

Klasyczny Model Regresji Liniowej

Stanisław Cichocki

Natalia Nehrebecka

Wykład 8

Plan wykładu

- ▶ 1. Założenia KMRL
- ▶ 2. Własności estymatora MNK w KMRL
 - Twierdzenie Gaussa-Markowa
- ▶ 3. Estymator wariancji błędu losowego: podstawowa macierz idempotentna

Plan wykładu

- ▶ 1. Założenia KMRL
- ▶ 2. Własności estymatora MNK w KMRL
 - Twierdzenie Gaussa-Markowa
- ▶ 3. Estymator wariancji błędu losowego: podstawowa macierz idempotentna

Klasyczny model regresji liniowej

- ▶ Na poprzednich wykładach pokazaliśmy, iż estymator MNK daje oszacowania parametrów, które są najlepiej dopasowane do danych
- ▶ Obecnie zajmiemy się własnościami statystycznymi tego estymatora i w tym celu przyjmujemy pewne dodatkowe założenia
- ▶ Najprostszym i najpopularniejszym układem założeń jest KMRL

Założenia klasycznego modelu regresji liniowej

- ▶ 1. Związek pomiędzy zmienną zależną a zmiennymi niezależnymi opisany jest równaniem:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_K x_{Ki} + \varepsilon_i \quad i = 1, 2, 3, \dots, n$$

- ▶ 2. Zmienne objaśniające $x_{2i}, x_{3i}, \dots, x_{Ki}$ są nielosowe dla $i = 1, 2, 3, \dots, n$
- ▶ 3. Wartość oczekiwana błędu losowego jest równa zeru:

$$E(\varepsilon) = \mathbf{0}$$

- ▶ 4. Zaburzenia losowe ε są **sferyczne**. Oznacza to, że warunkowa macierz wariancji-kowariancji wektora zaburzeń przy danej macierzy X ma postać:

$$Var(\varepsilon) = \sigma^2 \mathbf{I}$$

- ▶ gdzie \mathbf{I} oznacza macierz jednostkową.

Liniowość w modelu

- ▶ Liniowość w modelu względem:
- ▶ Po pierwsze - **zmiennych objaśniających** X_{Ki} , które są w pierwszej potędze,

$$y_i = \beta_1 + \beta_2 X_{2i}^1 + \dots + \beta_K X_{Ki}^1 + \varepsilon_i$$

- ▶ a po drugie - **względem parametrów** β_k , które są również w pierwszej potędze.

$$y_i = \beta_1^1 + \beta_2^1 X_{2i} + \dots + \beta_K^1 X_{Ki} + \varepsilon_i$$

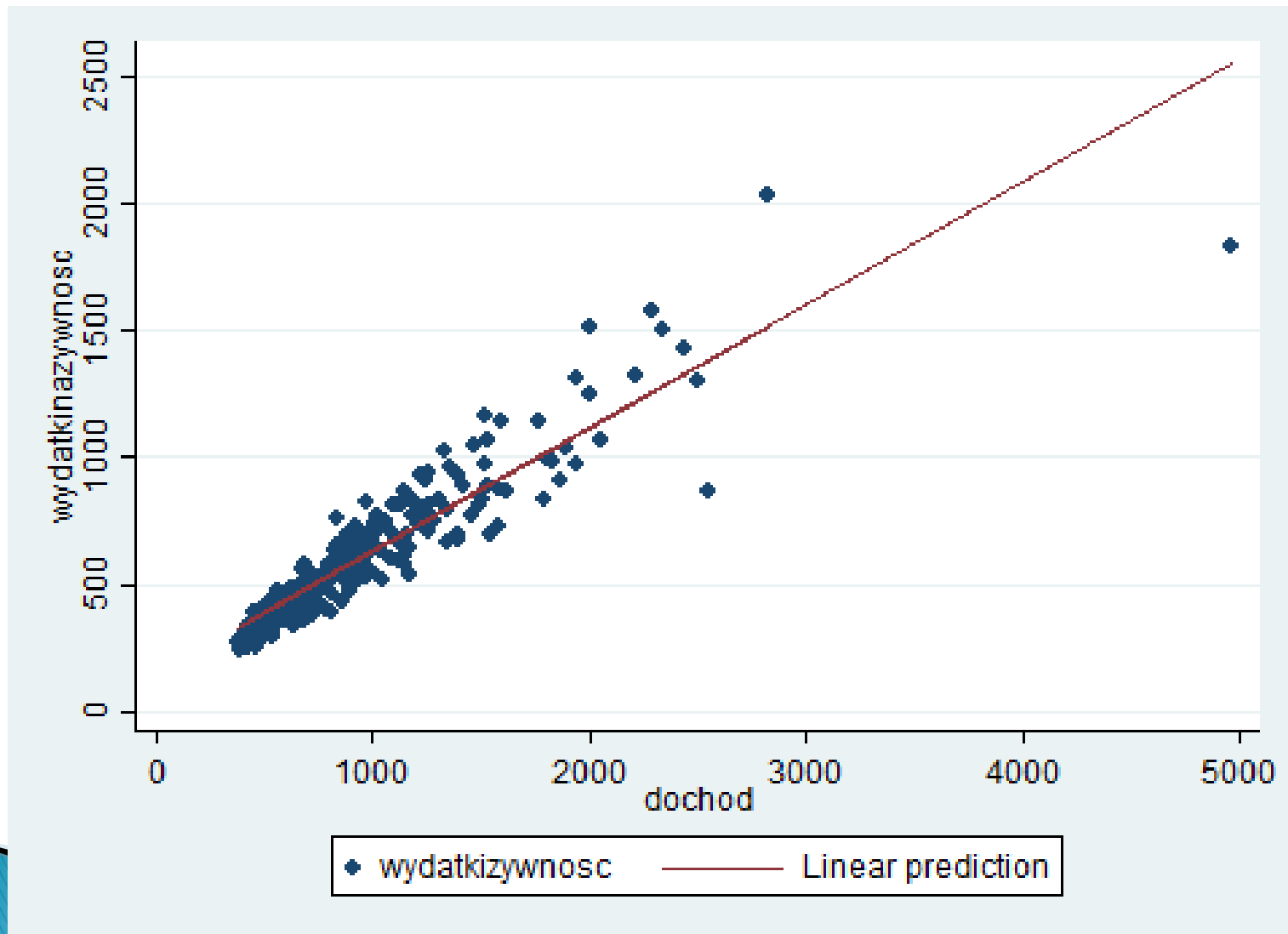
Założenia klasycznego modelu regresji liniowej

