

Metoda najmniejszych kwadratów

Model ekonometryczny

Wykształcenie a wynagrodzenia

- Hipoteza badawcza: Istnieje zależność między poziomem wykształcenia a poziomem wynagrodzenia
- Teoria ekonomiczna
- Dane empiryczne

Wykształcenie a wynagrodzenia

| poziom wykształcenia | przeciętne zarobki |
|-------------------------|-----------------------|
| wyższe | 677,78 |
| policealne | 564,37 |
| średnie zawodowe | 584,30 |
| średnie ogólne | 586,45 |
| zasadnicze zawodowe | 555,43 |
| podstawowe | 545,85 |
| bez podstawowego | 543,75 |

Źródło: Obliczenia własne na podstawie BAEL (2000)

Model

- Zależność funkcyjna

$$y = f(x_1, x_2, \dots, x_k)$$

- Model liniowy zjawiska

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Elementy modelu

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

- y - zmienna objaśniana (zależna, endogeniczna)
- x_j - zmienne objaśniające (niezależne, egzogeniczne)
- β_j - nieznane parametry modelu wymagające oszacowania
- ε - błąd losowy

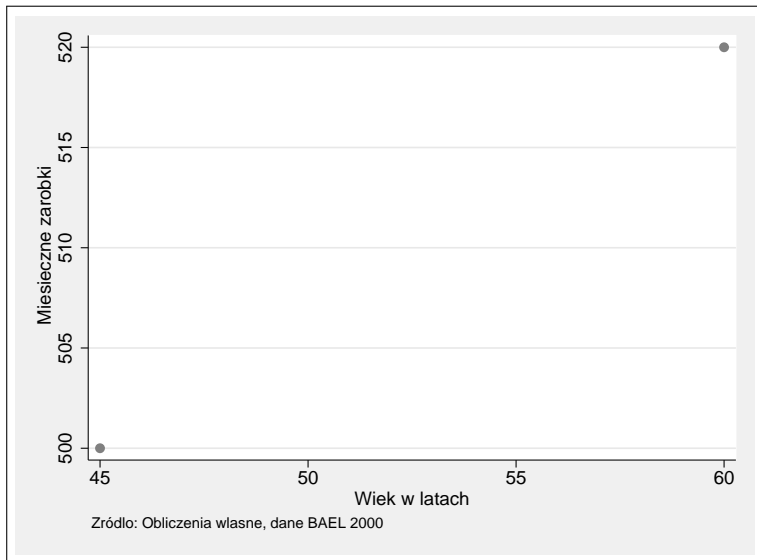
Przyczyny istnienia błędu losowego

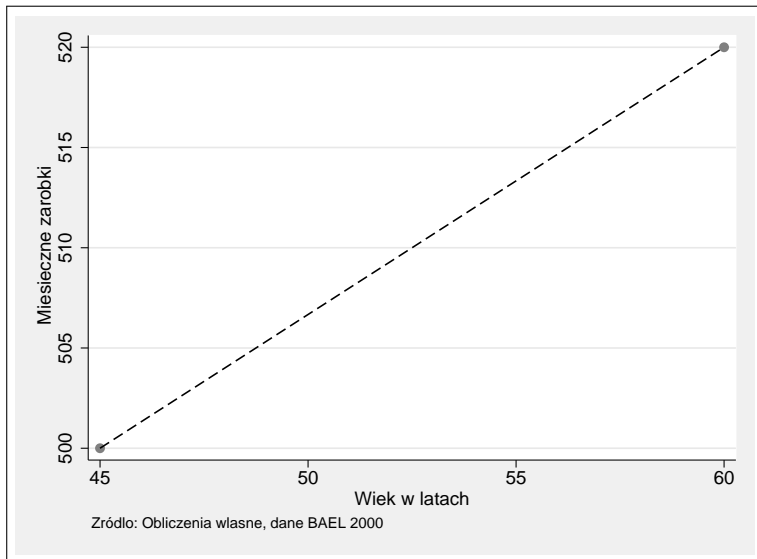
- Pominięte czynniki
- Zjawisko nie musi przyjmować wszystkich wartości
- Błędy pomiaru
- Błędy w danych empirycznych

