

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
НАЦИОНАЛЬНАЯ МЕТАЛЛУРГИЧЕСКАЯ АКАДЕМИЯ
УКРАИНЫ**

Г.Г. ШВАЧИЧ, А.В. СОБОЛЕНКО, А.И. ОРЖЕХ

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ И
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Часть 2. Математическая статистика

**Утверждено на заседании Ученого совета академии
в качестве учебного пособия**

Днепропетровск НМетАУ 2006

ББК 22.142

Швачич Г.Г., Соболенко А.В., Оржех А.И. Теория вероятностей и математическая статистика. Часть 2. Математическая статистика: Учебное пособие. - Днепропетровск: НМетАУ, 2006.- 56 с.

Содержит рабочую программу по математической статистике, основной теоретический материал по основным разделам дисциплины. Представлен широкий круг примеров и задач, приведены варианты индивидуальных заданий.

Предназначено для студентов экономических специальностей заочной формы обучения.

Табл. 4. Библиогр.: 13 наим.

Печатается в авторской редакции.

Ответственный за выпуск Г.Г.Швачич, канд. техн. наук, проф.

Рецензенты: Ю.Н. Головки, канд. физ-мат. наук, доц. (УДХТУ)
Ю.Е. Чернявский, канд. физ-мат. наук, доц. (ДУБП)

© Национальная металлургическая академия Украины, 2006

© Швачич Г.Г., 2006

© Соболенко А.В., 2006

© Оржех А.И., 2006

СОДЕРЖАНИЕ

Программа дисциплины «теория вероятностей и математическая статистика» (раздел «математическая статистика») для студентов экономических специальностей.....	4
Тема 1. Задачи математической статистики. Эмпирические законы распределения.....	5
Тема 2. Статистическое оценивание параметров.....	12
Тема 3. Проверка статистических гипотез.....	16
Тема 4. Дисперсионный анализ.....	26
Тема 5. Корреляционный анализ.....	32
6. Указания к выполнению индивидуальных заданий.....	41
7. ЛИТЕРАТУРА.....	43
8. Задачи для выполнения индивидуальных заданий.....	44
8.1. Эмпирические законы распределения.....	44
8.2. Статистическое оценивание параметров.....	45
8.3. Проверка статистических гипотез.....	45
8.4. Дисперсионный анализ.....	45
8.5. Корреляционный анализ.....	47
9. Таблица вариантов индивидуальных заданий.....	48
10. Приложения.....	49
10.1. Таблица значений функции Лапласа.....	49
10.2. Таблица значений критических точек распределения Стьюдента.....	51
10.3. Таблица значений критических точек распределения χ^2	52
10.4. Таблица значений критических точек распределения Фишера.....	53

**ПРОГРАММА ДИСЦИПЛИНЫ «ТЕОРИЯ ВЕРОЯТНОСТЕЙ И
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА» (РАЗДЕЛ
«МАТЕМАТИЧЕСКАЯ СТАТИСТИКА») ДЛЯ СТУДЕНТОВ
ЭКОНОМИЧЕСКИХ СПЕЦИАЛЬНОСТЕЙ**

1. Эмпирические законы распределения. Выборочный метод. Вариационные ряды. Статистические ряды. Полигон. Гистограмма. Эмпирическая функция распределения.
2. Статистическое оценивание параметров распределений. Статистическая оценка параметра. Точечные оценки. Их свойства. Точечные оценки математического ожидания и дисперсии. Доверительная вероятность и доверительный интервал.
3. Проверка статистических гипотез. Понятие статистической гипотезы. Критерий. Уровень значимости. Проверка однородности дисперсий. Критерий согласия χ^2 - Пирсона.
4. Дисперсионный анализ. Общая идея дисперсионного анализа. Предпосылки дисперсионного анализа. Двухфакторный дисперсионный анализ.
5. Корреляционный анализ. Функциональная и статистическая зависимости. Корреляционная таблица. Виды корреляции. Простая линейная корреляция. Метод наименьших квадратов.

ТЕМА 1. ЗАДАЧИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. ЭМПИРИЧЕСКИЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ.

Математическая статистика – это раздел математики, изучающий методы сбора, систематизации и обработки результатов наблюдений массовых случайных явлений с целью выявления закономерностей.

Значительная часть математической статистики связана с необходимостью описания и изучения большой совокупности объектов. Выборочный метод состоит в том, что из большой совокупности объектов выбирается какая-нибудь часть.

Вся подлежащая изучению совокупность однородных элементов называется *генеральной совокупностью*.

Например, генеральной совокупностью могут быть все жители Москвы, месячная продукция завода, студенты данного потока и т.д.

Генеральная совокупность может быть конечной и бесконечной. При конечной генеральной совокупности каждый её элемент представляется реально существующим индивидуумом или имеющимся значением случайной величины. Объём конечной генеральной совокупности равен числу проведённых наблюдений (опытов) над случайной величиной. Часто, особенно в технических приложениях, в качестве генеральной совокупности удобно рассматривать множество значений случайной величины, которые могли бы появиться при проведении наблюдений над случайной величиной. Число наблюдений над случайной величиной теоретически может быть сколь угодно большим, поэтому генеральная совокупность значений наблюдаемой случайной величины бесконечна.

Совокупность ограниченного числа значений случайной величины, полученных в результате эксперимента, называется *выборочной совокупностью*, или выборкой из генеральной совокупности.

Законы, построенный по выборочным данным, называются эмпирическими. Эмпирические законы устанавливают связь между выборочными значениями x_1, x_2, \dots, x_n и соответствующими

относительными частотами $P_i^* = \frac{m_i}{n}$, где n – объём выборки, т.е. $n = \sum_{i=1}^k m_i$,

а m_i – числа, показывающие сколько раз данное значение x_i встречалось среди выборочных данных. Относительная частота P_i^* является статистическим аналогом вероятности ($P_i^* = \xrightarrow{\text{вер. } n \rightarrow \infty} P_i$).

Если выборочные значения расположить в порядке их возрастания, то такой ряд называется **несгруппированным вариационным рядом**. Порядковый номер элемента в вариационном ряду называется *рангом*.

При большом числе наблюдений выборочные значения группируются.

Для построения интервального вариационного ряда необходимо определить длину интервала (Δx_i) в самую левую точку, от которой откладывается интервалы. Эти вопросы чаще всего решает экспериментатор, чтобы число интервалов было не слишком малым, но и не слишком большим.

Дискретный вариационный ряд имеет вид

x_i	x_1	x_2	...	x_k
m_i	m_1	m_2	...	m_k

, где $\sum_{i=1}^k m_i = n$.

Если все интервалы имеют одинаковую длину (Δx), то можно воспользоваться для определения формулой Стёрджеса

$$\Delta x = \frac{x_{\max} - x_{\min}}{1 + 3,31 \lg n},$$

где

x_{\min} – наименьшее значение в выборке;

x_{\max} – наибольшее значение в выборке.

Величину Δx можно округлять.

Левый конец первого интервала определяется

$$x^{(1)} = x_{\min} - \frac{\Delta x}{2}.$$

Затем получаем

$$x^{(2)} = x^{(1)} + \Delta x;$$

$$x^{(3)} = x^{(2)} + \Delta x;$$

...

пока величина $x^{(k+1)}$ станет больше x_{\max} .

Получим *интервальный вариационный ряд*

J_i	$[x^{(1)}; x^{(2)})$	$[x^{(2)}; x^{(3)})$...	$[x^{(k)}; x^{(k+1)})$
m_i	m_1	m_2	...	m_k

Здесь m_i показывает, сколько выборочных значений попадает в i -й интервал. Квадратная скобка обозначает, что здесь включается левый конец интервала и не включается правый (хотя можно договориться о включении правого и невключении левого). Имея вариационные ряды, легко подсчитать относительные частоты. Если таблицу сгруппированного вариационного ряда дополнить строкой относительных частот, то получим статистический ряд. *Дискретный статистический* ряд имеет вид.

x_i	x_1	x_2	...	x_k
m_i	m_1	m_2	...	m_k
P_i^*	P_1^*	P_2^*	...	P_i^*

, где $\sum_{i=1}^k m_i = n$,

$$P_i^* = \frac{m_i}{n}, \quad \sum_{i=1}^k P_i^* = 1.$$

Интервальный статистический ряд представляется таблицей вида:

J_i	$[x^{(1)}; x^{(2)})$	$[x^{(2)}; x^{(3)})$...	$[x^{(k)}; x^{(k+1)})$
m_i	m_1	m_2	...	m_k
P_i^*	P_1^*	P_2^*	...	P_i^*

Среднюю строку (строку m_i) можно опускать.

Графически статистические ряды изображаются в виде *полигона* (*дискретный статистический ряд*) или *полигона и гистограммы* (*интервальный статистический ряд*). *Полигон* – это ломаная, соединяющая точки с координатами (x_i, P_i^*) или (\bar{x}_i, P_i^*) , где \bar{x}_i – середина i -того интервала.

Гистограмма – это плоская фигура, состоящая из прямоугольников, основания которых равны длине интервала, а высота равна или пропорциональна относительной частоте. Чтобы площадь гистограммы равнялась единице, берут $h_i = \frac{P_i^*}{\Delta x_i}$.

Дискретный статистический ряд является аналогом ряда распределения, а полигон – аналогом многоугольника распределения.

С помощью интервального статистического ряда, а именно, воспользовавшись полигоном или гистограммой, можно судить о виде плотности распределения $f(x)$.

Статистическим аналогом функции распределения $F(x)$ является эмпирическая функция распределения $F^*(x) = P^*(X < x)$, равная частоте выборочных значений, меньших некоторого фиксированного значения x .

Эмпирическая функции задаётся так:

для дискретного статистического ряда

$$F^*(x) = \begin{cases} 0, & \text{при } x \leq x_1; \\ P_1^*, & \text{при } x_1 < x \leq x_2; \\ P_1^* + P_2^*, & \text{при } x_2 < x \leq x_3; \\ \dots & \\ 1, & \text{при } x_k < x < \infty; \end{cases}$$

для интервального статистического ряда

$$F^*(x) = \begin{cases} 0, & \text{при } x = x^{(1)}; \\ P_1^*, & \text{при } x = x^{(2)}; \\ P_1^* + P_2^*, & \text{при } x = x^{(3)}; \\ \dots & \\ 1, & \text{при } x = x^{(k+1)}. \end{cases}$$

Эмпирическая функция распределения

$$F^*(x) = \begin{cases} 0, & \text{при } x \leq 0; \\ 0,1, & \text{при } 0 < x \leq 1; \\ 0,1 + 0,4, & \text{при } 1 < x \leq 2; \\ 0,1 + 0,4 + 0,3, & \text{при } 2 < x \leq 3; \\ 1, & \text{при } 3 < x < \infty. \end{cases}$$

Пример 1.1.

1. X – количество работающих завалочных машин в цехе.

Исходные данные за 10 дней: 2, 3, 1, 0, 2, 2, 3, 1, 1, 1.

Построим статистический ряд.

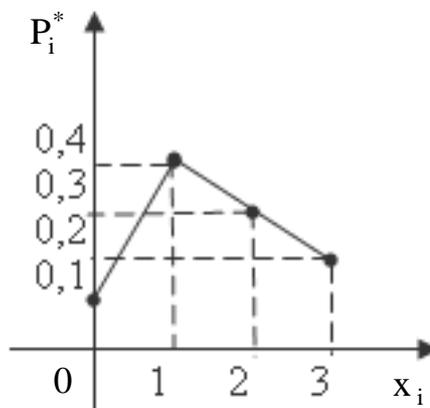
Для этого подсчитаем, сколько раз данное значение встретилось в выборке. Например, значение $x_1=0$, встречалось один раз, значит $m_1=0$. Затем разделим частоты на объём выборки (количество отобранных элементов). В нашем случае $n=10$.

x_i	0	1	2	3
m_i	1	4	3	2
P_i^*	0,1	0,4	0,3	0,2

$$\sum_{i=1}^4 m_i = 1 + 4 + 3 + 2 = 10;$$

$$P_i^* = \frac{m_i}{n}, \quad \sum_{i=1}^4 P_i^* = 1.$$

Полигон имеет вид



Эмпирическая функция распределения

$$F^*(x) = \begin{cases} 0 & \text{при } x \leq 0 ; \\ 0,1 & \text{при } 0 < x \leq 1; \\ 0,1 + 0,4 & \text{при } 1 < x \leq 2; \\ 0,1 + 0,4 + 0,3 & \text{при } 2 < x \leq 3; \\ 1 & \text{при } 3 < x < \infty. \end{cases}$$

2. X – производительность мартеновской печи.

Исходные данные по 40 плавкам:

16,0; 16,4; 16,1; 16,3; 16,4; 16,1; 16,2; 17,0; 16,8; 16,7; 16,6; 16,9; 16,7; 17,4;
17,5; 18,0; 17,6; 17,9; 15,0; 15,4; 15,3; 15,6; 15,9; 16,0; 15,7; 16,6; 16,8; 16,7;
16,5; 17,0; 15,2; 15,1; 16,4; 16,3; 16,2; 15,6; 15,9; 15,8; 15,9; 16,4.

Построим интервальный статистический ряд. Для этого определим длину интервала по формуле Стёрджеса.

$$\Delta x = \frac{x_{\max} - x_{\min}}{1 + 3,31 \lg n} = \frac{18,0 - 15,0}{1 + 3,31 \lg 40} = \frac{3,0}{1 + 3,3 \cdot 1,6} = 0,42.$$

Округлим шаг: $\Delta x = 0,4$

$$x^{(1)} = x_{\min} - \frac{\Delta x}{2} = 15,0 - \frac{0,4}{2} = 14,8;$$

$$x^{(2)} = x^{(1)} + \Delta x = 14,8 + 0,4 = 15,2;$$

$$x^{(3)} = x^{(2)} + \Delta x = 15,2 + 0,4 = 15,6;$$

...

Последнее $x^{(k+1)}$ должно быть больше или равно x_{\max} .

