

## Analiza podataka u obliku rangova: rešenja zadataka

Rešenja uradila Selma ustovi, studentkinja psihologije  
(generacija 2012/2013)

### Zadatak 1.

U fajlu uspdisc.sav nalaze se rangovi za uzorak u enika osnovne škole u pogledu njihovog uspeha u matematici (varijabla uspeh) i disciplinovanosti (varijabla disc).

- Ustanovite da li u populaciji iz koje je uzorak postoji monotona povezanost između disciplinovanosti u enika i njihovog uspeha u matematici;
- Izračunajte Brave-Pirsonov koeficijent linearne korelacije između disciplinovanosti u enika i njihovog uspeha u matematici (bez obzira na to što to ne bi trebalo raditi sa rangovanim podacima) i uporedite dobijeni koeficijent sa Spirmanovim ro-koeficijentom korelacije između dva niza rangova. Do kakvog ste zaključka došli?;

Korelaciju za dve varijable sa rangovima ćemo izračunati preko Analyze Correlate Bivariate, tako što ćemo u polje Variables prebaciti varijable disc i uspeh i uključiti testove Kendall's tau-b i Spearman.

Dobijamo ispis:

		uspeh	disc
uspeh	Correlation Coefficient	1,000	<b>,892</b>
	Sig. (2-tailed)	.	,000
	N	25	25
	Spearman's rho		
disc	Correlation Coefficient	,892**	1,000
	Sig. (2-tailed)	,000	.
	N	25	25

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

### Correlations

		uspeh	disc
Kendall's tau_b	uspeh	1.000	.720
	Correlation Coefficient		
	Sig. (2-tailed)		.000
	N	25	25
disc	uspeh	.720**	1.000
	Correlation Coefficient		
	Sig. (2-tailed)	.000	
	N	25	25

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

Pošto je Spirmanov koeficijent korelacije statistički značajan i iznosi 0.892, zaključujemo da između u disciplinovanosti na časovima i uspeha u matematici populaciji postoji jaka pozitivna monotona povezanost između ova dva niza rangovanih podataka. Takođe uočavamo da je za iste podatke Kendallov tau koeficijent korelacije aritmetički niži od Spirmanovog koeficijenta, ali su zaključci o značajnosti isti za oba.

Da bismo izračunali Brave-Pirsonov koeficijent korelacije u meniju Analyze Correlate Bivariate uključimo Pearson i varijable uspeh i disc prebaciti u desni okvir.

### Correlations

		uspeh	disc
uspeh	Pearson Correlation	1	.892**
	Sig. (2-tailed)		.000
	N	25	25
disc	Pearson Correlation	.892**	1
	Sig. (2-tailed)	.000	
	N	25	25

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

Kao što se može videti iz ispisa, kada se Spirmanov i Brave-Pirsonov koeficijent korelacije računaju na osnovu dva niza rangovanih podataka, oni su matematički jednaki. (Dakle, Spirmanov koeficijent je zapravo Pirsonov koeficijent izračunat na rangovima).

**Napomena: Važno je znati da Pirsonov i Spirmanov koeficijent neće biti jednaki ako ih računamo iz intervalnih podataka. U tom slučaju Brave-Pirsonov koeficijent će biti računat iz intervalnih podataka a Spirmanov koeficijent će biti izračunat na osnovu rangova koje će SPSS automatski napraviti iz intervalnih podataka.**

## Zadatak 2.

U fajlu rangovanje\_eseja.sav tri ocenjiva a su rangovala 15 eseja prema umetni kom utisku. Utvrditi koliko je slaganje medju ocenjiva ima.

Za odredivanje slaganja, odnosno povezanosti više od dva niza rangova koristimo Kendalov koeficijent slaganja. Za ovaj test potrebno je da podaci budu organizovani tako da ocenjiva i, odnosno kriterijumi rangovanja budu u redovima, a entiteti koji se procenjuju (u ovom slučaju eseji) u kolonama.

