{,}987)=e^{-0{,}112\cdot 0}-e^{-0{,}112\cdot 2{,}987}=0{,}2843;$$

$$p_2=P(2{,}987\le X<8{,}127)=e^{-0{,}112\cdot 2{,}987}-e^{-0{,}112\cdot 8{,}127}=0{,}3133;$$

$$p_3=P(8{,}127\le X<13{,}267)=e^{-0{,}112\cdot 8{,}127}-e^{-0{,}112\cdot 13{,}267}=0{,}1764;$$

$$p_4=P(13{,}267\le X<18{,}407)=e^{-0{,}112\cdot 13{,}267}-e^{-0{,}112\cdot 18{,}407}=0{,}099;$$

$$p_5=P(18{,}407\le X<23{,}547)=e^{-0{,}112\cdot 18{,}407}-e^{-0{,}112\cdot 23{,}547}=0{,}055;$$

$$p_6=P(23{,}547\le X<28{,}687)=e^{-0{,}112\cdot 23{,}547}-e^{-0{,}112\cdot 28{,}687}=0{,}032;$$

$$p_7=P(28{,}687\le X<33{,}827)=e^{-0{,}112\cdot 28{,}687}-e^{-0{,}112\cdot 33{,}827}=0{,}017;$$

$$p_8=P(33{,}827\le X<\infty)=e^{-0{,}112\cdot 33{,}827}-e^{-0{,}112\cdot\infty}=0{,}023.$$

Для контроля убедимся, что $\sum_{i=1}^{8}p_i=1$.

Определим значения теоретических частот и занесём их в расчётную таблицу:

$[C_i,C_{i+1})$	[0; 2,987)	[2,987; 8,127)	[8,127; 13,267)	[13,267; 18,407)	[18,407; 23,547)	[23,547; 28,687)	[28,687; 33,827)	[33,827; ∞)
m_i	15	15	9	4	3	2	1	1
p_i	0,2843	0,3133	0,1764	0,099	0,055	0,032	0,017	0,023
np_i	14,215	15,665	8,82	4,95	2,75	1,6	0,85	1,15

Поскольку значения np_i, соответствующие пяти последним интервалам разбиения, не превышают пяти единиц, объединим эти интервалы в один и для вычисления значения критерия χ^2 составим следующую расчётную таблицу:

$[C_i,C_{i+1})$	[0; 2,987)	[2,987; 8,127)	[8,127; 13,267)	[13,267; ∞)
m_i	15	15	9	11
p_i	0,2843	0,3133	0,1764	0,226
np_i	14,215	15,665	8,82	11,3

Вычислим значение критерия χ^2

$$\chi^2=\frac{(15-14{,}215)^2}{14{,}215}+\frac{(15-15{,}665)^2}{15{,}665}+\frac{(9-8{,}82)^2}{8{,}82}+\frac{(11-11{,}3)^2}{11{,}3}=0{,}083.$$

Получение близкого к нулю значения критерия χ^2 свидетельствует об очень хорошем согласовании выдвинутой гипотезы с выборочными данными.

Критическое значение критерия, соответствующее значениям $\alpha=0{,}05$ и $\nu=k-r-1=4-1-1=2$, определим с помощью таблицы приложения Г: $\chi^2_{\alpha,\nu}=\chi^2_{0{,}05;2}=5{,}99$.

Поскольку $\chi^2<\chi^2_{\alpha,\nu}$, можно сделать вывод о том, что выборочные данные не противоречат проверяемой гипотезе об экспоненциальном законе распределения изучаемой случайной величины *X*.

Порядок выполнения работы

1 Получить выборку значений исследуемой случайной величины X и записать её на диск (приложение А).

2 Произвести первичную обработку полученных статистических данных.

3 Выдвинуть гипотезу о виде закона распределения изучаемой случайной величины.

4 Проверить согласование сформулированной гипотезы с имеющимися выборочными данными (ручной расчёт):

– вычислить оценки параметров предполагаемого закона распределения;
– вычислить вероятности попадания значений случайной величины, распределённой по предполагаемому закону в частичные разряды разбиения (т. е. определить значения вероятностей $p_i = P(X = \widetilde{x}_i)$, если изучается дискретная случайная величина и вероятности $p_i = P(C_i \le X < C_{i+1})$, если исследуется непрерывная случайная величина, для всех значений $i = 1, 2, \ldots, k$);
– определить значения теоретических частот np_i, $i = 1, 2, \ldots, k$;
– вычислить выборочное значение критерия χ^2;
– сравнить выборочное значение критерия с критическим значением $\chi^2_{\alpha,\upsilon}$.

5 Проверить согласование выдвинутой гипотезы с имеющимися экспериментальными данными с помощью ЭВМ:

– вычислить значение критерия χ^2 (приложение А, п. 8);
– построить совместное графическое изображение эмпирического и предполагаемого теоретического распределений изучаемой случайной величины (приложение А, п. 9).

6 Сделать вывод о законе распределения вероятностей изучаемой случайной величины.

Лабораторная работа № 4

ПОСТРОЕНИЕ РЕГРЕССИОННОЙ МОДЕЛИ СИСТЕМЫ ДВУХ СЛУЧАЙНЫХ ВЕЛИЧИН

Цель работы: изучить основные методы регрессионного и корреляционного анализа; исследовать зависимость между двумя случайными величинами, заданными выборками.

Задание: по виду корреляционного поля выдвинуть гипотезу о форме регрессионной зависимости между двумя случайными величинами;
используя метод наименьших квадратов (МНК), найти параметры (коэффициенты) уравнения регрессии;
оценить тесноту связи между случайными величинами и проверить ее значимость.

Основные теоретические сведения

1 Введение.

Часто в практике инженерных и научных исследований результат испытания характеризуется сразу некоторой системой случайных величин, т. е. многомерной случайной величиной (МСВ).

Многомерные случайные величины характеризуются многомерными (k-мерными) законами распределения (заданными таблично, в виде функции распределения $F(x_1, x_2, ..., x_k)$, плотности распределения $f(x_1, x_2, ..., x_k)$ или другими способами) и числовыми характеристиками. Наряду с числовыми характеристиками, применимыми для одномерных с. в. (математическим ожиданием и др.), МСВ определяются числовыми характеристиками, описывающими зависимость между ее компонентами.

В дальнейшем ограничимся рассмотрением двумерной с. в. (X, Y), компоненты X и Y которой фактически являются одномерными с. в.

Зависимость между случайными величинами, при которой каждому значению x случайной величины X однозначно ставится в соответствие единственное значение y с. в. Y, называется функциональной. Например, зная напряжение U на участке электрической цепи сопротивлением R, можно однозначно определить величину тока ($U = R\,I$).

Однако часто на практике одному значению с. в. X может соответствовать не одно, а множество значений с. в. Y, характеризуемых для каждого $X = x$ условным распределением с плотностью вероятностей $f(Y \mid X = x)$. Такая зависимость называется ***статистической***. Примером статистической зависимости является зависимость технической скорости движения поезда (Y) от его массы (X).

Изменение условного закона распределения с. в. Y при изменении значения X может проявляться как при изменении вида распределения (рисунок 4.1, *а*), так и при изменении его числовых характеристик. В частности, зависимость условного математического ожидания одной с. в. $M(Y|X=x)$ от значения, которое примет другая с. в., $X=x$ называется ***регрессионной*** (рисунок 4.1, *б*).

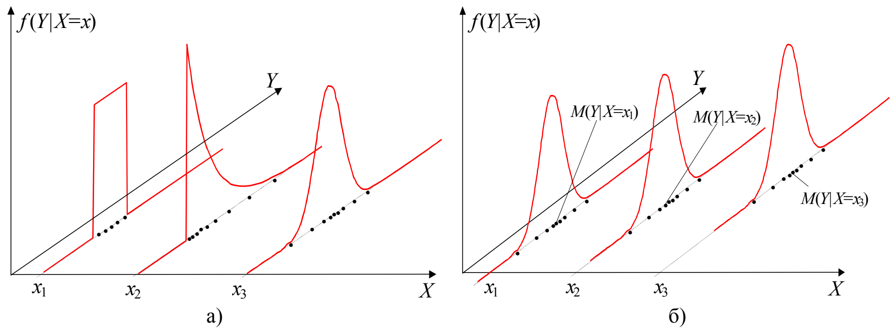


Рисунок 4.1 – Примеры статистической зависимости: а) с изменением вида закона распределения; б) с изменением только лишь математического ожидания

Исследование зависимостей между случайными величинами является наиболее часто используемым приложением математической статистики (МС) и применяется главным образом для предсказания значения одной с. в. по значению другой. Знание совместного закона распределения двух с. в. ($F(x, y)$, $f(x, y)$) или статистической зависимости между ними в вероятностном смысле полностью определяет зависимость между исследуемыми с. в. Однако получение такой информации является затруднительным. Поэтому в МС чаще ограничиваются изучением регрессионной зависимости, которая каждому значению одной с. в. ставит в соответствие условное математическое ожидание другой с. в. ($M(Y|X=x)$).

