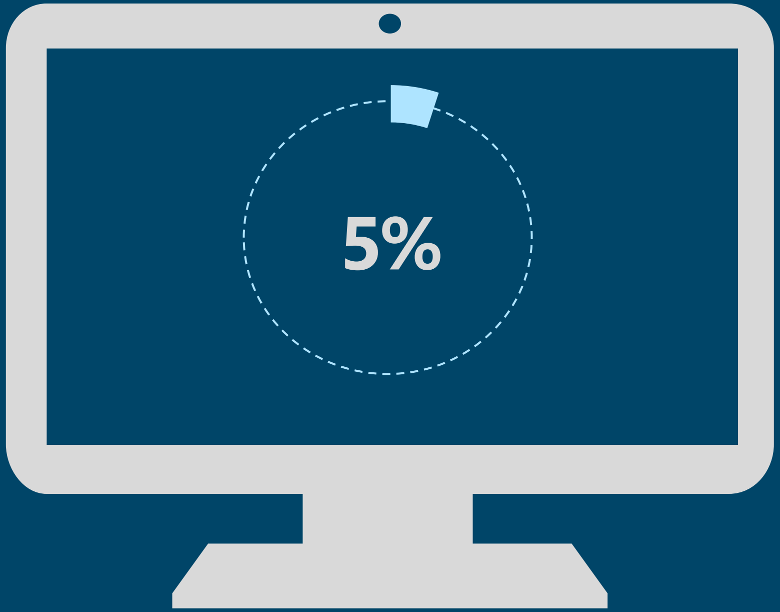
VJEŽBE IZ KVANTITATIVNIH METODA

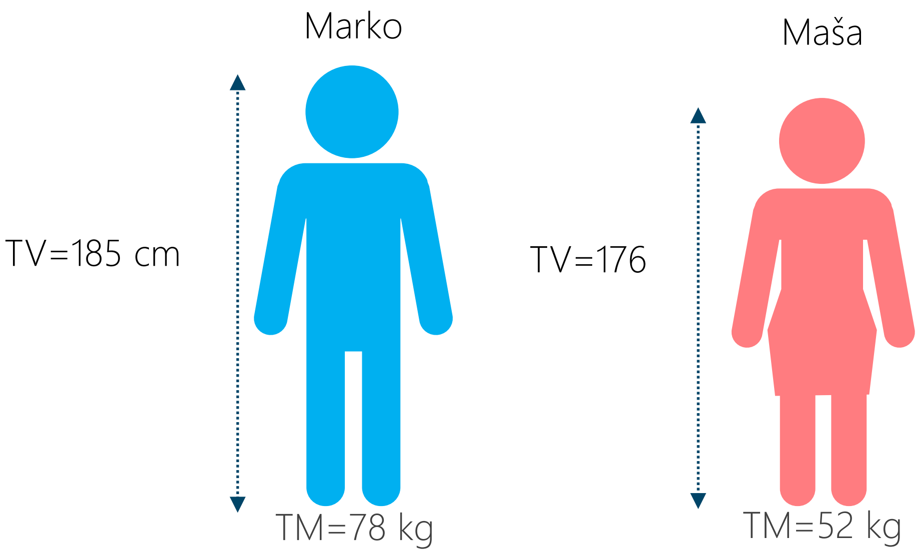
OSNOVNI STATISTIČKI POJMOVI

Vježba 1



Osnovni statistički pojmovi

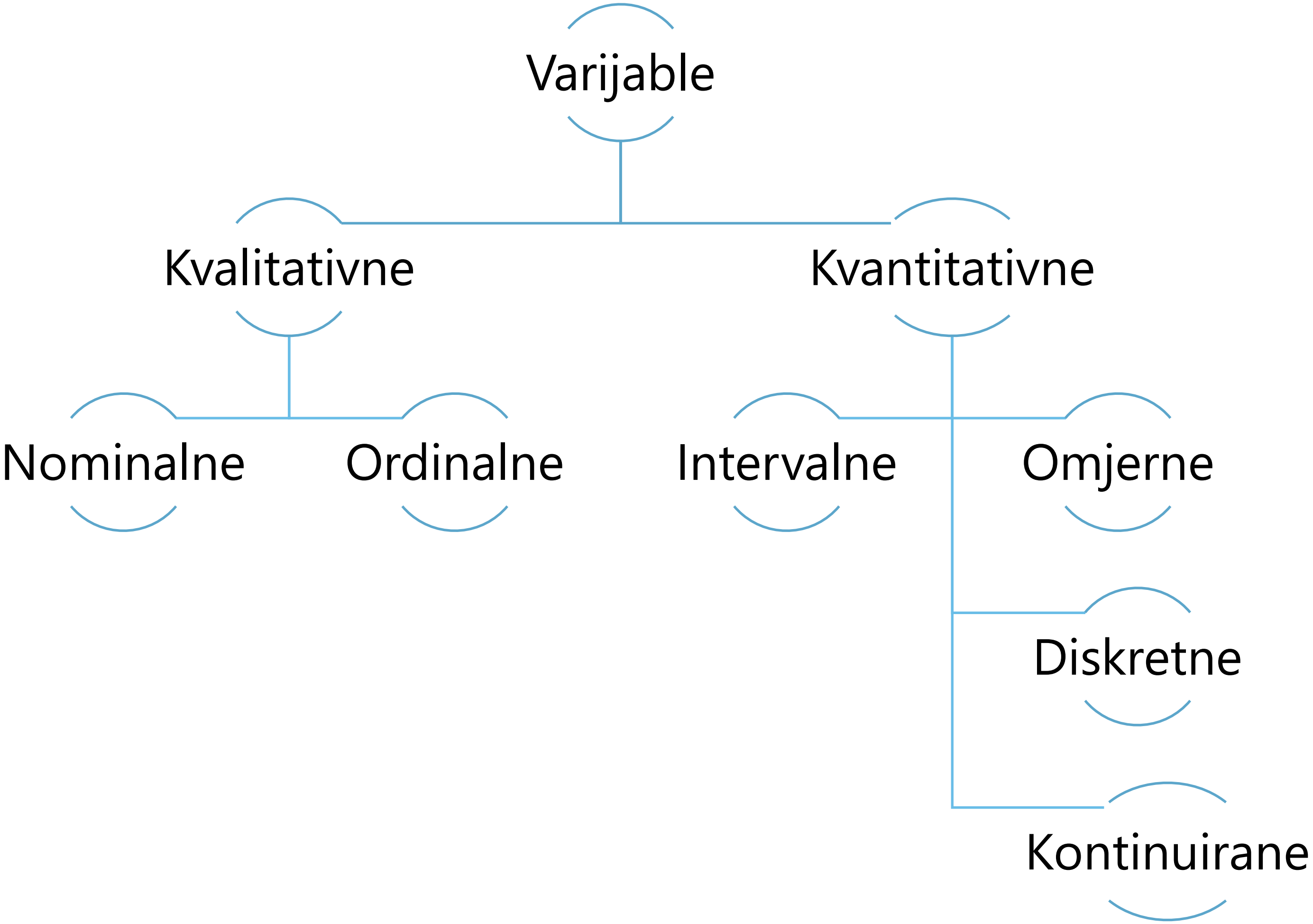
1. **Statistika** je skup metoda za uređivanje, analiziranje i grafičko prikazivanje podataka.
2. **Podatak** je kvantitativna ili kvalitativna vrijednost kojom je opisano određeno obilježje nekog entiteta.
3. **Entitet** je jedinka nekog skupa osoba,stvari, pojava i sl. koja je opisana određenim obilježjima.
4. **Varijabla** je svojstvo, obilježje, osobina, sposobnost, znanje, itd. po kojem se entiteti međusobno razlikuju.



ENTITET	SPOL	TV	TM
Marko	M	185	78
Maša	Ž	176	52
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

Osnovni statistički pojmovi

5 Vrste varijabli



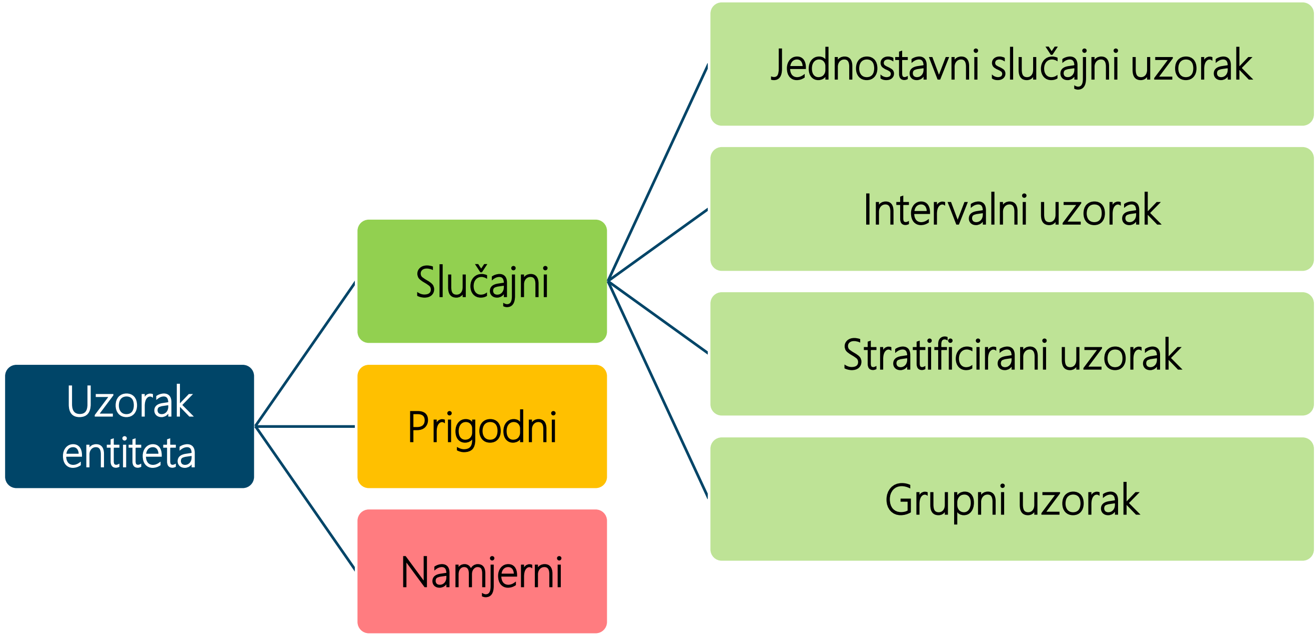
- Varijable
  - Kvalitativne
    - Nominalne
    - Ordinalne
  - Kvantitativne
    - Intervalne
    - Omjerne
    - Diskretne
    - Kontinuirane

Osnovni statistički pojmovi

6 **Populacija entiteta** je skup svih entiteta čija su obilježja predmet statističke analize.

7 **Uzorak entiteta** je podskup entiteta izabran iz populacije s ciljem da je što bolje reprezentira.

8 Vrste uzoraka entiteta



Uzorak entiteta

Slučajni

Prigodni

Namjerni

Jednostavni slučajni uzorak

Intervalni uzorak

Stratificirani uzorak

Grupni uzorak

Osnovni statistički pojmovi

9 **Populacija varijabli** predstavlja skup svih mogućih varijabli kojima se može opisati stanje nekog entiteta.

10 **Uzorak varijabli** je podskup varijabli izabran na temelju neke teorije iz populacije varijabli.

11 **Matrica podataka** je skup podataka dobivenih opisom skupa entiteta nekim skupom varijabli smještenih tako da svaki redak sadrži podatke kojima je pojedini entitet opisan svim varijablama, dok svaki stupac sadrži podatke svih entiteta u pojedinoj varijabli.

ENTITET	SPOL	TV	TM
Marko	M	185	78
Maša	Ž	176	52
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

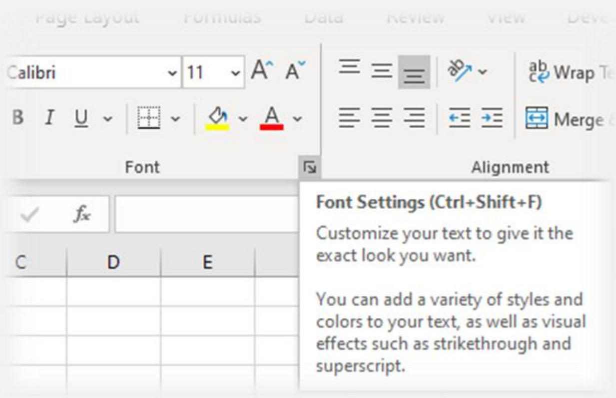
Microsoft Excel

Zadatak 1: Kreirajte sljedeću matricu podataka:

Entiteti	SPOL	TV	TM
Šime	m	192	87
Marko	m	174	68
Mate	m	186	75

1. ***UNOS PODATAKA:*** Podaci se unose putem tipkovnice pri čemu je moguće iskoristiti opcije *Copy* i *Paste* ili hvataljku za kopiranje sadržaja označenog polja u susjedna polja.

2. ***FORMATIRANJE ĆELIJA:*** Dijaloški okvir za formatiranje ćelija pokreće se odabirom opcije *Font Setting* izbornika *Home*. Može se odabrati tip podataka (*Number*), vršiti poravnavanje (*Alignment*) te izvršiti odabir obilježja fonta (*Font*), rubova polja (*Borders*) i pozadine polja (*Patterns*).



Font Settings (Ctrl+Shift+F)

Customize your text to give it the exact look you want.

You can add a variety of styles and colors to your text, as well as visual effects such as strikethrough and superscript.

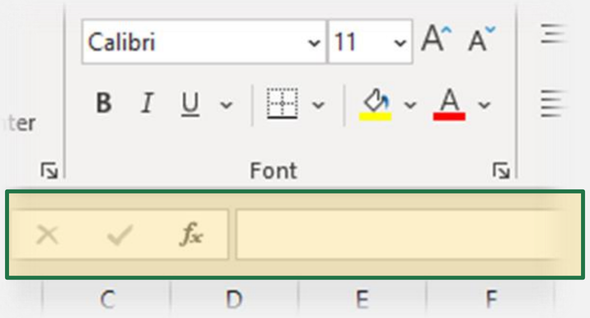
Microsoft Excel

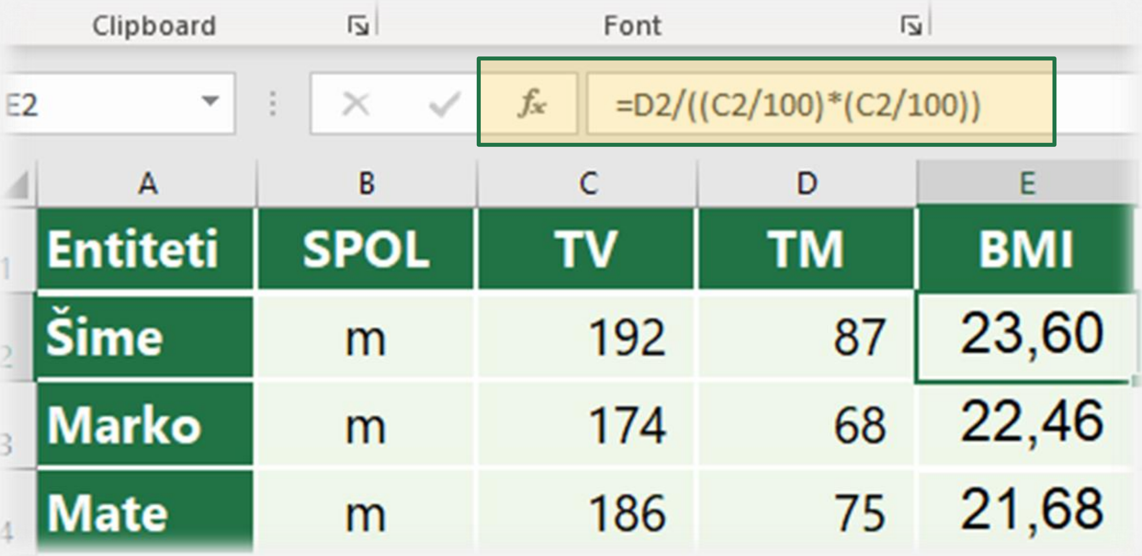
Zadatak 2: Dodati varijablu BMI i putem formule izračunati podatke.

Entiteti	SPOL	TV	TM	BMI
Šime	m	192	87	
Marko	m	174	68	
Mate	m	186	75	

$$\text{BMI}=\frac{TM(kg)}{TV(m)^2}$$

3 ***UNOS FORMULE:*** Formule za izračunavanje vrijednosti označenog polja se unose u traku *fx (npr. =C1/B1)*. U svrhu kopiranja formula iz polja u polje (opcije *Copy* i *Paste*) moguće je koristiti relativne (npr. A22), apsolutne ($A$2) ili kombinirane adrese ($A2), zavisno o potrebi.





=D2/((C2/100)*(C2/100))

	A	B	C	D	E
1	Entiteti	SPOL	TV	TM	BMI
2	Šime	m	192	87	23,60
3	Marko	m	174	68	22,46
4	Mate	m	186	75	21,68

Microsoft Excel

Zadatak 3: Pretvori kreirani raspon podataka u tablicu

Entiteti	SPOL	TV	TM	BMI
Šime	m	192	87	23.60
Marko	m	174	68	22.46
Mate	m	186	75	21.68

4 ***STVARANJE TABLICE:*** Nakon što označimo raspon s podacima koje želimo pretvoriti u tablicu, odaberemo opciju *Table* u izborniku *Insert*.

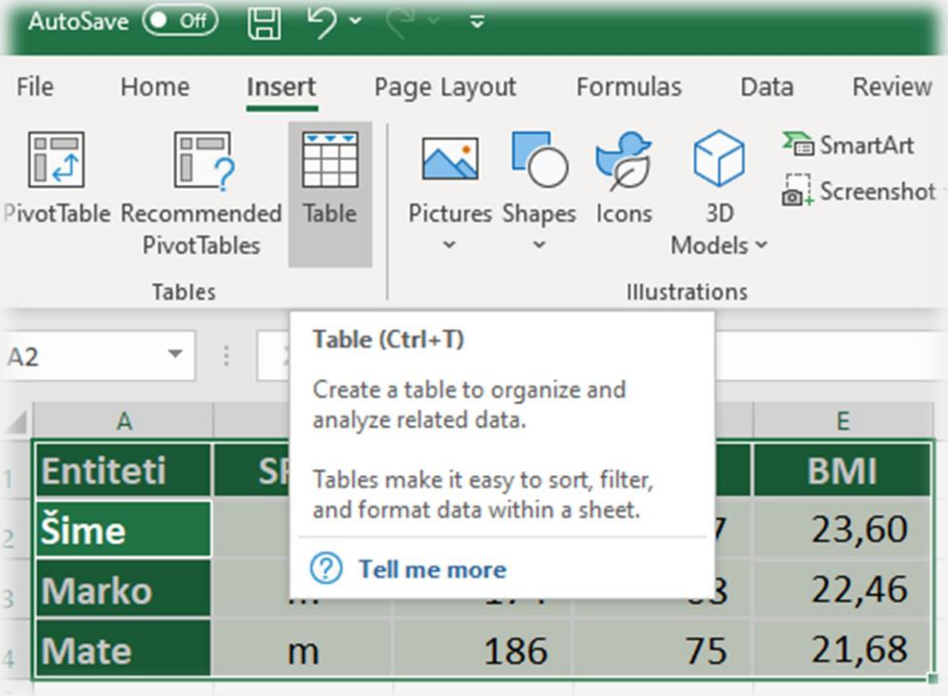


Table (Ctrl+T)

Create a table to organize and analyze related data.

Tables make it easy to sort, filter, and format data within a sheet.

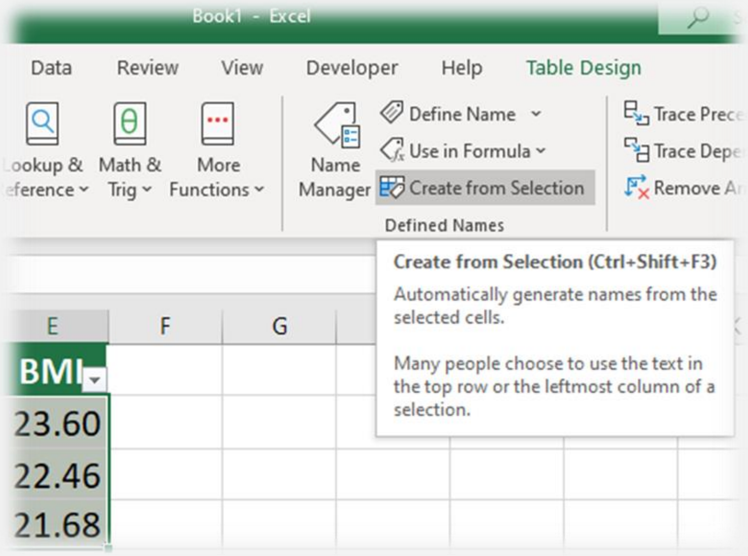
Tell me more

Microsoft Excel

Zadatak 4: Definiraj varijable SPOL, TV, TM i BMI u tablici s podacima.

Entiteti	SPOL	TV	TM	BMI
Šime	m	192	87	23.60
Marko	m	174	68	22.46
Mate	m	186	75	21.68

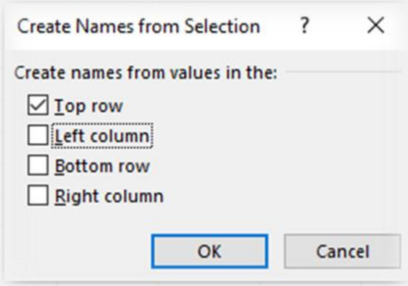
5 ***STVARANJE VARIJABLI:*** Nakon što označimo raspon s podacima (uključujući i nazive varijabli) koje želimo pretvoriti u varijable, odaberemo opciju *Create from Selection* u izborniku *Formulas te* odaberemo opciju *Top row.*



Create from Selection (Ctrl+Shift+F3)

Automatically generate names from the selected cells.

Many people choose to use the text in the top row or the leftmost column of a selection.



Create names from values in the:

☑ Top row
☐ Left column
☐ Bottom row
☐ Right column

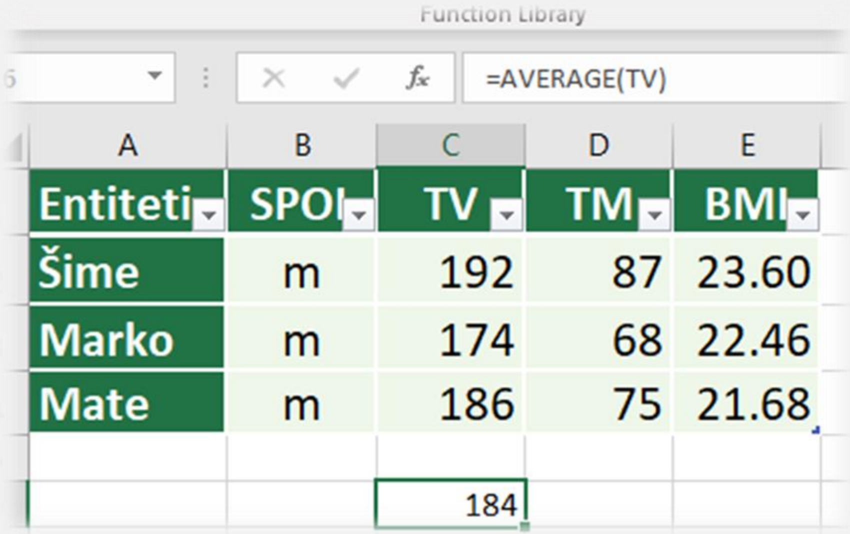
Microsoft Excel

Zadatak 5: Izračunaj aritmetičku sredinu varijable TV.

Entiteti	SPOL	TV	TM	BMI
Šime	m	192	87	23.60
Marko	m	174	68	22.46
Mate	m	186	75	21.68

AS_{TV} = (192+174+186)/3 = 184

6 ***KORIŠTENJE IMENA VARIJABLI U FORMULAMA:*** Označimo ćeliju u kojoj želimo izračunati neku vrijednost pomoću formule (npr. C6). Upišemo znak = zatim naziv funkcije (npr. AVERAGE) i varijable (npr. TV) te pritisnemo tipku *enter*.



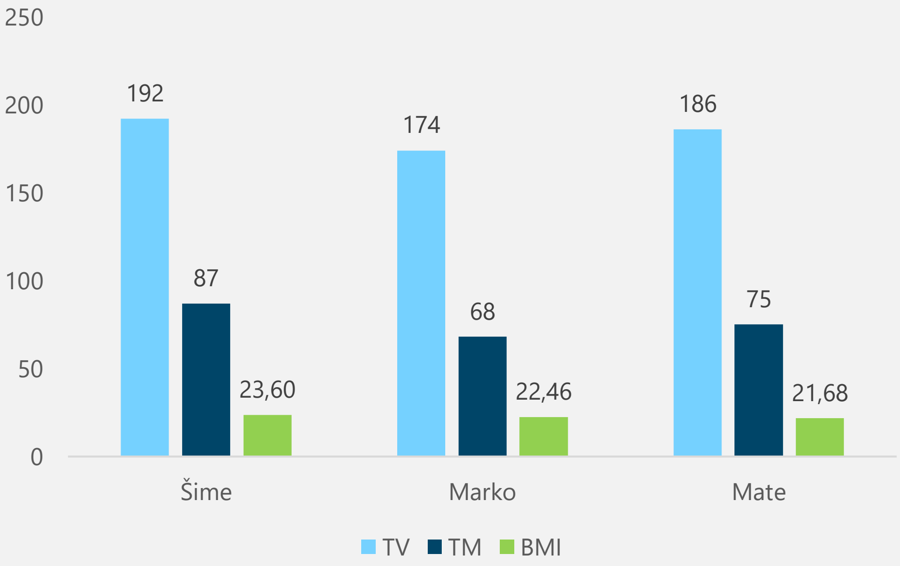
=AVERAGE(TV)

A	B	C	D	E
Entiteti	SPOL	TV	TM	BMI
Šime	m	192	87	23.60
Marko	m	174	68	22.46
Mate	m	186	75	21.68
		184		

Microsoft Excel

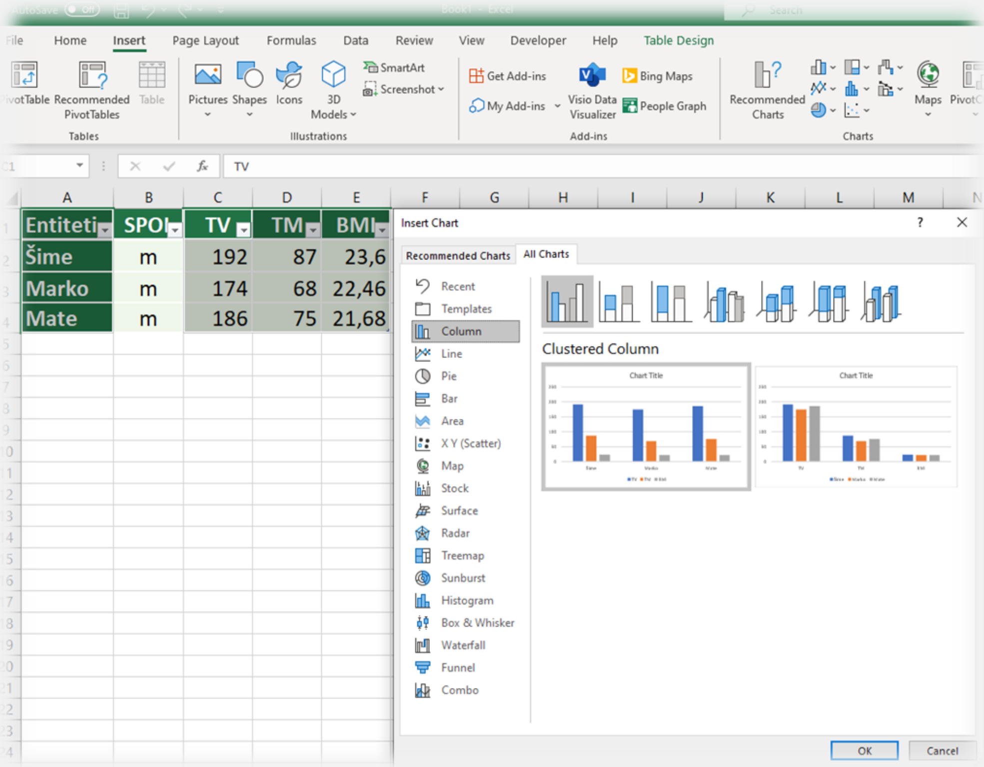
Zadatak 6: Grafikonom stupaca prikažite vrijednosti Šime, Mate i Marka u varijablama TV, TM i BMI.

Entiteti	SPOL	TV	TM	BMI
Šime	m	192	87	23.60
Marko	m	174	68	22.46
Mate	m	186	75	21.68



7

STVARANJE GRAFIKONA:
Odaberemo raspon podataka koje želimo prokazati grafikonom te odaberemo određenu vrstu grafikona u sekciji *Charts* izbornika *Insert*.

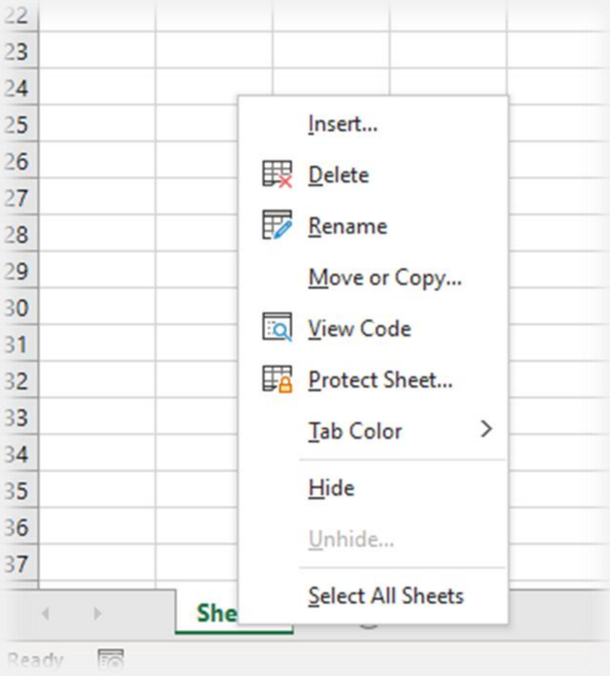


Entiteti	SPOL	TV	TM	BMI
Šime	m	192	87	23,6
Marko	m	174	68	22,46
Mate	m	186	75	21,68

Microsoft Excel

Zadatak 7: List s prethodno kreiranom matricom nazovite «KM», označite ga žutom bojom i kopirajte!

8 ***UPRAVLJANJE LISTOVIMA:***
Kontekstni izbornik za upravljanje listovima pokreće se desnim klikom miša na ime lista. Moguće je umetnuti novi list (*Insert...*), izbrisati ga (*Delete*), preimenovati (*Rename*), premjestiti ili kopirati (*Move or Copy...*) te označiti bojom (*Tab Color...)*.

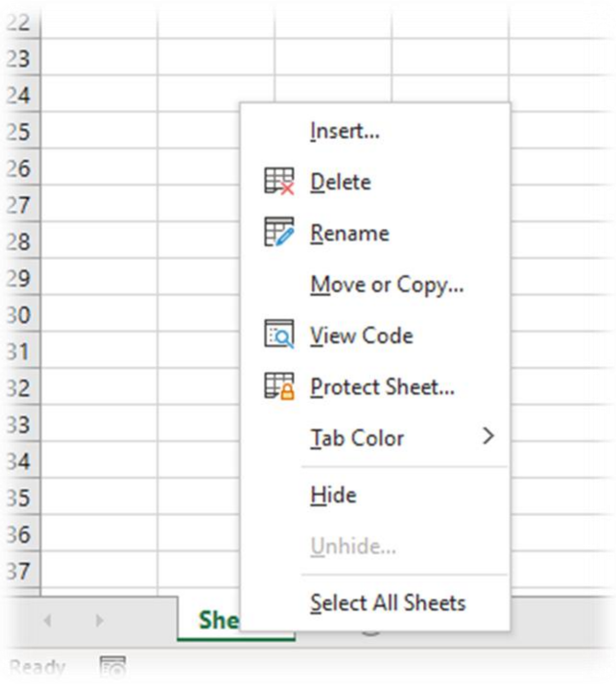


Microsoft Excel

Zadatak 7: List s prethodno kreiranom matricom nazovite «KM», označite ga žutom bojom i kopirajte!

8

UPRAVLJANJE LISTOVIMA:
Kontekstni izbornik za upravljanje listovima pokreće se desnim klikom miša na ime lista. Moguće je umetnuti novi list (*Insert...*), izbrisati ga (*Delete*), preimenovati (*Rename*), premjestiti ili kopirati (*Move or Copy...*) te označiti bojom (*Tab Color...)*.



Što je R?

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

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

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

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

RStudio

RStudio?

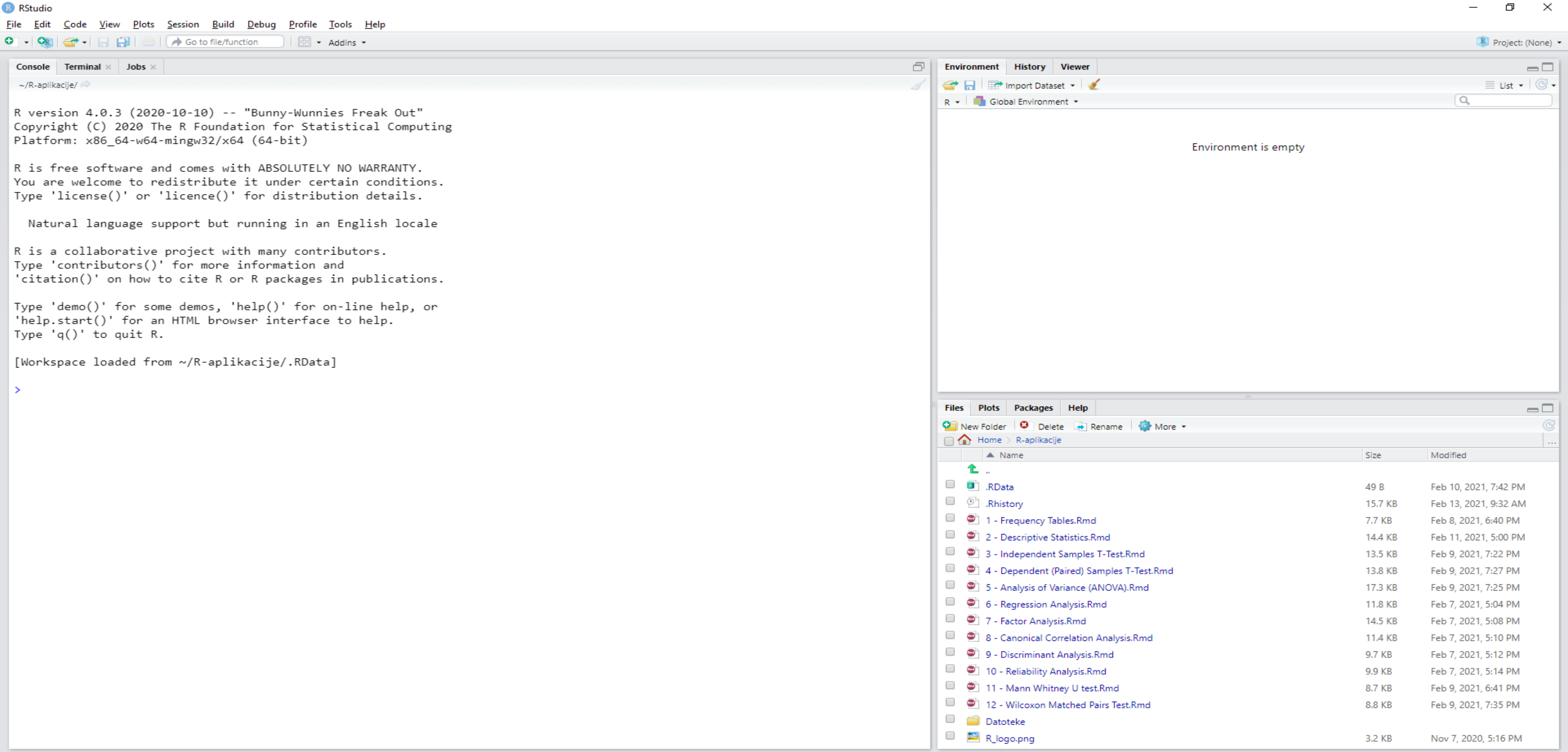
RStudio je korisničko sučelje, odnosno integrirano razvojno okruženje (engl. Integrated Development Invironment, IDE) za R koje se može preuzeti na web-stranici:

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

Za rad u RStudiju potrebno je prethodno instalirati R. Kada se RStudio prvi put pokrene otvori se prozor s trima oknima:

- Console
- Environment i
- Files.

RStudio



```
R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/R-aplikacije/.RData]

>
```

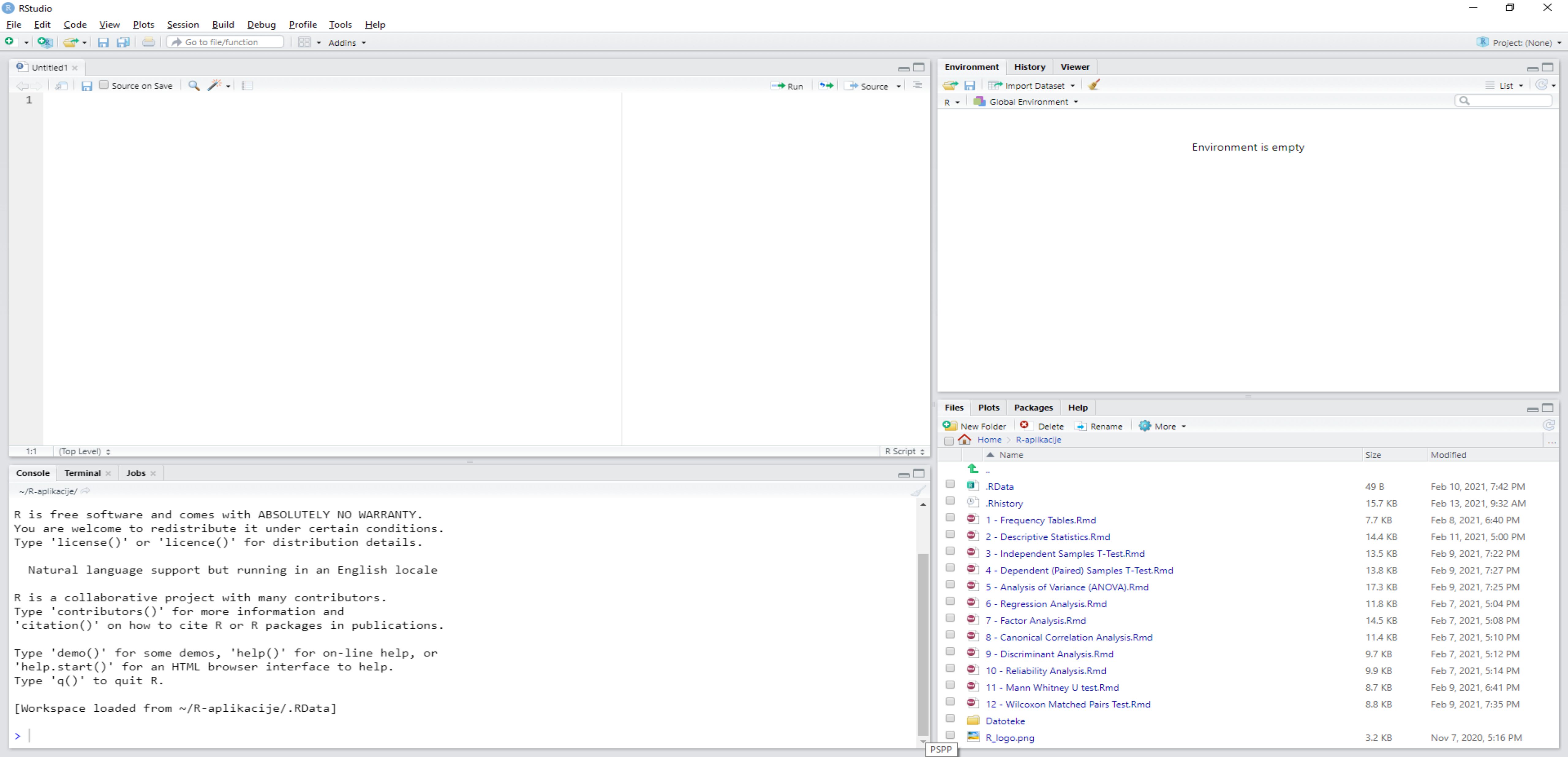
Environment is empty

Name	Size	Modified
..		
.RData	49 B	Feb 10, 2021, 7:42 PM
.Rhistory	15.7 KB	Feb 13, 2021, 9:32 AM
1 - Frequency Tables.Rmd	7.7 KB	Feb 8, 2021, 6:40 PM
2 - Descriptive Statistics.Rmd	14.4 KB	Feb 11, 2021, 5:00 PM
3 - Independent Samples T-Test.Rmd	13.5 KB	Feb 9, 2021, 7:22 PM
4 - Dependent (Paired) Samples T-Test.Rmd	13.8 KB	Feb 9, 2021, 7:27 PM
5 - Analysis of Variance (ANOVA).Rmd	17.3 KB	Feb 9, 2021, 7:25 PM
6 - Regression Analysis.Rmd	11.8 KB	Feb 7, 2021, 5:04 PM
7 - Factor Analysis.Rmd	14.5 KB	Feb 7, 2021, 5:08 PM
8 - Canonical Correlation Analysis.Rmd	11.4 KB	Feb 7, 2021, 5:10 PM
9 - Discriminant Analysis.Rmd	9.7 KB	Feb 7, 2021, 5:12 PM
10 - Reliability Analysis.Rmd	9.9 KB	Feb 7, 2021, 5:14 PM
11 - Mann Whitney U test.Rmd	8.7 KB	Feb 9, 2021, 6:41 PM
12 - Wilcoxon Matched Pairs Test.Rmd	8.8 KB	Feb 9, 2021, 7:35 PM
Datoteke		
R_logo.png	3.2 KB	Nov 7, 2020, 5:16 PM

RStudio

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

RStudio



```
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/R-aplikacije/.RData]

>
```

Name	Size	Modified
..		
.RData	49 B	Feb 10, 2021, 7:42 PM
.Rhistory	15.7 KB	Feb 13, 2021, 9:32 AM
1 - Frequency Tables.Rmd	7.7 KB	Feb 8, 2021, 6:40 PM
2 - Descriptive Statistics.Rmd	14.4 KB	Feb 11, 2021, 5:00 PM
3 - Independent Samples T-Test.Rmd	13.5 KB	Feb 9, 2021, 7:22 PM
4 - Dependent (Paired) Samples T-Test.Rmd	13.8 KB	Feb 9, 2021, 7:27 PM
5 - Analysis of Variance (ANOVA).Rmd	17.3 KB	Feb 9, 2021, 7:25 PM
6 - Regression Analysis.Rmd	11.8 KB	Feb 7, 2021, 5:04 PM
7 - Factor Analysis.Rmd	14.5 KB	Feb 7, 2021, 5:08 PM
8 - Canonical Correlation Analysis.Rmd	11.4 KB	Feb 7, 2021, 5:10 PM
9 - Discriminant Analysis.Rmd	9.7 KB	Feb 7, 2021, 5:12 PM
10 - Reliability Analysis.Rmd	9.9 KB	Feb 7, 2021, 5:14 PM
11 - Mann Whitney U test.Rmd	8.7 KB	Feb 9, 2021, 6:41 PM
12 - Wilcoxon Matched Pairs Test.Rmd	8.8 KB	Feb 9, 2021, 7:35 PM
Datoteke		
R_logo.png	3.2 KB	Nov 7, 2020, 5:16 PM

RStudio

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

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

RStudio

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

- **Files** je mjesto na kojem se može odabrati radni direktorij. Radni direktorij nekog procesa je hijerarhijska mapa dokumenata vezana za proces. Kad se unutar procesa stvara ili poziva neka datoteka, njen put (engl. file path) kreće od pozicije radnog direktorija, a ne iz korijenskog direktorija (engl. root directory).

- **Plots** služi za prikaz grafikona.

- **Packages** daje pregled svih dostupnih paketa. Kvačicom su označeni paketi koji su trenutno aktivni. Popis paketa nadopunjava se po potrebi naredbom install.packages()

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

- **Viewer** služi za pregled lokalnoga web-sadržaja.

RStudio

Podešavanje radnoga direktorija može se napraviti na nekoliko načina. Jedan način je preko kartice **Files**. Pronađemo i odaberemo mapu koju želimo postaviti kao radni direktorij, zatim kliknemo na ikonu **More** koja se nalazi na alatnoj traci unutar karitice i odaberemo **Set As Working Directory**.

Drugi način je pomoću naredbe

```
setwd()
```

Unutar nje u navodnicima upišemo lokaciju (engl. *file path*) željenoga radnog direktorija.

```
setwd("Desktop/R-podaci")
```

Nakon podešavanja radnoga direktorija, dobro je napraviti provjeru. To se može napraviti upitom

`getwd()`.

RStudio

Vrste datoteka u Rstudiju

- **R Script** je jednostavna tekstualna datoteka u koju se upisuju naredbe za R. Ekstenzija datoteka ovoga tipa je **.R**. Naredbe se izvršavaju istovremenim pritiskom tipki [CTRL]+[ENTER], a rezultat se prikazuje u konzoli, a grafikoni se prikazuju na kartici **Plots**. Izvršavanje svih naredbi odjednom moguće je pritiskom na tipke [CTRL]+[A], a nakon toga [CTRL]+[ENTER] ili istovremenim pritiskom tipki [CTRL]+[ALT]+[R]. Komentari se mogu pisati iza znaka #. Prednost R Scripta nad konzolom je u tome što sadrži pregled svih naredbi koje korisnik lako može ponoviti, ispraviti, doraditi i pohraniti.

RStudio

Vrste datoteka u Rstudiju

- **R Markdown** je vrsta R datoteke u kojoj je moguće kombinirati programski kôd, njegove rezultate i tekst te generirati izlazni dokument u HTML, PDF ili Word formatu. Ekstenzija datoteka ovog tipa je **.Rmd**. R Markdown dokument konvertira se u HTML, PDF, ili Word dokument pomoću gumba **Knit**. Ako se radi o **R Markdown Flex Dashboard Shiny** dokumentu, tada se pokreće (izvršava) klikom na gumb **Run Dokument**, a rezultati se prikazuju u web pregledniku.

- **Shiny Web Applications** su datoteke kreirane kao interaktivne web aplikacije čiju izradu i korištenje omogućava **shiny** paket. Dodatnu pogodnost nudi paket **shinydashboard** koji omogućava izradu interaktivnih nadzornih ploča (engl. dashboard) za prezentaciju rezultata aplikacije.

RStudio

Paketi

Sve funkcionalnosti sustava R grupirane su u paketima. R paketi sastoje se od skupa funkcija. Određeni paketi dolaze sa samom instalacijom sustava R, a ostale korisnik nadograđuje prema svojim potrebama. Osnovni paketi (**base**, **stats**, **datasets**, **graphics**, **grDevices**, **methods**, **utils**) dolaze sa sustavom R i sadrže osnovne funkcije za aritmetiku, statistiku, grafiku, ulaz i izlaz podataka itd.

Paketi

Paketi se preuzimaju pomoću naredbe **`install.packages()`**, a učitavaju pomoću naredbe **`library()`**.

Primjer:

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

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

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

Paketi

Glavni repozitorij ili spremište paketa nalazi se na web-stranici **CRAN - The Comprehensive R Archive Network** na kojoj postoji više od 15 000 paketa, a cijeli popis može se pronaći na poveznici: https://cran.r-project.org/web/packages/available_packages_by_name.html.

Postoji i pregledniji način pronalaska odgovarajućih paketa, na primjer, na sljedećim poveznicama grupirani su po temi:
https://cran.r-project.org/web/views/

R aplikacija Quantitative Methods

R aplikaciju Quantitative Methods možete preuzeti na ovoj poveznici:

https://github.com/Quantitative-Methods/R-aplications

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

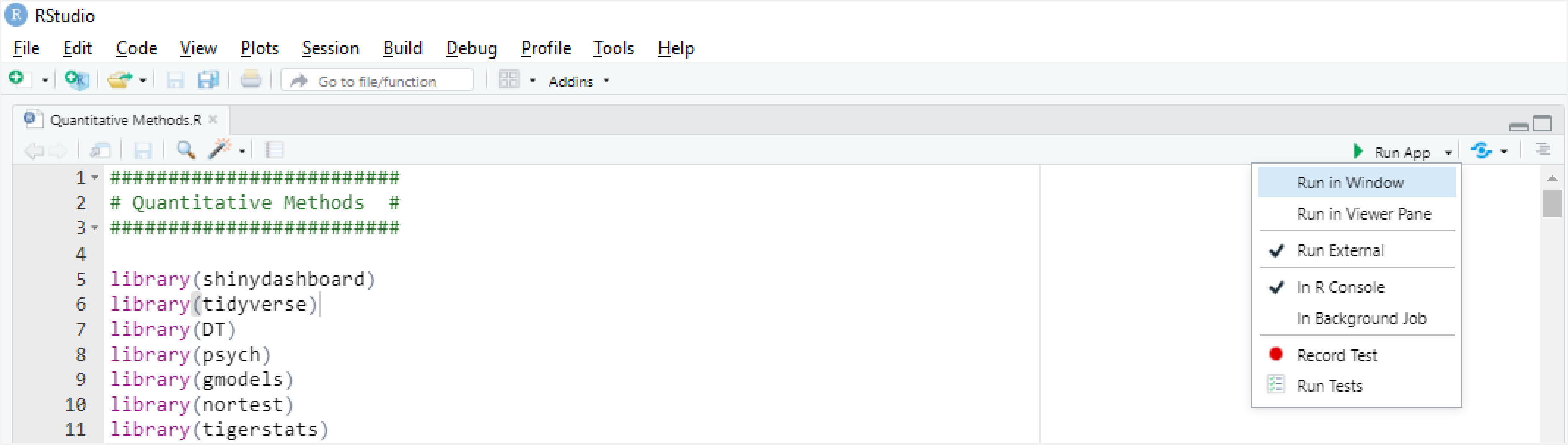
R aplikacija Quantitative Methods

Prije korištenja aplikacije **Quantitative Methods.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 **Quantitative Methods.R** spremna je za korištenje.

Klikom na opciju **Open File** padajućeg izbornika **File** otvorimo mapu u kojoj se nalazi **Quantitative Methods.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**.

R aplikacija Quantitative Methods

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 RStudia. **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 i sl.).

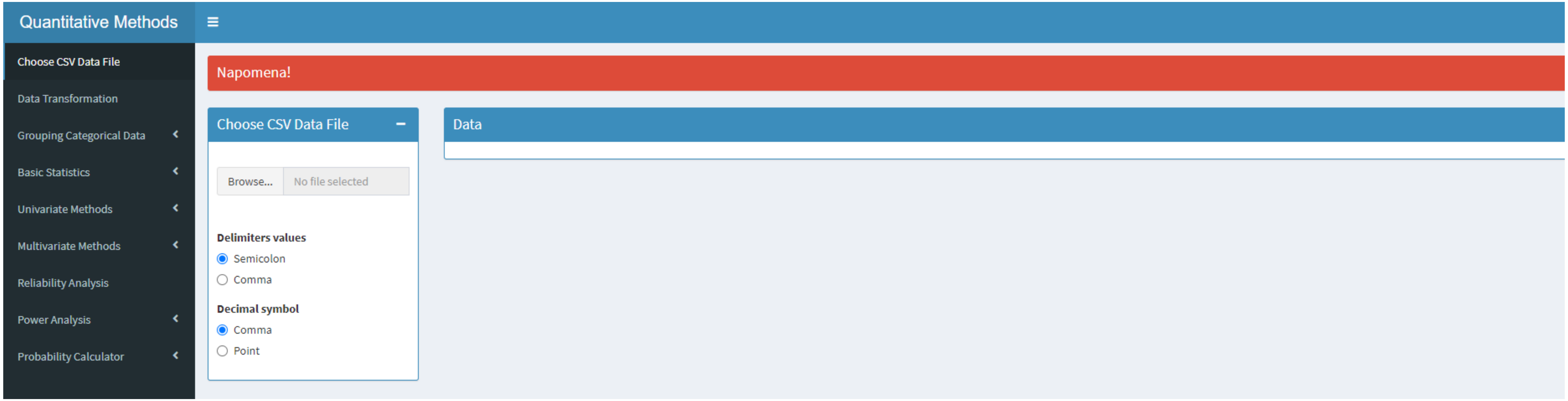


```
##########################
# Quantitative Methods   #
##########################

library(shinydashboard)
library(tidyverse)
library(DT)
library(psych)
library(gmodels)
library(nortest)
library(tigerstats)
```

Choose CSV Data File

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



Delimiters values

● Semicolon
○ Comma

Decimal symbol

● Comma
○ Point

Choose CSV Data File

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

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

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

Choose CSV Data File

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.

OSNOVNI POSTUPCI ZA UREĐIVANJE I PRIKAZIVANJE PODTAKA

Vježba 2

Osnovni postupci za uređivanje i prikazivanje podataka

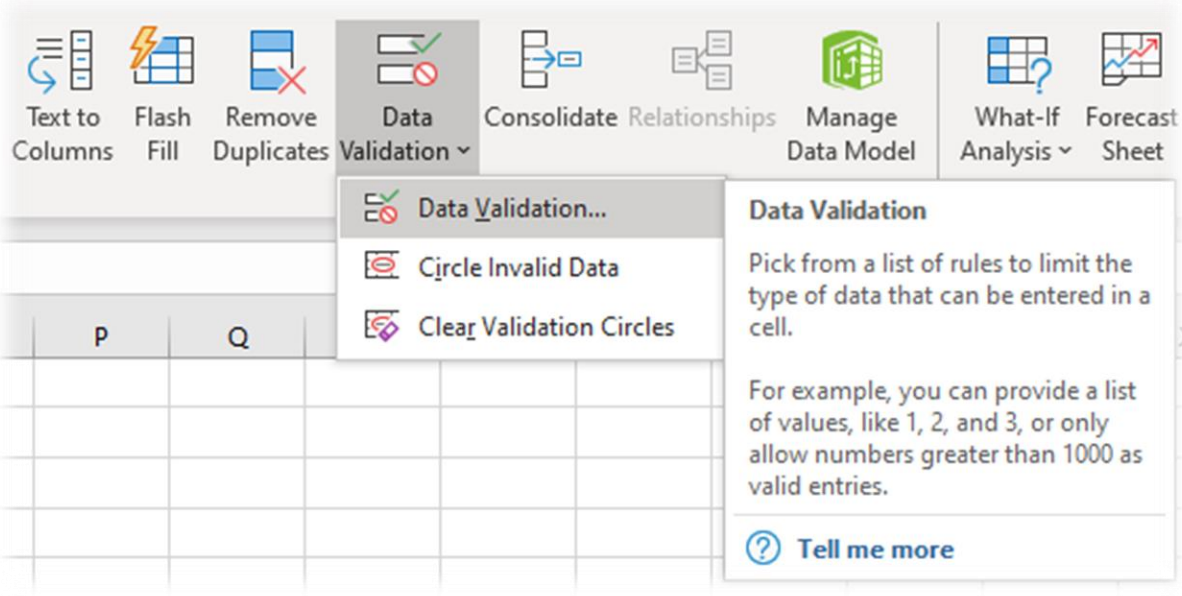
1. Provjera ispravnosti unosa podataka
2. Pretraživanje i izmjena podataka
3. Kodiranje podataka
4. Grupiranje kvalitativnih i kvantitativnih podataka
5. Jednodimenzionalno i višedimenzionalno grupiranje
6. Apsolutne, relativne i kumulativne frekvencije
7. Grafikoni stupaca, redaka i strukturni krug
8. Histogram i poligon frekvencija

Microsoft Excel

Zadatak 1: Kreirajte sljedeću matricu podataka uz korištenje uvjeta za provjeru ispravnosti unosa podataka

ENT.	SPOL	POZ	OKI
AV	M	B	4
EM	M	B	3
KV	M	B	4
MD	M	B	3
MM	M	K	3
NM	M	K	2
NK	M	K	3
SA	M	K	3
SS	M	C	2
VM	M	C	3

1. **Uvjeti za provjeru ispravnosti** za vrijeme unosa podataka postavljaju se putem dijaloškog okvira *Data Validation* koji se pokreće odabirom opcije *Data Validation* izbornika *Data*.



Data Validation

Pick from a list of rules to limit the type of data that can be entered in a cell.

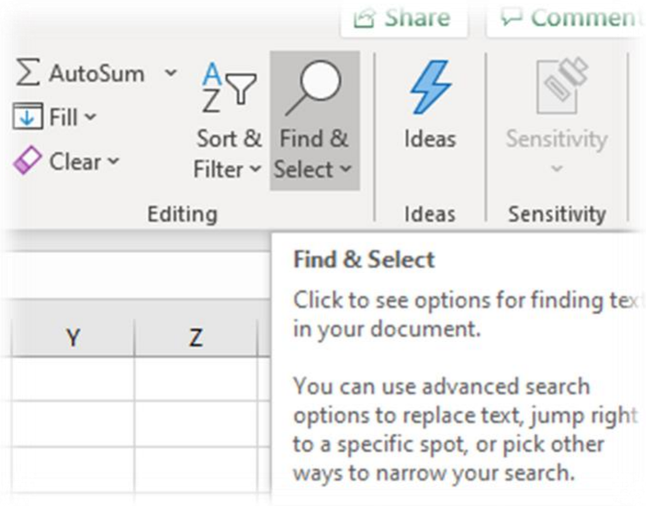
For example, you can provide a list of values, like 1, 2, and 3, or only allow numbers greater than 1000 as valid entries.

Microsoft Excel

Zadatak 2: U varijabli *SPOL* zamijenite sve kodove *Z* s oznakom *Žensko* te kodove *M* s *Muško*!

ENT.	SPOL	POZ	OKI
AV	M	B	4
EM	M	B	3
KV	M	B	4
MD	M	B	3
MM	M	K	3
NM	M	K	2
NK	M	K	3
SA	M	K	3
SS	M	C	2
VM	M	C	3

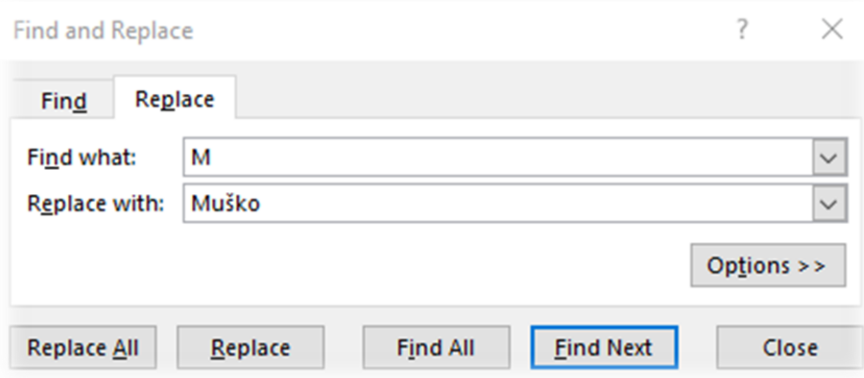
2 **Pretraživanje i izmjena** podataka vrši se putem dijaloškog okvira *Find and Replace* koji se pokreće kombinacijom tipki *Ctrl+H* ili odabirom opcije *Find & Select* izbornika *Home*.



Find & Select

Click to see options for finding tex in your document.

You can use advanced search options to replace text, jump right to a specific spot, or pick other ways to narrow your search.



Find what: M

Replace with: Muško

Osnovni postupci za uređivanje i prikazivanje podataka

Kodiranje podataka – U svrhu bržeg i točnijeg unosa podatke je moguće kodirati. Kodiranje je pridruživanje skraćene oznake ili broja odgovarajućoj kategoriji. Za prepoznavanje kodova potrebno je izraditi kodnu listu.

ENT.	SPOL	POZ	OKI
AV	M	B	4
EM	M	B	3
KV	M	B	4
MD	M	B	3
MM	M	K	3
NM	M	K	2
NK	M	K	3
SA	M	K	3
SS	M	C	2
VM	M	C	3

KRATKO IME VARIJABLE	DUGO IME VARIJABLE	OBLICI (VRIJEDNOSTI) VARIJABLE	KOD
SPOL	Spol	Muškarci	M
		Žene	Z
POZ	Pozicija u igri	Bek	B
		Krilo	K
		Centar	C
OKI	Ocjena kvalitete igrača	Vrlo slaba kvaliteta	1
		Slaba kvaliteta	2
		Dobra kvaliteta	3
		Vrlo dobara kvaliteta	4
		Izvrsna kvaliteta	5

Osnovni postupci za uređivanje i prikazivanje podataka

Grupiranje podataka - predstavlja statistički postupak razvrstavanja entiteta s istim oblikom obilježja u određeni broj disjunktnih podskupova.

Frekvencija - broj entiteta koji imaju isti oblik obilježja, odnosno, broj entiteta u određenoj grupi (klasi, kategoriji, razredu).

Jednodimenzionalno grupiranje - grupiranje entiteta po jednom obilježju (varijabli).

