

# Statistički i grafički prikaz bivarijacionih podataka: rešenja zadataka za vežbe

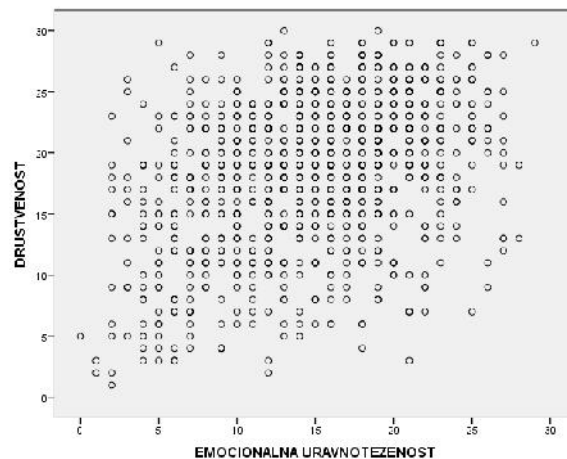
(uradio Luka Milošević, student psihologije, broj indeksa PS160032)

## Zadatak 1.

U fajlu **gzptbu.sav** nalaze se, izmeću ostalog, podaci o društvenosti (varijabla **drus**) i emocionalnoj uravnoteženosti (varijabla **eu**) slučajnog uzorka studenata.

- Grafički prikazati vezu između društvenosti i emocionalne uravnoteženosti (varijabla **eu** ide na X osu, a varijabla **drus** na Y osu);
- Izračunati kovarijansu između društvenosti i emocionalne uravnoteženosti;
- Izračunati koeficijent linearne korelacije između društvenosti i emocionalne uravnoteženosti;

Grafik prikazivanje veze između dve varijable vrši se odabirom menija **Graphs/Legacy Dialogs/Scatter**. Da bismo prikazali vezu između društvenosti i emocionalne uravnoteženosti, u polje **X Axis** unosimo **eu**, dok u polje **Y Axis** unosimo **drus**. Klikom na dugme **OK** dobija se traženi dijagram u prozoru za ispis.



Kovarijansa i koeficijent linearne korelacije između ovih varijabli izračunavaju se odabirom menija **Analyze/Correlate/Bivariate**.

U polje *Variables* prebacuju se iz spiska varijabli varijable **eu** i **drus**. Da bi se u ispisu dobila vrednost kovarijanse, neophodno je kliknuti na dugme *Options* i uključiti opciju *Cross-product deviations and covariances*. Nakon toga, za izračunavanje koeficijenta linearne korelacije potrebno je kliknuti na dugme *Continue* i u okviru *Correlation Coefficients* samo proveriti da li je uključena opcija *Pearson* (koja je inače uključena po defaultu). Klikom na dugme **OK** u ispisu se dobija sledeća tabela:

		EMOCIONALNA URAVNOTEŽENOST	DRUSTVENOST
EMOCIONALNA URAVNOTEŽENOST	Pearson Correlation	1	.361**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	32445.796	12273.587
	Covariance	33.833	12.798
	N	960	960
DRUSTVENOST	Pearson Correlation	.361	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	12273.587	35688.962
	Covariance	12.798	37.215
	N	960	960

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

Vrednosti kovarijanse i koeficijenta linearne korelacije nalaze se u ćeliji koja predstavlja presek kolone sa varijablom **eu** i reda sa varijablom **drus** (ove vrednosti označene su crvenim okvirom). Iste vrednosti nalaze se i u ćeliji koja se dobija u preseku reda u kojem je varijabla **eu** i kolone u kojoj je varijabla **drus**. Dakle, matrica interkorelacija varijabli je simetrična na dijagonali.

*Vidimo da su i kovarijanse i koeficijent linearne korelacije pozitivni. Iz veličine koeficijenta linearne korelacije ( $r = 0.361$ ) vidimo da je na ovom uzorku linearna povezanost između emocionalne uravnoteženosti i društvenosti umerenog intenziteta.*

## Zadatak 2.

Fajl je isti kao za zadatak 1.

