



REGRESJA LINIOWA

Dawid Słomian

- 
- Równanie regresji liniowej
 - Estymacja współczynników prostej regresji
 - metoda najmniejszych kwadratów
 - ważona metoda najmniejszych kwadratów
 - Przykłady równań regresji
 - regresja wieloraka
 - wielomiany
 - regresja logarytmiczna
 - Interpretacja przykładowych wyników

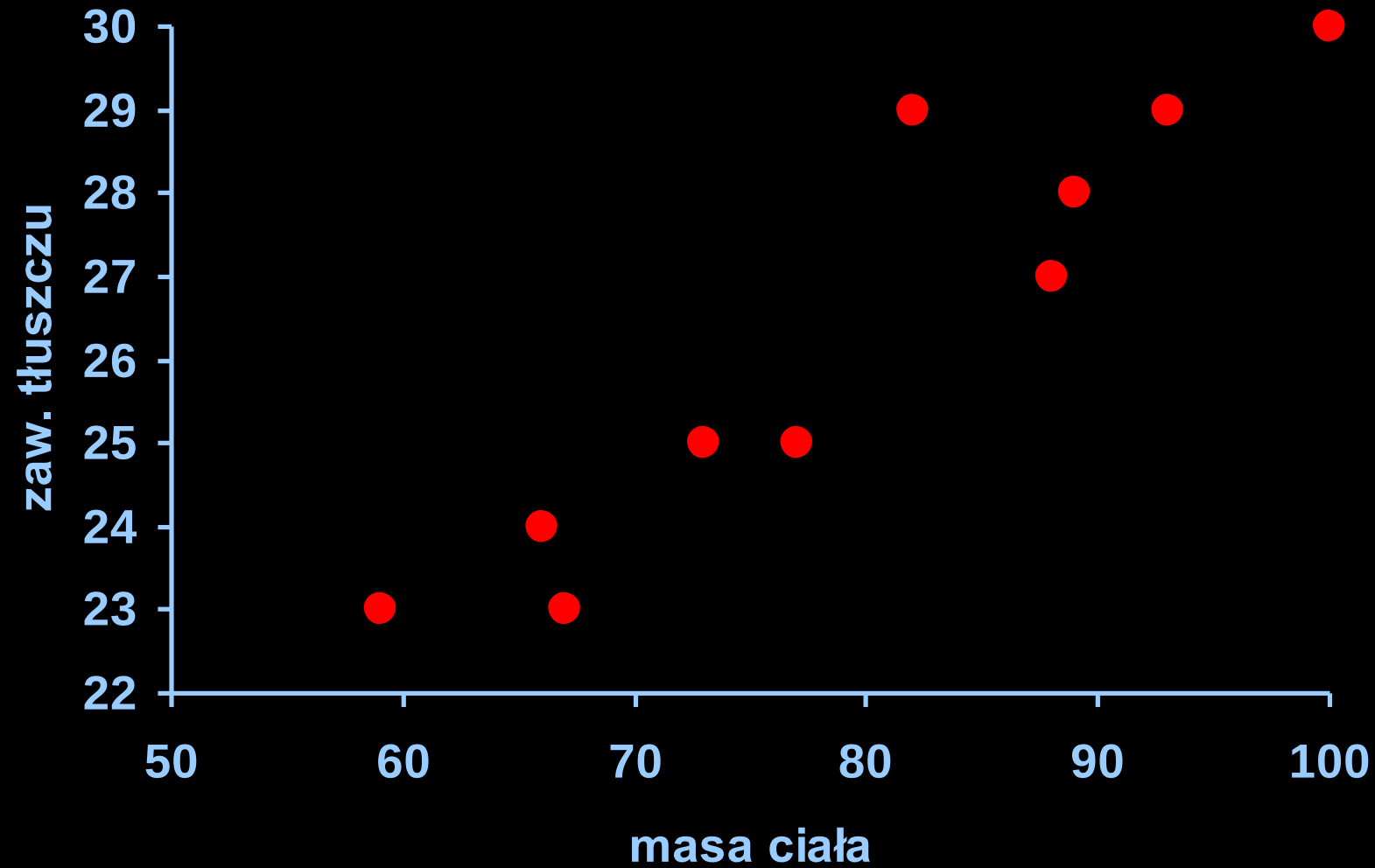
RÓWNANIA REGRESJI

- Regresja liniowa = wyznaczenie prostej, która pozwala na predykcję wartości zmiennej Y na podstawie znajomości zmiennej X
- Zmienna niezależna (X)
 - Dokładnie znane wartości / wartości eksperymentalne
- Zmienna zależna (Y)
 - Zakładamy błąd pomiarowy



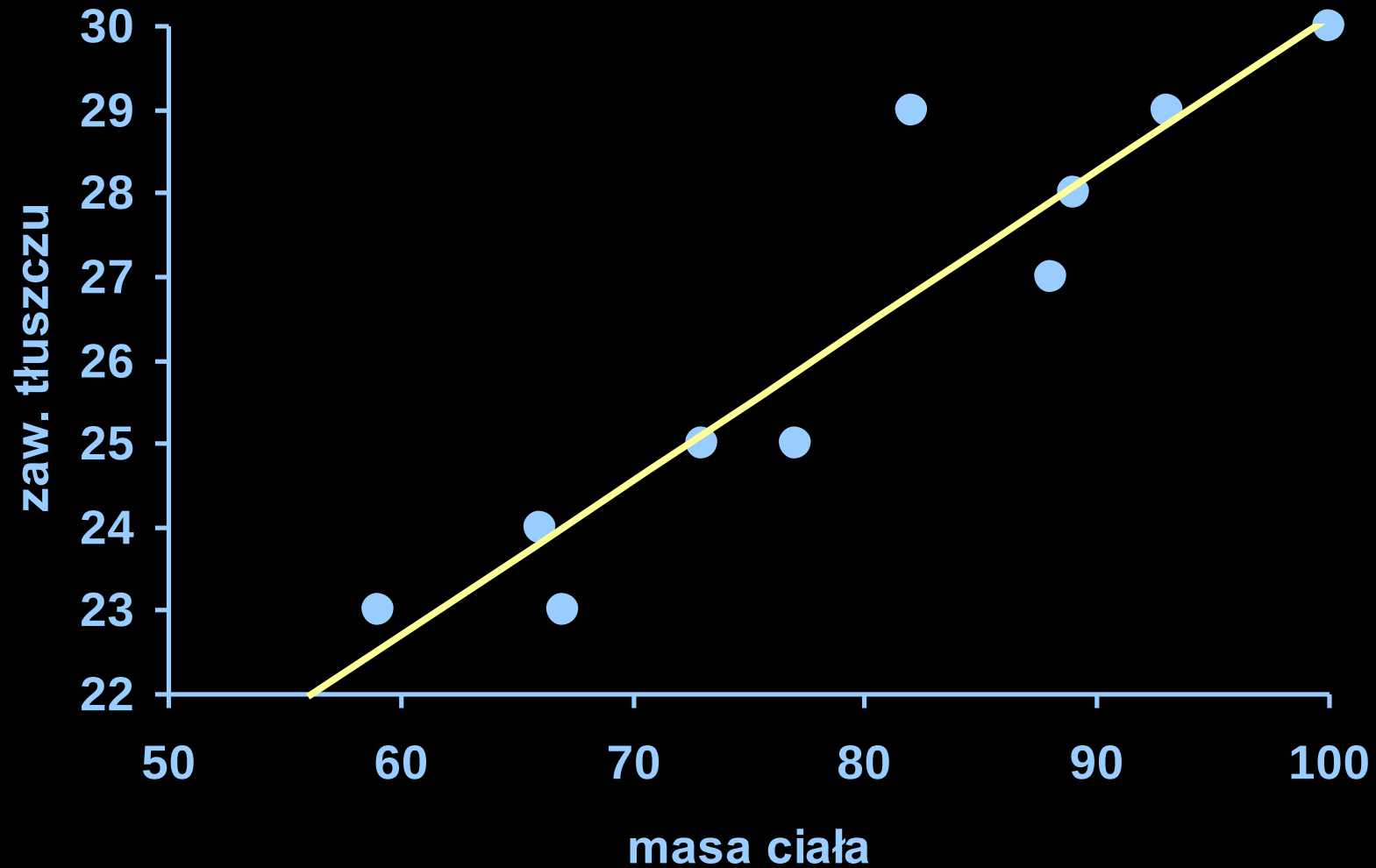
MASA CIAŁA	ZAW. TŁUSZCZU
89	28
88	27
66	24
59	23
93	29
73	25
82	29
77	25
100	30
67	23

