

## Loglinearni modeli: rešenja zadataka

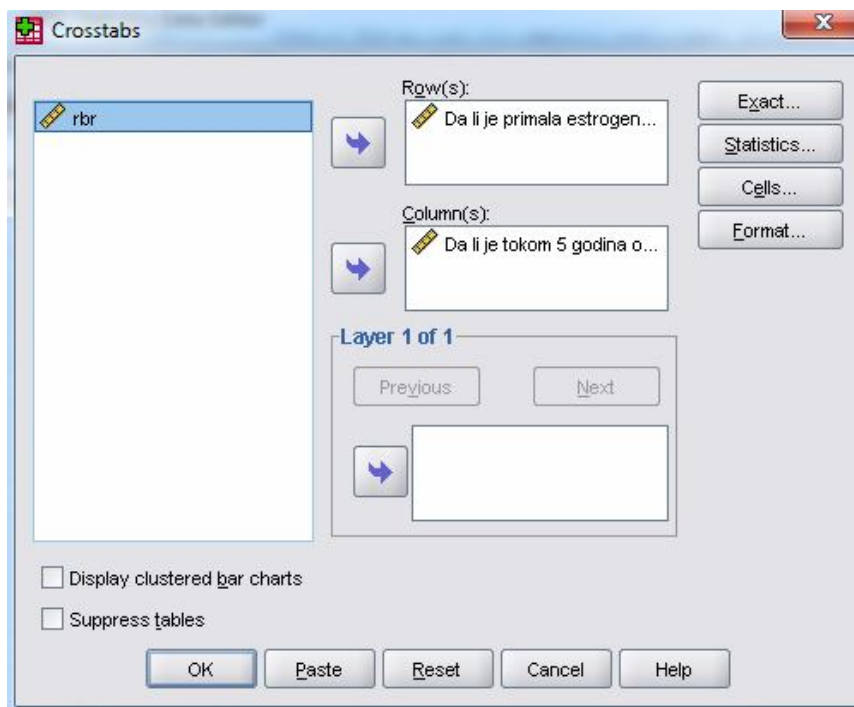
(Rešenja uradila Katarina Rančić, studentkinja psihologije, generacija 2012/2013)

### Zadatak 1.

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 dati su podaci o tome da li su dobijale estrogene tokom menopauze (0-NE; 1-DA) a u varijabli **alchajmb** da li su obolerle od Alchajmerove bolesti tokom 5-godišnjeg perioda praćenja (0-NE; 1-DA).

- 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 količnik šansi oboljevanja od Alchajmerove bolesti za žene koje nisu primale estrogen u odnosu na žene koje jesu primale estrogen;
- Utvrditi da li su oboljevanje od Alchajmerove bolesti i primanje estrogena tokom menopauze statistički nezavisne varijable.
- Utvrditi primenom loglinearnog modela da li je model koji pretpostavlja da nema asocijacije ovih varijabli u skladu sa podacima.

-Da bismo izračunali šanse, potrebno je da napravimo tabelu kontingencije, u našem slučaju tabelu 2 x 2. Idemo na **Analyze → Descriptive Statistics → Crosstabs**. Varijable *estrogen* i *alchajmb* prebacujemo na desnu stranu prozora, u redove (**Row(s)**) i kolone (**Column(s)**). Redosled ubacivanja nije bitan.



Pritiskom na **OK** dobijamo tabelu:

**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

Gledamo red gde se nalaze žene koje su primale estrogen, a tražene šanse računamo kao količnik/odnos onih koje su obolele i onih koje nisu (za označeni red!).

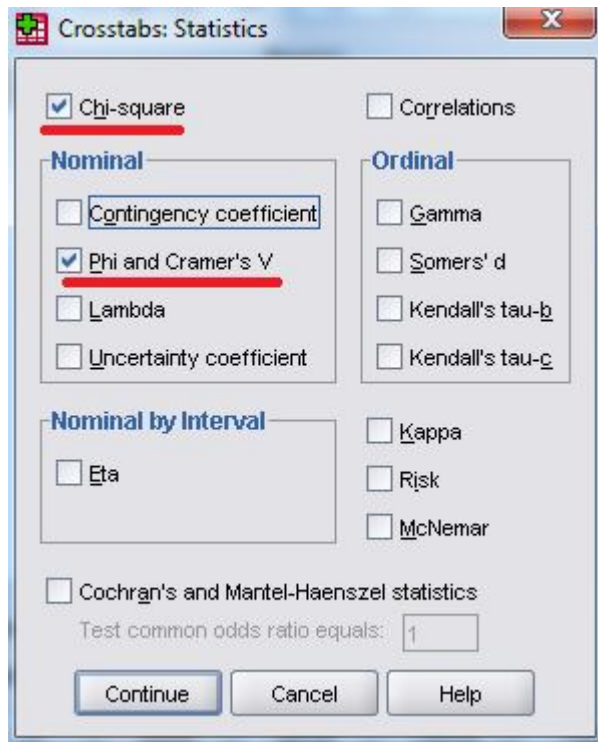
$\hat{S} = 9/147 \rightarrow \hat{S} = 0.061$  Postupak se može izvršiti pomoću kalkulatora ili komande **Compute**.

-Da bismo izračunali šanse oboljevanja za one koje nisu primale estrogen koristimo potpuno isti postupak opisan u prethodnom zadatku, ali za red IZNAD označenog-u redu žena koje nisu primale estrogen.  $\hat{S} = 158/810 = 0.195$

-Za računanje količnika šansi oboljevanja za žene koje nisu primale u odnosu na one koje su primale estrogen treba da podelimo izračunate šanse i to navedenim redosledom-šanse oboljevanja onih koje nisu primale estrogen/šanse oboljenja onih koje jesu primale

estrogen:  $K\check{S}=(158/810)/(9/147) =3.19$  što znači da su šanse oboljevanja za žene koje nisu primale estrogen više od 3 puta veće u odnosu na one koje su primale!

-Za testiranje nezavisnosti, u našem slučaju, dveju kategoričkih varijabli koristimo Hi-kvadrat test i to postupkom: **Analyze** → **Descriptive Statistics** → **Crosstabs** (kao u prvom delu zadatka), na isti način ubacimo varijable, a zatim u kartici **Statistics** označimo **Chi-square** i (eventualno) **Phi and Cramer's V** kao na sledećem prikazu:



Klikom na **Continue**, vraćamo se na prethodni prozor i kliknemo na opciju **Cells**. Označimo **Observed I Expected** u okviru **Counts**, kliknemo na **Continue** i zatim **OK**.



U ispisu dobijamo sledeće:

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

			Da li je tokom 5 godina obolela od Alchajmerove bolesti		Total
			ne	da	
Da li je primala estrogene tokom menopauze	ne	Count	810	158	968
		<u>Expected Count</u>	824,2	143,8	968,0
	da	Count	147	9	156
		<u>Expected Count</u>	132,8	23,2	156,0
Total	Count	957	167	1124	
	Expected Count	957,0	167,0	1124,0	

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
<u>Pearson Chi-Square</u>	<u>11,828<sup>a</sup></u>	1	<u>,001</u>		
Continuity Correction <sup>b</sup>	11,008	1	,001		
Likelihood Ratio	14,390	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	11,817	1	,001		
N of Valid Cases	1124				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 23,18.

b. Computed only for a 2x2 table

**Symmetric Measures**

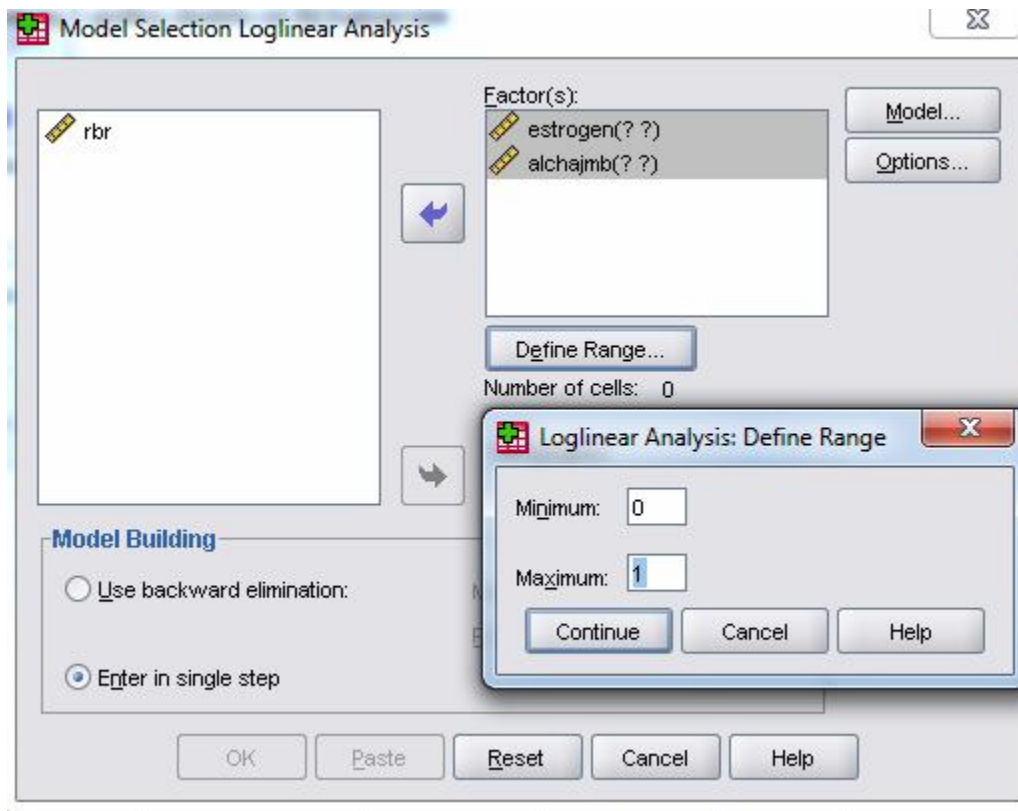
		Value	Approx. Sig.
Nominal by Nominal	Phi	-,103	<u>,001</u>
	Cramer's V	,103	<u>,001</u>
N of Valid Cases		1124	

U prvoj tabeli sa opaženim i očekivanim frekvencijama uočavamo da su očekivane frekvencije, prema fiksnim marginalnim, u približno istom odnosu po redovima i kolonama. Za redove je taj odnos 5.7, a za kolone 6.2. Takav odnos bi imale i dobijene frekvencije kada bi ove dve varijable bile nezavisne. O statističkoj nezavisnosti ovih varijabli nam govori Pirsonov statistik-Chi-square, odnosno njegova značajnost. Nulta hipoteza bi glasila da ove varijable jesu statistički nezavisne, tj. da se odnosi dobijenih frekvencija poklapaju sa odnosima očekivanih frekvencija. Medjutim, kada bi ova hipoteza bila tacna, verovatnoća da dobijemo Hi-kvadrat statistic jednak ili veći od 11.828 jednaka je 0.001. Stoga, na nivou 0.01 možemo odbaciti nultu hipotezu i reći da ove dve varijable nisu statistički nezavisne!

