

Hi-kvadrat test: zadaci sa rešenjima

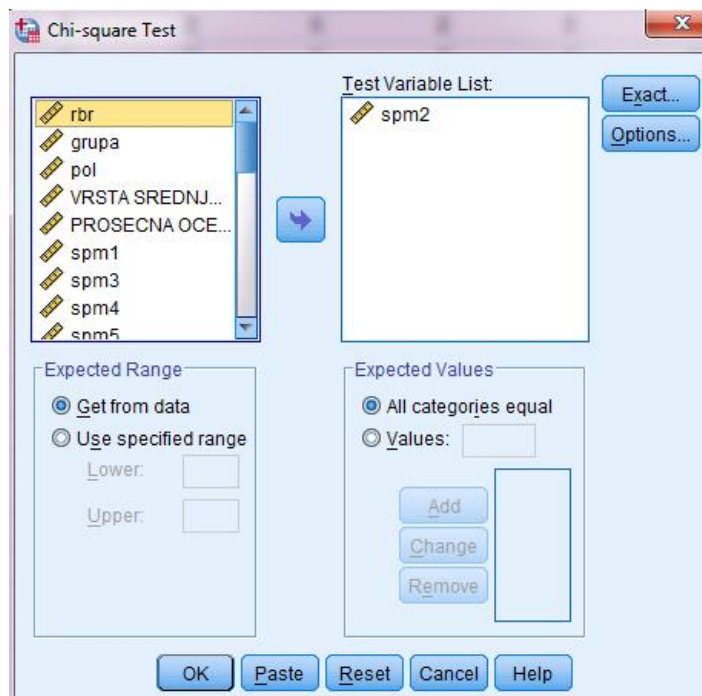
Rešenja uradila Marija Balovi , studentkinja psihologije

Zadatak 1.

U fajlu **stavmat.sav** nalaze se, izme u ostalog, odgovori na jedno pitanje iz upitnika za ispitivanje stava prema matematici (varijabla **spm2**) koje su ispitanici dali koriste i petostepenu skalu za odgovaranje.

- Testirajte pretpostavku da su ispitanici nasumi no odgovarali na ovo pitanje;
- Testirajte slede u pretpostavku: verovatno e biranja odgovora 1 ili biranja odgovora 5 su jednake i svaka iznosi 0.1, verovatno e biranja odgovora 2 ili biranja odgovora 4 su jednake i svaka iznosi 0.2, a verovatno a biranja odgovora 3 je 0.4;

Za testiranje pretpostavke da su ispitanici nasumi no odgovarali na ovo pitanje, koristimo *hi-kvadrat test za testiranje pretpostavke o distribuciji verovatno a jedne kategori ke varijable*. Dakle, odabrati meni **Analyze / Legacy Dialogs/ Nonparametric tests / Chi-square**. Varijablu **spm2** ubacujemo u okvir **Test Variable list**, a u okviru **Expected Values** ostavljamo uklju enu opciju **All categories equal**, jer testiramo nultu hipotezu da su ispitanici odgovarali na pitanje nasumi no, što zna i da je distribucija odgovora uniformna (verovatno a svakog mogu eg odgovora na pitanje jednaka 1/5, budu i da ima 5 mogu ih odgovora na pitanje).



U ispisu dobijamo tabele sa imenom varijable **spm2** i **Test Statistics**. U prvoj tabel vidimo opažene (**Observed N**) i očekivane (**Expected N**) frekvencije, kao i njihove razlike, tj. rezidualne. Iz tabele **Test Statistics** vidimo da hi-kvadrat statistik iznosi 331.635 i da je statistički značajan, zbog čega možemo odbaciti nultu hipotezu i zaključiti da ispitanici nisu nasumično odgovarali na pitanje.

