



Priručnik za korištenje web aplikacija

# Quantitative Methods

Dražan Dizdar & Darko Katović

Zagreb, 2022.

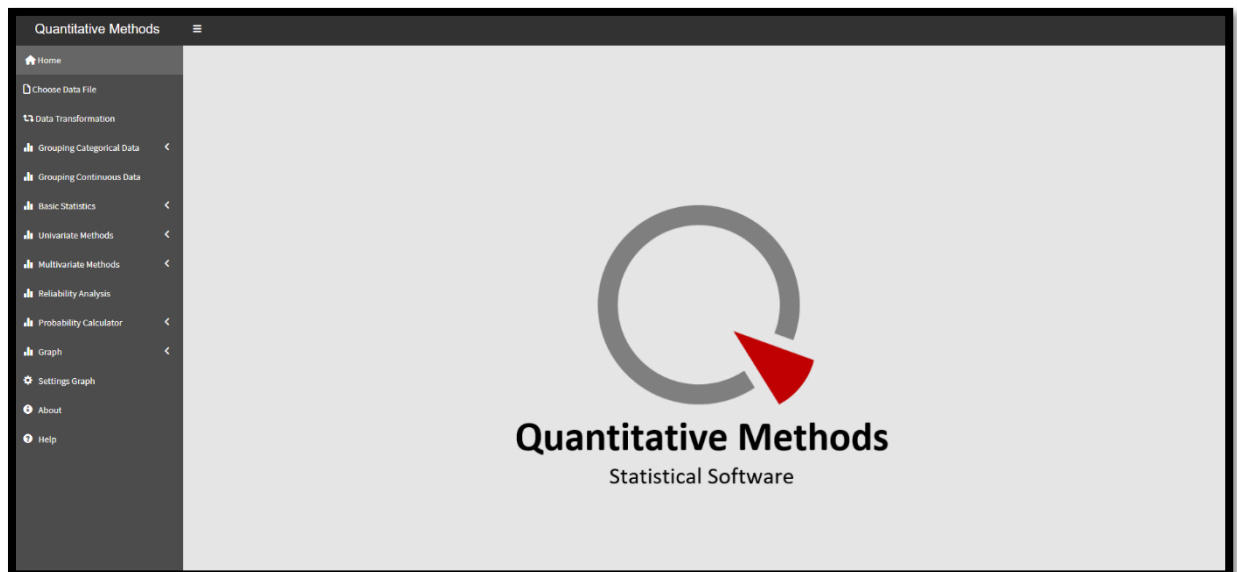
## Sadržaj

WEB APLIKACIJA QUANTITATIVE METHODS .....	2
CHOOSE CSV DATA FILE .....	2
DATA TRANSFORMATION.....	4
GROUPING CATEGORICAL DATA.....	5
FREQUENCY TABLES.....	5
CONTINGENCY TABLES .....	6
GROUPING CONTINUOUS DATA .....	7
BASICS STATISTICS.....	8
DESCIRPTIVE PARAMETERS.....	8
TESTING FOR NORMALITY .....	10
CI FOR POPULATION MEAN.....	11
CORRELATIONS ANALYSIS .....	12
UNIVARIATE METHODS .....	14
INDEPENDENT SAMPLES T-TEST .....	14
DEPENDENT SAMPLES T-TEST .....	18
ONE WAY ANOVA.....	21
MULTIVARIATE METHODS.....	24
REGRESSION ANALYSIS.....	24
FACTOR ANALYSIS.....	26
CANONICAL ANALYSIS.....	28
DISCRIMINANT ANALYSIS.....	30
RELIABILITY ANALYSIS .....	32
PROBABILITY CALCULATOR.....	35
NORMAL DISTRIBUTION .....	35
T – DISTRIBUTION.....	36
F – DISTRIBUTION .....	36
CHI SQUERE – DISTRIBUTION .....	37

## WEB APLIKACIJA QUANTITATIVE METHODS

Quantitative Methods je R Shiny web aplikacije za statističko-grafičku analizu podataka koja se besplatno može koristiti na sljedećim web adresama:

<https://dd-qm.shinyapps.io/QuantitativeMethods/>



### CHOOSE CSV DATA FILE

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati.

	ONT	OUZ	NEB	SKL	TRB	CUC	SDM	BML	T20m	Groups
Marko	25.4	5.4	4	7	16	98	160	510	5.14	I
Mate	14.7	4.9	14	12	20	132	135	330	5.38	I
Sime	16.4	3.9	10	13	52	150	175	550	4.21	I
Mile	15.3	4.2	11	14	30	164	170	650	3.72	I
Jure	17.4	4.1	3	10	13	107	155	520	3.79	I
Ante	15	3.7	10	20	35	128	170	500	4.11	I
Ive	15	4	4	13	32	110	165	390	4.32	I
Stipe	19.4	4.7	8	3	17	165	150	580	4.09	I
Tin	17.5	4	9	20	36	156	155	370	4.87	I
Dino	14.7	4.5	7	11	25	123	195	520	4.26	I
Darko	18	4.7	15	8	26	189	175	630	3.53	I
Stanko	15.2	4.3	8	12	28	190	150	620	4.31	I
Branko	16	4.1	2	23	37	207	180	500	4.09	I
Zarko	13.7	3.8	12	30	63	310	210	690	2.78	II
Sanimir	18.2	4.7	10	3	25	107	180	590	3.48	I
Toni	15.9	4	10	13	34	200	205	540	3.03	II
17	14.7	3.6	7	7	28	130	180	540	3.58	I
18	13.3	4	12	20	47	221	160	470	4.36	I
19	15.1	3.6	10	13	55	210	170	420	3.91	I
20	17.6	4.2	11	10	35	232	185	440	4.42	I

Podaci za statističko-grafičku obradu trebaju biti pripremljeni na sljedeći način:

- prvi stupac tablice s podacima sadrži nazive (oznake) entiteta
- ostali stupci tablice s podacima sadrže podatke entiteta u varijablama
- oznake (kodovi) modaliteta kvalitativnih varijabli ne smiju biti brojevi
- podaci su kreirani u CSV (*Comma Separated Values*) formatu gdje se kao separator između podataka koristi točka-zarez (;), a za decimalne brojeve zarez (,).

**.csv** je tekstualna datoteka s vrijednostima odvojenim zarezima (,) ili točka-zarezima (;). Ovaj format najčešće se koristi za uvoz (import) podataka iz jednog sustava pohrane u drugi, budući da sve aplikacije za statističku obradu podataka prepoznaju ovaj format.

**.ods** je datoteka kreirana u Calc programu za proračunske tablice koji se nalazi u okviru besplatnog programskog paketa LibreOffice (<https://www.libreoffice.org/>).

ENTITETI	SPOL	ATV	ATT	AOP	ANN
Maša	Z	127	26,2	18	6
Ana	Z	122,5	22	17	10
Marija	Z	126	27	19,2	10
Šime	M	122	26,8	19,3	13
Darko	M	125	22,5	17,8	11

Tablice s podacima za vježbu možete preuzeti na ovoj poveznici:

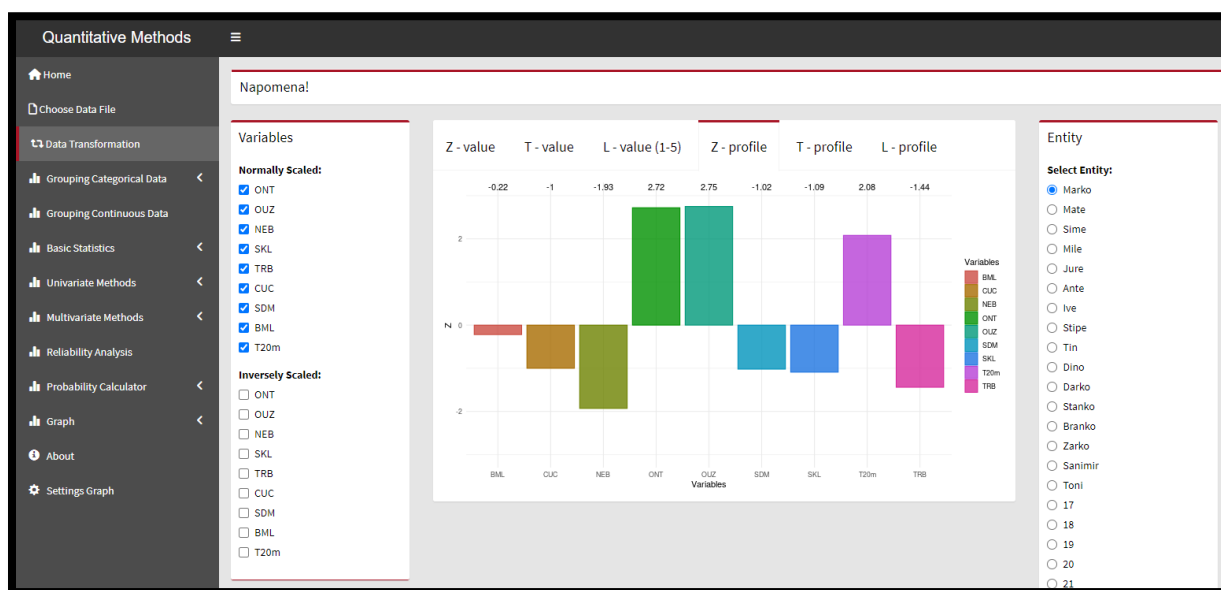
<https://github.com/Quantitative-Methods/Data>

Klikom na gumb **Code** odaberemo opciju **Download ZIP**. Nakon preuzimanja arhiviranu datoteku **Data-main.zip** treba „raspakirati“ i sve datoteke kopirati u prethodno kreiranu mapu. U nastavku teksta opisane su metode za statističko-grafičku analizu podataka koje se nalaze u okviru Web aplikacije Quantitative Methods.

## DATA TRANSFORMATION

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvantitativnu varijablu. Primjerice, u datoteci JUDO3F.csv nalaze se podaci 60 entiteta opisanih sa 9 kvantitativnih varijabli.

Odabirom opcije **Data Transformation** glavnog izbornika za odabrane normalno skalirane (**Normally Scaled**) ili obrnuto skalirane (**Inversely Scaled**) varijable prikazuje se tablica s podacima transformiranih u Z - vrijednosti (**Z - value**), T - vrijednosti (**T - value**), L - skala školskih ocjena (**L - value (1-5)**) s kondenziranim rezultatima svakog ispitanika u odabranim varijablama izračunatih aritmetičkom sredinom (**MEAN\_Z**, **MEAN\_T**, **MEAN\_L**) te grafikoni profila (**Z - profile**, **T - profile** i **L - profile**) u za odabranog entiteta (**Select Entity**).



Rezultati opcije Data Transformation aplikacije Quantitative Methods

# GROUPING CATEGORICAL DATA

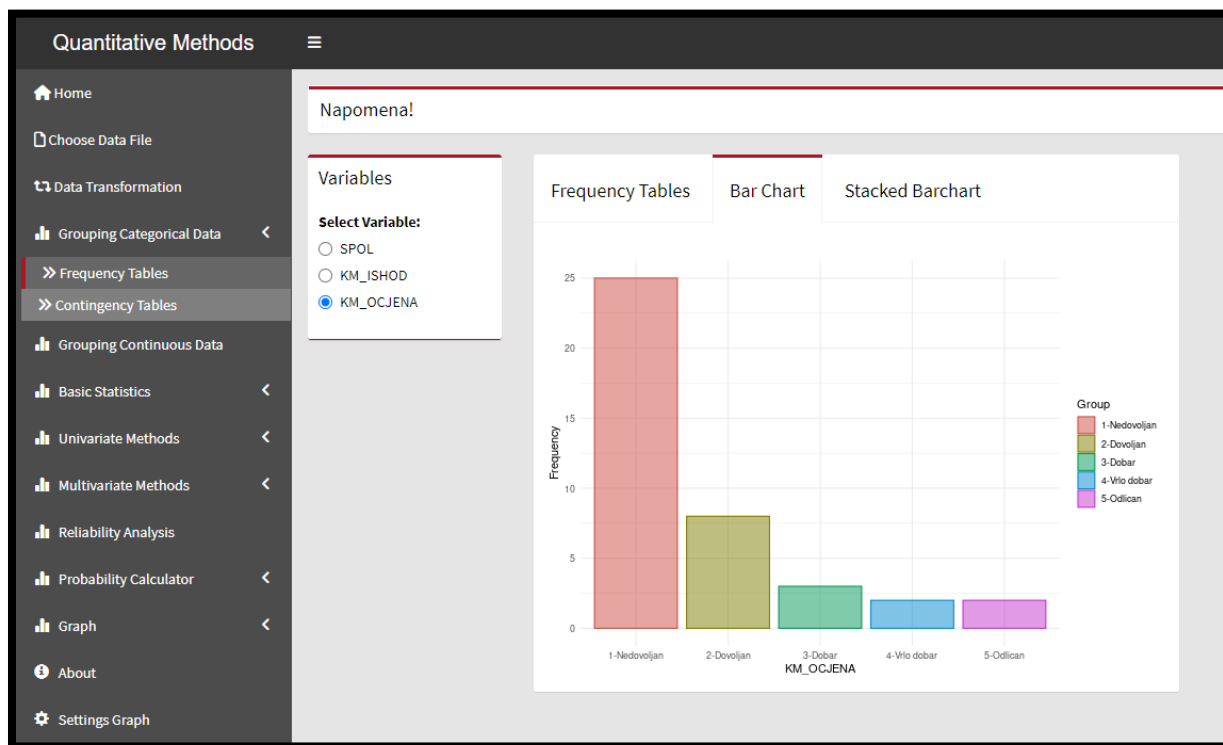
## FREQUENCY TABLES

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvalitativnu varijablu. Primjerice, u datoteci TERAPIJA.csv nalaze se podaci 90 entiteta opisanih s jednom kvalitativnom varijablom (TERAPIJA).

Odabirom opcije **Grouping Categorical Data** glavnog izbornika te opcije **Frequency Tables** za odabranu varijablu (**Select Variable**) prikazuje se tablica i grafikon s rezultatima jednodimenzionalnog grupiranja odabrane varijable te rezultati  $\chi^2$ -testa (Chi-square test = 6,467, df = 2, p = 0,039) za utvrđivanje statističke značajnosti razlika između opaženih i teoretskih frekvencija dobivenih uniformnom distribucijom. Tablica za svaku grupu entiteta prikazuje:

- **Frequency** – apsolutne frekvencije
- **Percent** – relativne postotne frekvencije
- **Cumulative** – kumulativne frekvencije
- **Cumulative Percent** – relativne postotne kumulativne frekvencije.

Na temelju dobivenih rezultata zaključujemo da pacijenti značajno više biraju plivanje kao najučinkovitiju terapiju, potom masažu, a znatno manje elektroterapiju uz pogrešku manju od 5% ( $p = 0,039 < 0,05$ ).



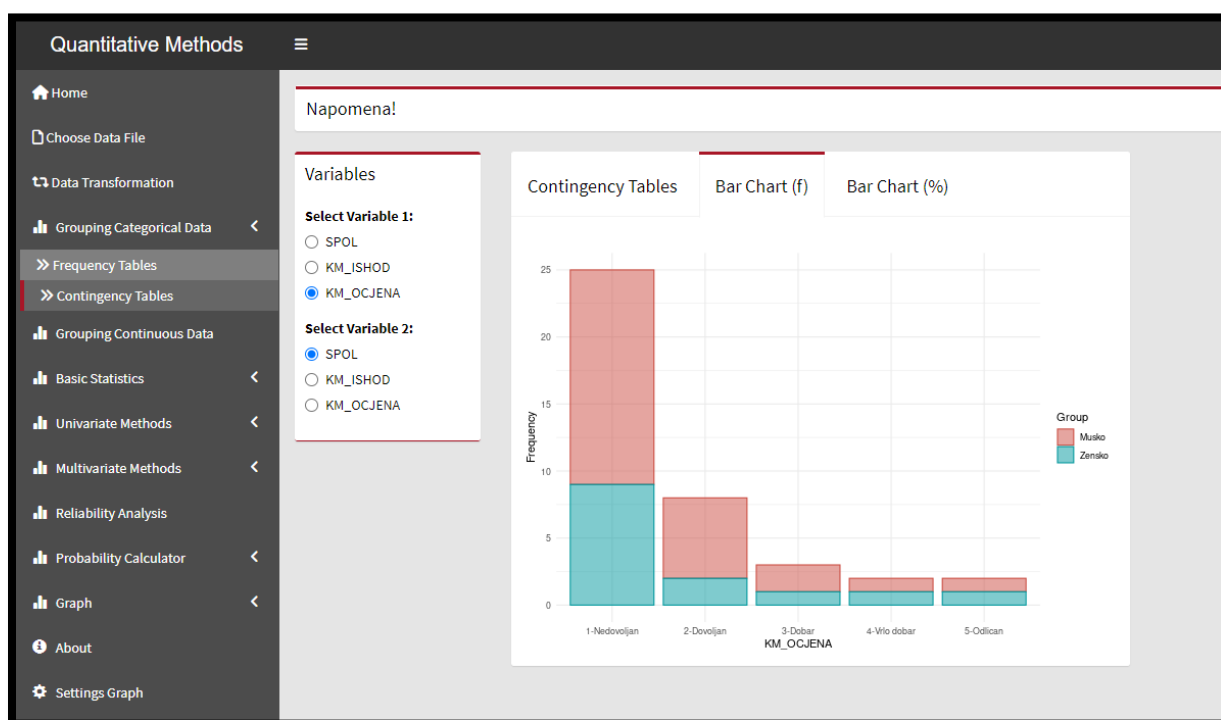
Rezultati opcije Frequency Tables aplikacije Quantitative Methods

## CONTINGENCY TABLES

Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje dvije kvalitativne varijable. Primjerice, u datoteci TEST.csv nalaze se podaci 200 entiteta opisanih s dvije kvalitativne varijable (SPOL i TEST).

