

Linearna korelacija i jednostruka regresija: rešenja zadataka sa vežbi

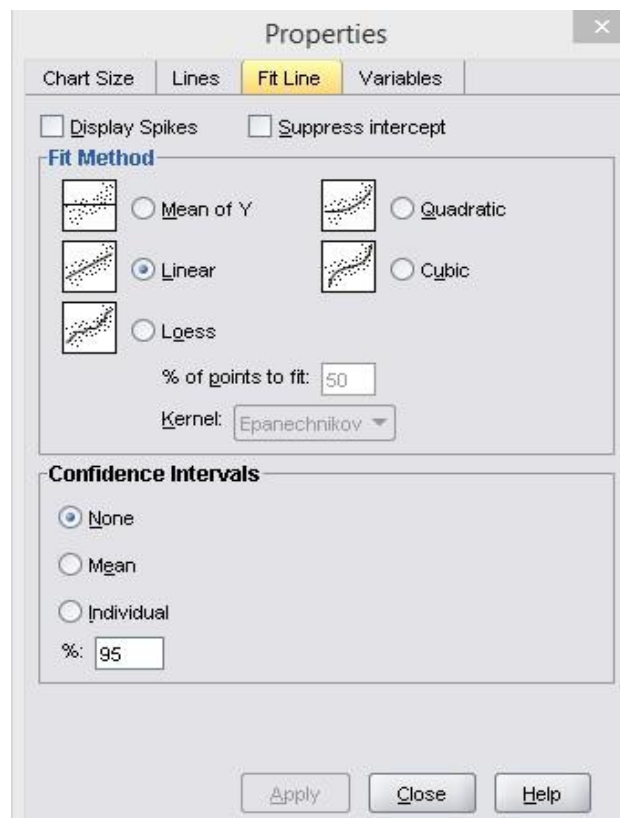
Rešenja uradila Kova evic Ana, PS 14/36, studentkinja psihologije

Zadatak 1:

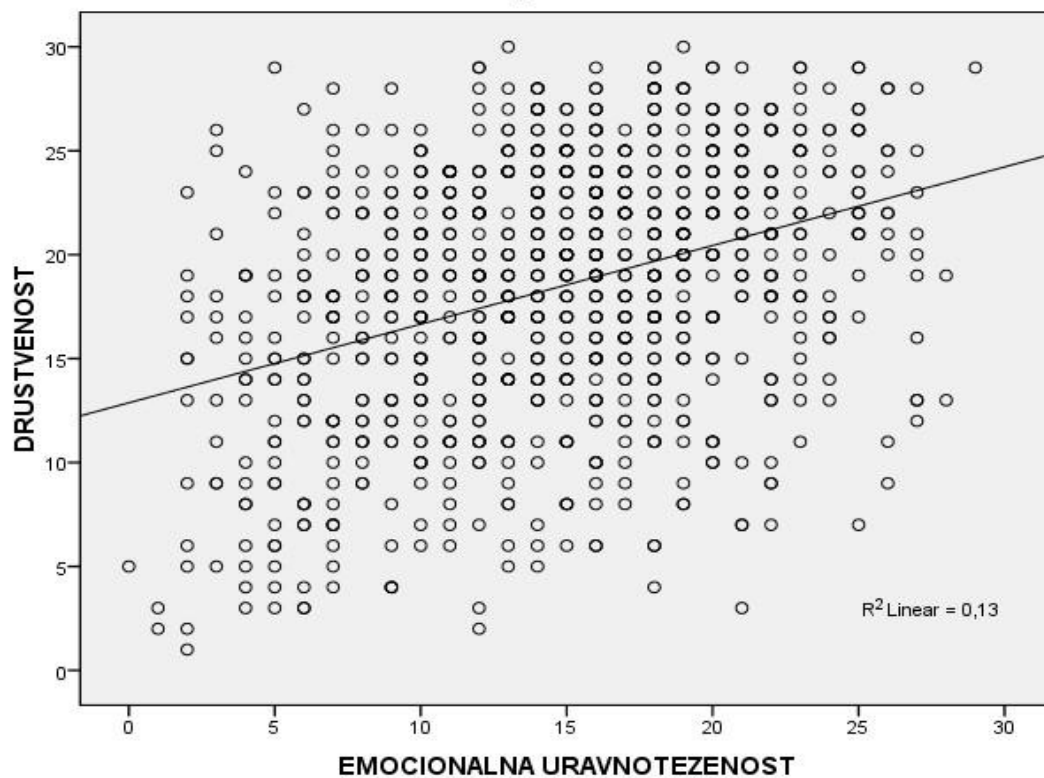
U fajlu G:\SUP\podaci\gzptbu.sav nalaze se, izme u ostalog, podaci o društvenosti (varijabla drus) i emocionalnoj uravnoteženosti (varijabla eu) slu ajnog uzorka studenata..

