

1) **Парная регрессия:** $\mathbb{E}(Y | X = x) = \beta_0 + \beta_1 x$

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ Y_i – зависимая;
 X_i – регрессор;
 ε_i – навязка

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \underset{\text{остаток}}{e_i} \quad RSS = \sum_1^n e_i^2 = \sum_1^n (Y_i - \hat{Y}_i)^2$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

2) **МНК:** $RSS \rightarrow \min$

$$FOC : \begin{cases} n\hat{\beta}_0 + \sum_1^n X_i \hat{\beta}_1 = \sum_1^n Y_i \\ \sum_1^n X_i \hat{\beta}_0 + \sum_1^n X_i^2 \hat{\beta}_1 = \sum_1^n X_i Y_i \end{cases} \Rightarrow \begin{cases} \hat{\beta}_1 = \frac{\sum_1^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_1^n (X_i - \bar{X})^2} = \frac{sCov(X, Y)}{sVar(X)} \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \end{cases}$$

3) **Леммы:**

1. Линия регрессии проходит через точку (\bar{X}, \bar{Y}) , т.е. $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$

$$\text{Док-во: } \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \Rightarrow \boxed{\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}}$$

$$2. \sum_1^n e_i = 0$$

Док-во:

$$e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

$$\sum_1^n e_i = \sum_1^n Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_1^n X_i$$

$$\frac{1}{n} \sum_1^n e_i = \bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} = 0 \Rightarrow \boxed{\sum_1^n e_i = 0}$$

$$3. \sum_1^n Y_i = \sum_1^n \hat{Y}_i$$

Док-во:

$$\sum_1^n e_i = \sum_1^n Y_i - \sum_1^n \hat{Y}_i = 0 \Rightarrow \boxed{\sum_1^n Y_i = \sum_1^n \hat{Y}_i}$$

Следствие:

$$\frac{1}{n} \sum_1^n e_i = \bar{Y} - \bar{\hat{Y}} = 0 \Rightarrow \boxed{\bar{Y} = \bar{\hat{Y}}}$$

$$4. \sum_1^n X_i e_i = 0$$

Док-во:

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = -2 \sum_1^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0 \Rightarrow \boxed{\sum_1^n X_i e_i = 0}$$

$$5. \sum_1^n \hat{Y}_i e_i = 0$$

Док-во:

$$\sum_1^n \hat{Y}_i e_i = \sum_1^n (\hat{\beta}_0 + \hat{\beta}_1 X_i) e_i = \hat{\beta}_0 \sum_1^n e_i + \hat{\beta}_1 \sum_1^n X_i e_i = 0 \Rightarrow \boxed{\sum_1^n \hat{Y}_i e_i = 0}$$

$$4) Y_i = \hat{Y}_i + e_i; \bar{Y} = \bar{\hat{Y}} \Rightarrow Y_i - \bar{Y} = \hat{Y}_i - \bar{\hat{Y}} + e_i$$

$$\sum_1^n (Y_i - \bar{Y})^2 = \sum_1^n (\hat{Y}_i - \bar{\hat{Y}})^2 + 2(\sum_1^n \hat{Y}_i e_i - \bar{\hat{Y}} \sum_1^n e_i) + \sum_1^n e_i^2$$

$$\sum_1^n (Y_i - \bar{Y})^2 = \sum_1^n (\hat{Y}_i - \bar{\hat{Y}})^2 + \sum_1^n e_i^2$$

$$TSS = ESS + RSS$$

5) Коэффициент детерминации R^2 :

$$R^2 = \frac{ESS}{TSS}; x_i = X_i - \bar{X}; y_i = Y_i - \bar{Y}; \bar{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

$$R^2 = \frac{\sum_1^n (\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_1^n y_i^2} = \frac{\sum_1^n (\hat{\beta}_0 + \hat{\beta}_1 X_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{X})^2}{\sum_1^n y_i^2} = \frac{\sum_1^n (\hat{\beta}_1 x_i)^2}{\sum_1^n y_i^2} = \frac{\hat{\beta}_1^2 \sum_1^n x_i^2}{\sum_1^n y_i^2} = \frac{\sum_1^n (x_i y_i)^2 \cdot \sum_1^n x_i^2}{(\sum_1^n x_i^2)^2 \cdot \sum_1^n y_i^2} =$$

$$= \frac{\sum_1^n (x_i y_i)^2}{\sum_1^n x_i^2 \cdot \sum_1^n y_i^2} = \frac{sCov^2(X, Y)}{sVar(X) \cdot sVar(Y)} = sCorr^2(X, Y)$$