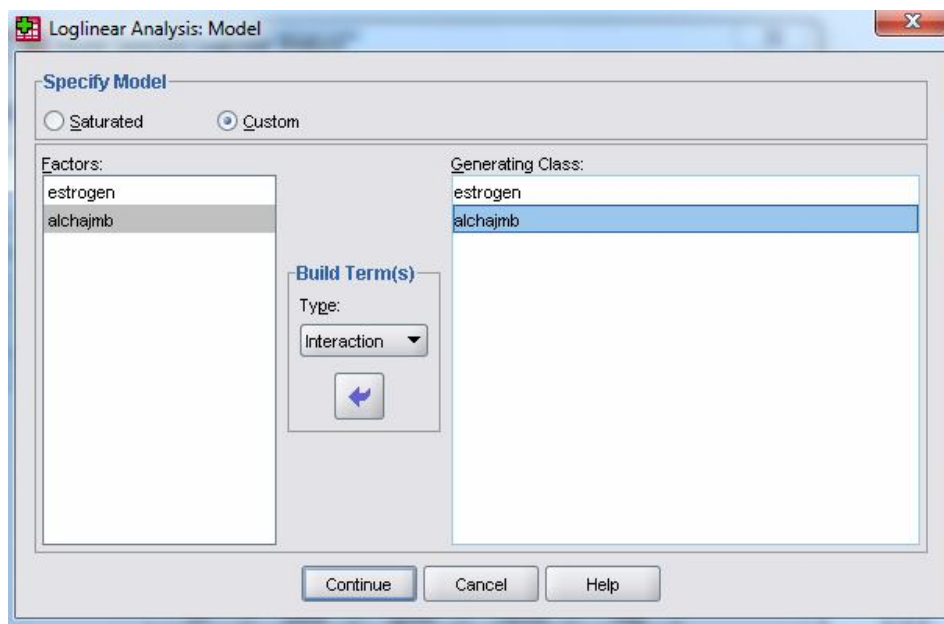
-Loglinearni model:

**Analyze → Loglinear → Model Selection;** U prozoru **Model Selection**, u okviru **Model Building**, obeležimo **Enter in single step**, varijable *estrogen* i *alchajmb* prebacimo u desni deo prozora **Factor(s)**, obeležimo obe varijable pošto imaju iste definisane

vrednosti (koje možemo proveriti u **Values Label**) i klikom na **Define Range** kao Minimum ubacimo vrednost 0, a kao Maximum vrednost 1. Uključimo opciju **Enter in single step**, kliknemo na **Continue** i na dugme **Model**.



U novom prozoru, u okviru **Specify model** obeležimo **Custom** i pošto se pretpostavlja da nema asocijacije izmedju varijabli, ne biramo njihovu interakciju, već jednu po jednu varijablu iz levog dela **Factors** prebacujemo na desnu stranu **Generating Class**.



Klikom na **Continue** i zatim **OK**, dobijamo ispis:

**Goodness-of-Fit Tests**

	Chi-Square	df	Sig.
Likelihood Ratio	14,390	1	,000
Pearson	11,828	1	,001

U poslednjoj tabeli koja je prikazana na slici stoji Hi-kvadrat statistik na principu verodostojnosti. Pošto je statistički značajan, model nije u saglasnosti sa našim podacima jer među varijablama ipak postoji povezanost, što smo upravo prethodno utvrdili i testom nezavisnosti ovih dveju varijabli.

## Zadatak 2.

U fajlu **pusenje\_simpson.sav** su podaci iz istraživanja na slučajnom uzorku žena koje je izvedeno u dve faze: U prvoj fazi su (između ostalog) prikupljeni podaci o tome u koju uzrasnu kategoriju ispitanica spada (varijabla **starost**) i da li ispitanica puši ili ne (varijabla **pusac**). Posle 20 godina ustanovljeno je preživljavanje ispitanica, tj. da li su ispitanice žive ili ne (varijabla **zive**).

- Napraviti tabelu kontingencije varijabli **pusac** i **zive**;
- Izračunati šanse preživljavanja za pušačice u odnosu na nepušačice bez uzimanja u obzir starosti;

- Izračunati šanse preživljavanja za pušačice u odnosu na nepušačice posebno po kategorijama starosti;
- Primenom loglinearnog modela ustanoviti da li je hipoteza o uzajamnoj nezavisnosti varijabli **starost**, **pusac** i **zive** u skladu sa podacima;
- Odabarti najjednostavniji hijerarhijski loglinearni model koji je u skladu sa podacima.

-Tabelu kontingencije pravimo prema već opisanom postupku iz prvog dela prvog zadatka: **Analyze** → **Descriptive Statistics** → **Crosstabs**. U redove u kolone ubacujemo samo varijable **pusac** i **zive** (ne uzimamo u obzir varijablu *starost*). U ispisu dobijamo sledeću tabelu kontingencije:

**pusac \* zive Crosstabulation**

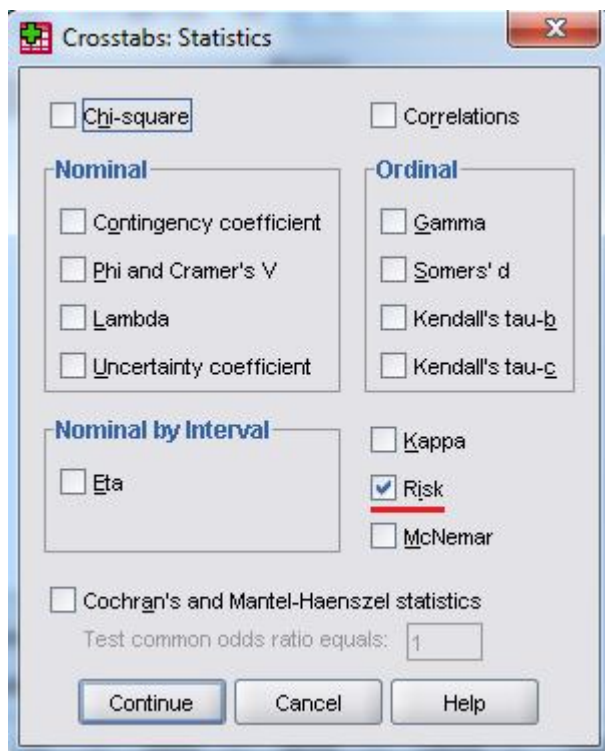
Count

		zive		Total
		zive	umrle	
pusac	ne	502	230	732
	da	443	139	582
Total		945	369	1314

Traže se šanse preživljavanja pušačica u ODNOSU na šanse preživljavanja nepušačica, što je zapravo količnik šansi.  $\hat{S}_1=443/139$ ,  $\hat{S}_2=502/230$ ,  $K\hat{S}=\hat{S}_1/\hat{S}_2=1.46$

Traženi odnos šansi možemo naći i na drugi način. Ulaskom u prozor **Crosstabs**, ostavljamo već ubačene varijable u redove i kolone, a zatim idemo na karticu **Statistics** i obeležavamo **Risk** kao na prikazu:





Nakon klika na **Continue** i zatim **OK**, u ispisu dobijamo tabelu:

**Risk Estimate**

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for <u>ne</u> / da)	<u>.685</u>	,535	,876
For cohort zive = zive	,901	,843	,963
For cohort zive = umrla	1,316	1,099	1,575
N of Valid Cases	1314		

Međutim, vidimo da je vrednost 0.685 za količnik šansi NEpušačica! Odnos šansi preživljavanja za pušačice i šansi za nepušačice dobijamo kao recipročnu vrednost dobijene vrednosti:  $1/0.685=1.46$ .

-Šanse preživljavanja za pušačice u odnosu na nepušačice za svaki nivo varijable *starost* dobijamo na sledeći način: U prozoru **Crosstabs** (do koga dolazimo kao i do sada opisanim postupkom), gde su ubačene varijable *pusac* i *zive*, u deo **Layer 1 of 1** ubacimo varijablu *starost*. U okviru kartice **Statistics** obeležimo **Risk** i u ispisu dobijamo:

**pusac \* zive \* starost Crosstabulation**

Count

starost			zive		Total
			zive	umrle	
18-34	pusac	ne	213	6	219
		da	174	5	179
	Total		387	11	398
35-44	pusac	ne	114	7	121
		da	95	14	109
	Total		209	21	230
45-64	pusac	ne	147	52	199
		da	167	78	245
	Total		314	130	444
65+	pusac	ne	28	165	193
		da	7	42	49
	Total		35	207	242

**Risk Estimate**

starost		Value	95% Confidence Interval	
			Lower	Upper
18-34	Odds Ratio for pusac (ne / da)	1,020	,306	3,399
	For cohort zive = zive	1,001	,968	1,034
	For cohort zive = umrle	,981	,304	3,161
	N of Valid Cases	398		
35-44	Odds Ratio for pusac (ne / da)	2,400	,931	6,189
	For cohort zive = zive	1,081	,993	1,176
	For cohort zive = umrle	,450	,189	1,075
	N of Valid Cases	230		
45-64	Odds Ratio for pusac (ne / da)	1,320	,872	2,000
	For cohort zive = zive	1,084	,962	1,221
	For cohort zive = umrle	,821	,610	1,104
	N of Valid Cases	444		
65+	Odds Ratio for pusac (ne / da)	1,018	,416	2,491
	For cohort zive = zive	1,016	,472	2,186
	For cohort zive = umrle	,997	,877	1,134
	N of Valid Cases	242		

Sve tražene količnike šansi za pojedini nivo varijable *starost*, dobićemo kao recipročnu vrednost dobijene vrednosti (**Value**) u tabeli **Risk Estimate** kao **Odds Ratio for pusac (ne/da)**. Na primer za prvi nivo starosti (18-24) KŠ je jednak 1/1.020, što iznosi 0.98

koliko dobijamo i standardnim računom “peške”. Za prvu kategoriju npr. Taj račun bi bio sledeći:  $K\check{S}=(174/5)/(213/6)=(174*6)/(5*213)=1044/1065=0.98$

-Po postupku koji je opisan u prethodnom zadatku, odemo na **Loglinear → Model Selection** i u datom prozoru prebacimo sve varijable sa leve na desnu stranu (Factor(s)). Pri definisanju vrednosti za svaku varijablu voditi računa jer se ovog puta za varijablu *starost* kao Minimum ubacuje 1, a kao Maximum 4, dok se za preostale dve varijable definišu iste vrednosti: 0 i 1. U **Model Buildingu** uključiti **Enter in single step**. U prozoru koji otvaramo klikom na **Model** i obeležavanjem **Custom** prebacujemo svaku varijablu posebno u desni prostor jer se testira NEZAVISNOST varijabli (da li je u skladu sa podacima).

