МИНОБРНАУКИ РФ
Государственное образовательное учреждение
высшего образования

Государственный университет телекоммуникаций

Методические указания к курсовой работе

по дисциплине «Методы исследования и моделирования информационных процессов и технологий»

Санкт-Петербург
2018

УЧЕБНОЕ ПОСОБИЕ

ДЛЯ ВЫПОЛНЕНИЯ КУРСОВОЙ РАБОТЫ

СанктПетербург

2018

Оглавление

2. Первичная обработка экспериментальных данных и точечное оценивание параметров распределения

2.1. Первичная обработка экспериментальных данных

Первичная обработка полученных в результате случайного эксперимента данных включает в себя:
– построение интервального (группированного) ряда,
– построение эмпирической функции распределения,
– получение точечных оценок параметров распределения,
– предварительное предположение о характере генерального распределения.

Пусть в результате случайного эксперимента наблюдаются (измеряются) значения случайной величины X с функцией распределения $F(x)$, а x_i – реализация эксперимента (в эксперименте имело место событие $\{X = x_i\}$). Если провести n экспериментов (измерений), то в результате получим n значений измеряемой случайной величины X, которые мы обозначим $x_1, x_2, ..., x_n$, и которые образуют *выборку объема* n .

Случайную величину X называют при этом *генеральной случайной величиной,* а множество всех возможных значений наблюдаемой случайной величины X, которые могут реализоваться в рамках проводимого эксперимента – *генеральной совокупностью.*

Случайная величина X может быть дискретной или непрерывной, принимать конечное или бесконечное множество значений. Соответственно и генеральная совокупность может представлять собой конечное или бесконечное множество, дискретное или непрерывное.

Числа $x_1, x_2, ..., x_n$ называются *наблюдениями* или *выборочными значениями* случайной величины.

О самой случайной величине X может быть ничего не известно или, например, известен закон ее распределения, но не известны параметры распределения.

ЗАМЕЧАНИЕ 1

Выборку $x_1, x_2, ..., x_n$ можно рассматривать как результат одновременного наблюдения n независимых одинаково распределенных случайных величин $X_1, X_2, \dots, X_n$.

Для того чтобы на основе выборки делать выводы о случайной величине X, объем выборки n должен быть достаточно большим, а условия эксперимента должны обеспечивать одинаковые условия реализации возможных значений наблюдаемой случайной величины. Такая выборка сохраняет свойства генеральной совокупности и называется *представительной* (репрезентативной).

ЗАМЕЧАНИЕ 2

В практических статистических расчетах будем считать выборки большими или малыми в зависимости от условий $n \ge 30$ или $n < 30$.

1. Построение интервального статистического ряда

При статистической обработке экспериментальных данных генеральное распределение заменяется так называемым *выборочным распределением*, т.е. распределением *выборочной* случайной величины X^*, принимающей выборочные значения $x_1, x_2, ..., x_n$ с вероятностями $\frac{1}{n}$.

На первом этапе следует по заданной выборке $x_1, x_2, ..., x_n$ объема n построить *интервальный (группированный)* статистический ряд (табл. 1.1).

Для этого необходимо:

1) упорядочить выборку, т.е. составить так называемый *вариационный ряд*

$$x_1^* \le x_2^* \le x_3^* \le \dots \le x_{n-1}^* \le x_n^*, \tag{2.1}$$

в котором упорядоченные значения $x_1^*, x_2^*, ..., x_n^*$ называют порядковыми статистиками;

2) найти *диапазон выборки* $\left[x_1^*;\, x_n^*\right]$ и *размах выборки* R_B по формуле:

$$R_B = x_n^* - x_1^*; \tag{2.2}$$

3) для заданного объема выборки n найти оптимальное число интервалов l, на которые разбивается диапазон выборки. Рекомендуется выбирать:

$$l = \begin{cases} l \le 5\lg n, \quad 6 \le l \le 20 \\ \left[\sqrt{n}\right], \quad 5 \le l \le 25 \\ \left[1 + 3{,}32\lg n\right], \quad 5 \le l \le 25 \end{cases}; \tag{2.3}$$

4) найти длину каждого интервала h по формуле

$$h = \frac{R_B}{l}; \tag{2.4}$$

5) чтобы исключить тот случай, когда какое–то выборочное значение попадет на границу интервала и надо решать к какому интервалу его отнести, весь диапазон выборки $\left[x_1^*;\, x_n^*\right]$ нужно разбить на полуоткрытые интервалы

$$I_i = \left[a_{i-1}, a_i\right),\ I_l = \left[a_{l-1}, a_l\right],$$

границы которых определяются формулами:

$$a_0 = x_1^*,\ a_i = a_0 + i \cdot h,\ i = 1, ..., l; \qquad (2.5)$$

ЗАМЕЧАНИЕ 1

Поскольку вычисленное по формуле (2.4) значение h является приближенным, то при округлении по известным правилам может оказаться, что $a_l < x_{\max}^*$. Чтобы этого не произошло, рекомендуется вычислять h с округлением в большую сторону. Это означает, что если вычисления проводятся с точностью до $\varepsilon = 0{,}01$, то при вычисленном $h = 0{,}7621...$, следует положить $h \approx 0{,}77$. Тем самым мы увеличим диапазон выборки, но обеспечим равную длину интервалов и попадание всех выборочных значений в диапазон выборки.

6) выписать полученные интервалы

$$I_1 = [a_0;\ a_1);\ I_2 = [a_1;\ a_2);\ I_3 = [a_2;\ a_3);\ ....;\ I_l = [a_{l-1};\ a_l];$$

7) для каждого интервала I_i с помощью вариационного ряда (2.1) вычислить n_i – количество выборочных значений попавших в этот интервал;

ЗАМЕЧАНИЕ 2

К этим интервалам можно добавить два интервала $I_0 = (-\infty;\ a_0)$ и $I_{l+1} = (a_l;\ +\infty)$, если есть основания предполагать, что генеральная случайная величина X распределена на всей числовой оси или даже положить $a_0 = -\infty$ и $a_l = +\infty$.

8) все выборочные значения, попавшие в интервал I_i $(i = 1, 2, ..., l)$, принять равными его середине, а середины интервалов группированного ряда $\overline{a}_i$ вычислить по формулам:

$$\overline{a}_i = \frac{a_{i-1} + a_i}{2},\ i = 1, 2, ..., l; \qquad (2.6)$$

9) вычислить частоты (статистические оценки вероятностей, с которыми выборочная случайная величина X^* принимает значения $\overline{a}_i$, $i = 1, 2, .., n$) по формуле:

$$p_i^* = \frac{n_i}{n}, \qquad (2.7)$$

где n_i – число выборочных значений (наблюдений), попавших в интервал I_i;

10) записать группированный (интервальный) статистический ряд (табл. 2.1), в котором указаны интервалы, а также середины интервалов и частоты, вычисленные по формулам (2.6) и (2.7) соответственно.

Таблица 2.1

I_i	I_1	I_2	....	I_l	$\sum$
$\overline{a}_i$	$\overline{a}_1$	$\overline{a}_2$	....	$\overline{a}_l$	–
p_i^*	p_1^*	p_2^*	....	p_l^*	1

2. Построение эмпирической функции распределения

Если интервальный ряд построен, то далее строится дискретный статистический ряд (табл. 2.2). В таблицу 2.2 вводятся вычисленные середины интервалов, соответствующие частоты из таблицы 2.1, а также накопленные *частоты* z_i, которые вычисляются по формулам:

$$z_1 = p_1^*;\ z_2 = p_1^* + p_2^*; ...;\ z_i = \sum_{k=1}^{i} p_k^*.$$

Таблица 2.2

$\overline{a}_i$	$\overline{a}_1$	$\overline{a}_2$	$\overline{a}_3$	....	$\overline{a}_l$	$\sum$
p_i^*	p_1^*	p_2^*	p_3^*	....	p_l^*	1
z_i	z_1	z_2	z_3	....	1	

Построенный дискретный статистический ряд представляет собой приближенное распределение выборочной случайной величины X^*, а частоты p_i^* являются статистическими оценками вероятностей того, что выборочное значение равно $\overline{a}_i$.

Для интервального ряда в качестве приближения функции распределения генеральной случайной величины X можно рассматривать так называемую *эмпирическую функцию распределения* $F_n^*(x)$, которая представляет собой функцию распределения выборочной случайной величины X^* и которая определяется формулой:

$$F_n^*(x) = \sum_{i:\overline{a}_i < x} p_i^* . \qquad (2.8)$$

Аналитическое выражение $F_n^*(x)$ через накопленные частоты z_i имеет вид:

$$F_n^*(x) = \begin{cases} 0, & x \le \overline{a}_1 \\ z_1, & \overline{a}_1 < x \le \overline{a}_2 \\ z_2, & \overline{a}_2 < x \le \overline{a}_3 \\ \dots\dots\dots\dots & \\ z_{l-1}, & \overline{a}_{l-1} < x \le \overline{a}_l \\ 1, & x > \overline{a}_l \end{cases} . \qquad (2.9)$$

Эмпирическая функция распределения представляет собой кусочно–постоянную функцию со скачками в серединах интервалов $\overline{a}_i$ $i = 1, 2, ..., l$. На рис. 2.1 дан пример ее графика при четырех интервалах статистического ряда.

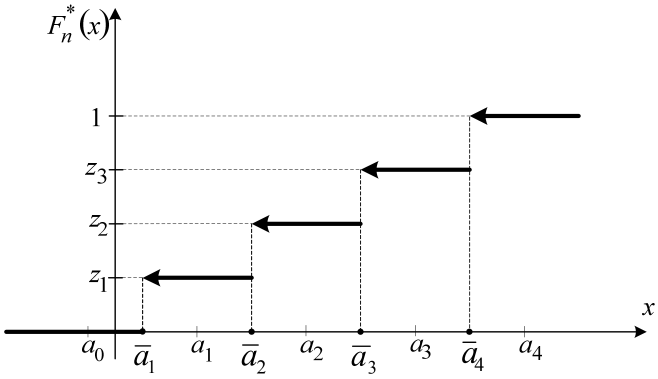

Рис. 2.1

Если X – непрерывная случайная величина, то в качестве статистического аналога функции распределения генеральной совокупности используют *кумуляту*.

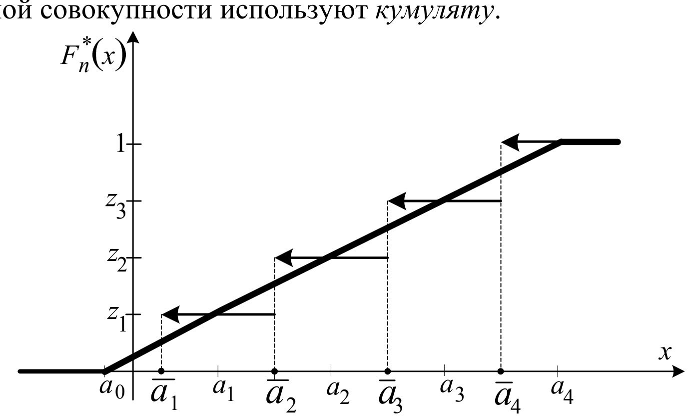

Рис. 2.2

Кумуляту строят как непрерывную ломаную линию, состоящую из отрезков, соединяющих точки $(a_0, 0)$ и $(a_1; z_1)$, $(a_1; z_1)$ и $(a_2; z_2)$, $(a_2; z_2)$ и $(a_3; z_3)$, ..., $(a_{l-1}; z_{l-1})$ и $(a_l; z_l)$, а также двух полупрямых: $y=0$, при $x \le a_0$ и $y=1$, при $x > a_l$. На рисунке 2.2 построена кумулята для графика функции распределения, показанной на рисунке 2.1.

3. Эмпирическая плотность распределения. Гистограмма и полигон

Статистическим аналогом плотности генерального распределения является эмпирическая плотность распределения $f_n^*(x)$, которая представляет собой плотность выборочной случайной величины X^* и вычисляется по формуле:

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

где частота p_i^* выбирается из таблицы 1.2; h – длина интервала.

График эмпирической плотности распределения $f_n^*(x)$ строится в виде кусочно–постоянной над интервалами $I_i = (a_{i-1}; a_i)$ линии и называется *гистограммой.*

Чтобы построить гистограмму следует:

1) вычислить значения эмпирической плотности по формуле (2.10);

2) внести в таблицу 2.3 середины интервалов $\bar{a}_i$ и значения эмпирической плотности $f_n^*(x)$;

Таблица 2.3

$\bar{a}_i$	$\bar{a}_1$	$\bar{a}_2$	………	$\bar{a}_l$
$f_i^*(x)$	$f_1^*(x)$	$f_2^*(x)$	………..	$f_l^*(x)$

3) записать аналитическое выражение для эмпирической плотности распределения $f_n^*(x)$ на каждом из интервалов I_i при всех $i = 1, 2, ..., l$, т.е.

$$f_n^*(x) = \begin{cases} 0, & x \le a_0 \\ f_1^*(x), & a_1 < x \le a_2 \\ f_2^*(x), & a_2 < x \le a_3 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ f_i^*(x), & a_i < x \le a_{i+1} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ 0, & x > a_l \end{cases} . \qquad (2.11)$$

Пример гистограммы для статистического ряда, содержащего четыре интервала, показан на рис. 2.3.

Гистограмма является статистическим аналогом кривой плотности распределения. Легко видеть, что площадь заштрихованной фигуры (площадь под гистограммой равна единице).

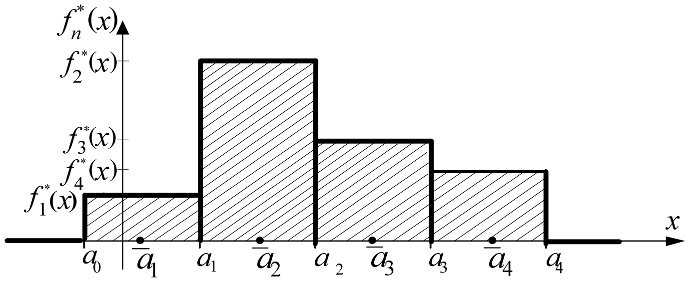

Рис. 2.3

Соединив точки гистограммы с абсциссами $\bar{a}_i = \dfrac{a_{i-1} + a_i}{2}$ при $i = 1, 2, ..., l$, на этом же рисунке можно построить *полигон частот.*

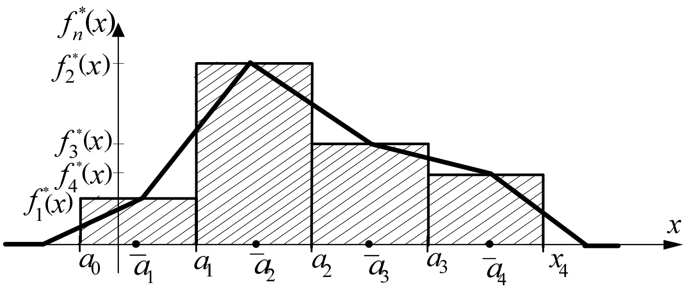

Рис. 2.4.