RÓWNANIE REGRESJI



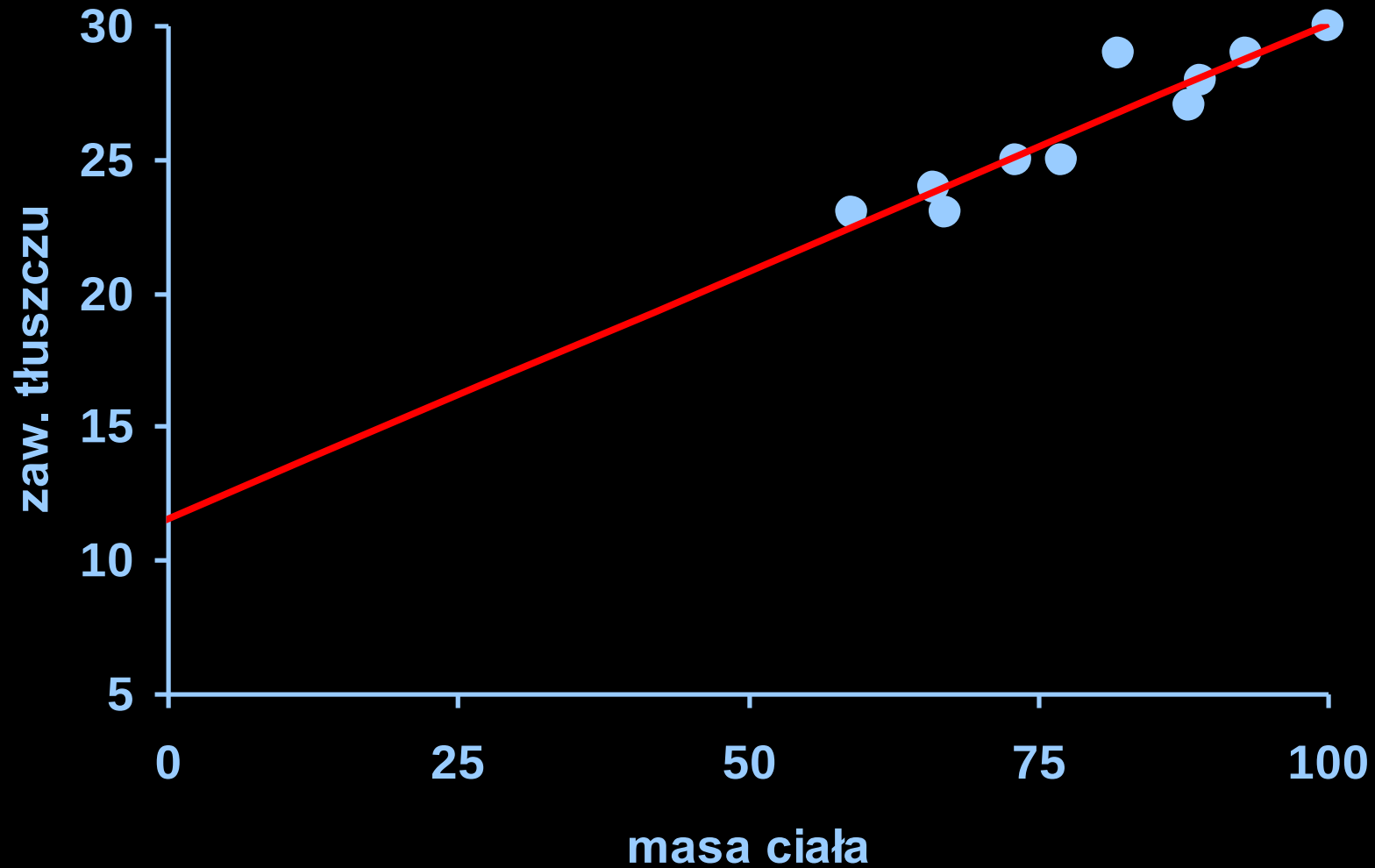
RÓWNANIE REGRESJI

elementy równania regresji: wyraz wolny
Intercept



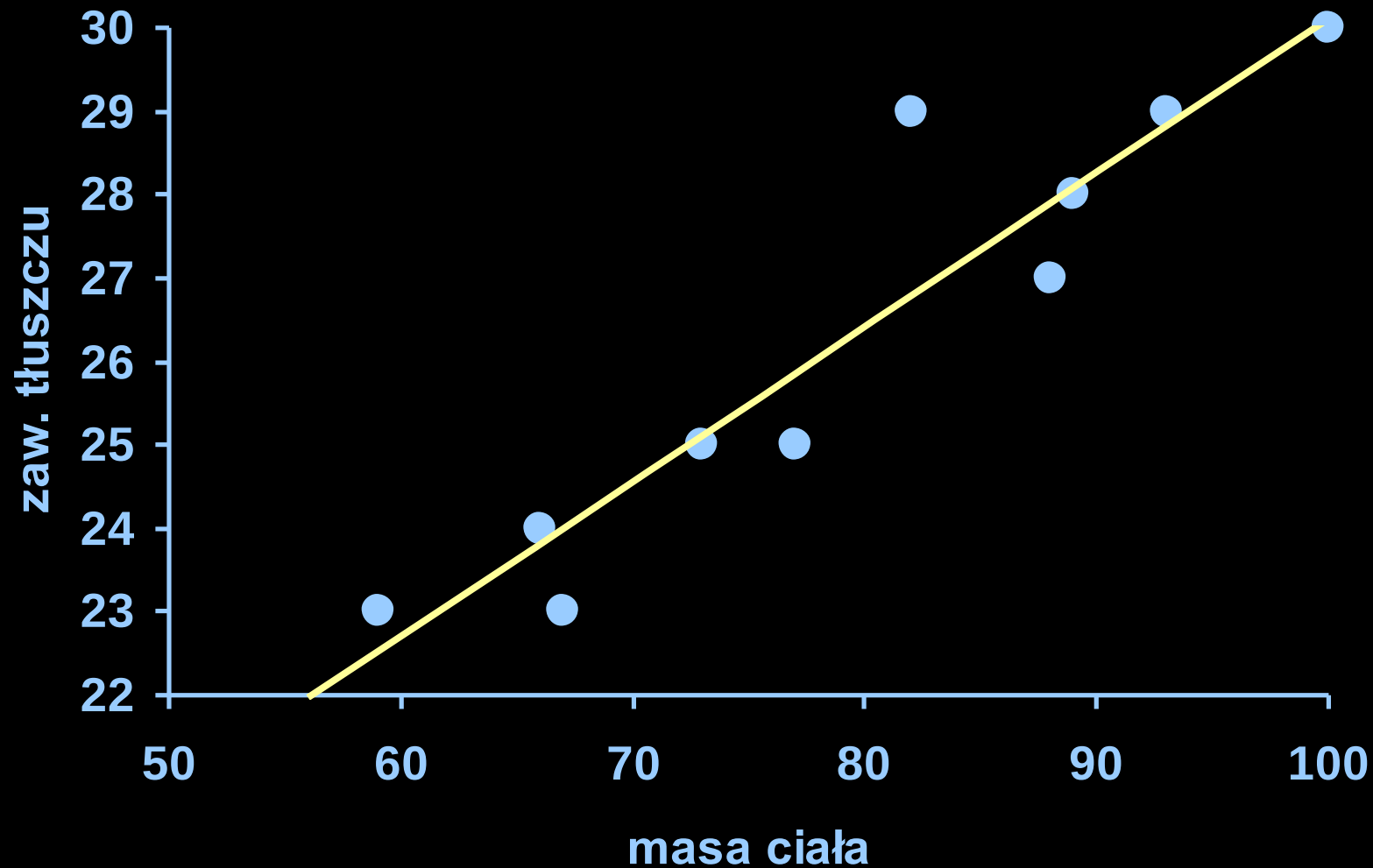
RÓWNANIE REGRESJI

elementy równania regresji: gradient Slope



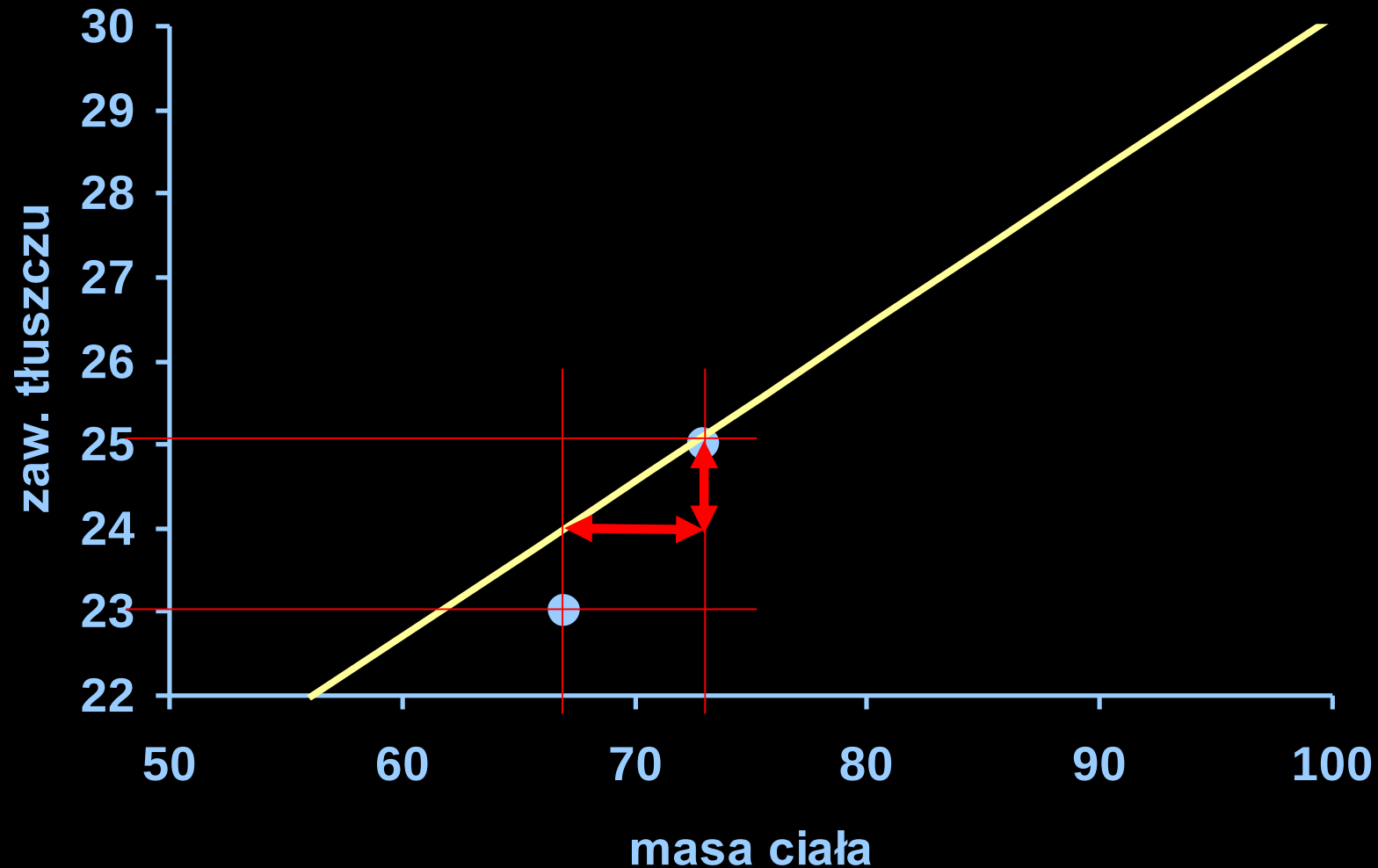
RÓWNANIE REGRESJI