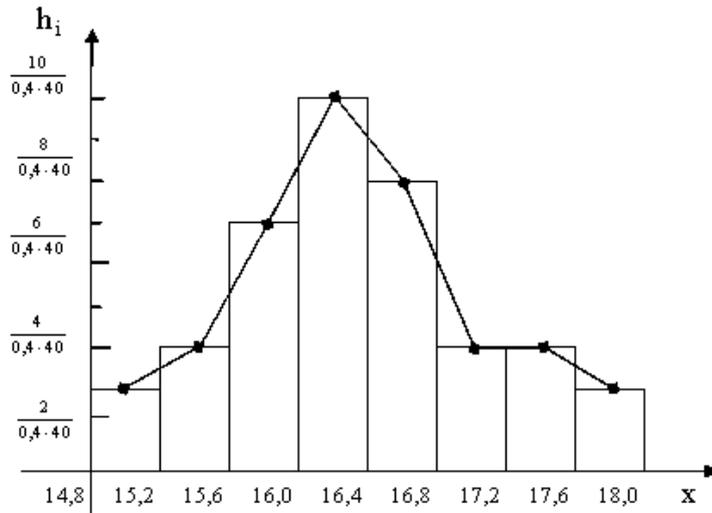
Запишем интервальный статистический ряд, подсчитав, сколько значений попадает в каждый интервал:

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	3	4	7	10	8	3	3	2
P_i^*	$\frac{3}{40} =$ 0,075	$\frac{4}{40} =$ 0,1	$\frac{7}{40} =$ 0,175	$\frac{10}{40} =$ 0,25	$\frac{8}{40} =$ 0,2	$\frac{3}{40} =$ 0,075	$\frac{3}{40} =$ 0,075	$\frac{2}{40} =$ 0,05

Графически интервальный статистический ряд можно изобразить в виде гистограммы и полигона. Высота прямоугольника $h_i = \frac{P_i^*}{\Delta x_i} = \frac{P_i^*}{0,4}$.

Полигон построим на том же чертеже, соединив середины верхних сторон прямоугольника, т.к. вид полигона не изменится, если вместо P_i^* взять

$$\frac{P_i^*}{\Delta x}$$



Эмпирическая функция распределения определяется своими значениями на концах интервалов:

$$F^*(14,8) = 0; F^*(15,2) = p_1^* = 3/40;$$

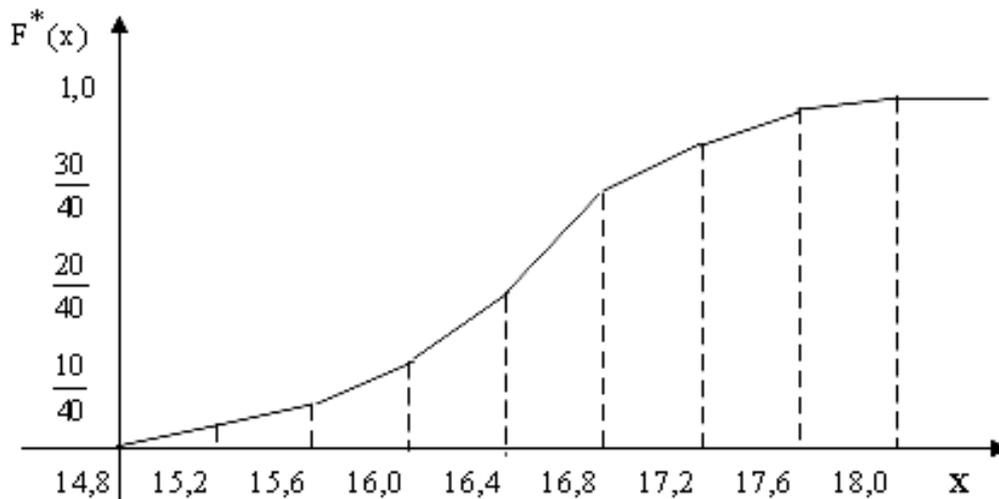
$$F^*(15,6) = \sum_1^2 p_i^* = 7/40; F^*(16,0) = \sum_1^3 p_i^* = 14/40;$$

$$F^*(16,4) = \sum_1^4 p_i^* = 24/40; F^*(16,8) = \sum_1^5 p_i^* = 32/40;$$

$$F^*(17,2) = \sum_1^6 p_i^* = 35/40; F^*(17,6) = \sum_1^7 p_i^* = 38/40;$$

$$F^*(18,0) = \sum_1^8 p_i^* = 1.$$

Соединив полученные точки плавной кривой или отрезками прямой, получим график эмпирической функции распределения $F^*(x)$:



ТЕМА 2. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Пусть X – случайный признак, подчиненный закону распределения $F(x, \theta)$, где θ – параметр распределения, числовое значение которого неизвестно. Истинное значение этого параметра мы можем определить, исследовав все элементы генеральной совокупности, что не представляется возможным. По выборочным данным мы можем определить только приближенное значение параметра, которое называют значением оценки параметра. Задача статистического оценивания параметров состоит в выборе такой оценивающей функции, значение которой наиболее близко к оцениваемому параметру. Таким образом, оценка параметра – это однозначно определенная функция выборочных данных (оценочная функция), а значение оценки параметра – это значение этой функции для конкретной выборки. Оценка параметра θ обозначается $\hat{\theta}$.

Такая оценка, значение которой определяется одним числом, называется точечной. Оценка, значение которой определяется двумя числами (концами интервала), называется интервальной.

Точечная оценка считается наилучшей, если она несмещенная $M(\tilde{\theta}) = \theta$, состоятельная $P(\tilde{\theta} - \theta) < \varepsilon) \rightarrow 1$ и эффективная $D(\tilde{\theta}) \rightarrow \min$.

Наилучшей оценкой математического ожидания генеральной совокупности m_x является среднее арифметическое значение выборочных данных, полученных из этой генеральной совокупности, т.е.

$$\tilde{m}_x = \bar{x} = \begin{cases} \frac{1}{n} \sum_{i=1}^n x_i, & \text{если } x_1, x_2, \dots, x_n; \\ \frac{1}{n} \sum_{i=1}^k m_i x_i, & \text{если } \begin{matrix} x_i | x_1 \ x_2 \ \dots \ x_k \\ m_i | m_1 \ m_2 \ \dots \ m_k \end{matrix}, \sum_{i=1}^k m_i = n; \\ \frac{1}{n} \sum_{i=1}^k m_i \bar{x}_i, & \text{если } \begin{matrix} J_i | [x^{(1)}, x^{(2)}] \dots [x^{(k)}, x^{(k+1)}] \\ m_i | m_1 \ \dots \ m_k \end{matrix}. \end{cases}$$

Выборочная дисперсия определяется так:

$$D_B = \begin{cases} \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{m}_x)^2, & \text{если данные } x_1, x_2, \dots, x_n; \\ \frac{1}{n} \sum_{i=1}^k m_i (x_i - \tilde{m}_x)^2, & \text{если данные } \begin{matrix} x_i | x_1 \ x_2 \ \dots \ x_k \\ m_i | m_1 \ m_2 \ \dots \ m_k \end{matrix}; \\ \frac{1}{n} \sum_{i=1}^k m_i (\bar{x}_i - \tilde{m}_x)^2, & \text{если данные } \begin{matrix} J_i | [x^{(1)}, x^{(2)}] \dots [x^{(k)}, x^{(k+1)}] \\ m_i | m_1 \ \dots \ m_k \end{matrix}. \end{cases}$$

Выборочная дисперсия не является несмещенной оценкой дисперсии генеральной совокупности, поэтому в качестве оценки D_x берется исправленная выборочная дисперсия, т.е.

$$\tilde{D}_x = \frac{n}{n-1} D_v.$$

При $n \geq 50$ можно брать $\tilde{D}_x = D_v$.

Для определения точечных оценок параметров распределения обычно применяют метод моментов или максимального правдоподобия. Оценка параметров распределений:

нормального $\tilde{m} = \tilde{m}_x; \tilde{\delta} = \sqrt{\tilde{D}_x}$;

показательного $\tilde{\lambda} = 1/\tilde{m}_x$;

равномерного $\tilde{a} = \tilde{m}_x - \sqrt{3} \cdot \sqrt{\tilde{D}_x}; \tilde{b} = \tilde{m}_x + \sqrt{3} \cdot \sqrt{\tilde{D}_x}$.

Для определения интегральных оценок задаются доверительной вероятностью β (обычно принимают $\beta = 0,95$) и определяют доверительный интервал, т.е. интервал возможных значений параметра, который с заданной доверительной вероятностью содержит значение искомого параметра.

Если дисперсия генеральной совокупности D_x неизвестна, то длина интервала определяется числом $\varepsilon = t_{\beta,n} \frac{\tilde{\delta}_x}{\sqrt{n}}$, где $\tilde{\delta}_x = \sqrt{\tilde{D}_x}$; $t_{\beta,n}$ определяется по таблицам Стьюдента, т.е. $t_{\beta,n} = t_{\text{табл.двуст.}}(\alpha=1-\beta; f=n-1)$. Доверительный интервал в этом случае имеет вид $(\tilde{m}_x - \varepsilon; \tilde{m}_x + \varepsilon)$.

Доверительный интервал для D_x определяется неравенствами

$$\frac{(n-1)\tilde{D}_x}{x_2^2} < D_x < \frac{(n-1)\tilde{D}_x}{x_1^2}, \text{ где}$$

$$x_1^2 = x_{\text{табл}}^2(\alpha_1 = 1 - \frac{1-\beta}{2}; f_1 = n-1);$$

$$x_2^2 = x_{\text{табл}}^2(L_2 = \frac{1-\beta}{2}; f_2 = n-1).$$

Пример 2.1.

1. X – число работающих завалочных машин в цехе.

Запишем дискретный вариационный ряд (из задания №2):

x_i	0	1	2	3
m_i	1	4	3	2

Определим объем выработки: $n = \sum_{i=1}^k m_i = 10$.

Среднее число работающих завалочных машин в цехе

$$\bar{m}_x = \frac{\sum_{i=1}^k m_i x_i}{n} = \frac{1 \cdot 0 + 4 \cdot 1 + 3 \cdot 2 + 2 \cdot 3}{10} = 1,6.$$

Отклонение от среднего значения

$$\begin{aligned} \tilde{D}_x &= \frac{n}{n-1} D_B = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^k m_i (x_i - \bar{m}_x)^2}{n} = \frac{\sum_{i=1}^k m_i (x_i - \tilde{m}_x)^2}{n-1} = \\ &= \frac{1 \cdot (0 - 1,6)^2 + 4 \cdot (1 - 1,6)^2 + 3 \cdot (2 - 1,6)^2 + 2 \cdot (3 - 1,6)^2}{10 - 1} \approx 0,93; \end{aligned}$$

$$\bar{\delta}_x = \sqrt{\tilde{D}_x} = \sqrt{0,93} \approx 0,96.$$

2. X – производительность мартеновской печи.

Запишем интервальный вариационный ряд (из задания №2):

\tilde{x}_i	15	15.4	15.8	16.2	16.6	17	17.4	17.8
I_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	3	4	7	10	8	3	3	2

Определим середину каждого интервала и запишем над каждым интервалом.

Например,

$$\tilde{x}_1 = \frac{x^{(1)} + x^{(2)}}{2} = \frac{14,8 + 15,2}{2} = 15, \text{ или}$$

$$\tilde{x}_1 = x^{(1)} + \frac{\Delta x}{2} = 14,8 + \frac{0,4}{2} = 14,8 + 0,2 = 15.$$

Определим объем выработки: $\tau = \sum_{i=1}^k m_i = 40$.

Средняя производительность мартеновской печи:

$$\tilde{m}_x = \frac{\sum_{i=1}^k m_i \tilde{x}_i}{n} = \frac{3 \cdot 15 + 4 \cdot 15,4 + 7 \cdot 15,8 + 10 \cdot 16,2 + 8 \cdot 16,6 + 3 \cdot 17,4 + 2 \cdot 17}{40} = 16,27.$$

Отклонение от среднего значения:

$$\begin{aligned} \tilde{D}_x &= \frac{\sum_{i=1}^k m_i (\tilde{x}_i - \tilde{m}_x)^2}{n-1} = \frac{3 \cdot (15 - 16,27)^2 + 4 \cdot (15,4 - 16,27)^2 + 7 \cdot (15,8 - 16,27)^2 +}{39} \\ &+ \frac{10 \cdot (16,2 - 16,27)^2 + 8 \cdot (16,6 - 16,27)^2 + 3 \cdot (17,4 - 16,27)^2 + 2 \cdot (17,8 - 16,27)^2}{39} = \\ &= 0,52; \end{aligned}$$

$$\tilde{\delta}_x = \sqrt{\tilde{D}_x} = \sqrt{0,52} = 0,72.$$

Чтобы построить доверительный интервал для m_x , необходимо знать \tilde{m}_x и ε

$$\tilde{m}_x = 16,27 \text{ (из п.2),}$$

$$\begin{aligned} t_{\beta, n} &= t_{\text{табл. двуст.}}(\alpha = 1 - 0,95 = 0,05; f = 40 - 1 = 39) = 2,02, \\ \varepsilon &= 2,02 \cdot 0,72 / 40 = 0,23. \end{aligned}$$

Доверительный интервал для m_x :

$$(\tilde{m}_x - \varepsilon; \tilde{m}_x + \varepsilon);$$

$$(16,27 - 0,23; 16,27 + 0,23);$$

$$I_{m_x} = (16,04; 16,5).$$

Определим доверительный интервал для D_x :

$$\frac{(n-1)\tilde{D}_x}{x_2^2} < D_x < \frac{(n-1)\tilde{D}_x}{x_1^2},$$

$$x_1^2 = x_{\text{табл}}^2(\alpha_1 = 1 - \frac{1-0,95}{2} = 0,975; f_1 = 40 - 1 = 39) = 23,6;$$

$$x_2^2 = x_{\text{табл}}^2(\alpha_2 = \frac{1-0,95}{2} = 0,025; f_2 = 40 - 1 = 39) = 58,1;$$

$$\frac{(40-1) \cdot 0,52}{58,1} < D_x < \frac{(40-1) \cdot 0,52}{23,6};$$

$$0,33 < D_x < 0,86; I_{D_x} = (0,33; 0,86).$$

ТЕМА 3. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Некоторое предложение относительно свойств генеральной совокупности, из которой произведена выборка, называется *статистической гипотезой*.

