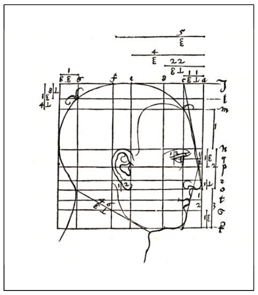
Metodologija psiholoških istraživanja 1

obrada koreg 1



10. decembar 2018.

IV. Obrada podataka
A. Frekvencijski nacrti
B. Faktorijalni nacrti
C. Korelaciono-regresioni nacrti (KRN)
1. bivarijatni korelaciono-regresioni nacrti

C. Korelaciono-regresioni nacrti (KRN) 2

- podsetimo se tipova nacrta čiju obradu razmatramo u ovom predmetu:
  - *A. frekvencijski nacrti:*
    - *3 tipa:* nacrti sa jednom, sa dve i sa tri varijable
  - *B. faktorijalni nacrti:*
    - *3 tipa:* nacrti sa dve, sa tri i sa četiri varijable
  - *C. korelaciono-regresioni nacrti:*
    - *2 tipa:* nacrti sa dve i sa tri varijable
      - nacrti sa dve varijable: *bivarijatni* korelaciono-regresioni nacrti
      - nacrti sa više od dve varijable: *multivarijatni* korelaciono-regresioni nacrti

C. Korelaciono-regresioni nacrti (KRN) 3

1. Bivarijatni korelaciono-regresioni nacrti (BKRN)

- BKRN su KRN nacrti sa dve varijable, X i Y
- vrste varijabli u BKRN:
  - prema načinu izražavanja vrednosti: numeričke ili kategoričke
    - međutim, u *tipičnim* BKRN obe varijable su numeričke
      - samo takvim slučajevima ćemo se baviti
    - ako su obe kategoričke: frekvencijski nacrt
    - ako je jedna numerička jedna kategorička: faktorijalni ili diskriminacioni nacrt
  - prema stepenu kontrole: nema ograničenja
    - manipulativne, selektivne ili registrovane
      - registrovane varijable su česte u istraživanjima u kojima se ne koriste ni manipulacija ni selekcija
- kao i svi KRN, i BKRN imaju dva aspekta:
  - *korelacioni aspekt:* ispitivanje povezanosti X i Y
  - *regresioni aspekt:* predviđanje X na osnovu Y, ili obrnuto

1. Bivarijatni KRN 4

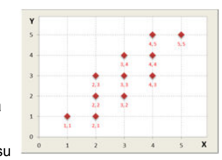
a. Organizacija i prikaz podataka

koristi se matrični i grafički prikaz

- matrica podataka: tipa objekti x varijable
  - *kolone:* dve varijable; *redovi:* za svaki objekt po jedan red
  - *PRIMER:* koristićemo dve *interpretacije* podataka:
    - (1) X: dužina učenja, Y: dobijena ocena
    - (2) X: ocena iz srpskog, Y: ocena iz matematike

#	X	Y
1.	1	1
2.	2	1
3.	2	2
4.	2	3
5.	3	2
6.	3	3
7.	3	4
8.	4	3
9.	4	4
10.	4	5
11.	5	5
zbir	*33*	*33*
prosek	*3*	*3*

- grafikon podataka
  - naziv: korelacioni dijagram ili skater-plot
  - principi grafičkog prikaza podataka:
    - svaki objekt je prikazan *markerom* (grafičkom oznakom) na grafikonu
    - *položaj* markera na grafikonu je određen vrednostima X i Y varijable datog objekta
    - te vrednosti su X i Y *ko-ordinate* markera
      - na prikazanom grafikonu, te ko-ordinate su naznačene brojevima (što se obično ne čini)



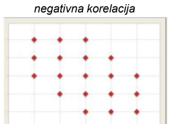
1. Bivarijatni KRN 5

- markeri u korelacionom dijagramu grade izvesne geometrijske *oblike*
  - u vrlo uprošćenim ali karakterističnim slučajevima, ti oblici su obrisi *elipsi*
- *PRIMERI:* obrisi markerskih elipsi za različite smerove korelacija

pozitivna korelacija

elipsa nagnuta od dole levo ka gore desno
kako raste X, tako raste Y

negativna korelacija



elipsa nagnuta od gore levo ka dole desno
kako raste X, tako opada Y

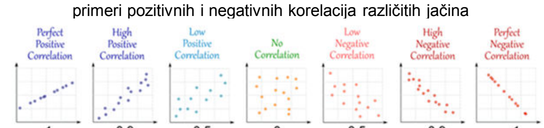
slaba ili nikakva korelacija

elipsa koja nije nagnuta
kako raste X, Y se ne menja na neki karakterističan način

