

KORELACIJA

Lenon

"Život je ono što ti se dešava dok ti praviš planove za život."

KORELACIJA

- ◆ Povezanost pojava ili svojstava koje se proučavaju
- ◆ Procijenjuje stupanj variranja varijabli
- ◆ Odnosi se na **intenzitet i smjer** povezanosti varijabli u tom odnosu

Pozitivna povezanost – porastom vrijednosti jednoga svojstva rastu i pripadajuće vrijednosti drugog svojstva

- ◆ visina i težina čovjeka
- ◆ brzina automobila i vrijeme potrebno za zaustavljanje
- ◆ broj cvjetova i broj zametnutih plodova
- ◆ duljina klipa i broj zrna
- ◆ broj etaža i broj mahuna
- ◆ gnojidba i prinos
- ◆ efikasnost pesticida i prinos
- ◆ razmak sjetve(sadnje) i visina biljke

Negativna povezanost – porastom vrijednosti jednoga svojstva opadaju pripadajuće vrijednosti drugog svojstva

- ◆ Starost osobe i broj zdravih zubi
- ◆ Visina prinosa i kvaliteta
- ◆ Postotak šećera i sadržaj kiselina u soku
- ◆ Postotak bjelančevina i postotak škroba
- ◆ Brzina sijačice i jednoličnost polaganja sjemena u tlo
- ◆ Promjer cijevi za odvodnju i preostala količina vode u tlu
- ◆ Koncentracija preparata i stupanj zaraženosti

- ◆ Karl Pearson (1857. – 1936.)
- ◆ *sir* Francis Galton(1822. – 1911.)

$$r = \frac{SP_{xy}}{\sqrt{SS_x \times SS_y}}$$

$$SS_x = \sum (x^2) - \frac{(\sum x)^2}{n}$$

$$SS_y = \sum (y^2) - \frac{(\sum y)^2}{n}$$

$$SP_{xy} = \sum (x \cdot y) - \frac{\sum x \cdot \sum y}{n}$$



sir Francis Galton



Karl Pearson



GRAFIČKI PRIKAZ

dijagram rasipanja - *scatter diagram*

Apscisa – nezavisna varijabla (samovoljno je mijenjamo ili bираmo)

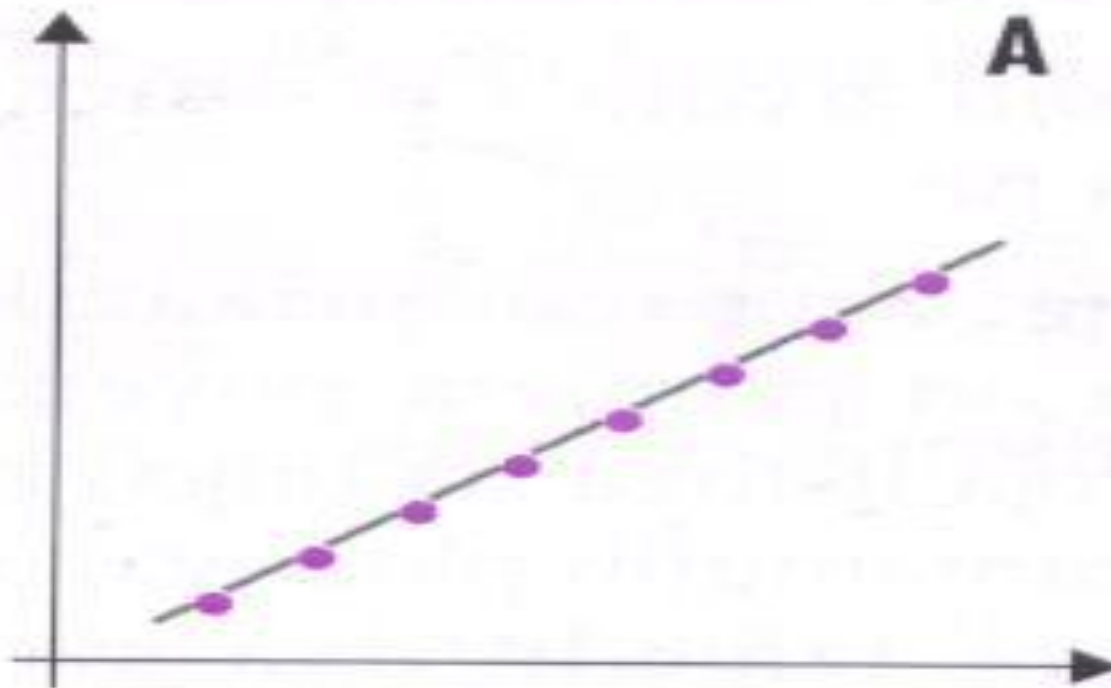
Ordinata – zavisna varijabla (istraživanjem je želimo ustanoviti)

Točkice predstavljaju parove vrijednosti x i y varijable

Praktičan savjet – grafički prikaz, a zatim analiza!

