МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ «ВОРОНЕЖСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

РУКОВОДСТВО К РЕШЕНИЮ ЗАДАЧ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Учебно-методическое пособие

ВОРОНЕЖ
2019

Утверждено научно-методическим советом факультета прикладной математики, информатики и механики 15 июня 2019 г., протокол №10

Составители: Ляликова В.Г., Новикова Н.М., Безрядин М.М.

Рецензент – доктор физико-математических наук, профессор, заведующий кафедрой цифровых технологий факультета компьютерных наук Воронежского государственного университета С.Д. Кургалин

Учебно-методическое пособие подготовлено преподавателями факультета ПММ Воронежского государственного университета

Рекомендуется для студентов 2 курса дневного отделения факультета прикладной математики, информатики и механики, обучающихся по направлениям: 02.03.02 – Фундаментальная информатика и информационные технологии, 01.03.02 – Прикладная математика и информатика

1. ВЫБОРКА И СПОСОБЫ ЕЕ ПРЕДСТАВЛЕНИЯ

Математическая статистика позволяет получать обоснованные выводы о параметрах, видах распределений и других свойствах случайных величин по конечной совокупности наблюдений над ними - выборке.

Выборка понимается следующим образом. Пусть случайная величина X наблюдается в случайном эксперимента ξ. Осуществив n независимых испытаний эксперимента (предполагается, что условия проведения эксперимента, а следовательно и распределение наблюдаемой случайной величины X не изменяются), мы получим последовательность $(x_1, x_2, ..., x_n)$, называемую выборкой объема n из генеральной совокупности с функцией распределения $F_X(x)$.

Закон распределения случайной величины X называется *распределением генеральной совокупности*, а случайный вектор $(x_1, x_2, ..., x_n)$ – *выборочным вектором*.

Расположив величины $(x_1, x_2, ..., x_n)$ в порядке возрастания получим *вариационный ряд* $x_1 \le x_2 \le ... \le_n$. Разность между максимальным и минимальным элементами выборки называется *размахом выборки* $w = x_n - x_1$.

Из исходной выборки сформируем выборку $(z_1, z_2, ..., z_k)$, где z_i встречается в исходной выборке n_i раз $(i = 1,..,k)$. Число n_i называется *частотой* элемента z_i. Очевидно, что $\sum_{i=1}^{k} n_i = n$.

Статистическим рядом называется последовательность пар (z_i, n_i), обычно его представляют в виде таблицы, первая строка которой содержит элементы , а вторая – их частоты.

При большом объеме выборки ее элементы объединяются в группы (разряды) и результаты опытов представляются в виде *группированного* *статического ряда*. Для этого интервал, содержащий все элементы выборки, разбивается на k не пересекающихся интервалов. Для упрощения, обычно длины интервалов выбирают одинаковыми. В этом случае длину интервала полагают равной $b = w / k$.

Для выбора оптимального количества интервалов k часто применяется правило Стерджеса $k = 1 + [log_2 n]$, где n – общее число наблюдений, log_2 – логарифм по основанию 2, $[x]$ – обозначает целую часть числа x. Также часто встречается правило, оценивающее оптимальное количество интервалов как $k = [\sqrt{n}]$.

Далее для каждого из интервалов определяют частоты - количество элементов n_i попавших в него. При этом элемент, совпадающий с верхней границей интервала, относят к последующему интервалу. Наряду с частотами одновременно подсчитывают: *относительные частоты* n_i / n, *накопленные частоты* $\sum_{j=1}^{i} n_j$ и *относительные накопленные частоты* $\sum_{j=1}^{i} n_j / n$. Результаты представляют в виде таблицы, которая называется *таблицей частот группированной выборки:*

№ интервала	Границы интервала	Середина интервала	Частота	Накопленная частота	Относительная частота	Относительная накопленная частота
..	…	…	…	…	…	…

Гистограммой частот группированной выборки называют кусочно-постоянную функцию, построенную на интервалах группировки и принимающую на каждом из них значения n_i / b соответственно *i=1,2,..,k*. Площадь ступенчатой фигуры расположенной под графиком гистограммы равна объему выборки n. Аналогичным образом определяется гистограмма относительных частот. Ее площадь будет равна 1.

Полигоном частот называется ломаная линия с вершинами в точках $(z_i, \frac{n_i}{b})$, $i=1,2,..,k$, а полигоном относительных частот – ломанная с вершинами в точках $(z_i, n_i / nb)$.

Если плотность распределения генеральной совокупности является достаточно гладкой функцией, то полигон относительных частот является более хорошим приближением плотности, чем гистограмма.

Эмпирической функцией называется функция распределения выборки $(x_1,...,x_n)$ из генеральной совокупности, принимающей значения $x_1,...,x_n$ с вероятностями 1/n и обозначается $F_n^*(x)$.

Эмпирическая функция распределения определяется по значениям накопленных частот соотношением

$$F_n^*(x) = \tfrac{1}{n} \sum_{z_i < x} n_i \tag{1.1}$$

где суммируются частоты тех элементов выборки, для которых выполнено неравенство $z_i < x$, z_i – середины интервалов группировки. Заметим что $F_n^*(x) = 0$ при $x \le x_1$ и $F_n^*(x)=1$ при $x > x_n$.

При большом объеме выборки эмпирическая функция распределения может служить приближенным значением (оценкой) функции распределения генеральной совокупности в каждой точке *x*.

Пример 1.1 Для приведённой ниже выборки определить размах, а также построить вариационный и статистический ряды: 6, 4, 8, 11, 6, 6, 3, 11, 8, 3, 8, 8, 4, 3, 5.

Решение. Объем выборки n=15. Упорядочив элементы выборки по величине, получим вариационный ряд: 3, 3, 3, 4, 5, 5, 6, 6, 6, 8, 8, 8, 8, 11, 11.

Размах выборки *w*=11-3=8.

Различными в заданной выборке являются элементы $z_1 = 3$, $z_2 = 4$, $z_3 == 5$, $z_4 = 6$, $z_5 = 8$, $z_6 = 11$; их частоты соответственно равны $n_1 = 3$, $n_2 == 1$, $n_3 = 2$, $n_4 = 3$, $n_5 = 4$, $n_6 = 2$. Следовательно, сстатистический ряд исходной выборки можно записать в виде следующей таблицы:

z_i	3	4	5	6	8	11
n_i	3	1	2	3	4	2

Для контроля правильности записи находим: $\sum_{i=1}^{k} n_i = 15$.

Пример 1.2. Выборка задана в виде распределения частот:

x_i	1	4	7
n_i	2	3	5

Найти распределение относительных частот.

Решение. Для того, чтобы найти относительные частоты, сначала найдем объем выборки: n=2+3+5=10. Теперь найдем относительные частоты:

$n_1/n=2/10=0{,}2$; $n_2/n=3/10=0{,}3$: $n_3/n=5/10=0{,}5$.

Напишем искомое распределение относительных частот

x_i	1	4	7
n_i /n	0,1	0,3	0,6

Контроль: 0,2+0,3+0,6=1.

Пример 1.3. Найти эмпирическую функцию по данному распределению выборки:

x_i	1	5	9
n_i	15	10	25

Решение. Используя формулу (1.1) будем искать эмпирическую функцию распределения. Сначала найдем объем выборки n=15+10+25=50.

Наименьшее значение равно 1, поэтому $F_n^*(x) = 0$ при $x \le 1$.

Значение X<5, а именно $x_1 = 1$ наблюдалось 15 раз, следовательно, $F_n^*(x) = 15/50 = 0{,}3$ при $1 < x \le 5$.

Значение X<9, а именно: $x_1 = 1$ и $x_2 = 5$ наблюдалось 15+10=25 раз, следовательно $F_n^*(x) = 25/50 = 0{,}5$ при $5 < x \le 9$.

Так как X=9 наибольшее значение, то $F_n^*(x)=1$ при x>9.

Запишем искомую эмпирическую функцию в виде:

$$F_n^*(x)=\begin{cases}0 \text{ при } x\le 1,\\ 0{,}3 \text{ при } 1<\mathrm{x}\le 5,\\ 0{,}5 \text{ при } 5<\mathrm{x}\le 9,\\ 1 \text{ при } \mathrm{x}>9.\end{cases}$$

Пример 1.4. Представить выборку 55 наблюдений в виде таблицы частот, используя 7 интервалов группировки. Построить гистограмму, полигон частот и график эмпирической функции распределения группированной выборки. Выборка

20.3 15.3 14.3 19.3 10.1 13.9 19.5 17.8 15.4 16.8 20.1 17.8 21.1 19.1
17.2 13.5 17.2 13.2 16.8 16.2 18.3 18.5 19.6 17.8 19.2 20.4 14.7 15.7
14.7 20.2 17.8 11.8 23.3 16.5 20.8 22.8 14.5 23.8 21.3 18.6 18.1 19.7
19.5 21.9 18.1 16.7 17.5 19.1 21.9 20.5 15.3 12.5 18.4 20.4 19.4

Решение. Размах выборки *w*=23,8-10,1=13,7. Длина интервала группировки $b=13{,}7/7\approx 2$. В качестве первого интервала удобно взять интервал 10-12. Результаты группировки представлены в таблице 1.1.

Таблица 1.1

Номер интер-вала	Границы интервала	Середина интервала	Частота	Накоп-ленная частота	Относительная частота, $\frac{\mathrm{n_i}}{\mathrm{n}}$	Накопленная относительная частота, $\sum_{\mathrm{j=1}}^{i}\frac{\mathrm{n_j}}{\mathrm{n}}$
1	10 –12	11	2	2	0,0364	0,0364
2	12–14	13	4	6	0,0727	0,1091
3	14–16	15	8	14	0,1455	0,2546
4	16–18	17	12	26	0,2182	0,4728
5	18–20	19	16	42	0,2909	0,7637
6	20–22	21	10	52	0,1818	0,9455
7	22–24	23	3	55	0,0545	1,000

По результатам группировки таблицы строим гистограмму частот (рисунок 1). Соединяя отрезками ломанной середины верхних оснований прямоугольников, из которых состоит полученная гистограмма, получаем полигон частот (рисунок 2).

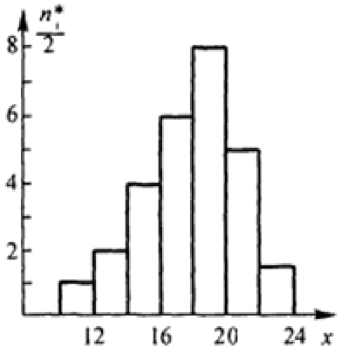


Рис.1. Гистограмма частот

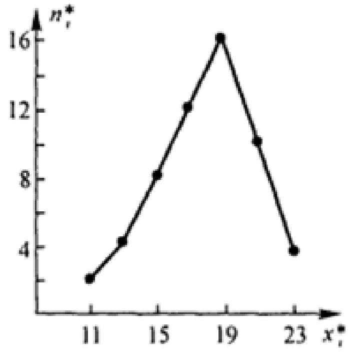


Рис.2. Полигон частот

Эмпирическая функция распределения определяется по значениям накопленных относительных частот $\sum_{j=1}^{i} \mathrm{n}_{\mathrm{j}} / n = F_n(x)$. Так как середина первого интервала группировки $z_1 = 11$, то $F_n^*(x) = 0$ при $x \leq 11$. Рассуждая аналогично, находим, что $F_n^*(x) = 1$ при $x > 23$. На полуинтервале (11,23] эмпирическую функцию распределения строим по данным третьего и последнего столбцов таблицы 1. $F_n^*(x)$ имеет скачки в точках, соответствующих серединам интервалов группировки. В результате получаем график $F_n^*(x)$, изображенный на рисунке 3.

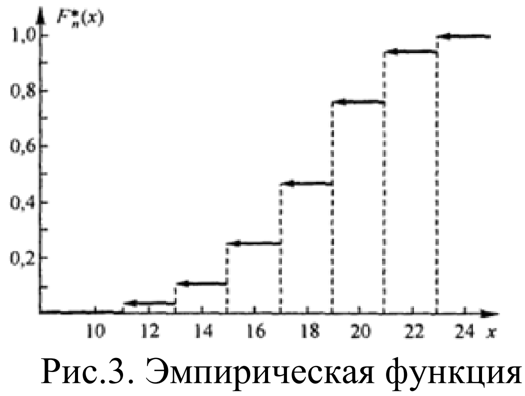


Рис.3. Эмпирическая функция распределения

Задачи для самостоятельной работы

1.1. По данному распределению выборки найти эмпирическую функцию и распределение относительных частот:

а)

x_i	2	5	7	8
n_i	1	3	2	4

б)

x_i	4	7	8
n_i	5	2	3

1.2 Для каждой из приведенных ниже выборок определить размах, а также построить вариационный и статистический ряды.

а) 11, 15, 12, 0, 16, 19, 6, 11, 12, 13, 16, 8, 9, 14, 5, 11, 3.

б) 17, 18, 16, 16, 7, 18, 19, 17, 15, 17, 19, 18, 16, 16, 18, 18

В задачах 1.3-1.5 построить графики эмпирических функций распределения, гистограммы и полигоны частот для выборок, представленных статистическими рядами

1.3.

x_i	2	3	4	5	6	7	8
n_i	1	3	4	6	5	2	1

1.4.

Границы интервалов	10–20	20–30	30–40	40–50	50–60	60–70	70–80
Частоты	1	2	7	18	12	8	2

1.5.

Границы интервалов	18–20	20–22	22–24	24–26	26–28	28–30	30–32	32–34
Частоты	4	3	3	2	4	7	12	5

1.6. Измерения емкости затвора – сток у 80 полевых транзисторов дали следующие результаты:

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

Построить гистограмму и полигон относительных частот по этой выборке, предварительно проведя группировку. В качестве длны интервала взять следующие значения: а) b=0,3; б) b=0,6.

2. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ВЫБОРОЧНОГО РАСПРЕДЕЛЕНИЯ

Определим основные числовые характеристики выборочного распределения. Пусть $x_1,..., x_n$ - выборка объема n из генеральной совокупности с функцией распределения $F_n^*(x)$. Рассмотрим выборочное распределение, т.е. распределение дискретной случайной величины, принимающей значения $x_1,..., x_n$ с вероятностями равными *1/n*. Числовые характеристики этого выборочного распределения называются выборочными (эмпирическими) числовыми характеристиками. Следует отметить, что выборочные числовые характеристики являются характеристиками данной выборки, но не являются характеристиками распределения генеральной совокупности.

Выборочное математическое ожидание или выборочное среднее определяется по формуле

$$m_x^* = \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{2.1}$$

Выборочная дисперсия определяется по формуле

$$D_x^* = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 \tag{2.2}$$

или

$$D_x^* = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \tag{2.3}$$

Выборочные начальные и центральные моменты порядка s, s-1,2,.. для не группированной выборки объема n определяются следующими формулами

$$\alpha_s^* = \frac{1}{n}\sum_{j=1}^{n} x_j^s \tag{2.4}$$

$$\mu_s^* = \frac{1}{n}\sum_{j=1}^{n}(x_j - \alpha_1^*)^s \tag{2.5}$$

Выборочной модой d_x^* унимодального (одновершинного) распределения называется элемент выборки, встречающийся с наибольшей частотой.

Выборочной медианой называется число h_x^*, которое делит вариационный ряд на две части, содержащие равное число элементов.

Если объем выборки нечетное число (т.е. n=2*L+1), то $h_x^* = x_{(L+1)}$ (2.6), то есть является элементом вариационного ряда со средним номером.

Если же n=2*L – объем выборки четное число, то

$$h_x^* = \frac{1}{2}(x_{(L)} + x_{(L+1)}) \tag{2.7}$$

Для группированной выборки объема n выборочные начальные и центральные моменты порядка s, s=1,2,... определяются по следующим формулам

$$\alpha_s^* = \frac{1}{n}\sum_{i=1}^{k} n_i z_i^s \tag{2.8}$$

$$\mu_s^* = \frac{1}{n}\sum_{i=1}^{k} n_i(z_i - \alpha_1^*)^s \tag{2.9}$$

Для упрощения вычислений выборочных среднего и дисперсии группированной выборки используют следующие формулы

$$\overline{x} = b\overline{u} + d_x^* \tag{2.10}$$

$$D_x^* = b^2 D_U^* \tag{2.11}$$

где $u_i = (z_i - d_x^*)/b, \quad i = 1,\ 2,\ \ldots\ ,k$ (2.12)

d_x^* - выборочная мода – элемент группированной выборки, b - длина интервала группировки.

Соотношение (2.12) показывает, что в выборку $z_1, z_2, ..., z_k$ внесена систематическая ошибка d_x^*, а результат подвергнут преобразованию масштаба с коэффициентом $k = 1/b$. Полученный в результате набор чисел $u_1, u_2, u_3, ..., u_k$ можно рассматривать как выборку из генеральной совокупности $U = \frac{1}{b}(x - d_x^*)$.

Пример 2.1. Вычислить моду, медиану, среднее и дисперсию следующей выборки: 8, 5, 3, 6, 4, 5, 3, 2, 1, 3.

Решение. Представим данные в виде вариационного ряда: 1, 2, 3, 3, 3, 4, 5, 5, 6, 8.

Элемент встречающийся с наибольшей частотой – это 3, значит мода $d_x^* = 3$. Так как n=10, то медиана $h_x^* = \frac{1}{2}(3+3) = 3$.

Выборочное среднее найдем по формуле (2.1):

$$\bar{x} = \frac{1}{10}(1+2+3+3+3+4+5+5+6+8) = 4{,}0\,.$$

Выборочную дисперсию найдем по формуле (2.3):

$$D_x^* = \frac{1}{10}(1^2 + 2^2 + 3^2 + 3^2 + 3^2 + 4^2 + 5^2 + 5^2 + 6^2 + 8^2 - 10 \cdot 4^2) = 3{,}8$$

Пример 2.2. Определить среднее, моду, медиану и дисперсию группированной выборки:

Границы интервалов	5–7	7–9	9–11	11–13	13–15	15–17
Частоты	8	14	40	26	6	4

Решение. Выборочное среднее группированной выборки найдем по формуле (2.8).

Середина интервала	6	8	10	12	14	16
Частоты	8	14	40	26	6	4

$$\alpha_1^* = \bar{x} = \frac{1}{98}(6 \cdot 8 + 8 \cdot 14 + 10 \cdot 40 + 12 \cdot 26 + 14 \cdot 6 + 16 \cdot 4) = 10{,}41\,.$$

Выборочную дисперсию группированной выборки найдем по формуле (2.9): $\mu_2^* = D_x^* = \frac{1}{98}(8\cdot(6-10{,}41)^2 + 14\cdot(8-10{,}41)^2 + 40\cdot(10-10{,}41)^2 +$

$+ 26\cdot(12-10{,}41)^2 + 6\cdot(14-10{,}41)^2 + 4\cdot(16-10{,}41)^2 = 5{,}22$

Так как в интервал 9-11 попадает наибольшее число элементов, то в качестве выборочной моды берем середину данного интервала, т.е. $d_x^* = 10$. Выборочная медиана делит вариационный ряд на две части, равные по числу вариант, значит $h_x^* = 10$.

Пример 2.3. Вычислить среднее и дисперсию, предварительно проведя группировку выборки с заданной длиной интервала. Вычислить среднее и дисперсию не группированной выборки, используя заданные значения. Положительные отклонения от номинального размера у партии деталей (в мм)

17	21	8	20	23	18	22	20	17	12
20	11	9	19	20	9	19	17	21	13
17	22	22	10	20	20	15	19	20	20
13	21	21	9	14	11	19	18	23	19

$n = 40; b = 2; \sum x_i = 689; \sum x_i^2 = 12635$

Решение: Проведем группировку выборки, используя указанную длину интервала *b*=2. Для удобства вычислений выборочного среднего и дисперсии для группированных данных введем столбцы u_i, $u_i \cdot n_i$, $n_i \cdot u_i^2$, $n_i \cdot (u_i + 1)^2$. Получим следующую таблицу 2.1

Таблица 2.1

Номер интервала	Границы интервала	Середина интервала	Частота n_i	u_i	$u_i \cdot n_i$	$n_i \cdot u_i^2$	$n_i \cdot (u_i + 1)^2$
1	8 – 10	9	4	-6	-24	144	100
2	10 – 12	11	3	-5	-15	75	48
3	12 – 14	13	3	-4	-12	48	27
4	14 – 16	15	2	-3	-6	18	8
5	16 – 18	17	4	-2	-8	16	4
6	18 – 20	19	7	-1	-7	7	0

7	20 – 22	21	12	0	0	0	12
8	22 - 24	23	5	1	5	5	20
$\sum$	-	-	40	-20	-67	313	219

Мода $d_x^* = 21$, следовательно по формуле (2.12) $u_i = \frac{1}{2}(z_i - 21)$.

Последний столбец служит для контроля вычислений при помощи тождества: $\sum n_i (u_i + 1)^2 = \sum n_i u_i^2 + 2\sum n_i u_i + \sum n_i$

Подставляя в тождество данные последней строки таблицы, получим $219 = 313 + 2\cdot(-67) + 40 = 219$. Следовательно, вычисления выполнены правильно. Тогда можем найти выборочное среднее и дисперсию, предварительно отыскав среднее $\bar{u}$.

По формуле (2.8) получим $\bar{u} = \frac{1}{n}\sum_{i=1}^{k} u_i n_i = \frac{1}{40}(-67) = -1{,}675$.

Тогда по формуле (2.10) найдем выборочное среднее, а по формуле (2.11) выборочную дисперсию.

$$\bar{x} = b\bar{u} + d_x^* = 2\cdot(-1{,}675) + 21 = 17{,}65$$

$$D_u^* = \frac{1}{n}\sum_{i=1}^{k} n_i u_i^2 - \bar{u}^2 = \frac{1}{40}\cdot 313 - 2{,}806 \approx 7{,}825 - 2{,}8056 \approx 5{,}02.\ D_x^* = b^2 D_u^* = 20{,}08.$$

Теперь найдем выборочное среднее и выборочную дисперсию негруппированной выборки, воспользовавшись формулами (2.1) и (2.3).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{40}\cdot 689 \approx 17{,}225$$

$$D_x^* = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{40}\cdot 12635 - 296{,}701 \approx 19{,}17.$$

Ответ: 1) $\bar{x} = 17{,}65$, $D_x^* = 20{,}08$, 2) $\bar{x} = 17{,}2$, $D_x^* = 19{,}7$.

Задачи для самостоятельной работы

Вычислить моду, медиану, среднее и исперсию следующих выборок:

2.1. 7, 3, 3, 6, 4, 5, 1, 2, 1, 3.

2.2. 3,1; 3,0; 1,5; 1,8; 2,5; 3,1; 2,4; 2,8; 1,3.

2.3. а) 1, 2, 3, 4, 5, 5, 9; б) 1, 2, 3, 4, 5, 5, 12. Сравнить полученные числовые результаты для выборок а) и б).

2.4. Определить среднее, моду, медиану и дисперсию группированной выборки:

Границы интервалов	1–3	3–5	5–7	7–9	9–11	11–13
Частоты	1	2	4	2	1	1

2.5. Определить среднее, моду, медиану и дисперсию группированной выборки:

Границы интервалов	0–4	4–8	8–12	12–16	16–20	20–24
Частоты	1	1	3	2	1	1

Для приведенных в задачах **2.6-2.7** выборок выполнить следующие задания:

1. Вычислить среднее и дисперсию, предварительно проведя группировку выборки с заданной длиной интервала; для упрощения вычислений преобразовать данные по формуле ().
2. Вычислить среднее и дисперсию не группированной выборки, используя заданные значения.

Сравнить результаты вычислений.

2.6. Время восстановления диодов из одной партии (в наносекундах)

69 73 70 68 61 73 70 72 67 70
66 70 76 68 71 71 68 70 64 65
72 70 70 69 66 70 77 69 71 74
72 72 72 68 70 67 71 67 72 69
66 75 76 69 71 67 70 73 71 64

$n=50; b=3; \sum x_i =3492; \sum x_i^2 =244342$

2.7. Время реакции (в секундах)

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

$n=36; b=1; \sum x_i =213{,}8; \sum x_i^2 =1316{,}82$

3. СТАТИСТИЧЕСКИЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ. ТОЧЕЧНЫЕ ОЦЕНКИ

Основная задача математической статистики состоит в нахождении распределения наблюдаемой случайной величины X по данным выборки. Во многих случаях вид распределения X можно считать известным, и задача сводится к получению приближенных значений неизвестных параметров этого распределения. Пусть $F_X(x,\theta)$ - функция распределения случайной величины X, содержащая один неизвестный параметр θ, а $x_1,x_2,..,x_n$ - выборка наблюдений этой случайной величины.

Статистической оценкой θ^* неизвестного параметра θ теоретического распределения называют функцию $f(x_1,x_2,..,x_n)$ от наблюдаемых случайных величин $X_1,X_2,..,X_n$.