Odabirom opcije **Grouping Categorical Data** glavnog izbornika te opcije **Contingency Tables** za odabrane varijable (**Select Variable 1** i **Select Variable 2**) prikazuje se tablice i grafikoni (apsolutnih i relativnih frekvencija) s rezultatima dvodimenzionalnog grupiranja odabranih varijabli (SPOL i TEST) te rezultati  $\chi^2$ -testa (Chi-square test = 0,63, df = 2, p = 0,73) za utvrđivanje statističke značajnosti razlika između opaženih i teoretskih frekvencija dviju kvalitativnih varijabli.

Na temelju dobivenih rezultata  $\chi^2$ -testa ne možemo tvrditi da se muškarci i žene statistički značajno razlikuju u testu iz Kvantitativnih metoda uz pogrešku od 0,05 ( $p = 0,73 > 0,05$ ).



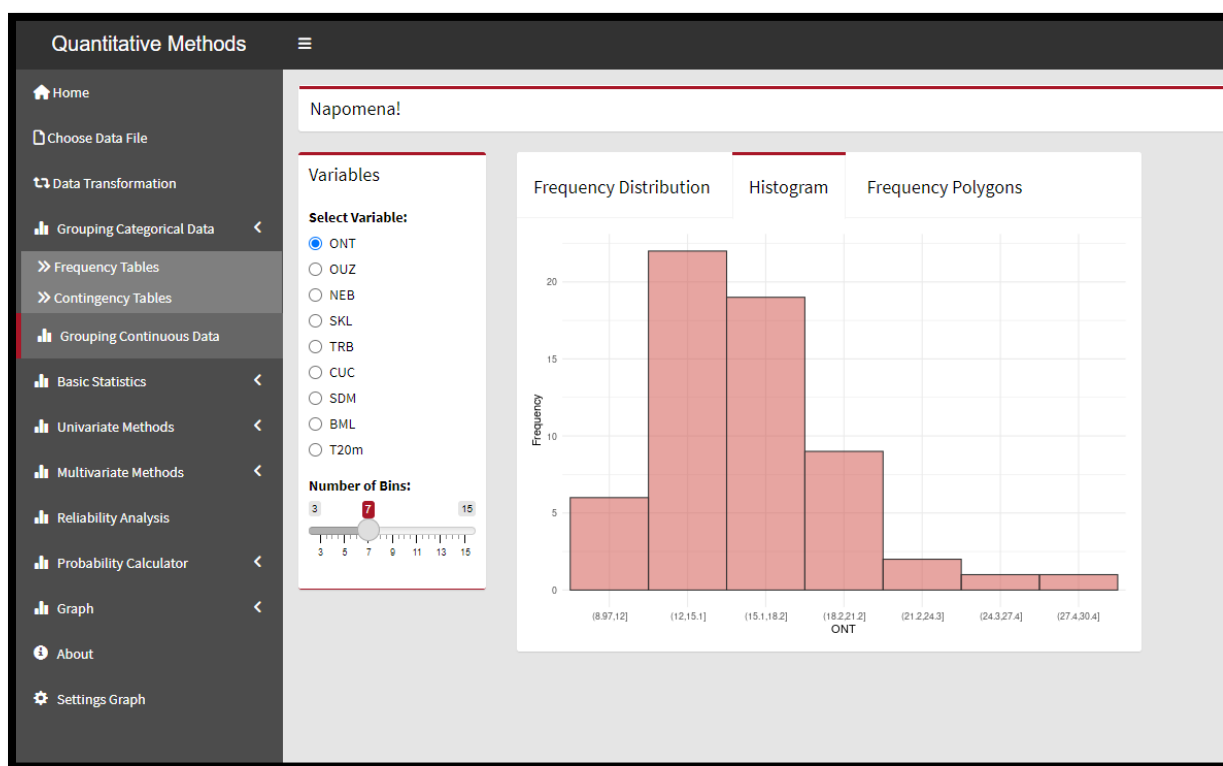
Rezultati opcije Contingency Tables aplikacije Quantitative Methods

## GROUPING CONTINUOUS DATA

Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvantitativnu varijablu. Primjerice, u datoteci JUDO3F.csv nalaze se podaci 60 entiteta opisanih s 9 kvantitativnih varijabli.

Odabirom opcije **Grouping Continuous Data** glavnog izbornika prikazuje se tablica **Frequency Distribution** u kojoj se nalaze:

- **Frequency** – apsolutne frekvencije
- **Percent** – relativne postotne frekvencije
- **Cumulative** – kumulativne frekvencije
- **Cumulative Percent** – relativne postotne kumulativne frekvencije.



*Rezultati opcije Grouping Continuous Data aplikacije Quantitative Methods*

U karici **Histogram** nalazi se histogram apsolutnih frekvencija, a u kartici **Frequency Polygons** poligon relativnih kumulativnih frekvencija za odabrani broj razreda (**Number of Bins**).



# BASICS STATISTICS

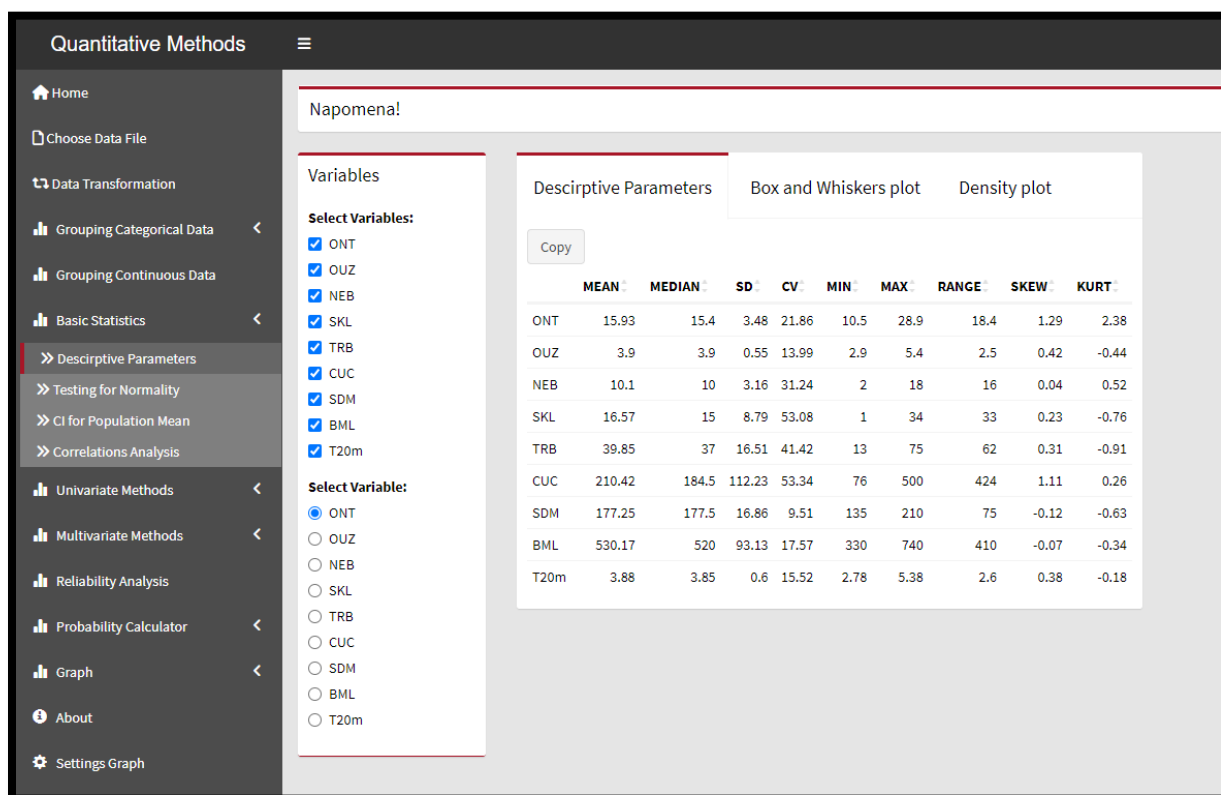
## DESCIRPTIVE PARAMETERS

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvantitativnu varijablu. Primjerice, u datoteci JUDO3F.csv nalaze se podaci 60 entiteta opisanih s 9 kvantitativnih varijabli.

Odabirom opcije **Basics Statistics** glavnog izbornika te opcije **Descirptive Parameters** za odabrane varijable (**Select Variables**) prikazuje se tablica deskriptivnih pokazatelja:

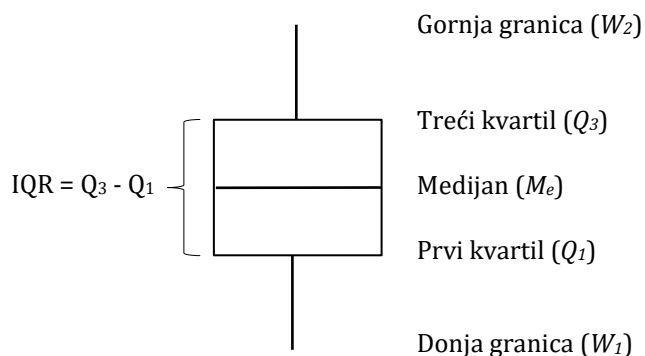
- **MEAN** – aritmetička sredina
- **MEDIAN** – medijan
- **RANGE** – totalni raspon
- **SD** – standardna devijacija
- **CV** – koeficijent varijabilnosti
- **MIN** – minimalna vrijadnost
- **MAX** – maksimalna vrijednosti
- **SKEW** – koeficijent asimetrije distribucije (*skewness*)
- **KURT** – stupanj spljoštenosti odnosno izduženosti distribucije (*kurtosis*).

Odabirom kartice **Box and Whiskers plot** ili **Density plot** prikazuju se odgovarajući grafikoni za odabranu varijablu **Select Variable**.



Rezultati opcije Descriptive Parameters aplikacije Quantitative Methods

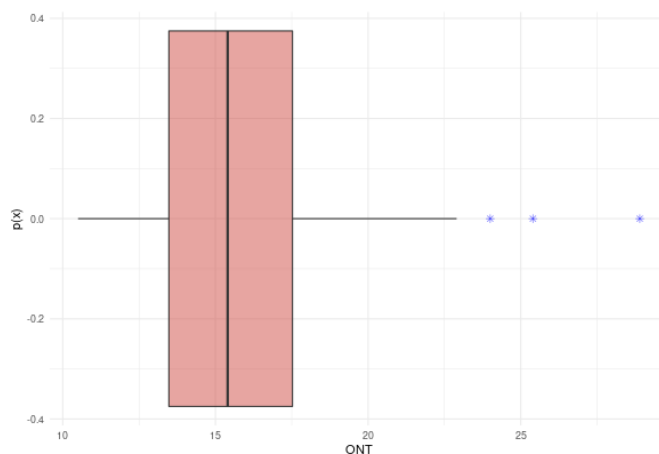
Kutijasti dijagram (engl. Box – whisker plot) prikazuje odnose pet statističkih pokazatelja temeljeg kojeg je moguće uočiti stupanj disperzije i asimetrije distribucije te outliere (vrijednosti koje ekstremno odstupaju od ostalih).



Kutijasti dijagram sastoji od pravokutnika čije stranice prikazuje vrijednosti prvog ( $Q_1$ ) i trećeg kvartila ( $Q_3$ ) unutar kojih se nalazi 50% svih rezultata. Crta unutar pravokutnika označava median ( $M_e$ ), dok se donja ( $W_1$ ) i gornja ( $W_2$ ) granica (engl. *whisker*) najčešće odredi na sljedeći način:

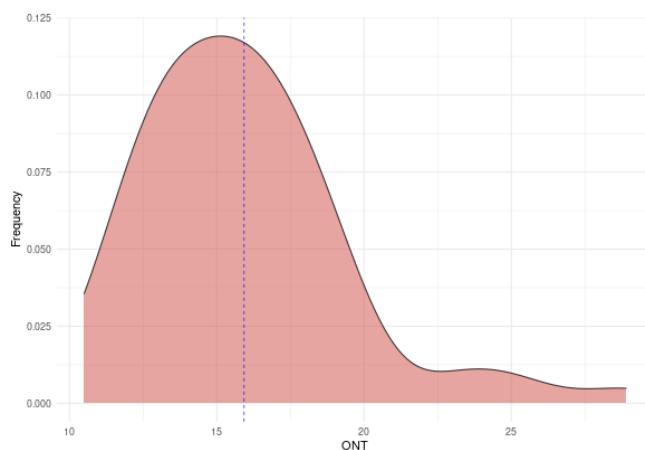
$$W_1 = \text{Min} \text{ ako je } \text{Min} > Q_1 - 1,5 \cdot IQR, \quad \text{inače je } W_1 = Q_1 - 1,5 \cdot IQR$$

$$W_2 = \text{Max} \text{ ako je } \text{Max} < Q_3 + 1,5 \cdot IQR, \quad \text{inače je } W_2 = Q_3 + 1,5 \cdot IQR$$



Kutijasti dijagram (Box-whisker plot)

Dijagram gustoće (engl. Density plot) „izglađena“ je verzija histograma procijenjena na temelju empirijskih podataka.

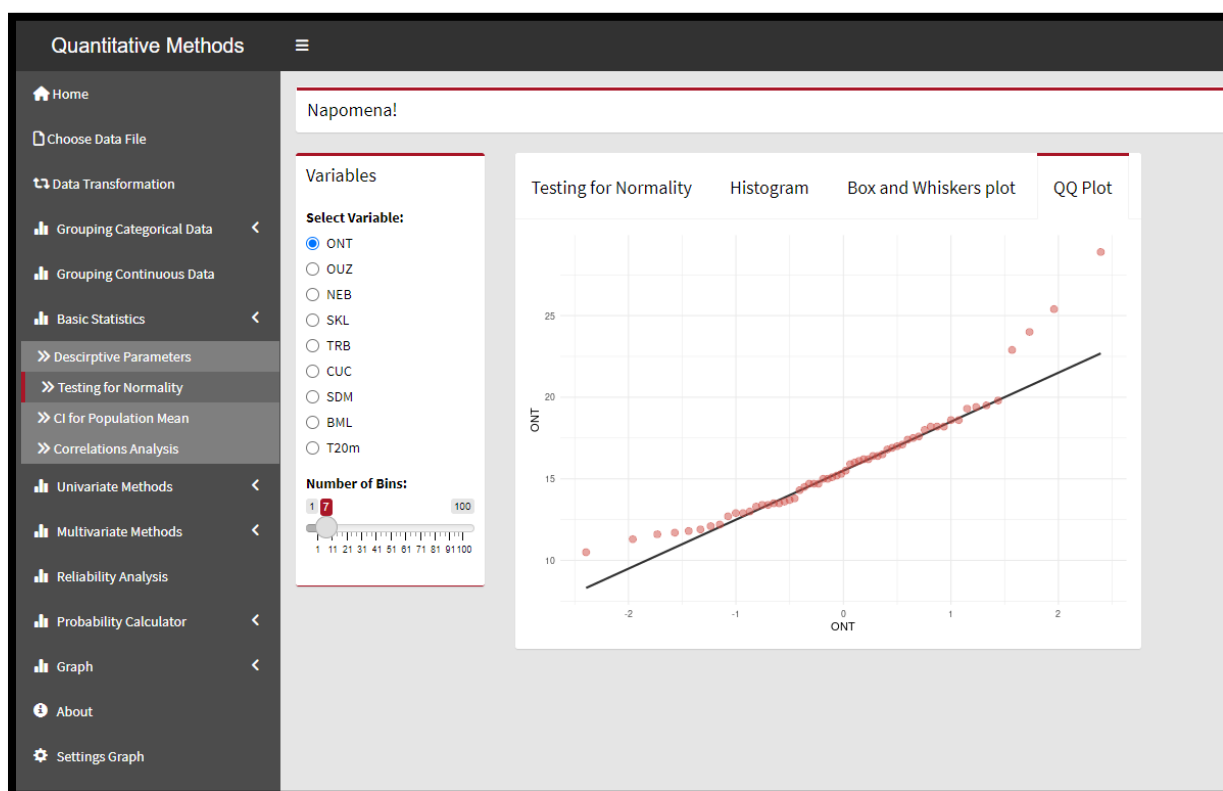


Dijagram gustoće (Density plot)

## TESTING FOR NORMALITY

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvantitativnu varijablu. Primjerice, u datoteci JUDO3F.csv nalaze se podaci 60 entiteta opisanih s 9 kvantitativnih varijabli.

