



# Korelacija

Korelacija opisuje jačinu i smer linearne veze između dve promenljive. Pomoću IBM SPSS-a može se izračunati više statističkih pokazatelja, koje korisnik bira u zavisnosti od nivoa merenja promenljivih i prirode podataka. U ovom poglavlju predstaviću postupke dobijanja i tumačenja koeficijenta Pirsonove linearne korelacije ( $r$ ) i koeficijenta Spirmanove korelacije ranga ( $r_o$ ). Pirsonov  $r$  je prikladan za intervalne (neprekidne) promenljive. Može se upotrebiti i kada imate jednu neprekidnu promenljivu (npr. rezultate merenja samopoštovanja) i jednu dihotomnu promenljivu (npr. pol: M/Ž). Spirmanov  $r_o$  je prikladan za ordinalne veličine ili veličine koje se mogu rangirati, i naročito je koristan kada podaci ne zadovoljavaju kriterijume za Pirsonovu korelaciju.

IBM SPSS izračunava obe vrste korelacije. Prvo ćete dobiti jednostavnu bivarijantnu korelaciju (tj. korelaciju između dve promenljive), poznatu i kao korelacija nultog reda (engl. *zero-order correlation*). U IBM SPSS-u takođe možete istraživati odnos između dve promenljive, uz kontrolu treće promenljive. Ta tehnika je poznata kao delimična (parcijalna) korelacija. U ovom poglavlju, predstavljena je procedura dobijanja koeficijenta Pirsonove korelacije proizvod/momenat i neparametarskog Spirmanovog koeficijenta  $r_o$ . Parcijalna korelacija je obrađena u poglavlju 12.

Koeficijenti Pirsonove korelacije ( $r$ ) mogu poprimiti samo vrednosti od  $-1$  do  $+1$ . Predznak pokazuje da li je korelacija pozitivna (obe promenljive zajedno i opadaju i rastu) ili negativna (jedna promenljiva opada kada druga raste i obrnuto). Apsolutna vrednost tog koeficijenta (kada zanemarimo njegov predznak) pokazuje jačinu veze. Savršena korelacija, koja iznosi  $1$  ili  $-1$ , pokazuje da se vrednost jedne promenljive može tačno utvrditi kada znamo vrednost druge. Dijagram rasturanja koji ilustruje takvu vezu prava je linija. S druge strane, korelacija jednaka nuli pokazuje da između te dve promenljive ne postoji nikakva veza. Poznavanje vrednosti jedne promenljive tada nimalo ne pomaže u predviđanju vrednosti druge. Dijagram rasturanja tada izgleda kao oblak nasumično raspoređenih tačaka.

Za tumačenje korelacije potrebno je znati odgovore na više pitanja. To su uticaj nelinearnih veza, netipičnih tačaka, ograničenosti opsega uzorka, odnos korelacije i kauzalnosti, i odnos statističke i praktične značajnosti izračunatih koeficijenata korelacije. Te teme su razmotrene u uvodu u četvrti deo knjige. Pre nego što nastavite čitanje ovog poglavlja, trebalo bi da pročitate taj uvod.

## Objašnjenje primera

Upotrebu korelacije pokazaću na primeru istraživanja uzajamnih veza nekih promenljivih sadržanih u datoteci s podacima **survey.sav**, s prateće veb lokacije ove knjige. Ta anketa je projektovana da istraži činioce koji utiču na psihološku prilagođenost i opšte raspoloženje ispitanika (celovit opis studije videti u dodatku). U ovom primeru, hoću da ocenim korelaciju između osećanja samokontrole ispitanika i njihovog nivoa subjektivno doživljenog stresa. Dve upotrebene promenljive detaljno su opisane u nastavku teksta. Ukoliko želite da pratite i odmah primenjujete navedena uputstva, pokrenite IBM SPSS i otvorite datoteku **survey.sav**.

**Primer istraživačkog pitanja:** Postoji li veza između nivoa samokontrole ljudi nad njihovim unutrašnjim stanjima i njihovog nivoa subjektivno doživljenog stresa? Da li osobe sa visokim nivoom samokontrole doživljavaju manje stresa?