Test radimo preko menija **Analyze** **Nonparametric Tests** **Legacy Dialogs** **K Related Samples**, tako što ćemo u prozoru za dijalog u **Test Variables** prebaciti sve varijable koje se odnose na eseje, ukupno njih 15 (od esej1 do esej15), a u **Test Type** štriklirati **Kendall's W**.

N	3
Kendall's W <sup>a</sup>	,550
Chi-Square	23,105
df	14
Asymp. Sig.	,059

Iz ispisa vidimo da Kendalov W koeficijent nije statistički značajan (mada je na ivici značajnosti) što znači da ne možemo odbaciti nultu hipotezu koja tvrdi da u populaciji nema slaganja. Prema tome, ne možemo tvrditi da postoji slaganje u ocenama koje su ocenjiva i davali esejima pa ne možemo zaključiti ni da su imali iste kriterijume pri rangovanju eseja.

---

## Zadatak 3.

U fajlu medijana.sav nalaze se rangovi koje su radnici dveju grupa dobili u pogledu uspešnosti u obavljanju posla (varijabla rang\_usp). Dve grupe su tome poslu obavane dvema različitim metodama: varijabla grupa sadrži podatak o tome kojom od dve metode je radnik obavljan poslu.

- Ustanovite pregledom rangova (varijabla rang\_usp) da li je rangovanje vršeno za sve ispitanike zajedno, bez obzira na

njihovu grupnu pripadnost, ili su ispitanici rangovani posebno unutar svake grupe;

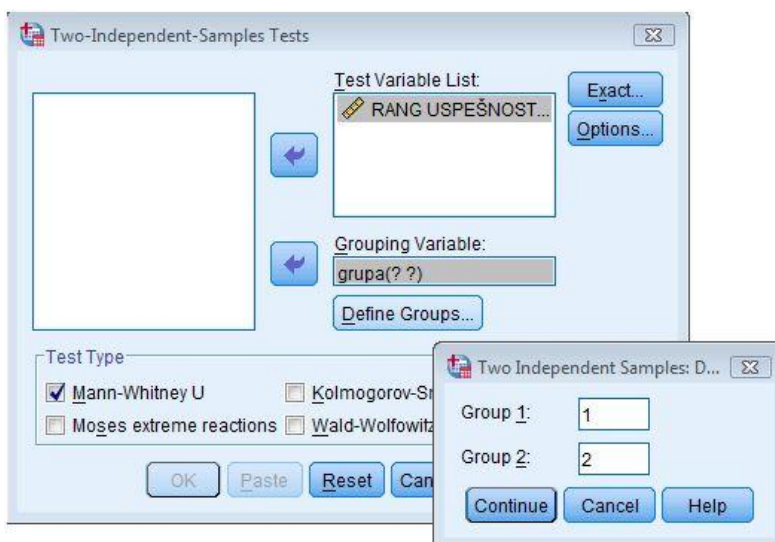
- Utvrdite da li ove dve različite metode obuke imaju različite efekte na uspeh u obavljanju datog posla.