Odabirom opcije **Basics Statistics** glavnog izbornika te opcije **Testing for Normality** za odabranu varijablu (**Select Variable**) dobije mogućnost testiranja statističke značajnosti odstupanja empirijske distribucije od normalne ili Gaussove distribucije statističkim postupcima kao što su: *Kolmogorov-Smirnov*, *Lilliefors*, *Anderson-Darling*, *Shapiro-Francia* i *Shapiro-Wilksov test*.



Rezultati opcije Testing for Normality aplikacije Quantitative Methods

Osnovni problem svih testova normaliteta distribucije jest u zavisnosti njihovih rezultata od veličine uzorka (što su uzorci veći, to je veća i vjerojatnost da razlika bude statistički značajna). Ako se u istraživanjima koriste veliki uzorci, to može rezultirati statistički značajnim testovima normalnosti distribucija i u slučajevima kada grafički prikazi empirijskih distribucija izgledaju gotovo identično teoretskoj normalnoj distribuciji. Nasuprot tome, na malim uzorcima testovima normaliteta distribucije nedostaje statističke snage kako bi se utvrdilo da su evidentno velika odstupanja neke empirijske distribucije statistički značajna. Test za kojeg se smatra da ima najveću statističku snagu od svih testova za provjeru hipoteze o normalnosti neke empirijske distribucije je Shapiro-Wilkov test.

Vrijednosti parametra W Shapiro-Wilkovog testa kreću se u intervalu od 0 do 1, pri čemu niže vrijednosti pokazuju veća, a više vrijednosti manja odstupanja od normalne distribucije. Shapiro-Wilkov test modificirao je Royston (1992) i proširio mogućnost korištenja za uzorke veće od 50 entiteta. Razali i Wah (2011) uspoređivali su snagu

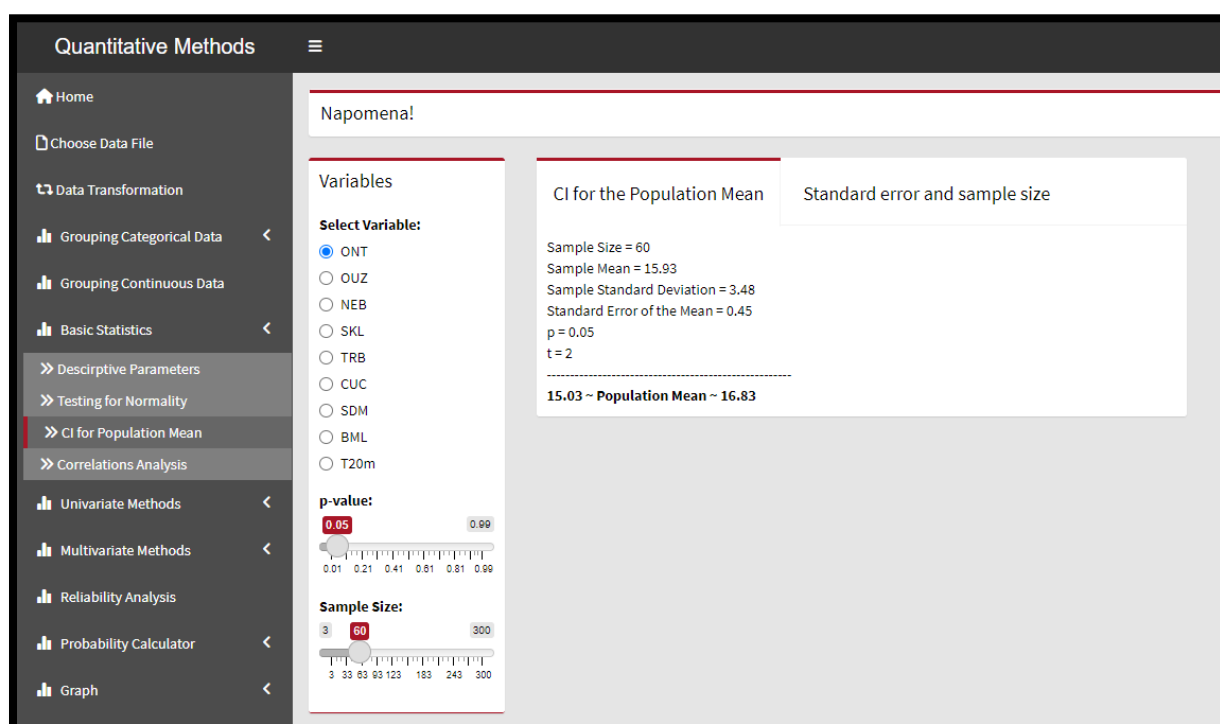
Shapiro-Wilkovog, Kolmogorov-Smirnovog, Lillieforsovog i Anderson-Darlingovog testa pomoću Monte Carlo metode na deset tisuća uzoraka različitih veličina koji su generirani iz populacija čije distribucije nisu normalne. Rezultati su pokazali da je Shapiro-Wilkov test naj snažaniji, dok je Kolmogorov-Smirnov test najmanje snažan. Međutim, snaga svih testova i dalje je niska za male uzroke ( $n < 30$ ). Slične rezultate dobili su Mendesa i Pala (2003) te Keskin (2006). Autori su predložili da se u praksi kombiniraju grafički prikazi (**Histogram**, **Box & Whiskers Plot** i **QQ Plot**), pokazatelji oblika distribucije (skewnes i kurtosis) te statistički testovi kako bi se što pouzdanije utvrdila razina i vrsta odstupanja neke empirijske od normalne distribucije.

Q-Q dijagram (**QQ Plot**) je jedan od najboljih načina za usporedbu empirijske distribucije s normalnom ili Gaussovom distribucijom. Ovaj grafikon prikazuje teoretske kvantile u odnosu na stvarne kvantile promatrane varijable. Što su sve točke bliže liniji (teoretska distribucija), to empirijska distribucija manje odstupa od normalne distribucije.

## CI FOR POPULATION MEAN

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvantitativnu varijablu. Primjerice, u datoteci JUDO3F.csv nalaze se podaci 60 entiteta opisanih s 9 kvantitativnih varijabli.

U okviru opcije **Basics Statistics** glavnog izbornika odabirom opcije **CI for Population Mean** izbornika za odabranu varijablu (**Select Variable**) aplikacija računa aritmetičku sredinu (**Sample Mean**) i standardnu devijaciju (**Sample Standard Deviation**) uzorka te zavisno o veličini uzorka računa standardnu pogrešku aritmetičke sredine (**Standard Error of Mean**) na temelju koje računa interval u kojem se nalazi aritmetička sredina populacije (**Population Mean**), zavisno o pogrešci statističkog zaključka (**p-value**).



Rezultati opcije CI for the Population Mean aplikacije Quantitative Methods

## CORRELATIONS ANALYSIS

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje dvije kvantitativne varijable. Primjerice, u datoteci JUDO3F.csv nalaze se podaci 60 entiteta opisanih s 9 kvantitativnih varijabli.

Odabirom opcije **Basics Statistics** glavnog izbornika te opcije **Correlations Analysis** za odabrane varijable (**Select Variables**) prikazuje se tablica korelacija.



*Rezultati opcije Correlations Analysis aplikacije Quantitative Methods*

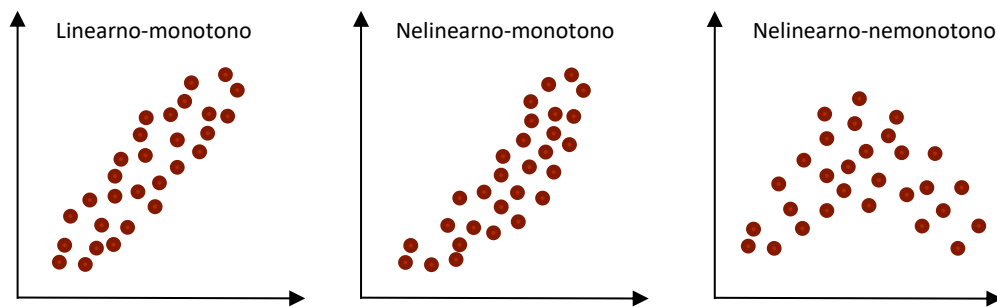
Aplikacija omogućava izračunavanje Pearsonovog, Spearmanovog i Kendallovog koeficijenta korelacije (**Correlations**), utvrđivanje njihove statističke značajnosti (**Level of Significance**) te grafičko prikazivanje korelativnih odnosa putem korelacijskog dijagrama (**Scatter Plot**).

Za utvrđivanje povezanosti između dviju varijabli najčešće se koristi Pearsonov koeficijent korelacija. Da bismo koristili Pearsonov koeficijent korelacije, podaci moraju ispunjavati sljedeće zahtjeve (pretpostavke):

- Podaci su dobiveni pomoću intervalne ili omjerne mjerne ljestvice (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, tapping rukom i sl.).
- Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ( $n > 30$ ).
- Podaci ne bi trebali imati značajne *outliere*, odnosno ekstremno visoke ili niske rezultate.
- Varijable su u linearnom odnosu.
- Varijable imaju normalnu ili Gaussovu distribuciju.

Ako su varijable mjerene na ordinalnoj, intervalnoj ili omjernoj mjerenoj ljestvici, koje nisu normalno distribuirane te se ne nalaze u linearnom, već monotonom odnosu, tada

se za utvrđivanje povezanosti između dviju varijabli koristi **Spearmanov koeficijent rang korelacije** (Spearman, 1904)<sup>1</sup>.



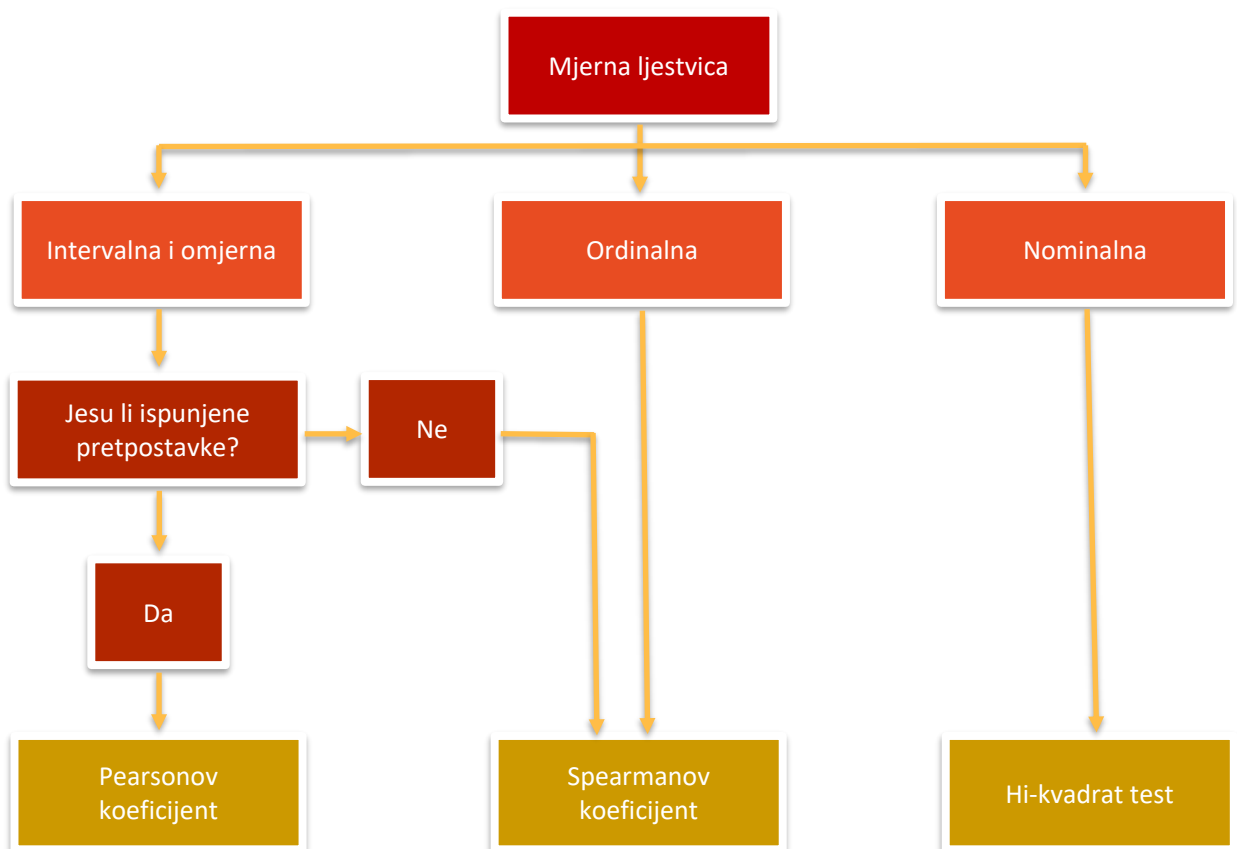
Spearmanov koeficijent rang korelacije izračunava se prema formuli

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)}$$

gdje je

- $d_i$  – razlika između rangova entiteta  $i$  u prvoj i drugoj varijabli
- $n$  – broj entiteta.

*Izbor odgovarajućeg koeficijenta korelacije ovisno o vrsti, distribuciji i odnosu između varijabli.*



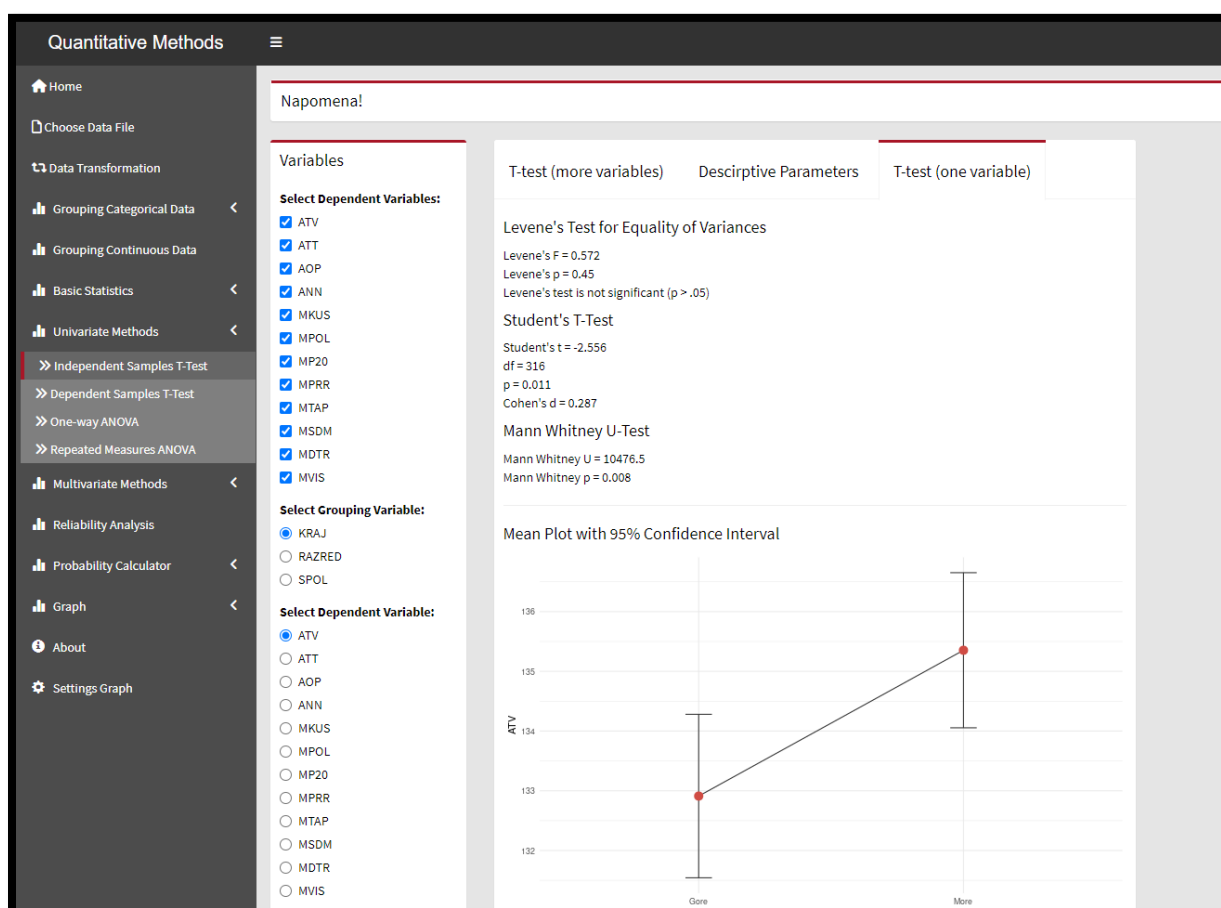
<sup>1</sup>Spearman C. (1904). The proof and measurement of association between two things. American Journal of Psychology. 15 (1): 72-101.

