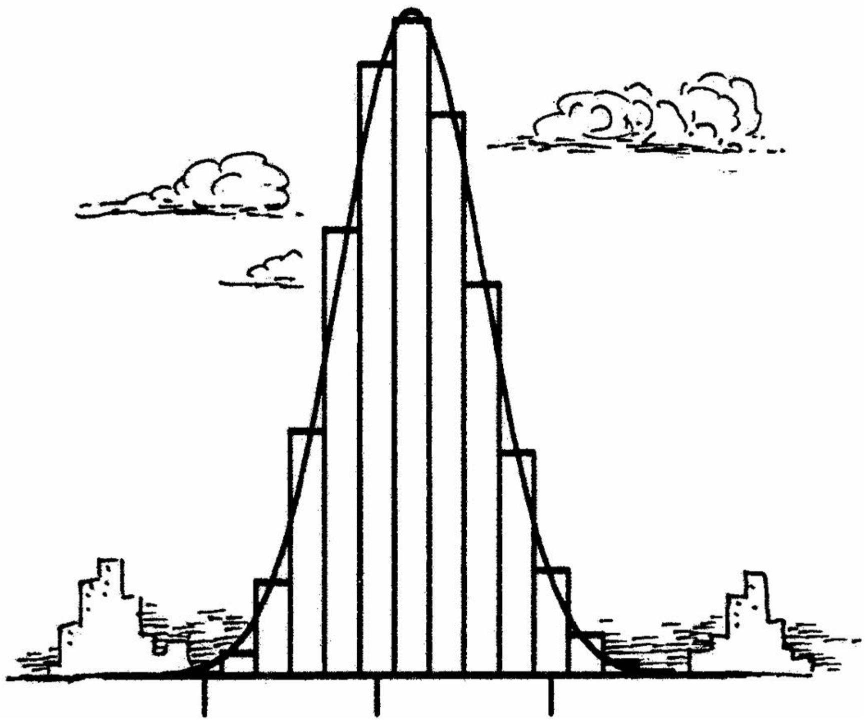
Razpršenost



Milena Kovač

12. oktober 2012

Mere razpršenosti

1. Razmiki
2. Absolutni odklon
3. Varianca
4. Standardni odklon
5. Standardna napaka ocena
6. Koeficient variabilnosti
7. Kvantili

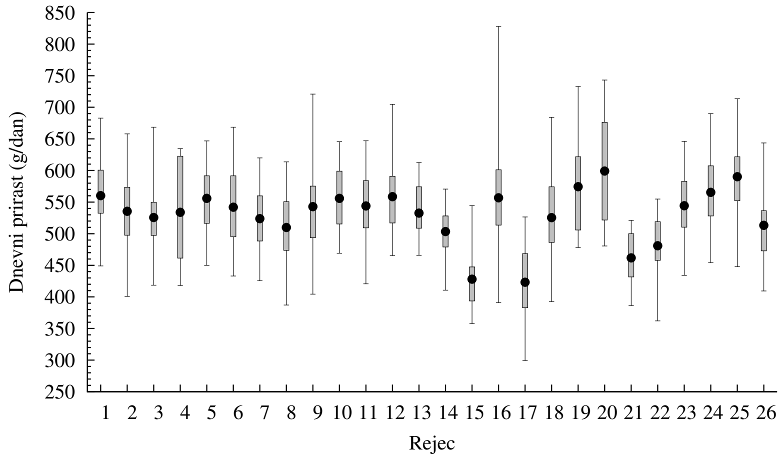
Razmiki

- Variacijska širina (VS): $VS = x_{max} - x_{min}$

- Kvartilni razmik (Q): $Q = Q_3 - Q_1$

```
|------------|------------|------------|------------|
xmin          Q1           Q2           Q3          xmax
  \__________/\______________________/\__________/
       25 %                50 %                25 %
```

Primer - Kvartilni razmik



Obrazložimo sliko!

Narišite porazdelitev za rejce 2, 4, 16 in 22.

Povprečni odklon?

$$\overline{O} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}) =$$

- Odklon = odstopanje meritve (x_i) od povprečja$(\overline{x})$
- Dobimo pozitivne in negativne odklone ...
- $= \frac{1}{n}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\overline{x}\right] = \frac{n\overline{x}}{n} - \frac{n\overline{x}}{n} = 0$
- Poleg povprečja še druge vrednosti, pri katerih je vsota odklonov 0
- Torej: **povsem neuporabna statistika**

Absolutni odklon

Absolutni odklon je povprečje absolutnih vrednosti odklonov.

$$AO = \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}|$$

- odklon = odstopanje meritve od povprečja
- vsi odkloni z enako težo
- za srednjo vrednost uporabimo mediano

Varianca

Varianca je povprečni kvadratni odklon.

