

Dražan Dizdar i Darko Katović



Priručnik za web aplikaciju

R-STAT

Statistical Software

Zagreb, 2023.

Sadržaj

Web aplikacija R-STAT	3
Home	3
Choose Data File.....	6
Data Transformation	7
Grouping Categorical Data.....	9
Chi-Square Goodness of Fit Test	9
Chi-Square Test of Independence.....	12
McNemar's Test.....	15
Grouping Continuous Data	17
Basics Statistics.....	20
Descirptive Parameters.....	20
Testing for Normality	24
CI for Population Mean	27
Correlations Analysis	29
Univariate Methods.....	32
Independent Samples T-Test	32
Dependent Samples T-Test.....	36
One Way ANOVA.....	40
Repeated Measures ANOVA.....	44
Multivariate Methods	46
Regression Analysis.....	46
Factor Analysis	49
Canonical Analysis	52
Discriminant Analysis.....	54
Reliability Analysis.....	56
Probability Calculator	59
Normal Distribution	59
T – Distribution	60
F – Distribution.....	60
Chi Squere – Distribution.....	61

Settings Graph.....	62
---------------------	----

Web aplikacija R-STAT

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

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

Home

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



Slika 1: Home

U glavnom izborniku moguće je odabrati sljedeće opcije:

- Choose Data File
- 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 koje želimo analizirati. Podaci za statističko-grafičku analizu trebaju biti pripremljeni na sljedeći način (slika 2):

- prvi stupac tablice s podacima sadrži jedinstvene¹ oznake entiteta
- ostali stupci tablice s podacima sadrže podatke entiteta u varijablama
- oznake (kodovi) modaliteta kvalitativnih varijabli ne smiju biti brojevi
- datoteke s podacima mogu biti kreirane u CSV (Comma Separated Values) formatu gdje se kao separator između podataka koristi točka-zarez (;), a za decimalne brojeve zarez (,) ili u ODS (Open Document Spreadsheet) formatu.

¹ Entiteti ne smiju imati istu oznaku.

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
9	155	48	21	4	37	21	170	35
10	171	59	24	11	37	21	180	39

Slika 2. Primjer ispravno kreirane datoteke s podacima

Tablice s podacima za vježbu možete preuzeti na ovoj poveznici:

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

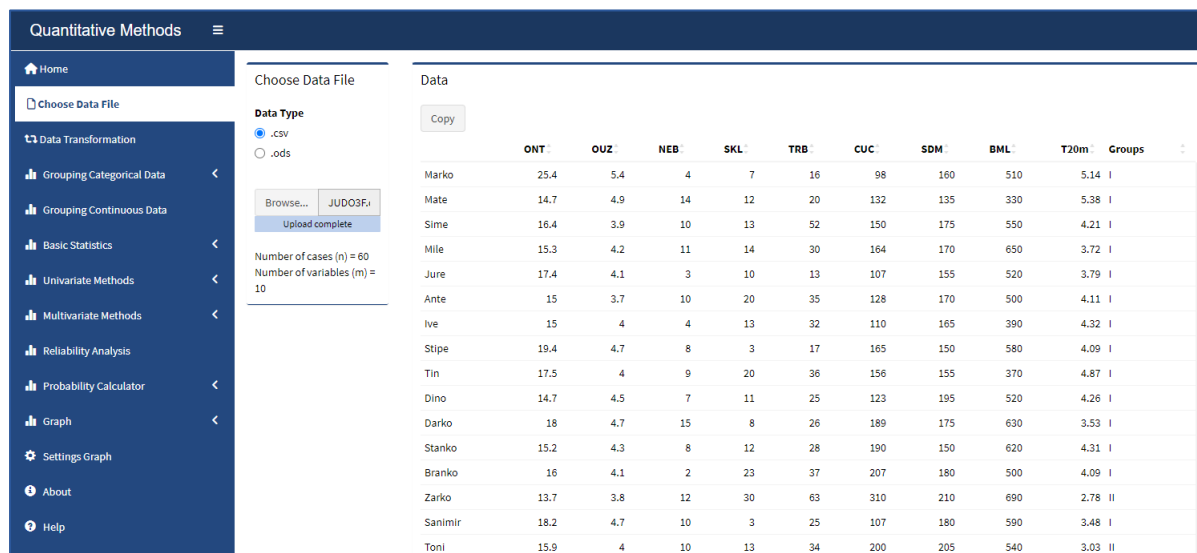
Klikom na gumb **Code** odaberemo opciju **Download ZIP**. Nakon preuzimanja arhiviranu datoteku **Data-main.zip** treba *raspakirati* i sve datoteke kopirati u prethodno kreiranu mapu.

NAPOMENA!

- ✓ Sve tablice koje se kreiraju u web aplikaciji Quantitative Methods moguće je kopirati u međuspremnik (clipboard) klikom na gumb **Copy** te ih potom premjestiti u neku drugu aplikaciju (npr. **MS Excel**, **LibreOffice Calc**).

Choose Data File

Odabirom opcije **Choose Data File** te klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Nakon odabira datoteka s podacima se prikazuje u tablici **Data** (slika 2).



The screenshot shows the 'Quantitative Methods' application. On the left is a sidebar with navigation options: Home, Choose Data File, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods, Multivariate Methods, Reliability Analysis, Probability Calculator, Graph, Settings Graph, About, and Help. The 'Choose Data File' dialog is open, showing 'Data Type' with radio buttons for '.CSV' (selected) and '.ods'. Below are 'Browse...' and 'JUDO3F...' buttons, and an 'Upload complete' button. It also displays 'Number of cases (n) = 60' and 'Number of variables (m) = 10'. The main 'Data' table is visible, containing 11 columns and 16 rows of data.

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

Slika 3: Choose Data File

Opcija **Data Type** omogućava učitavanje datoteke s podacima koje su kreirane u **.csv** ili **.ods** formatu:

- **.csv** – tekstualna datoteka s vrijednostima odvojenim zarezima (,) ili točka-zarezima (;). Ovaj format najčešće se koristi za uvoz (import) podataka iz jednog sustava pohrane u drugi, budući da sve aplikacije za statističku obradu podataka prepoznaju ovaj format
- **.ods** – datoteka kreirana u **Calc** programu za proračunske tablice koji se nalazi u okviru besplatnog programskog paketa LibreOffice.

<https://www.libreoffice.org/>

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

Data Transformation

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

	NEB	SKL	TRB	CUC	SDM	BML	ONT	OUZ	T20m	MEAN_Z
Marko	-1.933	-1.088	-1.445	-1.002	-1.023	-0.217	-2.72	-2.746	-2.082	-1.584
Mate	1.236	-0.519	-1.202	-0.699	-2.506	-2.149	0.353	-1.829	-2.48	-1.088
Sime	-0.032	-0.406	0.736	-0.538	-0.133	0.213	-0.135	0.003	-0.54	-0.092
Mile	0.285	-0.292	-0.597	-0.414	-0.43	1.287	0.181	-0.547	0.273	-0.028
Jure	-2.25	-0.747	-1.627	-0.921	-1.32	-0.109	-0.422	-0.363	0.157	-0.845
Ante	-0.032	0.39	-0.294	-0.734	-0.43	-0.324	0.267	0.37	-0.374	-0.129
Ive	-1.933	-0.406	-0.476	-0.895	-0.727	-1.505	0.267	-0.18	-0.722	-0.731
Stipe	-0.666	-1.543	-1.384	-0.405	-1.616	0.535	-0.997	-1.463	-0.341	-0.875
Tin	-0.349	0.39	-0.233	-0.485	-1.32	-1.72	-0.451	-0.18	-1.634	-0.665
Dino	-0.982	-0.633	-0.9	-0.779	1.053	-0.109	0.353	-1.096	-0.623	-0.413
Darko	1.553	-0.974	-0.839	-0.191	-0.133	1.072	-0.594	-1.463	0.588	-0.109
Stanko	-0.666	-0.519	-0.718	-0.182	-1.616	0.965	0.21	-0.73	-0.705	-0.44
Branko	-2.567	0.732	-0.173	-0.03	0.163	-0.324	-0.02	-0.363	-0.341	-0.325
Zarko	0.602	1.528	1.402	0.887	1.942	1.716	0.64	0.186	1.832	1.193
Sanimir	-0.032	-1.543	-0.9	-0.921	0.163	0.642	-0.652	-1.463	0.671	-0.448
Toni	-0.032	-0.406	-0.354	-0.093	1.646	0.106	0.009	-0.18	1.418	0.235

Slika 4: Data Transformation → Z-volute

Odabirom opcije **Data Transformation** glavnog izbornika za odabrane normalno skalirane (**Normally Scaled**) ili obrnuto skalirane (**Inversely Scaled**) varijable prikazuje se tablica s podacima (slika 4) transformiranih u:

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

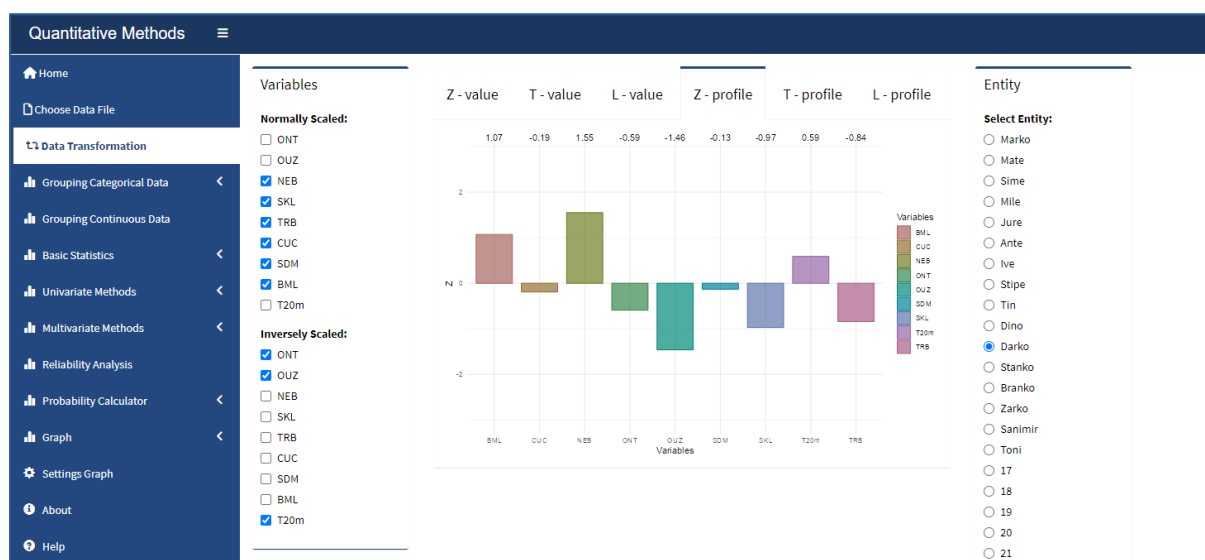
$$L_i = 3 + 0,83 \cdot z_i$$

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 rezultati 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 (**Z-profile**, **T-profile** i **L-profile**) u odabranim varijablama (slika 5).



Slika 5: Data Transformation → Z – profile

CILJ

- ✓ transformirati podatke dobivene mjerenje u Z-vrijednosti, T-vrijednosti i L- vrijednosti
- ✓ kondenzirati transformiranje podatke aritmetičkom sredinom
- ✓ grafički prikazati profile entiteta u transformiranim podacima.

UVJETI

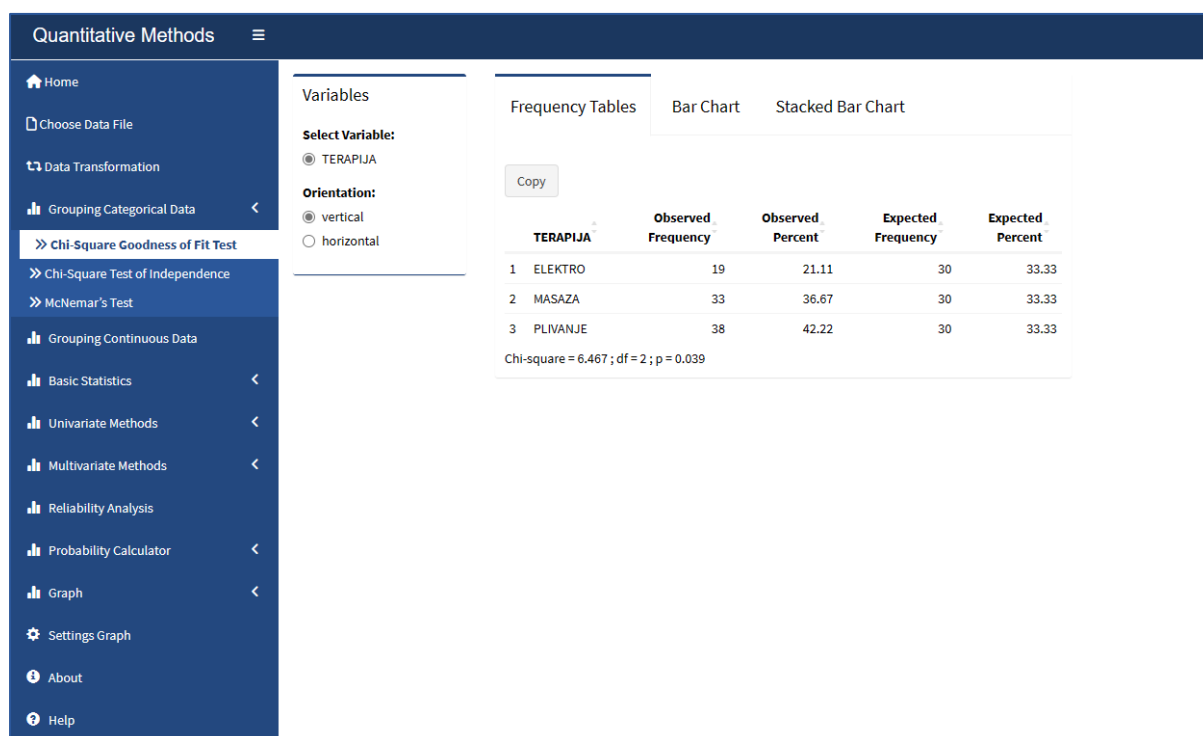
- ✓ podaci su dobiveni na intervalnoj ili omjernoj mjerne ljestvici.

Grouping Categorical Data

Chi-Square Goodness of Fit Test

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

Odabirom opcije **Grouping Categorical Data** glavnog izbornika te opcije **Chi-square Goodness of Fit Test** za odabranu varijablu (**Select Variable**) prikazuje kartica **Frequency Tables** s rezultatima jednodimenzionalnog grupiranja za odabranu varijablu 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 kartica **Frequency Tables** nalazi se tablica koja za svaki modalitet kvalitativne varijable prikazuje:

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

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

gdje je

- ✓ $fo(\%)_g$ relativna frekvencija modalitet g ($g = 1, \dots, k$) izražena u postotku
- ✓ fo_g 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 sukladnosti²

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

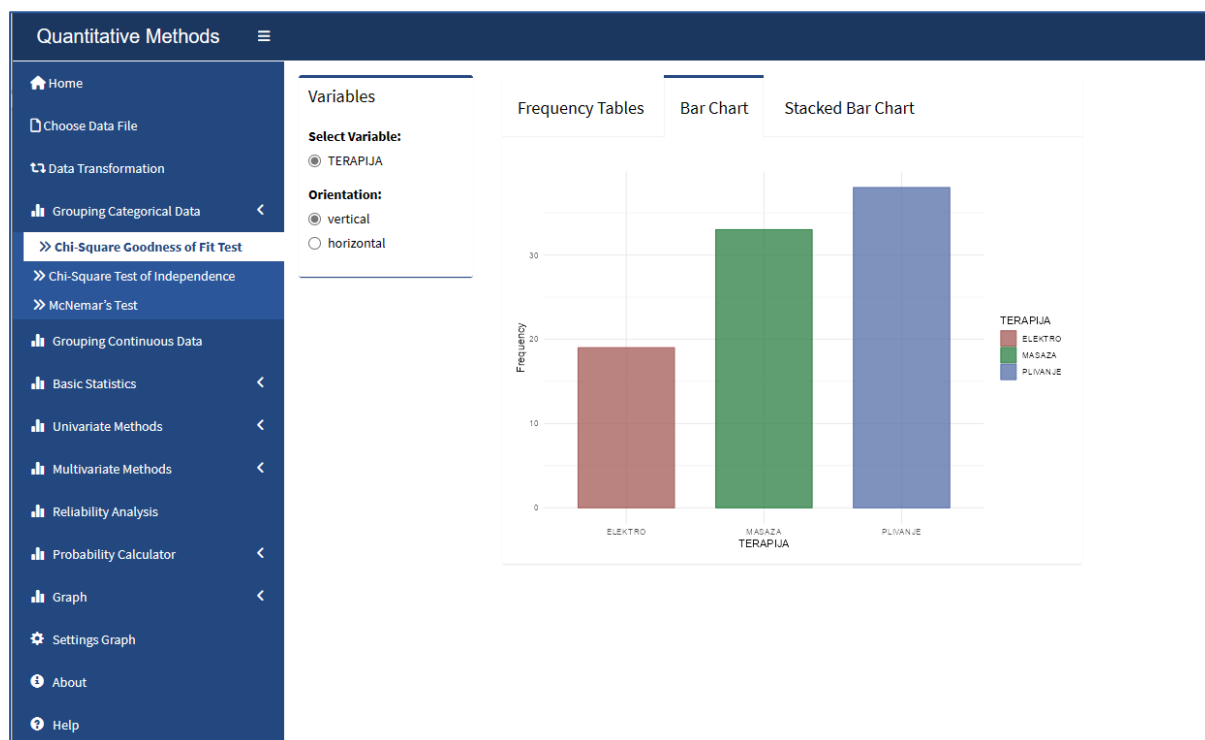
$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

gdje je

- ✓ χ^2 hi kvadrat vrijednost
 - ✓ f_o opažene frekvencije
 - ✓ f_e očekivane frekvencije.
- **df** – broj stupnjeva slobode ($df = k - 1$, gdje je k – broj kategorija, odnosno oblika mjerenog obilježja).
 - **p** – pogreška s kojom tvrdimo da je razlika između opaženih i očekivanih frekvencija statistički značajna.

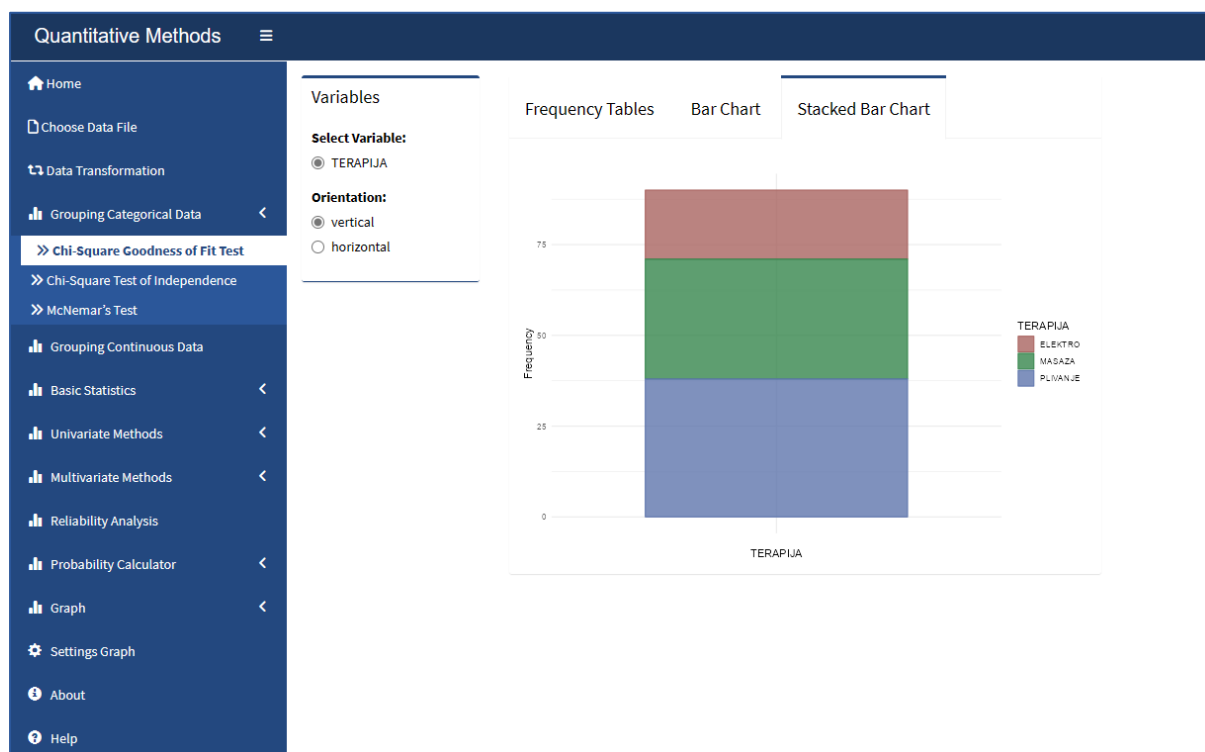
² U anglosaksonskoj literaturi test se naziva *goodness of fit* jer se testira koliko dobro (*good*) opažene frekvencije odgovaraju (*fit*) očekivanim frekvencijama.

