Министерство образования и науки РФ
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«ТОМСКИЙГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР) »

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания к лабораторным работам
для студентов направления 27.03.05 Инноватика.

Томск 2018

Стариков Виталий Иванович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА
Методические указания к лабораторным работам
для студентов направления 27.03.05. Инноватика

Содержание

Введение

Тематика работ охватывает все положения из рабочей программы учебной дисциплины «Теория вероятностей и математическая статистика» , уровень образовательной программы «бакалавриат», направление подготовки (специальность) - 27.03.05 Инноватика.

Цель проведения лабораторных работ – закрепить знания, полученные на лекциях, научить студентов решать задачи с использований различных программных сред, таких как “ Mathematica” “, Mathcad, “Excel”. Примеры, представленные в настоящих лабораторных работах, решены в программной среде “ Mathematica”. Обозначения, используемые в примерах, соответствуют этой программной среде.

Лабораторные работы могут быть выполнены с любым пакетом программ.

Форма отчетности для всех лабораторных работ- защита отчета.

Лабораторная работа «Практика расчета вероятностей сложных событий»

Цель работы: Проведение расчетов вероятностей сложных событий с использованием формулы классического определения вероятности. Применение формулы полной вероятности и формулы Бейеса.

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 2 часа.

Теоретические основы и примеры.

Обозначения.

$C_N^m = Binomial[N, m]$- число сочетаний из N элементов по m.

Основные формулы.

1. При классическом определении вероятность события A определяется равенством

$$P = m/n \qquad (1)$$

в котором m – число элементарных исходов испытания, благоприятствующих появлению события A , n – общее число возможных элементарных исходов испытания.

2. **Формула полной вероятности**. Вероятность $P(A)$ события A, которое может наступить лишь при появлении одного из несовместных событий (гипотез) H_1, H_2 , …H_n, образующих полную группу ($P(H_1) + P(H_2) +\ldots+ P(H_n) =1$) , равна

$$P(A) = P(H_1)\, P(A/H_1) + P(H_2)\, P(A/H_2) +\ldots+ P(H_n)\, P(A/H_n) \qquad (2)$$

В этой формуле $P(A/H_i)$ – условные вероятности, т.е. вероятности появления события A при условии, что событие H_i совершилось.

3. **Формула Бейеса**. Событие A может наступить при условии появления одного из несовместных событий (гипотез) H_1, H_2 , …H_n. Если событие A уже произошло, то вероятности гипотез могут быть вычислены по формуле

$$P(H_i/A) = P(H_i) \cdot P(A/H_i) / P(A) \qquad (3)$$

Стандартная задача 1. В ящике N деталей, помеченных номерами 1, 2, ...N . Наудачу отобраны m_1 деталей. Найти вероятность того, что среди отобранных деталей есть детали а) с номером k_1 б) с номерами k_1 и k_2.

Задача решается с использованием формулы (1), в которой общее число возможных элементарных исходов, число способов, которыми можно извлечь m_1 деталей из N деталей определяется как

$$n = \texttt{Binomial[N,}m_1\texttt{]}, \tag{4}$$

а число благоприятных исходов определяется как

$$m = \text{Binomial[N-k ,}m_1\text{ -k]}, \tag{5}$$

так что

$$P=\texttt{Binomial[N-k,}m_1\texttt{-k]/Binomial[N,}m_1\texttt{]}.$$

Пример 1.1 В ящике 10 одинаковых деталей, помеченных номерами 1, 2, ...10. Наудачу извлечены 6 деталей. Найти вероятность того, что среди извлеченных окажутся : A- деталь №1, B -детали №1 и №2.

Решение. В этом примере N =10, m_1 =6, k_1 =1, k_2 = 2. По формуле (4) n=Binomial[10,6] = 210. а) При событии A остальные 5 деталей имеют другие номера, а число способов, которыми можно отобрать 5 деталей из оставшихся 9 есть m=Binomial[9,5] =126. Это и есть число исходов, благоприятствующих появлению события A , следовательно, для события A $P = m/n$ = Binomial[9,5]/Binomial[10,6] = 3/6 =0.6. Аналогично для события B $P = m/n$ = Binomial[8,4]/Binomial[10,6] = 1/3 =0.333

Стандартная задача 2. В партии из N деталей имеется n_1 стандартных. Наудачу отобраны m_1 деталей. Найти вероятность того, что среди отобранных деталей ровно k стандартных.

Задача решается с использованием формулы (1), в которой

$$n = \texttt{Binomial[}N, m_1\texttt{]}, \tag{6}$$

а число благоприятствующих событий вычисляется по формуле

$$m = \texttt{Binomial}[n_1, k] * \texttt{Binomial}[N - n_1, m_1 - k] \quad (7)$$

Пример 1.2 На стеллаже библиотеки в случайном порядке расставлено N =15 учебников, причем n_1 =10 из них - в переплете. Студент наудачу берет m_1 =5 учебников. Найти вероятность того, что среди отобранных учебников хотя бы один окажется в переплете.

Решение. Задача решается двумя способами.

Способ 1. Событие A (хотя бы один учебник окажется в переплете), реализуется, если реализуются или событие B (k =1 учебник окажется в переплете), или C (k =2 учебника окажется в переплете), или D (k =3 учебника окажется в переплете), или E (k =4 учебника окажется в переплете), или L (k =5 учебника окажется в переплете), поэтому $P(A)=P(B)+P(C)+P(D)+P(E)+P(L)$. Полагая в (6) N =15, m_1 =5 находим n = 3003 и в (7) для n_1 =10, k = 1,2,3, 4,5 находим m = 50, 450, 1200, 1050 и 252, так что

$P(A)$= 0.01665+ 0.1498+ 0.3996 +0.3496+0.0839= 0.9996.

Способ 2. Находим вероятность противоположного события $\overline{A}$ - ни один из взятых учебников не имеет переплета. Полагаем в формуле (7) k =0 находим $P(\overline{A}) = 0.0004$, так что $P(A) = 1 - P(\overline{A}) = 0.9996$.

Пример 1.3. В каждой из 3-х урн содержится 6 черных и 4 белых шара. Из 1-й урны наудачу извлечен шар и перекладывается во 2-ю урну, после чего из второй урны наудачу извлечен шар и перекладывается в 3-ю урну. Найти вероятность того, что шар, наудачу извлеченный из 3-й урны, окажется белым.

Решение. Задача решается с помощью формулы (2) для полной вероятности. Гипотеза H_1 – из 1-й во 2-ю урну переложен белый шар и из 2-й- в 3-ю то же белый, условно Б $(1\rightarrow 2) \rightarrow$ Б$(2\rightarrow 3)$, гипотеза H_2 : Б $(1\rightarrow 2) \rightarrow$ Ч$(2\rightarrow 3)$ (из 2-й- в 3-ю переложен черный шар), гипотеза H_3: Ч (1→2) → Б(2→3), гипотеза H_4: Ч (1→2) → Ч(2→3). Результаты расчетов оформляем в виде таблицы.

Гипотезы	$P(H_i)$	$P(A/H_i)$	$P(H_i)\cdot P(A/H_i)$
H_1	(4/10)·(5/11)	5/11	10/121
H_2	(4/10)·(6/11)	4/11	48/605
H_3	(6/10)·(4/11)	5/11	12/121
H_4	(6/10)·(7/11)	4/11	84/605
			$\sum$=2/5=0.4

Пример 1.4. В пирамиде N= 8 винтовок, из которых n = 2 винтовок снабжены оптическим прицелом. Вероятность того, что мишень поражена винтовкой с оптическим прицелом равна p_1 = 0.98, а без прицела p_2= 0.75. Мишень была поражена. Найти вероятности событий: 1) стреляли из винтовки с оптическим прицелом, 2) стреляли из винтовки без оптического прицела.

Решение. Задача решается с помощью формулы Бейеса (3). Гипотеза H_1 – мишень поражена винтовкой с оптическим прицелом, гипотеза H_2 – без прицела. Число винтовок без прицела $N-n$ = 2. Вероятность $P(A) = (n/N)\cdot p_1 + [(N-n)/N]\cdot p_2 = 0.8075$;

По формуле Бейеса (3) $P(H_1/A) = P(H_1)\cdot P(A/\mathrm{H}_1)/P(A) = p_1\cdot(n/N)/P(A) = 0.303$;

$P(H_2/A) = P(H_2)\cdot P(A/\mathrm{H}_2)/P(A) = p_2\cdot[(N-n)/N]/P(A) = 0.696$.

Задачи к Лабораторной работе «Практика расчета вероятностей сложных событий»**.**

Задача 1.1 Брошены 2 игральные кости. Найти вероятности следующих событий: а) сумма выпавших очков равна N; б) сумма выпавших очков равна N_1, а разность n_1; в) сумма выпавших очков равна N_2, если известно, что их разность равна n_2; г) сумма выпавших очков равна N_3, а произведение $-n_3$. Данные взять из Таб. 1.1

Таблица 1.1 . Данные к задаче 1.1.

Вариант	N	N_1	n_1	N_2	n_2	N_3	n_3
1	7	8	4	8	4	5	4
2	5	9	3	5	3	6	3
3	4	10	2	6	2	7	2
4	7	11	4	8	4	8	5
5	8	12	5	9	5	9	6
6	9	6	6	10	6	10	7

Продолжение

Таблица 1.1 . Данные к задаче 1.1.

7	10	7	3	11	7	11	8
8	11	8	4	12	4	12	3
9	12	9	5	5	5	2	4
10	2	10	6	6	3	3	5
11	5	11	3	7	4	4	6
12	6	12	4	8	5	5	7
13	7	5	5	9	3	6	8
14	8	6	6	10	4	7	3
15	9	7	3	12	5	8	4
16	10	8	4	5	6	9	5
17	11	9	5	6	3	10	6
18	12	10	6	7	4	11	3
19	4	11	3	8	5	12	4
20	5	12	4	9	6	5	2

Задача 1. 2. В ящике N деталей, помеченных номерами 1, 2, ...N . Наудачу отобраны m деталей. Найти вероятность того, что среди отобранных деталей есть детали а) с номером k_1 б) с номерами k_1 и k_2. Данные взять из Таб. 1.2

Таблица 1.2 . Данные к задаче 1.2.

Вариант	N	m	k_1	k_2
1	10	6	4	5
2	12	5	2	3
3	8	4	2	6
4	7	4	3	2
5	8	5	2	3
6	8	5	1	3
7	10	6	2	3
8	10	5	4	1
9	12	9	5	4
10	10	7	3	4
11	11	8	3	7
12	6	4	4	2
13	8	5	3	4
14	8	6	6	2

Таблица 1.2 . Данные к задаче 1.2.

15	9	7	3	1
16	10	8	4	5
17	11	9	5	6
18	12	10	6	3
19	4	3	1	2
20	5	3	2	1

Задача 1.3. В партии из N деталей имеется n стандартных. Наудачу отобраны m деталей. Найти вероятность того, что среди отобранных деталей ровно k стандартных. Данные взять из Таб. 1.3

Таблица 1.3. Данные к задаче 1.3.

Вариант	N	n	m	k
1	10	6	3	1
2	11	7	4	2
3	12	8	5	3
4	13	9	6	4
5	14	10	7	5
6	15	11	8	6
7	16	12	9	7
8	17	13	10	8
9	18	14	11	9
10	19	15	12	10
11	20	16	13	11
12	21	17	14	12
13	22	18	15	13
14	23	19	16	14
15	24	20	17	15
16	25	21	18	16
17	26	22	19	17
18	27	23	20	18
19	28	24	21	19
20	29	25	22	20

Задача. 1.4 На стеллаже библиотеки в случайном порядке расставлено N учебников, причем n из них - в переплете. Студент наудачу берет m учебников. Найти вероятность того, что среди отобранных учебников хотя бы один окажется в переплете. Данные взять из Таб. 1.4

Таблица 1.4. Данные к задаче 1.4.

Вариант	N	N	m	k
1	10	6	3	1
2	11	7	4	2
3	12	8	5	3
4	13	9	6	4
5	14	10	7	5
6	15	11	8	6

Таблица 1.4. Данные к задаче 1.4.

7	16	12	9	7
8	17	13	10	8
9	18	14	11	9
10	19	15	12	10
11	20	16	13	11
12	21	17	14	12
13	22	18	15	13
14	23	19	16	14
15	24	20	17	15
16	25	21	18	16
17	26	22	19	17
18	27	23	20	18
19	28	24	21	19
20	29	25	22	20

Задача 1.5. В каждой из 3-х урн содержится n черных и m белых шаров. Из первой урны наудачу извлечен один шар и перемещен во вторую урну, после чего из второй урны наудачу извлечен один шар и переложен в третью урну. Найти вероятность того, что шар, наудачу извлеченный из третьей урны, окажется белым. Данные взять из Таб. 1.5

Таблица 1.5. Данные к задаче 1.5.

Вариант	n	M	Вариант	n	m
1	7	5	11	17	15
2	8	6	12	18	16
3	9	7	13	19	17
4	10	8	14	20	18
5	11	9	15	21	19
6	12	10	16	22	20
7	13	11	17	23	21
8	14	12	18	24	22
9	15	13	19	25	23
10	16	14	20	26	24

Задача 1.6. Две машинистки набирают один и тот же текст. Вероятность того, что первая машинистка допустит ошибку равна p_1 , а вторая - p_2 . При сверке текста была обнаружена ошибка. Найти вероятность того, что ошиблась первая машинистка. Данные взять из Таб. 1.6

Таблица 1.6. Данные к задаче 1.6.