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

Oznaka σ^2

Drugi parameter pri normalni porazdelitvi: $N(\mu, \sigma^2)$

Varianca

Varianca je povprečni kvadratni odklon

$$(x_i - \overline{x})$$

Varianca

Varianca je povprečni kvadratni odklon

$$(x_i - \overline{x})^2$$

Varianca

Varianca je povprečni kvadratni odklon

$$= \frac{\sum_{i=1}^{n}}{n-1}$$

Varianca

Varianca je povprečni kvadratni odklon

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

- pri vzorcu izračunamo varianco vzorca
- ocenjuje parameter varianco populacije

Varianca

- pri vzorcih iz različnih populacij upoštevamo različne pričakovane vrednosti

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - E(x_i))^2}{n-p}$$

p = število "različnih pričakovanih vrednosti, ki jih lahko izračunamo"[1]

[1] bolj natančno bomo to obdelali kasneje

Preuredimo enačbe

$$var\left(x\right)=\frac{\sum\left(x_i-\overline{X}\right)\left(x_i-\overline{X}\right)}{n-1}=$$

pomnožimo

$$\frac{\sum\left(x_ix_i-\overline{X}x_i-x_i\overline{X}+\overline{XX}\right)}{n-1}=$$

razstavimo

$$=\frac{\sum\left(x_i^2\right)-2\sum\left(\overline{X}x_i\right)+\sum\left(\overline{X}^2\right)}{n-1}=$$

Preuredimo enačbe (nadalj.)

povprečje prestavimo pred vsoto

$$= \frac{\sum(x_i^2) - 2\overline{X}\sum(x_i) + \overline{X}^2\sum(1)}{n-1} =$$

ker velja $\sum(x_i) = n\overline{X}$ in $\sum(1) = n$, lahko zapišemo

$$= \frac{\sum(x_i^2) - 2n\overline{X}^2 + n\overline{X}^2}{n-1} =$$

Še dve enačbi za varianco

sešteljemo zadnja dva člena

$$= \frac{\sum\left(x_i^2\right) - n\overline{X}^2}{n-1} = var\left(x\right)$$

zadnji člen lahko preoblikujemo

$$= \frac{\sum\left(x_i^2\right) - \frac{1}{n}\sum^2\left(x_i\right)}{n-1} = var\left(x\right) = \sigma^2$$