U karticama **Bar Chart** (slika 7) i **Stacked Bar 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** – 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 → Stacked Bar Chart

- **Bar Stacked Chart** – grafikon s razdjeljenim stupcima prikazuje udjele svakog modaliteta u ukupnoj frekvenciji.

CILJ

- ✓ jednodimenzionalno grupirati kvalitativne varijable
- ✓ utvrditi da li je razlika između opaženih i očekivanih frekvencija dobivenih uniformnom raspodjelom statistički značajna.

UVJETI

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

Odabirom opcije **Grouping Categorical Data** glavnog izbornika te opcije **Chi-Square Test of Independence** za odabrane varijable (**Select Variable 1** i **Select Variable 2**) prikazuju se tablice u kartici **Contingency Tables** s rezultatima dvodimenzionalnog grupiranja za odabrane varijable te rezultati χ^2 -testa za utvrđivanje veze između njih (slika 9).

χ^2 - test neovisnosti³ (*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

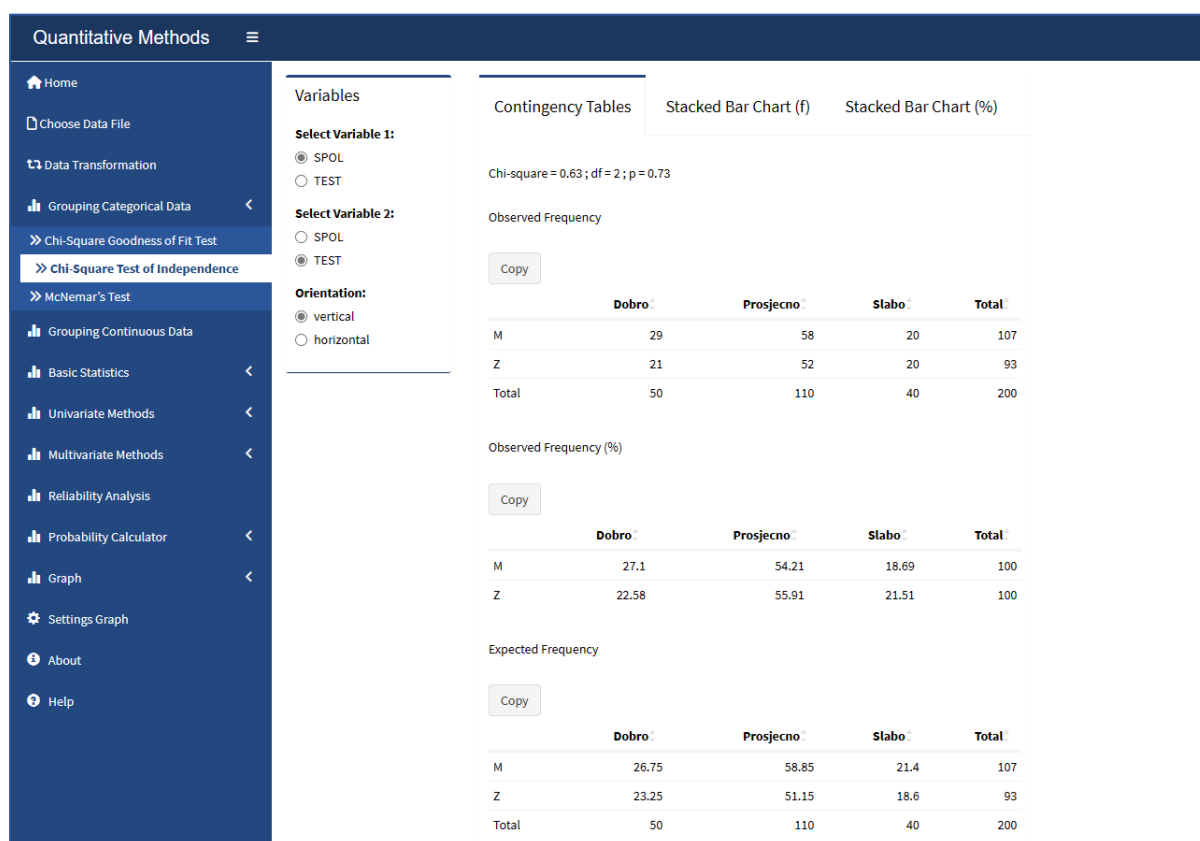
$$fe = \frac{\sum reda \cdot \sum stupca}{n}$$

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

³ U anglosaksonskoj literaturi test se naziva *Chi-Square Test of Independence* jer se njime testira jesu li dvije kvalitativne varijable međusobno povezane (ovisne).

✓ n ukupan broj entiteta.



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

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

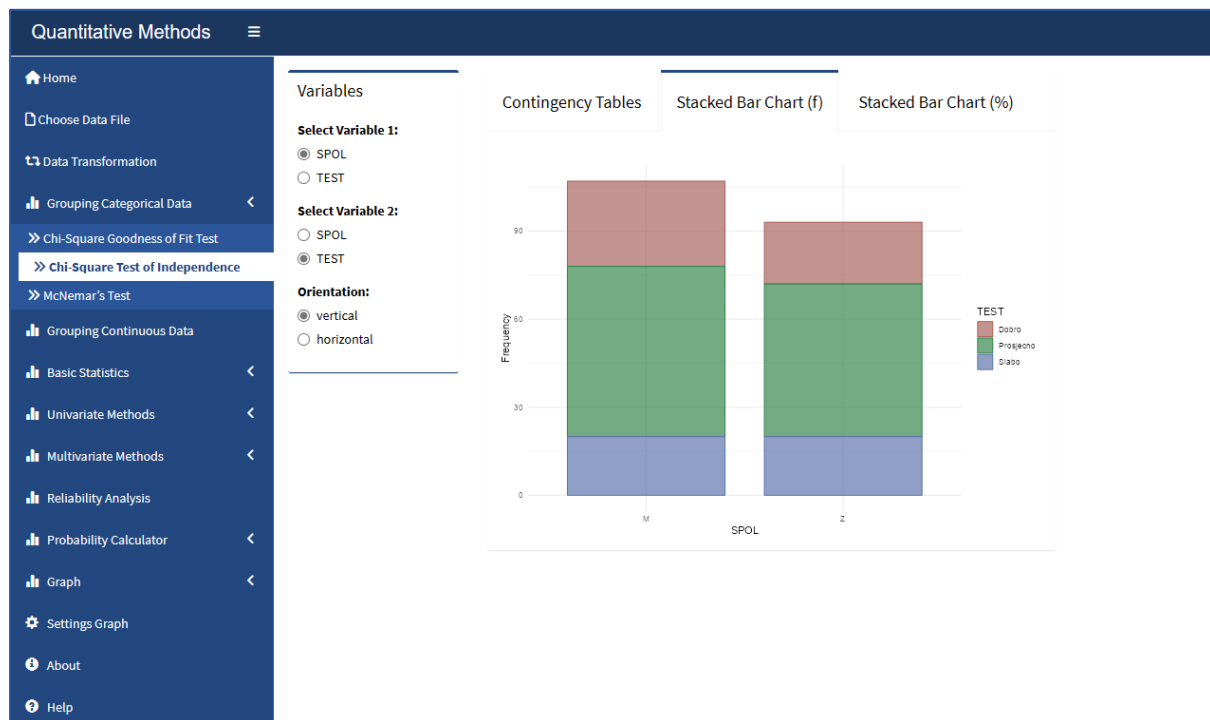
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{oij} - f_{eij})^2}{f_{eij}}$$

gdje je

- ✓ χ^2 hi kvadrat vrijednost
- ✓ f_{oij} 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)⁴
- **p** – pogreška s kojom tvrdimo da je razlika između opaženih i očekivanih frekvencija statistički značajna.

⁴ U broj redova i broj stupaca ne ubrajaju se zbirni red i zbirni stupac.

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**) orijentacije (**Orientation**).



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



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

- **Bar Stacked Chart (f)** – grafikon s razdjeljenim stupcima 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 stupcima 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

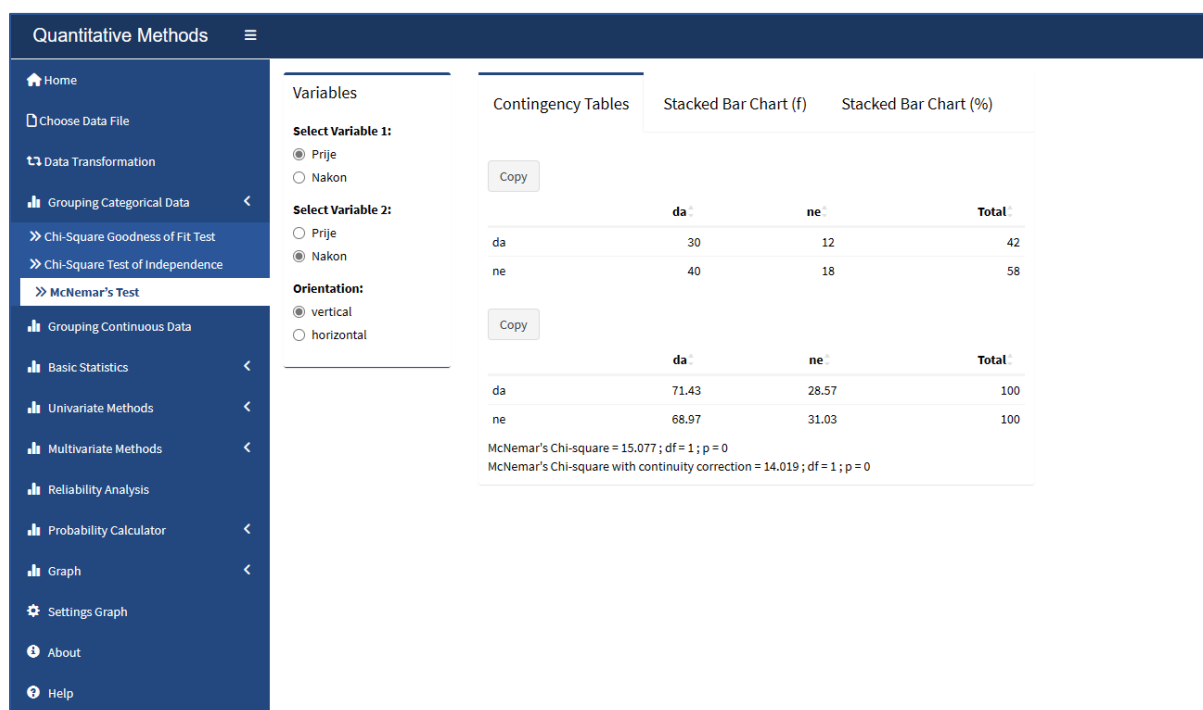
- ✓ dvodimenzionalno grupirati kvalitativne varijable
- ✓ utvrditi postoji li veza između dvije kvalitativne varijable.

UVJETI

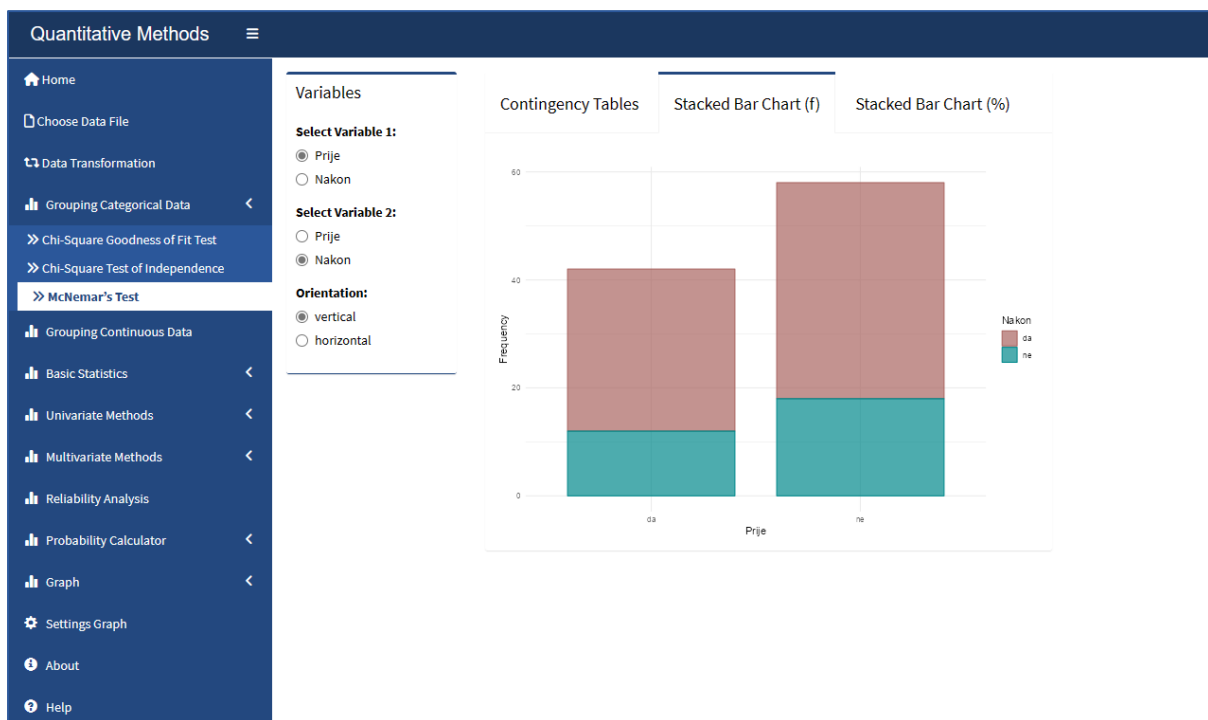
- ✓ podaci su dobiveni na nominalnoj ili ordinalnoj mjernoj ljestvici s dva ili više modaliteta
- ✓ podaci su dobiveni na slučajno odabranim i relativno velikim uzorcima entiteta ($n > 30$).

McNemar's Test

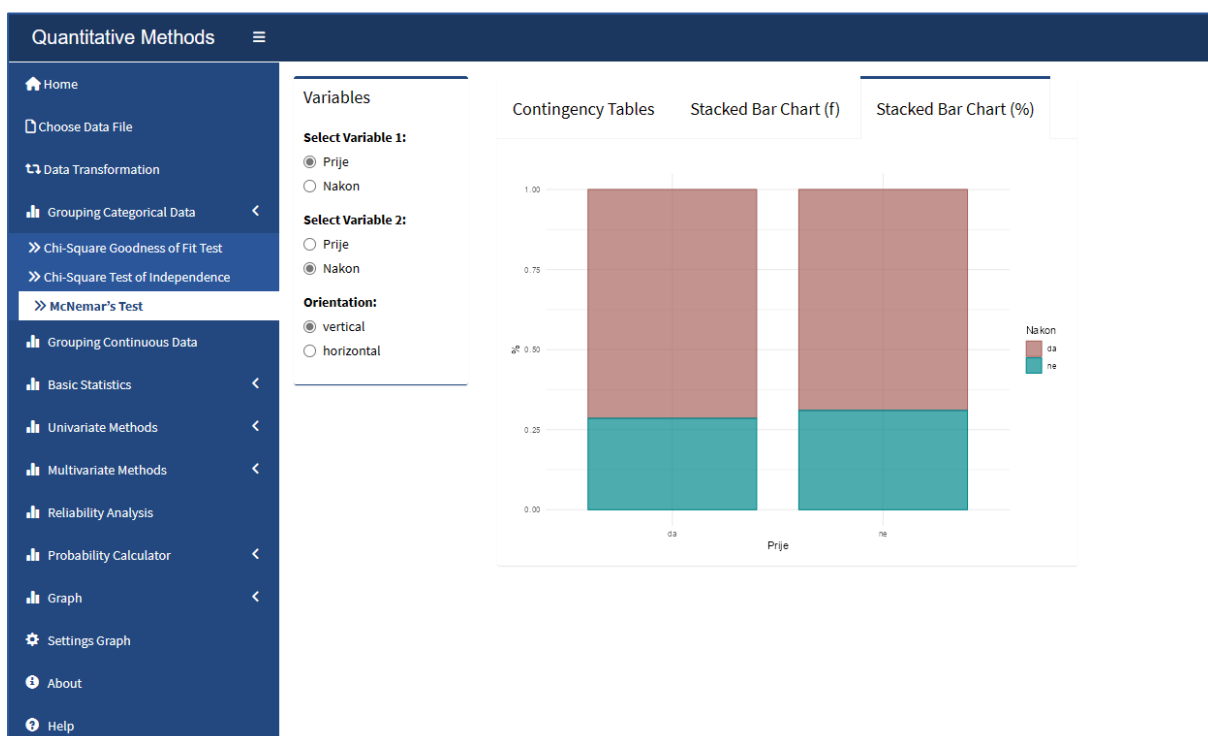
Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje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



Slika 12: McNemar's Test → Contingency Tables



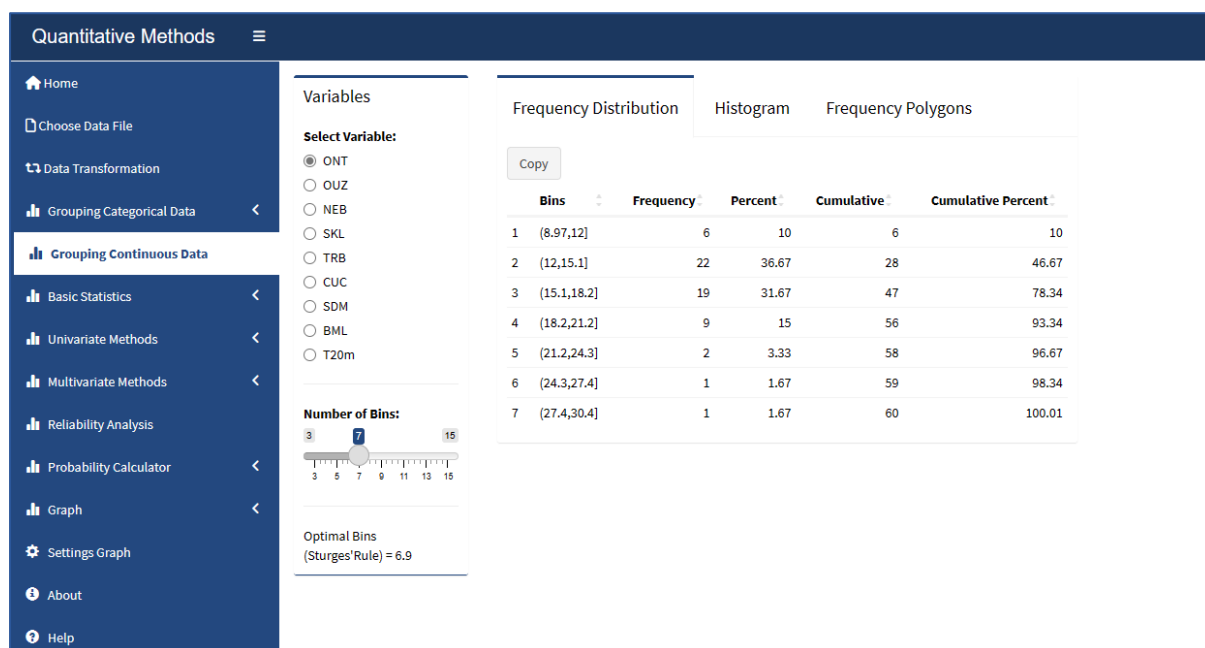
Slika 13: McNemar's Test → Stacked Bar Chart (f)



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

Grouping Continuous Data

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



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** – apsolutne 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.

