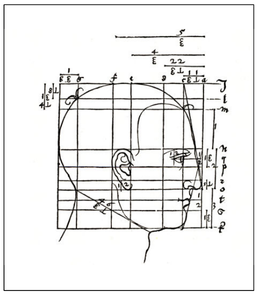
Metodologija psiholoških istraživanja

1

obrada koreg 2



11. decembar 2018.

C. Korelaciono-regresioni nacrti (KRN)

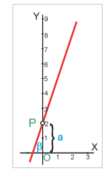
1. bivarijatni korelaciono-regresioni nacrti
2. multivarijatni korelaciono-regresioni nacrti

1. Bivarijatni KRN

2

interpretacija regresionih parametara a i b

- parametar a: intercept, tj. vrednost Y' kada je X nula
  - uočiti: ova vrednost nema realnog smisla ako X ne može biti nula
    - u primeru: za 0 min učenja dobila bi se ocena 0
- parametar b: veličina promene Y, kada se X menja za 1
  - ova vrednost ima uvek realni, a ne samo matematički smisao
    - u primeru: ako je b = 1, sledi da se za dodatni 1 min učenja prosečno dobija ocena za 1 (tj. b) viša
    - u primeru: ako je b = -1, sledi da se za dodatni 1 min učenja prosečno dobija ocena za 1 niža
  - ako je b=0, regresiona jednačina glasi $Y' = a + bX = a$
    - pritom važi da je: $a = M_Y – bM_X = My$
    - dakle, u proceni Y uopšte ne učestvuju vrednosti varijable X, tj. Y *ne zavisi* od X
    - za sve objekte, procena je da je vrednost varijable Y jednaka njenom proseku My

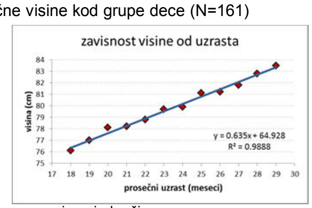


1. Bivarijatni KRN

3

- interpolacija i ekstrapolacija
- procene vrednosti Y za vrednosti X koje nisu ispitane
  - interpolacija: procena Y za vrednosti X *između* ispitanih vrednosti
    - npr. za 2.5 min učenja dobila bi se prosečna ocena 2.5
  - ekstrapolacija: procena Y za vrednosti X *izvan* ispitanih vrednosti
    - npr. za 15 min učenja dobila bi se prosečna ocena 15!
  - uočiti: ekstrapolacija je manje pouzdana i više riskantna
- *PRIMER:* odnos uzrasta i prosečne visine kod grupe dece (N=161)

uzrast (X)	prosečna visina (Y)	procenjena visina (Y')
18	76.1	76.4
19	77.0	77.0
20	78.1	77.6
21	78.2	78.3
22	78.8	78.9
23	79.7	79.5
24	79.9	80.2
25	81.1	80.8
26	81.2	81.4
27	81.8	82.1
28	82.8	82.7
29	83.5	83.3

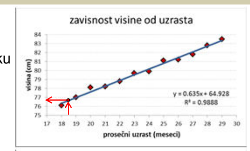


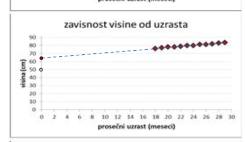
regresiona jednačina:
procenjena visina $Y' = 0.6*X + 64.9$

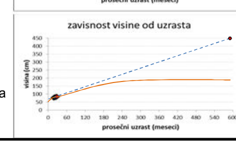
1. Bivarijatni KRN

4

- $Y' = 64.9 + 0.6*X$, tj.:
- $Y' = a + bX$, $a = 64.9$, $b = 0.6$, $r_{XY} = 0.99$
- svakog meseca deca rastu za 0.6 cm, u proseku
  - tj., kada se uzrast X uveća za 1 mesec visina Y se uveća za vrednost b=0.6 cm
- jednačina odlično procenjuje izmerene visine
- *primer procene:* za X=18, jednačina daje Y'=76.3cm, a prava vrednost je Y=76.1
- *primer interpolacije:*
  - za X=18.5, dobija se $Y'= 64.9+0.6*18.5=76.9$ cm
  - ta vrednost je prihvatljiva
- *primeri ekstrapolacije:*
  - za X = 0, dobija se Y' = 64.9 (intercept)
    - ta vrednost je precenjena!
    - naime, za novorođenčad važi X ≈ 50 cm
  - za X = 600 (50 godina), Y' = 446cm
    - ta vrednost je realno nemoguća!
    - problem: rast *nije* celim tokom života linearan, sa nagibom utvrđenim za period od 18-29 meseci
    - bebe rastu brže, odrasli prestaju da rastu



