Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет Информационных технологий и программирования

Типовой расчет по математической статистике

«Ковариация и регрессия. Построение выборочного уравнения

линии регрессии»

Вариант 4

Васильков Дмитрий Алексеевич

M3215

Санкт-Петербург

2024 г.

Исходные данные:

y_j^* / x_i^*	21	31	36	41
16	0	0	30	80
26	0	55	20	0
36	15	0	0	0

Ход работы:

Приступим к выполнению задания.

Сначала, при обработке нашей выборки предварительно проведем группировку значений X и Y подобно тому, как мы это делали в ТР1. При этом, для частичных интервалов $\Delta x_i = [x_{i-1}, x_i, i = 1, \dots, k]$ и $\Delta y_j = [y_{j-1}, y_j, j = 1, \dots, m]$ определим число элементов выборки n_{ij}, попавших в прямоугольник $\Delta x_i \times \Delta y_j$, и вычислим середины интервалов по формулам $x_i^* = \frac{x_{i-1}+x_i}{2}$, $y_j^* = \frac{y_{j-1}+j}{2}$. Все элементы выборки, попавшие в прямоугольник $\Delta x_i \times \Delta y_j$, будем считать равными (x_i^*, y_j^*), причем количество значений x_i^* будет равно $n_i = \sum_{j=1}^{m} n_{ij}$, а количество значений y_j^* будет равно $n_j = \sum_{i=1}^{k} n_{ij}$.

Объем выборки равен $n_i = \sum_{i=1}^{k} n_i = \sum_{j=1}^{m} n_j = \sum_{i=1}^{k} \sum_{j=1}^{m} n_{ij}$.

Занесем эти данные в таблицу:

y_j^* / x_i^*	21	31	36	41	$\boldsymbol{n_i}$
16	0	0	30	80	110
26	0	55	20	0	75
36	15	0	0	0	15
$\boldsymbol{n_j}$	15	55	50	80	**n=200**

Теперь, выполним построение выборочного уравнения линии линейной регрессии по нашей таблице группированных данных.

Формула выборочного уравнения линии регрессии:

$$y(x)=\overline{Y}_n+r_n(X,Y)\frac{\sigma_n(Y)}{\sigma_n(X)}(x-\overline{X}_n). \qquad (4)$$

Для расчета коэффициентов в выборочном уравнении линии регрессии (4) используют формулы:

$$\overline{X}_n=\frac{1}{n}\sum_{i=1}^{k}n_i x_i^*, \qquad \overline{Y}_n=\frac{1}{n}\sum_{j=1}^{m}n_j y_j^*, \qquad (5)$$

$$\sigma_n(X)=\sqrt{\frac{1}{n}\sum_{i=1}^{k}n_i(x_i^*)^2-\overline{X}_n^2}, \qquad \sigma_n(Y)=\sqrt{\frac{1}{n}\sum_{j=1}^{m}n_j(y_j^*)^2-\overline{Y}_n^2}, \qquad (6)$$

$$r_n(X,Y)=\frac{\sum_{i=1}^{k}\sum_{j=1}^{m}n_{ij}x_i^*y_j^*-n\cdot\overline{X}_n\cdot\overline{Y}_n}{n\cdot\sigma_n(X)\cdot\sigma_n(Y)}. \qquad (7)$$

По формулам (5) находим:

$\overline{X_{200}} = 21{,}25$

$\overline{Y_{200}} = 35{,}5$

По формулам (6) находим:

$\sigma_{200}(X) = 6{,}320$

$\sigma_{200}(Y) = 5{,}788$

По формуле (7) находим:

$r_{200}(X,Y) = -0{,}9193$

Подставив найденные величины в формулу (4), получим искомое выборочное уравнение линейной регрессии Y на X:

$$y = 35{,}5 - 0{,}9193\cdot\frac{5{,}788}{6{,}320}(x-21{,}25)$$

Или окончательно:

$y = 53{,}3907 - 0{,}841916\,x$ (8)

Сравним оценки условных математических ожиданий, вычисленные:

а) на основе последнего уравнения

б) по данным таблицы 7, полагая, как и ранее, $P(y_j^*) = p_j^* = n_{ij}/n_i$

Например, возьмем $x_i^* = 16$:

а) $E(Y|X = 16) = 53{,}3907 - 0{,}841916 * 16 = 39{,}920$

б) $E(Y|X = 16) = (30 * 36 + 80 * 41) / 110 = 39{,}636$

Как видно, соответствие удовлетворительное.

Заметим, что уравнения линейной регрессии (3) и выборочной линейной регрессии (4), (8) являются уравнениями, задающими прямую линию.

На этом выполнение типового расчета №2 **закончено**.