6) Регрессия без свободного члена:

$$Y_i = \beta_1 X_i + \varepsilon_i$$

$$\hat{Y}_i = \hat{\beta}_1 X_i \quad RSS \rightarrow \min : \hat{\beta}_1 = \frac{\sum_1^n x_i y_i}{\sum_1^n x_i^2}$$

$$e_i = Y_i - \hat{\beta}_1 X_i$$

Если:

$$a) \sum_1^n e_i \neq 0$$

$$b) \sum_1^n Y_i \neq \sum_1^n \hat{Y}_i$$

$$c) \bar{Y} \neq \bar{\hat{Y}},$$

то R^2 неприменим

7) Теорема Гаусса-Маркова для парной регрессии:

Если:

1. Модель правильно специфицирована
2. Все X_i детерминированы и не равны между собой
3. Ошибки не носят систематического характера, т.е. $\mathbb{E}(\varepsilon_i) = \forall i$
4. $\text{Var}(\varepsilon_i) = \sigma_\varepsilon^2 \forall i$
5. Ошибки некоррелированы

Тогда оценки МНК $\hat{\beta}_0, \hat{\beta}_1$ оптимальны (имеют наименьшую дисперсию) в классе линейных несмещенных оценок – являются BLUE.

8) МНК-оценки - случайные величины:

$$\hat{\beta}_1 = \frac{sCov(X, Y)}{sVar(X)} = \frac{sCov(X, \beta_0 + \beta_1 X + \varepsilon)}{sVar(X)} = \beta_1 \frac{sVar(X)}{sVar(X)} + \frac{sCov(X, \varepsilon)}{sVar(X)} = \beta_1 + \frac{sCov(X, \varepsilon)}{sVar(X)}$$

$$\begin{aligned} \mathbb{E}(\hat{\beta}_1) &= \mathbb{E} \left(\beta_1 + \frac{sCov(X, \varepsilon)}{sVar(X)} \right) = \beta_1 + \frac{1}{sVar(X)} \cdot \mathbb{E} \left(\frac{1}{n-1} \sum_1^n (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon}) \right) = \\ &= \beta_1 + \frac{1}{sVar(X)} \cdot \frac{1}{n-1} \cdot \sum_1^n (X_i - \bar{X})(\mathbb{E}(\varepsilon_i) - \mathbb{E}(\bar{\varepsilon})) \stackrel{\mathbb{E}(\varepsilon) = \mathbb{E}(\bar{\varepsilon})}{=} \beta_1 \end{aligned}$$

Утверждение 1: $\text{Var}(\hat{\beta}_1) = \frac{\sigma_\varepsilon^2}{\sum_1^n x_i^2}; \text{Var}(\hat{\beta}_0) = \frac{\sigma_\varepsilon^2 \sum_1^n x_i^2}{n \sum_1^n x_i^2}$

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_j x_j^2} = \sum_i \frac{x_i}{\sum_j x_j^2} y_i = \sum_i w_i y_i$$

$$\begin{aligned} a) \sum_i w_i &= 0 & b) \sum_i w_i x_i &= 1 & c) \sum_i w_i^2 &= \frac{1}{\sum_j x_j^2} \\ \sum_i w_i &= \frac{\sum_i x_i}{\sum_j x_j^2} = \frac{\sum_i (X_i - \bar{X})}{\sum_j x_j^2} = \frac{n\bar{X} - n\bar{X}}{\sum_j x_j^2} = 0 & \frac{\sum_i x_i \cdot x_i}{\sum_j x_j^2} &= 1 & \sum_i w_i^2 &= \frac{\sum_i x_i^2}{(\sum_j x_j^2)^2} = \frac{1}{\sum_j x_j^2} \end{aligned}$$