Точечной называют статистическую оценку, которая определяется одним числом $\theta^* = f(x_1,x_2,..,x_n)$, где $x_1,x_2,..,x_n$ - результаты n наблюдений над количественным признаком X (выборка). Т.е. точечная оценка – это приближенное значение неизвестного параметра θ, полученное по выборке.

Метод подстановки состоит в том, что в качестве оценки той или иной числовой характеристики (среднего, дисперсии и др.) генеральной совокупности берут соответствующую характеристику распределения выборки – выборочную характеристику.

Качество оценок характеризуется следующими основными свойствами.

1. ***Состоятельность.*** Оценка θ^* называется состоятельной, если она сходится по вероятности к оцениваемому параметру при $n\to\infty$,т.е. $\theta^* \xrightarrow[n\to\infty]{P} \theta$. Состоятельность оценки определяется следующей теоремой: если $M[\theta^*]\to\theta$ и $D[\theta^*]\to 0$ при $n\to\infty$, то θ^* - состоятельная оценка

параметра θ.***Несмещенность оценки*** θ^* заключается в том, что ее математическое ожидание равно оцениваемому параметру при любом объеме выборки, т.е. $M[\theta^*]=\theta$. Разность $M[\theta^*]-\theta\equiv b$ называется смещением. Для несмещенных оценок систематическая ошибка оценивания равна 0.

Выборочное среднее (2.1) является несмещенной оценкой генеральной средней (математического ожидания).

Выборочная дисперсия (2.2) является смещенной оценкой генеральной дисперсии. Несмещенной оценкой генеральной дисперсии служит исправленная выборочная дисперсия

$$S^2=\frac{n}{n-1}D_x^*=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{3.1}$$

2. ***Эффективность оценки.*** Мерой точности оценки θ^* считают ее дисперсию D[θ^*]. Пусть θ_1^* и θ_2^* - две различные несмещенные оценки параметра θ. Если D[θ_1^*]<D[θ_2^*], то считают, что θ_1^* более эффективна, чем θ_2^*

В предположении, что распределение случайной величины X и статистика θ^* удовлетворяют условиям регулярности, для дисперсии несмещенной оценки θ^* параметра θ выполняется неравенство Рао-Крамера:

$D[\theta^*]\geq\frac{1}{I_n(\theta)}$, где

$$I_n(\theta)=n\cdot I(\theta)=n\cdot M\left[\left(\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)^2\right] \tag{3.2}$$

информация Фишера, содержащаяся в выборке объема n относительно неизвестного параметра θ. В качестве $p(x,\theta)$ берется либо плотность распределения в точке x (для непрерывных случайных величин), либо вероятность принять значение x (для дискретных случайных величин). Говорят, что величина $I(\theta)$ определяет количество информации по Фишеру.

Также информацию Фишера можно представить в виде

$$I_n(\theta) = -n \cdot M\left[\frac{\partial^2}{\partial\theta^2}\ln p(x,\theta)\right] \tag{3.3}$$

Условия регулярности выполняются для статистик нормального, биномиального и пуассоновского распределений.

Несмещенная оценка θ_0^* параметра θ, дисперсия которой достигает своего наименьшего возможного значения $\frac{1}{I_n(\theta)}$, называется эффективной:

$$D[\theta_0^*] = \frac{1}{I_n(\theta)} \tag{3.4}$$

Несмещенная оценка θ^* называется асимптотически эффективной оценкой параметра θ, если $\lim\limits_{n\to\infty}\frac{1}{I_n(\theta)D[\theta^*]} = 1$.

Пример 3.1. Пусть $x_1, x_2, .., x_n$ - выборка из генеральной совокупности с конечным математическим ожиданием m и дисперсией σ^2. Используя метод подстановки, найти оценку m. Проверить несмещенность и состоятельность полученной оценки.

Решение. По методу подстановки в качестве оценки m^* математического ожидания возьмем математическое ожидание распределения выборки – выборочное среднее.

Получим $m^* = \bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$.Для проверки несмещенности и состоятельности выборочного среднего как оценки m, рассмотрим эту статистику как функцию выборочного вектора ($X_1, .., X_n$). По определению выборочного вектора имеем: $M[X_i] = m$ и $D[X_i] = \sigma^2$, *i=1,..,n*. Причем X_i – независимые в совокупности случайные величины. Следовательно,

$$M[\overline{X}] = M\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} M[X_i] = \frac{1}{n} nm = m$$

$$D[\overline{X}] = D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} D[X_i] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Получаем, что $\overline{X}$ - несмещенная оценка m, и т.к. $D[\overline{X}] \to 0$ при $n \to \infty$, то $\overline{X}$ является состоятельной оценкой математического ожидания m генеральной совокупности.

Пример 3.2. Пусть $(x_1,..,x_n)$ – выборка из нормально распределенной генеральной совокупности *N(m,σ)*. Показать, что выборочное среднее $\bar{x}$ является эффективной оценкой параметра m.

Решение. В примере 2.1 было показано, что $\bar{x}$ - несмещенная оценка параметра m, причем $D[\overline{X}] = \dfrac{\sigma^2}{n}$. Для проверки эффективности оценки математического ожидания m найдем информацию Фишера $I_n(\theta)$, воспользовавшись формулой (3.2).

Так как плотность нормального закона распределения имеет вид: $f_X(x,m) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$, тогда получим $\ln f_X(x,m) = -\dfrac{(x-m)^2}{2\sigma^2} - \ln(\sqrt{2\pi}\sigma)$.

Следовательно $\dfrac{\partial \ln f_X(x,m)}{\partial m} = \dfrac{x-m}{\sigma^2}$. Математическое ожидание случайной величины $\left((x-m)/\sigma^2\right)^2$ равно:

$M\left[\dfrac{(X-m)^2}{\sigma^4}\right] = \dfrac{1}{\sigma^4} M\left[(X-m)^2\right] = \dfrac{1}{\sigma^4}\sigma^2 = 1/\sigma^2$. Таким образом, информация Фишера

$I_n(m) = nM\left[\frac{(x-M)^2}{\sigma^4}\right] = \frac{n}{\sigma^2}$. Т.к. условие (3.4) выполнено, т.е. $D[\overline{X}] = \frac{\sigma^2}{n} = \frac{1}{I_n(m)}$, то для нормально распределенной генеральной совокупности $\bar{x}$ является эффективной оценкой математического ожидания m.

Пример. 3.4. Пусть $x_1,..,x_n$ – выборка из генеральной совокупности с известным средним m и неизвестной дисперсией σ^2. Показать, что несмещенной оценкой σ^2 будет статистика $s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$.

Решение. Для того, чтобы оценка была несмещенной, необходимо чтобы выполнялось следующее равенство $M[s_0^2] = M\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2\right] = \sigma^2$.

Проверим это.

$$M[s_0^2] = M\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2\right] = \frac{1}{n}\sum_{i=1}^{n}M\left[x_i^2 - 2mx_i + m^2\right] = \frac{1}{n}\sum_{i=1}^{n}M[x_i^2] - \frac{2m}{n}\sum_{i=1}^{n}M[x_i] +$$

$$+ m^2 = M[x_i^2] - 2m^2 + m^2 = M[x_i^2] - m^2 = D[x_i] = \sigma^2.$$

Таким образом, s_0^2 является несмещенной оценкой σ^2.

Пример 3.5. Пусть $x_1, x_2, ..., x_n$ - выборка из нормального распределения генеральной совокупности $N(m,\sigma)$. Найти информацию Фишера $I_n(\sigma^2)$

Решение: По условию задачи имеем $f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$, $D[X] = \sigma^2$.

Для нахождения информации Фишера воспользуемся формулой (3.3).

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

Следовательно,

$$\frac{\partial \ln f(x)}{\partial \sigma^2} = \left(-\ln\sqrt{\sigma^2} - \frac{(x-m)^2}{2\sigma^2} \right)^1_{\sigma^2} = -\frac{1}{2\sigma^2} + \frac{(x-m)^2}{2\sigma^4},$$

$$\frac{\partial^2 \ln f(x)}{\partial(\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{(x-m)^2}{\sigma^6},$$

$$I(\sigma^2) = -M\left[\frac{\partial^2 \ln f(x)}{\partial(\sigma^2)^2}\right] = -M\left[\frac{1}{2\sigma^4} - \frac{(x-m)^2}{\sigma^6}\right] = -\frac{1}{2\sigma^4} + \frac{\sigma^2}{\sigma^6} = \frac{1}{2\sigma^4},$$

$$I_n(\sigma^2) = n \cdot I\left(\sigma^2\right) = \frac{n}{2\sigma^4}.$$

Ответ: $\frac{n}{2\sigma^4}$

Пример 3.6. Показать, что выборочное среднее, вычисленное по выборке из генеральной совокупности, имеющей распределение Пуассона с параметром λ, будет несмещенной, состоятельной и эффективной оценкой этого параметра.

Решение. Распределение Пуассона имеет вид: $P\{X = k\} \approx \frac{\lambda^k}{k!}e^{-\lambda}$, при этом известно, что $M[X] = \lambda$ и $D[X] = \lambda$. Пусть $\lambda^* = \bar{x}$ - оценка параметра λ. Проверим ее на несмещенность и состоятельность.

$M[\lambda^*] = M\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} M[X_i] = \frac{1}{n}n\lambda = \lambda$, т.е. оценка является несмещенной, т.к. $M[\lambda^*] = \lambda$.

$$D[\lambda^*] = D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} D[X_i] = \frac{1}{n^2}n\lambda = \frac{\lambda}{n}.$$

Получаем, что λ^* - несмещенная оценка, т.к. $M[\lambda^*] = \lambda$, и т.к. $D[\lambda^*] \to 0$ при $n \to \infty$, то λ^* является состоятельной оценкой.

Теперь проверим эффективность оценки. Найдем информацию Фишера $I_n(\lambda)$ воспользовавшись формулой (3.2), т.е. $I(\lambda) = M\left(\frac{\partial \ln p(x,\lambda)}{\partial \lambda}\right)^2$.

Так как $P\{X = x, \lambda\} = p(x,\lambda) \approx \frac{\lambda^x}{x!}e^{-\lambda}$ получим

$\ln p(x,\lambda) = x\ln\lambda - \ln x! - \lambda$,

$$\frac{\partial \ln p(x,\lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}.$$

$M\left[\frac{\partial \ln p(x,\lambda)}{\partial \lambda}\right]^2 = M\left[\frac{x-\lambda}{\lambda}\right]^2 = \frac{1}{\lambda^2}M[(x-\lambda)^2] = \frac{1}{\lambda^2}\lambda = \frac{1}{\lambda}$, следовательно

$I_n(\lambda) = \frac{n}{\lambda}$, и таким образом $D[\lambda^*] = \frac{1}{I_n(\lambda)}$. Значит, выборочное среднее является эффективной оценкой праметра λ.

Задачи для самостоятельной работы

3.1. Пусть $x_1, x_2, .., x_n$ - выборка наблюдений случайной величины X. Используя метод подстановки, найти оценку следующих числовых характеристик случайной величины X: дисперсии, медианы, асимметрии, эксцесса.

3.2. Пусть $x_1, x_2, .., x_n$ - выборка из генеральной совокупности с конечным начальным моментом α_{2l} . Используя метод подстановки, найти оценку начального момента α_l. Показать, что полученная оценка является несмещенной и состоятельной.

3.3. Предположим, что выборка $(x_1, .., x_n)$ получена из генеральной совокупности с конечным математическим ожиданием m и дисперсией σ^2. Показать, что несмещенная оценка дисперсии генеральной совокупности задается статистикой: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Показать, что статистика S^2 является состоятельной оценкой дисперсии генеральной совокупности.

3.4. Пусть $\tilde{\theta}$ - несмещенная оценка параметра θ, $D[\tilde{\theta}] < \infty$. Показать, что $\tilde{\theta}^2$ является смещенной θ^2, и вычислить смещение.

3.5. Пусть $x_1, x_2, .., x_n$ - выборка из генеральной совокупности с конечным математическим ожиданием m и дисперсией σ^2. Показать, что

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

является смещенной и состоятельной оценкой параметра σ.

3.6. Пусть $x_1, x_2, .., x_n$ - выборка из генеральной совокупности, имеющей равномерное распределение R(0,1). Показать, что статистика $\tilde{m} = \frac{1}{2}\left(x^{(1)} + x^{(2)}\right)$ является более эффективной оценкой математического ожидания, чем выборочное среднее.

3.7. В результате проведения n независимых экспериментов в одних и тех же условиях случайное событие A произошло x раз. Показать, что относительная частота $h = x/n$ появления события A будет несмещенной и состоятельной оценкой вероятности события A: $P(A) = p$ в одном эксперименте.

3.8. В условиях предыдущей задачи показать, что относительная частота появления события A в n независимых испытаниях является эффективной оценкой вероятности p появления события A в одном испытании.

3.9. Пусть $x_1,..,x_n$ – выборка из генеральной совокупности с известным средним m и неизвестной дисперсией σ^2. Показать, что статистика $s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$ является эффективной оценкой σ^2.

3.10 Из генеральной совокупности извлечена выборка объема $n=50$:

x_i	2	5	7	10
n_i	16	12	8	14

Найти несмещенную оценку генеральной средней.

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

x_i	1	3	6	26
n_i	8	40	1	2

Найти несмещенную оценку генеральной средней.

3.12. По выборке объема n=41 найдена смещенная оценка $D_x^* = 3$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

3.13. По выборке объема n=51 найдена смещенная оценка $D_x^* = 5$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

3.14. В итоге 5 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.

3.15. В итоге 4 измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: 8; 9; 11; 12. Найти: а) выборочную среднюю результатов измерений; б) выборочную и исправленную дисперсии ошибок прибора.

3.16. Восемь независимых измерений расстояния между двумя геодезическими знаками дали следующие результаты: 369, 378, 315, 420, 385, 401, 372, 383 м. Систематическая ошибка отсутствует. Найти несмещенную оценку дисперсии ошибок измерительного метода, если длина измеряемого расстояния $\bar{x} = 375$ м.

4. МЕТОДЫ ПОЛУЧЕНИЯ ТОЧЕЧНЫХ ОЦЕНОК

4.1. Метод максимального правдоподобия

Метод максимального правдоподобия точечной оценки неизвестных параметров заданного распределения сводится к отысканию максимума функции одного или нескольких оцениваемых параметров.

Пусть X – дискретная случайная величина, которая в результате n опытов приняла возможные значения $x_1, x_2,...,x_n$. Допустим, что вид закона распределения величины X задан, но неизвестен параметр θ, которым определяется этот закон, требуется найти его точечную оценку $\theta^* = \theta^*(x_1, x_2,..,x_n)$.

Обозначим вероятность того, что в результате испытания величина X примет значение x_i через $p(x_i, \theta)$.

Функция правдоподобия дискретной случайной величины X называют функцию аргумента θ:

$$L(x_1, x_2,..,x_n;\theta) = p(x_1;\theta) \cdot p(x_2;\theta)....p(x_n;\theta) = \prod_{i=1}^{n} p(x_i;\theta)$$

Оценкой максимального правдоподобия параметра θ называют такое его значение θ^*, при котором функция правдоподобия достигает максимума.

Функции L и LnL достигают максимума при одном и том же значении θ, поэтому вместо отыскания максимума функции L ищут, что удобнее, максимум функции lnL. Логарифмической функцией правдоподобия называют функцию lnL.

Пусть X – непрерывная случайная величина, которая в результате n испытаний приняла значения $x_1, x_2,...,x_n$. Допустим, что вид плотности распределения – функции f(x) - задан, но неизвестен параметр θ, которым определяется эта функция.

Функцией правдоподобия непрерывной случайной величины X называют функцию аргумента θ:

$$L(x_1, x_2,..,x_n;\theta) = f(x_1;\theta) \cdot f(x_2;\theta)....f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Оценку максимального правдоподобия неизвестного параметра распределения непрерывной случайной величины ищут так же, как в случае дискретной случайной величины.

Точку максимума функции lnL аргумента θ можно искать, например, следующим образом:

1. Найти производную $\frac{d\ln L}{d\theta}$
2. Приравнять производную нулю и найти критическую точку θ^* - корень полученного уравнения (его называют уравнением правдоподобия)
3. Найти вторую производную $\frac{d^2\ln L}{d\theta^2}$. Если вторая производная при $\theta=\theta^*$ отрицательна, то θ^* - точка максимума.

Найденную точку максимума θ^* принимают в качестве оценки максимального правдоподобия параметр θ

Максимально правдоподобные оценки состоятельны, асимптотически эффективны и асимптотически нормально распределены. Если для параметра θ существует эффективная оценка, то метод максимального правдоподобия дает именно эту оценку.

Пример 4.1. Найти методом максимального правдоподобия оценку параметра λ распределения Пуассона.

Решение. Пусть $x_1,...,x_n$ - выборка наблюдений случайной величины X, имеющей распределение Пуассона с неизвестным параметром λ, т.е. $P(X=x)=\frac{\lambda^x}{x!}e^{-\lambda}$, x=0,1,2,...

Составим функцию правдоподобия:

$L=p(x_1;\theta)\cdot p(x_2;\theta)...p(x_n;\theta)$.

Учитывая, что $\theta=\lambda$ и $P(X=x_i)=\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}$, получим

$$L=\left[\frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\right]\cdot\left[\frac{\lambda^{x_2}}{x_2!}e^{-\lambda}\right]....\left[\frac{\lambda^{x_n}}{x_n!}e^{-\lambda}\right] \text{ или } L=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{x_1!x_2!..x_n!}e^{-\lambda n}$$

Найдём логарифмическую функцию правдоподобия:

$$L = \sum_{i=1}^{n} x_i \cdot \ln\lambda - \ln(x_1!x_2!...x_n!) - \lambda n .$$

Найдем первую производную по λ:

$$\frac{d\ln L}{d\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n .$$

Приравняв первую производную нулю и решив полученное уравнение, получим критическую точку

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} .$$

Найдем вторую производную по λ:

$$\frac{d^2\ln L}{d\lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} .$$

Легко убедиться, что при $\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i$ вторая производная отрицательна; следовательно, эта точка есть точка максимума и ее надо принять в качестве оценки максимального правдоподобия неизвестного параметра λ распределения Пуассона. Т.е. $\lambda^* = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Пример 4.2. По выборке $x_1, x_2,.., x_n$ объема n найти методом максимального правдоподобия оценку параметра m нормального распределения с параметрами *N(m,1).*

Решение. По условию задачи плотность распределения с параметрами m и 1 имеет вид: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2}}$.

Составим функцию правдоподобия:

$$L = f(x_1;\theta)\cdot f(x_2;\theta)...f(x_n;\theta) .$$

Учитывая, что $\theta = m$, получим

$$L=\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_1-m)^2}{2}}\right]\cdot\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_2-m)^2}{2}}\right]\ldots\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_m-m)^2}{2}}\right]$$ или

$$L=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_i-m)^2}{2}}=\frac{1}{(2\pi)^{n/2}}e^{-\frac{\sum_{i=1}^{n}(x_i-m)^2}{2}}$$

Найдём логарифмическую функцию правдоподобия:

$\ln \mathrm{L}=-\frac{\mathrm{n}}{2}\ln 2\pi-\frac{\sum_{i=1}^{n}(x_i-m)^2}{2}$ Найдем первую производную по m:

$$\frac{\partial\,\mathrm{Ln\,L}}{\partial \mathrm{m}}=\sum_{i=1}^{n}(x_i-m).$$

Приравняв первую производную нулю и решив полученное уравнение, получим критическую точку: $m=\frac{1}{n}\sum x_i=\bar{x}$.

Вторая производная отрицательна: $\frac{d^2\ln L}{dm}=-n,$ следовательно, точка $m^*=\frac{1}{n}\sum x_i=\bar{x}$ есть точка максимума и ее надо принять в качестве оценки максимального правдоподобия неизвестного параметра m нормального распределения с параметрами *N(m,1)*.

4.2. Метод моментов

Метод моментов нахождения точечной оценки неизвестных параметров заданного распределения состоит в приравнивании теоретических моментов соответствующим эмпирическим (выборочным) моментам того же порядка.

Если распределение определяется одним параметром, то для его отыскания приравнивают один теоретический момент одному эмпирическому моменту того же порядка. Например, можно приравнять начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка: $\alpha_1 = \alpha_1^*$. Учитывая, что $\alpha_1 = M[X]$ и $\alpha_1^* = \bar{x}$, получим

$$M[X] = \bar{x}. \qquad (*)$$

Математическое ожидание является функцией от неизвестного параметра заданного распределения, поэтому, решив уравнение (*) относительно неизвестного параметра, тем самым получим его оценку.

Если распределение определяется двумя параметрами, то приравнивают два теоретических момента двум соответствующим эмпирическим моментам того же порядка. Например, можно приравнять начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка и центральный теоретический момент второго порядка центральному эмпирическому моменту второго порядка

$$\alpha_1 = \alpha_1^*, \; \mu_2 = \mu_2^*.$$

Учитывая, что $\alpha_1 = M[X]$ и $\alpha_1^* = \bar{x}$, $\mu_2 = D[X]$ и $\mu_2^* = D_x^*$, имеем

$$\begin{cases} M[X] = \bar{x} \\ D[X] = D_x^* \end{cases} \qquad (**)$$

Левые части этих равенств являются функциями от неизвестных параметров, поэтому, решив систему (**) относительно неизвестных параметров, тем самым получим их точечные оценки.

Разумеется, для вычисления выборочной средней $\bar{x}$ и выборочной дисперсии D_x^* надо располагать выборкой $x_1, x_2, .., x_n$.

Пример 4.3. Найти методом моментов оценку параметра p (вероятности) геометрического распределения $P(X = x_i) = (1-p)^{x_i - 1} p$, где $x_1, x_2, .., x_n$ – число испытаний, произведенных до появления события; p-вероятность появления события в одном испытании.

Решение. Требуется оценить один параметр, поэтому достаточно иметь одно уравнение относительно этого параметра. Приравняем начальный теоретический момент первого порядка α_1 начальному эмпирическому моменту первого порядка α_1^*: $\alpha_1 = \alpha_1^*$.

Приняв во внимание, что $\alpha_1 = M[X]$, $\alpha_1^* = \bar{x}$, получим $M[X] = \bar{x}$. Т.е. для решения задачи необходимо вычислить математическое ожидание геометрического распределения. Из курса теории вероятности известно, что $M[X] = \frac{1}{p}$. Окончательно имеем $\frac{1}{p} = \bar{x} \Rightarrow p = \frac{1}{\bar{x}}$. Итак, точечной оценкой параметра p геометрического распределения служит следующая характеристика $p^* = \frac{1}{\bar{x}}$.

Пример 4.4. Случайная величина X (отклонение контролируемого изделия от номинала) подчинена нормальному закону распределения с неизвестными параметрами m и σ. Ниже приведено эмпирическое распределение отклонения от номинала n=200 изделий (в первой строке указано отклонение x_i (мм); во второй строке приведена частота n_i - количество изделий, имеющих отклонение x_i):

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,2	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

Найти методом моментов точечные оценки неизвестных параметров m и σ нормального распределения.

Решение. Для отыскания двух неизвестных параметров необходимо иметь два уравнения. Приравняем начальный теоретический момент первого порядка α_1 начальному эмпирическому моменту первого порядка α_1^* и центральный теоретический момент второго порядка μ_2 центральному эмпирическому моменту второго порядка μ_2^*:

$$\alpha_1 = \alpha_1^*, \ \mu_2 = \mu_2^*.$$

Учитывая, что $\alpha_1 = M[X]$ и $\alpha_1^* = \bar{x}$, $\mu_2 = D[X]$ и $\mu_2^* = D_x^*$, имеем

$$\begin{cases} M[X] = \bar{x} \\ D[X] = D_x^* \end{cases} \qquad (*)$$

Математическое ожидание и дисперсия нормального распределения равны $M[X] = m$, $D[X] = \sigma^2$, поэтому (*) можно записать в виде

$$\begin{cases} m = \bar{x} \\ \sigma^2 = D_x^* \end{cases}.$$

Итак, окончательно получаем точечные оценки параметров m и σ нормального распределения

$$\begin{cases} m^* = \bar{x} \\ \sigma^* = \sqrt{D_x^*} \end{cases} \qquad (**)$$

По заданному распределению легко найти выборочную среднюю и выборочную дисперсию по формулам (9) и (10).

$$\bar{x} = \frac{1}{200}(1{,}8 + 4{,}5 + 18{,}2 + 22{,}5 + 33 + 33{,}8 + 31{,}5 + 40{,}8 + 38 + 17{,}6 + 11{,}5) = 1{,}266$$

$$D_x^* = \frac{1}{200}(6 \cdot 0{,}966^2 + 9 \cdot 0{,}766^2 + 26 \cdot 0{,}566^2 + 25 \cdot 0{,}366^2 + 30 \cdot 0{,}166^2 +$$

$$+ 26 \cdot 0{,}034^2 + 21 \cdot 0{,}234^2 + 24 \cdot 0{,}434^2 + 20 \cdot 0{,}634^2 + 8 \cdot 0{,}934^2 + 5 \cdot 1{,}034^2) = 0{,}25$$

Следовательно, учитывая (**), получаем $m^* = 1{,}266$, $\sigma^* = \sqrt{0{,}25} = 0{,}5$.