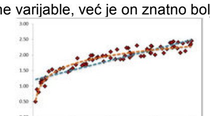


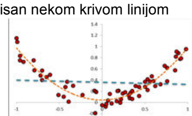


1. Bivarijatni KRN

5

- varijante linearne regresije
  - dosadašnja izlaganja: kako na osnovu mera X proceniti mere Y
    - odgovor: $Y' = a + bX$
  - novo pitanje: kako na osnovu devijacija x procenjivati devijacije y
    - odgovor: $y' = bx$
      - naime, a=0, jer je $a = M_Y – bM_X$, a za devijacije važi $M_y = M_x = 0$
  - novo pitanje: kako na osnovu stand. vred. z_x procenjivati stand. vred z_y
    - odgovor: $z_y' = \beta z_x$ (može se pokazati da je $\beta = r_{XY}$)
- nelinearna regresija
  - za bilo koji slučaj bivarijatne regresije, uvek se mogu izračunati vrednosti regresionih koeficijenata a i b, i postaviti regresiona prava
  - međutim, ponekad prava linija ne opisuje dobro odnos između nezavisne i zavisne varijable, već je on znatno bolje opisan nekom krivom linijom





1. Bivarijatni KRN

6

- iako može postojati jasna povezanost dve varijable, u takvim slučajevima koeficijent *linearne* korelacije r_{XY} biće blizak nuli!
  - naime, metoda linearne regresije je prilagođena situacijama kod kojih je zavisnost Y od X približno *linearna,* tj. adekvatno prikaziva pravom linijom
- u takvim slučajevima, odnos X i Y biće bolje prikazan nekom *nelinearnom* funkcijom, koristeći postupke nelinearne regresije
  - postupci su u principu slični kao u linearnoj regresiji, ali se koriste *nelinearne transformacije* varijabli, npr. *kvadratna* transformacija, *logaritamska* transfor. itd
- *PRIMER:* kvadratna regresija: $Y' = a + bX + cX^2$
  - koristi se kada nelinearna funkcija ima jedan pregib, kao kod parabole
  - cX^2 je *kvadratni* član jednačine (bX je *linearni* član)
  - parametar c odražava *zakrivljenost* funkcije
- *PRIMER:* kubna regresija: $Y' = a + bX + cX^2 + dX^3$
  - funkcija ima dva pregiba (npr. sigmoidalna fun.) i koristi se *kubna* transformacija
- opšti slučaj: polinomska regresija
  - polinom: zbir *stepenovanih* funkcija ($Y' = aX^0 + bX^1 + cX^2 + dX^3 + eX^4 + ...= \Sigma b_i x^i$)
- koriste se i druge transformacije, kao što je logaritmovanje (log X), itd.
  - logaritmovanjem varijabli (bilo X, bilo Y, bilo obe) može se ponekad transformisati nelinearan odnos dve varijable u linearan odnos

1. Bivarijatni KRN

c. Značajnost rezultata

- statistička značajnost odnosi se i na korelacioni i na regresioni aspekt nacrta
  - *korelacioni aspekt*: koeficijent korelacije može biti numerički različit od nule na *uzorku*, ali je pitanje da li korelacija postoji i u *populaciji*
  - *regresioni aspekt:* regresiona prava može u *uzorku* biti nagnuta, ali da u *populaciji* bude statistički horizontalna, tj. da Y ne zavisi od X
- analiza značajnosti kod bivarijatnih korelaciono-regresionih nacrta (BKRN) je slična kao kod bivalentnih jednofaktorskih nacrta (BJFN), uz neke razlike
- bitna razlika: tip *nezavisne* varijable u dva nacrta:
  - BJFN: nezavisna varijabla A je *kategorička*, i ima dva nivoa (kategorije), a1 i a2
  - BKRN: nezavisna varijabla X je *numerička,* nema kategorije
- *Faza I: Podaci i deskriptivne mere*
- *PRIMER:* vrlo sličan kao primer za BJFN, uz bitnu razliku u pogledu NV:
  - u ranijem primeru, kod BJFN: nezavisna varijabla A je kategorička
    - NV je pušenje, izraženo *kvalitativno* sa dva nivoa ('pušači' i 'nepušači')
  - u sadašnjem primeru, kod BKRN: nez. var. X je *numerička*
    - NV je pušenje, izraženo *kvantitativno*, brojem dnevno popušenih cigareta
  - zavisna varijabla Y: broj rešenih matematičkih zadataka (isto kao u BJFN)

1. Bivarijatni KRN

bivalentni jednofaktorski nacrt (BJFN)

Subj.	mere NV X (nivoi)	1. mere zav. var. Y		2. grupni proseci M		3. opšti prosek M_Y
		Y_1	Y_2	M_1	M_2	M_Y
O_1	a_1 (puš.)	4		3		5
O_2	a_1 (puš.)	2		3		5
O_3	a_2 (nep.)		8		7	5
O_4	a_2 (nep.)		6		7	5

bivarijatni kor-reg nacrt (BKRN)

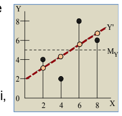
Subj.	mere nez. var. X	mere zav. var. Y	procene zav. var. Y'	prosek zav. var. M_Y
O_1	2	4	3.2	5
O_2	4	2	4.4	5
O_3	6	8	5.6	5
O_4	8	6	6.8	5