**Cell Counts and Residuals**

starost	pusac	zive	Observed		Expected		Residuals	Std. Residuals
			Count	%	Count	%		
18-34	ne	zive	213,000	16,2%	159,454	12,1%	53,546	4,240
		umrle	6,000	,5%	62,263	4,7%	-56,263	-7,130
	da	zive	174,000	13,2%	126,779	9,6%	47,221	4,194
		umrle	5,000	,4%	49,504	3,8%	-44,504	-6,325
35-44	ne	zive	114,000	8,7%	92,147	7,0%	21,853	2,277
		umrle	7,000	,5%	35,981	2,7%	-28,981	-4,831
	da	zive	95,000	7,2%	73,264	5,6%	21,736	2,539
		umrle	14,000	1,1%	28,608	2,2%	-14,608	-2,731
45-64	ne	zive	147,000	11,2%	177,883	13,5%	-30,883	-2,316
		umrle	52,000	4,0%	69,459	5,3%	-17,459	-2,095
	da	zive	167,000	12,7%	141,432	10,8%	25,568	2,150
		umrle	78,000	5,9%	55,226	4,2%	22,774	3,065
65+	ne	zive	28,000	2,1%	96,954	7,4%	-68,954	-7,003
		umrle	165,000	12,6%	37,858	2,9%	127,142	20,664
	da	zive	7,000	,5%	77,087	5,9%	-70,087	-7,983
		umrle	42,000	3,2%	30,101	2,3%	11,899	2,169

**Goodness-of-Fit Tests**

	Chi-Square	df	Sig.
Likelihood Ratio	671,306	10	,000
Pearson	737,070	10	,000

Jasno se vidi statistička značajnost statistika koji testira hipotezu o usaglašenosti našeg modela (nezavisnih varijabli) i podataka. Pod uslovom da je nulta hipoteza tačna, verovatnoća da Hi-kvadrat statistik verodostojnosti bude veći ili jednak dobijenom gotovo je jednaka nuli, te na nivou od 0.01 možemo odbaciti nultu hipotezu i reći da ovakav model nije u saglasnosti sa našim podacima. To se može uočiti i na osnovu tabele **Cell Counts and Residuals** gde su u poslednjoj koloni, svi standardizovani reziduali po apsolutnoj vrednosti veći od 1.96!

-U poslednjem delu 2. zadatka traži se da izaberemo NAJJEDNOSTAVNIJI hijerarhijski model koji je u skladu sa podacima. Videli smo iz prethodnog dela zadatka da model koji predviđa nezavisnost varijabli nije saglasan sa našim podacima, te je najbolje da isprobamo sledeći model po složenosti. Najviši model u hijerarhiji bi obuhvatio interakciju trećeg reda, odnosno da su sve 3 varijable u interakciji.

Nakon isprobavanja različitih hijerarhijskih modela dolazimo i do modela koji predviđaju dve asocijacije. Od toga samo je jedan model sa dve interakcije u skladu sa našim podacima!

U okviru kartice **Model**, u Generating Class prebacimo dve asocijacije i to: *pusac\*starost* i *starost\*zive*.

**Goodness-of-Fit Tests**

	Chi-Square	df	Sig.
Likelihood Ratio	5,220	4	,265
Pearson	5,173	4	,270

Značajnost statistika koji testira usaglašenost modela sa našim podacima iznosi 0.265(ne odbacuje se nulta hipoteza koja pretpostavlja da je model u skladu sa našim podacima). Model je izabran kao najjednostavniji, a da je u skladu sa podacima!

### Zadatak 3.

U fajlu **delinkv\_testost\_ses.sav** nalaze se podaci o socioekonomskom statusu (varijabla **SES**), nivou testosterona (varijabla **Testost** – normalan, gornjih 10%), i učešću u delinkventnim radnjama (varijabla **delinkve**) za slučaj an uzorak muškaraca.

- Odabarti najjednostavniji i najpodesniji hijerarhijski loglinearni model koji je u skladu sa podacima.

Prema opisanom postupku za Loglinearni model i nakon ubacivanja I definisanja svih varijabli, u prozoru **Model** se u desni deo prebaci svaka varijabla posebno (bez interakcija). To bi bio prvi model koji testiramo a koji pretpostavlja da nema nikakvih asocijacija među varijablama.

Cell Counts and Residuals

				Observed		Expected		Residuals	Std. Residuals
				Count	%	Count	%		
Socialnoekonomski status	Nizak	Normalan	NE	1104,000	40,4%	1194,152	43,6%	-90,152	-2,609
			DA	190,000	6,9%	151,482	5,5%	38,518	3,130
			Visok (gornjih 10%)		NE	140,000	5,1%	133,438	6,562
					DA	62,000	2,3%	16,927	45,073
Visok	Normalan		NE	1114,000	40,7%	989,805	36,2%	124,195	3,948
			DA	53,000	1,9%	125,560	4,6%	-72,560	-6,475
			Visok (gornjih 10%)		NE	70,000	2,6%	110,604	-40,604
					DA	3,000	,1%	14,030	-11,030

Goodness-of-Fit Tests

	Chi-Square	df	Sig.
Likelihood Ratio	185,825	4	,000
Pearson	218,034	4	,000

Iz **Cell Counts and Residuals** tabele vidimo da je 7 od 8 standardizovanih reziduala po apsolutnoj vrednosti veće od 1.96, a u tabeli **Goodness-of-Fit Tests**, verovatnoća da dobijemo Chi-square kakav smo dobili pod uslovom da je nulta hipoteza o saglasnosti ovakvog modela i naših podataka tačna, gotovo je jednaka nuli! Na nivou od 0.01 možemo odbaciti nultu hipotezu. Model nije u skladu sa našim podacima, tako da moramo isprobati novi.

Došavši na isti način do prozora **Model**, vraćamo sve varijable sa desne na levu stranu i ovog puta ubacujemo njihove međusobne interakcije (ne računajući trostruku interakciju). Dakle, kao po prethodno opisanom postupku iz prethodnog zadatka dugme „Interaction“ promenimo u „All 2-way“, obeležimo sve varijable i prebacimo na desnu stranu njihove interakcije.

Cell Counts and Residuals

				Observed		Expected		Residuals	Std. Residuals
				Count	%	Count	%		
Socialnoekonomski status	Nizak	Normalan	NE	1104,000	40,4%	1100,322	40,2%	3,678	,111
			DA	190,000	6,9%	193,678	7,1%	-3,678	-,264
			Visok (gornjih 10%)		NE	140,000	5,3%	-3,725	-,311
					DA	62,000	2,1%	3,725	,488
Visok	Normalan		NE	1114,000	40,7%	1117,688	40,9%	-3,688	-,110
			DA	53,000	1,9%	49,312	1,8%	3,688	,525
			Visok (gornjih 10%)		NE	70,000	2,4%	3,735	,459
					DA	3,000	,2%	-3,735	-1,439

Goodness-of-Fit Tests

	Chi-Square	df	Sig.
Likelihood Ratio	3,518	1	,061
Pearson	2,986	1	,084

Ovog puta dobijamo model koji je u skladu sa podacima što zaključujemo na osnovu standardizovanih reziduala gde nijedan nije veći po apsolutnoj vrednosti od 1.96 i na osnovu značajnosti Hi-kvadrata(0.061), na osnovu koga ne možemo da odbacimo nultu hipotezu i uslovno rečeno „prihvatamo“ da je dati model, koji pretpostavlja asocijacije

svake varijable sa svakom drugom, ali ne i trostruku asocijaciju u saglasnosti sa našim podacima.

#### Zadatak 4.

U fajlu **stavmat.sav** nalaze se, između ostalog, odgovori na jedno pitanje iz upitnika (varijabla **spm2**) koje su ispitanici dali koristeći petostepenu skalu za odgovaranje.

- Testirajte sledeću pretpostavku: verovatnoće biranja odgovora 1 ili biranja odgovora 5 su jednake i svaka iznosi 0.1, verovatnoće biranja odgovora 2 ili biranje odgovora 4 su jednake i svaka iznosi 0.2, a verovatnoća biranja odgovora 3 je 0.4;

Test sprovodimo preko komande: **Analyze** → **Nonparametric Tests** → **Chi-square**. U **Test Variable List** ubacujemo varijablu za koju testiramo pretpostavku o verovatnoći biranja odgovora, dakle *spm2*. Pošto imamo tačno određene verovatnoće za svaku ocenu koje želimo testirati, u **Expected Values** obeležimo **Values**. U odgovarajući prostor se upisuju redom (po veličini cifre kojom su označeni odgovori, dakle prvo za 1, pa za 2 i tako redom do 5) očekivane verovatnoće 0.1, 0.2, 0.4, 0.2 i 0.1 ili 10, 20, 40, 20, 10 (tada se tretiraju kao procenti). Posle svake verovatnoće obavezno kliknuti **Add** i po završenom ubacivanju verovatnoća **OK**.

#### Frequencies

spm2			
	Observed N	Expected N	Residual
1	2	24,1	-22,1
2	1	48,2	-47,2
3	14	96,4	-82,4
4	78	48,2	29,8
5	146	24,1	121,9
Total	241		

#### Test Statistics

spm2	
Chi-Square	771,925 <sup>a</sup>
df	4
Asymp. Sig.	,000

Nulta hipoteza glasi da će verovatnoće davanja ocena imati raspodelu koju smo uneli, dakle da će po 10% ispitanika davati ocene 1 i 5, po 20 % ocene 2 i 4 i 40 % ispitanika

daće ocenu 3. Pod uslovom da je nulta hipoteza tačna, verovatnoća da dobijemo Hi-kvadrat 771.925 ili veći od toga gotovo je jednaka nuli, tako da na nivou od 0.01 možemo odbaciti nultu hipotezu i reći da naši podaci odstupaju od ovakve raspodele, što se između ostalog vidi i iz opserviranih frekvencija u tabeli **Frequencies** za **spm2**. Jasno se uočava tendencija davanja viših ocena. Što je ocena viša veći je broj onih koji se opredeljuju za tu ocenu. (Bitno je povesti računa da se redosled ubačenih verovatnoća za svaku ocenu poklapa sa redosledom ocena u ispisu, što u našem slučaju potvrđuje tabela sa frekvencijama-redosled ocena je isti kao za raspodelu koju smo testirali).

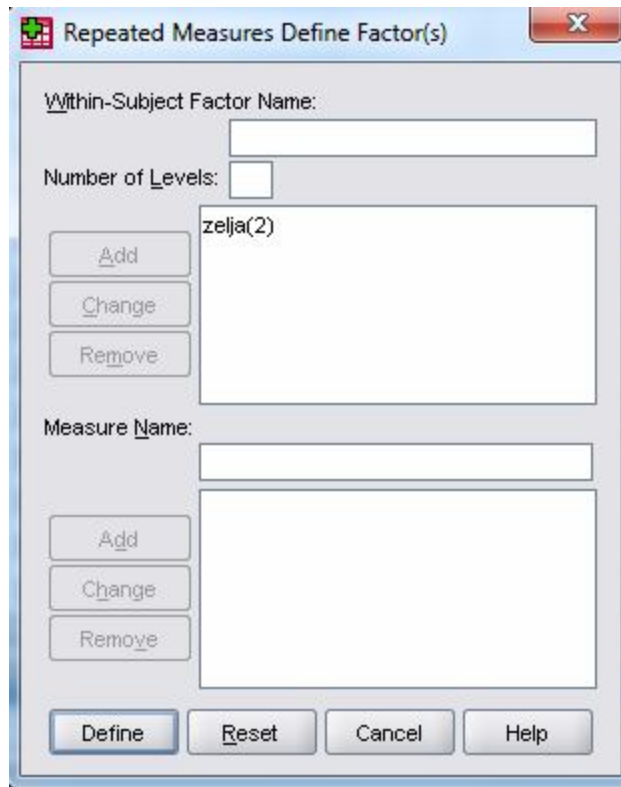
## Zadatak 5

U fajlu **pusenje\_odvikavanje.sav** dati su podaci za iste ispitanike o procenjenoj želji za cigaretom pre (varijabla **pre**) i posle (varijabla **posle**) odvikavanja od pušenja. Odvikavanje od pušenja je izvodjeno na tri različita načina (varijabla **tretman**) pri čemu je ista grupa učesnika izložena samo jednom od tretmana (smanjenje= 1; prekid=2; averzivna terapija = 3).