4.3. Метод наименьших квадратов

Метод наименьших квадратов применяется, когда необходимо найти зависимость между двумя переменными заданными выборочными данным. В этом случае речь идет об отклонениях теоретических значений от экспериментальных. Метод наименьших квадратов заключается в том, что оценка определяется из условия минимизации суммы квадратов отклонений выборочных данных от определяемой оценки.

Пусть переменная Y зависит от одной переменной x. При этом предполагается, что переменная x принимает заданные (фиксированные) значения, а зависимая переменная Y имеет случайный разброс из-за ошибок измерения. Каждому значению переменной x соответствует некоторое вероятностное распределение случайной величины Y. Для построения формы зависимости переменной Y от X задана некоторая исходная случайная выборка данных (x_i, y_i), где y_i- значение зависимой переменной Y при заданном значении x_i, $i = 1,2,..,n$. Тогда каждое наблюдаемое в опыте значение y_i можно представить в виде

$$y_i = a + bx_i + \varepsilon_i \tag{4.1}$$

Здесь a и b - неизвестные параметры, $\varepsilon_1, \varepsilon_2,.., \varepsilon_n$ - ненаблюдаемые не коррелируемые случайные величины, причем предполагается $M[\varepsilon_i] = 0$, $D[\varepsilon_i] = \sigma^2$ (σ^2 - постоянная, но неизвестная дисперсия), т.е.

$$K_{\varepsilon_i,\varepsilon_j} = \begin{cases} 0, & i \neq j \\ \sigma^2, & i = j,\ i,j = 1,2,..,n \end{cases} \tag{4.2}$$

Также предполагается что случайные ошибки наблюдений имеют нормальное распределение, т.е. $\varepsilon_i \sim N(0, \sigma)$.

Это значит, что предполагается справедливой следующая модель зависимости случайной переменной Y от переменной X:

$$Y = a + bX + \varepsilon \tag{4.3}$$

Такая модель зависимости Y от X называется простой линейной регрессией Y по X.

Для получения оценок параметров a и b используется метод наименьших квадратов, который минимизирует сумму квадратов отклонений наблюдаемых значений y_i от суммы квадратов ошибок.

$$Q(a,b)=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}\left[y_i-(a+bx_i)\right]^2 \qquad (4.4)$$

Из необходимых условий минимума функции (4.3)

$$\frac{\partial Q}{\partial a}=0,\ \frac{\partial Q}{\partial b}=0,$$

Получаем оценки параметров линейной регрессии.

$$b^*=\frac{\sum_{1=1}^{n}y_i\cdot x_i-n\cdot\bar{x}\cdot\bar{y}}{\sum_{i=1}^{n}x_i^2-n\cdot\bar{x}^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x}^2)} \qquad (4.5)$$

$$a^*=\bar{y}-b^*\bar{x} \qquad (4.6),$$

где $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$, $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$.

Пусть s_x^2 - выборочная дисперсия для выборки $x_1,x_2,..,x_n$, s_y^2 - выборочная дисперсия для выборки $y_1,y_2,..,y_n$, а r^* - выборочный коэффициент корреляции для этих выборок. Тогда

$$b^*=r^*\cdot\frac{s_y}{s_x} \qquad (4.7)$$

Оценки параметров линейной регрессии, получаемые по методу наименьших квадратов, при любом законе распределения ошибок наблюдений ε_i при условиях (4.2) имеют следующие свойства:

1. Оценки a^* и b^* несмещенными оценками параметров, т.е. $M[a^*]=a$, $M[b^*]=b$.
2. Оценки a^* и b^*являются состоятельными.
3. Оценки a^* и b^* являются эффективными, то есть они имеют минимальные дисперсии в классе несмещенных оценок, являющихся линейными функциями результатов наблюдений.

4. Оценки a^* и b^* являются не коррелируемыми случайными величинами, то есть их ковариация $\operatorname{cov}(a^*, b^*) = 0$

Пример 4.5. Подобрать зависимость вида $y = a + bx$, найти оценки a^*, b^* для следующих опытных значений (x_i, y_i):

x_i	0	1	2	3	4	5
y_i	1,1	3,3	4,8	7,4	8,6	11,5

Предполагается, что дисперсии величин y_i одинаковы, величины x_i ошибок не содержат.

Решение. Пользуясь формулами (4.5) и (4.6) найдем параметры a^* и b^* линейной регрессии. Имеем

$$\bar{x} = \frac{1}{6}\left(0 + 1 + 2 + 3 + 4 + 5 + 6\right) = 2{,}5,$$

$$\bar{y} = \frac{1}{6}(1{,}1 + 3{,}3 + 4{,}8 + 7{,}4 + 8{,}6 + 11{,}5) = 6{,}12$$

$$b^* = \frac{\sum\limits_{1=1}^{4} y_i \cdot x_i - n \cdot \bar{x} \cdot \bar{y}}{\sum\limits_{i=1}^{4} x_i^2 - n \cdot \bar{x}^2} = \frac{1 \cdot 3{,}3 + 2 \cdot 4{,}8 + 3 \cdot 7{,}4 + 4 \cdot 8{,}6 + 5 \cdot 11{,}5 - 6 \cdot 2{,}5 \cdot 6{,}12}{1^2 + 2^2 + 3^2 + 4^2 + 5^2 - 6 \cdot 2{,}5^2} =$$

$$= \frac{35{,}25}{17{,}5} = 2{,}014, \text{ т.е. } a^* = 6{,}12 - 2{,}014 \cdot 2{,}5 = 1{,}085.$$

Получаем уравнение линейной регрессии

$y = 1{,}085 + 2{,}014x$.

Ответ : $a^* = 1{,}085,\ b^* = 2{,}014,\ y = 1{,}085 + 2{,}014x$.

Задачи для самостоятельной работы

4.1. Найти методом максимального правдоподобия оценку параметра σ по выборке объема n из нормально распределенной генеральной совокупности с известным математическим ожиданием m. Показать, что полученная оценка является смещенной.

4.2. Найти методом моментов и методом максимального правдоподобия точечную оценку неизвестного параметра p (вероятность появления события в одном испытания) биномиального распределения $P_m(x_i) = C_m^{x_i} p^{x_i}(1-p)^{m-x_i}$, где x_i-число появлений события в *i*-ом опыте (*i=1,2,..,n*), m-количество испытаний в одном опыте, n – число опытов. Показать, что полученная оценка является несмещенной, состоятельной и эффективной.

4.3. Найти методом моментов и методом максимального правдоподобия по выборке $x_1, x_2,.., x_n$ точечную оценку неизвестного параметра λ показательного распределения, плотность которого $f(x) = \lambda e^{-\lambda x}\ (x \geq 0)$. Является ли полученная оценка несмещенной, состоятельной и эффективной?

4.4. Найти методом максимального правдоподобия по выборке $x_1, x_2,.., x_n$ точечную оценку неизвестного параметра λ показательного распределения, плотность которого $f(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}x}\ (x \geq 0)$. Показать, что полученная оценка является несмещенной и состоятельной.

4.5. Найти методом моментов и методом максимального правдоподобия по выборке $x_1, x_2,.., x_n$ точечные оценки неизвестных параметров a и σ нормального распределения, плотность которого

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

Проверить несмещенность и состоятельность полученных оценок.

4.6. Найти методом максимального правдоподобия по выборке $x_1, x_2,.., x_n$ точечную оценку параметра p геометрического распределения:

$$P(X = x_i) = (1-p)^{x_i - 1} p,$$

где x_i – число испытаний, произведенных до появления события; *p*-вероятность появления события в одном испытании.

4.7. Найти методом максимального правдоподобия по выборке $x_1, x_2, .., x_n$ точечную оценку параметра β гамма распределения (параметр α известен), плотность которого

$$f(x) = \frac{1}{\beta^{\alpha+1}(\alpha+1)} x^{\alpha} e^{-x/\beta} \ (\alpha > 1, \beta > 0, x \geq 0).$$

4.8. Найти методом моментов по выборке $x_1, x_2, .., x_n$ точечные оценки параметров a и b равномерного распределения, плотность которого $f(x) = 1/(b-a)$ $(b > a)$.

4.9. Методом моментов найти оценки неизвестных параметров a и b для Г-распределения с плотностью $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, x > 0$.

4.10. Найти методом моментов оценку параметра λ распределения Пуассона.

4.11. Найти методом максимального правдоподобия по выборке $x_1, x_2, .., x_n$ точечную оценку параметра a распределения Рэлея, плотность которого $f(x) = \frac{x}{a^2} e^{-\frac{x^2}{2a^2}} \ (x > 0)$.

4.12. Случайная величина X (число семян сорняков в пробе зерна) распределена по закону Пуассона. Получено следующее распределение семян сорняков в n=1000 пробах зерна (x_i – количество сорняков в одной пробе, n_i – число проб, содержащих x_i семян сорняков) :

x_i	0	1	2	3	4	5	6
n_i	405	366	175	40	8	4	2

Найти методом моментов точечную оценку неизвестного параметра распределения Пуассона.

4.13. Случайная величина X (время работы элемента) имеет показательное распределение $f(x) = \lambda e^{-\lambda x} \ (x \geq 0)$. Ниже приведено эмпирическое распределение среднего времени работы n=200 элементов (x_i

– среднее время работы элемента в часах, n_i – количество элементов, проработавших в среднем x_i часов) :

x_i	2,5	7,5	12,5	17,5	22,5	27,5
n_i	133	45	15	4	2	1

Найти методом моментов точечную оценку неизвестного параметра показательного распределения.

5. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

Иногда при обработке статистической обработке результатов наблюдений необходимо найти не только оценку θ^* неизвестного параметра θ, но и охарактеризовать точность этой оценки. С этой целью вводится понятие доверительного интервала.

Интервальной называют оценку, которая определяется двумя числами – концами интервала, покрывающего оцениваемый параметр.

Доверительным интервалом для параметра θ называется интервал (θ_1, θ_2) содержащий истинное значение θ с заданной вероятностью $\gamma = 1 - \alpha$, т.е. $P[\theta_1 < \theta < \theta_2] = 1 - \alpha = \gamma$.

Число $\gamma = 1 - \alpha$ называется доверительной вероятностью, а значение α - уровнем значимости. Статистики $\theta_1 = \theta_1(x_1,..,x_n)$ и $\theta_2 = \theta_2(x_1,..,x_n)$, определяемые по выборке $(x_1,..,x_n)$ из генеральной совокупности с неизвестным параметром θ называется соответственно нижней и верхней границами доверительного интервала.

Длина доверительного интервала, характеризующая точность интервального оценивания, зависит от объема выборки n и доверительный вероятности $1 - \alpha$: при увеличении объема выборки длина доверительного интервала уменьшается, а с приближением доверительной вероятности к единице – увеличивается. Обычно используют значения γ равные: 0,9;0,95;0,99.

При решении некоторых задач применяются односторонние доверительные интервалы, границы которых определяются из условий:

$$P[\theta < \theta_2] = 1 - \alpha \text{ или } P[\theta_1 < \theta] = 1 - \alpha .$$

Эти интервалы называются соответственно левосторонними и правосторонними доверительными интервалами.

1. Доверительным интервалом (с надежностью γ) математического ожидания m нормально количественного признака X по выборочной средней $\bar{x}$ *при известном среднем квадратическим отклонением* σ генеральной совокупности служит доверительный интервал

$$\bar{x} - \frac{\sigma}{\sqrt{n}} u_{\frac{1+\gamma}{2}} < m < \bar{x} + \frac{\sigma}{\sqrt{n}} u_{\frac{1+\gamma}{2}} \qquad (5.1)$$

где n-объем выборки, $u_{\frac{1+\gamma}{2}}$-квантили нормального распределения N(0,1) (приложение 1).

2. Доверительным интервалом (с надежностью γ) математического ожидания m нормально количественного признака X по выборочной средней $\bar{x}$ *при неизвестном среднем квадратическим отклонением* σ генеральной совокупности служит доверительный интервал

$$\bar{x} - \frac{S}{\sqrt{n\text{-}1}} t_{\gamma,(n-1)} < m < \bar{x} + \frac{S}{\sqrt{n\text{-}1}} t_{\gamma,(n-1)} \qquad (5.2)$$

где n-объем выборки, S - несмещенное выборочное среднее квадратическое отклонение $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, $t_{\gamma,n-1}$ - квантиль распределения Стьюдента с n-1 степенью свободы, который находят по таблице приложения 3 по заданным n и γ.

3. Доверительным интервалом (с надежностью γ) среднего квадратического отклонения σ нормально распределенного

количественного признака X по выборочной дисперсии $s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$ *при известном* $m = \overline{x}$ служит доверительный интервал

$$\frac{ns_0^2}{\chi^2_{n,\frac{1+\gamma}{2}}} < \sigma^2 < \frac{ns_0^2}{\chi^2_{n,\frac{1-\gamma}{2}}} \tag{5.3}$$

где n-объем выборки, $\chi^2_{n,\frac{1-\gamma}{2}}$ - квантиль распределения χ^2, который находят по таблице приложения 5 по заданным n и γ.

4. Доверительным интервалом (с надежностью γ) среднего квадратического отклонения σ нормально распределенного количественного признака X по несмещенной выборочной дисперсии $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ *при неизвестном* $m = \overline{x}$ служит доверительный интервал

$$\frac{(n-1)\cdot S^2}{\chi^2_{n-1,\frac{1+\gamma}{2}}} < \sigma^2 < \frac{(n-1)\cdot S^2}{\chi^2_{n-1,\frac{1-\gamma}{2}}} \tag{5.4}$$

Пример 5.1. Найти 90% и 99% доверительные интервалы для математического ожидания (среднего) емкости конденсатора, если $\overline{x}$ =20 мкФ, n=16, среднеквадратичное отклонение известно и равно 4мкФ.

Решение: По условию задачи имеем: $\overline{x} = 20$, *n=16*, $\sigma = 4$, $\gamma_1 = 0{,}9$, $\gamma_2 = 0{,}99$. Требуется найти доверительные интервалы для математического ожидания при известном среднем квадратическим отклонением, значит будем использовать формулу (5.1).

Сначала найдем 90% доверительный интервал.

$$20 - \frac{4}{\sqrt{16}}u_{\frac{1+0,9}{2}} < m < 20 + \frac{4}{\sqrt{16}}u_{\frac{1+0,9}{2}}.$$

Квантиль $u_{\frac{1+0,9}{2}}$ найдем из таблицы приложения 1: $u_{0,95} = 1,645$.

Таким образом 90% доверительный интервал для математического ожидания имеет следующий вид:

$18,355 < m < 21,645$

Теперь найдем 99% доверительный интервал.

$$20 - \frac{4}{\sqrt{16}} u_{0,995} < m < 20 + \frac{4}{\sqrt{16}} u_{0,995}$$

Квантиль $u_{0,995}$ также найдем из таблицы приложения 1: $u_{0,995} = 2,576$.

99% доверительный интервал для математического ожидания

$17,424 < m < 22,576$.

Ответ: (18,355; 21,645), (17,424;22,576).

Пример 5.2. Найти 90% и 99% доверительные интервалы для математического ожидания (среднего) диаметра вала, если $\overline{x}$ =30 мм, n=9, несмещенная выборочная дисперсия равна 9 мм^2.

Решение: По условию задачи имеем: $\overline{x} = 30$, $n=9$, $S^2 = 9$, $\gamma_1 = 0,9$, $\gamma_2 = 0,99$.

Требуется найти доверительные интервалы для математического ожидания при неизвестном среднем квадратическим отклонением, значит будем использовать формулу (5.2).

Сначала найдем 90% доверительный интервал.

$$30 - \frac{\sqrt{9}}{\sqrt{9}} t_{0,9;8} < m < \overline{x} + \frac{\sqrt{9}}{\sqrt{9}} t_{0,9;8}.$$

Квантиль $t_{0,9,8}$ найдем из таблицы приложения 3: $t_{0,9,8} = 1,86$.

Таким образом 90% доверительный интервал для математического ожидания имеет следующий вид: $28,14 < m < 31,86$

Теперь найдем 99% доверительный интервал.

$$30-\frac{\sqrt{9}}{\sqrt{9}}t_{0,99;8}<m<\overline{x}+\frac{\sqrt{9}}{\sqrt{9}}t_{0,99;8}$$

Квантиль $t_{0,99;8}$ также найдем из таблицы приложения 3: $t_{0,99;8}=3{,}36$.

99% доверительный интервал для математического ожидания

$26{,}64<m<33{,}36$.

Ответ: (28,14; 31,86), (26,64;33,36).

Пример 5.3. Произведено 12 измерений одним прибором (без систематических ошибок) некоторой физической величины, причем “исправленное” среднее квадратическое отклонение случайных ошибок измерений равно 0,6. Найти точность прибора с надежностью 0,99. Предполагается, что результаты измерений распределены нормально.

Решение. Точность прибора характеризуется средним квадратическим отклонением случайных ошибок измерений. Поэтому задача сводится к отысканию доверительного интервала, покрывающего σ с заданной надежностью γ=0,99, когда математическое ожидание неизвестно. Используем формулу (5.4), по условию задачи имеем: n=12, s=0,6, $\gamma=0{,}99$.

$$\frac{0{,}6^2\cdot 11}{\chi^2_{0,995(11)}}<\sigma^2<\frac{0{,}6^2\cdot 11}{\chi^2_{0,005(11)}}$$

Квантили $\chi^2_{0,995(11)}$ и $\chi^2_{0,005(11)}$ найдем из таблицы приложения 5: $\chi^2_{0,995(11)}=26{,}8$, $\chi^2_{0,005(11)}=2{,}6$.

Таким образом доверительный интервал имеет следующий вид: $0{,}15<\sigma^2<1{,}52$ Ответ. $0{,}15<\sigma^2<1{,}52$.

Пример 5.5. Из генеральной совокупности извлечена выборка объеман=10:

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

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

Решение. Для нахождения доверительного интервала воспользуемся формулой (5.2). Выборочную среднюю и несмещенное среднее квадратическое отклонение найдем, воспользовавшись формулами (2.8) и (2.9), т.е.: $\overline{x} = \frac{1}{10}\sum_{i=1}^{6} n_i x_i$, $s = \sqrt{\frac{1}{10-1}\sum_{i=1}^{6} n_i (x_i - \overline{x})^2}$.

Подставив в эти формулы данные задачи, получим $\overline{x} = 2$, $s=2,4$.

Теперь найдем искомый доверительный интервал, учитывая, что $\overline{x} = 2$, $s=2,4$, $n=10$, $t_{0,95,9} = 2,26$: $0,3 < m < 3,7$.

Ответ: $0,3 < m < 3,7$.

Задачи для самостоятельной работы.

Выборочные оценки в задачах 5.1-5.4 определялись по результатам n наблюдений. Предполагается, что выборка объема n получена из генеральной совокупности, имеющей либо нормальное распределение, либо распределение, достаточно близкое к нормальному.

5.1. Найти 90% и 99% доверительные интервалы для математического ожидания (среднего) время безотказной работы электронной лампы, если $\overline{x} = 500$, n=1000, среднеквадратичное отклонение известно и равно 10ч.

5.2. Найти 90% и 99% доверительные интервалы для математического ожидания (среднего) содержания углерода в единице продукта, если $\overline{x} = 18$ г, n=25, $s^2 = 16\,\text{г}^2$.

5.3. Найти 90% и 95% доверительные интервалы для дисперсии диаметра вала, если $\overline{X}$ =29 мм, n=16, $s^2 = 4,5\,\text{мм}^2$.

5.4. Найти 90% и 99% доверительные интервалы для дисперсии содержания углерода в единице продукта, если $\overline{x} = 18,8$ г, n=9, $s^2 = 20\,\text{г}^2$.

5.5. По результатам 10 измерений емкости конденсатора прибором, не имеющим систематической ошибки, дали такие отклонения от номинала (пкФ): 5,4; -13,9; -11; 7,2; -15,6; 29,2; 1,4; -0,3; 6,6; -9,9. Найти 90% доверительный интервал для дисперсии и среднеквадратического отклонения.

5.6. Результаты наблюдений средней температуры января имеют значения:

-19,2; -14,8; -19,6; -11,1; -9,4; -16,9; -13,7; -4,9; -13,9; -9,4; -8,3; -7,9; -5,3

Найти 95% доверительный интервал для дисперсии. Предполагается, что результаты измерений распределены нормально.

5.7. Найти 95% доверительный интервал для математического ожидания средней температуры января города Саратова за 13 лет, если $\overline{x} = -13{,}75$, $D^*(x) = 20{,}09$. Предполагается, что результаты измерений распределены нормально.

5.8. Найти 95% доверительный интервал для дисперсии средней температуры января г. Саратова за 13 лет, если $\overline{x} = -11{,}87$, $D^*(x) = 22{,}14$. Предполагается, что результаты измерений распределены нормально.

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

5.10. Из генеральной совокупности извлечена выборка объема n=12:

x_i	-0,5	-0,4	-0,2	0	0,2	0,6	0,8	1	1,2	1,5
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.

5.11. По данным выборки объема n из генеральной совокупности нормально распределенного признака найдено «исправленное» среднее квадратическое отклонение s. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,99, если: а) n=10, s=5,1; б) n=50, s=14.

5.12. По данным выборки 16 независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\overline{x_в} = 42{,}8$ и «исправленное» среднее квадратическое отклонение s=8. Оценить истинное значение измеряемой величины с надежностью 0,999.

5.13. Произведено 10 измерений одним прибором (без систематических ошибок) некоторой физической величины, причем “исправленное” среднее квадратическое отклонение случайных ошибок измерений равно 0,8. Найти точность прибора с надежностью 0,95. Предполагается, что результаты измерений распределены нормально.

5.14. Для отрасли, включающей 1200 фирм, составлена случайная выборка из 19 фирм. По выборке оказалось, что исправленное среднее квадратическое отклонение для числа работающих на фирме составляет s=25 человек. Построить 90% доверительный интервал для среднего квадратического отклонения числа работающих на фирме по всей отрасли.

6. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

6.1. Основные понятия

Во многих случаях результаты наблюдений используются для проверки предположений (гипотез) относительно тех или иных свойств распределения генеральной совокупности. Статистической гипотезой называется любое предположение о свойствах распределения вероятностей, лежащего в основе наблюдаемых явлений.

Например, статистическими являются гипотезы:

1. Генеральная совокупность распределена по закону Пуассона;
2. Дисперсии двух нормальных совокупностей равны между собой.

В первой гипотезе сделано предположение о виде неизвестного распределения, во второй – о параметрах двух известных распределений.

Гипотезы обозначаются большими латинскими буквами H_0, H_1,.., H_k . Гипотеза H_0 называется основной в том смысле, что необходимо убедиться в справедливости именное ее. Основная гипотеза одна.

Гипотезы H_1, H_2,.., H_k противопоставлены H_0 и называются альтернативными (конкурирующими). Далее будем рассуждать только об одной альтернативе - H_1.

Принятие гипотезы H_0 или ее альтернативы основано на исследовании выборочных данных $(x_1,..,x_n)$ из некоторой генеральной совокупности.

Гипотеза H - простая, если она содержит только одно предположение. Например, если λ - параметр показательного распределения, то гипотеза $H_0 : \lambda = 5$ - простая. При этом конкурирующая гипотеза, в частности, может состоять в предположении, что $H_1 : \lambda \neq 5$ - тоже простая.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез. Например, сложная гипотеза $H : \lambda > 5$ состоит из бесчисленного множества простых вида $H_i : \lambda = b_i$, где b_i - любое число, большее 5.

Гипотезу проверяют на основании выборки, полученной из генеральной совокупности. Из-за случайности выборки в результате проверки могут возникнуть ошибки и принимаются неправильные решения. Различают ошибки первого и второго рода.

Ошибка первого рода состоит в том, что будет отвергнута правильная нулевая гипотеза H_0 и принята конкурирующая гипотеза H_1. Вероятность ошибки первого рода называют уровнем значимости и обозначают α .

Ошибка второго рода состоит в том, что будет принята неправильная нулевая гипотеза H_0 . Вероятность ошибки второго рода обозначают через β.

Правило К, по которому гипотеза принимается или отвергается, называется статистическим критерием (или просто критерием).

Наблюдаемым (эмпирическим) значением $K_{набл}$ называют то значения критерия, которое вычислено по выборкам. После выбора определенного критерия множество его возможных значений разбивают на два непересекающихся подмножества:

- подмножества значений критерия, при которых нулевая гипотеза H_0принимается (не отвергается), называется областью принятия гипотезы (допустимой областью);
- подмножества значений критерия, при которых нулевая гипотеза H_0 отвергается (отклоняется) и принимается альтернативная гипотеза H_1, называется критической областью.

