



OSNOVNI KONCEPT ZA ANALIZU KORELACIJE I REGRESIJE

Prof. dr. sc. Velimir Sušić
Prof.dr.sc. Ivo Karadjole



CILJEVI

- na koja pitanja (probleme) odgovore možemo dobiti analizom korelacije i regresije
- definicije korelacije i regresije i statistička osnova za njihovu analizu
- grafički prikaz i izmjeritelji korelacije odnosno regresije i njihova interpretacija



PITANJA (PROBLEMI)

- Dva različita testa (test A i test B) trebaju pokazati razumijevanje i usvojeno znanje studenata iz predmeta Osnove statistike u veterinarskoj medicini.
 - Postoji li povezanost rezultata koje studenti ostvare na testu A s rezultatima na testu B?
 - Može li se predvidjeti mogući rezultat testa B na temelju rezultata testa A?
- Pri selekciji krava na obilježje mliječnosti u obzir se najčešće uzimaju svojstva količine pomuzenog mlijeka i udjela (%) mliječne masti.
 - Postoji li povezanost između količine pomuzenog mlijeka i udjela mliječne masti?
 - Može li se na temelju podatka o količini pomuzenog mlijeka predvidjeti mogući podatak za udio mliječne masti?



PRIKUPLJANJE PODATAKA

BROJ BODOVA ZA POJEDINOG STUDENTA

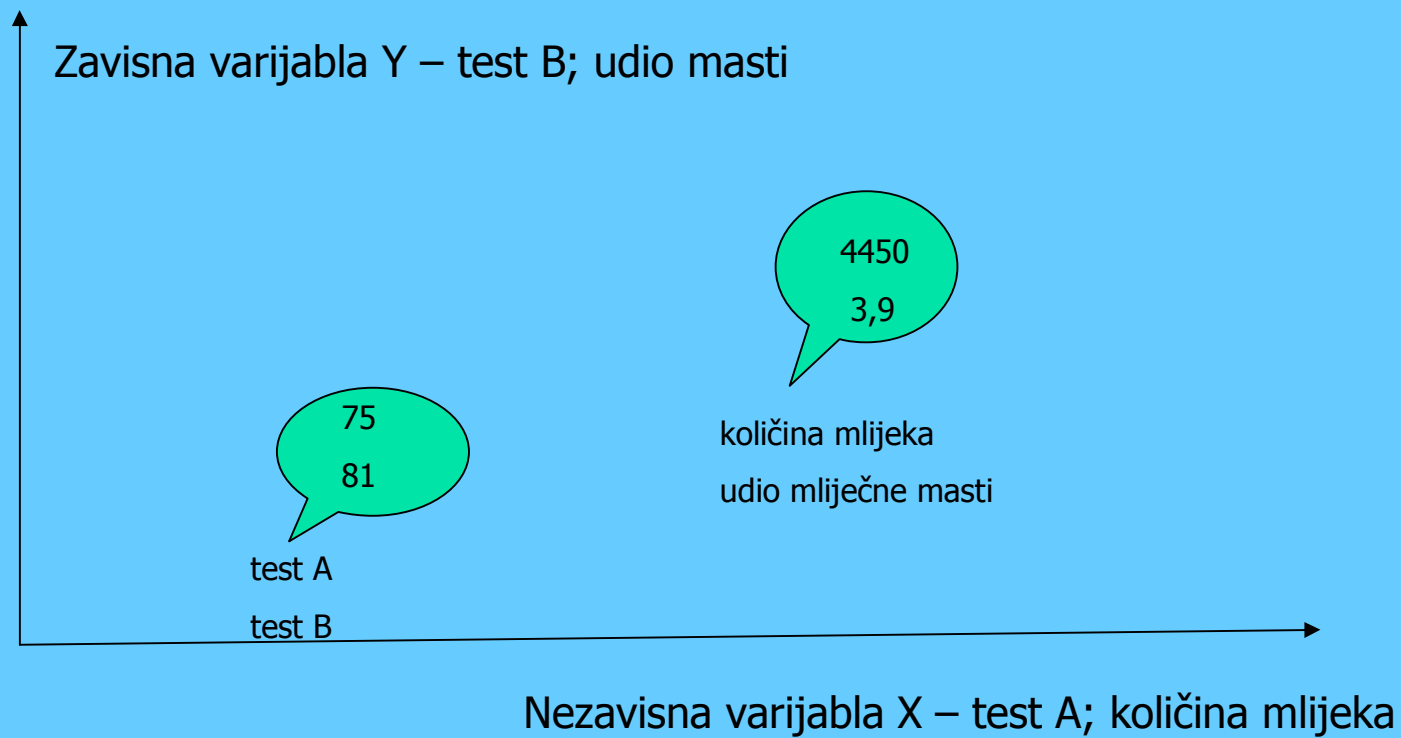
Test A	75	78	88	92	95	67	55	73	74	80	75	95
Test B	81	73	85	85	89	73	66	81	81	80	81	80

