Ekonometria

Natalia Nehrebecka
Stanisław Cichocki

Wykład 2

Plan wykładu

- 1. Model liniowy
  - Postać modelu liniowego – przypadek jednej zmiennej
- 2. Estymacja modelu
  - Wstęp
  - Wartość teoretyczna (dopasowana)
  - Reszty
  - Metoda Najmniejszych Kwadratów
- 3. MNK – przypadek jednej zmiennej

Plan wykładu

- 1. Model liniowy
  - Postać modelu liniowego – przypadek jednej zmiennej
- 2. Estymacja modelu
  - Wstęp
  - Wartość teoretyczna (dopasowana)
  - Reszty
  - Metoda Najmniejszych Kwadratów
- 3. MNK – przypadek jednej zmiennej

Postać modelu liniowego

teoria ekonomiczna ⟺ dane empiryczne

↓

zależności ilościowe między zmiennymi

↓

badanie ekonometryczne

Postać modelu liniowego – równanie regresji w populacji

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_K x_{Ki} + \varepsilon_i$$

- y_i − zmienna objaśniana (*zależna, endogeniczna, regresant*),
- $x_{1i}, \dots, x_{Ki}$− zmienne objaśniające (*niezależne, egzogeniczne, regresory*),
- ε_i − błąd losowy,
- $\beta_0, \beta_{1, \dots,} \beta_K$− nieznane parametry,
- $i = 1, \dots, N$
- i − indeks obserwacji,
- N − liczba obserwacji.

Postać modelu liniowego – przypadek jednej zmiennej

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

- y_i – zmienna objaśniana (*zależna, endogeniczna, regresant*),
- x_{1i} – zmienna objaśniająca (*niezależna, egzogeniczna, regresor*),
- ε_i – błąd losowy,
- β_0, β_1 – nieznane parametry,
- $i = 1, \dots, N$
- i – indeks obserwacji,
- N – liczba obserwacji.

Pytanie

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

$$wydatki_i = \beta_0 + \beta_1 dochód_i + \varepsilon_i \qquad (1)$$

$$dochód_i = \beta_0 + \beta_1 wydatki_i + \varepsilon_i \qquad (2)$$

Przykład

- Związek przyczynowo-skutkowy ≠ korelacja

Przykład

Stwierdzono dodatnią korelację między wielkością spożycia lodów w danym dniu i liczbą utonięć w tym dniu. **Czy po zjedzeniu lodów nie powinno się wchodzić do wody?**

Odpowiedź

Więcej utonięć zdarza się w ciepłe dni (kąpie się wtedy więcej osób). W takie dni jest też większe spożycie lodów.

Czyli: *występuje korelacja między zdarzeniami, ale nie ma między nimi związku przyczynowo-skutkowego*

Zatem…

- Zależność między analizowanymi zmiennymi jest **liniowa**

- Istnieje zależność przyczynowo-skutkowa między zmiennymi (≠korelacja)
  - zależność zwykle wynika z teorii (powinna)

- Pewna część zmienności zmiennej objaśnianej pozostaje niewyjaśniona, bo:
  - nieuwzględnienie pewnych zmiennych objaśniających
  - losowy charakter zmienności zmiennej objaśnianej

Plan wykładu

- 1. Model liniowy
  - Postać modelu liniowego
  - Zapis macierzowy modelu liniowego
- 2. Estymacja modelu
  - Wstęp
  - Wartość teoretyczna (dopasowana)
  - Reszty
  - Metoda Najmniejszych Kwadratów
- 3. MNK – przypadek jednej zmiennej

Estymacja - wstęp

- Teoria zwykle nie dostarcza informacji nt. wielkości parametrów modelu (β_0 , β_1).

$$\Downarrow$$

- Wielkość nieznanych parametrów należy oszacować (*estymować*) na podstawie danych empirycznych (*próby*).

$$\Downarrow$$

- Oszacowane wielkości parametrów (*estymatory*) (b_0, b_1) są niedokładne (*losowe*), zależą od próby.