Изучением зависимости между с. в. занимается регрессионный и корреляционный анализ. Предметом регрессионного анализа является нахождение вида уравнения регрессионной зависимости между случайными величинами. Предметом же корреляционного анализа является оценка тесноты связи между с. в.

2 Регрессионный анализ.

Пусть дана выборка значений двумерной с. в. $(X,Y)=\{(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)\}$, где n – объем двумерной выборки. Первым шагом в поиске или подборе уравнения регрессии между с. в. является графическое отображение значений двумерной с. в. в виде точек $(x_1, y_1), \ldots, (x_n, y_n)$ на плоскости X-Y. Такое изображение называется ***диаграммой рассеяния*** (корреляционным полем). Визуальный анализ диаграммы рассеяния позволяет сделать предположение о виде линии регрессии.

Если предполагается, что зависимость между с. в. X и Y линейна (рисунок 4.2, *а*), то теоретическая модель регрессионной зависимости между с. в. задается уравнением 4.1 – моделью простой линейной регрессии Y по X.

$$M(Y|X=x)=\beta_0+\beta_1 x, \qquad (4.1)$$

т. е. для каждого $X=x_i$ имеется условное распределение с. в. Y со средним значением $(\beta_0+\beta_1 x_i)$. Таким образом, для каждого i-го наблюдения справедлива следующая зависимость:

$$\bar{y}_i=\beta_0+\beta_1 x_i+e_i,\ i=\overline{1,n}, \qquad (4.2)$$

где y_i – i-е выборочное значение с. в. Y; β_0 и β_1 – параметры линейной регрессии, требующие определения; x_i – i-е выборочное значение с. в. X; e_i – ошибка, вызванная отклонением i-го наблюдения с. в. Y от условного среднего $M(У|X=x_i)$.

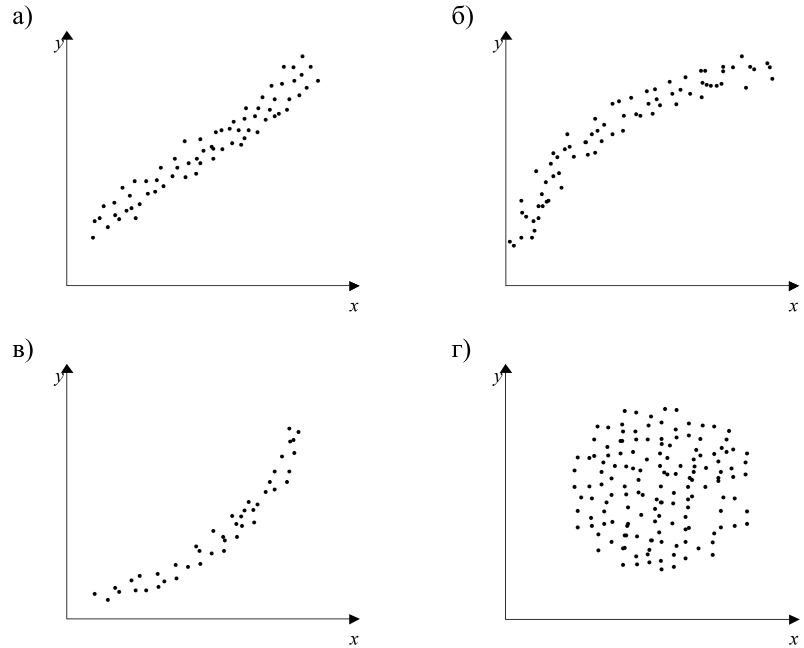


Рисунок 4.2 – Диаграмма рассеяния, соответствующая линейной (а), экспоненциальной (б), параболической (в) регрессионной зависимости и отсутствию регрессионной зависимости (г)

3 Метод наименьших квадратов (МНК).

Наилучшие (несмещенные, состоятельные и эффективные) оценки неизвестных параметров линейной регрессии β_0 и β_1 могут быть получены методом наименьших квадратов, смысл которого заключается в следующем.

Рассмотрим функцию S (4.3), равную сумме квадратов отклонений выборочных значений y_i случайной величины Y от значения $\bar{y}(x_i)$, предсказанного уравнением регрессии в точке $X = x_i$ (рисунок 4.3).

$$S = \sum_{i=1}^{n} (y_i - \bar{y}(x_i))^2 = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 \cdot x_i)^2 = S(\beta_0, \beta_1) \qquad (4.3)$$

Фактически эти отклонения в каждой точке x_i равны e_i, т. е. той ошибке, которая обусловлена упрощением вида зависимости Y от X (без учёта дополнительных влияющих факторов), а также возможной ошибкой в выборе формы регрессии (в действительности она может описываться другим уравнением).

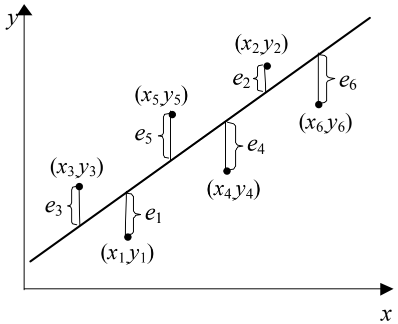


Рисунок 4.3 – Отклонения результатов совместного измерения случайных величин *X* и *Y* от уравнения регрессии

Оценки β_0 и β_1 параметров регрессии β_0 и β_1 получаются минимизацией соответственно по β_0 и β_1 суммы квадратов отклонений e_i выборочных значений с. в. Y от предполагаемой линии регрессии, т. е. функции S. Из курса математического анализа известно, что для нахождения минимума функции $S(\beta_0, \beta_1)$ необходимо приравнять нулю частные производные (по неизвестным β_0 и β_1) и решить систему уравнений, называемых нормальными:

$$\begin{cases} \dfrac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0; \\ \dfrac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0. \end{cases} \qquad (4.4)$$

Система линейных уравнений (4.4) имеет единственное решение, если определитель матрицы ее коэффициентов не равен нулю. Полученные значения $\hat{\beta}_0$ и $\hat{\beta}_1$, являющиеся решением системы (4.4), называются оценками параметров регрессии. Для предполагаемой ***линейной*** регрессионной зависимости (4.1) оценки минимизируют ошибку, возникающую при аппроксимации выборки ***прямой***, и вычисляются по формулам:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}, \qquad (4.5)$$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i. \qquad (4.6)$$

В результате оценка уравнения линейной регрессии (эмпирическое уравнение линейной регрессии, прямая МНК) будет иметь вид: $\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$.

В предположении, что значения ошибок $\{e_1, e_2, \ldots, e_n\}$, возникающих при аппроксимации выборки уравнением регрессии, являются взаимно независимыми случайными величинами с нормальным распределением, нулевым математическим ожиданием и постоянной дисперсией, оценки МНК параметров таких уравнений регрессии являются несмещенными, состоятельными и эффективными.

4 Пошаговый регрессионный анализ.

В случае, если по виду диаграммы рассеяния сложно выдвинуть предположение о виде регрессионной зависимости, рекомендуется использовать пошаговый регрессионный анализ. Для этого уравнение регрессии выбирают как можно более сложным, содержащим большое количество слагаемых. Например,

$$\bar{y}(x) = \beta_0 + \beta_1 / x + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + \beta_5 e^x. \qquad (4.7)$$

Стандартная пошаговая процедура позволяет в записанном уравнении регрессии выбрать те слагаемые, совокупность которых достаточно качественно предсказывает среднее значение зависимой с. в., и отбросить те слагаемые, которые существенно не улучшают предсказание с. в. Y.

Особенно полезен пошаговый регрессионный анализ в случае множественной регрессии, когда с. в. Y зависит от нескольких независимых случайных величин X_1, X_2, X_3 и т. д. При этом, пошаговая процедура позволяет из множества влияющих (независимых) с. в. исключить несущественные и тем самым упростить уравнение регрессии.

5 Корреляционный анализ.

Помимо предположения о форме регрессионной зависимости между случайными величинами и нахождения его параметров исследователю требуется оценить насколько удачно выбранное уравнение регрессии объясняет существующую зависимость между исследуемыми с. в.

Коэффициент корреляции. Основной числовой характеристикой, определяющей тесноту ***линейной*** регрессионной зависимости (4.1) между двумя случайными величинами, является коэффициент корреляции

$$r = \frac{M(XY) - M(X)M(Y)}{\sigma(X)\sigma(Y)}, \tag{4.8}$$

где *M(XY)* – математическое ожидание произведения с. в. *X* и *Y*.

Коэффициент корреляции является безразмерной величиной и может принимать значения из интервала: $-1 \le r \le 1$. Для линейно независимых случайных величин *X* и *Y* коэффициент корреляции равен нулю. Экстремальные значения 1 или –1 коэффициента корреляции соответствуют линейной функциональной зависимости между двумя с. в. (положительной и отрицательной соответственно).

