# Множественный регрессионный анализ

**Обозначения:** $y$ - отклик $y=\varphi(x_1,x_2,...,x_k)$

$x_1,x_2,...,x_k$ - факторы - модель

$$y_{набл}=\varphi(x_1,x_2,...,x_k)+\varepsilon$$

**МНК** – метод наименьших квадратов

$$\sum(y_{набл}-y_{мод})^2\to\min$$

Требования:

1) $M(\varepsilon_i)=0$

2) $D(\varepsilon_i)=\sigma_\varepsilon{}^2$

3) $M(\varepsilon_i\varepsilon_j)=0$ $\quad i\neq j$

4) $\varepsilon_i\sim N(0,\sigma_\varepsilon)_1$ $\quad i=\overline{1,n}$

# Множественная линейная регрессия

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

- линейная модель

$y_i$ - $i$-ое наблюдаемое значение отклика

$x_{i1}, x_{i2}, \ldots, x_{ik}$ - $i$-ые наблюдаемые значения факторов

$$i = \overline{1, n}$$

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_k x_{ik} + \varepsilon_i$$

**МНК:** $$\sum \varepsilon_i^2 \to \min$$

# Множественная линейная регрессия

Модель: $y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$

$x_0 \equiv 1 \Rightarrow \quad y = b_0 x_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$

Наблюдения:

$$y_i = b_0 x_{i0} + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_k x_{ik} + \varepsilon_i \qquad i = \overline{1,n}$$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} x_{10} & x_{11} & \ldots & x_{1k} \\ x_{20} & x_{21} & \ldots & x_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ x_{n0} & x_{n1} & \ldots & x_{nk} \end{pmatrix} \quad b = \begin{pmatrix} b_0 \\ b_1 \\ \ldots \\ b_k \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}$$

$(n \times 1) \qquad (n \times (k+1)) \qquad ((k+1) \times 1) \qquad (n \times 1)$

# Множественная линейная регрессия

$$y = Xb + \varepsilon$$

**МНК:** $$\sum \varepsilon_i^2 \to \min$$

$$Q(b) = \sum \varepsilon_i^2 = \varepsilon^T \cdot \varepsilon =$$

$$\varepsilon^T = (\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_n)^T$$

$$\varepsilon = y - Xb$$

$$= (y - Xb)^T (y - Xb) =$$

$$= \left( y^T - b^T X^T \right)(y - Xb) =$$

$$= y^T y - b^T X^T y - y^T Xb + b^T X^T Xb$$

# Множественная линейная регрессия

$$y = Xb + \varepsilon$$

**МНК:** $\sum \varepsilon_i^2 \to \min$

$$Q(b) = \sum \varepsilon_i^2 = y^T y - b^T X^T y - y^T Xb + b^T X^T Xb$$

$$\nabla_b Q(b) = -X^T y - X^T y + X^T Xb + X^T Xb = 0$$

$$X^T Xb = X^T y$$

$$\hat{b} = \left( x^T x \right)^{-1} x^T y$$

$$\nabla_b c = \left( \frac{\partial c}{\partial b_0}, \frac{\partial c}{\partial b_1}, \ldots, \frac{\partial c}{\partial b_k} \right)^T$$

$$\nabla_b \left( b^T A \right) = A \qquad \nabla_b \left( Ab \right) = A^T$$

$$\nabla_b \left( b^T Ab \right) = Ab + A^T b$$

# Множественная линейная регрессия

$$X^T X b = X^T y$$ - система нормальных уравнений

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix} \cdot \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}$$

$$X^T X = \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1} x_{ik} \\ \dots & \dots & \dots & \dots \\ \sum x_{ik} & \sum x_{i1} x_{ik} & \dots & \sum x_{ik}^2 \end{pmatrix}$$

# Множественная линейная регрессия

$$X^T Xb = X^T y \qquad \hat{b} = \left( X^T X \right)^{-1} X^T y$$

$$X^T Xb = \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1} x_{ik} \\ \dots & \dots & \dots & \dots \\ \sum x_{ik} & \sum x_{i1} x_{ik} & \dots & \sum x_{ik}^2 \end{pmatrix} \cdot \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_k \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_{i1} y_i \\ \dots \\ \sum x_{ik} y_i \end{pmatrix}$$

# Множественная линейная регрессия

$$X^T X b = X^T y \qquad \hat{b} = \left( X^T X \right)^{-1} X^T y$$

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$\begin{cases} n b_0 + b_1 \sum x_{i1} + b_2 \sum x_{i2} = \sum y_i \\ b_0 \sum x_{i1} + b_1 \sum x_{i1}^2 + b_2 \sum x_{i1} x_{i2} = \sum x_{i1} y_i \\ b_0 \sum x_{i2} + b_1 \sum x_{i1} x_{i2} + b_2 \sum x_{i2}^2 = \sum x_{i2} y_i \end{cases}$$