- mere NV: u BJFN kategoričke vrednosti, u BKRN: kvantitativne vrednosti
- mere ZV: u oba nacrta kvantitativne vrednosti (u primeru su iste u oba nacrta)
- u BJFN: opšti prosek My, tj prosek ZV za oba nivoa, u BKRN: prosek ZV (My)
- u BJFN: proseci nivoa (grupa), u BKRN: procenjene vrednosti ZV (Y')
  - računanje vrednosti b i a: $b = \Sigma xy/\Sigma x^2 = 12/20 = 0.6$, $a = M_Y - bM_X = 5\text{-}0.6{*}5 = 2$
  - regresiona jednačina glasi $Y' = a+bX$, a u ovom slučaju: $Y' = 2 + 0.6X$

#	X	Y	x	y	xy	x^2	y^2	z_x	z_y	$z_x z_y$
O1	2	4	-3	-1	3	9	1	-1.34	-0.45	0.6
O2	4	2	-1	-3	3	1	9	-0.45	-1.34	0.6
O3	6	8	1	3	3	1	9	0.45	1.34	0.6
O4	8	6	3	1	3	9	1	1.34	0.45	0.6
zbir	20	20	0	0	12	20	20	0	0	2.4
pros.	5	5	0	0	3	5	5	0	0	r=0.6

računanje $r_{XY} = 0.6$

- na osnovu dobijene jednačine postavlja se regresiona prava
- za svako X računa se Y':
- za $X=2$, $Y'=2+0.6{*}2=3.2$
  za $X=4$, $Y'=2+0.6{*}4=4.4$, itd
- te vrednosti su u gornjoj tabeli, i na desnom grafikonu



1. Bivarijatni KRN

- *Faza II: nulte hipoteze, očekivane vrednosti, devijacije*
- postoje *dve* nulte hipoteze, koje su *ekvivalentne*, tj. imaju isti test značajnosti
  - *korelaciona H0:* u populaciji nema korelacije, tj. populacioni $r_{XY} = 0$
  - *regresiona H0:* regresiona linija je horizontalna, tj. u populaciji $b = 0$
- devijacije: kao kod BJFN, postoje tri vrste (međusobne razlike Y, Y', i My)
  - *totalne devijacije:* razlika ind. mera (Y) i opšteg proseka (My): $y = Y - My$
    - koliko opservirane mere ZV, Y, odstupaju od proseka svih mera ZV, My
  - *regresione greške (reziduali):* razlika ind. mera (Y) i procena ZV (Y'): $g = Y - Y'$
    - koliko *opservirane* mere ZV, Y, odstupaju od *procenjenih* mera ZV, Y'
    - odgovaraju *unutargrupnim* devijacijama ($g = Y - M$) u BJFN
  - *regresioni efekti:* razlika procena ZV (Y'), i opšteg proseka (My): $e = Y' - My$
    - koliko procenjene mere ZV, Y', odstupaju od proseka svih mera ZV, My
    - odgovaraju *međugrupnim* devijacijama ($e = M - My$) u BJFN

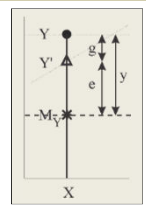
Subj.	mere nez. var. X	mere zav. var. Y	procene zav. var. Y'	prosek zav. var. M_Y
O_1	2	4	3.2	5
O_2	4	2	4.4	5
O_3	6	8	5.6	5
O_4	8	6	6.8	5

tot. odstup. $y = Y - M_Y$	reg. greške $g = Y - Y'$	reg. efekti $e = Y' - M_Y$
4-5 = -1	4 - 3.2 = 0.8	3.2 - 5 = -1.8
2-5 = -3	2 - 4.4 = -2.4	4.4 - 5 = -0.6
8-5 = 3	8 - 5.6 = 2.4	5.6 - 5 = 0.6
6-5 = 1	6 - 6.8 = -0.8	6.8 - 5= 1.8
-1+(- 3)+3+1 = 0	0.8+(-2.4)+2.4+(-0.8)=0	-1.8+(-0.6)+0.6+1.8=0

- zbirovi devijacija ($\Sigma y, \Sigma g, \Sigma e$) su nulti:

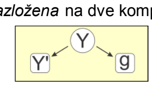
1. Bivarijatni KRN

- u BKRN važe iste jednačine kao u BJFN:
  - *devijaciona jednačina:* $y = e + g$
  - *strukturna jednačina*: $Y = My + e + g$
- ovi odnosi se mogu i grafički prikazati;
  - vrednosti NV X odgovara vrednost ZV Y
  - prikazane su i vrednosti Y' i My
  - devijacije: $y = Y\text{-}My$, $g = Y\text{-}Y'$, $e = Y'\text{-}My$
  - može se videti da je $y = e + g$ i $Y = My + e + g$
    - pretpostavljeno je da su sve devijacije pozitivne



- nova formula: $Y = Y' + g$
  - ovo sledi iz definicije $g = Y\text{-}Y'$
  - a slaže se i sa strukt. jedn., jer je $e = Y'\text{-}My$
- kaže se: postupkom regresije, ZV Y je *razložena* na dve komponente, Y' i g

