

# Analiza podataka korištenjem neparametarskih metoda

---

Levec, Rene

Undergraduate thesis / Završni rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Organization and Informatics / Sveučilište u Zagrebu, Fakultet organizacije i informatike**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:211:680339>

Rights / Prava: [Attribution-NoDerivs 3.0 Unported/Imenovanje-Bez prerada 3.0](#)

Download date / Datum preuzimanja: **2024-12-26**



Repository / Repozitorij:

[Faculty of Organization and Informatics - Digital Repository](#)



**SVEUČILIŠTE U ZAGREBU  
FAKULTET ORGANIZACIJE I INFORMATIKE  
VARAŽDIN**

**Rene Levec**

**ANALIZA PODATAKA KORIŠTENJEM  
NEPARAMETARSKIH METODA**

**ZAVRŠNI RAD**

**Varaždin, 2020.**

**SVEUČILIŠTE U ZAGREBU**  
**FAKULTET ORGANIZACIJE I INFORMATIKE**  
**V A R A Ž D I N**

**Rene Levec**

**Matični broj: 0016100329**

**Studij: Poslovni sustavi**

**ANALIZA PODATAKA KORIŠTENJEM NEPERAMETARSKIH**  
**METODA**

**ZAVRŠNI RAD**

**Mentor/Mentorica:**

Prof. dr. sc. Jasminka Dobša

**Varaždin, rujan 2020.**

*Rene Levec*

### **Izjava o izvornosti**

Izjavljujem da je moj završni/diplomski rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristio drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

*Autor/Autorica potvrdio/potvrdila prihvaćanjem odredbi u sustavu FOI-radovi*

---

## **Sažetak**

Tema ovog rada su neparametrijske metode u statistici. Bilo mi je zanimljivo vidjeti u kojim situacijama i sa kojim podacima su primjenjive metode koje se rjeđe koriste u statistici.

Ovisno o dostupnim podacima odabrao sam nekoliko neparametrijskih metoda sa naglaskom na rješavanju stvarnog problema, dokazivanja hipoteze i „rješavanja“ stvarnog pitanja koje si istraživanjem postavljam.

Koristio sam se online istraživanjem / online anketom kao i dostupnim statističkim podacima sa nekoliko web stranica koje su navedene u izvorima.

**Ključne riječi:** neparametrijske metode, test predznaka, MWW test, test homogenosti niza, X2 test, Kruskal Wallisov test, Friedmanov test

# Sadržaj

Sadržaj.....	III
1. Uvod.....	1
2. Metode i tehnike rada.....	2
3. Odabrani neparametrijski testovi.....	3
3.1. Test predznaka.....	4
3.1.1. Istraživanje.....	4
3.2. MWW Test.....	7
3.2.1. Istraživanje.....	7
3.2.1.1. Izračun.....	9
3.3. Test homogenosti niza.....	10
3.3.1. Istraživanje.....	10
3.3.1.1. Izračun.....	12
3.4. $\chi^2$ Test.....	14
3.4.1. Istraživanje.....	14
3.4.1.1. Izračun.....	15
3.5. Kruskal Wallisov test.....	17
3.5.1. Istraživanje.....	17
3.5.1.1. Izračun.....	18
3.6. Friedmanov test.....	20
3.6.1. Istraživanje.....	20
3.6.1.1. Izračun.....	21
4. Zaključak.....	23
Popis literature:.....	24
Popis slika.....	25
Popis tablica.....	25

# Uvod

Kada pokušavamo saznati nešto više o pojedinoj populaciji koristeći statističke metode gotovo uvijek postavljamo određenu hipotezu. Hipoteze su usko povezane sa pretpostavkom o distribuciji te populacije, npr. rezultati testa na nekom kolegiju će najčešće imati normalnu distribuciju; jednostavnim rječnikom rečeno možemo pretpostaviti da će najveći broj studenata imati neku „srednju“ ocjenu (dobar 3) dok će manji broj studenata imati ocjene vrlo dobar i odličan, ili prema drugom kraju skale dovoljan i nedovoljan. Time pokušavamo dokazati hipotezu da će najveći broj studenata koji su adekvatno ispunjavali svoje studentske obveze u vidu predavanja i kolokvija bez velikih napora postići zadovoljavajuću ocjenu. Ovaj vrlo pojednostavljen princip primjenjiv je na većinu populacija koje istražujemo, ali naravno ne može se primijeniti uvijek i svugdje i na svaku populaciju.

Vrlo često u statističkom istraživanju nailazimo na podatke koji su distribuirani na nepoznat način i za koje ne možemo pretpostaviti da pripadaju distribuciji normalnog oblika. Podatci takvih distribucija su često mjereni kvalitativno te je na njih nemoguće primijeniti parametarske postupke testiranja. Statistička istraživanja takvih skupova ovise o neparametrijskim metodama. Neparametrijska statistika koristi se kod podataka koji imaju mali broj ispitanika u istraživanju, podaci su izraženi na nominalnim ili ordinalnim mjernim ljestvicama, raspodjela se značajno razlikuje od normalne. (Šošić, 2004., str. 331.)

Parametarski testovi se služe mjernim podacima koji se normalno distribuiraju i upotrebljavaju se kod podataka izraženih intervalnom i mjernom skalom dok kod neparametarskih testova nije važno je li populacija normalno distribuirana i koriste se kvalitativni podaci koji se mjere nominalnom i ordinalnom skalom. Neparametarske testove dopušteno je i moguće primijeniti i pri obradi podataka koji su izraženi intervalnim i mjernim skalama te su normalno distribuirani, ali bi takav postupak bio neracionalan jer na taj način se namjerno gubi niz informacija i precizniji postupak se zamjenjuje aproksimativnim postupkom. Statističari kažu da je snaga neparametarskih testova manja od snage parametarskih testova. Ako između dvije populacije postoji razlika, to će se obično preciznije i uspješnije ustanoviti pomoću parametarskih nego neparametarskih testova. (Petz, Kolesarić, Ivanec, 2012., str. 321.)