- a) Ustanoviti postoji li promena u procenjenoj želji za cigaretom pre i posle odvikavanja od pušenja bez obzira na tretman (bez uzimanja u obzir tretmana);
- b) Ustanoviti postoji li efekat tretmana na promenu u procenjenoj želji za cigaretom;
- c) Ustanoviti postoji li promena u procenjenoj želji za cigaretom pre i posle tretmana s obzirom na tretman.

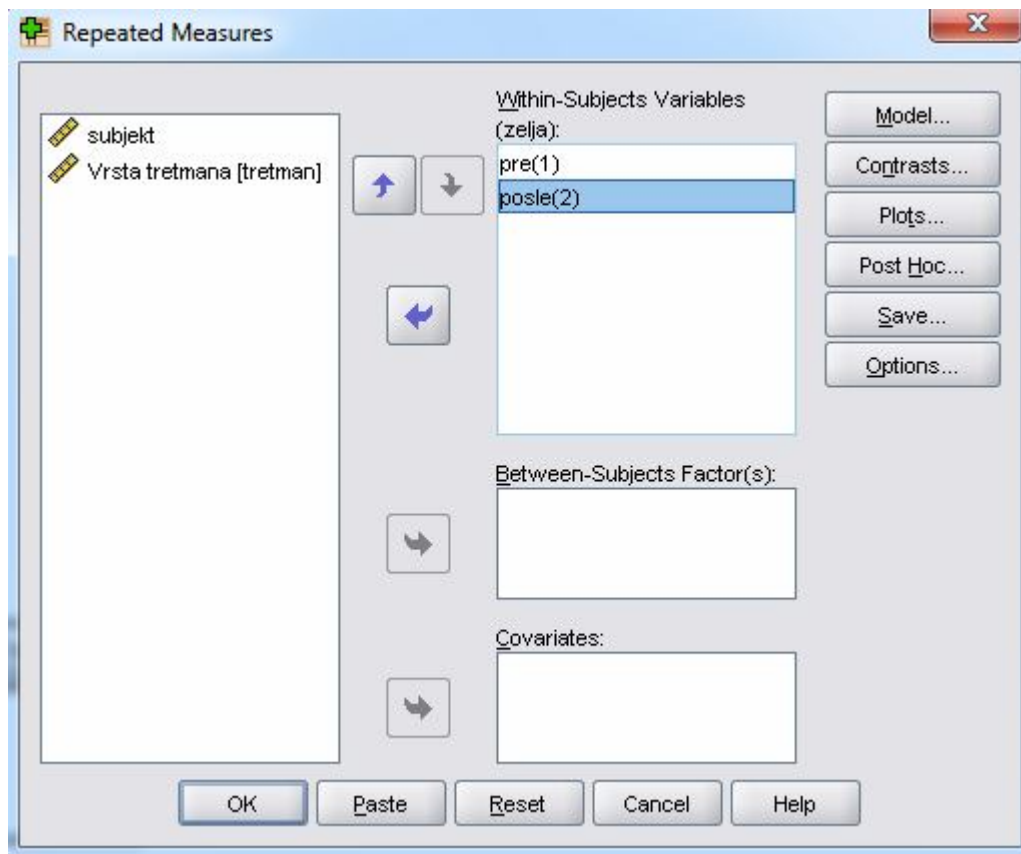
- a) **postoji li promena u procenjenoj želji za cigaretom pre i posle odvikavanja od pušenja bez obzira na tretman (bez uzimanja u obzir tretmana)**

-Za potrebe ovog dela zadatka mogli bismo koristiti t-test za zavisne uzorke (**Paired-samples t-test**) ali ćemo mi ovde u sklopu celokupnog zadatka koristiti analizu varijanse za ponovljena merenja jer su podaci na varijabli *pre* i *posle* podaci istih ispitanika( ocena želje za cigaretom pre i nakon odvikavanja). Idemo na **Analyze → General Linear Model → Repeated Measures**. U okviru polja **Within-Subject Factor Name** imenujemo factor npr. kao “želja”. U **Number of Levels** unesemo broj 2 jer imamo 2 nivoa ove varijable-pre i posle. Kliknemo na **Add**. Zatim idemo na **Define**.

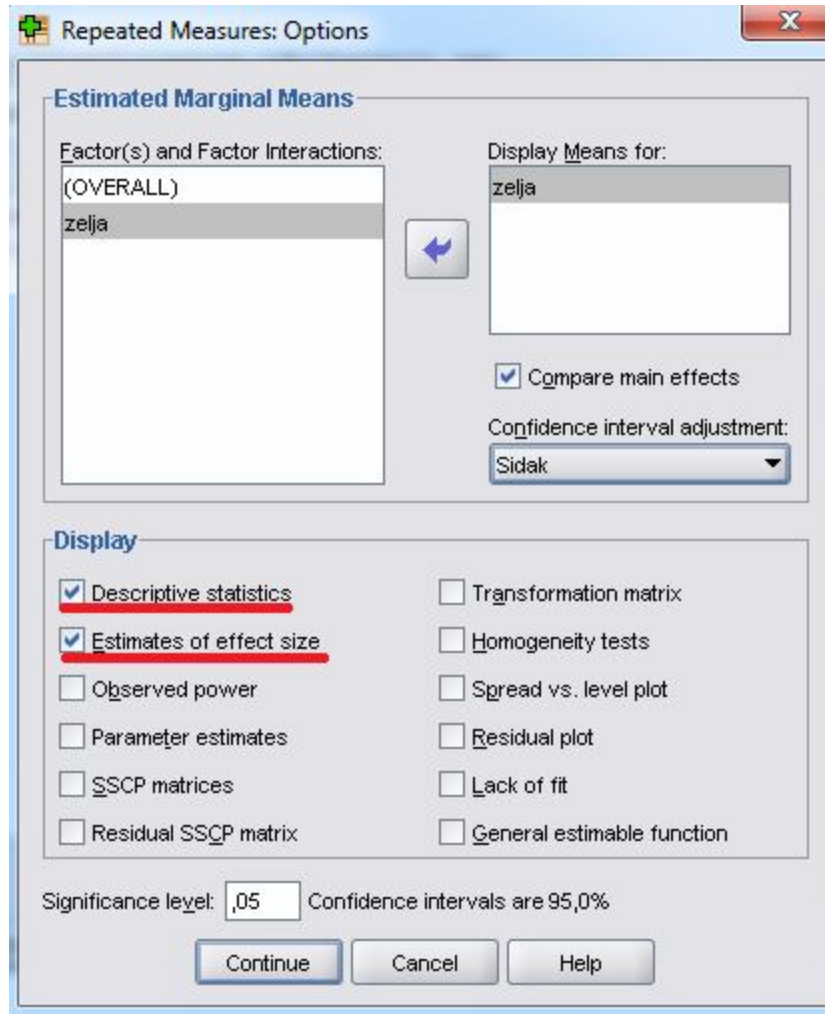


U desni prostor **Within-Subjects Variables** (**zelja**) ubacimo redom varijable sa leve strane: *pre* i *posle*.

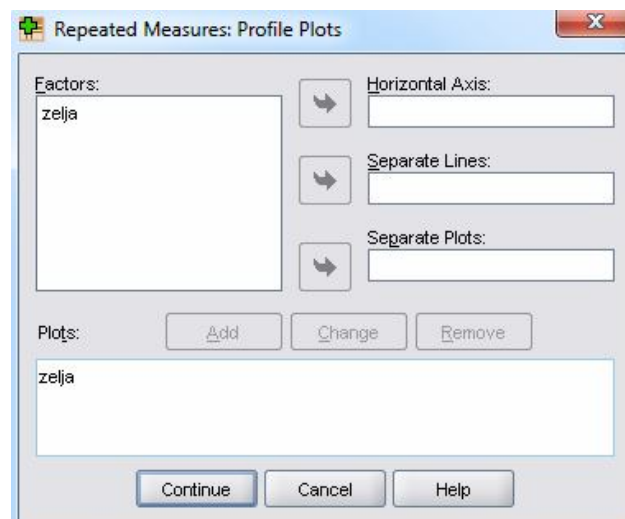




Idemo na karticu **Options**; U okviru **Display** označimo **Descriptive Statistics** i, poželjno, **Estimates of effect size**, a ukoliko želimo poređenje prosečnih vrednosti po nivoima date varijable, odnosno glavni efekat ove varijable po nivoima, u desni prostor **Display Means for** ubacimo varijablu *zelja*, označimo opciju **Compare main effects**, a za opciju **Confidence interval adjustment** izaberemo „Sidak“. (Ovo poslednje nije neophodno kada imamo samo dva nivoa faktora).



Kliknemo na **Continue** i idemo na karticu **Plots** kako bismo dobili i grafički prikaz. Ubacimo faktor „zelja“ u **Horizontal Axis**, kliknemo **Add** i zatim **Continue** i **OK**



U ispisu u tabeli **Tests of Within-Subjects Effects**, gledamo F-statistik i njegovu značajnost za factor “zelja” u prvom redu-Sphericity Assumed:

**Tests of Within-Subjects Effects**

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
zelja	Sphericity Assumed	16,133	1	16,133	32,893	,000	,701
	Greenhouse-Geisser	16,133	1,000	16,133	32,893	,000	,701
	Huynh-Feldt	16,133	1,000	16,133	32,893	,000	,701
	Lower-bound	16,133	1,000	16,133	32,893	,000	,701
Error(zelja)	Sphericity Assumed	6,867	14	,490			
	Greenhouse-Geisser	6,867	14,000	,490			
	Huynh-Feldt	6,867	14,000	,490			
	Lower-bound	6,867	14,000	,490			

Nulta hipoteza bi glasila da nema razlika na nivoima faktora “zelja”, tj. da nema efekta odvikavanja od pušenja na promenu želje za cigaretom i da su prosečne procene želje za cigaretom iste na oba nivoa faktora. Međutim verovatnoća da dobijemo F-statistik veći ili jednak onome koji smo dobili na uzorku, pod uslovom da je nulta hipoteza tačna, gotovo je jednaka nuli. Prema tome, odbacujemo nultu hipotezu na nivou značajnosti od 0.01 i kažemo da efekat faktora “zelja” postoji, odnosno da se prosečne procene želje razlikuju po nivoima datog faktora!

