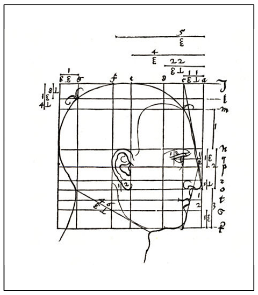
Metodologija psiholoških istraživanja

obrada koreg 3



17. decembar 2018.

C. Korelaciono-regresioni nacrti (KRN)

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

D. Završne napomene

V. Izveštaj o istraživanju

2. Multivarijatni KRN

(2) regresioni aspekt multivarijatnih KRN

- osnovni zadatak: za date X1, X2, X3 itd., proceniti odn. predvideti Y
  - X1, X2, X3, ... : prediktorske odn. nezavisne varijable
  - Y: kriterijumska (zavisna) varijabla
    - napomena: razmotrićemo detaljnije samo slučaj sa *dve* predikt. var., X1 i X2
    - on liči na dvofaktorski nacrt sa dva faktora (X1 i X2) i zav. varijablom Y
- *PRIMERI:*
  - procena uspeha na studijama (Y) na osnovu uspeha na prijemnom ispitu (X1) i školskog uspeha (X2)
  - procena inteligencije dece (Y) u zavisnosti od inteligencije očeva (X1) i majci (X2)
  - procena ocene na testu (Y) na osnovu vremena učenja (X1) i inteligencije (X2)
- statistički postupak kojim se vrši predikcija naziva se multivarijatna regresiona analiza ili multipla (višestruka) regresija
- multivarijatna regresija se zasniva na sličnim principima kao ranije opisana bivarijatna regresija, ali sadrži i neke nove momente
- u narednim izlaganjima ćemo se stalno:
  - prvo *podsećati* na određene aspekte bivarijatne regresije (BKRN) ...
  - ... da bi zatim izložili odgovarajuće *nove* multivarijatne sadržaje (MKRN)

2. Multivarijatni KRN

- *algebarski oblici regresije u BKRN i MKRN*
- *setimo se:* u BKRN se na osnovu X (predikt. var.) procenjuje Y (kriterij. var)
  - *opšti* oblik procene Y na osnovu X zapisuje se kao: Y' = f(X)
    - f označava bilo koju funkciju od X, linearnu ili nelinearnu
  - primeri *konkretnih* oblika procene Y:
    - bivarijatna *linearna* regresija (BLR): Y' = a + b*X
    - bivarijatna *nelinearna* regresija: $Y' = a + b*X + c*X^2$ (*kvadratna* regresija)
    - postoje i brojni drugi matematički oblici procene, ali je BLR najvažnija
- *novo:* oblici procene u MKRN (u slučaju *dva* prediktora, X1 i X2)
- postoje dva aspekta:
- prvi aspekt: Y se može proceniti na *dva* načina, predikcijom pomoću *dve pojedinačne bivarijatne* regresije, jedne po X1, i druge po X2
  - to su: procena Y na osnovu X1: Y'1 = f(X1); procena Y na osnovu X2: Y'2 = f(X2)
  - primeri *konkretnih* oblika procene Y:
    - dve bivarijatne *linearne* regresije: jedna po X1, druga po X2
      - Y'1 = a1 + b1*X1; Y'2 = a2 + b2*X2
    - moguće je koristiti i bivarijatne *nelinearne* regresije (ali se time ne bavimo)

2. Multivarijatni KRN

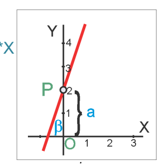
- drugi aspekt: Y se može proceniti pomoću *jedne multiple* regresije, predikcijom na osnovu *oba* prediktora X1 i X2 *zajedno*
  - *opšti* oblik procene Y: Y' = f(X1, X2)
  - *primer konkretnog* oblika procene Y:
    - multipla *linearna* regresija: Y' = a + b1*X1 + b2*X2
      - kaže se da je Y' *linearna kombinacija* X1 i X2, sa dodatom konstantom a
      - regresija je *linearna* jer su prediktori podignuti na prvi stepen ($X1^1$, $X2^1$)
    - postoji i multipla *nelinearna* regresija: pomenućemo je kasnije
- *članovi regresionih jednačina u BKRN i MKRN*
- *setimo se:* u BKRN, linearna regresiona jednačina Y' = a + b*X sadrži četiri *člana*, tj. dve varijable (X, Y') i dva parametra (a, b)
  - varijable: X: nez. var. odn. prediktor; Y': procena zav. var. odn. kriterijuma
  - parametri: a: intercept (Y' kada je X=0); b: nagib odn. regres. koeficijent
- *novo:* u MKRN, linearna regresiona jednačina Y' = a + b1*X1 + b2*X2 sadrži šest članova, tri varijable (X1, X2, Y) i tri parametra (a, b1, b2)
  - X1, X2: prediktori; Y': procena zavisne odn. kriterijumske varijable
  - a: intercept (označava se i sa b0); b1, b2: nagibi odn. regres. koeficijenti

