

Типовой расчет

выполнил ст. гр. ***** Петров Ю.А.

Задача №11

Вариант XX

1 Условие

По выборке двумерной случайной величины:

- вычислить оценку коэффициента корреляции;
- проверить гипотезу об отсутствии корреляционной зависимости ($\alpha = 0.05$);
- вычислить параметры линии регрессии α_0 и α_1 ;
- построить диаграмму рассеивания и линию регрессии;

Исходные данные для варианта XX приведены в таблице 1.

Таблица 1 – Двумерная выборка

(-0.68; -0.26)	(-4.03; -2.32)	(-0.72; 0.47)	(1.25; 0.82)	(1.27; -0.81)
(-3.57; 0.46)	(3.00; -2.85)	(-2.19; 2.71)	(-4.72; 0.48)	(4.38; 2.77)
(1.16; 2.37)	(-1.04; 2.03)	(-0.63; 1.74)	(-0.07; -0.30)	(-1.55; 1.85)
(1.57; -0.10)	(-0.27; -0.84)	(-1.92; -0.17)	(-0.80; -0.27)	(-0.30; 3.87)
(-2.51; -1.20)	(0.21; 0.36)	(2.99; 2.78)	(2.26; 2.43)	(1.95; 0.79)
(3.27; 0.62)	(-0.40; 2.71)	(-0.53; 1.01)	(0.16; 2.11)	(3.07; 0.47)
(-0.87; -2.17)	(2.41; -0.85)	(-0.52; -1.54)	(0.99; -0.26)	(0.57; 1.41)
(1.47; -0.41)	(5.76; -1.11)	(-1.16; 0.95)	(-1.22; -3.60)	(3.13; 2.46)
(0.90; 0.79)	(0.77; -3.32)	(-0.80; -1.46)	(1.48; -0.69)	(0.18; 0.25)
(2.08; 2.50)	(-0.99; -2.73)	(-1.33; 1.70)	(-2.36; -2.75)	(-1.82; -2.29)

2 Решение

2.1 Вычисление точечных оценок параметров двумерной выборки

Вычислим оценки математических ожиданий по каждой переменной.

$$\begin{aligned} m_X^* = \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i, & m_Y^* = \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i, \\ m_X^* &= 0.1856, & m_Y^* &= 0.2122. \end{aligned} \quad (1)$$

Вычислим оценки дисперсий по каждой переменной.

$$\begin{aligned} D_X^* = S_0^2(x) &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \\ D_Y^* = S_0^2(y) &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2, \\ D_X^* &= 4.4893, \quad D_Y^* = 3.3579. \end{aligned} \quad (2)$$

Вычислим оценку корреляционного момента:

$$\begin{aligned} K_{XY}^* &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \\ K_{XY}^* &= 0.7663. \end{aligned} \quad (3)$$

Найдем оценку коэффициента корреляции:

$$\begin{aligned} R_{XY}^* &= \frac{K_{XY}^*}{\sqrt{S_0^2(x) \cdot S_0^2(y)}}, \\ R_{XY}^* &= \frac{0.7663}{\sqrt{4.4893 \cdot 3.3579}} = 0.1974. \end{aligned} \quad (4)$$

2.2 Вычисление интервальной оценки коэффициента корреляции

Вычислим интервальную оценку коэффициента корреляции с надёжностью $\gamma = 0.95$ по следующей формуле:

$$I_{\gamma}(R_{XY}) = \left[\frac{e^{2a} - 1}{e^{2a} + 1}; \frac{e^{2b} - 1}{e^{2b} + 1} \right]. \quad (5)$$

Для этого в таблице функции Лапласа найдем значение, равное $\frac{\gamma}{2} = 0.475$ и определим значение аргумента, ему соответствующее:

$$z_{0.95} = \arg\Phi(0.475) = 1.96.$$

Для вычисления интервальной оценки коэффициента корреляции найдем вспомогательные значения a, b по следующим формулам:

$$\begin{aligned} a &= 0.5 \cdot \ln \left(\frac{1 + R_{XY}^*}{1 - R_{XY}^*} \right) - \frac{z_{\gamma}}{\sqrt{n - 3}}, \\ b &= 0.5 \cdot \ln \left(\frac{1 + R_{XY}^*}{1 - R_{XY}^*} \right) + \frac{z_{\gamma}}{\sqrt{n - 3}}, \end{aligned} \quad (6)$$

$$a = -0.0859, \quad b = 0.4859.$$

Таким образом, доверительный интервал для коэффициента корреляции имеет вид:

$$\begin{aligned} I_{\gamma}(R_{XY}) &= \left[\frac{e^{2a} - 1}{e^{2a} + 1}; \frac{e^{2b} - 1}{e^{2b} + 1} \right], \\ I_{\gamma}(R_{XY}) &= [-0.0857; 0.4509] \end{aligned} \quad (7)$$

2.3 Проверка гипотезы об отсутствии корреляционной зависимости

Выдвинем двухальтернативную гипотезу об отсутствии корреляционной зависимости между величинами X и Y :

- $H_0 - R_{XY} = 0$: между величинами X и Y корреляционная зависимость отсутствует;
- $H_1 - R_{XY} \neq 0$: между величинами X и Y существует корреляционная зависимость.