Postoji promena u želji za cigaretom pre i posle odvikavanja od pušenja, ne uzimajući u analizu efekat tretmana!

**Pairwise Comparisons**

Measure: MEASURE\_1

(I) zelja	(J) zelja	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
					Lower Bound	Upper Bound
1	2	1,467 <sup>*</sup>	,256	,000	,918	2,015
2	1	-1,467 <sup>*</sup>	,256	,000	-2,015	-,918

Based on estimated marginal means

\*. The mean difference is significant at the ,05 level.

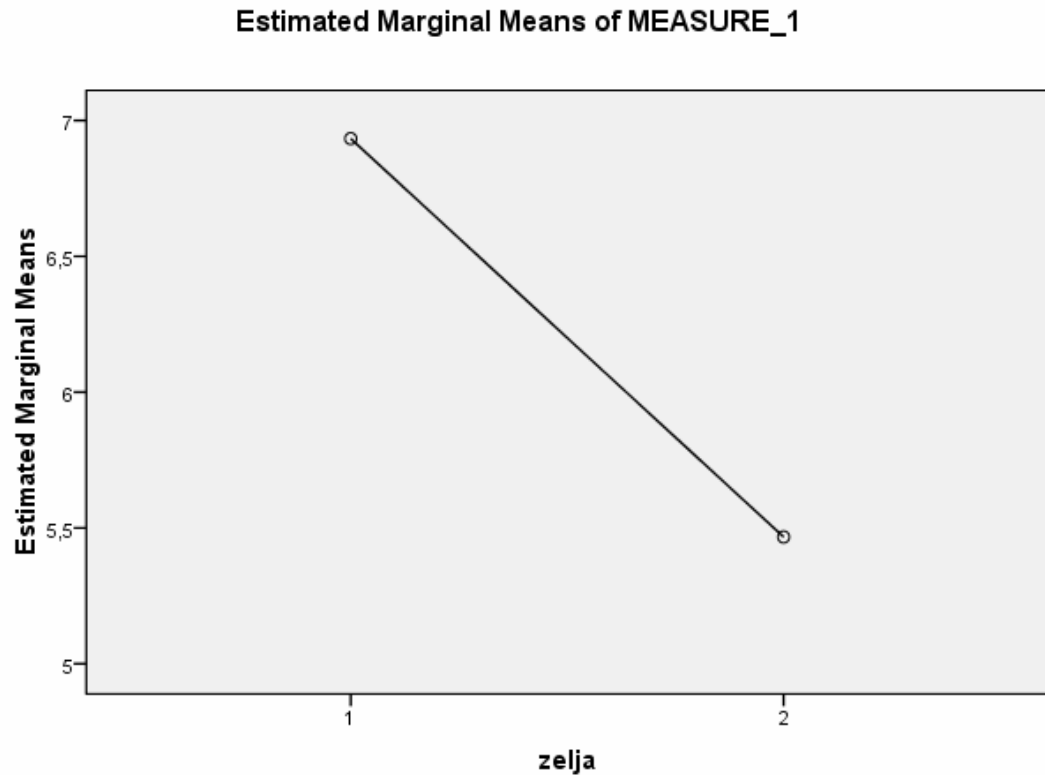
a. Adjustment for multiple comparisons: Sidak.

Na osnovu Šidakovog testa, vidimo da postoje značajne razlike u prosecima između dva nivoa ove varijable.

NAPOMENA: Ova analiza nam u suštini nije potrebna (redundantna je) jer imamo samo dva nivoa, ali je korisna kada postoje više od dva nivoa faktora.

Potvrdu dobijamo i na osnovu izgleda grafika koji dobijamo u ispisu:

## Profile Plots



Vidimo da je zelja za cigaretom znatno veća pre(1) nego posle(2) odvikavanja.

### b) postoji li efekat tretmana na promenu u procenjenoj želji za cigaretom

-U okviru komande **Compute**, treba da napravimo novu varijablu *promena* i to kao razliku varijabli *posle* i *pre*. U okviru **Numeric Expression** u komandi **Compute** prebacimo iz spiska varijabli varijablu *posle*, a zatim od nje oduzmemo varijablu *pre* koju na isti način prebacimo.

Kada smo dobili novu varijablu, efekat tretmana na promenu testiramo preko **One-Way ANOV-e**. U Dependent List prebacujemo varijablu *promena*, a u Faktor stavljamo vrstu tretmana. U okviru kartice **Options** označimo Descriptive i Homogeneity of variance test.

U ispisu dobijemo:

### Test of Homogeneity of Variances

promena1

Levene Statistic	df1	df2	Sig.
,400	2	12	,679

### ANOVA

promena1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4,133	2	2,067	2,583	,117
Within Groups	9,600	12	,800		
Total	13,733	14			

Leveneov statistik nije značajan, tako da je uslov homoscedastičnosti ispunjen.

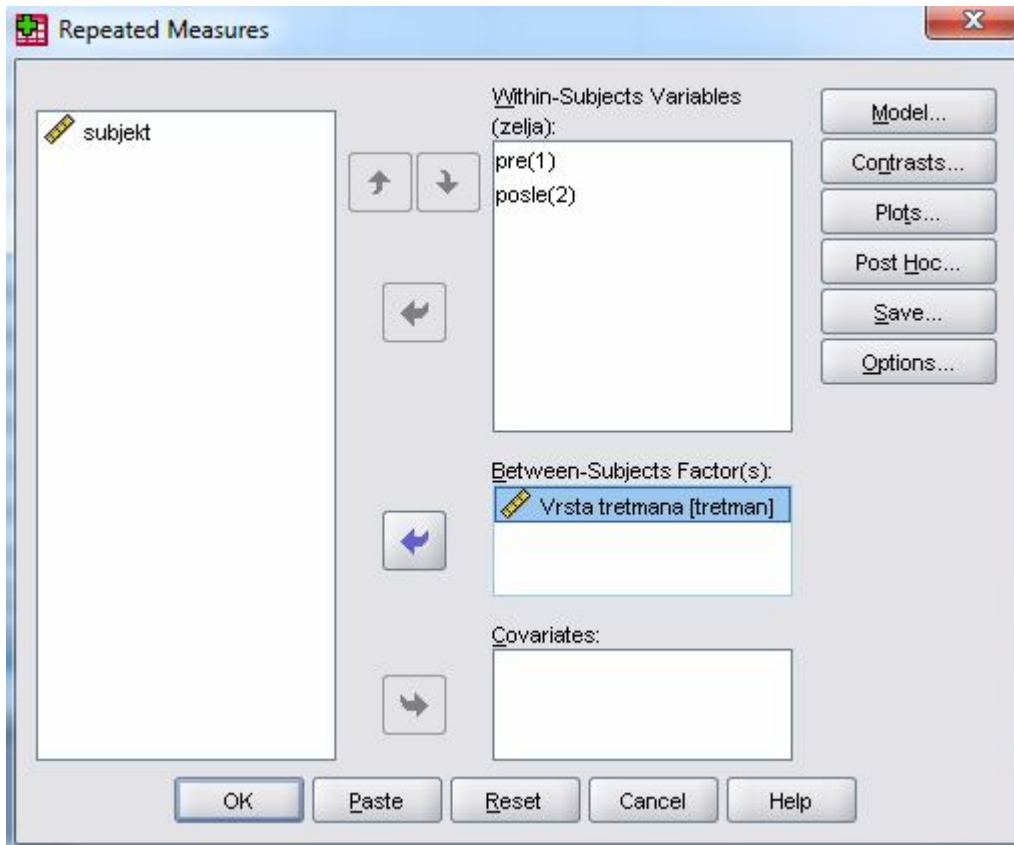
F-statistik za efekat tretmana iznosi 2.583 i nije statistički značajan ( $p=0.117$ ).

Dakle, možemo zaključiti da ne postoji efekat tretmana na promenu u procenjenoj želji za cigaretom i da su razlike u proceni nastale slučajno.

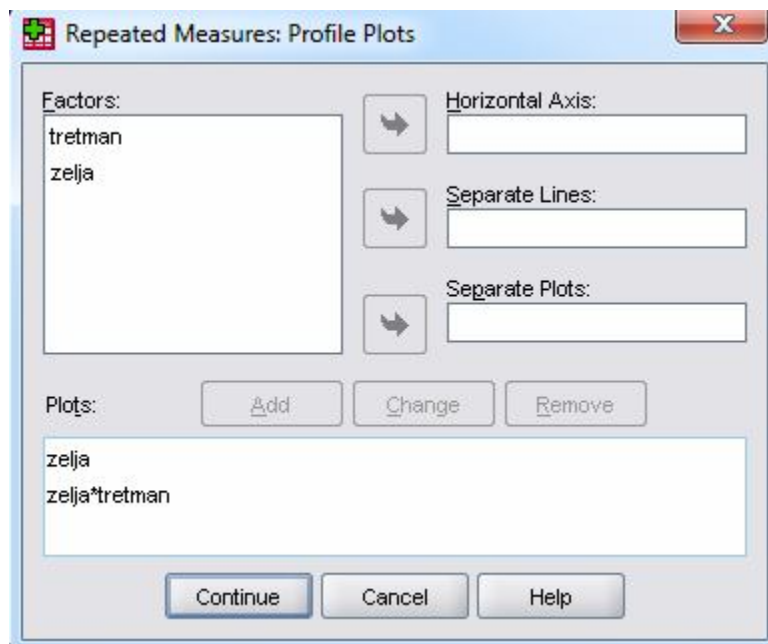
### c) postoji li promena u procenjenoj želji za cigaretom s obzirom na tretman

-Za testiranje veličine promene zelje za cigaretom s obzirom na tretman koristimo istu komandu i iste korake kao pod a) dok ne dodjemo do prozora gde definišemo nivoe navedenog faktora (koje je potpuno isto kao pod a).