Полигон частот для гистограммы, показанной на рисунке 2.3, дан на рисунке 2.4.

ЗАМЕЧАНИЕ

По виду полигон частот больше «походит» на график плотности распределения, однако площадь под полигоном частот уже не равна единице.

По виду полигона выдвигается основная гипотеза H_0 о виде распределения генеральной случайной величины X: нормальное (гауссовское) распределение, показательное (экспоненциальное) распределение, равномерное распределение, и.т.д.

- Для кривой на рис. 2.4, основной гипотезой H_0 будет, очевидно, гипотеза о нормальном законе распределения.
- Для кривой на рис. 2.5, основной гипотезой H_0 будет, очевидно, гипотеза о показательном распределении генеральной совокупности, и.т.д.

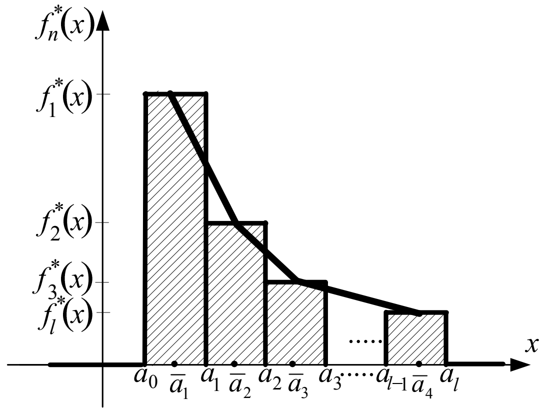

Рис. 2.5

2.2. Получение точечных статистических оценок

Если не известны параметры распределения генеральной случайной величины X, такие, например, как математическое ожидание и дисперсия, то при статистической обработке выборочных данных получают приближенные значения этих параметров, т.е. их *статистические оценки*. Статистические оценки разделяются на точечные статистические оценки и интервальные.

Точечной называется статистическая оценка, которая определяется одним числом. Основными точечными статистическими оценками являются: выборочное среднее $\bar{x}$, выборочная дисперсия s^2, выборочное среднее квадратическое отклонение (СКВО) s, выборочная медиана, выборочная асимметрия A^* и выборочный эксцесс E^*.

К точечным статистическим оценкам предъявляется ряд требований. Если θ^* – статистическая оценка параметра θ, то она должна удовлетворять следующим условиям:

1) быть несмещенной, что означает, что $M\left[\theta^*\right] = \theta$;

2) быть состоятельной, т.е. $\theta^* \underset{n\to\infty}{\overset{P.}{\to}} \theta$;

3) быть эффективной, т.е. дисперсия $D\left[\theta^*\right]$ – наименьшая или быть асимптотическая эффективной, что означает, что $\lim\limits_{n\to\infty} D\left[\theta^*\right]=0$.

Для вычисления точечных статистических оценок справедливы формулы (2.12) – (2.17), которые приведены ниже.

Выборочное среднее $\bar{x}$ вычисляется по формуле

$$\bar{x}=\sum_{i=1}^{l}\bar{a}_i p_i^* , \tag{2.12}$$

где $\bar{a}_i$ – середины интервалов, p_i^* – частоты.

ЗАМЕЧАНИЕ 1

Выборочное среднее является несмещенной, состоятельной и асимптотически эффективной оценкой математического ожидания генеральной совокупности.

Выборочная дисперсия s^2 и *выборочное СКВО* s вычисляются по формулам

$$s^2=\sum_{i=1}^{l}\bar{a}_i^2 p_i^* -\bar{x}^2,\ s=\sqrt{s^2} \tag{2.13}$$

где $\bar{a}_i$ – середины интервалов, p_i^* – частоты, $\bar{x}$ – выборочное среднее.

ЗАМЕЧАНИЕ 2

Выборочная дисперсия является состоятельной, но смещенной оценкой дисперсии генеральной совокупности. Несмещенной оценкой дисперсии является исправленная выборочная дисперсия.

Исправленная выборочная дисперсия $\bar{s}^2$ и *исправленное выборочное СКВО* $\bar{s}$ вычисляются по формулам

$$\bar{s}^2=\frac{n}{n-1}\cdot s^2,\ \bar{s}=\sqrt{\bar{s}^2} , \tag{2.14}$$

где s^2 – выборочная дисперсия, n – объем выборки.

Модальный интервал – это интервал статистического ряда с наибольшей частотой; *выборочная мода* – середина этого интервала.

Медианный интервал – интервал интервального ряда (табл. 2.2), которому соответствует накопленная частота, большая 0,5; *медиана* – середина этого интервала.

Для *выборочной асимметрии* A^* справедлива формула

$$A^*=\frac{1}{\bar{s}^3}\sum_{i=1}^{l}\left(\bar{a}_i-\bar{x}\right)^3 p_i^* , \tag{2.15}$$

Выборочный эксцесс E^* определяется формулой

$$E^*=\frac{1}{\bar{s}^4}\sum_{i=1}^{l}\left(\bar{a}_i-\bar{x}\right)^4 p_i^* -3, \tag{2.16}$$

где $\bar{a}_i$ – середины интервалов, $\bar{s}$ – *исправленное выборочное СКВО*, p_i^* – частоты, $\bar{x}$ – выборочное среднее.

ЗАМЕЧАНИЕ 3

Если $A^*=0$, то распределение имеет симметричную форму. При $A^*>0$ и $A^*<0$ говорят о положительной (правосторонней) или отрицательной (левосторонней) асимметрии. Эксцесс является показателем «крутости» полигона по сравнению с нормальным распределением.

ЗАМЕЧАНИЕ 4

Поскольку асимметрия и эксцесс являются характеристиками формы кривой распределения, то по величине выборочных асимметрии и эксцесса можно делать предположения о его виде. Если выборочные асимметрия и эксцесс достаточно малы, т.е. близки к нулю, то можно выдвигать гипотезу о нормальном законе распределения генеральной совокупности. Можно даже указать интервалы для этих выборочных характеристик, при попадании в которые они «не существенно» отклоняются от соответствующих характеристик нормального распределения. Для асимметрии: $\left(-\sigma_A;\sigma_A\right)$, где $\sigma_A=\sqrt{\dfrac{6(n-1)}{(n+1)(n+3)}}$. Для эксцесса: $\left(-\sigma_E;\sigma_E\right)$, где $\sigma_E=\sqrt{\dfrac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}$.

Выводы о виде распределения генеральной совокупности

В результате выполнения первой части работы следует сделать выводы о предполагаемом (гипотетическом) распределении генеральной случайной величины. Эти предположения делаются на основе:

– анализа гистограммы и полигона, т.е. сравнением их с кривыми плотностей известных распределений;

– величины выборочных асимметрии и эксцесса, которые в случае гипотетического нормального распределения должны быть малыми;

– сравнения *теоретической кривой*, т.е. графика плотности гипотетического (предполагаемого) распределения, в котором за параметры распределения приняты их статистические оценки, с построенным полигоном.

3. Проверка статистических гипотез

Если выдвинута статистическая гипотеза H_0 о законе распределения генеральной совокупности, то проверка ее истинности может проводиться на основе так называемого критерия согласия.

Смысл любого критерия согласия состоит в том, что рассматривается некоторая статистика (функция от выборочных данных)

$$\zeta_n=\zeta_n\left(x_1,x_2,\ldots,x_n\right),$$

закон распределения которой известен в предположении, что выдвинутая гипотеза верна. Если вычисленное по выборке значение принятой в данном критерии статистики попадает в *область допустимых значений* (область больших вероятностей), то гипотеза принимается, если это значение не попадет в область допустимых значений, т.е. попадает в *критическую область*, то гипотеза отвергается.

Например, пусть известна плотность распределения выбранной статистики $f_{\zeta_n}(x)$ и ее график показан на рисунке 3.1. Вся область под графиком $f_{\zeta_n}(x)$ разбивается на область больших вероятностей D и критическую область $\overline{D}$. Если выдвинутая гипотеза H_0 верна, то значения статистики ζ_n попадают в область больших вероятностей D, т.е. вероятность $P\left\{\zeta_n\in D/H_0\right\}$ велика, а вероятность $P\left\{\zeta_n\in\overline{D}/H_0\right\}$ мала.

Критическая область может быть двусторонней (рис. 3.1) и односторонней (в данном случае – правосторонняя) (рис. 3.2)

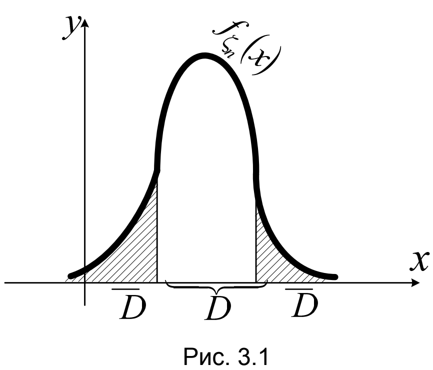

Рис. 3.1

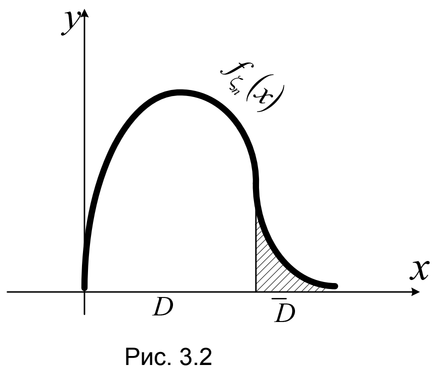

Рис. 3.2

3.1. Проверка гипотезы о законе распределения генеральной совокупности по критерию Пирсона

Если используется критерий Пирсона, то в качестве статистики ζ_n выбирается следующая функция выборочных данных

$$\chi^2_{набл} = \sum_{i=1}^{l} \frac{(n_i - n p_i)^2}{n p_i}, \tag{3.1}$$

в которой p_i – вероятности попадания генеральной случайной величины X в интервал $I_i = [a_{i-1}, a_i)$, l – количество интервалов статистического ряда, n_i – количества выборочных значений, попавших в интервал I_i.

Если основная гипотеза H_0 верна, то эта статистика $\chi^2_{набл}$ распределена асимптотически по закону χ^2 с m степенями свободы. Число степеней свободы определяется числом интервалов l и числом неизвестных параметров гипотетического распределения r по формуле

$$m = l - r - 1. \tag{3.2}$$

Ранее был построен интервальный ряд и вычислено количество выборочных значений, попавших в каждый интервал. Запишем полученные данные в таблицу 3.1.

Таблица 3.1

I_i	$[a_0;\, a_1)$	$[a_1;\, a_2)$	………	$[a_{l-1};\, a_l]$	$\sum$
n_i	n_1	n_2	………	n_l	n

Проверка гипотезы о законе распределения генеральной совокупности по критерию Пирсона проводится следующим образом.

1. По виду гистограммы и теоретической кривой выдвигается *основная гипотеза* H_0 о виде распределения генеральной совокупности X.

2. *Альтернативная гипотеза* H_1 заключается в том, что основная гипотеза H_0 не выполнена.

3. Прежде чем использовать критерий Пирсона выясняется, в каждом ли интервале ряда количество наблюдений больше пяти. Если в каком–то интервале это не так, то его объединяют с одним из соседних интервалов. При этом количество интервалов l уменьшается, а количества соответствующих выборочных значений n_i складываются.

4. Далее вычисляются p_i – вероятности попадания генеральной случайной величины X в интервалы $I_i = [a_{i-1}, a_i)$ по формулам:

$$p_i = P\{a_{i-1} \le X < a_i\} = F(a_i) - F(a_{i-1}),\ i = 1, 2, \ldots, l, \tag{3.3}$$

где $F(x)$ – гипотетическая функция распределения.

- Если основная гипотеза H_0 состоит в том, что случайная величина X распределена по нормальному закону, то вероятности p_i определяются по формулам:

  $$p_i = \Phi_0\left(\frac{a_i - \bar{x}}{\bar{s}}\right) - \Phi_0\left(\frac{a_{i-1} - \bar{x}}{\bar{s}}\right),\ i = 1, 2, \ldots, l, \tag{3.4}$$

  где $\bar{x}$ – среднее выборочное, $\bar{s}$ – исправленное выборочное СКВО, $\Phi_0(x)$ – функция Лапласа, значения которой находится из таблицы 8.2.

- Если основная гипотеза H_0 состоит в том, что случайная величина X распределена по показательному закону, то вероятности p_i определяются по формулам:

  $$p_i = e^{-\lambda a_{i-1}} - e^{-\lambda a_i},\ i = 1, 2, \ldots, l, \tag{3.5}$$

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

- Если основная гипотеза H_0 состоит в том, что случайная величина X распределена равномерно на отрезке $[a_0; a_l]$ (диапазон выборки), то вероятности p_i определяются по формулам:

$$p_i = \frac{1}{a_l - a_0}(a_i - a_{i-1}),\ i = 1, 2, ..., l. \quad (3.6)$$

5. Вычисляется статистика $\chi^2_{набл}$ по формуле (3.1). Статистика $\chi^2_{набл}$ используется для выяснения отклонения гипотетического распределения от генерального распределения. Она распределена асимптотически по закону χ^2 с m степенями свободы. Число степеней свободы при этом равно: $m = l_1 - r - 1$, где l_1 – новое число интервалов; r – число оцениваемых параметров.

- Если проверяется гипотеза о нормальном распределении, то оцениваемых параметров два: математическое ожидание a и среднее квадратическое отклонение σ.
- Если проверяется гипотеза о показательном распределении, то оцениваемый параметр один: λ – число обратное математическому ожиданию.
- Если проверяется гипотеза о равномерном на интервале (a, b) распределении, то оцениваемых параметров два: a и b, и.т.д.

Все вычисления вносятся в таблицу 3.2.

Таблица 3.2

№	$[a_{i-1}, a_i)$	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
1	$[a_0, a_1)$	n_1	p_1	np_1	$\frac{(n_1 - np_1)^2}{np_1}$
2	$[a_1, a_2)$	n_2	p_2	np_2	$\frac{(n_2 - np_2)^2}{np_2}$
…	…	…	…	…	…
l	$[a_{l-1}, a_l]$	n_l	p_l	np_l	$\frac{(n_l - np_l)^2}{np_l}$
$\sum$	—	$\sum = n$	$\sum = 1$	$\sum = n$	$\chi^2_{набл}$

Для использования критерия Пирсона задается число α, называемое *уровнем значимости* и равное вероятности отвергнуть истинную гипотезу H_0. Уровень значимости должен быть малым. Мы рекомендуем выбирать $\alpha = 0{,}01 - 0{,}05$.

Критерий Пирсона может использоваться с односторонней или двусторонней критической областью.