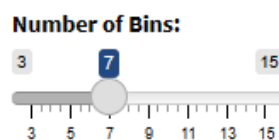
Za uspješno grupiranje potrebno je odrediti optimalan broj i interval razreda. Broj razreda prije svega ovisi o broju entiteta i najčešće se kreće između pet i petnaest. Za procjenu optimalnog broja razreda moguće je koristiti Sturgesovo pravilo

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

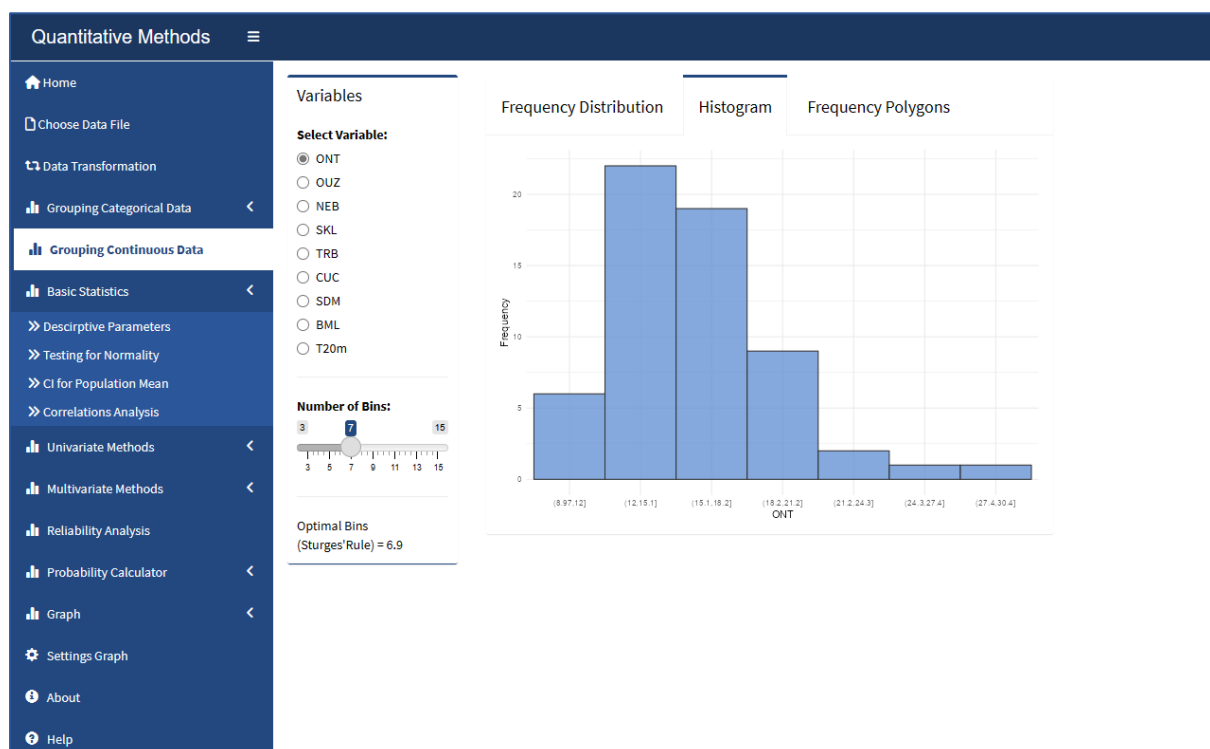
gdje je:

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

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

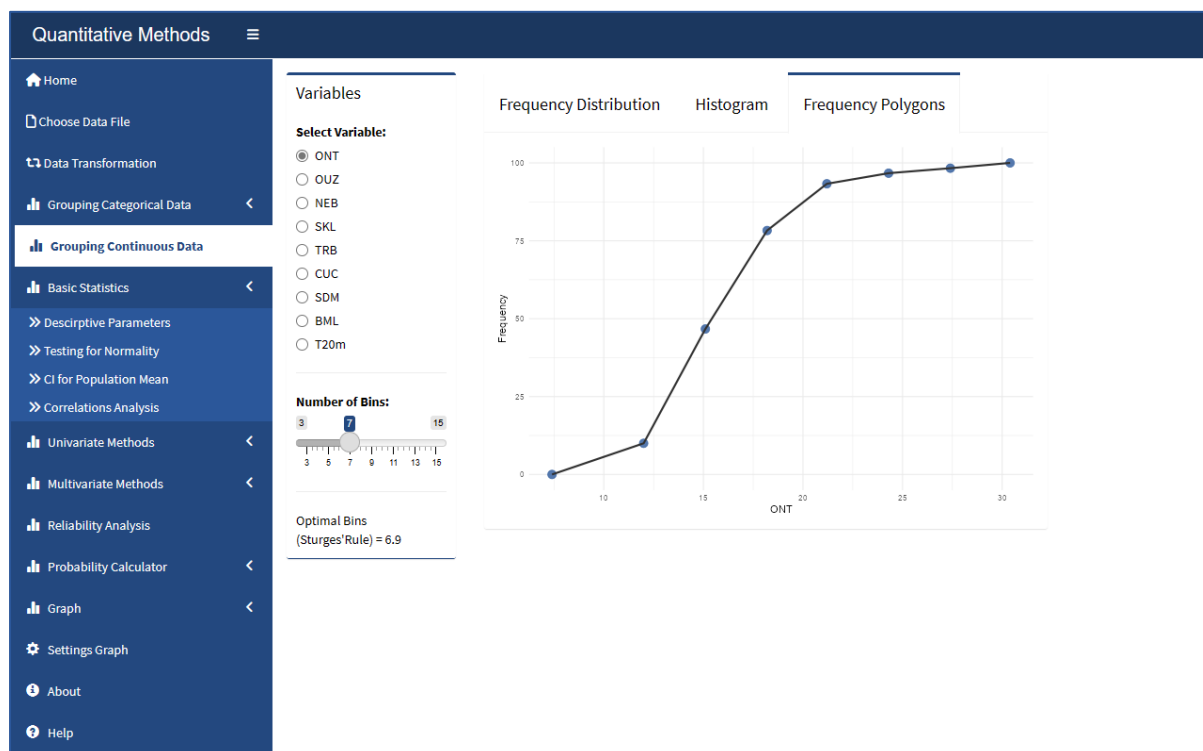


U kartici **Histogram** nalazi se histogram apsolutnih frekvencija (slika 16), a u kartici **Frequency Polygons** poligon relativnih postotnih kumulativnih frekvencija (slika 17).



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.

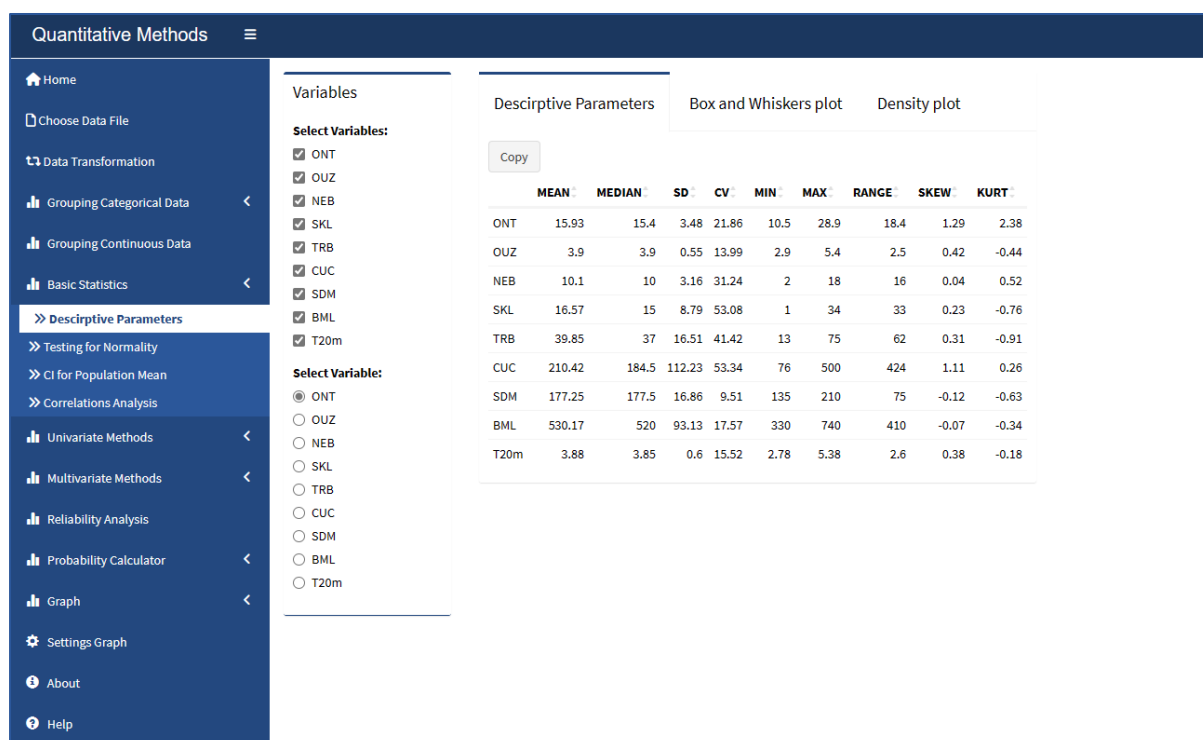
UVJETI

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



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.

$$\mu_e = \frac{x_{r1} + x_{r2}}{2} \text{ gdje je } r_1 = \frac{n}{2}, \text{ a } r_2 = \frac{n}{2} + 1$$

- **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 vrijadnost
- **MAX** – maksimalna vrijednosti
- **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 distirbucija 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.

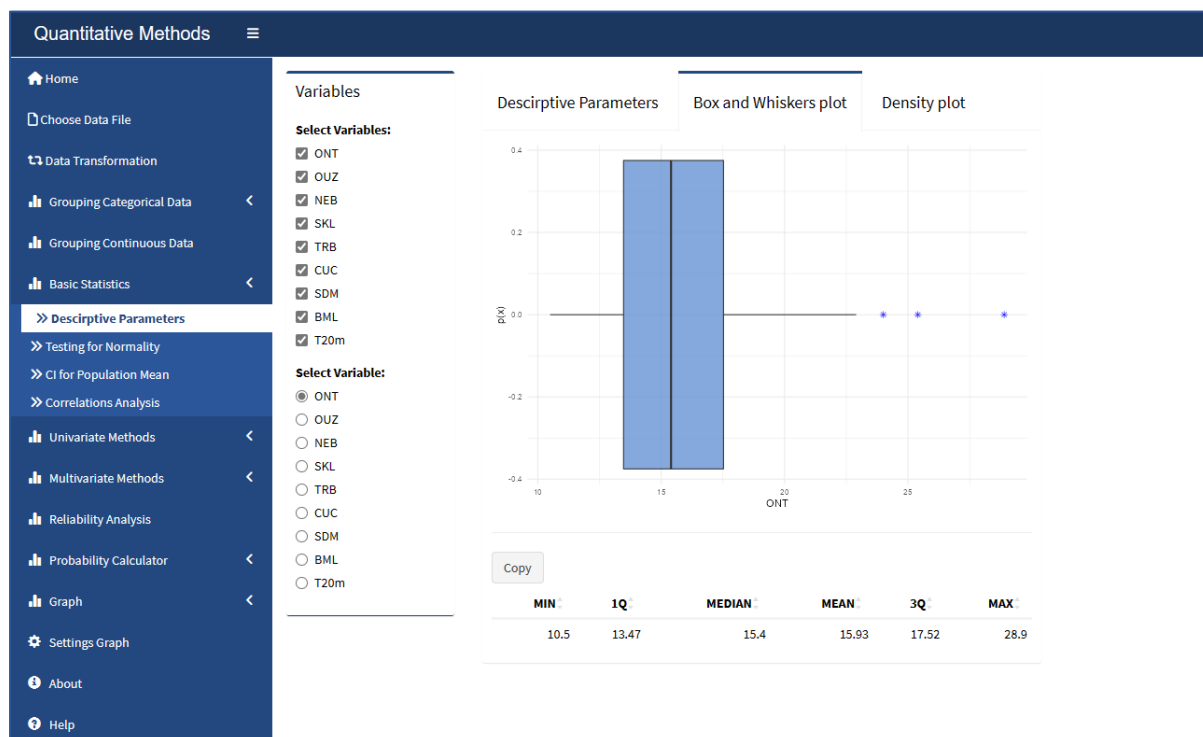
Korigirani koeficijet a_4 (tip 2) računa se formulom

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

Ako je koeficijent spljoštenosti:

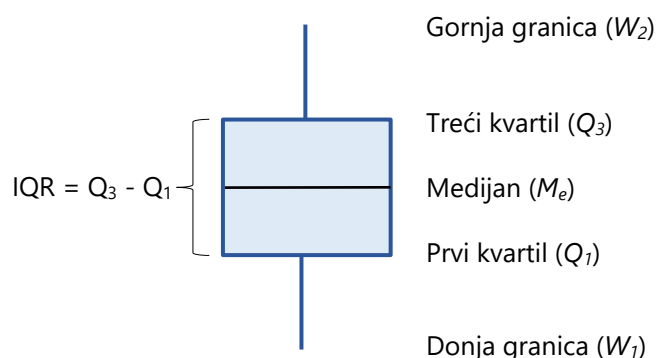
- ✓ $a_4 = 0$ - distribucija je mezokurtična - normalna
- ✓ $a_4 > 0$ - distribucija je leptokurtična - izdužena
- ✓ $a_4 < 0$ - distribucija je platikurtična - spljoštena.

Odabirom kartice **Box and Whiskers plot** (slika 19) ili **Density plot** (slika 20) 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 temeljeg kojeg je moguće uočiti stupanj disperzije i asimetrije distribucije te outliere (vrijednosti koje ekstremno odstupaju od ostalih).

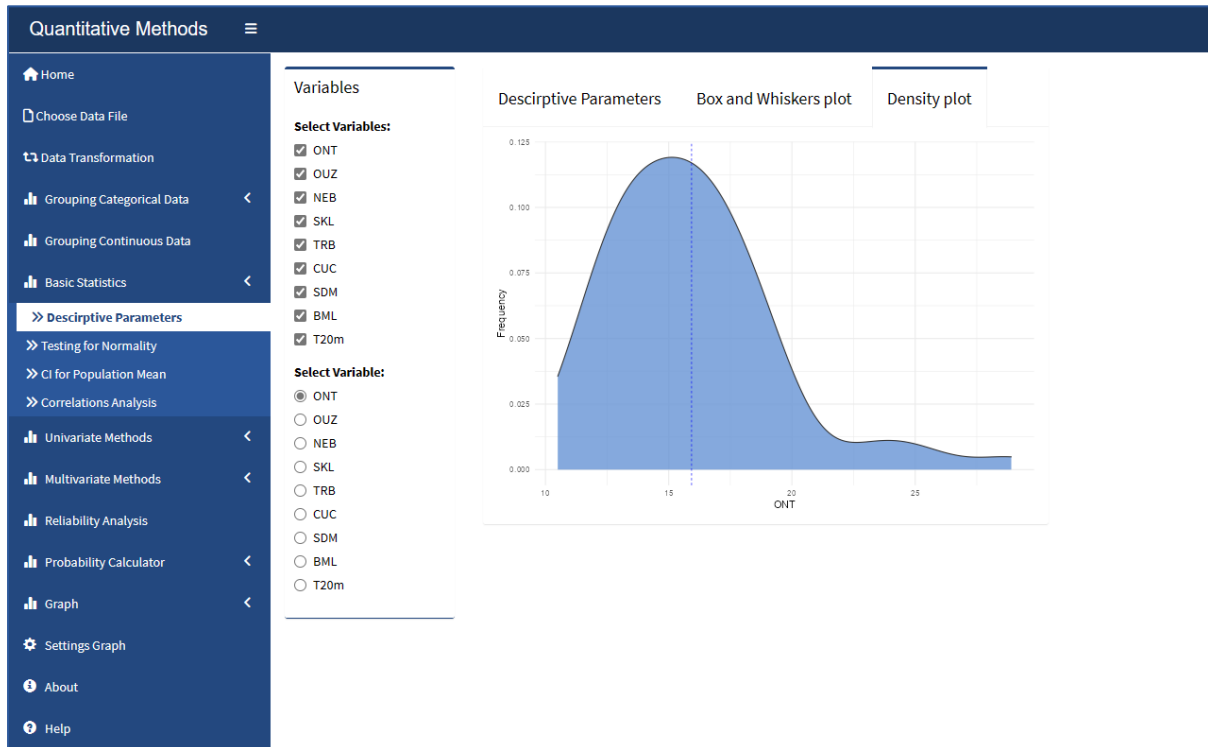


Kutijasti dijagram sastoji od pravokutnika čije stranice prikazuje vrijednosti prvog (Q_1) i trećeg kvartila (Q_3) unutar kojih se nalazi 50% svih rezultata. Crta unutar pravokutnika označava median (M_e), dok se donja (W_1) i gornja (W_2) granica (engl. *whisker*) najčešće odredi na sljedeći način:

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

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

- Dijagram gustoće (engl. *Density plot*) „izglađena“ je verzija histograma procijenjena na temelju empirijskih podataka (slika 20).