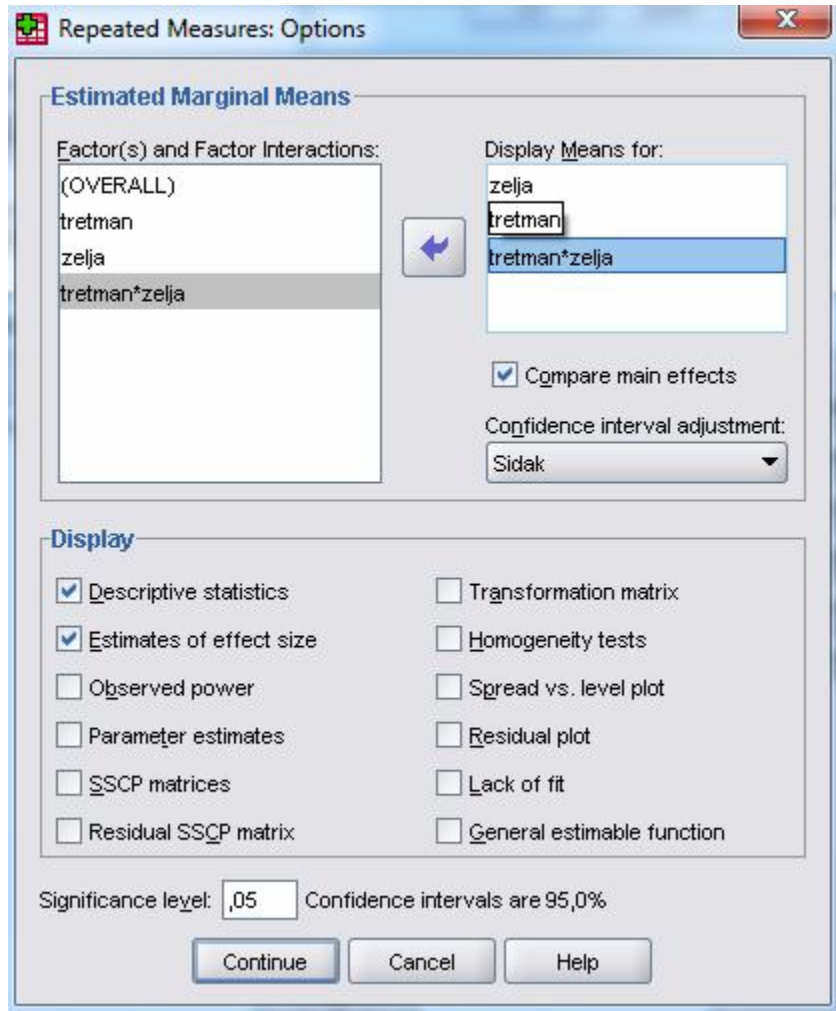
U **Between-Subjects Factor(s)** ubacimo varijablu *tretman*:



Ako odemo na karticu **Plots** “zelju” ubacujemo u **Horizontal Axis**, a “tretman” u **Separate Lines**, kliknemo na **Add I Continue**.



U okviru kartice **Options** označene stavke ostaju iste kao prethodne, eventualno u okviru **Display Means for** ubacimo i “tretman” i interakciju tretmana i zelje kao što je prikazano:



U ispisu u tabeli:

**Tests of Within-Subjects Effects**

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
zelja	Sphericity Assumed	16,133	1	16,133	40,333	,000	,771
	Greenhouse-Geisser	16,133	1,000	16,133	40,333	,000	,771
	Huynh-Feldt	16,133	1,000	16,133	40,333	,000	,771
	Lower-bound	16,133	1,000	16,133	40,333	,000	,771
zelja * tretman	Sphericity Assumed	2,067	2	1,033	<u>2,583</u>	<u>,117</u>	,301
	Greenhouse-Geisser	2,067	2,000	1,033	2,583	,117	,301
	Huynh-Feldt	2,067	2,000	1,033	2,583	,117	,301
	Lower-bound	2,067	2,000	1,033	2,583	,117	,301
Error(zelja)	Sphericity Assumed	4,800	12	,400			
	Greenhouse-Geisser	4,800	12,000	,400			
	Huynh-Feldt	4,800	12,000	,400			
	Lower-bound	4,800	12,000	,400			

vidimo da promena zelje pre i posle odvikavanja s obzirom na vrstu tretmana nije statistički značajna. Dakle nema interakcije između zelje pre i posle odvikavanja i vrste tretmana.

Glavni efekat varijable “tretman” se gleda u tabeli:

**Tests of Between-Subjects Effects**

Measure: MEASURE\_1  
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	1153,200	1	1153,200	427,111	,000	,973
<u>tretman</u>	1,400	2	,700	<u>,259</u>	<u>,776</u>	,041
Error	32,400	12	2,700			

Odavde vidimo da ne postoji glavni efekat tretmana (nije statistički značajan)-nema razlike na različitim nivoima ove varijable.

## Zadatak 6.

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

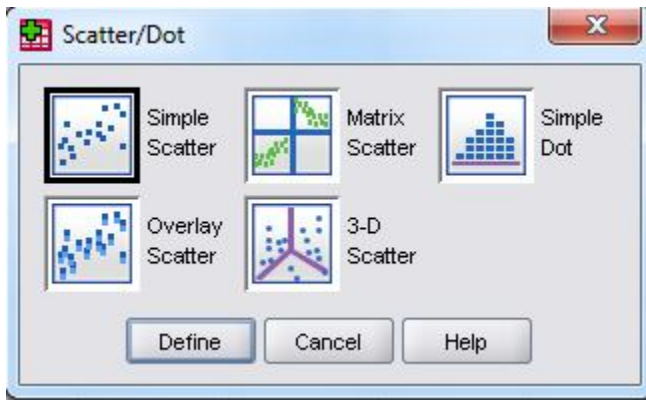
- Grafički prikazati vezu između ovih varijabli
- Izračunati kovarijansu između motivisanosti i uspeha;
- Izračunati koeficijent linearne korelacije između motivisanosti i uspeha;
- Testirati statističku značajnost koeficijenta linearne korelacije.

Šta se na osnovu koeficijenta linearne korelacije i njegove statističke značajnosti može zaključiti?

-Za grafički prikaz veze ove dve varijable koristimo sledeću komandu:

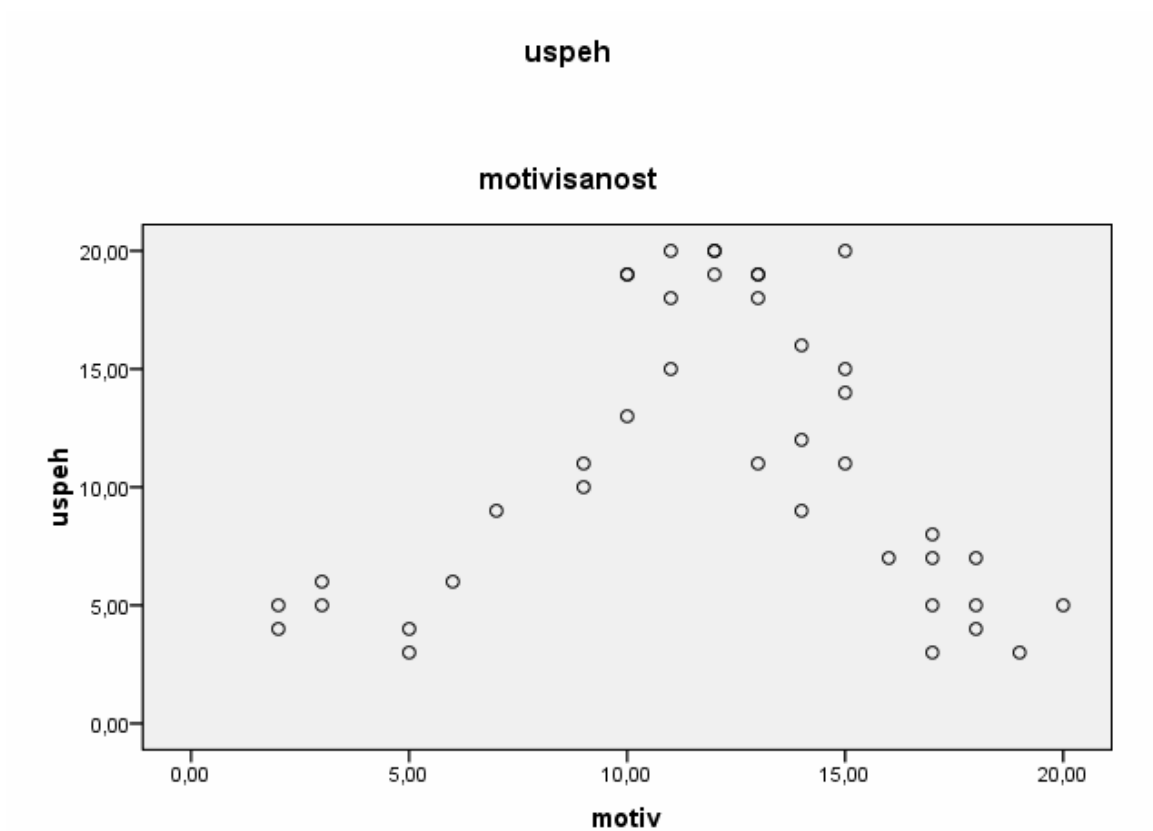
**Graphs** → **Legacy Dialogs** → **Scatter/Dot**. Obeležimo Simple Scatter kao što je prikazano:





i idemo na **Define**.

U Y Axis ubacimo „uspeh“, u X Axis ubacimo „motiv“, a u kartici **Titles** u Line 1 i Line 2 upišemo redom npr. „uspeh“ i „motivisanost“ radi preglednosti grafika. Dobijamo sledeće:

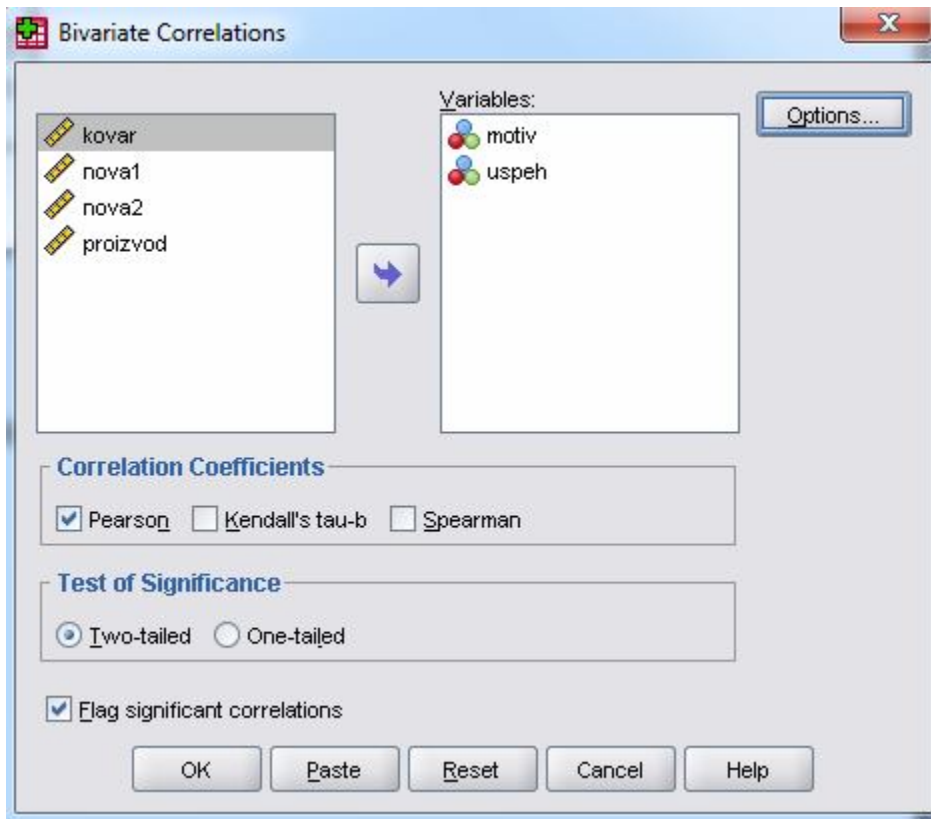


Uočavamo nelinearnost veze uspeha i motivacije. Grafik podseća na obrnuto-U krivu, gde se najveći uspeh postiže pri srednjim vrednostima motivacije, dok slaba ili prevelika motivacija dovode do lošeg uspeha.