# UNIVARIATE METHODS

## INDEPENDENT SAMPLES T-TEST

Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvalitativnu varijablu s dva modaliteta i jednu ili više kvantitativnih varijabli. Primjerice, u datoteci UCENICI.csv nalaze se podaci 318 entiteta opisanih s tri kvalitativne varijable (KRAJ, RAZRED i SPOL) te 12 kvantitativnih varijabli pri čemu varijable KRAJ i SPOL imaju po dva modaliteta.

Odabirom opcije **Univariate Methods** glavnog izbornika te opcije **Independent Samples T-Test** za odabrane varijable (**Select Dependent Variables** i **Select Grouping Variable**) prikazuju se rezultati t – testa za nezavisne uzorke.



*Rezultati opcije Independent Samples T-Test aplikacije Quantitative Methods*

Da bismo koristili t - test za nezavisne uzorke, podaci moraju ispunjavati sljedeće zahtjeve (pretpostavke):

- Nezavisnu varijablu čine podaci prikupljeni na dihotomnoj (s dva modaliteta) nominalnoj ili ordinalnoj mjernoj ljestvici (npr. muškarci-žene, uspješni-neuspješni, pobjednici-poraženi, eksperimentalna grupa-kontrolna grupa i sl.).
- Zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjernoj mjernoj skali (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, taping rukom i sl.).
- Podaci su prikupljeni na slučajno odabranim i relativno velikim uzorcima entiteta ( $n > 30$ ).

- Podaci u zavisnoj varijabli imaju normalnu ili Gaussovu distribuciju. Ukoliko ova pretpostavka nije ispunjena onda se umjesto t-testa za nezavisne uzorke može koristiti neparametrijski Mann-Whitney U test koji ne zahtjeva normalnost distribucije podataka u zavisnoj varijabli.
- Podaci u zavisnoj varijabli dviju grupa entiteta imaju homogene varijance. Ukoliko ova pretpostavka nije ispunjena, onda se računa korigirana Welchova t-vrijednost.
- Podaci u zavisnoj varijabli ne bi trebali imati značajne outliere, odnosno ekstremno visoke ili niske rezultate.

U kartici **T-Test (more variables)** prikazuje se tablica sa sljedećim rezultatima:

- **MEAN1** – aritmetička sredina prve grupe
- **MEAN2** – aritmetička sredina druge grupe
- **DIFF** – razlika aritmetičkih sredina prve i druge grupe (MEAN1-MEAN2)
- **SED** – standardna pogreška razlika aritmetičkih sredina
- **T** – Studentova t-vrijednost koja pokazuje koliko je puta razlika između aritmetičkih sredina veća od standardne pogreške razlika.
- **DF** – broj stupnjeva slobode ( $n - 2$ , gdje je  $n$  broj entiteta prve i druge grupe)
- **P** – pogreška kojom tvrdimo da je razlika između aritmetičkih sredina prve i druge grupe statistički značajna.
- **DIFF-CI95%** – donja granica intervala u kojem se s vjerojatnošću 95% nalazi razlika aritmetičkih sredina populacije prve i druge grupe
- **DIFF+CI95%** – gornja granica intervala u kojem se s vjerojatnošću 95% nalazi razlika aritmetičkih sredina populacije prve i druge grupe
- **Levene's F** – Levenova F vrijednost temeljem koje se utvrđuje homogenost varijanci, odnosno jesu li su varijance analiziranih grupa statistički značajno različite.
- **Leven's p** – pogreška s kojom tvrdimo da su varijance analiziranih grupa statistički značajno različite. Ako je  $p < 0,05$ , zaključujemo da varijance analiziranih grupa nisu homogene, odnosno da se statistički značajno razlikuju uz pogrešku manju od 5%. U tom slučaju se umjesto Studentove t-vrijednosti, računa Welchova t-vrijednost.
- **Cohen's d** – Cohenova mjera veličine učinka (effect size).

U kartici **Descirptive Parameters** prikazuje se tablica sa sljedećim pokazateljima za svaku grupu entiteta:

- **MEAN** – aritmetička sredina
- **SD** – standardna devijacija
- **SEM** – standardna pogreška aritmetičke sredine
- **MEAN-CI95%** – donja granica intervala u kojem se s vjerojatnošću 95% nalazi aritmetička sredina populacije
- **MEAN+CI95%** – gornja granica intervala u kojem se s vjerojatnošću 95% nalazi aritmetička sredina populacije
- **Shapiro-Wilk W** – vrijednost Shapiro-Wilkovog testa za procjenu normaliteta distribucije
- **P** – pogreška s kojom tvrdimo da se empirijska distribucija statistički značajno razlikuje od normalne ili Gaussove.

Ako je  $P < 0,05$ , zaključujemo da empirijska distribucija statistički značajno odstupa od normalne ili Gaussove distribucije uz pogrešku  $P$  te da bi umjesto t-testa za nezavisne uzorke bilo primjerenije koristiti Mann-Whitney U test koji ne zahtjeva normalnost distribucije podataka u zavisnoj varijabli.

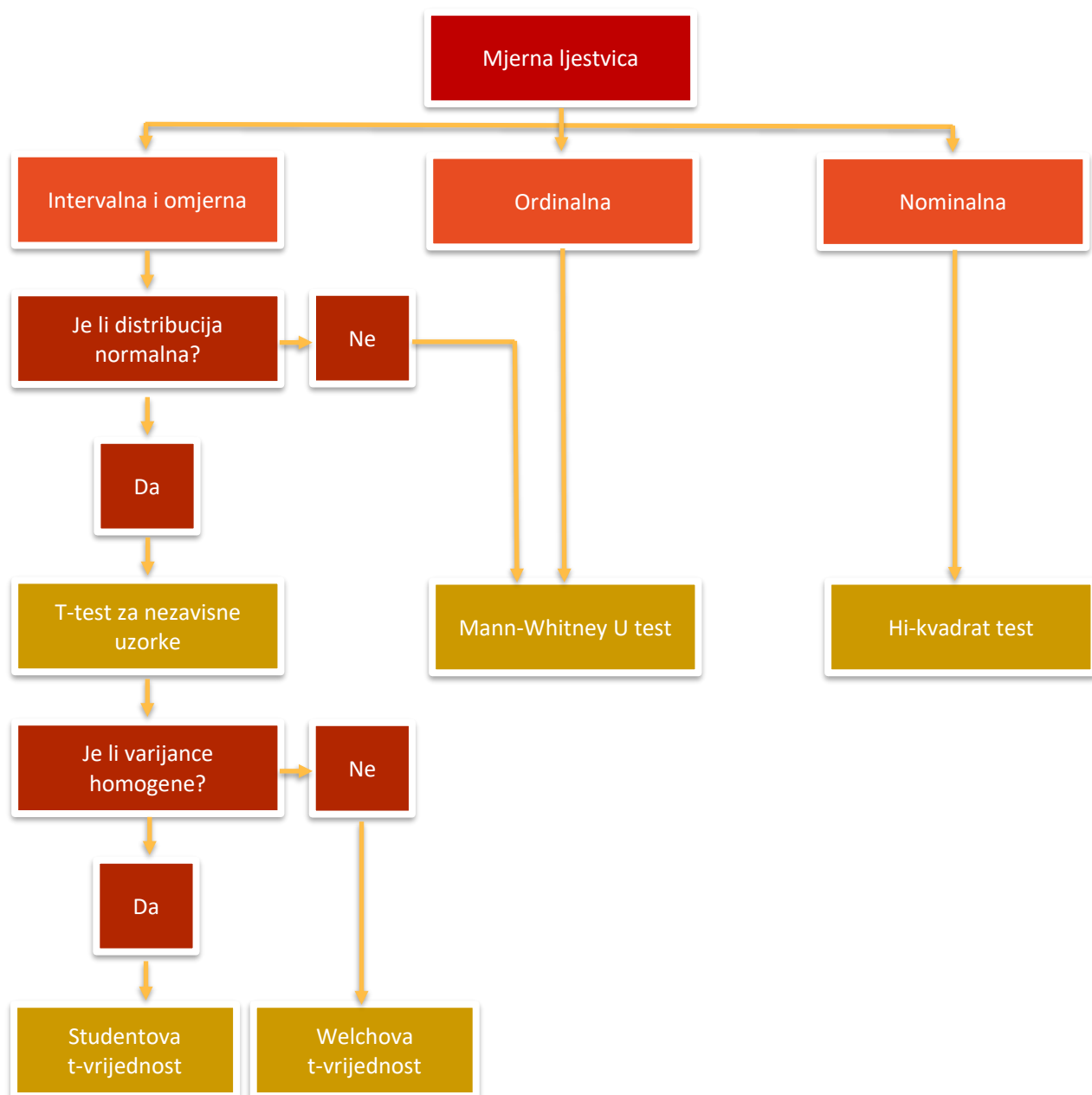
Osim toga, aplikacija omogućava prikazivanje podataka putem grafikona (**Box and Whiskers plot** i **Density plot**) za svaku grupu entiteta.



U kartici (**T-Test one variable**) prikazuju se sljedeći rezultati za svaku odabranu zavisnu varijablu:

- **Levene's F** – Levenova F vrijednost temeljem koje se utvrđuje homogenost varijanci, odnosno jesu li su varijance analiziranih grupa statistički značajno različite.
- **Leven's p** – pogreška s kojom tvrdimo da su varijance analiziranih grupa statistički značajno različite. Ako je  $p < 0,05$ , zaključujemo da varijance analiziranih grupa nisu homogene, odnosno da se statistički značajno razlikuju uz pogrešku manju od 5%. U tom slučaju se umjesto Studentove t-vrijednosti, računa Welchova t-vrijednost.
- **Student's t** – Studentova t-vrijednost koja pokazuje koliko je puta razlika između aritmetičkih sredina veća od standardne pogreške razlika.
- **df** – broj stupnjeva slobode ( $n - 2$ , gdje je n broj entiteta prve i druge grupe)
- **p** – pogreška kojom tvrdimo da je razlika između aritmetičkih sredina prve i druge grupe statistički značajna.
- **Cohen's d** – Cohenova mjera veličine učinka (effect size).
- **Mann Whitney U** – U vrijednost Mann Whitney testa.
- **Mann Whitney p** – pogreška kojom tvrdimo da je razlika između prve i druge grupe statistički značajna.
- **Mean Plot with 95% Confidence Interval** – grafički prikaz vrijednosti aritmetičkih sredina te donje i gornje granice intervala u kome se nalaze aritmetičke sredine populacija analiziranih grupa entiteta uz pogrešku do 5%.

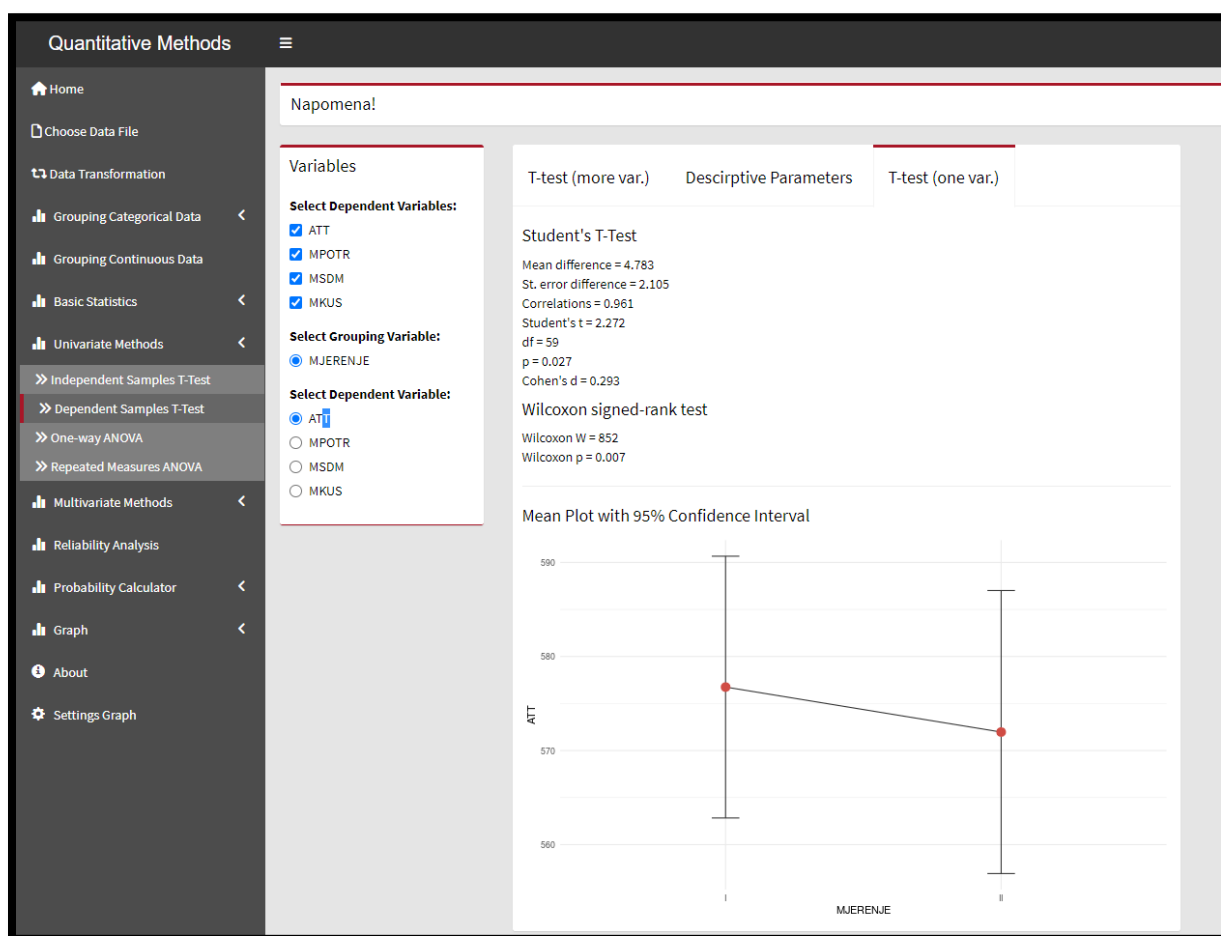
*Testiranje razlika između dviju nezavisnih grupa, ovisno o vrsti i distribuciji varijabli.*



## DEPENDENT SAMPLES T-TEST

Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvalitativnu varijablu s dva modaliteta i jednu ili više kvantitativnih varijabli. Primjerice, u datoteci POD.csv nalaze se podaci 60 entiteta mjerenih u dvije vremenske točke (MJERENJE) u 4 kvantitativne varijable.

Odabirom opcije **Univariate Methods** glavnog izbornika te opcije **Dependent Samples T-Test** za odabrane varijable (**Select Dependent Variables** i **Select Grouping Variable**) prikazuju se rezultati t – testa za zavisne uzorke.



*Rezultati opcije Dependent Samples T-Test aplikacije Quantitative Methods*

Da bismo koristili t - test za zavisne uzorke, podaci moraju ispunjavati sljedeće zahtjeve (pretpostavke):

- Nezavisnu varijablu čine upareni (zavisni) podaci (redoslijed entiteta prvog i drugog mjerenja mora biti isti) na nominalnoj ljestvici (npr. prvo mjerenje - drugo mjerenje, prije tretmana - poslije tretmana i sl.).
- Zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjernoj mjerne skali (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, taping rukom i sl.).
- Podaci su prikupljeni na slučajno odabranim i relativno velikim uzorcima entiteta ( $n > 30$ ).
- Podaci u varijabli razlika prvog i drugog mjerenja u zavisnoj varijabli imaju normalnu ili Gaussovu distribuciju. Ukoliko ova pretpostavka nije ispunjena, onda se

umjesto t-testa za zavisne uzorke može koristiti neparametrijski Wilcoxonov test ekvivalentnih parova (Wilcoxon Matched Pairs Test) koji ne zahtijeva normalnost distribucije podataka u zavisnoj varijabli.