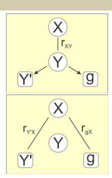
Sub.	mere zav. var. Y	procene zav. var. Y'	reg. greške $g = Y - Y'$
O_1	4 = 3.2 + 0.8	3.2	0.8
O_2	2 = 4.4 - 2.4	4.4	-2.4
O_3	8 = 5.6 + 2.4	5.6	2.4
O_4	6 = 6.8 – 0.8	6.8	-0.8



- kaže se takođe da su komponente Y' i g iz Y *ekstrahovane* ili *parcijalizovane ili izdvojene*

1. Bivarijatni KRN

- šematski prikaz odnosa X, Y, Y' i g:
  - između X i Y postoji korelacija, označena sa r_{XY}
  - razmotrićemo još neke korelacije između ovih varijabli
- kakve su korelacije varijabli Y' i g sa varijablom X?
  - označićemo ih sa $r_{Y'X}$ i r_{gX}



- Y' je procena zav. var. Y linearnom regresijom: $Y' = a + bX$
  - dakle, Y' je *linearna transformacija* varijable X
  - sledi: korelacija Y' i X je *maksimalna* moguća: $r_{Y'X} = 1$ ili $r_{Y'X} = -1$
  - interpretacija:
    - Y' predstavlja onaj *deo* Y, odn. onu komponentu ekstrahovanu iz Y, koja *maksimalno* korelira sa nezavisnom varijablom X
- g je greška, odstupanje procene Y' od opservirane vrednosti Y: $g = Y– Y'$
  - može se dokazati da je korelacija između g i X *minimalna* moguća: $r_{gX} = 0$
  - interpretacija:
    - g predstavlja onaj deo Y, odn. onu komponentu koja je u Y *preostala* nakon što je iz Y izdvojeno Y' (tj. onaj deo koji linearno korelira sa X)
    - kaže se i da se izračunavanjem varijable g korelacija Y sa X 'statistički kontroliše' ili 'uzima u obzir'

1. Bivarijatni KRN

- *Faza III: test statistik*
- važe iste jednačine kao i u BJFN, samo uz izvesne razlike u oznakama:
- jednačina zbirova kvadrata: $\Sigma y^2 = \Sigma e^2 + \Sigma g^2$ odn. $SS_{TOT} = SS_{reg} + SS_{rez}$
  - međugrupnom zbiru kvadrata SS_A odgovara *regresioni zbir kvadrata* SS_{reg}
  - unutargrupnom zbiru kvadrata SS_e odgovara *rezidualni zbir kvadrata* SS_{rez}

tot. odstup. $y = Y - M_Y$	reg. greške $g = Y - Y'$	reg. efekti $e = Y' - M_Y$
$(-1)^2 = 1$	$(0.8)^2 = 0.64$	$(-1.8)^2 = 3.24$
$(-3)^2 = 9$	$(-2.4)^2 = 5.76$	$(-0.6)^2 = 0.36$
$(3)^2 = 9$	$(2.4)^2 = 5.76$	$(0.6)^2 = 0.36$
$(1)^2 = 1$	$(-0.8)^2 = 0.64$	$(1.8)^2 = 3.24$
$\Sigma y^2 = 1+9+9+1 = 20$	$\Sigma g^2 = 0.64+5.76+5.76+0.64 = 12.80$	$\Sigma e^2 = 3.24+0.36+0.36+3.24 = 7.20$

  - uočiti: $SS_{TOT} = 20$; $SS_{reg} = 7.2$; $SS_{rez} = 12.8$
  - dakle, važi: $SS_{TOT} = SS_{reg} + SS_{rez}$, budući da je: $20 = 7.2 + 12.8$
- kao i kod BJFN, jednačina zbirova kvadrata se transformiše na dva načina
- (1) prva transformacija: deljenje jednačine sa SS_{TOT}
  - važi: $SS_T / SS_T = SS_{reg} / SS_T + SS_{rez} / SS_T$
  - slično kao u BJFN, važe oznake $SS_{reg} / SS_T = r^2$, i $SS_{rez} / SS_T = q^2$
  - u datom primeru: $20/20 = 7.2/20 + 12.8/20$ odn., $1 = 0.36 + 0.64$