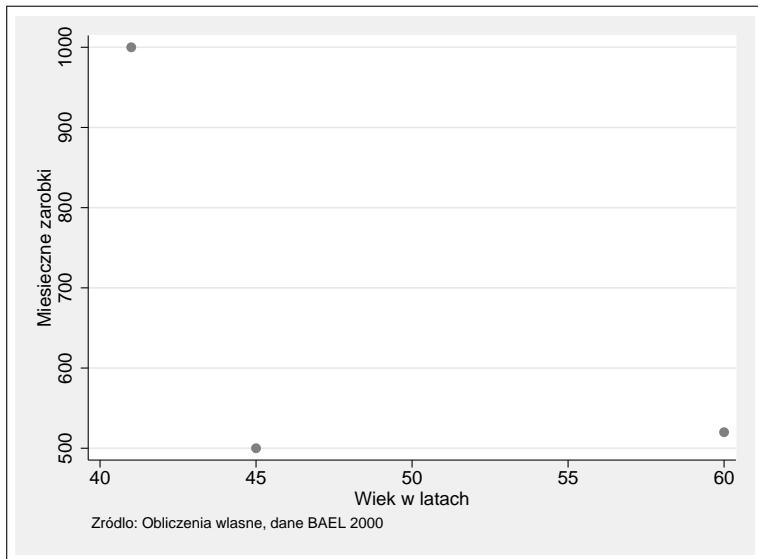
Model Galtona

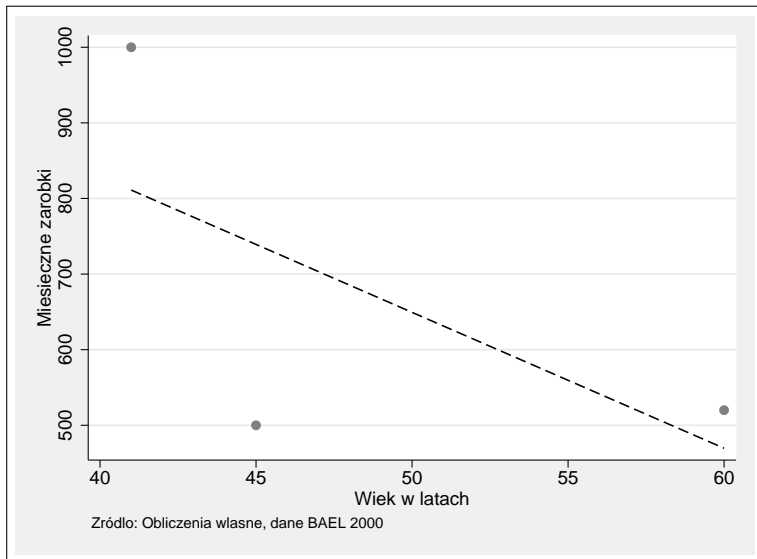
$$\text{wzrost dzieci} = \beta_0 + \beta_1 \text{ wzrost rodziców} + \varepsilon$$

