# Регресія

# Формули

## Функція регресії I роду

Функцією регресії I роду $g(x)$ визначаємо, приписуючи значенням незалежної ознаки (найчастіше *X*) умовне середне залежного ознаки (найчастіше *Y*).

$$g(x_i) = E(Y \mid X = x_i)$$

Емпірічну лінію регресії рисуємо на графіку поєднуючи відповідні точки відрізками.

## Функція регресії II роду

Якщо $Y$ залежить від $X$ , визначення **функції регресії II роду** має вигляд:

$\hat{y}_i = a_0 + a_1 x_i$ ,

де:

$$a_1 = \frac{\sum\left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\sum\left(x_i - \overline{X}\right)^2}$$

$$a_0 = \overline{Y} - a_1 \overline{X}$$

Якщо $X$ залежить від $Y$ , визначенння **функції регресії II роду** має вигляд:

$\hat{x}_i = b_0 + b_1 y_i$ ,

де:

$$b_1 = \frac{\sum\left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\sum\left(y_i - \overline{Y}\right)^2}$$

$$b_0 = \overline{X} - b_1 \overline{Y}$$

**Уваги**

Необхідно пам’ятати, що суми у формулах означають суму всіх одиниць вибірки і в деяких завданнях може знадобитися додатково помножити на числа $n_i$ .

**Коефіцієнт лінійної кореляції Пірсона**: $r_{xy} = \pm\sqrt{a_1 \cdot b_1}$ , знак перед квадратом завжди збігається з коефіцієнтами $a_1$ та $b_1$ .

Додатково коефіцієнти $a_1$ та $b_1$ можна знайти за формулами:

$$a_1 = r_{xy} \frac{S(Y)}{S(X)} \qquad b_1 = r_{xy} \frac{S(X)}{S(Y)}$$

## Перевірка точності відповідності оціненої функції регресії

### Залишкова дисперсія

$S^2(u_i) = \frac{\sum (y_i - \hat{y}_i)^2}{n-k}$ , якщо $Y$ залежить від $X$

$S^2(z_i) = \frac{\sum (x_i - \hat{x}_i)^2}{n-k}$ , якщо $X$ залежить від $Y$

, де *k* визначає кількість параметрів в функції регресії (в разі прямої $k = 2$)

### Коефіцієнт збіжності $\varphi^2$

$\varphi_{yx}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum \left(y_i - \overline{Y}\right)^2}$ , якщо $Y$ залежить від $X$

$\varphi_{xy}^2 = \frac{\sum (x_i - \hat{x}_i)^2}{\sum \left(x_i - \overline{X}\right)^2}$ , якщо $Y$ залежить від $X$

Цей коефіцієнт показує, в скількох відсотках зміна вартості однієї змінної не буде визначена регресією.

Відбувається:

$\varphi_{xy}^2 + r_{xy}^2 = 1$

## Кореляція багатьох змінних

### Часткова кореляція

Перевіряємо взаємну кореляцію двох ознак, нехтуючи впливом інших.

**Коефіцієнт часткової кореляції:**

$r_{ij.abzc...} = \dfrac{-P_{ij}}{\sqrt{P_{ii} \cdot P_{jj}}}$, де:

$r_{ij.abzc...}$ - коефіцієнт часткової кореляції між об'єктами з ідексами $i$ та $j$, пропускаючи об'єкти з індексами $a, b, z, c, \dots$.

$P_{ij}, P_{ii}, P_{jj}$ - відповідне алгебраїчне доповнення до матриць коефіцієнтов кореляції лінійної Пірсона: $P = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 \end{bmatrix}$.

### Множинна кореляція

Перевіряємо кореляцію обраної характеристики з усіма інними разом**:**

$R_w = R_{i.abc...} = \sqrt{1 - \dfrac{|P|}{|R|}}$, де:

$R_{i.abc...}$ - коефіцієнт множинної кореляції характеристики з індексом $i$ та всіх інших характеристик

$|P|$ - визначник матриці лінійних коефіцієнтів кореляції Пірсона $P = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 \end{bmatrix}$

$|R|$ - визначник матриці лінійних коефіцієнтів кореляції Пірсона, з якої видалено рядок із номером $i$ та стовпчик з номером $i$.

## Регресія для трьох змінних

Якщо $Y$ залежить від двох змінних $X_1$ та $X_2$, визначення **функції регресії II роду** має вигляд:

$$\hat{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i}.$$

де:

$$a_1 = \frac{S_1}{S_2} \cdot \frac{r_{12} - r_{13} \cdot r_{23}}{1 - r_{23}^2}$$

$$a_2 = \frac{S_1}{S_3} \cdot \frac{r_{13} - r_{12} \cdot r_{23}}{1 - r_{23}^2}$$

$$a_0 = \overline{Y} - a_1 \overline{X}_1 - a_2 \overline{X}_2$$

Індекс $1$ має коефіцієнт для змінної $Y$, індекс $2$ коефіцієнт для змінної $X_1$, а індекс $3$ коефіцієнт для змінної $X_2$.

**Коефіцієнт множинної кореляції:**

$$R_W = R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$$

**Дисперсія залишкова:**

$$S^2(u) = S_1^2\left(1 - R_W^2\right)$$