Вариант	p_1	p_2	Вариант	p_1	p_2
1	0.05	0.1	11	0.1	0.05
2	0.15	0.2	12	0.2	0.15
3	0.25	0.3	13	0.3	0.25
4	0.35	0.4	14	0.4	0.35
5	0.45	0.5	15	0.5	0.45
6	0.55	0.6	16	0.6	0.55
7	0.65	0.7	17	0.7	0.65
8	0.75	0.8	18	0.8	0.75
9	0.85	0.9	19	0.9	0.85
10	0.95	1.0	20	1.0	0.95

Лабораторная работа «Числовые характеристики случайных величин».

Цель работы:

1. Построение ряда распределения для дискретной случайной величины.

   Определение функции распределения для непрерывной случайной величины.

2. Расчет числовых характеристик дискретной или непрерывной случайной величины.

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 2 часа.

Теоретические основы и примеры.

$M(X) \equiv m_x$ - математическое ожидание дискретной случайной величины X. Вычисляется по формуле

$$M(X) = \sum_{i=1}^{n} x_i p_i \qquad (1)$$

$D(X) \equiv D_x$ - дисперсия дискретной случайной величины, вычисляется по формуле

$$D(X) = M(X^2) - [M(X)]^2 = \sum_{i=1}^{n} x_i^2 p_i - (\sum_{i=1}^{n} x_i p_i)^2 \qquad (2)$$

Среднеквадратичное отклонение дискретной случайной величины вычисляется по формуле

$$\sigma(X) = [D(X)]^{1/2} \qquad (3)$$

Функция распределения вероятностей $F(x)$ определяется соотношением

$$F(x) = P(X<x) \qquad (4)$$

и обладает свойствами:

1)

$$0 \le F(x) \le 1 \qquad (5а)$$

2)

$$F(x_2) \ge F(x_1) \text{ для } x_2 > x_1 \qquad (5б)$$

3) Если все возможные значения X принадлежат интервалу (a, b), то

$$F(x) = 0 \text{ при } x \le a, \; F(x) = 1 \text{ при } x \ge b. \qquad (5г)$$

Для непрерывной случайной величины

$$M(X)=\int\limits_{-\infty}^{\infty} x\rho(x)dx \tag{6}$$

$$D(X)=\int\limits_{-\infty}^{\infty}[x-M(X)]^2\rho(x)dx=\int\limits_{-\infty}^{\infty}x^2\rho(x)dx-[M(X)]^2 \tag{7}$$

где $\rho(x)$- плотность распределения вероятностей непрерывной случайной величины с условием нормировки

$$\int\limits_{-\infty}^{\infty}\rho(x)dx=1 \tag{8}$$

Функции $F(x)$ и $\rho(x)$ связаны соотношением

$$F(x)=\int\limits_{-\infty}^{x}\rho(x)dx \tag{9}$$

$$\rho(x)=F'(x) \tag{10}$$

Пример 2.1 В урне 6 белых и 4 черных шара. Из нее извлекают 3 шара. Случайная величина X - число белых шаров среди извлеченных. Найти: 1) ряд распределения X; 2) функцию распределения $F(x)$; $F(0.2)$; $F(2.5)$; 3) m_x, D_x ;4) $P(0.2 < X < 2.5)$.

Решение.

1) Случайная величина X может принимать значения $k =$ 0, 1, 2, 3. Вероятности $P(X=k)$ вычисляются с помощью формул (6) и (7) Лабораторной работы «Практика расчета вероятностей сложных событий» , а именно

$$P(X=k)=\frac{C_n^k\cdot C_{N-n}^{m-k}}{C_N^m}$$

В нашем примере N =10, n =6, m =3, так что ряд распределения X имеет вид

X	0	1	2	3
p	1/30	3/30	½	1/6

2) По формуле (4) определяем (к предыдущей значению прибавляем последующее значение для вероятности, начиная с 0) значения $F(x)$

$$F(x)=\begin{cases}0, & \text{если } x\le 0\\ 1/30, & \text{если } 0<x\le 1\\ 1/3, & \text{если } 1<x\le 2\\ 5/6, & \text{если } 2<x\le 3\\ 1, & \text{если } x>1\end{cases}$$

$F(0.2)$ = 1/30 и $F(2.5)$ =5/6 находим из соответствующих строчек для $F(x)$.

3) m_x = 9/5, D_x = 14/25 находятся по формулам (1) и (2).

4) $P(0.2 < X < 2.5) = F(2.5) - F(0.2) = 5/6 - 1/30 = 4/5$.

Пример 2.2 Дана функция распределения случайной величины X

$$F(x) = \begin{cases} 0, & \text{если } x \le 0 \\ ½\, x, & \text{если } 0 < x \le 1 \\ 1/3 + Ax^2, & \text{если } 1 < x \le 2 \\ B, & \text{если } x > 2 \end{cases}$$

Найти: 1**)** Константы A и B ; 2) плотность распределения $\rho(x)$; 3)m_x , D_x ; 4) $P(1/4 < X < 2.0)$.

1) Константа B =1.0 находится из свойства (5с) $F(x)$ =1 для $x \ge 2$.

2) плотность $\rho(x)$ определяем по формуле (10) через производную от $F(x)$

$$\rho(x) = F'(x) = \begin{cases} 0, & \text{если } x < 0 \\ ½\,, & \text{если } 0 < x \le 1 \\ 2Ax, & \text{если } 1 < x \le 2 \\ 0, & \text{если } x > 2 \end{cases}$$

Из условия нормировки (8) $\int_{-\infty}^{\infty} \rho(x)dx = 1 = \int_0^1 (1/2)dx + \int_1^2 2Axdx = 1/2 + 3A = 1$ находим A =1/6.

3) По формуле (6) $m_x = \int_{-\infty}^{\infty} x\rho(x)dx = \int_0^1 x \cdot (1/2)dx + \int_1^2 x \cdot 2 \cdot \frac{1}{6} xdx = 37/36$, аналогично по формуле (7) D_x = 467/1296.

4) $P(1/4 < X < 2.0) = F(2) - F(1/4) = \{1/3 + (1/6)\cdot 4\} - \{(1/2)\cdot(1/4)\} = 7/8$

Задачи к лабораторной работе «Числовые характеристики случайных величин».

Задача 2.1 В партии из N деталей имеется n стандартных. Наудачу отобраны m деталей. X – случайная величина- число стандартных деталей среди отобранных. Найти:

1) Ряд распределения для X ; 2) функцию распределения (построить графически); 3) $M(X)$ и $D(X)$. Входные данные взять из Таблицы 2.1

Указание. Воспользоваться решением задачи 1. 3 Лабораторной работы «Практика расчета вероятностей сложных событий»

Таблица 2. 1. Данные к задаче 2.1.

Вар.	*N*	*n*	*m*
1	10	4	3
2	8	5	3
3	8	4	2
4	9	5	2
5	12	6	3
6	12	5	4
7	11	5	3
8	10	4	2
9	14	8	2
10	15	9	2
11	6	3	2
12	9	3	3
13	6	2	2
14	8	5	2
15	10	5	3
16	12	4	3
17	12	6	4
18	9	8	2
19	8	6	3
20	8	6	4

Задача 2.2 Задана плотность распределения случайной величины *X*

$$\rho(x) = \begin{cases} 0, \text{ если } x \le 0 \\ A\cdot\ (x^a + d\cdot x^b), \quad \text{если } 0 < x \le \text{e} \\ 0, \text{ если } x\ > e \end{cases}$$

Найти 1) константу A ;2) функцию $F(x)$; 3) $F(1/2)$; 4) $m_x \equiv M(X)$;, $D\ (X) \equiv D_x$. Входные данные взять из Таблицы 2.2.

Таблица 2. 2. Данные к задаче 2.2.

Вар.	*a*	*b*	*d*	*e*
1	2	1	2	1
2	2	1	2	2
3	2	1	3	2
4	2	1	3	3
5	2	1	3	4
6	2	1	3	5
7	2	1	4	4
8	2	1	4	5
9	1	0	4	5
10	2	0	1	1
11	2	0	1	2
12	2	0	1	3
13	2	0	1	4
14	2	0	2	1
15	2	0	3	1
16	2	0	4	1
17	3	0	1	1
18	3	1	1	1
19	3	2	1	1
20	3	2	1	2

Лабораторная работа «Числовые характеристики двумерной случайной величины» .

Цель работы

Построение рядов распределения для составляющих X и Y из двумерной дискретной случайной величины (С.В.) (X, Y) , отыскание числовых характеристик системы (X, Y) , нахождение условных законов распределения вероятностей составляющих дискретной и непрерывной случайной величины.

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 2 часа.

Теоретические основы и примеры.

А. Дискретная двумерная С.В. задается таблицей

Y / X	x_1	x_2	...	x_n
y_1	p_{11}	p_{21}	...	p_{n1}
y_2	p_{12}	p_{22}	...	p_{n2}
…	…	….		….
y_m	p_{1m}	p_{2m}	...	p_{nm}

с вероятностями p_{ij} того, что С.В. (X, Y) примет значение $(X=x_i, Y=y_j)$; индекс i =1,2,… n нумерует столбцы (для X) , индекс j = 1,2,… m нумерует строчки для (Y)

1. Ряды распределения для X и Y отдельно определяются как

$$\begin{array}{l} X \;\; x_1\, x_2 \,\dots\, x_n \\ p \;\; p_1\, p_2 \,\dots\, p_n \end{array} \qquad\qquad \begin{array}{l} Y \;\; y_1\, y_2 \,\dots\, y_m \\ p \;\; p_1\, p_2 \,\dots\, p_m \end{array} \tag{1}$$

Здесь $p_i = \sum_{j=1}^{m} p_{ij}$ (суммируем вероятности по столбцам), $p_j = \sum_{i=1}^{n} p_{ij}$ (суммируем вероятности по строчкам).

2. Условные законы распределения находятся по условным вероятностям составляющих X и Y

$$\begin{array}{c} P(Y=y_j/ X= x_i) \equiv p(y_j /x_i) = P(X= x_i , Y=y_i)/ P(X= x_i) \equiv p(x_i ,y_j)/ p(x_i) \\ P(X= x_i / Y=y_j) \equiv p(x_i/ y_j) = P(X= x_i \;\; Y=y_i)/P(Y= y_j) \equiv p(x_i ,y_j)/ p(y_j) \end{array} \tag{2}$$

3. Условные математические ожидания

$$M[Y / X = y_i] = \sum_{j=1}^{m} y_j\, p(y_j / x_i) = \sum_{i=1}^{n} y_j\, p_{ji} / p_i \tag{3}$$

$$M[X / Y = y_j] = \sum_{i=1}^{n} x_i\, p(x_i / y_j) = \sum_{i=1}^{n} x_i\, p_{ij} / p_j$$

4. Ковариация С.В.

$$Cov(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i-m_x)(y_j-m_y)p_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_{ij}-m_x\cdot m_y \tag{4}$$

5. Коэффициент корреляции

$$r_{xy}=Cov(X,Y)/(\sigma_x\cdot\sigma_y) \tag{5}$$

6. Мат.ожидание, дисперсия и среднеквадратичные отклонения

$$m_x=\sum_{i=1}^{n}x_i p_i\ ,\ D_x=\sum_{i=1}^{n}x_i^2 p_i-m_x^2\quad \sigma_x=D_x^{1/2}$$
$$m_y=\sum_{j=1}^{m}y_j p_j\ ,\ D_y=\sum_{j=1}^{m}y_i^2 p_j-m_y^2\quad \sigma_y=D_y^{1/2} \tag{6}$$

Б. Непрерывная двумерная С.В.

1. Плотность распределения одной из составляющей

$$\rho_1(x)=\int_{-\infty}^{\infty}\rho(x,y)dy\quad \rho_2(y)=\int_{-\infty}^{\infty}\rho(x,y)dx \tag{7}$$

2. Математические ожидания

$$m_x=\int_{-\infty}^{\infty}x\cdot\rho_1(x)dx\ ,\ m_y=\int_{-\infty}^{\infty}y\cdot\rho_2(y)dy \tag{8}$$

3 Дисперсия

$$D_x=\int_{-\infty}^{\infty}x^2\cdot\rho_1(x)dx-m_x^2\ ,\ D_y=\int_{-\infty}^{\infty}y^2\cdot\rho_2(y)dy-m_y^2 \tag{9}$$

4 Ковариация

$$Cov(X,Y)_x=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\cdot y\cdot\rho(x,y)dxdy-m_x\cdot m_y \tag{10}$$

5. Коэффициент корреляции

$$r_{xy}=Cov(X,Y)/(\sigma_x\cdot\sigma_y) \tag{11}$$

где $\sigma_x=D_x^{1/2}$, $\sigma_y=D_y^{1/2}$

6. Условные плотности распределения

$$\rho(x/y)=\rho(x,y)/\rho_2(y),\quad \rho(y/x)=\rho(x,y)/\rho_1(x) \tag{12}$$

Пример 3.1 Дана матрица распределения вероятностей системы (*X*, *Y*)

Y	*X*		
	1	2	3
1	0.25	0.11	0.16
2	0.13	0.2	0.15

Найти: 1) ряды распределения для X и Y, 2) m_x, m_y, D_x, D_y, 3) $Cov(X,Y)$, 4) r_{xy} , 5) $M\,[Y/X=1]$.

Решение.

1) По формулам (1)

X	1	2	3
p	0.38	0.31	0.31=0.16+0.15

Y	1	2
p	0.52	0.48=0.13+0.2+0.15

2) Для каждой случайной величины X или Y математическое ожидание и дисперсия вычисляется по формулам (1) и (2) Лабораторной работы «Числовые характеристики случайных величин» , например, для X используем полученный ряд распределения и находим

$m_x = 1\cdot 0.38+2\cdot 0.31+3\cdot 0.31=1.93;$

$D_x = 1^2\cdot 0.38+2^2\cdot 0.31+3^2\cdot 0.31-(m_x)^2 = 0.685$

Аналогично $m_y = 1.48$; $D_y = 0.2496$;

3) По формуле (4)

$Cov(X,Y) = 0.0536.$

4) По формуле (5)

$r_{xy} = 0.13$

5) По формуле (3)

$M[Y/X=1] = 1\cdot 0.25/0.38+2\cdot 0.13/0.38=1.342$

Пример 3.2 Дана плотность распределения вероятностей системы (X, Y)

$\rho\,(x,y) = C$ внутри треугольника $O(0,0)$, $A(-3,0)$, $B(0,2)$ и $\rho\,(x,y)=0$ в других точках.