- Izračunati pointbiserijsku korelaciju između maskuliniteta i interesovanja i osećanja (varijabla **mask**) i pola ispitanika uobičajenim obrascem;
- Izračunati koeficijent linearne korelacije između maskuliniteta i interesovanja i osećanja i pola ispitanika i uporediti ga sa pointbiserijskom korelacijom;

Koefijent point-biserijske korelacije računava se po obrascu:

Obrazac za računanje pointbiserijske korelacije je sledeći:

$$r_{pbis} = \frac{M_Q - M_P}{S \sqrt{p * q}}$$

U ovom obrascu oznake imaju sledeća značenja:

$M_P$  je aritmetička sredina kvantitativne varijable (u ovom slučaju varijable maskulinitet) za ispitanike koji su ušli u kategoriju P dihotomne varijable (u ovom slučaju je za kategoriju P dihotomne varijable uzeta prva kategorija varijable pol, tj. muško);

$M_Q$  je aritmetička sredina kvantitativne varijable za ispitanike koji su ušli u kategoriju Q dihotomne varijable (u ovom slučaju je za kategoriju Q dihotomne varijable uzeta druga kategorija varijable pol, tj. žensko);

$S$  je standardna devijacija mera na kvantitativnoj varijabli za sve ispitanike;

$p$  je proporcija ispitanika u kategoriji P dihotomne varijable;

$q$  je proporcija ispitanika u kategoriji Q dihotomne varijable.

Za izračunavanje pointbiserijskog koeficijenta korelacije, neophodno je prvo izračunati vrednosti svake od ovih mera.

Da bi se izračunala proporcija ispitanika u obe kategorije varijable **pol**, treba odabrati meni **Analyze/Descriptive Statistics/Frequencies** i u polje *Variable(s)* ubaciti **pol**. Klikom na dugme **OK** u ispisu se dobija tabela koja pokazuje raspodelu ispitanika po kategorijama tražene varijable.

**pol**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid MUSKARCI	480	50.0	50.0	50.0
ZENE	480	50.0	50.0	100.0
Total	960	100.0	100.0	

U toj tabeli su dati (u koloni **Percent**) procenti ispitanika za obe kategorije. Deljenjem tih vrednosti procenata sa 100 dobijaju se tražene proporcije **p** i **q**. U ovom zadatku, obe imaju istu vrednost koja je jednaka 0.5.

Da bi se izra unale aritmeti ke sredine  $M_P$  i  $M_Q$  treba odabrati meni *Data/Split File*. Potom treba kliknuti na dugme *Compare Groups*, u polje *Groups Based on:* staviti **pol** i kliknuti na dugme **OK**. Nakon toga, potrebno je odabrati meni *Analyze/Descriptive Statistics/Frequencies* i u polje *Variable(s)* staviti **mask**. Zatim treba kliknuti na dugme *Statistics* i u okviru *Central Tendency* izabrati *Mean*. Klikom na dugme *Continue*, a potom i na dugme **OK**, u ispisu se dobija tabela u kojoj su izra unate tražene vrednosti (u redovima **Mean**).

**Statistics**

MASKULINOST INTERES. I OSEC.

MUSKARCI	N	Valid	480
		Missing	0
	Mean		14.97
ZENE	N	Valid	480
		Missing	0
	Mean		9.93

Kao što se iz tabele vidi,  $M_P$  je 14.97, dok je  $M_Q$  9.93.

Radi ra unanja **S** neophodno je u meniju *Data/Split File*, kliknuti na dugme *Analyze all cases, do not create groups*, a zatim kliknuti na dugme **OK**. Time smo poništili dejstvo prethodne komande **Split file**. Nakon toga treba odabrati meni *Analyze/Descriptive Statistics/Frequencies* i u polje *Variable(s)* ubaciti **mask**. Zatim treba kliknuti na dugme *Statistics* i u okviru *Dispersion* izabrati *Std. Deviation*. Klikom na dugme *Continue*, a potom i na **OK**, u ispisu se dobija tabela u kojoj je vrednost standardne devijacije.

MASKULINOST INTERES. I OSE

N	Valid	960
	Missing	0
Std. Deviation		4.607

Kao što se iz tabele vidi, **S** je **4.607**.

