Metody Ekonometryczne
Metoda Najmniejszych Kwadratów

Jakub Mućk
Szkoła Główna Handlowa w Warszawie

Sprawy organizacyjne

Literatura podstawowa

- Greene W.H., (2011), *Econometric Analysis*, Prentice Hall.

Literatura uzupełniająca

- Hill R.C., Griffiths W. E. i Lim G.C, (2011), *Principle of Econometrics*, Willey.
- Wooldridge J. M., (2012), *Introductory Econometrics: A Modern Approach*, Cengage Learning.

Oprogramowanie

- R.

- **Raport** opisujący badanie ekonometryczne: 50%.
- **Egzamin**: 50%.
- **Brak raportu oznacza ocenę niedostateczną**.

1. Metoda najmniejszych kwadratów.
2. Uogólniona metoda najmniejszych kwadratów.
3. Metoda zmiennych instrumentalnych.
4. Modele wielorównaniowe.

Wprowadzenie

Ekonometria to zbiór metod statystycznych i matematycznych pozwalających na empiryczną weryfikację teorii ekonomicznej.

Inaczej

Ekonometria pozwala na pomiar siły (istotności) i kierunku zjawisk i procesów ekonomicznych.

Załóżmy, że zmienna y jest funkcją x_1, x_2, ..., x_k, t. że

$$y = \mathcal{F}\left(x_1, x_2, \ldots, x_k\right) \tag{1}$$

Zauważmy, że równanie (1) określa funkcję **deterministyczną**. Wprowadźmy, element losowy, tj. ε:

$$y = f\left(x_1, x_2, \ldots, x_k, \varepsilon\right) \tag{2}$$

Szczególnym przypadkiem (1) jest liniowa postać. Wtedy mowa o **modelu regresji liniowej**:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \tag{3}$$

gdzie:

- y – zmienna objaśniana.
- $x_1, \ldots, x_k$ – zmienne objaśnianające, (egzogeniczne).
- $\beta_1, \ldots, \beta_k$ – parametry strukturalne modelu.
- β_0 – wyraz wolny.
- ε – składnik losowy.

- **Dane eksperymentalne i nieeksperymentalne** (*experimental and nonexperimental data*).
- Dane przekrojowe (*cross-section data*), szeregi czasowe (*time series*) oraz dane panelowe (*panel data*).
- Dane mikroekonomiczne oraz makroekonomiczne.
- Przypływy (*flow*) i stany (*stock*).
- Dane ilościowe (*quantitative*) i jakościowe (*qualitative*)

Konsekwencje

Rodzaj danych ma znaczenie w wyborze zarówno i) postaci funkcyjnej oraz ii) metody estymacji.

1. Sformułowanie celu badawczego i/lub hipotez badawczych.
2. Wybór modelu ekonomicznego.
3. Kolekcjonowanie danych i wybór odpowiednich metod ekonometrycznych lub statystycznych.
4. Szacowanie nieznanych parametrów w celu odpowiedzi na kluczowe pytania badawcze lub realizacji celów badawczych.
5. Weryfikacja zasadności założeń wykorzystanych w procesie estymacji.

Rachunek prawdopodobieństwa i statystyka matematyczna

Zmienna losowa (*random variable*)

jest zmienną, której wartości są nieznane dopóki nie jest obserowana.

Zmienna losowa (*random variable*)

jest zmienną, której wartości są nieznane dopóki nie jest obserowana.

- **Dyskretna zmienna losowa** – przyjmuje wartości z ograniczonego i przeliczalnego zbioru potencjalnych wyników.
- **Binarna zmienna losowa** – przyjmuje wyłącznie wartości 0 lub 1.
- **Ciagła zmienna losowa** – przyjmuje dowolne wartości (lub z pewnego zakresu).

- **Rozkład prawdopodbieństwa (*probability density function, pdf*)** dla zmiennej losowej opsiuje potencjalne wyniki oraz towarzyszące im prawdopodobieństwa.
- Rozkład prawdopodbieństwa *pdf* dla dyskretnej zmiennej losowej X :

$$f(x) = P(X = x), \tag{4}$$

  i $\sum_i^k f(x_i) = 1$.
- Ponieważ dla ciągłej zmiennej losowe $P(X = x) = 0$ *pdf* wykorzystuje się funkcję gęstości, która jest określona dla pewnego zbioru.

$$P(a \leq X \leq b) = \int_a^b f(x)dx, \tag{5}$$

  i $\int_{-\infty}^{\infty} f(x)dx = 1$.