Primjer: Na praktični dio ispit iz K.M. izišlo je 40 studenata, 15 ih je položilo, a 25 nije.

OBLIK OBILJEŽJA	FREKVENCIJA
NISU POLOŽILI	25
POLOŽILI	15
UKUPNO	40

Osnovni postupci za uređivanje i prikazivanje podataka

Višedimenzionalno grupiranje - grupiranje entiteta po više obilježja (varijabli).

Primjer: Od 26 studenata na praktičnom dijelu ispita iz K.M. položilo je 10, a od 14 studentica položilo ih je 5.

SPOL	NISU POLOŽILI	POLOŽILI	UKUPNO
MUŠKARCI	16	10	26
ŽENE	9	5	14
Ukupno	25	15	40

Osnovni postupci za uređivanje i prikazivanje podataka

Grupiranje i grafičko prikazivanje kvalititivnih podataka

S obzirom da se kvalitativne varijable (obilježja) dijele na nominalna i ordinalna (redosljedna) moguće je razlikovati grupiranje prema *nominalnim* i *ordinalnim* varijablama.

USPJEH NA ISPITU	FREKVENCIJA	%
NEDOVOLJAN	25	62,5
DOVOLJAN	8	20
DOBAR	3	7,5
VRLO DOBAR	2	5
ODLIČAN	2	5
UKUPNO	40	100

Osnovni postupci za uređivanje i prikazivanje podataka

Grupiranje i grafičko prikazivanje kvalititivnih podataka

Relativna frekvencija - predstavlja omjer između frekvencije određene kategorije i zbroja frekvencija svih kategorija (ukupan broj entiteta).

$$p_g = \frac{f_g}{n} \quad \%_g = \frac{f_g}{n} \cdot 100$$

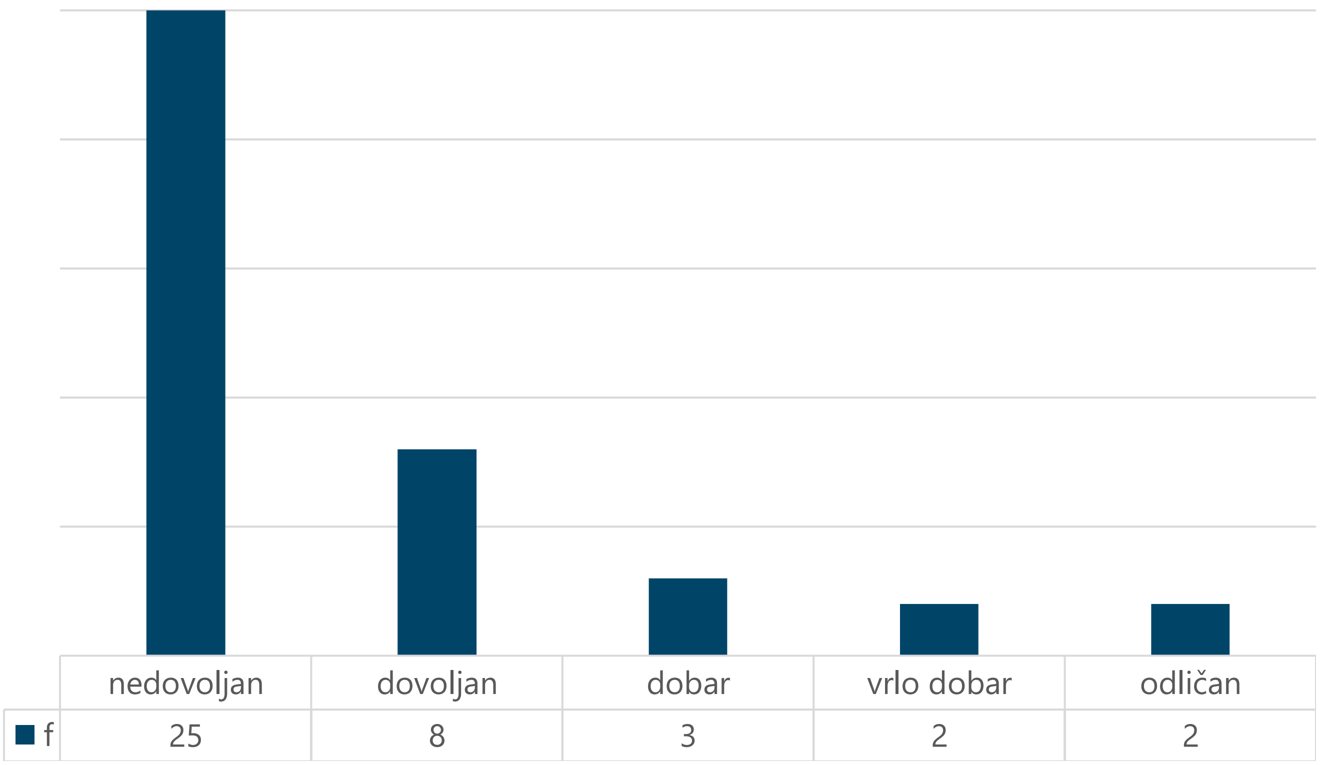
gdje je:

- p_g - relatvna frekvencija izražena u proporciji grupe g
- f_g - frekvencija u grupi g
- $\%_g$ - relativna frekvencija izražena u postotku
- n - ukupan broj entiteta
- k - broj kategorija (grupa).

Osnovni postupci za uređivanje i prikazivanje podataka

Grupiranje i grafičko prikazivanje kvalititivnih podataka

Kvalitativni podaci najčešće se prikazuju pomoću *grafikona stupaca*, *grafikona redaka* i *strukturnim krugom*.

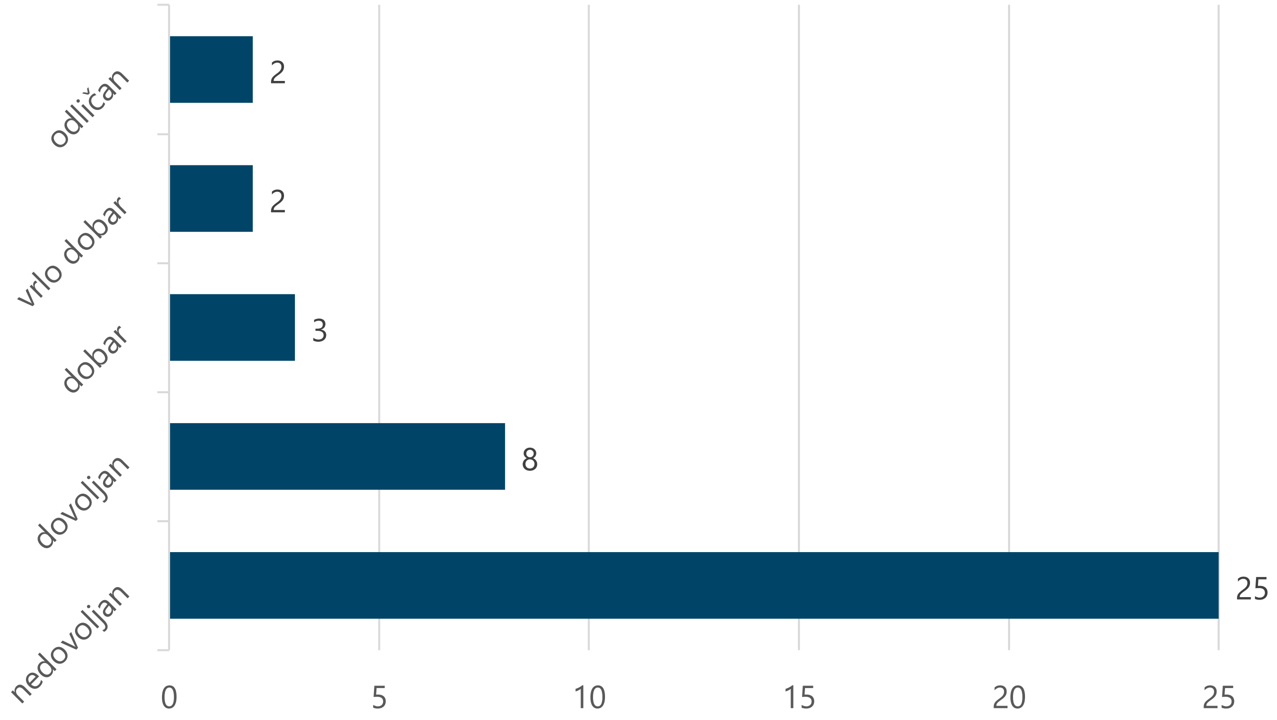


	nedovoljan	dovoljan	dobar	vrlo dobar	odličan
■ f	25	8	3	2	2

Osnovni postupci za uređivanje i prikazivanje podataka

Grupiranje i grafičko prikazivanje kvalititivnih podataka

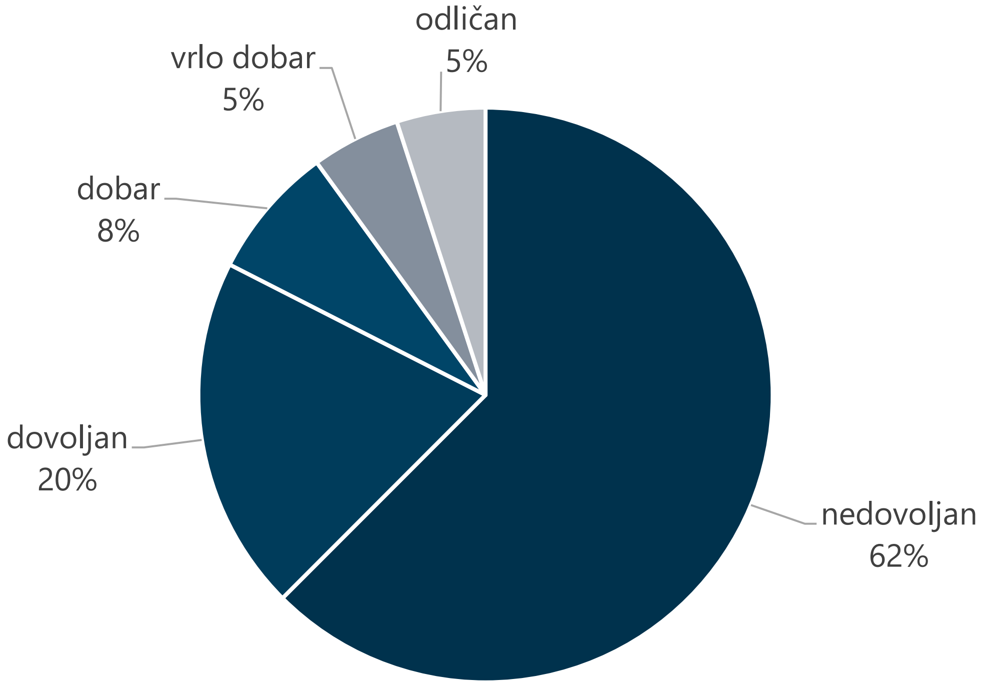
Kvalitativni podaci najčešće se prikazuju pomoću *grafikona stupaca*, ***grafikona redaka*** i *strukturnim krugom*.



Osnovni postupci za uređivanje i prikazivanje podataka

Grupiranje i grafičko prikazivanje kvalititivnih podataka

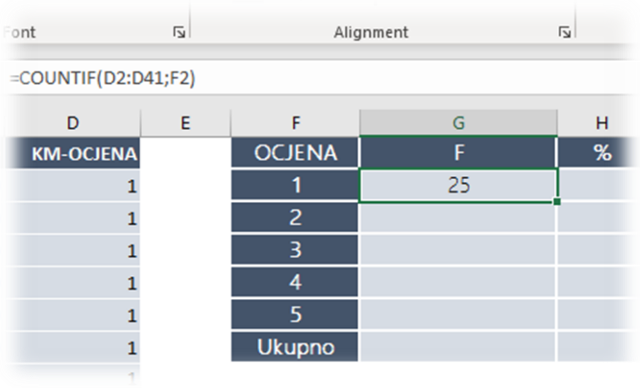
Kvalitativni podaci najčešće se prikazuju pomoću *grafikona stupaca*, *grafikona redaka* i ***strukturnim krugom***.



Microsoft Excel

Zadatak 3: U datoteci *KM.xls* utvrdite frekvencije i relativne frekvencije (u postotku) za varijablu KM-OCJENA te ih prikažite pomoću grafikona stupaca vertikalne i horizontalne orijentacije te strukturnim krugom.

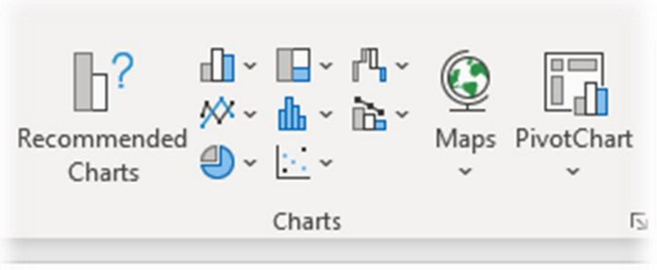
3 **Prebrojavanje podataka:** Utvrđivanje frekvencije pojedine kategorije vrši se prebrojavanjem pomoću funkcije *Countif*.



=COUNTIF(D2:D41;F2)

D	E	F	G	H
KM-OCJENA		OCJENA	F	%
1		1	25	
1		2		
1		3		
1		4		
1		5		
1		Ukupno		

4 **Grafikoni stupaca i strukturni krug:** Iscrtavanje grafikona stupaca vertikalne orijentacije (*Column*), grafikona stupaca horizontalne orijentacije (*Bar*) i strukturnog kruga (*Pie*) vrši se odabirom opcije *Chart...* padajućeg izbornika *Insert*.

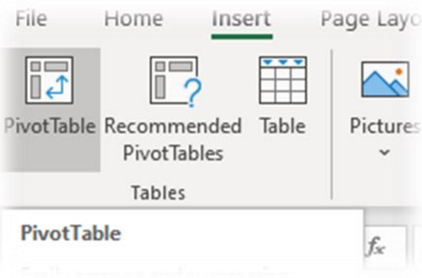


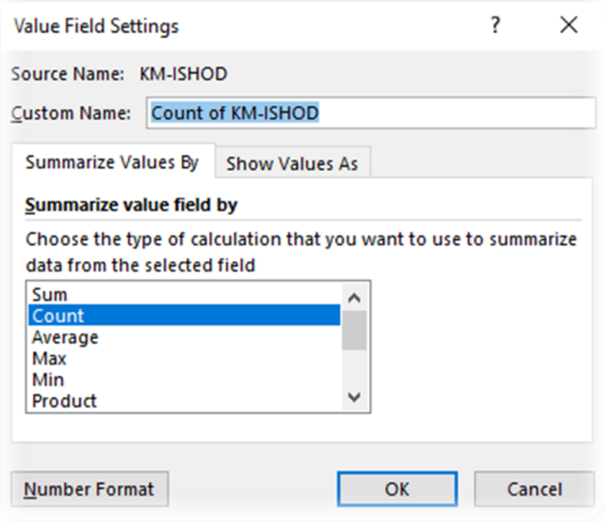
Microsoft Excel

Zadatak 4: U datoteci *KM.xls* grupirajte entitete prema varijabli SPOL i KM-ISHOD koristeći opciju *Pivot Table*.

SPOL	NISU POLOŽILI	POLOŽILI	UKUPNO
MUŠKARCI	16	10	26
ŽENE	9	5	14
Ukupno	25	15	40

5 **Pivot Table:** Dvodimenzionalno grupiranje moguće je odabirom opcije *Pivot Table* izbornika *Insert*.




Value Field Settings

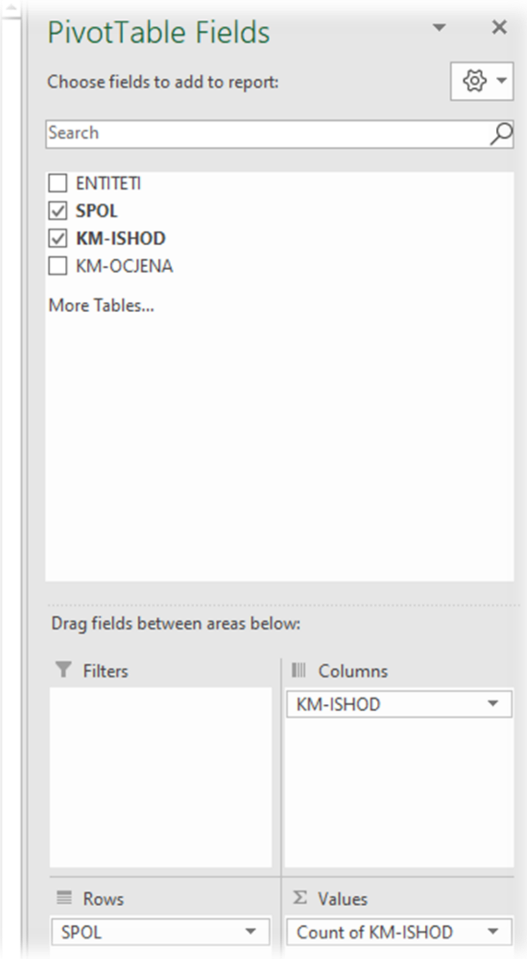
Source Name: KM-ISHOD

Custom Name: Count of KM-ISHOD

Summarize value field by

Choose the type of calculation that you want to use to summarize data from the selected field

- Sum
- Count
- Average
- Max
- Min
- Product


PivotTable Fields

Choose fields to add to report:

☐ ENTITETI
☑ SPOL
☑ KM-ISHOD
☐ KM-OCJENA

Drag fields between areas below:

Filters:

Columns: KM-ISHOD

Rows: SPOL

Values: Count of KM-ISHOD

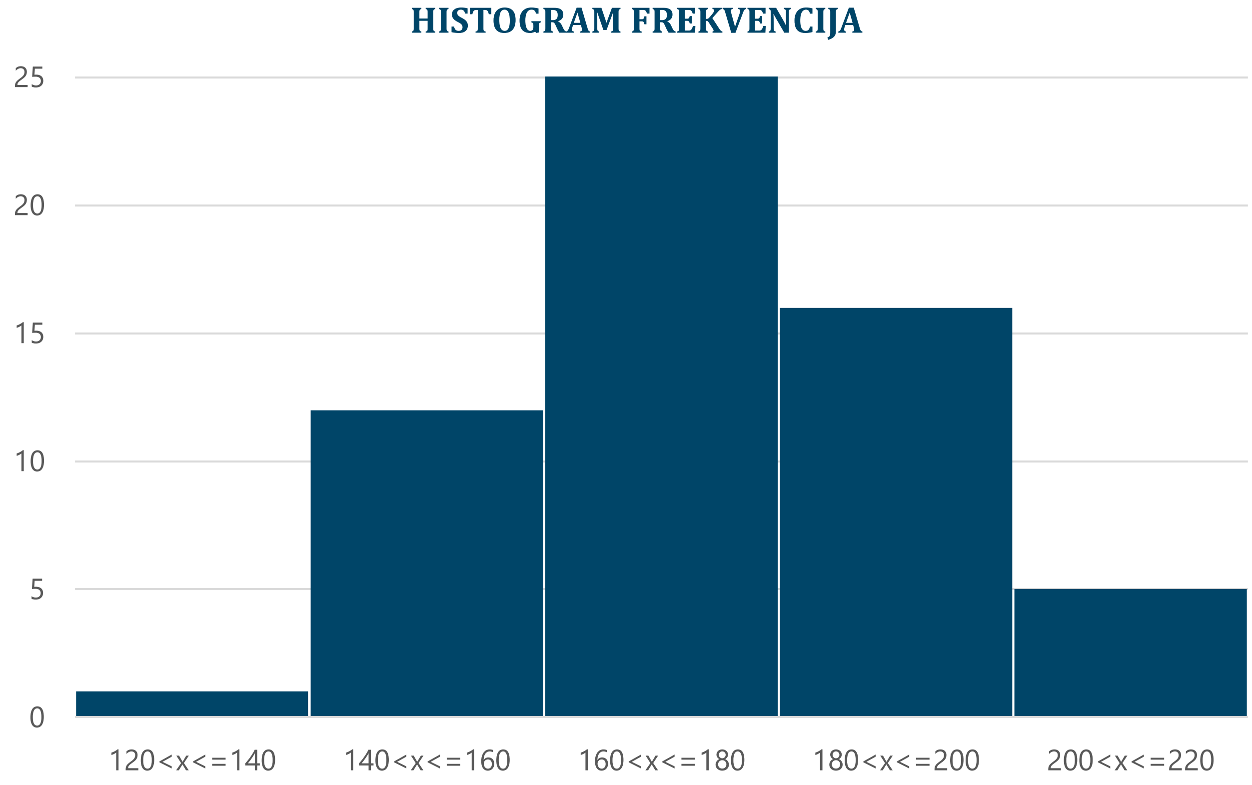
Osnovni postupci za uređivanje i prikazivanje podataka

Grupiranje i grafičko prikazivanje kvantitativnih podataka

Ako diskretna varijabla ima veliki broj pojavnih oblika ili se radi o kontinuiranoj varijabli onda se podaci grupiraju u manji broj razreda. Za uspješno grupiranje potrebno je odrediti prikladan *broj razreda* i njihovu veličinu - *interval razreda*.

Primjer: Grupiranje 60 dječaka judaša u 5 razreda prema varijabli *Skok udalj s mjesta*.

INTERVALI RAZREDA	F	RF (%)
120<x<=140	1	1,67
140<x<=160	12	20,00
160<x<=180	26	43,33
180<x<=200	16	26,67
200<x<=220	5	8,33



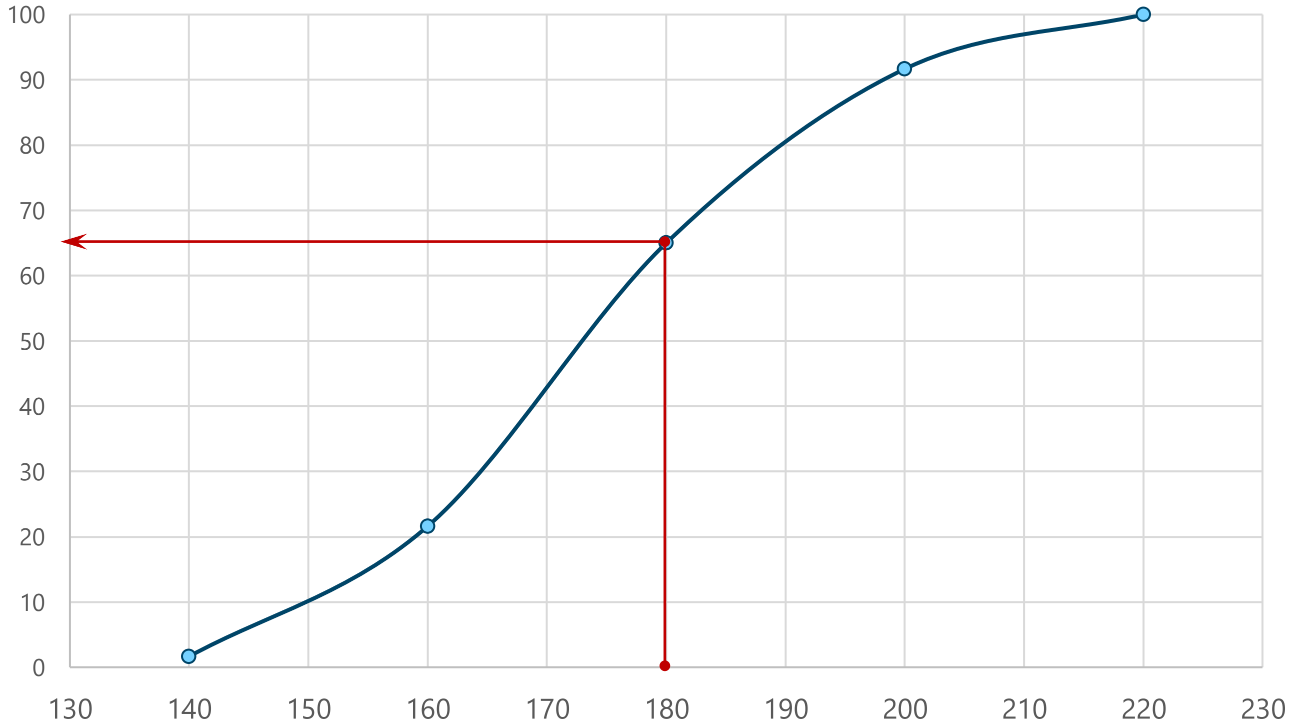
Osnovni postupci za uređivanje i prikazivanje podataka

Grupiranje i grafičko prikazivanje kvantitativnih podataka

Ako se frekvencije (apsolutne ili relativne) svakog narednog razreda zbrajaju sa sumom predhodnih razreda dobiju se tzv. kumulativne frekvencije.

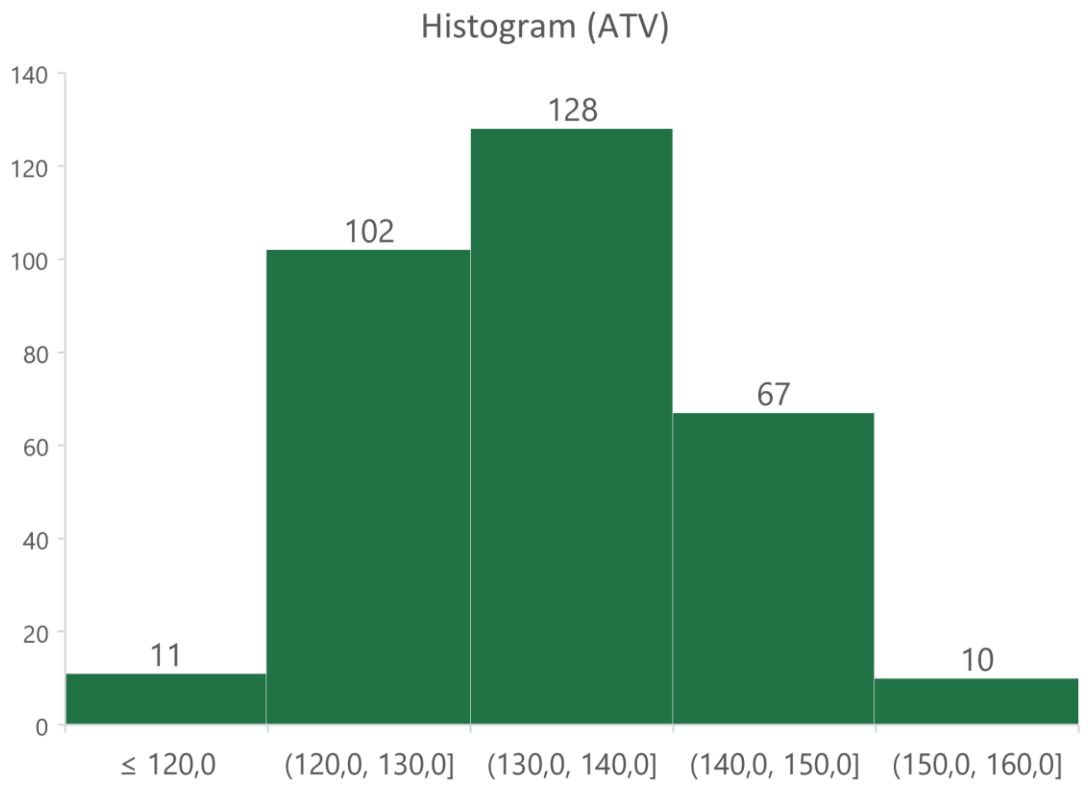
INTERVALI RAZREDA	CF	CRF (%)
120<x<=140	1	1,67
140<x<=160	13	21,61
160<x<=180	39	65,00
180<x<=200	55	91,67
200<x<=220	60	100,00

POLIGON KUMULATIVNIH RELATIVNIH FREKVENCIJA



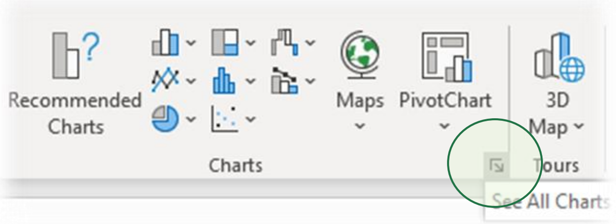
Microsoft Excel

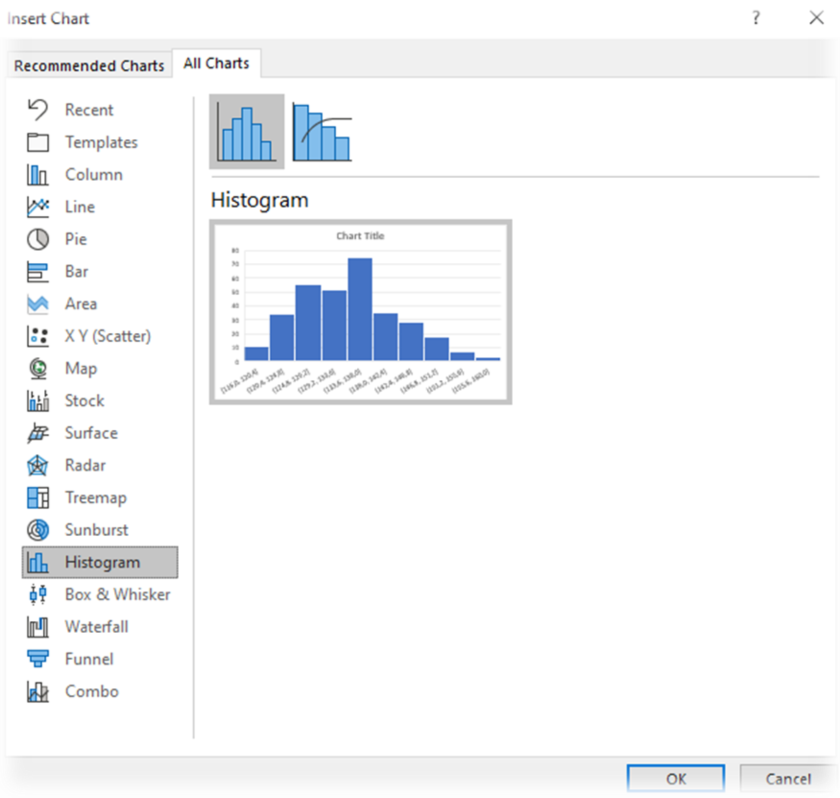
Zadatak 5: U datoteci *Ucenici.xls* grupirajte podatke varijable ATV u 5 razreda i prikažite rezultate histogramom frekvencija.



6 **Histogram frekvencija** dobije se odabirom opcije *Histogram* kartice *All Charts* dijaloškog okvira *Insert Chart* nakon odabira *See All Charts*.

.



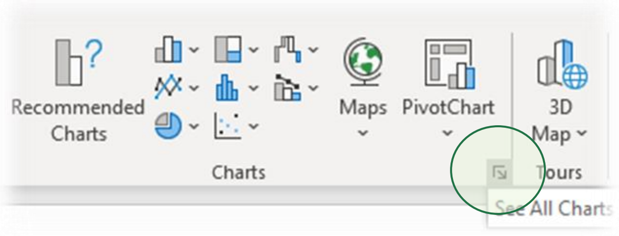


Microsoft Excel

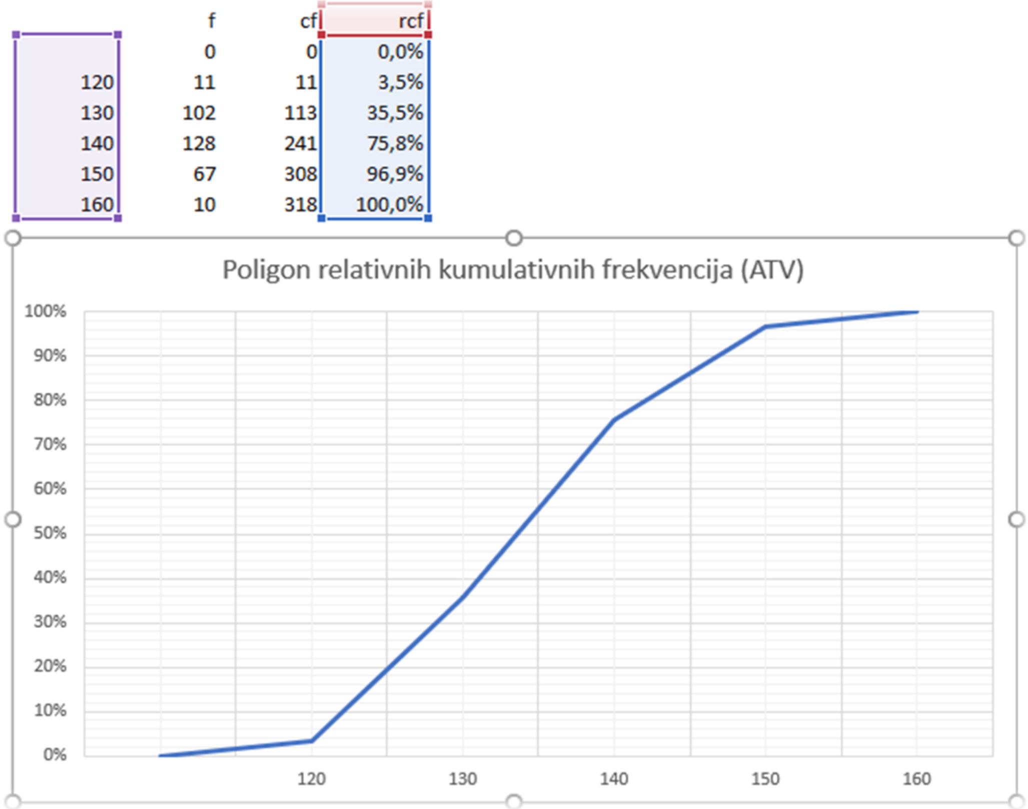
Zadatak 6: U datoteci *Ucenici.xls* grupirajte podatke varijable ATV u 5 razreda te izračunajte kumulativne i relativne kumulativne frekvencije.

RAZREDI	F	CF	RCF (%)
x<=120	11	11	3,5
120<x<=130	102	113	35,5
130<x<=140	128	241	75,8
140<x<=150	67	308	96,9
150<x<=160	10	318	100

7 **Poligon frekvencija** dobije se odabirom opcije *Line* kartice *All Charts* dijaloškog okvira *Insert Chart* nakon odabira *See All Charts*.



.

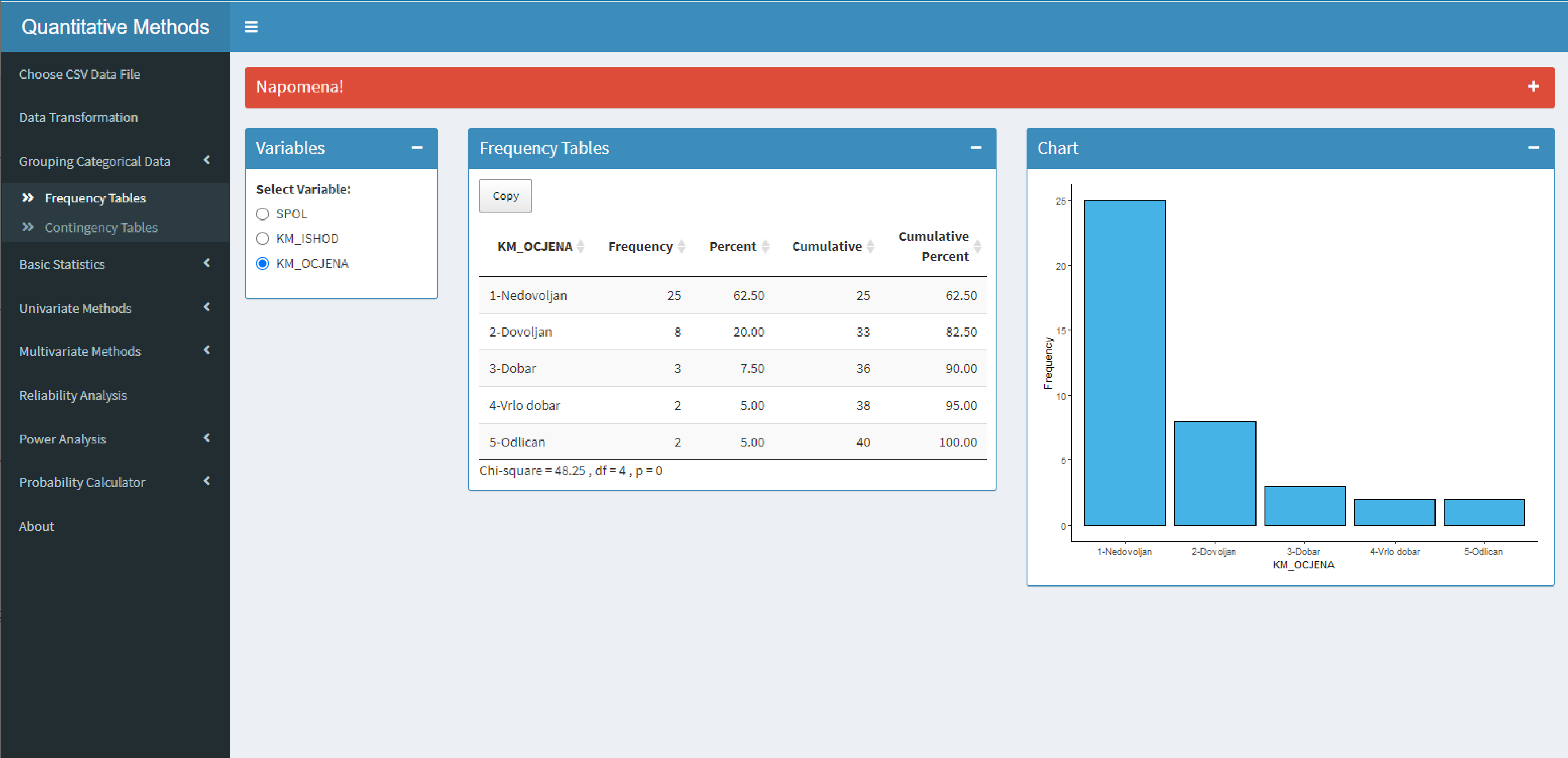


	f	cf	rcf
	0	0	0,0%
120	11	11	3,5%
130	102	113	35,5%
140	128	241	75,8%
150	67	308	96,9%
160	10	318	100,0%

RStudio – Frequency Tables

Zadatak 6: U datoteci *KM.csv* utvrdite frekvencije i relativne frekvencije (u postotku) za varijablu KM_OCJENA te ih prikažite pomoću grafikona.

RStudio – Frequency Tables



Select Variable:
- ○ SPOL
- ○ KM_ISHOD
- ● KM_OCJENA

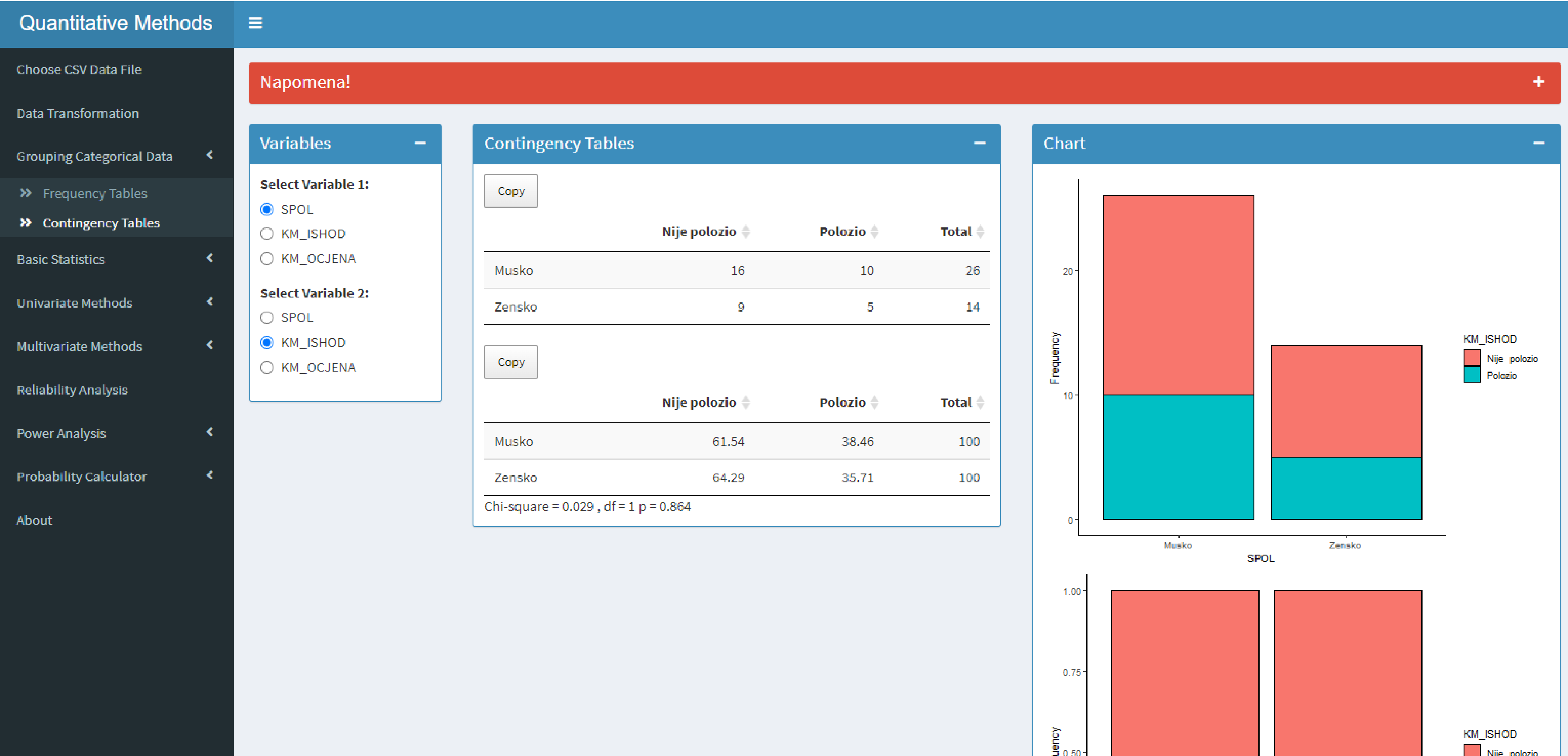
KM_OCJENA	Frequency	Percent	Cumulative	Cumulative Percent
1-Nedovoljan	25	62.50	25	62.50
2-Dovoljan	8	20.00	33	82.50
3-Dobar	3	7.50	36	90.00
4-Vrlo dobar	2	5.00	38	95.00
5-Odlican	2	5.00	40	100.00

Chi-square = 48.25 , df = 4 , p = 0

RStudio – Contingency Tables

Zadatak 7: U datoteci *KM.csv* grupirajte entitete prema varijabli SPOL i KM_ISHOD.

RStudio – Contingency Tables



Select Variable 1:

● SPOL
○ KM_ISHOD
○ KM_OCJENA

Select Variable 2:

○ SPOL
● KM_ISHOD
○ KM_OCJENA

	Nije polozio	**Polozio**	**Total**
Musko	16	10	26
Zensko	9	5	14

	Nije polozio	**Polozio**	**Total**
Musko	61.54	38.46	100
Zensko	64.29	35.71	100

Chi-square = 0.029 , df = 1 p = 0.864

DESKRIPTIVNI POKAZATELJI

Vježba 3

Deskriptivni pokazatelji

1. Mjere centralne tendencije ili središnje mjere
2. Mjere varijabilnosti ili disperzije
3. Mjere oblika distribucije
4. Kutijasti dijagram (Box – whisker plot)
5. Normalna ili Gaussova distribucija

Mjere centralne tendencije ili središnje mjere

Aritmetička sredina

Najčešće korištena mjera centralne tendencije. Izračunava se kao omjer zbroja svih vrijednosti neke varijable i ukupnog broja entiteta.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

gdje je

- *i* = *1, …,n,*
- *n* - broj entiteta.

Mjere centralne tendencije ili središnje mjere

Primjer: Naka je 10 entiteta postiglo sljedeće rezultate u nekom motoričkom testu: 1, 2, 2, 3, 3, 3, 3, 4, 4, 5.

Aritmetička sredina je

$$\bar{x} = \frac{1+2+2+3+3+3+3+4+4+5}{10} = \frac{30}{10} = 3$$

Računa se samo za kvantitativne podatke i ima sljedeća svojstva:

- $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$
- $\sum_{i=1}^{n}(x_i - \bar{x})^2 = min$
- $x_{min} \leq \bar{x} \leq x_{max}$

Mjere centralne tendencije ili središnje mjere

Mod ili dominantna vrijednost

Mod ili dominantna vrijednost - vrijednost kvalitativne ili kvantitativne varijable koja se najčešće pojavljuje.

Primjer: Naka je *10* entiteta postiglo sljedeće rezultate u nekom motoričkom testu: 1, 2, 2, 3, 3, 3, 3, 4, 4, 5.

OCJENA	F
1	1
2	2
3	4
4	2
5	1

Mjere centralne tendencije ili središnje mjere

Medijan ili središnja vrijednost

Medijan ili centralna vrijednost - vrijednost koja se nalazi na sredini uređenog niza podataka (uzlazno ili silazno sortiranog), odnosno vrijednost koja uređeni niz podataka dijeli na dva jednakobrojana dijela.

Primjer: Naka je 15 entiteta (neparan niz) izmjereno nekim testom čiji su rezultati uređeni po veličini:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
1	2	2	2	2	3	3	3	3	3	4	4	4	5	5

Primjer: Ako je broj podataka (entiteta) paran onda je medijan jednak aritmetičkoj sredini vrijednosti dvaju središnjih članova uređenog niza.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
1	1	2	2	2	3	3	3	3	3	4	4	4	4	5	5

$$\mu_e = \frac{x_8 + x_9}{2} = \frac{3 + 3}{2} = 3$$

Mjere disperzije ili varijabilnosti

Mjerama disperzije ili varijabilnosti ukazuje se na veličinu međusobnog razlikovanja rezultata entiteta u nakoj varijabli. To su:

- totalni raspon
- varijanca
- standardna devijacija
- koeficijent varijablinosti.

Totalni raspon

Totalni raspn predstavlja razliku između maksimalne (x_{max}) i minimalne (x_{min}) vrijednosti.

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

Vrlo je nesigurana mjera varijabilnosti, jer jedan ekstremni rezultata znatno utječe na njegovu vrijednost. Povećenjem entiteta u uzorku obično se povećava i totalni raspon jer se povećava vjerojatnost uključivanja entiteta s ekstremnim (maksimalnim i minimalnim) vrijednostima.

Mjere disperzije ili varijabilnosti

Varijanca i standardna devijacija

Procjena stupnja disperzije moguća je i putem odstupanja vrijednosti članova niza od neke središnje vrijednosti, najčešće aritmetičke sredine.

$$d_i = x_i - \bar{x}$$

$$\sum_{i=1}^{n} d_i = 0$$

Varijanca – prosječno kvadratno odstupanje rezultata entiteta od aritmetičke sredine.

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

Standardna devijacija – korijen iz varijance.

Mjere disperzije ili varijabilnosti

Koeficijent varijabilnosti

Za usporedbu različitih pojava (varijabli) koristi se *koeficijent varijabilnosti* koji pokazuje koliki postotak vrijednosti aritmetičke sredine iznosi standardna devijacija

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

gdje je

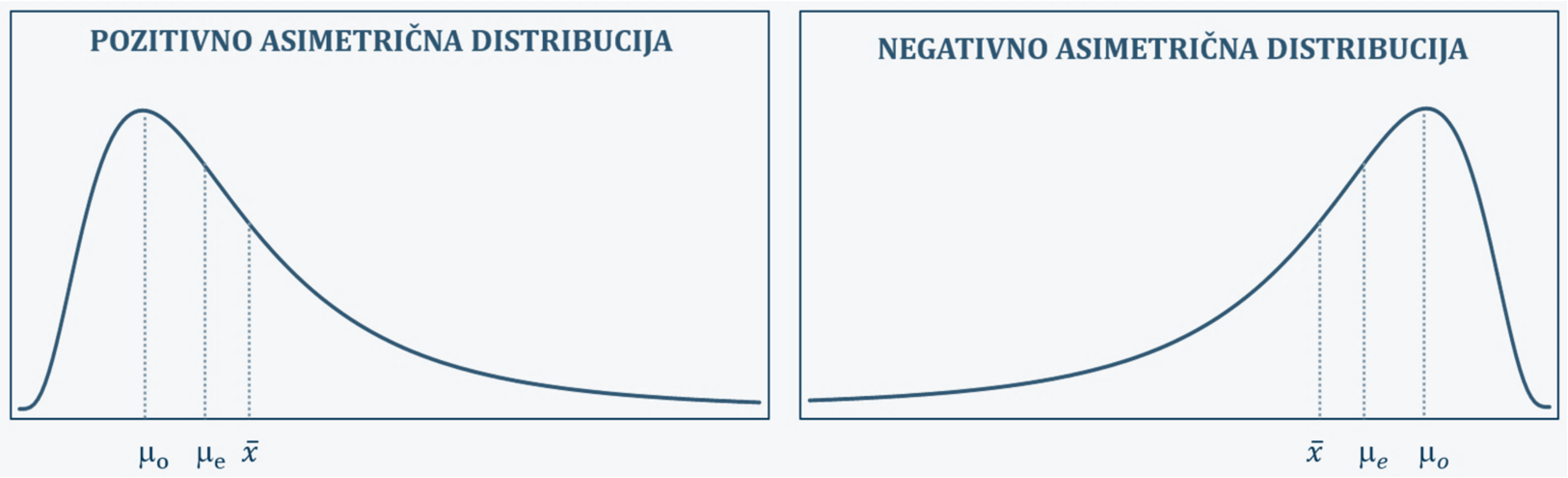
- *V* - keoficijent varijabilnosti
- *s* - standardna devijacija
- $\bar{x}$- aritmetička sredina.

Mjere oblika distribucije

Skewness - mjera asimetrije distribucije

Koeficijent asimetrije se izračunava preko *trećeg momenta oko sredine* (m_3) i standardne devijacije podignute na treću potenciju (σ^3)

$$a_3 = \frac{m_3}{s^3}, \text{ gdje je } m_3 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n} \text{ treći moment oko sredine.}$$



Mjere oblika distribucije

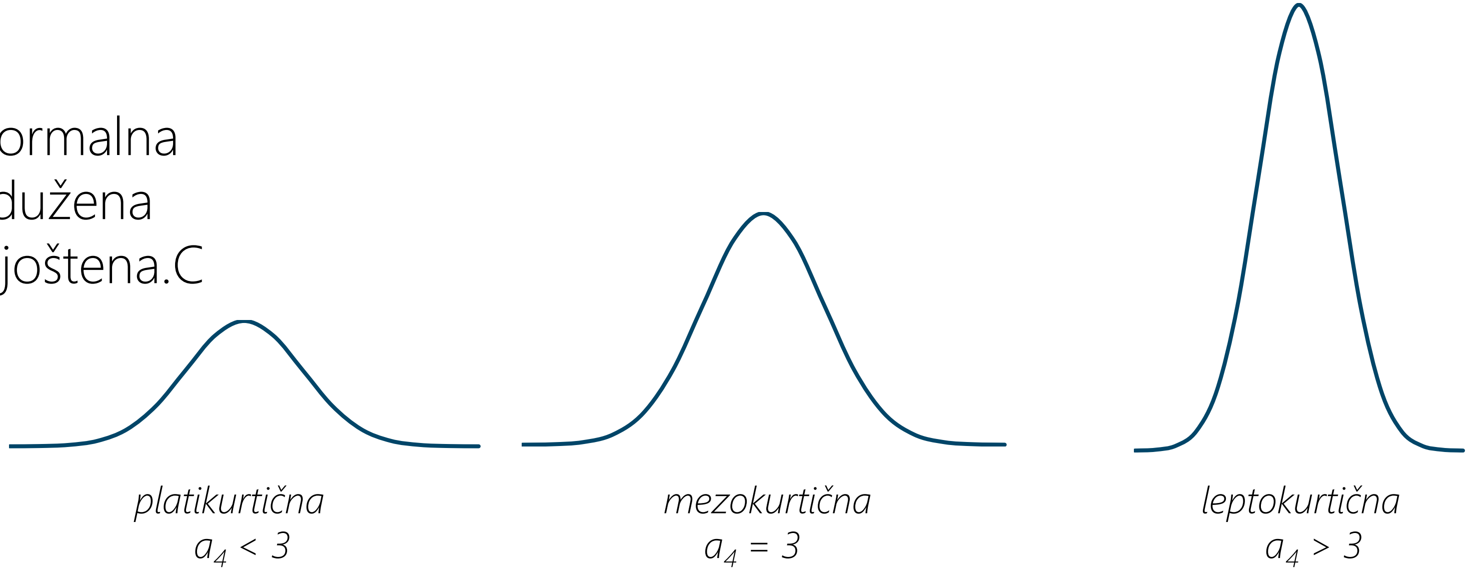
Kurtosis - mjera izduženosti distribucije

Stupanj spljoštenosti ili izduženosti distribucije izražava se koeficijentom a_4, a se izračunava preko *četvrtog momenta oko sredine* (m_4) i standardne devijacije podignute na četvrtu potenciju (σ^4).

$a_4 = \frac{m_4}{s^4}$, gdje je $m_4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}$ treći moment oko sredine.

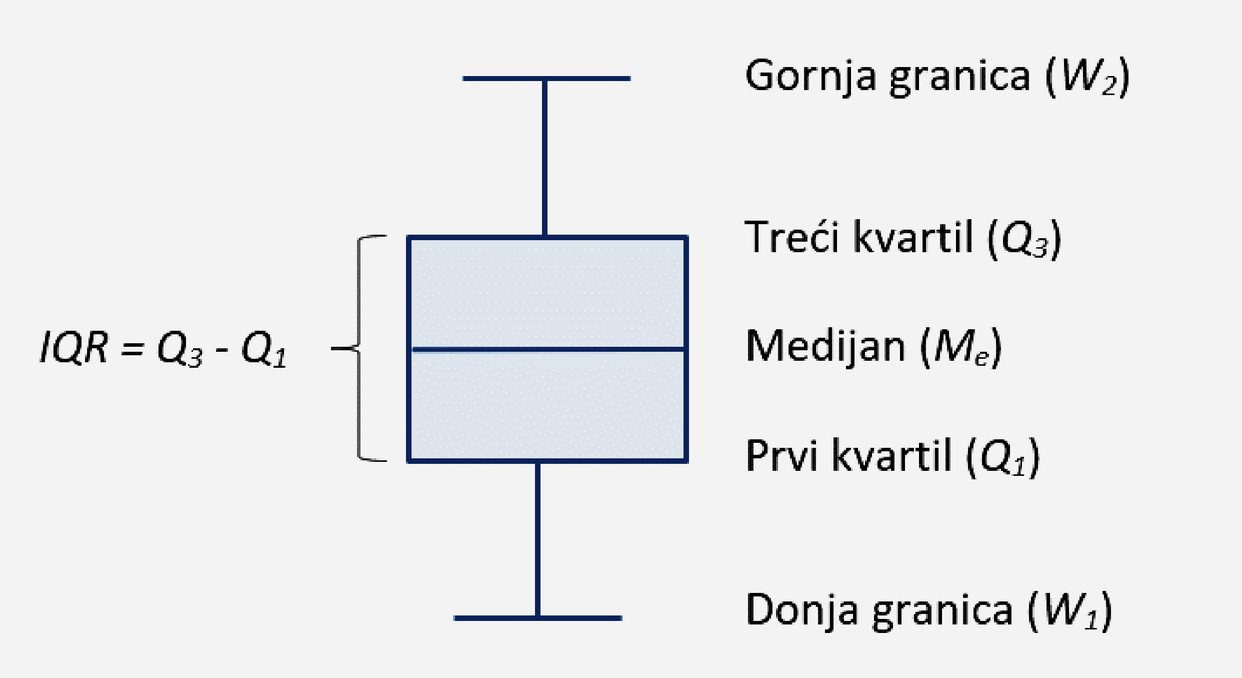
Ako je koeficijent spljoštenosti:

- $a_4 = 3$ distribucija je *mezokurtična* - normalna
- $a_4 > 3$ distribucija je *leptokurtična* - izdužena
- $a_4 < 3$ distribucija je *platikurtična* - spljoštena.C



Kutijasti dijagram (*Box – whisker 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 (*Box – whisker plot*)

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

- $W_1 = Min$ ako je $Min > Q_1 - 1{,}5 \cdot IQR$, inače je $W_1 = Q_1 - 1{,}5 \cdot IQR$
- $W_2 = Max$ ako je $Max < Q_3 + 1{,}5 \cdot IQR$, inače je $W_2 = Q_3 + 1{,}5 \cdot IQR$

Normalna ili Gaussova distribucija

Za slučajnu kontinuiranu varijablu *x* kaže se da ima normalnu distribuciju s parametrima μ i σ^2 ako *je*

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

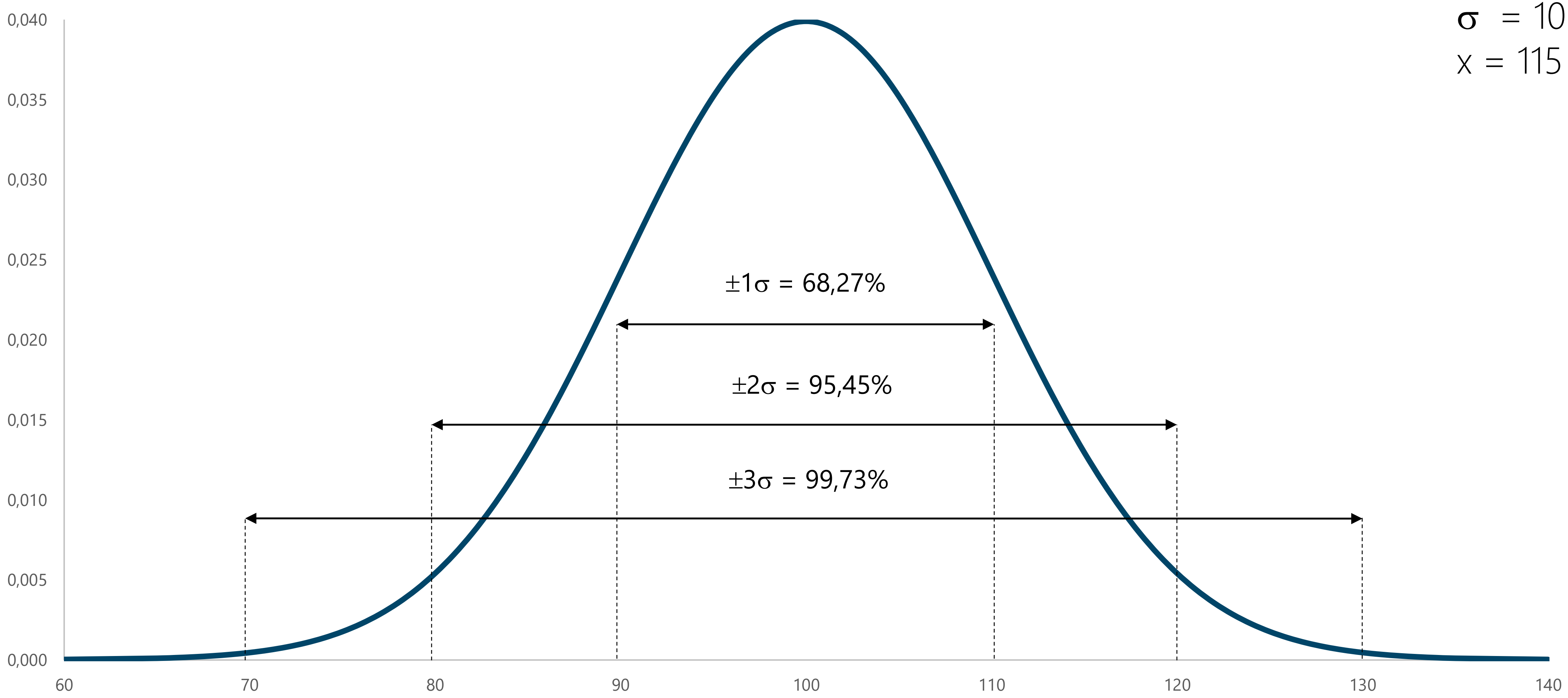
gdje je

- μ - aritmetička sredina
- σ - standardna devijacija
- Π *=3.14159..*
- *e* - baza prirodnog logaritma (*e =2,71828…*).

