Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего образования
«Омский государственный технический университет»

Факультет информационных технологий и компьютерных систем
Кафедра «Прикладная математика и фундаментальная информатика»

Лабораторная работа
по дисциплине «Теория вероятностей и математическая статистика»

Студента	Курпенова Куата Ибраимовича
	фамилия, имя, отчество полностью
Курс	3, группа ФИТ-222
Направление	02.03.02 Фундаментальная информатика и информационные технологии
	код, наименование
Руководитель	доц., канд. тех. наук
	должность, ученая степень, звание
	Болдовская Т. Е.
	фамилия, инициалы
Выполнил	
	дата, подпись студента
Проверил	
	дата, подпись руководителя

Омск 2025

Корреляционно-регрессионный анализ статистических данных

Вариант 12

Результаты наблюдения двумерной случайной величины (X, Y):

X/Y	1	2	3	4	5	6
1	2	1	-	-	-	-
2	1	2	-	-	-	-
3	-	3	1	-	-	-
4	-	1	3	1	-	-
5	-	-	2	2	2	1
6	-	-	-	1	1	1

Задание:

1. Найти групповое среднее $\overline{y_i}$ переменной Y.
2. В прямоугольной системе координат построить точки $(x_i, \overline{y_i})$ и ломаную линию регрессии Y на X.
3. Найти генеральные средние $\overline{x}$ и $\overline{y}$.
4. Составить уравнение линейной регрессии X на Y и Y на X. Построить график регрессии.
5. По выбранному значению переменной X сделать прогноз ожидаемого среднего значения переменной.
6. Установить тесноту связи между переменными величинами X и Y.
7. Оценить существенность выборочного коэффициента корреляции.
8. Найти 95%-й доверительные интервалы для среднего значения и коэффициентов уравнения регрессии.

1 Найти групповое среднее

1. $x_1 = 1; y_1 = \frac{1 \cdot 2 + 2 \cdot 1}{1+2} = \frac{4}{3}$
2. $x_2 = 2; y_2 = \frac{1 \cdot 1 + 2 \cdot 2}{1+2} = \frac{5}{3}$
3. $x_3 = 3; y_3 = \frac{2 \cdot 3 + 3 \cdot 1}{3+1} = \frac{9}{4}$
4. $x_4 = 4; y_4 = \frac{2 \cdot 1 + 3 \cdot 3 + 4 \cdot 1}{1+3+1} = \frac{15}{5} = 3$
5. $x_5 = 5; y_5 = \frac{3 \cdot 2 + 4 \cdot 2 + 5 \cdot 2 + 6 \cdot 1}{2+2+2+1} = \frac{30}{7}$
6. $x_6 = 6; y_6 = \frac{4 \cdot 1 + 5 \cdot 1 + 6 \cdot 1}{1+1+1} = \frac{15}{3} = 5$