Таким образом, можно говорить, что коэффициент корреляции характеризует степень линейной регрессионной зависимости между двумя с. в., близость ее к линейной функциональной зависимости.

По заданной двумерной выборке, оценку коэффициента корреляции $\hat{r}$ можно найти по формуле

$$\hat{r} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}. \tag{4.9}$$

При большом числе независимых наблюдений, подчиняющихся одному и тому же распределению, близкому к нормальному, оценка $\hat{r}$ близка к истинному значению коэффициента корреляции *r*.

Коэффициент детерминации. Для характеристики тесноты связи между случайными величинами, описываемой нелинейной функцией регрессии, используется коэффициент детерминации. Данный коэффициент характеризует качество описания зависимости между двумя с. в. выбранным уравнением регрессии. Очевидно, чем теснее наблюдения примыкают к линии регрессии, тем лучше регрессия описывает соответствующую зависимость и потому с большей надежностью может быть применена для практических расчетов.

Коэффициент детерминации рассчитывается по формулам (4.10 – 4,12) и может принимать значения в интервале от нуля до единицы ($0 \le R^2 \le 1$):

$$\hat{R}^2 = \frac{\sum_{i=1}^{n}(\bar{y}(x_i) - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{4.10}$$

$$\hat{R}^2 = \frac{\frac{1}{n-1}\cdot\sum_{i=1}^{n}(\bar{y}(x_i) - \bar{y})^2}{\frac{1}{n-1}\cdot\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{S^2_{\hat{y}}}{S^2_Y}, \tag{4.11}$$

$$\hat{R}^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{4.12}$$

где *n* – объем выборок; $\bar{y}(x_i)$ – среднее значение с. в. *Y*, вычисленное по уравнению регрессии для заданного x_i (условное среднее); $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – безусловное среднее с. в. *Y*; y_i – *i*-е выборочное значение с. в. *Y*; $S^2_{\hat{y}}$ – дисперсия линии регрессии относительно средней; S^2_Y – общая дисперсия; e_i – отклонения фактических значений с. в. $Y = y_i$ от расчетных, полученных по уравнению регрессии $\bar{y}(x_i)$.

В первой формуле числитель характеризует рассеяние условных средних $\hat{y}_i$, определяемых уравнением регрессии, около безусловного $\bar{y}$, в знаменателе – рассеяние опытных данных y_i около безусловного среднего $\bar{y}$.

Анализ (4.11) показывает, что коэффициент детерминации характеризует насколько хорошо уравнение регрессии предсказывает значения зависимой с. в. *Y*, т. е. объясняет ее рассеяние в общей величине рассеяния (дисперсии) зависимой с. в. Таким образом, коэффициент детерминации рассматривается как мера качества описания зависимости между с. в. с помощью уравнения регрессии, т. е. характеризует насколько удачно выбранная модель (уравнение) регрессии описывает действительную зависимость между с. в.

Отметим, что коэффициент детерминации не является мерой какой-либо зависимости априорно (как это справедливо для коэффициента корреляции, который всегда характеризует степень линейной зависимости между двумя случайными величинами), он лишь оценивает степень приближения выбранного уравнения регрессионной зависимости к действительной зависимости между двумя с. в.

Равенство коэффициента детерминации нулю указывает на то, что выбранное уравнение регрессии (модель зависимости) никак не объясняет действительную зависимость между с. в. Равенство же коэффициента детерминации единице указывает на то, что зависимость между случайными величинами является функциональной, описываемой уравнением регрессии, т. е.

выбранное уравнение регрессии полностью (однозначно) определяет зависимость между с. в.

Если значение коэффициента детерминации больше 0,7, то считают, что выбранное уравнение регрессии хорошо описывает зависимость, существующую между случайными величинами. Если же коэффициент детерминации меньше 0,3, то уравнение регрессии незначительно описывает зависимость между случайными величинами, если таковая существует.

Замечание: При анализе связи между случайными величинами X и Y, для которых значение одной из с. в. (обычно X) задаются исследователем, коэффициент корреляции (или детерминации) нельзя рассматривать как строгую меру взаимосвязи явлений (как в случае, где X – неконтролируемая величина), поскольку здесь большую роль играет выбор самих значений x_i. В этом случае коэффициент корреляции (детерминации) характеризует лишь меру близости эмпирических точек к линии регрессии.

6 Проверка значимости $\hat{r}$.

Оценки коэффициентов корреляции и детерминации сами являются случайными величинами, так как для различных выборок из одной и той же генеральной совокупности могут принимать различные значения. При малых объемах выборок эти различия будут особенно существенными. Поэтому более надежным является нахождение интервальных оценок коэффициентов корреляции (или детерминации).

Достаточно часто при нахождении оценок коэффициентов корреляции и детерминации используется проверка значимости этих оценок, которая позволяет сделать вывод о существенности описания действительной зависимости уравнением регрессии. Фактически, проверка значимости коэффициентов корреляции и детерминации соответствует нахождению интервальных оценок этих коэффициентов и анализу принадлежности этому интервалу значения $r=0$ и $R^2=0$ соответственно, но выполняется по более простому алгоритму. Проверка значимости позволяет сделать вывод либо о существенности описания зависимости уравнением регрессии, либо о том, что данное уравнение практически никак не определяет существующую зависимость между с. в., а ненулевые значения коэффициентов корреляции и/или детерминации обусловлены только лишь случайностью выборки.

Чтобы сделать статистический вывод о значимости коэффициента корреляции (при проверке линейности регрессионной зависимости) выдвигается нулевая гипотеза об отсутствии линейной зависимости между исследуемыми с. в. (т. е. H_0: $r=0$, H_a: $r\neq 0$). Если гипотеза H_0 отклоняется, то считается, что уравнение регрессии Y по X действительно имеет линейный вид (4.1-4.2).

Для проверки гипотезы H_0 вычисляется t-статистика

$$t=\hat{r}\sqrt{\frac{n-2}{1-\hat{r}^2}}\,. \tag{4.13}$$

При условии справедливости гипотезы H_0 рассчитанная t-статистика имеет распределение Стьюдента с $n-2$ степенями свободы. Найденное по формуле (4.13) значение t' сравнивается с критическим значением $t_{\alpha,\nu}$ при $\nu=n-2$ степенях свободы (см. приложение Д). Если расчетное значение t' по абсолютной величине превосходит табличное для заданного уровня значимости, то нулевая гипотеза H_0 о линейной независимости двух с. в. отклоняется.

7 Проверка значимости $\hat{R}^2$.

Характеристикой, измеряющей тесноту связи двух с. в., близость ее к выбранной функции регрессии $\bar{y}(x_i)$, отличной от линейной, является коэффициент детерминации.

При выполнении процедуры проверки значимости коэффициента детерминации выдвигается нулевая гипотеза о том, что предложенное уравнение регрессии никак не отражает реальную зависимость между с. в., т. е. H_0: $R^2=0$. Альтернативная гипотеза заключается в том, что выбранная модель зависимости (уравнение регрессии) $\bar{y}(x_i)$ в достаточной степени объясняет действительную зависимость между случайными величинами, т. е. H_0: $R^2>>0$.

Для оценки значимости коэффициента детерминации используется статистика

$$F=\hat{R}^2\,\frac{n-2}{1-\hat{R}^2}\,, \tag{4.14}$$

имеющая F-распределение Фишера с $\nu_1=m=1$ и $\nu_2=n-m-1=n-2$ степенями свободы. Здесь m – число «независимых» величин, в данном случае – одна (X). Значение статистики, вычисленное по (4.14) сравнивается с критическим значением $F_{\nu1,\,\nu2,\,\alpha}$, найденным по таблицам (см. приложение Е) при заданном уровне значимости и соответствующем числе степеней свободы. Если $F>F_{\nu1,\,\nu2,\,\alpha}$, то нулевая гипотеза отклоняется, вычисленный коэффициент детерминации значимо отличается от нуля, и значит выбранное уравнение регрессии может использоваться в дальнейших исследованиях.

Пример 4.1. По десяти предприятиям лёгкой промышленности были получены следующие данные об объёмах вкладываемых в предприятие инвестиций X, млрд руб., и соответствующих размерах прибыли Y, млрд руб.

Таблица 4.1 – **Исходные данные**

X	1,0	1,5	4,0	2,5	1,0	0,5	3,0	2,0	0,5	5,0
Y	7,34	11,46	15,35	16,02	3,50	4,95	13,43	7,31	–0,87	17,54

Требуется исследовать зависимость между величинами инвестиций и прибылью, получаемой предприятиями лёгкой промышленности, с целью оптимизации вложения средств.