- **Dystrybuanta rozkładu prawdopodobieństwa (*cumulative distribution function, cdf*)** jest alternatywną miarą opisu potencjalnych wyników wraz z towarzyszącym im prawdopodobieństwami. Dla dowolnej stałe x dystrybuanta jest zdefiniowana następująco:
$$F(x) = P(X \leq x). \tag{6}$$
  - W przypadku dyskretnych zmiennych losowych, dystrybuanta *cdf* to po prostu suma *pdf* dla wszystkich x_i (spełniających odpowiedni warunek).
  - Dla ciągłej zmiennej losowej, $F(x)$ jest powierzchnią *pdf na lewo* od punktu x.
- Dystrybuanta rozkładu prawdopodbieństwa jest przydatna w obliczaniu prawdopodob
W szczególności,

$$P(X > x) = 1 - P(X \leq x) = 1 - F(x), \tag{7}$$
$$P(a < X \leq b) = F(b) - F(a). \tag{8}$$

- **Rozkład łączny (*joint probability density function*** zawiera informacje związaną z potencjalnymi wynikami (przynjamniej dwóch) zmiennych losowych oraz towarzyszącym im prawdopodobieństwom. Dla zmienncyh dyskretnych:

$$f(x, y) = P(X = x, Y = y), \tag{9}$$

i dla zmiennych ciągłych:

$$P(a \leq X \leq b, c \leq X \leq d) = \int_a^b \int_c^d f(x, y) dy dx. \tag{10}$$

- **Rozkład brzegowy** ***marginal distribution*** pozwala uzyskać informację o indywidualnym rozkładzie prawdopodbieństwa zmiennej losowej (ponizej dla X):

$$f_X(x) = P(X = x) = \sum_y f(x, y). \tag{11}$$

- **Warunkowy rozkład prawdopodobieństwa (*conditional distribution*)** jest rozkładem prawdopodbieństwa zmiennej Y w przypadku, gdy znana X jest znane:

$$f(x|y) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}. \tag{12}$$

- Zmienne losowe są niezależne wtedy i tylko wtedy gdy rozkłąd łączny prawdopodobieńs jest równy iloczynowi indywidualnych rozkładów prawdopodobieństwa *pdfs*. Przykładowo, dla dwóch zmiennych losowych:

$$f(x,y) = f_X(x) f_Y(y). \tag{13}$$

- **Wartość oczekiwana (*expected value, expectation*)** zmiennej losowej X jest średnią ważoną (rozkładem prawdopodobieństwa) wszystkich potencjalnych wyników X.
- Wartość oczekiwana dla zmiennej dyskretnej X:

$$\mathbb{E}\left(X\right) = \sum_{i=1}^{N} X_i f\left(x_i\right), \tag{14}$$

gdzie $f\left(x\right)$ jest rozkładem prawdopodobieństwa$X$.

- Wartość oczekiwana jest również nazywana **średnią z populacji (*population mean*)** ($\neq$ średnia z próby (*sample average*)).
- Wartość oczekiwana dla zmiennej ciągłej X:

$$\mathbb{E}\left(X\right) = \int_{-\infty}^{\infty} x f(x) dx. \tag{15}$$

- Własności wartości oczekiwanej
  1. Dla dowolnej stałej c:

$$\mathbb{E}\left(c\right) = c. \tag{16}$$

2. Dla dowolnych stałych a i b:

$$\mathbb{E}\left(aX+b\right)=a\mathbb{E}\left(X\right)+b. \tag{17}$$

3. Dla dowolnych stałych $a_1, \ldots, a_k$ i zmiennych losowych $X_1, \ldots X_k$:

$$\mathbb{E}\left(\sum_i^k a_i X_i\right)=\sum_i^k a_i \mathbb{E}\left(X_i\right), \tag{18}$$

i jeżeli $a_i = 1$ dla wszystkich i wtedy wartość oczekiwana sumy jest sumą wartości oczekiwanych.

