

REGRESJA LINIOWA

EKONOMETRIA WNE 2022/23

Sebastian Zalas

Uniwersytet Warszawski
s.zalas@uw.edu.pl

26 października 2023

EKONOMETRIA

- Teoria ekonomii + Metody statystyczne + Dane ekonomiczne
- Najczęściej stosowane narzędzie ekonometryczne: **Metoda Najmniejszych Kwadratów** (MNK), również znana jako **regresja**.
ang. Ordinary Least Squares, (OLS).

EKONOMETRIA

- Teoria ekonomii + Metody statystyczne + Dane ekonomiczne
- Najczęściej stosowane narzędzie ekonometryczne: **Metoda Najmniejszych Kwadratów** (MNK), również znana jako **regresja**.
ang. Ordinary Least Squares, (OLS).

MODEL LINIOWY

$$Y \sim X$$

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y - zmienna zależna
- X - zmienna niezależna
- β_0 - stała, parametr modelu
- β_1 - współczynnik opisujący zależność X i Y , parametr modelu
- ε - błąd losowy

MODEL LINIOWY

$$Y \sim X$$

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y - zmienna zależna
- X - zmienna niezależna
- β_0 - stała, parametr modelu
- β_1 - współczynnik opisujący zależność X i Y , parametr modelu
- ε - błąd losowy

MODEL LINIOWY

$$Y \sim X$$

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y - zmienna zależna
- X - zmienna niezależna
- β_0 - stała, parametr modelu
- β_1 - współczynnik opisujący zależność X i Y , parametr modelu
- ε - błąd losowy

MODEL LINIOWY

$$Y \sim X$$

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y - zmienna zależna
- X - zmienna niezależna
- β_0 - stała, parametr modelu
- β_1 - współczynnik opisujący zależność X i Y , parametr modelu
- ε - błąd losowy

MODEL LINIOWY

$$Y \sim X$$

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- znamy Y oraz X
- nie znamy $\beta_0, \beta_1 \Rightarrow$ jak je znaleźć?
- nie możemy ich wyznaczyć, ale możemy **oszacować**

MODEL LINIOWY

$$Y \sim X$$

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- znamy Y oraz X
- nie znamy $\beta_0, \beta_1 \Rightarrow$ **jak je znaleźć?**
- nie możemy ich wyznaczyć, ale możemy **oszacować**

