METODY STATYSTYCZNE W BIOLOGII

1. Wykład wstępny
2. Populacje i próby danych
3. Testowanie hipotez i estymacja parametrów
4. Planowanie eksperymentów biologicznych
5. Najczęściej wykorzystywane testy statystyczne I
6. Najczęściej wykorzystywane testy statystyczne II
7. Regresja liniowa
8. Regresja nieliniowa
9. Określenie jakości dopasowania równania regresji liniowej i nieliniowej
10. Korelacja
11. Elementy statystycznego modelowania danych
12. Porównywanie modeli
13. Analiza wariancji
14. Analiza kowariancji
15. Podsumowanie materiału, wspólna analiza przykładów, dyskusja

data i termin egzaminu

Wstęp

1. Równanie regresji liniowej

2. Estymacja współczynników prostej regresji
   - metoda najmniejszych kwadratów
   - ważona metoda najmniejszych kwadratów

3. Przykłady równań regresji
   - regresja wieloraka
   - wielomiany
   - regresja logarytmiczna

4. Interpretacja przykładowych wyników

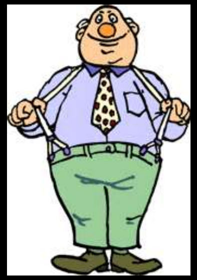
elementy równania regresji

Równanie regresji

próba danych

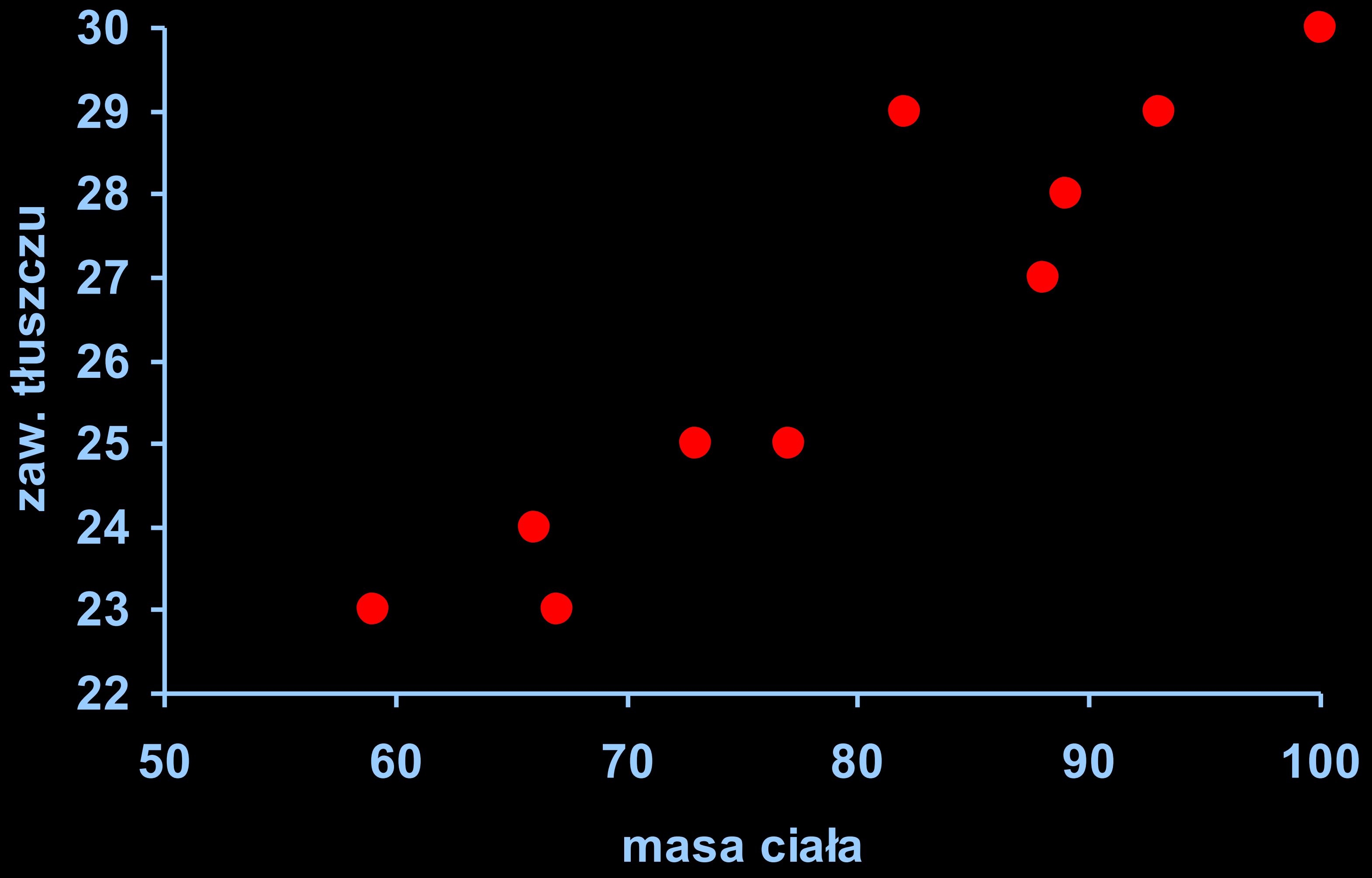
MASA CIAŁA	ZAW. TŁUSZCZU
89	28
88	27
66	24
59	23
93	29
73	25
82	29
77	25
100	30
67	23

