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

Plan wykładu

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

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:

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

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

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

- 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 **I** oznacza macierz jednostkową.

Założenia klasycznego modelu regresji liniowej

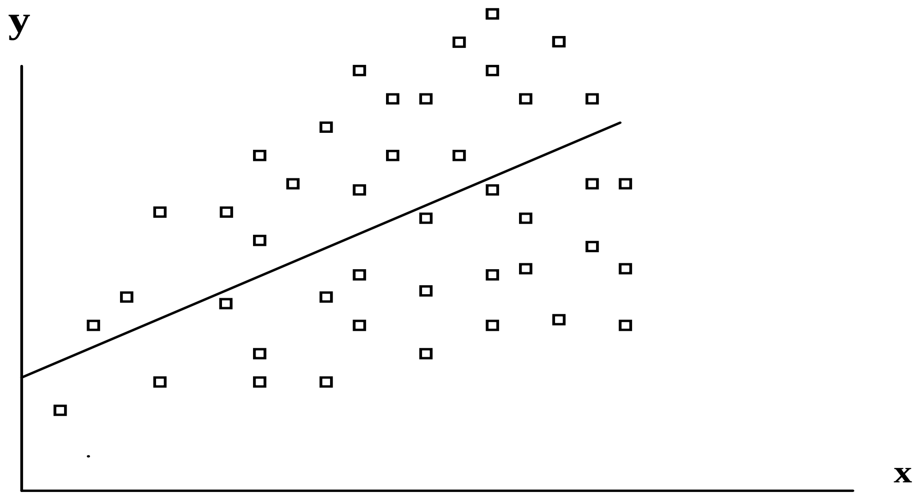
$$Var(\varepsilon) = \sigma^2 \mathbf{I}$$

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

- 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;
- 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.

Założenia klasycznego modelu regresji liniowej

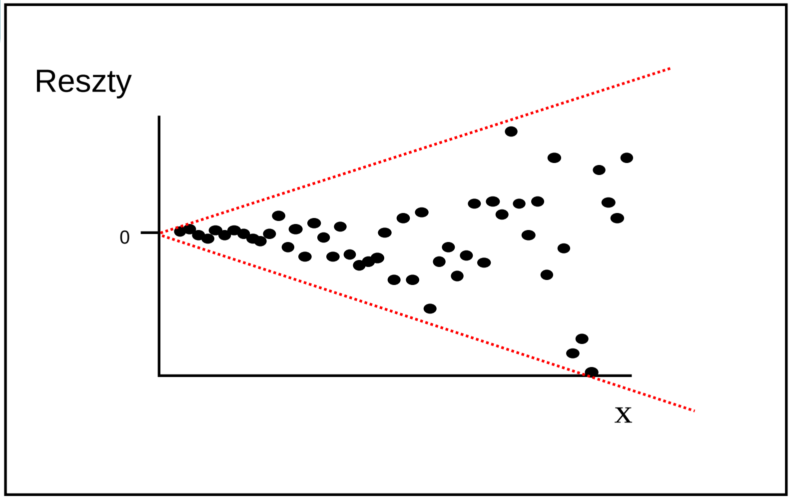
- Stałość wariancji zaburzeń nazywamy homoskedastycznością zaburzeń. **Oznacza to, że zaburzenia losowe są jednakowo rozproszone wokół zerowej wartości oczekiwanej.** Jeśli wariancje nie byłyby jednakowe, to sytuację taką nazywamy heteroskedastycznością.



Rys.1. H**eteroskedastyczność**

Założenia klasycznego modelu regresji liniowej

Oznacza to, że zaburzenia losowe są jednakowo rozproszone wokół zerowej wartości oczekiwanej.