Обычно она состоит в том, что одному или нескольким параметрам генеральной совокупности приписывается некоторое значение или делается предложение относительно вида закона распределения генеральной совокупности или различных ее свойств. Одна из гипотез выбирается в качестве исходной и обозначается H_0 . Например $H_0: m_x=0,5$. Здесь двоеточие аналогично союзу «что». Гипотеза, которая противоречит нулевой, называется конкурирующей (альтернативной) гипотезой и обозначает H_1 .

Путем статистической проверки необходимо установить, насколько данные, полученные из выборки, согласуются с высказанной гипотезой. Процедура проверки гипотезы – это правило, по которому гипотеза принимается или отвергается. Абсолютно надежное решение относительно проверяемой гипотезы получить нельзя. Необходимо заранее допустить возможность ошибочного решения, причем возможны ошибки двух родов: отвергнуть верную гипотезу (ошибка 1-го рода) и принять верную гипотезу (ошибка 2-го рода).

Вероятность ошибки первого рода обозначают α и называют уровнем значимости. Это достаточно малая вероятность, при которой в данной задаче событие можно считать практически невозможным. Величина α выбирается экспериментатором в зависимости от практических ситуаций. При решении большинства технических задач выбирается $\alpha=0,05$.

Процедура проверки H_0 заключается в следующем: выбирается некоторая подходящая функция выборочных данных $K(x_1, x_2, \dots, x_n; H_0)$, которая определяется выборкой и выдвинутой гипотезой H_0 и называется статистическим критерием. Функция $K(x_1, x_2, \dots, x_n; H_0)$ выбирается такой, чтобы закон ее распределения при справедливости H_0 был известен. Затем устанавливается область, вероятность попадания значения статистического критерия в которую в случае справедливости

гипотезы H_0 равна α , т.е. эти значения являются невозможными в однократном случае.

Эта область называется критической. Если вычисленное по выборке значение статистического критерия попадает в критическую область, гипотеза H_0 отвергается, в противном случае - нет оснований для отклонения гипотезы. Точки, отделяющие критическую область от области принятия гипотезы, называются критическими. Различают односторонние и двухсторонние критические области. Критерии для проверки гипотезы о виде закона распределения генеральной совокупности называется критерием согласия.

На практике чаще всего применяется критерий согласия χ^2 – Пирсона. В критерии согласия χ^2 – Пирсона за меру расхождения между теоретическим (предлагаемым) и эмпирическим законами распределения принимается сумма квадратов отклонения относительной частоты от соответствующей вероятности, взятых с определенным весом $c_i = n/p_i$ т.е.

$$\sum_{i=1}^K (P_i^* - P_i)^2 \cdot C_i = \sum_{i=1}^K \left(\frac{m_i}{n} - P_i\right)^2 \cdot \frac{n}{P_i} = \sum_{i=1}^K \frac{(m_i - np_i)^2}{np_i},$$

где

K – число интервалов;

$n = \sum_{i=1}^K m_i$ – объем выработки;

m_i – количество значений признака, попавших в i -й интервал

P_i вероятность попадания значений признака в i -й интервал

Эта случайная величина, если $m_i \geq 5$ подчиняется закону распределения χ^2 с числом степеней свободы $f = k - 1$ (постоянная связь $n = \sum_{i=1}^K m_i$).

Для конкретной выборки мы получим определенное (расчетное, наблюдение) значение критерия $\chi^2_{\text{набл}} = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i}$.

Если параметры предлагаемого распределения неизвестны, то они определяются по выборочным данным, а значит число степеней свободы уменьшится на количество параметров, т.е. $f = k - s - 1$, где S – количество

параметров, определяемых по выборочным данным. Критическое значение определяется по таблице χ^2 распределения:

$$\chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha; f=k-s-1).$$

Если $\chi^2_{\text{набл.}} > \chi^2_{\text{крит}}$ то гипотеза отвергается; если $\chi^2_{\text{набл.}} \leq \chi^2_{\text{крит}}$ то делаем вывод, что данные противоречат гипотезе.

Для вычисления $\chi^2_{\text{набл}}$ обычно используют следующую таблицу:

J_i		m_i	p_i	np_i	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
$\chi^{(i)}$	$\chi^{(i+1)}$					
		$\sum_{i=1}^K m_i = n$	$\sum_{i=1}^K p_i = 1$	$\chi^2_{\text{набл}} = \Sigma$		

Здесь $p_i = p(\chi^{(i)} < \chi < \chi^{(i+1)})$ вычисляется по определенным формулам в зависимости от вида предлагаемого распределения.

Итак, проверку гипотезы с помощью критерия χ^2 – Пирсона производят следующим образом:

1. Строится статистический ряд распределения.
2. Статистический ряд изображается графически в виде гистограммы или полигона.
3. По виду эмпирического распределения выдвигается гипотеза о виде закона распределения генеральной совокупности.
4. Определяются оценки параметров предлагаемого закона распределения.
5. Вычисляется $\chi^2_{\text{набл.}}$.
6. Задаем уровень значимости α и вычисляем число степеней свободы $f=k-s-1$.
7. По таблице χ^2 – распределения находим $\chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha; f=k-s-1)$
8. Сравниваем $\chi^2_{\text{набл.}}$ и $\chi^2_{\text{крит}}$ делаем выводы о принятии или непринятии H_0 .

Рассмотрим применение критерия χ^2 - Пирсона.

При уровне значимости $\alpha=0,05$ высказывать гипотезу о виде закона распределения производительности мартеновской печи и проверять ее с помощью критерия χ^2 – Пирсона (использовать интервальный статистический ряд, построенный ранее).

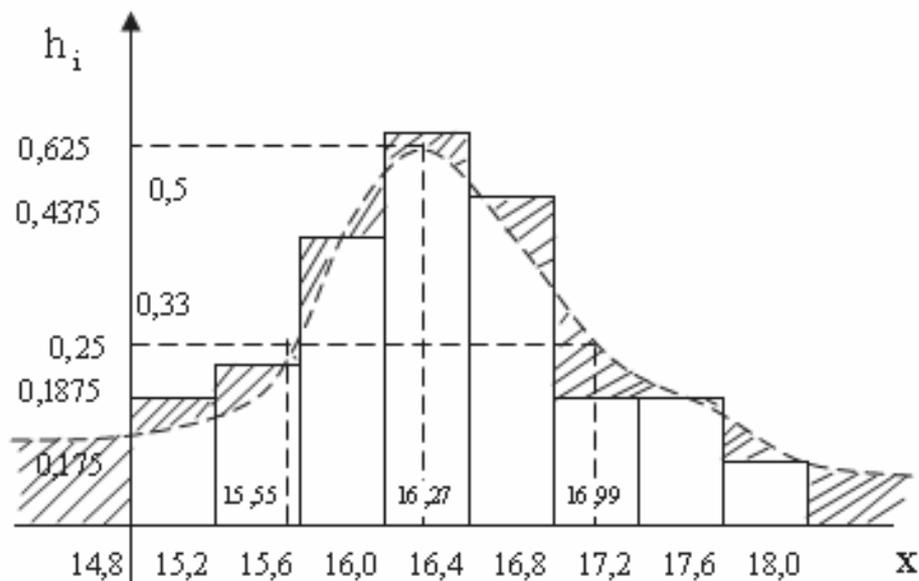
Пример 3.1.

1) запишем интервальный статистический ряд, дополнив строкой

$h_i = \frac{P_i^*}{\Delta x}$, которая понадобится при построении гистограммы:

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	3	4	7	10	8	3	3	2
P_i^*	0,075	0,1	0,175	0,25	0,2	0,075	0,075	0,05
$\frac{P_i^*}{\Delta x}$	0,1875	0,25	0,4375	0,625	0,5	0,1875	0,1875	0,175

2) строим гистограмму:



3) по виду гистограммы можно высказать H_0 : генеральная совокупность имеет нормальное распределение;

4) нормальное распределение имеет два параметра δ и m .

Их оценки:

$$\tilde{m} = \tilde{m}_x = \frac{1}{n} \sum_{i=1}^k m_i \tilde{x}_i = 16,27;$$

$$\tilde{\delta} = \tilde{\delta}_x = \sqrt{\tilde{D}_x} = \sqrt{\frac{1}{n-1} \sum_{i=1}^k m_i (\tilde{x}_i - \tilde{m}_x)^2} = \sqrt{0,52} = 0,72.$$

Зная оценки параметров находим точку (16,27; 0,56) и точки перегиба (15,55; 0,33), (16,99; 0,33).

5) $\chi^2_{\text{набл.}}$ определяется заштрихованной площадью (чем она больше, тем больше расхождение). Для вычисления $\chi^2_{\text{набл.}}$ составим таблицу объединив интервалы с $m_i < 5$.

Таблица 1

n=40

J _i		m _i	p _i	np _i	(m _i -np _i) ²	$\frac{(m_i - np_i)^2}{np_i}$
$\chi^{(i)}$	$\chi^{(i+1)}$					
14,8	15,6	7	0,1162	4,648	5,5319	1,1902
15,6	16,0	7	0,1983	7,932	0,8686	0,1095
16,0	16,4	10	0,1969	7,876	4,5114	0,5128
16,4	17,2	11	0,3301	13,204	4,8576	0,3679
17,2	18,0	5	0,0985	3,94	1,1236	0,2852

$$\sum p_i = 0,94$$

$$\chi^2_{\text{набл.}} = 2,4656$$

Вероятность p_i вычисляем по формуле:

$$P_i = \Phi\left(\frac{\chi^{(i+1)} - \tilde{m}}{\tilde{\delta}}\right) - \Phi\left(\frac{\chi^{(i)} - \tilde{m}}{\tilde{\delta}}\right).$$

Причем так как для нормального распределения $-\infty < \chi < \infty$ то $\chi^{(1)} = -\infty$, а $\chi^{(6)} = +\infty$,

$$\begin{aligned} P_1 &= \Phi\left(\frac{15,6-16,27}{0,72}\right) - \Phi\left(\frac{-\infty-16,27}{0,72}\right) = \Phi(-0,93) - \Phi(-\infty) = -\Phi(0,93) + \Phi(\infty) = \\ &= -0,3238 + 0,5 = 0,1162; \end{aligned}$$

Значение $\Phi(x)$ определяем по таблице 1.

$$P_2 = \Phi\left(\frac{16,0-16,27}{0,72}\right) - \Phi\left(\frac{15,6-16,27}{0,72}\right) = \Phi(-0,32) - \Phi(-0,93) = -\Phi(0,32) + \Phi(0,93) =$$

$$= -0,1255 + 0,3238 = 0,1983;$$

$$P_3 = \Phi\left(\frac{16,4-16,27}{0,72}\right) - \Phi\left(\frac{16,0-16,27}{0,72}\right) = \Phi(0,18) - \Phi(-0,32) = \Phi(0,18) + \Phi(0,32) =$$

$$= 0,0714 + 0,1255 = 0,1969;$$

$$P_4 = \Phi\left(\frac{17,2-16,27}{0,72}\right) - \Phi\left(\frac{16,4-16,27}{0,72}\right) = \Phi(1,29) - \Phi(0,18) = 0,4015 - 0,0714 = 0,3301;$$

$$P_5 = \Phi\left(\frac{\infty-16,27}{0,72}\right) - \Phi\left(\frac{17,2-16,27}{0,72}\right) = \Phi(\infty) - \Phi(1,29) = 0,5 - 0,4015 = 0,0985.$$

б) полагаем $\alpha=0,05$. Число степеней свободы $f=k-s-1=5-2-1=2$, так как осталось после объединения 5 интервалов, а число параметров $s=2$ (m и δ).

$$7) \chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha=0,05; f=2) = 6,0;$$

$$(\chi^2_{\text{набл}} = 2,4656) < (\chi^2_{\text{крит}} = 6,0).$$

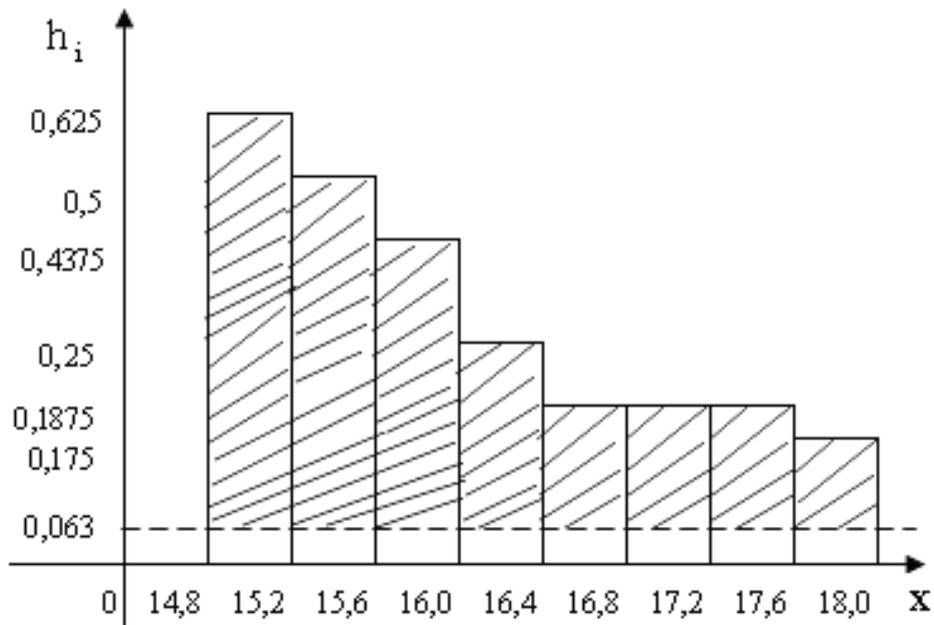
Вывод: выборочные данные не противоречат гипотезе о нормальном распределении производительности данной мартеновской печи.