Найти: 1) константу C ;2) $\rho_1(x)$, $\rho_2(y)$; 3) m_x, m_y, D_x, D_y, ; 4) r_{xy} .

Решение.

1) Рисуем область (рис. 3.1) ,в которой задана плотность $\rho\,(x,y) = C$ и составляем уравнения прямых (по двум точкам), ограничивающих эту область : $y=0$ (нижняя

граница), $y = 2x/3 + 2$ (верхняя граница) , $x = -3$ (левая) и $x = 0$ (правая) границы области.

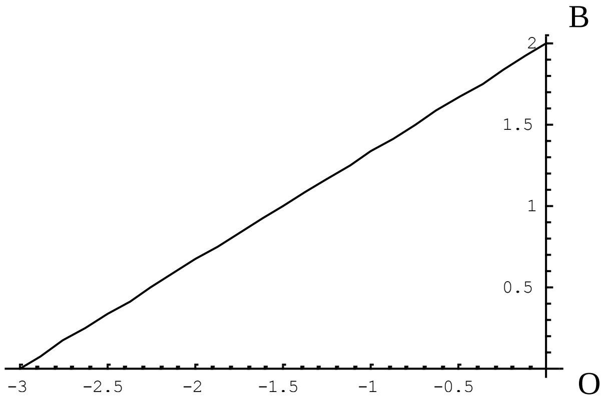

Рис. 3.1 Область ρ $(x,y) = C$, ограниченная треугольником с вершинами O(0,0), A(-3,0), B(0,2).

Используем условия нормировки $\iint\limits_D \rho(x,y)dxdy = 1$. Расставляя пределы интегрирования, находим

$$\iint\limits_D \rho(x,y)dxdy = \iint\limits_D Cdxdy = C\cdot\int_{-3}^{0}dx\int_{0}^{2x/3+2}dy = C\cdot\int_{-3}^{0}(2x/3+2)dx = 3\cdot C = 1$$

Откуда $C = 1/3$.

2) По формулам (7)

$$\rho_1(x) = \int_{-\infty}^{\infty}\rho(x,y)dy = \int_{0}^{2x/3+2}Cdy = C\cdot(2x/3+2) = 2x/9 + 2/3$$

$$\rho_2(y) = \int_{-\infty}^{\infty}\rho(x,y)dx = \int_{3y/2-3}^{0}Cdx = -C\cdot(3y/2-3) = -y/2+1$$

По формулам (8)

$$m_x = \int_{-\infty}^{\infty}x\cdot\rho_1(x)dx = \int_{-3}^{0}x\cdot(2x/9+2/3) = -1$$

$$m_y = \int_{-\infty}^{\infty}y\cdot\rho_2(y)dy = \int_{0}^{2}y\cdot(-y/2+1)dy = 2/3$$

По формулам (9)

$$D_x = \int_{-\infty}^{\infty}x^2\cdot\rho_1(x)dx - m_x^2 = \int_{-3}^{0}x^2\cdot(2x/9+2/3) - 1 = 1/2$$

$$D_y = \int_{-\infty}^{\infty}y^2\cdot\rho_2(y)dy - m_y^2 = \int_{0}^{2}y^2\cdot(-y/2+1)dy - 4/9 = 2/9$$

По формуле (10)

$$Cov(X,Y)_x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot \rho(x,y)dxdy - m_x \cdot m_y = C \cdot \int_{-3}^{0} xdx \int_{0}^{2x/3+2} ydy + \frac{2}{3} = \frac{1}{3}\int_{-3}^{0} x \cdot (2x/3+2)^2 \frac{1}{2} + \frac{2}{3} = \frac{1}{6}$$

По формуле (11)

$$r_{xy} = Cov(X,Y)/\ (\sigma_x \cdot \sigma_y) = 1/2$$

Задачи к Лабораторной работе «Числовые характеристики двумерной случайной величины»

Задача 3.1 Дана матрица распределения вероятностей системы (X, Y)

Найти**:** 1) ряды распределения для X и Y; 2) m_x, m_y, D_x, D_y; 3) $Cov(X,Y)$; 4) r_{xy};6) $M[Y/X=x_1]$.

Вариант 1				Вариант 2				Вариант 3			
Y	*X*			*Y*	*X*			*Y*	*X*		
	3	10	12		2	5	8		10	14	18
4	0.17	0.13	0.25	0.4	0.15	0.30	0.35	8	0.25	0.15	0.32
5	0.10	0.30	0.05	0.8	0.05	0.12	0.013	6	0.10	0.05	0.13

Вариант 4				Вариант 5				Вариант 6			
Y	*X*			*Y*	*X*			*Y*	*X*		
	1	2	3		3	10	12		4	3	2
1	0.1	0.19	0.2	2	0.17	0.13	0.25	2	0.17	0.13	0.25
2	0.16	0.20	0.15	4	0.30	0.10	0.05	1	0.10	0.30	0.05

Вариант 7				Вариант 8				Вариант 9			
Y	*X*			*Y*	*X*			*Y*	*X*		
	2	3	4		0	2	3		3	5	7
0.4	0.10	0.30	0.05	2	0.05	0.10	0.30	4	0.30	0.05	0.10
0.5	0.17	0.13	0.25	4	0.25	0.17	0.13	2	0.13	0.25	0.17

Вариант 10				Вариант 11				Вариант 12			
Y	*X*			*Y*	*X*			*Y*	*X*		
	2	0	3		4	5	6		6	4	2
4	0.16	0.20	0.15	5	0.20	0.16	0.15	3	0.20	0.15	0.16
5	0.1	0.19	0.2	4	0.19	0.1	0.2	2	0.19	0.2	0.1

Вариант 13				Вариант 14				Вариант 15			
Y	*X*			*Y*	*X*			*Y*	*X*		
	6	4	2		5	3	1		4	2	0
4	0.10	0.05	0.13	0.5	0.13	0.10	0.05	2	0.05	0.13	0.10
5	0.25	0.15	0.32	0.4	0.32	0.25	0.15	3	0.15	0.32	0.25

Вариант 16				Вариант 17				Вариант 18			
Y	*X*			*Y*	*X*			*Y*	*X*		
	6	0	3		0	2	1		3	4	2
4	0.10	0.30	0.05	4	0.05	0.10	0.30	3	0.30	0.05	0.10
5	0.17	0.13	0.25	2	0.25	0.17	0.13	2	0.13	0.25	0.17

Продолжение

Задача 1.1

Вариант 19				Вариант 20				Вариант 21			
Y	X			Y	X			Y	X		
	2	4	1		3	2	1		4	6	5
4	0.16	0.20	0.15	5	0.20	0.16	0.15	6	0.16	0.15	0.20
5	0.1	0.19	0.2	3	0.19	0.1	0.2	4	0.1	0.2	0.19

Задача 3.2 Дана плотность распределения вероятностей системы (X, Y)

$\rho\ (x,y) = C$ внутри треугольника O(0,0), $A(a,0)$, $B(0,b)$ и $\rho\ (x,y) = 0$ в других точках.

Найти: 1) константу C ;2) $\rho_1(x)$, $\rho_2(y)$; 3) m_x, m_y, D_x, D_y, ; 4) r_{xy} . Данные взять из Таб. 3.2

Таб.3.2 Данные для задачи 3.2

Вариант	a	b	Вариант	a	b
1	-3	1	11	2	-1
2	-4	3	12	3	2
3	-5	4	13	2	3
4	2	1	14	-3	2
5	2	2	15	4	3
6	3	2	16	2	5
7	3	3	17	5	2
8	-4	5	18	-5	2
9	-3	3	19	-6	2
10	-3	4	20	-2	6

Лабораторная работа «Линейная корреляция»**.**

Цель работы:

Научиться выявлять зависимости между признаками по значению выборочного коэффициента корреляции, научиться получать линейные уравнения регрессии и применять формулы для функции регрессии для прогноза изменения одного признака при изменении другого.

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 2 часа.

Теоретические основы и примеры.

Обозначения:

1**.** *Корреляционная* таблица. Пусть (X, Y) - система из двух случайных величин X и Y .

Данные наблюдений двух признаков , т.е. пары чисел (x_i, y_j), группируют в

корреляционную таблицу

Y	X			
	x_1	…	x_l	
y_1	n_{11}		n_{l1}	m_1
…	…	…	…	
y_l	n_{q1}		n_{ql}	m_q
	n_1		n_l	$n= \sum n_i$ $=\sum m_j$

с указанием частот n_{ij} появления пар чисел (x_i, y_j), i =1,2,…l, j =1,2,…q ,

с указанием частот $n_i = \sum_{j=1}^{q} n_{ij}$, появления значений x_i (сумма n_{ij} по столбцам) , с указанием частот $m_j = \sum_{i=1}^{l} n_{ij}$ появления значений y_j (сумма n_{ij} по строчкам) и с указанием $n = \sum_{i=1}^{l} n_j = \sum_{j=1}^{q} m_i$

Замечание. Вместо x_i и y_j могут быть даны интервалы, тогда x_i и y_j - центры интервалов.

2. *Выборочные* средние признаков X и Y

$$\bar{x} = \sum_{i=1}^{l} x_i n_i / n, \ \bar{y} = \sum_{j=1}^{q} y_j m_j / n \qquad (1)$$

3. *Выборочные* дисперсии признаков X и Y

$$D_x = \sum_{i=1}^{l} x_i^2 n_i / n - \bar{x}^2, \ D_y = \sum_{j=1}^{q} y_j^2 m_j / n - \bar{y}^2 \qquad (2)$$

4. *Выборочные* средние квадратичные отклонения

$$\sigma_x = (D_x)^{1/2}, \sigma_y = (D_y)^{1/2} \qquad (3)$$

5. *Выборочный* коэффициент корреляции.

$$r_v = \frac{\sum_{i=1}^{l}\sum_{j=1}^{q} x_i \cdot y_j \cdot n_{ij} / n - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \qquad (4)$$

6. *Выборочное уравнение прямой линии регрессии* Y на X имеет вид

$$\bar{y}_x - \bar{y} = r_v \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x}) \qquad (5)$$

7. *Выборочное уравнение прямой линии регрессии* X на Y имеет вид

$$\bar{x}_y - \bar{x} = r_v \cdot \frac{\sigma_x}{\sigma_y}(y - \bar{y}) \qquad (6)$$

Пример 4.1 Данные об объеме выпуска продукции (Y) и стоимости основных промышленных фондов (X) по 60 предприятиям сгруппированы в Таблицу 4.1. Найти выборочное уравнение прямой линии регрессии Y на X. Определить объем выпуска продукции при стоимости основных фондов 14 млн. р.

Таб. 4.1

		i	1	2	3	4	5= l	
j	Y		0-2	2-4	4-6	6-8	8-10	m_j , Сумма частот по строчкам
	интервал	*центры* x_j y_j	1	3	5	7	9	
1	0-0.2	0.1	2	2				4= m_1
2	0.2-0.4	0.3	2	7	10			19= m_2
3	0.4-0.6	0.5		2	17	7		26= m_3
4	0.6-0.8	0.7			4	3	2	9= m_4
5=q	0.8-1.0	0.9					2	2= m_5
n_i - Сумма частот по столбцам			4= n_1	11= n_2	31= n_2	10= n_3	4= n_4	$n = \sum n_i = \sum_j m_j$ =60

Решение.

По формулам (1) находим $\bar{x} = 4.96667$, $\bar{y} = 0.453333$;

По формулам (2) и (3) находим D_x = 3.53222, D_y = 0.03248; σ_x = 1.87942, σ_y = 0.180247;

По формуле (4) $r_v = 0.743235$;

Подставляем найденные значения в формулу (5) и получаем уравнение прямой линии

$$\bar{y}_x = 0.0993 + 0.0713 \cdot x$$

При $x = 14.0$ $\bar{y}_x = 1.0972$

Задачи к Лабораторной работе «Линейная корреляция»

Задача 4.1 . При выборке объема n, извлеченной из двумерной нормальной совокупности (X, Y) найти выборочные уравнения прямых регрессии Y на X и X на Y . Данные взять из Таб. 4.2.

Таб. 4.2. Данные к задаче 4.1.

Вариант 1						Вариант 2					
Y	*X*					*Y*	*X*				
	20	25	30	35	40		21	26	31	36	41
16	4	6	0	0	0	17	5	7	0	0	0
26	0	8	10	0	0	27	0	9	10	0	0
36	0	0	32	3	9	37	0	0	33	4	10
46	0	0	4	12	6	47	0	0	5	13	7
56	0	0	0	1	5	57	0	0	0	2	6
Вариант 3						Вариант 4					
Y	*X*					*Y*	*X*				
	22	27	32	37	42		22	27	32	37	42
18	6	8	0	0	0	18	6	8	0	0	0
28	0	10	11	0	0	28	2	10	11	2	0
38	0	0	34	5	11	38	0	2	34	5	11
48	0	0	6	14	8	48	0	0	6	14	8
58	0	0	0	2	7	58	0	0	2	2	7
Вариант 5						Вариант 6					
Y	*X*					*Y*	*X*				
	23	28	33	38	43		23	28	33	38	43
19	6	8	0	0	0	19	7	9	0	0	0
29	2	10	11	2	0	29	3	11	12	3	0
39	0	2	34	5	11	39	0	2	35	6	12
49	0	0	6	14	8	49	0	0	6	15	9
59	0	0	2	2	7	59	0	0	2	3	8
Вариант 7						Вариант 8					
Y	*X*					*Y*	*X*				
	24	29	34	39	44		24	29	34	39	44
20	7	9	0	0	0	20	8	10	0	0	0
30	3	11	12	3	0	30	4	12	13	4	0
40	0	2	35	6	12	40	0	2	36	7	13
50	0	0	6	15	9	50	0	0	7	16	10
60	0	0	2	3	8	60	0	0	3	4	9
Вариант 7						Вариант 8					
Y	*X*					*Y*	*X*				

Продолжение

Таб. 4.2. Данные к задаче 4.1.