Основной принцип проверки статистических гипотез можно сформулировать так: если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если наблюдаемое значения критерия принадлежит области принятия гипотезы – гипотезу принимают. Так как критерий К- одномерная случайная величина, то критическая область и область принятия гипотезы являются интервалами и, следовательно, существуют точки, которые их разделяют.

Критическими точками (границами) $k_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы. Различают два вида критических областей:

- односторонняя (правосторонняя или левосторонняя) критическая область. Правосторонняя область определяется неравенством

$K > k_{кр}$, где $k_{кр}$ - положительное число. Левосторонняя область определяется неравенством $K < k_{кр}$, где $k_{кр}$ - отрицательное число

- двусторонняя критическая область. Двусторонняя область определяется неравенством $K < k_1$, $K > k_2$, где $k_2 > k_1$. В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется неравенством $|K| > k_{кр}$.

Для отыскания критической области задаются уровнем значимости α и ищут критические точки, исходя из следующих соотношений:

- для правосторонней критической области

$$P(K > k_{кр}) = \alpha \ \ (k_{кр} > 0);$$

- для левосторонней критической области

$$P(K < k_{кр}) = \alpha \ \ (k_{кр} < 0);$$

- для двусторонней симметричной области

$$P(K > k_{кр}) = \alpha / 2 \ \ (k_{кр} > 0), \ P(K < -k_{кр}) = \alpha / 2$$

Мощностью критерия называют вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза. Другими словами, мощность критерия есть вероятность того что нулевая гипотеза будет отвергнута, если верна конкурирующая гипотеза.

6.2. Проверка гипотез о значениях параметров распределений (параметрические гипотезы)

Сравнение двух дисперсий нормальных генеральных совокупностей

Гипотезы о дисперсиях возникают довольно часто, поскольку дисперсии характеризуют такие важные показатели, как точность приборов, степень однородности признаков, риск, связанный с отклонением доходности от заданного уровня, разброс уровня успеваемости обучающихся относительно среднего, и т.д.

Пусть имеются независимые выборки $(x_1,..,x_{n_1})$ и $(y_1,..,y_{n_2})$ объемом n_1 и n_2 из нормальных генеральных совокупностей $N(a_x,\sigma_x)$ и $N(a_y,\sigma_y)$. Оценкой для σ_x^2 является исправленная выборочная дисперсия S_x^2, для σ_y^2 - S_y^2 . Требуется сравнить эти дисперсии.

Рассмотрим гипотезу $H_0:\sigma_x^2=\sigma_x^2$. Критерием служит случайная величина

$$F_{набл.}=\frac{S_Б^2}{S_М^2} \qquad (6.1)$$

отношение большей исправленной дисперсии к меньшей, которая при условии справедливости нулевой гипотезы имеет распределение Фишера Снедекора со степенями свободы $k_1=n_1-1$ (k_1 - число степеней свободы большей исправленной дисперсии) и $k_2=n_2-1$. Критическая область зависит от вида конкурирующей гипотезы:

1. Если $H_1:\sigma_x^2>\sigma_y^2$, то критическая область правосторонняя:

$p(F>F_{кр}(\alpha,k_1,k_2))=\alpha$, где α - уровень значимости . Тогда критическая точка

$$F_{кр}(\alpha,k_1,k_2) \qquad (6.2)$$

находится по таблице критических точек распределения Фишера Снедекора. Если $F_{набл.}=\frac{S_Б^2}{S_М^2}<F_{кр}$ нулевая гипотеза принимается, в противном случае отвергается.

2. Если $H_1:\sigma_x^2\neq\sigma_y^2$, тогда критическая область двусторонняя: $p(F<F_1)=\frac{\alpha}{2}$, $p(F>F_2)=\frac{\alpha}{2}$. При этом достаточно найти

$$F_2=F_{кр}\left(\frac{\alpha}{2},k_1,k_2\right) \qquad (6.3).$$

Если $F_{набл.} = \dfrac{S_Б^2}{S_М^2} < F_{кр}$, то нет оснований отвергнуть нулевую гипотезу. Если $F_{набл.} > F_{кр}$ - нулевую гипотезу отвергают.

Пример 6.1. 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: $H_1 : D(X) > D(Y)$.

Решение. Пользуясь формулой (6.1) найдем отношение большей исправленной дисперсии к меньшей. Имеем: $F_{набл.} = \dfrac{34{,}02}{12{,}15} = 12{,}8$.

Конкурирующая гипотеза имеет вид $H_1 : D(X) > D(Y)$, поэтому критическая область – правосторонняя. По таблице приложения 4, по уровню значимости $\alpha = 0{,}01$ и числам степеней свободы $k_1 = 9 - 1 = 8$ и $k_2 = 16 - 1 = 15$ находим критическую точку $F_{кр}(0{,}01;8;15) = 4$. Так как $F_{набл.} < F_{кр}$ - нет оснований отвергнуть гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные дисперсии различаются незначимо.

Пример 6.2. По двум независимым выборкам, объемы которых $n_1 = 9$, и $n_2 = 6$, извлеченным из нормальных генеральных совокупностей X и Y, найдены выборочные дисперсии $D_x^* = 14{,}4$ и $D_y^* = 20{,}5$. При уровне значимости 0,1 проверить нулевую гипотезу $H_0 : D(X) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1 : D(X) \neq D(Y)$.

Решение. Для того, чтобы найти $F_{набл.}$ по формуле (6.1), сначала необходимо вычислить исправленные выборочные дисперсии по следующим формулам: $S_x^2 = \frac{n_1}{n_1 - 1} D_x^*$, $S_Y^2 = \frac{n_1}{n_1 - 1} D_y^*$. Имеем: $S_X^2 = \frac{9}{8} \cdot 14{,}4 = 16{,}2$, $S_2^2 = \frac{6}{5} \cdot 20{,}5 = 24{,}6$. Теперь можем найти $F_{набл.}$: $F_{набл.} = \frac{24{,}6}{16{,}2} = 1{,}52$. По условию конкурирующая гипотеза имеет вид $H_1 : D(X) \neq D(Y)$, поэтому критическая область – двусторонняя. По таблице приложения 4 , по уровню значимости $\frac{\alpha}{2} = \frac{0{,}1}{2} = 0{,}05$ и числам степеней свободы $k_1 = 6 - 1 = 5$ и $k_2 = 9 - 1 = 8$ находим критическую точку $F_{кр}(0{,}05;5;8) = 3{,}69$. Так как $F_{набл.} < F_{кр}$ - нет оснований отвергнуть гипотезу о равенстве генеральных дисперсий.

Сравнение двух средних генеральных совокупностей, дисперсии которых известны

На практике часто встречаются ситуации, когда среднее значение данных одного эксперимента отличается от среднего значения данных другого (проводимого при тех же условиях) эксперимента. Тогда возникает вопрос, можно ли считать это расхождение незначимым, то есть чисто случайным, или оно вызвано существенным различием двух генеральных совокупностей.

Пусть X имеет нормальное распределение с параметрами a_x и σ_x, Y – такое же распределение с параметрами a_y и σ_y и дисперсии σ_x^2 и σ_y^2 известны. Имеются независимые выборки $(x_1,..,x_{n_1})$ и $(y_1,..,y_{n_2})$ объемом n (n>30) и m (m>30) из генеральных совокупностей X и Y. По ним найдены выборочные средние $\bar{x}$ и $\bar{y}$.

Рассмотрим гипотезу $H_0 : a_x = a_y$ о равенстве математических ожидании (генеральных средних). Критерием служит случайная величина

$$Z_{набл} = \frac{\overline{x} - \overline{y}}{\sqrt{\sigma_x^2 / n + \sigma_y^2 / m}} \tag{6.4}$$

которая при условии справедливости нулевой гипотезы имеет нормальное распределение с параметрами с нулевым математическим ожиданием и единичной дисперсией. Для заданного уровня доверия γ по таблицам интеграла Лапласа и уравнения $\Phi(z) = \gamma$ находим $z_{кр}$, где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ - функция Лапласа, а уровень γ и вид критической области зависит от типа конкурирующей гипотезы:

1. При альтернативе $H_1 : a_x \neq a_y$ критическая область двусторонняя. Тогда критическая точка $Z_{кр}$ определяется по таблице функции Лапласа из равенства

$$\Phi(z_{кр}) = (1 - \alpha) / 2 \tag{6.5}$$

Если $|z_{набл}| < z_{кр}$ - нет оснований отвергнуть нулевую гипотезу, т.е. нулевая гипотеза принимается, в обратном случае – отвергается.

2. При конкурирующей гипотезе $H_1 : a_x > a_y$ критическая область правосторонняя. Критическая точка $z_{кр}$ определяется по таблице функции Лапласа из равенства

$$\Phi(z_{кр}) = (1 - 2\alpha) / 2 \tag{6.6}$$

Если $z_{набл} < z_{кр}$ - нулевая гипотеза принимается. Если $z_{набл} > z_{кр}$ - нулевую гипотезу отвергают.

3. При конкурирующей гипотезе $H_1 : a_x < a_y$ критическая область левосторонняя. Критическая точка $z_{кр}$ вычисляется так же, как и в

предыдущем случае (пункт 2). Если $z_{набл} > -z_{кр}$ - нулевая гипотеза принимается. Если $z_{набл} < -z_{кр}$ - нулевую гипотезу отвергают.

Пример 6.3. По выборке объема n=50 найден средний размер $\overline{x} = 20{,}1$ мм диаметра валиков, изготовленных автоматом №1. По выборке объема m=50 найден средний размер $\overline{y} = 19{,}8$ мм диаметра валиков, изготовленных автоматом №2. Генеральные дисперсии известны $\sigma_x^2 = 1{,}750$ мм^2, $\sigma_y^2 = 1{,}375$ мм^2. Требуется, при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : M(X) = M(Y)$ при конкурирующей гипотезе $H_1 : M(X) \neq M(Y)$. Предполагается, что случайные величины X и Y распределены нормально и выборки независимы.

Решение. Из условия задачи имеем: *n=50*, $\overline{x} = 20{,}1$, $\sigma_x^2 = 1{,}750$; *m=50*, $\overline{y} = 19{,}8$, $\sigma_y^2 = 1{,}375$ мм^2. По формуле (6.4) найдем наблюдаемое значение критерия:

$$z_{набл} = \frac{20{,}1 - 19{,}8}{\sqrt{1{,}75/50 + 1{,}375/50}} = \frac{0{,}3}{\sqrt{0{,}035 + 0{,}0275}} = \frac{0{,}3}{\sqrt{0{,}0625}} = \frac{0{,}3}{0{,}25} = 1{,}2.$$

Так как по условию задачи конкурирующая гипотеза имеет вид $H_1 : M(X) \neq M(Y)$, поэтому критическая область - двусторонняя. Правую критическую точку найдем, воспользовавшись равенством (6.5): $\Phi(z_{кр}) = (1-\alpha)/2 = (1-0{,}05)/2 = 0{,}475$. $\Phi(z_{кр}) = 0{,}475$. По таблице функции Лапласа (приложение 2) находим $z_{кр} = 1{,}96$. Так $|z_{набл}| < z_{кр}$ - нулевая гипотеза принимается. Выборочные средние различаются незначимо.

Сравнение двух средних генеральных совокупностей, дисперсии которых неизвестны

Пусть генеральные совокупности X и Y имеют нормальное распределение с параметрами (a_x, σ_x) и $(a_y,\ \sigma_y)$. Имеются независимые выборки $(x_1, .., x_{n_1})$ и $(y_1, .., y_{n_2})$ объемом n (n<30) и m (m<30) из генеральных совокупностей X и Y, для которых предполагается, что генеральные дисперсии одинаковы. По данным выборкам найдены выборочные средние $\bar{x}$ и $\bar{y}$ и исправленные выборочные дисперсии S_x^2 и S_y^2.

Рассмотрим гипотезу $H_0: a_x = a_y$ о равенстве математических ожидании (генеральных средних) с неизвестными, но одинаковыми дисперсиями. Критерием служит случайная величина

$$T_{набл.} = \frac{\bar{x}-\bar{y}}{\sqrt{(n-1)S_x^2+(m-1)S_Y^2}}\sqrt{\frac{mn(n+m-2)}{n+m}} \qquad (6.7)$$

которая при условии справедливости нулевой гипотезы имеет t-распределение Стьюдента с *k=n+m-2* степенями свободы.

Вид критической области зависит от типа конкурирующей гипотезы:

1. При альтернативе $H_1: a_x \neq a_y$ критическая область двусторонняя: $p(|T| > t_{двуст.кр}(\alpha,k)) = \alpha$. Тогда критическая точка

$$t_{двуст.кр}(\alpha,k) \qquad (6.8)$$

находится по таблице критических точек распределения Стьюдента (приложение 3) по заданному уровню значимости α. Если $|T_{набл}| < t_{двуст.кр}(\alpha,k)$ - гипотезу H_0 принимают. Если $|T_{набл}| > t_{двуст.кр}(\alpha,k)$ - гипотезу H_0 отвергают.

2. При альтернативе $H_1: a_x > a_y$ критическая область правосторонняя: $p(T > t_{правост.кр}(\alpha,k)) = \alpha$. Критическая точка $t_{правост.кр}(\alpha,k)$ находится по таблице критических точек распределения Стьюдента (приложение 3) по заданному уровню значимости α. Если $T_{набл.} < \mathrm{t}_{\text{правост.кр}}(\alpha,k)$ – нулевую гипотезу H_0 принимают. Если $T_{набл.} > \mathrm{t}_{\text{правост.кр}}(\alpha,k)$ – нулевую гипотезу H_0 отвергают.

3. При альтернативе $H_1 : a_x < a_y$ критическая область левосторонняя: $p(T < -t_{правост.кр}(\alpha, k)) = \alpha$. Критическая точка $t_{правост.кр}(\alpha, k)$ вычисляется так же, как в предыдущем случае (пункт 2) и полагают $t_{левост.кр}(\alpha, k) = -t_{правост.кр}(\alpha, k)$. Если $T_{набл.} >$ -t$_{\text{правост.кр}}(\alpha, k)$ – нулевую гипотезу H_0 принимают. Если $T_{набл.} <$ -t$_{\text{правост.кр}}(\alpha, k)$ – нулевую гипотезу H_0 отвергают.

Пример 6.4. По двум независимым выборкам n=12, m=18, извлеченным из нормальных генеральных совокупностей X и Y, найдены выборочные средние $\bar{x} = 31{,}2$, $\bar{y} = 29{,}2$ и исправленные дисперсии $S_x^2 = 0{,}84$ и $S_y^2 = 0{,}4$. Требуется при α=0,05 проверить нулевую гипотезу $H_0 : M(X) = M(Y)$ при альтернативе $H_1 : M(X) \neq M(Y)$.

Решение. Так как выборочные дисперсии различны, предварительно необходимо проверить гипотезу о равенстве генеральных дисперсий, используя критерий Фишера-Снедекора (см.п.5.2). Найдем отношение большей исправленной дисперсии к меньшей:

$$F_{набл.} = \frac{0{,}84}{0{,}4} = 2{,}1.$$

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 4 , по уровню значимости $\alpha = 0{,}05$ и числам степеней свободы $k_1 = 12 - 1 = 11$ и $k_2 = 18 - 1 = 17$ находим критическую точку $F_{кр}(0{,}05;11;17) = 2{,}41$. Так как $F_{набл.} < F_{кр}$ - нет оснований отвергнуть гипотезу о равенстве генеральных дисперсий. Предположение о равенстве генеральных дисперсий выполняется, поэтому сравним средние. Вычислим наблюдаемое значение критерия Стьюдента по формуле (6.7):

$$T_{набл.}=\frac{31,2-29,2}{\sqrt{11*0,84+17*0,4}}\sqrt{\frac{12*18(12+18-2)}{30(12+18)}}=\frac{2}{\sqrt{9,24+6,8}}\sqrt{\frac{36*28}{5}}=\frac{2}{\sqrt{16,04}}\sqrt{201,6}=7,1$$

$T_{набл.}$=7,1. По условию конкурирующая гипотеза имеет вид $H_1: M(X)\neq M(Y)$, поэтому критическая область – двусторонняя. По уровню значимости 0,05 и числу степеней свободы *k=n+m-2=12+18-2=28* находим по таблице приложения 3 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 для самостоятельной работы

6.1. 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 $H_0: D(X)=D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе: $H_1: D(X)>D(Y)$.

6.2. По двум независимым выборкам, объемы которых $n_1=14$ и $n_2=10$, извлеченным из нормальных генеральных совокупностей X и Y, найдены выборочные дисперсии $S_x^2=0{,}84$ и $S_y^2=2{,}52$. При уровне значимости 0,1 проверить нулевую гипотезу $H_0: D(X)=D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1: D(X)\neq D(Y)$.

6.3. Два токарных автомата изготовляют детали по одному чертежу. Из продукции первого станка было отобрано n_1=9 деталей, а из продукции второго n_2=11 деталей. Выборочные дисперсии контрольного размера, определенные по этим выборкам, S_1^2=5,9 мкм^2, S_2^2=23,2 мкм^2. Проверить гипотезу о равенстве дисперсий при $\alpha = 0{,}1$, если альтернативная гипотеза утверждает, что дисперсии не равны.

6.4. Два токарных автомата изготовляют детали по одному чертежу. Из продукции первого станка было отобрано n_1=11 деталей, а из продукции второго n_2=9 деталей. Выборочные дисперсии контрольного размера, определенные по этим выборкам, S_1^2=23,2 мкм^2, S_2^2=5,9 мкм^2, Проверить гипотезу о равенстве дисперсий при $\alpha = 0{,}01$, если альтернативная гипотеза утверждает, что дисперсии размера для первого больше, чем для второго станка.

6.5. Для сравнения точности двух станков-автоматов взяты две пробы (выборки), объемы которых n_1=10 и n_2=8. В результате измерения контролируемого размера отобранных изделий получены следующие результаты:

x_i	1,08	1,10	1,12	1,14	1,15	1,25	1,36	1,38	1,40	1,42
y_i	1,11	1,12	1,18	1,22	1,33	1,35	1,36	1,38		

Можно ли считать, что станки обладают одинаковой точностью [$H_0 : D(X) = D(Y)$], если принять уровень значимости 0,1 и в качестве конкурирующей гипотезы $H_1 : D(X) \neq D(Y)$.

6.6. Двумя методами проведены измерения одной и той же физической величины. Получены следующие результаты:

x_i	9,6	10,0	9,8	10,2	10,6
y_i	10,4	9,7	10,0	10,3	

Можно ли считать, что оба метода обеспечивают одинаковую точность измерений, если принять уровень значимости $\alpha = 0{,}1$? Предполагается, что результаты измерений распределены нормально и выборки независимы.

6.7. По двум независимым выборкам, объемы которых n=40 и m=50, извлеченным из нормальных генеральных совокупностей. Найдены выборочные средние: $\bar{x} = 130$ и $\bar{y} = 140$ мм диаметра валиков, изготовленных автоматом №2. Генеральные дисперсии известны $\sigma_x^2 = 80$, $\sigma_y^2 = 100$. Требуется, при уровне значимости 0,01 проверить нулевую гипотезу $H_0 : M(X) = M(Y)$ при конкурирующей гипотезе $H_1 : M(X) \neq M(Y)$.

6.8. По выборке объема n=30 найден средний вес $\bar{x} = 130$ г изделий, изготовленных на первом станке; по выборке объема m=40 найден средний вес $\bar{y} = 125$ г изделий, изготовленных на втором станке. Генеральные дисперсии известны: $\sigma_x^2 = 60$ г², $\sigma_y^2 = 80$ г². Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : M(X) = M(Y)$ при конкурирующей гипотезе $H_1 : M(X) \neq M(Y)$. Предполагается, что случайные величины X и Y распределены нормально и выборки независимы.

6.9. По двум независимым малым выборкам, объемы которых n=10, m=8, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние $\bar{x} = 142{,}3$, $\bar{y} = 145{,}3$ и исправленные дисперсии $S_x^2 = 2{,}7$ и $S_y^2 = 3{,}2$. При уровне значимости 0,01 проверить нулевую гипотезу $H_0 : M(X) = M(Y)$ при альтернативе $H_1 : M(X) \neq M(Y)$.

6.10. Из двух партий изделий, изготовленных на двух одинаково настроенных станках, извлечены малые выборки, объемы которых n=10 и m=12. Получены следующие результаты:

x_i (контролируемый размер изделий 1 станка)	3,4	3,5	3,7	3,9
n_i (число изделий)	2	3	4	1

y_i (контролируемый размер изделий 2 станка)	3,2	3,4	3,6
n_i (число изделий)	2	2	8

Требуется при уровне значимости 0,02 проверить гипотезу $H_0 : M(X) = M(Y)$ о равенстве средних размеров изделий при конкурирующей гипотезе $H_1 : M(X) \neq M(Y)$. Предполагается, что случайные величины X и Y распределены нормально.

6.11. При уровне значимости 0,05 проверить нулевую гипотезу $H_0: M(X)=M(Y)$ о равенстве генеральных средних нормальных совокупностей X и Y при конкурирующей гипотезе $H_1: M(X)>M(Y)$ по малым независимым выборкам, объемы которых n=10 и m=16. Получены следующие результаты:

x_i	12,3	12,5	12,8	13,0	13,5
n_i	1	2	4	2	1

y_i	12,2	12,3	13,0
m_i	6	8	2

6.12. Давление в камере контролируется по двум манометрам. Для сравнения точности этих приборов одновременно фиксируются их показания. По результатам 10 замеров выборочные оценки оказались следующими: $\overline{x}=15{,}3$, $\overline{y}=16{,}1$, $S_x^2=0{,}2$ и $S_y^2=0{,}15$. Используя двусторонний и односторонний критерии, проверить при $\alpha=0{,}1$: а) гипотезу о равенстве дисперсий; б) гипотезу о равенстве средних.

6.13. На двух станках A и B производят одну и ту же продукцию, контролируемую по внутреннему диаметру изделия. Из продукции станка A была взята выборка из 16 изделий, а из продукции станка B – выборка из 25 изделий. Выборочные оценки средних и дисперсий контролируемых размеров $\overline{x}_A=37{,}5$ мм при $S_A^2=1{,}21$ мм^2 и $\overline{x}_B=36{,}8$ мм при $S_B^2=1{,}44$ мм^2. Используя двусторонний критерий, проверить гипотезу о равенстве математических ожиданий контролируемых размеров в продукции обоих станков, если а) $\alpha=0{,}05$; б) $\alpha=0{,}1$.

7. ПРОВЕРКА ГИПОТЕЗ О ЗАКОНЕ РАСПРЕДЕЛЕНИЯ

7.1. Критерии согласия χ^2 - Пирсона

Одним из наиболее распространенных критериев проверки гипотез о виде закона распределения изучаемой случайной величины является *критерий согласия* χ^2 *- Пирсона*. Данный критерий позволяет проверить значимость расхождения эмпирических (наблюдаемых) и теоретических (ожидаемых) частот. Таким образом с помощью данного критерия можно проверить гипотезу о принадлежности наблюдаемой выборки некоторому теоретическому закону распределения.

Пусть $x_1,..,x_n$– выборка наблюдений случайной величины X объемом n. Проверяется гипотеза H_0, утверждающая, что X имеет функцию распределения $F_\xi(x)$. Проверка гипотезы H_0 при помощи критерия χ^2 осуществляется по следующей схеме.

1. По выборке наблюдений находят оценки неизвестных параметров (например $\bar{x}$, σ) предполагаемого закона распределения случайной величины X.

2. Область возможных значений случайной величины X разбивается на r множеств $\Delta_1, \Delta_2, .., \Delta_r$, например, r интервалов в случае, когда X – непрерывная случайная величина X, или r групп, состоящих из отдельных значений, для дискретной случайной величины X.

3. Пусть n_k - число элементов выборки, принадлежащих множеству Δ_k, $k = 1,2,..,r$. Очевидно, что $\sum_{k=1}^{r} n_k = n$. Используя предполагаемый закон распределения случайной величины X, находят вероятности p_k того, что значение X принадлежит множеству Δ_k, т.е. $p_k = P[X \in \Delta_k]$, $k = 1,2,..,r$. Очевидно, что $\sum_{k=1}^{r} p_k = 1$.