**Šta vam treba:** Dve promenljive, obe neprekidne; ili jedna neprekidna, a druga dihotomna (može imati samo dve vrednosti).

**Šta se postiže:** Korelacija opisuje vezu između dve neprekidne promenljive – i jačinu te veze i njen smer.

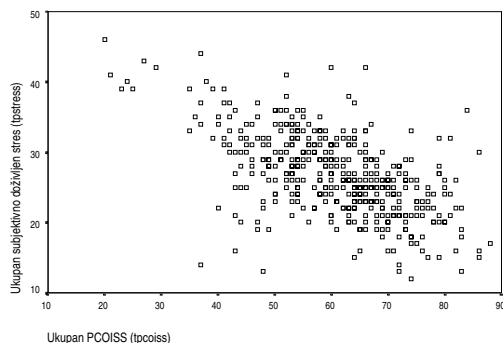
**Pretpostavke:** Videti uvod u četvrti deo knjige.

**Neparametarska alternativa:** Spirmanova korelacija ranga (ro).

## Preliminarne analize za korelaciju

Pre izračunavanja korelacije treba nacrtati dijagram rasturanja. Na njemu proverite da li su zadovoljene pretpostavke o linearnosti i homogenosti varijanse (videti uvod u četvrti deo knjige). Pregledom dijagrama rasturanja steći ćete i bolju predstavu o prirodi veze između ispitivanih promenljivih.

Da biste nacrtali dijagram rasturanja koji prikazuje zavisnost između vaše nezavisne promenljive (Total PCOISS) i zavisne promenljive (Ukupan doživljen stres, tpstress), pratite detaljno uputstvo iz poglavlja 7. Ovde je prikazan rezultat generisan pomoću ovog postupka:



## Tumačenje dijagrama rasturanja

Na dijagramu se može sagledati više svojstava raspodele dve ispitivane promenljive.

### Korak 1: traženje netipičnih tačaka

Pogledajte ima li na dijagramu rasturanja netipičnih tačaka, tj. tačaka izdvojenih od ostalih, izuzetaka, eventualno mnogo većih ili manjih od ostalih. Kada netipične tačke imaju ekstremne vrednosti, treba proveriti da li su informacije uopšte bile ispravno unete? Da to nisu greške? Netipične tačke znatno utiču na neke analize, pa to вреди ispitati. Pojedini udžbenici statistike preporučuju da se ekstremne netipične tačke izbacе iz skupa podataka. Drugi predlažu da im se promene šifre (kodovi) tako da im vrednost ne bude toliko ekstremna (videti poglavlje 6).

Kada prepoznate netipičnu tačku i poželite da je identifikujete pomoću ID broja datog slučaja (opservacije), upotrebite ikonicu **Data Label Mode** u prozoru **Chart Editor**. Dvapat pritisnite dijagram da biste aktivirali prozor **Chart Editor**. Pritisnite ikonicu koja pomalo liči na centar streljačke mete (ili izaberite **Data Label Mode** u meniju **Elements**) i dovedite kursor na tačku dijagrama koju želite da identifikujete. Pritisnite tačku jednom i prikazaće se njen ID broj (ako ste u koraku 5 prethodnog postupka izabrali ID) odnosno broj slučaja (engl. *case number*) koji mu je dodelio IBM SPSS. Prikazivanje brojeva isključićete ako ponovo pritisnete istu ikonicu.

### Korak 2: pregled raspodele vrednosti

Raspodela vrednosti kazuje više stvari:

- Da li su tačke raspršene posvuda? To bi značilo da je korelacija veoma mala.
- Da li su tačke lepo raspoređene u obliku tanke cigare? To bi značilo da je korelacija veoma velika.
- Može li se povući prava linija kroz glavnu gomilu tačaka ili bi tačke bolje predstavila neka kriva linija? Ako je kriva linija očigledna (što sugeriše krivolinijsku vezu između promenljivih), ne treba koristiti Pirsonovu korelaciju pošto ona počiva na pretpostavci da je veza između promenljivih linearna (pravolinijska).
- Koji je oblik glavne gomile tačaka? Da li je jednake širine od jednog kraja do drugog? Ili je na početku uska pa postaje sve šira? U tom slučaju, podaci možda ne zadovoljavaju pretpostavku o homogenosti varijanse (engl. *homoscedasticity, homogeneity of variance*).