-Koefficient linearne korelacije, kao i kovarijansu motivacije i uspeha možemo dobiti preko:

Analyze → Correlate → Bivariate.

U **Variables** ubacimo varijable za koje računamo korelaciju I kovarijansu.



U okviru kartice **Options** označimo **Cross-product deviations and covariances**.

U ispisu dobijemo:

**Correlations**

		motiv	uspeh
motiv	<b>Pearson Correlation</b>	1	<b>,062</b>
	Sig. (2-tailed)		,702
	Sum of Squares and Cross-products	942,975	72,900
	<b>Covariance</b>	24,179	<b>1,869</b>
	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

Koeficijent linearne korelacije je 0.062, a kovarijansa je 1.869.

-Vidimo da je statistička značajnost koeficijenta linearne korelacije 0.702, pa možemo reći da među ovim varijablama nema linearne povezanosti. Dakle nulta hipoteza bi glasila da između ove dve varijable nema linearne povezanosti. Verovatnoća da se dobije koeficijent lin. korelacije veći ili jednak dobijenom na našem uzorku je 0.702 što je daleko iznad 0.05, tako da ne odbacujemo nultu hipotezu!

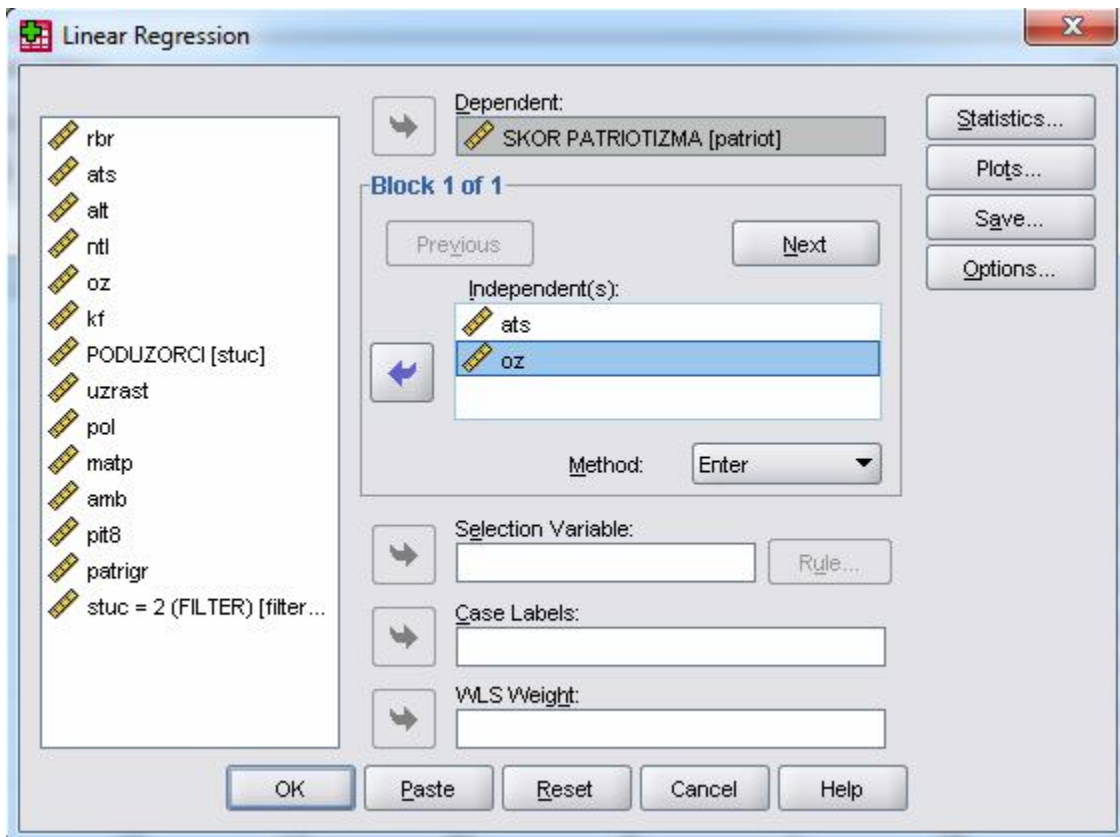
Šta nam ovo govori? Između motiva i uspeha nema **linearne** povezanosti, ali to ne znači da nema nikakve povezanosti! Sa grafika smo videli da rezultati imaju tendenciju grupisanja u vidu **nelinearne** funkcije, pa nam testiranje **linearne** korelacije zapravo ne ukazuje na povezanost.

## Zadatak 7.

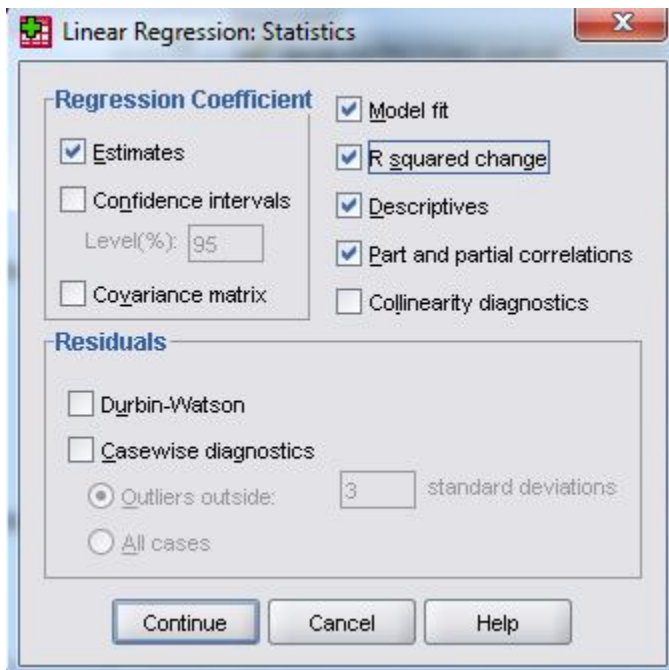
U fajlu **patrioti.sav** nalaze se, između ostalog, podaci o patriotizmu (varijabla **patriot**), autoritarnosti (varijabla **ats**) i stepenu nacionalne vezanosti (varijabla **oz**) slučajnog uzorka studenata.

- Napraviti model za predviđanje patriotizma na osnovu autoritarnosti i nacionalne vezanosti uzetih zajedno.
- Utvrditi koliki je dodatni doprinos nacionalne vezanosti objašnjenju individualnih razlika (varijabilnosti) patriotizma kada se već uzme u obzir doprinos autoritarnosti objašnjenju individualnih razlika (varijabilnosti) patriotizma studenata;

-Do predviđanja dolazimo preko **Analyze → Regression → Linear**. U **Dependent** stavljamo varijablu *patriot* koju predviđamo. U **Independent(s)** stavljamo prediktorske varijable *ats* i *oz*.



U okviru opcije **Statistics** obeležiti R squared change, Descriptives i eventualno Part and Partial correlations.