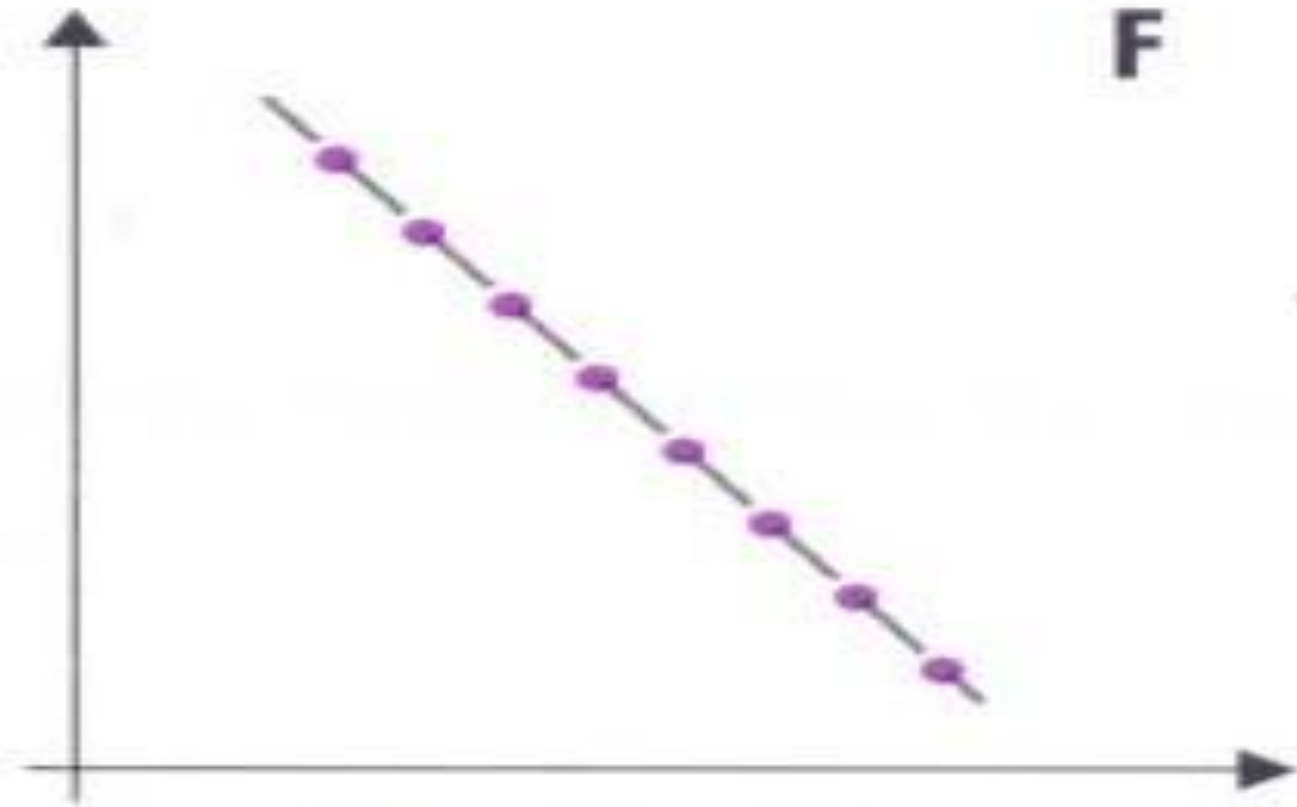


Korelacije različitih smjerova i jačina



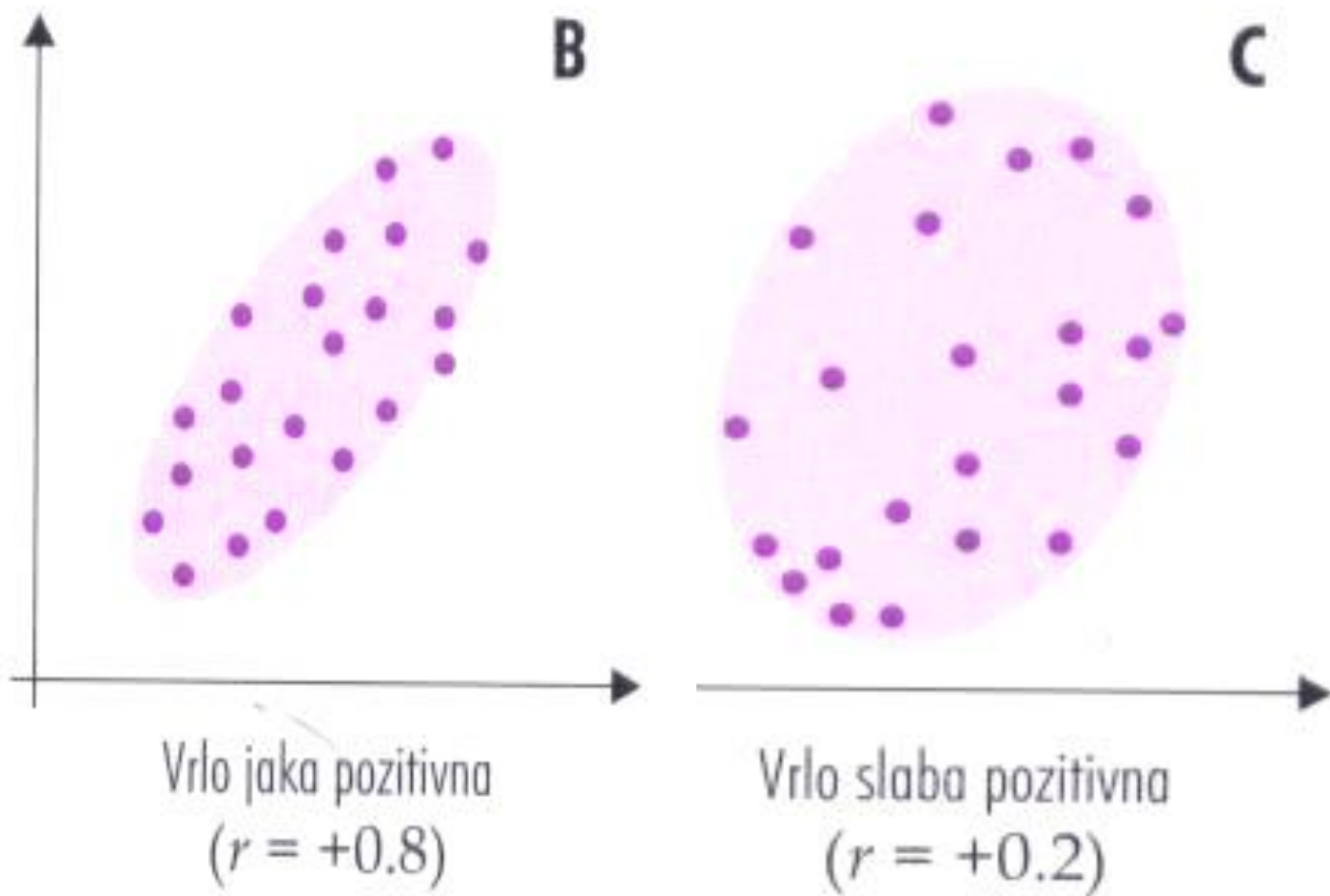
Potpuna pozitivna
($r = +1.0$)

npr. odnos između polumjera i opsega kruga

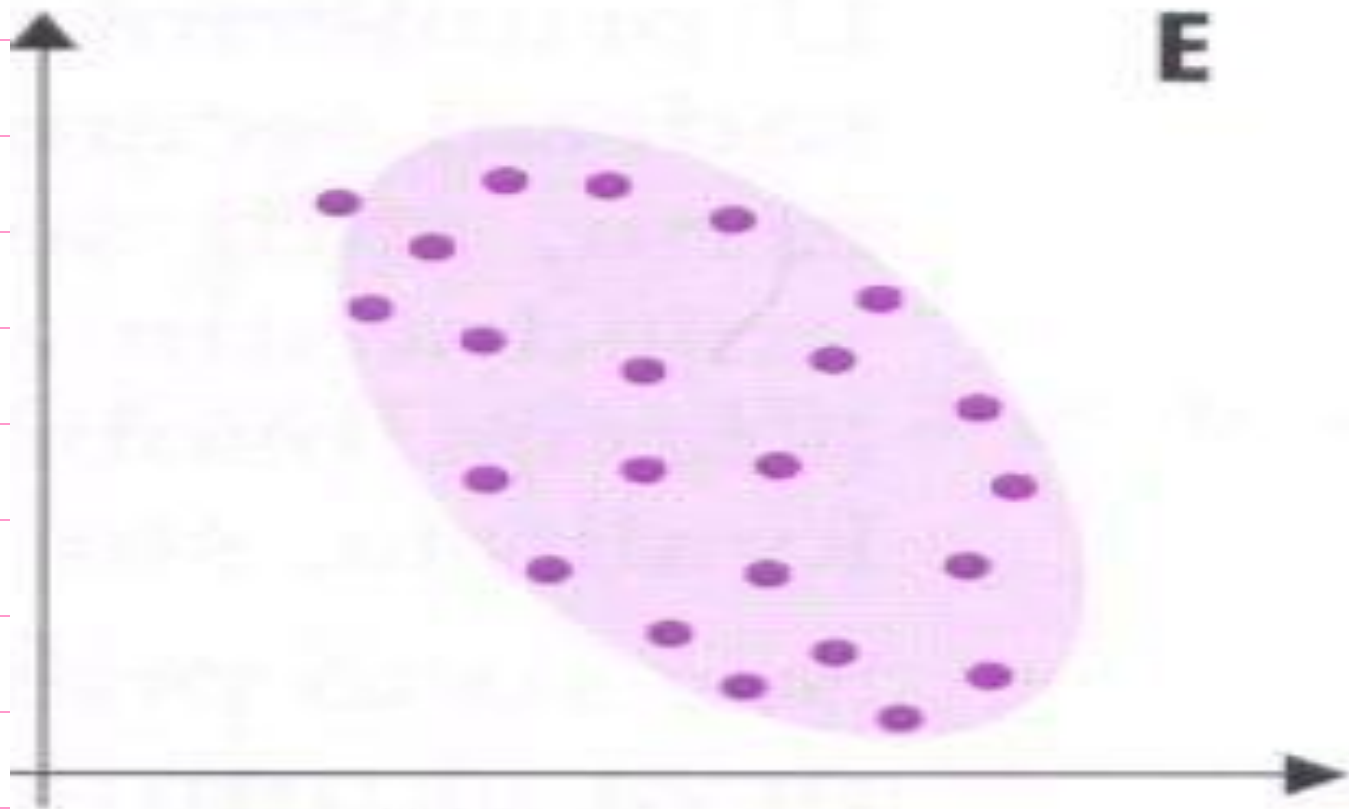


Potpuna negativna
($r = -1.0$)

npr. odnos između vremena proteklog od ispaljivanja metka uvis i brzine tog metka