1. Bivarijatni KRN

13

- dakle, kao i u BJFN važi *proporciona jednačina:* $r^2 + q^2 = 1$
- r^2: koeficijent determinacije
  - u primeru: $r^2 = 0.36$ odn. 36%
  - r^2 kvantitativno izražava *objašnjenu* varijabilnost zav. var. Y
    - takođe nazvana *zajednička* varijabilnost nez. var. X i zav. var. Y
    - *PRIMER:* u datom primeru, to je onaj deo variranja matematičke sposobnosti koji se može povezati sa pušenjem, a iznosi 36%
  - r^2 je jednak *kvadratu* koeficijenta korelacije r varijabli X i Y
  - r^2 je uvek pozitivan broj (za razliku od r) ili nula (kad je $r = 0$)
  - r^2 je uvek *manji* broj od r, izuzev za $r = \pm 1$ i za $r = 0$
    - npr., kada je $r = 0.60$, kao u datom primeru, tada je $r^2 = (0.60)^2 = 0.36 < 0.60$
- q^2: koeficijent nedeterminacije (u datom primeru: 0.64 odn. 64%)
  - q^2 kvantitativno izražava *neobjašnjenu* varijabilnost zav. var. Y
    - takođe nazvana *specifična* varijabilnost zav. var Y
    - *PRIMER:* u datom primeru, to je onaj deo variranja matematičke sposobnosti koji se ne može povezati sa pušenjem i ostaje neobjašnjen, *a* iznosi 64%
    - neobjašnjenost se odnosi na dato istraživanje, i ne znači da je ta varijabilnost u principu neobjašnjiva!

1. Bivarijatni KRN

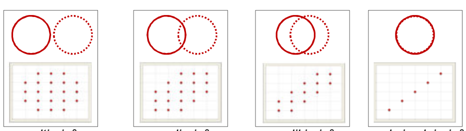
14

- (2) druga transformacija jednačine zbirova kvad.: kao i u BJFN, deljenjem *zbirova kvadrata (SS)* sa *step. slobode (df)* dobijaju se *prosečni kvadrati (MS)*
- setimo se: BJFN: $df_A = 1$; u nep. nacrtu: $df_e = 2(N-1)$; u pon. nacrtu: $df_e = N-1$
- u BKRN važi: $df_{reg} = 1$; $df_{rez} = N-2$
  - samo jedna grupa, a gube se 2 stepena slobode, jer ima *dve* numeričke varijable, X i Y
- jednačine za prosečne kvadrate imaju isti oblik kao kod BJFN:
  - $MS_{reg} = SS_{reg}/df_{reg}$, $MS_{rez} = SS_{rez}/df_{rez}$
- odgovarajući F-količnik uzima oblik: $F = \dfrac{MS_{reg}}{MS_{rez}} = \dfrac{\Sigma e^2/1}{\Sigma g^2/(N-2)}$
  - formula je slična kao kod BJFN
    - formalne razlike se odnose na stepene slobode
    - postoje i ekvivalentne alternativne formule
  - iz formule se može videti da će F biti utoliko veće ukoliko je:
    - manje Σg^2, odn. ukoliko procenjene vrednosti Y' manje odstupaju od opserviranih vrednosti zavisne varijable Y, tj. što je predikcija bolja
    - veće N, odn. ukoliko je veća veličina uzorka
    - veći Σe^2, odn. ukoliko su veća odstupanja procenjenih vrednosti Y' od My, odn. ukoliko regresiona prava više odstupa od horizontale
  - *PRIMER:* za date podatke: $MS_{reg} = 7.2$; $MS_{rez} = 12.8/(2(4-2)) = 6.4$; $F(1,2) = 7.2/6.4 = 1.13$

1. Bivarijatni KRN

15

- grafički prikazi varijabilnosti i korelacije varijabli X i Y
  - varijabilnosti X i Y se prikazuju pomoću dva kruga
  - veličina *zajedničke* varijabilnosti se prikazuje *preklopljenim* delom dva kruga



nulti r i r² | *mali r i r²* | *veliki r i r²* | *maksimalni r i r²*

- postoji i 'prilagođeni' koeficijent determinacije r^2 (engl.: 'adjusted r-square')
  - setimo se: može se pokazati da neke mere dobijene na *uzorku* (npr. SD i V) nisu najbolje moguće procene odgovarajućih mera u *populaciji*
  - slično tome, može se pokazati da koeficijent determinacije izračunat na uzorku *precenjuje* odgovarajući koeficijent u populaciji
  - stoga se, prema određenoj formuli (koju ovde ne dajemo), računa *prilagođeni* r^2
    - to je realističnija (i obično niža) mera objašnjene varijanse u populaciji
    - u primeru, izračunati r^2 iznosi 0.36, dok prilagođeni r^2 iznosi 0.04, odn. 4%

1. Bivarijatni KRN

16

- *Faza IV: p-vrednost*
  - kompjuter računa p-vrednost prema odgovarajućim formulama
  - *PRIMER:* u datom primeru, za $F(1,2) = 7.2/6.4 = 1.13$, dobija se $p = 0.4$
- *Faza V: značajnost:* isto kao u BJFN
  - rezultat je statistički značajan ako je $p < 0.05$
  - *PRIMER:* u datom primeru, p je znatno veće od 0.05
  - statistički formulisan zaključak: H0 se ne može odbaciti
  - istraživački formulisan zaključak: nema osnova da se tvrdi da matematička sposobnost zavisi od pušenja
- *Uslovi korišćenja F-testa*
  - kao i kod BJFN: traži se slučajnost i nezavisnost, kao i jedna varijanta normalnosti
    - naime: traži se tzv. *bivarijatna* normalna raspodela
    - tj.: X je normalno distribuiran za svaku vrednost Y, i obratno
- Alternative F-testu
  - kao kod BJFN, može se koristiti i ekvivalentni t-test