4. Dla funkcji tworzącej nową zmienną losową $g(\cdot)$:

$$\mathbb{E}\left(g(X)\right)=\sum_i^k g(x_i) f_x(x_i), \tag{19}$$

i w przypadku ciągłej zmiennej losowej:

$$\mathbb{E}\left(g(X)\right)=\int_{-\infty}^{\infty} g(x_i) f_x(x_i), \tag{20}$$

- **Warunkowa wartość oczekiwana (*conditional expected value*)** jest wartością oczekiwaną X w przypadku gdy Y jest znane. Dla dyskretnych zmiennych losowych:

$$\mathbb{E}(X|Y=y) = \sum_{i=1}^{k} x_i f(x_i, y). \tag{21}$$

- **Wariancja (*variance*)** mierzy stopień zmienności (*variability*) zmiennej losowej:

$$Var(X) = \mathbb{E}\left[(X - \mu)^2\right], \tag{22}$$

where $\mathbb{E}(X) = \mu$.

- Wariancja jest zazwyczaj oznaczana jako σ^2 (lub σ_X^2 for X).
- Alternatywnie, wariancję zmiennej losowej można zapisać jako:

$$\begin{aligned} Var(X) &= \mathbb{E}\left(X^2 - 2X\mu + \mu^2\right) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) - \mu^2 \\ &= \mathbb{E}\left(X^2\right) - \mu^2. \end{aligned}$$

- Dla dowolnych stałych wartości a i b:

$$Var(aX + b) = a^2 Var(X). \tag{23}$$

- **Odchylenie standardowe (*standard deviation*, sd)**jest pierwiastkiem wariancji

$$sd(X) = \sqrt{Var(X)} \tag{24}$$

- Dla dowolnych stałych a i b:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + cov(X, Y). \tag{25}$$

- **Kowariancja (*covariance*)** mierzy zależność pomiedzy zmiennymi losowymi:

$$cov\left(X,Y\right)=\mathbb{E}\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right], \tag{26}$$

where $\mathbb{E}(Y)=\mu_Y$ i $\mathbb{E}(X)=\mu_X$.

- Kowariancja pomiędzy X i Y jest również oznaczana jako σ_{XY}.
- Kowariancja może zostać alternatywnie zapisana jako:

$$\begin{aligned} cov\left(X,Y\right) &= \mathbb{E}\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]\mathbb{E}\left[X\left(Y-\mu_Y\right)\right] \\ &= \mathbb{E}\left[\left(X-\mu_X\right)Y\right]=\mathbb{E}\left(XY\right)-\mu_X\mu_Y \end{aligned}$$

- Gdy X i Y są niezależne to $cov(X,Y)=0$ i $\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)$.
- Na podstawie kowariancji trudno ocenić siłę zależności pomiędzy zmiennymi losowymi. **Korelacja *correlation* (ρ)** uwzględnia potencjalne różnice w wariancji zmiennych losowych:

$$\rho=\frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}, \tag{27}$$

and $\rho\in<0,1>$.

- Dlasze własności wariancji:

$$\begin{aligned} Var\left(aX+bY\right) &= a^2Var\left(X\right)+b^2Var\left(Y\right)+2abCov\left(X,Y\right), && (28) \\ Var\left(aX-bY\right) &= a^2Var\left(X\right)+b^2Var\left(Y\right)-2abCov\left(X,Y\right). && (29) \end{aligned}$$

- Estymator $\hat{\theta}$ jest funkcją zmiennej losowej (zmiennych losowych) oraz próby (danych empirycznych) mającą na celu pomiar (empiryczny) parametru θ.
- Nieobciążoność (*unbiasedness*) estymatora:

$$\mathbb{E}\left(\hat{\theta}\right) = \theta \qquad \text{lub} \quad Bias\left(\hat{\theta}\right) = \mathbb{E}\left(\hat{\theta}\right) - \theta = 0, \tag{30}$$

