

DRAŽAN DIZDAR

DARKO KATOVIĆ



Version: 2.0

ZAGREB, 2023.

Sadržaj

Web aplikacija R-STAT.....	2
Home.....	4
Choose Data File	7
View Data.....	8
Data Transformation.....	9
Grouping Categorical Data.....	11
Chi-Square Goodness of Fit Test	11
Chi-Square Test of Independence.....	15
McNemar's Test	18
Grouping Continuous Data	20
Basics Statistics	23
Descriptive Parameters	23
Testing for Normality	27
CI for Population Mean	30
Correlations Analysis	32
Univariate Methods	35
Independent Samples T-Test.....	35
Dependent Samples T-Test	39
One Way ANOVA.....	43
Repeated Measures ANOVA	47
Multivariate Methods	49
Regression Analysis	49
Factor Analysis.....	53
Canonical Analysis	58
Discriminant Analysis	61
Reliability Analysis.....	65
Probability Calculator.....	68
Normal Distribution.....	68
T – Distribution.....	69
F – Distribution.....	70
Chi Squere – Distribution	71
Settings Graph.....	72

Web aplikacija R-STAT

R-STAT je web aplikacije za statističko-grafičku analizu podataka koja se besplatno može koristiti na sljedećoj web adresi:

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

R-STAT je kreiran korištenjem integriranog razvojnog okruženja **RStudio** (engl. Integrated Development Environment, IDE) za programski jezik **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**.

R je programski jezik i softversko okruženje za statističko-grafičku analizu podataka koje se može besplatno 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, multivariatne metode te specijalizirane pakete za genetiku, psihologiju, ekonometriju, itd.

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.

R se može nadograđivati paketima (engl. packages) koji predstavljaju skup funkcija za jednostavnije izvršavanje određenih zadatka, što je u velikoj mjeri pridonijelo njegovoj popularizaciji. Osnovni paketi (**base**, **stats**, **datasets**, **graphics**, **grDevices**, **methods** i **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("shiny")
library(shiny)
```

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/>

Kôd R-STAT aplikacije možete preuzeti na poveznici:

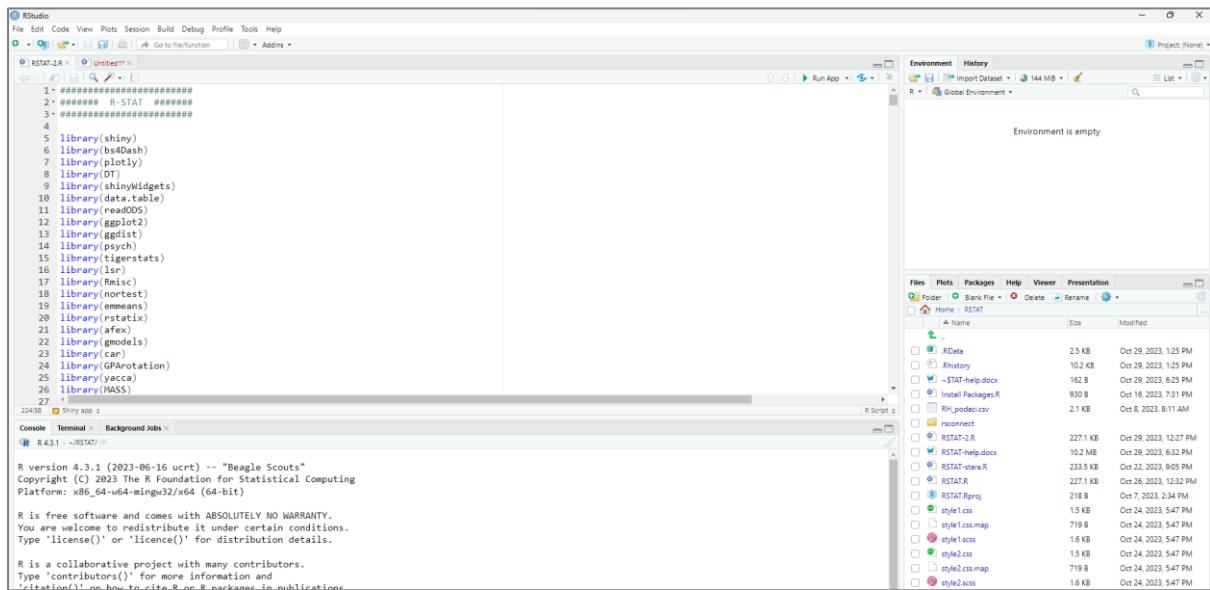
<https://github.com/Quantitative-Methods/R-applications>

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

Prije korištenja aplikacije **R-STAT.R** potrebno je instalirati pakete koji se nalaze u datoteci **Install Packages.R**. Nakon što se datoteka učita 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 **R-STAT.R** spremna je za korištenje.

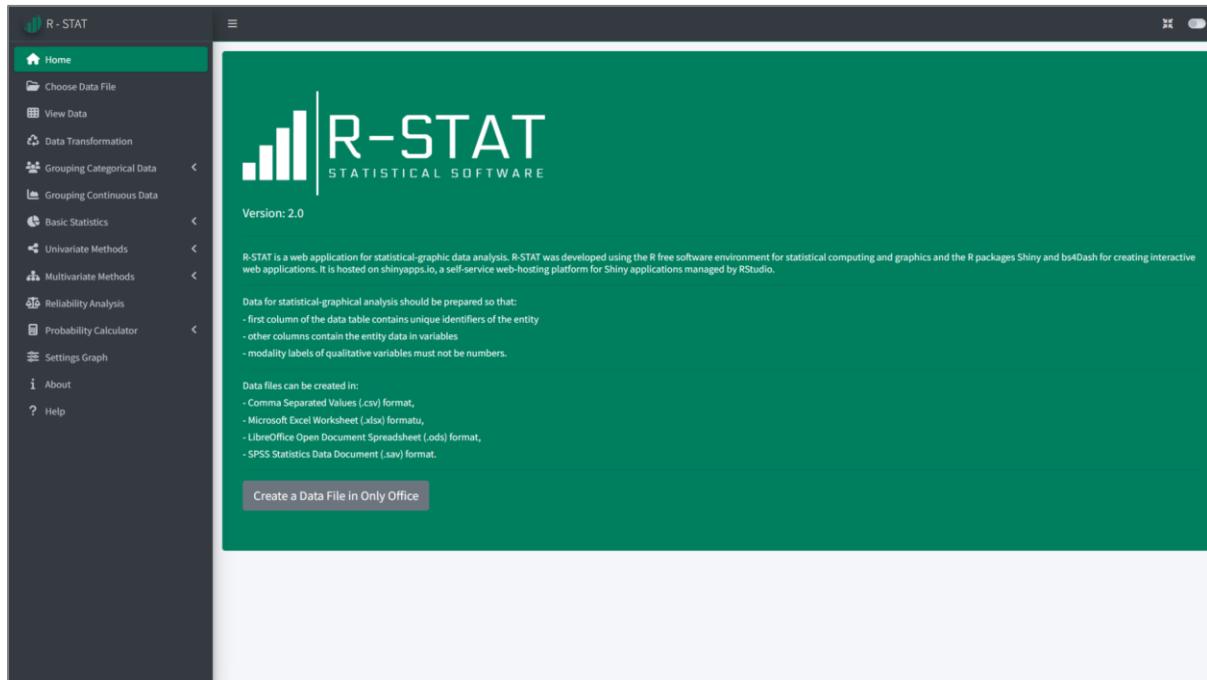
Klikom na opciju **Open File** padajućeg izbornika **File** otvorimo mapu u kojoj se nalazi datoteka **R-STAT.R** te je odaberemo i kliknemo na gumb **Open**. Nakon što se datoteka učita i prikaže u oknu **Source** potrebno je kliknuti na gumb **Run App**.

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 RStudioa. **Run in Viewer Pane** omogućava korištenje aplikacije u oknu **File** kartice **Viewer**, dok opcija **Run External** pokreće aplikaciju u nekom vanjskom u web pregledniku (Google Chrome, Firefox, Microsoft Edge i sl.).



Home

Nakon pokretanja web aplikacije R-STAT prikazuje se opcija **Home** glavnog izbornika (slika 1).



Slika 1: Home

U glavnom izborniku moguće je odabratи sljedeće opcije:

- Choose Data File
- View Data
- Data Transformation
- Grouping Categorical Data
 - Chi-Square Goodness of Fit Test
 - Chi-Square Test of Independence
 - McNemar's Test
- Grouping Continuous Data
- Basic Statistics
 - Descriptive Parameters
 - Testing for Normality
 - CI for Population Mean
 - Correlations Analysis
- Univariate Methods
 - Independent Samples T-Test
 - Dependent Samples T-Test
 - One-way ANOVA
 - Repeated Measures ANOVA
- Multivariate Methods

- Regression Analysis (simple)
- Regression Analysis (multiple)
- Factor Analysis
- Canonical Analysis
- Discriminant Analysis
- Cluster Analysis
- Reliability Analysis
- Probability Calculator
 - Normal Distribution
 - T – Distribution
 - F – Distribution
 - Chi Square – Distribution
- Settings Graph
- About
- Help

Prije odabira određene statističko-grafičke metode potrebno je odabrati datoteku u kojoj se nalaze **podaci**¹ koje želimo analizirati. Podaci za statističko-grafičku analizu trebaju biti pripremljeni na sljedeći način (tablica 1):

- prvi stupac tablice s podacima sadrži jedinstvene oznake **entiteta**², odnosno entiteti ne smiju imati istu oznaku
- ostali stupci tablice s podacima sadrže podatke entiteta u **varijablama**³
- oznake (kodovi) modaliteta kvalitativnih varijabli ne smiju biti brojevi
- datoteke s podacima mogu biti kreirane u:
 - ✓ **Comma Separated Values (.csv)**,
 - ✓ **Microsoft Excel Worksheet (.xlsx)**,
 - ✓ **LibreOffice Open Document Spreadsheet (.ods)** i
 - ✓ **SPSS Statistics Data Document (.sav)** formatu.

¹ Podatak ili informacija je kvantitativna ili kvalitativna vrijednost kojom je opisano određeno obilježje nekog objekta, stvari, osobe, pojave, procesa..., odnosno, entiteta.

² Entitet je jedinka nekog skupa osoba, stvari, pojave i sl. koja je opisana određenim obilježjima, odnosno varijablama.

³ Varijabla je svojstvo, obilježje, osobina, sposobnost, znanje, itd. po kojem se entiteti međusobno razlikuju.

ENTITETI	VISI	TEZI	OBPO	NABN	TAPI	POLI	SDAL	POTR
1	176	60	24	8	35	25	160	21
2	160	48	21	7	44	13	205	44
3	170	66	22	12	36	23	160	28
4	165	66	23	12	42	28	155	26
5	166	65	22	16	33	29	155	21
6	165	60	21	10	35	22	160	28
7	160	56	24	18	38	21	165	30
8	153	52	23	7	35	22	170	38

Tablica 1: Podaci u 8 entiteta opisanih s 8 varijabli.

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.

VAŽNO!

Sve tablice koje se kreiraju u web aplikaciji R-STAT koriste '.' kao decimalnu oznaku.

Tablice je moguće je kopirati u međuspremnik (**clipboard**) klikom na gumb **Copy** te ih potom premjestiti (**Paste**) u neku drugu aplikaciju (npr. **MS EXCEL**, **LIBREOFFICE CALC**, **ONLYOFFICE**). Kako bi kopiranje podataka u navedene aplikacije bilo valjano potrebno je da je i u njima namjestiti '.' kao decimalnu oznaku.

Choose Data File

Odabirom opcije **Choose Data File** te klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci za statističku analizu. Nakon odabira datoteke s podacima te klikom na gumb **Import data**, u prozoru **Select Entities** prikazuju se varijable temeljem kojih je moguće izvršiti selekciju entiteta. Selekcija entiteta za kvantitativne varijable vrši se pomoću klizača kojima se određuje raspon podataka u kojem se selekcionirani entiteti nalaze, dok se za kvalitativne varijable vrši odabirom određenog modaliteta. U prozoru **Update & select variables** moguće je odabrati varijable te promjeniti njihov naziv (**Name**) i tip (**New class**).

The screenshot shows the R-STAT application interface. On the left, there is a sidebar with various statistical analysis options like Home, Choose Data File, View Data, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods, Multivariate Methods, Reliability Analysis, Probability Calculator, Settings Graph, About, and Help. The main area is divided into two sections: 'Choose Data File' and 'Update & Select Variables'. In the 'Choose Data File' section, a file 'JUDO03.ods' is selected, and a message says 'Data ready to be imported! data has 60 obs. of 11 variables. First five rows are shown below.' Below this is a 'Select Entities' section with a table showing observations for groups like ONT, OUZ, NEB, SKL, TRB, and CUC. In the 'Update & Select Variables' section, there is a table titled 'Update & select variables' showing variables like Groups, ONT, OUZ, NEB, SKL, TRB, and CUC with their current class (character or numeric) and new class (character or numeric). At the bottom, there is an 'Edit Data' section with a table showing data for individuals Marko, Mate, and Šime across various variables.

Slika 2: Choose Data File

Ispod gumba **Apply changes** nalazi se sekcija **Edit Data** koja omogućava ispravljanje podataka (**Update**) i brisanje svih podataka određenog retka (**Delete**) te dodavanje novog retka (**Add row**). Klikom na gumb **Excel** ili **CSV** vrši se preuzimanje datoteke u .xlsx ili .csv formatu.

VAŽNO!

Bez obzira jesu li se, ili nisu, vršile bilo kakve promjene na odabranoj datoteci potrebno je kliknuti gumb **Apply changes** kako bi datoteka postala aktivna.

View Data

Odabirom opcije **View Data** prikazuje se trenutno aktivna datoteka s podacima.

Groups	ONT	OUZ	NEB	SKL	TRB	CUC	SDM	BML	T20m
Marko	I	25.4	5.4	4	7	16	98	160	510
Mate	I	14.7	4.9	14	12	20	132	135	330
Sime	I	16.4	3.9	10	13	52	150	175	550
Mile	I	15.3	4.2	11	14	30	164	170	650
Jure	I	17.4	4.1	3	10	13	107	155	520
Ante	I	15	3.7	10	20	35	128	170	500
Ivo	I	15	4	4	13	32	110	165	390
Stipe	I	19.4	4.7	8	3	17	165	150	580
Tin	I	17.5	4	9	20	36	156	155	370
Dino	I	14.7	4.5	7	11	25	123	195	520
Darko	I	18	4.7	15	8	26	189	175	630
Stanko	I	15.2	4.3	8	12	28	190	150	620
Branko	I	16	4.1	2	23	37	207	180	500
Zarko	II	13.7	3.8	12	30	63	310	210	690

Slika 3: View Data

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

Data Transformation

Za korištenje opcije **Data Transformation** glavnog izbornika odabrana datoteka podataka mora imati najmanje jednu kvantitativnu varijablu. Primjerice, u datoteci JUDO3F.csv nalaze se podaci 60 entiteta opisanih s 9 kvantitativnih varijabli (slika 4).

The screenshot shows the R-STAT software interface with the 'Data Transformation' tab selected. On the left, there's a sidebar with various statistical methods like Grouping Categorical Data, Basic Statistics, and Multivariate Methods. The main area has two sections: 'Normally Scaled:' and 'Inversely Scaled:', each listing the same nine variables: ONT, OÜZ, NEB, SKL, TRB, CUC, SDM, BML, and T20m. The 'Normally Scaled:' section has checkboxes checked for ONT, OÜZ, NEB, SKL, TRB, CUC, SDM, BML, and T20m. The 'Inversely Scaled:' section has checkboxes unchecked for all variables. Below these sections is a table with 60 rows, each representing an entity (Marko, Mate, Sime, Mile, Jure, Ante, Ive, Stipe, Tin, Dino, Darko, Stanko, Branko, Zarko, Sanimir, Toni, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38) and 10 columns representing the scaled values for each variable. A green bar at the bottom indicates the mean Z-value for each entity.

	ONT	OÜZ	NEB	SKL	TRB	CUC	SDM	BML	T20m	MEAN_Z
Marko	2.72	2.746	-1.933	-1.088	-1.445	-1.002	-1.023	-0.217	2.082	0.093
Mate	-0.353	1.829	1.236	-0.519	-1.202	-0.699	-2.506	-2.149	2.48	-0.209
Sime	0.135	-0.093	-0.032	-0.406	0.736	-0.538	-0.133	0.213	0.54	0.057
Mile	-0.181	0.547	0.285	-0.292	-0.597	-0.414	-0.43	1.287	-0.273	-0.097
Jure	0.422	0.363	-2.25	-0.747	-1.627	-0.921	-1.32	-0.109	-0.157	-0.705
Ante	-0.267	-0.37	-0.032	0.39	-0.294	-0.734	-0.43	-0.324	0.374	-0.187
Ive	-0.267	0.18	-1.933	-0.406	-0.476	-0.895	-0.727	-1.505	0.722	-0.59
Stipe	0.997	1.463	-0.666	-1.543	-1.384	-0.405	-1.616	0.535	0.341	-0.251
Tin	0.451	0.18	-0.349	0.39	-0.233	-0.485	-1.32	-1.72	1.634	-0.161
Dino	-0.353	1.096	-0.982	-0.633	-0.9	-0.779	1.053	-0.109	0.623	-0.109
Darko	0.594	1.463	1.553	-0.974	-0.839	-0.191	-0.133	1.072	-0.588	0.217
Stanko	-0.21	0.73	-0.666	-0.519	-0.718	-0.182	-1.616	0.965	0.705	-0.168
Branko	0.02	0.363	-2.567	0.732	-0.173	-0.03	0.163	-0.324	0.341	-0.164
Zarko	-0.64	-0.186	0.602	1.528	1.402	0.887	1.942	1.716	-1.832	0.602
Sanimir	0.652	1.463	-0.032	-1.543	-0.9	-0.921	0.163	0.642	-0.671	-0.127
Toni	-0.009	0.18	-0.032	-0.406	-0.354	-0.093	1.646	0.106	-1.418	-0.042

Slika 4: Data Transformation → Z-valute

Nakon učitavanja datoteke podataka odabirom opcije **Data Transformation** glavnog izbornika za odabrane normalno skalirane (**Normally Scaled**) ili obrnuto skalirane varijable (**Inversely Scaled**) u prikazuju se tablice s podacima (slika 4) transformiranih u:

- Z – vrijednosti (kartica **Z – value**)

$$Z_i = \frac{x_i - \bar{x}}{s}$$

gdje je

- ✓ z_i standardizirani rezultat entiteta i
- ✓ x_i originalna vrijednost ispitanika i
- ✓ \bar{x} aritmetička sredina
- ✓ s standardna devijacija.

- T – vrijednosti (kartica **T – value**)

$$T_i = 50 + 10 \cdot z_i$$

gdje je

- ✓ T_i transformirani rezultat entiteta i na skali čija je aritmetička sredina 50, a standardna devijacija 10
- ✓ z_i standardizirani rezultat entiteta i

- L – vrijednosti (kartica **L – value**)

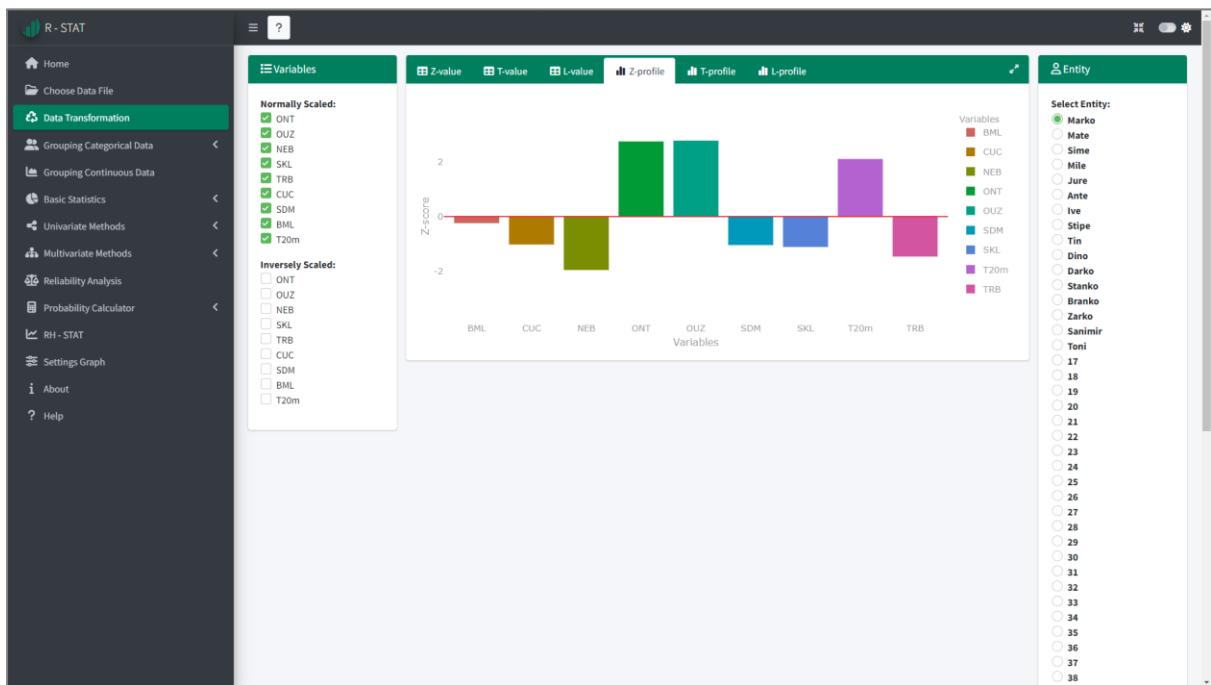
$$L_i = 3 + 0,83 \cdot z_i$$

gdje je

- ✓ L_i transformirani rezultat entiteta i na skali čija je aritmetička sredina 3, a standardna devijacija 0,83 (skala školskih ocjena)
- ✓ z_i standardizirani rezultat entiteta i .

Kondenzirani rezultatati svakog ispitanika u odabranim varijablama izračunati su aritmetičkom sredinom (**MEAN_Z**, **MEAN_T**, **MEAN_L**).

Pored toga, za određenog entiteta (**Select Entity**) prikazuju se grafikoni profila (kartice **Z-profile**, **T-profile** i **L-profile**) u odabranim varijablama (slika 5).



Slika 5: Data Transformation → Z – profile

CILJ

- ✓ Transformirati podatke dobivene mjeranjem u Z-vrijednosti, T-vrijednosti i L-vrijednosti.
- ✓ Kondenzirati transformiranje podatke aritmetičkom sredinom.
- ✓ Grafički prikazati profile entiteta u transformiranim podacima.

UVJETI

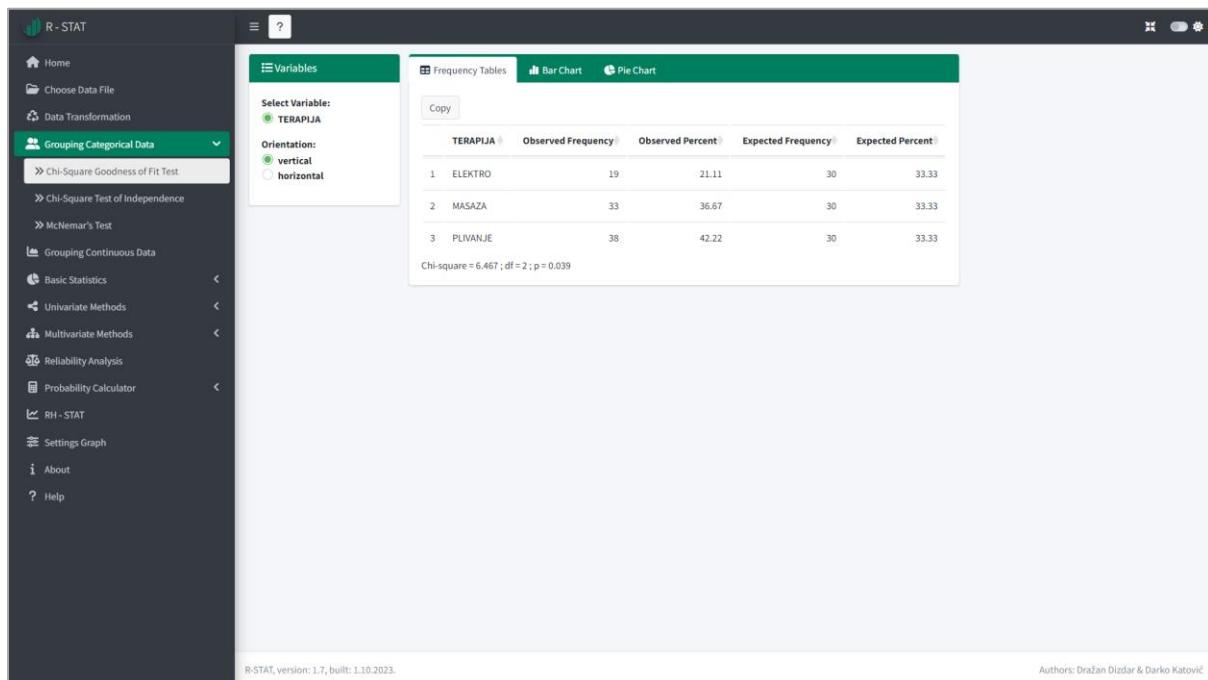
- ✓ Podaci su dobiveni na intervalnoj ili omjernoj mjernoj ljestvici.

Grouping Categorical Data

Chi-Square Goodness of Fit Test

Za korištenje opcije **Chi-square Goodness of Fit** 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 opocije **Chi-square Goodness of Fit** za odabranu varijablu (**Select Variable**) prikazuje se kartica **Frequency Tables** s rezultatima jednodimenzionalnog grupiranja te rezultati χ^2 -testa (hi kvadrat) za utvrđivanje statističke značajnosti razlika između opaženih i očekivanih frekvencija dobivenih uniformnom raspodjelom (slika 6).



Slika 6: Chi-square Goodness of Fit Test → Frequency Tables

U kartici **Frequency Tables** nalazi se tablica koja za svaki modalitet kvalitativne varijable prikazuje:

- **Observed Frequency** – opažene frekvencije
- **Observed Percent** – relativne postotne opažene frekvencije

$$fo(\%)_g = \frac{f_{og}}{n} \cdot 100, \quad g = 1, \dots, k$$

gdje je

- ✓ $fo(\%)_g$ relativna frekvencija modalitet g ($g = 1, \dots, k$) izražena u postotku
- ✓ f_{og} opažena frekvencija modaliteta g
- ✓ n ukupan broj entiteta

- ✓ k broj modalitet.
- **Expected Frequency** – očekivane frekvencije prema uniformnoj raspodjeli

$$fe_g = \frac{n}{k}, \quad g = 1, \dots, k$$
 gdje je
 - ✓ fe_g očekivana frekvencija modaliteta g
 - ✓ n ukupan broj entiteta
 - ✓ k broj modaliteta.
- **Expected Percent** – relativne očekivane frekvencije prema uniformnoj raspodjeli izražene u postotku

$$fe(\%)_g = \frac{fe_g}{n} \cdot 100, \quad g = 1, \dots, k$$

gdje je

- ✓ fe_g očekivana frekvencija modaliteta g ($g = 1, \dots, k$) izražena u postotku
- ✓ n ukupan broj entiteta
- ✓ k broj modaliteta.

te rezultati χ^2 –testa sukladnosti (eng. *Chi-square Goodness of Fit*)⁴:

- **Chi-square** – χ^2 (hi kvadrat) vrijednost

$$\chi^2 = \sum \frac{(f_o - fe)^2}{f_o}$$

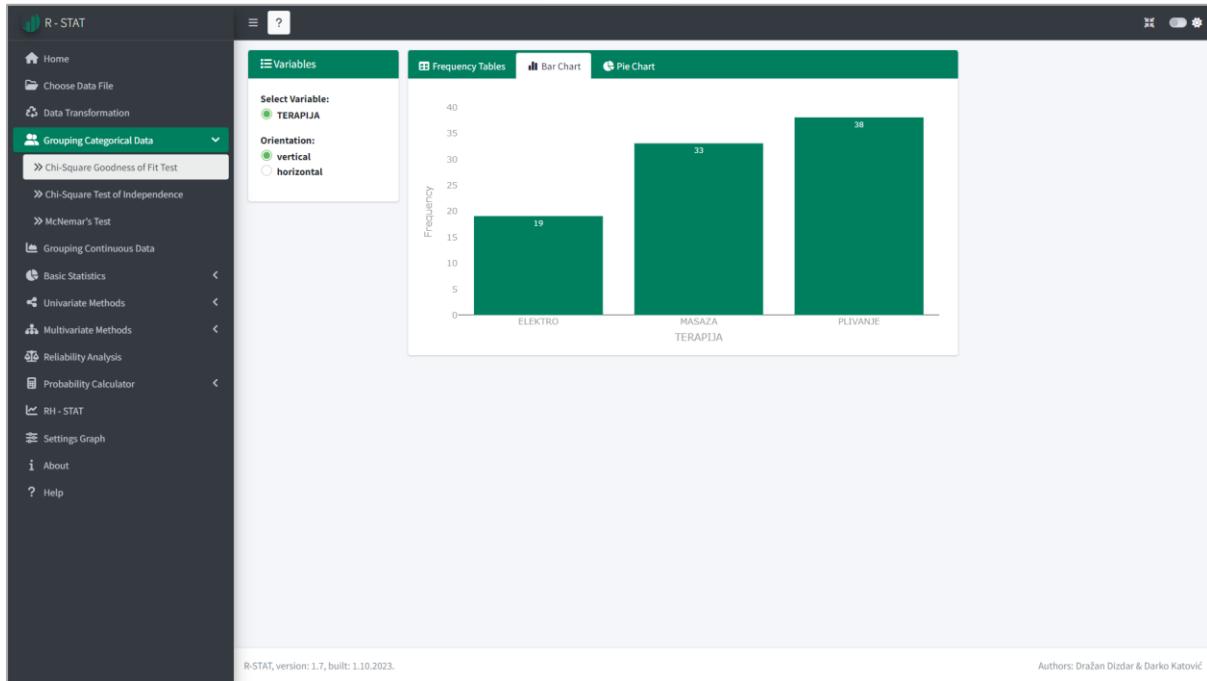
gdje je

- ✓ χ^2 hi kvadrat vrijednost
- ✓ f_o opažene frekvencije
- ✓ fe očekivane frekvencije.

- **df** – broj stupnjeva slobode ($df = k - 1$, gdje je k – broj kategorija, odnosno oblika mјerenog obilježja).
- **p** – pogreška s kojom tvrdimo da je razlika između opaženih i očekivanih frekvencija statistički značajna.

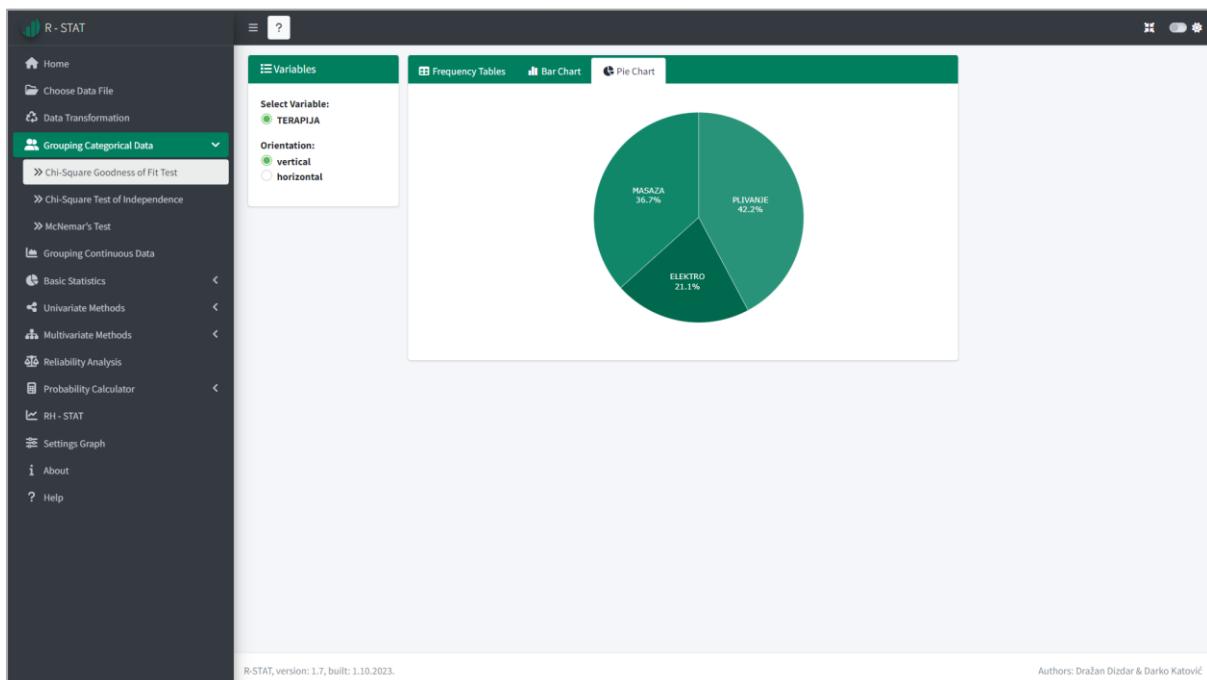
⁴ U anglosaksonskoj literaturi test se naziva *goodness of fit* jer se njime testira koliko dobro (*good*) opažene frekvencije (*observed frequency*) odgovaraju (*fit*) očekivanim frekvencijama (*expected frequency*).

U karticama **Bar Chart** (slika 7) i **Pie Chart** (slika 8) nalaze se grafikoni koji mogu biti vertikalne i horizontalne orijentacije.



Slika 7: Chi-square Goodness of Fit Test → Bar Chart

Bar Chart je grafikon stupaca u kojem se na osi x nalaze se oblici obilježja (modaliteti), a na osi y frekvencije (**vertical**) ili obrnuto (**horizontal**).



Slika 8: Chi-square Goodness of Fit Test → Pie Chart

Pie Chart je strukturni krug koji prikazuje udjele svakog modaliteta u ukupnoj frekvenciji.

CILJ

- ✓ Jednodimenzionalno grupirati kvalitativnih varijabli.
- ✓ Utvrditi da li je razlika između očekivanih i očekivanih frekvencija dobivenih uniformnom raspodjelom statistički značajna.

UVJETI

- ✓ Podaci su dobiveni na nominalnoj ili ordinalnoj mjerenoj ljestvici s dva ili više modaliteta.
- ✓ Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ($n > 30$).
- ✓ Niti jedna očekivana frekvencija ne smije biti manja od 5.

Chi-Square Test of Independence

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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 **Chi-Square Test of Independence** za odabранe varijable (**Select Variable 1** i **Select Variable 2**) prikazuju se tablice u kartici **Contingency Tables** s rezultatima dvodimenzionalnog grupiranja za odabранe varijable te rezultati χ^2 -testa za utvrđivanje veze između njih (slika 9).

χ^2 - test neovisnosti⁵ (*Chi-Square Test of Independence*) isto kao i χ^2 - test sukladnosti (*Chi-square Goodness of Fit Test*) utvrđuje razlikuju li se statistički značajno opažene (**Observed Frequency**) i očekivane frekvencije (**Expected Frequency**) koje se kod χ^2 - test neovisnosti računaju prema formuli

$$fe = \frac{\sum reda \cdot \sum stupca}{n}$$

gdje je

- ✓ fe očekivana frekvencija polja u kontigencijskoj tablici
- ✓ $\sum reda$ predstavlja zbroj svih frekvencija u redu
- ✓ $\sum stupca$ predstavlja zbroj svih frekvencija u stupcu
- ✓ n ukupan broj entiteta.

⁵ U anglosaksonskoj literaturi test se naziva *Chi-Square Test of Independence* jer se njime testira jesu li dvije kvalitativne varijable međusobno povezna (ovisne).

The screenshot shows the R-STAT software interface. On the left, there is a sidebar with the following menu items:

- Home
- Choose Data File
- Data Transformation
- Grouping Categorical Data** (selected)

 - Chi-Square Goodness of Fit Test
 - Chi-Square Test of Independence (selected)
 - McNemar's Test

- Grouping Continuous Data
- Basic Statistics
- Univariate Methods
- Multivariate Methods
- Reliability Analysis
- Probability Calculator
- RH - STAT
- Settings Graph
- About
- Help

The main panel has a title bar with tabs: Contingency Tables, Stacked Bar Chart (f), and Stacked Bar Chart (%). The Contingency Tables tab is selected. It displays the following information:

Chi-square = 0.63 ; df = 2 ; p = 0.73

Observed Frequency

	Dobro	Prosječno	Slabo	Total
M	29	58	20	107
Z	21	52	20	93
Total	50	110	40	200

Observed Frequency (%)

	Dobro	Prosječno	Slabo	Total
M	27.1	54.21	18.69	100
Z	22.58	55.91	21.51	100

Expected Frequency

	Dobro	Prosječno	Slabo	Total
M	26.75	58.85	21.4	107
Z	23.25	51.15	18.6	93

Slika 9: Chi-Square Test of Independence → Contingency Tables

- **Chi-squere – χ^2** (hi kvadrat) vrijednost testa neovisnosti računa se prema formuli

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{0ij} - f_{eij})^2}{f_{eij}}$$

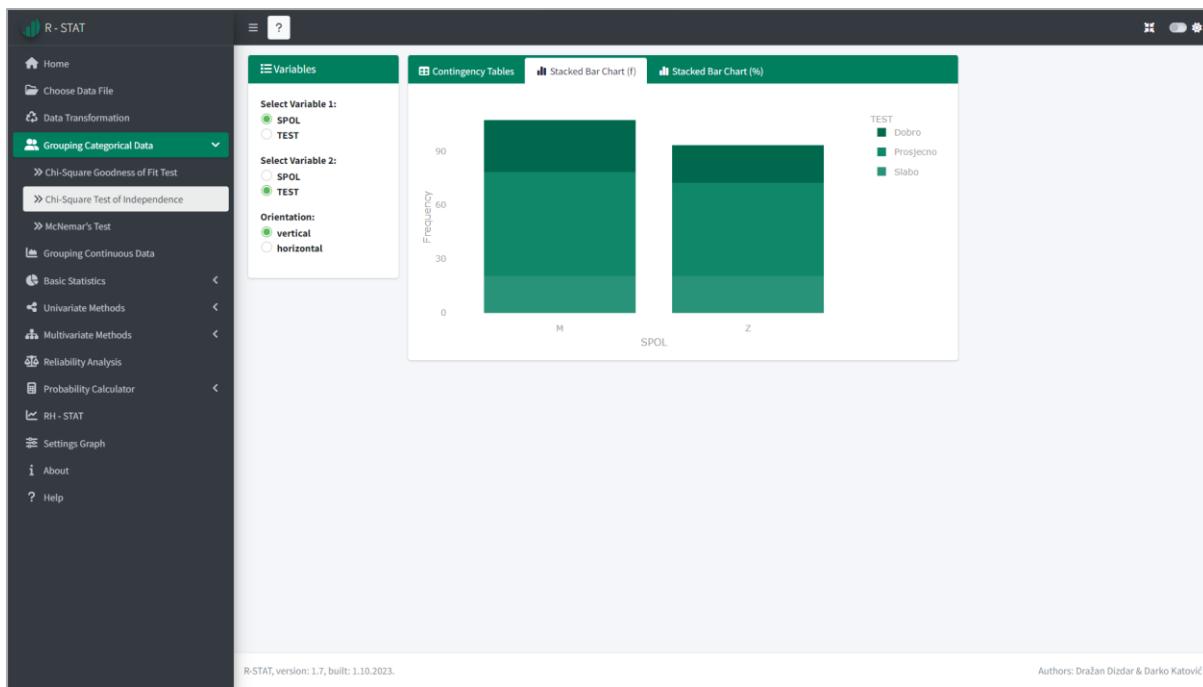
gdje je

- ✓ χ^2 hi kvadrat vrijednost
- ✓ f_{0ij} opažene frekvencije polja ij
- ✓ f_{eij} očekivane frekvencije polja ij
- ✓ r broj redova u kontigencijskoj tablici
- ✓ c broj stupaca u kontigencijskoj tablici.

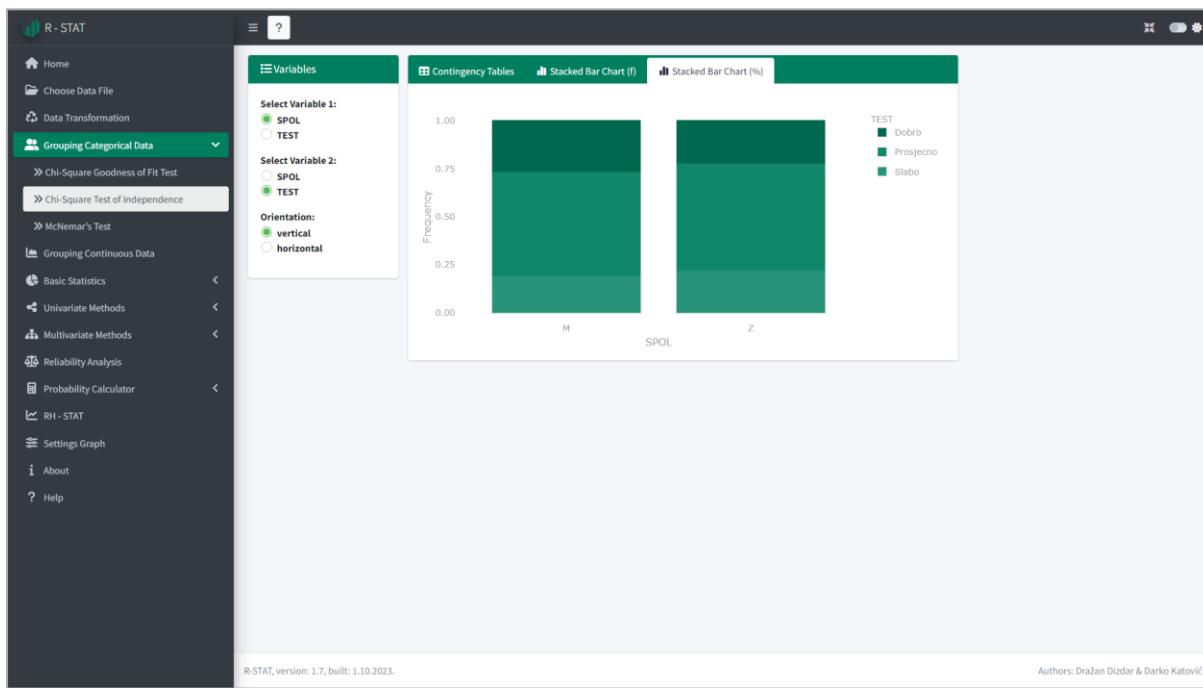
- **df** – broj stupnjeva slobode ($df = (r - 1) \cdot (c - 1)$, gdje je r – broj redova, a c – broj stupaca)⁶
- **p** – pogreška s kojom tvrdimo da je razlika između opaženih i očekivanih frekvecija statistički značajna.

U karticama **Stacket Bar Chart (f)** (slika 10) i **Stacket Bar Chart (%)** (slika 11) nalaze se grafikoni koji mogu biti vertikalne (**vertical**) i horizontalne (**horizontal**) orientacije (**Orientation**).

⁶ U broj redova i broj stupaca ne ubrajam se zbirni red i zbirni stupac.



Slika 10: Chi-Square Test of Independence → Stacked Bar Chart (f)



Slika 11: Chi-Square Test of Independence → Stacked Bar Chart (%)

- **Bar Stacket Chart (f)** – grafikon s razdjeljenim stupacima prikazuje frekvencije dviju kvalitativnih varijabli, gdje se prva varijabla prikazuje na osi x, a druga se varijabla razdjeljuje se prema modalitetima prve varijable.

- **Bar Stacked Chart (%)** – grafikon s razdjeljenim stupacima prikazuje relativne frekvencije dviju kvalitativnih varijabli, gdje se prva varijabla prikazuje na osi x, a druga se varijabla razdjeljuje se prema modalitetima prve varijable.

CILJ

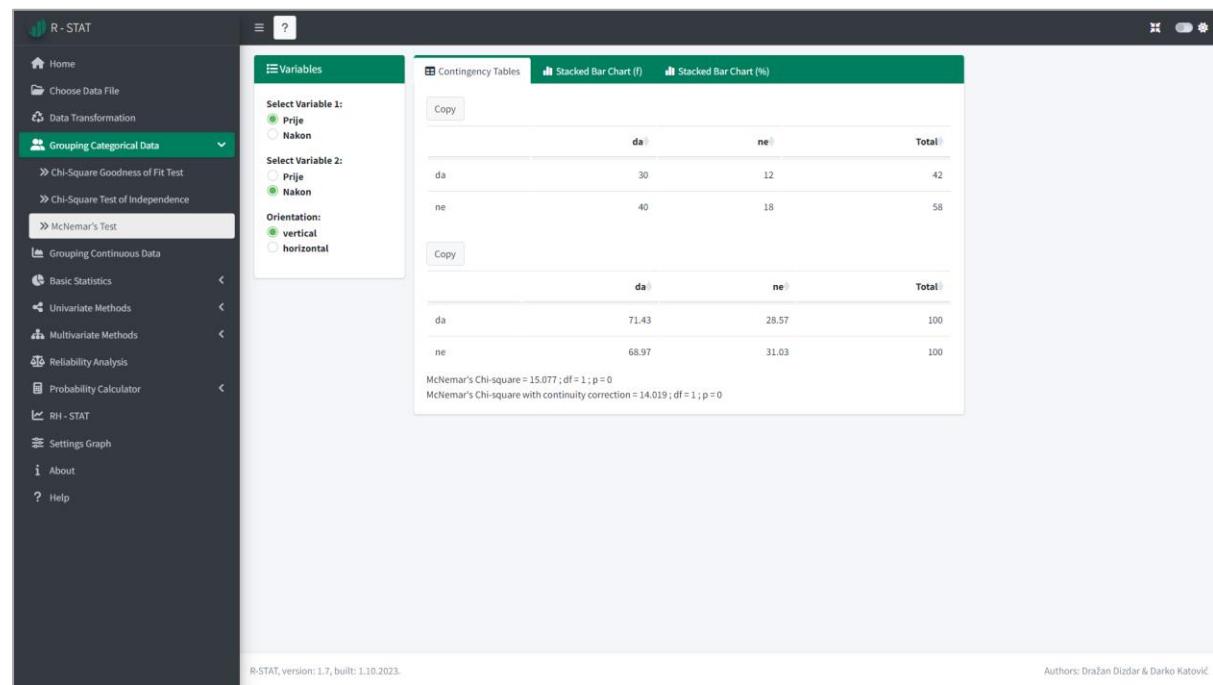
- ✓ Dvodimenzionalno grupirati kvalitativne varijable.
- ✓ Utvrditi postoji li veza između dvije kvalitativne varijable.

UVJETI

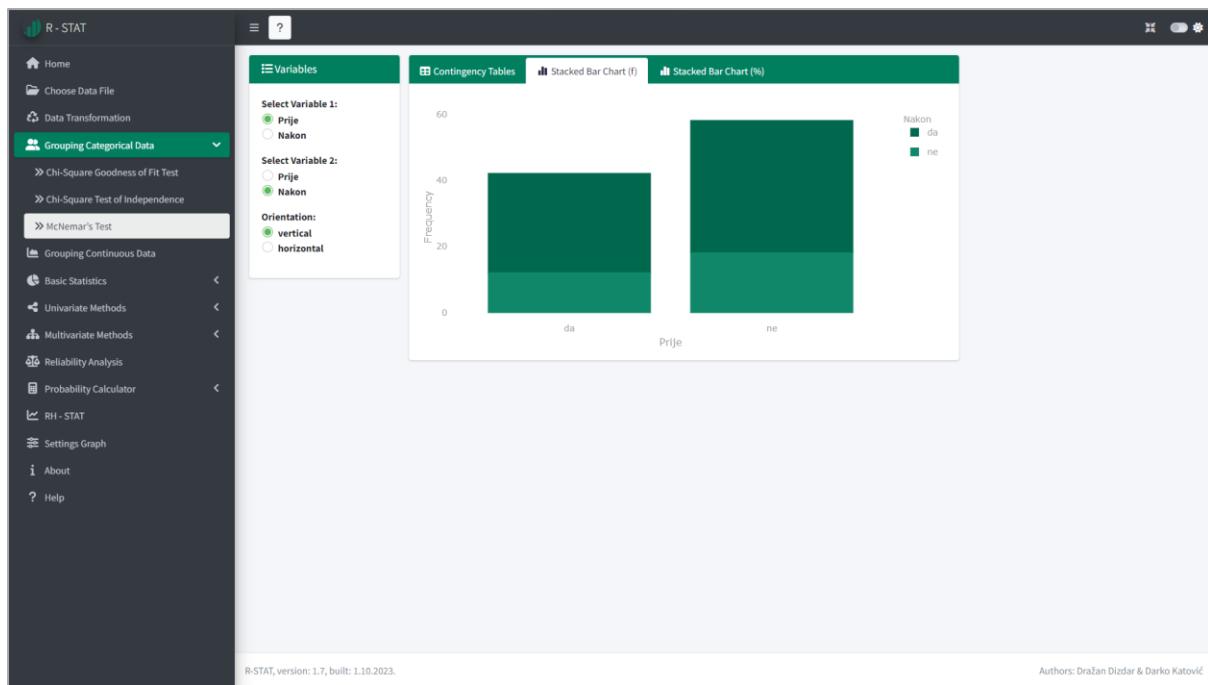
- ✓ Podaci su dobiveni na nominalnoj ili ordinalnoj mjerenoj ljestvici s dva ili više modaliteta.
- ✓ Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ($n > 30$).

McNemar's Test

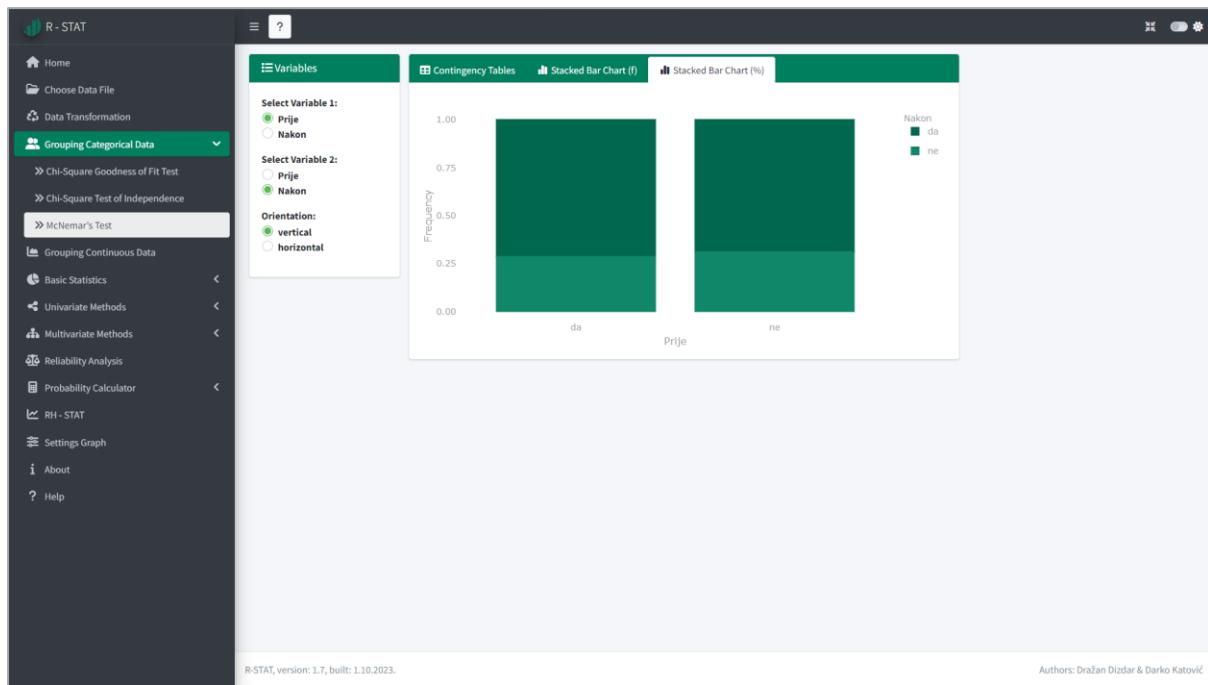
Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje dvije kvalitativne varijable. Primjerice, u datoteci MCT.csv nalaze se podaci 100 entiteta testiranih prije i nakon tretmana.



Slika 12: McNemar's Test → Contingency Tables



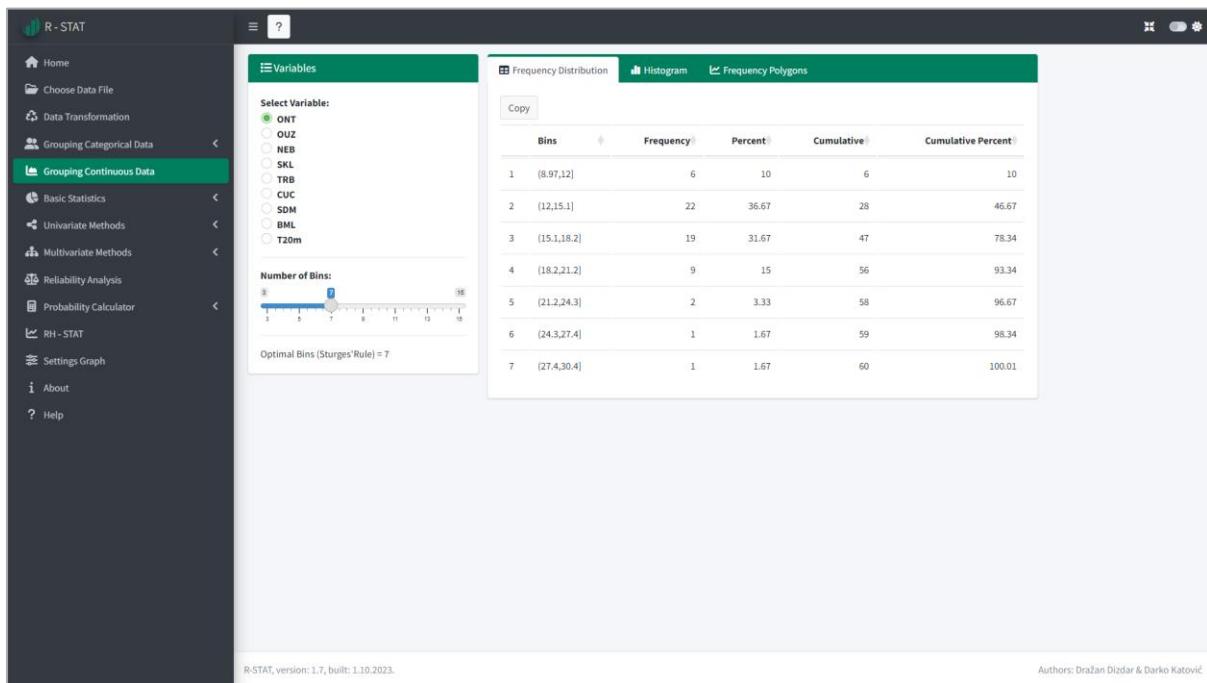
Slika 13: McNemar's Test → Stacked Bar Chart (f)



Slika 14: McNemar's Test → Stacked Bar Chart (%)

Grouping Continuous Data

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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.



Slika 15: Grouping Continuous Data → Frequency Distribution

Odabirom opcije **Grouping Continuous Data** glavnog izbornika prikazuje se tablica u kartici **Frequency Distribution** (slika 15) u kojoj se nalaze:

- **Bins** – donja i gornja granica razreda pri čemu vrijednosti jednake donjoj granici ne pripadaju tom razredu, već prethodnom
- **Frequency** – absolutne frekvencije
- **Percent** – relativne postotne frekvencije

$$f(\%)_g = \frac{f_g}{n} \cdot 100, \quad g = 1, \dots, k$$

gdje je

- ✓ $f(\%)_g$ relativna postotna frekvencija razreda g ($g = 1, \dots, k$)
- ✓ f_g frekvencija razreda g
- ✓ n ukupan broj entiteta
- ✓ k broj razreda.

- **Cumulative** – kumulativne frekvencije pokazuju koliko je entiteta (apsolutno ili relativno) kojima je vrijednost jednaka ili manja od gornje granice razreda čija je frekvencija ušla u kumulativni niz

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

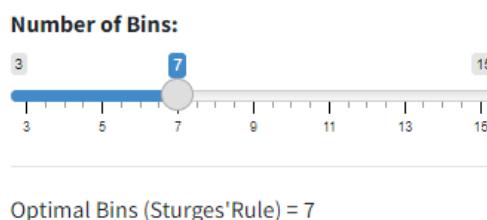
Za uspješno grupiranje potrebno je odrediti optimalan broj i interval razreda. Broj razreda prije svega ovisi o broju entiteta i najčešće se kreće između pet i petnaest. Za procjenu optimalnog broja razreda moguće je koristiti Sturgersovo pravilo

$$k \approx 1 + 3,3 \cdot \log(n)$$

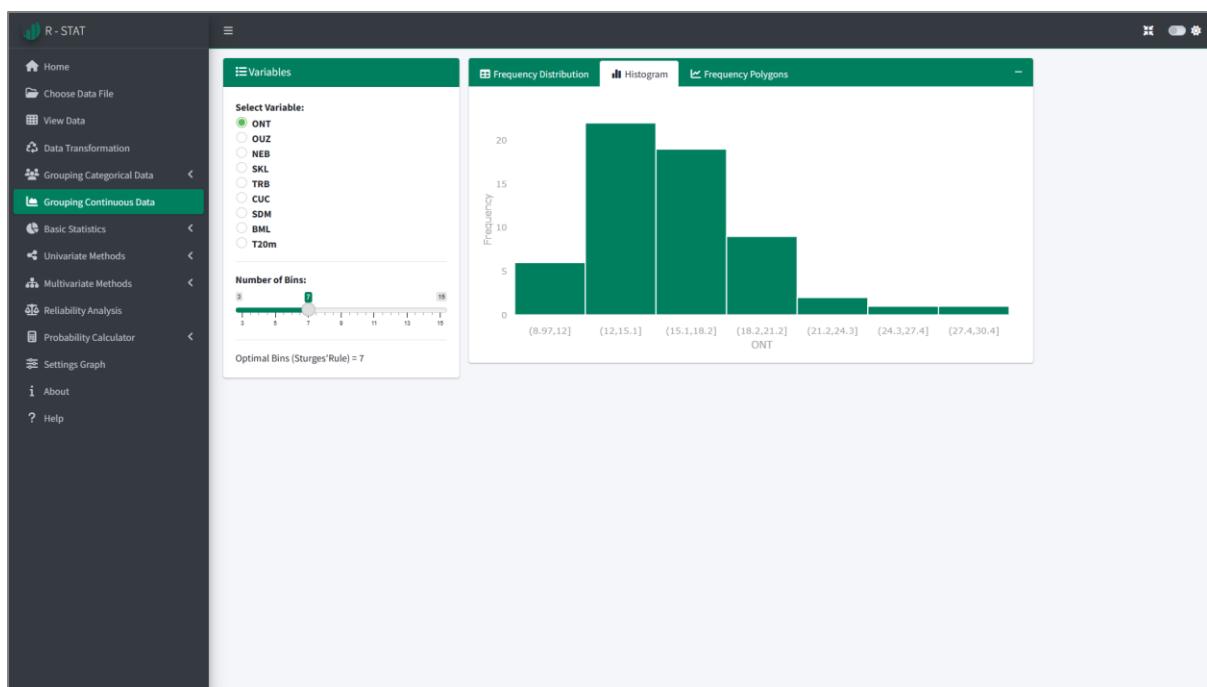
gdje je:

- ✓ k – optimalan broj razreda
- ✓ n – broj entiteta.

Klizačem (**Number of Bins**) se određuje broj razreda.

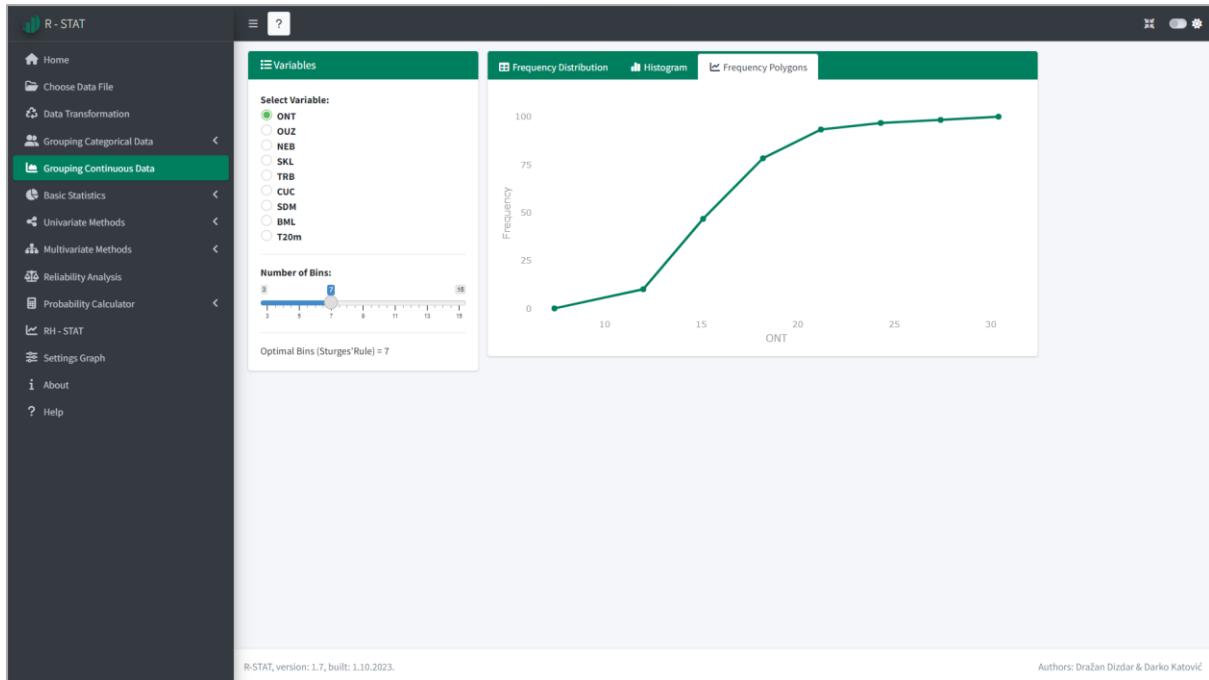


U karici **Histogram** nalazi se histogram apsolutnih frekvencija (slika 16), a u kartici **Frequency Polygons** poligon relativnih postotnih kumulativnih frekvencija (slika 17).



Slika 16: Grouping Continuous Data → Histogram

- **Histogram frekvencija** je površinski grafički prikaz distribucije frekvencija. Crta se tako da osnovicu pravokutnika određuje interval razreda, a visinu frekvencija pojedinog razreda.



Slika 17: Grouping Continuous Data → Frequency Polygons

- **Poligon frekvencija** je linijski grafički prikaz koji nastaje spajanjem točaka čije su koordinate određene gornjom granicom razreda i frekvencijom tog razreda.

CILJ

- ✓ Grupirati kvantitativne varijable u optimalan broj razreda i grafički prikazati distribuciju frekvencija.

UVJETI

- ✓ Podaci su dobiveni na intervalnoj ili omjernoj mjernoj ljestvici.
- ✓ Podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ($n > 30$).

Basics Statistics

Descriptive Parameters

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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.

The screenshot shows the R-STAT application interface. On the left, there is a sidebar with various statistical analysis options like Home, Choose Data File, Data Transformation, etc. The 'Basic Statistics' section is selected, and 'Descriptive Parameters' is highlighted. The main area displays a table of descriptive statistics for nine variables: ONT, OUZ, NEB, SKL, TRB, CUC, SDM, BML, and T20m. The table includes columns for Mean, Median, SD, CV, Min, Max, Range, Skew, and Kurt. Below the table, there is a 'Box and Whiskers plot' button. At the bottom of the main area, it says 'R-STAT, version: 1.7, built: 1.10.2023.' and 'Authors: Dražan Dizdar & Darko Katović'.

	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	8.79	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

Slika 18: Basic Statistics → Descriptive Parameters → Descriptive Parameters

Odabirom opcije **Basics Statistics** glavnog izbornika te opcije **Descriptive Parameters** za odabrane varijable (**Select Variables**) prikazuje se tablica deskriptivnih pokazatelja (slika 18):

- **MEAN** – aritmetička sredina izračunava se omjer zbroja svih vrijednosti neke varijable i ukupnog broja entiteta

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ gdje je } i = 1, \dots, n, \text{ a } n \text{ predstavlja broj entiteta.}$$

- **MEDIAN** – vrijednost koja se nalazi na sredini uređenog niza podataka (uzlazno ili silazno sortiranog), odnosno vrijednost koja uređeni niz podataka dijeli na dva jednakobrojna dijela. Ako je broj podataka neparan, onda medijan predstavlja vrijednost središnjeg člana tj. entiteta (x_r)

$$\mu_e = x_r \text{ gdje je } r = \frac{n+1}{2}$$

Ako je broj podataka paran, onda je medijan jednak aritmetičkoj sredini vrijednosti dvaju središnjih članova uređenog niza.

$$\mu_e = \frac{x_{r_1} + x_{r_2}}{2} \text{ gdje je } r_1 = \frac{n}{2}, \text{ a } r_2 = \frac{n}{2} + 1$$

- **RANGE** – totalni raspon (R_{tot}) je najjednostavnija mjera varijabilnosti, a utvrđuje se kao razlika između maksimalne (x_{max}) i minimalne (x_{min}) vrijednosti.

$$R_{tot} = x_{max} - x_{min}$$

- **SD** – standardna devijacija korjen iz prosječnog kvadratnog odstupanja rezultata entiteta od aritmetičke sredine.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

- **CV** – koeficijent varijabilnosti izračuna se omjer standardne devijacije (s) i aritmetičke sredine (\bar{x}) pomnožen sa 100.

$$V = \frac{s}{\bar{x}} \cdot 100$$

- **MIN** – minimalna vrijednost
- **MAX** – maksimalna vrijednost
- **SKEW** – skewness koeficijent asimetrije (tip 1) izračunava se kao omjer trećeg momenta oko sredine (m_3) i standardne devijacije podignute na treću potenciju (s^3)

$$a_3 = \frac{m_3}{s^3}$$

gdje je $m_3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n}$ treći moment oko sredine.

Mnogi statistički programi izračunavaju korigirani koeficijent asimetrije (tip 2) s obzirom na veličinu uzorka (koeficijent korekcije se približava 1 kako raste veličina uzorka) formulom

$$a_3 = \frac{m_3}{s^3} \cdot \frac{\sqrt{n \cdot (n-1)}}{(n-2)}$$

Ako je a_3 biti veći od nule distribucija je pozitivno asimetričnu distribuciju, a ako je a_3 manji od nule, onda se radi o negativnoj asimetriji distribucije.

- **KURT** – kurtosis stupanj spljoštenosti ili izduženosti distribucije (tip 1) izračunava se preko četvrtog momenta oko sredine (m_4) i standardne devijacije podignute na četvrtu potenciju (s^4).

$$a_4 = \frac{m_4}{s^4} - 3$$

gdje je $m_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}$ četvrti moment oko sredine.

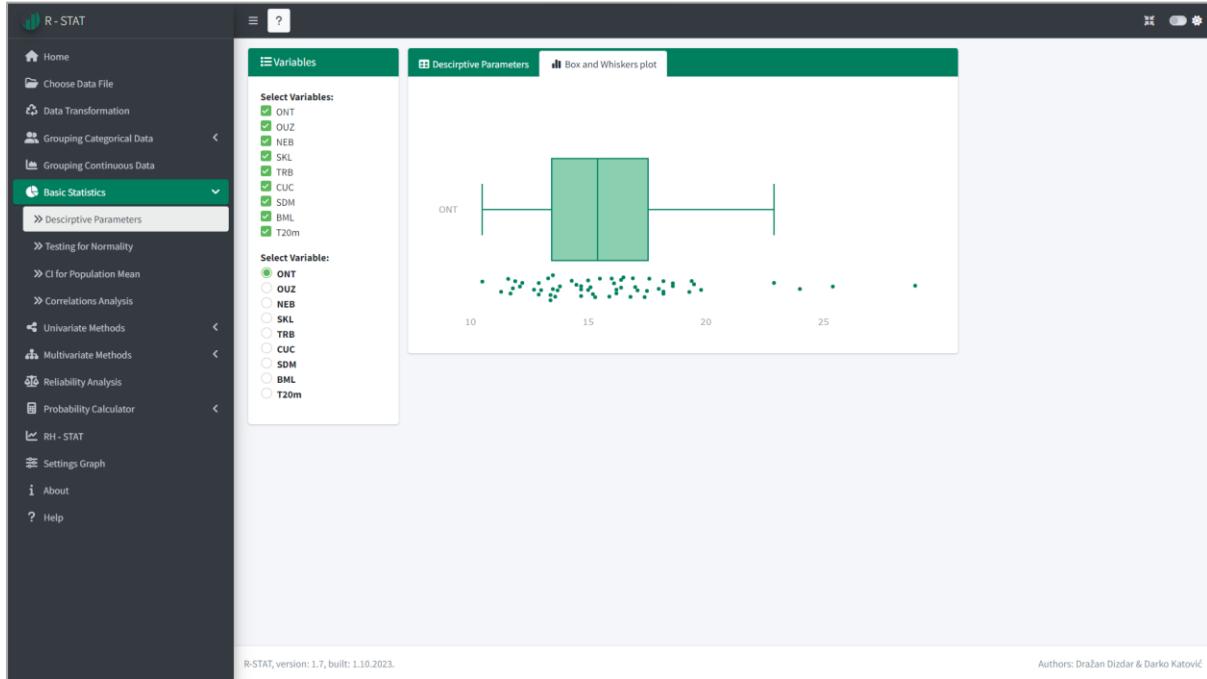
Korigirani koeficijet a_4 (tip 2) računa se formulom

$$a_4 = \left\{ \frac{n-1}{(n-2) \cdot (n-3)} \cdot [(n+1) \cdot a_4 + 6] \right\} - 3$$

Ako je koeficijent spljoštenosti:

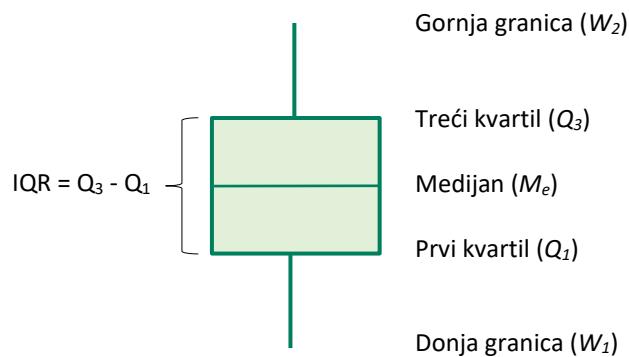
- ✓ $a_4 = 0$ - distribucija je mezokurtična - normalna
- ✓ $a_4 > 0$ - distribucija je leptokurtična - izdužena
- ✓ $a_4 < 0$ - distribucija je platikurtična - spljoštena.

Odabirom kartice **Box and Whiskers plot** (slika 19) prikazuju se odgovarajući grafikoni za odabranu varijablu **Select Variable**.



Slika 19: Basic Statistics → Descriptive Parameters → Box and Whiskers plot

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

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

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

CILJ

- ✓ Opis kvantitativnih varijabli pomoću deskriptivnih pokazatelja (mjere centralne tendencije, mjere disperzije i mjere oblika distribucije).

UVJETI

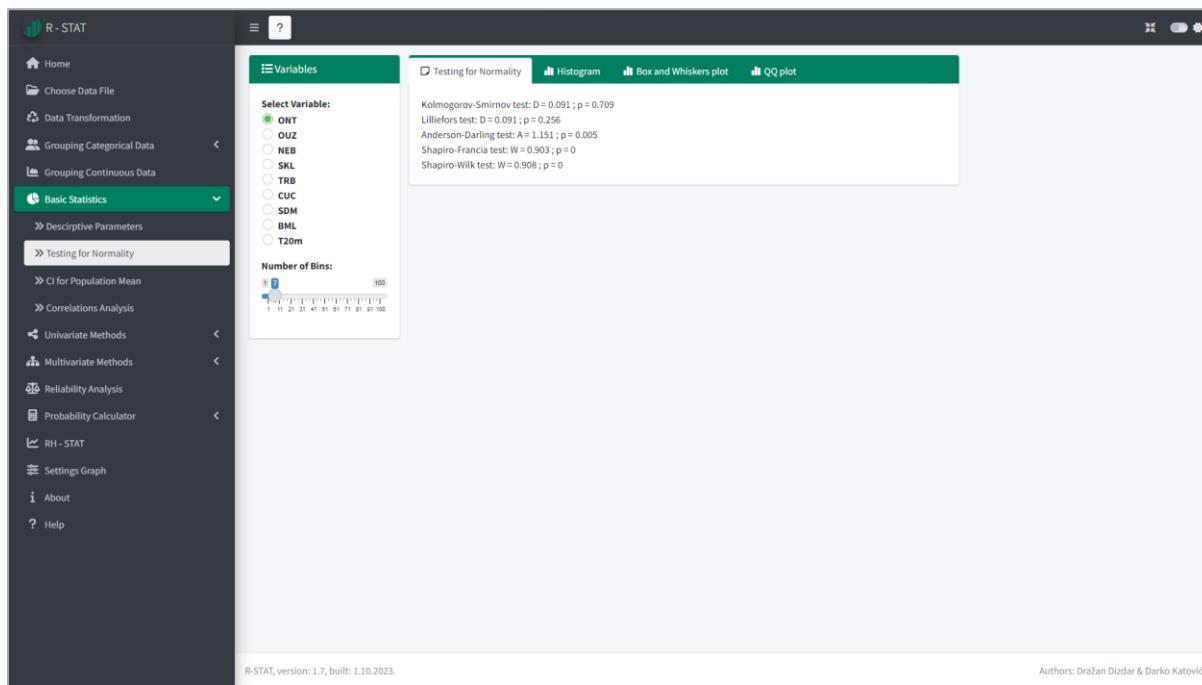
- ✓ Podaci su dobiveni na intervalnoj ili omjernoj mjerenoj ljestvici

Testing for Normality

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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-Wilkov test.



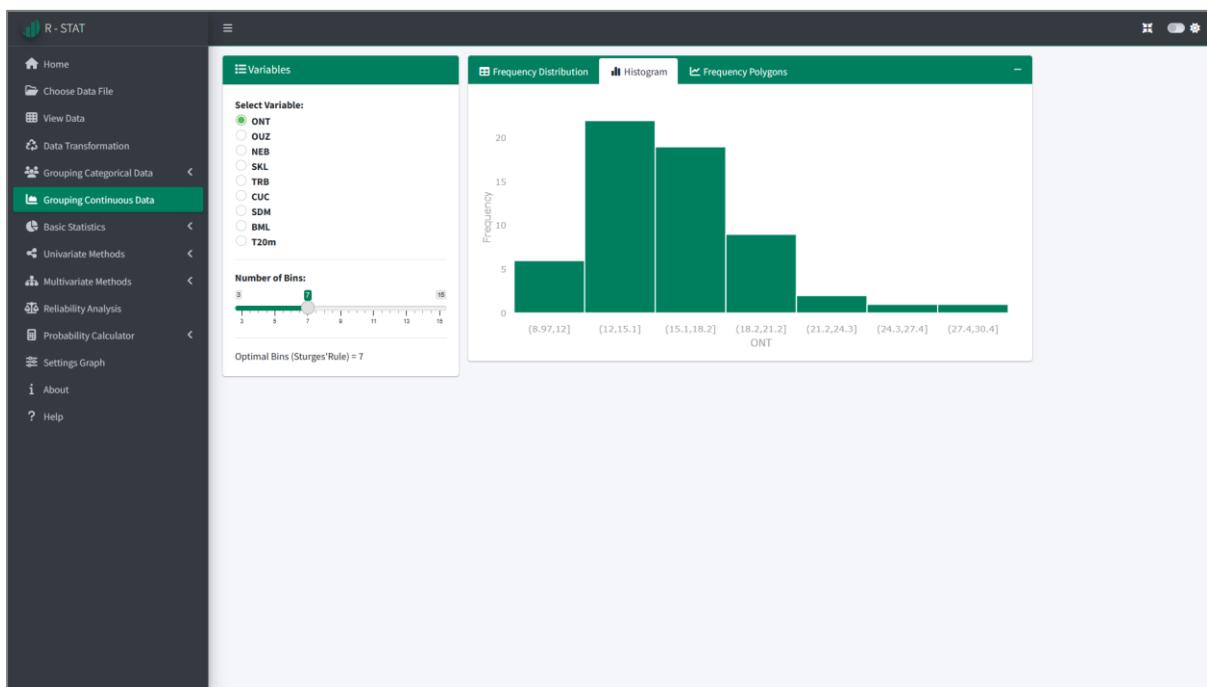
Slika 21: Basic Statistics → Testing for Normality

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 uzroci, 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.

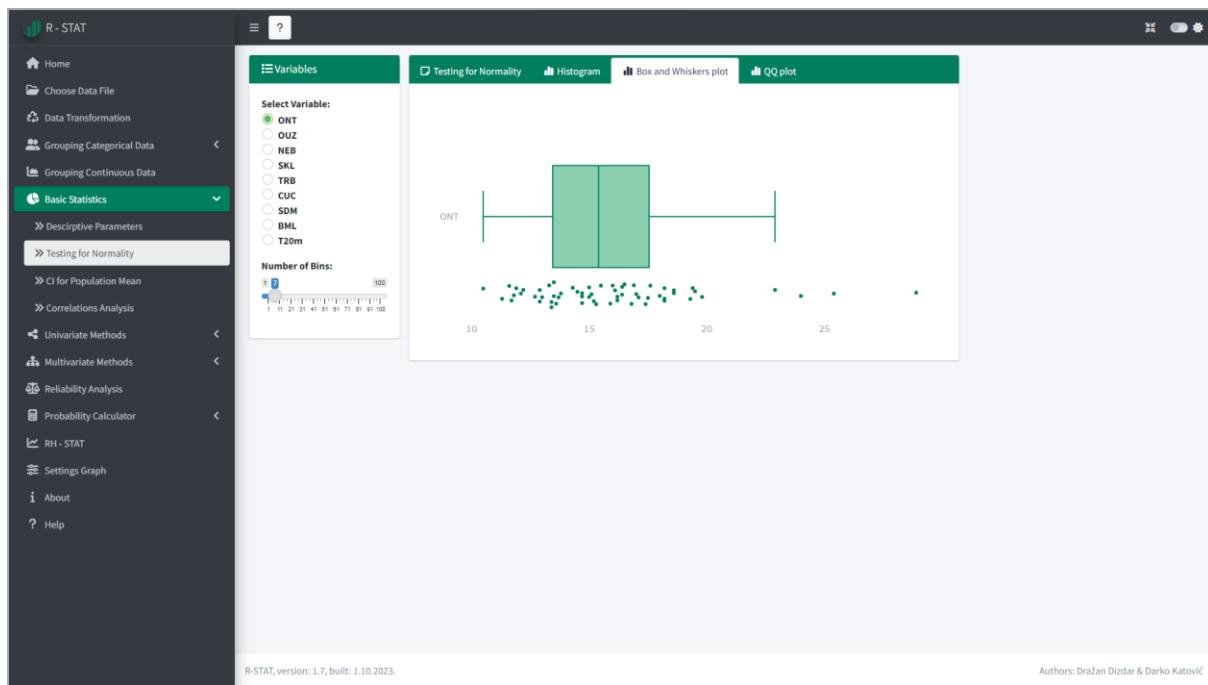
Vrijednosti parametra W Shapiro-Wilkovog testa kreću se u intervalu od 0 do 1, pri čemu niže vrijednosti pokazuju veća, a više vrijednosti manja odstupanja od normalne distribucije. Shapiro-Wilkov test modificirao je Royston (1992) i proširio mogućnost korištenja za uzorke veće od 50 entiteta. Razali 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 najsnazniji, 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 (skewness i kurtosis) te statističkim testovima kako bi se što pouzdano utvrdila razina i vrsta odstupanja neke empirijske od normalne distribucije.

Odabirom kartice **Histogram** (slika 22), **Box and Whiskers plot** (slika 23) ili **QQ plot** (slika 24) prikazuju se odgovarajući grafikoni za odabranu varijablu (**Select Variable**).



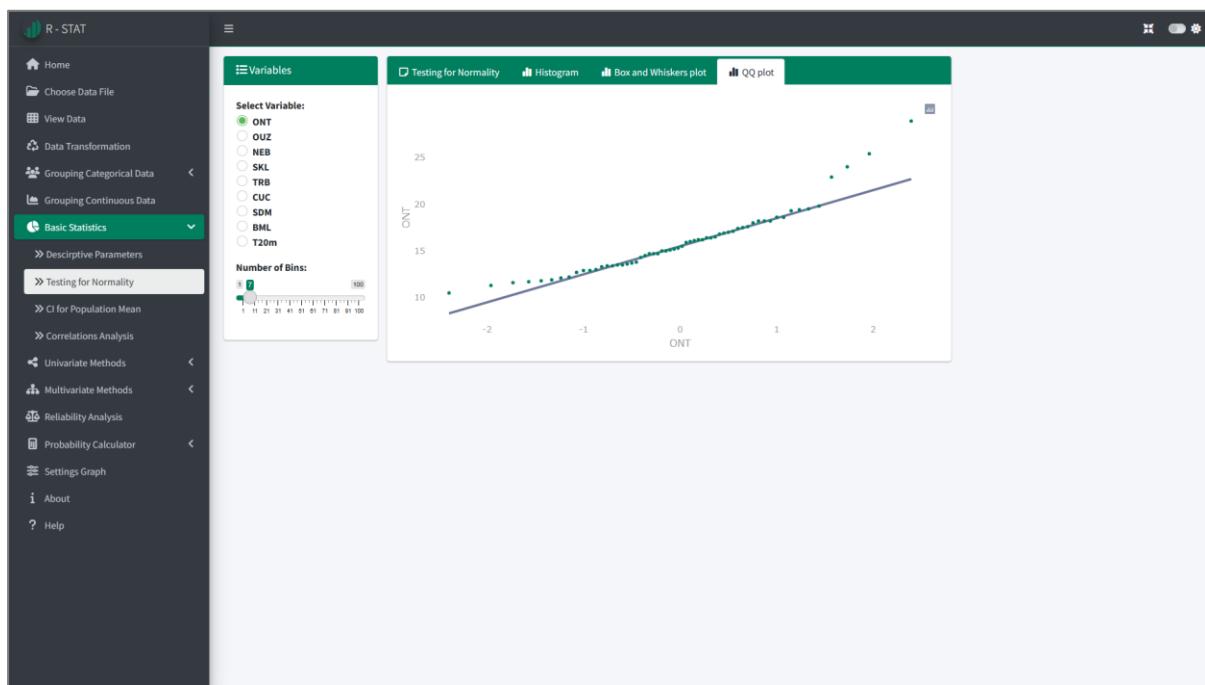
Slika 22: Basic Statistics → Testing for Normality → Histogram

- **Histogram** je površinski grafički prikaz distribucije frekvencija. Crta se tako da osnovicu pravokutnika određuje interval razreda, a visinu frekvencija pojedinog razreda.



Slika 23: Basic Statistics → Testing for Normality → Box and Whiskers plot

- **Box and whisker plot** je kutijasti dijagram koji prikazuje odnose pet statističkih pokazatelja temeljem kojeg je moguće uočiti stupanj disperzije i asimetrije distribucije te outliere (vrijednosti koje ekstremno odstupaju od ostalih). Za detaljniji opis vidi str. 23-24.



Slika 24: Basic Statistics → Testing for Normality → QQ Plot

- **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.

CILJ

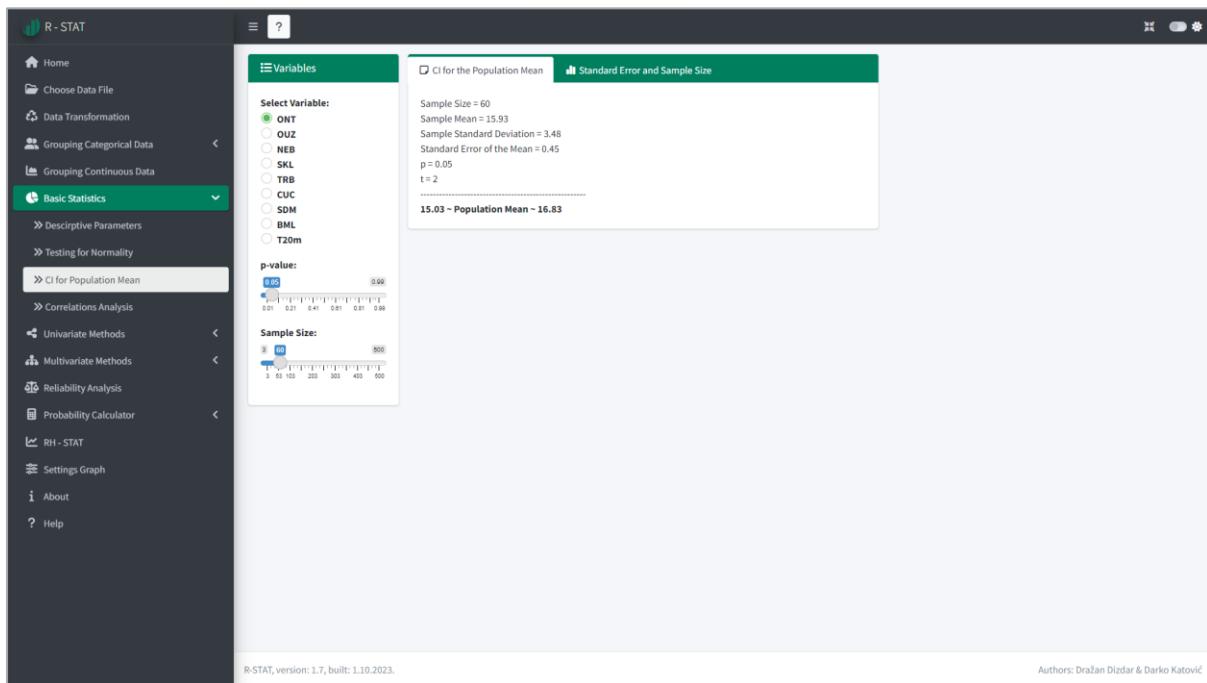
- ✓ Utvrditi odstupa li statistički značajno empirijska distribucija neke kvantitativne varijable od normalne ili Gaussove distribucije.

UVJETI

- ✓ Podaci su dobiveni na intervalnoj ili omjernoj mjerenoj ljestvici

CI for Population Mean

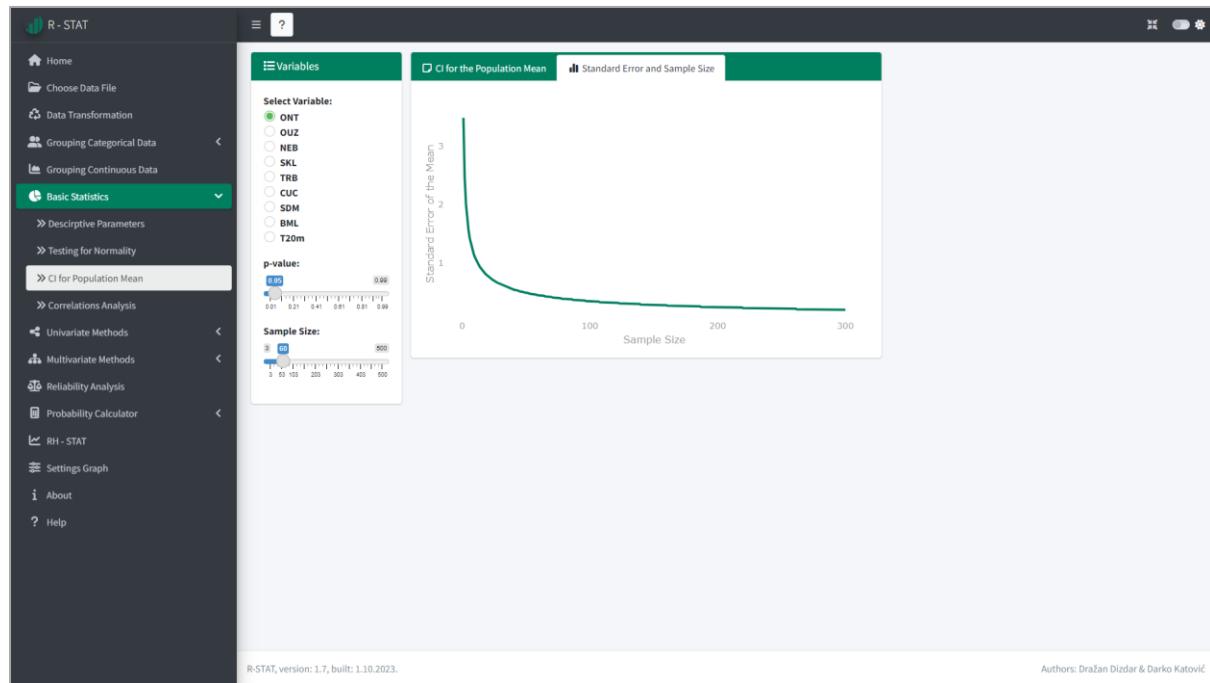
Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje jednu kvantitativnu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



Slika 25: Basic Statistics → CI for the Population Mean → CI for the Population Mean

U okviru opcije **Basics Statistics** glavnog izbornika odabriom opcije **CI for Population Mean** za odabranu varijablu (**Select Variable**) aplikacija računa (slika 25) aritmetičku sredinu (**Sample Mean**) i standardnu devijaciju (**Sample Standard Deviation**) uzorka te zavisno o veličini uzorka računa standardnu pogrešku aritmetičke sredine (**Standard Error of the Mean**) na temelju koje računa interval u kojem se nalazi aritmetička sredina populacije (**Population Mean**), zavisno o pogrešci statističkog zaključka (**p-value**).

Pored toga, u kartici **Standard Error and Sample Size** (slika 26) nalazi se grafikon koji pokazuje odnos standardne pogreške aritmetičke sredine (**Standard Error od Mean**) i veličine uzorka (**Sample Size**) s obzirom na standardnu devijaciju odabrane varijable (**Select Variable**).



Slika 26: Basic Statistics → CI for the Population Mean → Standard Error and Sample Size

CILJ

- ✓ Na temelju slučajno odabranog uzorka entiteta uz pogrešku p procjeniti interval u kojem se nalazi aritmetička sredina populacije.

UVJETI

- ✓ Podaci su dobiveni na intervalnoj ili omjernoj mjerenoj ljestvici.

Correlations Analysis

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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 (slika 27).

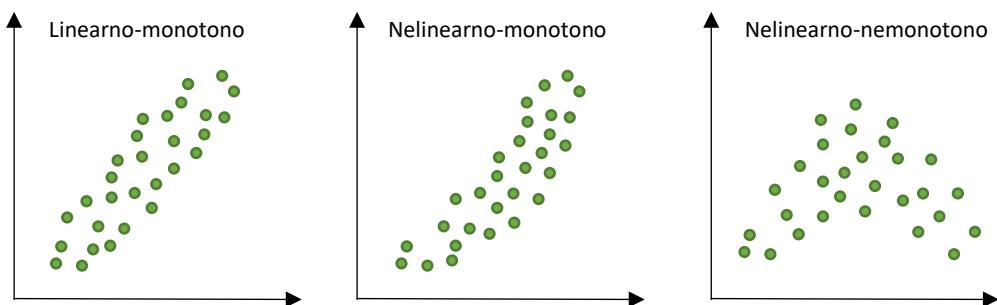
ONT	OUZ	NEB	SKL	TRB	CUC	SDM	BML	T20m	
1.000									
OUZ	0.657								
NEB	-0.357	-0.211	1.000						
SKL	-0.348	-0.620	0.152	1.000					
TRB	-0.259	-0.532	0.298	0.726	1.000				
CUC	-0.057	-0.214	0.140	0.408	0.749	1.000			
SDM	-0.383	-0.609	0.136	0.483	0.471	0.272	1.000		
BML	-0.117	-0.279	0.043	0.089	0.004	-0.147	0.456	1.000	
T20m	0.343	0.648	-0.094	-0.366	-0.359	-0.172	-0.776	-0.663	1.000

Slika 27: Basic Statistics → Correlations Analysis → Correlations

Aplikacija omogućava izračunavanje Pearsonovog (**Pearson**), Spearmanovog (**Spearman**) i Kendallovog (**Kendall**) koeficijenta korelacije (kartica **Correlations**, slika 27), utvrđivanje njihove statističke značajnosti (kartica **Level of Significance**, slika 28) te grafičko prikazivanje korelativnih odnosa putem korelacijskog dijagrama (kartica **Scatter Plot**, slika 29).

Za utvrđivanje povezanosti između dviju varijabli najčešće se koristi Pearsonov koeficijent korelacija. Da bismo koristili Pearsonov koeficijent korelacije, podaci ne bi trebali imati značajne *outliere*, odnosno ekstremno visoke ili niske rezultate te bi trebale biti u linearном odnosu i imati normalnu ili Gaussovnu distribuciju.

Ako su varijable mjerene na ordinalnoj, intervalnoj ili omjernoj mjernoj ljestvici, koje nisu normalno distribuirane te se ne nalaze u linearном, već monotonom odnosu, tada se za utvrđivanje povezanosti između dviju varijabli koristi Spearmanov koeficijent rang korelacijske.

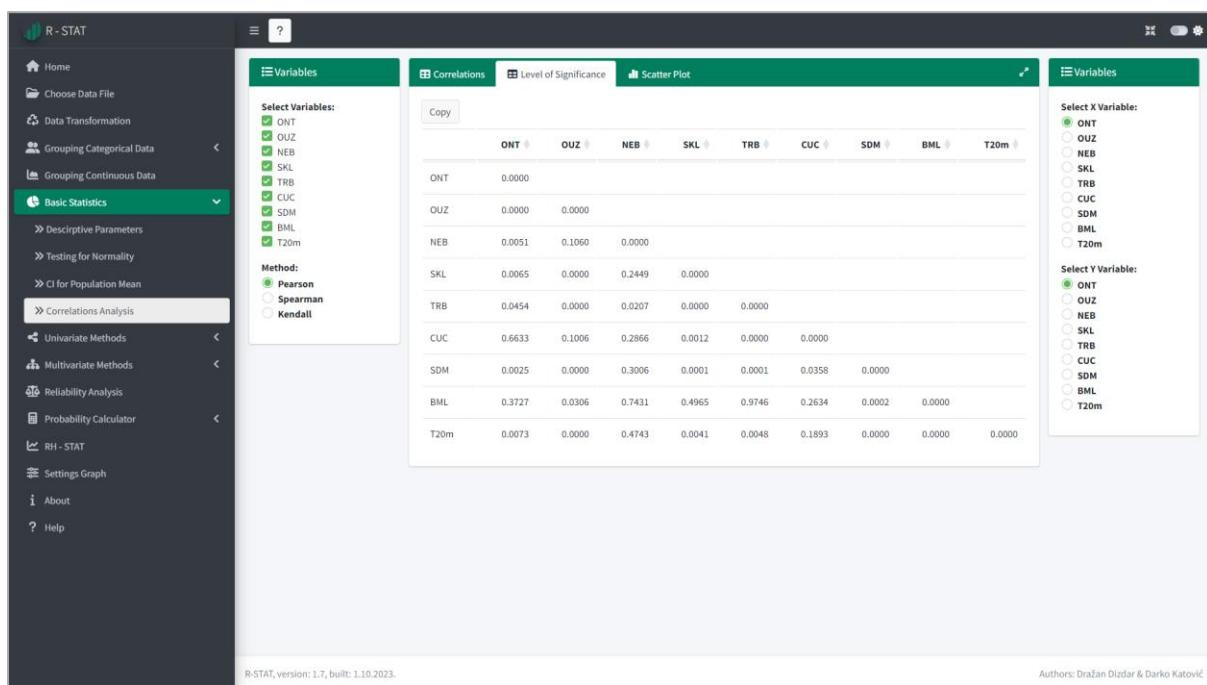


Spearmanov koeficijent rang korelacijske izračunava se prema formuli

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

gdje je

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



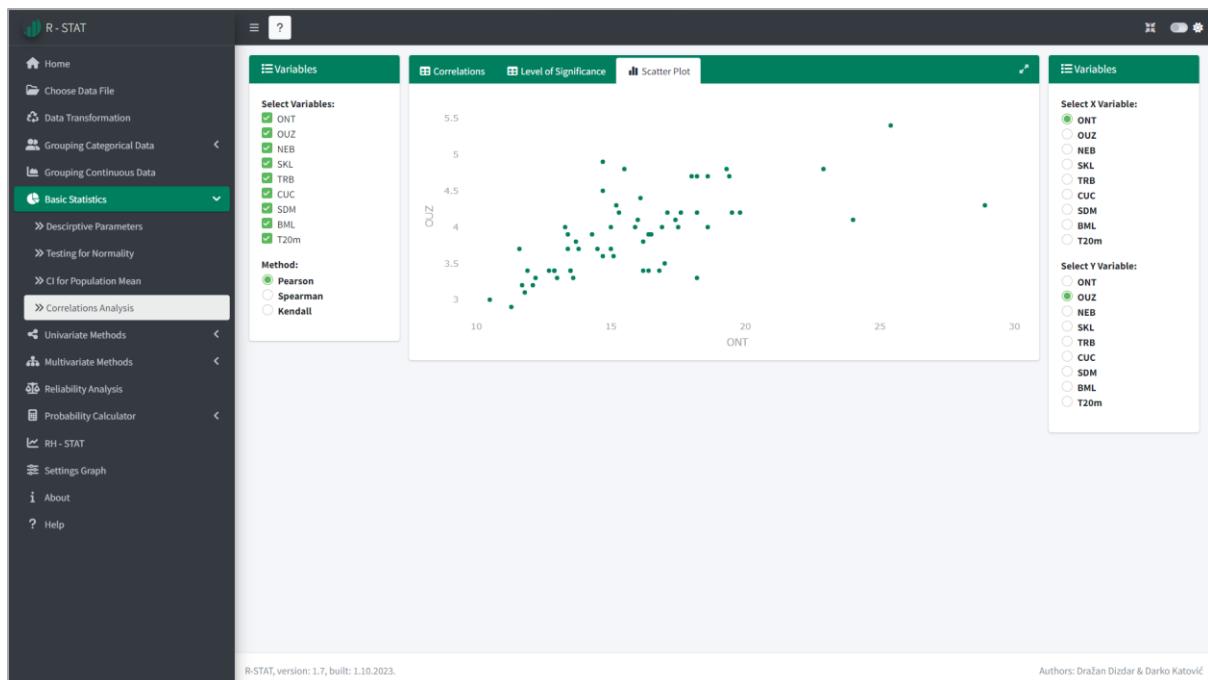
Slika 28: Basic Statistics → Correlations Analysis → Level of Significanse

CILJ

- ✓ utvrditi povezanost između varijabli

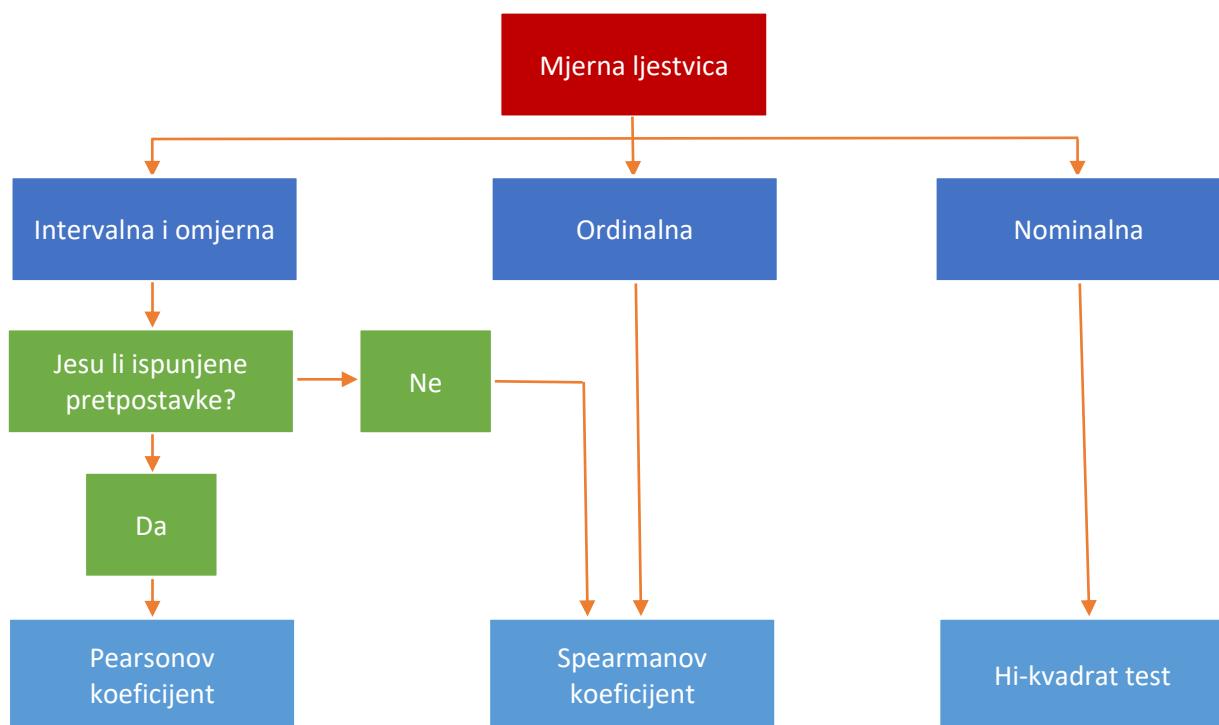
UVJETI

- ✓ podaci su dobiveni na ordinalnoj, intervalnoj ili omjernoj mjernoj ljestvici
- ✓ podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ($n > 30$).



Slika 29: Basic Statistics → Correlations Analysis → Scatter Plot

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



Univariate Methods

Independent Samples T-Test

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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.

The screenshot shows the R-STAT interface. On the left, there's a sidebar with various statistical methods: Home, Choose Data File, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods (selected), Independent Samples T-Test (selected), Dependent Samples T-Test, One-way ANOVA, Repeated Measures ANOVA, Multivariate Methods, Reliability Analysis, Probability Calculator, RH-STAT, Settings Graph, About, and Help. The main area has tabs for 'Variables' (selected), 'T-test (more variables)' (selected), 'Descriptive Parameters', and 'T-test (one variable)'. Under 'Select Dependent Variables', several variables are checked: ATV, ATT, AOP, ANN, MKUS, MPOL, MP20, MPRR, MTAP, MSDM, MDTR, and MVIS. Under 'Select Grouping Variable', 'KRAJ' is selected. The 'T-test (more variables)' tab displays a table of results:

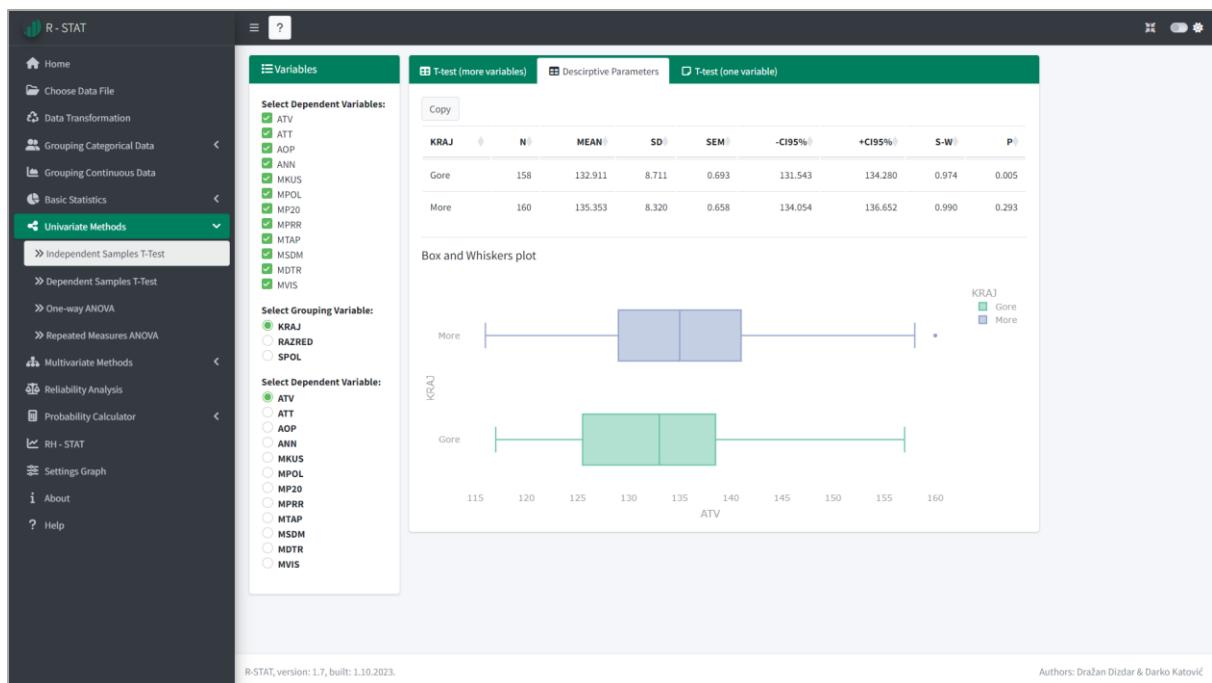
	MEAN1	MEAN2	SED	T	DF	P	CI-95%	CI+95%	Levene's F	P	Cohen's D
ATV	132.911	135.353	0.955	2.556	316	0.011	-4.321	-0.562	0.572	0.45	0.287
ATT	29.91	31.472	0.754	2.072	316	0.039	-3.046	-0.079	0.542	0.462	0.232
AOP	19.297	19.471	0.191	0.911	316	0.363	-0.551	0.202	0.204	0.652	0.102
ANN	11.956	11.981	0.525	0.049	316	0.961	-1.058	1.007	4.163	0.042	0.005
MKUS	164.07	166.069	4.698	0.426	316	0.671	-11.243	7.245	0.2	0.655	0.048
MPOL	240.886	228.613	9.027	1.36	316	0.175	-5.487	30.035	2.194	0.14	0.152
MP20	65.918	62.206	3.3	1.125	316	0.262	-2.782	10.205	0.761	0.384	0.126
MPRR	55.373	43.206	1.868	6.512	316	0	8.491	15.843	0.42	0.517	0.73
MTAP	21.013	21.688	0.537	1.256	316	0.21	-1.732	0.383	0.002	0.961	0.141
MSDM	117.627	133.175	2.879	5.4	316	0	-21.214	-9.883	1.596	0.207	0.606
MDTR	24.652	23.288	0.947	1.441	316	0.15	-0.498	3.227	0.463	0.497	0.162
MVIS	25.253	27.212	2.679	0.731	316	0.465	-7.23	3.311	4.1	0.044	0.082

Slika 30: Univariate Methods → Independent Samples T-Test → T-test (more variables)

U kartici **T-Test (more variables)** prikazuje se tablica (slika 30) u kojoj je:

- **MEAN1** – aritmetička sredina prve grupe
- **MEAN2** – aritmetička sredina druge grupe
- **SED** – standardna pogreška razlika aritmetičkih sredina
- **T** – 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.

- **CI-95%** – donja granica intervala u kojem se s vjerojatnošću 95% nalazi razlika aritmetičkih sredina populacije prve i druge grupe
- **CI+95%** – 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).

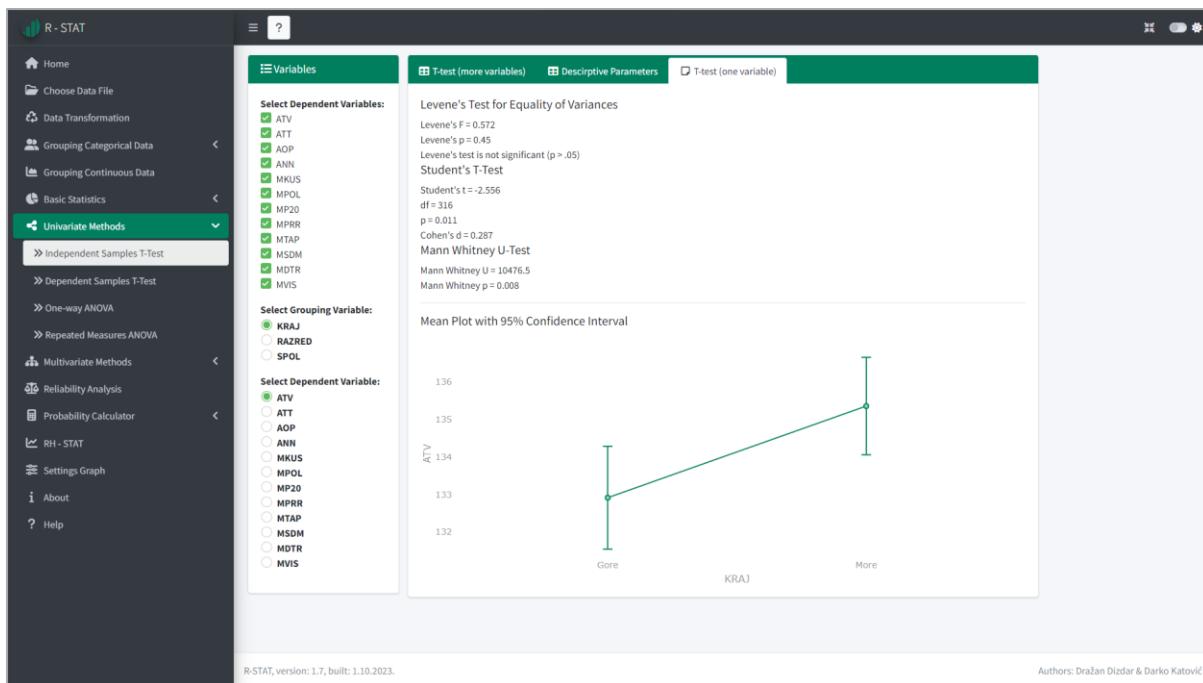


Slika 31: Univariate Methods → Independent Samples T-Test → Descriptive Parameters

U kartici **Descriptive Parameters** prikazuje se tablica (slika 31) u kojoj je:

- **MEAN** – aritmetička sredina.
- **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
- **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.

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



Slika 32: Univariate Methods → Independent Samples T-Test → T-Test (one variable)

U kartici **T-Test (one variable)** prikazuju se sljedeći rezultati za svaku odabranu zavisnu varijablu (slika 32):

- **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%.

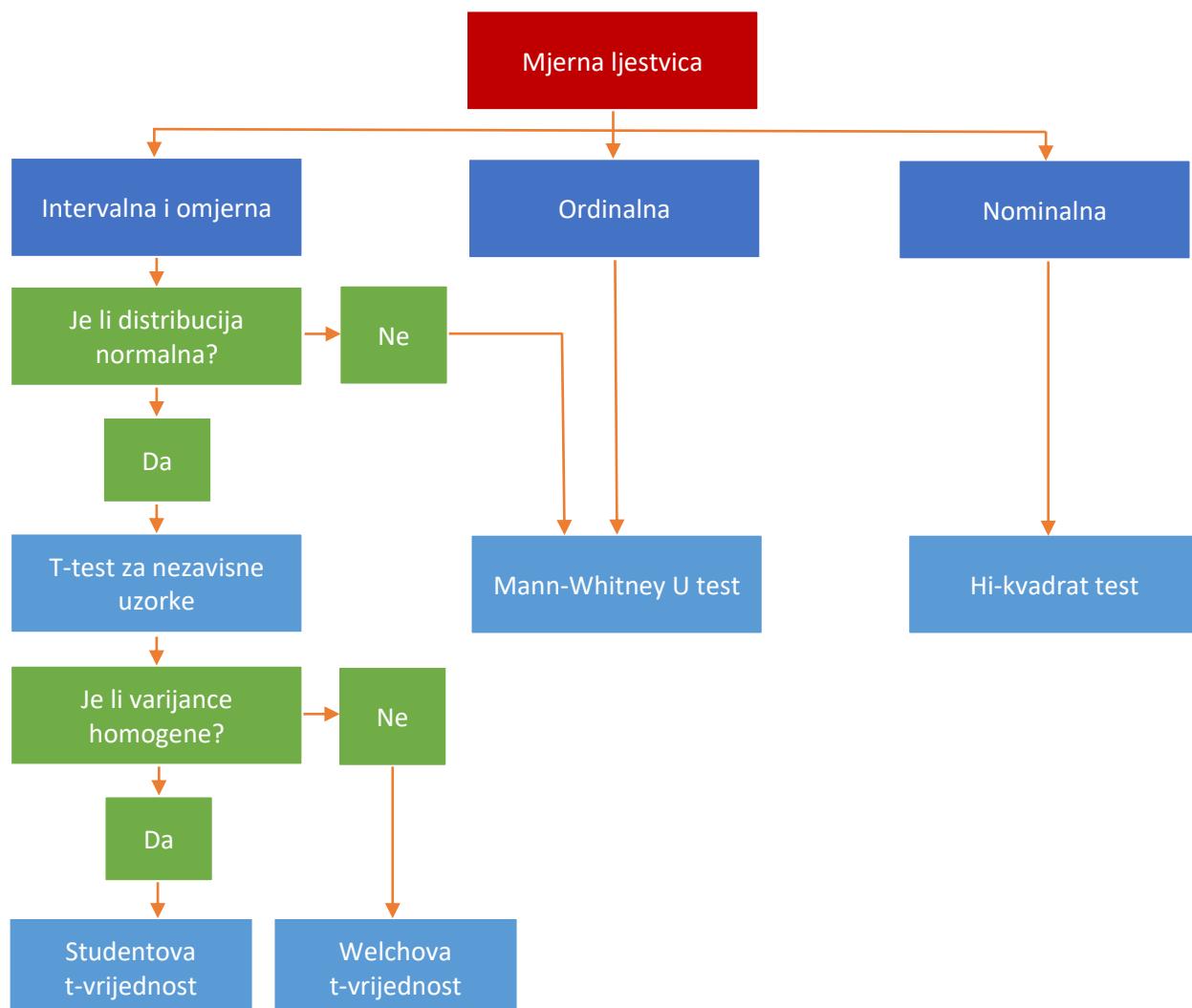
CILJ

- ✓ Uz pogrešku p utvrditi razliku li se statistički značajno aritmetičke sredine dviju grupa entiteta u jednoj kvantitativnoj varijabli.

UVJETI

- ✓ Nezavisnu varijablu čine podaci prikupljeni na dihotomnoj (s dva modaliteta) nominalnoj ili ordinalnoj mjerenoj ljestvici.
- ✓ Zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjerno mjerenoj skali.
- ✓ Podaci su prikupljeni na slučajno odabranim i relativno velikim uzorcima entiteta ($n>30$).
- ✓ Podaci u zavisnoj varijabli imaju normalnu ili Gaussovou distribuciju (ukoliko ova pretpostavka nije ispunjena onda se 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.

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



Dependent Samples T-Test

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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 mjerjenih 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.

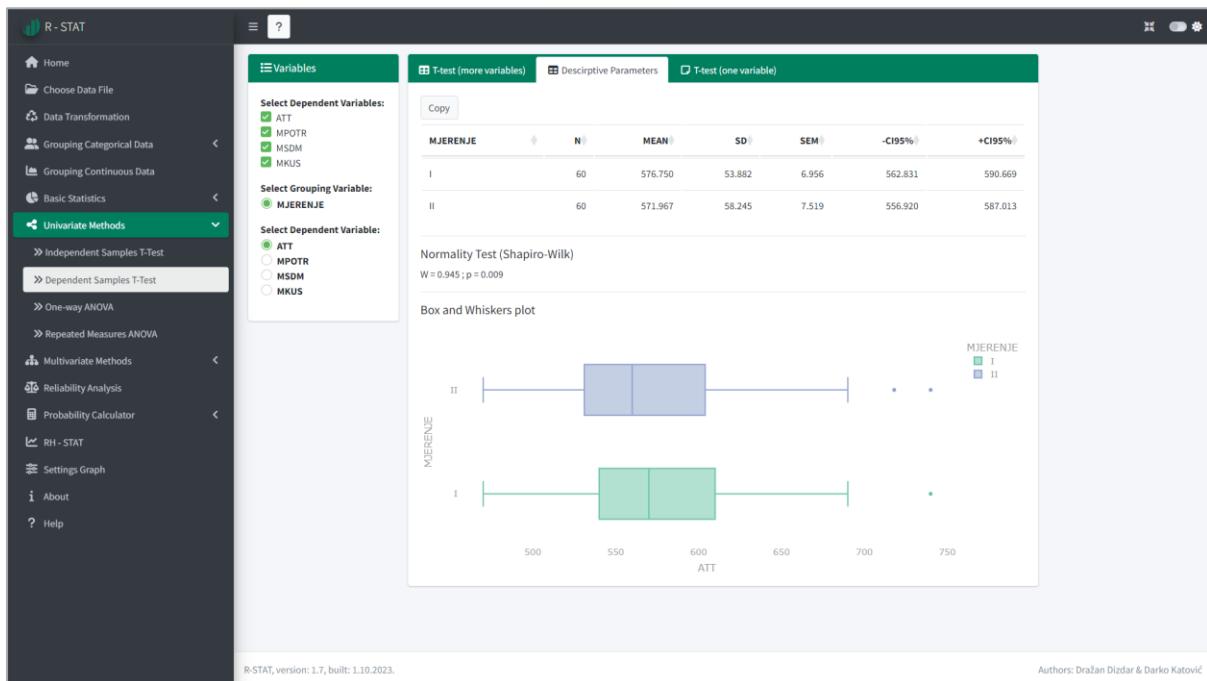
	DIFF	SED	T	DF	P	CI-95%	CI+95%	Cohen's D
ATT	4.783	2.105	2.272	59	0.027	0.571	8.996	0.293
MPOTR	-2.2	0.653	3.371	59	0.001	-3.506	-0.894	0.435
MSDM	2.178	1.565	1.392	59	0.169	-0.953	5.31	0.18
MKUS	56.205	8.299	6.773	59	0	39.599	72.811	0.874

Slika 33: Univariate Methods → Dependent Samples T-Test → T-test (more var.)

U kartici T-Test (more var.) prikazuje se tablica (slika 33) sa sljedećim rezultatima:

- **DIFF** – razlika aritmetičkih sredina prvog i drugog mjerjenja
- **SED** – standardna pogreška razlika aritmetičkih sredina prvog i drugog mjerjenja
- **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 mjerjenja statistički značajna.
- **CI-95%** – donja granica intervala u kojem se s vjerojatnošću 95 % nalazi razlika aritmetičkih sredina populacije prvog i drugog mjerjenja
- **CI+95%** – gornja granica intervala u kojem se s vjerojatnošću 95 % nalazi razlika aritmetičkih sredina populacije prvog i drugog mjerjenja

- **Cohen's d** – Cohenova mjera veličine učinka (effect size).



Slika 34: Univariate Methods → Dependent Samples T-Test → Descriptive Parameters

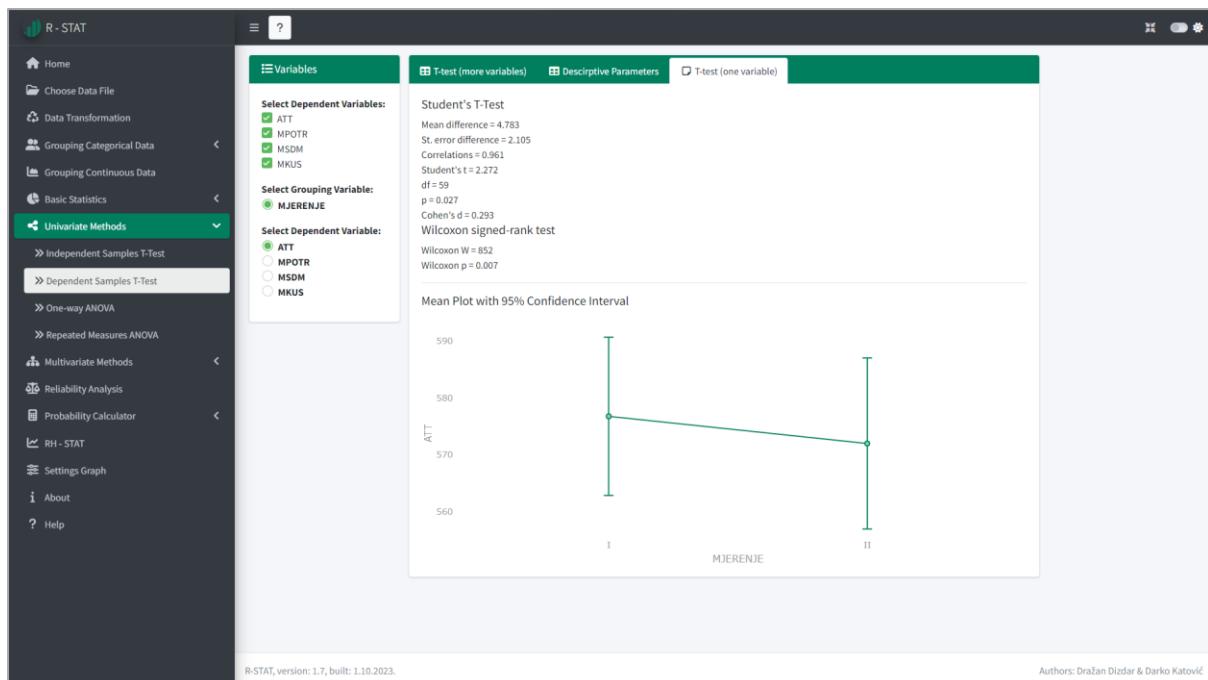
U kartici **Descriptive Parameters** prikazuje se tablica (slika 34) u kojoj je:

- **MEAN** – aritmetička sredina
- **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.

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

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



Slika 35: Univariate Methods → Dependent Samples T-Test → T-Test (one var.)

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 mjerjenja
- **Correlations** – korelacija između prvog i drugog mjerjenja
- **Student's t** – Studentova t- vrijednost koja pokazuje koliko je puta razlika između aritmetičkih sredina veća od standardne pogreške razlika
- **df** – broj stupnjava slobode ($n - 1$, gdje je n broj entiteta)
- **p** – pogreška s kojom tvrdimo da je razlika između aritmetičkih sredina prvog i drugog mjerjenja 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 mjerjenja 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 mjerjenja uz pogrešku do 5%.

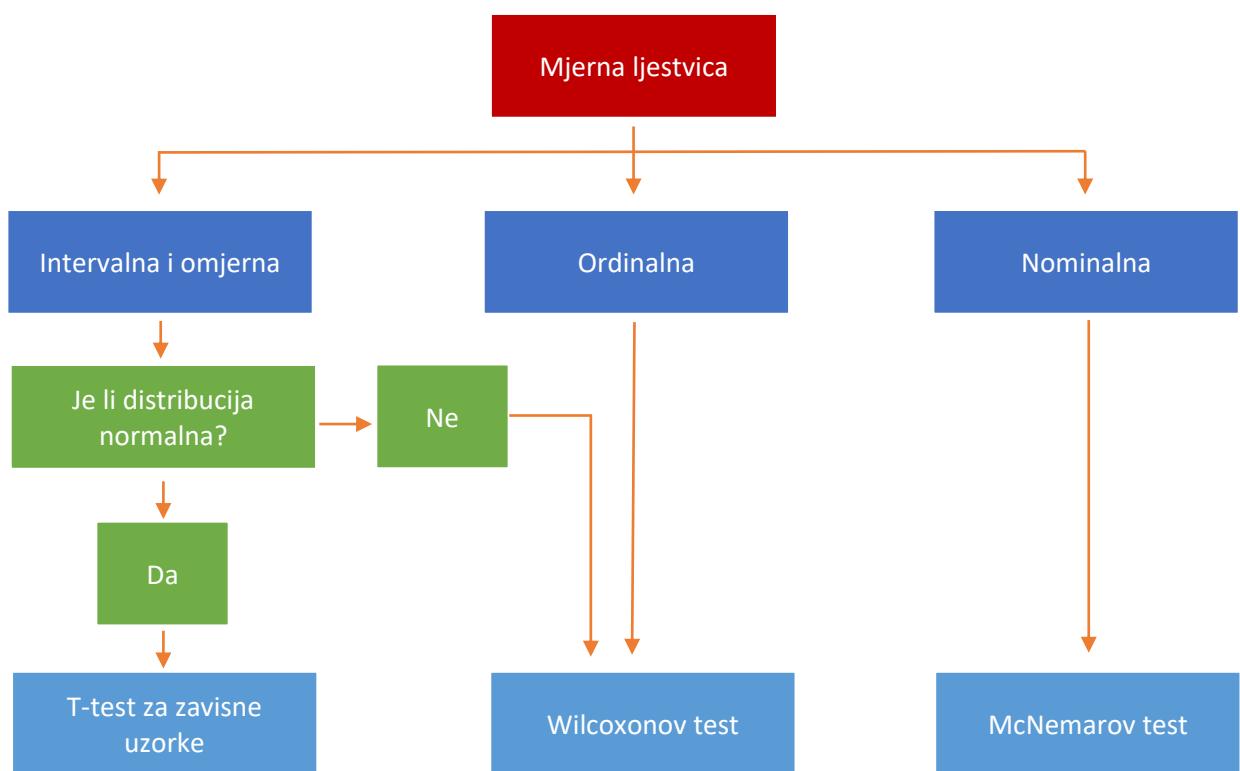
CILJ

- ✓ Uz pogrešku p utvrditi razliku li se statistički značajno aritmetičke sredine dvaju mjerenja u jednoj kvantitativnoj varijabli.

UVJETI

- ✓ Nezavisnu varijablu čine upareni (zavisni) podaci (redoslijed entiteta prvog i drugog mjerenja mora biti isti).
- ✓ Zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjernoj mjerenoj skali.
- ✓ 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.

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



One Way ANOVA

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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.

The screenshot shows the R-STAT application window. On the left, the navigation menu is open, with 'Univariate Methods' selected under 'One-way ANOVA'. In the center, the 'Variables' panel is displayed. Under 'Select Dependent Variables', 'ATV' is checked. Under 'Select Grouping Variable', 'KRAJ' is checked. The main area shows a table titled 'ANOVA (more variable)' with the following data:

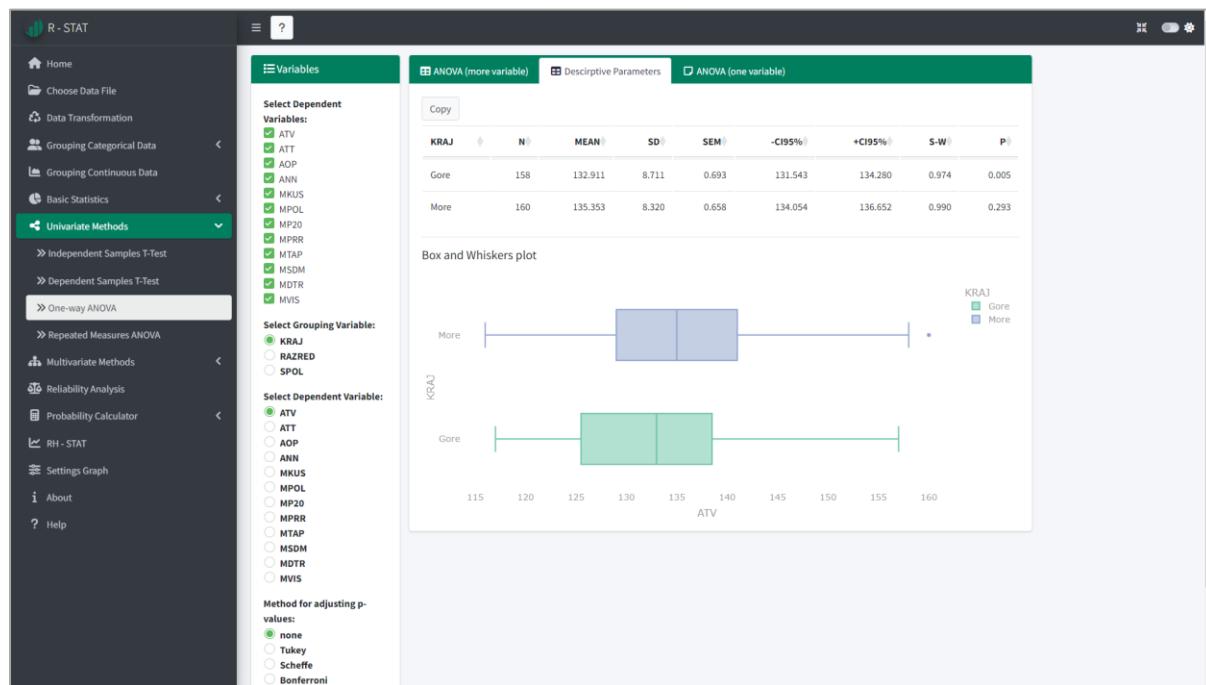
	Fisher's F	DF1	DF2	Fisher's p	Levene's F	Levene's p
ATV	6.534	1	316	0.011	0.572	0.45
ATT	4.294	1	316	0.039	0.542	0.462
AOP	0.83	1	316	0.363	0.204	0.652
ANN	0.002	1	316	0.961	4.163	0.042
MKUS	0.181	1	316	0.671	0.2	0.655
MPOL	1.849	1	316	0.175	2.194	0.14
MP20	1.265	1	316	0.262	0.761	0.384
MPRR	42.411	1	316	0	0.42	0.517
MTAP	1.577	1	316	0.21	0.002	0.961
MSDM	29.159	1	316	0	1.596	0.207
MDTR	2.078	1	316	0.15	0.463	0.497
MVIS	0.535	1	316	0.465	4.1	0.044

Slika 36: Univariate Methods → One Way ANOVA → ANOVA (more variable)

U kartici **ANOVA (more variable)** prikazuju se tablica sa sljedećim rezultatima za svaku odabranu zavisnu varijablu (slika 36):

- **Fisher's F** - vrijednost koja pokazuje omjer varijance između i unutar grupa
- **DF1 i DF2** – broj stupnjava 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.

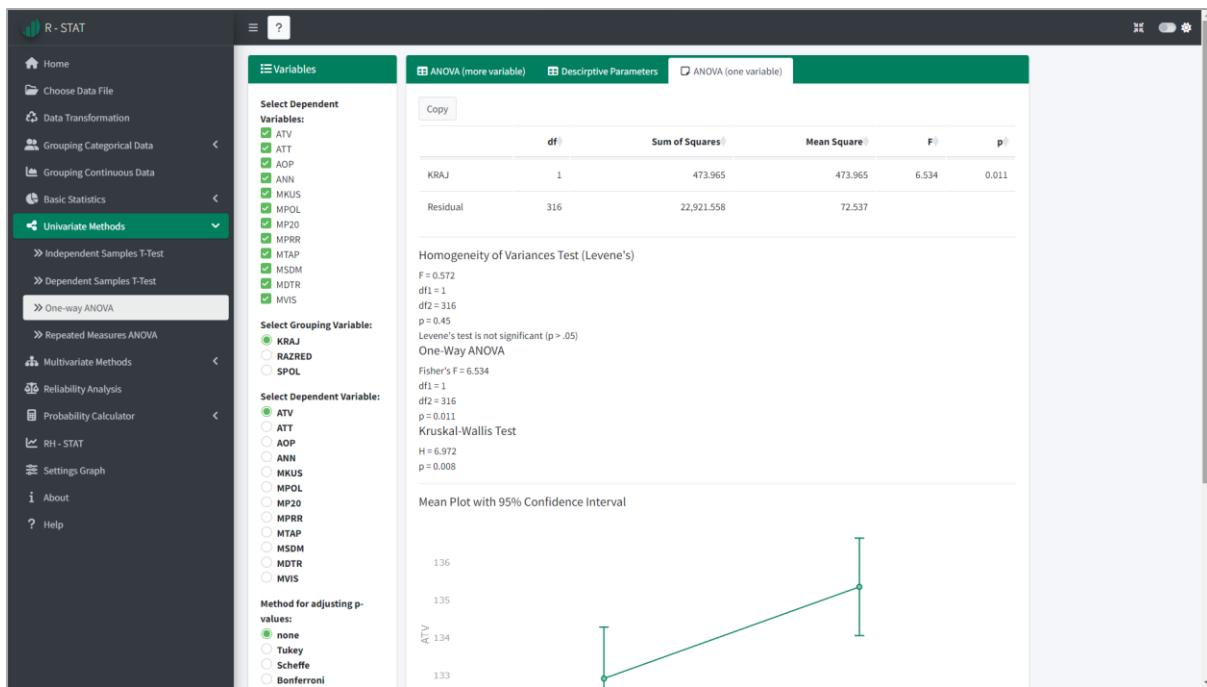


Slika 37: Univariate Methods → One Way ANOVA → Descriptive Parameters

U kartici **Descriptive Parameters** prikazuje se tablica (slika 37) u kojoj je:

- **N** – broj entiteta
- **MEAN** – aritmetička sredina
- **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
- **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.

Osim toga, aplikacija omogućava grafičko prikazivanje podataka pomoću **Box and Whiskers plota** za analizirane grupe.



Slika 38: Univariate Methods → One Way ANOVA → ANOVA (one var.)

U kartici **ANOVA (one variable)** prikazuju se sljedeći rezultati za svaku odabranu zavisnu varijablu (slika 38):

- **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
- **df_1 i df_2** – broj stupnjava 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
- **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%.
- **Post Hoc Test for Multiple Comparisons** – tablica rezultata post hoc testa za višestruko uspoređivanje⁷

⁷Ako 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.).

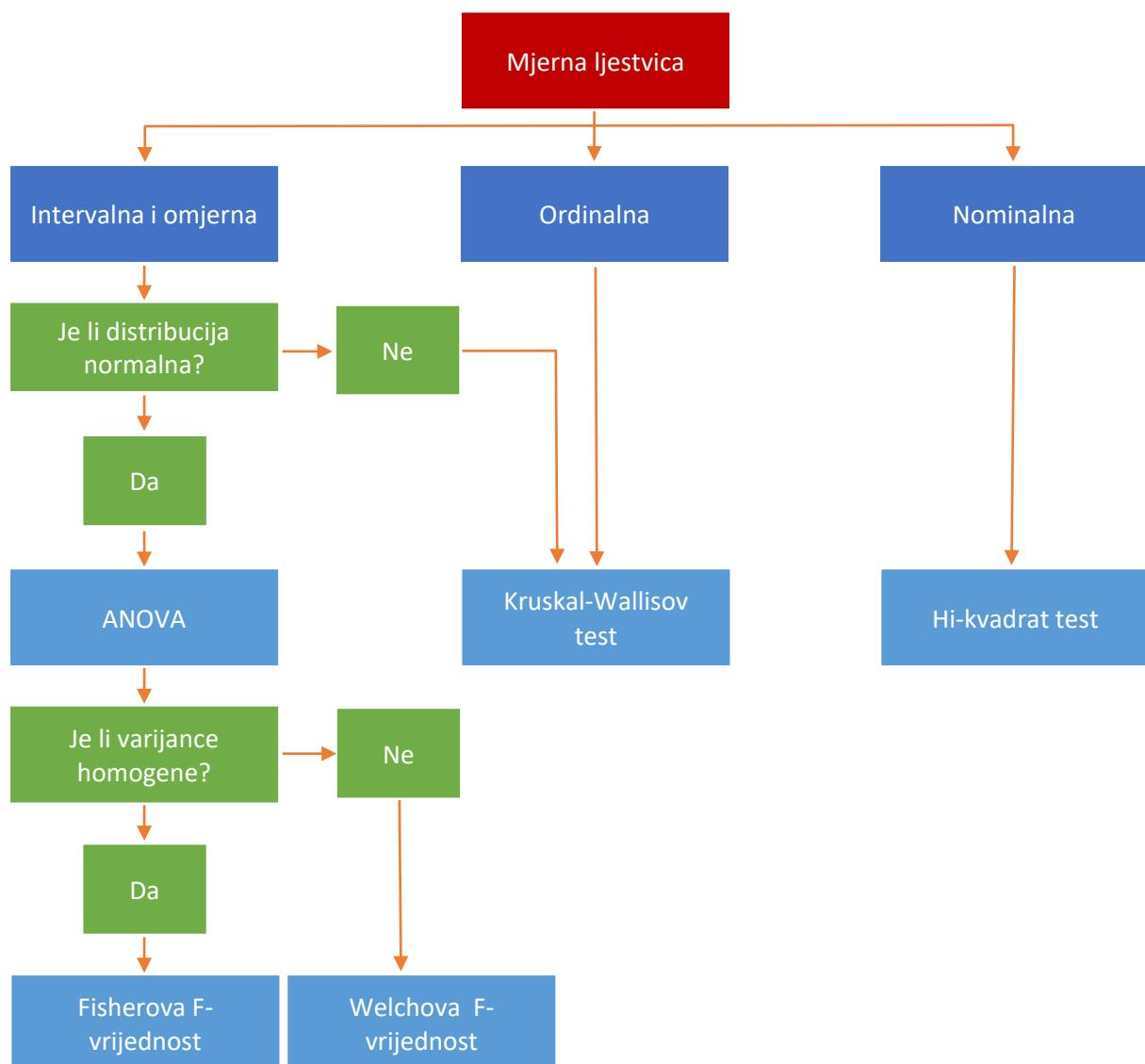
CILJ

- ✓ Uz pogrešku p utvrditi razliku li se statistički značajno aritmetičke sredine dviju ili više grupa entiteta u jednoj kvantitativnoj varijabli.

UVJETI

- ✓ Nezavisnu varijablu čine podaci prikupljeni na nominalnoj ili ordinalnoj mjerne ljestvici s dva ili više modaliteta.
- ✓ Zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjerenoj mjerne skali.
- ✓ 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.

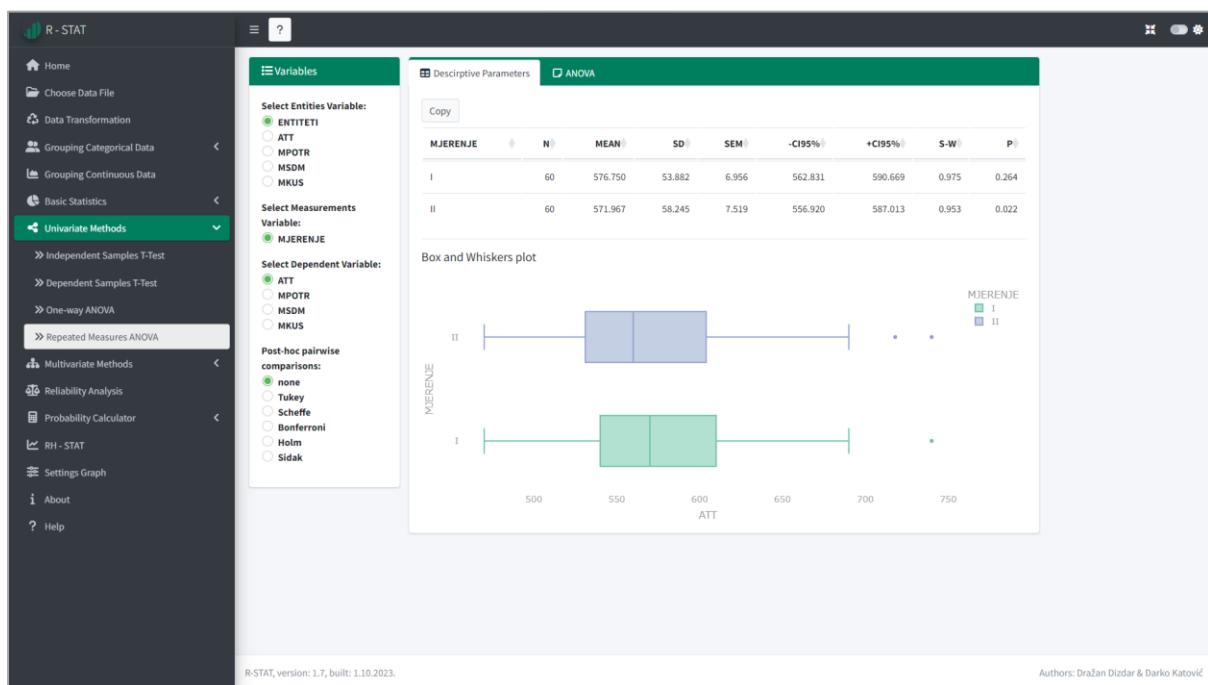
Testiranje razlika između dviju ili više nezavisnih grupa ovisno o vrsti i distribuciji varijabli.



Repeated Measures ANOVA

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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 mjerjenih u dvije vremenske točke (**MJERENJE**) u 4 kvantitativne varijable.

Odabirom opcije **Univariate Methods** glavnog izbornika te opcije **Repeated Measures ANOVA** za odabrane varijable (**Select Dependent Variables** i **Select Grouping Variable**) prikazuju se rezultati univariatne analize varijance za ponavljanja mjerjenja.



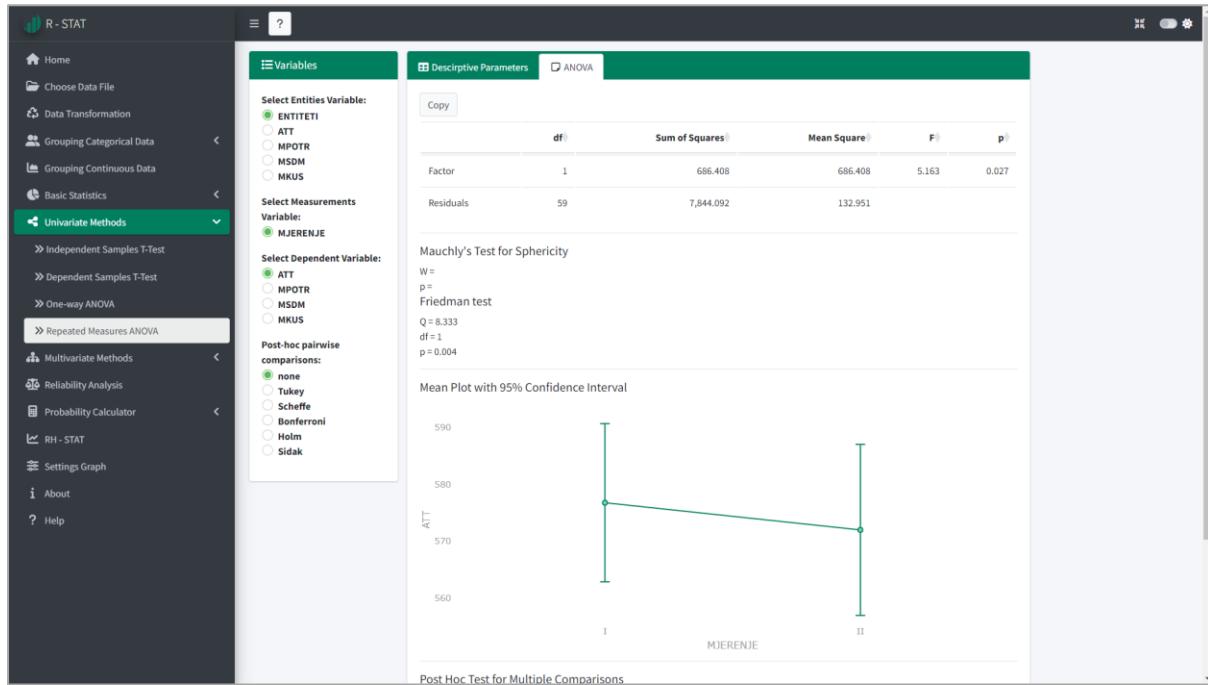
Slika 39: Univariate Methods → Repeated Measures ANOVA → Descriptive Parameters

U kartici **Descriptive Parameters** prikazuje se tablica (slika 39) u kojoj je:

- **N** – broj entiteta
- **MEAN** – aritmetička sredina
- **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
- **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.

Osim toga, aplikacija omogućava grafičko prikazivanje podataka pomoću **Box and Whiskers plota** za analizirane grupe.

U kartici **ANOVA** prikazuju se rezultati analize za svaku odabranu zavisnu varijablu (slika 40):



Slika 40: Univariate Methods → Repeated Measures ANOVA → ANOVA

Multivariate Methods

Regression Analysis

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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.

The screenshot shows the R-STAT application window. On the left, there is a sidebar with various menu items under 'Multivariate Methods' (e.g., Regression Analysis (simple), Regression Analysis (multiple), Factor Analysis, Canonical Analysis, Discriminant Analysis, Cluster Analysis, Reliability Analysis, Probability Calculator). The main panel is titled 'Variables' and contains two sections: 'Select Independent Variables:' (with checkboxes for VISI, TEZI, OBPO, NABN, TAPI, POLI, SDAL, POTR, PRR, IZVI, TR6M) and 'Select Dependent Variable:' (with checkboxes for VISI, TEZI, OBPO, NABN, TAPI, POLI, SDAL, POTR, PRR, IZVI, TR6M, where TR6M is selected). At the top of the main panel, there are tabs for 'Reg. Results' (selected), 'Graph (P)', 'Graph (Beta, Part R, R)', and 'Reg. Scores'. Below the tabs, the output text reads: 'R0 = 0.83 ; R02 = 0.681 ; SEE = 133.48 ; F-value = 18.972 ; p-value = 0 ; Autocorrelation = 0.081 ; Durbin-Watson d = 1.83 ; p = 0.332'. A 'Copy' button is present. Below the text is a table with columns: B, SE(B), Beta, Part_R, R, P, Tolerance, t, and p. The table rows represent the coefficients for each independent variable. At the bottom of the main panel, it says 'R-STAT, version: 1.7, built: 1.10.2023.' and 'Authors: Dražan Dizdar & Darko Katović'.

	B	SE(B)	Beta	Part_R	R	P	Tolerance	t	p
B0	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
NABN	-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

Slika 41: Multivariate Methods → Regression Analysis (multiple) → Reg. Results

U kartici **Reg. Results** (slika 41) prikazuju se sljedeći pokazatelji za odabranu zavisnu i nezavisne varijable:

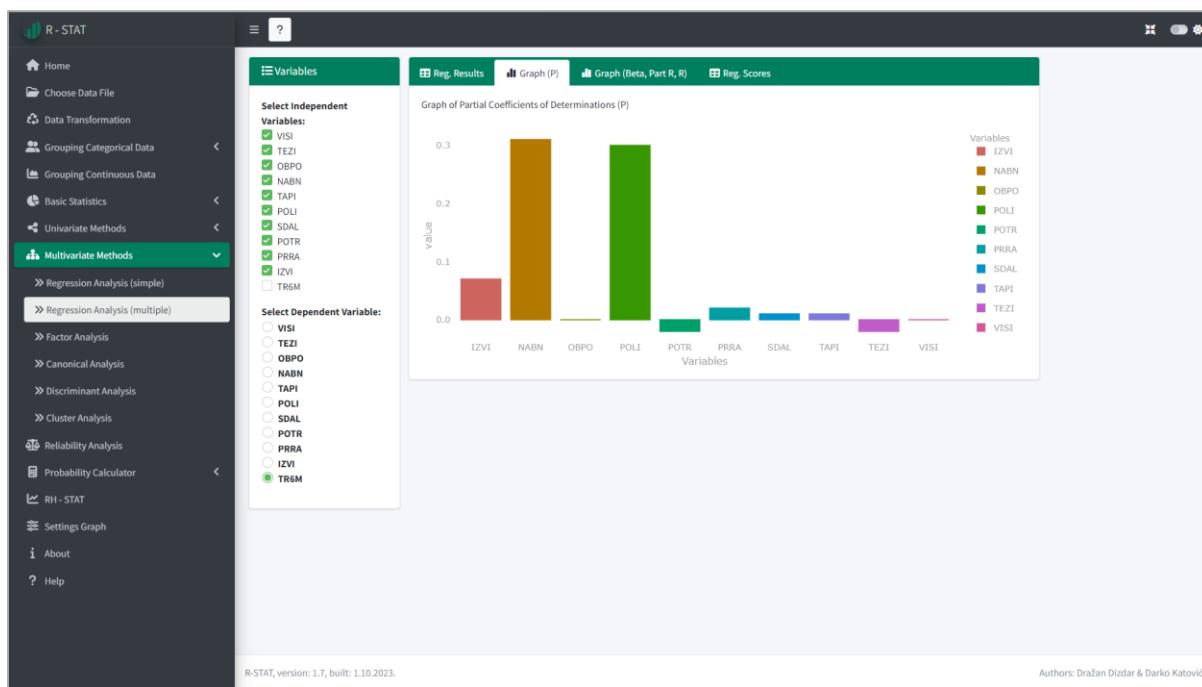
- **RO** – koeficijent multiple korelacija
- **RO2** – koeficijent determinacije multiple korelacija
- **SEE** – standardna pogreška prognoze
- **F-value** – F-vrijednost kojom se izračunava pri testiranju statističke značajnosti multiple korelacijske
- **p-value** – pogreška kojom tvrdimo da je multipla korelacija statistički značajna
- **Autocorrelation** – autokorelacija prvog reda za n-1 parova rezidualnih vrijednosti
- **Durbin-Watson** – vrijednost Durbin-Watsonovog testa statističke značajnosti autokorelacijske prvog reda

- **p** – pogreške s kojima tvrdimo da je autokorelacija prvog reda rezidualnih vrijednosti statistički značajna.

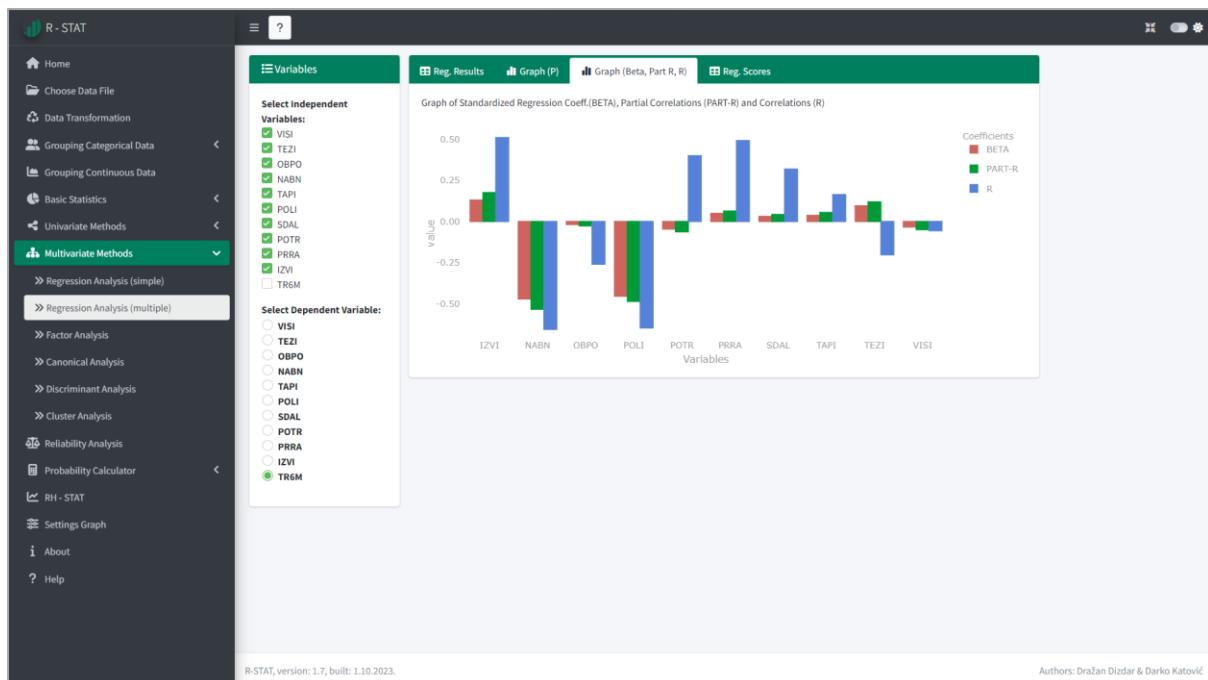
te tablica u kojoj su prikazani sljedeći pokazatelji:

- **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 objašnjrenom 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 (P)** (slika 42) i standardiziranih regresijskih koeficijenata, parcijalnih korelacija i korelacija **Graph (Beta), Partial R, R** (slika 43).



Slika 42: Multivariate Methods → Regression Analysis (multiple) → Graph (P)



Slika 43: Multivariate Methods → Regression Analysis (mlultiple) → Graph (Beta, Partial R, R)

U kartici **Reg. Scores** prikazuje se tablica s izmjeranim, prognoziranim i rezidualnim vrijednostima zavisne varijable (slika 44).

	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	1,032.395	9.105
16	1,330.000	1,155.324	174.676
17	1,333.000	1,351.446	-18.446

Slika 44: Multivariate Methods → Regression Analysis (mlultiple) → Reg. Scores

CILJ

- ✓ Utvrditi funkcionalnu vezu između jedne zavisne (kriterijske) varijable i dvije ili više nezavisnih (prediktorskih) varijabli.

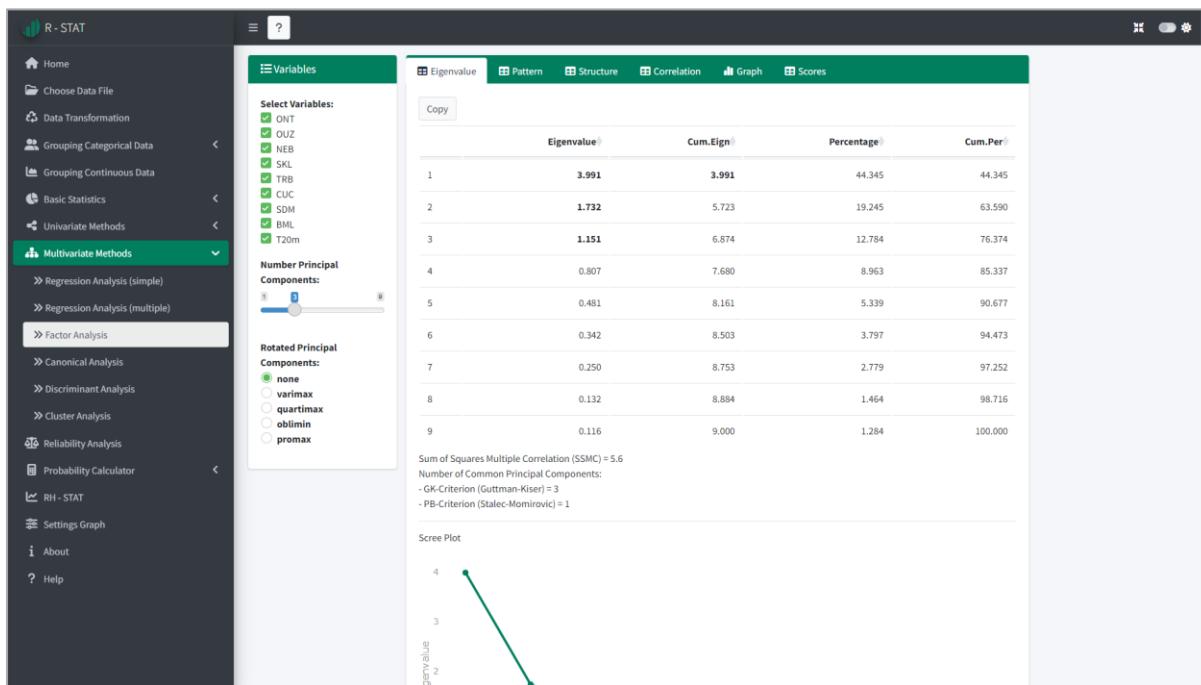
UVJETI

- ✓ 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 linearном odnosu.
- ✓ Varijable imaju normalnu ili Gaussovou 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 Gaussovou distribuciju.

Factor Analysis

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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.



Slika 45: Multivariate Methods → Factor Analysis → Eigenvalue

Određivanje broja značajnih glavnih moguće je putem **GK** i **PB** kriterija te **Scree Plot**, a za dobivanje jednostavne strukture faktora koriste se ortogonalne (**varimax** i **quartimax**) i neortogonalne (**oblimin** i **promax**) rotacije.

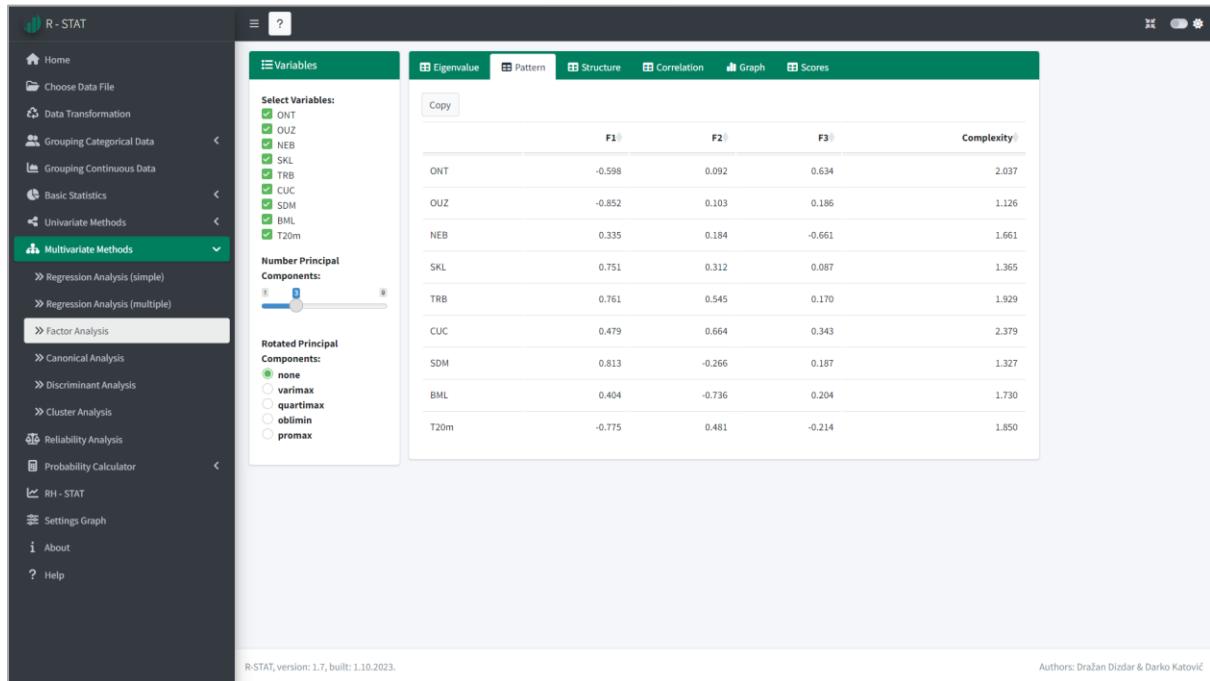
U kartici Eigenvalue prikazuje se tablica sa sljedećim rezultatima:

- **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.
-



Slika 46: Multivariate Methods → Factor Analysis → Pattern

U kartici **Pattern** (slika 46) nalazi se matrica sklopa, odnosno paralelnih projekcija manifestnih varijabli na faktore te Hoffmanov index kompleksiteta (**Complexity**) svake manifestne varijable⁸. 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

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

⁸ Hofmann, R. J. (1978). Complexity and simplicity as objective indices descriptive of factor solutions. *Multivariate Behavioral Research*, 13(2), 247–250

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

U kartici **Structure** (slika 47) 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**).

The screenshot shows the R-STAT software interface. On the left, there is a sidebar with various statistical methods: Home, Choose Data File, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods, Multivariate Methods (selected), Regression Analysis (simple), Regression Analysis (multiple), Factor Analysis (selected), Canonical Analysis, Discriminant Analysis, Cluster Analysis, Reliability Analysis, Probability Calculator, RH-STAT, Settings Graph, About, and Help. The version information "R-STAT, version: 1.7, built: 1.10.2023." and authors "Authors: Dražan Dizdar & Darko Katović" are at the bottom.

The main window has tabs: Eigenvalue, Pattern, Structure (selected), Correlation, Graph, and Scores. The Structure tab displays a table of factor loadings and communality values. The columns are F1, F2, F3, and Communality. The rows are variables: ONT, OUZ, NEB, SKL, TRB, CUC, SDM, BML, and T20m. The data is as follows:

	F1	F2	F3	Communality
ONT	-0.598	0.092	0.634	0.768
OUZ	-0.852	0.103	0.186	0.770
NEB	0.335	0.184	-0.661	0.584
SKL	0.751	0.312	0.087	0.669
TRB	0.761	0.545	0.170	0.904
CUC	0.479	0.664	0.343	0.789
SDM	0.813	-0.266	0.187	0.766
BML	0.404	-0.736	0.204	0.747
T20m	-0.775	0.481	-0.214	0.877

Slika 47: Multivariate Methods → Factor Analysis → Structure

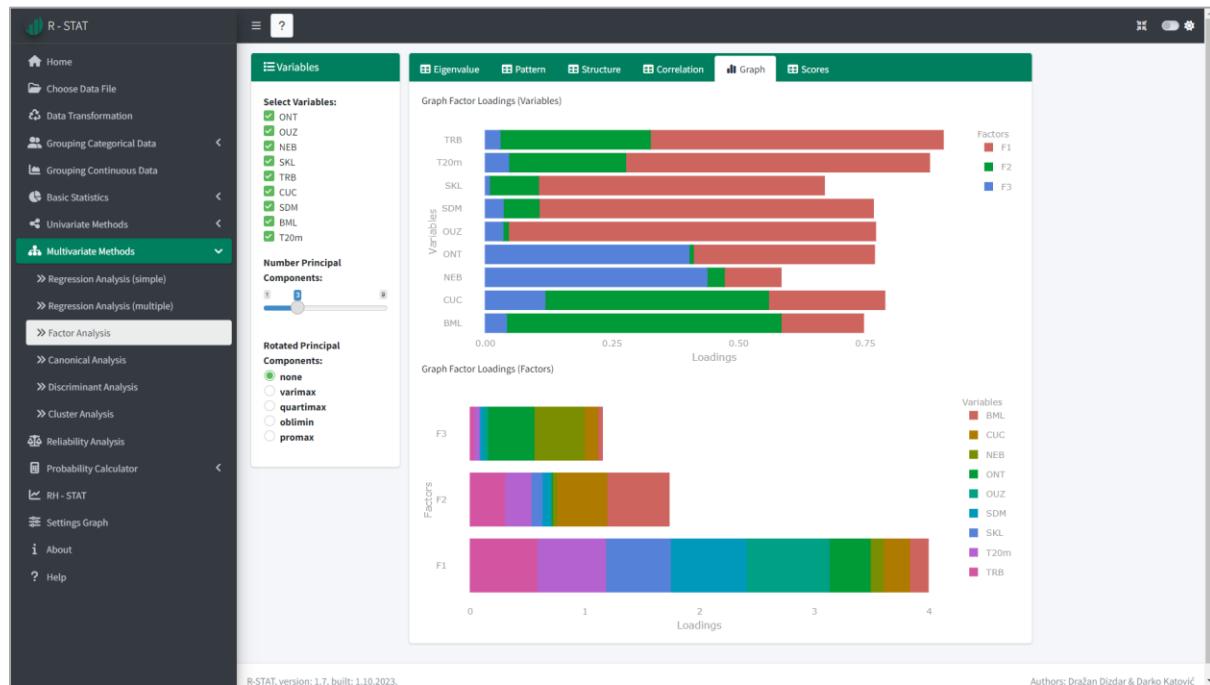
U kartici **Correlation** (slika 48) nalazi se matrica korelacija između faktora.

The screenshot shows the R-STAT software interface, identical to Slika 47 but with the Correlation tab selected instead of Structure. The main window has tabs: Eigenvalue, Pattern, Structure, Correlation (selected), Graph, and Scores. The Correlation tab displays a correlation matrix for the three factors F1, F2, and F3. The diagonal elements are 1.000, and the off-diagonal elements are 0.000. The data is as follows:

	F1	F2	F3
F1	1.000	0.000	-0.000
F2	0.000	1.000	0.000
F3	-0.000	0.000	1.000

Slika 48: Multivariate Methods → Factor Analysis → Correlation

U kartici **Graph** (slika 49) nalazi se grafikon strukture manifestnih varijabli (**Factor Loadings (Variables)**), grafikon strukture faktora (**Graph Factor Loadings (Factors)**).



Slika 49: Multivariate Methods → Factor Analysis → Graph

U kartici **Scores** (slika 50) nalazi se tablica s rezultatima entiteta u faktorima.

The screenshot shows the R-STAT interface with the 'Multivariate Methods' section selected. In the 'Factor Analysis' sub-section, the 'Scores' tab is active. On the left, there are sections for 'Select Variables' (checkboxes for OÜT, OÜZ, NEB, SKL, TRB, CUC, SDM, BML, T20m) and 'Number Principal Components' (set to 3). Below these are 'Rotated Principal Components' options: none, varimax, quartimax, oblimin, or promax. The main area displays a table of scores for entities (Marko, Mate, Sime, Mile, Jure, Ante, Ivo, Stipe, Tin, Dino, Darko, Stanko, Branko, Zarko, Sanimir, Toni) across three factors (F1, F2, F3). The table includes a 'Copy' button at the top. At the bottom, it says 'R-STAT, version: 1.7, built: 1.10.2023.' and 'Authors: Dražan Dizdar & Darko Katović'.

	F1	F2	F3
Marko	-2.390	-0.105	1.868
Mate	-1.853	1.469	-2.285
Sime	-0.133	0.035	-0.075
Mile	-0.188	-0.902	-0.200
Jure	-1.140	-0.989	0.809
Ante	-0.147	-0.036	-0.618
Ivo	-0.875	0.177	0.106
Stipe	-1.461	-0.683	0.495
Tin	-0.924	1.196	-0.495
Dino	-0.565	-0.696	0.171
Darko	-0.442	-0.827	-0.305
Stanko	-0.805	-0.393	-0.038
Branko	-0.261	0.022	1.473
Zarko	1.772	-0.461	0.819
Sanimir	-0.757	-1.280	0.355
Toni	0.426	-0.904	0.482

Slika 50: Multivariate Methods → Factor Analysis → Scores

CILJ

- ✓ Veći broj manifestnih varijabli kondenzirati na manji broj latentnih dimenzija.

UVJETI

- ✓ 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 linearном odnosu.
- ✓ Varijable imaju normalnu ili Gaussovu distribuciju.

Canonical Analysis

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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.

CF	Rc	Chi-sq.	df	p-level
CF1	0.832	125.654	18	0
CF2	0.399	16.129	8	0.041

Slika 51: Multivariate Methods → Canonical Analysis → Canonical Factor

U kartici **Canonical Factor** (slika 51) prikazuju se:

- **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 korelacijske statistički značajni.

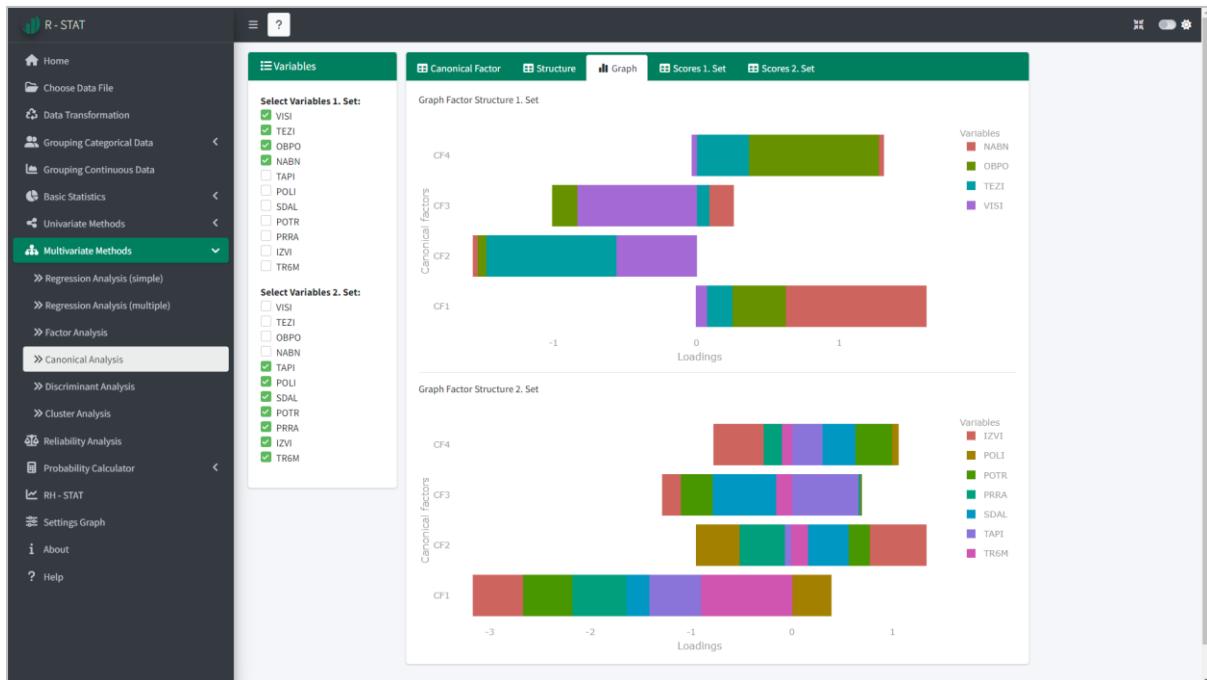
U kartici **Structure** (slika 52) nalaze se korelacija manifestnih varijabli prvog skupa s kanoničkim faktorima prvog skupa (**Factor Structure 1. Set**) i korelacija manifestnih varijabli drugog skupa s kanoničkim faktorima drugog skupa (**Factor Structure 2. Set**).

	CF1	CF2	CF3	CF4
VISI	0.071	-0.555	-0.828	-0.029
TEZI	0.177	-0.91	0.087	0.364
OBPO	0.374	-0.06	-0.177	0.908
NABN	0.984	-0.035	0.171	0.036

	CF1	CF2	CF3	CF4
TAPI	-0.514	-0.063	0.659	0.3
POLI	0.389	-0.434	0.012	0.066
SDAL	-0.228	0.403	-0.631	0.326
POTR	-0.49	0.21	-0.315	0.364
PRRA	-0.537	-0.449	0.021	-0.181
IZVI	-0.499	0.564	-0.187	-0.499
TR6M	-0.894	0.156	-0.149	-0.092

Slika 52: Multivariate Methods → Canonical Analysis → Structure

U kartici **Graph** (slika 53) nalaze se grafikoni strukture kanoničkih faktora prvog (**Graf Factor Structure 1. Set**) i drugog prvog (**Graf Factor Structure 2. Set**) skupa.



Slika 53: Multivariate Methods → Canonical Analysis → Graph

U karticama **Scores 1. Set** i **Scores 2. Set** nalaze se tablice s rezultatima entiteta u kanoničkim faktorima prvog i drugug skupa (slika 54).

	CF1	CF2	CF3	CF4
1	-0.609	-0.454	-1.804	1.419
2	-1.144	0.707	0.412	-0.524
3	0.097	-1.732	0.288	-0.331
4	-0.026	-1.268	0.858	0.753
5	1.058	-1.433	0.922	-0.583
6	-0.468	-1.091	0.757	-0.854
7	1.652	0.674	0.621	1.035
8	-1.388	0.986	1.432	1.576
9	-2.08	0.813	1.17	0.017
10	0.059	-0.138	-1.025	1.316
11	0.819	-1.27	-0.471	-1.636
12	1.022	0.891	1.366	0.564
13	-0.336	0.767	0.022	0.853
14	0.727	-1.119	-0.488	1.081
15	-0.638	-0.059	0.421	-0.782
16	-0.578	0.249	-0.209	0.873

Slika 54: Multivariate Methods → Canonical Analysis → Scores 1. Set

CILJ

- ✓ Utvrditi relacije između dva skupa varijabli.

UVJETI

- ✓ 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 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 linearном odnosu.
- ✓ Varijable imaju normalnu ili Gaussovou distribuciju.

Discriminant Analysis

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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.

The screenshot shows the R-STAT interface. On the left, a sidebar menu is open under the 'Multivariate Methods' section, with 'Discriminant Analysis' selected. The main panel displays the 'DF' tab of the Discriminant Analysis results. The 'Select Dependent Variables:' section contains checked boxes for: ATV, ATT, AOP, ANN, MKUS, MPOL, MP20, MPRR, MTAP, MSDM, MDTR, and MVIS. The 'Select Grouping Variable:' section contains a checked box for RAZRED and unchecked boxes for KRAJ and SPOL. Below these sections is a table with the following data:

	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

Slika 55: Multivariate Methods → Discriminant Analysis → DF

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** (slika 55) nalaze se korelacije zavisnih varijabli i diskriminacijskih funkcija.

	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
MKUS	-0.492	0.087	0.28
MPOL	-0.533	0.531	0.175
MP20	0.157	-0.138	0.286
MPRR	0.506	0.244	-0.568
MTAP	0.641	-0.032	-0.16
MSDM	0.558	0.017	0.084
MDTR	0.443	-0.25	0.293
MVIS	0.429	0.345	0.457

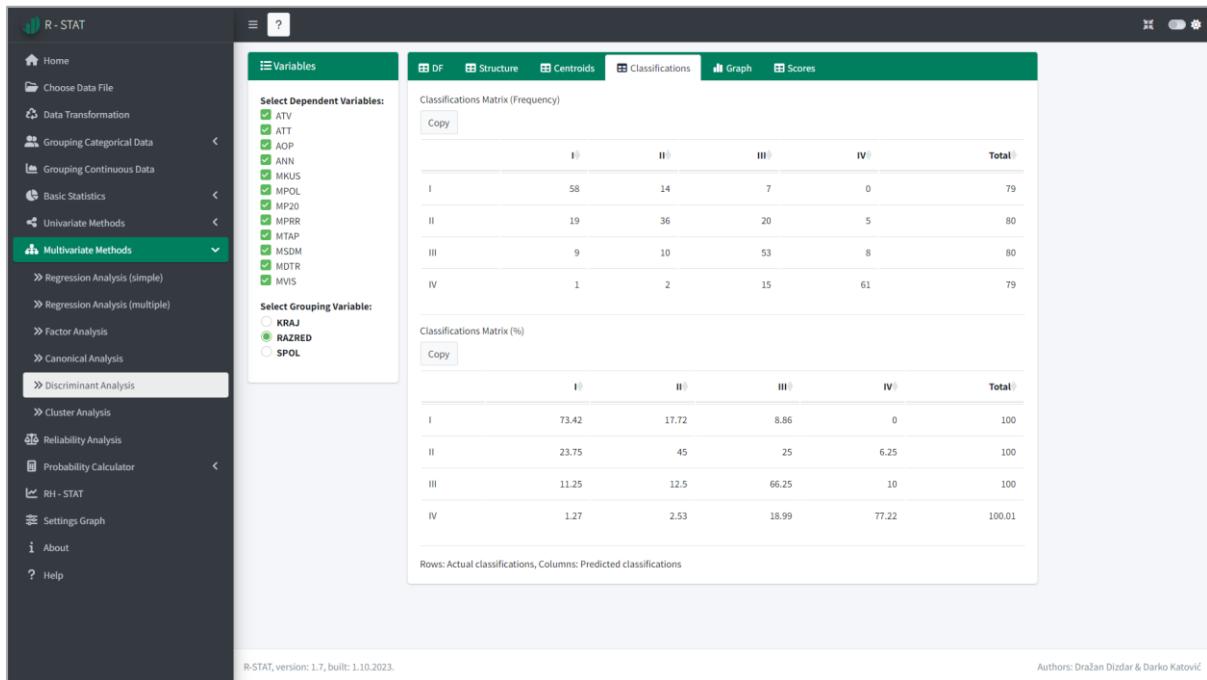
Slika 55: Multivariate Methods → Discriminant Analysis → Structure

U kartici **Centroids** (slika 56) nalaze se aritmetičke sredine analiziranih grupa u diskriminacijskim funkcijama.

RAZRED	FREQ.	DF1	DF2	DF3
I	79	-1.848	-0.012	0.312
II	80	-0.634	0.434	-0.339
III	80	0.356	-0.651	-0.163
IV	79	2.129	0.231	0.196

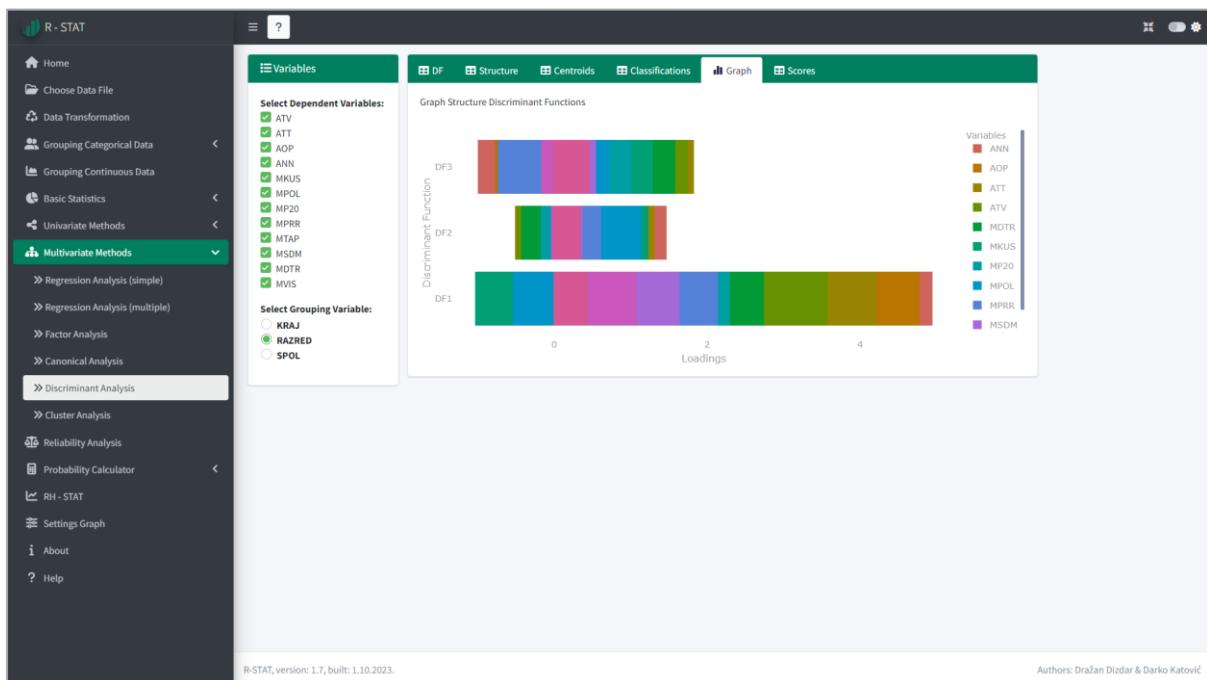
Slika 56: Multivariate Methods → Discriminant Analysis → Centroids

U kartici **Clasifications** (slika 57) nalaze se tablice koje pokazuju broj i postotak ispravno i neispravno klasificiranih entiteta na temelju diskriminacijskih funkcija.



Slika 57: Multivariate Methods → Discriminant Analysis → Clasifications

U kartici **Graph** (slika 58) nalazi se grafikon strukture diskriminacijskih funkcija, odnosno korelacija manifestnih varijabli s diskriminacijskim funkcijama.



Slika 58: Multivariate Methods → Discriminant Analysis → Graph

U kartici **Scores** (slika 59) nalazi rezultati entiteta u diskriminacijskim funkcijama.

The screenshot shows the R-STAT software interface. The left sidebar has a dark theme with various menu items like Home, Choose Data File, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods, Multivariate Methods (selected), Regression Analysis (simple), Regression Analysis (multiple), Factor Analysis, Canonical Analysis, Discriminant Analysis (selected), Cluster Analysis, Reliability Analysis, Probability Calculator, RH-STAT, Settings Graph, About, and Help. The main window title is "Scores". It displays a table of discriminant analysis scores for 17 entities across three discriminant functions (DF1, DF2, DF3). The dependent variables listed are RAZRED, ATT, AOP, ANN, MKUS, MPOL, MP20, MPRR, MTAP, MSDM, MDTR, and MVIS. The grouping variable selected is RAZRED. The table includes a header row and 17 data rows, with the last row being a summary.

	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
12	I	-2.083	-1.27	-0.311
13	I	-1.832	-0.269	1.047
14	I	-2.669	-0.576	0.485
15	I	-0.377	-0.597	0.569
16	I	-2.073	-0.362	-0.04
17	I	-2.682	0.632	0.315

Slika 59: Multivariate Methods → Discriminant Analysis → Scores

CILJ

- ✓ Utvrditi razlike između dvije ili više grupa entiteta opisanih s dvije ili više varijabli.

UVJETI

- ✓ Nezavisnu varijablu čine podaci prikupljeni na nominalnoj ili ordinalnoj mjerenoj ljestvici s dva ili više modaliteta (npr. muškarci-žene, bek-krilo-centar i sl.).
- ✓ Zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjernoj mjerenoj 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 *outliere*, odnosno ekstremno visoke ili niske rezultate.
- ✓ Varijable su u linearnom odnosu.
- ✓ Varijable imaju normalnu ili Gaussovou distribuciju.

Reliability Analysis

Nakon pokretanja aplikacije **R-STAT** u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose 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 (čestice testa). Primjerice, u datoteci GIM.csv nalaze se podaci 103 entiteta opisanih s 12 varijabli (čestica) dobivenih na 4 testa za procjenu motoričkih sposobnosti (FEDSM, BFTAP, FLPRRA i AGKUS).

The screenshot shows the R-STAT application interface. On the left, a sidebar lists various statistical methods under 'Multivariate Methods'. The main area is titled 'Reliability Analysis Results' and contains three panels: 'Reliability Coefficients', 'Intraclass Correlation', and 'Condensed Data'. The 'Reliability Coefficients' panel displays Cronbach's alpha (0.919), Spearman-Brown alpha (0.922), Kaiser-Caffrey alpha (0.922), and Average interitem correlation (0.799). The 'Intraclass Correlation' panel shows Intraclass Corr. Coeff. (ICC) = 0.788; F-value = 12.177; p - value = 0, with a 95% Conf. Interval for ICC = 0.722 - 0.844. The 'Condensed Data' panel shows a table with 13 rows of data, each containing X-mean, Z-mean, and PC1 values.

	X-mean	Z-mean	PC1
1	125.000	1.237	1.330
2	140.000	2.051	2.206
3	116.667	0.795	0.854
4	59.333	-2.339	-2.515
5	71.667	-1.668	-1.794
6	114.000	0.651	0.699
7	86.667	-0.864	-0.928
8	81.667	-1.175	-1.261
9	99.000	-0.178	-0.192
10	105.000	0.178	0.190
11	102.667	0.060	0.064
12	66.667	-1.923	-2.068
13	75.000	-1.469	-1.580

Slika 60: Multivariate Methods → Reliability Analysis

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

- **Cronbach's alpha** – Cronbachov koeficijent pouzdanosti
- **Spearman-Brown alpha** – Sperman-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)⁹ 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 ocjenjuju svi ocjenjivači (nema generaliziranja na šиру populaciju)

⁹ 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>

dok se drugi broj odnosi na konačnu ocjenu:

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

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 ¹⁰
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	$\frac{MS_R - MS_E}{MS_R}$

¹⁰ MS_R - srednji kvadrat za retke; MS_W - srednji kvadrat za rezidualne izvore varijance; MS_E - srednja vrijednost kvadrat za grešku; MS_C - srednji kvadrat za stupce; n - broj ispitanika; k - broj ocjenjivača.

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 linearom 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.

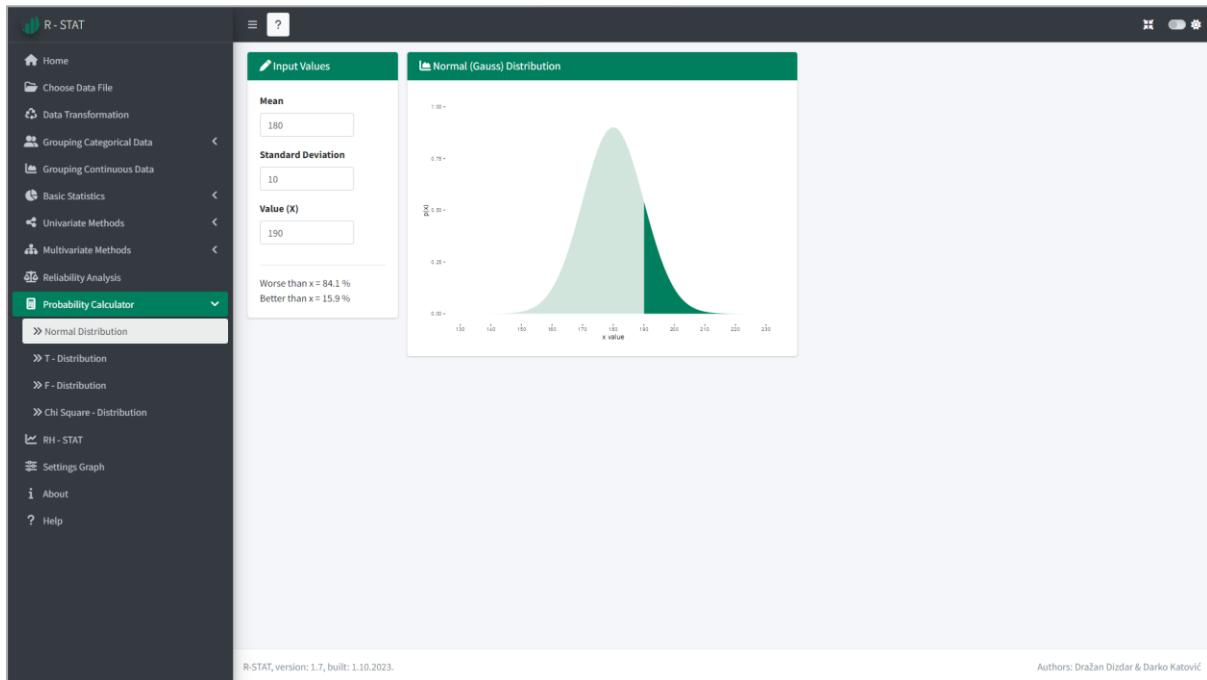
Koeficijenti pouzdanosti koji se utvrđuju metodom interne konzistencije odabiru se s obzirom na način kondenzacije rezultata, odnosno izračunavanje ukupnog rezultata ispitanika na temelju pripadajućih rezultata u česticama testa. Ako se kondenzacija rezultata nekog kompozitnog testa vrši:

- jednostavnom linearom kombinacijom originalnih rezultata (zbroj ili aritmetička sredina), onda je mjera pouzdanosti Cronbachova α
- jednostavnom linearom kombinacijom standardiziranih rezultata (zbroj ili aritmetička sredina), onda je mjera pouzanosti Spearman-Brownova α
- prvom glavnom komponentom, onda je mjera pouzdanosti Kaiser-Caffreyeva α .

Probability Calculator

Normal Distribution

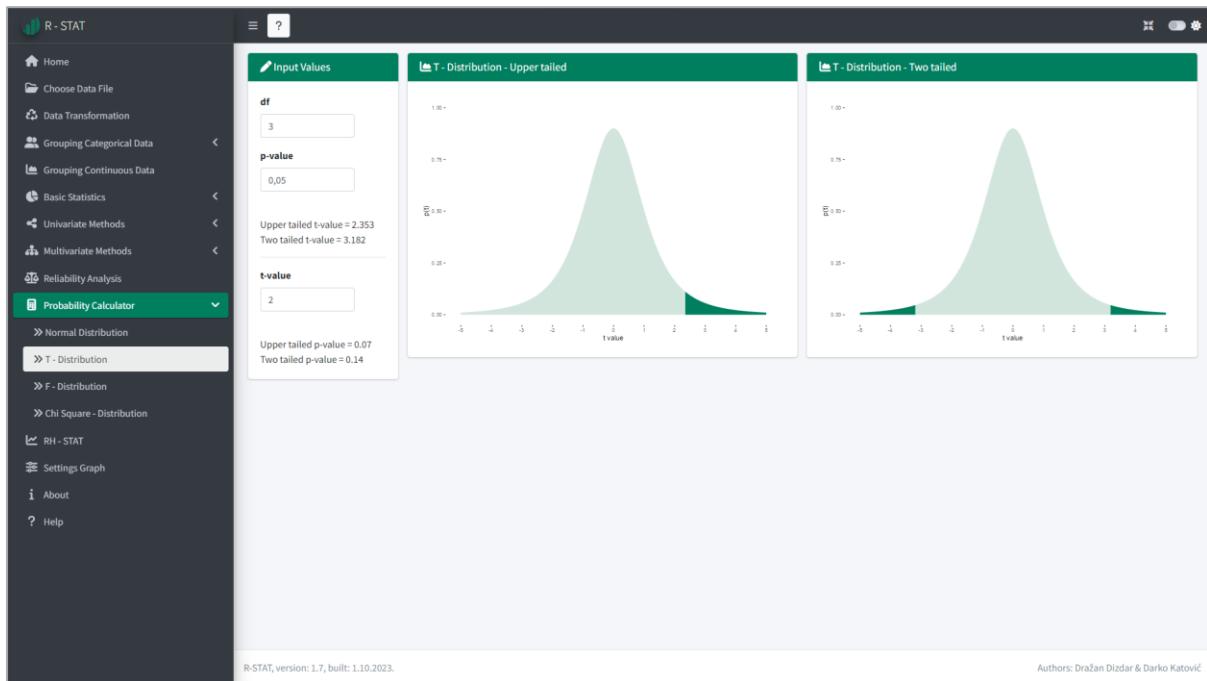
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **Normal Distribution** omogućava računanje postotka boljih (**Better than x**) i lošijih (**Worse than x**) entiteta od određenog rezultata (**X value**) u populaciji entiteta čije obilježje (varijabla) ima normalnu ili Gaussovu distribuciju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



Slika 61: Probability Calculator → Normal Distribution

T – Distribution

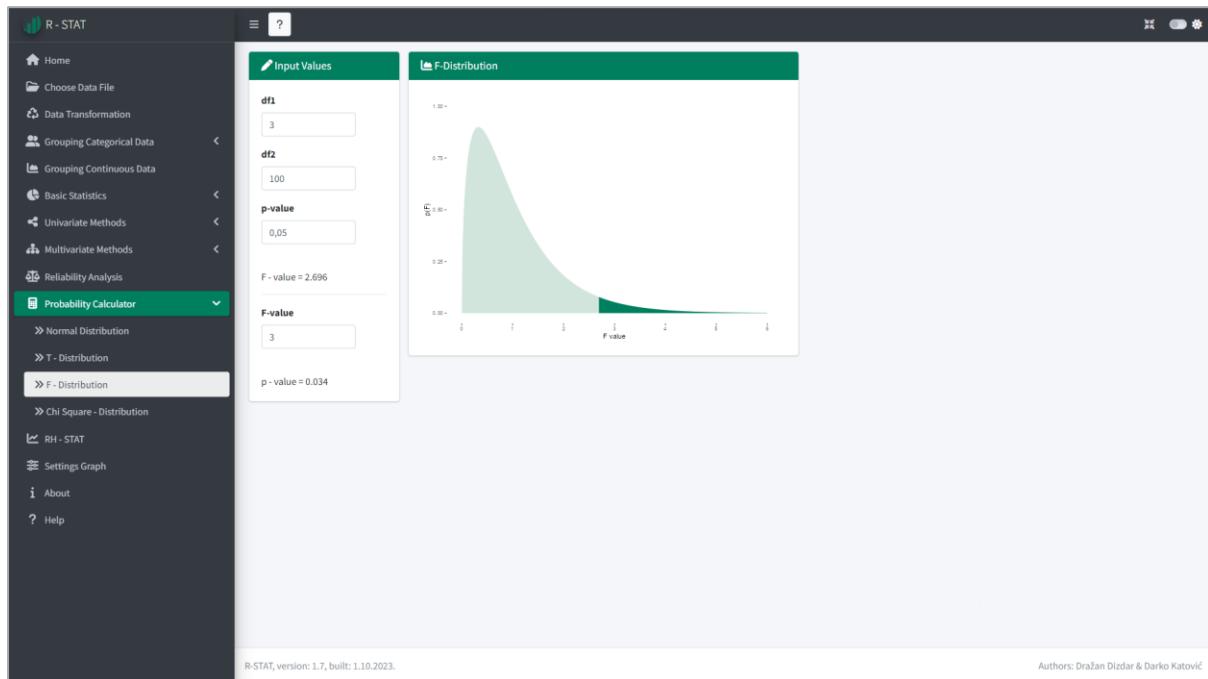
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **T-Distribution** omogućava računanje odgovarajuće t-vrijednosti (**Two tailed t-value** i **Upper tailed t-value**) za zadatu p-vrijednost (**p-value**) i broj stupnjeva slobode (**df**).



Slika 62: Probability Calculator → T- Distribution

F – Distribution

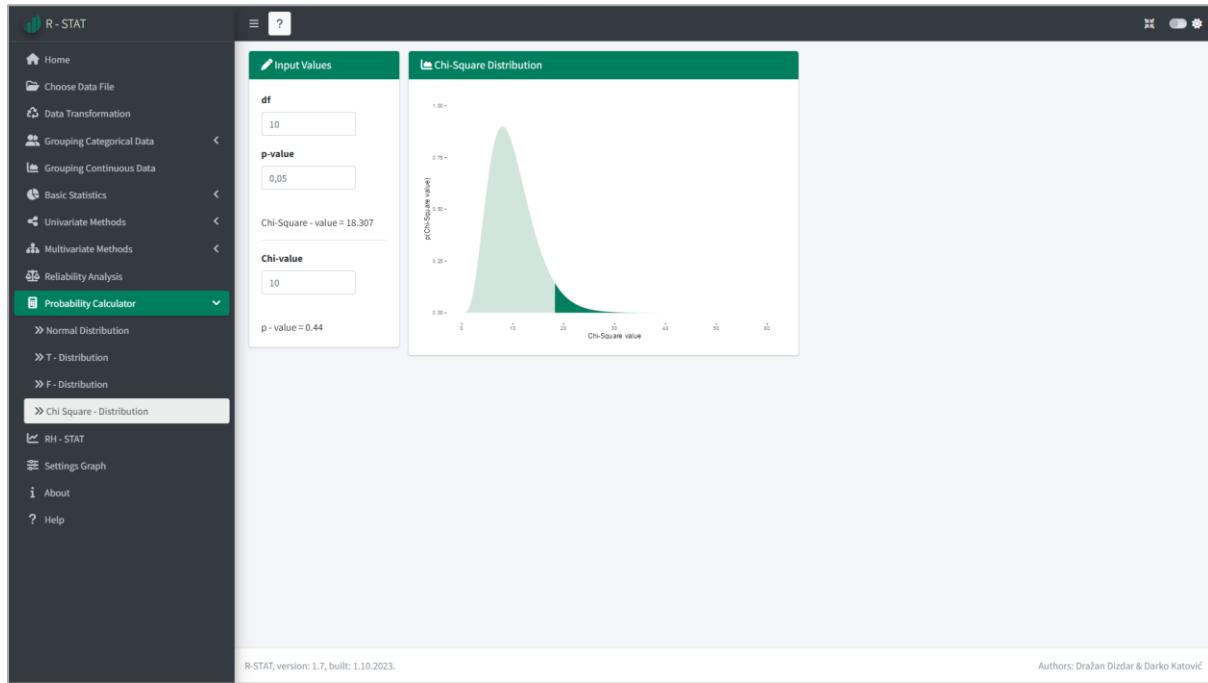
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **F-Distribution** omogućava računanje odgovarajuće F-vrijednosti (**F-value**) za zadanu p-vrijednost (**p-value**) i broj stupnjeva slobode (**df1** i **df2**).



Slika 63: Probability Calculator → F - Distribution

Chi Squere – Distribution

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

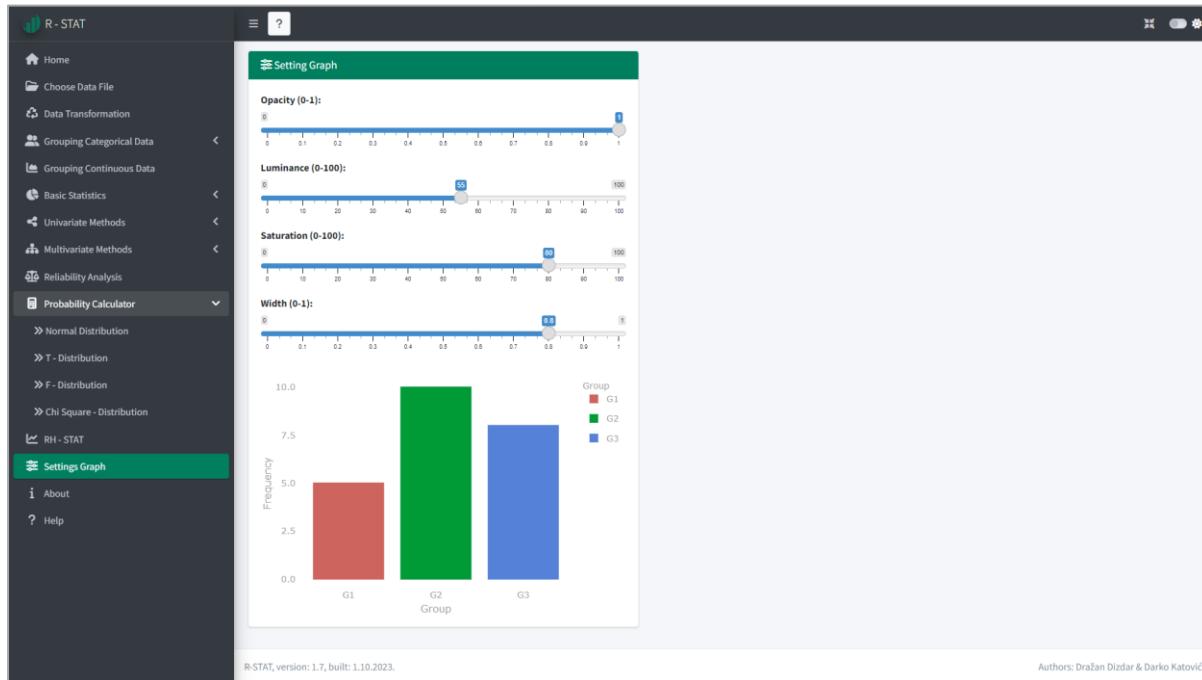


Slika 64: Probability Calculator → Chi Squere - Distribution

Settings Graph

Odabirom opcije Settings Graph glavnog izbornika moguće je odraditi postavke grafikona:

- **Width** - širina osnovice
- **Opacity** - prozirnost
- **Luminance** - svjetlina i
- **Saturation** - saturacija boja.



Slika 65: Settings Graph