Carl Friedrich Gauss
(1777- 1855)

Normalna ili Gaussova distribucija

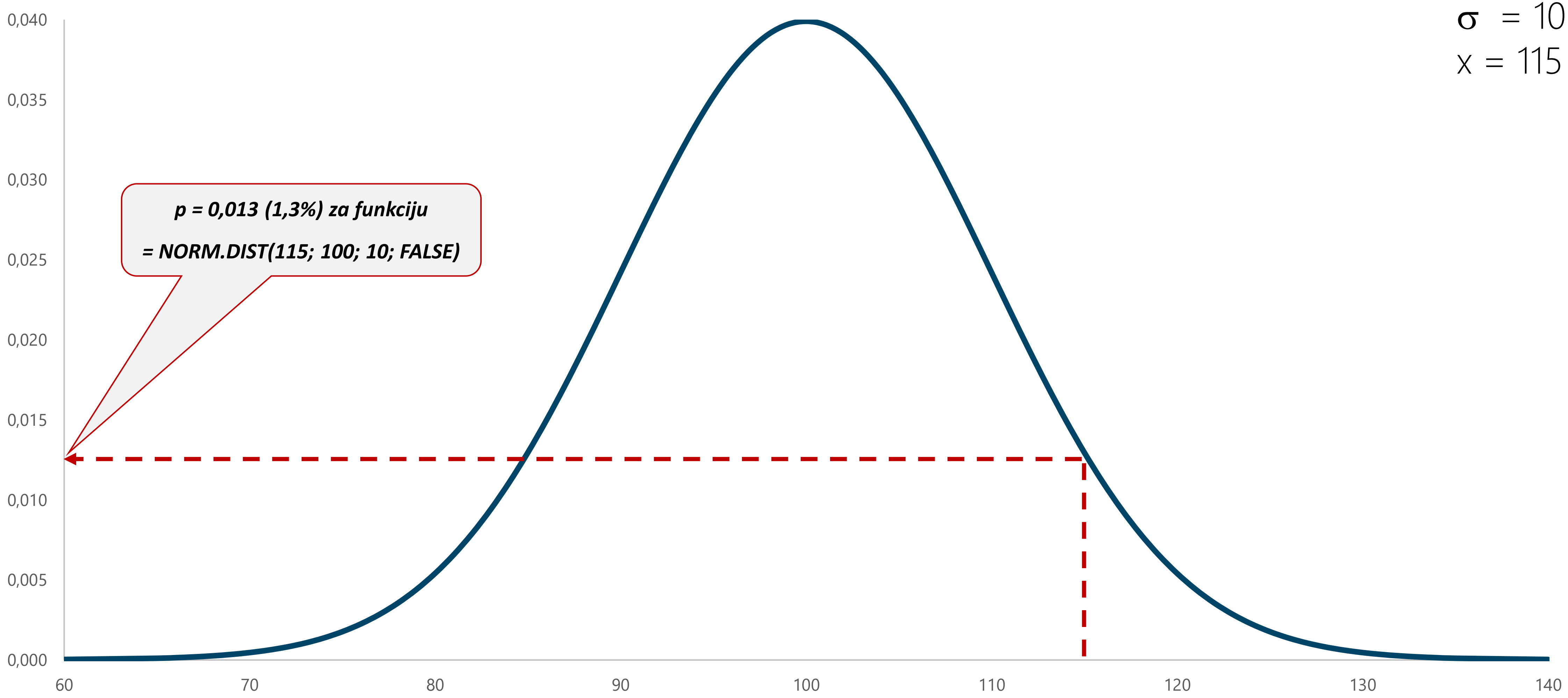
$\mu = 100$
$\sigma = 10$
$x = 115$



Normalna ili Gaussova distribucija

=NORM.DIST(x; μ; σ; FALSE)

$\mu = 100$
$\sigma = 10$
$x = 115$



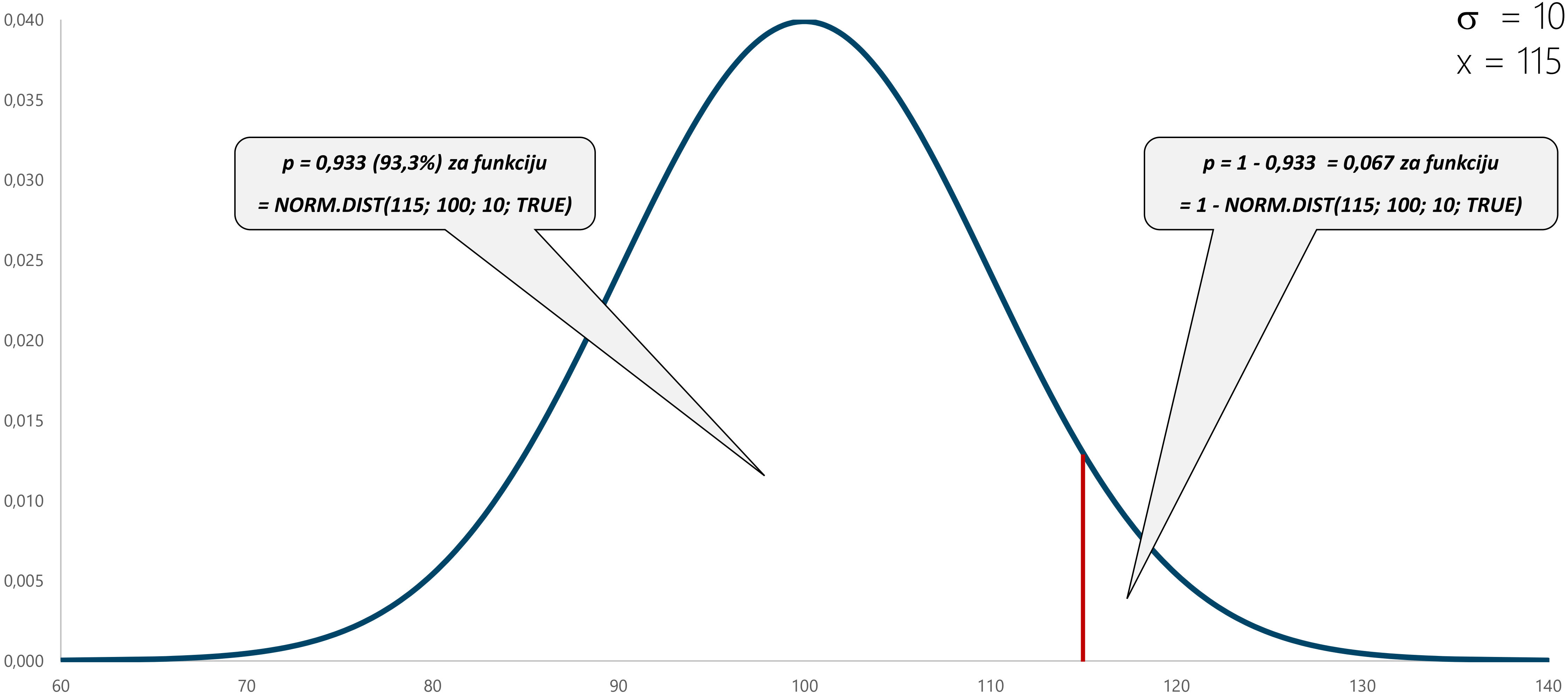
p = 0,013 (1,3%) za funkciju

= NORM.DIST(115; 100; 10; FALSE)

Normalna ili Gaussova distribucija

=NORM.DIST(x; μ; σ; TRUE)
=1 - NORM.DIST(x; μ; σ; TRUE)

μ = 100
σ = 10
x = 115



p = 0,933 (93,3%) za funkciju

= NORM.DIST(115; 100; 10; TRUE)

p = 1 - 0,933 = 0,067 za funkciju

= 1 - NORM.DIST(115; 100; 10; TRUE)

Microsoft Excel

Zadatak 1: U datoteci *JUDO3F.xls* izračunajte aritmetičku sredinu, mod, medijan, minimum, maksimum, raspon, varijancu, standardnu devijaciju, koeficijent varijabilnosti, skewness i kurtosis za sve kvantitativne varijable!

Napomena: Rezultate prikažite tako da su u recima varijable, a u stupcima oznake deskriptivnih pokazatelja.

1

IZRAČUNAVANJE DESKRIPTIVNIH POKAZATELJA vrši se pomoću funkcija:

- ***Average*** (aritmetička sredina)
- ***Mode*** (mod)
- ***Median*** (medijan)
- ***Min*** (minimum)
- ***Max*** (maksimum)
- ***Stdev*** (standardna devijacija)
- ***Var*** (varijanca)
- ***Skew*** (skewness) i
- ***Kurt*** (kurtosis).

=AVERAGEA(b2:b61)

Funkcije za izračunavanje vrijednosti označenog polja unose se u traku *fx. Npr. =AVERAGE(b2:b61).*

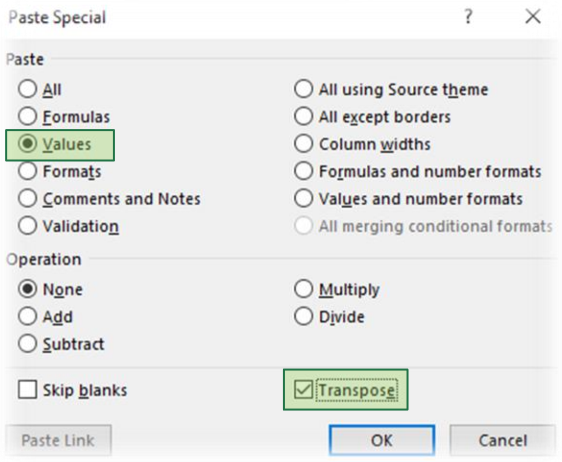
Microsoft Excel

Zadatak 1: U datoteci *JUDO3F.xls* izračunajte aritmetičku sredinu, mod, medijan, minimum, maksimum, raspon, varijancu, standardnu devijaciju, koeficijent varijabilnosti, skewness i kurtosis za sve kvantitativne varijable!

Napomena: Rezultate prikažite tako da su u recima varijable, a u stupcima oznake deskriptivnih pokazatelja.

2 **TRANSPONIRANJE MATRICE** vrši se pomoću funkcija *Trasnpose*. Postupak se provodi na sljedeći način:

1. Označite ćeliju u koju želite prikazati trasnponiranu matricu (tablicu).
2. Označite matricu (tablicu) koju želite transponirati (zamjeniti redke stupcima, a stupce redcima) te odaberete kombinaciju tipki Ctrl + C.
3. U izborniku *Home* odaberete opciju *Paste*, pa potom opciju *Paste Specijal...*
4. U dijaloškom okviru *Past Specijal* odaberete opcije *Values* i *Transpose*.


Paste Special

Paste

○ All
○ Formulas
● Values
○ Formats
○ Comments and Notes
○ Validation
○ All using Source theme
○ All except borders
○ Column widths
○ Formulas and number formats
○ Values and number formats
○ All merging conditional formats

Operation

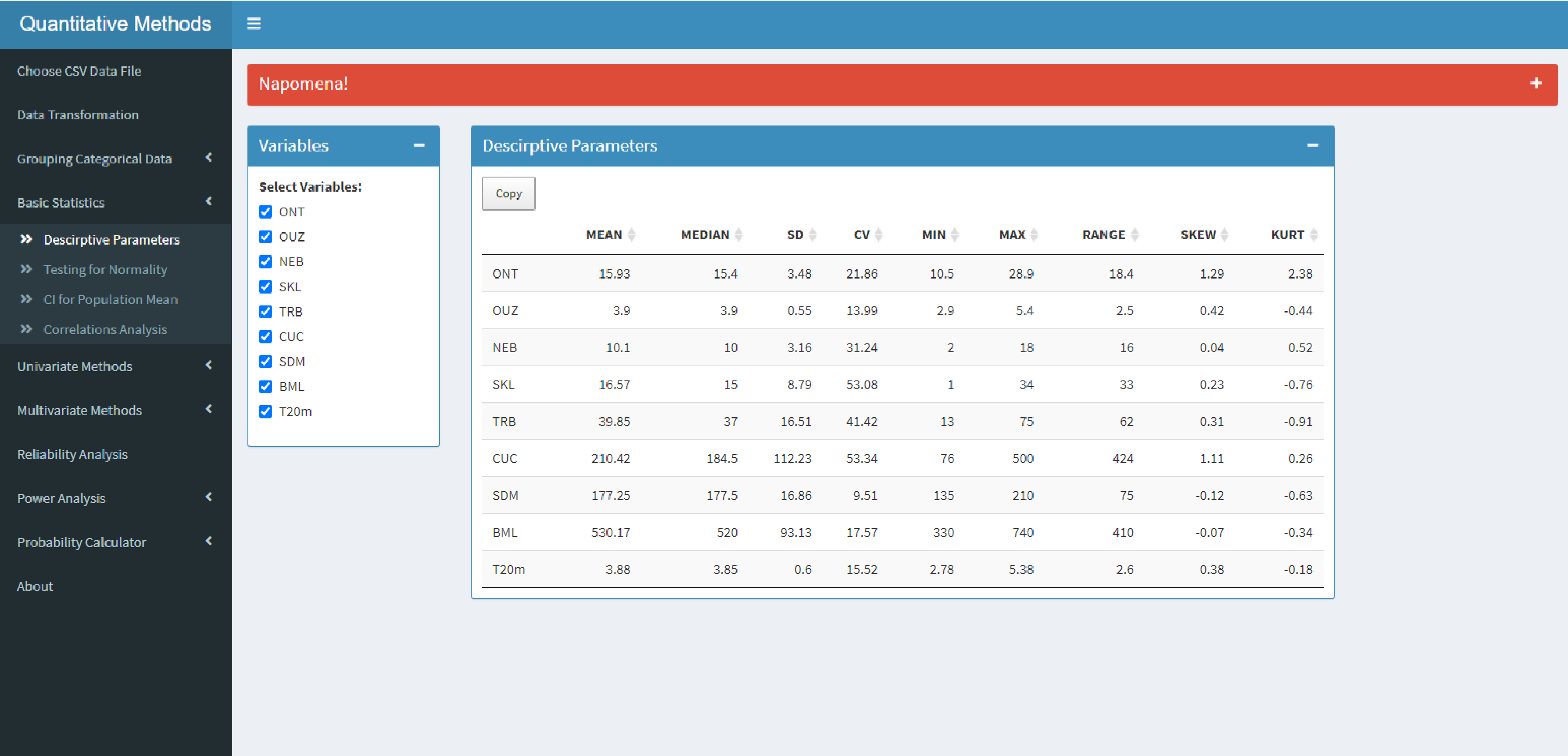
● None
○ Add
○ Subtract
○ Multiply
○ Divide

☐ Skip blanks
☑ Transpose

RStudio – Descriptive Statistics

Zadatak 2: U datoteci *JUDO3F.xls* izračunajte aritmetičku sredinu, mod, medijan, minimum, maksimum, raspon, varijancu, standardnu devijaciju, koeficijent varijabilnosti, skewness i kurtosis za sve kvantitativne varijable!

RStudio – Descriptive Statistics

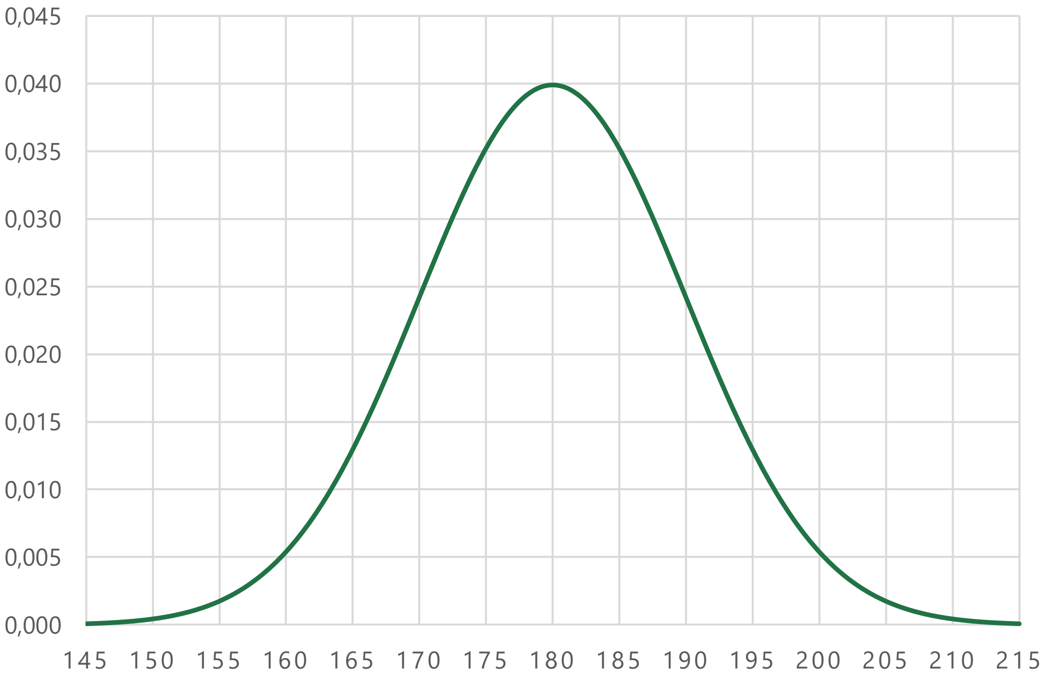


Select Variables:

☑ ONT
☑ OUZ
☑ NEB
☑ SKL
☑ TRB
☑ CUC
☑ SDM
☑ BML
☑ T20m

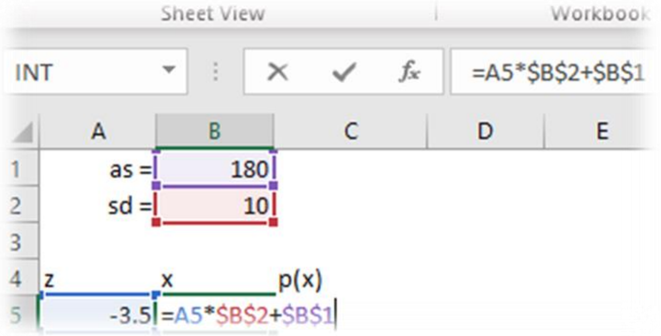
	MEAN	MEDIAN	SD	CV	MIN	MAX	RANGE	SKEW	KURT
ONT	15.93	15.4	3.48	21.86	10.5	28.9	18.4	1.29	2.38
OUZ	3.9	3.9	0.55	13.99	2.9	5.4	2.5	0.42	-0.44
NEB	10.1	10	3.16	31.24	2	18	16	0.04	0.52
SKL	16.57	15	8.79	53.08	1	34	33	0.23	-0.76
TRB	39.85	37	16.51	41.42	13	75	62	0.31	-0.91
CUC	210.42	184.5	112.23	53.34	76	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

Zadatak 3: Prikažite grafički Gaussovu distribuciju s aritmetičkom sredinom (as= 180 cm) i standardnom devijacijom (sd=10 cm).



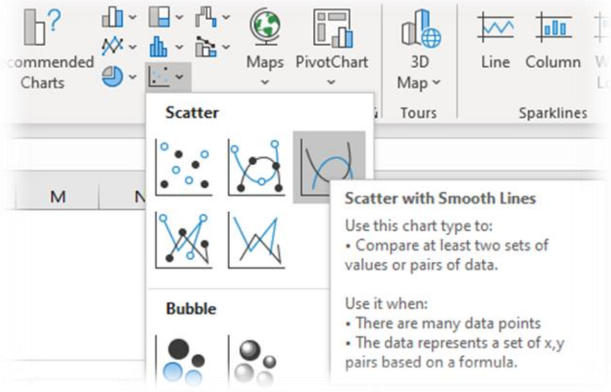
3 GRAFIČKI PRIKAZ NORMALNE DISTRIBUCIJE:

- kreirati varijablu „z” čiji se rezultati kreću od -3,5 do 3,5
- kreirati varijablu „x” formulom x = z*sd + as
- kreirati varijablu „p(x)” pomoću funkcije =NORM.DIST(x; as; sd; FALSE)
- označiti varijable x i p(x) te odabrati grafikon *Scatter with Smooth Line*



=A5*$B$2+$B$1

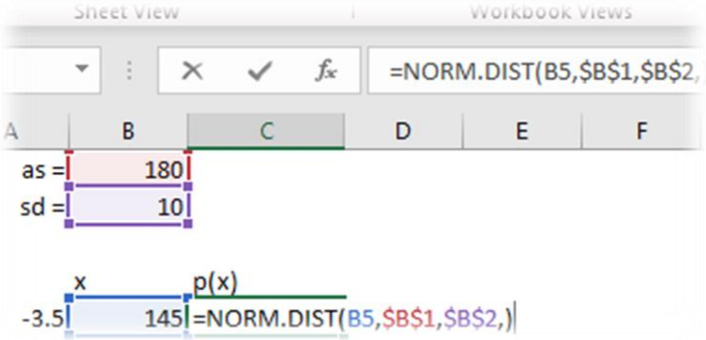
	A	B	C
1	as =	180	
2	sd =	10	
3			
4	z	x	p(x)
5	-3.5	=A5*$B$2+$B$1	



Scatter with Smooth Lines

Use this chart type to:
- Compare at least two sets of values or pairs of data.

Use it when:
- There are many data points
- The data represents a set of x,y pairs based on a formula.



=NORM.DIST(B5,$B$1,$B$2,

	B	C
as =	180	
sd =	10	
	x	p(x)
-3.5	145	=NORM.DIST(B5,$B$1,$B$2,)

Microsoft Excel

Zadatak 4: 257 dječaka je izmjereno testom za procjenu eksplozivne snage *Skok udalj s mjesta*. Aritmetička sredina iznosi 215 cm, a standardna devijacija 12 cm. Učenik *XY* postigao je rezultat 230 cm. Potrebno je procijeniti postotak (%) i broj učenika koji postižu lošije i bolje rezultate od učenika *XY*.

$$z_{XY} = \frac{230-215}{12} = 1{,}25s$$

p=NORM.S.DIST(1,25; TRUE)

$p = 0{,}8944 \rightarrow 89{,}44\%$ *lošijih*

$1 - p = 0{,}1056 \rightarrow 10{,}56\%$ *boljih*

$p \cdot n = 0{,}1056 \cdot 257 \approx 27 -$ *boljih*

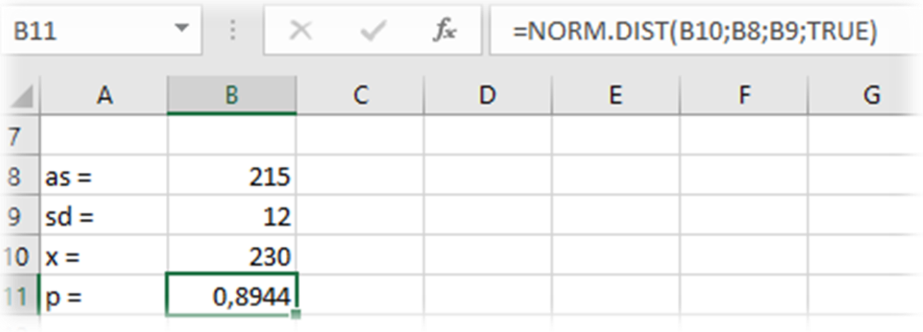
$257 - 27 = 230$ *lošijih*

3 RAČUNANJE POVRŠINE ISPOD NORMALNE DISTRIBUCIJE:

Površina ispod normalne distribucije od x do lijevog kraja distribucije računa se formulom:

= NORM.DIST(x; as; sd; TRUE) i predstavlja vjerojatnost

postizanja većih rezultata od x.



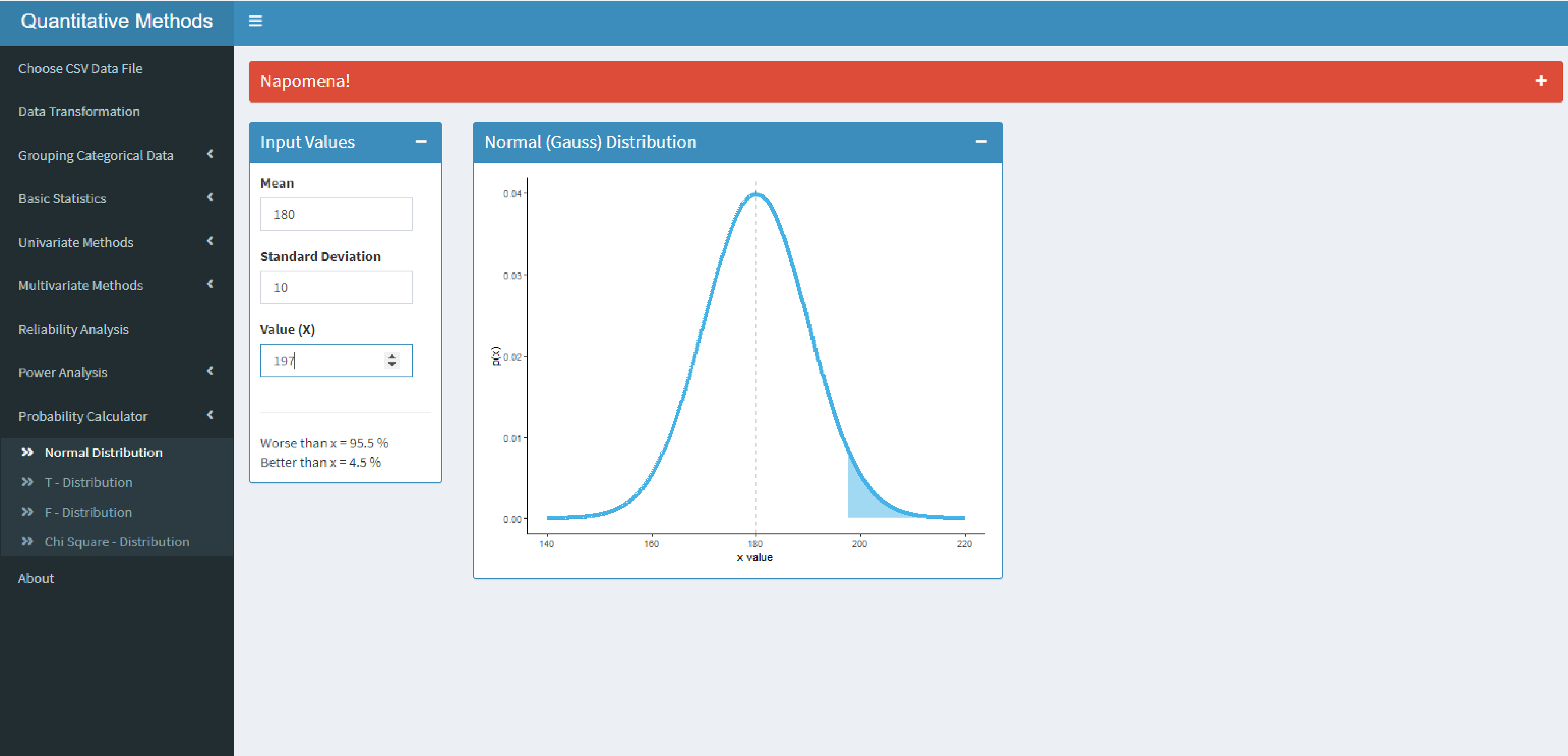
B11 | =NORM.DIST(B10;B8;B9;TRUE)

	A	B	C	D	E	F	G
7							
8	as =	215					
9	sd =	12					
10	x =	230					
11	p =	0,8944					

RStudio – Normal Distribution

Zadatak 5: Izračunajte postotak ispitanika čiji je rezultat veći od 197 cm u grupi s aritmetičkom sredinom (as= 180 cm) i standardnom devijacijom (sd=10 cm).

RStudio – Normal Distribution



Mean: 180

Standard Deviation: 10

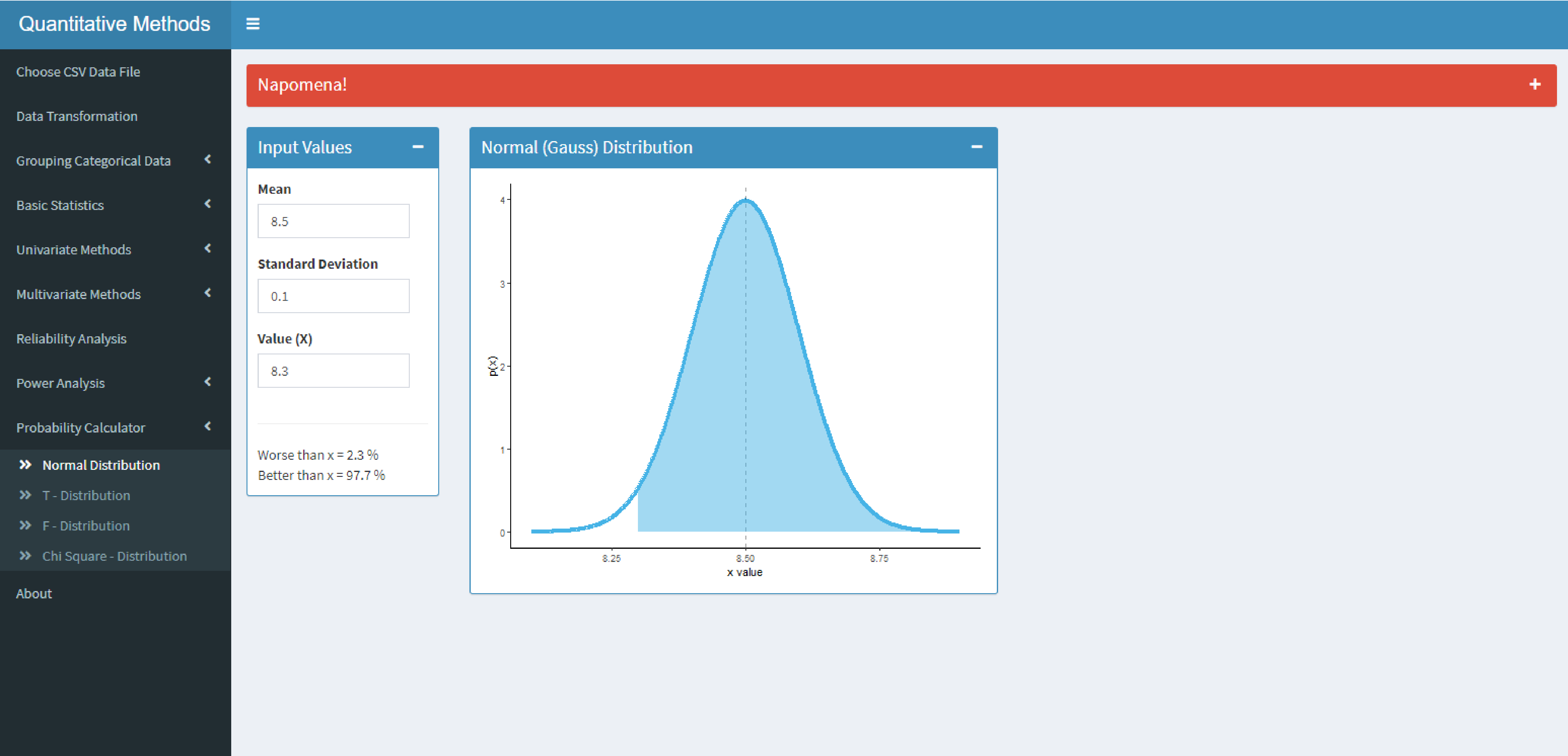
Value (X): 197

Worse than x = 95.5 %
Better than x = 4.5 %

RStudio – Normal Distribution

Zadatak 6: Aritmetička sredina svih službenih skokova udalj nekog atletičara iznosi 8,5 metara, a standardna devijacija 0,1 metar. Putem normalne distribucije procijenite kolika je vjerojatnost da ovaj atletičar na natjecanju izvede skok kraći od 8,3 metara (klasifikacijska norma za OI).

RStudio – Normal Distribution



Mean
8.5

Standard Deviation
0.1

Value (X)
8.3

Worse than x = 2.3 %
Better than x = 97.7 %

TRANSFORMACIJE PODATAKA

Vježba 4

Transformacije podataka

1. Rangiranje
2. Z - vrijednosti
3. Utvrđivanje vjerojatnosti putem normalne distribucije
4. Ostale transformacije podataka (skala školskih ocjena skala T – skorova)

Transformacije podataka

Rangiranje

Rangiranje je transformacija kvantitativne (intervalne ili omjerne) varijable u ordinalnu, odnosno zamjena rezultata odgovarajućim rangovima.

Rang je redni broj entiteta utvrđen sortiranjem prema nekoj kvantitativnoj varijabli.

Uobičajeno je da se najbolji rezultat zamjenjuje rangom *1*. Sukladno tome rang se entiteta u *normalno skaliranim varijablama* utvrđuje silaznim sortiranjem, a u *obrnuto skaliranim varijablama* uzlaznim sortiranjem.

Normalno skalirana varijabla - najveća numerička vrijednost predstavlja najbolji rezultat (npr. *Skok uvis*)

Obrnuto skalirana varijabla - najmanja numerička vrijednost predstavlja najbolji rezultat (npr. *Sprint 100 metara*)

Transformacije podataka

Rangiranje

Primjer: 10 sportaša je izmjereno testom Skok udalj s mjesta. Rangiranje je izvršeno sljedećim postupkom:

	SDM
D.P.	315
J.K.	260
B.V.	330
G.I.	265
I.M.	240
T.K.	290
D.R.	295
I.K.	270
S.J.	235
Z.P.	270

silazno sortiranje →

	SDM
B.V.	330
D.P.	315
D.R.	295
T.K.	290
Z.B.	275
D.Š.	275
I.K.	270
Z.P.	270
G.I.	265
J.K.	260

utvrđivanje rangova →

	RANG
B.V.	1
D.P.	2
D.R.	3
T.K.	4
Z.B.	5
D.Š.	5
I.K.	7
Z.P.	7
G.I.	9
J.K.	10

Transformacije podataka

Z - vrijednosti

Postupak standardizacije provodi se pomoću formule

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

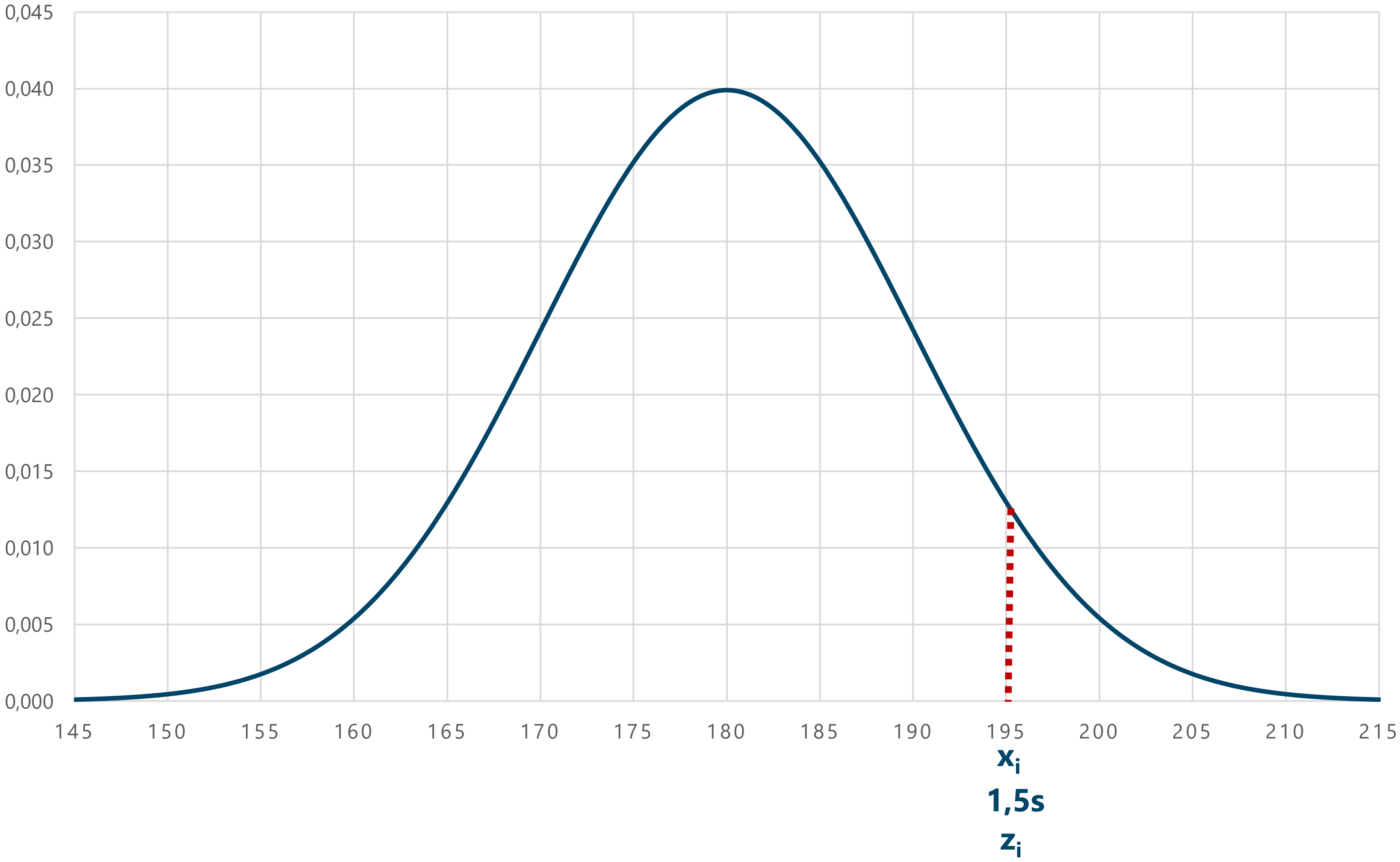
gdje je:
z_i – standardizirani rezultat entiteta i
x_i – originalna vrijednost entiteta i
$\bar{x}$ – aritmetička sredina
s – standardna devijacija.

Transformacije podataka

Z - vrijednosti

$\bar{x}$=180 cm
s =10 cm
x_i = 195 cm

$$z_i = \frac{195-180}{10} = 1{,}5s$$



Transformacije podataka

Z - vrijednosti

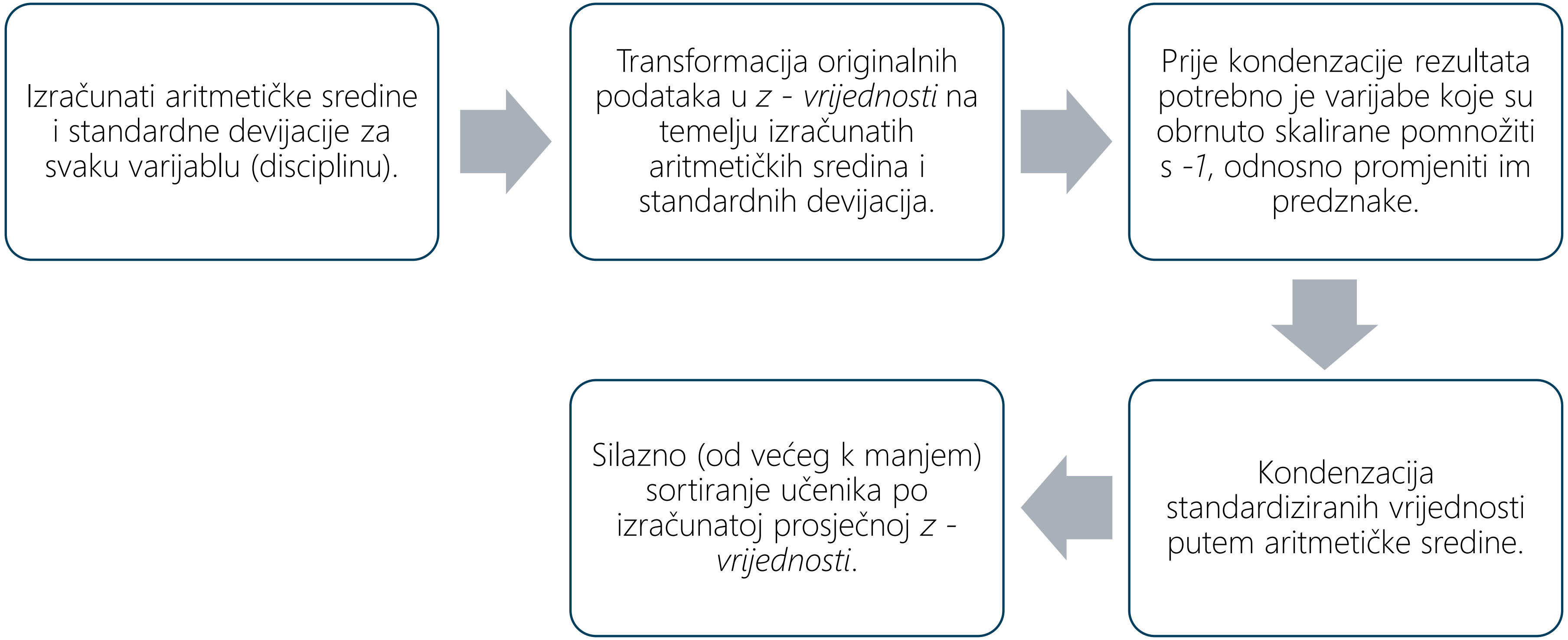
Primjer: Deset učenika natjecalo se u tri atletske discipline:
- skok udalj (SD),
- trčanje na 100 metara (T100m) i
- bacanje kugle (BK) i postiglo rezultate navedene u tablici.

Potrebno je utvrditi ukupan poredak (rang) ovog natjecanja ?

Učenik	SD	T100m	BK
AB	359	13,6	561
DF	321	13,9	550
JG	346	13,7	538
KL	332	14,0	490
DD	450	12,2	518
ED	314	14,1	551
TB	410	12,5	589
ZN	425	12,3	602
RG	369	13,5	547
EN	378	13,8	510

Transformacije podataka

Z - vrijednosti



Izračunati aritmetičke sredine i standardne devijacije za svaku varijablu (disciplinu).

Transformacija originalnih podataka u *z - vrijednosti* na temelju izračunatih aritmetičkih sredina i standardnih devijacija.

Prije kondenzacije rezultata potrebno je varijabe koje su obrnuto skalirane pomnožiti s *-1*, odnosno promjeniti im predznake.

Kondenzacija standardiziranih vrijednosti putem aritmetičke sredine.

Silazno (od većeg k manjem) sortiranje učenika po izračunatoj prosječnoj *z - vrijednosti*.

Transformacije podataka

Z - vrijednosti

Prvi korak: Izračunati aritmetičke sredine i standardne devijacije za svaku varijablu (disciplinu).

	SD	T100m	BK
$\bar{x}$	370,4	13,36	545,6
s	45,66	0,73	34,21

Učenik	SD	T100m	BK
AB	359	13,6	561
DF	321	13,9	550
JG	346	13,7	538
KL	332	14,0	490
DD	450	12,2	518
ED	314	14,1	551
TB	410	12,5	589
ZN	425	12,3	602
RG	369	13,5	547
EN	378	13,8	510

Transformacije podataka

Z - vrijednosti

Drugi korak: Transformacija originalnih podataka u *z - vrijednosti* na temelju izračunatih aritmetičkih sredina i standardnih devijacija.

	SD	T100m	BK
$\bar{x}$	370,4	13,36	545,6
s	45,66	0,73	34,21

$$z_{AB,SD} = \frac{359-370,4}{45,66} = -0,25s$$

Učenik	SD	T100m	BK
AB	**359**	13,6	561
DF	321	13,9	550
JG	346	13,7	538
KL	332	14,0	490
DD	450	12,2	518
ED	314	14,1	551
TB	410	12,5	589
ZN	425	12,3	602
RG	369	13,5	547
EN	378	13,8	510

Transformacije podataka

Z - vrijednosti

Treći korak: Prije kondenzacije rezultata varijable koje su obrnuto skalirane pomnožiti s *-1*, odnosno rezultatima promijeniti predznake.

Učenik	SD	T100m	BK
AB	-0,25	0,33	0,45
DF	-1,08	0,74	0,13
JG	-0,53	0,46	-0,22
KL	-0,84	0,87	-1,63
DD	1,74	-1,58	-0,81
ED	-1,24	1,01	0,16
TB	0,87	-1,17	1,27
ZN	1,20	-1,44	1,65
RG	-0,03	0,19	0,04
EN	0,17	0,60	-1,04

Transformacije podataka

Z - vrijednosti

Četvrti korak: Kondenzacija standardiziranih vrijednosti putem aritmetičke sredine.

$$\bar{z}_{AB} = \frac{z_{AB,SD} + z_{AB,T100m} + z_{AB,BK}}{3} =$$

$$= \frac{-0{,}25 + (-0{,}33) + 0{,}45}{3} = -0{,}04$$

Učenik	SD	T100m	BK	$\bar{Z}$
AB	-0,25	-0,33	0,45	**-0,04**
DF	-1,08	-0,74	0,13	
JG	-0,53	-0,46	-0,22	
KL	-0,84	-0,87	-1,63	
DD	1,74	1,58	-0,81	
ED	-1,24	-1,01	0,16	
TB	0,87	1,17	1,27	
ZN	1,20	1,44	1,65	
RG	-0,03	-0,19	0,04	
EN	0,17	-0,60	-1,04	

Transformacije podataka

Z - vrijednosti

Četvrti korak: Kondenzacija standardiziranih vrijednosti putem aritmetičke sredine.

$$\bar{z}_{AB} = \frac{z_{AB,SD} + z_{AB,T100m} + z_{AB,BK}}{3} =$$

$$= \frac{-0{,}25 + (-0{,}33) + 0{,}45}{3} = -0{,}04$$

Učenik	**SD**	**T100m**	**BK**	$\overline{Z}$
AB	-0,25	-0,33	0,45	**-0,04**
DF	-1,08	-0,74	0,13	**-0,56**
JG	-0,53	-0,46	-0,22	**-0,41**
KL	-0,84	-0,87	-1,63	**-1,11**
DD	1,74	1,58	-0,81	**0,84**
ED	-1,24	-1,01	0,16	**-0,70**
TB	0,87	1,17	1,27	**1,10**
ZN	1,20	1,44	1,65	**1,43**
RG	-0,03	-0,19	0,04	**-0,06**
EN	0,17	-0,60	-1,04	**-0,49**

Transformacije podataka

Z - vrijednosti

Peti korak: Silazno (od većeg k manjem) sortiranje učenika po izračunatoj prosječnoj *z - vrijednosti*.

Učenik	SD	$\overline{z}$
ZN	1	1,43
TB	2	1,10
DD	3	0,84
AB	4	-0,04
RG	5	-0,06
JG	6	-0,41
EN	7	-0,49
DF	8	-0,56
ED	9	-0,70
KL	10	-1,11

Transformacije podataka

Ostale transformacije

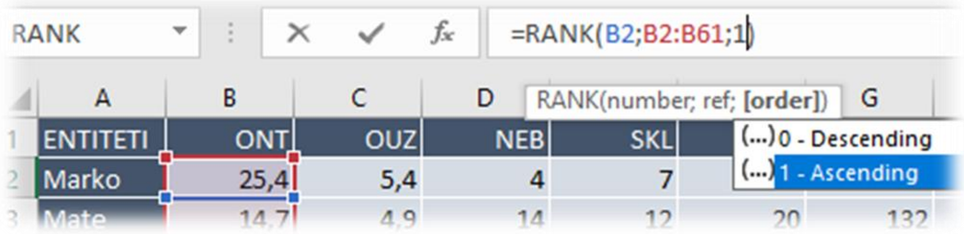
Standardizirane rezultate moguće je transformirati na različite načine, zavisno o potrebi. Najčešće se transformiraju u vrijednosti na sljedećim skalama:

- skala školskih ocjena (1-5): $\boldsymbol{ocjena_i = 3 + 0{,}83 \cdot \ z_i}$
- skala T – skorova (20-80): $\boldsymbol{T - skor_i = 50 + 10 \cdot \ z_i}$

Microsoft Excel

Zadatak 1: U datoteci *JUDO3F.xls* utvrdite rangove judaša na temelju rezultata u varijabli ONT (okretnost na tlu). Iskoristite funkciju Rank.

1 **UTVRĐIVANJE RANGA:** Utvrđivanje rangova vrši se pomoću funkcije *Rank*. Funkcija se unosi u označeno polje matrice odabirom opcije *Function…*. U traku *Order* dijaloškog okvira za unos ove funkcije potrebno je upisati *1* ako se utvrđuje rang u obrnuto skaliranoj varijabli.



```
=RANK(B2;B2:B61;1)
```

	A	B	C	D	E	F	G
1	ENTITETI	ONT	OUZ	NEB	SKL		
2	Marko	25,4	5,4	4	7		
3	Mate	14,7	4,9	14	12	20	132

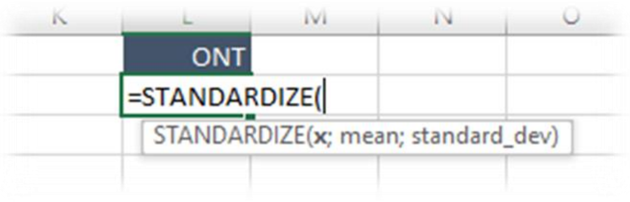
Microsoft Excel

Zadatak 2: U datoteci *JUDO3F.xls* izračunajte z-vrijednosti za sve varijable i rangirajte entitete po prosječnoj z-vrijednosti svih varijabli.

Napomena: Varijable ONT, OUZ i T20m su obrnuto skalirane.

2

RAČUNANJE Z - VRIJEDNOSTI: Izračunavanje z-vrijednosti vrši se pomoću funkcije *Standardize*. Funkcija se unosi u označeno polje matrice odabirom opcije *Function...* Prethodno je potrebno izračunati aritmetičku sredinu (funkcija *Average*) i standardnu devijaciju (funkcija *Stdev*) varijable.



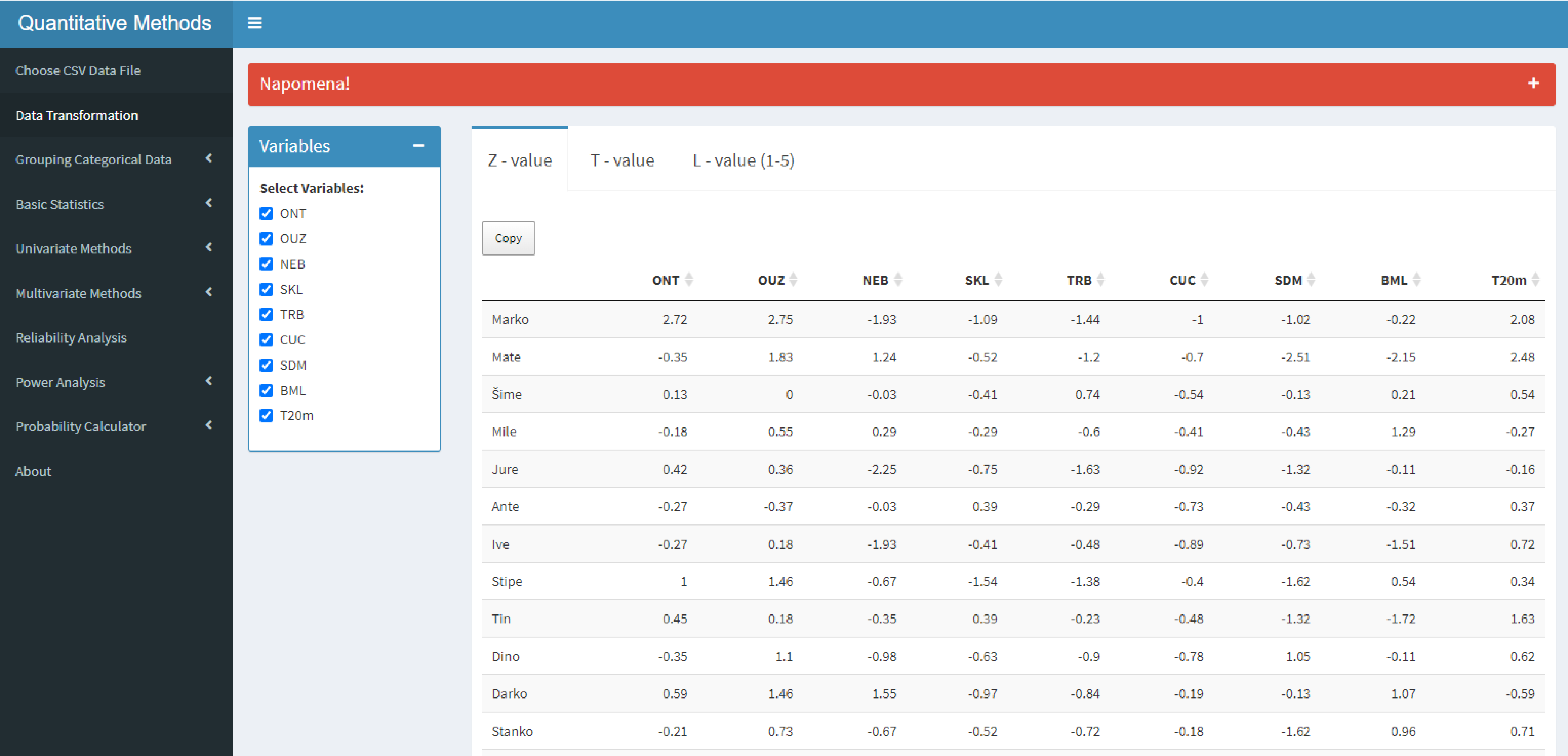
ONT

`=STANDARDIZE(`

RStudio – Data Transformation

Zadatak 5: U datoteci *JUDO3F.csv* izvršite standardizaciju svih varijabli.

RStudio – Data Transformation



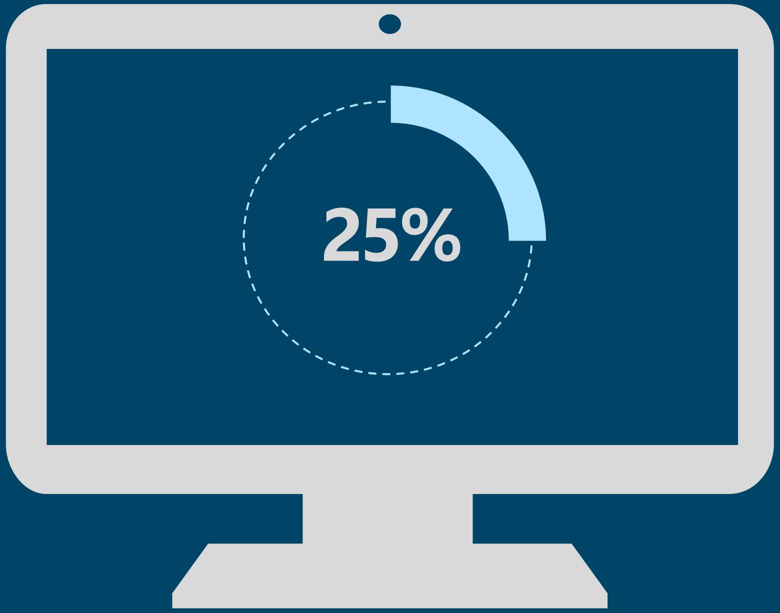
Select Variables:

- ☑ ONT
- ☑ OUZ
- ☑ NEB
- ☑ SKL
- ☑ TRB
- ☑ CUC
- ☑ SDM
- ☑ BML
- ☑ T20m

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

PROCJENA ARITMETIČKE SREDINE POPULACIJE

Vježba 5



Procjena aritmetičke sredine populacije

Statističke metode dijele se na:

- *metode deskriptivne statistike*, tj. postupke za utvrđivanje statističkih parametara koji se odnose isključivo na promatrani uzorak entiteta i
- *metode inferencijalne statistike*, tj. postupke kojima se na temelju statističkih parametara utvrđenih na uzorku entiteta zaključci proširuju na populaciju koje je promatrani uzorak reprezentant.

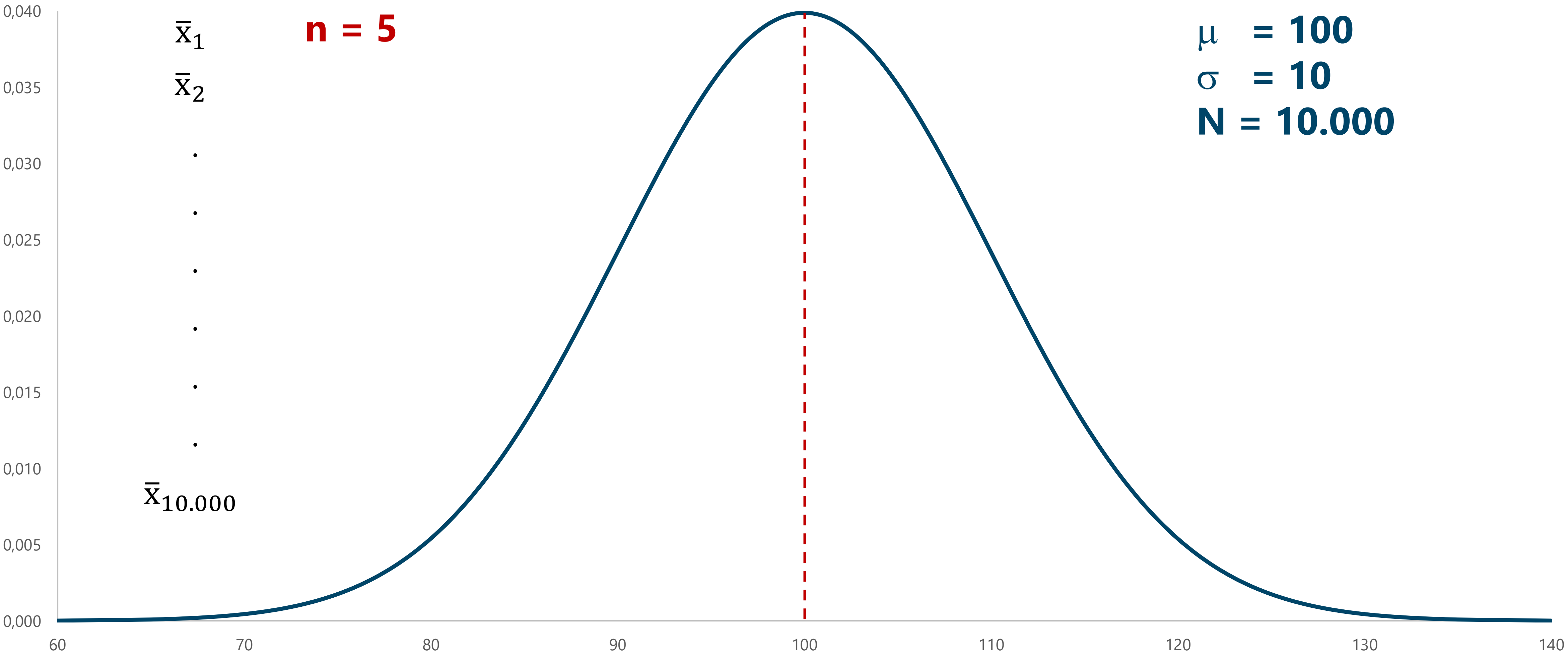
Procjena aritmetičke sredine populacije

Reprezentativnost uzorka utječe na pogrešku s kojom se zaključci generaliziraju na populaciju, a zavisi o načinu odabira entiteta u uzorak i veličini, tj. broju entiteta u uzorku.

Generalizacija zaključaka s uzorka na populaciju bit će ispravna samo ako se uzorak bira na način da svi entiteti iz populacije imaju jednaku vjerojatnost da budu izabrani u uzorak, odnosno ako se radi o slučajnom uzorku.

Reprezentativnost uzorka će biti veća što je broj entiteta u uzorku veći, odnosno bliži broju entiteta u populaciji.

Procjena aritmetičke sredine populacije



$\bar{x}_1$

$\bar{x}_2$

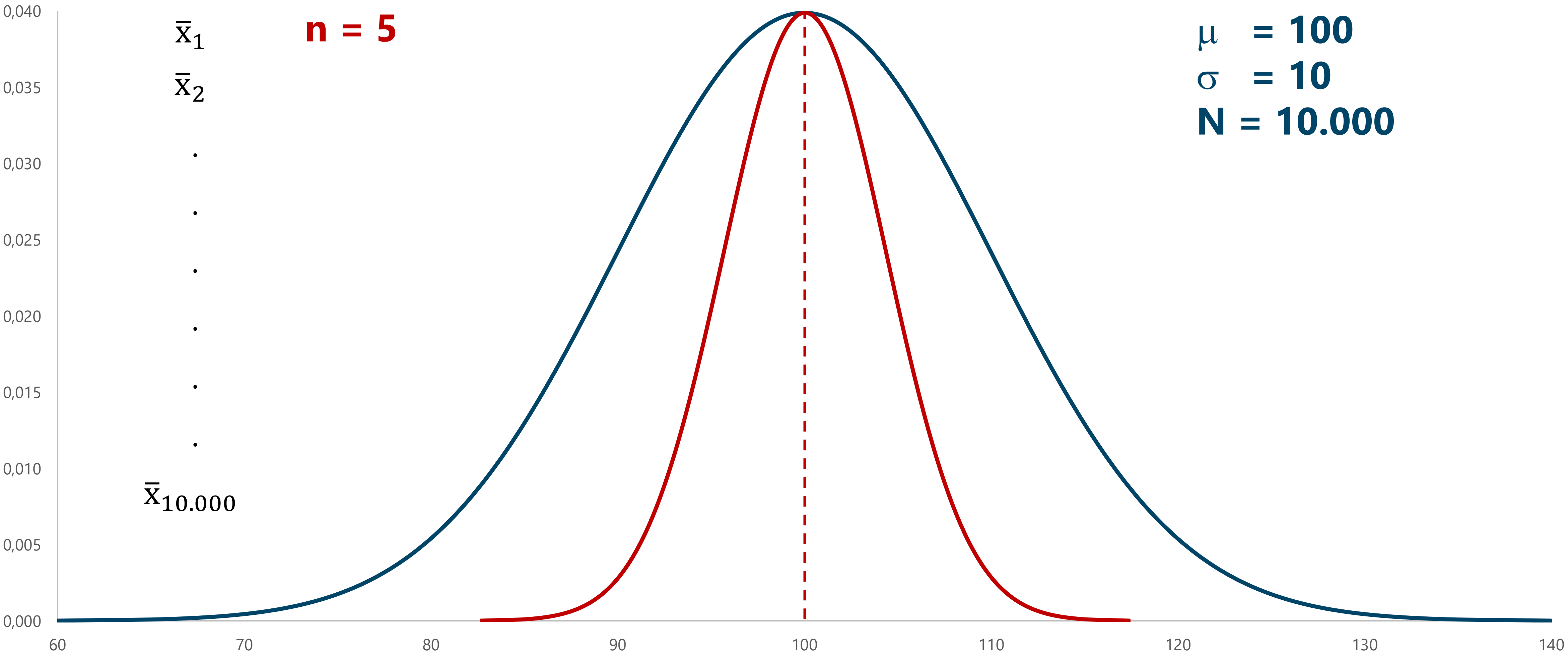
.
.
.
.
.
.

