Министерство народного образования БССР

БЕЛОРУССКИЙ ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ
ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ ИМЕНИ С.М.КИРОВА

Кафедра высшей математики

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ ПО ТЕМЕ "МАТЕМАТИЧЕСКАЯ СТАТИСТИКА" КУРСА "ПРИКЛАДНАЯ МАТЕМАТИКА"

для студентов спец. 26.01, 26.02

Минск 1988

УДК 519.2

Рассмотрены и рекомендованы к изданию Редакционно-издательским советом института.

Составители: А.С.Семенкова, Ж.Н.Горбатович, В.И.Янович.

Научный редактор доцент Е.А.Островский.

Рецензент доцент кафедры вычислительной техники С.Н.Губарь.

По тематическому плану внутривузовских изданий учебно-методической литературы на 1988 год. Поз. 55.(47).

Для студ. спец. 26.01 "Лесоинженерное дело", 26.02 "Технология деревообработки".

ВВЕДЕНИЕ

Стратегический курс на ускорение социально-экономического развития страны предполагает интенсификацию производства на базе научно-технического прогресса и внедрения эффективных форм управления. Это выдвигает новые задачи, связанные с разработкой научных методов исследования народнохозяйственных процессов. Современный инженер должен не только обладать техническими знаниями, но и уметь разрабатывать формальные модели реальных систем и процессов.

Материал в указаниях разбит на темы. По каждой теме приводятся необходимые теоретические сведения, решение типовых задач, набор задач для индивидуальной работы в аудитории и для самостоятельной работы. Задачи повышенной трудности, относящиеся к различным темам, даны в заключительном разделе. В приложениях приведен необходимый справочный материал.

IА. МЕТОДЫ СТАТИСТИЧЕСКОГО ОПИСАНИЯ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

I. Теоретические сведения

Выборкой объема n из генеральной совокупности с функцией распределения $F(x)$ называется совокупность $x_1, x_2, \ldots, x_n$ наблюдаемых значений случайной величины ξ (СВ ξ), соответствующих n независимым повторениям эксперимента.

Совокупность значений $x_1, x_2, \ldots, x_n$, упорядоченная по неубыванию, называется вариационным рядом.

Размах выборки w – разность между максимальным и минимальным значениями элементов выборки:

$$w = x_{max} - x_{min}.$$

Статистический ряд – совокупность пар (x_i, n_i), $i = \overline{1, k}$, где x_i – различные элементы выборки, n_i – частота появления выборочного значения x_i. Очевидно, что $\sum_{i=1}^{k} n_i = n$.

Обычно статистический ряд записывают в виде таблицы:

x_1	x_2	$\ldots$	x_i	$\ldots$	x_k
n_1	n_2	$\ldots$	n_i	$\ldots$	n_k

При большом объеме выборки её элементы объединяют в группы (разряды), получая при этом интервальный статистический ряд:

Интервалы	$[x_{min}, x_2[$	$[x_2, x_3[$	...	$[x_k, x_{max}]$
Частоты	n_1	n_2	...	n_k

Если все интервалы имеют одну длину h, то $h = w/k$; n_i - число элементов выборки, попавших в i-й интервал. Середину интервала будем обозначать x_i^*, а относительную частоту для каждого интервала - n_i/n; $\sum_{i=1}^{k} n_i/n = 1$.

Обычно число интервалов выбирают от 5 до 20 и чаще всего одинаковой длины. Для определения числа интервалов иногда используют формулу

$$k = 1 + 3{,}2 \lg n. \qquad (1)$$

Полигоном относительных частот выборки называется ломаная с вершинами в точках $(x_i^*, n_i/n)$, $i = \overline{1, k}$; по оси абсцисс откладываются выборочные значения, а по оси ординат - относительные частоты.

Гистограмма относительных частот интервального статистического ряда - это ступенчатая фигура, составленная из прямоугольников, построенных на интервалах группировки так, что площадь каждого прямоугольника равна n_i/n, $i = \overline{1, k}$. Сумма площадей прямоугольников равна I.

Эмпирической (статистической) функцией распределения СВ ξ, полученной по данной выборке, называется функция

$$F^*(x) = \sum_{x_i < x} n_i / n. \qquad (2)$$

Обычно считают, что $F^*(x)$ непрерывна слева. Функция $F^*(x)$ обладает всеми свойствами функции распределения. Если объем выборки большой, то значение эмпирической функции распределения в каждой точке x может служить приближенным значением (оценкой) теоретической функции распределения $F(x) = P(\xi < x)$ в этой точке.

Задача IA.I. Для выборки I2, II, I5, I2, 0, I6, I9, 6, II, I2, I6, II, I6, I5, I2 построить вариационный и статистический ряды, полигоны частот и относительных частот, эмпирическую функцию распределения.

Решение. Объем выборки n = I5. Упорядочив элементы выборки по величине, получим вариационный ряд: 0, 6, II, II, II,

I2, I2, I2, I2, I5, I5, I6, I6, I6, I9. Размах выборки w = = I9 - 0 = I9.

Статистический ряд дополним строкой относительных частот:

x_i	0	6	II	I2	I5	I6	I9
n_i	I	I	3	4	2	3	I
n_i/n	I/I5	I/I5	3/I5	4/I5	2/I5	3/I5	I/I5

$$\sum_{i=1}^{7} n_i = n, \quad \sum_{i=1}^{7} n_i/n = 1.$$

Полигоны частот и относительных частот представлены на рис. I и 2.

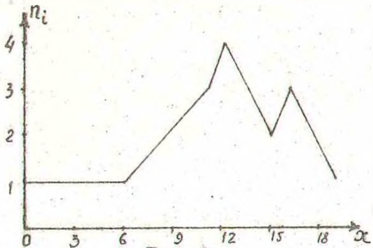

Рис.I

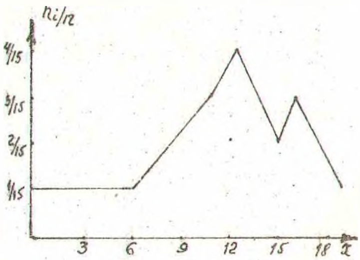

Рис.2

Эмпирическую функцию распределения $F^*(x)$ определим по формуле (2):

$$F^*(x) = \begin{cases} 0, & x \le 0, \\ I/I5, & 0 < x \le 6, \\ 2/I5, & 6 < x \le II, \\ 5/I5, & II < x \le I2, \\ 9/I5, & I2 < x \le I5, \\ II/I5, & I5 < x \le I6, \\ I4/I5, & I6 < x \le I9, \\ I, & x > I9. \end{cases}$$

График $F^*(x)$ представлен на рис. 3

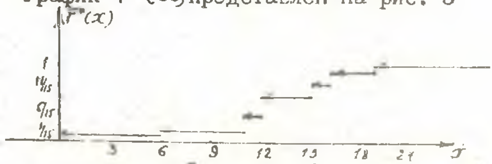

Рис.3

IA.2. Построить интервальный статистический ряд, гистограмму и полигон относительных частот, эмпирическую функцию распределения для выборки, объем которой n = I00:

32	27	43	42	36	25	43	20	42	32
36	39	49	22	36	38	45	3I	36	33
50	39	26	39	53	28	32	42	37	29
35	40	46	38	3I	35	54	I9	40	35
47	52	24	36	44	45	33	4I	30	I6
30	39	33	38	48	24	47	28	37	46
37	44	3I	55	32	26	4I	36	39	25
28	27	5I	2I	46	49	30	I8	43	34
I7	45	34	50	36	28	34	37	44	4I
25	29	24	34	26	27	33	25	34	I5

Решение. Так как объем выборки достаточно большой, строим интервальный статистический ряд. Для определения числа интервалов k используем формулу (I). Принимаем k = 7. Размах выборки w = 55-I5 = 40; длина интервала h = 40/7 ≈ 6.

Результаты группировки сведем в таблицу:

Интервал	Середина интервала x_i^*	Число наблюдаемых значений в интервале n_i	Относительная частота n_i/n
[I5, 2I[	I8	6	0,06
[2I, 27[	24	II	0,II
[27, 33[	30	20	0,20
[33, 39[	36	27	0,27
[39, 45[	42	I8	0,I8
[45, 5I[	48	I3	0,I3
[5I, 57]	54	5	0,05
		$\sum_i n_i$ = I00	$\sum_i n_i/n$ = I

Для построения гистограммы относительных частот на оси абсцисс откладываем интервалы, а на оси ординат – n_i/nh. На каждом i -м интервале строим прямоугольник, площадь которого равна его относительной частоте n_i/n (рис.4). Площадь гистограммы равна I.

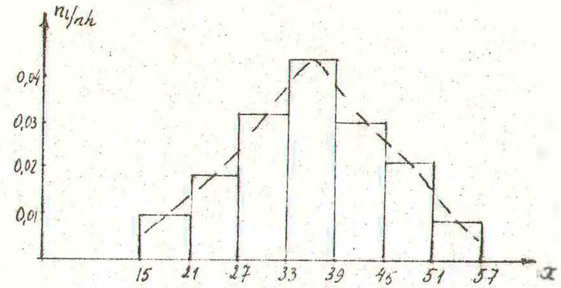

Рис.4

При большем значении объема n и достаточно малом h высоты построенных прямоугольников можно рассматривать в качестве приближенных значений для плотности p в средних точках соответствующих интервалов, т.е. верхнюю границу гистограммы можно рассматривать как статистический аналог плотности распределения наблюдаемой СВ.

В данном примере, судя по виду гистограммы, можно высказать предположение о нормальном распределении СВ (штриховая линия рис.4). Эмпирическая функция распределения представлена на рис.5.

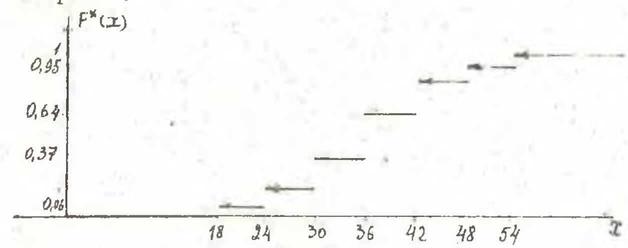

Рис.5

$$F^*(x) = \begin{cases} 0, & x \le 18, \\ 0{,}06, & 18 < x \le 24, \\ 0{,}17, & 24 < x \le 30, \\ 0{,}37, & 30 < x \le 36 \\ 0{,}64, & 36 < x \le 42, \\ 0{,}82, & 42 < x \le 48, \\ 0{,}95, & 48 < x \le 54, \\ \mathrm{I}, & x > 54. \end{cases}$$

2. Задачи для аудиторной работы

IA.3. Составить вариационный, статистический ряды для случайной величины – длины заготовок, отобранных случайным образом: 42, 44, 43, 43, 46, 44, 47, 45, 44, 44, 46, 45, 42, 43, 45, 46, 44, 43, 44, 42, 45, 45, 44, 45, 43, 44, 46, 44, 42, 43. Построить полигоны частот и относительных частот, график эмпирической функции распределения.

IA.4. Даны результаты измерения количества пиломатериалов, полученных из бревен лиственных пород, не удовлетворяющих техническим требованиям ГОСТа (% от сырья):

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

Составить интервальный статистический ряд распределения частот и относительных частот. Построить график эмпирической функции распределения и гистограмму относительных частот.

IA.5. В результате измерения шероховатости поверхности пиломатериалов получено I40 значений высот неровностей разрушения (мкм) на поверхности досок:

Границы интервалов	400-500	500-600	600-700	700-800	800-900	900-I000	I000-II00	II00-I200	I200-I300
Частоты	I2	23	25	25	23	I6	5	6	5

Построить гистограмму относительных частот. Найти эмпирическую функцию распределения и построить её график.

IA.6. Даны отклонения от номинального размера диаметра (мкм) I00 роликов:

Интервал	-0,I3; -0,II	-0,II; -0,09	-0,09; -0,07	-0,07; -0,05	-0,05; -0,03	-0,03; -0,0I	-0,0I; 0,0I
Частота	2	I5	2I	25	20	I4	3

Построить гистограмму относительных частот. Найти эмпирическую функцию распределения и построить ее график.

3. Задачи для самостоятельного решения

<u>Задание I</u>. Ответить на вопросы с. II.

<u>Задание 2</u>. Для группированных выборок построить гистограммы относительных частот, полигоны и эмпирические функции распределения (задачи IA.7 – IA.2I).

IA.7. Даны значения процентного выхода пиломатериалов 2-го сорта из бревен I-го сорта хвойных пород:

x_i, %	2I – 25	25 – 29	29 – 33	33 – 37	37 – 4I
n_i	7	I9	42	24	8

IA.8. Даны значения процентного выхода пиломатериалов 2-го сорта из бревен I-го сорта лиственных пород:

x_i, %	I0-II,2	II,2-I2,4	I2,4-I3,6	I3,6-I4,8	I4,8-I6	I6-I7,2
n_i	6	I0	I8	34	I6	II

IA.9. Даны результаты измерения (м) длин хлыстов хвойных пород, поступающих на деревообрабатывающее предприятие для раскроя на пиловочные бревна:

x_i, м	II-I2,4	I2,4-I3,8	I3,8-I5,2	I5,2-I6,6	I6,6-I8,0	I8,0-I9,4
n_i	4	I6	28	32	I5	5

IA.I0. Даны результаты измерения количества пиломатериалов, полученных из бревен лиственных пород, не удовлетворяющих техническим требованиям ГОСТа (% от сырья):

x_i, %	4,00-4,20	4,20-4,40	4,40-4,60	4,60-4,80	4,80-5,00
n_i	6	20	46	23	5

IA.II. Даны результаты измерения (м) фанерной зоны хлыстов березы (к данной зоне относится часть хлыста с диаметром в вершине не менее I6 см):

x_i, м	5,4 - 6,0	6,0 - 6,6	6,6 - 7,2	7,2 - 7,8	7,8 - 8,4
n_i	15	75	100	50	10

1А.12. Даны размеры диаметров 100 отверстий (мм), просверленных одним и тем же сверлом:

x_i, мм	40,10-40,20	40,20-40,30	40,30-40,40	40,40-40,50	40,50-40,60
n_i	7	24	34	26	9

1А.13. Даны размеры 200 деталей (мм) после шлифовки:

x_i, мм	3,45-3,65	3,65-3,85	3,85-4,05	4,05-4,25	4,25-4,45
n_i	7	36	99	50	8

1А.14. С автомата, обрабатывающего втулки диаметром d = 20 мм, была взята выборка объемом 100 единиц:

d_i, мм	20,00-20,04	20,04-20,08	20,08-20,12	20,12-20,16	20,16-20,20
n_i	7	14	55	15	9

1А.15. Даны результаты измерения объемного выхода (%) коротких досок из бревен лиственных пород (короткими считаются доски от 1,0 м до 1,5 м включительно):

x_i, %	1,0 - 1,1	1,1 - 1,2	1,2 - 1,3	1,3-1,4	1,4 - 1,5
n_i	7	20	44	21	8

1А.16. Даны результаты измерения твердости (по шкале HRC) 100 сверл:

x_i	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
n_i	9	20	45	19	7

1А.17. Даны результаты измерения твердости 100 фрез (по шкале HRC):

x_i	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
n_i	7	20	44	21	8

1А.18. Даны результаты испытания стойкости 100 сверл (ч):

x_i, ч	17,5-22,5	22,5-27,5	27,5-32,5	32,5 - 37,5	37,5-42,5
n_i	6	21	45	21	7

1А.19. Даны результаты испытания стойкости 100 фрез (ч):

x_i, ч	22,5 - 27,5	27,5 - 32,5	32,5 - 37,5	37,5-42,5	42,5-74,5
n_i	7	22	44	21	6

1А.20. Даны результаты измерения величины объема (м^3) бессучковой зоны хлыстов хвойных пород, поступающих на д/о предприятие:

x_i, м^3	1 - 2	2 - 3	3 - 4	4 - 5	5 - 6
n_i	6	23	38	25	8

1А.21. Даны результаты испытания стойкости 200 удлиненных сверл диаметром 4 мм (ч):

x_i, ч	3 - 3,2	3,2 - 3,4	3,4 - 3,6	3,6 - 3,8	3,8 - 4,0
n_i	17	49	70	46	18

Вопросы к теме 1

1. Какие способы представления первичного статистического материала Вы знаете?
2. Что называется гистограммой относительных частот?
3. Дайте определение и перечислите свойства эмпирической функции распределения.
4. В чем состоит связь эмпирической функции распределения $F^*(x)$ и функции распределения $F(x)$?
5. Какие вероятностные аналоги полигона относительных частот, гистограммы относительных частот, эмпирической функции распределения Вы знаете?

2А. ТОЧЕЧНЫЕ И ИНТЕГРАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

1. Теоретические сведения

Всякая случайная величина, являющаяся функцией выборки, называется статистикой. Статистика $\tilde{\theta}$, значение которой при каждой реализации выборки принимают за приближенное значение параметра θ, называется оценкой параметра θ.

Оценка $\tilde{\theta}$ называется состоятельной, если при $n \to \infty$ $\tilde{\theta}$ сходится по вероятности к истинному параметру θ, т.е.

$$P(|\tilde{\theta}-\theta|<\varepsilon)\xrightarrow[n\to\infty]{}1.$$

Оценка $\tilde{\theta}$ называется несмещенной оценкой параметра θ, если математическое ожидание оценки равно значению параметра θ, т.е. $M\tilde{\theta}=\theta$.

