

Klasyczny Model Regresji Liniowej

Wykład 8

Natalia Nehrebecka Stanisław Cichocki

26 listopada 2018

Plan zajęć

- 1 Klasyczny Model Regresji Liniowej
 - Założenia KMRL
 - Dodatkowo założenie klasycznego modelu regresji liniowej
 - Własności estymatora MNK w KMRL
- 2 Estymator macierzy wariancji i kowariancji estymatora MNK
 - Estymator wariancji błędu losowego
 - Estymator macierzy kowariancji \mathbf{b}
 - Kombinacja liniowa wektora \mathbf{b}
- 3 Pytania teoretyczne

- 1 $y_i = x_{1i}\beta_1 + \dots + x_{ki}\beta_k + \varepsilon_i$, zapisie macierzowym:
 $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- 2 Zmienne objaśniające x_{1i}, \dots, x_{ki} są nielosowe dla $i = 1, \dots, N$
- 3 $E(\varepsilon_i) = 0$ dla $i = 1, \dots, N$, zapisie macierzowym: $E(\varepsilon) = \mathbf{0}$
- 4 $Cov(\varepsilon_i, \varepsilon_j) = 0$ dla $i \neq j$ (brak *autokorelacji* błędu losowego)
- 5 $Var(\varepsilon_i) = \sigma^2$ dla $i = 1, \dots, N$ (*homoskedastyczność*
 \neq *heteroskedastyczność* błędu losowego)

- 1 $y_i = x_{1i}\beta_1 + \dots + x_{ki}\beta_k + \varepsilon_i$, zapisie macierzowym:
 $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- 2 Zmienne objaśniające x_{1i}, \dots, x_{ki} są nielosowe dla $i = 1, \dots, N$
- 3 $E(\varepsilon_i) = 0$ dla $i = 1, \dots, N$, zapisie macierzowym: $E(\varepsilon) = \mathbf{0}$
- 4 $Cov(\varepsilon_i, \varepsilon_j) = 0$ dla $i \neq j$ (brak *autokorelacji* błędu losowego)
- 5 $Var(\varepsilon_i) = \sigma^2$ dla $i = 1, \dots, N$ (*homoskedastyczność*
 \neq *heteroskedastyczność* błędu losowego)

- 1 $y_i = x_{1i}\beta_1 + \dots + x_{ki}\beta_k + \varepsilon_i$, zapisie macierzowym:
 $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- 2 Zmienne objaśniające x_{1i}, \dots, x_{ki} są nielosowe dla $i = 1, \dots, N$
- 3 $E(\varepsilon_i) = 0$ dla $i = 1, \dots, N$, zapisie macierzowym: $E(\varepsilon) = \mathbf{0}$
- 4 $Cov(\varepsilon_i, \varepsilon_j) = 0$ dla $i \neq j$ (brak *autokorelacji* błędu losowego)
- 5 $Var(\varepsilon_i) = \sigma^2$ dla $i = 1, \dots, N$ (*homoskedastyczność*
 \neq *heteroskedastyczność* błędu losowego)

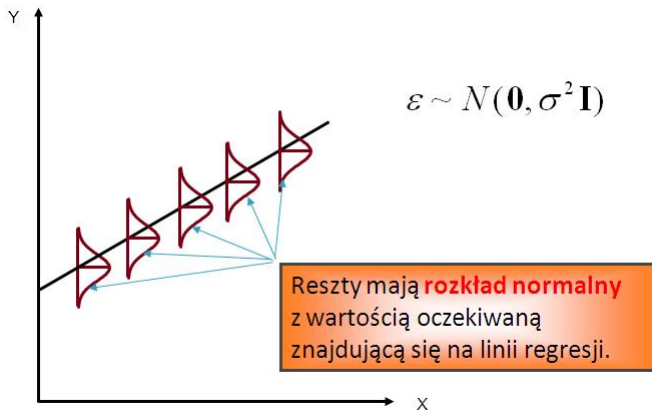
- 1 $y_i = x_{1i}\beta_1 + \dots + x_{ki}\beta_k + \varepsilon_i$, zapisie macierzowym:
 $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- 2 Zmienne objaśniające x_{1i}, \dots, x_{ki} są nielosowe dla $i = 1, \dots, N$
- 3 $E(\varepsilon_i) = 0$ dla $i = 1, \dots, N$, zapisie macierzowym: $E(\varepsilon) = \mathbf{0}$
- 4 $Cov(\varepsilon_i, \varepsilon_j) = 0$ dla $i \neq j$ (brak *autokorelacji* błędu losowego)
- 5 $Var(\varepsilon_i) = \sigma^2$ dla $i = 1, \dots, N$ (*homoskedastyczność*
≠heteroskedastyczność błędu losowego)