Izra unavanjem vrednosti svih potrebnih mera mogu e je izra unati i pointbiserijski koeficijent korelacije. Ovo se vrši odabirom menija *Transform/Compute Variable*. U polje *Target Variable*

potrebno je upisati ime nove varijable u koju e biti upisana vrednost ove korelacije (npr. **rpb**). U polje *Numeric Expression* neophodno je upisati  $((14.97-9.93)/4.607)*(\text{Sqrt}(0.5*0.5))$ . Klikom na dugme **OK**, u prozoru za podatke u opciji *Data View* se nalazi tražena vrednost u “varijabli” **rpb**. Ako se uzme u obzir i tre a decimala vrednost koju smo dobili u “varijabli” **rpb** iznosi 0.547.

glm	clm	rpb
8	6	.55
15	1	.55
9	4	.55
15	6	.55
10	1	.55
5	6	.55
9	5	.55
14	5	.55
12	4	.55

Za ra unanje koeficijenta linearne korelacije izme u varijabli **mask** i **pol** odabere se meni *Analyze/Correlate/Bivariate* i u polje *Variables* ubacuju **mask** i **pol**. Nakon toga, u okviru *Correlation Coefficients* treba samo proveriti da li je uklju ena opcija *Pearson* i nakon toga kliknuti na dugme **OK**.

**Correlations**

		MASKULINOST INTERES. I OSEC.	pol
MASKULINOST INTERES. I OSEC.	Pearson Correlation	1	-.547**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	20353.796	-1210.000
	Covariance	21.224	-1.262
	N	960	960
pol	Pearson Correlation	-.547**	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	-1210.000	240.000
	Covariance	-1.262	.250
	N	960	960

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

Kao što vidimo iz tabele, ako se zanemari predznak, koeficijent linearne korelacije koji smo dobili u meniju *Correlate/Bivariate* jednak je point-biserijskom koeficijentu korelacije koji smo dobili komandom *Compute*. Predznak korelacije koja se dobija u meniju *Correlate/Bivariate* je

arbitraran i zavisi od toga kako su označene grupe na dihotomnoj varijabli pol. Budući da muškarci imaju veći u aritmetičku sredinu od žena na varijabli mask, a na kategoriji koja varijabli pol označeni su manjom cifrom (muškarci = 1, žene = 2) izračunati koeficijent u meniju *Correlate/Bivariate* ima negativan predznak. Da smo žene označile ili cifrom 1 a muškarce cifrom 2 koeficijent koji bismo izračunali u meniju *Correlate/Bivariate* imao bi pozitivan predznak. Prema tome, predznak koeficijenta korelacije između u dihotomne i kvantitativne varijable koji računamo u proceduri *Correlate/Bivariate* zavisi od toga kako su označene kategorije na dihotomnoj varijabli. Ako je kategorija koja ima veći u aritmetičku sredinu označena manjom cifrom predznak će biti negativan, a ako je kategorija koja ima veći u aritmetičku sredinu označena većom cifrom na dihotomnoj varijabli predznak će biti pozitivan.

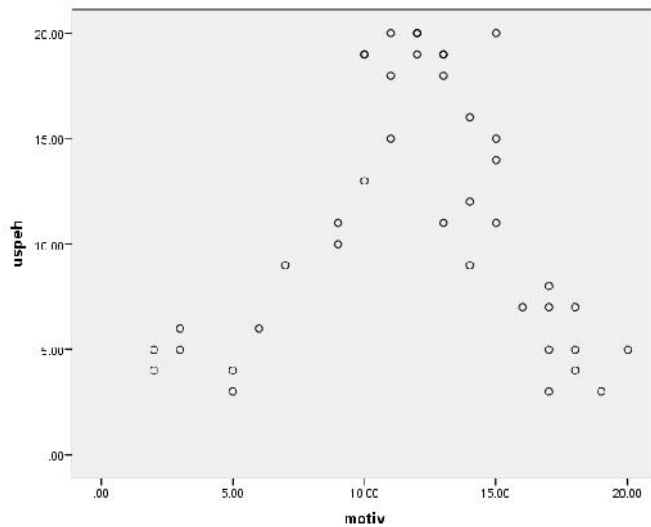
*Dakle, pointbiserijski koeficijent korelacije predstavlja samo poseban slučaj koeficijenta linearne korelacije između u dihotomne i kvantitativne varijable. Računanjem koeficijenta linearne korelacije između u takvih dveju varijabli u meniju Correlate/Bivariate programa SPSS dobija se pointbiserijska korelacija.*