elementy równania regresji: gradient Slope

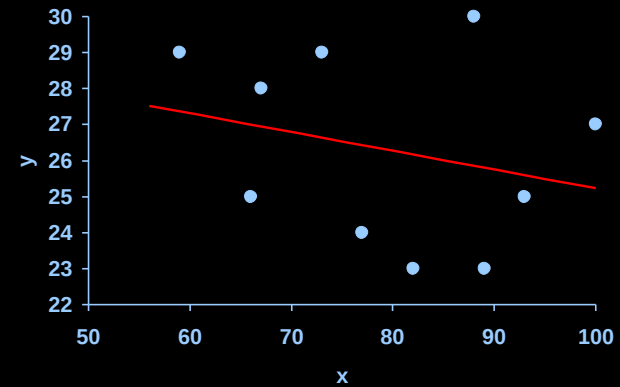
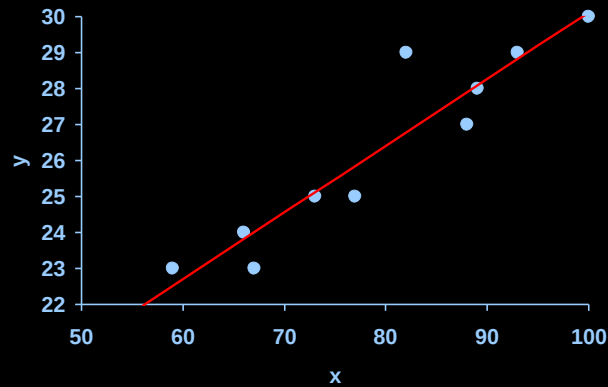
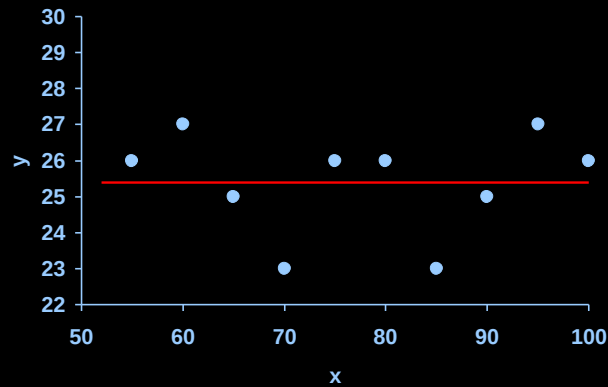


RÓWNANIE REGRESJI

elementy równania regresji: gradient Slope



RÓWNANIE REGRESJI



O ile zmieni się wartość zmiennej zależnej jeżeli zmienna niezależna wzrośnie o 1

