Н.О.Фастовец, М.А.Попов

Математическая статистика

примеры, задачи и типовые задания

учебное пособие

для нефтегазового образования

Москва 2012

Введение

Основное содержание математической статистики составляют методы систематизации, обработки и использования статистических данных, выявление статистических закономерностей.

Поскольку исходной базой для всех построений математической статистики является рассмотрение конечной совокупности результатов опытов или наблюдений как выборки из некоторой генеральной совокупности, то математическая теория выборки является центральным разделом математической статистики.

Выборка называется случайной , если из генеральной совокупности элементы берутся наугад и в выборку каждый из них может попасть с одинаковой вероятностью.

Если случайная выборка такова, что по её распределению по некоторому признаку можно судить о распределении по этому же признаку неизвестной генеральной совокупности, то такая выборка называется репрезентативной, т.е. хорошо представляющей генеральную совокупность.

Основные задачи при изучении статистических совокупностей, рассматриваемые в этой работе:

- получение рационально выбранных числовых характеристик, которые дали бы общее представление о всей совокупности (1 параграф);
- графическое представление эмпирического материала, дающее приближённые выражения для функции распределения или плотности распределения вероятности (2 параграф);
- интервальное оценивание, позволяющее по данным выборки указать интервал, в котором с высокой вероятностью следует искать истинное, но не известное значение параметра распределения генеральной совокупности (3 параграф);
- статистические методы оценки зависимости между двумя изучаемыми признаками (4 параграф);
- статистическая проверка гипотез, т.е. предположений, относящихся к рассматриваемым распределениям наблюдений. Например, если требуется выяснить, не противоречит ли известному теоретическому значению или ранее проведенным опытам полученный некоторый экспериментальный результат (5 параграф);

- регрессионный анализ, представляющий собой специальный случай метода наименьших квадратов и служащий анализу влияния независимой переменной на зависимую (6 параграф).

В каждом параграфе пособия приводятся основные определения и соответствующие формулы, излагаются решения типовых примеров и затем следуют задачи для решения на практических занятиях. Кроме того, предусмотрены типовые домашние задания по всем темам. Авторы надеются, что выполнение этих заданий в сочетании с использованием компьютерных методов расчётов окажутся полезными студентам всех специальностей и форм обучения. Пособие будет также полезным магистрантам, аспирантам и специалистам в качестве справочного материала при решении практических задач.

§1. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ СТАТИСТИЧЕСКИХ ДАННЫХ

1.1. Числовые характеристики (параметры) выборочного распределения

Последовательность n значений $x_1, x_2, ..., x_n$, полученных в результате наблюдения некоторого процесса, мы будем рассматривать как совокупность значений одинаково распределенных независимых случайных величин $\xi_1, \xi_2, ..., \xi_n$, представляющих собой n экземпляров одной и той же случайной величины ξ. Эта последовательность значений называется ***выборкой.*** В этом случае говорят, что выборка взята из ***генеральной совокупности*** случайной величины ξ. Если величина ξ следует закону распределения $F(x)$, то мы будем говорить, что генеральная совокупность распределена по закону $F(x)$.

Пусть $x_1, x_2, ..., x_n$ − выборка объема n из некоторой генеральной совокупности. По этой выборке можно оценить основные числовые характеристики генеральной совокупности. Различные элементы выборки x_i называются ***вариантами***. Ряд вариант, расположенных в порядке возрастания их значений называется ***вариационным рядом***. Им пользуются, в основном, при малых n. Если n велико, то ряд преобразуют в группировки по отдельным значениям признака x (дискретная группировка) или по интервалам изменения признака (интервальная группировка), для чего разбивают диапазон изменения признака x, называемый размахом $R = x_{\max} - x_{\min}$, на K равных интервалов. Для определения количества интервалов рекомендуется правило $K \approx \sqrt{n}$, где $5 \le K \le 20$. Можно пользоваться и другими эмпирическими формулами, например, формулой Стерджеса $K \approx 1 + 3{,}322 \lg n$, но они дают приблизительно одинаковый результат.

Иногда данные для обработки поступают уже в интервальной группировке, или представляется невозможным использовать одинаковые интервалы (например, в экономике).

Результат группировки представляют рядом вариант или интервалов вариант, расположенных в порядке их возрастания и рядом соответствующих ***частот***. Под ***частотой*** m_i признака или интервала понимают число членов выборки с данной вариантой x_i или, соответственно, число членов выборки, варианты которых лежат в $i-$ м интервале. ***Относительной частотой*** h_i называется отношение частоты m_i к объему выборки: $h_i = \frac{m_i}{n}$. Таким образом, если проведена группировка, то значению x_i или $i-$ му интервалу ($i=1,...,k$) будут отвечать частоты m_i и относительные частоты $h_i = \frac{m_i}{n}$, при этом $\sum_{i=1}^{k} m_i = n$, $\sum_{i=1}^{k} h_i = 1$, а все выборочные значения, попавшие в $i-$ ый интервал, заменяют серединой интервала u_i.

Пример 1. В обувном магазине за день продали 30 пар мужской обуви следующих размеров:
39, 41, 40, 42, 41, 40, 42, 44, 42, 43, 42, 41, 43, 39, 42, 39, 41, 43, 41, 38, 43, 42, 41, 40, 41, 38, 44, 40, 39, 44.

Решение: Проведем группировку по отдельным значениям признака, то есть по размеру обуви (дискретная группировка):

x_i	38	39	40	41	42	43	44	$\sum$
m_i	2	4	4	7	6	4	3	30
h_i	0,067	0,133	0,133	0,233	0,2	0,133	0,1	1

Пример 2. Дана выборка объема $n = 30$. Сделать интервальную группировку этой выборки.

20.3; 15,4; 17,2; 19,2; 23,1; 18,1; 21,9; 15,3; 16,8; 13,2; 20,4; 16,5; 19,7; 20,5; 14,3; 20,1; 16,8; 14,7; 20,8; 19,5; 15,4; 19,3; 17,8; 16,2; 15,7; 22,8; 21,9; 12,5; 10,1; 21,1.

Решение: Диапазон изменения признака x (размах выборки) $R = x_{\max} - x_{\min} = 23{,}1 - 10{,}1 = 13$. Количество интервалов K можно взять равным 5 ($K \approx \sqrt{30}$), ширина интервала $d = \dfrac{R}{K} = \dfrac{13}{5} = 2{,}6$. Таблица интервальной группировки будет выглядеть следующим образом:

Интервалы	10,1–12,7	12,7–15,3	15,3–17,9	17,9–20,5	20,5–23,1
m_i	2	4	9	9	6
h_i	0,067	0,133	0,3	0,3	0,2
середины интервалов u_i	11,4	14,0	16,6	19,2	21,8

Варианты, попадающие на границу интервала, отнесены к левому интервалу (можно отнести их и к правому интервалу, а в том случае, если на границу попадает много вариант, можно их поделить пополам между соседними интервалами).

Иногда статистические данные, полученные для обработки, уже сгруппированы по интервалам. Например, распределение роста 1000 взрослых мужчин имеет вид:

Рост (см)	143-149	149-155	155-161	161-167	167-173	173-179	179-185	185-191
Число мужчин, m_i	3	34	185	382	290	92	13	1

В этом случае ширина интервала и количество интервалов заданы.

Итак, пусть $x_1, x_2, ..., x_n$ – выборка объема n из генеральной совокупности, имеющей функцию распределения $F(x)$. Числовые характеристики выборки называются ***выборочными*** (эмпирическими) числовыми характеристиками. Введем основные числовые характеристики:

– ***среднее арифметическое (среднее)***

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i ; \tag{1.1}$$

Если сделана дискретная группировка, то

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{K} m_i x_i = \sum_{i=1}^{K} h_i x_i , \tag{1.2}$$

при интервальной группировке

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{K} m_i u_i = \sum_{i=1}^{K} h_i u_i , \tag{1.3}$$

где m_i – частота, K – количество интервалов, u_i – середина i -го интервала.

– ***выборочная (эмпирическая) дисперсия***

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\cdot\left(\sum_{i=1}^{n} x_i^2 - n\left(\overline{x}\right)^2\right). \tag{1.4}$$

При дискретной группировке

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{K} m_i (x_i - \overline{x})^2 = \frac{1}{n-1}\cdot\left(\sum_{i=1}^{K} m_i x_i^2 - n\left(\overline{x}\right)^2\right), \tag{1.5}$$

при интервальной группировке

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{K} m_i (u_i - \overline{x})^2 = \frac{1}{n-1}\cdot\left(\sum_{i=1}^{K} m_i u_i^2 - n\left(\overline{x}\right)^2\right). \quad (1.6)$$

ЗАМЕЧАНИЕ. Множитель $\frac{1}{n-1}$ объясняется тем, что $M[S^2] = \sigma^2$, т.е. S^2 является несмещенной оценкой σ^2. Если использовать формулу $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, то $M[S_1^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$, т.е. S_1^2 является смещенной оценкой дисперсии, уменьшающей ее.

– ***стандартное (среднее квадратическое) отклонение*** является корнем из выборочной дисперсии:

$$S = \sqrt{\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\frac{1}{n-1}\cdot\left(\sum_{i=1}^{n} x_i^2 - n\left(\overline{x}\right)^2\right)}. \quad (1.7)$$

Для сгруппированных данных, соответственно, имеем формулы:

$$S = \sqrt{\frac{1}{n-1}\cdot\left(\sum_{i=1}^{K} m_i x_i^2 - n\left(\overline{x}\right)^2\right)} \quad (1.8)$$

$$S = \sqrt{\frac{1}{n-1}\cdot\left(\sum_{i=1}^{K} m_i u_i^2 - n\left(\overline{x}\right)^2\right)} \quad (1.9)$$

– ***коэффициент вариации***

$$V = \frac{S}{\overline{x}}, \quad \overline{x} \neq 0; \quad (1.10)$$

– ***выборочные начальные и центральные моменты порядка*** ℓ ($\ell = 1, 2, \ldots$), соответственно, определяются формулами

$$\alpha_\ell = \frac{1}{n}\sum_{i=1}^{n} x_i^\ell, \qquad \beta_\ell = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^\ell \qquad (1.11)$$

Если данные сгруппированы, то получаем следующие формулы:

при дискретной группировке

$$\alpha_\ell = \frac{1}{n}\sum_{i=1}^{K} m_i x_i^\ell, \qquad \beta_\ell = \frac{1}{n}\sum_{i=1}^{K} m_i (x_i - \overline{x})^\ell \qquad (1.12)$$

при интервальной группировке

$$\alpha_\ell = \frac{1}{n}\sum_{i=1}^{K} m_i u_i^\ell, \qquad \beta_\ell = \frac{1}{n}\sum_{i=1}^{K} m_i (u_i - \overline{x})^\ell \qquad (1.13)$$

– ***оценка коэффициента асимметрии*** S_k (характеризует симметричность распределения относительно среднего $\overline{x}$) определяется по формуле

$$\hat{S}_k = \frac{\beta_3}{(\beta_2)^{\frac{3}{2}}}. \qquad (1.14)$$

– ***оценка эксцесса*** E_x (меры островершинности распределения по сравнению с нормальным распределением)

$$\hat{E}_x = \frac{\beta_4}{\beta_2^2} - 3. \qquad (1.13)$$

Если $\hat{E}_x > 0$, то вершина более острая, а если $\hat{E}_x < 0$, то более плоская, чем у нормального распределения. У нормального распределения $E_x = 0$.

– ***выборочная мода*** $\hat{M}_0$

Для дискретного вариационного ряда (дискретная группировка) мода определяется как значение варианты с наибольшей частотой, если выборка достаточно большая.

При интервальной группировке выбирается интервал, которому соответствует наибольшая частота. Пусть это k–й интервал $(t_{k-1} - t_k)$, его частота равна m_k, а ширина d, тогда

$$\hat{M}_0 = t_{k-1} + d\frac{m_k - m_{k-1}}{2m_k - m_{k-1} - m_{k+1}} \qquad (1.14)$$

– ***выборочная медиана*** $\hat{M}_e$

Определяется, как значение признака, относительно которого выборка делится на две равные по объему части. Если выборка объема n представлена вариационным рядом, то

$$\hat{M}_e = \begin{cases} x_{k+1}, n = 2k+1 \\ \dfrac{x_k + x_{k+1}}{2}, n = 2k. \end{cases} \qquad (1.15)$$

При интервальной группировке (интервальный вариационный ряд) сначала находят так называемый медианный интервал $(t_{s-1} - t_s)$, номер s которого определяют из неравенств:

$$\sum_{i<s} m_i \le \frac{n}{2}, \quad \sum_{i \le s} m_i > \frac{n}{2},$$

где $\sum\limits_{i<s} m_i$ – сумма частот всех интервалов левее медианного, $\sum\limits_{i \le s} m_i$ – сумма частот, включающая частоту медианного интервала.

Медиану оценивают с помощью следующей интерполяционной формулы:

$$\hat{M}_e = t_{s-1} + d\left(\frac{\frac{n}{2} - \sum\limits_{i<s} m_i}{m_s}\right), \qquad (1.16)$$

где m_S – частота медианного интервала.

Пример 3. Дана выборка объема $n = 40$ с интервальной группировкой:

Интервалы	5–7	7–9	9–11	11–13	13–15	15–17	17–19
m_i	4	8	11	7	5	3	2

Найти оценки моды и медианы для этой выборки.

Решение:

Оценка моды. Ширина интервала равна 2, а наибольшая частота отвечает третьему интервалу (9 – 11) и равна $m_3 = 11$, поэтому согласно формуле (1.14)

$$\hat{M}_0 = 9 + 2\frac{11-8}{2\cdot 11 - 8 - 7} = 9,86.$$

Оценка медианы. $\frac{n}{2} = \frac{40}{2} = 20$. Определим медианный интервал:

$$m_1 + m_2 = 4 + 8 = 12 < \frac{n}{2} = 20,$$

$$m_1 + m_2 + m_3 = 4 + 8 + 11 = 23 > \frac{n}{2} = 20,$$

следовательно, третий интервал (9–11) – медианный. Поэтому согласно формуле (1.16) получаем оценку медианы

$$\hat{M}_e = 9 + 2\left(\frac{20-(4+8)}{11}\right) = 10,45.$$

1.2. Предварительная проверка на нормальность

С помощью вычисленных числовых характеристик можно определить, является ли выборочное распределение близким к нормальному. Если выборочное распределение близко к ***нормальному*** (или является таковым), то:

1. в интервалы $\overline{x} \pm s$, $\overline{x} \pm 2s$ и $\overline{x} \pm 3s$ должны попадать соответственно приблизительно 68%, 95% и 100% выборочных значений;

2. в не слишком маленькой выборке величина коэффициента вариации V должна быть не более 33% , т.е. $V < 0,33$ (см., например, [4]);

3. Оценка эксцесса $\hat{E}_x$ и коэффициента асимметрии $\hat{S}_k$ должны быть близки к нулю;

4. $\bar{x} \approx \hat{M}_e$.

Пример 4. В результате измерения температуры раздела фракции бензин-авиакеросин на установке первичной переработки нефти были получены значения температур, приведенные в таблице 1 (в градусах Цельсия).

Таблица 1

N	Значение	N	Значение	N	Значение	N	Значение
1	133,5	14	141,5	27	144,0	40	137,5
2	142,0	15	139,0	28	142,5	41	141,5
3	145,5	16	140,5	29	139,0	42	141,0
4	144,5	17	139,0	30	137,0	43	142,5
5	134,5	18	143,5	31	136,0	44	143,5
6	138,5	19	139,5	32	137,0	45	141,0
7	144,0	20	140,5	33	138,5	46	147,0
8	141,0	21	140,0	34	139,0	47	139,5
9	141,5	22	138,5	35	139,5	48	136,5
10	139,5	23	135,0	36	140,5	49	142,0
11	140,0	24	139,5	37	139,5	50	140,0
12	145,0	25	139,0	38	140,0		
13	141,5	26	138,0	39	140,5		

По этой выборке вычислены основные числовые характеристики для объема выборки $n = 50$:

$\bar{x} = 140,2$; $\hat{S}_k = 0{,}0083$;

$S^2 = 7,755$; $\hat{E}_x = 0{,}19$;

$S = 2,78$; $\hat{M}_e = 140{,}0$.

$$V = \frac{S}{\bar{x}} = 0,02$$

Проведена предварительная проверка на нормальность распределения:

1. в интервалы $\bar{x} \pm s,\ \bar{x} \pm 2s,\ \bar{x} \pm 3s$, которые в этом случае равны соответственно:

$$140,2 \pm 2,78,\ (137,40 \div 142,98);$$
$$140,2 \pm 5,56,\ (134,64 \div 145,76);$$
$$140,2 \pm 8,34,\ (131,86 \div 148,54),$$

попало, соответственно, 70%, 94% и 100 % выборки;

2. величина $V = \dfrac{S}{\bar{x}} = 0,02 < 0,33$;

3. $\hat{E}_x = 0{,}19$ и $\hat{S}_k = 0{,}0083$ можно считать близкими к нулю;

4. $\bar{x} = 140{,}2 \approx \hat{M}_e = 140{,}0$.

Предварительный анализ показывает, что распределение температуры раздела фракции бензин-авиакеросин не противоречит предположению о нормальности.

1.3. Задачи

Для каждой из приведенных выборок вычислить основные числовые характеристики. Провести предварительную проверку на нормальность распределения в задачах **1.1. − 1.3**.

1.1. 11, 15, 12, 0, 16, 19, 6, 11, 12, 13, 16, 8, 9, 14, 5, 11, 3.

1.2. 3,1; 3,0; 1,5; 1,8; 2,5; 3,1; 2,4; 2,8; 1,3.

1.3. Распределение скорости автомобилей на одном из участков шоссе (км/час).

Интервалы	61 − 69	69 − 77	77 − 85	85 − 93	93 −101
Частота	5	13	23	7	2

1.4. Как изменятся выборочные среднее и дисперсия, если результаты наблюдения подвергнуть преобразованию масштаба, т.е. увеличить или уменьшить одновременно в k раз?

1.5. Найти выборочную дисперсию и коэффициент вариации признака по данному распределению.

Интервалы	9 – 12	12 – 15	15 – 18	18 – 21	21 – 24	24 – 27
m_i	6	12	33	22	19	8

1.6. Дано распределение. Найти оценки асимметрии и эксцесса.

x_i	-4	-3	-2	-1	0	2	3
m_i	1	5	10	7	4	2	1

§2. ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ВЫБОРОЧНОГО (ЭМПИРИЧЕСКОГО) РАСПРЕДЕЛЕНИЯ

Наиболее распространенными способами графического представления эмпирических данных (выборки) являются ***гистограмма***, ***полигон частот*** и ***эмпирическая функция распределения*** (накопленные относительные частоты).

Пусть $x_{\min}$ и $x_{\max}$ – соответственно наименьшее и наибольшее значения вариант выборки. Величина $R = x_{\max} - x_{\min}$ называется ***размахом*** выборки. Размах делится на число интервалов K (интервальная группировка), которое можно вычислить по одной из следующих формул:

$$K \approx \sqrt{n},$$

$$K < 5 \cdot \lg n \qquad ,$$

$$K \approx 1 + 3{,}322 \cdot \lg n \quad \text{(ôî ðì óëà Ñòåðäæåñà)}$$

округлив до целого числа.

Обычно предполагают, что количество интервалов должно удовлетворять условию $5 \le K \le 20$.

Ширина каждого интервала d вычисляется по формуле $d = \dfrac{R}{K}$. После разбиения на интервалы определяют:

- абсолютные частоты m_i, $i = 1,..,K$, где m_i – количество элементов выборки, попавших в $i-$й интервал (элемент, попавший на границу интервала, относят к какому-нибудь выбранному интервалу, например, левому, или правому; если на границу интервала попадает много элементов выборки, то их делят пополам между левым и правым интервалами);
- относительные частоты $h_i = \dfrac{m_i}{n}$;

- относительные накопленные частоты $\sum_{j=1}^{i}\frac{m_j}{n}=\sum_{j=1}^{i}h_j, i=1,...,K;$;
- середины интервалов u_i.

Все полученные результаты сводятся в таблицу.

Номер интервала	Границы интервалов	m_i	h_i	$\sum_{j=1}^{i}h_j$	u_i
1		m_1	h_1	h_1	z_1
2		m_2	h_2	h_1+h_2	z_2
…		…	…	…	…

При этом $\sum_{i=1}^{K}m_i=n, \sum_{i=1}^{K}h_i=1$.