$\bar{x}_{10.000}$

n = 5

μ = 100
σ = 10
N = 10.000

Procjena aritmetičke sredine populacije



$\bar{x}_1$

$\bar{x}_2$

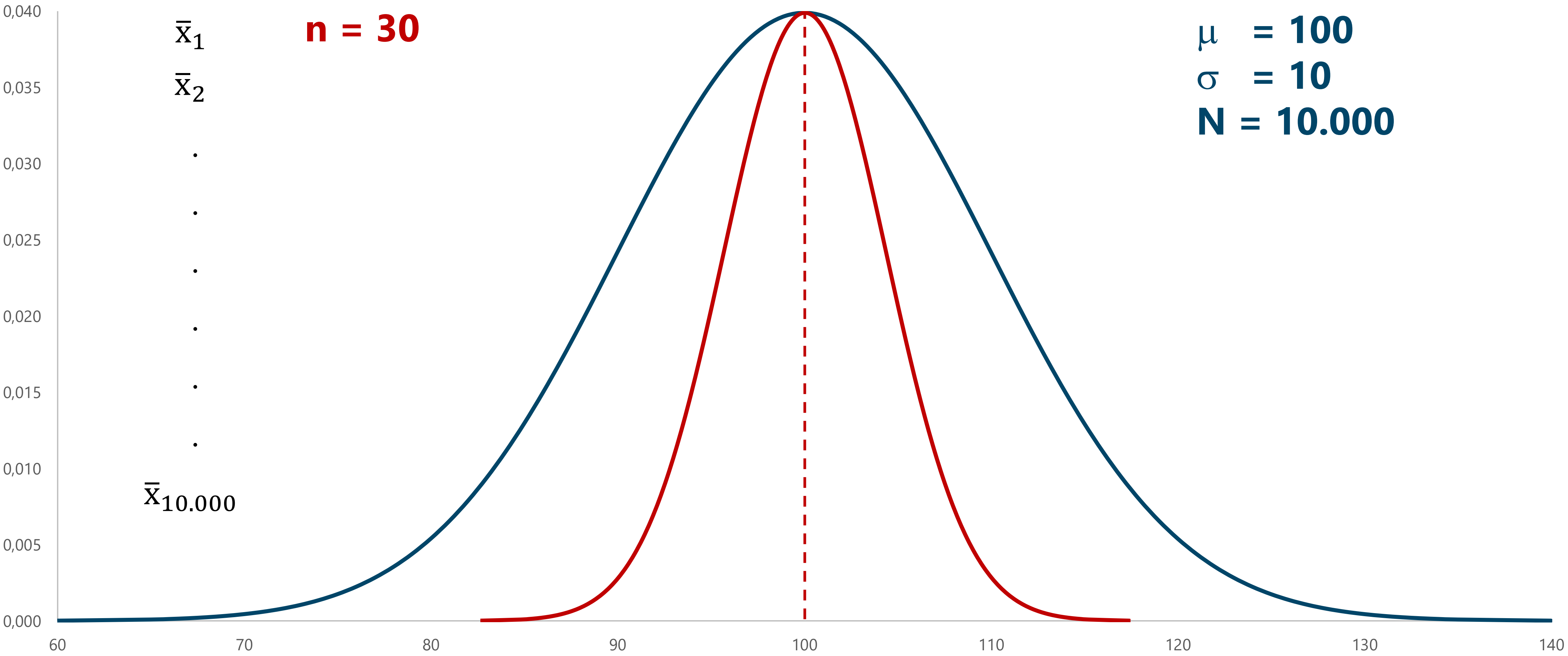
.
.
.
.
.
.

$\bar{x}_{10.000}$

n = 5

μ **= 100**
σ **= 10**
N = 10.000

Procjena aritmetičke sredine populacije



$\bar{x}_1$

$\bar{x}_2$

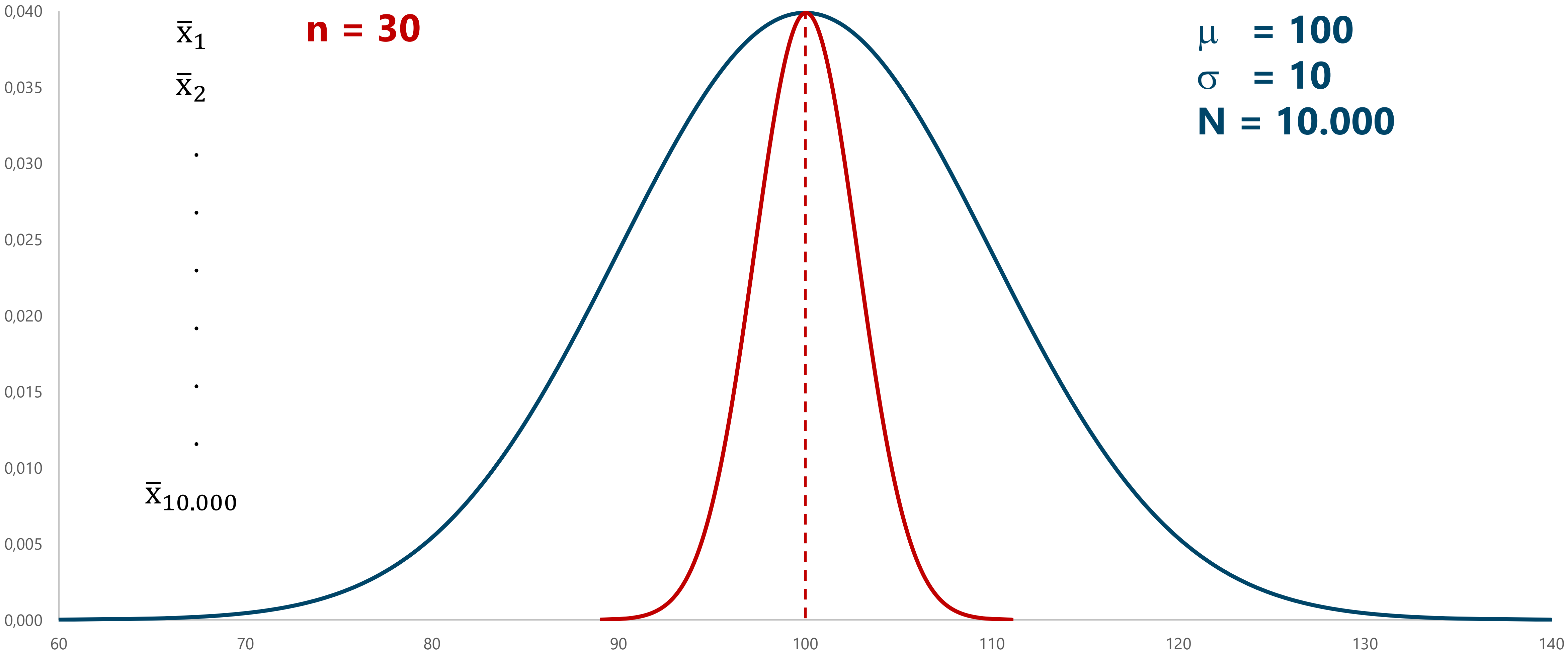
.
.
.
.
.
.

$\bar{x}_{10.000}$

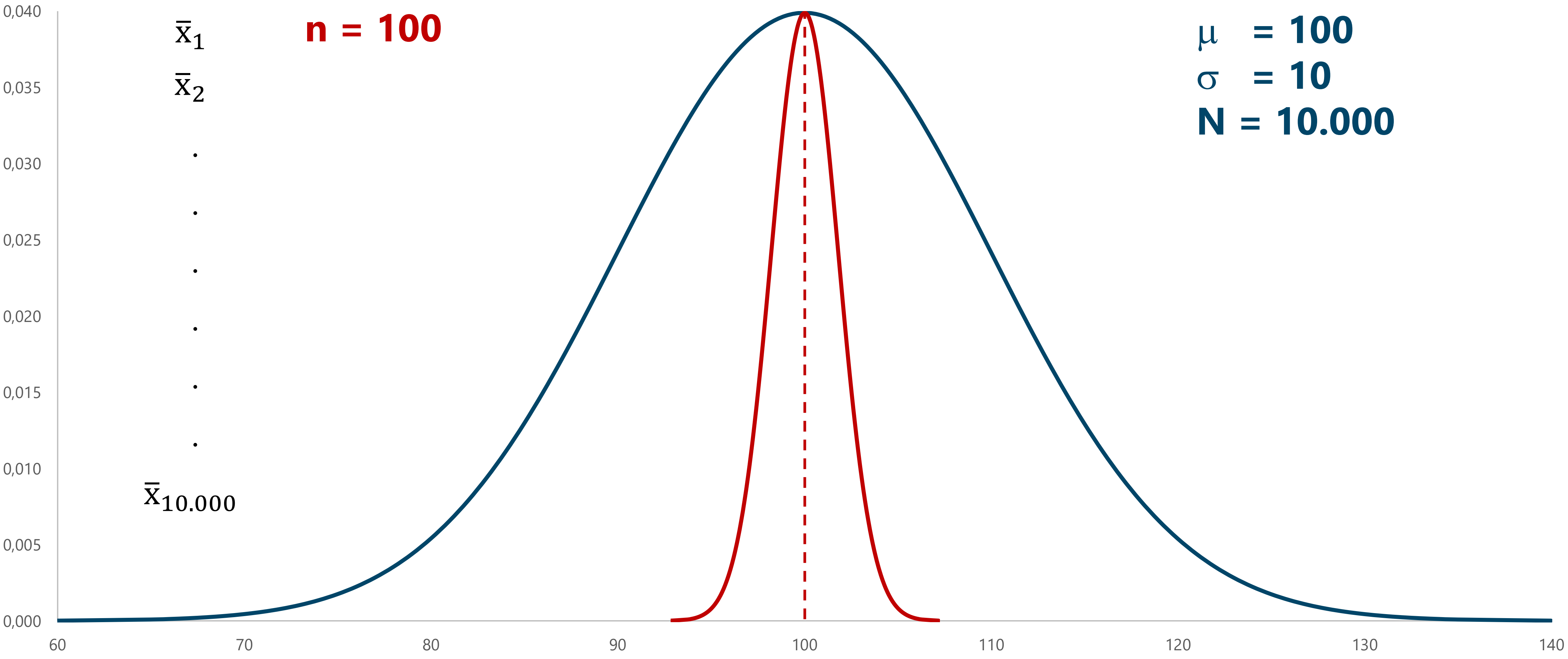
$n = 30$

$\mu = 100$
$\sigma = 10$
$N = 10.000$

Procjena aritmetičke sredine populacije



Procjena aritmetičke sredine populacije



$\bar{x}_1$
$\bar{x}_2$
.
.
.
.
.
.
$\bar{x}_{10.000}$

n = 100

μ = 100
σ = 10
N = 10.000

Procjena aritmetičke sredine populacije

Zaključci:

- aritmetičke sredine slučajnih uzoraka variraju
- distribucija aritmetičkih sredina slučajno odabranih uzoraka iste veličine bit će normalna ili Gaussova
- aritmetička sredina aritmetičkih sredina slučajno odabranih uzoraka jednake veličine tendirat će aritmetičkoj sredini populacije
- standardna devijacija aritmetičkih sredina slučajnih uzoraka biti će manja što je varijabilnost obilježja u populaciji manja i što je broj entiteta u uzorku veći.

Procjena aritmetičke sredine populacije

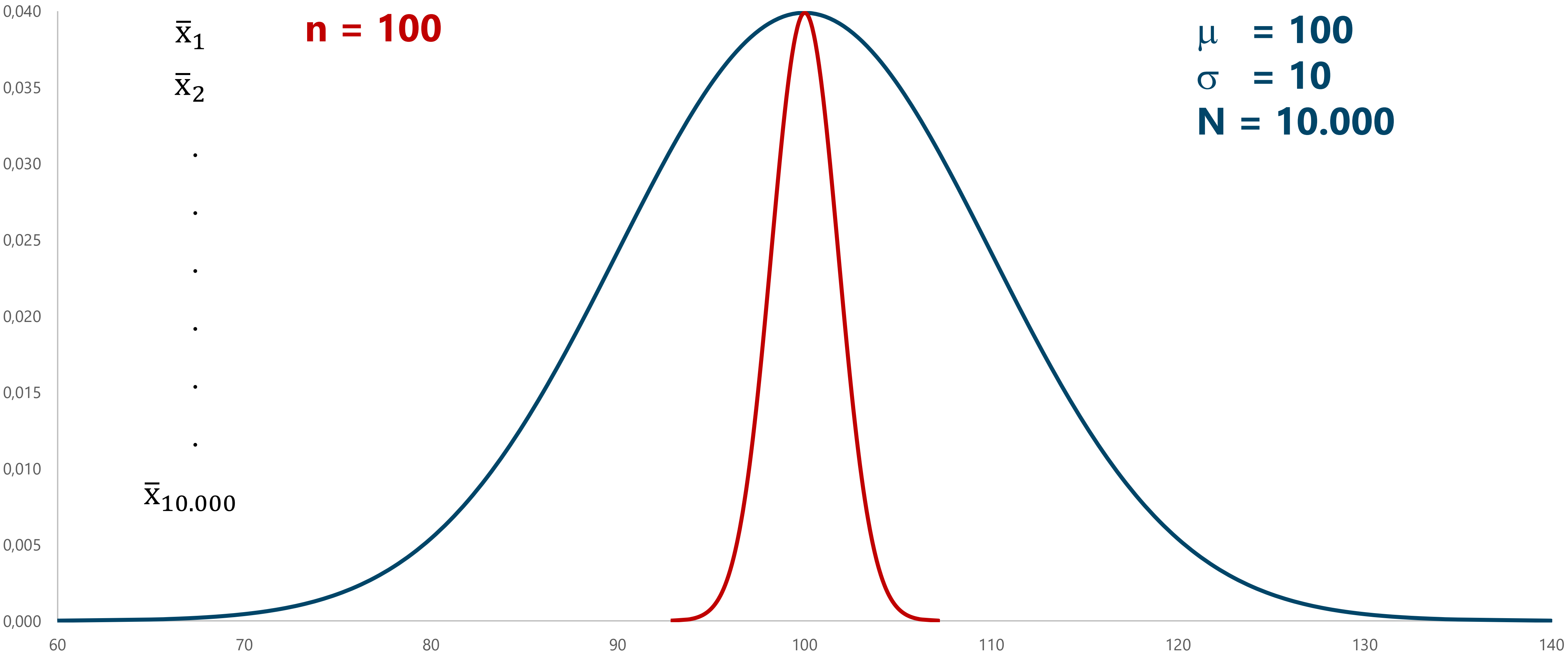
Standardna pogreška aritmetičke sredine

Standardna devijacija aritmetičkih sredina slučajnih uzoraka naziva se ***standardna pogreška aritmetičke sredine***, a izračunava se formulom

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = s_e$$

gdje je
s – standardna devijacija uzorka
n – broj entiteta u uzorku.

Procjena aritmetičke sredine populacije



$\bar{x}_1$

$\bar{x}_2$

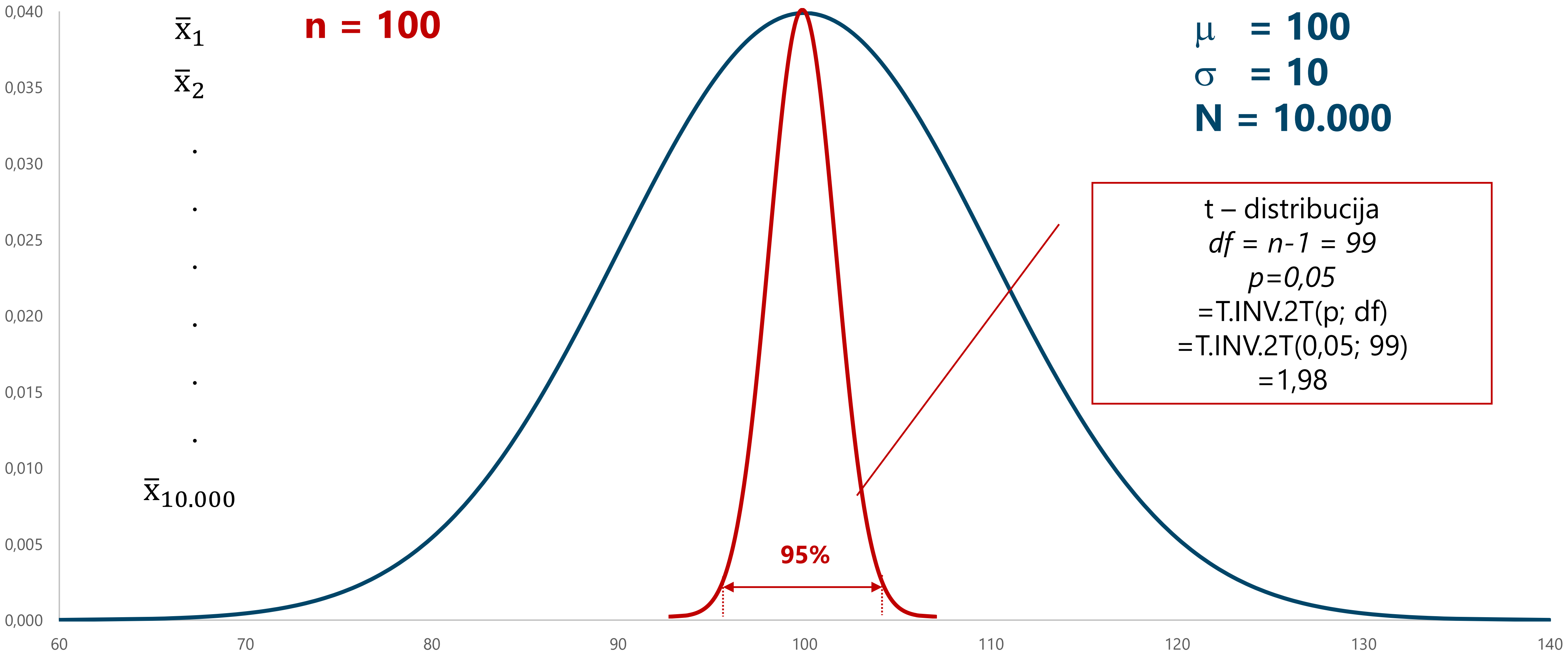
.
.
.
.
.
.

$\bar{x}_{10.000}$

n = 100

μ = 100
σ = 10
N = 10.000

Procjena aritmetičke sredine populacije



$\bar{x}_1$

$\bar{x}_2$

.
.
.
.
.
.

$\bar{x}_{10.000}$

n = 100

μ = 100
σ = 10
N = 10.000

t – distribucija
df = n-1 = 99
p=0,05
=T.INV.2T(p; df)
=T.INV.2T(0,05; 99)
=1,98

$\bar{x}$ - 1,98s_e ~ μ ~ $\bar{x}$ + 1,98s_e

Procjena aritmetičke sredine populacije

Interval u kojem se s određenom vjerojatnošću nalazi aritmetička sredina populacije moguće je procijeniti formulom

$$\bar{x} - t_{df,p} \cdot s_{\bar{x}} < \mu < \bar{x} + t_{df,p} \cdot s_{\bar{x}}$$

gdje je:

- $\bar{x}$ - aritmetička sredina uzorka
- $s_{\bar{x}}$ - standardna pogreška aritmetičke sredine
- $t_{df,p}$ - vrijednost koja se za pogrešku *p* (u statističkom zaključivanju najčešće se koriste pogreške *0,01* ili *1%*, i *0,05* ili *5%*) i određeni broj stupnjeva slobode (*df*=*n*-*1*) odredi na temelju Studentove t-distribucije.

Procjena aritmetičke sredine populacije

Primjer: Na slučajno odabranom uzorku veličine 100 entiteta izračunata je aritmetička sredina 180 *cm* i standardna devijacija 10 *cm*. Potrebno je procijeniti interval u kojem se s vjerojatnošću od 0,95 nalazi aritmetička sredina populacije.

$n = 100$

$\bar{x} = 180\ cm$

$s = 10\ cm$

✓$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{10}{\sqrt{100}} = \frac{10}{10} = 1cm$

✓$t_{99,0,05} = 1{,}98$

✓$\bar{x} - t_{df,p} \cdot s_{\bar{x}} < \ \mu \ < \bar{x} + t_{df,p} \cdot s_{\bar{x}}$

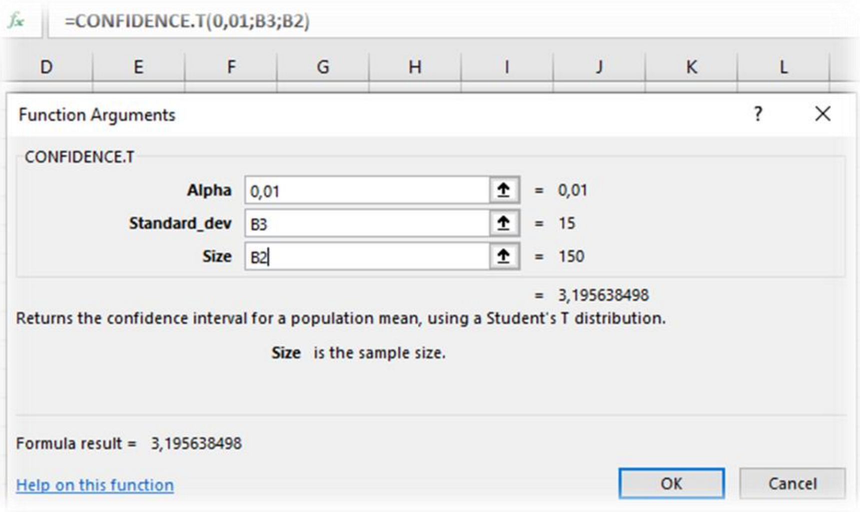
✓$180 - 1{,}98 \cdot 1 < \ \mu \ < 180 + 1{,}98 \cdot 1$

✓$178{,}02\ cm < \mu < 181{,}98\ cm$

Statistički zaključak:
Aritmetička sredina populacije nalazi se u intervalu od 178,02 do 181,98 cm s vjerojatnošću od 95%, odnosno uz pogrešku od 5%.

Zadatak 1: *Neka je* $n = 150$, $\bar{x} = 210$ *cm*, $s = 15$ *cm*. Uz pogrešku p = 0,01 izračunajte interval u kojem se nalazi aritmetička sredina populacije.

1 Interval u kojem se nalazi airtmetička sredina populacije utvrđuje se pomoću funkcije ***CONFIDENCE.T***. Funkcija se unosi u označeno polje odabirom opcije *Function...* U traku *Alpha* dijaloškog okvira za unos ove funkcije potrebno je upisati pogrešku *p*, u traku *Standatd_dev* vrijednost standardne devijacije, te u traku *Size* veličinu uzorka. Dobivena vrijednost se oduzme i zbroji s vrijednošću aritmetičke sredine.



=CONFIDENCE.T(0,01;B3;B2)

Function Arguments

CONFIDENCE.T

Alpha 0,01 = 0,01

Standard_dev B3 = 15

Size B2 = 150

= 3,195638498

Returns the confidence interval for a population mean, using a Student's T distribution.

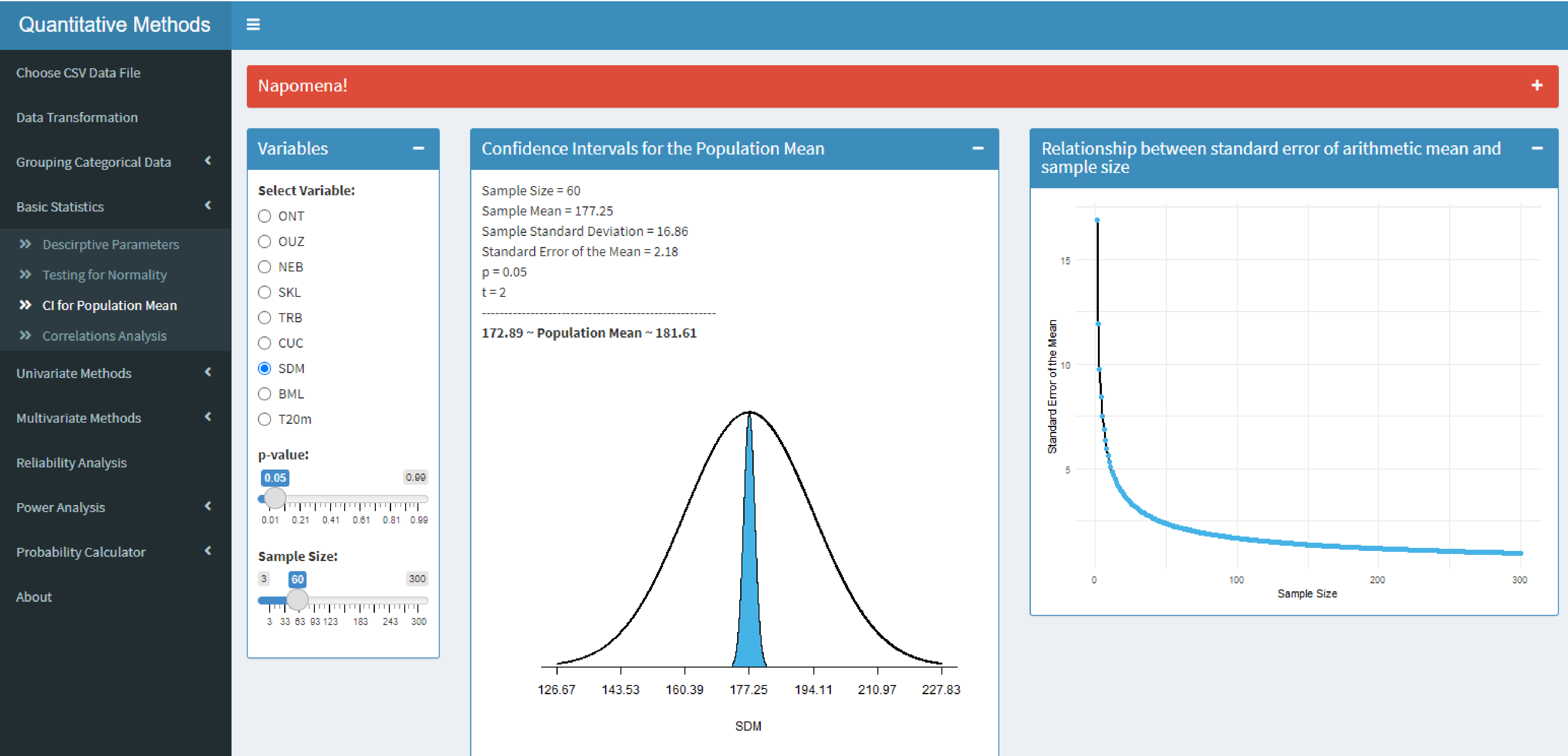
Size is the sample size.

Formula result = 3,195638498

RStudio – CI for Population Mean

Zadatak 2: Izračunajte interval u kojem se nalazi aritmetička sredina populacije za varijablu SDM koja se nalazi u datoteci JUDO3F.xls.

RStudio – CI for Population Mean



Select Variable:

○ ONT
○ OUZ
○ NEB
○ SKL
○ TRB
○ CUC
● SDM
○ BML
○ T20m

p-value: 0.05 (0.05–0.99)

Sample Size: 60 (3–300)

```
Sample Size = 60
Sample Mean = 177.25
Sample Standard Deviation = 16.86
Standard Error of the Mean = 2.18
p = 0.05
t = 2
-------------------------------------------------------
172.89 ~ Population Mean ~ 181.61
```

T - TEST

Vježba 6

T -test

Wiliam Sealy Gosset
(1876. – 1937.)

T-test je statistički postupak koji je osmislio Wiliam Sealy Gosset - Student, a kojim se utvrđuje statistička značajnost razlike između dviju aritmetičkih sredina.

Statistički značajna razlika je razlika utvrđena na uzorku entiteta, a za koju s visokim stupnjem sigurnosti (95% ili 99%) možemo tvrditi da se nije dogodila slučajno (kao posljedica slučajnog variranja aritmetičkih sredina uzoraka u odnosu na aritmetičku sredinu populacije).

T-test za nezavisne uzorke je postupak kojim se utvrđuje statistička značajnost razlike između aritmetičkih sredina dvaju uzoraka (dvaju nezavisnih skupova entiteta).

Pri tome je moguće postaviti sljedeće hipoteze:

H_0: $\overline{x}_1 = \overline{x}_2$ - Uz pogrešku p ne možemo tvrditi da je razlika između aritmetičkih sredina analiziranih uzoraka statistički značajna.

H_1: $\overline{x}_1 \neq \overline{x}_2$ - Razlika između aritmetičkih sredina analiziranih uzoraka je statistički značajna uz pogrešku p.

T -test

T-test za zavisne uzorke je postupak kojim se utvrđuje statistička značajnost razlike između aritmetičkih sredina jednog uzorka mjerenog u dvije vremenske točke.

Pri tome je moguće postaviti sljedeće hipoteze:

H_0: $\overline{x}_1 = \overline{x}_2$ - Uz pogrešku *p* ne možemo tvrditi da je razlika između aritmetičkih sredina prvog i drugog mjerenja statistički značajna.

H_1: $\overline{x}_1 \neq \overline{x}_2$ - Razlika između aritmetičkih sredina prvog i drugog mjerenja je statistički značajna uz pogrešku *p*.

T -test

Pretpostavimo da iz iste populacije formiramo veliki broj parova slučajnih uzoraka (npr. *10.000*) veličine 5 entiteta te izračunamo razliku aritmetičkih sredina svakog para. Da li su sve razlike jednake nuli? Kako su distribuirane “slučajne” razlike između aritmetičkih sredina?

$\bar{x}_1 - \bar{x}_2 = d_1$

$\bar{x}_2 - \bar{x}_2 = d_2$

. . .

. . .

. . .

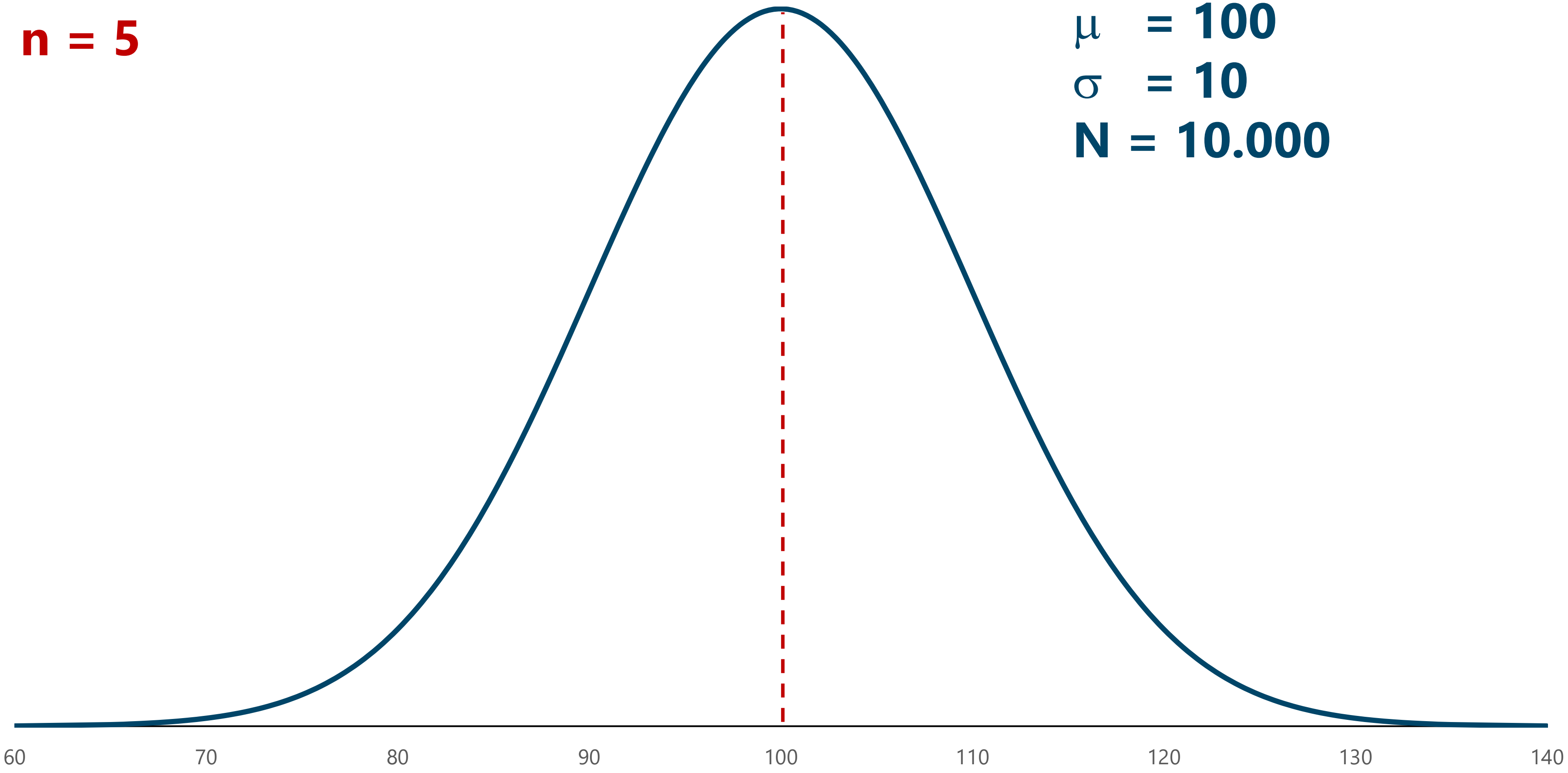
. . .

. . .

. . .

$\bar{x}_1 - \bar{x}_2 = d_{10.000}$

n = 5



μ = 100
σ = 10
N = 10.000

T -test

Pretpostavimo da iz iste populacije formiramo veliki broj parova slučajnih uzoraka (npr. *10.000*) veličine 5 entiteta te izračunamo razliku aritmetičkih sredina svakog para. Da li su sve razlike jednake nuli? Kako su distribuirane “slučajne” razlike između aritmetičkih sredina?

$\bar{x}_1 - \bar{x}_2 = d_1$

$\bar{x}_1 - \bar{x}_2 = d_2$

. . .

. . .

. . .

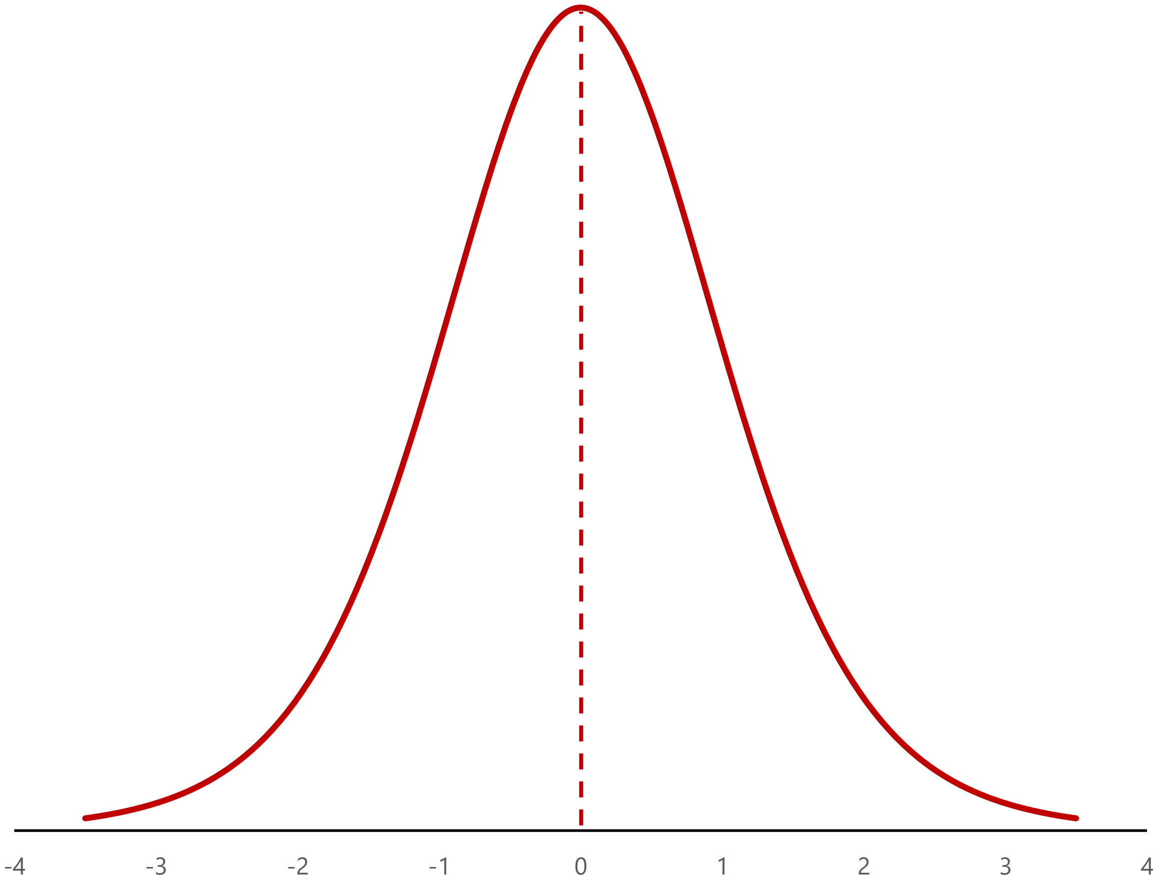
. . .

. . .

. . .

$\bar{x}_1 - \bar{x}_2 = d_{10.000}$

n = 5



T -test

U slučaju da između aritmetičkih sredina populacija ne postoji razlika na temelju prethodno izračunatih razlika između aritmetičkih sredina slučajnih uzoraka možemo zaključiti sljedeće:

- razlika između aritmetičkih sredina slučajnih uzoraka nije uvijek jednaka nuli
- razlike između aritmetičkih sredina slučajnih uzoraka variraju
- aritmetička sredina “slučajnih” razlika, tj. razlika aritmetičkih sredina uzoraka tendirat će nuli
- distribucija razlika između aritmetičkih sredina slučajno odabranih uzoraka iste veličine bit će *normalna*.

T -test

Pretpostavimo da iz iste populacije formiramo veliki broj parova slučajnih uzoraka (npr. *10.000*) veličine 10 entiteta te izračunamo razliku aritmetičkih sredina svakog para. Da li razlike aritmetičkih sredina uzoraka veličine 10 entiteta variraju više ili manje nego razlike aritmetičkih sredina uzoraka veličine 5 entiteta?

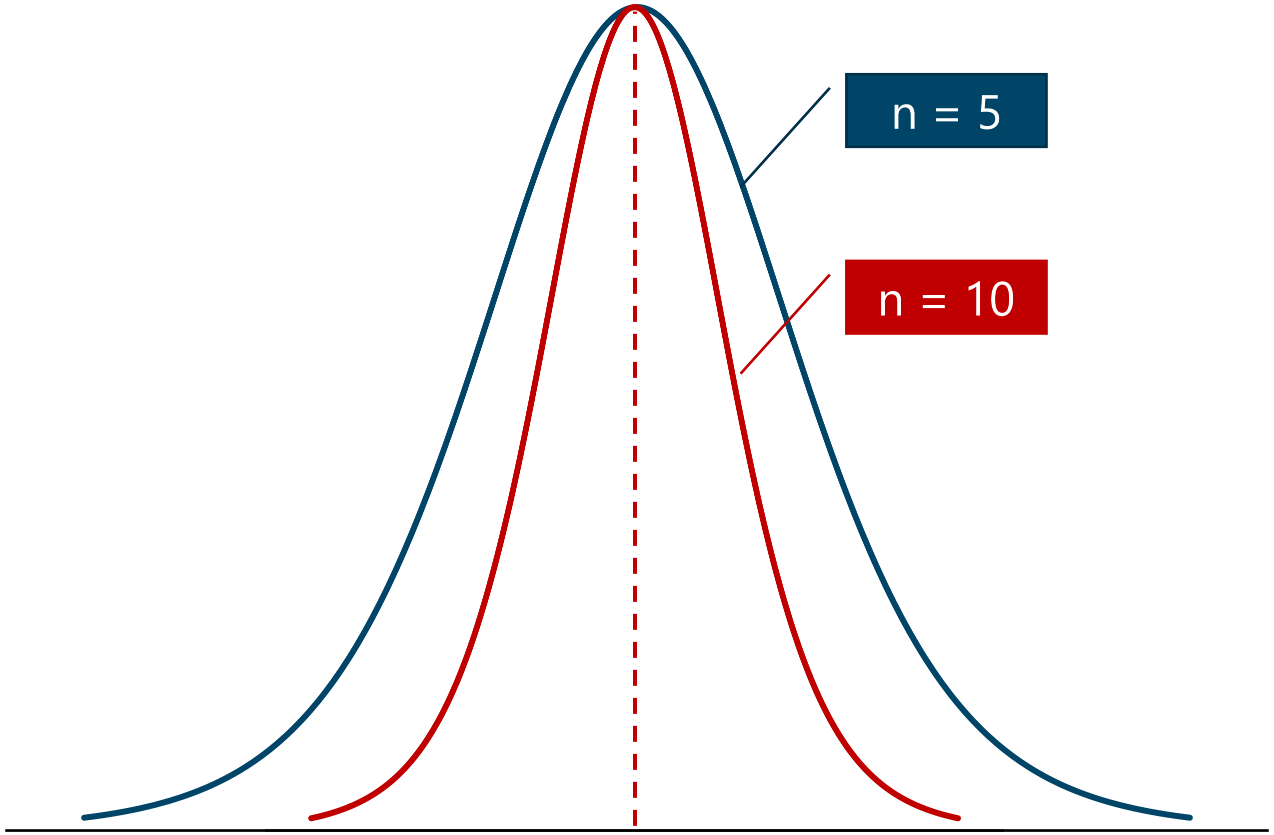
$$\bar{x}_1 - \bar{x}_2 = d_1$$

$$\bar{x}_1 - \bar{x}_2 = d_2$$

$$\vdots \quad \vdots \quad \vdots$$

$$\bar{x}_1 - \bar{x}_2 = d_{10.000}$$

n = 10



T -test

Usporedbom varijabli razlika između aritmetičkih sredina parova slučajnih uzoraka veličine 5 entiteta i parova slučajnih uzoraka veličine 10 entiteta moguće je zaključiti da je standardna devijacija "slučajnih" razlika manja što je broj entiteta u uzorku veći.

Osim o broju entiteta u uzorku, standardna devijacija "slučajnih" razlika zavisi i o varijabilnosti istraživanog obilježja (varijable) u populaciji. Što je varijabilnost obilježja u populaciji veća to će i "slučajne" razlike više varirati.

Standardna devijacija "slučajnih" razlika naziva se standardna pogreška razlika aritmetičkih sredina, a označava se simbolom $s_{\bar{x}_1-\bar{x}_2}$

T -test

Standardna pogreška razlika aritmetičkih sredina u **t-testu za nezavisne uzorke** izračunava se formulom

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

gdje je

- s_1 - standardna devijacija prvog uzorka
- s_2 - standardna devijacija drugog uzorka
- n_1 - broj entiteta u prvom uzorku
- n_2 - broj entiteta u drugom uzorku.

T -test

Standardna pogreška razlika aritmetičkih sredina u t-testu za zavisne uzorke izračunava se formulom

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n} - 2 \cdot r \cdot \frac{s_1}{\sqrt{n}} \cdot \frac{s_2}{\sqrt{n}}}$$

gdje je

- s_1 - standardna devijacija prvog mjerenja
- s_2 - standardna devijacija drugog mjerenja
- n - broj entiteta u uzorku
- r - korelacija između varijabli prvog i drugog mjerenja.

T -test

Ako je razlika između dviju aritmetičkih sredina ${}_{df}t_{p}$ puta veća od standardne pogreške razlika aritmetičkih sredina onda je vjerojatnost da razlika ne postoji u populaciji manja od p . Omjer razlike između dviju aritmetičkih sredina i standardne pogreške razlika naziva se t-omjer ili t-vrijednost, označava se s t, a testira se putem t-distribucije.

T-omjer, tj. t-vrijednost (t) izračunava se jednako u t-testu za nezavisne uzorke i u t-testu za zavisne uzorke i to pomoću sljedeće formule

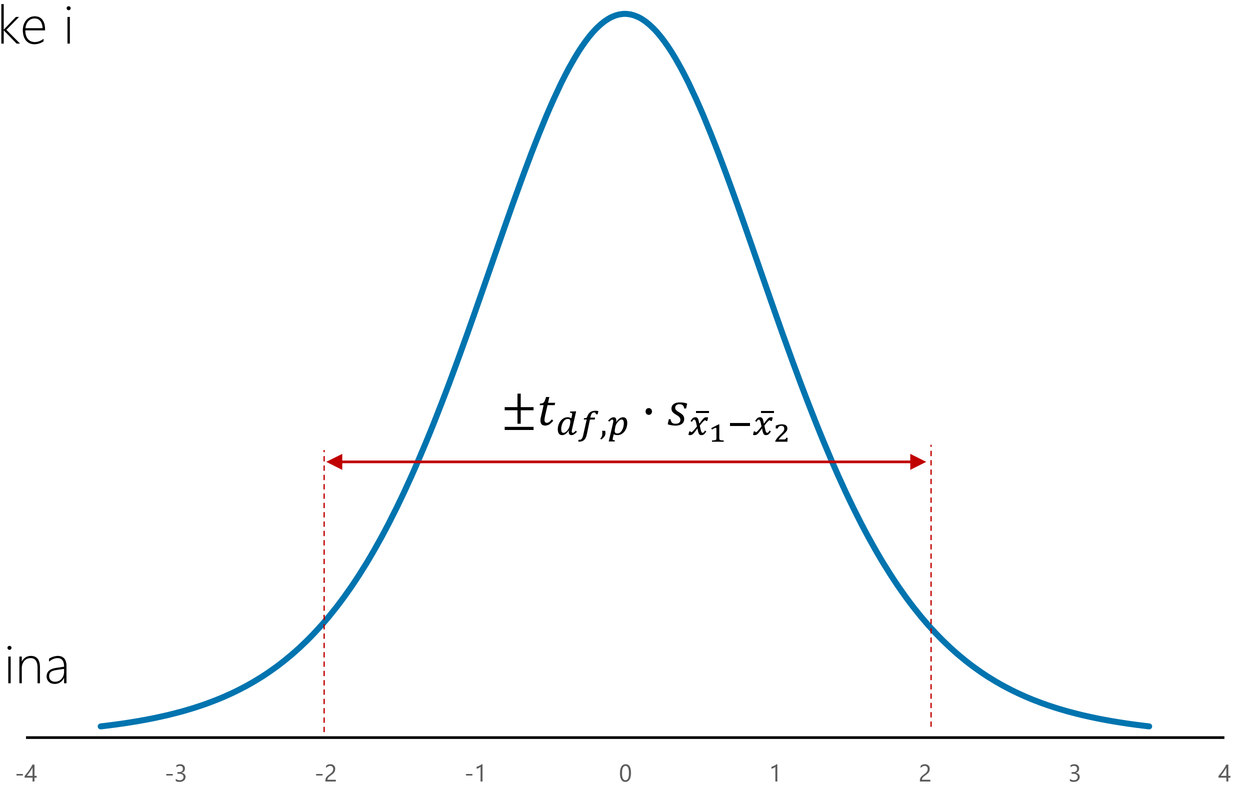
$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

gdje je

$\bar{x}_1$ - aritmetička sredina prvog uzorka/mjerenja

$\bar{x}_1$ - aritmetička sredina drugog uzorka/mjerenja

$s_{\bar{x}_1 - \bar{x}_2}$- standardna pogreška razlika aritmetičkih sredina



T -test

U **t-testu za nezavisne uzorke** kritična t-vrijednost (${}_{df}t_p$) određuje se na temelju broja stupnjeva slobode $df=n_1-1+n_2-1$ i pogreške statističkog zaključka p (npr. $p<0{,}05$ ili $p<0{,}01$).

Testiranje hipoteza vrši se na sljedeći način:

$|t| < {}_{df}t_p \Rightarrow \mathrm{H}_0\text{:}\ \bar{x}_1 = \bar{x}_2$

(H0: Uz pogrešku p ne možemo tvrditi da je razlika između aritmetičkih sredina analiziranih uzoraka statistički značajna)

$|t| > {}_{df}t_p \Rightarrow \mathrm{H}_1\text{:}\ \bar{x}_1 \neq \bar{x}_2$

(H1: Razlika između aritmetičkih sredina analiziranih uzoraka statistički je značajna uz pogrešku p)

T -test

U **t-testu za zavisne uzorke** kritična t-vrijednost ($_{df}t_p$) određuje se na temelju broja stupnjeva slobode $df=n-1$ i pogreške statističkog zaključka p (npr. $p<0{,}05$ ili $p<0{,}01$).

Testiranje hipoteza vrši se na sljedeći način:

$|t| < {}_{df}t_p \Rightarrow \mathrm{H}_0\!: \bar{x}_1 = \bar{x}_2$

(H0: *Uz pogrešku p ne možemo tvrditi da je razlika između aritmetičkih sredina prvog i drugog mjerenja statistički značajna*)

$|t| > {}_{df}t_p \Rightarrow \mathrm{H}_1\!: \bar{x}_1 \neq \bar{x}_2$

(H1: *Razlika između aritmetičkih sredina prvog i drugog mjerenja statistički je značajna uz pogrešku p*)

T -test

Primjer: Učenici 1.a i 1.b razreda testirani su na kraju školske godine testom za procjenu fleksibilnosti *Pretklon raskoračno* i dobiveni su sljedeći rezultati:

$n_{1a} = 25 \qquad n_{1b} = 27$

$\bar{x}_{1a} = 34 \qquad \bar{x}_{1b} = 38$

$s_{1a} = 10 \qquad s_{1b} = 18$

Utvrdi da li se aritmetičke sredine 1.a i 1.b razreda statistički značajno razlikuju uz pogrešku $p<0{,}01$!

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{10^2}{25} + \frac{18^2}{27}} = 4$$

$$|t| = \left|\frac{\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2}{\mathrm{s}_{\bar{\mathrm{x}}_1-\bar{\mathrm{x}}_2}}\right| = \left|\frac{34-38}{4}\right| = |-1| = 1$$

$$df = n_1 - 1 + n_2 - 1 = 25 - 1 + 27 - 1 = 50$$

$${}_{df}t_p = {}_{50}t_{0,01} = 2{,}68$$

$|t| < {}_{df}t_p \Rightarrow H_0\text{:}\ \bar{x}_1 = \bar{x}_2$

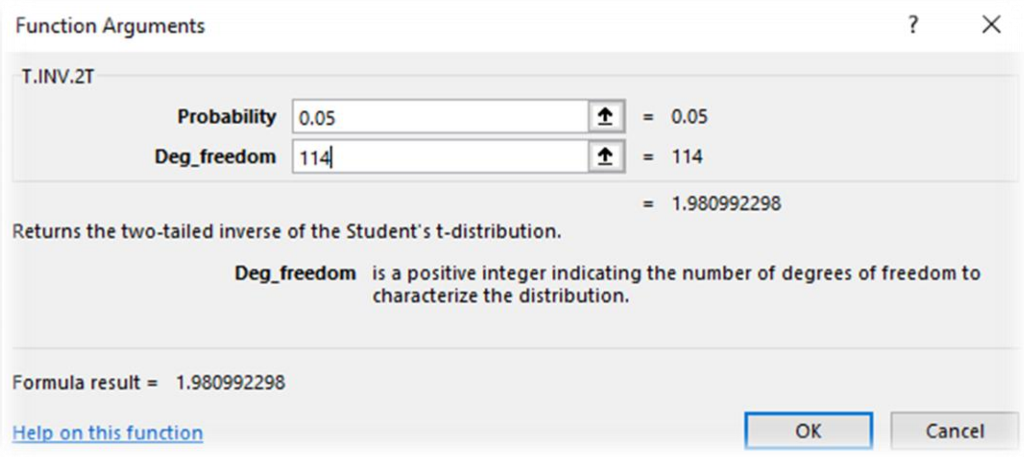
H0: Uz pogrešku 0,01 ne možemo tvrditi da je razlika između aritmetičkih sredina 1.a i 1.b razreda statistički značajna.

Zadatak 1: Uz pogrešku $p = 0{,}05$ testirajte da li se aritmetičke sredine ovih uzoraka statistički značajno razlikuju.

$n_1 = 55,\ \bar{x}_1 = 220,\ s_1 = 15$

$n_2 = 61,\ \bar{x}_2 = 195,\ s_2 = 10.$

1 **UTVRĐIVANJE KRITIČNE T-VRIJEDNOSTI** vrši se pomoću funkcije ***=T.INV.2T()***. Funkcija se unosi u označeno polje matrice odabirom opcije *Function...* padajućeg izbornika *Insert*. U traku *Probability* dijaloškog okvira za unos ove funkcije potrebno je upisati pogrešku *p*, a u traku *Deg_freedom* broj stupnjeva slobode.



T.INV.2T

Probability 0.05 = 0.05

Deg_freedom 114 = 114

= 1.980992298

Returns the two-tailed inverse of the Student's t-distribution.

Deg_freedom is a positive integer indicating the number of degrees of freedom to characterize the distribution.

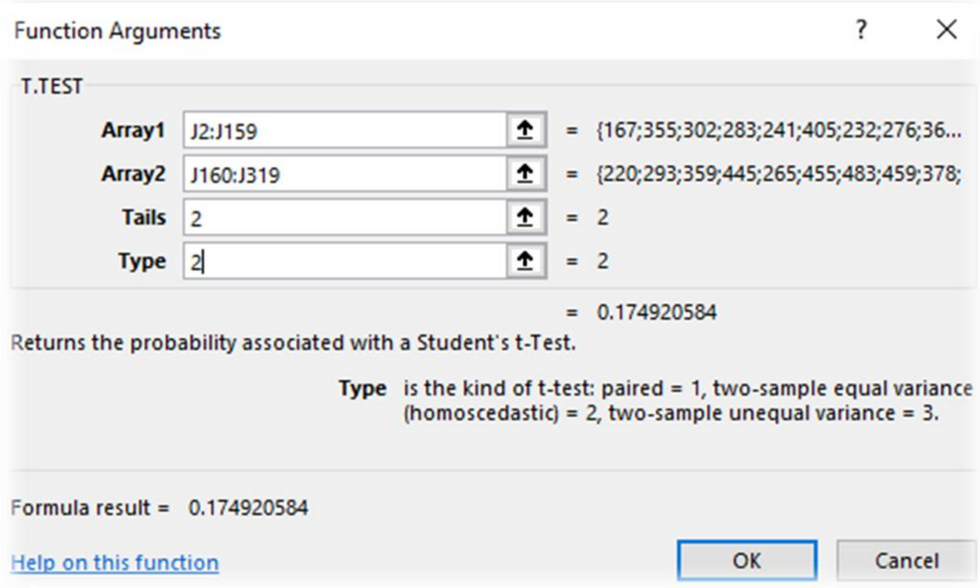
Formula result = 1.980992298

Microsoft Excel

Zadatak 2: U datoteci *UCENICI.xls* uz pogrešku *p=0,01* testirajte da li se aritmetičke sredine učenika Gorskog (Gore) i Primorskog (More) kraja u varijabli *MPOL* statistički značajno razlikuju!

2

T-TEST: Utvrđivanje pogreške s kojom je moguće zaključiti da se dvije aritmetičke sredine statistički značajno razlikuju vrši se pomoću funkcije ***=T.TEST***. Putem traka *Array1* i *Array2* potrebno je definirati niz podataka prvog odnosno drugog uzorka/mjerenja, u traku *Tails* unijeti vrijednost *2*, a u traku *Type* vrijednost *1* ako se provodi t-test za zavisne uzorke, odnosno *2* ako se provodi t-test za nezavisne uzorke.



Function Arguments

T.TEST

Array1: J2:J159 = {167;355;302;283;241;405;232;276;36...

Array2: J160:J319 = {220;293;359;445;265;455;483;459;378;

Tails: 2 = 2

Type: 2 = 2

= 0.174920584

Returns the probability associated with a Student's t-Test.

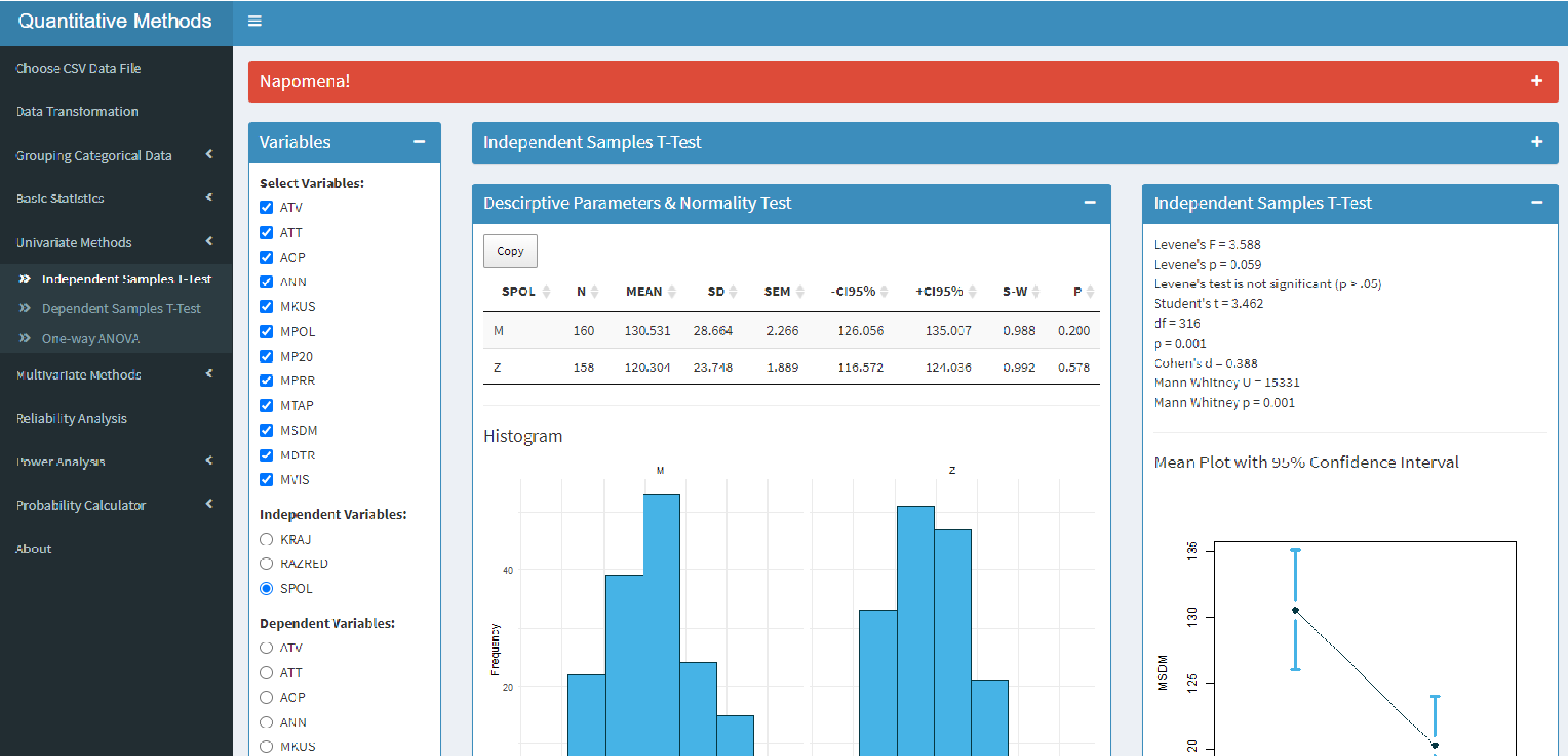
Type is the kind of t-test: paired = 1, two-sample equal variance (homoscedastic) = 2, two-sample unequal variance = 3.

Formula result = 0.174920584

RStudio – Independent Samples T-Test

Zadatak 3: U datoteci *UCENICI.csv* uz pogrešku *p=0,01* testirajte da li se aritmetičke sredine učenika (M) i učenica (Z) u varijabli *MSDM* statistički značajno razlikuju!