≈ 0

Brak zależności

gradient = 0

> 0

Zmiany w tym samym kierunku

gradient > 0

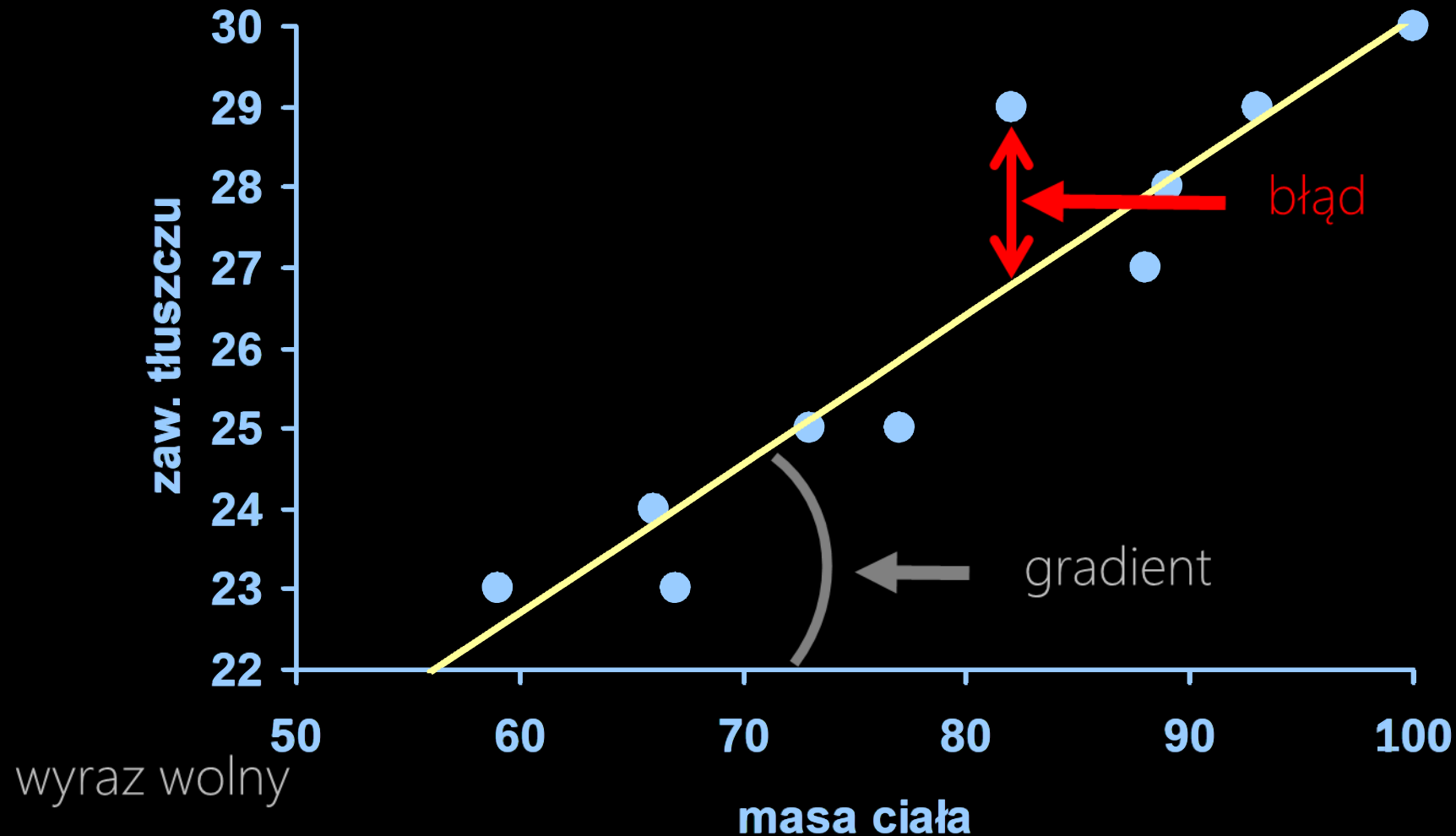
< 0

Zmiany w różnych kierunkach

gradient < 0

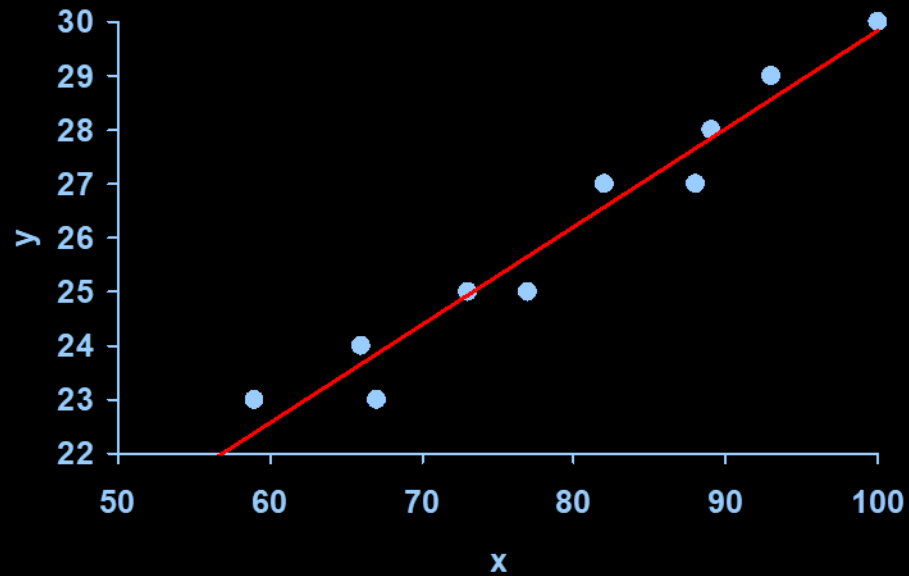
RÓWNANIE REGRESJI

elementy równania regresji: błąd Error, Residual

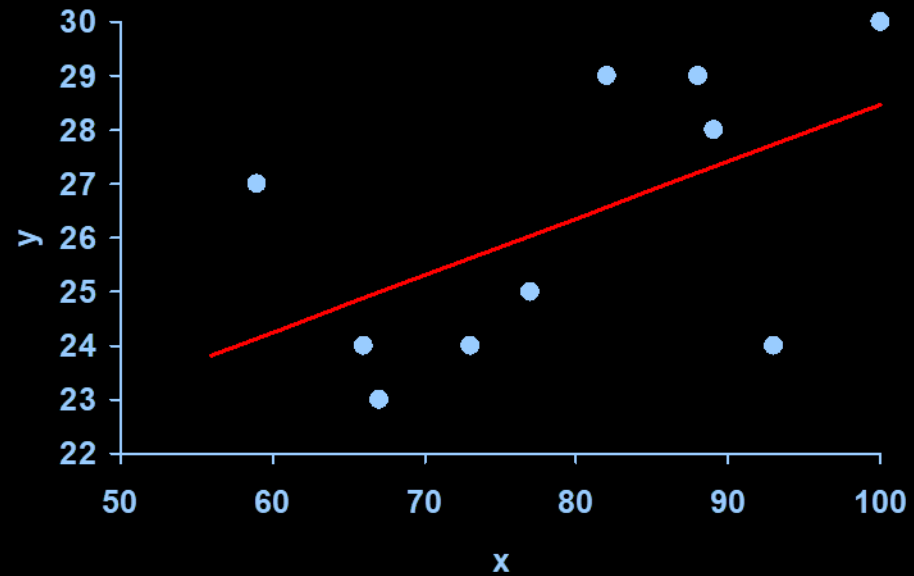


RÓWNANIE REGRESJI

elementy równania regresji: błąd Error, Residual



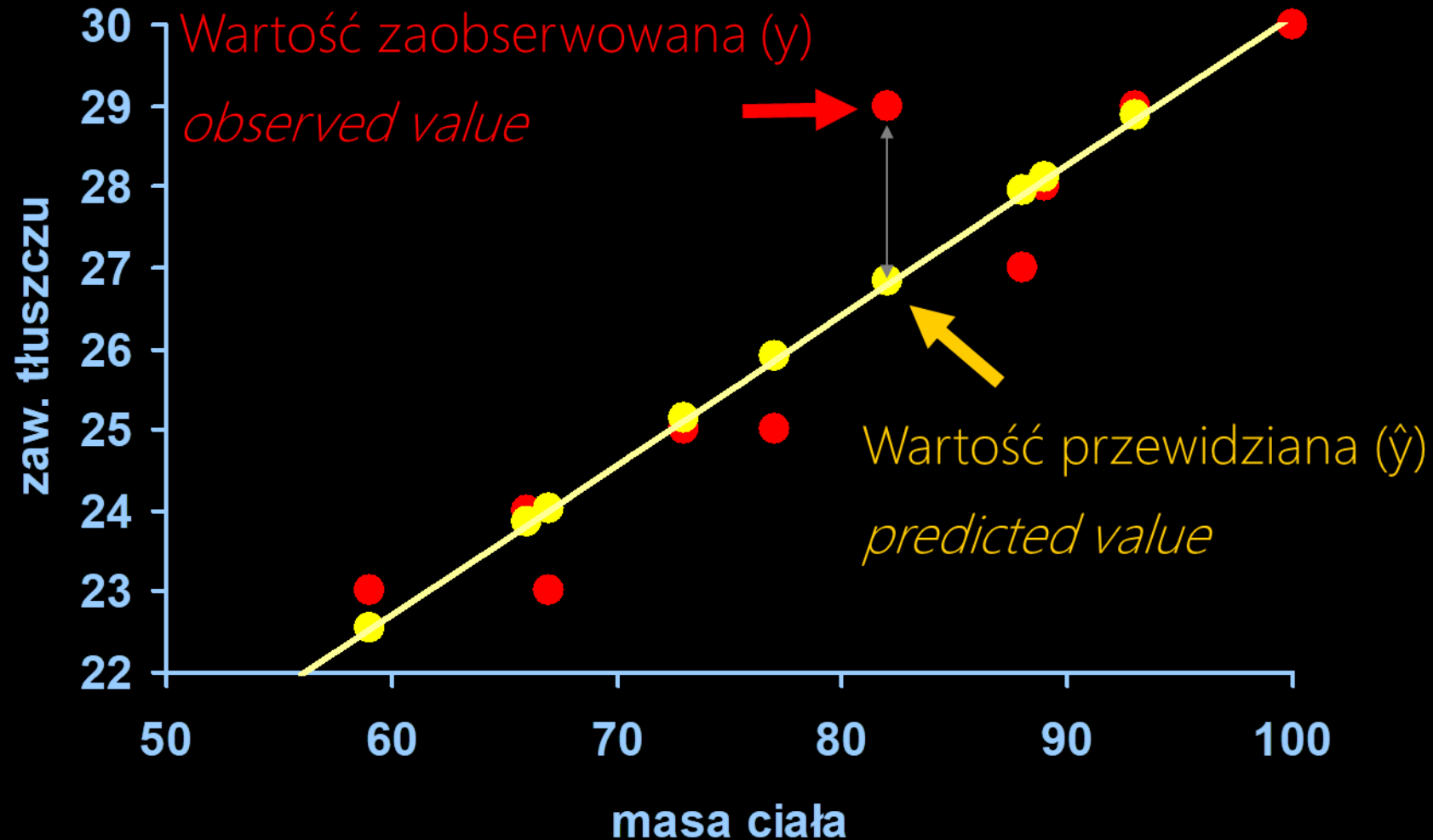
mały błąd



duży błąd

RÓWNANIE REGRESJI

elementy równania regresji: błąd Error, Residual



RÓWNANIE REGRESJI

Elementy równania regresji

wyraz wolny

gradient

błąd

$$y = \beta_0 + \beta_1 x + e$$

Zmienna zależna
Dependent variable

Zmienna niezależna
Independent variable

RÓWNANIE REGRESJI

Elementy równania regresji

$$y = \beta_0 + \beta_1 x + e$$

Zmienna zależna