2.1. Гистограмма и полигон частот

Гистограмма строится следующим образом. На оси абсцисс откладываются интервалы, и на каждом из них строится прямоугольник, площадь которого равна относительной частоте, соответствующей этому интервалу, т.е. высота прямоугольника (ордината) равна $\frac{h_i}{d}=\frac{m_i}{nd}$, так что полная площадь гистограммы равна 1. (Гистограмма является эмпирическим аналогом плотности распределения). Так как множители $\frac{1}{nd}$ (или $\frac{1}{d}$) можно рассматривать как масштабные, то по оси ординат можно откладывать частоты m_i или h_i (правда, в этом случае площадь всех прямоугольников будет равна nd или d).

Полигон частот - ломаная линия, которая получается, если из середины каждого интервала u_i восстановить перпендикуляр высотой h_i (или m_i) и соединить вершины этих перпендикуляров. Полигон частот чаще используют при дискретной группировке.

2.2. Эмпирическая функция распределения

Эмпирическую функцию распределения $F_n(x)$ получают построением ступенчатой кривой относительных накопленных частот; $F_n(x)$ имеет скачки в точках, соответствующих серединам интервалов u_i.

По гистограмме и полигону частот судят о виде плотности распределения исследуемой непрерывной случайной величины или о распределении вероятностей дискретной случайной величины. Эмпирическая функция распределения дает представление о функции распределения и используется в основном в статистической проверке гипотез. Кроме того, эмпирическая функция распределения используется для определения эмпирических (выборочных) квантилей. **Квантилем** порядка p $(0 < p < 1)$ называется величина x_p, определяемая из соотношения $P\{\xi < x_p\} = F(x_p) = p$. **Выборочный квантиль** x_p определяется из соотношения $F_n(x_p) = p$. Задавая p, можно по графику эмпирической функции распределения оценить, например, значение x_p, которое исследуемая величина не превзойдет с вероятностью p, либо, задавая x_p, по тому же графику оценить соответствующую вероятность p. По $F_n(x)$ можно оценить медиану из соотношения $F_n(\hat{M}_e) = 0{,}5$.

Пример 1. Построение гистограммы, полигона частот и эмпирической функции распределения по данным примера 4 из §1.

$x_{\min} = 133{,}5$,

$x_{\max} = 147{,}0$,

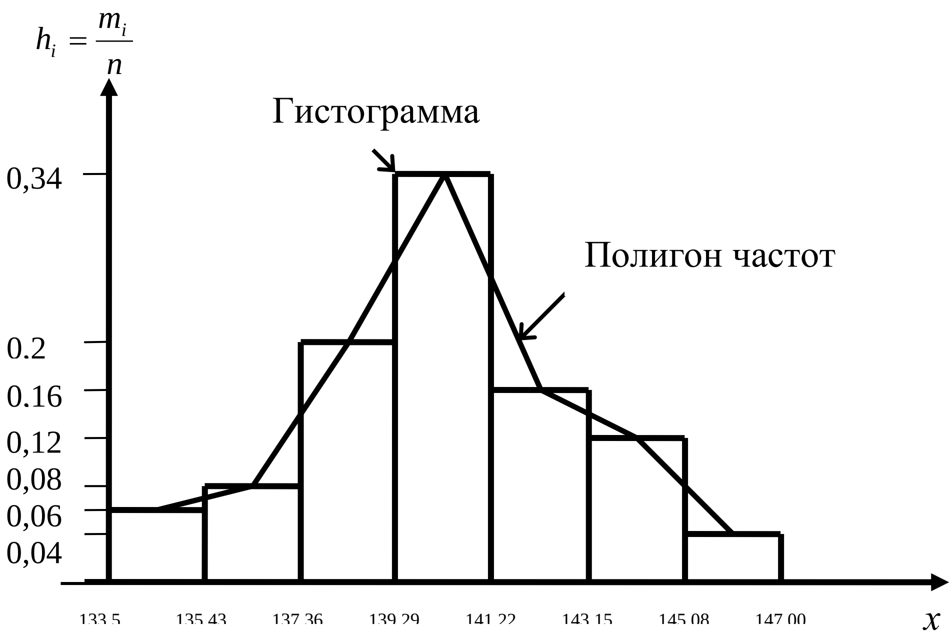
$R = 13{,}5$,

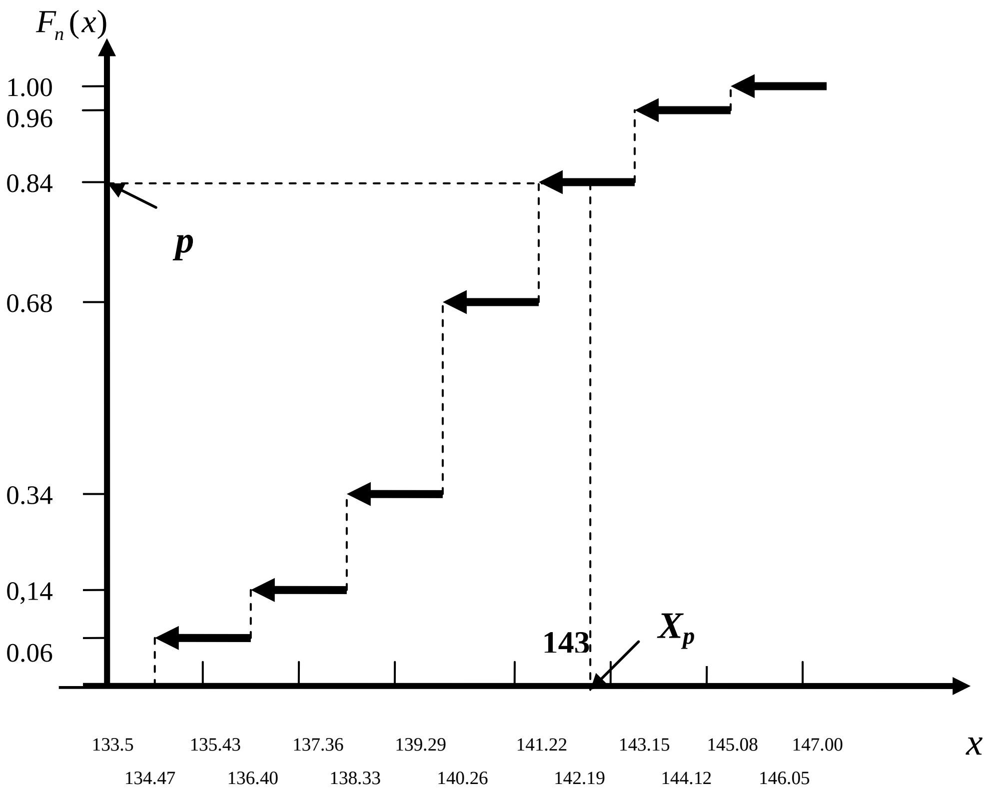
$n = 50$.

Количество интервалов: $K \approx \sqrt{n},\ K = 7$.

Ширина интервала: $d = \dfrac{R}{K} = 1,93$.

Номер интервала	Границы	m_i	$h_i = m_i / n$	$\sum_{j=1}^{i} h_j$	u_i
1	133,50 − 135,43	3	0,06	0,06	134,47
2	135,43 − 137,36	4	0,08	0,14	136,40
3	137,36 − 139,29	10	0,20	0,34	138,33
4	139,29 − 141,22	17	0,34	0,68	140,26
5	141,22 − 143,15	8	0,16	0,84	142,19
6	143,15 − 145,08	6	0,12	0,96	144,12
7	145,08 − 147,00	2	0,04	1	146,05





Эмпирическая функция распределения

По эмпирической функции распределения можно оценить, например, вероятность того, что температура раздела фракции бензин-авиакеросин будет ниже $143^{o}C$:

$$P\{T < 143\} \approx 0{,}84.$$

2.3. Задачи

2.1. Построить гистограмму и эмпирическую функцию распределения по данным задачи 1.3. Оценить вероятность того, что скорость превысит 80 км/час.

2.2. Построить полигон частот и эмпирическую функцию распределения для распределения 45 пар мужской обуви, проданных магазином за день:

39, 41, 40, 42, 41, 40, 42, 44, 40, 43, 42, 41, 43, 39, 42, 41, 42, 39, 41, 37, 43, 41, 38, 43, 42, 41, 40, 41, 38, 44, 40, 39, 41, 40, 42, 40, 41, 42, 40, 43, 38, 39, 41, 41, 42.

Оценить по эмпирической функции распределения медиану.

2.3. Через каждый час измерялось напряжение в электросети. При этом были получены следующие значения (в вольтах):

227, 219, 215, 230, 232, 223, 220, 222, 218, 219, 222, 221, 227, 226, 226, 209, 211, 215, 218, 220, 216, 220, 221, 225, 224, 212, 217, 219, 220.

Построить гистограмму, полигон частот, эмпирическую функцию распределения; оценить вероятность того, что напряжение не превосходит 220 В.

§3. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ

В §1 мы рассмотрели оценки параметров распределения с помощью числовых значений, которые находятся по выборке и называются точечными. К сожалению, точечные оценки параметров не дают представления о степени точности, т.е. близости оценки к оцениваемому параметру, и о надежности полученной оценки. Особенно часто вопрос о точности и надежности оценки возникает тогда, когда объем выборки недостаточно велик. В таких случаях для того, чтобы получить представление о точности и надежности оценки $\hat{\theta}$ для параметра θ, строят так называемые интервальные оценки, которые еще называют доверительными интервалами.

Пусть $\hat{\theta}_1(x_1,...,x_n)$ и $\hat{\theta}_2(x_1,...,x_n)$ − такие функции от выборочных значений $x_1,...,x_n$, что

$$P\{\hat{\theta}_1 \le \theta \le \hat{\theta}_2\} \ge 1-\alpha, \qquad (3.1)$$

где α – небольшое число: $\alpha = 0,01;\ \ 0,05;\ \ 0,1.$ Обычно величину $1-\alpha$ обозначают через γ.

Интервал $\left(\hat{\theta}_1, \hat{\theta}_2\right)$ называется доверительным интервалом для параметра θ с коэффициентом доверия (надежности) $1-\alpha$.

Рассмотрим несколько случаев построения доверительных интервалов, если выборка $x_1,...,x_n$ взята из нормально распределенной генеральной совокупности.

3.1. Доверительный интервал для математического ожидания при известной дисперсии

А) При построении доверительного интервала, используется точечная оценка параметра μ – математического ожидания, т.е.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Из свойств математического ожидания и дисперсии следует, что

$$M\overline{x} = \mu;\ \ D\overline{x} = \frac{\sigma^2}{n} \qquad (3.2)$$

Поэтому, нормированная случайная величина

$$\frac{\left(\overline{x}-\mu\right)}{\sigma}\cdot\sqrt{n} \qquad (3.3)$$

имеет нормированное нормальное распределение $N(0,1)$. Границы доверительного интервала будем находить из соотношения

$$P\left\{\left|\overline{x}-\mu\right| \le \delta\right\} = 1-\alpha, \qquad (3.4)$$

которое, используя предыдущую формулу (3.3), преобразуем к виду

$$P\left\{\frac{|\bar{x}-\mu|}{\sigma}\sqrt{n} \le \frac{\delta\sqrt{n}}{\sigma}\right\} = 1-\alpha. \qquad (3.5)$$

Отсюда получаем:

$$2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) = 1-\alpha \qquad (3.6)$$

Здесь $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-t^2/2}dt$ − функция Лапласа.

Из последнего соотношения (3.6) по таблице значений функции Лапласа (*Таблица II* Приложения), находим значение z_α:

$$z_\alpha = \frac{\delta\sqrt{n}}{\sigma}, \qquad (3.7)$$

из которого определяем величину

$$\delta = z_\alpha \frac{\sigma}{\sqrt{n}}. \qquad (3.8)$$

Обычно используются стандартные значения (Таблица 2)

Таблица 2.

α	0,01	0,05	0,1
z_α	2,576	1,96	1,64

Величина δ задает границы доверительного интервала, т.е. определяет точность интервальной оценки. Коэффициент доверия $1-\alpha=\gamma$ имеет следующий смысл: если мы будем повторять выборку и для каждой из них находить доверительный интервал, то в среднем на 100 выборок доля тех интервалов, которые накроют оцениваемый параметр, составит $(1-\alpha)\cdot 100\%$.

Итак, мы получили доверительный интервал для параметра μ при известной дисперсии σ^2, как следует из (3.4):

$$\left|\overline{x}-\mu\right| \le z_\alpha \frac{\sigma}{\sqrt{n}}$$

или в более принятой форме:

$$\overline{x}-z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x}+z_\alpha \frac{\sigma}{\sqrt{n}} \qquad (3.9)$$

Из анализа формулы (3.9) можно сделать следующие выводы:

а) при увеличении объема выборки n точность интервальной оценки увеличивается, так как величина $\delta = z_\alpha \frac{\sigma}{\sqrt{n}}$ уменьшается. При больших n хорошей оценкой для μ становится $\overline{x}$, т.е. точечная оценка;

б) в силу того, что функция $\Phi(x)$ является неубывающей, при увеличении надежности γ растет величина δ, т.е. уменьшается точность (интервал становится шире);

в) для фиксированных значений надежности $1-\alpha=\gamma$ и точности δ из формулы (3.8) можно определить необходимый объем выборки, обеспечивающий заданное значение $1-\alpha$ и δ. Следует запомнить, что при неизменном объеме выборки одновременно увеличивать точность и надежность оценки **нельзя**.

Пример 1. При измерении уровня шума вырубочного пресса ПВГ-18 были получены следующие значения (дБ): 121,7; 117; 132,4; 117,9; 103,5 $(n=5)$. Считая дисперсию известной и равной $\sigma^2=26$, найти доверительный интервал для математического ожидания уровня шума с надежностью $\gamma=0,95$ ($\alpha=1-\gamma=0,05$).

Решение. Вычислим: $\overline{x}=\dfrac{121,7+117+132,4+117,9+103,5}{5}=118,5$.

Для $\alpha=0,05$ соответствующее значение $z_{\alpha}=1,96$:

$$\frac{\delta\sqrt{n}}{\sigma}=1,96 \qquad \sqrt{n}=2,24 \qquad \sigma=5,1$$

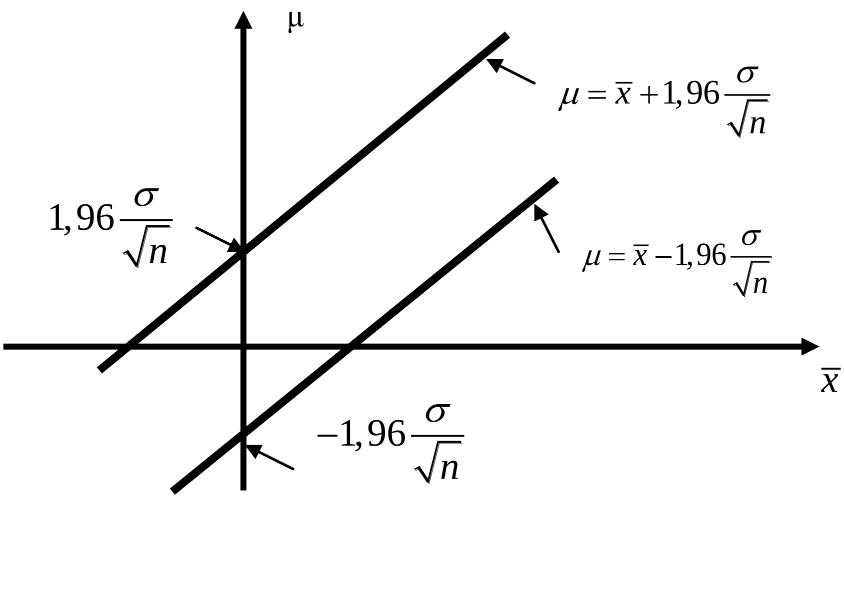
$$\delta=z_{\alpha}\frac{\sigma}{\sqrt{n}} \qquad \delta=1,96\frac{5,1}{2,24}=4,9$$

Окончательно получаем:

$$118,5-4,9\le\mu\le 118,5+4,9$$

$$113,6\le\mu\le 123,4$$

Б) Доверительный интервал для данных значений σ и n можно представить с помощью доверительной полосы графически



Границы задаются уравнениями:

$$\mu = \bar{x} + 1,96\frac{\sigma}{\sqrt{n}}$$

$$\mu = \bar{x} - 1,96\frac{\sigma}{\sqrt{n}}$$

Для любого вычисленного значения $\bar{x}$ по выборке данного объема n, используя рисунок, можно определить границы доверительного интервала для μ. При росте n границы доверительной полосы будут стремиться к линии $\mu = \bar{x}$. Рассмотренный доверительный интервал симметричен относительно $\bar{x}$. Кроме того, вероятность превзойти левую, либо правую границу интервала одинакова и равна $\frac{\alpha}{2} = 0,025$.

В) Для той же надежности γ можно построить и другой интервал, несимметричный. Вероятность того, что оцениваемый параметр лежит вне интервала, равна $\alpha = 1 - \gamma = 0,05$, но вероятность того, что он лежит левее левой границы равна, например, 0,01, а вероятность того, что он лежит правее правой границы, равна 0,04 (0,01+0,04=0,05).

В этом случае вместо соотношения (3.5) будем использовать

$$P\left\{z_1 \le \frac{(\bar{x} - \mu)}{\sigma}\sqrt{n} \le z_2\right\} = 0,95, \qquad (3.10)$$

а значения z_1 и z_2 найдем из соотношений:

$$P\left\{\frac{(\bar{x} - \mu)}{\sigma}\sqrt{n} \le z_1\right\} = F(z_1) = 0,01,,$$

$$P\left\{\frac{(\bar{x} - \mu)}{\sigma}\sqrt{n} \le z_2\right\} = F(z_2) = 0,96,$$

где $F(x)$ – функции распределения $N(0,1)$. Поскольку существует очевидное соотношение: $F(x) = 0,5 + \Phi(x)$, (значения $\Phi(x)$ берутся из *Таблицы II* Приложения), получаем, что

$$z_1 = -2,3 \qquad \left(F(-2,3) = 0,0107\right),$$

$$z_2 = 1,75 \qquad \left(F(1,75) = 0,9599\right).$$

Так как $z_1 \neq z_2$, то интервал несимметричен относительно $\overline{x}$. Следует отметить, что ширина симметричного интервала для $\gamma = 0,95$ равна $3,92 \cdot \dfrac{\sigma}{\sqrt{n}}$, а для несимметричного интервала $(z_2 - z_1) \cdot \dfrac{\sigma}{\sqrt{n}} = 4,05 \cdot \dfrac{\sigma}{\sqrt{n}}$, т.е. при прочих равных условиях симметричный интервал уже, а, значит, оценка точнее.

Однако, в некоторых случаях, в целях гарантии безопасности, имеет смысл строить несимметричный интервал, например, при оценке содержания в воздухе токсичного вещества, вредного в больших дозах. Здесь допустима ошибка только в безопасную сторону. Опасностью является превышение истинного значения содержания вещества над оцениваемым. При заданной надежности γ вероятность того, что истинное значение содержания вещества превзойдет правую, т.е. опасную границу доверительного интервала, должно быть близким к нулю, т.е.

$$P\left\{\mu \geq \overline{x} + \delta\right\} \approx 0.$$

Если же оценивается прочность конструкции, то важно, чтобы прочность конструкции была не меньше допустимой:

$$P\left\{\mu \leq \overline{x} - \delta\right\} \approx 0.$$

3.2. Доверительный интервал для математического ожидания при неизвестной дисперсии

В этом случае, помимо точечной оценки для μ, будем еще использовать точечную оценку для дисперсии:

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

Для построения доверительного интервала используется величина

$$\frac{(\overline{x}-\mu)}{s}\sqrt{n},$$

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

Доверительный интервал будем находить из соотношения (3.11):

$$P\left\{\frac{|\overline{x}-\mu|}{s}\sqrt{n} \le \frac{\delta\sqrt{n}}{s}\right\} = 1-\alpha. \qquad (3.11)$$

Теперь для определения величины δ воспользуемся таблицами распределения Стьюдента (*Таблица III* Приложения)

откуда

$$\frac{\delta\sqrt{n}}{s} = t_{\alpha,n-1}; \quad \delta = t_{\alpha,n-1}\cdot\frac{s}{\sqrt{n}}$$

и, следовательно, доверительный интервал для μ запишется в следующем виде:

$$\overline{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}} \le \mu \le \overline{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}} \qquad (3.12)$$

ЗАМЕЧАНИЕ. Распределение Стьюдента при $n>50$ близко к стандартному нормальному распределению. Поэтому, при $n>50$ для определения величины δ можно пользоваться не значением $t_{\alpha,n-1}$, а соответствующим значением z_α.Например, при $n=60$:

α	0,1	0,05	0,01
z_α	1,64	1,96	2,576
$t_{\alpha,59}$	1,67	2,00	2,66

Выводы а) − в) остаются в силе и для соотношения (3.12).