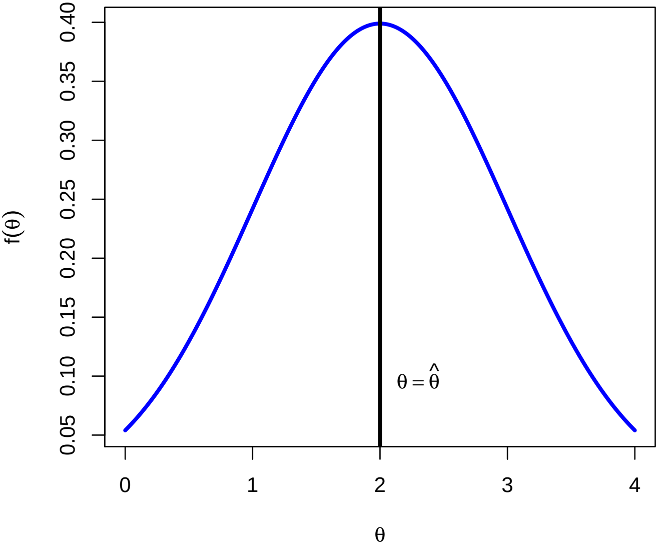
  gdzie $Bias\left(\hat{\theta}\right)$ to obciążenie estymatora.
- Wariancja estymatora $\hat{\theta}$:

$$Var\left(\hat{\theta}\ \right) = \mathbb{E}\left[\left(\hat{\theta} - \mathbb{E}\left(\hat{\theta}\right)\right)^2\right], \tag{31}$$

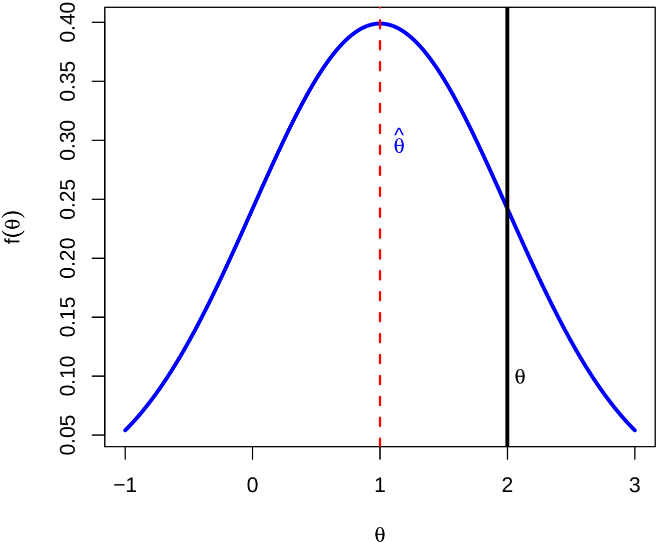
  wariancja może służyć jako miara efektywności estymatora.
- Zgodność (*consistency*) estymatora:

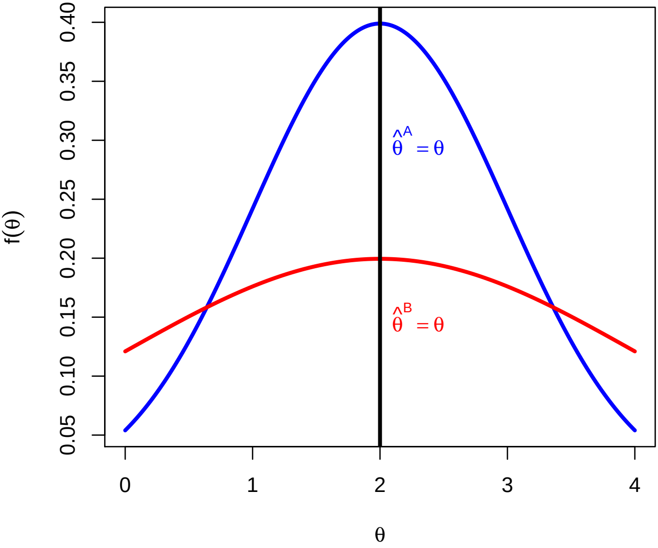
$$\text{plim}_{n\to\infty}\hat{\theta}_n = \theta, \tag{32}$$