Left Hand Side

LHS

Zmienna niezależna

Right Hand Side

RHS

RÓWNANIE REGRESJI

Elementy równania regresji

$$y = \beta_0 + \beta_1 x + e$$

Komponent systematyczny

$$\beta_0 + \beta_1 x$$

Komponent losowy

$$e \sim N(0, \sigma_e^2)$$

RÓWNANIE REGRESJI

Elementy równania regresji

wyraz wolny

gradient

$$\text{tłuszcz} = 11.57 + 0.19\text{masa_ciała}$$

Zmienna zależna

Zmienna niezależna

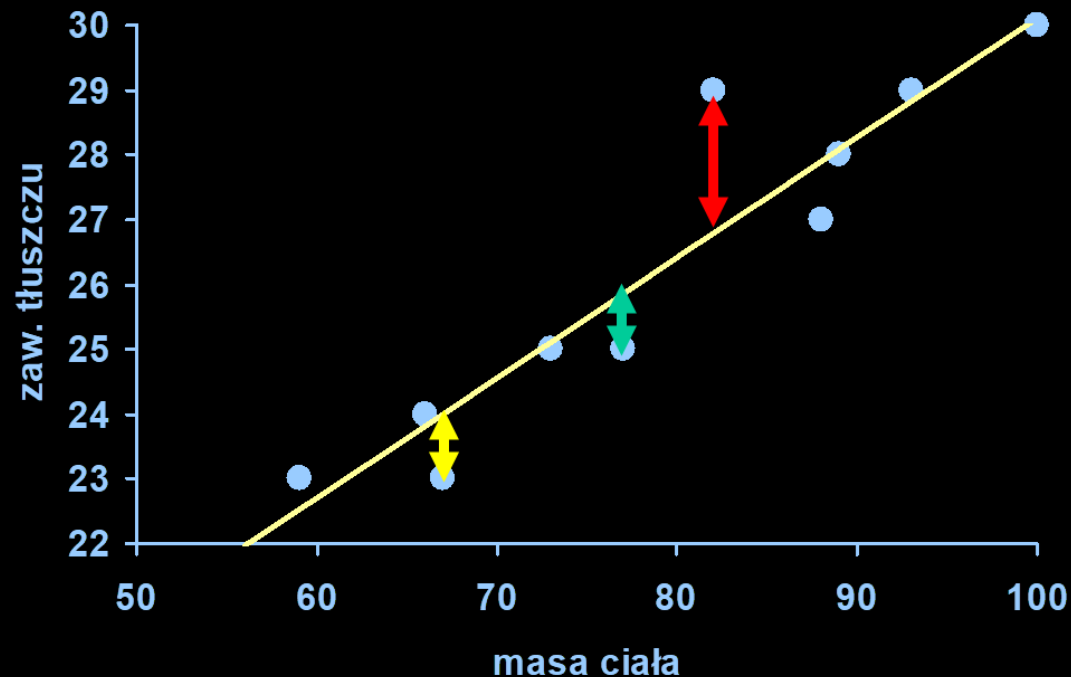
Co mówią nam poszczególne wartości?

METODA NAJMNIJSZYCH KWADRATÓW

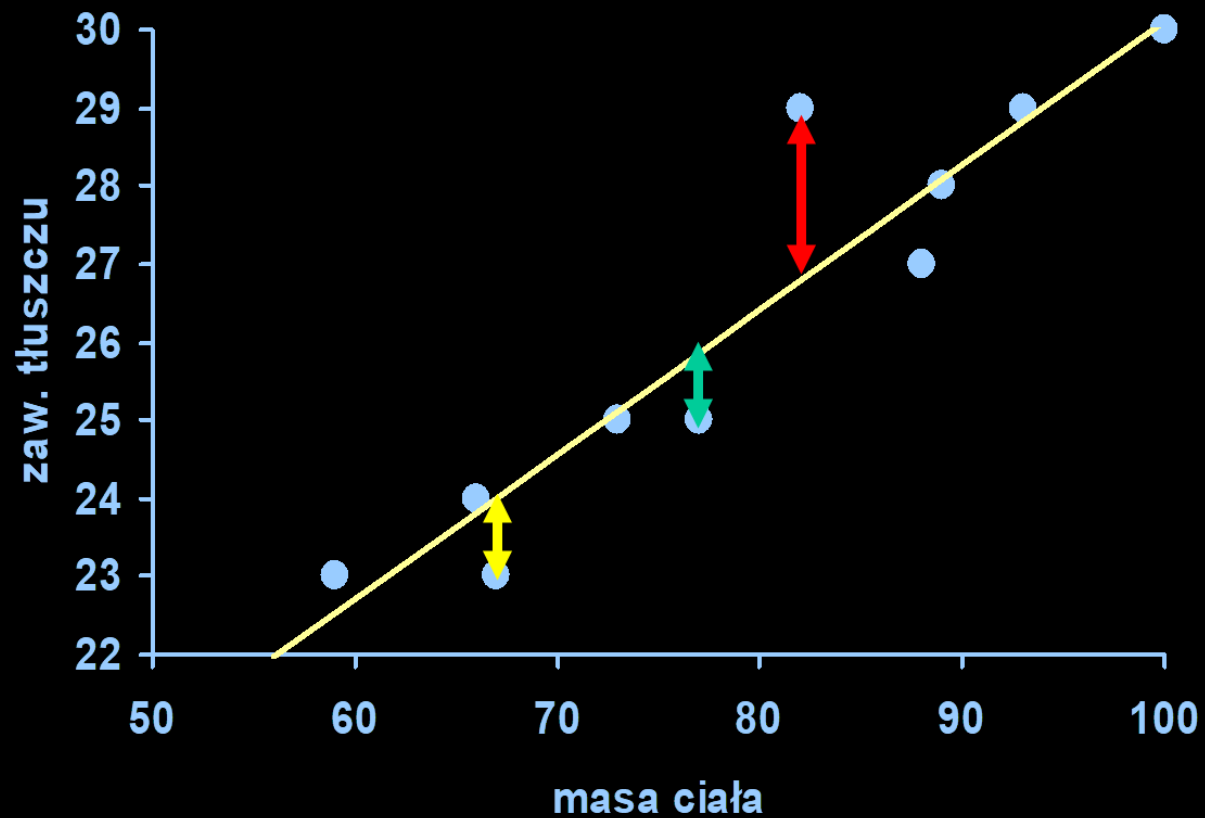
Jak wyznaczyć parametry równania regresji ?

$$y = \beta_0 + \beta_1 x + e$$

Tak aby odległości obserwacji od prostej były jak najmniejsze →
metoda najmniejszych kwadratów



METODA NAJMNIJSZYCH KWADRATÓW



$$\sum (\text{↕})^2 \rightarrow \text{minimum}$$

METODA NAJMNIJSZYCH KWADRATÓW

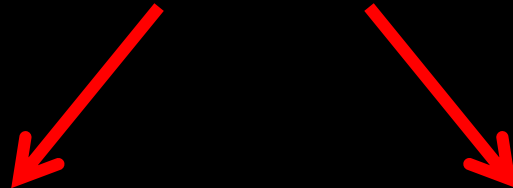
$$\sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N [y_i - (\beta_0 + \beta_1 x_i)]^2$$

$$\frac{\partial \sum_{i=1}^N [y_i - (\beta_0 + \beta_1 x_i)]^2}{\partial \beta_0} = 0 \quad i \quad \frac{\partial \sum_{i=1}^N [y_i - (\beta_0 + \beta_1 x_i)]^2}{\partial \beta_1} = 0$$

$$-2 \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i) = 0 \quad i \quad -2 \sum_{i=1}^N x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

METODA NAJMNIEJSZYCH KWADRATÓW

$$y = \beta_0 + \beta_1 x$$


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

$$\frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

WAŻONA METODA NAJMNIEJSZYCH KWADRATÓW

Jak wyznaczyć parametry równania regresji jeżeli zmienna zależna y mierzona jest ze zróżnicowaną dokładnością?