- Grafi ki prikazati vezu izme u društvenosti i emocionalne uravnoteženosti;
- Izra unati kovarijansu izme u društvenosti i emocionalne uravnoteženosti;
- Izra unati koeficijent linearne korelacije izme u društvenosti i emocionalne uravnoteženosti;
- Testirati statisti ku zna ajnost koeficijenta linearne korelacije.
- Napraviti 99% interval poverenja za koeficijent linearne korelacije
- Odredite veli inu efekta (koeficijent determinacije *100) korelisanosti dveju varijabli.

Kako bismo grafi ki prikazali vezu izme u društvenosti i emocionalne uravnoteženosti treba kliknuti na Graphs, a zatim u Legacy dialogs izabrati Scatter/Dot, zatim kliknuti na Simple Scatter pa i i na Define. U novom prozoru Simple Scatterplot unosimo u Y axis jednu od varijabli (na primer emocionalnu uravnoteženost), a u X axis varijablu društvenost. Klikom na dugme Titles možemo upisati naslov dijagrama. Kada napravimo dijagram, dvostrukim klikom na njega otvaramo novi prozor Chart Editor. Kliknemo na Elements i odaberemo Fit Line at Total, a u novom prozoru Properties pod karticom Fit line biramo u



okviru Fit Method opciju Linear.



Mada dijagram ne pokazuje sve pojedina ne ispitanike u uzorku (jer ih je mnogo) može se uociti pozitivna korelisanost ove dve varijable na uzorku.

Da bismo izračunali kovarijansu, koeficijent linearne korelacije i njegovu statističku značajnost kliknemo na Analyze, Correlate, Bivariate i u desni prozor ubacujemo varijable društvenost i emocionalna uravnoteženost. Kako je Pearsonov koeficijent već štrikliran, ulazimo u Options i uključimo Cross-product deviations and covariances (da bismo dobili kovarijansu).

U ispisu dobijamo tabelu:

Correlations			
		DRUSTVENOST	EMOCIONALNA URAVNOTEZENOST
DRUSTVENOST	Pearson Correlation	1	,361**
	Sig. (2-tailed)		,000
	Sum of Squares and Cross-products	35688,962	12273,587
	Covariance	37,215	12,798
	N	960	960
EMOCIONALNA URAVNOTEZENOST	Pearson Correlation	,361**	1
	Sig. (2-tailed)	,000	
	Sum of Squares and Cross-products	12273,587	32445,796
	Covariance	12,798	33,833
	N	960	960

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

Iz tabele vidimo da je koeficijent umerenog intenziteta i iznosi 0,361, a vidimo i da je pozitivan što smo uo ili i na grafiku. Koeficijent je statistiki značajan na nivou 0,01, a kovarijansa iznosi 12,798.

U tabeli smo dobili tačku ocenu parametra, tj. koeficijenta linearne korelacije u populaciji, ali ocenjivanje parametra možemo vršiti i preko intervala poverenja. Kako u programu SPSS ne postoji opcija za računavanje intervala poverenja koeficijenta linearne korelacije (ako se izuzme mogućnost samouzrkovanja preko opcije Bootstrap) to se mora računati postupno. Interval se može dobiti preko komande Transform, Compute Variable. Ukratko o samom postupku pravljenja intervala: ideja je da se r koeficijent pretvori u Fišerov z statistik, zatim da se za Fišerov statistik odrede granice za interval (gornja i donja) a da se zatim te granice izražene u metrici Fišerovog statistika transformišu u gornju i donju granicu intervala poverenja za koeficijent linearne korelacije. Postupak je sledeći (svaka varijabla se dobija korišćenjem komande Transform, Compute Variable):

1. Prva varijabla koju pravimo je Fišerov z statistik, FZ. U Numeric Expression kucamo sledeću formulu $0.5 \cdot \ln((1+r)/(1-r))$. Pri tome, umesto r unosimo određenu vrednost koeficijenta linearne korelacije (najbolje da je kopiramo iz tabele).
2. Druga varijabla je standardna greška (SEZ) za prethodno izračunati statistik i dobijamo je formulom $1/\sqrt{N-3}$ pri čemu je N veličina uzorka tako da umesto N u Numeric Expression treba uneti konkretan broj za veličinu uzorka.
3. Treća i četvrta varijabla jesu donja (fzd) i gornja (fzg) granica intervala za Fišerov statistik i dobijamo ih formulama: za fzd u Numeric Expression treba uneti $FZ - 2.58 \cdot SEZ$ a za fzg treba uneti $FZ + 2.58 \cdot SEZ$. Biramo 2.58 zato što se u zadatku traži 99% interval poverenja. Za 95% interval poverenja bismo umesto 2.58 uneli 1.96.
4. Peta i šesta varijabla jesu donja (rd) i gornja granica (rg) intervala poverenja za koeficijent linearne korelacije: za donju granicu u Numeric Expression treba uneti $(\exp(2 \cdot fzd) - 1) / (\exp(2 \cdot fzd) + 1)$, a za gornju granicu intervala poverenja $(\exp(2 \cdot fzg) - 1) / (\exp(2 \cdot fzg) + 1)$.