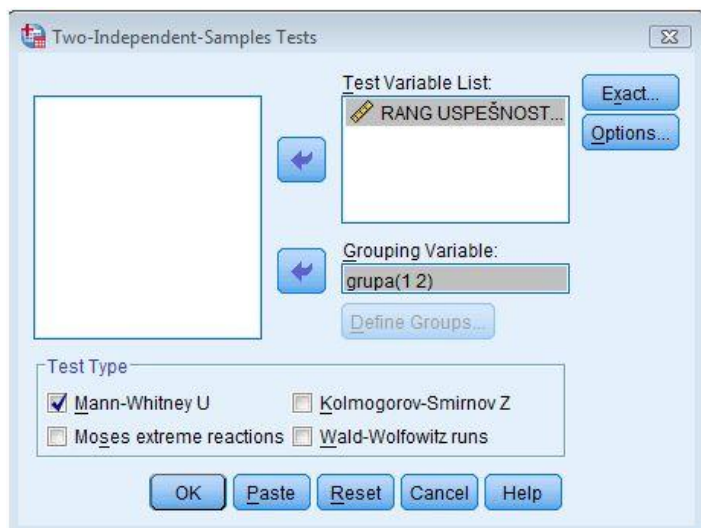
Pregledom varijable rang\_usp vidimo da je broj ispitanika 40 a po rangovima vidimo da su svi ispitanici rangovani zajedno. Da li su obe grupe rangovane zajedno možemo proveriti i preko obrasca  $n*(n+1)/2$ , koji će nam dati sumu svih brojeva od 1 do n. Ako je ovaj broj (u ovom slučaju 820) jednak sumi svih rangova, zaključimo da su ispitanici rangovani zajedno.

Sumu varijable rang\_usp dobijamo kada u Descriptive Statistics > Descriptives u Options... označimo opciju Sum, ime dobijamo broj 820.

Da li primenjene metode imaju različite efekte, proverimo Mann-Vitnijevim testom, koji će nam pokazati da li ova dva uzorka pripadaju populacijama sa istom distribucijom. Kada odemo na Analyze > Nonparametric Tests > Legacy Dialogs > 2 Independent Samples, u prozoru za dijalog, u polje Test Variable List prebacujemo varijablu rang\_usp, a u polje Grouping Variable varijablu grupa za koju je potrebno definisati opseg. To ćemo uraditi klikom na Define Groups..., tako što ćemo u polje Group 1 uneti vrednost 1, kojim je označena prva metoda, a u polje Group 2 vrednost 2, koja označava drugu metodu.

U delu Test Type uobičajeno je već po defaultu (podrazumevano) uključeno Mann-Whitney U test. Ako slučajno nije onda treba štriklirati ovu opciju.





Klikom na OK dobijamo:

#### Ranks

	grupa	N	Mean Rank	Sum of Ranks
RANG USPEŠNOSTI U OBAVLJANJU POSLA	Metoda 1	24	23.04	553.00
	Metoda 2	16	16.69	267.00
	Total	40		

#### Test Statistics<sup>a</sup>

	RANG USPEŠNOSTI U OBAVLJANJU POSLA
Mann-Whitney U	131.000
Wilcoxon W	267.000
Z	-1.687
Asymp. Sig. (2-tailed)	.092
Exact Sig. [2*(1-tailed Sig.)]	.095 <sup>b</sup>

a. Grouping Variable: grupa

b. Not corrected for ties.

Nulta hipoteza tvrdi da dva uzorka pripadaju istoj populaciji u pogledu uspešnosti u datom poslu, odnosno da oba uzorka pripadju populaciji sa istom distribucijom. Budu i da dobijeni statistik Z nije statisti ki zna ajan, ne možemo odbaciti nultu hipotezu i zaklju ujem da ne možemo tvrditi

da ove dve različite metode obuke pokazuju različite efekte na uspešnost u obavljanju datog posla.