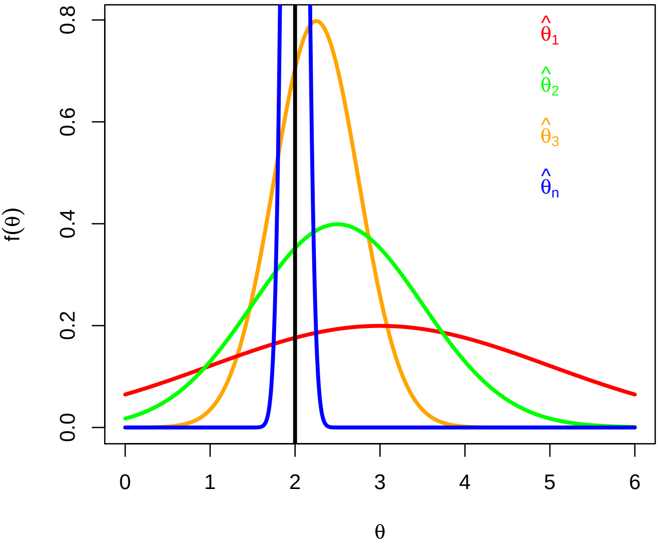
  gdzie $\text{plim}_{n\to\infty}$ to granica stochastyczna.
- Nieobciążoność **nie implikuje** zgodności
  i zgodność **nie implikuje** nieobciążoności.


f(θ)
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0
1
2
3
4
θ
θ = θ̂

Przykład - obciążenie



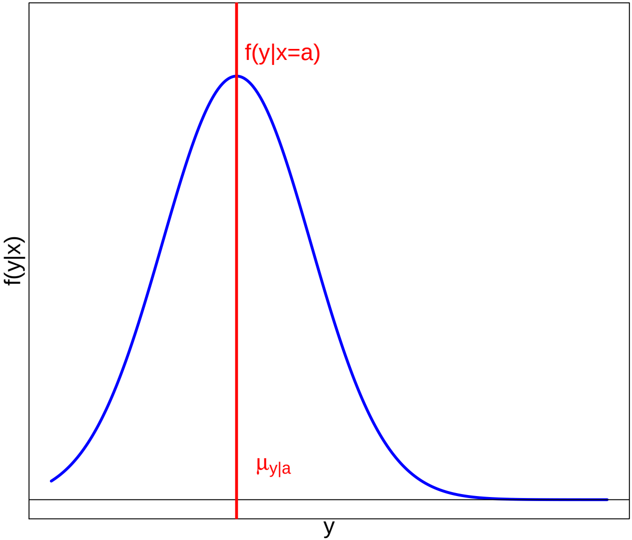

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
f(θ)
0
1
2
3
4
θ
$\hat{\theta}^A = \theta$
$\hat{\theta}^B = \theta$

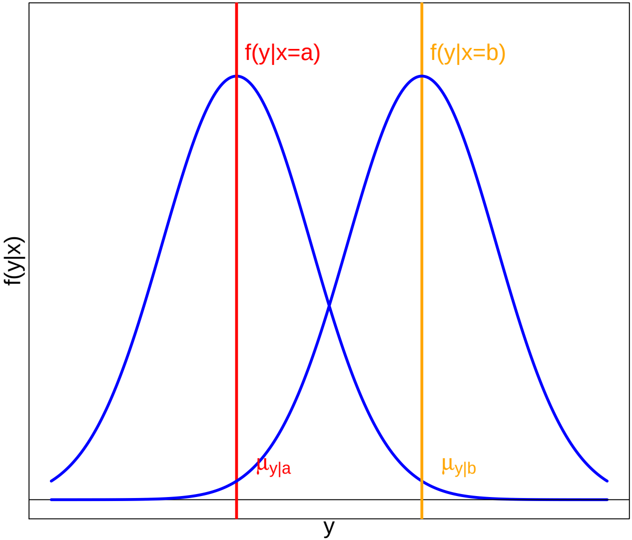


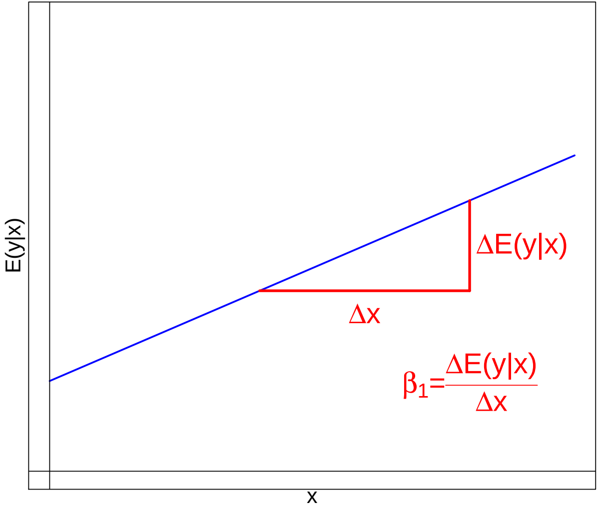
$\hat{\theta}_1$
$\hat{\theta}_2$
$\hat{\theta}_3$
$\hat{\theta}_n$
f(θ)
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
θ