Пример 3.2.

1) пусть исходные данные имеют вид

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	10	8	7	4	3	3	3	2
P_i^*	0.25	0.2	0.175	0.1	0.075	0.075	0.075	0.05
$\frac{P_i^*}{\Delta x}$	0.625	0.5	0.4375	0.25	0.1875	0.1875	0.1875	0.175

2) построим гистограмму



3) по виду гистограммы можно высказать H_0 : генеральная совокупность имеет показательный закон распределения;

4) показательное распределение имеет один параметр λ

Его оценка

$$\tilde{\lambda} = \frac{1}{\tilde{m}_x} = \frac{1}{\frac{1}{n} \sum_{i=1}^k m_i x_i} = \frac{n}{\sum_{i=1}^k m_i x_i} =$$

$$= \frac{40}{10 \cdot 15 + 8 \cdot 15,4 + 7 \cdot 15,8 + 4 \cdot 16,2 + 3 \cdot 16,6 + 3 \cdot 17 + 3 \cdot 17,4 + 2 \cdot 17,8} = \frac{40}{631,2} = 0,063.$$

5) зная λ построим теоретическую кривую распределения; заштрихуем расхождения между эмпирическим законом распределения и предполагаемым теоретическим. Для вычисления $\chi^2_{\text{набл}}$ составим таблицу, объединив интервалы с $m_i < 5$.

J _i		m _i	p _i	np _i	(m _i -np _i) ²	$\frac{(m_i - np_i)^2}{np_i}$
$\chi^{(i)}$	$\chi^{(i+1)}$					
14,8	15,2	10	0,6162	24,6480	214,5639	8,1051
15,2	15,6	8	0,0095	0,3800	58,0644	152,8010
15,6	16,4	11	0,0184	0,7360	109,3496	143,1381
16,4	17,2	6	0,0115	0,4600	30,6916	66,7209
17,2	18,0	5	0,3384	13,5360	72,8633	5,3829

$$\sum p_i = 0,994$$

$$\chi^2_{\text{набл}} = 376,148$$

Вероятность $p_i = P(\chi^{(i)} < \chi < \chi^{(i+1)})$ вычисляем по формуле:

$$p_i = e^{-\lambda \chi^{(i)}} - e^{-\lambda \chi^{(i+1)}},$$

причем, так как $0 < \chi < \infty$, то $\chi^{(1)} = 0$; $\chi^{(6)} = \infty$.

Итак:

$$P_1 = e^{-0,063 \cdot 0} - e^{-0,063 \cdot 15,2} = 1 - 0,3838 = 0,6162;$$

$$P_2 = e^{-0,063 \cdot 15,2} - e^{-0,063 \cdot 15,6} = 0,3838 - 0,3743 = 0,0095$$

$$P_3 = e^{-0,063 \cdot 15,6} - e^{-0,063 \cdot 16,4} = 0,3743 - 0,3559 = 0,0184;$$

$$P_4 = e^{-0,063 \cdot 16,4} - e^{-0,063 \cdot 17,2} = 0,3559 - 0,3384 = 0,0115;$$

$$P_5 = e^{-0,063 \cdot 17,2} - e^{-0,063 \cdot \infty} = 0,3384.$$

6) полагаем $\alpha = 0,05$. Число степеней свободы $f = k - s - 1 = 5 - 1 - 1 = 3$, так как осталось после объединения $k = 5$ интервалов, а число параметров $s = 1$ (λ).

$$7) \chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha = 0,05; F = 3) = 7,8;$$

$$(\chi^2_{\text{набл}} = 376,148) > (\chi^2_{\text{крит}} = 7,8).$$

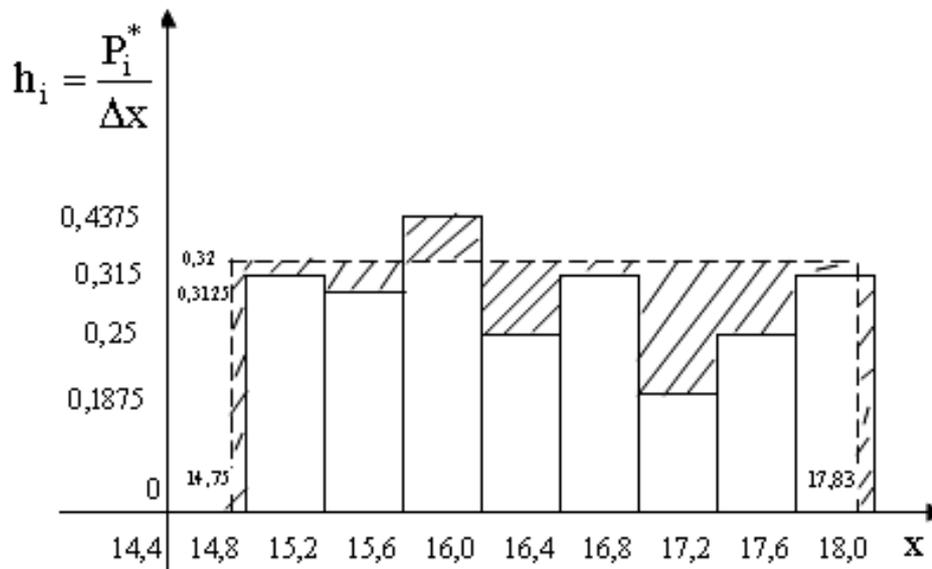
Вывод: гипотеза о показательном распределении производительности данной мартеновской печи отвергается. Очевидно, здесь более подходящим было бы β – распределение.

Пример 3.3.

1) пусть исходные данные имеют вид

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	5	5	7	4	6	3	4	5
P_i^*	0,15	0,125	0,175	0,1	0,15	0,075	0,1	0,125
$\frac{P_i^*}{\Delta x}$	0,315	0,3125	0,4375	0,25	0,315	0,1875	0,25	0,3125

2) построим гистограмму



3) по виду гистограммы можно высказать H_0 : генеральная совокупность имеет равномерный закон распределения.

4) равномерный закон распределения имеет два параметра a и b . Их оценки:

$$\bar{a} = \bar{m}_x - \sqrt{3} \times \sqrt{\bar{D}_x},$$

$$\bar{b} = \bar{m}_x + \sqrt{3} \times \sqrt{\bar{D}_x}.$$

Если \tilde{m}_x и \tilde{D}_x неизвестны, то их вычисляют по формулам:

$$\tilde{m}_x = \frac{1}{n} \sum_{i=1}^k m_i \tilde{x}_x = \frac{6 \cdot 15 + 5 \cdot 15,4 + 7 \cdot 15,8 + 4 \cdot 16,2 + 6 \cdot 16,6 + 3 \cdot 17 + 4 \cdot 17,4 + 5 \cdot 17,8}{40} =$$

$$= \frac{651}{40} = 16,29;$$

$$\tilde{D}_x = \frac{1}{n-1} \sum_{i=1}^k m_i (\tilde{x}_i - \tilde{m}_i)^2 = \frac{31,016}{39} = 0,7953;$$

$$\tilde{\delta}_x = \sqrt{\tilde{D}_x} = \sqrt{0,7953} = 0,8918;$$

$$\tilde{a} = 16,29 - \sqrt{3} \cdot 0,8918 = 14,75;$$

$$\tilde{b} = 16,29 + \sqrt{3} \cdot 0,8918 = 17,83.$$

5) построим кривую равномерного распределения и заштрихуем расхождения между эмпирическим законом распределения и теоретическим. Кривая равномерного распределения представляет собой прямую проведенную на высоте $\frac{1}{\tilde{b} - \tilde{a}} = \frac{1}{17,83 - 14,75} = 0,32$ от точки

$\tilde{a} = 14,75$ до точки $\tilde{b} = 17,83$.

Для вычисления $\chi^2_{\text{набл}}$ составим таблицу, объединив интервалы с $m_i < 5$:

J_i		m_i	p_i	np_i	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
$\chi^{(i)}$	$\chi^{(i+1)}$					
14,8	15,2	6	0,1461	5,844	0,0243	
15,2	15,6	5	0,1298	5,192	0,0388	0,0071
15,6	16,0	7	0,1298	5,192	3,2689	0,6296
16,0	16,8	10	0,2596	10,384	0,1475	0,0142
16,8	17,6	7	0,2596	10,384	11,4515	1,1028
17,6	18,0	5	0,0747	2,987	4,0522	1,3566

$$\sum p_i = 0,9937$$

$$\chi^2_{\text{набл}} = 3,1103$$

Вероятности $p_i = P(\chi^{(i)} < \chi < \chi^{(i+1)})$ вычисляем по формуле:

$$p_i = \frac{\chi^{(i+1)} - \chi^{(i)}}{\tilde{b} - \tilde{a}},$$

причем, так как $\tilde{a} \leq \chi \leq \tilde{b}$, то $\chi^{(1)} = \tilde{a}$; $\chi^{(6)} = \tilde{b}$.

Итак,

$$P_1 = \frac{x^{(2)} - \tilde{a}}{\tilde{b} - \tilde{a}} = \frac{15,2 - 14,75}{17,83 - 14,75} = 0,1461;$$

$$P_2 = P_3 = \frac{0,8}{17,83 - 14,75} = \frac{0,8}{3,08} = 0,1298;$$

$$P_4 = P_5 = \frac{0,8}{17,83 - 14,75} = \frac{0,8}{3,08} = 0,2596;$$

$$P_6 = \frac{\tilde{b} - X^{(6)}}{\tilde{b} - \tilde{a}} = \frac{17,83 - 17,6}{17,83 - 14,75} = 0,0747.$$

б) полагаем $\alpha=0,05$. Число степеней свободы $f=k-s-1=6-2-1=3$, так как осталось после объединения $k=6$ интервалов а число параметров $s=2$ (а и б).

$$7) \chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha=0,05; f=3)=7,8;$$

$$(\chi^2_{\text{набл}}=3,1103) < (\chi^2_{\text{крит}}=7,8).$$

Вывод: выборочные данные не противоречат гипотезе о равномерном распределении производительности данной мартеновской печи.

ТЕМА 4. ДИСПЕРСИОННЫЙ АНАЛИЗ

Начиная исследование плохо изученной сложной системы, следует выделить существенные факторы из большого числа независимых факторов, сократив тем самым общее количество исследуемых факторов.

Так, например, при выполнении на автоматической линии некоторой операции обработки параллельно на нескольких станках важно для правильного построения последующей обработки знать, являются ли однотипным средние размеры деталей, т.е. определить влияние одного фактора – станка.

Если обработка ведется несколькими операторами, то нас интересует влияние двух факторов – станка и оператора.

Научно обоснованное решение подобных задач при некоторых предположениях составляет предмет дисперсионного анализа.

Основоположником дисперсионного анализа считают известного английского математика Р.Фишера (1900 – 1962г.г.), который ввел в статистику сами термины «дисперсия» и «дисперсионный анализ».

Дисперсионный анализ – это статистический метод анализа результатов, зависящих от нескольких одновременно действующих факторов, определение значимости влияния этих факторов и выбор наиболее важных из них.

В зависимости от количества факторов, включенных в анализ, различают однофакторный, двухфакторный и многофакторный анализы.

Идея дисперсионного анализа заключается в разложении суммы квадратов отклонений от среднего для всей выборки ($S_{\text{общ}}$) на составляющие, соответствующие влиянию изучаемых факторов ($S_{\text{факт}}$) и влиянию неучтенных факторов ($S_{\text{ост}}$), т.е.

$$S_{\text{общ}} = S_{\text{факт}} + S_{\text{ост}},$$

это равенство называют основным уравнением дисперсионного анализа. Для однофакторного дисперсионного анализа

$$S_{\text{общ}} = S_a + S_{\text{ост}},$$

для двухфакторного

$$S_{\text{общ}} = S_A + S_B + S_{\text{ост}},$$

где S_{AB} – сумма квадратов отклонений, вызванных взаимодействием факторов А и В.

Предпосылки дисперсионного анализа: результаты наблюдений должны быть независимыми случайными величинами, имеющими нормальное распределение, что позволяет оценить значимость оценок дисперсий; дисперсия генеральных совокупностей должны быть равны (хотя и не известны нам), что гарантирует воспроизводимость опытов с равной точностью, а это дает возможность правильно рассчитывать оценку остаточной дисперсии. Только при выполнении этих условий применим аппарат дисперсионного анализа.

Рассмотрим двухфакторный дисперсионный анализ, который чаще всего применяется на практике.

Для проведения анализа необходимо иметь результаты наблюдений над изучаемым признаком, которые представляются в виде таблицы наблюдений.

A	A₁	A₂	...	A_k	\bar{y}_{*j}
B					
B₁	y_{11}	y_{21}	...	y_{k1}	\bar{y}_{*1}
B₂	y_{12}	y_{22}	...	y_{k2}	\bar{y}_{*2}
...
B_n	y_{1n}	y_{2n}	...	y_{kn}	\bar{y}_{*n}
\bar{y}_{i*}	\bar{y}_{1*}	\bar{y}_{2*}	...	\bar{y}_{k*}	\bar{y}

Здесь y_{ij} – значение признака,

A – фактор с k уровнями: A_1, A_2, \dots, A_k ,

B – фактор с n уровнями: B_1, B_2, \dots, B_n ,

$$\tilde{y}_{*j} = \frac{1}{k} \sum_{i=1}^k y_{ij} \text{ – среднее значение по строкам,}$$

$$\tilde{y}_{i*} = \frac{1}{n} \sum_{j=1}^n y_{ij} \text{ – среднее значение по столбцам,}$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n \bar{y}_{*j} = \frac{1}{k} \sum_{i=1}^k \bar{y}_{i*} = \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n y_{ij} \text{ – общее среднее значение.}$$

Очевидно что \bar{y}_{*1} является оценкой математического ожидания т.е.

$$\tilde{m}_1^B = \bar{y}_{*1}; \tilde{m}_2^B = \bar{y}_{*2}; \dots; \tilde{m}_n^B = \bar{y}_{*n}.$$

$$\text{Аналогично } \tilde{m}_1^A = \bar{y}_{1*}; \tilde{m}_2^A = \bar{y}_{2*}; \dots; \tilde{m}_k^A = \bar{y}_{k*}.$$

Необходимо проверить гипотезу о равенстве математических ожиданий генеральных совокупностей т.е для фактора A:

$$H_0^A : m_1^A = m_2^A = \dots = m_k^A; H_1^A : m_1^A \neq m_2^A \dots \neq m_k^A.$$

Для фактора B:

$$H_0^B : m_1^B = m_2^B = \dots = m_n^B; H_1^B : m_1^B \neq m_2^B \neq \dots \neq m_n^B.$$

Эти гипотезы мы проверяем по известным оценкам \tilde{m}_i .