- Podaci u zavisnoj varijabli ne bi trebali imati značajne outliere, odnosno ekstremno visoke ili niske rezultate.

U kartici **T-Test (more var.)** prikazuje se tablica sa sljedećim rezultatima:

- **DIFF** – razlika aritmetičkih sredina prvog i drugog mjerenja
- **SED** – standardna pogreška razlika aritmetičkih sredina prvog i drugog mjerenja
- **T** – Studentova t-vrijednost koja pokazuje koliko je puta razlika između aritmetičkih sredina veća od standardne pogreške razlika.
- **DF** – broj stupnjeva slobode ( $n - 1$ , gdje je  $n$  broj entiteta)
- **P** – pogreška kojom tvrdimo da je razlika između aritmetičkih sredina prvog i drugog mjerenja statistički značajna.
- **DIFF-CI95%** – donja granica intervala u kojem se s vjerojatnošću 95 % nalazi razlika aritmetičkih sredina populacije prvog i drugog mjerenja
- **DIFF+CI95%** – gornja granica intervala u kojem se s vjerojatnošću 95 % nalazi razlika aritmetičkih sredina populacije prvog i drugog mjerenja
- **Cohen's d** – Cohenova mjera veličine učinka (effect size).

U kartici **Descirptive Parameters** prikazuje se tablica sa sljedećim deskriptivnim pokazateljima za prvo i drugo mjerenje:

- **MEAN** – aritmetička sredina
- **SD** – standardna devijacija
- **SEM** – standardna pogreška aritmetičke sredine
- **-CI 95%** – donja granica intervala u kojem se s vjerojatnošću 95% nalazi aritmetička sredina populacije
- **+CI 95%** – gornja granica intervala u kojem se s vjerojatnošću 95% nalazi aritmetička sredina populacije.

te rezultati **Shapiro-Wilkovog test** kojim se utvrđuje da li se varijabla razlika prvog i drugom mjerenja statistički značajno razlikuje od normalne ili Gaussove distribucije:

- **W** – vrijednost Shapiro-Wilkovog testa
- **p** – pogreška s kojom tvrdimo da se varijabla razlika prvog i drugog mjerenja statistički značajno razlikuje od normalne ili Gaussove.

Ako je  $p < 0,05$ , zaključujemo da varijabla razlika prvog i drugog mjerenja statistički značajno odstupa od normalne ili Gaussove distribucije uz pogrešku  $p$  te da bi umjesto t-testa za zavisne uzorke bilo primjerenije koristiti Wilcoxonov test ekvivalentnih parova (Wilcoxon Matched Pairs Test).

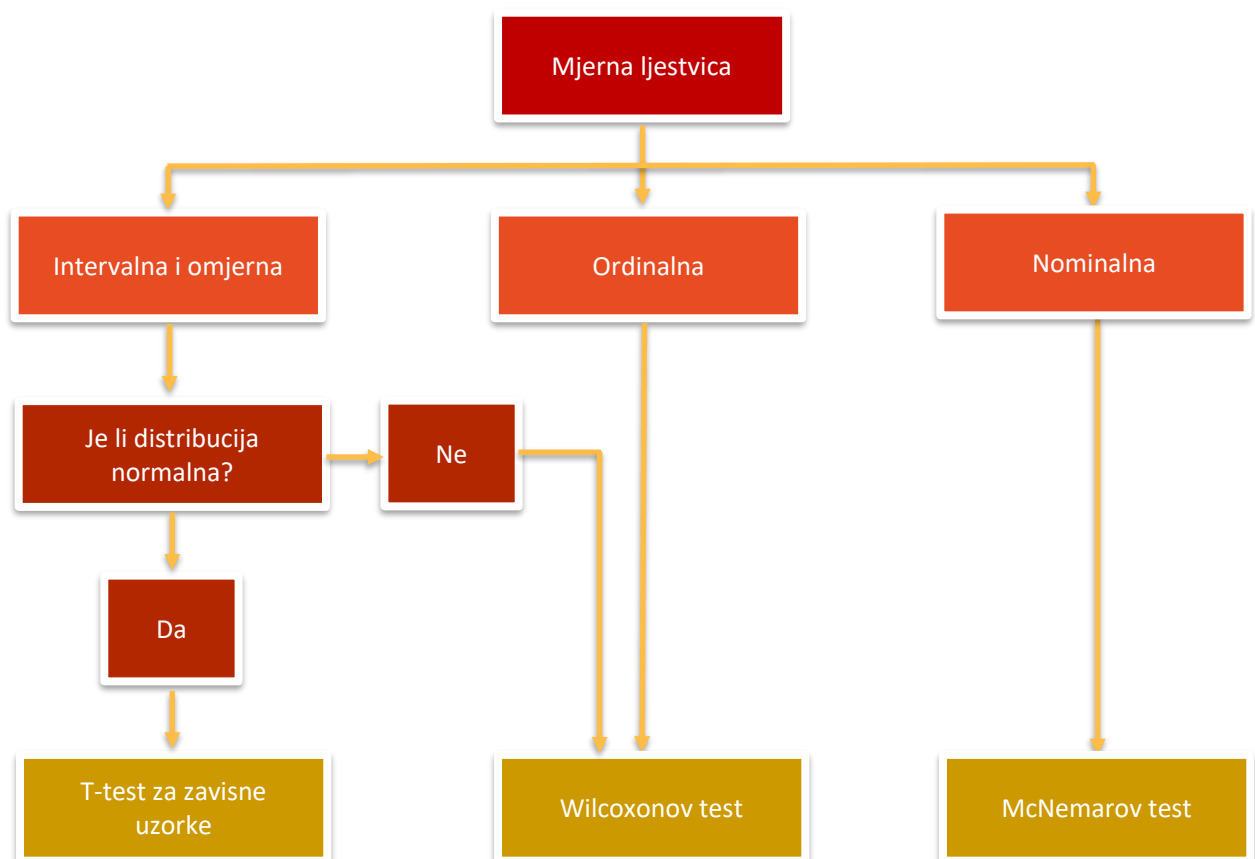
Osim toga, aplikacija omogućava grafičko prikazivanje podataka putem (**Box and Whiskers plot** i **Density plot**) za prvo i drugo mjerenje.

U kartici **T-Test (one var.)** prikazuju se sljedeći rezultati za svaku odabranu zavisnu varijablu:

- **Mean difference** – razlika između aritmetičkih sredina prvog i drugog mjerenja
- **St. error difference** – standardna pogreška razlika prvog i drugom mjerenja
- **Correlations** – korelacija između prvog i drugog mjerenja

- **Student's t** – Studentova t- vrijednost koja pokazuje koliko je puta razlika između aritmetičkih sredina veća od standardne pogreške razlika
- **df** – broj stupnjeva slobode ( $n - 1$ , gdje je n broj entiteta)
- **p** – pogreška s kojom tvrdimo da je razlika između aritmetičkih sredina prvog i drugog mjerenja statistički značajna
- **Cohen's d** – Cohenova mjera veličine učinka (effect size)
- **Wilcoxon W** – W vrijednost Wilcoxonovog test ekvivalentnih parova
- **Wilcoxon p** – pogreška kojom tvrdimo da je razlika između prvog i drugog mjerenja statistički značajna.
- **Mean Plot with 95 % Confidence Interval** – grafički prikaz vrijednosti aritmetičkih sredina te donje i gornje granice intervala u kome se nalaze aritmetičke sredine populacija prvog i drugog mjerenja uz pogrešku do 5%.

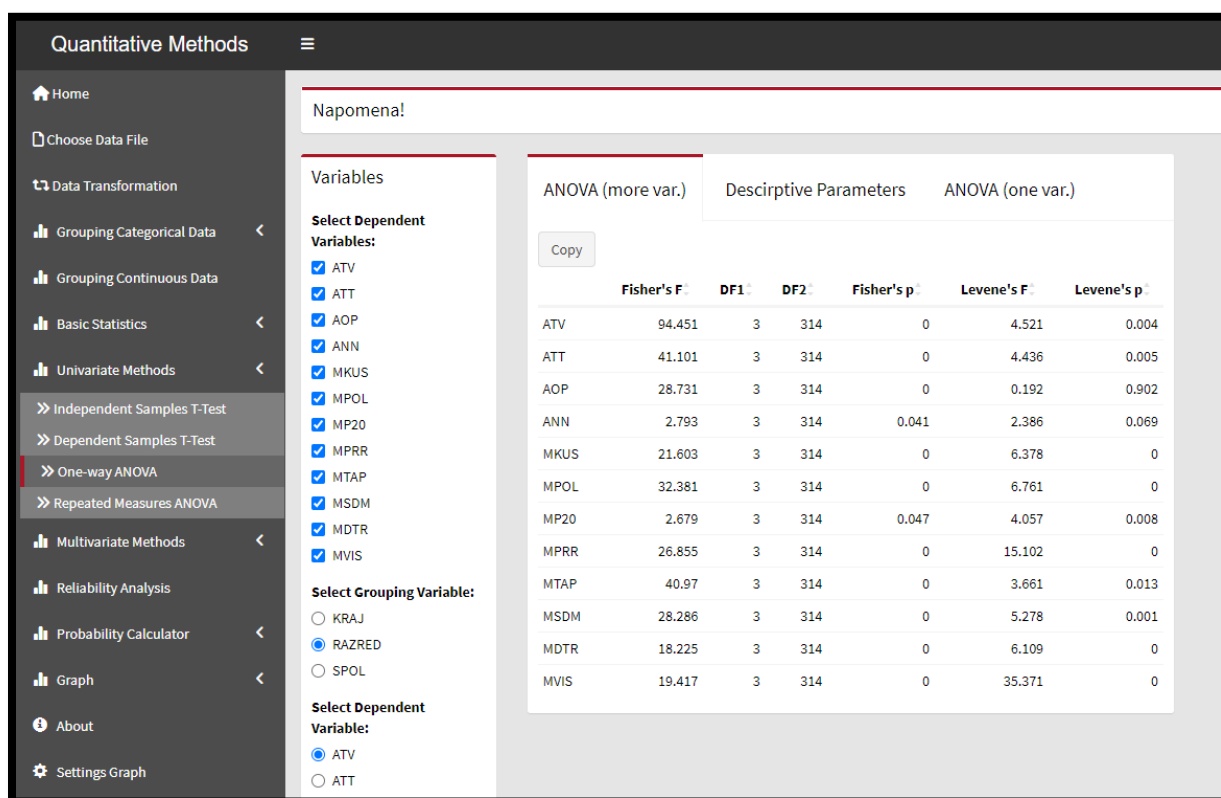
*Testiranje razlika između dviju zavisnih grupa ovisno o vrsti i distribuciji varijabli.*



## ONE WAY ANOVA

Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvalitativnu varijablu s dva ili više modaliteta i jednu ili više kvantitativnih varijabli. Primjerice, u datoteci UCENICI.csv nalaze se podaci 318 entiteta opisanih s tri kvalitativne varijable (KRAJ, RAZRED i SPOL) te 12 kvantitativnih varijabli pri čemu varijable KRAJ i SPOL imaju po dva, a varijabla RAZRED četiri modaliteta.

Odabirom opcije **Univariate Methods** glavnog izbornika te opcije **One Way ANOVA** za odabrane varijable (**Select Dependent Variables** i **Select Grouping Variable**) prikazuju se rezultati jednofaktorske analize varijance.



Rezultati opcije One Way ANOVA aplikacije Quantitative Methods

Da bismo koristili jednofaktorsku analizu varijance, podaci moraju ispunjavati sljedeće zahtjeve (pretpostavke):

- Nezavisnu varijablu čine podaci prikupljeni na nominalnoj ili ordinalnoj mjernoj ljestvici s dva ili više modaliteta (npr. muškarci-žene, bek-krilo-centar i sl.).
- Zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjernoj mjernoj skali (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, taping rukom i sl.).
- Podaci su prikupljeni na slučajno odabranim i relativno velikim uzorcima entiteta ( $n > 30$ ).
- Podaci u zavisnoj varijabli imaju normalnu ili Gaussovu distribuciju. Ukoliko ova pretpostavka nije ispunjena onda se umjesto One Way ANOVA može koristiti neparametrijski Kruskal-Wallis test koji ne zahtijeva normalnost distribucije podataka u zavisnoj varijabli.
- Podaci u zavisnoj varijabli dviju grupa entiteta imaju homogene varijance. Ukoliko ova pretpostavka nije ispunjena, onda se računa korigirana Welchova F-vrijednost.

- Podaci u zavisnoj varijabli ne bi trebali imati značajne outliere, odnosno ekstremno visoke ili niske rezultate.

U kartici **ANOVA (more var.)** prikazuju se tablica sa sljedećim rezultatima za svaku odabranu zavisnu varijablu:

- **Fisher's F** - vrijednost koja pokazuje omjer varijance između i unutar grupa
- **DF1 i DF2** – broj stupnjeva slobode ( $df_1 = k - 1$ ,  $df_2 = n - k$ , gdje je  $n$  broj entiteta,  $k$  – broj grupa)
- **Fisher's p** – pogreška kojom tvrdimo da je razlika između aritmetičkih sredina analiziranih grupa statistički značajna
- **Levene's F** – Levenova F-vrijednost temeljem koje se utvrđuje homogenost varijanci, odnosno jesu li varijance analiziranih grupa statistički značajno različite
- **Leven's p** – pogreška s kojom tvrdimo da su varijance analiziranih grupa statistički značajno različite. Ako je  $p < 0,05$ , zaključujemo da varijance analiziranih grupa nisu homogene, odnosno da se statistički značajno razlikuju uz pogrešku manju od 5%. U tom slučaju umjesto Fisherove F-vrijednosti, računa se Welchova F-vrijednost.

U kartici **Descrptive Parameters** prikazuje se tablica sa sljedećim deskriptivnim pokazateljima za prvo i drugo mjerenje:

- **MEAN** – aritmetička sredina
- **SD** – standardna devijacija
- **SEM** – standardna pogreška aritmetičke sredine
- **MEAN-CI95%** – donja granica intervala u kojem se s vjerojatnošću 95% nalazi aritmetička sredina populacije
- **MEAN+CI95%** – gornja granica intervala u kojem se s vjerojatnošću 95% nalazi aritmetička sredina populacije
- **S-W** – vrijednost Shapiro-Wilkovog testa za procjenu normaliteta distribucije
- **P** – pogreška s kojom tvrdimo da se empirijska distribucija statistički značajno razlikuje od normalne ili Gaussove.

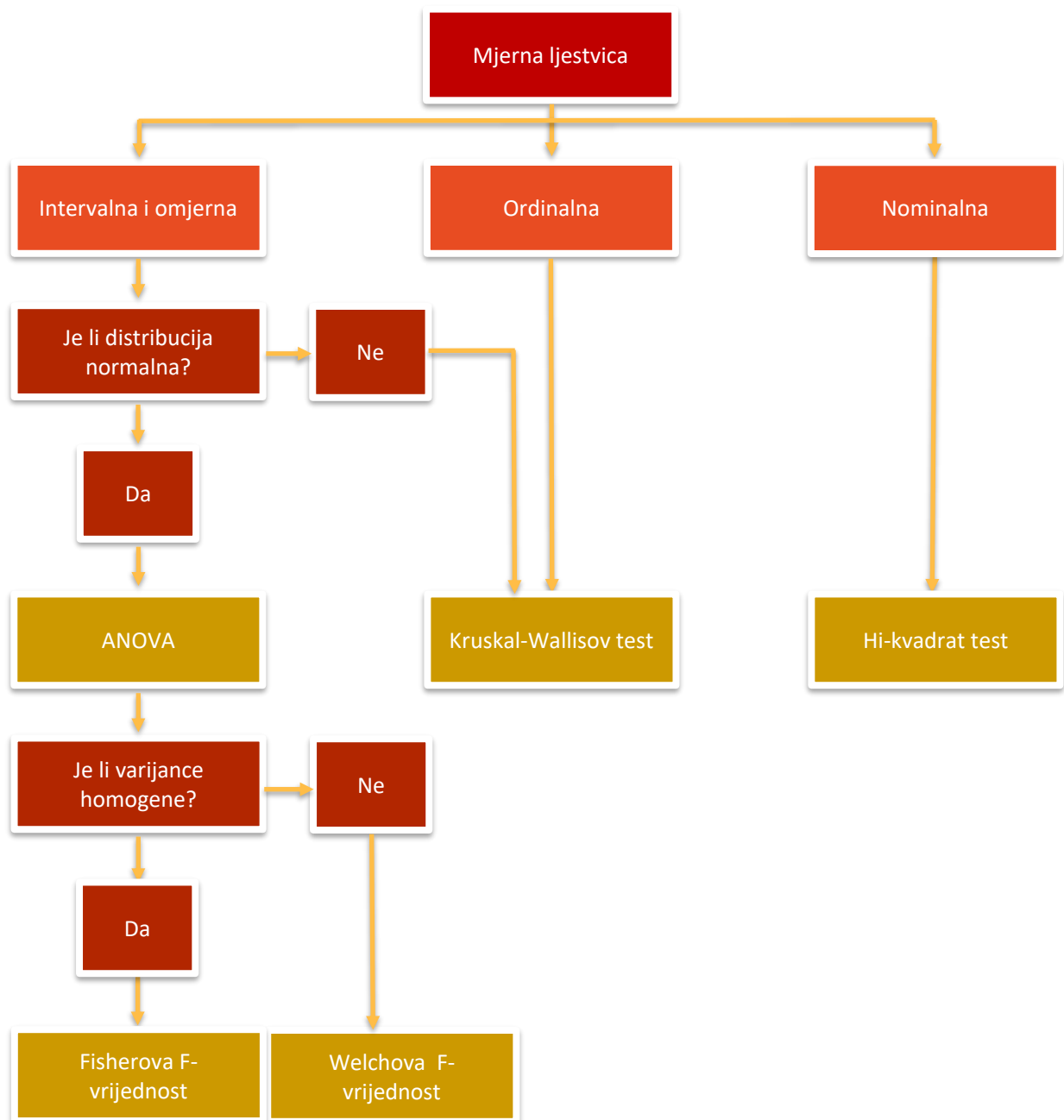
Ako je  $P < 0,05$ , zaključujemo da empirijska distribucija statistički značajno odstupa od normalne ili Gaussove distribucije uz pogrešku  $P$  tada bi umjesto One Way ANOVA bilo primjereno koristiti Kruskal - Wallisov test. Osim toga, aplikacija omogućava grafičko prikazivanje podataka (**Box and Whiskers plot i Density plot**) za analizirane grupe.