- 1 $y_i = x_{1i}\beta_1 + \dots + x_{ki}\beta_k + \varepsilon_i$, zapisie macierzowym:
 $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- 2 Zmienne objaśniające x_{1i}, \dots, x_{ki} są nielosowe dla $i = 1, \dots, N$
- 3 $E(\varepsilon_i) = 0$ dla $i = 1, \dots, N$, zapisie macierzowym: $E(\varepsilon) = \mathbf{0}$
- 4 $Cov(\varepsilon_i, \varepsilon_j) = 0$ dla $i \neq j$ (brak *autokorelacji* błędu losowego)
- 5 $Var(\varepsilon_i) = \sigma^2$ dla $i = 1, \dots, N$ (*homoskedastyczność*
 \neq *heteroskedastyczność* błędu losowego)

- Zaburzenia losowe ε są sferyczne.

$$\text{Var}(\varepsilon) = \sigma^2 I = \begin{bmatrix} \text{Var}(\varepsilon_1) & \text{Cov}(\varepsilon_2, \varepsilon_1) & \cdots & \text{Cov}(\varepsilon_N, \varepsilon_1) \\ \text{Cov}(\varepsilon_1, \varepsilon_2) & \text{Var}(\varepsilon_2) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\varepsilon_1, \varepsilon_N) & \text{Cov}(\varepsilon_2, \varepsilon_N) & \cdots & \text{Var}(\varepsilon_N) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}$$



1 **Nieobciążoność:** $E(b) = \beta$

2 **Efektywność:**

Twierdzenie (Gausa-Markowa) Dla spełnionych założeń KMRL estymator MNK jest najlepszym estymatorem parametrów β w klasie liniowych i nieobciążonych estymatorów tego parametru.

- Wektor oszacowań (estymator) parametrów jest wektorem **losowym**, może więc odbiegać od prawdziwych wartości parametrów
- Precyzję oszacowania mierzy się za pomocą miar rozrzutu, dyspersji
- Najpopularniejszą miarą rozrzutu jest macierz **wariancji-kowariancji**

-

Wyprowadzenie wzoru

- Ważne założenia KMRL: *brak autokorelacji* ($cov(\varepsilon_i, \varepsilon_j) = 0$ dla $i \neq j$) i *homoskedastyczność* ($var(\varepsilon_i) = \sigma^2$ dla $i = 1, \dots, N$)

$$\begin{aligned} \text{Var}(\mathbf{b}) &= \text{Var}((X'X)^{-1}X'y) = \text{Var}((X'X)^{-1}X'(X\beta + \varepsilon)) \\ &= \text{Var}(\beta + (X'X)^{-1}X'\varepsilon) = (X'X)^{-1}X' \underbrace{\text{Var}(\varepsilon)}_{\sigma^2 I} X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1} = \Sigma \end{aligned}$$

$$\Sigma = \begin{bmatrix} \text{Var}(b_1) & \cdots & \text{cov}(b_p, b_q) \\ \vdots & \ddots & \vdots \\ \text{cov}(b_q, b_p) & \cdots & \text{Var}(b_k) \end{bmatrix}$$

- Oszacowanie wariancji estymatora parametrów Σ uzyskamy, jeśli uda się znaleźć estymator wariancji błędu losowego σ^2

- Oszacowanie błędów losowych w MNK - reszty
- Dlatego estymator wariancji błędów losowych buduje się na podstawie wariancji reszt

Nieobciążony estymator wariancji

Ponieważ

$$E\left(\frac{\mathbf{e}'\mathbf{e}}{N-K}\right) = \sigma^2$$

Dlatego nieobciążonym estymatorem wariancji błędu losowego jest

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{N-K}$$

$$\hat{\Sigma} = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

- $se(b_k)$ - oszacowanie odchylenia błędu standardowego b_k , które wykorzystuje się do mierzenia precyzji oszacowań parametrów.

$$\hat{se}(b_k) = \sqrt{[\hat{\Sigma}_b]_{kk}}$$

- Ponieważ estymator b jest liniowy, to analogiczne własności posiada również dowolna kombinacja liniowa wektora b .
- Weźmy na przykład wektor złożony ze stałych v o nie wszystkich elementach jednocześnie równych 0:

$$\bullet v = \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix}$$

- i utwórzmy kombinację liniową wektora v i b , a więc:
 - $v'b = v_1 b_1 + \dots + v_k b_k$.
- Ta kombinacja liniowa jest również najlepszym liniowym i nieobciążonym estymatorem kombinacji liniowej $v'\beta$.

- 1 Wymienić założenia Klasycznego Modelu Regresji Liniowej (KMRL).
- 2 Udowodnić, że, w KMRL estymator \mathbf{b} jest nieobciążony.
- 3 Wyprowadzić postać macierzy wariancji kowariancji \mathbf{b} i podać interpretację jej elementów.
- 4 Podać (słowami) treść twierdzenia Gaussa-Markowa i wyjaśnić jego znaczenie.
- 5 Pokazać, że s^2 jest nieobciążonym estymatorem σ^2 .
- 6 Udowodnić, że $s^2(X'X)^{-1}$ jest nieobciążonym estymatorem $\mathbf{Var}(\mathbf{b})$.