Osnove verjetnosti in statistike FRI - vaje 10

Naloga 1. Slučajna spremenljivka X je na večtisočglavi populaciji porazdeljena normalno s povprečjem μ in standardnim odklonom σ. Vzeli smo vzorec moči 50 in s pomočjo njega izračunali vzorčno povprečje $\bar{X} = 12,24$ in popravljen vzorčni standardni odklon $\hat{S} = 0,6$. Določite intervala zaupanja za μ pri stopnjah zaupanja $0,95$ in $0,90$.

```
> n=50
> s=0.6
> m=12.24
> gamma=0.95
> a=qnorm((1+gamma)/2)*s/sqrt(n)
> interval=c(m-a,m+a)
> interval
[1] 12.07369 12.40631
>
> gamma=0.9
> ...
```

Naloga 2. S pomočjo vzorca moči 20 smo izračunali popravljen vzorčni standardni odklon $\hat{S} = 0,7$. Poišči intervala zaupanja za standardni odklon σ za stopnji zaupanja $0,95$ in $0,99$.

```
> n=20
> s=0.7
> gamma=0.95
> h1=qchisq((1-gamma)/2,df=n-1)
> h2=qchisq((1+gamma)/2,df=n-1)
> interval=c(s*sqrt(n-1)/sqrt(h2),s*sqrt(n+1)/sqrt(h2))
> interval
[1]0.5323433 1.022400
>
> gamma=0.99
...
```

Naloga 3. Na veliki populaciji smo dvanajstkrat izmerili količino X in pri tem dobili naslednje vrednosti.

$$-0,5 \quad 1,2 \quad 0 \quad 0,8 \quad 1,2 \quad -0,4 \quad 0,2 \quad -0,2 \quad 1,5 \quad 0,6 \quad -0,4 \quad 1$$

Izračunaj vzorčno povprečje in popravljen vzorčni standardni odklon ter nato izračunaj interval zaupanja za $E(X)$ s stopnjo zaupanja 0.96. Zaradi majhnosti vzorca uporabi Studentovo porazdelitev.

```
> v=c(-0.5, 1.2, 0, 0.8, 1.2, -0.4, 0.2, -0.2, 1.5, 0.6, -0.4, 1)
> m=mean(v)
> s=sd(v)
> n=length(v)
> gamma=0.96
> a=qt((1+gamma)/2,df=n-1)*s/sqrt(n)
> interval=c(m-a,m+a)
> interval
```

```
[1] -0.06698479 0.90031812
>
> t.test(v,conf.level=0.96)
...
```

Naloga 4. V enostavnem slučajnem vzorcu 200 državljanov Slovenije je bilo 183 desničarjev, ostali levičarji. Oceni delež levičarjev med Slovenci s 95% natančnostjo, ob predpostavki, da je bilo v času vzorčenja 2050000 državljanov.

```
> n=200
> N=2050000
> p=17/200
> s=sqrt(p*(1-p)/n)*sqrt((N-n)/(N-1))
> gamma=0.95
> a=qnorm((1+gamma)/2)*s/sqrt(n)
> interval=c(p-a,p+a)
> interval
[1] 0.08226714 0.08773286
>
```

Naloga 5. Oglej si dokumentacijo v R vgrajenih podatkov pod imenom "faithful", ki prikazujejo statistiko izbruhov znanega gejzira v Yellowstonskem narodnem parku v zvezni državi Wyoming, ZDA.

- Izračunaj vzorčni povprečji in popravljena vzorčna standardna odklona za čase med posameznimi izbruhi ter za čase trajanja izbruhov.
- Oceni povprečni čas trajanja izbruha z 90% gotovostjo in povprečni čas med izbruhi x 99% gotovostjo.

```
> ?faithful
> izbruhi=faithful$eruptions
> pavze=faithful$waiting
> c(mean(izbruhi),sd(izbruhi))
[1] 3.487783 1.141371
> c(mean(pavze),sd(pavze))
[1] 70.89706 13.59497
>
> t.test(izbruhi,conf.level=0.9)
...
> t.test(pavze,conf.level=0.99)
...
```

Naloga 6. Spodaj so naštete dolžine repov kuščarjev zelencev v vzorcu.

6,2	6,6	7,1	7,4	7,6	7,9	8	8,3
8,4	8,5	8,6	8,8	8,8	9,1	9,2	9,4
9,4	9,7	9,9	10,2	10,4	10,8	11,3	11,9

- Izračunaj vzorčno povprečje in popravljen vzorčni standardni odklon za dolžine repov kuščarjev zelencev.

- Predpostaviš lahko, da so dolžine repov kuščarjev zelencev normalno porazdeljene z enakim povprečjem in standardnim odklonom, kot si ga izračunal pri prejšnji točki naloge. Pripravi si seznam dolžine 10^6 z elementi, ki bodo enako porazdeljeni kot dolžine repov kuščarjev. Ta seznam naj predstavlja populacijo. Izračunaj povprečje μ in standardni odklon σ te populacije.
- Pripravi si funkciji, ki bosta sprejeli vzorec in iz njega izračunali intervala zaupanja za povprečje, ena uptevajoč normalno distribucijo, druga pa Studentovo t-distribucijo.
- Iz populacije, zgenerirane prej, izberi 50 vzorcev velikosti 10. Izračunaj intervale zaupanja za povprečje pri 95% gotovosti, enkrat s predpostavko normalne in drugič s predpostavko Studentove distribucije. Intervale grafično prikaži. Koliko intervalov, izračunanih s predpostavko normalne distribucije, "zgreši" pravo povprečje populacije? Kaj pa intervali po Studentovi t-distribuciji? Komentiraj rezultat.

Po priporočilih Polone naj bi to nalogo naredili na vseh vajah.

```
> lizard=c(6.2, 6.6, 7.1, 7.4, 7.6, 7.9, 8, 8.3, 8.4, 8.5, 8.6, 8.8,
+ 8.8, 9.1, 9.2, 9.4, 9.4, 9.7, 9.9, 10.2, 10.4, 10.8, 11.3, 11.9)
> mul=mean(lizard)
> sdl=sd(lizard)
>
> pop=rnorm(10^6,mean=mul,sd=sdl)
> m=mean(pop)
> s=sd(pop)
>
> confIntN=function(smpl,prob=0.95)
+ { n=length(smpl)
+   m=mean(smpl)
+   s=sd(smpl)
+   a=qnorm((1+prob)/2)*s/sqrt(n)
+   c(m-a,m+a) }
>
> confIntT=function(smpl,prob=0.95)
+ { n=length(smpl)
+   m=mean(smpl)
+   s=sd(smpl)
+   a=qt((1+prob)/2)*s/sqrt(n)
+   c(m-a,m+a) }
>
> intN=matrix(nrow=50,ncol=10)
> for (i in 1:50) intN[i,]=sample(pop,10,replace=T)
> intT = intN
> confN=apply(intN,1,confIntN)
> confT=apply(intT,1,confIntT)
>
> plot(range(confN), c(0,51), type="n")
> for (i in 1:50) lines(confN[,i], c(i,i), lwd=2)
```

```
> abline(v=m,lwd=2,lty=2)
> sum(confN[1,]<=m & confN[2,]>=m)
[1] 47
>
> plot(range(confT), c(0,51), type="n")
> for (i in 1:50) lines(confT[,i], c(i,i), lwd=2)
> abline(v=m,lwd=2,lty=2)
> sum(confT[1,]<=m & confN[2,]>=m)
[1] 49
...
```