	24	29	34	39	44		24	29	34	39	44
20	7	9	0	0	0	20	8	10	0	0	0
30	3	11	12	3	0	30	4	12	13	4	0
40	0	2	35	6	12	40	0	2	36	7	13
50	0	0	6	15	9	50	0	0	7	16	10
60	0	0	2	3	8	60	0	0	3	4	9
Вариант 9						Вариант 10					
Y	*X*					*Y*	*X*				
	24	29	34	39	44		24	29	34	39	44
20	9	11	0	0	0	20	9	11	0	0	0
30	5	13	14	5	0	30	5	13	14	5	2
40	2	3	37	8	14	40	2	3	37	8	14
50	0	0	8	17	11	50	0	0	8	17	11
60	0	0	4	5	10	60	0	0	4	5	10
Вариант 11						Вариант 12					
Y	*X*					*Y*	*X*				
	24	29	34	39	44		24	29	34	39	44
20	9	11	0	0	0	20	10	12	0	0	0
30	5	10	14	5	2	30	5	10	14	5	2
40	2	3	30	8	14	40	3	4	30	8	14
50	0	0	8	17	11	50	0	0	8	17	11
60	0	0	4	5	10	60	0	0	4	5	10
Вариант 13						Вариант 14					
Y	*X*					*Y*	*X*				
	14	19	24	29	34		14	19	24	29	34
10	10	12	0	0	0	15	10	12	0	0	0
20	5	10	14	5	2	25	5	10	14	5	2
30	3	4	30	8	14	35	3	4	30	8	14
40	0	0	8	17	11	45	0	0	8	17	11
50	0	0	4	5	10	55	0	0	4	5	10
Вариант 15						Вариант 16					
Y	*X*					*Y*	*X*				
	19	24	29	34	39		20	25	30	35	40
15	10	12	0	0	0	15	10	12	0	0	0
25	5	10	14	5	2	25	5	10	14	5	2
35	3	4	30	8	14	35	3	4	30	8	14
45	0	0	8	17	11	45	0	0	8	17	11
55	0	0	4	5	10	55	0	0	4	5	10
Вариант 17						Вариант 18					
Y	*X*					*Y*	*X*				
	20	25	30	35	40		20	25	30	35	40
15	10	12	0	0	0	15	10	12	0	0	0
25	5	10	14	0	0	25	5	10	14	0	0
35	3	4	30	0	0	35	3	4	30	0	0
45	0	0	8	17	11	45	0	0	8	20	10
55	0	0	4	5	10	55	0	0	4	5	10

Продолжение

Таб. 4.2. Данные к задаче 4.1.

Вариант 19						Вариант 20					
Y	*X*					*Y*	*X*				
	20	25	30	35	40		30	35	40	45	50
15	10	15	0	0	0	25	10	15	0	0	0
25	5	10	15	0	0	35	5	10	15	0	0
35	3	4	30	0	0	45	3	4	30	0	0
45	0	0	8	20	10	55	0	0	8	20	10
55	0	0	4	5	10	65	0	0	4	5	10

Лабораторная работа «Статистика ошибок в параметрах модели».

Цель работы:

1.В методе распространения ошибок научиться проводить расчет ошибок определения параметров модели при известных ошибках в некоррелированных наблюдаемых.

2. Научиться проводить расчет ошибок в предсказании наблюдаемых .

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 4 часа.

Теоретические основы и примеры.

Определения.

Y_i (i =1, 2, …n) –в общем случае некоррелированные экспериментальные (наблюдаемые) данные для физической величины Y ;

x_j (j =1, 2, …m)-параметры аналитической модели, $m \leq n$;

t- контролируемая переменная

$$\overline{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i \tag{1}$$

-среднее значение случайной величины Z;

$$\sigma^2(Z) = \mathrm{var}(Z) = \frac{1}{N}\sum_{i=1}^{N}(Z_i - \overline{Z})^2 \ , \tag{2}$$

$$\sigma(Y_1) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{1,i} - \overline{Y}_1)^2} \tag{3}$$

- *среднеквадратичная ошибка* (*rms*) *i* –го наблюдаемого,

$\sigma(x_i)$ – *среднеквадратичная ошибка* (*rms*) *i* –го параметра; $\sigma(x_i)$ = *rms* = $\{\mathrm{var}(x_i)\}^{1/2}$, в которой $\mathrm{var}(x_i)$ = вариация этого параметра. Эту ошибку необходимо вычислить.

Задача решается с помощью формул из Таб. 5.1.

Таблица 5.1. Распространение стандартных отклонений 2-х некоррелированных наблюдаемых на стандартные отклонения и корреляционный коэффициент 2-х параметров с использованием производных ($\partial Y_i/ \partial x_j$).

$\sigma(x_1) = \{(\partial Y_2/ \partial x_2)^2\, \sigma^2(Y_1)+ (\partial Y_1/ \partial x_2)^2\, \sigma^2(Y_2)\}^{1/2} / \,|J|$

$\sigma(x_2) = \{(\partial Y_2/ \partial x_1)^2\, \sigma^2(Y_1)+ (\partial Y_1/ \partial x_1)^2\, \sigma^2(Y_2)\}^{1/2} / \,|J|$

$\rho(x_1, x_2) = - [J^2\, \sigma(x_1) \cdot\sigma(x_2)]^{-1}\, \{(\partial Y_2/ \partial x_2)(\partial Y_2/ \partial x_1)\sigma^2(Y_1)+ (\partial Y_1/ \partial x_1)(\partial Y_1/ \partial x_2)\sigma^2(Y_2)\}^{1/2}$

$J = \begin{vmatrix} \dfrac{\partial Y_1}{\partial x_1} & \dfrac{\partial Y_1}{\partial x_2} \\ \dfrac{\partial Y_2}{\partial x_1} & \dfrac{\partial Y_2}{\partial x_2} \end{vmatrix},$ $|J|$ -абсолютное (т.е.положительное) значение Якобиана J.

В формулах из таблицы 5.1 $\partial Y_i/\partial x_j$ - это производные от аналитической модели по параметрам, т.е. эта аналитическая формула, в которую нужно подставить наблюдаемые «экспериментальных» значения для Y_i.

Поясним смысл «экспериментальных» значений $Y_1^{\text{эксп.}}$, $Y_2^{\text{эксп.}}$ и величин $\sigma(Y_1)$ и $\sigma(Y_2)$. Это значит, что для конкретного значения контролируемой переменной t_1 проведено N_1 измерений величины Y, в каждом из котором эта величина принимала значение $Y_{1,i}$, i= 1, 2 … N_1 . Тогда по формулам (1) и (3)

$$Y_1^{\text{эксп.}} = \overline{Y}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} Y_{1,i}\ , \tag{4}$$

$$\sigma(Y_1) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{1,i} - \overline{Y}_1)^2} \tag{5}$$

Пример 5.1 Пример расчета стандартных отклонений $\sigma(x_1)$, $\sigma(x_2)$ и коэффициента корреляции $\rho(x_1, x_2)$ с использованием формул из Таблицы 5.1. Константа химической реакции K в законе Аррениуса (пример взят из [4])

$$K = A\cdot \exp[-E/(R\cdot T)] = A\cdot \exp[-E/(8.1314\cdot T)] \tag{6}$$

(R –газовая постоянная) зависит от двух параметров x_1=A, $x_2 = E$ (энергия активации) и переменной $t = T$, в качестве которой выступает температура T .

Основные этапы расчета.

Шаг 1.

Пусть для двух температур T_1 = 300K и T_2 = 350K получены 2 серии из N_1 =10 и N_2 =10 наблюдений константы K :

Для T_1=300.0 ;

Y_1=0.580;Y_2=0.582;Y_3=0.589;Y_4=0.588;Y_5=0.590;Y_6=0.593;Y_7=0.594;Y_8=0.595;Y_9=0.585;

Y_{10}=0.591;

т.е. получено N_1 =10 значений.

Для T= 350.

Y_1=9.05 ;Y_2=8.95;Y_3=9.06;Y_4=9.10;Y_5=9.12;Y_6=9.15;Y_7=9.20;Y_8=8.98;Y_9=9.10;Y_{10}=8.99;

Шаг 2. По формулам (5) и (6) вычисляем значения $\overline{Y}_1 = K_1^{\text{эксп.}} = K(T_1$=300K) = 0.5887, $\sigma(Y_1)$ = 0.0047 , $\overline{Y}_2 = K_2^{\text{эксп.}} = K$ (T_2 =350K)= 9.07, $\sigma(Y_2)$ = 0.0750 (размерности величин опущены). Значения $K_1^{\text{эксп.}}$ =0.5887 и $K_2^{\text{эксп.}}$ =9.07 принимаются за две наблюдаемые величины Y_1 и Y_2, т.е. $Y_1 = K_1^{\text{эксп.}}$, $Y_2 = K_2^{\text{эксп.}}$

Шаг 3. Из двух известных значений $K_1^{\text{эксп.}}$ и $K_2^{\text{эксп.}}$ определяем значения двух параметров $x_1 = A$ и $x_2 = E$. Из формулы (6) для $T = T_1$ и $T = T_2$ находим

$$A = K_1^{\text{эксп.}} \cdot \exp\,[E /(8.314 \cdot T_1)] = 1.213 \cdot 10^8 \qquad (7)$$

$$E = 8.314 \cdot (T_2 \cdot T_1) \cdot \mathrm{Ln}(K_2^{\text{эксп.}} / K_1^{\text{эксп.}})/ (T_2 - T_1) = 4.669 \cdot 10^4,$$

Шаг 4. Из формулы (6) находим производные

$$(\partial Y_i / \partial x_1) = (\partial Y_i / \partial A) = (\partial K_i / \partial A) = \exp[-E/(8.1314 \cdot T_i)] \qquad (8)$$

$$(\partial Y_i / \partial x_2) = (\partial Y_i / \partial E) = -(A/(8.1314 \cdot T_i) \exp[-E/(8.1314 \cdot T_i)] .$$

Подставляем в эти формулы найденные значения параметров A =1.219·10^8 и E = 4.669·10^4 , находим численные значения производных

$$\partial Y_1 / \partial A = 4.85 \cdot 10^{-9},\ \partial Y_2 / \partial A = 4.11 \cdot 10^{-8},$$

$$\partial Y_1 / \partial E = -2.4 \cdot 10^{-4},\ \partial Y_2 / \partial E = -3.2 \cdot 10^{-3},$$

из которых формируем и вычисляем Якобиан

$$J = \begin{vmatrix} 4.85 \cdot 10^{-9} & -2.4 \cdot 10^{-4} \\ 4.11 \cdot 10^{-8} & -3.2 \cdot 10^{-3} \end{vmatrix} = 2.58 \cdot 10^{-18} \tag{9}$$

Шаг 5. Подставляя найденные численные значения для производных и Якобиана J в формулы для $\sigma(x_1\!=\!A)$, $\sigma(x_2\!=\!E)$ и $\rho(x_1, x_2)$ из таблицы 5.1, и учитывая, что $\sigma(Y_1)$ = 0.0047 , $\sigma(Y_2)$ = 0.0750, находим

$$\sigma(x_1\!=\!A) = 9.20 \cdot 10^6, \tag{10}$$

$$\sigma(x_2\!=\!E) = 1.98 \cdot 10^2,$$

$$\rho(x_1, x_2) = 0.997.$$

Шаг 6. Результат определения параметров A и E (7) и стандартных отклонений $\sigma(A)$, $\sigma(E)$ записываем как

$$A = (1.219 \pm 0.092) \cdot 10^8, \tag{11}$$

$$E = (4.669 \pm 0.019) \cdot 10^4$$

Формулы (11) и есть окончательный ответ для решаемой задачи1 .

Задача расчета ошибок в предсказании наблюдаемых, в случае 2-х наблюдаемых и двух параметров модели, решается с применением формулы

$$\sigma(Y_i) = [(\frac{\partial Y_i}{\partial x_1})^2 \sigma^2(x_1) + (\frac{\partial Y_i}{\partial x_2})^2 \sigma^2(x_2) + 2(\frac{\partial Y_i}{\partial x_1}\frac{\partial Y_i}{\partial x_2})\sigma(x_1)\sigma(x_2)\rho(x_1, x_2)]^{1/2} \tag{12}$$

Например, для нахождения значения константы химической реакции в (6) для конкретной температуры T =100 нужно для этой температуры вычислить все входящие в (12) величины, после чего можно найти $K(T=100) = 1.38 \cdot 10^{-17}$, и $\sigma(K\!=\!100) = 2.33 \cdot 10^{-18}$.

Задачи к Лабораторной работе «Статистика ошибок в параметрах модели»

Задача 5.1. Для закона Аррениуса (6) определить значения параметров A , E их стандартные отклонения $\sigma(A)$, $\sigma(E)$, коэффициент корреляции $\rho(A, E)$, провести расчет константы химической реакции К для температур T =100, 200, 300, 400, 500, 600 К и стандартного отклонения $\sigma(K)$. Данные взять из Таблицы 5.2.

Таблица 5.1. Данные для решения задачи 5.1.