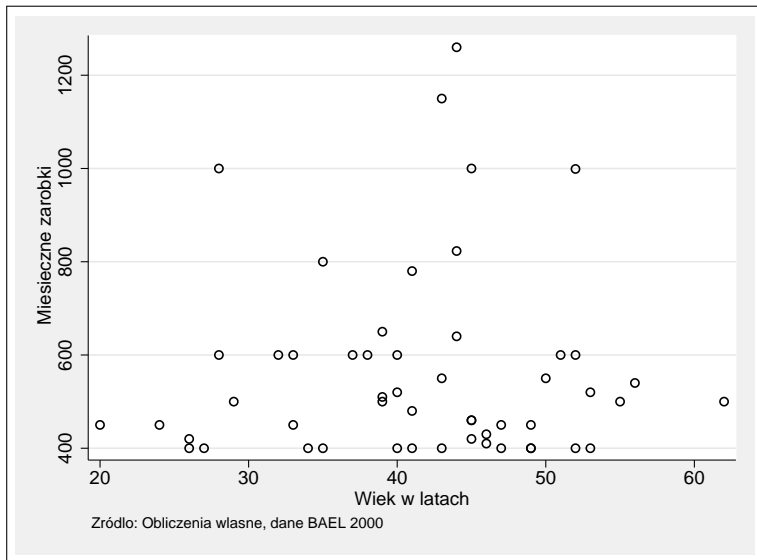
- Brak symetrii

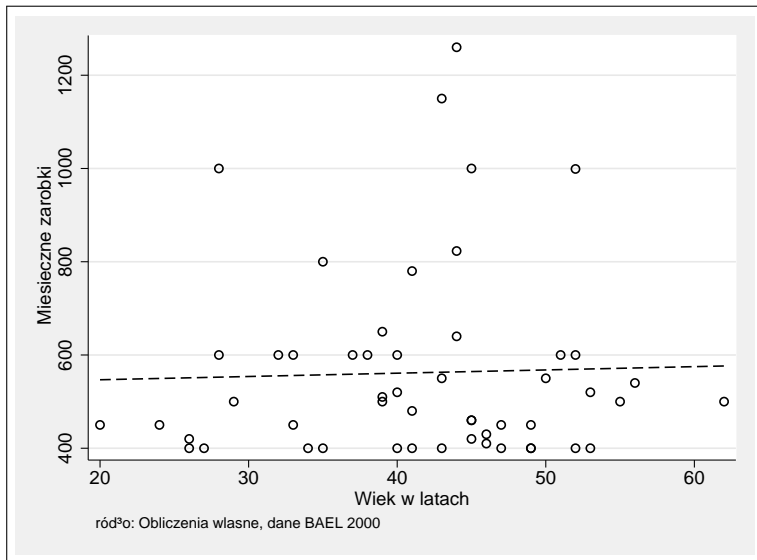


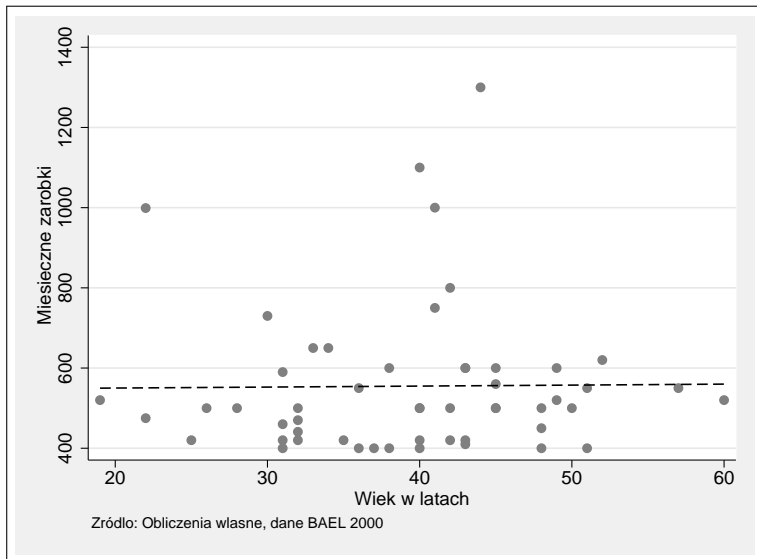


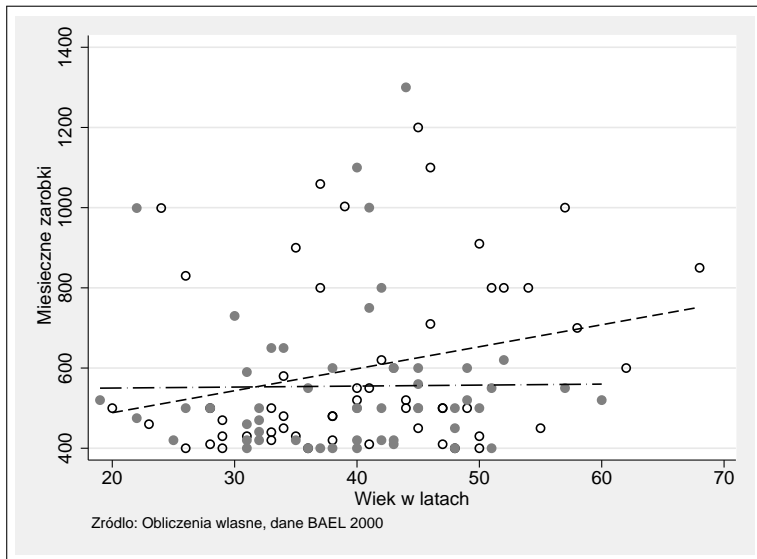












Ważny wniosek

Oszacowania nielosowych parametrów hiperpłaszczyzny (linii) regresji są losowe.