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




Równanie regresji

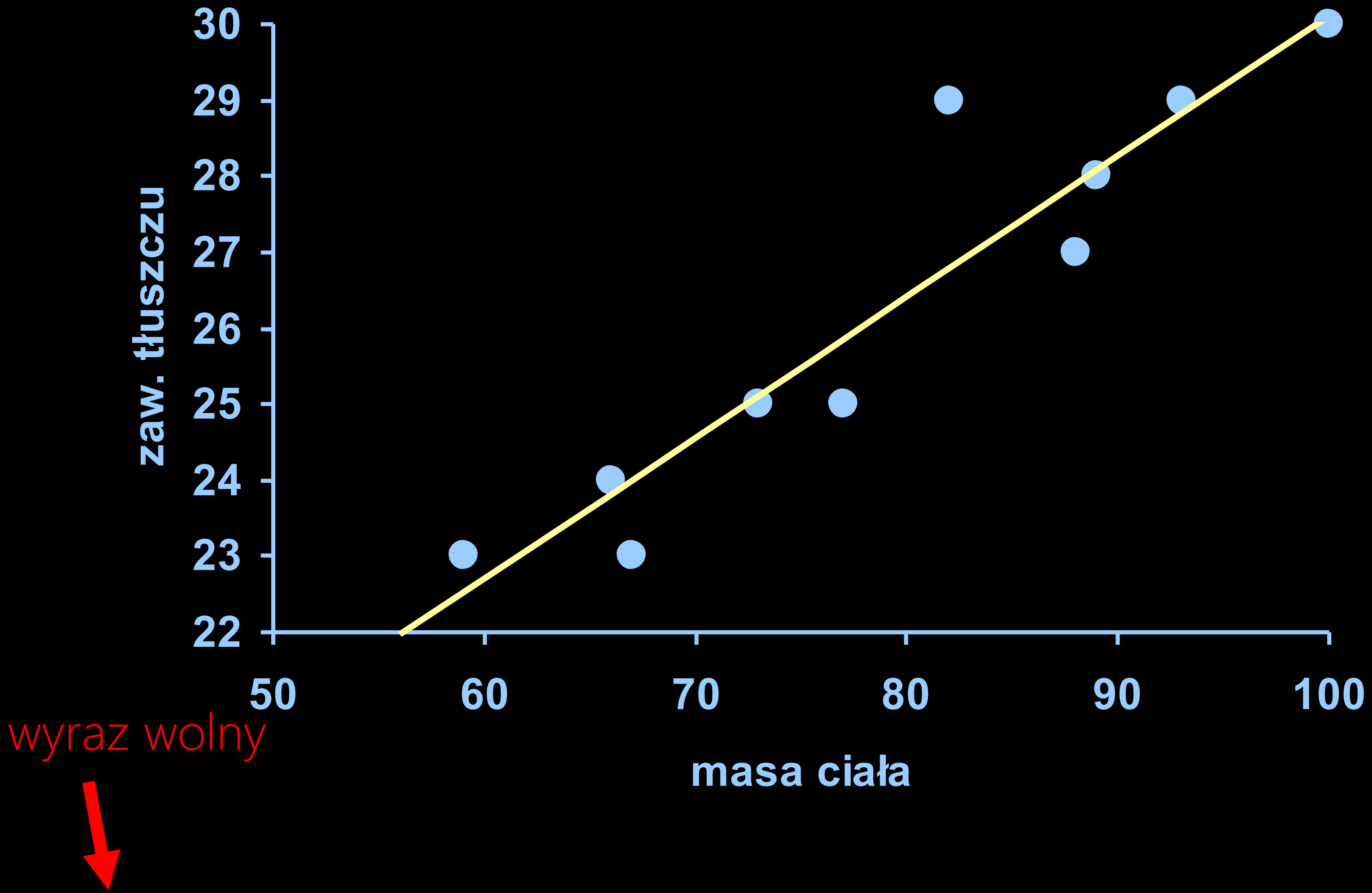
próba danych



Równanie regresji

elementy równania regresji: wyraz wolny

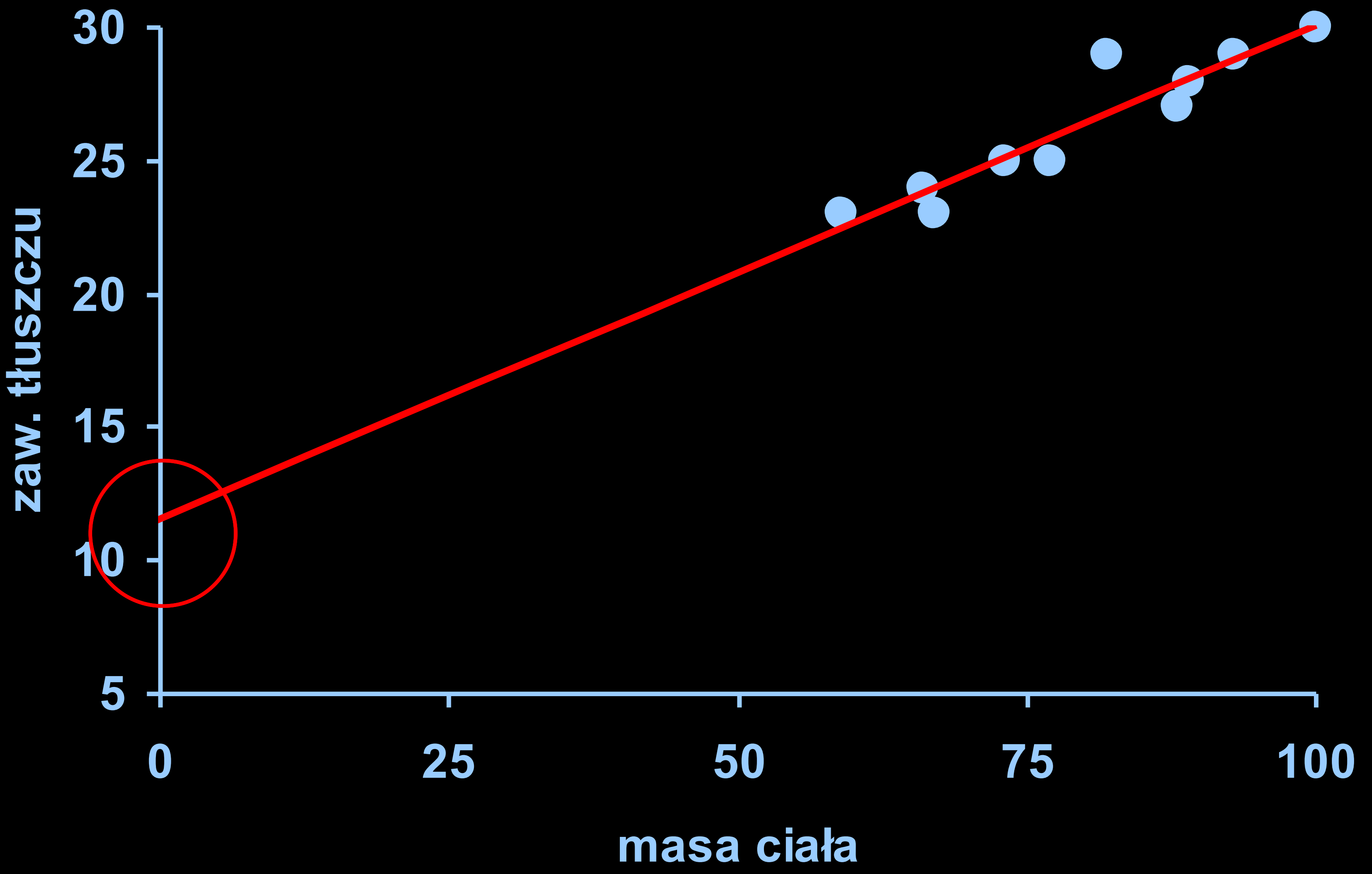
Intercept



Równanie regresji

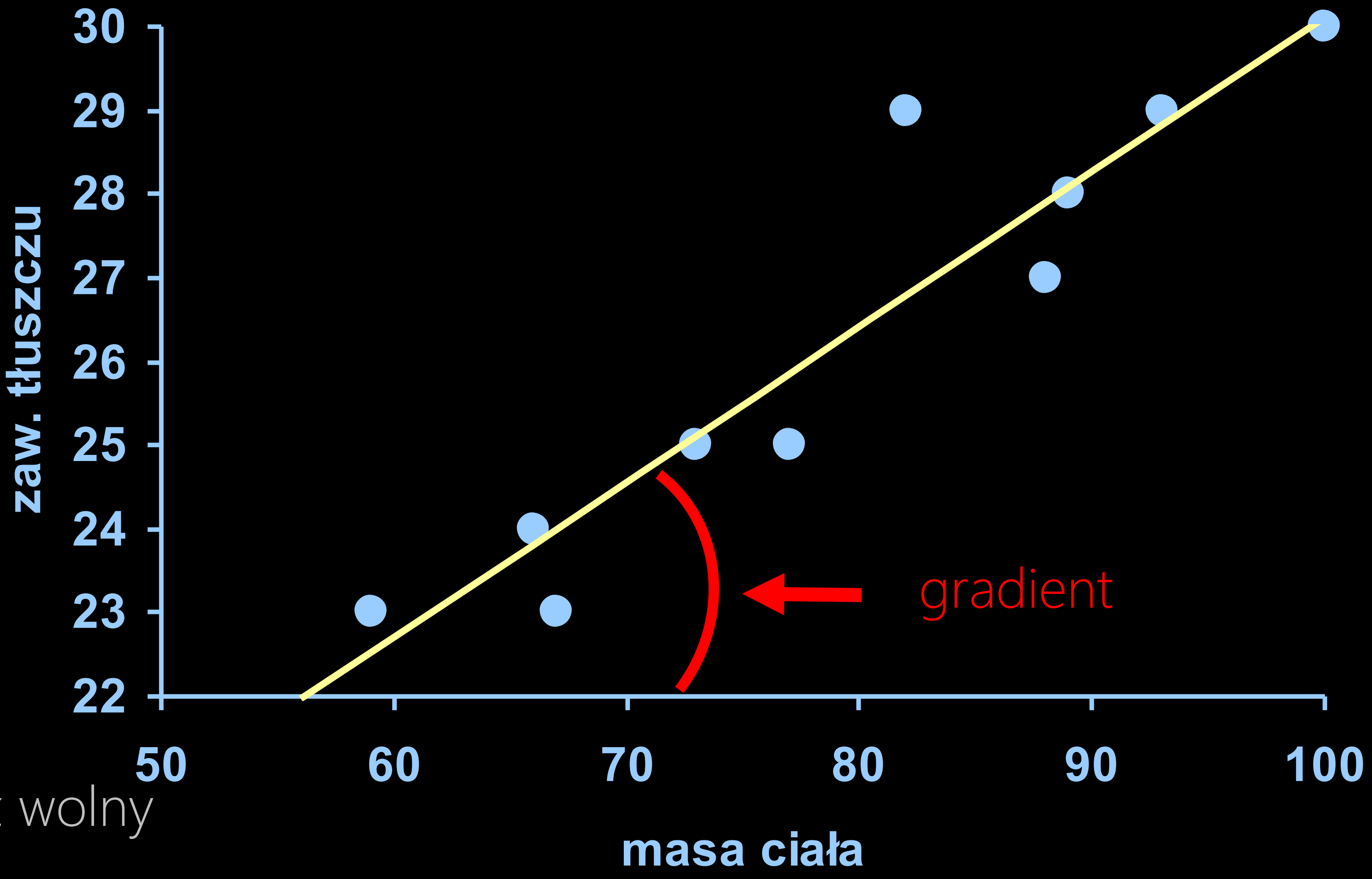
elementy równania regresji: wyraz wolny

Intercept



Równanie regresji

elementy równania regresji: gradient
Slope