Zbog njihove specifičnosti izbor iz neparametrijskih metoda mi se učinio zanimljivom temom za rad.

## Metode i tehnike rada

Za izradu ovog rada koristio sam metodu online ankete, obzirom na specifičnosti metoda koje sam želio obraditi uzeo sam za neke od njih specifične teme u odgoju djece čime prikupljeni podatci imaju više kvalitativna nego kvantitativna svojstva. Online anketu sam postavio u nekoliko grupa vezanih za odgoj djece i na taj način prikupio podatke. I ovim putem zahvaljujem svim anonimnim sudionicama na njihovom doprinosu.

Za druge teme sam istraživao tri stranice vezane uz statističke podatke:

- Stranice Državnog zavoda za statistiku
- Portal otvorenih podataka Republike Hrvatske
- Portal o kvaliteti života Numbeo



# Odabrani neparametrijski testovi

Statistička i općenito znanstvena istraživanja temelje se na hipotezama. Kada želimo saznati više o nekoj populaciji ili uzorku postavljamo neku hipotezu: djeca u urbanim sredinama pokazuju bolje rezultate na standardiziranim testovima od djece u ruralnim sredinama (ukoliko testiramo kvalitetu školovanja u urbanim/ruralnim sredinama), studenti koji redovito izvršavaju svoje obveze (predavanja/vježbe/kolokvije) postižu u prosjeku bolje rezultate od kolega koji samo pristupaju ispitima u ispitnim rokovima, kvaliteta života je bolja u prigradskim urbanim sredinama od kvalitete života u strogo urbanim sredinama, itd...

Hipoteze koje istraživač postavlja usko su vezane uz populaciju (podatke) koje promatra, da bismo ocjenili kvalitetu života u nekoj sredini moramo definirati koji su to parametri (podatci) koji čine kvalitetu života (niska stopa kriminala, cijene prehrambenih i neprehrambenih artikala, kvaliteta i blizina škola i vrtića, dostupnost sportskih sadržaja, gustoća prometa, gustoća naseljenosti, kvaliteta komunalnih usluga...), ukoliko su podatci kvalitativne prirode (zadovoljstvo stanara određenim kvartom) moramo naći način da ih kvantitativno izrazimo kako bismo ih mogli statistički obraditi. Vrlo često hipoteza ovisi i o pretpostavci distribucije tih podataka, tj. na temelju pretpostavke o distribuciji podataka možemo kvalitetno postaviti hipotezu. Ali što napraviti kada je pretpostavka da podatci neće imati „uobičajenu“ distribuciju?

Kako postaviti hipotezu o povezanosti između „dobrog“ imuniteta i dojenja kod djece? Ili kako „obraniti“ hipotezu da su djeca koja su nošena od strane roditelja motorički spretnija i naprednija od druge djece? Obzirom na ogroman broj varijabli koji utječu na kompleksnost imuniteta ili motorike svjestan sam ograničenja u ovom statističkom istraživanju, kao i činjenice da ozbiljan znanstveni rad koji bi pokušao dokazati ili opovrgnuti neku od ovih hipoteza bi morao sadržavati puno veći broj parametara o djeci koja sudjeluju u istraživanju kako bi dao imalo respektabilne rezultate.

S druge strane obzirom na odabranu temu trudio sam se naći podatke koji bi na najbolji način odgovarali metodi koju obrađujem, mada sam bio svjestan da ovisno o podacima i distribuciji treba odabrati statističku metodu koja će dati najtočniji rezultat.

Problem koji me najviše intrigirao pri pisanju rada je kako kvalificirati kvalitativne podatke kako bi respektabilno obrazložio svaku od odabranih metoda.

## Test predznaka

Test predznaka ili test hipoteze o medijanu osnovnog skupa. Medijan dijeli uređen niz vrijednosti numeričke ili rang varijable na dva jednakobrojna dijela. Ako raspoloživi podatci čine uzorak pomoću podataka iz uzorka može se testirati hipoteza o pretpostavljenoj vrijednosti medijana osnovnog skupa. Test može biti dvosmjerni ili jednosmjerni. Odluka se donosi pomoću binomne distribucije vjerojatnosti. (Šošić, Zbirka zadataka iz statistike, 1998.)

Koristi se za testiranje jesu li dvije grupe jednake ili ne. Test predznaka koristi se kada su ovisni uzorci poredani u parovima, pri čemu su slučajne varijable međusobno neovisne. Temelji se na smjeru znaka plus i minus promatranja, a ne na njihovoj brojčanoj veličini. (Šošić, Zbirka zadataka iz statistike, 1998., str. 210.)

## Istraživanje

Proveo sam statističko istraživanje metodom online ankete<sup>1</sup> gdje su majke s dvoje ili više djece od kojih je bar jedno nošeno prema principima Babywearinga<sup>2</sup> odgovarale na pitanja o motorici kod djece. Hipoteza je da su djeca koja su nošena prema principima Babywearinga (ispravna nosiljka ili marama koja podržava anatomski položaj djeteta) motorički naprednija od djece koja nisu nošena, vožena su u kolicima ili su ih roditelji nosili u naručju.

Svjestan sam da bi u ovakvom primjeru istraživanje bilo bolje ukoliko bi se radilo o blizancima od kojih je jedan nošen a jedan ne, ili bar o braći i sestrama (tj. dvoje djece u istom odgojnom/životnom okruženju) od kojih je jedno nošeno, a jedno ne, jer su svi drugi parametri koji utječu na razvoj motorike kod djeteta, počevši od genetskih predispozicija, vremena kada je dijete rođeno, prohodalo, životnog okruženja, navika roditelja (koje sigurno nisu iste sa jednim i dvoje ili više djece) toliko raznolike i raznovrsne te da s toliko varijabli istraživanje nikako ne može biti u potpunosti točno. Ali obzirom da je cilj rada predstaviti neparametarsku statističku metodu testa predznaka, ipak sam se radije odlučio za ovu varijantu nego za izmišljen primjer.