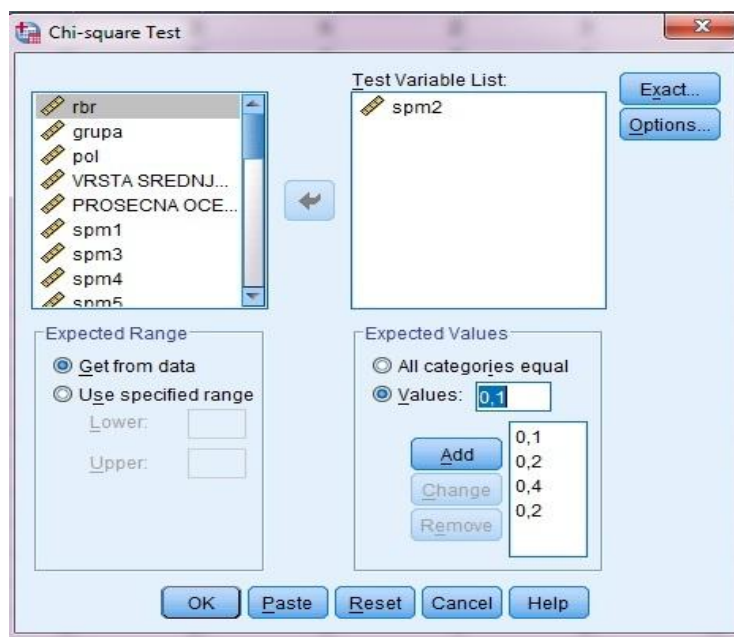
spm2			
	Observed N	Expected N	Residual
1	2	48,2	-46,2
2	1	48,2	-47,2
3	14	48,2	-34,2
4	78	48,2	29,8
5	146	48,2	97,8
Total	241		

Test Statistics

	spm2
Chi-Square	331,635 ^a
df	4
Asymp. Sig.	,000

a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 48,2.

Postupak za testiranje druge pretpostavke se razlikuje samo po tome što u okviru **Expected Values**, umesto **All categories equal**, treba štriklirati **Values**, a zatim redom unositi vrednosti 0.1 , 0.2 , 0.4 , 0.2 , 0.1 tako što posle svake vrednosti treba kliknuti na **Add** da bi se vrednost upisala u prazan okvir u desnom uglu.



Ispis iz procedure izgleda ovako:

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 ^a
df	4
Asymp. Sig.	,000

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

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 24,1.

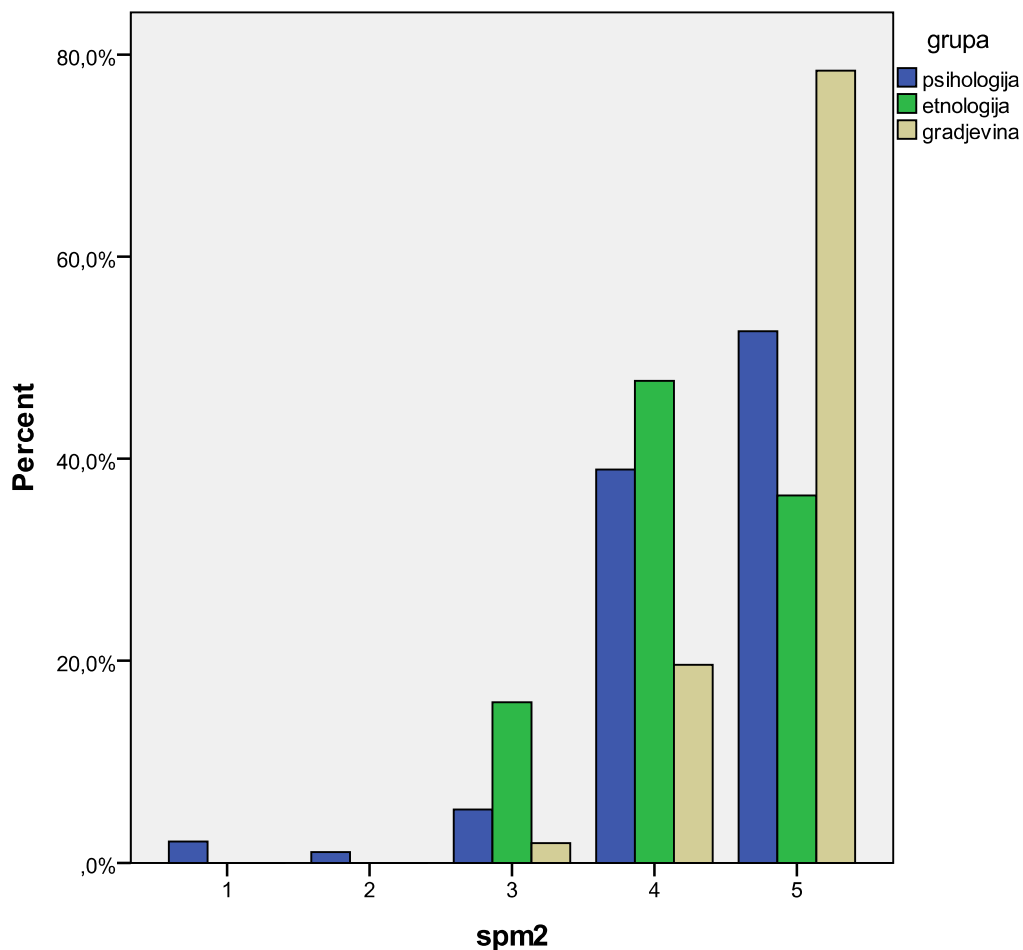
U ispisu vidimo da je hi-kvadrat statistika značajna jer podaci ne podržavaju ni drugu pretpostavku.