Критерий Пирсона с односторонней критической областью

По заданному уровню значимости α и числу степеней свободы m из таблицы распределения χ^2 (табл. 8.3) определяют границу (рис. 3.3) односторонней критической области $\chi^2_{кр} = \chi^2_{m,\alpha}$, так, что $\alpha = P\{\chi^2_{набл} \ge \chi^2_{кр}\}$.

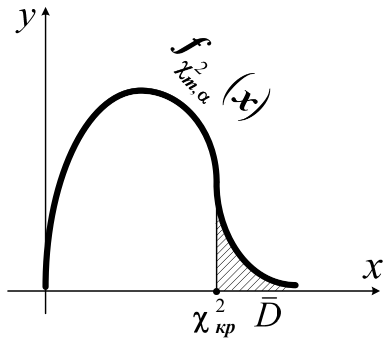

Рис. 3.3

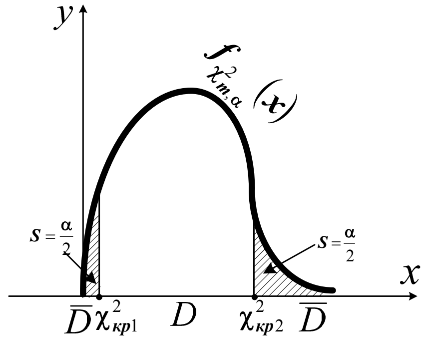

Рис. 3.4

Гипотеза о выбранном законе распределения случайной величины X принимается, если $\chi^2_{набл}$ попадает в область принятия гипотезы, т.е. при $\chi^2_{набл} < \chi^2_{кр}$.

Если $\chi^2_{набл}$ попадает в критическую область, т.е. $\chi^2_{набл} \geq \chi^2_{кр}$, то гипотеза отвергается и принимается альтернативная гипотеза: распределение генеральной случайной величины X не совпадает с гипотетическим распределением.

Критерий Пирсона с двусторонней критической областью

По заданному значению уровня значимости α и по вычисленному числу степеней свободы m из таблицы распределения χ^2 (табл. 8.3) определяют границы (рис. 3.4) двусторонней критической области

$$\chi^2_{кр1} = \chi^2_{m,1-\frac{\alpha}{2}} \text{ и } \chi^2_{кр2} = \chi^2_{m,\frac{\alpha}{2}},$$

так, что

$$P\left\{\chi^2_{набл} \geq \chi^2_{кр2}\right\} = \frac{\alpha}{2} \text{ и } P\left\{\chi^2_{набл} \leq \chi^2_{кр1}\right\} = 1 - P\left\{\chi^2_{набл} \geq \chi^2_{кр1}\right\} = 1 - \frac{\alpha}{2}.$$

Согласно критерию согласия Пирсона, гипотеза о выбранном теоретическом законе распределения случайной величины X принимается, если $\chi^2_{набл}$ попадает в область принятия гипотезы, т.е. при выполнении условия:

$$\chi^2_{кр1} < \chi^2_{набл} < \chi^2_{кр2}.$$

Если $\chi^2_{набл}$ попадает в критическую область, т.е. $\chi^2_{набл} \leq \chi^2_{кр1}$ или $\chi^2_{набл} \geq \chi^2_{кр2}$, то гипотеза отвергается и принимается альтернативная гипотеза: распределение генеральной случайной величины X не совпадает с гипотетическим распределением.

3.2. Проверка гипотезы о законе распределения генеральной совокупности по критерию Колмогорова

Пусть $F(x)$ – функция распределения генеральной случайной величины X, а $F_n^*(x)$ – эмпирическая функция распределения, построенная по выборке $x_1, x_2, ..., x_n$. Известно, что случайная величина

$$\rho_n = \sqrt{n} \max_x \left| F_n^*(x) - F(x) \right|$$

асимптотически распределена по закону Колмогорова, т.е.

$$F_{\rho_n}(x) = P\{\rho_n < x\} \underset{n\to\infty}{\to} K(x), \tag{3.7}$$

где $K(x)$ – функция распределения Колмогорова, значения которой можно найти в таблице 8.4.

Проверка гипотезы о законе распределения генеральной совокупности по критерию Колмогорова проводится следующим образом.

1. Выдвигается основная гипотеза H_0 о виде генерального распределения, т.е.

$$F(x) = F_0(x),$$

где $F_0(x)$ – гипотетическая функция распределения.

2. Выбирается статистика $\zeta_n = \sqrt{n}\max\limits_x \left|F_n^*(x) - F_0(x)\right|$, совпадающая с ρ_n, распределение которой удовлетворяет (3.7), в случае истинности гипотезы H_0.

3. Вычисляются значения статистики ζ_n в точках a_i $(i = 0, 1, ..., l)$ – границах интервального ряда. В подавляющем большинстве случаев $\max\left|F_n^*(x) - F_0(x)\right|$ достигается в этих точках (рис. 3.5).

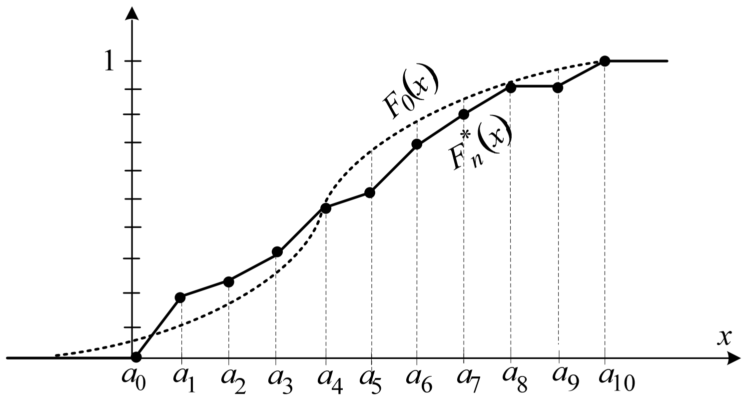

Рис. 3.5

При вычислении статистики ζ_n значения эмпирической функции распределения $F_n^*(a_i)$ определяют по формуле (2.9). Значения $F_0(a_i)$ вычисляются в зависимости от выдвинутой гипотезы H_0:

– если выдвинута гипотеза о нормальном распределении, то

$$F_0(a_i) = 0{,}5 + \Phi_0\left(\frac{a_i - \bar{x}}{\bar{s}}\right), \tag{3.8}$$

где $\bar{x}$ – выборочное среднее; $\bar{s}$ – исправленное выборочное СКВО; $\Phi_0(x)$ – функция Лапласа, значения которой берутся из таблицы 8.2.

– если выдвинута гипотеза о показательном распределении, то

$$F_0(a_i) = \begin{cases} 0, & a_i < a_0 \\ 1 - e^{-\lambda \cdot a_i}, & a_i \ge a_0 \end{cases}, \tag{3.9}$$

где параметр $\lambda = \dfrac{1}{\bar{x}}$, $\bar{x}$ – среднее выборочное.

– если выдвинута гипотеза о равномерном распределении, то

$$F_0(a_i) = \frac{a_i - a_0}{a_l - a_0}, \; a_0 \le a_i \le a_l. \tag{3.10}$$

где a_i – границы интервалов интервального ряда.

Все вычисления заносятся в таблицу 3.3. Выбирая в последнем столбце этой таблицы наибольшее значение $\left|F_n^*(a_i) - F_0(a_i)\right|$, легко вычислить статистику $\zeta_n = \sqrt{n}\cdot\max\limits_i\left|F_n^*(a_i) - F_0(a_i)\right|$.

Таблица 3.3.

a_i	$F_n^*(a_i)$	$F_0(a_i)$	$\left|F_n^*(a_i)-F_0(a_i)\right|$
a_0	$F_n^*(a_0)$	$F_0(a_0)$	$\left|F_n^*(a_0)-F_0(a_0)\right|$
a_1	$F_n^*(a_1)$	$F_0(a_1)$	$\left|F_n^*(a_1)-F_0(a_1)\right|$
………..	…………………	…………	…………..
a_l	$F_n^*(a_l)$	$F_0(a_l)$	$\left|F_n^*(a_l)-F_0(a_l)\right|$

4. На следующем шаге задается уровень значимости q. Критическая область $\overline{D}$ определяется условием

$$P\{\zeta_n \in \overline{D}/H_0\} = q,$$

и может быть односторонней: $\overline{D} = \left[\lambda_q; +\infty\right)$ (рис. 3.6) или двусторонней: $\overline{D} = \left(0; \lambda_{1-\frac{q}{2}}\right] \cup \left[\lambda_{\frac{q}{2}}; +\infty\right)$ (рис. 3.7).

В первом случае по заданному значению q определяется λ_q, такое, что $P\{\zeta_n \geq \lambda_q\} = q$ из таблицы распределения Колмогорова 8.4. Если значение статистики $\zeta_n \in \overline{D} = \left[\lambda_q; +\infty\right)$, т.е. $\zeta_n \geq \lambda_q$, то основная гипотеза отвергается. Если $\zeta_n < \lambda_q$, то гипотеза H_0 принимается, т.е. генеральное распределение считается совпадающим с гипотетическим.

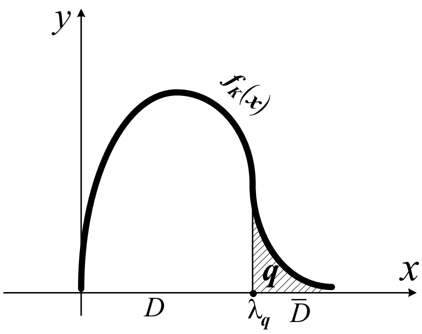

Рис. 3.6.

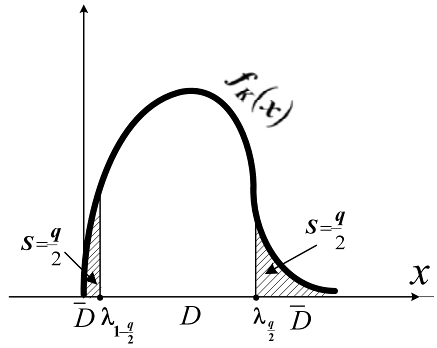

Рис. 3.7.

В случае двусторонней критической области ее границы $\lambda_{1-\frac{q}{2}}$ и $\lambda_{\frac{q}{2}}$ также выбираются из таблицы распределения Колмогорова – таблицы 8.4, в которой вместо значения q на входе задаются значения $1-\frac{q}{2}$ и $\frac{q}{2}$ соответственно, поскольку из рисунка 3.7. ясно, что

$$P\left\{\xi_n \geq \lambda_{\frac{q}{2}}\right\} = \frac{q}{2},\ P\left\{0 \leq \xi_n \leq \lambda_{1-\frac{q}{2}}\right\} = 1 - P\left\{\xi_n > \lambda_{1-\frac{q}{2}}\right\} = 1 - \left(1-\frac{q}{2}\right) = \frac{q}{2}.$$

Если значение статистики

$$\xi_n \in \overline{D},\ \text{т.е. } \xi_n \geq \lambda_{\frac{q}{2}} \text{ или } \xi_n \leq \lambda_{1-\frac{q}{2}},$$

то основная гипотеза отвергается. Если $\lambda_{1-\frac{q}{2}} < \xi_n < \lambda_{\frac{q}{2}}$, то гипотеза H_0 принимается, т.е. генеральное распределение совпадает с гипотетическим распределением.

3.3. Интервальное оценивание параметров распределения

Интервальное оценивание параметров распределения предполагает построение *доверительных интервалов* для этих параметров, т.е. построение интервалов, в которые с заданной *доверительной вероятностью* β попадают неизвестные параметры генерального распределения.

Если принята гипотеза о нормальном распределении генеральной случайной величины, то необходимо построить доверительные интервалы для двух параметров распределения: математического ожидания a и среднего квадратичного отклонении σ.

1. Доверительный интервал для математического ожидания a ***при известном среднем квадратическом отклонении*** σ

Интервал

$$\left(\bar{x}-\frac{\sigma}{\sqrt{n}}t_\beta;\,\bar{x}+\frac{\sigma}{\sqrt{n}}t_\beta\right), \tag{3.11}$$

в котором $\bar{x}$ – среднее выборочное, σ^2 – известная дисперсия, n – объем выборки, t_β определяется из таблицы 8.2 через функцию $\Phi_0^{-1}(x)$ – обратную функции Лапласа формулой

$$t_\beta=\Phi_0^{-1}\left(\frac{\beta}{2}\right), \tag{3.12}$$

покрывает с заданной доверительной вероятностью (надежностью) β неизвестный параметр a – математическое ожидание гауссовского (нормального) генерального распределения.

Доверительный интервал (3.11) иногда записывают в виде:

$$(\bar{x}-\varepsilon;\,\bar{x}+\varepsilon) \tag{3.13}$$

где величину

$$\varepsilon=\frac{\sigma}{\sqrt{n}}t_\beta \tag{3.14}$$

часто называют *точностью*.

2. Доверительный интервал для математического ожидания a ***при неизвестном среднем квадратическом отклонении*** σ

Если СКВО σ неизвестно, то оно заменяется исправленным выборочным СКВО $\bar{s}$. Тогда определенная формулой (3.14) точность будет иметь вид:

$$\varepsilon=\frac{\bar{s}}{\sqrt{n}}t_{n-1,\beta}=\frac{s}{\sqrt{n-1}}t_{n-1,\beta}, \tag{3.14$'$}$$

где s – СКВО, число $t_{n-1,\beta}$ определяют по заданной доверительной вероятности β и числу степеней свободы m из таблицы значений $t_{m,\beta}$ распределения Стьюдента, удовлетворяющих условию

$$\beta=P\{|\tau_m|<t_{m,\beta}\}.$$

Следовательно, интервал

$$\left(\bar{x}-\frac{s}{\sqrt{n-1}}t_{n-1,\beta};\,\bar{x}+\frac{s}{\sqrt{n-1}}t_{n-1,\beta}\right), \tag{3.15}$$

в котором $\bar{x}$ – среднее выборочное, s^2 – выборочная дисперсия, n – объем выборки, значения $t_{n-1,\beta}$ берутся из таблицы 8.5, покрывает с надежностью β неизвестный параметр a – математическое ожидание гауссовского (нормального) генерального распределения.

Чтобы построить доверительный интервал (3.15) необходимо:

1) задать доверительную вероятность (надежность) $\beta=0{,}92\div0{,}98$;
2) по заданной надежности β и числу степеней свободы $m=n-1$ из таблицы 8.5 найти $t_{n-1,\beta}$;
3) вычислить точность $\varepsilon=\dfrac{s}{\sqrt{n-1}}t_{n-1,\beta}$;
4) выписать доверительный интервал $(\bar{x}-\varepsilon;\,\bar{x}+\varepsilon)$ для параметра a.

