

Тема: Линейная регрессия по корреляционной таблице

ЗАДАНИЕ. Найти выборочное уравнение прямой $y = r \frac{\delta_y}{\delta_x} (x - \bar{x}) + \bar{y}$ регрессии Y на X по заданной корреляционной таблице

Y/X	10	15	20	25	30	35	n_y
30	2	6					8
40		4	4				8
50			7	35	8		50
60			2	10	8		20
70				5	6	3	14
n_x	2	10	13	50	22	3	n=100

РЕШЕНИЕ. Построим ряды распределений для X и Y, вычислим их характеристики (выборочное среднее и выборочное среднее квадратическое отклонение).

x_i	n_i	$x_i \cdot n_i$	$(x_i - \bar{x})^2 \cdot n_i$
10	2	20	417,605
15	10	150	893,025
20	13	260	257,4325
25	50	1250	15,125
30	22	660	677,655
35	3	105	333,9075

Сумма 100 2445 2594,75

Выборочная средняя $\bar{x} = \frac{1}{n} \sum x_i n_i = \frac{1}{100} 2445 = 24,45$

Выборочная дисперсия $\bar{D}_x = \frac{1}{n} \sum (x_i - \bar{x})^2 n_i = \frac{1}{100} 2594,75 = 25,9475$

Выборочное квадратическое отклонение $\sigma_x = \sqrt{\bar{D}_x} \approx 5,09$

y_i	n_i	$y_i \cdot n_i$	$(y_i - \bar{y})^2 \cdot n_i$
30	8	240	4014,08
40	8	320	1230,08
50	50	2500	288
60	20	1200	1155,2
70	14	980	4336,64

Сумма 100 5240 11024

Выборочная средняя $\bar{y} = \frac{1}{n} \sum y_i n_i = \frac{1}{100} 5240 = 52,4$

Выборочная дисперсия $\bar{D}_y = \frac{1}{n} \sum (y_i - \bar{y})^2 n_i = \frac{1}{100} 11024 = 110,24$

Выборочное квадратическое отклонение $\sigma_y = \sqrt{\bar{D}_y} \approx 10,5$

Коэффициент корреляции вычислим по формуле $r_a = \frac{\sum n_{xy} x_i y_i - n \bar{x} \bar{y}}{n \sigma_x \sigma_y}$.

Найдем сумму $\sum n_{xy} x_i y_i = 132150$. Тогда $r_a = \frac{132150 - 100 \cdot 24,45 \cdot 52,4}{100 \cdot 5,09 \cdot 10,5} \approx 0,754$.

Уравнение регрессии Y на X имеет вид $\bar{y}_x = r_a \frac{\sigma_y}{\sigma_x} (x - \bar{x}) + \bar{y}$. Подставляем все величины:

$$\bar{y}_x = 0,754 \frac{10,5}{5,09} (x - 24,45) + 52,4$$

$$\bar{y}_x = 1,555x + 14,37$$

Построим корреляционное поле и линию регрессии.