Zadatak 2.

Podaci isti kao za zadatak 1.

Grafik prikazati raspodelu odgovora na pitanje iz upitnika (varijabla **spm2**) za svaku od grupa studenata (varijabla **grupa**) na istom grafiku;

- Izabrati **Graphs/Legacy Dialogs/Bar/Clustered**, nakon toga u okviru prozora **Define Clustered Bar** varijablu **spm2** treba ubaciti u odeljak **Category Axis**, a varijablu **grupa** u odeljak **Define Clusters by**. U okviru **Bars Represent** najbolje je uključiti opciju **% of cases** kako bi štapići na štapićastom dijagramu predstavili relativne frekvencije u procentima.



Na dobijenom grafiku se može, između ostalog, uočiti da je relativno najveći procenat studenata etnologije odabrao odgovor pod rednim brojem 4 i 3, dok se relativno najveći procenat studenata psihologije i gradjevine odlučio za ponudjeni odgovor 5. Inače, studenti gradjevine su generalno preferirali odgovor 5 jer je čak oko 80% njih odgovorilo na taj način. Odgovor 1 i 2 se pojavljuje samo kod studenata psihologije i to u veoma retko.

Zadatak 3.

Fajl **estrogen_alhajm.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 **alhajmb** da li su oboljele 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 koliki su šansi 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.

Šanse računamo iz kontingencijske tabele koju dobijamo preko menija **Analyze/Descriptive Statistics/Crosstabs** pa potom u okvir **Row(s)** ubaciti varijablu estrogen a u **Column(s)** varijablu **Alchajmb**.

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

- Šanse da se oboli od Alchajmerove bolesti za žene koje su primale estrogen dobijamo koliki su frekvencija **primalaObolela/primalaNeobolela** ili 9/147 što je 0,061.
- Šanse da se oboli od Alchajmera za žene koje nisu primale estrogen dobijamo koliki su frekvencija **nijeprimalaObolela/nijeprimalaNeobolela** ili 158/810 što je 0.195.
- Deljenjem šansi **da se oboli od Alchajmerove bolesti ako se ne uzima estrogen** sa **šansama da se oboli od te bolesti ako se uzima estrogen** dobijamo koliki su šansi oboljevanja od Alchajmerove bolesti za žene koje nisu primale estrogen u odnosu na žene koje jesu primale estrogen. Taj koliki su šansi iznosi 3.196, što znači da su šanse da se oboli ako se ne uzima estrogen tokom menopauze oko 3 puta veće nego ako se uzima estrogen.
- Za testiranje statističke nezavisnosti možemo upotrebiti hi-kvadrat test: Odabrati **Analyze/Descriptive Statistics/Crosstabs** pa potom u okvir **Row(s)** ubaciti varijablu estrogen a u **Column(s)** varijablu **Alchajmb**. Zatim klikom na dugme **Statistics** štriklirati opcije **Chi-square**, **Contingency coefficient** i **Phi and Cramer's V**. Posle klika na **Continue** kliknuti na dugme **Cells** pa štriklirati u okviru **Percentages** opcije **Raw** i **Column**. Definiciju procedure treba završiti sa **Continue** i **Ok**.