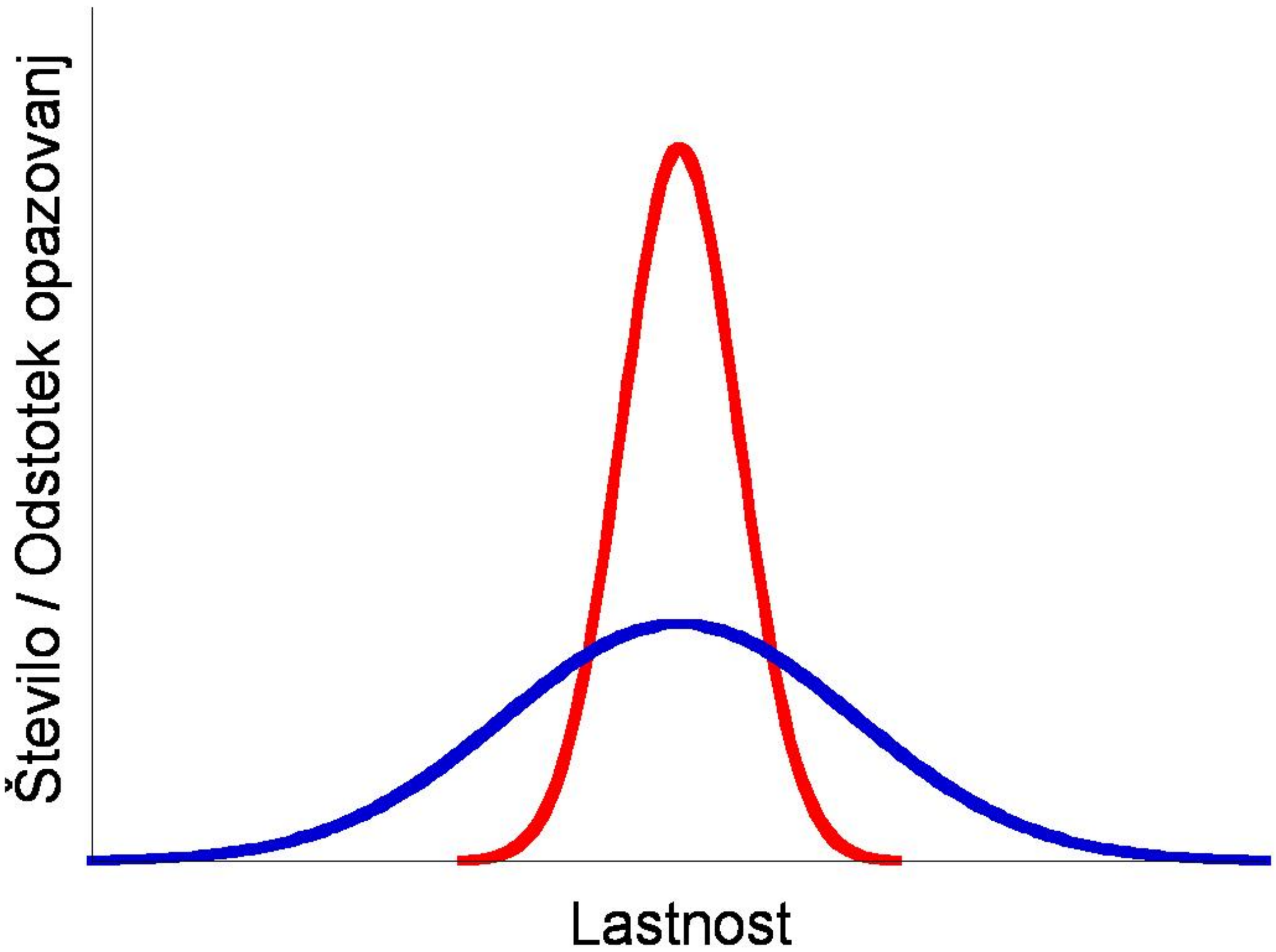
Standardni odklon

Standardni odklon oziroma standardna deviacija je pozitivna vrednost kvadratnega korena iz variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{var(x)}$$

- primerna za ponazoritev razpršenosti

Ponazoritev standardnega odklona



Komentar:

Masa rogovja jelenov

Jeleni: Vrnimo se k poskusu z jeleni!

Revir	Jelen	Masa (kg)	Revir	Jelen	Masa (kg)
1	1	13.0	2	9	10.6
1	2	10.3	2	10	10.2
1	3	12.0	2	11	13.7
1	4	13.3	2	12	12.7
1	5	14.0	2	13	14.1
1	6	12.4	2	14	14.2
1	7	14.2	2	15	12.0
1	8	13.5	2	16	10.4

Izračunajmo!

Skupina	Revir	n	Povp.	Varianca	Stand. odklon
A	1	8	12.8		
B	2	8	12.2		
C	3	16	12.5		

Masa rogovja jelenov

Jeleni: Vrnimo se k poskusu z jeleni!

Revir	Jelen	Masa (kg)	Revir	Jelen	Masa (kg)
1	1	13.0	2	9	10.6
1	2	10.3	2	10	10.2
1	3	12.0	2	11	13.7
1	4	13.3	2	12	12.7
1	5	14.0	2	13	14.1
1	6	12.4	2	14	14.2
1	7	14.2	2	15	12.0
1	8	13.5	2	16	10.4

Izračunajmo!

Skupina	Revir	n	Povp.	Varianca	Stand. odklon
A	1	8	12.8	1.60	1.27
B	2	8	12.2	2.85	1.69
C	3	16	12.5	2.17	1.47

Komentar:

Standardna napaka ocene

Standardna napaka ocene prikaže zanesljivost ocene povprečja

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- prikaže zanesljivost tudi drugih ocen
- odvisna od variance in števila opazovanj
- jo vedno navajamo pri prikazovanju rezultatov

$$\overline{X} \pm \sigma_{\bar{x}}$$

Dogovori in nekaj pomembnih enačb

oblike zapisa variance

$$var\left(x_i\right) = \sigma_x^2 = cov\left(x_i, x_i\right)$$

dve obliki zapisa kovariance

$$cov\left(x_i, x_{i'}\right) = \sigma_{ii'}$$

varianca identično porazdeljenih spremenljivk

$$var\left(x_1\right) = var\left(x_2\right) = \cdots = var\left(x_i\right)$$

kovarianca neodvisnih spremenljivk

$$cov\left(x_i, x_{i'}\right) = 0$$

... in še nekaj

varianca vsote

$$var\,(x_1 + x_2) =$$
$$var\,(x_1) + var\,(x_2) + 2 * cov\,(x_1, x_2)$$

$$= 2var\,(x) = 2\sigma_x^2 = 2\sigma^2$$

varianca s konstanto pomnožene spremenljivke

$$var\,(kx) = k^2 var\,(x)$$

Izpeljava variance povprečja

$$\sigma^2_{\bar{x}} = var(\bar{x}) = var\left(\frac{x_1 + x_2 + \cdots + x_i + \cdots + x_n}{n}\right) =$$

$$= \frac{1}{n^2} var(x_1 + x_2 + \cdots + x_i + \cdots + x_n) =$$

$$\frac{1}{n^2}\left(\sum_{i=1}^{n} var(x_i) + 2\underbrace{\sum_{i=1}^{n}\sum_{i'=i+1}^{n} cov(x_i, x_{i'})}\right) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n}{n^2}\sigma^2 = \frac{1}{n}\sigma^2$$

