

Семинар 6

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{— парная регрессия}$$

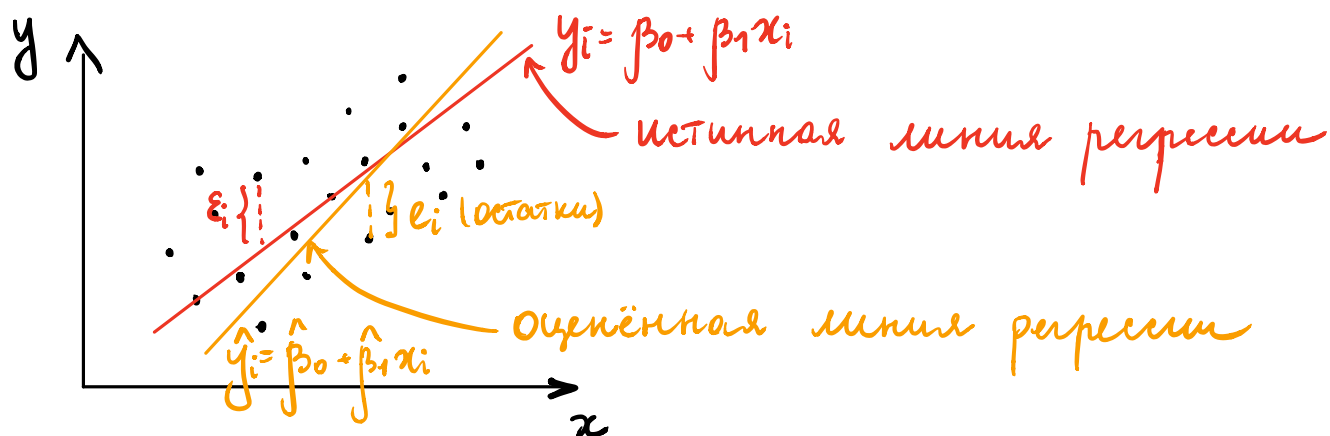
$$y = X\beta + \varepsilon \quad \text{— множественная регрессия}$$

$n \times 1 \quad n \times k \quad k \times 1 \quad n \times 1$

Важно: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ — истинная модель

$$y_i = \beta_0 + \beta_1 x_i \quad \text{— линия истинной регрессии}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad \text{— оценённая линия регрессии}$$



Пример 1

Модель: $y_i = \beta x_i + u_i$

Найти $\hat{\beta}_{\text{МНК}}$

x_i	y_i
1	1
2	2
2	4

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min_{\hat{\beta}}$$

$$-2 \sum_{i=1}^n x_i (y_i - \hat{\beta} x_i) = 0$$

$$\hat{\beta}_{\text{МНК}} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{1+4+8}{1+4+4} = \frac{13}{9}.$$

Для $y_i = \beta_0 + \beta_1 x_i + u_i$
 $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
 $\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$

Пример 2

Модель Гаусса
 Гаусса $y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$
 Параметры $\hat{\beta}_{1\text{МНК}}, \dots, \hat{\beta}_{k\text{МНК}}$

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min_{\hat{\beta}_1, \dots, \hat{\beta}_k}$$