Regresja prosta – intuicja

Punkt wyjścia – rozkład warunkowy zmiennej Y względem zmiennej X.




f(y|x=a)
f(y|x=b)
f(y|x)
μy|a
μy|b
y



Regresja prosta:

$$\mathbb{E}(y|x) = \beta_0 + \beta_1 x$$

Estymator MNK

Zapis ogólny:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon. \tag{33}$$

Zapis macierzowy

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{34}$$

gdzie

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \ldots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \ldots & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \ldots & x_{n,k} \end{bmatrix}_{n\times(k+1)},$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1},$$

k – liczba zmiennych objaśniających; n – liczba obserwacji.

- β oznacza **prawdziwą (nieznaną)** wartość wektora parametru a $\hat{\beta}$ jego oszacowanie (szacunek), tj. wynik estymacji.
- $\mathbf{y}$ jest obserwowaną (empriczyną) wartością zmiennej objaśnanej, a $\hat{\mathbf{y}}$ wartością teoretyczną, a więc wynikającą z oszacowań, tj.:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}. \tag{35}$$

- ε to składnik losowy. Jego realizacje stanowi $\mathbf{e}$, czyli tzw. reszty (składnik losowy):

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}. \tag{36}$$

Dla następującego modelu liniowego:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{37}$$

załóżmy, że

1. **Brak współliniowości:** $rz(\mathbf{X}) = k + 1 \leq n$;
   macierz obserwacji $\mathbf{X}$ jest pełnego rzędu kolumnowego.
2. Zmienne x_i są **ściśle egzogeniczne**, a więc są niezależne od składnika losowego, tj. $\mathbb{E}(\mathbf{X}\varepsilon) = 0$.
3. $\mathbb{E}(\varepsilon) = 0$.
4. **Sferyczna macierz wariancji-kowariancji składnika losowego:**
   $Var(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon^T) = I\sigma^2$.
5. $\varepsilon_i \sim N(0, \sigma^2)$.

Kluczowe są założenia (1)-(4). Założenie o normalności składnika losowego jest pomocnicze.

- Załóżmy, że badane zjawisko można opisać modelem postaci [lub *prawdziwy proces generujący zmienną y jest następujący*]:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{38}$$

- Nieznane parametry można uzyskać przy pomocy **Metody Najmniejszych Kwadra (*OLS – ordinary least squares*)**. Idea tej metody polega na znalezieniu takich wartości nieznanego wektora parametrów β, który minimalizują sumę kwadratów reszt, czyli różnic pomiędzy wartościami obserwowanymi a teoretycznymi:

$$\hat{\beta}^{OLS} = \arg\min_{\beta} \mathbf{e}^T\mathbf{e}, \tag{39}$$

gdzie $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}$.

- Przyjmijmy sumę kwadratów reszt (*SSE - error sum of squares*) jako funkcję szukanego wektora parametrów β, którą chcemy minimalizować:

$$SSE(\beta) = \mathbf{e}^T\mathbf{e} = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta), \tag{40}$$

- Po przemnożeniu macierzy w (40), otrzymujemy:

$$SSE(\beta) = \mathbf{y}\mathbf{y}^T - 2\mathbf{y}^T\mathbf{X}\beta + \beta^T\mathbf{X}^T\mathbf{X}\beta, \tag{41}$$

- Zróżniczkujmy wyrażenie (41) po β, wtedy otrzymamy:

$$\frac{\partial SSE(\beta)}{\partial \beta} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\beta, \tag{42}$$

- Przyrównując pierwszą pochodną, tj. (42), otrzymujemy układ równań

$$\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\beta, \tag{43}$$

Ostatecznie, wykorzystując założenie o pełnym rzędzie kolumnowym macierzy $\mathbf{X}$:

$$\hat{\beta}^{OLS} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}. \tag{44}$$

- Model regresji prostej (alternatywnej notacji):

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, N, \tag{45}$$

gdzie i oznacza indeks obserwacji.

- Estymator MNK dla parametrów równania (45) można wyrazić następująco:

$$\hat{\beta}_0^{OLS} = \bar{y} - \hat{\beta}_1^{OLS} \bar{x}, \tag{46}$$

$$\hat{\beta}_1^{OLS} = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^N (x_i - \bar{x})^2}. \tag{47}$$

- Interpretacja?