SVOJSTVA MLIJEČNOSTI UTVRĐENA KOD POJEDINIH KRAVA

KOLIČINA MLIJEKA (kg)	4450	4050	3950	3810	4230	4200	4900	5100	3850
UDIO MLIJEČNE MASTI (%)	3,9	4,1	4,2	4,4	3,9	3,7	3,4	3,2	4,6

GRAFIČKI PRIKAZ PODATAKA

- KOORDINATNI SUSTAV





KORELACIJA - definicija

- sukladnost u variranju vrijednosti dvije (ili više) varijabli
- označava povezanost između varijabli
- na osnovi same veličine korelacije ne može se ništa zaključivati o uzročno-posljedičnom odnosu među varijablama



KORELACIJA – grafički prikaz i interpretacija

- najjednostavniji oblik primjene korelacijske analize je kada su varijable (napr. varijabla X i varijabla Y) u linernom odnosu

KORELACIJA – grafički prikaz i interpretacija

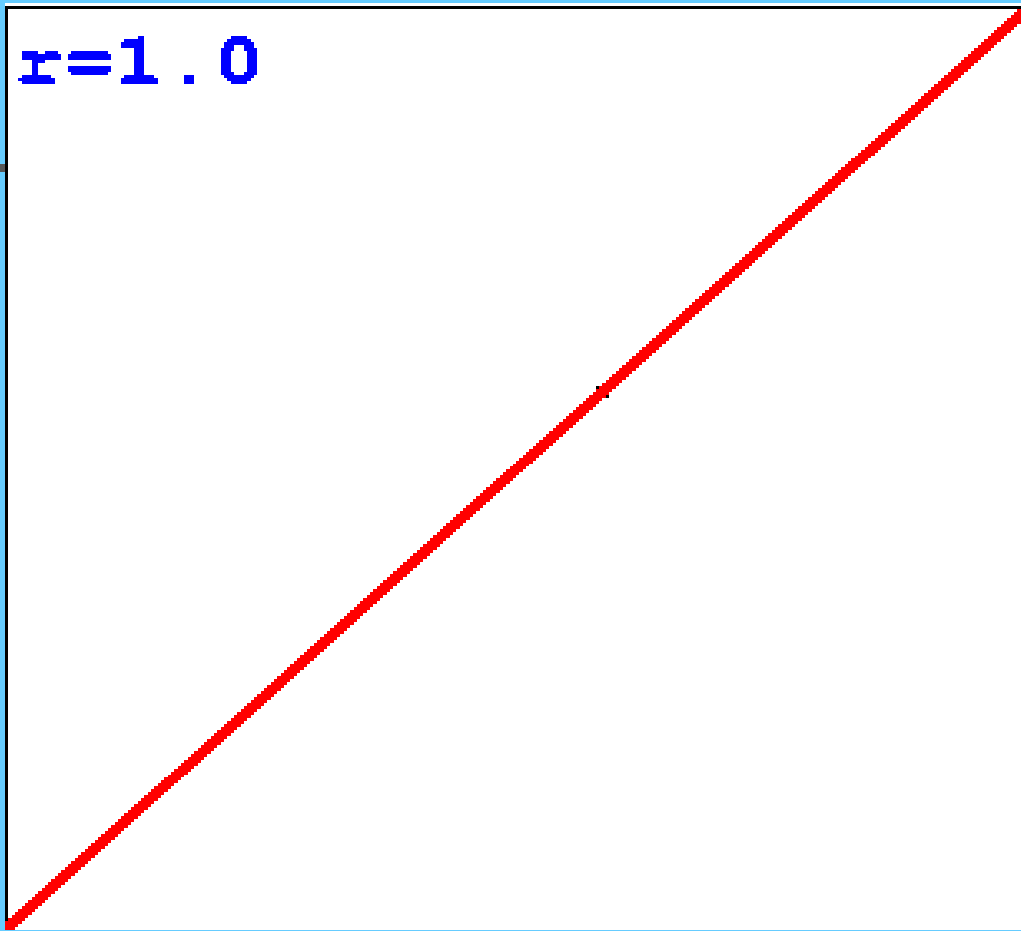
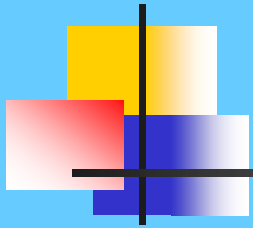


- najveći stupanj sukladnosti u variranju, tzv. potpuna korelacija ili funkcionalna veza, postoji kad svakoj vrijednosti u jednoj varijabli (napr. varijabli X) odgovara samo jedna vrijednost u drugoj varijabli (napr. varijabli Y)
- djelomična korelacija znači da određenoj vrijednosti varijable X odgovara više različitih vrijednosti varijable Y



KORELACIJA – grafički prikaz i interpretacija

- korelacija je manja što ima više različitih vrijednosti varijable Y koje vežemo uz određenu vrijednost varijable X
- korelacija je veća što ima manje različitih vrijednosti varijable Y koje vežemo uz određenu vrijednost varijable X

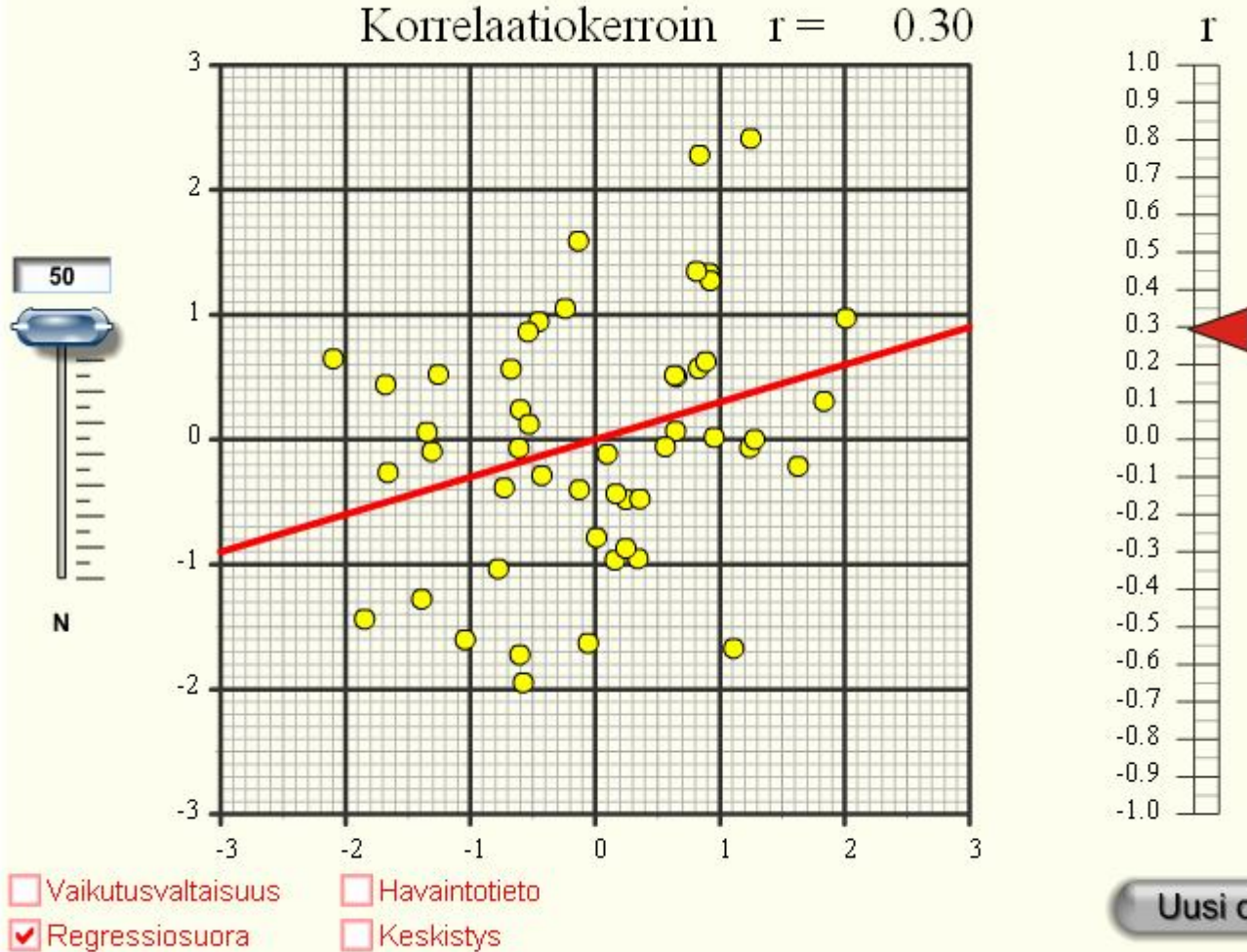


Y

X

Korelacija

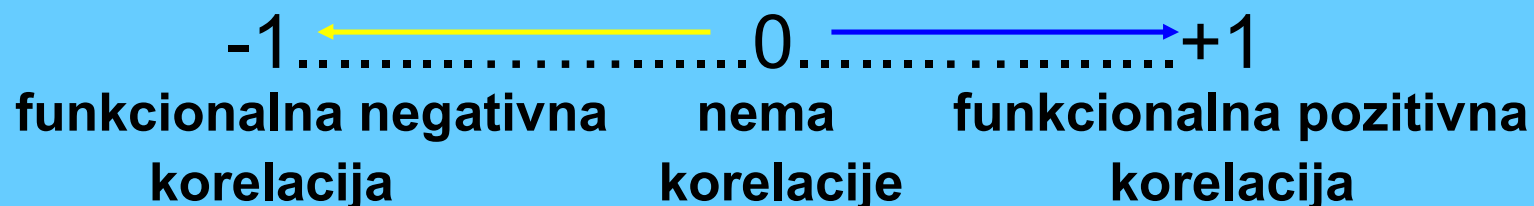
Korrelaatiokerroin $r = 0.30$



IZMJERITELJ KORELACIJE:

Koeficijent korelacije “ r “

- pokazuje stupanj (i smjer) korelacije



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



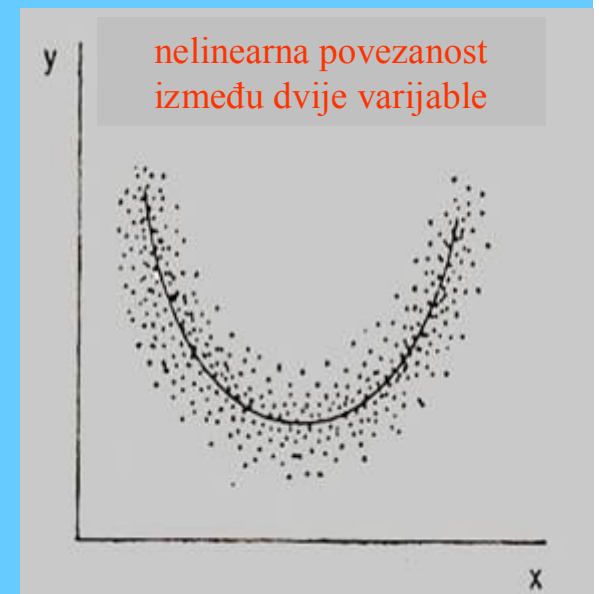
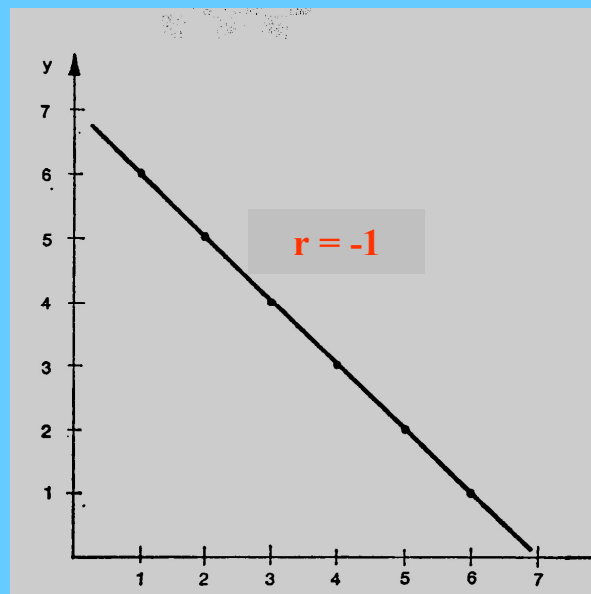
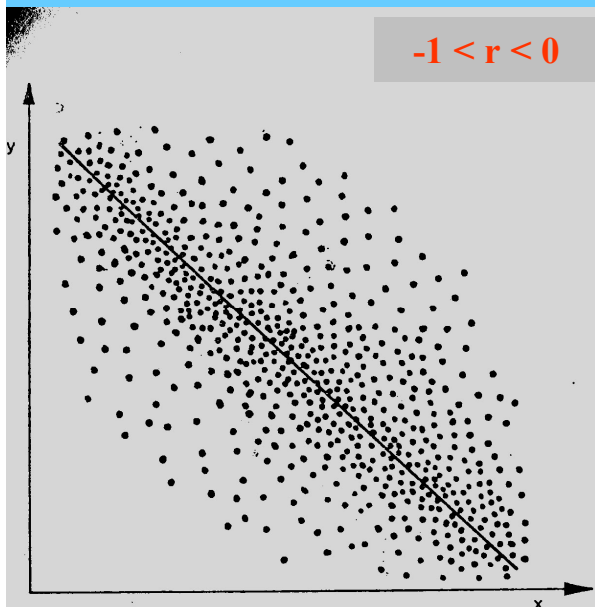
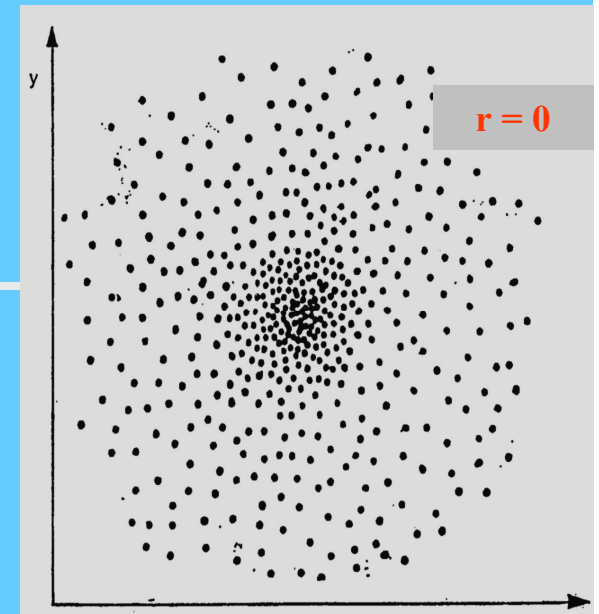
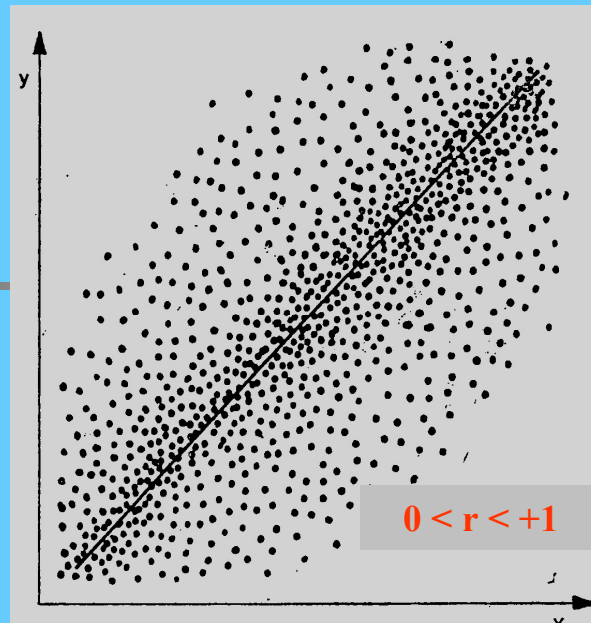
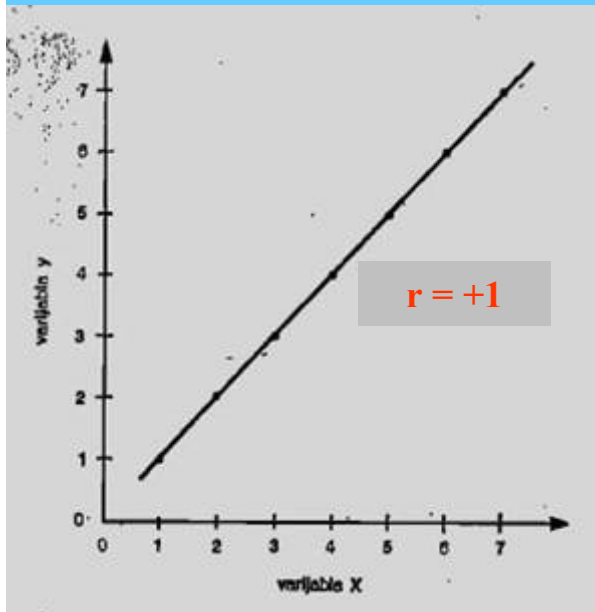
PREDZNAK KOEFICIJENTA KORELACIJE