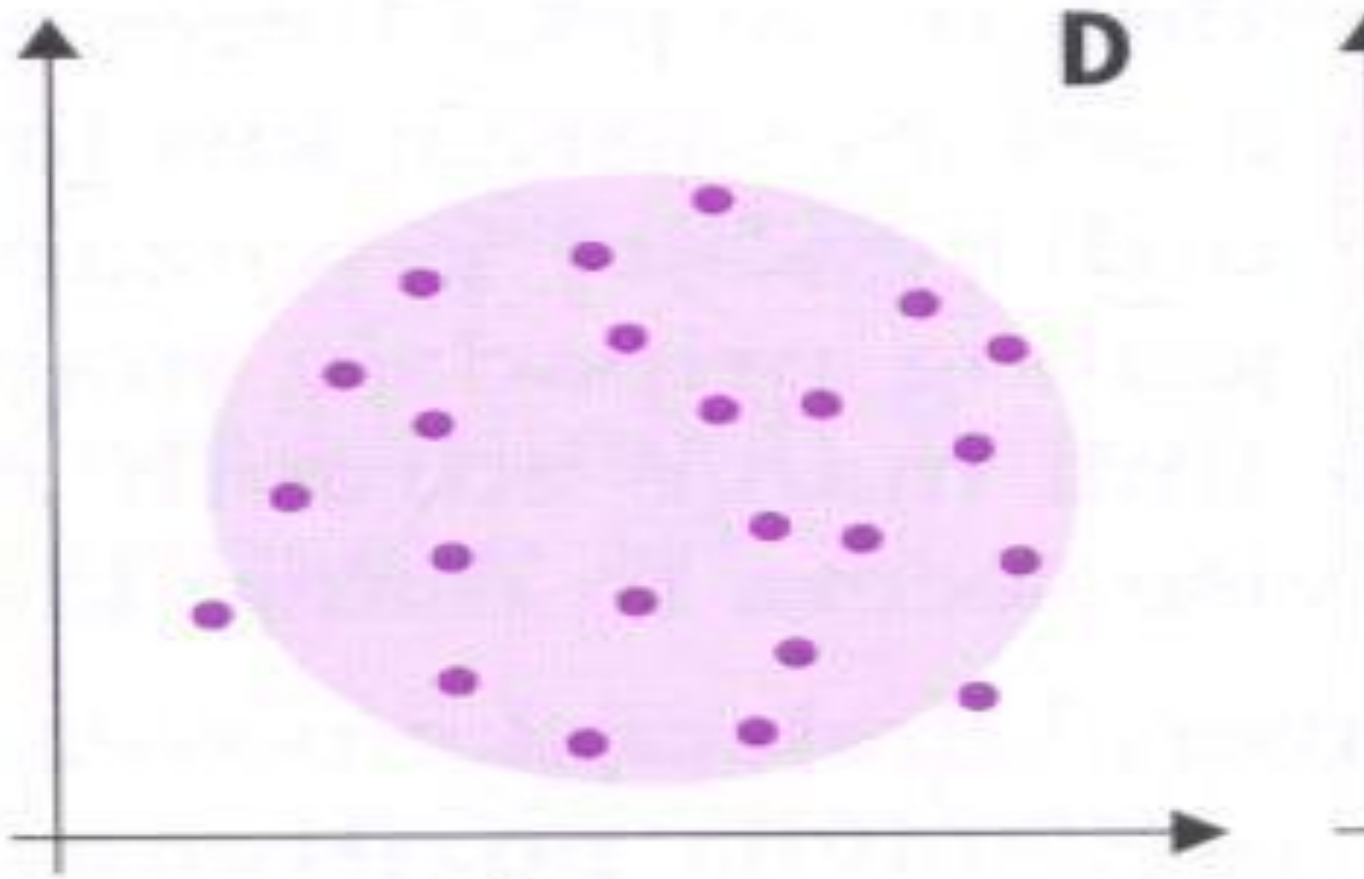
npr. odnos između visine i težine ljudi



Jaka negativna
($r = +0.5$)