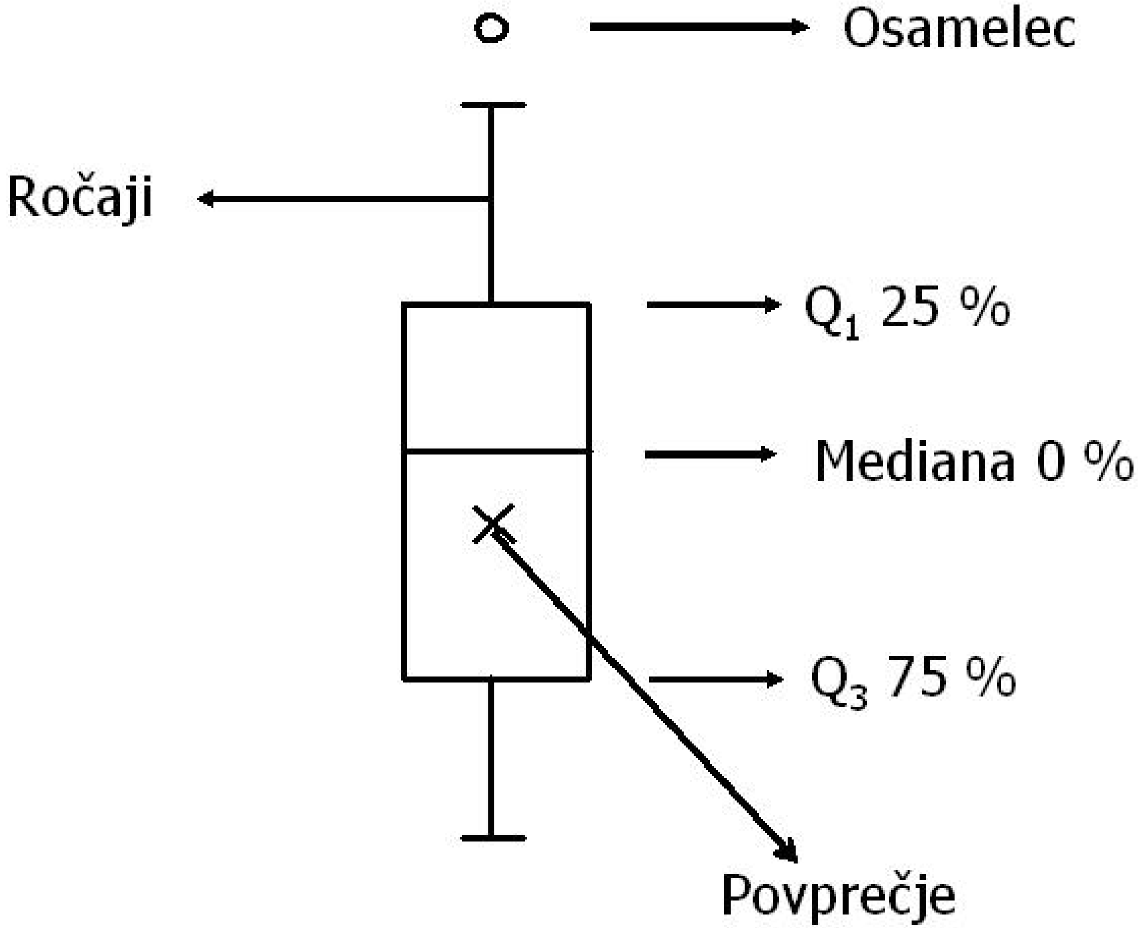
Koeficient variabilnosti

- mera za variabilnost
- primerjamo standardni odklon s povprečno vrednostjo

$$KV = \frac{\sigma}{\bar{x}} * 100$$

- vrednosti navajamo v odstotkih
- v starejši literaturi je pogosto uporabljena ...
- če $\overline{X} \to 0 \Rightarrow KV \to \infty$

Kvantili



Mere podobnosti

1. Kovarianca

2. Korelacija

3. Regresija

Kovarianca

= povprečni produkt odklonov obeh spremenljivk.

$$\sigma_{xy} = cov\,(x, y) = \frac{\sum(x_i - \overline{X})(y_i - \overline{Y})}{n-1}$$

- vrednosti so od $-\infty$ do ∞
- kovarianca $= 0 \Rightarrow$ spremenljivki sta neodvisni, **nista podobni!**
- kovarianca $> 0 \Rightarrow$ spremenljivki sta povezani: **obe večji ali obe manjši**
- kovarianca $< 0 \Rightarrow$ spremenljivki sta povezani: **ena večja druga manjša**

Preuredimo enačbo!