Оценка $\tilde{\theta}$ параметра θ называется эффективной, если ее дисперсия минимальна.

Оценки параметров подразделяются на точечные и интервальные.

Точечной оценкой для $M\xi$ служит выборочное среднее $\bar{x}$:

$$M\xi=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{n}\sum_{i=1}^{k}x_i n_i. \qquad (3)$$

Несмещенной оценкой дисперсии $D\xi$ является S^2 :

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n-1}\sum_{i=1}^{k}n_i(x_i-\bar{x})^2=$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{k}n_i x_i^2-\frac{\left(\sum_{i=1}^{k}n_i x_i\right)^2}{n}\right). \qquad (4)$$

Доверительным интервалом для параметра θ называется интервал $]\alpha_1, \alpha_2[$, содержащий истинное значение параметра с заданной вероятностью $\gamma=1-\alpha$, т.е. $P(\alpha_1<\theta<\alpha_2)=\gamma$. Число $\gamma=1-\alpha$ называется доверительной вероятностью, а значение α - уровнем значимости. На практике обычно используют уровни значимости 0,I; 0,05; 0,0I.

Доверительный интервал для математического ожидания нормально распределенной случайной величины ξ при заданном уровне значимости α :

При известной дисперсии $D\xi=\sigma^2$

$$\left(\bar{x}-t_\gamma\frac{\sigma}{\sqrt{n}}, \bar{x}+t_\gamma\frac{\sigma}{\sqrt{n}}\right), \qquad (5)$$

где t_γ определяется из условий

$$2\Phi(t_\gamma)=\gamma=1-\alpha, \quad \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt -$$

функция Лапласа (прилож.I).

При неизвестной дисперсии

$$\left(\bar{x}-t_{\alpha,n-1}\frac{s}{\sqrt{n}}, \bar{x}+t_{\alpha,n-1}\frac{s}{\sqrt{n}}\right) \qquad (6)$$

где $t_{\alpha,n-1}$ определяется с помощью таблицы распределения Стьюдента по заданному числу степеней свободы $(n-1)$ и уровню значимости α (прилож.2).

Задача 2А.I. На телефонной станции производились наблюдения за числом неправильных соединений в минуту. Наблюдения в течение часа дали следующие результаты: 3, I, 3, 4, 2, I, 2, 4, 0, 3, 0, 2, 2, 0, 2, I, 4, 3, 3, I, 4, 2, 2, I, I, 2, I, 0, 3, 4, I, 3, 2, 7, 2, 0, 0, I, 3, 3, I, 2, 4, 2, 0, 2, 3, I, 2, 5, I, I, 0, I, I, 2, 2, I, I, 5. Найти выборочное

среднее и несмещенную оценку дисперсии.

Решение. Представим данные в виде статистического ряда:

x_i	0	I	2	3	4	5	7
n_i	8	I7	I6	I0	6	2	I

Выборочное среднее находим по формуле (3):

$\bar{x}=\frac{1}{n}\sum_{i=1}^{7}x_i n_i$ = (0·8 + I·I7 + 2·I6 + 3·I0 + 4·6 + 5·2 + 7·I)/60 = 2.

Несмещенная оценка дисперсии $S^2=\frac{1}{n-1}\sum_{i=1}^{7}(x_i-\bar{x})^2 n_i$ = $\left[(0-2)^2\cdot 8+(I-2)^2\cdot I7+\ldots+(5-2)^2\cdot 2+(7-2)^2\cdot I\right]$ /6 = 2,I.

2А.2. Известны результаты измерения роста (см) случайно отобранных I00 студентов:

Рост	I54-I58	I58-I62	I62-I66	I66-I70	I70-I74	I74-I78	I78-I82
Число студ.	I0	I4	26	28	I2	8	2

Найти выборочное среднее $\bar{x}$ и несмещенную оценку дисперсии S^2.

Решение. Выборочную среднюю определяем по формуле: $\bar{x}=\frac{1}{n}\sum_{i=1}^{k}x_i^* n_i$: $\bar{x}$ =(I56·I0 + I60·I4 + ... + I76·8 +I80·2)/I00= = I66 (см); $S^2=\frac{1}{n-1}\sum_{i=1}^{k}n_i(x_i^*-\bar{x})^2$ = ((I56-I66)^2·I0 + (I60-I66)^2·I4 + ... + (I80-I66)^2·2) /99 = 33,44(см^2).

2А.3. В течение продолжительного срока при анализе данного материала на содержание железа установлено стандартное отклонение 0,I2 %. Найти с доверительной вероятностью 0,95 доверительный интервал для истинного содержания железа в образце, если по результатам 6 анализов среднее содержание железа составило 32,56 %.

Решение. По данным задачи имеем $\bar{x}$ = 32,56, σ = 0,I2, n = 6; $2\Phi(t_\gamma)$= 0,95 $\Rightarrow$ t_γ = I,96. Тогда по формуле (5) получим доверительный интервал (32,46; 32,66).

Смысл полученного результата: если произвести достаточно большое число выборок по 6 образцам, то в 95 % из них среднее содержание железа будет находиться в найденном интервале и только в 5 % случаев вне его.

2А.4. С помощью микроскопа была определена толщина прозрачного покрытия h (мкм) на древесной подложке. На образце покрытом полиэфирным лаком, было выполнено I0 замеров. В результате получены следующие значения величины h: 470, 354, 402, 434, 35I, 4I3, 465, 448, 540, 393. Требуется рассчитать точечную оценку и доверительный интервал для математического ожидания, взяв уровень значимости α равным 0,05 и 0,0I. (Считается, что толщина прозрачного покрытия имеет нормальное распределение).

Решение. Точечную оценку Mh найдем по формуле (3):

$$Mh = \bar{h} = (470+354+402+ \ldots +540+393)/10 = 427 \text{ (мкм)}.$$

Для нахождения S^2 используем формулу (4):

$$S^2 = \left((470-427)^2 + (354-427)^2 + \ldots + (540-427)^2 + (393-427)^2\right)/10 = 3293 \text{ мкм}^2.$$

Отсюда $S = \sqrt{3293} = 57,4$ мкм.

При $\alpha = 0,05$ и числе степеней свободы $n - 1 = 9$ по таблице прилож.2 найдем $t_{\alpha; n-1} = 2,26$. Подставляя найденные значения в формулу (6), получим доверительный интервал для математического ожидания: $\left(427 - \frac{2,26 \cdot 57,4}{10};\ 427 + \frac{2,26 \cdot 57,4}{10}\right)$, т.е. (386;468).

При $\alpha = 0,01$, $t_{\alpha; n-1} = 3,25$ доверительный интервал - (368;486), т.е. с большей надежностью можно гарантировать только более широкий доверительный интервал для математического ожидания при тех же опытных данных.

2А.5. Найти минимальный объем выборки, на основании которой можно было бы оценить математическое ожидание СВ ξ с ошибкой, не превышающей 0,2, и надежностью 0,98, если предположить, что ξ имеет нормальное распределение с $\sigma = 4$.

Решение. Так как $\varepsilon = t_\gamma \frac{\sigma}{\sqrt{n}}$, а $2\Phi(t_\gamma) = 0,98$, то

$$0,2 = 4 \cdot 2,23/\sqrt{n} \quad \text{и} \quad n = 16 \cdot 2,33^2/0,2^2 = 2170.$$

2. Задачи для аудиторной работы

2А.6. Найти оценку математического ожидания и несмещенную оценку дисперсии по выборке задачи IА.3.

2А.7. Найти несмещенную оценку дисперсии по выборке задачи IА.4.

2А.8. Найти несмещенную оценку дисперсии по выборке задачи IА.5.

2А.9. Найти несмещенную оценку дисперсии по выборке задачи IА.6.

В задачах 2А.I0-2А.I2 найти 95% -ные и 99%-ные доверительные интервалы для математического ожидания (предполагая нормальность распределения генеральных совокупностей):

2А.I0. По выборке задачи IА.4, используя результаты задачи 2А.7.

2А.II. По выборке задачи IА.5, используя результаты задачи 2А.8.

2А.I2. По выборке задачи IА.6, используя результаты задачи 2А.9.

3. Задачи для самостоятельного решения

Задание I. Для выборок задач IА.7-IА.2I: а) найти числовые характеристики; б) предполагая, что исследуемые случайные величины имеют нормальное распределение, найти 95%-ные и 99%-ные доверительные интервалы для математического ожидания.

Задание 2. Ответить на вопросы к теме 2А.

Вопросы к теме 2А

I. Как изменится выборочное среднее, если каждый член выборки: а) увеличить (уменьшить) на число α; б) увеличить (уменьшить) в K раз.

2. Определить, как изменится дисперсия выборки, если каждый член выборки: а) увеличить(уменьшить) на число α; б) увеличить (уменьшить) в K раз.

3. От каких величин зависит длина доверительного интервала?

4. Как изменится длина доверительного интервала: а) при увеличении доверительной вероятности? б) При увеличении объема выборки?

5. Перечислите свойства точечных оценок. В чем суть этих свойств?

6. Являются ли концы доверительного интервала постоянными величинами? Случайными?

7. Доказать, что $\frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2 =$

$$\frac{1}{n-1}\left(\sum_{i=1}^{k} n_i x_i^2 - \frac{\left(\sum_{i=1}^{k} n_i x_i\right)^2}{n}\right).$$

3А. СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ КРИТЕРИЙ χ^2 (ПИРСОНА)

I. Теоретические сведения

Статистической гипотезой H называется предположение относительно параметров или вида распределения $CB\ \xi$. Статистическая гипотеза называется простой, если она однозначно определяет распределение $CB\ \xi$. Например, простой гипотезой является предположение о том, что $CB\ \xi$ распределена по нормальному закону $N(0,I)$; если же высказывается предположение, что $CB\ \xi$ имеет нормальное распределение $N(m, 1)$, где $a \le m \le b$, то это сложная гипотеза.

Правило, по которому проверяемая гипотеза H принимается или отвергается, называется статистическим критерием проверки гипотезы H.

Одним из наиболее распространенных является критерий χ^2 (Пирсона). Пусть $x_1, x_2, \ldots, x_n$ – выборка наблюдений $CB\xi$. Проверяется гипотеза H_0, утверждающая, что ξ имеет закон распределения $F(x)$. Схема применения критерия χ^2 для проверки гипотезы состоит в следующем.

I. По выборке находим оценки неизвестных параметров предполагаемого закона распределения.

2. Если ξ – дискретная СВ, то определяем частоты n_i, $i=\overline{1,k}$, с которыми каждое значение или группа значений встречается в выборке.

Если ξ – непрерывная СВ, то разбиваем область ее значений на k интервалов и определяем число элементов выборки n_i, $i=\overline{1,k}$, принадлежащих каждому интервалу.

3. Если ξ – дискретная СВ, используя предполагаемый закон $F(x)$, находим вероятности p_i, $i=\overline{1,k}$, с которыми СВ ξ принимает каждое значение или вероятность появления группы значений.

Если ξ – непрерывная СВ, определяем вероятности p_i, $i=\overline{1,k}$ попадания в каждый интервал, исходя из предполагаемого закона $F(x)$.

4. Вычисляем выборочное значение статистики критерия χ^2:

$$\chi^2_{набл} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}. \qquad (7)$$

5. Принимаем статистическое решение: гипотеза H_0 не

противоречит выборке наблюдений на заданном уровне значимости α, если $\chi^2_{набл} < \chi^2_{\alpha,\nu}$, где $\nu = k - \ell - 1$ – число степеней свободы, а ℓ – число параметров распределения $F(x)$, определяемых по выборке; если же $\chi^2_{набл} \geq \chi^2_{\alpha,\nu}$, то гипотеза H_0 отклоняется.

Задача 3А.I. Проверить гипотезу о распределении СВ ξ по нормальному закону при условиях задачи IА.2 (α = 0,05).

Решение. Диапазон наблюдаемых значений разбит на 7 интервалов, а оценки параметров распределения $M\xi$ и $\sigma\xi$: $\bar{x}$ =35,94, S = 9,2. Тогда

$$p(x) = \frac{1}{9{,}2\sqrt{2\pi}}\, e^{-\frac{(x-35{,}94)^2}{2\cdot 9{,}2^2}}$$

и можно вычислить теоретические вероятности p_i попадания СВ в каждый i-й интервал:

$$p_i = P(x_{i-1} \le \xi < x_i) = \Phi\left(\frac{x_i-\bar{x}}{S}\right) - \Phi\left(\frac{x_{i-1}-\bar{x}}{S}\right).$$

Итак: $p_1 = P(-\infty < \xi < 21) = \Phi\left(\frac{21-35{,}94}{9{,}2}\right) - \Phi\left(\frac{-\infty-35{,}94}{9{,}2}\right) =$
$= \Phi(+\infty) - \Phi(1{,}62) = 0{,}5 - 0{,}44738 = 0{,}05262;$

$p_2 = P(21 \le \xi < 27) = \Phi\left(\frac{27-35{,}94}{9{,}2}\right) - \Phi\left(\frac{21-35{,}94}{9{,}2}\right) = \Phi(1{,}62) -$
$- \Phi(0{,}97) = 0{,}44738 - 0{,}33398 = 0{,}11340;$

$p_3 = P(27 \le \xi < 33) = \Phi(0{,}97) - \Phi(0{,}32) = 0{,}33398 - 0{,}12552 =$
$= 0{,}20846;$

$p_4 = P(33 \le \xi < 39) = \Phi(0{,}33) + \Phi(0{,}32) = 0{,}12930 + 0{,}12552 =$
$= 0{,}25482;$

$p_5 = P(39 \le \xi < 45) = \Phi(0{,}98) - \Phi(0{,}33) = 0{,}33646 - 0{,}12930 =$
$= 0{,}20716;$

$p_6 = P(45 \le \xi < 51) = \Phi(1{,}64) - \Phi(0{,}98) = 0{,}44950 - 0{,}33646 =$
$= 0{,}11304;$

$p_7 = P(51 \le \xi < \infty) = \Phi(\infty) - \Phi(1{,}64) = 0{,}5 - 0{,}44950 =$
$= 0{,}05050.$

Все вычисления сводим в таблицу:

Интервалы	Частота n_i	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i-np_i)^2}{np_i}$
]-∞ , 21[	6	0,05262	5,262	0,738	0,5446	0,1035
[21 ,27[	II	0,II340	II,340	-0,340	0,II56	0,0102
[27, 33[	20	0,20846	20,846	-0,846	0,7I57	0,0343

Продолж. таблицы

Интервалы	Частота n_i	p_i	np_i	n_i-np_i	$(n_i-np_i)^2$	$\frac{(n_i-np_i)^2}{np_i}$
[33, 39[	27	0,25482	25,482	-1,518	2,3043	0,0904
[39, 45[	18	0,20716	20,716	-2,716	7,5766	0,3561
[45, 51[	13	0,11304	11,304	1,696	2,8764	0,2544
[51, ∞[	5	0,0505	5,05	- 0,05	0,0025	0,0005

$\sum_{i=1}^{7} p_i = 1,0000$ $\qquad \chi^2_{набл} = 0,8491$

$\chi^2_{набл} = 0,8491$, $\chi^2_{0,05;4} = 9,49$ (находим по таблице прилож.3 при $\alpha = 0,05$ и числу степеней свободу $\nu = k - \ell - 1 = 7-2-1= =4$; ℓ - число определяемых параметров). Так как $\chi^2_{набл} < \chi^2_{0,05;4}$, то нет основания для отклонения гипотезы о нормальном распределении. Построим график функции $y = p(x)$. Для этого из середины частичных интервалов (рис.3) восстановим перпендикуляры высотой p_i/h_i (h_i - длина интервала). Полученные точки соединим плавной кривой. Сравнение гистограммы с нормальной кривой показывает, что гистограмма относительных частот сглажена хорошо.

2. Задачи для аудиторной работы

3А.2. Результаты измерения высот микронеровностей на поверхности доски представлены в виде интервального статистического ряда (n =140):

Интервалы высот	Частоты n_i	Интервалы высот	Частоты n_i
400 - 500	13	800 - 900	23
500 - 600	22	900 - 1000	15
600 - 700	25	1000 - 1100	5
700 - 800	26	1100 - 1200	6
		1200 - 1300	5

Проверить гипотезу о нормальном распределении высот микронеровностей ($\alpha = 0,05$).

3А.3. Результаты исследования прочности на сжатие 200 образцов бетона представлены в виде интервального статистического ряда (n =200):

Интервалы прочности, $кгс/см^2$	Частоты n_i	Интервал прочности	Частоты n_i
19 - 20	10	22 - 23	64
20 - 21	26	23 - 24	30
21 - 22	56	24 - 25	14

Проверить гипотезу о нормальном распределении прочности на сжатие (α =0,05).

3А.4. Рост 1004 девушек в возрасте 16 лет (см):

Интервал	134-137	137-140	140-143	143-146	146-149	149-152	152-155
Частота n_i	1	4	16	53	121	197	229

Проверить гипотезу о том, что выборка получена из нормально распределенной генеральной совокупности, приняв α =0,1.

3. Задачи для самостоятельного решения

Задание 1. Применяя критерий χ^2, проверить гипотезу о нормальном распределении СВ по данным задач 1А.7 - 1А.21.

Задание 2. Ответить на вопросы к теме 3А.

