

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Testiranje statističke značajnosti razlika prema nominalnim varijablama vrši se pomoću

- ➔ hi-kvadrat testa za jedan uzorak,
- ➔ hi-kvadrat testa za dva ili više nezavisnih uzoraka i
- ➔ McNemarov hi-kvadrat testa za dva zavisna uzorka.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Hi-kvadrat (χ^2) testom za jedan uzorak utvrđuje se statistička značajnost odstupanja frekvencija utvrđenih jednodimenzionalnim grupiranjem od frekvencija očekivanih pod određenom hipotezom.

Pri tome se najčešće testira da li se udjeli entiteta u pojedinim kategorijama statistički značajno razlikuju, odnosno da li utvrđene frekvencije odstupaju statistički značajno od frekvencija očekivanih pri posve slučajnoj raspodjeli.

Prije provođenja hi-kvadrat testa za jedan uzorak potrebno je izvršiti grupiranje entiteta prema analiziranoj varijabli.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Primjer: Praktičnom dijelu ispita iz KM-a pristupilo je 40 studenata; 15 ih je položilo, a 25 nije.

USPJEH NA ISPITU	f_o	%
NISU POLOŽILI	25	62,5
POLOŽILI	15	37,5
UKUPNO	40	100

Da li se udio studenata koji su položili praktični dio ispita iz KM-a statistički značajno ($p < 0,05$) razlikuje od udjela studenata koji nisu uspjeli položiti?

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Najprije je potrebno izračunati teoretske frekvencije. Teoretske frekvencije se pod pretpostavkom posve slučajne raspodjele izračunavaju na način da se ukupan broj entiteta, tj. suma frekvencija svih kategorija podijeli s brojem kategorija (u ovom primjeru $f_t = 40/2 = 20$).

USPJEH NA ISPITU	f_t	%
NISU POLOŽILI	20	50
POLOŽILI	20	50
UKUPNO	40	100

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Statistička značajnost razlike između opaženih frekvencija i teoretskih frekvencija se testira putem hi-kvadrat distribucije s $df=N-1$ stupnjeva slobode gdje N označava ukupan broj polja. Hi-kvadrat vrijednost (χ^2) se izračunava sljedećom formulom

$$\chi^2 = \sum_{g=1}^N \frac{(f_{og} - f_{tg})^2}{f_{tg}}$$

gdje je

- ➔ f_{og} - opažena frekvencija polja g ($g=1, \dots, N$)
- ➔ f_{tg} - teoretska frekvencija polja g ($g=1, \dots, N$)
- ➔ N - ukupan broj polja

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

U ovom primjeru hi-kvadrat vrijednost se izračunava na sljedeći način:

$$\chi^2 = \sum_{g=1}^N \frac{(f_{o_g} - f_{t_g})^2}{f_{t_g}} = \frac{(25 - 20)^2}{20} + \frac{(15 - 20)^2}{20} = 2,5$$

$$df = (2 - 1) = 1$$

$$\chi^2 = 2,5 \quad df = 1 \quad p < 0,114$$

Izračunata hi-kvadrat vrijednost i pripadajuća minimalna pogreška statističkog zaključka upućuju da se udjeli studenata ne razlikuju statistički značajno na razini pogreške 5%.

STATISTICA 7

Jednodimenzionalno grupiranje kvalitativnih podataka

Jednodimenzionalno grupiranje kvalitativnih podataka izvodi se slijedom koraka: padajući izbornik *Statistics* → *Basic Statistics/Tables* → *Frequency Tables* → *Advanced* → *All distinct values*.

Zadatak - Pokrenite datoteku *TA.sta*! Utvrdite frekvencije tjelesno aktivnih i tjelesno neaktivnih osoba u inicijalnom testiranju (*TA-I*)!

STATISTICA 7

Hi-kvadrat test za jedan uzorak

Prije provođenja hi-kvadrat testa za jedan uzorak potrebno je kreirati novu matricu podataka te u jedan stupac unijeti opažene frekvencije utvrđene jednodimenzionalnim grupiranjem podataka, a u drugi teoretske frekvencije. Hi- kvadrat test za jedan uzorak izvodi se slijedom koraka: padajući izbornik *Statistics* → *Nonparametrics* → *Observed versus expected X2* .

Zadatak - Na temelju prethodno provedenog grupiranja podataka hi-kvadrat testom utvrdite da li se udjeli tjelesno aktivnih i tjelesno neaktivnih osoba statistički značajno razlikuju na razini pogreške 5% !