Непосредственная проверка этих гипотез не рациональна. Р.Фишер

предложил вместо этих гипотез проверять однородность оценок дисперсий а именно: разброс значений \tilde{m}_i^A ($i = 1, 2, \dots, k$) можно измерить с помощью факторной дисперсии которую обозначают \tilde{D}_A разброс значений \tilde{m}_j^B ($j = 1, 2, \dots, n$) можно измерить с помощью факторной дисперсии которую обозначают \tilde{D}_B . Чем сильнее различаются между собой средние величины тем больше факторная дисперсия. Кроме влияния факторов на выходную величину влияют случайные причины (ошибки опыта). Разброс значений за счет случайных причин измеряется дисперсией, которая называется остаточной и обозначается $\tilde{D}_{\text{ост}}$.

Итак нулевые и противоположные гипотезы для факторов А и В можно записать так:

$$H_0^A : m_1^A = m_2^A = \dots = m_k^A \Leftrightarrow D_A = D_{\text{ост}} - \text{фактор А влияет незначимо};$$

$$H_1^A : m_1^A \neq m_2^A \neq \dots \neq m_k^A \Leftrightarrow D_A > D_{\text{ост}} - \text{фактор А влияет значимо};$$

$$H_0^B : m_1^B = m_2^B = \dots = m_n^B \Leftrightarrow D_B = D_{\text{ост}} - \text{фактор В влияет незначимо};$$

$$H_1^B : m_1^B \neq m_2^B \neq \dots \neq m_n^B \Leftrightarrow D_B > D_{\text{ост}} - \text{фактор В влияет значимо}.$$

Очевидно, что легче проверять однородность дисперсий чем большого числа (равного числу уравнений фактора математических ожиданий).

Проверим гипотезы:

$$H_0^A : D_A = D_{\text{ост}} ; H_0^B : D_B = D_{\text{ост}} ,$$

$$H_1^A : D_A > D_{\text{ост}} ; H_1^B : D_B > D_{\text{ост}} .$$

Однородность дисперсий проверяется по критерию Фишера:

$$F_{\text{набл}}^A = \frac{\tilde{D}_A}{\tilde{D}_{\text{ост}}} ; F_{\text{набл}}^B = \frac{\tilde{D}_B}{\tilde{D}_{\text{ост}}} ,$$

где

$$\tilde{D}_A = \frac{n \sum_{i=1}^k (\bar{y}_{i^*} - \bar{y})^2}{k-1}; \tilde{D}_B = \frac{k \sum_{j=1}^n (\bar{y}_{*j} - \bar{y})^2}{n-1};$$

$$\tilde{D}_{ост} = \frac{\sum_{i=1}^n \sum_{j=1}^k (y_{ij} - \bar{y}_{i^*} - \bar{y}_{*j} + \bar{y})^2}{(k-1)(n-1)}.$$

Критическое значение определяется по таблице Фишера:

$$F_{крит}^A = F_{табл}^A(\alpha; f_1=k-1; f_2=(k-1)(n-1));$$

$$F_{крит}^B = F_{табл}^B(\alpha; f_1=n-1; f_2=(k-1)(n-1)).$$

Если $F_{набл} \geq F_{крит}$, дисперсии неоднородны, а это значит, что фактор влияет значимо.

Критерий Фишера применяется, если $\tilde{D}_{факт} > \tilde{D}_{ост}$; если $\tilde{D}_{факт} < \tilde{D}_{ост}$ то без проверки делается вывод о незначимости влияния факторов.

Пример 4.1.

Y – производительность труда;

A – концентрация производства;

B – тип оборудования.

Таблица наблюдений:

B	B ₁	B ₂	\bar{y}_{i^*}
A			
A ₁	25	30	27,5
A ₂	50	60	55
A ₃	80	90	85
\bar{y}_{*j}	51,7	60	$\bar{y} = 55,8$

При уровне значимости $\alpha=0,05$ проверить значимо ли влияние факторов A и B на выходную величину Y, т.е. проверить гипотезы:

$$H_0^A : m_1^A = m_2^A = m_3^A \Leftrightarrow D_A = D_{ост} \Leftrightarrow \text{фактор A влияет незначимо};$$

$$H_1^A : m_1^A \neq m_2^A \neq m_3^A \Leftrightarrow D_A > D_{ост} \Leftrightarrow \text{фактор A влияет значимо};$$

$H_0^B : m_1^B = m_2^B \Leftrightarrow D_B = D_{ост} \Leftrightarrow$ фактор В влияет незначимо;

$H_1^B : m_1^B \neq m_2^B \Leftrightarrow D_B > D_{ост} \Leftrightarrow$ фактор В влияет значимо;

Вычислим оценки дисперсий

$$\tilde{D}_A = \frac{n \sum_{i=1}^k (\bar{y}_{i*} - \bar{y})^2}{k-1} = \frac{2[(27,5 - 55,8)^2 + (55,8 - 55,8)^2 + (85 - 55,8)^2]}{3-1} = 1654,2;$$

$$\tilde{D}_B = \frac{k \sum_{j=1}^n (\bar{y}_{*j} - \bar{y})^2}{n-1} = \frac{3[(51,7 - 55,8)^2 + (60 - 55,8)^2]}{2-1} = 103,35;$$

$$\tilde{D}_{ост} = \frac{\sum_{i=1}^n \sum_{j=1}^k (y_{ij} - \bar{y}_{i*} - \bar{y}_{*j} + \bar{y})^2}{(k-1)(n-1)} = \frac{1}{(3-1)(2-1)} [(25 - 27,5 - 57,7 + 55,8)^2 +$$

$$+ (30 - 27,5 - 60 + 55,8)^2 + (50 - 55 - 57,7 + 55,8)^2 + (60 - 55 - 60 + 55,8)^2 +$$

$$+ (50 - 55 - 57,7 + 55,8)^2 + (60 - 55 - 60 + 55,8)^2 + (80 - 85 - 57,7 + 55,8)^2 +$$

$$+ (90 - 85 - 60 + 55,8)^2] = \frac{8,35}{2} \approx 4,2;$$

$$F_{набл}^A = \frac{\tilde{D}_A}{\tilde{D}_{ост}} = \frac{1654,2}{4,2} = 393,9;$$

$$F_{набл}^B = \frac{\tilde{D}_B}{\tilde{D}_{ост}} = \frac{103,35}{4,2} = 24,6;$$

$$F_{крит}^A = F_{табл}^A(\alpha=0,05; f_1=k-1=3-1=2; f_2=(k-1)(n-1)=2 \cdot 1=2)=19;$$

$$F_{крит}^B = F_{табл}^B(\alpha=0,05; f_1=n-1=2-1=1; f_2=(k-1)(n-1)=2)=18,51;$$

$(F_{набл}^A=393,9) > (F_{крит}^A=19) \Leftrightarrow$ фактор А влияет значимо;

$(F_{набл}^B=24,6) > (F_{крит}^B=18,51) \Leftrightarrow$ фактор В влияет значимо.

Вывод: на производительность труда в этой группе предприятий значимо влияет и концентрация производства и тип оборудования.

ТЕМА 5. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Корреляционный анализ базируется на определенной модели исследуемого явления, согласно которой все переменные, описывающие явление, представляют собой случайные величины и имеют совместное (двумерное или многомерное) нормальное распределение.

Если же окажется, что фактическое положение вещей не соответствует такой модели, то результаты корреляционного анализа теряют свою достоверность, которая, очевидно, будет тем выше, чем лучше исследуемое явление соответствует модели.

При нарушении предположения о совместном нормальном распределении переменных регрессия их не будет являться линейной функцией, т.е. корреляционная связь между переменными будет нелинейной, и, следовательно, коэффициент корреляции не будет являться достаточной характеристикой корреляционной связи.

Предположение о случайности переменных, в противоположность предыдущему, является не строгим, так как наличие ошибок при измерении переменных в процессе экспериментирования, а также влияние неконтролируемых и неучитываемых факторов всегда делает их «частично случайными» даже при неслучайности их природы.

Корреляционный анализ решает следующие задачи:

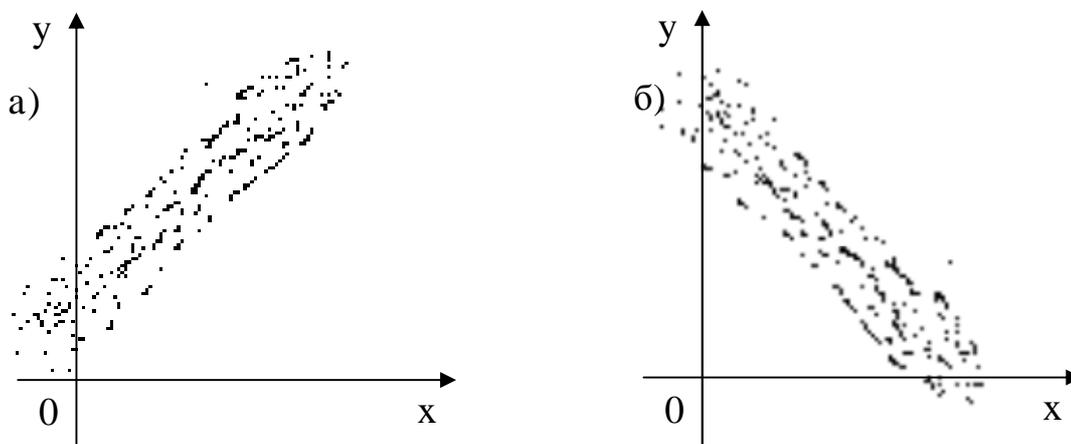
- установление существования корреляционной зависимости между переменными;
- определение величины степени связности (тесноты связи);
- отбор факторов, оказывающих наиболее существенное влияние на результативный признак (на зависимую переменную).

В дополнение к этим задачам, которые являются основными, решается также задача нахождения эмпирических уравнений регрессии, которые используются для прогнозирования средних значений одних величин по заданным значениям других.

Рассмотрим различные виды корреляции.

Относительно характера корреляции различают положительную корреляцию и отрицательную. При положительной корреляции с

увеличением переменной другая соответственно увеличивается, а при уменьшении уменьшается (а); при отрицательной с увеличением одной переменной значения другой уменьшаются, а при уменьшении увеличиваются (б).



Относительно числа переменных различают простую, или парную, корреляцию и множественную. **Простая корреляция** – это корреляция между двумя переменными. Например, между доходом и потреблением, между прибылью и себестоимостью и т.д.

Множественная корреляция – это корреляция между более чем двумя переменными. Например, между производительностью труда, уровнем механизации производства, квалификацией рабочих, уровнем использования машинного времени и др.

Относительно формы связи различают линейную корреляцию и нелинейную. При линейной корреляции между исследуемыми переменными существуют линейные соотношения, при нелинейной – нелинейные.

Рассмотрим линейную корреляционную зависимость между двумя переменными X и Y , т.е. простую линейную корреляцию. Результаты наблюдения записываем в виде

$$(x_1; y_1); (x_2; y_2); \dots; (x_n; y_n).$$

Если число наблюдений большое, то исходные данные группируют и представляют в виде дискретной корреляционной таблицы (а) или в виде интервальной корреляционной таблицы (б)

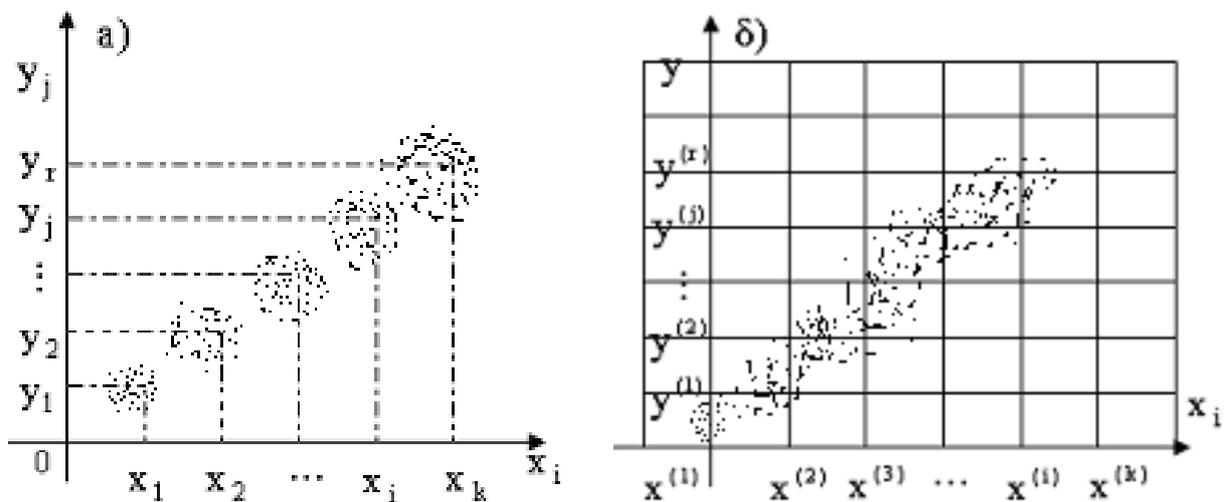
а)	x_i	x_1	x_2	...	x_k	n_y
	y_i	n_{11}	n_{21}	...	n_{k1}	n_{y1}
	y_1	n_{11}	n_{21}	...	n_{k1}	n_{y1}
	y_2	n_{12}	n_{22}	...	n_{k2}	n_{y2}

	y_r	n_{1r}	n_{2r}	...	n_{kr}	n_{yr}
n_x	n_{x1}	n_{x2}	...	n_{xk}	n	

б)	x	$x^{(1)};x^{(2)}$	$x^{(2)};x^{(3)}$...	$x^{(k-1)};x^{(k)}$	n_y
	y	$y^{(1)};y^{(2)}$	$y^{(2)};y^{(3)}$...	$y^{(r-1)};y^{(r)}$	$n_{\bar{y}1}$
	$y^{(1)};y^{(2)}$	n_{11}	n_{21}	...	n_{k1}	$n_{\bar{y}1}$
	$y^{(2)};y^{(3)}$	n_{12}	n_{22}	...	n_{k2}	$n_{\bar{y}2}$

	$y^{(r-1)};y^{(r)}$	n_{1r}	n_{2r}	...	n_{kr}	$n_{\bar{y}r}$
n_x	$n_{\bar{x}1}$	$n_{\bar{x}2}$...	$n_{\bar{x}k}$	n	

Корреляционная таблица может быть изображена графически в виде **корреляционного поля**, которое даёт представление о виде и тесноте корреляционной зависимости.