Вопросы к теме 3А

1. Что называется критерием согласия?
2. На основании каких признаков можно произвести предварительный выбор закона распределения?
3. Могут ли результаты эксперимента одновременно согласовываться с несколькими распределениями?
4. Опишите схему проверки гипотезы о виде функции распределения с помощью критерия χ^2.
5. Какие достоинства и недостатки имеет критерий χ^2 ?

4А. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

1. Теоретические сведения

Одной из важнейших задач математической статистики является нахождение связи между двумя случайными величинами ξ и η. Во многих случаях одна из переменных может быть неслучайной, т.е. принимать заданные значения, в то время как другая переменная имеет случайные флуктуации, обусловленные ошибками измерений или другими причинами.

Предположим, что функциональная зависимость между переменными, называемая моделью, известна из предварительных сведений с точностью до параметров β_0, β_1, ..., β_k и имеет вид $y=f(x;\beta_0,\beta_1,\dots,\beta_k)$. Требуется по результатам наблюдений $(x_i; y_i)$, $i=\overline{1,n}$ найти оценки параметров – $\tilde\beta_0$, $\tilde\beta_1$, ... $\tilde\beta_k$. По МНК в качестве оценок этих параметров (МНК-оценки) принимаются значения $\tilde\beta_0$, $\tilde\beta_1$, ..., $\tilde\beta_k$, дающие минимум функции

$$S(\tilde\beta_0,\tilde\beta_1,\dots,\tilde\beta_k)=\sum_{i=1}^{n}\left(y_i-f(x_i,\tilde\beta_0,\tilde\beta_1,\dots,\tilde\beta_k)\right)^2=\sum_{i=1}^{n}\varepsilon_i^2.$$

Из необходимых условий экстремума функции S находим, что МНК-оценки являются решениями системы

$$\frac{\partial S(\tilde\beta_0,\tilde\beta_1,\dots,\tilde\beta_k)}{\partial\tilde\beta_j}=0,\quad j=\overline{0,k}. \tag{8}$$

В общем случае эти уравнения нелинейны. Наиболее часто используют линейные по параметрам модели, например:

$$y=\beta_0+\beta_1 x+\dots+\beta_k x^k.$$

Если $y=\beta_0+\beta_1 x$, то система (8) примет вид

$$\begin{cases} n\tilde\beta_0+\tilde\beta_1\sum\limits_{i=1}^{n}x_i=\sum\limits_{i=1}^{n}y_i;\\ \tilde\beta_0\sum\limits_{i=1}^{n}x_i+\tilde\beta_1\sum\limits_{i=1}^{n}x_i^2=\sum\limits_{i=1}^{n}x_i y_i, \end{cases} \tag{9}$$

а для $y=\beta_0+\beta_1 x+\beta_2 x^2$ —

$$\begin{cases} n\tilde\beta_0+\tilde\beta_1\sum\limits_{i=1}^{n}x_i+\tilde\beta_2\sum\limits_{i=1}^{n}x_i^2=\sum\limits_{i=1}^{n}y_i;\\ \tilde\beta_0\sum\limits_{i=1}^{n}x_i+\tilde\beta_1\sum\limits_{i=1}^{n}x_i^2+\tilde\beta_2\sum\limits_{i=1}^{n}x_i^3=\sum\limits_{i=1}^{n}x_i y_i;\\ \tilde\beta_0\sum\limits_{i=1}^{n}x_i^2+\tilde\beta_1\sum\limits_{i=1}^{n}x_i^3+\tilde\beta_2\sum\limits_{i=1}^{n}x_i^4=\sum\limits_{i=1}^{n}x_i^2 y_i. \end{cases} \tag{10}$$

Систему можно решать любым из известных способов: по правилу Крамера, методом Гаусса, матричным способом.

Если $y=f(x,\beta_0,\beta_1,\dots,\beta_k)$ – не многочлен, а например, $y=ae^{bx}$, то, используя замену переменной, проводим линеаризацию: $\ln y=\ln a+bx\ln e$, $Y=\ln y$, $X=x$, $b_0=\ln a$, $b_1=b$ и получим $Y=b_0+b_1X$. При других видах зависимостей также используют метод замены переменных (метод выравнивания).

Задача 4А.I. Считая зависимость между коэффициентом трения пары шпон-сталь и нормальным давлением линейной, найти МНК-оценки параметров зависимости по результатам испытаний,

приведенным в таблице:

p_i, кг/см^2	I	2	3	4	5
f_i	0,I5	0,I7	0,20	0,2I	0,24

Построить точечную диаграмму.

Решение. Считаем, что $f=\beta_0+\beta_1 p$. Результаты промежуточных вычислений оформим в виде таблицы:

p_i	f_i	p_i^2	$p_i f_i$
I	0,I5	I	0,I5
2	0,I7	4	0,34
3	0,20	9	0,60
4	0,2I	I6	0,84
5	0,24	25	I,20
I5	0,97	55	3,I3

Согласно (9), получаем систему для определения $\tilde\beta_0$ и $\tilde\beta_1$:

$$\begin{cases} 5\tilde\beta_0+15\tilde\beta_1=0,97;\\ 15\tilde\beta_0+55\tilde\beta_1=3,13. \end{cases}$$

Решение системы $\tilde\beta_0=0,I3$; $\tilde\beta_1=0,022$.

Значит, f = 0,I3+0,022 P. Точечная диаграмма представлена на рис.6.

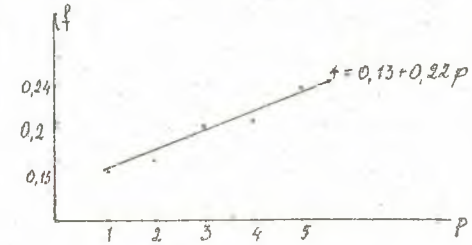

Рис.6

4А.2. Измерение температуры корпуса работающего аппарата, производимое с интервалом 5 мин, дало результаты:

t_i, мин	5	I0	I5	20	25
T_i, °C	59,3	59,8	60,I	64,9	70,2

Считая, что зависимость между ними имеет вид $T=\beta_0+\beta_1 t+\beta_2 t^2$, найти оценки $\tilde\beta_0$, $\tilde\beta_1$, $\tilde\beta_2$ по МНК.

Решение. Предварительно введем новые переменные (для упрощения вычислений): $x=\frac{t-15}{5}$, $y=10(T-60)$ и найдем $\tilde\beta_0$, $\tilde\beta_1$, $\tilde\beta_2$ модели $y=\beta_0+\beta_1 x+\beta_2 x^2$.

Промежуточные вычисления сведем в таблицу:

t_i	x_i	T_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y_i$	x_i^3	x_i^4
5	-2	59,3	-7	I4	4	-28	-8	I6
I0	-I	59,8	-2	2	I	-2	-I	I
I5	0	60,I	0	0	0	0	0	0
20	I	64,9	49	49	I	49	I	I
25	2	70,2	I02	204	4	408	8	I6
Σ	0	-	I43	269	I0	427	0	34

Система уравнений для определения β_i примет вид (см.(I0)):

$$\begin{cases} 5\tilde\beta_0 + 10\tilde\beta_2 = 143; \\ 10\tilde\beta_1 = 269; \\ 10\tilde\beta_0 + 34\tilde\beta_2 = 427. \end{cases}$$

Отсюда $\tilde\beta_0$ = 8,457, $\tilde\beta_1$ = 26,9, $\tilde\beta_2$ = I0,07 и y =8,457+26,9x + + I0,07 x^2, а T = 6I,84 - 0,67t + 0,04 t^2.

Решение можно проводить, не вводя новых переменных.

4А.3. По МНК найти оценки параметров для модели $y=\beta_0+\beta_1 \ln x$. Данные исследований приведены в таблице (рис.7):

x_i	I	2	3	4	5	6
y_i	2,II	2,45	2,6I	2,73	2,75	2,8I

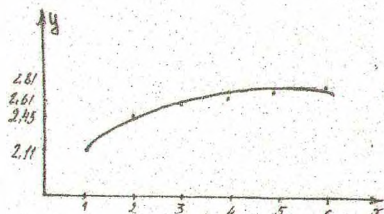

Рис.7

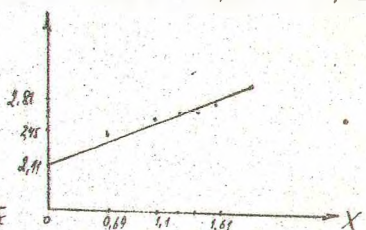

Рис.8

Решение. Пусть $Y=y$, $X=\ln x$, тогда $Y=\beta_0+\beta_1 X$ (рис.8).

$X_i=\ln x_i$	0	0,693	I,099	I,386	I,609	I,792
Y_i	2,II	2,45	2,6I	2,73	2,75	2,8I

Промежуточные вычисления поместим в таблицу, предварительно сделав замену: X_1 = I0(X -I,386), Y_1 = I0(Y -2,73). Тогда $Y_1=\alpha_0+\alpha_1 X_1$.

X_i	X_{1i}	Y_i	Y_{1i}	X_{1i}^2	$X_{1i} Y_{1i}$
0	-I3,86	2,II	-6,2	I92,I0	85,93
0,693	- 6,93	2,45	-2,8	48,02	I9,40
I,099	- 2,87	2,6I	-I,2	8,24	3,44
I,386	0	2,73	0	0	0
I,609	2,23	2,75	0,2	4,97	0,45
I,792	4,06	2,8I	0,8	I6,48	3,25
Σ	-I7,37	-	-9,2	269,8I	II2,47

Находим $\tilde\alpha_0$ и $\tilde\alpha_1$ по МНК:

$$\tilde\alpha_0 = \frac{\sum Y_{1i}\sum X_{1i}^2 - \sum X_{1i}\sum X_{1i}Y_{1i}}{n\sum X_{1i}^2 - (\sum X_{1i})^2} = \frac{-9{,}2\cdot 269{,}8I-(-I7{,}37)\cdot II2{,}47}{2\cdot 269{,}8I - I7{,}37^2} = -0{,}40I;$$

$$\tilde\alpha_1 = \frac{n\sum X_{1i}Y_{1i} - \sum X_{1i}\sum Y_{1i}}{n\sum X_{1i}^2 - (\sum X_{1i})^2} = \frac{6\cdot II2{,}47-(-I7{,}37)(-9{,}2)}{2\cdot 269{,}8I - I7{,}37^2} = 0{,}39I.$$

Итак, Y_1 = -0,40I+0,39IX_1, откуда I0Y-27,3 = -0,40I+ +0,39I·I0(X-I,386). Приведя подобные, получаем Y =0,39IX + +2,I5 и y = 0,39I $\ln x$ +2,I5.

2. Задачи для аудиторной работы

4А.4. Найти МНК-оценки параметров зависимости растворимости соли от температуры для модели $y=\beta_0+\beta_1 t$. Результаты измерений приведены в таблице:

t_i, °C	0	4	I0	I5	2I
y_i, %	66,7	7I,0	76,3	80,6	85,7

Построить точечную диаграмму.

4А.5. Найти МНК-оценки параметров зависимости жесткости М гнуто-клееных образцов из шпона березы от давления прессования Р для модели $M=\beta_0+\beta_1 P$. Результаты испытаний даны в таблице:

P_i, кг/см²	0,5	I	I,5	2,0	2,5
M_i, кг/см²	I500	I582	I650	I750	I800

Построить точечную диаграмму.

4А.6. По результатам испытаний найти линейную зависимость предела прочности при изгибе σ от плотности ρ древесностружечных плит:

ρ_i, г/см^3	0,55	0,60	0,65	0,7	0,75	0,8
σ_i, МПа	15	20	24	29,5	36	42

4А.7. Найти МНК-оценки параметров зависимости температуры смазочного масла в двигателе от скорости движения для модели $t=\beta_0+\beta_1 x+\beta_2 x^2$. Результаты испытаний даны в таблице:

x_i, км/ч	20	25	30	35	40	45	50	55	60
t_i, °C	43,5	43,9	44,2	45,0	46,0	46,9	47,5	49,0	51,0

4А.8. Исследование зависимости продолжительности t решения систем линейных уравнений от порядка n дало следующие результаты:

n_i	2	3	4	5	6	7	8	9	10
t_i, мин	12	35	75	130	210	315	445	600	800

Предполагая, что $t=A\cdot n^{\gamma}$, найти оценки параметров A и γ.

3. Задачи для самостоятельного решения

Задание 1. Построить корреляционное поле (точечную диаграмму). По характеру расположения точек на поле выбрать математическую модель регрессионной зависимости и найти оценки коэффициентов.

В задачах 4А.9-4А.25 рекомендуется использовать линейную модель, в задачах 4А.26-4А.32 - параболическую, в задачах 4А.33-4А.40 модель указана.

Задание 2. Ответить на вопросы к теме 4А.

4А.9. σ - прочность гнуто-клееных заготовок; R - радиус закругления углов.

R_i, мм	5	10	15	20	25	30
σ_i, кг/см^2	680	800	760	1000	1100	1010

4А.10. x - электровооруженность труда рабочего; y - выпуск продукции.

x_i, кВт·ч	3,0	3,5	4,0	4,5	5,0	5,5
y_i, т.руб	4,3	4,8	5,0	5,7	6,5	7,0

4А.11. f - коэффициент трения пары сталь-шпон (березовый); p - нормальное давление.

p_i, кг/см^2	1	2	3	4	5	6
f_i	0,206	0,224	0,192	0,189	0,188	0,179

4А.12. p - нормальное давление; f - коэффициент трения пары сталь-шпон (ольхи).

p_i, кг/см^2	1	2	3	4	5	6
f_i	0,198	0,189	0,219	0,192	0,180	0,187

4А.13. t - время роста кристаллов, с; l - длина кристалла, измеренная вдоль оси роста в микронах.

t_i, с	50	60	70	80	90	95	100	110	120	130
l_i, мк	19	21	22	28	25	25	29	28	36	32

4А.14. q - усилие прижима стального вальца, МПа; σ - деформация древесины в статике, мм.

q_i, МПа	1,0	1,5	2,0	2,5	3,5	4
σ_i, мм	0,11	0,16	0,22	0,33	0,4	0,38

4А.15. x - вес детали; t - время закрепления детали на токарном станке.

x_i, кг	0,7	0,8	1,0	1,2	1,3	1,4
t_i, с	2,2	2,3	2,4	2,5	2,6	2,7

4А.16. n - число слоев; M - условный модуль упругости пакета из березового шпона толщиной в 1,5 мм при параллельном расположении волокон.

n_i	5	9	15	21	27	31
M_i, кг/см	123800	138000	166000	174500	196500	211200

4А.17. v - скорость движения автомобиля ВАЗ-2301; S - длина тормозного пути.

v_i, км/ч	21	30	42	40	55	48	64	59
S_i, м	2,0	2,6	3,9	5,2	7,0	6,2	7,0	8,6

4А.18. M – прочность гнуто-клееных образцов из березового шпона; x – начальная влажность шпона.

x_i, %	8	12	16	20	25	30
M_i, кг/см^2	747	605	423	214	100	10

4А.19. t – температура при прессовании болтов из стекловолокна; y – предел их прочности.

t_i, °C	1,3	1,35	1,40	1,45	1,50	1,60	1,65	1,70
y_i, 10^7 Па	10,8	10,2	9,2	8,9	8,3	8,0	7,3	6,7

4А.20. P – давление прессования; M – условный предел прочности при разгибе гнуто-клееных образцов из березового шпона.

P_i, кг/см^2	0,5	1,0	1,5	2,0	2,5	3,0
M_i, кг/см^2	625	597	620	755	745	700

4А.21. P – давление прессования; M – жесткость гнуто-клееных образцов из шпона ольхи.

P_i, кг/см^2	0,5	1,0	1,5	2,0	2,5	3,0
M_i, кг/см^2	1245	1186	1296	1288	1284	1205

4А.22. x – средний балл студента I-го курса БТИ, y – средний балл по итогам двух первых семестров.

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

4А.23. P – давление прессования; M – условный предел прочности при разгибе гнуто-клееных образцов из шпона ольхи.

P_i, кг/см^2	0,5	1,0	1,5	2,0	2,5	3,0
M_i, кг/см^2	660	638	679	804	758	746

4А.24. x – масса груза; y – удлинение шнура, на котором подвешен груз.

x_i, кг	0,05	0,07	0,10	0,13	0,15	0,18	0,20
y_i, см	0,05	0,08	0,12	0,16	0,17	0,25	0,27

4А.25. P – давление прессования; M – жесткость гнуто-клееных образцов из шпона ольхи.