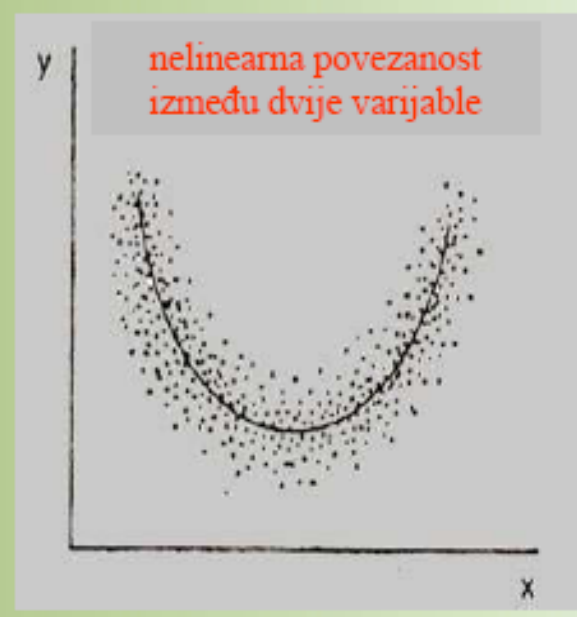
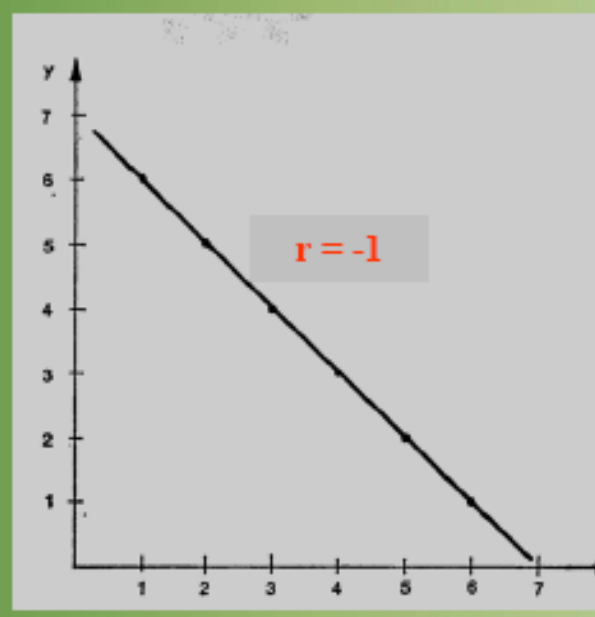
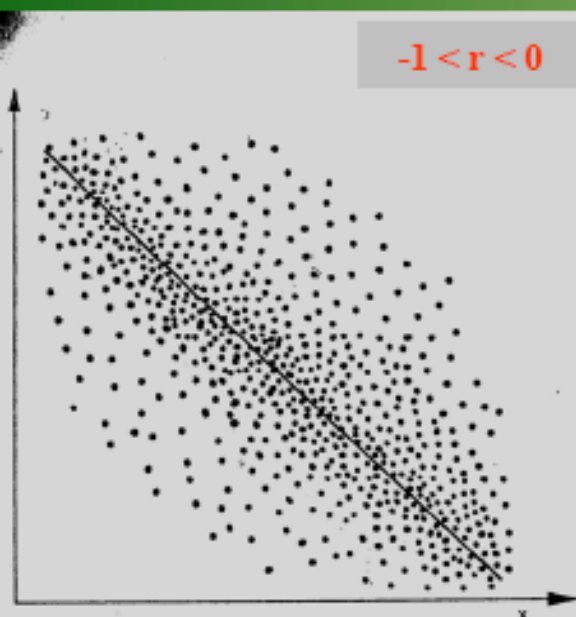
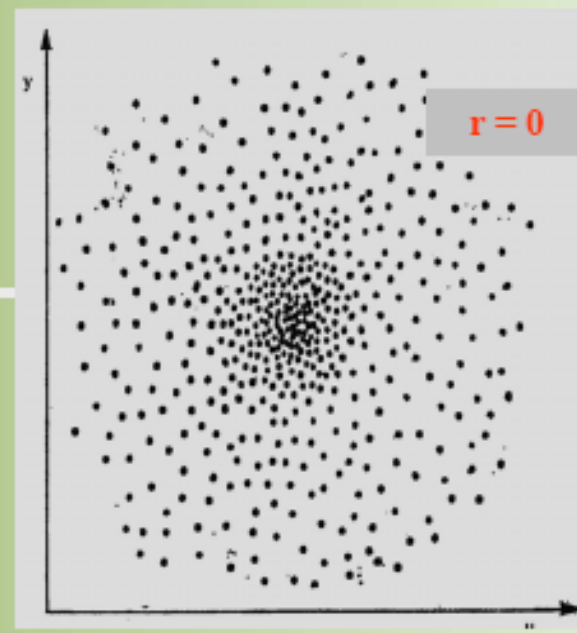
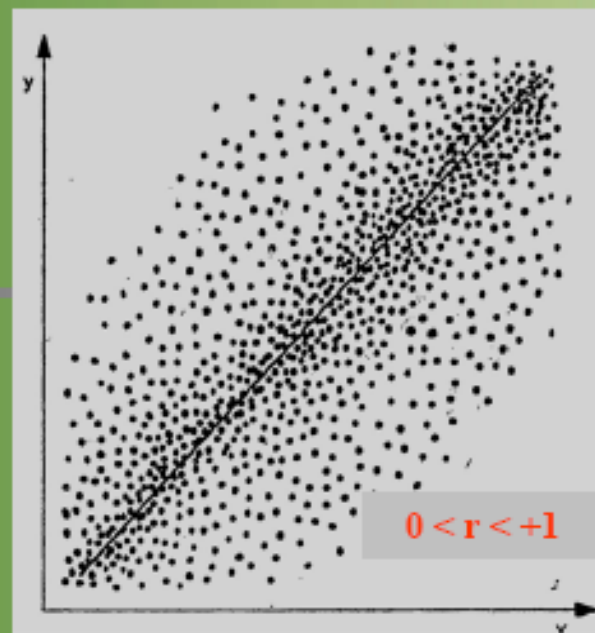
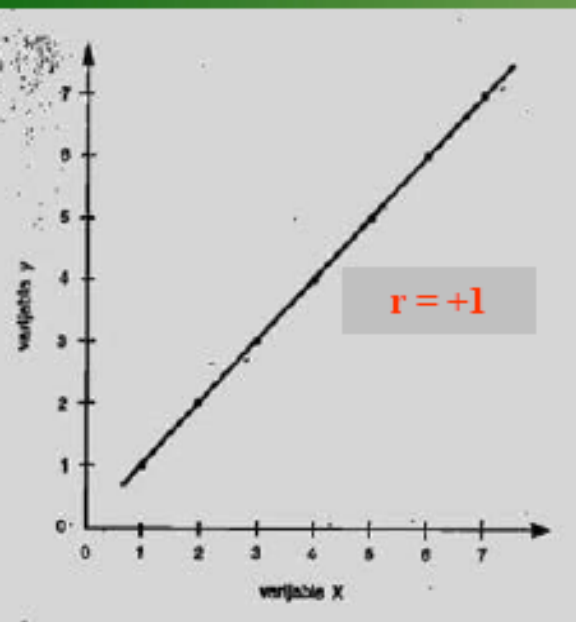
E

npr. odnos između stupnja treniranosti i frekvencije pulsa u prvoj minuti oporavka



Korelacije nema
($r = 0.0$)

npr. odnos između duljine kose i krvnog tlaka




Korelacija može biti:

- po smjeru: **pozitivna ili negativna**
- po jačini: **od vrlo slabe ili nikakve do potpune**

$$-1,00 \leq r \leq +1,00$$



Roemer-Orphal-ova tablica



korelacijski koeficijent (r)	jačina korelacije
0.00 – 0.10	nema
0.10 – 0.25	vrlo slaba
0.25 – 0.40	slaba
0.40 – 0.50	srednja
0.50 – 0.75	jaka
0.75 – 0.90	vrlo jaka
0.90 – 1.00	potpuna

KOEFICIJENT KORELACIJE “ r “

- pokazuje stupanj (i smjer) korelacije



- zaključke o vrijednosti koeficijenta korelacije donosimo uz određenu razinu vjerojatnosti ($p < 0,05$; $p < 0,01$)

Testiranje opravdanosti korelacijskog koeficijenta

$$H_0 : r = 0$$

$$t_{\text{exp}} = \frac{r}{s_r}$$

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

t_{tab} - očitati na temelju $n-2$ slobodnih varijanata
za $p = 5\%$ i za $p = 1\%$