Решение. Естественно, что на величину прибыли предприятия, помимо величины инвестиций, влияние оказывает большое количество дополнительных факторов (эффективность расходования средств, конъюнктура рынка, величина основных фондов предприятия и др.). Поэтому при исследовании зависимости между величиной инвестиций и прибыли ограничимся отысканием регрессионной зависимости между ними.

Корреляционное поле, построенное по результатам наблюдений исследуемых величин, представлено на рисунке 4.4.

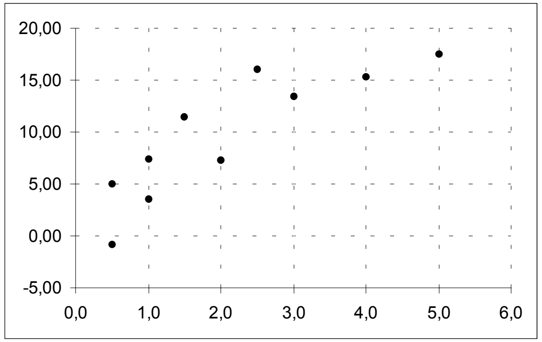


Рисунок 4.4 – Корреляционное поле случайных величин *X* и *Y*

Характер расположения точек на корреляционном поле позволяет сделать предположение о логарифмической регрессионной зависимости вида

$$\bar{y}(x) = \beta_0 + \beta_1 \ln x . \tag{4.15}$$

Параметры β_0 и β_1 найдём методом наименьших квадратов. Для этого составим функцию S (см. (4.3)), которая в случае логарифмической регрессии примет вид

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \bar{y}(x_i))^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \ln x_i)]^2 . \tag{4.16}$$

Для отыскания оценок параметров β_0 и β_1, минимизирующих функцию S составим и решим систему нормальных уравнений (4.4).

$$\begin{cases} \dfrac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0; \\ \dfrac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0. \end{cases} \Rightarrow \begin{cases} -2\sum\limits_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \ln x_i)] = 0; \\ -2\sum\limits_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \ln x_i)] \ln x_i = 0; \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \sum\limits_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \ln x_i)] = 0; \\ \sum\limits_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \ln x_i)] \ln x_i = 0; \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \sum\limits_{i=1}^{n} y_i - \beta_0 n - \beta_1 \sum\limits_{i=1}^{n} \ln x_i = 0; \\ \sum\limits_{i=1}^{n} y_i \ln x_i - \beta_0 \sum\limits_{i=1}^{n} \ln x_i - \beta_1 \sum\limits_{i=1}^{n} \ln^2 x_i = 0. \end{cases} \tag{4.17}$$

Здесь $\sum\limits_{i=1}^{n} y_i = 7{,}34 + 11{,}46 + \cdots + 17{,}54 = 96{,}04;$

$$\sum_{i=1}^{n} \ln x_i = \ln 1{,}0 + \ln 1{,}5 + \ln 4{,}0 + \cdots + \ln 5{,}0 = 4{,}723;$$

$$\sum_{i=1}^{n} \ln^2 x_i = \ln^2 1{,}0 + \ln^2 1{,}5 + \ln^2 4{,}0 + \cdots + \ln^2 5{,}0 = 8{,}1644;$$

$$\sum_{i=1}^{n} y_i \ln x_i = 7{,}34 \ln 1{,}0 + 11{,}46 \ln 1{,}5 + \cdots + 17{,}54 \ln 5{,}0 = 85{,}815.$$

Результаты вычислений сведём в таблицу 4.2.

Таблица 4.2 – **Результаты промежуточных вычислений**

N	x_i	y_i	$\ln x_i$	$\ln^2 x_i$	$y_i \ln^2 x_i$	$y(x_i) = \beta_0 + \beta_1 \ln x_i$
1	1,0	7,34	0,0000	0,0000	0,000	6,383
2	1,5	11,46	0,4055	0,1644	4,648	9,148
3	4,0	15,35	1,3863	1,9218	21,286	15,835
4	2,5	16,02	0,9163	0,8396	14,678	12,631
5	1,0	3,50	0,0000	0,0000	0,000	6,383
6	0,5	4,95	–0,6931	0,4805	–3,433	1,657
7	3,0	13,43	1,0986	1,2069	14,749	13,874
8	2,0	7,31	0,6931	0,4805	5,065	11,109
9	0,5	–0,87	–0,6931	0,4805	0,600	1,657
10	5,0	17,54	1,6094	2,5903	28,224	17,357
Σ		96,04	4,7230	8,1644	85,815	

После подстановки значений система уравнений (4.17) примет вид:

$$\begin{cases} 96{,}04 - \beta_0 \cdot 10 - \beta_1 \cdot 4{,}723 = 0; \\ 85{,}815 - \beta_0 \cdot 4{,}723 - \beta_1 \cdot 8{,}1644 = 0; \end{cases} \Rightarrow \begin{cases} 10\beta_0 + 4{,}723\beta_1 = 96{,}04; \\ 4{,}723\beta_0 + 8{,}1644\beta_1 = 85{,}815. \end{cases}$$

Выразив β_0 из первого уравнения и подставив во второе, получим:

$$\begin{cases} \beta_0 = \dfrac{96{,}04 - 4{,}723\beta_1}{10} = 9{,}604 - 0{,}4723\beta_1; \\ 4{,}723(9{,}604 - 0{,}4723\beta_1) + 8{,}1644\beta_1 = 85{,}815, \end{cases} \quad \text{откуда} \quad \beta_1 = \frac{48{,}8688}{5{,}9337} = 6{,}8183;$$

$$\beta_0 = 9{,}604 - 0{,}4723 \cdot 6{,}8183 = 6{,}3833.$$

Таким образом, оценки параметров уравнения регрессии $\hat{\beta}_0 = 6{,}3833$; $\hat{\beta}_1 = 6{,}8183$.

На рисунке 4.5 представлено корреляционное поле случайных величин X и Y с нанесённой на него линией регрессии.

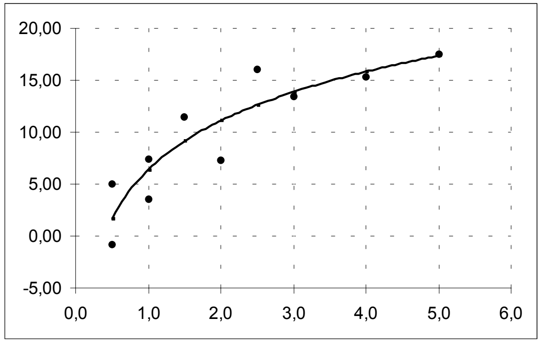


Рисунок 4.5 – Корреляционное поле случайных величин *X* и *Y* с нанесённой линией регрессии

Оценим качество описания зависимости между величиной инвестиций (X) и прибыли (Y), полученным уравнением регрессии с помощью коэффициента детерминации (4.10),

$$\hat{R}^2 = \sum_{i=1}^{n} (\bar{y}(x_i) - \bar{y})^2 \Big/ \sum_{i=1}^{n} (y_i - \bar{y})^2,$$

где $\bar{y}(x) = \beta_0 + \beta_1 \ln x$ – значение величины прибыли, предсказываемое уравнением регрессии, соответствующее значению x_i инвестиции; $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – среднее значение величины прибыли;

$$\hat{R}^2 = \frac{(6{,}383 - 9{,}604)^2 + (9{,}148 - 9{,}604)^2 + \dots + (17{,}357 - 9{,}604)^2}{(7{,}34 - 9{,}604)^2 + (11{,}46 - 9{,}604)^2 + \dots + (17{,}54 - 9{,}604)^2} = \frac{275{,}8568}{334{,}081} =$$
$$= 0{,}8257.$$

Расчётное значение коэффициента детерминации $\hat{R}^2 = 0{,}8257$ указывает на удовлетворительность описания зависимости между величиной инвестиций (X) и прибыли (Y), выбранным уравнением регрессии. Проверим, однако, значимость коэффициента детерминации с помощью статистики Фишера

$$F = 08257 \cdot \frac{10 - 2}{1 - 0{,}8257} = 37{,}9.$$

Критическое значение статистики Фишера для степеней свободы $\upsilon_1 = 1$ и $\nu_2 = 10 - 1 - 1 = 8$ и уровня значимости $\alpha = 0{,}05$ составляет $F_{0{,}05;1;8} = 5{,}317$. Поскольку расчётное значение статистики Фишера больше критического ($F = 37{,}9 > 5{,}317 = F_{0{,}05;1;8}$), то вычисленный коэффициент детерминации значимо отличается от нуля, и выбранное уравнение регрессионной зависимости между величинами инвестиций и прибыли может быть использовано в экономических исследованиях (в частности, для предсказания величины прибыли).

Для отыскания параметров уравнения регрессионной зависимости между величинами прибыли и инвестиций на ЭВМ воспользуемся пунктом Multiple Regression раздела REGRESSION ANALYSIS пакета StatGraphics (см. приложение *А*, п. 11). В поле зависимой переменной (Dep. Var.) введем имя выборки с.в. Y (например, LAB4.y), а в поле независимых переменных (Ind. Vars.) – выражение LOG(LAB4.x), что соответствует предложенному уравнению регрессии (4.15).