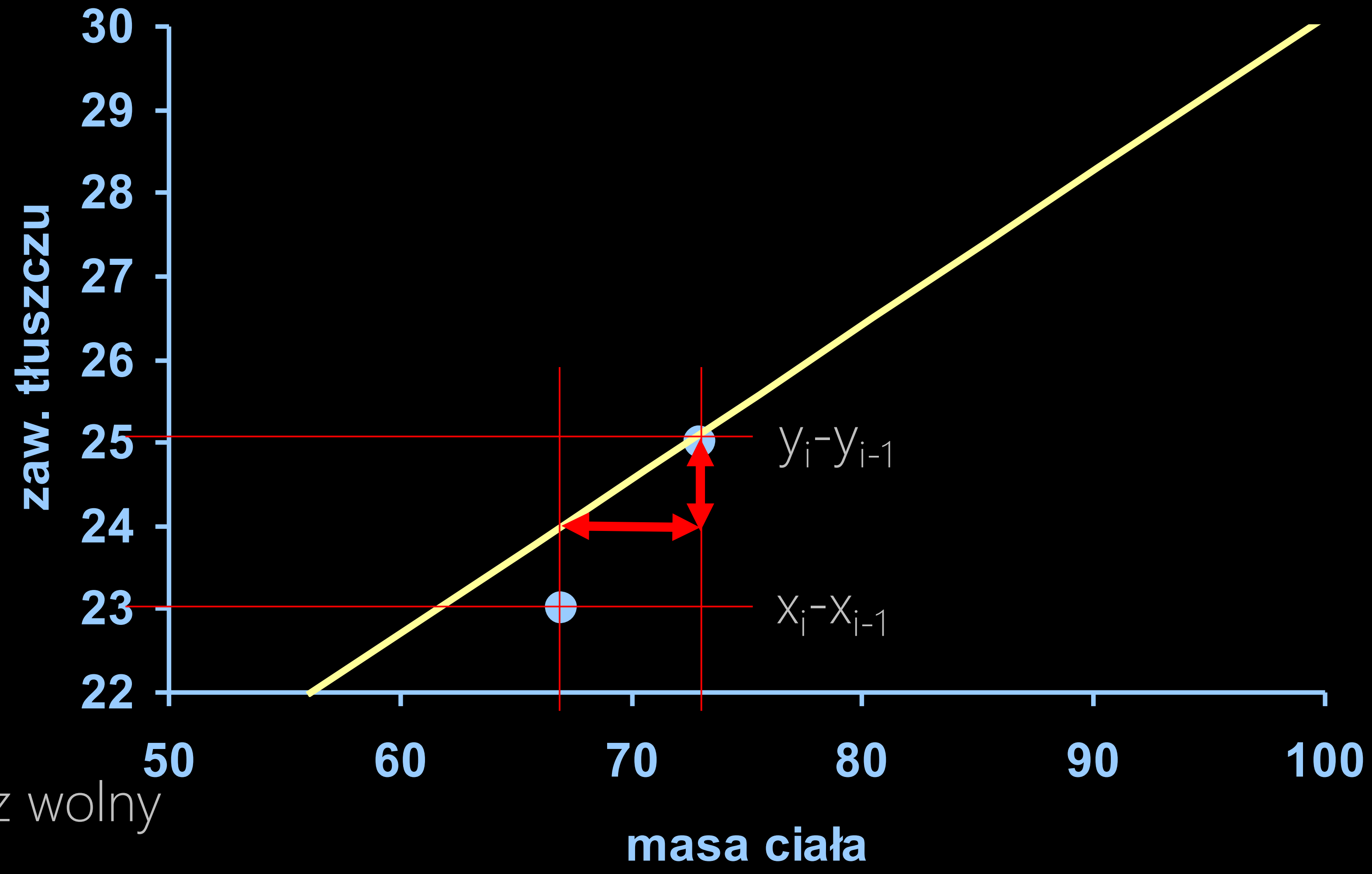


wyraz wolny

Równanie regresji

elementy równania regresji: gradient

Slope

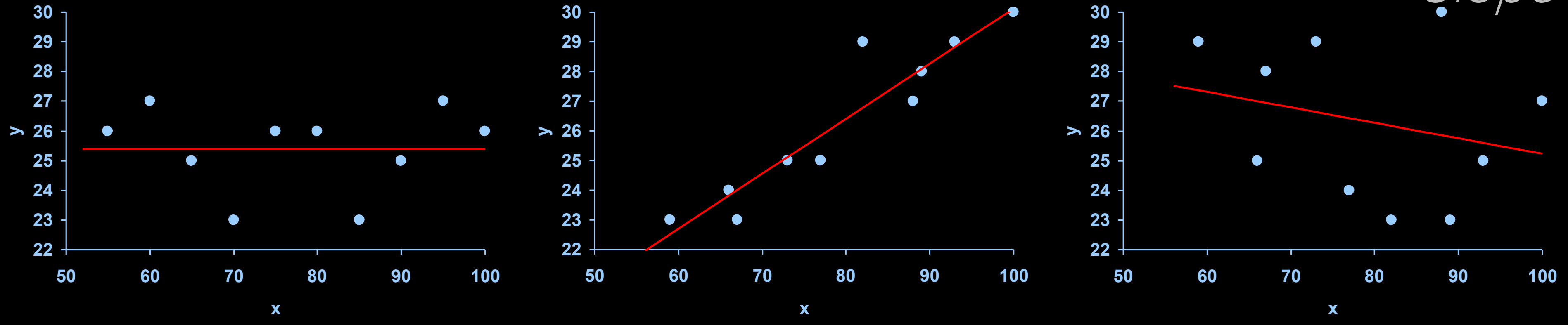


wyraz wolny

Równanie regresji

elementy równania regresji: gradient

Slope



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

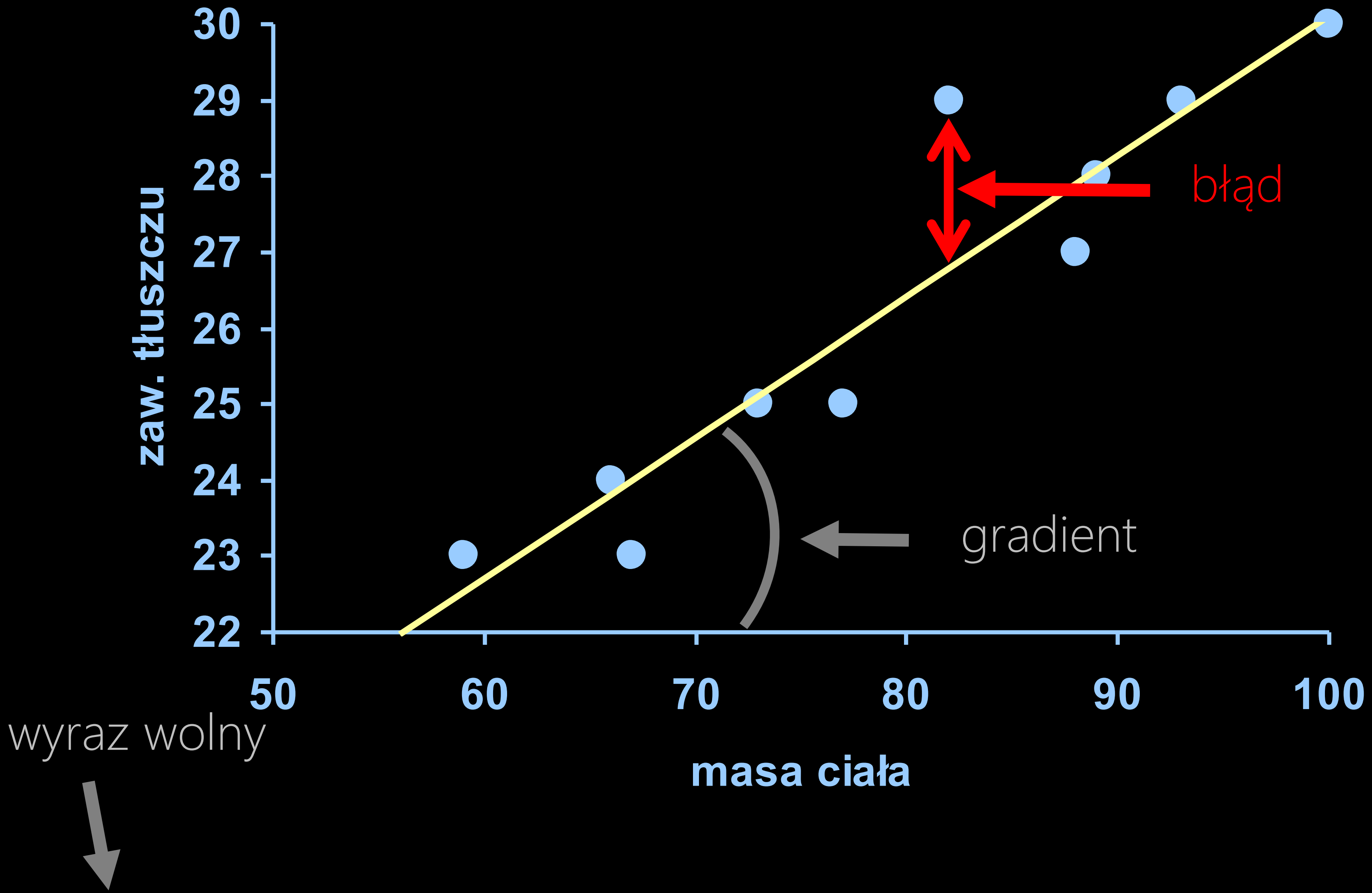
≈ 0 Brak zależności gradient = 0	> 0 Zmiany w tym samym kierunku gradient > 0	< 0 Zmiany w różnych kierunkach gradient < 0