MODEL LINIOWY

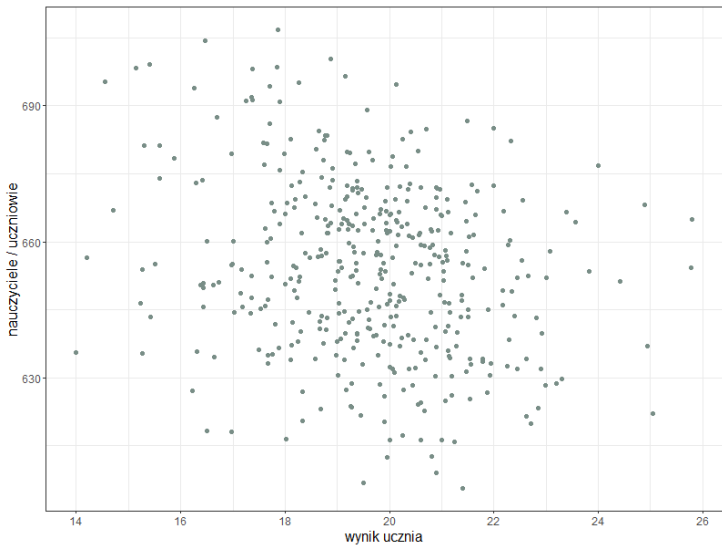
$$Y \sim X$$

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$$

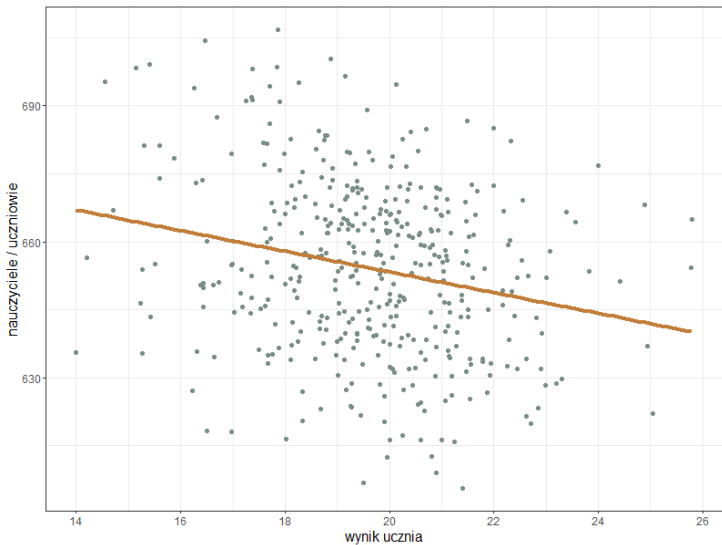
$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- znamy Y oraz X
- nie znamy $\beta_0, \beta_1 \Rightarrow$ **jak je znaleźć?**
- nie możemy ich wyznaczyć, ale możemy **oszacować**

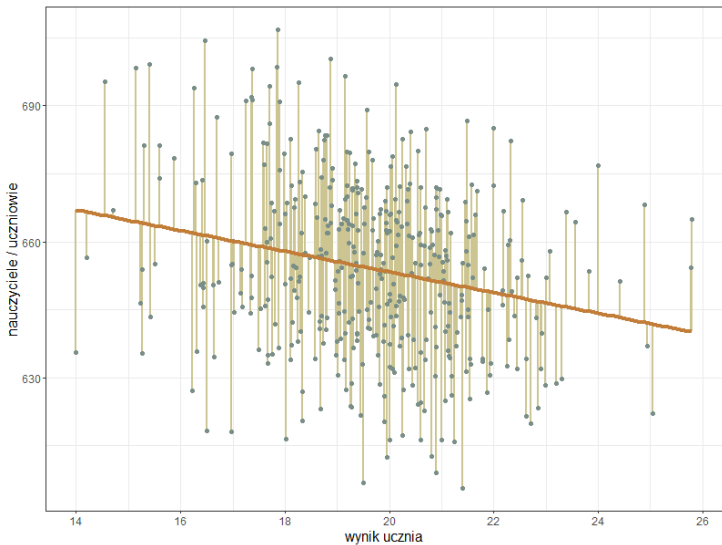
PRZYKŁAD



PRZYKŁAD



PRZYKŁAD



METODA NAJMNIEJSZYCH KWADRATÓW

- szukamy estymatorów: $\hat{\beta}_0 \rightarrow \beta_0$ oraz $\hat{\beta}_1 \rightarrow \beta_1$
- minimalizuje sumę kwadratów e_i **reszt** (odchyłeń) od danych:

$$\min \left\{ \sum_{i=1}^n e_i^2 \right\}$$

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

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \left\{ \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \right\}$$

- Korzystamy z warunków pierwszego rzędu.

Dokładne rozwiązanie na końcu prezentacji

METODA NAJMNIEJSZYCH KWADRATÓW

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

- estymator MNK parametru β_1 , $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2} = \frac{S_{X,Y}}{S_X}$$

- estymator MNK wyrazu wolnego β_0 , $\hat{\beta}_0$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- estymator jest funkcją danych, więc sam jest zmienną losową.

OSZACOWANIE MODELU LINIOWEGO

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

- Y - zmienna zależna
- X - zmienna niezależna
- $\hat{\beta}_0$ - oszacowanie stałej
- $\hat{\beta}_1$ - oszacowanie parametru β_1
- e - reszty

PREDYKCJA

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

- \hat{Y} wartości teoretyczne (dopasowane) zmiennej zależnej:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

- e reszty:

$$e = Y - \hat{Y} = Y - (\hat{\beta}_0 + \hat{\beta}_1 X)$$

- reszty $u \neq \varepsilon$ składnik losowy!

