Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Ставропольский государственный аграрный университет»

Кафедра «Математика»

ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Учебное пособие

Ставрополь 2019

ДК 51 (075.8)
ББК 22.1я73
Л64

Литвин, Д.Б.
Л64 Основы математической статистики : учебное пособие / Д. Б. Литвин ; Ставропольский гос. Аграрный ун-т. – Ставрополь, 2019. – 82 с.

Пособие предназначено для студентов экономических направлений обучения. Содержание материала в целом соответствует второй части дисциплины «Теория вероятностей и математическая статистика».

ОГЛАВЛЕНИЕ

1. ПЕРВИЧНАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

1.1. Вариационный и статистический ряды

Пусть для изучения некоторого количественного признака X из всей объема N генеральной совокупности однородных объектов извлечена выборка объема n. Наблюдавшиеся значения $x_1, x_2, ..., x_k$ признака X называют *вариантами*, а последовательность вариант, записанных в возрастающем порядке, – *вариационным рядом*.

Статистическим распределением (рядом) выборки называют перечень вариант x_i вариационного ряда и соответствующих им частот n_i (сумма всех частот равна объему выборки n) или относительных частот $w_i = n_i / n$ (сумма всех относительных частот равна единице).

Различают *дискретный (точечный) и интервальный (сгруппированный)* статистический ряд.

Дискретный статистический ряд

Пример 1.1. Дана выборка, состоящая из чисел: 3.2, 4.1, 8.1, 8.1, 6.7, 4.4, 4.4, 3.2, 5.0, 6.7, 6.7, 7.5, 3.2, 4.4, 6.7, 6.7, 5.0, 5.0, 4.4, 8.1. Составить статистический ряд распределения абсолютных и относительных частот.

Решение. Объем выборки n = 20. Перепишем варианты в порядке возрастания: 3.2, 3.2, 3.2, 4.4, 4.4, 4.4, 4.4, 4.4, 5.0, 5.0, 5.0, 6.7, 6.7, 6.7, 6.7, 6.7, 7.5, 8.1, 8.1, 8.1.

Из данного *вариационного ряда* видно, что выборка состоит всего из шести вариант. Поэтому составим т.н. *дискретный (точечный) статистический ряд*:

Таблица 1.1

i	1	2	3	4	5	6	
x_i	3,2	4,4	5,0	6,7	7,5	8,1	
n_i	3	5	3	5	1	3	$\sum n_i = 20$
$n_i^{нак}$	3	8	11	16	17	20	
w_i	0,15	0,25	0,15	0,25	0,05	0,15	$\sum w_i = 1$
$w_i^{нак}$	0,15	0,4	0,55	0,8	0,85	1	

Здесь ряд дополнен накопленными частотами и частостями, где

n_i - частота (абсолютная);

$w_i = n_i / n$ - относительная частота (частость);

$n_1^{нак} = n_1$; $n_i^{нак} = n_i + n_{i-1}^{нак}$, для i>1 - накопленные частоты;

$w_1^{нак} = w_1$; $w_i^{нак} = w_i + w_{i-1}^{нак}$, для i>1 - накопленные частости. (1)

По статистическому ряду можно построить *полигон* абсолютных или относительных частот, а также *эмпирическую (выборочную) функцию распределения* признака

$$F_n^*(x) = \frac{n_x}{n}, \tag{2}$$

где n_x – количество элементов выборки, меньших чем x.

$F_n^*(x)$ есть относительная частота появления события $A = \{X < x\}$ в n независимых испытаниях. Главное различие между $F(x)$ и $F_n^*(x)$ состоит в том, что $F(x)$ определяет вероятность события A, а $F_n^*(x)$ – накопленную относительную частоту этого события.

Для дискретного статистического ряда функция $F_n^*(x)$ является "ступенчатой" и задается аналитически следующим соотношением:

$$F_n^*(x) = \begin{cases} 0 \quad \text{при} \quad x \le x_1; \\ w_i^{нак} \quad \text{при} \quad x_i < x \le x_{i+1}, \quad i = 1, 2, \ldots, n-1; \\ 1 \quad \text{при} \quad x_n < x. \end{cases} \tag{3}$$

На рисунке 1 представлены полигон и эмпирическая функция распределения $F_n^*(x)$ относительных частот для статистического ряда примера 1.1

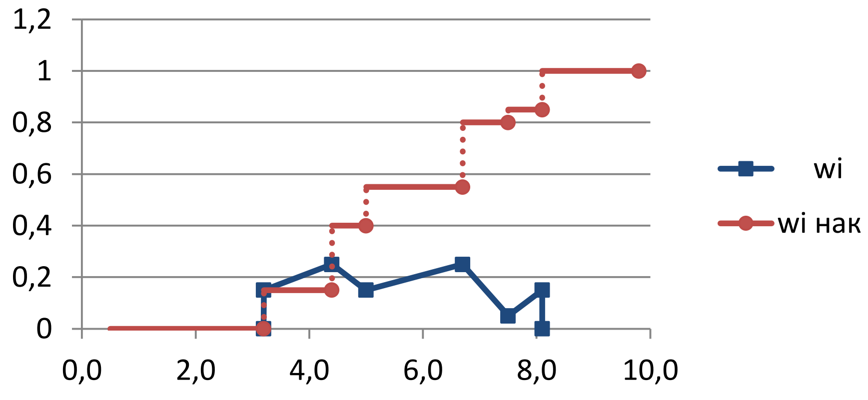


Рисунок 1.1 - Эмпирическая функция распределения

Интервальный статистический ряд

Если имеется выборка значений *непрерывного* количественного признака, где число вариант очень велико, то составляется *сгруппированный (интервальный) статистический ряд*. Для его получения интервал (a, b), содержащий все варианты, делится на k равных частей длины h, и в качестве абсолютных частот выступают количества вариант, попавших в данный интервал.

Количество интервалов k следует выбирать так, чтобы построенный ряд не был громоздким, но в то же время позволял выявлять характерные изменения случайной величины.

Для вычисления k рекомендуется использовать формулу Стерджеса:

$$k = 1 + \log_2 n \quad \text{или} \quad k = 1 + \log_2 10 \cdot \lg n, \text{ где } \log_2 10 \approx 3{,}322 \qquad (4)$$

с округлением k до ближайшего целого значения.

Необходимо, чтобы интервал (a, b) статистического ряда длины kh с небольшим "нахлестом" перекрывал вариационный размах наблюдаемого признака

$$kh = b - a \geq x_{\max} - x_{\min}, \text{ т.е. чтобы } \quad a \leq x_{\min}, \qquad b \geq x_{\max}, \qquad (5)$$

поэтому длину частного интервала выбирают из неравенства

$$h \geq \frac{x_{\max} - x_{\min}}{k}, \qquad (6)$$

где $x_{\max}, x_{\min}$ – наибольшее и наименьшее значения признака.

После нахождения частных интервалов определяется, сколько значений случайной величины попало в каждый конкретный интервал

$$[x_i, x_i + h), \quad i = 1, 2, ..., k, \qquad (7)$$

где k – число интервалов.

При этом в интервал включают значения, большие или равные нижней границе и меньшие верхней границы.

Для наглядного представления распределения наблюдаемого непрерывного признака X, исследуемого по выборке, используется ***гистограмма*** – столбчатая диаграмма, состоящая из прямоугольников, основания которых – частичные интервалы длины h, а высоты – плотности абсолютных n_i/h или относительных w_i/h частот (частостей). Гистограмма является статистическим аналогом плотности распределения *f(x)*, при этом общая площадь гистограммы относительных частот (частостей) равна единице, а гистограммы абсолютных частот - объему выборки.

Эмпирическая функция распределения $F_n^*(x)$ для интервального статистического ряда является кусочно-линейной. При этом, она равна нулю в начале первого частного интервала, а в конце каждого частного интервала i принимает значения соответствующих накопленных частостей $w_i^{нак}$.

Пример 1.2. Отклонение диаметра вала после шлифовки от номинального значения в мм представлено следующей выборкой (объемом $n = 69$):

-0,84	0,26	0,88	-0,33	0,72	-0,44	0,93
-0,14	0,71	-0,99	1,21	0,31	-0,29	0,79
0,34	0,38	-0,54	1,81	1,13	0,28	1,22
0,79	-0,89	0,89	-0,49	-0,03	0,44	-0,35
-0,3	1,34	-0,8	-0,32	1,15	-0,41	0,76
0,25	-0,18	-0,41	0,96	-0,63	0,86	0,8
-0,61	-0,65	-0,03	1,72	1,96	0,45	-0,6
1,15	0,19	0,35	0,5	0,77	0,91	-0,26
0,51	1,36	-0,01	0,42	0,63	-0,14	0,1
-0,08	-0,97	0,55	0,38	0,86	-0,57	

Необходимо построить интервальный вариационный ряд, гистограмму относительных частот и эмпирическую функцию распределения (кумуляту).

Решение.

1. Определим количество частных интервалов по формуле (4):

$$k = 1 + \log_2 n = 1 + \log_2 69 = 7{,}108524. \quad \text{Примем } k = 7.$$

2. Выполним ранжирование вариант в порядке возрастания.
3. Определим длину частного интервала по формуле (6):

$$x_{\max} = 1{,}96; \quad x_{\min} = -0{,}99;$$

$$h \geq \frac{x_{\max} - x_{\min}}{k} = \frac{1{,}96 - (-0{,}99)}{7} = \frac{2{,}95}{7} = 0{,}421. \quad \text{Примем } h = 0{,}43.$$

4. Проверим выполнение условия (5):

$$kh \geq x_{\max} - x_{\min}; \quad 7 \cdot 0{,}43 \geq 1{,}96 - (-0{,}99); \quad 3{,}01 \geq 2{,}95.$$

При этом, "нахлест" диапазона статистического ряда составляет $3{,}01\text{-}2{,}95 = 0{,}06$. Это значение определяет максимальное смещение влево от

$x_{\min}$ начала интервала (a, b) статистического ряда. В рассматриваемом примере смещение удобно принять *0,01*, т.е. левую границу примем

$$a = -1 \le x_{\min} = -0,99,$$

тогда правая граница

$$b = a + kh = 2,01 \ge x_{\max} = 1,96.$$

5. Определим границы частичных интервалов по формуле (7):

-1	-0,57	-0,14	0,29	0,72	1,15	1,58	2,01

6. Подсчитав количество вариант, попавших в соответствующие интервалы, получим искомый интервальный статистический ряд:

Таблица 1.2

i	1	2	3	4	5	6	7
X	[-1,00; -0,57)	[-0,57; -0,14)	[-0,14; 0,29)	[0,29; 0,72)	[0,72; 1,15)	[1,15; 1,58)	[1,58; 2,01)
n_i	9	13	11	13	14	6	3
$n_i^{нак}$	9	22	33	46	60	66	69
w_i	0,130	0,188	0,159	0,188	0,203	0,087	0,043
w_i/h	0,303	0,438	0,371	0,438	0,472	0,202	0,101
$w_i^{нак}$	0,130	0,319	0,478	0,667	0,870	0,957	1

7. Используя интервальный ряд (см. табл. 1), построим гистограмму плотностей относительных частот w_i/h, и эмпирическую функцию распределения $F_n^*(x)$, которые представлены на рисунке 1.2.

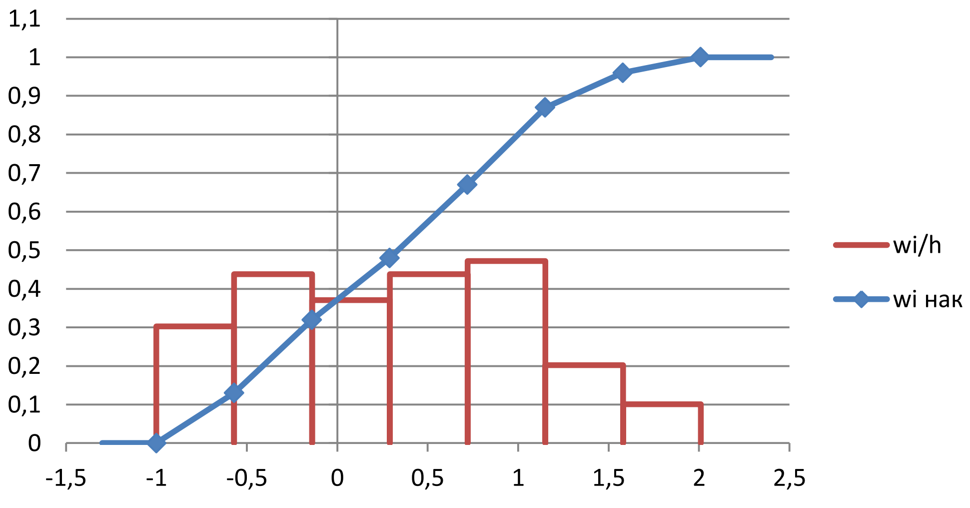


Рисунок 1.2 - Гистограмма и эмпирическая функция распределения

1.2. Первичная обработка в Excel

После внесения информации в электронную таблицу необходимо определить минимальный и максимальный варианты, размах вариационного ряда, количество вариант и число частных интервалов, например по формуле Стерджеса, и их границы. Затем вычислить частоты n_i. Эти операции можно сделать с помощью обычной сортировки. Существуют также встроенные специальные статистические функции.

Для вычисления частот n_i можно использовать ***функцию ЧАСТОТА***, обращение к которой имеет вид:

=ЧАСТОТА(*массив_данных;массив_интервалов*),

где *массив_данных* – адреса ячеек, для которых вычисляется частота n_i - количество вариант, меньших или равных граничному значению $(z_{i-1};z_i]$, $i=1,2,...,k+1$ (сравните с (7)); *массив_границ* – адреса ячеек, в которых размещаются упорядоченные по возрастанию значения z_i, $i=1,2,...,k+1$, где k – число интервалов.

Особенности.

Количество элементов в возвращаемом массиве на единицу больше числа элементов в массиве "*массив_интервалов*". Дополнительный элемент в возвращаемом массиве содержит количество значений, превышающих верхнюю границу интервала, содержащего наибольшие значения.

Функция ЧАСТОТА вводится как формула массива, т.е. предварительно выделяется интервал ячеек, в который будут помещены вычисленные частоты (число ячеек должно быть на 1 больше числа границ), затем вводится функция ЧАСТОТА с соответствующими аргументами, потом одновременно нажимаются клавиши [Ctrl] + [Shift] + [Enter].

Функция МАКС вычисляет максимальное значение из заданных аргументов. Обращение к ней имеет вид:

=МАКС(*арг1; арг2; …; арг255*),

где *арг1; арг2; …; арг255* – числовые константы или адреса ячеек, содержащих числовые величины.

Функция МИН вычисляет минимальное значение из заданных аргументов. Обращение к ней имеет вид:

=МИН(*арг1; арг2; …; арг255*),

где *арг1; арг2; …; арг255* – числовые константы или адреса ячеек, содержащих числовые величины.

Для подсчета количества элементов выборки (т.е. объема выборки) использовалась ***функция СЧЁТ***, обращение к которой имеет вид:

СЧЁТ(*массив_данных*),

где *массив_данных* – адреса ячеек или числовые константы.

Построение гистограммы частот и частостей возможно как с использованием стандартных инструментов Excel (Вставка-Гистограмма), так и с помощью специального инструмента из вкладки *Данные-Анализ данных,* в которой следует выбрать пункт *Гистограмма*.

Появится окно гистограммы, показанное на рисунке 1.3.

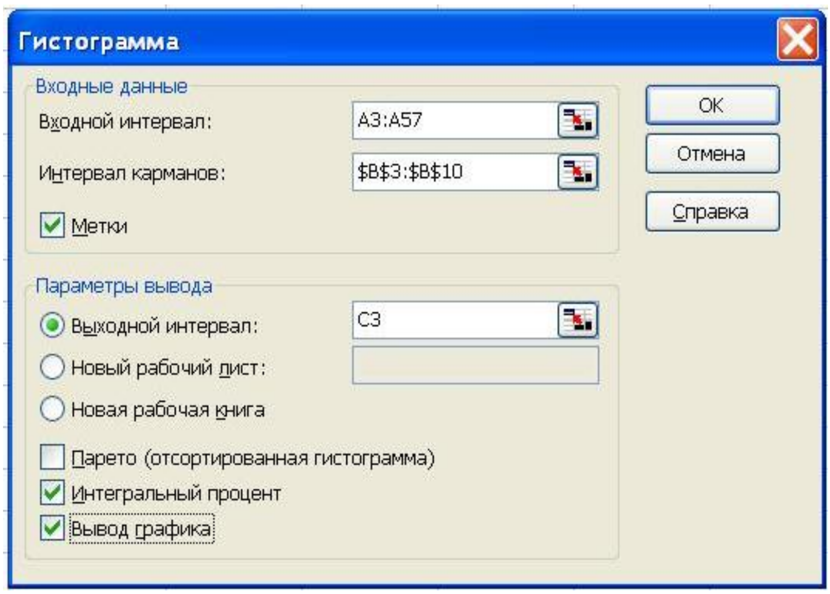


Рисунок 1.3 - Диалоговое окно режима ***Гистограмма***

В окне задаются следующие параметры:

Входной интервал: – адреса ячеек, содержащие выборочные данные.

Интервал карманов: (необязательный параметр) – адреса ячеек, содержащие границы интервалов (кармана). Эти значения должны быть введены в возрастающем порядке.

Если границы интервалов не заданы, то автоматически будет создан набор интервалов с одинаковой длиной

$$h = \frac{x_{\max} - x_{\min}}{[k] - 1},$$

где $[k]$ – целая часть величины $k = 1 + 3{,}322 \cdot \lg n$, n – объем выборки.

Метки – флажок, включаемый, если первая строка во входных данных содержит заголовки. Если заголовки отсутствуют, то флажок следует выключить.

Парето (*отсортированная гистограмма*) – устанавливается в активное состояние, чтобы представить w_i в порядке их убывания. Если параметр выключен, то w_i приводятся в порядке следования интервалов.

Интегральный процент – устанавливается в активное состояние для расчета выраженных в процентах накопленных относительных частот (процентный аналог значений выборочной функции распределения.

Результатом использования описываемого инструмента для данных **примера 1.2** является гистограмма, представленная на рисунке 1.4 (сравните этот результат с таблицей 1.1 и рисунком 1.2).

Таблица 1.2

h	*Частота*	*Интегральный %*
-1	0	0,00%
-0,57	8	11,76%
-0,14	13	30,88%
0,29	11	47,06%
0,72	14	67,65%
1,15	15	89,71%
1,58	4	95,59%
2,01	3	100,00%
Еще	0	100,00%

Особенность.

Столбцы гистограммы и "Интегральный %" строятся по серединам частных интервалов.

Построенная гистограмма является ненормированной: высоты прямоугольников в ней равны частостям w_i, а не их плотностям w_i/h. В этом случае единице равна сумма высот всех прямоугольников, а не сумма их площадей. Поэтому ненормированная гистограмма не может служить оценкой для плотности распределения случайной величины, из значений которой была сформирована выборка (особенно в случае неравных длин интервалов) [5].

Поэтому для построения гистограммы и эмпирической функции распределения рекомендуется использовать стандартные инструменты Excel.

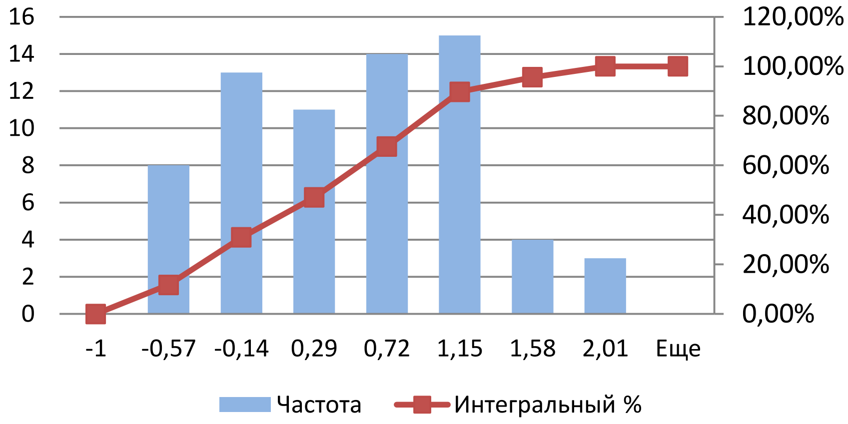


Рисунок 1.4 - График построенной гистограммы

ЗАДАЧИ

ЗАДАЧА 1. Построить полигон относительных частот. Найти и построить эмпирическую функцию по данному дискретному распределению:

i	1	2	3	4
x_i	2	5	7	8
n_i	1	3	2	4
$n_i^{нак}$				
w_i				
$w_i^{нак}$				

ЗАДАЧА 2. Построить гистограмму относительных частот и эмпирическую функцию по данному интервальному распределению:

Номер интервала	i	1	2	3	4	5
Частный интервал	$[x_i - x_{i+1})$	[10; 15)	[15; 20)	[20; 25)	[25; 30)	[30; 35)
Частоты	n_i	2	4	7	6	1
	$n_i^{нак}$					
Частости	w_i					
	$w_i^{нак}$					
Плотности	n_i/h					
	w_i/h					

2. ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

2.1. Требования к оценкам

Любая оценка (приближенное значение) параметра распределения вычисляется в статистике как функция случайных вариант наблюдаемого признака, а потому сама является в определенной мере случайной.