P_i, кг/см^2	0,5	1,0	1,5	2,0	2,5	3,0
M_i, кг/см^2	1758	1825	1801	1822	1888	1965

4А.26. M – средняя величина условного модуля упругости; n – число слоев березового шпона в пакете (толщина слоя 1,15 мм и волокна склеиваемых слоев перпендикулярны).

n_i	5	9	15	21	27	31
M_i, кг/см^2	96200	86900	92700	98700	125500	144000

4А.27. M – средняя величина условного модуля упругости; n – число слоев березового шпона в пакете (толщина слоя 1,5 мм и волокна склеиваемых слоев перпендикулярны).

n	5	9	15	21	27	31
M_i, кг/см^2	140000	150000	170000	178000	198500	202500

4А.28. M – прочность гнуто-клееных образцов из березового шпона; q – расход клея на основе смолы М-60.

q_i, г/м^2	65	75	85	95	115	125	145
M_i, кг/см^2	144	135	126	148	206	260	427

4А.29. σ – прочность склеивания древесного шпона с поверхностью древесностружечной плиты в зависимости от вязкости клея ρ:

ρ, $\frac{кг}{м \cdot с}$	60	90	120	150	180
σ_i, кН/м	2,64	3,48	3,76	3,80	3,64

4А.30. σ – прочность склеивания древесного шпона с поверхностью древесностружечной плиты; q – расход клея.

q_i, г/м^2	700	100	130	160	190
σ_i, кН/м	1,76	3,06	3,82	3,70	3,08

4А.31. σ – прочность склеивания древесного шпона с поверхностью древесностружечной плиты; p – давление запрессовки.

p_i, МПа	0,2	0,5	0,8	1,1	1,4
σ_i, кН/м	3,21	3,88	3,76	3,42	2,84

4А.32. x – предел прочности образцов стали; y – выносливость этих образцов.

x_i, гПа	0,70	0,80	0,90	1,00	1,10	1,20	1,30
y_i, гПа	0,33	0,40	0,44	0,49	0,49	0,51	0,55

4А.33. p – давление насыщенного пара; v – удельный объем.

v_i, см³	3,34	1,63	0,86	0,42	0,26	0,17	
p_i, Н/см²	0,48	1,03	2,03	4,25	7,16	11,48	модель $p=\beta v^{\alpha}$

4А.34. t – температура; y – удельная электропроводность (модель $y=b\,e^{at}$).

t_i, °С	14,5	30,5	64,5	74,5	86,7
y_i, $\frac{ом \cdot мм^2}{м}$	0	0,004	0,018	0,051	0,138

4А.35. p – давление пара; v – объем (модель $p=b\,e^{av}$).

v_i, см³	30	25	20	15	10	5
p_i, кг/см²	0,52	1,95	2,67	3,70	6,77	16,64

4А.36. v – скорость резания; S – подача (модель $v=b\,S^a$).

S_i, мм	0,5	0,6	0,7	0,8	0,9	1,0
v_i, м/с	38,0	33,8	30,6	28,0	25,8	24,1

4А.37. μ – коэффициент трения в подшипнике; t – температура (модель $\mu=\beta e^{-at}$).

t_i, °С	60	70	80	90	100	110
μ_i	0,0148	0,0124	0,0102	0,0085	0,0071	0,0059

4А.38. A – амплитуда колебаний маятника; t – время (модель $A=a\cdot e^{bt}$).

t_i, с	I	2	3	4	5	6
A_i	4,97	2,47	I,22	0,6I	0,30	I,I4

4А.39. v - скорость резания; t - глубина резания (модель $v = a\cdot t^{b}$).

t_i, мм	I	2	3	4	5	6
v_i, м/с	3I,6	27,8	25,7	24,2	23,3	22,5

4А.40. По данным эксперимента найти оценки параметров модели $y = a\cdot x^{n}$.

x_i	I	2	3	4	5
y_i	28,6	79,4	I82	3I8	589

Вопросы к теме 4А

I. В чем суть метода наименьших квадратов?

2. Вывести систему уравнений для МНК-оценок параметров зависимости $y = \beta_0 + \beta_1 x + \beta_2 x^2$.

3. Как линеаризовать зависимости: $y = a x^{b}$, $y = a\cdot e^{bx}$, $y = a + \frac{b}{x}$, $y = a + b\cdot \ln x$.

5А. СТАТИСТИЧЕСКОЕ ОПИСАНИЕ ДВУМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ (ξ, η). ЛИНЕЙНАЯ РЕГРЕССИЯ

I. Теоретические сведения

Пусть исход эксперимента описывается двумя случайными величинами ξ и η. Предварительное представление о зависимости ξ и η можно получить, нанося элементы выборки (x_i, y_i), $i = \overline{1, n}$ в виде точек на плоскость с выбранной системой координат. Эта точечная диаграмма называется <u>корреляционным полем</u>.

Распределение системы (ξ, η) характеризуется числовыми параметрами m_ξ, m_η, которые определяют положение центра распределения; дисперсиями σ_ξ, σ_η, которые определяют рассеивание распределения относительно центра; корреляционным моментом (ковариацией) $K_{\xi\eta} = M((\xi - M\xi)(\eta - M\eta))$ и коэффициентом корреляции $\rho = K_{\xi\eta}/\sigma_\xi \sigma_\eta$. Коэффициент корреляции определяет степень линейной зависимости между ξ и η: чем ближе $|\rho|$ к единице, тем теснее связь между ξ и η; при $\rho = 1$ корреляционная зависимость становится функциональной (линейной).

Функцией регрессии η на ξ назовем $M(\eta/\xi=x) = \int\limits_{-\infty}^{\infty} y\,p(y|x)dy = \varphi(x)$. Функция регрессии η на ξ позволяет прогнозировать среднее значение СВ η по фиксированным значениям СВ ξ. Если СВ (ξ, η) распределена по нормальному закону, то $\varphi(x)$ имеет линейный вид:

$$M(\eta|x) = \varphi(x) = \beta_0 + \beta_1 x = m_\eta + \rho \frac{\sigma_\eta}{\sigma_\xi}(x - m_\xi).$$

По функции регрессии η на ξ можно делать точечный прогноз значений y_i по фиксированным значениям x_i СВ ξ с ошибкой ε_i: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \in N(0, \sigma_\varepsilon)$, если $CB(\xi,\eta) \in N(\mu, \Sigma)$.

Как правило, при обработке результатов эксперимента распределение СВ (ξ, η) неизвестно, поэтому зависимость между ξ и η характеризуют эмпирическими формулами.

По виду корреляционного поля подбирают либо линейную, либо криволинейную функции регрессии.

Оценку линейной функции регрессии $M(\eta|\xi=x) = \beta_0 + \beta_1 x$ называют эмпирической функцией регрессии η на ξ и обозначают $\tilde{y}_x = \tilde{\beta}_0 + \tilde{\beta}_1 x$, где $\tilde{\beta}_0, \tilde{\beta}_1$ – оценки параметров β_0, β_1, определяемых по данным эксперимента (x_i, y_i), $i = \overline{1, n}$; $\tilde{\beta}_0$ и $\tilde{\beta}_1$ находим по МНК из условия минимума $\sum\limits_{i=1}^{n} (y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i))^2$.

Решая систему нормальных уравнений (9) матричным методом, получим

$$\begin{pmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{pmatrix} = \begin{pmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{pmatrix}, \text{ где } C = \begin{pmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix}^{-1} = \frac{1}{n\Sigma x_i^2 - (\Sigma x_i)^2} \begin{pmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{pmatrix} -$$

ковариационно-дисперсионная матрица:

Тогда: $$\tilde{a}_{y_1} = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2}, \qquad (11)$$

$$\tilde{\beta}_0 = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \qquad (12)$$

<u>При выборке небольшого объема</u> n оценки параметров распределения удобно проводить в следующем порядке. Вычисляют суммы Σx_i, Σy_i, Σx_i^2, Σy_i^2, $\Sigma x_i y_i$. Затем вычисляют суммы квадратов отклонений от среднего и произведений отклонений от среднего: $Q_x = \Sigma(x_i - \bar{x})^2 = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$, $Q_y = \Sigma(y_i - \bar{y})^2 = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}$, $Q_{xy} = \Sigma(x_i - \bar{x})(y_i - \bar{y}) = \Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}$.

Тогда $\bar{x} = \frac{\sum x_i}{n}$, $\bar{y} = \frac{\sum y_i}{n}$, $S_x^2 = \frac{Q_x}{n-1}$, $S_y^2 = \frac{Q_y}{n-1}$,

$$\widehat{K}_{\xi\eta} = \frac{Q_{xy}}{n-1}, \quad \widehat{\rho}_{\xi\eta} = r = \frac{Q_{xy}}{\sqrt{Q_x Q_y}} . \qquad (13)$$

Формулы для оценок $\widehat{\beta}_1$, $\widehat{\beta}_0$ (II,I2) можно упростить с учетом (I3): $\widehat{\beta}_1 = \frac{Q_{xy}}{Q_x} = r\frac{S_y}{S_x}$, $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$. (I4)

Итак, эмпирическая линейная функция регрессии η на ξ может быть представлена в виде $\bar{y}_x = \widehat{\beta}_0 + \widehat{\beta}_1 x$ или с учетом (I4):
$\bar{y}_x - \bar{y} = r\frac{S_y}{S_x}(x - \bar{x})$.

<u>Выборку большого объема</u> представляют в виде корреляционной таблицы. С этой целью группируют реализации СВ ξ и η по интервалам длины b_x и b_y, а в клетки таблицы записывают число пар исходной выборки (т.е. частоты) для каждой комбинации интервалов. В дальнейших вычислениях используют середины интервалов и соответствующие частоты.

Для упрощения вычислений вместо середин интервалов x_i^*, y_j^* вводят числа (условные нули) u_i и v_j:

$$u_i = \frac{x_i^* - d_x}{h_x}, \; i = \overline{1,k}; \quad v_j = \frac{y_j^* - d_y}{h_y}, \; j = \overline{1,l},$$

где d_x и d_y - середины наиболее часто встречающихся интервалов; $h_x = x_i^* - x_{i-1}^*$; $h_y = y_j^* - y_{j-1}^*$.

Оценки параметров распределения находят в следующем порядке. Сначала вычисляют суммы для u и v по формулам, аналогичным (I3). Тогда $\bar{x} = h_x \bar{u} + d_x$, $\bar{y} = h_y \bar{v} + d_y$; $S_x^2 = \frac{Q_u}{n-1} b_x^2$;
$S_y^2 = \frac{Q_v}{n-1} b_y^2$; $r = \frac{Q_{uv}}{\sqrt{Q_u Q_v}}$.

Чтобы охарактеризовать, насколько хорошо полученная функция $\bar{y}_x = \widehat{\beta}_0 + \widehat{\beta}_1 x$ отражает зависимость между средними значениями y и фиксированными значениями x, проверяют адекватность модели экспериментальным данным.

Для этого вычисляют:

I) остатки $e_i = y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i)$;

2) среднее квадратическое отклонение, характеризующее рассеивание экспериментальных точек относительно эмпирической линии регрессии: $S_e = \sqrt{\sum e_i^2/(n-2)}$;

3) средние квадратические отклонения эмпирических коэффициентов регрессии: $S_{\widehat{\beta}_0} = S_e\sqrt{C_{00}}$, $S_{\widehat{\beta}_1} = S_e\sqrt{C_{11}}$, где C_{00}, C_{11} - диагональные элементы ковариационно-дисперсионной матрицы C;

4) интервальные оценки коэффициентов β_0 и β_1:

$$\tilde{\beta}_i - t_{\alpha,n-2} S_{\tilde{\beta}_i} < \beta_i < \tilde{\beta}_i + t_{\alpha,n-2} S_{\tilde{\beta}_i}\ ,\ i = 0, 1, \qquad (15)$$

где $t_{\alpha,n-2}$ - квантиль распределения Стьюдента;

5) эмпирический коэффициент корреляции:

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} = \frac{Q_{xy}}{\sqrt{Q_x Q_y}}\ ;$$

$R = r^2$ называется коэффициентом детерминации и показывает процент рассеивания y_i , объясняемый линейной моделью;

6) проверка согласия линейной регрессии с результатами наблюдений сводится к проверке гипотезы об отсутствии линейной связи между переменными η и ξ , т.е. гипотезы $H_0 : \beta_1 = 0$. Гипотеза H_0 принимается на уровне значимости α , если доверительный интервал для β_1 накрывает нуль. В этом случае считаем, что линейная модель не согласуется с опытными данными, и должна быть использована другая модель.

Для проверки гипотезы H_0 можно использовать и статистику $F_H = \frac{\tilde{\beta}_1 Q_{xy}}{S_e^2}$. H_0 верна, если $F_H \geq F_{1-\alpha}(1, n-2)$, где $F_{1-\alpha}(1, n-2)$ определяется по распределению Фишера. Согласие линейной регрессии с результатами наблюдений может быть установлено и при проверке гипотезы о значимости коэффициента корреляции, т.е. гипотезы $H_0 : \rho = 0$.

Вычисляем $t_H = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$. Сравниваем t_H с t табл. $= t_{\alpha,n-2}$, определяемым по таблице распределения Стьюдента. Если $t_H > t_{\alpha,n-2}$, то нулевая гипотеза отвергается, и линейная модель согласуется с результатами наблюдений.

Задача 5А.I. По результатам выборки:

x_i	5	II	I8	20	25
y_i	I0	22	33	39	5I

I) построить корреляционное поле; 2) найти эмпирические уравнения регрессии η на ξ и ξ на η ; 3) доверительные интервалы для β_0 и β_1 ; 4) проверить гипотезу $H_0 : \rho = 0$ (уровень значимости $\alpha = 0{,}05$).

Решение. I) Строим корреляционное поле (рис.9).

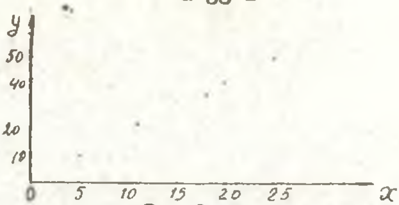

Рис.9.

2) Для нахождения уравнений регрессии используем формулы (13). Результаты оформим в виде таблицы:

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	5	10	25	100	50
2	11	22	121	484	242
3	18	33	324	1089	594
4	20	39	400	1521	780
5	25	51	625	2601	1275
Σ	79	155	1495	5795	2941

$\bar{x} = \dfrac{\sum x_i}{n} = 79/5 = 15.8$; $\bar{y} = \dfrac{\sum y_i}{n} = 155/5 = 31$; $Q_x = \sum x_i^2 - \dfrac{(\sum x_i)^2}{n} = 1495 - 79^2/5 = 246.8$; $Q_y = \sum y_i^2 - \dfrac{(\sum y_i)^2}{n} = 5795 - 155^2/5 = 990$; $s_x^2 = \dfrac{Q_x}{n-1} = \dfrac{246.8}{4} = 61.7$; $s_x = 7.85$; $s_y^2 = \dfrac{Q_y}{n-1} = 990/4 = 247.5$; $s_y = 15.73$; $r = \dfrac{Q_{xy}}{\sqrt{Q_x \cdot Q_y}} = \dfrac{492}{246.8 \cdot 990} = 0.995$.

Эмпирическое уравнение регрессии η на ξ: $\bar{y}_x - \bar{y} = r \dfrac{s_y}{s_x}(x - \bar{x})$; $\bar{y}_x - 31 = 0,995 \dfrac{15.73}{7,85}(x - 15,8)$ или $\bar{y}_x = 1,999x - 0,50 \approx 2x - 0,5$.

Это же уравнение можно получить, вычисляя коэффициенты $\tilde{\beta}_0$ и $\tilde{\beta}_1$ уравнения $y = \tilde{\beta}_0 + \tilde{\beta}_1 x$ по формулам (11),(12):

$$\tilde{\beta}_0 = \frac{1.55 \cdot 1495 - 79 \cdot 2941}{5 \cdot 1495 - (79)^2} = -0,498;$$

$$\tilde{\beta}_1 = \frac{5 \cdot 2941 - 79 \cdot 155}{5 \cdot 1495 - (79)^2} = 1,994;$$

$y = 1,994x - 0,498 \approx 2x - 0,5$ (сравните с предыдущим уравнением). Эмпирическое уравнение регрессии ξ на η: $\bar{x}_y - \bar{x} = r \dfrac{s_x}{s_y}(y - \bar{y})$. $\bar{x}_y - 15.8 = 0,995 \cdot 7,85/15,73 \cdot (y - 31)$ или

$\bar{x}_y = 0{,}496\,y + 0{,}41$.

3) Доверительные интервалы находим по формулам (15):

$$e_i = y_i - (\hat\beta_0 + \hat\beta_1 x_i) = y_i - (2x_i - 0.5)$$

x_i	y_i	$2x_i - 0,5$	e_i	e^2
5	10	9,5	0,5	0,25
11	22	21,5	0,5	0,25
18	33	35,5	-2,5	6,25
20	39	39,5	-0,5	0,25
25	51	49,5	1,5	2,25

$\sum e_i^2 = 9{,}25; \quad S_e = \sqrt{\sum e_i^2/(n-2)} = 3.08 = 1{,}755;$

$$C_{00} = \frac{\sum x_i^2}{n\sum x_i^2 - (\sum x_i)^2} = \frac{1495}{1234} = 1{,}21; \quad S_{\hat\beta_0} = S_e\sqrt{C_{00}} = 1{,}93;$$

$$C_{11} = \frac{n}{n\sum x_i^2 - (\sum x_i)^2} = \frac{5}{1234} = 0{,}004; \quad S_{\hat\beta_1} = S_e\sqrt{C_{11}} = 0{,}112;$$

$$t_{\alpha,n-2} = t_{0,05;3} = 3{,}182; \quad \hat\beta_i - t_{\alpha,n-2} S_{\hat\beta_i} < \beta_i < \hat\beta_i + t_{\alpha,n-2} S_{\hat\beta_i}, \; i = 0,1.$$

Таким образом, доверительные интервалы для β_i:

$$-6{,}642 < \beta_0 < 5{,}646;$$

$$1{,}628 < \beta_1 < 2{,}350.$$

Так как доверительный интервал для β_0 накрывает нуль, то β_0 незначим.