По данным корреляционной таблицы могут быть построены линии регрессии. Для их построения определяем условные средние \bar{y}_x и \bar{x}_y и строим ломаные, соединяющие для эмпирической линии регрессии Y на X точки с координатами $(x_j; \bar{y}_{xj})$, а X на Y - $(y_j; \bar{x}_{yj})$.

Построенные линии регрессии содержат элемент случайности ввиду ограниченности объёма выборки. Поэтому эту эмпирическую линию регрессии сглаживают такой линией, которая наилучшим образом описывала бы распределение точек.

Такая линия называется *теоретической линией регрессии*. Прежде чем её строить, необходимо убедиться в достаточно тесной связи между точками. Аналитически тесноту линейной корреляционной зависимости характеризует коэффициент корреляции.

$$r_{xy} = \frac{K_{xy}}{\delta_x \cdot \delta_y} = \frac{M[(X - m_x)(Y - m_y)]}{\delta_x \cdot \delta_y} = \frac{M[XY] - m_x \cdot m_y}{\delta_x \cdot \delta_y}.$$

По выборочным данным определяем оценку коэффициента корреляции, которая называется коэффициентом корреляции:

$$r_B = \tilde{r}_{xy} = \frac{\tilde{M}[XY] - \tilde{m}_x \cdot \tilde{m}_y}{\tilde{\delta}_x \cdot \tilde{\delta}_y}.$$

Значимость выборочного коэффициента корреляции проверяется по критерию Стьюдента. Если он значим ($r_{xy} \neq 0$), то переходим к построению теоретических линий регрессии.

При простой линейной корреляции линии регрессии описываются уравнениями:

$$\bar{y}_x = ax + b;$$

$$\bar{x}_y = cy + d.$$

Коэффициенты a, b, c, d определяются чаще всего по методу наименьших квадратов. По этому методу коэффициенты определяются таким образом, чтобы сумма квадратов отклонений эмпирических (опытных) точек от теоретических (лежащих на прямой) была минимальной, т.е.

$$S(a, b) = \sum_{i=1}^n (y_i - \bar{y}_{x_i})^2 = \sum_{i=1}^n (y_i - ax_i - b)^2 \rightarrow \min.$$

Из математического анализа известно, что для этого нужно составить систему:

$$\begin{cases} \frac{dS}{da} = 0; \\ \frac{dS}{db} = 0. \end{cases}$$

Решая эту систему уравнений, получим:

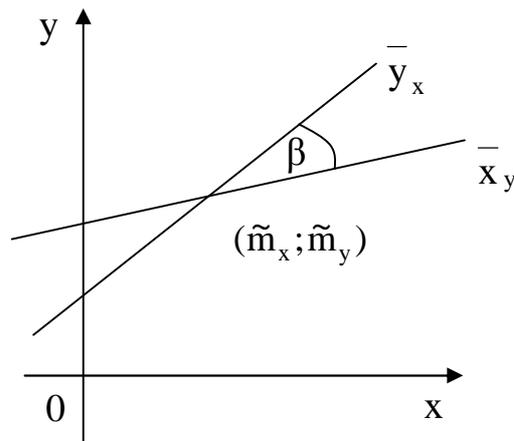
$$a = r_B \frac{\tilde{\delta}_y}{\tilde{\delta}_x}; \quad b = \tilde{m}_y - r_B \frac{\tilde{\delta}_y}{\tilde{\delta}_x} \tilde{m}_x.$$

Поставим эти значения в уравнения регрессии

$$\bar{y}_x = ax + b = \tilde{m}_y - r_B \frac{\tilde{\sigma}_y}{\tilde{\sigma}_x} (x - \tilde{m}_x);$$

$$\bar{x}_y = cy + d = \tilde{m}_x - r_B \frac{\tilde{\sigma}_x}{\tilde{\sigma}_y} (y - \tilde{m}_y);$$

Линии регрессии пересекаются в точке с координатами $(\tilde{m}_x; \tilde{m}_y)$



Обозначим через β угол, образованный обеими линиями регрессии.

Если $r_{xy}=1$ или $r_{xy}=-1$, то $\beta=0$, т.е. обе прямые сливаются в одну (функциональная зависимость). При $r_{xy}=0$, угол $\beta=90^\circ$, т.е. две прямые оказываются взаимно перпендикулярными. Это означает отсутствие линейной корреляционной зависимости. Во всех остальных случаях прямые образуют угол между 0° и 90° . Чем меньше этот угол, тем сильнее линейная корреляционная зависимость.

Пример 5.1.

x – скорость завалки лома на данной металлургической печи;

y – производительность мартеновской печи.

Исходные данные:

x_i	1	1	3	3	3	5	5	6
y_j	17	18	16	17	18	16	17	16
n_{xy}	3	10	20	25	30	4	6	2

1. Строим корреляционную таблицу.

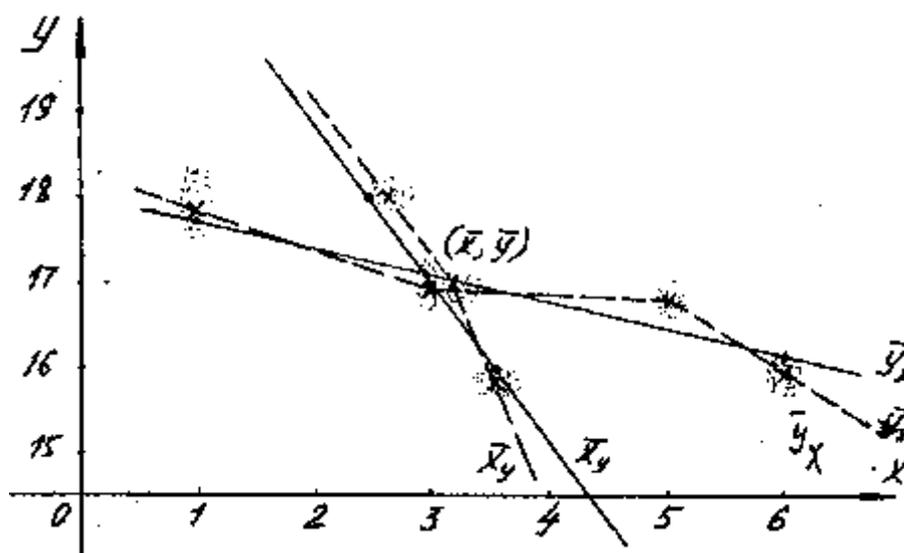
Для этого запишем по одному разу в возрастающем порядке x_i в верхней горизонтальной строке а y_j в правой вертикальной; в клеточках запишем, сколько раз данная точка встречалась в выборке, т.е. числа n_{xy} :

	1	2	5	6	n_y
16		20	4	2	26
17	3	25	6		34
18	10	30			40
n_x	13	75	10	2	$n=100$

n_x показывает сколько раз в выборке встречалось данное значение x с любым y , т.е. n_x равно сумме n_{xy} по столбцам; аналогично n_y показывает сколько раз в выборке встречалось данное значение y с любым x , т.е. n_y равно сумме n_{xy} по строкам. Объем выборки $n = \sum_x n_x = \sum_y n_y = 13 + 75 + 10 + 2 = 26 + 34 + 40 = 100$.

2. Чтобы визуальнo определить наличие линейной корреляционной зависимости, построим корреляционное поле и эмпирические линии регрессии.

Корреляционное поле – это совокупность точек $(x_i; y_i)$ на плоскости XoY .



Эмпирические линии регрессии – это ломаные, соединяющие точки $(x; \bar{y}_x)$ и $(y; \bar{x}_y)$.

Сделаем необходимые вычисления для построения эмпирической линии \bar{y}_x :

$$x = 1; \bar{y}_{x=1} = \frac{17 \cdot 3 + 18 \cdot 10}{13} = 17,7;$$

$$x = 3; \bar{y}_{x=3} = \frac{16 \cdot 20 + 17 \cdot 25 + 18 \cdot 30}{75} = 17,1;$$

$$x = 5; \bar{y}_{x=5} = \frac{16 \cdot 4 + 17 \cdot 6}{10} = 16,6;$$

$$x = 6; \bar{y}_{x=6} = \frac{16 \cdot 2}{2} = 16,0.$$

Получим точки $(1; 17,8)$, $(3; 17,1)$, $(5; 16,6)$, $(6; 16)$, отложим их на графике и соединим отрезками ломаной.

Аналогично строим эмпирическую линию \bar{x}_y

$$y = 16; \bar{x}_{y=16} = \frac{3 \cdot 20 + 5 \cdot 4 + 6 \cdot 2}{26} = 3,5;$$

$$y = 17; \bar{x}_{y=17} = \frac{1 \cdot 3 + 3 \cdot 25 + 5 \cdot 6}{34} = 3,2;$$

$$y = 18; \bar{x}_{y=18} = \frac{1 \cdot 10 + 3 \cdot 20}{40} = 2,5.$$

Полученные точки (3,5;16),(3,2;17), (2,5;18), соединяем их отрезками ломаной;

3. Аналитически тесноту линейной корреляционной зависимости определим, вычислив выборочный коэффициент корреляции, который является оценкой теоретического коэффициента корреляции r_{xy} :

$$\tilde{r}_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\tilde{\delta}_x \cdot \tilde{\delta}_y};$$

$$\bar{x} = \tilde{m}_x = \frac{1}{n} \sum_{i=1}^k n_x x_i = \frac{13 \cdot 1 + 75 \cdot 3 + 10 \cdot 5 + 2 \cdot 6}{100};$$

$$\bar{y} = \tilde{m}_y = \frac{1}{n} \sum_{j=1}^l n_y y_j = \frac{26 \cdot 16 + 34 \cdot 17 + 40 \cdot 18}{100} = 17,4;$$

$$\begin{aligned} \overline{xy} &= \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^l n_{xy} x_i y_j = \frac{20 \cdot 3 \cdot 16 + 4 \cdot 5 \cdot 16 + 2 \cdot 6 \cdot 16 + 3 \cdot 1 \cdot 17}{100} + \\ &+ \frac{25 \cdot 3 \cdot 17 + 6 \cdot 5 \cdot 17 + 10 \cdot 1 \cdot 18 + 30 \cdot 3 \cdot 18}{100} = 51,08; \end{aligned}$$

$$\tilde{\delta}_x = \sqrt{\tilde{D}_x} = \sqrt{\frac{\sum n_x (x_i - \bar{x})^2}{n}} = \sqrt{\frac{13(1-3)^2 + 75(3-3)^2 + 10(5-3)^2 + 2(6-3)^2}{100}} = 1,05;$$

$$\tilde{\delta}_y = \sqrt{\tilde{D}_y} = \sqrt{\frac{\sum n_y (y_j - \bar{y})^2}{n}} = \sqrt{\frac{26(16-17,1)^2 + 34(17-17,1)^2 + 40(18-17,1)^2}{100}} = 0,8;$$

$$\tilde{r}_{xy} = \frac{51,08 - 3 \cdot 17,1}{1,05 \cdot 0,8} = -0,40.$$

4. Проверим значимость выборочного коэффициента корреляции по критерию Стьюдента, т.е. проверим гипотезу:

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

$$H_1: r_{xy} \neq 0;$$

$$t_{\text{набл}} = \frac{|\tilde{r}_{xy}| \sqrt{n-2}}{\sqrt{1-\tilde{r}_{xy}^2}} = \frac{0,40 \sqrt{98}}{\sqrt{1-(-0,40)^2}} = 4,32;$$

$$t_{\text{крит}} = t_{\text{табл.дв.}} (\alpha = 0,05; f = n - 2 = 98 = 1,984);$$

$(t_{\text{набл}} = 4,32) > (t_{\text{крит}} = 1,984)$ H_0 отвергается, принимается $H_1: r_{xy} \neq 0$, т.е. линейная корреляционная зависимость значима и её можно описать с помощью уравнений регрессии;

5. Запишем уравнение регрессии Y на X

$$\bar{y}_x - \bar{y} = \tilde{r}_{xy} \frac{\tilde{\delta}_y}{\tilde{\delta}_x} (x - \bar{x});$$

$$\bar{y}_x - 17,4 = -0,40 \frac{0,8}{1,05} (x - 3);$$

$$\bar{y}_x = -0,3x + 18,05.$$

Запишем уравнение регрессии X на Y

$$\bar{x}_y - \bar{x} = \tilde{r}_{xy} \frac{\tilde{\delta}_x}{\tilde{\delta}_y} (y - \bar{y});$$

$$\bar{x}_y - 3 = -0,40 \frac{1,05}{0,8} (y - 17,4);$$

$$\bar{x}_y = -0,53y + 12,0.$$

6. Построим линии регрессии на корреляционном поле. Для этого на каждой прямой определим по две точки:

У _x	x	y
	1	17,75
	6	16,25
т.(1;17,75),(6;16,25)		

У _y	y	x
	16	3,52
	18	2,46
т.(3,52;16),(2,46;18)		

Линии регрессии пересекаются в точке $(\bar{x}; \bar{y})$, т.е. в точке (3;17,4).