### Korak 3: određivanje smera veze između promenljivih

Dijagram rasturanja kazuje da li je veza između dve promenljive pozitivna ili negativna. Kada bi se kroz tačke nacrtala linija, koji smer bi pokazivala – sleva nadesno, da li bi pokazivala naviše ili naniže? Trend naviše prikazuje pozitivnu vezu, tj. da velikim vrednostima X odgovaraju velike vrednosti Y. Linija naniže kazuje da se radi o negativnoj korelaciji, tj. da malim vrednostima X odgovaraju velike vrednosti Y. U ovom primeru, izgleda da je korelacija negativna i umereno jaka.

Pošto ste istražili raspodelu rezultata na dijagramu rasturanja i utvrdili da je veza između promenljivih približno linearna i da su rezultati ravnomerno raspoređeni u obliku cigare, možete nastaviti izračunavanje koeficijenta Pirsonove ili Spirmanove korelacije.

Pre nego što počnete sledeći postupak, u glavnom meniju izaberite **Edit**, zatim **Options**, pa proverite da li je na kartici **General**, u odeljku **Output**, potvrđena opcija **No scientific notation for small numbers in tables**.

### Postupak izračunavanja Pirsonovog koeficijenta r ili Spirmanovog ro

1. U meniju na vrhu prozora pritisnite **Analyze**, zatim **Correlate**, pa **Bivariate**.
2. Izaberite svoje dve promenljive i prebacite ih u polje **Variables** (npr. ukupan subjektivno doživljen stres: tpstress, ukupan PCOISS:tpcoiss). Ovde se može zadati cela lista promenljivih, ne samo dve. U rezultujućoj matrici (tabeli) biće navedene korelacije svih mogućih parova promenljivih. Tabela će biti velika ako u listi zadate više od nekoliko promenljivih.
3. U odeljku **Correlation Coefficients**, podrazumevana opcija je **Pearson** (što znači da se podrazumevano izračunava ta korelacija). Ako hoćete da izračunate i Spirmanovu korelaciju ranga (neparametarsku alternativu), potvrdite i polje **Spearman rho**.
4. Pritisnite dugme **Options**. Za nedostajuće vrednosti (odeljak **Missing Values**), potvrdite opciju **Exclude cases pairwise**. Na kartici **Options** može se zadati i izračunavanje srednje vrednosti i standardnog odstupanja.
5. Pritisnite **Continue** i zatim **OK** (odnosno dugme **Paste** da biste komandu snimili u **Syntax Editor**).

Opisani postupak generiše sledeće komande:

```
CORRELATIONS /VARIABLES=tpstress tpcoiss
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE .
NONPAR CORR /VARIABLES=tpstress tpcoiss
/PRINT=SPEARMAN TWOTAIL NOSIG
/MISSING=PAIRWISE .
```

Rezultat prethodnog postupka (sa obe izračunate korelacije – i Pirsonovom i Spirmanovom) prikazan je u tabelama.

Correlations

		Total perceived stress	Total PCOISS
Total perceived stress	Pearson Correlation	1	-.581**
	Sig. (2-tailed)		.000
	N	433	426
Total PCOISS	Pearson Correlation	-.581**	1
	Sig. (2-tailed)	.000	
	N	426	430

\*\* Correlation is significant at the 0.01 level (2-tailed).

## Neparametarske korelacije

Correlations

			Total perceived stress	Total PCOISS
Spearman's rho	Total perceived stress	Correlation Coefficient	1.000	-.556**
		Sig. (2-tailed)	.	.000
	N		433	426
	Total PCOISS	Correlation Coefficient	-.556**	1.000
Sig. (2-tailed)		.000	.	
N		426	430	

\*\* Correlation is significant at the 0.01 level (2-tailed).