Usporedba t_{exp} i t_{tab}

$$t_{\text{exp}} < t_{\text{tab}}$$

H_0 se prihvaća

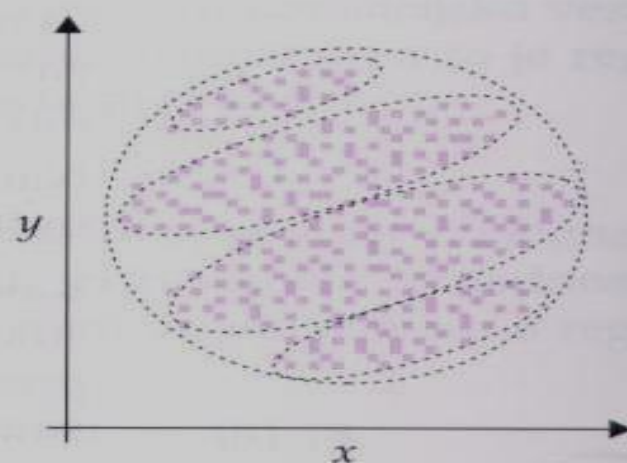
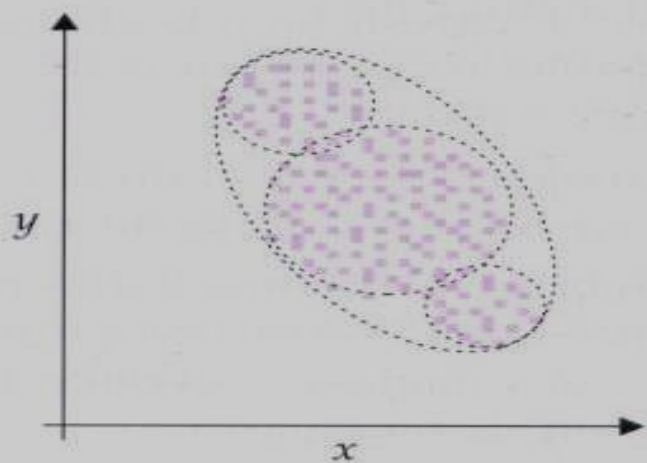
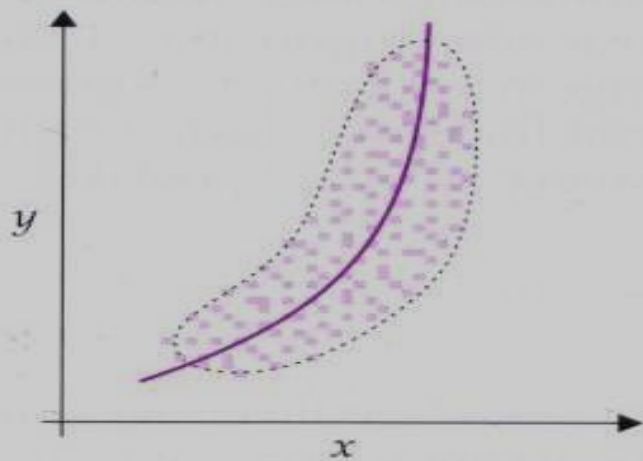
$$t_{\text{exp}} > t_{\text{tab}}$$

H_0 se odbacuje

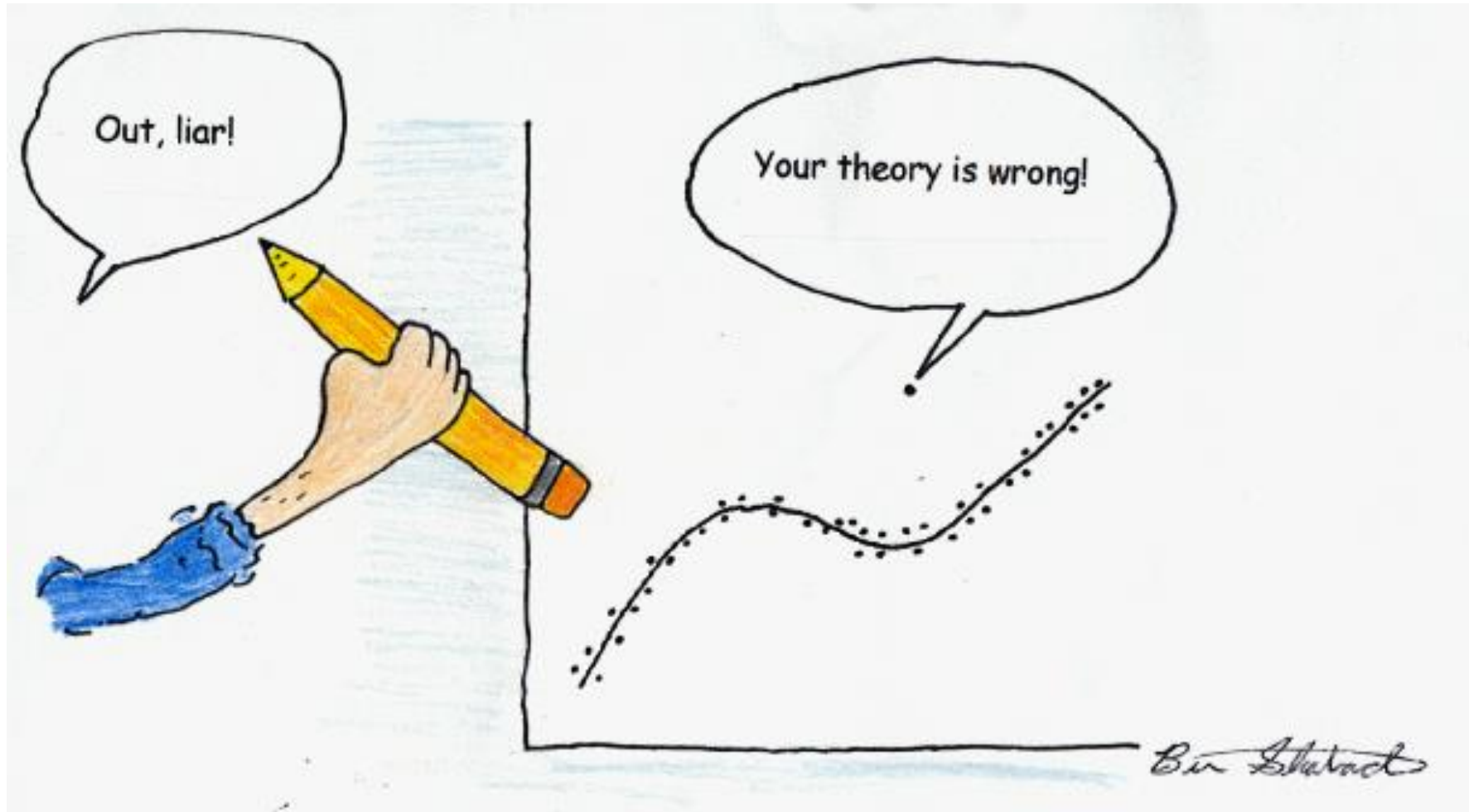
uz $p = 5\%$ ili $p = 1\%$



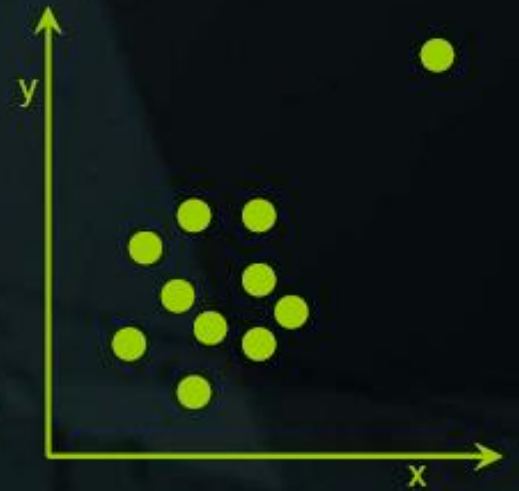
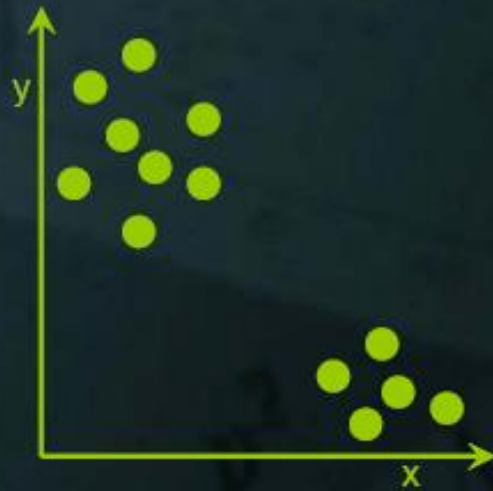
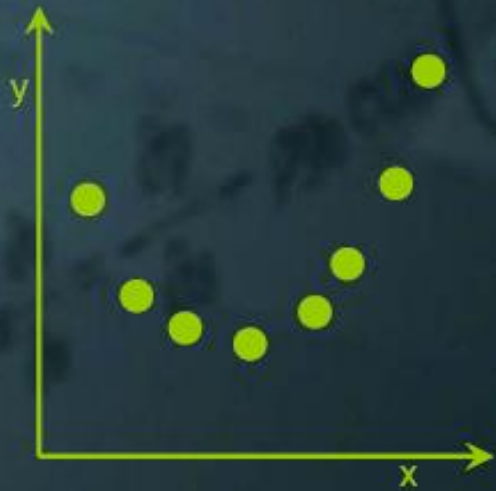
Mogući razlozi pogrešne analize i interpretacije korelacije



