

Тема: Линейная регрессия, коэффициенты корреляции и детерминации.

Задача.

Постройте линии регрессии Y на x и X на y для двумерной с.в. (X, Y) , закон распределения которой задан таблицей, рассчитайте коэффициенты корреляции и детерминации.

X / Y	-1	0	1	Σ
0	0,1	0,15	0,2	0,45
1	0,15	0,25	0,15	0,55
Σ	0,25	0,4	0,35	1

Решение:

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

x_i	n_i	$x_i \cdot n_i$	$(x_i - \bar{x})^2 \cdot n_i$
0	0,45	0	0,14
1	0,55	0,55	0,11
Сумма	1	0,55	0,2475

$$\text{Выборочная средняя } \bar{x} = \frac{1}{n} \sum x_i n_i = 0,55$$

$$\text{Выборочная дисперсия } \bar{D}_x = \frac{1}{n} \sum (x_i - \bar{x})^2 n_i = 0,2475$$

$$\text{Выборочное квадратическое отклонение } \sigma_x = \sqrt{\bar{D}_x} = \sqrt{0,2475} = 0,497$$

y_i	n_i	$y_i \cdot n_i$	$(y_i - \bar{y})^2 \cdot n_i$
-1	0,25	-0,25	0,303
0	0,4	0	0,004
1	0,35	0,35	0,284
Сумма	1	0,1	0,59

Выборочная средняя $\bar{y} = \frac{1}{n} \sum y_i n_i = 0,1$

Выборочная дисперсия $\bar{D}_y = \frac{1}{n} \sum (y_i - \bar{y})^2 n_i = 0,59$

Выборочное квадратическое отклонение $\sigma_y = \sqrt{\bar{D}_y} = \sqrt{0,59} = 0,768$

Найдем сумму $\sum n_{xy} x_i y_i = 0$.

X / Y	-1	0	1	$y_j n_{ij}$	$x_i y_j n_{ij}$
0	0,1	0,15	0,2	0,1	0
1	0,15	0,25	0,15	0	0
				Сумма	0

Коэффициент корреляции вычислим по формуле

$$r = \frac{\sum n_{xy} x_i y_i - n \bar{x} \bar{y}}{n \sigma_x \sigma_y} = \frac{0 - 1 * 0,55 * 0,1}{1 * 0,497 * 0,768} = -0,144$$

Связь слабая, отрицательная по направлению.

Коэффициент детерминации для линейной связи равен квадрату коэффициента корреляции:

$$R^2 = (-0,144)^2 = 0,021$$

Следовательно, только 2,1% вариации Y объясняется вариацией X и наоборот.

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

величины:

$$\bar{Y}_x = (-0,144) * \frac{0,768}{0,497} * (x - 0,55) + 0,1$$

$$\bar{Y}_x = -0,222x + 0,222$$

Уравнение регрессии X на y имеет вид $\bar{X}_y = r \frac{\sigma_x}{\sigma_y} (y - \bar{y}) + \bar{x}$. Подставляем все

величины:

$$\bar{X}_y = (-0,144) * \frac{0,497}{0,768} * (y - 0,1) + 0,55$$

$$\bar{X}_y = -0,093y + 0,559$$