К статистическим оценкам обычно предъявляются требования:

- *состоятельности* - при увеличении числа наблюдений n она должна приближаться (сходиться по вероятности) к истинному значению параметра;

- *несмещенности* - ее математическое ожидание должно равняться значению параметра (отсутствие систематической ошибки);

- *эффективности* - среди всех несмещенных оценок выбранная должна обладать наименьшей дисперсией.

2.2. Точечные оценки M(X) и D(X)

По имеющейся выборке можно дать оценку математического ожидания и дисперсии генеральной совокупности. Несмещенной оценкой математического ожидания служит ***выборочное среднее***

$$\overline{x}_B = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} = \frac{1}{n}\sum_i n_i x_i, \quad n = n_1 + n_2 + \ldots + n_k, \tag{8}$$

то есть среднее арифметическое всех элементов выборки.

Оценкой дисперсии может служить ***выборочная дисперсия***

$$D_B = \sum_{i=1}^{k} \frac{n_i (x_i - \overline{x}_B)^2}{n} = \overline{\left(x - \overline{x}_B\right)^2}. \tag{9}$$

Более удобна формула - *средний квадрат минус квадрат среднего*:

$$D_B = \sum_{i=1}^{k} \frac{n_i x_i^2}{n} - \overline{x}_B^2 = \overline{x^2} - \left(\overline{x}_B\right)^2. \tag{10}$$

Выборочная дисперсия - смещенная в сторону занижения оценка генеральной дисперсии $D_Г$, и ее математическое ожидание $M(D_B) = \frac{n-1}{n} D_Г$. Поэтому вводится несмещенная оценка генеральной дисперсии – ***исправленная выборочная дисперсия***

$$s^2 = \frac{n}{n-1} D_B. \tag{11}$$

Соответственно $s=\sqrt{s^2}$ является ***исправленным выборочным средним квадратическим отклонением.***

Замечание 1. Если первоначальные варианты x_i — большие числа, то для упрощения расчета целесообразно вычесть из каждой варианты одно и то же число C, т. е. перейти к *условным вариантам* $u_i = x_i - C$ (в качестве C выгодно принять число, близкое к выборочной средней; поскольку выборочная средняя неизвестна, число C выбирают «на глаз»).

Тогда для выборочного среднего справедливо

$$u_i = x_i - C \qquad \to \qquad \overline{x}_B = C + \frac{1}{n}\sum_i n_i u_i \quad \text{или} \quad \overline{x}_B = C + \overline{(x-C)}; \qquad (12)$$

выборочная дисперсия при этом не изменится

$$u_i = x_i - C \qquad \to \qquad D_B(X) = D_B(u) = \overline{u^2} - \left(\overline{u}\right)^2 \qquad (13)$$

Замечание 2. Если первоначальные варианты являются десятичными дробями с k десятичными знаками после запятой, то, чтобы избежать действий с дробями, умножают первоначальные варианты на постоянное число $C=10^k$, т.е. переходят к *условным вариантам* $u_i = C\cdot x_i$. При этом выборочное среднее увеличится в C раз, а дисперсия - в C^2 раз, поэтому справедливы выражения:

$$u_i = C\cdot x_i \qquad \to \qquad \overline{x}_B = \frac{\overline{u}_B}{C}; \qquad D_B(X) = \frac{D_B(u)}{C^2}. \qquad (14)$$

Пример 2.1. Найти выборочное среднее, исправленную выборочную дисперсию и исправленное выборочное среднее квадратическое отклонение для выборок, заданных в примере 1.1.

x_i	3,2	4,4	5,0	6,7	7,5	8,1
n_i	3	5	3	5	1	3
w_i	0,15	0,25	0,15	0,25	0,05	0,15

Решение.

$$\overline{x}_B = \frac{3,2\cdot 3 + 4,4\cdot 5 + 5\cdot 3 + 6,7\cdot 5 + 7,5\cdot 1 + 8,1\cdot 3}{20} = 5,595;$$

$$s^2 = \frac{(3,2-5,595)^2\cdot 3 + \ldots + (8-5,595)^2\cdot 3}{19} = 2,84; \; s=\sqrt{2,84} = 1,69.$$

Замечание. В интервальном статистическом ряде вариантами следует считать середины частичных интервалов.

2.3. Другие характеристики вариационного ряда

Кроме выборочной средней и выборочной дисперсии применяются и другие характеристики вариационного ряда. Укажем главные из них.

Медианой *Me* называют варианту, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно, т. е. *n=2k+1*, то $Me = x_{k+1}$, если четно, т.е. *n* = *2k*, то $Me = (x_k + x_{k+1})/2$.

Для примера 1.1: *n* = *20*, *k=10*, $Me = (x_{10} + x_{11})/2 = (6,7+6,7)/2 = 6,7$.

Для интервального ряда медиане соответствует значение кумуляты частостей равное 0,5, поэтому используют формулу (15), которая получена из подобия треугольников, показанных на рис. 2.1

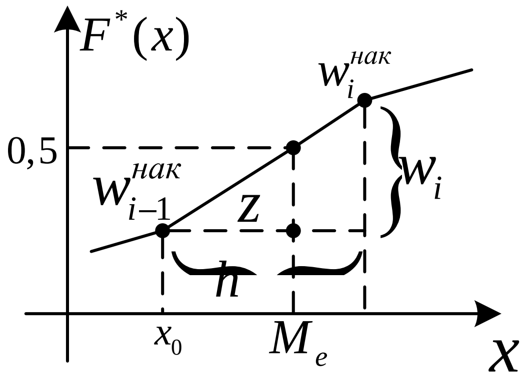


Рисунок 2.1 - К определению медианы *Me*

$$\frac{z}{h} = \frac{0,5 - w_{i-1}^{нак}}{w_i}; \qquad Me = x_0 + z;$$

$$Me = x_0 + h\frac{0,5 - w_{i-1}^{нак}}{w_i}, \qquad (15)$$

где x_0 - начало интервала, содержащего медиану; w_i, $w_{i-1}^{нак}$ - частость и накопленная частость медианного и предмедианного интервалов соответственно.

Для примера 1.2 (см. табл.1.2) получим по формуле (15):

$$Me = 0,29 + 0,43\frac{0,5 - 0,478}{0,188} = 0,34.$$

Модой *Mo* называют варианту, которая имеет наибольшую частоту.

Для интервального ряда используют формулу (16), которая получена из подобия треугольников, показанных на рис. 2.2.

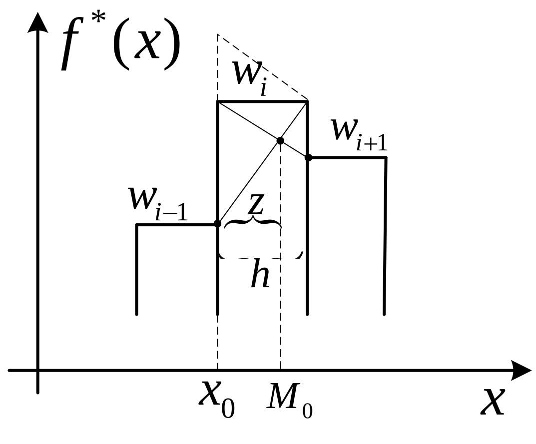


Рисунок 2.2 - К определению моды *Mo*

$$\frac{z}{h} = \frac{w_i - w_{i-1}}{(w_i - w_{i-1}) + (w_i - w_{i+1})};$$

$$M_0 = x_0 + z;$$

$$M_0 = x_0 + h\frac{w_i - w_{i-1}}{2w_i - w_{i-1} - w_{i+1}}, \qquad (16)$$

где x_0 - начало интервала, содержащего моду; w_{i-1}, w_i, w_{i+1} - частоты предмодального, модального и постмодального интервалов соответственно.

В примере 1.1 имеется две моды $M_{0\_1} = 4,4$; $M_{0\_2} = 6,7$.

Для примера 1.2 (см. табл.1.2) получим по формуле (16):

$$M_0 = 0,72 + 0,43 \frac{0,203 - 0,188}{2 \cdot 0,203 - 0,188 - 0,087} = 0,769.$$

Размахом варьирования *R* называют разность между наибольшей и наименьшей вариантами:

$$R = x_{\max} - x_{\min}.$$

Коэффициент вариации *V* - безразмерная величина, которая характеризует в процентах долю выборочного СКО от выборочной средней:

$$V = \frac{\sigma_B}{\overline{x}} 100\%.$$

2.4. Вычисление точечных оценок в Excel

Для вычисления выборочного среднего (8) используется ***функция СРЗНАЧ***, обращение к которой имеет вид:

=СРЗНАЧ(*арг1; арг2; ...; арг255*),

где *арг1; арг2; ...; арг255* – числа или адреса ячеек (не более 225), содержащих числовые данные. Если ячейка содержит текстовые, логические значения или ячейка пуста, то такие ячейки игнорируются.

Для вычисления выборочной (9) и исправленной (11) дисперсий используются ***функции ДИСПР*** и ***ДИСП соответственно,*** обращение к которым имеет вид:

=ДИСПР(*арг1; арг2; ...; арг255*);

=ДИСП(*арг1; арг2; ...; арг255*).

Для вычисления суммы квадратов отклонений $\sum_{i=1}^{n}(x_i - \overline{x}_B)^2$ используется ***функция КВАДРОТКЛ***, обращение к которой имеет вид:

=КВАДРОТКЛ(*арг1; арг2; ...; арг255*).

Для вычисления ***выборочного и исправленного СКО*** используются соответственно функции:

=СТАНДОТКЛОНП (*арг1; арг2; ...; арг255*);

=СТАНДОТКЛОН (*арг1; арг2; ...; арг255*).

Находят применение также следующие встроенные функции.

Функция ЭКСЦЕСС вычисляет оценку

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}_в}{d_в}\right)^2-\frac{3(n-1)^2}{(n-2)(n-3)}$$

для характеристики *эксцесс* $\frac{\mu_4}{\sigma^4}-3$, которая определяет островершинность или плосковершинность плотности распределения.

Функция МОДА вычисляет наиболее часто встречающееся значение в заданных аргументах функции, т.е. значение, встречающееся в выборке с максимальной частотой.

Если в заданных значениях аргументов ***нет повторяющихся значений***, то функция возвращает признак ошибки #Н/Д.

Функция МЕДИАНА вычисляет значение выборки, приходящееся на середину упорядоченной выборочной совокупности. Если выборка имеет четное число элементов, то значение функции будет равно среднему двух значений, находящихся по середине упорядоченной выборочной совокупности.

Функция СКОС вычисляет оценку

$$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\frac{\left(x_i-\overline{x}_в\right)^3}{d_в^{3/2}}$$

для характеристики асимметрии $\frac{\mu_3}{\sigma^3}$, которая для симметричной плотности распределения равна 0.

Основные характеристики положения, разброса и асимметрии можно также вычислить, используя вкладку ***Данные - Анализ данных - Описательная статистика*** *пакета анализа*.

В появившемся диалоговом окне Описательная статистика параметр *Уровень надежности* включается, если необходимо вычислить доверительный интервал для математического ожидания с задаваемым ($в$ %) уровнем надежности γ. *Уровень надежности* – определяет величину $\Delta_{\overline{x}}$, от которой зависит доверительный интервал для математического ожидания, имеющий вид

$$\left[\overline{x}_в-\Delta_{\overline{x}},\overline{x}_в+\Delta_{\overline{x}}\right],$$

где $\overline{x}_в$ – выборочное среднее (подробнее см. Интервальные оценки).

Назначение остальных параметров достаточно очевидно.

ЗАДАЧИ

ЗАДАЧА 1. Из генеральной совокупности извлечена выборка объема $n = 60$:

x_1	1	3	6	26
n_i	8	40	10	2

Найти несмещенную оценку генеральной средней.

ЗАДАЧА 2. Найти выборочную среднюю по данному распределению выборки объема $n = 10$:

x_1	1250	1270	1280
n_i	2	5	3

(Перейти к условным вариантам $u_i = x_i - 1270$)

ЗАДАЧА 3. По выборке объема $n = 51$ найдена смещенная оценка $D_B = 5$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

ЗАДАЧА 4. В итоге пяти измерений длины стержня одним прибором (без систематических ошибок) получены следующие результаты (в мм): 92; 94; 103; 105; 106. Найти: а) выборочную среднюю длину стержня; б) выборочную и исправленную дисперсии ошибок прибора.

Указание. Перейти к условным вариантам $u_i = x_i - C$.

ЗАДАЧА 5. Ниже приведены результаты измерения роста (в см) случайно отобранных 100 студентов.

Рост	154-158	158-162	162-166	166-170	170-174	174-178	178-182
среднее							
Число студентов	10	14	26	28	12	8	2
u = x-170							
u^2							

Найти выборочную среднюю и выборочную дисперсию роста обследованных студентов.

Указание. Найти середины интервала и принять их в качестве вариант.

ЗАДАЧА 6. Найти выборочное среднее и выборочную дисперсию по данному распределению выборки объема $n=10$:

x_i	0,01	0,04	0,08
n_i	5	3	2

Указание. Перейти к условным вариантам $u_i = 100x_i$.

2.5. Интервальные оценки

Точечная оценка при малом объеме выборки может существенно отличаться от оцениваемого параметра, поэтому важно знать, насколько истинное значение параметра может отклоняться от найденной точечной оценки. Интервал вида $|\theta-\theta^*|<\delta,$ где Θ - истинное значение оцениваемого параметра, а Θ^* - его точечная оценка, называется *доверительным интервалом*, а вероятность $\gamma = p\left(|\theta-\theta^*|<\delta\right)$ - *доверительной вероятностью* или *надежностью*.

Границы доверительного интервала являются случайными величинами и с вероятностью γ накрывают истинное значение оцениваемого параметра.

Для построения доверительного интервала требуется знать закон распределения исследуемой случайной величины $f\left(\theta^*\right)$, как показано на рисунке 2.3.

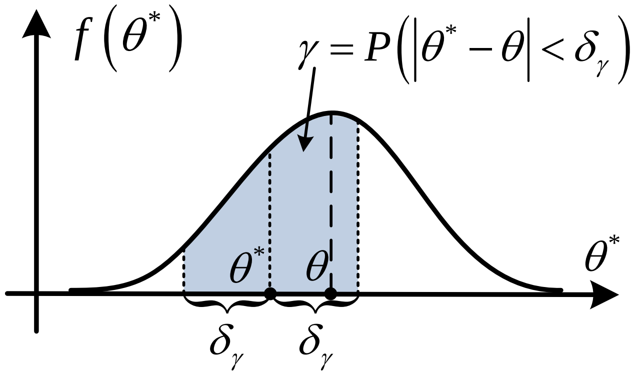


Рисунок 2.3 - К определению интервальной оценки θ^*

Некоторые статистические распределения

Генеральные совокупности часто имеют нормальный закон распределения. В этом случае многие выборочные характеристики, в том числе $\bar{X}_в, D_в, S^2$, выражаются через небольшое число специальных распределений. Как правило, в математической статистике используются не плотности этих распределений, а некоторые их числовые характеристики, представленные таблицами. Чаще всего в качестве такой характеристики выступает либо квантиль распределения [5], либо критическая точка распределения.

Квантилем уровня p $(0<p<1)$ или *p-квантилем* случайной величины X называется такое число d_p, что вероятность $p=P\left(X<d_p\right)$ равна заданной величине *p*.

Критической точкой распределения случайной величины X (для правосторонней критической области) называют такое число $x_{kp}(\alpha)$, что вероятность $\alpha = P\left(X > x_{kp}(\alpha)\right)$ равна заданной величине α - *уровню значимости*.

Таким образом квантиль определяет границу площади "левого хвоста" под плотностью распределения, а критическая точка - подобную границу площади, но "правого хвоста" под плотностью распределения, как показано на рисунке 2.4.

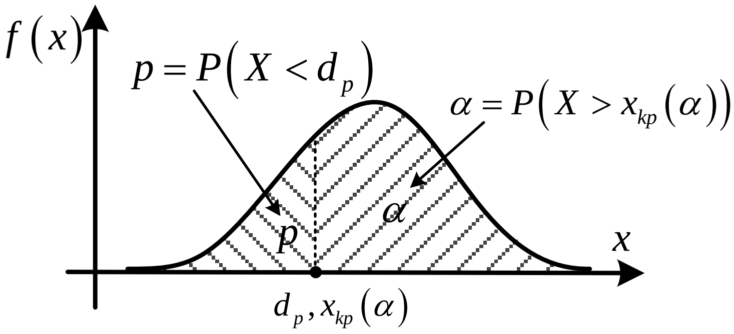


Рисунок 2.4 - К определению квантиля d_p и критической точки $x_{kp}(\alpha)$ распределения (если $d_p = x_{kp}(\alpha)$, как на рисунке, то $\alpha = 1 - p$)

Распределение χ^2 (распределение К. Пирсона).

Рассмотрим несколько распределений, которым подчиняются выборочные характеристики и которые используются для построения интервальных оценок.

Пусть $N_1, \ldots, N_n$ – независимые нормально распределенные случайные величины с параметрами N(0,1). Распределение случайной величины

$$\chi_n^2 = N_1^2 + N_2^2 + N_3^2 + \ldots + N_n^2 \tag{17}$$

называется *распределением χ^2 с n степенями свободы*.

Заметим, что количество степеней свободы *n* является единственным параметром χ^2-распределения и значения χ^2 неотрицательны, т.е. $P(\chi_n^2 < 0) = 0$.

Определим математическое ожидание величины χ^2.

По определению (17) имеем

$$M(\chi_n^2) = M\left(\sum_{i=1}^{n} N_i^2\right) = \sum_{i=1}^{n} M(N_i^2) = \sum_{i=1}^{n}\left[D(N_i) + M^2(N_i)\right],$$

так как $D(X) = M(X^2) - M^2(X)$. Но $D(N_i) = 1$, $M(N_i) = 0$, а значит, $M(\chi_n^2) = n$.

Вычислим дисперсию случайной величины χ_n^2. Так как случайные величины $N_1^2,...,N_n^2$ независимы, то

$$D(\chi_n^2) = nD(N_1^2) = n\left[M(N_1^4) - M(N_1^2)\right]. \qquad (18)$$

Плотность распределения случайной величины N_1 равна $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, значит,

$$M(N_1^4) = \int_{-\infty}^{\infty} x^4 f(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2}} = 3.$$

Последний интеграл вычисляется методом интегрирования по частям. Далее, так как $M(N_1^2) = 1$, то $D(\chi_n^2) = n(3-1) = 2n$. Таким образом, χ^2-распределение с n степенями свободы имеет следующие числовые характеристики:

$$M[\chi_n^2] = n; \quad D[\chi_n^2] = 2n. \qquad (19)$$

Согласно центральной предельной теореме, если случайные величины $N_1^2, N_2^2, ..., N_n^2$ независимы, одинаково распределены и имеют конечные дисперсии, то последовательность $\chi_n^2 = N_1^2 + ... + N_n^2$ асимптотически нормальна. Другими словами, при больших значениях n распределение случайной величины χ_n^2 близко к нормальному распределению с параметрами $a = n, \sigma^2 = 2n$. Однако при малых значениях n функция плотности случайной величины χ_n^2 значительно отличается от кривой Гаусса.

На рисунке 2.5 показаны плотности распределения $f(x)$ случайной величины χ_n^2 при $n=1$, $n=5$ и $n=15$.

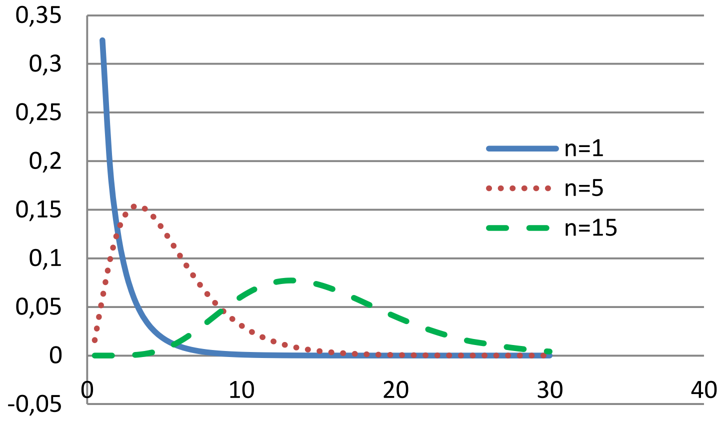


Рисунок 2.5 - Плотность распределения χ_n^2

Видно, что при увеличении n плотность $f(x)$ "приближается" к плотности нормального распределения.

Замечательное свойство распределения χ_n^2: сумма независимых случайных величин $\chi_n^2 + \chi_m^2$ также распределена по закону χ^2 с $(n+m)$ степенями свободы.

Распределение Стьюдента (t-распределение).

Пусть $Z = N(0,1)$ – нормально распределенная случайная величина с параметрами $a = 0, \sigma = 1$ (z-статистика), а χ_n^2 – независимая от $N(0,1)$ случайная величина, подчиняющаяся распределению χ^2 с n степенями свободы. Тогда распределение случайной величины

$$T_n = \frac{N(0,1)\sqrt{n}}{\sqrt{\chi_n^2}} \tag{20}$$

называется *t-распределением* или *распределением Стьюдента.* Сама случайная величина (20) называется *t-величиной с n степенями свободы.*

При больших значениях n кривая $f_n(t)$ близка к кривой нормального распределения $N(0,1)$. Поэтому в практических расчетах при n > 30 часто считают, что $f_n(t) \approx f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.

Функция плотности $f_n(t)$ симметрична относительно оси ординат.