2. Multivarijatni KRN

17

2. Multivarijatni korelaciono-regresioni nacrti (MKRN)

- zovu se i *multipli* ili *višestruki* (korelaciono-)regresioni nacrti
- sadrže više od dve varijable
- varijable su tipično numeričke (ali mogu biti kategoričke)

a. Organizacija podataka

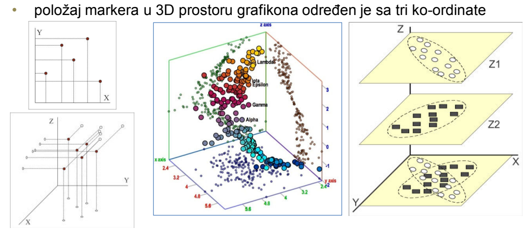
- koristi se matrični i grafički prikaz
- matrica podataka: tipa objekti x varijable
  - kolone: za svaku varijablu po jedna
  - redovi: za svaki objekt istraživanja po jedan
  - *PRIMER:* istraživanje osobina studenata

#	Visina	Težina	Pol	Pušenje	Školski uspeh	Prijemni ispit
1.	170	80	1	1	4.4	25
2.	195	95	1	0	3.9	17
3.	150	55	2	0	5.0	29
4.	165	65	2	1	4.1	23
5.	...	...	...	...	...	...

2. Multivarijatni KRN

18

- grafikoni podataka
  - obično se prikazuju 2D korelacioni dijagrami *parova* varijabli
    - npr. visina i težina, školski uspeh i uspeh na prijemnom ispitu, itd.
    - setimo se: položaj markera u ravni grafikona je određen sa dve ko-ordinate
  - retko se prikazuju 3D korelacioni dijagrami odnosa *tri* varijable, X, Y i Z
    - položaj markera u 3D prostoru grafikona određen je sa tri ko-ordinate



uočiti: projekcije skupa markera na tri ko-ordinatne ravni čine tri 2D korelaciona dijagrama *parova* varijabli (X i Y, Y i Z, Z i Y)

slučaj kad je Z dihotomija, tako da su markeri samo na dva nivoa, z1 i z2

2. Multivarijatni KRN

b. Deskriptivne statističke mere, prikaz, struktura

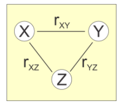
(1) korelacioni aspekt multivarijatnih KRN

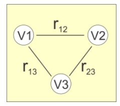
- osnovna deskriptivna mera je koeficijent korelacije (KK)
  - kako ima više od dve varijable, ima i više od dva KK
  - npr. za tri varijable, X, Y, Z (odn. V1, V2, V3), ima tri KK: r_{XY}, r_{XZ}, r_{YZ}, odn. r_{12}, r_{13}, r_{23}
  - npr.: $r_{12} = 0.25$, $r_{13} = 0.35$, $r_{23} = 0.45$
- prikaz KK se obično vrši pomoću *matrice korelacija*
- *PRIMER:* matrica korelacija tri varijable

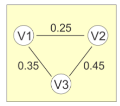
	V_1	V_2	V_3
V_1	1	r_{12}	r_{13}
V_2	r_{21}	1	r_{23}
V_3	r_{31}	r_{32}	1

	V_1	V_2	V_3
V_1	1	0.25	0.35
V_2	0.25	1	0.45
V_3	0.35	0.45	1

- matrica korelacija je *simetrična* (sastoji se od dva jednaka trougla), budući da je $r_{12} = r_{21}$, $r_{13} = r_{31}$ itd.
- na dijagonali su jedinice, jer je $r_{11} = 1$, $r_{22} = 1$ itd.





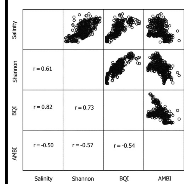


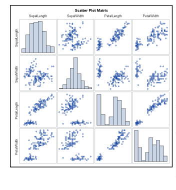
2. Multivarijatni KRN

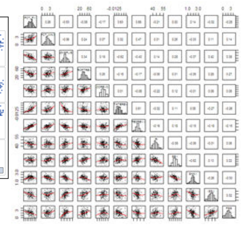
- dijagonala i jedan trougao ne moraju se prikazati, budući da ne donose novu informaciju

	V_1	V_2	V_3
V_1			
V_2	0.25		
V_3	0.35	0.45	

- u drugom trouglu se mogu prikazati odgovarajući korelacioni dijagrami
- na dijagonali se mogu prikazati distribucije varijabli
- moguće je na istom grafikonu prikazati sve tri informacije: i KK, i korelacione dijagrame, i distribucije