### **Zadatak 3.**

U fajlu **nlin.sav** nalaze se podaci o motivisanosti za obavljanje jednog zadatka (varijabla **motiv**) i o uspešnosti u obavljanju tog zadatka (varijabla **uspeh**).

- Grafički prikazati vezu između u ovih varijabli (varijabla **motiv** ide na X osu, a varijabla **uspeh** na Y osu).
- Izračunati kovarijansu između u motivisanosti i uspeha;
- Izračunati koeficijent linearne korelacije između u motivisanosti i uspeha;

Grafički prikaz veze između u ove dve varijable vrši se odabirom menija *Graphs/Legacy Dialogs/Scatter*. Da bismo prikazali vezu između u motivisanosti za obavljanje jednog zadatka i uspešnosti u obavljanju tog zadatka, u polje *X Axis* unosimo **motiv**, dok u polje *Y Axis* unosimo **uspeh**. Klikom na dugme **OK** dobija se traženi dijagram u prozoru za ispis.



Kovarijansa i koeficijent linearne korelacije izme u ovih varijabli izra unavaju se odabirom menija *Analyze/Correlate/Bivariate*.

U polje *Variables* prebacuju se **motiv** i **uspeh**. Da bi se u ispisu dobila vrednost kovarijanse, neophodno je kliknuti na dugme *Options* i uklju iti opciju *Cross-product deviations and covariances*. Nakon toga, za izra unavanje koeficijenta linearne korelacije, potrebno je kliknuti na dugme *Continue* i u polju *Correlation Coefficients* proveriti da li je uklju ena opcija *Pearson*. Klikom na dugme **OK** u ispisu se dobija tabela u kojoj su prikazane tražene vrednosti.

Correlations

		motiv	uspeh
motiv	Pearson Correlation	1	.062
	Sig. (2-tailed)		.702
	Sum of Squares and Cross-products	942.975	72.900
	Covariance	24.179	1.869
	N	40	40
uspeh	Pearson Correlation	.062	1
	Sig. (2-tailed)	.702	
	Sum of Squares and Cross-products	72.900	1447.600
	Covariance	1.869	37.118
	N	40	40

Vidimo da je koeficijent linearne korelacije veoma blizu nuli mada iz dijagrama raspršenja uo avamo da postoji dosta jaka veza izme u ovih dveju varijabli! Prema tome, kada iz dijagrama raspršenja jasno uo imo da postoji **nelinearni** tip veze izme u varijabli nema smisla ra unati koeficijent **linearne** korelacije.

#### Zadatak 4.

Fajl isti kao za zadatak 3.

- Napraviti varijablu **motiv2** = **-6.82+4.13\*motiv -0.19\* motiv\*\*2**.
- Izra unati koeficijent linearne korelacije izme u varijable motiv2 i uspeha;

Pravljene potrebne varijable se vrši tako što se odabire meni **Transform/Compute Variable**. U polju *Target Variable* unosimo ime nove varijable, što je u ovom slu aju **motiv2**. Nakon toga, u polje *Numeric Expression* unosimo formulu te nove varijable, a ona glasi **-6.82+4.13\*motiv -0.19\* motiv\*\*2**. Klikom na dugme **OK**, u prozoru za podatke, u *Data View* se stvara varijabla **motiv2** i sa njom elije sa vrednostima ove varijable za svakog ispitanika.



	motiv	uspeh	motiv2
1	2.00	4.00	.68
2	3.00	5.00	3.86
3	5.00	3.00	9.08
4	3.00	6.00	3.86
5	2.00	5.00	.68
6	7.00	9.00	12.78
7	9.00	11.00	14.96
8	9.00	10.00	14.96
9	13.00	11.00	14.76
10	10.00	19.00	15.48
11	14.00	9.00	13.76
12	14.00	12.00	13.76
13	15.00	11.00	12.38
14	15.00	15.00	12.38
15	15.00	14.00	12.38

