# Statistika

SIIT / IIS / SIIT LO

školska 2016/17

# Vežbe 6

## Pokretanje R-a

1. Unutar foldera OpenStatistics napraviti svoj folder sa oznakom XXNNYY, gde je XX oznaka smera (SW/SL), NN broj indeksa (7→07) i YY godina upisa (2014→14)

2. Kopirati ikonu R-a u svoj direktorijum

3. Podesiti Properties kopirane ikone: u Start In uneti punu putanju do svog foldera (na primer c:\Users\statistika\Desktop\OpenStatistics\SW0014)

4. Pokrenuti R.

5. Isprobati u komandnom promptu

   ```
   x<-sin(1)
   q()
   ```

   Save Workspace Image: Yes
   (biće sačuvan u fajlu .RData, istorija komandi u .Rhistory)

6. Ponovo pokrenuti R

```
objects()
x
```

treba da je $x = 0.841471$.

7. Izračunati u R-u zadatak 4. a)

```
1-pbinom(21,40,.5)
```

8. Izračunati u R-u zadatak 4. c)

```
1-ppois(21,40*.5)
```

9. Približno rešiti zadatak 3. simulacijom u R-u.

```
n<-10000
s<-numeric(n)
for(k in 1:n){s[k]<-sum(runif(30))<17}
p<-mean(s)
```

# Vežbe 7

## Deskriptivna statistika

1. Neka je dat uzorak

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 87 | 103 | 130 | 160 | 180 | 195 | 132 | 145 | 211 | 105 |
| 145 | 153 | 152 | 138 | 87 | 99 | 93 | 119 | 129 | 145 |

   a) Odrediti modus i medijanu.

```
x<-c(87,103,130,160,180,195,132,145,211,105,145,153,152,138,87,99,93,119,129,145)
Me<-median(x)
tx<-table(x)
Mo<-as.numeric(names(tx)[which.max(tx)])
```

   b) Izračunati aritmetičku sredinu i standardnu devijaciju uzorka.

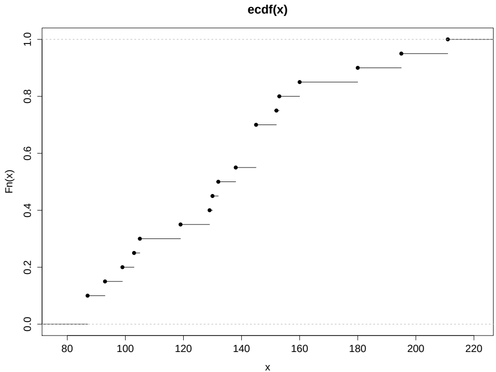
```
xn<-mean(x)
n<-length(x)
sn<-sqrt(sd(x)^2*(n-1)/n)
```

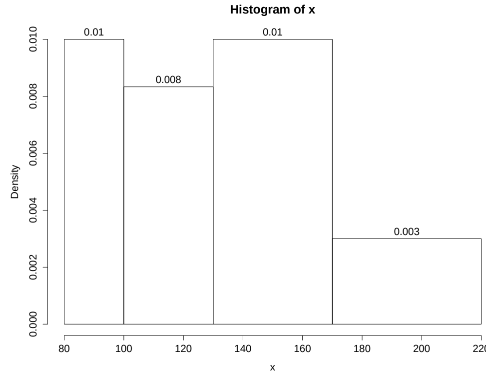
c) Odrediti koeficijent spljoštenosti i koeficijent asimetrije.

```
s4<-numeric(n)
for(k in 1:n) s4[k]<-(x[k]-xn)^4
mi4<-mean(s4)
s3<-numeric(n)
for(k in 1:n) s3[k]<-(x[k]-xn)^3
mi3<-mean(s3)
s2<-numeric(n)
for(k in 1:n) s2[k]<-(x[k]-xn)^2
mi2<-mean(s2)

Ks<-mi4/mi2^2
Ka<-mi3/mi2^(3/2)

#momenti
library(e1071)
m4<-moment(x, order=4, center=TRUE)
m3<-moment(x, order=3, center=TRUE)
m2<-moment(x, order=2, center=TRUE)
```

d) Nacrtati uzoračku funkciju raspodele, histogram i poligon.

