

Домашняя самостоятельная работа №2 по математической статистике.

Даны две выборки, относящиеся к наблюдениям за одной и той же двумерной случайной величиной.

Объем первой выборки составляет $n_1 = 15$ значений, выборочное среднее равно $(\bar{x}_1; \bar{y}_1) = (4.2; -1.2)$, выборочная дисперсия $(\tilde{s}_{\bar{X}_1}^2; \tilde{s}_{\bar{Y}_1}^2) = (6.2; 3.2)$, выборочная ковариация $\tilde{K}_{XY}^{(1)} = -2.4$.

Вторая выборка известна полностью:

$(1; 0), (3; 1), (5; -2), (2; -1), (4; 0), (7; -3), (2; 0), (3; -1), (1; -1), (5; 0)$.

Найти выборочный коэффициент корреляции по двум выборкам.

Решение.

1. Найдем сначала выборочные числовые характеристики для второй выборки.

Её объем $n_2 = 10$.

Выборочные средние:

$$a. \bar{x}_2 = \frac{1+3+5+2+4+7+2+3+1+5}{10} = 3.3.$$

$$\bar{y}_2 = \frac{0+1-2-1+0-3+0-1-1+0}{10} = -0.7.$$

Выборочные средние квадраты:

$$a. \overline{x_2^2} = \frac{1^2+3^2+5^2+2^2+4^2+7^2+2^2+3^2+1^2+5^2}{10} = 14.3$$

$$\tilde{s}_{\bar{X}_2}^2 = \overline{x_2^2} - (\bar{x}_2)^2 = 14.3 - 3.3^2 = 3.41$$

Выборочные дисперсии:

$$a. \overline{y_2^2} = \frac{0^2+1^2+(-2)^2+(-1)^2+0^2+(-3)^2+0^2+(-1)^2+(-1)^2+0^2}{10} = 1.7$$

$$\tilde{s}_{\bar{Y}_2}^2 = \overline{y_2^2} - (\bar{y}_2)^2 = 1.7 - (-0.7)^2 = 1.21$$

Выборочное среднее произведения компонент:

$$\langle x_2 y_2 \rangle = \frac{1 \cdot 0 + 3 \cdot 1 + 5 \cdot (-2) + 2 \cdot (-1) + 4 \cdot 0 + 7 \cdot (-3) + 2 \cdot 0 + 3 \cdot (-1) + 1 \cdot (-1) + 5 \cdot 0}{10} = -3.$$

Произведение выборочных средних:

$$a. \langle x_2 \rangle \langle y_2 \rangle = 3.3 \cdot (-0.7) = -2.31$$

Выборочная ковариация:

$$a. \tilde{K}_{XY}^{(2)} = \langle x_2 y_2 \rangle - \langle x_2 \rangle \langle y_2 \rangle = -3.4 - (-2.31) = -1.09$$

2. Теперь найдем выборочные числовые характеристики объединенной выборки:

a. Выборочные средние:

$$\langle \bar{x}_i \rangle = \frac{\sum_{i=1}^k n_i \bar{x}_i}{\sum_{i=1}^k n_i} = \frac{15 \cdot 4.2 + 10 \cdot 3.3}{15 + 10} = 3.84$$

$$\langle \bar{y}_i \rangle = \frac{\sum_{i=1}^k n_i \bar{y}_i}{\sum_{i=1}^k n_i} = \frac{15 \cdot (-1.2) + 10 \cdot (-0.7)}{15 + 10} = -1.0$$

b. Средние арифметические групповых дисперсий:

$$\overline{\tilde{s}_{\bar{X}}^2} = \frac{\sum_{i=1}^k \tilde{s}_{\bar{X}_i}^2 n_i}{\sum_{i=1}^k n_i} = \frac{6.2 \cdot 15 + 3.41 \cdot 10}{15 + 10} = \frac{93 + 34.1}{25} = \frac{127.1}{25} = 5.084$$

$$\overline{\tilde{S}_Y^2} = \frac{\sum_{i=1}^k \tilde{S}_{Y_i}^2 n_i}{\sum_{i=1}^k n_i} = \frac{3.2 \cdot 15 + 1.21 \cdot 10}{15 + 10} = 2.404$$