RStudio – Independent Samples T-Test



Variables

Select Variables:

☑ ATV
☑ ATT
☑ AOP
☑ ANN
☑ MKUS
☑ MPOL
☑ MP20
☑ MPRR
☑ MTAP
☑ MSDM
☑ MDTR
☑ MVIS

Independent Variables:

○ KRAJ
○ RAZRED
● SPOL

Dependent Variables:

○ ATV
○ ATT
○ AOP
○ ANN
○ MKUS

Independent Samples T-Test

Descirptive Parameters & Normality Test

SPOL	N	MEAN	SD	SEM	-CI95%	+CI95%	S-W	P
M	160	130.531	28.664	2.266	126.056	135.007	0.988	0.200
Z	158	120.304	23.748	1.889	116.572	124.036	0.992	0.578

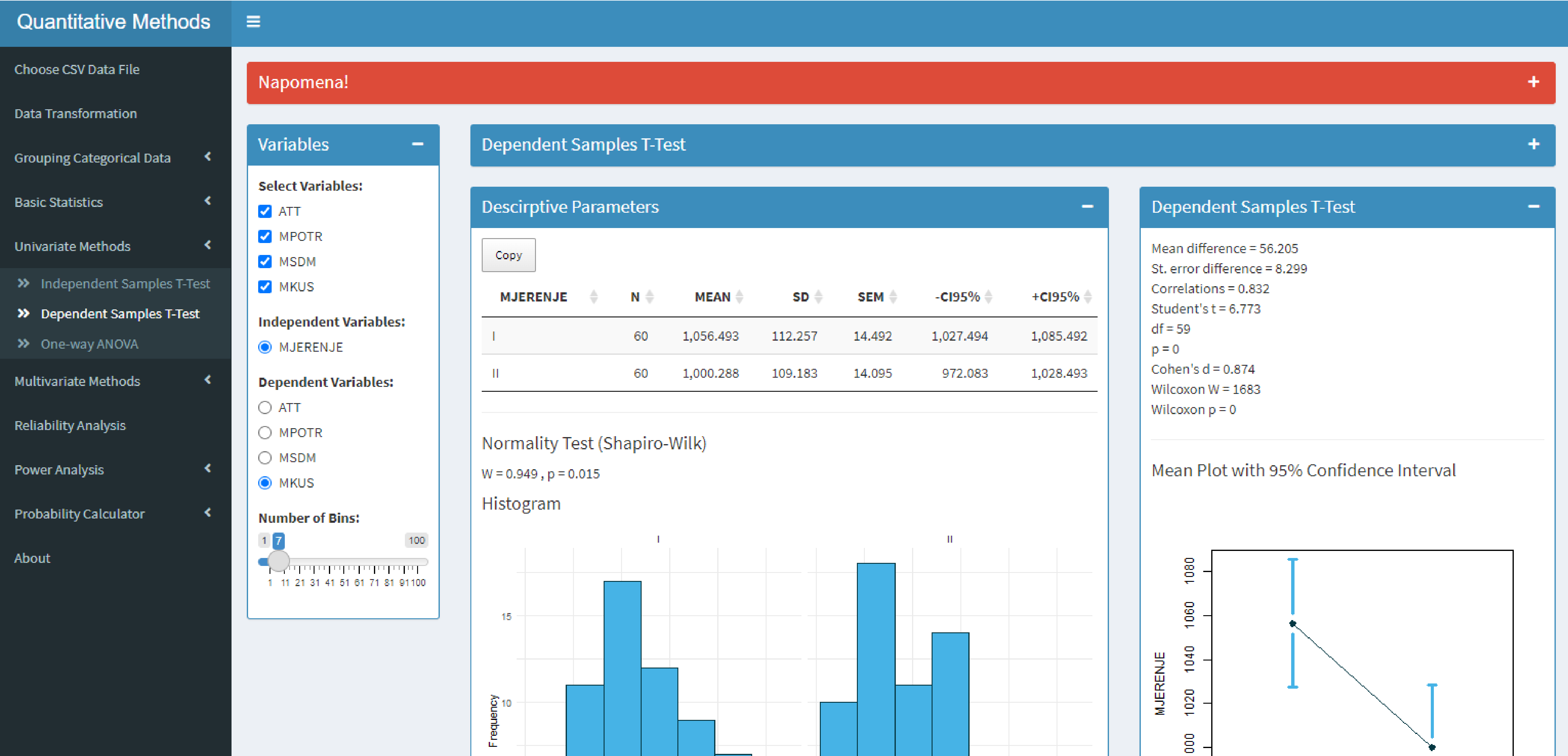
Independent Samples T-Test

Levene's F = 3.588
Levene's p = 0.059
Levene's test is not significant (p > .05)
Student's t = 3.462
df = 316
p = 0.001
Cohen's d = 0.388
Mann Whitney U = 15331
Mann Whitney p = 0.001

RStudio – Dependent (Paired) Samples T-Test

Zadatak 4: U datoteci *POD.csv* uz pogrešku *p=0,01* testirajte da li se aritmetičke sredine prvog mjerenja i drugog mjerenja u varijabli *MKUS* statistički značajno razlikuju!

RStudio – Dependent (Paired) Samples T-Test



Select Variables:

- ☑ ATT
- ☑ MPOTR
- ☑ MSDM
- ☑ MKUS

Independent Variables:

- ● MJERENJE

Dependent Variables:

- ○ ATT
- ○ MPOTR
- ○ MSDM
- ● MKUS

Number of Bins: 7 (1–100)

MJERENJE	N	MEAN	SD	SEM	-CI95%	+CI95%
I	60	1,056.493	112.257	14.492	1,027.494	1,085.492
II	60	1,000.288	109.183	14.095	972.083	1,028.493

Normality Test (Shapiro-Wilk)

W = 0.949 , p = 0.015

Histogram

Mean difference = 56.205
St. error difference = 8.299
Correlations = 0.832
Student's t = 6.773
df = 59
p = 0
Cohen's d = 0.874
Wilcoxon W = 1683
Wilcoxon p = 0

KORELACIJA

Vježba 7

Korelacija

Začetnikom korelacijske i regresijske analize smatra se engleski antropolog Francis Galton koji je pod utjecajem rođaka Charlesa Darwina istraživao utjecaj naslijeđa na razvoj čovjekovih karakteristika.

Surađujući s Galtonom, Karl Perason je razvio računski postupak za utvrđivanje povezanosti između dviju varijabli i nazvao ga ***produkt -moment koeficijent korelacije***.

Pearsonov ***produkt - moment koeficijent korelacije*** (*r*) predstavlja mjeru međusobne linearne povezanosti rezultata dviju standardiziranih varijabli, a izračunava se formulom

$$r = \frac{\sum_{i=1}^{n} x_{ci} y_{ci}}{\sqrt{\sum_{i=1}^{n} x_{ci}^2} \cdot \sqrt{\sum_{i=1}^{n} y_{ci}^2}}$$

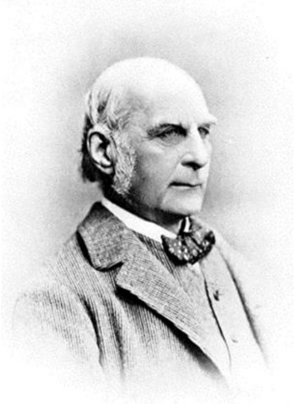
gdje je

$x_{ci} = x_i - \bar{x}$ (centrirani rezultat entiteta *i* u varijabli *x*)

$\bar{x}$ - aritmetička sredina varijable *x*

$y_{ci} = y_i - \bar{y}$ (centrirani rezultat entiteta *i* u varijabli *y*)

$\bar{y}$ - aritmetička sredina varijable *y*



Francis Galton
(1822. – 1911.)

Karl Pearson
(1857. – 1936.)

Korelacija

Pearsonov koeficijent korelacije može se izračunati i iz originalnih (necentriranih) rezultata pomoću formule

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right)\left(\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right)}}$$

gdje je

x_i - rezultat entiteta *i* u varijabli *x*

y_i - rezultat entiteta *i* u varijabli *y*

n - broj entiteta

Ako se rezultati ispitanika u varijablama *x* i *y* standardiziraju, onda formula poprima sljedeći oblik

$$r = \frac{\sum_{i=1}^{n} (z_{x_i} z_{y_i})}{n}$$

gdje je

z_{x_i} - standardizirani rezultat entiteta *i* u varijabli *x*

z_{y_i}- standardizirani rezultat entiteta *i* u varijabli *y*

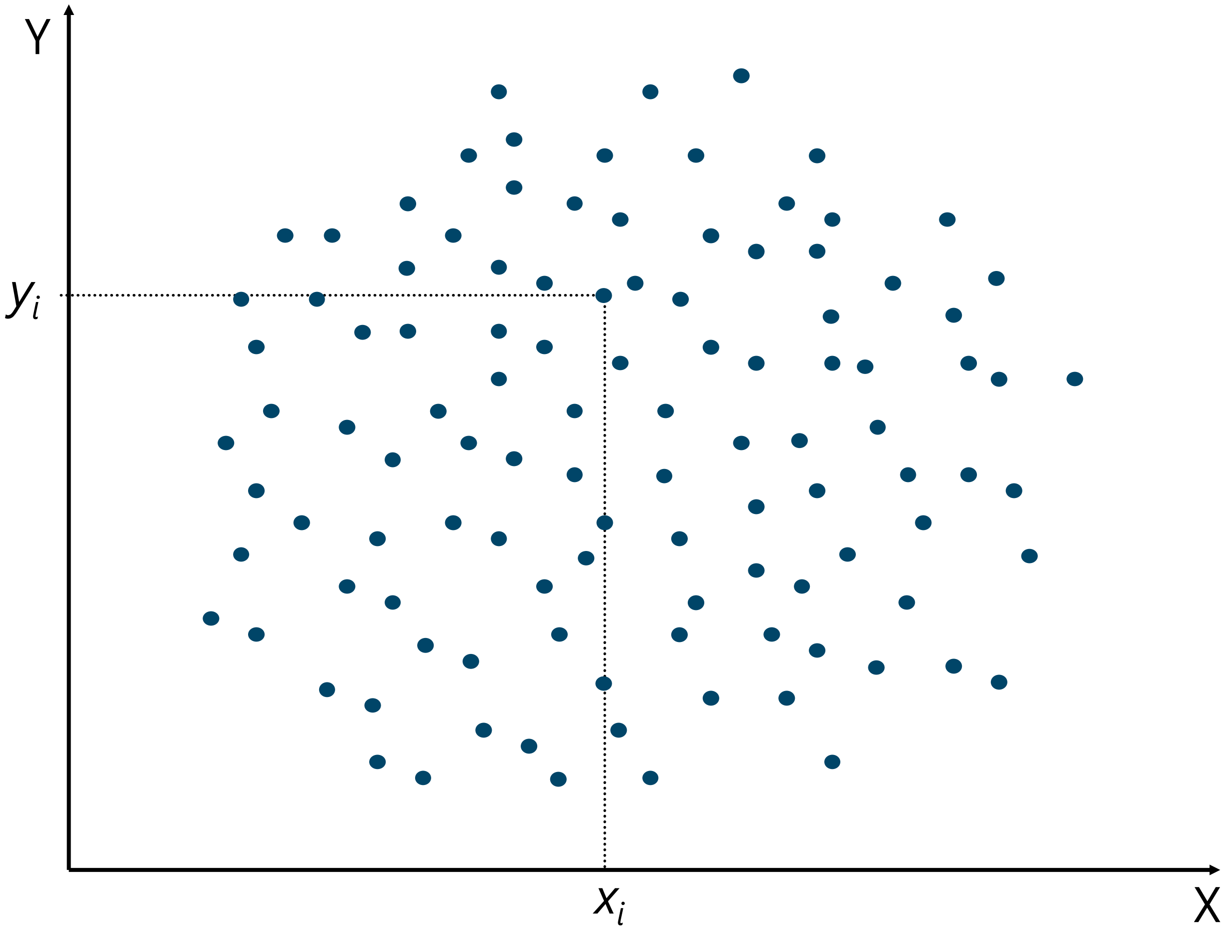
n - broj entiteta

Pearsonov koeficijent korelacije (izračunat bilo kojim postupkom) uvijek se nalazi u intervalu od *-1* do *1* .

Korelacija

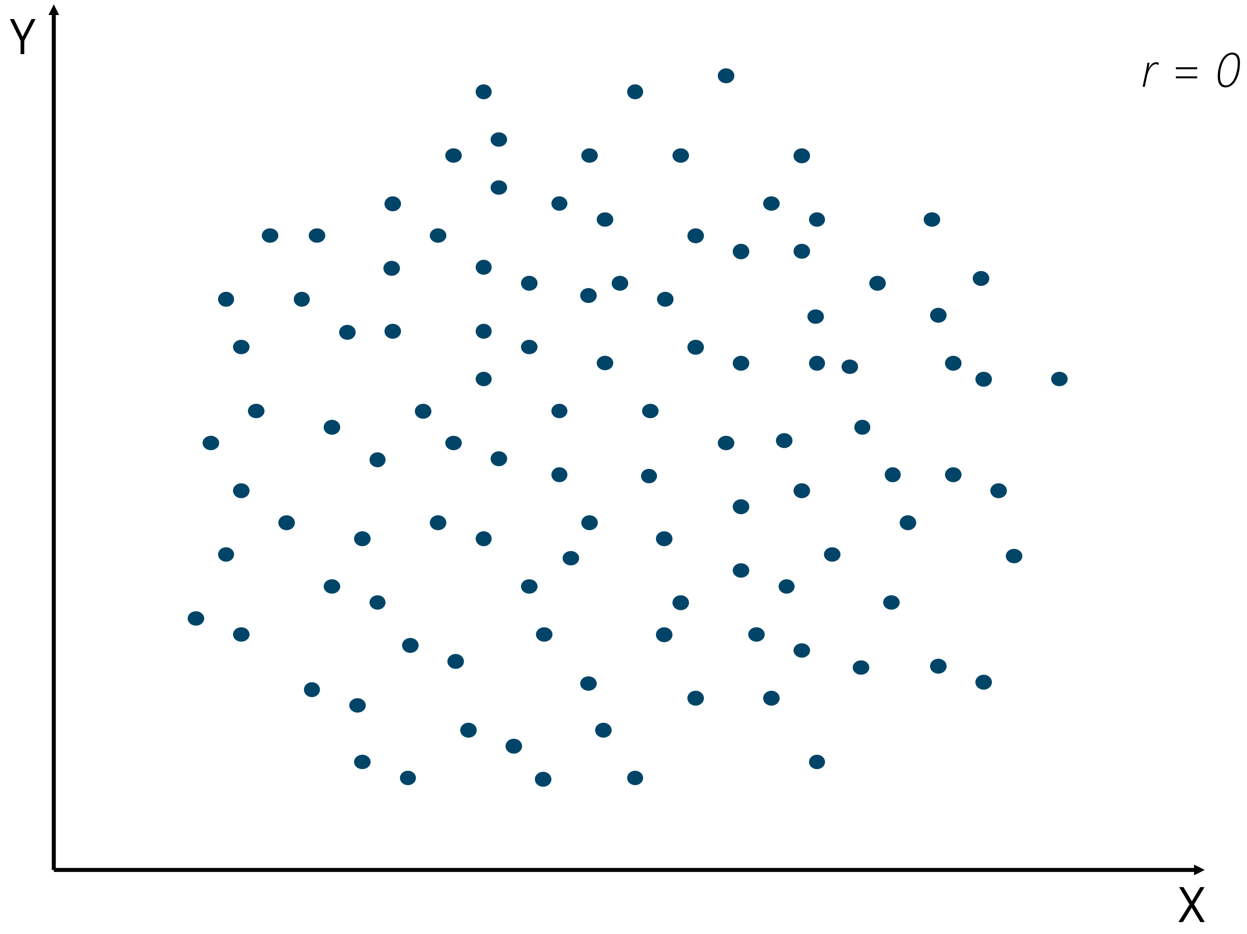
Dvodimenzionalni ***korelacijski dijagram*** je grafički način prikazivanja povezanosti rezultata dviju varijabli, a iscrtava se na način da se za svakog ispitanika odredi položaj u koordinatnom sustavu pri čemu se položaj na apscisi određuje sukladno rezultatu ispitanika u jednoj varijabli, a položaj na ordinati sukladno rezultatu u drugoj varijabli.

Oblik korelacijskog dijagrama zavisan je o smjeru i veličini povezanosti varijabli, odnosno o predznaku i veličini koeficijenta korelacije.



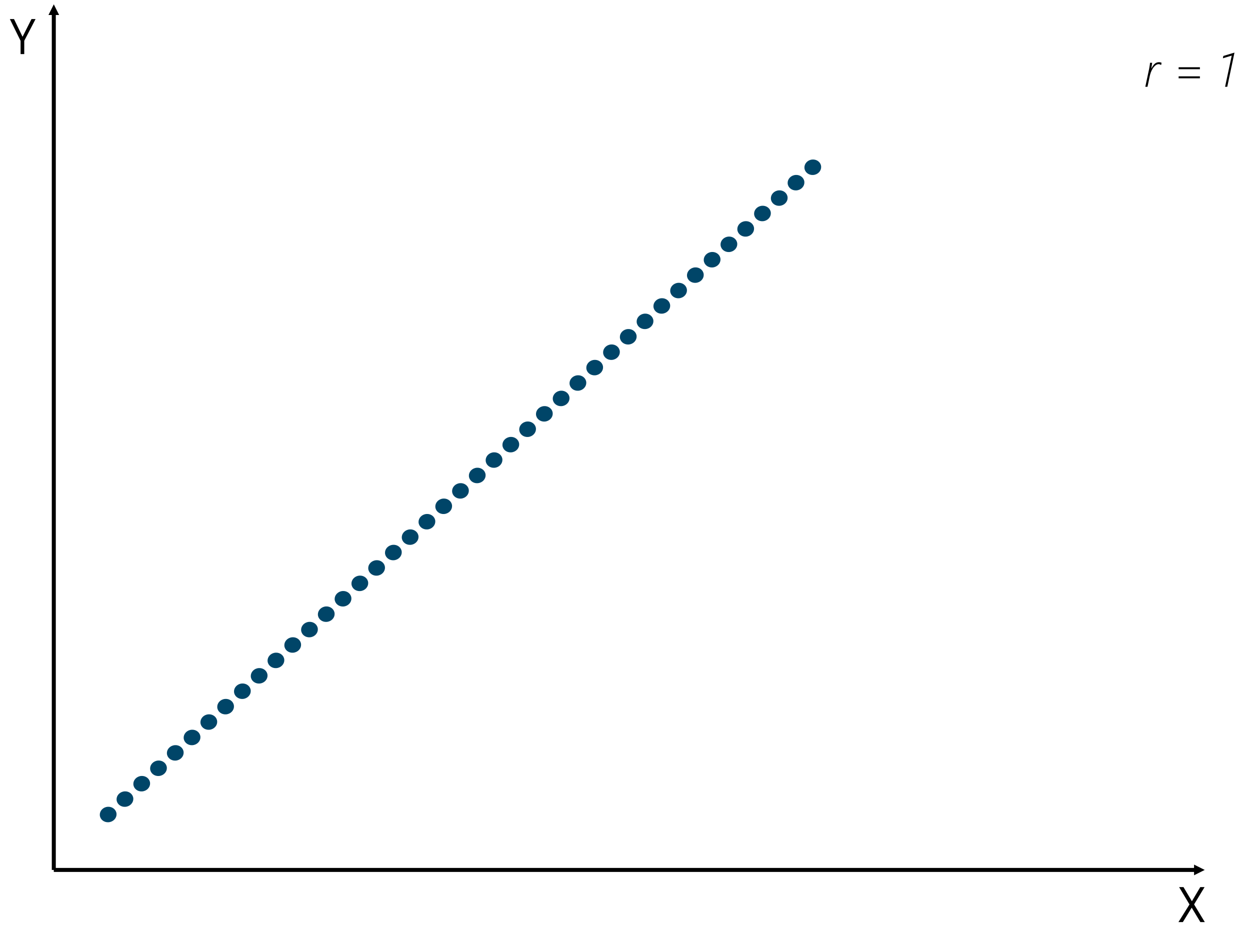
Korelacija

Nulta korelacija (*r*=*0*) označava takav odnos između dviju varijabli u kojem svakom rezultatu u jednoj varijabli može odgovarati bilo koji rezultat u drugoj varijabli.



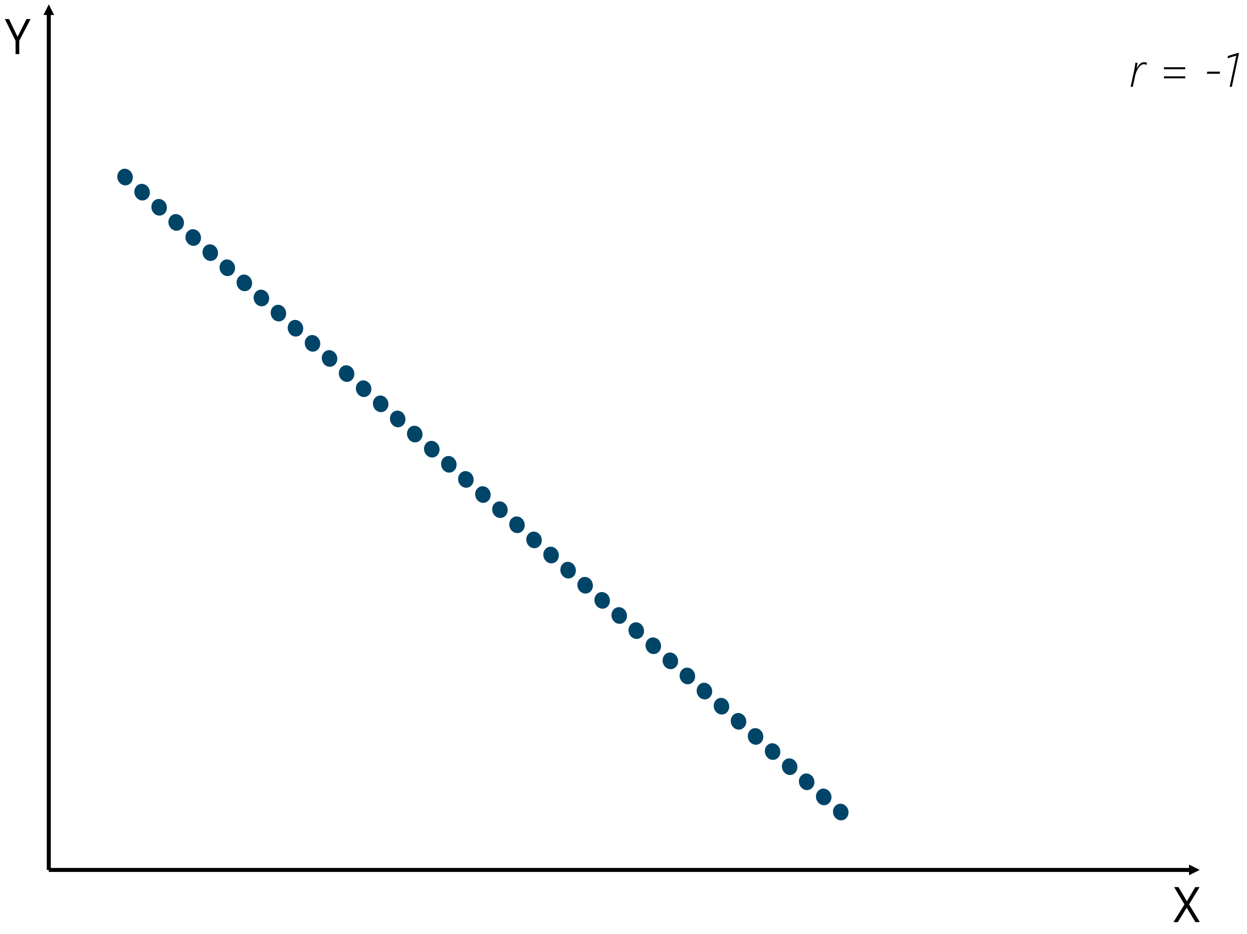
Korelacija

Potpuna pozitivna korelacija (*r*=*1*) označava takav odnos između dviju varijabli u kojem svakom iznadprosječnom rezultatu u jednoj varijabli odgovara jednako iznadprosječan rezultat u drugoj varijabli, odnosno svakom ispodprosječnom rezultatu u jednoj varijabli odgovara jednako ispodprosječan rezultat u drugoj varijabli.



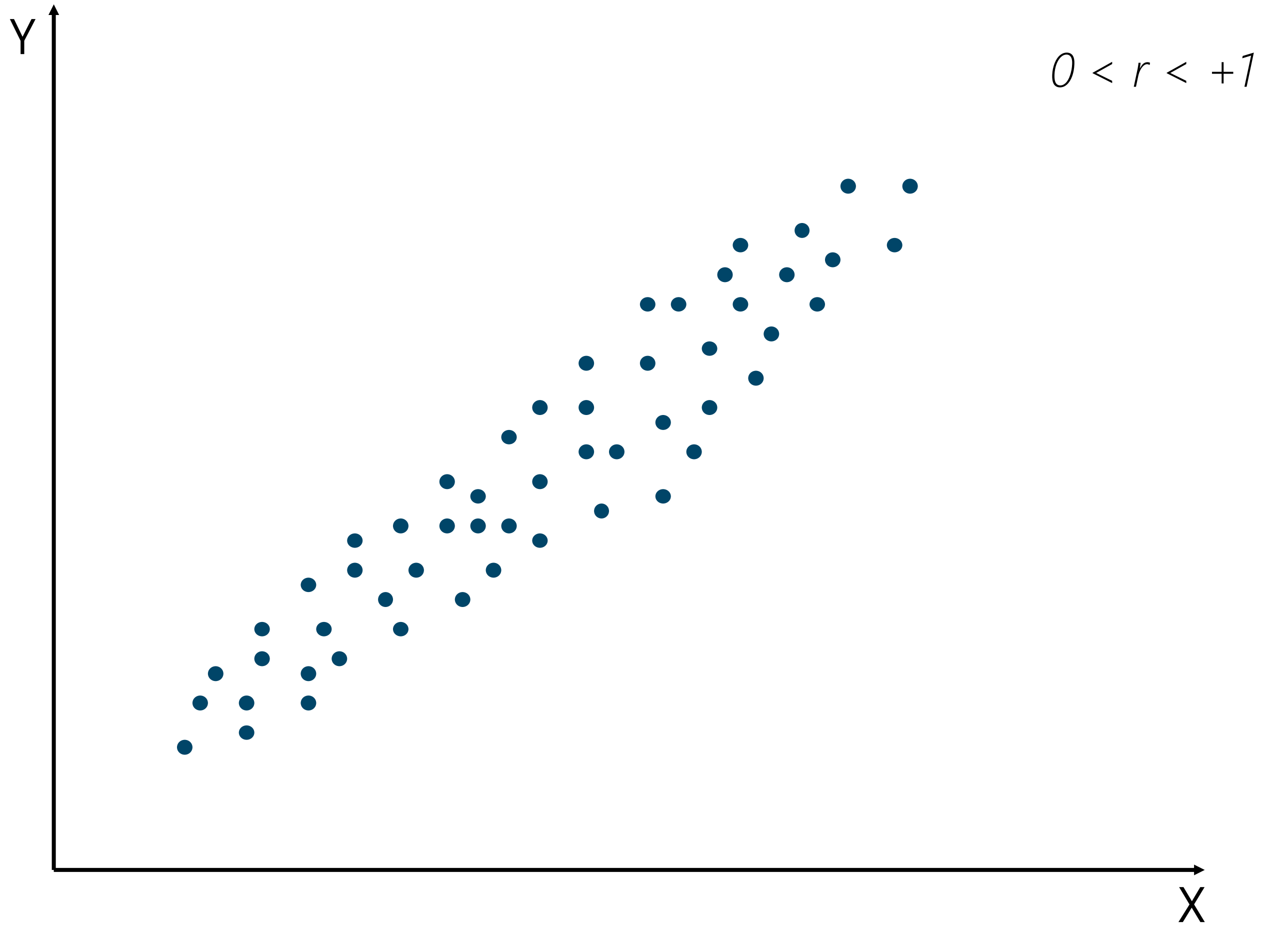
Korelacija

Potpuna negativna korelacija (*r*=-1) označava takav odnos između dviju varijabli u kojem svakom iznadprosječnom rezultatu u jednoj varijabli odgovara jednako ispodprosječan rezultat u drugoj varijabli, odnosno svakom ispodprosječnom rezultatu u jednoj varijabli odgovara jednako iznadprosječan rezultat u drugoj varijabli.



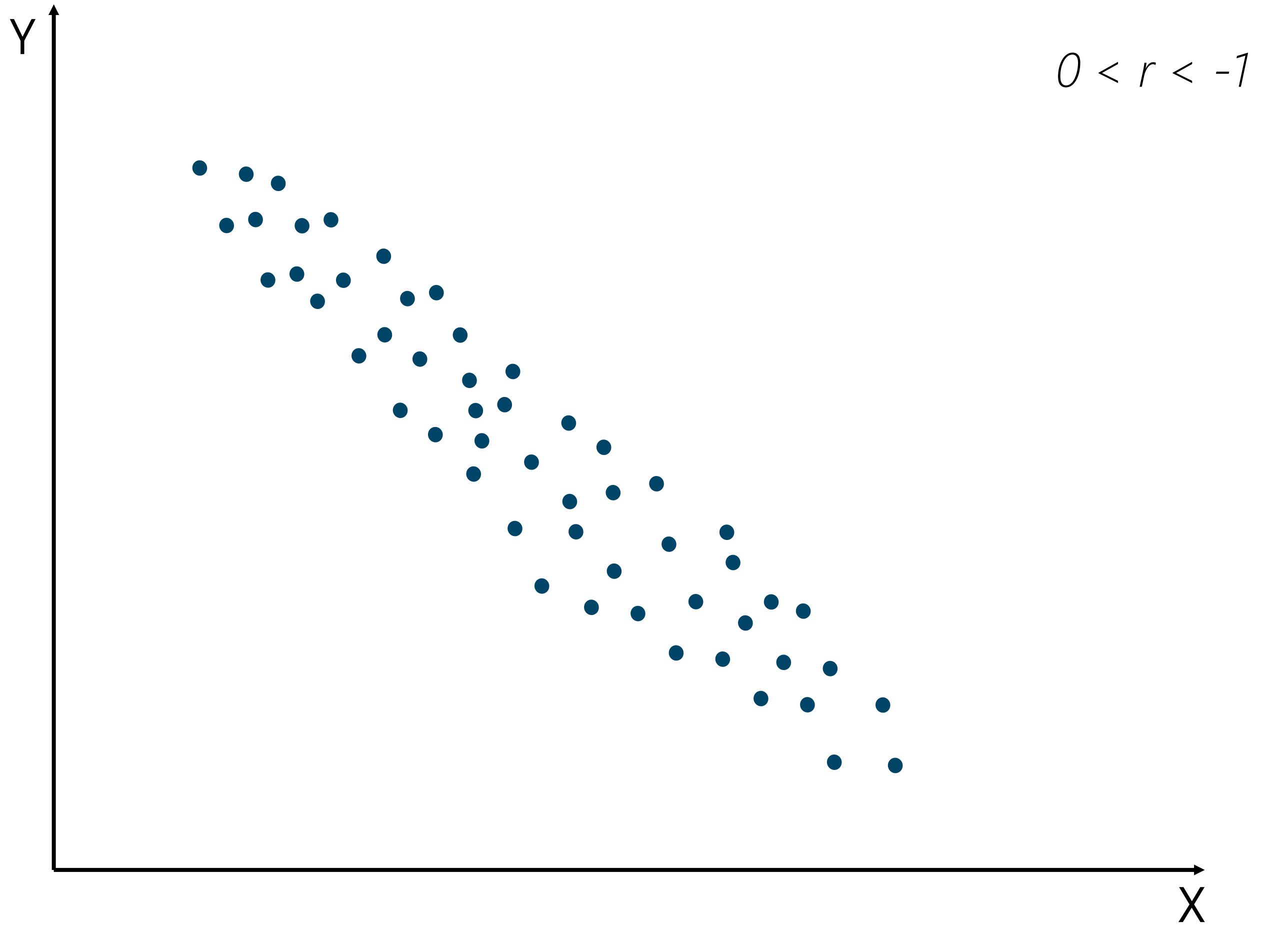
Korelacija

Nepotpuna pozitivna korelacija (0<*r*<1) označava takav odnos između dviju varijabli u kojem svakom iznadprosječnom rezultatu u jednoj varijabli najvjerojatnije odgovara jednako iznadprosječan rezultat u drugoj varijabli, odnosno svakom ispodprosječnom rezultatu u jednoj varijabli najvjerojatnije odgovara jednako ispodprosječan rezultat u drugoj varijabli.



Korelacija

Nepotpuna negativna korelacija (0>*r*>-1) označava takav odnos između dviju varijabli u kojem svakom iznadprosječnom rezultatu u jednoj varijabli najvjerojatnije odgovara jednako ispodprosječan rezultat u drugoj varijabli, odnosno svakom ispodprosječnom rezultatu u jednoj varijabli najvjerojatnije odgovara jednako iznadprosječan rezultat u drugoj varijabli.



Korelacija

Ako se povezanost između dviju varijabli utvrđuje na uzorku ispitanika, potrebno je testirati statističku značajnost koeficijenta korelacije odnosno utvrditi vjerojatnost da se korelacija nije dogodila slučajno. Pri testiranju statističke značajnosti koeficijenta korelacije moguće je postaviti sljedeće hipoteze

$H_0 : r=0$ - korelacija nije statistički značajna uz pogrešku p

$H_1 : r \neq 0$ - korelacija je statistički značajna uz pogrešku p

Statistička značajnost koeficijenta korelacije testira se putem t-distribucije pri čemu se kritična t vrijednost određuje na temelju pogreške statističkog zaključka p i broja stupnjeva slobode $df=n-2$. Vrijednost koja se uspoređuje s kritičnom t-vrijednosti izračunava se formulom

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

gdje je
t - vrijednost koja se distribuira prema t-distribuciji za $df=n-2$
r - koeficijent korelacije
n - broj entiteta

Iz prethodno navedene formule moguće je uočiti kako je vjerojatnost da se korelacija u uzorku dogodila slučajno iako u populaciji ne postoji manja što je broj entiteta u uzorku veći i što je apsolutna vrijednost izračunate korelacije veća.

Korelacija

Rezultati korelacijske analize najčešće se prikazuju pomoću ***korelacijske matrice***. U dijagonali korelacijske matrice su varijance varijabli, a izvandijagonalni elementi su korelacije svake varijable sa svakom. Korelacijska matrica je simetrična što znači da se koeficijent korelacije između svake dvije varijable nalazi i s gornje i s donje strane glavne dijagonale.

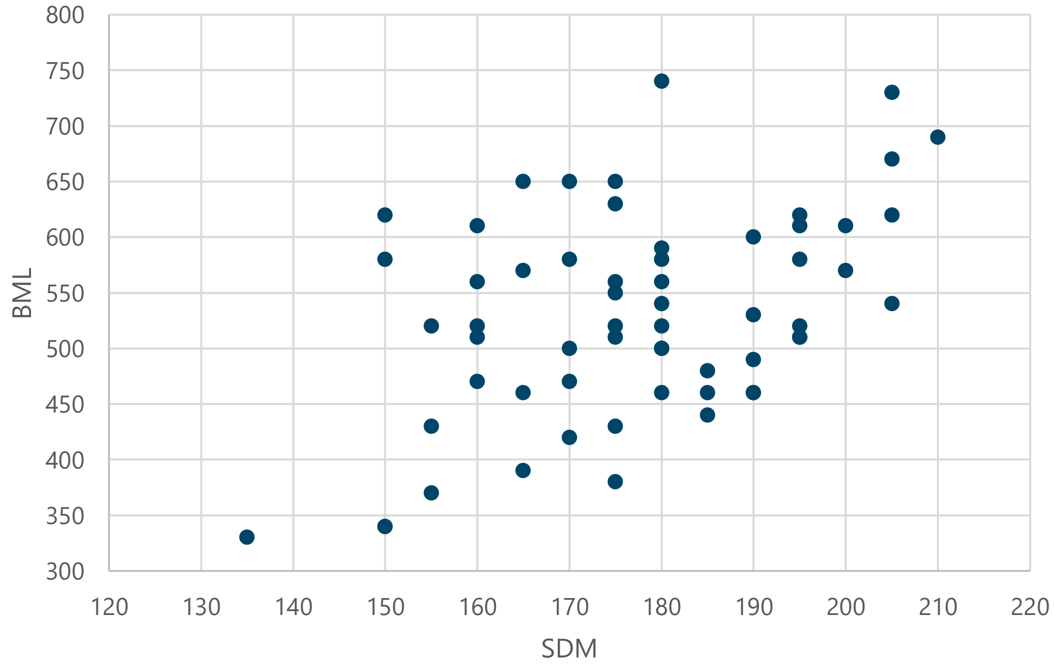
Formalni prikaz korelacijske matrice s četiri varijable

	v1	*v2*	*v3*	*v4*
v1	*1*	$r_{v1,v2}$	$r_{v1,v3}$	$r_{v1,v4}$
v2	$r_{v2,v1}$	*1*	$r_{v2,v3}$	$r_{v2,v4}$
v3	$r_{v3,v1}$	$r_{v3,v2}$	*1*	$r_{v3,v4}$
v4	$r_{v4,v1}$	$r_{v4,v2}$	$r_{v4,v3}$	*1*

Primjer: U korelacijskoj matrici prikazani su Pearsonovi koeficijenti korelacije između četiri motorička testa. Označene korelacije su statistički značajne uz pogrešku *p=0,05*)

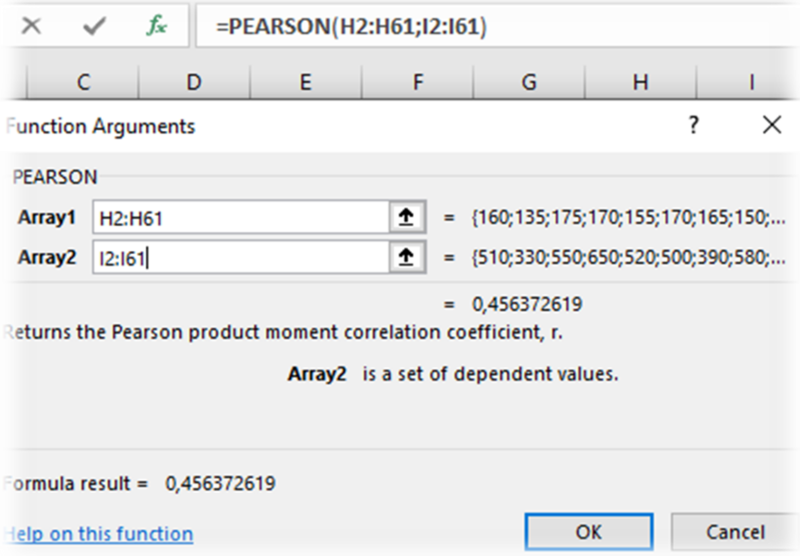
	ONT	OUZ	NEB	SKL
ONT	*1*	*0,66**	*-0,36**	*-0,35**
OUZ	*0,66**	*1*	*-0,21*	*-0,62**
NEB	*-0,36**	*-0,21*	*1*	*0,15*
SKL	*-0,35**	*-0,62**	*0,15*	*1*

Zadatak 1: U datoteci *JUDO3F.xls* izračunajte koliki je Pearsonov koeficijent korelacije između varijabli *SDM* i *BML te* kreirajte korelacijski dijagram (*Scatter plot*).



1 IZRAČUNAVANJE PEARSONOVOG KOEFICIJENTA KORELACIJE

vrši se pomoću funkcije ***PEARSON***. Funkcija se unosi u označeno polje odabirom opcije *Function...* Putem traka *Array1* i *Array2* potrebno je definirati niz podataka prve, odnosno druge varijable.



=PEARSON(H2:H61;I2:I61)

unction Arguments

PEARSON

Array1 H2:H61 = {160;135;175;170;155;170;165;150;...

Array2 I2:I61 = {510;330;550;650;520;500;390;580;...

= 0,456372619

eturns the Pearson product moment correlation coefficient, r.

Array2 is a set of dependent values.

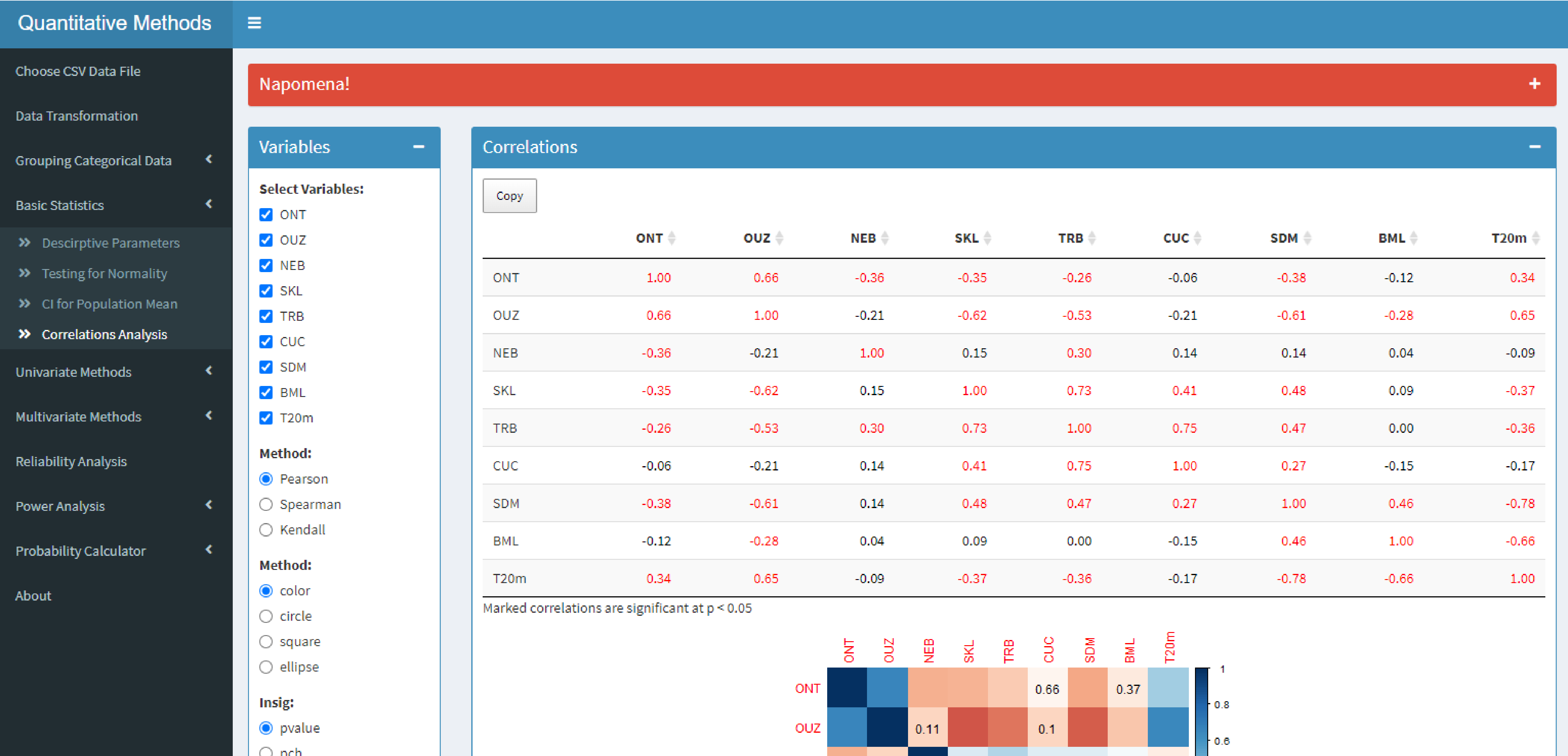
ormula result = 0,456372619

elp on this function

RStudio – Correlation Analysis

Zadatak 2: U datoteci *JUDO3F.csv* izračunajte Pearsonove koeficijente korelacije za sve kvantitativne varijable.

RStudio – Correlation Analysis



Select Variables:
☑ ONT
☑ OUZ
☑ NEB
☑ SKL
☑ TRB
☑ CUC
☑ SDM
☑ BML
☑ T20m

Method:
● Pearson
○ Spearman
○ Kendall

Method:
● color
○ circle
○ square
○ ellipse

Insig:
● pvalue
○ pch

	ONT	OUZ	NEB	SKL	TRB	CUC	SDM	BML	T20m
ONT	1.00	0.66	-0.36	-0.35	-0.26	-0.06	-0.38	-0.12	0.34
OUZ	0.66	1.00	-0.21	-0.62	-0.53	-0.21	-0.61	-0.28	0.65
NEB	-0.36	-0.21	1.00	0.15	0.30	0.14	0.14	0.04	-0.09
SKL	-0.35	-0.62	0.15	1.00	0.73	0.41	0.48	0.09	-0.37
TRB	-0.26	-0.53	0.30	0.73	1.00	0.75	0.47	0.00	-0.36
CUC	-0.06	-0.21	0.14	0.41	0.75	1.00	0.27	-0.15	-0.17
SDM	-0.38	-0.61	0.14	0.48	0.47	0.27	1.00	0.46	-0.78
BML	-0.12	-0.28	0.04	0.09	0.00	-0.15	0.46	1.00	-0.66
T20m	0.34	0.65	-0.09	-0.37	-0.36	-0.17	-0.78	-0.66	1.00

Marked correlations are significant at $p < 0.05$

DESKRIPTIVNA ANALIZA PROMJENA

Vježba 8

Deskriptivna analiza promjena

Deskriptivna analiza promjena je skup postupaka za analizu grupnih ili individualnih promjena putem deskriptivnih statističkih parametara.

Grupne promjene podrazumijevaju razlike u razini jedne ili više karakteristika grupe entiteta u dvije ili više vremenskih točaka.

Individualne promjene podrazumijevaju razlike u razini jedne ili više karakteristika jednog entiteta u dvije ili više vremenskih točaka.

Deskriptivna analiza grupnih promjena

Primjer: Na nekoj grupi polaznika fitnes centra primijenjen je tromjesečni program za povećanje mišićne mase. Korisnicima programa je izmjerena tjelesna masa prije početka programa (x_1) i po završetku provođenja programa (x_2) te je za svakog ispitanika izračunata razlika između inicijalnog i finalnog stanja tjelesne mase (d).

Izračunati su sljedeći deskriptivni parametri: aritmetička sredina ($\bar{x}$), standardna devijacija (s), minimum (*min*), maksimum (*max*), totalni raspon (R), korelacija varijabli inicijalnog i finalnog stanja ($r_{x1,x2}$) i korelacija varijabli inicijalnog stanja i promjene stanja ($r_{x1,d}$).

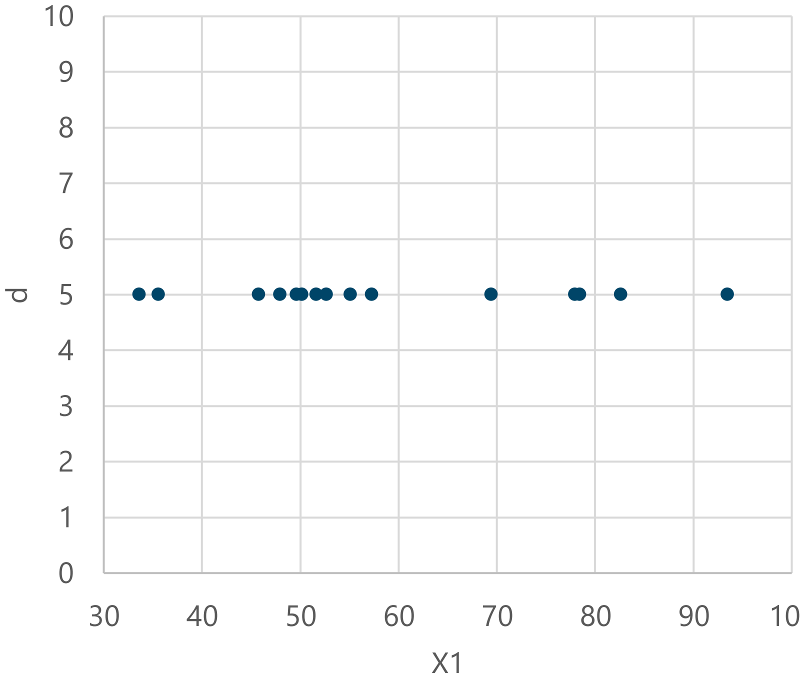
Deskriptivna analiza grupnih promjena

Deskriptivna analiza grupnih promjena - primjer 1:

x_1	x_2	d
82,61	87,61	5,00
93,51	98,51	5,00
78,46	83,46	5,00
55,14	60,14	5,00
49,65	54,65	5,00
45,82	50,82	5,00
50,21	55,21	5,00
51,65	56,65	5,00
69,45	74,45	5,00
57,32	62,32	5,00
35,62	40,62	5,00
47,95	52,95	5,00
33,65	38,65	5,00
52,69	57,69	5,00
77,95	82,95	5,00

	$\bar{x}$	s	min	max	R	$r_{x1,x2}$	$r_{x1,d}$
x1	58,77	17,63	33,65	93,51	59,86		
x2	63,77	17,63	38,65	98,51	59,86	1	0
d	5	0	5	5	0		

Tjelesna masa svakog od sudionika programa povećala se za 5 kilograma. Program je bio primjeren za povećanje tjelesne mase svih ispitanika.



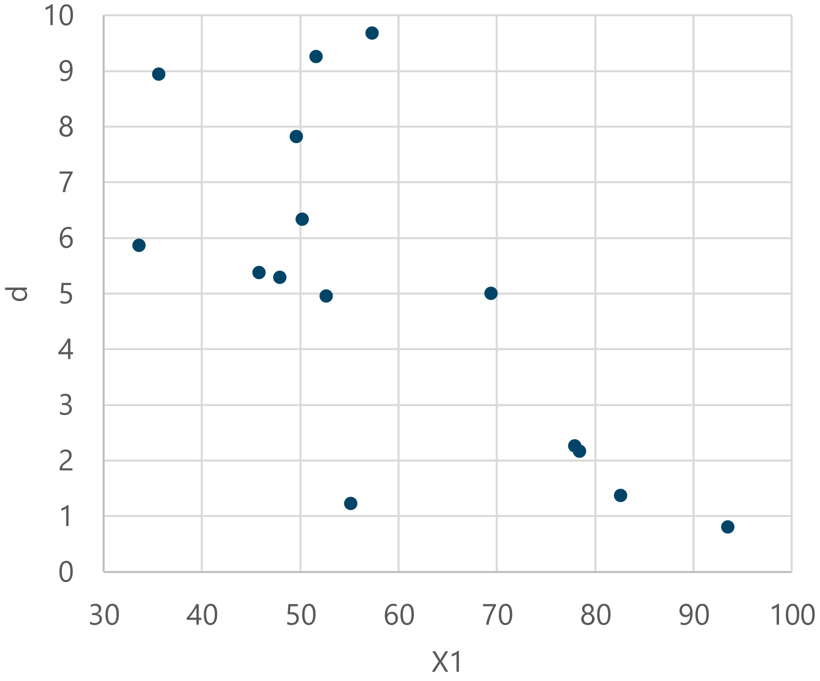
Deskriptivna analiza grupnih promjena

Deskriptivna analiza grupnih promjena - primjer 2:

x_1	x_2	d
82,61	83,98	1,37
93,51	94,31	0,80
78,46	80,62	2,16
55,14	56,36	1,22
49,65	57,47	7,82
45,82	51,19	5,37
50,21	56,54	6,33
51,65	60,91	9,26
69,45	74,45	5,00
57,32	67,00	9,68
35,62	44,56	8,94
47,95	53,24	5,29
33,65	39,51	5,86
52,69	57,64	4,95
77,95	80,21	2,26

	$\bar{x}$	s	min	max	R	$r_{x1,x2}$	$r_{x1,d}$
x1	58,77	17,63	33,65	93,51	59,86		
x2	63,86	15,67	39,51	94,31	54,80	0,99	-0,7
d	5,08	3,01	0,80	9,68	8,88		

Tjelesna masa sudionika programa u prosjeku se povećala za 5,08 kilograma. Program je imao veći učinak na povećanje tjelesne mase ispitanika manje početne tjelesne mase.



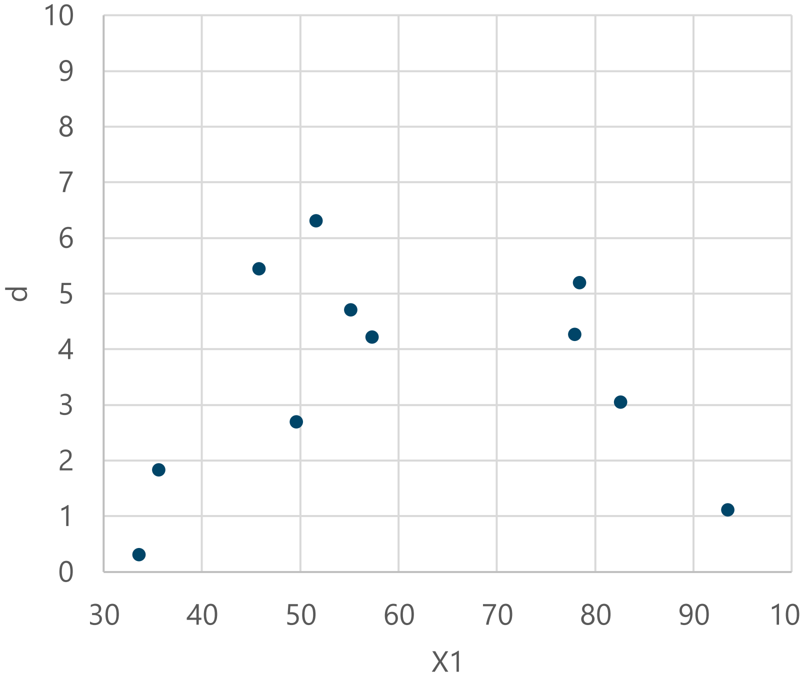
Deskriptivna analiza grupnih promjena

Deskriptivna analiza grupnih promjena - primjer 3:

x_1	x_2	d
82,61	85,65	3,04
93,51	94,62	1,11
78,46	83,65	5,19
55,14	59,84	4,70
49,65	52,34	2,69
45,82	51,26	5,44
50,21	61,24	11,03
51,65	57,95	6,30
69,45	80,65	11,20
57,32	61,53	4,21
35,62	37,45	1,83
47,95	58,31	10,36
33,65	33,95	0,30
52,69	64,25	11,56
77,95	82,21	4,26

	$\bar{x}$	s	min	max	R	$r_{x1,x2}$	$r_{x1,d}$
x1	58,77	17,63	33,65	93,51	59,86		
x2	64,32	17,73	33,95	94,62	60,67	0,98	-0,08
d	5,54	3,79	0,30	11,56	11,26		

Tjelesna masa sudionika programa u prosjeku se povećala za 5,54 kilograma. Program je imao neravnomjeran učinak na povećanje tjelesne mase ispitanika.



Deskriptivna analiza grupnih promjena

Deskriptivna analiza grupnih promjena - primjeri 1-3:

x_1	${}_1x_2$	${}_2x_2$	${}_3x_2$	d_1	d_2	d_3
33,65	38,65	39,51	33,95	5,00	5,86	0,30
35,62	40,62	44,56	37,45	5,00	8,94	1,83
45,82	50,82	51,19	51,26	5,00	5,37	5,44
47,95	52,95	53,24	58,31	5,00	5,29	10,36
49,65	54,65	57,47	52,34	5,00	7,82	2,69
50,21	55,21	56,54	61,24	5,00	6,33	11,03
51,65	56,65	60,91	57,95	5,00	9,26	6,30
52,69	57,69	57,64	64,25	5,00	4,95	11,56
55,14	60,14	56,36	59,84	5,00	1,22	4,70
57,32	62,32	67,00	61,53	5,00	9,68	4,21
69,45	74,45	74,45	80,65	5,00	5,00	11,20
77,95	82,95	80,21	82,21	5,00	2,26	4,26
78,46	83,46	80,62	83,65	5,00	2,16	5,19
82,61	87,61	83,98	85,65	5,00	1,37	3,04
93,51	98,51	94,31	94,62	5,00	0,80	1,11

Nakon uzlaznog sortiranja entiteta prema rezultatima inicijalnog stanja lakše je uočiti eventualnu zavisnost učinka programa o inicijalnom stanju subjekta.

Aritmetička sredina varijable razlika između dvaju stanja opisuje efikasnost primijenjenog programa.

Standardna devijacija razlika između dvaju stanja opisuje variranje učinka primijenjenog programa među ispitanicima. Ako je cilj grupnih programa ravnomjeran napredak svih sudionika, velika standardna devijacija može upućivati na slabu primjerenost programa pojedinim sudionicima. U interpretaciji variranja učinaka programa korisno je pregledati i minimalnu i maksimalnu vrijednost promjene stanja.

Deskriptivna analiza grupnih promjena

Deskriptivna analiza grupnih promjena - primjeri 1-3:

x_1	${}_1x_2$	${}_2x_2$	${}_3x_2$	d_1	d_2	d_3
33,65	38,65	39,51	33,95	5,00	5,86	0,30
35,62	40,62	44,56	37,45	5,00	8,94	1,83
45,82	50,82	51,19	51,26	5,00	5,37	5,44
47,95	52,95	53,24	58,31	5,00	5,29	10,36
49,65	54,65	57,47	52,34	5,00	7,82	2,69
50,21	55,21	56,54	61,24	5,00	6,33	11,03
51,65	56,65	60,91	57,95	5,00	9,26	6,30
52,69	57,69	57,64	64,25	5,00	4,95	11,56
55,14	60,14	56,36	59,84	5,00	1,22	4,70
57,32	62,32	67,00	61,53	5,00	9,68	4,21
69,45	74,45	74,45	80,65	5,00	5,00	11,20
77,95	82,95	80,21	82,21	5,00	2,26	4,26
78,46	83,46	80,62	83,65	5,00	2,16	5,19
82,61	87,61	83,98	85,65	5,00	1,37	3,04
93,51	98,51	94,31	94,62	5,00	0,80	1,11

Ako je korelacija inicijalnog stanja i varijable razlika između dvaju stanja ($r_{x1,d}$) jednaka nuli to znači da je primijenjeni program bio primjeren svim ispitanicima, nezavisno o njihovom inicijalnom stanju.

Što je korelacija bliža *1* , to je primijenjeni program primjereniji ispitanicima s višim rezultatima inicijalnog stanja. Što je korelacija bliža *-1* , to je primijenjeni program primjereniji ispitanicima s nižim rezultatima inicijalnog stanja.

OPREZ kod analize obrnuto skaliranih varijabli!!!

Deskriptivna analiza individualnih promjena

Promjene stanja jednog subjekta analiziraju se putem ***analize vremenskih nizova***.

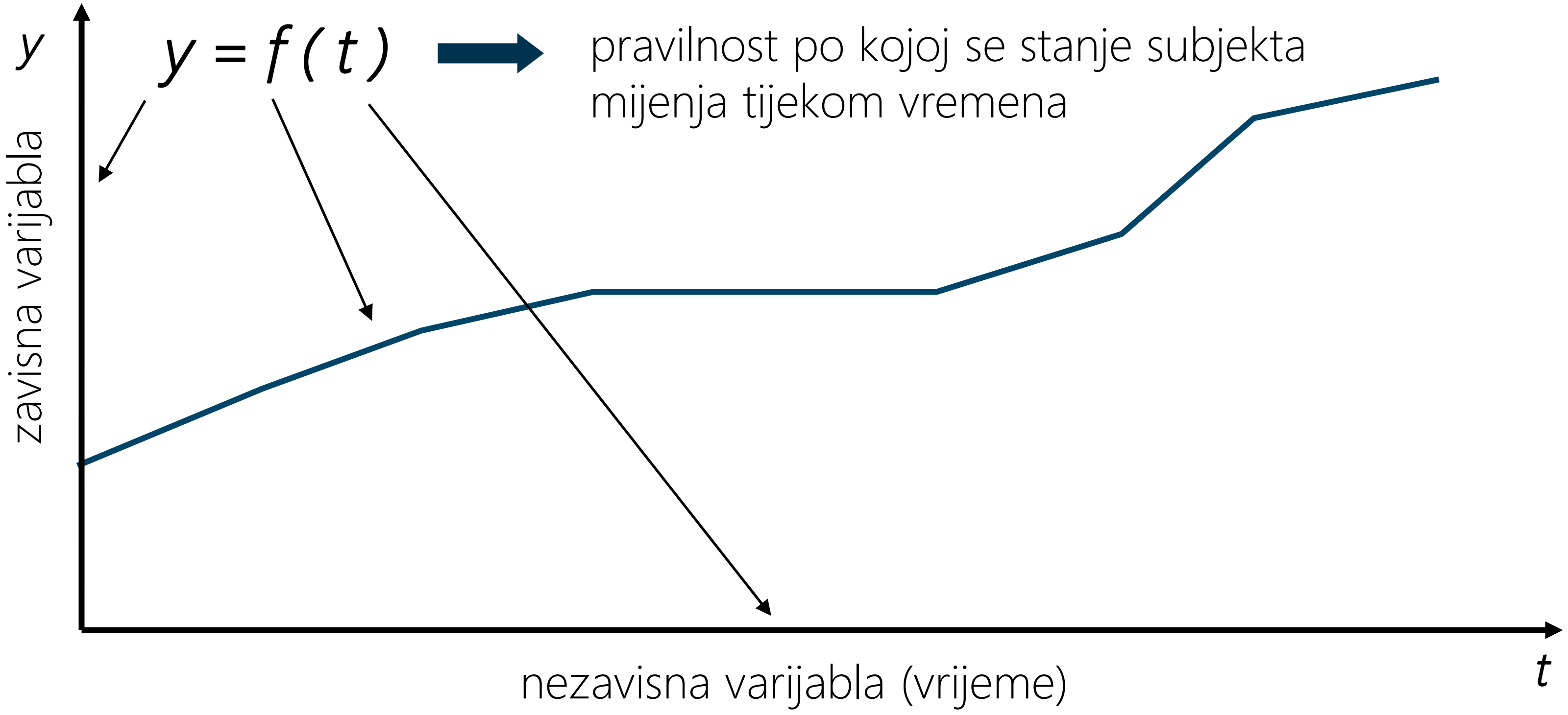
Dinamička analiza ili ***analiza vremenskih nizova*** služi za analizu promjena stanja subjekta kroz određeno vremensko razdoblje. Pri tome se utvrđuje zavisnost stanja subjekta (zavisna varijabla) o vremenu provođenja programa (nezavisna varijabla).