ZAPIS MACIERZOWY

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\mathbf{X}_{(n \times 2)}$ - obserwacje zm. niezależnej (objaśniającej)
- $\mathbf{y}_{(n \times 1)}$ - obserwacje zm. zależnej (objaśnianej)
- $\boldsymbol{\varepsilon}_{(n \times 1)}$ - wektor składników losowych
- $\boldsymbol{\beta}_{(2 \times 1)}$ wektor parametrów β_0, β_1 do oszacowania

ZAPIS MACIERZOWY

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{array}{ccccccc} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} & = & \begin{bmatrix} 1 & x_{1,1} \\ 1 & x_{2,1} \\ \vdots & \vdots \\ 1 & x_{n,1} \end{bmatrix} & \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2 \times 1} & + & \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} \\ \mathbf{y} & = & \mathbf{X}' & \boldsymbol{\beta} & + & \boldsymbol{\varepsilon} \end{array}$$

MOŻEMY ROZSZERZYĆ MODEL DO WIELU ZMIENNYCH

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- k - liczba zmiennych niezależnych (objaśniających)
- $\mathbf{X}_{(n \times k+1)}$ - obserwacje zm. niezależnych (objaśniających)
- $\mathbf{y}_{(n \times 1)}$ - obserwacje zm. zależnej (objaśnianej)
- $\boldsymbol{\varepsilon}_{(n \times 1)}$ - wektor składników losowych
- $\boldsymbol{\beta}_{(k+1 \times 1)}$ wektor $k + 1$ parametrów $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ do oszacowania

MOŻEMY ROZSZERZYĆ MODEL DO WIELU ZMIENNYCH

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & x_{1,1} & \dots & x_{1,k} \\ 1 & x_{2,1} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,k} \end{bmatrix}_{n \times (k+1)} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

$\mathbf{y} \quad = \quad \mathbf{X} \quad \boldsymbol{\beta} \quad + \quad \boldsymbol{\varepsilon}$

MNK - MODEL Z WIELOMA ZMIENNYMI

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

wyznaczamy wektor **reszt**:

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

Aby wyznaczyć estymator MNK $\hat{\boldsymbol{\beta}}$, minimalizujemy sumę kwadratów reszt:

$$\begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = e_1^2 + e_2^2 + \cdots + e_n^2 = \mathbf{e}'\mathbf{e}$$

MNK - MODEL Z WIELOMA ZMIENNYMI

Przekształćmy reszty:

$$\begin{aligned}\mathbf{e}'\mathbf{e} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}\end{aligned}$$

Minimalizujemy reszty:

$$\min_{\hat{\boldsymbol{\beta}}} \left\{ \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \right\}$$

Warunek pierwszego rzędu:

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

MNK - MODEL Z WIELOMA ZMIENNYMI

Przekształćmy reszty:

$$\begin{aligned}\mathbf{e}'\mathbf{e} &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}\end{aligned}$$

Minimalizujemy reszty:

$$\min_{\hat{\beta}} \left\{ \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \right\}$$

Warunek pierwszego rzędu:

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial \hat{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0$$

MNK - MODEL Z WIELOMA ZMIENNYMI

Przekształćmy reszty:

$$\begin{aligned}\mathbf{e}'\mathbf{e} &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}\end{aligned}$$

Minimalizujemy reszty:

$$\min_{\hat{\beta}} \left\{ \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \right\}$$

Warunek pierwszego rzędu:

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial \hat{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0$$

MNK - MODEL Z WIELOMA ZMIENNYMI

Z warunku pierwszego rzędu otrzymujemy układ równań normalnych:

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

Gdy \mathbf{X} ma pełen rząd, $\mathbf{X}'\mathbf{X}$ jest dodatnio określona, odwracalna \Rightarrow układ równań ma jednoznaczne rozwiązanie

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

Wyznaczamy $\hat{\beta}$:

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

MNK - MODEL Z WIELOMA ZMIENNYMI

Z warunku pierwszego rzędu otrzymujemy układ równań normalnych:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