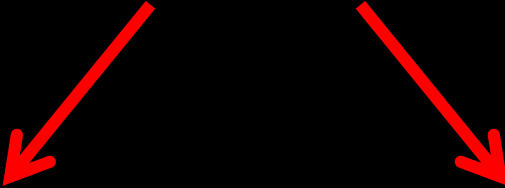
$$y = \beta_0 + \beta_1 x + e$$

ważenie obserwacji



WAŻONA METODA NAJMNIEJSZYCH KWADRATÓW

$$\frac{\partial \sum_{i=1}^N w_i [y_i - (\beta_0 + \beta_1 x_i)]^2}{\partial \beta_0} = 0 \quad i \quad \frac{\partial \sum_{i=1}^N w_i [y_i - (\beta_0 + \beta_1 x_i)]^2}{\partial \beta_1} = 0$$

$$y = \beta_0 + \beta_1 x$$


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

$$\frac{\sum_{i=1}^N w_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N w_i (x_i - \bar{x})^2}$$

REGRESJA WIELORAKA

MASA CIAŁA	WIEK	ZAW. TŁUSZCZU
89	35	28
88	43	27
66	24	24
59	56	23
93	48	29
73	31	25
82	61	29
77	27	25
100	31	30
67	47	23

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\text{tluszcz} = \beta_0 + \beta_1 \text{masa_c} + \beta_2 \text{wiek}$$



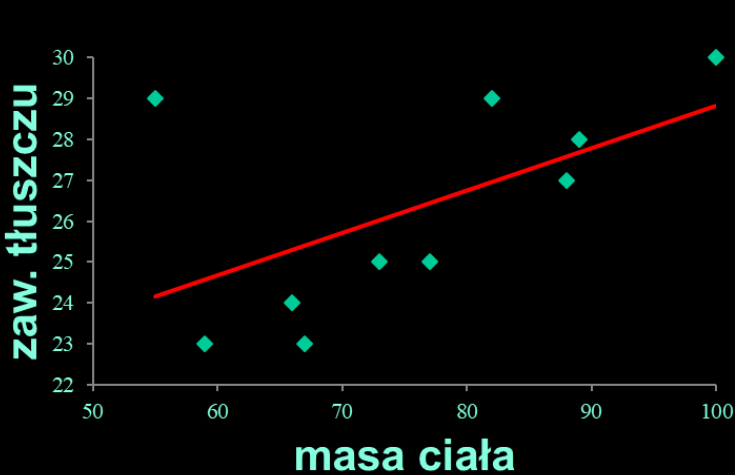
WIELOMIANY

$$y = \beta_0 + \beta_1 x$$

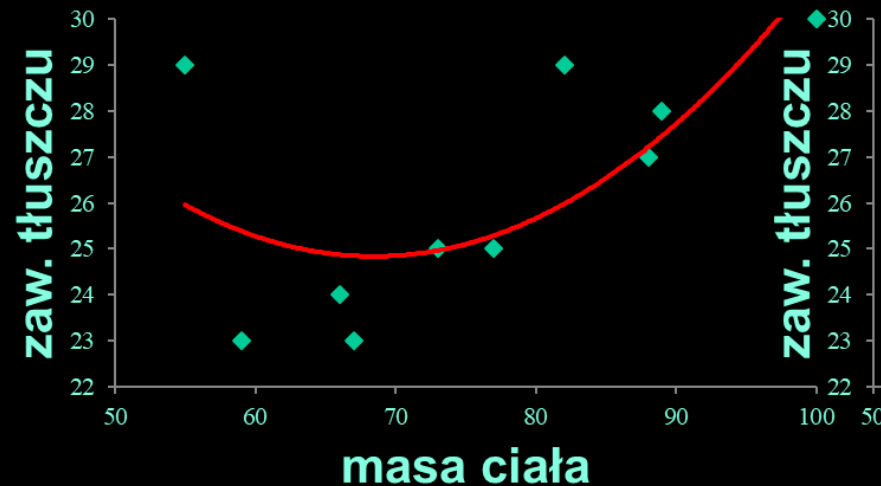
$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