$$\sigma_{xy} = cov(x, y) = \frac{\sum\left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{n-1} =$$

pomnožimo

$$\frac{\sum\left(x_i y_i - \overline{X} y_i - x_i \overline{Y} + \overline{XY}\right)}{n-1} =$$

in razstavimo

$$= \frac{\sum(x_i y_i) - \sum\left(\overline{X} y_i\right) - \sum\left(x_i \overline{Y}\right) + \sum\left(\overline{XY}\right)}{n-1} =$$

povprečja lahko izpostavimo pred vsoto

$$\frac{\sum(x_i y_i) - \overline{X}\sum(y_i) - \overline{Y}\sum(x_i) + \overline{XY}\sum(1)}{n-1} =$$

Preuredimo enačbo! (nadalj.)

$$= \frac{\sum(x_i y_i) - \overline{X}\sum(y_i) - \overline{Y}\sum(x_i) + \overline{X}\,\overline{Y}\sum(1)}{n-1} =$$

ker velja $\sum(y_i) = n\overline{Y}$

in $\sum(1) = n$,

lahko zapišemo

$$= \frac{\sum(x_i y_i) - n\overline{X}\,\overline{Y} - n\overline{X}\,\overline{Y} + n\overline{X}\,\overline{Y}}{n-1} =$$

Preuredimo enačbo! (nadalj.)

$$= \frac{\sum(x_i y_i) - n\overline{XY} - n\overline{XY} + n\overline{XY}}{n-1} =$$

zadnje tri člene lahko sedaj seštejemo

$$= \frac{\sum(x_i y_i) - n\overline{XY}}{n-1} = \sigma_{xy}$$

preuredimo zadnji člen

$$= \frac{\sum(x_i y_i) - \frac{1}{n}\sum(x_i)\sum(y_i)}{n-1} = \sigma_{xy}$$

Primer: masa rogovja in starost losov

V enem revirju so tehtali rogovje losov. Našli so 5 rogovij in na osnovi le-teh določili starost živali. Podatki so navedeni spodaj.

Jelen	Masa (kg)	Starost (let)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	13.0	3					
2	8.5	2					
3	12.0	3					
4	16.5	4					
5	10.0	3					
Σ							

Izračunajte kovarianco med maso rogovja in starostjo!

Primer: masa rogovja in starost losov

V enem revirju so tehtali rogovje losov. Našli so 5 rogovij in na osnovi le-teh določili starost živali. Podatki so navedeni spodaj.

Jelen	Masa (kg)	Starost (let)	$x_i y_i$
1	13.0	3	
2	8.5	2	
3	12.0	3	
4	16.5	4	
5	10.0	3	
$\sum$			

Izračunajte kovarianco med maso rogovja in starostjo!

Primer: masa rogovja in starost losov

V enem revirju so tehtali rogovje losov. Našli so 5 rogovij in na osnovi le-teh določili starost živali. Podatki so navedeni spodaj.

Jelen	Masa (kg)	Starost (let)	x_i^2	y_i^2	$x_i y_i$
1	13.0	3	169.00	9	39.0
2	8.5	2	72.25	4	17.0
3	12.0	3	144.00	9	36.0
4	16.5	4	272.25	16	66.0
5	10.0	3	100.00	9	30.0
$\sum$	60.0	15	757.50	47	188.0