---

#### Zadatak 4.

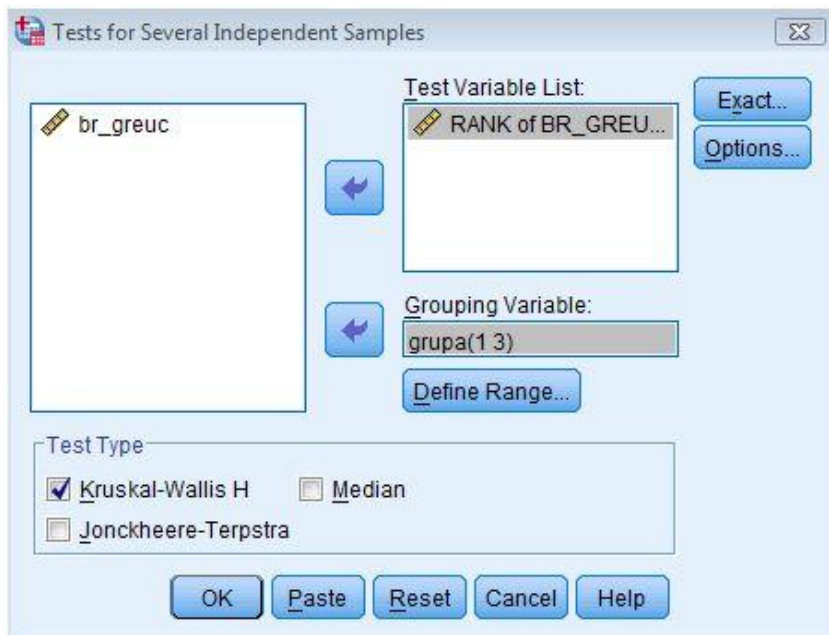
U fajlu `dislek.sav` nalaze se rangovi koji su napravljeni na osnovu uestalosti grešaka određene vrste koje pri ispitivanju prave ispitanici koji imaju neku od tri različite vrste disleksije (fonološka, površinska dubinska). Rangovi se nalaze u varijabli `rang_brgr`, a pripadnost ispitanika grupi u pogledu vrste disleksije u varijabli `grupa`.

- Ustanovite pregledom podataka (varijabla `rang_brgr`) da li je rangovanje vršeno za sve ispitanike zajedno, bez obzira na njihovu grupnu pripadnost, ili su ispitanici rangovani posebno unutar svake grupe;
- Utvrdite da li se prema uestalosti ovog tipa greške u ispitivanju razlikuju disleksionari koji imaju tri različite vrste disleksije.

Da li su ispitanici rangovani zajedno proverimo na isti način kao u prethodnom zadatku. I vidimo da jesu.

Ovde ćemo koristiti Kruskal-Wallisov test kojim ispitujemo da li se 2 ili više nezavisnih uzoraka razlikuju u pogledu centralne tendencije, odnosno da li uzorci pripadaju istoj populaciji u pogledu ispitivane osobine.

U `Analyze > Nonparametric Tests > Legacy Dialogs > K Independent Samples`, popunjavamo prozor za dijalog tako što u polje `Test Variable List` prebacujemo varijablu `rang_brgr`, a u polje `Grouping Variable` varijablu `grupa`. Varijabli `grupa` je potrebno definisati opseg, a to ćemo učiniti preko opcije `Define Range...` Kako imamo tri grupe disleksionara označene brojevima od 1 do 3, u polje `Minimum` upisujemo cifru 1, a u polje `Maximum` cifru 3.



U ispisu dobijamo:

	VRSTA DISLEKSIJE	N	Mean Rank
RANK of BR_GREUC	DUBINSKA	5	4,40
	FONOLOŠKA	5	7,40
	POVRŠINSKA	4	11,50
	Total	14	

	RANK of BR_GREUC
Chi-Square	6,406
df	2
Asymp. Sig.	,041

a. Kruskal Wallis Test

b. Grouping Variable: VRSTA

DISLEKSIJE

Dobijeni statistik Chi-square je statistički značajan na nivou 0.05, pa odbacujemo nultu hipotezu koja tvrdi da uzorci pripadaju istoj populaciji u pogledu ispitivane osobine i zaključujemo da se tri grupe disleksijske razlikuju po učestalosti ovog tipa greške. Budući da se rangovi od 1 nadalje dodeljuju idući od najnižeg rezultata iz tabele Ranks vidimo da

disleksiji su sa dubinskom disleksijom prave najmanje grešaka ove vrste (prose ni rang u koloni Mean Rank je najniži za ovu grupu), a površinski disleksiji su prave najviše grešaka ove vrste.