Problem ekstremnih vrijednosti (*outliers*)



Kada NE računati r



Važne napomene

- ◆ Uz korelacijski koeficijent navesti granice x i y varijable
- ◆ Korelacijski koeficijent na temelju samo dva para podataka bit će uvijek $+1$ ili -1
- ◆ Za nelinearne veze transformirati podatke
- ◆ Odnos između dvije varijable je jednostavna ili ukupna korelacija



Istraživanje ovisnosti između dvije ili više varijabli ne podrazumjeva automatski postojanje uzročno posljedične veze između njih. Kauzalnost nije statistički, već eksperimentalni i logički problem.

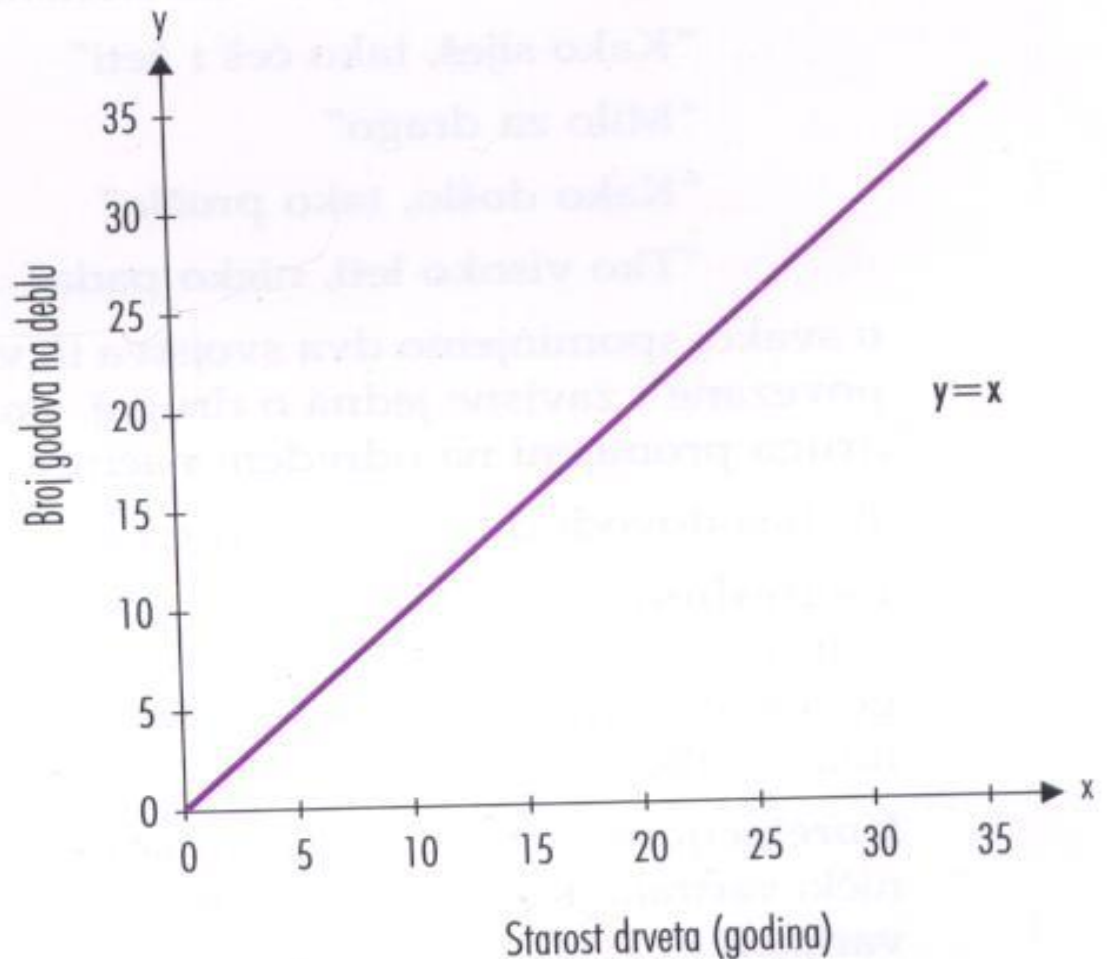
REGRESIJA



REGRESIJA

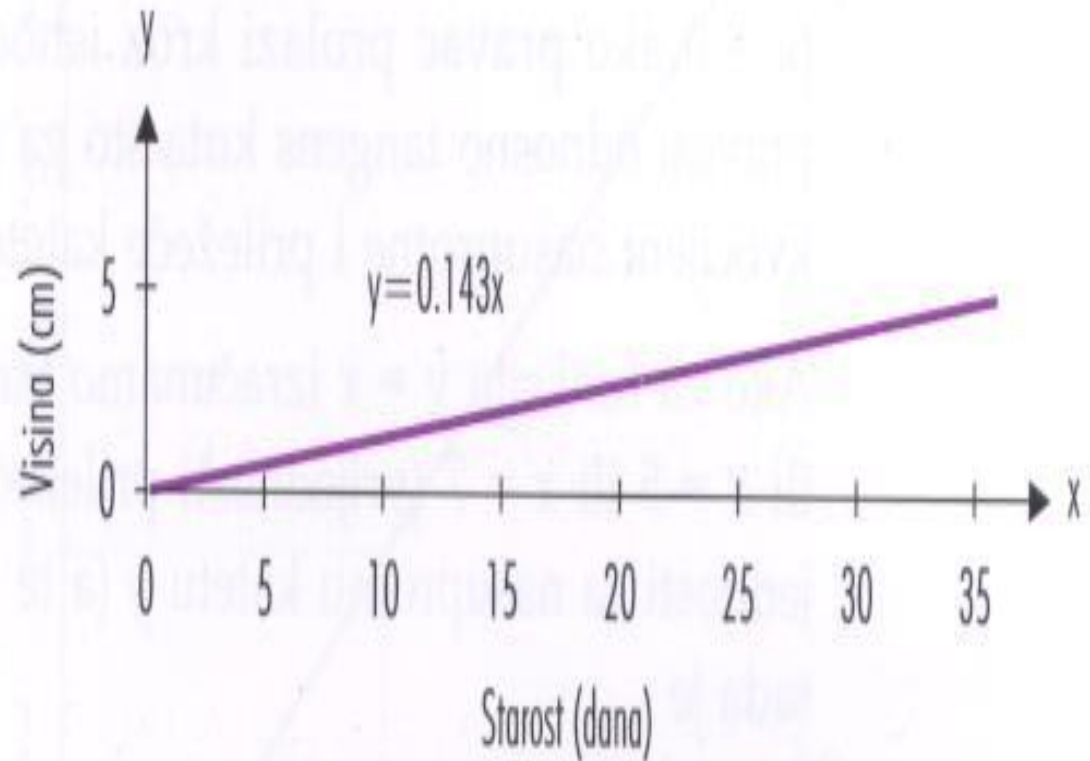
- ◆ Funkcionalni odnos dva svojstva
- ◆ $Y = f(x)$
- ◆ Zavisna varijabla – promjena zavisi o drugoj varijabli
- ◆ Nezavisna varijabla – uzrok promjeni prve varijable

REGRESIJA: $y=x$



SLIKA 23. Odnos starosti drveta (x) i broja godova na deblu (y)

REGRESIJA: $y=1/7 x$ ili $y = 0.143 x$



SLIKA 24.