Пример 2. При замере освещенности в одной из лабораторий были получены следующие значения в лк. 356,4; 353,3; 354,3; 350,5; 357,2. Найти доверительные границы для математического ожидания уровня освещенности при коэффициенте доверия $\gamma=0,95$ ($n=5$).

Решение. $\overline{x}=354,9$; $n=5$; $\gamma=0,95$; $\alpha=0,05$; $\sqrt{5}=2,236$;

$$s^2=\frac{1}{4}\sum_{i=1}^{5}(x_i-354,9)^2=6,86; \quad s=2,62.$$

Для $\gamma=0,95$ и четырех степеней свободы по таблицам распределения Стьюдента находим $t_{\alpha,n-1}=2,776$. Следовательно, доверительный интервал, определяемый по формуле (3.12), запишется в виде:

$$354,9-2,776\frac{2,62}{2,236}\le\mu\le 354,9+2,776\frac{2,62}{2,236}$$

или

$$352,5\le\mu\le 357,3.$$

3.3. Доверительный интервал для дисперсии

При построении доверительного интервала для дисперсии воспользуемся тем, что величина $\frac{(n-1)s^2}{\sigma^2}$ принадлежит распределению Пирсона (χ^2) с $(n-1)$ степенями свободы.

Доверительный интервал будем находить из следующего соотношения:

$$P\left\{\left|s^2-\sigma^2\right|\le\delta\right\}=1-\alpha=\gamma\,, \tag{3.13}$$

которое с помощью тождественных преобразований сводится к виду:

$$P\left\{\underline{x}^2\le\frac{(n-1)s^2}{\sigma^2}\le\overline{x}^2\right\}=\gamma\,. \tag{3.14}$$

Здесь $\underline{x}^2$ – нижняя граница доверительного интервала для случайной величины $\frac{(n-1)s^2}{\sigma^2}$; $\overline{x}^2$ – верхняя граница.

Значение $\underline{x}^2$ и $\overline{x}^2$ находим из таблиц распределения χ^2 (*Таблица III* Приложения) чаще всего из следующих соотношений:

$$P\left\{\chi^2_{n-1}\ge\overline{x}^2\right\}=\frac{\alpha}{2};\;\; P\left\{\chi^2_{n-1}\ge\underline{x}^2\right\}=1-\frac{\alpha}{2} \tag{3.15}$$

Найденные значения $\underline{x}^2$ и $\overline{x}^2$ подставляем в выражение

$$\underline{x}^2\le\frac{(n-1)s^2}{\sigma^2}\le\overline{x}^2, \tag{3.16}$$

откуда находим искомый доверительный интервал для σ^2:

$$\frac{(n-1)\cdot s^2}{\overline{x}^2} \le \sigma^2 \le \frac{(n-1)\cdot s^2}{\underline{x}^2} \qquad (3.17)$$

При $n > 30$ распределение χ_n^2 близко к $N(n, 2n)$. Поэтому в тех случаях, когда объем выборки $n > 30$, можно пользоваться для нахождения границ доверительного интервала стандартными значениями z_α нормального распределения, предварительно сделав нормировку случайной величины $\chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma^2}$.

Итак,

$$\frac{(n-1)s^2}{\sigma^2} \in N(n-1; 2(n-1))$$

Нормируем случайную величину $\frac{(n-1)s^2}{\sigma^2}$:

$$\frac{\frac{(n-1)s^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}} \in N(0;1)$$

Для данного значения надежности $1-\alpha$ находим значение z_α.

Тогда

$$P\left\{\left|\frac{\frac{(n-1)s^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}}\right| \le z_\alpha\right\} = 1-\alpha$$

Путем тождественных преобразований выражения, стоящего в фигурных скобках, получаем доверительный интервал для σ^2:

$$\frac{(n-1)\cdot s^2}{(n-1) + z_\alpha\sqrt{2(n-1)}} \le \sigma^2 \le \frac{(n-1)\cdot s^2}{(n-1) - z_\alpha\sqrt{2(n-1)}}. \qquad (3.18)$$

Например, при $n = 33$ и $\alpha = 0{,}05, z_\alpha = 1{,}96$

$\underline{x}^2 = 16{,}8,$ а $(n-1) - z_\alpha \cdot \sqrt{2(n-1)} \approx 16{,}32;$

$\overline{x}^2 = 47{,}0,$ а $(n-1) + z_\alpha \cdot \sqrt{2(n-1)} \approx 47{,}68.$

Пример 3. Используя данные предыдущего примера, построить доверительный интервал для σ^2.

Решение. $n = 5$; $\gamma = 0{,}95$; $s^2 = 6{,}86.$

$$P\left\{\chi_4^2 \geq \overline{x}^2\right\} = 0{,}025; \quad P\left\{\chi_4^2 \geq \underline{x}^2\right\} = 0{,}975$$

Воспользовавшись таблицей, находим (*Таблица III* Приложения)

$$\underline{x}^2 = 0{,}484; \ \overline{x}^2 = 11{,}1$$

$$\frac{4 \cdot 6{,}86}{11{,}1} \leq \sigma^2 \leq \frac{4 \cdot 6{,}86}{0{,}484}$$

$$2{,}47 \leq \sigma^2 \leq 57{,}1$$

Доверительный интервал получился очень широким, что, безусловно, объясняется очень маленьким объемом выборки.

3.4. Доверительный интервал для параметра *p* биномиального распределения

Точечной оценкой параметра p биномиального распределения является относительная частота $h = \dfrac{m}{n}$, где n – число опытов, а m – число наступлений рассматриваемого события (число успехов), при этом

$M[h] = p$, $D[h] = \dfrac{pq}{n}$.

Доверительный интервал для неизвестного параметра *p*

$$p_1 < p < p_2 \qquad (3.19)$$

при больших *n* находим, пользуясь тем, что

$$\frac{h-p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n\to\infty]{} N(0;1),$$

т.е. нормальной асимптотикой.

Для заданного значения $\gamma = 1-\alpha$ находят z_α и из соотношения

$$(h-p)^2 < \frac{z_\alpha^{\ 2}}{n} p(1-p) \qquad (3.20)$$

получают значения p_1 и p_2

$$\left.\begin{aligned} p_1 &= \frac{m + \frac{1}{2}z_\alpha^{\ 2} - z_\alpha\sqrt{h(1-h)n + \frac{1}{4}z_\alpha^{\ 2}}}{n + z_\alpha^{\ 2}} \\ p_2 &= \frac{m + \frac{1}{2}z_\alpha^{\ 2} + z_\alpha\sqrt{h(1-h)n + \frac{1}{4}z_\alpha^{\ 2}}}{n + z_\alpha^{\ 2}} \end{aligned}\right\} \qquad (3.21)$$

Если число опытов значительно велико $\left(\frac{z_\alpha^{\ 2}}{n}\right.$ стремится к нулю, обе величины *np* и *n(1-p)* больше 10 $\left.\right)$, то формулы (3.21) упрощаются и доверительный интервал можно находить в виде

$$h - z_\alpha\sqrt{\frac{h(1-h)}{n}} < p < h + z_\alpha\sqrt{\frac{h(1-h)}{n}} \qquad (3.22)$$

Пример. Частота успеха в серии из *n*=100 опытов оказалась *h*=0,78. Определим доверительный интервал для *p* с надёжностью $1-\alpha = \gamma = 0{,}9$, при этом $z_\alpha = 1{,}64$ и по формулам (3.21) получаем

$p_1 = 0{,}705$, $p_2 = 0{,}840$, т.е.

0,705<*p*<0,840.

Полагая ориентировочно $np \approx nh = 78$ и $n(1-p) \approx n(1-h) = 22$, применим формулу (3.22) и получим

$0{,}712 < p < 0{,}848$

3.5. Задачи

3.1. Построить доверительные интервалы для μ и σ^2 по данным задачи 1.1 для разных значений γ: $0{,}9$; $0{,}95$; $0{,}99$.

3.2. Построить доверительные интервалы для μ и σ^2 по данным задач 2.2 и 2.3. В задаче 2.2. использовать и нормальную асимптотику, $\gamma = 0{,}95$.

3.3. Построить доверительные интервалы для вероятности успеха p в одном опыте:

а) $n = 60; m = 15; \gamma = 0{,}95;$

б) $n = 200; m = 70; \gamma = 0{,}9;$

§4. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Для характеристики многомерного эмпирического распределения вычисляют его числовые параметры, позволяющие делать выводы о существовании зависимостей между компонентами.

4.1. Коэффициент корреляции

Мерой силы (тесноты) и направления линейной связи между двумя переменными x и y, вычисленной по ряду из n пар (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), является эмпирический коэффициент корреляции:

$$r_{xy} = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\left(\sum\limits_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum\limits_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}} \qquad (4.1)$$

При пользовании эмпирическим коэффициентом корреляции подразумевается, что выборка $(x_i, y_i),\ i = 1, 2, ..., n$, получена из двумерной нормальной генеральной совокупности.

Если рассматривать одновременно три и более переменные, имеющие совместные нормальные распределения, то по формулам, аналогичным выше приведенной, можно вычислить коэффициенты корреляции между любыми двумя переменными. При большом количестве переменных их удобно обозначать числами 1, 2, 3, … . В этом случае данные удобно представить в виде корреляционной матрицы. При этом $r_{ii} = 1, r_{ij} = r_{ji}$, поэтому заполняется обычно только верхняя половина матрицы. Например, при четырёх переменных матрица будет иметь вид $\begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} \\ & 1 & r_{23} & r_{24} \\ & & 1 & r_{34} \\ & & & 1 \end{pmatrix}$.

4.2. Коэффициент ранговой корреляции Спирмена

Для определения взаимозависимости между переменными, имеющими произвольное двумерное распределение, можно пользоваться коэффициентом ранговой корреляции Спирмена. Ряды измерений x_i и y_i, $i = 1, 2, ..., n$ преобразуются с помощью рангов следующим образом. Каждому значению x_i ставится в соответствие ранг R_i, т.е. номер элемента x_i в вариационном ряду; аналогичным образом определяются ранги R_i' элементов y_i. Таким образом, каждой паре (x_i, y_i) соответствует пара рангов (R_i, R_i'). Ранговый коэффициент корреляции Спирмена вычисляется по формуле:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} (R_i - R_i')^2}{n \cdot (n^2 - 1)}. \qquad (4.2)$$

Пример 1. Бегуны, ранги которых при построении по росту были 1, 2, 3, …, 10, заняли на состязаниях следующие места: 6, 5, 1, 4, 2, 7, 8, 10, 3, 9. Как велика ранговая корреляция между ростом и быстротой бега?

Решение. Обозначим R_i – ранги роста, R_i' – ранги быстроты бега участников. И составим таблицу.

R_i	1	2	3	4	5	6	7	8	9	10
R_i'	6	5	1	4	2	7	8	10	3	9
$R_i - R_i'$	–5	–3	2	0	3	–1	–1	–2	6	1

По формуле $r_s = 1 - \frac{6 \cdot \sum_{i=1}^{10} (R_i - R_i')^2}{10 \cdot (10^2 - 1)}$ вычислим ранговый коэффициент корреляции $r_s = 1 - \frac{6 \cdot (25 + 9 + 4 + 9 + 1 + 1 + 4 + 36 + 1)}{10 \cdot (100 - 1)} \approx 0,455$.

4.3. Доверительный интервал для коэффициента корреляции r_0

Если выборка получена из генеральной совокупности, имеющей двумерное нормальное распределение, то при достаточно больших n $(n > 25)$ статистика

$$w = \frac{1}{2}\ln\frac{1+r}{1-r} \qquad (4.3)$$

имеет приближенно нормальное распределение с параметрами

$$M[w] = \frac{1}{2}\ln\frac{1+r_0}{1-r_0}, \quad D[w] = \frac{1}{n-3}.$$

Здесь r – эмпирический коэффициент корреляции, n – объем выборки. Таким образом, случайная величина

$$z = \frac{w - M[w]}{\sqrt{D[w]}} \qquad (4.4)$$

приближенно удовлетворяет нормированному нормальному распределению $N(0;1)$. Для различных значений $\alpha = 1-\gamma$ определяем z_α (см. таблицу 2, стр. 22) и получаем приближенный доверительный интервал для коэффициента корреляции r_0 в виде (4.5)

$$\operatorname{th}\left(\frac{1}{2}\ln\frac{1+r}{1-r} - z_\alpha\frac{1}{\sqrt{n-3}}\right) < r_0 < \operatorname{th}\left(\frac{1}{2}\ln\frac{1+r}{1-r} + z_\alpha\frac{1}{\sqrt{n-3}}\right). \quad (4.5)$$

(это следует из того, что функция $y = \frac{1}{2}\ln\frac{1+x}{1-x}$ монотонно возрастающая, а обратная функция определяется формулой $x = \frac{e^{2y}-1}{e^{2y}+1} = \operatorname{th} y$ – гиперболический тангенс).

Пример 2. Выборочный коэффициент корреляции, вычисленный по выборке объема $n = 50$, равен $r = 0,687$; $\gamma = 0,95$. Найти доверительный интервал для коэффициента корреляции.

Решение. $\alpha = 1-\gamma = 0,05$; $z_\alpha = 1,96$; $\dfrac{1}{\sqrt{n-3}} = \dfrac{1}{\sqrt{47}} = 0,146$;

$$\operatorname{th}\left(\frac{1}{2}\ln\frac{1+0,687}{1-0,687} - 1,96\cdot 0,146\right) < r_0 < \operatorname{th}\left(\frac{1}{2}\ln\frac{1+0,687}{1-0,687} + 1,96\cdot 0,146\right)$$

Отсюда получаем

$$0,505 < r_0 < 0,810$$

Обычно доверительный интервал для r_0 вычисляется с помощью таблиц гиперболического тангенса (*Таблица V* Приложения) ($w = \operatorname{Arth} z$) из соотношения

$$\operatorname{Arth} r - z_\alpha \frac{1}{\sqrt{n-3}} < \operatorname{Arth} r_0 < \operatorname{Arth} r + z_\alpha \frac{1}{\sqrt{n-3}} \qquad (4.6)$$

В нашем примере получаем

$$0,842 - 1,96\cdot 0,146 < \operatorname{Arth} r_0 < 0,842 + 1,96\cdot 0,146$$

$$0,556 < \operatorname{Arth} r_0 < 1,128$$

$$0,505 < r_0 < 0,810$$

4.4. Задачи

4.1. Вычислить выборочный коэффициент корреляции для выборки (x_i, y_i) $i = 1, 2, \ldots, 5$: (8;1), (10;3), (5;1), (8;2), (9;3).

4.2. По двум предметам выставлены рейтинги знаний 10-ти студентов. Найти выборочный коэффициент ранговой корреляции Спирмена между знаниями по этим предметам.

82	78	63	45	76	86	68	92	75	49
78	84	45	56	80	91	72	79	65	57

4.3. Мастерство 8-ми спортсменов оценивается двумя арбитрами; получены две последовательности рангов:

арбитр А	1	2	3	4	5	6	7	8
арбитр В	3	5	7	2	8	1	6	4

Оценить, как согласуются оценки арбитров.

4.4. По двумерной выборке объема $n = 40$ вычислен коэффициент корреляции $r = 0,78$. Найти доверительные интервалы для коэффициента корреляции для различных α: $0,1;\ 0,05;\ 0,01$.

§5. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Статистической гипотезой H называется предположение относительно параметров или вида распределения случайной величины. Проверяемая гипотеза называется нулевой гипотезой и обозначается H_0. Наряду с гипотезой H_0 рассматривают одну из альтернативных (конкурирующих) гипотез H_1. Правило, согласно которому принимается или отклоняется гипотеза H_0, называется критерием. Так как решение принимается на основе выборки, строится подходящая статистика t (функция элементов выборки), называемая статистикой критерия.

Проверка статистических гипотез основывается на принципе, в соответствии с которым принимаются события, имеющие большую вероятность, и отвергаются – имеющие малую вероятность. Для осуществления этого принципа перед проверкой фиксируется некоторая малая вероятность α, называемая уровнем значимости критерия.

Множество всех значений статистики критерия t, при которых принимается решение отклонить гипотезу H_0, называется критической областью. Множество значений t, при которых гипотеза H_0 принимается, называется областью допустимых значений. Уровень значимости α определяет «размер» критической области. Если цель проверки состоит в том, чтобы установить различие двух генеральных совокупностей, и знак предполагаемого различия не существенен, то говорят о двусторонней альтернативной гипотезе H_1 (двусторонний критерий), а если основная гипотеза позволяет высказать определенное предположение о знаке ожидаемого различия, то критерий называется односторонним.

Таким образом, проверка статистических гипотез может быть разбита на следующие основные этапы:

1) формулировка проверяемой (H_0) и альтернативной (H_1) гипотез;

2) назначение уровня значимости α (стандартные значения α: 0,1; 0,05; 0,01);

3) выбор статистики критерия t (тип критерия);

4) определение критической области (критические значения статистики, найденные по соответствующим таблицам);

5) принятие решения в зависимости от того, в какую область попадает вычисленное по выборке значение статистики $\hat{t}$ (сравнение табличных и вычисленных значений статистики критерия).

Проверка статистических гипотез с использованием критериев может быть проведена на основе **доверительных интервалов**. При этом двустороннему критерию, когда альтернативная гипотеза имеет вид:

$H_1: \mu_1 \neq \mu_0; H_1: \mu_1 \neq \mu_2;\ H_1: \sigma_1^{\ 2} \neq \sigma_2^{\ 2}$ и т.п., соответствует двусторонний доверительный интервал, а если критерий односторонний, т.е. альтернативная гипотеза вида $H_1: \mu > \mu_0; H_1: \mu < \mu_0;\ H_1: \mu_1 > \mu_2$ и т.п., то доверительный интервал будет односторонним.

5.1. Проверка гипотезы о среднем значении μ нормально распределенной совокупности

$H_0: \ \mu = \mu_0,$

$H_1: \ \mu \neq \mu_0,$

либо $H_1: \ \mu > \mu_0,$

либо $H_1: \ \mu < \mu_0.$

Статистика критерия t имеет распределение Стьюдента с $\nu = (n-1)$ (ν – число степеней свободы):

$$t = \frac{\left(\overline{x} - \mu_0\right) \cdot \sqrt{n}}{s}. \qquad (5.1)$$

Если вычисленное значение статистики попадает в критическую область, то гипотеза H_0 отвергается. Это происходит по следующей схеме:

а) Если $H_1: \ \mu \neq \mu_0$, то H_0 отклоняется при условии, что $\left|\hat{t}\right| > t_{êð}$ (двусторонний критерий), $t_{êð}$ находим по таблице критерия Стьюдента для заданного α и $\nu = n-1$.

Если используется доверительный интервал для математического ожидания, то гипотеза H_0 отклоняется, если μ_0 не попадает в интервал

$$\overline{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}};$$

б) Если $H_1: \ \mu > \mu_0$, то H_0 отклоняется при условии $\hat{t} > t_{êð}$ (правосторонний критерий);

в) Если $H_1: \mu < \mu_0$, то H_0 отклоняется при условии $\hat{t} < -t_{êð}$ (левосторонний критерий).

Замечание. Если дисперсия σ^2 генеральной совокупности известна, то статистика критерия

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \qquad (5.2)$$

имеет стандартное нормальное распределение и для нахождения критических значений можно использовать таблицу функции Лапласа $\hat{O}(\mathrm{x})$:

для двустороннего критерия $\hat{O}(\mathrm{t}_{êð}) = \dfrac{(1-\alpha)}{2}$;

для одностороннего критерия $\hat{O}(\mathrm{t}_{êð}) = \dfrac{(1-2\alpha)}{2}$.

Пример 1. Из нормальной генеральной совокупности с известной дисперсией $\sigma^2 = 25$ извлечена выборка объёма n=36 и по ней найдено $\bar{x} = 17$. При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу $H_0: \mu_0 = 20$ при

1) $H_1 : \mu_0 \neq 20$; 2) $H_1 : \mu_0 < 20$.

Решение.

1) Вычисляем значение статистики

$$|\hat{t}| = \frac{|17-20| \cdot 6}{5} = 3{,}6$$

Так как критерий двусторонний, то $t_{êð}$ находим из соотношения $\hat{O}(\mathrm{t}_{êð}) = \dfrac{(1-\alpha)}{2}$ =0,475, из таблицы функции Лапласа следует, что $t_{êð}$ =1,96 и, следовательно,

$$|\hat{t}| = 3{,}6 > t_{êð} = 1{,}96.$$

Гипотеза $H_0: \mu_0 = 20$ отклоняется.