$$\text{Var}(\hat{\beta}_1) = \text{Var}\left(\sum_i w_i y_i\right) = \sum_i w_i^2 \text{Var}(y_i) + \sum_i \sum_j w_i w_j \text{Cov}(y_i, y_j) = \sum_i w_i^2 \sigma_\varepsilon^2 = \boxed{\frac{\sigma_\varepsilon^2}{\sum_j x_j^2}}$$

$$\begin{aligned} \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} = \frac{1}{n} \sum_1^n Y_i - \sum_1^n w_i (Y_i - \bar{Y}) \cdot \bar{X} = \sum_1^n \left(\frac{1}{n} - w_i \bar{X} \right) Y_i + \bar{X} \bar{Y} \sum_1^n w_i = \sum_1^n \left(\frac{1}{n} - w_i \bar{X} \right) Y_i \\ \mathbb{E}(\beta_0) &= \mathbb{E}\left(\sum_1^n \left(\frac{1}{n} - w_i \bar{X} \right) Y_i\right) = \sum_1^n \left(\frac{1}{n} - w_i \bar{X} \right) \mathbb{E}(Y_i) = 0?? \end{aligned}$$

$$\begin{aligned} \text{Var}(\hat{\beta}_0) &= \sum_1^n \left(\frac{1}{n^2} - \frac{2}{n} w_i \bar{X} + w_i^2 \bar{X}^2 \right) \sigma_\varepsilon^2 = \sigma_\varepsilon^2 \left(\frac{1}{n} - \frac{2}{n} \bar{X} \sum_1^n w_i + \bar{X}^2 \sum_1^n w_i^2 \right) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_1^n x_i^2} \right) = \\ &= \sigma_\varepsilon^2 \frac{\sum_1^n x_i^2 + n \bar{X}^2}{n \sum_1^n x_i^2} = \boxed{\sigma_\varepsilon^2 \frac{\sum_1^n X_i^2}{n \sum_1^n x_i^2}} \end{aligned}$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = ?$$

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}; \text{Var}(\bar{Y}) = \frac{\sigma_\varepsilon^2}{n}$$

$$\text{Var}(\hat{\beta}_0 + \hat{\beta}_1 \bar{X}) = \text{Var}(\hat{\beta}_0) + 2\bar{X} \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \bar{X}^2 \text{Var}(\hat{\beta}_1)$$

$$\frac{\sigma_\varepsilon^2}{n} = \sigma_\varepsilon^2 \frac{\sum_1^n X_i^2}{n \sum_1^n x_i^2} + 2\bar{X} \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \bar{X}^2 \frac{\sigma_\varepsilon^2}{\sum_j x_j^2}$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{\sigma_\varepsilon^2 \left(\frac{1}{n} - \frac{1}{n} - \frac{\bar{X}^2}{\sum_1^n x_i^2} - \frac{\bar{X}^2}{\sum_1^n x_i^2} \right)}{2\bar{X}} = \boxed{\frac{-\sigma_\varepsilon^2 \cdot 2\bar{X}^2}{2\bar{X} \sum_1^n x_i^2}}$$

Утверждение 2: при выполнении условий ТГМ оценки $\hat{\beta}_0, \hat{\beta}_1$ являются лучшими, т.е. имеют наименьшую дисперсию в классе всех линейных несмещенных оценок

Пусть $\tilde{\beta}_1 = \sum_1^n \tilde{w}_i Y_i$ – другая несмещенная оценка, т.е. $\mathbb{E}(\tilde{\beta}_1) = \beta_1$

$$\mathbb{E}(\tilde{\beta}_1) = \sum_1^n \tilde{w}_i \mathbb{E}(Y_i) = \sum_1^n \tilde{w}_i \mathbb{E}(\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_0 \sum_1^n \tilde{w}_i + \beta_1 \sum_1^n \tilde{w}_i X_i \equiv \beta_1 \Rightarrow$$

$$(1) \sum_1^n \tilde{w}_i = 0$$

$$(2) \sum_1^n \tilde{w}_i x_i = 1$$

Т.е. необходимо решить задачу: $\text{Var}(\tilde{\beta}_1) = \sigma_\varepsilon^2 \sum_1^n \tilde{w}_i^2 \rightarrow \min$ при ограничениях (1) и (2)