przykłady

Równanie regresji

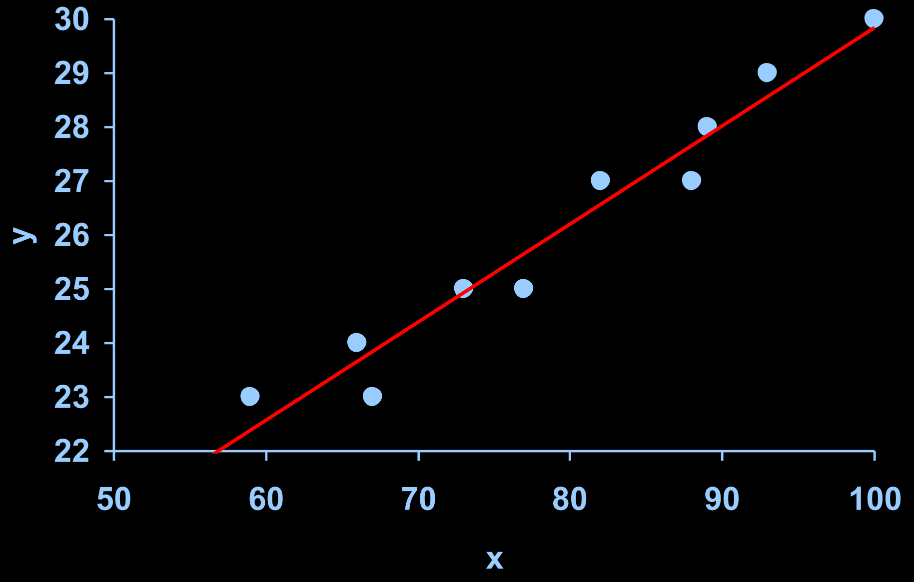
elementy równania regresji: błąd

Error, Residual

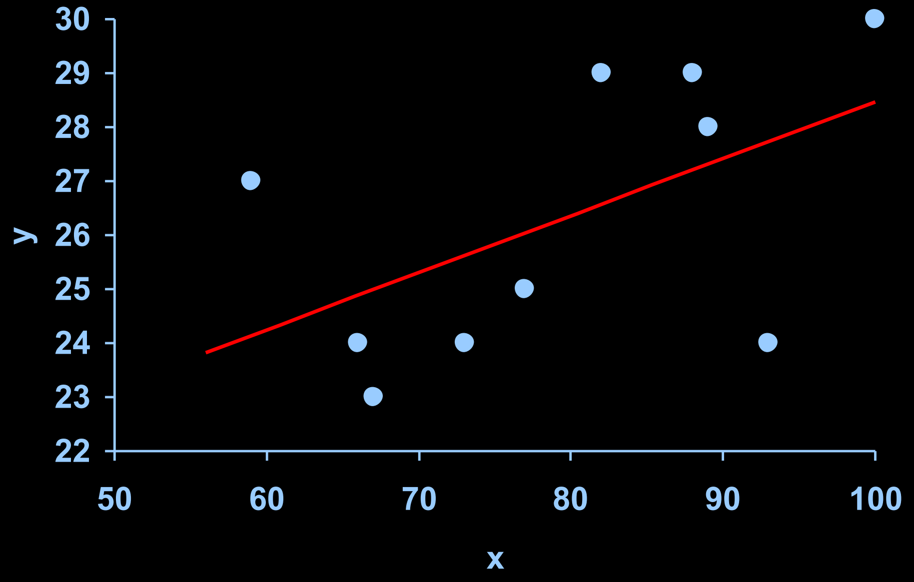


Równanie regresji

elementy równania regresji: błąd
Error, Residual



mały błąd

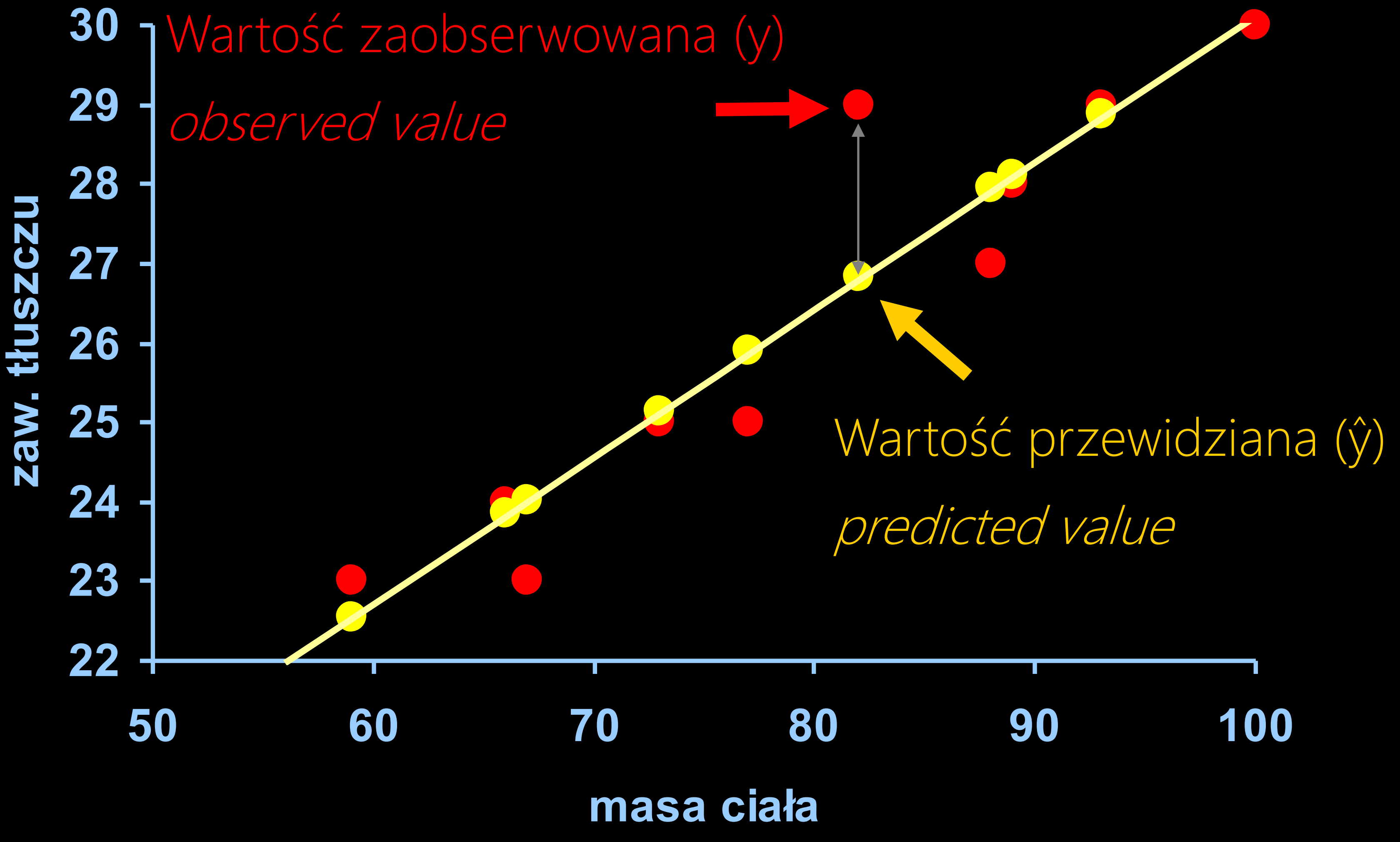


duży błąd

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

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