2. Multivarijatni KRN

- *PRIMER:* procena inteligencije dece (Y) na osnovu inteligencije očeva (X1) i inteligencije majci (X2), preko jednačine Y' = a + b1*X1 + b2*X2
- prikazaćemo dva primera mogućih veličina regresionih koeficijenata
- pretpostavićemo da:
  - (1) inteligencija dece zavisi u jednakoj meri od inteligencije oca i inteligencije majke, odn. da je jednaka *proseku* inteligencije roditelja
    - Y' = (X1 +X2)/2 = 0.5*X1 + 0.5*X2
    - dakle, parametri regresione jednačine su: b1 = 0.5, b2 = 0.5, a = 0
    - na pr., za X1 = 120 i X2 = 100 sledi da je Y' = 0.5*100 + 0.5*120 = 110
  - (2) inteligencija dece zavisi tri puta više od jednog roditelja nego od drugog, a pritom su deca inteligentnija od roditelja za još 5 poena
    - Y' = 0.25*X1 + 0.75*X2 + 5
    - dakle, parametri su: b1 = 0.25, b2 = 0.75, a = 5
    - na pr., za X1 = 120 i X2 = 100 sledi da je Y' = 0.25*120 + 0.75*100 + 5 = 110
- važna napomena:
  - u primerima su veličine parametara odabrane na osnovu određenih hipoteza
  - u istraživanjima se veličine parametara izračunavaju automatski određenim formulama, na osnovu datih vrednosti varijabli (slično kao *a* i *b* kod BKRN)

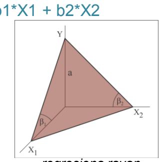
2. Multivarijatni KRN

- *geometrijski oblik regresije u BKRN i MKRN*
- *setimo se:* u bivarijatnoj linearnoj regresiji, sa X i Y:
  - geometrijski odraz algebarske jednačine Y' = a + b*X je *regresiona prava* u ravni (2D)
  - geometrijsko značenje članova jednačine:
    - X: horizontalna ko-ordinatna osa
    - Y: vertikalna ko-ordinatna osa
    - a: presek regresione prave sa Y-osom (intercept)
    - b: tangens ugla regresione prave sa X-osom



regresiona prava

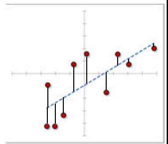
- *novo:* u multivarijatnoj linearnoj regresiji, sa X1, X2 i Y:
  - geometrijski odraz algebarske jednačine: Y' = a + b1*X1 + b2*X2 je *regresiona ravan* u 3D prostoru
  - geometrijsko značenje članova jednačine:
    - X1, X2: horizontalne ko-ordinatne ose
    - Y: vertikalna ko-ordinatna osa
    - a (ili b0): presek regresione ravni sa Y osom
    - b1, b2: tangensi uglova β1, β2 preseka regresione ravni sa X1-osom i X2-osom



regresiona ravan

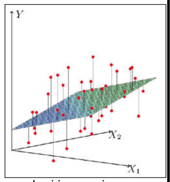
2. Multivarijatni KRN
7

- *aproksimacija linearnim jednačinama u BKRN i MKRN*
- *setimo se:* u bivarijatnoj linearnoj regresiji:
  - *idealno:* svi markeri se nalaze tačno na regres. pravoj
  - *realno:* neki markeri su iznad, a neki ispod regresione prave (predikcija nije egzaktna već približna)
  - *često:* podaci se grupišu u oblik *elipse*, koju *aproksimira* regresiona prava $Y' = a + b*X$
  - *regresioni parametri a i b:* određuje se metodom najmanjih kvadrata: $b = \Sigma xy/\Sigma x^2$, $a = M_Y - bM_X$



markeri i regresiona prava

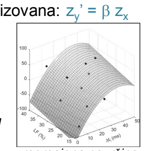
- novo: u multivarijatnoj linearnoj regresiji:
  - *idealno:* svi markeri se nalaze na regresionoj ravni
  - *realno:* neki markeri su iznad, a neki ispod regresione ravni (predikcija nije egzaktna već približna)
  - *često:* podaci se grupišu u oblik *elipsoida*, koji *aproksimira* regresiona ravan $Y' = b0 + b1*X1 + b2*X2$
  - *regresioni parametri b0, b1, b2:* određuju se metodom najmanjih kvadrata
    - nećemo razmatrati njihove formule



markeri i regresiona ravan

2. Multivarijatni KRN
8

- *varijante linearne regresije u BKRN i MKRN*
- *setimo se:* postoje tri varijante bivarijatne linearne regresije
  - osnovna: $Y' = a + b*X$; devijaciona: $y' = bx$; standardizovana: $z_y' = \beta z_x$
- *novo:* u multiploj linearnoj regresiji:
  - osnovna verzija: $Y' = b0 + b1*X1 + b2*X2$
  - devijaciona verzija: $y' = b1*x1 + b2*x2$
  - standardizovana verzija: $z_y' = \beta 1*z_{x1} + \beta 2*z_{x2}$



regresiona površina

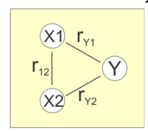
- *varijante nelinearne regresije u BKRN i MKRN*
- *setimo se:* postoji bivarijatna *nelinearna* regresija
  - npr. kvadratna regresija $Y' = a + b*X + c*X^2$ ima nelinearni kvadratni član $c*X^2$
- *novo:* postoji i *multipla nelinearna regresija*
  - naime, ponekad se dobijeni podaci ne mogu prikladno aproksimirati preko *ravni*
  - geometrijsko rešenje: aproksimacija nekom *zakrivljenom* površinom
  - algebarsko rešenje: uvođenje *nelinearnih članova* u regresionu jednačinu
  - primer: $Y' = b0 + b1*X1 + b2*X2 + b3*X1^2 + b4*X2^2 + b5*X1*X2$
    - $b3*X1^2$ i $b4*X2^2$ su *kvadratni* članovi
    - $b5*X1*X2$ je *mešoviti* član, koji odgovara *interakciji* varijabli X1 i X2