7. Используем полученные данные уравнения регрессии для прогнозирования средней производительности мартеновской печи

$$\bar{Y}_{x=4} = -0,3 \cdot 4 + 18,05 = 16,85(\text{т/ч}).$$

Определим среднее время завалки лома, необходимое, чтобы производительность мартеновской печи равнялась 19 т/ч

$$\bar{X}_{y=19} = -0,53 \cdot 19 + 12,0 = 1,93(\text{мин}).$$

6. УКАЗАНИЯ К ВЫПОЛНЕНИЮ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ

Учебным планом дисциплины «Теория вероятностей и математическая статистика» предусмотрено выполнение индивидуальных заданий. Количество задач индивидуальных заданий определяется преподавателем. Перед решением соответствующих задач необходимо соответствующий раздел теоретического материала.

При выполнении индивидуальных заданий студент должен строго придерживаться следующих правил.

1. Выполнять контрольные работы строго в соответствии с вариантом, номер которого указывает преподаватель.

2. Каждое индивидуальное задание выполняется в отдельной тетради в клеточку чернилами любого цвета, кроме красных. В тетради должны быть отведены поля для рецензента; в конце тетради должно быть оставлено несколько чистых страниц для дополнений и исправлений в соответствии с замечаниями рецензента.

3. Оформление титульного листа должно соответствовать принятому образцу.

4. Перед решением каждой задачи указывается ее условие. Размещение задач необходимо выполнять в порядке увеличения их номеров.

5. Решение задач обязательно сопровождается пояснениями, необходимыми рисунками или графиками. Обязательна ссылка на соответствующие теоретические положения и формулы.

6. После получения прорецензированной работы студент должен внимательно изучить рецензию и выполнить все замечания рецензента.

7. Работа, выполненная с любыми нарушениями перечисленных выше требований, не засчитывается и возвращается студенту для переработки.

8. Если индивидуальное задание после проверки имеет замечания рецензента, то необходимо выполнить и исправить ошибки на дополнительных листах задания в конце тетради. Вносить изменения в

текст индивидуальных заданий после проверки рецензентом категорически запрещается.

9. Индивидуальные задания должны выполняться самостоятельно. Контроль за самостоятельно выполненной работой студента осуществляется преподавателем в форме собеседования со студентом или методом тестирования.

10. Номер варианта контрольной работы выбирается в соответствии с двумя последними цифрами зачетной книжки или студенческого билета. При этом, если этот номер превышает цифру 50, то номер контрольной работы определяется следующим образом: от цифры 100 вычитается цифра, которая соответствует двум последним цифрам зачетной книжки или студенческого билета. Например, если номер двух последних цифр зачетной книжки соответствует цифре 48, то студент выполняет 48 вариант. Если номер двух последних цифр зачетной книжки 85, то вариант контрольной работы выбирается следующим образом: $100-85=15$ и студент выполняет 15 вариант контрольной работы.

7. ЛИТЕРАТУРА

Основная

1. Жлуктенко В.І., Наконечний С.І. Теорія ймовірностей і математична статистика. – Ч.1. Теорія ймовірностей. – К.: КНЕУ, 2000.
2. Барковський В.В., Барковська Н.В., Лопатін О.К. Математика для економістів: Теорія ймовірностей та математична статистика. – К.: Національна академія управління, 1997.
3. Вентцель Е.С. Теория вероятностей. – М.: Наука, 1969 .
4. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. – М.: Наука, 1988.
5. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1997
6. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 1997.

Дополнительная

1. Колемаев В.А., Староверов О.В., Турундаевский В.Б. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1991.
2. Гнеденко Б.В., Хинчин А.Я. Элементарное введение в теорию вероятностей. – К.: Вища школа, 1991.
3. Крамер Г. Математические методы статистики. – М.: Мир, 1975 .
4. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики. – М.: Наука, 1969.
5. Колде Я.К. Практикум по теории вероятностей и математической статистике. – М.: Высшая школа, 1991.
6. Гурский Е.И. Теория вероятностей с элементами математической статистики. – М.: Высшая школа, 1971.
7. Микулик Н.А., Рейзина Г.Н. Решение технических задач по теории вероятностей и математической статистике. – Мн.: Вышэйшая школа, 1991.

8. ЗАДАЧИ ДЛЯ ВЫПОЛНЕНИЯ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ

1. ЭМПИРИЧЕСКИЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ.

1. Построить дискретный статистический ряд, полигон в $F^*(x)$ для случайного признака X – количества работающих завалочных машин в каждом цехе, если имеются сведения по 10 цехам за 15 дней:

1. 2,1,2,1,2,1,1,3,2,3,3,1,2,1,2.
2. 1,2,4,1,3,2,2,1,2,2,1,1,3,2,3.
3. 2,1,4,1,1,2,1,3,2,1,2,1,3,1,2.
4. 1,2,2,3,2,3,3,2,4,4,1,1,2,3,2.
5. 2,1,2,1,2,2,1,4,2,3,4,1,2,1,3.
6. 1,3,3,2,3,2,2,4,3,4,1,2,4,3,3.
7. 1,2,1,1,2,1,4,2,3,2,3,3,4,2,1.
8. 3,4,2,3,1,3,2,4,4,3,2,1,2,3,4.
9. 2,1,2,1,1,3,4,1,2,1,4,3,4,2,3.
10. 3,2,3,4,3,1,3,2,3,4,1,2,1,3,4.

2. Построить интервальный статистический ряд, гистограмму, полигон и $F^*(x)$ для случайного признака X – производительности мартеновской печи (в т/час), если имеются данные по 40 плавкам для каждой из 10 мартеновской печей:

- 1) 16,1; 16,2; 16,4; 15,3; 15,4; 15,0; 15,6; 15,7; 15,6; 15,8
16,6; 16,7; 16,6; 16,9; 16,9; 16,8; 16,7; 17,1; 17,2; 17,6;
17,8; 16,9; 16,7; 16,9; 16,6; 17,8; 16,1; 16,1; 16,2; 15,9;
15,8; 15,9; 16,9; 16,8; 16,6; 16,9; 16,4; 16,4; 16,3; 16,1;
- 2) 18,0; 15,9; 15,8; 15,2; 15,9; 16,1; 16,4; 15,3; 16,3; 16,2
16,9; 16,7; 16,6; 16,8; 17,2; 17,1; 17,4; 15,0; 15,1; 15,4;
17,6; 16,6; 16,5; 15,7; 15,8; 15,6; 15,1; 15,4; 15,5; 15,0;
15,0; 17,4; 17,0; 15,0; 15,4; 15,3; 15,2; 15,3; 15,2; 15,1
- 3) 15,4; 15,0; 15,1; 15,2; 15,5; 15,6; 15,7; 15,7; 16,0; 16,2
16,3; 16,4; 16,5; 16,4; 16,6; 16,9; 16,9; 17,0; 16,7; 17,1
17,3; 17,2; 17,4; 17,5; 17,6; 17,7; 17,8; 17,9; 16,4; 15,0;
15,2; 15,2; 15,4; 16,0; 16,8; 16,7; 17,3; 18,0; 17,6; 17,8.
- 4) 17,9; 17,8; 15,0; 15,1; 15,2; 15,1; 16,3; 16,2; 17,3; 17,1;
16,6; 16,7; 16,6; 15,4; 15,1; 15,2; 15,6; 15,9; 15,9; 17,7;
17,8; 16,3; 16,1; 16,2; 15,9; 15,8; 15,9; 16,9; 16,5; 16,8;
15,1; 15,1; 15,5; 15,5; 16,0; 16,5; 16,5; 17,0; 17,5; 17,5.
- 5) 17,4; 15,0; 15,1; 15,4; 15,3; 15,7; 15,9; 16,0; 16,2; 16,4;
15,2; 16,5; 15,0; 15,2; 15,3; 15,5; 15,3; 15,4; 18,0; 17,6;
17,1; 15,2; 17,5; 16,6; 16,7; 16,9; 15,9; 15,7; 15,2; 15,3;
16,0; 16,4; 16,2; 15,9; 16,0; 15,0; 16,3; 16,2; 15,5; 15,0.
- 6) 15,4; 15,0; 15,2; 15,1; 15,5; 15,6; 15,9; 16,0; 15,7; 15,8;
17,0; 17,7; 17,9; 17,1; 17,2; 16,1; 16,3; 16,2; 16,4; 16,2;
16,8; 16,9; 16,6; 17,0; 17,5; 17,4; 15,4; 15,5; 15,2; 15,9;

- 15,9; 15,7; 16,4; 16,9; 16,8; 15,0; 15,4; 15,2; 15,9; 17,0
- 7) 15,1; 15,0; 15,7; 15,8; 16,0; 16,4; 16,0; 16,1; 16,5; 16,9;
17,0; 17,4; 17,2; 17,6; 17,9; 17,7; 18,0; 15,0; 16,3; 15,4;
16,1; 16,2; 16,3; 16,3; 17,4; 17,3; 16,7; 16,9; 16,9; 17,9;
17,8; 15,4; 15,3; 17,0; 18,0; 17,8; 17,3; 17,4; 15,5; 18,0.
- 8) 17,1; 17,3; 17,4; 15,0; 15,3; 15,2; 15,4; 15,9; 16,8; 16,7;
16,1; 16,2; 16,3; 16,4; 17,9; 17,6; 16,9; 16,7; 16,7; 17,0;
17,6; 17,7; 17,6; 16,6; 17,0; 16,7; 16,9; 16,4; 16,3; 17,4;
15,8; 15,6; 15,7; 15,9; 16,6; 16,9; 18,8; 17,9; 17,8; 17,4.
- 9) 16,1; 0; 15,6; 15,7; 15,9; 15,0; 15,2; 15,1; 18,0; 17,9;
16,9; 16,8; 17,0; 17,1; 15,6; 15,7; 15,8; 15,3; 15,4; 15,3;
16,2; 16,4; 16,3; 16,7; 16,6; 16,9; 17,4; 17,9; 17,7; 16,4;
16,3; 16,2; 17,0; 16,9; 15,4; 15,3; 15,2; 15,7; 15,9; 15,9;
- 10) 16,4; 16,2; 16,3; 16,1; 16,2; 16,9; 16,8; 16,7; 16,5; 16,6;
17,1 17,0; 17,2; 17,9; 15,0; 15,9; 15,7; 15,6; 15,8; 16,1;
16,2; 16,2; 16,3; 16,4; 17,4; 17,3; 16,9; 16,8; 16,6; 16,7;
15,7; 15,6; 16,0; 16,9; 16,7; 16,3; 16,2; 17,0; 16,5; 16,9;

2. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ

По приведенным выше (разд.1) выборочным данным определить:

1) среднее число работающих завалочных машин в цехе (\tilde{m}_x) и отклонение от среднего значения ($\tilde{D}_x, \tilde{\sigma}_x$), используя дискретный вариационный ряд;

2) среднюю производительность мартеновской печи (\tilde{m}_x) и отклонение от среднего значения ($\tilde{D}_x, \tilde{\sigma}_x$) по интервальному вариационному ряду;

3) построить доверительные интервалы для m_x и D_x случайного признака X – производительности мартеновской печи при $\beta = 0,98$.

3. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

По приведенным выше (разд.1) выборочным данным случайного признака X (производительности мартеновской печи) высказать гипотезу о виде закона распределения этого признака и проверить ее с помощью критерия согласия χ^2 - Пирсона при уровне значимости $\alpha = 0,01$

4. ДИСПЕРСИОННЫЙ АНАЛИЗ

Определить, значимо ли при $\alpha = 0,025$ влияние концентрации производства (фактор А с уровнями: A_1 – мелкие предприятия, A_2 – средние, A_3 - крупные) и типа оборудования (фактор В с уровнями: B_1 – прогрессивное, B_2 - обычное) на производительность труда (выходная

величина Y в рублях, уменьшенная в 100 раз), в следующих группах предприятий, если известно, что предпосылка дисперсионного анализа выполняются, а таблицы наблюдений имеют следующий вид:

1.

B A	B_1	B_2
A_1	28	32
A_2	53	61
A_3	84	81

2.

B A	B_1	B_2
A_1	33	37
A_2	53	65
A_3	93	91

3.

B A	B_1	B_2
A_1	27	33
A_2	52	60
A_3	85	80

4.

B A	B_1	B_2
A_1	30	33
A_2	61	53
A_3	82	94

5.

B A	B_1	B_2
A_1	36	27
A_2	64	51
A_3	90	83

6.

B A	B_1	B_2
A_1	25	38
A_2	59	66
A_3	86	94

7.

B A	B_1	B_2
A_1	24	38
A_2	56	69
A_3	91	87

8.

B A	B_1	B_2
A_1	34	23
A_2	56	57
A_3	95	83

9.

B A	B_1	B_2
A_1	38	26
A_2	62	55
A_3	95	81

10.

B A	B_1	B_2
A_1	23	34
A_2	58	69
A_3	88	93

5. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Методом корреляционного анализа исследовать зависимость производительности мартеновской печи (Y , т/ч) от скорости завалки лома (X , мин) по паспортам 100 плавов для каждой печи, если известно, что все предпосылки корреляционного анализа выполняются; определить среднюю производительность исследуемой мартеновской печи, если время завалки лома будет равно 2,5 мин, а также среднее время завалки лома, чтобы производительность печи равнялась 17,5 т/час.