После нажатия клавиши F6 на экране появятся результаты расчета параметров регрессии:

Model fitting results for: LR4.y
Результаты расчетов модельного уравнения регрессии для с.в. Y

Independent variable *Независимая переменная*	coefficient *знач. коэф-та*	std. Error *ст. ошибка*	t-value *t-значение*	sig. level *уров. значимости*
CONSTANT *Значение параметра β_0*	6.381536	1.000421	6.3789	0.0002
LOG(LR4.x) *Значение параметра β_1*	6.820868	1.107186	6.1605	0.0003

R-SQ. (ADJ.) = 0.8041 *Коэф-т детерминации (приведенный)*		SE=2.697031 *Станд. ошибка оценивания R^2*	MAE = 2.027380	DurbWat = 2.343 *Коэф-т Дурбина-Уотсона*
Previously: *Предварительно:*	0.0000	0.000000	0.000000	0.000

10 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
10 наблюдений используются для предсказания 0 значений независимой переменной

Таблица результатов дисперсионного анализа (Analysis of Variance):

Analysis of Variance for the Full Regression
Дисперсионный анализ для полной регрессии

Source *Источник*	Sum of Squares *Сумма квадратов*	DF *Степ. своб.*	Mean Square *Средн. квадраты*	F-Ratio *F-отношение*	P-value *P-значение*
Model *модель*	276.064	1	276.064	37.9523	.0003
Error *остатки e_i*	58.1918	8	7.27397		
Total (Corr.) *Всего*	334.256	9			

R-squared = 0.825907
Коэф-т детерминации

Stnd. error of est. = 2.69703
Станд.ошибка оценивания R^2

R-squared (Adj. for d.f.) = 0.804145
Коэф-т детерминации (приведенный)

Durbin-Watson statistic = 2.34342
Коэф-т Дурбина-Уотсона

Порядок выполнения работы

1 Записать на диск выборки двух случайных величин, зависимость между которыми требуется исследовать (приложение А, п. 1).

2 По заданной двумерной выборке построить (на ЭВМ) корреляционное поле (приложение А, п. 10).

3 По виду корреляционного поля выдвинуть гипотезу о форме регрессионной зависимости между двумя случайными величинами.

4 Найти параметры гипотетического (предполагаемого) уравнения регрессии:
- вручную методом наименьших квадратов (МНК);
- на ЭВМ с помощью процедуры *REGRESSION ANALYSIS* (приложение А, п. 11). Сделать распечатку найденных значений;
- сравнить полученные уравнения;
- распечатать диаграмму рассеяния случайной величины (X, Y) с нанесенной на нее линией регрессии $У$ по X (приложение А, п. 11).

5 Оценить тесноту связи исследуемых случайных переменных с помощью коэффициента корреляции (в случае линейной зависимости) или с помощью коэффициента детерминации (в случае нелинейной зависимости).

6 Проверить значимость коэффициента корреляции (коэффициента детерминации).

7 Сформулировать выводы о зависимости между исследуемыми случайными величинами.

ПРИЛОЖЕНИЕ А
(*справочное*)
Работа с пакетом STATGRAPHICS

Для того чтобы начать работу с пакетом STATGRAPHICS, необходимо на диске D: или C: войти в каталог STAT и запустить на выполнение пакетный файл st.bat (для нормальной работы программы на рабочем диске должно быть не менее 2 Mb свободного места).

Файлы, созданные пользователями при работе с пакетом программ STATGRAPHICS, записываются в подкаталог DATA (STAT\DATA).

1 СОЗДАНИЕ ФАЙЛА ПЕРЕМЕННЫХ

1.1 В главном меню выбрать «A. DATA MANAGEMENT».

1.2 В подменю выбрать «2. File operations».

1.3 В поле «File name» ввести имя файла (латинские буквы и цифры, не более 8 символов).

1.4 В поле «Desired operation» ввести символ «B» (Create – создать) и нажать клавишу F6.

1.5 Выйти в главное меню, нажимая клавишу Esc.

2 ЗАПИСЬ ПЕРЕМЕННОЙ (ВЫБОРКИ) НА ДИСК

2.1 В главном меню выбрать «A. DATA MANAGEMENT».

2.2 В подменю выбрать «2. File operations».

2.3 В поле «File name» ввести имя созданного файла (для выбора файла из имеющегося списка использовать клавиши Ctrl + F7).

2.4 В поле «Desired operation» ввести символ «J» (Update – обновить) и нажать клавишу F6.

2.5 В окне просмотра содержимого файла ввести символ «N» (New – новый).

2.6 В поле «Enter name of new variable:» ввести имя переменной (латинские буквы и цифры, до 10 символов).

2.7 В появившееся окно ввести элементы выборки (целая часть числа от дробной отделяется точкой, друг от друга числа отделяются пробелами). После ввода всех чисел нажать «ENTER».

2.8 Введите текст комментария (до 25 символов; можно и не вводить). Нажать «ENTER».

2.9 Для ввода новой переменной (выборки) повторить всё с п. 2.3.

2.10 Выйти в главное меню, нажимая клавишу Esc – до тех пор, пока не появится главное меню системы.

3 ВЫЧИСЛЕНИЕ ОЦЕНОК ЧИСЛОВЫХ ХАРАКТЕРИСТИК

3.1 В главном меню выбрать «F. DESCREPTIVE METHODS».

3.2 В подменю выбрать «1. Summary statistics».

3.3 В поле «Data vectors:» ввести имя переменной (нажать F7, выбрать из каталога нужное имя, нажать ENTER).

3.4 Нажать клавишу F6.

3.5 На экране появится таблица с результатами вычислений. Перевод английских терминов приводится ниже.

Sample Size	Объём выборки
Average	Среднее
Median	Медиана
Mode	Мода
Geometric mean	Геометрическое среднее
Variance	Дисперсия
Std. Deviation	Стандартное отклонение
Std. Error	Стандартная ошибка
Minimum	Минимальное значение выборки
Maximum	Максимальное значение выборки
Range	Размах выборки
Lower quartile	Нижняя квартиль
Upper quartile	Верхняя квартиль
Interquar. range	Интерквартильная широта распределения
Skewness	Коэффициент асимметрии
Std. skewness	Отклонение коэффициента асимметрии от канонического значения
Kurtosis	Коэффициент эксцесса
Std. kurtosis	Отклонение коэффициента эксцесса от канонического значения

3.6 Для вывода результатов на печать подготовить принтер и нажать клавишу F4.

4 ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ И ДИСПЕРСИИ

4.1 В главном меню выбрать «G. Estimation and Testing».

4.2 В подменю выбрать «1. One-Sample Analysis».

4.3 В поле «Data vectors:» задать имя переменной, для которой строятся доверительные интервалы (нажать F7, выбрать из каталога нужное имя, нажать ENTER).

4.4 Нажать клавишу F6.

4.5 В появившемся меню можно изменить предлагаемую информацию в «окнах»:

Confidence Interval for Mean:	95 Percent	Доверительный интервал для среднего	95 %
Confidence Interval for Variance:	95 Percent	Доверительный интервал для дисперсии	95 %

4.6 Нажать клавишу F6 (2 раза).

4.7 На экране появятся результаты вычислений:

Sample Statistics:	Number of Obs.	Объём выборки
	Average	Среднее
	Variance	Дисперсия
	Std. Deviation	Среднеквадратическое отклонение
	Median	Медиана

Confidence Interval for Mean:		95 Percent
	Sample 1	Границы доверительного интервала для математического ожидания и соответствующее число степеней свободы (D.F.)
Confidence Interval for Variance:		95 Percent
	Sample 1	Границы доверительного интервала для дисперсии и соответствующее число степеней свободы (D.F.)

4.8 Для вывода результатов на печать подготовить принтер и нажать клавишу F4.

5 ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ ЗАДАННОМУ ЗНАЧЕНИЮ

5.1 В главном меню выбрать пункт «G. Estimation and Testing».

5.2 В подменю выбрать пункт «1. One-Sample Analysis».

5.3 Заполнить поле «Data vector», поместив в него имя выборки (по клавише F7). Нажать клавишу F6.

5.4 Заполнить поля:

Название поля	Назначение	
Confidence Interval for Mean	Доверительная вероятность для построения интервальной оценки математического ожидания	
Confidence Interval for variance	Доверительная вероятность для построения интервальной оценки дисперсии	
Hypothesis test for H0: Mean	Гипотетическое значение (a_0) математического ожидания для проверки гипотезы $H_0: a = a_0$	
Vs Alt:	Вид альтернативной гипотезы:	
	NE	$H_a: a \neq a_0$
	GT	$H_a: a > a_0$
	LT	$H_a: a < a_0$
At Alpha	Уровень значимости для проверки гипотезы	

5.5 После нажатия клавиши F6 просмотреть или распечатать получившиеся результаты (формулировка «so reject H0» после поля для ввода уровня значимости означает, что нулевая гипотеза для заданного уровня значимости отклоняется в пользу альтернативной; «so do not reject H0» – введенные данные не дают основания для отклонения нулевой гипотезы). При проверке гипотез генеральная дисперсия предполагается неизвестной.