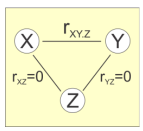




2. Multivarijatni KRN

korelacije višeg reda

- u multivarijatnim KRN se osim većeg *broja* korelacija među varijablama (r_{12}, r_{13}, ...) pojavljuju i nove *vrste* korelacija
- to su: *parcijalna korelacija*, *semiparcijalna korelacija*, i *multipla korelacija*
- parcijalna korelacija
  - to je korelacija varijabli X i Y, pri kojoj se treća varijabla, Z, 'kontroliše', 'uzima u obzir', ili 'drži konstantnom'
  - naime, to je korelacija X i Y koja *bi* među njima postojala, kada *ne bi* postojale:
    - ni korelacija r_{XZ} među varijablama X i Z
    - ni korelacija r_{YZ} među varijablama Y i Z
  - oznaka: $r_{XY.Z}$ (može se čitati kao 'XY bez Z') odn. $r_{12.3}$
  - *PRIMER:* parcijalna korelacija varijabli *broj cipele* (X) i *matematička sposobnost* (Y)
    - to je korelacija među njima kada ni broj cipele ni matematička sposobnost ne bi korelirali sa uzrastom (Z), tj. kada bi bilo $r_{XZ} = 0$ i $r_{YZ} = 0$
    - takva korelacija $r_{XY.Z}$ bi bila nulta, ili veoma mala

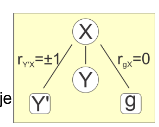


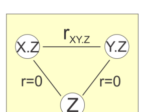
2. Multivarijatni KRN

- sve ranije opisane korelacije (r_{XY}, r_{XZ}, r_{YZ} itd.) se nazivaju *obične korelacije* ili *korelacije nultog reda*
- parcijalna korelacija spada u *korelacije prvog reda*
  - *prvog* reda: zato što se statistički kontroliše *jedna* dodatna varijabla (Z)
  - postoje i odgovarajuće korelacije drugog, trećeg itd. reda (njima se ne bavimo)
- kako se utvrđuje koeficijent parcijalne korelacije?
  - postoje dve vrste rešenja, metodološko i statističko
- *metodološko rešenje:* pogodnim nacrtom istraživanja učiniti da korelacije r_{XZ} i r_{YZ} budu nulte
  - računati KK između X i Y kod skupa objekata istraživanja kod kojih je vrednost varijable Z *konstantna,* tako da ona ne može da korelira ni sa kojom varijablom
  - na pr., korelacije broja cipele (X) i matematičke sposobnosti (Y), pri čemu su svi ispitanici istog uzrasta (Z)
  - problem: ovakav postupak može da bude nepogodan ili nemoguć, a važi samo za datu konstantnu vrednost Z (npr. samo za dati uzrast)
- *statističko rešenje:* izračunavanje $r_{XY.Z}$ pomoću određene formule
  - u toj formuli se pojavljuju sve tri korelacije nultog reda, tj. r_{XY}, r_{XZ}, r_{YZ}
  - objasnićemo princip računanja

2. Multivarijatni KRN

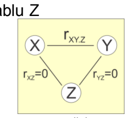
- setimo se analize u BKRN: $Y = Y' + g$
  - pritom važi $r_{gX} = 0$, tj. varijabla $g = Y\text{-}Y'$ je ona komponenta varijable Y koja *ne korelira* sa var. X
    - stoga se g označava i kao Y.X ('Y bez X')
  - dakle: na osnovu X, koja korelira sa Y, konstruisana je iz Y nova varijabla, Y.X = Y-Y', koja ne korelira sa X
- na analogan način se postupa u MKRN
  - na osnovu Z, koja korelira i sa X i Y, konstruišu se, iz X i iz Y, *dve* nove varijable koje ne koreliraju sa Z
    - varijabla 'X bez Z': X.Z = X – X'
    - varijabla 'Y bez Z': Y.Z = Y – Y'
- korelacija varijabli X.Z i Y.Z je tražena parcijalna korelacija $r_{XY.Z}$ varijabli X i Y
  - naime: $r_{XY.Z}$ je korelacija X i Y kada one ne bi korelirale sa Z
- formula glasi: $r_{xy.z} = \dfrac{r_{xy} - \left(r_{xz} \times r_{yz}\right)}{\sqrt{\left(1-r_{xz}^2\right)\left(1-r_{yz}^2\right)}}$
  - uočimo: u ovoj formuli se pojavljuje ne samo r_{XY} već i ostale korelacije nultog reda, tj. r_{XZ}, r_{YZ}