4) Адекватность модели (проверку гипотезы $H_0 : \rho = 0$) устанавливают с помощью статистики $t_н = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$:

$$t_н = \frac{0{,}995\ 3}{1-0{,}995} = 16{,}2; \qquad t_{\alpha,n-2} = t_{0{,}05;3} = 3{,}182.$$

Так как $t_н > t_{\alpha;n-2}$, то линейная регрессионная модель выбрана удачно.

Коэффициент детерминации $R = r^2 = 0{,}99$ и связь между ξ и η близка к функциональной.

5А.2. Исследовался прогиб доски y (мм) в зависимости от нагрузки x (кг). Найти средние значения прогиба, нагрузки, коэффициент корреляции между этими признаками. Определить для каждого x среднее значение $\bar{y}_x$ и для каждого y – $\bar{x}_y$. Найти уравнения регрессии, проверить гипотезу $H_0 : \rho = 0$; $\alpha = 0{,}05$.

v_j	u_i	-2	-1	0	1	2	
	y_i \ x_i	0,5-1,5 1	1,5-2,5 2	2,5-3,5 3	3,5-4,5 4	4,5-5,5 5	n_j
-2	13,5-14,5 14	10	8				18
-1	14,5-15,5 15		12	7			19
0	15,5-16,5 16			28	6		34
1	16,5-17,5 17				8	9	17
2	17,5-18,5 18					12	12
	n_i	10	20	35	14	21	$\Sigma n_{ij} = 100$

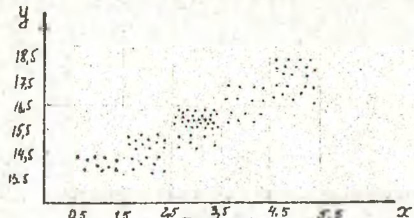

Рис.10

Решение. Для того, чтобы упростить вычисления, введем новые величины: u_i и v_j : $u_i = \frac{x_i^* - d_x}{h_x}$, $v_j = \frac{y_j^* - d_y}{h_y}$,

где $d_x = 3$, $d_y = 16$, $h_x = 1$, $h_y = 1$. Под каждым интервалом для x и y стоят x_i^* и y_j^* . Значения для u_i и v_j помещены в таблицу. Тогда

$$\bar{u} = \frac{\sum n_i u_i}{n} = \left[(-2)\cdot 10+(-1)\cdot 20+1\cdot 14+2\cdot 21\right]/100 = 0,16;$$

$$\bar{v} = \frac{\sum n_j v_j}{n} = \left[(-2)\cdot 18+(-1)\cdot 19+1\cdot 17+2\cdot 12\right]/100 = -0,14;$$

$$Q_u = \sum n_i u_i^2 - \frac{(\sum n_i u_i)^2}{n} = \sum n_i u_i^2 - n\bar{u}^2 =$$

$$= 158 - 100\cdot 0,16^2 = 155,44;$$

$$Q_v = \sum n_j v_j^2 - n\bar{v}^2 = 156 - 100\cdot 0,14^2 = 154,04;$$

$$Q_{uv} = \sum\sum u_i v_j n_{ij} - \frac{(\sum n_i u_i)(\sum n_j v_j)}{n} = (-2)(-2)\cdot$$

$$\cdot 10+8(-2)(-1)+(-1)(-1)\cdot 12+0(-1)\cdot 7+0\cdot 0\cdot 28+0\cdot 1\cdot 6+1\cdot 1\cdot 8+1\cdot 2\cdot 9+$$

$$+2\cdot 2\cdot 12 - 16\cdot(-14)/100 = 124,24;$$

$$S_u^2 = \frac{Q_{uv}}{\sqrt{Q_u Q_v}} = \frac{155.44}{99} = 1,570; \quad S_v^2 = \frac{Q_v}{n-1} = \frac{154,04}{99} = 1,556;$$

$r = \frac{Q_{uv}}{n-1} = \frac{124,24}{155,44 \cdot 154,04} = 0,93$; $\bar{x} = d_x + \bar{u} h_x = 3+0,16 \cdot 1 = 3,16$; $\bar{y} = d_y + \bar{v} h_y = 16-0,14 = 15,86$; $S_x = S_u \cdot h_x = \sqrt{1,570} \cdot 1 = 1,253$; $S_y = S_v \cdot h_y = \sqrt{1,556} \cdot 1 = 1,247$;

Найдем для каждого x среднее значение y — $\bar{y}_x$:

для $x = 1$, $\bar{y}_1 = \frac{14 \cdot 10}{10} = 14$;

$x = 2$, $\bar{y}_2 = (14 \cdot 8 + 15 \cdot 12)/20 = 14,6$;

$x = 3$, $\bar{y}_3 = (15 \cdot 7 + 16 \cdot 28)/35 = 15,8$;

$x = 4$, $\bar{y}_4 = (16 \cdot 6 + 17 \cdot 8)/14 = 16,6$;

$x = 5$, $\bar{y}_5 = (17 \cdot 9 + 18 \cdot 12)/21 = 17,6$ (рис.11).

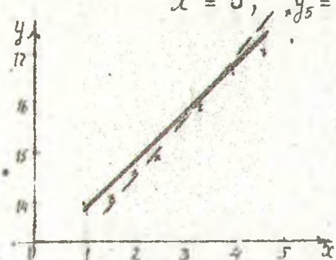

Рис.11.

Зависимость $\bar{y}_x$ от x найдем по формуле $\bar{y}_x - \bar{y} = r \frac{S_y}{S_x} (x - \bar{x})$,

$\bar{y}_x - 15,86 = 0,93 \frac{1,247}{1,253} (x - 3,16)$,

или $\bar{y}_x = 0,925x + 12,937$

(рис.11, сплошная прямая).

Средние значения x при каждом фиксированном y:

$y = 14$, $\bar{x}_1 = (1 \cdot 10 + 8 \cdot 2)/18 = 1,44$;

$y = 15$, $\bar{x}_2 = (2 \cdot 12 + 3 \cdot 7)/19 = 2,37$;

$y = 16$, $\bar{x}_3 = (28 \cdot 3 + 6 \cdot 4)/34 = 3,18$;

$y = 17$, $\bar{x}_4 = (8 \cdot 4 + 9 \cdot 5)/17 = 4,53$;

$y = 18$, $\bar{x}_5 = 12 \cdot 5/12 = 5$ (рис.11).

Эмпирическое уравнение регрессии ξ на η $\bar{x}_y - \bar{x} = r \frac{S_x}{S_y}(y - \bar{y})$: $\bar{x}_y - 3,16 = 0,93 \frac{1,253}{1,247}(y - 15,86)$, или $\bar{x}_y = 0,934y - 11,65$ (рис.11, штриховая прямая).

Проверим гипотезу H_0: $\rho = 0$. Вычислим $t_k = \frac{0,93 \sqrt{98}}{\sqrt{1-0,93^2}} = 24,9$. Находим $t_{табл.} = t_{\alpha, n-2} = t_{0,05, n-2} = 1,98$.

Так как $t_k > t_{табл.}$, то линейная модель выбрана удачно.

Промежуточные вычисления удобно оформлять в виде таблицы:

u_i	n_i	$n_i u_i$	$n_i u_i^2$
-2	I0	-20	40
-I	20	-20	20
0	35	0	0
I	I4	I4	I4
2	2I	42	84
Σ	I00	I6	I58

v_j	n_j	$n_j v_j$	$n_j v_j^2$
-2	I8	-36	72
-I	I9	-I9	I9
0	34	0	0
I	I7	I7	I7
2	I2	24	48
Σ	I00	-I4	I56

2. Задачи для аудиторного решения

В задачах 5А.3-5А.5 найти: I) центр рассеивания ($\bar{x}$, $\bar{y}$); 2) оценки средних квадратических отклонений S_x, S_y; 3) коэффициент корреляции r; 4) построить корреляционное поле; 5) найти уравнения регрессии; 6) проверить адекватность модели (α = 0,05).

5А.3. m - масса детали, кг; t - время, затраченное на закрепление детали на токарном станке, с:

m_i, кг	0,7	0,8	I,0	I,2	I,3	I,4	I,5	I,7	I,8	I,9
t_i, с	2,2	2,3	2,4	2,5	2,6	2,7	2,8	3,0	3,I	3,4

5А.4. M - прочность гнуто-клееных образцов из березового шпона; t - вязкость клея на основе смолы М-60.

t_i, $\frac{кг}{м \cdot с}$	30	60	I20	I50	200	240
M_i, кг/см^2	I30	I35	202	243	300	347

5А.5. ℓ - количество ремонтных слесарей, шт.; m - число станко-смен.

m \ ℓ	I0-I5	I5-20	20-25	25-30	30-35	35-40
0, - 0,2	4	-	-	-	-	-
0,2-0,4	2	2	-	-	-	-
0,4-0,6	-	-	2	-	-	-
0,6-0,8	-	6	-	4	4	-
0,8-I,0	-	-	-	-	6	6
I,0-I,2	-	-	-	-	-	4

3. Задачи для самостоятельного решения

Задание I. В задачах 5А.6-5А.20 найти: I) центр рассеивания ($\bar{x}$, $\bar{y}$); 2) оценки средних квадратических отклонений S_x,

S_y; 3) коэффициент корреляции r; 4) построить корреляционное поле; 5) найти уравнения регрессии; 6) проверить адекватность модели (α = 0,05).

Задание 2. Ответить на вопросы к теме 5А.

5А.6. x и y - содержание кремния и алюминия в образцах (в условных единицах)

y \ x	40-50	50-60	60-70	70-80
10 - 11	2	11	3	2
11 - 12	1	19	2	4
12 - 13	3	6	27	6
13 - 14	2	3	3	8

5А.7. x - ширина, см; y - длина плитки, штампуемой автоматом, см.

y \ x	1	3	5	7	9
7	-	-	-	2	4
10	-	-	3	4	6
13	-	1	8	3	2
16	-	2	9	6	-
19	-	3	5	7	2
22	8	4	1	-	-

5А.8. y - прочность стальной проволоки, кН/м; x - диаметр проволоки, мм.

y \ x	2,3	2,6	2,9	3,2	3,5	3,8
7,1	5	-	-	-	-	-
7,3	4	12	-	-	-	-
7,5	-	8	5	4	-	-
7,7	-	1	5	7	12	-
7,9	-	-	-	-	1	1

5А.9. x - овальность колец после обточки, y - после термической обработки.

y \ x	5	15	25	35	45	55	65
4	2	-	2	-	-	-	-
8	-	1	4	-	-	-	-
12	-	4	13	10	-	-	-

Продолжение таблицы

$y \backslash x$	5	I5	25	35	45	55	65
I6	-	2	-	22	II	6	-
20	-	-	-	-	I5	4	2
22	-	-	-	-	I	I	-

5А.I0. x – количество выпускаемых деталей, тыс.шт.; y – затраты на их изготовление, руб.

$y \backslash x$	5	I0	I5	20	25	30	35
I00	2	I	-	-	-	-	-
I20	3	4	3	-	-	-	-
I40	-	-	5	I0	8	-	-
I60	-	-	-	I5	-	6	2
I80	-	-	-	-	-	I	-

5А.II. x – сумма разрывных усилий отдельных проволок, составляющих стальной канат, кг; y – разрывное усилие всего каната, кг.

$y \backslash x$	I8	23	28	33	38	43
I25	-	I	-	-	-	-
I50	I	2	5	-	-	-
I75	-	3	I2	I2	-	-
200	-	-	II	I8	7	-
225	-	-	-	I3	3	-
250	-	-	-	-	I	I

5А.I2. x – угловые колебания ведущего моста автомобиля, Гц; y – угловые колебания подрессорной массы, Гц.

$y \backslash x$	I0	20	30	40	50	60
I5	5	7	-	-	-	-
25	-	20	23	-	-	-
35	-	-	30	47	2	-
45	-	-	I0	II	20	6
55	-	-	-	9	7	3

5А.I3. x – предел прочности, кг/мм^3; y – предел текучести, кг/мм^3.

y \ x	5	I0	I5	20
I0	2	-	-	-
20	5	4	I	-
30	3	8	6	3
40	-	3	6	6
50	-	-	2	I

5А.I4. x - высота дерева, см; y - диаметр, см.

y \ x	65	95	I25	I55	I85
I0	5	-	-	-	-
20	4	I2	-	-	-
30	-	8	I5	4	-
40	-	I	5	7	2
50	-	-	-	-	2

5А.I5. x - износ резца, мм; y - время работы, ч.

y \ x	I	2	3	4	5
I4	I0	8	-	-	-
I5	-	I2	7	-	-
I6	-	-	28	6	-
I7	-	-	-	8	9
I8	-	-	-	-	I2

5А.I6. x - ширина, см; y - длина плитки, штампуемой автоматом, см

y \ x	I	2	3	4
I	I	3	3	I
2	I	3	3	I
3	I	3	3	I
4	I	3	3	I
5	I	3	3	I

5А.I7. σ - предел прочности при изгибе, МПа; ρ - прочность древесностружечных плит, г/см^3.

ρ, г/см^3	0,55	0,6	0,65	0,7	0,75	0,8
σ, МПа	I5	20	24	29,5	36	42

5A.I8. x - остаточная деформация в статике древесины стальным вальцом при подаче рейсмусового станка, мм; y - остаточная деформация в динамике, мм.

x_i	005	002	005	0,I	007	007	005	005	0I5	0I4	0I2	0085	0085	0,I	0II
y_i	002	00I5	003	005	003	003	003	003	002	0085	005	005	004	004	005

5A.I9. Данные измерений двух переменных x и y :

$y \backslash x$	5	I0	I5	20	25	30	35
I00	-	-	-	-	-	6	I
I20	-	-	-	-	-	4	2
I40	-	-	8	I0	I5	-	-
I60	3	4	I3	-	-	-	-
I80	2	I	-	I	-	-	-

5A.20. Данные измерений двух переменных x и y :

$y \backslash x$	0,I	0,3	0,5	0,7	0,9
I2,5	4	2	-	-	-
I7,5	-	2	-	6	-
22,5	-	-	I2	-	-
27,5	-	-	-	I4	-
32,5	-	-	-	I4	6
37,5	-	-	-	I0	-

Вопросы к теме 5А

I. В чем состоит разница между функциональной и статистической зависимостью двух величин?

2. Какой вид имеют модельные функции регрессии η на ξ и ξ на η , если СВ (ξ, η) распределена по нормальному закону?

3. Через какую характерную точку проходят эмпирические уравнения регрессии?

4. Запишите систему нормальных уравнений для определения коэффициентов эмпирического уравнения регрессии $\bar{y}_x = b_0 + b_1 x$ по МНК.

5. Представьте схему алгоритма регрессионного анализа между двумя переменными.

6A. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

I. Теоретические сведения

Во многих задачах деревообработки изучается зависимость случайной величины η от нескольких переменных $x_1, x_2, \ldots, x_k$. Так, например, шероховатость поверхности зависит от подачи на зуб, жесткости, толщины, пути резания зубьев пил, высоты пропила, шага зубьев пил.

Функцией регрессии назовем условное математическое ожидание $M(\eta | \xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_k = x_k) = \varphi(x_1, \ldots, x_k)$. Простейшая регрессионная линейная модель $\varphi(x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$, $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ - вектор-столбец коэффициентов регрессии. Параметры множественной функции регрессии находят по методу наименьших квадратов (МНК). Предположим, что $n > k$, x_{ij} - j-е наблюдение i-го фактора, η - распределена по нормальному закону с постоянным средним квадратическим отклонением. Система нормальных уравнений по МНК для нахождения оценок $\tilde{\beta}$ вектора β имеет вид

$$(X'X)\tilde{\beta} = X'Y, \qquad (16)$$

где $Y' = (y_1, y_2, \ldots, y_n)$ - вектор наблюдений;

матрица

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{k1} \\ 1 & x_{12} & x_{22} & \ldots & x_{k2} \\ \vdots & & & & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{kn} \end{pmatrix}, \quad |X'X| \neq 0, \quad X'X = \begin{pmatrix} n & \sum x_{1i} & \ldots & \sum x_{ki} \\ \sum x_{1i} & \sum x_{1i}^2 & \ldots & \sum x_{ki} x_{1i} \\ \ldots & \ldots & \ldots & \ldots \\ \sum x_{ki} & \sum x_{ki} x_{1i} & \ldots & \sum x_{ki}^2 \end{pmatrix},$$

$$X'Y = \begin{pmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \vdots \\ \sum x_{ki} y_i \end{pmatrix}.$$

Несмещенную оценку $\tilde{\beta}$ вектора β коэффициентов регрессии находим по формуле $\tilde{\beta} = b = (X'X)^{-1} X'Y$. (17)

Дисперсия оценок $D(\tilde{\beta}) = \sigma^2 (X'X)^{-1}$, где $D(\eta | x_1, x_2, \ldots, x_k) = \sigma^2$.

Оценку вектора ошибок ε в модели $Y = X'\beta + \varepsilon$ определяем по формуле

$$S^2 = \frac{1}{n-k-1} (Y'Y - \tilde{\beta}' X'Y), \qquad (18)$$

так как обычно дисперсия ошибок σ^2 неизвестна.

Доверительный интервал для коэффициента β_j определяется:

$$\tilde{\beta}_j - t_{табл} S_{\tilde{\beta}_j} < \beta_j < \tilde{\beta}_j + t_{табл} S_{\tilde{\beta}_j}, \quad j = \overline{0, k}, \qquad (19)$$

где $t_{табл.}$ - находим с помощью таблицы распределения Стьюдента по числу степеней свободы $\nu = n-k-1$ и уровню значимости α, $t_{табл} = t_{\alpha, n-k-1}$; $S_{\tilde{\beta}_j} = S\sqrt{c_{jj}}$; c_{jj} - $(j+1)$-й диагональный элемент матрицы $(X'X)^{-1}$.