Slika 20: Basic Statistics → Descriptive Parameters → Density plot

CILJ

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

UVJETI

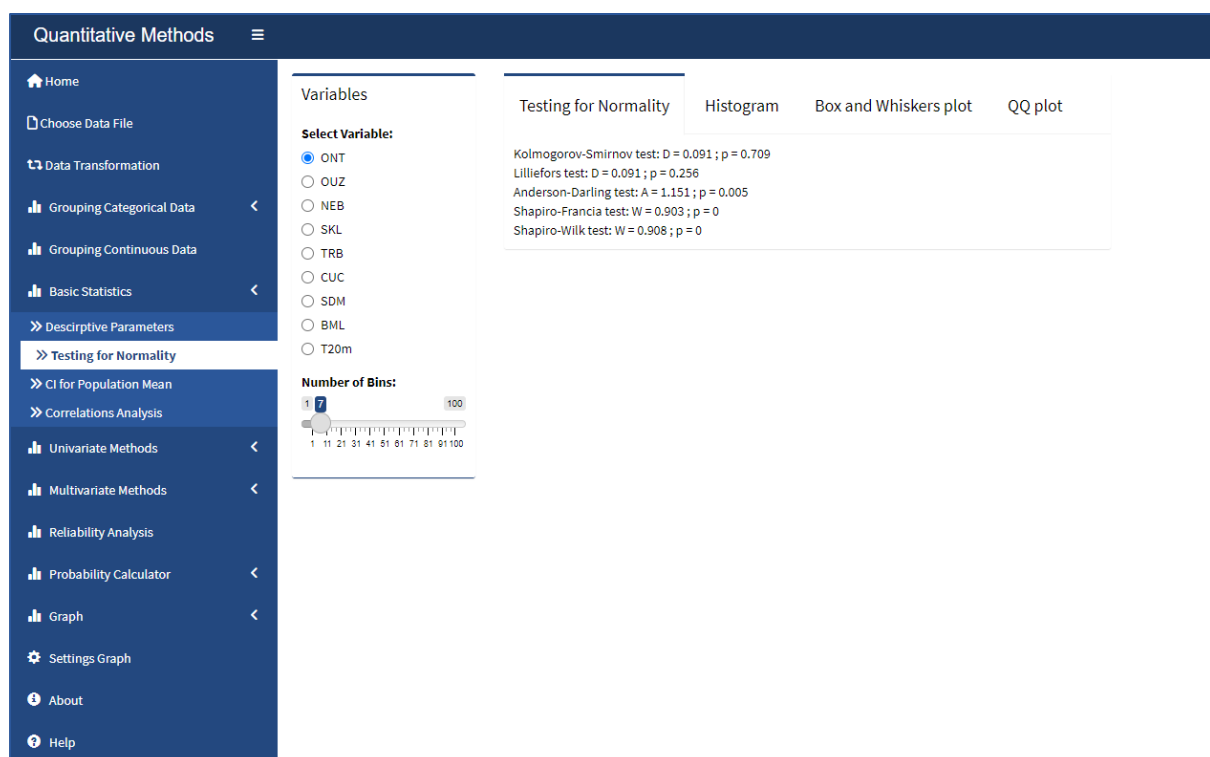
- ✓ podaci su dobiveni na intervalnoj ili omjernoj mjernoj ljestvici

Testing for Normality

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

Odabirom opcije **Basics Statistics** glavnog izbornika te opcije **Testing for Normality** za odabranu varijablu (**Select Variable**) dobije mogućnost testiranja statističke značajnosti odstupanja empirijske distribucije od normalne ili Gaussove distribucije statističkim postupcima kao što su:

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



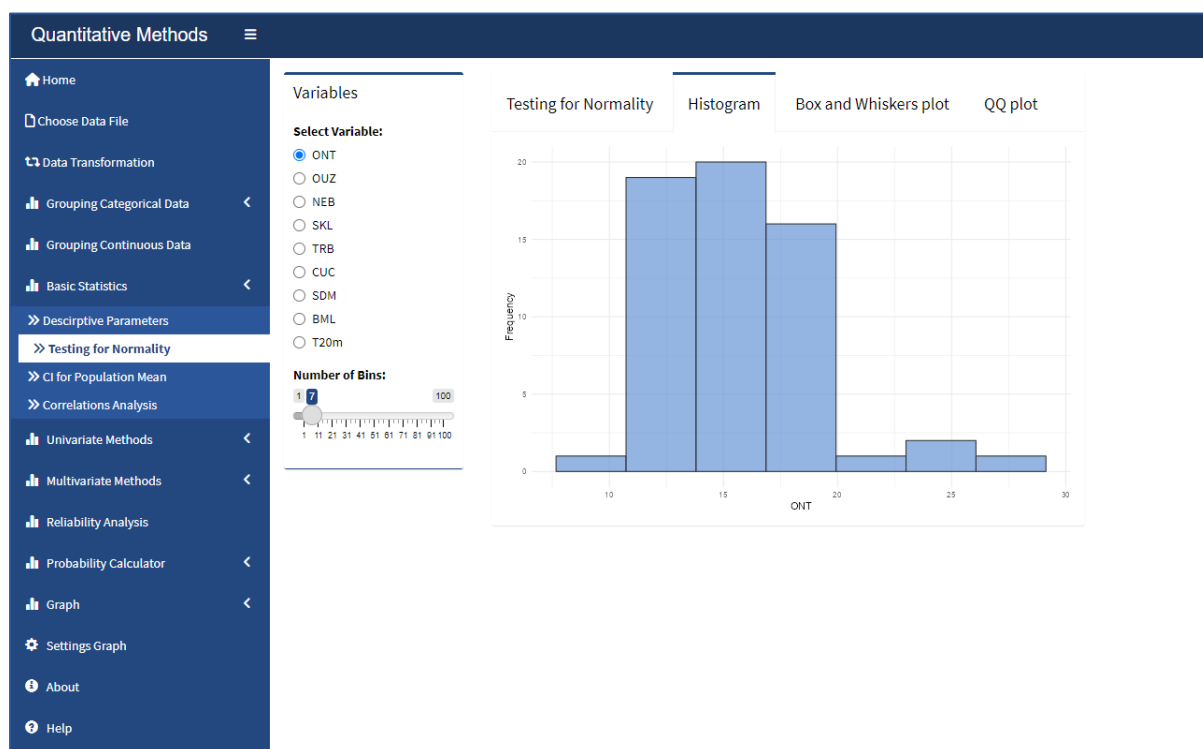
Slika 21: Basic Statistics → Testing for Normality

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 uzorci, to može rezultirati statistički značajnim testovima normalnosti distribucija i u slučajevima kada grafički prikazi empirijskih distribucija izgledaju gotovo identično teoretskoj normalnoj distribuciji. Nasuprot tome, na malim uzorcima testovima normaliteta distribucije nedostaje statističke snage kako bi se utvrdilo da su evidentno velika odstupanja neke empirijske

distribucije statistički značajna. Test za kojeg se smatra da ima najveću statističku snagu od svih testova za provjeru hipoteze o normalnosti neke empirijske distribucije je Shapiro-Wilkov test.

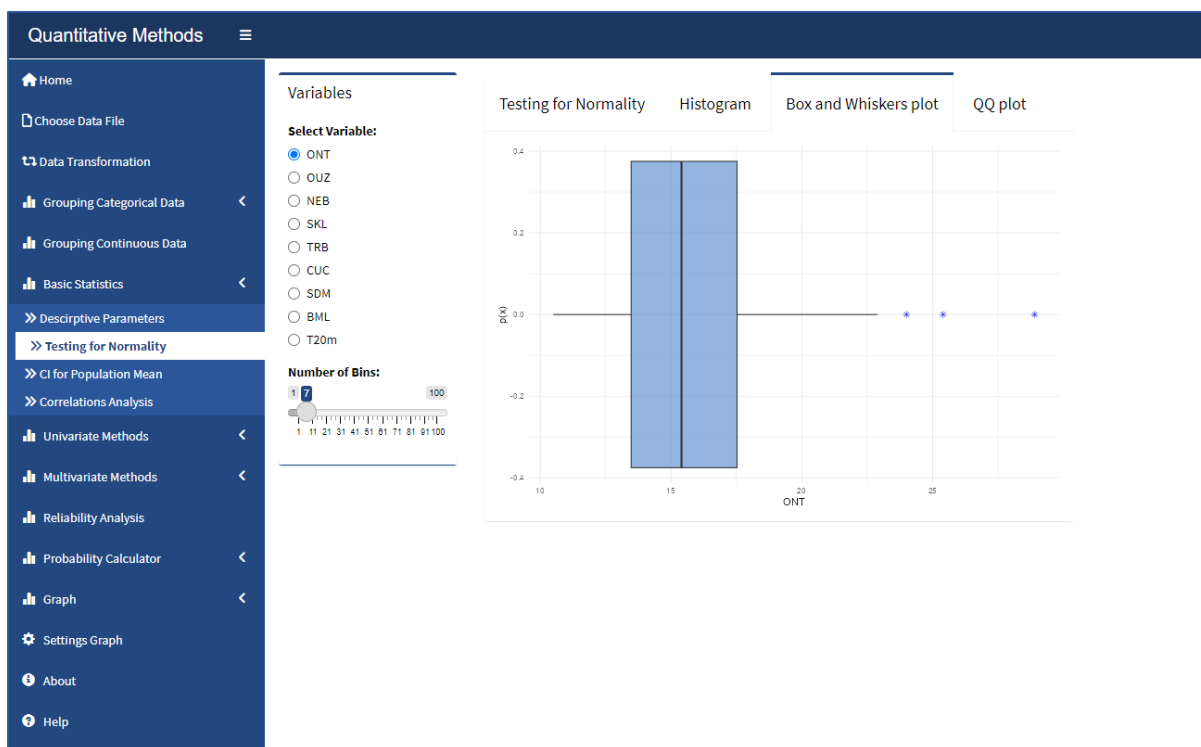
Vrijednosti parametra W Shapiro-Wilkovog testa kreću se u intervalu od 0 do 1, pri čemu niže vrijednosti pokazuju veća, a više vrijednosti manja odstupanja od normalne distribucije. Shapiro-Wilkov test modificirao je Royston (1992) i proširio mogućnost korištenja za uzorke veće od 50 entiteta. Razali i Wah (2011) uspoređivali su snagu Shapiro-Wilkovog, Kolmogorov-Smirnovog, Lillieforsovog i Anderson-Darlingovog testa pomoću Monte Carlo metode na deset tisuća uzoraka različitih veličina koji su generirani iz populacija čije distribucije nisu normalne. Rezultati su pokazali da je Shapiro-Wilkov test naj snažaniji, dok je Kolmogorov-Smirnov test najmanje snažan. Međutim, snaga svih testova i dalje je niska za male uzroke ($n < 30$). Slične rezultate dobili su Mendesa i Pala (2003) te Keskin (2006). Autori su predložili da se u praksi kombiniraju grafički prikazi (Histogram, Box & Whiskers Plot i QQ Plot), pokazatelji oblika distribucije (skewnes i kurtosis) te statističkim testovima kako bi se što pouzdanije 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 temeljeg kojeg je moguće uočiti stupanj disperzije i asimetrije distribucije te outliere (vrijednosti koje ekstremno odstupaju od ostalih). Za detaljniji opis vidi str. 21-22.



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

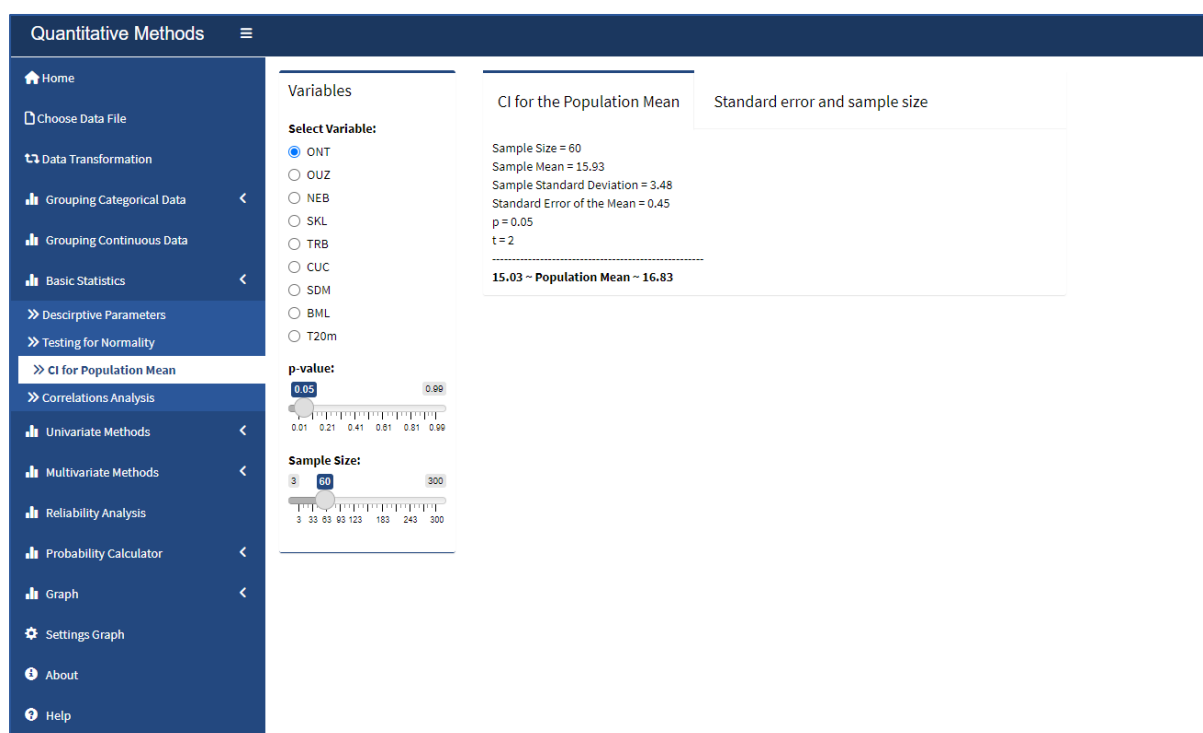
- ✓ utvrditi odstupa li statistički značajno empirijska distribucija neke kvantitativne varijable od normalne ili Gaussove distribucije.

UVJETI

- ✓ podaci su dobiveni na intervalnoj ili omjernoj mjernoj ljestvici

CI for Population Mean

Nakon pokretanja aplikacije Quantitative Methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** bismo 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.

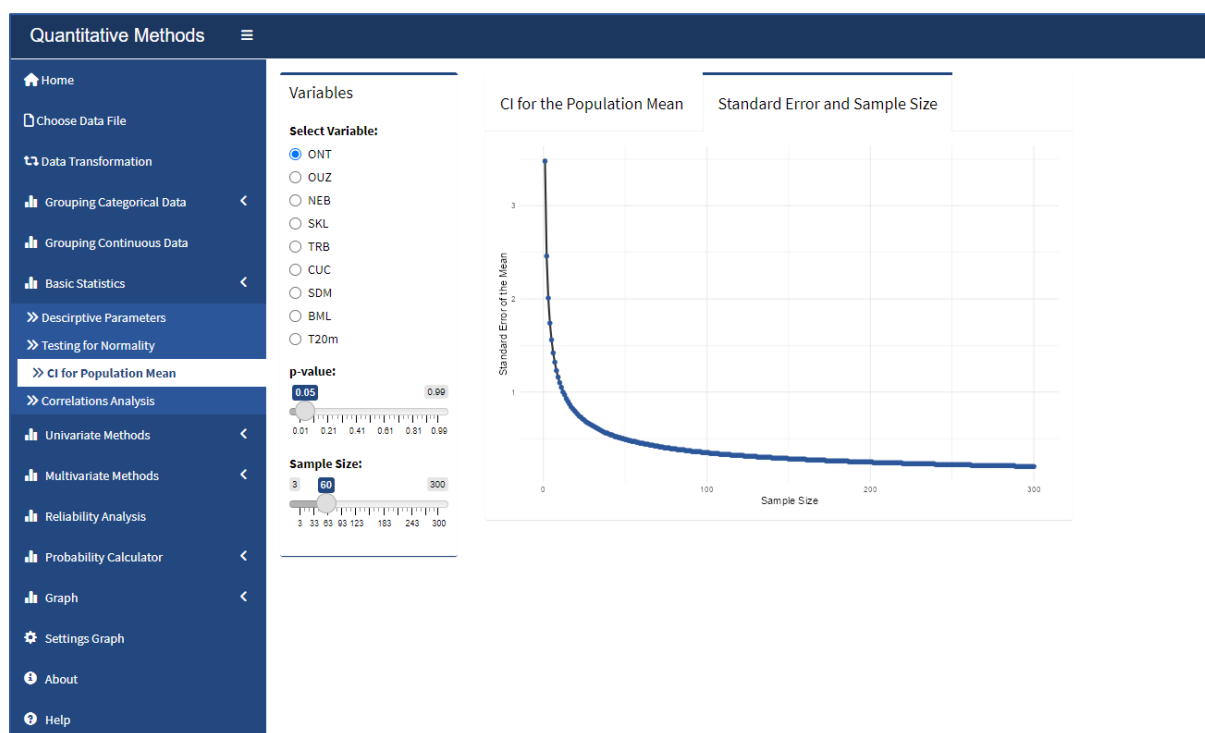


Slika 25: Basic Statistics → CI for the Population Mean → CI for the Population Mean

U okviru opcije **Basics Statistics** glavnog izbornika odabirom 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 od 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 procijeniti interval u kojem se nalazi aritmetička sredina populacije.

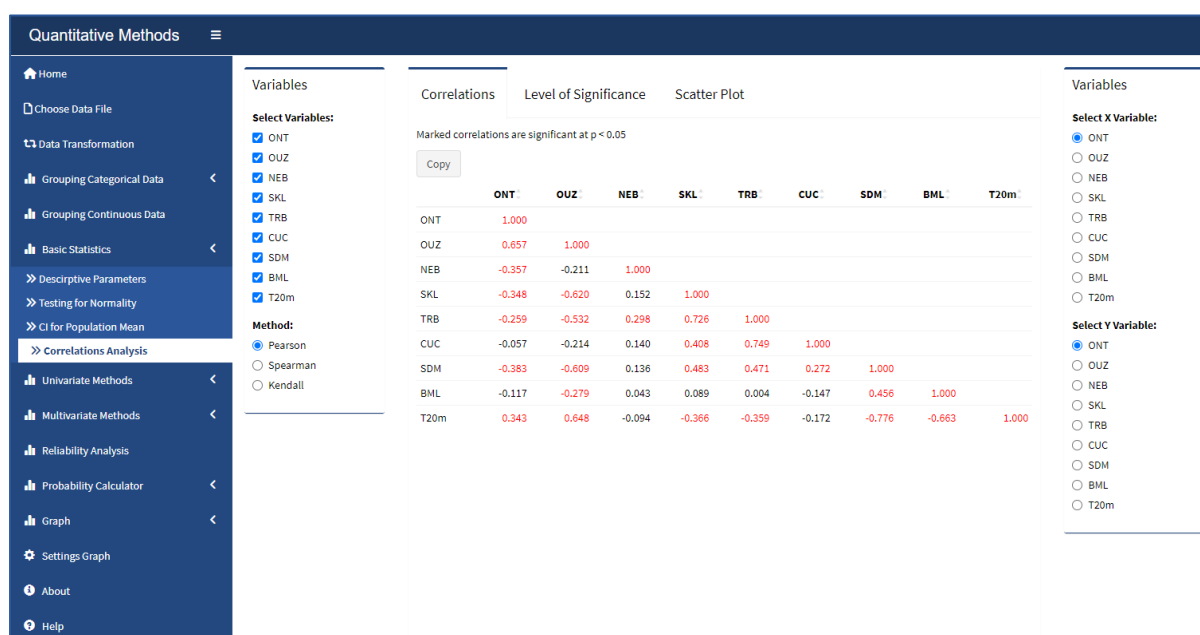
UVJETI

- ✓ podaci su dobiveni na intervalnoj ili omjernoj mjernoj ljestvici

Correlations Analysis

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

Odabirom opcije **Basics Statistics** glavnog izbornika te opcije **Correlations Analysis** za odabrane varijable (**Select Variables**) prikazuje se tablica korelacija (slika 27).

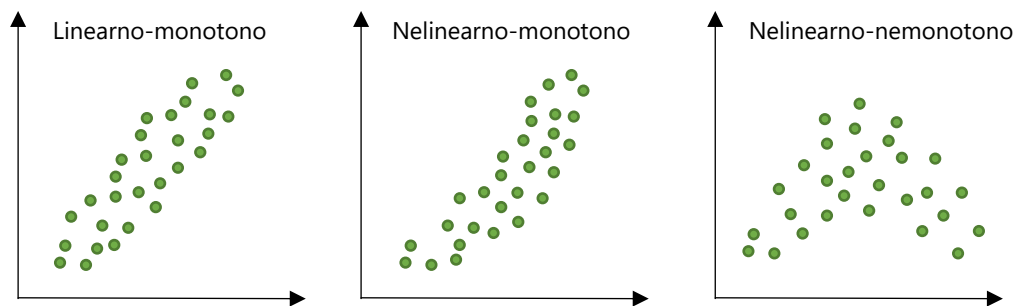


Slika 27: Basic Statistics → Correlations Analysis → Correlations

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 linearnom odnosu i imati normalnu ili Gaussovu distribuciju.