U ispisu iz procedure ključne su sledeće 3 tabele:

1. Tabela kontingencije

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			
		ne	da	Total	
Da li je primala estrogene tokom menopauze	ne	Count	810	158	968
		% within Da li je primala estrogene tokom menopauze	83,7%	16,3%	100,0%
		% within Da li je tokom 5 godina obolela od Alchajmerove bolesti	84,6%	94,6%	86,1%
	da	Count	147	9	156
		% within Da li je primala estrogene tokom menopauze	94,2%	5,8%	100,0%
		% within Da li je tokom 5 godina obolela od Alchajmerove bolesti	15,4%	5,4%	13,9%
Total		Count	957	167	1124
		% within Da li je primala estrogene tokom menopauze	85,1%	14,9%	100,0%
		% within Da li je tokom 5 godina obolela od Alchajmerove bolesti	100,0%	100,0%	100,0%

Iz Tabele **Chi Square tests** vidimo da je Pirsonov hi-kvadrat statistik 11.828 i statistički je značajan. Prema tome, nultu hipotezu prema kojoj su varijable statistički nezavisne odbacujemo i kažemo da postoji veza ili asocijacija između ovih varijabli u populaciji.

2. Tabela sa značajnošću u hi-kvadrata

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	11,828 ^a	1	,001		
Continuity Correction ^b	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

3. Tabela sa koeficijentima asocijacije

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	

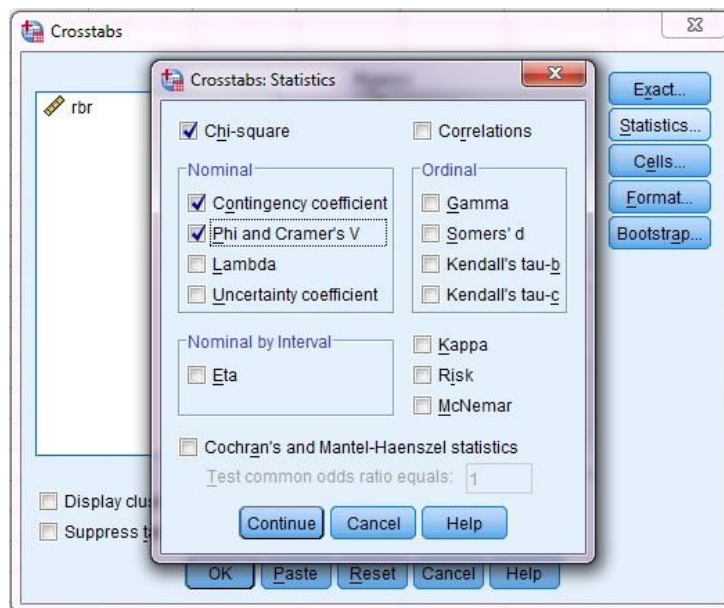
Iz tabele **Symmetric Measures** vidimo da je C-koeficijent 0.102, a Kramerov V koeficijent 0.103 (u slučaju tabele 2x2 jednak je Fi-koeficijentu jedino što SPSS za Fi-koeficijent dodeljuje predznak, ovog puta minus. Predznak fi-koeficijenta se određuje na osnovu raspodele frekvencija u dijagonalama tabele ali nije uvek koristan za tumačenje tabele). Iz tabele kontingencije vidimo da je među ženama koje nisu primale estrogen tokom menopauze relativno veći procenat žena koje su obolele od Alchajmerove bolesti (16.3%) nego među ženama koje jesu primale estrogen (5.8%). Drugim rečima, žene koje nisu primale estrogen tokom menopauze u većem su riziku da obole u narednih pet godina od Alchajmerove bolesti nego žene koje jesu primale estrogen.

Zadatak 4.