2. Multivarijatni KRN
9

- multipla regresija sa više od dve nezavisne varijable
  - često se u istraživanjima koristi veći broj prediktora
  - primer: multipla linearna regresija sa tri prediktora: X1, X2, X3
    - algebarski izraz: regresiona jednačina $Y' = b0 + b1*X1 + b2*X2 + b3*X3$
    - geometrijski odraz: ovo je jednačina tzv. *hiper-ravni u 4D prostoru!*
  - postoje i složenije varijante (još viši prostori i nelinearni članovi)
- mere korelacije *u BKRN i MKRN*
- *setimo se:* u bivarijatnoj linearnoj regresiji postoje dve mere:
  - koeficijent korelacije r_{XY}: izražava jačinu veze varijabli X i Y
  - koeficijent determinacije r_{XY}^2: izražava proporciju *objašnjene* varijabilnosti varijable Y varijablom X
    - ta veličina jednaka je kvadratu koeficijenta korelacije r_{XY}
- *novo:* u multiploj linearnoj regresiji postoji više odgovarajućih mera:
  - dva koeficijenta korelacije: $r_{Y,X1}$ i $r_{Y,X2}$ odn., kraće: r_{Y1} i r_{Y2}
  - dva odgovarajuća koeficijenta determinacije: r_{Y1}^2 i r_{Y2}^2
- ove veličine su *bivarijatni* koeficijenti
  - oni odražavaju relacije svakog od dva prediktora X1 i X2 za zavisnom varijablom Y *ponaosob*, bez obzira na postojanje drugog prediktora

2. Multivarijatni KRN
10

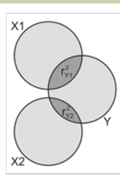
- pored r_{Y1} i r_{Y2}, u multiploj regresiji postoji još jedan novi bivarijatni koeficijent:
  - koeficijent korelacije dve *nezavisne* varijable, X1 i X2: r_{X1X2} odn. r_{12}



- pored *bivarijatnih*, u multiploj regresiji takođe postoje i *multipli* koeficijenti:
  - *koeficijent multiple determinacije*, označen sa R^2
    - izražava proporciju varijabilnosti varijable Y koja je objašnjena prediktorima X1 i X2 *zajedno*
    - za razliku od r_{Y1}^2 i r_{Y2}^2 koji izražavaju proporcije objašnjene varijablama X1 i X2 *ponaosob*, tj. svake za sebe
    - detaljno ćemo razmotriti odnos R^2, r_{Y1}^2 i r_{Y2}^2
  - *koeficijent multiple korelacije*, označen sa R
    - izražava stepen koreliranosti varijable Y sa oba prediktora *zajedno*
    - ova veličina jednaka je korenu koeficijenta multiple determinacije

2. Multivarijatni KRN
11

- postojanje odn. nepostojanje r_{12} veoma je važno za analizu rezultata MKRN
- postoje *dve* mogućnosti: (a) $r_{12} = 0$, (b) $r_{12} \neq 0$
  - pođimo od prvog, jednostavnijeg slučaja:
- (a) ne postoji korelacija varijabli X1 i X2: $r_{12} = 0$
  - ovaj slučaj se može prikazati šemom sa krugovima
  - krugovi predstavljaju *varijabilnost* varijabli X1, X2 i Y
    - odsustvo preseka X1 i X2 odražava činjenicu da je $r_{12} = 0$
    - prisustvo preseka X1 i Y, i X2 i Y, odražava postojanje korelacija r_{Y1} i r_{Y2}



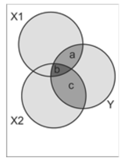
  - preseci krugova X1 i Y, i X2 i Y označeni su sa r^2_{Y1} i r^2_{Y2}
  - to su bivarijatni koeficijenti determinacije, koji odražavaju objašnjenu odn. zajedničku varijabilnost ovih parova varijabli
    - r^2_{Y1} izražava proporciju varijabilnosti varijable Y objašnjenu prediktorom X1
    - r^2_{Y2} izražava proporciju varijabilnosti varijable Y objašnjenu prediktorom X2
    - ostatak površine kruga izražava neobjašnjenu varijabilnost varijable Y
  - u ovom slučaju (tj. kada je $r_{12} = 0$) važi: $R^2 = r^2_{Y1} + r^2_{Y2}$
  - naime: variranje Y objašnjeno *zajedničkim* dejstvom dva prediktora, X1 i X2, tj. R^2, jednako je *zbiru* njihovih *pojedinačnih* doprinosa, $r^2_{Y1} + r^2_{Y2}$