3. Доверительный интервал для среднего квадратического отклонения σ

Чтобы получить доверительный интервал для СКВО σ нормального генерального распределения, вычисляют значения γ_1 и γ_2 с помощью таблицы распределения χ^2 (табл. 8.3) так, чтобы (рис. 3.8):

$$\gamma_2 = \chi^2_{n-1,\frac{1-\beta}{2}} \text{ или } P\left\{\chi^2_{n-1} \geq \gamma_2\right\} = \frac{1-\beta}{2}, \qquad (3.16)$$

$$\gamma_1 = \chi^2_{n-1,\frac{1+\beta}{2}} \text{ или } P\left\{\chi^2_{n-1} \leq \gamma_1\right\} = 1 - P\left\{\chi^2_{n-1} > \gamma_1\right\} = 1 - \frac{1-\beta}{2} = \frac{1+\beta}{2}. \qquad (3.17)$$

Это означает, что для определения γ_1 в таблице 8.3 нужно положить $\alpha = \dfrac{1+\beta}{2}$ и $m = n-1$, а для определения γ_2 в таблице 8.3 нужно положить $\alpha = \dfrac{1-\beta}{2}$ и $m = n-1$.

Тогда, при вычисленных γ_1 и γ_2, интервал

$$\left(s\sqrt{\frac{n-1}{\gamma_2}};\, s\sqrt{\frac{n-1}{\gamma_1}}\right), \qquad (3.18)$$

где s^2 – выборочная дисперсия, n – объем выборки, γ_1 и γ_2, покрывает с надежностью β параметр σ генерального нормального распределения.

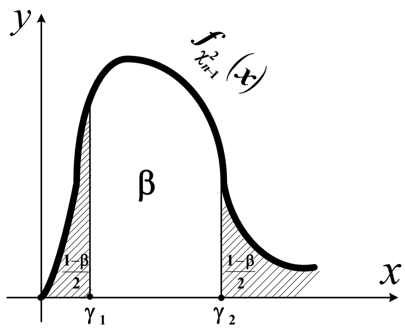

Рис.3.8

Если принята гипотеза о показательном распределении, то следует найти доверительные интервалы для неизвестных математического ожидания и среднего квадратичного отклонения генерального распределения по формулам (3.12) и (3.15), которые можно использовать, если выборка достаточно большого объема. Аналогично следует поступать, если основная гипотеза окажется отвергнутой, т.е. характер распределения останется неизвестным.

4. Примеры выполнения вычислений в курсовой работе

4.1. Вычисления на калькуляторе

Задание: по извлеченной случайной выборке

$0{,}70$; $-0{,}28$; $1{,}24$; $2{,}28$; $2{,}20$; $2{,}73$; $-1{,}18$; $0{,}77$; $2{,}10$; $-0{,}09$; $0{,}31$; $-0{,}69$; $-0{,}85$; $0{,}02$; $0{,}23$; $-1{,}12$; $0{,}43$; $0{,}60$; $1{,}13$; $0{,}63$; $0{,}67$; $0{,}63$; $2{,}34$; $0{,}91$; $0{,}81$; $0{,}49$; $2{,}97$; $1{,}66$; $3{,}38$; $0{,}35$; $2{,}66$; $-0{,}61$; $1{,}54$; $1{,}90$; $1{,}72$; $0{,}92$; $0{,}48$; $1{,}68$; $0{,}62$; $1{,}76$; $0{,}44$; $0{,}15$; $0{,}52$; $0{,}64$; $0{,}97$; $1{,}03$; $0{,}68$; $3{,}10$; $-0{,}74$; $0{,}26$.

генеральной непрерывной случайной величины X:

1) составить группированный (интервальный) ряд распределения;
2) построить эмпирическую функцию распределения, ее график и кумуляту;
3) вычислить эмпирические плотности распределения, построить гистограмму и полигон;
4) получить точечные статистические оценки параметров распределения;

5) построить теоретическую кривую и выдвинуть гипотезу о законе генерального распределения.

Первичная обработка статистических данных

1. Построение интервального статистического ряда

1) Упорядочим выборку, т.е. составим вариационный ряд:

$-1{,}18$; $-1{,}12$; $-0{,}85$; $-0{,}74$; $-0{,}69$; $-0{,}61$; $-0{,}28$; $-0{,}09$; $0{,}02$; $0{,}15$; $0{,}23$; $0{,}26$; $0{,}31$; $0{,}35$; $0{,}43$; $0{,}44$; $0{,}48$; $0{,}49$; $0{,}52$; $0{,}60$; $0{,}62$; $0{,}63$; $0{,}63$; $0{,}64$; $0{,}67$; $0{,}68$; $0{,}70$; $0{,}77$; $0{,}81$; $0{,}91$; $0{,}92$; $0{,}97$; $1{,}03$; $1{,}13$; $1{,}24$; $1{,}54$; $1{,}66$; $1{,}68$; $1{,}72$; $1{,}76$; $1{,}90$; $2{,}10$; $2{,}20$; $2{,}28$; $2{,}34$; $2{,}66$; $2{,}73$; $2{,}97$; $3{,}10$; $3{,}38$.

2) $x_{\min} = -1{,}18$; $x_{\max} = 3{,}38$. Объем выборки $n = 50$.

3) Диапазон выборки $[-1{,}18; 3{,}38]$.

4) Размах выборки $R_B = 3{,}38 + 1{,}18 = 4{,}56$.

5) Количество интервалов, вычисленное по формуле (2.3): $l = [1 + 3{,}32 \cdot \lg 50] = 6$.

6) Длина интервала, вычисленная по формуле (2.4): $h = \dfrac{4{,}56}{6} = 0{,}76$:

7) Границы интервалов статистического ряда (2.5):

$$a_0 = -1{,}18,\ a_1 = -0{,}42,\ a_2 = 0{,}34,\ a_3 = 1{,}10,\ a_4 = 1{,}86,\ a_5 = 2{,}62,\ a_6 = 3{,}38.$$

8) Интервалы:

$$I_0 = (-\infty; -1{,}18),\ I_1 = [-1{,}18;\ -0{,}42);\ I_2 = [-0{,}42; 0{,}34);\ I_3 = [0{,}34; 1{,}10);\ I_4 = [1{,}10; 1{,}86),$$
$$I_5 = [1{,}86; 2{,}62),\ I_6 = [2{,}62; 3{,}38].$$

9) Число выборочных значений, попавших в каждый интервал:

$$n_1 = 6,\ n_2 = 7,\ n_3 = 20,\ n_4 = 7,\ n_5 = 5,\ n_6 = 5.$$

Контроль: $\sum_{i=1}^{6} n_i = 6 + 7 + 20 + 7 + 5 + 5 = 50$.

10) Частоты интервального ряда, вычисленные по формуле (2.7):

$$p_1^* = \frac{6}{50} = 0{,}12,\ p_2^* = \frac{7}{50} = 0{,}14,\ p_3^* = \frac{20}{50} = 0{,}40,\ p_4^* = \frac{7}{50} = 0{,}14,\ p_5^* = \frac{5}{50} = 0{,}10,$$
$$p_6^* = \frac{5}{50} = 0{,}10.$$

Контроль: 0,12+0,14+0,4+0,14+0,1+0,1=1.

10) Середины интервалов, вычисленные по формуле (2.6):

$$\overline{a}_1 = \frac{-1{,}18 - 0{,}42}{2} = -0{,}80,\ \overline{a}_2 = \frac{-0{,}42 + 0{,}34}{2} = -0{,}04,\ \overline{a}_3 = \frac{0{,}34 + 1{,}10}{2} = 0{,}72,$$
$$\overline{a}_4 = \frac{1{,}10 + 1{,}86}{2} = 1{,}48,\ \overline{a}_5 = \frac{1{,}86 + 2{,}62}{2} = 2{,}24,\ \overline{a}_6 = \frac{2{,}62 + 3{,}38}{2} = 3{,}00.$$

11) Интервальный статистический ряд

Таблица 4.1

I_i	I_1	I_2	I_3	I_4	I_5	I_6	$\sum$
$\overline{a}_i$	–0,80	–0,04	0,72	1,48	2,24	7,00	–
p_i^*	0,12	0,14	0,40	0,14	0,10	0,10	

2. Построение эмпирической функции распределения

1) Накопленные частоты:

$$z_1 = p_1^* = 0{,}12,\ z_2 = p_1^* + p_2^* = 0{,}12 + 0{,}14 = 0{,}26,\ z_3 = p_1^* + p_2^* + p_3^* = 0{,}26 + 0{,}40 = 0{,}66,$$

$$z_4 = p_1^* + p_2^* + p_3^* + p_4^* = 0{,}66 + 0{,}14 = 0{,}80,\ z_5 = p_1^* + p_2^* + p_3^* + p_4^* + p_5^* = 0{,}80 + 0{,}10 = 0{,}90,$$

$$z_6 = p_1^* + p_2^* + p_3^* + p_4^* + p_5^* + p_6^* = 0{,}90 + 0{,}10 = 1{,}00.$$

2) Дискретный статистический ряд, в котором указаны середины интервалов $\overline{a}_i$, частоты p_i^* и накопленные частоты z_i, дан в таблице 4.2.

Таблица 4.2

$\overline{a}_i$	$-0{,}8$	$-0{,}04$	0,72	1,48	2,24	3,00	$\sum$
p_i^*	0,12	0,14	0,40	0,14	0,10	0,10	1
z_i	0,12	0,26	0,66	0,80	0,9	1	

4) Эмпирическая функция распределения $F_n^*(x)$, вычисленная по формуле (2.9) через накопленные частоты z_i имеет вид:

$$F_n^*(x) = \begin{cases} 0, & x \le -0{,}80 \\ 0{,}12, & -0{,}80 < x \le -0{,}04 \\ 0{,}26, & -0{,}04 < x \le 0{,}72 \\ 0{,}66, & 0{,}72 < x \le 1{,}48 \\ 0{,}80, & 1{,}48 < x \le 2{,}24 \\ 0{,}90, & 2{,}24 < x \le 3{,}00 \\ 1{,}00, & x > 3{,}00 \end{cases}.$$

5) График эмпирической функции распределения

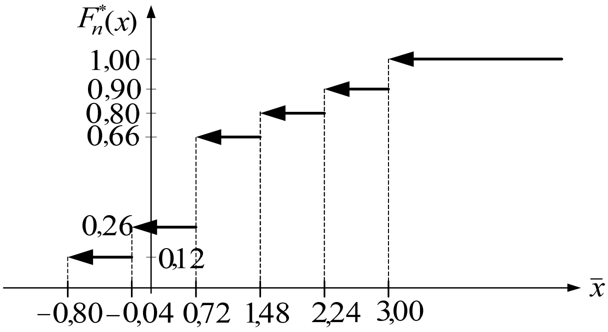

Рис. 4.1

6) Кумулята

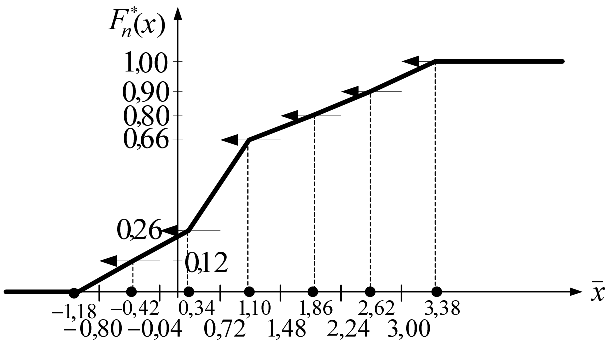

Рис. 4.2

3. Вычисление эмпирической плотности распределения. Построение гистограммы и полигона

1) Эмпирические плотности распределения, вычисленные по формулам (2.10), равны:

$$f_1^*(x)=\frac{0{,}12}{0{,}76}=0{,}16,\ f_2^*(x)=\frac{0{,}14}{0{,}76}=0{,}18,\ f_3^*(x)=\frac{0{,}40}{0{,}76}=0{,}53,\ f_4^*(x)=\frac{0{,}14}{0{,}76}=0{,}18,$$

$$f_5^*(x)=\frac{0{,}10}{0{,}76}=0{,}13,\ f_6^*(x)=\frac{0{,}10}{0{,}76}=0{,}13.$$

2) Аналитическое выражение эмпирической плотности распределения $f_n^*(x)$:

$$f_n^*(x)=\begin{cases}0, & x\le -1{,}18\\ 0{,}16, & -1{,}18<x\le -0{,}42\\ 0{,}18, & -0{,}42<x\le 0{,}34\\ 0{,}53, & 0{,}34<x\le 1{,}10\\ 0{,}18, & 1{,}10<x\le 1{,}86\\ 0{,}13, & 1{,}86<x\le 2{,}62\\ 0{,}13, & 2{,}62<x\le 3{,}38\\ 0, & x>3{,}38\end{cases}.$$

3) Приближенное эмпирическое распределения

Таблица 4.3

$\overline{a}_i$	$-0{,}8$	$-0{,}04$	0,72	1,48	2,24	3,00
$f_i^*(x)$	0,16	0,18	0,53	0,18	0,13	0,13

4) Гистограмма:

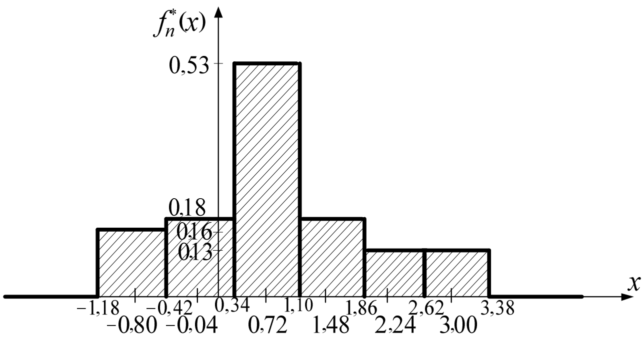

Рис. 4.3

5) Гистограмма и полигон:

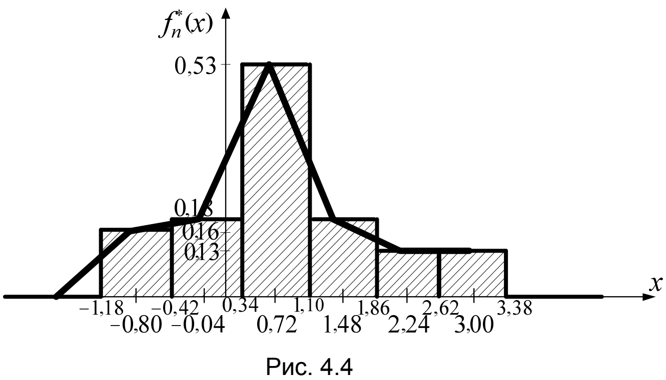

Рис. 4.4

Получение точечных статистических оценок

1) Выборочное среднее $\overline{x}$, вычисленное по формуле (2.12):

$$\overline{x}=-0{,}8\cdot 0{,}12-0{,}04\cdot 0{,}14+0{,}72\cdot 0{,}4+1{,}48\cdot 0{,}14+2{,}24\cdot 0{,}10+3\cdot 0{,}10=0{,}92.$$

2) Выборочная дисперсия s^2, вычисленная по формуле (2.13):

$$s^2=(-0{,}8)^2\cdot 0{,}12+(-0{,}04)^2\cdot 0{,}14+0{,}72^2\cdot 0{,}4+$$
$$+1{,}48^2\cdot 0{,}14+2{,}24^2\cdot 0{,}10+3^2\cdot 0{,}10-0{,}92^2=1{,}12.$$

3) Исправленная выборочная дисперсия $\bar{s}^2$ и исправленное выборочное СКВО $\bar{s}$ вычисленные по формуле (2.14):

$$\bar{s}^2 = \frac{50}{49}s^2 = \frac{50}{49}\cdot 1{,}12 = 1{,}14\,,\ \bar{s} = \sqrt{\bar{s}^2} = \sqrt{1{,}14} = 1{,}07\,.$$

4) Поскольку в таблице 4.2 накопленная частота $z_3 = 0{,}66 > 0{,}5$, то I_3 – медианный интервал, а выборочная медиана – $\bar{a}_3 = 0{,}72$. Этот же интервал является и модальным, поскольку ему соответствует наибольшая частота. Следовательно, выборочная мода равна выборочной медиане и равна: $\bar{a}_3 = 0{,}72$.