- *promer* elipse odražava *jačinu* korelacije
  - što je elipsa tanja, korelacija je jača
- slučaj ekstremne korelacije: $r_{XY} = 1$ (ili -1):
  - markerska 'elipsa' 'degeneriše' u pravu liniju

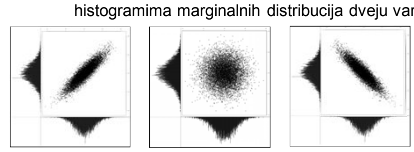
1. Bivarijatni KRN 6

primeri pozitivnih i negativnih korelacija različitih jačina



KK: 1 | 0.9 | 0.5 | 0 | -0.5 | -0.9 | -1

primeri korelacionih dijagrama kombinovanih sa histogramima marginalnih distribucija dveju varijabli



pozitivna korelacija | *nulta korelacije* | *negativna korelacija*

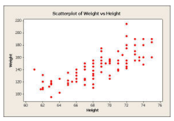
uočiti: samo na osnovu *marginalnih* distribucija ne može se ništa reći o tipu korelacije

- napomena: korelacioni dijagrami u *realnim* slučajevima često su složeniji, 'neuredniji', i teži za interpretaciju

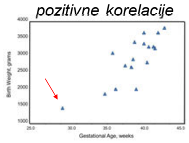
1. Bivarijatni KRN
7

primeri korelacionih dijagrama različitih istraživanja

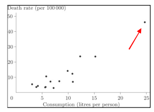
pozitivne korelacije



visina i težina osoba

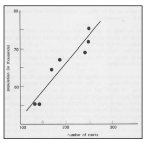


gestaciona starost i težina novorođenčeta

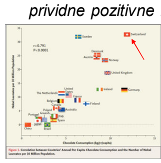


konzumacija alkohola i stopa smrtnosti u državama

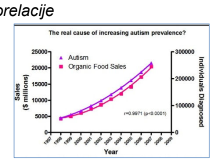
prividne pozitivne korelacije



broj stanovnika i broj roda u gradovima



potrošnja čokolade i relativni broj Nobelovih nagrada u državama

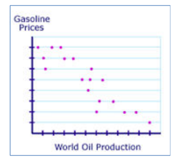


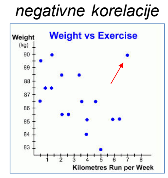
uočiti: dva preklopljena dijagrama, sa istom x-osom i dve y-ose

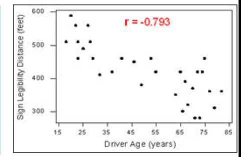
1. Bivarijatni KRN
8

primeri korelacionih dijagrama različitih istraživanja

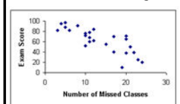
negativne korelacije

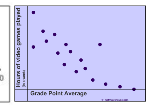




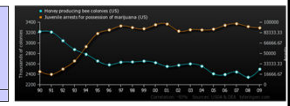


navodne negativne korelacije





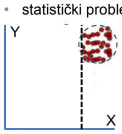
prividna negativna korelacija



uočiti: raznovrsnost objekata istraživanja u primerima (ljudi, gradovi, države..)

1. Bivarijatni KRN
9

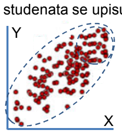
- stepen korelacije zavisi od izbora *raspona vrednosti* varijabli
- *PRIMER 1:* selekcija raspona na jednoj od dve varijable
  - X: uspeh na prijemnom ispitu; Y: uspeh na studijama
  - korelacija može da bude, neočekivano, relativno niska
  - statistički problem: samo deo prijavljenih studenata se upisuje (za X>kriterijum)



uzak raspon vrednosti X

korelacioni dijagram samo sa stvarno upisanim studentima: samo za X>kriterijum, tj. svi imaju visoki X

korelacija slaba

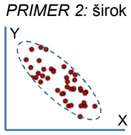


širok raspon vrednosti X

korelacioni dijagram kada bi se upisali svi koji su polagali prijemni ispit, tj., X uzima *sve* vrednosti

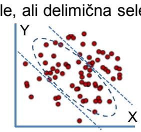
korelacija postoji

- *PRIMER 2:* širok raspon na obe varijable, ali delimična selekcija



korelacioni dijagram kada bi se *zanemarili* slučajevi kod kojih su i X i Y mali, i kod kojih su i X i Y veliki

korelacija postoji



korelacioni dijagram kada bi se prikazali *svi* slučajevi, bez ograničenja veličine