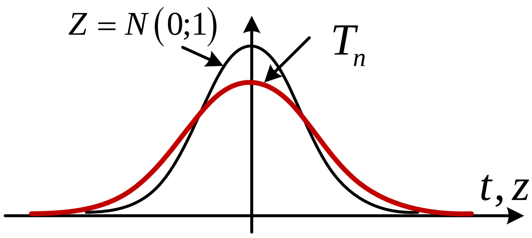


Рисунок 2.6 - Плотности z- и t-распределений

Распределение Фишера (F-распределение).

Пусть χ_n^2 и χ_m^2 – независимые случайные величины, имеющие χ^2-распределение с n и m степенями свободы соответственно. Распределение случайной величины

$$F_{n,m} = \frac{\chi_n^2 / n}{\chi_m^2 / m} \qquad (21)$$

называется *F-распределением* или *распределением Фишера* с n и m степенями свободы, а сама величина (21) $- F_{n,m}$ величиной. Так как случайные величины $\chi_n^2 \geq 0$ и $\chi_m^2 \geq 0$, то $F_{n,m} \geq 0$.

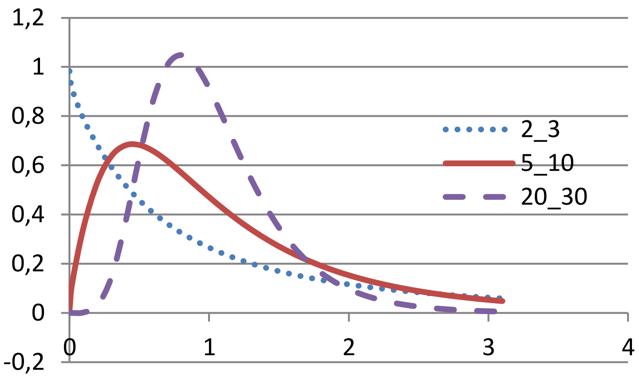


Рисунок 2.7 - Плотность F-распределений со степенями свободы *2 и 3, 5 и 10, 20 и 30* соответственно.

Из рисунка 2.7 видно, что с увеличением или одной или другой степени свободы плотность распределения становится более узкой и высокой со средним значением равным единице.

Теорема о распределении выборочных характеристик

Теорема о распределении выборочных характеристик $\bar{X}_в$ и $D_в$ (доказана Р. Фишером) [5].

Если генеральная совокупность X распределена по нормальному закону с параметрами $N(a;\sigma)$, то:

а) случайная величина $\bar{X}_в$ распределена нормально с параметрами

$$\bar{X}_в \to N\left(a; \frac{\sigma}{\sqrt{n}}\right); \qquad (22)$$

б) $\dfrac{D_в}{\sigma^2/n} = \dfrac{nD_в}{\sigma^2}$ имеет распределение χ_{n-1}^2

$$\frac{nD_в}{\sigma^2} \to \chi_{n-1}^2; \qquad (23)$$

в) случайные величины $\bar{X}_в$ и $D_в$ независимы.

Ограничимся доказательством утверждения а). Очевидно, что $\bar{X}_в$ есть линейная комбинация

$$\bar{X}_в = \frac{1}{n}\left(X_1 + X_2 + ... + X_n \right)$$

независимых, нормально распределенных случайных величин. Как отмечалось в курсе теории вероятностей, в этом случае случайная величина $\bar{X}_в$ распределена нормально. Имеем

$$M(\bar{X}_в) = M\left(\frac{x_1 + x_2 + ... + x_n}{n} \right) = \frac{M(x_1) + ... + M(x_n)}{n} = \frac{na}{n} = a, \qquad (24)$$

$$D(\bar{X}_в) = D\left(\frac{x_1 + ... + x_n}{n} \right) = \frac{D(x_1) + ... + D(x_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \qquad (25)$$

Тем самым первое утверждение теоремы доказано.

Как следует из в), используя случайные величины $\bar{X}_в$ и $D_в$, можно составить случайную величину T_{n-1}. Действительно, пронормировав $\bar{X}_в$, получим $\frac{(\bar{X}_в - a)\sqrt{n}}{\sigma} = N(0,1)$. Так как $\bar{X}_в$ и $D_в$ независимы, то по (20)

$$T_{n-1} = \frac{(\bar{X}_в - a)\sqrt{n}\sqrt{n-1}}{\sigma} : \sqrt{\frac{nD_в}{\sigma^2}} = \frac{(\bar{X}_в - a)\sqrt{n-1}}{\sqrt{D_в}}.$$

Итак, мы получили

Следствие. Если условия теоремы о распределении выборочных характеристик выполнены, то случайная величина

$$\frac{(\bar{X}_в - a)\sqrt{n-1}}{\sqrt{D_в}} = \frac{(\bar{X}_в - a)\sqrt{n}}{\sqrt{S^2}} = T_{n-1} \qquad (26)$$

имеет распределение Стьюдента с $(n-1)$ степенями свободы.

Здесь использовано выражение для исправленной дисперсии $S^2 = \frac{n}{n-1} D_в$.

Интервальные оценки математического ожидания нормального распределения

Пусть генеральная совокупность X распределена ***по нормальному закону*** $N(a,\sigma)$, причем ***параметр*** σ ***известен***, а параметр $a = M(X)$ требуется оценить с надежностью γ. По теореме о распределении выборочных характеристик случайная величина $Z = \frac{\left(\bar{X}_в - a \right)\sqrt{n}}{\sigma}$ распределена по закону $N(0,1)$ с плотностью вероятностей $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$ и называется *z-статистикой*.

На рисунке 2.8 изображены графики плотностей случайной величины $\overline{X}_в$, распределенной по $N\left(a,\frac{\sigma}{\sqrt{n}}\right)$ и *z-статистики,* распределенной по $N(0,1)$.

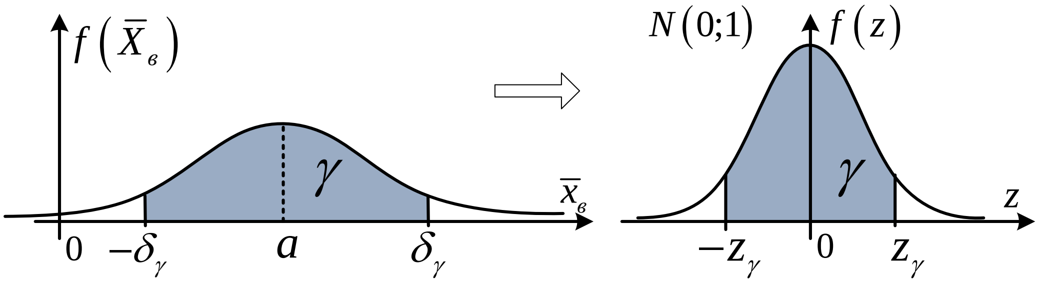


Рисунок 2.8 - К построению доверительных интервалов

Зададимся требуемой надежностью γ оценки *a* и, полагая известными σ и объем выборки *n*, определим точность оценки δ_γ - ширину доверительного интервала

$$\gamma=P\left(\left|\overline{X}_в-a\right|<\delta_\gamma\right)=P\left(\left|\frac{\left(\overline{X}_в-a\right)\sqrt{n}}{\sigma}\right|<\delta_\gamma\frac{\sqrt{n}}{\sigma}\right)=P\left(|Z|<z_\gamma\right), \tag{27}$$

где $\delta_\gamma=z_\gamma\frac{\sigma}{\sqrt{n}}$ - искомый доверительный интервал.

Выражение (27) эквивалентно следующему

$$P\left(\overline{X}_в-z_\gamma\frac{\sigma}{\sqrt{n}}<a<\overline{X}_в+z_\gamma\frac{\sigma}{\sqrt{n}}\right)=\gamma, \tag{28}$$

которое определяет ***доверительный интервал (интервальную оценку) для математического ожидания а*** с точностью $\delta_\gamma=z_\gamma\frac{\sigma}{\sqrt{n}}$ и надежностью γ.

Значение $z=z_\gamma$ находится с использованием интегральной функции Лапласа $\Phi(z)$, представленной в Приложении 2. Действительно,

$$P\left(-z<N(0,1)<z\right)=\Phi(z)-\Phi(-z)=2\Phi(z)=\gamma. \tag{29}$$

По табл. П2 определяем значение z, удовлетворяющее уравнению

$$\Phi(z)=\frac{\gamma}{2};\qquad z=\Phi^{-1}\left(\frac{\gamma}{2}\right). \tag{30}$$

Пример 2.1. Дана выборка значений нормально распределенной случайной величины: 2, 3, 3, 4, 2, 5, 5, 5, 6, 3, 6, 3, 4, 4, 4, 6, 5, 7, 3, 5. Найти с доверительной вероятностью γ = 0,95 границы доверительных интервалов для математического ожидания, если известно СКО распределения $\sigma = 1{,}37$.

Решение.

Поскольку распределение нормальное и дисперсия известна $\sigma^2 = 1{,}37^2 = 1{,}88$, то для оценки $\overline{x}_B$ используем z-распределение. Найдем n=20, $\overline{x}_B$ = 4,25. По таблице Прил.2 определим аргумент $z=1{,}96$, при котором функции Лапласа $\Phi(z) = \dfrac{0{,}95}{2} = 0{,}475$. Тогда из формулы (28)

$$4{,}25 - 1{,}96\frac{1{,}37}{\sqrt{20}} < a < 4{,}25 + 1{,}96\frac{1{,}37}{\sqrt{20}};$$

$$4{,}25 - 0{,}6 < a < 4{,}25 + 0{,}6; \qquad 3{,}65 < a < 4{,}85.$$

Минимальный объем выборки n, при котором оценку математического ожидания a можно получить с заданной надежностью γ и точностью δ_γ.

Интервальная оценка a зависит от трех взаимосвязанных параметров: надежности γ, точности δ_γ и объема выборки n. Задаваясь двумя из них можно определить оставшийся.

Пусть $\delta_\gamma = \left|\overline{X}_в - a\right|$. Тогда на основании формулы (27) ***минимальный объем выборки n***, гарантирующий оценку математического ожидания a с заданной надежностью γ и точностью δ_γ определяется неравенствами

$$\delta_\gamma \ge z_\gamma \frac{\sigma}{\sqrt{n}}, \qquad \delta_\gamma\sqrt{n} \ge z_\gamma\sigma, \qquad n \ge \left(\sigma\frac{z_\gamma}{\delta_\gamma}\right)^2. \tag{31}$$

Пример 2.2. Найти минимальный объем выборки, при котором с надежностью 0,95 точность оценки математического ожидания a генеральной совокупности по выборочной средней равна $\delta = 0{,}5$, если известно СКО $\sigma = 1{,}37$ нормально распределенной генеральной совокупности.

Решение.

По таблице Прил.2 определим аргумент z=1,96, при котором функции Лапласа $\Phi(z) = \dfrac{0{,}95}{2} = 0{,}475$. Тогда по формуле (31)

$$n \geq \left(\sigma \frac{z_\gamma}{\delta_\gamma} \right)^2 = \left(1{,}37 \frac{1{,}96}{0{,}5} \right)^2 = (5{,}37)^2 = 28{,}84. \quad \text{Т.о., } n_{\min} = 29.$$

При неизвестном СКО и объеме выборки $n < 30$, для нормального ЗР, на основании теоремы (26) имеем

$$\frac{\left(\bar{X}_в - a \right)\sqrt{n}}{s} = T_{n-1}.$$

Поэтому доверительный интервал для математического ожидания при заданной надежности γ определяется так:

$$\bar{x}_B - \frac{t_\gamma \cdot s}{\sqrt{n}} < a < \bar{x}_B + \frac{t_\gamma \cdot s}{\sqrt{n}}. \tag{32}$$

Здесь s – исправленное выборочное среднее квадратическое отклонение, а $t_\gamma = t_\gamma(\gamma, k)$ – критическая точка распределения Стьюдента, определяемая выражением $P\left(\left|T_k\right| < t\left(\gamma, k\right)\right) = \gamma$, где $k = n - 1$ – количество степеней свободы. Значения t_γ можно найти из таблицы Приложения 3 по известным n и γ.

При неизвестном СКО и объеме выборки $n \geq 30$, для нормального ЗР, полагают, что распределение Стьюдента несущественно отличается от нормального. Поэтому для оценки математического ожидания применяют формулу (28), где вместо σ используют s.

Пример 2.3. Дана выборка значений нормально распределенной случайной величины: 2, 3, 3, 4, 2, 5, 5, 5, 6, 3, 6, 3, 4, 4, 4, 6, 5, 7, 3, 5. Найти с доверительной вероятностью γ = 0,95 границы доверительного интервала для математического ожидания.

Решение.

Поскольку объем выборки небольшой n = 20 и дисперсия не известна, то для оценки $\bar{x}_B$ используем распределение Стьюдента. Найдем $\bar{x}_B$ = 4,25, s = 1,37. По таблице Прил.3 определим $t_\gamma\left(0{,}95;\ 19\right) = 2{,}093$. Тогда по формуле (32)

$$4{,}25 - \frac{2{,}093 \cdot 1{,}37}{\sqrt{20}} < a < 4{,}25 + \frac{2{,}093 \cdot 1{,}37}{\sqrt{20}}; \quad 3{,}64 < a < 4{,}86$$

- доверительный интервал для математического ожидания.

Интервальные оценки дисперсии нормального распределения

Как и при построении интервальных оценок для математического ожидания, в данном случае также необходимо определить статистику (функцию от наблюдаемых вариант), распределение которой было бы известно и включало бы оцениваемый параметр σ.

В соответствии с теоремой о распределении выборочных характеристик (23) такой статистикой может быть случайная величина $\dfrac{nD_в}{\sigma^2}$ или $\dfrac{(n-1)S^2}{\sigma^2}$, распределенная по закону χ^2_{n-1} с $(n-1)$ степенями свободы. Заметим, что распределение χ^2, в отличие от распределения Стьюдента, не является симметричным, поэтому для доверительного интервала целесообразно выбрать два предела $\chi^2_{лев,\gamma}$ и $\chi^2_{пр,\gamma}$ так, чтобы площади "хвостов" были равными $\dfrac{\alpha}{2}$ [5]

$$P\left(\chi^2_{n-1} < \chi^2_{лев,\gamma}\right) = P\left(\chi^2_{n-1} > \chi^2_{пр,\gamma}\right) = \frac{\alpha}{2}, \tag{33}$$

где $\alpha = 1-\gamma$, α, γ – уровень значимости и надежность оценки;

$\chi^2_{лев,\gamma}$ – критическая точка χ^2_{n-1}-распределения уровня $1-\alpha/2$ (или квантиль уровня $\alpha/2$), $\chi^2_{пр,\gamma}$ – критическая точка уровня $\alpha/2$ (или квантиль уровня $1-\alpha/2$). Таблица критических точек представлена в Приложении 4.

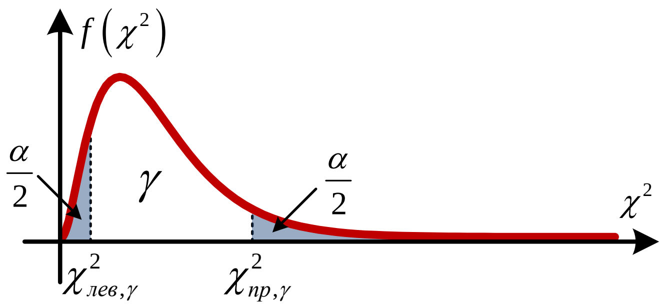


Рисунок 2.9 - К построению доверительных интервалов

На основании равенств

$$P\left(\chi^2_{лев,\gamma} < \frac{nD_в}{\sigma^2} < \chi^2_{пр,\gamma}\right) = \gamma \quad \text{или} \quad P\left(\frac{nD_в}{\chi^2_{лев,\gamma}} > \sigma^2 > \frac{nD_в}{\chi^2_{пр,\gamma}}\right) = \gamma, \tag{34}$$

интервал

$$\frac{nD_в}{\chi^2_{пр,\gamma}} < \sigma^2 < \frac{nD_в}{\chi^2_{лев,\gamma}} \quad \text{или} \quad \frac{n-1}{\chi^2_{пр,\gamma}} s^2 < \sigma^2 < \frac{n-1}{\chi^2_{лев,\gamma}} s^2 \tag{35}$$

является интервальной оценкой для σ^2 с надежностью γ.

Границы интервалов (35) являются случайными величинами и с вероятностью γ накрывают неизвестную дисперсию σ^2.

Пример 2.4. По выборке объема *n* = 20 из нормально распределенной генеральной совокупности вычислено значение выборочной дисперсии $D_в = 1{,}5$. Построить интервальную оценку для параметра σ^2 с надежностью γ = 0,96.

Решение. Значения $\chi^2_{лев,\gamma}$, $\chi^2_{пр,\gamma}$ находим из условий (33):

$$\alpha = 0{,}04; \quad P\left(\chi^2_{19} > \chi^2_{лев,\gamma}\right) = 0{,}98; \quad P\left(\chi^2_{19} > \chi^2_{пр,\gamma}\right) = 0{,}02.$$

Т.е. $\chi^2_{лев,\gamma}$ есть критическая точка χ^2-распределения с 19 степенями свободы уровня 0,98, а $\chi^2_{пр,\gamma}$ – критическая точка уровня 0,02.

По табл. Приложения 4 критических точек χ^2-распределения находим

$$\chi^2_{лев,\gamma} = 8{,}6; \qquad \chi^2_{пр,\gamma} = 33{,}7.$$

Тогда интервальная оценка σ^2 принимает вид (35)

$$\left(\frac{20}{33{,}7}D_в;\ \frac{20}{8{,}6}D_в\right) = \left(0{,}59\,D_в;\ 2{,}33\,D_в\right).$$

Подставляя вычисленное значение $D_в = 1{,}5$, получаем

$$0{,}89 < \sigma^2 < 3{,}488,$$

откуда оценка СКО $\quad 0{,}94 < \sigma < 1{,}868.$

Интервальная оценка вероятности события

Точечной оценкой вероятности *p* события является частность $w = m/n$, где *n* – общее число независимых испытаний, а *m* – число испытаний, в которых произошло событие *A*.

Зададимся надежностью интервальной оценки γ и найдем числа $p_{лев,\gamma}$, $p_{пр,\gamma}$ такие, чтобы выполнялось соотношение

$$P\left(p_{лев,\gamma} < p < p_{пр,\gamma}\right) = \gamma. \qquad (36)$$

Интервальная оценка вероятности при большом числе испытаний. Если $n > 30$ и $np > 10$, то распределение случайной величины $w = \frac{m}{n}$ можно аппроксимировать нормальным распределением $N\left(p, \sqrt{pq/n}\right)$. Следовательно, при этих же условиях распределение величины $\dfrac{(w-p)}{\sqrt{pq/n}}$ близко к нормальному с нулевым математическим ожиданием и единичной дисперсией, т.е.

$$\frac{w-p}{\sqrt{pq/n}} = N(0,1).$$

По аналогии с (27), найдем такое число z_γ, для которого справедливо равенство

$$P\left(-z_\gamma < \frac{w-p}{\sqrt{pq/n}} < z_\gamma\right) = \gamma. \tag{37}$$

Это число z_γ является корнем уравнения $\Phi(z_\gamma) = \gamma/2$, где $\Phi(z)$ – функция Лапласа, и корень может быть найден с помощью табл. П2.

Неравенство, стоящее в скобках выражения (37), разрешим относительно *p*. Для этого неравенство перепишем в виде эквивалентного неравенства $\left|\dfrac{w-p}{\sqrt{pq/n}}\right| < z$. Возведем в квадрат, в результате получим $(w-p)^2 < \dfrac{p(1-p)}{n} z^2$.

Далее, возведя в квадрат $(w-p)$ и перенеся все члены влево, получим

$$\left(1+\frac{z^2}{n}\right)p^2 - \left(2w+\frac{z^2}{n}\right)p + w^2 < 0.$$

Корни p_1 и p_2 этого квадратного трехчлена ***являются границами интервальной оценки*** (36) ***вероятности события*** и определяются выражениями

$$\begin{aligned} p_1 = p_{лев,\gamma} &= \frac{n}{z^2+n}\left[w + \frac{z^2}{2n} - z\sqrt{\frac{w(1-w)}{n} + \left(\frac{z}{2n}\right)^2}\,\right], \\ p_2 = p_{пр,\gamma} &= \frac{n}{z^2+n}\left[w + \frac{z^2}{2n} + z\sqrt{\frac{w(1-w)}{n} + \left(\frac{z}{2n}\right)^2}\,\right], \end{aligned} \tag{38}$$

где *n* — общее число испытаний; *m* — число появлений события; *w* — относительная частота; z — значение аргумента функции Лапласа (приложение 2), при котором $\Phi(z) = \dfrac{\gamma}{2}$ (γ — заданная надежность).

Если $n >> 100$, то в формулах (38) слагаемым $\frac{z}{2n}$ можно пренебречь, тогда для вычисления p_1, p_2 можно использовать приближенные формулы, аналогичные (28):

$$p_1 = w - z\sqrt{\frac{w(1-w)}{n}},\ p_2 = w + z\sqrt{\frac{w(1-w)}{n}}. \tag{39}$$

Пример 2.5. Событие A в серии из $n = 100$ испытаний произошло $m = 78$ раз. Построить интервальную оценку для вероятности p события с надежностью $\gamma = 0,9$.