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

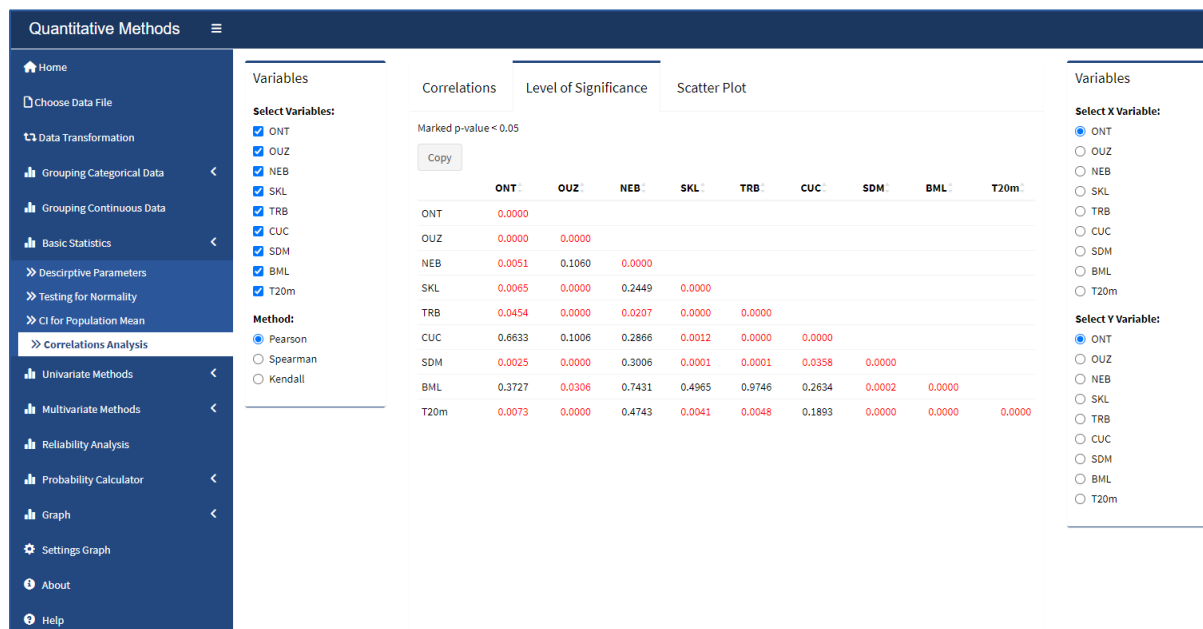


Spearmanov koeficijent rang korelacije izračunava se prema formuli

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

gdje je

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



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



Slika 29: Basic Statistics → Correlations Analysis → Scatter Plot

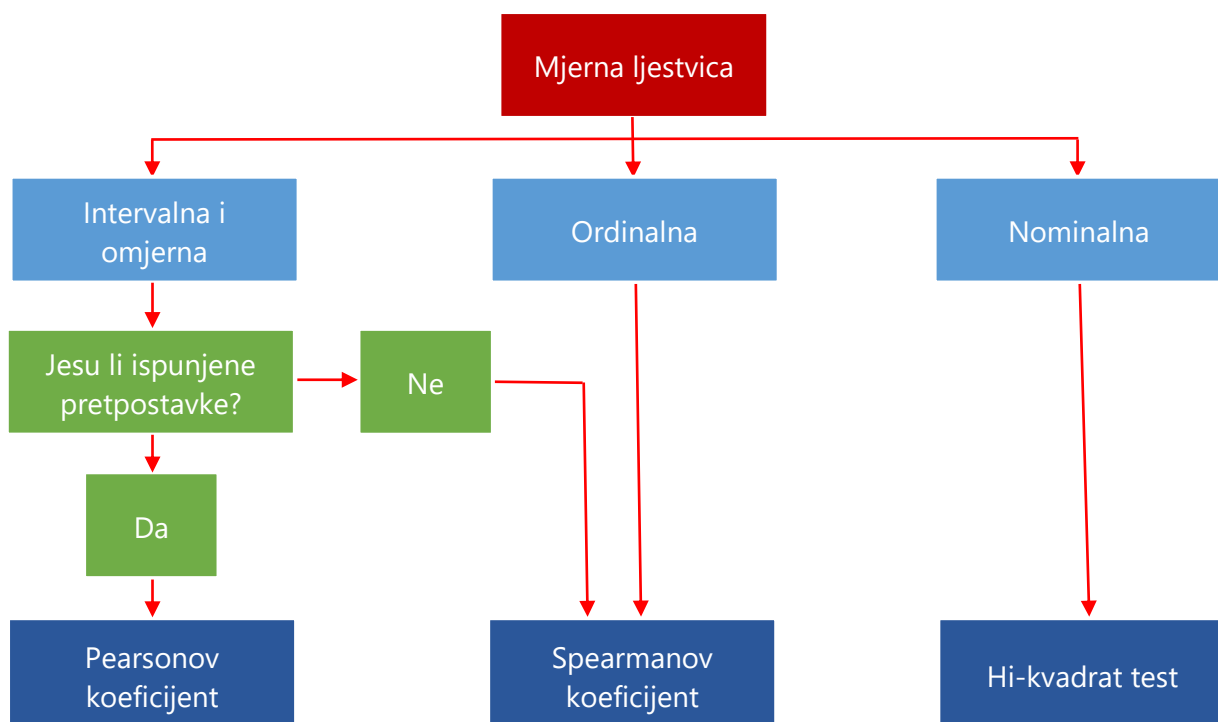
CILJ

- ✓ 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 ($n > 30$).

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



Univariate Methods

Independent Samples T-Test

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

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

The screenshot shows the 'Quantitative Methods' application interface. On the left is a sidebar menu with options like 'Home', 'Choose Data File', 'Data Transformation', 'Grouping Categorical Data', 'Grouping Continuous Data', 'Basic Statistics', 'Univariate Methods', 'Multivariate Methods', 'Reliability Analysis', 'Probability Calculator', 'Graph', 'Settings Graph', 'About', and 'Help'. The 'Univariate Methods' section is expanded, showing 'Independent Samples T-Test' as the selected option. The main area is divided into three tabs: 'T-test (more variables)', 'Descriptive Parameters', and 'T-test (one variable)'. The 'T-test (more variables)' tab is active, displaying a table of results. The table has columns for variables (ATV, ATT, AOP, ANN, MKUS, MPOL, MP20, MPRR, MTAP, MSDM, MDTR, MVIS), MEAN1, MEAN2, SED, T, DF, P, CI-95%, CI+95%, Levene's F, P, and Cohen's D. The data rows show the results for each variable, with KRAJ selected as the grouping variable.

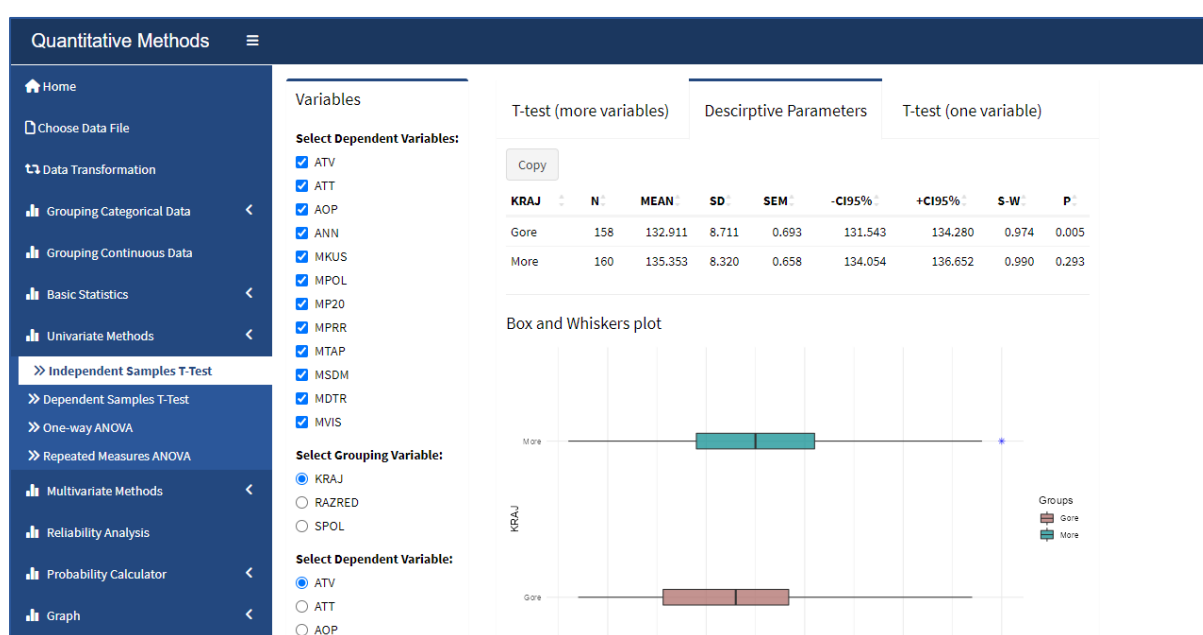
	MEAN1	MEAN2	SED	T	DF	P	CI-95%	CI+95%	Levene's F	P	Cohen's D
ATV	132.911	135.353	0.955	2.556	316	0.011	-4.321	-0.562	0.572	0.45	0.287
ATT	29.91	31.472	0.754	2.072	316	0.039	-3.046	-0.079	0.542	0.462	0.232
AOP	19.297	19.471	0.191	0.911	316	0.363	-0.551	0.202	0.204	0.652	0.102
ANN	11.956	11.981	0.525	0.049	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).

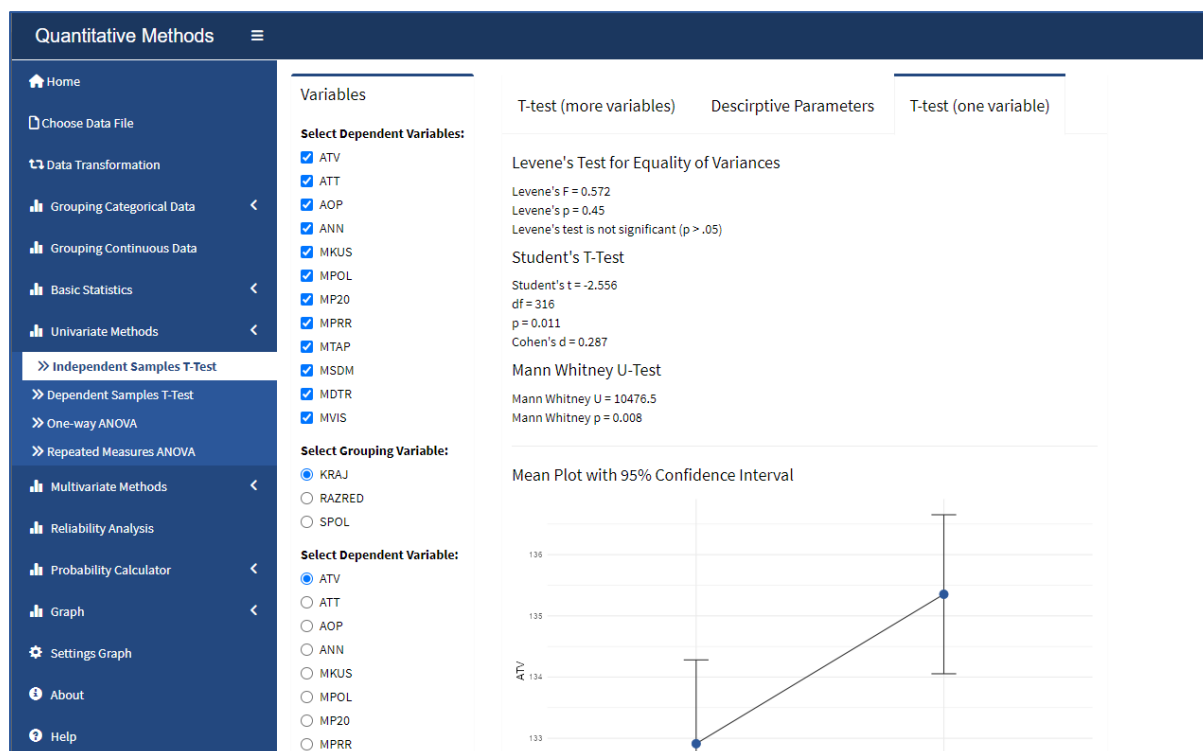


Slika 31: Univariate Methods → Independent Samples T-Test → Descriptive Parameters

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%.

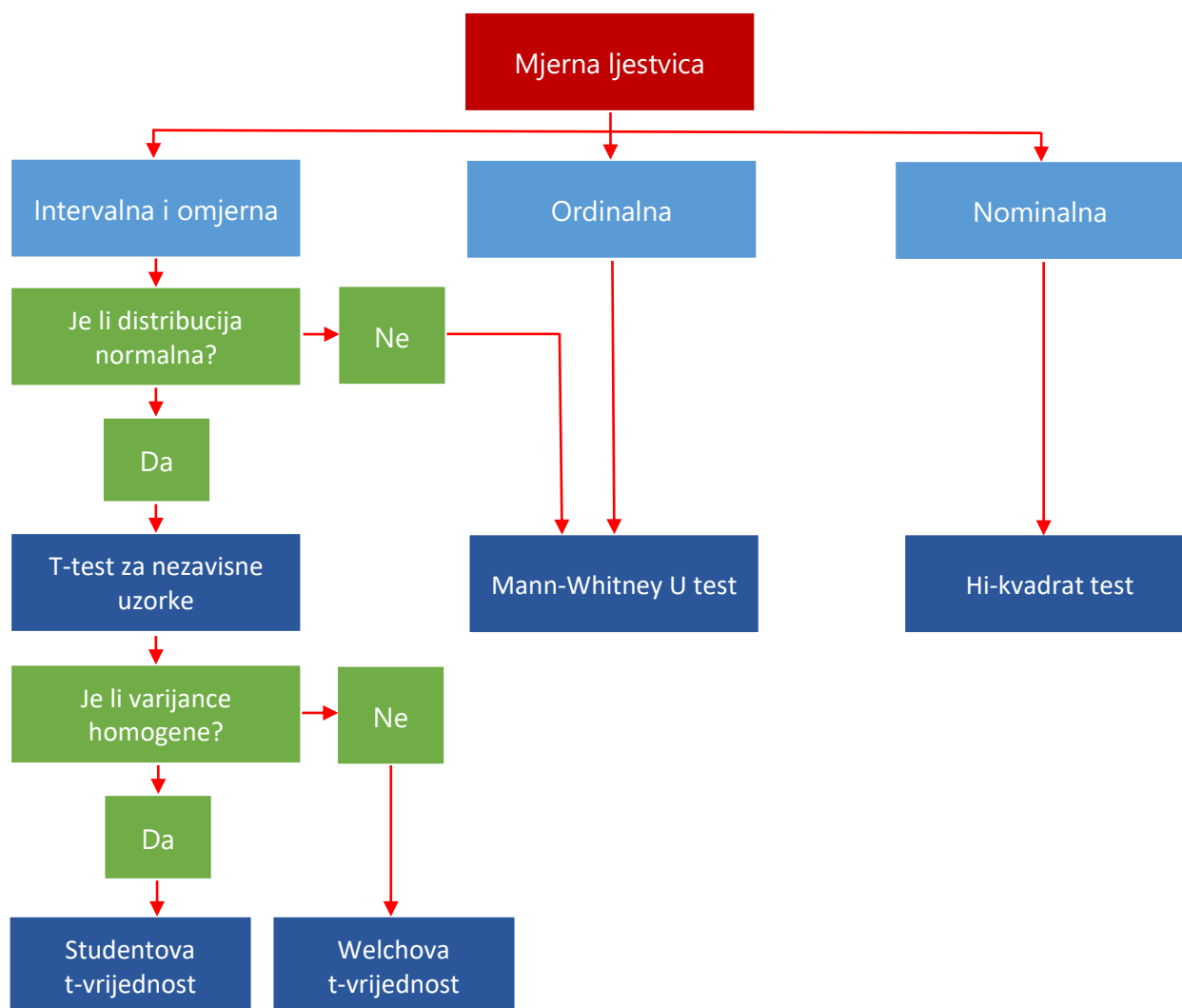
CILJ

- ✓ uz pogrešku p utvrditi razlikuju 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 mjernoj ljestvici
- ✓ zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjernoj mjernoj 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 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.

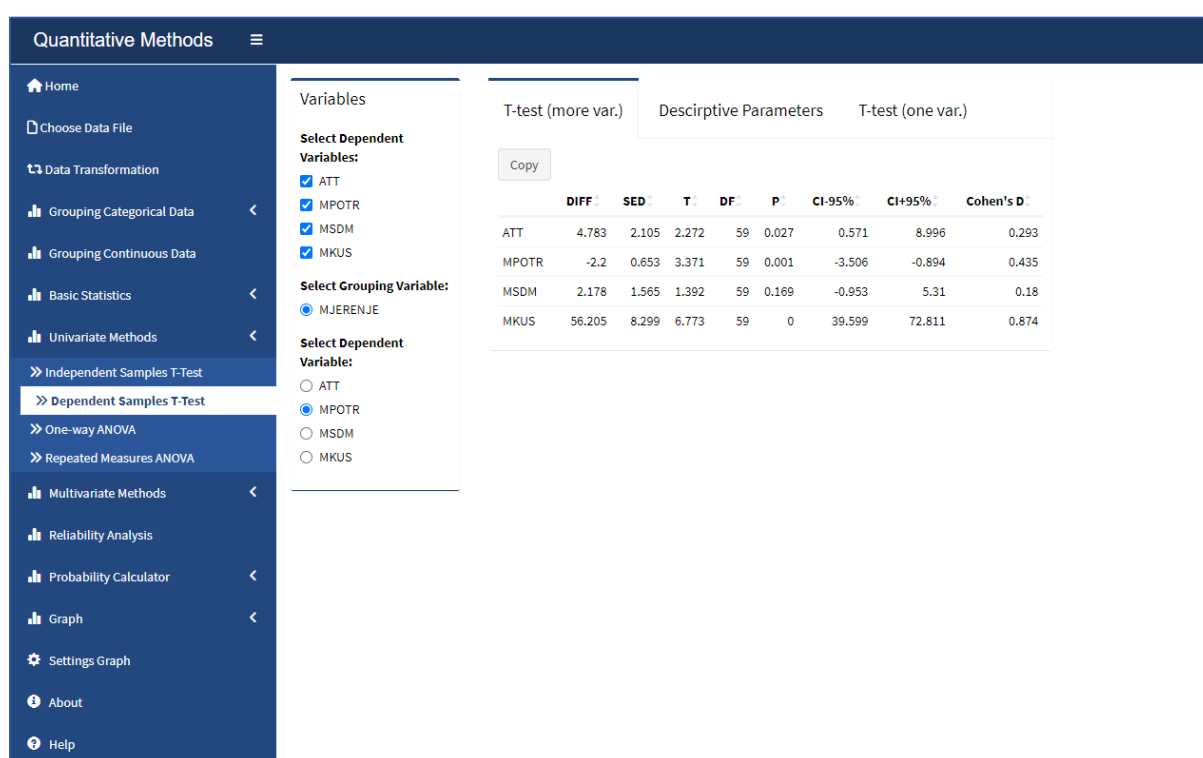
Testiranje razlika između dviju nezavisnih grupa, ovisno o vrsti i distribuciji varijabli.



Dependent Samples T-Test

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

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



The screenshot shows the 'Quantitative Methods' web application. The sidebar on the left contains navigation links: Home, Choose Data File, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods, Independent Samples T-Test, **Dependent Samples T-Test** (highlighted), One-way ANOVA, Repeated Measures ANOVA, Multivariate Methods, Reliability Analysis, Probability Calculator, Graph, Settings Graph, About, and Help.

The main content area is titled 'Variables' and shows the results of a 'T-test (more var.)'. The 'Select Dependent Variables' section lists ATT, MPOTR, MSDM, and MKUS, all of which are selected. The 'Select Grouping Variable' section lists MJERENJE, which is selected. The 'Select Dependent Variable' section lists ATT, MPOTR, MSDM, and MKUS, all of which are unselected.