2. Multivarijatni KRN
12

- *PRIMER*: setimo se istraživanja sa jednom NV i jednom ZV
  - X: vreme učenja (1 min do 5 min), Y: ocena (1 do 5)
  - za ranije date podatke, $r_{XY} = 0.84$, tako da je: $r^2_{XY} = 0.70$
    - tj. 70% varijabilnosti ocena može se objasniti dužinom učenja
    - dakle, 30% varijabilnosti ostaje neobjašnjeno
    - jedna mogućnost je da ocene zavise i od inteligencije subjekata
  - istraživanje: dve NV (vreme učenja, inteligencija) i jedna ZV (ocena)
    - npr.: X1: vreme učenja (1-5 min), X2: nivoi inteligencije (IQ = 90, 100, 110)
    - odgovara dvofaktorskom nacrtu tipa 3x5, sa 15 situacija (grupa ispitanika)
    - ako je nacrt *balansiran* (jednak broj ispitanika po grupi), tada da je $r_{12} = 0$
  - mogući rezultat: $r^2_{Y1} = 0.7$; $r^2_{Y2} = 0.3$; $R^2 = r^2_{Y1} + r^2_{Y2} = 0.7 + 0.3 = 1$
    - u ovom slučaju, celokupne razlike u ocenama mogle bi se objasniti razlikama u vremenu učenja (70%) i inteligenciji (30%)
  - napomena: ovakvi ishodi, tj. da se u istraživanju objasni *celokupna* varijabilnost zavisne varijable, su nerealni
  - primer realističnijeg ishoda:
    - X1 objašnjava 30%, X2 objašnjava 20%, njihova interakcija objašnjava 10%, a 40% varijabilnosti ostaje neobjašnjeno nezavisnim varijablama

2. Multivarijatni KRN

- druga mogućnost u vezi korelacije r_{12}:
- (b) *postoji* korelacija varijabli X1 i X2: $r_{12} \neq 0$
  - ovo je češći i tipičniji, ali složeniji slučaj
  - i ovaj slučaj se može prikazati odgovarajućom šemom
    - preklapanje X1 i X2 odražava činjenicu da je $r_{12} \neq 0$
  - pritom je:
  - unija područja a i b: r^2_{Y1} (zajednička varijabilnost X1 i Y)
  - unija područja b i c: r^2_{Y2} (zajednička varijabilnost X2 i Y)
  - unija područja a, b i c: R^2 (zajednička varijabilnost X1, X2 i Y)

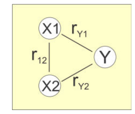


- uočimo: ako je $r_{12} \neq 0$, onda obično važi: $R^2 < r^2_{Y1} + r^2_{Y2}$
  - naime: R^2 je *veći* od r^2_{Y1} i r^2_{Y2} pojedinačno, ali je *manji* od *zbira* $r^2_{Y1} + r^2_{Y2}$
    - razlog je to što se područja r^2_{Y1} i r^2_{Y2} *preklapaju* (imaju zajednički deo b)
- *PRIMER*: zavisnost inteligencije dece od inteligencije roditelja
- pretpostavimo da postoji pozitivna korelacija inteligencija majki i očeva ($r_{12} \neq 0$)
- tada će varijabilnost inteligencije dece biti objašnjena na osnovu:
  - samo inteligencije *majki* (područje a), samo inteligencije *očeva* (područje c), ali i inteligencije *roditelja* (područ. b), koje se ne može *jednoznačno* pripisati ni majkama ni očevima, budući da postoji preklapanje
  - a jednim delom će ostati neobjašnjena u datom istraživanju (Y izvan a, b i c)

2. Multivarijatni KRN

uticaj uvođenja novih prediktora na moć predikcije

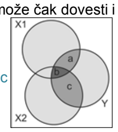
- pretpostavimo istraživanje po BKRN, sa X1 kao NV, i Y kao ZV
  - *PRIMER*: zavisnost uspeha na studijama (Y) od uspeha na prijemnom ispitu (X1)
  - dobijeni r^2_{Y1} izražava procent objašnjene varijabilnosti Y pomoću X1
- proširimo nacrt u MKRN, uvodeći X2 kao drugu NV, npr. školski uspeh
  - novi prediktor će takođe objasniti deo varijabilnosti kriterijumske varijable Y
- ključno pitanje: koliko će X2 *doprineti* objašnjenju Y, povrh i preko dela koji je *već* objašnjen pomoću X1?
- odgovor: to će zavisiti od *dva* činioca:



  - (a) korelacija X2 sa Y:
    - što je jača korelacija X2 sa Y, doprinos X2 će biti *veći*
  - (b) korelacija X2 sa X1
    - što je jača korelacija X2 sa X1, doprinos X2 će biti *manji*
- *PRIMER*: pretpostavimo da je školski uspeh (X2) visoko koreliran sa uspehom na prijemnom ispitu (X1)
  - u tom slučaju nam informacija o školskom uspehu neće mnogo *dodatno* pomoći da predvidimo uspeh na studijama, povrh i preko dela koji je *već* objašnjen uspehom na prijemnom ispitu

2. Multivarijatni KRN

- pouka: za dobru predikciju ne pomaže nužno imati što *više* prediktora, već imati međusobno što *nezavisnije* (manje korelirane) prediktore
  - *PRIMER*: ako bi školski uspeh (X2) *savršeno* korelirao sa uspehom na prijemnom ispitu (X1), bilo bi *beskorisno* koristiti i X2 kao prediktor, pored X1
  - napomena: ako su prediktori međusobno visoko korelirani, to može čak dovesti i do računskih problema prilikom statističke obrade podataka
- formalni opis prethodnih razmatranja
  - doprinos X2 nije celo područje r^2_{Y2} (tj. b+c) već samo područje c
  - ta veličina je *koeficijent semiparcijalne determinacije*, $r^2_{Y(2.1)}$
    - ona je kvadrat semiparcijalne korelacije Y i X2: $r_{Y(2.1)}$
    - to je korelacija Y sa var. X2.X1 (komponentom X2 koja ne korelira sa X1)
  - analogno tome, postoji i $r^2_{Y(1.2)}$, što odgovara području a