Так как объем выборки велик ($n \geq 50$), то вычислим значение критерия по формуле:

$$Z = \frac{|R_{XY}^*| \cdot \sqrt{n}}{1 - (R_{XY}^*)^2}, \quad (8)$$

$$Z = 1.4236.$$

Определим значение Z_α из таблицы функции Лапласа ($\alpha = 0.05$):

$$Z_{0.05} = 1.96.$$

Вывод: так как $Z = 1.4236 < Z_{0.05}$, то гипотеза H_0 об отсутствии корреляционной зависимости между величинами X и Y принимается (отклоняется).

2.4 Построение линии регрессии

Уравнение линии регрессии имеет следующий вид:

$$\bar{y}(x) = a_0^* + a_1^*x, \quad (9)$$

где $a_1^* = \frac{K_{XY}^*}{S_0^2(x)}$, $a_0^* = \bar{y} - a_1^* \cdot \bar{x}$ – коэффициенты линии регрессии.

Найдем значения a_1^*, a_0^* :

$$a_1^* = 0.1707, \quad a_0^* = 0.1805$$

Таким образом, линия регрессии примет вид:

$$\bar{y}(x) = 0.1805 + 0.1707 \cdot x \quad (10)$$

График линии регрессии изображен на рисунке 1.

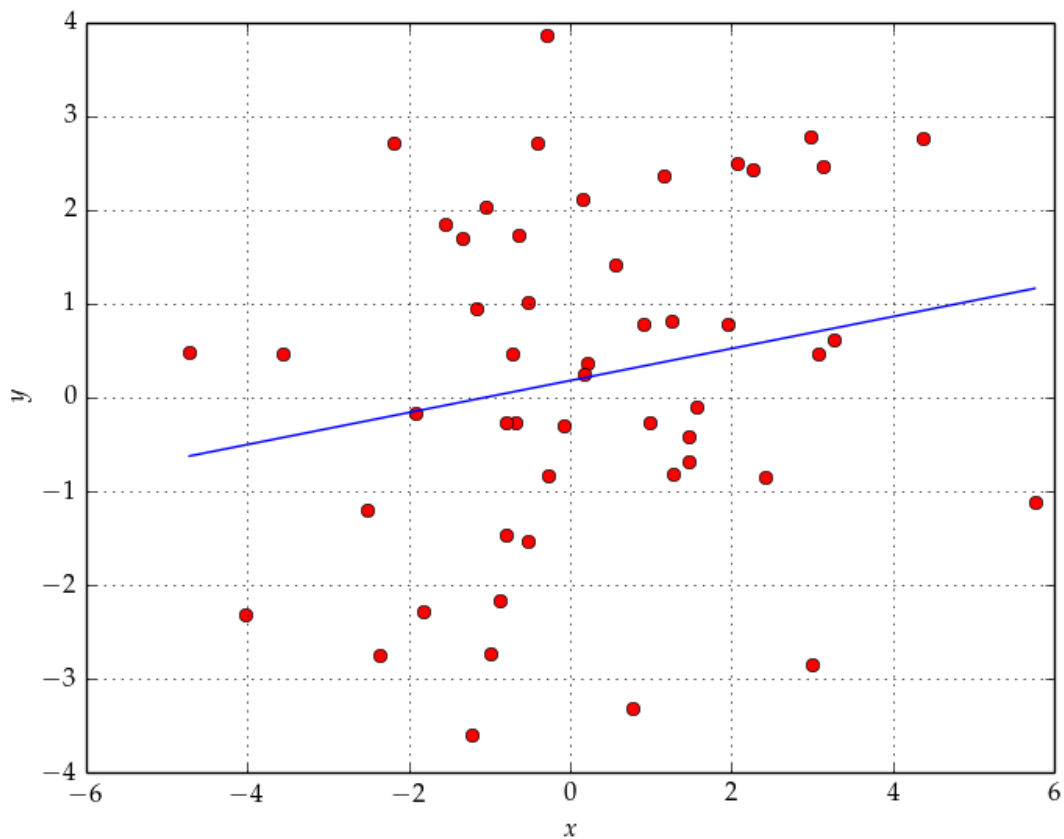


Рисунок 1 – График линии регрессии для двумерной случайной величины