

2. Deskriptivna statistika i grafičko prikazivanje podataka

TEME

- | | |
|----|--|
| 1. | Tablice frekvencija |
| 2. | Deskriptivni statistički pokazatelji |
| 3. | Deskriptivna statistika zasebno za svaku skupinu |
| 4. | Grafičko prikazivanje podataka |
| 5. | Grafički prikazi zasebno za svaku skupinu |

2.1. Tablice frekvencija

Radimo na primjerima u datoteci pod nazivom:
02_deskriptivna_statistika.sav

Proceduru pokrećemo u *Descriptives* meniju, preko opcije *Descriptive Statistics*.

Za primjer ćemo zatražiti tablicu frekvencija za varijablu *Seksizam01*.

The screenshot shows the SPSS interface. On the left, the 'Descriptive Statistics' dialog box is open, with 'Seksizam1' in the 'Variables' list. The 'Frequency tables' checkbox is checked and circled in red. On the right, the 'Results' window displays the following data:

Descriptive Statistics

Seksizam1	
Valid	161
Missing	4
Mean	2.325
Std. Deviation	1.191
Minimum	1.000
Maximum	5.000

Frequency Tables

Frequencies for Seksizam1

Seksizam1	Frequency	Percent	Valid Percent	Cumulative Percent
uopće se ne slažem	41	24.848	25.466	25.466
ne slažem se	60	36.364	37.267	62.733
ne znam	28	16.970	17.391	80.124
slažem se	29	17.576	18.012	98.137
potpuno se slažem	3	1.818	1.863	100.000
Missing	4	2.424		
Total	165	100.000		

Opciju ispisa tablice frekvencija moguće je zatražiti samo za varijable nominalnog i ordinalnog tipa.

U lijevom okviru popis je dostupnih varijabli. One varijable za koje želimo izračunati tablice frekvencija i deskriptivnu statistiku valja prebaciti u desni okvir.

- Niz uzastopnih varijabli najlakše je selektirati tako da kliknemo mišem na prvu varijablu u nizu, pritisnemo i držimo tipku Shift te potom kliknemo na posljednju varijablu u nizu.
- Želimo li odabrati više varijabli koje nisu uzastopne, to ćemo najlakše učiniti tako da kliknemo mišem na prvu, potom pritisnemo i držimo tipku Control te nastavimo s odabirom varijabli.

Opcijom **Frequency tables** odabiremo želimo li u ispisu rezultata tablice frekvencija ili želimo dobiti samo statističke pokazatelje, bez tablica frekvencija. Ova opcija nije inicijalno uključena.

Opcijom **Plots** odabiremo želimo li – i kako – grafički prikazati distribucije selektiranih varijabli.

Opcijom **Statistics** određujemo koje deskriptivne statističke pokazatelje želimo.

U tablici **Frequency Tables** može se razlikovati ukupan broj ispitanika za koje su na ovoj varijabli izračunate frekvencije (Total = 165) te broj ispitanika koji nemaju valjanu vrijednost ili uopće nemaju vrijednost na toj varijabli (tzv. nedostajuće vrijednosti (engl. *missing values*) – u ovom ih slučaju imamo 4).

- Stupac **Frequency** daje frekvencije (učestalost pojedinih vrijednosti) varijable.
- Stupac **Percent** iskazuje učestalost pojedinih vrijednosti u postocima, u odnosu na ukupan broj ispitanika.
- Stupac **Valid Percent** iskazuje učestalost pojedinih vrijednosti u postocima, u odnosu na broj ispitanika koji na varijabli imaju valjanu vrijednost (ovdje: $165 - 4 = 161$).
- Stupac **Cumulative Percent** daje kumulativne postotke (izračunate na temelju valjanih postotaka).

2.2. Deskriptivni statistički pokazatelji

Želimo li u ispisu rezultata dobiti – osim tablica frekvencija – i deskriptivnu statistiku, valja odabrati opciju **Statistics**, gdje možemo odabrati željene statističke pokazatelje.