U kartici **ANOVA (one var.)** prikazuju se sljedeći rezultati za svaku odabranu zavisnu varijablu:

- **Levene's F** – Levenova F-vrijednost temeljem koje se utvrđuje homogenost varijanci, odnosno jesu li varijance analiziranih grupa statistički značajno različite
- **Leven's p** – pogreška s kojom tvrdimo da su varijance analiziranih grupa statistički značajno različite. Ako je  $p < 0,05$ , zaključujemo da varijance analiziranih grupa nisu homogene, odnosno da se statistički značajno razlikuju uz pogrešku manju od 5%. U tom slučaju umjesto Fisherove F-vrijednosti, računa se Welchova F-vrijednost
- **Fisher's F** - vrijednost koja pokazuje omjer varijance između i unutar grupa
- **df1 i df2** – broj stupnjeva slobode ( $df_1 = k - 1$ ,  $df_2 = n - k$ , gdje je  $n$  broj entiteta,  $k$  – broj grupa)

- **p** – pogreška kojom tvrdimo da je razlika između aritmetičkih sredina analiziranih grupa statistički značajna
- **Post Hoc Test for Multiple Comparisons** – tablica rezultata post hoc testa za višestruko uspoređivanje<sup>2</sup>
- **Mean Plot with 95% Confidence Interval** – grafički prikaz vrijednosti aritmetičkih sredina te donje i gornje granice intervala u kome se nalaze aritmetičke sredine populacija analiziranih grupa uz pogrešku do 5%.

*Testiranje razlika između dviju ili više nezavisnih grupa ovisno o vrsti i distribuciji varijabli.*



<sup>2</sup>Ako smo utvrdili da analizirane grupe ne pripadaju istoj populaciji (statistički se značajno razlikuju), postavlja se pitanje koje se od njih međusobno statistički značajno razlikuju. Za odgovor na to pitanje morali bismo međusobno uspoređivati sve parove uzoraka. U tu svrhu koristimo testove za višestrukog uspoređivanja ili post-hoc testove (Tukey, Scheffe, Duncan i dr.).

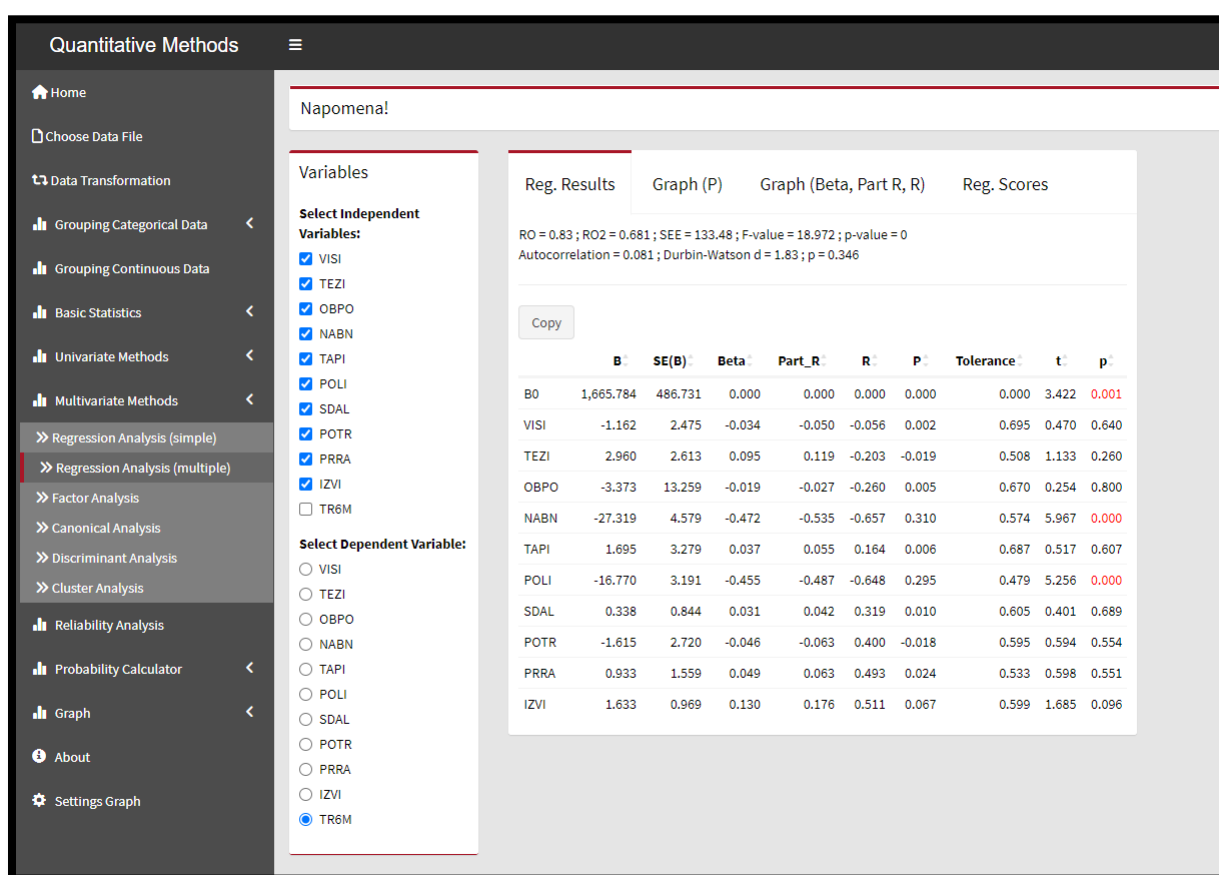


# MULTIVARIATE METHODS

## REGRESSION ANALYSIS

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje dvije kvantitativne varijable. Primjerice, u datoteci SKOLA.csv nalaze se podaci 100 entiteta opisanih s 11 kvantitativnih varijabli.

Odabirom opcije **Multivariate Methods** glavnog izbornika te opcije **Regression Analysis** za odabrane nezavisne varijable (**Select Independent Variables**) i zavisnu varijablu (**Select Dependent Variable**) prikazuje se rezultati regresijske analize.



Rezultati opcije Regression Analysis aplikacije Quantitative Methods

Da bismo koristili regresijsku analizu, podaci moraju ispunjavati sljedeće zahtjeve (pretpostavke):

- Varijable čine podaci dobiveni pomoću intervalne ili omjerne mjerne ljestvice (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, taping rukom i sl.).
- Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ( $n > 30$ ) jer rezultati ne moraju pokazati stabilnost pri ponovljenoj analizi na drugom uzroku iz iste populacije
- Podaci u varijablama ne bi trebali imati značajne *outliere*, odnosno ekstremno visoke ili niske rezultate.
- Varijable su u linearnom odnosu.

- Varijable imaju normalnu ili Gaussovu distribuciju.
- Rezidualne vrijednosti međusobno su nezavisne (autokorelacija prvog reda nije statistički značajna) te su nezavisne u odnosu na zavisnu i nezavisne varijabe. Ova pretpostavka provjerava se pomoću Durbin-Watsonovog testa.
- Varijanca rezidualnih vrijednosti konstantna je za sve vrijednosti u zavisnoj varijabli (homoscedastičnost).
- Rezidualne vrijednosti imaju normalnu ili Gaussovu distribuciju.

U kartici **Regression Results** prikazuju se sljedeći pokazatelji za odabranu zavisnu i nezavisne varijable:

- **RO** – koeficijent multiple korelacije
- **RO2** – koeficijent determinacije multiple korelacije
- **SEE** – standardna pogreška prognoze
- **F-value** – F-vrijednost kojom se izračunava pri testiranju statističke značajnosti multiple korelacije
- **p-value** – pogreška kojom tvrdimo da je multipla korelacija statistički značajna
- **Autocorrelation** – autokorelacija prvog reda za  $n-1$  parova rezidualnih vrijednosti
- **Durbin-Watson** – vrijednost Durbin-Watsonovog test statističke značajnosti autokorelacije prvog reda
- **p** – pogreške s kojima tvrdimo da je autokorelacija prvog reda rezidualnih vrijednosti statistički značajna.

te tablica u kojoj su prikazani sljedeći pokazatelji:

- **B** – regresijski koeficijenti
- **SE(B)** – standardne pogreške regresijskih koeficijenata
- **Beta** – standardizirani regresijski koeficijenti
- **Part R** – koeficijenti parcijalne korelacije
- **R** – koeficijenti korelacije
- **P** – parcijalni koeficijenti determinacije (relativni udio svake prediktorske varijable u objašnjenom varijabilitetu kriterijske varijable)
- **Tolerance** – neobjašnjeni dio varijance svake prediktorske varijable u odnosu na ostale (količina nezavisnih informacija)
- **t** – t-vrijednosti koje se izračunavaju pri testiranju statističke značajnosti regresijskih koeficijenata
- **p** – pogreške s kojima tvrdimo da su regresijski koeficijenti statistički značajni.

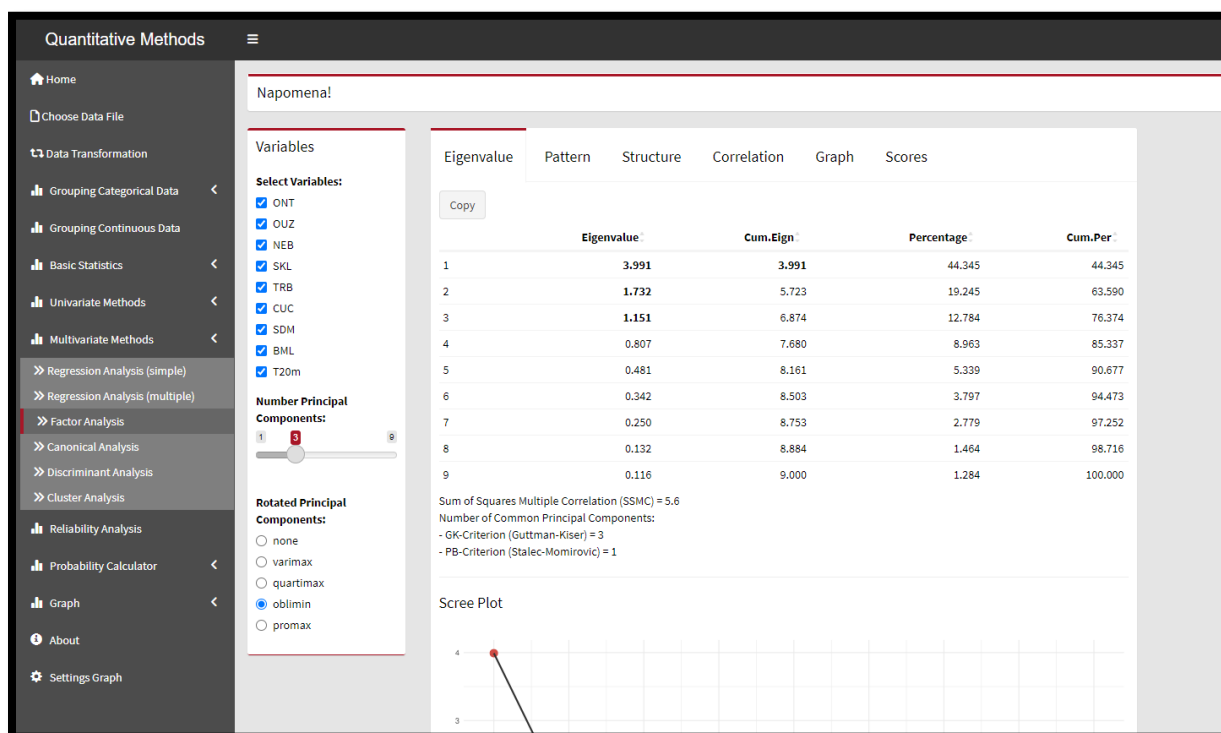
Osim toga, prikazuju se i grafikoni parcijalnih koeficijenata determinacije (**Graph of Partial Coefficients of Determinations (P)**) i standardiziranih regresijskih koeficijenata, parcijalnih korelacija i korelacija (**Standardized Regression Coefficients, Partial Correlations and Correlations**).

U kartici **Regression Scores** prikazuje se tablica s izmjerenim, prognoziranim i rezidualnim vrijednostima zavisne varijable.

## FACTOR ANALYSIS

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje tri kvantitativne varijable. Primjerice, u datoteci JUDO3F.csv nalaze se podaci 60 entiteta opisanih s 9 kvantitativnih varijabli.

Odabirom opcije **Multivariate Methods** glavnog izbornika te opcije **Factor Analysis** za odabrane varijable (**Select Variables**) prikazuje se rezultati komponentnog modela faktorske analize koju je predložio američki ekonomist i statističar Harold Hotelling.



Rezultati opcije Factor Analysis aplikacije Quantitative Methods

Određivanje broja značajnih glavnih moguće je putem **GK**, **PB** i **Scree plot**a, a za dobivanje jednostavne strukture faktora koriste se ortogonalne (**varimax** i **quatimax**) i neortogonalne (**oblimin** i **promax**) rotacije. Da bismo koristili faktorsku analizu, podaci moraju ispunjavati sljedeće zahtjeve (pretpostavke):

- Varijable čine podaci dobiveni pomoću intervalne ili omjerne mjerne ljestvice (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, taping rukom i sl.).
- Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta (5 od 10 puta više entiteta od broja varijabli).
- Podaci u varijablama ne bi trebali imati značajne outliere, odnosno ekstremno visoke ili niske rezultate.
- Varijable su u linearnom odnosu.
- Varijable imaju normalnu ili Gaussovu distribuciju.

U kartici **Eigenvalue** prikazuje se tablica sa sljedećim rezultatima:

- **Eigenvalue** – svojstvene vrijednosti, odnosno varijance glavnih komponenata
- **Cum. Eign.** – kumulativne svojstvene vrijednosti
- **Percentage** – relativne postotne svojstvene vrijednosti
- **Cum. Per.** – kumulativne relativne postotne svojstvene vrijednosti.

Ispod tablice prikazuje se:

- **Sum of Squares Multiple Correlation (SSMC)** – suma kvadrata multiplih korelacija svake manifestne varijable u odnosu na preostale
- **Number of Common Principal Components** – broj značajnih glavnih komponenata koji se dobije primjenom:
- **GK-Criterion (Guttman-Kaiser)** – Guttman-Kaiserovog kriterija prema kojem su značajne sve glavne komponente čija je varijanca, odnosno svojstvena vrijednost veća ili jednaka 1
- **PB-Criterion (Štalec-Momirović)** – PB – kriterija prema kojem je broj značajnih glavnih komponenata jednak broju svojstvenih vrijednosti poredanih po veličini, čiji zbroj ne prelazi SSMC (sumu kvadrata multiplih korelacija)
- **Scree Plot** – Na scree plotu se subjektivnom procjenom odredi točka nakon koje se svojstvene vrijednosti smanjuju u skladu s blagim linearnim trendom. Značajnima se smatraju sve prethodne glavne komponente.