Вар.		Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
1	T_1=300	0.585	0.592	0.582	0.590	0.583	0.594	0.585	0.595	0.581
	T_2=350	9.150	8.967	8.999	9.012	9.014	8.995	8.996	9.013	9.014
2	T_1=350	9.160	8.997	8.899	9.112	9.114	8.895	8.896	9.113	9.114
	T_2=400	70.05	71.0	71.05	70.50	71.06	70.20	72.00	71.05	71.15
3	T_1=400	70.05	71.0	71.05	70.50	71.06	70.20	72.00	71.05	71.15
	T_2=450	350.0	351.2	352.0	349.6	350.6	353.1	353.8	354.4	355.5
4	T_1=450	351.0	352.2	353.0	349.8	351.6	354.1	353.4	354.2	355.1
	T_2=500	$1.23\cdot10^{3}$	$1.21\cdot10^{3}$	$1.20\cdot10^{3}$	$1.24\cdot10^{3}$	$1.25\cdot10^{3}$	$1.26\cdot10^{3}$	$1.27\cdot10^{3}$	$1.28\cdot10^{3}$	$1.26\cdot10^{3}$
5	T_1=500	$1.20\cdot10^{3}$	$1.20\cdot10^{3}$	$1.22\cdot10^{3}$	$1.22\cdot10^{3}$	$1.24\cdot10^{3}$	$1.23\cdot10^{3}$	$1.26\cdot10^{3}$	$1.25\cdot10^{3}$	$1.23\cdot10^{3}$
	T_2=550	$3.60\cdot10^{3}$	$3.62\cdot10^{3}$	$3.63\cdot10^{3}$	$3.62\cdot10^{3}$	$3.64\cdot10^{3}$	$3.63\cdot10^{3}$	$3.66\cdot10^{3}$	$3.65\cdot10^{3}$	$3.63\cdot10^{3}$
6	T_1=550	$3.62\cdot10^{3}$	$3.63\cdot10^{3}$	$3.64\cdot10^{3}$	$3.60\cdot10^{3}$	$3.61\cdot10^{3}$	$3.62\cdot10^{3}$	$3.61\cdot10^{3}$	$3.61\cdot10^{3}$	$3.65\cdot10^{3}$
	T_2=600	$8.60\cdot10^{3}$	$8.62\cdot10^{3}$	$8.63\cdot10^{3}$	$8.62\cdot10^{3}$	$8.64\cdot10^{3}$	$8.63\cdot10^{3}$	$8.66\cdot10^{3}$	$8.65\cdot10^{3}$	$8.63\cdot10^{3}$
7	T_1=600	$8.61\cdot10^{3}$	$8.63\cdot10^{3}$	$8.64\cdot10^{3}$	$8.61\cdot10^{3}$	$8.62\cdot10^{3}$	$8.61\cdot10^{3}$	$8.62\cdot10^{3}$	$8.61\cdot10^{3}$	$8.63\cdot10^{3}$
	T_2=650	$1.80\cdot10^{4}$	$1.81\cdot10^{4}$	$1.80\cdot10^{4}$	$1.84\cdot10^{4}$	$1.85\cdot10^{4}$	$1.86\cdot10^{4}$	$1.87\cdot10^{4}$	$1.88\cdot10^{4}$	$1.86\cdot10^{4}$
8	T_1=650	1.8210^{4}	$1.82\cdot10^{4}$	$1.83\cdot10^{4}$	$1.81\cdot10^{4}$	$1.82\cdot10^{4}$	$1.82\cdot10^{4}$	$1.82\cdot10^{4}$	$1.83\cdot10^{4}$	$1.88\cdot10^{4}$
	T_2=700	$3.40\cdot10^{4}$	$3.41\cdot10^{4}$	$3.40\cdot10^{4}$	$3.44\cdot10^{4}$	$3.45\cdot10^{4}$	$3.46\cdot10^{4}$	$3.47\cdot10^{4}$	$3.48\cdot10^{4}$	$3.46\cdot10^{4}$
9	T_1=200	$4.0\cdot10^{-5}$	$4.1\cdot10^{-5}$	$4.2\cdot10^{-5}$	$4.4\cdot10^{-5}$	$4.3\cdot10^{-5}$	$4.4\cdot10^{-5}$	$4.5\cdot10^{-5}$	$4.6\cdot10^{-5}$	$4.7\cdot10^{-5}$
	T_2=250	$1.3\cdot10^{-2}$	$1.4\cdot10^{-2}$	$1.2\cdot10^{-2}$	$1.4\cdot10^{-2}$	$1.3\cdot10^{-2}$	$1.5\cdot10^{-2}$	$1.5\cdot10^{-2}$	$1.6\cdot10^{-2}$	$1.7\cdot10^{-2}$
10	$T_1$250	$1.2\cdot10^{-2}$	$1.3\cdot10^{-2}$	$1.1\cdot10^{-2}$	$1.3\cdot10^{-2}$	$1.2\cdot10^{-2}$	$1.4\cdot10^{-2}$	$1.4\cdot10^{-2}$	$1.5\cdot10^{-2}$	$1.6\cdot10^{-2}$
	T_2=300	0.585	0.592	0.582	0.590	0.583	0.594	0.585	0.595	0.581

Лабораторная работа «Вычисление среднеквадратичных ошибок в параметрах модели в методе наименьших квадратов».

Цель работы:

1. Научиться проводить расчет ошибок определения параметров модели, нелинейно зависящей от параметров, при известных ошибках в некоррелированных наблюдаемых.

2. Построение доверительных интервалов

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 4 часа.

Теоретические основы и примеры.

Определения:

x_1, x_2 ... x_m - оптимальные параметры модели, которые считаются известными. Они получены в методе наименьших квадратов;

y_i - теоретические значения наблюдаемой, полученные с оптимальными параметрами;

Δy_i = Y_i(экспериментальные значения) – y_i (теоретические значения, полученные с оптимальным набором параметров, которые минимизируют сумму квадратов отклонений).Вычисление среднеквадратичной ошибки $\sigma(x_i)$ в m параметрах модели из n экспериментальных данных проводится по формулам Таблицы 6.1

Таблица 6.1. Вычисление среднеквадратичной ошибки $\sigma(x_i)$ в m параметрах из n экспериментальных данных.

$$\sigma(x_i) = \sigma \cdot [(P)_{ii}]^{1/2} \quad (1)$$

$$\rho(x_i, x_j) = \rho_{ij} = (P)_{ij} / [(P)_{ii} (P)_{jj}]^{1/2} \quad (2)$$

-элементы корреляционной матрицы R

$(P)_{ij}$ - элементы матрицы

$$P = B^{-1} = (A' \cdot W \cdot A)^{-1} \quad (3)$$

W – весовая матрица, A' – транспонированная для A матрица

$$\sigma = [(\Delta y)' \cdot W \cdot \Delta y / (n-m)]^{1/2} \quad (4)$$

или

$$\sigma = \{ S_{cp} / (n-m) \}^{1/2} \quad (5)$$

$$S_{cp} = \sum_{i=1}^{n} (Y_i - y_i)^2 \quad (6)$$

Элементы матрицы A определяются соотношением

$$(A)_{ij} = \partial Y_i / \partial x_j \quad (7)$$

Применение формул из Таблицы 6.1 рассмотрим на конкретном примере.

Пример 6.1 Для подгонки $n = 15$ экспериментальных данных Y_i из второй колонки Таблицы 6.2 использовалась двухпараметрическая ($m = 2$) формула

$$y(t) = se^{pt}, \quad (8)$$

в которой t – независимая переменная, а $x_1 = s = 58.603$ и $x_2 = p = -0.04$ – параметры, определенные методом наименьших квадратов. Снова предполагается, что все Y_i некоррелированы и весовая матрица W есть единичная матрица 15×15.

Таблица 6. 2. Вычисление среднеквадратичных ошибок параметров из формулы (8).

i	t_i	Y_i	$a_{i1} = \partial y_i/\partial s \times 10^{-2}$	$a_{i2} = \partial y_i/\partial p \times 10^{-2}$	$\Delta y_i = Y_i - y_i$
1	2.0	54.0	0.923	108.2	0.14
2	5.0	50.0	0.818	239.9	-1.92
3	7.0	45.0	0.756	310.0	-0.58
4	10.0	37.0	0.670	392.8	2.44
5	14.0	35.0	0.571	468.6	-1.32
6	19.0	25.0	0.467	520.7	2.62
7	26.0	20.0	0.353	538.5	0.94
8	31.0	16.0	0.289	525.7	1.17
9	34.0	18.0	0.256	511.4	-2.74
10	38.0	13.0	0.219	487.0	0.02
11	45.0	8.0	0.165	435.9	1.86
12	52.0	11.0	0.125	380.7	-3.52
13	53.0	8.0	0.120	372.8	-0.81
14	60.0	4.0	0.090	318.9	1.45
15	65.0	6.0	0.0743	282.9	-1.52

Шаг 1. Находим формулы для производных от $y(t)$ (8) по параметрам и вычисляем элементы матрицы A

$$a_{i1} = \partial y_i/\partial s = exp(p \cdot t_i), \quad (9)$$

$$a_{i2} = \partial y_i/\partial p = s \cdot t_i \cdot exp(p \cdot t_i),$$

Они приведены в 3-й и 4-й колонках таблицы 6.2. Матрица A имеет размерность (15×2), а нормальная матрица $B=(A'\cdot W\cdot A) = (A'A)$ - размерность (2×2). Ее явный вид следующий.

$$A'A = \begin{bmatrix} 3.48 & 2099.3 \\ 2099.3 & 2.53\cdot 10^{-6} \end{bmatrix} \tag{10}$$

Обращение этой матрицы дает матрицу $P = B^{-1} = (A'\cdot W\cdot A)^{-1}$

$$P = \begin{bmatrix} P_{11} = 0.575 & P_{12} = -0.00047 \\ P_{21} = -0.00047 & P_{22} = 7.96\cdot 10^{-7} \end{bmatrix} \tag{11}$$

Шаг 2. По формуле (6) вычисляем сумму квадратов отклонений S_{cp} с помощью последней колонки Таблицы 6.2, находим S_{cp} =49.28 и величину $\sigma = (S_{cp}/(n-m))^{1/2} = (S_{cp}/13)^{1/2} = 1.95$. Среднеквадратичные ошибки в параметрах аналитической формулы (8) определяются по формуле (1) как

$$\sigma(s) = \sigma\cdot [(P)_{11}]^{1/2} = 1.47;\ \sigma(p) = \sigma\cdot [(P)_{22}]^{1/2} = 0.0017 \tag{12}$$

Корреляционная матрица R с элементами $\rho(x_i, x_j) = \rho_{ij} = (P)_{ij} / [(P)_{ii}(P)_{jj}]^{1/2}$

имеет вид

$$R = \begin{bmatrix} 1.0 & -0.7 \\ -0.7 & 1.0 \end{bmatrix} \tag{13}$$

Шаг 3. Построение доверительных интервалов. Знание матрицы P вполне достаточно, чтобы сделать определенные заключения о параметрах x_k выбранной аналитической модели. Верхняя $x_k^{(+)}$ и нижняя $x_k^{(-)}$ границы полученных параметров определяются соотношением

$$x_k^{(\pm)} = x_k \pm t_\gamma\ \sigma(x_k), \tag{14}$$

в котором t_γ находится по таблице Стьюдента по заданному γ. Часто в научной литературе используют значение γ =0. 75 , при этом t_γ =1.0, и доверительный интервал полностью определяется значениями $\sigma(x_k)$.

Задачи к Лабораторной работе «Вычисление среднеквадратичных ошибок в параметрах модели в методе наименьших квадратов».

Задача 6.1. Для аналитической модели

$$y(t) = \frac{1+st}{p+ut} , \tag{15}$$

с найденными в методе наименьших квадратов значениями параметров $x_1 = s$, $x_2 = p$, $x_3 = u$ и известными при t_i = {0.35,0.60,1.0,2.0,3.0,4.0,4.5, 5.0,5.5} значениями Y_i найти среднеквадратичные отклонения для параметров и построить корреляционную матрицу. Данные взять из Таб. 6.3.

Таблица 6. 3. Данные к задаче 6.1.