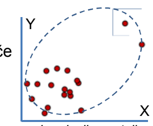
korelacija slaba

primer: izbor partnera

1. Bivarijatni KRN
10

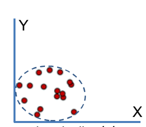
- stepen korelacije zavisi od prisustva *autlajera (iznimaka)*
  - autlajeri mogu da štrče po obe varijable, ili po samo jednoj, ili ni po jednoj

PRIMER 1: autlajeri koji štrče po obe varijable (i po X i po Y)



korelacioni dijagram kada su autlajeri uključeni

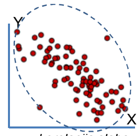
korelacija postoji



korelacioni dijagram kada su autlajeri isključeni

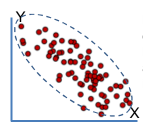
korelacija slaba

PRIMER 2: autlajeri koji ne štrče ni po jednoj varijabli, ali štrči njihova *kombinacija*



korelacioni dijagram kada su autlajeri uključeni

korelacija slaba



korelacioni dijagram kada su autlajeri isključeni

korelacija postoji

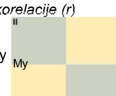
- uočiti: prisustvo autlajera se lako primećuju *vizuelnim* uvidom u grafikon, a ne može se detektovati ako se samo izračuna *numerička* vrednost r_{XY}
- naravoučenije: uvek pogledati korelacioni dijagram!

1. Bivarijatni KRN
11

b. Deskriptivne statističke mere, prikaz, struktura

(1) korelacioni aspekt BKRN

- mera korelacije: broj koji izražava smer i stepen povezanosti dve varijable
- do sada razmatrane mere (pokazatelji) korelacije varijabli X i Y:
  - bivarijatni frekvencijski nacrti (BFN):
    - i X i Y su dihotomije (tip 2x2): *fi-koeficijent, količnik šansi*
    - složenji nacrti: *Kramerov fi-koeficijent*
  - bivalentni jednofaktorski nacrti (BJFN)
    - X je dihotomija, Y je kvantitativna: *point biserijalni koeficijent*
- novo: mere korelacije u bivarijat. korelaciono-regresionim nacrtima (BKRN)
    - i X i Y su kvantitativne
  - opisaćemo tri mere korelacije u BKRN
    - *zbir proizvoda (SP), kovarijansa (C), Pirsonov koeficijent korelacije (r)*
- razmotrićemo nekoliko korelacionih diagrama
  - u njima će biti posebno naznačene *prosečne* vrednosti Mx i My
    - Mx: vertikalna linija; My: horizontalna linija
    - linije dele dijagram na četiri kvadranta: I, II, III i IV

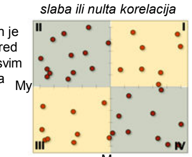


1. Bivarijatni KRN
12

jačina i smer korelacije, i raspored markera po kvadrantima

slaba ili nulta korelacija

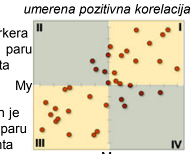
uočiti: sličan je broj i raspored markera u svim kvadrantima



umerena pozitivna korelacija

više markera je u I i III paru kvadranta

manje ih je u II i IV paru kvadranta



jaka pozitivna korelacija

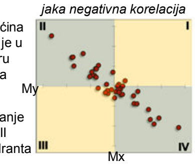
velika većina markera je u I i III paru kvadranta

mnogo manje ih je u II i IV paru kvadranta

jaka negativna korelacija

velika većina markera je u II i IV paru kvadranta

mnogo manje ih je u I i III paru kvadranta



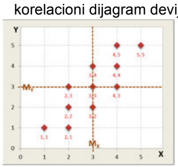
uočiti: *jačina* korelac. zavisi od koncentracije markera u parovima kvadranta (I i III, II i IV)
smer korelac. (+ ili -) zavisi od toga u *kojem* paru (I i III, ili II i IV) je koncentracija

1. Bivarijatni KRN

13

- opisane osobine rasporeda markera mogu se izraziti računanjem *devijacija* x i y varijabli X i Y
- x-devijacije: $x = X - M_X$
  - horizontalna odstojanja markera od proseka M_X
  - važi: za $X > M_X$, $x > 0$; za $X < M_X$, $x < 0$
- y-devijacije: $y = Y – M_Y$
  - vertikalna odstojanja markera od proseka M_Y
  - važi: za $Y > M_Y$, $y > 0$; za $Y < M_Y$, $y < 0$
- korelacioni dijagram devijacija: isti izgled kao za mere!