2) Критерий левосторонний, поэтому $t_{êð}$ находим из соотношения $\hat{O}(\mathrm{t}_{êð}) = \dfrac{(1-2\alpha)}{2}$ =0,45.

$t_{êð}$ =1,64 ; значение статистики $\hat{t}$ =-3,6 и так как

$\hat{t}$ =-3,6<-$t_{êð}$ =-1,64, то

Гипотеза H_0 отклоняется в пользу гипотезы H_1.

5.2 Проверка гипотезы о равенстве средних двух независимых выборок из нормально распределенных совокупностей

$H_0: \ \mu_1 = \mu_2,$

$H_1: \ \mu_1 \neq \mu_2,$

$H_1: \ \mu_1 > \mu_2,$

$H_1: \ \mu_1 < \mu_2.$

Статистика критерия t имеет распределение Стьюдента с $\nu = n_1 + n_2 - 2$, где n_1 и n_2 – объемы выборок:

$$t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \qquad (5.3)$$

где

$$s^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}. \qquad (5.4)$$

Схема принятия решения аналогична рассмотренной в предыдущем пункте.

Пример 2. Проверка гипотезы о равенстве средних двух нормально распределенных совокупностей с помощью критерия Стьюдента.

Изучая физические свойства нефтяного пласта, исследователь получил следующие значения пористости пласта в зонах расположения двух достаточно удаленных скважин на залежи:

Первая скважина	5,9	8,2	6,0	7,7	6,5	5,2			$n_1 = 6$
Вторая скважина	7,5	7,8	8,5	10,3	8,1	7,6	8,8	9,9	$n_2 = 8$

Можно ли считать, что коллектор однороден, т.е. его пористость в среднем меняется незначительно?

Решение.

$H_0: \ \ \mu_1 = \mu_2,$

$H_1: \ \ \mu_1 \neq \mu_2,$

$\alpha = 0,05.$

Критическое значение $t_{êð}$ находим по таблицам двустороннего критерия Стьюдента (*Таблица IV* Приложения) для α = 0.05 и числа степеней свободы $\nu = n_1 + n_2 - 2 = 6 + 8 - 2 = 12$. Это значение $t_{êð} = 2,179$. Вычислим значение статистики t:

$$\bar{x} = 6,58; \ \ \bar{y} = 8,56; \ \ s_1^{\ 2} = \frac{6,59}{5}; \ \ s_2^{\ 2} = \frac{7,72}{7}; \ \ s^2 = \frac{(n_1 - 1) \cdot s_1^{\ 2} + (n_2 - 1) \cdot s_2^{\ 2}}{n_1 + n_2 - 2} = 1,19;$$

$$\hat{t} = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = 3,36.$$

Так как $|\hat{t}| > t_{êð}$, то гипотезу H_0 можно отвергнуть, т.е. с надежностью $\gamma = 1 - \alpha = 0,95$ можно утверждать, что различие в пористости между двумя скважинами не является случайным.

5.3 Проверка гипотезы о значении дисперсии нормально распределённой совокупности

$H_0 : \sigma^2 = \sigma_0{}^2$
$H_1 : \sigma^2 \neq \sigma_0{}^2$
$H_1 : \sigma^2 > \sigma_0{}^2$
$H_1 : \sigma^2 < \sigma_0{}^2$

Статистика критерия

$$t = \frac{(n-1)s^2}{\sigma_0{}^2} \qquad (5.5)$$

имеет распределение Пирсона χ^2 с $\nu = n-1$ степенями свободы.

а) $H_1 : \sigma^2 \neq \sigma_0{}^2$

По таблицам распределения χ^2 для данных α и $\nu = n-1$ находят два критических значения $\underline{x^2}$ и $\overline{x^2}$ (аналогично построению доверительного интервала). Если таблица распределения дает значения вероятностей $P\{\chi^2 > x^2\}$, то критические значения находим из соотношений

$$P\{\chi_{n-1}{}^2 > \underline{x^2}\} = 1 - \frac{\alpha}{2}\ ;\ P\{\chi_{n-1}{}^2 > \overline{x^2}\} = \frac{\alpha}{2} \qquad (5.6)$$

Если вычисленное значение статистики $\hat{t}$ удовлетворяет условию

$$\underline{x^2} < \hat{t} < \overline{x^2}, \text{ то}$$

нет оснований отклонять гипотезу H_0.

б) $H_1 : \sigma^2 > \sigma_0{}^2$

Гипотеза H_0 отклоняется, если

$$\hat{t} > \overline{x^2}, \text{ где } \overline{x^2} \text{ находится}$$

из соотношения $P\{\chi_{n-1}{}^2 > \overline{x^2}\} = \alpha$ (правосторонний критерий).

в) $H_1 : \sigma^2 < \sigma_0{}^2$

Гипотеза H_0 отклоняется, если

$$\hat{t} < \underline{x^2}, \text{ где } \underline{x^2} \text{ находится}$$

из соотношения $P\{\chi_{n-1}{}^2 > \underline{x^2}\} = 1 - \alpha$ (левосторонний критерий).

Пример 3. Из нормально распределённой совокупности извлечена выборка n=21, для которой вычислено значение $s^2 = 10{,}3$. Требуется проверить гипотезу $H_0 : \sigma_0{}^2 = 9$ при альтернативной гипотезе $H_1 : \sigma_0{}^2 > 9$. $\alpha = 0{,}05$.

Решение.

Вычисляем значение статистики

$$\hat{t} = \frac{20 \cdot 10{,}3}{9} \approx 22{,}89$$

Критерий правосторонний, поэтому находим $t_{\hat{e}\eth} = \overline{\mathrm{x}^2}$ из соотношения

$P\left\{\chi_{20}{}^2 > \overline{x^2}\right\} = \alpha = 0{,}05;\ \ t_{\hat{e}\eth} = \overline{x^2} = 31{,}4$

Так как $\hat{t} = 22{,}89 < t_{\hat{e}\eth} = 31{,}4$, то нет оснований отклонять гипотезу H_0.

5.4. Сравнение двух выборочных дисперсий из нормально распределенных совокупностей

$$H_0 :\ \ \sigma_1^2 = \sigma_2^2,$$
$$H_1 :\ \ \sigma_1^2 > \sigma_2^2.$$

Статистика критерия

$$t = \frac{s_1^2}{s_2^2} \tag{5.7}$$

имеет распределение Фишера с $\nu_1 = n_1 - 1$ и $\nu_2 = n_2 - 1$(*Таблица VI* Приложения).

Гипотеза H_0 отвергается, если вычисленное значение статистики t больше $t_{\hat{e}\eth}$, найденного по таблицам критерия Фишера для данных α, ν_1 и ν_2.

ЗАМЕЧАНИЕ. В числителе статистики всегда стоит большая выборочная дисперсия.

Пример 4. Проверка гипотезы о равенстве двух дисперсий

По двум независимым выборкам, объемы которых $n_1 = 11$ и $n_2 = 14$, найдены выборочные дисперсии:

$s_1^2 = 0,76; \quad s_2^2 = 0,38;$

$H_0: \ \sigma_1^2 = \sigma_2^2,$

$H_1: \ \sigma_1^2 > \sigma_2^2,$

$\alpha = 0,05.$

Решение.

Критическое значение статистики находим по таблицам критерия Фишера для $\alpha = 0,05$ (*Таблица VI* Приложения) и чисел степеней свободы $\nu_1 = 10$ и $\nu_2 = 13$: $t_{\hat{e}\eth} = 2,67$. Вычисленное значение статистики:

$$\hat{t} = \frac{s_1^{\ 2}}{s_2^{\ 2}} = \frac{0,76}{0,38} = 2.$$

Так как $|t| < t_{\hat{e}\eth}$, делаем вывод, что данные выборки гипотезе H_0 не противоречат, т.е. с надежностью $\gamma = 1 - \alpha = 0,95$ можно утверждать, что дисперсии различаются незначимо.

5.5. Проверка на значимость коэффициентов корреляции r и r_s

Проверка на значимость вычисленных выборочных коэффициентов корреляции представляет собой проверку следующей гипотезы: существенно ли (значимо ли) отличается от нуля рассчитанный по ряду измерений объема n эмпирический коэффициент корреляции?

Проверка на значимость коэффициента корреляции.

$H_0: \ r_{xy} = 0;$

$H_1: \quad r_{xy} \neq 0.$

Статистика критерия

$$t = \frac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1-\left(r_{xy}\right)^2}} \qquad (5.8)$$

имеет распределение Стьюдента с $\nu = n-2$. Гипотеза H_0 отвергается, если вычисленное значение статистики $|\hat{t}| > t_{êð}$, найденного по таблице для данных α и ν. Если гипотеза H_0 принимается, то это позволяет сделать вывод о независимости случайных величин x и y. (Этот вывод делается только в предположении, что выборка взята из двумерной нормальной совокупности). В общем случае говорят о некоррелированности величин.

Проверка на значимость рангового коэффициента корреляции.

$H_0: r_s = 0.$

$H_1: r_s \neq 0$

При $n \geq 10$ проверка осуществляется с помощью статистики

$$t = r_s \cdot \sqrt{\frac{n-2}{1-r_s^2}}, \qquad (5.9)$$

имеющей распределение Стьюдента с $\nu = n-2$ степенями свободы.

При $n \geq 30$ можно использовать статистику

$$t = r_s \cdot \sqrt{n-1}, \qquad (5.10)$$

имеющую приближенно стандартное нормальное распределение $N(0;1)$.

Если вычисленное значение $|\hat{t}|$ больше $t_{êð}$, гипотеза H_0 отвергается.

Стандартные значения для нормального распределения

α	0,1	0,05	0,01
z_α	1,64	1,96	2,576

Пример 5. Проверка на значимость коэффициента корреляции.

Проверить на значимость коэффициент корреляции между забойным и пластовым давлением фонтанирующих скважин по выборке (x_i, y_i), $i = 1, 2, ..., n$, $n = 30$.

Решение.

Вычисленное значение коэффициента корреляции равно $r_{xy} = 0,913$. Проверим его на значимость при $\alpha = 0,01$.

$H_0: \;\; r_{xy} = 0;$

$H_1: \;\; r_{xy} > 0.$

(односторонний критерий)

Значение статистики $t = \dfrac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1-\left(r_{xy}\right)^2}} = 12{,}1.$

Для $\alpha = 0,01$ и $\nu = n - 2 = 28$ находим по таблице $t_{\hat{e}\eth} = 2,457$ (*Таблица IV* Приложения).

Так как $\hat{t} > t_{\hat{e}\eth}$ (рассматривается односторонний критерий), гипотеза H_0 отвергается. Следовательно, забойное и пластовое давления зависимы (предполагается нормальное распределение).

Пример 6. Проверка на значимость рангового коэффициента корреляции Спирмена.

На 30 нефтеперерабатывающих заводах были собраны данные выработки на одного рабочего в отчетном году в процентах к предыдущему году (x) и выпуска валовой продукции в том же году в процентах к предыдущему году (y). Определим их корреляцию с помощью рангового коэффициента Спирмена, поскольку нет предположений о законе распределения. Вычисленное значение коэффициента $r_S = -0{,}058$.

Проверим этот коэффициент на значимость:

$H_0: \; r_S = 0;$

$H_1: \; r_S \neq 0.$

$\alpha = 0{,}01$

Так как $n = 30$, можно считать, что статистика $t = r_S \cdot \sqrt{n-1}$ имеет приближенно нормальное распределение. Для $\alpha = 0{,}01$ и двустороннего критерия $z_\alpha = t_{êð} = 2{,}576$, а вычисленное значение статистики $|\hat{t}| = 0{,}058 \cdot \sqrt{n-1} = 0{,}058 \cdot \sqrt{29} = 0{,}312$. И так как $|\hat{t}| < t_{êð}$, нет основания отвергать гипотезу H_0, т.е. можно считать, что рассматриваемые величины не коррелированны.

5.6. Критерий согласия Пирсона $\left(\chi^2\right)$ для проверки гипотезы о законе распределения

H_0: выборка принадлежит распределению $F_0(x)$;

H_1: выборка не принадлежит распределению $F_0(x)$.

Проверка гипотезы H_0 осуществляется с помощью критерия Пирсона (χ^2) по следующей схеме.

Если данные представлены интервальной группировкой (K интервалов), то подсчитываются количества m_i выборочных значений, попавших в i-ый интервал ($i = 1,\dots,K$) и вычисляются по теоретическому распределению $F_0(x)$ вероятности p_i попадания в i-ый интервал ($i = 1,\dots,K$).

Для дискретной группировки данных, состоящей из отдельных K значений, подсчитываются абсолютные частоты этих значений m_i $(i=1,...,K)$ и вычисляются теоретические вероятности p_i этих значений. И в том и в другом случаях $\sum_{i=1}^{K} m_i = n$.

Статистика критерия

$$t=\sum_{i=1}^{k}\frac{(m_i-np_i)^2}{np_i}. \qquad (5.11)$$

Эта статистика имеет распределение Пирсона $\left(\chi^2\right)$ с $\nu=k-1$ степенями свободы. Если для распределения $F_0(x)$ предварительно по выборке вычислялись L параметров, то $\nu=k-1-L$. Этот критерий требует, чтобы были выполнены следующие соотношения для всех i: $np_i \geq 5$. Если в некоторых интервалах это требование нарушено, то соседние интервалы надо объединять в один. Гипотеза H_0 отвергается, если вычисленное значение $\hat{t}$ больше $t_{\hat{e}\eth}$, найденного по таблицам для данных α и ν.

Пример 7. Проверить гипотезу о нормальном распределении выборки объема $n=50$ при $\alpha=0,05$.

$$H_0: F(x)=F_0(x)=\frac{1}{\sqrt{2\pi}s}\int_{-\infty}^{x} e^{-\frac{(t-40,6)^2}{2s^2}}dt \quad (\xi \in N(40,6;3))$$

По данным выборки вычислены $\overline{x}=40,6$; $s^2=3$; $s=1,73$; размах выборки разбит на K=6 интервалов и подсчитано количества m_i выборочных значений, попавших в каждый интервал (i=1,2,…,6); вычислены теоретические вероятности p_i того, что значение случайной величины принадлежит i-ому интервалу, по формуле

$$p_i = P\left\{a_i < \xi < b_i\right\} = \hat{O}\left(\frac{\mathrm{b_i} - \bar{x}}{s}\right) - \hat{O}\left(\frac{\mathrm{a_i} - \bar{x}}{s}\right), i = 1,2,...,6. \qquad (5.12)$$

(здесь a_i и b_i – границы i-ого интервала, а значения $\Phi(x)$ - значения функции Лапласа). Для удобства вычислений все данные сведены в таблицу.

i	m_i	p_i	np_i	$m_i - np_i$	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
1	6	0,1	5	1	1	0,20
2	5	0,14	7	–2	4	0,57
3	10	0,22	11	–1	1	0,09
4	13	0,22	11	2	4	0,36
5	9	0,16	8	1	1	0,12
6	7	0,14	7	0	0	0,00
	$\sum = 50$					$\sum = 1{,}34$

Следовательно, значение статистики критерия χ^2 (формула (5.11))

$$\hat{t} = 1{,}34\,.$$

Число степеней свободы $\nu = 6 - 1 - 2 = 3$. По *Таблице IV* Приложения находим $t_{\hat{e}\eth} = 7{,}81$ и т.к. $\hat{t} < t_{\hat{e}\eth}$, то данные выборки не противоречат гипотезе о нормальном распределении.

Пример 8. Проверить гипотезу о биномиальном распределении, т.е. что при проведении n независимых опытов вероятность успеха в каждом опыте одна и та же и равна $p_0 = 0{,}2$. Иными словами,

$$H_0 : \; P_n\{\xi = l\} = C_n^l 0{,}2^l \cdot 0{,}8^{n-l} \qquad (5.13)$$

$\alpha = 0{,}05$ (ξ - число успехов в n опытах).

Для проверки этой гипотезы было проведено $n = 50$ серий опытов по $k = 4$ опытов в каждой серии и получены следующие результаты:

i	0	1	2	3	4
m_i	20	18	10	1	1

, где i - количество успехов в серии из $k = 4$ опытов, а m_i - количество серий, в которых этот результат был получен. ($n = \sum_{i=0}^{4} m_i = 50$).

По формуле (5.13) вычислены теоретические вероятности и величины np_i

$P_4\{\xi = 0\} = C_4^0 \cdot 0{,}2^0 \cdot 0{,}8^4 = 0{,}4096 = p_0; np_0 = 20{,}48$

$P_4\{\xi = 1\} = C_4^1 \cdot 0{,}2^1 \cdot 0{,}8^3 = 0{,}4096 = p_1; np_1 = 20{,}48$

$P_4\{\xi = 2\} = C_4^2 \cdot 0{,}2^2 \cdot 0{,}8^2 = 0{,}1536 = p_2; np_2 = 7{,}68$

$P_4\{\xi = 3\} = C_4^3 \cdot 0{,}2^3 \cdot 0{,}8^1 = 0{,}0256 = p_3; np_3 = 1{,}28$

$P_4\{\xi = 4\} = C_4^4 \cdot 0{,}2^4 \cdot 0{,}8^0 = 0{,}0016 = p_4; np_4 = 0{,}08$

Так как np_3 и np_4 очень малы (по условию $np_i \geq 5$), то их объединяем с np_2 и составляем таблицу

i	m_i	np_i	$m_i - np_i$	$\frac{(m_i - np_i)^2}{np_i}$
0	20	20,48	-0,48	0,01125
1	18	20,48	-2,48	0,30031
≥ 2	12	9,04	2,96	0,9692

$$t = \sum_{i=0}^{2} \frac{(m_i - np_i)^2}{np_i}$$

Вычисленное значение $\hat{t} = 1{,}28 (k = 3)$.

По таблице для распределения χ^2 находим

$t_{êð} = \chi^2_{0{,}05;3\text{-}1} = 6{,}0$.

И так как $\hat{t} = 1{,}28 < t_{êð} = 6{,}0$, то нет оснований отклонять гипотезу H_0 о биномиальном распределении.

5.7. Задачи

5.1. По выборке $n = 16$ из нормальной генеральной совокупности найдены $\bar{x} = 12{,}4$ и среднее квадратическое отклонение s =1,2. Проверить нулевую гипотезу H_0: $\mu = 11{,}8$ при конкурирующих гипотезах H_1 : $\mu \neq 11{,}8$ и H_1 : $\mu > 11{,}8$ при разных уровнях значимости: $\alpha = 0{,}1; 0{,}05; 0{,}01$.

5.2. Используя данные задачи 5.1 проверить гипотезу H_0: $\sigma^2 = 1$, H_1 : $\sigma^2 \neq 1, \alpha = 0{,}05$.

5.3. Из нормальной генеральной совокупности с известной дисперсией $\sigma^2 = 1{,}44$ извлечена выборка объёма n=49 и по ней найдено $\bar{x} = 3{,}8$. Требуется при уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу H_0: $\mu = 3$ при конкурирующей гипотезе H_1 : $\mu > 3$.

5.4. По двум выборкам n_1 и n_2, извлечённым из нормальных генеральных совокупностей, найдены $\bar{x}, \bar{y}, s_1^2$ и s_2^2. Проверить нулевую гипотезу H_0 : $\mu_1 = \mu_2$ при конкурирующей гипотезе H_1 : $\mu_1 \neq \mu_2$, если $n_1 = 30; n_2 = 20; \bar{x} = 10{,}2; \bar{y} = 12{,}5; s_1^2 = 12; s_2^2 = 10; \alpha = 0{,}05.$

5.5. По двум независимым выборкам объемов $n = 5$ и $m = 6$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние $\bar{x} = 15{,}9$ и $\bar{y} = 14{,}1$, и выборочные дисперсии $s_1^2 = 14{,}76$ и $s_2^2 = 4{,}92$. При уровне значимости $\alpha = 0{,}05$ проверить следующие гипотезы:

а) $\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2, \\ H_1 : \sigma_1^2 > \sigma_2^2. \end{array}$ б) $\begin{array}{l} H_0 : \mu_1 = \mu_2, \\ H_1 : \mu_1 \neq \mu_2. \end{array}$ в) $\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{array}.$

5.6. Проверить на значимость коэффициенты корреляции, полученные в задачах 1,2 и 3 §4, $\alpha = 0{,}1; \alpha = 0{,}05$.

5.7. В таблице даны результаты измерений чувствительности приемника (в μкв). Проверить согласие результатов измерения с нормальным законом распределения при $\alpha = 0{,}01$.