U kartici:

- **Pattern** nalazi se matrica sklopa, odnosno paralelnih projekcija manifestnih varijabli na faktore te Hoffmanov index kompleksiteta (**Complexity**) svake manifestne varijable<sup>3</sup>. Indeks se kreće u intervalu od 1 do  $k$ , gdje je  $k$  broj faktora. Što je indeks bliži 1, to je kompleksitet varijable manji, a što je bliži broju faktora ( $k$ ) to je veći. Ako je vrijednost indeksa jednaka 1, varijabla objašnjava samo 1 faktor. Ako je jednak  $k$  onda u jednakoj mjeri objašnjava svaki od  $k$  faktora

$$Complexity = \frac{(\sum_{j=1}^k a_j^2)^2}{\sum_{j=1}^k a_j^4}$$

gdje je  $k$  broj faktora, a  $a_j$  faktorsko opterećenje faktora  $j$ .

- **Structure** nalazi se matrica strukture, odnosno ortogonalnih projekcija (korelacija) manifestnih varijabli na faktore te kumunaliteti, odnosno dio varijance manifestne varijable koji je moguće objasniti sa značajnim glavnim komponentama (**Communality**).
- **Correlation** – matrica korelacija između faktora.
- **Graph** – grafikon strukture manifestnih varijabli (**Factor Loadings (Variables)**), grafikon strukture faktora (**Graph Factor Loadings (Factors)**).
- **Scores** nalazi se tablica s rezultatima entiteta u faktorima.

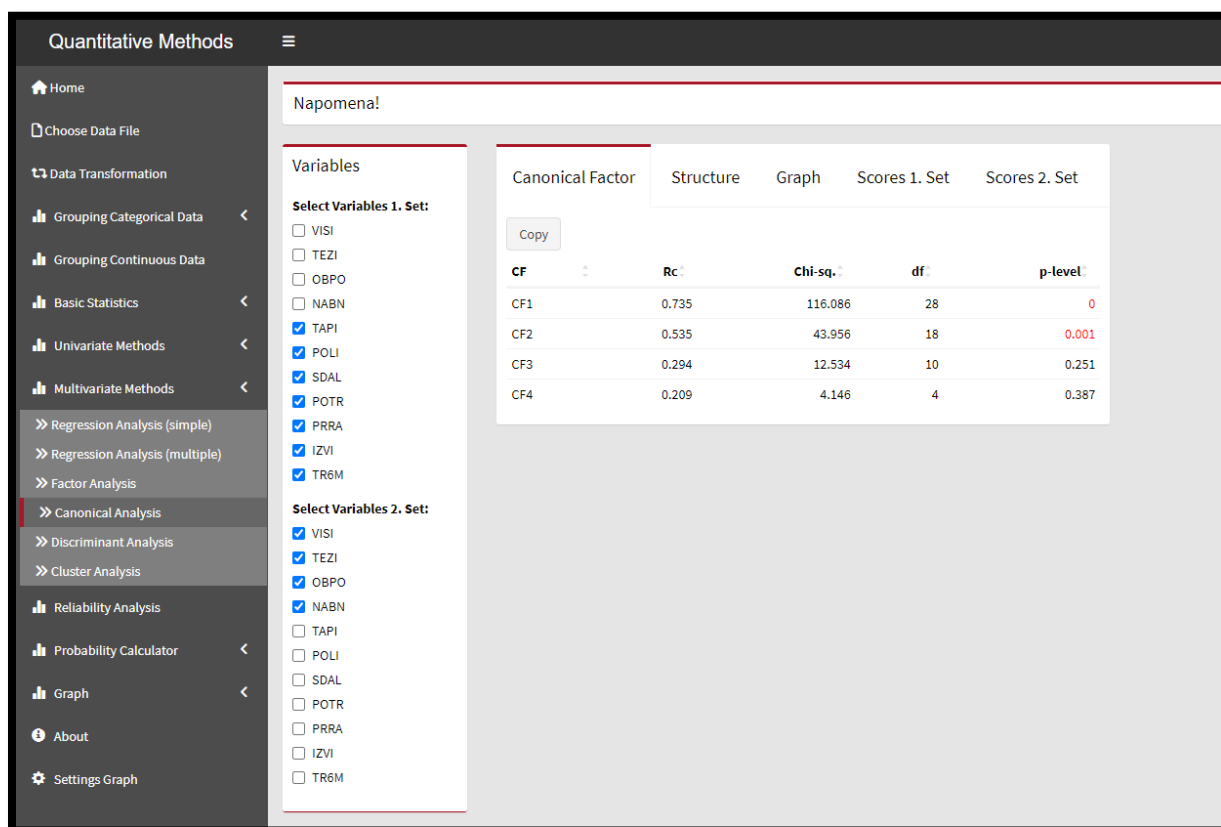
---

<sup>3</sup>Hofmann, R. J. (1978). Complexity and simplicity as objective indices descriptive of factor solutions. *Multivariate Behavioral Research*, 13(2), 247–250

## CANONICAL ANALYSIS

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje tri kvantitativne varijable. Primjerice, u datoteci SKOLA.csv nalaze se podaci 100 entiteta opisanih s 11 kvantitativnih varijabli.

Odabirom opcije **Multivariate Methods** glavnog izbornika te opcije **Canonical Analysis** za odabrane varijable (**Select Variables 1. Set** i **Select Variables 2. Set**) prikazuje se rezultati kanoničke analize.



Rezultati opcije Canonical Analysis aplikacije Quantitative Methods

Da bismo koristili kanoničku analizu, podaci moraju ispunjavati sljedeće zahtjeve (pretpostavke):

- Varijable čine podaci dobiveni pomoću intervalne ili omjerne mjerne ljestvice (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, tapping rukom i sl.).
- Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta (5 do 10 puta više entiteta od broja varijabli).
- Podaci u varijablama ne bi trebali imati značajne *outliere*, odnosno ekstremno visoke ili niske rezultate.
- Varijable su u linearnom odnosu.
- Varijable imaju normalnu ili Gaussovu distribuciju.

U kartici:

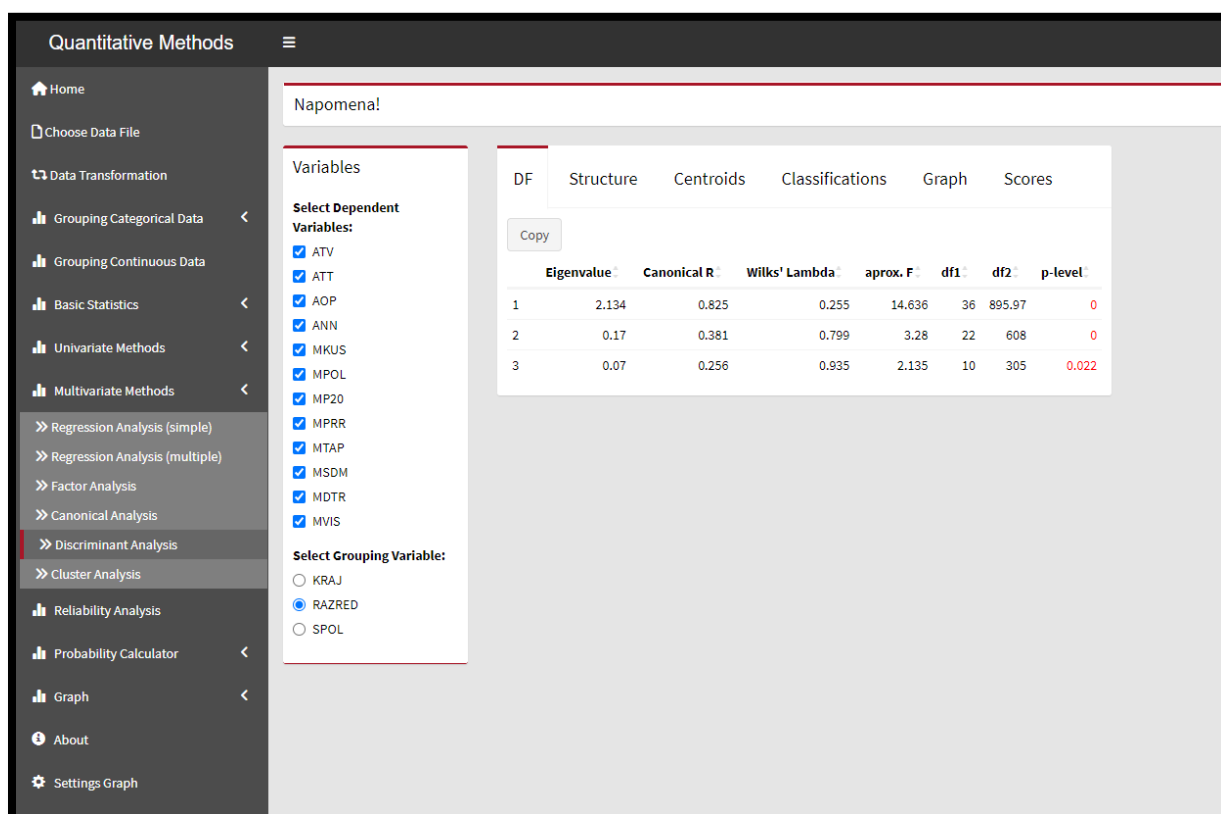
- **Canonical Factor** prikazuju se:
  - **Rc** – koeficijenti kanoničkih korelacija

- **Chi-sq.** – vrijednosti Bartlettovog  $\chi^2$  – testa za testiranje statističke značajnosti kanoničkih korelacija
- **df** – stupnjevi slobode
- **p** – pogreške s kojima tvrdimo da su koeficijenti kanoničke korelacije statistički značajni.
- **Structure nalaze se** korelacija manifestnih varijabli prvog skupa s kanoničkim faktorima prvog skupa (**Factor Structure 1. Set**) i korelacija manifestnih varijabli drugog skupa s kanoničkim faktorima drugog skupa (**Factor Structure 2. Set**).
- **Graph nalaze se** grafikoni strukture kanoničkih faktora prvog (**Graf Factor Structure 1. Set**) i drugog prvog (**Graf Factor Structure 2. Set**) skupa.
- **Scores – 1. Set** i **Scores – 2. Set** nalaze se tablice s rezultatima entiteta u kanoničkim faktorima prvog i drugog skupa.

## DISCRIMINANT ANALYSIS

Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvalitativnu varijablu s dva ili više modaliteta i jednu ili više kvantitativnih varijabli. Primjerice, u datoteci UCENICI.csv nalaze se podaci 318 entiteta opisanih s tri kvalitativne varijable (KRAJ, RAZRED i SPOL) te 12 kvantitativnih varijabli pri čemu varijable KRAJ i SPOL imaju po dva, a varijabla RAZRED četiri modaliteta.

Odabirom opcije **Multivariate Methods** glavnog izbornika te opcije **Discriminant Analysis** za odabrane varijable (**Select Dependent Variables** i **Select Grouping Variable**) prikazuju se rezultati diskriminacijske analize.



Rezultati opcije Discriminant Analysis aplikacije Quantitative Methods

Da bismo koristili diskriminacijsku analizu, podaci moraju ispunjavati sljedeće zahtjeve (pretpostavke):

- Nezavisnu varijablu čine podaci prikupljeni na nominalnoj ili ordinalnoj mjernoj ljestvici s dva ili više modaliteta (npr. muškarci-žene, bek-krilo-centar i sl.).
- Zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjernoj mjernoj skali (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, taping rukom i sl.).
- Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta podjednake veličine (3 od 5 puta više entiteta u svakoj grupi od broja varijabli).
- Podaci u varijablama ne bi trebali imati značajne outlieri, odnosno ekstremno visoke ili niske rezultate.
- Varijable su u linearnom odnosu.
- Varijable imaju normalnu ili Gaussovu distribuciju.

U kartici **DF** nalazi se tablica sa sljedećim rezultatima:

- **Eigenvalue** – svojstvene vrijednosti, odnosno varijance diskriminacijskih funkcija
- **Canonical R** – koeficijenti kanoničke korelacije (diskriminacije), odnosno korelacija diskriminacijskih funkcija s nezavisnom (selektorskom) varijablom
- **Wilks' Lambda** – Wilksove lambde kreću se u intervalu od 0 do 1, a što im je vrijednost manja to je veća vjerojatnost da je razlika između analiziranih grupa statistički značajna
- **aprox. F** – aproksimativne F-vrijednosti temeljem kojih se utvrđuje statistička značajnost diskriminacijskih funkcija
- **df1 i df2** – broj stupnjeva slobode
- **p-level** – pogreške kojima tvrdimo da pojedina diskriminacijska funkcija statistički značajno razlikuje analizirane grupe.

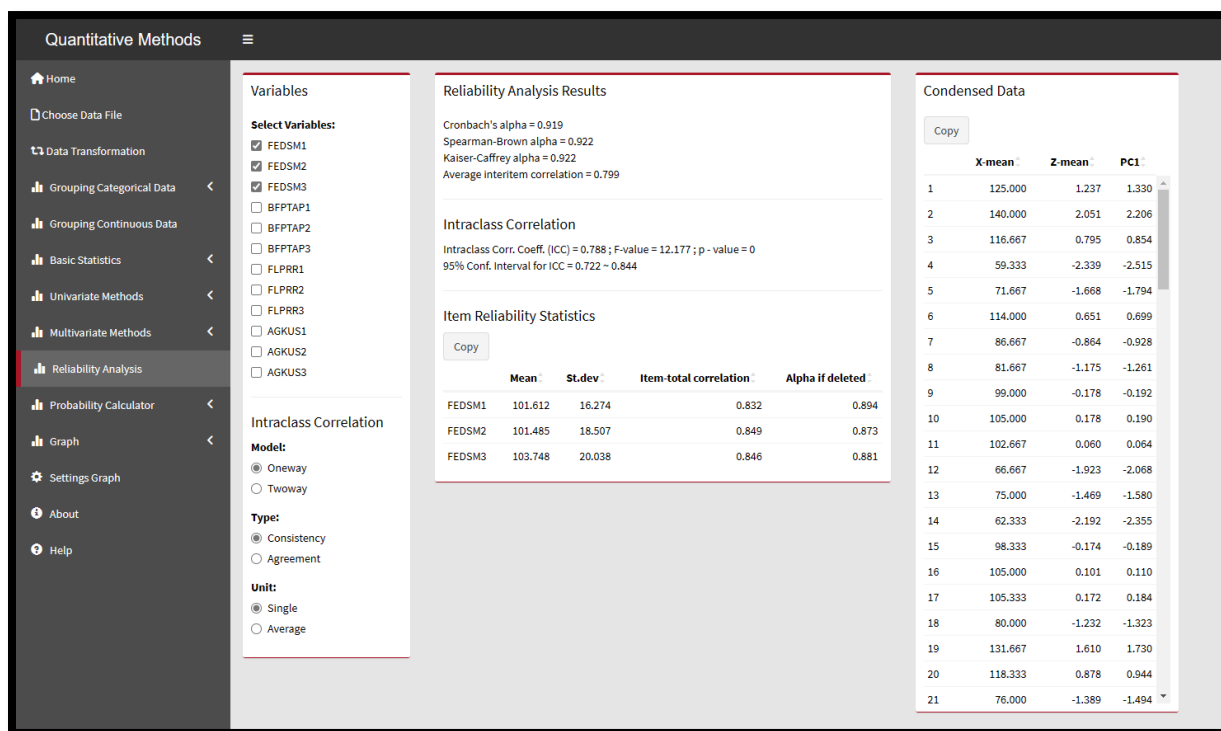
U kartici:

- **Structure nalaze se** korelacije zavisnih varijabli i diskriminacijskih funkcija
- **Centroids** nalaze se aritmetičke sredine analiziranih grupa u diskriminacijskim funkcijama
- **Clasifications** nalaze se tablice koje pokazuju broj i postotak ispravno i neispravno klasificiranih entiteta na temelju diskriminacijskih funkcija
- **Scores** nalazi rezultati entiteta u diskriminacijskim funkcijama.



## RELIABILITY ANALYSIS

Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje dvije kvalitativne varijable (čestice testa). Primjerice, u datoteci GIM.csv nalaze se podaci 103 entiteta opisanih s 12 varijablama (čestica) dobivenih na 4 testa za procjenu motoričkih sposobnosti (FEDSM, BFPTAP, FLPRRA i AGKUS).