$$var(x) = 9.4\,kg^2 \quad var(y) = 0.5 let^2 \quad cov(x,y) = 2 kglet$$

Korelacija

- razmerje med kovarianco in geometrijskim povprečjem varianc

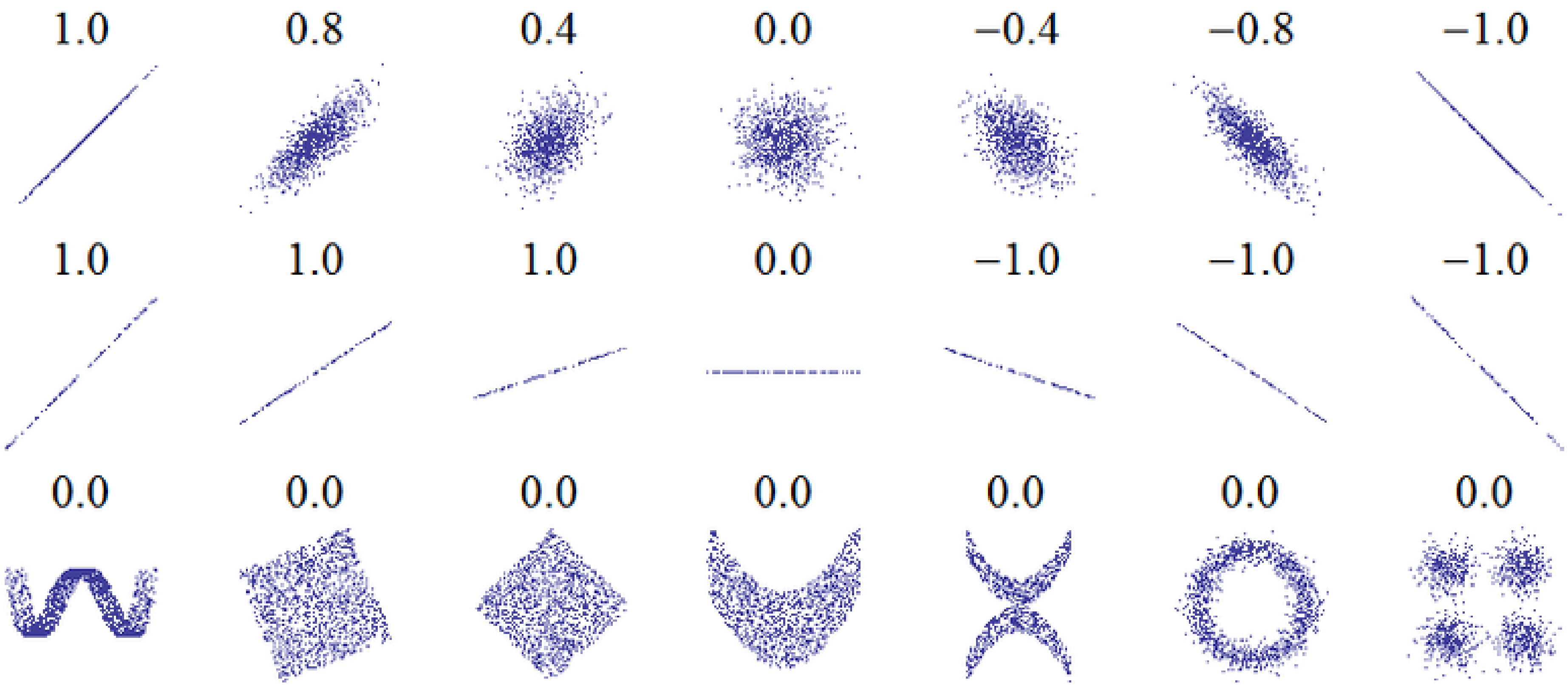
- standardizirana kovarianca (torej brez enot!)

- zaloga vrednosti: od -1 do 1

$$r_{xy} = corr(x, y) = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}$$

$$= \frac{\sum(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum(x_i - \overline{X})^2 \sum(y_i - \overline{Y})^2}}$$

Korelacija



Regresija (linearna)