- zaključak: važi: $R^2 = (a + b) + c = r^2_{Y1} + r^2_{Y(2.1)}$, kao i $R^2 = (b + c) + a = r^2_{Y2} + r^2_{Y(1.2)}$
- uočimo: razmatrali smo slučajeve: $R^2 = r^2_{Y1} + r^2_{Y2}$ i $R^2 < r^2_{Y1} + r^2_{Y2}$
- ima i *treća* mogućnost, koja se ne može prikazati krugovima: $R^2 > r^2_{Y1} + r^2_{Y2}$
- ovaj slučaj je složen za analizu, a može se pojaviti na više načina, npr.:
  - $r_{Y1} > 0$; $r_{Y2} > 0$; $r_{12} < 0$
  - $r_{Y2} = 0$; $r_{Y(2.1)} \neq 0$

2. Multivarijatni KRN

c. Značajnost rezultata

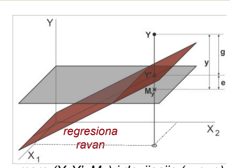
- *Faza I: Podaci i deskriptivne mere*
  - *PRIMER:* zavisnost uspeha na studijama (Y) od uspeha na testu na prijemnom ispitu (X1) i od školskog uspeha (X2)
  - nećemo navoditi konkretne podatke
  - utvrđuju se:
    - vrednosti varijabli X1, X2 i Y za svaki objekt istraživanja
    - prosečna vrednost My varijable Y
    - procenjene vrednosti Y' na osnovu jednačine $Y' = a + b1{*}X1 + b2{*}X2$
- *Faza II: nulte hipoteze, očekivane vrednosti, devijacije*
- postoje dve glavne vrste nultih hipoteza, *multiple* i *parcijalne*
- *multipla H0:* odnosi se na oba prediktora zajedno
  - X1 i X2 *zajedno* ne doprinose objašnjenju varijabilnosti varijable Y
  - R i R^2 su u populaciji nulti, regresiona ravan je horizontalna, na nivou My
- *PRIMER:* u populaciji, uspeh na prijemnom ispitu i školski uspeh, uzeti zajedno, ne koreliraju sa uspehom na studijama, niti objašnjavaju statistički značajan deo njegovog variranja

2. Multivarijatni KRN

- *parcijalne H0:* odnose se na pojedinačne prediktore
  - u ovim analizama bitan je *redosled* uvođenja varijabli X1 i X2 u analizu
  - kod faktorijalnih nacrta to nije bilo bitno, jer varijable nisu bile korelirane
- *prva parcijalna H0:* odnosi se na X2
  - H0 glasi: ako se var. X2 uvede u analizu posle var. X1, ona ne doprinosi objašnjenju varijabilnosti Y *preko* doprinosa X1
    - *PRIMER:* ako se u predikciju uspeha na studijama (Y) posle uspeha na prijemnom ispitu (X1) u analizu uključi i školski uspeh (X2), on ne doprinosi značajno povećanju objašnjenja variranja Y
  - to znači da će u populaciji biti nulte mere koje se odnose na X2: semiparcijalna korelacija $r_{Y(2.1)}$, semiparcijalna determinacija $r^2_{Y(2.1)}$, kao i b2, koeficijent uz X2
- *druga parcijalna H0:* istog tipa kao prva, ali se odnosi na X1
  - H0 glasi: ako se var. X1 uvede u analizu posle var. X2, ona ne doprinosi objašnjenju varijabilnosti Y *preko* doprinosa X2
    - *PRIMER:* ako se u predikciju uspeha na studijama (Y) posle školskog uspeha (X2) u analizu uključi i uspeh na prijemnom ispitu (X1), on ne doprinosi značajno povećanju objašnjenja variranja Y
  - to znači da će u populaciji biti nulte mere koje se odnose na X1: semiparcijalna korelacija $r_{Y(1.2)}$, semiparcijalna determinacija $r^2_{Y(1.2)}$, kao i koeficijent b1
- dalja analiza odvija se analogno kao kod BKRN

2. Multivarijatni KRN

- devijacije: formule su iste kao i kod BKRN
  - totalne devijacije: $y = Y – My$
  - regresioni efekti: $e = Y' – My$
  - regresione greške: $g = Y – Y'$
- jednačine: iste kao i kod BKRN
  - devijaciona jednačina: $y = e + g$
  - strukturna jednačina: $Y = My + e + g$



mere (Y, Y', My) i devijacije (y, g, e)

- *Faza III: test statistik*
  - jednačina zbirova kvadrata: ista kao i kod BKRN
    - $\Sigma y^2 = \Sigma e^2 + \Sigma g^2$ odn. $SS_{TOT} = SS_{reg} + SS_{rez}$
  - deljenjem sa SS_{TOT} : dobija se proporciona jednačina:
    - $R^2 + Q^2 = 1$
    - prvi član je *koeficijent multiple determinacije*
  - deljenjem sa df: prosečni kvadrati: $MS_{reg} = SS_{reg}/df_{reg}$, $MS_{rez} = SS_{rez}/df_{rez}$
  - stepeni slobode: budući da ima dve NV i 1 ZV: $df_{reg} = 2$; $df_{rez} = N\text{-}3$
    - u slučaju k nezavisnih varijabli, $df_{reg} = k$; $df_{rez} = N\text{-}k\text{-}1$
  - F-količnik: $F = \dfrac{MS_{reg}}{MS_{rez}} = \dfrac{\Sigma e^2/2}{\Sigma g^2/(N\text{-}3)}$