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Hi-kvadrat testom za dva ili više nezavisnih uzoraka utvrđuje se statistička značajnost razlika između dviju ili više grupa entiteta prema nekoj nominalnoj varijabli.

Prije provođenja hi-kvadrat testa za dva ili više nezavisnih uzoraka potrebno je izvršiti dvodimenzionalno grupiranje entiteta pri čemu je jedna varijabla selektorska (određuje pripadnost entiteta određenoj grupi), a druga varijabla ona prema kojoj se žele testirati razlike između grupa.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Primjer: Od 26 studenata koji su pristupili praktičnom dijelu ispita iz KM- a položilo ih je 10, a od 14 studentica položilo ih je 5.

SPOL	NISU POLOŽILI	POLOŽILI	UKUPNO
MUŠKARCI	16 (62%)	10 (38%)	26
ŽENE	9 (64%)	5 (36%)	14
UKUPNO	25	15	40

Da li se studenti statistički značajno razlikuju od studentica prema uspješnosti na praktičnom ispitu iz KM-a?

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Najprije je potrebno izračunati teoretske frekvencije. Teoretske frekvencije se pod pretpostavkom posve slučajne raspodjele izračunavaju na način da se suma frekvencija retka u kojoj se nalazi teoretska frekvencija pomnoži sa sumom frekvencija stupca u kojoj se nalazi teoretska frekvencija te se izračunati produkt podijeli sa sumom frekvencija svih kategorija.

NEPARAMETRIJSKE METODE II.

*Analiza razlika prema nominalnim varijablama
opažene frekvencije*

SPOL	NISU POLOŽILI	POLOŽILI	UKUPNO
MUŠKARCI	16 (62%)	10 (38%)	26
ŽENE	9 (64%)	5 (36%)	14
UKUPNO	25	15	40

$$26 \times 25 / 40 = 16,25 \quad 26 \times 15 / 40 = 9,75 \quad 14 \times 25 / 40 = 8,75 \quad 14 \times 15 / 40 = 5,25$$

teoretske frekvencije

SPOL	NISU POLOŽILI	POLOŽILI	UKUPNO
MUŠKARCI	16,25 (62,5%)	9,75 (37,5%)	26
ŽENE	8,75 (62,5%)	5,25 (37,5%)	14
UKUPNO	25	15	40

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Statistička značajnost razlike između opaženih frekvencija i teoretskih frekvencija testira se pomoću hi-kvadrat distribucije s $df=(n-1) \times (m-1)$ stupnjeva slobode gdje n označava broj redaka, a m broj stupaca. Hi-kvadrat vrijednost (χ^2) se izračunava sljedećom formulom

$$\chi^2 = \sum_{g=1}^N \frac{(f_{og} - f_{tg})^2}{f_{tg}}$$

gdje je

- ➔ f_{og} - opažena frekvencija polja g ($g=1, \dots, N$)
- ➔ f_{tg} - teoretska frekvencija polja g ($g=1, \dots, N$)
- ➔ N - ukupan broj polja

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

U ovom primjeru hi-kvadrat vrijednost se izračunava na sljedeći način:

$$\chi^2 = \sum_{g=1}^N \frac{(f_{og} - f_{tg})^2}{f_{tg}} = \frac{(16-16,25)^2}{16,25} + \frac{(10-9,75)^2}{9,75} + \frac{(9-8,75)^2}{8,75} + \frac{(5-5,25)^2}{5,25} = 0,03$$

$$df = (2-1) \cdot (2-1) = 1$$

$$\chi^2 = 0,03 \quad df = 1 \quad p < 0,864$$

Izračunata hi-kvadrat vrijednost i pripadajuća minimalna pogreška statističkog zaključka upućuju da se udjeli studenata i studentica ne razlikuju statistički značajno.

STATISTICA 7

Višedimenzionalno grupiranje kvalitativnih podataka

Višedimenzionalno grupiranje kvalitativnih podataka izvodi se slijedom koraka: padajući izbornik *Statistics* → *Basic Statistics/Tables* → *Tables and banners*. Odabirom opcije *Specify Tables (select variables)* potrebno je odabrati varijable prema kojima će se izvršiti grupiranje.