1) X_i	2	2	4	4	5	5	6
Y_i	16	17	15	16	15	16	15
n_{xy}	4	9	21	24	19	20	3
2) X_i	1	1	2	2	3	3	6
Y_i	16	17	15	16	15	16	15
n_{xy}	5	10	20	23	15	22	5
3) X_i	3	3	4	4	5	5	6
Y_i	16	17	15	16	15	16	15
n_{xy}	3	15	12	25	13	12	20
4) X_i	1	1	2	2	5	5	6
Y_i	16	17	15	16	15	16	15
n_{xy}	11	10	12	28	14	21	4
5) X_i	2	2	3	3	4	4	5
Y_i	16	17	15	16	15	16	15
n_{xy}	6	5	8	32	18	12	19
6) X_i	2	2	3	3	5	5	6
Y_i	16	17	15	16	15	16	15
n_{xy}	7	3	12	38	30	5	5
7) X_i	1	1	2	2	4	4	5
Y_i	16	17	15	16	15	16	15
n_{xy}	6	5	8	32	18	12	19
8) X_i	2	2	3	3	5	5	6
Y_i	16	17	15	16	15	16	15
n_{xy}	2	18	29	31	13	2	5
9) X_i	2	2	3	3	4	4	5
Y_i	16	17	15	16	15	16	15
n_{xy}	6	4	10	40	10	24	6
10) X_i	1	1	2	2	5	5	6
Y_i	16	17	15	16	15	16	15
n_{xy}	5	15	21	34	20	1	4

9. ТАБЛИЦА ВАРИАНТОВ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ

№№ вар	№№ заданий				
1	1.3	2.2	3.9	4.2	5.4
2	1.1	2.8	3.4	4.4	5.3
3	1.2	2.4	3.3	4.5	5.10
4	1.4	2.6	3.1	4.7	5.9
5	1.9	2.5	3.5	4.8	5.7
6	1.2	2.3	3.9	4.9	5.8
7	1.5	2.1	3.4	4.6	5.2
8	1.4	2.9	3.8	4.3	5.5
9	1.10	2.7	3.3	4.1	5.4
10	1.8	2.10	3.4	4.10	5.6
11	1.6	2.1	3.9	4.5	5.1
12	1.1	2.5	3.2	4.8	5.2
15	1.8	2.9	3.7	4.3	5.4
14	1.3	2.7	3.6	4.6	5.9
15	1.4	2.5	3.10	4.9	5.6
16	1.6	2.3	3.8	4.1	5.7
17	1.7	2.2	3.1	4.4	5.3
18	1.10	2.8	3.10	4.7	5.10
19	1.4	2.4	3.6	4.8	5.5
20	1.9	2.6	3.7	4.9	5.3
21	1.4	2.5	3.9	4.6	5.1
22	1.7	2.2	3.5	4.3	5.6
23	1.3	2.8	3.4	4.2	5.4
24	1.1	2.10	3.9	4.1	5.5
25	1.6	2.1	3.3	4.4	5.4

№№ вар	№№ заданий				
26	1.5	2.3	3.1	4.7	5.9
27	1.10	2.4	3.5	4.1	5.7
28	1.1	2.9	3.2	4.2	5.8
29	1.8	2.7	3.6	4.5	5.5
30	1.6	2.10	3.4	4.10	5.6
31	1.7	2.4	3.5	4.4	5.4
32	1.5	2.6	3.10	4.5	5.3
33	1.2	2.5	3.3	4.1	5.2
34	1.9	2.1	3.8	4.5	5.6
35	1.3	2.6	3.5	4.8	5.9
36	1.4	2.3	3.2	4.7	5.4
37	1.6	2.8	3.9	4.10	5.5
38	1.7	2.2	3.2	4.6	5.3
39	1.5	2.9	3.5	4.7	5.2
40	1.4	2.3	3.7	4.1	5.6
41	1.10	2.5	3.8	4.8	5.10
42	1.4	2.6	3.1	4.6	5.5
43	1.7	2.3	3.6	4.5	5.6
44	1.3	2.9	3.5	4.2	5.3
45	1.2	2.2	3.1	4.3	5.5
46	1.6	2.2	3.4	4.4	5.10
47	1.5	2.4	3.2	4.5	5.9
48	1.5	2.5	3.6	4.10	5.5
49	1.3	2.10	3.3	4.4	5.6
50	1.1	2.8	3.10	4.7	5.9

10. ПРИЛОЖЕНИЯ

Таблица 1

Таблица значений функции $\Phi(X) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$

X	Φ(X)								
0.00	0.000	0.25	0.0987	0.50	0.1915	0.75	0.2734	0.00	0.3413
0.01	0.0040	0.26	0.1026	0.51	0.1950	0.76	0.2764	0.01	0.3438
0.02	0.0080	0.27	0.1064	0.52	0.1985	0.77	0.2794	1.02	0.3461
0.03	0.0120	0.28	0.1103	0.53	0.2019	0.78	0.2823	1.03	0.3485
0.04	0.0160	0.29	0.1141	0.54	0.2054	0.79	0.2852	1.04	0.3508
0.05	0.0199	0.30	0.1179	0.55	0.2088	0.80	0.2881	1.05	0.3531
0.06	0.0239	0.31	0.1217	0.56	0.2123	0.81	0.2910	1.06	0.3554
0.07	0.0269	0.32	0.1255	0.57	0.2157	0.82	0.2939	1.07	0.2577
0.08	0.0319	0.33	0.1293	0.58	0.2190	0.83	0.2967	1.08	0.3599
0.09	0.0359	0.34	0.1331	0.59	0.2224	0.84	0.2995	1.09	0.3621
0.10	0.0398	0.35	0.1368	0.60	0.2257	0.85	0.3023	1.10	0.3643
0.11	0.0438	0.36	0.1406	0.61	0.2291	0.86	0.3051	1.11	0.3665
0.12	0.0478	0.37	0.1443	0.62	0.2324	0.87	0.3078	1.12	0.3683
0.13	0.0517	0.38	0.1480	0.63	0.2557	0.88	0.3106	1.13	0.3708
0.14	0.0557	0.39	0.1517	0.64	0.2389	0.89	0.3133	1.14	0.3729
0.15	0.0596	0.40	0.1554	0.65	0.2422	0.90	0.3159	1.15	0.3749
0.16	0.0636	0.41	0.1591	0.66	0.2454	0.91	0.3136	1.16	0.3770
0.17	0.0675	0.42	0.1528	0.67	0.2486	0.92	0.3113	1.17	0.3790
0.18	0.0714	0.43	0.1664	0.68	0.2517	0.93	0.3138	1.18	0.3810
0.19	0.0753	0.44	0.1700	0.69	0.2549	0.94	0.3264	1.19	0.3830
0.20	0.0793	0.45	0.1736	0.70	0.2580	0.95	0.3289	1.20	0.3849
0.21	0.0832	0.46	0.1772	0.71	0.2611	0.96	0.3315	1.21	0.3869
0.22	0.0871	0.47	0.1808	0.72	0.2642	0.97	0.3340	1.22	0.3883
0.23	0.0910	0.48	0.1844	0.73	0.2673	0.98	0.3365	1.23	0.3907
0.24	0.0948	0.49	0.1879	0.74	0.2703	0.99	0.3389	1.24	0.3925

Продолжение табл. 1

X	$\Phi(X)$								
1.25	0.3944	1.50	0.4332	1.75	0.4599	2.00	0.4772	0.50	0.4938
1.26	0.3962	1.51	0.4345	1.76	0.4608	2.02	0.4783	2.52	0.4941
1.27	0.3880	1.52	0.4357	1.77	0.4616	2.04	0.4793	2.54	0.4945
1.28	0.3997	1.53	0.4370	1.78	0.4625	2.06	0.4804	2.56	0.4948
1.29	0.4015	1.54	0.4382	1.79	0.4633	2.08	0.4812	2.58	0.4951
1.30	0.4032	1.55	0.4394	1.80	0.4641	2.10	0.4821	2.60	0.4953
1.31	0.4049	1.56	0.4406	1.81	0.4649	2.12	0.4830	2.62	0.4956
1.32	0.4066	1.57	0.4418	1.82	0.4656	2.14	0.4838	2.64	0.4959
1.33	0.4082	1.58	0.4429	1.83	0.4664	2.16	0.4846	2.66	0.4961
1.34	0.4099	1.59	0.4441	1.84	0.4671	2.18	0.4854	2.68	0.4963
1.35	0.4115	1.60	0.4452	1.85	0.4673	2.20	0.4861	2.70	0.5965
1.36	0.4131	1.61	0.4462	1.86	0.4686	2.22	0.4868	2.72	0.4967
1.37	0.4147	1.62	0.4474	1.87	0.4693	2.24	0.4875	2.74	0.4969
1.38	0.4162	1.63	0.4484	1.88	0.4699	2.26	0.4881	2.76	0.4971
1.39	0.4177	1.64	0.4495	1.89	0.4706	2.28	0.4887	2.78	0.4973
1.40	0.4192	1.65	0.4505	1.90	0.4713	2.30	0.4893	2.80	0.4974
1.41	0.4207	1.66	0.4515	1.91	0.4719	2.32	0.4898	2.82	0.4978
1.42	0.4222	1.67	0.4525	1.92	0.4726	2.34	0.4904	2.84	0.4977
1.43	0.4236	1.68	0.4535	1.93	0.4732	2.26	0.4909	2.86	0.4979
1.44	0.4251	1.69	0.4545	1.94	0.4738	2.38	0.4913	2.88	0.4980
1.45	0.4265	1.70	0.4554	1.95	0.4744	2.40	0.4918	2.90	0.4981
1.46	0.4279	1.71	0.4564	1.96	0.4750	2.42	0.4922	2.92	0.4982
1.47	0.4292	1.72	0.4573	1.97	0.4756	2.44	0.4927	2.94	0.4984
1.48	0.4306	1.73	0.4582	1.98	0.4761	2.46	0.4931	3.00	0.4986
1.49	0.4319	1.74	0.4591	1.99	0.4767	2.48	0.4934	4.00	0.49996

Таблица 2

Критические точки распределения Стьюдента

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

Таблица 3

Критические точки распределения χ^2

Число степеней свободы f	Уровень значимости α (двуст. крит. область)					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,0	17,5	15,5	2,73	2,13	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2

Таблица 4

Критические точки распределения F (Фишера).

Уровень значимости $\alpha = 0,025$									
$f_1 \backslash f_2$	1	2	3	4	5	6	7	8	9
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28
2	38,51	39,00	39,16	39,25	39,30	39,33	39,35	39,37	39,39
3	17,44	16,04	15,44	15,10	14,88	14,74	14,62	14,54	14,47
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68
6	8,81	7,26	6,60	6,23	5,99	5,82	5,69	5,60	5,52
7	8,70	6,54	5,89	5,52	5,28	5,12	4,99	4,90	4,82
8	7,57	6,06	5,42	5,05	4,82	4,65	4,63	4,43	4,36
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03
10	6,94	5,46	4,82	4,47	4,24	4,07	3,95	3,85	3,78
11	6,72	5,26	4,63	4,27	4,04	3,88	3,76	3,66	3,59
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44
13	6,41	4,96	4,35	3,99	3,77	3,60	3,48	3,39	3,31
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,28	3,21
15	6,20	4,76	4,15	3,80	3,58	3,41	3,29	3,20	3,12
16	6,11	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,00	2,93
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84
21	5,83	4,42	3,82	3,47	3,25	3,09	2,97	2,87	2,80
22	5,79	4,38	3,78	3,44	3,21	3,05	2,93	2,84	2,76
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68
26	5,66	4,26	3,67	3,33	3,10	2,94	2,82	2,73	2,64
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63
28	5,61	4,22	3,63	3,29	3,6	2,90	2,78	2,69	2,61
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45
60	5,29	3,92	3,34	3,01	2,79	2,63	2,51	2,41	2,33
120	5,15	3,80	3,23	2,89	2,67	2,51	2,39	2,30	2,22
∞	5,02	3,69	3,12	2,78	2,57	2,41	2,29	2,19	2,11

Продолжение табл. 4

Уровень значимости $\alpha = 0,025$										
$f_2 \backslash f_1$	10	12	15	20	24	30	40	60	120	∞
1	968,63	976,71	984,87	993,10	997,25	1001,4	1005,6	1009,8	1014,0	1018,3
2	39,40	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,50
3	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95	13,90
4	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26
5	6,62	6,52	6,42	6,38	6,28	6,23	6,17	6,12	6,07	6,01
6	5,46	5,36	5,27	5,17	5,12	5,06	5,01	4,96	4,90	4,85
7	4,76	4,66	4,57	4,47	4,41	4,36	4,31	4,25	4,20	4,14
8	4,29	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67
9	3,96	3,86	3,77	3,67	3,61	3,56	3,50	3,45	3,39	3,33
10	3,72	3,62	3,52	3,42	3,36	3,31	3,25	3,20	3,14	3,08
11	3,52	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
12	3,37	3,26	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
13	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60
14	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49
15	3,06	2,96	2,86	2,76	2,70	2,64	2,58	2,52	2,46	2,32
16	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32
17	2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32	2,75
18	2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26	2,19
19	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13
20	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09
21	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04
22	2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00
23	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97
24	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94
25	2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98	1,91
26	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95	1,88
27	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93	1,85
28	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91	1,83
29	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89	1,81
30	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79
40	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64
60	2,27	2,17	2,06	1,94	1,85	1,82	1,74	1,67	1,59	1,48
120	2,16	2,05	1,95	1,82	1,76	1,69	1,61	1,53	1,43	1,31
∞	2,05	1,94	1,83	1,71	1,64	1,67	1,48	1,39	1,27	1,00

Окончание табл. 4

Уровень значимости $\alpha = 0,05$									
$f_1 \backslash f_2$	1	2	3	4	5	6	7	8	9
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,39
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,95	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,89	2,74	2,63	2,54	2,42	2,41
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

Учебное издание

Швачич Геннадий Григорьевич
Соболенко Александр Викторович
Оржех Александр Игоревич

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Раздел 2. Математическая статистика

Учебное пособие

Тем. план 2006, поз. 105

Подписано к печати 24.01.2006. Формат 60x84 1/16. Бумага типогр. Печать плоская. Уч.-изд. л. 3,29. Усл. печ. л. 3,25. Тираж 450 экз. Заказ №

Национальная металлургическая академия Украины
49600, Днепропетровск-5, пр. Гагарина, 4

Редакционно-издательский отдел НМетАУ