U Data View možemo videti da varijable rd i rg iznose 0.29 i 0.43, što znači da sa 99% sigurnosti možemo reći da smo ovim intervalom obuhvatili parametar.

Da bismo odredili veličinu efekta koristimo formulu $r^2 \cdot 100$, pri čemu u polje Numeric expression umesto r unosimo određenu vrednost koeficijenta linearne korelacije (najbolje da je kopiramo iz tabele). Inače, r kvadrat smo dobili i na grafiku. Dobija se srednji efekat koji iznosi 13.01%.

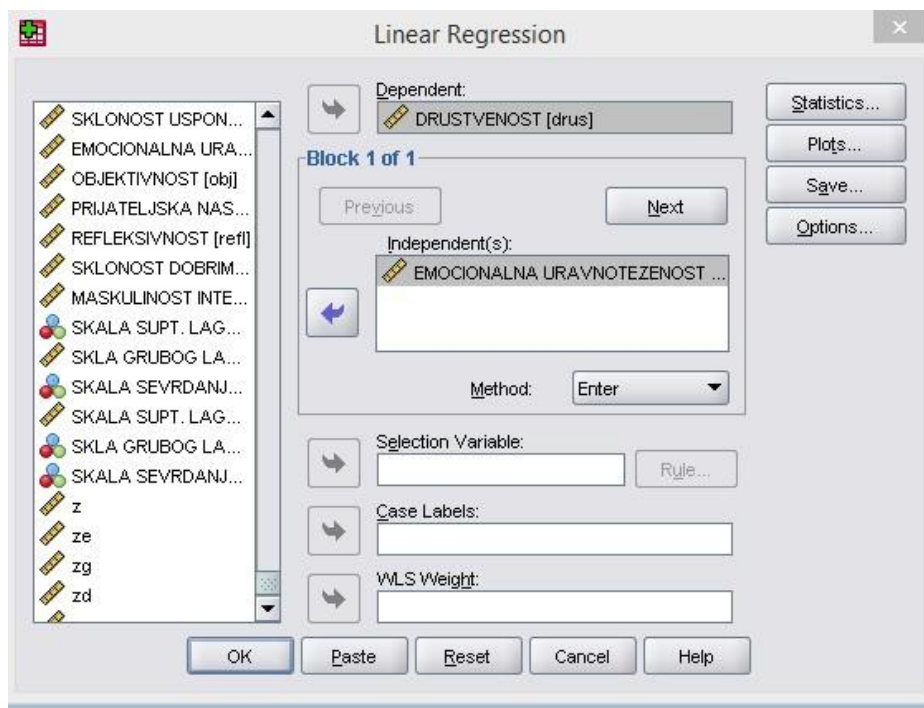
Zadatak 2:

Fajl isti kao za zadatak 1.

- Napraviti linearni regresioni model za predviđanje društvenosti na osnovu emocionalne uravnoteženosti studenata;
- Napraviti linearni regresioni model za predviđanje emocionalne uravnoteženosti na osnovu društvenosti studenata;
- Izračunati koeficijent linearne korelacije između društvenosti i emocionalne uravnoteženosti samo korišćenjem nestandardizovanih regresionih koeficijenata za predviđanje emocionalne uravnoteženosti na osnovu društvenosti studenata i za predviđanje društvenosti na osnovu emocionalne uravnoteženosti studenata;
- Izračunati koeficijent linearne korelacije između društvenosti i emocionalne

uravnoteženost samo koriš enjem standardizovanih regresionih koeficijenata za predvi anje emocionalne uravnoteženosti na osnovu društvenosti studenata i za predvi anje društvenosti na osnovu emocionalne uravnoteženosti studenata.

Da bismo napravili modele koji se traže u prve dve ta ke treba i i na Analize, Regression, Linear i zatim (za prvu ta ku) ubaciti u Dependent varijablu drustvenost, a u



Independent(s) emocionalnu uravnotezenost:

Ispis:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,361 ^a	,130	,129	5,693

a. Predictors: (Constant), EMOCIONALNA URAVNOTEZENOST

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4642,850	1	4642,850	143,266	,000 ^a
	Residual	31046,113	958	32,407		
	Total	35688,962	959			

a. Predictors: (Constant), EMOCIONALNA URAVNOTEZENOST

b. Dependent Variable: DRUSTVENOST

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12,878	,500		25,754	,000
	EMOCIONALNA URAVNOTEZENOST	,378	,032	,361	11,969	,000

a. Dependent Variable: DRUSTVENOST

Prva tabela Model Summary: R je koeficijent multiple korelacije, a kako u analizi imamo dve varijable on je jednak obi nom Brave-Pirsonovom koeficijentu korelacije koji je dobijen u prvom zadatku; R square je koeficijent multiple determinacije koji kad se pomnoži sa 100 pokazuje koliko procenata varijabilnosti u zavisnoj varijabli možemo da objasnimo (predvidimo) na osnovu nezavisne; tre a vrednost je korigovani koeficijent multiple determinacije, a predstavlja isto što i koeficijent determinacije samo što se odnosi na populaciju jer predstavlja manje pristrasnu ocenu koeficijenta determinacije u populaciji. Poslednja vrednost predstavlja standardnu gresku ocene koja govori o tome koliko verovatno prose no grešimo kada na osnovu emocionalne uravnoteženosti predvi amo društvenost. Iskazana je u jedinicama kriterijumske (zavisne) varijable, u ovom slu aju društvenosti.