с. Средняя арифметическая групповых ковариаций:

$$\langle \tilde{K}_{XY}^{(i)} \rangle = \frac{\sum_{i=1}^k n_i \tilde{K}_{XY}^{(i)}}{\sum_{i=1}^k n_i} = \frac{15 \cdot (-2.4) + 10 \cdot (-1.09)}{15 + 10} = -1.876$$

d. Теперь найдем межгрупповые числовые характеристики:

Межгрупповая дисперсия:

$$\begin{aligned} \text{i. } \delta_X^2 &= \frac{\sum_{i=1}^k (\bar{x}_i - \bar{x})^2 n_i}{\sum_{i=1}^k n_i} = \frac{(4.2 - 3.84)^2 \cdot 15 + (3.3 - 3.84)^2 \cdot 10}{15 + 10} = \\ &= \frac{0.36^2 \cdot 15 + 0.54^2 \cdot 10}{25} = 0.1944 \end{aligned}$$

$$\begin{aligned} \delta_Y^2 &= \frac{\sum_{i=1}^k (\bar{y}_i - \bar{y})^2 n_i}{\sum_{i=1}^k n_i} = \frac{((-1.2) - (-1.0))^2 \cdot 15 + ((-0.7) - (-1.0))^2 \cdot 10}{15 + 10} = \\ &= \frac{(-0.2)^2 \cdot 15 + (0.3)^2 \cdot 10}{15 + 10} = 0.06 \end{aligned}$$

Межгрупповая ковариация:

i. среднее произведение групповых выборочных средних:

$$\langle \bar{x}_i \bar{y}_i \rangle = \frac{\sum_{i=1}^k n_i \bar{x}_i \bar{y}_i}{\sum_{i=1}^k n_i} = \frac{15 \cdot 4.2 \cdot (-1.2) + 10 \cdot 3.3 \cdot (-0.7)}{15 + 10} = -3.948$$

произведение групповых выборочных средних:

$$\langle \bar{x}_i \rangle \cdot \langle \bar{y}_i \rangle = 3.84 \cdot (-1.0) = -3.84$$

межгрупповая ковариация:

$$\langle \tilde{K}_{\bar{x}_i \bar{y}_i} \rangle = \langle \bar{x}_i \bar{y}_i \rangle - \langle \bar{x}_i \rangle \langle \bar{y}_i \rangle = -3.948 - (-3.84) = -0.108$$

e. Теперь найдем выборочные числовые характеристики объединенной выборки:

i. Общая дисперсия \tilde{S}^2 равна сумме средней арифметической $\overline{\tilde{S}_i^2}$ групповых дисперсий и межгрупповой дисперсии δ^2 , т.е.

$$\tilde{S}_X^2 = \overline{\tilde{S}_{X_i}^2} + \delta_X^2 = 5.084 + 0.1944 = 5.2784$$

$$\tilde{S}_Y^2 = \overline{\tilde{S}_{Y_i}^2} + \delta_Y^2 = 2.404 + 0.06 = 2.464.$$

Ковариация:

$$\tilde{K}_{XY} = \langle \tilde{K}_{XY}^{(i)} \rangle + \langle \tilde{K}_{\bar{x}_i \bar{y}_i} \rangle = -1.876 + (-0.108) = -1.984$$

f. Выборочный коэффициент корреляции

$$\text{i. } \tilde{r}_{xy} = \frac{\tilde{K}_{XY}}{\tilde{S}_X \tilde{S}_Y} = \frac{-1.984}{\sqrt{5.2784} \sqrt{2.464}} = \frac{-1.984}{2.29747687692 \cdot 1.56971334963} = -0.550136128664$$