Anketa je dala sljedeći rezultat, s tim da su uzeti u obzir samo odgovori majki sa dvoje ili više djece:

---

<sup>1</sup><https://docs.google.com/forms/d/e/1FAIpQLSe7Nn3wNagg4nuzNqiApJ41-8y0RjSxKvWxgEdbrBb-8rao5A/viewform>

<sup>22</sup><https://www.babycentre.co.uk/a562815/babywearing-for-beginners>

Tablica 1: rezultati istraživanja

Par	Da li ste nosili (babywearing) bar jedno od vaše djece?	Uočavate li, na temelju vašeg iskustva, da djeca koja su nošena (babywearing) imaju bolje motoričke vještine od djece koja nisu nošena?	Predznak
1	Da	Da	+
2	Da	Da	+
3	Da	Da	+
4	Da	Da	+
5	Da	Da	+
6	Da	Da	+
7	Da	Da	+
8	Da	Da	+
9	Da	Ne	-
10	Da	Ne	-
11	Da	Ne	-
12	Da	Da	+
13	Da	Da	+
14	Da	Ne	-
15	Da	Da	+
16	Da	Da	+
17	Da	Da	+
18	Da	Da	+
19	Da	Ne	-

Hipoteze ovog istraživanja su sljedeće,  $H_0$  da ne postoji razlika između djece koja su nošena i one koja nisu, a  $H_1$  je da razlika postoji.

Koristeći metodu testa predznaka formirao sam 19 parova od kojih jedan dio para čini nošeno dijete a drugi dio para dijete koje nije nošeno. Svjestan sam da je usporedba pod velikim subjektivnim utjecajem majke kao i da postoje izuzetno labave veze između parova (najčešće u dobu/spolu/godišnjem dobu rođenja djeteta). Par kod kojih je nošeno dijete koje je motorički naprednije označili smo oznakom +, a ono gdje nije oznakom -.

Kao što vidimo kod 14 parova smo dobili predznak +, a kod 5 predznak minus. Razina signifikantnosti od 5% jednaka je pouzdanosti od 95%, pa izračunavamo:

$$\alpha=0,05$$

$$\alpha/2=0,025$$

$$0,5-0,025=0,475$$

$$z=1,96$$

Binomnim testom provjeravamo vjerojatnost da binomna slučajna varijabla poprimi vrijednost manju ili jednaku 5, obzirom da imamo 5 negativnih predznaka. Ako je ta vjerojatnost veća od 0,05 (signifikantnosti), ostajemo pri nultoj hipotezi, u suprotnom se odbacuje nulta hipoteza i prihvaća alternativna hipoteza da postoji razlika između grupa.

Pri testiranju koristim excel funkciju =BINOM.DIST što prikazujemo u postupku niže:

$$C+ = 14$$

$$C- = 5$$

$$n = 19$$

$$=BINOM.DIST(5;19;0,5;1)$$

$$p\text{-vrijednost} = 0,03178 > 0,025$$

$$\alpha = 0.025$$

Za test na donju granicu vrijedi:

H0 - ne postoji razlika (DONJA GRANICA)

H1 - postoji razlika

Kako je p-vrijednost veća od 0,025, zaključujemo da ne postoji razlika između djece koja su nošena i onih koja nisu. Važno je napomenuti da je izračunata p-vrijednost vrlo bliska graničnoj, pa bi test trebalo ponoviti s većim brojem ispitanika.

## MWW Test

Mann-Whitney test (koji se naziva i Mann-Whitney-Wilcoxon (MWW) test) je neparametarski test za procjenu da li dva uzorka opažanja potječu iz iste distribucije. Donekle je sličan testu homogenih nizova, ali koristi više informacija pa se može smatrati jačim testom.

Ovim testom možemo testirati statistički značajnost razlike između medijana dvaju skupina. Koristimo ga kada imamo podatke koji nemaju normalnu distribuciju. Testiramo razliku između dviju grupa koja je veća od razlike koja nastaje zbog slučajne varijacije uzorka. Nulta hipoteza je da su dva uzorka uzeta iz populacija sa istim medijanom. Svi rezultati se rangiraju od najmanjih prema najvećim u obje grupe te se računa suma rangova za svaku grupu i uspoređuju se. Što su dvije grupe sličnije to će i suma rangova biti sličnija. Ako postoji razlika, jedna grupa će imati manji rang a druga veći i time zaključujemo da uzorci pripadaju populacijama kod kojih postoji statistički značajna razlika.

Hipoteze se oblikuju na različite načine. Primjerice, nultom se hipotezom pretpostavlja da su funkcije distribucija dvaju populacija jednake, a alternativnom da su različite. Pri tome nije nužno specificirati oblik distribucija. (Šošić, Zbirka zadataka iz statistike, 1998., str. 219.)

## Istraživanje

Proveo sam statističko istraživanje metodom online ankete<sup>3</sup> gdje su majke s dvoje ili više djece od kojih je bar jedno dojeno odgovarale na pitanja o imunitetu kod djece. Hipoteza je da su djeca koja su dojena imaju općenito bolje zdravstveno stanje i bolji imunitet od djece koja nisu dojena. Ova hipoteza je postavljena uzimajući u obzir 5% signifikantnost.

$H_0$  hipoteza ovog istraživanja je da ne postoji razlika u imunitetu djece koja su dojena i koja nisu, dok je  $H_1$  hipoteza da takva razlika postoji.

Podijelili smo rezultate ispitivanja u dvije skupine, prva skupina su djeca čije majke smatraju da dojena djeca (po vlastitom iskustvu) imaju bolji imunitet i opće zdravstveno stanje, a druga skupina su djeca čije majke misle da dojena djeca nemaju bolji imunitet i opće zdravstveno stanje ili su se odlučile za odgovor „Ne mogu reći“.