Rezultati opcije Reliability Analysis aplikacije Quantitative Methods

Odabirom opcije **Reliability Analysis** za odabrane varijable (čestice testa) prikazuju se sljedeći rezultati:

- **Cronbach's alpha** – Cronbachov koeficijent pouzdanosti
- **Spearman-Brown alpha** – Spearman-Brownov koeficijent pouzdanosti
- **Kaiser-Caffrey alpha** – Kaiser-Caffreyjev koeficijent pouzdanosti
- **Average interitem correlation** – prosječna korelacija između čestica testa
- **Interklasni koeficijent korelacije (ICC)**

Shrout i Fleiss (1979)<sup>4</sup> su definirali šest oblika ICC-a, koji se označavaju s dva broja u zagradama. Prvi broj se odnosi na ocjenjivače:

- 1 – svaki subjekt ocjenjuje drugi ocjenjivač
- 2 – svaki subjekt ocjenjuje slučajno odabran uzorak od  $k$  ocjenjivača
- 3 – svaki subjekt ocjenjuje svi ocjenjivači (nema generaliziranja na širu populaciju)

dok se drugi broj odnosi na konačnu ocjenu:

- 1 - jedan ocjenjivač
- $k$  - aritmetička sredina  $k$  ocjenjivača.

<sup>4</sup> Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420. <https://doi.org/10.1037/0033-2909.86.2.420>

Odabirom odgovarajućih opcija:

- **model** (oneway ili twoway)
- **type** (consistency ili agreement)
- **unit** (single ili average)

računa se odgovarajući ICC.

Ako svaki subjekt ocjenjuje drugi ocjenjivač treba odabrati opciju **oneway** pri čemu opcije **consistency** i **agreement** daju iste rezultate.

Ako svaki subjekt ocjenjuje slučajno odabran uzorak od  $k$  ocjenjivača treba odabrati opciju **twoway** pri čemu umjesto opcije **consistency** treba odabrati opciju **agreement** ukoliko su razlike u prosječnim ocjenama ocjenjivača bitne.

Ako konačna ocjena predstavlja aritmetičku sredinu dobivenu od više ocjenjivača treba odabrati opciju **average**, a ako ne onda **single**.

*Vrste ICC-a prema Shrout i Fleiss (1979).*

Oznaka ICC-a	Odabir opcija za računanje ICC-a	Formula za računanje ICC <sup>5</sup>
ICC(1,1)	oneway, consistency or agreement, single	$\frac{MS_R - MS_W}{MS_R + (k - 1)MS_W}$
ICC(2,1)	twoway, agreement, single	$\frac{MS_R - MS_E}{MS_R + (k - 1)MS_E + \frac{k}{n}(MS_C + MS_E)}$
ICC(3,1)	twoway, consistency, single	$\frac{MS_R - MS_E}{MS_R + (k - 1)MS_E}$
ICC(1,k)	oneway, consistency or agreement, average	$\frac{MS_R - MS_W}{MS_R}$
ICC(2,k)	twoway, agreement, average	$\frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$
ICC(3,k)	twoway, consistency, average	$\frac{MS_R - MS_E}{MS_R}$

Tablica **Item Reliability Statistics** u kojoj se nalaze sljedeći rezultati:

- **Mean** – aritmetičke sredine čestica testa
- **St.dev.** – standardne devijacije čestica testa
- **Item-total correlation** – korelacije čestica s jednostavnom linearnom kombinacijom preostalih
- **Alpha if deleted** – Cronbachov koeficijent pouzdanosti izračunat bez pripadajuće čestice.

<sup>5</sup>MS<sub>R</sub> - srednji kvadrat za retke; MS<sub>W</sub> - srednji kvadrat za rezidualne izvore varijance; MS<sub>E</sub> - srednja vrijednost kvadrat za grešku; MS<sub>C</sub> - srednji kvadrat za stupce; n - broj ispitanika; k - broj ocjenjivača.

U prozoru **Condensed Data** nalaze se kondenzirani rezultati entiteta u testu izračunati kao:

- **X-mean** – aritmetička sredina originalnih rezultata
- **Z-mean** – aritmetička sredina rezultata transformiranih u z-vrijednosti
- **PC1** – prva glavna komponenta.

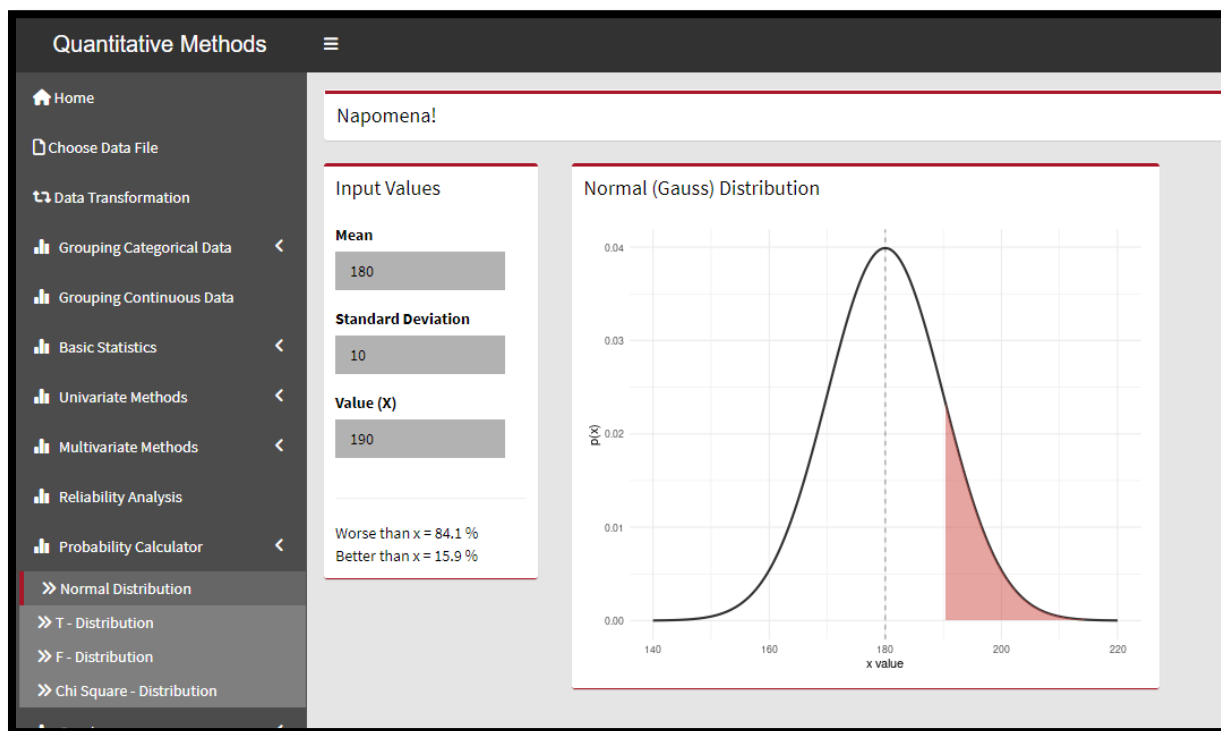
Koeficijenti pouzdanosti koji se utvrđuju metodom interne konzistencije odabiru se s obzirom na način kondenzacije rezultata, odnosno izračunavanje ukupnog rezultata ispitanika na temelju pripadajućih rezultata u česticama testa. Ako se kondenzacija rezultata nekog kompozitnog testa vrši:

- jednostavnom linearnom kombinacijom originalnih rezultata (zbroy ili aritmetička sredina), onda je mjera pouzdanosti Cronbachova  $\alpha$
- jednostavnom linearnom kombinacijom standardiziranih rezultata (zbroy ili aritmetička sredina), onda je mjera pouzdanosti Spearman-Brownova  $\alpha$
- prvom glavnom komponentom, onda je mjera pouzdanosti Kaiser-Caffreyeva  $\alpha$ .

# PROBABILITY CALCULATOR

## NORMAL DISTRIBUTION

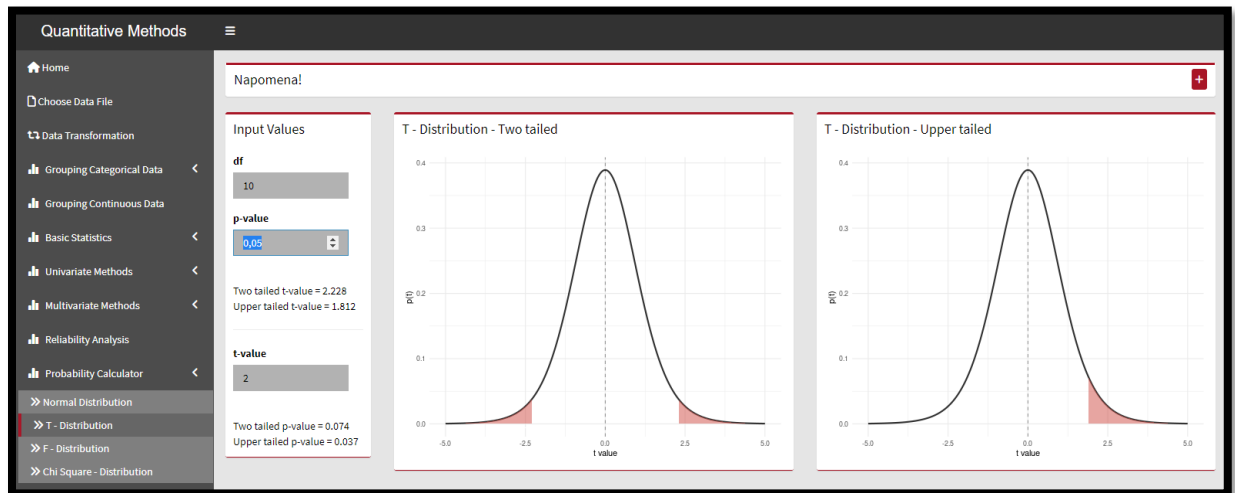
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **Normal Distribution** omogućava računanje postotka boljih (**Better than x**) i lošijih (**Worse than x**) entiteta od određenog rezultata (**X value**) u populaciji entiteta čije obilježje (varijabla) ima normalnu ili Gaussovu distribuciju s određenom aritmetičkom sredinom (**Mean**) i standardnom devijacijom (**Standard Deviation**).



*Rezultati opcije Normal Distribution aplikacije Quantitative Methods*

## T – DISTRIBUTION

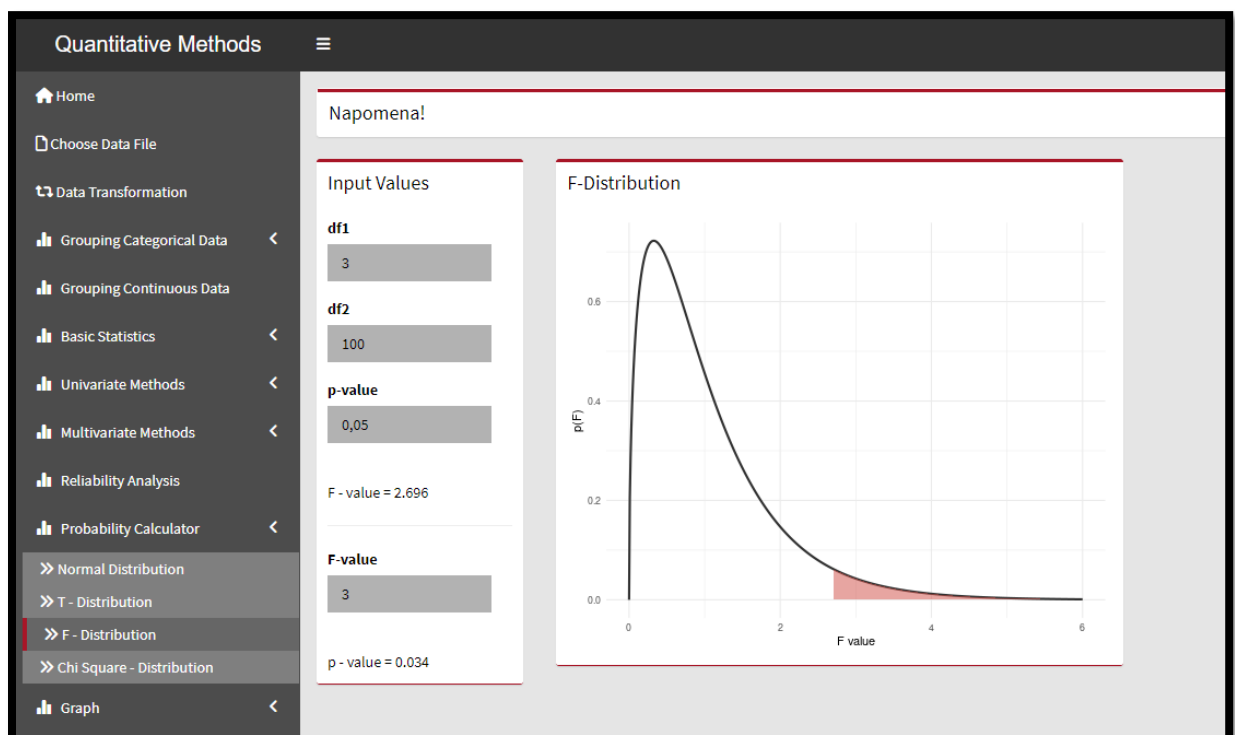
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **T-Distribution** omogućava računanje odgovarajuće t-vrijednosti (**Two tailed t-value** i **Upper tailed t-value**) za zadanu p-vrijednost (**p-value**) i broj stupnjeva slobode (**df**).



*Rezultati opcije T- Distribution aplikacije Quantitative Methods*

## F – DISTRIBUTION

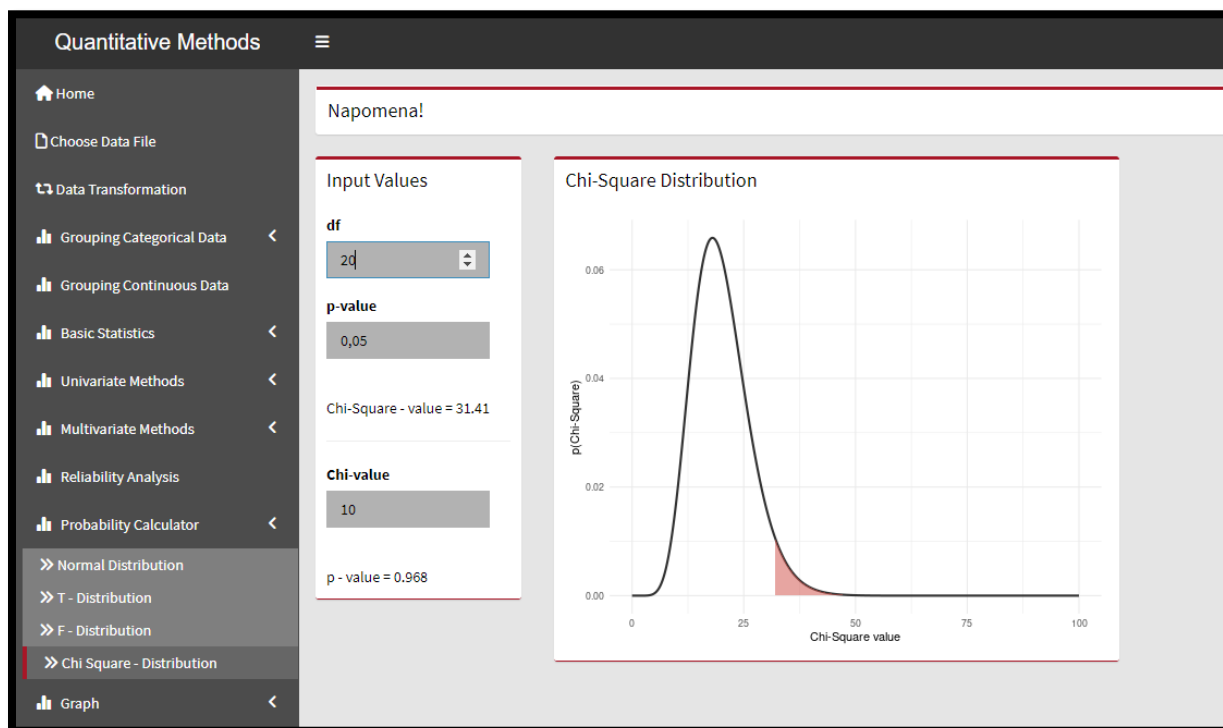
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **F-Distribution** omogućava računanje odgovarajuće F-vrijednosti (**F-value**) za zadanu p-vrijednost (**p-value**) i broj stupnjeva slobode (**df<sub>1</sub>** i **df<sub>2</sub>**).



*Rezultati opcije F- Distribution aplikacije Quantitative Methods*

## CHI SQUARE – DISTRIBUTION

Odabir opcije **Probability Calculator** glavnog izbornika te opcije **Chi Square-Distribution** omogućava računanje odgovarajuće  $\chi^2$ -vrijednosti (**Chi Square-value**) za zadanu p-vrijednost (**p-value**) i broj stupnjeva slobode (**df**).



*Rezultati opcije Chi-Distribution aplikacije Quantitative Methods*