▼ Statistics

Percentile Values

- Quartiles
- Cut points for: 4 equal groups
- Percentiles:

Central Tendency

- Mean
- Median
- Mode
- Sum

Dispersion

- S. E. mean
- Std. deviation
- MAD
- MAD Robust
- IQR
- Variance
- Range
- Minimum
- Maximum

Distribution

- Skewness
- Kurtosis
- Shapiro-Wilk test

Pozicijske mjere distribucije frekvencija

Percentile Values (vrijednosti percentila):

- **Quartiles** (kvartili) – ispisuju se vrijednosti 25., 50. i 75. percentila;
- **Cut points for o equal groups** (granične vrijednosti za o jednakih skupina) valja upisati u koliko jednakih skupina (jednakih prema broju ispitanika u njima) želimo podijeliti ispitanike; u ispisu rezultata daju se granične vrijednosti na temelju kojih se oblikuju specificirane skupine ispitanika (npr. upišemo li da želimo 10 jednakih skupina, u ispisu rezultata dobit ćemo, zapravo, vrijednosti decila);
- **Percentiles** (percentili) – nakon aktiviranja ove opcije valja upisati percentile za koje želimo dobiti vrijednosti u ispisu rezultata; željene percentile valja razdvojiti zarezom.

Mjere centralne tendencije – opisivanje gomilanja podataka u kvantitativnim varijablama

Central Tendency (mjere centralne tendencije):

- **Mean** – aritmetička sredina (prosječna vrijednost);
- **Median** – medijan (srednja vrijednost);
- **Mode** – mod (tipična ili najčešća vrijednost);
- **Sum** – zbroj vrijednosti svih ispitanika na određenoj varijabli (ne računajući one koji na toj varijabli nemaju valjanu vrijednost).

Mjere disperzije – opisivanje raspršenja podataka oko mjera centralne tendencije

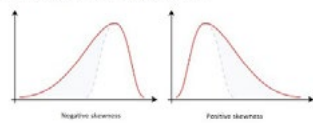
Dispersion (mjere disperzije):

- **S.E. mean** (engl. *standard error of the mean*) – standardna pogreška aritmetičke sredine;
- **Std. deviation** – standardna devijacija;
- **MAD** (engl. *median absolute deviation*) – srednje apsolutno raspršenje, mjera disperzije relativno neosjetljiva na nenormalnost distribucija. Izvještavanje o Mdn +/- MAD za podatke koji se ne distribuiraju normalno ekvivalentno je izvještavanju o aritmetičkoj sredini +/- SD za normalno distribuirane podatke;
- **MAD Robust** – srednje apsolutno odstupanje podataka, prilagođeno faktorom za asimptotsku normalnu konzistenciju;
- **IQR** (engl. *interquartile range*) – interkvartilni raspon;
- **Variance** – varijanca;
- **Range** – raspon varijacija (razlika između najveće i najmanje registrirane vrijednosti varijable);
- **Minimum** – najmanja (registrirana) vrijednost varijable;
- **Maximum** – najveća (registrirana) vrijednost varijable.

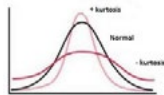
Mjere oblika distribucije – asimetrije i zaobljenosti

DISTRIBUTION

Skewness describes the shift of the distribution away from a normal distribution. Negative skewness shows that the mode moves to the right resulting in a downturned left tail. Positive skewness shows that the mode moves to the left resulting in a downturned right tail.



Kurtosis describes how heavy or light the tails are. Positive kurtosis results in an increase in the "peakedness" of the distribution with heavier (longer) tails while negative kurtosis exhibit a much more uniform or flatter distribution with light (shorter) tails.



Izvor: Goss-Sampson (2019:15)

Distribution (oblik distribucije – asimetričnost i zaobljenost distribucije):

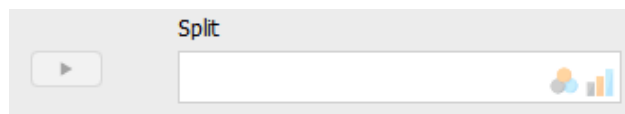
- **Skewness** – mjera asimetrije distribucije; za normalnu distribuciju ima vrijednost 0; za lijevo asimetrične distribucije (dugačak desni "rep" distribucije; točka gomilanja rezultata nalazi ispod sredine raspona varijacija) ima pozitivne, a za desno asimetrične distribucije (dugačak lijevi "rep" distribucije; točka gomilanja rezultata nalazi iznad sredine raspona varijacija) negativne vrijednosti;
- **Kurtosis** – mjera zaobljenosti distribucije; za normalnu distribuciju ima vrijednost 0; pozitivne vrijednosti karakteriziraju šiljate (leptokurtične), a negativne vrijednosti karakteriziraju spljoštene (platikurtične) distribucije. U ispisu rezultata se uz oba statistika daje njihova standardna pogreška.
- **Shapiro-Wilk** – test normalnosti distribucije, nulta hipoteza glasi da je distribucija varijable u populaciji normalna; p-vrijednost manja od odabrane razine rizika (npr. 0,05 za 5% rizika) znači da distribucija varijable čiji oblik testiramo, u populaciji (vrlo vjerojatno) nije normalna.