elementy równania regresji

wyraz wolny

gradient

$$tluszcz = 11.57 + 0.19 masa\_ciała$$

Zmienna zależna

Zmienna niezależna

Co mówią nam poszczególne wartości?

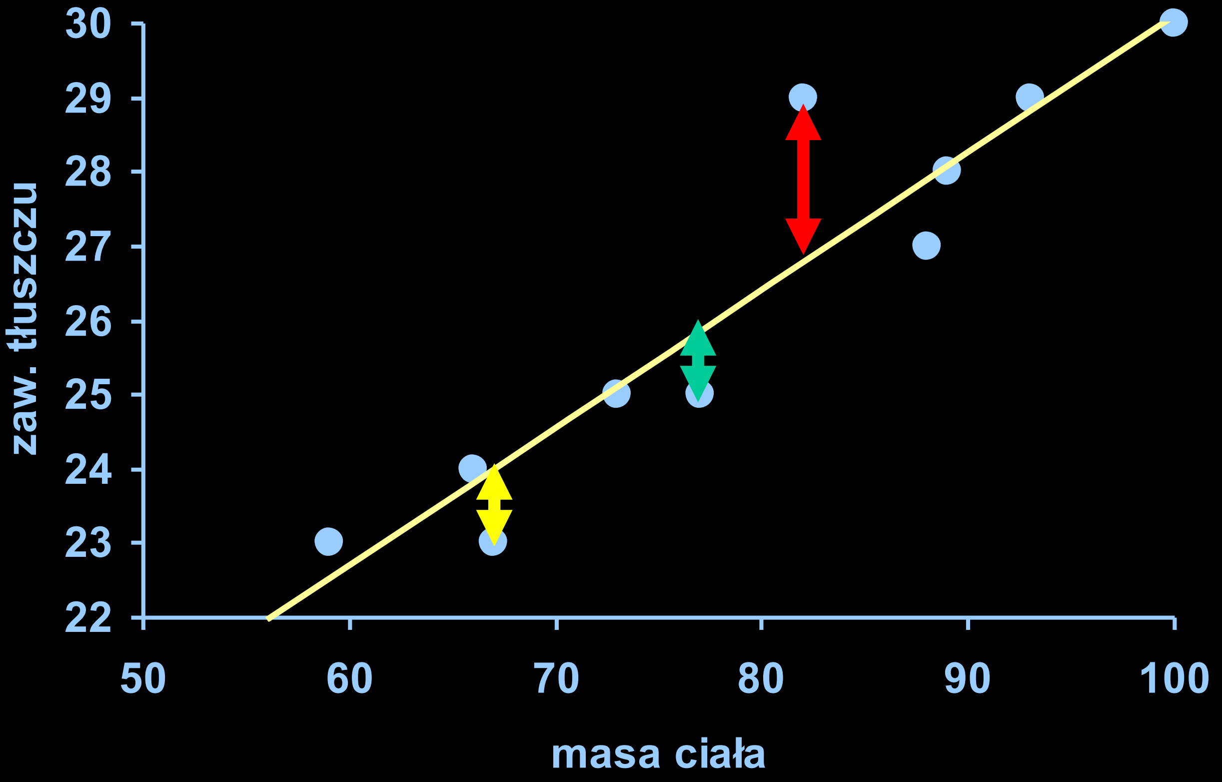
estymacja współczynników równania regresji

Metoda najmniejszych kwadratów

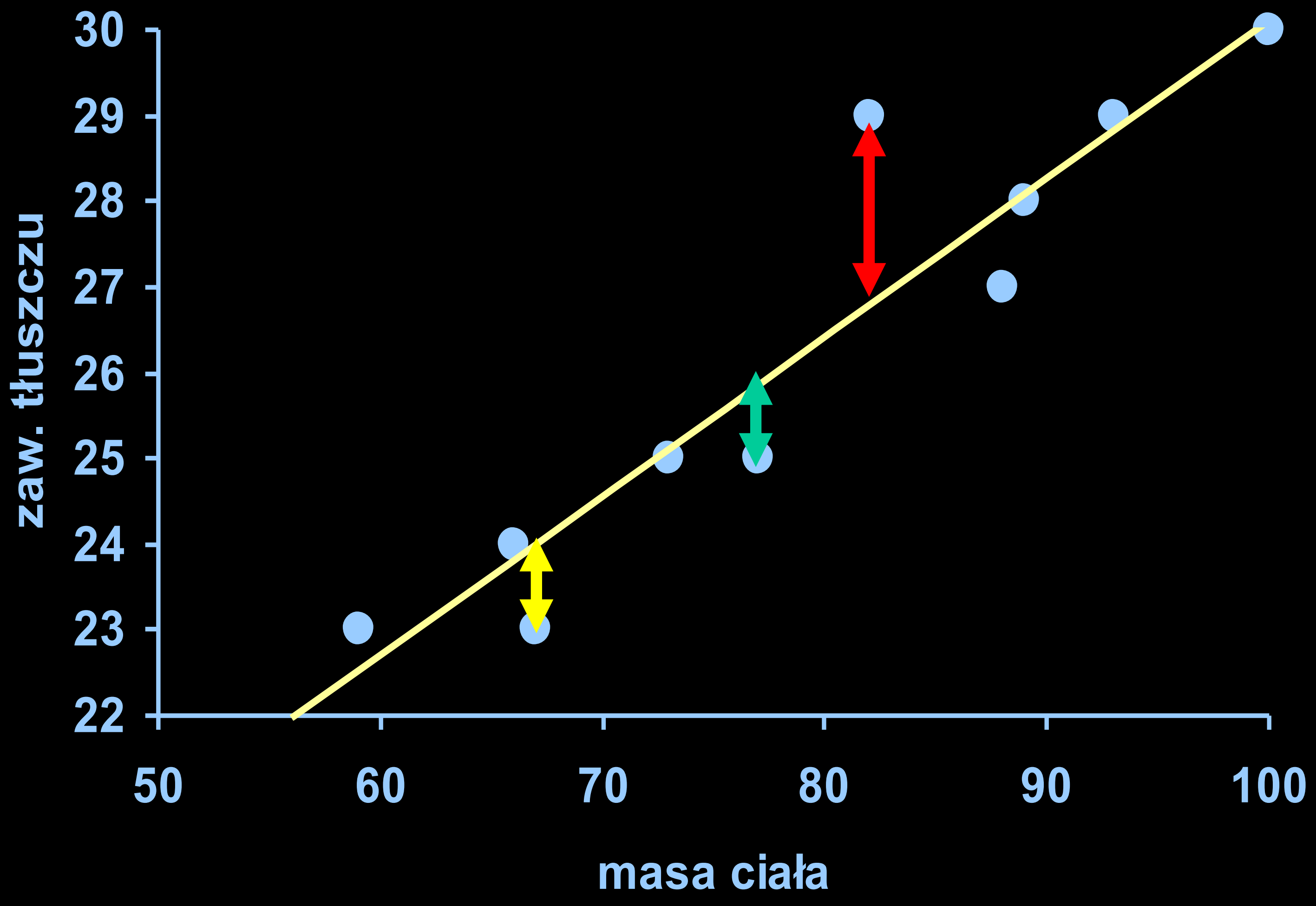
Jak wyznaczyć parametry równania regresji ?

