



KVANTITATIVNE METODE ZA RSTUDIO

PRIRUČNIK

DRAŽAN DIZDAR
DARKO KATOVIĆ

Što je R?	3
Što je RStudio?.....	3
Vrste datoteka u RStudiju	5
R aplikacija Quantitative Methods.....	7
Choose CSV Data File	8
Data Transformation.....	9
Grouping Categorical Data	10
Frequency Tables	10
Contingency Tables.....	11
Basic Statistics	12
Descrptive Parameters	12
Testing for Normality	13
CI for the Population Mean.....	15
Correlation Analysis	16
Univariate Methods.....	18
Independent Samples T-Test.....	18
Dependent Samples T-Test.....	22
One – Way ANOVA	25
Multivariate Methods	28
Regression Analysis.....	28
Factor Analysis	30
Canonical Analysis	32
Discriminant Analysis	33
Reliability Analysis	35
Podaci	37
GIM.csv (n = 103, m = 12)	37
JUDO3F.csv (n = 60, m =9).....	37
KM.csv (n = 40, m = 3).....	38
KOS.csv (n = 26, m = 8).....	38
MUHCS20m.csv (n = 31, m = 3).....	38
POD.csv (n=120, m=5).....	39
SKOLA.csv (n = 100, m = 11).....	39
TEST.csv (n = 200, m = 2)	39

UCENICI.csv (n = 318, m = 16).....	40
VZA.csv (n = 60, m = 12).....	41

ŠTO JE R?

R je programski jezik i softversko okruženje za statističko-grafičku analizu podataka koje se može slobodno koristiti i distribuirati te pripada grupi softverskih alata otvorenog kôda (engl. open-source).

R sadrži veliki broj statističkih i grafičkih alata kao što su osnovne statističke metode, multivarijatne metode te specijalizirane pakete za genetiku, psihologiju, ekonometriju, itd.

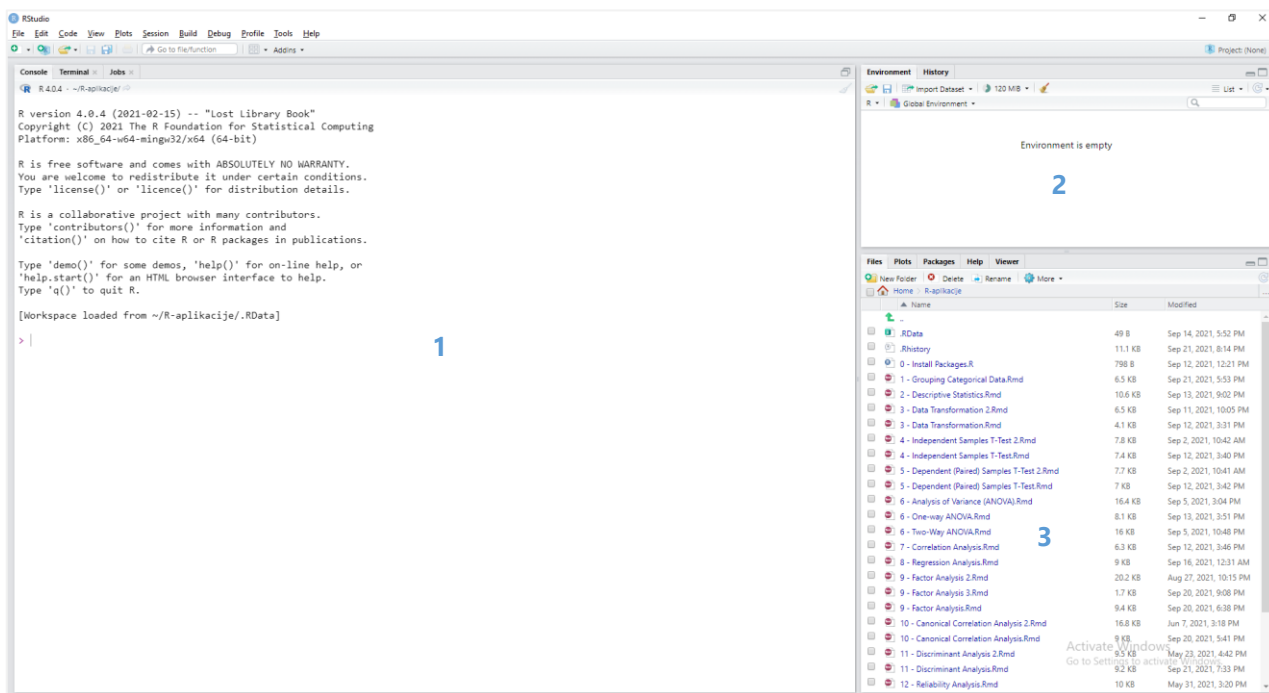
R se može nadograđivati paketima (engl. packages) koji predstavljaju skup funkcija za jednostavnije izvršavanje određenih zadataka, što je u velikoj mjeri pridonijelo njegovoj popularizaciji. Danas se R koristi u gotovo svim granama znanosti i industrije. Više o sustavu R može se pronaći na poveznici <https://www.r-project.org/>.

Za instalaciju R-a potrebno je na web-stranici <http://cran.r-project.org/bin/> odabrati odgovarajuću verziju (Windows, Mac, Linux) te je instalirati na osobno računalo.

ŠTO JE RSTUDIO?

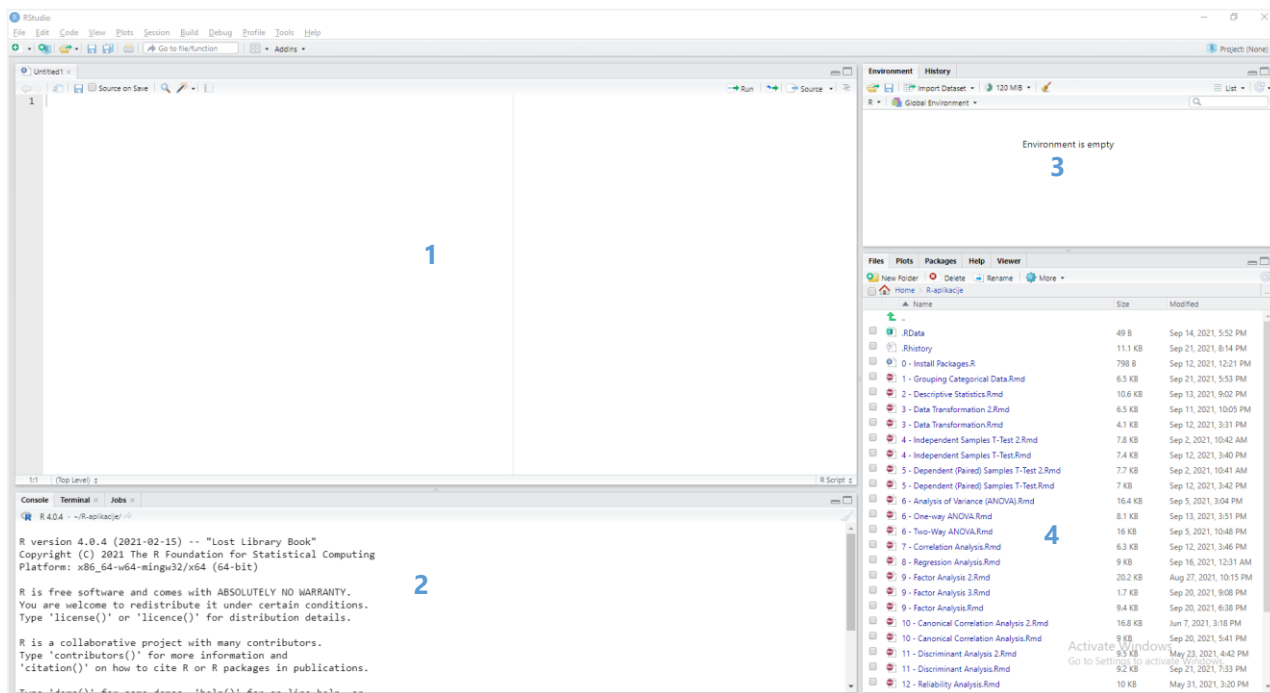
RStudio je korisničko sučelje, odnosno integrirano razvojno okruženje (engl. Integrated Development Environment, IDE) za R koje se može preuzeti na web-stranici:

<https://www.rstudio.com/products/rstudio/download/#download>



Za rad u RStudiju potrebno je prethodno instalirati R. Kada se RStudio prvi put pokrene otvori se prozor s trima oknima: **Console (1)**, **Environment (2)** i **Files (3)**. Ako u izborniku

odaberemo opciju **File -> New File** te odgovarajući tip datoteke (R Script, R Notebook, R Markdown, Shiny Web App...) u sučelju programa RStudio otvori se i okno **Source** s sadržajem datoteke s kojom želimo raditi.



Console (2) je okno preko kojeg se izvršavaju sve naredbe i prikazuju rješenja sustava R. Naredbe koje se upisuju direktno u konzolu izvršavaju se nakon pritiska tipke [ENTER]. Istovremeno se može napisati i izvršiti samo jedna naredba. Sav kôd iz konzole nestaje nakon prekida rada. U slučaju da se u konzoli pojavi simbol +, to je znak da naredba iz konzole ili okna Source nije dovršena i da se čeka nastavak unosa.

Okno **Environment** (3) ili radno okruženje jest preglednik svih objekata (varijabli, skupova podataka, funkcija...) koji su kreirani tijekom rada. U kartici **History** nalazi se popis od maksimalno 512 prethodno izvršenih naredbi koje se mogu proslijediti u **Source** dokument ili u **Console** na izvršavanje.

Okno **Files** sastoji se od sljedećih kartica:

- **Files** je mjesto na kojem se može odabrati radni direktorij. Radni direktorij nekog procesa je hijerarhijska mapa dokumenata vezana za proces. Kad se unutar procesa stvara ili poziva neka datoteka, njen put (engl. file path) kreće od pozicije radnog direktorija, a ne iz korijenskog direktorija (engl. root directory).
- **Plots** služi za prikaz grafikona.
- **Packages** daje pregled svih dostupnih paketa. Kvačicom su označeni paketi koji su trenutno aktivni. Popis paketa nadopunjava se po potrebi naredbom `install.packages()`

- **Help** je kartica za pomoć za bilo koji paket, funkciju, ugrađeni podatkovni skup i za sve ostalo za što postoji službena R dokumentacija. Ako nas na primjer, zanima kako se koristi funkcija `plot()`, upišemo u tražilicu `plot` i dobit ćemo dokumentaciju o toj funkciji.
- **Viewer** služi za pregled lokalnoga web-sadržaja.

VRSTE DATOTEKA U RSTUDIJU

- **R Script** je jednostavna tekstualna datoteka u koju se upisuju naredbe za R. Ekstenzija datoteka ovoga tipa je **.R**. Naredbe se izvršavaju istovremenim pritiskom tipki `[CTRL]+[ENTER]`, a rezultat se prikazuje u konzoli, a grafikoni se prikazuju na kartici **Plots**. Izvršavanje svih naredbi odjednom moguće je pritiskom na tipke `[CTRL]+[A]`, a nakon toga `[CTRL]+[ENTER]` ili istovremenim pritiskom tipki `[CTRL]+[ALT]+[R]`. Komentari se mogu pisati iza znaka `#`. Prednost R Scripta nad konzolom je u tome što sadrži pregled svih naredbi koje korisnik lako može ponoviti, ispraviti, doraditi i pohraniti.
- **R Markdown** je vrsta R datoteke u kojoj je moguće kombinirati programski kôd, njegove rezultate i tekst te generirati izlazni dokument u HTML, PDF ili Word formatu. Ekstenzija datoteka ovog tipa je **.Rmd**. S obzirom na to da je cilj svake analize podataka predstaviti dobivene rezultate u nekoj vrsti dokumenta, R Markdown datoteka omogućava dokumentiranje cijelog postupka analize podataka te njegovo ponavljanje i dorađivanje. Naredbe se pišu unutar blokova (engl. chunk) koji se stvaraju istovremenim pritiskom tipki `[CTRL]+[ALT]+[I]`. Pokretanje jedne naredbe unutar bloka izvršava se istovremenim pritiskom tipki `[CTRL]+[ENTER]`, a pokretanje svih naredbi iz jednog bloka izvršava se istovremenim pritiskom na `[CTRL]+[SHIFT]+[ENTER]`. Tekst u **.Rmd** datoteci koji je napisan izvan bloka piše se u običnom formatu, ali postoji sintaksa (jezik Markdown) koja omogućava oblikovanje teksta u izlaznoj datoteci. Više o markdown sintaksi može se pronaći na poveznici:

<https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>.