2. Multivarijatni KRN

19

- opisani F-količnik vrši *multipli* F-test, tj. koristi se za testiranje *multiple* H0
- za testiranje *parcijalnih* H0 postoje odgovarajući *parcijalni testovi* (ne navodimo ih)
  - ovi testovi mogu koristiti F-količnik ili t-količnik (češće)
- *Faza IV: p-vrednost, Faza V: značajnost:* isto kao u BKRN
  - ako je multipli F-test značajan ($p<0.05$):
    - koeficijenti R i R^2 su statistički značajni
    - X1 i X2 zajedno, u obliku linearne kombinacije $a + b1*X1 + b2*X2$ objašnjavaju statistički značajan deo varijabilnosti zav. var. Y
    - međutim, test je *globalan*, tj. ne govori o pojedinačnim doprinosima X1 i X2
  - ako je prvi parcijalni test (odnosi se na doprinos X2) statistički značajan:
    - mere povezane sa X2 su statistički značajne: $r_{Y(2.1)}$, $r^2_{Y(2.1)}$, $b2$
    - zaključak: X2 statistički značajno povišava procent objašnjene varijabilnosti zav. var. Y, povrh i preko doprinosa X1
  - ako je drugi parcijalni test (odnosi se na X1) statistički značajan:
    - mere povezane sa X1 su statistički značajne: $r_{Y(1.2)}$, $r^2_{Y(1.2)}$, $b1$
    - zaključak: X1 statistički značajno povišava procent objašnjene varijabilnosti zav. var. Y, povrh i preko doprinosa X2

2. Multivarijatni KRN

20

- postoje različite moguće strukture ishoda testova značajnosti:
  - primer 1: R^2 značajan, $r^2_{Y(2.1)}$ značajan, $r^2_{Y(1.2)}$ značajan
    - obe varijable zajedno, a i svaka pojedinačno, povrh one druge, doprinose objašnjavanju variranja Y
  - primer 2: R^2 značajan, $r^2_{Y(2.1)}$ značajan, $r^2_{Y(1.2)}$ neznačajan
    - obe varijable zajedno doprinose, i X2 doprinosi povrh i preko X1, ali X1 ne doprinosi povrh i preko X2
  - primer 3: R^2 značajan, $r^2_{Y(2.1)}$ neznačajan, $r^2_{Y(1.2)}$ neznačajan
    - obe varijable *zajedno* doprinose, ali nijedna ne doprinosi povrh druge
    - može se desiti ako su X1 i X2 međusobno visoko korelirane
- ponekad ima opravdanih razloga da se pojedini prediktori uvode u analizu određenim *redosledom* (engl.: 'stepwise regression')
  - na pr., prvo se uvede X1 u bivarijatnoj analizi, a zatim X2 u multivarijatnoj, a ne razmatra se obrnuta mogućnost (prvo X2 pa onda X1)
  - tada se preklopljena oblast, b, može jednoznačno pripisati dejstvu X1
- *PRIMER:* pretpostavimo da se školski uspeh *obavezno* mora koristiti pri upisu, a da je mogućnost prijemnog ispita ostavljena na izbor pojedinim fakultetima
  - tada bi samo imalo smisla istraživati da li uspeh na prijemnom ispitu doprinosi objašnjavanju uspeha na studijama, povrh i preko doprinosa školskog uspeha

D. Završne napomene

21

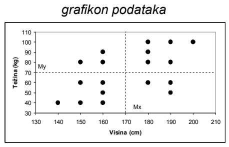
1. Transformacije nacrta

- isti problem se može obraditi pomoću različitih tipova nacrta
  - tip nacrta će zavisiti od tipa varijabli, tj. da li su kategoričke ili numeričke
  - tip varijable može zavisiti od izbora istraživača
- *PRIMER:* odnos visine i težine, na tri načina
- 1. obrada pomoću bivarijatnog korelaciono-regresionog nacrta
  - varijabla X1 *visina*: numerička; varijabla X2 *težina:* numerička

matrica podataka

#	VISINA	TEŽINA
1.	140	40
2.	150	80
...	...	...
17.	190	60
18.	200	100

grafikon podataka



- prosečna visina: $Mx = 170$ cm; prosečna težina: $My = 70$ kg
- postoji korelacije visine i težine, Pirsonov koeficijent : $r_{XY} = 0.53$

D. Završne napomene

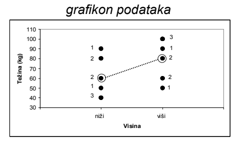
22

- 2. obrada pomoću bivalentnog jednofaktorskog nacrta
  - dihotomizovaćemo visinu
    - transformisaćemo je iz numeričke varijable u kategoričku varijablu
    - nivo a1: *niski* (niži od proseka); nivo a2: *visoki* (viši od proseka)
  - visina: kategorička varijabla; težina: numerička varijabla

matrica podataka

#	VISINA	TEŽINA
1.	nizak	40
2.	nizak	80
...	...	...
17.	visok	60
18.	visok	100

grafikon podataka



uočiti: markeri podataka nisu raspršeni po celom dijagramu, već se nalaze samo na dvema vertikalama (dva nivoa NV)