Вариант	$x_1 = s$, $x_2 = p$, $x_3 = u$	Y
1	s =0.78; p=14.5; u= 4.3	{0.08;0.084;0.097;0.108;0.124;0.129;0.132;0.135;0.137}
2	s =0.71; p=14.1; u= 4.1	{0.05;0.081;0.091;0.101;0.121;0.121;0.131;0.131;0.131}
3	s =0.72 p=14.2 u= 4.2	{0.06;0.082;0.092;0.102;0.122;0.122;0.132;0.132;0.132}
4	s =0.73; p=14.3; u= 4.3	{0.07;0.083;0.093;0.103;0.123;0.123;0.133;0.133;0.133}
5	s =0.74; p=14.4; u= 4.4	{0.05;0.084;0.094;0.104;0.124;0.124;0.134;0.134;0.134}
6	s =0.75; p=14.5; u= 4.5	{0.06;0.085;0.095;0.105;0.125;0.125;0.135;0.135;0.135}
7	s =0.76; p=14.6; u= 4.6	{0.07;0.085;0.096;0.106;0.126;0.126;0.136;0.136;0.136}
8	s =0.77; p=14.7; u= 4.7	{0.05;0.086;0.097;0.107;0.127;0.127;0.137;0.137;0.137}
9	s =0.78; p=14.8; u= 4.8	{0.06;0.088;0.098;0.108;0.128;0.128;0.138;0.138;0.138}
10	s =0.79; p=14.9; u= 4.9	{0.07;0.089;0.099;0.109;0.129;0.129;0.139;0.139;0.139}
11	s =0.71; p=14.1; u= 4.1	{0.05;0.081;0.091;0.101;0.121;0.129;0.131;0.131;0.131}
12	s =0.72; p=14.2; u= 4.2	{0.06;0.082;0.092;0.102;0.122;0.121;0.132;0.132;0.132}
13	s =0.73; p=14.3; u= 4.3	{0.07;0.083;0.093;0.103;0.123;0.122;0.133;0.133;0.133}
14	s =0.74; p=14.4; u= 4.4	{0.05;0.084;0.094;0.104;0.124;0.123;0.134;0.134;0.134}
15	s =0.75; p=14.5; u= 4.5	{0.06;0.085;0.095;0.105;0.125;0.124;0.135;0.135;0.135}
16	s =0.76; p=14.6; u= 4.6	{0.07;0.086;0.096;0.106;0.126;0.125;0.136;0.136;0.136}
17	s =0.77; p=14.7; u= 4.7	{0.05;0.087;0.097;0.107;0.127;0.126;0.137;0.137;0.137}
18	s =0.78; p=14.8; u= 4.8	{0.06;0.088;0.098;0.108;0.128;0.127;0.138;0.1380.138}
19	s =0.79; p=14.1; u= 4.9	{0.07;0.089;0.099;0.109;0.129;0.1280.131;0.139;0.139}
20	s =0.71; p=14.2; u= 4.1	{0.05;0.081;0.091;0.101;0.121;0.121;0.132;0.131;0.131}

Лабораторная работа «Определение параметров линейной модели методом наименьших квадратов и построение для них доверительных интервалов» **.**

Цель работы:

Определение параметров линейной модели методом наименьших квадратов. Построение доверительных интервалов для параметров.

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 2 часа.

Теоретические основы и примеры.

Определения:

Те же, что и в Лабораторной работе «Статистика ошибок в параметрах модели».

Пусть для каждого значения Y_i (i=1, 2,3…n) проведена серия из k наблюдений и получены экспериментальные значения Y_{ij} (i=1, 2,3…$n;j$ = 1,2,..k).

Для ряда $\{Y_1, Y_2, \ldots Y_n\}$ выбирается простейшая теоретическая модель

$$y(t) = a + b \cdot t \qquad (1)$$

с параметрами $x_1 = a$ и $x_2 = b$. Необходимо определить эти параметры с учетом стандартных отклонений для Y_k (с учетом весов), и найти ошибку определения этих параметров.

Пример 7.1 .

Для каждого $t_i \equiv i$ 1, 2, 3…8 проведены $k = 4$ измерения и получены значения Y_{t1} = {5.,8.,11.,14.,17.,20.,23.,26.}, Y_{t2} = {4.95, 7.85, 10.75, 13.65, 16.55, 19.45, 22.35, 25.25}, Y_{t3} = {4.92,7.89,10.86,13.83,16.8,19.77,22.74,25.71}, Y_{t4} = {4.91,7.92,10.93,13.94,16.95,19.96,22.97,25.98}.

По этим данным нужно определить параметры аналитической модели (1) и построить для них доверительные интервалы.

Шаг 1.

По формулам $Y_i = \overline{Y}_i = \frac{1}{k}\sum_{j=1}^{k} Y_{ij}$, $\sigma(Y_i) = \{\frac{1}{k}\sum_{j=1}^{k}(Y_{ij} - \overline{Y}_i)^2\}^{1/2}$ определяем экспериментальные значения $Y^{экс} = Y_i$, среднеквадратичные отклонения $\sigma(Y_i)$ и веса $w_{ii} = 1/\sigma^2(Y_i)$.

Результаты вычислений занесены в соответствующие колонки Таблицы 1.

Таблица 7. 1. Результаты вычислений параметров и их стандартных отклонений в линейной модели (1).

t_i	Y_i	$\sigma(Y_i)$	$w_{ii} = 1/\sigma^2(Y_i)$	$a_{i1} = \partial Y_i/\partial x_1 = 1$	$a_{i2} = \partial Y_i/\partial x_2 = t_i$
1	4.945	0.035	816.327	1	1
2	7.915	0.055	330.579	1	2
3	10.885	0.092	117.302	1	3
4	13.855	0.133	56.417	1	4
5	16.825	0.175	32.653	1	5
6	19.5875	0.238	17.630	1	6
7	22.765	0.259	14.809	1	7
8	24.235	1.643	0.370	1	8

Шаг 2. По формулам из Таблицы 7.2 определяем параметры $a = 1.988$ и $b = 2.960$.

Таблица 7. 2 . Формулы для определения параметров a и b из уравнения (1) в методе наименьших квадратов

$$b = \frac{1}{d}[(\sum_{i=1}^{n} w_{ii})(\sum_{i=1}^{n} w_{ii}t_iY_i) - (\sum_{i=1}^{n} w_{ii}t_i)(\sum_{i=1}^{n} w_{ii}Y_i)],$$

$$a = \frac{1}{d}[(\sum_{i=1}^{n} w_{ii}t_i^2)(\sum_{i=1}^{n} w_{ii}Y_i) - (\sum_{i=1}^{n} w_{ii}t_i)(\sum_{i=1}^{n} w_{ii}t_iY_i)],$$

$$d = (\sum_{i=1}^{n} w_{ii}t_i^2)(\sum_{i=1}^{n} w_{ii}) - (\sum_{i=1}^{n} w_{ii}t_i)^2 ,$$

$$w_{ii} = 1/\sigma^2(Y_i)$$

Шаг 3. С найденными параметрами a и b вычисляем взвешенную сумму квадратов отклонений

$$S_{cp} = \sum_{i=1}^{n} w_{ii}\cdot(Y_i - y_i)^2 = \sum_{i=1}^{n} w_{ii}\cdot(Y_i - (a + b\cdot t_i)^2 \qquad (2)$$

И величину $\sigma = \{ S_{cp}/(n\text{-}m)\}^{1/2}$ с $n = 8$ (число экспериментальных данных), $m = 2$ (число параметров модели). Получаем $S_{cp} = 0.823$, $\sigma = 0.370$.

Шаг 4 . Дальнейшие расчеты можно провести по формулам Таблицы 5.1 . Однако для случая 2 – х параметров расчеты можно провести по формулам из Таблицы 7.3. По этим формулам определяем среднеквадратичные отклонения $\sigma(x_k)$ в параметрах $x_1 = a$ и $x_2 = b$.

Таблица 7. 3. Уравнения для *вычисления* ошибок в 2-х параметрах модели, полученных из n наблюдаемых.

$\sigma(x_1) = \sigma\cdot(c_2/ d)^{1/2}$, $\sigma(x_2) = \sigma\cdot(c_1/ d)^{1/2}$, $\rho(x_1, x_2) = -c_{12}/ (c_1 c_2)^{1/2}$,

$\sigma = \{S_{cp} / (n-2)\}^{1/2}$,

$a_{i1} = \partial Y_i/ \partial x_1$, $a_{i2} = \partial Y_i/ \partial x_2$,

$$c_1 = \sum_{i=1}^{n} w_{ii} a_{i1}^2 , \quad c_2 = \sum_{i=1}^{n} w_{ii} a_{i2}^2 , \quad c_{12} = \sum_{i=1}^{n} w_{ii} a_{i1} \cdot a_{i2} ,$$

$d = c_1 c_2 - (c_{12})^2$

w_{ii} – веса, приписываемые Y_i, S_{cp}- взвешенная сумма квадратов отклонений (2), определяемая с найденными параметрами x_1 и x_2.

Так как использовалась линейная модель (1), то $a_{i1} = \partial Y_i/ \partial x_1 = 1.0$, $a_{i2} = \partial Y_i/ \partial x_2$, $= t_i$.

В результате вычислений получено: $c_1 = 1386.12$; $c_2 = 3.63\cdot10^7$, $c_{12} = 180183.$; $d = 1.79\cdot10^{10}$, $\sigma(x_1) = \sigma(a) = 0.021$, $\sigma(x_1) = \sigma(b) = 0.00013$, $\rho(x_1, x_2) = -0.802$.

Шаг 5. **Построение доверительных интервалов**. Верхняя $x_k^{(+)}$ и нижняя $x_k^{(-)}$ границы полученных параметров определяются соотношением (14) из предыдущей Лабораторной работы, т.е.

$$x_k^{(\pm)} = x_k \pm t_\gamma \ \sigma(x_k),$$

в котором t_γ находится по таблице Стьюдента по заданному γ. При $\gamma = 0.75$ $t_\gamma = 1.0$, и доверительный интервал полностью определяется значениями $\sigma(x_k)$. Результат расчета параметров a и b и их доверительных интервалов записываем для $t_\gamma = 1.0$ в виде

$$a = 1.988 \pm 0.021,$$
$$b = 2.960 \pm 0.00013$$

Это и есть окончательный ответ.

Задачи к Лабораторной работе «Определение параметров линейной модели методом наименьших квадратов и построение для них доверительных интервалов» **.**

Задача 7.1 Для заданных рядов наблюдений при t_i =1,2,3,4,5,6,7,8 методом наименьших квадратов найти неизвестные параметры a и b в линейной модели $y(t) = a$

$+b\cdot t$ и построить для них доверительные интервалы для γ =0. 75 (t_γ =1.0). Данные взять из Таб. 7..4.

Таблица 7.4. Данные к задаче 7.1.

Вар.	Ряд данных Y =
1	k=1, Y_{i1} = {4.07,6.18,8.35,10.58,12.87,15.22,17.63,20.1} k=2, Y_{i2} = {4.1,6.21,8.38,10.61,12.9,15.25,17.66,20.13} k=3, Y_{i3} = {3.99,6.06,8.19,10.38,12.63,14.94,17.31,19.74} k=4, Y_{i4} = {3.96,6.03,8.16,10.35,12.6,14.91,17.28,19.71}
2	k=1, Y_{i1} = {5.07,8.18,11.35,14.58,17.87,21.22,24.63,28.1} k=2, Y_{i2} = {5.1,8.21,11.38,14.61,17.9,21.25,24.66,28.13} k=3, Y_{i3} = {4.99,8.06,11.19,14.38,17.63,20.94,24.31,27.74} k=4, Y_{i4} = {4.96,8.03,11.16,14.35,17.6,20.91,24.28,27.71}
3	k=1, Y_{i1} = {6.07,9.18,12.35,15.58,18.87,22.22,25.63,29.1} k=2, Y_{i2} = {6.1,9.21,12.38,15.61,18.9,22.25,25.66,29.13} k=3, Y_{i3} = {5.99,9.06,12.19,15.38,18.63,21.94,25.31,28.74} k=4, Y_{i4} = {5.96,9.03,12.16,15.35,18.6,21.91,25.28,28.71}
4	k=1, Y_{i1} = {5.07,7.18,9.35,11.58,13.87,16.22,18.63,21.1} k=2, Y_{i2} = {5.1,7.21,9.38,11.61,13.9,16.25,18.66,21.13} k=3, Y_{i3} = {4.99,7.06,9.19,11.38,13.63,15.94,18.31,20.74} k=4, Y_{i4} = {4.96,7.03,9.16,11.35,13.6,15.91,18.28,20.71}
5	k=1, Y_{i1} = {7.08,11.22,15.44,19.74,24.12,28.58,33.12,37.74} k=2, Y_{i2} = {7.11,11.25,15.47,19.77,24.15,28.61,33.15,37.77} k=3, Y_{i3} = {7.,11.1,15.28,19.54,23.88,28.3,32.8,37.38} k=4, Y_{i4} = {6.97,11.07,15.25,19.51,23.85,28.27,32.77,37.35}
6	k=1, Y_{i1} = {7.08,10.22,13.44,16.74,20.12,23.58,27.12,30.74} k=2, Y_{i2} = {7.11,10.25,13.47,16.77,20.15,23.61,27.15,30.77} k=3, Y_{i3} = {7.,10.1,13.28,16.54,19.88,23.3,26.8,30.38} k=4, Y_{i4} = {6.97,10.07,13.25,16.51,19.85,23.27,26.77,30.35}
7	k=1, Y_{i1} = {8.08,12.22,16.44,20.74,25.12,29.58,34.12,38.74} k=2, Y_{i2} = {8.11,12.25,16.47,20.77,25.15,29.61,34.15,38.77} k=3, Y_{i3} = {8.,12.1,16.28,20.54,24.88,29.3,33.8,38.38} k=4, Y_{i4} = {7.97,12.07,16.25,20.51,24.85,29.27,33.77,38.35}
8	k=1, Y_{i1} = {9.08,14.22,19.44,24.74,30.12,35.58,41.12,46.74} k=2, Y_{i2} = {9.11,14.25,19.47,24.77,30.15,35.61,41.15,46.77} k=3, Y_{i3} = {9.,14.1,19.28,24.54,29.88,35.3,40.8,46.38} k=4, Y_{i4} = {8.97,14.07,19.25,24.51,29.85,35.27,40.77,46.35}
9	k=1, Y_{i1} = {9.08,13.22,17.44,21.74,26.12,30.58,35.12,39.74} k=2, Y_{i2} = {9.11,13.25,17.47,21.77,26.15,30.61,35.15,39.77} k=3, Y_{i3} = {9.,13.1,17.28,21.54,25.88,30.3,34.8,39.38} k=4, Y_{i4} = {8.97,13.07,17.25,21.51,25.85,30.27,34.77,39.35}
10	k=1, Y_{i1} = {10.08,15.22,20.44,25.74,31.12,36.58,42.12,47.74} k=2, Y_{i2} = {10.11,15.25,20.47,25.77,31.15,36.61,42.15,47.77} k=3, Y_{i3} = {10.,15.1,20.28,25.54,30.88,36.3,41.8,47.38} k=4, Y_{i4} = {9.97,15.07,20.25,25.51,30.85,36.27,41.77,47.35}
11	k=1, Y_{i1} = {11.08,17.22,23.44,29.74,36.12,42.58,49.12,55.74} k=2, Y_{i2} = {11.11,17.25,23.47,29.77,36.15,42.61,49.15,55.77} k=3, Y_{i3} = {11.,17.1,23.28,29.54,35.88,42.3,48.8,55.38} k=4, Y_{i4} = {10.97,17.07,23.25,29.51,35.85,42.27,48.77,55.35}

Продолжение

Таблица 7.4. Данные к задаче 7.1.