5) Выборочная асимметрия A^*, вычисленная по формуле (2.15):

$$A^* = \frac{1}{1{,}07^3}\Big[(-0{,}8-0{,}92)^3\cdot 0{,}12 + (-0{,}04-0{,}92)^3\cdot 0{,}14 +$$

$$+(0{,}72-0{,}92)^3\cdot 0{,}4 + (1{,}48-0{,}92)^3\cdot 0{,}14 + (2{,}24-0{,}92)^3\cdot 0{,}10 + (3-0{,}96)^3\cdot 0{,}10\Big] = 0{,}34\,.$$

6) Выборочный эксцесс E^*, вычисленный по формуле (2.16):

$$E^* = \frac{1}{1{,}07^4}\Big[(-0{,}8-0{,}92)^4\cdot 0{,}12 + (-0{,}04-0{,}92)^4\cdot 0{,}14 +$$

$$+(0{,}72-0{,}92)^4\cdot 0{,}4 + (1{,}48-0{,}92)^4\cdot 0{,}14 + (2{,}24-0{,}92)^4\cdot 0{,}10 + (3-0{,}92)^4\cdot 0{,}10\Big] - 3 = -0{,}23\,.$$

5. Предположение о характере генерального распределения

Поскольку для нормального распределения эксцесс и асимметрия равны нулю, то достаточно малые значения E^* и A^*, а также вид гистограммы и то, что мода и медиана совпадают и близки к среднему выборочному, позволяют выдвинуть гипотезу о нормальном распределении генеральной совокупности.

6. Построение теоретической кривой

Поскольку выдвинута гипотеза: генеральная совокупность распределена по нормальному закону, то в точках, являющихся серединами интервалов

$$\bar{a}_1 = -0{,}80\,,\ \bar{a}_2 = -0{,}04\,,\ \bar{a}_3 = 0{,}72\,,\ \bar{a}_4 = 1{,}48\,,\ \bar{a}_5 = 2{,}24\,,\ \bar{a}_6 = 3{,}00\,,$$

вычисляются значения плотностей нормального распределения с параметрами $a = \bar{x} = 0{,}92$ и $\sigma = \bar{s} = 1{,}07$ по формуле

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}} \quad \text{или} \quad f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right),$$

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

ЗАМЕЧАНИЕ

Значения функции $\varphi(x)$ можно найти в таблице 8.1. При использовании этой таблицы следует учитывать, что функция $\varphi(x)$ – четная. Поэтому ее значения даны только для положительных значений x.

Значения плотностей теоретического распределения:

$$f(\bar{a}_1) = 0{,}109\,,\ f(\bar{a}_2) = 0{,}252\,,\ f(\bar{a}_3) = 0{,}360\,,\ f(\bar{a}_4) = 0{,}317\,,\ f(\bar{a}_5) = 0{,}173\,,\ f(\bar{a}_6) = 0{,}058\,.$$

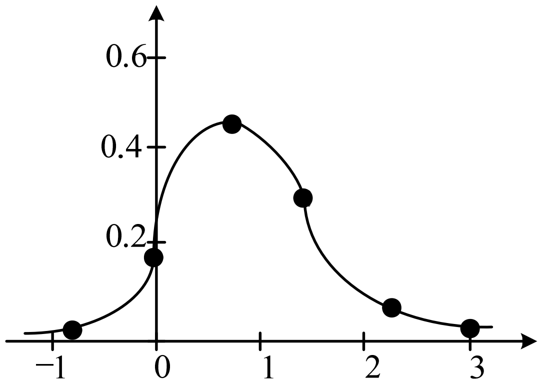

Рис. 4.5

Теоретическая кривая нормального распределения $f(x)$ построена на рисунке 4.5.

Выводы

Поскольку вычисленные значения асимметрии и эксцесса являются малыми, а также из вида гистограммы, полигона и полигона и теоретической кривой можно предположить, что генеральное распределение является нормальным.

Статистическая проверка истинности выдвинутой гипотезы о нормальном генеральном распределении

По виду гистограммы была выдвинута основная (нулевая) гипотеза H_0: генеральная совокупность X распределена по нормальному закону. Альтернативная гипотеза H_1: гипотеза H_0 не выполнена.

1. Проверка истинности гипотезы H_0 по критерию Пирсона

1) Во всех интервалах статистического ряда абсолютные частоты $n_i \geq 5$. Поэтому объединение интервалов проводить не нужно.

2) Вычислим по формуле (3.4) вероятности p_i – вероятности попадания генеральной случайной величины X в каждый из интервалов статистического ряда, используя функцию Лапласа. При этом учтем, что границы интервалов статистического ряда:

$$a_0 = -1{,}18,\ a_1 = -0{,}42,\ a_2 = 0{,}34,\ a_3 = 1{,}10,\ a_4 = 1{,}86,\ a_5 = 2{,}62,\ a_6 = 3{,}38,$$

а соответствующие значения функции Лапласа определяются из таблицы 8.2.

$$p_1 = P\{a_0 \leq \xi \leq a_1\} = \Phi_0\left(\frac{a_1 - \bar{x}}{\bar{s}}\right) - \Phi_0\left(\frac{a_0 - \bar{x}}{\bar{s}}\right) = \Phi_0\left(\frac{-0{,}42 - 0{,}92}{1{,}07}\right) - \Phi_0\left(\frac{-1{,}18 - 0{,}92}{1{,}07}\right) =$$

$$= -\Phi_0(1{,}25) - \Phi_0(1{,}96) = -0{,}39 + 0{,}47 = 0{,}08;$$

$$p_2 = \Phi_0\left(\frac{0{,}34 - 0{,}92}{1{,}07}\right) - \Phi_0\left(\frac{-0{,}42 - 0{,}92}{1{,}07}\right) = -\Phi_0(0{,}54) + \Phi_0(1{,}25) = -0{,}21 + 0{,}39 = 0{,}18;$$

$$p_3 = \Phi_0\left(\frac{1{,}1 - 0{,}92}{1{,}07}\right) - \Phi_0\left(\frac{0{,}34 - 0{,}92}{1{,}07}\right) = \Phi_0(0{,}17) + \Phi_0(0{,}54) = 0{,}07 + 0{,}21 = 0{,}28;$$

$$p_4 = \Phi_0\left(\frac{1{,}86 - 0{,}92}{1{,}07}\right) - \Phi_0\left(\frac{1{,}1 - 0{,}92}{1{,}07}\right) = \Phi_0(0{,}88) - \Phi_0(0{,}17) = 0{,}31 - 0{,}07 = 0{,}24;$$

$$p_5 = \Phi_0\left(\frac{2{,}62 - 0{,}92}{1{,}07}\right) - \Phi_0\left(\frac{1{,}86 - 0{,}92}{1{,}07}\right) = \Phi_0(1{,}59) - \Phi_0(0{,}88) = 0{,}44 - 0{,}31 = 0{,}13;$$

$$p_6 = \Phi_0\left(\frac{3{,}38 - 0{,}92}{1{,}07}\right) - \Phi_0\left(\frac{2{,}62 - 0{,}92}{1{,}07}\right) = \Phi_0(2{,}29) - \Phi_0(1{,}59) = 0{,}49 - 0{,}44 = 0{,}05.$$

Таблица 4.4

№	$[a_{i-1}, a_i)$	n_i	p_i	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-1{,}18; -0{,}42)$	6	0,08	1,00
2	$[-0{,}42; 0{,}34)$	7	0,18	0,57
3	$[0{,}34; 1{,}1)$	20	0,017	2,57
4	$[1{,}1; 1{,}86)$	7	0,24	3,57
5	$[1{,}86; 2{,}62)$	5	0,13	0,37
6	$[2{,}62; 2{,}38]$	5	0,57	2,5
$\sum$	–	50	–	10,58

3) Вычисляется статистика $\chi^2_{набл}$ по формуле (3.1). Все вычисления внесем в таблицу 4.4.

4) В последней строке и в последнем столбце таблицы 4.4 сумма $\sum_{i=1}^{l} \frac{(n_i - np_i)^2}{np_i}$, т.е. значение статистики $\chi^2_{набл}$. Следовательно, $\chi^2_{набл} = 10{,}58$.

5) Поскольку интервалов шесть, а оцениваемых параметров два (a и σ), то число степеней свободы m по формуле (3.2) равно:

$$m = 6 - 2 - 1 = 3.$$

6) По заданному уровня значимости $\alpha = 0{,}01$ и по вычисленному числу степеней свободы $m = 3$ из таблицы распределения χ^2 (табл. 8.3) определим границу критической области, которую выберем односторонней

$$\chi^2_{кр} = 11{,}34.$$

7) Гипотеза о нормальном законе распределения генеральной случайной величины ξ принимается, так как

$$\chi^2_{набл} = 10{,}58 \le \chi^2_{кр}.$$

2. Проверка истинности гипотезы H_0 по критерию Колмогорова

Согласно критерию Колмогорова вычисляется статистика

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

в граничных точках интервального ряда, т.е. в точках

$$a_0 = -1{,}18,\ a_1 = -0{,}42,\ a_2 = 0{,}34,\ a_3 = 1{,}10,\ a_4 = 1{,}86,\ a_5 = 2{,}62,\ a_6 = 3{,}38.$$

1) Эмпирическая функция распределения $F_n^*(x)$ была вычислена ранее, т.е.

$$F_n^*(x) = \begin{cases} 0, & x \le -0{,}80 \\ 0{,}12, & -0{,}80 < x \le -0{,}04 \\ 0{,}26, & -0{,}04 < x \le 0{,}72 \\ 0{,}66, & 0{,}72 < x \le 1{,}48 \\ 0{,}80, & 1{,}48 < x \le 2{,}24 \\ 0{,}90, & 2{,}24 < x \le 3{,}00 \\ 1{,}00, & x > 3{,}00 \end{cases}.$$

2) Значения $F_0(a_i)$ вычисляются с учетом того, что выдвинута гипотеза о нормальном распределении, т.е.

$$F_0(a_i) = 0{,}5 + \Phi_0\left(\frac{a_i - \bar{x}}{\bar{s}}\right).$$

Все вычисления заносятся в таблицу 4.5. Для удобства вычисления $F_0(a_i)$ введем в таблицу столбец значений $\frac{a_i - \bar{x}}{\bar{s}}$ с вычисленными $\bar{x} = 0{,}92$ и $\bar{s} = 1{,}07$.

3) Из последнего столбца таблицы ясно, что

$$\max_i \left| F_n^*(a_i) - F_0(a_i) \right| = 0{,}093.$$

Тогда значение статистики Колмогорова

$$\zeta_n = \sqrt{n} \cdot \max_i \left| F_n^*(a_i) - F_0(a_i) \right| = \sqrt{50} \cdot 0{,}093 = 0{,}658.$$

Таблица 4.5.

a_i	$F_n^*(a_i)$	$\frac{a_i - \bar{x}}{\bar{s}}$	$F_0(a_i)$	$\left\lvert F_n^*(a_i) - F_0(a_i) \right\rvert$
–1,18	0,00	–1,96	0,025	0,025
–0,42	0,12	–1,25	0,105	0,015
0,34	0,26	–0,54	0,295	0,035
1,10	0,66	0,17	0,567	0,093
1,86	0,80	0,88	0,810	0,010
2,62	0,90	1,59	0,944	0,044
3,38	1,00	2,30	0,989	0,011

4) Зададим уровень значимости $q = 0{,}01$. По заданному значению q определим λ_q, из таблицы распределения Колмогорова – таблицы 8.4.

5) Поскольку значение статистики $\zeta_n = 0{,}658$, меньше значения $\lambda_q = 1{,}627$, то основная гипотеза H_0 принимается, т.е. генеральное распределение считается нормальным.

3. Интервальное оценивание параметров распределения

Поскольку принята гипотеза о нормальном законе генерального распределения, то следует найти доверительные интервалы для параметров распределения, т.е. параметра a – математического ожидания и σ – среднего квадратического отклонения.

Доверительный интервал для параметра a

1) Пусть доверительная вероятность (надежность) $\beta = 0{,}99$.

2) По таблице распределения Стьюдента (табл. 8.5) определим $t_\beta = 2{,}405$ для $\beta = 0{,}99$ и числа степеней свободы $n - 1 = 49$.

3) Вычислим по формуле $(3.14')$ точность ε:

$$\varepsilon = t_\beta \cdot \frac{s}{\sqrt{n-1}} = 2{,}405 \cdot \frac{1{,}06}{\sqrt{49}} = 0{,}364.$$

4) Доверительный интервал для математического ожидания a нормального распределения, построенный по формуле (3.13) имеет вид:

$$(\bar{x} - \varepsilon;\ \bar{x} + \varepsilon) = (0{,}92 - 0{,}364;\ 0{,}92 + 0{,}364) = (0{,}538;\ 1{,}284).$$

Доверительный интервал для параметра σ

1) По таблице распределения χ_n^2 (табл. 8.3) для числа степеней свободы $n - 1 = 49$ и для значений

$$\alpha_1 = \frac{1+\beta}{2} = 0{,}995 \text{ и } \alpha_2 = \frac{1-\beta}{2} = 0{,}005$$

вычисляются:

$$\gamma_1 = \chi^2_{n-1,\alpha_1} = 27{,}249 \text{ и } \gamma_2 = \chi^2_{n-1,\alpha_2} = 78{,}231.$$

3) Построенный по формуле (3.18) доверительный интервал для параметра σ нормального распределения имеет вид:

$$\left(s\sqrt{\frac{n-1}{\gamma_2}};\, s\sqrt{\frac{n-1}{\gamma_1}}\right)=\left(1{,}06\cdot\sqrt{\frac{49}{78{,}231}};\, 1{,}06\cdot\sqrt{\frac{49}{27{,}249}}\right)=(0{,}839;\, 1{,}421).$$

Выводы

Выдвинутая основная гипотеза принимается, т.е. генеральная совокупность распределена по нормальному закону. Проверка выдвинутой гипотезы по критерию Пирсона и Колмогорова привели к одинаковому результату.