Vremenski niz je niz podataka o određenoj karakteristici subjekta prikupljenih u uzastopnim vremenskim točkama (npr. inicijalno stanje, prvo tranzitivno stanje, drugo tranzitivno stanje, finalno stanje).

Svrha analize vremenskih nizova je:

- ***praćenje*** vremenskog razvoja neke karakteristike subjekta
- ***utvrđivanje zakonitosti*** razvoja promatrane karakteristike
- ***predviđanje*** daljnjeg razvoja promatrane karakteristike.

Deskriptivna analiza individualnih promjena



$y = f(t)$ ➡ pravilnost po kojoj se stanje subjekta mijenja tijekom vremena

Deskriptivna analiza individualnih promjena

Vremenski niz se može analizirati putem *pokazatelja dinamike s promjenjivom bazom* ili *pokazatelja dinamike sa stalnom bazom*.

Pokazatelji dinamike s promjenjivom bazom izražavaju odstupanje stanja subjekta u određenoj vremenskoj točki u odnosu na stanje u prethodnoj vremenskoj točki.

Pokazatelji dinamike sa stalnom bazom izražavaju odstupanje stanja subjekta u određenoj vremenskoj točki u odnosu na početno stanje.

Deskriptivna analiza individualnih promjena

Pokazatelji dinamike s promjenjivom bazom

Apsolutna stopa promjene (Δy) s promjenjivom bazom izražava razliku rezultata subjekta u određenoj vremenskoj točki od rezultata u prethodnoj vremenskoj točki, a izračunava se formulom

$$\Delta y_i = y_i - y_{i-1}$$

gdje je

- y_i - rezultat subjekta u vremenskoj točki i
- $i = 2,\ldots,k$
- k - broj vremenskih točaka

Deskriptivna analiza individualnih promjena

Pokazatelji dinamike s promjenjivom bazom

Verižni indeks (V_i) pokazuje koliko puta je rezultat subjekta u određenoj vremenskoj točki veći od rezultata u prethodnoj vremenskoj točki, a izračunava se formulom

$$V_i = \frac{y_i}{y_{i-1}} \cdot 100$$

gdje je

- y_i - rezultat subjekta u vremenskoj točki i
- $i = 2,\dots,k$
- k - broj vremenskih točaka

Deskriptivna analiza individualnih promjena

Pokazatelji dinamike s promjenjivom bazom

Relativna stopa promjene (S_i) s promjenjivom bazom izražava postotak promjene rezultata subjekta u određenoj vremenskoj točki u odnosu na rezultat u prethodnoj vremenskoj točki, a izračunava se formulom

$$S_i = \frac{y_i}{y_{i-1}} \cdot 100 - 100 \qquad \text{ili} \qquad S_i = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100$$

gdje je

- y_i - rezultat subjekta u vremenskoj točki i
- $i = 2,\ldots,k$
- k - broj vremenskih točaka

Deskriptivna analiza individualnih promjena

Pokazatelji dinamike s promjenjivom bazom

Primjer: Na nekom sportašu primijenjen je program za povećanje mišićne mase. Kroz vremenski period od 11 mjeseci praćeno je stanje sportaša pri čemu je tjelesna masa izmjerena prije početka programa (inicijalno stanje), svakih mjesec dana tijekom provođenja programa (10 tranzitivnih stanja) i po završetku programa (finalno stanje).

Izračunati su apsolutni i relativni pokazatelji dinamike s promjenjivom bazom.

Deskriptivna analiza individualnih promjena

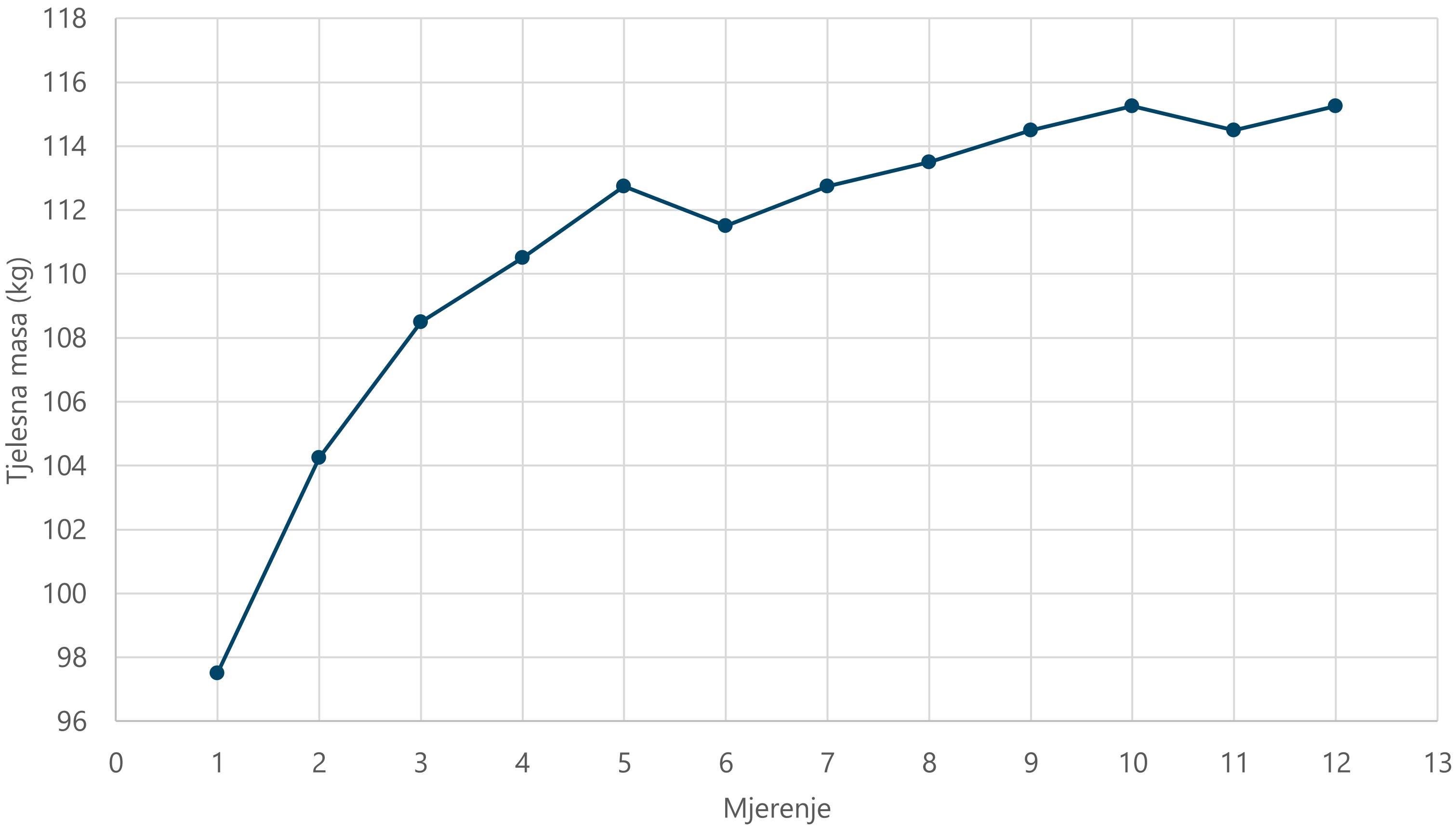
Pokazatelji dinamike s promjenjivom bazom

datum	mjerenje	tjelesna masa (kg)	apsolutna stopa promjene	verižni indeks	relativna stopa promjene
01.01.2007.	1	97,5			
01.02.2007.	2	104,25	6,75	107	6,92
01.03.2007.	3	108,5	4,25	104	4,08
01.04.2007.	4	110,5	2	102	1,84
01.05.2007.	5	112,75	2,25	102	2,04
01.06.2007.	6	111,5	-1,25	99	-1,11
01.07.2007.	7	112,75	1,25	101	1,12
01.08.2007.	8	113,5	0,75	101	0,67
01.09.2007.	9	114,5	1	101	0,88
01.10.2007.	10	115,25	0,75	101	0,66
01.11.2007.	11	114,5	-0,75	99	-0,65
01.12.2007.	12	115,25	0,75	101	0,66

Deskriptivna analiza individualnih promjena

Pokazatelji dinamike s promjenjivom bazom

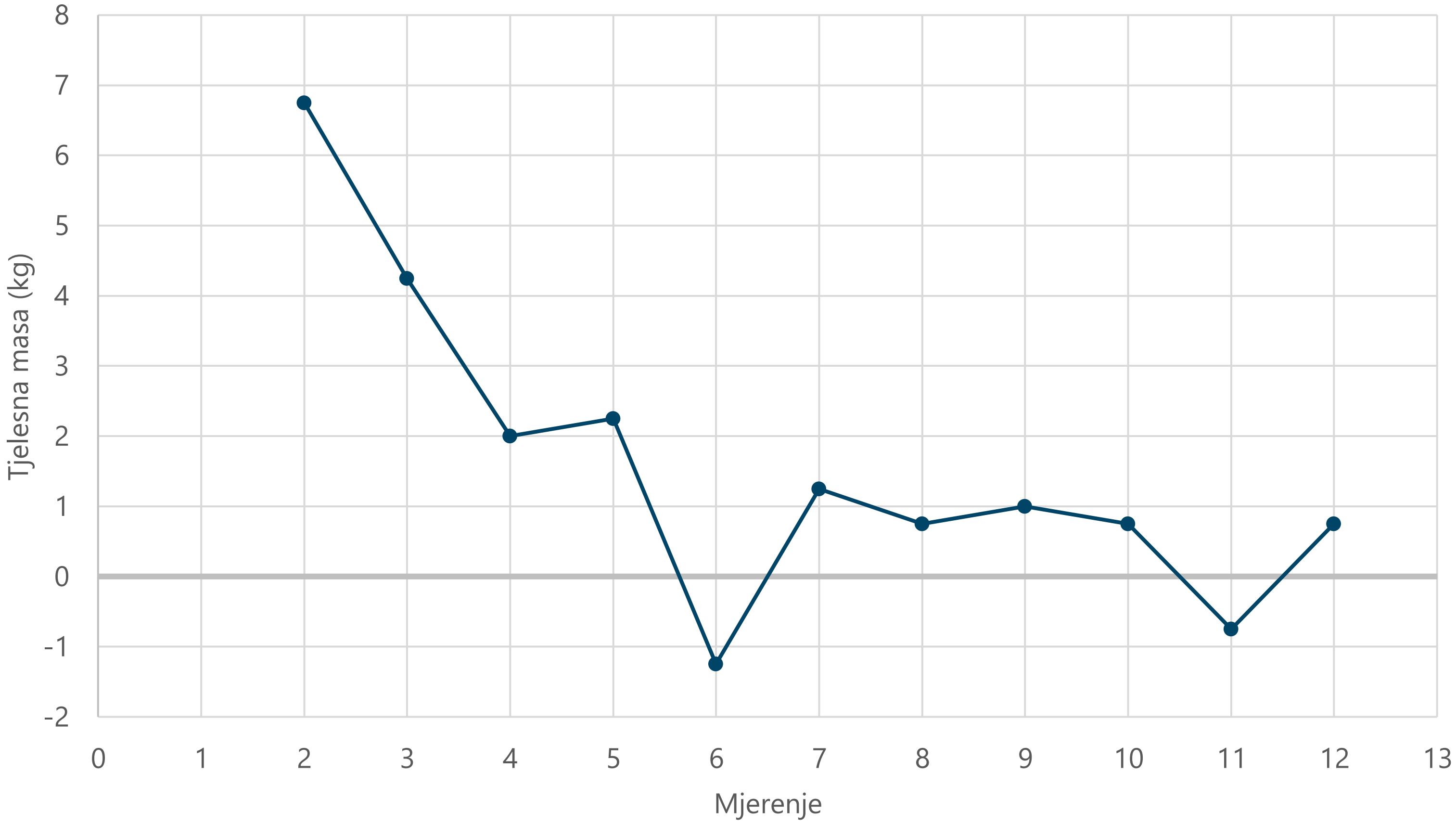
(Prikaz promjena tjelesne mase kroz vrijeme putem linijskog grafikona)



Deskriptivna analiza individualnih promjena

Pokazatelji dinamike s promjenjivom bazom

(Prikaz variranja apsolutne stope promjene s promjenjivom bazom kroz vrijeme putem linijskog grafikona)



Deskriptivna analiza individualnih promjena

Pokazatelji dinamike sa stalnom bazom

Apsolutna stopa promjene (Δy) sa stalnom bazom izražava razliku rezultata subjekta u određenoj vremenskoj točki od rezultata u početnoj vremenskoj točki, a izračunava se formulom

$$\Delta y_i = y_i - y_1$$

gdje je

- y_i - rezultat subjekta u vremenskoj točki i
- $i = 2,\ldots,k$
- k - broj vremenskih točaka

Deskriptivna analiza individualnih promjena

Pokazatelji dinamike sa stalnom bazom

Bazni indeks (I_i) pokazuje koliko puta je rezultat subjekta u određenoj vremenskoj točki veći od rezultata u početnoj vremenskoj točki, a izračunava se formulom

$$I_i = \frac{y_i}{y_1} \cdot 100$$

gdje je

- y_i - rezultat subjekta u vremenskoj točki i
- $i = 2,\ldots,k$
- k - broj vremenskih točaka

Deskriptivna analiza individualnih promjena

Pokazatelji dinamike sa stalnom bazom

Relativna stopa promjene (S_i) sa stalnom bazom izražava postotak promjene rezultata subjekta u određenoj vremenskoj točki u odnosu na rezultat u početnoj vremenskoj točki, a izračunava se formulom

$$S_i = \frac{y_i}{y_1} \cdot 100 - 100 \qquad \text{ili} \qquad S_i = \frac{y_i - y_1}{y_1} \cdot 100$$

gdje je

- y_i - rezultat subjekta u vremenskoj točki i
- $i = 2,\ldots,k$
- k - broj vremenskih točaka

Deskriptivna analiza individualnih promjena

Pokazatelji dinamike sa stalnom bazom

datum	mjerenje	tjelesna masa (kg)	apsolutna stopa promjene	bazni indeks	relativna stopa promjene
01.01.2007.	1	97,5			
01.02.2007.	2	104,25	6,75	107	6,92
01.03.2007.	3	108,5	11	111	11,28
01.04.2007.	4	110,5	13	113	13,33
01.05.2007.	5	112,75	15,25	116	15,64
01.06.2007.	6	111,5	14	114	14,36
01.07.2007.	7	112,75	15,25	116	15,64
01.08.2007.	8	113,5	16	116	16,41
01.09.2007.	9	114,5	17	117	17,44
01.10.2007.	10	115,25	17,75	118	18,21
01.11.2007.	11	114,5	17	117	17,44
01.12.2007.	12	115,25	17,75	118	18,21

Microsoft Excel

Zadatak 1: U datoteci *POD.xls* izračunajte prosječnu promjenu stanja tjelesne mase ispitanika između inicijalnog mjerenja (*ATT_I*) i finalnog mjerenja (*ATT_F*). Utvrdite da li je program jednako djelovao na tjelesnu masu svih ispitanika!

1. Izračunavanje deskriptivnih pokazatelja u svrhu analize grupnih promjena vrši se pomoću funkcija: ***AVERAGE*** (aritmetička sredina), ***MIN*** (minimum), ***MAX*** (maksimum) i ***STDEV*** (standardna devijacija). Funkcija se unosi u označeno polje matrice odabirom opcije ***Function***... padajućeg izbornika ***Insert***.

2. Izračunavanje Pearsonovog koeficijenta korelacije u svrhu analize grupnih promjena se vrši pomoću funkcije ***PEARSON***. Putem traka ***Array1*** i ***Array2*** potrebno je definirati niz podataka prve, odnosno druge varijable. Funkcija se unosi u označeno polje matrice odabirom opcije ***Function***... padajućeg izbornika ***Insert***.

Zadatak 2: U datoteci *TREND.xls* izračunajte verižne i bazne indekse te apsolutne i relativne stope promjene s promjenjivom i stalnom bazom za vremenski niz podataka o tjelesnoj masi ispitanika (*TEZ*)!

3

U svrhu izračunavanja pokazatelja dinamike s promjenjivom ili stalnom bazom formula za izračunavanje vrijednosti označenog polja unosi se u traku *fx (npr. =C3/C2)*. Pri unosu formula za izračunavanje pokazatelja dinamike sa stalnom bazom korisno je upotrijebiti apsolutne adrese polja (npr. =*C3/$C$2* ako je bazna vremenska točka u drugom retku matrice).

MATRIČNA ALGEBRA 1

Vježba 9

Matrična algebra 1

Matrična algebra je dio matematike koji se bavi računskim operacijama s matricama.

Matrica predstavlja skup brojeva smještenih u *n* redaka i *m* stupaca.

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

Oznaka matrice

Element a_{1m} nalazi se u retku 1 i stupcu m

Matrice se označavaju velikim masno otisnutim (bold) slovima (**A, B, C…**), a elementi matrice s malim slovima i indeksima (a_{11}, a_{12},…).

Matrična algebra 1

Matrica koja ima više redaka i jedan stupac naziva se **vektor stupca** ili samo **vektor**, dok se matrica s više stupaca i jednim retkom naziva **vektor retka** ili **transponirani vektor**.

$$\mathbf{a} = \begin{vmatrix} a_1 \\ . \\ . \\ a_n \end{vmatrix} \qquad \mathbf{a}^{T} = \begin{vmatrix} a_1 & . & . & a_n \end{vmatrix}$$

Vektori se označavaju malim masno otisnutim (bold) slovima (**a, b, c,…**), a transponirani vektori tako da se oznaci vektora doda eksponent ^{T} ($\mathbf{a}^T$, $\mathbf{b}^T$, $\mathbf{c}^T$,…).

Matrična algebra 1

VRSTE MATRICA

Matrica s jednakim brojem redaka i stupaca naziva se **kvadratna matrica**.

Primjer: Kvadratna matrica **A**

$$\mathbf{A} = \begin{vmatrix} 2 & 3 & 5 \\ 1 & 4 & 2 \\ 5 & 4 & 3 \end{vmatrix}$$

Matrična algebra 1

VRSTE MATRICA

Matricu koja je dobivena iz neke matrice A zamjenom stupaca redcima, a redaka stupcima naziva se **transponirana matrica** i označava se s **A^T**. Opisani postupak zove se **transponiranje matrice.**

Primjer: Matrica **A^T** je dobivena transponiranjem matrice **A**

$$\mathbf{A} = \begin{vmatrix} 2 & 1 & 3 & 4 \\ 3 & 2 & 3 & 6 \\ 4 & 7 & 1 & 0 \end{vmatrix} \qquad \mathbf{A}^{\mathbf{T}} = \begin{vmatrix} 2 & 3 & 4 \\ 1 & 2 & 7 \\ 3 & 3 & 1 \\ 4 & 6 & 0 \end{vmatrix}$$

Matrična algebra 1

VRSTE MATRICA

Ako je matrica **A** jednaka transponiranoj matrici $\mathbf{A}^\mathrm{T}$ naziva se **simetrična matrica**.

Primjer: Simetrične matrice **A** i $\mathbf{A}^\mathrm{T}$

$$\mathbf{A} = \begin{vmatrix} 2 & 1 & 5 \\ 1 & 4 & 2 \\ 5 & 2 & 3 \end{vmatrix} \qquad \mathbf{A}^\mathbf{T} = \begin{vmatrix} 2 & 1 & 5 \\ 1 & 4 & 2 \\ 5 & 2 & 3 \end{vmatrix}$$

Matrična algebra 1

VRSTE MATRICA

Matrica koja u dijagonali ima elemente različite od nule, dok su svi ostali elementi jednaki nuli, naziva se **dijagonalna matrica**.

Primjer: Dijagonalna matrica D

$$\mathbf{D} = \begin{vmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{vmatrix}$$

Matrična algebra 1

VRSTE MATRICA

Skalarna matrica je poseban slučaj dijagonalne matrice kod koje su dijagonalni elementi jednaki.

Primjer: Skalarna matrica **S**

$$\mathbf{S} = \begin{vmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{vmatrix}$$

Matrična algebra 1

VRSTE MATRICA

Poseban slučaj skalarne matrice u kojoj su dijagonalni elementi jednaki jedinici naziva se **matrica identiteta**.

Primjer: Matrica identiteta **I**

$$\mathbf{I} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

Matrična algebra 1

RAČUNSKE OPERACIJE S MATRICAMA

Zbrajanje i oduzimanje matrica može se provoditi uz uvjet da matrice imaju jednak broj redaka i stupaca. Operacija se vrši tako da zbrojimo, odnosno oduzmemo odgovarajuće elemente matrica.

$$\mathbf{A}+\mathbf{B}=\mathbf{C}=\begin{vmatrix} a_{11}+b_{11} & a_{12}+b_{12} & . & . & a_{1m}+b_{1m} \\ a_{21}+b_{21} & a_{22}+b_{22} & . & . & a_{2m}+b_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & . & . & a_{nm}+b_{nm} \end{vmatrix}$$

Matrična algebra 1

RAČUNSKE OPERACIJE S MATRICAMA

Primjer: Zbrajanje matrica **A** i **B**

$$\mathbf{A} = \begin{vmatrix} 2 & 1 & 5 \\ 1 & 4 & 2 \\ 5 & 2 & 3 \end{vmatrix} \quad \mathbf{B} = \begin{vmatrix} 1 & 1 & 4 \\ 1 & 5 & 2 \\ 3 & 1 & 0 \end{vmatrix} \quad \mathbf{A} + \mathbf{B} = \mathbf{C} = \begin{vmatrix} 3 & 2 & 9 \\ 2 & 9 & 4 \\ 8 & 3 & 3 \end{vmatrix}$$

Matrična algebra 1

RAČUNSKE OPERACIJE S MATRICAMA

Množenje matrica provodi se tako da se zbrajaju produkti elemenata redaka prve matrice i odgovarajućih elemenata stupaca druge matrice uz uvjet da je broj stupaca prve matrice jednak broju redaka druge matrice.

$$\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}, \qquad \mathbf{B} = \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{vmatrix}$$

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{C} = \begin{vmatrix} a_{11} \cdot b_{11} + a_{12} \cdot b_{21} = c_{11} & a_{11} \cdot b_{12} + a_{12} \cdot b_{22} = c_{12} & a_{11} \cdot b_{13} + a_{12} \cdot b_{23} = c_{13} \\ a_{21} \cdot b_{11} + a_{22} \cdot b_{21} = c_{21} & a_{21} \cdot b_{12} + a_{22} \cdot b_{22} = c_{22} & a_{21} \cdot b_{13} + a_{22} \cdot b_{23} = c_{23} \\ a_{31} \cdot b_{11} + a_{32} \cdot b_{21} = c_{31} & a_{31} \cdot b_{12} + a_{32} \cdot b_{22} = c_{32} & a_{31} \cdot b_{13} + a_{32} \cdot b_{23} = c_{33} \end{vmatrix}$$

Matrična algebra 1

RAČUNSKE OPERACIJE S MATRICAMA

Primjer: Množenje matrica **A** i **B**

$$\mathbf{A} = \begin{vmatrix} 1 & 3 & 1 \\ 4 & 1 & 4 \end{vmatrix} \qquad \mathbf{B} = \begin{vmatrix} 2 & 1 & 3 & 4 \\ 3 & 2 & 3 & 6 \\ 4 & 7 & 1 & 0 \end{vmatrix}$$

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{C} = \begin{vmatrix} 1 \cdot 2 + 3 \cdot 3 + 1 \cdot 4 = 15 & 1 \cdot 1 + 3 \cdot 2 + 1 \cdot 7 = 13 & 1 \cdot 3 + 3 \cdot 3 + 1 \cdot 1 = 13 & 1 \cdot 4 + 3 \cdot 6 + 1 \cdot 0 = 22 \\ 4 \cdot 2 + 1 \cdot 3 + 4 \cdot 4 = 27 & 4 \cdot 1 + 1 \cdot 2 + 4 \cdot 7 = 34 & 4 \cdot 3 + 1 \cdot 3 + 4 \cdot 1 = 19 & 4 \cdot 4 + 1 \cdot 6 + 4 \cdot 0 = 22 \end{vmatrix}$$

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{C} = \begin{vmatrix} 15 & 13 & 13 & 22 \\ 27 & 34 & 19 & 22 \end{vmatrix}$$

Matrična algebra 1

RAČUNSKE OPERACIJE S MATRICAMA

Množenjem nekog vektorom **a** (vektor stupca) s nekim transponiranim vektorom $\mathbf{b}^{\mathrm{T}}$ (vektor retka) uvijek se dobije matrica.

Primjer: Množenje vektora **a** i $\mathbf{b}^{\mathrm{T}}$

$$\mathbf{a} = \begin{vmatrix} 3 \\ 2 \\ 0 \end{vmatrix} \qquad \mathbf{b}^{\mathrm{T}} = \begin{vmatrix} 1 & -3 & 4 \end{vmatrix} \qquad \mathbf{a} \cdot \mathbf{b}^{\mathrm{T}} = \mathbf{c} = \begin{vmatrix} 3 & -9 & 12 \\ 2 & -6 & 8 \\ 0 & 0 & 0 \end{vmatrix}$$

Matrična algebra 1

RAČUNSKE OPERACIJE S MATRICAMA

Množenjem nekog transponiranog vektora a^T (vektor retka) s nekim vektorom b (vektor stupca) uvijek se dobije skalar.

$$\mathbf{a}^{\mathrm{T}} \cdot \mathbf{b} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{a_i b_i}$$

Primjer: Množenje vektora $\mathbf{a}^T$ i **b**

$$\mathbf{a}^{\mathrm{T}} = \begin{vmatrix} 1 & -3 & 4 \end{vmatrix} \qquad \mathbf{b} = \begin{vmatrix} 3 \\ 2 \\ 0 \end{vmatrix} \qquad \mathbf{a}^{\mathrm{T}} \cdot \mathbf{b} = -3$$

Matrična algebra 1

RAČUNSKE OPERACIJE S MATRICAMA

Množenje matrice skalarom vrši se tako da se svaki element matrice pomnoži skalarom.

$$\varphi \cdot \mathbf{A} = \begin{bmatrix} \varphi \cdot a_{11} & \cdots & \varphi \cdot a_{1m} \\ \vdots & \ddots & \vdots \\ \varphi \cdot a_{n1} & \cdots & \varphi \cdot a_{nm} \end{bmatrix}$$

Primjer: Množenje matrice **A** skalarom 3

$$\mathbf{A} = \begin{vmatrix} 2 & 4 & 2 \\ 5 & 2 & 6 \\ 1 & 3 & 3 \\ 6 & 5 & 5 \end{vmatrix} \quad 3 \cdot \mathbf{A} = \begin{vmatrix} 3 \cdot 2 & 3 \cdot 4 & 3 \cdot 2 \\ 3 \cdot 5 & 3 \cdot 2 & 3 \cdot 6 \\ 3 \cdot 1 & 3 \cdot 3 & 3 \cdot 3 \\ 3 \cdot 6 & 3 \cdot 5 & 3 \cdot 5 \end{vmatrix} = \begin{vmatrix} 6 & 12 & 6 \\ 15 & 6 & 18 \\ 3 & 9 & 9 \\ 18 & 15 & 15 \end{vmatrix}$$

MATRIČNA ALGEBRA 2

Vježba 10

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Trag matrice je zbroj elemenata glavne dijagonale matrice.

Primjer: Trag matrice **A**

$$\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$\mathbf{A} = \begin{vmatrix} 2 & 1 & 5 \\ 1 & 4 & 2 \\ 5 & 2 & 3 \end{vmatrix}$$

$\mathrm{trag}(\mathbf{A}) = a_{11} + a_{22} + a_{33}$

$\mathrm{trag}(\mathbf{A}) = 2 + 4 + 3 = 9$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Duljina ili norma vektora dobije se operacijom

$$|\mathbf{a}| = \sqrt{\mathrm{a}_1^2 + \mathrm{a}_2^2 + \mathrm{a}_3^2 + \ldots + \mathrm{a}_\mathrm{n}^2} = \left(\sum_{i=1}^{n} \mathrm{a}_\mathrm{i}^2\right)^{1/2}$$

odnosno

$$|\mathbf{a}| = (\mathbf{a}^\mathrm{T}\mathbf{a})^{1/2}$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Vektor čija je duljina jednaka 1 naziva se **normirani vektor**. Normirani vektor (**â**) se izračunava postupkom **normiranja**, odnosno dijeljenjem vektora sa svojom duljinom:

$$\hat{\mathbf{a}} = \frac{\mathbf{a}}{|\mathbf{a}|}$$

odnosno

$$\hat{\mathbf{a}} = \mathbf{a}(\mathbf{a}^{\mathrm{T}}\mathbf{a})^{-1/2}$$

RAČUNSKE OPERACIJE S MATRICAMA

Primjer: Normiranje vektora **a**

$$\mathbf{a} = \begin{vmatrix} 2 \\ 4 \\ 3 \\ 1 \\ 5 \end{vmatrix}$$

$$\begin{aligned} |\mathbf{a}| &= \sqrt{2^2 + 4^2 + 3^2 + 1^2 + 5^2} \\ &= \sqrt{4 + 16 + 9 + 1 + 25} \\ &= \sqrt{55} \\ &= 7{,}42 \end{aligned}$$

$$\hat{\mathbf{a}} = \begin{vmatrix} 2 \\ 4 \\ 3 \\ 1 \\ 5 \end{vmatrix} \cdot \frac{1}{7{,}42} = \begin{vmatrix} 0{,}27 \\ 0{,}54 \\ 0{,}4 \\ 0{,}13 \\ 0{,}67 \end{vmatrix}$$

$$|\hat{\mathbf{a}}| = \sqrt{0{,}27^2 + 0{,}54^2 + 0,4^2 + 0{,}13^2 + 0{,}67^2} = \sqrt{1} = 1$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Udaljenost između dvaju vektora istog reda izračunava se kao norma razlike dvaju vektora, a naziva se **euklidska udaljenost.**

$$\mathbf{d} = |\mathbf{a} - \mathbf{b}| = [(\mathbf{a} - \mathbf{b})^T(\mathbf{a} - \mathbf{b})]^{1/2}$$

odnosno

$$\mathbf{d} = \left(\sum_{i=1}^{n}(a_i - b_i)^2\right)^{1/2}$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Primjer: Euklidska udaljenost između vektora **a** i **b**

$$\mathbf{a} = \begin{vmatrix} 2 \\ 4 \\ 3 \end{vmatrix} \qquad \mathbf{b} = \begin{vmatrix} 4 \\ 2 \\ 2 \end{vmatrix} \qquad \mathbf{a} - \mathbf{b} = \begin{vmatrix} -2 \\ 2 \\ 1 \end{vmatrix} \qquad (\mathbf{a} - \mathbf{b})^{\mathrm{T}} = \begin{vmatrix} -2 & 2 & 1 \end{vmatrix}$$

$$\mathbf{d} = \left[(\mathbf{a} - \mathbf{b})^{\mathrm{T}}(\mathbf{a} - \mathbf{b})\right]^{\frac{1}{2}} = \left[\begin{vmatrix} -2 & 2 & 1 \end{vmatrix} \cdot \begin{vmatrix} -2 \\ 2 \\ 1 \end{vmatrix}\right]^{\frac{1}{2}} = \sqrt{9} = 3$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Kosinus kuta između dvaju vektora istog reda izračunava se kao omjer skalarnog produkta dvaju vektora i umnoška njihovih normi.

$$\cos\alpha = \left(\mathbf{a}^{\mathrm{T}}\mathbf{b}\right)\left(\mathbf{a}^{\mathrm{T}}\mathbf{a}\right)^{-1/2}\left(\mathbf{b}^{\mathrm{T}}\mathbf{b}\right)^{-1/2}$$

odnosno

$$\cos\alpha = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

RAČUNSKE OPERACIJE S MATRICAMA

Primjer: Kosinus kuta između vektora **a** i **b**

$$\mathbf{a} = \begin{vmatrix} 2 \\ 4 \\ 3 \end{vmatrix} \qquad \mathbf{b} = \begin{vmatrix} 4 \\ 5 \\ 3 \end{vmatrix}$$

$$|\mathbf{a}| = \sqrt{2^2 + 4^2 + 3^2} = \sqrt{29} = 5{,}39$$

$$|\mathbf{b}| = \sqrt{4^2 + 5^2 + 3^2} = \sqrt{50} = 7{,}07$$

$$\mathbf{a}^{\mathrm{T}}\mathbf{b} = \begin{vmatrix} 2 & 4 & 3 \end{vmatrix} \cdot \begin{vmatrix} 4 \\ 5 \\ 3 \end{vmatrix} = (2 \cdot 4) + (4 \cdot 5) + (3 \cdot 3) = 37$$

$$\cos\alpha = \frac{37}{5{,}39 \cdot 7{,}07} = \frac{37}{38{,}11} = 0{,}97$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Ako se rezultati u varijablama centriraju

$$a_{ci} = a_i - \bar{a}$$
$$b_{ci} = b_i - \bar{b}$$

tada je kosinus kuta α jednak

$$\cos\alpha = \frac{\sum_{i=1}^{n} a_{ci} b_{ci}}{\sqrt{\sum_{i=1}^{n} a_{ci}^2} \cdot \sqrt{\sum_{i=1}^{n} b_{ci}^2}}$$

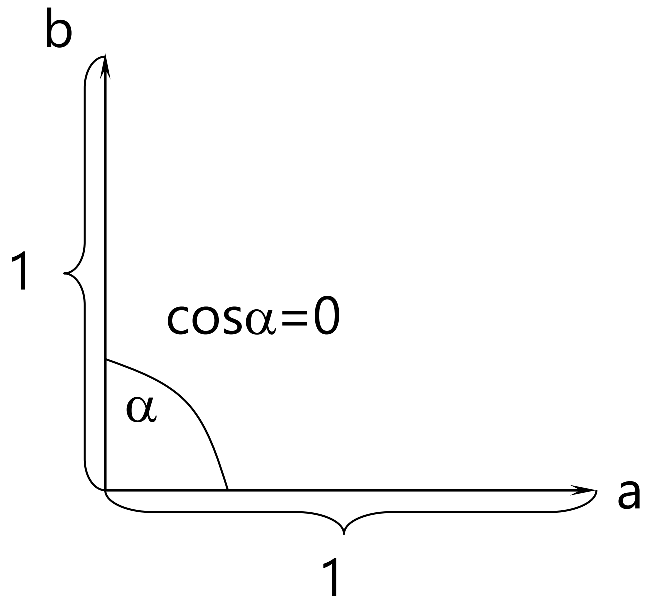
što je formula za izračunavanje koeficijenta korelacije, pa je

$$r_{ab} = \cos\alpha_{ab}$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

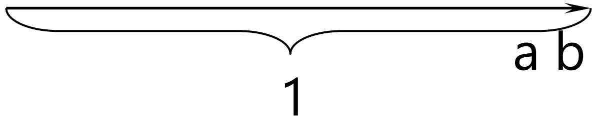
Ako je korelacija između varijabli jednaka nuli ($r_{xy}=0$), onda su dva centrirana vektora pod kutem od 90°.



RAČUNSKE OPERACIJE S MATRICAMA

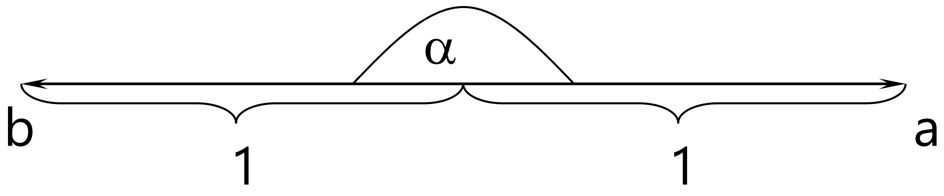
Ako je korelacija između varijabli potpuna pozitivna ($r_{xy}=1$), onda je kut između dvaju centriranih vektora 0°.

$\cos\alpha = 1$



Ako je korelacija između varijabli potpuna negativna ($r_{xy}=-1$), onda je kut između dvaju centriranih vektora 180°

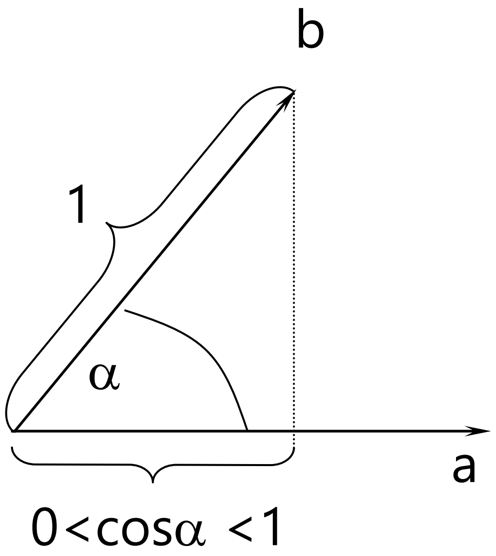
$\cos\alpha = -1$



Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

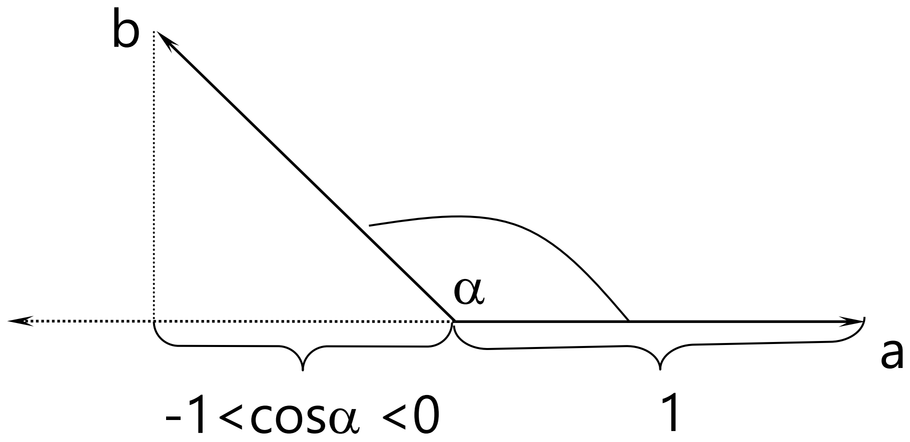
Ako je korelacija nepotpuna pozitivna ($0<r_{xy}<1$), onda je kut između dvaju centriranih vektora veći od 0°, a manji od 90°.



Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Ako je korelacija nepotpuna negativna ($-1<r_{xy}<0$), onda je kut između dvaju centriranih vektora veći od 90°, a manji od 180°.



Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Linearna kombinacija vektora je vektor koji je nastao zbrajanjem drugih vektora ponderiranih pripadajućim skalarima.

Ako su $\mathbf{a}_j$ (j=1,...,m) vektori istog reda, a β_j pripadajući skalari onda je vektor **b** linearna kombinacija vektora $\mathbf{a}_j$

$$\mathbf{b} = \sum_{j=1}^{m} \beta_j \mathbf{a}_\mathrm{j} = \beta_1 \mathbf{a_1} + \beta_2 \mathbf{a_2} + \ldots + \beta_m \mathbf{a_m}$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Jednostavna linearna kombinacija je vektor nastao zbrajanjem drugih vektora istog reda ponderiranih jednakim skalarima (**ponderima**).

Diferencijalno ponderirana linearna kombinacija je vektor nastao zbrajanjem drugih vektora istog reda ponderiranih različitim skalarima.

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Ako su **a**, **b** i **c** vektori istog reda, a α, β i γ skalari, novi vektor **d** nastao je linearnom kombinacijom vektora **a**, **b** i **c** ponderiranih skalarima α, β i γ.

$$\mathbf{d} = \alpha \begin{vmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{vmatrix} + \beta \begin{vmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{vmatrix} + \gamma \begin{vmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{vmatrix} = \begin{vmatrix} \alpha \cdot a_1 + \beta \cdot b_1 + \gamma \cdot c_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha \cdot a_n + \beta \cdot b_n + \gamma \cdot c_n \end{vmatrix}$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Za neki redak ili stupac matrice kažemo da je **linearno zavisan** ako se može izraziti kao linearna kombinacija drugih redaka ili stupaca te matrice.

Rang matrice jednak je minimalnom broju redaka ili stupaca u matrici čijom se linearnom kombinacijom mogu izraziti svi ostali redci ili stupci te matrice.

Primjer: Rang matrice **A** je 2 jer se, primjerice, prvi stupac može izračunati kao linearna kombinacija elemenata drugog i trećeg stupca.

$$\mathbf{A} = \begin{vmatrix} 4 & 3 & 1 \\ 3 & 0 & 3 \\ 9 & 4 & 5 \end{vmatrix}$$

$$\begin{aligned} a_{11} &= a_{12} + a_{13} = 3 + 1 = 4 \\ a_{21} &= a_{22} + a_{23} = 0 + 3 = 3 \\ a_{31} &= a_{32} + a_{33} = 4 + 5 = 9 \end{aligned}$$

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Inverz matrice u matričnoj algebri odgovara recipročnoj vrijednosti broja (skalara) u skalarnoj algebri. Matrica $\mathbf{A}^{-1}$ je inverz matrice **A** ako je

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{I}$$

Inverz matrice moguće je izračunati samo ako je matrica kvadratna i punog ranga.

Matrična algebra 2

RAČUNSKE OPERACIJE S MATRICAMA

Inverz matrice može se koristiti za **rješavanje sustava linearnih jednadžbi u matričnom obliku** na sljedeći način:

$$\mathbf{A}\,\mathbf{x} = \mathbf{y} \quad / \, \mathbf{A}^{-1}$$

$$\mathbf{A}^{-1}\,\mathbf{A}\,\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{y}$$

$$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{y}$$

gdje je

- **A** kvadratna matrica reda *n* x *n* poznatih vrijednosti
- **x** vektor stupca reda n x 1 nepoznatih vrijednosti
- **y** vektor stupca reda n x 1 poznatih vrijednosti

REGRESIJSKA ANALIZA

Vježba 11

Regresijska analiza

Regresijska analiza je matematičko-statistički postupak kojim se utvrđuje odgovarajuća funkcionalna veza (relacija) između jedne **zavisne** ili **kriterijske varijable** i jedne ili više **nezavisnih** ili **prediktorskih varijabli**.

Zavisna (kriterijska) varijabla je varijabla čiji se varijabilitet objašnjava putem nezavisnih varijabli.

Nezavisne (prediktorske) varijable su varijable na temelju kojih se objašnjava varijabilitet zavisne varijable.

Regresijska analiza se u kineziologiji najčešće koristi u svrhu:

- utvrđivanja utjecaja jedne varijable ili skupa varijabli na neku kriterijsku varijablu (npr. utvrđivanje utjecaja građe tijela na rezultat u bacanju kugle) i

- utvrđivanje trenda razvoja rezultata u nekom sportu (npr. utvrđivanje trenda razvoja najboljih rezultata u bacanju kugle na svjetskim prvenstvima)

Regresijska analiza

Funkcionalna veza između prediktorskih varijabli i kriterijske varijable definira se utvrđivanjem odgovarajuće **regresijske jednadžbe**. Opći oblik regresijske jednadžbe izgleda ovako:

$$y = f(x_1, x_2,\ldots, x_m) + \mathrm{e}$$

- y - zavisna (kriterijska) varijabla
- f - odgovarajuća funkcija
- $x_1, x_2,\ldots, x_m$ - nezavisne (prediktorske) varijable
- e - greška prognoze.

Regresijska analiza

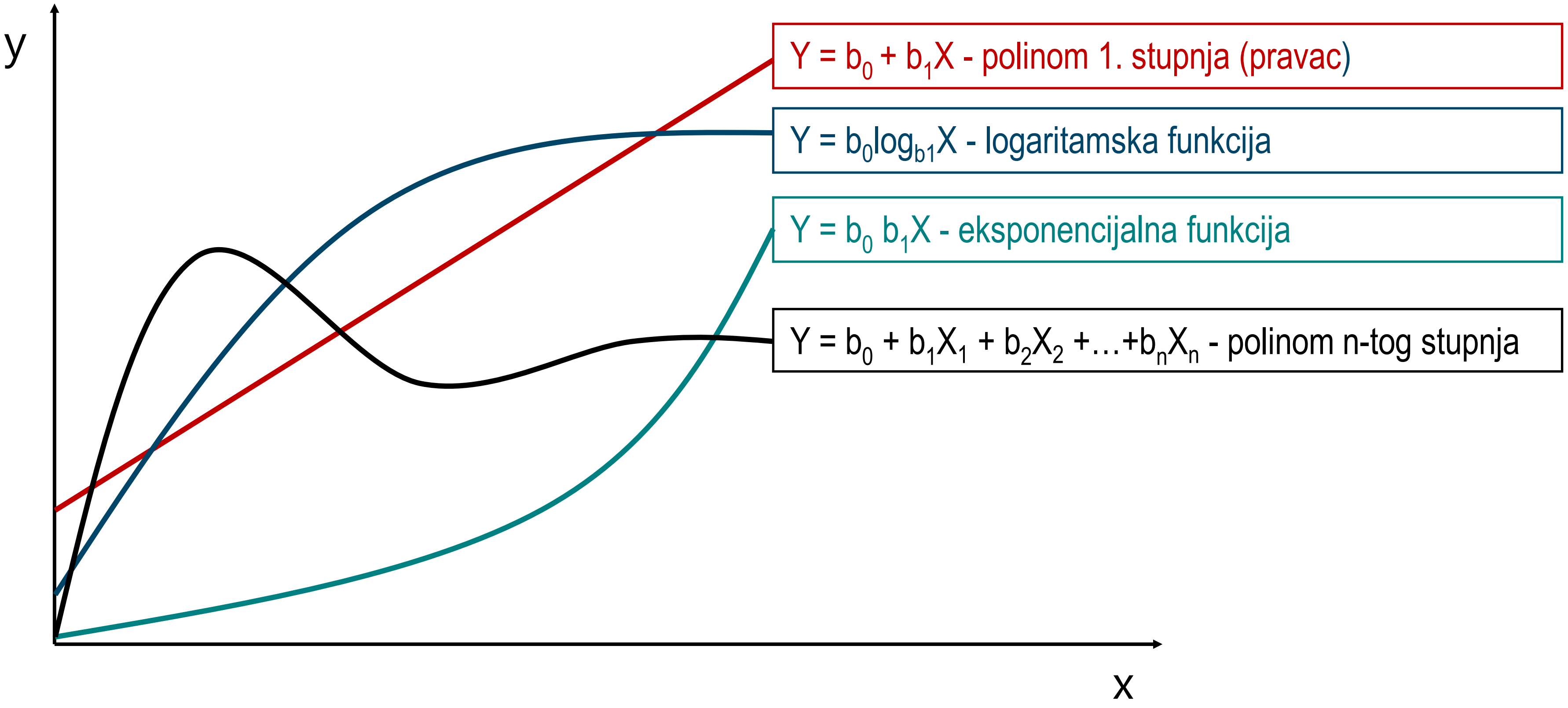
Regresijske modele moguće je generalno podijeliti na temelju dvaju kriterija i to:

- prema broju nezavisnih varijabli na:
    - ✓ **jednostavne (simple) regresijske modele** i
    - ✓ **višestruke (multiple) regresijske modele**, te

- prema odnosu između zavisne i nezavisnih varijabli na:
    - ✓ **linearne regresijske modele** i
    - ✓ **nelinearne regresijske modele**.

Regresijska analiza

Linearni i nelinearni modeli jednostavne regresijske analize:



$Y = b_0 + b_1X$ - polinom 1. stupnja (pravac)

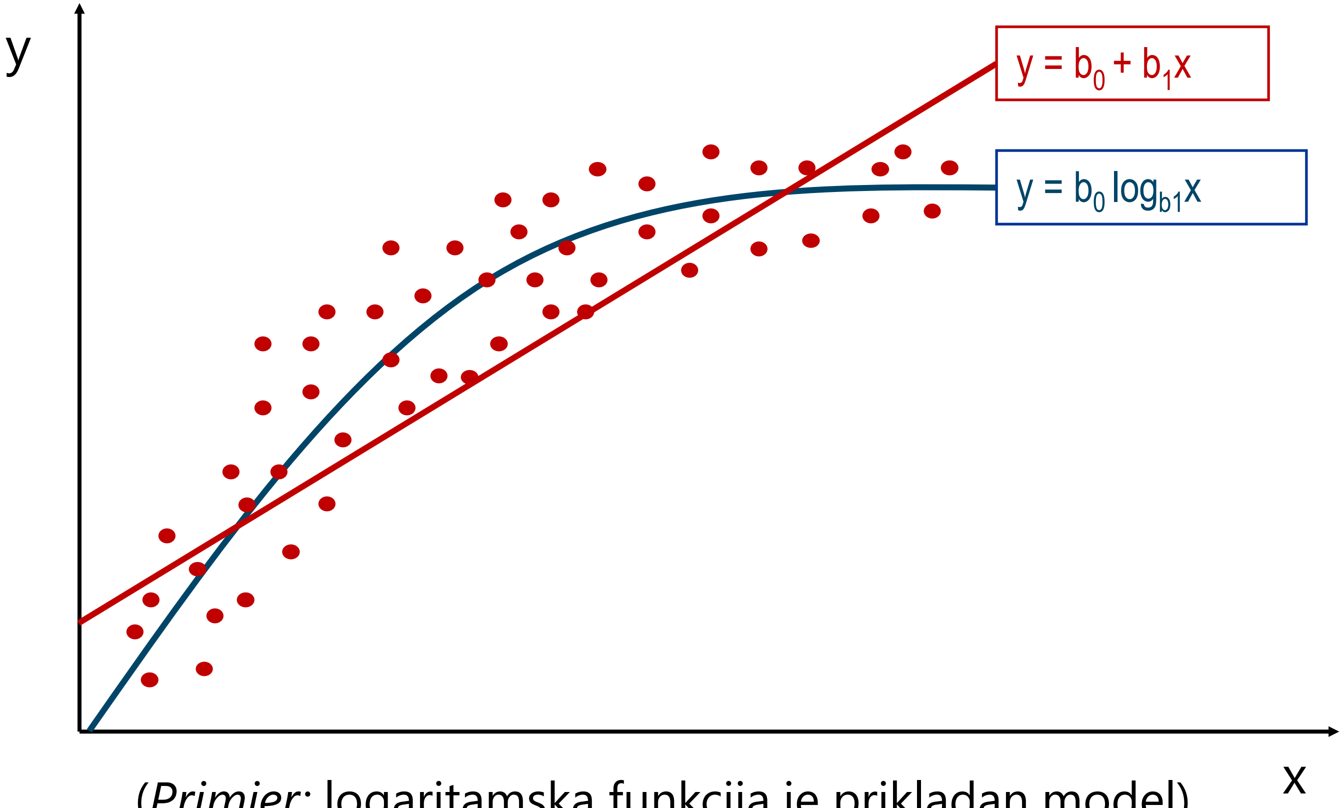
$Y = b_0 \log_{b1}X$ - logaritamska funkcija

$Y = b_0\ b_1X$ - eksponencijalna funkcija

$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n$ - polinom n-tog stupnja

Regresijska analiza

Odabir modela jednostavne regresijske analize vrši se pomoću korelacijskog dijagrama.



(*Primjer*: logaritamska funkcija je prikladan model)

Regresijska analiza

Jednostavna linearna regresijska analiza

Jednostavnom linearnom regresijskom analizom utvrđuje se linearna povezanost između jedne nezavisne (prediktorske) i jedne zavisne (kriterijske) varijable pri čemu regresijska jednadžba ima sljedeći oblik:

$$y_i = b_0 + b_1 x_i + e_i$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- b_0 i b_1 - regresijski koeficijenti
- x_i - rezultat entiteta *i* u prediktorskoj varijabli
- e_i - rezidualna vrijednost entiteta *i*
- $i = 1,\ldots,n$
- *n – broj entiteta*

Regresijska analiza

Jednostavna linearna regresijska analiza

Regresijski koeficijenti omogućavaju prognoziranje rezultata entiteta u kriterijskoj varijabli na temelju rezultata u prediktorskoj varijabli putem sljedeće formule:

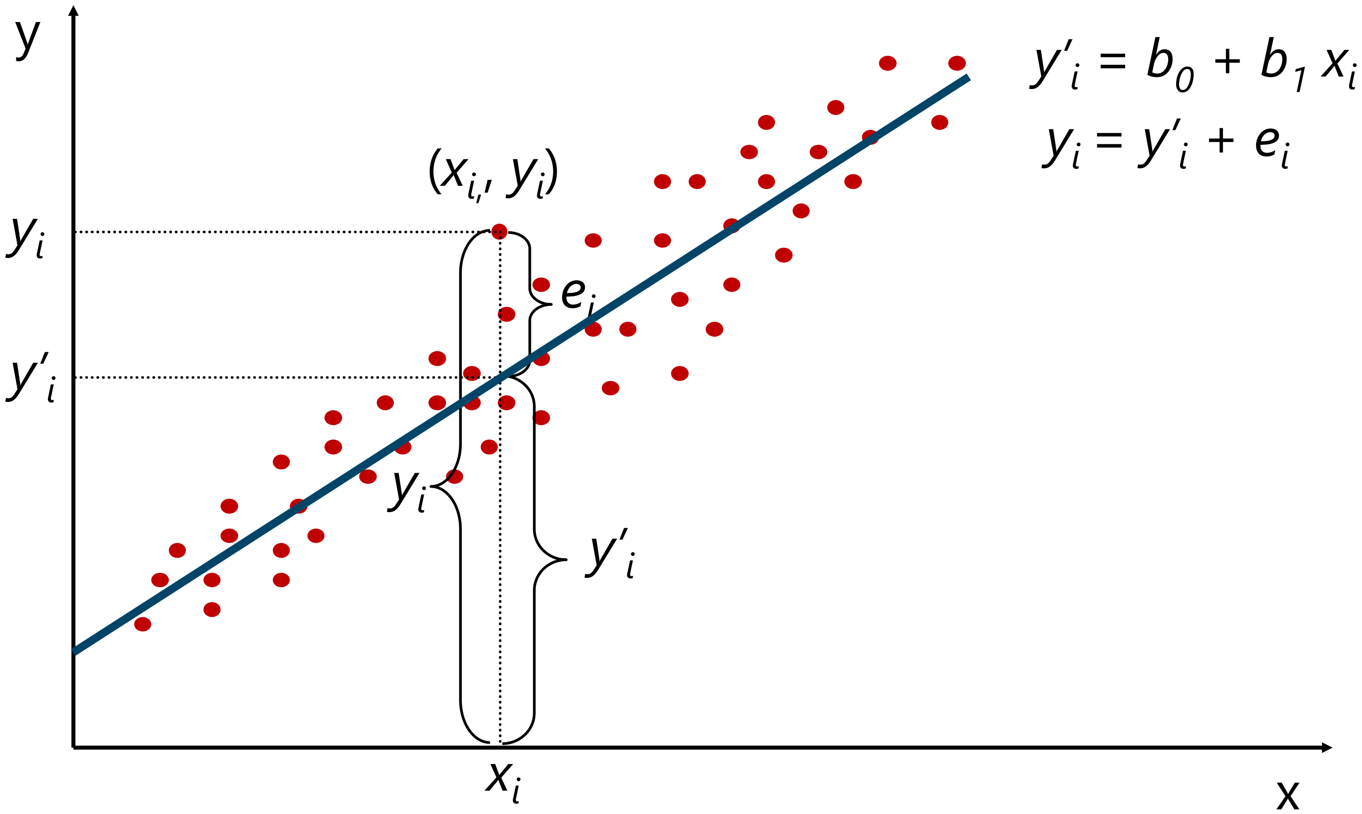
$$y'_i = b_0 + b_1 x_i$$

gdje je

- y'_i - prognozirani rezultat entiteta i u kriterijskoj varijabli
- b_0 i b_1 - regresijski koeficijenti
- x_i - rezultat entiteta i u prediktorskoj varijabli.

Regresijska analiza

Jednostavna linearna regresijska analiza



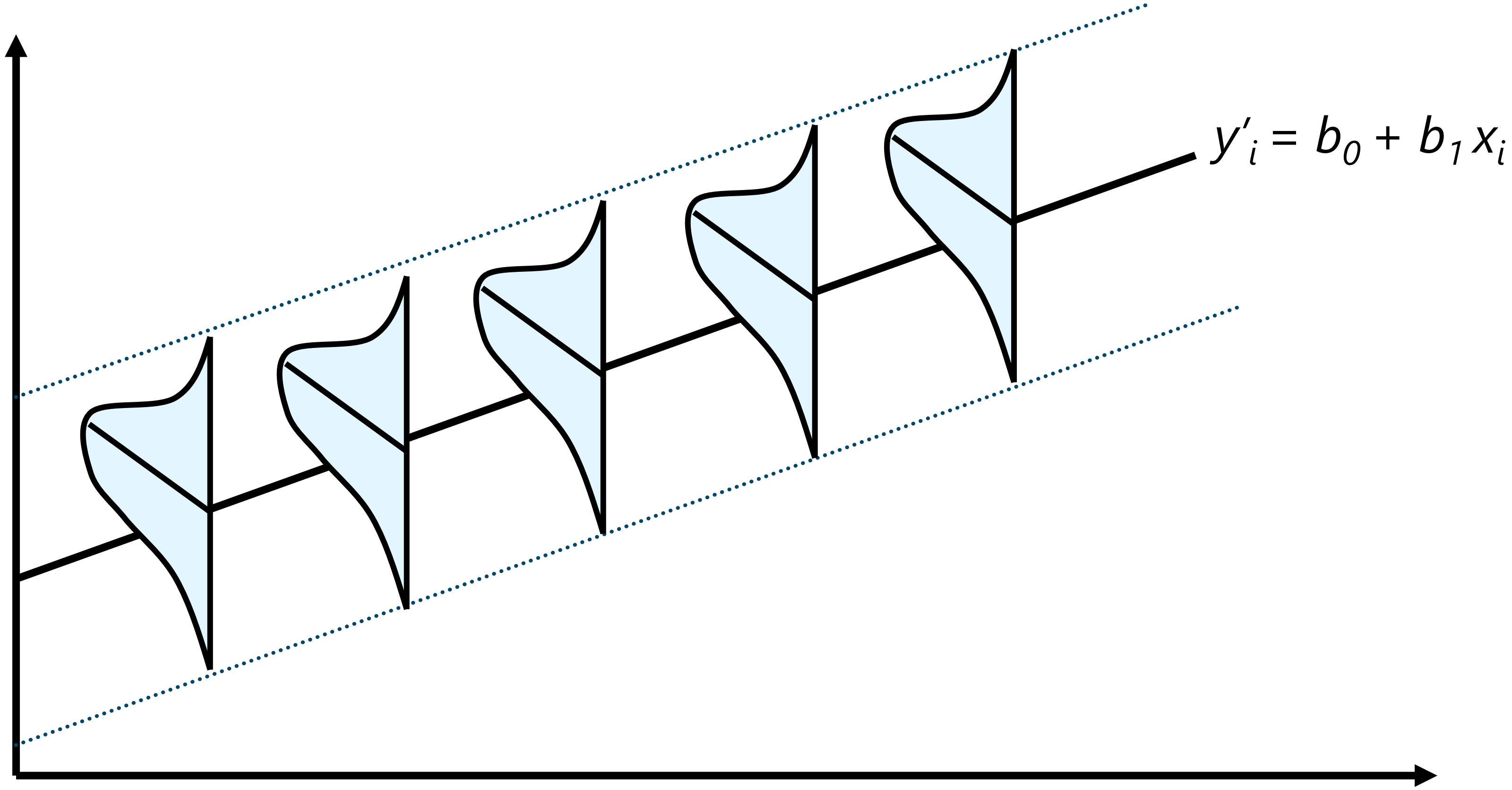
$$y'_i = b_0 + b_1\,x_i$$

$$y_i = y'_i + e_i$$

(Prikaz regresijskog pravca, originalnih i prognoziranih rezultata u kriterijskoj varijabli i rezidualnih vrijednosti)

Regresijska analiza

Jednostavna linearna regresijska analiza



(Distribucija rezidualnih vrijednosti oko regresijskog pravca)

Regresijska analiza

Jednostavna linearna regresijska analiza

Koeficijenti regresijskog pravca utvrđuju se **metodom najmanjih kvadrata**.

Metoda najmanjih kvadrata temelji se na uvjetu da je suma kvadrata rezidualnih vrijednosti minimalna

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - y'_i)^2 = \min$$

gdje je

- e_i - rezidualna vrijednost entiteta i
- y_i - rezultat entiteta i u kriterijskoj varijabli
- y'_i - prognozirani rezultat entiteta i u kriterijskoj varijabli

Regresijska analiza

Jednostavna linearna regresijska analiza

Regresijski koeficijent b$_0$ predstavlja odsječak na osi zavisne varijable y, odnosno, vrijednost zavisne varijable y ukoliko je vrijednost nezavisne varijable x = 0 .

$$b_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} {x^2}_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} {x^2}_i - (\sum_{i=1}^{n} x_i)^2}$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- x_i - rezultat entiteta *i* u prediktorskoj varijabli

Regresijska analiza

Jednostavna linearna regresijska analiza

Regresijski koeficijent b_1 određuje nagib pravca, odnosno, pokazuje koliko se u prosjeku linearno mijenja vrijednost zavisne varijable *y* za jedinični porast vrijednosti nezavisne varijable *x*.