2 Точки и линия регрессии

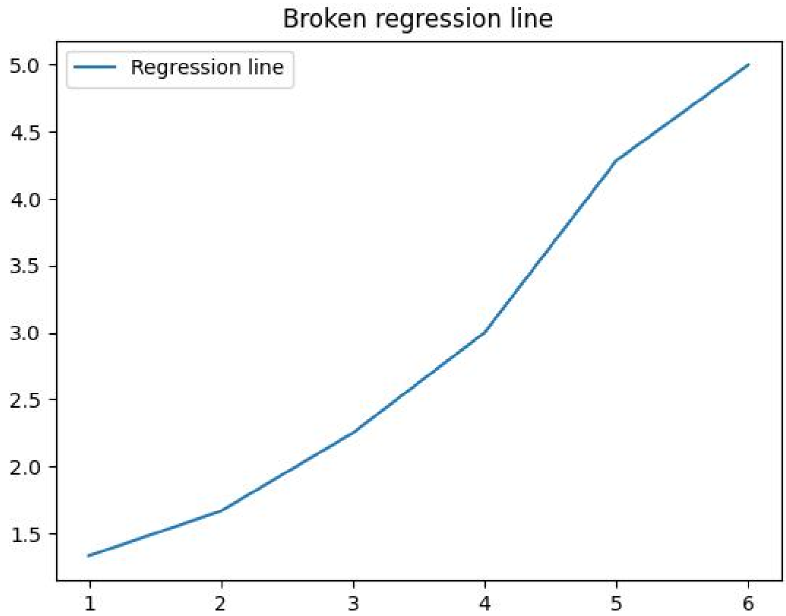

Рисунок 1 – Ломаная линия регрессии

3 Найти генеральные средние

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot n_{x_i} = \frac{1}{25}(1 \cdot 3 + 2 \cdot 3 + 3 \cdot 4 + 4 \cdot 5 + 5 \cdot 7 + 6 \cdot 3) = \frac{94}{25} = 3.76$$

$$\overline{y} = \frac{1}{n}\sum_{j=1}^{n} y_i \cdot n_{y_i} = \frac{1}{25}(1\cdot 3 + 2\cdot 7 + 3\cdot 6 + 4\cdot 4 + 5\cdot 3 + 6\cdot 2) = \frac{78}{25} = 3.12$$

4 Составить уравнение линейной регрессии

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 \cdot n_{x_i} = \frac{1}{25}(1^2\cdot 3 + 2^2\cdot 3 + 3^2\cdot 4 + 4^2\cdot 5 + 5^2\cdot 7 + 6^2\cdot 3) = \frac{414}{25} = 16.56$$

$$\overline{y^2} = \frac{1}{n}\sum_{j=1}^{n} {y_i}^2 \cdot n_{y_i} = \frac{1}{25}(1^2\cdot 3 + 2^2\cdot 7 + 3^2\cdot 6 + 4^2\cdot 4 + 5^2\cdot 3 + 6^2\cdot 2) = \frac{296}{25} = 11.8$$

$$\overline{xy} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} x_i y_i n_{ij} = \frac{1}{25}(1\cdot 1\cdot 2 + 1\cdot 2\cdot 1 + 2\cdot 1\cdot 1 + 2\cdot 2\cdot 2 + 3\cdot 2\cdot 3$$
$$+ 3\cdot 3\cdot 1 + 4\cdot 2\cdot 1 + 4\cdot 3\cdot 3 + 4\cdot 4\cdot 1 + 5\cdot 3\cdot 2 + 5\cdot 4\cdot 2 + 5\cdot 5\cdot 2$$
$$+ 5\cdot 6\cdot 1 + 6\cdot 4\cdot 1 + 6\cdot 5\cdot 1 + 6\cdot 6\cdot 1) = \frac{341}{25} = 13.64$$

Уравнение Y на X:

$$cov(X,Y) = \overline{xy} - \bar{x}\bar{y} = 13.64 - 3.76\cdot 3.12 = 1.91$$

$${\sigma_x}^2 = \overline{x^2} - \bar{x}^2 = 16.56 - (3.76)^2 = 2.42$$

$$a = \rho_{yx} = \frac{1.91}{2.42} = 0.79$$

$$b = \overline{y} - a\overline{x} = 3.12 - 0.79\cdot 3.76 = 0.15$$

$$y = ax + b = 0.79x + 0.15$$

Уравнение X на Y:

$$cov(X,Y) = \overline{xy} - \bar{x}\bar{y} = 13.64 - 3.76 \cdot 3.12 = 1.91$$

$$\sigma_y{}^2 = \overline{y^2} - \bar{y}^2 = 11.8 - (3.12)^2 = 2.0656$$

$$c = \rho_{xy} = \frac{1.91}{2.0656} = 0.925$$

$$d = \overline{x} - c\overline{y} = 3.76 - 0.925 \cdot 3.12 = 0.873$$

$$x = cy + d = 0.925y + 0.873$$

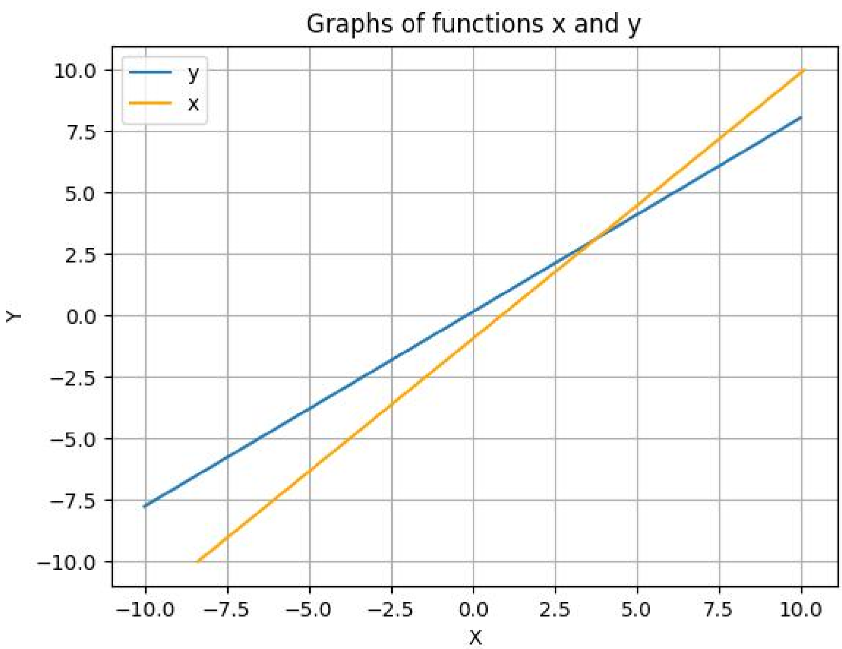

Рисунок 2 – Графики уравнений

5 Прогноз переменной

При $x = 3.76$, $\overline{y}$ равен: $\overline{y} = 0.79 * 3.76 + 0.15 = 3.1204$

6 Установить тесноту связи

$$\sigma_x = \sqrt{\sigma_x^2} = 1.556$$

$$\sigma_y = \sqrt{\sigma_y^2} = 1.437$$

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y} = \frac{13.64 - 3.76 \cdot 3.12}{1.556 \cdot 1.437} = 0.854$$

По шкале Чеддока оценка силы связи тесная.

7 Оценить существенность выборочного коэффициента корреляции

Примем, что: $H_0 : r = 0$ и $H_1 : r \neq 0$, тогда:

$$|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{|0.854| \cdot \sqrt{25-2}}{\sqrt{1-0.854^2}} = 7.885$$

При $\alpha = 0.05$, $t_{0.05;23} = 2.07$ по таблице Стьюдента.

7.88 > 2.07, следовательно, H_0 отвергается и признаётся статистическая значимость и надёжность уравнения. Линейная корреляционная связь между переменными присутствует.

8 Найти доверительный интервал

Найдём выборочную остаточную дисперсию:

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} (\overline{y_i} - y_i)^2 = \frac{1}{23}(\frac{1}{9} + \frac{1}{9} + \frac{9}{16} + 1 + \frac{25}{49} + 1) = 0.143$$

$$S = \sqrt{S^2} = \sqrt{0.143} = 0.378$$

Размах доверительных интервалов:

$$\Delta a = t_{a;n-2} \frac{S}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} =$$
$$2.07 \frac{0.378}{\sqrt{(-2.76)^2 + (-1.76)^2 + (-0.76)^2 + 0.24^2 + 1.24^2 + 2.24^2}} = 0.185$$

$$\Delta b = \Delta a \sqrt{\overline{x}} = 0.185 \cdot \sqrt{3.76} = 0.35$$

Получаем: $0.606 < a < 0.974$ и $-0.21 < b < 0.51$.

$$S_{y\ x=1}^{\ 2} = S^2\left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right) = 0.378\left(\frac{1}{25} + \frac{(1 - 3.76)^2}{17.9056}\right) = 0.176$$

$$S_{y_{x=1}} = 0.4194$$

$$t_{a;n} = t_{0.05;25} = 2.07$$

Доверительный интервал при $x = 1$: $0.0718 < \overline{y} < 1.808$.