---

<sup>3</sup><https://docs.google.com/forms/d/e/1FAIpQLSe7Nn3wNagg4nuzNqiApJ41-8y0RjSXkVWxgEdbrBb-8rao5A/viewform>

Ovaj test koristi jednostavnu linearnu statistiku ranga (S).

$$S = \sum_{j=1}^n c_j a(R_j)$$

Gdje je  $R_j$  = rang j-te opservacije,  $a(R_j)$  = bodovi postignuti na rangu  $R_j$ ,  $c_j$  = indikator kojem uzorku pripada opservacija i  $n$  = ukupan broj opservacija.

Statistiku ranga standardiziramo u varijablu

$$Z = \frac{S - E_0(S)}{\sqrt{Var_0(S)}}$$

Tablica 2: rezultati istraživanja

Imate li dvoje ili više djece?	Da li ste dojili barem jedno od Vaše djece?	Uočavate li, na temelju Vašeg iskustva, da dojena djeca imaju bolji imunitet ili općenito bolje zdravstveno stanje od djece koja nisu dojena?	Ispitanik
Da	Da	Da	A
Da	Da	Da	B
Da	Da	Da	C
Da	Da	Da	D
Da	Da	Ne mogu reći	E
Da	Da	Ne mogu reći	F
Da	Da	Da	G
Da	Da	Ne mogu reći	H
Da	Da	Ne mogu reći	I
Da	Da	Ne	J
Da	Da	Da	K
Da	Da	Da	L
Da	Da	Ne	M
Da	Da	Da	N
Da	Da	Ne	O
Da	Da	Da	P
Da	Da	Ne mogu reći	R
Da	Da	Da	S
Da	Da	Da	T
Da	Da	Ne	U

Obje skupine smo rangirali te svakom ispitaniku dodijelili rang. Svjestan sam da bi opće zdravstveno stanje i imunitet djece prilikom rangiranja zahtijevao puno više parametara od ovih koje smo prikupili online anketom, npr. broj dana koliko je dijete bilo bolesno, težina bolesti, reakcije na cijepljenje, opći napredak djeteta i sl. Obzirom na upitnost etičnosti takvog istraživanja kao i na dostupnost podataka ovdje smo pribjegli izmišljenim rezultatima gdje smo svakom djetetu dodijelili rang prema općem zdravstvenom stanju i imunitetu.

Tablica 3: rangovi

Skupina „Da“	Rang	Skupina „Ne“	Rang
A	1	E	3
B	7	F	15
C	4	H	6
D	11	I	12
G	2	J	17
K	14	M	8
L	10	O	19
N	15	R	9
P	5	U	20
S	18		
T	13		
N1	11	N2	9
T1	100	T2	109

### Izračun

Koristeći izraz

$$z = \frac{|2T_i - N_i(N + 1)| - 2}{\sqrt{\frac{N_1 N_2 (N + 1)}{3}}}$$

gdje je  $T_i$  suma rangova po grupama, a  $N_i$  broj ispitanika u skupini iz koje smo uzeli  $T_i$ , te  $N$  ukupan broj ispitanika, možemo za Skupinu „Da“ izračunati da je:

$$z = \frac{|2T_i - N_i(N + 1)| - 2}{\sqrt{\frac{N_1 N_2 (N + 1)}{3}}} = \frac{|2 * 100 - 11(20 + 1)| - 2}{\sqrt{\frac{11 * 9(20 + 1)}{3}}} = 1,10$$

Budući da na razini značajnosti od 5%, statistički značajnima možemo smatrati samo one rezultate veće od 1,96, možemo zaključiti da je naš  $z$  premalen i da ostajemo kod nul hipoteze. Budući da je testna statistika manja od kritične teorijske vrijednosti za nivo signifikantnosti od 5% (1,96), ostajemo kod nulte hipoteze da ne postoji razlika između promatranih grupa u imunitetu.

## Test homogenosti niza

Test homogenosti niza ili test o slučajnoj seriji omogućuje nam da pretpostavimo da se članovi statističkog niza sastoje od dva modaliteta jednog obilježja ili je neko obilježje s više modaliteta pretvoreno u alternativno ili je riječ o dva znaka i tome slično. Homogenim podnizom smatra se svaka skupina (sekvenca) jednoga (istoga) oblika obilježja ili istoga znaka kojoj prethodi različiti oblik obilježja i koju slijedi različit oblik obilježja. Podniz se može sastojati od jednoga ili više članova. Test se temelji na odgovarajućoj sampling distribuciji. (Šošić, Zbirka zadataka iz statistike, 1998., str. 221.)

Zaključak o nekoj populaciji relevantan je samo ako je uzorak slučajan. Kako bismo bili sigurni da je uzorak slučajan možemo koristiti ovaj test.

Ako u nekom nizu od  $N$  članova postoje dvije varijable uzoraka ( $n=2$ ) koje se nasumično izmjenjuju, onda ne može biti manje od  $n$  broja nizova niti više od  $N$  broja nizova. Da bi dobili  $n$  nizova ( $n=2$ ) svi članovi iste varijable moraju biti grupirani skupa, ili u suprotnosti, ako se pravilno izmjenjuju onda ćemo dobiti  $N$  nizova. Obje te kombinacije su malo vjerojatne u stvarnom životu pa je za pretpostaviti da se te kombinacije neće dogoditi slučajno i da takav uzorak nikako nije slučajan. Znači ukoliko je broj nizova u populaciji ili izrazito mali ili izrazito visok možemo ustanoviti da postoji uzrok takvom uzorku.

Obično testiramo da li je broj nizova manji od onog broja koji bismo mogli smatrati slučajnošću i to najčešće uz 5% signifikatnost.

## Istraživanje

Koristeći online istraživanje o percepciji dobrobiti dojenja pokušao sam ustanoviti postoji li povezanost između pozitivnog mišljenja o dobrobitima dojenja kod žena koje su dojile i kod onih koje nisu dojile. Promatram dva niza sa pretpostavkom da žene koje doje od prije imaju pozitivno mišljenje o dobrobitima dojenja i da je njihovo mišljenje i nakon dojenja barem jednog djeteta i dalje pozitivno, tj da se nije promijenilo.

Hipoteza  $H_0$  glasi da ne postoji razlika u imunitetu djece koja su dojena u odnosu na djecu koja nisu dojena.