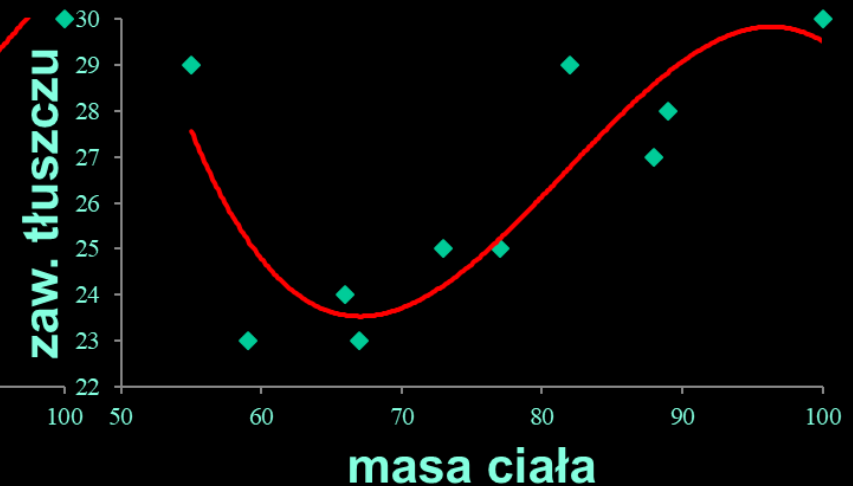
wielomian
1go stopnia



wielomian
2go stopnia

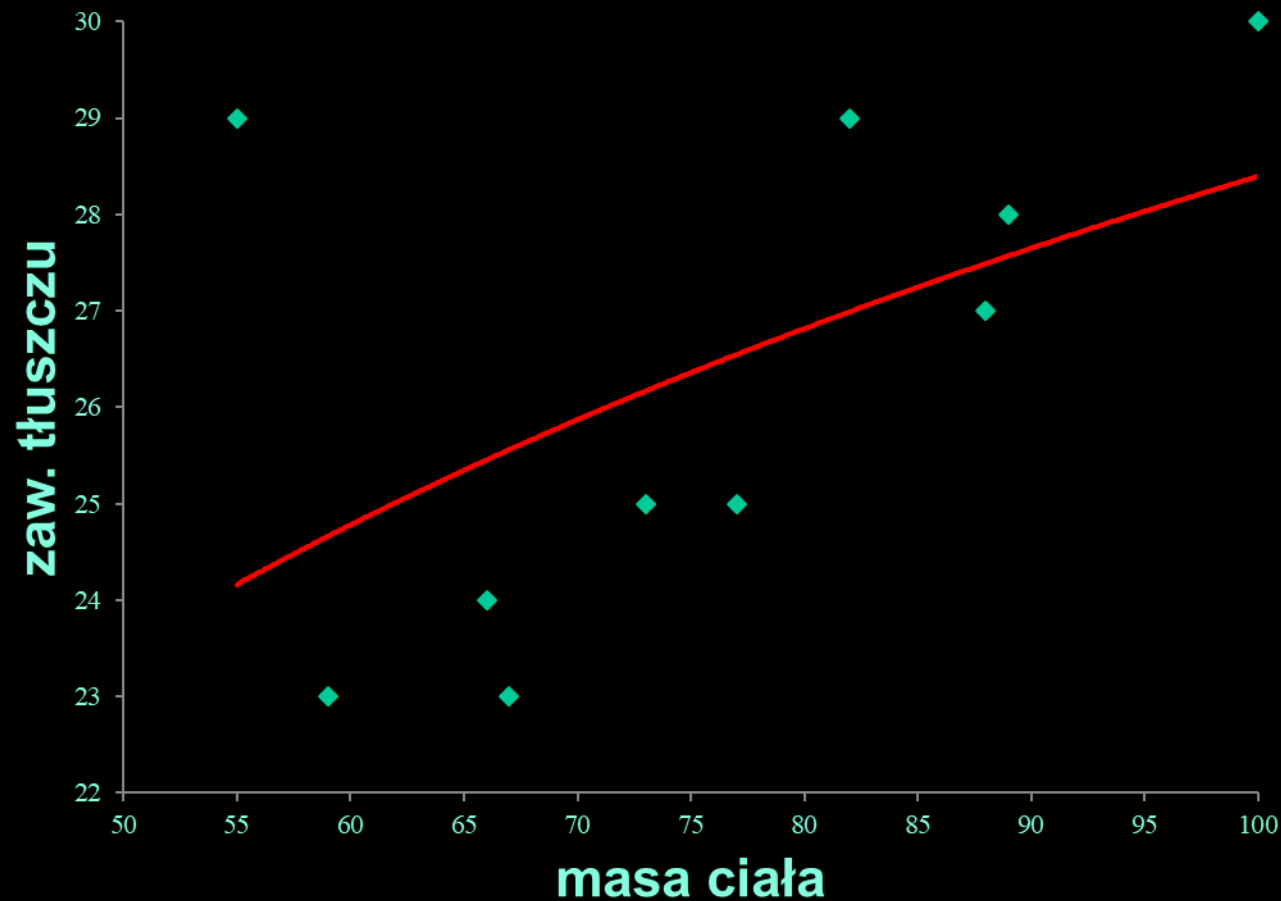


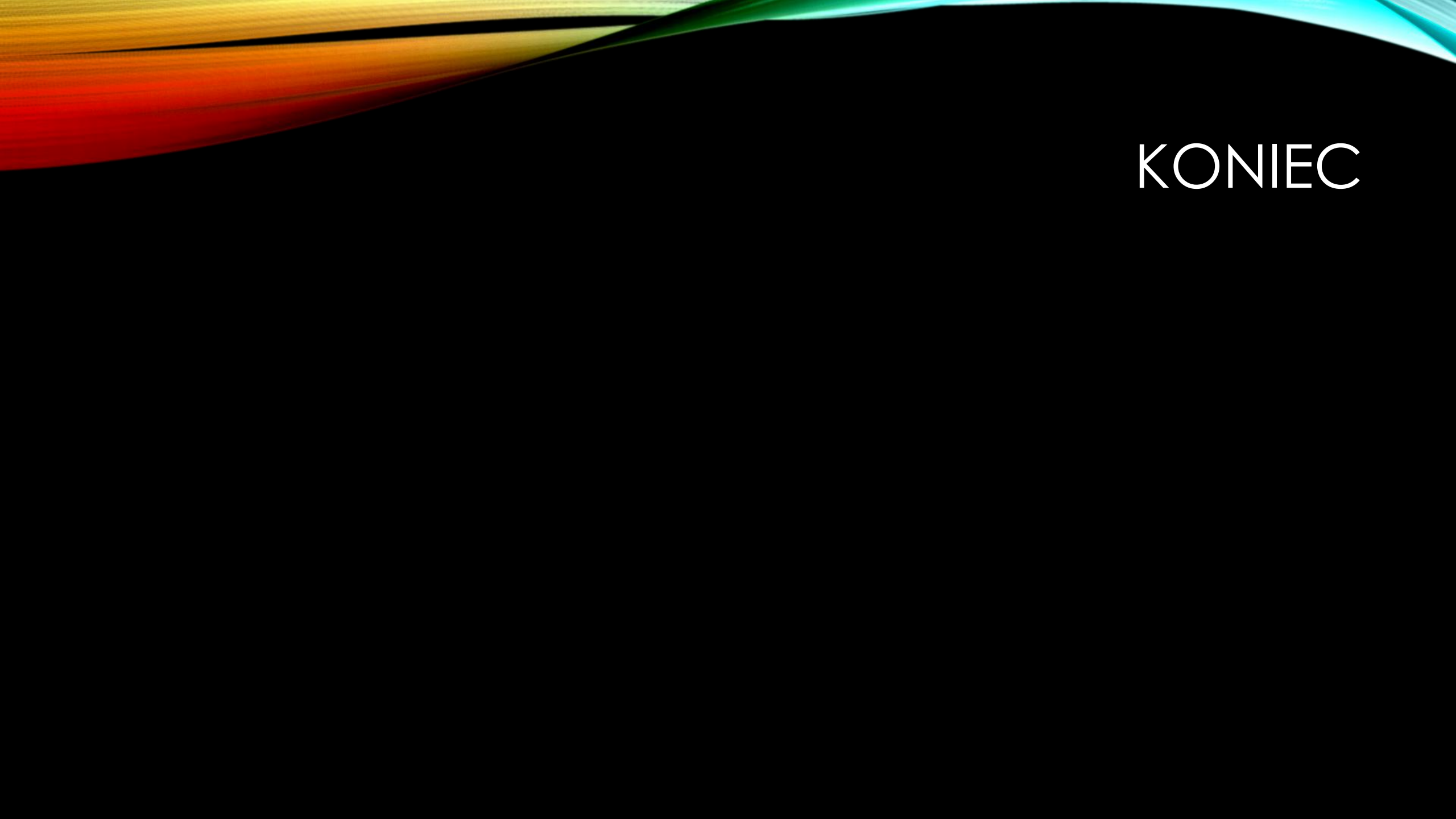
wielomian
3go stopnia



REGRESJA LOGARYTMICZNA

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





KONIEC