4.2. Вычисления в среде Mathcad

По извлеченной случайной выборке

$0{,}70$; $-0{,}28$; $1{,}24$; $2{,}28$; $2{,}20$; $2{,}73$; $-1{,}18$; $0{,}77$; $2{,}10$; $-0{,}09$; $0{,}31$; $-0{,}69$; $-0{,}85$; $0{,}02$; $0{,}23$; $-1{,}12$; $0{,}43$; $0{,}60$; $1{,}13$; $0{,}63$; $0{,}67$; $0{,}63$; $2{,}34$; $0{,}91$; $0{,}81$; $0{,}49$; $2{,}97$; $1{,}66$; $3{,}38$; $0{,}35$; $2{,}66$; $-0{,}61$; $1{,}54$; $1{,}90$; $1{,}72$; $0{,}92$; $0{,}48$; $1{,}68$; $0{,}62$; $1{,}76$; $0{,}44$; $0{,}15$; $0{,}52$; $0{,}64$; $0{,}97$; $1{,}03$; $0{,}68$; $3{,}10$; $-0{,}74$; $0{,}26$.

генеральной непрерывной случайной величины X провести первичную обработку экспериментальных данных. Для этого:

1) составить группированный (интервальный) ряд распределения;
2) построить эмпирическую функцию распределения, ее график и кумуляту;
3) вычислить эмпирические плотности распределения, построить гистограмму и полигон;
4) получить точечные статистические оценки параметров распределения;
5) построить теоретическую кривую;
6) на основе анализа гистограммы, вычисленных выборочных моментов и вида теоретической кривой выдвинуть предположение о характере генерального распределения.

Первичная обработка статистических данных

1. Задание выборочных значений, построение интервального ряда

Для этого нужно описать вектор выборки с помощью оператора присваивания (на панели ***Calculator***)

$$x := \blacksquare$$

и кликнуть мышью на первую кнопку панели ***Matrix***. Откроется окно, в котором нужно указать число строк и столбцов матрицы (в данном случае 1 столбец и 50 строк). В открывшийся шаблон для вектора ввести заданные выборочные значения.

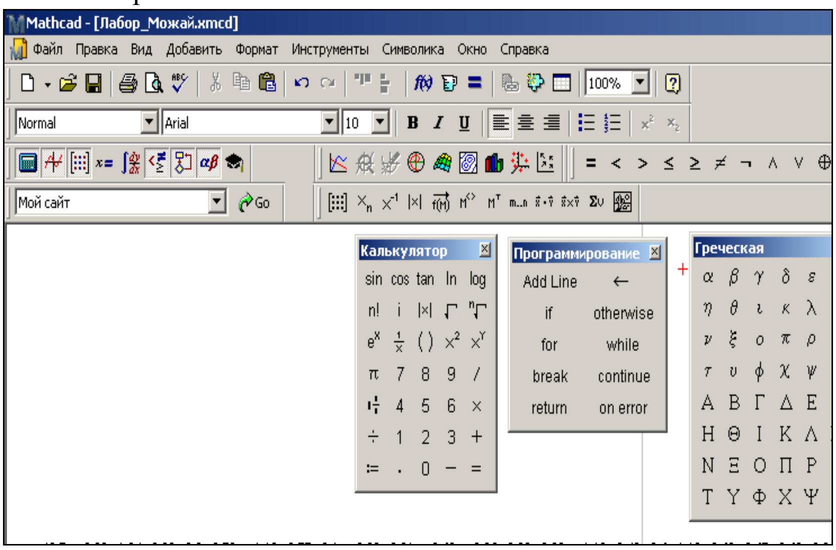

Рис. 4.6. Окно приложения Mathcad

Удобно вводить выборочные значения в виде одной строки, а затем провести транспонирование, кликнув мышью на иконку M^T панели ***Matrix*** (рис. 4.6). Увидеть набранный набор выборочных значений в виде вектора можно, если набрать с клавиатуры $\boxed{x =}$ На рисунке 4.7 помещен листинг окна приложения, где показана выборка, набранная в виде строки с последующим транспонированием и вектор выборочных значений x. Команда ORIGIN:=1 устанавливает во всем приложении нумерацию элементов векторов и матриц с единицы (в противном случае нумерация начиналась бы с нуля).

$x := (0.7\ \ -0.28\ \ 1.24\ \ 2.28\ \ 2.2\ \ 2.73\ \ldots\ 0.26\ \ 0.52\ \ 0.64)^T$

ORIGIN= 1

x=

1	0.7
2	-0.28
3	1.24
4	2.28
5	2.2
6	2.73
7	-1.18
8	0.77
9	2.1
10	-0.09
11	0.31
12	-0.69
13	-0.85
14	0.02
15	0.23
16	...

$n := 50 \quad l := \mathrm{trunc}(1 + 3.32 \cdot \log(50))$

$l = 6 \quad \min(x) = -1.18 \quad \max(x) = 3.38$

$f := \mathrm{histogram}(l, x) \quad h := \dfrac{\max(x) - \min(x)}{l}$

$h = 0.76 \quad p := \dfrac{f^{\langle 2 \rangle}}{n} \quad f1 := \dfrac{p}{h}$

$$f = \begin{pmatrix} -0.8 & 6 \\ -0.04 & 7 \\ 0.72 & 20 \\ 1.48 & 7 \\ 2.24 & 5 \\ 3 & 5 \end{pmatrix} \quad p = \begin{pmatrix} 0.12 \\ 0.14 \\ 0.4 \\ 0.14 \\ 0.1 \\ 0.1 \end{pmatrix} \quad f1 = \begin{pmatrix} 0.158 \\ 0.184 \\ 0.526 \\ 0.184 \\ 0.132 \\ 0.132 \end{pmatrix}$$

Рис. 4.7. Первичная обработка экспериментальных данных

В окне приложения (рис. 4.7) задаются и вычисляются:

- x – вектор выборочных значений.
- n – объем выборки;
- l – количество интервалов;
- min(x) – минимальное выборочное значение;
- max(x) –максимальное выборочное значение;
- h – длина интервала интервального ряда;
- f – матрица, первым столбцом которой являются середины интервалов интервального ряда, а вторым столбцом – количества выборочных значений в каждом интервале.

ЗАМЕЧАНИЕ

Обратите внимание, что переменная или вектор задаются с помощью оператора присваивания, а оператор $\boxed{=}$ возвращает значение заданной переменной.

При вычислении переменной l используется встроенная функция $\mathrm{trunc}(z)$ – целая часть числа z. При вычислении минимального и максимального элементов выборки используются функции $\min(x)$ и $\max(x)$, где x – вектор выборочных значений. Статистическая функция $\mathrm{histogramm}(l, x)$, где l – число интервалов, x – вектор выборочных значений, вычисляет середины интервалов статистического ряда (1 столбец) и количества выборочных значений для каждого интервала (2 столбец) интервального ряда.

2. Построение гистограммы и полигона

На листинге рисунка 4.7. вычислен вектор частот **p,** по формуле $p = \dfrac{n_i}{n}$, в которой количества выборочных значений для каждого интервала n_i получены выделением в матрице f второго столбца набором соответствующего индекса с панели матриц кнопкой $M^{\langle\ \rangle}$. Вектор fl – вектор эмпирических плотностей, который вычисляется по формуле $fl = \dfrac{p}{h}$.

На рис. 4.8 показано вычисление границ и середин интервалов. Вектор int – вектор середин интервалов, получается из матрицы f выделением в ней первого столбца набором соответствующего индекса с панели матриц кнопкой $M^{\langle\ \rangle}$. Вектор int1 – вектор границ интервалов, получаемых из середин интервалов сдвигом на $\frac{h}{2}$ влево. Последним элементом столбца int1 является максимальный элемент выборки.

$$int := f^{\langle 1 \rangle} \qquad int1 := int - \frac{h}{2} \qquad int1_7 := \max(x)$$

$$int = \begin{pmatrix} -0.8 \\ -0.04 \\ 0.72 \\ 1.48 \\ 2.24 \\ 3 \end{pmatrix} \qquad int1 = \begin{pmatrix} -1.18 \\ -0.42 \\ 0.34 \\ 1.1 \\ 1.86 \\ 2.62 \\ 3.38 \end{pmatrix}$$

Рис. 4.8. Вычисление середин и границ интервалов

ЗАМЕЧАНИЕ 1

Номер элемента вектора указывается в виде нижнего индекса $x_{\blacksquare}$, который вводится прямо с главной панели (соответствующая иконка проявляется в окне главной панели после набора идентификатора переменной).

ЗАМЕЧАНИЕ 2

Обратите внимание, что для обозначения переменных не используется переменная с индексом. Это обусловлено тем, что индексированная переменная в среде Mathcad используется для элементов массивов – матриц или векторов.

Далее строится гистограмма и полигон. Для этого открывается шаблон с панели ***Graph (Графики)*** (первая иконка панели рис. 4.6). В черный квадрат по горизонтали введем название вектора середин интервалов int, а в черный квадрат по вертикали – вектор эмпирических плотностей fl. Остальные местозаполнители предназначены для шаблона, их указывать не обязательно.

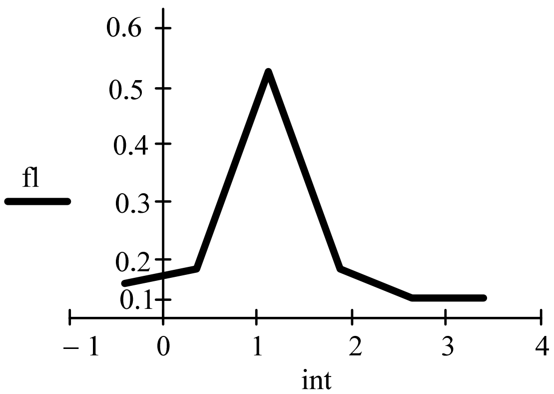

Рис. 4.9. Полигон

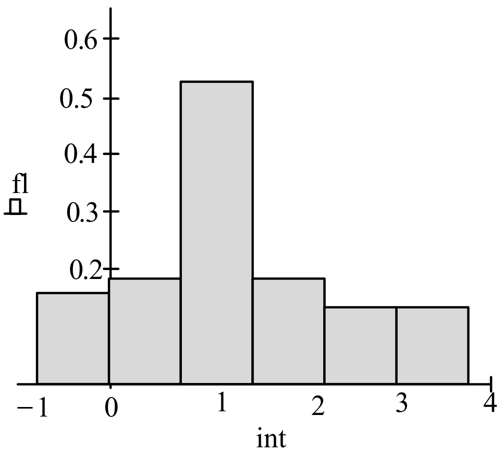

Рис. 4.10. Гистограмма

На рис. 4.9 показан построенный полигон. Чтобы построить гистограмму, следует этот график соответствующим образом отформатировать. Для этого нужно кликнуть на рисунок правой кнопкой мыши. В открывшемся окне кликнуть на кнопку **Формат.** Откроется панель форматирования графиков (рис. 4.6). На этой панели в окне **Графики** нужно указать в опции **Тип линии – Панель заливок.** Тогда тот же график, который показан на рис. 4.9 примет вид гистограммы (рис. 4.10).

Получение точечных характеристик. Построение теоретической кривой

Точечные характеристики (выборочные моменты) вычисляются с помощью встроенных функций:

- mean (x) – среднее выборочное $\bar{x}$;
- median (x) – выборочная медиана μ^*;
- Var (x) – исправленная выборочная дисперсия $\bar{s}^2$;
- Stdev (x) – исправленное выборочное среднее квадратичное отклонение;
- kurt (x) – выборочный эксцесс E^*;
- Skev (x) – выборочная асимметрия A^*;
- x – вектор выборочных значений.

Вычисление выборочных моментов дано на рис. 4.11.

Достаточно малые величины коэффициентов эксцесса и асимметрии, а также построенные гистограмма и полигон позволяют предположить, что генеральное распределение является нормальным. Поэтому в серединах интервального ряда вычислены значения плотности нормального распределения (вектор f) через встроенную функцию $\mathrm{dnorm}(\mathrm{int}, a, \sigma)$, где в качестве параметров a и σ взяты вычисленные среднее выборочное и выборочное среднее квадратичное отклонение.

$\mathrm{mean}(x) = 0.902$

$\mathrm{median}(x) = 0.675$ $\quad\quad$ $\mathrm{kurt}(x) = -0.23$

$\mathrm{Var}(x) = 1.198$

$\mathrm{Stdev}(x) = 1.094$ $\quad\quad$ $\mathrm{skew}(x) = 0.313$

$$fp := \mathrm{dnorm}(\mathrm{int}, 0.902, 1.094)$$

$$fp = \begin{pmatrix} 0.109 \\ 0.252 \\ 0.36 \\ 0.317 \\ 0.173 \\ 0.058 \end{pmatrix}$$

Рис. 4.11. Вычисление точечных характеристик и вычисление плотностей гипотетического распределения

На одном рисунке 4.12 построены гистограмма и теоретическая кривая – плотность нормального распределения. При построении графиков двух функций в одном шаблоне нужно в квадрат по вертикали внести название одной функции (fl), нажать на клавиатуре запятую, а затем ввести название второй функции (fp).

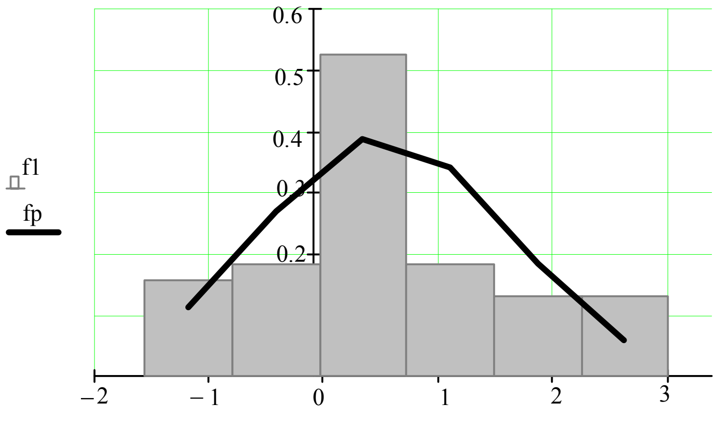

Рис. 4.12. Гистограмма и теоретическая кривая

Построение эмпирической функции распределения

$$kum_1 := 0 \qquad i := 1..6 \qquad kum_{i+1} := kum_i + p_i$$

$$kum = \begin{pmatrix} 0 \\ 0.12 \\ 0.26 \\ 0.66 \\ 0.8 \\ 0.9 \\ 1 \end{pmatrix} \qquad femp(z) := \left| \begin{array}{l} 0 \ \text{ if } \ z \le \min(x) \\ kum_1 \ \text{ if } \ \min(x) < z \le int_1 \\ kum_2 \ \text{ if } \ int_1 < z \le int_2 \\ kum_3 \ \text{ if } \ int_2 < z \le int_3 \\ kum_4 \ \text{ if } \ int_3 < z \le int_4 \\ kum_5 \ \text{ if } \ int_4 < z \le int_5 \\ kum_6 \ \text{ if } \ int_5 < z \le int_6 \\ kum_7 \ \text{ if } \ z > int1_6 \end{array} \right.$$

Рис. 4.13. Вычисление эмпирической функции распределения

Для построения кумуляты формируется вектор kum – вектор накопленных частот. Формирование этого вектора показано на листинге рисунка 4.13. Элементы вектора kum заданы как *ранжированная* переменная, значения которой вычисляются при i от 1 до 6 шагом 1. Следует учитывать, что в операторе $i := 1..6$ многоточие набирается на клавиатуре английского регистра – русская буква **ж.** Графическое представление этого вектора в виде кусочно–линейной кривой, соединяющей точки $(\mathrm{int1}_i, \mathrm{kum}_i)$ дано на рисунке 4.14.