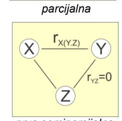


2. Multivarijatni KRN

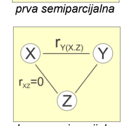
- pored parcijalne korelacije, postoji još jedan tip korelacija prvog reda:
- semiparcijalna korelacija varijabli X i Y s obzirom na varijablu Z
  - ponekad se naziva i part korelacija
- *semi*parcijalna korelacija je korelacija među X i Y kada bi samo *jedna* od njihovih korelacija sa Z bila nula
  - tj., kada važi ili $r_{YZ} = 0$ ili $r_{XZ} = 0$
    - za razliku od parcijalne korelacije $r_{XY.Z}$, kod koje su *obe* ove korelacije nulte, tj. važi i $r_{YZ} = 0$ i $r_{XZ} = 0$
- postoje dve varijante semiparcijalne korelacije:
  - 1. korelacija varijabli X i Y koja bi postojala kada ne bi bilo korelacije između Y i Z, tj. kada je $r_{YZ} = 0$
    - to je korelacija između X i Y.Z (tj. Y bez Z)
    - oznaka: $r_{X(Y.Z)}$
  - 2. korelacija varijabli X i Y koja bi postojala kada ne bi bilo korelacije između X i Z, , tj. kada je $r_{XZ} = 0$
    - to je korelacija između Y i X.Z (tj. X bez Z)
    - oznaka: $r_{Y(X.Z)}$
- kasnije ćemo ukazati na značaj i ulogu semiparcijalnih korelacija u regresionoj analizi



parcijalna



prva semiparcijalna



druga semiparcijalna

2. Multivarijatni KRN

25

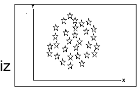
struktura korelacija

- međusobni korelacioni odnosi više varijabli mogu biti složeni
- u narednim primerima razmotrićemo odnose tri varijable, i to samo u slučaju kada su dve, X i Y, numeričke, a treća, Z, dihotomija
  - *PRIMER:* X: ocena iz matematike; Y: ocena iz srpskog; Z: pol
- počećemo od pitanja odnosa X i Y bez obzira na Z, odn. postojanja ili nepostojanja *marginalne korelacije* X i Y
  - *PRIMER:* da li postoji korelacija između ocene iz matematike (X) i ocene iz srpskog (Y), bez obzira na pol (Z)
- zatim ćemo uključiti Z kao *moderatorsku* varijablu, i razmotriti postojanje *parcijalnih korelacija* X i Y, u okviru pojedinih kategorija varijable Z
  - *PRIMER:* da li postoji korelacija između ocene iz matematike (X) i ocene iz srpskog (Y), ali posebno kod devojčica, i posebno kod dečaka
- koristićemo dve grupe primera
  - u prvoj grupi pretpostavićemo da *nema* marginalne korelacije između X i Y
  - u drugoj grupi pretpostavićemo da *ima* marginalne korelacije između X i Y

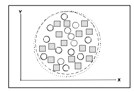
2. Multivarijatni KRN

26

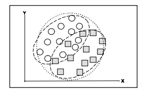
- prva grupa primera
  - pretpostavićemo da *nema* marginalne korelacije između X i Y (ili da je vrlo mala)
    - npr., kada se svi subjekti zajedno uzmu u obzir, ocene iz matematike (X) i ocene iz srpskog (Y) nisu korelirane



  - *zatim* ćemo uvesti pol kao treću, moderatorsku varijablu
    - razmotrićemo *četiri* mogućnosti za marginalne korelacije ocena, posebno kod muškaraca (M) i posebno kod žena (Ž)
    - muškarci će biti označeni kvadratićima, žene kružićima

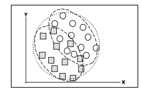


1. *nema* korelacije između X i Y, ni kod M ni kod Ž



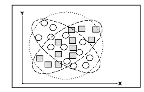
2. *pozitivna* korelacija između X i Y i kod M i kod Ž

uočiti: M bolji na X, Ž bolje na Y



3. *negativna* korelacija između X i Y i kod M i kod Ž

uočiti: Ž bolje od M, i na X i na Y

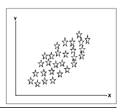


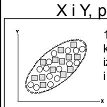
4. *pozitivna* korelacija između X i Y kod M, *negativna* korelacija između X i Y kod Ž

2. Multivarijatni KRN

27

- druga grupa primera
  - pretpostavićemo da *postoji* marg. korelacija među X i Y
    - npr. korelacija ocena iz matematike i srpskog je pozitivna
  - razmotrićemo *šest* mogućnosti za parcijalne korelacije X i Y, posebno kod muškaraca i posebno kod žena



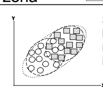


1. pozitivna korelacija između X i Y, i kod M i kod Ž

korelacije su *iste* kao marginalna

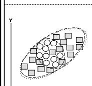
2. pozitivna korelacija između X i Y, i kod M i kod Ž

korelacije su *veće* nego marginalna



3. pozitivna korelacija između X i Y, i kod M i kod Ž

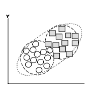
korelacije su *manje* nego marginalna



4. pozitivna korelacija između X i Y kod M, ali *nulta* korelacija između X i Y kod Ž

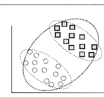
iako je marginalna korelacija pozitivna

slično specifikaciji



5. nulta korelacija između X i Y, i kod M i kod Ž

iako je marginalna korelacija pozitivna!



6. negativna korelacija između X i Y, i kod M i kod Ž

iako je marginalna korelacija pozitivna!!

- opšti zaključak: struktura parcijalnih korelacija može biti veoma raznolika, bez obzira da li marginalna korelacija postoji ili ne postoji