4. Вычисляются теоретические частоты np_k. Для этого удобно полученные результаты представить в виде следующей таблицы:

Число наблюдений в интервале Δ_i	Наблюдаемое значение	Ожидаемое значение
Δ_1	n_1	np_1
Δ_2	n_2	np_2
…	…	…
Δ_r	n_r	np_r
Всего	n	n

5. Сравниваются эмпирические и теоретические частоты с помощью критерия, для этого необходимо:

а) Вычислить выборочное значение статистики критерия χ^2 (наблюдаемое значение)

$$\chi^2_{набл.} = \sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k} \tag{7.1}$$

б) Принять решение о том, что X имеет функцию распределения $F_x(x)$, если $\chi^2_{набл.} < \chi^2_{кр.1-\alpha}(r-l-1)$, где $\chi^2_{1-\alpha}(r-l-1)$ - квантиль порядка $1\text{-}\alpha$ распределения χ^2 с $r\text{-}l\text{-}1$ степенями свободы, а l – число неизвестных параметров распределения, оцениваемых по выборке, r – число групп выборки.

Замечание 1. Малочисленные частоты ($np_k<5$) следует объединить, соответствующие им теоретические частоты также надо сложить.

Замечание 2. При проверке гипотезы о нормальном распределении изучаемой случайной величины теоретические частоты для дискретного вариационного ряда вычисляются по формуле:

$$np_k = \frac{nh}{\sigma}\varphi\left(\frac{x_i - \overline{x}}{\sigma}\right) \tag{7.2}$$

где n-объем выборки (сумма всех частот), h – шаг (разность между двумя соседними измерениями), а значения функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ можно найти в таблице приложения 1.

Для интервального вариационного ряда теоретические частоты можно найти по формуле

$$np_k = n \cdot \left[\Phi\left(\frac{z_{k+1} - \overline{x}}{\sigma}\right) - \Phi\left(\frac{z_k - \overline{x}}{\sigma}\right)\right] \tag{7.3}$$

где n-объем выборки (сумма всех частот), z_k и z_{k+1} - соответственно нижняя и верхняя границы интервалов, причем наименьшее значение интервала полагают равным $-\infty$, а наибольшее $+\infty$. Значения функции Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$ можно найти в таблице приложения 2.

Пример 7.1. В итоге проверки на нестандартность 200 ящиков консервов получено следующее эмпирическое распределение (в первой строке указано количество x_i нестандартных коробок консервов в одном ящике; во второй строке – частота n_i, т.е. число ящиков, содержащих x_i коробок нестандартных консервов):

x_k	0	1	2	3	4
n_k	132	43	20	3	2

Требуется при уровне значимости 0,05 проверить гипотезу о том, что случайная величина X – число нестандартных коробок – распределена по закону Пуассона.

Решение. 1. Найдем оценку параметра λ распределения Пуассона. В примере 2.6. было получено, что $\lambda=\bar{x}$.

Выборочное среднее $\bar{x}$ найдем воспользовавшись формулой (9). Получим $\lambda=\bar{x}=(132\cdot 0+43\cdot 1+20\cdot 2+3\cdot 3+4\cdot 2)/200=0{,}5$.

Следовательно, предполагаемый закон распределения Пуассона $P_n(k)=\frac{\lambda^k}{k!}e^{-\lambda}$ имеет вид $P_{200}(k)=\frac{(0{,}5)^k}{k!}e^{-0{,}5}$.

2. Область значений случайной величины уже задана в виде статистического ряда, поэтому переходим к пункту 3.

3. Положив *k=0, 1, 2, 3, 4,* найдем вероятности P_k появления *k* нестандартных коробок консервов в 200 ящиках:

$$P_0=P_{200}(0)=\frac{(0{,}5)^0}{0!}e^{-0{,}5}=0{,}60653;\ P_1=P_{200}(1)=\frac{(0{,}5)^1}{1!}e^{-0{,}5}=0{,}30327;$$

$$P_2=P_{200}(2)=\frac{(0{,}5)^2}{2!}e^{-0{,}5}=0{,}07582;\ P_4=P_{200}(4)=\frac{(0{,}5)^4}{4!}e^{-0{,}5}=0{,}00158.$$

4. Найдем теоретические частоты по формуле $np_k = n \cdot P_k = 200 \cdot P_k$. Подставив в эту формулу найденные значения вероятностей P_k, получим:

$np_0 = 200 \cdot P_0 = 121{,}306$; $np_1 = 200 \cdot P_1 = 60{,}654$; $np_2 = 200 \cdot P_2 = 15{,}164$; $np_3 = 200 \cdot P_3 = 2{,}529$; $np_4 = 200 \cdot P_4 = 0{,}316$.

Составим расчетную таблицу 5.1. Учитывая замечание 1, объединим малочисленные частоты (3+2=5) и соответствующие теоретические частоты (2,529+0,316=2,845).

Таблица 7.1.

k	n_k	np_k	$n_k - np_k$	$(n_k - np_k)^2$	$(n_k - np_k)^2 / np_k$
0	132	121,306	10,694	114,362	0,943
1	43	60,654	-17,654	311,664	5,138
2	20	15,164	4,836	23,387	1,542
3	5	2,845	2,155	4,644	1,632
$\sum$	200				$\chi^2_{набл.} = 9{,}255$

5. Сравним эмпирические и теоретические частоты с помощью критерия χ^2 - Пирсона.

а) Из расчетной таблицы находим наблюдаемое значение критерия Пирсона: $\chi^2_{набл.} = 9{,}255$.

б) По выборке оценивается один параметр λ, значит число степеней свободы *k=4-1-1=2*. По таблице критических точек распределения χ^2 (приложение 5), по уровню значимости $1-\alpha = 0{,}95$ и числу степеней свободы k=2 находим критическую точку $\chi^2_{0,95}(2) = 5{,}99$, т.е. $\chi^2_{набл.} < \chi^2_{кр.}$ и гипотеза о распределении случайной величины X по закону Пуассона принимается.

Пример 7.2. При уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема n=200.

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Решение. 1. Для проверки гипотезы о нормальном распределении найдем оценки математического ожидания и дисперсии, то есть выборочное среднее и выборочную дисперсию, воспользовавшись формулами (9) и (10):

$$\overline{x}=\frac{1}{200}\sum_{i=1}^{9}x_i\cdot n_i=\frac{1}{200}\cdot(5\cdot 15+7\cdot 26+9\cdot 25+11\cdot 30+13\cdot 26+15\cdot 21+$$

$$+17\cdot 24+19\cdot 20+21\cdot 13)=\frac{2526}{200}=12{,}63$$

$D^{*}=\sigma^2=\dfrac{1}{200}\sum\limits_{i=1}^{9}(x_i-\overline{x})^2\cdot p_i$ =24,5025. Тогда выборочное среднеквадратическое отклонение есть $\sigma=\sqrt{24{,}5025}=4{,}695$.

2. Область значений случайной величины уже задана в виде интервалов, тогда переходим к пункту 3.

3. Вычислим теоретические частоты, учитывая, что n=200, h=2, $\sigma=4{,}695$ по формуле $np_k=\dfrac{nh}{\sqrt{\sigma}}\varphi(u_k)=\dfrac{200\cdot 2}{4{,}695}\varphi(u_k)=85{,}2\cdot\varphi(u_k)$,

где $u_k=(x_i-\overline{x})/\sigma$.

Составим расчетную таблицу 7.2, значения функции $\varphi(u_k)$ помещены в приложении 1.

4. Сравним эмпирические и теоретические частоты с помощью критерия χ^2 - Пирсона.

Таблица 7.2.

K	x_k	$u_k=\frac{x_k-\overline{x}}{\sigma}$	$\varphi(u_k)$	np_k	n_k	n_k-np_k	$(n_k-np_k)^2$	$(n_k-np_k)^2/np_k$
1	5	-1,62	0,1074	9,1	15	5,9	34,81	3,8
2	7	-1,20	0,1942	16,5	26	9,5	90,25	5,5
3	9	-0,77	0,2966	25,3	25	-0,3	0,09	0,0
4	11	-0,35	0,3752	32,0	30	-2,0	4,00	0,1
5	13	0,08	0,3977	33,9	26	-7,9	62,41	1,8
6	15	0,51	0,3503	29,8	21	-8,8	77,44	2,6
7	17	0,93	0,2589	22,0	24	2,0	4,00	0,2
8	19	1,36	0,1582	13,5	20	6,5	42,25	3,1
9	21	1,78	0,0818	7,0	13	6,0	36,00	5,1
$\sum$					200			$\chi^2_{набл.}=22{,}2$

Из расчетной таблицы находим наблюдаемое значение критерия Пирсона: $\chi^2{}_{набл.} = 22{,}2$.

По таблице критических точек распределения χ^2 при уровне значимости $1-\alpha = 0{,}95$ и числу степеней свободы k=s-3=9-3=6 находим критическую точку $\chi^2_{0,95}(6) = 12{,}6$. Т.е. $\chi^2_{набл.} > \chi^2_{кр}$ - следовательно, гипотезу о нормальном распределении генеральной совокупности отвергаем. Другими словами, эмпирические и теоретические частоты различаются значимо.

Пример 7.3. При уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема n=100, приведенным в таблице 4.3.

Таблица 7.3.

Номер интервала	Границы интервала	Частота с
1	3-8	6
2	8-13	8
3	13-18	15
4	18-23	40
5	23-28	16
6	28-33	8
7	33-38	7

Решение. Вычислим выборочное среднее и выборочное среднеквадратическое отклонение. Для этого перейдём от заданного интервального распределения к статистическому распределению, в качестве x_i^* примем среднее арифметическое концов интервала $x_i^* = (x_i + x_{i+1})/2$. В итоге получим распределение:

Середина интервала x_i^*	5,5	10,5	15,5	20,5	25,5	30,5	35,5
Частота x_i^*	6	8	15	40	16	8	7

Найдем теоретические частоты np_k, учитывая, что $np_k = n \cdot \left[\Phi\left(\frac{x_{k+1} - \bar{x}}{\sigma} \right) - \Phi\left(\frac{x_k - \bar{x}}{\sigma} \right) \right]$. Для этого составим расчетную таблицу 7.4 (левый конец первого интервала примем равным $-\infty$, а правый конец последнего интервала $+\infty$).

Таблица 7.4

К	Границы интервала	$z_k = \frac{x_k - \bar{x}}{\sigma}$	$z_{k+1} = \frac{x_{k+1} - \bar{x}}{\sigma}$	$\Phi(z_k)$	$\Phi(z_{k+1})$	$\Phi(z_{k+1})$ -$\Phi(z_k)$	np_k
1	$-\infty$ - 8	$-\infty$	-1,74	-0,5000	-0,4591	0,0409	4,09
2	8-13	-1,74	-1,06	-0,4591	-0,3554	0,1037	10,37
3	13-18	-1,06	-0,37	-0,3554	-0,1443	0,2111	21,11
4	18-23	-0,37	0,32	-0,1443	0,1255	0,2698	26,98
5	23-28	0,32	1,00	0,1255	0,3413	0,2158	21,58
6	28-33	1,00	1,69	0,3413	0,4545	0,1132	11,32
7	33-$+\infty$	1,69	$+\infty$	0,4545	0,5000	0,0455	4,55
$\sum$						1	100

Сравним эмпирические и теоретические частоты с помощью критерия χ^2 - Пирсона. Для этого составим расчетную таблицу 7.5, значения функции $\Phi(z_k)$ помещены в приложении 2.

Таблица 7.5

К	n_k	np_k	$n_k - np_k$	$(n_k - np_k)^2$	$\frac{(n_k - np_k)^2}{np_k}$	n_k^2	n_k^2 / np_k
1	6	4,09	1,91	3,6481	0,8920	36	8,8019
2	8	10,37	-2,37	5,6169	0,5416	64	6,1716
3	15	21,11	-6,11	37,3321	1,7684	225	10,6584
4	40	26,98	13,02	169,5204	6,2833	1600	59,3052
5	16	21,58	-5,58	31,1364	1,4428	256	11,8628
6	8	11,32	-3,32	11,0224	0,9737	64	5,6537
7	7	4,55	2,45	6,0025	1,3192	49	10,7692
$\sum$	100	100			$\chi^2_{набл.} = 13{,}22$		113,22

Из расчетной таблицы находим наблюдаемое значение критерия Пирсона: $\chi^2_{набл.} = 13{,}22$.

Столбцы 7 и 8 служат для контроля вычислений по формуле

$\chi^2_{набл.} = \sum(n_k^2 / np_k) - n$. Контроль: $\chi^2_{набл.} = \sum_{k=1}^{7}(n_k^2 / np_k) - n = 113{,}2 - 100 = 13{,}22$. Вычисления произведены правильно.

По таблице критических точек распределения χ^2 при уровне значимости $1-\alpha = 0{,}95$ и числу степеней свободы *k=s-3=7-3=4* (число интервалов) находим критическую точку $\chi^2_{0,95}(4) = 9{,}5$. Т.е. $\chi^2_{набл.} > \chi^2_{кр}$ - следовательно, гипотезу о нормальном распределении генеральной совокупности отвергаем. Другими словами, эмпирические и теоретические частоты различаются значимо. Это означает. Что данные наблюдений не согласуются с гипотезой о нормальном распределении генеральной совокупности.

7.2. Критерий согласия Колмогорова

Критерий согласия Колмогорова применяют для проверки гипотез о законах распределения только непрерывных случайных величин. Проверяем гипотезу $H_0 : F_\xi(x) = F(x)$ против альтернативы $H_1 : F_\xi(x) \neq F(x)$, где *F(x)* – однозначно заданная и непрерывная функция. Критерий основан на том факте, что распределение супремума разности между теоретической и эмпирической функциями распределения $d_n = \sup|F_n(x) - F(x)|$ одинаково для любой $F(x)$. Величину d_n называют статистикой Колмогорова.

Пусть заданы уровень значимости α и случайная выборка независимых наблюдений $x_1, x_2, .., x_n$. Из данной выборки строится вариационный ряд $x_{(1)}, x_{(2)}, .., x_{(n)}$ и вычисляется выборочное значение $d_{n,набл}$ статистики d_n для этого ряда по формуле

$$d_n = \max_{1 \le k \le n}\left|\frac{k}{n} - F(x_{(k)}), F(x_{(k)} - \frac{k-1}{n})\right| \qquad (7.4)$$

Выборочное значение $d_{n,набл}$ сравнивают с полученным из таблиц критических точек.

При малых n для статистики d_n существуют таблицы критических точек $d_{кр}$, (приложение 6). Если $d_{n,набл} < d_{кр}$, то гипотеза H_0 принимается, иначе отвергается. При больших n используют предельное распределение Колмогорова: $P(\sqrt{n}D_n < x) \to Q(x) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2x^2}$, $n \to \infty$.

Для распределения Колмогорова $Q(x)$, предельного для статистики $\lambda = \sqrt{n}d_n$, также существуют таблица критических точек $\lambda_{кр}$. Практически их используют уже при $n > 20$. Если $\lambda < \lambda_{кр}$, то гипотеза H_0 принимается, иначе отвергается.

Пример 7.4. Имеется следующая выборка из 10 наблюдений: 0,01; 0,09; 0,1; 0,25; 0,33; 0,35; 0,52; 0,73; 0,76; 0,86. Проверить с помощью критерия Колмогорова гипотезу о том, что это выборка из генеральной совокупности, равномерно распределенной на отрезке [0,1]. Уровень значимости $\alpha = 0{,}05$

Решение. С учетом того, что по условию задачи $F_0(x) = x/10$, $0 \le x \le 10$, построим таблицу:

k	$x_{(k)}$	$F_n(x) = \frac{k}{n}$	$F(x_k) = t_k$	$\frac{k}{n} - t_k$	$t_k - \frac{k-1}{n}$
1	0,01	0,1	0,01	0,09	0,01
2	0,09	0,2	0,09	0,11	-0,01
3	0,1	0,3	0,1	0,2	-0,1
4	0,25	0,4	0,25	0,15	-0,05
5	0,33	0,5	0,33	0,17	0,07
6	0,35	0,6	0,35	0,25	-0,15
7	0,52	0,7	0,52	0,18	-0,08
8	0,73	0,8	0,73	0,07	0,03
9	0,76	0,9	0,76	0,14	-0,04
10	0,86	1	0,86	0,14	-0,04

Из таблицы получаем $d_{n,набл} = 0{,}25$. По уровню значимости $\alpha = 0{,}05$ находим критическую точку $d_{кр} = 0{,}41$. Поскольку $d_{набл} < d_{кр}$, то гипотеза принимается.

7.3. Критерий однородности χ^2

Критерий однородности χ^2 используют для проверки однородности дискретных данных, т.е. когда в опытах наблюдается некоторый переменный признак, принимающий конечное число различных значений Известно, что к такой схеме можно свести любую другую модель, применяя предварительно метод группировки данных. Поэтому критерий χ^2 применим к анализу любых данных, т.е. является универсальным.

Пусть имеется k серий опытов, состоящих из наблюдений за случайной величиной ξ, которая может принимать одно из m возможных значений (χ^2 - дискретная случайная величина). Серии k наблюдений характеризуются выборками

$X = (x_1, x_2, .., x_{n1})$ из генеральной совокупности с $F_1(u)$,

$Y = (y_1, y_2, .., y_{n2})$ из генеральной совокупности с $F_2(u)$,

…………………………………………………………,

$Z = (z_1, z_2, .., z_{nk})$ из генеральной совокупности с $F_k(u)$

Требуется проверить гипотезу H_0 о том, что все наблюдения производились над одной и той же случайной величиной

$H_0 : F_1(u) = F_2(u) = ... = F_k(u)$

Против альтернативы H_1: хотя бы одно распределение не равно остальным.

Статистикой критерия является величина:

$$G = n\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{v_{ij}^2}{v_i n_j} - 1\right) \tag{7.4}$$

где v_{ij} - число появлений i-го значения в j-ой серии; $v_i = \sum_{j=1}^{k} v_{ij}$ - общее число появлений i-го значения во всех сериях; n_j - объем j-ой серии; $n = \sum_{j=1}^{k} n_j$ - общее число наблюдений во всех сериях; n_j - объем j – ой серии;

При $n \to \infty$ и справедливости основной гипотезы статистика G будет иметь распределение χ^2 с (m-1)(k-1) степенями свободы, т.е. $G \underset{n\to\infty}{\overset{F}{\longrightarrow}} \chi^2_{(m-1)(k-1)}$

Отсюда следует критерий проверки гипотезы:

1. Вычислить значение статистики $g_{набл.}$.
2. При заданном уровне значимости α по таблицам найти значение $\chi^2_{кр} = \chi^2_{1-\alpha,(m-1)(k-1)}$ - квантиль распределения χ^2 с числом степеней свободы *(m-1)(k-1)* порядка $1-\alpha$.
3. Если $g_{набл.} \geq \chi^2_{кр}$, то гипотеза H_0 отвергается, если $g_{набл.} < \chi^2_{кр}$ гипотеза H_0 принимается.

Замечание. Критерий χ^2 применяется для непрерывной случайной величин ξ, тогда область возможных значений ξ можно разбить на подинтервалы и v_{ij} будет числом значений в j-той серии, попавших в i-ый подинтервал.

Пример 7.5. Два игрока бросали монету по 100 раз. У первого - герб выпал 57 раз; у второго - 48. Проверить гипотезу о том, что монеты идентичны, при уровне значимости 0.05.

Решение. Предполагаем, что исследуется случайная величина ξ, показывающая число выпавших гербов при одном подбрасывании.

ξ_i	*0*	*1*
p_i	*1-p*	*p*

Число возможных значений случайной величины ξ *m=2,* число серий (число игроков) *k=2,* $n_1 = n_2 = 100$, $n = n_1 + n_2 = 200$. Следовательно

$\nu_{11} = 57$ $\nu_{12} = 48$ - выпадение герба

$\nu_{21} = 43$ $\nu_{22} = 52$ - выпадение решки

$\nu_1 = 57 + 48$ - число появлений герба во всех опытах

$\nu_1 = 43 + 52$ - число появлений решки во всех опытах

Теперь можно вычислить значение статистики $g_{набл.}$. Имеем

$g_{набл.} = 200 \cdot \left(\frac{\nu_{11}^2}{n_1\nu_1} + \frac{\nu_{12}^2}{n_2\nu_1} + \frac{\nu_{21}^2}{n_1\nu_2} + \frac{\nu_{22}^2}{n_2\nu_2} - 1 \right) \approx 16$. По таблицам распределения χ^2 найдем $\chi^2_{кр} = \chi^2_{0,95,1} = 3{,}8$. Так как $g_{набл.} \geq \chi^2_{кр}$, значит основная гипотеза H_0 отвергается и принимается альтернативная, т.е. монеты нельзя считать одинаковыми.

7.4. Критерий однородности Колмогорова-Смирнова

Пусть $X = (x_1, x_2, .., x_n)$ и $Y = (y_1, y_2, .., y_n)$ - выборки из генеральных совокупностей с некоторыми неизвестными функциями $F_1(x)$ и $F_2(x)$ соответственно. Требуется проверить гипотезу о том, что выборки взяты из одной и той же генеральной совокупности, т.е. выдвигаются гипотезы

$H_0 : F_1(x) = F_2(x)$,

$H_1 : F_1(x) \neq F_2(x)$

Критерий Колмогорова-Смирнова применяют в случае непрерывных распределений. Он использует статистику

$$d_{nm} = \sqrt{\frac{n \cdot m}{n + m}} \sup\left| F_n(x) - F_m(x) \right| \qquad (7.5)$$

где $F_n(x)$ и $F_m(x)$ - эмпирические функции распределения, построенные по выборкам X и Y соответственно.

Отсюда следует критерий проверки гипотезы:

1. Вычислить значение статистики (5.5) $d_{набл.}$.

2. По заданному уровню значимости α найти значение $k_{кр} = K_{\alpha}$ - квантиль распределения Колмогорова порядка α .

3. Если $d_{набл.} \geq k_{кр}$, то гипотеза H_0 отвергается, если $d_{набл.} < k_{кр}$ гипотеза H_0 принимается.

Задачи для самостоятельной работы

7.1. При 50 подбрасываниях монеты герб появился 20 раз. Можно ли считать монету симметричной? Принять α =0,10.

7.2. При 120 бросаниях игральной кости шестерка выпала 40 раз. Согласуется ли этот результат с утверждением, что кость правильная? Принять α =0,05.

7.3. Метод получения случайных чисел был применен 250 раз. При этом получены следующие результаты:

Цифра	0	1	2	3	4	5	6	7	8	9
Частота появления	27	18	23	31	21	23	28	25	22	32

Можно ли считать, что примененный метод действительно дает случайные числа? Принять α =0,1.

7.4. На экзамене студент отвечает только на один вопрос по одной из трех частей курса. Анализ вопросов, заданных 60 студентам, показал, что 23 студента получили вопросы из первой, 15 – из второй и 22 – из третьей части курса. Можно ли считать, что студент, идущий на экзамен, с равной вероятностью получит вопрос по любой из трех частей курса? Принять $\alpha = 0{,}01$.

7.5. Для определения зависимости цвета волос жителей от их местожительства были обследованы три группы людей из районов А, В, С. Свидетельствуют ли приводимые ниже результаты обследования о зависимости цвета волос жителей от их местожительства? Принять $\alpha = 0{,}05$

.

Район	Цвет волос		
Частота появления	Рыжий	Светлый	Темный
А	2	9	9
В	3	6	21
С	15	15	20

7.6. В цехе с 10 станками регистрировалось число вышедших из строя станков. Всего было проведено 200 наблюдений, результаты которых приведены ниже:

Число выбывших станков	0	1	2	3	4	5	6	7	8	9	10
Число зарегистрированных случаев	41	62	45	22	16	8	4	2	0	0	0

Проверить гипотезу H_0 о том, что число выбывших из строя станков имеет распределение Пуассона. Принять $\alpha = 0{,}05$.

7.7. Для определения засоренности партии семян клевера семенами сорняков было проверено 1000 случайно отобранных проб и получено следующее эмпирическое распределение (в первой строке указано количество x_i семян сорняков в одной пробе; во второй строке – частота n_i, т.е. число проб, содержащих x_i семян сорняков):

x_k	0	1	2	3	4	5	6
n_k	405	366	175	40	8	4	2

Требуется при уровне значимости 0,01 проверить гипотезу о том, что случайная величина X (число семян сорняков) распределен по закону Пуассона.

7.8. В результате проверки 500 контейнеров со стеклянными изделиями установлено, что число поврежденных изделий X имеет следующее эмпирическое распределение (в первой строке указано количество x_i поврежденных изделий в одном контейнере; во второй строке – частота n_i, т.е. число контейнеров, содержащих x_i поврежденных изделий):

x_k	0	1	2	3	4	5	6	7
n_k	199	169	87	31	9	3	1	1

Требуется при уровне значимости 0,01 проверить гипотезу о том, что случайная величина X - число поврежденных изделий распределена по закону Пуассона.

7.9. Используя критерий Пирсона, при уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема n=200:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

7.10. Используя критерий Пирсона, при уровне значимости 0,01 установить, случайно или значимо расхождение между эмпирическими частотами n_i и теоретическими частотами n'_i , которые вычислены, исходя из гипотезы о нормальном распределении генеральной совокупности X.