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x^2{}_i - \left(\sum_{i=1}^{n} x_i\right)^2}$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- x_i - rezultat entiteta *i* u prediktorskoj varijabli

Regresijska analiza

Jednostavna linearna regresijska analiza

Regresijski koeficijenti se također mogu izračunati i rješavanjem regresijske jednadžbe u matričnom obliku:

$$\overbrace{\begin{vmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{vmatrix}}^{\boldsymbol{y}} = \overbrace{\begin{vmatrix} 1 & x_1 \\ . & . \\ . & . \\ . & . \\ 1 & x_n \end{vmatrix}}^{\boldsymbol{X}} \cdot \overbrace{\begin{vmatrix} b_0 \\ b_1 \end{vmatrix}}^{\boldsymbol{b}} + \overbrace{\begin{vmatrix} e_1 \\ . \\ . \\ . \\ e_n \end{vmatrix}}^{\boldsymbol{e}}$$

gdje je

- $\boldsymbol{y}$ - vektor *n* rezultata entiteta u kriteriju
- $\boldsymbol{X}$ - matrica reda $n{\cdot}2$ rezultata entiteta u prediktoru
- $\boldsymbol{b}$ - vektor regresijskih koeficijenata
- $\boldsymbol{e}$ - vektor *n* rezidualnih vrijednosti

Regresijska analiza

Jednostavna linearna regresijska analiza

$\mathbf{y} = \mathbf{X\,b} \qquad / \, \mathbf{X}^T$

$\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\,\mathbf{b} \qquad / \, (\mathbf{X}^T\mathbf{X})^{-1}$

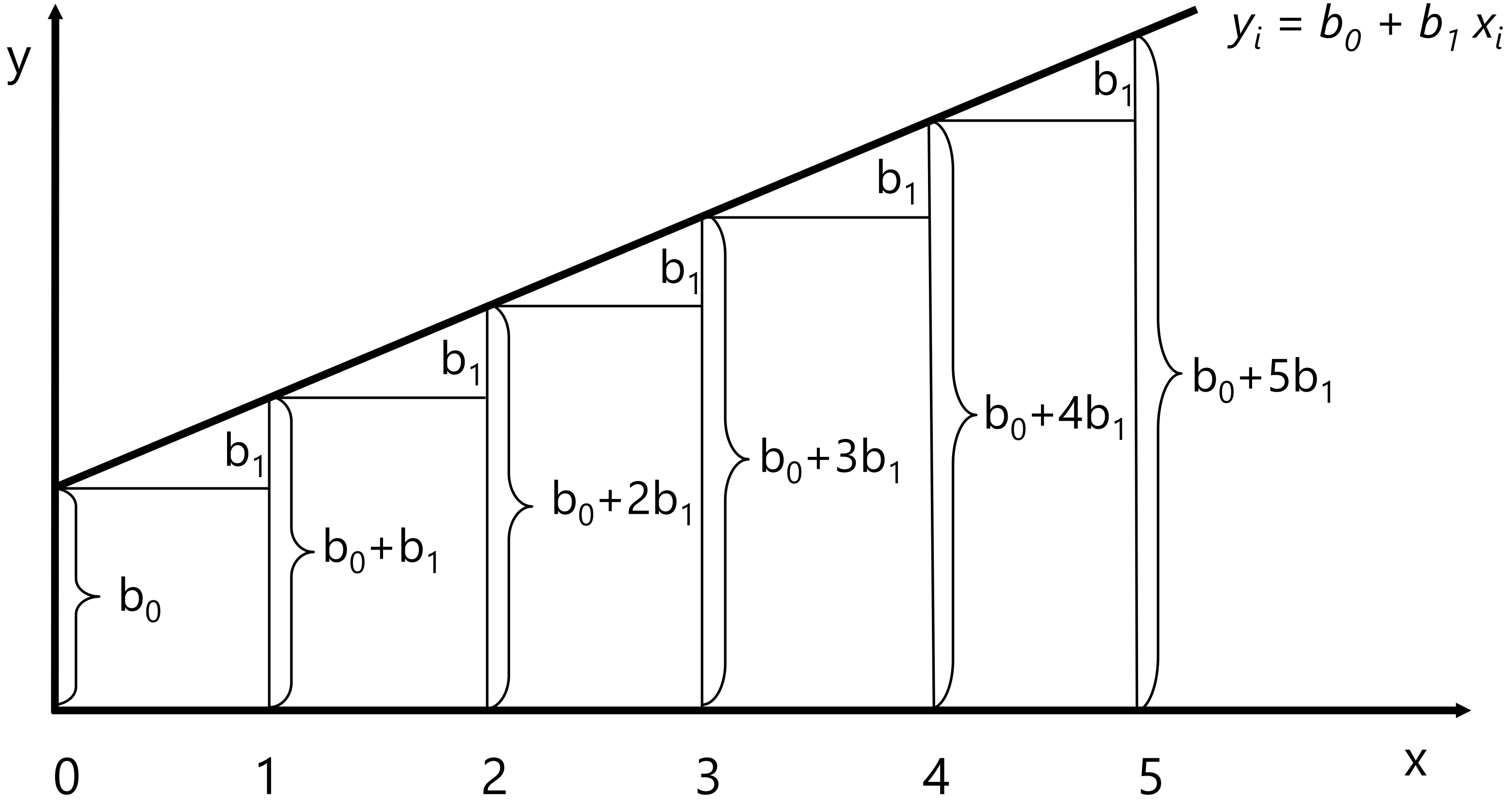
$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\mathbf{y}$

$\mathbf{y}' = \mathbf{Xb}$

$\mathbf{e} = \mathbf{y} - \mathbf{y}'$

Regresijska analiza

Jednostavna linearna regresijska analiza



(Prikaz regresijskih koeficijenata b_0 i b_1)

Regresijska analiza

Jednostavna linearna regresijska analiza

Standardna pogreška prognoze (σ_e) je drugi korijen iz varijance rezidualnih vrijednosti, a predstavlja mjeru reprezentativnosti regresijskog modela.

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y'_i)^2}{n-2}}$$

gdje je

- y_i - rezultat entiteta i u kriterijskoj varijabli
- y'_i - prognozirani rezultat entiteta i u kriterijskoj varijabli

Regresijska analiza

Jednostavna linearna regresijska analiza

Koeficijent korelacije između kriterijske i prediktorske varijable izražava veličinu njihove linearne povezanosti.

Kada je $r_{x,y} = 0$ to znači da nezavisna varijabla x nema nikakav utjecaj na varijabilitet kriterijske varijable y .

Ako koeficijent korelacije ima maksimalnu vrijednost $r_{x,y} = 1$, to znači da je cjelokupan varijabilitet varijable y moguće pripisati utjecaju varijable x.

Kvadrat koeficijenta korelacije (r^2) naziva se **koeficijent determinacije**, a predstavlja proporciju varijance kriterijske varijable koju je moguće objasniti putem prediktorske varijable.

Regresijska analiza

Višestruka linearna regresijska analiza

Višestrukom linearnom regresijskom analizom utvrđuje se linearna povezanost između dviju ili više nezavisnih (prediktorskih) i jedne zavisne (kriterijske) varijable pri čemu regresijska jednadžba ima sljedeći oblik:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_m x_{im} + e_i$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- $b_0,\ldots,b_m$ - regresijski koeficijenti
- $x_{i1},\ldots,x_{im}$ - rezultati entiteta *i* u *m* prediktorskih varijabli
- e_i - rezidualna vrijednost entiteta *i*
- $i = 1,\ldots,n$ (*n* - broj entiteta), a *m* - broj prediktora

Regresijska analiza

Višestruka linearna regresijska analiza

Regresijski koeficijenti mogu se izračunati rješavanjem regresijske jednadžbe u matričnom obliku:

$$\boldsymbol{y} = \boldsymbol{X} \boldsymbol{b} + \boldsymbol{e}$$

$$\begin{vmatrix} y_1 \\ . \\ . \\ y_n \end{vmatrix} = \begin{vmatrix} 1 & x_{11} & . & . & x_{1m} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_{n1} & . & . & x_{nm} \end{vmatrix} \cdot \begin{vmatrix} b_0 \\ b_1 \\ . \\ b_m \end{vmatrix} + \begin{vmatrix} e_1 \\ . \\ . \\ e_n \end{vmatrix}$$

Regresijska analiza

Višestruka linearna regresijska analiza

$$\mathbf{y} = \mathbf{X}\,\mathbf{b} \qquad / \; \mathbf{X}^{T}$$

$$\mathbf{X}^{T}\mathbf{y} = \mathbf{X}^{T}\mathbf{X}\mathbf{b} \qquad / \; (\mathbf{X}^{T}\mathbf{X})^{-1}$$

$$\mathbf{b} = (\mathbf{X}^{T}\mathbf{X})^{-1}\,\mathbf{X}^{T}\mathbf{y}$$

$$\mathbf{y}' = \mathbf{X}\mathbf{b}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{y}'$$

Regresijska analiza

Višestruka linearna regresijska analiza

Regresijski koeficijent b_0 predstavlja vrijednost zavisne varijable y ukoliko je vrijednost svih nezavisnih varijabli jednaka 0 .

Regresijski koeficijenti $b_1,\ldots,b_m$ pokazuju koliko se u prosjeku linearno mijenja vrijednost zavisne varijable y za jedinični porast vrijednosti odgovarajuće nezavisne varijable ($x_1,\ldots,x_m$) uz uvjet da su vrijednosti ostalih nezavisnih varijabli konstantne.

Regresijska analiza

Višestruka linearna regresijska analiza

Ako se kriterijska i prediktorske varijable prethodno standardiziraju regresijska jednadžba poprima sljedeći oblik:

$$k_i = \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_m z_{im} + \varepsilon_i$$

gdje je

- k_i - standardizirani rezultat entiteta i u kriterijskoj varijabli
- $\beta_1,\ldots,\beta_m$ - standardizirani regresijski koeficijenti
- $z_{i1},\ldots,z_{im}$ - standardizirani rezultati entiteta i u m prediktorskih varijabli
- ε_i - standardizirana rezidualna vrijednost entiteta i
- $i = 1,\ldots,n$ (n - broj entiteta), a m - broj prediktora

Regresijska analiza

Višestruka linearna regresijska analiza

Standardizirani regresijski koeficijenti mogu se izračunati rješavanjem sljedeće jednadžbe u matričnom obliku:

$$k \quad = \qquad Z \qquad \beta \; + \; \varepsilon$$

$$\begin{vmatrix} k_1 \\ . \\ . \\ k_n \end{vmatrix} = \begin{vmatrix} z_{11} & . & . & z_{1m} \\ . & . & . & . \\ . & . & . & . \\ z_{n1} & . & . & z_{nm} \end{vmatrix} \cdot \begin{vmatrix} \beta_1 \\ . \\ . \\ \beta_m \end{vmatrix} + \begin{vmatrix} \varepsilon_1 \\ . \\ . \\ \varepsilon_n \end{vmatrix}$$

Regresijska analiza

Višestruka linearna regresijska analiza

$$\mathbf{k} = \mathbf{Z}\,\beta \qquad / \, \mathbf{Z}^{T} n^{-1}$$

$$\mathbf{Z}^{T}\mathbf{k}n^{-1} = \mathbf{Z}^{T}\mathbf{Z}n^{-1}\beta$$

$$\mathbf{r} = \mathbf{R}\,\beta \qquad / \, \mathbf{R}^{-1}$$

$$\beta = \mathbf{R}^{-1}\,\mathbf{r}$$

gdje je

- $\mathbf{k}$ - vektor n standardiziranih rezultata entiteta u kriteriju
- $\mathbf{Z}$ - matrica reda $n \cdot m$ standardiziranih rezultata entiteta u m prediktora
- $\boldsymbol{\beta}$ - vektor standardiziranih regresijskih koeficijenata
- $\mathbf{r}$ - vektor korelacija m prediktora s kriterijem
- $\mathbf{R}$ - matrica međusobnih korelacija m prediktora

Regresijska analiza

Višestruka linearna regresijska analiza

Standardizirani regresijski koeficijenti $\beta_1,\ldots,\beta_m$ su relativni koeficijenti utjecaja, a predstavljaju veličinu promjene zavisne varijable izraženu u dijelovima standardne devijacije za jedinični porast standardizirane vrijednosti odgovarajuće nezavisne varijable ($z_1,\ldots,z_m$) uz uvjet da su vrijednosti preostalih nezavisnih varijabli konstantne.

Statistička značajnost svakog pojedinog regresijskog koeficijenta se testira putem Studentove t-distribucije.

Pri tome je za svaki regresijski koeficijent moguće postaviti sljedeću alternativnu (H1), odnosno nultu (H0) hipotezu:

- H0: Uz pogrešku p ne možemo tvrditi da je utjecaj prediktora j na kriterijsku varijablu statistički značajan.
- H1: Utjecaj prediktora j na kriterijsku varijablu je statistički značajan uz pogrešku p.

Regresijska analiza

Višestruka linearna regresijska analiza

Standardna pogreška prognoze (σ_e) je drugi korijen iz varijance rezidualnih vrijednosti, a predstavlja mjeru reprezentativnosti regresijskog modela.

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y'_i)^2}{n - (m + 1)}}$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- y'_i - prognozirani rezultat entiteta *i* u kriterijskoj varijabli
- $i = 1,\ldots,n$
- n - broj entiteta, m - broj prediktorskih varijabli

Regresijska analiza

Višestruka linearna regresijska analiza

Koeficijent multiple korelacije (ρ) je korelacija između kriterijske varijable i varijable prognoziranih rezultata, a izražava veličinu linearne povezanosti skupa prediktorskih varijabli s kriterijem.

Koeficijent multiple korelacije se kreće u intervalu od *0* do *1* pri čemu *0* označava nikakavu, a *1* potpunu zavisnost kriterijske varijable o skupu prediktorskih varijabli.

Kvadrat koeficijenta korelacije (ρ^2) naziva se **koeficijent multiple determinacije**, a predstavlja proporciju varijance kriterijske varijable koju je moguće objasniti putem skupa prediktorskih varijabli.

Regresijska analiza

Višestruka linearna regresijska analiza

Statistička značajnost koeficijenta multiple korelacije (ρ) se testira putem Snedecorove F-distribucije.

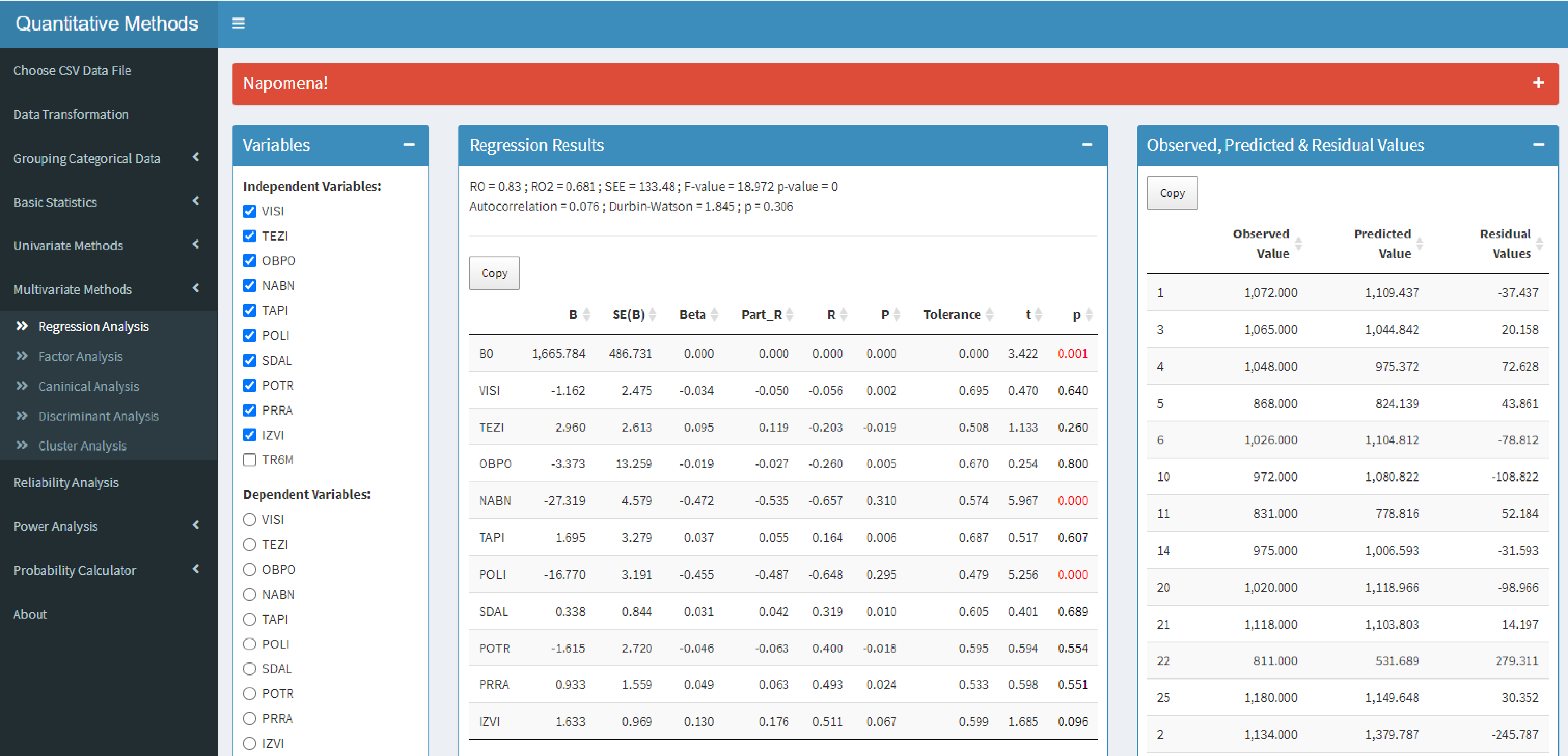
Pri tome je moguće postaviti sljedeću alternativnu (H1), odnosno nultu (H0) hipotezu:

- H0: $\rho = 0$ - Uz pogrešku *p* ne možemo tvrditi da je povezanost između skupa prediktora i kriterijske varijable statistički značajna.
- H1: $\rho \neq 0$ - Povezanost između skupa prediktora i kriterijske varijable je statistički značajna uz pogrešku *p*.

RStudio – Regression Analysis

Zadatak: U datoteci *SKOLA.csv* utvrdite relacije između morfoloških obilježja (VISI, TEZI, OBPO, NABN), testova za procjenu motoričkih sposobnosti (TAPI, POLI, SDAL, POTR, PRRA) i testa *trčanje 6 minuta* (TR6M).

RStudio – Regression Analysis



Independent Variables:

☑ VISI
☑ TEZI
☑ OBPO
☑ NABN
☑ TAPI
☑ POLI
☑ SDAL
☑ POTR
☑ PRRA
☑ IZVI
☐ TR6M

Dependent Variables:

○ VISI
○ TEZI
○ OBPO
○ NABN
○ TAPI
○ POLI
○ SDAL
○ POTR
○ PRRA
○ IZVI

RO = 0.83 ; RO2 = 0.681 ; SEE = 133.48 ; F-value = 18.972 p-value = 0
Autocorrelation = 0.076 ; Durbin-Watson = 1.845 ; p = 0.306

	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

	Observed Value	Predicted Value	Residual Values
1	1,072.000	1,109.437	-37.437
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
10	972.000	1,080.822	-108.822
11	831.000	778.816	52.184
14	975.000	1,006.593	-31.593
20	1,020.000	1,118.966	-98.966
21	1,118.000	1,103.803	14.197
22	811.000	531.689	279.311
25	1,180.000	1,149.648	30.352
2	1,134.000	1,379.787	-245.787

FAKTORSKA ANALIZA

Vježba 12

Faktorska analiza

Faktorska analiza je zajedničko ime za više metoda kojima je cilj kondenzacija većeg broja **manifestnih varijabli**, među kojima postoji povezanost (korelacija), na manji broj **latentnih dimenzija (faktora)** koje su izvor te povezanosti.

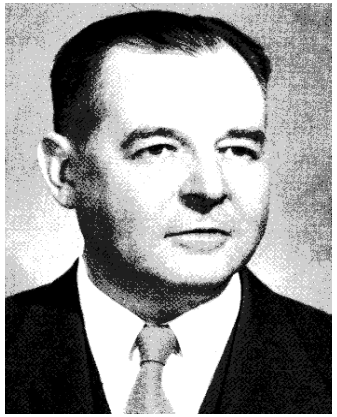
Manifestne varijable su varijable, odnosno obilježja koja se mogu izravno mjeriti postojećim mjernim instrumentima (npr. skok udalj s mjesta, bacanje medicinke iz ležanja na leđima, bacanje kugle).

Latentne dimenzije su varijable, odnosno obilježja koja nisu izravno mjerljiva postojećim mjernim instrumentima, već se izračunavaju kao linearna kombinacija manifestnih varijabli (npr. eksplozivna snaga).

Faktorska analiza

Komponentni model faktorske analize

Komponentni model ili **metoda glavnih komponenata** je metoda ekstrakcije, odnosno izračunavanja latentnih dimenzija koju je 1933. godine predložio američki ekonomist i statističar **Harold Hotelling.**



Harold Hotelling
(1895. – 1973.)

Metodom glavnih komponenata se iz skupa od *m* manifestnih varijabli na temelju nereducirane korelacijske matrice izračuna *m* latentnih dimenzija koje su međusobno linearno nezavisne, a nazivaju se **glavne komponente**.

Glavne komponente su linearne kombinacije manifestnih varijabli izračunate na način da prva glavna komponenta objašnjava maksimalan moguć dio ukupne varijance manifestnih varijabli te da druga, kao i svaka sljedeća glavna komponenta, objašnjava najveći dio preostale varijance manifestnih varijabli, odnosno najveći dio varijance manifestnih varijabli koji nije objašnjen prethodnim glavnim komponentama.

Faktorska analiza

Komponentni model faktorske analize

Glavne komponente izračunavaju se na sljedeći način:

Neka su u matrici B podaci skupa $E = \{e_i \,;\, i = 1,\ldots,n\}$ entiteta koji su opisani skupom $V = \{v_j \,;\, j = 1,\ldots,m\}$ manifestnih varijabli. Operacijom

$$\boldsymbol{Z} = \boldsymbol{B_c} \, \boldsymbol{V}^{-1}$$

gdje je

- $\boldsymbol{B_c}$ - matrica centriranih podataka dobivenih operacijom $\boldsymbol{B_c} = \boldsymbol{B}^{-1}\boldsymbol{m}^T$
- $\boldsymbol{m}$ - vektor aritmetičkih sredina varijabli matrice $\boldsymbol{B}$
- $\boldsymbol{V}^{-1}$ - dijagonalna matrica standardnih devijacija varijabli iz matrice $\boldsymbol{B}$

izračuna se matrica standardiziranih podataka ($\boldsymbol{Z}$).

Faktorska analiza

Komponentni model faktorske analize

Operacijom

$$\boldsymbol{R} = \boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{Z}n^{-1} = \begin{vmatrix} 1 & r_{12} & . & . & r_{1m} \\ r_{21} & 1 & . & . & r_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ r_{m1} & r_{m2} & . & . & 1 \end{vmatrix}$$

izračuna se matrica korelacija manifestnih varijabli (***R***) u čijoj su glavnoj dijagonali jedinice, odnosno varijance standardiziranih varijabli. Zbroj varijanci svih standardiziranih varijabli, odnosno ukupna varijanca skupa standardiziranih varijabli jednaka je *m*, odnosno broju manifestnih varijabli.

Faktorska analiza

Komponentni model faktorske analize

Operacijom spektralne dekompozicije matrice korelacija

$$\boldsymbol{R} \qquad = \qquad \boldsymbol{X} \qquad \qquad \boldsymbol{\lambda} \qquad \qquad \boldsymbol{X}^{T}$$

$$\begin{vmatrix} 1 & \mathrm{r}_{12} & . & . & \mathrm{r}_{1\mathrm{m}} \\ \mathrm{r}_{21} & 1 & . & . & \mathrm{r}_{2\mathrm{m}} \\ . & . & . & . & . \\ . & . & . & . & . \\ \mathrm{r}_{\mathrm{m}1} & \mathrm{r}_{\mathrm{m}2} & . & . & 1 \end{vmatrix} = \begin{vmatrix} x_{11} & x_{12} & . & . & x_{1m} \\ x_{21} & x_{22} & . & . & x_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{m1} & x_{m2} & . & . & x_{mm} \end{vmatrix} \cdot \begin{vmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & . & & \\ & & & . & \\ & & & & \lambda_m \end{vmatrix} \cdot \begin{vmatrix} x_{11} & x_{21} & . & . & x_{m1} \\ x_{12} & x_{22} & . & . & x_{m2} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{1m} & x_{2m} & . & . & x_{mm} \end{vmatrix}$$

izračunaju se **matrica svojstvenih vektora** (X) i **matrica svojstvenih vrijednosti** (λ).

Faktorska analiza

Komponentni model faktorske analize

Matrica svojstvenih vektora (**X**) je kvadratna matrica reda $m \times m$ za koju vrijedi da je $\mathbf{X}^T\mathbf{X} = \mathbf{X}\mathbf{X}^T = \mathbf{I}$ tj. da su vektori matrice linearno nezavisni, a u čijim su stupcima ponderi za izračunavanje glavnih komponenata iz standardiziranih varijabli.

Matrica svojstvenih vrijednosti (**λ**) je dijagonalna matrica reda m×m za koju vrijedi da je $\lambda_1 + \lambda_2 + \ldots + \lambda_m = m$, tj. da je zbroj svojstvenih vrijednosti jednak ukupnoj varijanci standardiziranih varijabli, te da je $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$, tj. da se svojstvene vrijednosti nižu od najveće prema najmanjoj.

Faktorska analiza

Komponentni model faktorske analize

Operacijom

$$\mathbf{K} \quad = \quad \mathbf{Z} \quad \mathbf{X}$$

$$\begin{vmatrix} k_{11} & k_{12} & . & . & k_{1m} \\ k_{21} & k_{22} & . & . & k_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ k_{n1} & k_{n2} & . & . & k_{nm} \end{vmatrix} = \begin{vmatrix} z_{11} & z_{12} & . & . & z_{1m} \\ z_{21} & z_{22} & . & . & z_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ z_{n1} & z_{n2} & . & . & z_{nm} \end{vmatrix} \cdot \begin{vmatrix} x_{11} & x_{12} & . & . & x_{1m} \\ x_{21} & x_{22} & . & . & x_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{m1} & x_{m2} & . & . & x_{mm} \end{vmatrix}$$

izračuna se **matrica glavnih komponenata** (**K**).

Faktorska analiza

Komponentni model faktorske analize

Matrica glavnih komponenata (**K**) je matrica reda n×m koju čine rezultati n ispitanika u m linearno nezavisnih (ortogonalnih) glavnih komponenata. Varijanca prve glavne komponente jednaka je λ_1, varijanca druge glavne komponente λ_2, itd.

Pošto vrijedi da je $\lambda_1 + \lambda_2 + \ldots + \lambda_m = m$ te da je $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$, može se zaključiti da je ukupna varijanca m standardiziranih varijabli raspodijeljena tako da prva glavna komponenta objašnjava najveći dio ukupne varijance, druga glavna komponenta najveći dio preostale varijance, itd.

Faktorska analiza

Komponentni model faktorske analize

Operacijom

$$\mathbf{H} = \mathbf{X} \quad \boldsymbol{\lambda}^{1/2}$$

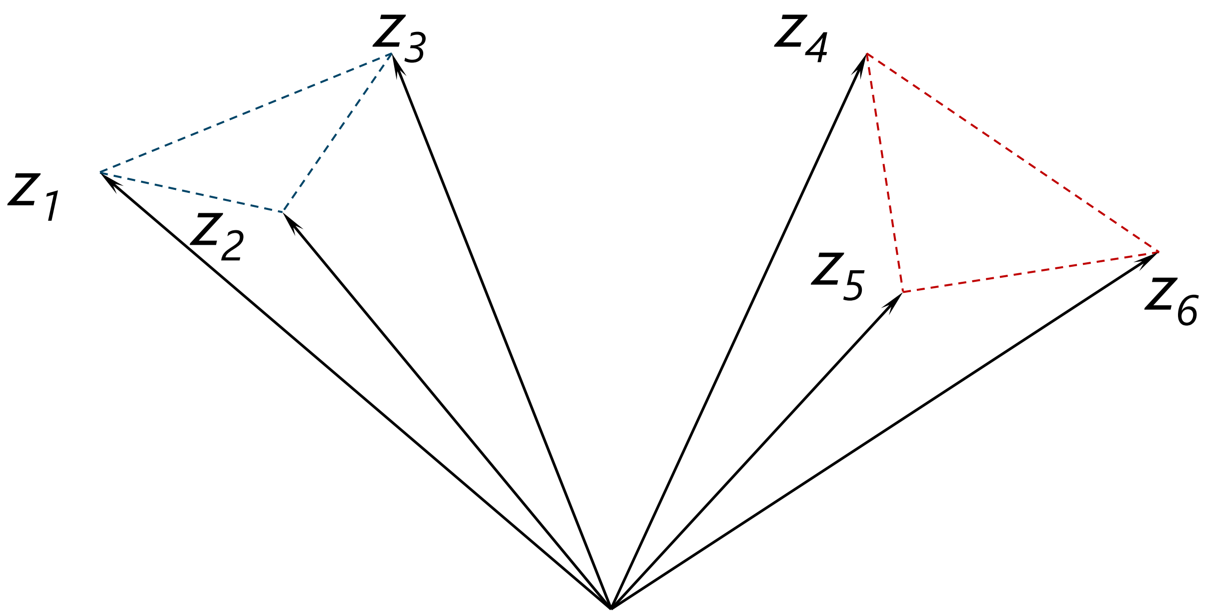
$$\begin{vmatrix} h_{11} & h_{12} & . & . & h_{1m} \\ h_{21} & h_{22} & . & . & h_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ h_{m1} & h_{m2} & . & . & h_{mm} \end{vmatrix} = \begin{vmatrix} x_{11} & x_{12} & . & . & x_{1m} \\ x_{21} & x_{22} & . & . & x_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{m1} & x_{m2} & . & . & x_{mm} \end{vmatrix} \cdot \begin{vmatrix} \sqrt{\lambda_1} & & & & \\ & \sqrt{\lambda_2} & & & \\ & & . & & \\ & & & . & \\ & & & & \sqrt{\lambda_m} \end{vmatrix}$$

izračuna se **matrica glavnih osovina** (H). Matrica glavnih osovina je matrica korelacija manifestnih varijabli i glavnih komponenata. Elementi ove matrice ukazuju koliki je doprinos svake pojedine manifestne varijable pri formiranju svake pojedine glavne komponente.

Faktorska analiza

Komponentni model faktorske analize

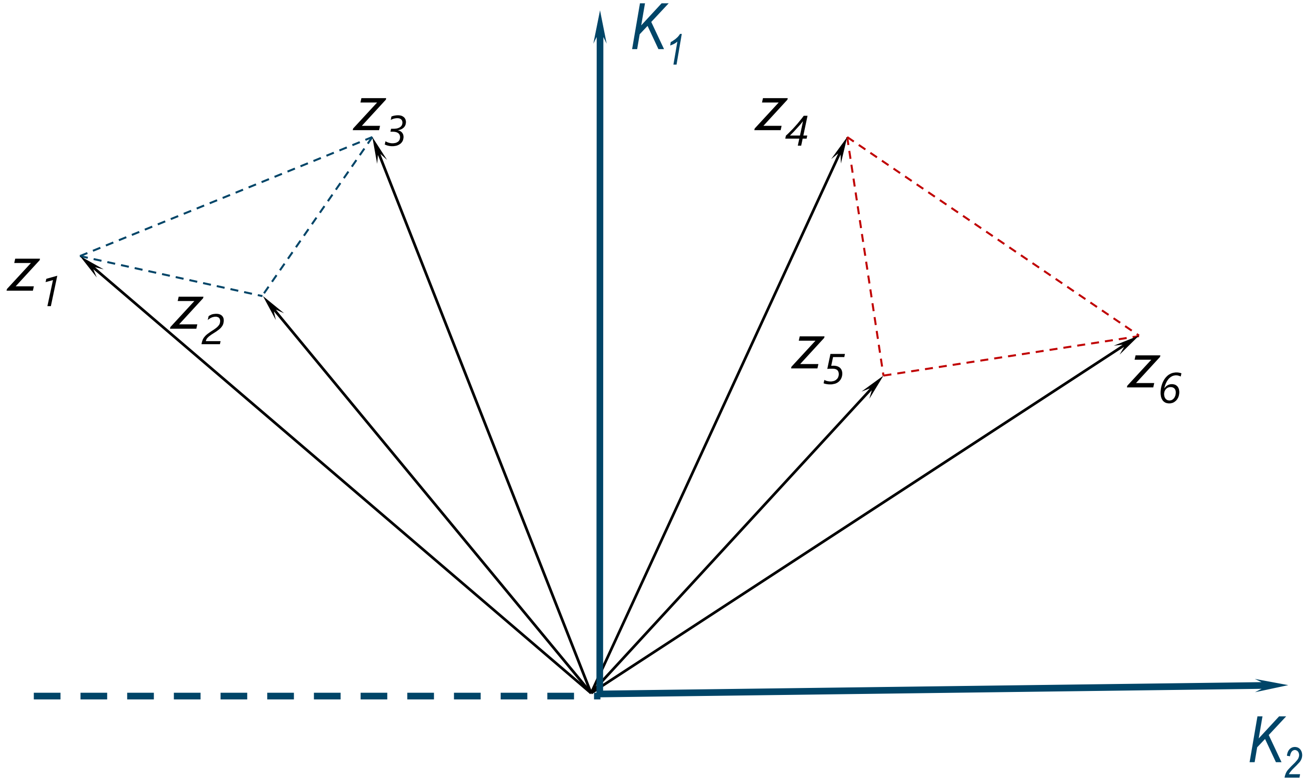
Neka je skup od *n* entiteta opisan sa 6 manifestnih varijabli ($v_1,\ldots,v_6$). Na temelju korelacijske matrice može se utvrditi odnos odgovarajućih standardiziranih vektora u *n*-dimenzionalnom prostoru.



Faktorska analiza

Komponentni model faktorske analize

Na temelju matrice glavnih osovina može se utvrditi prostorni odnos standardiziranih manifestnih vektora i standardiziranih glavnih komponenata.



Faktorska analiza

Kriteriji za odabir značajnih glavnih komponenata

Pošto je osnovni cilj faktorske analize kondenzacija većeg broja manifestnih varijabli na manji broj latentnih dimenzija, potrebno je izvršiti **redukciju broja glavnih komponenata**.

Redukcija broja glavnih komponenata odnosno određivanje broja značajnih glavnih komponenata vrši se putem različitih kriterija kao što su **GK-kriterij**, **PB-kriterij** i **Scree-test**.

Prema **GK-kriteriju** glavna komponenta j je značajna ako je njena varijanca (λ_j) veća ili jednaka 1.

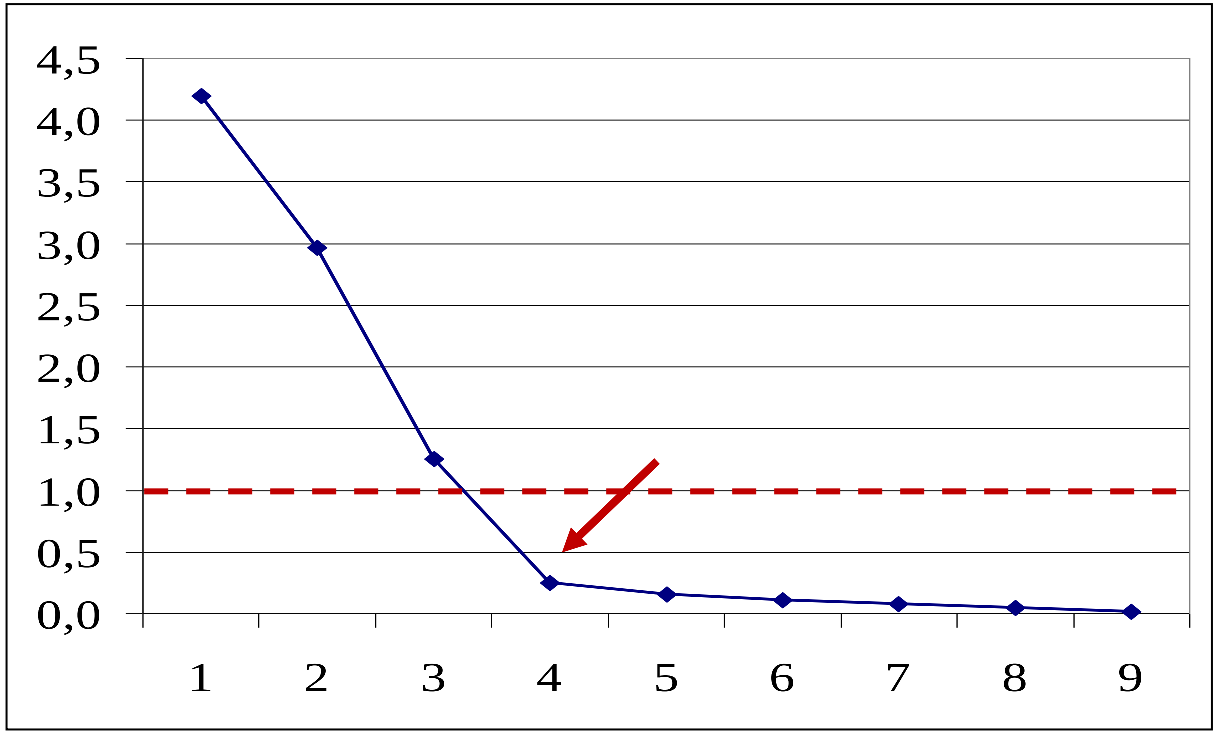
Prema **PB-kriteriju** broj značajnih glavnih komponenata jednak je broju svojstvenih vrijednosti poredanih po veličini čiji zbroj ne prelazi Σsmc (sumu kvadrata multiplih korelacija).

Scree-test je grafički kriterij. 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.

Faktorska analiza

Kriteriji za odabir značajnih glavnih komponenata

Scree plot



GK - kriterij

Faktorska analiza

Kriteriji za odabir značajnih glavnih komponenata

Komunaliteti i unikviteti

$$\mathbf{H} = \left| \begin{array}{cccc:cccc}
\multicolumn{4}{c}{\overbrace{\text{značajne glavne komponente}}} & \multicolumn{4}{c}{\overbrace{\text{glavne komponente koje nisu značajne}}} \\
h_{11} & . & . & h_{1k} & . & . & . & h_{1m} \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
h_{m1} & . & . & h_{mk} & . & . & . & h_{mm}
\end{array} \right|$$

(Matrica **H** prije i nakon redukcije broja glavnih komponenata)

Faktorska analiza

Kriteriji za odabir značajnih glavnih komponenata

Komunaliteti i unikviteti

Varijancu svake manifestne varijable moguće je dekomponirati na **komunalitet** (h^2) i **unikvitet** (u^2) pri čemu je

$$s_j^2 = 1 = h_j^2 + u_j^2$$

Komunalitet je dio varijance manifestne varijable j koji je moguće objasniti s k značajnih glavnih komponenata.

$$h_j^2 = \sum_{p=1}^{k} h_{jp}^2 = h_{j1}^2 + h_{j2}^2 + \ldots + h_{jk}^2$$

Unikvitet je dio varijance manifestne varijable j koji nije moguće objasniti s k značajnih glavnih komponenata

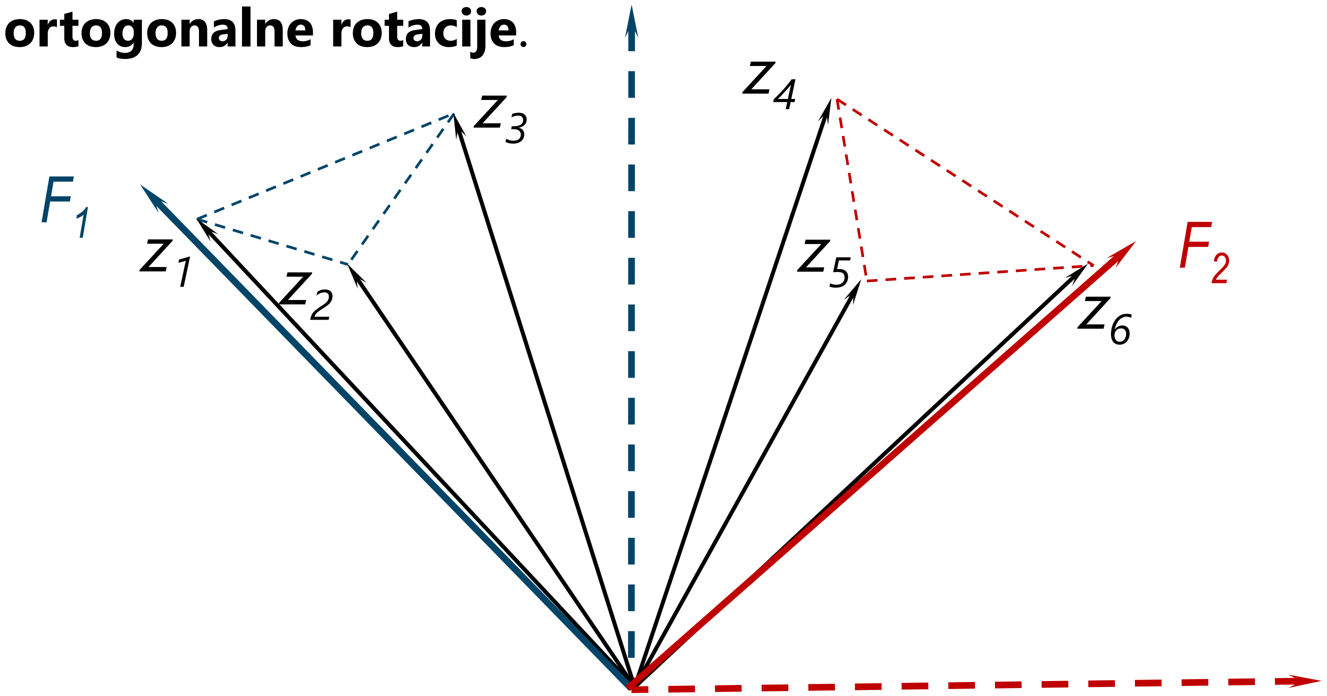
$$u_j^2 = 1 - h_j^2$$

Faktorska analiza

Rotacije

Pravi uvid u strukturu međusobnih odnosa manifestnih varijabli često nije moguće steći putem glavnih komponenata. U takvim se slučajevima koriste transformacije glavnih komponenata čija je svrha postizanje jednostavne faktorske strukture, a koje se nazivaju **rotacije**.

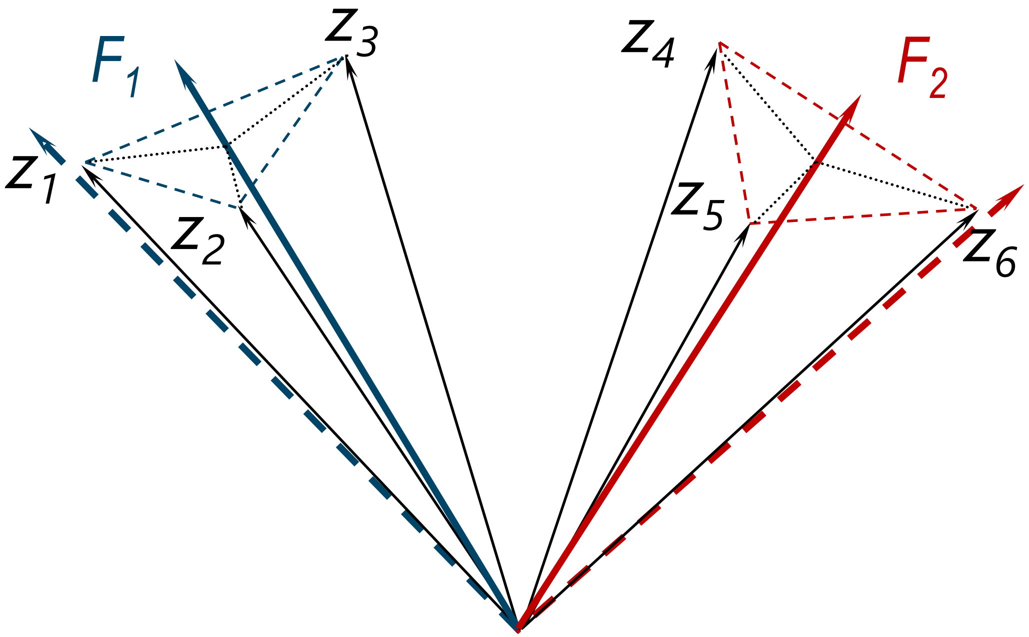
Transformacije glavnih komponenata koje se provode uz uvjet zadržavanja njihove linearne nezavisnosti nazivaju se **ortogonalne rotacije**.



Faktorska analiza

Rotacije

Transformacije glavnih komponenata koje se provode bez uvjeta zadržavanja njihove linearne nezavisnosti nazivaju se **neortogonalne** ili **kosokutne rotacije**.



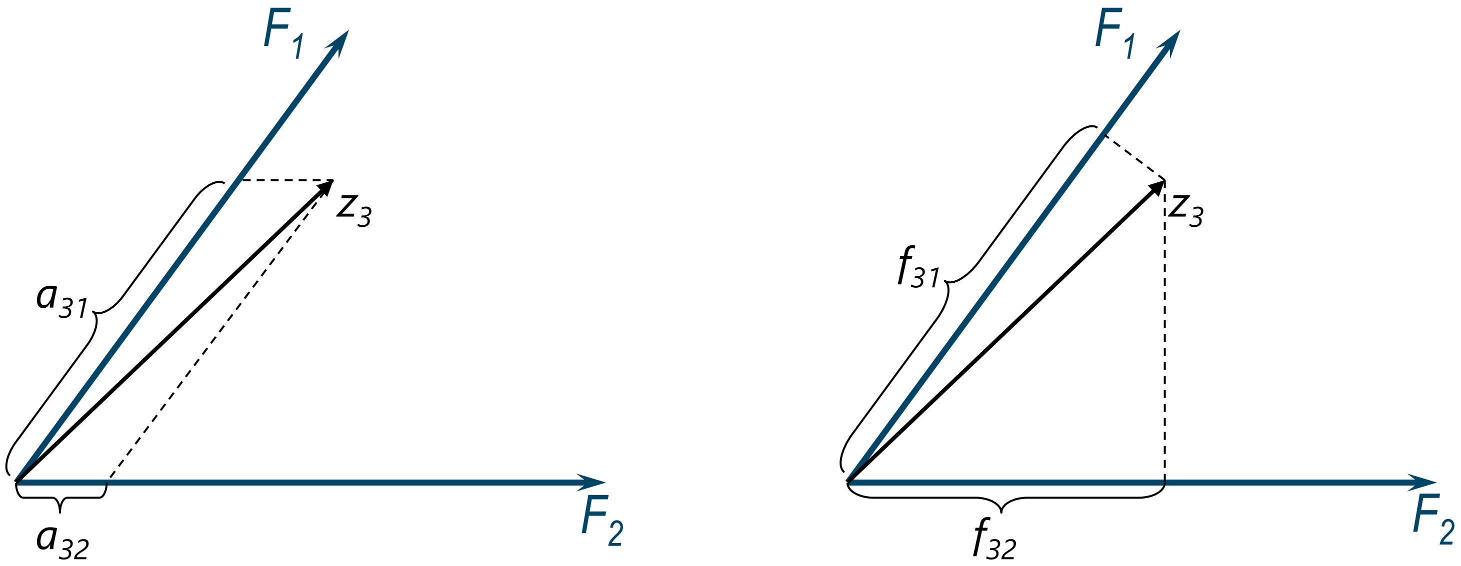
Faktorska analiza

Rotacije

Interpretacija faktora nakon rotacije vrši se putem **matrice strukture** (**F**), **matrice sklopa** (**A**), i **matrice korelacija među faktorima** (**M**).

Matricu strukture čine korelacije manifestnih varijabli s faktorima, a **matricu sklopa** čine paralelne projekcije manifestnih varijabli na faktore.

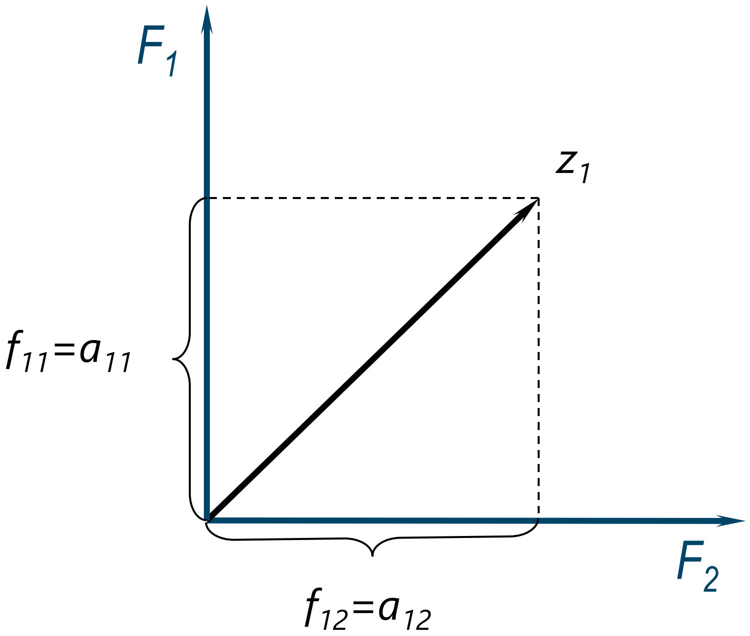
Matrica sklopa često jasnije pokazuje koje varijable određuju pojedine faktore nego matrica strukture.



Faktorska analiza

Rotacije

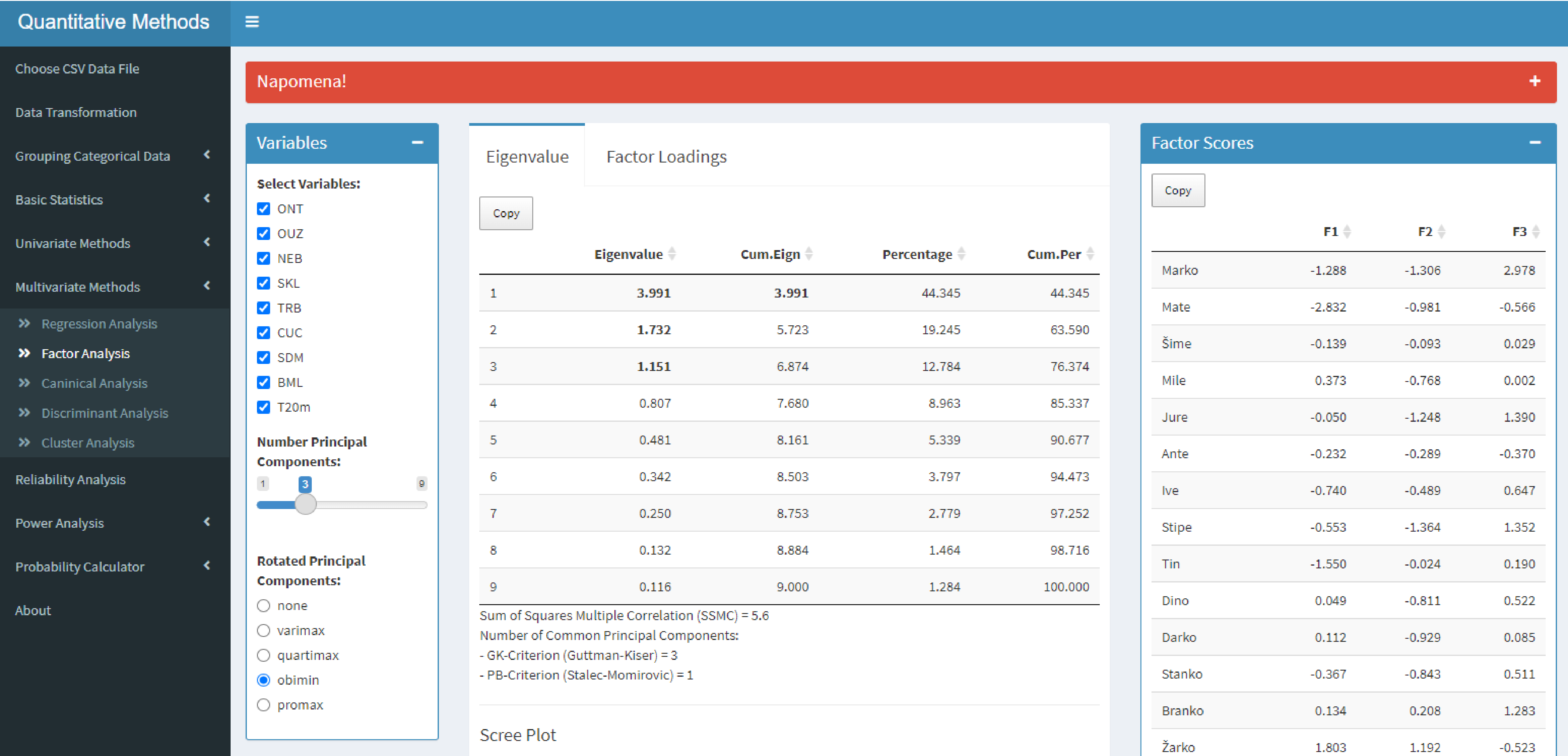
Paralelne projekcije i ortogonalne projekcije manifestnih varijabli na faktore bit će sličnije što su korelacije među faktorima manje. Matrica strukture bit će jednaka matrici sklopa ako su faktori potpuno nezavisni.



RStudio – Factor Analysis

Zadatak: Na varijablama matrice *JUDO3F.csv* provedite komponentni model faktorske analize uz *GK-kriterij* redukcije i neortogonalnu *oblimin* rotaciju! Izračunajte matricu faktorske strukture i matricu rezultata entiteta na faktorima!

RStudio – Factor Analysis



Variables

Select Variables:

- ☑ ONT
- ☑ OUZ
- ☑ NEB
- ☑ SKL
- ☑ TRB
- ☑ CUC
- ☑ SDM
- ☑ BML
- ☑ T20m

Number Principal Components: 1 — 3 — 9

Rotated Principal Components:

- ○ none
- ○ varimax
- ○ quartimax
- ● obimin
- ○ promax

Eigenvalue | Factor Loadings

	Eigenvalue	Cum.Eign	Percentage	Cum.Per
1	**3.991**	**3.991**	44.345	44.345
2	**1.732**	5.723	19.245	63.590
3	**1.151**	6.874	12.784	76.374
4	0.807	7.680	8.963	85.337
5	0.481	8.161	5.339	90.677
6	0.342	8.503	3.797	94.473
7	0.250	8.753	2.779	97.252
8	0.132	8.884	1.464	98.716
9	0.116	9.000	1.284	100.000

Sum of Squares Multiple Correlation (SSMC) = 5.6
Number of Common Principal Components:
- GK-Criterion (Guttman-Kiser) = 3
- PB-Criterion (Stalec-Momirovic) = 1

Factor Scores

	F1	F2	F3
Marko	-1.288	-1.306	2.978
Mate	-2.832	-0.981	-0.566
Šime	-0.139	-0.093	0.029
Mile	0.373	-0.768	0.002
Jure	-0.050	-1.248	1.390
Ante	-0.232	-0.289	-0.370
Ive	-0.740	-0.489	0.647
Stipe	-0.553	-1.364	1.352
Tin	-1.550	-0.024	0.190
Dino	0.049	-0.811	0.522
Darko	0.112	-0.929	0.085
Stanko	-0.367	-0.843	0.511
Branko	0.134	0.208	1.283
Žarko	1.803	1.192	-0.523

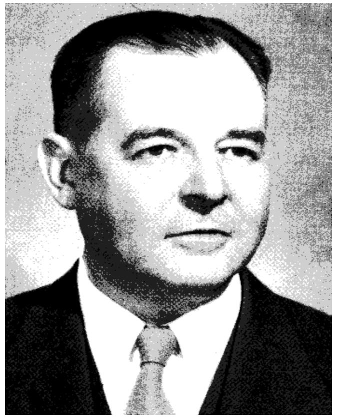
KANONIČKA ANALIZA

Vježba 13

Kanonička analiza

Kanoničku analizu je 1936. godine u časopisu Biometrika predložio Harold Hotelling.

Kanonička analiza je metoda za utvrđivanje relacija između dva skupa varijabli (npr. relacije morfoloških karakteristika i motoričkih sposobnosti).



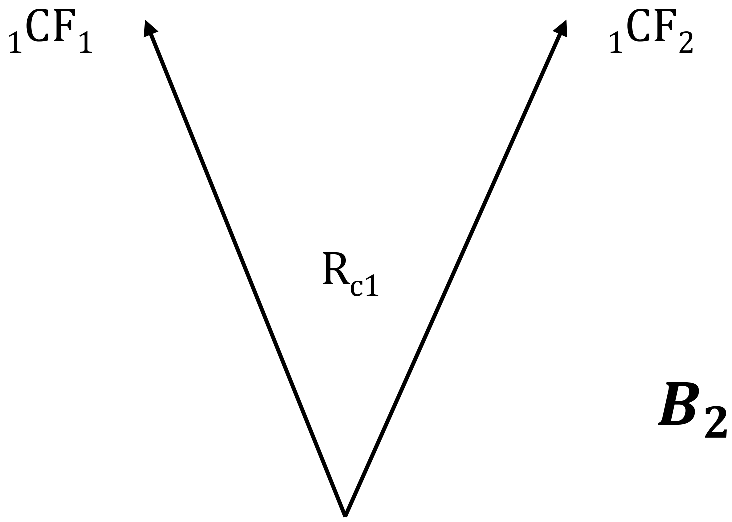
Harold Hotelling
(1895. – 1973.)

$$\boldsymbol{B_1} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} \longleftrightarrow \boldsymbol{B_2} = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

Kanonička analiza

Relacije se utvrđuju pomoću **parova kanoničkih faktora**.

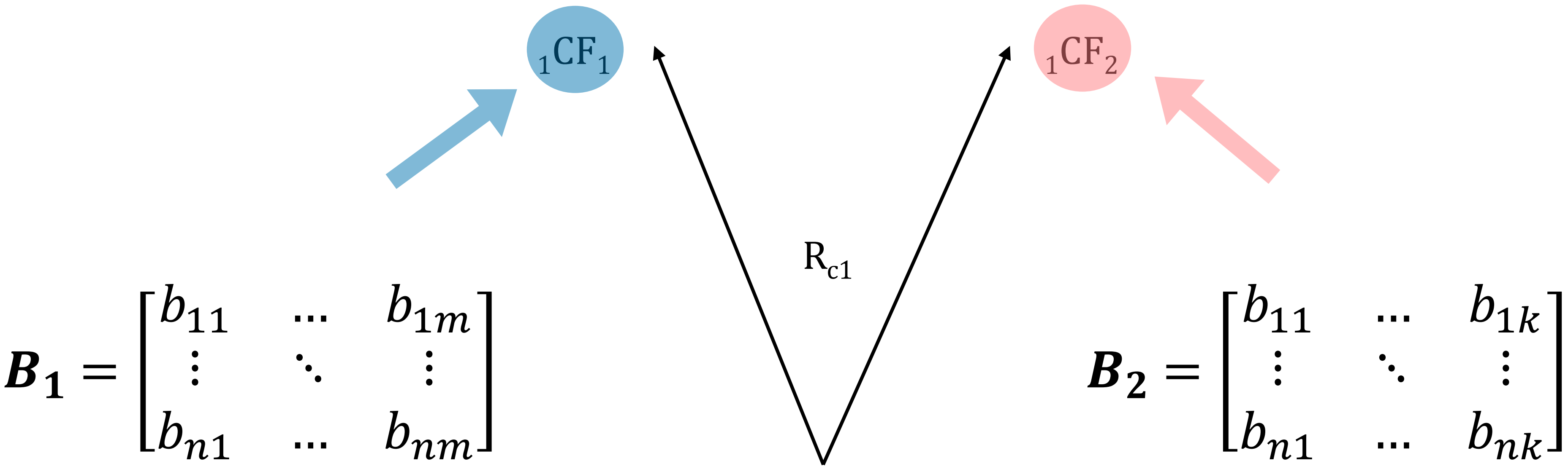
$$\boldsymbol{B_1} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix}$$



$$\boldsymbol{B_2} = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

Kanonička analiza

Svaki par kanoničkih faktora čine jedan faktor iz prvog i jedan faktor iz drugog skupa manifestnih varijabli. Broj parova kanoničkih faktora jednak je broju manifestnih varijabli u manjem skupu.