Проверку гипотезы $H_0: \beta_j = 0$ против $H_1: \beta_j \neq 0$ можно провести, вычисляя $t_{\text{набл.}} = |\beta_j|/S_{\beta_j}$ и сравнивая $t_{\text{набл.}}$ с $t_{\text{табл.}}$, вычисленным из таблицы распределения Стьюдента по ν и α. Если $t_{\text{набл.}} \geqslant t_{\text{табл.}}$, то коэффициенты регрессии значимы, т.е. гипотеза H_0 отклоняется.

Незначимые факторы, при которых коэффициенты β_j несущественно отклоняются от нуля, следует исключить из модели и заново пересчитать линейную множественную регрессионную модель с меньшим числом факторов.

Для предсказания качества регрессионной модели вычисляем коэффициент множественной корреляции:

$$R = \sqrt{1-(y'y-\hat{\beta}'X'y)(y'y-\frac{1}{n}(e'y)^2}\,,$$

где e - единичный вектор-столбец.

Обычно при $R > 0{,}7$ модель считается работоспособной.

Для случая $K = 2$ при ручном счете удобно находить оценки для модели вида $y-\bar{y} = \beta_1(x_1-\bar{x}_1)+\beta_2(x_2-\bar{x}_2)$. Оценка параметра β_0 вычисляется по формуле $\hat{\beta}_0 = \bar{y}-\hat{\beta}_1\bar{x}_1-\hat{\beta}_2\bar{x}_2$.

Оценки для параметров β_1, β_2 находим в этом случае в следующем порядке.

Вычисляем: $Q_y = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$; $Q_{x_j} = \sum\limits_i x_{ji}^2 - \frac{(\sum\limits_i x_{ji})^2}{n}$, $j=1,2$;

$$Q_{x_1x_2} = \sum_i x_{1i}x_{2i} - \frac{(\sum\limits_i x_{1i})(\sum\limits_i x_{2i})}{n};\quad Q_{x_jy} = \sum_i x_{ji}y_i - \frac{(\sum\limits_i x_{ji})(\sum\limits_i y_i)}{n},\ j=1,2;$$

$$A'A = \begin{pmatrix} Q_{x_1} & Q_{x_1x_2} \\ Q_{x_1x_2} & Q_{x_2} \end{pmatrix};\ (A'A)^{-1} = \frac{1}{|A'A|}\begin{pmatrix} Q_{x_2} & -Q_{x_1x_2} \\ -Q_{x_1x_2} & Q_{x_1} \end{pmatrix};\ A'y = \begin{pmatrix} Q_{x_1y} \\ Q_{x_2y} \end{pmatrix};$$

$$\bar{x}_1 = \frac{\sum x_{1i}}{n},\quad \bar{x}_2 = \frac{\sum x_{2i}}{n},\quad \bar{y} = \frac{\sum\limits_i y_i}{n}.$$

Вектор оценок:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (A'A)^{-1}A'y = \begin{pmatrix} Q_{x_2} & -Q_{x_1x_2} \\ -Q_{x_1x_2} & Q_{x_1} \end{pmatrix}\frac{1}{|A'A|}\begin{pmatrix} Q_{x_1y} \\ Q_{x_2y} \end{pmatrix}.$$

Статистический анализ эмпирического уравнения регрессии позволяет выяснить качество регрессионной модели - ее точность и адекватность. Остаточная сумма квадратов Q_e вычисляется по формуле $Q_e = Q_y - \hat{\beta}'A'y$.

Для проверки линейности модели проверяем гипотезу: $H_0 = \beta_1 = \beta_2 = 0$. Вычисляем статистику $F_H = \frac{\hat{\beta}'A'y/2}{Q_e/n-3}$. По уровню значимости α и числу степеней свободы ν_1 и ν_2 с помощью таблицы распределения Фишера вычисляем $F_\alpha(\nu_1,\nu_2)=F_\alpha(K, n-K-1)=F_\alpha(2, n-3)$. Если $F_H > F_\alpha(2, n-3)$, то нулевая гипотеза H_0 отвергается, т.е. уравнение регрессии значимо (адекватная модель). В противном случае линейной взаимосвязи с переменными x_1 и x_2 нет.

Значимость коэффициентов регрессии проверяем по критерию Стьюдента. Вычисляем $t_j^H = |\hat\beta_j| / S_{\hat\beta_j}$, где $S_{\hat\beta_j} = \sqrt{S_{ост}^2\, C_{jj}} = \sqrt{\frac{Q_e}{n-3} C_{jj}}$, ($j$ = I,2), C_{jj} - диагональный элемент матрицы $(A'A)^{-1}$.

По уровню значимости α и числу степеней свободы $\nu = n-K-1$ определяем t табл. t - критерия Стьюдента (прилож. 2). Если $t_j^H > t_{табл}$, $j=1,2$, то коэффициенты регрессии β_j статистически значимы. Если какой-то коэффициент β_j незначим, то отбрасываем j -переменную и снова по методу наименьших квадратов уточняем значимые коэффициенты регрессии.

С помощью t -критерия Стьюдента находим доверительный интервал для произвольного коэффициента регрессии β_j:

$$\hat\beta_j - t_{табл} S_{\hat\beta_j} < \beta_j < \hat\beta_j + t_{табл} S_{\hat\beta_j}$$

Для проверки гипотезы $H_0^{(j)}$: $\beta_j = 0$, j = I,2 значимости каждой переменной x_j воспользуемся доверительным интервалом для β_j (j =I,2). Если доверительный интервал для β_j (j =I,2) накрывает нуль, то гипотеза $H_0^{(j)}$: β_j = 0 принимается, в противном случае - отклоняется. Другими словами, фактор x_j либо незначим (в случае принятия гипотезы $H_0^{(j)}$: β_j =0, j = I,2), либо значим.

Коэффициент множественной корреляции

$$R = \sqrt{\hat\beta' A' Y / Q_y}$$

показывает процент рассеивания СВ η , объясняемый линейной регрессионной моделью.

2. Задачи для аудиторного решения

В задачах 6А.I - 6А.3 для приведенных данных: I) найти уравнение регрессии $y=\beta_0+\beta_1 x_1+\beta_2 x_2$ и доверительные интервалы для параметров β_1, β_2 ; 2) проверить гипотезу H_0 : $\beta_1=\beta_2=0$; 3) проверить гипотезы $H_0^{(j)}$: β_j = 0, j = I,2; 4) вычислить коэффициент множественной корреляции.

Уровень значимости принять α = 0,05.

6А.I. Результаты измерений прочности σ (МПа) ДСП в зависимости от температуры t ^{0}С и продолжительности прессования

τ (мин/мм толщины ковра):

t	I38	I72	I40	I83	I40	I55	I45	I60	I65	I70
τ	0,5	0,49	0,9I	0,9	0,5	0,6	0,7	0,8	0,9	0,5
σ	36,2	33,3	36,5	33,9	36,2	35,3	36,I	35,I	35	34,I

Продолжение таблицы

t	I75	I80	I42	I52	I62	I72	I54	I64	I74	I84
τ	0,5	0,6	0,7	0,8	0,9	0,9	0,8	0,7	0,6	0,5
σ	34	32,5	36,2	36	35,2	34,5	35,6	34,6	34	33,I

Решение. I) Вычислим величины Q_σ, Q_t, Q_τ, $Q_{t\sigma}$, $Q_{\tau\sigma}$, $Q_{t\tau}$. Результаты вычислений сведем в таблицу:

Номер опыта	$t=x_1$	$\tau=x_2$	$\sigma=y$	t^2	τ^2	$t\sigma$	$\tau\sigma$	σ^2	$t\tau$
I	2	3	4	5	6	7	8	9	I0
I	I38	0,5	36,2	I9044	0,25	4995,5	I8,I	I310,44	69
2	I72	0,49	33,3	3204I	0,236I	5960,7	I6,3I7	II08,89	87,7I
3	I40	0,9I	36,5	I9600	0,828I	5II0	33,2I5	I332,25	I27,4
4	I83	0,9	33,9	33489	0,8I	6203,7	30,5I	II49,2I	I64,7
5	I40	0,5	36,2	I9600	0,25	5068	I8,I	I3I0,44	70
6	I55	0,6	35,3	24025	0,36	542I,5	2I,I8	I246,09	93
7	I45	0,7	36,I	2I025	0,49	5234,5	25,27	I303,2I	I0I,5
8	I60	0,8	35,I	25600	0,64	56I6	28,08	I232,0I	I28
9	I65	0,9	35	27225	0,8I	5775	3I,5	I225	I48,5
I0	I70	0,5	34,I	28900	0,25	5797	I7,05	II62,8I	85
II	I75	0,5	34	30625	0,25	5950	I7	II56	87,5
I2	I80	0,6	32,5	32400	0,36	5850	I9,5	I056,25	I08
I3	I42	0,7	36,2	20I64	0,49	5I40,4	25,34	I3I0,44	99,4
I4	I52	0,8	36	23I04	0,64	5472	28,8	I296	I2I,6
I5	I62	0,9	35,2	26244	0,8I	5702,4	3I,68	I239,04	I45,8
I6	I72	0,9	34,5	29584	0,8I	5934	3I,05	II90,25	I54,8
I7	I54	0,8	35,6	237I6	0,64	5482,4	28,48	I267,36	I23,2
I8	I64	0,7	34,6	26896	0,49	5674,4	24,22	II97,I6	II4,8
I9	I74	0,6	34	30276	0,36	59I6	20,4	II56	I04,4
20	I84	0,5	33,I	33856	0,25	6090,4	I6,55	I095,6I	92
Σ	3234	I3,8	697,4	5324I4	I0,0242	II244I	482,342	24544,I6	2226,3I

Вычисляем: $Q_t = 527414 - \frac{(3234)^2}{20} = 527414 - 522937,8 =$
$= 4476,2;\ Q_\tau = 10,0242 - \frac{(13.8)^2}{20} = 10,0242 - 9,522 = 0,5022;$

$Q_\sigma = 24344,46 - \frac{(697,4)^2}{20} = 26,122;\ Q_{t\tau} = 2226,31 - \frac{3234 \cdot 33,8}{20} =$
$= 2226,31 - 22,31,46 = -5,15;\ Q_{t\sigma} = 112444 - \frac{3234 \cdot 697.4}{20} =$
$= 112444 - 112769,58 = -325,58;\ Q_{\tau\sigma} = 482,342 - \frac{13.8 \cdot 697.4}{20} =$
$= 482,342 - 481,206 = 1,136;$

$A'A = \begin{pmatrix} 4476,2 & -5,15 \\ -5,15 & 0,5022 \end{pmatrix},\ |A'A| = 4476,2 \cdot 0,5022 - (5,15)^2 = 2247,94 - 26,5225 = 2221,4251.$

$(A'A)^{-1} = \frac{1}{2221,4251} \begin{pmatrix} 0,5022 & 5.15 \\ 5,15 & 4476,2 \end{pmatrix} = \begin{pmatrix} 0,000226 & 0,002318 \\ 0,002318 & 2,015 \end{pmatrix}$

$A'y = A'\sigma = \begin{pmatrix} -325,58 \\ 1,136 \end{pmatrix},\ \bar{t} = \frac{\sum t_i}{n} = \frac{3234}{20} = 161,7;$

$\bar{\tau} = \frac{\sum \tau_i}{n} = \frac{13,8}{20} = 0,69;\ \bar{\sigma} = \frac{\sum \sigma_i}{n} = \frac{697,4}{20} = 34,87.$

Оценка $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (A'A)^{-1}(A'y) = \frac{1}{2221,4251} \cdot$
$\begin{pmatrix} 0,5022 & 5,15 \\ 5,15 & 4476,2 \end{pmatrix} \begin{pmatrix} -325,88 \\ 1,136 \end{pmatrix} = \frac{1}{2221,4251} \cdot$
$\begin{pmatrix} 0,5022 \cdot (-325,58) + 5,15 \cdot 1,136 \\ -325,58 \cdot 5,15 + 4476,2 \cdot 1,136 \end{pmatrix} = \frac{1}{2221,4251} \begin{pmatrix} -157,65587 \\ 3408,2262 \end{pmatrix} =$
$= \begin{pmatrix} -0,07097 \\ 1,53425 \end{pmatrix}.$

Итак, $\hat{\beta}_1 = -0,071;\ \hat{\beta}_2 = 1,534;\ \hat{\beta}_0 = 34,87 - (-0,071) \cdot 161,7 -$
$- 1,534 \cdot 0,69 = 45,29.$

Таким образом, уравнение регрессии:

$$\sigma = 45,29 - 0,071t + 1,534\tau .$$

Вычислим остаточную сумму квадратов Q_e и несмещенную оценку дисперсии ошибок S^2:

$Q_e = Q_y - \hat{\beta}' A' y = 26,122 - (-0,071 \cdot 1,534) \begin{pmatrix} -325,58 \\ 1,136 \end{pmatrix} = 26,122 -$
$- (0,071 \cdot 325,58 + 1,534 \cdot 1,136) = 1,6632;$

$S^2 = \frac{Q_e}{n-3} = \frac{1,6632}{17} = 0,0978;\ S = 0,3127.$

Точечные несмещенные оценки средних квадратических отклонений коэффициентов регрессии $\hat{\beta}_1, \hat{\beta}_2$:

$S_{\hat{\beta}_1} = S\sqrt{C_{11}} = 0,3127\ 0,000226 = 0,3127 \cdot 0,01503 = 0,0047;$

$S_{\hat{\beta}_2} = S\sqrt{C_{22}} = 0,3127 \cdot 2,013 = 0,3127 \cdot 1,4195 = 0,44388 = 0,4439$

где c_{11}, c_{22} – диагональные элементы матрицы $(A'A)^{-1}$.

С помощью таблицы распределения Стьюдента (прилож.2) по уровню значимости α и числу степеней свободы $n-k-1=n-3$ определяем $t_{\alpha, n-3}$. Тогда доверительные интервалы для параметров β_j, $j = 1,2$ определяим по формуле:

$$\tilde{\beta}_j - t_{\alpha,n-3} S_{\tilde{\beta}_j} < \beta_j < \tilde{\beta}_j + t_{\alpha,n-3} S_{\tilde{\beta}_j}, \quad j=1,2.$$

Следовательно, доверительные интервалы для коэффициентов β_j, $j = 1,2$ линейного уравнения регрессии равны
$-0,071-2,11\cdot0,0047 < \beta_1 < -0,071+2,11\cdot0,0047$; $-0,081 < \beta_1 < -0,061$;
$1,534-2,11\cdot0,4439 < \beta_2 < 1,534+2,11\cdot0,4439$; $0,597 < \beta_2 < 2,471$.

Итак, параметры β_1 и β_2 с доверительной вероятностью 0,95 накрываются соответственно интервалами:]-0,081;-0,061[и]0,597;2,471[.

2) Для проверки гипотезы $H_0: \beta_1 = \beta_2 = 0$ вычисляем выборочное значение статистики F_H. Так как $\tilde{\beta}'A'Y =$
$= (-0,071;1,534)\begin{pmatrix}-325,58\\1,136\end{pmatrix} = 24,8588$, а $S^2 = Q_e/n-3 = 0,0978$, то $F_H = \dfrac{24,8588/2}{0,0978} = 127,08$. Так как F_H больше $F_{0,95}(2,17) = 3,59$, то гипотеза H_0 отклоняется.

3) Проверим значимость переменных t и τ, т.е. оказывают ли существенное влияние температура t и продолжительность прессования τ на прочность ДСП при статическом изгибе.

Проверить нулевые гипотезы $H_0^{(j)}: \beta_j = 0$, $j = 1,2$ можно непосредственно по доверительным интервалам для параметров β_1 и β_2, а именно: доверительные интервалы для параметров β_1 и β_2 не накрывают нуль. Следовательно, переменные t и τ значимы.

Значимость коэффициентов регрессии β_1 и β_2 можно проверить и используя распределение Стьюдента. Проверим гипотезу $H_0^{(1)}: \beta_1 = 0$ против конкурирующей гипотезы $H_1^{(1)}: \beta_1 \neq 0$. Находим $t_{\text{набл.}} = \dfrac{|\tilde{\beta}_1|}{S_{\tilde{\beta}_1}} = \dfrac{0,071}{0,0047} = 15,11$. Из распределения Стьюдента (прилож.2) по уровню значимости $\alpha = 0,05$ и числу степеней свободы $\nu = n-k-1 = 17$ находим $t_{\text{табл.}} = t_{\alpha, n-k-1} = t_{0,05,17} = 2,11$. Так как $t_{\text{набл.}} = 15,11 > 2,11$, то нулевая гипотеза $H_0^{(1)}$ отклоняется. Это означает, что температура t оказывает существенное влияние на прочность ДСП.

Аналогично проверяется нулевая гипотеза $H_0^{(2)}: \beta_2 = 0$ против $H_1^{(2)}: \beta_2 \neq 0$. Имеем $t_{\text{набл.}} = \dfrac{|\tilde{\beta}_2|}{S_{\tilde{\beta}_2}} = \dfrac{1,534}{0,4439} = 3,46$. Так как $t_{табл.} = 3,46 > 2,II$, то нулевая гипотеза $H_0^{(2)}$ отклоняется, т.е. фактор продолжительности прессования τ также оказывает существенное влияние на прочность ДСП.

