

1. V datoteki `popis_2002.txt` imamo podane frekvence starosti državljanov Republike Slovenije iz popisa leta 2002.

(a) Nariši frekvenčne poligone. Opazimo, da so združeni ljudje, ki so starejši od 85 let, iz podatkov odstrani to skupino ljudi.

```
> popis=popis[1:85,1:4]
> popis=read.table("popis_2002.txt", header=TRUE)
> names(popis)
[1] "STAROST" "SKUPAJ" "MOSKI" "ZENSKE"
> plot(popis$SKUPAJ,type="l")
> dim(popis)
[1] 86 4
> popis = popis[1:85,1:4]
> plot(popis$SKUPAJ,type="l")
```

(b) Določi povprečno starost vseh prebivalcev in določi standardni odklon (vseh, moških, žensk).

```
> sum(popis$STAROST*popis$SKUPAJ)/sum(popis$SKUPAJ)
[1] 38.4422
> sum(popis$STAROST*popis$MOSKI)/sum(popis$MOSKI)
[1] 36.90545
> povp = sum(popis$STAROST*popis$MOSKI)/sum(popis$MOSKI)
> sd = sqrt(sum(popis$STAROST^2*popis$SKUPAJ/sum(popis$SKUPAJ)) - povp^2)
> sd
[1] 23.76707
```

(c) Oglej si `help(weighted.mean)` in izračunaj povprečja iz prejšnje točke.

```
> ?weighted.mean
> weighted.mean(popis$STAROST,popis$SKUPAJ)
[1] 38.4422
```

(d) Iz celotne populacije ljudi izberi enostavno slučajni vzorec stotih oseb. Nariši nekaj histogramov. Izračunaj povprečno vrednost in standardni odklon.

```
> p = popis$SKUPAJ / sum(popis$SKUPAJ)
> v = popis$STAROST
> sample(v,100,prob=p,replace=TRUE)
[1] 7 21 46 12 17 4 42 45 7 44 ...
> s100 = sample(v,100,prob=p,replace=TRUE)
> hist(s100)
> hist(s100,breaks=100)
> hist(s100,breaks=20)
> mean(s100)
[1] 37.28
> sd(s100)
[1] 21.28122
```

- (e) Nariši histograme za vzorčno povprečje 10/100/1000/5000 slučajno enostavnih vzorcev velikosti $n = 100$. Kaj opaziš? Poračunaj še povprečje in standardne odklone teh vzorčenih povprečij.

```
> vzorec10=c()
> for (i in 1:10) vzorec10 = c(vzorec10,mean(sample(v,100,prob=p,
replace=TRUE)))
> vzorec10
 [1] 40.49 37.67 37.06 ...
> hist(vzorec10,breaks=20)
> vzorec100 = c()
> for (i in 1:100) vzorec100 = c(vzorec100,mean(sample(v,100,prob=p,
replace=TRUE)))
> hist(vzorec100,breaks=20)
> vzorec1000 = c()
> for (i in 1:1000) vzorec1000 = c(vzorec1000,mean(sample(v,100,prob=p,
replace=TRUE)))
> hist(vzorec1000,breaks=20)
> vzorec5000 = c()
> for (i in 1:5000) vzorec5000 = c(vzorec5000,mean(sample(v,100,prob=p,
replace=TRUE)))
> hist(vzorec5000,breaks=20)
```

2. Podatki o sončevih pegah med leti 1700-1988 so spravljani v podatkovni bazi sunspots.year. Z ukazom `?sunspots.year` si oglej kaj ti podatki predstavljajo.

- (a) Izpiši podatke o sončevih pegah in jih grafično prikaži.
 (b) Prikaži histogram podatkov.
 (c) Izračunaj povprečno število sončevih peg na leto.
 (d) Preberi `help(sd)` in izračunaj standardni odklon.

```
> ?sunspots.year
> sunspots.year
> plot(sunspot.year)
> hist(sunspot.year)
> length(sunspot.year)
 [1] 289
> sum(sunspot.year)/289
 [1] 48.61349
> mean(sunspot.year)
 [1] 48.61349
> ?sd
> std_odklon = (289-1)/289*sd(sunspot.year)
> std_odklon
 [1] 39.33751
```

3. V datoteki `ocene_ovs.txt` so zbrane ocene rezultatov 1. kolokvija OVS 2010.

- (a) Prikaži frekvenčno tabelo ocen in nariši histogram.
- (b) Izračunaj povprečno oceno in mediano.
- (c) Izračunaj 0.1-kvantil in vse tri kvartile.
- (d) Izbriši študente, ki so dosegli manj kot 5 točk in ponovi vajo.

```
> ocene=read.table("ocene_ovs.txt", header=TRUE)
> names(ocene)
[1] "Student" "Ocena"
> oc = ocene$Ocena
> oc
> table(oc)
> hist(oc,20)
> hist(oc,breaks=20)
> mean(oc)
[1] 41.27805
> median(oc)
[1] 42
> quantile(oc,probs=0.1)
10%
 7
> quantile(oc,c(0.25,0.5,0.75))
25% 50% 75%
20 42 62
> summary(oc)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00  20.00   42.00   41.28  62.00  100.00
> ocs = sort(oc)
> ocs
> length(ocs)
[1] 205
> ocs = ocs[20:205]
> summary(ocs)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 7.00  28.25   45.00   45.25  63.75  100.00
```

4. Populacija šteje 3000 študentov (moških). Predpostavimo, da so njihove višine porazdeljene s povprečjem 180 cm in standardnim odklonom 5 cm. Naključno izberemo vzorec 25 študentov in izmerimo povprečno višino vzorca.

- Določi matematično upanje za povprečno višino vzorca.
- Določi standardni odklon za povprečno višino vzorca.
- Oceni verjetnost, da bo imel naključno izbrani vzorec povprečje višin

- med 178 in 182 cm.
- več kot 183 cm.
- manj kot 175 cm.

```
> E = 180
> D = 5^2*(3000-25)/(25*(3000-1))
> sd = sqrt(D)
> sd
> pnorm(182,180,5/sqrt(25))-pnorm(178,180,5/sqrt(25))
[1] 0.9544997
> 1-pnorm(183,180,5/sqrt(25))
[1] 0.001349898
> pnorm(175,180,5/sqrt(25))
[1] 2.866516e-07
```

5. Vzorec doseženih točk pri maturi je 12, 18, 21, 25, 26, 27, 27, 30, 33. Izračunaj vzorčno povprečje, vzorčno disperzijo in popravljeno vzorčno disperzijo.

```
> v = c(12, 18, 21, 25, 26, 27, 27, 30, 33)
> mean(v)
[1] 24.33333
> s = 8/9*sd(v)
> sh = sd(v)
> s
[1] 5.691666
> sh
[1] 6.403124
```