Izra unavanje koeficijenta linearne korelacije izme u varijabli motiv2 i uspeh vrši se odabirom menija *Analyze/Correlate/Bivariate*. U polje *Variables* se ubacuju **motiv2** i **uspeh** i kao i u prethodnim zadacima, proveriti se da li je uklju ena opcija *Pearson* u okviru *Correlation Coefficients*. Klikom na dugme **OK** u ispisu se dobija slede a tabela:

**Correlations**

		uspeh	motiv2
uspeh	Pearson Correlation	1	.791**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	1447.600	911.300
	Covariance	37.118	23.367
	N	40	40
motiv2	Pearson Correlation	.791	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	911.300	915.900
	Covariance	23.367	23.485
	N	40	40

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

Vidimo da je sada koeficijent linearne korelacije veoma visok. Dakle, odgovaraju om transformacijom varijable **motiv** u varijablu **motiv 2** linearizovali smo vezu izme u motivacije i uspeha te sada koeficijent linearne korelacije realnije predstavlja ja inu veze izme u ovih dveju varijabli. Prema tome, koeficijent linearne korelacije nije adekvatna mera povezanosti dveju varijabli ukoliko me u tim varijablama veza nije linearna. Videli smo da je koeficijent linearne korelacije izme u netransformisane motivacije (varijabla **motiv**) i uspeha veoma blizu nuli mada

je iz grafika jasno da postoji jaka **nelinearna** povezanost. (Nelinearna veza u obliku slova U koja postoji između u motivisanosti i uspeha poznata je u psihologiji kao Jerks-Dodsonov zakon).

### Zadatak 5.

U fajlu **estrogen\_alchajm.sav** nalaze se podaci iz medicinskog istraživanja čiji rezultati su objavljeni u poznatom časopisu *Lancet* (Tang, Jacobs, Stern, 1996): U varijabli **estrogen** za svaku od 1124 žene iz slučajnog uzorka dani su podaci o tome da li su dobijale estrogene tokom menopauze (0-NE; 1-DA) a u varijabli **alchajmb** da li su obolele od Alchajmerove bolesti tokom 5-godišnjeg perioda praćenja (0-NE; 1-DA).

- Tabelom predstaviti bivarijatnu raspodelu varijabli **estrogen** i **alchajmb** i izračunati mere asocijacije ovih dveju varijabli (C koeficijent i Kramerov V koeficijent=phi koeficijent u ovom slučaju). (Varijablu **estrogen** stavite u redove, a varijablu **alchajmb** u kolone tabele kontingencije).
- Izračunati šanse da se oboli od Alchajmerove bolesti za žene koje su primale estrogen tokom menopauze;
- Izračunati šanse da se oboli od Alchajmerove bolesti za žene koje nisu primale estrogen tokom menopauze;
- Izračunati koliki su šansi za žene koje nisu primale estrogen u odnosu na žene koje jesu primale estrogen;

Tabelu kojom se predstavlja bivarijatna raspodela dveju kategoričkih varijabli dobija se tako što se odabire meni *Analyze/Descriptive Statistics/Crosstabs*. U polje *Row(s)* stavljamo **estrogen**, dok u polje *Column(s)* stavljamo **alchajmb**. Da bi se izračunali C koeficijent i Kramerov V koeficijent neophodno je kliknuti na dugme *Statistics* i u okviru *Nominal* izabrati *Contingency coefficient* i *Phi and Cramer's V*. Posle toga treba kliknuti na dugme *Continue* i na **OK**.

U ispisu se dobijaju dve tabele.