$$\sum_1^n \tilde{w}_i^2 = \sum_1^n (\tilde{w}_i - w_i + w_i)^2 = \sum_1^n (\tilde{w}_i - w_i)^2 + 2 \sum_1^n (\tilde{w}_i - w_i) w_i + \sum_1^n w_i^2$$

$$\begin{aligned} \sum_i (\tilde{w}_i - w_i) w_i &= \sum_i \tilde{w}_i \cdot w_i - \sum_i w_i^2 = \sum_i \tilde{w}_i \frac{(X_i - \bar{X})}{\sum_j x_j^2} - \sum_i w_i^2 = \frac{1}{\sum_j x_j^2} \left(\sum_i \tilde{w}_i X_i - \bar{X} \sum_1^n \tilde{w}_i \right) - \frac{1}{\sum_j x_j^2} = \\ &= \frac{1}{\sum_j x_j^2} - \frac{1}{\sum_j x_j^2} = 0 \Rightarrow \sum_i \tilde{w}_i^2 = \sum_1^n (\tilde{w}_i - w_i)^2 + \frac{1}{\sum_j x_j^2} \end{aligned}$$

$\Rightarrow \sum_1^n \tilde{w}_i^2$ достигает минимума при $\tilde{w}_i = w_i$, т.е. для оценок МНК.

Для β_0 доказательство аналогично

9) МНК - линейны по Y?

$$\hat{\beta}_1 = \frac{sCov(X, Y)}{sVar(X)} \quad \text{Если } \tilde{Y} = nY, \quad \hat{\tilde{\beta}} = \frac{sCov(X, \tilde{Y})}{sVar(X)} = n\hat{\beta}_1$$

10) Гипотезы о конкретном значении:

В рамках линейной модели усилим ТГМ утверждением $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$

Утверждение 1: Если $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, то $\hat{\beta}_0 \sim N(\beta_0, \sigma_{\hat{\beta}_0}^2)$, $\hat{\beta}_1 \sim N(\beta_1, \sigma_{\hat{\beta}_1}^2)$

$$H_0 : \beta_1 = \beta_1^0$$

Тогда Что-то тут написано $z = \frac{\hat{\beta}_1 - \beta_1^0}{\sigma_{\hat{\beta}_1}} \sim N(0, 1)$

В $\sigma_{\hat{\beta}_1}$ входит σ_ε^2 и она неизвестна. Заменим оценкой $\hat{\sigma}_\varepsilon^2$

$$z = \frac{\hat{\beta}_1 - \beta_1^0}{\hat{\sigma}_{\hat{\beta}_1}} \sim ?$$

Утверждение 2:

$$\sigma_\varepsilon^2 = \frac{RSS}{n-2} - \text{несмещенная оценка } \sigma_\varepsilon^2$$

Утверждение 3:

$$\sigma_\varepsilon^2 = \frac{RSS}{n-k} - \text{общий случай для } k \text{ регрессоров (включая свободный член)}$$

$n - k$ - число независимых слагаемых RSS . Степени свободы

Утверждение 4:

$$\frac{RSS}{\sigma_\varepsilon^2} \sim \chi_{n-k}^2, \text{ где } k - \text{количество регрессоров}$$

Утверждение 5: Оценки β_1 и $\hat{\sigma}_\varepsilon^2$ независимы

$$\hat{\sigma}_{\hat{\beta}_0^2} = \sqrt{\hat{\sigma}_{\varepsilon^2} \cdot \frac{\sum X_i^2}{n \sum X_i^2}}$$

$$\hat{\sigma}_{\hat{\beta}_1^2} = \sqrt{\frac{\hat{\sigma}_{\varepsilon^2}}{\sum X_i^2}}$$

$$\frac{\hat{\beta}_1 - \beta_1^0}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{\frac{\hat{\beta}_1 - \beta_1^0}{\sigma_{\hat{\beta}_1}}}{\frac{\sigma_{\hat{\beta}_1}}{\sigma_{\hat{\beta}_1}}} = \text{продолжение следует...}$$