12	k=1, Y_{i1} = {11.08,16.22,21.44,26.74,32.12,37.58,43.12,48.74} k=2, Y_{i2} = {11.11,16.25,21.47,26.77,32.15,37.61,43.15,48.77} k=3, Y_{i3} = {11.,16.1,21.28,26.54,31.88,37.3,42.8,48.38} k=4, Y_{i4} = {10.97,16.07,21.25,26.51,31.85,37.27,42.77,48.35}
13	k=1, Y_{i1} = {12.08,18.22,24.44,30.74,37.12,43.58,50.12,56.74} k=2, Y_{i2} = {12.11,18.25,24.47,30.77,37.15,43.61,50.15,56.77} k=3, Y_{i3} = {12.,18.1,24.28,30.54,36.88,43.3,49.8,56.38} k=4, Y_{i4} = {11.97,18.07,24.25,30.51,36.85,43.27,49.77,56.35}
14	k=1, Y_{i1} = {13.09,20.26,27.53,34.9,42.37,49.94,57.61,65.38} k=2, Y_{i2} = {13.12,20.29,27.56,34.93,42.4,49.97,57.64,65.41} k=3, Y_{i3} = {13.01,20.14,27.37,34.7,42.13,49.66,57.29,65.02} k=4, Y_{i4} = {12.98,20.11,27.34,34.67,42.1,49.63,57.26,64.99}
15	k=1, Y_{i1} = {14.09,21.26,28.53,35.9,43.37,50.94,58.61,66.38} k=2, Y_{i2} = {14.12,21.29,28.56,35.93,43.4,50.97,58.64,66.41} k=3, Y_{i3} = {14.01,21.14,28.37,35.7,43.13,50.66,58.29,66.02} k=4, Y_{i4} = {13.98,21.11,28.34,35.67,43.1,50.63,58.26,65.99}
16	k=1, Y_{i1} = {15.11,23.34,31.71,40.22,48.87,57.66,66.59,75.66} k=2, Y_{i2} = {15.14,23.37,31.74,40.25,48.9,57.69,66.62,75.69} k=3, Y_{i3} = {15.03,23.22,31.55,40.02,48.63,57.38,66.27,75.3} k=4, Y_{i4} = {15.,23.19,31.52,39.99,48.6,57.35,66.24,75.27}
17	k=1, Y_{i1} = {15.11,22.34,29.71,37.22,44.87,52.66,60.59,68.66} k=2, Y_{i2} = {15.14,22.37,29.74,37.25,44.9,52.69,60.62,68.69} k=3, Y_{i3} = {15.03,22.22,29.55,37.02,44.63,52.38,60.27,68.3} k=4, Y_{i4} = {15.,22.19,29.52,36.99,44.6,52.35,60.24,68.27}
18	k=1, Y_{i1} = {16.11,24.34,32.71,41.22,49.87,58.66,67.59,76.66} k=2, Y_{i2} = {16.14,24.37,32.74,41.25,49.9,58.69,67.62,76.69} k=3, Y_{i3} = {16.03,24.22,32.55,41.02,49.63,58.38,67.27,76.3} k=4, Y_{i4} = {16.,24.19,32.52,40.99,49.6,58.35,67.24,76.27}
19	k=1, Y_{i1} = {17.11,26.34,35.71,45.22,54.87,64.66,74.59,84.66} k=2, Y_{i2} = {17.14,26.37,35.74,45.25,54.9,64.69,74.62,84.69} k=3, Y_{i3} = {17.03,26.22,35.55,45.02,54.63,64.38,74.27,84.3} k=4, Y_{i4} = {17.,26.19,35.52,44.99,54.6,64.35,74.24,84.27}
20	k=1, Y_{i1} = {17.12,25.38,33.8,42.38,51.12,60.02,69.08,78.3} k=2, Y_{i2} = {17.15,25.41,33.83,42.41,51.15,60.05,69.11,78.33} k=3, Y_{i3} = {17.04,25.26,33.64,42.18,50.88,59.74,68.76,77.94} k=4, Y_{i4} = {17.01,25.23,33.61,42.15,50.85,59.71,68.73,77.91}

Лабораторная работа «Предварительный анализ и сглаживание временного ряда. Выявление тренда»

Цель работы:

1.Выявление и устранение аномальных уровней ряда.

2.Выявление наличия тренда.

3.Сглаживание временных рядов.

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 4 часа.

Теоретические основы и примеры.

Определения:

$y_1, y_2, \ldots y_n$ - временной ряд, состоящий из n – уровней,

$$y_i = u_i + \varepsilon_i \quad (1)$$

u_i - тренд, ε_i - ошибки измерений y_i , случайные составляющие результатов измерений y_i

Для заданного ряда работа по выявлению наличия тренда выполняется в несколько этапов.

Этап 1 . Выявление аномальных уровней ряда. Процедура проводится с использованием статистического критерия Ирвина. Для ряда $\{y_i\}$ вычисляются

$$\lambda_i = \frac{|y_i - y_{i-1}|}{\sigma(y)} \quad , i = 2,3,.. n \quad (2)$$

которые сравниваются с табличными значениями критерия Ирвина λ_α . В (2) $\sigma(y)$ вычисляется по формуле (5) Лабораторной работы «Статистика ошибок в параметрах модели» с $N = n$ -1. Если какое- либо λ_i оказывается больше табличных, то соответствующее значение y_i уровня ряда считается аномальным. Значения λ_α для уровня значимости α =0.05 приведено в Таблице 8. 1.

Таблица 8. 1. Значения критерия Ирвина λ_α для α =0.05.

n	2	3	10	20	30	50	100
λ_α	2.8	2.3	1.5	1.3	1.2	1.1	1.0

Пример 8.1. Для ряда y= {10.0; 12.0; 9.0 ;8.0 ;9.0 ;4.0 ;6.0 ;9.0 ; 10.0 ; 8.0} определить аномальные уровни с 5% ошибкой.

Для данного ряда n =10, $\bar{y} = 8.5$ и $\sigma(y)$= 2, 22. По формуле (2) вычисляем значения λ_i :

i	1	2	3	4	5	6	7	8	9	10
λ_i	-	0.89	1.34	0.44	0.44	2.24	0.89	1.34	0.44	0.89

Из Таб. 8.1 для n =10 значение λ_α =1.5. В наших λ_i значение $\lambda_{i=6}$ =2.24 > λ_α =1.5, поэтому уровень ряда y_i с номером i =6 следует считать аномальным.

Этап 2. Проверка наличия тренда. Осуществляется в несколько шагов.

1.Исходный временной ряд разбивается на две примерно равные по числу уровней части. Первая часть содержит n_1 уровней, а вторая n_2 – уровней, $n = n_1 + n_2$.

2. Для каждой из частей вычисляются средние значения $\bar{y}_1$, $\bar{y}_2$ и дисперсии $\sigma^2(y_1) \equiv \sigma^2_1$, $\sigma^2(y_2) \equiv \sigma^2_2$.

3. Проверяется равенство (однородности) дисперсий с помощью – критерия Фишера, для чего вычисляется значение

$$F = \sigma^2(y_1) / \sigma^2(y_2), \text{ если } \sigma^2(y_1) > \sigma^2(y_2),$$
$$F = \sigma^2(y_2) / \sigma^2(y_1), \text{ если } \sigma^2(y_1) < \sigma^2(y_2), \quad (1)$$

и сравнивается с табличным (критическим) значением критерия Фишера F_α с заданным уровнем значимости α. Если вычисленное значение F меньше табличного F_α , то гипотеза о равенстве дисперсий принимается. Если F больше F_α , то гипотеза о равенстве дисперсий отклоняется и делается вывод, что данный метод для определения наличия тренда ответа не дает.

Пример 8.2. Проверить однородность ряда y= {10.0; 12.0; 11.0 ;12.0;15.0 ;18.0; 20.0} с помощью критерия Фишера . Разбиваем ряд на 2 части {10.0; 12.0; 11.0 ;12.0} с n_1 =4 и {15.0 ;18.0; 20.0} с n_2= 3. Определяем средние значения $\bar{y}_1 = 11.25$, $\bar{y}_2 = 17.67$ и дисперсии $\sigma^2(y_1)$ =0.917, $\sigma^2(y_1)$ =6.333. Так как $\sigma^2(y_1)$ > $\sigma^2(y_1)$, то

$F = \sigma^2(y_2) / \sigma^2(y_1)$=6.333/0.917= 6.9062.

Табличное значение F_α = F_α (n_2= 3, n_1 =4) = 19.1643. Так как $F < F_\alpha$, то гипотеза о равенстве дисперсий принимается.

4. Проверяется гипотеза об отсутствии тренда с использованием t - критерия Стьюдента. Для этого вычисляется критерий

$$t = \frac{|\bar{y}_1 - \bar{y}_2|}{\sigma\sqrt{1/n_1 + 1/n_2}} \tag{4}$$

$$\sigma = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}} \tag{5}$$

и сравнивается с табличным критерием t_α . Если $t < t_\alpha$,то гипотеза об отсутствии тренда принимается (тренда нет), в противном случае тренд есть.

Для примера 8.2

$$\sigma = \sqrt{\frac{3 \cdot 0.917 + 2 \cdot 6.333}{(n_1 + n_2 - 2) = 5}} = 1.7559\,,\ t = 4.003.$$

По таблице критерия Стьюдента t_α =2.015 для уровня значимости α = 0.05. Для этого уровня значимости расчетное значение t =4.003 > t_α =2.015 и можно предположить наличие тренда.

Этап 3. Сглаживание временных рядов. Применяется несколько методов.

1. Метод простой скользящей средней.

Вычисляется среднее значение для первых $2к$+1 уровней. Это и будет первое сглаженное значение уровня ряда. Затем интервал сглаживания сдвигается на один уровень вправо и т.д. Расчет проводится по формуле

$$\bar{y}_i = \left(\frac{1}{2k+1}\right)\sum_{j=i-k}^{i+k} y_j \tag{6}$$

В результате получаются n -2k сглаженных значений $\bar{y}_i$, первые и последние k уровней теряются.

2. Метод взвешенной скользящей средней. Отличается от предыдущего метода тем, что в сумме (6) используются веса, так что расчет проводится по формуле

$$\bar{y}_i = \frac{\sum_{j=i-k}^{i+k} \rho_j y_j}{\sum_{j=i-k}^{i+k} \rho_j} \tag{7}$$

Веса для различных степеней аппроксимирующих полиномов и интервалов сглаживания приведены в Таблице 2.

Таблица 8. 2. Веса для аппроксимирующих полиномов.

Порядок полинома	Интервал сглаживания k	Последовательность весов
2 или 3	2	{-3,12,17,12,-3}
2 или 3	3	{-2,3,6,7,6,3,-2}
4 или 5	3	{5,-30,75,131,75,-30,5}

3. Метод экспоненциального сглаживания. Сглаженные значения s_i определяются по формуле

$$s_i = \alpha\, y_i + (1-\alpha)\, s_{i-1} \tag{8}$$

где α - параметр сглаживания, $0 < \alpha < 1.0$.

Пример 8.3. Для ряда y ={12.0; 23.0;25.0;24.0;25.0;22.0;21.0;18.0;16.0; 14}получить сглаженные значения методом простой скользящей средней и методом экспоненциального сглаживания.

По методу простой скользящей среднее первый уровень ряда отсутствует, второй уровень ряда вычисляется по формуле (6) как

$$s_2=(1/3)(y_1+y_2+y_3);$$

далее в этой формуле нижний индекс увеличивается на единицу, предпоследний уровень ряда вычисляется как

$$s_9=(1/3)(y_8+y_9+y_{10}) ;$$

В результате получается сглаженный ряд

s ={-; 20.0; 24.0; 24.66; 23.66; 22.66; 20.33; 18.33; 16.; - }

Последний уровень ряда, как и первый, теряется.

Расчет по формуле (8) с α =0.1 и с $s_0 = s_2=(1/3)(y_1+y_2+y_3)$ дает ряд

s ={ 19.2; 18.28; 17.55; 17.09; 16.78; 16.60; 16.34; 16.01; 15.91; 15.92}

Задачи к Лабораторной работе «Предварительный анализ и сглаживание временного ряда. Выявление тренда»

Задача 8. 1 Проверить ряд на наличие выбросов с использованием критерия Ирвина (этап1) , проверить ряд на наличие тренда (этап 2), сгладить 3-мя рассмотренными выше методами (этап 3). В методе скользящих средних взять k =2, в методе экспоненциального сглаживания взять α =0.2 и начальное значение s_0 в формуле (8) взять в виде среднего первых трех значений ряда, т.е. $s_0 = (y_1 + y_2 + y_3)/3$

Представить результаты графически (графики исходного и сглаженного ряда). Данные взять из Таб. 8.3

Таблица 8.3. Данные к задаче 8.1.