---

#### Zadatak 5.

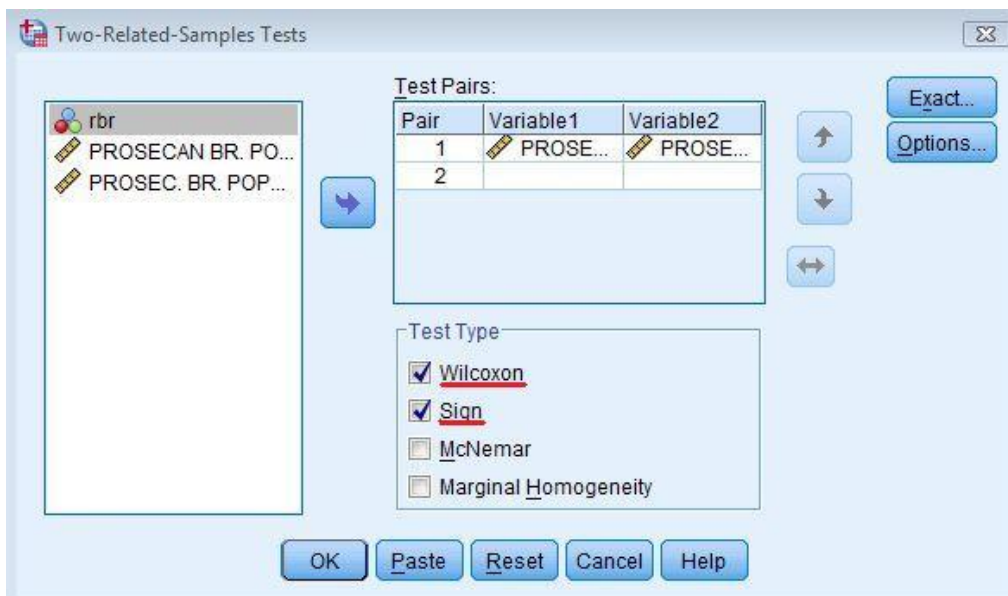
U fajlu puseenje.sav

nalaze se podaci o prose i broju popušenih cigareta na dan i to pre (varijabla pbpc\_pre) i posle kampanje protiv puseenja (varijabla pbpc\_pos) za jedan slu ajni uzorak iz populacije puša a koji su bili izloženi kampanji protiv pušenja.

- Utvrditi (testom predznaka i Wilkoxsonovim testom ekvivalentnih parova) da li je ova kampanja imala efekta.

U pitanju su dva zavisna uzorka, pa ćemo testove vršiti preko **Analyze > Nonparametric Tests > Legacy Dialogs > 2 Related Samples**.

Varijable pbpc\_pre i pbpc\_pos ćemo prebaciti u polje Test Pairs. Bitno je da nakon prebacivanja one budu u istom paru. U polju Test Type ćemo štriklirati testove Wilcoxon i Sign (test predznaka).





Dobijamo:

### Wilcoxon Signed Ranks Test

		Ranks		
		N	Mean Rank	Sum of Ranks
PROSEC. BR. POPUS.	Negative Ranks	7 <sup>a</sup>	8,57	60,00
CIGAR. POSLE k. -	Positive Ranks	7 <sup>b</sup>	6,43	45,00
PROSECAN BR. POPUS.	Ties	1 <sup>c</sup>		
CIGARETA-PRE	Total	15		

a. PROSEC. BR. POPUS. CIGAR. POSLE k. < PROSECAN BR. POPUS. CIGARETA-PRE

b. PROSEC. BR. POPUS. CIGAR. POSLE k. > PROSECAN BR. POPUS. CIGARETA-PRE

c. PROSEC. BR. POPUS. CIGAR. POSLE k. = PROSECAN BR. POPUS. CIGARETA-PRE

Test Statistics <sup>a</sup>	
	PROSEC. BR. POPUS. CIGAR. POSLE k. - PROSECAN BR. POPUS. CIGARETA- PRE
Z	-,474 <sup>b</sup>
Asymp. Sig. (2-tailed)	,635