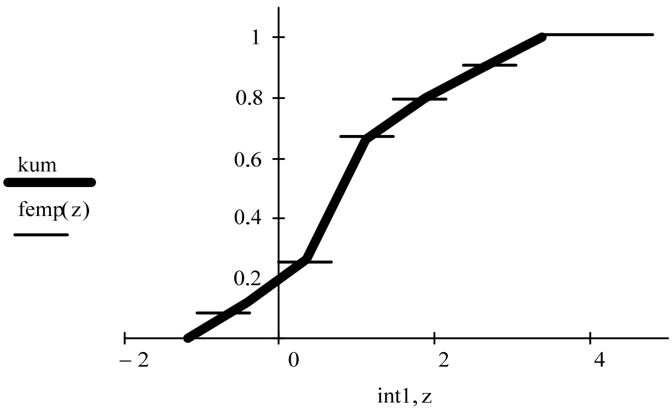

Рис. 4.14. Кумулята и эмпирическая функция распределения

Построение эмпирической функции распределения $F_n^*(x)$ проводится через функцию $\mathrm{femp}(z)$, задание которой показано на листинге рисунка 4.13 с помощью условных операторах присваивания – кнопки $\mathrm{Add\,Line}$ и if на панели ***Программирование (Programming)***, в которых логические операторы $\le$ и $>$ набираются на шкале ***Логические операторы (Boolean.***

График функции $\mathrm{femp}(z)$ показан на рисунке 4.14 в виде кусочно–постоянной линии с разрывами в серединах интервалов. Это достигается указанием в окне **Формат** в опции **Тип линии – линии точек.**

Выводы:

На основе анализа гистограммы и теоретической кривой, а также вследствие малости выборочных эксцесса и асимметрии можно выдвинуть гипотезу о нормальном распределении генеральной случайной величины.

Статистическая проверка истинности выдвинутой нулевой гипотезы

Поскольку по виду гистограммы было выдвинуто предположение о нормальном распределении генеральной совокупности, то это предположение – основная выдвинутая гипотеза H_0. Конкурирующая гипотеза: генеральное распределении не является нормальным.

1. Проверка истинности гипотезы H_0 по критерию Пирсона

1) Вычисляются вероятности p_i – вероятности попадания генеральной случайной величины X в каждый из интервалов статистического ряда, используя функцию Лапласа, т.е.

$$p_i = P\{a_{i+1} \le \xi \le a_i\} = \Phi_0\left(\frac{a_{i+1} - \bar{x}}{\bar{s}}\right) - \Phi_0\left(\frac{a_i - \bar{x}}{\bar{s}}\right). \quad (4.1)$$

На рисунке 4.15 – окно приложения **Mathcad,** где формируется вектор pt – вектор вероятностей попадания в интервалы статистического ряда, заданные вектором int1. Эти вероятности подсчитываются по формуле (4.1) с помощью функции распределения нормального закона $\mathrm{pnorm}(\mathrm{int1}, a, \sigma)$, в которой в качестве параметров распределения взяты их точечные оценки, т.е. $a = \bar{x} = 0{,}902$, $\sigma = \bar{s} = 1{,}094$. На этом же листинге вычислен вектор pt1 – вектор значений $p_i \cdot n$.

$$i := 1..6$$

$$pt_i := \mathrm{pnorm}(intl_{i+1}, 0.902, 1.094) - \mathrm{pnorm}(intl_i, 0.902, 1.094)$$

$$pt1 := pt \cdot n$$

$$pt = \begin{pmatrix} 0.085 \\ 0.19 \\ 0.268 \\ 0.238 \\ 0.132 \\ 0.046 \end{pmatrix} \qquad pt1 = \begin{pmatrix} 4.247 \\ 9.514 \\ 13.404 \\ 11.879 \\ 6.622 \\ 2.32 \end{pmatrix}$$

Рис. 4.15

2) Вычисляется статистика $\chi^2_{набл}$ по формуле

$$\chi^2_{набл} = \sum_{i=1}^{l} \frac{(n_i - np_i)^2}{np_i}, \quad (4.2)$$

Для этого выделяется второй столбец матрицы f (рис. 4.7) – вектор pn – вектор количеств выборочных значений n_i в каждом интервале статистического ряда. Вектор pt1 – вектор значений $p_i \cdot n$. Вектор ptx – вектор квадратичных невязок – слагаемых в сумме для статистики $\chi^2_{набл}$, вычисляемой по формуле (4.2).

Формирование векторов pn и ptx показаны на листинге рисунка 4.16. Статистика $\chi^2_{набл}$ критерия Пирсона вычислена оператором суммирования $\sum$ элементов вектора ptx, который находится на панели ***Matrix.***

$$pn := f^{\langle 2 \rangle} \qquad ptx := \frac{(pn - pt1)^2}{pt1}$$

$$pn = \begin{pmatrix} 6 \\ 7 \\ 20 \\ 7 \\ 5 \\ 5 \end{pmatrix} \qquad ptx = \begin{pmatrix} 0.724 \\ 0.664 \\ 3.246 \\ 2.004 \\ 0.397 \\ 3.094 \end{pmatrix}$$

Наблюдаемое значение статистики критерия Пирсона

$$\sum ptx = 10.13$$

Граница критической области

$$\mathrm{qchisq}(0.99, 3) = 11.345$$

Рис. 4.16. Вычисление статистики критерия Пирсона и границы критической области

На этом же рисунке вычисляется граница критической области $\chi^2_{кр}$ через встроенную функцию $\mathrm{qchisq}(\beta, m)$, где β – доверительная вероятность, m – число степеней свободы. Доверительная вероятность выбрана равной $0{,}99$, соответствующей уровню значимости $\alpha = 0{,}01$, а число степеней свободы $m = l - 1 - r = 6 - 1 - 2 = 3$.

Поскольку $\chi^2_{набл} = 10{,}13$ меньше $\chi^2_{кр} = 11{,}345$, то основная гипотеза о нормальном распределении генеральной случайной величины принимается.

2. Проверка истинности гипотезы H_0 по критерию Колмогорова

Вычисляется статистика критерия Колмогорова

$$\xi_K = \sqrt{n}\max_x \left| F_n^*(x) - F_0(x) \right| \tag{4.3}$$

в граничных точках интервального ряда – вектор $\mathrm{int1}$, вычисленный ранее. Эмпирическая функция распределения $\mathrm{femp}(z)$ также построена (рис. 4.13). На рисунке 4.17 показано вычисление статистики Колмогорова – ξ_K. Для этого в точках, являющихся границами интервалов ($\mathrm{int1}$), вычислены значения эмпирической функции распределения – вектор pv и значения функции распределения нормального закона – вектор po, через встроенную функцию $\mathrm{pnorm}(x, a, \sigma)$. Затем вычислен вектор pm – вектор модулей разностей значений функций распределения – эмпирической $F_n^*(x)$ и гипотетической $F_0(x)$, который затем упорядочен (вектор pp) по возрастанию через функцию $\mathrm{sort}(x)$. Поскольку вектор pp упорядочен по возрастанию, то ясно, что $\max\limits_x \left| F_n^*(x) - F_0(x) \right| = pp_7$. На основании этого по формуле (4.3) вычислена статистика Колмогорова.

$i := 1..7$

$\mathrm{femp}(\mathrm{int1}_i) =$

0
0.12
0.26
0.66
0.8
0.9
1

$\mathrm{po}_i := \mathrm{pnorm}(\mathrm{int1}_i, 0.902, 1.094)$ $\quad \mathrm{pv}_i := \mathrm{femp}(\mathrm{int1}_i)$ $\quad \mathrm{pm}_i := \left| \mathrm{pv}_i - \mathrm{po}_i \right|$

$$\mathrm{po} = \begin{pmatrix} 0.029 \\ 0.113 \\ 0.304 \\ 0.572 \\ 0.809 \\ 0.942 \\ 0.988 \end{pmatrix} \qquad \mathrm{pp} := \mathrm{sort}(\mathrm{pm}) \qquad \mathrm{pp} = \begin{pmatrix} 6.555\times 10^{-3} \\ 9.4\times 10^{-3} \\ 0.012 \\ 0.029 \\ 0.042 \\ 0.044 \\ 0.088 \end{pmatrix}$$

$\xi_K := \sqrt{n}\cdot \mathrm{pp}_7 \qquad \xi_K = 0.624$

Рис. 4.17. Вычисление статистики Колмогорова

Для заданного уровня значимости $q = 0{,}01$ вычисляется граница критической области (табл. 8.4) $\lambda_q = 1{,}627$. Поскольку $\xi_K = 0{,}624 < \lambda_q = 1{,}627$, то гипотеза о нормальном распределении генеральной совокупности принимается.

Интервальные оценки параметров распределения

1. Доверительный интервал для математического ожидания

Поскольку принята гипотеза о нормальном генеральном распределении, то для параметров этого распределения a и σ, являющихся математическим ожиданием и средним квадратичным отклонением, справедливы формулы (3.15) и (3.18).

На рисунке 4.18 – листинг вычисления доверительного интервала для математического ожидания. По заданной надежности $\beta = 0{,}99$ и по числу степеней свободы $n - 1$, где $n = 50$ (объем выборки), вычисляется граница критической области $\mathrm{t}\beta$ – квантиль уровня 0,99 распределения Стьюдента через встроенную функцию $\mathrm{qt}(\beta, n-1)$.

Доверительный интервал, вычисленный по формуле (3.15) имеет вид:

$$\left(\bar{x} - t\beta \cdot \frac{s}{\sqrt{n-1}};\, \bar{x} + t\beta \cdot \frac{s}{\sqrt{n-1}}\right),$$

где $\bar{x}$ и s – среднее выборочное и выборочное среднее квадратичное отклонение. Тогда его границы

$$a1 = \bar{x} - t\beta \cdot \frac{s}{\sqrt{n-1}}, \qquad a2 = \bar{x} + t\beta \cdot \frac{s}{\sqrt{n-1}}.$$

Доверительный интервал для математического ожидания

Надежность $\beta := 0.99$

Граница критической области $t\beta := qt(\beta, n-1)$ $t\beta = 2.405$

Точность $eps := t\beta \cdot \frac{stdev(x)}{\sqrt{n-1}}$ $eps = 0.372$

Границы доверительного интервала

$a1 := mean(x) - eps$ $a2 := mean(x) + eps$

$a1 = 0.53$ $a2 = 1.274$

Рис.4.18. Вычисление доверительного интервала для математического ожидания

2. Доверительный интервал для среднего квадратичного отклонения

Доверительный интервал для среднего квадратичного отклонения

Надежность $\beta := 0.99$

Двусторонняя критическая область

$\alpha 1 := \frac{1+\beta}{2}$ $\alpha 2 := \frac{1-\beta}{2}$ $\alpha 1 = 0.995$ $\alpha 2 = 5 \times 10^{-3}$

Критические точки распределения χ^2

$\gamma 1 := qchisq(1 - \alpha 1, n - 1)$ $\gamma 2 := qchisq(1 - \alpha 2, n - 1)$

$\gamma 1 = 27.249$ $\gamma 2 = 78.231$

Границы доверительного интервала

$\sigma 1 := stdev(x) \cdot \sqrt{\frac{n-1}{\gamma 2}}$ $\sigma 2 := stdev(x) \cdot \sqrt{\frac{n-1}{\gamma 1}}$

$\sigma 1 = 0.857$ $\sigma 2 = 1.453$

Рис. 4.19. Вычисление доверительного интервала для среднего квадратичного отклонения

На рисунке 4.19 – листинг вычисления доверительного интервала для среднего квадратичного отклонения. По заданной доверительной вероятности $\beta = 0{,}99$ вычисляются критические точки распределения χ^2 – границы двусторонней критической области γ_1 и γ_2. Для вычисления границ критической области используется встроенная функция $\mathrm{qchisq}(p, n-1)$, где p – вероятность попадания в соответствую часть критической области ($p = 1 - \alpha 1$ и $p = 1 - \alpha 2$), $n-1$ – число степеней свободы, n – объем выборки.

Границы доверительного интервала вычисляются по формуле (3.18)

$$\left(s\sqrt{\frac{n-1}{\gamma_2}};\, s\sqrt{\frac{n-1}{\gamma_1}}\right),$$

где γ_1 и γ_2 – границы двусторонней критической области, s – выборочное среднее квадратическое отклонение.

Выводы

На основе критериев Пирсона и Колмогорова, которые дали аналогичные результаты, гипотезу о нормальном распределении генеральной случайной величины следует принять.