Gdy \mathbf{X} ma pełen rząd, $\mathbf{X}'\mathbf{X}$ jest dodatnio określona, odwracalna \Rightarrow układ równań ma jednoznaczne rozwiązanie

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Wyznaczamy $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

MNK - MODEL Z WIELOMA ZMIENNYMI

Z warunku pierwszego rzędu otrzymujemy układ równań normalnych:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

Gdy \mathbf{X} ma pełen rząd, $\mathbf{X}'\mathbf{X}$ jest dodatnio określona, odwracalna \Rightarrow układ równań ma jednoznaczne rozwiązanie

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Wyznaczamy $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

PREDYKCJA - MODEL Z WIELOMA ZMIENNYMI

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$$

- $\hat{\mathbf{y}}$ wartości teoretyczne (dopasowane) zmiennej zależnej:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

- \mathbf{e} wektor reszt:

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

KLUCZOWE POJĘCIA

model linowy, metoda najmniejszych kwadratów, parametry, oszacowania, wartości dopasowane (teoretyczne), reszty, składnik losowy, MNK w zapisie macierzowym, minimalizacja sumy kwadratów reszt, estymator MNK, wariancja & kowariancja z próby (empiryczna), własności estymatora MNK (patrz zadanie 1 z listy), macierz \mathbf{X} , zmienna zależna (objaśniania), zmienna niezależna (objaśniająca)

Pytania? Wątpliwości?
Dziękuję!

e: s.zalas@uw.edu.pl

WYPROWADZENIE MNK

Metoda najmniejszych kwadratów polega na znalezieniu takich wartości $\hat{\beta}_0$, $\hat{\beta}_1$, które minimalizują sumę kwadratów reszt:

$$\min \sum_{i=1}^n (\hat{e}_i)^2$$
$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Warunki pierwszego rzędu:

$$\frac{\partial W}{\partial \hat{\beta}_0} = \sum_{i=1}^n -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (1)$$

$$\frac{\partial W}{\partial \hat{\beta}_1} = \sum_{i=1}^n -2x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (2)$$

WYPROWADZENIE MNK

Przekształćmy równanie (1):

$$\sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_0 - \sum_{i=1}^n \hat{\beta}_1 x_i = 0$$

zauważmy, że $n\bar{x} = \sum_{i=1}^n x_i$ oraz $n\bar{y} = \sum_{i=1}^n y_i$

$$n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1\bar{x} = 0$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \quad (3)$$

otrzymaliśmy wzór na oszacowanie wyrazu wolnego.

WYPROWADZENIE MNK

Przekształćmy równanie (2):

$$\begin{aligned}\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \hat{\beta}_0 - \sum_{i=1}^n x_i \hat{\beta}_1 x_i &= 0 \\ \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i (\bar{y} - \hat{\beta}_1 \bar{x}) - \sum_{i=1}^n x_i \hat{\beta}_1 x_i &= 0 \\ \sum_{i=1}^n x_i y_i - n \bar{y} \bar{x} + \hat{\beta}_1 n \bar{x}^2 - \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= 0 \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}\end{aligned}\tag{4}$$

Otrzymaliśmy wzór na oszacowanie $\hat{\beta}_1$, pokażemy że jest ono równe stosunkowi kowariancji x_i oraz y_i z próby, do wariancji x_i z próby.

WYPROWADZENIE MNK

Przekształćmy licznik równania (4). Pokażemy, że

$$\sum_{i=1}^n x_i y_i - n \bar{y} \bar{x} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

przekształcając prawą stronę:

$$\begin{aligned} &= \sum_{i=1}^n (x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x} \bar{y}) \\ &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \bar{y} - \sum_{i=1}^n y_i \bar{x} + \sum_{i=1}^n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} - n \bar{x} \bar{y} + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \end{aligned}$$

WYPROWADZENIE MNK

Przekształcimy mianownik (4). Pokażemy, że

$$\sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

przekształcając prawą stronę:

$$\begin{aligned} &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 \end{aligned}$$