- prosečna težina niskih: $M1 = 60$ kg; prosečna težina visokih: $M2 = 80$ kg
- grafikon podataka ima neuobičajen oblik
  - naime, obično se koristi linijski grafikon proseka (naznačen na grafikonu)
  - ovde su prikazani i markeri podataka; brojevi uz markere: broj takvih podataka
  - svrha ovakvog prikaza je lakše poređenje sa prikazima druga dva nacrta
- niži su lakši od viših; point-biserijalni koeficijent korelacije: $r_{pb} = 0.48$

D. Završne napomene

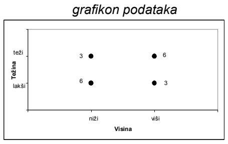
23

- 3. obrada pomoću bivarijatnog frekvencijskog nacrta
  - dihotomizovaćemo pored visine i težinu
    - transformisaćemo je iz numeričke varijable u kategoričku varijablu
    - nivo b1: *laki* (lakši od proseka); nivo b2: *teški:* (teži od proseka)
  - *visina:* kategorička varijabla; *težina:* kategorička varijabla

matrica podataka

#	VISINA	TEŽINA
1.	nizak	lak
2.	nizak	težak
...	...	...
17.	visok	lak
18.	visok	težak

grafikon podataka



uočiti: markeri podataka nisu raspršeni po celom dijagramu, već se nalaze samo na četiri mesta (po dve kategorije za svaku od dve varijable)

- frekvence: teži niži: $a=3$; teži viši: $b=6$; lakši niži: $c=6$; lakši viši: $d=3$
- grafikon podataka ima neuobičajen oblik, radi poređenja
  - grafikon podataka, sa naznačenim brojevima, je ustvari isti kao matrica rezultata tipa 2x2, koja se uobičajeno koristi da prikaže rezultate bivarijatnih nacrta
- fi-koeficijent visine i težine: $\phi = 0.33$
  - korelacija je pozitivna, u smislu da su niže osobe mahom lakše, a više teže

D. Završne napomene

24

- poređenje tri vrste nacrta
  - za date podatke, najprimereniji nacrt je korelaciono-regresioni
  - naime, dihotomizacijom numeričkih varijabli *gubi* se već dobijena informacija, drugim rečima 'bacaju' se podaci
    - korišćenje dobijenih *numeričkih* vrednosti visine i težine je *informativnije* nego korišćenje *kategoričkih* podataka utvrđenih dihotomizacijom, da je data mera iznad ili ispod proseka
  - doduše, dihotomizacijom se može na izvestan način *olakšat*i interpretacija rezultata
    - lakše se razmišlja u suprotnostima tipa 'viši-niži', 'lakši-teži' itd nego sa stvarnim numeričkim vrednostima
    - stoga se u analizama rezultata dihotomizacija numeričkih varijabli često ipak koristi u istraživanjima, iako je to uglavnom neprimereno
- napomena
  - iako su fi-koeficijent i point-biserijalni koeficijent samo specijalni slučajevi Pirsonovog koeficijenta, njihove vrednosti se u primerima razlikuju:
    - $r_{XY} = 0.53$, $r_{pb} = 0.48$, $\phi = 0.33$
  - razlog: ne računaju se na istim već na transformisanim podacima

D. Završne napomene

25

- sledeća *pregledna* poglavlja (8 strana) pročitaćete sami:
- 2. Strukture rezultata u tri vrste nacrta (str. 458-463)
  - veoma sažet pregled mogućih struktura rezultata frekvencijskih, faktorijalnih i korelaciono-regresionih nacrta
- 3. Značajnost rezultata u tri vrste nacrta (str. 464-466)
  - veoma sažet pregled utvrđivanja statističke značajnosti u frekvencijskim, faktorijalnim i korelaciono-regresionim nacrtima

V. Izveštaj o istraživanju

26

naučno istraživanje postaje stvarno deo nauke tek kada je *javno saopšteno*, u obliku izveštaja

postoje više vrsta naučnih izveštaja:

- 1. usmeni izveštaj
  - izlaganje na naučnom skupu ili predavanje po pozivu
- 2. pismeni izveštaj
  - članak u naučnom časopisu
- *3. poster*
  - izlaganje na naučnom skupu pomoću postera

V. Izveštaj o istraživanju

27

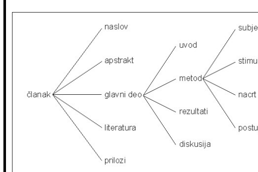
- 1. usmeni izveštaj: izlaganje na naučnom skupu ili po pozivu

- izveštaj ima oblik predavanja, uz kompjutersku prezentaciju
- vreme je ograničeno, čega se izlagač na skupu mora strogo pridržavati
  - obično: oko 15 minuta
  - duža, tzv. plenarna izlaganja: oko 45 minuta
- izveštaj ne treba da sastoji od čitanja napisanog teksta
- izveštaj treba da bude jasan i sažet, bez zalaženja u nepotrebne detalje
- izveštaj treba da ima određenu strukturu
  - u najkraćem: uvodni deo, glavni deo, završni deo
- nakon izlaganja obično slušaoci postavljaju pitanja

V. Izveštaj o istraživanju

28

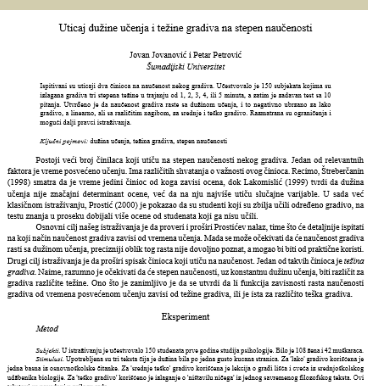
- 2. pismeni izveštaj: članak u naučnom časopisu
  - naučni rad treba da ima određenu strukturu (sklop delova)