граница интервалов	200 – 250	250 – 300	300 – 350	350 – 400	400 – 450	450 – 500	500 – 550	550 – 600
m_i	1	2	11	20	19	13	8	3

5.8. Испытания 200 ламп на продолжительность работы (в часах) дали следующие результаты:

интервалы	0–300	300 – 600	600 – 900	900 – 1200	1200 – 1500	1500 – 1800
m_i	58	41	30	22	16	12
интервалы	1880 – 2100	2100 – 2400	2400 – 2700	2700 –3000	3000 – 3300	св. 3300
m_i	9	7	5	3	2	0

Проверить гипотезу о согласии этих данных с экспоненциальным законом распределения, $\alpha = 0{,}05$.

5.9. В течение 100 дней для проверки качества продукции отбирались случайным образом партии из k=10 изделий и фиксировалось количество нестандартных изделий в каждой партии. В результате проверки получены следующие данные

i	0	1	2	3	4	5
m_i	55	28	10	4	2	1

$$\left(\sum_{i=0}^{5} m_i = 100\right)$$

i - количество нестандартных изделий в партии, m_i - количество дней, когда фиксировалось i нестандартных изделий.

Проверить гипотезу о том, что данное предприятие выпускает 90% стандартных изделий ($p = 0{,}9$), при $\alpha = 0{,}05$.

5.10. При испытании радиоэлектронной аппаратуры фиксировалось число отказов за определённое время. В результате 50 испытаний были получены следующие данные

Число отказов	0	1	2	3
Число испытаний	38	7	3	2

Проверить гипотезу H_0 о том, что число отказов имеет распределение Пуассона при $\alpha = 0{,}05$ (предварительно оценить параметр распределения).

§6. РЕГРЕССИОННЫЙ АНАЛИЗ

6.1. Модели регрессионного анализа

Во многих ситуациях интерес представляет зависимость между двумя признаками продукта, материала, процесса и т.п.

При одновременном изучении двух признаков получают двумерную выборку (x_i, y_i), $i = 1, ..., n$. Пары точек (x_i, y_i) наносят на координатную сетку, получая так называемое **облако точек**, которое дает предварительное представление о рассеянии и форме зависимости между признаками.

На основании регрессионного анализа наблюдаемое облако аппроксимируется уравнением регрессии. Теоретически уравнение регрессии является условным математическим ожиданием одной случайной переменной (зависимой), при условии, что вторая переменная (независимая) принимает заданные (фиксированные) значения. Таким образом, регрессия – это всегда «зависимость в среднем». Если эта зависимость задается уравнением прямой

$$y = a + bx, \qquad (6.1)$$

то говорят о линейной регрессии или **линейной регрессионной модели.** Параметры a и b оцениваются в большинстве случаев на основе метода наименьших квадратов (МНК).

Если обе переменные принадлежат двумерному нормальному распределению, то уравнение регрессии может быть записано в виде уравнения прямой линии

$$y = \mu_y + r\frac{\sigma_y}{\sigma_x}(x - \mu_x), \qquad (6.2)$$

Либо

$$x = \mu_x + r\frac{\sigma_x}{\sigma_y}(y - \mu_y). \qquad (6.3)$$

Оценки этих линейных уравнений имеют вид:

$$y = \overline{y} + \hat{r}\frac{s_y}{s_x}(x - \overline{x}) \qquad (6.4)$$

$$x = \overline{x} + \hat{r}\frac{s_x}{s_y}(y - \overline{y}) \qquad (6.5)$$

Прямые пересекаются в точке $(\overline{x}, \overline{y})$ и образуют «ножницы», причем тем уже, чем больше $|r|$. При $|r| = 1$ обе прямые регрессии совпадают, а при $|r| = 0$ прямые регрессии перпендикулярны осям координат (переменные независимы).

Во многих случаях графическое представление данных, т.е. облако точек, показывает, что зависимость между переменными не может быть описана прямой линией. В этом случае можно попробовать представить искомое уравнение в виде

$$y = a + bx + cx^2,$$

т.е. взять уравнение второго порядка, либо в виде

$$y = a + bx + c\sqrt{x},$$

либо другим способом.

Коэффициенты a, b и c находятся методом наименьших квадратов. При пользовании МНК важно, чтобы искомые коэффициенты входили в уравнение линейно, поэтому при более сложных зависимостях прибегают к преобразованиям одной или обеих переменных, чтобы получить зависимость линейного вида.

В заключение рассмотрим несколько примеров линеаризации зависимостей между x и y.

1º. $y = a \cdot b^x$.

Прологарифмируем это выражение и введем новые переменные

$$\ln y = \ln a + x \cdot \ln b$$

$$\ln y = z, \quad \ln a = \alpha, \quad \ln b = \beta.$$

В результате получаем линейную зависимость:

$$z = \alpha + \beta \cdot x.$$

2º. $y = a + \dfrac{b}{x}$.

Введем новую переменную

$$z = \frac{1}{x}.$$

В результате получаем зависимость вида:

$$y = a + b \cdot z.$$

3º. $y = \dfrac{a}{b + x}$.

В этом случае сделаем следующее преобразование:

$$\frac{1}{y} = \frac{b + x}{a}$$

и соответствующие замены переменных

$$\frac{1}{y} = z, \quad \frac{b}{a} = \alpha, \quad \frac{1}{a} = \beta.$$

Получена линейная зависимость

$$z = \alpha + \beta \cdot x.$$

Получив в результате преобразований линейное уравнение, с помощью МНК находят коэффициенты, а затем возвращаются к исходным переменным и коэффициентам, т.е. искомой зависимости.

6.2. Построение линейной регрессионной модели методом наименьших квадратов (МНК)

Рассмотрим МНК для получения оценок параметров линейной регрессионной модели $y(x) = a + bx$ по результатам наблюдений $(x_i, y_i),\ i = 1, ..., n$. В соответствии с этим методом составляют сумму квадратов разностей между y_i и $y(x_i)$

$$U(a,b) = \sum_{i=1}^{n}\left(y_i - y(x_i)\right)^2 = \sum_{i=1}^{n}\left(y_i - a - bx_i\right)^2. \qquad (6.6)$$

Определяют a и b, доставляющие минимум $U(a,b)$, решением системы

$$\begin{cases} \dfrac{\partial U}{\partial a} = 0, \\ \dfrac{\partial U}{\partial b} = 0, \end{cases}$$

которая согласно (6.6) перепишется в виде

$$\begin{cases} \sum\limits_{i=1}^{n}(y_i - a - bx_i) = 0, \\ \sum\limits_{i=1}^{n} x_i y_i - a\sum\limits_{i=1}^{n} x_i - b\sum\limits_{i=1}^{n} x_i^2 = 0, \end{cases}$$

или

$$\begin{cases} a + b \cdot \overline{x} = \overline{y}, \\ a \cdot n \cdot \overline{x} + b\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i. \end{cases} \qquad (6.7)$$

Полученную систему легко решить, например, методом Крамера

$$\Delta = \begin{vmatrix} 1 & \overline{x} \\ n \cdot \overline{x} & \sum_{i=1}^{n} x_i^2 \end{vmatrix} = \sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2,$$

$$\Delta_b = \begin{vmatrix} 1 & \overline{y} \\ n \cdot \overline{x} & \sum_{i=1}^{n} x_i y_i \end{vmatrix} = \sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y},$$

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2} = \frac{n \cdot \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right) \cdot \left(\sum_{i=1}^{n} y_i \right)}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}. \qquad (6.8)$$

Оценку для $\hat{a}$ найдем из первого уравнения системы (6.7)

$$\hat{a} = \overline{y} - \hat{b} \cdot \overline{x} = \frac{1}{n} \left(\sum_{i=1}^{n} y_i - \hat{b} \cdot \sum_{i=1}^{n} x_i \right) \qquad (6.9)$$

Таким образом, получена оценка регрессии

$$\hat{y} = \hat{a} + \hat{b} \cdot x.$$

где $\hat{a}$ и $\hat{b}$ определяются формулами (6.8) и (6.9).

6.3. Определение качества аппроксимации

Мера ошибки, которая получается при оценке или предсказании η по заданным значениям ξ с помощью уравнения регрессии, называется стандартной ошибкой оценивания или стандартной ошибкой предсказания.

Квадрат стандартной ошибки называется остаточной дисперсией $S^2_{ост}$ и определяется по формуле

$$S^2_{ост} = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{a} - \hat{b}\cdot x_i\right)^2 . \qquad (6.10)$$

Для того, чтобы проверить значимость построенной линейной регрессионной модели, осуществляется проверка значимости коэффициента регрессии: проверяется гипотеза $H_0 : b = 0$, т.е. проверяется, отличается ли статистически значимо оценка коэффициента регрессии от нуля. Проверка проводится с помощью статистики

$$t = \frac{\hat{b}}{s_{\hat{b}}}, \qquad (6.11)$$

имеющей распределение Стьюдента с $(n-2)$ степенями свободы. Если вычисленное по формуле (6.11) значение статистики $\hat{t}$ удовлетворяет условию $\left|\hat{t}\right| > t_{кр}$, где $t_{кр}$ найдено по *Таблице III* Приложения для $(n-2)$ и заданного α, то гипотеза H_0 отвергается, т.е. $\hat{b}$ значимо отличается от нуля.

В формуле (6.11) $s_{\hat{b}}$ – стандартное отклонение $\hat{b}$ определяется следующим образом:

$$s_{\hat{b}} = \sqrt{\frac{n\cdot S^2_{ост}}{n\cdot\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}}$$

Пример 1. Даны три пары наблюдений величин x и y: (2; 5), (4; 3), (6; 7). Построим линейную регрессионную модель и вычислим коэффициент корреляции между x и y. Для удобства вычислений построим следующую таблицу:

N	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2	5	10	4	25
2	4	3	12	16	9
3	6	7	42	36	49
Σ	12	15	64	56	83

Решение. Расчет линейной регрессии и коэффициента корреляции по приведенным данным:

$$\hat{b} = \frac{n \cdot \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{3 \cdot 64 - 12 \cdot 15}{3 \cdot 56 - 12^2} = \frac{1}{2},$$

$$\hat{a} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \hat{b} \cdot \sum_{i=1}^{n} x_i\right) = \frac{1}{3}\left(12 - \frac{1}{2} \cdot 15\right) = 3.$$

Оценка регрессии $\hat{y}$ по x:

$$\hat{y} = 3 + \frac{1}{2} x.$$

Аналогично можно вычислить оценку регрессии $\hat{x}$ по y:

$$\hat{b} = \frac{n \cdot \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2} = \frac{3 \cdot 64 - 12 \cdot 15}{3 \cdot 83 - 15^2} = \frac{1}{2},$$

$$\hat{a}=\frac{1}{n}\left(\sum_{i=1}^{n}x_i-\hat{b}\cdot\sum_{i=1}^{n}y_i\right)=\frac{1}{3}\left(12-\frac{1}{2}\cdot 15\right)=\frac{3}{2}.$$

Получаем

$$\hat{x}=\frac{3}{2}+\frac{1}{2}y.$$

Коэффициент корреляции можно вычислить по формуле

$$\hat{r}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})\cdot(y_i-\overline{y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)\left(\sum_{i=1}^{n}(y_i-\overline{y})^2\right)}}=$$

$$=\frac{n\cdot\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\cdot\sum_{i=1}^{n}y_i}{\sqrt{\left(n\cdot\sum_{i=1}^{n}x_i{}^2-\left(\sum_{i=1}^{n}x_i\right)^2\right)\left(n\cdot\sum_{i=1}^{n}y_i{}^2-\left(\sum_{i=1}^{n}y_i\right)^2\right)}}=$$

$$=\frac{3\cdot 64-12\cdot 15}{\sqrt{\left(3\cdot 56-12^2\right)\cdot\left(3\cdot 83-15^2\right)}}=\frac{1}{2}.$$

Пример 2. По данным **примера 1** вычислить остаточную дисперсию и проверить на значимость коэффициент регрессии $\hat{b}$ в уравнении $\hat{y}=3+\frac{1}{2}x,$

$H_0 : b=0;$
$H_1 : b\neq 0, \alpha=0{,}1.$

Решение. Остаточная дисперсия $s^2_{\hat{\imath}\,\tilde{n}\grave{o}}$ вычисляется по формуле:

$$S^2_{\text{ост}} = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{a} - \hat{b}\cdot x_i\right)^2 = \frac{1}{3-2}\left(1^2 + (-2)^2 + 1^2\right) = 6\,.$$

При проверке на значимость коэффициента регрессии $\hat{b}$ вычислим стандартное отклонение

$$s_{\hat{b}} = \sqrt{\frac{n\cdot s^2_{\text{ост}}}{n\cdot\sum\limits_{i=1}^{n}x_i^2 - \left(\sum\limits_{i=1}^{n}x_i\right)^2}} = \sqrt{\frac{3\cdot 6}{3\cdot 56 - 12^2}} = \frac{\sqrt{3}}{2}$$

и значение статистики критерия t

$$\hat{t} = \frac{\hat{b}}{s_{\hat{b}}} = \frac{1}{\sqrt{3}} \approx 0{,}58\,.$$

Статистика критерия имеет распределение Стьюдента с $(n-2)$ степенями свободы. Для уровня значимости $\alpha = 0{,}1$ и $(3-2) = 1$ степени свободы по *Таблице III* Приложения получаем

$$t_{\text{кр}} = 6{,}31$$

и т.к. $\left|\hat{t}\right| < t_{\text{кр}}$, то гипотезу H_0 о незначимости коэффициента регрессии надо принять.

6.4. Задачи

6.1. Зависимость признака η от признака ξ характеризуется таблицей:

x_i	13	17	10	17	20	11	15
y_i	12	17	11	13	16	14	15

Построить уравнение линейной регрессии и проверить на значимость коэффициент регрессии, $\alpha = 0,05$.

6.2. Восемь раз при различных значениях признака ξ было измерено значение признака η. Получены следующие результаты:

x_i	0,30	0,91	1,50	2,00	2,20	2,62	3,00	3,30
y_i	0,20	0,43	0,35	0,52	0,81	0,68	1,15	0,85

Получить линейную регрессионную модель $y(x) = a + b \cdot x$, оценить коэффициент корреляции между ξ и η.

6.3. Пользуясь методом наименьших квадратов найти параметры a, b и c нелинейной регрессионной модели $y(x) = a + bx + cx^2$, если зависимость между ξ и η задана таблицей:

x_i	–2	–1	0	1	2	3
y_i	–2	–3	–3	–1	3	7

6.4. Зависимость между двумя величинами описывается экспонентой вида $y = a \cdot b^x$. Оценить по МНК параметры a и b, если данные измерений описаны в таблице:

x_i	1	2	3	4	5
y_i	4	1	3	5	6

6.5. Исследование зависимости продолжительности (t) решения систем линейных уравнений одинаковой степени трудности от порядка системы (n) дано следующей таблицей:

n	2	3	4	5	6	7	8	9	10
t	12	35	75	130	210	315	445	600	800

Предполагая, что $t = a \cdot n^b$, найти a и b методом наименьших квадратов.

Указание. В задачах 6.4 и 6.5 провести линеаризацию зависимости.

§7. ВАРИАНТЫ ТИПОВЫХ ЗАДАНИЙ

Задание №1.

1. Вычислить числовые характеристики выборки: $\overline{x}$, s^2, s, V, s_k, E_x, Me, Mo.

2. Сделать предварительную проверку выборки на нормальность распределения.

3. Построить эмпирическую функцию распределения, гистограмму и полигон частот.

Задание №2.

Построить доверительные интервалы для параметров нормального распределенной генеральной совокупности:

1) доверительный интервал для параметра μ при известной дисперсии (дисперсию задать самостоятельно). $\gamma = 0,95$;

2) доверительный интервал для параметра μ при неизвестной дисперсии. $\gamma = 0,9;\ \ 0,95;\ \ 0,99$;

3) доверительный интервал для дисперсии. $\gamma = 0,9;\ \ 0,95;\ \ 0,99$.

4) по критерию Пирсона проверить гипотезу о законе распределения.

Задание №3.

1. Вычислить эмпирический коэффициент корреляции r_{xy};

2. Вычислить ранговый коэффициент корреляции Спирмена r_S;

3. Проверить на значимость r_{xy} и r_s, $\alpha = 0,1; 0,05$;

4. Проверить гипотезы о равенстве средних и дисперсий двух выборок, $\alpha = 0,05$.

Задание №4.

1. Построить линейную регрессионную модель;

2. Проверить на значимость коэффициент регрессии. $\alpha = 0,05$.

§8. СТАТИСТИЧЕСКИЙ МАТЕРИАЛ ДЛЯ РАСЧЕТОВ

Нижеперечисленные данные для статистической обработки собраны на основе реальной промысловой информации. В начале каждой выборки указан ее объем.

8.1. Одномерные выборки

1. $n = 35$ 43; 21; 32; 18; 28; 26; 48; 28; 28; 61; 56; 26; 17; 22; 25; 52; 16; 52; 20; 45; 48; 91; 38, 98; 88; 58; 27; 49; 53; 15; 22; 60; 44; 60; 42.

2. $n = 35$ 71; 73; 19; 47; 78; 28; 35; 22; 48; 86; 27; 50; 27; 109; 20; 54; 58; 64; 56; 98; 55; 12; 52; 24; 24; 22; 67; 71; 23; 58; 19; 68; 31; 41; 95.

3. $n = 40$ 101; 102; 103; 104; 105; 106; 208; 210; 211; 212; 213; 214; 215; 216; 217; 218; 109; 110; 111; 219; 220; 221; 112; 113; 114; 115; 116; 117; 222; 223; 224; 118; 119; 120; 121; 124; 126; 130; 131; 132.

4. $n = 35$ 42; 21; 31; 18; 27; 28; 48; 26; 28; 62; 56; 26; 18; 22; 26; 52; 16; 53; 20; 46; 48; 81; 38; 88; 86; 58; 49; 53; 18; 22; 60; 44; 60; 44; 71.

5. $n = 35$ 73; 19; 47; 78; 28; 35; 22; 48; 86; 27; 50; 27; 29; 20; 54; 16; 20; 11; 15; 43; 29; 28; 33; 50; 50; 48; 49; 62; 31; 23; 24; 56; 54; 54; 25.

6. $n=30$ 9192; 9161; 9162; 9163; 9128; 9114; 9113; 9126; 9127; 9115; 9122; 9111; 9121; 9137; 9112; 9064; 9074; 9072; 9073; 9098; 9086; 9088; 9099; 9096; 9097; 9125; 9036; 9034; 9033; 9028.

7. $n=30$ 9217; 9165; 9155; 9160; 9367; 9143; 9045; 9149; 9148; 9150; 9077; 9078; 9101; 9100; 9061; 9035; 9324; 9046; 9036; 9037; 9055; 9325; 9258; 9280; 9218; 9050; 9056; 9234; 9137; 9158.

8. $n=40$ 120; 240; 72; 240; 144; 145; 120; 72; 73; 144; 96; 144; 96; 168; 121; 96; 98; 192; 144; 192; 149; 168; 145; 312; 288; 168; 120; 292; 168; 144; 72; 144; 146; 144; 96; 120; 120; 144; 168; 122.

9. $n=40$ 8; 25; 4; 5; 6; 16; 10; 12; 32; 12; 9; 23; 31; 12; 7; 48; 7; 8; 10; 4; 4; 50; 9; 4; 40; 5; 20; 24; 11; 42; 11; 11; 7; 10; 5; 10; 14; 13; 6; 4.

10. $n=40$ 8; 40; 10; 9; 8; 5; 3; 44; 5; 6; 6; 7; 5; 33; 25; 7; 5; 4; 7; 14; 8; 27; 31; 35; 15; 8; 6; 2; 41; 12; 17; 18; 34; 45; 44; 21; 9; 8; 10; 5.

11. $n=40$ 92; 44; 28; 31; 59; 57; 55; 37; 89; 98; 36; 77; 33; 11; 79; 52; 52; 33; 23; 32; 19; 48; 62; 31; 46; 33; 33; 52; 75; 77; 100; 36; 29; 31; 85; 89; 32; 37; 26; 22.

12. $n=40$ 56; 48; 39; 42; 47; 32; 18; 41; 33; 29; 60; 32; 66; 68; 33; 47; 30; 34; 40; 33; 58; 35; 63; 55;

20; 32; 17; 38; 56; 44; 44; 42; 21; 36; 46; 39; 40; 37; 60; 60.

13. $n = 35$ 39; 6; 40; 38; 25; 14; 75; 64; 45; 43; 39; 59; 30; 66; 57; 37; 15; 47; 10; 33; 88; 61; 53; 12; 33; 55; 49; 34; 54; 58; 36; 41; 24; 34; 34.