$H_1$  hipoteza glasi da postoji razlika u imunitetu djece koja su dojena u odnosu na djecu koja nisu dojena.



Tablica 4: rezultati istraživanja

Imate li dvoje ili više djece?	Da li ste dojili barem jedno od Vaše djece?	Uočavate li, na temelju Vašeg iskustva, da dojena djeca imaju bolji imunitet ili općenito bolje zdravstveno stanje od djece koja nisu dojena?	Ispitanik
Da	Da	Da	A
Da	Da	Da	B
Da	Ne	Da	C
Da	Da	Da	D
Da	Da	Ne mogu reći	E
Da	Da	Ne mogu reći	F
Da	Da	Da	G
Da	Da	Ne mogu reći	H
Da	Da	Ne mogu reći	I
Da	Da	Ne	J
Da	Ne	Da	K
Da	Ne	Da	L
Da	Da	Ne	M
Da	Da	Da	N
Da	Da	Ne	O
Da	Da	Da	P
Da	Da	Ne mogu reći	R
Da	Da	Da	S
Da	Da	Da	T
Da	Da	Ne	U

Obje skupine smo rangirali te svakom ispitaniku dodijelili rang kao u prethodnom primjeru. Kao što sam već napomenuo svjestan sam da bi opće zdravstveno stanje i imunitet djece prilikom rangiranja zahtijevao puno više parametara od ovih koje smo prikupili online anketom, npr. broj dana koliko je dijete bilo bolesno, težina bolesti, reakcije na cijepljenje, opći napredak djeteta i sl. Obzirom na upitnost etičnosti takvog istraživanja kao i na dostupnost podataka ovdje smo pribjegli izmišljenim rezultatima gdje smo svakom djetetu dodijelili rang prema općem zdravstvenom stanju i imunitetu.

Slaganjem odgovora prema rangu djeteta dobivamo sljedeće podatke:

Tablica 5: rangovi

Skupina „Da“	Rang	Skupina „Ne“	Rang
A	1	E	3
B	7	F	15
C	4	H	6
D	11	I	12
G	2	J	17
K	14	M	8
L	10	O	19
N	15	R	9
P	5	U	20
S	18		
T	13		
N1	11	N2	9
T1	100	T2	109

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	15	17	18	19	20
Ispitanik	A	G	E	C	P	H	B	M	R	L	D	I	T	K	F	N	J	S	O	U
Pitanje 1	+	+	+	-	+	+	+	+	+	-	+	+	+	-	+	+	+	+	+	+
Pitanje 2	+	+	-	+	+	-	+	-	-	+	+	-	+	+	-	+	-	+	-	-

Koristeći test homogenosti niza promatram dva niza u kojima je

- N=20,
- pitanje 1 glasi „Da li ste dojili barem jedno od Vaše djece?“,
- pitanje 2: „Uočavate li, na temelju Vašeg iskustva, da dojena djeca imaju bolji imunitet ili općenito bolje zdravstveno stanje od djece koja nisu dojena?“
- odgovori „Da“ su zamijenjeni znakom „+“
- odgovori „Ne“ i „Ne mogu reći znakom „-“,

te time nizovi izgledaju ovako:

Tablica 6: nizovi za test homogenosti

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	15	17	18	19	20	Br. nizova	+	-
Ispitanik	A	G	E	C	P	H	B	M	R	L	D	I	T	K	F	N	J	S	O	U		m	n
Pitanje 1	+	+	+	-	+	+	+	+	+	-	+	+	+	-	+	+	+	+	+	+	7	17	3
Pitanje 2	+	+	-	+	+	-	+	-	-	+	+	-	+	+	-	+	-	+	-	-	14	11	9

## Izračun

Koristeći se metodom u kojoj prvo određujem aritmetičku sredinu nizova:

$$\bar{X} = \frac{2 * m * n}{m + n} + 1$$

$$\bar{X}_1 = \frac{2 * 17 * 3}{17 + 3} + 1 = 6,1$$

$$\bar{X}_2 = \frac{2 * 11 * 9}{11 + 9} + 1 = 10,9$$

te nakon toga standardnu devijaciju nizova:

$$s = \sqrt{\frac{2mn(2mn - m - n)}{(m + n)^2(m + n - 1)}}$$

$$s_1 = \sqrt{\frac{2 * 17 * 3 * (2 * 17 * 3 - 17 - 3)}{(17 + 3)^2 * (17 + 3 - 1)}} = 1,05$$

$$s_2 = \sqrt{\frac{2 * 11 * 9 * (2 * 11 * 9 - 11 - 1)}{(11 + 9)^2 * (11 + 9 - 1)}} = 2,20$$

Te nakon toga izračunavamo vrijednost z:

$$z = \frac{\text{Broj nizova} - \bar{X}}{s}$$

$$z_1 = \frac{7 - 6,1}{1,05} = 0,857$$

$$z_2 = \frac{14 - 10,9}{2,20} = 1,409$$

Kako je iznos dobivenog z u oba slučaja manji od 1,654 možemo zaključiti da oba uzorka pripadaju istoj populaciji, tj. da populacije iz kojih su uzete imaju ista obilježja tj. da ne postoji razlika u imunitetu djece koja su dojena u odnosu na djecu koja nisu dojena,

## $\chi^2$ Test

Hi kvadrat test primjenjuje se u različitim statističkim postupcima. Na primjer, taj se test rabi da bi se ispitala zavisnosti dvaju kvalitativnih obilježja. Obilježja A i B mogu biti bilo koje vrste. Grupiraju li se članovi osnovnog skupa istodobno prema oblicima obilježja A i obilježja B, dobiva se dvodimenzionalni raspored koji se predločuje u dvodimenzionalnoj tablici. (Šošić, Primjenjena statistika, 2004., str. 354.)