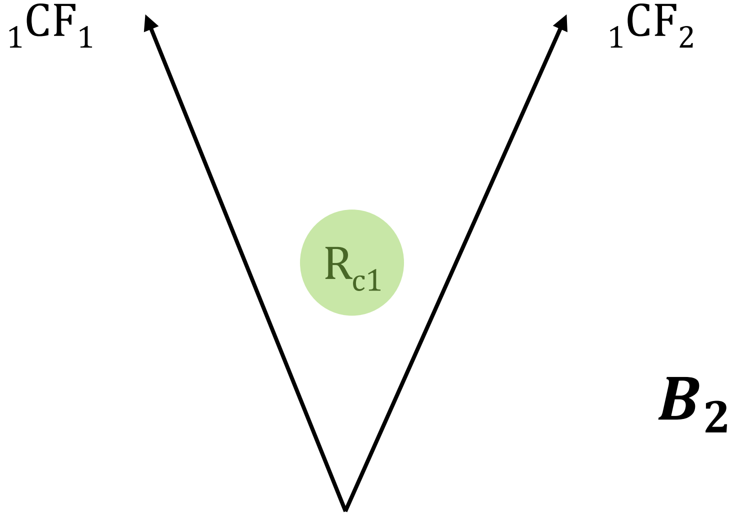
$$\boldsymbol{B_1} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix}$$

$$\boldsymbol{B_2} = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

Kanonička analiza

Koeficijent kanoničke korelacije predstavlja mjeru međusobne povezanosti (korelacije) između kanoničkih faktora koji čine jedan par.

$$\boldsymbol{B_1} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix}$$

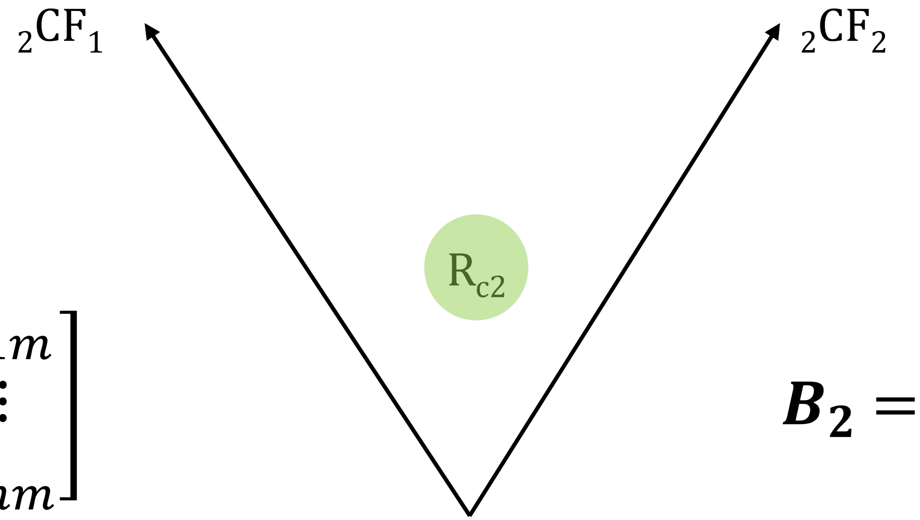


$$\boldsymbol{B_2} = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

Kanonička analiza

Prvi par kanoničkih faktora je u najvećoj korelaciji, dok je svaki sljedeći u jednakoj ili manjoj korelaciji u odnosu na prethodni.

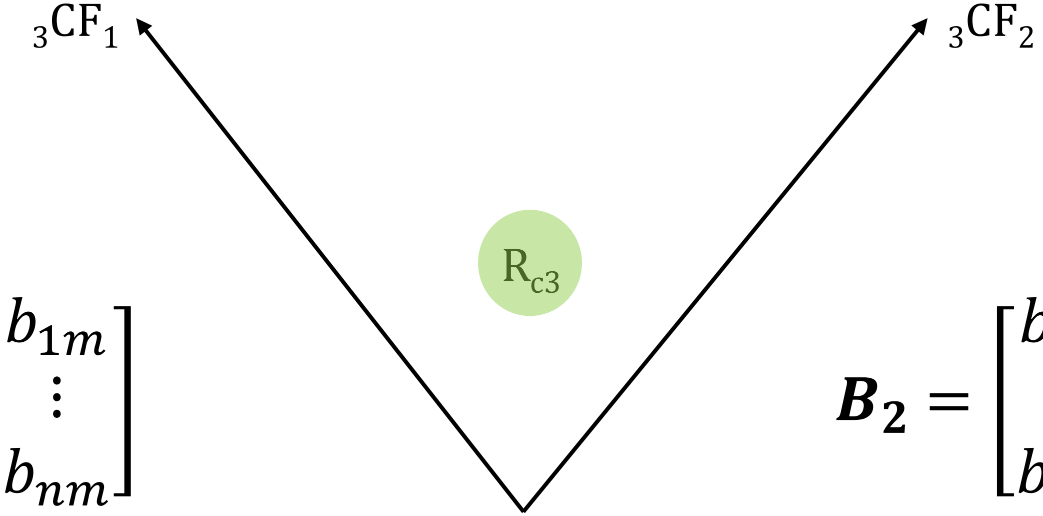
$$\boldsymbol{B_1} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix}$$



$$\boldsymbol{B_2} = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

Kanonička analiza

Prvi par kanoničkih faktora je u najvećoj korelaciji, dok je svaki sljedeći u jednakoj ili manjoj korelaciji u odnosu na prethodni.



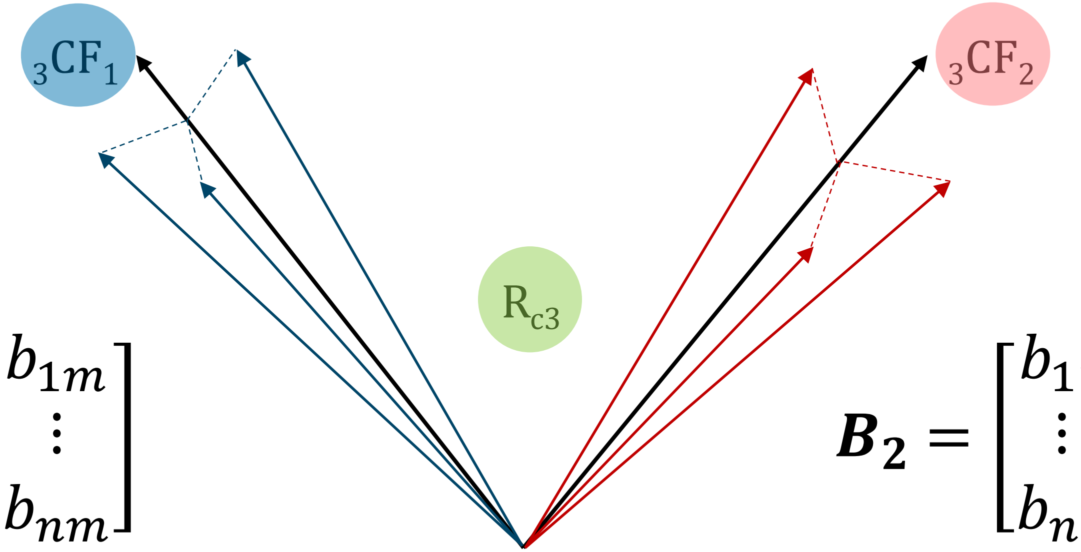
$$\boldsymbol{B_1} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix}$$

$$\boldsymbol{B_2} = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

Kanonička analiza

Struktura kanoničkih faktora utvrđuje se računanjem korelacija između manifestnih varijabli prvog skupa s kanoničkim faktorima prvog skupa, odnosno manifestnih varijabli drugog skupa s kanoničkim faktorima drugog skupa.

$$\boldsymbol{B_1} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix}$$



$$\boldsymbol{B_2} = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

Kanonička analiza

Testiranje statističke značajnosti kanoničkih korelacija vrši se *Bartlettovim* testom putem χ^2 - *distribucije* s brojem stupnjeva slobode $df = (m - g) \times (k - g)$. χ^2 vrijednost se izračunava sljedećom formulom

$$\chi^2 = -\left[(n-1) - \left(\frac{m+k+1}{2}\right)\right] \cdot \log_e \prod_{j=g+1}^{k} (1 - \lambda_j)$$

gdje je

- n - broj entiteta
- m - broj varijabli prvog skupa
- k - broj varijabli drugog skupa
- λ_j - koeficijent kanoničke determinacije (R^2_{cj}) j-tog para kanoničkih faktora
- g - broj prethodno testiranih kanoničkih korelacija

Kanonička analiza

Statistička značajnost kanoničkih korelacija testira se redom od prve do posljednje.

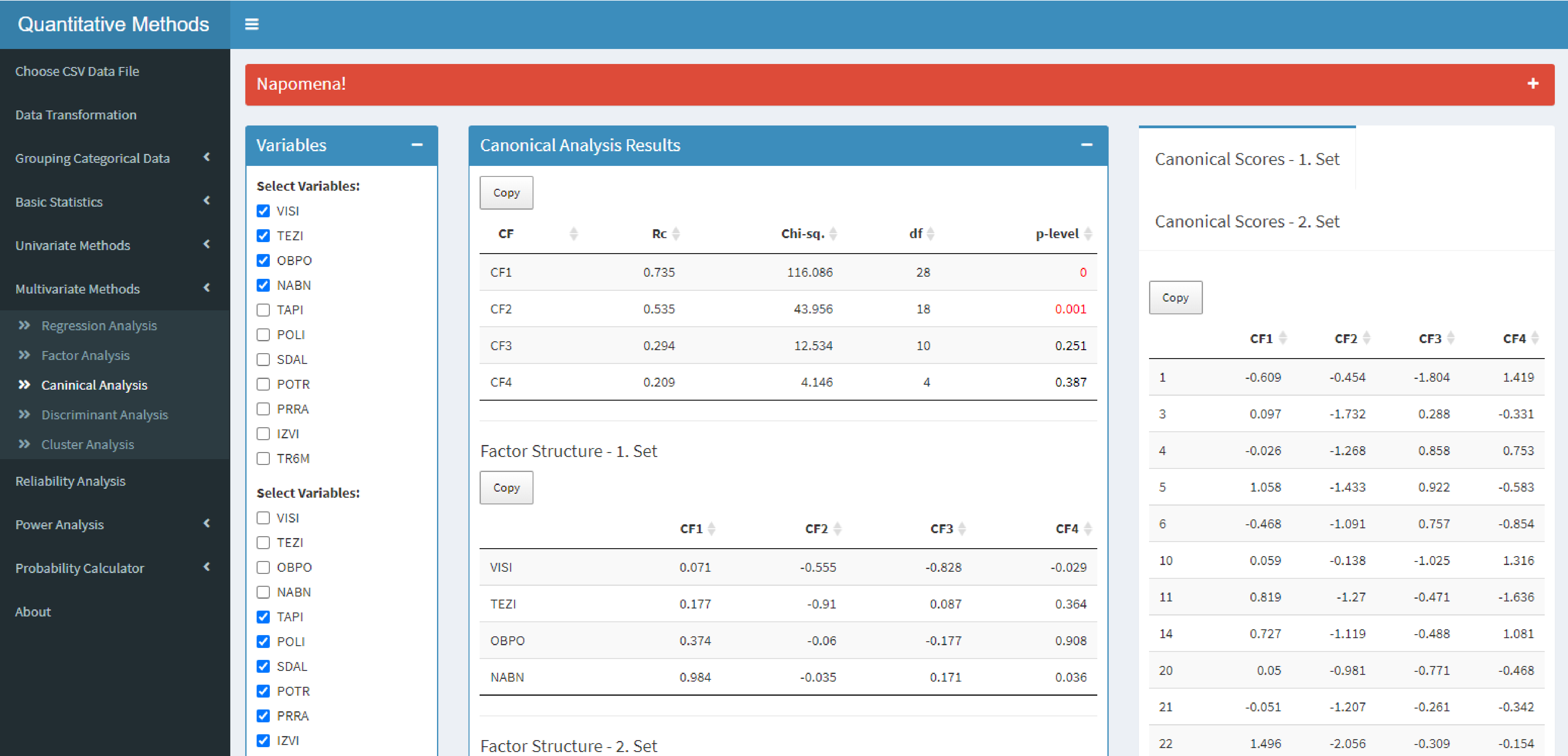
Utvrdi li se da neka kanonička korelacija nije statistički značajna, tada niti jedna sljedeća kanonička korelacija ne može biti statistički značajna.

Interpretiraju se samo oni parovi kanoničkih faktora čija je kanonička korelacija statistički značajna.

RStudio – Canonical Analysis

Zadatak: U matrici *SKOLA.csv* utvrdite relacije između skupa antropometrijskih varijabli (*VISI*, *TEZI*, *OBPO* i *NABN*) i skupa motoričkih testova (*TAPI*, *POLI*, *SDAL*, *POTR*, *PRRA*, *IZVI* i *TRGM*)! Izračunajte kanoničke korelacije i testirajte njihovu statističku značajnost! Interpretirajte matrice strukture kanoničkih faktora.

RStudio – Canonical Analysis



Variables

Select Variables:

☑ VISI
☑ TEZI
☑ OBPO
☑ NABN
☐ TAPI
☐ POLI
☐ SDAL
☐ POTR
☐ PRRA
☐ IZVI
☐ TR6M

Select Variables:

☐ VISI
☐ TEZI
☐ OBPO
☐ NABN
☑ TAPI
☑ POLI
☑ SDAL
☑ POTR
☑ PRRA
☑ IZVI

Canonical Analysis Results

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

Factor Structure - 1. 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

Factor Structure - 2. Set

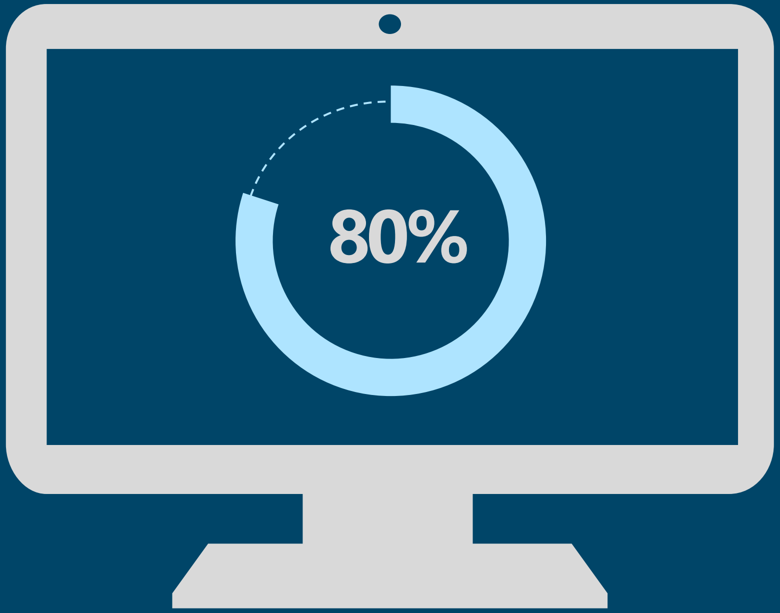
Canonical Scores - 1. Set

Canonical Scores - 2. Set

	CF1	CF2	CF3	CF4
1	-0.609	-0.454	-1.804	1.419
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
10	0.059	-0.138	-1.025	1.316
11	0.819	-1.27	-0.471	-1.636
14	0.727	-1.119	-0.488	1.081
20	0.05	-0.981	-0.771	-0.468
21	-0.051	-1.207	-0.261	-0.342
22	1.496	-2.056	-0.309	-0.154

DISKRIMINACIJSKA ANALIZA

Vježba 14



Diskriminacijska analiza

Diskriminacijskom analizom može se utvrditi statistička značajnost razlika između centroida dvije ili više grupa ispitanika te doprinos pojedinih varijabli razlikovanju među grupama.

Iz skupa od *m* manifestnih izračuna se *k-1* **diskriminacijskih funkcija** ako je broj varijabli (*m*) veći ili jednak broju grupa (*k*), odnosno *m* diskriminacijskih funkcija ako je broj varijabli manji od broja grupa.

Diskriminacijske funkcije su linearne kombinacije manifestnih varijabli koje se izračunavaju uz uvjet da se centroidi grupa entiteta na njima što je moguće više razlikuju i uz uvjet da su međusobno potpuno linearno nezavisne.

Diskriminacijska analiza

Neka su u matricama $\boldsymbol{B_g}$ (gdje je $g=1,2,\ldots,k$, a k broj grupa entiteta) podaci skupova $E_g = \{e_{gi}\ ;\ i = 1,\ldots,n_g\}$ entiteta opisanih skupom $V = \{v_j\ ;\ j = 1,\ldots,m\}$ varijabli, a u matrici $\boldsymbol{B}$ podaci svih entiteta odnosno skupa $E = \{e_i\ ;\ i = 1,\ldots,n\}$ entiteta opisanih skupom $V = \{v_j\ ;\ j = 1,\ldots,m\}$ varijabli.

$$\begin{tikzcd}
\boldsymbol{B_1} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n_1 1} & \cdots & b_{n_1 m} \end{bmatrix} \arrow[dr] & \\
\boldsymbol{B_2} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n_2 1} & \cdots & b_{n_2 m} \end{bmatrix} \arrow[r] & \boldsymbol{B} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} \\
\vdots & \\
\boldsymbol{B_k} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n_k 1} & \cdots & b_{n_k m} \end{bmatrix} \arrow[uur] &
\end{tikzcd}$$

Diskriminacijska analiza

Operacijom

$$\boldsymbol{m} = \boldsymbol{B}^T \boldsymbol{1}\, n^{-1}$$

Izračuna se **zajednički centroid**, odnosno vektor aritmetičkih sredina varijabli koji u obzir uzima sve entiteta (gdje je $\boldsymbol{1}$ vektor stupca sa n jedinica).

Matrica odstupanja rezultata entiteta od zajedničkog centroida dobije se operacijom

$$\boldsymbol{D} = \boldsymbol{B} - \boldsymbol{1}\, \boldsymbol{m}^T$$

, a opracijom

$$\boldsymbol{T} = \boldsymbol{D}^T \boldsymbol{D}$$

izračuna se **matrica suma kvadrata i krosprodukata za total**.

Diskriminacijska analiza

Operacijom

$$\boldsymbol{m_g} = \boldsymbol{B_g}^{\boldsymbol{T}}\, \boldsymbol{1_g}\, n_g^{-1},$$

izračuna k vektora aritmetičkih sredina varijabli gdje se u obzir uzimaju samo entiteti iz pojedine grupe (n_g), gdje je $g = 1,\dots,k$, a $\boldsymbol{1_g}$ vektor stupca sa n_g jedinica.

Operacijom

$$\boldsymbol{W_g} = \boldsymbol{D_g}^{\boldsymbol{T}} \boldsymbol{D_g}$$

izračunju se **matrice suma kvadrata i krosprodukata za svaku grupu**, *a* sumacijom svih dobivenih matrica dobije se **matrica suma kvadrata i krosprodukata unutar grupa**

$$\boldsymbol{W} = \sum_{g=1}^{k} \boldsymbol{W}_g$$

Diskriminacijska analiza

Matrica suma kvadrata i krosprodukata između grupa može se izračunati operacijom

$$\mathbf{A} = \mathbf{T} - \mathbf{W}$$

jer vrijedi da je

$$\mathbf{T} = \mathbf{W} + \mathbf{A}$$

Matricu **diskriminacijskih funkcija** (ψ) čine rezultati entiteta u *k-1* diskriminacijskih funkcija, a izračunava se operacijom

$$\psi = \mathbf{D}\,\mathbf{X}$$

gdje je

X - matrica svojstvenih vektora matrice $\mathbf{W}^{-1}\mathbf{A}$

Diskriminacijska analiza

Operacijom

$$\Delta = \psi^T \psi n^{-1}$$

se izračuna dijagonalna matrica **varijanci diskriminacijskih funkcija**, a operacijom

$$\phi = \psi \Delta^{-1/2}$$

standardizirani rezultati entiteta na diskriminacijskim funkcijama, odnosno matrica **standardiziranih diskriminacijskih funkcija**.

Koeficijent kanoničke diskriminacije (R_{cj}) je mjera diskriminacijske moći odgovarajuće diskriminacijske funkcije, a predstavlja korelaciju **diskriminacijske funkcije** sa **selektorskom varijablom** odnosno varijablom koja određuje pripadnost entiteta grupama.

$$R_{cj} = \sqrt{\frac{\lambda_j}{1 + \lambda_j}}$$

Koeficijenti kanoničke diskriminacije se kreću u intervalu od *0* do *1*. Što je koeficijent kanoničke diskriminacije veći, to znači da pripadajuća diskriminacijska funkcija bolje razlikuje grupe entiteta.

Diskriminacijska analiza

Statistička značajnost diskriminacijskih funkcija testira se redom od prve do posljednje. Utvrdi li se da neka diskriminacijska funkcija nije statistički značajna, tada niti jedna sljedeća diskriminacijska funkcija ne može biti statistički značajna.

Ako se želi zaključivati s uzorka na populaciju entiteta interpretiraju se samo one diskriminacijske funkcije za koje je utvrđeno da su statistički značajne.

Testiranje statističke značajnosti diskriminacijskih funkcija vrši se putem **χ^2 - distribucije** s brojem stupnjeva slobode *df=(m-p)×(k-p-1)*. χ^2 vrijednost se izračunava sljedećom formulom

$$\chi^2 = -\left(\mathrm{n} - \frac{\mathrm{m} + \mathrm{k}}{2} - 1\right)\log_e\left(\prod_{j=p+1}^{q}\frac{1}{1+\lambda_\mathrm{j}}\right)$$

gdje je

- n - broj entiteta
- k - broj grupa
- q - broj diskriminacijskih funkcija
- p - broj prethodno testiranih diskriminacijskih funkcija
- λ_j - svojstvene vrijednosti matrice $\boldsymbol{W}^{-1}\boldsymbol{A}$

Diskriminacijska analiza

Doprinos pojedinih manifestnih varijabli razlikovanju centroida grupa određuje se na temelju **matrice strukture diskriminacijskih funkcija** (**F**) koja se izračunava operacijom

$$\mathbf{F} = \mathbf{Z}^{\mathbf{T}} \phi \;\; \mathbf{n}^{-1}$$

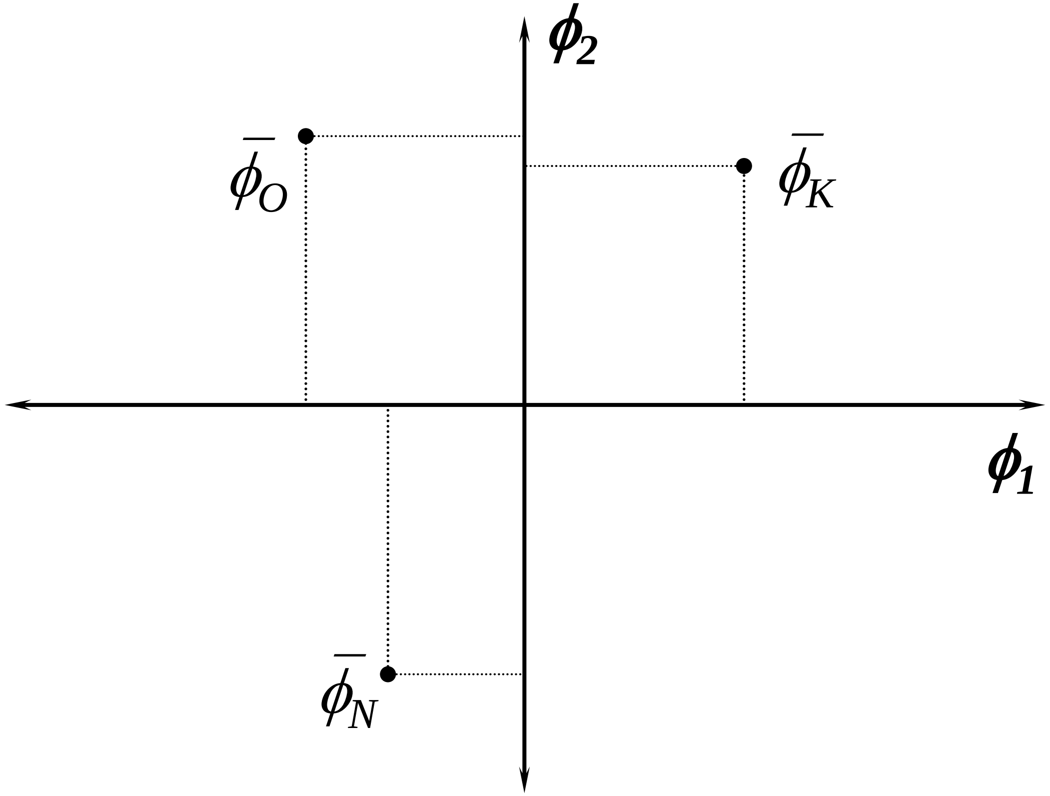
Matricu strukture diskriminacijskih funkcija čine korelacije manifestnih varijabli s diskriminacijskim funkcijama.

U interpretaciji razlika potrebno je utvrditi i ***centroide grupa*** na diskriminacijskim funkcijama kao i njihove međusobne udaljenosti. ***Centroidi grupa*** ($\bar{\phi}_g$) u prostoru diskriminacijskih funkcija izračunaju se operacijom

$$\bar{\phi}_{\mathbf{g}} = \phi_{\mathbf{g}}^{T} \, \mathbf{1}_{g} \, n_{g}^{-1}$$

Diskriminacijska analiza

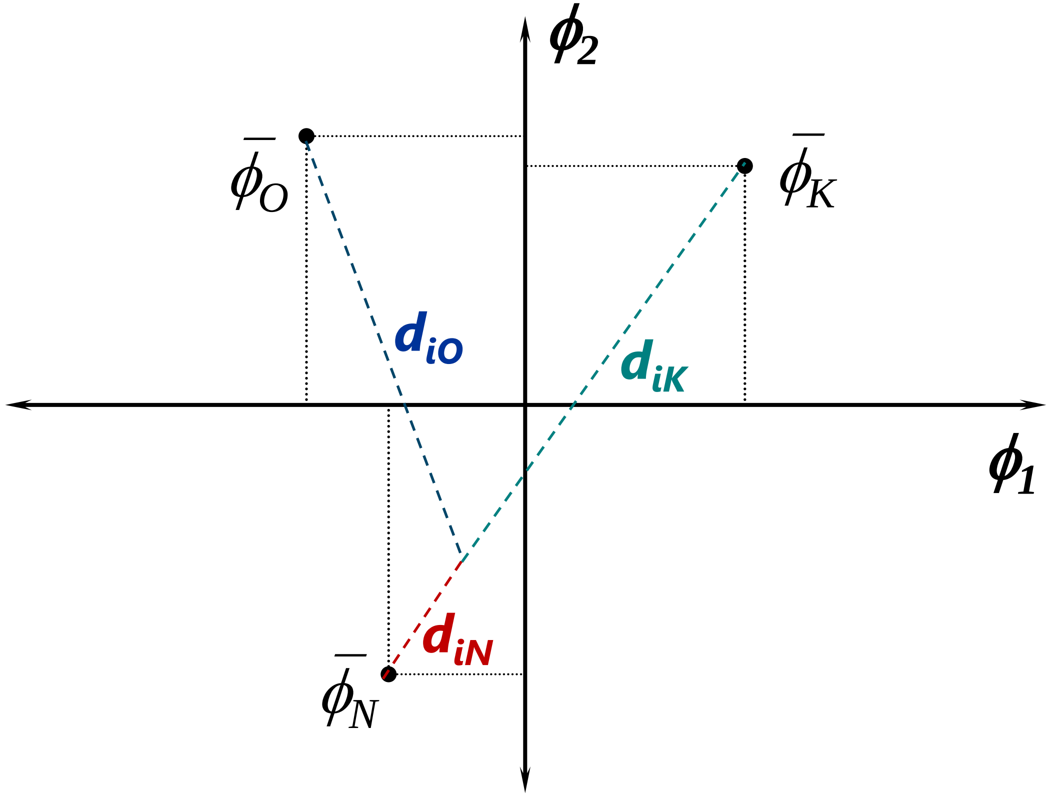
Prikaz koordinata centroida grupa K, N i O u koordinatnom sustavu dviju diskriminacijskih funkcija



Diskriminacijska analiza

Prikaz udaljenosti rezultata entiteta *i* od centroida grupa K, N i O u koordinatnom sustavu dviju diskriminacijskih funkcija

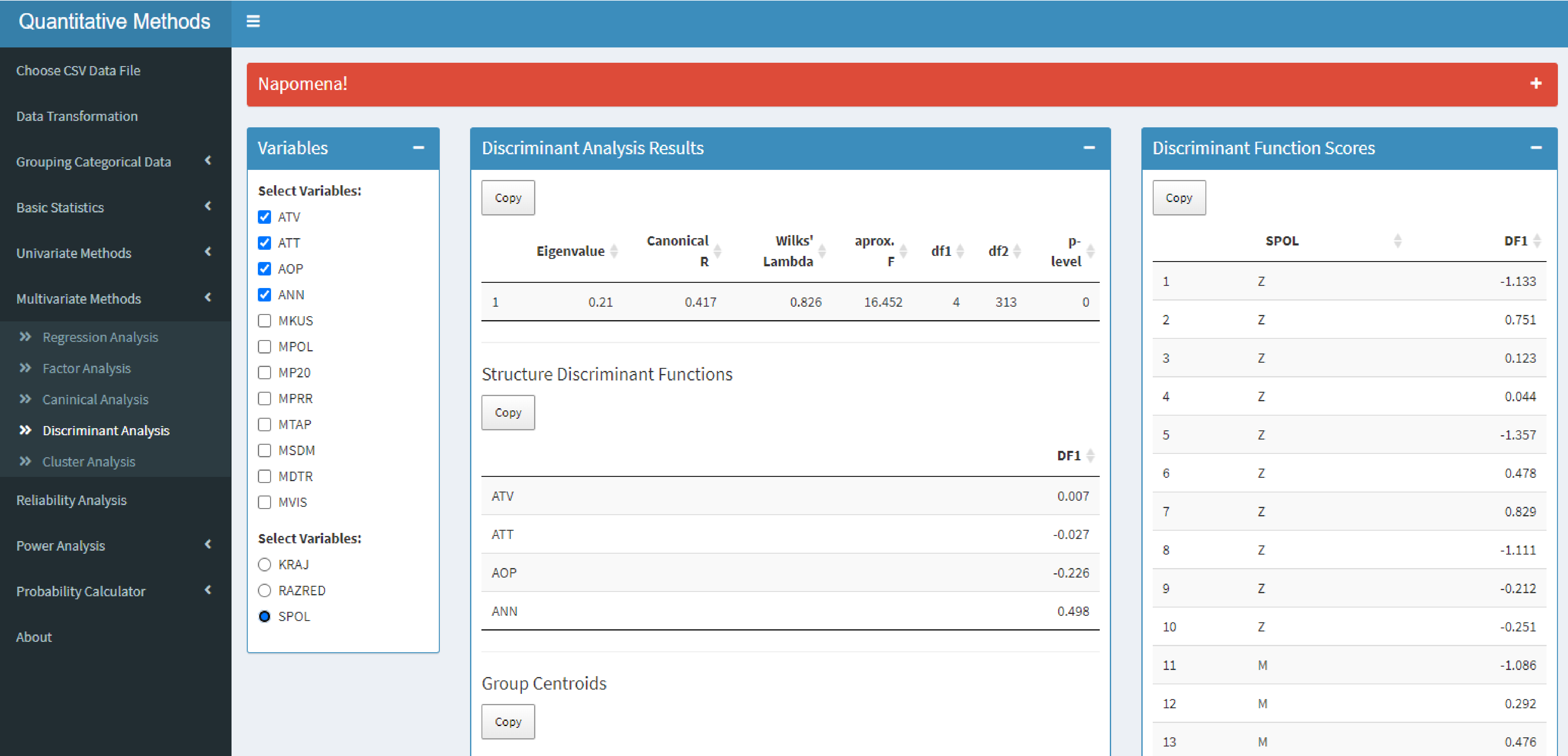
Pripadnost nekog entiteta određenoj grupi može se procijeniti na temelju udaljenosti vektora njegovih rezultata na diskriminacijskim funkcijama od centroida svake pojedine grupe. Što je udaljenost od centroida grupe manja to je veća vjerojatnost da entitet pripada toj grupi.



RStudio – Discriminant Analysis

Zadatak: U datoteci *UCENICI.csv* utvrdite da li se učenici i učenice 1., 2., 3. i 4. razreda statistički značajno razlikuju prema građi tijela (varijable *ATV, ATT, AOP i ANN)*! Testirajte statističku značajnost razlika, izračunajte koeficijente kanoničke diskriminacije, matricu strukture i centroide grupa!

RStudio – Discriminant Analysis



Variables

Select Variables:

☑ ATV
☑ ATT
☑ AOP
☑ ANN
☐ MKUS
☐ MPOL
☐ MP20
☐ MPRR
☐ MTAP
☐ MSDM
☐ MDTR
☐ MVIS

Select Variables:

○ KRAJ
○ RAZRED
● SPOL

Discriminant Analysis Results

	Eigenvalue	Canonical R	Wilks' Lambda	aprox. F	df1	df2	p-level
1	0.21	0.417	0.826	16.452	4	313	0

Structure Discriminant Functions

	DF1
ATV	0.007
ATT	-0.027
AOP	-0.226
ANN	0.498

Group Centroids

Discriminant Function Scores

	SPOL	DF1
1	Z	-1.133
2	Z	0.751
3	Z	0.123
4	Z	0.044
5	Z	-1.357
6	Z	0.478
7	Z	0.829
8	Z	-1.111
9	Z	-0.212
10	Z	-0.251
11	M	-1.086
12	M	0.292
13	M	0.476

METRIJSKE KARAKTERISTIKE 1

Vježba 15

Metrijske karakteristike

Nakon konstrukcije preliminarne forme mjernog instrumenta potrebno je testirati njegovu djelotvornost, odnosno utvrditi **metrijske karakteristike**.

Metrijske karakteristike su:
- pouzdanost
- objektivnost
- homogenost
- valjanost i
- osjetljivost.

Metrijske karakteristike

Pouzdanost

Pouzdanost mjernog instrumenta je nezavisnost mjerenja od nesistematskih, tj. slučajnih pogrešaka.

Pouzdanost mjernog instrumenta može se kretati u intervalu od *0* do *1* pri čemu 1 označava potpuno odsustvo slučajne pogreške mjerenja.

Pouzdanost se može utvrditi:
- test-retest metodom
- metodom tau-ekvivalentnih testova i
- metodom interne konzistencije.

Test-retest metoda podrazumijeva primjenu mjernog instrumenta na istoj grupi ispitanika u dva navrata, pri čemu se koeficijentom pouzdanosti smatra korelacija između rezultata prvog i ponovljenog mjerenja.

Period između testa i retesta mora biti dovoljno dug da prvo mjerenje ne utječe na drugo (npr. odgovaranje prema sjećanju, upala mišića od prethodnog testiranja), odnosno dovoljno kratak da se razvijenost predmeta mjerenja ne promijeni.

Metrijske karakteristike

Pouzdanost

Metoda tau-ekvivalentnih testova podrazumijeva primjenu dva mjerna instrumenta, koji su slični po sadržaju, broju i obliku zadataka, na istoj grupi ispitanika, pri čemu se koeficijentom pouzdanosti smatra korelacija između rezultata primijenjenih testova.

Vrijednost koeficijenta pouzdanosti utvrđenog ovom metodom je nešto niža od stvarne pouzdanosti testa jer se korištenjem tau-ekvivalentnih testova u određenoj mjeri ipak aktiviraju ponešto različiti predmeti mjerenja.

Metoda interne konzistencije se koristi za utvrđivanje pouzdanosti kompozitnih mjernih instrumenata, a temelji se na pretpostavci da su sve čestice testa međusobno paralelni testovi.

Koeficijenti pouzdanosti koji se utvrđuju metodom interne konzistencije odabiru se s obzirom na način *kondenzacije rezultata*.

Kondenzacija rezultata je izračunavanje ukupnog rezultata ispitanika na temelju pripadajućih rezultata u česticama testa.

Metrijske karakteristike

Pouzdanost

Načini kondenzacije rezultata i odgovarajući koeficijenti pouzdanosti

kondenzacija rezultata	**koeficijent pouzdanosti**
jednostavnom linearnom kombinacijom originalnih rezultata	Cronbach α
jednostavnom linearnom kombinacijom standardiziranih rezultata	standardizirana ili Spearman-Brown α
prvom glavnom komponentom	Kaiser-Caffrey α

Metrijske karakteristike

Pouzdanost

Cronbachov koeficijent pouzdanosti ($\alpha_{Cronbach}$) se izračunava formulom:

$$\alpha_{\mathrm{Cronbach}} = \frac{\mathrm{m}}{\mathrm{m}-1}\left(1 - \frac{\sum_{\mathrm{j}=1}^{\mathrm{m}} \sigma_{\mathrm{j}}^{2}}{\sigma^{2}}\right)$$

gdje je
- m - broj čestica testa
- σ_j^2 - varijanca čestice j
- σ^2 - varijanca testa

Metrijske karakteristike

Pouzdanost

Primjer: izračunavanje Cronbachovog koeficijenta pouzdanosti

ISP.	x_1	x_2	x_3	x
1	70	115	115	**300**
2	150	145	142	**437**
3	100	120	122	**342**
4	120	92	100	**312**
5	105	82	87	**274**
6	94	116	117	**327**
7	120	100	105	**325**
8	60	70	76	**206**
9	99	104	105	**308**
10	100	110	108	**318**
$\bar{x}$	**101,80**	**105,40**	**107,70**	**314,90**
σ^2	**647,73**	**446,49**	**336,46**	**3301,21**

$$\alpha_{\text{Cronbach}} = \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \sigma_j^2}{\sigma^2}\right)$$

$$\alpha_{\text{Cronbach}} = \frac{3}{2}\left(1 - \frac{647{,}73 + 446{,}49 + 336{,}46}{3301{,}21}\right)$$

$$\alpha_{\text{Cronbach}} = 0{,}85$$

Metrijske karakteristike

Pouzdanost

Spearman-Brownov koeficijent pouzdanosti (*SB*) se izračunava formulom:

$$\mathrm{SB} = \frac{\mathrm{m}}{\mathrm{m}-1}\left(1 - \frac{\mathrm{m}}{\sigma^2}\right)$$

gdje je

- m - broj čestica testa
- σ^2 - varijanca testa

Metrijske karakteristike

Pouzdanost

Primjer: izračunavanje Spearman-Brownovog koeficijenta pouzdanosti

ISP.	**z_1**	**z_2**	**z_3**	**z**
1	-1,249	0,454	0,398	**-0,397**
2	1,894	1,874	1,870	**5,638**
3	-0,071	0,691	0,780	**1,400**
4	0,715	-0,634	-0,420	**-0,339**
5	0,126	-1,107	-1,129	**-2,110**
6	-0,306	0,502	0,507	**0,702**
7	0,715	-0,256	-0,147	**0,312**
8	-1,642	-1,675	-1,728	**-5,046**
9	-0,110	-0,066	-0,147	**-0,323**
10	-0,071	0,218	0,016	**0,163**
$\bar{x}$	**0**	**0**	**0**	**0**
σ^2	**1**	**1**	**1**	**7,18**

$$\mathrm{SB} = \frac{\mathrm{m}}{\mathrm{m}-1}\left(1 - \frac{m}{\sigma^2}\right)$$

$$\mathrm{SB} = \frac{3}{2}\left(1 - \frac{3}{7{,}18}\right)$$

$$SB = 0{,}87$$

Metrijske karakteristike

Pouzdanost

Kaiser-Caffreyev koeficijent pouzdanosti ($\alpha_{Kaiser\text{-}Caffrey}$) se izračunava formulom:

$$\alpha_{\mathrm{Kaiser-Caffrey}} = \frac{\mathrm{m}}{\mathrm{m}-1}\left(1 - \frac{1}{\lambda_1}\right)$$

gdje je
- m - broj čestica testa
- λ_1 - prva svojstvena vrijednost matrice korelacija među česticama

Metrijske karakteristike

Pouzdanost

Primjer: izračunavanje Kaiser-Caffreyevog koeficijenta pouzdanosti

ISP.	z_1	z_2	z_3	k_1
1	-1,249	0,454	0,398	**-0,046**
2	1,894	1,874	1,870	**2,078**
3	-0,071	0,691	0,780	**0,565**
4	0,715	-0,634	-0,420	**-0,199**
5	0,126	-1,107	-1,129	**-0,853**
6	-0,306	0,502	0,507	**0,308**
7	0,715	-0,256	-0,147	**0,061**
8	-1,642	-1,675	-1,728	**-1,865**
9	-0,110	-0,066	-0,147	**-0,119**
10	-0,071	0,218	0,016	**0,071**
$\bar{x}$	**0**	**0**	**0**	**0**
σ^2	**1**	**1**	**1**	**2,42**

$$\alpha_{\text{Kaiser}-\text{Caffrey}} = \frac{\text{m}}{\text{m}-1}\left(1-\frac{1}{\lambda_1}\right)$$

$$\alpha_{\text{Kaiser}-\text{Caffrey}} = \frac{3}{2}\left(1-\frac{1}{2{,}42}\right)$$

$$\alpha_{\text{Kaiser}-\text{Caffrey}} = 0{,}88$$

Metrijske karakteristike

Objektivnost

Objektivnost je mjerna karakteristika kojom se određuje nezavisnost rezultata mjerenja od mjeritelja.

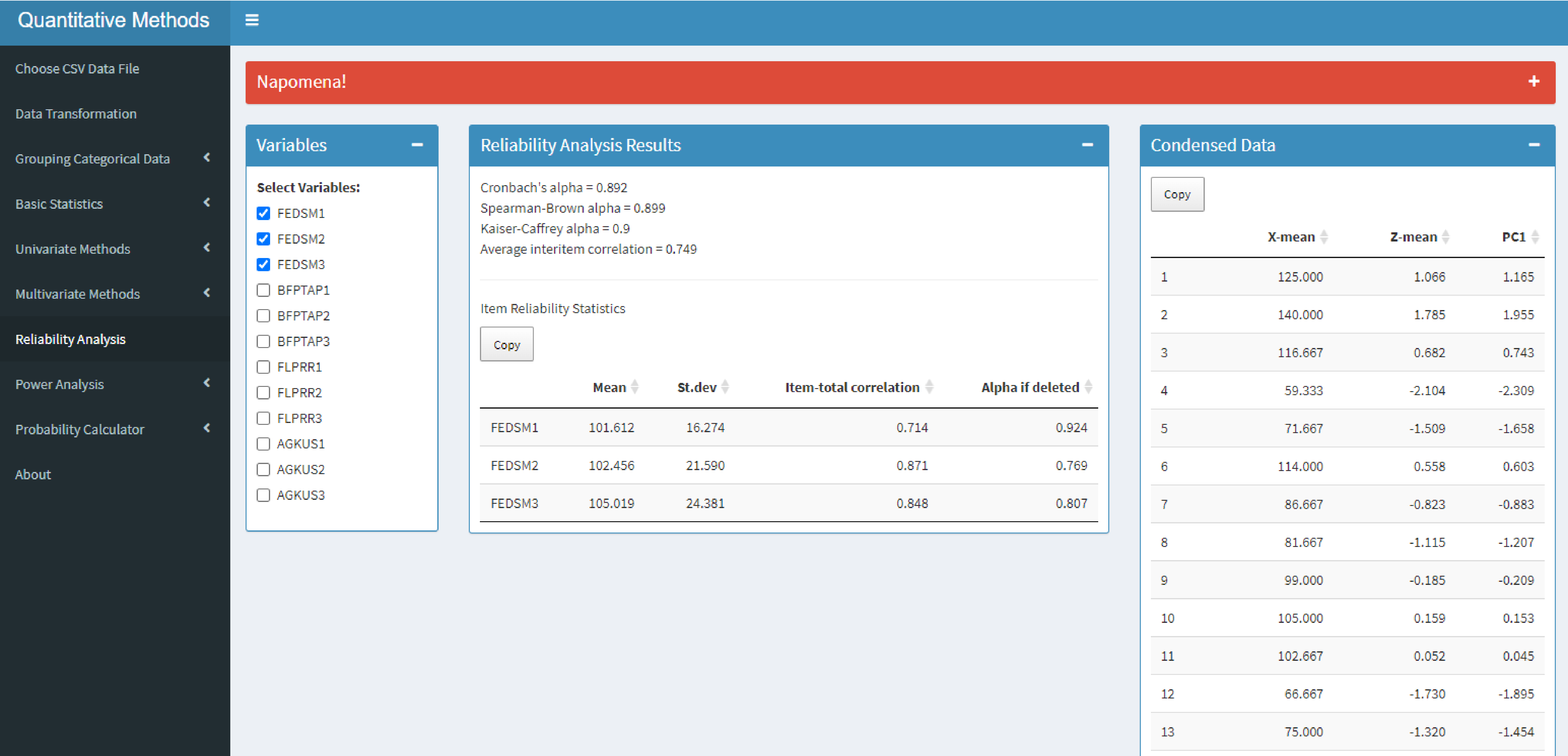
Koeficijent objektivnosti se kreće u intervalu od *0* do *1*, a vrijednost mu je to veća što je veći broj mjeritelja i što je veći stupanj slaganja između rezultata ispitanika utvrđenih od strane različitih mjeritelja.

Postupak za utvrđivanje objektivnosti nekog mjerenja u kome sudjeluje veći broj mjeritelja identičan je metodi interne konzistencije za utvrđivanje pouzdanosti kompozitnih mjernih instrumenata, pri čemu su čestice mjerenja mjeritelji, odnosno suci.

RStudio – Reliability Analysis

Zadatak: U datoteci *GIM.csv* izračunajte *Cronbachov*, *Spearman-Brownov* i *Kaiser-Caffreyev* koeficijent pouzdanosti koeficijent pouzdanosti za test *FEDSM*!

RStudio – Reliability Analysis



Select Variables:

- ☑ FEDSM1
- ☑ FEDSM2
- ☑ FEDSM3
- ☐ BFPTAP1
- ☐ BFPTAP2
- ☐ BFPTAP3
- ☐ FLPRR1
- ☐ FLPRR2
- ☐ FLPRR3
- ☐ AGKUS1
- ☐ AGKUS2
- ☐ AGKUS3

Cronbach's alpha = 0.892
Spearman-Brown alpha = 0.899
Kaiser-Caffrey alpha = 0.9
Average interitem correlation = 0.749

Item Reliability Statistics

	Mean	St.dev	Item-total correlation	Alpha if deleted
FEDSM1	101.612	16.274	0.714	0.924
FEDSM2	102.456	21.590	0.871	0.769
FEDSM3	105.019	24.381	0.848	0.807

	X-mean	Z-mean	PC1
1	125.000	1.066	1.165
2	140.000	1.785	1.955
3	116.667	0.682	0.743
4	59.333	-2.104	-2.309
5	71.667	-1.509	-1.658
6	114.000	0.558	0.603
7	86.667	-0.823	-0.883
8	81.667	-1.115	-1.207
9	99.000	-0.185	-0.209
10	105.000	0.159	0.153
11	102.667	0.052	0.045
12	66.667	-1.730	-1.895
13	75.000	-1.320	-1.454

METRIJSKE KARAKTERISTIKE 2

Vježba 16

Metrijske karakteristike

Homogenost

Homogenost je mjerna karakteristika kompozitnih mjernih instrumenata koja pokazuje koliko rezultati ispitanika u svim česticama zavise od istog predmeta mjerenja ili identične kombinacije različitih predmeta mjerenja.

Koeficijent homogenosti se kreće u intervalu od *0* do *1*, a najčešće se utvrđuje izračunavanjem **prosječne korelacije među česticama**.

Prosječna korelacija među česticama se izračunava sljedećom formulom

$$\bar{r} = \frac{\sum_{j,k=1}^{m} r_{j,k}}{m}, j \neq k$$

gdje je

- *r* - korelacija među česticama *j* i *k*
- *m* - broj čestica testa

Metrijske karakteristike

Homogenost

Primjer: izračunavanje prosječne korelacije među česticama

ISP.	**z_1**	**z_2**	**z_3**
1	-1,249	0,454	0,398
2	1,894	1,874	1,870
3	-0,071	0,691	0,780
4	0,715	-0,634	-0,420
5	0,126	-1,107	-1,129
6	-0,306	0,502	0,507
7	0,715	-0,256	-0,147
8	-1,642	-1,675	-1,728
9	-0,110	-0,066	-0,147
10	-0,071	0,218	0,016

$$\mathrm{r}_{\mathrm{z1,z2}} = 0{,}53 \qquad \bar{r} = \frac{\sum_{j,k=1}^{m} r_{zj,zk}}{m}, j \neq k$$

$$\mathrm{r}_{\mathrm{z1,z3}} = 0{,}57 \qquad \bar{r} = \frac{\mathrm{r}_{\mathrm{z1,z2}} + \mathrm{r}_{\mathrm{z1,z3}} + r_{z2,z3}}{3}$$

$$r_{z2,z3} = 0{,}99 \qquad \bar{r} = \frac{0{,}53 + 0{,}57 + 0{,}99}{3} = 0{,}70$$

Metrijske karakteristike

Osjetljivost

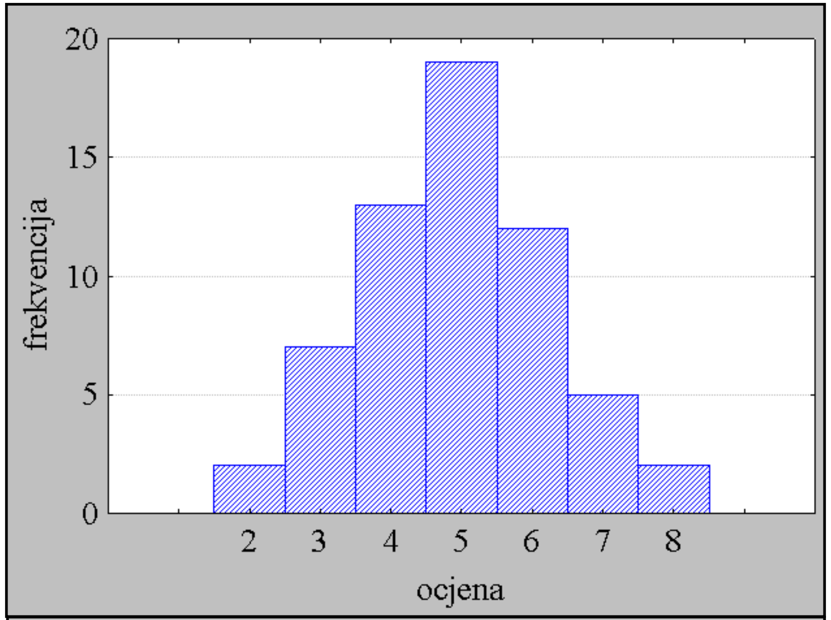
Osjetljivost je metrijska karakteristika koja pokazuje koliko uspješno mjerni instrument razlikuje ispitanike po predmetu mjerenja.

Osjetljivost kineziologijskih mjernih instrumenta se procjenjuje na temelju mjera disperzije i oblika distribucije rezultata.

Metrijske karakteristike

Osjetljivost

Primjer 1: 60 odraslih veslača je izmjereno nekim testom koordinacije u ritmu i dobiveni su sljedeći rezultati:



$\bar{\chi}= 4,92$ $s=1,36$ $V=27,64\%$

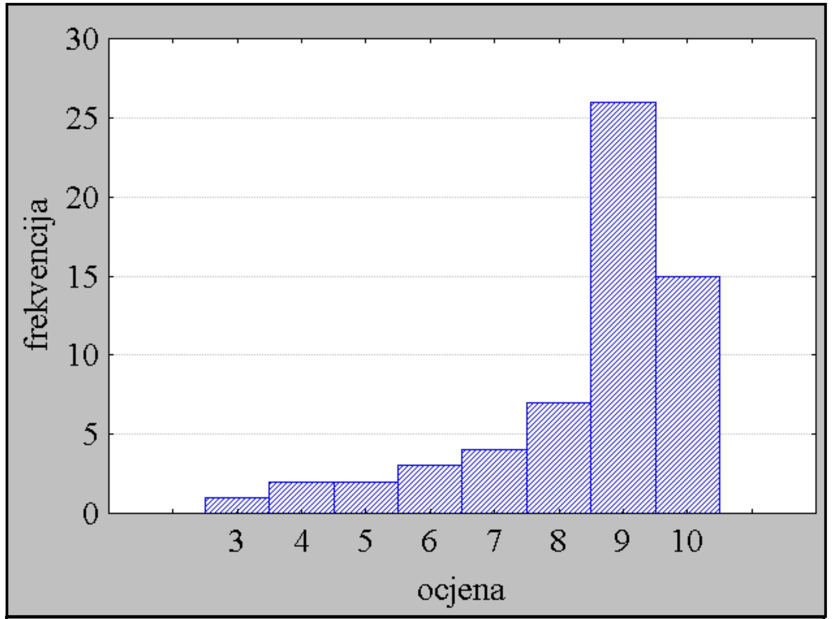
$R_{tot}=6$ $a3=0,07$ $a4=1,83$

Mjere varijabilnosti i koeficijent zakrivljenosti distribucije (*a4*) upućuju na zadovoljavajuću osjetljivost mjernog instrumenta. Simetrija distribucije koja se može uočiti iz oblika histograma frekvencija, kao i iz koeficijenta asimetrije distribucije (*a3*), upućuje na prilagođenost težine testa izmjerenom uzorku ispitanika.

Metrijske karakteristike

Osjetljivost

Primjer 2: 60 odraslih plesača je izmjereno nekim testom koordinacije u ritmu i dobiveni su sljedeći rezultati:



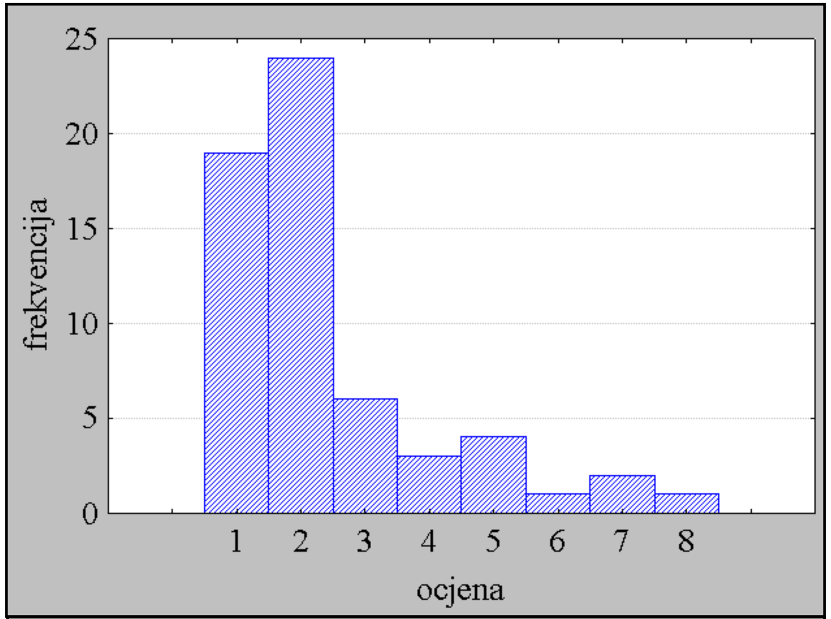
$\bar{x}$= 8,45	s=1,67	V=19,76%
R_{tot}=7	$a3$=-1,57	$a4$=2,04

Negativna asimetrija distribucije koja se može uočiti iz oblika histograma frekvencija kao i iz koeficijenta asimetrije distribucije (*a3*) upućuje na zaključak da je test prelagan za izmjereni uzorak ispitanika. Visoka vrijednost koeficijenta zakrivljenosti distribucije (*a4*) upućuje na slabu osjetljivost testa.

Metrijske karakteristike

Osjetljivost

Primjer 3: 60 učenika prvog razreda osnovne škole je izmjereno nekim testom koordinacije u ritmu i dobiveni su sljedeći rezultati:



$\bar{x}$= 2,42 *s=1,67* *V=69,01%*
R_{tot}=7 *a3=1,65* *a4=2,37*

Pozitivna asimetrija distribucije koja se može uočiti iz oblika histograma frekvencija kao i iz koeficijenta asimetrije distribucije (*a3*) upućuje na zaključak da je test pretežak za izmjereni uzorak ispitanika. Visoka vrijednost koeficijenta zakrivljenosti distribucije (*a4*) upućuje na slabu osjetljivost testa.

Metrijske karakteristike

Dijagnostička valjanost

Dijagnostička valjanost je metrijska karakteristika koja pokazuje koliko stvarni predmet mjerenja odgovara ciljanom predmetu mjerenja.

Dijagnostičku valjanost je prema načinu utvrđivanja moguće podijeliti na *apriornu valjanost* i *faktorsku valjanost*.

Apriorna valjanost se utvrđuje promišljanjem stručnjaka o stvarnom predmetu mjerenja, a **faktorska valjanost** eksperimentalnom provjerom pretpostavke o predmetu mjerenja.

Metrijske karakteristike

Dijagnostička valjanost

Faktorska valjanost se može izraziti:

- **korelacijom rezultata testa s rezultatima referentnog mjernog instrumenta**, odnosno testa za kojeg je potvrđena visoka faktorska valjanost za procjenu ciljanog predmeta mjerenja (primjerice, korelacija rezultata upitnika za procjenu tjelesne aktivnosti nepoznate faktorske valjanosti s rezultatima sedmodnevnog dnevnika tjelesne aktivnosti).
- **korelacijom testa s prvom glavnom komponentom** izračunatom na temelju rezultata tri ili više testova za koje je potvrđena visoka faktorska valjanost za procjenu ciljanog predmeta mjerenja (primjerice, korelacija rezultata testa za procjenu eksplozivne jakosti nepoznate faktorske valjanosti s prvom glavnom komponentom izračunatom na temelju rezultata testova *MFESDM - Skok udalj s mjesta*, *MFESVM - Skok uvis s mjesta* i *MFEBML - Bacanje medicinke iz ležanja na leđima*)
- **korelacijom testa s faktorom koji predstavlja ciljani predmet mjerenja**, a koji je utvrđen faktorskom analizom u koju su, osim testova za procjenu ciljanog predmeta mjerenja, uključeni i testovi drugih predmeta mjerenja (primjerice, korelacija rezultata testa za procjenu eksplozivne jakosi nepoznate faktorske valjanosti s faktorom koji predstavlja eksplozivnu jakost utvrđenim faktorskom analizom u koju je uključen skup testova za procjenu različitih motoričkih sposobnosti)

Metrijske karakteristike

Prognostička valjanost

Prognostička valjanost je metrijska karakteristika koja pokazuje s kolikom se sigurnošću može predvidjeti uspješnost u nekoj praktičnoj aktivnosti na temelju rezultata u jednom ili više testova. Može se utvrditi prognostička valjanost:

- **Prognostička valjanost jednog testa za jednodimenzionalni kvantitativni kriterij** (npr. uspješnost u bacanju kugle određena najboljim hicem) utvrđuje se **Pearsonovim koeficijentom korelacije** između rezultata testa i rezultata u kriterijskoj varijabli ili **jednostavnom regresijskom analizom**.

- **Prognostička valjanost skupa testova za jednodimenzionalni kvantitativni kriterij** utvrđuje se **višestrukom regresijskom analizom** pri čemu je mjera valjanosti skupa testova *koeficijent multiple korelacije*, dok su *standardizirani regresijski koeficijenti* mjere valjanosti svakog pojedinog testa.

- **Prognostička valjanost skupa testova za višedimenzionalni kvantitativni kriterij** (npr. uspješnost u košarkaškoj igri određena nizom parametara situacijske učinkovitosti) utvrđuje se **kanoničkom analizom** pri čemu se mjerom valjanosti skupa testova smatraju *koeficijenti kanoničke korelacije*.

Metrijske karakteristike

Prognostička valjanost

Prognostička valjanost je metrijska karakteristika koja pokazuje s kolikom se sigurnošću može predvidjeti uspješnost u nekoj praktičnoj aktivnosti na temelju rezultata u jednom ili više testova. Može se utvrditi prognostička valjanost:

- **Prognostička valjanost jednog testa za jednodimenzionalni kvalitativni kriterij** (npr. uspješnost tenisača u prvom kolu turnira opisana s dvije kategorije - *pobjeda* i *poraz*) utvrđuje se **t-testom za nezavisne uzorke** ili **univarijatnom analizom varijance**, a mjera valjanosti testa je *statistička značajnost* i *veličina razlike/a između aritmetičkih sredina* .

- **Prognostička valjanost skupa testova za jednodimenzionalni kvalitativni kriterij** se utvrđuje **diskriminacijskom analizom** pri čemu su mjera valjanosti skupa testova *koeficijenti kanoničke diskriminacije*, a mjera valjanosti svakog pojedinog testa *korelacije testa s diskriminacijskim funkcijama*.

Metrijske karakteristike

Prognostička valjanost

BROJ TESTOVA	TIP KRITERIJA	METODA
jedan	jednodimenzionalni kvantitativni	korelacija, jednostruka regresijska analiza
dva ili više	jednodimenzionalni kvantitativni	višestruka regresijska analiza
dva ili više	višedimenzionalni kvantitativni	kanonička analiza
jedan	jednodimenzionalni kvalitativni	t-test, anova
dva ili više	jednodimenzionalni kvalitativni	diskriminacijska analiza

(Metode za utvrđivanje prognostičke valjanosti s obzirom na broj testova i tip kriterija)