14. $n = 40$ 40,1; 46,34; 50,5; 45,0; 31,8; 44,8; 43,7; 44,0; 28,0; 80,0; 19,9; 11,8; 28,4; 53,0; 40,5; 56,0; 79,6; 23,9; 50,7; 59,4; 61,7; 32,4; 0,3; 46,1; 60,5; 56,8; 28,3; 34,0; 47,4; 68,6; 40,5; 42,0; 63,4; 56,5; 75,2; 47,8; 21,6; 30,9; 27,6; 14,5.

15. $n = 40$ 64,9; 72,7; 52,5; 39,2; 54,0; 39,2; 57,3; 32,0; 62,1; 73,1; 112,9; 106,4; 79,8; 92,2; 68,7; 112,2; 69,0; 20,1; 117,6; 105,3; 105,4; 109,4; 80,6; 59,2; 86,0; 70,0; 31,1; 80,7; 58,2; 68,2; 36,2; 93,0; 68,7; 69,7; 86,3; 0,4; 76,7; 73,9; 94,3; 92,6.

16. $n = 40$ 89,7; 10,5; 82,2; 54,5; 1,2; 59,0; 61,8; 77,3; 50,5; 41,7; 65,0; 32,8; 60,8; 61,8; 57,1; 42,0; 39,8; 43,2; 46,3; 47,8; 36,1; 16,6; 57,5; 41,5; 26,0; 44,5; 43,0; 27,3; 34,2; 64,1; 26,8; 30,7; 25,8; 59,3; 46,3; 44,3; 43,4; 19,8; 36,1; 61,2.

17. $n = 40$ 28, 30, 32, 27, 18, 28, 26, 22, 31, 25, 24, 26, 28, 32, 31, 33, 35, 24, 22, 32, 29, 28, 24, 22, 31, 28, 31, 28, 35, 27, 28, 29, 30, 30, 30, 35, 22, 30, 24, 27.

18. $n = 35$ 888, 1320, 792, 456, 1160, 2880, 1048, 792, 1260, 1160, 288, 552, 576, 744, 576, 1440, 360, 1400, 528, 480, 1100, 1160, 912, 600, 744, 1080, 432, 1230, 768, 1080, 1920, 888, 792, 500, 502.

19. $n = 35$ 216, 96, 192, 144, 312, 144, 216, 150, 120, 96, 150, 96, 360, 150, 144, 150, 150, 144, 96, 192, 192, 168, 120, 120, 128, 96, 96, 120, 144, 96, 168, 120, 72, 240, 96.

20. $n = 35$ 1664, 1344, 1200, 1480, 3152, 1992, 2424, 1248, 3000, 1680, 1288, 912, 2160, 1752, 1320, 2280, 1656, 1152, 456, 1010, 1360, 1248, 3288, 3312, 2978, 2960, 1700, 1744, 1416, 1072, 2258, 1408, 2088, 2640, 1600.

21. $n = 40$ 70,0; 31,1; 80,7; 58,2; 68,2; 36,2; 93,0; 68,7; 69,7; 86,3; 80,4; 76,7; 73,9; 94,3; 92,6; 89,7; 10,5; 82,2; 54, 5; 41,2; 59,0; 61,8; 77,3; 50,5; 41,7; 65,0; 32,8; 60,8; 61,8; 57,1; 42,0; 39,8; 43,2; 46,3; 47,8; 36,1; 16,6; 57,5; 41,5; 26,0.

22. $n = 40$ 600; 168; 576; 196; 120; 1128; 168; 120; 120; 192; 512; 240; 120; 552; 120; 196; 288; 144; 144; 168; 120; 144; 240; 198; 384; 144; 198; 192; 150; 744; 144; 120; 144; 192; 264; 240; 150; 150; 267; 144.

23. $n = 40$ 867; 866; 1299; 1578; 1053; 1054; 579; 1316; 1314; 1315; 1313; 1312; 1094; 2416; 1096; 2407; 1501; 2404; 1011; 1508; 2422; 1507; 2424; 1790; 1791; 2418; 2419; 2511; 1603; 2512; 2423; 2425; 1140; 2515; 2516; 2428; 2427; 1605; 1606; 1510.

24. $n = 40$ 180; 47; 465; 50; 40; 135; 130; 330; 40; 20; 340; 200; 140; 200; 90; 192; 171; 363; 472; 440; 550; 78; 20; 450; 440; 20; 520; 280; 227; 78; 20; 40; 120; 126; 459; 20; 18; 438; 194; 100.

25. $n = 35$ 2512; 2423; 2425; 1140; 2515; 2516; 2428; 2427; 1605; 1606; 1510; 2517; 2426; 1168; 817; 892; 891; 1169; 1322; 1321; 1320; 1319; 2405; 2501; 1665; 1666; 2401; 2402; 2403; 2406; 1502; 1623; 2502; 1503; 2503.

26. $n = 35$ 795; 560; 342; 588; 367; 432; 215; 123; 327; 1004; 1342; 534; 858; 610; 708; 1183; 1933; 1537; 939; 685; 1594; 631; 893; 756; 624; 1232; 1173; 1211; 1006; 961; 1128; 1302; 1118; 737; 280.

27. $n = 40$ 920; 1200; 1305; 695; 1304; 1306; 1129; 737; 819; 1307; 818; 780; 1170; 821; 1171; 1277; 1279; 1275; 1273; 1276; 1274; 1280; 1278; 1270; 1269; 1267; 1268; 1265; 1266; 830; 1139; 828; 867; 866; 1299; 578; 1053; 1054; 579; 1316.

28. $n = 40$ 714; 716; 1978; 433; 776; 775; 735; 1128; 734; 1127; 1162; 889; 500; 455; 434; 568; 816; 815; 814; 1167; 1165; 443; 1108; 1005; 667; 1302; 1303; 633; 632; 590; 631; 1325; 589; 921; 920; 1200; 1305; 695; 1304; 1306.

29. $n = 35$ 591; 789; 592; 441; 810; 850; 875; 1152; 1161; 420; 421; 422; 442; 458; 1169; 1106; 1318; 811; 1012; 673; 849; 857; 855; 840; 674; 736; 777; 1317; 1223; 1224; 708; 714; 716; 1178; 433.

30. $n = 35$ 120; 72; 72; 216; 96; 96; 72; 144; 432; 72; 72; 120; 48; 24; 72; 96; 96; 144; 96; 72; 96; 72; 72; 120; 72; 72; 168; 48; 144; 144; 199; 384; 408; 96; 72.

31. $n = 35$ 120; 144; 168; 48; 120; 72; 144; 96; 72; 72; 136; 120; 96; 192; 72; 72; 96; 72; 96; 72; 164; 72; 72; 168; 72; 96; 96; 48; 120; 72; 96; 72; 120; 72; 72.

32 . $n = 40$ 61,7; 32,4; 20,3; 46,1; 60,5; 56,8; 28,3; 34,0; 47,4; 68,6; 40,5; 42,0; 63,4; 56,5; 75,2; 47,8; 21,6; 30,9; 27,6; 14,5; 64,9; 72,7; 52,5; 39,2; 54,0; 39,2; 57,3; 32,0; 62,1; 73,1; 112,9; 106,4; 79,8; 92,2; 68,7; 112,2; 69,0; 20,1; 117,6; 105,3.

8.2. Двумерные выборки

1. $n = 30$ (4,570; 3,558), (3,017; 3,825), (3,511; 3,499), (4,393; 5,793), (5,522; 3,975), (3,066; 4,913), (4,657; 5,036), (5,143; 4,547), (3,824; 5,904), (3,248; 6,784), (3,105; 3, 708), (3,857; 5,002), (3,701; 3,124), (3,662 3,725), (5,194; 3,165), (3,190; 3,103), (2,405; 3,271), (2,807; 3,128), (3,824; 2,958), (3,631; 6,284), (4,879; 3,372), (6,959; 3,533), (4,354; 3,143), (3,651; 5,197), (5,426; 4,478), (3,229; 3,528), (3,547; 5,927), (3,296; 5,231), (4,025; 3,502), (6,285; 5,717).

2. $n = 40$ (11,49; 8,52), (10,28; 11,31), (11,65; 10,36), (11,39; 10,81), (12,15; 10,35), (9,49; 15,58), (9,92; 11,62), (11,00; 13,60), (11,78; 9,76), (12,92; 12,82), (9,76; 9,61), (12,37; 10,23), (9,46; 10,35), (10,45; 9,13), (15,72; 12,40), (12,84; 10,53), (13,00; 11,28), (12,51; 10,23), (14,07; 13,14), (10,46; 12,46), (11,75; 10,45), (12,09; 11,69), (12,72; 10,92), (15,49; 11,43), (12,14; 12,41), (11,26; 13,49), (11,81; 12,17), (9,13; 12,89), (12,24; 11,14), (13,59; 12,98), (9,55; 13,06), (15,88; 12,28), (13,65; 9,82), (9,64; 12,45), (10,18; 8,91), (11,15; 12,21), (9,98; 10,75), (9,27; 14,97), (10,75; 11,01), (12,60; 12,43).

3. $n = 40$ (3,96; 2,61), (3,55; 3,15), (3,66; 3,92), (2,93; 2,89), (4,61; 3,51), (2,58; 6,15), (2,99; 4,35), (4,40; 5,35), (4,86; 3,24), (4,27; 4,67), (3,32; 2,09), (5,38; 3,11), (1,93; 3,36), (3,67; 2,64), (6,27; 5,17), (4,14; 2,81), (5,43; 4,22), (4,31; 2,95), (5,29; 4,88), (4,20; 5,54), (3,72; 2,46), (3,99; 4,37), (4,59; 2,71), (6,17; 3,51), (4,22; 4,06), (4,86; 4,78), (3,62; 5,50), (2,60; 4,69), (4,12; 3,14), (6,03; 4,42), (2,86; 5,71), (6,28; 4,46), (5,41; 3,27), (2,82; 6,06), (2,42; 2,44), (3,39; 3,13), (2,50; 2,95), (2,91; 5,43), (2,99; 3,64), (4,20; 5,25).

4. $n = 30$ (18,20; 13,99), (16,06; 14,05), (13,97; 15,34), (15,62; 17,94), (18,36; 15,25), (14,97; 16,98), (13,11; 16,90), (16,41; 17,43), (14,44; 16,93), (16,19; 20,97), (14,90; 11,96), (16,09; 17,66), (15,27; 14,07), (14,66; 15,44), (20,00; 15,73), (11,84; 15,59), (12,52; 15,06), (17,91; 15,21), (11,12; 14,49), (17,22; 20,84), (18,12; 16,24), (19,94; 20,89), (16,29; 15,38), (17,44; 17,10), (17,48; 17,34), (15,09; 12,08), (11,76; 18,00), (15,74; 18,74), (16,03; 15,63), (19,86; 18,63).

5. $n = 40$ (264; 120), (144; 48), (48; 48), (552; 48), (72; 24), (288; 48), (240; 48), (336; 168), (24; 528), (72; 96), (72; 48), (48; 72), (168; 96), (72; 48), (96; 48), (96; 48), (24; 96), (168; 96), (48; 48), (72; 264), (72; 96), (24; 72), (48; 48), (480; 144), (24; 72), (48; 144), (96; 168), (144; 216), (336; 24), (48; 168), (456; 48), (48; 552), (96; 24), (72; 144), (192; 96) , (48; 24), (24; 24), (24; 48), (24; 96), (24; 96).

6. $n = 30$ (9,0; 10,0), (5,0; 17,0), (8,0; 8,6), (6,0; 10,5), (3,0; 5,0), (5,3; 4,0), (5,0; 3,0), (4,0; 3,0), (14,0; 13,5), (5,8; 4,5), (8,5; 7,5), (5,0; 5,2), (16,0; 9,0), (19,1; 6,5), (3,9; 23,0), (6,0; 5,0), (24,0; 4,0), (22,0; 8,0), (8,0; 14,0), (4,5; 5,0), (3,0; 8, 6), (7,5; 8,0), (5,0; 1,1), (10,0; 6,5), (5,0; 7,0), (4,0; 9,3), (14,5; 4,5), (7,0; 9,0), (9,0; 7,0), (6,0; 9,0).

7. $n = 30$ (250, 530), (620, 395), (471, 25), (370, 70), (95, 0), (90, 260), (1027, 0), (695, 105), (385, 522), (260, 35), (445, 360), (125, 100), (230, 60), (275, 725), (70, 40), (970, 445), (534, 325), (100, 439), (1140, 20), (0, 690), (280, 247), (440, 91), (300, 140), (360, 320), (85, 130), (337, 1133), (1140, 0), (165, 723), (95, 240), (53, 450).

8. $n = 40$ (156; 18), (43; 29), (83; 54), (44; 58), (27; 32), (48; 81), (48; 42), (28; 91), (45; 98), (52; 49), (142; 20), (60; 54), (19; 61), (25; 156), (32; 79), (36; 80), (88; 21), (50; 19), (78; 52), (12; 118), (28; 41), (26; 48), (22; 83), (22; 30), (109; 42), (36; 35), (54; 41), (47; 69), (142; 20), (21; 14), (58; 68), (67; 31), (35; 32), (43; 17), (71; 29), (14; 34), (59; 20), (37; 20), (61; 23), (26; 24).

9. $n = 30$ (-304; -386), (35; -305), (-330; -105), (-400; -234), (-185; -160), (-160; -285), (-370; -343), (65; -35), (-51; 45), (-380; -388), (-68; 10), (48; -340), (-361; -475), (-2; -320), (-395; -240), (-356; -67), (35; -398), (-268; 70), 19; (-362; 0), (73; -10), (-192; -310), (-285; -404), (-300; 60), (-400; 5), (-349; -305), (21; -400), (-375; -80), (-365; -272), (-355; -363), (-380; -266).

10. $n = 40$ (28; -111), (115; -111), (-203; -32), (440; 98), (-353; 29), (360; 77), (79; -361), (330; -300), (-363; -105), (250; -329), (-302; 182), (-475; -322), (-276; -201), (-145; 0), (238; -115), (455; -46), (0; 0), (-109; -236), (0; 275), (86; 58), (-354; 40), (-398; 76), (-106; 95), (-185; -233), (95; 0), (-345; 0), (92; -158), (-97; -350), (200; 0), (109; -329), (254; -345), (227; -371), (370; 280), (0; -90), (95; -203), (-112; 52), (158; -70), (-142; 260), (-282; -358), (142; -299).

11. $n = 35$ (405; 142), (115; 190), (180; 90), (440; 280), (25; 382), (360; 160), (443; 270), (330; 270), (0; 360), (250; 490), (70; 395), (90; 440), (105; 50), (225; 65), (238; 273), (455; 60), (0; 545), (280; 35), (0; 180), (458; 0), (25; 260), (0; 325), (320; 0), (180; 150), (460; 275), (30; 450), (475; 440), (293; 450), (200; 475), (499; 160), (254; 0), (227; 0), (370; 220), (0; 90), (455; 0).

12. $n = 40$ (96; 216), (96; 48), (72; 72), (72; 120), (48; 96), (24; 48), (96; 144), (240; 48), (168; 72), (96; 72), (72; 48), (168; 48), (48; 120), (216; 72), (168; 96), (144; 48), (96; 192), (96; 48), (48; 144), (72; 96), (96; 120), (72; 96), (144; 72), (72; 48), (48; 168), (48; 192), (96; 216), (96; 120), (72; 48), (96; 96), (72; 144), (168; 72), (72; 120), (48; 144), (120; 72), (72; 72), (72; 48), (96; 96), (72; 96), (48; 96).

13. $n = 35$ (2,96; 3,26), (5,93; 3,09), (4,11; 2,22), (2,61; 3,95), (2,91; 3,33), (1,75; 2,79), (3,07; 3,49), (3,30; 2,10), (2,68; 1,71), (2,97; 5,22), (2,65; 2,66), (2,48; 3,55), (3,45; 3,76), (2,86; 1,77), (3,62; 3,09), (3,11; 4,10), (4,12; 4,85), (3,66; 3,10), (3,32; 2,63), (3,26; 2,13), (3,10; 3,66), (4,81; 2,59), (7,14; 3,25), (2,63; 3,32), (2,66; 3,23), (4,13; 4,36), (3,75; 4,56), (3,91; 6,26), (3,63; 7,36), (3,66; 3,20), (2,61; 0,00), (2,86; 8,28), (3,19; 4,29), (2,85; 2,45), (1,89; 3,91).

14, $n = 40$ (13,0; 5,0), (7,0; 5,0), (7,0; 4,0), (9,0; 7,0), (14,0; 12,0), (5,0; 13,0), (6,0; 3,0), (6,0; 4,5), (6,5; 11,0), (5,5; 3,0), (5,5; 5,0), (7,0; 20,0), (6,0; 6,0), (2,5; 6,0), (6,8; 6,0), (8,0; 10,0), (4,0; 8,0), (2,0; 7,0), (7,0; 1,0), (7,0; 9,0), (5,0; 5,0), (3,0; 13,5), (7,0; 11,0), (20,0; 9,0), (3,0; 5,0), (6,0; 10,0), (5,0; 7,8), (7,0; 6,0), (2,0; 6,0), (5,5; 7,0), (5,0; 7,0), (5,0; 7,0), (2,5; 5,0), (4,8; 6,0), (7,0; 6,5), (4,0; 4,5), (2,0; 4,0), (3,0; 7,0), (6,0; 5,0), (4,0; 5,0).

15, $n = 40$ (10,0; 3,0), (4,0; 3,0), (19,0; 1,0), (5,0; 10,0), (8,0; 2,5), (3,0; 5,0), (3,0; 2,8), (3,0; 1,0), (10,5; 3,0), (5,0; 6,0), (5,5; 6,0), (3,8; 3,0), (4,0; 6,0), (16,5; 5,0), (16,0; 20,0), (14,0; 4,5), (4,0; 3,0), (6,0; 4,0), (3,0; 4,0), (3,5; 2,0), (5,5; 8,0), (7,0; 4,0), (15,0; 5,0), (5,0; 5,0), (4,0; 3,5), (8,0; 7,5), (5,0; 8,0), (15,5; 3,5), (2,0; 4,0), (5,0; 3,5), (3,5; 4,5), (22,8; 6,8), (4,5; 4,0), (4,0; 4,0), (7,5; 4,0), (3,0; 3,8), (4,3; 3,0), (6,0; 3,0), (5,0; 23,0), (8,8; 5,0).

16. $n = 35$ (16,03; 16,74), (19,99; 15,82), (17,15; 14,89), (14,69; 17,09), (17,56; 16,64), (13,53; 15,53), (15,19; 16,07), (15,84; 15,27), (16,85; 13,62), (15,86; 20,37), (16,68; 16,21), (16,38; 17,47), (17,96; 19,69), (16,66; 15,84), (17,05; 17,24), (13,89; 18,10), (19,11; 18,86), (16,20; 17,35), (16,49; 15,49), (15,89; 14,12), (17,43; 17,72), (19,46; 15,05), (20,95; 16,03), (16,11; 15,60), (14,74; 16,05), (18,40; 17,28), (17,01; 18,33), (17,45; 21,91), (16,06; 23,38), (17,52; 16,04), (14,86; 23,38), (16,13; 24,31), (15,46; 19,54), (15,29; 15,06), (14,21; 16,65).

17. $n = 30$ (142,0; 59,0), (38,0; 42,0), (16,0; 26,0), (41,0; 26,0), (48,0; 15,0), (23,0; 24,0), (21,0; 65,0), (43,0; 29,0), (11,5; 50,0), (30,5; 20,0), (20,0; 12,0), (44,0; 48,0), (21,0; 78,0), (52,0; 50,0), (80,0; 89,0), (80,0; 68,5), (72,0; 33,0), (36,0; 47,0), (14,0; 12,0), (26,0; 90,0), (38,0; 39,0), (36,0; 41,0), (69,0; 29,0), (15,0; 43,0), (48,5; 81,0), (66,0; 24,0), (11,0; 44,0), (79,0; 32,0), (30,0; 26,0), (43,5; 27,0).

18. $n = 40$ (158,7; 161,7), (183,6; 207,4), (177,6; 188,9), (110,8; 119,1), (190,3; 201,4), (119,0; 123,0), (206,9; 213,9), (110,0; 119,0), (120,0; 123,0), (179,4; 205,9), (117,3; 126,5), (104,8; 120,0), (171,4; 201,0), (165,3; 170,0), (119,8; 126,7), (117,5; 124,4), (163,0; 170,3), (127,2; 140,5), (141,4; 147,3), (186,0; 201,0), (134,0; 149,0), (127,0; 136,0), (112,0; 131,0), (122,0; 153,0), (134,0; 143,0), (140,0; 151,0), (132,0; 144,0), (147,0; 163,0), (122,6; 134,0), (130,0; 160,0), (126,0; 144,0), (139,9; 156,7), (118,0; 127,0), (129,8; 155,2), (125,2; 136,3), (112,0; 131,0), (161,3; 169,1), (200,7; 209,0), (167,0; 129,4), (113,9; 137,0), (123,6; 122,2), (105,1; 123,8), (176,6; 180,2), (132,2; 147,0), (132,0; 153,6).