Descriptive Statistics

	Seksizam_ad
Valid	157
Missing	8
Mean	9.191
Std. Deviation	2.507
Skewness	-0.274
Std. Error of Skewness	0.194
Kurtosis	-0.227
Std. Error of Kurtosis	0.385
Shapiro-Wilk	0.977
P-value of Shapiro-Wilk	0.009

2.3. Deskriptivna statistika zasebno za svaku skupinu

U proceduri 'Descriptive statistics' također je korisna funkcija 'Split' pomoću koje možete prikazati zatražene frekvencije, grafike i deskriptivne parametre s obzirom na kategorije neke nominalne ili ordinalne varijable.



— primjer: deskripcija rezultata i tablice frekvencija rezultata na varijabli *Seksizam1* prema *spolu*:

Descriptive Statistics

Descriptive Statistics	Seksizam1	
	Muski	Zenski
Valid	81	80
Missing	2	2
Mean	2.469	2.200
Std. Deviation	1.130	1.060
Minimum	1.000	1.000
Maximum	5.000	5.000

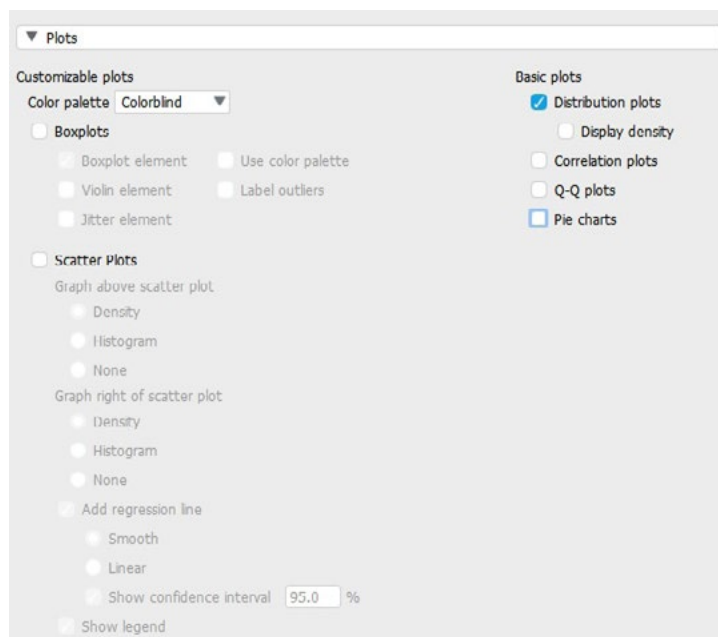
Frequency Tables

Frequencies for Seksizam1

Spol	Seksizam1	Frequency	Percent	Valid Percent	Cumulative Percent
Muski	uopće se ne slažem	17	20.482	20.988	20.988
	ne slažem se	31	37.349	38.272	59.259
	ne znam	13	15.663	16.048	75.309
	slažem se	18	21.687	22.222	97.531
	potpuno se slažem	2	2.410	2.489	100.000
	Missing	2	2.410		
Total		83	100.000		
Zenski	uopće se ne slažem	24	29.268	30.000	30.000
	ne slažem se	29	35.366	36.250	66.250
	ne znam	15	18.293	18.750	85.000
	slažem se	11	13.415	13.750	98.750
	potpuno se slažem	1	1.220	1.250	100.000
	Missing	2	2.439		
Total		82	100.000		

2.4. Grafičko prikazivanje podataka

Želimo li grafički prikazati distribucije frekvencija, valja odabrati opciju Plots, gdje možemo odabrati željeni tip grafičkog prikaza.



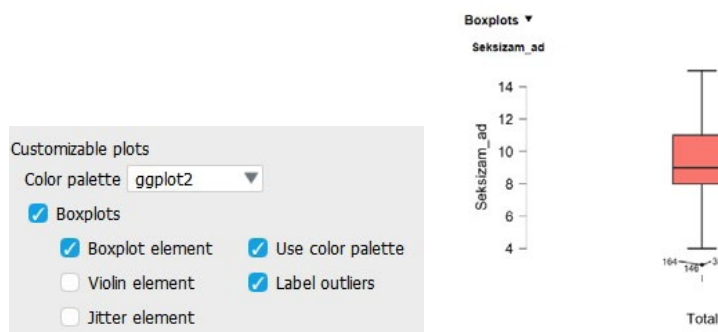
Kvantitativne varijable (univarijatno)

Boxplot – zove se još i engl. *box-and-whisker plot* te hrv. *kutijasti dijagram*. Prikazuje odnos minimuma, maksimuma, donjega i gornjega kvartila te medijana podataka. Može se prikazati samo za kvantitativne varijable.