Tabela ANOVA pokazuje ishod testa statisti ke zna ajnosti koeficijenta linearne korelacije. Kada se sume kvadrata (Sum of Squares) podele stepenima slobode koji se nalaze u redu sa njima (za regresiju je $df=m$, a za rezidualne je $N-m-1$) dobijaju se prose ni kvadrati odstupanja (Mean Square) iji koli nik predstavlja F-statistik. U koloni Sig se nalazi verovatno a da se dobije F-statistik ve i ili jednak dobijenom ako je nulta hipoteza ta na. U zadatku je dobijena verovatno a manja od 0,05 tako da se odbacuje nulta hipoteza koja pretpostavlja da je linearna korelacija ove dve varijable u populaciji jednaka nuli.

U tabeli Coefficients su date vrednosti nestandardizovanih regresionih koeficijenata (B), standardizovanih regresionih koeficijenata (BETA), t-statistik (dobija se formulom B/SeB , tj. $B/Std.Error$), i verovatno e kojima se testira znacajnost regresionih koeficijenata (kolona Sig). Regresioni koeficijenti izražavaju nagib prave koja se po principu metoda najmanjih kvadrata povla i kroz dijagram raspršenja. Nestandardizovani regresioni koeficijent pokazuje za koliko se mernih jedinica prose no promeni kriterijumska varijabla ako se prediktorska promeni za jednu mernu jedinicu, dok standardizovani regresioni koeficijenti pokazuju isto to samo preko standardnih devijacija. Ta nije, za koliko se standardnih devijacija prose no promeni kriterijumska varijabla ako se prediktorska promeni za jednu standardnu devijaciju. Constant predstavlja intercept, tj. vrednost kriterijumske varijable kada je prediktorska

varijabla jednaka nuli.

Po istom principu se radi i predvi anje emocionalne uravnoteženosti na osnovu društvenosti samo što se u Analyze-Regression-Linear u Dependent stavlja emocionalna uravnoteženost, dok se u Independent(s) ubacuje društvenost. Zbog simetri nosti linearne korelacije ove dve varijable (jednaka predvi anja društvenosti na osnovu emocionalne uravnoteženosti i obrnuto) dobijaju se iste vrednosti u tabeli Model Summary osim

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,361 ^a	,130	,129	5,428

a. Predictors: (Constant), DRUSTVENOST

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4220,939	1	4220,939	143,266	,000 ^a
	Residual	28224,856	958	29,462		
	Total	32445,796	959			

a. Predictors: (Constant), DRUSTVENOST

b. Dependent Variable: EMOCIONALNA URAVNOTEZENOST

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8,372	,558		14,999	,000
	DRUSTVENOST	,344	,029	,361	11,969	,000

a. Dependent Variable: EMOCIONALNA URAVNOTEZENOST

standardne greške ocene, tj. standardne greške za regresiju.

F-statistik i njegova statistička značajnost su isti u obema tabelama.

Naravno, u tabelama Coefficients konstante, tj. intercepti i nestandardizovani regresioni koeficijenti nisu isti ali su standardizovani regresioni koeficijenti isti.

Koeficijent linearne korelacije je jednak geometrijskoj sredini koeficijenata b (za predvi anje x na osnovu y i za predvi anje y na osnovu x). Za nestandardizovane koeficijente je $r = (b_{yx} * b_{xy}) = (0.344 * 0.378) = 0.361$, a za standardizovane je $r = BETA = 0.361$.

Regresione jednačine za naš zadatak su:

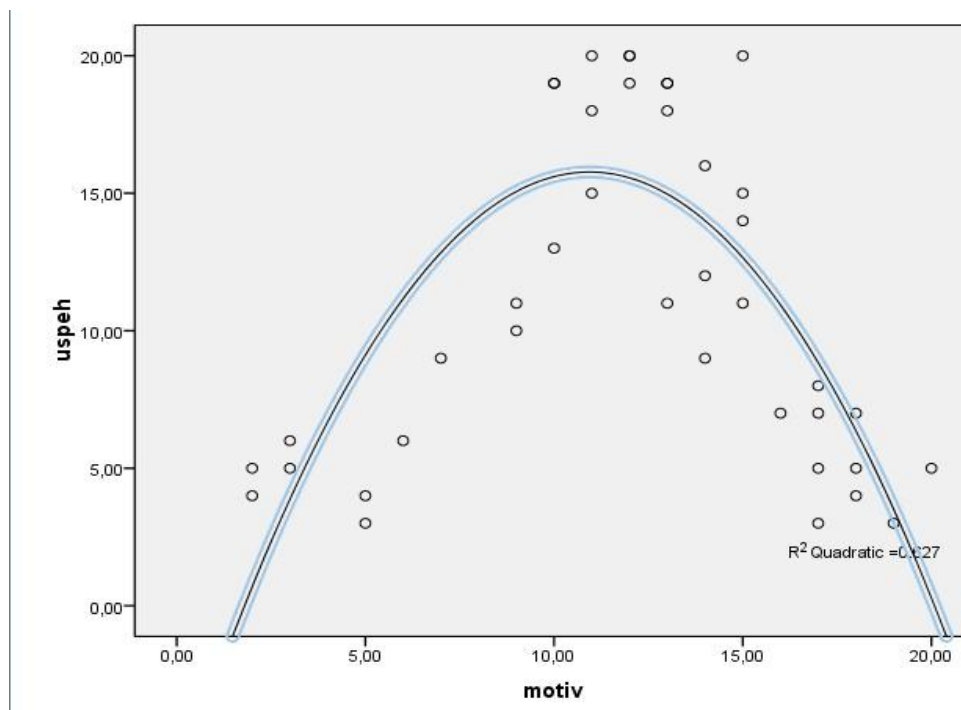
- $drus = 12.878 + 0.378 * eu$
- $eu = 8.372 + 0.344 * drus$

Zadatak 3:

U fajlu G:\SUP\podaci\mlin.sav nalaze se podaci o motivisanosti za obavljanje jednog zadatka (varijabla motiv) i o uspešnosti u obavljanju tog zadatka (varijabla uspeh).

- Grafi ki prikazati vezu izme u ovih varijabli
- Izra unati kovarijansu izme u motivisanosti i uspeha;
- Izra unati koeficijent linearne korelacije izme u motivisanosti i uspeha;
- Testirati statisti ku zna ajnost koeficijenta linearne korelacije. Šta se na osnovu koeficijenta linearne korelacije i njegove statisti ke zna ajnosti može zaklju iti?

Pravimo grafik po postupku koji je pokazan u prvom zadatku. U Y axis unosimo varijablu uspeh, a u X axis varijablu motiv. Jos jedna razlika u odnosu na prvi zadatak je ta što dvostrukim klikom na grafik, pa zatim Fit Line at Total (nalazi se u kartici Elements) i u novom prozoru Properties, u kartici Fit line sada biramo QUADRATIC.



Odabirom opcije Quadratic mozemo uo iti da postoji povezanost motiva i uspeha, a ona se može predstaviti dobro obrnutom U funkcijom. Sa porastom motiva raste i uspeh, ali do odre ene ta ke, od koje ce daljim porastom motiva opadati skor na varijabli uspeh.

Racunamo kovarijansu i koeficijent linearne korelacije kao što je ve objašnjeno i dobijamo u ispisu:

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

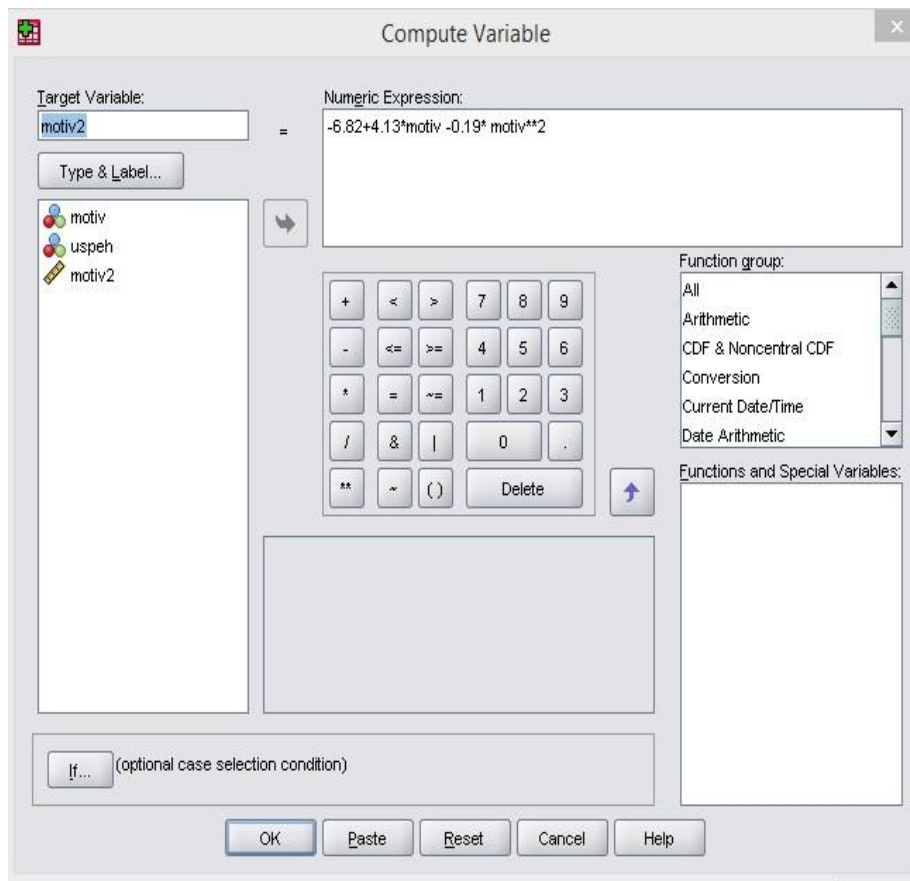
Kovarijansa iznosi 1,869, koefijent linearne korelacije je 0,062 što predstavlja vrlo nizak intenzitet, a na osnovu verovatnoće koja iznosi 0,702 zaključujemo da nije statistički značajan. To znači da ne postoji linearna povezanost ovih varijabli, ali na osnovu toga ne smemo zaključiti da između ovih varijabli ne postoji nikakva povezanost! U ovakvim slučajevima se kao mera povezanosti može koristiti korelacioni razmer.