Вар.	Ряд данных y =
1	{4,8,11,11,17,20,24,26,29,32}
2	{5,8,9,11,11,16,17,19,21,23}
3	{8,10,13,16,19,21,25,28,31,34}
4	{7,12,15,19,23,27,31,35,39,43}
5	{11,13,23,29,35,41,47,53,59,65}
6	{11,15,21,26,31,36,41,46,51,56}
7	{15,22,31,39,47,55,63,71,79,87}
8	{15,22,28,36,43,50,57,64,71,78}
9	{19,29,38,49,59,69,79,89,99,109}
10	{19,28,38,46,55,64,73,82,91,100}
11	{23,35,48,59,71,83,95,107,119,131}
12	{23,34,46,56,67,78,89,100,111,122}
13	{27,41,56,69,83,97,111,125,139,153}
14	{27,40,54,66,79,92,105,118,131,144}
15	{31,47,64,79,95,111,127,143,159,175}
16	{33,50,68,84,101,118,135,152,169,186}
17	{33,49,66,81,97,113,129,145,161,177}
18	{37,56,76,94,113,132,151,170,189,208}
19	{37,55,74,91,109,127,145,163,181,199}
20	{41,62,84,104,125,146,167,188,209,230}

Лабораторная работа «Определение параметров трендовой модели» **.**

Цель работы:

1.Определение параметров линейной трендовой модели методом наименьших квадратов.

2.Оценка адекватности и точности трендовой модели.

3. Прогнозирование динамики на основе трендовой модели.

Форма проведения: Выполнение индивидуального задания.

Продолжительность выполнения работы: 4 часа.

Задача определение параметров трендовой модели решается в несколько этапов.

Этап 1. Нахождение параметров простейшей модели.

Шаг 1.

Для временного ряда $\{y_1, y_2, \dots y_n\}$ выбирается простейшая теоретическая модель

$$y(t) = a + b \cdot t \quad (1)$$

Шаг 2.

Методом наименьших квадратов, из системы уравнений

$$\begin{aligned} a \cdot n + b \cdot \sum_{i=1}^{n} t_i &= \sum_{i=1}^{n} y_i \\ a \cdot \sum_{i=1}^{n} t_i + b \cdot \sum_{i=1}^{n} t_i^2 &= \sum_{i=1}^{n} t_i y_i \end{aligned} \quad (2)$$

определяются a и b по формулам из Таблицы 7.2 . В частном случае $w_{ii} = 1$

$$\begin{gathered} a = (\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} t_i^2 - \sum_{i=1}^{n} t_i \cdot \sum_{i=1}^{n} t_i \cdot y_i) / \Delta \\ b = (n \cdot \sum_{i=1}^{n} t_i \cdot y_i - \sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} t_i) / \Delta \\ \Delta = n \cdot \sum_{i=1}^{n} t_i^2 - \sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} t_i \end{gathered} \quad (3)$$

Функция $y(t) = a + b \cdot t$ объявляется трендом.

Пример 9.1 Для временного ряда

y= {20.0, 23.0, 25.0, 24.0, 25.0, 22.0, 21.0, 17.0, 15.0, 11.0}

n =10, $\sum_{i}^{10} t_i = 55$**,** $\sum_{i}^{10} t_i^2 = 385$**,** по формулам (3) определяем a =26.6, b = -1.145,

и функция $y(t)$ = 26.6-1.145·t объявляется трендом.

Этап 2. Проверка адекватности и точности трендовой модели.

Шаг 1. С найденными параметрами a и b вычисляются разности

$$\varepsilon_i = y_i - (a + b \cdot t_i) \tag{4}$$

и одновременно

$$\varepsilon^2 = (\sum_{i=1}^{n} \varepsilon_i^2)/n \tag{5}$$
$$\varepsilon^3 = (\sum_{i=1}^{n} \varepsilon_i^3)/n$$
$$\varepsilon^4 = (\sum_{i=1}^{n} \varepsilon_i^4)/n$$

Для ряда из примера 9.1 значения $y^{теор.}{}_i = y(t_i)$ = 23.4-0.709·t_i , ε_i , ε_i^2 , ε_i^3 и ε_i^4

представлены в таблице 9.1.

Таблица 9.1. Результаты расчетов отклонений ε_i , ε_i^2 , ε_i^3, ε_i^4 и поворотных точек p_i для примера 9.1

t_i	y	$y^{теор}$	ε_i	p_i	ε_i^2	ε_i^3	ε_i^4
1	20.0	25.45	-5.45	-	29.75	-162.28	885.18
2	23.0	24.30	-1.30	-	1.71	-2.24	2.93
3	25.0	23.16	1.83	-	3.37	-.19	11.37
4	24.0	22.01	1.98	-	3.92	7.78	15.42
5	25.0	20.87	4.12	1	17.03	70.30	290.17
6	22.0	19.72	2.27	1	5.16	11.73	26.68
7	21.0	18.58	2.41	1	5.84	14.14	34.19
8	17.0	17.43	-0.43	-	0.19	-0.08	0.036
9	15.0	16.29	-1.29	-	1.66	-2.15	2.77
10	11.0	15.14	-4.14	-	17.18	-71.23	295.31
$\sum \varepsilon_i^k/n$			$1.2 \cdot 10^{-15}$	3	8.58	12.78	156.41

Шаг 2. Проверка случайности колебаний $\varepsilon_{i.}$. Для $\varepsilon_{i.}$ находят число пиков p (поворотных точек). Для пиков $\varepsilon_{t.}$ выполняется условие $\varepsilon_{t-1.} < \varepsilon_{t.} > \varepsilon_{t+1}$ или $\varepsilon_{t-1} > \varepsilon_{t.} < \varepsilon_{t+1}$

Для ряда из примера 9.1 поворотные точки показаны в 5-м столбце Таб. 9.1. Их число p =3.

Далее вычисляются

$$p_{cp} = (n - 2) / 3, \qquad (6)$$

$$(\sigma_p)^2 = (16n - 29)/90$$

и проверяется равенство

$$p > [p_{cp} - 1.96\sigma_p]$$

в котором квадратные скобки обозначают целую часть числа. Если оно выполняется, то с 5% уровнем значимости пики распределены случайно.

Для ряда из примера 9.1 n = 10, p_{cp} = (10-2)/3= 5.33, $(\sigma_p)^2$ = (16·10 -29)/90=1.45, p_{cp} - 1.96σ_p = 2.96 и [2.96]=2.0. Т.к. p >2 , то свойство случайности колебаний ε_i. принимается.

Шаг 3. Проверка соответствия распределения ε_i нормальному закону распределения.

Для этого вычисляются выборочные характеристики асимметрии (γ_1) и эксцесса (γ_2) и их ошибки $\sigma(\gamma_1)$, $\sigma(\gamma_2)$

$$\gamma_1 = \varepsilon^3 / (\varepsilon^2)^3, \qquad \sigma(\gamma_1) = \{6(n-2)/[(n+1)(n+3)]\}^{1/2} \qquad (7)$$

$$\gamma_2 = \varepsilon^4 / (\varepsilon^2)^2 - 3 \qquad \sigma(\gamma_1) = \{6(n-2)/[(n+1)(n+3)]\}^{1/2}$$

Если одновременно выполняются неравенства

$$|\gamma_1| < 1.5\sigma(\gamma_1)\ ; \quad |\gamma_2 + 6/(n+1)| < 1.5\sigma(\gamma_2)\ ; \qquad (8)$$

то гипотеза о нормальном распределении ε_i принимается. Если выполняется хотя бы одно из неравенств

$$|\gamma_1| > 1.5\sigma(\gamma_1)\ ; \quad |\gamma_2 + 6/(n+1)| > 1.5\sigma(\gamma_2)\ , \qquad (9)$$

то такая гипотеза отвергается и трендовая модель считается неадекватной.

Для ряда из примера 9.1

$\sigma(\gamma_1)$= 0.579, $\sigma(\gamma_2)$= 0.238, 1.5·$\sigma(\gamma_1)$ =0.869, 1.5 ·$\sigma(\gamma_2)$= 0.356, γ_1 = 0.508, γ_2 = -0.878, $|\gamma_2 + 6/(n+1)|$= 0.332. Из этих чисел видно, что оба неравенства (9) выполняются и гипотеза о нормальном распределении случайной компоненты ε_i принимается.

Шаг 4. Проверка гипотезы $\bar{\varepsilon} = 0$

Рассчитывается критерий

$$t = \bar{\varepsilon} \cdot \sqrt{n} / S_\varepsilon \tag{10}$$

в котором $\bar{\varepsilon}$ - среднее значение для ε_i , а $S_\varepsilon = [\sum (\varepsilon_i)^2 / n]^{1/2}$. Если расчетное значение t меньше табличного t_α статистики Стьюдента с заданным уровнем значимости α и числом степеней свободы n -1, то гипотеза о равенстве нулю математического ожидания ε принимается, в противном случае она отвергается и модель считается неадекватной.

Для ряда из примера 9.1, согласно Таб. 9.1, $\bar{\varepsilon} = 1.5 \cdot 10^{-15}$. Эта величина близка к нулю и можно без использования статистики Стьюдента принять гипотезу $\bar{\varepsilon} = 0$.

Шаг 5. Проверка независимости значений ε_i Проверка делается с использованием критерия Дарбина-Уотсона (ДУ). Для этого рассчитывается критерий

$$d = \sum_{i=2}^{n} (\varepsilon_i - \varepsilon_{i-1})^2 / \sum_{i=1}^{n} \varepsilon_i^2 \tag{11}$$

Если d попало в интервал от 2 до 4, то его нужно преобразовать к d' =4- d. Расчетное значение d (или d') нужно сравнить с верхним d_2 или нижним d_1 критическими значениями статистики ДУ из Таблицы 9.2.

Таблица 9.2. Критические значения статистики Дарбина-Уотсона (для уровня значимости α=0.05)

n	d_1	d_2
15	1.08	1.36
20	1.20	1.41
30	1.35	1.49

Если расчетное значение d больше табличного d_2 , то гипотеза о независимости ε_i принимается. Если расчетное значение d меньше нижнего табличного значения d_1 , то гипотеза отвергается и модель неадекватна. Если расчетное значение d лежит в интервале от d_1 до d_2 , включая сами эти значения, то никакого вывода не делается.

Если все четыре проверки относительно ε_i дали положительный результат, то делается вывод об адекватности трендовой модели.

Для ряда из примера 9.1 расчетное значение d = 0.607. Это значение меньше нижнего значения d_1 из табл. 9. (для n =15) , оно меньше и для нашего случая.

Шаг 6. Оценка точности модели. Проводится с использованием величины

$$\bar{\varepsilon}_o = \frac{1}{n}\sum_{i=1}^{n}\left|\varepsilon_i / y_i\right| \cdot 100\% \tag{12}$$

определяющей ошибку аппроксимации ряда трендом. Для ряда из примера 9.1

$\bar{\varepsilon}_o = 13.57\%$. Для удовлетворительного уровня точности $\bar{\varepsilon}_o < 5.\%$

Этап 3. Интервальный прогноз на базе трендовой модели. Рассчитывается доверительный интервал для y_{n+L} (L – период упреждения).

Шаг 1. Рассчитывается величина

$$S_y = \sqrt{\frac{\sum_{i=1}^{n}(\varepsilon_i)^2}{n-k}} \tag{13}$$

где k =2 – число параметров модели. Доверительный интервал прогноза U_y определяется выражением

$$U_y = y_{n+L} \pm K \cdot S_y \tag{14}$$

где y_{n+L} - точечный прогноз для модели на $(n + L)$ –й шаг, а

$$K = t_\alpha \cdot \sqrt{1 + \frac{1}{n} + \frac{3(n+2L-1)^2}{n(n2-1)}} \tag{15}$$

в котором t_α - табличное значение критерия Стьюдента для уровня значимости α и для числа степеней свободы n -2. Значения K для α=0.05 представлено в Таблице 9.3.

Таблица 9. 3. Значения величины K (15) для оценки доверительного интервала прогноза линейного тренда для α=0.05 .

Число уровней n	Период упреждения L					
	1	2	3	4	5	6
7	1.932	2.106	2.300	2.510	2.733	2.965
10	1.692	1.774	1.865	1.964	2.069	2.180
13	1.581	1.629	1.682	1.738	1.799	1.863
15	1.536	1.572	1.611	1.653	1.697	1.745

Пример 9.2. Пусть для ряда из примера 9.1 необходимо сделать точечный и интервальный прогноз на три шага вперед. Точечный прогноз, т.е. значения y_{11} = 14, y_{12} = 12.85, y_{13} = 11.70, получаются из тренда $y^{теор.}{}_i = y(t_i) = 23.4\text{-}0.709 \cdot t_i$ при t_i =11, 12 и 13, соответственно. Для интервального прогнозирования вычисляем по формуле (13)

S_y = = 3.27, из таблицы 9.3 для L =1, 2 и 3 и для n =10 находим K=1.69, K=1.77 и K=1.86, соответственно. После этого по формуле (14) определяем нижнюю и верхнюю границу для прогнозируемого значения. Результат расчета можно представить в виде:

y_{11} = 14± 5.53, , y_{12} = 12.85±5.79, y_{13} = 11.70±6.09.

Задачи к Лабораторной работе «Определение параметров трендовой модели» **.**

Задача 9.1 Для временных рядов из Таблицы 8.3 определить трендовую модель в виде полинома первой степени, дать точечный и интервальный прогноз на 3 шага вперед. Результат представить графически.

Литература.

1.В.Е. Гмурман. Теория вероятностей и математическая статистика. М.Высшая школа. 1999. 478 с.

2. В.Е. Гмурман. Теория вероятностей и математическая статистика школа. М. . Высшая школа. 1999. 478 с.

3. В.Н. Калинина, В.Ф. Панкин. Математическая статистика. М. Высшая школа. 1998. 336 с.

4. З. Брант. Анализ данных. М. Мир. 2003. 686 с.

5. A.A.Clifford. Multivariate error analysis . A Handbook of error propagation and calculation in many-parameters system. Applied Science Publishers, Ltd. London,1973.

6. М. Г. Сидоренко. Математические методы в экономике. Учебное пособие. ТУСУР. Томск. 2000. 129 с.