6 ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ ДВУХ ВЫБОРОК

6.1 В главном меню выбрать пункт «G. Estimation and Testing».

6.2 В подменю выбрать пункт «2. Two-Sample Analysis».

6.3 Заполнить поля «Sample1» и «Sample2», поместив в него имена выборок (по клавише F7). Нажать клавишу F6.

6.4 Заполнить поля:

<table>
<tr><th>Название поля</th><th colspan="2">Назначение</th></tr>
<tr><td>Confidence Interval for Diff. In Means</td><td colspan="2">Доверительная вероятность для построения интервальной оценки разности математических ожиданий</td></tr>
<tr><td>Confidence Interval for Ratio of Variances</td><td colspan="2">Доверительная вероятность для построения интервальной оценки отношения дисперсий</td></tr>
<tr><td>Hypothesis test for H0: Diff=</td><td colspan="2">Гипотетическое значение (a_0) разности математических ожиданий для проверки гипотезы $H_0:\ a_1 - a_2 = a_0$, где a_1, a_2 – значения математических ожиданий для первой и второй генеральных совокупностей</td></tr>
<tr><td rowspan="4">Vs Alt:</td><td colspan="2">Вид альтернативной гипотезы:</td></tr>
<tr><td>NE</td><td>$H_a:\ a_1 - a_2 \neq a_0$</td></tr>
<tr><td>GT</td><td>$H_a:\ a_1 - a_2 > a_0$</td></tr>
<tr><td>LT</td><td>$H_a:\ a_1 - a_2 < a_0$</td></tr>
<tr><td>At Alpha</td><td colspan="2">Уровень значимости для проверки гипотезы</td></tr>
</table>

6.5 После нажатия клавиши F6 просмотреть или распечатать получившиеся результаты (п. 5.5). При проверке гипотез предполагается, что генеральные дисперсии неизвестны и равны между собой.

7 ПОСТРОЕНИЕ ГИСТОГРАММЫ

7.1 В главном меню выбрать «F. DESCREPTIVE METHODS».

7.2 В подменю выбрать «3. Frequency Histogram».

7.3 В окне «Data:» ввести имя переменной (нажать F7, выбрать из каталога нужное имя, нажать ENTER).

7.4 Нажать клавишу F6.

7.5 В появившемся меню при необходимости можно изменить предлагаемую информацию в окнах (переход между окнами – клавишей ENTER):

Type	Тип гистограммы (Continuous – для вещественных данных; Discrete – для целочисленных; выбор осуществляется клавишей пробела).
Lower limit	Нижняя граница первого интервала разбиения.
Upper limit	Верхняя граница последнего интервала разбиения.
No. of classes	Число интервалов разбиения.

7.6 Нажать клавишу F6.

7.7 Для вывода полученного графика на печать подготовить принтер, нажать клавишу F4, затем − ENTER.

8 ПОДБОР ЗАКОНА РАСПРЕДЕЛЕНИЯ

8.1 В главном меню выбрать «H. Distribution Functions».

8.2 В подменю выбрать «1. Distribution Fitting».

8.3 В поле «Data vector» ввести имя переменной (нажать F7, выбрать

нужное имя, нажать ENTER). В поле «Distribution Number» указать номер требуемой функции распределения.

8.4 Нажать F6.

8.5 На экране появились значения параметров выбранного закона распределения.

8.6 Нажать F6.

8.7 В появившемся подменю выбрать «CHI-SQUARE test».

8.8 В появившемся меню можно изменить предлагаемые значения в окнах

Lower limit	Нижняя граница первого интервала разбиения
Upper limit	Верхняя граница последнего интервала разбиения
No. of classes	Число интервалов разбиения

8.9 Нажать F6.

8.10 В появившейся на экране таблице приведены результаты расчётов:

Lower limit	Нижняя граница интервала разбиения
Upper limit	Верхняя граница интервала разбиения
Observed Frequency	Наблюденные частоты
Expected Frequency	Теоретические (ожидаемые) частоты

Под таблицей указано расчётное значение критерия χ^2 (Chisquare), число степеней свободы (d. f.), соответствующий полученному значению χ^2 уровень значимости (sig. level).

Полагают, что нет оснований для отклонения проверяемой гипотезы, если выполняются условия:

а) расчётное (выборочное) значение критерия χ^2 не превышает критического значения $\chi^2_{\alpha,\nu}$;

б) полученное на ЭВМ значение sig. level больше 0,1.

Если указанные условия не выполняются, проверяемая гипотеза должна быть отвергнута. Для проверки другой гипотезы повторить операции начиная с п. 8.3.

8.11 Для вывода полученной таблицы на печать, подготовить принтер, нажать клавишу F4, затем – Enter.

9 ПОСТРОЕНИЕ СОВМЕСТНОГО ГРАФИЧЕСКОГО ИЗОБРАЖЕНИЯ ЭМПИРИЧЕСКОГО (ВЫБОРОЧНОГО) И ТЕОРЕТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

9.1 а) Находясь в условиях п. 8.10, два раза нажать клавишу Esc, в появившемся подменю выбрать пункт «Histogram» или

б) находясь в главном меню, выполнить операции пп. 8.1 – 8.6, в появившемся меню выбрать пункт «Histogram».

9.2 Если изучается непрерывная случайная величина, можно таким же образом, как и в п. 8.8, изменить значения, указанные в окнах, нажать клавишу F6.

9.3 Для вывода полученного графика на печать подготовить принтер, нажать клавиши F4, Enter.

10 ПОСТРОЕНИЕ КОРРЕЛЯЦИОННОГО ПОЛЯ

10.1 В главном меню выбрать пункт «E. Plotting Functions».

10.2 В подменю выбрать «1. X-Y Line and Scatterplots».

10.3 Заполнить поля:

X – имя независимой переменной (через клавишу F7);

Y – имя зависимой переменной (через клавишу F7).

10.4 Нажать клавишу F6.

11 РАСЧЕТ ПАРАМЕТРОВ УРАВНЕНИЯ РЕГРЕССИИ

11.1 В главном меню выбрать «Regression Analysis» (регрессионный анализ).

11.2 В подменю выбрать «Simple Regression» (простая регрессия).

11.3 Заполнить поля:

Dependent variable	– имя зависимой переменной
Independent variable	– имя независимой переменной
Model:	
Linear	– линейная регрессионная модель $g(x) = \beta_0 + \beta_1 x$
Exponential	– экспоненциальная регрессионная модель $g(x) = \exp(\beta_0 + \beta_1 x)$
Multiplicative	– мультипликативная регрессионная модель $g(x) = \beta_0 x\hat{}\beta_1$
Reciprocal	– гиперболическая регрессионная модель $g(x) = 1/(\beta_0 + \beta_1 x)$

11.4 Нажать клавишу *F*6.

11.5 Распечатать найденные значения параметров регрессии и коэффициента корреляции (детерминации), нажав клавишу F4.

11.6 Нажать клавишу F6. В появившемся меню выбрать пункт «Plot fitted line» для вывода на экран диаграммы рассеяния и рассчитанной линии регрессии. Снова нажать F6.

11.7 Распечатать данный рисунок, предварительно нажав клавишу «Esc», затем – F3.

Если в пункте «Simple Regression» отсутствует нужная регрессионная модель, то в подменю (шаг 11.2.) следует выбрать «Stepwise VariableSelection» (пошаговый регрессионный анализ).

11.3. Заполнить поля:

Dependent variable – имя зависимой переменной

Independent variables – список возможных слагаемых уравнения регрессии. Например, если независимая переменная имеет имя «LAB1. x», то для уравнения регрессии (4.6) такими слагаемыми будут:

LAB1.x^(–1)   LAB1.x   LAB1.x^2   LAB1.x^3   exp(LAB1.x)

11.4 Нажать клавишу F6.

11.5 Нажимать клавишу «Enter» до тех пор, пока на экране не появятся результаты расчетов («Model fitting results for:»).

11.6 Распечатать значения найденных параметров регрессии и коэффициента корреляции (детерминации) с помощью клавиши F4.

11.7 Нажать «Esc». В появившемся меню выбрать пункт «Interval plots» и нажать F6.

11.8 Нажать клавишу «V».

11.9 Ввести цифру «1».

11.10 Ввести имя независимой переменной.

11.11 Распечатать полученную диаграмму рассеяния с нанесенной на нее линией регрессии, нажав клавишу F3.