Estymator wariancji-kowariancji wektora oszacowań I

SGH

- Ogólny wzór dla estymatora macierzy wariancji-kowariancji oszacowań ($\hat{\beta}^{OLS}$):

$$Var(\hat{\beta}^{OLS}) = \mathbb{E}\left[\left(\hat{\beta}^{OLS} - \beta\right)\left(\hat{\beta}^{OLS} - \beta\right)^T\right]. \tag{48}$$

- Zauważmy, że

$$\hat{\beta}^{OLS} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\left(\mathbf{X}\beta + \varepsilon\right) = \beta + \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon, \tag{49}$$

- wtedy

$$\begin{aligned}
Var(\hat{\beta}^{OLS}) &= \mathbb{E}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\right)^T\right] \\
&= \mathbb{E}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\varepsilon^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right] \\
&= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbb{E}\left[\varepsilon\varepsilon^T\right]\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}
\end{aligned}$$

- Następnie korzystając z założenie o **sferyczności macierzy wariancji-kowariancji składnika losowego**, tj. $\mathbb{E}(\varepsilon\varepsilon^T) = \sigma^2 I$, można uprościć wzór na estymator wariancji kowariancji oszacowań do:

$$Var(\hat{\beta}^{OLS}) = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}. \tag{50}$$

- W przypadku MNK, możemy wyznaczyć estymator macierzy wariancji-kowariancji dla parametrów $\hat{\beta}^{OLS}$ korzystając z szacunku wariancji wariancji składnika losowego ($\mathbb{S}^2_\varepsilon$):
$$Var(\hat{\beta}^{OLS}) = \mathbb{S}^2_\varepsilon (\mathbf{X}^T\mathbf{X})^{-1}, \tag{51}$$
gdzie
$$\mathbb{S}^2_\epsilon = \frac{\mathbf{e}^T\mathbf{e}}{n-(k+1)} = \frac{SSE(\hat{\beta}^{OLS})}{df} \tag{52}$$
gdzie $SSE(\hat{\beta}^{OLS})$ to suma kwadratów reszt, a df to **liczba stopni swobody**.
- Element diagonalne macierzy wariancji-kowariancji (oznaczmy jako $\hat{d}_{ii}$), stanowią wariancję estymowanych parametrów. Zatem **błąd szacunku dla i-tego parametru** jest równy:
$$\mathbb{S}(\hat{\beta}_i) = \sqrt{d_{ii}}. \tag{53}$$
- **Względny błąd szacunku**
$$\left|\frac{\mathbb{S}(\hat{\beta}_i)}{\hat{\beta}_i}\right|. \tag{54}$$

Założenia:

1. $rz(\mathbf{X}) = k + 1 \leq n$
2. $\mathbb{E}(\mathbf{X}\varepsilon) = 0$
3. $\mathbb{E}(\varepsilon) = 0$
4. $Var(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon^T) = I\sigma^2$
5. $\varepsilon_i \sim N(0, \sigma^2)$

Twierdzenie Gaussa - Markowa

Estymator $\hat{\beta}$ uzyskany Klasyczną Metodą Najmniejszych Kwadratów jest estymatorem **BLUE** [***best linear unbiased estimator***], tj. zgodnym, nieobciążonym i najefektywniejszy w klasie liniowych estymatorów wektora β.

- nieobciążoność, czyli: $\mathbb{E}(\hat{\beta}^{OLS}) = \beta$
- najefektywniejszy, czyli posiadający najmniejszą wariancję w swojej klasie
- zgodny, czyli: $\text{plim}_{n\to\infty}(\hat{\beta}^{OLS}) = \beta$.