	mere		devijacije	
#	X	Y	x	y
1.	1	1	-2	-2
2.	2	1	-1	-2
3.	2	2	-1	-1
4.	2	3	-1	0
5.	3	2	0	-1
6.	3	3	0	0
7.	3	4	0	1
8.	4	3	1	0
9.	4	4	1	1
10.	4	5	1	2
11.	5	5	2	2
zbir	33	33	0	0
prosek	3	3	0	0



korelacioni dijagram mera X i Y

korelac. dijagram devijacija x i y

setimo se: devijacija je AT (translacija), a skup devijacija je na brojnoj x-osi centriran u nuli

skup devijacija x i y na korelac. dijagramu je *dvostruko* transliran (na obe ose) i centriran u nuli

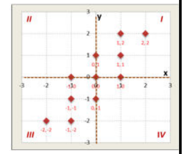
- devijacije x i y su pogodne za matematičko definisanje mera korelacije
- u tu svrhu x i y se *množe*, tj. koriste se *proizvodi devijacija* xy, koji se *sabiraju* (Σxy)

1. Bivarijatni KRN

14

- 1. zbir proizvoda devijacija: $SP_{XY} = \Sigma xy$
- osobine proizvoda devijacija xy, i njihovog zbira Σxy
  - u kvadrantu I: $x > 0, y > 0$, pa je i proizvod $xy > 0$
  - u kvadrantu II: $x < 0, y > 0$, pa je proizvod $xy < 0$
  - u kvadrantu III: $x < 0, y < 0$, pa je proizvod $xy > 0$
  - u kvadrantu IV: $x > 0, y < 0$, pa je proizvod $xy < 0$
  - ako je x ili y nula (na osama koor. sistema): $xy = 0$
- dakle: u I i III: $\Sigma xy > 0$; u II i IV: $\Sigma xy < 0$ (odn: $-\Sigma xy > 0$)
- ukupan zbir proizvoda SP = Σxy (za I, II, III i IV) biće:
  - *(a) pozitivan:* ako je Σxy za I i III *veći* od $-\Sigma xy$ za II i IV
  - *(b) negativan:* ako je Σxy za I i III *manji* od $-\Sigma xy$ za II i IV
  - *(c) nulti (ili mali):* ako je Σxy za I i III *jednak (ili sličan)* kao $-\Sigma xy$ za II i IV
- odavde sledi da je Σxy *mera korelacije* X i Y! naime:
  - (a) će uglavnom biti slučaj kada je više markera u I i III, a manje u II i IV, tj. kad je korelacija *pozitivna*
  - (b) će uglavnom biti slučaj kada je obrnuto, tj. kada korelacija *negativna*
  - (c) će biti slučaj ako je korelacija *nulta* ili *mala*

#	X	Y	x	y	xy
1.	1	1	-2	-2	4
2.	2	1	-1	-2	2
3.	2	2	-1	-1	1
4.	2	3	-1	0	0
5.	3	2	0	-1	0
6.	3	3	0	0	0
7.	3	4	0	1	0
8.	4	3	1	0	0
9.	4	4	1	1	1
10.	4	5	1	2	2
11.	5	5	2	2	4
zbir	33	33	0	0	14
prosek	3	3	0	0	1.3



1. Bivarijatni KRN

15

- 2. kovarijansa: *prosečni* zbir proizvoda devijacija: $C_{XY} = \Sigma xy/N$
- uočiti: kada je zbir proizvoda Σxy pozitivan (+), negativan (-), ili nulti (0), tada je i kovarijansa $\Sigma xy/N$ pozitivna, negativna, ili nulta
- kovarijansa je i po imenu i po strukturi srodna *varijansi*
  - *kovarijansa:* prosek zbira *proizvoda* devijacija: $C_{XY} = \Sigma xy/N$
  - *varijansa:* prosek zbira *kvadrata* devijacija: $V = \Sigma x^2/N$
  - razlika: varijansa se odnosi na jednu varijablu, kovarijansa na dve
  - uočiti: varijansa je kovarijansa varijable 'sa sobom', tj. $V = C_{XX}$ (tj. slučaj y = x)
  - takođe: SS (zbir kvadrata) je SP varijable same sa sobom, tj. $SS = SP_{XX}$
- jednačine sa N u imeniocu odnose se na *uzorak*
- *populacione* jednačine sadrže N-1 (stepeni slobode), tj. $C_{XY}=\Sigma xy/(N-1)$
- 'mana' kovarijanse: veličina zavisi od mernih jedinica varijabli
- *PRIMER:* korelacija visine i težine, izražene u različitim jedinicama
  - visina: metri, santimetri, inči; težina: kilogrami, grami, funte
  - što su brojke koje izražavaju mere veće, veće su i devijacije, a time i kovarijanse
  - međutim, radi se o korelaciji istih pojava, i bilo bi pogodnije da ona bude izražena istom brojkom, bez obzira na korišćene merne jedinice