Zadatak 4.

Fajl isti kao za zadatak 3.

- Napraviti varijablu $\text{motiv2} = -6.82 + 4.13 * \text{motiv} - 0.19 * \text{motiv}^2$.
- Izračunati koeficijent linearne korelacije između varijabli motiv2 i uspeha ;
- Testirati statističku značajnost koeficijenta linearne korelacije. Šta se na osnovu koeficijenta linearne korelacije i njegove statističke značajnosti može zaključiti?

Preko Transform-Compute pravimo novu varijablu motiv2 :



Ra unanjem koeficijenta linearne korelacije dobijamo:

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).

Kako je varijabla motiv2 kvadratna (nelinearna) funkcija varijable motiv, na osnovu nje se dobro predvi a uspeh, a o jakoj povezanosti varijable motiv2 i uspeha govori i koeficijent linearne korelacije koji u ovom slu aju iznosi 0.791 i statisticki je zna ajan na nivou 0,01. Dakle, kako je motiv2 kvadratna (nelinearna) funkcija varijable motiv, zaklju ujemo da izme u motivisanosti i uspeha postoji povezanost ali je ona nelinearna! Ne smemo

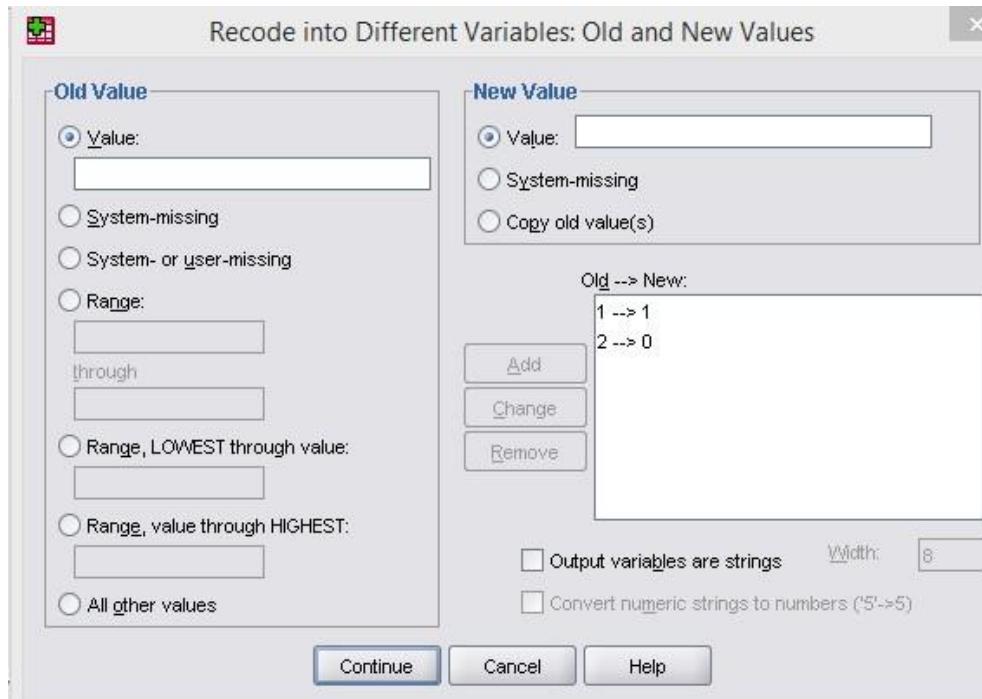
zaključiti da između motivisanosti i uspeha postoji linearna povezanost jer je visok i statistički značajan koeficijent linearne korelacije dobijen između u nelinearno transformisane motivisanosti i uspeha.