- je smerni koeficient premice

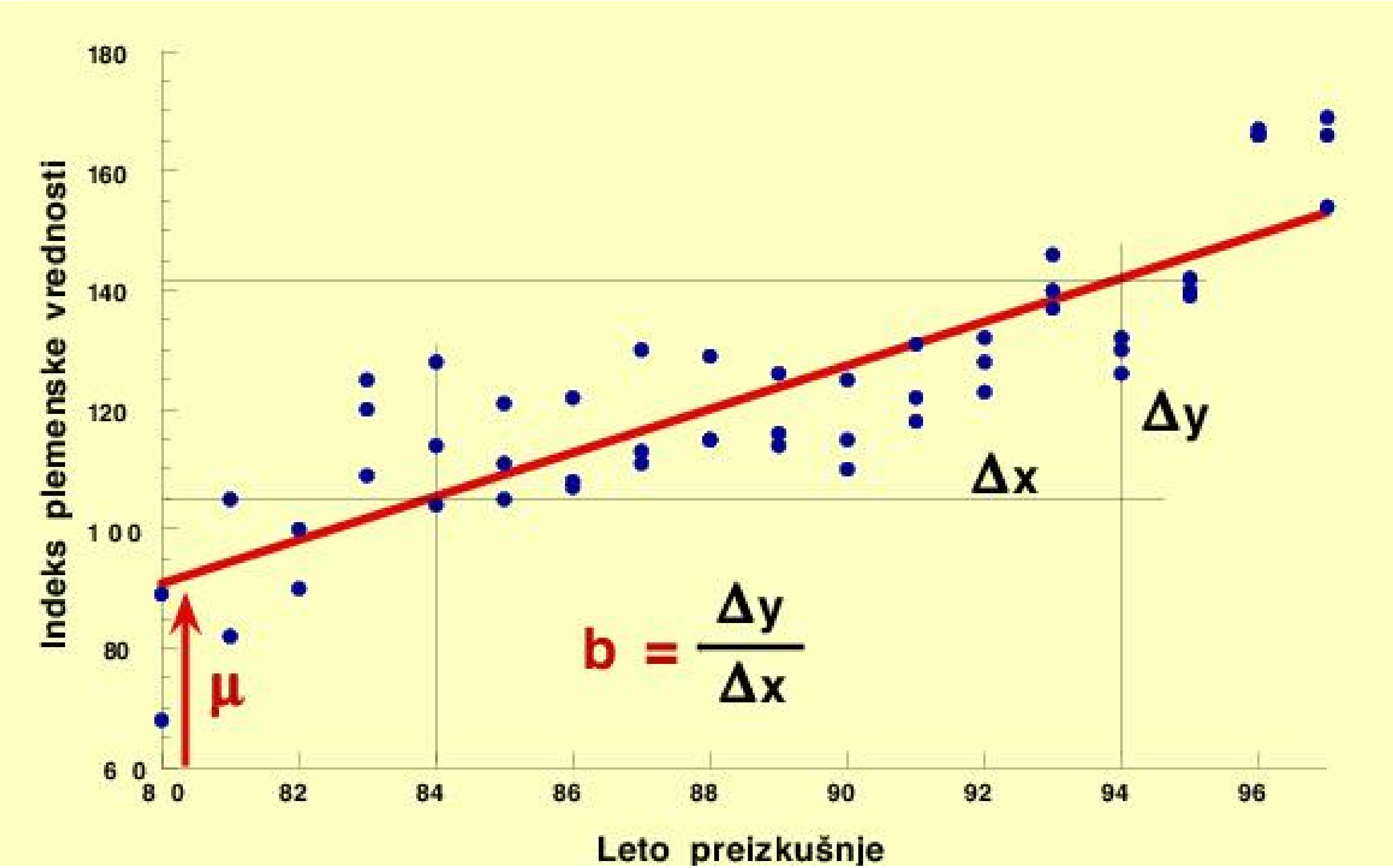
- v statistiki - regresijski koeficient

$$b_{xy} = \frac{cov(x,y)}{var(x)}$$
$$= \frac{\sum\left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\sum\left(x_i - \overline{X}\right)^2} =$$
$$= r_{xy}\sigma_y$$