U fajlu **abortus_velporpor.sav** su dati podaci sa slu ajnog uzorka o veli ini porodice porekla ispitanika – varijabla **velporpor** (iskazanoj prema broju dece: 1 –sa jednim detetom, 2- sa dva deteta 3 –sa tri deteta 4- sa etvoro ili više dece) i o odnosu ispitanika prema abortusu - varijabla **abortus** (kategorije: 1- svaka žena ima pravo da odlu uje o tome da li e abortirati; 2- odobravam abortus samo u odre enim slu ajevima kad je to medicinski opravdano; 3- abortus uopšte ne treba dozvoliti).

- Utvrditi da li u populaciji iz koje je uzorak postoji povezanost izme u veli ine porodice porekla i odnosa prema abortusu;
- Ako povezanost postoji izra unati koeficijent asocijacije ovih varijabli.

Postupak zapo injemo sa **Analyze/Descriptive statistics/Crosstabs**. U **Row(s)** ubacujemo varijablu **velporpor**, a u **Column(s)** varijablu **abortus**. Zatim kliknemo na **Statistics** i u okviru prozora **Crosstabs: Statistics** štrikliramo **Chi-square, Contingency coefficient** i **Phi and Cramer's V**.



U tabeli Chi-Square tests vidimo da je verovatno a da hi-kvadrat statistik (ako je nulta hipoteza ta na) bude 16.315 ili ve i jednaka 0.012 što je manje od 0.05 i stoga zaklju ujemo da se nulta hipoteza može odbaciti i da u populaciji postoji povezanost izme u veli ine porodice porekla i odnosa prema abortusu.

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16,315 ^a	6	,012
Likelihood Ratio	17,811	6	,007
Linear-by-Linear Association	7,413	1	,006
N of Valid Cases	180		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 9,63.

Najbolji pokazatelj asocijacije u tabeli **Symmetric measures** je Kramerov V koeficijent čija je vrednost u ovom slučaju 0.213.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,301	,012
	Cramer's V	,213	,012
	Contingency Coefficient	,288	,012
N of Valid Cases		180	

Zadatak 5.

Fajl **odgnastp.sav** nalaze se podaci o tačnosti odgovora na jedno statističko pitanje (varijabla **tacn_odg**) koje je dao slučajni uzorak studenata, asistenata, docenata i profesora sa različitih odeljenja za psihologiju u Americi (podatak kojim od grupa ispitanika pripada je u varijabli **grupa**).

- Izračunati verovatnoću tačnog odgovora pod uslovom da ispitanici pripadaju grupi docenata;
- Izračunati verovatnoću tačnog odgovora na pitanje;
- Izračunati verovatnoću da se tačnost odgovora na pitanje i da se istovremeno pripada grupi docenata ako su tačnost odgovaranja i pripadnost grupi statistički nezavisne varijable.

Verovatno e ocenjujemo relativnim frekvencijama koje ra unamo na osnovu frekvencija iz kontingencijske tabele. Dakle **Analyze/Descriptive statistics/Crosstabs**. U **Row(s)** ubacujemo varijablu **grupa**, a u **Column(s)** varijablu **tacn_odg**.

grupa * tacnost odgovora Crosstabulation

Count

		tacnost odgovora		Total
		tacno	netacno	
Grupa	student	10	7	17
	asistent	59	116	175
	docent	58	76	134
	profesor	93	89	182
Total		220	288	508

- Verovatno a ta nog odgovora za ispitanike koji pripadaju grupi docenata ra una se kao uslovna verovatno a $(58/508)/(134/508)$ ili, što je isto, deljenjem broja docenata koji su ta no odgovorili sa ukupnim brojem docenata. Dakle $58/134$, a što iznosi 0.433 ili 43.3%.
- Verovatno a da se ta no odgovori na pitanje dobija se deljenjem broja svih ispitanika koji su ta no odgovorili (dakle, uzevši u obzir sve grupe), sa ukupnim brojem ispitanika. Darkle, deljenjem 220 sa 508 što iznosi 0.433 ili 43.3%.
- Ako su ta nost odgovora i pripadnost grupi statisti ki nezavisne varijable onda su i doga aji da se ta no odgovori na pitanje i da se bude u grupi docenata statisti ki nezavisni doga aji pa je verovatno a njihovog istovremenog dešavanja jednaka proizvodu verovatno a svakog od tih doga aja. Verovatno a da se ta no odgovori na pitanje je 0.433 (to smo izra unali u prethodnom zahtevu), a verovatno a da se bude u grupi docenata je $134/508 = 0.264$. Prema tome verovatno a istovremenog dešavanja ova dva doga aja, ako su ta nost odgovora i pripadnost grupi statisti ki nezavisne varijable, jeste $0.433 \cdot 0.264$, a to je 0.114 ili 11.4%.