Wartość teoretyczna (dopasowana)

- Wartości dopasowane: wartości zmiennej objaśnianej (y_i) przewidywane na podstawie oszacowanego modelu - regresji liniowej y_i na x_{1i}:

$$\hat{y}_i = b_0 + b_1 x_{1i}$$

- Różnią się od wartości rzeczywistych, bo:
  - zamiast nieznanych prawdziwych wielkości parametrów (β_0, β_1) używamy ich estymatorów (b_0, b_1)
  - pomijamy błąd losowy (ε_i)

Reszty

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

$$e_i = y_i - \hat{y}_i = y_i - b_0 - x_{1i}b_1$$

- Zależność między resztami, obserwacjami i oszacowaniami parametrów:

$$y_i = \hat{y}_i + e_i = b_0 + b_1 x_{1i} + e_i$$

Estymatory i reszty

- **Estymatory** (b_0, b_1) są to oszacowania (β_0, β_1), ale nie są im równe

- **Reszty** (e_i) są to oszacowania (ε_i), ale nie są im równe

Przykład

- Analizujemy wydatki na żywność w gospodarstwach małżeństw pracowniczych z dwójką dzieci w zależności od dochodu

$$wydatki_i = \beta_0 + \beta_1 doch\acute{o}d_i + \varepsilon_i$$

Przykład

- $$\widehat{wydatki}_i = 463 + 0{,}08 \cdot dochód_i$$

- Dla pierwszej jednostki w badaniu:

Id	Wydatki	Dochód
1	639,1	890,6

$$\hat{y}_1 = 463 + 0{,}08 \cdot 890{,}6 = 534{,}2$$

$$e_1 = 639{,}1 - 534{,}2 = 104{,}9$$

Metoda Najmniejszych Kwadratów

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

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

Metoda Najmniejszych Kwadratów

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

- Funkcja ta jest **ciągła** i **różniczkowalna** dla wszystkich e_i, dzięki czemu można znaleźć jej minimum względem wielkości parametrów poprzez rozwiązanie standardowych warunków pierwszego rzędu.

Pytanie

- Jaką znasz inną funkcję odległości?
- Dlaczego trudno jest ją stosować w procesie estymacji?

Plan wykładu

- 1. Model liniowy
  - Postać modelu liniowego
  - Zapis macierzowy modelu liniowego
- 2. Estymacja modelu
  - Wstęp
  - Wartość teoretyczna (dopasowana)
  - Reszty
  - Metoda Najmniejszych Kwadratów
- 3. MNK – przypadek jednej zmiennej

Zadanie

- Zapisz model teoretyczny, wartości dopasowane oraz reszty dla modelu linowego zawierającego jedną zmienną objaśniająca i stałą

MNK dla modelu z jedną zmienną

- Model teoretyczny:

$$y_i = \beta_0 + x_{1i}\beta_1 + \varepsilon_i$$

- Wartość dopasowana (teoretyczna):

$$\hat{y}_i = b_0 + x_{1i}b_1$$

- Reszta:

$$e_i = y_i - \hat{y}_i = y_i - b_0 - x_{1i}b_1$$

MNK dla modelu z jedną zmienną

- Oszacowania b_0 i b_1 powinny być dobrane tak, by suma kwadratów reszt była jak najmniejsza

$$S(b_0, b_1) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - b_0 - x_i b_1)^2 =$$

$$= \sum_{i=1}^{N} \left(y_i^2 - 2y_i b_0 - 2y_i b_1 x_i + 2b_0 b_1 x_i + b_0^2 + b_1^2 x_i^2 \right)$$

MNK dla modelu z jedną zmienną

- Policz pochodne cząstkowe względem parametrów b_0 i b_1 powyższego równania i przyrównaj je do zera

$$\begin{cases} \dfrac{\partial S(b_0, b_1)}{\partial b_0} = 0 \\ \dfrac{\partial S(b_0, b_1)}{\partial b_1} = 0 \end{cases}$$

MNK dla modelu z jedną zmienną

- Licząc pochodne dla poszczególnych równań uzyskujemy układ równań zwany **układem równań normalnych**

$$\begin{cases} \sum_{i=1}^{N}[-2y_i + 2b_0 + 2b_1x_i] = 0 \\ \sum_{i=1}^{N}\left[-2y_ix_i + 2b_0x_i + 2b_1x_i^2\right] = 0 \end{cases}$$

MNK dla modelu z jedną zmienną

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{\frac{\sum_{i=1}^{N} y_i x_i}{N} - \overline{yx}}{\frac{\sum_{i=1}^{N} x_i^2}{N} - \bar{x}^2}$$

MNK dla modelu z jedną zmienną

- Przypomnij wzór na wariancję (s^2_x) i kowariancję (s_{xy}) empiryczną.