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

### Sign Test

		Frequencies	
		N	
PROSEC. BR. POPUS.	Negative Differences <sup>a</sup>	7	
CIGAR. POSLE k. -	Positive Differences <sup>b</sup>	7	
PROSECAN BR. POPUS.	Ties <sup>c</sup>	1	
CIGARETA-PRE	Total	15	

a. PROSEC. BR. POPUS. CIGAR. POSLE k. < PROSECAN BR. POPUS. CIGARETA-PRE

b. PROSEC. BR. POPUS. CIGAR. POSLE k. > PROSECAN BR. POPUS. CIGARETA-PRE

c. PROSEC. BR. POPUS. CIGAR. POSLE k. = PROSECAN BR. POPUS. CIGARETA-PRE

Test Statistics <sup>a</sup>	
	PROSEC. BR. POPUS. CIGAR. POSLE k. - PROSECAN BR. POPUS. CIGARETA- PRE
Exact Sig. (2-tailed)	1,000 <sup>b</sup>

a. Sign Test

b. Binomial distribution used.

Iz ispisa vidimo da ni jedan ni drugi statistik nisu statistički značajni, što znači da ne odbacujemo nultu hipotezu koja tvrdi da uzorci pripadaju istoj populaciji u pogledu ispitivane osobine i zaključujemo da ne možemo tvrditi da postoji razlika u prosečnom broju popušanih cigareta pre i posle kampanje protiv pušenja, što znači da ne možemo tvrditi da je kampanja imala efekta.

### Zadatak 6.

U fajlu cokolade.savdati su podaci za 30 degustatora cokolada iste vrste od različitih proizvođača. Svaki degustator je prema kvalitetu ukusa rangovao cokolade ova 4 proizvođača od 1 (najukusnija) do 4 (najmanje ukusna od svih).

- Utvrditi da li su prema proceni degustatora distribucije ukusnosti cokolada iste vrste ovih proizvođača jednake, tj. Da li degustatori cokolade različitih proizvođača procenjuju kao podjednako ukusne. Za stepen ukusnosti cokolade ne bismo mogli pretpostaviti normalnu raspodelu u populaciji.

U pitanju su zavisni uzorci i stoga analizu radimo pomoću u Analyze Nonparametric Tests Legacy Dialogs K Related Samples i koristimo Friedmanov test. U polje Test Variables prebacujemo varijable cokolad1,2,3 i 4, štikliramo test Friedman i dobijamo:

**Ranks**

	Mean Rank
cokolad1	1,75
cokolad2	2,43
cokolad3	2,62
cokolad4	3,20

**Test Statistics<sup>a</sup>**

N	30
Chi-Square	19,866
df	3
Asymp. Sig.	,000

a. Friedman Test

Vidimo da je Hi-kvadrat statistika statistički značajan, što znači da odbacujemo nultu hipotezu koja tvrdi da se grupe ne razlikuju po ispitivanoj osobini, tj. da degustatori sve vrste čokolade procenjuju kao podjednako ukusne. Kako se četiri vrste čokolade po proceni degustatora razlikuju po ukusu, koja je ocenjena kao najukusnija možemo videti u tabeli Ranks gde su dati prosečni rangovi ocena za svaku vrstu čokolade. Pošto je najniži rang dodeljivan onoj čokoladi koju su degustatori ocenili kao najukusniju, iz tabele zaključujemo da je to prva vrsta čokolade koja ima najniži prosečni rang.

---

### Zadatak 7.

U fajlu stavmat.sav postoje podaci o stavu prema matematici (varijable STPMATU - "uživanje u bavljenju matematikom" i STPMATV - "vrednovanje, tj. procena važnosti matematike") novoupisanih studenata sa tri fakulteta (varijabla GRUPA).