4) Коэффициент множественной корреляции $R = \sqrt{\frac{\beta' A' y}{Q_r}}$, характеризующий отклонение результатов наблюдений от линейной модели, равен $R = \sqrt{24,8588/26,I22} = 0,98$. Так как $R > 0,7$, то считается, что модель достаточно хорошая (адекватная).

6А.2. Результаты замеров температуры объекта y, ^{o}С, температуры x_1 компоненты А в теплоносителе и температуры x_2 окружающей среды:

x_1, ^{o}С	I	-I	2	I	-I	-4	7	0	8	3	6	-2
x_2, ^{o}С	-I	I	-2	-6	-8	5	3	3	0	-I0	2	7
y, ^{o}С	2	0	I	-4	-8	4	II	-2	9	8	I0	5

6А.3. Результаты замеров величин y, x_1, x_2:

x_1	I	4	0	5	-3	3	-5	-I	2	-2
x_2	4	-6	2	-4	I2	-2	I4	6	0	8
y	-4	-5	4	-I	4	0	5	I	2	7

3. Задачи для самостоятельного решения

В задачах 6А.4-6А.I8 для приведенных данных: I) найти уравнение регрессии $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ и доверительные интервалы для параметров β_1, β_2; 2) проверить гипотезу $H_0: \beta_1 = \beta_2 = 0$; 3) проверить гипотезы $H_0^{(j)}$: $\beta_j = 0$, $j = I,2$; 4) вычислить коэффициент множественной корреляции.

Уровень значимости принять $\alpha = 0,05$.

6А.4. Проведено экспериментальное исследование зависимости предела прочности древесины σ, МПа (выходная величина) от изменения температуры t, ^{o}С и влажности w, %. Результаты опытов сведены в таблицу:

i	I	2	3	4	5	6	7	8	9	I0
t, ^{o}С	40	40	40	80	80	80	50	60	70	75
w %	6	I8	30	6	I8	30	I2	24	I2	24
σ, МПа	9	5,5	3,0	7,5	4,2	2	7,I	3,8	6,3	3,4

6А.5. Исследовался процесс пиления древесины ели цепными моторными пилами ЭП4-3 на эстакаде нижнего склада леспромхоза. Приведены результаты замеров производительности пиления П,$см^2$/с,диаметра резца x_1 , см и рабочей длины пильного аппарата x_2 , см:

x_1 ,см	30,5	53	30,5	53	30,5	53	30,5	53
x_2 ,см	48	48	66	66	48	48	66	66
П ,$см^2$/с	24	42,2	33,8	33,8	57,8	5I	5I,7	54,6

6А.6. Исследовался процесс образования прочносвязанного комплекса древесины и её компонентов с фенолфурфуролформальдегидным полимером. Результаты эксперимена:

для целлюлозы

x_1	Количество введенной смолы, %	60	60	20	20	40	50	30	40
x_2	Температура модифицирования, К	333	3I3	333	3I3	323	323	333	333
y	Количество прочносвязанного комплекса, %	5I,3	21,5	I4,6	I4,3	25,3	22,5	26,3	3I,3

6А.7 для древесины

x_1	Количество введенной смолы, %	60	60	20	20	40	50	30	40
x_2	Температура модифицирования, К	333	3I3	333	3I3	323	323	333	333
y	Количество прочносвязанного комплекса, %	I8,4	I4,2	I7,8	I2,9	I6,I5	I6,2	I8,65	I8,8

6А.8. Изучалось влияние вязкости смолы x_1, °Э и давления прессования x_2 , МПа прессованной фанеры на скалывание по клееному слою. Данные замеров приведены в таблице:

x_1	Вязкость смолы, ͨЭ	200	50	200	50	I25	I25	200	50
x_2	Давление прессования, МПа	2,2	2,2	I,6	I,6	I,9	I,6	I,9	I,9
y	Предел прочности, МПа	I,2I	I,00	I,3I	I,22	I,07	I,0I	I,I5	0,99

6А.9. Изучалось влияние вязкости смолы x_1, ͨЭ и температуры прессования x_2, ^oC на предел прочности y на скалывание по клееному слою. Приведены данные измерений:

x_1	Вязкость смолы, ͨЭ	200	200	50	50	I25	I25	200	50
x_2	Температура прессования, ^oC	I50	I30	I50	I30	I40	I30	I40	I40
y	Предел прочности, МПа	I,2I	I,05	I,00	I,42	I,08	I,03	I,I6	I,00

6А.I0. P_z - касательная сила резания, Н; δ - угол резания на основной кромке, град.; α - задний угол резания, град.:

Угол резания δ, град.	50	40	50	40	50	40	55	35	55	35
Задний угол α, град.	7,5	7,5	2,5	2,5	I0	0	7,5	2,5	I0	0
Касательная сила резания P_z, Н	475	364	498	389	475	393	534	37I	5I6	383

6А.II. P_z - касательная сила резания, Н; δ - угол резания на основной кромке, град; α - задний угол резания, град.:

Угол резания δ, град.	45	35	55	45	45	35	40	50	40
Задний угол α, град	5	5	5	0	I0	0	0	2,5	7,5
Касательная сила резания P_z, Н	4I8	359	535	44I	409	383	393	498	364

6А.I2. y - прочность шпона на растяжение поперек волокон, МПа при обжиме древесины в %; x_1 - температура древесины, ^oC; x_2 - толщина шпона, мм:

Температура древесины x_1, °С	40	50	60	70	40	50	60	70	70	50
Толщина шпона x_2, мм	4	4	4	4	2,5	3	5	6	3,5	5
Прочность шпона y, МПа	0,40	0,42	0,44	0,46	0,45	0,43	0,44	0,5	0,45	0,4

6А.I3. y - расход пара в месяц z; x_1 - среднемесячная температура, °С; x_2 - число рабочих дней в месяце:

x_1	I9,7	I6,3	I6,5	32,3	33,7	39,I	40,9	42,I	38,7	3I,6	25,5	I5,8
x_2	20	20	23	20	2I	22	II	23	2I	20	20	2I
y	I098	III3	I25I	8,4	9,27	8,73	6,36	8,5	7,82	9,I4	8,24	I2I9

6А.I4. x_1 - величина зазора затворной пластинки, мм; x_2 - ее температура, °С; y - процент заготовок, которые благополучно проходят контрольную проверку на машинах по упаковке брикета.

Зазор затворной пластинки x_1, мм	I30	I74	I34	I9I	I65	I94	I43	I86	I88	I90	I75	I48
Температура затвор- пластинки x_2, °С	I90	I76	205	2I0	230	I92	220	235	230	220	200	225
Процент заготовок, проходящих зазор y, %	35	8I7	425	983	527	82	345	954	848	833	943	5I,7

6А.I5. L_A - уровень звука при резании в круглопильных станках для продольной распиловки, дб; V - скорость резания, м/с; U_z - подача на зуб, при заданном числе зубьев z = 36 и заданной высоте пропила H = I8 мм, мм.

Скорость резания v, м/с	40	50	60	40	50	60	45	55
Подача на зуб U_z, мм	0,2	0,5	0,8	0,5	0,2	0,5	0,2	0,8
L_1 Уровень шума, дБ	96,03	98,28	I0053	97,66	96,65	98,9	96,84	99,7I

6A.I6. L_1 - уровень звука при резании в круглопильных станках для продольной распиловки при постоянном числе зубьев и постоянной подаче на зуб, дБ; V - скорость резания, м/с; H - высота пропила, мм.

Скорость резания v, м/с	40	50	60	40	50	60	45	55
Высота пропила H, мм	I8	49	80	49	80	I8	49	I8
L_1 Уровень шума, дБ	95,92	95,I3	94,34	94,39	93,6	97,4	94,76	9I,03

6A.I7.

y	476	457	540	55I	575	698	545	574	645	390	560	562
x_1	III	92	90	I07	98	I50	II8	II0	II7	9I	I09	II8
x_2	68	46	50	59	50	66	54	5I	59	44	66	6I

6A.I8.

y	2,6	2,8	2,8	2,9	2,9	3	3,4	3,6	3,7	3,9
x_1	3,9	4,2	4,2	2,6	2,6	9	3,I	9,5	0,6	7,0
x_2	83	64	69	56	73	59	68	92	66	60

Б. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

IБ.I. Даны результаты измерения объемного выхода (%) коротких досок из бревен лиственных пород (короткими считаются доски длиной до I,5 м включительно):

2.I	3.I	4.6	3.8	3.2	3.9	2.4	2.7	4.2	0.9
0.9	3.2	0.9	3.I	3.I	2.8	2.5	I.9	4.3	I.I
3.2	I.3	I.I	3.2	2.5	2.I	3.I	I.6	I.I	I.9
0.8	I.I	I.7	4.I	2.I	2.6	0.9	I.5	I.2	I.I

I.5	0.9	2.3	4.3	4.3	0.7	I.2	0.8	I.8	2.9
4.I	2.9	2.4	4.5	4.5	0.9	I.4	5.I	2.2	3.I
I.7	3.I	0.9	0.8	3.3	I.2	2.6	4.0	2.3	3.2
I.9	3.8	I.3	0.7	2.3	I.I	2.5	4.I	2.7	4.0

Построить гистограммы относительных частот, взяв в качестве длины интервалы следующих значений: а) $h = 0,3$; б) $h = 0,6$; с) $h = 0,9$.

IБ.2. Используя таблицу случайных чисел и таблицу функции $\Phi(x)$, получить выборку объема 50 из генеральной совокупности с нормальным распределением $N(0,I)$. Проведя группировку данных, построить гистограмму, полигон частот и полигон относительных накопленных частот. Найти квантили $x_{0,25}$; $x_{0,5}$ и сравнить с теоретическими значениями.

<u>Указание</u>. Отобрать 50 случайных чисел и разделить каждое на I00. Использовать их как заданные значения $\Phi(x)$ для нахождения соответствующих им значений аргумента. Полученная последовательность чисел образует выборку из генеральной совокупности с нормальным распределением $N(0,I)$.

2Б.I. Доказать следующие свойства выборочного среднего:
а) $\sum_i (x_i-\bar{x})=0$; б) $\sum_i (x_i-a)^2 \quad \sum_i (x_i-\bar{x})^2$, $a\in R$, $a\neq\bar{x}$.

<u>Указание</u> к решению б). Найти значение параметра a, при котором $\sum_{i=1}^{n} (x_i-a)^2$ принимает наименьшее значение.

2Б.2. Доказать равносильность соотношений, определяющих оценку дисперсии:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2;\quad S^2=\frac{n\sum x_i^2-(\sum x_i)^2}{n(n-1)};$$

$$S^2=\frac{1}{n-1}\left(\sum x_i^2-n\bar{x}^2\right).$$

2Б.3. Амплитуда колебаний определялась двумя лаборантами. Первый лаборант по I0 наблюдениям получил среднее значение амплитуды $\bar{x}_1$ = 8I мм, а второй - по I5 наблюдениям - $\bar{x}_2$ = 84 мм. В предположении, что среднее квадратическое отклонение измерений известны и равны σ_1 = 8 мм и σ_2 = 9 мм для первого и второго лаборанта соответственно, найти 99 % доверительный интервал для разности средних $\bar{x}_1$ и $\bar{x}_2$. Можно ли считать, что результаты лаборантов действительно различаются?

2Б.4. Выборка А имеет параметры: n_1 = I0, $\bar{x}_1$ = 3, S_1^2 = 2, а выборка В - n_2 = I5, $\bar{x}_2$ = 2, S_2^2 = 3. Найти среднее и дисперсию выборки С, полученной объединением выборок А и В.

<u>Указание</u>. $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$, $S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$.

2Б.5. Показать, что для элементов выборки системы двух случайных величин выполняется равенство

$$\sum_i x_i(y_i-\bar{y}) = \sum_i y_i(x_i-\bar{x}),$$

$$\sum_i (x_i-\bar{x})(y_i-\bar{y}) = \sum_i x_i y_i - n\bar{x}\bar{y}.$$

2Б.6. Для определения содержания воды (%) в некотором продукте из большой партии взята выборка $n = 25$. Среднее и дисперсия дали $\bar{X} = 7,2$, $S_1^2 = 1.85$. Чтобы уточнить результаты, из этой же партии взяли еще одну выборку $n = 20$, причем оказалось, что $\bar{x}_2 = 7,5$, $S_2^2 = 1,15$. Найти $\bar{x}$ и S^2 для объединенной выборки.

Пусть $f = f(x,\theta)$ – плотность распределения. Тогда функция правдоподобия ФП – $L(x_1,x_2,\dots,x_n,\theta) = f(x_1,\theta)\cdot f(x_2,\theta)\cdots f(x_n,\theta)$, где x_1, x_2, ..., x_n – значения СВ в выборке. Для дискретной СВ ξ

x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

$L(x_1,x_2,\dots,x_m,\theta) = p_1(\theta)\cdot p_2(\theta)\cdots p_m(\theta)$, где $p_i(\theta) = P(\xi = x_i)$.

За оценку максимального правдоподобия параметра θ принимается решение уравнения $\frac{\partial L}{\partial\theta} = 0$.

2Б.7. Осуществлено две серии из n_1 и n_2 независимых испытаний. В первой серии событие А наступило m_1 раз, а во второй – m_2 раз. Найти методом максимального правдоподобия оценку для неизвестной вероятности события А в каждом испытании, считая эту вероятность одной и той же в обеих сериях.

<u>Указание</u>. ФП имеет вид – $L(m_1,m_2,p) = C_{n_1}^{m_1} C_{n_2}^{m_2}\cdot p^{m_1+m_2}\cdot(1-p)^{n_1+n_2-m_1-m_2}$

2Б.8. Выравнять опытные данные при помощи закона распределения с равномерной плотностью:

I	]-1,1]	]1,3]	]3,5]	]5,7]	]7,9]
n	6	7	4	5	8

<u>Указание</u>. $M = \frac{a+b}{2}$; $D = \frac{(b-a)^2}{12}$; a и b находятся из системы $\frac{a+b}{2} = \bar{m}$, $\frac{(b-a)^2}{12} = S^2$.

3Б.9. Приводятся данные об отказах аппаратуры за I0 000 часов:

Число отказов x_i	0	I	2	3	4	5	6
Количество случаев, в которых наблюдалось n_i отказов	427	235	72	2I	I	I	0 .

Проверить гипотезу о том, что число отказов имеет распределение Пуассона ($\alpha = 0,0I$).

Решение. Оценка параметра λ равна среднему числу отказов:

$$\lambda = \bar{x} = \frac{\sum x_i n_i}{n} = \frac{0\cdot 427+I\cdot 235+2\cdot 72+3\cdot 2I+4\cdot I+5\cdot I+6\cdot 0}{757} = 0,6.$$

Вычислим теоретические вероятности p_i появления числа x_i отказов по формуле Пуассона:

$$P_i = P_n(x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \quad ; \quad x_i = 0,\ I,\ 2,\ 3,\ 4,\ 5,\ 6,\ \ldots$$

Вычисления сведём в таблицу:

Число отказов x_i	0	I	2	3	4	5	6
p_i	0,5488I	0,32929	0,09879	0,0I976	0,00296	0,0036	0,0004
np_i	4I6	249	75	I5	2	0	0

Для x_i = 4, 5 и 6 значения $np_i < 5$, поэтому объединим столбцы для x = 3, 4, 5, 6. В результате получим таблицу:

x_i	n_i	np_i	$(n_i - np_i)^2/np_i$
0	427	4I6	0,29I
I	235	249	0,787
2	72	75	0,I20
3	23	I7	2,II8
			$\chi^2_н$ = 3,3I6

Так как по выборке оценивался один параметр λ, то ℓ = I, число степеней свободы ν = 4-I-I = 2. По таблице прилож.3 находим $\chi^2_{0,0I;2}$ = 9,2; следовательно, гипотеза о распределении числа отказов по закону Пуассона принимается.

3Б.I0. В течение I00 дней фиксировалось количество аварий водопроводно-канализационной сети в некотором районе. Получены данные ($n = \sum n_i$ = I00):

Число аварий	Частота	Число аварий	Частота
0	8	3	I8
I	28	4	9
2	3I	5	6

Проверить гипотезу о том, что число аварий имеет распределение Пуассона $\alpha = 0{,}05$.

3Б.II. Во время Второй мировой войны на Лондон упало 537 самолетов-снарядов. Вся территория Лондона была разделена на 576 участков площадью по 0,25 км^2. В таблице приведено число участков i, на которые упало n_i снарядов:

i	0	I	2	3	4	5
n_i	229	2II	93	35	7	I

Согласуются ли эти данные с гипотезой о том, что число снарядов, упавших на каждый из участков, имеет распределение Пуассона? Принять $\alpha = 0{,}05$.

4Б.I2. Для представления некоторых данных предполагается использовать модель $y = \beta_0 + \beta_1 x$, где значение β_1 известно. I) Найти оценку параметра β_0. 2) При известном значении β_0 найти оценку параметра β_1.

4Б.I3. В результате исследований установлено, что между овальностью колец после их обточки x и термической обработкой y существует корреляционная связь, представленная таблицей.

y_i \ x_i	5	I0	I5	20
I0	2	-	-	-
20	5	4	I	-
30	3	8	6	3
40	-	3	6	6
50	-	-	2	I

Найти уравнение регрессии Y на X и оценить тесноту связи признаков X и Y.