- **(+)** pozitivan smjer korelacije pokazuje da porast vrijednosti varijable X prati porast vrijednosti varijable Y
- **(-)** negativan smjer korelacije pokazuje da porast vrijednosti varijable X prati opadanje vrijednosti varijable Y



STUPANJ KOEFICIJENTA KORELACIJE (r)

- kreće se u rasponu od 0 do ± 1 (ovisno o smjeru)
- 0 – nema korelacije; + 1 ili – 1 znači apsolutnu korelaciju (relativno rijetko !!!!!)
- najčešće djelomična korelacija – veća od 0 i manja od ± 1





INTERPRETACIJA KOEFICIJENTA KORELACIJE

- prema Petzu (2004.) gruba aproksimacija visine povezanosti – koeficijenta korelacije je:
 - r od $\pm 0,00$ do $\pm 0,20$ nikakva ili neznatna povezanost
 - r od $\pm 0,20$ do $\pm 0,40$ laka povezanost
 - r od $\pm 0,40$ do $\pm 0,70$ značajna povezanost
 - r od $\pm 0,70$ do $\pm 1,00$ visoka ili vrlo visoka povezanost (matematička)

IZMJERITELJ KORELACIJE

Koeficijent determinacije (R)

- označuje se sa $R = r^2$
- R je pokazatelj zajedničkih faktora - udjela kod dva obilježja X i Y koja su uključena u korelacijsku analizu
- npr. $r = 0,3^2 = 0,09 = R$, ili npr.
 $r = 0,6^2 = 0,36 = R$ – koeficijent determinacije
- što je korelacija manja npr. $\pm 0,3$ koeficijent determinacije je značajno manji nego kad je korelacija veća npr. $\pm 0,6$ ($R = 9\%$, odnosno 36%)



REGRESIJE

- znači odnos između odabranih vrijednosti jedne varijable (varijable X) i s njima povezanih vrijednosti druge varijable (varijable Y)
- koristimo je kada želimo istražiti mogućnost procjene vrijednosti jedne varijable (varijabla Y) na temelju poznate vrijednosti druge varijable (varijable X)



REGRESIJSKA ANALIZA

- vrijednosti mjerenja predstavljene su kao parovi rezultata koji se sastoje od podatka za nezavisnu varijablu (obično varijablu X) i od podatka za zavisnu varijablu (obično varijablu Y)
- na osi X su vrijednosti za varijablu X , a na osi Y vrijednosti za varijablu Y sjecišta parova rezultata su prikazana kao točke u koordinatnom sustavu



JEDNADŽBA PRAVCA REGRESIJE

- u koordinatnom sustavu pravac “spaja” točke čiji je položaj određen mjerenjem varijable X i varijable Y
- crta regresije pokazuje tip odnosa između varijabli X i Y:
 - ako je pravac, onda je povezanost linearna i određena jednadžbom pravca regresije
 - ako nije pravac, tada se radi o zakrivljenoj korelaciji



JEDNADŽBA PRAVCA REGRESIJE

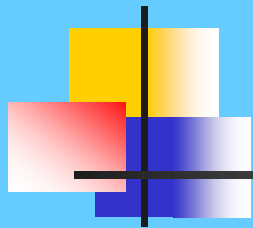
- kad se crta regresije, prikazana u koordinatnom sustavu, može opravdano opisati pravcem tada se i matematički može opisati jednadžbom pravca:

$$Y = a + bX$$

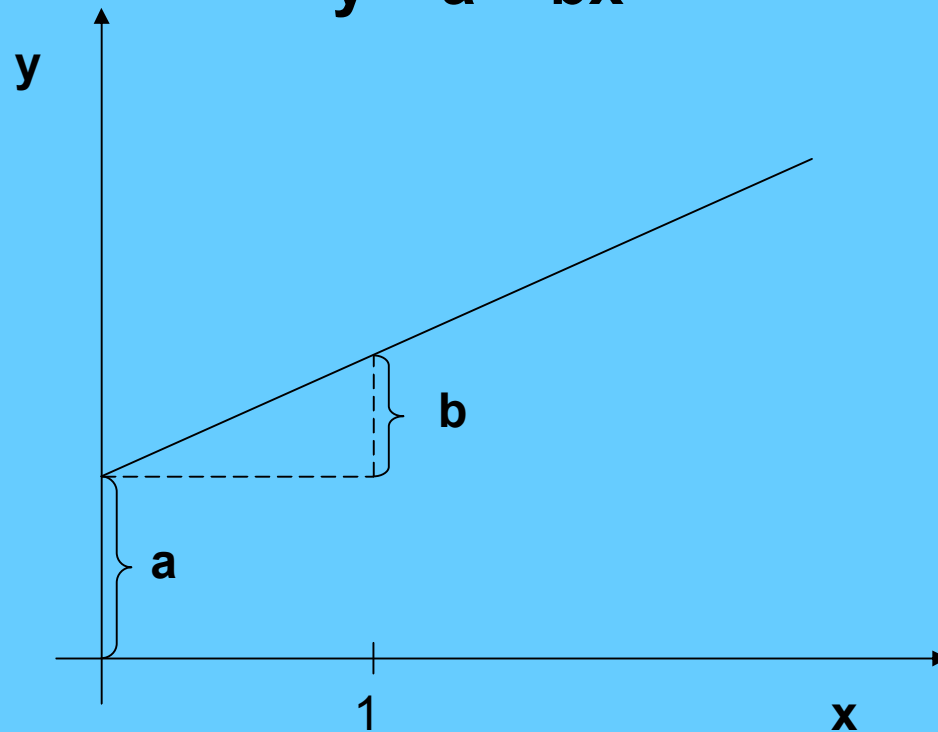
Y = najvjerojatniji rezultat varijable y za određeni rezultat varijable X

a = odsječak na osi Y (konstanta)

b = koeficijent nagiba (konstanta)



