

# Оценки параметров регрессии

## Линейная регрессия

# Вид модели. Парная регрессия

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

$y$  — отклик

$x$  — фактор

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, i = 1, 2, \dots, N$$

$\theta_0, \theta_1$  — ошибка наблюдения (случайная величина)

$\varepsilon_i$  — неизвестные параметры

# Вид модели. Множественная регрессия

Выборка  $(x_{11}, x_{12}, \dots, x_{1k}, y_1), (x_{21}, x_{22}, \dots, x_{2k}, y_2), \dots, (x_{N1}, x_{N2}, \dots, x_{Nk}, y_N)$

$$y_i = \theta_0 + \theta_1 x_{i1} + \dots + \theta_k x_{ik} + \varepsilon_i, i = 1, 2, \dots, N$$

# Предположения о модели

$$M[\varepsilon_i] = 0, i = 1, 2, \dots, N$$

$$D[\varepsilon_i] = \sigma^2, i = 1, 2, \dots, N$$

$$M[\varepsilon_i \varepsilon_j] = 0, i \neq j$$

$$\varepsilon_i \sim N(0, \sigma^2), i = 1, 2, \dots, N$$

# Метод наименьших квадратов

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n (y_i - f(x_i, \theta))^2$$

# Как вычислить оценки для простой парной регрессии методом наименьших квадратов?

$$F = \sum_{i=1}^N (y_i - (\theta_0 + \theta_1 x_i))^2 \rightarrow \min_{\theta_0, \theta_1}$$

$$\hat{\theta}_0 = \frac{1}{N} \sum_{i=1}^N y_i - \frac{1}{N} \sum_{i=1}^N x_i \hat{\theta}_1$$

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

# Как вычислить оценки для множественной регрессии методом наименьших квадратов?

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \Theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}, E = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}, X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \ddots & \vdots \\ \vdots & \vdots & & \\ 1 & x_{N1} & \dots & x_{Nk} \end{bmatrix}$$

$$Y = X\Theta + E$$

$$\hat{\Theta} = (X^T X)^{-1} X^T Y$$

$y$  — отклик

$x$  — фактор

$\Theta$  — вектор неизвестных параметров