Семинар 6

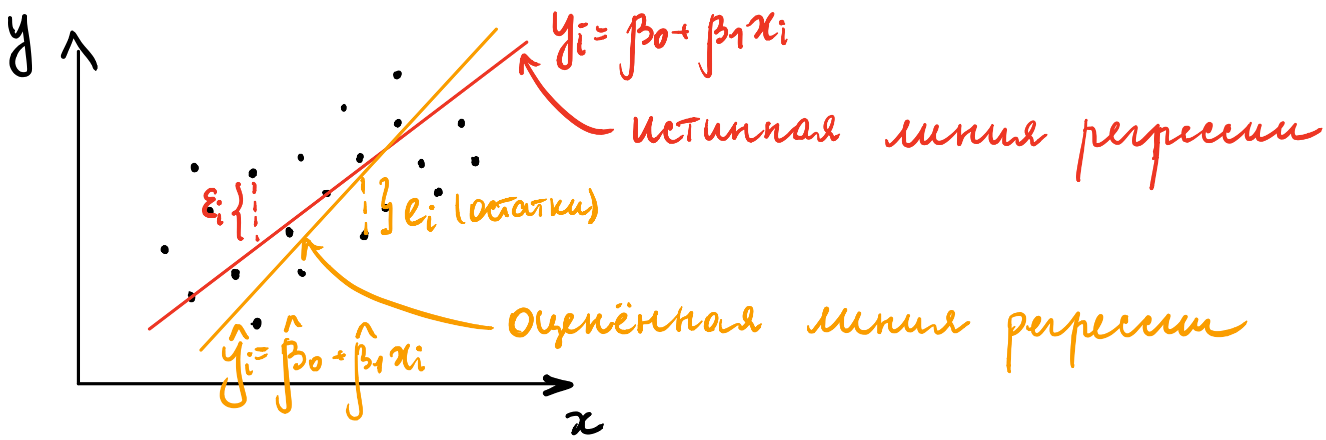
$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ – парная регрессия

$\underset{n\times1}{y} = \underset{n\times k}{X}\underset{k\times1}{\beta} + \underset{n\times1}{\varepsilon}$ – множественная регрессия

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

$y_i = \beta_0 + \beta_1 x_i$ – линия истинной регрессии

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



Пример 1

Модель: $y_i = \beta x_i + u_i$

Найти $\hat{\beta}_{МНК}$

x_i	y_i
1	1
2	2
2	4

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

Для $y_i=\beta_0+\beta_1 x_i+u_i$

$$\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

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

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

Пример 2

Модель $y_i=\beta_1+\beta_2 x_{1i}+\beta_2 x_{2i}+\dots+\beta_3 x_{ki}+\varepsilon_i$

Найти $\hat{\beta}_{1МНК},\dots,\hat{\beta}_{kМНК}$

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

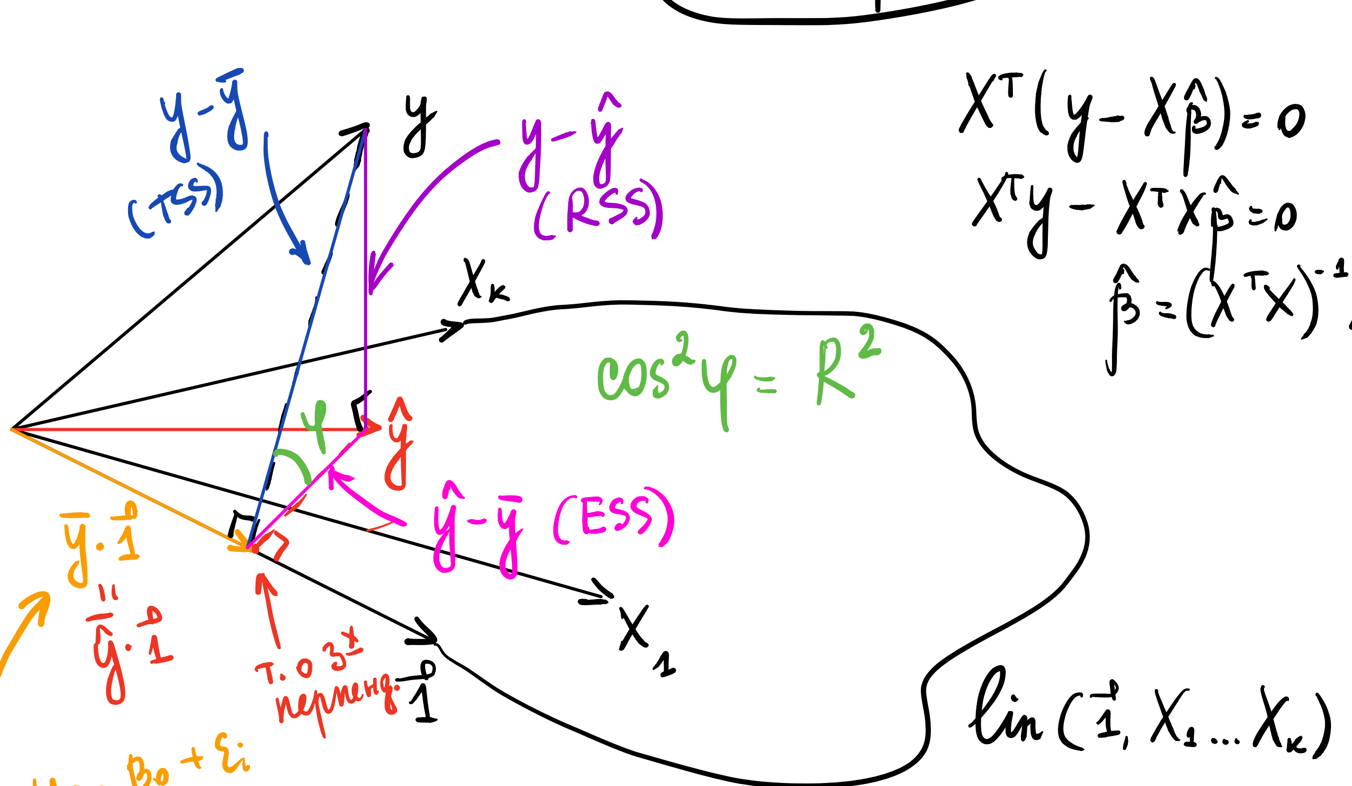
$$X=\begin{bmatrix}1 & | & | & & | \\ \vdots & x_1 & x_2 & \dots & x_k \\ 1 & | & | & & |\end{bmatrix}$$

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

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

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

геометрия



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

$y: \; y_i = \beta_0 + \varepsilon_i$

$\hat{y} = \hat{\beta}_0$

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0)^2 \to \min_{\hat{\beta}_0}$$

$$-2\sum_{i=1}^{n} (y_i - \hat{\beta}_0) = 0$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$$

По т. Пифагора:

$$\|y-\bar{y}\|^2 = \|\hat{y}-\bar{y}\|^2 + \|y-\hat{y}\|^2$$

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{\substack{\text{TSS} \\ \text{total sum of squares}}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{y}_i)^2}_{\substack{\text{ESS} \\ \text{estimated sum of squares}}} + \underbrace{\sum_{i=1}^{n}(y_i-\hat{y})^2}_{\substack{\text{RSS} \\ \text{residual sum of squares}}}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$\shortparallel$$

$$\frac{\widehat{Var}(\hat{y})}{\widehat{Var}(y)}$$

выбор. корреляция

$$R^2 = \text{scorr}^2(y, \hat{y})$$

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

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