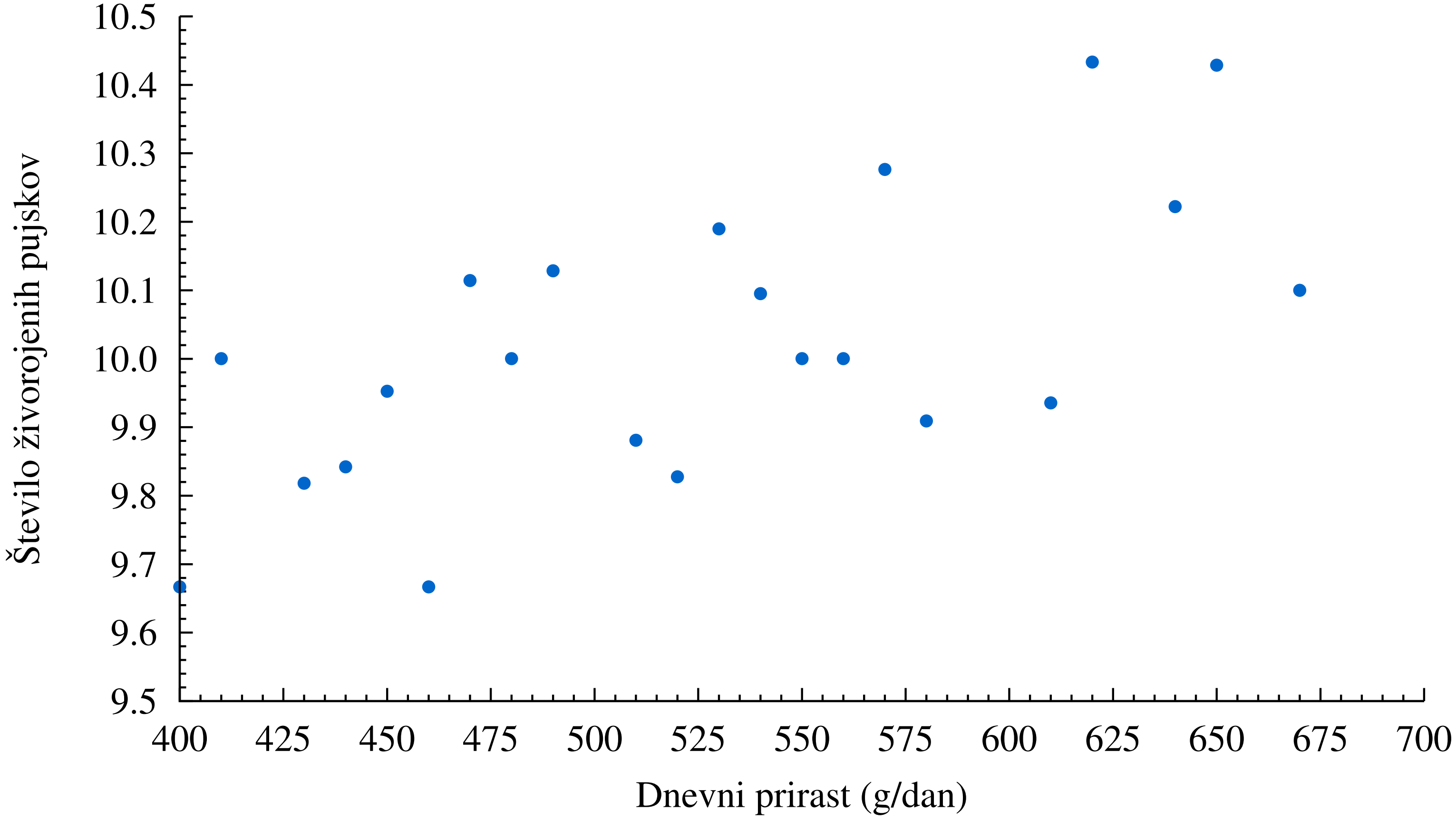
- zaloga vrednosti: od $-\infty$ do ∞

- enote: enota za y/enota za x

Primer: smerni koeficient premice

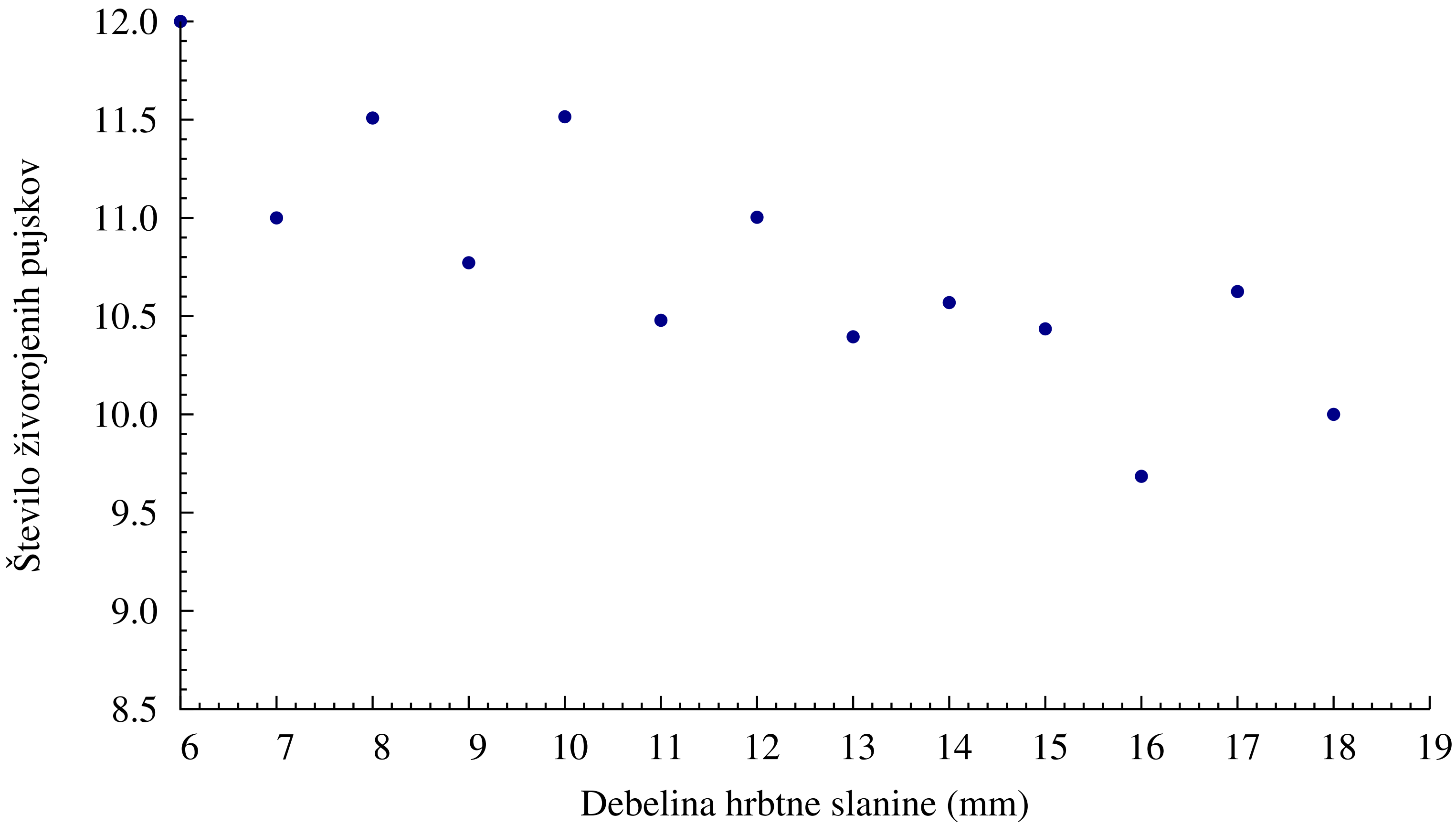


Primer: smerni koeficient premice



Vrišite premico, ki se prilega podatkom, in ocenite smerni koeficient!

Primer: smerni koeficient premice



Vrišite premico, ki se prilega podatkom, in ocenite smerni koeficient!