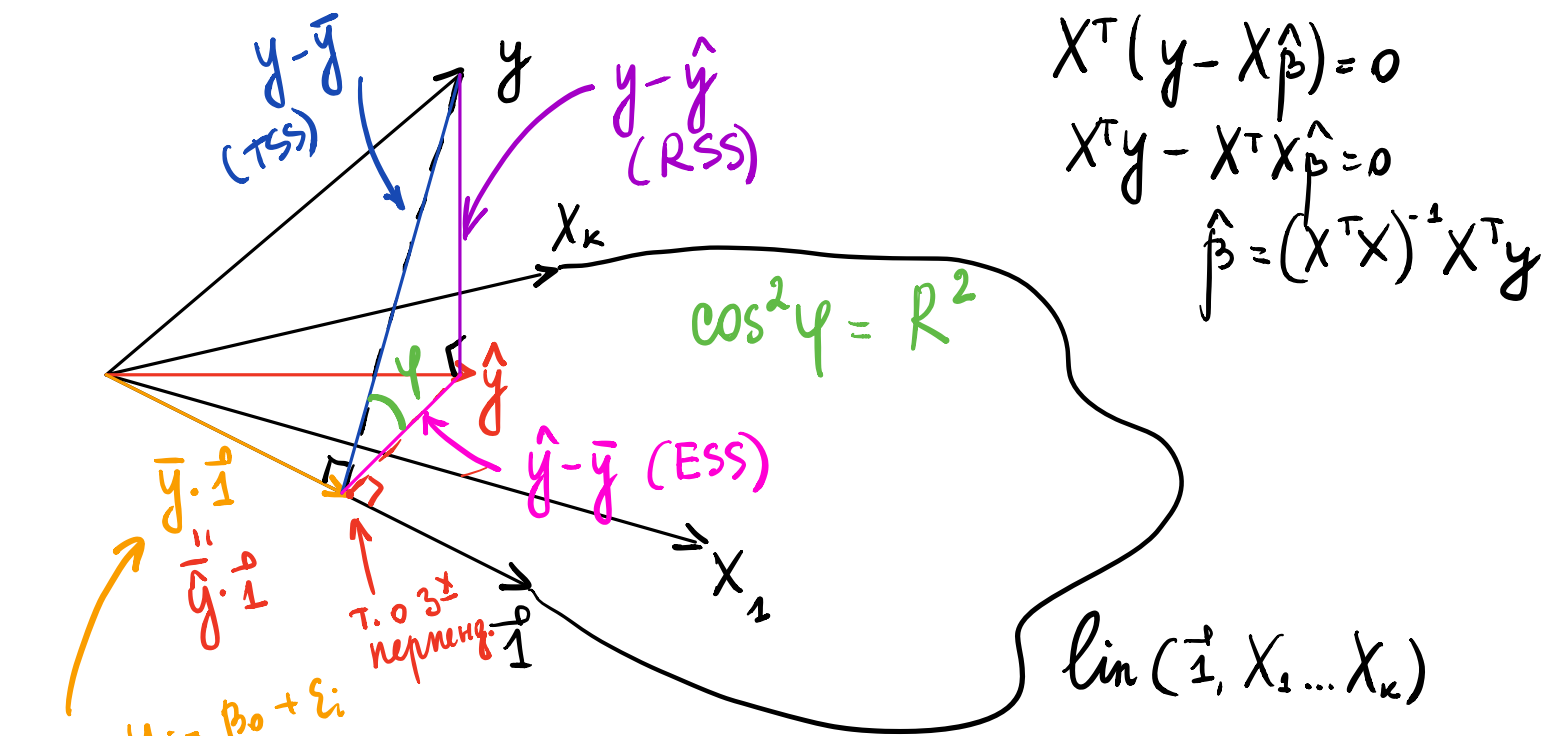
$$X = \begin{bmatrix} 1 & | & | & | & | \\ \vdots & x_1 & x_2 & \dots & x_k \\ 1 & | & | & | & | \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$y = X\beta + \varepsilon$$

$$(y - \hat{y})^T (y - \hat{y}) \rightarrow \min_{\hat{\beta}}$$

геометрия



$$\begin{aligned}
 X^T(y - X\hat{\beta}) &= 0 \\
 X^T y - X^T X \hat{\beta} &= 0 \\
 \hat{\beta} &= (X^T X)^{-1} X^T y
 \end{aligned}$$

$y: y_i = \beta_0 + \varepsilon_i$
 $\hat{y} = \hat{\beta}_0$
 $\sum_{i=1}^n (y_i - \hat{\beta}_0)^2 \rightarrow \min_{\hat{\beta}_0}$
 $-2 \sum_{i=1}^n (y_i - \hat{\beta}_0) = 0$
 $\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}$

По т. Пирсона:

$$\begin{aligned}
 \|y - \bar{y}\|^2 &= \|\hat{y} - \bar{y}\|^2 + \|y - \hat{y}\|^2 \\
 \underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{\text{TSS}} &= \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{\text{ESS}} + \underbrace{\sum_{i=1}^n (y_i - \hat{y})^2}_{\text{RSS}} \\
 &\quad \begin{array}{l} \text{total sum} \\ \text{of squares} \end{array} \quad \begin{array}{l} \text{estimated} \\ \text{sum of} \\ \text{squares} \end{array} \quad \begin{array}{l} \text{residual} \\ \text{sum of} \\ \text{squares} \end{array}
 \end{aligned}$$

$$\begin{aligned}
 R^2 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{ESS}}{\text{TSS}} = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = \\
 &\quad \frac{\widehat{\text{Var}}(\hat{y})}{\widehat{\text{Var}}(y)} = 1 - \frac{\text{RSS}}{\text{TSS}}
 \end{aligned}$$

$$R^2 = \overset{\text{коэфф. корреляции}}{\text{scorr}^2(y, \hat{y})}$$

$$\text{scorr}(y, \hat{y}) = \frac{\langle y - \bar{y}, \hat{y} - \bar{y} \rangle}{\|y - \bar{y}\| \|\hat{y} - \bar{y}\|} = \frac{\langle \hat{y} + e - \bar{y}, \hat{y} - \bar{y} \rangle}{\|y - \bar{y}\| \|\hat{y} - \bar{y}\|} =$$

$$= \frac{\|\hat{y} - \bar{y}\|^2 + \underbrace{\langle e, \hat{y} - \bar{y} \rangle}_{=0 \text{ (см. картинку)}}}{\|y - \bar{y}\| \|\hat{y} - \bar{y}\|} = \frac{\|\hat{y} - \bar{y}\|^2}{\|y - \bar{y}\| \|\hat{y} - \bar{y}\|} = \frac{\|\hat{y} - \bar{y}\|}{\|y - \bar{y}\|} = \sqrt{\frac{\text{ESS}}{\text{TSS}}}$$