а)

n_i	8	16	40	72	36	18	10
n'_i	6	18	36	76	39	18	7

б)

n_i	14	18	32	70	20	36	10
n'_i	10	24	34	80	18	22	12

В задачах для приведенных группированных выборок, приняв 10%-ный уровень значимости проверить гипотезу о том, что они получены из нормально распределенной генеральной совокупности.

7.11. 200 отклонений размера вала от номинального значения (мкм):

Середина интервала	-0,14	-0,12	-0,10	-0,08	-0,06	-0,04	-0,02
частота	3	8	11	20	27	36	29

Середина интервала	0,00	0,02	0,04	0,06	0,08	0,10	0,12
частота	18	17	17	8	4	1	1

7.12. Входное сопротивление 130 электронных ламп (Ом):

Границы интервала	3,0-3,6	3,6-4,2	4,2-4,8	4,8-5,4	5,4-6,0	6,0-6,6	6,6-7,2
Частота	3	8	11	20	27	36	29

7.13. Рост 1004 девушек в возрасте 16 лет (см)

Границы интервала	134-137	137-140	140-143	143-146	146-149	149-152	152-155
частота	1	4	16	53	121	197	229

Границы интервала	155-158	158-161	161-164	164-167	167-170	170-173
частота	186	121	53	17	5	1

7.14. Пассажир, приходящий в случайные моменты времени на автобусную остановку, в течение пяти поездок фиксировал время ожидания автобуса: 5,1; 3,7; 1,2; 9,2; 4,8 мин. Проверить гипотезу о том, что время ожидания равномерно распределено на отрезке [0,10]. На уровне значимости $\alpha = 0{,}05$.

7.15. За 6 рабочих дней спрос на некоторый товар составлял: 104; 80; 96; 120; 113; 82. На уровне значимости $\alpha = 0{,}1$ проверить гипотезу о том, что спрос равномерно распределен на отрезке [75,125].

7.16. В таблице приведены сгруппированные данные по срокам службы x_i для 1000 изделий:

x_i, час	0-10	10-20	20-30	30-40	40-50	50-60	60-70
n_i	365	245	150	100	70	45	25

На уровне значимости $\alpha = 0{,}05$ проверить гипотезу, что срок службы изделия имеет показательное распределение.

7.17. В таблице приведены условные данные о заработной плате $n_1 = 100$ и $n_2 = 100$ служащих двух отраслей народного хозяйства. Проверить с помощью критерия Колмогорова гипотезу о том, что распределение заработной платы служащих первой отрасли $(F_1(x))$ совпадает с распределением заработной платы служащих второй отрасли $(F_2(x))$. Уровень значимости $\alpha = 0{,}05$.

Зарплата	x_i	y_i
150	4	1
170	4	1
200	15	8
250	51	43
300	22	34
350	3	7
400	1	3
500	-	3

7.18. При изучении творческой активности студентов были получены результаты экспериментальных и контрольных групп. Являются ли значимыми различия между контрольной и экспериментальной группами при $\alpha = 0{,}05$?

Уровень усвоения	Хороший	Приблизительный	Плохой
Частота в экспериментальной группе	172	36	15
Частота в контрольной группе	120	49	36

8. НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

В практике обработки результатов наблюдений часто распределение генеральной совокупности неизвестно либо (для непрерывных случайных величин) отличается от нормального распределения. В этих случаях применяют методы, не зависящие от распределения генеральной совокупности, называемые также непараметрическими методами.

Данные методы используют не сами численные значения элементов выборки, а структурные свойства выборки (например, отношения порядка между ее элементами). В связи с этим теряется часть информации, содержащаяся в выборке, поэтому мощность непараметрических критериев меньше, чем мощность их аналогов. Однако непараметрические методы могут применяться при более общих предположениях и более просты с точки зрения выполнения вычислений.

Большая группа непараметрических критериев используется для проверки гипотезы о принадлежности двух выборок $x_1, x_2,.., x_n$ и $y_1, y_2,.., y_n$ одной и той же генеральной совокупности, т.е. $H_0 : F_1(x) = F_2(x)$. Такие генеральные совокупности называют однородными. В качестве альтернатив рассматриваются гипотезы: $H_1 : F_1(x) \neq F_2(x)$; или $H_1 : F_1(x) < F_2(x)$; или $H_1 : F_1(x) > F_2(x)$. Заметим, что принятие конкурирующих гипотез $H_1 : F_1(x) < F_2(x)$ и $H_1 : F_1(x) > F_2(x)$ означает, что X>Y и X<Y соответственно.

8.1. Критерий знаков

Данный критерий одним из наиболее простых и грубых критериев проверки гипотезы об однородности распределения случайных величин X и Y. Критерий знаков применяется, когда обе выборки имеют одинаковый объем (n=m). Происходит попарное сравнение элементов выборок: (x_1, y_1), (x_2, y_2),.., (x_n, y_m). Фактически здесь имеется выборка для двумерной случайной величины *(X,Y)*. Пары (x_i, y_i) *i=1,2,..,n* являются взаимно независимыми, но компоненты внутри пары могут зависеть друг от друга.

Итак, проверяемая гипотеза утверждает, что x_i и y_i распределены одинаково, т.е. *P(X<Y)=P(X>Y)=1/2,* или *P(Z<0)=P(Z>0)=1/2,*
где $z_i = x_i - y_i$, $i = 1,2..,l$, l – число ненулевых разностей. Следовательно, с равной вероятностью *1/2* знак любого элемента последовательности $z_1, z_2,.., z_n$ может быть положительным и отрицательным.

Задача сводится к проверке гипотезы $H_0 : p = 1/2$ при одной из альтернатив групп $H_1^{(1)} : p > 1/2$, $H_1^{(2)} : p < 1/2$, $H_1^{(3)} : p \neq 1/2$. Критерием r здесь является число положительных знаков z_i. Если гипотеза верна, то r имеет биномиальное распределение с параметрами *l* и *p=1/2*, т.е. *B(l,1/2)*. Часто более удобно проводить проверку гипотезы H_0, используя статистику Фишера

Гипотеза H_0 отклоняется, если

1. при альтернативе $H_1^{(1)} : p > 1/2$ выполняется неравенство

   $$F_в = \frac{r}{l - r + 1} \geq F_{1-\alpha}(k_1, k_2) \qquad (8.1),$$

   где $k_1 = 2(l - r + 1)$ и $k_2 = 2r$
2. при альтернативе $H_1^{(2)} : p < 1/2$ выполняется неравенство

   $$F_в = \frac{l - r}{r + 1} \geq F_{1-\alpha}(k_1, k_2) \qquad (8.2),$$

   где $k_1 = 2(r + 1)$ и $k_2 = 2(l - r)$
3. при альтернативе $H_1^{(3)} : p \neq 1/2$ должно выполняться одно из неравенств (8.1)-(8.2) с заменой α на 2α .

Если при соответствующих альтернативных гипотезах равенства (8.1)-(8.2) не выполняются, то гипотеза H_0 не противоречит результатам наблюдений и принимается на уровне значимости α .

8.4. Критерий однородности Вилкоксона

Критерий применяется для сравнения двух независимых выборок $x_1, x_2, .., x_n$ и $y_1, y_2, .., y_m$, и проверяет гипотезу H_0, утверждающую, что выборки получены из однородных генеральных совокупностей.

Нулевая гипотеза утверждает, что функции распределения равны $H_0 : F_1(x) = F_2(x)$. Альтернативными являются следующие гипотезы: $H_1 : F_1(x) \neq F_2(x)$; $H_1 : F_1(x) < F_2(x)$; $H_1 : F_1(x) > F_2(x)$. Предполагается, что объем первой выборки меньше объема второй: $n \leq m$. Если это не так, то выборки можно перенумеровать (поменять местами).

Для того, чтобы при заданном уровне значимости $\alpha = 2Q$ проверить нулевую гипотезу $H_0 : F_1(x) = F_2(x)$, необходимо

а) Расположить обе выборки в возрастающем порядке, т.е. в виде одного вариационного ряда и найти наблюдаемое значение критерия $W_{набл.}$ - сумму порядковых номеров элементов первой выборки

б) Найти по таблице приложения 8 нижнюю критическую точку $w_{ниж.кр.}(Q,n,m)$, где $Q = \alpha / 2$;

в) найти верхнюю критическую точку по формуле

$w_{верх.кр.} = (n+m+1)n - w_{нижн.кр.}$ (8.3)

1. При альтернативе $H_1 : F_1(x) \neq F_2(x)$ критическая область двусторонняя. Если $W_{набл.} < w_{нижн.кр.}$ или $W_{набл.} > w_{верх.кр.}$ - нулевая гипотеза отвергается. Если $w_{нижн.кр.} < W_{набл.} < w_{верхн.кр.}$ - нулевая гипотеза принимается.

2. При альтернативе $H_1 : F_1(x) > F_2(x)$ необходимо найти по таблице нижнюю критическую точку $w_{ниж.кр.}(Q,n,m)$, где $Q = \alpha$. Если $W_{набл.} > w_{нижн.кр.}$ - нулевая гипотеза принимается. Если $W_{набл.} < w_{нижн.кр.}$ - нулевую гипотезу отвергают.

3. При альтернативе $H_1 : F_1(x) < F_2(x)$ необходимо найти верхнюю критическую точку $w_{верх.кр.}(Q,n,m) = (n+m+1)n - w_{нижн.кр.}(Q,n,m)$ (8.4), где $Q = \alpha$. Если $W_{набл.} < w_{верх.кр.}$ - нулевую гипотезу принимают. Если $W_{набл.} > w_{верх.кр.}$ - нулевую гипотезу отвергают.

Замечание 1. Если несколько элементов только одной выборки одинаковы, то в общем вариационном ряду их нумеруют как различные числа. Если совпадают элементы разных выборок, то им присваивают один и тот же порядковый номер, равный среднему арифметическому порядковых номеров, которые имели бы эти элементы до совпадения.

Замечание 2. Если объем хотя бы одной из выборок больше 25, тогда для нахождения критической точки используется формула

$$w_{ниж.кр.}(Q,n,m\ )=\left[\frac{(n+m+1)n-1}{2}-z_{кр}\sqrt{\frac{nm(n+m+1)}{12}}\right] \qquad (8.5)$$

1. При альтернативе $H_1: F_1(x) \neq F_2(x)$ нижняя критическая точка определяется по формуле (6.5), где $Q=\alpha/2$; $z_{кр}$ находят по таблицам функции Лапласа по равенству $\Phi(z_{кр})=\dfrac{(1-\alpha)}{2}$, знак [ ] означает целую часть числа.

2. При альтернативах $H_1: F_1(x) > F_2(x)$ или $H_1: F_1(x) < F_2(x)$ нижнюю критическую точку находят по формуле (8.5), положив $Q=\alpha$; $z_{кр}$ находят по таблице функции Лапласа по равенству $\Phi(z_{кр})=\dfrac{(1-2\alpha)}{2}$.

8.5. Критерий Манна-Уитни

Также, как и критерий Вилкоксона, применяется для проверки однородности двух независимых выборок. В качестве статистики U критерий Манна-Уитни выступает статистика

$$U=\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi(x_i,y_j),\ \text{где}\ \varphi(a,b)=\begin{cases}1, \textit{если}\ a<b,\\ 0, \textit{в противном случае}\end{cases}$$

Статистику U можно вычислять следующим образом. Для каждой пары значения x_i и y_j смотрим, какое из них меньше: если x_i меньше, то паре приписываем единицу; а если y_j меньше, то – 0. Складываем нули и единицы и обозначаем сумму U. Применение критериев Вилкоксона и Манна Уитни дает одинаковые результаты в силу того, что для всех m и n статистики U и W этих критериев связаны равенством

$$W = U + \frac{n(n+1)}{2}, \text{ или } W + U = nm + \frac{n(n+1)}{2}$$

Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0 : F_1(x) = F_2(x)$, необходимо

1. Расположить обе выборки в порядке возрастания, т.е. в виде одного вариационного ряда. Из общего вариационного ряда определить сумму R_x порядковых номеров первой выборки, и сумма R_y порядковых номеров второй выборки.

2. Вычислить значения U_x и U_y по формулам:

$$U_1 = nm + \frac{1}{2}n(n+1) - R_1 \qquad (8.6)$$

$$U_2 = nm + \frac{1}{2}m,(m+1) - R_2. \qquad (8.7)$$

Правильность вычислений проверяется по формуле $U_1 + U_2 = nm$ (8.8).

3. Найти наблюдаемое значение критерия $U_{набл.}$ - меньшее из чисел U_1 и U_2, т.е. $U_{набл.} = \min(U_1, U_2)$.

4. Найти по таблице статистики U критерия Манна-Уитни (см. Приложение 9) критическую точку $U_{кр}(\alpha, n, m)$.

5. если $U_{набл.} > U_{кр}$, то на уровне значимости α принимается гипотеза H_0. Если же $U_{набл.} \le U_{кр}$, то гипотеза H_0 отвергается на уровне значимости α (принимается альтернативная гипотеза). Достоверность различий тем выше, чем меньше значение $U_{набл.}$.

Замечание. При достаточно большом объеме выборочных данных (n>19) проверку гипотезы H_0 можно проводить, используя статистику

$$Z_{набл.} = \frac{U_{набл.} - M_u}{\sigma_u}, \ (8.9)$$

где $M_u = \dfrac{nm}{2}$, $\sigma_u = \sqrt{\dfrac{nm(n+m+1)}{12}}$

имеющую (при условии, что верна гипотеза H_0) приблизительно нормальное распределение *N(0,1)*. В этом случае гипотеза H_0 отклоняется на уровне значимости α, если $Z_{набл.} < u_\alpha$ $(Z_{набл.} > u_{1-\alpha})$ при левосторонней (правосторонней) альтернативной гипотезе H_1 и если $|Z_{набл.}| < u_{1-\alpha/2}$ при двусторонней альтернативной гипотезе H_1. u_α - квантиль нормального распределения *N(0,1)*.

Пример 8.1. Предполагается, что один из двух приборов, определяющих скорость автомобиля, имеет систематическую ошибку. Для проверки этого предположения определили скорость 10 автомобилей, причем скорость каждого фиксировалась одновременно двумя приборами. В результате получены следующие данные:

υ_1, км/ч	70	85	63	54	65	80	75	95	52	55
υ_2, км/ч	72	86	62	55	63	80	78	90	53	57

Позволяют ли эти результаты утверждать, что второй прибор действительно дает завышенные значения скорости при $\alpha = 0{,}1$.

Решение. В предположении, что скорости движения автомобилей не зависят друг от друга, задачу можно решить, применяя критерий знаков.

Составим последовательность знаков разностей $\upsilon_1 - \upsilon_2$: - , -, +, -, +, 0, -, +, -, -. Число ненулевых разностей *l=9*, число положительных разностей *r=3*. Проверим гипотезу о том, что различие в показаниях приборов вызвано случайными ошибками, т.е. гипотезу $H_0 : p = 1/2$. Альтернативная гипотеза предполагает, что показания второго прибора имеют положительное смещение; в этом случае вероятность появления положительных разностей должна быть меньше ½. Таким образом, альтернативная гипотеза формулируется так: $H_1 : p < 1/2$. Для проверки гипотезы H_0 используем неравенство (8.2). Имеем

$$k_1 = 2 \cdot (3+1) = 8,\ k_2 = 2 \cdot (9-3) = 12,\ F_в = \frac{9-3}{3+1} = 1{,}5,\ F_{кр}(0{,}09;8;12) = 2{,}24.$$

Так как $F_{кр} > F_{в}$, значит гипотеза H_0 не противоречит результатам наблюдений. Следует считать, что различие в показаниях приборов вызвано случайными ошибками.

Пример 8.2. При уровне значимости $\alpha = 0{,}05$ проверить гипотезу $H_0 : F_1(x) = F_2(x)$ об однородности двух выборок объемов $n_1 = 6$, $n_2 = 8$ при альтернативе $H_1 : F_1(x) \neq F_2(x)$.

x_i	15	23	25	26	28	29		
y_i	12	14	18	20	22	24	27	30

Решение. Их двух выборок построим один вариационный ряд, элементы пронумеруем (перенумеруем).

Порядковый номер	1	2	$\underline{3}$	4	5	6	$\underline{7}$	8	$\underline{9}$	$\underline{10}$	11	$\underline{12}$	$\underline{13}$	14
Вариационный ряд	12	14	15	18	20	22	23	24	25	26	27	28	29	30

Найдем наблюдаемое значение критерия Вилкоксона – сумму порядковых номеров (они подчеркнуты) элементов первой выборки.

$W_{набл.}$=3+7+9+10+12+13=54

Найдем по таблицам приложения 8 нижнюю критическую точку, учитывая, что $Q = \alpha / 2 = 0{,}05 / 2 = 0{,}025$, $n_1 = 6$, $n_2 = 8$

$$w_{ниж.кр.}(0{,}025;6;8) = 29$$

Найдем верхнюю критическую точку

$$w_{верх.кр.} = (n_1 + n_2 + 1)n_1 - w_{нижн.кр.} = (6 + 8 + 1) \cdot 6 - 29 = 61.$$

Так как 29<54<61, т.е. $w_{нижн.кр.} < W_{набл.} < w_{верхн.кр.}$ - гипотеза об однородности выборок принимается.

Пример 8.3. Измерялось напряжение пробоя у диодов, отобранных случайным образом из двух партий. Результаты измерения (в вольтах) следующие

1-я партия	39	50	61	67	40	40	54	
2-я партия	60	53	42	41	40	54	63	69

Решение. Составим вариационный ряд, отмечая принадлежность элемента к первой партии чертой снизу. В результате получим следующую ранжированную последовательность:

Элемент	$\underline{39}$	$\underline{40}$	$\underline{40}$	40	41	42	$\underline{50}$	53	$\underline{54}$	54	60	$\underline{61}$	63	$\underline{67}$	$\underline{69}$
Ранг	1	3	3	3	5	6	7	8	9,5	9,5	11	12	13	14	15

Сумма рангов первой выборки $R_1 = 49{,}5$, сумма рангов второй выборки $R_2 = 70{,}5$, $n_1 = 7$, $n_2 = 8$. По формулам (6.6) и (6.7) находим

$$U_1 = 7 \cdot 8 + \frac{7 \cdot (7+1)}{2} - 49{,}5 = 34{,}5,$$

$$U_2 = 7 \cdot 8 + \frac{8 \cdot (8+1)}{2} - 70{,}5 = 21{,}5.$$

Используя соотношение (8.8), проверяем правильность вычислений:

34,5+21,5=56.

Теперь найдем наблюдаемое значение критерия $U_{набл.} = \min(34{,}5;21{,}5) = 21{,}5$

Найдем по таблицам приложения 9 критическую точку $U_{кр}(0{,}01{,}7{,}8) = 6$. Так как $U_{набл.} > U_{кр}$, значит гипотеза H_0 не противоречит результатам измерений. Следовательно, результаты измерений не дают оснований считать, что напряжение пробоя у диодов второй партии выше, чем у диодов первой партии.

Задачи для самостоятельной работы

8.1. Ниже приводится время (в секундах) решения контрольных задач одиннадцатью учащимися до и после специальных упражнений по устному счету. Можно ли считать, что эти упражнения улучшили способности учащихся в решении задач? Принять $\alpha = 0{,}1$.

До упражнений	87	61	98	90	93	74	83	72	81	75	83
После упражнений	50	45	79	90	88	65	52	79	84	61	52

8.2. Для 10 человек была предложена специальная диета. После двухнедельного питания по этой диете масса тела изменилась следующим образом? Принять $\alpha = 0{,}1$.

Масса до диеты (кг)	68	80	92	81	70	79	78	66	57	76
Масса после диеты (кг)	60	84	87	79	74	71	72	67	57	70

Можно ли рекомендовать эту диету для людей, желающих похудеть.

8.3. Сравнивалось действие двух экстрактов вируса табачной мозаики. Для этого каждая из половин листа натиралась соответствующим препаратом. Число пораженных мест приводится ниже:

Экстракт А	20	39	43	13	28	26	17	49	36
Экстракт В	31	22	45	6	21	13	17	46	31

Можно ли считать, что действие этих экстрактов различно? Принять $\alpha = 0{,}1$.

8.4. Предложены два метода (А и В) увеличения выхода продукции. При уровне значимости 0,05 проверить нулевую гипотезу об их одинаковой эффективности по двум выборкам объемов $n_1 = 6$, $n_2 = 9$ (в первой строке приведены проценты прироста продукции в каждом опыте по методу А; во второй строке – по методу В)

x_i	0,2	0,3	0,5	0,8	1,0	1,3			
y_i	0,1	0,4	0,6	0,7	0,9	1,4	1,7	1,8	1,9

Принять в качестве конкурирующей гипотезу: эффективность методов А и В различна.

8.5. Производительность труда двух смен завода характеризуется выборками объемов $n_1 = 9$, $n_2 = 10$:

Первая смена	28	33	39	40	41	42	45	46	47	
Вторая смена	34	40	41	42	43	44	46	48	49	52

Используя критерий Вилкоксона, при уровне значимости 0,1 проверить нулевую гипотезу об одинаковой производительности обеих смен, приняв в качестве конкурирующей гипотезу: производительность труда смен различна.

8.6. Эффективность каждого из двух рационов (А и В) откорма скота характеризуется выборками объемов $n_1 = 10$, $n_2 = 12$ (в первой строке приведен вес (в кг) животных, которых откармливали по рациону А, во второй строке – по рациону В):

x_i	24	26	27	27	30	32	33	34	35	36		
y_i	21	21	22	23	25	25	25	25	27	27	29	31

Используя критерий Вилкоксона, при уровне значимости 0,05 проверить нулевую гипотезу об одинаковой эффективности рационов А и В, приняв в качестве конкурирующей гипотезу: рацион А эффективнее рациона В ($H_1 : F_1(x) < F_2(x)$, т.е. X>Y).

8.7. При уровне значимости 0,05 проверить нулевую гипотезу об однородности двух выборок объемов $n_1 = 25$, $n_2 = 30$ при конкурирующей гипотезе $H_1 : F_1(x) \neq F_2(x)$:

Варианты 1–ой выборки	12	14	15	18	21	25	26	27	30	31	32	35	38	
	41	43	46	48	52	56	57	60	65	68	73	75		
Варианты 2–ой выборки	11	13	16	17	19	20	22	23	24	26	28	29	33	
	34	36	37	39	40	42	44	45	47	49	51	53	55	
	58	61	63	66										

8.8. У полевых транзисторов из двух партий, изготовленных с применением различных технологий, измерялось дифференциальное сопротивление канала R_i. Результаты измерений (в микроомах) следующие:

Технология А	0,01	0,02	0,12	0,30	0,29	0,15	0,21	
Технология В	0,15	0,07	0,25	0,15	0,22	0,18	0,18	0,27

Влияет ли технология изготовления на величину дифференциального сопротивления канала R_i? Принять $\alpha = 0{,}05$.

8.9. В биохимическом исследовании, проведенном методом меченых атомов, по результатам 8 препаратов контрольной серии получены следующий показания счетчика импульсов (в импульсах в минуту):

Опыт	340	343	322	349	332	320	313	304
Контроль	318	321	318	301	312			

Можно ли считать, что полученные значения опытной и контрольной серий различны? Принять $\alpha = 0{,}10$.

8.10. Длина тела личинок щелкуна, обитающих в посевах озимой ржи и проса (выраженная в мм), варьируется следующим образом:

В посевах ржи	7	10	14	15	12	16	12
В посевах проса	11	12	16	13	18	15	

На основании этих проб создается впечатление о более крупных размерах личинок щелкунов, обитающих на просе. Проверить это предположение, используя критерий Манна-Уитни. Принять $\alpha = 0{,}10$.

9. ОТВЕТЫ

ОТВЕТЫ К ГЛАВЕ 1

1.1

а)

$$F^*(x) = \begin{cases} 0 \text{ при } x \le 2 \\ 0{,}1 \text{ при } 2 < \text{x} \le 5 \\ 0{,}4 \text{ при } 5 < \text{x} \le 7 \\ 0{,}6 \text{ при } 7 < \text{x} \le 8 \\ 1 \text{ при x} > 8 \end{cases}$$

x_i	2	5	7	8
n_i	0,1	0,3	0,2	0,4

б)

$$F^*(x) = \begin{cases} 0 \text{ при} x \le 4 \\ 0{,}5 \text{ при } 4 < \text{x} \le 7 \\ 0{,}7 \text{ при } 7 < \text{x} \le 8 \\ 1 \text{ при x} > 8 \end{cases}$$

x_i	4	7	8
n_i	0,5	0,2	0,3

1.2 а) 0,3,5,6,8,9,11,11,11,12,12,13,14,15,16,16,19 –вариационный ряд,

x_i	0	3	5	6	8	9	11	12	13	14	15	16	19
n_i	1	1	1	1	1	1	3	2	1	1	1	2	1

б) а) 15,16,16,16,16,17,17,17,17,18,18,18,18,18,19,19 –вариационный ряд,

x_i	15	16	17	18	19
n_i	1	4	4	5	2

1.3. График эмпирической функции распределения изображен на рис.4, гистограмма частот – на рис.5, полигон частот – на рис.6.

