

Ponovitev snovi s predavanj

1. Vrste spremenljivk

- Izpolnite tabelo s primeri ustreznih spremenljivk z vašega področja. Ob vsakem primeru zapišite nekaj vrednosti, ki jih navedena spremenljivka zavzame ter skicirajte ustrezen grafični prikaz.

Tip spremenljivke		Primer	Grafični prikaz

- Kako rečemo opisnim spremenljivkam, ki imajo samo dve vrednosti (navedite tudi primer)?

2. Mere središčnosti in variabilnosti

Oglejte si spletno anketo (https://docs.google.com/spreadsheets/viewform?usp=drive_web&formkey=dGY4MWxXTU5nRDQ5c01jdUVSTjZRUnc6MQ#gid=0). Na anketo je odgovorilo 140 študentov (iz 1.letnika veterinarske fakultete in 2. letnika medicinske fakultete).

- Kaj so statistične enote v tem primeru?
- Koliko jih je?
- Število spremenljivk je:

Teža prvih 6 študentov, ki so odgovorili na anketo je bila: **62, 62, 58, 71, 52, 55** (v kg).

- Izpolnite tabelo (peš in s pomočjo programa R)

Povprečje	Mediana	Modus	Standardni odklon	Razpon	Prvi kvartil

- Katere od teh mer so primerne za posamezne vrste spremenljivk?

Osnove dela s statističnim programom R in opisna statistika

1. Osnove dela z R-om - vnašanje podatkov

Poženite program R (z dvoklikom na ikono njegove bližnjice na namizju ali preko menija Start). Ustvarite novo datoteko s podatki o porodih, ki naj za tri statistične enote (primere) vsebuje podatke o štirih statističnih spremenljivkah:

- porodna teža otroka (PTO) - vrednost v gramih (ponavadi med 2000 in 4500)
- spol otroka (SPOL) - oznaka 1 za dečka in 2 za deklico
- gestacijska starost (GESTAC) - vrednost v tednih (ponavadi med 35 in 42)
- starost matere ob porodu (STAROST) - vrednost v letih (ponavadi med 16 in 45)

2. Osnove dela z R-om - opisna statistika

Odprite datoteko **Ankete1011.txt** (podatki ankete izpred nekaj let)

```
mydata <- read.table("Ankete1011.txt", header = T, dec = ",", sep = "\t", fill = T)
```

Pozor: podatki morajo biti v delovnem direktoriju!). Opisne statistike za spremenljivke lahko izračunate s tem postopkom

```
summary(mydata)
```

in si jih lahko ogledate tudi grafično

```
hist(mydata$Starost)
```

```
barplot(table(mydata$Spol)/sum(table(mydata$Spol))*100, ylab = "Relativna frekvenca")
```

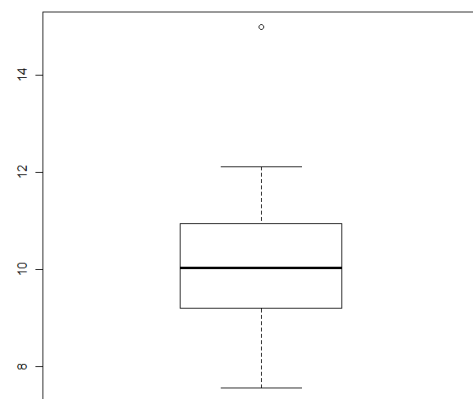
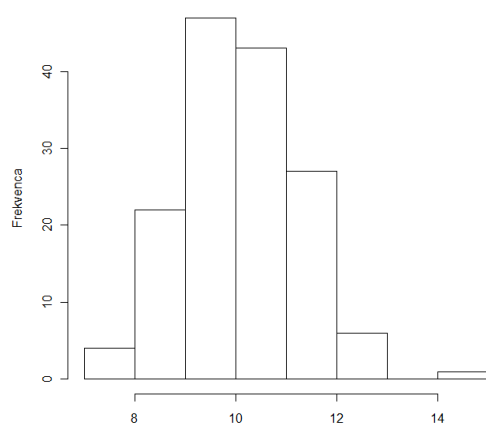
Dopolnite spodnji tabeli (kjer je to smiselno)

Spremenljivka	Število različnih vrednost	Vrsta spremenljivke	Najnižja vrednost	Najvišja vrednost	Smislen grafični prikaz
barva oči					
igrice					
Kajenje					
Internet					
Domače živali					

	<i>internet</i>	<i>domače živali</i>	<i>barva oči</i>
Povprečje	$\bar{x} =$	$\bar{x} =$	$\bar{x} =$
Mediana	$Me =$	$Me =$	$Me =$
Modus	$Mo =$	$Mo =$	$Mo =$
Razpon	max - min =	max - min =	max - min =
standardni odklon	$s =$	$s =$	$s =$
Varianca	$s^2 =$	$s^2 =$	$s^2 =$
interkvartilni razmik	$IQR =$	$IQR =$	$IQR =$

- Skica grafičnih prikazov za *internet* in *barva oči*:

3. Histogram in okvir z ročaji



```
hist(x, ...)
boxplot(x, ...)
```

- Kaj pomeni, da je številska spremenljivka zvezna?
- Izmislite si primer zvezne spremenljivke s prikazano porazdelitvijo.

Odčitajte oz. izračunajte navedene količine (v vzorcu je 150 enot). Pripišite, iz katerega grafa ste dobili rezultat.