5Б.I4. Данные о количестве выпускаемых деталей x (тыс. шт.) и полных затратах на их изготовление y (сотен руб.), полученные на I5 машиностроительных заводах, приведены в таблице:

$x_i \backslash y$	3	4	5	7	8	10	12	13	14	19	20	24	26
2	-	-	-	-	-	-	-	-	-	1	1	1	1
4	1	1	-	-	1	-	-	1	1	-	-	-	-
9	1	-	-	1	-	1	1	-	-	-	-	-	-
10	-	-	1	-	1	-	-	-	-	-	-	-	-

Для каждого значения x найти частную среднюю $\bar{y}_x$; построить эмпирическую кривую (x, $\bar{y}_x$), установить по ней возможность представления рассматриваемой зависимости в виде уравнения регрессии $\bar{y}_x = \frac{a}{x} + b$ и найти его; с помощью корреляционного отношения $\eta_{y/x}$ оценить тесноту связи.

6Б.15. По условию задачи 6А.1 найти уравнение регрессии в виде $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 = \beta_0 + \beta_1 t + \beta_2 \tau = x'\beta$, используя матричную форму записи уравнений.

Решение. Учитывая, что предполагаемая зависимость σ от t и τ линейная (с точностью до случайного слагаемого ε), получаем $\sigma_i = \beta_0 + \beta_1 t_i + \beta_2 \tau_i + \varepsilon_i$, $i = \overline{1,n}$, $\varepsilon \in N(0, \sigma)$, σ считаем неизвестной. В матричном виде это уравнение примет вид $y = \sigma = x'\beta + \varepsilon$, где $y' = $ (36,2; 33,3; 36,5; 33,9; 36,2; 35,3; 36,1; 35,1; 35; 34,1; 34; 32,5; 36,2; 36; 35,2; 34,5; 35,6; 34,6; 34; 33,1);

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad X'X = \begin{pmatrix} n & \Sigma x_{1i} & \Sigma x_{2i} \\ \Sigma x_{1i} & \Sigma x_{1i}^2 & \Sigma x_{1i} x_{2i} \\ \Sigma x_{2i} & \Sigma x_{1i} x_{2i} & \Sigma x_{2i}^2 \end{pmatrix} =$$

$$= \begin{pmatrix} 20 & 3234 & 13,8 \\ 3234 & 52744 & 2226,31 \\ 13,8 & 2226,31 & 10,0242 \end{pmatrix}$$

$$X'y = \begin{pmatrix} \Sigma y_i \\ \Sigma x_{1i} y_i \\ \Sigma x_{2i} y_i \end{pmatrix} = \begin{pmatrix} \Sigma \sigma_i \\ \Sigma t_i \sigma_i \\ \Sigma \tau_i \sigma_i \end{pmatrix} = \begin{pmatrix} 697,4 \\ 112444 \\ 482,342 \end{pmatrix}$$

Вычислим обратную матрицу $(X'X)^{-1}$, для чего сначала вычислим определитель этой матрицы:

$$|X'X| = 20\begin{vmatrix} 527414 & 2226,31 \\ 2226,31 & 10,0242 \end{vmatrix} - 3234\begin{vmatrix} 3234 & 2226,31 \\ 13,8 & 10,0242 \end{vmatrix} +$$

$$+ 13,8\begin{vmatrix} 3234 & 527414 \\ 13,8 & 2226,31 \end{vmatrix} = 20(5286903,4 - 4156456,2) -$$

$$- 3234(32418,262-30723,078) + 13,8(7199886,5-7278313,2) =$$

$$= 20(330447,2)-3234\cdot 1615,184 + 13,8(-78426,7) = 6608944 -$$

$$5482225 - 1082288,4 = 44430,6 \neq 0.$$

Для определения обратной матрицы $(X'X)^{-1}$ найдем алгебраические дополнения A_{ij} элементов матрицы $(X'X)$:

$$A_{11} = \begin{vmatrix} 527414 & 2226,31 \\ 2226,31 & 10,0242 \end{vmatrix} = 330447,2; \quad A_{21} = -\begin{vmatrix} 3234 & 13,8 \\ 2226,31 & 10,0242 \end{vmatrix} = -1695,18$$

$$A_{31} = \begin{vmatrix} 3234 & 527414 \\ 13,8 & 2226,31 \end{vmatrix} = -78426,7; \quad A_{12} = -\begin{vmatrix} 3234 & 2226,31 \\ 13,8 & 10,0242 \end{vmatrix} = -1695,184;$$

$$A_{22} = \begin{vmatrix} 20 & 13,8 \\ 13,8 & 10,0242 \end{vmatrix} = 10,044; \quad A_{32} = -\begin{vmatrix} 20 & 13,8 \\ 3234 & 2226,31 \end{vmatrix} = 103;$$

$$A_{13} = \begin{vmatrix} 3234 & 527414 \\ 13,8 & 2226,31 \end{vmatrix} = -78426,7; \quad A_{23} = -\begin{vmatrix} 20 & 3234 \\ 13,8 & 2226,31 \end{vmatrix} = 103;$$

$$A_{33} = \begin{vmatrix} 20 & 3234 \\ 3234 & 527414 \end{vmatrix} = 89524;$$

$$(X'X)^{-1} = \frac{1}{44430,6} \begin{pmatrix} 330447,2 & -1695,184 & -78426,7 \\ -1695,184 & 10,044 & 103 \\ -78426,7 & 103 & 89524 \end{pmatrix} =$$

$$= \begin{pmatrix} 7.4374 & -0,13815 & -1,76515 \\ -0,03815 & 0,00226 & 0,0023182 \\ -1,76515 & 0,0023182 & 2,037 \end{pmatrix}.$$

Точечная оценка:

$$\tilde{\beta} = (X'X)^{-1}X'y = \frac{1}{44430,6} \begin{pmatrix} 330447,2 & -1695,184 & -78426,7 \\ -1625,185 & 10,044 & 103 \\ -78426,7 & 103 & 89524 \end{pmatrix}$$

$$\begin{pmatrix} 697,4 \\ 112444 \\ 482,342 \end{pmatrix} = \begin{pmatrix} 45,287 \\ -0,071 \\ 1,533 \end{pmatrix}.$$

Итак, $\sigma = 45,287 - 0,071\,t + 1,533\,\tau$.

Точечная оценка коэффициента $\tilde{\beta}_1$ уравнения регрессии означает, что возрастание температуры прессования на 1 % приводит к уменьшению прочности ДСП на 0,071 (МПа), а при неизменных остальных параметрах увеличение продолжительности прессования на 1 мин/мм приводит к увеличению при неизменной температуре прочности ДСП на 1,523 (МПа).

Для нахождения доверительных интервалов для β_0, β_1 и β_2 найдем сначала их средние квадратические отклонения. Точечная оценка S^2 дисперсии σ^2: $S^2 = \frac{y'y - \tilde{\beta}'X'y}{n-k-1}$.

Имеем: $y'y$ = 24344,46; $\tilde{\beta}'X'y$ = (45,287-0,071; 1,533) .

$$\begin{pmatrix} 697,4 \\ 112444 \\ 482,342 \end{pmatrix} = 24339,067;$$

$$S^2 = \frac{24344,46 - 24339,067}{20 - 2 - 1} = \frac{5,4}{17} = 0,3176; \quad S = 0,56.$$

Точечные несмещённые оценки средних квадратических отклонений коэффициентов регрессии $\hat{\beta}_0$, $\tilde{\beta}_1$, $\tilde{\beta}_2$:

$S_{\hat{\beta}_0} = S\sqrt{C_{00}} = 0{,}56\sqrt{7{,}437} = 0{,}56\cdot 2{,}727 = 1{,}527;$

$S_{\hat{\beta}_1} = S\sqrt{C_{11}} = 0{,}56\sqrt{0{,}000226} = 0{,}56\cdot 0{,}015 = 0{,}0084;$

$S_{\hat{\beta}_2} = S\sqrt{C_{22}} = 0{,}56\sqrt{2{,}0149} = 0{,}56\cdot 1{,}427 = 0{,}795,$

где C_{00}, C_{11}, C_{22} - диагональные элементы матрицы $C=(X'X)^{-1}$.

Доверительные интервалы для коэффициентов β_j линейного уравнения регрессии при уровне значимости $\alpha = 0{,}05$ равны:

$\hat{\beta}_j - t_{\alpha,n-1} S_{\hat{\beta}_j} < \beta_j < \hat{\beta}_j + t_{\alpha,n-1} S_{\hat{\beta}_j}$, $j=0,1,2$, $t_{\alpha,n-k-1}$ находим по таблице распределения Стьюдента: $t_{\alpha,n-1} = 2{,}11$,

$45{,}287-2{,}11\cdot 1{,}527 < \beta_0 < 45{,}287+2{,}11\cdot 1{,}527$; $42{,}065 < \beta_0 < 48{,}509$;

$-0{,}071-2{,}11\cdot 0{,}0084 < \beta_1 < -0{,}071+2{,}11\cdot 0{,}0084$; $-0{,}89 < \beta_1 < -0{,}53$;

$1{,}533-2{,}11\cdot 0{,}795 < \beta_2 < 1{,}533+2{,}11\cdot 0{,}795$; $-0{,}144 < \beta_2 < 2{,}21$.

ПРИЛОЖЕНИЯ

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

x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,35	0,I368
0,0I	0,0040	0,36	0,I406
0,02	0,0080	0,37	0,I443
0,03	0,0I20	0,38	0,I480
0,04	0,0I60	0,39	0,I5I7
0,05	0,0I99	0,40	0,I554
0,06	0,0239	0,4I	0,I59I
0,07	0,0279	0,42	0,I628
0,08	0,03I9	0,43	0,I664
0,09	0,0359	0,44	0,I700
0,I0	0,0398	0,45	0,I736
0,II	0,0438	0,46	0,I772
0,I2	0,0478	0,47	0,I808
0,I3	0,05I7	0,48	0,I844
0,I4	0,0557	0,49	0,I879
0,I5	0,0596	0,50	0,I9I5
0,I6	0,0636	0,5I	0,I950
0,I7	0,0675	0,52	0,I985
0,I8	0,07I4	0,53	0,20I9
0,I9	0,0753	0,54	0,2054
0,20	0,0793	0,55	0,2088
0,2I	0,0832	0,56	0,2I23
0,22	0,087I	0,57	0,2I57
0,23	0,09I0	0,58	0,2I90
0,24	0,0948	0,59	0,2224
0,25	0,0987	0,60	0,2257
0,26	0,I026	0,6I	0,229I
0,27	0,I064	0,62	0,2324
0,28	0,II03	0,63	0,2357
0,29	0,II4I	0,64	0,2389
0,30	0,II79	0,65	0,2422
0,3I	0,I2I7	0,66	0,2454
0,32	0,I255	0,67	0,2486
0,33	0,I293	0,68	0,25I7
0,34	0,I33I	0,69	0,2549

Продолжение прилож. I

x	$\Phi(x)$	x	$\Phi(x)$
0,70	0,2580	I,05	0,353I
0,7I	0,26II	I,06	0,3554
0,72	0,2642	I,07	0,3577
0,73	0,2673	I,08	0,3599
0,74	0,2703	I,09	0,362I
0,75	0,2734	I,I0	0,3643
0,76	0,2764	I,II	0,3665
0,77	0,2794	I,I2	0,3686
0,78	0,2823	I,I3	0,3708
0,79	0,2852	I,I4	0,3729
0,80	0,288I	I,I5	0,3749
0,8I	0,29I0	I,I6	0,3770
0,82	0,2939	I,I7	0,3790
0,83	0,2967	I,I8	0,38I0
0,84	0,2995	I,I9	0,3830
0,85	0,3023	I,20	0,3849
0,86	0,305I	I,2I	0,3869
0,87	0,3078	I,22	0,3883
0,88	0,3I06	I,23	0,3907
0,89	0,3I33	I,24	0,3925
0,90	0,3I59	I,25	0,3944
0,9I	0,3I86	I,26	0,3962
0,92	0,32I2	I,27	0,3980
0,93	0,3238	I,28	0,3997
0,94	0,3264	I,29	0,40I5
0,95	0,3289	I,30	0,4032
0,96	0,33I5	I,3I	0,4049
0,97	0,3340	I,32	0,4066
0,98	0,3365	I,33	0,4082
0,99	0,3389	I,34	0,4099
I,00	0,34I3	I,35	0,4II5
I,0I	0,3438	I,36	0,4I3I
I,02	0,346I	I,37	0,4I47
I,03	0,3485	I,38	0,4I62
I,04	0,3508	I,39	0,4I77

Продолжение прилож. I

x	$\Phi(x)$	x	$\Phi(x)$
2,20	0,486I	2,68	0,4963
2,22	0,4868	2,70	0,4965
2,24	0,4875	2,72	0,4967
2,26	0,488I	2,74	0,4969
2,28	0,4887	2,76	0,497I
2,30	0,4893	2,78	0,4973
2,32	0,4898	2,80	0,4974
2,34	0,4904	2,82	0,4976
2,36	0,4909	2,84	0,4977
2,38	0,49I3	2,86	0,4979
2,40	0,49I8	2,88	0,4980
2,42	0,4922	2,90	0,498I
2,44	0,4927	2,92	0,4982
2,46	0,493I	2,94	0,4984
2,48	0,4934	2,96	0,4985
2,50	0,4938	2,98	0,4986
2,52	0,494I	3,00	0,49865
2,54	0,4945	3,20	0,4993I
2,56	0,4948	3,40	0,49966
2,58	0,495I	3,60	0,49984I
2,60	0,4953	3,80	0,499928
2,62	0,4956	4,00	0,499968
2,64	0,4959	4,50	0,499997
2,66	0,496I	5,00	0,499997

2. t – распределение Стьюдента

Число степеней свободы, N	Уровень значимости (двусторонняя критическая область)		Число степеней свободы, N	Уровень значимости (двусторонняя критическая область)	
	0,I0	0,05		0,I0	0,05
I	2	3	4	5	6
I	6,3I	I2,7	5	2,0I	2,57
2	2,92	4,30	6	I,94	2,45
3	2,35	3,I8	7	I,89	2,36
4	2,I3	2,78	8	I,86	2,3I

Продолжение прилож. 2

I	2	3	4	5	6
9	I,83	2,26	22	I,72	2,07
I0	I,8I	2,23	23	I,7I	2,07
II	I,80	2,20	24	I,7I	2,06
I2	I,78	I,I8	25	I,7I	2,06
I3	I,77	2,I6	26	I,7I	2,06
I4	I,76	2,I4	27	I,7I	2,05
I5	I,75	2,I3	28	I,70	2,05
I6	I,75	2,I2	29	I,70	2,05
I7	I,74	2,II	30	I,70	2,04
I8	I,73	2,I0	40	I,68	2,02
I9	I,73	2,09	60	I,67	2,00
20	I,73	2,09	I20	I,66	I,98
2I	I,72	2,08		I,64	I,96
	0,05	0,025		0,05	0,025
ν	Уровень значимости (односторонняя критическая область) α		ν	Уровень значимости (односторонняя критическая область) α	

3. χ^2- распределение

Число степеней свободы.	Уровень значимости			Число степеней свободы,	Уровень значимости		
	0,0I	0,025	0,05		0,0I	0,025	0,05
I	6,6	5,0	3,8	I6	32,0	28,8	26,3
2	9,2	7,4	6,0	I7	33,4	30,2	27,6
3	II,3	9,4	7,8	I8	34,8	3I,5	28,9
4	I3,3	II,I	9,5	I9	36,2	32,9	30,I
5	I5,I	I2,8	II,I	20	37,6	34,2	3I,4
6	I6,8	I4,4	I2,6	2I	38,9	35,5	32,7
7	I8,5	I6,0	I4,I	22	40,3	36,8	33,9
8	20,I	I7,5	I5,5	23	4I,6	38,I	35,2
9	2I,7	I9,0	I6,9	24	43,0	39,4	36,4
I0	23,2	20,5	I8,3	25	44,3	40,6	37,7
II	24,7	2I,9	I9,7	26	45,6	4I,9	38,9
I2	26,2	23,3	2I,0	27	47,0	43,2	40,I
I3	27,7	24,7	22,4	28	48,3	44,5	4I,3
I4	29,I	26,I	23,7	29	49,6	45,7	42,6
I5	30,6	27,5	25,0	30	50,9	47,0	43,9

ЛИТЕРАТУРА

1. Бородич Л.И., Герасимович А.И., Кеда Н.П., Млечко И.Н. Справочное пособие по приближенным методам решения задач высшей математики. - Мн.: Вышэйшая школа, 1986. - 189 с.

2. Сборник задач по математике для втузов. Специальные курсы. - М.: Наука, 1984. - 608 с.

3. Статистические методы обработки наблюдений. - Л.: ЛТИ, 1977. - 73 с.

4. Пижурин А.А., Розенблит М.С. Исследование процессов деревообработки. - М.: Лесная пром-сть, 1984. - 231 с.

5. Айвазян С.А. и др. Прикладная статистика. - М.: Финансы и кредит. Т.I. - Первичная обработка, 1983. - 471 с. Т. 2. - Исследование зависимостей, 1985. - 487 с.

6. Поллард Дж. Справочник по вычислительным методам статистики. - М.: Финансы и статистика, 1982. - 344 с.

7. Гмурман В.Е. Теория вероятностей и математическая статистика. - М.: Высшая школа, 1977. - 479 с.

ОГЛАВЛЕНИЕ