- *naslov:* informativan, ne predugačak
  - autori, afilijacije
- *apstrakt*: sažet paragraf koji opisuje istraživanje
- *glavni deo*
- *literatura:* spisak konsultovanih radova (prema određenim pravilima)
- *prilozi*: ako su potrebni
- *uvod:* opšta tema, konkretni problem kojim se bavi istraživanje, pregled prethodnih relevantnih radova
- *metod:* informacije o subjektima, stimulusima (ako ih ima), nacrtu istraživanja, i toku ispitivanja subjekata
- *rezultati:* struktura (šta se dobilo) i statistička značajnost (da li je pouzdano)
- *diskusija:* prikaz i tumačenje rezultata: koji su novi empirijski nalazi, na koji način su teoretski relevantni, kakva su moguća dalja istraživanja

29

- primer izveštaja o istraživanju
  - naslov
  - autori
  - afilijacija
  - apstrakt
  - ključne reči
  - glavni deo
  - metod
  - subjekti
  - stimulusi
  - ...



Uticaj dužine učenja i težine gradiva na stepen naučenosti

Jovan Jovanović i Petar Petrović
Šumadijski Univerzitet

Ispitivani su uticaji dva činioca na naučenost nekog gradiva. Učestvovalo je 150 subjekata kojima su izlagana gradiva tri stepena težine u trajanju od 1, 2, 3, 4, ili 5 minuta, a zatim je zadavan test sa 10 pitanja. Utvrđeno je da naučenost gradiva raste sa dužinom učenja, i to negativno ubrzano za lako gradivo, a linearno, ali sa različitim nagibom, za srednje i teško gradivo. Razmatrana su ograničenja i mogući dalji pravci istraživanja.

Ključni pojmovi: dužina učenja, težina gradiva, stepen naučenosti

Postoji veći broj činilaca koji utiču na stepen naučenosti nekog gradiva. Jedan od relevantnih faktora je vreme posvećeno učenju. Ima različitih shvatanja o važnosti ovog činioca. Recimo, Štreberčanin (1998) smatra da je vreme jedini činioc od koga zavisi ocena, dok Lakomislić (1999) tvrdi da dužina učenja nije značajni determinant ocene, već da na nju najviše utiču slučajne varijable. U sada već klasičnom istraživanju, Prostić (2000) je pokazao da su studenti koji su zbilja učili određeno gradivo, na testu znanja u proseku dobijali više ocene od studenata koji ga nisu učili.

Osnovni cilj našeg istraživanja je da proveri i proširi Prostićev nalaz, time što će detaljnije ispitati na koji način naučenost gradiva zavisi od vremena učenja. Mada se može očekivati da će naučenost gradiva rasti sa dužinom učenja, precizniji oblik tog rasta nije dovoljno poznat, a mogao bi biti od praktične koristi. Drugi cilj istraživanja je da proširi spisak činioca koji utiču na naučenost. Jedan od takvih činioca je *težina gradiva*. Naime, razumno je očekivati da će stepen naučenosti, uz konstantnu dužinu učenja, biti različit za gradiva različite težine. Ono što je zanimljivo je da se utvrdi da li funkcija zavisnosti rasta naučenosti gradiva od vremena posvećenom učenju zavisi od težine gradiva, ili je ista za različito teška gradiva.

Eksperiment

Metod

Subjekti. U istraživanju je učestvovalo 150 studenata prve godine studija psihologije. Bilo je 108 žena i 42 muškaraca.

Stimulusi. Upotrebljena su tri teksta čija je dužina bila po jedna gusto kucana stranica. Za 'lako' gradivo korišćena je jedna basna iz osnovnoškolske čitanke. Za 'srednje teško' gradivo korišćena je lekcija o građi lišća i cveća iz srednjoškolskog udžbenika biologije. Za 'teško gradivo' korišćeno je izlaganje o 'ništavilu ničega' iz jednog savremenog filozofskog teksta. Ovi tekstovi su navedeni u prilogu rada.

V. Izveštaj o istraživanju

30

- 3. poster: usmeno izlaganje uz pomoć velikog plakata (postera)

- poster omogućava direktnu interakciju sa zainteresovanim kolegama
- poster treba da ima određenu strukturu
  - treba da prati i potpomaže usmeni opis autora istraživanja
    - slično kompjuterskim prezentacijama
  - ali, treba da prenese informaciju i bez usmenog izlaganja, slično članku
  - međutim, ne treba da bude ni kopija prezentacije ni kopija članka

dodatak: uočene greške u udžbeniku

31

- strana 137

  stoji: ... kod manipulativnih selektivnih varijabli ...

  treba: ... kod manipulativnih nezavisnih varijabli ...

- strana 383

  na grafikonu IV.B.33a u marginalnoj matrici A nedostaje crtež i oznaka za test $A_{2,3,4}$. Ovaj nedostatk je popravljen u prezentaciji met-18 (faktorijalni nacrti 4)

- strane 440-441

  stoji: u paragrafu sa primerima, tačka (4) odnosi se na grafikon IV.C.17d, a tačka (5) odnosi se na grafikon IV.C.17e

  treba: obrnuto, tj. tačka (4) odnosi se na grafikon IV.C.17e, a tačka (5) na grafikon IV.C.17d