Klasyczny Model Regresji Liniowej

Wykład 8

Natalia Nehrebecka Stanisław Cichocki

21 listopada 2018

Plan zajęć

1. $y_i = x_{1i}\beta_1 + \cdots + x_{Ki}\beta_K + \varepsilon_i$, zapisie macierzowym: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
2. Zmienne objaśniające $x_{1i}, \ldots, x_{ki}$ są nielosowe dla $i = 1, \ldots, N$
3. $E(\varepsilon_i) = 0$ dla $i = 1, \ldots, N$, 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 $i = 1, \ldots, N$ (*homoskedastyczność* $\neq$*heteroskedastyczność* błędu losowego)

1. $y_i = x_{1i}\beta_1 + \cdots + x_{Ki}\beta_K + \varepsilon_i$, zapisie macierzowym: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
2. Zmienne objaśniające $x_{1i}, \ldots, x_{ki}$ są nielosowe dla $i = 1, \ldots, N$
3. $E(\varepsilon_i) = 0$ dla $i = 1, \ldots, N$, 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 $i = 1, \ldots, N$ (*homoskedastyczność* $\neq$*heteroskedastyczność* błędu losowego)

1. $y_i = x_{1i}\beta_1 + \cdots + x_{Ki}\beta_K + \varepsilon_i$, zapisie macierzowym: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
2. Zmienne objaśniające $x_{1i}, \ldots, x_{ki}$ są nielosowe dla $i = 1, \ldots, N$
3. $E(\varepsilon_i) = 0$ dla $i = 1, \ldots, N$, 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 $i = 1, \ldots, N$ (*homoskedastyczność* $\neq$*heteroskedastyczność* błędu losowego)

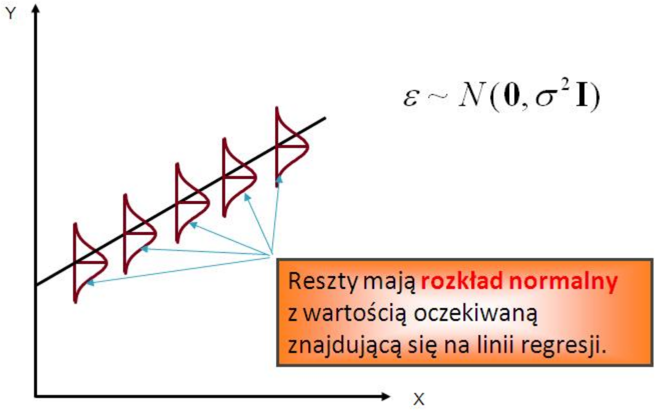
1. $y_i = x_{1i}\beta_1 + \cdots + x_{Ki}\beta_K + \varepsilon_i$, zapisie macierzowym: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
2. Zmienne objaśniające $x_{1i}, \ldots, x_{ki}$ są nielosowe dla $i = 1, \ldots, N$
3. $E(\varepsilon_i) = 0$ dla $i = 1, \ldots, N$, zapisie macierzowym: $E(\varepsilon)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

1. $y_i = x_{1i}\beta_1 + \cdots + x_{Ki}\beta_K + \varepsilon_i$, zapisie macierzowym: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
2. Zmienne objaśniające $x_{1i}, \ldots, x_{ki}$ są nielosowe dla $i = 1, \ldots, N$
3. $E(\varepsilon_i) = 0$ dla $i = 1, \ldots, N$, 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 $i = 1, \ldots, N$ (*homoskedastyczność* $\neq$*heteroskedastyczność* błędu losowego)

- Zaburzenia losowe ε są sferyczne.

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

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


Y
X
$\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$
Reszty mają rozkład normalny
z wartością oczekiwaną
znajdującą się na linii regresji.

1. **Nieobciążoność**: $E(b) = \beta$
2. **Efektywność**:
   Twierdzenie (Gaussa-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 $i = 1, \dots, N$)

$$Var(\mathbf{b}) = Var((X'X)^{-1}X'y) = Var((X'X)^{-1}X'(X\beta + \varepsilon)$$

$$= Var(\beta + (X'X)^{-1}X'\varepsilon) = (X'X)^{-1}X' \underbrace{Var(\varepsilon)}_{\sigma^2 I} X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1} = \Sigma$$

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

- 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.

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

1. Wymienić założenia Klasycznego Modelu Regresji Liniowej (KMRL).
2. Udowodnić, że, w KMRL estymator **b** jest nieobciążony.
3. Wyprowadzić postać macierzy wariancji kowariancji **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.