The results table is as follows:

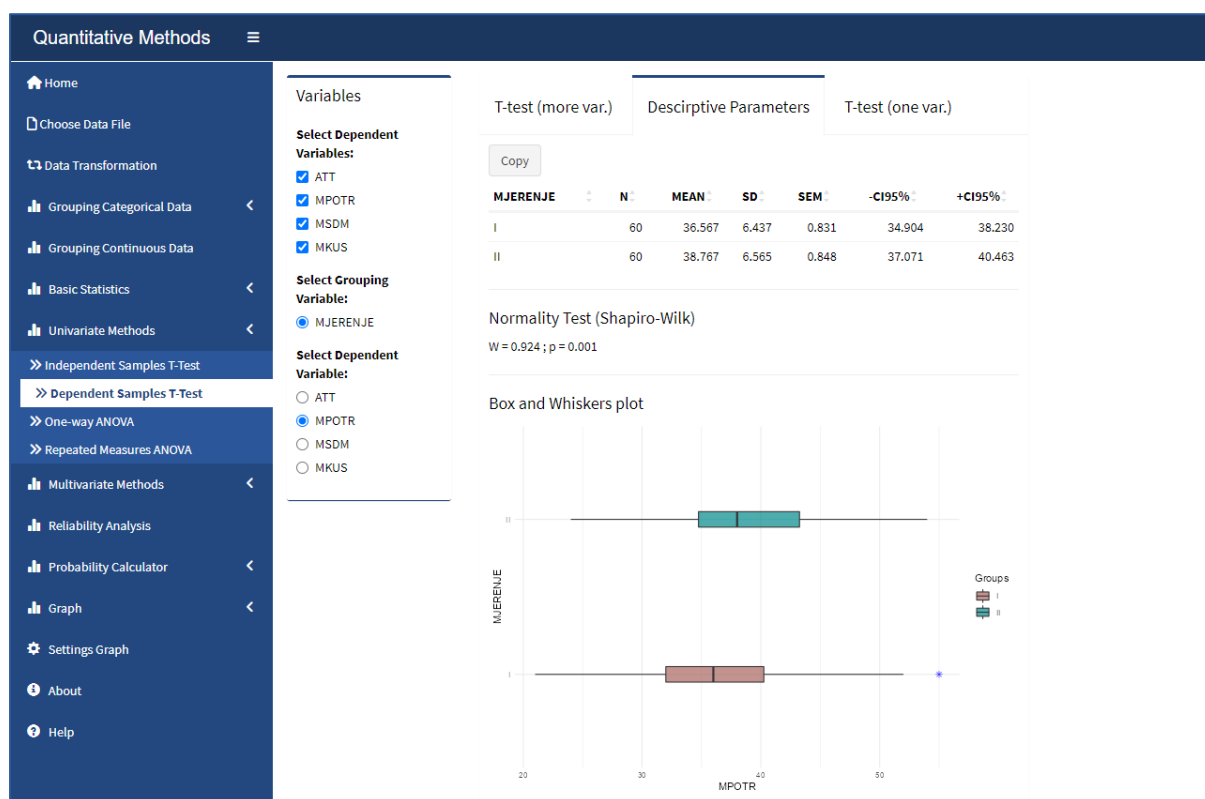
	DIFF	SED	T	DF	P	CI-95%	CI+95%	Cohen's D
ATT	4.783	2.105	2.272	59	0.027	0.571	8.996	0.293
MPOTR	-2.2	0.653	3.371	59	0.001	-3.506	-0.894	0.435
MSDM	2.178	1.565	1.392	59	0.169	-0.953	5.31	0.18
MKUS	56.205	8.299	6.773	59	0	39.599	72.811	0.874

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 mjerenja
- **SED** – standardna pogreška razlika aritmetičkih sredina prvog i drugog mjerenja
- **T** – Studentova t-vrijednost koja pokazuje koliko je puta razlika između aritmetičkih sredina veća od standardne pogreške razlika.
- **DF** – broj stupnjeva slobode ($n - 1$, gdje je n broj entiteta)
- **P** – pogreška kojom tvrdimo da je razlika između aritmetičkih sredina prvog i drugog mjerenja statistički značajna.
- **CI-95%** – donja granica intervala u kojem se s vjerojatnošću 95 % nalazi razlika aritmetičkih sredina populacije prvog i drugog mjerenja

- **CI+95%** – gornja granica intervala u kojem se s vjerojatnošću 95 % nalazi razlika aritmetičkih sredina populacije prvog i drugog mjerenja
- **Cohen's d** – Cohenova mjera veličine učinka (effect size).



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

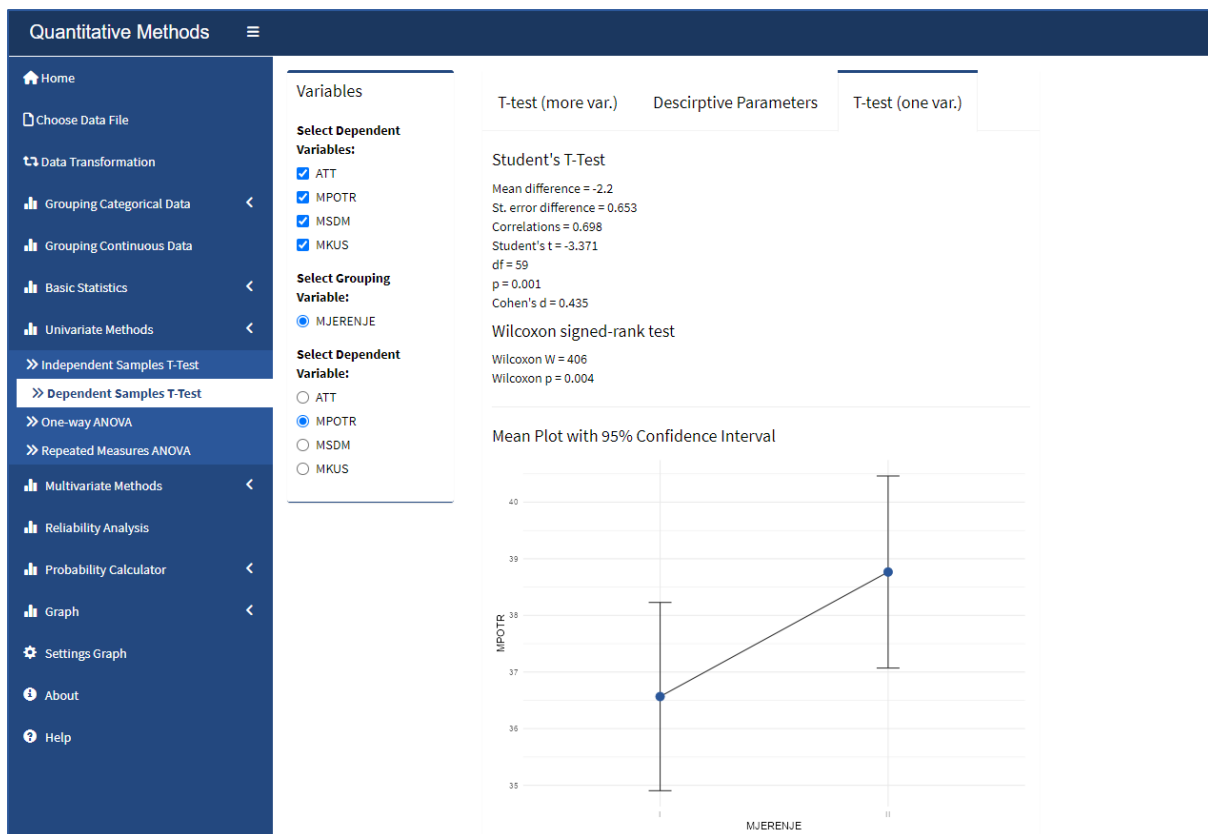
U kartici **Descriptive Parameters** prikazuje se tablica (slika 34) u kojoj je:

- **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 mjerenja statistički značajno razlikuje od normalne ili Gaussove distribucije:

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

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



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

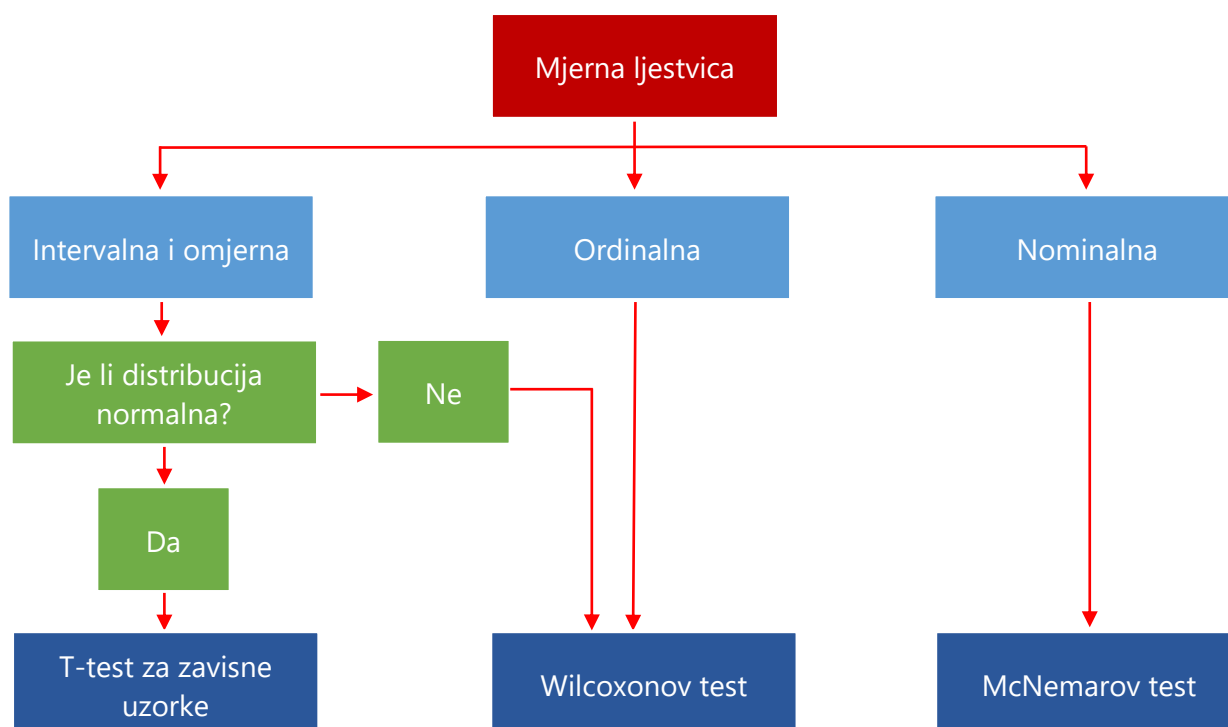
CILJ

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

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

The screenshot shows the 'Quantitative Methods' web application interface. The left sidebar contains a navigation menu with options like Home, Choose Data File, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods, Independent Samples T-Test, Dependent Samples T-Test, One-way ANOVA, Repeated Measures ANOVA, Multivariate Methods, Reliability Analysis, Probability Calculator, Graph, Settings Graph, About, and Help. The 'One-way ANOVA' option is selected. The main content area displays the 'ANOVA (more var.)' results. It includes a 'Variables' section with 'Select Dependent Variables' (ATV, ATT, AOP, ANN, MKUS, MPOL, MP20, MPRR, MTAP, MSDM, MDTR, MVIS) and 'Select Grouping Variable' (KRAJ, RAZRED, SPOL). The 'ANOVA (more var.)' tab is active, showing a table of results for each dependent variable. The table columns are Fisher's F, DF1, DF2, Fisher's p, Levene's F, and Levene's p. The data is as follows:

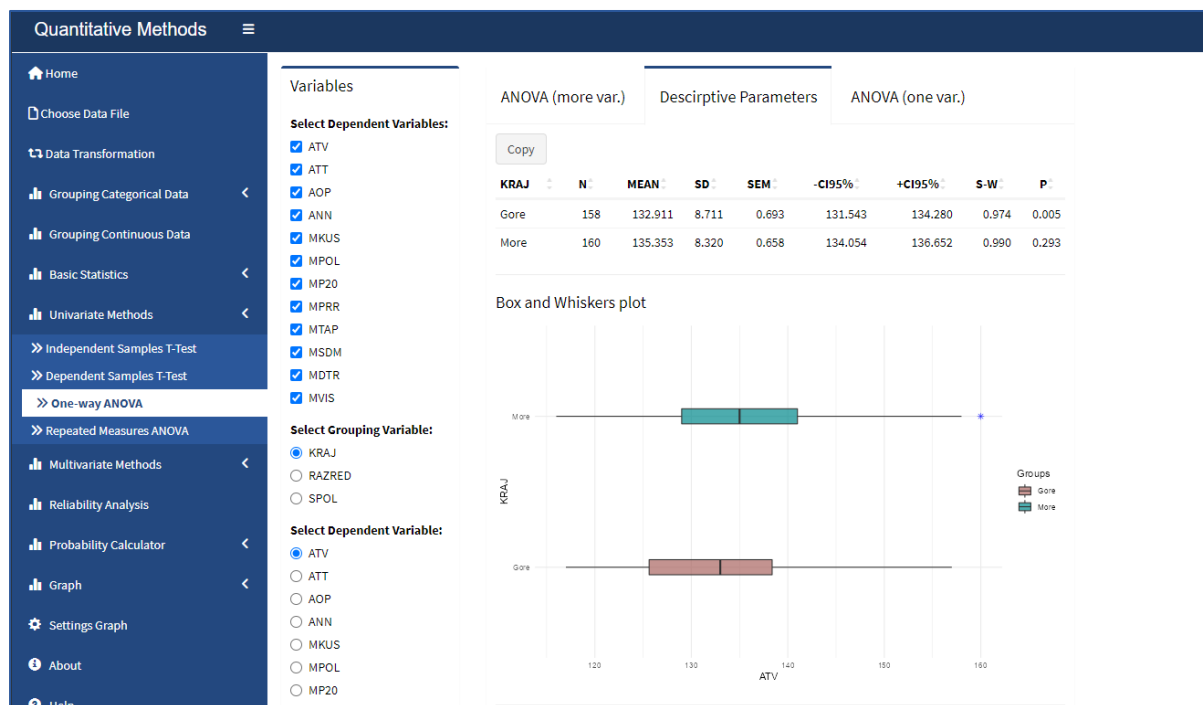
	Fisher's F	DF1	DF2	Fisher's p	Levene's F	Levene's p
ATV	6.534	1	316	0.011	0.572	0.45
ATT	4.294	1	316	0.039	0.542	0.462
AOP	0.83	1	316	0.363	0.204	0.652
ANN	0.002	1	316	0.961	4.163	0.042
MKUS	0.181	1	316	0.671	0.2	0.655
MPOL	1.849	1	316	0.175	2.194	0.14
MP20	1.265	1	316	0.262	0.761	0.384
MPRR	42.411	1	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 var.)

U kartici **ANOVA (more var.)** 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 stupnjeva slobode ($df_1 = k - 1$, $df_2 = n - k$, gdje je n broj entiteta, k – broj grupa)
- **Fisher's p** – pogreška kojom tvrdimo da je razlika između aritmetičkih sredina analiziranih grupa statistički značajna
- **Levene's F** – Levenova F-vrijednost temeljem koje se utvrđuje homogenost varijanci, odnosno jesu li varijance analiziranih grupa statistički značajno različite

- **Leven's p** – pogreška s kojom tvrdimo da su varijance analiziranih grupa statistički značajno različite. Ako je $p < 0,05$, zaključujemo da varijance analiziranih grupa nisu homogene, odnosno da se statistički značajno razlikuju uz pogrešku manju od 5%. U tom slučaju umjesto Fisherove F-vrijednosti, računa se Welchova F-vrijednost.

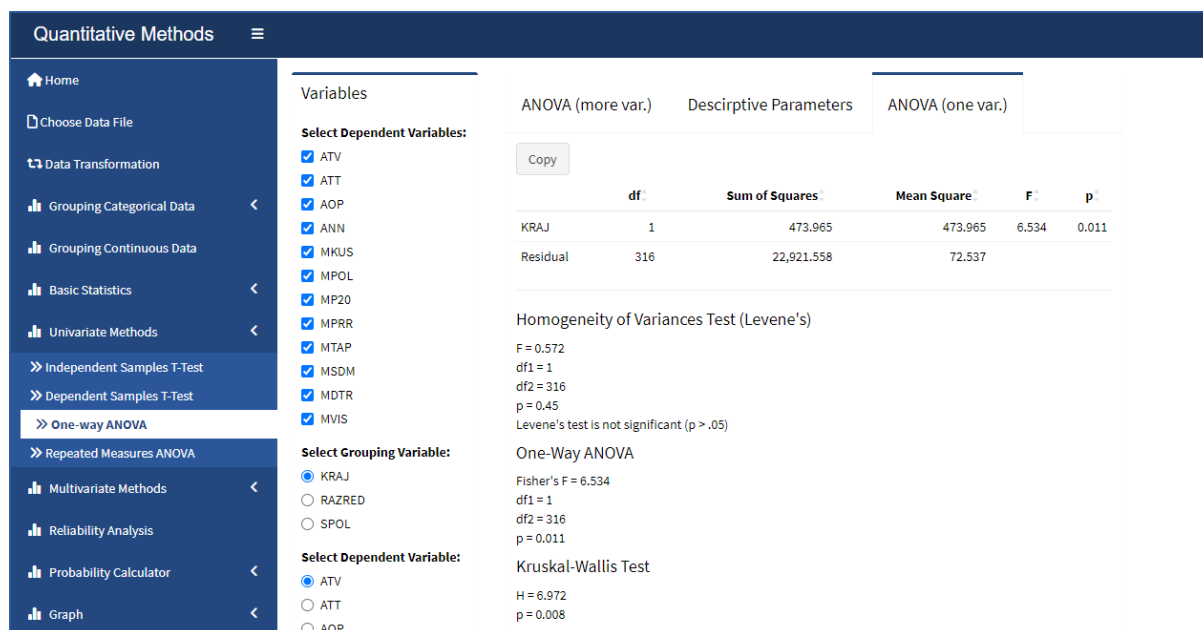


Slika 37: Univariate Methods → One Way ANOVA → Descriptive Parameters

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 (**Box and Whiskers plot** i **Density plot**) za analizirane grupe.



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

U kartici **ANOVA (one var.)** 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
- **df₁** i **df₂** – broj stupnjeva slobode ($df_1 = k - 1$, $df_2 = n - k$, gdje je n broj entiteta, k – broj grupa)
- **p** – pogreška kojom tvrdimo da je razlika između aritmetičkih sredina analiziranih grupa statistički značajna
- **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 uspoređivanje⁵

⁵Ako 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.).