Odnos starosti biljke (x) i visine biljke (y)

REGRESIJA: $y=1/7 x$ ili $y=0.143 x$

◆ $Y=1/7^* x$ $x = 5$ dana

◆ $Y=1/7 * 5 = 0.71$ cm

◆ $Y=1/7^* x$ $x = 10$ dana

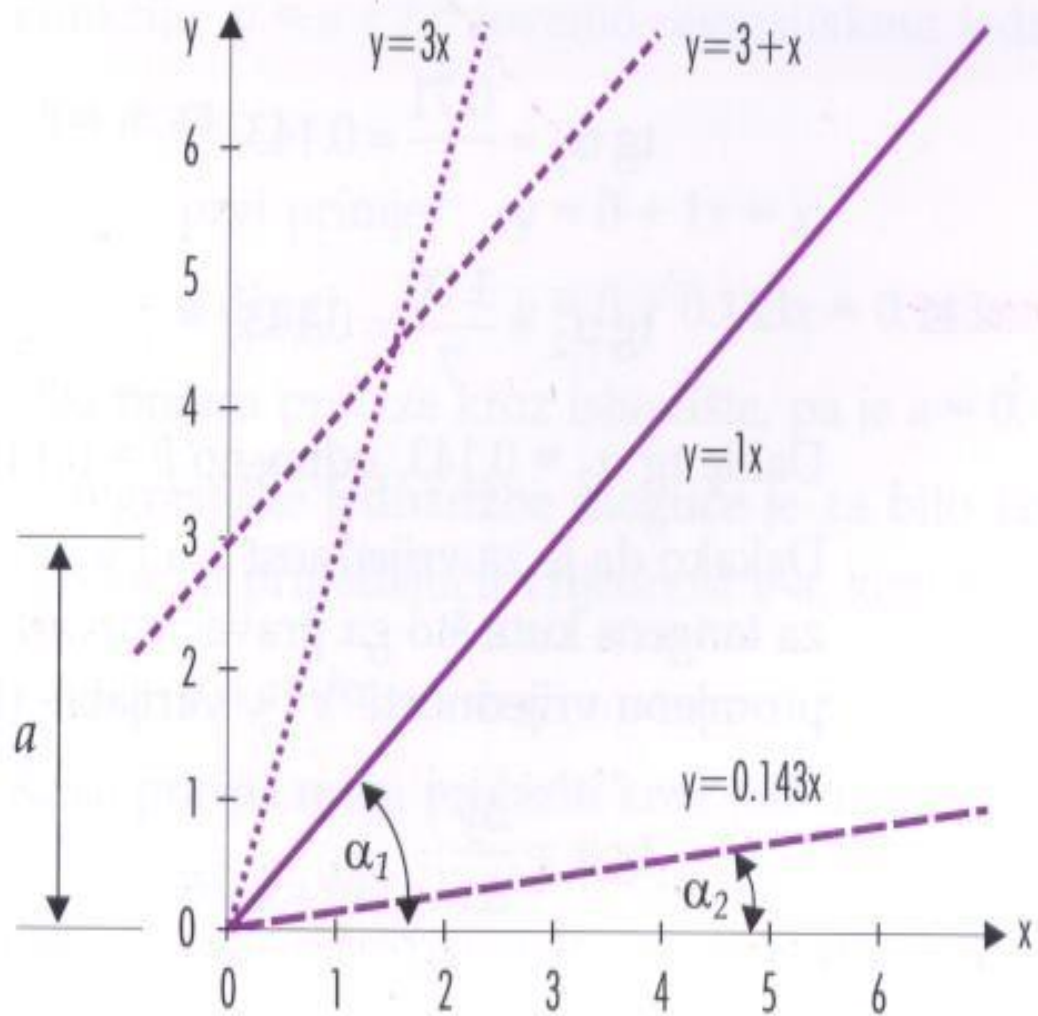
◆ $Y=1/7 * 10 = 1.43$ cm



$$Y = a + bx$$

- ◆ Pravec se može izraziti funkcijom
- ◆ a - odsječak na ordinati
- ◆ b - koeficijent smjera pravca ili tangens kuta što ga pravac zatvara sa osi x
(kvocijent nasuprotne i priležeće katete)

Koeficijent b određuje nagib pravca



SLIKA 25. Pravci različitih nagiba ovisno o koeficijentu smjera (b)