Решение. Значение точечной оценки вероятности p равно $w = 78/100 = 0,78$. По табл. П2 определяем для $\Phi(z) = 0,9/2 = 0,45$ $z = 1,64$ и вычисляем по формулам (38) значения p_1, p_2 при $w = 0,78$: $p_1 = 0,705,\quad p_2 = 0,848$. Таким образом, доверительный интервал для вероятности p события A следующий: (0,705; 0,848).

Интервальная оценка вероятности при малом числе испытаний (n<30). При малом числе испытаний n предположение о приближенном распределении случайной величины m по нормальному закону становится несправедливым. Для описания распределения величины m необходимо использовать формулу Бернулли:

$$P(m = x) = C_n^x p^x (1-p)^{n-x}, \quad x = 0,1,...,n.$$

Можно показать, что граничные точки интервальной оценки (36) являются решениями следующих нелинейных уравнений:

$$\sum_{x=0}^{m-1} C_n^x p_{лев,\gamma}^x (1-p_{лев,\gamma})^{n-x} = \frac{1+\gamma}{2}; \qquad \sum_{x=0}^{m} C_n^x p_{пр,\gamma}^x (1-p_{пр,\gamma})^{n-x} = \frac{1-\gamma}{2},$$

где γ – надежность интервальной оценки.

Корни этих уравнений могут быть найдены одним из численных методов решения нелинейных уравнений. Кроме этого, существуют специальные таблицы для нахождения $p_{лев,\gamma}, p_{пр,\gamma}$. В данном пособии они не приводятся.

2.6. Вычисление границ доверительных интервалов в Excel

Вычисление величины z_γ, входящей в доверительный интервал $P\left(|Z| < z_\gamma\right)$ оценки математического ожидания (27).

В Excel реализована не функция Лапласа (см. Прил.2), а интегральная функции стандартного нормального распределения $F(z) = P\left(Z < z\right)$, где $\left(-5 < z < 5\right)$. Поэтому величина z_γ (27) вычисляется с помощью функции НОРМСТОБР(p) следующим образом:

$z_\gamma = \text{НОРМСТОБР}\left(1-\alpha/2\right)$, где γ – надежность интервальной оценки.

Вычисление величины $\delta_\gamma = z_\gamma \dfrac{\sigma}{\sqrt{n}}$ (28) осуществляется с помощью функции ДОВЕРИТ:

$$\delta_\gamma = z_\gamma \frac{\sigma}{\sqrt{n}} = \text{ДОВЕРИТ}(\alpha; \sigma; n),$$

где $\alpha = 1-\gamma$, σ – известное СКО, n – объем выборки.

Вычисление $t_\gamma(\gamma, k)$ **критической точки распределения Стьюдента** (32), определяемой выражением $P\left(|T_k| < t\left(\gamma, k\right)\right) = \gamma$, осуществляют с использованием функции СТЬЮДРАСПОБР, обращение к которой имеет вид:

$$t_\gamma(\gamma, k) = \text{СТЬЮДРАСПОБР}(\alpha; k),$$

где $\alpha = 1-\gamma$, $k = n-1$ – *число степеней свободы*.

Вычисление критических точек распределения Пирсона $\chi^2_{лев,\gamma}$, $\chi^2_{пр,\gamma}$, входящих в доверительный интервал (35), для дисперсии σ^2 выполняется с использованием функции ХИ2ОБР:

$$\chi^2_{лев,\gamma} = \text{ХИ2ОБР}\left(1-\alpha/2; k\right); \qquad \chi^2_{пр,\gamma} = \text{ХИ2ОБР}\left(\alpha/2; k\right),$$

где $\alpha = 1-\gamma$, γ – надежность интервальной оценки.

ЗАДАЧИ

ЗАДАЧА 1. Найти доверительный интервал для оценки с надежностью 0,95 неизвестного математического ожидания *a* нормально распределенного признака *X* генеральной совокупности, если генеральное СКО $\sigma = 5$, выборочная средняя $\overline{x_B} = 14$ и объем выборки $n = 25$.

ЗАДАЧА 2. Найти доверительный интервал для оценки с надежностью 0,99 неизвестного математического ожидания a нормально распределенного признака X генеральной совокупности, если известны генеральное среднее квадратическое отклонение σ, выборочная средняя $\overline{x_B}$ и объем выборки n: а) $\sigma = 4$, $\overline{x_B} = 10,2$, $n = 16$; б) $\sigma = 5$, $\overline{x_B} = 16,8$, $n = 25$.

ЗАДАЧА 3. Найти минимальный объем выборки, при котором с надежностью 0,975 точность оценки математического ожидания a генеральной совокупности по выборочной средней равна $\delta = 0,3$, если известно СКО $\sigma = 1,2$ нормально распределенной генеральной совокупности.

ЗАДАЧА 4. Найти минимальный объем выборки, при котором с надежностью 0,925 точность оценки математического ожидания нормально распределенной генеральной совокупности по выборочной средней равна 0,2, если известно среднее квадратическое отклонение генеральной совокупности $\sigma = 1{,}5$.

ЗАДАЧА 5. Из генеральной совокупности извлечена выборка объема $n = 10$:

варианта	x_i	- 2	1	2	3	4	5
частота	n_i	2	1	2	2	2	1

Оценить с надежностью 0,95 математическое ожидание *a* нормально распределенного признака генеральной совокупности по выборочной средней при помощи доверительного интервала.

ЗАДАЧА 6. Из генеральной совокупности извлечена выборка объема $n = 12$:

варианта x_i	-0,5	-0,4	-0,2	0	0,2	0,6	0,8	1	1,2	1,5
частота n_i	1	2	1	1	1	1	1	1	2	1

Оценить с надежностью 0,95 математическое ожидание a нормально распределенного признака генеральной совокупности с помощью доверительного интервала.

ЗАДАЧА 7. По данным девяти независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\overline{x_B} = 30{,}1$ и «исправленное» среднее квадратическое отклонение $s = 6$. Оценить истинное значение измеряемой величины с помощью доверительного интервала c надежностью $\gamma = 0{,}99$. Предполагается, что результаты измерений распределены нормально.

ЗАДАЧА 8. По данным выборки объема $n=16$ из генеральной совокупности найдено «исправленное» среднее квадратическое отклонение $s=1$ нормально распределенного количественного признака. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,95.

ЗАДАЧА 9. Произведено 12 измерений одним прибором (без систематической ошибки) некоторой физической величины, причем «исправленное» среднее квадратическое отклонение s случайных ошибок измерений оказалось равным 0,6. Найти точность прибора с надежностью 0,98. Предполагается, что результаты измерений распределены нормально.

ЗАДАЧА 10. При испытаниях 1000 элементов зарегистрировано 100 отказов. Найти доверительный интервал, покрывающий неизвестную вероятность p отказа элемента с надежностью: а) 0,95; б) 0,99.

3. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

3.1. Линейная корреляция

Если в результате осуществления некоторого эксперимента наблюдаются две величины X и Y, то *выборочный корреляционный момент (ковариация)* $\mu^*_{x,y}$ величин X и Y определяется формулой:

$$\mu^*_{x,y} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \overline{xy} - \overline{x}\cdot\overline{y} \tag{40}$$

где $(x_1, y_1),\ldots,(x_n, y_n)$ — n пар наблюденных значений, полученных в n независимых повторениях эксперимента, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n}y_i$.

Выборочный коэффициент корреляции r^*_B:

$$r^*_B = \frac{\mu^*_{x,y}}{\sigma^*_x \cdot \sigma^*_y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}, \tag{41}$$

где σ^*_x и σ^*_y - выборочные СКО

$$\sigma^*_x = \sqrt{D^*(X)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}, \quad \sigma^*_y = \sqrt{D^*(Y)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}.$$

Выборочные коэффициенты регрессий

Y *на* X *:*
$$\rho^*_{y/x} = r^*_B\frac{\sigma^*_y}{\sigma^*_x} = \frac{\mu^*_{x,y}}{(\sigma^*_x)^2}; \tag{42}$$

X *на* Y *:*
$$\rho^*_{x/y} = r^*_B\frac{\sigma^*_x}{\sigma^*_y} = \frac{\mu^*_{x,y}}{(\sigma^*_y)^2}. \tag{43}$$

Выборочные уравнения линейных регрессий (среднеквадратических)

Y *на* X :
$$y - \overline{y} = \rho^*_{y/x}(x - \overline{x}) \quad \text{или} \quad \frac{y - \overline{y}}{\sigma^*_y} = r^*_B\frac{x - \overline{x}}{\sigma^*_x}, \tag{44}$$

X *на* Y :
$$\frac{x - \overline{x}}{\sigma^*_x} = r^*_B\frac{y - \overline{y}}{\sigma^*_y} \quad \text{или} \quad \frac{y - \overline{y}}{\sigma^*_y} = \frac{1}{r^*_B}\frac{x - \overline{x}}{\sigma^*_x}. \tag{45}$$

Пример 3.1. Для выборки двумерной случайной величины

Таблица 3.1

i	1	2	3	4	5	6	7	8	9	10
x_i	1,2	1,5	1,8	2,1	2,3	3	3,6	4,2	5,7	6,3
y_i	4,6	5,8	7,3	10,4	12,30	14,4	14,9	14,8	15,2	16,5

вычислить выборочные средние наблюдаемых признаков, выборочные средние квадратические отклонения, выборочный коэффициент корреляции и составить выборочные уравнения линейных регрессий. Представить корреляционное поле и линейные регрессии на графике.

Решение.

Удобно исходную таблицу дополнить строкой $x_i\, y_i$ *и столбцом* ср.знч.

Таблица 3.2

i	1	2	3	4	5	6	7	8	9	10	ср.знч
x_i	1,2	1,5	1,8	2,1	2,3	3	3,6	4,2	5,7	6,3	3,17
y_i	4,6	5,8	7,3	10,4	12,30	14,4	14,9	14,8	15,2	16,5	11,62
$x_i\, y_i$	5,52	8,7	13,14	21,84	28,29	43,2	53,64	62,16	86,64	103,95	42,71

Тогда искомые выборочные параметры распределения

$$\overline{x}_B = \frac{1,2+1,5+...+6,3}{10} = 3,17; \;\; \overline{y}_B = \frac{4,6+5,8+..+16,5}{10} = 11,62.$$

$$D_B(X) = 0,1(1,2^2+1,5^2+...+6,3^2) - 3,17^2 = 2,79; \;\; \sigma_X = \sqrt{2,79} = 1,67.$$

$$D_B(Y) = 0,1(4,6^2+5,8^2+...+16,5^2) - 11,62^2 = 16,90; \;\; \sigma_Y = \sqrt{16,90} = 4,11.$$

$$\frac{1}{10}\sum_{i=1}^{10} x_i y_i = \frac{1}{10}\left(1,2\cdot 4,6 + 1,5\cdot 5,8 + ... + 6,3\cdot 16,5\right) = 42,71.$$

$$\mu^*_{x,y} = 42,71 - 3,17\cdot 11,62 = 5,87; \qquad r^*_B = \frac{42,71 - 3,17\cdot 11,62}{1,67\cdot 4,11} = 0,85.$$

Выборочное уравнение линейной регрессии Y/X имеет вид (44):

$$y - 11,62 = 0,85\cdot\frac{4,11}{1,67}(x - 3,17) \;\; \text{или} \;\; y = 2,1x \; + \; 4,95.$$

Выборочное уравнение линейной регрессии X/ Y имеет вид (45):

$$y - 11,62 = \frac{1}{0,85}\cdot\frac{4,11}{1,67}(x - 3,17) \;\; \text{или} \;\; y = 2,88x \; + \; 2,45.$$

Поскольку линии регрессии прямые, то строим их по двум крайним точкам

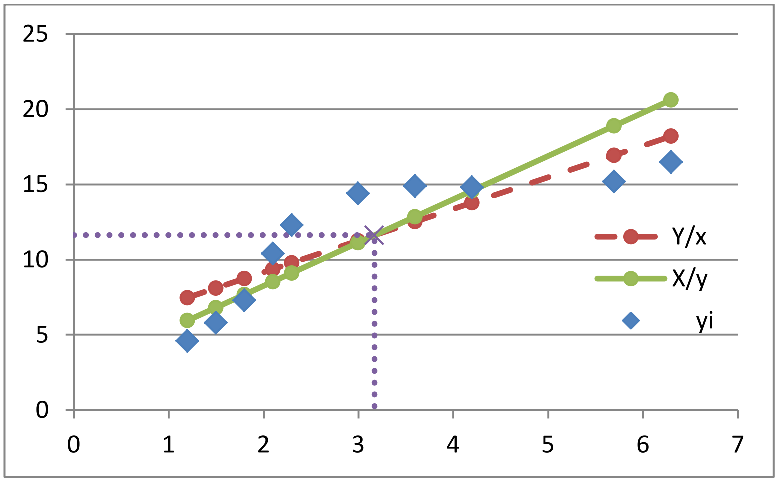


Рисунок 3.1 - Линейные регрессии

Для вычисления корреляционного момента (ковариации) и коэффициента корреляции в табличном процессоре ***Excel*** используются функции ***КОВАР(массив1;массив2)*** и ***КОРРЕЛ(массив1;массив2)*** соответственно.

3.2. Нелинейная корреляция

При совместном исследовании двух случайных величин по имеющейся выборке (x_1, y_2), (x_2, y_2),…,(x_k, y_k) возникает задача определения "наиболее подходящей", в том числе и нелинейной, зависимости между ними. Если вид функции $y = f(x,a,b,...)$ задан, то требуется найти такие значения коэффициентов a, b,..., при которых y_i "наименее" отличаются от $f(x_i)$. В качестве критерия оптимальности часто используют минимум суммы квадратов ошибок

$$\sum_{i=1}^{k}\left(y_i - f\left(x_i\right)\right)^2 \to \min. \qquad (46)$$

Коэффициенты a, b,... функции $y = f(x,a,b,...)$ и сама функция, обеспечивающие минимум критерия (46) называются оптимальными в смысле метода наименьших квадратов для этого класса функций. При этом, для другого класса функций, например $y = g(x,a,b,...)$, критерий (46) может принять значение еще более близкое к нулю. В этом случае функция $y = g(x,a,b,...)$, очевидно, лучше описывает статистическую взаимосвязь наблюдаемых признаков, нежели функция $y = f(x,a,b,...)$.

В качестве оценки тесноты нелинейной корреляционной связи используют коэффициент детерминации R^2 (корреляционное отношение), который показывает долю объясненной изменением факторного признака дисперсии от общей дисперсии:

$$R^2 = \frac{\sigma_e^2}{\sigma^2} = 1 - \frac{\sigma_u^2}{\sigma^2}, \tag{47}$$

где $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\Delta y_i^{\,2}$ - общая (total) дисперсия результативного признака y_i - относительно общей средней $\bar{y}$.

$\sigma_e^2 = \frac{1}{n}\sum_{i=1}^{n}\left(f\left(x_i\right) - \bar{y}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}e_i^2$ - объясненная (межгрупповая, explained) дисперсия - дисперсия точек линии регрессии $f\left(x_i\right)$ относительно общей средней $\bar{y}$;

$\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f\left(x_i\right)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}u_i^{\,2}$- остаточная (внутригрупповая, unexplained) дисперсия - дисперсия признака y_i относительно модельной линии регрессии $f\left(x_i\right)$.

Смысл коэффициента детерминации поясняется на рисунке 3.1.

Здесь $y_i - \bar{y} = \Delta y_i = \left(y_i - y_x\right) + \left(y_x - \bar{y}\right) = u_i + e_i$.

Доказывается, что $\sum \Delta y_i = \sum u_i + \sum e_i$, т.к. $2\sum u_i e_i = 0$.

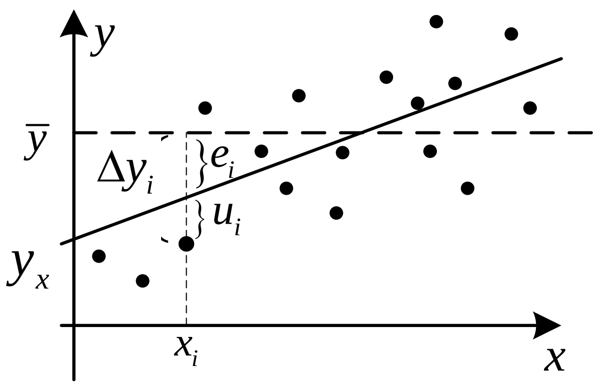


Рисунок 3.1 -
К пояснению суммы дисперсий

Поэтому имеет место *правило сложения дисперсий* - общая дисперсия σ^2 результативного признака равна сумме объясненной (межгрупповой) σ_e^2 и остаточной (необъясненной, внутригрупповой) σ_u^2 дисперсии:

$$\sigma^2 = \sigma_e^2 + \sigma_u^2 \tag{48}$$

Из выражения (47) следует пределы изменения значений R^2:

$$0 \le R^2 \le 1.$$

Рассмотрим методику определения оптимальных по методу наименьших квадратов (МНК) параметров для некоторых классов функций (модельных регрессий).

а) Пусть модельной линией регрессии является линейная зависимость

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

В качестве критерия тесноты корреляционной связи используем сумму квадратов отклонений значений признака y_i относительно линии регрессии $y(x_i)$

$$F(a_0, a_1) = \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x) \right)^2 \to \min.$$

Определим неизвестные параметры регрессии (a_0, a_1) из условия минимума $F(a_0, a_1)$: $F'_{a_0} = F'_{a_1} = 0$.

$$F'_{a_0} = 2\sum \left(y_i - a_0 - a_1 x \right) \cdot (-1) = 0; \qquad \sum \left(-y_i + a_0 + a_1 x \right) = 0;$$

$$\sum y_i = n \cdot a_0 + a_1 \sum x_i\ ; \qquad \overline{y} = a_0 + \overline{x} \cdot a_1.$$

$$F'_{a_1} = 2\sum \left(y_i - a_0 - a_1 x \right) \cdot (-x_i) = 0; \qquad \sum \left(-x_i y_i + a_0 x_i + a_1 x_i^2 \right) = 0;$$

$$\sum x_i y_i = a_1 \sum x_i^2 + a_0 \sum x_i\ ; \qquad \overline{xy} = \overline{x} \cdot a_0 + \overline{x^2} \cdot a_1.$$

$$\begin{cases} \overline{y} = a_0 + \overline{x} \cdot a_1 \\ \overline{xy} = \overline{x} \cdot a_0 + \overline{x^2} \cdot a_1 \end{cases} \quad \text{или} \quad \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}. \tag{49}$$

.

б) Квадратичная зависимость $y(x) = a_0 + a_1 x + a_2 x^2$. Приравнивая частные производные суммы квадратов ошибок к нулю, по аналогии с пунктом а) получим:

$$\begin{pmatrix} \overline{y} \\ \overline{xy} \\ \overline{x^2 y} \end{pmatrix} = \begin{pmatrix} 1 & \overline{x} & \overline{x^2} \\ \overline{x} & \overline{x^2} & \overline{x^3} \\ \overline{x^2} & \overline{x^3} & \overline{x^4} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}. \tag{50}$$

в) Логарифмическая зависимость $y(x)=a_0+a_1\ln x$. Вводя новую переменную $z=\ln x$ приводим логарифмическую зависимость к линейной. Тогда на основании выражений (49) получим:

$$\begin{cases} \overline{y}=a_0+\overline{z}\cdot a_1 \\ \overline{zy}=\overline{z}\cdot a_0+\overline{z^2}\cdot a_1 \end{cases} \quad \text{или} \quad \begin{pmatrix} \overline{y} \\ \overline{zy} \end{pmatrix}=\begin{pmatrix} 1 & \overline{z} \\ \overline{z} & \overline{z^2} \end{pmatrix}\begin{pmatrix} a_0 \\ a_1 \end{pmatrix}. \qquad (51)$$

Пример 3.2. Найти параметры и построить линии регрессий y от x для выборки из предыдущего примера 3.1 (табл.3.2) для случаев:

1) линейной зависимости $y(x)=a_0+a_1x$;

2) квадратичной зависимости $y(x)=a_0+a_1x+a_2x^2$;

3) логарифмической зависимости $y(x)=a_0+a_1\ln x$.

Решение.

В соответствие с (49)- (51) дополним исходную таблицу необходимыми строками и последним столбцом - ср.знч. После заполнения таблица примет вид (вычисления проводились в MS Excel):

Таблица 3.3

i	1	2	3	4	5	6	7	8	9	10	ср.знач
x_i	1,2	1,5	1,8	2,1	2,3	3	3,6	4,2	5,7	6,3	3,17
y_i	4,6	5,8	7,3	10,4	12,30	14,4	14,9	14,8	15,2	16,5	11,62
xy	5,52	8,70	13,14	21,84	28,29	43,20	53,64	62,16	86,64	103,95	42,71
(x^2)y	6,62	13,05	23,65	45,86	65,07	129,60	193,10	261,07	493,85	654,89	188,68
x^2	1,44	2,25	3,24	4,41	5,29	9,00	12,96	17,64	32,49	39,69	12,84
x^3	1,73	3,38	5,83	9,26	12,17	27,00	46,66	74,09	185,19	250,05	61,53
x^4	2,07	5,06	10,50	19,45	27,98	81,00	167,96	311,17	1055,60	1575,30	325,61
z=lnx	0,18	0,41	0,59	0,74	0,83	1,10	1,28	1,44	1,74	1,84	1,01
z^2	0,03	0,16	0,35	0,55	0,69	1,21	1,64	2,06	3,03	3,39	1,31
zy	0,84	2,35	4,29	7,72	10,24	15,82	19,09	21,24	26,46	30,37	13,84
Yлин	7,48	8,11	8,74	9,37	9,79	11,26	12,52	13,79	16,94	18,20	11,62
Yкв	4,79	6,59	8,23	9,73	10,65	13,36	15,04	16,13	16,31	15,36	11,62
Yлог	5,56	7,18	8,51	9,63	10,30	12,23	13,56	14,68	16,91	17,63	11,62

Искомые векторы неизвестных параметров определим из систем линейных уравнений (49)- (51), например методом Крамера или матричным методом, с использованием последнего столбца таблицы 3.3.

