

### Тема: Линейная регрессия, коэффициент корреляции

ЗАДАНИЕ. С целью анализа взаимного влияния зарплаты и текучести рабочей силы на пяти однотипных фирмах с одинаковым числом работников проведены измерения уровня месячной зарплаты  $X$  и числа уволившихся за год рабочих  $Y$ :

X	100	150	200	250	300
Y	60	35	20	20	15

Найти линейную регрессию  $Y$  на  $X$ , выборочный коэффициент корреляции.

РЕШЕНИЕ. Сначала найдем характеристики случайных величин  $X$  и  $Y$  (выборочное среднее и выборочное среднее квадратическое отклонение).

						<b>Сумма</b>
$x_i$	100	150	200	250	300	<b>1000</b>
$(x_i - \bar{x})^2$	10000	2500	0	2500	10000	<b>25000</b>

$$\text{Выборочная средняя } \bar{x} = \frac{1}{n} \sum x_i = \frac{1}{5} 1000 = 200$$

$$\text{Выборочная дисперсия } \bar{D}_x = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{5} 25000 = 5000$$

$$\text{Выборочное квадратическое отклонение } \sigma_x = \sqrt{\bar{D}_x} \approx 70,711$$

						<b>Сумма</b>
$y_i$	60	35	20	20	15	<b>150</b>
$(y_i - \bar{y})^2$	900	25	100	100	225	<b>1350</b>

$$\text{Выборочная средняя } \bar{y} = \frac{1}{n} \sum y_i = \frac{1}{5} 150 = 30$$

$$\text{Выборочная дисперсия } \bar{D}_y = \frac{1}{n} \sum (y_i - \bar{y})^2 = \frac{1}{5} 1350 = 270$$

$$\text{Выборочное квадратическое отклонение } \sigma_y = \sqrt{\bar{D}_y} \approx 16,432$$

Осталось подсчитать  $\sum x_i y_i = 24750$ . Подсчеты занесем в таблицу:

$x_i$	100	150	200	250	300	
$y_i$	60	35	20	20	15	
$x_i y_i$	6000	5250	4000	5000	4500	<b>Сумма</b>
						<b>24750</b>

Коэффициент корреляции вычислим по формуле

$$r_{\hat{a}} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n \sigma_x \sigma_y} = \frac{24750 - 5 \cdot 200 \cdot 30}{5 \cdot 70,711 \cdot 16,432} \approx -0,904.$$

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

$$\bar{y}_x - 30 = -0,904 \frac{16,432}{70,711} (x - 200)$$

$$\bar{y}_x = -0,21x + 72,015$$