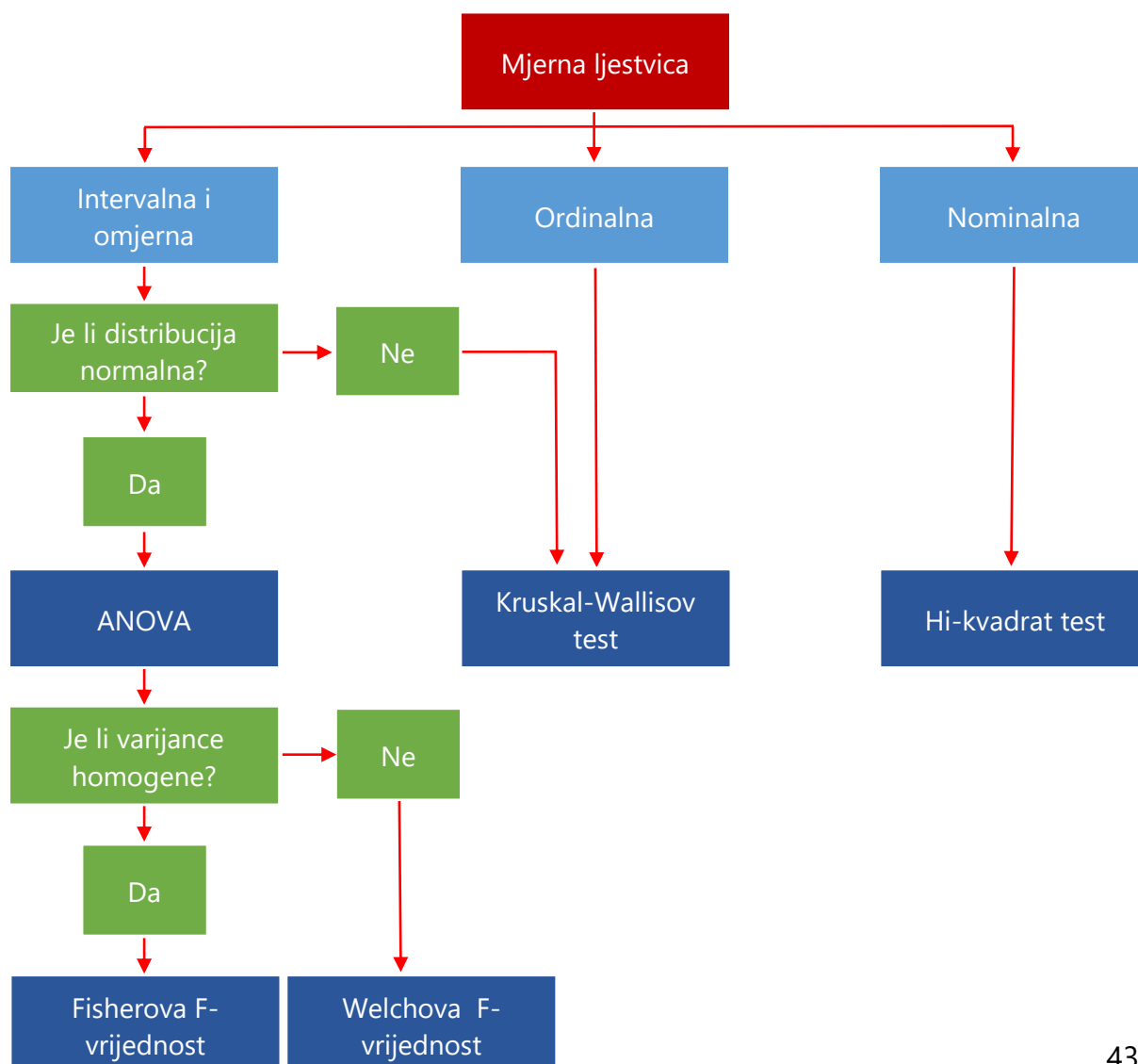
CILJ

- ✓ uz pogrešku p utvrditi razlikuju 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 mjernoj ljestvici s dva ili više modaliteta
- ✓ zavisnu varijablu čine podaci prikupljeni na intervalnoj ili omjernoj mjernoj 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.

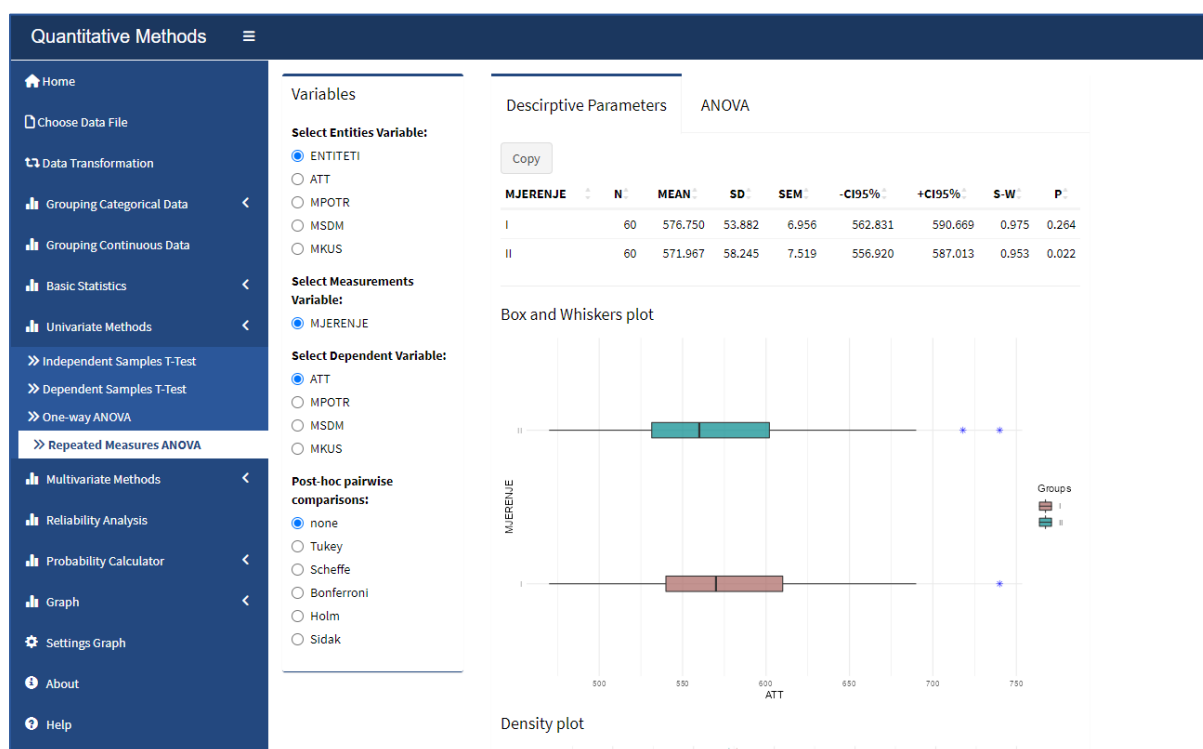
Testiranje razlika između dviju ili više nezavisnih grupa ovisno o vrsti i distribuciji varijabli.



Repeated Measures ANOVA

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

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



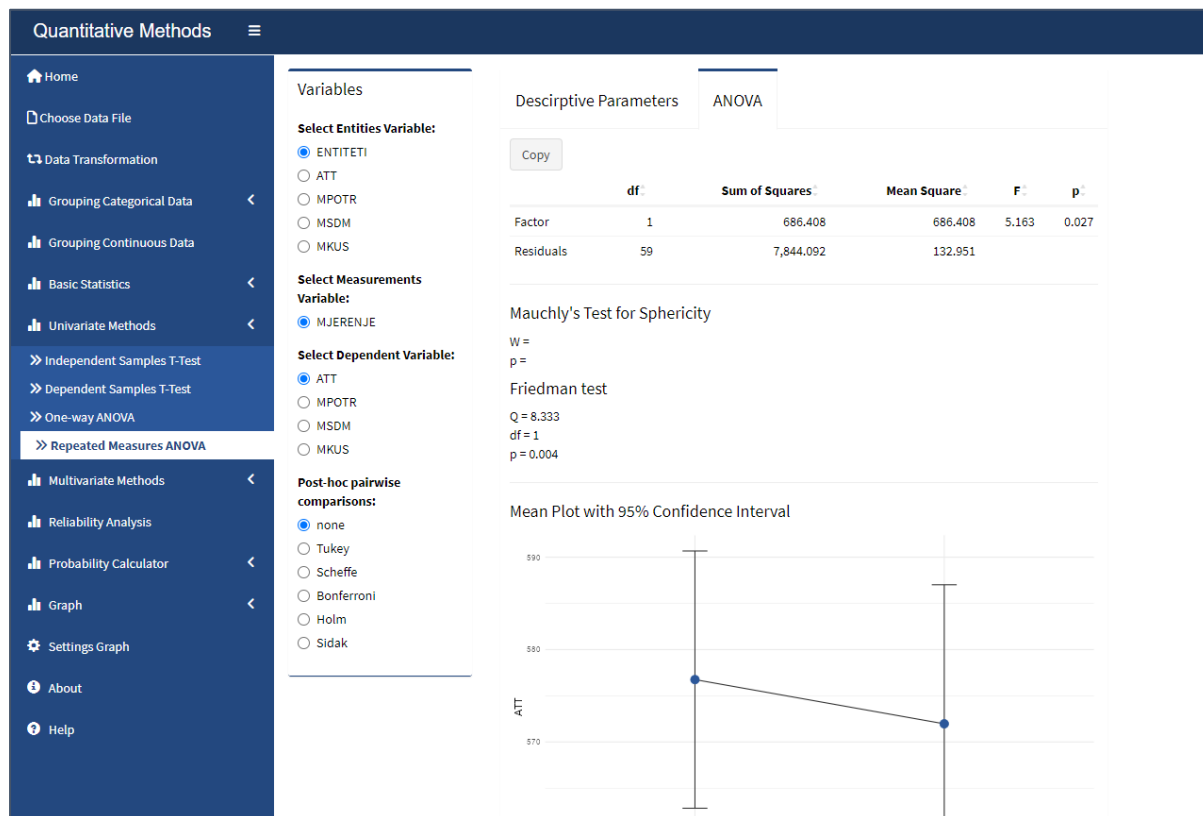
Slika 39: Univariate Methods → Repeated Measures ANOVA → Descriptive Parameters

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 (**Box and Whiskers plot** i **Density plot**) za analizirane grupe.



Slika 40: Univariate Methods → Repeated Measures ANOVA → ANOVA

Multivariate Methods

Regression Analysis

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

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

The screenshot shows the 'Quantitative Methods' web application. The sidebar menu on the left includes options like Home, Choose Data File, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods, Multivariate Methods, Regression Analysis (simple), Regression Analysis (multiple), Factor Analysis, Canonical Analysis, Discriminant Analysis, Cluster Analysis, Reliability Analysis, Probability Calculator, Graph, Settings Graph, About, and Help. The 'Regression Analysis (multiple)' option is selected. The main area shows the 'Variables' section with 'Select Independent Variables' (VISI, TEZI, OBPO, NABN, TAPI, POLI, SDAL, POTR, PRRA, IZVI, TR6M) and 'Select Dependent Variable' (VISI, TEZI, OBPO, NABN, TAPI, POLI, SDAL, POTR, PRRA, IZVI, TR6M). The 'Reg. Results' tab is active, displaying the following statistics: RO = 0.83; RO2 = 0.681; SEE = 133.48; F-value = 18.972; p-value = 0; Autocorrelation = 0.081; Durbin-Watson d = 1.83; p = 0.274. Below this is a table with 11 columns: B, SE(B), Beta, Part_R, R, P, Tolerance, t, and p. The table lists the regression coefficients for each variable.

	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)

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

- **RO** – koeficijent multiple korelacije
- **RO2** – koeficijent determinacije multiple korelacije
- **SEE** – standardna pogreška prognoze

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

te tablica u kojoj su prikazani sljedeći pokazatelji:

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

Osim toga, prikazuju se i grafikoni parcijalnih koeficijenata determinacije (**Graph (P)**) i standardiziranih regresijskih koeficijenata, parcijalnih korelacija i korelacija (**Beta, Partial R, R**).

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

CILJ

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

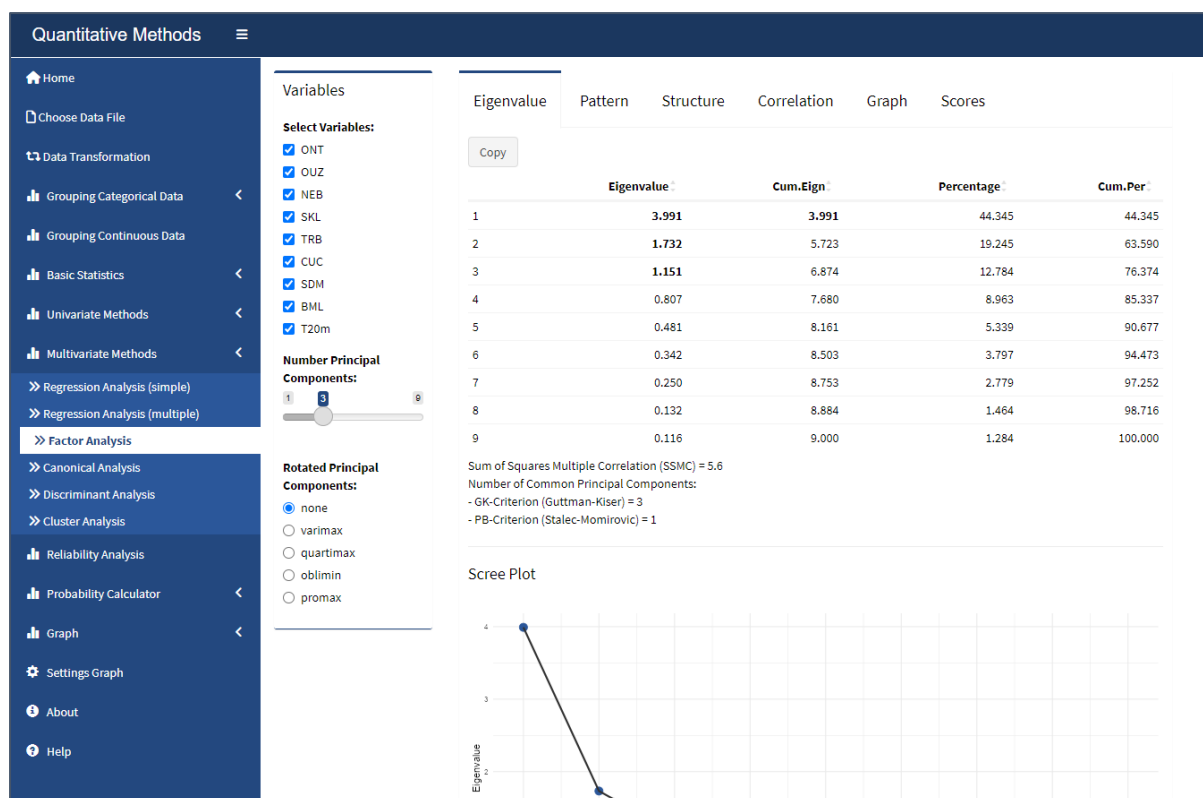
UVJETI

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

Factor Analysis

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

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



Slika 42: Multivariate Methods → Factor Analysis

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 **quatimax**) 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-Kaiser)** – Guttman-Kaiserovog kriterija prema kojem su značajne sve glavne komponente čija je varijanca, odnosno svojstvena vrijednost veća ili jednaka 1
- **PB-Criterion (Štalec-Momirović)** – PB – kriterija prema kojem je broj značajnih glavnih komponenata jednak broju svojstvenih vrijednosti poredanih po veličini, čiji zbroj ne prelazi SSMC (sumu kvadrata multiplih korelacija)
- **Scree Plot** – Na scree plotu se subjektivnom procjenom odredi točka nakon koje se svojstvene vrijednosti smanjuju u skladu s blagim linearnim trendom. Značajnima se smatraju sve prethodne glavne komponente.

U kartici:

- **Pattern** nalazi se matrica sklopa, odnosno paralelnih projekcija manifestnih varijabli na faktore te Hoffmanov index kompleksiteta (**Complexity**) svake manifestne varijable⁶. Indeks se kreće u intervalu od 1 do k , gdje je k broj faktora. Što je indeks bliži 1, to je kompleksitet varijable manji, a što je bliži broju faktora (k) to je veći. Ako je vrijednost indeksa jednaka 1, varijabla objašnjava samo 1 faktor. Ako je jednak k onda u jednakoj mjeri objašnjava svaki od k faktora

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

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

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

⁶Hofmann, R. J. (1978). Complexity and simplicity as objective indices descriptive of factor solutions. *Multivariate Behavioral Research*, 13(2), 247–250

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, 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 linearnom odnosu.
- ✓ varijable imaju normalnu ili Gaussovu distribuciju.

Canonical Analysis

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

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

The screenshot shows the 'Quantitative Methods' web application. The sidebar menu on the left includes options like Home, Choose Data File, Data Transformation, Grouping Categorical Data, Grouping Continuous Data, Basic Statistics, Univariate Methods, Multivariate Methods, Regression Analysis (simple), Regression Analysis (multiple), Factor Analysis, Canonical Analysis, Discriminant Analysis, Cluster Analysis, Reliability Analysis, Probability Calculator, Graph, Settings Graph, About, and Help. The 'Canonical Analysis' option is selected. The main area displays the 'Variables' section with two sets of variables: 'Select Variables 1. Set' (VISI, TEZI, OBPO, NABN, TAPI, POLI, SDAL, POTR, PRRA, IZVI, TR6M) and 'Select Variables 2. Set' (VISI, TEZI, OBPO, NABN, TAPI, POLI, SDAL, POTR, PRRA, IZVI, TR6M). The 'Canonical Factor' tab is active, showing a table of results. The table has columns for CF, Rc, Chi-sq., df, and p-level. The results are as follows:

CF	Rc	Chi-sq.	df	p-level
CF1	0.735	116.086	28	0
CF2	0.535	43.956	18	0.001
CF3	0.294	12.534	10	0.251
CF4	0.209	4.146	4	0.387

Slika 43: Multivariate Methods → Canonical Analysis

U kartici:

- **Canonical Factor** 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 korelacije statistički značajni.
- **Structure** nalaze se korelacija manifestnih varijabli prvog skupa s kanoničkim faktorima prvog skupa (**Factor Structure 1. Set**) i korelacija manifestnih varijabli drugog skupa s kanoničkim faktorima drugog skupa (**Factor Structure 2. Set**).

- **Graph** nalaze se grafikoni strukture kanoničkih faktora prvog (**Graf Factor Structure 1. Set**) i drugog prvog (**Graf Factor Structure 2. Set**) skupa.
- **Scores – 1. Set** i **Scores – 2. Set** nalaze se tablice s rezultatima entiteta u kanoničkim faktorima prvog i drugog skupa.

CILJ

- ✓ utvrditi relacije između dva skupa varijabli

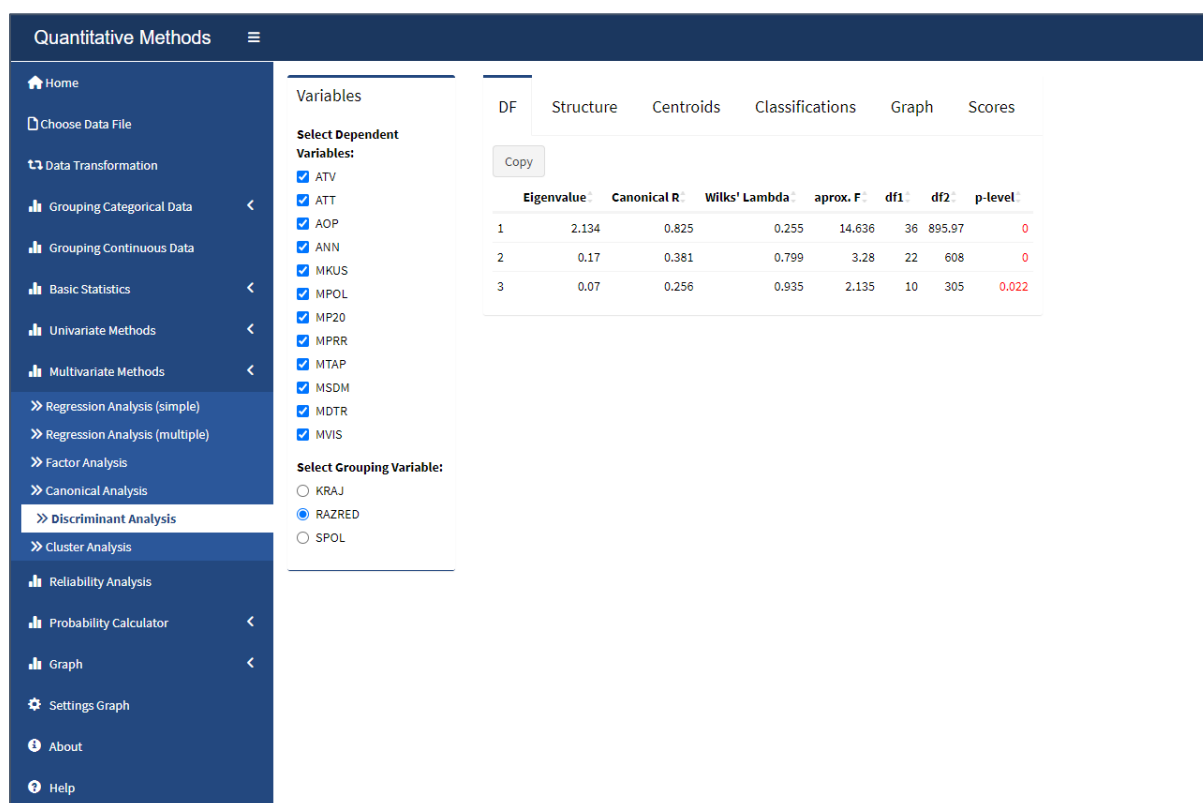
UVJETI

- ✓ 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 linearnom odnosu.
- ✓ varijable imaju normalnu ili Gaussovu distribuciju.