- Niebciążoność estymatora:
$$\mathbb{E}\left(\hat{\beta}^{OLS}\right) = \beta. \tag{55}$$
- Korzystając z tożsamości $y = \mathbf{X}\beta + \varepsilon$ postać estymatora KMNK może zostać zapisana jako:
$$\hat{\beta}^{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\left(X\beta + \varepsilon\right). \tag{56}$$
- Korzystając z faktu, że $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T X = \mathbf{I}$
$$\hat{\beta}^{OLS} = \beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\varepsilon. \tag{57}$$
- Wprowadźmy wartość oczekiwaną:
$$\mathbb{E}\left(\hat{\beta}^{OLS}\right) = \mathbb{E}\left(\beta\right) + \mathbb{E}\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\varepsilon\right], \tag{58}$$
- Wiedząc, że $\mathbb{E}\left(\beta\right) = \beta$ oraz wykorzystując nielosowość zmiennych objaśniających, tj. $\mathbb{E}\left(\mathbf{X}\right) = X$
$$\mathbb{E}\left(\hat{\beta}^{OLS}\right) = \beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}(\varepsilon), \tag{59}$$
a przy założeniu, że $\mathbb{E}(\varepsilon\mathbf{X}) = 0$ otrzymujemy:
$$\mathbb{E}\left(\hat{\beta}^{OLS}\right) = \beta. \tag{60}$$

- Na podstawie wszcześniejszych obliczeń:

$$\hat{\beta}^{OLS} = \beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\varepsilon. \tag{61}$$

- Założmy, że $\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Wtedy

$$\hat{\beta}^{OLS} = \beta + \mathbf{A}\varepsilon, \tag{62}$$

- a więc $\hat{\beta}^{OLS}$ jest liniową funkcją składnika losowego. Zgodnie z założeniami, które przyjeliśmy $\hat{\beta}^{OLS}$ jest zatem estymatorem liniowym.

- Wprowadźmy, inny nieobciążony estymator liniowy parametru β, np. $\hat{\mathcal{B}} = C\mathbf{y}$. Zatem

$$\mathbb{E}(\hat{\mathcal{B}}) = \mathbb{E}\left(C\mathbf{X}\beta + C\varepsilon\right) = \beta \tag{63}$$

- Warto zauważyć, żę $C\mathbf{X} = \mathbf{I}$.
- Wariancja estymatora $\hat{\mathcal{B}}$:

$$Var(\hat{\mathcal{B}}) = \sigma^2 CC^T. \tag{64}$$

- Porównując wariancję, wprowadźmy wyrażenie $D = C - \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$.
- Wariancja estymatora $\hat{\mathcal{B}}$:

$$Var(\hat{\mathcal{B}}) = \sigma^2\left[\left(D + \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\left(D + \left(\mathbf{X}^T\mathbf{X}\right)^{-1}X^T\right)^T\right], \tag{65}$$

- $D\mathbf{X} = 0$ ze względu na tożsamość $C\mathbf{X} = \mathbf{I} = D\mathbf{X} + \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{X}\right)$. Zatem wariancję można zapisać:

$$Var(\hat{\mathcal{B}}) = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1} + \sigma^2 DD^T = Var(\hat{\beta}^{OLS}) + \sigma^2 DD^T. \tag{66}$$

- Zgodność estymatora MNK:

$$\text{plim}_{n\to\infty}\hat{\beta}^{OLS} = \beta \tag{67}$$

- Korzystając z wcześniejszej tożsamości ($\mathbb{E}\left(\hat{\beta}^{OLS}\right) = \beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}(\varepsilon)$):

$$\text{plim}_{n\to\infty}\hat{\beta}^{OLS} = \beta + \text{plim}_{n\to\infty}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}(\varepsilon). \tag{68}$$

- Mnożymy wyrażenie przez jeden, tj. $1 = 1/n \times n$:

$$\text{plim}_{n\to\infty}\hat{\beta}^{OLS} = \beta + \text{plim}_{n\to\infty}\left(\frac{1}{n}\mathbf{X}^T\mathbf{X}\right)^{-1}\frac{1}{n}\mathbf{X}^T\mathbb{E}(\varepsilon). \tag{69}$$

- Korzystając z egzogeniczności:

$$\text{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^T\mathbb{E}(\varepsilon) = 0, \tag{70}$$

- Ponadto, zauważmy, że o ile wyrażenie $\text{plim}_{n\to\infty}(\mathbf{X}^T\mathbf{X})$ trudno ograniczyć to $\text{plim}_{n\to\infty}(1/n\mathbf{X}^T\mathbf{X})$ można ograniczyć przez pewną wartość $\mathcal{C}$. Wtedy:

$$\text{plim}_{n\to\infty}\hat{\beta}^{OLS} = \beta + \mathcal{C}\times 0 = \beta. \tag{71}$$