19. $n = 35$ (2071,7; 2078,1), (1981,0; 1971,0), (2125,4; 2110,0), (2023,1; 2019,0), (2034,9; 2035,0), (2317,5; 2318,0), (2303,3; 2300,0), (1989,8; 1998,0), (2037,0; 2057,0), (2294,9; 2308,0), (2275,0; 2290,0), (2232,7; 2240,0), (2193,1; 2187,0), (2218,6; 2230,0), (2438,8; 2430,0), (2288,8; 2895,0), (2272,2; 2283,3), (1885,3; 1906,0), (2410,6; 2415,8), (2212,0; 2219,4), (2503,8; 2503,5), (2520,7; 2532,0), (2052,9; 2041,9), (2097,1; 2101,2), (2014,9; 1999,0), (2100,2; 2084,1), (2012,3; 2005,9), (2308,9; 2303,6), (2227,0; 2225,7), (1584,2; 1593,2), (2103,5; 2113,4), (2455,8; 2478,3), (2301,3; 2311,3), (2106,7; 2105,3), (2015,7; 2027,0).

20. $n = 35$ (2077,9; 2081,4), (2086,1; 2091,3), (2225,6; 2228,8), (1319,0; 1325,0), (2416,3; 2420,5), (2033,5; 2037,0), (2044,0; 2055,9), (2433,2; 2420,8), (2335,6; 2346,2), (1995,8; 2003,6), (2272,8; 2276,4), (2125,3; 2127,1), (1732,9; 1734,6), (1901,6; 1899,1), (2049,1; 2044,6), (2171,6; 2170,3), (1988,3; 1985,4), (2121,2; 2120,6), (2043,6; 2040,9), (2302,2; 2305,1), (2519,3; 2515,1), (2084,4; 2081,0), (2501,2; 2505,3), (2012,0; 2010,3), (1814,7; 1810,1), (1820,3; 1823,2), (2201,1; 2206,0), (2505,4; 2507,3), (1919,1; 1916,2), (2004,3; 2000,6), (2332,3; 2336,1), (2221,1; 2119,3), (1858,1; 1861,3), (2103,3; 2108,3), (2428,1; 2429,1).

21. $n = 30$ (0,301; -2,00), (0,477; -1,8), (0,602; -0,15), (0,699; -0,14), (0,778; -0,98), (0,845; -0,35), (0,903; -0,06), (0,9542; 0,17), (1,00; 0,18), (1,041; 0,86), (1,079; 0,73), (1,114; 0,69), (1,146; 0,95), (1,176; 1,25), (1,204; 1,35), (1,230; 1,52), (1,255; 1,57), (1,301; 1,87), (1,342; 1,81), (1,362; 2,23), (1,38; 2,17), (1,398; 2,23), (1,416; 2,45), (1,431; 2,34), (1,447; 2,82), (1,462; 2,47), (1,477; 2,52), (1,491; 2,55), (1,544; 3,11), (1,579; 3,33).

22. $n = 50$ (1800; 20), (2150; 24), (1850; 24), (1600; 24), (2100; 20), (1550; 16), (1800; 22), (2000; 20), (1700; 16), (1300; 20), (1250; 21), (1600; 18), (1700; 18), (1800; 20), (2050; 21), (2050; 22), (2000; 24), (1700; 22), (1650; 20), (1450; 20), (1650; 16), (1600; 16), (1300; 17), (1600; 17), (1600; 21), (1700; 21), (1650; 19), (1900; 21), (2200; 23), (1900; 24), (1900; 20), (1500; 16), (1400; 21), (1100; 20), (1350; 18), (1350; 18), (1200; 20), (1300; 16), (1600; 18), (1400; 18), (1400; 19), (1650; 20), (1250; 33), (1300; 33), (1200; 22), (1200; 20), (1100; 16), (1150; 20), (1100; 22), (1200; 17).

23. $n = 50$ (59, 0; 2, 10), (118, 0; 2, 79), (66, 0; 4, 23), (43, 0; 2, 65), (142, 0; 3, 76), (142, 0; 2, 63), (61, 0; 4, 53), (34, 0; 2, 59), (21, 0; 5, 50), (20, 0; 2, 39), (42, 0; 2, 86), (29, 0; 4, 31), (17, 0; 3, 09), (80, 0; 6, 73), (32, 0; 3, 63), (41, 0; 2, 61), (43, 0; 3, 30), (15, 0; 1, 89), (48, 0; 2, 63), (78, 0; 2, 13), (29, 0; 3, 20), (28, 0; 3, 95), (50, 0; 3, 66), (35, 0; 3, 49), (48, 0; 2, 91), (42, 0; 1, 71), (142, 0; 2, 96), (23, 0; 1, 75), (36, 0; 3, 66), (14, 0; 3, 32), (36, 0; 4, 81), (32, 0; 3, 91), (79, 0; 3, 91), (30, 0; 3, 63), (43, 5; 3, 66), (69, 0; 7, 14), (26, 0; 5, 22), (32, 0; 4, 06), (69, 0; 3, 51), (61, 0; 3, 56), (39, 0; 6, 26), (41, 0; 7, 36), 81, 0; 8, 28), (89, 0; 2, 59), (68, 5; 3, 25), (32, 0; 3, 33), (50, 0; 2, 22), (54, 0; 2, 61), (50, 0; 4, 10), (15, 0; 2, 63),

24. $n = 40$ (101; 1220), (102; 1220), (103; 1200), (104; 1170), (105; 1160), (106; 1160), (208; 1200), (210; 1200), (211; 1200), (212; 1190), (213; 1200), (214; 1200), (215; 1190), (216; 1190), (217; 1190), (218; 1180), (109; 1240), (110; 1260), (111; 1230), (219; 1210), (220; 1210), (221; 1210), (112; 1520), (113; 1500), (114; 1500), (115; 1490), (116; 1500), (117; 1500), (222; 1150), (223; 1160), (224; 1160), (118; 1230), (119; 1230), (120; 1220), (121; 1220), (124; 1240), (126; 1240), (130; 1220), (131; 1220), (132; 1220).

25. $n = 30$ (11; 1320), (12; 1260), (13; 1260), (14; 1250), (15; 1260), (31; 1270), (32; 1280), (61; 1280), (62; 1280), (63; 1280), (64; 1270), (65; 1280), (71; 1280), (72; 1320), (73; 1310), (74; 1310), (81; 1210), (82; 1300), (83; 1200), (84; 1210), (121; 1470), (122; 1480), (151; 1320), (152; 1320), (161; 1450), (162; 1400), (171; 1550), (172; 1510), (173; 1300), (151; 1500).

26. $n = 35$ (2064; 1150), (2079; 1150), (2060; 1140), (2062; 1140), (2157; 1160), (2209; 1160), (2796; 1190), (2588; 1150), (2175; 1200), (2156; 1190), (2177; 1150), (2252; 1200), (2264; 1150), (2228; 1150), (2207; 1160), (2262; 1140), (2755; 1190), (2759; 1180), (2724; 1140), (2247; 1200), (2222; 1140), (2291; 1150), (2290; 1150), (2236; 1180), (2144; 1170), (2154; 1140), (1417; 1170), (3131; 1140), (2158; 1180), (1312; 1200), (2078; 1160), (2080; 1160), (2061; 1140), (2059; 1160), (2065; 1140).

27. $n = 30$ (2524,3; 2521,2), (2408,0; 2403,3), (2301,1; 2300,0), (1917,3; 1913,0), (2104,0; 2100,2), (2032,2; 2031,0), (2203,2; 2201,1), (2248,0; 2243,6), (2004,8; 2006,1), (1803,9; 1805,3), (1903,8; 1901,2), (2104,6; 2108,3), (2085,0; 2087,2), (2119,3; 2116,0), (2408,3; 2403,6), (1915,1; 1919,0), (2311,0; 2310,0), (2211,4; 2216,3), (2406,3; 2402,7), (2114,8; 2116,3), (1813,3; 1817,1), (1921,3; 1924,0), (2308,8; 2307,0), (2114,3; 2117,0), (1913,1; 1917,0), (2301,8; 2300,3), (2412,0; 2411,0), (1913,8; 1917,0), (2138,0; 2137,0), (2138,0; 2138,0).

28. $n = 40$ (904; 312), (291; 96), (895; 120), (1148; 149), (904; 149), (904; 72), (258; 192), (1268; 120), (1104; 120), (450; 149), (762; 120), (1112; 120), (1363; 144), (696; 120), (833; 72), (299; 144), (164; 120), (639; 120), (392; 240), (1877; 72), (728; 240), (1689; 144), (1027; 144), (1103; 120), (885; 72), (190; 77), (1394; 144), (904; 96), (1253, 144), (874, 96), (904; 168), (904; 120), (553; 96), (1681; 96), (1010; 192), (920; 144), (546; 192), (1682; 149), (782; 168), (753; 144).

29. $n = 40$ (32; 29), (23; 19), (35; 32), (31; 28), (35; 30), (37; 32), (32; 27), (0; 18), (32; 28), (30; 26), (25; 22), (34; 31), (30; 25), (29; 24), (30; 26), (30; 28), (36; 32), (35; 31), (36; 33), (38; 35), (27; 24), (27; 22), (35; 32), (34; 29), (31; 28), (25; 24), (25; 22), (35; 31), (32; 28), (34; 31), (32; 28), (36; 35), (30; 27), (31; 28), (33; 29), (33; 30), (33; 30), (32; 30), (38; 35), (25; 22).

30. $n = 40$ (33; 30), (26; 24), (31; 27), (28; 25), (27; 25), (21; 19), (27; 25), (37; 35), (32; 30), (34; 31), (28; 25), (27; 25), (27; 25), (28; 25), (27; 25), (38; 36), (30; 25), (28; 25), (28; 25), (28; 25), (36; 27), (37; 30), (28; 25), (33; 30), (36; 33), (33; 33), (35; 32), (36; 34), (35; 32), (34; 32), (28; 24), (28; 24), (32; 29), (34; 28), (38; 35), (28; 25), (28; 25), (33; 30), (28; 25), (38; 34).

31. $n = 40$ (36; 33), (26; 22), (30; 27), (28; 25), (28; 25), (38; 34), (28; 24), (33; 32), (36; 33), (36; 32), (38; 35), (36; 33), (30; 26), (30; 25), (32; 29), (33; 31), (36; 32), (0; 30), (30; 26), (34; 31), (37; 34), (36; 33), (38; 34), (32; 29), (0; 30), (38; 33), (35; 31), (38; 34), (31; 26), (26; 22), (26; 22), (32; 28), (29; 27), (28; 25), (37; 33), (32; 29), (26; 23), (37; 34), (29; 27), (38; 34).

32. $n = 40$ (9; 53), (11; 44), (6; 57), (10; 92), (10; 44), (22; 28), (13; 31), (13; 59), (13; 57), (6; 55), (4; 37), (7; 89), (3; 108), (12; 36), (12; 77), (8; 33), (11; 11), (5; 79), (11; 52), (1; 52), (8; 33), (8; 23), (7; 32), (6; 19), (8; 48), (7; 62), (12; 31), 8; 46), (1; 33), (6; 33), (20; 52), (9; 75), (3; 77), (3; 100), (10; 36), (8; 29), (5; 31), (12; 85), (6; 109), (7; 32).

§9. Приложения

9.1. Перечень распределений, используемых в пособии

1. Биномиальное

$$P_n\{\xi = k\} = C_n^k p^k (1-p)^{n-k}, k = 0,1,2,\dots n$$

$$M\xi = np, D\xi = npq.$$

2. Пуассона

$$P\{\xi = k\} = e^{-a}\frac{a^k}{k!}; k = 0,1,2,\dots$$

$$M\xi = D\xi = a$$

3. Экспоненциальное

$$f(x) = \begin{cases} 0, x < 0 \\ \lambda e^{-\lambda x}, x \ge 0 \end{cases} \quad (\lambda > 0)$$

$$M\xi = \frac{1}{\lambda};\; D\xi = \frac{1}{\lambda^2}$$

4. Гауссовское (нормальное, $N(\mu,\sigma^2)$)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\; -\infty < x < +\infty$$

$$M\xi = \mu; D\xi = \sigma^2$$

5. Пирсона (χ_n^2)

$$f(x) = \begin{cases} 0, x \le 0 \\ \dfrac{1}{2^{\frac{n}{2}}\tilde{A}\left(\frac{n}{2}\right)} x^{\frac{n-2}{2}} \cdot e^{-\frac{x}{2}}, x > 0 \end{cases}$$

$$M[\chi_n^2] = n; D[\chi_n^2] = 2n$$

6. Стьюдента (t_n)

$$f(x) = \frac{\tilde{A}\left(\frac{n+1}{2}\right)}{\tilde{A}\left(\frac{n}{2}\right)\sqrt{\pi n}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < x < +\infty$$

$$M\left[t_n\right] = 0; D\left[t_n\right] = \frac{n}{n-2}, n > 2$$

7. Фишера ($F_{m,n}$ **)**

$$f(x) = \begin{cases} 0, x \le 0 \\ \dfrac{\tilde{A}\left(\frac{m+n}{2}\right)}{\tilde{A}\left(\frac{m}{2}\right)\tilde{A}\left(\frac{n}{2}\right)} \cdot \left(\dfrac{m}{n}\right)^{\frac{m}{2}} \cdot \dfrac{1}{\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} x^{\frac{m}{2}-1}, x > 0 \end{cases}$$

$$M\left[F_{m,n}\right] = \frac{n}{n-2}, n > 2$$

$$D\left[F_{m,n}\right] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, n > 4$$

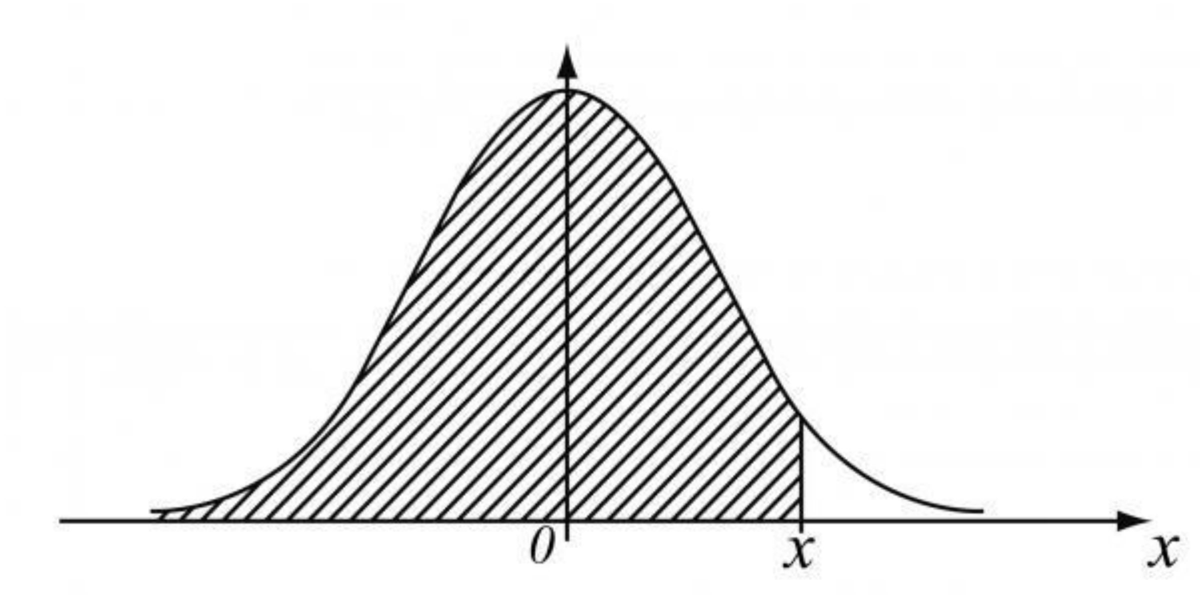
$$\tilde{A}(\mathrm{x}) = \int_0^{+\infty} t^{x-1} e^{-t} dt; \tilde{A}(\mathrm{n}) = (\mathrm{n}-1)!$$

9.2. Таблицы математической статистики

Таблица I

N(0,1), т.е.

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

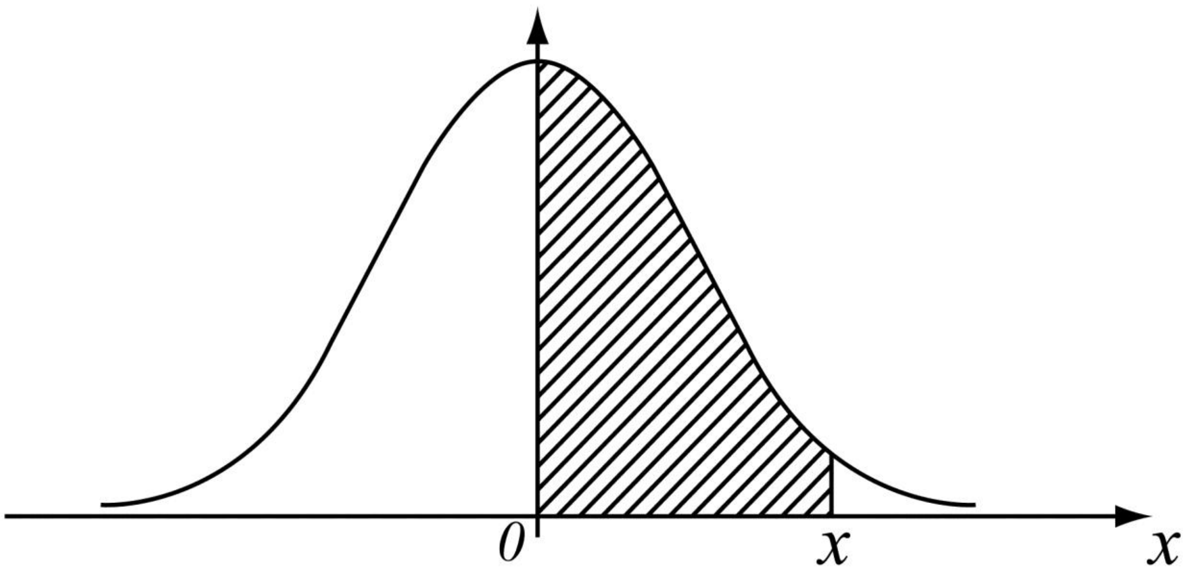


	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Таблица II

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

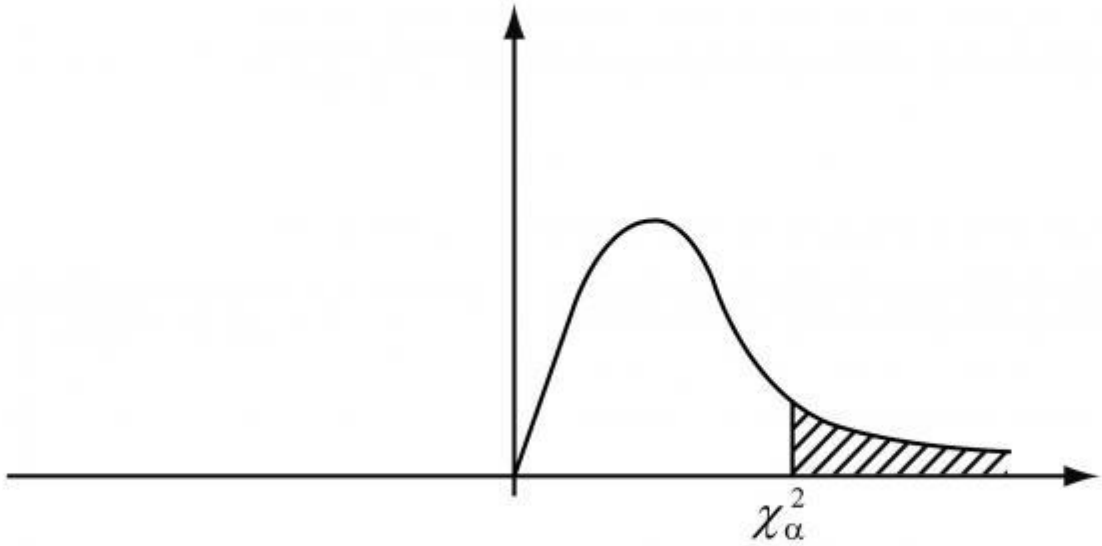


x	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,00	0,0000	0,004	0,008	0,012	0,016	0,0199	0,0239	0,0279	0,0319	0,0359
0,10	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,20	0,0793	0,0832	0,0871	0,091	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,30	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,148	0,1517
0,40	0,1554	0,1591	0,1628	0,1664	0,17	0,1736	0,1772	0,1808	0,1844	0,1879
0,50	0,1915	0,195	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,219	0,2224
0,60	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,70	0,258	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,80	0,2881	0,291	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,90	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,334	0,3365	0,3389
1,00	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,10	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,377	0,379	0,381	0,383
1,20	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,398	0,3997	0,4015