В результате получим уравнения регрессий:

1) $\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 1 & 3{,}17 \\ 3{,}17 & 12{,}84 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 11{,}62 \\ 42{,}71 \end{pmatrix} = \begin{pmatrix} 4{,}95 \\ 2{,}10 \end{pmatrix}$; линейной $y(x) = 4{,}95 + 2{,}10x$;

2) $\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & 3{,}17 & 12{,}84 \\ 3{,}17 & 12{,}84 & 61{,}53 \\ 12{,}84 & 61{,}53 & 325{,}61 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 11{,}62 \\ 42{,}71 \\ 188{,}68 \end{pmatrix} = \begin{pmatrix} -3{,}84 \\ 8{,}17 \\ -0{,}81 \end{pmatrix}$;

квадратичной $y(x) = -3{,}84 + 8{,}17x - 0{,}81x^2$;

3) $\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 1 & 1{,}01 \\ 1{,}01 & 1{,}31 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 11{,}62 \\ 13{,}84 \end{pmatrix} = \begin{pmatrix} 4{,}23 \\ 7{,}28 \end{pmatrix}$;

логарифмической $y(x) = 4{,}23 + 7{,}28 \ln x$.

Построим графики всех полученных уравнений регрессии по точкам x_i (последние строки в таблице 3.3), как показано на рисунке 3.3

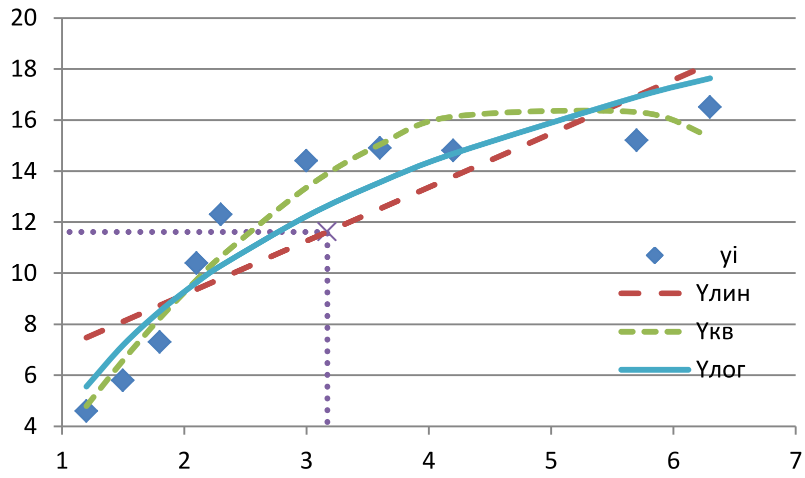


Рисунок 3.3 - Корреляционное поле и линии регрессий

Пример 3.3. Определите общую, объясненную и остаточную дисперсии для рассмотренных уравнений регрессий, а также коэффициенты детерминации.

Решение.

Вычислим общую, объясненную и остаточную дисперсии для рассмотренных уравнений регрессий с помощью MS Excel. Для этого составим следующую таблицу:

Таблица 3.4

i	1	2	3	4	5	6	7	8	9	10	ср.знач
y_i	4,6	5,8	7,3	10,4	12,30	14,4	14,9	14,8	15,2	16,5	11,62
Yлин$= y_{x1}$	7,48	8,11	8,74	9,37	9,79	11,26	12,52	13,79	16,94	18,20	11,62
Yкв$= y_{x2}$	4,79	6,59	8,23	9,73	10,65	13,36	15,04	16,13	16,31	15,36	11,62
Yлог$= y_{x3}$	5,56	7,18	8,51	9,63	10,30	12,23	13,56	14,68	16,91	17,63	11,62
$\Delta y_i = y_i - \bar{y}$	-7,02	-5,82	-4,32	-1,22	0,68	2,78	3,28	3,18	3,58	4,88	0,00
$\Delta y_i^{\ 2}$	49,280	33,872	18,662	1,488	0,462	7,728	10,758	10,112	12,816	23,814	16,90
$e_1 = y_{x1} - \bar{y}$	-4,14	-3,51	-2,88	-2,25	-1,83	-0,36	0,90	2,17	5,32	6,58	0,00
$e_2 = y_{x2} - \bar{y}$	-6,83	-5,03	-3,39	-1,89	-0,97	1,74	3,42	4,51	4,69	3,74	0,00
$e_3 = y_{x3} - \bar{y}$	-6,06	-4,44	-3,11	-1,99	-1,32	0,61	1,94	3,06	5,29	6,01	0,00
$e_1^{\ 2}$	17,17	12,34	8,30	5,06	3,35	0,13	0,82	4,69	28,32	43,34	12,35
$e_2^{\ 2}$	46,58	25,32	11,46	3,55	0,93	3,02	11,68	20,38	22,00	13,95	15,89
$e_3^{\ 2}$	36,73	19,68	9,66	3,94	1,75	0,37	3,76	9,38	27,94	36,17	14,94
$u_1 = y_i - y_{x1}$	-2,88	-2,31	-1,44	1,03	2,51	3,14	2,38	1,01	-1,74	-1,70	0,00
$u_2 = y_i - y_{x2}$	-0,19	-0,79	-0,93	0,67	1,65	1,04	-0,14	-1,33	-1,11	1,14	0,00
$u_3 = y_i - y_{x3}$	-0,96	-1,38	-1,21	0,77	2,00	2,17	1,34	0,12	-1,71	-1,13	0,00
$u_1^{\ 2}$	8,27	5,32	2,07	1,06	6,30	9,84	5,64	1,03	3,03	2,90	4,55
$u_2^{\ 2}$	0,04	0,62	0,87	0,44	2,71	1,09	0,02	1,78	1,23	1,31	1,01
$u_3^{\ 2}$	0,92	1,92	1,47	0,59	4,01	4,70	1,80	0,01	2,91	1,29	1,96

Легко проверить выполнение правила сложения дисперсий (48):

$$\sigma^2_{лин} = \sigma^2_{e1} + \sigma^2_{u1} = 12,35 + 4,55 = 16,90; \qquad \sigma^2_{кв} = \sigma^2_{e2} + \sigma^2_{u2} = 15,89 + 1,01 = 16,90;$$

$$\sigma^2_{лог} = \sigma^2_{e3} + \sigma^2_{u3} = 14,94 + 1,96 = 16,90.$$

Коэффициенты детерминации примут значения (47):

$$R^2_{лин} = \frac{\sigma^2_{e1}}{\sigma^2} = \frac{12,35}{16,90} = 0,73; \qquad R^2_{кв} = \frac{\sigma^2_{e2}}{\sigma^2} = \frac{15,89}{16,90} = 0,94;$$

$$R^2_{лог} = \frac{\sigma^2_{e3}}{\sigma^2} = \frac{14,94}{16,90} = 0,88, \quad \text{где } \sigma^2 = \sigma^2_{лин} = \sigma^2_{кв} = \sigma^2_{лог} = 16,90.$$

Максимальный коэффициент детерминации соответствует квадратичной регрессии, значит она и является наиболее удачной из рассмотренных.

4. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Под *статистической гипотезой* понимается всякое высказывание о генеральной совокупности (случайной величине X), проверяемое по выборочной совокупности (по результатам наблюдений).

Для проверки гипотезы определяют специальную величину (статистику) K с известным законом распределения, как функцию выборочных данных $K = K(x_1, x_2, ..., x_n)$. Величину K называют *статистическим критерием.* Полагая, что распределение признака X подчинено гипотезе H_0, рассчитывают область возможных значений (с заданной вероятностью) статистики K. Эта *область принятия основной гипотезы*. *Критической называют область* практически невозможных (с вероятностью α) значений статистики K, если верна гипотеза H_0.

По имеющимся наблюдениям рассчитывают выборочное значение K. Если оно принадлежит критической области, то основную гипотезу отвергают.

Решение о том, можно ли считать гипотезу H_0 верной для генеральной совокупности, принимается по выборочным данным, т.е. по ограниченному объему информации. Следовательно, это решение может быть ошибочным. При этом может иметь место ошибка двух родов (рисунок 4.1):

— *ошибка первого рода* совершается при отклонении верной гипотезы H_0 . Вероятность такой ошибки называют *уровнем значимости критерия* $\alpha = P(H_1 / H_0)$;

— *ошибка второго рода* совершается при принятии неверной гипотезы H_1. Вероятность ошибки второго рода обозначим как $\beta = P(H_0 / H_1)$.

Значение $1 - \beta = P(H_1 / H_1)$ - вероятность отвергнуть основную гипотезу когда она неверна, называют *мощностью критерия.*

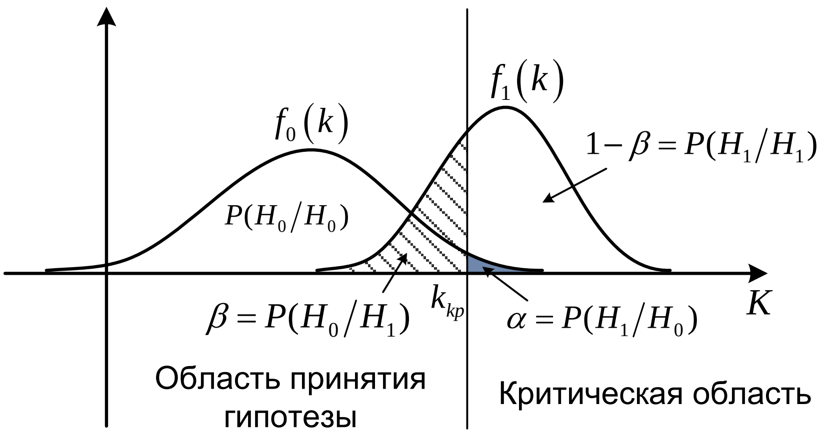
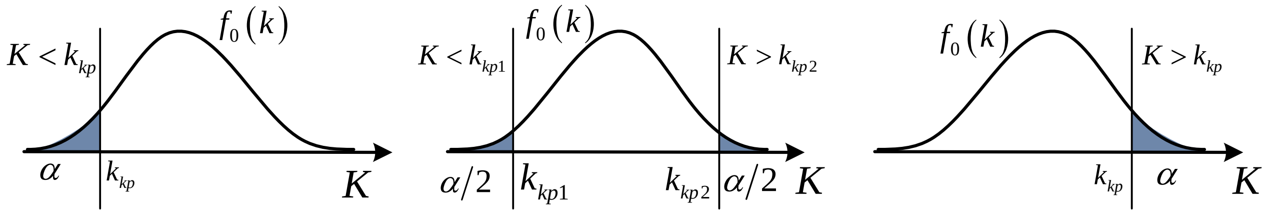


Рисунок 4.1 - К проверке гипотез

Критическую область следует выбирать из условия обеспечения максимума мощности критерия при заданном уровне значимости. Критическая область может быть *левосторонней,* если она задается неравенством ($K < k_{кр}$), *двусторонней* $\left(k_{кр1} > K\right) \cup \left(K > k_{кр2}\right)$ или *правосторонней* ($K > k_{кр}$) (рис. 4.2).



Левосторонняя Двусторонняя Правосторонняя

Рисунок 4.2 - Критические области

4.1. Проверка гипотез о числовых значениях

Проверка гипотезы о равенстве математического ожидания нормального распределения заданному значению

Проверка гипотезы о числовом значении МО ***при известной дисперсии****.*

Предполагается, что $X = N(a,\sigma)$, причем значение математического ожидания a неизвестно, а числовое значение дисперсии σ^2 известно.

По известной выборочной средней $\bar{x}$ необходимо проверить при уровне значимости α нулевую гипотезу $H_0 : a = a_0$ о том, что математическое ожидание генеральной совокупности $\boldsymbol{a}$ равно предполагаемому значению a_0

Если дисперсия генеральной совокупности известна (или неизвестна, но выборка достаточно большая $n > 30$), то в качестве критерия *K* проверки нулевой гипотезы принимают *z-статистику* - нормально распределенную случайную величину с параметрами $N\left(0;1\right)$.

Вычислим наблюдаемое значение *z-критерия*:

$$Z_{набл} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} = N\left(0;1\right).$$

Критическую точку $z_{кр}$ определим в зависимости от альтернативной гипотезы (типа критической области) с использованием функции Лапласа (Приложение 2).

Правило 1. При конкурирующей гипотезе $H_1: a \neq a_0$, *найдем критическую точку* $z_{кр}$ *двусторонней критической области из равенства*

$$\Phi(z_{кр}) = \frac{1}{2} - \frac{\alpha}{2} = \frac{(1-\alpha)}{2}. \qquad (52)$$

Если $|Z_{набл}| < z_{кр}$ нет оснований отвергнуть нулевую гипотезу. Если $|Z_{набл}| > z_{кр}$ - нулевую гипотезу отвергают.

Правило 2. При конкурирующей гипотезе $H_1: a > a_0$ критическую точку правосторонней критической области находят из равенства

$$\Phi(z_{кр}) = \frac{1}{2} - \alpha = \frac{(1-2\alpha)}{2}. \qquad (53)$$

Если $Z_{набл} < z_{кр}$ нет оснований отвергнуть нулевую гипотезу. Если $Z_{набл} > z_{кр}$ - нулевую гипотезу отвергают.

Правило 3. При конкурирующей гипотезе $H_1: a < a_0$ сначала находят вспомогательную критическую точку $z_{кр}$, как в пункте 2), а затем полагают границу левосторонней критической области $z_{левост.кр} = -z_{кр}$.

Если $Z_{набл} > -z_{кр}$ - нет оснований отвергнуть нулевую гипотезу. Если $Z_{набл} < -z_{кр}$ - нулевую гипотезу отвергают.

Пример 574.

Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 5,2$ извлечена выборка объема $n = 100$ и по ней найдена выборочная средняя $\overline{x} = 27,56$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: a = a_0 = 26$ при конкурирующей гипотезе $H_1: a \neq 26$.

Решение. Найдем наблюдаемое значение критерия:

$$z_{набл} = \frac{(\overline{x} - a_0)\sqrt{n}}{\sigma} = \frac{(27,56 - 26)\sqrt{100}}{5,2} = 3.$$

По условию, конкурирующая гипотеза имеет вид $a \neq a_0$, поэтому критическая область — двусторонняя. Найдем критическую точку из равенства (52)

$$\Phi(z_{кр}) = \frac{(1-\alpha)}{2} = \frac{(1-0,05)}{2} = 0,475.$$

По таблице функции Лапласа (см. приложение 2) находим $z_{кр} = 1,96$.

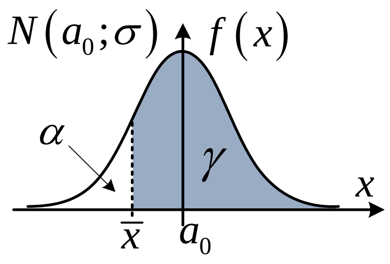
Так как $Z_{набл} > z_{кр}$ — нулевую гипотезу отвергаем. Другими словами, математическое ожидание генеральной совокупности ***a*** значимо отличается от гипотетического значения $a_0 = 26$.

В табличном процессоре Excel для проверки рассмотренной выше гипотезы используется функция ZTECT с обязательным указанием СКО. Обращение к ней имеет вид:

=ZTECT(*массив*; a_0; [σ]),

где *массив* – адреса ячеек, содержащих выборочные данные;

a_0 и σ – имеют прежний смысл, при этом, если параметр σ опущен, то используется исправленная выборочная дисперсия *s*, вычисленная по той же выборке.



Функция ZTECT численно равна вероятности того, что $P\left(X_0 > \bar{x}\right)$,

где $X_0 = N\left(a_0;\sigma\right)$ - гипотетическая случайная величина.

$$ZTECT(массив; a_0; [\sigma]) = P\left(X_0 > \bar{x}\right) = \gamma.$$

Проверка гипотезы о числовом значении МО ***при неизвестной дисперсии****.*

Предполагается, что $X = N(a,\sigma)$, значения математического ожидания a и дисперсии σ^2 неизвестны.

Если дисперсия генеральной совокупности неизвестна (например, в случае малых выборок), то в качестве критерия проверки нулевой гипотезы принимают *t-статистику* - случайную величину

$$K = T_{n-1} = \frac{\left(\bar{X}_в - a_0\right)\sqrt{n}}{s},$$

которая имеет распределение Стьюдента с числом степеней свободы $k = n - 1$.

Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0 : a = a_0$ о равенстве неизвестной генеральной средней ***a*** гипотетическому значению a_0, вначале надо вычислить наблюдаемое значение критерия

$$T_{набл} = \frac{\left(\bar{x} - a_0\right)\sqrt{n}}{s}. \quad (54)$$

Затем, в зависимости от конкурирующей гипотезы определить критическую точку указанного распределения и критерий принятия гипотезы.

Правило 1. При конкурирующей гипотезе $H_1: a \neq a_0$, по заданному уровню значимости α, помещенному ***в верхней строке таблицы*** (прилож.5), и числу степеней свободы $k = n-1$ найти критическую точку $t_{двуст.кр}(\alpha, k)$.

Если $|T_{набл}| < t_{двуст.кр}(\alpha, k)$ — нет оснований отвергнуть нулевую гипотезу.

Если $|T_{набл}| > t_{двуст.кр}(\alpha, k)$ — нулевую гипотезу отвергают.

Правило 2. При конкурирующей гипотезе $H_1: a > a_0$ по уровню значимости α, помещенному ***в нижней строке таблицы*** (в 2 раза меньше) приложения 5, и числу степеней свободы $k = n-1$ находят критическую точку $t_{правст.кр}(\alpha, k)$ правосторонней критической области.

Если $T_{набл} < t_{правст.кр}(\alpha, k)$ — нет оснований отвергнуть нулевую гипотезу.

Если $T_{набл} > t_{правст.кр}(\alpha, k)$ — нулевую гипотезу отвергают.

Правило 3. При конкурирующей гипотезе $H_1: a < a_0$ сначала находят «вспомогательную» критическую точку (по правилу 2) $t_{правст.кр}(\alpha, k)$ и полагают границу левосторонней критической области $t_{левост.кр}(\alpha, k) = -t_{правст.кр}(\alpha, k)$.

Если $T_{набл} > -t_{правст.кр}(\alpha, k)$ — нет оснований отвергнуть нулевую гипотезу.

Если $T_{набл} < -t_{правст.кр}(\alpha, k)$ — нулевую гипотезу отвергают.

Примечание.

Для проверки указанной гипотезы можно применять критерий Стьюдента (Приложение 3), использовавшийся для интервального оценивания.

В этом случае

$$P\left(|T_{n-1}| < t_\gamma\right) = \gamma; \qquad P\left(|T_{n-1}| > t_\gamma\right) = 1-\gamma = \alpha. \tag{55}$$

Поэтому определим вероятность γ следующим образом:

для двусторонней критической области $1-\gamma = \alpha \quad \to \quad \gamma = 1-\alpha$.

для односторонней критической области $1-\gamma = 2\alpha \quad \to \quad \gamma = 1-2\alpha$;

1) Гипотеза $H_1: a \neq a_0$ - критическая область двусторонняя $t_{kp} = t_\gamma(1-\alpha, k)$.

При $|T_{набл}| > t_{kp} = t_\gamma(1-\alpha, k)$ — нулевую гипотезу отвергают.

2) Гипотеза $H_1: a > a_0$ - критическая область правосторонняя.

При $T_{набл} > t_{kp} = t_\gamma(1-2\alpha, k)$ — нулевую гипотезу отвергают.

3) Гипотеза $H_1: a < a_0$ - критическая область левосторонняя.

При $T_{набл} < -t_{kp} = -t_\gamma(1-2\alpha, k)$ — нулевую гипотезу отвергают.

Пример 4.1. Хронометраж затрат времени на сборку узла машины $n = 20$ слесарей показал, что $\bar{x}_в = 77$ мин, а $s^2 = 4$ мин^2. В предположении о нормальности распределения решить вопрос: можно ли на уровне значимости $\alpha = 0.05$ считать 80 мин нормативом (математическим ожиданием) трудоемкости?

Решение. В качестве основной гипотезы принимается $H_0: a = 80$ мин, в качестве альтернативной $H_1: a \neq 80$ мин, т.е. имеем случай двусторонней критической области, при этом $a_0 = 80$.

Используя таблицу Приложения 5, находим по верхней строчке $\alpha = 0.05$ для $k = n - 1 = 19$ критическое значение $t_{двуст.кр} = 2{,}09$. Таким образом, критерий отклонения нулевой гипотезы - принадлежность наблюдаемого значения критической области $|T_{набл}| > 2{,}09$.

С использованием таблицы Приложения 3 получим тот же критерий отклонения нулевой гипотезы $|T_{набл}| > t_\gamma(1-\alpha, k) = t_\gamma(0{,}95;19) = 2{,}093$.

По формуле (54) вычисляем $T_{набл} = \dfrac{(77-80)\sqrt{20}}{2} = -6{,}71$.

Так как число –6,71 попадает в критическую область (конкретно в интервал $(-\infty, -2.09)$, то гипотеза $H_0: a = 80$ мин отвергается.

В табличном процессоре Excel для проверки рассмотренной выше гипотезы используется уже известная функция ZTECT без указания СКО. Обращение к ней имеет вид:

=ZTECT(*массив*; a_0; [σ]),

где *массив* – адреса ячеек, содержащих выборочные данные;

a_0 и σ – имеют прежний смысл, при этом, если параметр σ опущен, то используется исправленная выборочная дисперсия *s*, вычисленная по той же выборке.

Проверка гипотезы о числовом значении дисперсии нормального распределения

Полагаем, что $X = N(a, \sigma)$. При этом по выборке объема n найдена исправленная дисперсия s^2.

Необходимо при заданном уровне значимости α проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2$ о равенстве неизвестной генеральной дисперсии σ^2 гипотетическому (предполагаемому) значению σ_0^2.

В качестве критерия возьмем случайную величину

$$K = \chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma_0^2}, \tag{56}$$

которая подчиняется χ^2-распределению с числом степеней свободы $k = n - 1$.

После вычисления наблюдаемого значения критерия (56) определим его критические значения в зависимости от альтернативной гипотезы H_1 при заданном уровне значимости α и числу степеней свободы $k = n - 1$ (Прил. 4):

Правило 1. При конкурирующей гипотезе $H_1 : \sigma^2 > \sigma_0^2$ критическая область правосторонняя с критической точкой $\chi_{кр}^2(\alpha; k)$.

Если $\chi_{набл}^2 > \chi_{кр}^2$ — нулевую гипотезу отвергают. В противном случае, если $\chi_{набл}^2 < \chi_{кр}^2$, — нет оснований отвергнуть нулевую гипотезу.

Правило 2. При конкурирующей гипотезе $H_1 : \sigma^2 \neq \sigma_0^2$ критическая область двусторонняя. Находят левую $\chi_{лев.кр}^2(1 - \alpha/2; k)$ и правую $\chi_{прав.кр}^2(\alpha/2; k)$ критические точки.

Если $\chi_{набл}^2 < \chi_{лев.кр}^2$ или $\chi_{набл}^2 > \chi_{прав.кр}^2$ — нулевую гипотезу отвергают. Если $\chi_{лев.кр}^2 < \chi_{набл}^2 < \chi_{прав.кр}^2$ — нет оснований отвергнуть нулевую гипотезу.

Правило 3. При конкурирующей гипотезе $H_1 : \sigma^2 < \sigma_0^2$ критическая область левосторонняя. с критической точкой $\chi_{кр}^2(1 - \alpha; k)$.

Если $\chi_{набл}^2 < \chi_{кр}^2(1 - \alpha; k)$ — нулевую гипотезу отвергают. Если $\chi_{набл}^2 > \chi_{кр}^2(1 - \alpha; k)$ — нет оснований отвергнуть нулевую гипотезу.

З а м е ч а н и е . Если число степеней свободы $k > 30$, то критическую точку $\chi_{kp}^2(\alpha, k)$ можно найти из равенства Уилсона—Гильферти [4]:

$$\chi^2_{kp}\left(\alpha,k\right)=k\left(1-\frac{2}{9k}+z_a\sqrt{\frac{2}{9k}}\right)^3, \qquad (57)$$

где z_a находят, используя функцию Лапласа, из равенства

$$\Phi\left(z_a\right)=\left(1-2\alpha\right)/2.$$

Пример 560.

Из нормальной генеральной совокупности извлечена выборка объема $n=21$ и по ней найдена исправленная выборочная дисперсия $s^2=16,2$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0:\sigma^2=\sigma_0^2=15$, приняв в качестве конкурирующей гипотезы $H_1:\sigma_0^2>15$.

Решение. Найдем наблюдаемое значение критерия:

$$\chi^2_{набл}=\frac{\left(n-1\right)s^2}{\sigma_0^2}=\frac{\left(21-1\right)16,2}{15}=21,6.$$

По условию, конкурирующая гипотеза имеет вид $\sigma^2>15$, поэтому критическая область — правосторонняя (правило 1). По таблице приложения 4, по уровню значимости 0,01 и числу степеней свободы $k=n-1=21-1=20$ находим критическую точку $\chi^2_{kp}\left(0,01;20\right)=37,6$

Так как $\chi^2_{набл}<\chi^2_{kp}$ — нет оснований отвергнуть нулевую гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2=15$. Другими словами, различие между исправленной дисперсией и гипотетической генеральной дисперсией незначимо.

Пример 565.

Партия изделий принимается, если дисперсия контролируемого размера значимо не превышает 0,2. Исправленная выборочная дисперсия, найденная по выборке объема $n=21$, оказалась равной $s_x^2=0,3$. Можно ли принять партию при уровне значимости 0,01?

Решение. Нулевая гипотеза $H_0:\sigma^2=\sigma_0^2=0,2$. Найдем наблюдаемое значение критерия:

$$\chi^2_{набл}=\frac{\left(n-1\right)s^2}{\sigma_0^2}=\frac{120\cdot 0,3}{0,2}=180.$$

Конкурирующая гипотеза имеет вид $H_1:\sigma^2>0,2$, следовательно, критическая область правосторонняя. Поскольку в таблице приложения 4 не содержится числа степеней свободы $k=120$, найдем критическую точку приближенно из равенства Уилсона — Гильферти (57):

$$\chi^2_{kp}(\alpha,k)=k\left(1-\frac{2}{9k}+z_a\sqrt{\frac{2}{9k}}\right)^3.$$

Найдем предварительно (учитывая, что по условию $\alpha=0,01$) $z_\alpha=z_{0,01}$ из равенства

$$\Phi\left(z_{0,01}\right)=(1-2\alpha)/2=(1-2\cdot 0,01)/2=0,49.$$

По таблице функции Лапласа (см. приложение 2), используя линейную интерполяцию, находим: $z_{0,01}=2,326$. Подставив $k=120, z_\alpha=2,326$ в формулу Уилсона — Гильферти, получим $\chi^2_{кр}(0,01;120)=158,85$. (Это приближение достаточно хорошее: в более полных таблицах χ^2 приведено значение 158,95). Так как $\chi^2_{набл}>\chi^2_{кр}$ — нулевую гипотезу отвергаем. Партию принять нельзя.

Проверка гипотезы о числовом значении вероятности события

Пусть по достаточно большому числу *n* независимых испытаний, в каждом из которых вероятность *p* появления события постоянна, но неизвестна, найдена относительная частота *m*/*n*. Требуется при заданном уровне значимости α проверить нулевую гипотезу $H_0: p=p_0$ о равенстве неизвестной вероятности *p* некоторому гипотетическому значению p_0.

В качестве критерия возьмем величину

$$Z=\frac{(m/n-p_0)\sqrt{n}}{\sqrt{p_0q_0}}=N(0,1), \qquad (58)$$

значение которой подчиняется стандартному нормальному распределению.

Критическое значение критерия z_{kp} определим в зависимости от альтернативной гипотезы (типа критической области) с использованием функции Лапласа (Приложение 2) по формулам (52), (53) для МО.

Замечание. Удовлетворительные результаты обеспечивает выполнение неравенства $np_0q_0>9$.

Пример 586. По 100 независимым испытаниям найдена относительна частота $m/n=0{,}14$. При уровне значимости 0,05 требуется проверить нулевую гипотезу $H_0: p=p_0=0,20$ при конкурирующей гипотезе $H_1: p\neq 0,20$.

Решение.

Найдем наблюдаемое значение критерия, учитывая, что $q_0 = 1 - p_0 = 0,8$:

$$Z_{набл} = \frac{(0,14 - 0,2)\sqrt{100}}{\sqrt{0,2 \cdot 0,8}} = -1,5.$$

По условию, конкурирующая гипотеза имеет вид $H_1 : p \neq 0,2$, поэтому критическая область — двусторонняя. Найдем критическую точку z_{kp} по равенству (52)

$$\Phi(z_{кр}) = \frac{(1-\alpha)}{2} = \frac{1-0,05}{2} = 0,475.$$

По таблице функции Лапласа (см. приложение 2) находим $z_{kp} = 1,96$.

Так как $|Z_{набл}| < z_{кр}$ — нет оснований отвергнуть нулевую гипотезу. Другими словами, наблюдаемая относительная частота 0,14 незначимо отличается от гипотетической вероятности 0,20.

Проверка гипотезы о значимости выборочного коэффициента корреляции

Пусть двумерная генеральная совокупность (X, Y) распределена нормально. Из этой совокупности извлечена выборка объема *n* и по ней найден выборочный коэффициент корреляции $r_B \neq 0$. Требуется проверить нулевую гипотезу $H_0 : r_Г = 0$ о равенстве нулю генерального коэффициента корреляции.

Если нулевая гипотеза принимается, то это означает, что X и Y некоррелированы; в противном случае — коррелированы.

Для того чтобы при уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе $H_1 : r_Г \neq 0$, надо вычислить наблюдаемое значение критерия

$$K = T_{набл} = r_B \sqrt{n-2} \Big/ \sqrt{1 - r_B^2} \tag{59}$$

и по таблице критических точек распределения Стьюдента, по заданному уровню значимости α и числу степеней свободы $k = n - 2$ найти критическую точку $t_{kp}(\alpha, k)$ двусторонней критической области. Если $|T_{набл}| < t_{kp}$ - нет оснований отвергнуть нулевую гипотезу. Если $|T_{набл}| > t_{kp}$ - нулевую гипотезу отвергают.

Пример 610.

По выборке объема *n=100*, извлеченной из двумерной нормальной генеральной совокупности (X, Y), найден выборочный коэффициент корреляции

$r_B = 0,2$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: r_Г = 0$ о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе $H_1: r_Г \neq 0$.

Решение.

Найдем наблюдаемое значение критерия (59)

$$T_{набл} = 0,2\sqrt{100-2} \Big/ \sqrt{1-0,2^2} = 2,02.$$

По условию, конкурирующая гипотеза $H_1: r_Г \neq 0$, поэтому критическая область - двусторонняя.

По таблице критических точек *t*-распределения Стьюдента (прил.5), по уровню значимости $\alpha = 0,05$ в верхней части таблицы и числу степеней свободы *k=100-2=98* методом линейной интерполяции находим критическую точку двусторонней критической области $t_{kp}(0,05;98) = 1,99$.

Поскольку $|T_{набл}| > t_{kp}$, то нулевую гипотезу отвергаем. Генеральный коэффициент корреляции значимо отличается от нуля, а генеральные совокупности (X, Y) коррелированы.

ЗАДАЧИ

ЗАДАЧА 1. а) Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 40$ извлечена выборка объема $n = 64$ и по ней найдена выборочная средняя $\overline{x} = 136,5$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу H_0: $a = a_0 = 130$ при конкурирующей гипотезе H_1: $a \neq 130$.

б) Решить эту задачу при конкурирующей гипотез H_1: $a > 130$.

в) Установлено, что средний вес таблетки лекарства сильного действия должен быть равен $a_0 = 0,50$ мг. Выборочная проверка 121 таблетки полученной партии лекарства показала, что средний вес таблетки этой партии $\overline{x} = 0,53$ мг. Требуется при уровне значимости 0,01 проверить нулевую гипотезу H_0: $a = a_0 = 0,50$ при конкурирующей гипотезе $H_1 : a > 0,50$. Многократными предварительными опытами по взвешиванию таблеток, поставляемых фармацевтическим заводом, было установлено, что вес таблеток распределен нормально со средним квадратическим отклонением $\sigma = 0,11$ мг.

ЗАДАЧА 2. а) По выборке объема *n=16*, извлеченной из нормальной генеральной совокупности, найдены выборочная средняя $\overline{x} = 118,2$ и «исправленное» среднее квадратическое отклонение *s=3,6*. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : a = a_0 = 120$ при конкурирующей гипотезе $H_1 : a \neq 120$.

а) Решить эту задачу, приняв в качестве конкурирующей гипотезы $H_1 : a < 120$.

ЗАДАЧА 3. Из нормальной генеральной совокупности извлечена выборка объема *n=17* и по ней найдена исправленная выборочная дисперсия $s^2 = 0,24$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2 = 0,18$, приняв в качестве конкурирующей гипотезы $H_1 : \sigma^2 > 0,18$.

ЗАДАЧА 4. Партия изделий принимается, если дисперсия контролируемого размера значимо не превышает 0,2. Исправленная выборочная дисперсия, найденная по выборке объема $n = 21$, оказалась равной $s_x^2 = 0,3$. Можно ли принять партию при уровне значимости 0,05?

ЗАДАЧА 5. Партия изделий принимается, если вероятность того, что изделие окажется бракованным, не превышает 0,03. Среди случайно отобранных 400 изделий оказалось 18 бракованных. Можно ли принять партию?

Указание. Принять нулевую гипотезу $H_0 : p = p_0 = 0,03,$ а в качестве конкурирующей $H_1 : p > 0,03$; уровень значимости $\alpha = 0,05$.

ЗАДАЧА 6. Завод рассылает рекламные каталоги возможным заказчикам. Как показал опыт, вероятность того, что организация, получившая каталог, закажет, рекламируемое изделие, равна 0,08. Завод разослал 1000 каталогов новой улучшенной формы и получил 100 заказов. Можно ли считать, что новая форма рекламы оказалась значимо эффективнее первой?

Указание. Принять нулевую гипотезу $H_0 : p = p_0 = 0{,}08$*, а в качестве конкурирующей* $H_1 : p > 0{,}08$*; уровень значимости* $\alpha = 0{,}05$.

ЗАДАЧА 7. В результате длительных наблюдений установлено, что вероятность полного выздоровления больного, принимавшего лекарство *А*, равна 0,8. Новое лекарство *В* назначено 800 больным, причем 660 из них полностью выздоровели. Можно ли считать новое лекарство значимо эффективнее лекарства *А* на пятипроцентном уровне значимости?

Указание. Принять $H_0 : p = 0{,}8$*;* $H_1 : p \neq 0{,}8$.

ЗАДАЧА 8. По выборке объема $n = 62$, извлеченной из нормальной двумерной генеральной совокупности (X, Y), найден выборочный коэффициент корреляции $r_B = 0,03$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0 : r_Г = 0$ о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе $H_1 : r_Г \neq 0$.

ЗАДАЧА 9. По выборке объема $n = 120$, извлеченной из нормальной двумерной генеральной совокупности (X, Y), найден выборочный коэффициент корреляции $r_B = 0,4$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : r_Г = 0$ о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе $H_1 : r_Г \neq 0$.

4.2. Проверка гипотез о равенстве числовых характеристик

Проверка гипотезы о равенстве дисперсий двух нормальных распределений

Пусть X и Y – две случайные величины, имеющие нормальные распределения и неизвестные дисперсии σ_X^2 и σ_Y^2. Пусть найдены исправленные выборочные дисперсии s_x^2 и s_y^2, по которым требуется сравнить эти дисперсии, т.е. проверить при заданном уровне значимости α нулевую гипотезу

$$H_0 : \sigma_X^2 = \sigma_Y^2. \tag{60}$$

Имеют место следующие распределения:

$$\frac{(n-1)S_X^2}{\sigma_X^2} = \chi_{n-1}^2; \qquad \frac{(m-1)S_Y^2}{\sigma_Y^2} = \chi_{m-1}^2.$$

Поэтому в соответствии с определением F-распределения отношение $\dfrac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)}$ или отношение $\dfrac{(n-1)S_X^2}{\sigma_X^2(n-1)} \Big/ \dfrac{(m-1)S_Y^2}{\sigma_Y^2(m-1)}$ будет иметь распределение Фишера с $n-1$ и $m-1$ степенями свободы, т.е.

$$F_{n-1,m-1} = \frac{S_X^2}{\sigma_X^2} \Big/ \frac{S_Y^2}{\sigma_Y^2}. \tag{61}$$

Если гипотеза (60) верна, то из (61) непосредственно получаем критерий Фишера-Снедекора (отношение большей исправленной дисперсии к меньшей)

$$F_{набл} = \frac{S_Б^2}{S_М^2} = F(k_1, k_2), \tag{62}$$

где $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$ - число степеней свободы большей и меньшей дисперсий соответственно.

Правило 1. Для определения $F_{кр}$ при конкурирующей гипотезе $H_1 : D(X) > D(Y)$ надо по таблице критических точек распределения Фишера-Снедекора (Приложение 6) по заданному уровню значимости α и степеням свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$ (k_1 - для большей дисперсии) отыскать критическую точку $F_{кр}(\alpha, k_1, k_2)$. Если $F_{набл} > F_{кр}$, то нулевую гипотезу отвергают.

Правило 2. При конкурирующей гипотезе $H_1 : D(X) \neq D(Y)$ критическую точку $F_{кр}(\alpha/2, k_1, k_2)$ ищут по уровню значимости $\alpha/2$ (вдвое меньшему заданного) и числам степеней свободы k_1 и k_2 (k_1—число степеней свободы большей дисперсии). Если $F_{набл} < F_{кр}$ — нет оснований отвергнуть нулевую гипотезу. Если $F_{набл} > F_{кр}$ — нулевую гипотезу отвергают.

Пример 554. По двум независимым выборкам, объемы которых $n_1 = 11$ и $n_2 = 14$, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии $s_x^2 = 0,76$ и $s_y^2 = 0,38$. При уровне значимости 0,05, проверить нулевую гипотезу о равенстве генеральных дисперсий, при конкурирующей гипотезе $H_1 : D(X) > D(Y)$.

Решение. Найдем отношение большей исправленной дисперсии к меньшей:

$$F_{набл} = \frac{0,76}{0,38} = 2.$$

По условию конкурирующая гипотеза имеет вид $H_1 : D(X) > D(Y)$, поэтому критическая область — правосторонняя.

По таблице приложения 6, по уровню значимости 0,05 и числам степеней свободы $\mathrm{k}_1 = \mathrm{n}_1 - 1 = 10$ и $\mathrm{k}_2 = \mathrm{n}_2 - 1 = 13$ находим критическую точку

$$F_{kp}(0,05;\ 10;\ 13) = 2,67.$$

Так как $F_{набл} < F_{кр}$ — нет оснований отвергнуть гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные дисперсии различаются незначимо.

Для проверки данной гипотезы (60) ***в Excel*** в качестве границ критической области выступают квантили распределения Фишера. Для вычисления этих квантилей используется функция FРАСПОБР, обращение к которой имеет вид:

=FРАСПОБР(вероятность;степени_свободы1;степени_свободы2)

где вероятность – уровень значимости α при построении правосторонней критической области.

Граница правосторонней критической области вычисляется с помощью выражения

$$F_{пр,\alpha} = \mathrm{FРАСПОБР}(\alpha; n_1; n_2).$$

Граница при построении двухсторонней критической области вычисляется с помощью выражения

$$F_{пр,\alpha/2} = \mathrm{FРАСПОБР}(\alpha/2; n_1; n_2).$$

Проверить гипотезу о равенстве дисперсий двух случайных величин $X = N(a_X, \sigma_X)$, $Y = N(a_Y, \sigma_Y)$ можно с использованием режима ***Двухвыборочный F-тест для дисперсии***. Для вызова режима необходимо обратиться к соответствующему пункту *Пакета анализа*.

Проверка гипотезы о равенстве математических ожиданий нормальных распределений

*Сравнение двух средних генеральных совокупностей **при известных дисперсиях** (большие независимые выборки).*

Обозначим через n и m объемы больших ($n>30$, $m>30$) независимых выборок, по которым найдены соответствующие выборочные средние $\overline{x}$ и $\overline{y}$. Генеральные дисперсии *D(X)* и *D(Y)* известны.

Гипотеза $H_0 : M(X) = M(Y)$ о равенстве МО генеральных совокупностей.

Доказывается, что случайная величина

$$K = Z = \frac{\overline{x} - \overline{y}}{\sqrt{D(X)/n + D(Y)/m}} = N(0,1), \qquad (63)$$

при выполнении гипотезы H_0 подчиняется нормальному распределению $N(0,1)$, поэтому возьмем ее в качестве критерия.

Критическое значение критерия z_{kp} определим в зависимости от альтернативной гипотезы (типа критической области) с использованием функции Лапласа (Приложение 2) по формулам (52), (53) для МО.

Для проверки этой гипотезы ***с использованием Excel*** используется режим работы ***Двухвыборочный z-тест для средних***. Для вызова этого режима необходимо обратиться к меню ***Данные***, команде *Пакет анализа*.

Сравнение двух средних генеральных совокупностей при неизвестных, но одинаковых дисперсиях (малые независимые выборки).

Обозначим через n и m объемы малых ($n<30$, $m<30$) независимых выборок, по которым найдены соответствующие выборочные средние $\overline{x}$ и $\overline{y}$ и исправленные выборочные дисперсии s_x^2, s_y^2. Генеральные дисперсии *D(X)* и *D(Y)* *не известны, но предполагаются одинаковыми (предварительно проверить равенство дисперсий по критерию Фишера-Снедокора).*

Гипотеза $H_0 : M(X) = M(Y)$ о равенстве МО генеральных совокупностей.

Доказывается, что случайная величина

$$T_{n+m-2} = \frac{\overline{x} - \overline{y}}{\sqrt{nD_{вx} + mD_{вy}}} \times \sqrt{\frac{nm(n+m-2)}{n+m}}, \qquad D_в = s^2 \cdot \frac{(n-1)}{n} \tag{64}$$

при выполнении гипотезы H_0 подчиняется t-распределению Стьюдента с $k = n + m - 2$ степенями свободы, которую и возьмем в качестве критерия.

Ранее уже рассматривался критерий (54), имеющий распределение Стьюдента с $k = n - 1$ степенями свободы. Никаких принципиальных различий в алгоритмы построения критических областей не вносится.

Критическое значение критерия t_{kp} определим в зависимости от альтернативной гипотезы (типа критической области) с использованием распределения Стьюдента с $k = n + m - 2$ степенями свободы по таблице Приложения 5. Возможно использование Приложения 3. При этом, если

1) Гипотеза $H_1 : M(X) \neq M(Y)$ - критическая область двусторонняя.

2) Гипотеза $H_1 : M(X) > M(Y)$ - критическая область правосторонняя.

3) Гипотеза $H_1 : M(X) < M(Y)$ - критическая область левосторонняя.

Для проверки этой гипотезы в ***Excel*** используется режим ***Двухвыборочный t-тест с одинаковыми дисперсиями*** *Пакета анализа* или функция

=ТТЕСТ(массив1;массив2;хвосты;тип).

Пример 570. По двум независимым малым выборкам, объемы которых $n = 12$ и $m = 18$, извлеченным из нормальных генеральных совокупностей X и Y, найдены выборочные средние $\overline{x} = 31{,}2$, $\overline{y} = 29{,}2$ и исправленные дисперсии: $s_x^2 = 0{,}84$ и $s_y^2 = 0{,}40$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : M(X) = M(Y)$ при конкурирующей гипотезе $H_1 : M(X) \neq M(Y)$.

Решение. Исправленные дисперсии различны, поэтому проверим предварительно гипотезу о равенстве генеральных дисперсий, используя критерий Фишера-Снедекора.

Найдем отношение большей дисперсии к меньшей: $F_{набл} = 0{,}84 / 0{,}40 = 2{,}1$. Дисперсия s_x^2 значительно больше дисперсии s_y^2, поэтому в качестве конкурирующей примем гипотезу $H_1 : D(X) > D(Y)$.

В этом случае критическая область — правосторонняя. По таблице приложения 6, по уровню значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = n - 1 = 12 - 1 = 11$ и $k_2 = m - 1 = 18 - 1 = 17$ находим критическую точку $F_{кр}(0,05;11;17) = 2,41$.

Так как $F_{набл} < F_{кр}$ оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий.

Предположение о равенстве генеральных дисперсий выполняется, поэтому сравним средние. Вычислим наблюдаемое значение критерия Стьюдента (64). Подставив числовые значения входящих в эту формулу величин, получим $T_{набл} = 7,1$.

По условию, конкурирующая гипотеза имеет вид $M(X) \neq M(Y)$, поэтому критическая область — двусторонняя. По уровню значимости 0,05 и числу степеней свободы $k_1 = n + m - 2 = 12 + 18 - 2 = 28$ находим по таблице приложения 5 критическую точку $t_{двуст.кр}(0,05;28) = 2,05$.

Так как $T_{набл} > t_{двуст.кр}$ — нулевую гипотезу о равенстве средних отвергаем. Другими словами, выборочные средние различаются значимо.

ЗАДАЧИ

ЗАДАЧА 1. По двум независимым выборкам, объемы которых $n_1 = 9$ и $n_2 = 16$, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии $s_X^2 = 34,02$ и $s_Y^2 = 12,15$. При уровне значимости 0,01, проверить нулевую гипотезу H_0: $D(X) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе H_1: $D(X) > D(Y)$.

ЗАДАЧА 2. По двум независимым выборкам, объемы которых $n=40$ и $m=50$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние: $\overline{x}=130$ и $\overline{y}=140$. Генеральные дисперсии известны: *D(X)=80*, *D(Y)=100*. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0: M(X)=M(Y)$ при конкурирующей гипотезе $H_1: M(X)\neq M(Y)$.

ЗАДАЧА 3. По двум независимым малым выборкам, объемы которых *n=10* и *m=8*, извлеченным из нормальных генеральных совокупностей X и Y, найдены выборочные средние: $\overline{x}=142,3$, $\overline{y}=145,3$ и исправленные дисперсии: $s_x^2=2,7$ и $s_y^2=3,2$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0: M(X)=M(Y)$ при конкурирующей гипотезе $H_1: M(X)\neq M(Y)$. *(предварительно проверить равенство дисперсий по критерию Фишера-Снедокора).*

ВАРИАНТЫ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ

1. Первичная обработка информации. Оценки параметров.

1.1. i-му варианту соответствуют только элементы выборки "x" соответствующей строки (объем выборки при этом $n = 16$).

1.2. Построить интервальный вариационный ряд; гистограмму относительных частот; эмпирическую функцию распределения.

1.3. Определить выборочное среднее, выборочную и исправленную дисперсии, СКО, моду, медиану, вариационный размах и коэффициент вариации.

1.4. Полагая закон распределения вариант нормальным, найти интервальную оценку математического ожидания с надежностью γ_m (см. таблицу).

1.5. Найти минимальный объем выборки, обеспечивающий с надежностью γ_m точность δ_m оценки математического ожидания.

1.6. Найти интервальные оценки дисперсии и СКО генеральной совокупности с надежностью γ_D (см. таблицу).

2. Регрессионный анализ и проверка гипотез.

2.1. i-му варианту соответствуют выборки случайных величин X и Y, расположенные в соответствующих строках.

2.2. По данным выборки определить:

— оценку вектора математического ожидания;

— оценку вектора выборочной дисперсии;

— выборочную ковариацию;

— выборочный коэффициент корреляции;

— выборочные коэффициенты линейных регрессий Y на X и X на Y;

— выборочные уравнения линейных регрессий Y на X и X на Y.

2.3. Используя метод наименьших квадратов, найти параметры линейной, квадратичной и логарифмической регрессий.

2.4. Определите общую, объясненную и остаточную дисперсии для рассмотренных уравнений регрессий, а также коэффициенты детерминации.

2.5. Построить поле корреляции и найденные линии регрессии.

2.6. Полагая в дальнейшем, что наблюдаемые признаки X и Y распределены по нормальному закону, проверить гипотезы о равенстве их математических ожиданий гипотетическим значениям $H_0 : a = a_0$ с уровнями значимости α_m при альтернативе $H_1 : a$ "*знак*" a_0 (см. таблицу).

2.7. Проверить гипотезы о равенстве генеральных дисперсий гипотетическим значениям $H_0 : \sigma^2 = D_0$ с уровнями значимости α_m при альтернативе $H_1 : \sigma^2 \neq D_0$ (см. таблицу).

2.8. Проверить гипотезу $H_0 : D(X) = D(Y)$ о равенстве генеральных дисперсий признаков X и Y по их выборкам при конкурирующей гипотезе об их неравенстве при уровне значимости α_S (см. таблицу).

2.9. Проверить гипотезу о значимости выборочного коэффициента корреляции при уровне значимости α_S.

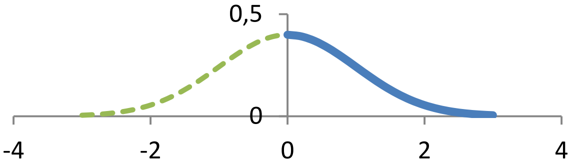
Вариант		Номер измерения																Интервал. оц.			Проверка гипотез				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	γ_m	δ_m	γ_D	α_m	зн	a_0	D_0	α_S
1	x	2,3	2,1	4,6	4,9	3,7	5	2,8	1,6	4,8	3,7	1,1	2,4	4,3	4,5	3,2	2,4	*0,95*	*0,16*	*0,95*	*0,05*	*<4,8*		*0,75*	
1	y	-1,9	-0,2	0,0	1,6	0,6	-0,9	0,0	-2,4	0,4	3,2	-4,6	-1,9	3,0	2,6	2,8	1,0				*0,02*	*>-0,3*		*7,35*	*0,05*
2	x	1,3	1,1	3,6	3,9	2,7	4	1,8	0,6	3,8	2,7	0,1	1,4	3,3	3,5	2,2	1,4	*0,99*	*0,21*	*0,98*	*0,02*	*><3,3*		*0,75*	
2	y	1,1	2,8	0,9	-1,2	2,9	0,7	2,0	1,4	-1,1	1,0	-1,5	-0,2	0,7	-1,6	0,2	0,5				*0,05*	*<1,35*		*2,48*	*0,01*
3	x	2,3	2,1	4,6	4,9	3,7	5	2,8	1,6	4,8	3,7	1,1	2,4	4,3	4,5	3,2	2,4	*0,98*	*0,19*	*0,95*	*0,05*	*<4,8*		*0,75*	
3	y	-0,6	-1,1	0,6	1,7	3,4	1,6	-0,6	-1,8	0,6	2,1	-4,6	1,0	0,6	-0,1	1,8	-0,2				*0,02*	*><0,15*		*4,43*	*0,05*
4	x	1,3	1,1	3,6	3,9	2,7	4	1,8	0,6	3,8	2,7	0,1	1,4	3,3	3,5	2,2	1,4	*0,95*	*0,16*	*0,98*	*0,05*	*<3,3*		*0,75*	
4	y	-1,3	0,3	-3,1	-1,8	-0,2	-3,9	0,4	-2,8	-2,1	-0,5	-2,4	1,7	0,8	-3,0	1,0	1,5				*0,1*	*><-1,05*		*2,22*	*0,01*
5	x	2,3	2,1	4,6	4,9	3,7	5	2,8	1,6	4,8	3,7	1,1	2,4	4,3	4,5	3,2	2,4	*0,99*	*0,21*	*0,95*	*0,05*	*<4,8*		*0,75*	
5	y	-1,4	-0,5	1,2	1,5	2,1	0,6	1,8	-4,1	2,1	2,8	-6,6	0,4	-1,0	2,4	1,9	-0,4				*0,02*	*><-1,05*		*6,89*	*0,05*
6	x	2,3	2,1	4,6	4,9	3,7	5	2,8	1,6	4,8	3,7	1,1	2,4	4,3	4,5	3,2	2,4	*0,98*	*0,19*	*0,98*	*0,02*	*<4,8*		*0,75*	
6	y	0,5	0,0	-1,1	0,2	-0,9	-2,1	0,8	4,1	-2,9	-3,6	8,3	1,6	-0,6	-2,5	-0,8	-1,2				*0,1*	*><0,45*		*12,63*	*0,01*
7	x	1,3	1,1	3,6	3,9	2,7	4	1,8	0,6	3,8	2,7	0,1	1,4	3,3	3,5	2,2	1,4	*0,95*	*0,16*	*0,95*	*0,05*	*<3,3*		*0,75*	
7	y	-2,0	0,3	1,4	0,7	-0,9	1,4	-1,0	-1,6	0,0	-3,3	1,2	-3,0	0,8	-0,1	0,0	-3,2				*0,02*	*><-1,5*		*2,07*	*0,05*
8	x	2,3	2,1	4,6	4,9	3,7	5	2,8	1,6	4,8	3,7	1,1	2,4	4,3	4,5	3,2	2,4	*0,99*	*0,21*	*0,98*	*0,05*	*><4,8*		*0,75*	
8	y	1,5	-1,1	-0,9	-0,6	-2,0	-3,1	-1,8	1,9	-3,5	-2,5	4,2	-0,5	-0,6	-2,5	-2,5	-2,3				*0,02*	*><-1,2*		*4,51*	*0,01*
9	x	1,3	1,1	3,6	3,9	2,7	4	1,8	0,6	3,8	2,7	0,1	1,4	3,3	3,5	2,2	1,4	*0,98*	*0,19*	*0,95*	*0,05*	*<3,3*		*0,75*	
9	y	-0,8	1,7	1,4	4,4	0,4	4,4	-0,1	1,1	1,8	0,2	1,2	0,2	0,7	2,7	0,3	0,7				*0,02*	*><1,35*		*2,48*	*0,05*
10	x	2,3	2,1	4,6	4,9	3,7	5	2,8	1,6	4,8	3,7	1,1	2,4	4,3	4,5	3,2	2,4	*0,95*	*0,16*	*0,98*	*0,02*	*<4,8*		*0,75*	
10	y	-0,5	2,7	1,1	-1,9	-1,7	-0,7	-0,9	2,4	-1,8	0,4	7,2	-0,9	-0,2	0,9	1,0	0,5				*0,05*	*<2,1*		*9,93*	*0,01*
11	x	17,2	23,2	29,5	29,1	17,2	30,6	7,2	6,1	29,3	25,6	6,5	9,1	31,4	24,3	32	16,4	*0,95*	*0,16*	*0,975*	*0,05*	*<25,05*		*44,72*	
11	y	-8,4	-9,9	-12,7	-10,3	-9,7	-12,9	-1,4	-0,4	-10,9	-12,9	0,6	-5,2	-14,0	-10,2	-13,0	-9,3				*0,02*	*><-11,85*		*11,80*	*0,05*
12	x	16,2	22,2	28,5	28,1	16,2	29,6	6,2	5,1	28,3	24,6	5,5	8,1	30,4	23,3	31	15,4	*0,99*	*0,21*	*0,99*	*0,05*	*><23,55*		*44,72*	
12	y	-1,8	-1,3	-4,3	-4,9	0,1	-5,9	3,5	4,5	-3,9	-2,6	4,4	2,5	-5,8	-5,1	-5,0	0,8				*0,02*	*><0,15*		*9,09*	*0,01*
13	x	17,2	23,2	29,5	29,1	17,2	30,6	7,2	6,1	29,3	25,6	6,5	9,1	31,4	24,3	32	16,4	*0,98*	*0,19*	*0,975*	*0,05*	*<25,05*		*44,72*	
13	y	0,0	-2,4	-2,1	-2,9	0,4	-4,9	7,0	7,4	-3,9	-1,4	5,8	2,7	-1,4	-3,0	-3,4	1,8				*0,02*	*><1,5*		*5,42*	*0,05*

Вариант		Номер измерения																Интервал. оц.			Проверка гипотез				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	γ_m	δ_m	γ_D	α_m	зн	a_0	D_0	α_S
14	x	16,2	22,2	28,5	28,1	16,2	29,6	6,2	5,1	28,3	24,6	5,5	8,1	30,4	23,3	31	15,4	0,95	0,16	0,99	0,02	<23,55		44,72	
14	y	-1,5	0,1	-0,3	-0,7	-0,7	-3,5	2,8	5,6	-3,0	-2,5	4,7	3,0	-3,7	0,2	-1,8	-0,2				0,05	<1,2		3,06	0,01
15	x	17,2	23,2	29,5	29,1	17,2	30,6	7,2	6,1	29,3	25,6	6,5	9,1	31,4	24,3	32	16,4	0,99	0,21	0,975	0,05	><25,05		44,72	
15	y	-0,5	-0,1	-4,0	-1,6	-0,9	-3,9	1,6	4,5	-1,0	-0,7	3,1	4,0	-2,7	-2,2	-1,3	-1,3				0,02	><0		5,33	0,05
16	x	17,2	23,2	29,5	29,1	17,2	30,6	7,2	6,1	29,3	25,6	6,5	9,1	31,4	24,3	32	16,4	0,98	0,19	0,99	0,01	<25,05		44,72	
16	y	10,9	9,8	11,4	10,2	8,7	12,4	0,2	-1,2	13,9	12,7	0,6	4,6	12,9	9,7	13,1	8,1				0,1	><11,1		8,28	0,01
17	x	16,2	22,2	28,5	28,1	16,2	29,6	6,2	5,1	28,3	24,6	5,5	8,1	30,4	23,3	31	15,4	0,95	0,16	0,975	0,05	<23,55		44,72	
17	y	0,4	3,0	2,8	5,4	-0,2	4,6	-3,7	-7,7	6,2	4,3	-6,3	-4,0	4,1	3,9	3,1	2,6				0,02	><0,75		8,83	0,05
18	x	17,2	23,2	29,5	29,1	17,2	30,6	7,2	6,1	29,3	25,6	6,5	9,1	31,4	24,3	32	16,4	0,99	0,21	0,99	0,02	><25,05		44,72	
18	y	-0,3	2,1	4,0	3,6	-0,6	2,5	-4,9	-8,7	2,6	4,0	-7,7	-3,4	3,0	0,4	4,3	-1,4				0,05	>-0,6		5,44	0,01
19	x	16,2	22,2	28,5	28,1	16,2	29,6	6,2	5,1	28,3	24,6	5,5	8,1	30,4	23,3	31	15,4	0,98	0,19	0,975	0,05	<23,55		44,72	
19	y	1,1	2,9	0,3	0,1	-0,7	0,1	-3,1	-4,4	1,5	2,5	-2,7	-0,8	3,8	2,9	2,6	-1,5				0,02	><-1,2		6,03	0,05
20	x	17,2	23,2	29,5	29,1	17,2	30,6	7,2	6,1	29,3	25,6	6,5	9,1	31,4	24,3	32	16,4	0,95	0,16	0,99	0,02	<25,05		44,72	
20	y	2,6	2,7	0,9	1,2	0,5	4,6	-2,9	-3,3	2,7	3,2	-3,3	-3,5	1,7	1,8	3,6	-1,2				0,05	>0,15		5,01	0,01
21	x	3	5	3,2	1,4	3,4	4,1	2,3	4,1	4,9	3,8	4,8	2,4	1,8	1,1	2,6	4,3	0,99	0,21	0,975	0,05	><4,8		0,75	
21	y	2,5	2,8	0,0	-3,6	0,0	3,1	0,0	2,2	-0,3	3,4	0,9	0,7	-4,8	-7,3	0,4	2,8				0,02	<1,05		13,31	0,05
22	x	2	4	2,2	0,4	2,4	3,1	1,3	3,1	3,9	2,8	3,8	1,4	0,8	0,1	1,6	3,3	0,98	0,19	0,99	0,02	<3,3		0,75	
22	y	3,4	-1,9	1,8	-0,7	1,7	0,8	2,0	1,0	-2,7	1,2	-0,2	3,3	2,1	-2,6	3,1	-1,2				0,1	><1,2		2,93	0,01
23	x	3	5	3,2	1,4	3,4	4,1	2,3	4,1	4,9	3,8	4,8	2,4	1,8	1,1	2,6	4,3	0,95	0,16	0,975	0,05	<4,8		0,75	
23	y	0,4	-0,3	3,5	-1,1	3,6	0,9	-1,6	2,8	0,5	1,5	2,1	-0,3	-1,1	-1,0	2,3	1,5				0,02	><1,35		5,20	0,05
24	x	2	4	2,2	0,4	2,4	3,1	1,3	3,1	3,9	2,8	3,8	1,4	0,8	0,1	1,6	3,3	0,99	0,21	0,99	0,01	><3,3		0,75	
24	y	1,4	-3,9	-0,2	-2,4	2,0	-1,1	2,2	-1,5	-2,8	-0,3	-0,6	-0,8	-0,2	-4,0	-0,7	0,7				0,05	><-0,6		3,12	0,01
25	x	3	5	3,2	1,4	3,4	4,1	2,3	4,1	4,9	3,8	4,8	2,4	1,8	1,1	2,6	4,3	0,98	0,19	0,975	0,05	<4,8		0,75	
25	y	-1,6	1,5	2,1	-5,8	-0,4	0,6	-1,4	1,4	-1,0	-0,5	2,3	0,2	-2,3	-4,2	-1,1	-0,5				0,02	><-0,9		6,97	0,05

ПРИЛОЖЕНИЕ 1 - **Нормированная функция Гаусса** $\varphi(x)$

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

$$\varphi(-x)=\varphi(x)$$



=НОРМРАСП(x;среднее;стандартное_откл;интегральная)

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	*x*
0,0	**0,3989**	3989	3989	3988	3986	3984	3982	3980	3977	3973	**0,0**
0,1	3970	3965	3961	3856	3951	3945	3939	3932	3925	3918	**0,1**
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825	**0,2**
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697	**0,3**
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538	**0,4**
0,5	**0,3521**	3503	3485	3467	3448	3429	3410	3391	3372	3352	**0,5**
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144	**0,6**
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920	**0,7**
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685	**0,8**
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444	**0,9**
1,0	**0,2420**	2396	2371	2347	2323	2299	2275	2251	2227	2203	**1,0**
1,1	2179	2155	2131	2107	2083	2059	2036	2021	1989	1965	**1,1**
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736	**1,2**
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518	**1,3**
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315	**1,4**
1,5	**0,1295**	1276	1257	1238	1219	1200	1182	1163	1145	1127	**1,5**
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957	**1,6**
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804	**1,7**
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0631	0669	**1,8**
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551	**1,9**
2,0	**0,0540**	0529	0519	0508	0498	0488	0478	0468	0459	0449	**2,0**
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363	**2,1**
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290	**2,2**
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0299	**2,3**
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180	**2,4**
2,5	**0,0175**	0171	0167	0163	0158	0154	0151	0147	0134	0139	**2,5**
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107	**2,6**
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081	**2,7**
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061	**2,8**
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046	**2,9**
3,0	**0,0044**	0043	0042	0040	0039	0038	0037	0036	0035	0034	**3,0**
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025	**3,1**
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018	**3,2**
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013	**3,3**
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009	**3,4**
3,5	**0,0009**	0008	0008	0008	0008	0007	0007	0007	0007	0006	**3,5**
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004	**3,6**
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003	**3,7**
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002	**3,8**
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001	**3,9**

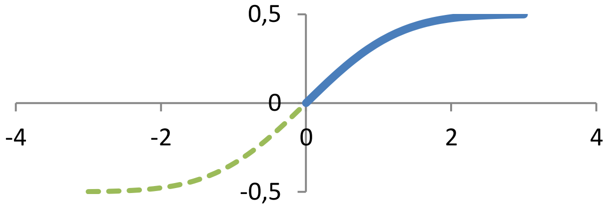
ПРИЛОЖЕНИЕ 2 - Функция Лапласа *Ф(z)*

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

$$\Phi(-z)=-\Phi(z);$$

$$0\le\Phi(z)\le\frac{1}{2};$$

=НОРМСТРАСП(z)-0,5



z	Ф(z)	z	Ф(z)	z	Ф(z)	z	Ф(z)
0,00	0,000000	***0,32***	0,125516	***0,64***	0,238914	***0,96***	0,331472
0,01	0,003989	***0,33***	0,129300	***0,65***	0,242154	***0,97***	0,333977
0,02	0,007978	***0,34***	0,133072	***0,66***	0,245373	***0,98***	0,336457
0,03	0,011966	***0,35***	0,136831	***0,67***	0,248571	***0,99***	0,338913
0,04	0,015953	***0,36***	0,140576	***0,68***	0,251748	***1,00***	0,341345
0,05	0,019939	***0,37***	0,144309	***0,69***	0,254903	***1,01***	0,343752
0,06	0,023922	***0,38***	0,148027	***0,70***	0,258036	***1,02***	0,346136
0,07	0,027903	***0,39***	0,151732	***0,71***	0,261148	***1,03***	0,348495
0,08	0,031881	***0,40***	0,155422	***0,72***	0,264238	***1,04***	0,350830
0,09	0,035856	***0,41***	0,159097	***0,73***	0,267305	***1,05***	0,353141
0,10	0,039828	***0,42***	0,162757	***0,74***	0,270350	***1,06***	0,355428
0,11	0,043795	***0,43***	0,166402	***0,75***	0,273373	***1,07***	0,357690
0,12	0,047758	***0,44***	0,170031	***0,76***	0,276373	***1,08***	0,359929
0,13	0,051717	***0,45***	0,173645	***0,77***	0,279350	***1,09***	0,362143
0,14	0,055670	***0,46***	0,177242	***0,78***	0,282305	***1,10***	0,364334
0,15	0,059618	***0,47***	0,180822	***0,79***	0,285236	***1,11***	0,366500
0,16	0,063559	***0,48***	0,184386	***0,80***	0,288145	***1,12***	0,368643
0,17	0,067495	***0,49***	0,187933	***0,81***	0,291030	***1,13***	0,370762
0,18	0,071424	***0,50***	0,191462	***0,82***	0,293892	***1,14***	0,372857
0,19	0,075345	***0,51***	0,194974	***0,83***	0,296731	***1,15***	0,374928
0,20	0,079260	***0,52***	0,198468	***0,84***	0,299546	***1,16***	0,376976
0,21	0,083166	***0,53***	0,201944	***0,85***	0,302337	***1,17***	0,379000
0,22	0,087064	***0,54***	0,205401	***0,86***	0,305105	***1,18***	0,381000
0,23	0,090954	***0,55***	0,208840	***0,87***	0,307850	***1,19***	0,382977
0,24	0,094835	***0,56***	0,212260	***0,88***	0,310570	***1,20***	0,384930
0,25	0,098706	***0,57***	0,215661	***0,89***	0,313267	***1,21***	0,386861
0,26	0,102568	***0,58***	0,219043	***0,90***	0,315940	***1,22***	0,388768
0,27	0,106420	***0,59***	0,222405	***0,91***	0,318589	***1,23***	0,390651
0,28	0,110261	***0,60***	0,225747	***0,92***	0,321214	***1,24***	0,392512
0,29	0,114092	***0,61***	0,229069	***0,93***	0,323814	***1,25***	0,394350
0,30	0,117911	***0,62***	0,232371	***0,94***	0,326391		
0,31	0,121720	***0,63***	0,235653	***0,95***	0,328944		

z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)
1,26	0,3962	**1,59**	0,4441	**1,92**	0,4726	**2,50**	0,4938
1,27	0,3980	**1,60**	0,4452	**1,93**	0,4732	**2,52**	0,4941
1,28	0,3997	**1,61**	0,4463	**1,94**	0,4738	**2,54**	0,4945
1,29	0,4015	**1,62**	0,4474	**1,95**	0,4744	**2,56**	0,4948
1,30	0,4032	**1,63**	0,4484	**1,96**	0,4750	**2,58**	0,4951
1,31	0,4049	**1,64**	0,4495	**1,97**	0,4756	**2,60**	0,4953
1,32	0,4066	**1,65**	0,4505	**1,98**	0,4761	**2,62**	0,4956
1,33	0,4082	**1,66**	0,4515	**1,99**	0,4767	**2,64**	0,4959
1,34	0,4099	**1,67**	0,4525	**2,00**	0,4772	**2,66**	0,4961
1,35	0,4115	**1,68**	0,4535	**2,02**	0,4783	**2,68**	0,4963
1,36	0,4131	**1,69**	0,4545	**2,04**	0,4793	**2,70**	0,4965
1,37	0,4147	**1,70**	0,4554	**2,06**	0,4803	**2,72**	0,4967
1,38	0,4162	**1,71**	0,4564	**2,08**	0,4812	**2,74**	0,4969
1,39	0,4177	**1,72**	0,4573	**2,10**	0,4821	**2,76**	0,4971
1,40	0,4192	**1,73**	0,4582	**2,12**	0,4830	**2,78**	0,4973
1,41	0,4207	**1,74**	0,4591	**2,14**	0,4838	**2,80**	0,4974
1,42	0,4222	**1,75**	0,4599	**2,16**	0,4846	**2,82**	0,4976
1,43	0,4236	**1,76**	0,4608	**2,18**	0,4854	**2,84**	0,4977
1,44	0,4251	**1,77**	0,4616	**2,20**	0,4861	**2,86**	0,4979
1,45	0,4265	**1,78**	0,4625	**2,22**	0,4868	**2,88**	0,4980
1,46	0,4279	**1,79**	0,4633	**2,24**	0,4875	**2,90**	0,4981
1,47	0,4292	**1,80**	0,4641	**2,26**	0,4881	**2,92**	0,4982
1,48	0,4306	**1,81**	0,4649	**2,28**	0,4887	**2,94**	0,4984
1,49	0,4319	**1,82**	0,4656	**2,30**	0,4893	**2,96**	0,4985
1,50	0,4332	**1,83**	0,4664	**2,32**	0,4898	**2,98**	0,4986
1,51	0,4345	**1,84**	0,4671	**2,34**	0,4904	**3,00**	0,49865
1,52	0,4357	**1,85**	0,4678	**2,36**	0,4909	**3,20**	0,49931
1,53	0,4370	**1,86**	0,4686	**2,38**	0,4913	**3,40**	0,49966
1,54	0,4382	**1,87**	0,4693	**2,40**	0,4918	**3,60**	0,499841
1,55	0,4394	**1,88**	0,4699	**2,42**	0,4922	**3,80**	0,499928
1,S6	0,4406	**1,89**	0,4706	**2,44**	0,4927	**4,00**	0,499968
1,57	0,4418	**1,90**	0,4713	**2,46**	0,4931	**4,50**	0,499997
1,58	0,4429	**1,91**	0,4719	**2,48**	0,4934	**5,00**	0,499999

$$P\left(\alpha < X < \beta\right) = \Phi\left(\frac{\beta - m_x}{\sigma_x}\right) - \Phi\left(\frac{\alpha - m_x}{\sigma_x}\right);$$

$$P\left(\left|X - m_x\right| < \alpha\right) = \Phi\left(\frac{\alpha}{\sigma_x}\right) - \Phi\left(\frac{-\alpha}{\sigma_x}\right) = 2\Phi\left(\frac{\alpha}{\sigma_x}\right).$$

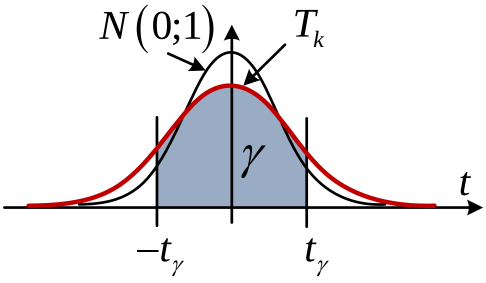
ПРИЛОЖЕНИЕ 3 - Таблица значений критерия Стьюдента

Таблица значений критерия Стьюдента $t_\gamma = t(\gamma, k)$, определяемых выражением

$$P\left(\left|T_k\right| < t_\gamma\right) = \gamma,$$

где k – число степеней свободы.

$$=СТЬЮДРАСПОБР\left(1-\gamma; k\right)$$



k \ γ	**0,95**	**0,98**	**0,99**	k \ γ	**0,95**	**0,98**	**0,99**
4	2,776	3,747	4,604	**25**	2,060	2,485	2,787
5	2,571	3,365	4,032	**26**	2,056	2,479	2,779
6	2,447	3,143	3,707	**28**	2,048	2,467	2,763
7	2,365	2,998	3,499	**30**	2,042	2,457	2,750
8	2,306	2,896	3,355	**32**	2,037	2,449	2,738
9	2,262	2,821	3,250	**34**	2,032	2,441	2,728
10	2,228	2,764	3,169	**36**	2,028	2,434	2,719
11	2,201	2,718	3,106	**38**	2,024	2,429	2,712
12	2,179	2,681	3,055	**40**	2,021	2,423	2,704
13	2,160	2,650	3,012	**45**	2,014	2,412	2,690
14	2,145	2,624	2,977	**50**	2,009	2,403	2,678
15	2,131	2,602	2,947	**60**	2,000	2,390	2,660
16	2,120	2,583	2,921	**70**	1,994	2,381	2,648
17	2,110	2,567	2,898	**80**	1,990	2,374	2,639
18	2,101	2,552	2,878	**90**	1,987	2,368	2,632
19	2,093	2,539	2,861	**100**	1,984	2,364	2,626
20	2,086	2,528	2,845	**120**	1,980	2,358	2,617
22	2,074	2,508	2,819	**1000**	1,962	2,330	2,581
24	2,064	2,492	2,797	**НЗР**	1,645	2,054	2,326

ПРИЛОЖЕНИЕ 4 - Критические точки распределения χ^2 Пирсона

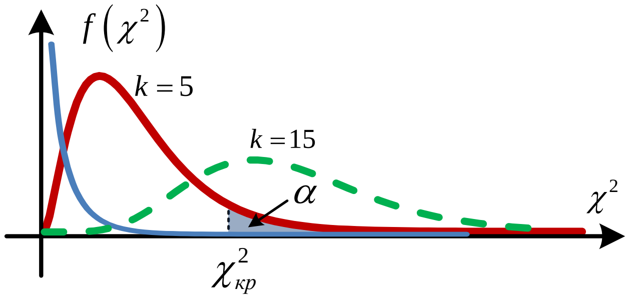
Критические точки распределения χ^2 Пирсона

$P\left(\chi^2 > \chi^2_{kp}\right) = \alpha$,

где k – число степеней свободы

$\chi^2_{лев,\gamma} = ХИ2ОБР\left(1-\alpha/2;k\right)$

$\chi^2_{пр,\gamma} = ХИ2ОБР\left(\alpha/2;k\right)$



степеней свободы k	Уровень значимости α (площадь правого хвоста)							
	0,01	**0,02**	**0,025**	**0,05**	**0,95**	**0,975**	**0,98**	**0,99**
1	6,6349	5,4119	5,0239	3,8415	0,00393	0,00098	0,00063	0,00016
2	9,2103	7,8240	7,3778	5,9915	0,1026	0,0506	0,0404	0,0201
3	11,3449	9,8374	9,3484	7,8147	0,3518	0,2158	0,1848	0,1148
4	13,2767	11,6678	11,1433	9,4877	0,7107	0,4844	0,4294	0,2971
5	15,0863	13,3882	12,8325	11,0705	1,1455	0,8312	0,7519	0,5543
6	16,8119	15,0332	14,4494	12,5916	1,6354	1,2373	1,1344	0,8721
7	18,4753	16,6224	16,0128	14,0671	2,1673	1,6899	1,5643	1,2390
8	20,0902	18,1682	17,5345	15,5073	2,7326	2,1797	2,0325	1,6465
9	21,6660	19,6790	19,0228	16,9190	3,3251	2,7004	2,5324	2,0879
10	23,2093	21,1608	20,4832	18,3070	3,9403	3,2470	3,0591	2,5582
11	24,7250	22,6179	21,9200	19,6751	4,5748	3,8157	3,6087	3,0535
12	26,2170	24,0540	23,3367	21,0261	5,2260	4,4038	4,1783	3,5706
13	27,6882	25,4715	24,7356	22,3620	5,8919	5,0088	4,7654	4,1069
14	29,1412	26,8728	26,1189	23,6848	6,5706	5,6287	5,3682	4,6604
15	30,5779	28,2595	27,4884	24,9958	7,2609	6,2621	5,9849	5,2293
16	31,9999	29,6332	28,8454	26,2962	7,9616	6,9077	6,6142	5,8122
17	33,4087	30,9950	30,1910	27,5871	8,6718	7,5642	7,2550	6,4078
18	34,8053	32,3462	31,5264	28,8693	9,3905	8,2307	7,9062	7,0149
19	36,1909	33,6874	32,8523	30,1435	10,1170	8,9065	8,5670	7,6327
20	37,5662	35,0196	34,1696	31,4104	10,8508	9,5908	9,2367	8,2604
21	38,9322	36,3434	35,4789	32,6706	11,5913	10,2829	9,9146	8,8972
22	40,2894	37,6595	36,7807	33,9244	12,3380	10,9823	10,6000	9,5425
23	41,6384	38,9683	38,0756	35,1725	13,0905	11,6886	11,2926	10,1957
24	42,9798	40,2704	39,3641	36,4150	13,8484	12,4012	11,9918	10,8564
25	44,3141	41,5661	40,6465	37,6525	14,6114	13,1197	12,6973	11,5240
26	45,6417	42,8558	41,9232	38,8851	15,3792	13,8439	13,4086	12,1981
27	46,9629	44,1400	43,1945	40,1133	16,1514	14,5734	14,1254	12,8785
28	48,2782	45,4188	44,4608	41,3371	16,9279	15,3079	14,8475	13,5647
29	49,5879	46,6927	45,7223	42,5570	17,7084	16,0471	15,5745	14,2565
30	50,8922	47,9618	46,9792	43,7730	18,4927	16,7908	16,3062	14,9535
31	52,1914	49,2264	48,2319	44,9853	19,2806	17,5387	17,0423	15,6555
32	53,4858	50,4867	49,4804	46,1943	20,0719	18,2908	17,7827	16,3622

ПРИЛОЖЕНИЕ 5 - Критические точки распределения Стьюдента

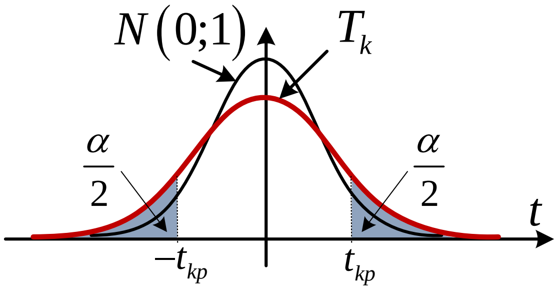
Критические точки распределения Стьюдента, определяемые выражением

$$P\left(\left|T_k\right| > t_{kp}\right) = \alpha,$$

где k – число степеней свободы.

=СТЬЮДРАСПОБР$(\alpha; k)$ *двустор. кр. обл.*

=СТЬЮДРАСПОБР$(2\alpha; k)$ *одностор.*



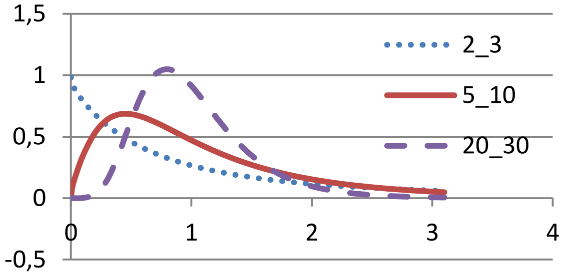
k	*Уровень значимости* α *(двусторонняя критическая область)*						
	0,001	0,002	0,005	0,01	0,02	0,05	0,1
4	8,610	7,173	5,598	4,604	3,747	2,776	2,132
5	6,869	5,893	4,773	4,032	3,365	2,571	2,015
6	5,959	5,208	4,317	3,707	3,143	2,447	1,943
7	5,408	4,785	4,029	3,499	2,998	2,365	1,895
8	5,041	4,501	3,833	3,355	2,896	2,306	1,860
9	4,781	4,297	3,690	3,250	2,821	2,262	1,833
10	4,587	4,144	3,581	3,169	2,764	2,228	1,812
11	4,437	4,025	3,497	3,106	2,718	2,201	1,796
12	4,318	3,930	3,428	3,055	2,681	2,179	1,782
13	4,221	3,852	3,372	3,012	2,650	2,160	1,771
14	4,140	3,787	3,326	2,977	2,624	2,145	1,761
15	4,073	3,733	3,286	2,947	2,602	2,131	1,753
16	4,015	3,686	3,252	2,921	2,583	2,120	1,746
17	3,965	3,646	3,222	2,898	2,567	2,110	1,740
18	3,922	3,610	3,197	2,878	2,552	2,101	1,734
20	3,850	3,552	3,153	2,845	2,528	2,086	1,725
22	3,792	3,505	3,119	2,819	2,508	2,074	1,717
24	3,745	3,467	3,091	2,797	2,492	2,064	1,711
26	3,707	3,435	3,067	2,779	2,479	2,056	1,706
28	3,674	3,408	3,047	2,763	2,467	2,048	1,701
30	3,646	3,385	3,030	2,750	2,457	2,042	1,697
34	3,601	3,348	3,002	2,728	2,441	2,032	1,691
40	3,551	3,307	2,971	2,704	2,423	2,021	1,684
50	3,496	3,261	2,937	2,678	2,403	2,009	1,676
60	3,460	3,232	2,915	2,660	2,390	2,000	1,671
70	3,435	3,211	2,899	2,648	2,381	1,994	1,667
80	3,416	3,195	2,887	2,639	2,374	1,990	1,664
100	3,390	3,174	2,871	2,626	2,364	1,984	1,660
120	3,373	3,160	2,860	2,617	2,358	1,980	1,658
1000	3,300	3,098	2,813	2,581	2,330	1,962	1,646
НЗР	3,090	2,878	2,576	2,326	2,054	1,645	1,282
	0,0005	0,001	0,0025	0,0050	0,01	0,025	0,05
	Уровень значимости α *(односторонняя критическая область)*						

ПРИЛОЖЕНИЕ 6 - Критические точки распределения Фишера-Снедекора

$$F_{k_1,k_2}=\frac{\chi^2_{k_1}/k_1}{\chi^2_{k_2}/k_2},\quad P\left(F_{k_1,k_2}>F\right)=\alpha,$$

(k_1 и k_2 — число степеней свободы большей и меньшей дисперсии соответственно)

=FРАСПОБР($\alpha;k_1;k_2$)



Уровень значимости $\alpha=0,01$

k2\k1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,41	99,42	99,42	99,43	99,43	99,44
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,13	27,05	26,98	26,92	26,87	26,83
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37	14,31	14,25	14,20	14,15
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89	9,82	9,77	9,72	9,68
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56	7,52
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31	6,28
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61	5,56	5,52	5,48
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,01	4,96	4,92
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65	4,60	4,56	4,52
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25	4,21
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01	3,97
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82	3,78
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66	3,62
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52	3,49
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41	3,37

Уровень значимости $\alpha=0,05$

k2\k1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43	19,43
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70	8,69
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86	5,84
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,49
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,20
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,99
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,60
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,51
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,44
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,38
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,33

ЛИТЕРАТУРА

1. **Вентцель Е.**С. Теория вероятностей: учебник для студентов вузов. – 10 изд. – М.: Издательский центр «Академия», 2005. – 576с.
2. **Выск Н.Д., Селиванов Ю.В., Титаренко В.И.** Вероятность и случайные величины. Методические указания и варианты курсовых заданий по теории вероятностей. М., МАТИ, 2004.
3. **Гмурман В.Е.** Теория вероятностей и математическая статистика: уч. пособие. – 12-е изд. – М.: Высшее образование, 2008. - 479с.
4. **Гмурман В.Е.** Руководство к решению задач по теории вероятностей и математической статистике: уч. пособие для студентов вузов. – 4-е изд. – М.: Высшая школа, 1997. - 400с.
5. **Воскобойников Ю.Е.** Математическая статистика (с примерами в Excel): учеб. пособие / Ю.Е.Воскобойников, Е.И.Тимошенко; Новосиб. гос. архитектур.-строит. ун-т (Сибстрин). – 2-е изд., перераб. и доп. – Новосибирск: НГАСУ (Сибстрин), 2006. – 152 с.
6. **Гулай Т.А., Долгополова А.Ф., Литвин Д.Б., Мелешко С.В.** Теория вероятностей и математическая статистика. - 2-е изд. - Ставрополь : Агрус, 2013. - 256с.
7. **Литвин Д.Б., Таволжанская О.Н.** Элементы математической статистики: Учебное пособие. – Ставрополь: ООО «Респект», 2015. – 52 с.
8. **Литвин Д.Б., Мелешко С.В.** Элементы математической статистики: Учебное пособие. – Ставрополь: Сервисшкола, 2016. – 80 с.