# Интервальные оценки

$$y = Xb + \varepsilon$$

$$\hat{b} = \left( X^T X \right)^{-1} X^T y$$

Оценка остаточной дисперсии

$$S_\varepsilon^2 = \frac{Q\left(\hat{b}\right)}{n - \left(k+1\right)}$$

$$S_\varepsilon^2 = \frac{1}{n-\left(k+1\right)} \sum_{i=1}^{n} \left( y_i - \left( \hat{b}_0 + \hat{b}_1 x_{i1} + \ldots + \hat{b}_k x_{ik} \right) \right)^2$$

$$b_i = \hat{b}_i \pm t_{табл} \cdot s_{b_i} \qquad s_{b_i} = s_\varepsilon \cdot \sqrt{c_{ii}} \qquad i = \overline{1,n}$$

$$t_{табл} = t_{\left( \frac{\alpha}{2};\ n-(k+1) \right)} \qquad C = \left( X^T X \right)^{-1}$$

# Множественная линейная регрессия

## Адекватность модели

$$y = Xb + \varepsilon$$

Оценка остаточной дисперсии

$$S_\varepsilon^2 = \frac{Q(\hat{b})}{n-(k+1)}$$

$$S_\varepsilon^2 = \frac{1}{n-(k+1)}\sum_{i=1}^{n}\left(y_i - \left(\hat{b}_0 + \hat{b}_1 x_{i1} + \ldots + \hat{b}_k x_{ik}\right)\right)^2$$

Оценка общей дисперсии

$$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$H_0: \; D_y = D_\varepsilon$$
$$H_1: \; D_y > D_\varepsilon$$

$$\frac{S_y^2}{S_\varepsilon^2} \sim F(n-1; n-(k+1))$$

# Полиномиальная регрессия

$$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \ldots + b_k x^k$$

Замена: $x \to x_1 \quad x^2 \to x_2 \quad \ldots \quad x^k \to x_k$

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

$x_0 \equiv 1 \qquad y = b_0 x_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$

В матричной форме $y_{мод} = Xb \qquad y_{набл} = Xb + \varepsilon$

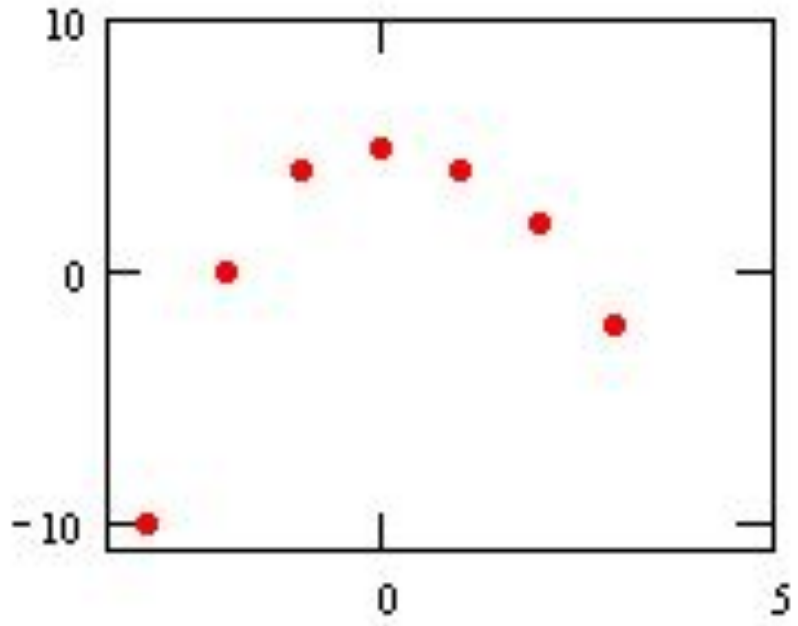
**мнк:** $\sum \varepsilon_i^2 \to \min$

$$\hat{b} = \left( X^T X \right)^{-1} X^T y$$

# Полиномиальная регрессия

Пример. Найти оценки параметров модели, проверить ее адекватность $y = b_0 + b_1 x + b_2 x^2$

| $x$ | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| $y$ | -10 | 0 | 4 | 5 | 4 | 2 | -2 |




$$x \to x_1 \qquad x^2 \to x_2$$

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$y = b_0 x_0 + b_1 x_1 + b_2 x_2$$

$$(x_0 \equiv 1)$$

Пример (продолжение).

| $x$ | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| $y$ | -10 | 0 | 4 | 5 | 4 | 2 | -2 |

