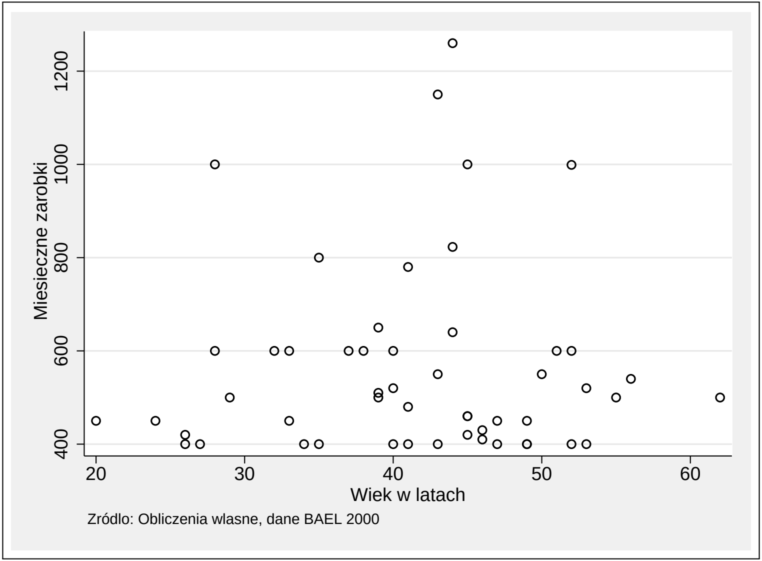
Metoda najmniejszych kwadratów

Szacowanie parametrów


Miesieczne zarobki
400
600
800
1000
1200
20
30
40
50
60
Wiek w latach
Zródlo: Obliczenia wlasne, dane BAEL 2000

- Model teoretyczny zjawiska

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

- Oszacowanie

$$y = b_0 + b_1 x_1 + e$$

Ważne:

Elementy y, x_1 oraz e są wektorami, β_1 i b_1 są skalarami.

Tabela: Dane o płacach i wieku

Numer obserwacji	zarobki	wiek
1	520	60
2	430	36
3	415	61
⋮	⋮	⋮
n	400	53

- Reszty dla i-tej obserwacji

$$e_i = y_i - \hat{y}_i = y_i - b_0 - x_1 b_1$$

- Minimalizowana jest suma kwadratów reszt

$$S(b_0, b_1) \to \min$$

Warunki pierwszego rzędu

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

Rozwiązanie

- Wzory dla oszacowań parametrów

$$\begin{cases} b_0 = \bar{y} - b_1\bar{x} \\ b_1 = \frac{S_{yx}}{S_x^2} \end{cases}$$

- Pozostaje pokazać, iż rzeczywiście jest to minimum.

- Równanie ogólne

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + e$$

- Równanie dla i-tej obserwacji

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_{ki} x_k + e_i$$

- Dla N obserwacji można zapisać układ N równań

$$\left\{\begin{array}{ccccccccccc} y_1 & = & b_0 & + & b_1 x_{11} & + & \ldots & + & b_k x_{k1} & + & e_1 \\ y_2 & = & b_0 & + & b_2 x_{11} & + & \ldots & + & b_k x_{k2} & + & e_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_N & = & b_0 & + & b_1 x_{1N} & + & \ldots & + & b_k x_{kN} & + & e_N \end{array}\right.$$

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} x_{11} & x_{21} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nk} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}}_{\varepsilon}$$

- Zapis macierzowy

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

- Zapis z wykorzystaniem **wektorów wierszowych**

$$y_i = x_i\beta + \varepsilon_i$$

$$e'e = \begin{bmatrix} e_1 & e_2 & \dots & e_N \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

Wyprowadzenie estymatora MNK

- $$S(b) = e'e = (y - Xb)'(y - Xb)$$
- $$S(b) = y'y - 2y'Xb + b'X'Xb$$
- Warunki pierwszego rzędu
$$-2X'y + 2X'Xb = 0$$
- Układ równań normalnych
$$X'Xb = X'y$$

Wyprowadzenie estymatora MNK

Wzór na estymator MNK

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