1. Bivarijatni KRN

16

- 3. koeficijent korelacije: $r_{XY} = \Sigma z_x z_y/N$
  - populaciona verzija (koja se koristi u testovima): $\Sigma z_x z_y/(N-1)$
  - duži naziv: *Pirsonov produkt-moment koeficijent linearne korelacije*
  - daleko najpopularnija mera korelacije
- formula je ista kao za kovarijansu, osim što su x i y zamenjeni sa z_x i z_x
  - z_x i z_x su *standardne vrednosti* mera X i Y, tj. $z_x = (X\text{-}Mx)/SDx$, $z_Y = (X\text{-}My)/SDy$
- kada su zbir proizvoda SP i kovarijansa C_{XY} +, -, ili 0, tada je i r_{XY} +, -, ili 0
- može se pokazati da veličina r_{XY}:
  - ne zavisi od mernih jedinica
    - npr. ista je bilo da se koriste metri i kilogrami, ili santimetri i grami, itd.
  - mora ležati između -1 i +1
- *PRIMER:* računanje r_{XY}:
  - moraju se računati x^2 i y^2, radi računanja SDx i SDy, pa preko njih z_x i z_x, i na kraju:
  - $r_{XY} = 0.84$
- postoje i brojne druge, *ekvivalentne* formule za r_{XY}

#	X	Y	x	y	xy	x^2	y^2	z_x	z_y	$z_x z_y$
1.	1	1	-2	-2	4	4	4	-1.77	-1.48	2.63
2.	2	1	-1	-2	2	1	4	-0.89	-1.48	1.32
3.	2	2	-1	-1	1	1	1	-0.89	-0.74	0.66
4.	2	3	-1	0	0	1	0	-0.89	0	0
5.	3	2	0	-1	0	0	1	0	-0.74	0
6.	3	3	0	0	0	0	0	0	0	0
7.	3	4	0	1	0	0	1	0	0.74	0
8.	4	3	1	0	0	1	0	0.89	0	0
9.	4	4	1	1	1	1	1	0.89	0.74	0.66
10.	4	5	1	2	2	1	4	0.89	1.48	1.32
11.	5	5	2	2	4	4	4	1.77	1.48	2.63
zbir	33	33	0	0	14	14	20	0	0	9.20
pros.	3	3	0	0	1.3	1.27	1.82	0	0	r=0.84

1. Bivarijatni KRN

17

korelacije i transformacije

- kako na koeficijent korelacije r_{XY} utiču *transformacije* varijabli X i Y?
  - transformacije: aditivna (AT), multiplikativna (MT), i linearna (LT)
- prikazaćemo dve grupe primera uticaja različitih transformacija kojima se uvodi nova varijabla Z, razmatranjem veličine r_{XZ}

1. slučajevi kada je varijabla Z transformacija varijable X

X	Z	transformacija	r_{XZ}
X	Z = A + X	AT	1
X	Z = X	AT za slučaj A = 0	1
X	Z = x	AT za slučaj A = -Mx	1
X	Z = BX	MT	±1
X	Z = -X	MT za slučaj B = -1	-1
X	Z = A + BX	LT	±1
X	$Z = z_X$	LT za stand. mere	1

2. slučajevi kada je varijabla Z transformacija varijable Y

X	Z	transformacija	r_{XZ}
X	Z = A + Y	AT	r_{XY}
X	Z = Y	AT za slučaj A = 0	r_{XY}
X	Z = y	AT za slučaj A = -My	r_{XY}
X	Z = BY	MT	$\pm r_{XY}$
X	Z = -Y	MT za slučaj B = -1	$-r_{XY}$
X	Z = A + BY	LT	$\pm r_{XY}$
X	$Z = z_Y$	LT za stand. mere	r_{XY}

- *zaključak:* linearna transformacija LT (uključujući i njene posebne slučajeve, AT i MT) ne utiče na veličinu korelacije (sem eventualne promene predznaka, i to za B<0)
  - *drugim rečima:* linearna transformacija neke varijable se u korelacijama 'ponaša' skoro isto kao sama ta varijabla (izuzev što se menja predznak korelacije ako je B<0)