## Tumačenje rezultata korelacije

Rezultate obe korelacije – i Pirsonove i Spirmanove – IBM SPSS prikazuje u tabeli koeficijena korelacije između svakog para navedenih promenljivih, uz nivo značajnosti i broj analiziranih slučajeva (opservacija) u uzorku. Rezultati Pirsonove korelacije dati su u odeljku **Correlation**. Kada zatražite da se izračuna Spirmanov koeficijent  $\rho$  (*rho*), ti rezultati se prikazuju u odeljku **Nonparametric Correlations**. Rezultati obe vrste korelacije, i parametarske i neparametarske, tumače se jednako.

### Korak 1: provera informacija o uzorku

U tabeli **Correlations** najpre treba pogledati N, tj. broj slučajeva (opservacija) u uzorku. Da li je taj broj tačan? Ako mnogo podataka nedostaje, treba da utvrdite zašto je tako. Da niste zaboravili da potvrdite opciju **Exclude cases pairwise** u odeljku **Missing Data** na kartici **Options**? Druga opcija (**Exclude cases listwise**) znači da su iz svih analiza isključeni svi slučajevi kojima nedostaje vrednost bilo koje promenljive. Time se N često preoštro (i nepotrebno) smanjuje. U gornjem primeru imamo 426 slučajeva sa vrednostima na obe skale upotrebene u ovoj analizi. Naravno, iz nje su morali da se izbace svi slučajevi kojima nedostaje podatak za bilo koju od te dve promenljive.

### Korak 2: utvrđivanje smera veze

Drugo što treba pogledati jeste smer veze između promenljivih. Da li je ispred vrednosti koeficijenta korelacije negativan predznak? Ako jeste, to znači da između te dve promenljive postoji negativna korelacija (tj. da velike vrednosti na jednoj skali prate male vrednosti na drugoj). Tumačenje ovoga se menja u zavisnosti od načina dodeljivanja vrednosti promenljivama. Uvek proverite šta piše u upitniku i ne zaboravite da mnoge skale imaju negativno formulisane stavke koje treba obrnuti pre dodeljivanja vrednosti. Šta veliki brojevi (vrednosti) zapravo znače? To često zbunjuje studente, pa vam mora biti savršeno jasno pre nego što se usudite da tumačite rezultate.

U ovom primeru, koeficijent Pirsonove korelacije (-0,58) i Spirmanov koeficijent ro (-0,56) negativni su, što pokazuje negativnu korelaciju između subjektivno doživljene samokontrole i stresa. Što *više* samokontrole osoba oseća da ima, to *manje* stresa misli da doživljava.

### Korak 3: utvrđivanje jačine veze

Treće što treba pogledati u rezultatima jeste veličina koeficijenta korelacije. On može imati vrednost od -1 do 1. Taj broj pokazuje jačinu veze između te dve promenljive. Vrednost 0 pokazuje da nikakva veza (korelacija) ne postoji; vrednost 1 pokazuje da je korelacija potpuna i pozitivna, dok vrednost -1 pokazuje da je korelacija potpuna i negativna.

Kako tumačimo korelacije čije su vrednosti između 0 i 1? Razni autori daju različita tumačenja; međutim, Cohen (1988, str. 79–81) daje sledeće smernice za veličinu korelacije:

mala	$r=0,10$ do $0,29$
srednja	$r=0,30$ do $0,49$
velika	$r=0,50$ do $1,0$

Navedene smernice važe bez obzira na to da li ispred koeficijenta  $r$  stoji negativan predznak. Ne zaboravite, negativan predznak pokazuje smer veze, a ne njenu jačinu. *Jačina* korelacije je jednaka kada su koeficijenti  $r=0,5$  odnosno  $r=-0,5$ . Razlikuje se samo njen *smer*.

U gornjem primeru, između dve promenljive postoji jaka korelacija (iznad 0,5), što govori u prilog tvrdnje da je veza između subjektivno doživljene samokontrole i stresa jaka.