ПРИЛОЖЕНИЕ Б

(*справочное*)

Таблица значений функции плотности стандартного нормального распределения

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

ПРИЛОЖЕНИЕ В

(*справочное*)

Таблица значений функции Лапласа

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

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	0,3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	0,4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3,0	0,49865									
3,1	0,49903									
3,2	0,49931									
3,3	0,49952									
3,4	0,49966									
3,6	0,499841									
3,8	0,499928									
4,0	0,499968									
4,5	0,499997									
5,0	0,4999997									

ПРИЛОЖЕНИЕ Г

(*справочное*)

Критические точки распределения χ^2

Степени свободы ν	Уровень значимости α								
	0,001	0,01	0,025	0,05	0,1	0,9	0,95	0,975	0,99
1	10,827	6,635	5,024	3,841	2,706	0,016	0,0039	0,00098	0,00016
2	13,815	9,210	7,378	5,991	4,605	0,211	0,103	0,051	0,020
3	16,266	11,345	9,348	7,815	6,251	0,584	0,352	0,216	0,115
4	18,466	13,277	11,143	9,488	7,779	1,064	0,711	0,484	0,297
5	20,515	15,086	12,832	11,070	9,236	1,610	1,145	0,831	0,554
6	22,457	16,812	14,449	12,592	10,645	2,204	1,635	1,237	0,872
7	24,321	18,475	16,013	14,067	12,017	2,833	2,167	1,690	1,239
8	26,124	20,090	17,535	15,507	13,362	3,490	2,733	2,180	1,647
9	27,877	21,666	19,023	16,919	14,684	4,168	3,325	2,700	2,088
10	29,588	23,209	20,483	18,307	15,987	4,865	3,940	3,247	2,558
11	31,264	24,725	21,920	19,675	17,275	5,578	4,575	3,816	3,053
12	32,909	26,217	23,337	21,026	18,549	6,304	5,226	4,404	3,571
13	34,527	27,688	24,736	22,362	19,812	7,041	5,892	5,009	4,107
14	36,124	29,141	26,119	23,685	21,064	7,790	6,571	5,629	4,660
15	37,698	30,578	27,488	24,996	22,307	8,547	7,261	6,262	5,229
16	39,252	32,000	28,845	26,296	23,542	9,312	7,962	6,908	5,812
17	40,791	33,409	30,191	27,587	24,769	10,085	8,672	7,564	6,408
18	42,312	34,805	31,526	28,869	25,989	10,865	9,390	8,231	7,015
19	43,819	36,191	32,852	30,144	27,204	11,651	10,117	8,907	7,633
20	45,314	37,566	34,170	31,410	28,412	12,443	10,851	9,591	8,260
21	46,796	38,932	35,479	32,671	29,615	13,240	11,591	10,283	8,897
22	48,268	40,289	36,781	33,924	30,813	14,041	12,338	10,982	9,542
23	49,728	41,638	38,076	35,172	32,007	14,848	13,091	11,689	10,196
24	51,179	42,980	39,364	36,415	33,196	15,659	13,848	12,401	10,856
25	52,619	44,314	40,646	37,652	34,382	16,473	14,611	13,120	11,524
26	54,051	45,642	41,923	38,885	35,563	17,292	15,379	13,844	12,198
27	55,475	46,963	43,195	40,113	36,741	18,114	16,151	14,573	12,878
28	56,892	48,278	44,461	41,337	37,916	18,939	16,928	15,308	13,565
29	58,301	49,588	45,722	42,557	39,087	19,768	17,708	16,047	14,256
30	59,702	50,892	46,979	43,773	40,256	20,599	18,493	16,791	14,953
31	61,098	52,191	48,232	44,985	41,422	21,434	19,281	17,539	15,655
32	62,487	53,486	49,480	46,194	42,585	22,271	20,072	18,291	16,362
33	63,869	54,775	50,725	47,400	43,745	23,110	20,867	19,047	17,073
34	65,247	56,061	51,966	48,602	44,903	23,952	21,664	19,806	17,789
35	66,619	57,342	53,203	49,802	46,059	24,797	22,465	20,569	18,509
36	67,985	58,619	54,437	50,998	47,212	25,643	23,269	21,336	19,233
37	69,348	59,893	55,668	52,192	48,363	26,492	24,075	22,106	19,960
38	70,704	61,162	56,895	53,384	49,513	27,343	24,884	22,878	20,691
39	72,055	62,428	58,120	54,572	50,660	28,196	25,695	23,654	21,426
40	73,403	63,691	59,342	55,758	51,805	29,051	26,509	24,433	22,164

ПРИЛОЖЕНИЕ Д

(*справочное*)

Критические точки распределения Стьюдента

Степени свободы ν	Уровень значимости α (односторонняя критическая область)							
	0,2	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
1	1,376	3,078	6,314	12,706	31,821	63,656	318,29	636,58
2	1,061	1,886	2,920	4,303	6,965	9,925	22,328	31,600
3	0,978	1,638	2,353	3,182	4,541	5,841	10,214	12,924
4	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,920	1,476	2,015	2,571	3,365	4,032	5,894	6,869
6	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,862	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,689
28	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,660
30	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,849	1,299	1,676	2,009	2,403	2,678	3,261	3,496
60	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
70	0,847	1,294	1,667	1,994	2,381	2,648	3,211	3,435
80	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,416
90	0,846	1,291	1,662	1,987	2,368	2,632	3,183	3,402
100	0,845	1,290	1,660	1,984	2,364	2,626	3,174	3,390
150	0,844	1,287	1,655	1,976	2,351	2,609	3,145	3,357
200	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,340
500	0,842	1,283	1,648	1,965	2,334	2,586	3,107	3,310
∞	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,290
Степени свободы	0,4	0,2	0,1	0,05	0,02	0,01	0,002	0,001
ν	Уровень значимости α (двусторонняя критическая область)							

ПРИЛОЖЕНИЕ Е

(*справочное*)

Критические точки распределения Фишера

(число степеней свободы бóльшей дисперсии – ν_1, меньшей – ν_2)

Уровень значимости $\alpha = 0{,}05$												
ν_2	ν_1											
	1	2	3	4	5	6	8	12	16	24	50	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	238,88	243,90	246,47	249,05	251,77	254,31
2	18,513	19,000	19,164	19,247	19,296	19,329	19,371	19,412	19,433	19,454	19,476	19,496
3	10,128	9,552	9,277	9,117	9,013	8,941	8,845	8,745	8,692	8,638	8,581	8,526
4	7,709	6,944	6,591	6,388	6,256	6,163	6,041	5,912	5,844	5,774	5,699	5,628
5	6,608	5,786	5,409	5,192	5,050	4,950	4,818	4,678	4,604	4,527	4,444	4,365
6	5,987	5,143	4,757	4,534	4,387	4,284	4,147	4,000	3,922	3,841	3,754	3,669
7	5,591	4,737	4,347	4,120	3,972	3,866	3,726	3,575	3,494	3,410	3,319	3,230
8	5,318	4,459	4,066	3,838	3,688	3,581	3,438	3,284	3,202	3,115	3,020	2,928
9	5,117	4,256	3,863	3,633	3,482	3,374	3,230	3,073	2,989	2,900	2,803	2,707
10	4,965	4,103	3,708	3,478	3,326	3,217	3,072	2,913	2,828	2,737	2,637	2,538
11	4,844	3,982	3,587	3,357	3,204	3,095	2,948	2,788	2,701	2,609	2,507	2,404
12	4,747	3,885	3,490	3,259	3,106	2,996	2,849	2,687	2,599	2,505	2,401	2,296
13	4,667	3,806	3,411	3,179	3,025	2,915	2,767	2,604	2,515	2,420	2,314	2,206
14	4,600	3,739	3,344	3,112	2,958	2,848	2,699	2,534	2,445	2,349	2,241	2,131
15	4,543	3,682	3,287	3,056	2,901	2,790	2,641	2,475	2,385	2,288	2,178	2,066
16	4,494	3,634	3,239	3,007	2,852	2,741	2,591	2,425	2,333	2,235	2,124	2,010
17	4,451	3,592	3,197	2,965	2,810	2,699	2,548	2,381	2,289	2,190	2,077	1,960
18	4,414	3,555	3,160	2,928	2,773	2,661	2,510	2,342	2,250	2,150	2,035	1,917
19	4,381	3,522	3,127	2,895	2,740	2,628	2,477	2,308	2,215	2,114	1,999	1,878
20	4,351	3,493	3,098	2,866	2,711	2,599	2,447	2,278	2,184	2,082	1,966	1,843
21	4,325	3,467	3,072	2,840	2,685	2,573	2,420	2,250	2,156	2,054	1,936	1,812
22	4,301	3,443	3,049	2,817	2,661	2,549	2,397	2,226	2,131	2,028	1,909	1,783
23	4,279	3,422	3,028	2,796	2,640	2,528	2,375	2,204	2,109	2,005	1,885	1,757
24	4,260	3,403	3,009	2,776	2,621	2,508	2,355	2,183	2,088	1,984	1,863	1,733
25	4,242	3,385	2,991	2,759	2,603	2,490	2,337	2,165	2,069	1,964	1,842	1,711
26	4,225	3,369	2,975	2,743	2,587	2,474	2,321	2,148	2,052	1,946	1,823	1,691
27	4,210	3,354	2,960	2,728	2,572	2,459	2,305	2,132	2,036	1,930	1,806	1,672
28	4,196	3,340	2,947	2,714	2,558	2,445	2,291	2,118	2,021	1,915	1,790	1,654
29	4,183	3,328	2,934	2,701	2,545	2,432	2,278	2,104	2,007	1,901	1,775	1,638
30	4,171	3,316	2,922	2,690	2,534	2,421	2,266	2,092	1,995	1,887	1,761	1,622
40	4,085	3,232	2,839	2,606	2,449	2,336	2,180	2,003	1,904	1,793	1,660	1,509
50	4,034	3,183	2,790	2,557	2,400	2,286	2,130	1,952	1,850	1,737	1,599	1,438
60	4,001	3,150	2,758	2,525	2,368	2,254	2,097	1,917	1,815	1,700	1,559	1,389
70	3,978	3,128	2,736	2,503	2,346	2,231	2,074	1,893	1,790	1,674	1,530	1,353
80	3,960	3,111	2,719	2,486	2,329	2,214	2,056	1,875	1,772	1,654	1,508	1,325
90	3,947	3,098	2,706	2,473	2,316	2,201	2,043	1,861	1,757	1,639	1,491	1,302
100	3,936	3,087	2,696	2,463	2,305	2,191	2,032	1,850	1,746	1,627	1,477	1,283
150	3,904	3,056	2,665	2,432	2,274	2,160	2,001	1,817	1,711	1,590	1,436	1,223
200	3,888	3,041	2,650	2,417	2,259	2,144	1,985	1,801	1,694	1,572	1,415	1,189
500	3,860	3,014	2,623	2,390	2,232	2,117	1,957	1,772	1,664	1,539	1,376	1,113
∞	3,841	2,996	2,605	2,372	2,214	2,099	1,938	1,752	1,644	1,517	1,350	1,000