Discriminant Analysis

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

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



Slika 44: Multivariate Methods → Discriminant Analysis

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

- **Eigenvalue** – svojstvene vrijednosti, odnosno varijance diskriminacijskih funkcija
- **Canonical R** – koeficijenti kanoničke korelacije (diskriminacije), odnosno korelacija diskriminacijskih funkcija s nezavisnom (selektorskom) varijablom
- **Wilks' Lambda** – Wilksove lambde kreću se u intervalu od 0 do 1, a što im je vrijednost manja to je veća vjerojatnost da je razlika između analiziranih grupa statistički značajna
- **aprox. F** – aproksimativne F-vrijednosti temeljem kojih se utvrđuje statistička značajnost diskriminacijskih funkcija

- **df1** i **df2** – broj stupnjeva slobode
- **p-level** – pogreške kojima tvrdimo da pojedina diskriminacijska funkcija statistički značajno razlikuje analizirane grupe.

U kartici:

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

CILJ

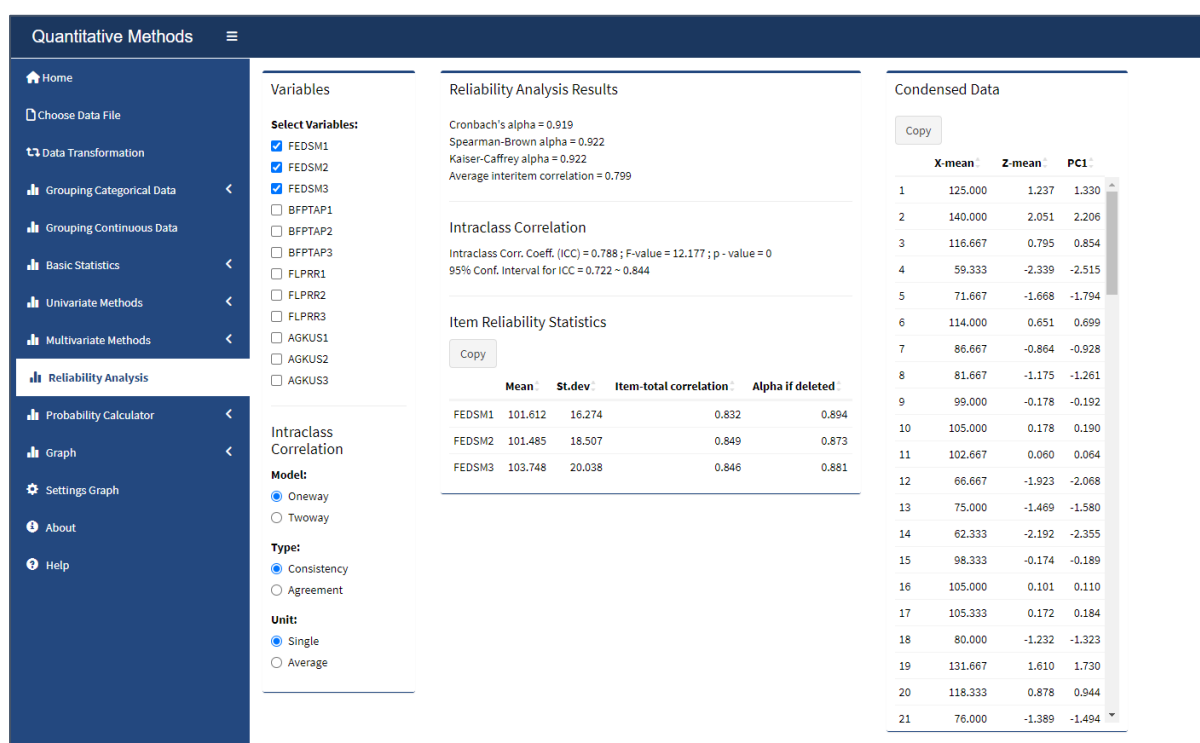
- ✓ utvrditi razlike između dvije ili više grupa entiteta opisanih s dvije ili više varijabli.

UVJETI

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

Reliability Analysis

Nakon pokretanja aplikacije Quantitative methods u web pregledniku se prikazuje glavni izbornik u okviru kojeg odaberemo opciju **Choose CSV Data File**. Klikom na gumb **Browse...** biramo mapu i datoteku u kojoj se nalaze podaci koje želimo statistički analizirati. Datoteka s podacima mora imati najmanje dvije 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).



Slika 45: 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)⁷ su definirali šest oblika ICC-a, koji se označavaju s dva broja u zagradama. Prvi broj se odnosi na ocjenjivače:

⁷ 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>

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

dok se drugi broj odnosi na konačnu ocjenu:

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

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 ICC8
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	$\frac{MS_R - MS_W}{MS_R}$

⁸MS_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.

ICC(2,k)	twoway, agreement, average	$\frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$
ICC(3,k)	twoway, consistency, average	$\frac{MS_R - MS_E}{MS_R}$

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

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

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

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

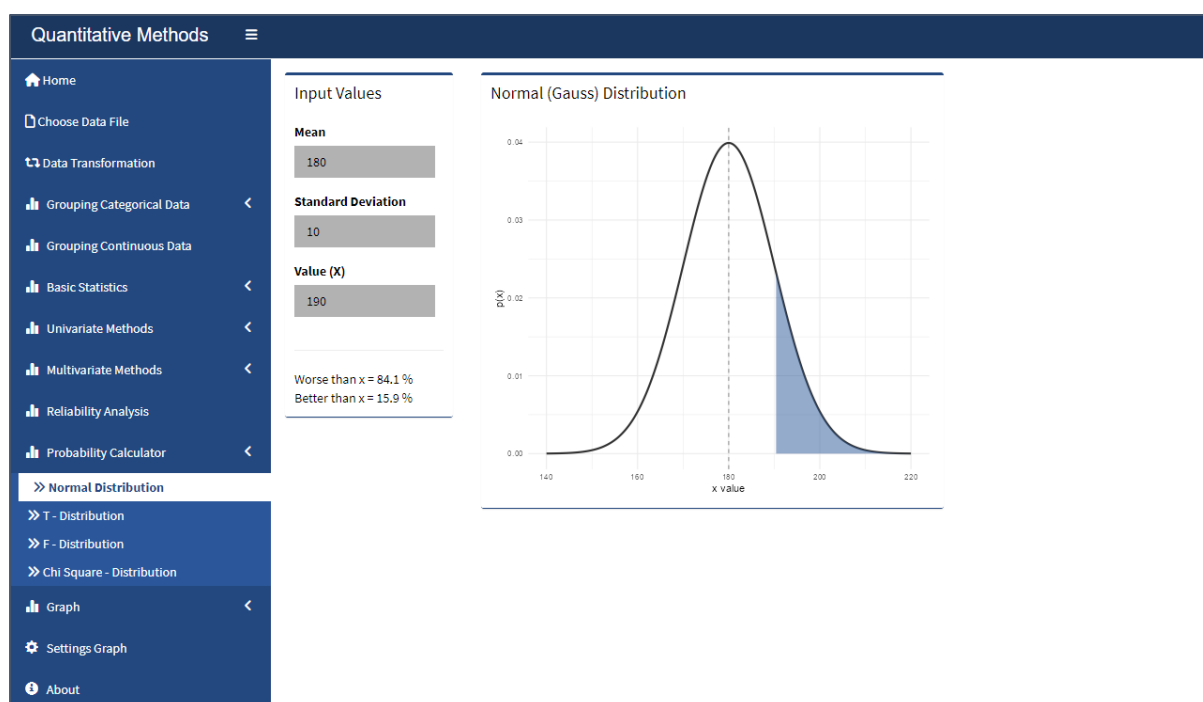
Koeficijenti pouzdanosti koji se utvrđuju metodom interne konzistencije odabiru se s obzirom na način kondenzacije rezultata, odnosno izračunavanje ukupnog rezultata ispitanika na temelju pripadajućih rezultata u česticama testa. Ako se kondenzacija rezultata nekog kompozitnog testa vrši:

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

Probability Calculator

Normal Distribution

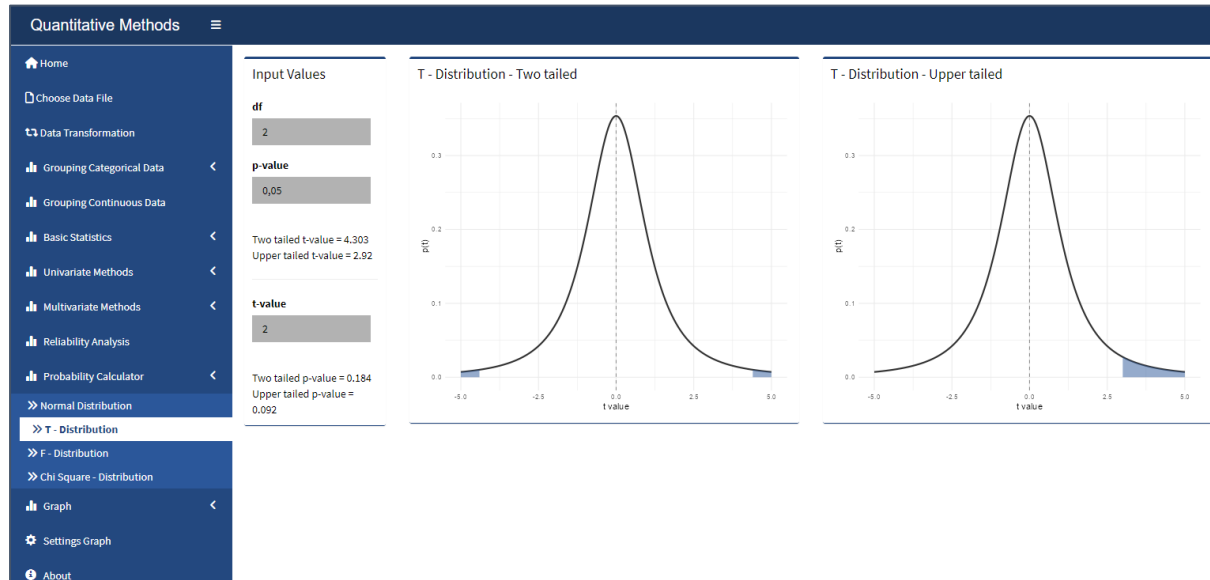
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **Normal Distribution** omogućava računanje postotka boljih (**Better than x**) i lošijih (**Worse than x**) entiteta od određenog rezultata (**X value**) u populaciji entiteta čije obilježje (varijabla) ima normalnu ili Gaussovu distribuciju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



Slika 46: Probability Calculator → Normal Distribution

T – Distribution

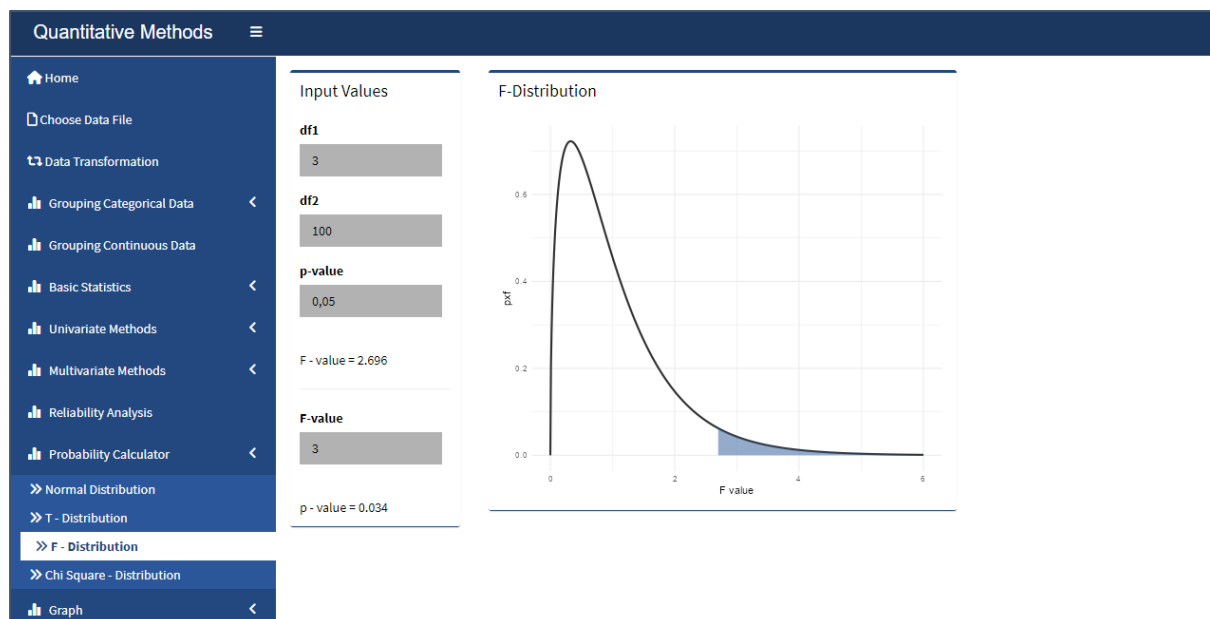
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **T-Distribution** omogućava računanje odgovarajuće t-vrijednosti (**Two tailed t-value** i **Upper tailed t-value**) za zadanu p-vrijednost (**p-value**) i broj stupnjeva slobode (**df**).



Slika 47: Probability Calculator → T- Distribution

F – Distribution

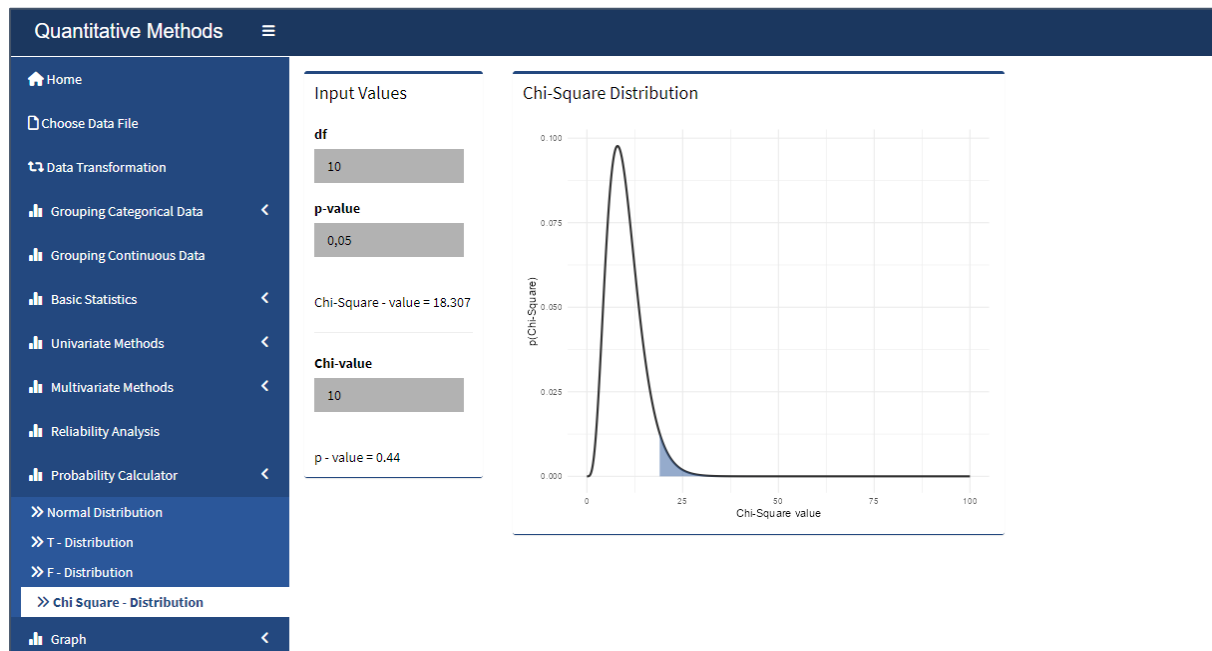
Odabir opcije **Probability Calculator** glavnog izbornika te opcije **F-Distribution** omogućava računanje odgovarajuće F-vrijednosti (**F-value**) za zadanu p-vrijednost (**p-value**) i broj stupnjeva slobode (**df1** i **df2**).



Slika 48: Probability Calculator → F- Distribution

Chi Square – Distribution

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

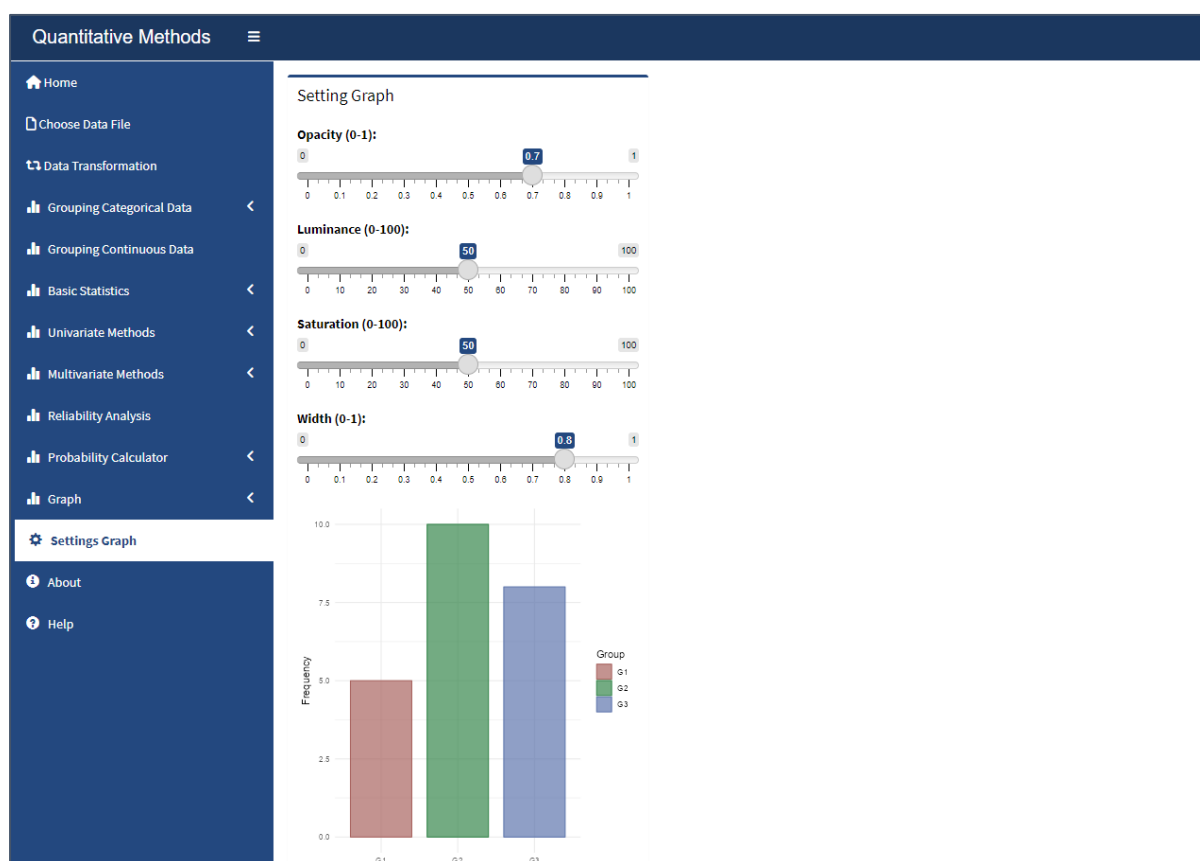


Slika 49: Probability Calculator → Chi Square- Distribution

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 52: Settings Graph