- Rangovati sve studente u pogledu varijable STPMATV, tj. vrednovanja matematike, bez obzira na to sa kojeg su fakulteta;
- Koriste li varijablu u kojoj su rangovi (varijabla RSTPMATV?) utvrditi da li kod studenata postoji efekat fakulteta koji se studira (varijabla GRUPA) na procenu važnosti matematike. Utvrditi da li kod studenata postoji efekat fakulteta koji se studira korišćenjem originalnih podataka u varijabli STPMATV i uporedite zaključak do kojeg ste dosli koriste li rangove sa

odgovorom koji ste dobili koriste i jednofaktorijalnu analizu varijanse.

Iako SPSS sam rangira podatke pri analizi podataka ako se traži metoda za rangove a podaci nisu rangovani, napravi emo rangovanje. Rangira emo ispitanike po varijabli stavmatv pomo u opcije Rank Cases u meniju Transform. U polje Variable(s) prebacujemo varijablu stavmatv i u Assign Rank 1 to ostavimo uklju eno Smallest value što zna i da e najniži skor na varijabli bobiti rang 1 jer to radi i SPSS kada primenjujemo Kraskal-Volisov test.

Analizu radimo preko Analyze Nonparametric Tests Legacy Dialogs K Independent Samples. U Test Variable List unosimo novodobijenu varijablu Rstavmatv, a u Grouping Variable varijablu grupa.

[prozor za dijalog je isti kao u zadatku 4.]

Klikom na OK dobijamo ispis:

### Kruskal-Wallis Test

Ranks			
	grupa	N	Mean Rank
Rank of stpmatv	psihologija	96	112,93
	etnologija	44	78,72
	gradjevina	102	148,02
	Total	242	

Test Statistics <sup>a,b</sup>	
Rank of stpmatv	
Chi-Square	32,621
df	2
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: grupa

Hi-kvadrat statistik je statisti ki zna ajan što zna i da odbacujemo nultu hipotezu koja tvrdi da nema razlika izme u grupa u populaciji, odnosno da nema efekta fakulteta na procenu važnosti matematike. Pošto pri rangovanju najniži rezultat dobija rang 1, onda e i grupa koja najmanje vrednuje matematiku imati najniži prose ni rang u tabeli Ranks, a to je

etnologija. Za razliku od njih, studenti gra evine, sa najvišim prose nim rangom, najviše vrednuju matematiku.

Napomena: Ukoliko bismo umesto varijable sa rangovima koristili u Kraskal-Volisovom testu originalnu varijablu stavmatv sa intervalnim podacima SPSS bi originalne podatke automatski rangovao i ishod procedure bi bio isti.

Na originalnim intervalnim podacima mogli bismo koristiti jednofaktorijalnu analizu varijanse preko menija Analyze CompareMeans One-Way ANOVA. U polje Dependent List prebacujemo varijablu stavmatv, a u polje Factor varijablu grupa.

#### ANOVA

STAV PREMA MATEMATICI-Vrednost matematike

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1235.757	2	617.878	20.387	.000
Within Groups	7243.499	239	30.308		
Total	8479.256	241			

Iz ispisa vidimo da je F statistik statisti ki zna ajan što zna i da odbacujemo nultu hipotezu koja tvrdi da su aritmeti ke sredine subpopulacija jednake, odnosno da sve grupe podjednako vrednuju matematiku, bez obzira na to sa kog su fakulteta. Dakle, opet zaklju ujemo da kod studenata postoji efekat fakulteta koji se studira na procenu vrednosti matematike.

Svaka parametrijska metoda ima svoju neparametrijsku zamenu u slu ajevima kada uslovi za njenu primenu nisu ispunjeni. Za jednofaktorijalnu analizu varijanse sa neponovljenim faktorom to je Kraskal-Volisov test. U ovom sluaju ispunjeni su uslovi za jednofaktorijalnu analizu varijanse, a isti zaklju ak dobijen je primenom obeju metoda.