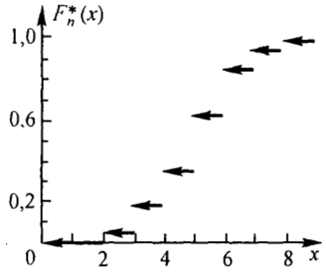


Рис.4

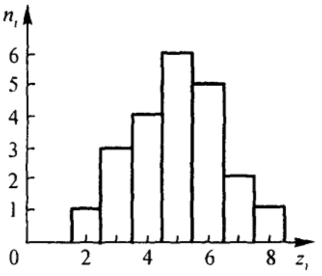


Рис.5

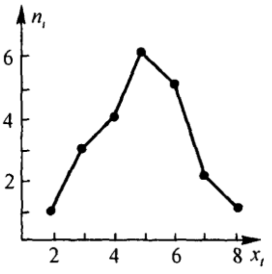


Рис.6

1.4. График эмпирической функции распределения изображен на рис.7, гистограмма частот – на рис.8, полигон частот – на рис.9.

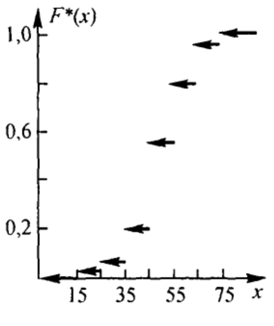


Рис.7

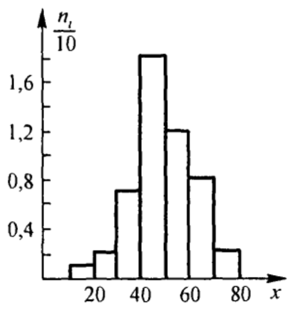


Рис.8

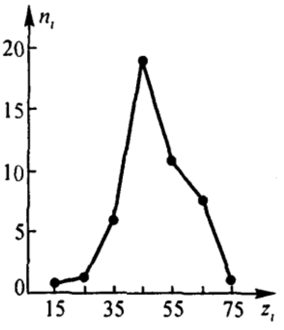


Рис.9

1.5. График эмпирической функции распределения изображен на рис.10., гистограмма частот – на рис.11., полигон частот – на рис.12.

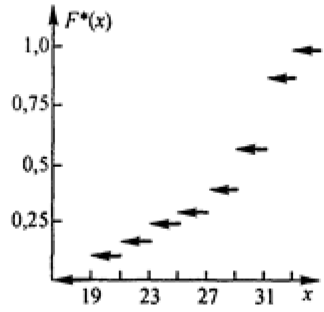


Рис.10.

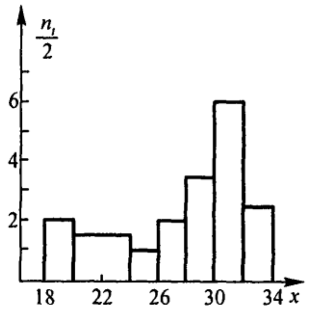


Рис.11.

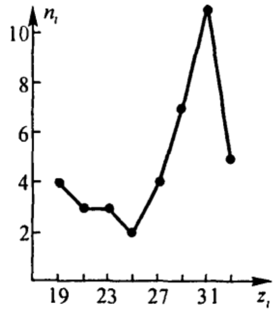


Рис.12.

1.6. а) Гистограмма изображена на рис.13, полигон накопленных частот – на рис.14.

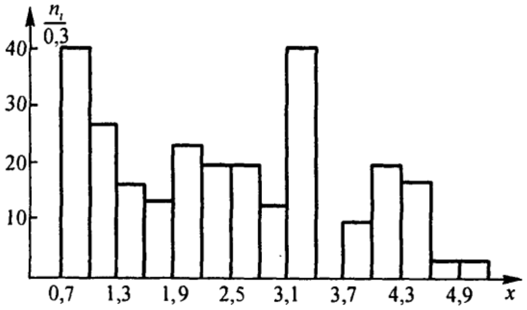


Рис.13.

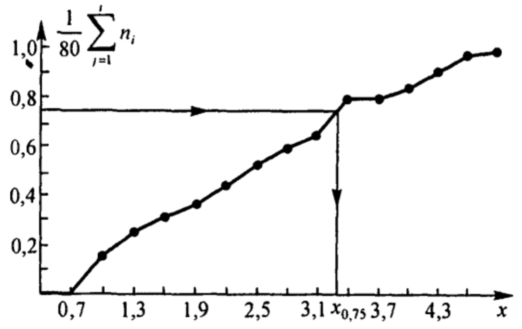


Рис.14.

б) Гистограмма изображена на рис.15, полигон накопленных частот – на рис.16.

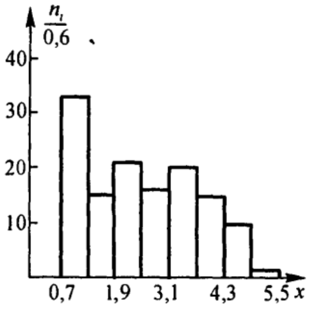


Рис.15.

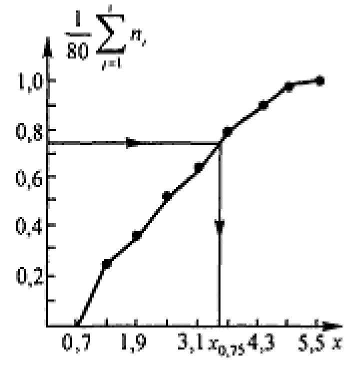


Рис.16.

ОТВЕТЫ К ГЛАВЕ 2

2.1. $d_x^* = h_x^* = 3, \bar{x} = 3{,}5$, $D_x^* = 3{,}65$. **2.2.** $d_x^* = 3{,}1$, $h_x^* = 2{,}5$ $\bar{x} = 2{,}39$, $D_x^* = 0{,}43$. **2.3.** а) $d_x^* = 5$, $h_x^* = 4$ $\bar{x} = 4{,}14$, $D_x^* = 5{,}84$; б) $d_x^* = 5$, $h_x^* = 4$ $\bar{x} = 4{,}57$, $D_x^* = 11{,}10$. Для выборки б) среднее и дисперсия увеличились, а мода и медиана не изменились.

2.4. $\bar{x} \approx 6{,}54$, $d_x^* = 6$, $h_x^* = 6$, $D_x^* \approx 7{,}34$. **2.5**. $\bar{x} \approx 11{,}78$, $d_x^* = 10$, $h_x^* = 10$, $D_x^* \approx 32{,}39$ **2.6.** 1) $\bar{x} \approx 70{,}4$ $D_x^* \approx 9{,}6$; 2) $\bar{x} \approx 69{,}8$ $D_x^* \approx 9{,}2$. **2.7**. 1) $\bar{x} \approx 5{,}96$ $D_x^* \approx 1{,}35$; 2) $\bar{x} \approx 5{,}94$ $D_x^* \approx 1{,}31$.

ОТВЕТЫ К ГЛАВЕ 3

3.1 $\tilde{\sigma}^2 = D_x^* = \frac{1}{n}\sum (x_i - \bar{x})^2$, $\tilde{h}_x = h_x^*$, $\tilde{\alpha}_x = \alpha_x^*$, $\tilde{e}_x = e_x^*$. **3.2.** $\tilde{\alpha}_l = \alpha_l^* = \frac{1}{n}\sum x_i^l$. 3**.4.** Смещение $\tilde{\theta}^2$ равно $D[\tilde{\theta}]$. Указание. Рассмотреть тождество $D[\tilde{\theta}] = M[(\tilde{\theta} - \theta)^2]$. **3.5**. Указание. Использовать указание к задаче 3.4. **3.10.** $\bar{x} = 5{,}76$. 2.11. $\bar{x} = 4$. **3.12.** $S^2 = 3{,}075$. **3.13.** $S^2 = 5{,}1$. **3.14.** а) $\bar{x} = 100$; б) $D_x^* = 34$; $S^2 = 42{,}5$. **3.15.** а) $\bar{x} = 10$; б) $D_x^* = 2{,}5$; $S^2 = 10/3$. **3.16**. 814,87 $м^2$.

ОТВЕТЫ К ГЛАВЕ 4

4.2. $p^* = \frac{\bar{x}}{m}$. **4.3.** $\lambda^* = \frac{1}{\overline{x}}$. **4.4.** $\lambda^* = \bar{x}$ **.4.5**. $a^* = \bar{x}$, $\sigma^* = \sqrt{D_x^*}$. **4.6**. $p^* = \frac{1}{\overline{x}}$. 4**.7.** $\beta^* = \frac{\bar{x}}{\overline{\alpha + 1}}$. 4**.8.** $a^* = \bar{x}_в - \sqrt{3D_в}$, $b^* = \bar{x}_в + \sqrt{3D_в}$. **4.9**. $a^* = \frac{\bar{x}^2}{D_x^*}$, $b^* = \frac{\bar{x}}{D_x^*}$. **4.10.** $\lambda^* = \bar{x}$. **4.11.** $a^* = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} x_i^2}$. **4.12.** $\lambda^* = \bar{x} = 0{,}9$. **4.13.** $\lambda^* = 1/\bar{x} = 0{,}2$.

ОТВЕТЫ К ГЛАВЕ 5

5.1. (498,35;501,64), (497,42;502,58). **5.2.** (16,63;19,37), (15,76;20,24). **5.3.** (2,70;9,3), (2,45;10,78). **5.4.** (7,27;119,40), (10,32;58,61).**5.5.** (96,81;491,34), (9,84;22,17). **5.6.** (11,5;63,85). **5.7.** (-16,7;10,81). **5.8.** (11,5;63,85). **5.9. а)** (7,63;12,776); б) (14,23;19,37). **5.10.** (-0,04;0,88). **5.11.** (3,15;11,63); (8,65;49,16). **5.12.** (34,66;50,94). **5.13.** (0,3; 2,13). **5.14.** (19,74; 34,61)

ОТВЕТЫ К ГЛАВЕ 6

6.1. $F_{набл.} = 2$; $F_{кр}(0{,}05;10;13) = 2{,}67$. Нет оснований отвергнуть нулевую гипотезу. **6.2**. $F_{набл.} = 3$; $F_{кр}(0{,}05;9;13) = 2{,}72$. Нулевая гипотеза отвергается. 6**.3.** $F_{набл.} = 3{,}93$; $F_{кр} = 3{,}34$. Нулевая гипотеза отвергается. **6.4.** $F_{набл.} = 3{,}93$; $F_{кр} = 5{,}82$. Нет оснований отвергнуть нулевую гипотезу. **6.5.** $S_x^2 = 188{,}67$, $S_y^2 = 124{,}84$,

$F_{набл.} = 1{,}51$, $F_{кр} = 3{,}68$. Гипотеза принимается. **6.6.** $F_{набл.} = 1{,}48$, $F_{кр} = 9{,}12$. Нет оснований отвергнуть нулевую гипотезу. **6.7.** $z_{набл.} = -5$, $z_{кр} = 2{,}58$. Нулевая гипотеза отвергается. **6.8.** $z_{набл.} = 2{,}5$, $z_{кр} = 1{,}96$. Нулевая гипотеза отвергается. Средний вес изделий различается значимо. **6.9.** $F_{набл.} = 1{,}19$; $F_{кр} = 5{,}62$. Нет основний отвергнуть гипотезу о равенстве дисперсий. Имеем $|T_{набл.}| = 3{,}7$, $t_{двуст.кр} = 2{,}92$. Нулевая гипотеза равенстве средних отвергается. **6.10.** $S_x^2 = 0{,}0267$, $S_y^2 = 0{,}0255$. $F_{набл.} = 1{,}05$, $F_{кр} = 4{,}63$. Дисперсии равны. , $T_{набл.} = 1{,}45$ $t_{кр} = 2{,}53$. Гипотеза принимается. **6.11.** $\bar{x} = 12{,}8$, $\bar{y} = 12{,}35$, $S_x^2 = 0{,}11$, $S_y^2 = 0{,}07$. $F_{набл.} = 1{,}57$, $F_{кр} = 2{,}59$. Дисперсии равны. , $T_{набл.} = 3{,}83$ $t_{правост.кр} = 1{,}71$. Нулевая гипотеза отвергается. **6.12.** а) Гипотеза принимается; б) Гипотеза отклоняется. **6.13.** а) Гипотеза принимается; б) Гипотеза отклоняется.

ОТВЕТЫ К ГЛАВЕ 7

7.1. Да. **7.2.** Нет. **7.3**. Да. **7.4**. Да. **7.5.** Гипотеза отклоняется. **7.6.** Основная гипотеза отвергается. **7.7.** Гипотеза принимается. **7.7.** Гипотеза принимается. **7.9.** Гипотеза принимается. **7.10.** а) Случайно; б) Значимо. **7.11.** Гипотеза принимается. **7.12.** Гипотеза отклоняется. **7.13.** Гипотеза принимается. **7.14.** Гипотеза принимается. **7.15.** Гипотеза принимается. **7.16.** Гипотеза отклоняется. **7.17.** Гипотеза отклоняется. **7.18.** Различия являются значимыми.

ОТВЕТЫ К ГЛАВЕ 8

8.1. Да. **8.2.** Нет. **8.3.** Нет. **8.4.** Основная гипотеза принимается. **8.5.** Нулевая гипотеза принимается. **8.6.** Нулевая гипотеза отвергается. Рацион А эффективнее рациона В. **8.7.** Нулевая гипотеза принимается. **8.8.** Нет. **8.9.** Да. **8.10.** Предположение отклоняется.

ПРИЛОЖЕНИЯ

Приложение 1

Таблица значений функции плотности нормального распределения N(0,1)

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

x	0	1	2	3	4	5	6	7	8	9
0	0,3989	3989	3989	3988	39986	3984	3982	3980	3977	3973
0,1	0,397	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	0,391	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	0,3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	0,3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	0,3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	0,3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	0,3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	0,2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	0,2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1	0,242	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	0,2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	0,1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	0,1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	0,1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	0,1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	0,1109	1092	1074	1057	1040	1023	1006	989	973	957
1,7	0,094	925	909	893	878	863	848	833	818	804
1,8	0,079	775	761	748	734	721	707	694	681	669
1,9	0,0656	644	632	620	608	596	584	573	562	551
2	0,054	529	519	508	498	488	478	468	459	449
2,1	0,044	431	422	413	404	396	387	379	371	363
2,2	0,0355	347	339	332	325	317	310	303	297	290
2,3	0,0283	277	270	264	258	252	246	241	235	229
2,4	0,0224	219	213	208	203	198	194	189	184	180
2,5	0,0175	171	167	163	158	154	151	147	143	139
2,6	0,0136	132	129	126	122	119	116	113	110	107
2,7	0,0104	101	99	96	93	91	88	86	84	81
2,8	0,0079	77	75	73	71	69	67	65	63	61
2,9	0,006	58	56	55	53	51	50	48	47	46

3	0,0044	43	42	40	39	38	37	36	35	34
3,1	0,0033	32	31	30	29	28	27	26	25	25
3,2	0,0024	23	22	22	21	20	20	19	18	18
3,3	0,0017	17	16	16	15	15	14	14	13	13
3,4	0,0012	12	12	11	11	10	10	10	9	9
3,5	0,0009	8	8	8	8	7	7	7	7	6
3,6	0,0006	6	6	5	5	5	5	5	5	4
3,7	0,0004	4	4	4	4	4	3	3	3	3
3,8	0,0003	3	3	3	3	2	2	2	2	2
3,9	0,0002	2	2	2	2	2	2	2	1	1

Квантили u_p нормального распределения *N(0,1)*.

p	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
u_p	1,282	1,645	1,96	2,326	2,576	3,090	3,291

Приложение 2

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

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0	0	0,32	0,1255	0,64	0,2389	0,96	0,3315
0,01	0,004	0,33	0,1293	0,65	0,2422	0,97	0,334
0,02	0,008	0,34	0,1331	0,66	0,2454	0,98	0,3365
0,03	0,012	0,35	0,1368	0,67	0,2486	0,99	0,3389
0,04	0,016	0,36	0,1406	0,68	0,2517	1	0,3413
0,05	0,0199	0,37	0,1443	0,69	0,2549	1,01	0,3438
0,06	0,0239	0,38	0,148	0,7	0,258	1,02	0,3461
0,07	0,0279	0,39	0,1517	0,71	0,2611	1,03	0,3485
0,08	0,0319	0,4	0,1554	0,72	0,2642	1,04	0,3508
0,09	0,0359	0,41	0,1591	0,73	0,2673	1,05	0,3531
0,1	0,0398	0,42	0,1628	0,74	0,2703	1,06	0,3554
0,11	0,0438	0,43	0,1664	0,75	0,2734	1,07	0,3577
0,12	0,0478	0,44	0,17	0,76	0,2764	1,08	0,3599
0,13	0,0517	0,45	0,1736	0,77	0,2794	1,09	0,3621
0,14	0,0557	0,46	0,1772	0,78	0,2823	1,1	0,3643
0,15	0,0596	0,47	0,1808	0,79	0,2852	1,11	0,3665
0,16	0,0636	0,48	0,1844	0,8	0,2881	1,12	0,3686
0,17	0,0675	0,49	0,1879	0,81	0,291	1,13	0,3708
0,18	0,0714	0,5	0,1915	0,82	0,2939	1,14	0,3729
0,19	0,0753	0,51	0,195	0,83	0,2967	1,15	0,3749
0,2	0,0793	0,52	0,1985	0,84	0,2995	1,16	0,377
0,21	0,0832	0,53	0,2019	0,85	0,3023	1,17	0,379
0,22	0,0871	0,54	0,2054	0,86	0,3051	1,18	0,381
0,23	0,091	0,55	0,2088	0,87	0,3078	1,19	0,383
0,24	0,0948	0,56	0,2123	0,88	0,3106	1,2	0,3849
0,25	0,0987	0,57	0,2157	0,89	0,3133	1,21	0,3869
0,26	0,1026	0,58	0,219	0,9	0,3159	1,22	0,3883
0,27	0,1064	0,59	0,2224	0,91	0,3186	1,23	0,3907
0,28	0,1103	0,6	0,2257	0,92	0,3212	1,24	0,3925
0,29	0,1141	0,61	0,2291	0,93	0,3238	1,25	0,3944
0,3	0,1179	0,62	0,2324	0,94	0,3264		
0,31	0,1217	0,63	0,2357	0,95	0,3289		

Продолжение приложения 2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,26	0,3962	1,59	0,4441	1,92	0,4726	2,5	0,4938
1,27	0,398	1,6	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
1,3	0,4032	1,63	0,4484	1,96	0,475	2,58	0,4951
1,31	0,4049	1,64	0,4495	1,97	0,4756	2,6	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	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	2,7	0,4965
1,37	0,4147	1,7	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	2,1	0,4821	2,76	0,4971
1,4	0,4192	1,73	0,4582	2,12	0,483	2,78	0,4973
1,41	0,4207	1,74	0,4591	2,14	0,4838	2,8	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	2,2	0,4861	2,86	0,4979
1,45	0,4265	1,78	0,4625	2,22	0,4868	2,88	0,498
1,46	0,4279	1,79	0,4633	2,24	0,4875	2,9	0,4981
1,47	0,4292	1,8	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	2,3	0,4893	2,96	0,4985
1,5	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	0,4987
1,52	0,4357	1,85	0,4678	2,36	0,4909	3,2	0,4993
1,53	0,437	1,86	0,4686	2,38	0,4913	3,4	0,4997
1,54	0,4382	1,87	0,4693	2,4	0,4918	3,6	0,4998
1,55	0,4394	1,88	0,4699	2,42	0,4922	3,8	0,4999
1,56	0,4406	1,89	0,4706	2,44	0,4927	4	0,5
1,57	0,4418	1,9	0,4713	2,46	0,4931	4,5	0,5
1,58	0,4429	1,91	0,4719	2,48	0,4934	5	0,5

Приложение 3

Критические точки распределения Стьюдента.

Число степеней свободы *k*	Уровень значимости α (двусторонняя критическая область)							
	0.20	**0.10**	**0.05**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	3.077	6.313	12.706	31.820	63.656	127.656	318.306	636.619
2	1.885	2.920	4.302	6.964	9.924	14.089	22.327	31.599
3	1.638	2.353	3.182	4.540	5.840	7.458	10.214	12.924
4	1.533	2.132	2.776	3.746	4.604	5.597	7.173	8.610
5	1.476	2.015	2.570	3.649	4.032	4.773	5.893	6.863
6	1.439	1.943	2.446	3.142	3.707	4.316	5.207	5.958
7	1.415	1.895	2.365	2.998	3.500	4.229	4.785	5.408
8	1.397	1.860	2.306	2.897	3.355	3.832	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.780
10	1.372	1.813	2.228	2.764	3.169	3.581	4.144	4.587
11	1.363	1.795	2.201	2.718	3.105	3.496	4.024	4.437
12	1.356	1.782	2.179	2.681	3.085	3.428	3.929	4.178
13	1.350	1.771	2.160	2.650	3.112	3.373	3.852	4.220
14	1.345	1.761	2.145	2.625	2.976	3.326	3.787	4.140
15	1.341	1.753	2.131	2.603	2.947	3.286	3.732	4.072
16	1.336	1.745	2.119	2.583	2.920	3.252	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	1.330	1.734	2.101	2.551	2.878	3.197	3.611	3.922
19	1.328	1.729	2.093	2.540	2.861	3.174	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	1.323	1.720	2.079	2.517	2.831	3.135	3.527	3.819
22	1.321	1.712	2.074	2.508	2.819	3.119	3.505	3.792
23	1.320	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	1.315	1.705	2.059	2.478	2.778	3.066	3.436	3.706
27	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.036	3.396	3.849
30	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	1.303	1.684	2.021	2.423	2.705	3.971	3.307	3.551
50	1.298	1.676	2.009	2.403	2.678	3.937	3.261	3.406
60	1.296	1.671	2.000	2.390	2.660	3.915	3.232	3.460
120	1.289	1.658	1.972	2.358	2.617	2.860	3.160	3.374
	0.10	**0.05**	**0.025**	**0.01**	**0.005**	**0.0025**	**0.001**	**0.0005**
	Уровень значимости α (односторонняя критическая область)							

Приложение 4

Критические точки распределения F Фишера-Снедекора

$$\left(\begin{array}{l} k_1 - \text{число степеней свободы большей выборочной дисперсии,} \\ k_2 - \text{число степеней свободы меньшей выборочной дисперсии} \end{array}\right)$$

Уровень значимости $\alpha = 0{,}01$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,3	99,34	99,36	99,36	99,4	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,2	18	16,69	15,98	15,52	15,21	14,98	14,8	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,1	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,8	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,2	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,4
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,5	4,39	4,3	4,22	4,16
13	9,07	6,7	5,74	5,2	4,86	4,62	4,44	4,3	4,19	4,1	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,8
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4	3,89	3,8	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,2	4,03	3,89	3,78	3,69	3,61	3,55
17	8,4	6,11	5,18	4,67	4,34	4,1	3,93	3,79	3,68	3,59	3,52	3,45
Уровень значимости $\alpha = 0{,}05$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19	19,16	19,25	19,3	19,33	19,36	19,37	19,38	19,39	19,4	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,7	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,1	4,06	4,03	4
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,6	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,5	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,1	3,07
10	4,96	4,1	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,2	3,09	3,01	2,95	2,9	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3	2,92	2,85	2,8	2,76	2,72	2,69
13	4,67	3,8	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,6
14	4,6	3,74	3,34	3,11	2,96	2,85	2,77	2,7	2,65	2,6	2,56	2,53
15	4,54	3,68	3,29	3,06	2,9	2,79	2,7	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,2	2,96	2,81	2,7	2,62	2,55	2,5	2,45	2,41	2,38