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

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



Metoda najmniejszych kwadratów



$$\Sigma\ (\updownarrow)^2 \rightarrow \text{minimum}$$

Metoda najmniejszych kwadratów

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

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

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

Metoda najmniejszych kwadratów

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

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

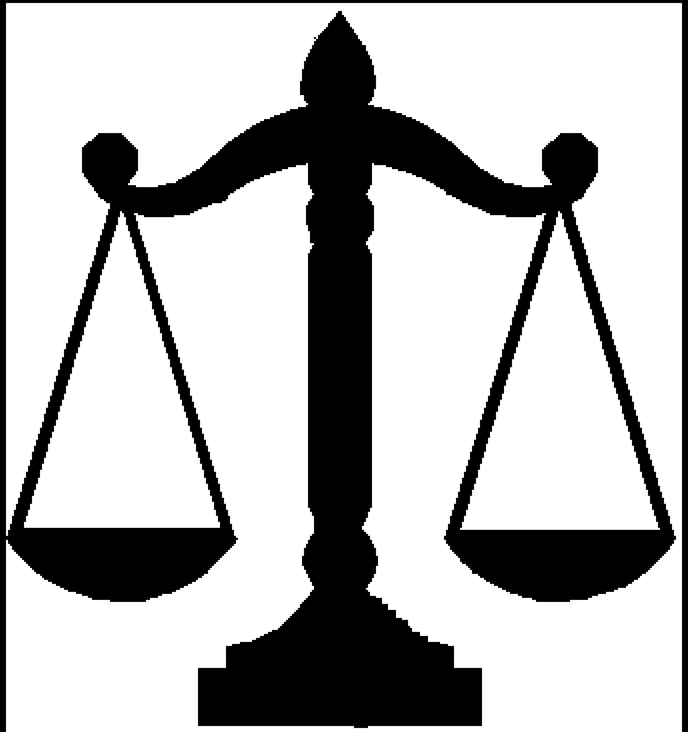
$$\frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Ważona metoda najmniejszych kwadratów

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

$$\mathrm{y} = \beta_0 + \beta_1 x + e$$

ważenie obserwacji



Ważona metoda najmniejszych kwadratów

ważenie obserwacji

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

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

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

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

przykłady równań regresji

Przykłady równań regresji

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

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

$$\mathrm{tluszcz} = \beta_0 + \beta_1 \mathrm{masa\_c} + \beta_2 \mathrm{wiek}$$

Przykłady równań regresji

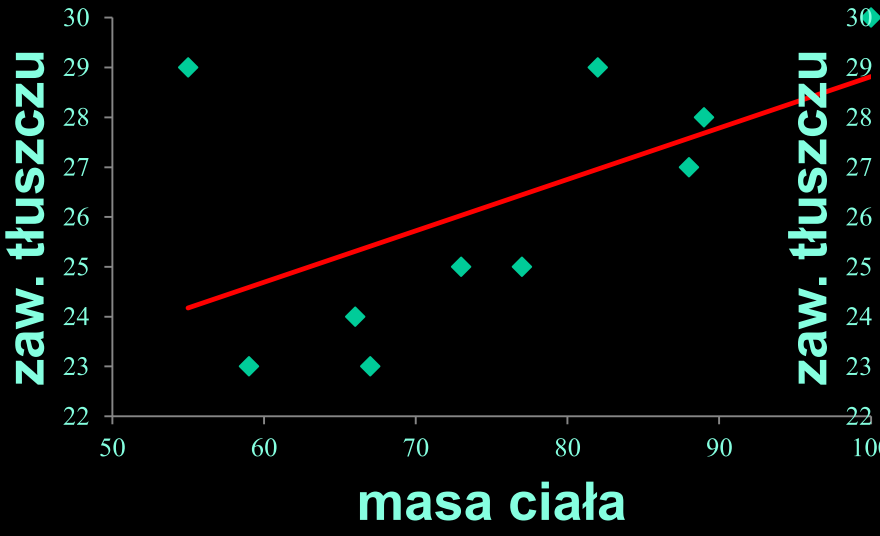
wielomiany

$$\boxed{y = \beta_0 + \beta_1 x}$$

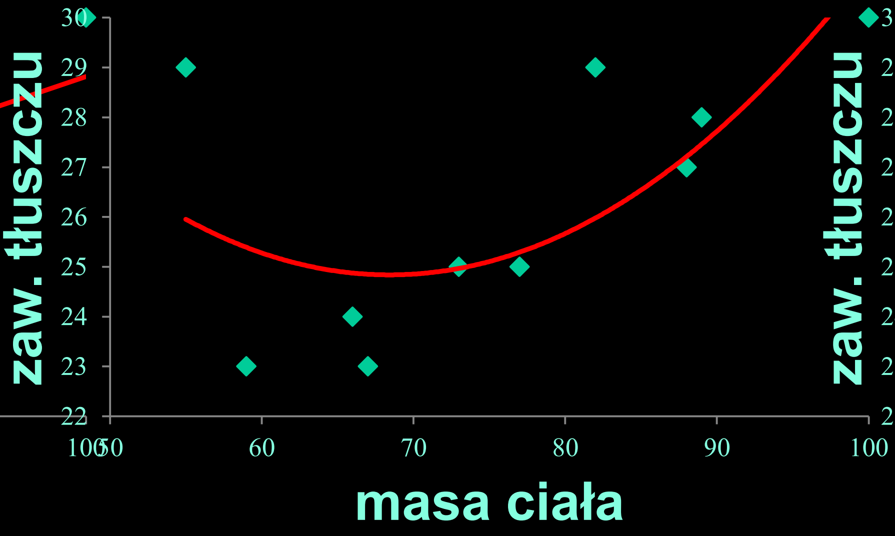
$$\boxed{y = \beta_0 + \beta_1 x + \beta_2 x^2}$$