Prva tabela predstavlja bivarijatnu raspodelu datih varijabli:

**Da li je primala estrogene tokom menopauze ^ Da li je tokom 5 godina obolela od Alchajmerove bolesti Crosstabulation**

Count

		Da li je tokom 5 godina obolela od Alchajmerove bolesti		Total
		ne	da	
Da li je primala estrogene tokom menopauze	ne	810	158	968
	da	147	9	156
Total		957	167	1124

U drugoj tabeli se dobijaju vrednosti C koeficijenta i Kramerovog V koeficijenta:

**Symmetric Measures**

		Value	Approx. Sig.
Nominal by Nominal	Phi	-.103	.001
	Cramer's V	.103	.001
	Contingency Coefficient	.102	.001
N of Valid Cases		1124	

Šansa da se oboli od Alchajmerove bolesti za žene koje su primale estrogen tokom menopauze dobija se tako što se frekvencija žena koje su primale estrogen tokom menopauze i obolele od Alchajmerove bolesti podeli sa frekvencijom žena koje su primale estrogen tokom menopauze ali nisu obolele od Alchajmerove bolesti.

Da bi se ovo uradilo potrebno je odabrati meni **Transform/Compute Variable** i u polju *Target Variable* upisati ime nove varijable ija e vrednost biti jednaka traženoj šansi (npr. **soape**). U polje *Numeric Expression* upisujemo **9/147**. Klikom na dugme **OK**, u prozoru za podatke u opciji *Data View* dobijamo tražene šansu u vidu nove varijable **soape**:

	rbr	estrogen	alchajmb	soape
1	1	1	0	.06
2	2	1	0	.06
3	3	1	0	.06
4	4	1	0	.06
5	5	1	0	.06
6	6	1	0	.06
7	7	1	0	.06
8	8	1	0	.06
9	9	1	0	.06
10	10	1	0	.06
11	11	1	0	.06

Šansa da se oboli od Alchajmerove bolesti za žene koje nisu primale estrogen tokom menopauze dobija se tako što se frekvencija žena koje nisu primale estrogen tokom menopauze, a pritom su obolele od Alchajmerove bolesti podeli sa frekvencijom žena koje nisu primale estrogen tokom menopauze i nisu obolele od Alchajmerove bolesti.

Da bi se ovo uradilo potrebno je opet odabrati meni *Transform/Compute Variable* i u polju *Target Variable* upisati ime varijable *soape* i vrednost biti jednaka traženim šansama (npr. *soanpe*). U polje *Numeric Expression* upisujemo **158/810**. Klikom na dugme **OK**, u prozoru za podatke u opciji *Data View* dobijamo tražene šanse u vidu nove “varijable”:

	rbr	estrogen	alchajmb	soape	soanpe
1	1	1	0	.06	.20
2	2	1	0	.06	.20
3	3	1	0	.06	.20
4	4	1	0	.06	.20
5	5	1	0	.06	.20
6	6	1	0	.06	.20
7	7	1	0	.06	.20

Koli nik šansi oboljevanja od Alchajmerove bolesti za žene koje nisu primale estrogen u odnosu na žene koje jesu primale estrogen izra unava se tako što se šanse da se oboli od Alchajmerove bolesti za žene koje nisu primale estrogen podeli sa šansama da se oboli od Alchajmerove bolesti za žene koje su primale estrogen.

Da bi se ova vrednost dobila potrebno je, kao i u prethodnim slučajevima, odabrati meni *Transform/Compute Variable* i u polju *Target Variable* upisati ime varijable *ks* i vrednost biti jednaka traženom količniku šansi (npr. **ks**). U polje *Numeric Expression* upisujemo **soanpe/soape**. Klikom na dugme **OK**, u prozoru za podatke u opciji *Data View* dobijamo traženi količnik šansi u vidu nove "varijable" **ks**:

	rbr	estrogen	alchajmb	soape	soanpe	ks
1	1	1	0	.06	.20	3.19
2	2	1	0	.06	.20	3.19
3	3	1	0	.06	.20	3.19
4	4	1	0	.06	.20	3.19
5	5	1	0	.06	.20	3.19
6	6	1	0	.06	.20	3.19

*Dakle, kada je o ženama iz ovog uzorka reč, žene koje nisu primale estrogen tokom menopauze imaju 3.19 puta veće šanse da obole od Alchajmerove bolesti nego žene koje jesu primale estrogen tokom menopauze.*