$$y = a + bx$$

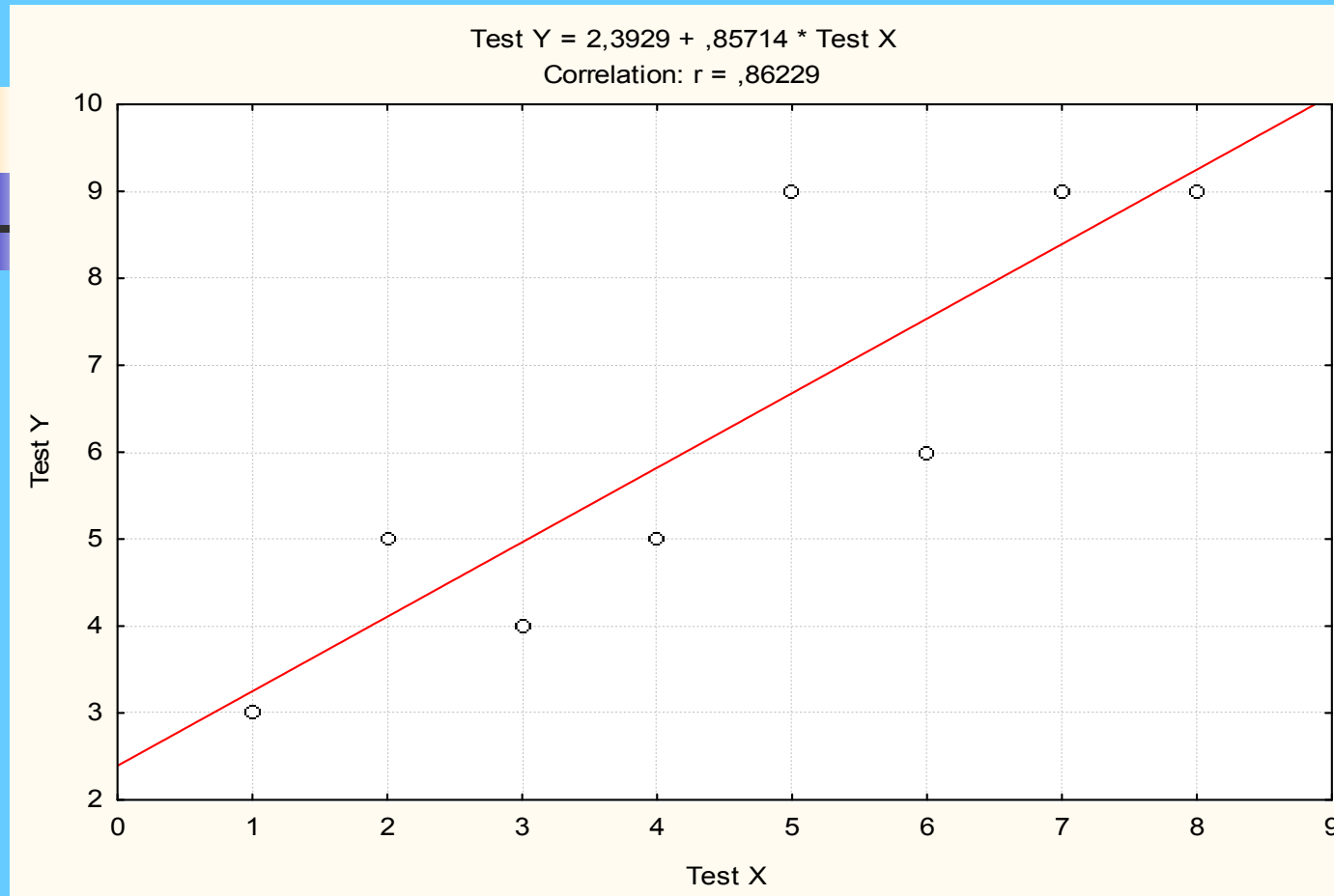


$y, x =$ varijable

$a =$ odsječak osi y u kojem pravac siječe ordinatu

$b =$ nagib pravca, porast na osi y za pojedinačni porast na osi x

Primjer: pravac regresije za rezultate ostvarene na testu X i testu Y (8 studenata)



Interpretacija: korelacija između rezultata na testu X i testu Y je pozitivna i visoka ($r = 0,86229$); više bodova na testu X znači više bodova na testu Y; na temelju rezultata na testu X možemo predvidjeti broj bodova na testu Y preko jednadžbe

$$\text{Broj bodova na testu Y} = 2,3929 + 0,85714 \times \text{broj bodova na testu X}$$