Zadatak 6.

Fajl isti kao za zadatak 5.

- Testirati pretpostavku da studenti, asistenti, docenti i profesori podjednako ta no, odnosno neta no odgovaraju na ovo statisti ko pitanje;
- Ako se prethodna pretpostavka odbacuje izra unati koeficijent aoscijacije, tj. povezanosti izme u varijabli **grupa** i **tacn_odg**.

Pretpostavku testiramo ve poznatim postupkom, hi-kvadrat testom, samo se sada taj test ne koristi kao test statisti ke nezavisnosti ovih dveju varijabli ve kao **test homogenosti** triju grupa u pogledu ta nosti odgovora na pitanje. (Ra unski sve je isto kao kod testa nezavisnosti, samo se tuma enje rezultata razlikuje). Odabiramo **Analyze/Descriptive statistics/Crosstabs**. Nakon prebacivanja varijable **grupa** u okvir **Row(s)**, a varijable **tacn_odg** u okvir **Column(s)**, klikom na **Statistics** u okviru prozora **Crosstabs:Statistics** štrikliramo **Chi-square**, **Contingency coefficient** i **Phi and Cramer's V**. Posle klika na **Continue** kliknemo na dugme **Cells** pa štrikliramo u okviru **Percentages** opcije **Raw** i **Column**. Definiciju procedure završavamo sa **Continue** i **Ok**.

grupa * tacnost odgovora Crosstabulation					
		<u>tacnost odgovora</u>			
		tacno	netacno	Total	
grupa	student	Count	10	7	17
		% within grupa	58,8%	41,2%	100,0%
		% within tacnost odgovora	4,5%	2,4%	3,3%
	asistent	Count	59	116	175
		% within grupa	33,7%	66,3%	100,0%
		% within tacnost odgovora	26,8%	40,3%	34,4%
	docent	Count	58	76	134
		% within grupa	43,3%	56,7%	100,0%
		% within tacnost odgovora	26,4%	26,4%	26,4%
	profesor	Count	93	89	182
		% within grupa	51,1%	48,9%	100,0%
		% within tacnost odgovora	42,3%	30,9%	35,8%
Total		Count	220	288	508
		% within grupa	43,3%	56,7%	100,0%
		% within tacnost odgovora	100,0%	100,0%	100,0%

U tabeli **Chi-Square Tests** uočavamo da je verovatnoća da na slučajnom uzorku dobijemo Pearsonov hi-kvadrat 12.727 ili veći jednaka 0.005, što je manje od 0.05, pa stoga zaključujemo da možemo odbaciti nultu hipotezu. Prema tome, asistenti, docenti i profesori ne daju podjednako iste tačne odgovore na ovo statističko pitanje. Iz prethodno prikazane tabele vidimo da asistenti i docenti relativno ređe (asistenti: 33.7%; docenti: 43.3%) daju tačne odgovore od studenata i profesora (studenti: 58.8% ; profesori: 51.1%).

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12,727 ^a	3	,005
Likelihood Ratio	12,824	3	,005
Linear-by-Linear Association	6,335	1	,012
N of Valid Cases	508		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 7,36.

O povezanosti između ove dve varijable nam govori Kramerov V koeficijent, čija vrednost vidimo iz tabele **Symmetric Measures**, a koja u ovom slučaju iznosi 0.158 i statistički je značajna (statistička značajnost Kramerovog V identična je uvek statističkoj značajnosti Pearsonovog hi-kvadrata).

Symmetric Measures			
		Value	Approx. Sig.
Nominal by Nominal	Phi	,158	,005
	Cramer's V	,158	,005
	Contingency Coefficient	,156	,005
N of Valid Cases		508	