Apsolutne frekvencije najčešće se označavaju  $m_{ij}$ ,  $n_i$ ,  $n_j$ , ili oznakama  $f_i$ ,  $f_{ij}$ , ili oznakama  $f_i$  i  $e_i$ , dok se oznakom  $p_{ij}$  označuje vjerojatnost izbora elemenata s modalitetom  $A_i$  i  $B_j$ . Ako su klasifikacije članova osnovnog skupa prema modalitetima neovisne, tada je vjerovatnost izbora člana s modalitetom obilježja  $A_i$  i  $B_j$  jednaka umnošku njihovih pojedinačnih vrijednosti.

$H_0$  hipoteza kod istraživanja glasila je distribucija osnovnog skupa specificirana oblika, a  $H_1$  da distribucija osnovnog skupa nije specificirana oblika. Testna statistika je prikazana sljedećom formulom:

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - e_i)^2}{e_i}$$

Gdje su  $f_i$  apsolutne frekvencije, a  $e_i$  očekivane apsolutne frekvencije prema distribuciji navedenoj u nultoj hipotezi.

Odluka se donosi usporedbom empirijske test-veličine s teorijskom vrijednosti hi-kvadrat distribucije. Ako je nulta hipoteza istinita, a uzorak dovoljno velik, empirijski hi-kvadrat pripada hi-kvadratnoj distribuciji s  $(r-1)(c-1)$  stupnjeva slobode. (Šošić, Primjenjena statistika, 2004., str. 354.)

## Istraživanje

Presumirat ćemo situaciju u kojoj banka ispituje učinkovitost poslovne procedure gdje klijent mora zakazati svoj dolazak u poslovnicu, nasuprot dosadašnje poslovne prakse da samo dođe u poslovnicu bez najave. Cilj je istražiti da li se korisnici koji se najave na razgovor ujedno i pojave na istom. Prema dosad provedenim istraživanjima dolazaka u banku, veliki broj klijenata (42%) odustane od samog razgovora kada dođe u banku i vidi da mora dugo čekati. Naša je pretpostavka da će se najavljuvanjem smanjiti gužve i red čekanja i da će broj odustajanja (tj. u ovom slučaju broj klijenata koji nisu došli na zakazani sastanak) biti osjetno manji, oko 10%.

$H_0$  hipoteza u ovom slučaju glasi da su obilježja najavljanja narazgovor i ostvarenja razgovora međusobno nezavisna.  $H_1$  hipoteza prema tome je suprotna i glasi da su obilježja najavljanja narazgovor i pojavljivanja na razgovoru međusobno zavisna.

Sljedeći podatci odgovaraju vremenskom razdoblju od 1 mjeseca klijenata banke za koje je zadužen jedan osobni bankar.

Tablica 7: dolasci najavljenih i nenajavljenih klijenata

	Najavljeni	Nenajavljeni	Ukupno
Ostvaren razgovor	57	15	72
Razgovor nije ostvaren	5	12	17
Ukupno	62	27	89

### Izračun

Koristeći izraz

$$\chi^2 = \sum \frac{(f_i - f_{ti})^2}{f_{ti}}$$

Gdje je  $f_i$  vrijednosti iz ispitivanja a  $f_{ti}$  su očekivane vrijednosti temeljem pretpostavke da obilježja najavljanja na razgovor i pojavljivanja na razgovoru nisu međusobno zavisna. Pretpostavka se temeljila na izraženoj volji klijenata (75% ih je u anketi odgovorilo da bi se rado najavilo na razgovor kod osobnog bankara) te na činjenici da osobni bankar u mjesecu obavi u prosjeku 80 takvih razgovora.

Tablica 8: pomoćna tablica za izračun  $\chi^2$ -testa

$f_i$	$f_{ti}$	$f_i - f_{ti}$	$(f_i - f_{ti})^2$	$(f_i - f_{ti})^2 / f_{ti}$
57	60	-3	9	0,15
5	6	-1	1	0,166667
15	20	-5	25	1,25
12	8	4	16	2
				<b>3,566667</b>

Uzevši u obzir tablicu graničnih vrijednosti, promatrajući signifikatnost od 5% i stupnjeve slobode  $v = 1$ , i rezultat od 3.566667 možemo zaključiti da su obilježja pretpostavljene i stvarne frekvencije međusobno nezavisna čime ostajemo pri  $H_0$  hipotezi.

$\nu$	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.837
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.478	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.457	15.655	17.538	19.280	21.433	41.422	44.985	48.231	52.190	55.000
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.814	17.073	19.046	20.866	23.110	43.745	47.400	50.724	54.774	57.646
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.191	18.508	20.569	22.465	24.796	46.059	49.802	53.203	57.340	60.272
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.584	19.960	22.105	24.075	26.492	48.363	52.192	55.667	59.891	62.880
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.994	21.425	23.654	25.695	28.196	50.660	54.572	58.119	62.426	65.473
40	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.766

Slika 1: Tablica graničnih vrijednosti

## Kruskal Wallisov test

Kruskal – Wallis test testira pripada li više uzoraka istoj populaciji tj. da li je broj od k nezavisnih uzoraka iz različitih populacija. Njime ispitujemo hipotezu koja tvrdi da broj od k uzoraka dolazi iz iste populacije ili iz identičnih populacija s obzirom na srednju vrijednost uz pretpostavku da su varijable kontinuirano distribuirane i da su varijable najmanje u ordinalnoj skali.

## Istraživanje

Koristeći podatke državnog zavoda za statistiku objavljene na stranicama portala Otvorenih podataka Republike Hrvatske<sup>4</sup> pokušao sam odgovoriti na pitanja da li svi subjekti u građevinarstvu u Republici Hrvatskoj imaju iste uvjete (brzina izdavanja građevinskih i drugih dozvola, cijene materijala i rada u građevinarstvu, interes investitora za novim građevinama i sl.) a čiji je zajednički rezultat vidljiv u ukupnoj vrijednosti izvršenih građevinskih radova. Promatrao sam uzorak za 2018. godinu (posljednji dostupan):

Tablica 9: vrijednost izvršenih građevinskih radova

Županija	2018.
Zagrebačka	1.707.927
Krapinsko-zagorska	352.223
Sisačko-moslavačka	570.267
Karlovačka	406.978
Varaždinska	460.956
Koprivničko-križevačka	253.840
Bjelovarsko-bilogorska	380.799
Primorsko-goranska	1.671.094
Ličko-senjska	261.285
Virovitičko-podravska	294.134
Požeško-slavonska	216.157
Brodsko-posavska	401.553
Zadarska	880.238
Osječko-baranjska	832.379
Šibensko-kninska	455.705
Vukovarsko-srijemska	457.938
Splitsko-dalmatinska	1.763.405
Istarska	2.067.400
Dubrovačko-neretvanska	987.604
Međimurska	362.604
Grad Zagreb	2.859.795

---

<sup>4</sup><http://data.gov.hr/dataset/statistika-u-nizu-gradovi-u-statistici>  
<https://www.dzs.hr/Hrv/publication/StatisticsInLine.htm>

## Izračun

$H_0$  hipoteza glasi da svi subjekti u građevinarstvu u republici Hrvatskoj imaju iste uvjete, a  $H_1$  da subjekti u različitim županijama imaju različite uvjete rada i poslovanja. Podijelio sam uzorak od 21 županije na tri ranga rangirajući županije prema BDPu:

Tablica 10: rangiranje županija prema BDPu

Županija	Bdp	rang
Ličko-senjska	3181012	1
Požeško-slavonska	3463447	2
Virovitičko-podravska	3711084	3
Bjelovarsko-bilogorska	6503754	4
Brodsko-posavska	7186996	5
Krapinsko-zagorska	7324017	6
Šibensko-kninska	7352828	7
Koprivničko-križevačka	7362155	8
Karlovačka	7660466	9
Međimurska	8249731	10
Vukovarsko-srijemska	8322099	11
Sisačko-moslavačka	9414295	12
Dubrovačko-neretvanska	11397675	13
Zadarska	12424221	14
Varaždinska	12745769	15
Osječko-baranjska	19189762	16
Grad Zagreb	21034276	17
Istarska	23092305	18
Splitsko-dalmatinska	30702304	19
Primorsko-goranska	31087676	20
Zagrebačka	125019655	21



Nakon toga napravio sam tri ranga po sedam županija i koristeći analizu varijance ANOVA u programu excel vršim sljedeći izračun:

Tablica 11: rangovi županija

Županija	2018.	rang	Županija	2018.	rang	Županija	2018.	rang
Ličko-senjska	261.285	1	Koprivničko-križevačka	253.840	8	Varaždinska	460.956	15
Požeško-slavonska	216.157	2	Karlovačka	406.978	9	Osječko-baranjska	832.379	16
Virovitičko-podravsk	294.134	3	Međimurska	362.604	10	Grad Zagreb	2.859.795	17
Bjelovarsko-bilogorska	380.799	4	Vukovarsko-srijemska	457.938	11	Istarska	2.067.400	18
Brodsko-posavska	401.553	5	Sisačkomoslavačka	570.267	12	Splitsko-dalmatinska	1.763.405	19
Krapinsko-zagorska	352.223	6	Dubrovačkoneretvanska	987.604	13	Primorsko-goranska	1.671.094	20
Šibensko-kinjska	455.705	7	Zadarska	880.238	14	Zagrebačka	1.707.927	21

Tablica 12: ANOVA izračun

Anova:  
Single  
Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	7,00	2361856,00	337408,00	7096696591,00
Column 2	7,00	3919469,00	559924,14	75339073202,14
Column 3	7,00	11362956,00	1623279,43	621969950204,95

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	6611974199378,00	2,00	3305987099689,00	14,08	0,00	3,55
Within Groups	4226434319988,57	18,00	234801906666,03			
Total	10838408519366,60	20,00				

Obzirom da je F veći od Fcrit, odbacujemo nul hipotezu. To znači da svi subjekti u građevinarstvu neamju iste uvjete poslovanja i rada u tom sektoru.

## Friedmanov test

Friedmanov test koristi se za usporedbe tri ili više parnih grupa. Osnova testa istovjetna je Kruskal-Wallis-ovom. Razlika je u tome što kada na istoj grupi ispitanika vršimo mjerenje u različitim uvjetima, Kruskal-Walisov test nam više nije adekvatan jer uzorci nisu nezavisni.

Friedmanov test se provodi na sljedeći način:

1. ispisati podatke u tablici sa K stupaca i N redova
2. rangirati parove u svakom redu od 1 do 3
3. zbrojiti rangove u svakom stupcu
4. izračunati veličinu testa  $\chi_r^2$
5. usporediti sa tablicom graničnih vrijednosti hi-kvadrat testa

## Istraživanje

Koristeći podatke državnog zavoda za statistiku objavljene na stranicama portala Otvorenih podataka Republike Hrvatske<sup>5</sup>, a nastavno na prethodno istraživanje, da li svi subjekti u građevinarstvu u Republici Hrvatskoj imaju iste uvjete, pokušao sam odgovoriti na pitanje da li su se uvjeti i stanje u građevinarstvu mijenjali u posljednje tri godine. Također sam kao mjerilo koristio ukupnu vrijednost izvršenih građevinskih radova. Promatrao sam uzorke za 2016., 2017. i 2018. godinu.

$H_0$  hipoteza glasi da svi subjekti imali iste uvjete i stanje u građevinarstvu u protekle tri godine, dok  $H_1$  hipoteza glasi da uvjeti i stanje u sektoru građevinarstva po županijama su se mijenjali po godinama.

---

<sup>5</sup><http://data.gov.hr/dataset/statistika-u-nizu-gradovi-u-statistici>

Tablica 13: vrijednost izvršenih građevinskih radova

Županija	2016.	2017.	2018.
Zagrebačka	1.308.815	1.577.023	1.707.927
Krapinsko-zagorska	394.947	331.986	352.223
Sisačko-moslavačka	545.719	454.691	570.267
Karlovačka	359.131	443.491	406.978
Varaždinska	346.755	365.347	460.956
Koprivničko-križevačka	486.706	341.122	253.840
Bjelovarsko-bilogorska	302.143	269.945	380.799
Primorsko-goranska	1.506.452	1.434.337	1.671.094
Ličko-senjska	285.329	267.436	261.285
Virovitičko-podravska	292.601	220.668	294.134
Požeško-slavonska	155.563	156.416	216.157
Brodsko-posavska	385.421	336.330	401.553
Zadarska	641.529	891.761	880.238
Osječko-baranjska	959.559	795.036	832.379
Šibensko-kninska	529.406	619.075	455.705
Vukovarsko-srijemska	583.819	566.656	457.938
Splitsko-dalmatinska	1.706.436	1.724.674	1.763.405
Istarska	1.332.276	1.650.832	2.067.400
Dubrovačko-neretvanska	908.008	797.346	987.604
Međimurska	345.363	312.270	362.604
Grad Zagreb	2.331.291	2.643.798	2.859.795

## Izračun

Dobivene rezultate sam rangirao unutar jedne županije po godinama. Najveća vrijednost građevinskih radova u toj županiji u protekle tri godine dobila je rang 1, manja rang 2, a najmanja rang 3.

Tablica 14: rangovi prema županijama

Županija	2016.	rang 2016.	2017.	rang 2017.	2018.	rang 2018.
Zagrebačka	1308815	3	1577023	2	1707927	1
Krapinsko-zagorska	394947	1	331986	3	352223	2
Sisačko-moslavačka	545719	2	454691	3	570267	1
Karlovačka	359131	3	443491	1	406978	2
Varaždinska	346755	3	365347	2	460956	1
Koprivničko-križevačka	486706	1	341122	2	253840	3
Bjelovarsko-bilogorska	302143	2	269945	3	380799	1
Primorsko-goranska	1506452	2	1434337	3	1671094	1
Ličko-senjska	285329	1	267436	2	261285	3
Virovitičko-podravska	292601	2	220668	3	294134	1
Požeško-slavonska	155563	3	156416	2	216157	1
Brodsko-posavska	385421	2	336330	3	401553	1
Zadarska	641529	3	891761	1	880238	2
Osječko-baranjska	959559	1	795036	3	832379	2
Šibensko-kninska	529406	2	619075	1	455705	3
Vukovarsko-srijemska	583819	1	566656	2	457938	3
Splitsko-dalmatinska	1706436	3	1724674	2	1763405	1
Istarska	1332276	3	1650832	2	2067400	1
Dubrovačko-neretvanska	908008	2	797346	3	987604	1
Međimurska	345363	2	312270	3	362604	1
Grad Zagreb	2331291	3	2643798	2	2859795	1
Σ		45		45		45

Koristeći izraz

$$\chi_r^2 = \frac{12}{Nk(k+1)} \sum T_i^2 - 3N(k+1)$$

Gdje je N broj redova (županija), k broj kolona (godina) a  $\sum T_i$  suma rangova, izračunavam:

$$\sum T_i^2 = 45^2 + 45^2 + 45^2 = 6075$$

$$\chi_r^2 = \frac{12}{21 * 3(3+1)} * 6075 - 3 * 21(3+1) = 37,28$$

Usporedbom sa tablicom graničnih vrijednosti hi-kvadrat testa (vidi sliku 2) uz sugnifikantnost od 5% i stupnjeve slobode (k-1) možemo odbaciti  $H_0$  hipotezu i zaključujemo da protekom godina situacija se nije jednako mijenjala u svim županijama, u nekim županijama je u području građevinarstva bolja bila 2016. a u nekima 2018.

## Zaključak

Tema neparametrijskih metoda u statistici učinila mi se zanimljivom i jedinstvenom kao tema za završni rad, ali moram priznati da mi je zadala dosta problema. Nisam bio pripremljen na specifičnosti svake od metoda i nedostatak stvarnih podataka prilagodljivih pojedinoj metodi, a ispočetka nisam htio pribjegavati izmišljenim podacima. Na poslijetku sam uvidio da su podatci ti koji uvjetuju izbor metode i da je prilično teško raditi obrnutim redoslijedom.

Ipak smatram da je najbolji način savladavanja nove razine znanja praktična primjena te se nadam da sam uspio uz nadasve odlične smjernice mentorice praktično primijeniti svaku od odabranih metoda.

## Popis literature:

*Državni zavod za statistiku.* (n.d.). Dohvaćeno iz <https://www.dzs.hr/>

*Numbeo - životni troškovi.* (n.d.). Dohvaćeno iz <https://www.numbeo.com/cost-of-living>

Petz, B., Kolesarić, V., & Ivanec, D. (2012.). *Petzova statistika.* Jastrebarsko: Naklada Slap.

*Portal otvorenih podataka RH.* (n.d.). Dohvaćeno iz <http://data.gov.hr/data/search>

Šošić, I. (2004.). *Primijenjena statistika.* Zagreb: Školska knjiga.

Šošić, I. (2006.). *Primijenjena statistika, 2. izmjenjeno izdanje.* Zagreb: Školska knjiga.

Šošić, I. (1998.). *Zbirka zadataka iz statistike.* Zagreb: Microrod.

## Popis slika

Slika 1: Tablica graničnih vrijednosti .....	16
--	----

## Popis tablica

Tablica 1: rezultati istraživanja .....	5
Tablica 2: rezultati istraživanja .....	8
Tablica 3: rangovi .....	9
Tablica 4: rezultati istraživanja .....	11
Tablica 5: rangovi .....	11
Tablica 6: nizovi za test homogenosti .....	12
Tablica 7: dolasci najavljenih i nenajavljenih klijenata .....	15
Tablica 8: pomoćna tablica za izračun $\chi^2$ -testa.....	15
Tablica 9: vrijednost izvršenih građevinskih radova .....	17
Tablica 10: rangiranje županija prema BDPu.....	18
Tablica 11: rangovi županija .....	19
Tablica 12: ANOVA izračun .....	19
Tablica 13: vrijednost izvršenih građevinskih radova.....	21
Tablica 14: rangovi prema županijama .....	22