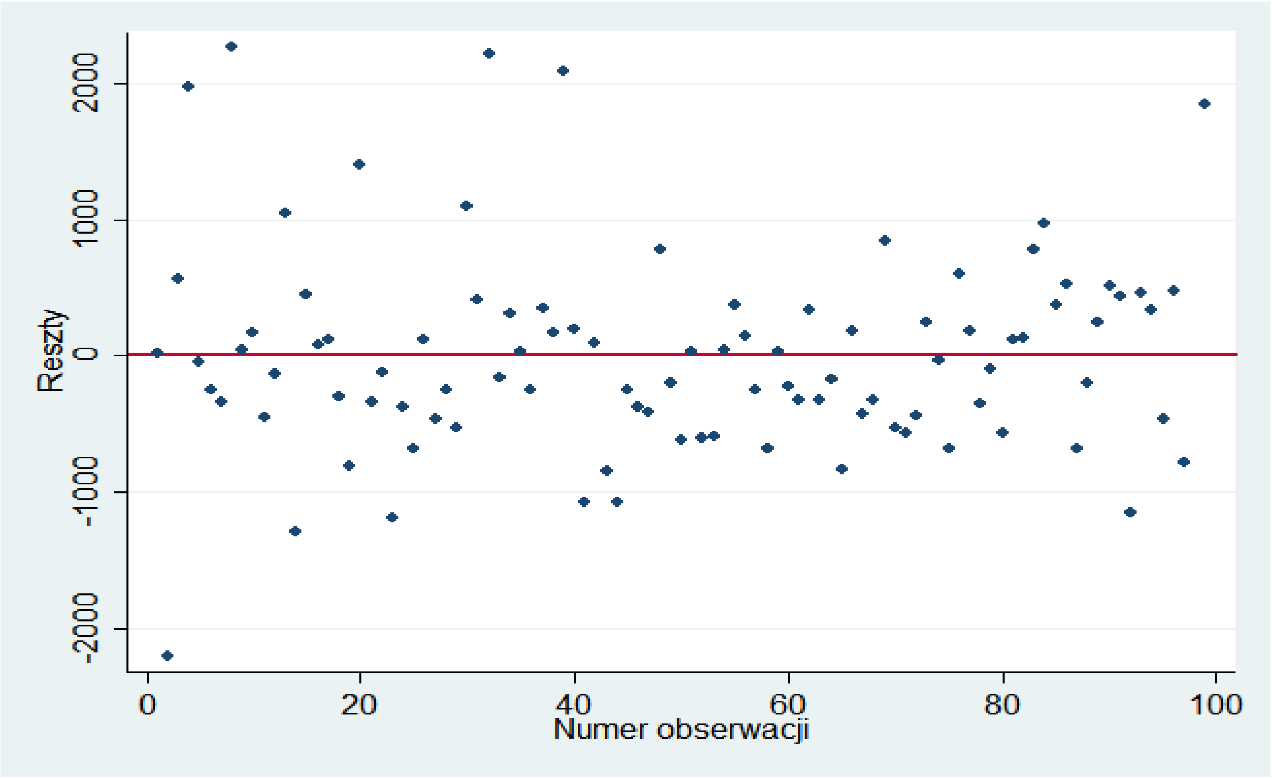


Homoskedastyczność: reszty zachowują się losowo.

Heteroskedastyczność: Wariancja reszt zmienia się wraz ze zmianą zmiennej niezależnej X.

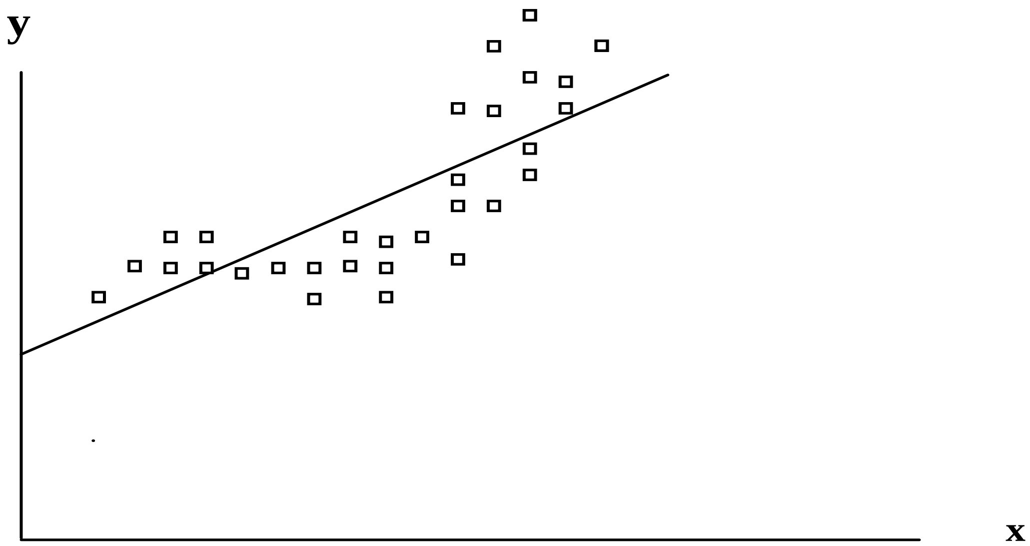
Założenia klasycznego modelu regresji liniowej

- Regresja wydatków na dochodzie



Założenia klasycznego modelu regresji liniowej

- Przypadek zerowych kowariancji dla różnych zaburzeń losowych ε_i oraz ε_j nazywamy brakiem autokorelacji zaburzeń. Oznacza to, że **zaburzenia losowe dla różnych obserwacji są niezależne**, a przez to nieskorelowane, a więc nie mają tendencji do gromadzenia się np. wokół dodatnich lub ujemnych (lub naprzemiennie dodatnich i ujemnych) wartości



Rys. 2. **Autokorelacja**

Plan wykładu

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

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

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

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

Twierdzenie Gaussa-Markowa

- 1. Estymator **b** jest estymatorem **liniowym**, gdyż jest liniową funkcją zmiennej losowej **y.**
- 2. **b** jest estymatorem **nieobciążonym**, to znaczy $E(\mathbf{b}) = \beta$.

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

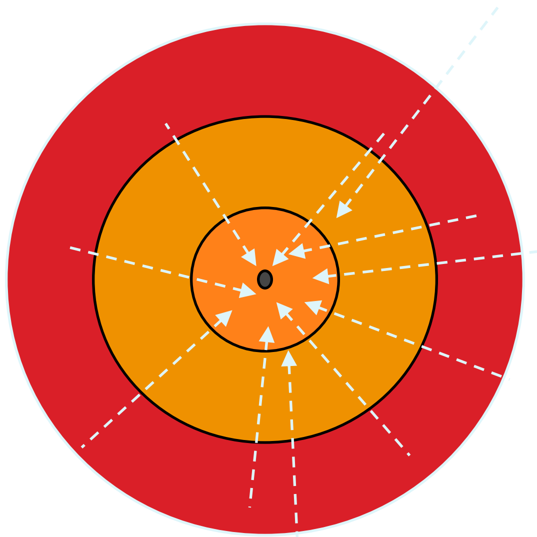
- i podstawiając za **y** $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- otrzymamy:

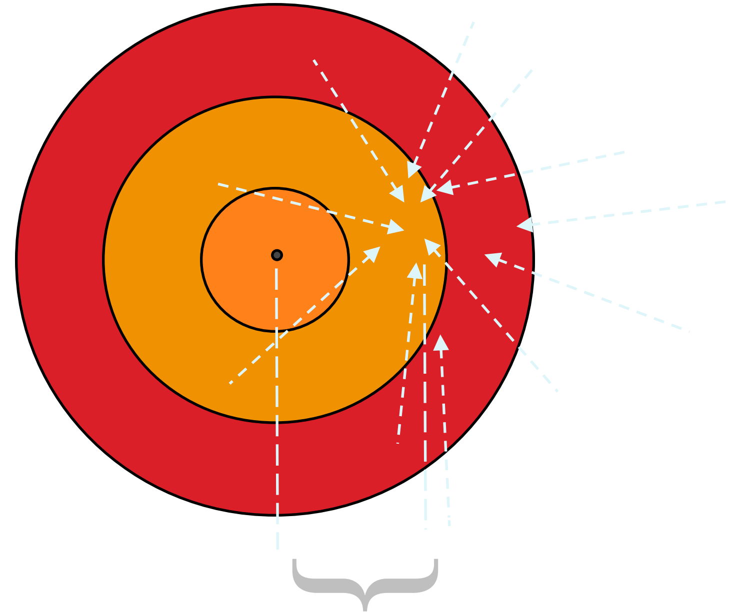
$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$$

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

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

Obciążony i nieobciążony estymator





Estymator nieobciążony

Estymator obciążony

Twierdzenie Gaussa-Markowa

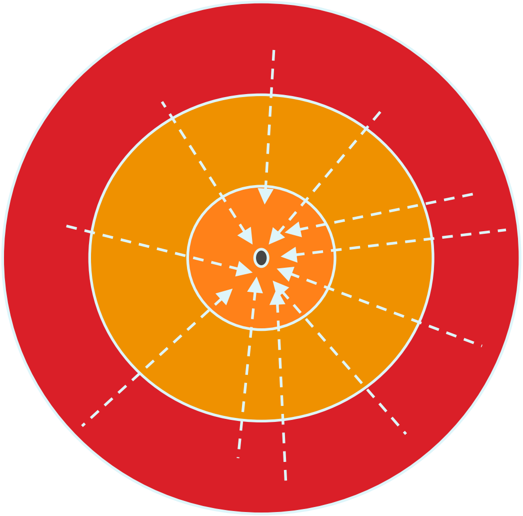
- Wariancja estymatora **b**

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

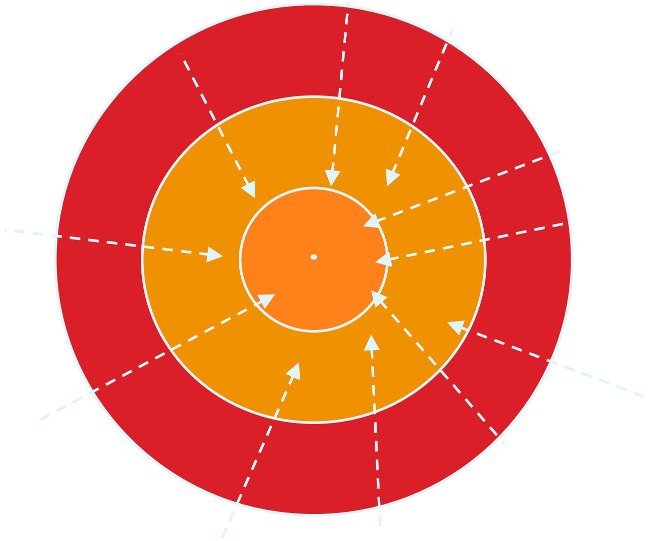
$$= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \Sigma$$

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

Dziękuję za uwagę