Приложение 5

Квантили χ – квадрат распределения $\chi_p^2(k)$

к	p											
	0,005	0,01	0,025	0,05	0,1	0,2	0,8	0,9	0,95	0,975	0,99	0,995
1			0,001	0,004	0,016	0,064	1,642	2,706	3,841	5,024	6,635	7,879
2	0,01	0,02	0,051	0,103	0,211	0,446	3,219	4,605	5,991	7,378	9,21	10,6
3	0,072	0,115	0,216	0,352	0,584	1,005	4,642	6,251	7,815	9,348	11,35	12,84
4	0,207	0,297	0,484	0,711	1,064	1,649	5,989	7,779	9,488	11,14	13,28	14,86
5	0,412	0,554	0,831	1,145	1,61	2,343	7,289	9,236	11,07	12,83	15,09	16,75
6	0,676	0,872	1,237	1,635	2,204	3,07	8,558	10,65	12,59	14,45	16,81	18,55
7	0,989	1,239	1,69	2,167	2,833	3,822	9,803	12,02	14,07	16,01	18,48	20,28
8	1,344	1,646	2,18	2,733	3,49	4,594	11,03	13,36	15,5	17,54	20,09	21,96
9	1,735	2,088	2,7	3,325	4,168	5,38	12,24	14,68	16,92	19,02	21,67	23,59
10	2,156	2,558	3,247	3,94	4,865	6,179	13,44	15,99	18,31	20,48	23,21	25,19
11	2,603	3,053	3,816	4,575	5,578	6,989	14,63	17,28	19,68	21,92	24,73	26,76
12	3,074	3,571	4,404	5,226	6,304	7,807	15,81	18,55	21,03	23,34	26,22	28,3
13	3,565	4,107	5,009	5,892	7,042	8,634	16,99	19,81	22,36	24,74	27,69	29,82
14	4,075	4,66	5,629	6,571	7,79	9,467	18,15	21,06	23,69	26,12	29,14	31,32
15	4,601	5,229	6,262	7,261	8,547	10,31	19,31	22,31	25	27,49	30,58	32,8
16	5,142	5,812	6,908	7,962	9,312	11,15	20,47	23,54	26,3	28,85	32	34,27
17	5,697	6,408	7,564	8,672	10,09	12	21,62	24,77	27,59	30,19	33,41	35,72
18	6,265	7,015	8,231	9,39	10,87	12,86	22,76	25,99	28,87	31,53	34,81	37,16
19	6,844	7,633	8,907	10,12	11,65	13,72	23,9	27,2	30,14	32,85	36,19	38,58
20	7,434	8,26	9,591	10,85	12,44	14,58	25,04	28,41	31,41	34,17	37,57	40
21	8,034	8,897	10,28	11,59	13,24	15,45	26,17	29,62	32,67	35,48	38,93	41,4
22	8,643	9,542	10,98	12,34	14,04	16,31	27,3	30,81	33,92	36,78	40,29	42,8
23	9,26	10,2	11,69	13,09	14,85	17,19	28,43	32	35,17	38,08	41,64	44,18
24	9,886	10,86	12,4	13,85	15,66	18,06	29,55	33,2	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	18,94	30,68	34,38	37,65	40,65	44,31	46,93
26	11,16	12,2	13,84	15,38	17,29	19,82	31,79	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	20,7	32,91	36,74	40,11	43,19	46,96	49,64
28	12,46	13,56	15,31	16,93	18,94	21,59	34,03	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	22,48	35,14	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,6	23,36	36,25	40,26	43,77	46,98	50,89	53,67
35	17,19	18,51	20,57	22,47	24,8	27,84	41,78	46,06	49,8	53,2	57,34	60,27
40	20,71	22,16	24,43	26,51	29,05	32,34	47,27	51,81	55,76	59,34	63,69	66,77
45	24,31	25,9	28,37	30,61	33,35	36,88	52,73	57,51	61,66	65,41	69,96	73,17
50	27,99	29,71	32,36	34,76	37,69	41,45	58,16	63,17	67,5	71,42	76,15	79,49
55	31,73	33,57	36,4	38,96	42,06	46,04	63,58	68,8	73,31	77,38	82,29	85,75
60	35,5	37,48	40,48	43,19	46,46	50,64	68,97	74,4	79,08	83,3	88,38	91,95
70	43,28	45,44	48,76	51,74	55,33	59,9	79,71	85,53	90,53	95,02	100,4	104,2
80	51,17	53,54	57,15	60,39	64,28	69,21	90,41	96,58	101,9	106,6	112,3	116,3
90	59,2	61,75	65,65	69,13	73,29	78,56	101	107,6	113,1	118,1	124,1	128,3
100	67,33	70,06	74,22	77,93	82,36	87,95	111,7	118,5	124,3	129,6	135,8	140,2

Приложение 6

Критические точки для статистики Колмогорова D_n

Объём выборки n	Уровень значимости α			
	0,1	0,05	0,02	0,01
1	0,95	0,98	0,99	0,995
2	0,78	0,84	0,9	0,93
3	0,64	0,71	0,78	0,83
4	0,57	0,62	0,69	0,73
5	0,51	0,56	0,62	0,67
6	0,47	0,52	0,58	0,62
7	0,44	0,48	0,54	0,58
8	0,41	0,45	0,51	0,54
9	0,39	0,43	0,48	0,51
10	0,07	0,41	0,46	0,49
11	0,35	0,39	0,44	0,47
12	0,34	0,38	0,42	0,45
13	0,33	0,36	0,4	0,43
14	0,31	0,35	0,39	0,42
15	0,3	0,34	0,38	0,4
16	0,29	0,33	0,37	0,39
17	0,29	0,32	0,36	0,38
18	0,28	0,31	0,34	0,37
19	0,27	0,3	0,34	0,36
20	0,26	0,29	0,33	0,35

Критические точки распределения Колмогорова

$$Q(\lambda) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2\lambda^2}$$

α	0,10	0,05	0,02	0,01
$\lambda_{\text{кр}}$	1,23	1,36	1,52	1,63

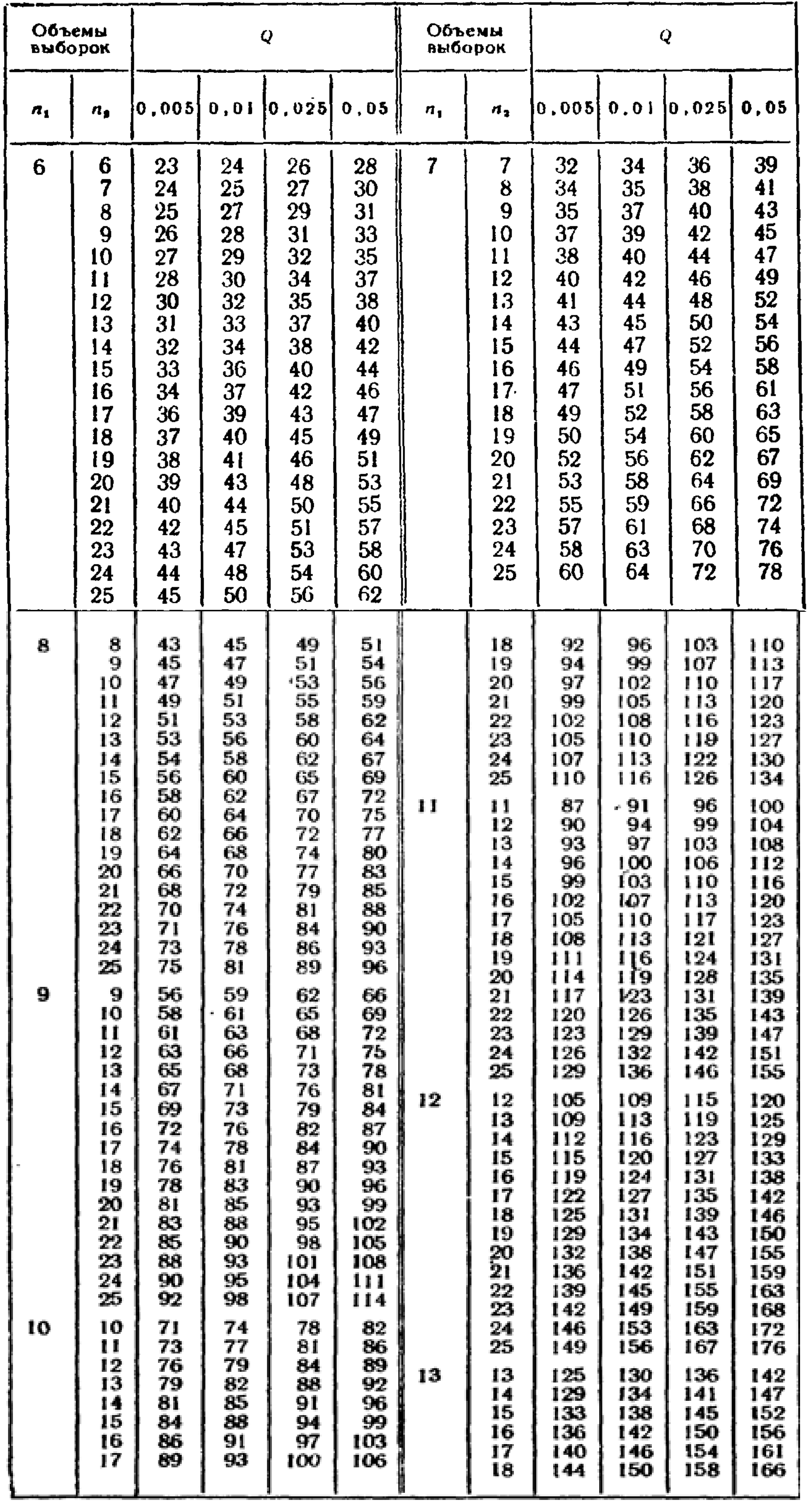
Приложение 7

Квантили распределения Фишера $F_p(k_1, k_2)$, где $p = 1 - \alpha$

k_2	k_1																		
	P=0,9																		
	1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120
1	39,8	49,	53,5	55,8	57,2	58,	58,9	59,4	59,8	60,1	60,	60,7	61,2	61,7	62	62,2	62,5	62,7	63,0
2	8,53	9	9,16	9,24	9,29	9,3	9,35	9,37	9,38	9,39	9,4	9,41	9,42	9,44	9,4	9,46	9,47	9,47	9,48
3	5,54	5,4	5,39	5,34	5,31	5,2	5,27	5,25	5,24	5,23	5,2	5,22	5,2	5,18	5,1	5,17	5,16	5,15	5,14
4	4,54	4,3	4,19	4,11	4,05	4,0	3,98	3,95	3,94	3,92	3,9	3,9	3,87	3,84	3,8	3,82	3,8	3,79	3,78
5	4,06	3,7	3,62	3,52	3,45	3,4	3,37	3,34	3,32	3,3	3,2	3,27	3,24	3,21	3,1	3,17	3,16	3,14	3,12
6	3,78	3,4	3,29	3,18	3,11	3,0	3,01	2,98	2,96	2,94	2,9	2,9	2,87	2,84	2,8	2,8	2,78	2,76	2,74
7	3,59	3,2	3,07	2,96	2,88	2,8	2,78	2,75	2,72	2,7	2,6	2,67	2,63	2,59	2,5	2,56	2,54	2,51	2,49
8	3,46	3,1	2,92	2,81	2,73	2,6	2,62	2,59	2,56	2,54	2,5	2,5	2,46	2,42	2,4	2,38	2,36	2,34	2,32
9	3,36	3,0	2,81	2,69	2,61	2,5	2,51	2,47	2,44	2,42	2,4	2,38	2,34	2,3	2,2	2,25	2,23	2,21	2,18
10	3,29	2,9	2,73	2,61	2,52	2,4	2,41	2,38	2,35	2,32	2,3	2,28	2,24	2,2	2,1	2,16	2,13	2,11	2,08
11	3,23	2,8	2,66	2,54	2,45	2,3	2,34	2,3	2,27	2,25	2,2	2,21	2,17	2,12	2,1	2,08	2,05	2,03	2
12	3,18	2,8	2,61	2,48	2,39	2,3	2,28	2,24	2,21	2,19	2,1	2,15	2,1	2,06	2,0	2,01	1,99	1,96	1,93
13	3,14	2,7	2,56	2,43	2,35	2,2	2,23	2,2	2,16	2,14	2,1	2,1	2,05	2,01	1,9	1,96	1,93	1,9	1,88
14	3,1	2,7	2,52	2,39	2,31	2,2	2,19	2,15	2,12	2,1	2,0	2,05	2,01	1,96	1,9	1,91	1,89	1,86	1,83
15	3,07	2,7	2,49	2,36	2,27	2,2	2,16	2,12	2,09	2,06	2,0	2,02	1,97	1,92	1,9	1,87	1,85	1,82	1,79
16	3,05	2,6	2,46	2,33	2,24	2,1	2,13	2,09	2,06	2,03	2,0	1,99	1,94	1,89	1,8	1,84	1,81	1,78	1,75
17	3,03	2,6	2,44	2,31	2,22	2,1	2,1	2,06	2,03	2	1,9	1,96	1,91	1,86	1,8	1,81	1,78	1,75	1,72
18	3,01	2,6	2,42	2,29	2,2	2,1	2,08	2,04	2	1,98	1,9	1,93	1,89	1,84	1,8	1,78	1,75	1,72	1,69
19	2,99	2,6	2,4	2,27	2,18	2,1	2,06	2,02	1,98	1,96	1,9	1,91	1,86	1,81	1,7	1,76	1,73	1,7	1,67
20	2,97	2,5	2,38	2,25	2,16	2,0	2,04	2	1,96	1,94	1,9	1,89	1,84	1,79	1,7	1,74	1,71	1,68	1,64
22	2,95	2,5	2,35	2,22	2,13	2,0	2,01	1,97	1,93	1,9	1,8	1,86	1,81	1,76	1,7	1,7	1,67	1,64	1,6
24	2,93	2,5	2,33	2,19	2,1	2,0	1,98	1,94	1,91	1,88	1,8	1,83	1,78	1,73	1,7	1,67	1,64	1,61	1,57
26	2,91	2,5	2,31	2,17	2,08	2,0	1,96	1,92	1,88	1,86	1,8	1,81	1,76	1,71	1,6	1,65	1,61	1,58	1,54
28	2,89	2,5	2,29	2,16	2,06	2	1,94	1,9	1,87	1,84	1,8	1,79	1,74	1,69	1,6	1,63	1,59	1,56	1,52
30	2,88	2,4	2,28	2,14	2,05	1,9	1,93	1,88	1,85	1,82	1,7	1,77	1,72	1,67	1,6	1,61	1,57	1,54	1,5
40	2,84	2,4	2,23	2,09	2	1,9	1,87	1,83	1,79	1,76	1,7	1,71	1,66	1,61	1,5	1,54	1,51	1,47	1,42
60	2,79	2,3	2,18	2,04	1,95	1,8	1,82	1,77	1,74	1,71	1,6	1,66	1,6	1,54	1,5	1,48	1,44	1,4	1,35
120	2,75	2,3	2,13	1,99	1,9	1,8	1,77	1,72	1,68	1,65	1,6	1,6	1,55	1,48	1,4	1,41	1,37	1,32	1,26

Приложение 8

Критические точки критерия Вилкоксона



Объемы выборок		Q				Объемы выборок		Q			
n_1	n_2	0,005	0,01	0,025	0,05	n_1	n_2	0,005	0,01	0,025	0,05
6	6	23	24	26	28	7	7	32	34	36	39
	7	24	25	27	30		8	34	35	38	41
	8	25	27	29	31		9	35	37	40	43
	9	26	28	31	33		10	37	39	42	45
	10	27	29	32	35		11	38	40	44	47
	11	28	30	34	37		12	40	42	46	49
	12	30	32	35	38		13	41	44	48	52
	13	31	33	37	40		14	43	45	50	54
	14	32	34	38	42		15	44	47	52	56
	15	33	36	40	44		16	46	49	54	58
	16	34	37	42	46		17	47	51	56	61
	17	36	39	43	47		18	49	52	58	63
	18	37	40	45	49		19	50	54	60	65
	19	38	41	46	51		20	52	56	62	67
	20	39	43	48	53		21	53	58	64	69
	21	40	44	50	55		22	55	59	66	72
	22	42	45	51	57		23	57	61	68	74
	23	43	47	53	58		24	58	63	70	76
	24	44	48	54	60		25	60	64	72	78
	25	45	50	56	62						
8	8	43	45	49	51		18	92	96	103	110
	9	45	47	51	54		19	94	99	107	113
	10	47	49	53	56		20	97	102	110	117
	11	49	51	55	59		21	99	105	113	120
	12	51	53	58	62		22	102	108	116	123
	13	53	56	60	64		23	105	110	119	127
	14	54	58	62	67		24	107	113	122	130
	15	56	60	65	69		25	110	116	126	134
	16	58	62	67	72	11	11	87	91	96	100
	17	60	64	70	75		12	90	94	99	104
	18	62	66	72	77		13	93	97	103	108
	19	64	68	74	80		14	96	100	106	112
	20	66	70	77	83		15	99	103	110	116
	21	68	72	79	85		16	102	107	113	120
	22	70	74	81	88		17	105	110	117	123
	23	71	76	84	90		18	108	113	121	127
	24	73	78	86	93		19	111	116	124	131
	25	75	81	89	96		20	114	119	128	135
9	9	56	59	62	66		21	117	123	131	139
	10	58	61	65	69		22	120	126	135	143
	11	61	63	68	72		23	123	129	139	147
	12	63	66	71	75		24	126	132	142	151
	13	65	68	73	78		25	129	136	146	155
	14	67	71	76	81	12	12	105	109	115	120
	15	69	73	79	84		13	109	113	119	125
	16	72	76	82	87		14	112	116	123	129
	17	74	78	84	90		15	115	120	127	133
	18	76	81	87	93		16	119	124	131	138
	19	78	83	90	96		17	122	127	135	142
	20	81	85	93	99		18	125	131	139	146
	21	83	88	95	102		19	129	134	143	150
	22	85	90	98	105		20	132	138	147	155
	23	88	93	101	108		21	136	142	151	159
	24	90	95	104	111		22	139	145	155	163
	25	92	98	107	114		23	142	149	159	168
10	10	71	74	78	82		24	146	153	163	172
	11	73	77	81	86		25	149	156	167	176
	12	76	79	84	89	13	13	125	130	136	142
	13	79	82	88	92		14	129	134	141	147
	14	81	85	91	96		15	133	138	145	152
	15	84	88	94	99		16	136	142	150	156
	16	86	91	97	103		17	140	146	154	161
	17	89	93	100	106		18	144	150	158	166

Продолжение приложения 8

Объемы выборок		Q			
n_1	n_2	0,005	0,01	0,025	0,05
	19	148	154	163	171
	20	151	158	167	175
	21	155	162	171	180
	22	159	166	176	185
	23	163	170	180	189
	24	166	174	185	194
	25	170	178	189	·199
14	14	147	152	160	166
	15	151	156	164	171
	16	155	161	169	176
	17	159	165	174	182
	18	163	170	179	187
	19	168	174	183	192
	20	172	178	188	197
	21	176	183	193	202
	22	180	187	198	207
	23	184	192	203	212
	24	188	196	207	218
	25	192	200	212	223
15	15	171	176	184	192
	16	175	181	190	197
	17	180	186	195	203
	18	184	190	200	208
	19	189	195	205	214
	20	193	200	210	220
	21	198	205	216	225
	22	202	210	221	231
	23	207	214	226	236
	24	211	219	231	242
	25	216	224	237	248
16	16	196	202	211	219
	17	201	207	217	225
	18	206	212	222	231
	19	210	218	228	237
	20	215	223	234	243
	21	220	228	239	249
	22	225	233	245	255
	23	230	238	251	261
	24	235	244	256	267
	25	240	249	262	273
17	17	223	230	240	249
	18	228	235	246	255
	19	234	241	252	262
	20	239	246	258	268
	21	244	252	264	274
	22	249	258	270	281
	23	255	263	276	287
	24	260	269	282	294
	25	265	275	288	300
18	18	252	259	270	280
	19	258	265	277	287
	20	263	271	283	294
	21	269	277	290	301
	22	275	283	296	307
	23	280	289	303	314
	24	286	295	309	321
	25	292	301	316	328
19	19	283	291	303	313
	20	289	297	309	320
	21	295	303	316	328
	22	301	310	323	335
	23	307	316	330	342
	24	313	323	337	350
	25	319	329	344	357
20	20	315	324	337	348
	21	322	331	344	356
	22	328	337	351	364
	23	335	344	359	371
	24	341	351	366	379
	25	348	358	373	387
21	21	349	359	373	385
	22	356	366	381	393
	23	363	373	388	401
	24	370	381	396	410
	25	377	388	404	418
22	22	386	396	411	424
	23	393	403	419	432
	24	400	411	427	441
	25	408	419	435	450
23	23	424	434	451	465
	24	431	443	459	474
	25	439	451	468	483
24	24	464	475	492	507
	25	472	484	501	517
25	25	505	517	536	552

Приложение 9

Таблица критических значения критерия Манна-Уитни

n_1	Уровень значимости $\alpha = 0{,}05$											
	n_2											
	7	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
3	1	2	2	3	3	4	4	5	5	6	6	7
4	3	4	4	5	6	7	8	9	10	11	11	12
5	5	6	7	8	9	11	12	13	14	15	17	18
6	6	8	10	11	13	14	16	17	19	21	22	24
7	8	10	12	14	16	18	20	22	24	26	28	30
8	10	13	15	17	19	22	24	26	29	31	34	36
9	12	15	17	20	23	26	28	30	34	37	39	42
10	14	17	20	23	26	29	33	36	39	42	45	48
11	16	19	23	26	30	33	37	40	44	48	51	55
12	18	22	26	29	33	37	41	45	49	53	57	61
13	20	24	28	33	37	41	45	50	54	59	63	67
14	22	26	31	36	40	45	50	55	59	64	67	74
15	24	29	34	39	44	49	54	59	64	70	75	80
16	26	31	37	42	47	53	59	64	70	75	81	86
17	28	34	39	45	51	57	63	67	75	81	87	93
18	30	36	42	48	55	61	67	74	80	86	93	99
19	32	38	45	52	58	65	72	78	85	92	99	106
20	34	41	48	55	62	69	76	83	90	98	105	112
	Уровень значимости $\alpha = 0{,}01$											
	n_2											
n_1	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
3			0	0	0	1	1	1	2	2	2	2
4	0	1	1	2	2	3	3	4	5	5	6	6
5	1	2	3	4	4	6	7	7	8	9	10	11
6	3	4	5	6	6	9	10	11	12	13	15	16
7	4	6	7	9	9	12	13	15	16	18	19	21
8	6	7	9	11	11	15	17	18	20	22	24	26
9	7	9	11	13	13	18	20	22	24	27	29	31
10	9	11	13	16	16	21	24	26	29	31	34	37
11	10	13	16	18	18	24	27	30	33	36	39	42
12	12	15	18	21	21	27	31	34	37	41	44	47
13	13	17	20	24	24	31	34	38	42	45	49	53
14	15	18	22	26	26	34	38	42	46	50	54	58
15	16	20	24	29	29	37	42	46	51	55	60	64
16	18	22	27	31	31	41	45	50	55	60	65	70
17	19	24	29	34	34	44	49	54	60	65	70	75
18	21	26	31	37	37	47	53	58	64	70	75	81
19	22	28	33	39	39	51	56	63	69	74	81	87
20	24	30	36	42	42	54	60	67	73	79	86	92

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Бочаров П.П. Теория вероятностей. Математическая статистика / П.П. Бочаров, А.В. Печинкин А.В. - М.: ФИЗМАТЛИТ, 2005. – 296 с.
2. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие для студентов втузов / В.Е. Гмурман. - М.: Высшая школа, 1979.- 400 с.
3. Ивченко Г.И. Математическая статистика / Г.И. Ивченко, Ю.И. Медведев. – М.: Высш.шк., 2003. – 357 с.
4. Кибзун А.И. Теория вероятностей и математическая статистика / А.И. Кибзун, Е.Р. Горяинова, А.В. Наумов, А.Н. Сиротин. - М.: ФИЗМАТЛИТ, 2002. – 224 с.
5. Крамер Г. Математические методы статистики / Г. Крамер. – М.: Мир, 1975. – 648 с.
6. Лагутин М.Б. Наглядная математическая статистика / М.Б. Лагутин. – М.: БИНОМ. Лабораитория знаний. 2007. -472 с.
7. Новикова Н.М. Прикладная математическая статистика учеб. Пособие / Н.М. Новикова, С.Л. Подвальный. – Воронеж: ФГБОУ ВПО «Воронежский государственный технический университет», 2012. Ч.1. – 164 с.
8. Новикова Н.М. Прикладная математическая статистика учеб. Пособие / Н.М. Новикова, С.Л. Подвальный. – Воронеж: ФГБОУ ВПО «Воронежский государственный технический университет», 2013. Ч.2. – 179 с.
9. Сборник задач по математике для втузов. В 4 частях. Ч. 4: Теория вероятностей и математическая статистика: Учеб. пособие для втузов / А.В. Ефимов, А.С. Поспелов.; Под ред. А. В. Ефимова - М.: Издательство Физико-математической литературы, 2003.- 432 с.
10. Статистический анализ данных в психологии: учебное пособие / В.К. Романко. – М.:БИНОМ. Лаборатория знаний, 2015. – 312 с.

11. Теория вероятностей и математическая статистика: учеб. пособие /Л.Н. Фадеева, А.В. Лебедев; под ред. Л.Н. Фадеевой – М.: Рид Групп, 2011. – 496 с.

СОДЕРЖАНИЕ