Уровень значимости $\alpha = 0{,}01$												
ν_2	ν_1											
	1	2	3	4	5	6	8	12	16	24	50	∞
1	4052,2	4999,3	5403,5	5624,3	5764,0	5859,0	5981,0	6106,7	6170,0	6234,3	6302,3	6365,6
2	98,502	99,000	99,164	99,251	99,302	99,331	99,375	99,419	99,437	99,455	99,477	99,499
3	34,116	30,816	29,457	28,710	28,237	27,911	27,489	27,052	26,826	26,597	26,354	26,125
4	21,198	18,000	16,694	15,977	15,522	15,207	14,799	14,374	14,154	13,929	13,690	13,463
5	16,258	13,274	12,060	11,392	10,967	10,672	10,289	9,888	9,680	9,466	9,238	9,020
6	13,745	10,925	9,780	9,148	8,746	8,466	8,102	7,718	7,519	7,313	7,091	6,880
7	12,246	9,547	8,451	7,847	7,460	7,191	6,840	6,469	6,275	6,074	5,858	5,650
8	11,259	8,649	7,591	7,006	6,632	6,371	6,029	5,667	5,477	5,279	5,065	4,859
9	10,562	8,022	6,992	6,422	6,057	5,802	5,467	5,111	4,924	4,729	4,517	4,311
10	10,044	7,559	6,552	5,994	5,636	5,386	5,057	4,706	4,520	4,327	4,115	3,909
11	9,646	7,206	6,217	5,668	5,316	5,069	4,744	4,397	4,213	4,021	3,810	3,602
12	9,330	6,927	5,953	5,412	5,064	4,821	4,499	4,155	3,972	3,780	3,569	3,361
13	9,074	6,701	5,739	5,205	4,862	4,620	4,302	3,960	3,778	3,587	3,375	3,165
14	8,862	6,515	5,564	5,035	4,695	4,456	4,140	3,800	3,619	3,427	3,215	3,004
15	8,683	6,359	5,417	4,893	4,556	4,318	4,004	3,666	3,485	3,294	3,081	2,868
16	8,531	6,226	5,292	4,773	4,437	4,202	3,890	3,553	3,372	3,181	2,967	2,753
17	8,400	6,112	5,185	4,669	4,336	4,101	3,791	3,455	3,275	3,083	2,869	2,653
18	8,285	6,013	5,092	4,579	4,248	4,015	3,705	3,371	3,190	2,999	2,784	2,566
19	8,185	5,926	5,010	4,500	4,171	3,939	3,631	3,297	3,116	2,925	2,709	2,489
20	8,096	5,849	4,938	4,431	4,103	3,871	3,564	3,231	3,051	2,859	2,643	2,421
21	8,017	5,780	4,874	4,369	4,042	3,812	3,506	3,173	2,993	2,801	2,584	2,360
22	7,945	5,719	4,817	4,313	3,988	3,758	3,453	3,121	2,941	2,749	2,531	2,305
23	7,881	5,664	4,765	4,264	3,939	3,710	3,406	3,074	2,894	2,702	2,483	2,256
24	7,823	5,614	4,718	4,218	3,895	3,667	3,363	3,032	2,852	2,659	2,440	2,211
25	7,770	5,568	4,675	4,177	3,855	3,627	3,324	2,993	2,813	2,620	2,400	2,169
26	7,721	5,526	4,637	4,140	3,818	3,591	3,288	2,958	2,778	2,585	2,364	2,131
27	7,677	5,488	4,601	4,106	3,785	3,558	3,256	2,926	2,746	2,552	2,330	2,097
28	7,636	5,453	4,568	4,074	3,754	3,528	3,226	2,896	2,716	2,522	2,300	2,064
29	7,598	5,420	4,538	4,045	3,725	3,499	3,198	2,868	2,689	2,495	2,271	2,034
30	7,562	5,390	4,510	4,018	3,699	3,473	3,173	2,843	2,663	2,469	2,245	2,006
40	7,314	5,178	4,313	3,828	3,514	3,291	2,993	2,665	2,484	2,288	2,058	1,805
50	7,171	5,057	4,199	3,720	3,408	3,186	2,890	2,563	2,382	2,183	1,949	1,683
60	7,077	4,977	4,126	3,649	3,339	3,119	2,823	2,496	2,315	2,115	1,877	1,601
70	7,011	4,922	4,074	3,600	3,291	3,071	2,777	2,450	2,268	2,067	1,826	1,540
80	6,963	4,881	4,036	3,563	3,255	3,036	2,742	2,415	2,233	2,032	1,788	1,494
90	6,925	4,849	4,007	3,535	3,228	3,009	2,715	2,389	2,206	2,004	1,759	1,457
100	6,895	4,824	3,984	3,513	3,206	2,988	2,694	2,368	2,185	1,983	1,735	1,427
150	6,807	4,749	3,915	3,447	3,142	2,924	2,632	2,305	2,122	1,918	1,665	1,331
200	6,763	4,713	3,881	3,414	3,110	2,893	2,601	2,275	2,091	1,886	1,629	1,279
500	6,686	4,648	3,821	3,357	3,054	2,838	2,547	2,220	2,036	1,829	1,566	1,164
∞	6,635	4,605	3,782	3,319	3,017	2,802	2,511	2,185	2,000	1,791	1,523	1,000

СПИСОК ПРИНЯТЫХ УСЛОВНЫХ ОБОЗНАЧЕНИЙ

ДСВ – дискретная случайная величина;
МНК – метод наименьших квадратов;
МС – математическая статистика;
МСВ – многомерная случайная величина;
НСВ – непрерывная случайная величина;
ПСП – пакет статистических программ;
с. в. – случайная величина, случайные величины.

СПИСОК ЛИТЕРАТУРЫ

1 *Афифи А., Эйзен С.* Статистический анализ: Подход с использованием ЭВМ. – М.: Мир, 1982. – 420 с.

2 *Герасимович А. И.* Математическая статистика. – Мн.: Вышэйшая школа, 1983. – 279 с.

3 *Математическая энциклопедия.* – М.: Советская энциклопедия. В 4 т. Т. 3, 4. 1982, 1984.

4 *Серегина В. С.* Решение инженерных задач методами математической статистики: Учеб. пособие. – Гомель: БелГУТ, 1994. – 107 с.

5 *Ферстер Э., Ренц Б.* Методы корреляционного и регрессионного анализа. – М.: Финансы и статистика, 1983. – 302 с.

6 *Четыркин Е. М., Калихман И. Л.* Вероятность и статистика. – М.: Финансы и статистика, 1982. – 319 с.

ОГЛАВЛЕНИЕ