Zadatak - Pokrenite datoteku *TA.sta*! Utvrdite frekvencije tjelesno aktivnih i tjelesno neaktivnih muškaraca, te tjelesno aktivnih i tjelesno neaktivnih žena u inicijalnom testiranju (*TA-I*)!

STATISTICA 7

Hi-kvadrat test za dva ili više nezavisnih uzoraka

Hi-kvadrat test za 2×2 tablice izvodi se slijedom koraka: padajući izbornik *Statistics* \rightarrow *Nonparametrics* \rightarrow *2 \times 2 Tables* (*X²/V²/Phi², McNemar, Fisher exact*). U predloženu tablicu je potrebno unijeti opažene frekvencije te odabrati opciju *Summary*. Iz tablice se može očitati χ^2 vrijednost i odgovarajuća pogreška statističkog zaključka (*Chi-square (df=1)*) ili Yatesova korigirana vrijednost (*Yates corrected Chi-square*).

Zadatak - Na temelju prethodno provedenog grupiranja podataka hi-kvadrat testom utvrdite da li se udjeli tjelesno aktivnih i tjelesno neaktivnih osoba statistički značajno razlikuju po spolu na razini pogreške 5% !

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

McNemarovim hi-kvadrat testom za dva zavisna uzorka utvrđuje se statistička značajnost razlika jedne grupe entiteta u dva mjerenja prema nekoj nominalnoj varijabli. Na ovaj način se mogu testirati promjene entiteta u određenom nominalnom svojstvu između dvije vremenske točke.

Prije provođenja hi-kvadrat testa za dva zavisna uzorka potrebno je izvršiti dvodimenzionalno grupiranje entiteta pri čemu svaka varijabla predstavlja jedno mjerenje.

Pri tome kategorija prvog retka tablice obvezno treba odgovarati kategoriji drugog stupca, a kategorija drugog retka kategoriji prvog stupca.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Primjer: Od ukupno 75 studenata prvi kolokvij iz KM-a je položilo 41, a drugi 58.

		2. kolokvij		UKUPNO
		nije položio	položio	
1. kolokvij	položio	7	34	41
	nije položio	10	24	34
	UKUPNO	17	58	75

Da li se uspješnost studenata na drugom kolokviju statistički značajno povećala u odnosu na prvi kolokvij?

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

2. kolokvij

		2. kolokvij			
		nije položio	položio	UKUPNO	
1. kolokvij	{	položio	A	B	A+B
		nije položio	C	D	C+D
	UKUPNO	A+C	B+D	A+B+C+D	

Frekvencija polja A i frekvencija polja D predstavljaju entitete koji su promijenili svoj uspjeh na drugom kolokviju u odnosu na prvi kolokvij. Pod pretpostavkom da nije došlo ni do kakvih promjena navedene frekvencije bi trebale biti jednake. S obzirom na to teoretska frekvencija za polja A i D iznosi $(A+D)/2$, a za polja B i C teoretske frekvencije su jednake opaženim frekvencijama.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Teoretske frekvencije kategorija pod pretpostavkom da nije došlo ni do kakvih promjena iznose

		2. kolokvij		UKUPNO
		nije položio	položio	
1. kolokvij	položio	15,5	34	41
	nije položio	10	15,5	34
	UKUPNO	17	58	75

zato jer je $(A+D)/2=(7+24)/2=15,5$.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Statistička značajnost razlike između opaženih frekvencija i teoretskih frekvencija se testira pomoću hi-kvadrat distribucije s $df=1$ stupnjem slobode. Hi-kvadrat vrijednost (χ^2) se izračunava sljedećom formulom

$$\chi^2 = \sum_{g=1}^N \frac{(f_{og} - f_{tg})^2}{f_{tg}}$$

gdje je

- ➔ f_{og} - opažena frekvencija polja g ($g=1, \dots, N$)
- ➔ f_{tg} - teoretska frekvencija polja g ($g=1, \dots, N$)
- ➔ N - ukupan broj polja

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

U ovom primjeru hi-kvadrat vrijednost izračunava se na sljedeći način:

$$\chi^2 = \sum_{g=1}^N \frac{(f_{og} - f_{tg})^2}{f_{tg}} = \frac{(7-15,5)^2}{15,5} + \frac{(34-34)^2}{34} + \frac{(10-10)^2}{10} + \frac{(24-15,5)^2}{15,5} = 8,26$$

$$df = 1$$

$$\chi^2 = 8,26 \quad df = 1 \quad p < 0,004$$

Izračunata hi-kvadrat vrijednost i pripadajuća minimalna pogreška statističkog zaključka upućuju da je došlo do statistički značajne ($p < 0,05$) promjene u uspješnosti.

STATISTICA 7

Višedimenzionalno grupiranje kvalitativnih podataka

Višedimenzionalno grupiranje kvalitativnih podataka izvodi se slijedom koraka: padajući izbornik *Statistics* → *Basic Statistics/Tables* → *Tables and banners*. Odabirom opcije *Specify Tables (select variables)* potrebno je odabrati varijable prema kojima će se izvršiti grupiranje.

Zadatak - Pokrenite datoteku *TA.sta*! Grupirajte podatke istovremeno prema tjelesnoj aktivnosti u inicijalnom mjerenju (*TA-I*) i tjelesnoj aktivnosti u finalnom mjerenju (*TA-F*)!

STATISTICA 7

Hi-kvadrat test za dva zavisna uzorka

Hi-kvadrat test za 2×2 tablice izvodi se slijedom koraka: padajući izbornik *Statistics* \rightarrow *Nonparametrics* \rightarrow *2 \times 2 Tables* (*X²/V²/Phi², McNemar, Fisher exact*). U predloženu tablicu je potrebno unijeti opažene frekvencije te odabrati opciju *Summary*. Iz tablice se mogu očitati rezultati hi-kvadrat testa za dva zavisna uzorka (*McNemar Chi-square (A/D)*).

Zadatak - Na temelju prethodno provedenog grupiranja podataka hi-kvadrat testom utvrdite da li se udjeli tjelesno aktivnih i tjelesno neaktivnih osoba statistički značajno razlikuju u inicijalno i finalnom mjerenju na razini pogreške 5% !

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Opća pravila za korištenje hi-kvadrat testa:

- ➔ polja moraju biti disjunktna, odnosno niti jedan entitet ne smije biti član više od jednog polja/kategorije, i
- ➔ najviše 20% teoretskih frekvencija smije biti manje od 5, a niti jedna ne smije biti manja od 1.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema nominalnim varijablama

Pravila za korištenje hi-kvadrat testa na 2×2 tablicama:

- ➔ sve teoretske frekvencije moraju biti veće od 10, i
- ➔ ako je neka teoretska frekvencija manja od 10, ali veća od 5, neki autori predlažu da se koristi Yates-ovu korekciju za kontinuitet (svaku opaženu frekvenciju koja je veća od teoretske potrebno je smanjiti za 0,5 te svaku opaženu frekvenciju koja je manja od teoretske povećati za 0,5).

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Odabir neparametrijske metode zavisit će o ciljnoj usmjerenosti obrade podataka, o tipu korištenih varijabli te o broju i zavisnosti/nezavisnosti uzoraka entiteta.

Primjena neparametrijskih metoda na intervalnim i omjernim varijablama najčešće uzrokuje gubitak informacija o entitetima (primjerice, ako se rezultati ispitanika u utrci na 100 metara transformiraju u rangove, gubi se informacija o veličini razlika među ispitanicima) stoga je takva primjena opravdana samo u slučaju kada prikupljeni podaci ne zadovoljavaju neku od pretpostavki odabrane parametrijske metode.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

PARAMETRIJSKA METODA	NEPARAMETRIJSKA METODA
t-test za nezavisne uzorke	Mann-Whitneyev U test
t-test za zavisne uzorke	Sign test / Wilcoxonov test ekvivalentnih parova
univarijatna analiza varijance	Kruskal-Wallisov test

(Parametrijske metode i odgovarajuće neparametrijske metode predviđene za obradu podataka izraženih ordinalnim, intervalnim ili omjernim skalama)

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama
Mann-Whitneyev U test je neparametrijska metoda kojom se utvrđuje statistička značajnost razlike između dva nezavisna uzorka entiteta u jednoj varijabli.

Originalni rezultati ispitanika se transformiraju u rangove te se izračuna suma rangova za svaku grupu zasebno. Što se grupe ispitanika prema sumi rangova više razlikuju, veća je vjerojatnost da je razlika između grupa statistički značajna.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Primjer: Šest učenika i četiri učenice su izmjereni nekim motoričkim testom x . Mann-Whitneyevim U testom je potrebno utvrditi statističku značajnost razlike između učenika i učenica prema rezultatima u testu x .

ISP.	x	<i>spol</i>
1	70	Z
2	150	M
3	100	M
4	120	M
5	105	Z
6	94	Z
7	120	M
8	60	Z
9	99	M
10	100	M

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Primjer - nastavak: Rezultati se transformiraju u rangove pri čemu rang 1 označava ispitanika s najmanjim rezultatom. Potom se zbroje rangovi za svaku grupu posebno. Suma rangova učenika (T_1) iznosi 42, a suma rangova učenica (T_2) 13. Sumu rangova korisno je u svrhu interpretacije podijeliti s brojem entiteta u odgovarajućoj grupi, tj. izračunati prosječan rang grupe. Prosječan rang učenika iznosi $42/6=7$, a učenica $13/4=3,25$.

ISP.	x	<i>spol</i>
1	2	Z
2	10	M
3	5,5	M
4	8,5	M
5	7	Z
6	3	Z
7	8,5	M
8	1	Z
9	4	M
10	5,5	M

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Statistička značajnost razlike između grupa testira se putem normalne distribucije pri čemu se odgovarajuća z-vrijednost izračunava formulom

$$z = \left| \frac{T_1 - \frac{n_1(n+1)}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}} \right| \quad \text{ili} \quad z = \left| \frac{T_2 - \frac{n_2(n+1)}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}} \right|$$

gdje je

- ➔ T_1 i T_2 - suma rangova prve, odnosno druge grupe
- ➔ n_1 i n_2 - broj entiteta u prvoj, odnosno drugoj grupi
- ➔ n - ukupan broj entiteta

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

U ovom primjeru z-vrijednost se izračunava na sljedeći način:

$$z = \left| \frac{T_1 - \frac{n_1(n+1)}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}} \right| = \left| \frac{42 - \frac{6(10+1)}{2}}{\sqrt{\frac{6 \cdot 4(10+1)}{12}}} \right| = 1,92$$

Izračunatoj z-vrijednosti odgovara minimalna pogreška statističkog zaključka $p=0,053$ što upućuje na zaključak da se učenici i učenice ne razlikuju statistički značajno ni na razini pogreške 5% ni na razini pogreške 1%.

STATISTICA 7

Mann-Whitneyev U test

Mann-Whitney U test izvodi se slijedom koraka: padajući izbornik *Statistics* → *Nonparametrics* → *Comparing two independent samples (groups)* → *Mann-Whitney U test*. U dijaloškom okviru koji se pokreće odabirom opcije *Variables* potrebno je označiti selektorsku varijablu (*Indep. (grouping) variable*) i jednu ili više zavisnih varijabli (*Dependent variable list*).

Zadatak - U datoteci *Pejcic-318.sta* uz pogrešku $p=0,01$ testirajte da li se učenici *III.* i *IV.* razreda (*RAZRED*) statistički značajno razlikuju prema rezultatima u varijabli *MPOL!*

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Sign test ili *test predznaka* je neparametrijska metoda kojom se utvrđuje statistička značajnost razlike između dva zavisna uzorka entiteta u jednoj varijabli (npr. jednog uzorka entiteta mjenenog u dvije vremenske točke).

Za svakog ispitanika se izračuna razlika između rezultata u prvom i drugom mjerenju te se utvrdi predznak svake pojedine razlike. Što je veći nesrazmjer u broju pozitivnih i negativnih predznaka razlika, to je veća vjerojatnost da je razlika statistički značajna.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Primjer: Ocijenjena je izvedba koluta naprijed desetorice učenika na početku (*I*) i na kraju (*F*) školske godine. Sign testom je potrebno utvrditi statističku značajnost razlike između ocjena učenika u inicijalnom i finalnom stanju.

ISP.	<i>I</i>	<i>F</i>
1	2	3
2	4	5
3	3	4
4	2	4
5	2	5
6	3	2
7	2	1
8	1	3
9	2	2
10	1	4

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Izračuna se razlika između ocjena u finalnom i inicijalnom stanju za svakog učenika te se utvrdi predznak svake razlike. Prebroje se pozitivni i negativni predznaci. Manji od dva prethodno utvrđena broja se testira pomoću tablice graničnih vrijednosti za test predznaka. Ako je utvrđeni broj predznaka s manjim udjelom manji ili jednak graničnoj vrijednosti u tablici prihvaća se alternativna hipoteza.

ISP.	<i>I</i>	<i>F</i>	<i>F-I</i>	<i>pr.</i>
1	2	3	1	+
2	4	5	1	+
3	3	4	1	+
4	2	4	2	+
5	2	5	3	+
6	3	2	-1	-
7	2	1	-1	-
8	1	3	2	+
9	2	2	0	
10	1	4	3	+

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

U ovom je primjeru broj pozitivnih predznaka bio 7, a broj negativnih predznaka 2. Granična vrijednost za 9 entiteta (u obzir se ne uzimaju entiteti bez predznaka, tj. bez promjene u rezultatu) na razini statističke pogreške 5% iznosi 1, a na razini 1% iznosi 0. Pošto utvrđeni broj predznaka s manjim udjelom iznosi 2, te je stoga veći od graničnih vrijednosti, nije moguće zaključiti da se učenici u inicijalnom i finalnom stanju statistički značajno razlikuju.

STATISTICA 7

Sign test

Sign test se izvodi slijedom koraka: padajući izbornik *Statistics* → *Nonparametrics* → *Comparing two dependent samples (variables)* → *Sign test*. U dijaloškom okviru koji se pokreće odabirom opcije *Variables* potrebno je označiti varijablu prvog mjerenja (*First variable list*) i varijablu drugog mjerenja (*Second variable list*).

Zadatak - U datoteci *POD.sta* uz pogrešku $p=0,01$ testirajte da li se rezultati prvog mjerenja (*MKUS_I*) i drugog mjerenja (*MKUS_F*) statistički značajno razlikuju!

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Wilcoxonov test ekvivalentnih parova je neparametrijska metoda kojom se utvrđuje statistička značajnost razlike između dva zavisna uzorka entiteta u jednoj varijabli (npr. jednog uzorka entiteta mjenenog u dvije vremenske točke).

Uvjet za provođenje Wilcoxonovog testa jest da je varijabla kvantitativna (intervalna ili omjerna). Za svakog ispitanika se izračuna razlika između rezultata u prvom i drugom mjerenju te se apsolutne vrijednosti razlika prema veličini rangiraju. Što je veći nesrazmjer između prvog i drugog mjerenja u sumi rangova apsolutnih vrijednosti pozitivnih i negativnih razlika, veća je vjerojatnost da je razlika statistički značajna.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Primjer: Deset učenika je testirano nekim testom motorike na početku (I) i na kraju (F) školske godine. Wilcoxonovim testom ekvivalentnih parova potrebno je utvrditi statističku značajnost razlike između rezultata učenika u inicijalnom i finalnom stanju.

ISP.	I	F
1	115	120
2	145	195
3	120	110
4	92	122
5	82	99
6	116	178
7	100	90
8	70	105
9	104	188
10	110	145

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Izračuna se razlika između rezultata u finalnom i inicijalnom stanju za svakog učenika. Apsolutne vrijednosti razlika (bez obzira na predznak) se rangiraju. Potom se zasebno zbroje rangovi razlika s pozitivnim predznakom i rangovi razlika s negativnim predznakom. Ako je manja suma rangova manja od granične vrijednosti u tablici graničnih vrijednosti za Wilcoxonov test prihvaća se alternativna hipoteza.

ISP.	<i>I</i>	<i>F</i>	<i>F-I</i>	<i>pr.</i>
1	115	120	5	1
2	145	195	50	8
3	120	110	-10	2,5
4	92	122	30	5
5	82	99	17	4
6	116	178	62	9
7	100	90	-10	2,5
8	70	105	35	6,5
9	104	188	84	10
10	110	145	35	6,5

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

U ovom primjeru suma rangova pozitivnih razlika iznosi 122, a negativnih 5. Granična vrijednost za 10 entiteta (u obzir se ne uzimaju entiteti bez promjene u rezultatu) na razini statističke pogreške 5% iznosi 8, a na razini 1% iznosi 3. Pošto utvrđena (manja) suma rangova iznosi 5, moguće je zaključiti da se učenici u inicijalnom i finalnom stanju statistički značajno razlikuju na razini pogreške 5%, ali ne i na razini pogreške 1%.

STATISTICA 7

Wilcoxonov test ekvivalentnih parova

Wilcoxonov test ekvivalentnih parova izvodi se slijedom koraka: padajući izbornik *Statistics* → *Nonparametrics* → *Comparing two dependent samples (variables)* → *Wilcoxon matched pairs test*. U dijaloškom okviru koji se pokreće odabirom opcije *Variables* potrebno je označiti varijablu prvog mjerenja (*First variable list*) i varijablu drugog mjerenja (*Second variable list*).

Zadatak - U datoteci *POD.sta* uz pogrešku $p=0,01$ testirajte da li se rezultati prvog mjerenja (*MKUS_I*) i drugog mjerenja (*MKUS_F*) statistički značajno razlikuju!

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama
Kruskal-Wallisov test je neparametrijska metoda kojom se utvrđuje statistička značajnost razlike između dva ili više nezavisnih uzoraka entiteta u jednoj varijabli.

Originalni rezultati ispitanika se transformiraju u rangove te se izračuna suma rangova za svaku grupu zasebno. Što se grupe ispitanika prema sumi rangova više razlikuju, veća je vjerojatnost da je razlika između grupa statistički značajna.

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Primjer: Pet sportaša i pet nesportaša je izmjereno nekim motoričkim testom x . Kruskal-Wallisovim testom potrebno je testirati statističku značajnost razlike između sportaša i nesportaša prema rezultatima u testu x .

ISP.	x	<i>sport</i>
1	115	DA
2	145	DA
3	120	DA
4	92	NE
5	82	NE
6	116	DA
7	100	NE
8	70	NE
9	104	NE
10	110	DA

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Primjer - nastavak: Rezultati se transformiraju u rangove pri čemu rang 1 označava ispitanika s najmanjim rezultatom. Potom se zbroje rangovi za svaku grupu zasebno. Suma rangova sportaša (T_s) iznosi 40, a suma rangova nesportaša (T_n) 15. Sumu rangova je u svrhu interpretacije korisno podijeliti s brojem entiteta u odgovarajućoj grupi, tj. izračunati prosječan rang grupe. Prosječan rang sportaša iznosi $40/5=8$, a nesportaša $15/5=3$.

ISP.	x	<i>sport</i>
1	7	DA
2	10	DA
3	9	DA
4	3	NE
5	2	NE
6	8	DA
7	4	NE
8	1	NE
9	5	NE
10	6	DA

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

Statistička značajnost razlike između grupa testira se pomoću hi-kvadrat distribucije s $k-1$ stupnjeva slobode gdje je k broj grupa. Odgovarajuća H vrijednost izračunava se formulom

$$H = \frac{12}{n(n+1)} \cdot \sum_{g=1}^k \frac{T_g^2}{n_g} - 3(n+1)$$

gdje je

→ n - ukupan broj entiteta

→ T_g - suma rangova grupe g ($g=1, \dots, k$)

→ n_g - broj entiteta grupe g ($g=1, \dots, k$)

NEPARAMETRIJSKE METODE II.

Analiza razlika prema ordinalnim i kvantitativnim varijablama

U ovom primjeru H-vrijednost izračunava se na sljedeći način:

$$H = \frac{12}{10(10+1)} \cdot \left(\frac{40^2}{5} + \frac{15^2}{5} \right) - 3(10+1) = 6,82$$

Izračunatoj H-vrijednosti odgovara minimalna pogreška statističkog zaključka $p=0,009$ što upućuje na zaključak da se sportaši i nesportaši statistički značajno razlikuju na razini pogreške 1%.

STATISTICA 7

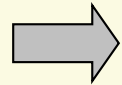
Kruskal-Wallisov test

Kruskal-Wallisov test izvodi se slijedom koraka: padajući izbornik *Statistics* → *Nonparametrics* → *Comparing multiple indep. samples (groups)* → *Summary*. U dijaloškom okviru koji se pokreće odabirom opcije *Variables* potrebno je označiti selektorsku varijablu (*Indep. (grouping) variable*) i jednu ili više zavisnih varijabli (*Dependent variable list*).

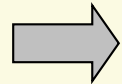
Zadatak - U datoteci *Pejcic-318.sta* uz pogrešku $p=0,01$ testirajte da li se rezultati učenika *I.*, *II.*, *III.* i *IV.* razreda (*RAZRED*) statistički značajno razlikuju prema rezultatima u varijabli *MPOL*!

NEPARAMETRIJSKE METODE II.

Literatura za pripremanje kolokvija



Petz, B. (2002). *Osnovne statističke metode za nematematičare*. Jastrebarsko: Naklada Slap, str. 249-273, 321-322, 327-330, 333-337, 338-341.



Šošić, I. (2004). *Primijenjena statistika*. Zagreb: Školska knjiga, str. 331-335, 340-348, 352-366, 369-377.