$$b_0 = \bar{y} - b_1\bar{x}$$

$$b_1 = \frac{S_{yx}}{S_x^2}$$

Przykład

- Estymacja modelu wyjaśniającym wielkość wydatków na żywność dochodami gospodarstwa

$$wydatki_i = \beta_0 + \beta_1 dochód_i + \varepsilon_i$$

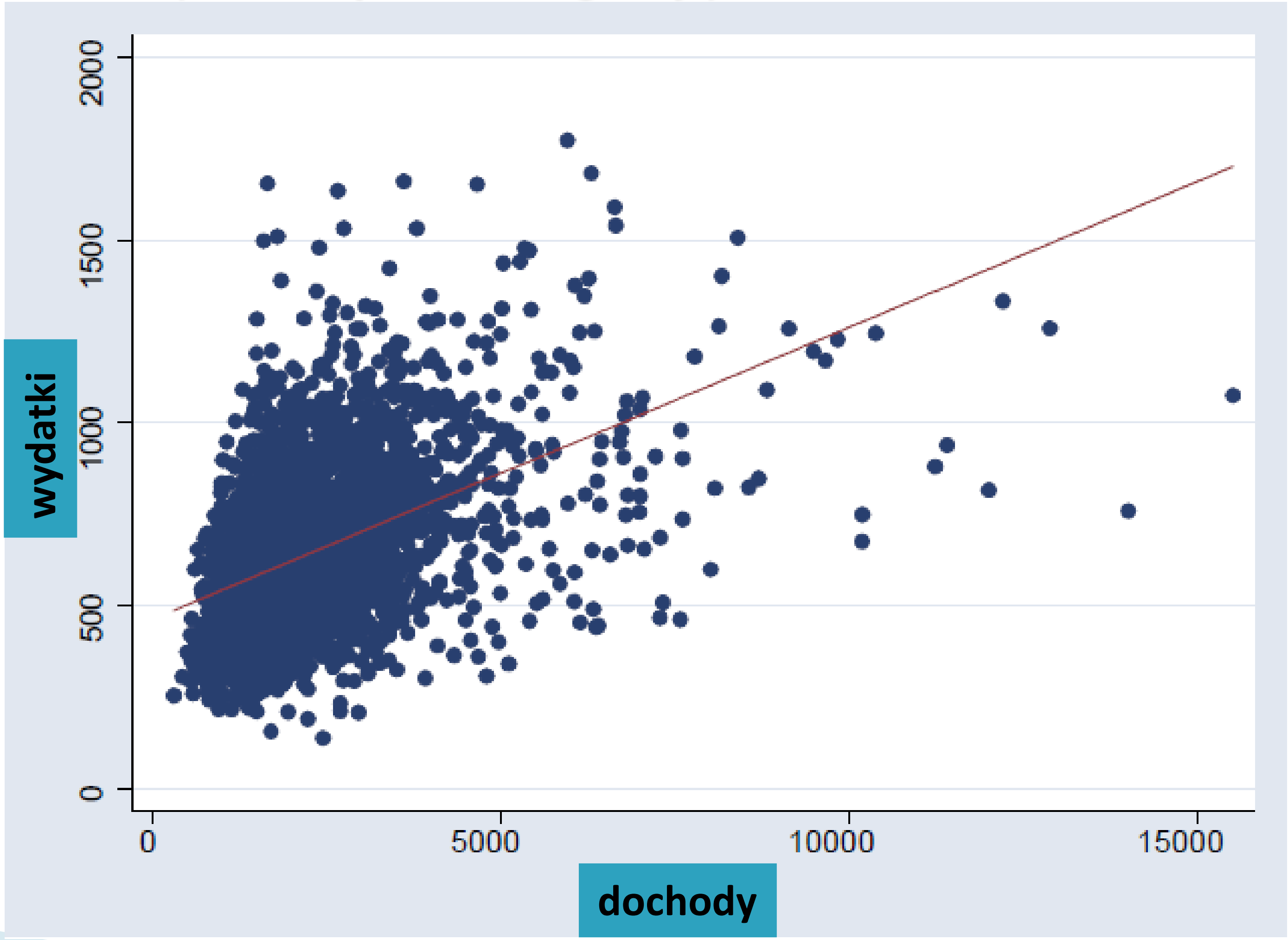
Zmienna	Średnia	Wariancja
Wydatki	644,18	46737
Dochód	2262,34	1584300

- Kowariancja empiryczna między zmiennymi jest równa 126211.

$$b_1 = \frac{126211}{1584300} = 0{,}079664$$

$$b_0 = 644{,}18 - 0{,}079664 \cdot 2262{,}34 = 463{,}95$$

Zależność wydatków na żywność od dochodu (dane i prosta regresji)



Dziękuję za uwagę