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

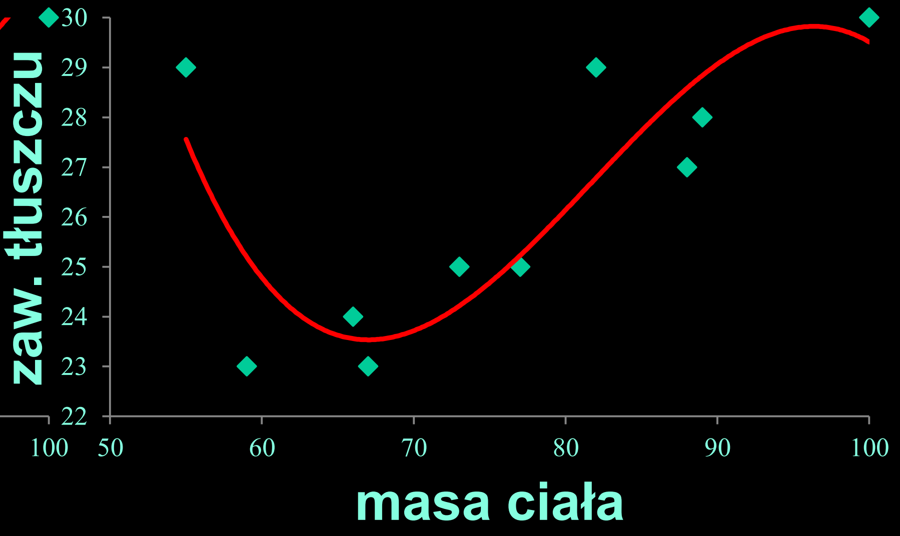
wielomian
1go stopnia



wielomian
2go stopnia



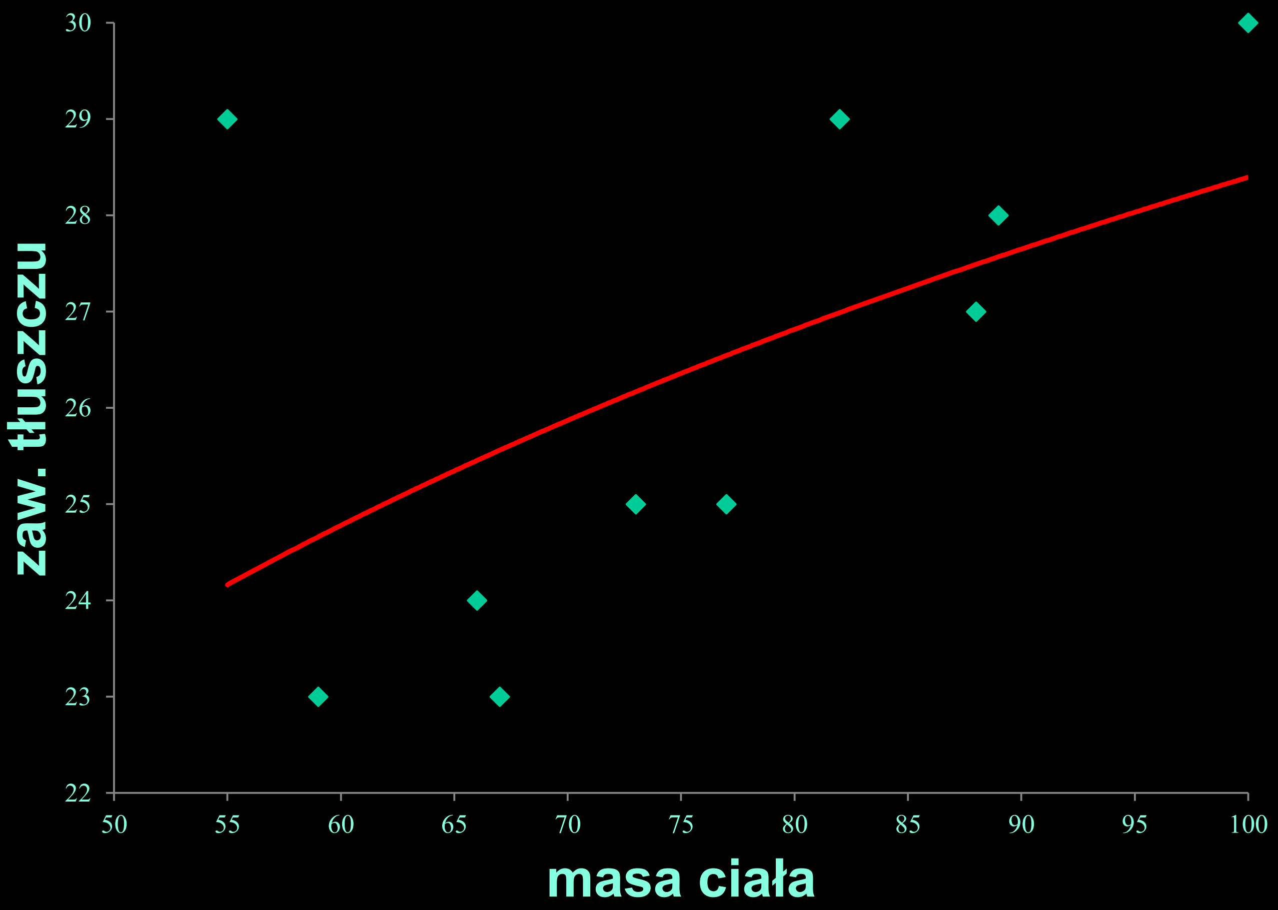
wielomian
3go stopnia



Przykłady równań regresji

regresja logarytmiczna

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

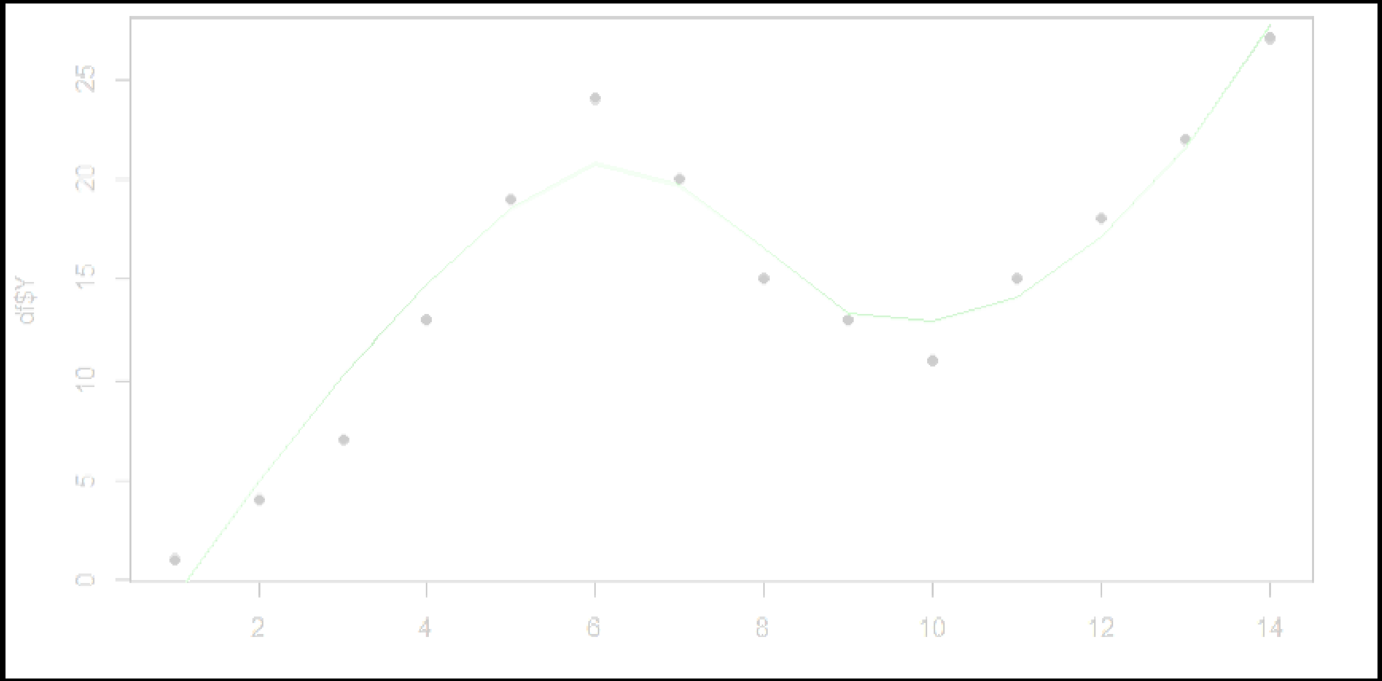


Przykłady równań regresji

LOESS
Locally Weighted Linear Regression

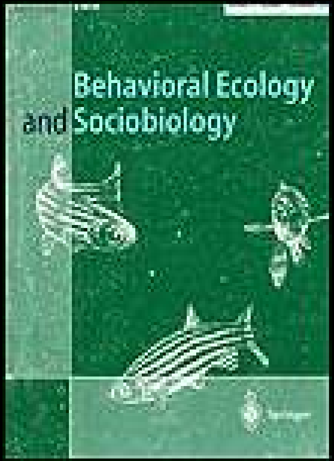
- Złożona zależność miedzy Y, a X
- Dopasowanie lokalnych równań regresji dla każdej wartości zmiennej X
- Wykorzystując tylko najbliżej położone wartości danych