1. Bivarijatni KRN

18

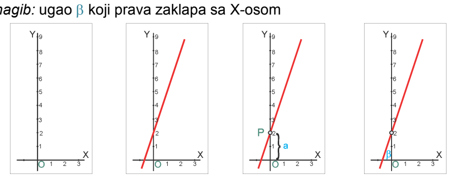
(2) regresioni aspekt BKRN

- osnovni zadatak: za dato X, proceniti odn. predvideti Y
  - X: prediktorska varijabla, nezavisna varijabla (NV)
  - Y: kriterijumska varijabla, zavisna varijabla (ZV)
  - napomena: uloge dve varijable mogu se obrnuti
    - tj., da prediktorska postane kriterijumska, i obrnuto
- statistički postupak: regresiona analiza odn. regresija
- osnovni problemi regresione analize:
  - kako na osnovu X *aproksimirati* (približno prikazati) Y?
  - kako zavisnost Y od X prikazati grafički?
  - kako zavisnost Y od X izraziti matematičkom formulom?
    - opšti, simbolički oblik formule: Y = f(X)
    - čita se: 'Y je f od X'; kaže se: 'Y je *funkcija* od X'
- u matematici postoji veoma veliki broj funkcija Y = f(X)
- među najprostije funkcije spada *linearna* funkcija
- linearna funkcija koristi se u postupku linearne regresije
  - opisaćemo *grafički* prikaz i *algebarsku* formulaciju linearne funkcije

1. Bivarijatni KRN
19

linearna funkcija

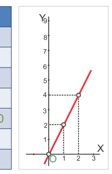
- grafički prikaz (geometrijska formulacija) linearnog odnosa X i Y
  - koristi se ko-ordinatni sistem sa osama X i Y, koje se pod pravim uglom seku u ko-ordinatnom početku O
  - zavisnost varijable Y od varijable X se grafički prikazuje preko *prave linije*
  - uočiti: postoji beskonačno mnogo pravih linija u ravni
  - položaj određene, konkretne prave linije se specifikuje pomoću dve konstante odn. dva *parametra*:
    - *intercept:* razmak a od ko-ordinatnog početka O do preseka P prave sa Y-osom
    - *nagib:* ugao β koji prava zaklapa sa X-osom



1. Bivarijatni KRN
20

- algebarska formulacija linearnog odnosa X i Y
  - algebarska jednačina prave linije glasi: Y = a + bX
    - uočimo: ova jednačina izražava *linearnu transformaciju* varijable X u Y
  - u jednačini se pojavljuju četiri elementa: dve varijable i dve konstante
    - varijable: X: nezavisna varijabla; Y: zavisna varijabla
    - konstante: *parametar a*: intercept; *parametar b*: to je tangens ugla β
    - zavisno od položaja prave u ravni, a i b mogu biti pozitivni, negativni, ili nulti
- *PRIMER:* neka je a = 0, b = 2, tako da jednačina glasi Y = 2X
  - ovo je multiplikativna transformacija
  - za različite vrednosti X mogu se izračunati odgovarajuće vrednosti Y
    - u ovom slučaju, svako Y je dva puta veće od odgovarajućeg X
  - važno: parovi vrednosti X i Y se mogu koristiti kao *ko-ordinate* tačaka u ravni
  - uočiti: sve takve tačke leže na pravoj liniji!
    - parametar b izražava nagib te prave
  - takva kombinacija grafičkog i algebarskog prikaza spada u *analitičku geometriju*

X	Y = 2*X
0	Y = 2*0 = 0
1	Y = 2*1 = 2
2	Y = 2*2 = 4
...	...
100	Y = 2*100 = 200
...	...
-1	Y = 2*(-1) = -2



1. Bivarijatni KRN
21

kako se *menja* prava, ako se u jednačini Y = bX menja vrednost parametra b?

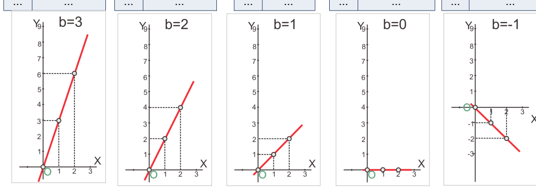
X	Y = 3*X
0	Y = 3*0 = 0
1	Y = 3*1 = 3
2	Y = 3*2 = 6
...	...

X	Y = 2*X
0	Y = 2*0 = 0
1	Y = 2*1 = 2
2	Y = 2*2 = 4
...	...

X	Y = X
0	Y = 0
1	Y = 1
2	Y = 2
...	...