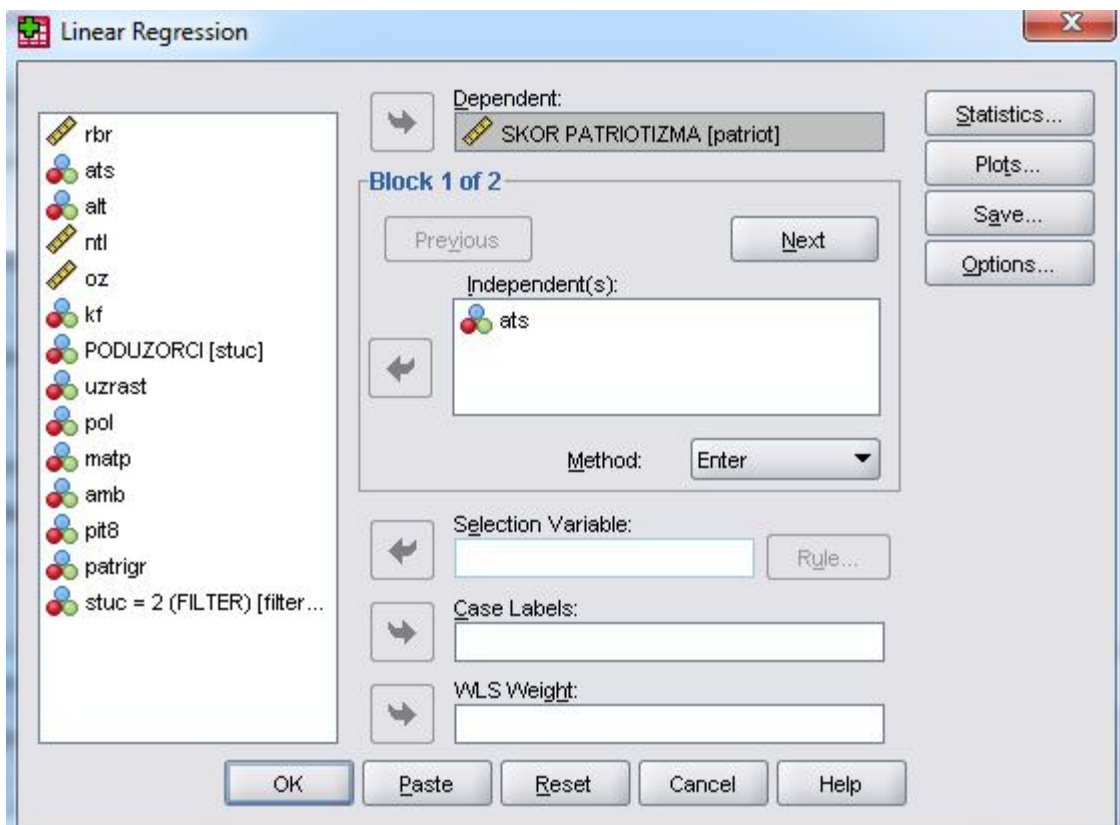
Model Summary

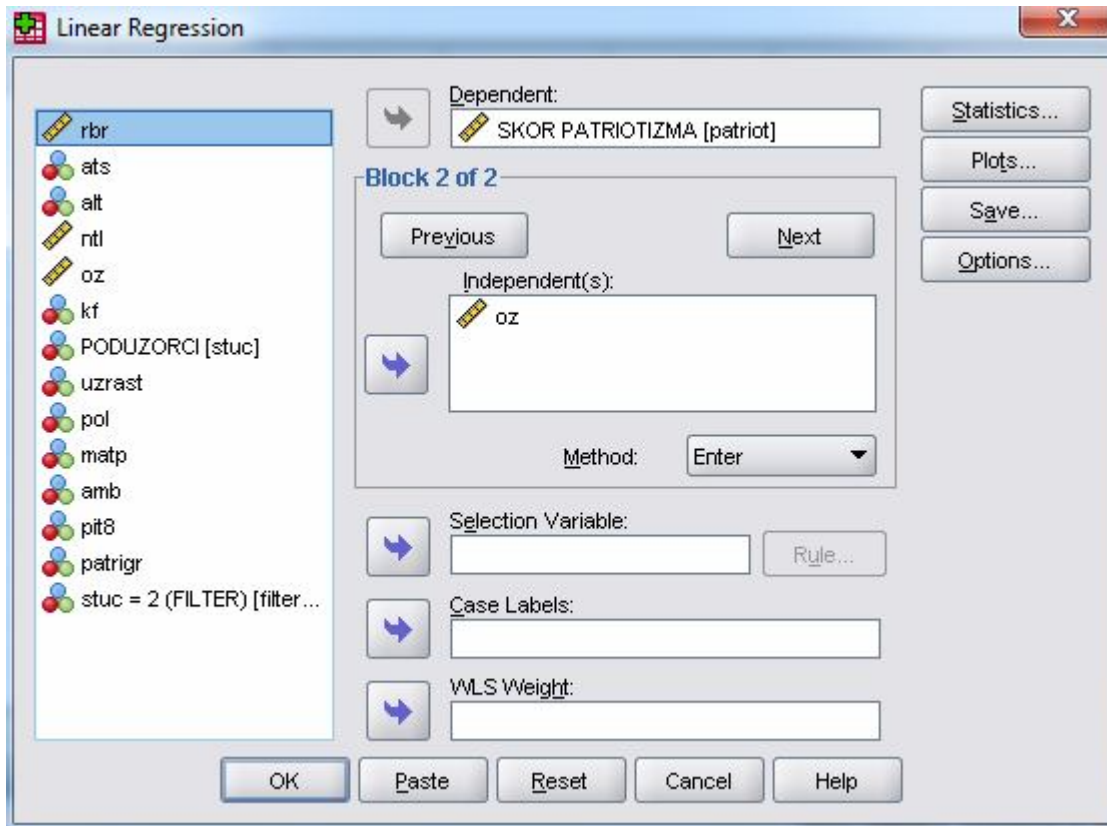
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,483 <sup>a</sup>	,234	,232	8,49590	,234	129,232	2	848	,000

a. Predictors: (Constant), oz, ats

Na osnovu ovog modela, može se predvideti 23.4 % varijabilnosti na UZORKU i 23.2 % varijabilnosti u POPULACIJI (prilagođeni R-kvadrat).

-Došavši do linearne regresije, prema opisanom postupku, isto kao i pre kao zavisna (kriterijumska) varijabla stavlja se *patriot*, dok se u prediktore stavlja *ats*. Zatim se ide na **Next** i u bloku 2 u prediktorske varijable se ubaci *oz*.





U okviru opcije Statistics uključiti isto što i u prethodnom zadatku.

Dodatni doprinos ove varijable vidi se iz sledeće tabele:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,329 <sup>a</sup>	,109	,107	9,15745
2	,483 <sup>b</sup>	,234	,232	8,49590

a. Predictors: (Constant), ats

b. Predictors: (Constant), ats, oz

Dok je uključena samo autoritarnost u model, može se objasniti 10.7 % varijabilnosti na uzorku. Kada se uključi i nacionalna vezanost taj procenat je 23.2. Dakle, 23.2-10.7 je 12.5 % varijabilnosti kao dodatni doprinos nacionalne vezanosti objašnjenju varijanse patriotizma.

Dodatni doprinos ove varijable vidi se i na osnovu semi-parcijalne korelacije nacionalne vezanosti i patriotizma kada se kontroliše autoritarnost:

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	67,726	2,100		32,257	,000			
	ats	,359	,059	,195	6,049	,000	,329	,203	,182
	oz	-,549	,047	-,378	-11,763	,000	-,448	-,375	-,354

a. Dependent Variable: SKOR PATRIOTIZMA

Semi-parcijalna korelacija nacionalne vezanosti iznosi -0.354, dakle dodatni doprinos u predviđanju patriotizma na osnovu nacionalne vezanosti izračunava se kao kvadrat ove korelacije:

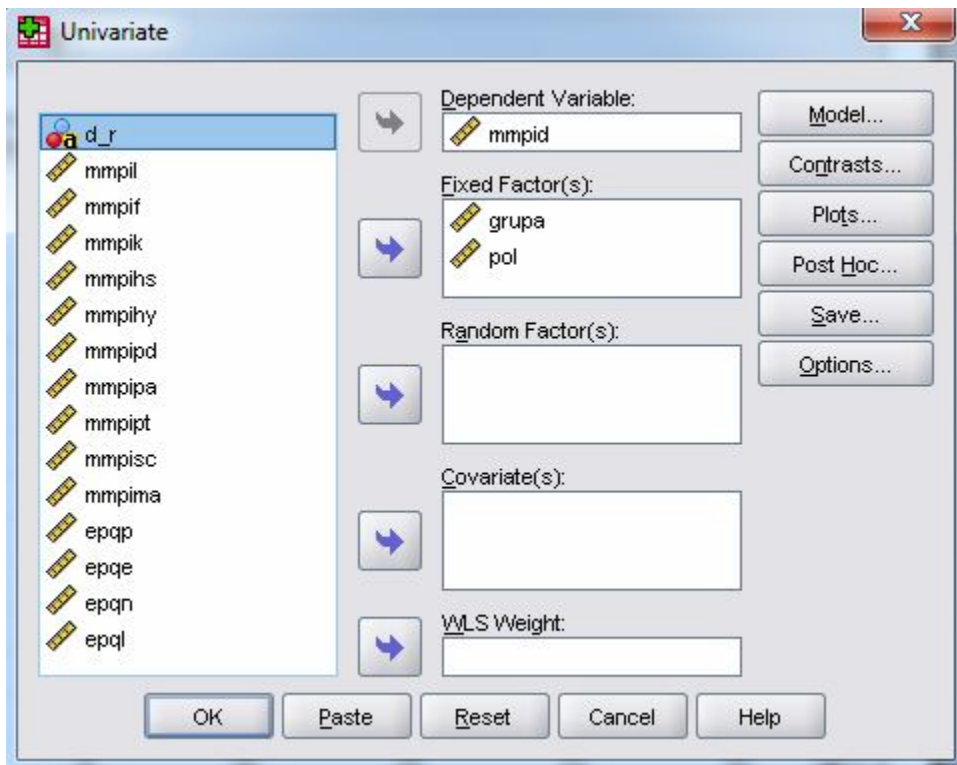
$(-0.354)^2 = 0.125$ . Množeći sa 100 dobijamo 12.5 % dodatno objašnjene varijabilnosti.

### Zadatak 8.

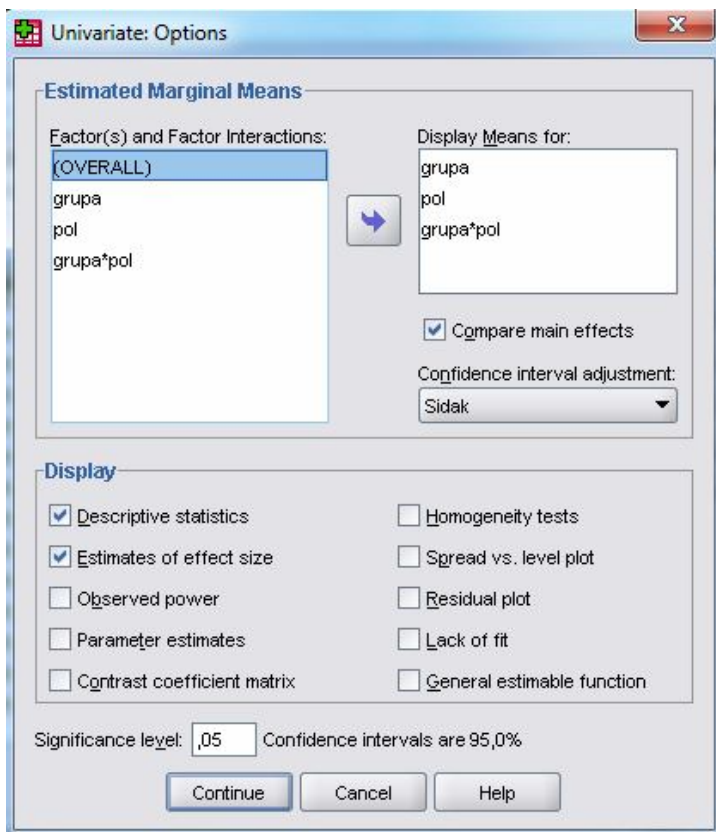
U fajlu **astma.sav** postoje podaci o izraženosti depresivnosti kao crte ličnosti (varijabla **MMPID**) koji su dobijeni na slučajnim uzorcima astmatičnih i neastmatičnih psihosomatskih bolesnika, i to na podjednakom broju muškaraca i žena u svakoj grupi (varijabla **GRUPA**: 1. astmatičari, 2. neastmatičari, varijabla **POL**: 1. muški, 2. ženski).

- Ustanoviti da li se astmatičari razlikuju od neastmatičnih psihosomatskih bolesnika u pogledu prosečne depresivnosti i, ako se razlikuju, da li je ova razlika ista kod muškaraca i žena ili, pak, zavisi od polne pripadnosti?
- Prikazati na istom grafiku aritmetičke sredine muškaraca i žena na depresivnosti (varijabla **MMPID**) odvojeno za grupe astmatičara i neastmatičara

-Koristimo dvofaktorijalnu univarijacionu analizu varijanse. Dakle, idemo na **Analyze** → **General Linear Model** → **Univariate**. Kao Dependent Variable se ubaci *mmpid*, a kao Fixed Factor(s)-grupa i *pol*.



U kartici **Options** se označi:





U kartici **Plots** kako bismo napravili grafik grupa se ubaci u Horizontal Axis, a pol kao Separate Lines. Klikne se na **Add** i zatim **Continue**, pa **OK**.

U ispisu u okviru tabele:

**Tests of Between-Subjects Effects**

Dependent Variable:mmpid

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3823,297 <sup>a</sup>	3	1274,432	69,395	,000	,452
Intercept	87098,766	1	87098,766	4742,693	,000	,950
grupa	3585,016	1	3585,016	195,211	,000	,437
pol	221,266	1	221,266	12,048	,001	,046
grupa * pol	17,016	1	17,016	,927	,337	,004
Error	4627,938	252	18,365			
Total	95550,000	256				
Corrected Total	8451,234	255				

a. R Squared = ,452 (Adjusted R Squared = ,446)

vidi se da postoje glavni efekti pola i grupe, odnosno da se prosečne vrednosti depresivnosti značajno razlikujuna na nivoima ovih varijabli. Medjutim interakcije pola i grupe nema. Dakle efekat faktora grupa je isti i kod muškaraca i kod žena. Iz aritmetičkih sredina u sledećoj tabeli i iz grafika koji sledi posle nje vidimo da su astmaticari prosečno depresivniji od neastmaticara i da to važi kako za muškarce, tako i za žene.

**3. grupa \* pol**

Dependent Variable:mmpid

grupa	pol	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
astmaticari	M	21,000	,536	19,945	22,055
	Z	23,375	,536	22,320	24,430
neastmaticari	M	14,031	,536	12,976	15,086
	Z	15,375	,536	14,320	16,430

-Grafik:

Estimated Marginal Means of mmpid