$$\text{Var}(\varepsilon) = \sigma^2 \mathbf{I}$$

$$\text{Var}(\varepsilon) = \begin{bmatrix} \text{Var}(\varepsilon_1) & \text{Cov}(\varepsilon_1, \varepsilon_2) & \cdots & \text{Cov}(\varepsilon_1, \varepsilon_n) \\ \text{Cov}(\varepsilon_2, \varepsilon_1) & \text{Var}(\varepsilon_2) & \cdots & \text{Cov}(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(\varepsilon_n, \varepsilon_1) & \text{Cov}(\varepsilon_n, \varepsilon_2) & \cdots & \text{Var}(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

- ▶ Założenie **sferyczności zaburzeń** oznacza:
- ▶ po pierwsze, że **wariacje** kolejnych zaburzeń (elementy na diagonalnej) są takie same dla wszystkich obserwacji i równe σ^2 , gdzie σ^2 jest nieznaną dodatnią stałą (założenie o homoskedastyczności);
- ▶ po drugie, że elementy pozadiagonalne, które są **kowariancjami** zaburzeń dla różnych obserwacji są równe zero, a więc zaburzenia dla różnych obserwacji są ze sobą nieskorelowane (założenie o braku autokorelacji).

Klasyczny model regresji liniowej



Klasyczny model regresji liniowej



Plan wykładu

- ▶ 1. Założenia KMRL
- ▶ 2. Własności estymatora MNK w KMRL
 - Twierdzenie Gaussa-Markowa
- ▶ 3. Estymator wariancji błędu losowego podstawowa macierz idempotentna

Twierdzenie Gaussa-Markowa

W klasycznym modelu regresji liniowej **najlepszym liniowym** i **nieobciążonym** estymatorem wektora parametrów β jest b wyznaczone za pomocą MNK

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

o macierzy wariancji-kowariancji

$$\text{Var}(b) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

$$\text{Var}(b) = \begin{bmatrix} \text{Var}(b_1) & \text{Cov}(b_2, b_1) & \cdots & \text{Cov}(b_n, b_1) \\ \text{Cov}(b_1, b_2) & \text{Var}(b_2) & \cdots & \text{Cov}(b_n, b_2) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(b_1, b_n) & \text{Cov}(b_2, b_n) & \cdots & \text{Var}(b_n) \end{bmatrix}$$

Twierdzenie Gaussa-Markowa

- ▶ 1. Estymator \mathbf{b} jest estymatorem **liniowym**, gdyż jest liniową funkcją zmiennej losowej \mathbf{y} .

- ▶ 2. \mathbf{b} jest estymatorem **nieobciążonym**, to znaczy $E(\mathbf{b}) = \beta$.

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

- ▶ i podstawiając za \mathbf{y} $\mathbf{y} = \mathbf{X}\beta + \varepsilon$

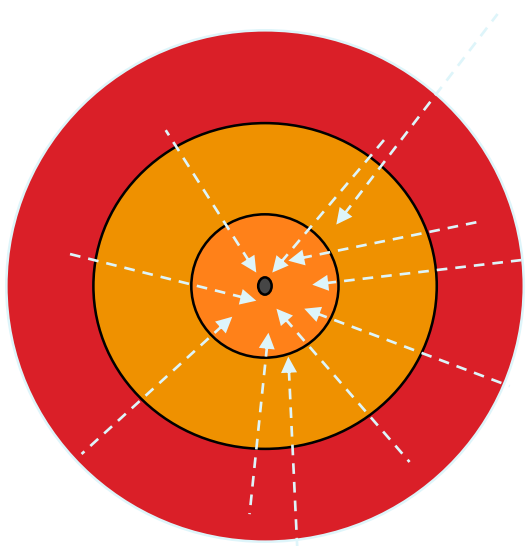
- ▶ otrzymamy:

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

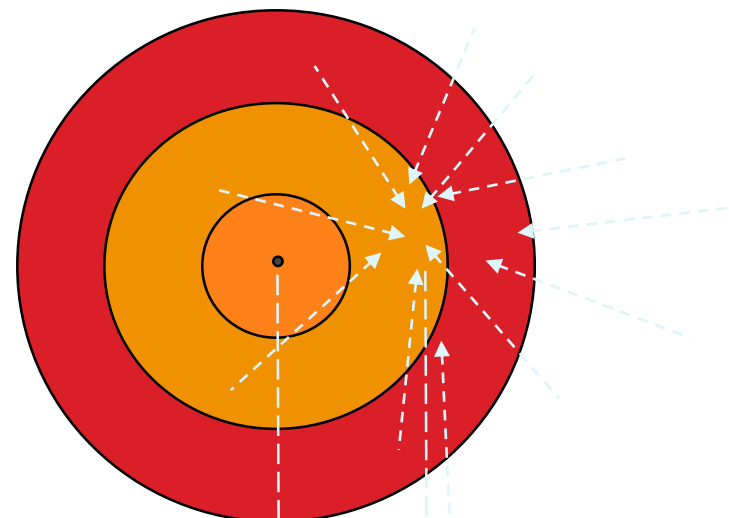
$$E(\mathbf{b}) = \beta + E((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\varepsilon) = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\varepsilon) = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{0} = \beta$$

- ▶ 3. Estymator \mathbf{b} jest estymatorem **najlepszym** w tym sensie, że każdy inny estymator liniowy i nieobciążony ma macierz wariancji-kowariancji większą od tej dla \mathbf{b} . Estymator taki nazywamy estymatorem **efektywnym**.

Obciążony i nieobciążony estymator



**Estymator
nieobciążony**



Obciążenie

**Estymator
obciążony**

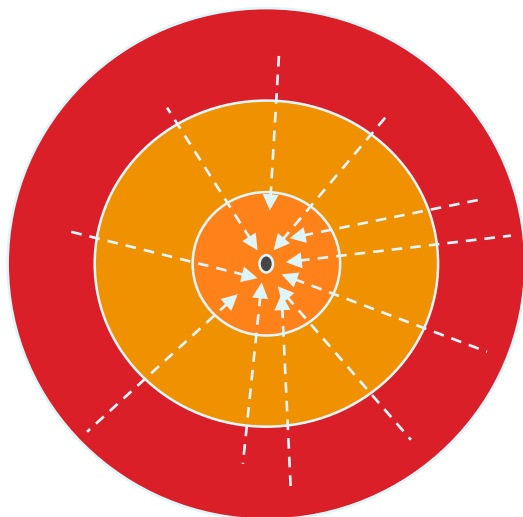
Twierdzenie Gaussa-Markowa

- ▶ Wariancja estymatora \mathbf{b}

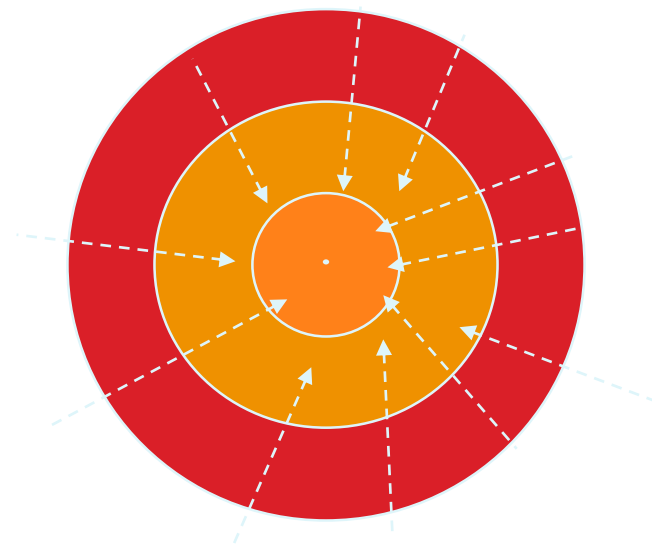
$$\begin{aligned} \text{Var}(b) &= \text{Var}(\beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \varepsilon) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \text{Var}(\varepsilon) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{I} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}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

Efektywność

Estymator jest efektywnym, jeśli ma najniższą wariancję i odchylenie standardowe.



Estymator efektywny



Estymator nieefektywny

Plan wykładu

- ▶ 1. Założenia KMRL
- ▶ 2. Własności estymatora MNK w KMRL
 - Twierdzenie Gaussa-Markowa
- ▶ 3. Estymator wariancji błędu losowego podstawowa macierz idempotentna

Pytania teoretyczne

1. Udowodnić, że w KMRL estymator \mathbf{b} jest nieobciążony.
2. Wyprowadzić postać macierzy wariancji kowariancji \mathbf{b} i podać interpretację jej elementów.
3. Podać treść twierdzenia Gaussa-Markowa i wyjaśnić jego znaczenie.

Dziękuję za uwagę