Zadatak 5:

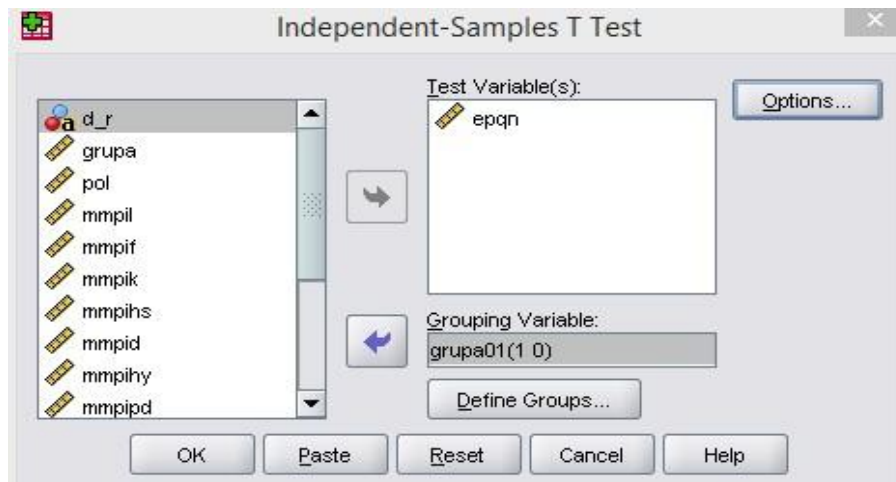
U fajlu G:\SUP\podaci\astma.sav postoje podaci o neuroticizmu (varijabla EPQN) za astmatičare i neastmatičare (varijabla GRUPA pokazuje kojoj od ovih kategorija ispitanik pripada: 1. astmatičari). Od varijable GRUPA napravite novu varijablu GRUPA01 u kojoj će astmatičari (GRUPA = 1) zadržati oznaku 1, a neastmatičari (GRUPA = 2) dobiti oznaku 0.

- Testirajte nultu hipotezu (astmatičari i neastmatičari u populaciji se ne razlikuju prema prosečno neuroticizmu) t-testom. Šta zaključujete?
- Napravite linearnu regresionu analizu u kojoj je kriterijumska varijabla EPQN, a prediktorska varijabla je GRUPA01. U Tabeli Model Summary pogledajte R i uporedite ga sa pointbiserijskom korelacijom između u varijabli GRUPA01 i EPQN. Šta zaključujete? U Tabeli Coefficients pogledajte Standardized coefficient Beta i uporedite ga sa pointbiserijskom korelacijom između u varijabli GRUPA01 i EPQN. Šta zaključujete? U Tabeli Coefficients pogledajte t statistiku za regresioni koeficijent za varijablu GRUPA01 i uporedite ga sa t statistikom iz t-testa (tabela Independent samples test). Šta zaključujete? Uporedite intercept iz regresionog modela sa aritmetičkom sredinom na varijabli EPQN za grupu koja je na varijabli GRUPA01 označena oznakom 0 (neastmatičari). A zatim uporedite razliku između u aritmetičkim sredinama astmatičara i neastmatičara za varijablu EPQN sa nestandardizovanim regresionim koeficijentom (kolona B u tabeli Coefficients) za varijablu GRUPA01. Šta zaključujete na osnovu svega što ste uradili? Da li će uvek postojati ovakva slaganja rezultata regresione analize i t testa za nezavisne uzorke? Pod kojim uslovima će se ovo dešavati?

Da bismo napravili varijablu GRUPA01 kliknemo na Transform-Recode into different variables, zatim u desni okvir prebacujemo varijablu GRUPA, u polje Name upisujemo ime nove varijable, a to je GRUPA01, kliknemo na Change, a potom na Old and New values. U novom prozoru se u delu za Old Value u polje Value upisuje stara oznaka prve kategorije (ako bismo prvo radili za astmatičare onda tu upisujemo 1), a u delu za New Value se u polje Value upisuje nova oznaka 1 (kada stara oznaka ostaje ista može se izabrati i opcija, u delu new value, Copy old value(s)). Zatim se klikne na Add. Za drugu kategoriju kao staru vrednost upisujemo 2, a kao novu 0, pa Add, Continue i OK.



Kako su astmaticari i neastmaticari nezavisni uzorci koristimo t-test za nezavisne uzorke. Kliknemo na Analyze-Compare means-Independent Samples T test. U polje Test variable ubacujemo EPQN, a u Grouping variable GRUPA01. Zatim u Define groups upisujemo oznake kategorija varijable GRUPA01 (na primer, prvu grupu označimo jedinicom, a drugu nulom).



drugu nulom).

Dobijamo ispis:

Group Statistics

	grupa01	N	Mean	Std. Deviation	Std. Error Mean
epqn	1,00	128	15,92	3,906	,345
	,00	128	10,77	3,842	,340

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
epqn	Equal variances assumed	,251	,617	10,632	254	,000	5,148	,484	4,195	6,102
	Equal variances not assumed			10,632	253,931	,000	5,148	,484	4,195	6,102

Kako je testiranjem Leveneovim testom dobijeno da se nulta hipoteza koja pretpostavlja homoscedasticnost ne može odbaciti, gledamo prvi red tabele. T-statistik iznosi 10.632, a dobijeno je i da je on statistički značajan, tako da se odbacuje nulta hipoteza koja pretpostavlja da se dve grupe ne razlikuju po prosečnom neuroticizmu. Značajnost se može videti i na osnovu intervala poverenja, zato što on u ovom slučaju ne obuhvata nulu.