Zadani izgled:



Dodatne opcije (boje i oznake ekstremnih vrijednosti):

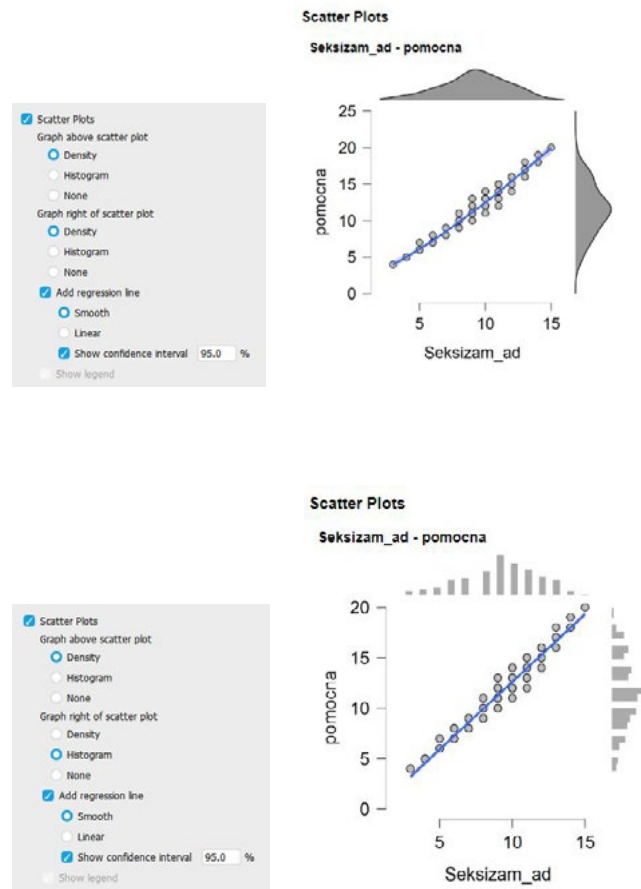


Odabirom opcije *Label outliers*, na grafičkom prikazu pojavit će se redni brojevi redaka u kojima se nalaze podaci prepoznati kao ekstremne vrijednosti.

Kvantitativne varijable (bivarijatno)

Scatter Plots – dijagrami raspršenja, prikazuju vrijednosti podataka dviju isključivo kvantitativnih varijabli na zajedničkom grafičkom prikazu, uz različite mogućnosti prikaza koje se namještaju za svaku varijablu.

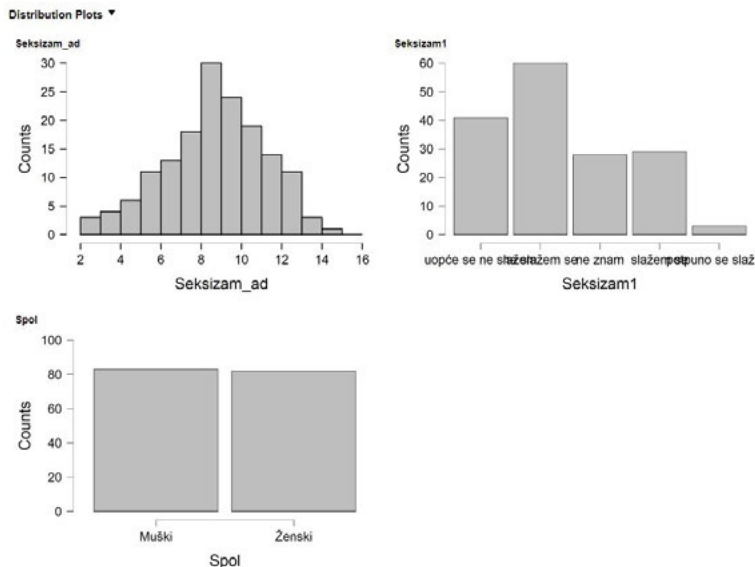
— Primjer: varijable *Seksizam_ad* i *Pomocna* s različitim opcijama:



**Sve varijable
(univarijatno)**

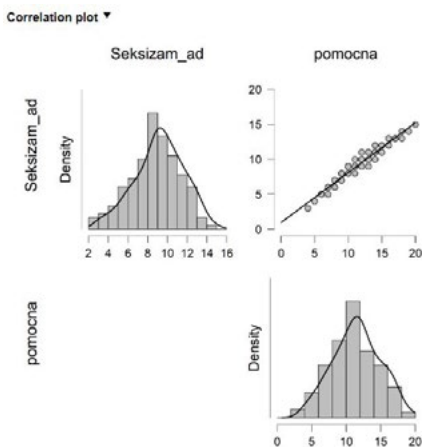
Basic Plots

- **Distribution plots** – stupčasti prikaz frekvencija podataka za sve vrste varijabli, ovdje tri varijable: *Seksizam_ad* (kvantitativna), *Seksizam1* (ordinalna) i *spol* (nominalna):

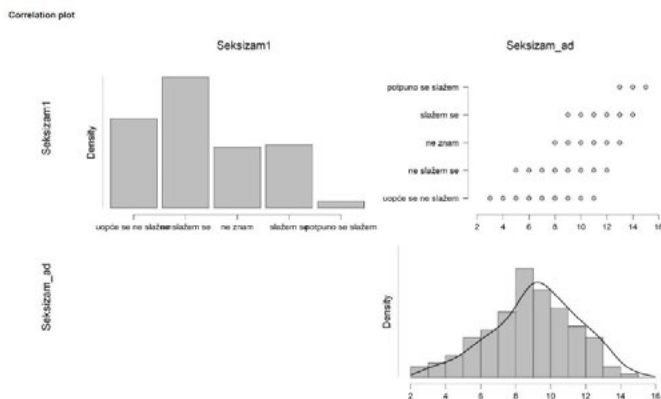


**Sve varijable
(bivarijatno)**

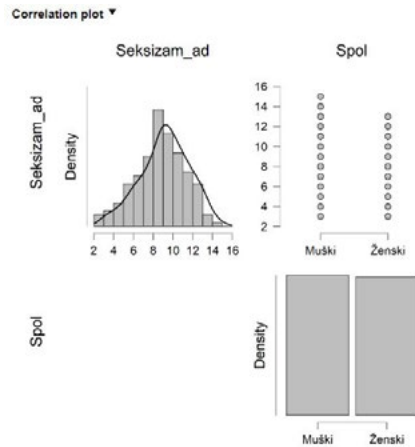
- **Correlation plots** – prikazuju dijagram raspršenja dviju varijabli (bilo koje vrste) uz zasebne grafičke prikaze svake od varijabli. — Primjer: *Correlation plots* za varijable *Seksizam_ad* i *Pomocna* (obje kvantitativne):



- Primjer: *Correlation plots* za varijable *Seksizam1* i *Seksizam_ad* (ordinalna i kvantitativna):



— Primjer: *Correlation plots* za varijable *Seksizam_ad* i *spol* (kvantitativna i nominalna):



Kvantitativne varijable (univarijatno – testiranje na normalnost)

- **Q-Q plots** – Q-Q grafikon (skraćeno od engl. *quantile-quantile**) može se koristiti za vizualnu procjenu potječe li skup podataka iz normalne distribucije. Q-Q grafikoni uzimaju podatke uzorka, sortiraju ih uzlaznim redoslijedom, a zatim ih crtaju prema kvantilima (percentilima) izračunatim iz teorijske raspodjele. Ako se podaci normalno distribuiraju, točke će pasti na referentnu liniju od 45 stupnjeva ili blizu nje. Ako se podaci ne distribuiraju normalno, točke će odstupiti od referentne linije.

Q-Q grafikon dijagram je raspršenja s teoretskim kvantilima normalne distribucije na osi X i istim kvantilima reziduala uzorka na osi Y. Služi za usporedbu oblika distribucije varijable na uzorku s normalnom teorijskom raspodjelom. Kada su podaci normalno distribuirani, tada se točke nalaze na dijagonalnoj referentnoj liniji ili vrlo blizu nje. Odstupanja od normalne distribucije izgledna su ako uočimo sljedeće pomake točaka od dijagonalne referentne linije:

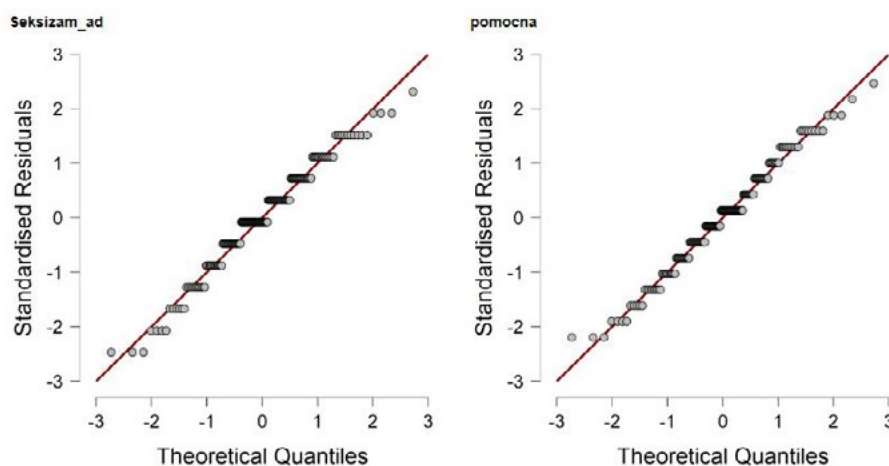
- 1) Ako samo nekoliko točaka ne leži na pravcu, to ukazuje na mogućnost ekstremnih vrijednosti (engl. *outlier*).
- 2) Ako su točke pomaknute iznad ili ispod linije, to ukazuje na odstupanje u zakrivljenosti distribucije u odnosu na normalnu.
- 3) Ako se točke raspršuju "zmijoliko" oko linije, to ukazuje na asimetričnost distribucije.

* Kvantili su točke koje dijele distribuciju frekvencija u intervale s jednakim vjerojatnostima (tj. jednakom količinom podataka ili jednakim površinama ispod distribucije). Vrste kvantila koje poznajete: kvantili, decili, percentili.

— Primjer: Q-Q grafikoni za varijable *Seksizam_ad* (Shapiro-Wilk $p = 0,009 \rightarrow$ distribucija statistički značajno odstupa od normalne) i *pomocna* (Shapiro-Wilk $p = 0,094 \rightarrow$ distribucija ne odstupa statistički značajno od normalne).

Descriptive Statistics		
	Seksizam_ad	pomocna
Valid	157	157
Missing	8	8
Mean	9.191	11.529
Std. Deviation	2.507	3.434
Skewness	-0.274	-0.011
Std. Error of Skewness	0.194	0.194
Kurtosis	-0.227	-0.457
Std. Error of Kurtosis	0.385	0.385
Shapiro-Wilk	0.977	0.985
P-value of Shapiro-Wilk	0.009	0.094

Q-Q Plots ▼



- **Pie charts** – strukturni krug ili pita, služi za prikaz kvalitativnih varijabli (nominalnih i ordinalnih)

— primjer: *pie chart* za varijablu *Seksizam1*:

Seksizam1



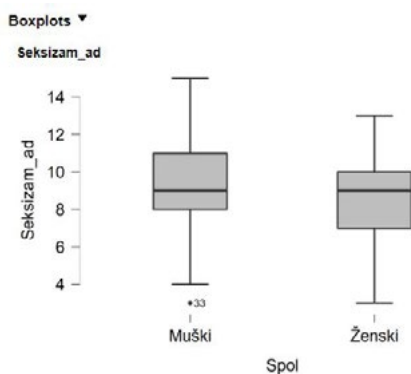
2.5. Grafički prikazi zasebno za svaku skupinu

Korištenjem opcije **Split** mogu se prikazati odvojeni grafikoni za svaku grupu odabrane nominalne varijable:

— primjer: *distribution plots* za varijablu *Seksizam1* prema spolu:



— primjer: *boxplots* za varijablu *Seksizam_ad* prema spolu:



Literatura

- Goss-Sampson, M. A. (2019). Statistical Analysis in JASP 0.10.2: A Guide for Students, poglavlje DESCRIPTIVE STATISTICS (str. 12-20). URL: <https://static.jasp-stats.org/Statistical%20Analysis%20in%20JASP%20-%20A%20Students%20Guide%20v0.10.2.pdf>
- Navarro, Danielle J., Foxcroft, David R. i Faulkenberry Thomas J. (2019). Learning Statistics with JASP: A Tutorial for Psychology Students and Other Beginners, poglavlje 4: Descriptive statistics. URL: <https://tomfaulkenberry.github.io/JASPbook/chapters/chapter4.pdf>