Regresja

Procedurę szukania wartości estymatora Metody Najmniejszych Kwadratów określamy terminem **regresja**.

- Model teoretyczny zjawiska

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

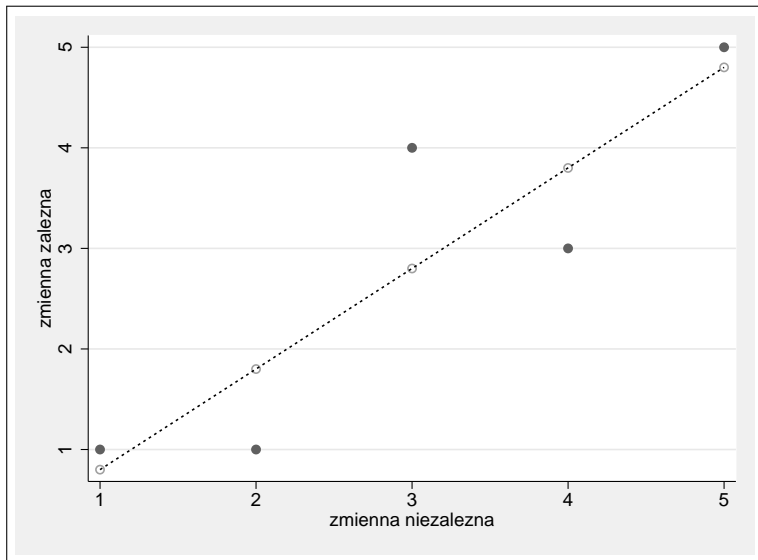
- Oszacowanie modelu

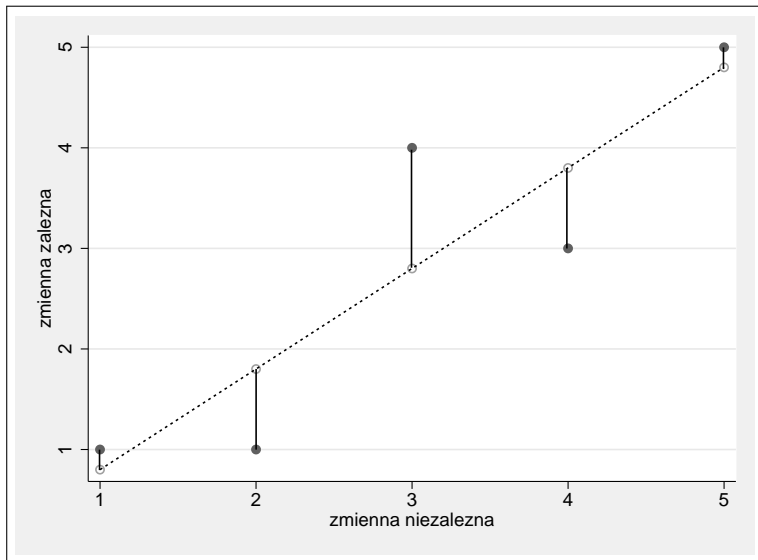
$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + e$$

- Wartość dopasowana

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$$

- Reszta a błąd losowy
- Dopasowanie modelu





Dane

| X | Y |
|---|---|
| 1 | 1 |
| 2 | 1 |
| 3 | 4 |
| 4 | 3 |
| 5 | 5 |

Wartości dopasowane

| X | Y | \hat{Y} |
|---|---|-----------|
| 1 | 1 | .8 |
| 2 | 1 | 1.8 |
| 3 | 4 | 2.8 |
| 4 | 3 | 3.8 |
| 5 | 5 | 4.8 |

Reszty

| X | Y | \hat{Y} | e |
|---|---|-----------|------|
| 1 | 1 | .8 | 0.2 |
| 2 | 1 | 1.8 | -0.8 |
| 3 | 4 | 2.8 | 1.2 |
| 4 | 3 | 3.8 | -0.8 |
| 5 | 5 | 4.8 | 0.2 |

Kryteria dopasowania

$$\sum_i |y_i - \hat{y}_i| = \sum_i |e_i|$$

$$\sum_i \text{dist}(y_i, \hat{y}_i)$$

$$\sum_i (y_i - \hat{y}_i)^2 = \sum_i e_i^2$$