R Markdown dokument konvertira se u HTML, PDF, ili Word dokument pomoću gumba **Knit**. Ako se radi o **R Markdown Flex Dashboard Shiny** dokumentu, tada se pokreće (izvršava) klikom na gumb **Run Dokument**, a rezultati se prikazuju u web pregledniku.

- **Shiny Web Applications** su datoteke kreirane kao interaktivne web aplikacije čiju izradu i korištenje omogućava **shiny** paket. Dodatnu pogodnost nudi paket

shinydashboard koji omogućava izradu interaktivnih nadzornih ploča (engl. dashboard) za prezentaciju rezultata aplikacije.

Sve funkcionalnosti sustava R grupirane su u paketima. R paketi sastoje se od skupa funkcija. Određeni paketi dolaze sa samom instalacijom sustava R, a ostale korisnik nadograđuje prema svojim potrebama. Osnovni paketi (**base**, **stats**, **datasets**, **graphics**, **grDevices**, **methods**, **utils**) dolaze sa sustavom R i sadrže osnovne funkcije za aritmetiku, statistiku, grafiku, ulaz i izlaz podataka itd. Ovaj skup paketa ne mijenja se između dviju verzija sustava R i ne mogu se pronaći izdvojeni na nekom repozitoriju paketa. Postoje paketi koji su također dio instalacije sustava R, a koji se mogu mijenjati (nadograđivati). Paketi se preuzimaju pomoću naredbe `install.packages()`, a učitavaju se pomoću naredbe `library()`.

```
install.packages("shinydashboard")  
library(shinydashboard)
```

Popis svih funkcija iz nekog paketa može se vidjeti na ovaj način:

```
library(help = "shinydashboard")
```

Glavni repozitorij ili spremište paketa nalazi se na web-stranici **CRAN - The Comprehensive R Archive Network** na kojoj postoji više od 15 000 paketa, a cijeli popis može se pronaći na poveznici:

https://cran.r-project.org/web/packages/available_packages_by_name.html.

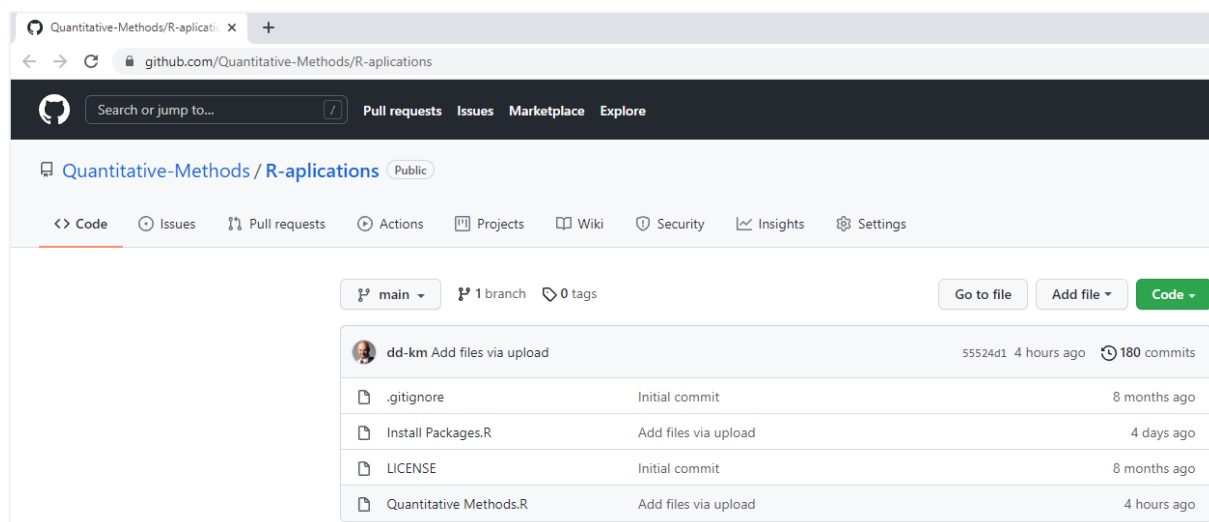
Postoji i pregledniji način pronalaska odgovarajućih paketa, na primjer, na sljedećim poveznicama grupirani su po temi:

<https://cran.r-project.org/web/views/>

R APLIKACIJA QUANTITATIVE METHODS

R aplikaciju **Quantitative Methods** možete preuzeti na ovoj poveznici:

<https://github.com/Quantitative-Methods/R-applications>

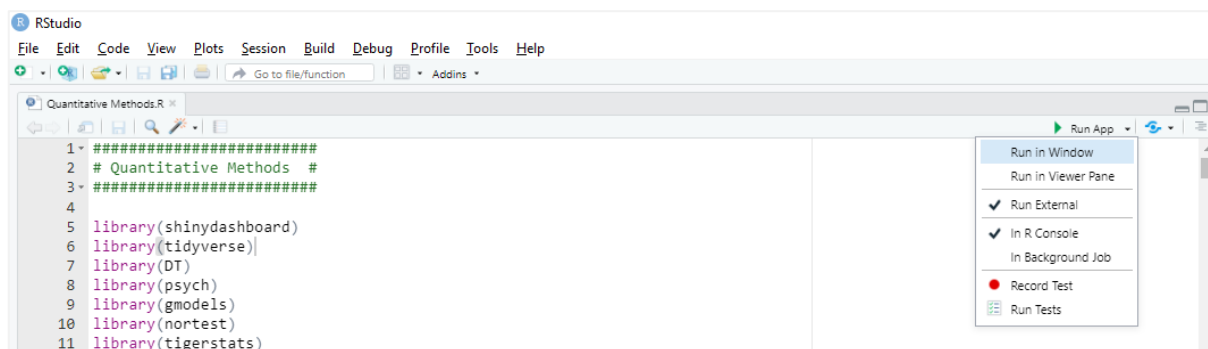


Klikom na gumb **Code** odaberemo opciju **Download ZIP**. Nakon preuzimanja arhiviranu datoteku **R-applications-main.zip** treba „raspakirati” i datoteke **Install Packages.R** i **Quantitative Methods.R** kopirati u neku prethodno kreiranu mapu.

Prije korištenja aplikacije **Quantitative Methods.R** potrebno je instalirati pakete koji se nalaze u datoteci **Install Packages.R**. Nakon što se datoteka učitava i prikaže u oknu **Source** potrebno je pokrenuti sve naredbe unutar datoteke istovremenim pritiskom tipki [CTRL]+[ALT]+[R]. Instalacija paketa potrajat će nekoliko minuta. Nakon instalacije potrebnih paketa aplikacija **Quantitative Methods.R** spremna je za korištenje.

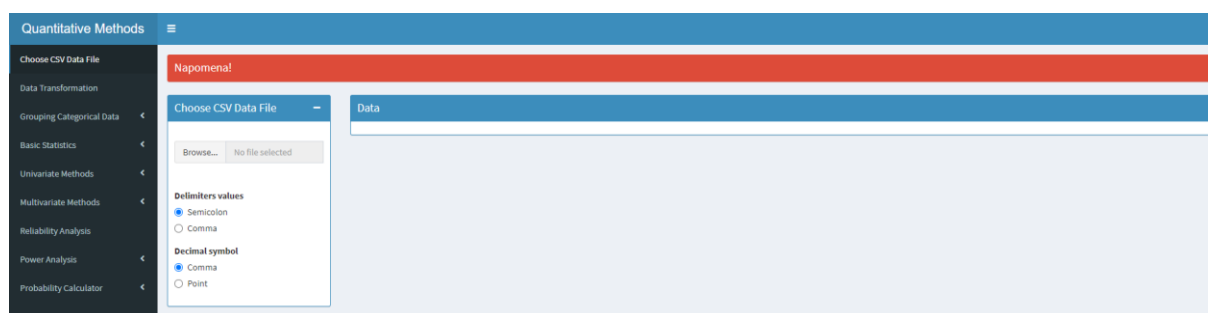
Klikom na opciju **Open File** padajućeg izbornika **File** otvorimo mapu u kojoj se nalazi datoteka **Quantitative Methods.R** te je odaberemo i kliknemo na gumb **Open**. Nakon što se datoteka učitava i prikaže u oknu **Source** potrebno je kliknuti na gumb **Run App**.

Klikom na maleni trokutić s vrhom okrenutim prema dolje koji se nalazi pored gumba **Run App** prikaže se izbornik u kojem možete odabrati web preglednik za prikazivanje rezultata. Opcija **Run in Window** omogućava korištenje aplikacije u web pregledniku RStudio. **Run in Viewer Pane** omogućava korištenje aplikacije u oknu **File** kartice **Viewer**, dok opcija **Run External** pokreće aplikaciju u nekom vanjskom web pregledniku (Google Chrome, Firefox i sl.).



CHOOSE CSV DATA FILE

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



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 (**C**omma **S**eparated **V**alues) formatu gdje se kao separator između podataka koristi točka-zarez (;), a za decimalne brojeve zarez (,).

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. Osnovne informacije o podacima nalaze se u poglavlju **Podaci** (str. 37-41).

U nastavku teksta opisane su metode za statističko-grafičku analizu podataka koje se nalaze u okviru R aplikacije **Quantitative Methods**.

DATA TRANSFORMATION

Data Transformation omogućava transformaciju originalnih podataka (podataka dobivenih mjerenjem) u:

- z-vrijednosti $z = (x_i - \bar{x})/s$, gdje je:
 - x_i originalna vrijednost entiteta i
 - \bar{x} aritmetička sredina
 - s standardna devijacija
- T-vrijednosti $T = 50 + z \cdot 10$
- L-vrijednosti (skala školskih ocjena) $L = 3 + z \cdot 0,83$

Quantitative Methods

Choose CSV Data File

Data Transformation

Grouping Categorical Data

Basic Statistics

Univariate Methods

Multivariate Methods

Reliability Analysis

Power Analysis

Probability Calculator

Napomena!

Variables

Select Variables:

☒ OHT

☒ OUZ

☒ NEB

☒ SKL

☒ TRB

☒ CUC

☒ SDM

☒ BML

☒ T20m

Copy

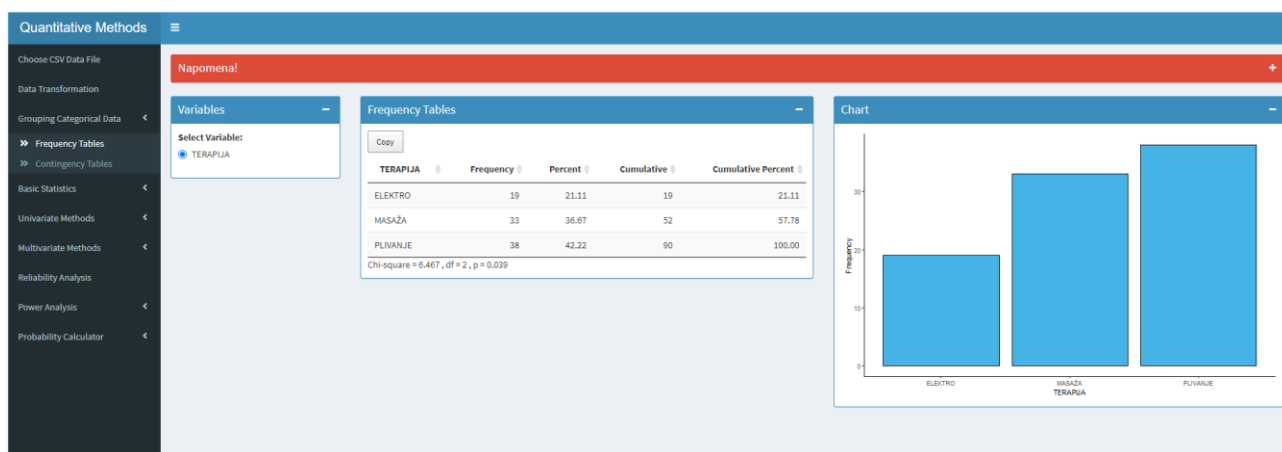
	Z - value	T - value	L - value (1-5)						
	OHT	OUZ	NEB	SKL	TRB	CUC	SDM	BML	T20m
Marko	2.72	2.75	-1.93	-1.09	-1.44	-1	-1.02	-0.22	2.08
Matte	-0.35	1.83	1.24	-0.52	-1.2	-0.7	-2.51	-2.15	2.48
Šime	0.13	0	-0.03	-0.41	0.74	-0.54	-0.13	0.21	0.54
Mile	-0.18	0.55	0.29	-0.29	-0.6	-0.41	-0.43	1.29	-0.27
Jure	0.42	0.36	-2.25	-0.75	-1.63	-0.92	-1.32	-0.11	-0.16
Ante	-0.27	-0.37	-0.03	0.39	-0.29	-0.73	-0.43	-0.32	0.37
Ive	-0.27	0.18	-1.93	-0.41	-0.48	-0.89	-0.73	-1.51	0.72
Stipe	1	1.46	-0.67	-1.54	-1.38	-0.4	-1.62	0.54	0.34
Tin	0.45	0.18	-0.35	0.39	-0.23	-0.48	-1.32	-1.72	1.63
Dino	-0.35	1.1	-0.98	-0.63	-0.9	-0.78	1.05	-0.11	0.62
Darko	0.59	1.46	1.55	-0.97	-0.84	-0.19	-0.13	1.07	-0.59
Stanko	-0.21	0.73	-0.67	-0.52	-0.72	-0.18	-1.62	0.96	0.71
Branko	0.02	0.36	-2.57	0.73	-0.17	-0.03	0.16	-0.32	0.34
Zarko	-0.64	-0.19	0.6	1.53	1.4	0.89	1.94	1.72	-1.83
Sanimir	0.65	1.46	-0.03	-1.54	-0.9	-0.92	0.16	0.64	-0.67

GROUPING CATEGORICAL DATA

Grupiranje podataka statistički je postupak razvrstavanja entiteta s istim oblikom obilježja u određen broj disjunktih podskupova. Rezultat grupiranja su frekvencije, odnosno broj entiteta u određenoj grupi (klasi, kategoriji, razredu). Kvalitativne (kategorijalne) varijable najčešće grupiramo prema jednom (jednodimenzionalno grupiranje) ili dva (dvodimenzionalno grupiranje) obilježja.

FREQUENCY TABLES

Odabirom opcije **Frequency Tables** prikazuje se tablica i grafikon s rezultatima jednodimenzionalnog grupiranja odabrane varijable te rezultati χ^2 -testa (**Chi-square test**) za utvrđivanje statističke značajnosti razlika između opaženih i teoretskih frekvencija dobivenih uniformnom distribucijom.



Tablica za svaku grupu entiteta prikazuje:

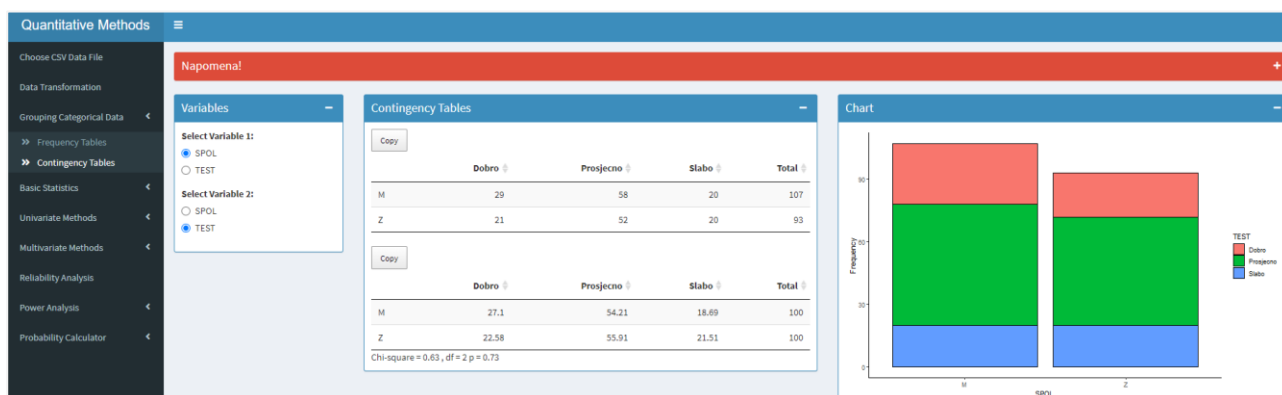
- **Frequency** – apsolutne frekvencije
- **Percent** – relativne postotne frekvencije
- **Cumulative** – kumulativne frekvencije
- **Cumulative Percent** – relativne postotne kumulativne frekvencije.

PRIMJER: Pitali smo 90 pacijenata koja im je terapija bila najučinkovitija. Elektroterapiju je odabralo 19, masažu 33, a plivanje 38 pacijenata. Zanima nas postoji li statistički značajna razlika između dobivene i slučajne (uniformne) distribucije odgovora kako bismo utvrdili preferiraju li pacijenti statistički značajno neku vrstu terapije u odnosu na druge. 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 % ($p = 0,039 < 0,05$).

CONTINGENCY TABLES

Odabirom opcije **Contingency Tables** prikazuje se tablica i grafikon s rezultatima dvodimenzionalnog grupiranja odabranih varijabli te rezultati χ^2 -testa (Chi-square test) za utvrđivanje statističke značajnosti razlika između opaženih i teoretskih frekvencija dviju kvalitativnih varijabli.

PRIMJER: Pomoću testa za procjenu znanja iz Kvantitativnih metoda testirano je 200 entiteta od čega je 107 muškog spola, a 93 ženskog spola. Nakon grupiranja podataka dobili smo sljedeće rezultate: 20 entiteta muškog spola ima slabe rezultate, 58 prosječne, a 29 dobre, dok 20 entiteta ženskog spola ima slabe rezultate, 52 prosječne, a 21 dobre. Postavlja se pitanje jesu li dobivene razlike između entiteta muškog i ženskog spola statistički značajne, odnosno, jesu li veće od razlika koje se mogu dobiti zbog slučajnog variranja entiteta u uzorku. Na temelju dobivenih rezultata χ^2 -testa ne možemo tvrditi da se muškarci i žene statistički značajno ne razlikuju u testu iz Kvantitativnih metoda uz pogrešku od 0,05 ($p = 0,73 > 0,05$).



BASIC STATISTICS

U okviru opcije **Basic Statistics** nalaze se statistički postupci koji omogućavaju opis podataka o istraživanom problemu putem deskriptivnih pokazatelja i grafikona, testiranje hipoteze o normalnosti empirijske distribucije podataka, procjenu aritmetičke sredine populacije te korelacijsku analizu.

DESCRPTIVE PARAMETERS

Deskriptivni pokazatelji koriste se za opis varijabli, a dijele se na mjere centralne tendencije ili središnje mjere, mjere varijabilnosti ili disperzije te mjere oblika distribucije.

	MEAN	MEDIAN	SD	CV	MIN	MAX	RANGE	SKEW	KURT
ONT	15.93	15.4	3.48	21.86	10.5	28.9	18.4	1.29	2.38
OUZ	3.9	3.9	0.55	13.99	2.9	5.4	2.5	0.42	-0.44
NEB	10.1	10	3.16	31.24	2	18	16	0.04	0.52
SKL	16.57	15	6.79	53.08	1	34	33	0.23	-0.76
TRB	39.85	37	16.51	41.42	13	75	62	0.31	-0.91
CUC	210.42	184.5	112.23	53.34	76	500	424	1.11	0.26
SDM	177.25	177.5	16.86	9.51	135	210	75	-0.12	-0.63
BML	530.17	520	93.13	17.57	330	740	410	-0.07	-0.34
T20m	3.88	3.85	0.6	15.52	2.78	5.38	2.6	0.38	-0.18

Mjere centralne tendencije ili središnje mjere predstavljaju jednu vrijednost koja bi trebala biti dobra zamjena za skup svih pojedinačnih vrijednosti, odnosno njihov najbolji reprezentant. U okviru ove aplikacije računaju se aritmetička sredina i medijan.

- **MEAN** - aritmetička sredina je najčešće korištena mjera centralne tendencije, a izračunava se kao omjer zbroja svih vrijednosti neke varijable i ukupnog broja entiteta.
- **MEDIAN** – medijan je vrijednost koja se nalazi na sredini uređenog niza podataka (uzlazno ili silazno sortiranog), odnosno vrijednost koja uređeni niz podataka dijeli na dva jednakobrojna dijela. Medijan za razliku od aritmetičke sredine, nije pod utjecajem ekstremno visokih ili niskih rezultata te je zato bolja mjera centralne tendencije za asimetrično distribuirane varijable.

Mjere varijabilnosti ili disperzije ukazuju na veličinu međusobnog razlikovanja rezultata entiteta u nekoj varijabli. U okviru ove aplikacije računaju se totalni raspon, standardna devijacija i koeficijent varijabilnosti.

- **RANGE** - totalni raspon je razlika između maksimalne (**MAX**) i minimalne (**MIN**) vrijednosti. Totalni raspon je nesigurna mjera varijabilnosti jer i jedan ekstremni rezultat može znatno povećati njegovu vrijednost. Povećanjem broja entiteta u

uzorku obično se povećava i totalni raspon jer se povećava vjerojatnost uključivanja entiteta s ekstremnim (maksimalnim i minimalnim) vrijednostima.

- **SD** - Standardna devijacija je drugi korijen iz varijance, odnosno drugi korijen iz prosječnog kvadratnog odstupanja rezultata entiteta od aritmetičke sredine.
- **CV** - Koeficijent varijabilnosti pokazuje koliki postotak vrijednosti aritmetičke sredine iznosi standardna devijacija, a koristi se za usporedbu varijabiliteta različitih varijabli.

Distribucija frekvencija je uređeni niz kvantitativnih vrijednosti s pripadajućim frekvencijama, a razlikujemo empirijske i teoretske distribucije. Empirijske distribucije su raspodjele eksperimentalno prikupljenih podataka, dok su teoretske distribucije matematičke funkcije koje omogućavaju utvrđivanje vjerojatnosti nekog slučajnog događaja u zadanim uvjetima. Oblik empirijske distribucije može se opisati mjerama asimetrije i izduženosti distribucije.

- **SKEW** - Skewness je koeficijent asimetrije distribucije. Ako je koeficijent asimetrije veći od nula, distribucija je pozitivno asimetrična. Ako je koeficijent asimetrije manji od nula, distribucija je negativno asimetrična.
- **KURT** - Kurtosis je stupanj spljoštenosti, odnosno izduženosti distribucije. Ako je kurtosis veći od nule, distribucija je izdužena ili leptokurtična. Ako je kurtosis manji od nule, distribucija je spljoštena ili platikurtična. Ako je kurtosis jednak nuli, distribucija je normalna ili Gaussova, odnosno mezokurtična.

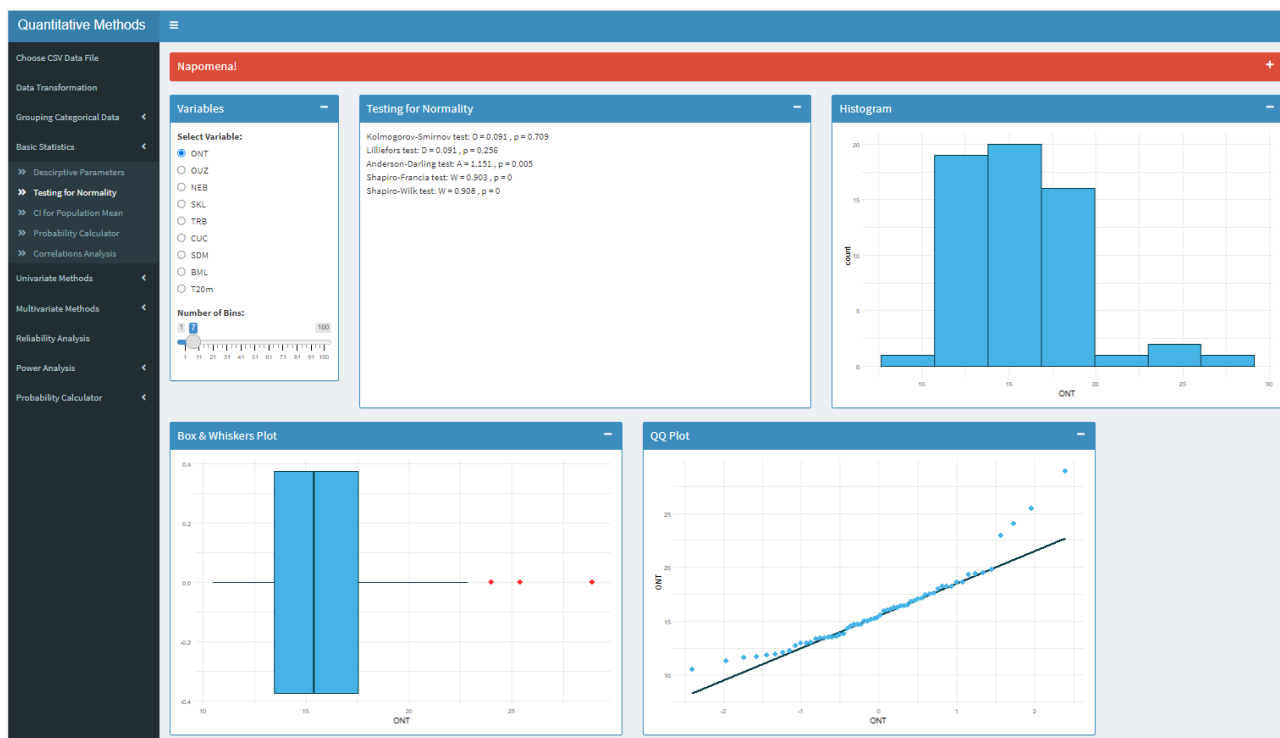
TESTING FOR NORMALITY

Normalitet distribucija varijabli, tj. sličnost empirijskih distribucija s normalnom ili Gaussovom distribucijom je uvjet za korištenje mnogih statističkih metoda. Veličina odstupanja empirijske distribucije od normalne ili Gaussove distribucije može se testirati statističkim postupcima kao što su:

- Kolmogorov-Smirnov test
- Lilliefors test
- Anderson-Darling test
- Shapiro-Francia test i
- Shapiro-Wilksov test.

Testovi za utvrđivanje normaliteta distribucije koriste se kako bi se utvrdilo jesu li podaci promatrane varijable nekog uzorka entiteta odabrani iz normalno distribuirane populacije. 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 najsnažaniji, 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-whiskerov dijagram, Q-Q dijagram), pokazatelji oblika distribucije (skewnes i kurtosis) te statistički testovi kako bi se što pouzdanije utvrdila razina i vrsta odstupanja neke empirijske od normalne distribucije.

- **Histogram** - histogram frekvencija je najčešće korišten način za brzu vizualizaciju distribucije frekvencija jedne varijable.

- **Box and Whiskers Plot** - kutijasti dijagram prikazuje odnose pet statističkih pokazatelja temeljem kojeg je moguće uočiti stupanj disperzije i asimetrije distribucije te outliere (vrijednosti koje ekstremno odstupaju od ostalih). Kutijasti dijagram sastoji se od pravokutnika čije stranice (na slici donja i gornja) prikazuje vrijednosti prvog (Q_1) i trećeg kvartila (Q_3) unutar kojih se nalazi 50 % svih rezultata ($IQR = Q_3 - Q_1$). Crta unutar pravokutnika označava median, dok se donja (W_1) i gornja (W_2) granica (engl. whisker) najčešće odredi na sljedeći način:

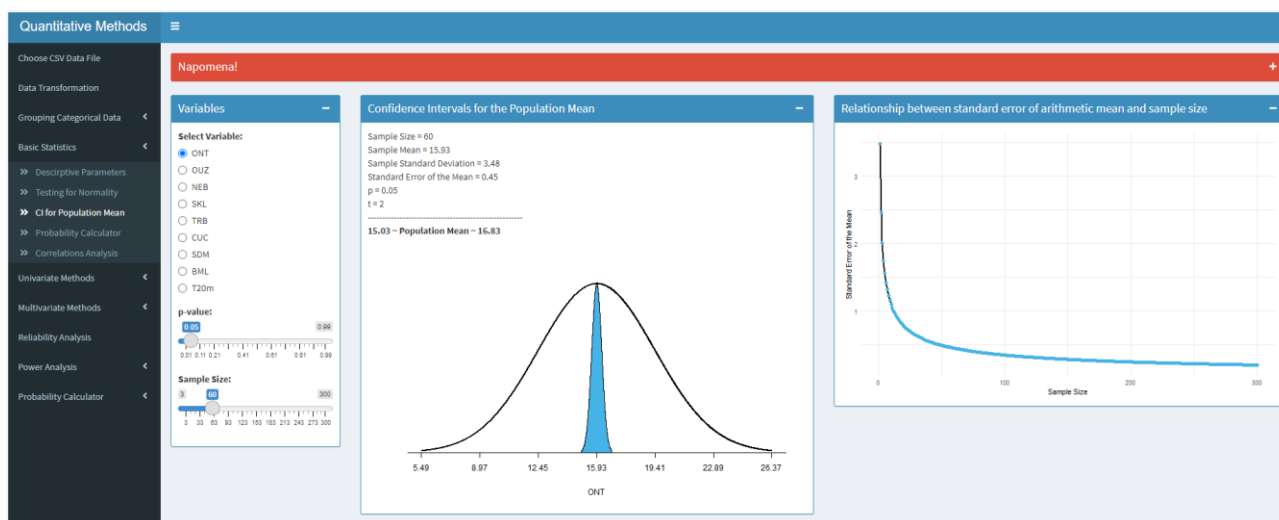
$$W_1 = \text{MIN ako je } \text{MIN} > Q_1 - 1,5 \cdot IQR, \text{ inače je } W_1 = Q_1 - 1,5 \cdot IQR$$

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

- **Q-Q Plot** - Q-Q dijagram (engl. Quantile vs Quantile 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 THE POPULATION MEAN

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**).

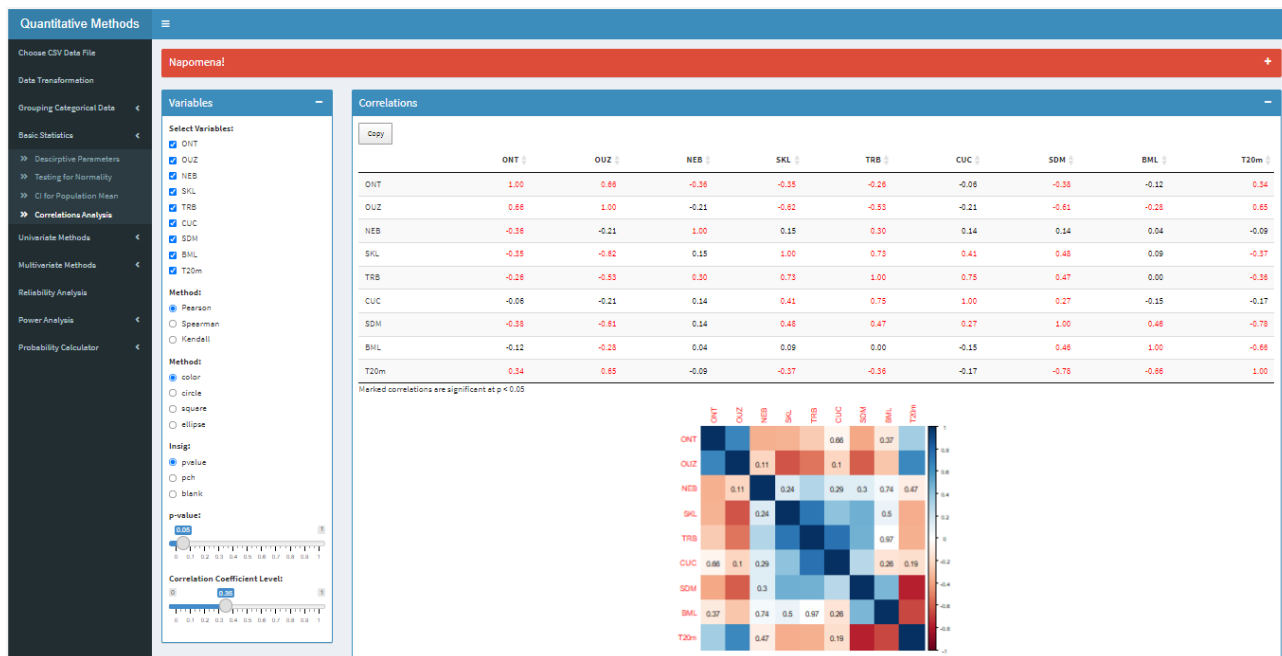


CORRELATION ANALYSIS

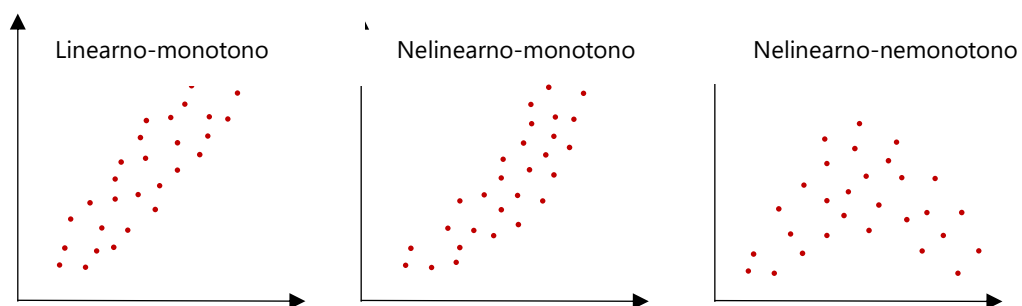
Korelacijskom analizom utvrđujemo povezanost između varijabli. Aplikacija omogućava izračunavanje Pearsonovog, Spearmanovog i Kendallovog koeficijenta korelacije, utvrđivanje njihove statističke značajnosti te grafičko prikazivanje korelativnih odnosa putem korelacijske mreže.

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

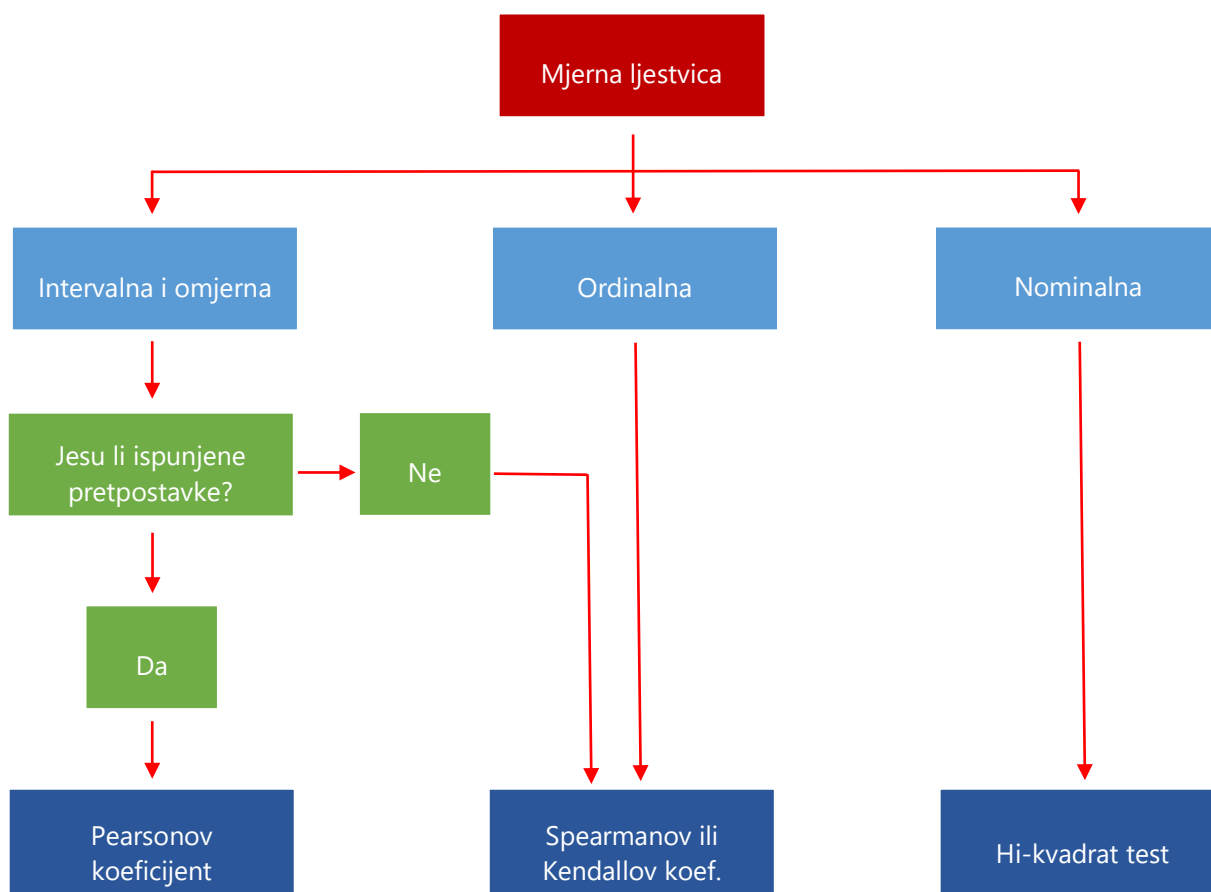
- Podaci su dobiveni pomoću intervalne ili omjerne mjerne ljestvice (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, taping rukom i sl.).
- Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ($n > 30$).
- Podaci u ne bi trebali imati značajne outliere, odnosno 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 mjernoj ljestvici, koje nisu normalno distribuirane te se ne nalaze u linearnom, već monotonom odnosu, tada se za utvrđivanje povezanosti između dviju varijabli koristi Spearmanov ili Kendallov koeficijent korelacije.



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



UNIVARIATE METHODS

U okviru opcije **Univariate Methods** glavnog izbornika nalaze se sljedeće univarijatne metode¹:

- t test za nezavisne uzorke
- t test za zavisne uzorke i
- jednofaktorska analiza varijance.

INDEPENDENT SAMPLES T-TEST

T-testom za nezavisne uzorke utvrđuje se statistička značajnost razlika aritmetičkih sredina dviju nezavisnih grupa entiteta (nezavisna varijabla) u nekoj kvantitativnoj varijabli (zavisna varijabla). 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 prozoru **Descriptive Parameters & Normality Test** prikazuje se tablica sa sljedećim pokazateljima za svaku grupu entiteta:

- **MEAN** – aritmetička sredina

¹Univarijatne metode koriste se za analizu podataka jedne varijable te testiranje odgovarajućih hipoteza koje se odnose na tu varijablu.

- **SD** – standardna devijacija
- **SEM** – standardna pogreška aritmetičke sredine
- **-CI95%** – donja granica intervala u kojem se s vjerojatnošću 95 % nalazi aritmetička sredina populacije
- **+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 zahtijeva normalnost distribucije podataka u zavisnoj varijabli.

Osim toga, aplikacija omogućava prikazivanje podataka putem histograma frekvencija (**Histogram**) i kutijastog dijagrama (**Box and Whiskers Plot**) za svaku grupu entiteta.

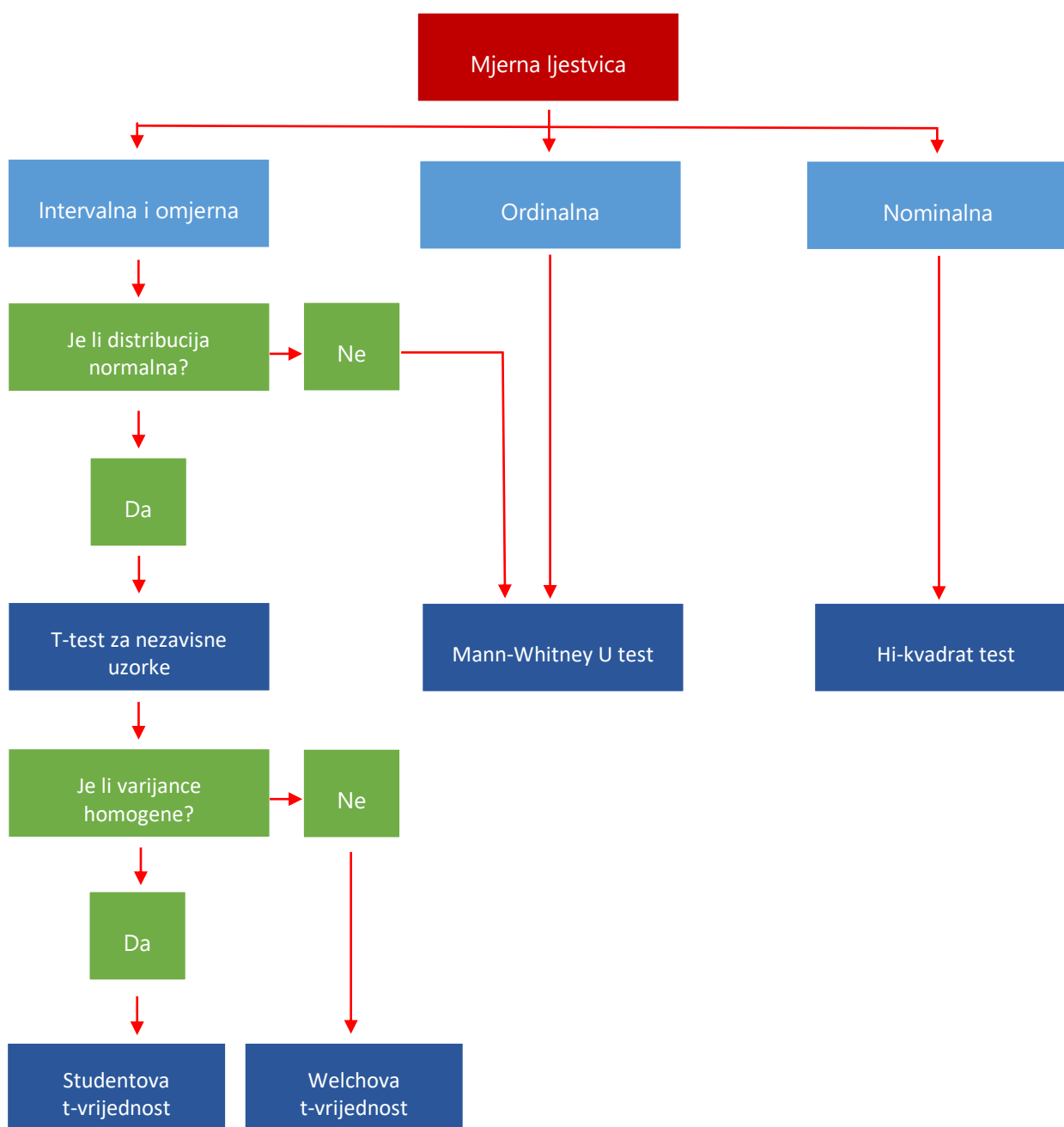
U prozoru **Independent Samples T-Test** prikazuju se sljedeći rezultati za svaku odabranu zavisnu varijablu:

- **Levene's F** – Levenova F vrijednost temeljem koje se utvrđuje homogenost varijanci, odnosno jesu li su varijance analiziranih grupa statistički značajno različite.
- **Levene'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).
- **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 %.

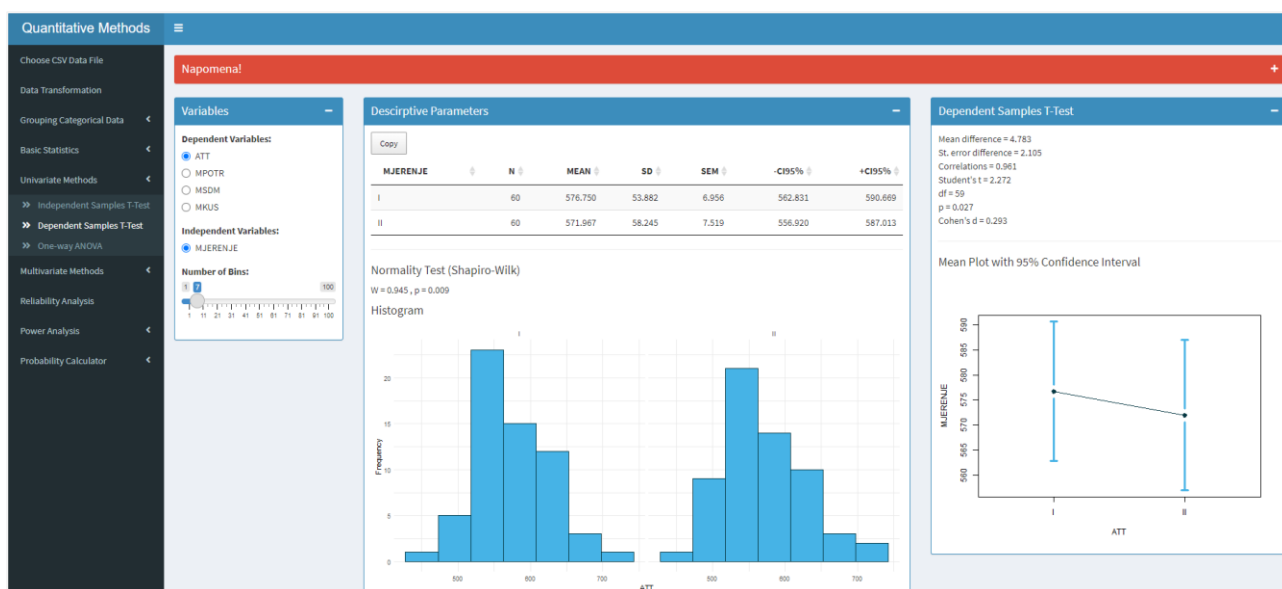
Testiranje razlika između dviju nezavisnih grupa, ovisno o vrsti i distribuciji varijabli.



DEPENDENT SAMPLES T-TEST

T-test za zavisne uzorke utvrđuje da li se aritmetičke sredine jedne grupe entiteta mjerene u dvije vremenske točke (nezavisna varijabla) statistički značajno razlikuju u nekoj kvantitativnoj varijabli (zavisna varijabla). 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 mjernoj skali (npr. tjelesna visina, tjelesna težina, skok u dalj s mjesta, tapping 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 outliers, odnosno ekstremno visoke ili niske rezultate.



U prozoru **Descriptive Parameters & Normality Test** 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.

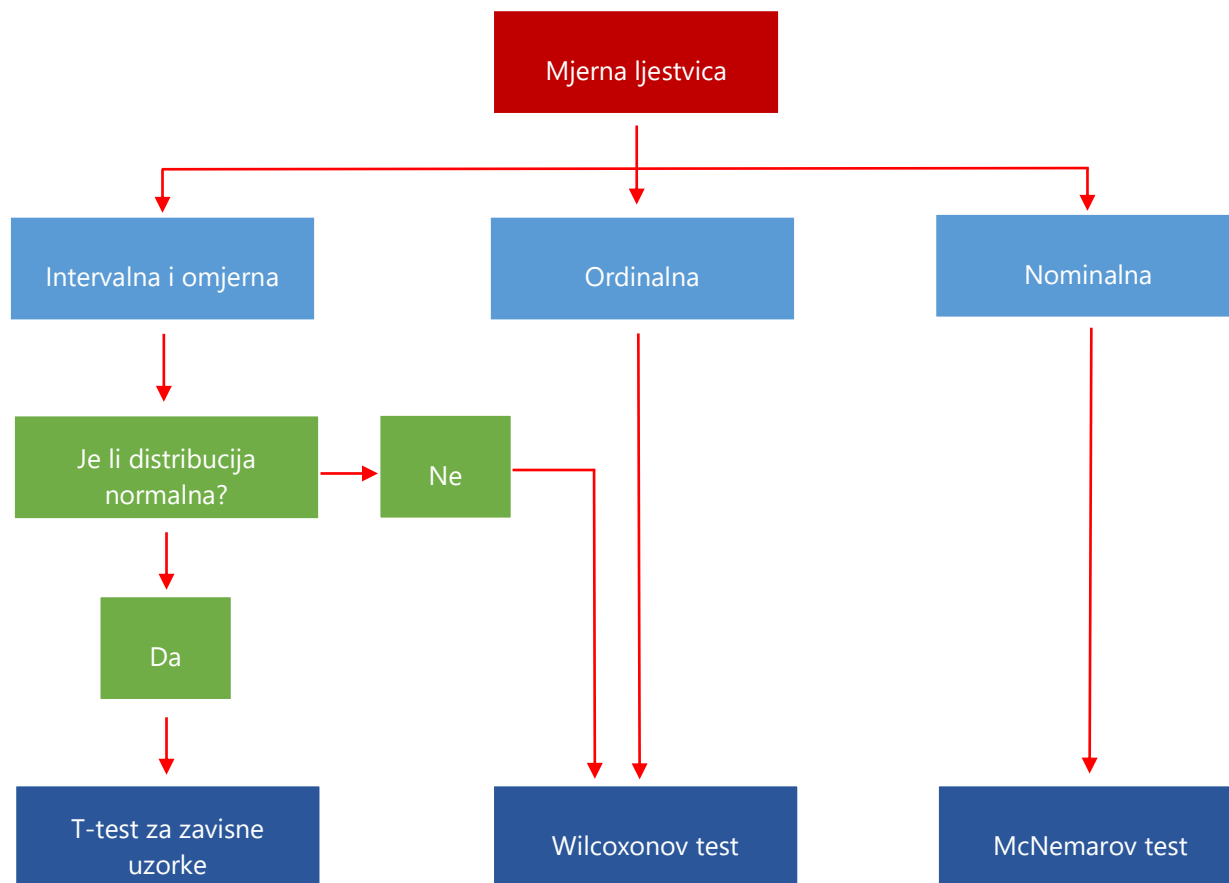
Ako je $p < 0,05$, zaključujemo da varijabla razlika prvog i drugog mjerenja statistički značajno odstupa od normalne ili Gaussove distribucije uz pogrešku p te da bi umjesto t-testa za zavisne uzorke bilo primjerenije koristiti Wilcoxonov test ekvivalentnih parova (**Wilcoxon Matched Pairs Test**).

Osim toga, aplikacija omogućava prikazivanje podataka putem histograma frekvencija (**Histogram**) i kutijastog dijagrama (**Box and Whiskers Plot**) za prvo i drugo mjerenje.

U prozoru **Dependent Samples T-Test** 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)
- **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 %.

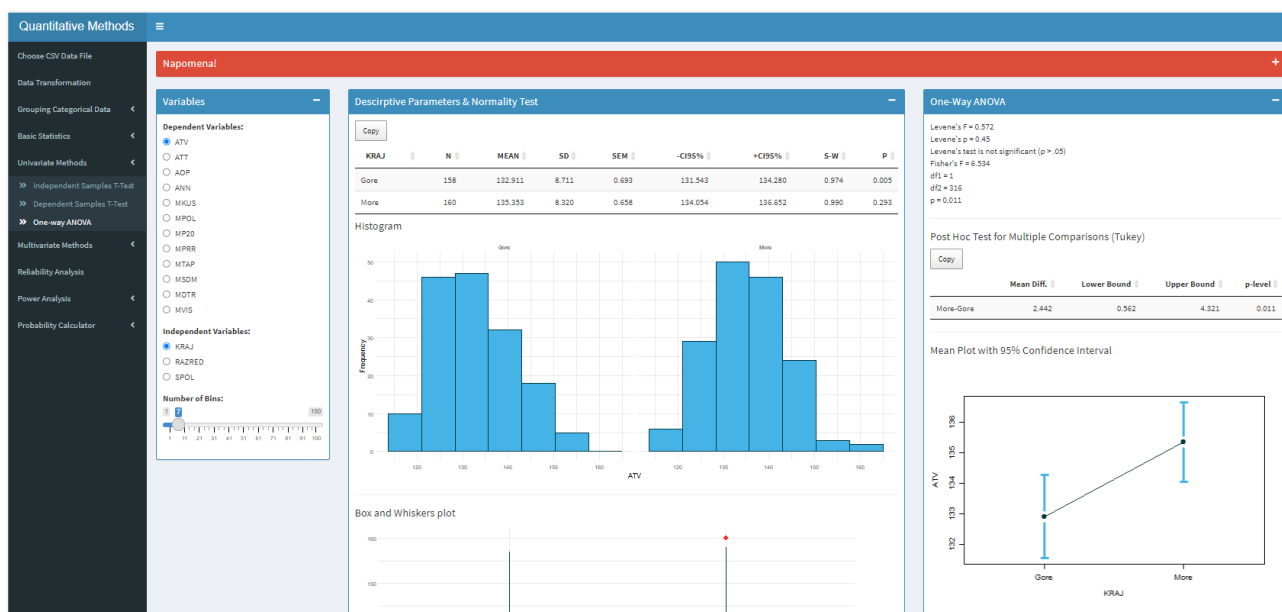
Testiranje razlika između dviju zavisnih grupa ovisno o vrsti i distribuciji varijabli.



ONE – WAY ANOVA

Jednofaktorska analiza varijance (**One - Way ANOVA**) utvrđuje da li se aritmetičke sredine dviju ili više nezavisnih grupa entiteta (nezavisna varijabla) statistički značajno razlikuju u nekoj kvantitativnoj varijabli (zavisna varijabla). 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 prozoru **Descriptive Parameters & Normality Test** 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
- **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 te da bi umjesto One - Way ANOVA bilo primjerenije koristiti **Kruskal - Wallisov test**. Osim toga, aplikacija omogućava prikazivanje podataka putem histograma frekvencija (**Histogram**) i kutijastog dijagrama (**Box and Whiskers Plot**) za analizirane grupe.

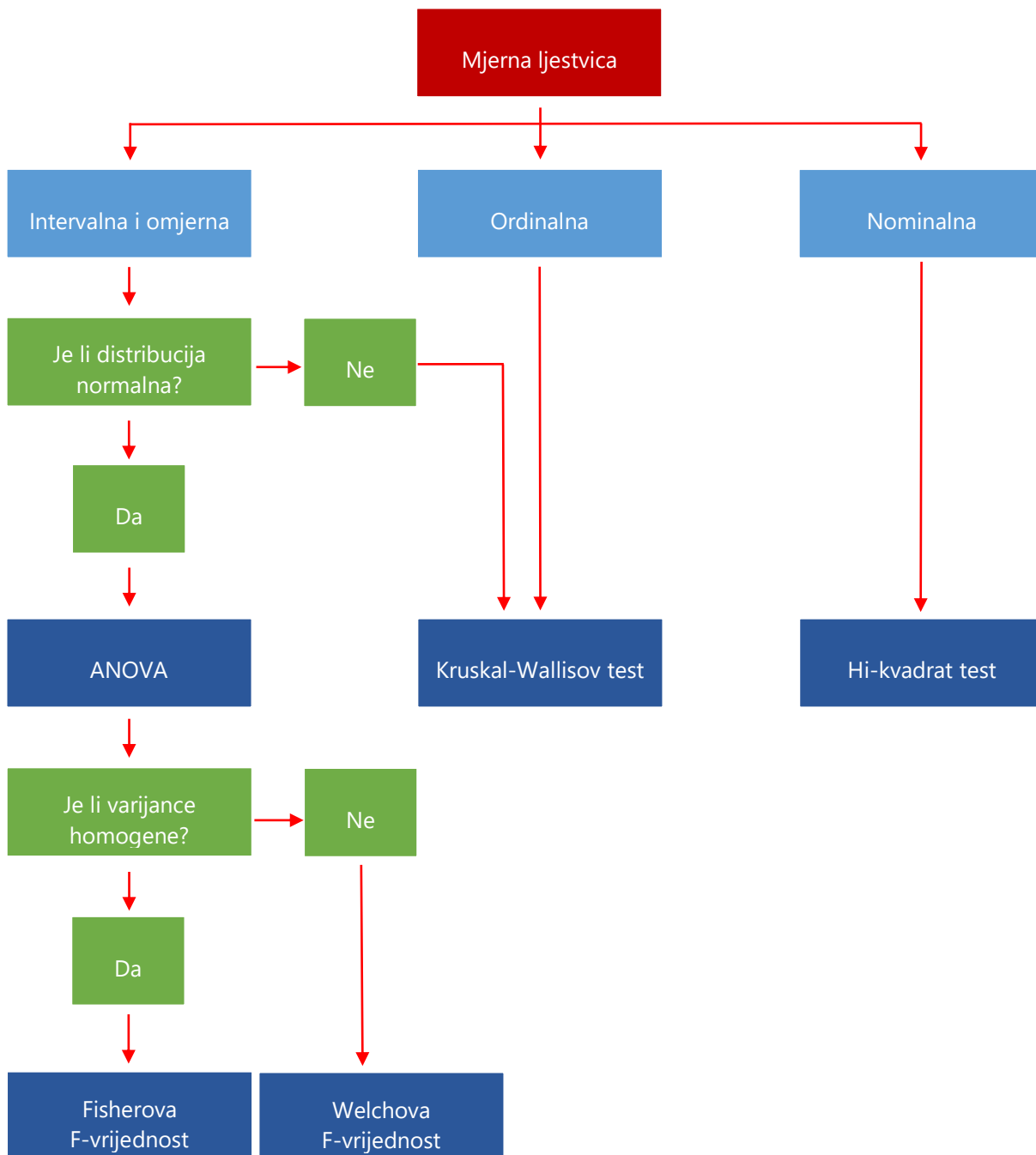
U prozoru **One-Way ANOVA** prikazuju se sljedeći rezultati za svaku odabranu zavisnu varijablu:

- **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 (Tukey)** – tablica rezultata Tukeyovog post hoc testa za višestruko uspoređivanje²

²Ako 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.).

- **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.



U okviru opcije **Multivariate Methods** glavnog izbornika nalaze sljedeće multivarijatne metode³:

- regresijska analiza
- faktorska analiza
- kanonička analiza i
- diskriminacijska analiza.

REGRESSION ANALYSIS

Regresijska analiza je matematičko-statistički postupak kojim se utvrđuje odgovarajuća funkcionalna veza (relacija) između jedne zavisne ili kriterijske varijable i jedne ili više nezavisnih ili prediktorskih varijabli. Aplikacija omogućava primjenu višestruke regresijske analize. 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, tapping rukom i sl.).
- Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ($n > 30$).
- 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.

³Multivarijatne metode koriste se za istovremenu analizu podataka dviju ili više varijabli uz uvažavanje njihovog međusobnog odnosa.

Quantitative Methods									
Napomena!									
Choose CSV Data File									
Data Transformation									
Grouping Categorical Data									
Basic Statistics									
Univariate Methods									
Multivariate Methods									
Regression Analysis									
Factor Analysis									
Canonical Analysis									
Discriminant Analysis									
Reliability Analysis									
Power Analysis									
Probability Calculator									
Variables									
Independent Variables:									
<input checked="" type="checkbox"/> VISI <input checked="" type="checkbox"/> TEZI <input checked="" type="checkbox"/> OBPO <input checked="" type="checkbox"/> NABIN <input checked="" type="checkbox"/> TAPI <input checked="" type="checkbox"/> SDAL <input checked="" type="checkbox"/> POTR <input checked="" type="checkbox"/> PRRA <input checked="" type="checkbox"/> IZVI <input type="checkbox"/> TR6M									
Dependent Variables:									
<input type="radio"/> VISI <input type="radio"/> TEZI <input type="radio"/> OBPO <input type="radio"/> NABIN <input type="radio"/> TAPI <input type="radio"/> POLI <input type="radio"/> SDAL <input type="radio"/> POTR <input type="radio"/> PRRA <input type="radio"/> IZVI <input checked="" type="radio"/> TR6M									
Regression Results									
RO = 0.83 ; RO2 = 0.681 ; SEE = 133.48 ; F-value = 18.972 p-value = 0 Autocorrelation = 0.081 ; Durbin-Watson = 1.83 ; p = 0.29									
Copy									
	B	SE(B)	Beta	Part_R	R	P	Tolerance	t	p
BO	1,665.784	486.731	0.000	0.000	0.000	0.000	0.000	3.422	0.001
VISI	-1.162	2.475	-0.034	-0.050	-0.056	0.002	0.695	0.470	0.640
TEZI	2.960	2.613	0.095	0.119	-0.203	-0.019	0.508	1.133	0.260
OBPO	-3.373	13.259	-0.019	-0.027	-0.260	0.005	0.670	0.254	0.800
NABIN	-27.319	4.579	-0.472	-0.535	-0.657	0.310	0.574	5.967	0.000
TAPI	1.695	3.279	0.037	0.055	0.164	0.006	0.687	0.517	0.607
POLI	-16.770	3.191	-0.455	-0.487	-0.648	0.295	0.479	5.256	0.000
SDAL	0.338	0.844	0.031	0.042	0.319	0.010	0.605	0.401	0.689
POTR	-1.615	2.720	-0.046	-0.063	0.400	-0.018	0.595	0.594	0.554
PRRA	0.933	1.559	0.049	0.063	0.493	0.024	0.533	0.598	0.551
IZVI	1.633	0.969	0.130	0.176	0.511	0.067	0.599	1.685	0.096
Graph of Partial Coefficients of Determinations (P)									
Observed, Predicted & Residual Values									
Copy									
	Observed Value	Predicted Value	Residual Values						
1	1,072.000	1,109.437	-37.437						
2	1,134.000	1,379.787	-245.787						
3	1,065.000	1,044.842	20.158						
4	1,048.000	975.372	72.628						
5	868.000	824.139	43.861						
6	1,026.000	1,104.812	-78.812						
7	1,128.000	922.374	205.626						
8	1,051.000	1,218.158	-167.158						
9	1,063.000	1,280.386	-217.386						
10	972.000	1,080.822	-108.822						
11	831.000	778.816	52.184						
12	895.000	872.483	22.517						
13	1,022.000	1,103.416	-81.416						
14	975.000	1,006.593	-31.593						
15	1,042.000	1,032.895	9.105						

U prozoru **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

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-value** – t-vrijednosti koje se izračunavaju pri testiranju statističke značajnosti regresijskih koeficijenata
- **p-value** – 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

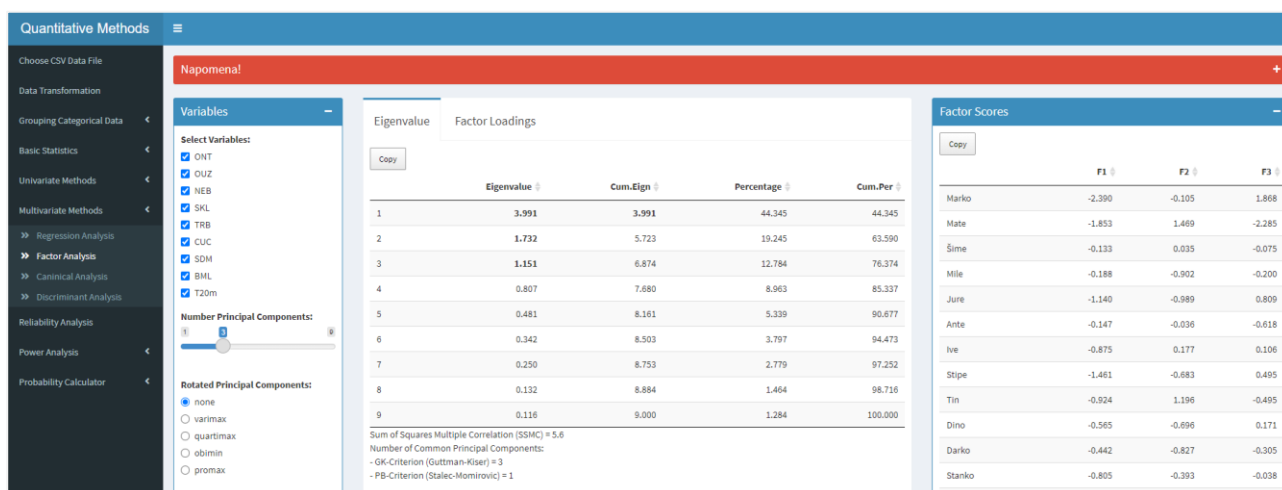
korelacija i korelacija (**Standardized Regression Coefficients, Partial Correlations and Correlations**).

U prozoru **Observed, Predicted & Residual Values** prikazuje se tablica s izmjerenim, prognoziranim i rezidualnim vrijednostima zavisne varijable.

FACTOR ANALYSIS

Faktorska analiza je zajedničko ime za više metoda kojima je cilj kondenzacija većeg broja manifestnih varijabli, među kojima postoji povezanost (korelacija), na manji broj latentnih dimenzija (faktora) koje su izvor te povezanosti. Aplikacija omogućava primjenu **metode glavnih komponenata** koju je predložio američki ekonomist i statističar Harold Hotelling, određivanje broja značajnih glavnih putem **GK, PB i Scree plot** te 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-Kiser)** – 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 **Factor Loadings** prikazuju se sljedeće tablice:

- **Pattern Matrix** – matrica sklopa, odnosno paralelnih projekcija manifestnih varijabli na faktore te Hoffmanov index kompleksiteta (**Complexity**) svake manifestne varijable⁴. 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

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

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

⁴Hofmann, R. J. (1978). Complexity and simplicity as objective indices descriptive of factor solutions. *Multivariate Behavioral Research*, 13(2), 247–250

- **Structure Matrix** – 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**)
- **Factor Correlation Matrix** – matrica korelacija između faktora
- **Graph Factor Loadings (Variables)** – grafikon strukture manifestnih varijabli
- **Graph Factor Loadings (Factors)** – grafikon strukture faktora.

U prozoru **Factor Scores** nalazi se tablica s rezultatima entiteta u faktorima.

CANONICAL ANALYSIS

Kanonička korelacijska analiza je metoda za utvrđivanje relacija između dvaju skupova varijabli. 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, taping 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.

Quantitative Methods

Choose CSV Data File

Data Transformation

Grouping Categorical Data

Basic Statistics

Univariate Methods

Multivariate Methods

Regression Analysis

Factor Analysis

Canonical Analysis

Discriminant Analysis

Reliability Analysis

Power Analysis

Probability Calculator

Napomena!

Variables

Select Variables:

☒ VISI

☒ TEZI

☒ OBPO

☒ NABN

☒ TAPI

☒ POLI

☒ SDAL

☒ POTR

☐ PRRA

☐ IZVI

☐ TRBM

Select Variables:

☐ VISI

☐ TEZI

☐ OBPO

☐ NABN

☐ TAPI

☐ POLI

☐ SDAL

☐ POTR

Canonical Analysis Results

Copy

CF	Rc	Chi-sq.	df	p-level
CF1	0.854	158.32	24	0
CF2	0.516	36.552	14	0.001
CF3	0.284	7.813	6	0.252

Factor Structure - 1. Set

Copy

	CF1	CF2	CF3
VISI	0.039	0.398	0.044
TEZI	0.196	0.848	-0.009
OBPO	0.326	0.151	-0.15
NABN	0.761	0.114	0.393
TAPI	-0.23	0.208	0.141

Canonical Scores - 1. Set

Copy

	CF1	CF2	CF3
1	0.814	0.342	-2.578
2	-1.101	-0.786	1.512
3	0.748	1.501	-0.142
4	1.257	1.275	-0.583
5	2.066	0.974	-0.321
6	0.457	0.793	-0.59
7	1.636	0	0.674
8	0.371	-1.05	-1.802
9	-0.257	-1.074	-1.76
10	0.607	0.127	0.161
11	2.353	0.753	-0.916

Canonical Scores - 2. Set

U prozoru **Canonical Analysis Results** prikazuju se sljedeći rezultati za odabrane varijable prvog i drugog skupa:

- **Rc** – koeficijenti kanoničkih korelacija

- **Chi-sq.** – vrijednosti Bartlettovog χ^2 – testa za testiranje statističke značajnosti kanoničkih korelacija
- **df** – stupnjevi slobode
- **p** – pogreške s kojima tvrdimo da su koeficijenti kanoničke korelacije statistički značajni
- **Factor Structure-First Set** – tablica faktorske strukture prvog skupa kanoničkih faktora, odnosno korelacija manifestnih varijabli prvog skupa s kanoničkim faktorima prvog skupa
- **Factor Structure-Second Set** – tablica faktorske strukture drugog skupa kanoničkih faktora, odnosno korelacija manifestnih varijabli drugog skupa s kanoničkim faktorima drugog skupa
- **Graph Factor Structure-First Set** – grafikon strukture kanoničkih faktora prvog skupa
- **Graph Factor Structure-Second Set** – grafikon strukture kanoničkih faktora drugog skupa.

U karticama **Canonical Scores - First Set** i **Canonical Scores - Second Set** nalaze se tablice s rezultatima entiteta u kanoničkim faktorima prvog i drugog skupa.

DISCRIMINANT ANALYSIS

Diskriminacijska analiza koristi se za utvrđivanje statističke značajnosti razlika više grupa entiteta (nezavisna varijabla) mjerenih u više varijabli (zavisne varijable), pri čemu se utvrđuje koliko se grupe međusobno razlikuju (centrirani grupa) i koliko pojedine varijable pridonose toj razlici (korelacije zavisnih varijabli s diskriminacijskim funkcijama). 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, tapping 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 outliere, odnosno ekstremno visoke ili niske rezultate.
- Varijable su u linearnom odnosu.
- Varijable imaju normalnu ili Gaussovu distribuciju.

U prozoru **Discriminant Analysis Results** prikazuje 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.

Quantitative Methods

Choose CSV Data File

Data Transformation

Grouping Categorical Data

Basic Statistics

Univariate Methods

Multivariate Methods

» Regression Analysis

» Factor Analysis

» Canonical Analysis

» Discriminant Analysis

Reliability Analysis

Power Analysis

Probability Calculator

Napomena!

Variables

Select Variables:

☒ ATV

☒ ATT

☒ AOP

☒ ANN

☒ MKUS

☒ MPOL

☒ MP20

☒ MPRR

☒ MTAP

☒ MSDM

☒ MDTR

☒ MVIS

Select Variables:

☐ KRAJ

☒ RAZRED

☐ SPOL

Discriminant Analysis Results

Copy

	Eigenvalue	Canonical R	Wilks' Lambda	aprox. F	df1	df2	p-level
1	2.134	0.825	0.255	14.636	36	895.97	0
2	0.17	0.381	0.799	3.28	22	608	0
3	0.07	0.256	0.935	2.135	10	305	0.022

Structure Discriminant Functions

Copy

	DF1	DF2	DF3
ATV	0.832	-0.07	0.167
ATT	0.642	0.083	0.078
AOP	0.562	-0.013	-0.037
ANN	0.168	0.156	-0.224

Discriminant Function Scores

Copy

	RAZRED	DF1	DF2	DF3
1	I	-0.221	-0.142	-0.063
2	I	-2.515	1.06	-0.03
3	I	-2.091	0.184	0.531
4	I	-2.649	0.346	0.499
5	I	-2.492	-0.59	0.164
6	I	-2.443	1.861	0.99
7	I	-1.087	0.777	-0.121
8	I	-1.478	0.933	0.443
9	I	-1.459	1.487	0.67
10	I	-1.761	1.5	0.956
11	I	-2.233	-0.87	0.519

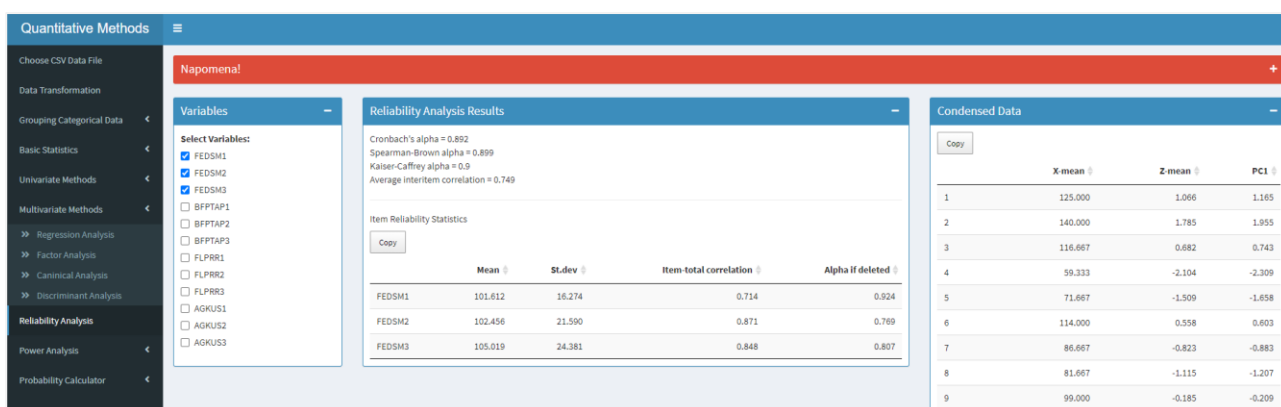
Osim toga, prikazuje se tablica i grafikon **Structure Discriminant Functions** s korelacijama zavisnih varijabli i diskriminacijskih funkcija te tablice **Group Centroids** s aritmetičkim sredinama analiziranih grupa u diskriminacijskim funkcijama i **Classifications Matrix** koje pokazuju broj i postotak ispravno i neispravno klasificiranih entiteta na temelju diskriminacijskih funkcija.

U prozoru **Discriminant Functions Scores** nalazi se tablica s rezultatima entiteta u diskriminacijskim funkcijama.

RELIABILITY ANALYSIS

Pouzdanost mjernog instrumenta je nezavisnost mjerenja od nesistematskih, tj. slučajnih pogrešaka. Aplikacija **Reliability Analysis** omogućava utvrđivanje pouzdanosti kompozitnih mjernih instrumenata **metodom interne konzistencije**. Kompozitni mjerni instrument sastoji se od tri ili više čestica, koje mogu biti:

- pitanja/zadaci (test tipa «papir-olovka»)
- uzastopna mjerenja (test tipa «aparatura za mjerenje» ili «primjena motoričkog zadatka»)
- mjeritelji/suci (test tipa «subjektivna procjena mjeritelja»).



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 α**
- jednostavnom linearnom kombinacijom standardiziranih rezultata (zbroy ili aritmetička sredina), onda je mjera pouzdanosti **Spearman-Brownova α**
- prvom glavnom komponentom, onda je mjera pouzdanosti **Kaiser-Caffreyeva α** .

Odabirom opcije **Reliability Analysis Results** 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

te 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.

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.

Osim toga, aplikacija **Quantitative Methods** nudi i analizu snage (**Power Analysis**) za t-test, univarijatnu analizu varijance i korelacijsku analizu te kalkulator za izračunavanje vjerojatnosti (**Probability Calculator**) za normalnu, t, F i hi-kvadrat distribuciju.

PODACI

GIM.CSV (N = 103, M = 12)

Kratko ime	Dugo ime	Vrsta varijable	Normalno skalirana
FEDSM1	Skok u dalj s mjesta 1. mjerenje	Kvantitativna	Da
FEDSM2	Skok u dalj s mjesta 2. mjerenje	Kvantitativna	Da
FEDSM3	Skok u dalj s mjesta 3. mjerenje	Kvantitativna	Da
BFPTAP1	Taping rukom 1. mjerenje	Kvantitativna	Da
BFPTAP2	Taping rukom 2. mjerenje	Kvantitativna	Da
BFPTAP3	Taping rukom 3. mjerenje	Kvantitativna	Da
FLPRR1	Pretklon raznožno 1. mjerenje	Kvantitativna	Da
FLPRR2	Pretklon raznožno 2. mjerenje	Kvantitativna	Da
FLPRR3	Pretklon raznožno 3. mjerenje	Kvantitativna	Da
AGKUS1	Koraci u stranu 1. mjerenje	Kvantitativna	Ne
AGKUS2	Koraci u stranu 2. mjerenje	Kvantitativna	Ne
AGKUS3	Koraci u stranu 3. mjerenje	Kvantitativna	Ne

JUDO3F.CSV (N = 60, M = 9)

Kratko ime	Dugo ime	Vrsta varijable	Normalno skalirana
ONT	Okretnost na tlu	Kvantitativna	Ne
OUZ	Okretnost u zraku	Kvantitativna	Ne
NEB	Neritmično bubnjanje	Kvantitativna	Da
SKL	Sklekovi	Kvantitativna	Da
TRB	Podizanje trupa	Kvantitativna	Da
CUC	Čučnjevi	Kvantitativna	Da
SDM	Skok u dalj s mjesta	Kvantitativna	Da
BML	Bacanje medicinske lopte iz ležanja	Kvantitativna	Da
T20m	Trčanje 20 metara	Kvantitativna	Ne

KM.CSV (N = 40, M = 3)

Kratko ime	Dugo ime	Modaliteti	Kod
SPOL	SPOL	Muškarac	m
		Žena	ž
KM_ISHOD	Ishod pismenog ispita iz predmeta Kvantitativne metode	Nije položio	0
		Položio	1
KM_OCJENA	Ocjena pismenog ispita iz predmeta Kvantitativne metode	Nedovoljan	1
		Dovoljan	2
		Dobar	3
		Vrlo dobar	4
		Odličan	5

KOS.CSV (N = 26, M = 8)

Kratko ime	Dugo ime	Vrsta varijable	Modalitet/Skaliranost	Kod
POZICIJA	Pozicija u igri	Nominalna	Bek	B
			Krilo	K
			Centar	C
MJERENJE	Mjerenje	Nominalna	Prvo mjerenje	I
			Drugo mjerenje	II
AVMT	Tjelesna masa	Kvantitativna	Normalno skalirana	
APMT	Postotak potkožnog masnog tkiva	Kvantitativna	Normalno skalirana	
FAEIJJOJO	JOJO test izdržljivosti	Kvantitativna	Normalno skalirana	
RVO2	Relativni primitak kisika	Kvantitativna	Normalno skalirana	
MAG20Y	Test agilnosti 20 jardi	Kvantitativna	Obrnuto skalirana	
S20m	Sprint 20 metara	Kvantitativna	Obrnuto skalirana	

MUHCS20M.CSV (N = 31, M = 3)

Kratko ime	Dugo ime	Vrsta varijable	Nor. Sk.
MUHCS20m1	Unilateralni horizontalni ciklički skokovi 1. mjerenje	Kvantitativna	Ne
MUHCS20m2	Unilateralni horizontalni ciklički skokovi 2. mjerenje	Kvantitativna	Ne
MUHCS20m3	Unilateralni horizontalni ciklički skokovi 3. mjerenje	Kvantitativna	Ne

POD.CSV (N=120, M=5)

Kratko ime	Dugo ime	Vrsta varijable	Modalitet (vrijednost)	Kod
MJERENJE	MJERENJE	Nominalna	Prvo	I
			Drugo	II
ATT	Tjelesna težina	Kvantitativna	Normalno skalirana	
MPOTR	Podizanje trupa	Kvantitativna	Normalno skalirana	
MSDM	Skok u dalj s mjesta	Kvantitativna	Normalno skalirana	
MKUS	Koraci u stranu	Kvantitativna	Obrnuto skalirana	

SKOLA.CSV (N = 100, M = 11)

Kratko ime	Dugo ime	Vrsta varijable	Normalno skalirana
VISI	Tjelesna visina	Kvantitativna	Da
TEZI	Tjelesna težina	Kvantitativna	Da
OBPO	Opseg podlaktice	Kvantitativna	Da
NABN	Nabor nadlaktice	Kvantitativna	Da
TAPI	Taping rukom	Kvantitativna	Da
POLI	Poligon natraške	Kvantitativna	Ne
SDAL	Skok u dalj s mjesta	Kvantitativna	Da
POTR	Podizanje trupa	Kvantitativna	Da
PRRA	Pretklon raznožno	Kvantitativna	Da
IZVI	Izdržaj u visu	Kvantitativna	Da
TR6M	Trčanje 6 minuta	Kvantitativna	Da

TEST.CSV (N = 200, M = 2)

Kratko ime varijable	Dugo ime varijable	Modalitet
SPOL	Spol	Muško
		Žensko
TEST	Test za procjenu eksplozivne jakosti	Slabo
		Prosječno
		Dobro

UCENICI.CSV (N = 318, M = 16)

Kratko ime	Dugo ime	Vrsta varijable	Modalitet/Skaliranost	Kod
KRAJ	Gorski i primorski	Nominalna	Gorski kraj	Gore
			Primorski kraj	More
RAZRED	Razred osnovne škole	Ordinalna	Prvi razred	I
			Drugi razred	II
			Treći razred	III
			Četvrti razred	IV
SPOL	Spol	Nominalna	Muškarci	M
			Žene	Z
ATV	Tjelesna visina	Kvantitativna	Normalno skalirana	
ATT	Tjelesna težina	Kvantitativna	Normalno skalirana	
AOP	Opseg podlaktice	Kvantitativna	Normalno skalirana	
ANN	Kožni nabor nadlaktice	Kvantitativna	Normalno skalirana	
MKUS	Koraci u stranu	Kvantitativna	Obrnuto skalirana	
MPOL	Poligon natraške	Kvantitativna	Obrnuto skalirana	
MP20	Test ravnoteže	Kvantitativna	Normalno skalirana	
MPRR	Pretklon raznožno	Kvantitativna	Normalno skalirana	
MTAP	Taping rukom	Kvantitativna	Normalno skalirana	
MSDM	Skok u dalj s mjesta	Kvantitativna	Normalno skalirana	
MDTR	Podizanje trupa	Kvantitativna	Normalno skalirana	
MVIS	Izdržaj u visu	Kvantitativna	Normalno skalirana	

VZA.CSV (N = 60, M = 12)

Kratko ime	Dugo ime	Vrsta varijable	Modalitet/Skaliranost	Kod
ISHOD	Ishod utakmice	Kvalitativna	Pobjeda	POB
			Poraz	POR
BIBL	Blok na igraču bez lopte	Kvantitativna	Normalno	
BOTK	Bez otkrivanja	Kvantitativna	Normalno	
LEPK	Igra leđima prema košu	Kvantitativna	Normalno	
LIPK	Igra licem prema košu	Kvantitativna	Normalno	
OTK	Otkrivanje	Kvantitativna	Normalno	
PP	Pick&pop	Kvantitativna	Normalno	
PR	Pick&roll	Kvantitativna	Normalno	
PUB	Šut nakon skoka u napadu	Kvantitativna	Normalno	
SLB	Slobodna bacanja	Kvantitativna	Normalno	
UR	Uručivanje	Kvantitativna	Normalno	
OA	Ostale akcije	Kvantitativna	Normalno	