X	Y = 0*X
0	Y = 0*0 = 0
1	Y = 0*1 = 0
2	Y = 0*2 = 0
...	...

X	Y = -X
0	Y = -1*0 = 0
1	Y = -1*1 = -1
2	Y = -1*2 = -2
...	...



- menja se nagib prave!
  - ako je b>0, nagib je pozitivan, ako je b<0, nagib je negativan, a ako je b=0, nagib je nulti
- uočiti: u svim gornjim primerima: a=0
  - to znači da je intercept nulti, tj. da sve te prave prolaze kroz ko-ordinatni početak

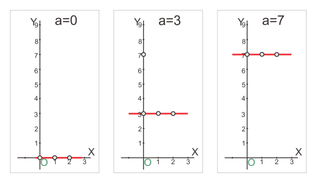
1. Bivarijatni KRN
22

- kako se menja prava, ako je Y = a, (tj. b=0), a menja se vrednost parametra a?

X	Y = 0
0	Y = 0
1	Y = 0
2	Y = 0
...	...

X	Y = 3
0	Y = 3
1	Y = 3
2	Y = 3
...	...

X	Y = 7
0	Y = 7
1	Y = 7
2	Y = 7
...	...

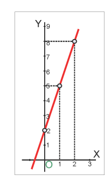


- prava je *horizontala*, a vrednost intercepta a određuje njenu *visinu* u odnosu na X-osu
- uočiti: Y se ne menja kada se X menja, drugim rečima Y ne zavisi od X

- kako izgleda prava ako je Y = a + b*X (pritom su i a i b različiti od nule)?

neka je:
a=2, b=3
tada je:
Y = 2 + 3*X

X	Y = 2 + 3*X
0	Y = 2+3*0 = 2
1	Y = 2+3*1 = 5
2	Y = 2+3*2 = 8
...	...



- b određuje *nagib* prave
- a određuje njenu *visinu*

1. Bivarijatni KRN
23

- *PRIMER:* formula idealne težine (izmišljena)
  - *'idealna težina (u kg) se dobija ako se od visine (u cm) oduzme 100'*
    - npr. za visinu od 160 cm idealna težina je 60 kg
  - odgovarajuća jednačina glasi: idealna težina = visina – 100
  - uočiti: radi se o linearnoj funkciji Y = a + bX
    - varijable: nezavisna var. X: visina; zavisna var. Y: idealna težina
    - parametri: aditivna konstanta a = -100; multiplikativna konstanta b = 1
- *PRIMER:* $E = Mc^2$ (Ajnštajnova jednačina odnosa mase i energije)
  - uočiti: i ovde se radi o linearnoj funkciji Y = a + bX
    - varijable: nezavisna var. X: masa M; zavisna var. Y: energija E
    - parametri: aditivna konstanta a = 0; multiplikativna konstanta $b = c^2$
- osim linearne funkcije Y = a + bX postoje i brojne *nelinearne* funkcije
  - njihov grafički odraz nisu prave već krive linije
  - u njihovoj algebarskoj formuli nezavisna var. X se pojavljuje u složenijem obliku
    - npr. stepenovana kao X^2, X^3, transformisana kao log(X), sin (X), itd
    - *PRIMER:* jednačina slobodnog pada: $d = gt^2/2$
      - nez. var. X: t; zav. var. Y: d; ad. konst. a: 0; mult. konst: b: g/2 (g je grav. konst.)
    - napomena: u linearnoj jednačini, X je stepenovana *jedinicom*, naime $X = X^1$

1. Bivarijatni KRN
24

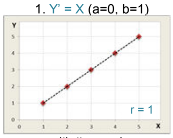
- primena linearne jednačine u statistici
- podaci iz bivarijatnog KRN predstavljaju se korelacionim dijagramom
  - *idealan slučaj:* ponekad (vrlo retko) svi markeri leže tačno na pravoj liniji
    - tada se odnos može *egzaktno* algebarski prikazati linearnom jednačinom
    - za dato X može se precizno i ispravno izračunati odgovarajuće Y
      - linearna jednačina opisuje *zakonitost* zavisnosti Y od X
  - *realan slučaj:* često (ne uvek) markeri zauzimaju obris sličan elipsi
    - elipsa se može *aproksimirati* (približno prikazati) pravom linijom
      - ta prava linija naziva se regresiona prava
    - regresiona prava se može algebarski izraziti jednačinom
      - ta jednačina se naziva regresiona jednačina i glasi: Y' = a + bX
        - X: nezavisna varijabla
        - Y': *procena* zavisne varijable Y, na osnovu regresione jednačine
          - često se označava sa $\hat{Y}$
        - a, b: regresioni parametri
    - procena varijable Y varijablom Y' praktično nikad neće biti egzaktna
      - pojaviće se *odstupanja* od predikcije, odn. greške
      - što su greške manje, procena je preciznija