### Korak 4: izračunavanje koeficijenta determinacije

Koeficijent determinacije pokazuje koliki je deo varijanse dve promenljive zajednički; još se kaže „koliki je deo varijanse jedne promenljive objašnjen (ili prouzrokovan) varijansom druge“. Zvuči impresivno, a izračunava se jednostavno; samo kvadirajte vrednost  $r$  (pomnožite  $r$  sa samim sobom). Da biste to pretvorili u 'procenat varijanse', dobijeni proizvod pomnožite sa 100 (pomerite decimalni zarez za dva mesta udesno). Na primer, dve promenljive s koeficijentom korelacije  $r=0,2$  imaju samo  $0,2 \times 0,2 = 0,04 = 4$  procenta zajedničke varijanse. Te dve promenljive nemaju mnogo veze. Međutim, korelacija  $r=0,5$  znači već 25 procenta zajedničke varijanse ( $0,5 \times 0,5 = 0,25$ ).

U našem primeru, Pirsonova korelacija iznosi 0,581; kada se digne na kvadrat, dobija se 33,76 procenta zajedničke varijanse. Subjektivno doživljena samokontrola objašnjava gotovo 34 procenta varijanse u odgovorima ispitanika na skali subjektivno doživljenog stresa. To je sasvim pristojan deo objašnjene varijanse, kada se uporedi s mnogim istraživanjima sprovedenim u društvenim naukama.

### Korak 5: ocena nivoa značajnosti

Sledeće što treba pogledati jeste nivo značajnosti (označen sa **Sig. 2 tailed**). Ovaj podatak se često pogrešno tumači, pa obratite pažnju. Nivo statističke značajnosti ne pokazuje jačinu veze između dve promenljive (nju pokazuje  $r$  odnosno  $r_o$ ), nego s koliko poverenja treba posmatrati dobijene rezultate. Na značajnost dobijenog iznosa koeficijenta  $r$  odnosno  $r_o$  jako utiče veličina uzorka. Iz malih uzoraka (npr.  $n=30$ ) mogu se izračunati umerene korelacije koje nisu statistički značajne na uobičajenom nivou  $p < 0,05$ . Međutim, već i vrlo male korelacije (npr.  $r=0,2$ ) izračunate iz velikih uzoraka ( $N=100+$ ) mogu biti statistički značajne. Naravno da statističku značajnost koeficijenta korelacije treba navesti u izveštaju o istraživanju, ali se usredsredite na jačinu veze i veličinu zajedničke varijanse (videti korak 4).

Pri objavljivanju radova iz nekih oblasti (posebno zdravstva i medicine), možda će se od vas tražiti da navedete intervale poverenja za svoje koeficijente korelacije. Drugim rečima, opseg vrednosti u kome smo 95 posto sigurni da se nalazi 'prava vrednost' (ukoliko smo mogli da je izmerimo!). Nažalost, IBM SPSS to ne omogućava; međutim, postoje veb lokacije s odgovarajućim kalkulatorima. Ako vam trebaju te vrednosti, predlažem da posetite lokaciju <http://vassarstats.net/rho.html>. Možete pregledati i celu lokaciju VassarStats (<http://vassarstats.net/>), koja sadrži niz alatki za obavljanje statističkih izračunavanja.

### Predstavljanje rezultata korelacije

Navešćemo kako bi se rezultati Pirsonove korelacije iz prethodno navedenog primera mogli predstaviti u izveštaju o istraživanju. (Ukoliko treba da predstavite rezultate Spirmanove korelacije, samo zamenite  $r$  vrednošću  $r_o$  datom u drugoj tabeli.)

Veza između subjektivno doživljene samokontrole unutrašnjih stanja (kako je meri skala PCOISS) i subjektivno doživljenog stresa (merenog na skali Perceived Stress) istražena je pomoću koeficijenta Pirsonove linearne korelacije. Obavljene su preliminarne analize da bi se dokazalo zadovoljenje pretpostavki o normalnosti, linearnosti i homogenosti varijanse. Izračunata je jaka negativna korelacija između te dve promenljive,  $r = -0,58$ ,  $n = 426$ ,  $p < ,001$ , pri čemu visoke nivoe subjektivno doživljene samokontrole prate niski nivooi subjektivno doživljenog stresa.

Korelacijom se često istražuju veze unutar cele grupe promenljivih, a ne samo između dve promenljive kao u gornjem primeru. U takvim slučajevima, bilo bi nezgrapno u jednom pasusu navoditi sve izračunate koeficijente korelacije; bolje je prikazati ih u tabeli. Evo kako se to može uraditi: