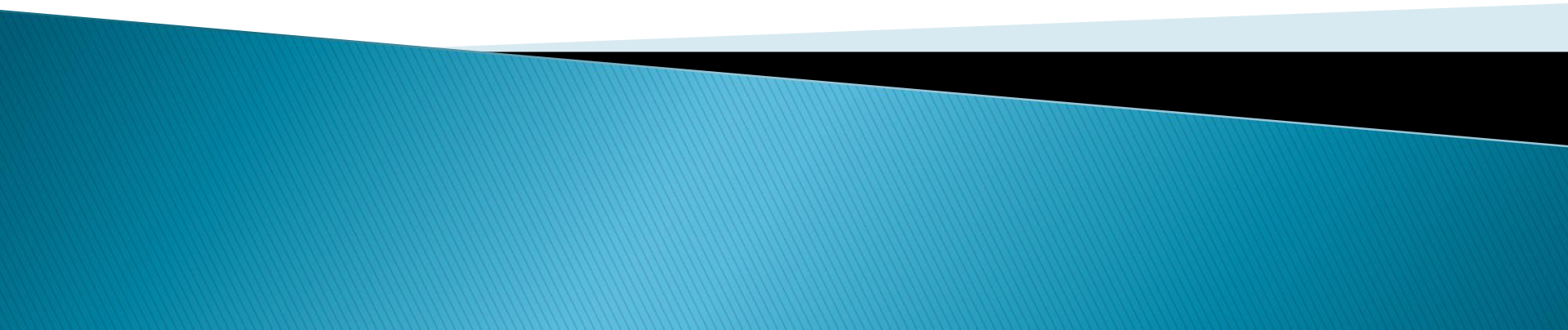


Metoda Najmniejszych Kwadratów **Własności hiperpłaszczyzny**

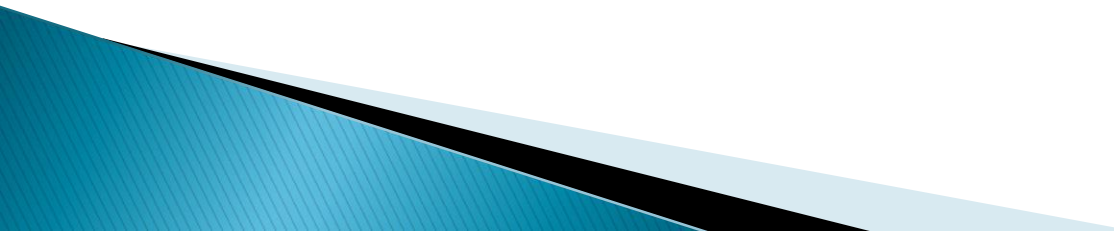
Stanisław Cichocki

Natalia Nehrebecka

Zajęcia 1-3



Plan zajęć

- ▶ 1. Model liniowy
 - Postać modelu liniowego
 - Zapis macierzowy modelu liniowego
 - ▶ 2. Estymacja modelu
 - Wartość teoretyczna (dopasowana)
 - Reszty
 - ▶ 3. MNK - przypadek wielu zmiennych
 - ▶ 4. Własności hiperpłaszczyzny regresji
- 

1. Model liniowy



Postać modelu liniowego

Postać modelu liniowego

teoria ekonomiczna



dane empiryczne



zależności ilościowe między zmiennymi



badanie ekonometryczne

Postać modelu

$$y_i = \beta_1 + x_{2i}\beta_2 + \dots + x_{Ki}\beta_K + \varepsilon_i$$

- y_i - zmienna objaśniana (zależna, endogeniczna),
- $x_2, \dots, x_K$ - zmienne objaśniające (niezależne, egzogeniczne),
- ε_i - błąd losowy,
- $\beta_1, \dots, \beta_K$ - nieznane parametry,
- $i=1, \dots, N$
- i – indeks obserwacji,
- N - liczba obserwacji.

1. Model liniowy



Zapis macierzowy modelu liniowego

Postać modelu - zapis macierzowy

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_y = \underbrace{\begin{bmatrix} 1 & x_{21} & \cdots & x_{K1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{2N} & \cdots & x_{KN} \end{bmatrix}}_X \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}}_{\varepsilon}$$

Stąd równanie macierzowe ma postać:

$$y = X\beta + \varepsilon$$

Zatem...

- ▶ Model liniowy zakłada, że:
 - zależność między analizowanymi zmiennymi jest liniowa (równanie regresji liniowej wyznacza hiperpłaszczyznę regresji)
 - istnieje zależność przyczynowo-skutkowa między zmiennymi (≠korelacja)
 - zmienne objaśniające są przyczyną zmienności zmiennej objaśnianej
 - zależność zwykle wynika z teorii (powinna)
 - pewna część zmienności zmiennej objaśnianej pozostaje niewyjaśniona, bo:
 - nieuwzględnienie pewnych zmiennych objaśniających
 - losowy charakter czynników wpływających na zmienną objaśnianą

Pytanie

- ▶ Który z modeli jest poprawny i dlaczego?
- ▶ Co jest zmienną objaśnianą a co objaśniającą?

$$wydatk_i = \beta_1 + \beta_2 dochód_i + \varepsilon_i$$

$$dochód_i = \beta_1 + \beta_2 wydatk_i + \varepsilon_i$$

2. Estymacja modelu



Estymacja - wstęp

- ▶ Teoria zwykle nie dostarcza informacji nt. wielkości parametrów modelu $(\beta_1, \dots, \beta_K)$.
- ▶ Wielkość nieznanych parametrów należy oszacować (estymować) na podstawie danych empirycznych (próby).
- ▶ Oszacowane wielkości parametrów (estymatory) $(b_1, \dots, b_K)$ są niedokładne (losowe), zależą od próby.

2. Estymacja modelu



Wartość teoretyczna (dopasowana)

Wartość teoretyczna (dopasowana)

- ▶ Wartości dopasowane: wartości zmiennej objaśnianej (y_i) przewidywane na podstawie oszacowanego modelu - regresji liniowej y_i na $x_{2i}, \dots, x_{Ki}$:

$$\hat{y}_i = b_1 + x_{2i}b_2 + \dots + x_{Ki}b_K$$

- ▶ Różnią się od wartości rzeczywistych, bo:
 - ▶ zamiast nieznanymi prawdziwymi wielkościami parametrów ($\beta_1, \dots, \beta_K$) używamy ich estymatorów ($b_1, \dots, b_K$)
 - ▶ pomijamy błąd losowy (ε_i)

2. Estymacja modelu



Reszty

Reszty

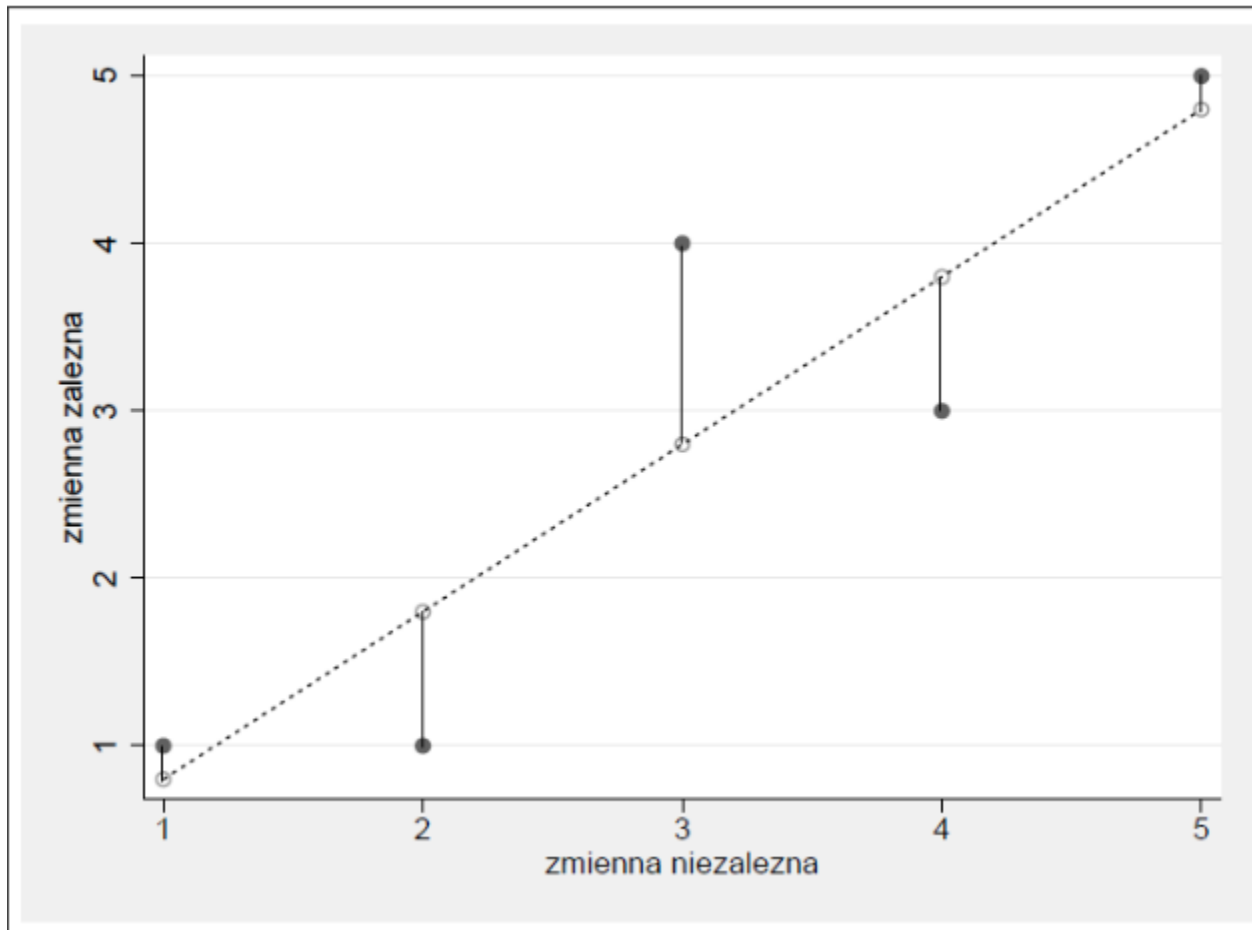
- ▶ Reszty: różnica między wartością rzeczywistą a dopasowaną zmiennej objaśnianej, są to oszacowania (ε_i) :

$$e_i = y_i - \hat{y}_i = y_i - b_1 - x_{2i}b_2 - \dots - x_{Ki}b_K$$

- ▶ Im mniejsza jest odległość wartości rzeczywistych od teoretycznych tym lepszy model 
estymatory parametrów modelu minimalizują sumę odległości y_i od $\hat{y}_i$:

$$\sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N e_i^2$$

Reszty



3. MNK - przypadek wielu zmiennych



Zadanie

- ▶ 1. Zapisz model teoretyczny, model wyestymowany, wartości dopasowane oraz reszty dla modelu linowego zawierającego K zmiennych objaśniających wraz ze stałą

MNK - przypadek wielu zmiennych

- ▶ Suma kwadratów reszt - zapis macierzowy

$$S(b) = \sum_{i=1}^N e_i^2 = \begin{bmatrix} e_1 & \cdots & e_N \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix} = e'e = (y - Xb)'(y - Xb) =$$

$$= y'y - y'Xb - b'X'y + b'X'Xb$$

- ▶ Ponieważ $y'Xb = b'X'y$

- ▶ Zatem: $S(b) = y'y - 2y'Xb + b'X'Xb$

MNK - przypadek wielu zmiennych

- ▶ Warunki pierwszego rzędu

$$\frac{\partial S(b)}{\partial b} = \frac{\partial y' y}{\partial b} - \frac{\partial 2y' Xb}{\partial b} + \frac{\partial b' X' Xb}{\partial b} =$$

$$= -2X' y + 2X' Xb = 0$$

- ▶ bo:

$$\frac{\partial \omega x}{\partial x} = \omega' \quad \frac{\partial x' A x}{\partial x} = (A + A')x$$

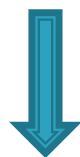
MNK - przypadek wielu zmiennych

- ▶ Układ równań normalnych

$$-2X'y + 2X'Xb = 0$$

$$X'Xb = X'y \quad / (X'X)^{-1}$$

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



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

MNK - przypadek wielu zmiennych

- ▶ Rozwiązanie układu równań normalnych :

$$X' Xb = X' y$$

- ▶ istnieje o ile macierz X ma pełny rząd kolumnowy, tzn. jej kolumny są liniowo niezależne, wtedy istnieje $(X'X)^{-1}$.
- ▶ MNK nie da się oszacować modelu w którym:
 - $(K > N)$ liczba zmiennych (parametrów) przekracza liczbę obserwacji.

4. Własności hiperpłaszczyzny regresji



Własności hiperpłaszczyzny regresji

1. $X'e = 0$

2. $\hat{y}'e = 0$

► Dodatkowo dla modelu ze stałą:

3. $\sum_{i=1}^N e_i = 0$

4. $\bar{y} = \bar{\hat{y}}$

Pytania teoretyczne



Pytania teoretyczne

1. Zapisać model liniowy. Podać interpretację poszczególnych elementów tego modelu.
2. Podać wzajemne relacje między wartościami obserwowanymi zmiennej zależnej, oszacowaniami parametrów, wartościami dopasowanymi i resztami.
3. Wyjaśnij różnicę między parametrami i oszacowaniami parametrów oraz między odchyleniami losowymi i resztami.
4. Skąd bierze się nazwa Metoda Najmniejszych Kwadratów?
5. Wymienić własności hiperpłaszczyzny regresji.

Dziękuję za uwagę