1. Bivarijatni KRN

25

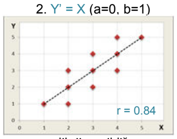
- *PRIMERI:* regresiona prava $Y' = a + bX$ (prikazana tačkastom linijom)

1. $Y' = X$ (a=0, b=1)



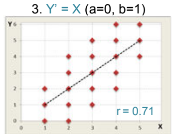
predikcije: egzaktne
greške: nulte
korelacija: savršena

2. $Y' = X$ (a=0, b=1)



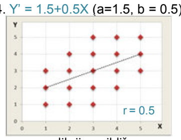
predikcije: približne
greške: relativno male
korelacija: vrlo visoka

3. $Y' = X$ (a=0, b=1)



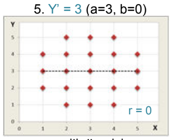
predikcije: približne
greške: veće
korelacija: visoka

4. $Y' = 1.5+0.5X$ (a=1.5, b = 0.5)



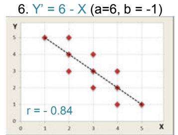
predikcije: približne
greške: još veće
korelacija: umerena

5. $Y' = 3$ (a=3, b=0)



predikcije: slabe
(ali bolje od nikakvih!)
korelacija: nulta

6. $Y' = 6 - X$ (a=6, b = -1)

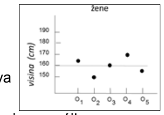


predikcije kao u slučaju 2.
nagib negativan

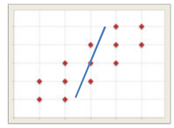
1. Bivarijatni KRN

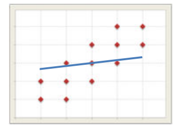
26

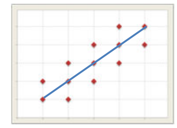
izračunavanje regresionih parametara a i b



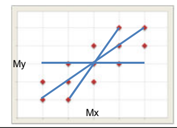
- kako provući *najbolju* pravu kroz skup markera?
  - kako naći 'centralnu pravu' skupa tačaka u ravni?
  - slično (lakše) pitanje: naći centr. tendenciju skupa brojeva
    - uočiti analogiju regresione prave i mera centr. tendencije
- rešenje pokušajima i pogreškama: tri primera od bezbroj mogućih:







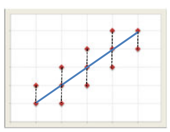
- sistematsko rešenje: zadati određene *uslove* koje treba da ispuni regresiona prava
- jedan uslov: regres. prava treba da uzme u obzir centralne tendencije obe varijable
  - dakle, treba da prolazi kroz tačku (Mx,My)
- OK, ali koji *nagib* treba da ima regresiona prava?



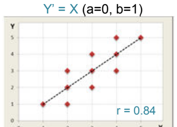
1. Bivarijatni KRN

27

- ključni matematički uslov: regresiona prava treba da bude ona prava od koje su odstupanja podataka *minimalna* (najmanja moguća)
  - odstupanja: rastojanja markera od regresione prave
- opšte rešenje dolazi iz više matematike (diferencijalni račun)
  - koristi se 'metoda najmanjih kvadrata'
    - traži se ona prava kod koje je *zbir kvadrata* svih vertikalnih rastojanja markera od prave najmanji
  - formule za a i b na osnovu te metode:
    - $b = \Sigma xy/\Sigma x^2$, $a = M_Y - bM_X$
  - ovo je opšte matematičko rešenje za intercept i nagib
    - nisu potrebni pokušaji i greške!



$Y' = X$ (a=0, b=1)



- *PRIMER:* podaci iz uvodnog primera
  - izračunato je: $\Sigma x^2 = 14$, $\Sigma xy = 14$, $M_X = 3$, $M_Y = 3$
  - sledi: $b = \Sigma xy/\Sigma x^2 = 14/14 = 1$; $a = M_Y - bM_X = 3-3 = 0$
  - prema tome: $Y' = a + bX$ ovde glasi: $Y' = X$
    - predviđanje: za 1 min učenja dobija se jedinica, za 2 min dobija se dvojka, itd
  - predviđanja nisu uvek tačna (r=0.84), ali su ovde najbolja moguća
  - procene su pogrešne samo za najviše ±1 ocenu