1,30	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,40	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,50	0,4332	0,4345	0,4357	0,437	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,60	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,70	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,80	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,90	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,475	0,4756	0,4761	0,4767
2,00	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,10	0,4821	0,4826	0,483	0,4834	0,4838	0,4842	0,4846	0,485	0,4854	0,4857
2,20	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,489
2,30	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,40	0,4918	0,492	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,50	0,4938	0,494	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,60	0,4953	0,4955	0,4956	0,4957	0,4959	0,496	0,4961	0,4962	0,4963	0,4964
2,70	0,4965	0,4966	0,4967	0,4968	0,4969	0,497	0,4971	0,4972	0,4973	0,4974
2,80	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,498	0,4981
2,90	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,00	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,499	0,499

Таблица III

Критические точки распределения Пирсона: $P\left\{\chi_n^2 > x_\alpha^2\right\} = \alpha$

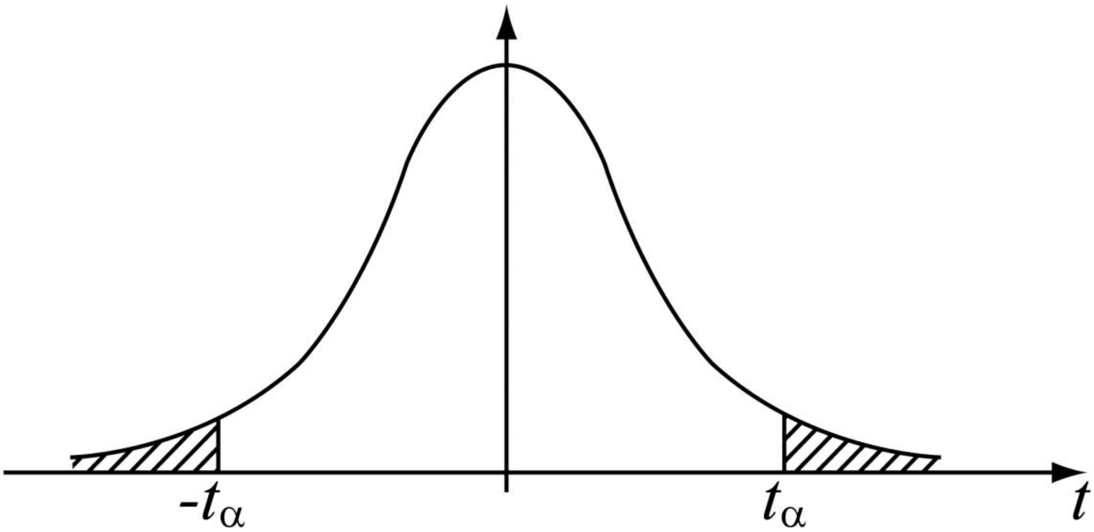


Число степеней свободы, *n*	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63490	5,02389	3,84146	0,003932	0,000982	0,000157
2	9,21034	7,37776	5,99147	0,102587	0,050636	0,020101
3	11,3449	9,34840	7,81473	0,351846	0,215795	0,114832
4	13,2767	11,1433	9,49773	0,710721	0,484419	0,297110
5	15,0863	12,8325	11,0705	1,145476	0,831211	0,554300
б	16,8119	14,4494	12,5916	1,63539	1,237247	0,872085
7	18,4753	16,0128	14,0671	2,16735	1,68987	1,239043
8	20,0902	17,5346	15,5073	2,73264	2,17973	1,646482
9	21,6660	19,0228	16,9190	3,32511	2,70039	2,087912
10	23,2093	20,4831	18,3070	3,94030	3,24697	2,55821
11	24,7250	21,9200	19,6751	4,57481	3,81575	3,05347
12	26,2170	23,3367	21,0261	5,22603	4,40379	3,57056

13	27,6883	24,7356	22,3621	5,89186	5,00874	4,10691
14	29,1413	26,1190	23,6848	6,57063	5,62872	4,66043
15	30,5779	27,4884	24,9952	7,26094	6,26214	5,22935
16	31,9999	28,8454	26,2962	7,96164	6,90766	5,81221
17	33,4087	30,1910	27,5871	8,67176	7,56418	6,40776
18	34,8053	31,5264	28,8693	9,39046	8,23075	7,01491
19	36,1908	32,8523	30,1435	10,1170	8,90655	7,63273
20	37,5662	34,1696	31,4104	10,8508	9,59083	8,26040
21	38,9321	35,4789	32,6705	11,5913	10,28293	8,89720
22	40,2894	36,7807	33,9244	12,3380	10,9823	9,54249
23	41,6384	38,0757	35,1725	13,0905	11,6885	10,19567
24	42,9798	39,3641	36,4151	13,8484	12,4011	10,8564
25	44,3141	40,6465	37,6525	14,6114	13,1197	11,5240
26	45,6417	41,9232	38,8852	15,3791	13,8439	12,1981
27	46,9630	43,1944	40,1133	16,1513	14,5733	12,8786
28	48,2782	44,4607	41,3372	16,9279	15,3079	13,5642
29	49,5879	45,7222	42,5569	17,7083	16,0471	14,2565
30	50,8922	46,9792	43,7729	18,4926	16,7908	14,9535
40	63,6907	59,3417	55,7585	26,5093	24,4331	22,1643
50	76,1539	71,4202	67,5048	34,7642	32,3574	29,7067
60	88,3794	83,2976	79,0819	43,1879	40,4817	37,4848
70	100,425	95,0231	90,5312	51,7393	48,7576	45,4418
80	112,329	106,629	101,879	60,3915	57,1532	53,5400
90	124,116	118,136	113,145	69,1260	65,6466	61,7541
100	135,807	129,561	124,342	77,9295	74,2219	70,0648
Число степеней свободы, n	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99

Таблица IV

Критические точки распределения Стьюдента: $P\{|t_n| > t_\alpha\} = \alpha$



Число степеней свободы, *n*	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,3138	12,7062	31,8205	63,6567	318,3081	636,6189
2	2,9200	4,3027	6,9646	9,9248	22,3271	31,5991
3	2,3534	3,1824	4,5407	5,8409	10,2145	12,9240
4	2,1318	2,7764	3,7469	4,6041	7,1732	8,6103
5	2,0150	2,5706	3,3649	4,0321	5,8934	6,8688
б	1,9432	2,4469	3,1427	3,7074	5,2076	5,9588
7	1,8946	2,3646	2,9980	3,4995	4,7853	5,4079
8	1,8595	2,3060	2,8965	3,3554	4,5008	5,0413
9	1,8331	2,2622	2,8214	3,2498	4,2968	4,7809
10	1,8125	2,2281	2,7638	3,1693	4,1437	4,5869
11	1,7959	2,2010	2,7181	3,1058	4,0247	4,4370
12	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	1,7709	2,1604	2,6503	3,0123	3,8520	4,2208
14	1,7613	2,1448	2,6245	2,9768	3,7874	4,1405
15	1,7531	2,1314	2,6025	2,9467	3,7328	4,0728

16	1,7459	2,1199	2,5835	2,9208	3,6862	4,0150
17	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	1,7341	2,1009	2,5524	2,8784	3,6105	3,9216
19	1,7291	2,0930	2,5395	2,8609	3,5794	3,8834
20	1,7247	2,0860	2,5280	2,8453	3,5518	3,8495
21	1,7207	2,0796	2,5176	2,8314	3,5272	3,8193
22	1,7171	2,0739	2,5083	2,8188	3,5050	3,7921
23	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	1,7109	2,0639	2,4922	2,7969	3,4668	3,7454
25	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251
26	1,7056	2,0555	2,4786	2,7787	3,4350	3,7066
27	1,7033	2,0518	2,4727	2,7707	3,4210	3,6896
28	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	1,6991	2,0452	2,4620	2,7564	3,3962	3,6594
30	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460
35	1,6896	2,0301	2,4377	2,7238	3,3400	3,5911
40	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510
45	1,6794	2,0141	2,4121	2,6896	3,2815	3,5203
50	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960
60	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602
70	1,6669	1,9944	1,3808	2,6479	3,2108	3,4350
80	1,6641	1,9901	2,3739	2,6387	3,1953	3,4163
90	1,6620	1,9867	2,3685	2,6316	3,1833	3,4019
100	1,6602	1,9840	2,3642	2,6259	3,1737	3,3905
120	1,6577	1,9799	2,3578	2,6174	3,1595	3,3735
200	1,6525	1,9719	2,3451	2,6006	3,1315	3,3398
∞	1,6449	1,9600	2,3263	2,5758	3,0902	3,2905
	0,05	0,025	0,01	0,005	0,0001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Таблица V

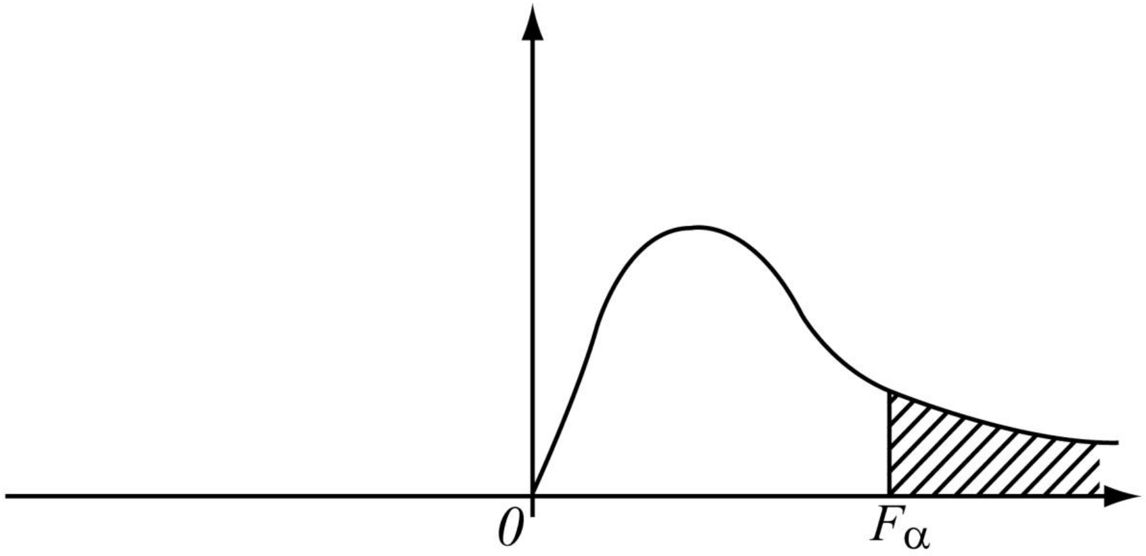
$$w = \operatorname{arcth} z$$

z	,000	,002	,004	,006	,008	z	,000	,002	,004	,006	,008
0,00	0000	0020	0040	0060	0080	0,50	5493	5520	5547	5573	5600
1	0100	0120	0140	0160	0180	1	5627	5654	5682	5709	5736
2	0200	0220	0240	0260	0280	2	5763	5791	5818	5846	5874
3	0300	0320	0340	0360	0380	3	5901	5929	5957	5985	6013
4	0400	0420	0440	0460	0480	4	6042	6070	6098	6127	6155
0,05	0500	0520	0541	0561	0581	0,55	6184	6213	6241	6270	6299
6	0601	0621	0641	0661	0681	6	6328	6358	6387	6416	6446
7	0701	0721	0741	0761	0782	7	6475	6505	6535	6565	6595
8	0802	0822	0842	0862	0882	8	6625	6655	6685	6716	6746
9	0902	0923	0943	0963	0983	9	6777	6807	6838	6869	6900
0,10	1003	1024	1044	1064	1084	0,60	6931	6963	6994	7026	7057
1	1104	1125	1145	1165	1186	1	7089	7121	7153	7185	7218
2	1206	1226	1246	1267	1287	2	7250	7283	7315	7348	7381
3	1307	1328	1348	1368	1389	3	7414	7447	7481	7514	7548
4	1409	1430	1450	1471	1491	4	7582	7616	7650	7684	7718
0,15	1511	1532	1552	1573	1593	0,65	7753	7788	7723	7858	7893
6	1614	1634	1655	1676	1696	6	7928	7964	7999	8035	8071
7	1717	1737	1758	1779	1799	7	8107	8144	8180	8217	8254
8	1820	1841	1861	1882	1903	8	8291	8328	8366	8404	8441
9	1923	1944	1965	1986	2007	9	8480	8518	8556	8595	8634
0,20	2027	2048	2069	2090	2111	0,70	8673	8712	8752	8792	8832
1	2132	2153	2174	2195	2216	1	8872	8912	8953	8994	9035
2	2237	2258	2279	2300	2321	2	9076	9118	9160	9202	9245
3	2342	2363	2384	2405	2427	3	9287	9330	9373	9417	9461
4	2448	2469	2490	2512	2533	4	9505	9549	9594	9639	9684
0,25	2554	2575	2597	2618	2640	0,75	0,973	0,978	0,982	0,987	0,991
6	2661	2683	2704	2726	2747	6	0,996	1,001	1,006	1,011	1,015
7	2769	2790	2812	2833	2855	7	1,020	1,025	1,030	1,035	1,040
8	2877	2899	2920	2942	2964	8	1,045	1,050	1,056	1,061	1,066
9	2986	3008	3029	3051	3073	9	1,071	1,077	1,082	1,088	1,093
0,30	3095	3117	3139	3161	3183	0,80	1,099	1,104	1,110	1,116	1,121
1	3205	3228	3250	3272	3294	1	1,127	1,133	1,139	1,145	1,151
2	3316	3339	3361	3383	3406	2	1,157	1,163	1,169	1,175	1,182
3	3428	3434	3473	3496	3518	3	1,188	1,195	1,201	1,208	1,214
4	3541	3564	3586	3609	3632	4	1,221	1,228	1,235	1,242	1,249
0,35	3654	3677	3700	3723	3746	0,85	1,256	1,263	1,271	1,278	1,286
6	3769	3792	3815	3838	3861	6	1,293	1,301	1,309	1,317	1,325
7	3884	3907	3931	3954	3977	7	1,333	1,341	1,350	1,358	1,367
8	4001	4024	4047	4071	4094	8	1,376	1,385	1,394	1,403	1,412
9	4118	4142	4165	4189	4213	9	1,422	1,432	1,442	1,452	1,462

Таблица VI

Критические значения распределения Фишера. F_{α, m_1, m_2} $\alpha = 0,05$

$$P\left\{F_{m_1,m_2} > F_\alpha\right\} = \alpha$$



m_1 \ m_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**20**	**24**	**30**	**40**	**50**
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	248,01	249,05	250,10	251,14	252,00
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,45	19,45	19,46	19,47	19,47
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,66	8,64	8,62	8,59	8,58
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,80	5,77	5,75	5,72	5,70
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,56	4,53	4,50	4,46	4,44
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,87	3,84	3,81	3,77	3,75
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,44	3,41	3,38	3,34	3,32
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,15	3,12	3,08	3,04	3,03
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	2,94	2,90	2,86	2,83	2,80
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,77	2,74	2,70	2,66	2,64
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,65	2,61	2,57	2,53	2,50
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,54	2,51	2,47	2,43	2,40
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,46	2,42	2,38	2,34	2,32
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,39	2,35	2,31	2,27	2,24
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,33	2,29	2,25	2,20	2,18
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,28	2,24	2,19	2,15	2,13
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,23	2,19	2,15	2,10	2,08
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,19	2,15	2,11	2,06	2,04
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,16	2,11	2,07	2,03	2,00

20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,12	2,08	2,04	1,99	1,96
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,10	2,05	2,01	1,96	1,93
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,07	2,03	1,98	1,94	1,91
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,05	2,01	1,96	1,91	1,88
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,03	1,98	1,94	1,89	1,86
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,01	1,96	1,92	1,87	1,84
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	1,99	1,95	1,90	1,85	1,82
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	1,97	1,93	1,88	1,84	1,80
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	1,96	1,91	1,87	1,82	1,78
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	1,94	1,90	1,85	1,81	1,77
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	1,93	1,89	1,84	1,79	1,76
32	4,15	3,30	2,90	2,67	2,51	2,40	2,32	2,25	2,19	2,14	1,91	1,86	1,82	1,76	1,74
34	4,13	3,28	2,88	2,65	2,49	2,38	2,30	2,23	2,17	2,12	1,89	1,84	1,80	1,74	1,71
36	4,11	3,26	2,86	2,63	2,48	2,36	2,28	2,21	2,15	2,10	1,87	1,82	1,78	1,72	1,69
38	4,10	3,25	2,85	2,65	2,46	2,35	2,26	2,19	2,14	2,09	1,85	1,80	1,76	1,71	1,67
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,84	1,79	1,74	1,69	1,66
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	1,81	1,76	1,72	1,66	1,63
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,14	2,09	2,04	1,80	1,75	1,71	1,65	1,62
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,78	1,74	1,69	1,63	1,60

Литература.

1. Сборник задач по математике для ВТУЗов (*под ред. А.В. Ефимова*). Ч.3. Теория вероятностей и математическая статистика. – М.: Наука, 1984.

2. *Фастовец Н.О.* «Элементы теории вероятностей и математической статистики». – М.: МИНХ и ГП им. И.М. Губкина, 1977.

3. *Григорьев Л.И., Подгорнов В.М., Фастовец Н.О.* «Основы математической статистики в задачах нефтегазовой отрасли» – М.: ГАНГ, 1995.

4. *Закс Л.* «Статистическое оценивание» – М.: Статистика, 1976.

5. *Хургин Я.И., Фастовец Н.О.* «Статистическое моделирование» – М.: ФГУП Изд-во «Нефть и газ» РГУ нефти и газа им. И.М. Губкина, 2003.

Оглавление

	Стр.
Введение	**2**
§1. Предварительный анализ статистических данных	**4**
1.1. Числовые характеристики (параметры) выборочного распределения	4
1.2. Предварительная проверка на нормальность	11
1.3. Задачи	13
§2. Графическое представление выборочного (эмпирического) распределения	**15**
2.1. Гистограмма и полигон частот	16
2.2. Эмпирическая функция распределения	17
2.3. Задачи	19
§3. Интервальное оценивание	**20**
3.1. Доверительный интервал для математического ожидания при известной дисперсии	21
3.2. Доверительный интервал для математического ожидания при неизвестной дисперсии	27
3.3. Доверительный интервал для дисперсии	29
3.4. Доверительный интервал для параметра p биномиального распределения	31
3.5. Задачи	33
§4. Корреляционный анализ	**33**
4.1. Коэффициент корреляции	33
4.2. Коэффициент ранговой корреляции Спирмена	34
4.3. Доверительный интервал для коэффициента корреляции r_0	35
4.4. Задачи	37

§5. Проверка статистических гипотез **38**

5.1. Проверка гипотезы о среднем значении μ нормально распределенной совокупности 40

5.2. Проверка гипотезы о равенстве средних двух независимых выборок у нормально распределенных совокупностей 42

5.3. Проверка гипотезы о значении дисперсии нормально распределенной совокупности 44

5.4. Сравнение двух выборочных дисперсий из нормально распределенных совокупностей 45

5.5. Проверка на значимость коэффициентов корреляции r и r_s 46

5.6. Критерий согласия Пирсона (χ^2) для проверки гипотезы о законе распределения 49

5.7. Задачи 53

§6. Регрессионный анализ **55**

6.1 Модели регрессионного анализа 55

6.2. Построение линейной регрессионной модели методом наименьших квадратов 58

6.3. Определение качества аппроксимации 59

6.4. Задачи 63

§7. Варианты типовых заданий **65**

§8. Статистический материал для расчетов **67**

8.1. Одномерные выборки 67

8.2. Двумерные выборки 72

§9. Приложения **82**

9.1. Перечень распределений, используемых в пособии 82

9.2. Таблицы математической статистики 84

- значения нормальной функции распределения 84

- значения функции Лапласа 86
- критические точки распределения Пирсона 88
- критические точки распределения Стьюдента 90
- значения функции $w = arcthz$ 92
- критические значения распределения Фишера 93
Литература **96**