6. Варианты расчетных заданий для курсовой работы

Вариант 1

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

Вариант 2

14,8	1,8	16,8	15,1	3,2	–4,6	10,5	10,6	16,0	11,2
6,9	7,0	14,2	15,7	9,5	2,7	0,4	9,4	1,0	15,8
14,4	6,4	1,3	11,8	6,4	11,9	21,7	1,2	6,2	1,6
1,9	3,5	4,3	0,3	–2,2	7,8	-0,9	15,4	5,3	15,6
5,2	14,3	11,3	12,0	7,6	4,7	12,3	4,0	8,2	12,3

Вариант 3

10,4	5,5	9,9	3,9	0,5	7,0	18,5	10,8	9,9	14,9
3,0	8,5	6,9	2,0	22,1	0,3	6,5	12,1	–1,8	3,8
6,6	4,3	6,9	12,2	11,5	5,5	6,0	14,7	4,1	27,3
16,4	9,7	21,9	9,6	10,7	9,6	3,8	–1,2	13,6	1,0
3,0	6,5	19,4	12,2	6,5	7,6	11,3	12,3	7,5	13,9

Вариант 4

3,3	9,9	7,3	18,6	8,9	9,8	12,6	1,6	5,8	10,9
10,0	18,8	-10,8	11,7	11,1	13,5	6,5	2,3	11,5	14,6
8,0	9,1	6,1	7,7	6,1	6,6	9,1	8,1	4,4	1,9
11,7	9,0	4,2	7,2	–2,2	5,9	13,6	5,3	3,9	8,0
3,8	5,9	23,5	8,3	16,3	11,2	4,2	12,0	8,5	12,1

Вариант 5

13,4	6,0	5,4	12,5	6,3	6,7	0,4	-1,0	11,2	19,3
14,9	13,4	1,3	18,1	0,5	7,7	6,0	10,2	8,3	11,6
5,9	14,2	2,3	6,9	17,8	3,5	2,2	8,4	14,5	4,8
3,1	10,9	7,6	6,6	5,1	–0,7	–9,8	4,1	17,5	4,2
7,3	0,8	14,9	9,7	1,6	7,0	–4,2	–9,2	–4,5	–5,0

Вариант 6

7,2	5,8	0,9	–5,8	7,1	6,6	8,7	–0,3	12,2	15,8
7,7	10,6	15,7	9,8	5,4	13,3	6,1	10,1	8,5	10,1
16,9	13,1	13,1	12,6	16,8	11,2	13,6	5,1	2,8	4,5
–1,9	9,5	–0,9	12,3	7,8	17,9	21,6	0,0	0,7	9,5
5,6	7,1	15,3	11,6	16,0	–3,4	17,1	–1,0	10,3	15,7

Вариант 7

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

Вариант 8

5,3	5,6	–1,4	2,7	10,7	–3,9	12,6	18,1	–0,7	6,1
2,1	2,8	–1,7	12,5	11,6	19,2	–1,1	6,6	11,4	9,6
8,6	4,4	13,0	10,2	10,8	16,7	6,2	–0,2	3,2	4,0
19,9	7,4	–3,7	14,0	10,5	7,0	12,6	1,7	4,1	8,8
6,9	–0,2	1,0	14,7	14,7	3,9	2,8	22,9	7,9	5,2

Вариант 9

9,0	–5,5	3,6	–0,8	–0,1	14,3	5,7	3,9	9,1	8,4
7,1	1,3	17,4	12,6	4,4	15,7	12,7	–1,6	5,5	–1,1
–5,4	11,2	15,2	9,3	4,0	15,2	11,9	6,4	12,9	15,7
–0,5	6,4	5,5	25,1	8,0	2,0	6,4	6,6	9,9	7,0
8,0	11,4	9,2	10,8	2,1	3,3	8,6	10,8	11,2	7,5

Вариант 10

12,5	11,5	2,3	–7,1	2,1	8,0	3,0	–1,9	16,0	14,0
14,7	12,2	5,3	16,9	9,8	3,9	3,8	5,2	4,5	1,0
21,4	22,1	6,8	9,7	11,2	9,7	4,5	12,8	3,4	–5,1
15,5	6,3	4,4	10,5	6,3	–1,0	6,0	2,2	19,8	7,8
–0,7	9,6	11,1	3,9	18,5	11,1	7,3	15,4	0,0	9,5

Вариант 11

4,5	–0,6	9,6	10,5	15,2	9,3	4,5	9,2	11,9	17,1
4,6	18,3	12,6	4,7	12,9	13,1	14,4	26,3	7,6	7,5
6,8	12,2	10,4	2,6	7,7	12,5	7,2	17,9	11,3	10,3
11,9	8,6	15,6	–0,5	11,1	3,0	9,7	–1,1	12,0	13,0
4,1	13,1	9,3	17,8	6,5	14,3	3,6	17,6	9,3	13,3

Вариант 12

1,01	0,98	1,10	0,07	0,96	0,28	0,06	0,048	1,11	0,33
0,03	0,32	0,80	0,62	0,13	1,56	0,31	0,954	0,26	0,30
0,48	0,54	0,31	0,03	0,55	0,99	0,33	0.274	1,77	0,79
0,06	0,19	0,55	0,06	0,56	0,61	0,17	0,178	0,95	0,23
0,59	1,30	0,94	0,25	0,22	1,49	2,11	0,694	1,14	1,58.

Вариант 13

3,0	12,6	18,3	12,6	8,1	6,1	1,7	8,2	7,1	16,0
4,6	12,3	3,1	9,8	8,2	14,3	22,8	15,8	4,3	2,4
–0,7	1,4	5,6	–1,0	12,0	18,9	10,5	10,6	6,7	3,1
3,4	1,4	9,2	18,5	5,7	0,4	–2,0	2,4	10,5	14,2
–0,2	3,8	14,9	–4,2	14,5	14,0	17,5	5,7	16,1	10,0

Вариант 14

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

Вариант 15

7,3	2,1	22,1	3,2	4,3	17,9	13,6	7,8	–7,0	6,1
10,7	7,8	10,6	8,1	9,8	8,9	15,5	14,1	6,6	4,1
4,1	8,7	6,0	9,6	–8,8	6,7	1,5	3,0	6,8	12,4
9,2	7,3	12,5	15,7	–0,7	15,5	16,8	4,3	9,6	–4,9
8,8	–4,2	12,3	7,2	1,0	0,3	–0,1	18,5	8,8	8,1

Вариант 16

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

Вариант 17

0,7	0,5	0,4	0,6	0,3	1,4	1,4	0,6	1,9	2,0
0,1	0,1	1,2	0,7	0,4	0,7	0,4	0,1	0,4	0,5
1,4	0,5	1,2	0,4	0,4	0,6	0,2	0,4	0,3	0,3
0,1	0,3	0,3	1,0	1,1	1,9	0,1	1,5	1,1	0,4
0,5	0,7	0,6	1,5	0,2	0,2	0,1	1,0	1,7	0,1.

Вариант 18

0,11	0,12	0.14	0,77	1,00	0,26	0,11	0,13	0,33	1,00
0,59	0,39	0,52	0,24	0,27	0,17	0,30	0,06	0,16	0,98
0,05	0,39	1,21	0,31	1,16	0,05	0,30	0,16	0,04	0,17
0,08	0,18	0,19	0,13	1,53	0,36	0,09	0,05	0,11	0,04
0,21	0,08	0,54	0,07	0,05	0,15	0,16	0,46	0,08	0,11.

Вариант 19

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

Вариант 20

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

Вариант 21

0,19	1,05	0,17	0,90	0,59	0,03	0,16	0,28	0,28	0,27
0,15	0,20	2,05	0,92	0,06	0,12	0,39	0,02	0,69	0,81
0,02	1,43	1,15	0,81	0,06	0.13	0,69	0,14	0,61	0,71
0,43	0,40	0,09	0,26	0,09	0,77	0,36	0,80	1,09	0,22
0,24	0,47	0,08	0,50	0,04	0,12	0,41	0,59	0,59	0,28.

Вариант 22

16,8	2,6	11,6	2,7	18,7	6,4	10,5	2,7	–1,3	4,0
4,8	10,7	5,9	4,9	9,4	8,4	7,3	9,8	5,5	12,6
7,1	10,8	3,1	10,2	9,4	9,4	10,4	14,5	12,5	15,1
3,8	16,9	8,3	6,9	6,0	19,9	18,3	12,8	21,5	12,4
7,0	8,1	4,7	11,9	–2,3	9,2	18,2	7,1	16,7	11,4

Вариант 23

12,7	15,0	4,9	3,4	6,6	21,8	6,6	6,0	4,1	9,1
5,7	4,2	19,5	7,1	16,4	6,5	–3,1	16,2	6,7	8,4
11,3	–5,7	12,6	6,1	10,3	14,6	15,9	14,6	12,5	0,8
0,8	13,2	14,0	6,4	5,5	7,5	6,5	8,6	–1,7	14,7
7,5	12,1	8,0	4,9	10,2	4,7	2,9	15,7	–3,3	15,0

Вариант 24

0,02	0,20	0.192	0.22	0,01	0,04	0,07	0,05	0,06	0,10
0,19	013	0,03	0,011	0,10	0,08	0,05	0.33	0.04	0,15
0,12	0,49	0,01	0,09	0,18	0,03	0,26	0,54	0,32	0,27
0,04	0,45	0,36	0,10	0,08	0,14	0,20	0,23	0,02	0,03
0,04	0,08	0,09	0,30	0,27	0,26	0,16	0,24	0,18	0,29

Вариант 25

2,4	2,7	21,2	–0,6	14,2	7,9	17,4	3,1	13,0	9,8
9,6	2,7	10,8	4,7	4,5	2,8	3,3	–0,2	6,7	15,2
16,1	8,7	14,6	11,6	12,4	20,5	1,1	7,1	7,8	1,3
22,0	1,5	12,7	4,9	11,7	16,1	1,4	2,8	4,6	6,4
13,5	24,6	8,2	1,1	6,4	–3,7	–3,4	1,7	8,3	6,6

Вариант 26

12,1	20,0	4,3	10,9	6,5	5,4	–7,0	–0,3	7,5	2,2
18,5	–0,4	1,2	9,0	9,1	14,1	16,1	5,9	10,3	8,8
–5,6	4,3	9,6	12,5	12,9	8,8	–0,8	12,4	7,1	2,5
1,5	17,0	5,5	12,4	23,0	17,8	3,3	6,3	4,5	14,5
8,2	7,7	16,2	11,5	0,0	14,4	11,1	17,7	13,0	3,0

Вариант 27

10,1	5,0	–4,8	–0,9	15,5	7,7	–3,9	7,8	4,5	7,5
6,0	5,3	11,8	4,6	20,2	5,8	14,6	15,4	4,2	–0,3
6,5	14,4	9,6	7,5	9,6	5,5	2,6	8,7	1,8	9,5
–6,2	8,0	7,2	–2,3	5,7	6,4	–2,7	3,7	7,4	4,3
–2,4	10,8	11,8	10,2	2,4	14,9	3,8	2,7	6,5	8,0

Вариант 28

0,04	0,16	0,28	0.16	0,07	0,01	0,05	0,38	0,61	0,49
0,17	0,29	0,24	0,14	0,06	0,30	0,05	0,21	0,29	0.11
0,07	0,07	0,18	0,02	0,27	0,55	0,08	0,39	0,13	0,04
0,07	0,08	0,23	0,25	0,16	0,03	0,02	0,12	0,44	0,17
0,22	0,18	0,02	0,02	0,08	0,12	0,03	0,51	0,35	0,46.

Вариант 29

2,4	10,7	0,4	15,0	6,7	10,3	10,0	7,7	6,4	3,4
12,0	7,1	1,3	8,7	17,1	21,0	6,0	10,0	–6,0	11,2
11,4	–2,9	9,2	5,0	20,9	27,9	6,5	10,0	12,8	12,4
18,8	–0,8	16,4	9,8	–3,9	–2,2	7,9	7,7	6,8	11,6
7,9	8,1	16,2	9,4	9,9	9,2	6,0	–0,9	0,3	–0,4

Вариант 30

4,9	8,4	14,5	8,6	16,0	14,1	10,7	6,3	12,6	1,1
11,7	14,7	4,7	15,2	23,7	0,3	10,4	5,6	13,1	8,0
11,5	–0,7	9,9	11,9	13,4	9,7	11,4	13,5	0,1	9,1
8,7	18,6	7,2	–5,4	3,3	8,3	7,8	9,7	10,0	7,3
10,8	6,9	13,4	10,1	0,1	11,4	–2,7	9,0	14,4	12,1

Вариант 31

3,46	0,52	2,90	3,88	3,79	3,81	2,92	2,80	4,70	–3,46
2,95	2,68	0,72	4,91	3,03	1,09	2,63	2,71	3,91	2,95
–1,26	3,66	2,82	1,59	2,40	1,27	3,81	3,09	2,59	2,30
–2,33	3,90	1,84	1,23	2,93	1,55	3,63	2,58	2,57	3,63
–3,06	3,68	–2,63	2,96	2,97	3,83	2,11	3,73	2,21	3,05

Вариант 32

2,59	1,54	2,72	3,57	0,03	1,84	2,68	4,67	2,44	1,39
3,77	4,24	3,18	4,19	0,84	2,13	3,024	1,64	4,98	0,64
2,52	1,81	1,58	2,32	1,94	0,70	2,69	2,93	4,37	3,13
0,39	1,26	2,76	2,93	0,21	3,78	2,28	2,58	4,38	4,96
1,55	3,19	1,42	4,43	2,32	4,60	2,09	4,28	4,66	4,82

7. Список рекомендуемой литературы

1. В.Е.Гмурман. Руководство к решению задач по теории вероятностей и математической статистике. – ВУЗ, 2011.
2. В.Е.Гмурман. Теория вероятностей и математическая статистика. – ВШ, 2002.
3. Д.Письменный. Конспект лекций по высшей математике. Ч. 1. Теория вероятностей и математическая статистика. Айрис Пресс, М. 2005.
4. Кирьянов Д.В. Самоучитель Mathcad 13, СПб.: БВХ – Петербург, 2006.

8. Приложения

Приложение А. Статистические таблицы

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

Таблица 8.1.

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988:	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3697
0,4	0,3683	0,3668	0,3652	0,2637	0,3621	0,3605	0,3589	0,3S72	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1738	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1569	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,00501	0,0048	0,0047	0,0043
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

Значения функции Лапласа $\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$

Таблица 8.2.

	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861

3,0	0,49865	**3,5**	0,49977	**4,0**	0,499968
3,1	0,49903	**3,6**	0,49984	**4,5**	0,499997
3,2	0,49931	**3,7**	0,49989	**5,0**	0,49999997
3,3	0,49952	**3,8**	0,49993		
3,4	0,49966	**3,9**	0,49995		

Значения $\chi^2_{m,\,\alpha}$ распределения χ^2 для числа степеней свободы m и вероятности

$$\alpha = P\left\{\chi^2_m > \chi^2_{m,\,\alpha}\right\}$$

Таблица 8.3.

m \ α	**0,01**	**0,025**	**0,05**	**0,95**	**0,975**	**0,99**
1	6,6	5,0	3,8	0,0039	0,000981	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,50
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,1	4,11
14	29,1	26,	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	41.2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0
47	72,4	67,8	64,0	32,3	29,9	27,4
48	73,7	69,0	65,2	33,1	30,7	28,2

Значения λ_q распределения Колмогорова для вероятности $q = P\left\{\xi_K \geq \lambda_q\right\}$

Таблица 8.4.

q	**0,50**	**0,40**	**0,30**	**0,20**	**0,10**	**0,05**	**0,02**	**0,01**	**0,001**
λ_q	0,828	0,895	0,975	1,073	1,224	1,358	1,520	1,627	1,950

Значения $t_{m,\beta}$ распределения Стьюдента для числа степеней свободы m и вероятности

$$\beta = P\left\{\left|\tau_m\right| < t_{m,\beta}\right\}$$

Таблица 8.5.

	0,1	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	3,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,95
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
47	1.68	2,01	2,41	2,68	3,27	3,51
48	1,68	2,01	2,40	2,68	3,27	3,50
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29