Regresionu analizu vršimo kao što smo pokazali u prethodnim zadacima. U Dependent ubacujemo EPQN, a u Independent(s) GRUPA01. Dobijamo ispis:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,555 ^a	,308	,305	3,874

a. Predictors: (Constant), grupa01

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1696,410	1	1696,410	113,045	,000 ^a
	Residual	3811,648	254	15,006		
	Total	5508,059	255			

a. Predictors: (Constant), grupa01

b. Dependent Variable: epqn

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10,773	,342		31,464	,000
	grupa01	5,148	,484	,555	10,632	,000

a. Dependent Variable: epqn

Koeficijent pointbiserijske korelacije dobijamo isto kao i Pirsonov koeficijent linearne korelacije.

Correlations

		epqn	grupa01
epqn	Pearson Correlation	1	,555**
	Sig. (2-tailed)		,000
	N	256	256
grupa01	Pearson Correlation	,555**	1
	Sig. (2-tailed)	,000	
	N	256	256

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

U Variables unosimo EPQN i GRUPA01.

Upoređivanjem R i pointbiserijskog koeficijenta korelacije uo avamo da su oni jednaki i iznose 0.555, kao i da su oba statisti ki zna ajna što govori o povezanosti testiranih varijabli, kao i da se na osnovu podatka o pripadnosti kategoriji astmati ara odnosno neastmati ara može predvi ati neuroticizam.

Pore enjem pointbiserijskog koeficijenta korelacije sa standardizovanim regresionim koeficijentom (BETA) možemo videti da su oni jednaki (0,555). Kako je poinbiserijski koeficijent korelacije matemati ki jednak koeficijentu linearne korelacije izme u kvantitativne (zavisna) i binarne kategoricke varijable (nezavisna), a kako je koeficijent linearne korelacije

jednak standardizovanom regresionom koeficijentu sledi da su i standardizovani regresioni koeficijent i pointbiserijski koeficijent korelacije u ovom slu aju jednaki.

T-statistik iz t-testa koji pretpostavlja homogenost varijansi i t-statistik iz regresione analize za testiranje zna ajnosti regresionog koeficijenta za prediktorsku varijablu grupa01 su jednaki. U obe analize se proverava nulta hipoteza koja pretpostavlja da ne postoji povezanost neuroticizma i astmaticnosti, tako da se u ovom slu aju nepovezanost može proveriti bilo kojim od ova dva modela. Postojanje razlika izme u dve grupe koje je utvr eno t-testom se može utvrditi i regresionom analizom, proverom zna ajnosti t-statistika za regresioni koeficijent za prediktorsku varijablu grupa01, jer se on u ovom slu aju svodi na pointbiserijski koeficijent korelacije, a on upravo pokazuje povezanost odnosno nepovezanost dveju varijabli.

Intercept iz regresione analize i aritmeti ka sredina na zavisnoj varijabli za kategoriju koja je ozna ena nulom (neastmatična) imaju jednake vrednosti što zna i da se u ovoj situaciji predvi ena vrednost na osnovu linearne regresije po principu najmanjih kvadrata kada je prediktorska binarna varijabla jednaka nuli svodi na aritmeti ke sredine grupe koja je ozna ena nulom. To sledi iz toga što je intercept u linearnoj regresiji predvi ena vrednost kriterijumske varijable kada je prediktorska varijabla jednaka nuli. Predvi ena vrednost u regresiji po principu najmanjih kvadrata je zapravo aritmeti ka sredina na kriterijumskoj varijabli za ispitanike sa odre enom (istom za sve) vrednoš u prediktorske varijable.

Razlike aritmeti kih sredina dveju kategorija jednake su nestandardizovanom regresionom koeficijentu za prediktorsku binarnu varijablu grupa01. To sledi otuda što nestandardizovani regresioni koeficijent pokazuje za koliko se menja predvi ena vrednost na zavisnoj varijabli sa promenom nezavisne za jednu jedinicu (a ta promena je u ovom slu aju u stvari promena iz kategorije 0 u kategoriju 1).

Iz ovoga se može izvesti zaklju ak da se regresiona analiza i testiranje statisti ke zna ajnosti razlika izme u aritmeti kih sredina dva nezavisna uzorka t-testom za homogene varijanse svodi na istu analizu ako se u regresionoj analizi zavisna varijabla iz t-testa koristi kao kriterijumska, a binarna kategoricka varijabla sa dve kategorije koja u t-testu definiše grupe koristi kao nezavisna (prediktorska) varijabla. Isto tako testiranje statisti ke zna ajnosti pointbiserijskog koeficijenta korelacije je isto kao t-test za testiranje zna ajnosti aritmeti kih sredina pod uslovom da se koristi t-test koji pretpostavlja homogenost varijansi (homoscedasti nost).

Dakle, važno je uo iti da sve opisane jednakosti t-testa sa linearnom regresionom analizom i pointbiserijskom korelacijom važe samo kada je re o t-testu koji pretpostavlja homogenost varijansi subpopulacija.