R: `loess(Y ~ X, data=df)`



interpretacja wyników

Interpretacja równań regresji



Behav Ecol Sociobiol (1986) 19:9–18

Behavioral Ecology and Sociobiology

The effect of social interactions on calling energetics in the gray treefrog (*Hyla versicolor*)

Kentwood D. Wells and Theodore L. Taigen

Department of Ecology and Evolutionary Biology, U-43, The University of Connecticut, Storrs, CT 06268, USA

Received June 12, 1985 / Accepted November 12, 1985

Summary. The vocal behavior of *Hyla versicolor* was studied in the field by means of behavioral observations and playback experiments, and these data were coupled with measurements of oxygen consumption in calling frogs to estimate the effect of social interactions on calling energetics. Male gray treefrogs have intense calls (median peak SPL = 109 dB, fast RMS SPL = 100 dB at 50 cm). At an air temperature of 23° C, males produced an average of 1,200–1,300 calls/h for 2–4 h per night. Calling rates and call durations differed among individuals, but were relatively constant for each male during periods of sustained calling. Males in dense choruses gave calls about twice as long as isolated males, but produced calls at about half the rate. Consequently, total calling effort and estimated aerobic costs were largely independent of chorus density. Playbacks of recorded calls to males in the field elicited increases in call duration

Introduction

Male animals engaged in competitive displays often change their behavior in response to the behavior of other males. Intrasexual competition may cause males to increase the frequency, intensity, duration, or complexity of visual or acoustic displays (e.g., fishes: Farr 1980; birds: Bradbury and Gibson 1983; insects: Alexander 1975; Greenfield and Shaw 1983; frogs: Rand and Ryan 1981; Schwartz and Wells 1984, 1985; Wells and Schwartz 1984). These socially mediated changes in behavior can affect a male's ability to obtain mates. For example, in both insects and frogs, females may be differentially attracted to males producing the longest, most intense, most frequent, or most complex signals (Rand and Ryan 1981; Arak 1983a, b; Sullivan 1983; Thornhill and Alcock 1983; Wells and Schwartz 1984; Forester and

Interpretacja równań regresji

1. Wraz ze wzrostem temperatury czas pomiędzy odgłosami skraca się
2. Wzrost temperatury o 1°C powoduje skrócenie odstępu o 0.21 s.
3. W temperaturze 10°C osobniki wydają dogłosy średnio co 6.26 s. :

$\text{interval} = 8.36 - 0.21 \cdot$ [illegible] $= 6.26$

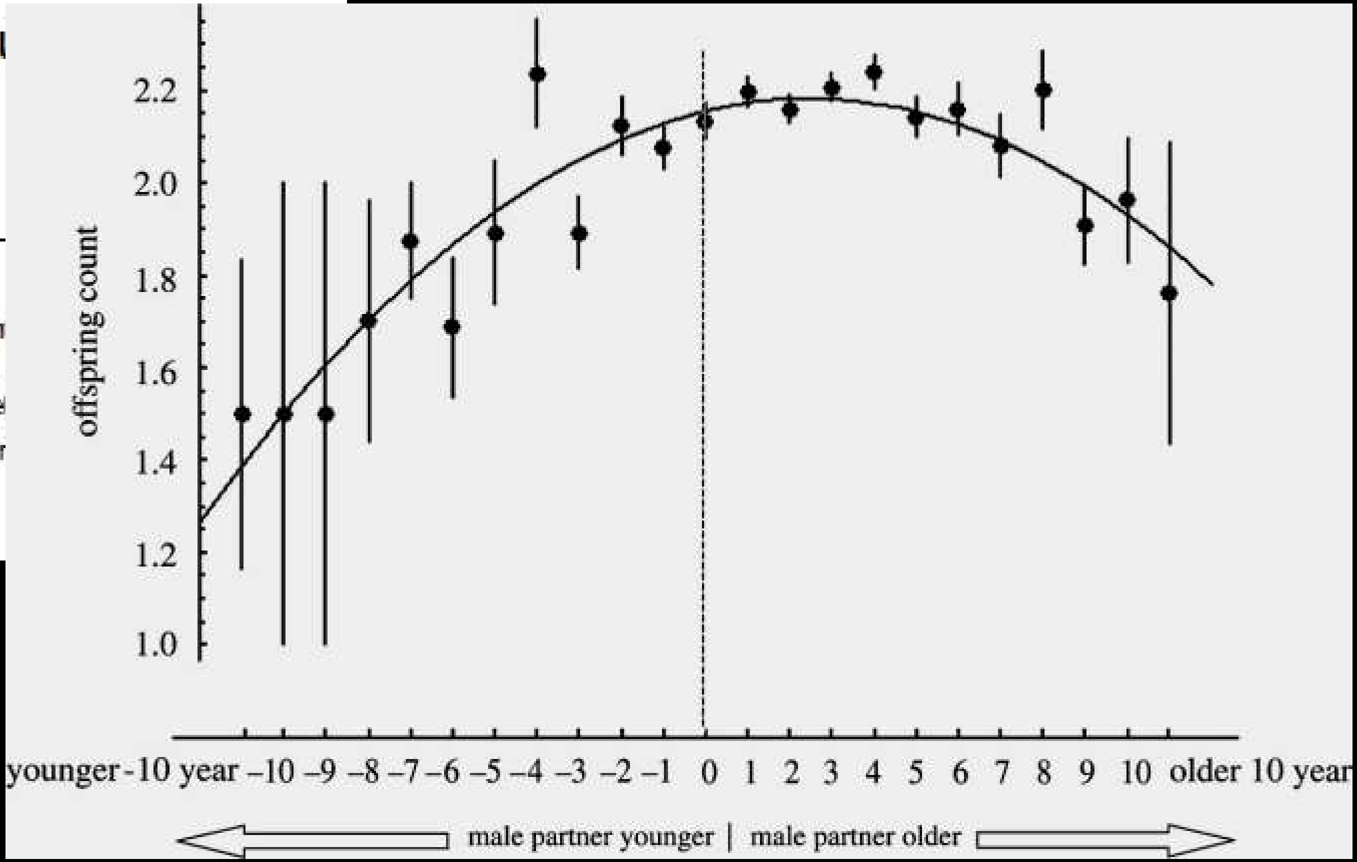
Interpretacja równań regresji

biology letters

Parental age difference and offspring count in hu

Martin Fieder and Susanne Huber

Biol. Lett. 2007 **3**, 689-691
doi: 10.1098/rsbl.2007.0324



Interpretacja równań regresji

1. Model:

$$n\_dzieci = 2.82 - 0.01wiek + 0.03r\_w - 0.002r\_w^2$$

2. Partnerzy, którzy w 2003 r. mieli średnio 40 lat oraz, u których mąż jest starszy o 15 mają średnio 2.42 dzieci:

$$n\_dzieci = 2.82 - 0.01 \cdot \boxed{\;} + 0.03 \cdot \boxed{\;} - 0.002 \cdot \boxed{\;}^2 = 2.42$$

3. Partnerzy, którzy w 2003 r. mieli średnio 25 lat oraz, u których mąż jest starszy o 4 mają średnio 2.66 dzieci:

$$n\_dzieci = 2.82 - 0.01 \cdot \boxed{\;} + 0.03 \cdot \boxed{\;} - 0.002 \cdot \boxed{\;}^2 = 2.66$$

1. Równanie regresji liniowej

2. Estymacja współczynników prostej regresji
   - metoda najmniejszych kwadratów
   - ważona metoda najmniejszych kwadratów

3. Przykłady równań regresji
   - regresja wieloraka
   - wielomiany
   - regresja logarytmiczna

4. Interpretacja przykładowych wyników