Перейдем к матричной форме

$$y = b_0 x_0 + b_1 x_1 + b_2 x_2$$

$$y = Xb + \varepsilon$$

$$y = \begin{pmatrix} -10 \\ 0 \\ 4 \\ 5 \\ 4 \\ 2 \\ -2 \end{pmatrix} \quad X = \begin{pmatrix} 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}$$

$$\hat{b} = \left( X^T X \right)^{-1} X^T y$$

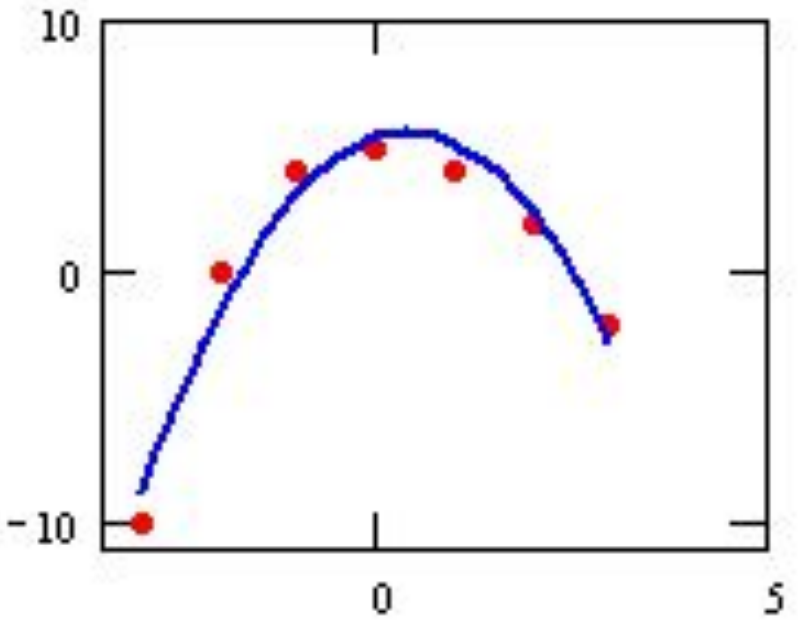
$$X^T X = \begin{pmatrix} 7 & 0 & 28 \\ 0 & 28 & 0 \\ 28 & 0 & 196 \end{pmatrix}$$

Пример
(продолжение).

$$X^T X = \begin{pmatrix} 7 & 0 & 28 \\ 0 & 28 & 0 \\ 28 & 0 & 196 \end{pmatrix} \qquad C = \left( X^T X \right)^{-1}$$

$$C = \begin{pmatrix} 0.33 & 0 & 0.05 \\ 0 & 0.04 & 0 \\ 0.05 & 0 & 0.01 \end{pmatrix}$$

$$\vec{b} = \left( X^T X \right)^{-1} X^T y$$




$$\vec{b} = \begin{pmatrix} 5.38 \\ 1 \\ -1.24 \end{pmatrix}$$

$$y = b_0 + b_1 x + b_2 x^2$$

$$y = 5.38 + x - 1.24x^2$$

Пример
(продолжение)

# Интервальные оценки

$$y = 5.38 + x - 1.24x^2$$

$$b_i = \hat{b}_i \pm t_{табл} \cdot s_{b_i}$$

$$s_{b_0} = 2.78 \cdot 1.3 \cdot \sqrt{0.33}$$

$$b_0 \in (3.29; 7.47)$$

$$s_{b_1} = 2.78 \cdot 1.3 \cdot \sqrt{0.04}$$

$$b_1 \in (0.31; 1.69)$$

$$s_{b_2} = 2.78 \cdot 1.3 \cdot \sqrt{0.01}$$

$$b_2 \in (-1.64; -0.84)$$

$$\hat{b} = \begin{pmatrix} 5.38 & 1 & -1.24 \end{pmatrix}^T$$

$$C = \begin{pmatrix} 0.33 & 0 & 0.05 \\ 0 & 0.04 & 0 \\ 0.05 & 0 & 0.01 \end{pmatrix}$$

$$S_\varepsilon^2 = \frac{Q(\hat{b})}{n - (k+1)}$$

$$S_\varepsilon^2 = \frac{6,78}{7 - (3)}$$

$$S_\varepsilon^2 = 1.695$$

$$S_\varepsilon = 1.3$$

$$t_{табл} = 2.78$$

$$\alpha = 0.05 \quad \nu = 4$$

Пример
(продолжение)

# Адекватность модели

$$y = 5.38 + x - 1.24x^2$$

$$H_0: \; D_y = D_\varepsilon$$
$$H_1: \; D_y > D_\varepsilon$$

$$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$\frac{S_y^2}{S_\varepsilon^2} \sim F_{(n-1;\,n-(k+1))}$$

$$S_\varepsilon^2 = \frac{Q(\vec{b})}{n-(k+1)}$$

$$S_\varepsilon^2 = \frac{6{,}78}{7-(3)}$$

$$S_\varepsilon^2 = 1.695$$

$$S_y^2 = \frac{163.71}{7-1} = 27.285$$

$$F_{набл} = 16.09$$

$$F_{табл} = 6.16$$

Вывод: модель адекватна