Tanges kuta α

Za funkciju $y = x$

za: $x = 2$ ili $x=5$ ili $x =7$

$$\text{Tg } \alpha = 2/2 = 1$$

$$\text{Tg } \alpha = 5/5 = 1$$

$$\text{Tg } \alpha = 7/7 = 1$$

$$\text{Tg } \alpha = 1 \text{ ili } b = 1$$

Tanges kuta alfa

Za funkciju $y = 1/7 x$

za: $x = 2$ ili $x=5$ ili $x =7$

$$y = 2/7 = 0.28$$

$$y = 5/7 = 0.71$$

$$y = 7/7 = 1.0$$

$$\text{tg } \alpha_2 = 0.28/2 = 0.143$$

$$\text{tg } \alpha_2 = 0.71/5 = 0.143$$

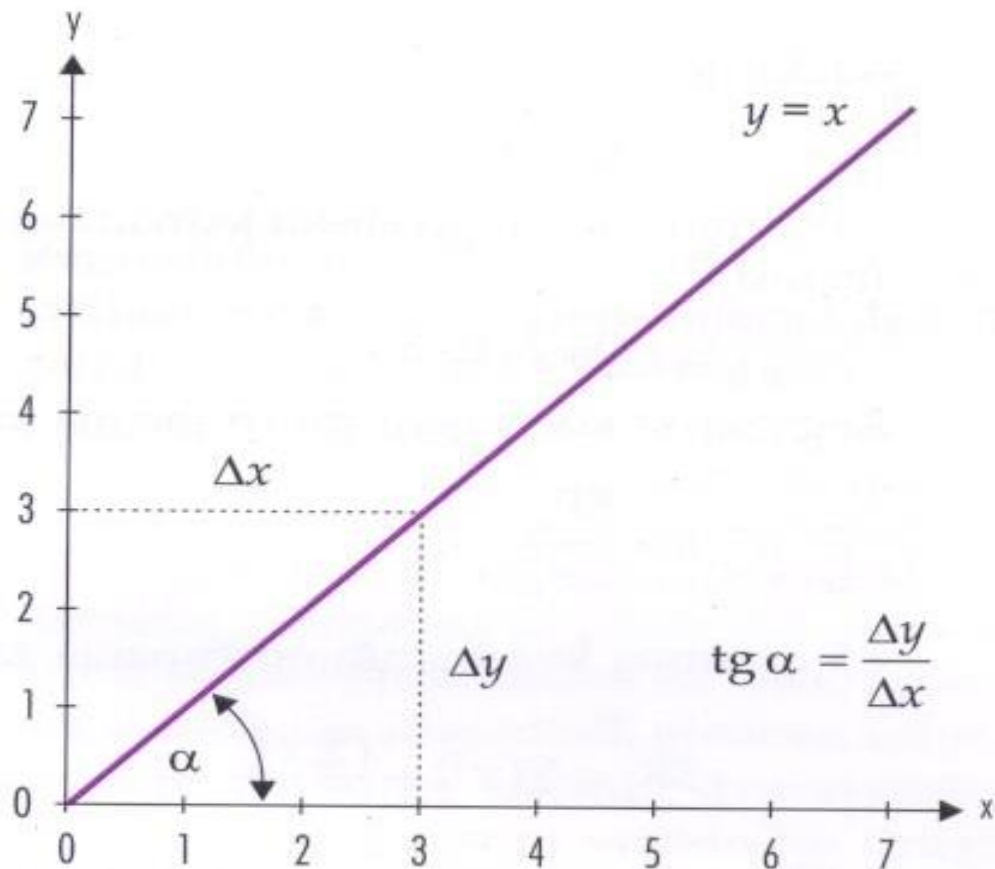
$$\text{tg } \alpha_2 = 1.00/7 = 0.143$$

$$\text{tg } \alpha_2 = 0.143 \text{ ili } b = 0.143$$

Tanges kuta alfa

- Za vrijednost x -a i y -a u bilo kojoj točki je ista vrijednost za tanges kuta što ga pravac zatvara sa x osi
- Za bilo koju promjenu vrijednosti x i y varijable tanges kuta je:
$$\text{tg } \alpha = \Delta y / \Delta x$$

tg α



SLIKA 26.

Koeficijent smjera pravca je tangens kuta α .

$$\text{za } \Delta x = 1 : b = \text{tg } \alpha = \Delta y$$

b- regresijski koeficijent

$Y = a + bx$ –regresijska

jednadžba

$$Y = 0 + 1x = x$$

$$Y = 0 + 0.143 x = 0.143 x$$

$$a = 0$$

Jednadžba linearne regresije

Za bilo koju vrijednost x -a moguće je izračunati pripadajuću vrijednost y -a

$$\hat{Y} = a + bx \quad T(\bar{x}, \bar{y})$$

$$\bar{y} = a + b\bar{x} \quad a = \bar{y} - b\bar{x}$$

$$\hat{y} = \bar{y} + b(x - \bar{x})$$

Regresijski koeficijent

$$b = \frac{SP_{xy}}{SS_x}$$

$$SP = \Sigma(xy) - \frac{\Sigma x \Sigma y}{n}$$



Kovarijanca-Cov

$$\mathbf{Cov} = \frac{SP_{xy}}{n - 1}$$

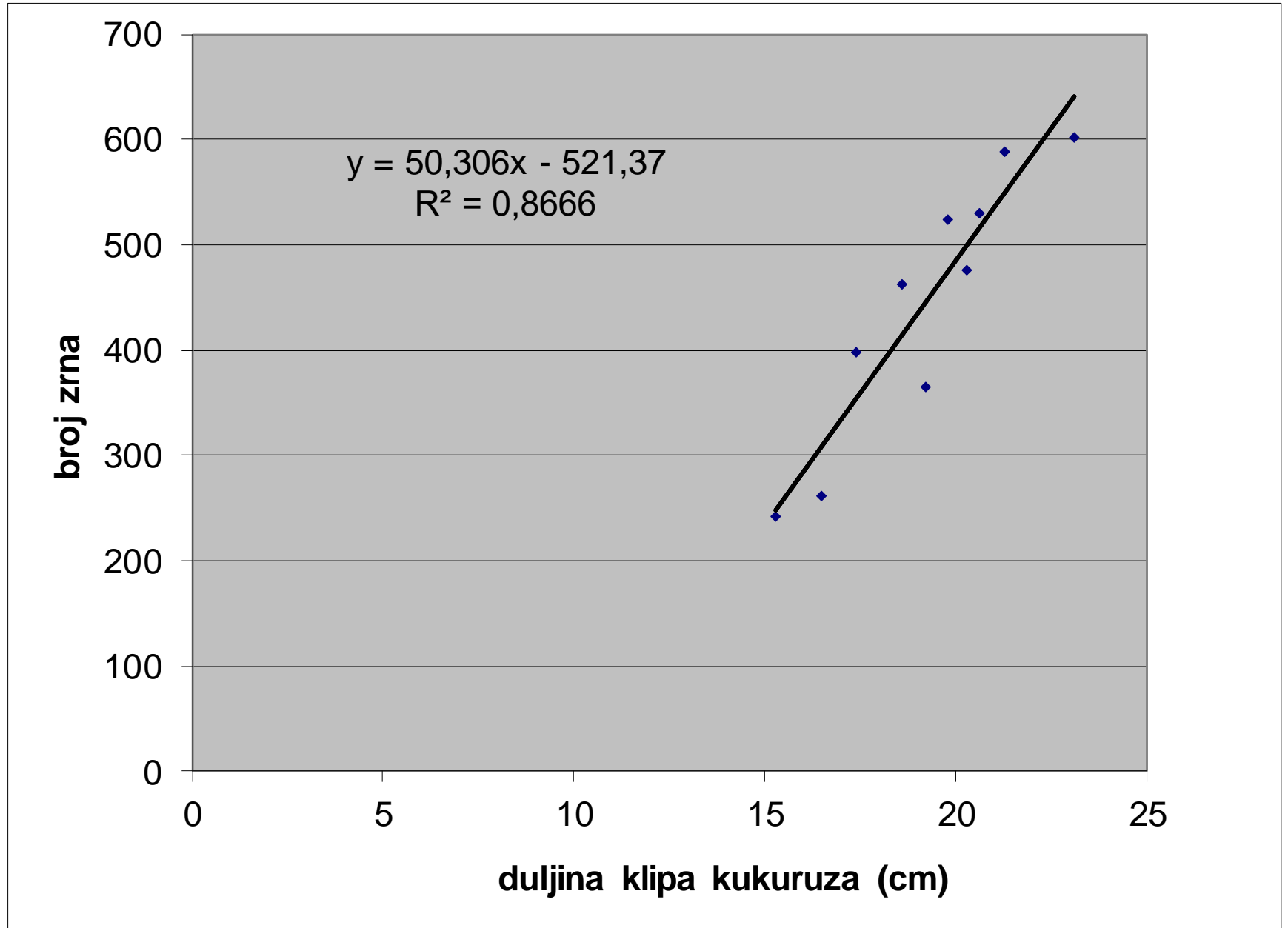
Primjer 1.

15,3	242
19,2	365
16,5	261
19,8	524
20,3	476
21,3	589
17,4	398
18,6	463
20,6	530
23,1	602

Primjer 2.

15,7	8,5
15,4	8,45
13,2	9,2
16,3	8,25
15,0	9,05
16,4	8,9
13,8	10,3
17,4	7,35
15,6	8,95
14,9	9,35
16,2	9,8
15,5	9,95

Duljina klipa kukuruza i broj zrna



% šećera i sadržaj kiselina(g/l)