- Mediano:
- Prvi kvartil:
- Delež primerov z vrednostjo do 11:
- Število primerov z vrednostjo do 11:
- Število primerov z vrednostjo med 11 in 12:
- Interkvartilni razmik:
- Najmanjšo vrednost:
- Največjo vrednost:
- Razpon:
- Približno ocenite povprečje:
- Približno ocenite standardni odklon:

Iz histograma ocenite še:

- kakšna je relativna frekvenca v najpogostejšem razredu (v vzorcu je 150 enot):
- kakšna je kumulativna frekvenca v najpogostejšem razredu:
- kakšna je kumulativna frekvenca v zadnjem razredu?
- Ali obstaja kakšen razred, kjer sta frekvenca in kumulativna frekvenca enaki?

4. Asimetrična porazdelitev

V tabeli so podane plače delavcev v podjetju Stečajko d.d.

Plača	600	900	1.800	15.000	30.000
Število delavcev	20	12	5	3	1

- Prikažite porazdelitev s stolpčnim diagramom (uporabite R).
- Izračunajte povprečno plačo:
- Izračunajte mediano plače:
- Kaj se zgodi s povprečjem oz. mediano, če si direktor poviša plačo?
- Skicirajte histogram, ki prikazuje porazdelitev plače.
- Kakšne so razlike med zgornjim stolpčnim diagramom in histogramom? Katerega bi uporabili? Zakaj?
- Porazdelitev plač je močno **asimetrična v desno**. Ugotovite, kaj se dogaja z mediano in povprečjem v asimetričnih porazdelitvah.
- Kako bi v tem primeru izgledal okvir z ročaji? Skicirajte ga!

- Navedite primer vsaj ene spremenljivke (realen), ki ima povprečje manjše kot mediano, ter skicirajte njeno porazdelitev (ne pozabite označiti osi)!

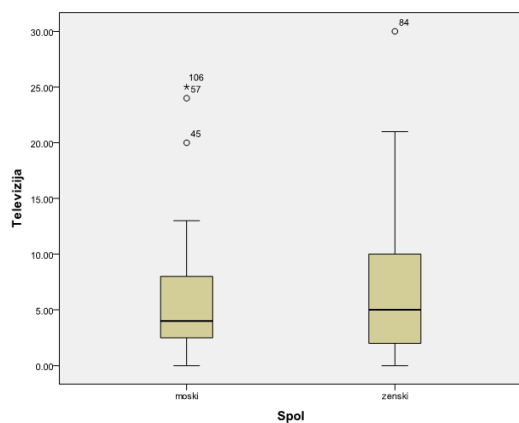
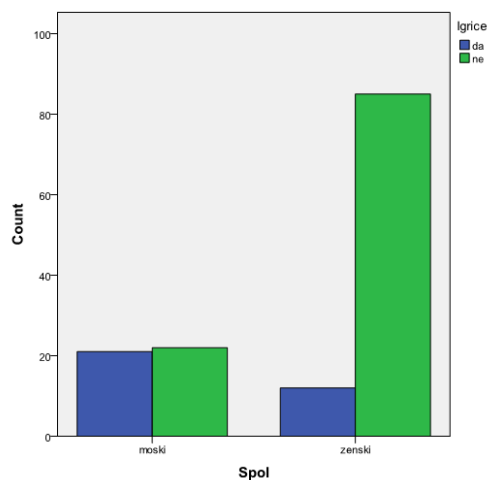
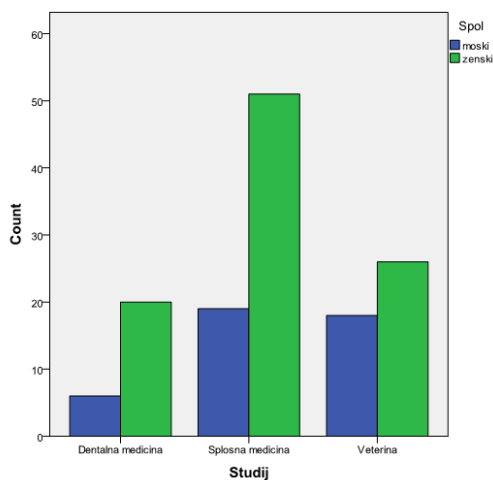
5. Primerjava podskupin

Grafično si oglejte porazdelitev spremenljivke višina za vse študente skupaj.

- Kaj lahko rečete o obliki porazdelitve?
- Oglejte si še porazdelitev višine glede na spol in komentirajte.

6. Obnašanje podatkov na vzorcu

- Oglejte si spodnje grafikone in jih komentirajte. Prikazali smo podatke našega vzorca; ali mislite, da se bodo tudi letos pojavile take razlike med spoloma (so letošnji študentje vzorec iz iste populacije, ali so razlike dovolj velike, da ne morejo biti plod naključja)



7. Izračun novih spremenljivk

Izračunajte indeks telesne teže: $ITM = [\text{teža (v kg)}] / [\text{višina}^2 \text{ (v m)}]$. Izračunajte tudi primerne opisne statistike in si porazdelitev grafično prikažite.

Povzetek

1. Katere opisne statistike je smiselno izračunati za opisne spremenljivke in katere za številske spremenljivke (obkljukajte)?

Statistika	Imenske opisne spremenljivke	Številske spremenljivke
Povprečje		
Standardni odklon		
Varianca		
Mediana		
Modus		
Interkvartilni razmik		
Razpon		

2. Histogram in stolpčni diagram

- Navedite razlike med stolpčnim diagramom in histogramom.
- Kako iz stolpčnega diagrama izračunamo povprečje?
- Kako iz stolpčnega diagrama izračunamo mediano?