---

### Zadatak 8.

U fajlu spavanje\_motorika.sav dati su podaci o broju grešaka u izvodjenju jednog motornog zadatka posle 24 sata nespavanja (varijabla mot24hns), 36 sati nespavanja (mot36hns) i 48 sati nespavanja (mot48hns) dobijeni na istoj grupi ispitanika u ove tri situacije.

Ustanoviti Fridmanovim testom postoji li efekat duzine nespavanja na uspesnost u izvodjenju motorne vestine i uporediti rezultate sa onima koji se dobijaju analizom varijanse za ponovljena merenja.

Fridmanov test radimo preko Analyze Nonparametric Tests Legacy Dialogs K Related Samples. U polje Test Variables prebacujemo varijable mot24hns, mot36hns i mot48hns, a u Test Type obeležimo opciju Friedman.

Dobijamo ispis:

	Mean Rank
Broj_gresaka_u_zadatku_motorne_vestine_posle_24_h_nespavanja	1,15
Broj_gresaka_u_zadatku_motorne_vestine_posle_36_h_nespavanja	2,00
Broj_gresaka_u_zadatku_motorne_vestine_posle_48_h_nespavanja	2,85

N	10
Chi-Square	16,056
df	2
Asymp. Sig.	,000

a. Friedman Test

Hi-kvadrat statistik je statisti ki zna ajan pa odbacujemo nultu hipotezu koja tvrdi da grupe pripadaju istoj populaciji u odnosu na ispitivanu

osobinu i zaključujemo da dužina nespavanja ima efekta na izvođenje motornog zadatka.

Na istim podacima uradimo i analizu varijanse sa ponovljenim merenjima preko menija Analyze → General Linear Model → Repeated Measures.

U prozoru za dijalog, u polje Within-Subjects Factor Name umesto factor1 upisujemo ime ponovljenog faktora (duznsp), a u polje Number of Levels broj nivoa ponovljenog faktora, u ovom slučaju 3, nakon čega treba kliknuti na opciju Add.

Odabirom opcije Define prelazimo u novi prozor za dijalog gde ćemo varijable mot24hns, mot36hns i mot48hns prebaciti u polje Within-Subjects Variables vodeći ih u redosledu nivoa.

Dobijamo:

Within-Subjects Factors	
Measure: MEASURE_1	
duznsp	Dependent Variable
1	mot24hns
2	mot36hns
3	mot48hns

Tests of Within-Subjects Effects						
Measure: MEASURE_1						
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
duznsp	Sphericity Assumed	68.067	2	34.033	23.622	.000
	Greenhouse-Geisser	68.067	1.330	51.185	23.622	.000
	Huynh-Feldt	68.067	1.473	46.210	23.622	.000
	Lower-bound	68.067	1.000	68.067	23.622	.001
Error(duznsp)	Sphericity Assumed	25.933	18	1.441		
	Greenhouse-Geisser	25.933	11.968	2.167		
	Huynh-Feldt	25.933	13.257	1.956		
	Lower-bound	25.933	9.000	2.881		

Kako je F statistika statistički značajna, odbacujemo nultu hipotezu koja tvrdi da u populaciji ne postoji efekat faktora i zaključujemo da dužina nespavanja utiče na uspešnost u izvođenju motornog zadatka.

Fridmanov test je neparametrijska zamena za analizu varijanse u nacrtima sa ponovljenim merenjima. I jedna i druga metoda ispituju postojanje razlika između više zavisnih nizova podataka. U ovom slučaju su nas obe analize dovele do istog zaključka, ali to ne mora uvek da bude slučaj.

Kada uslovi za analizu varijanse sa ponovljenim merenjima nisu ispunjeni podatke možemo analizirati Fridmanovim testom.