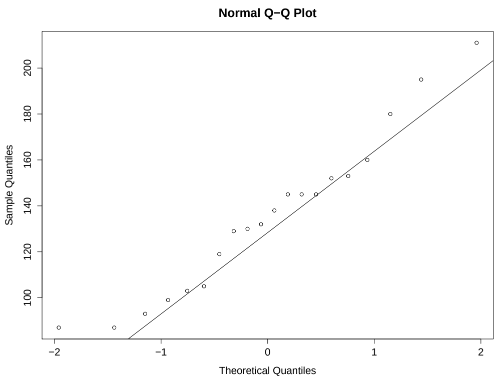
```
plot.ecdf(x)
hist(x)
mi<-c(80,100,130,170,220)
hist(x,breaks=mi,probability=TRUE,labels=TRUE)
#poligon rucno docrtamo
```

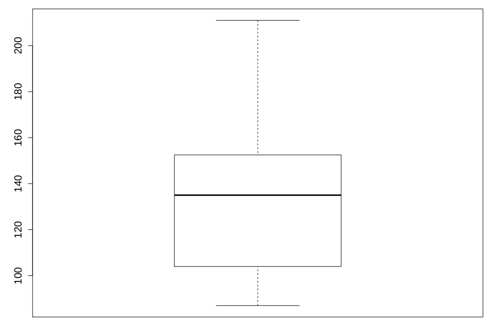


e) Naći kvartile i nacrtati *Q-Q plot* i *Box plot*.

```
summary(x)
IQR(x)
qx<-qnorm((1:n)/n)
plot(qx,sort(x))
qqnorm(x)
qqline(x)
boxplot(x)
```



2. Neka je dat intervalni uzorak

| $I_i$ | [0,1] | (1,2] | (2,3] | (3,5] | (5,10] | (10,20] |
|---|---|---|---|---|---|---|
| $f_i$ | 15 | 11 | 7 | 7 | 6 | 4 |

a) Nacrtati uzoračku funkciju raspodele, histogram i poligon.

| $I_i$ | [0,1] | (1,2] | (2,3] | (3,5] | (5,10] | (10,20] |
|---|---|---|---|---|---|---|
| $f_i$ | 15 | 11 | 7 | 7 | 6 | 4 |
| $x_i$ | 0.5 | 1.5 | 2.5 | 4 | 7.5 | 15 |
| $n_{x_i}$ | 15 | 26 | 33 | 40 | 46 | 50 |
| $f_n^*$ | 0.3 | 0.52 | 0.66 | 0.8 | 0.92 | 1 |

```
mi<-c(0,1,2,3,5,10,20)
fi <-c(15,11,7,7,6,4)
m0<-mi[1:length(mi)-1]
m1<-mi[2:length(mi)]
xi<-(m0+m1)/2
x<-rep(xi,fi)

f<-cumsum(fi)/sum(fi)
plot.ecdf(x)
hist(x,breaks=mi,probability=TRUE,labels=TRUE)
```

b) Odrediti modus i medijanu.

c) Izračunati aritmetičku sredinu i standardnu devijaciju uzorka.

d) Odrediti koeficijent spljoštenosti i koeficijent asimetrije.

e) Naći kvartile i nacrtati *Q-Q plot* i *Box plot*.

```
Me<-median(x)
tx<-table(x)
Mo<-as.numeric(names(tx)[which.max(tx)])
xn<-mean(x)
n<-length(x)
sn<-sqrt(sd(x)^2*(n-1)/n)
mi4<-moment(x, order=4, center=TRUE)
mi3<-moment(x, order=3, center=TRUE)
mi2<-moment(x, order=2, center=TRUE)
Ks<-mi4/mi2^2
Ka<-mi3/mi2^(3/2)
summary(x)
IQR(x)
qqnorm(x)
qqline(x)
boxplot(x)
```

# Domaći rad

Zadaci 141-150 iz zbirke

# Manipulisanje podataka

Pokretanje skript fajla iz R-a:

```
setwd("c:\Users\statistika\Desktop\OpenStatistics\SW0014")
source("zadatak1.R")
```

Čuvanje promenljivih

```
load(".RData")                    #ucitavanje promenljivih iz R formata
save.image("novo.RData")          #cuvanje Work space u fajl novo.RData
save(xn,sn, file ="xnsn.rda")     #cuvanje promenljivih xn i sn u fajl xnsn.rda
```

Uvoženje podataka u R

```
FlightDelays<-read.csv("FlightDelays.csv")      #ucitavanje csv tabele
summary(FlightDelays)                           #osnovna deskriptivna statistika
FlightDelays<-edit(FlightDelays)                #otvaranje data editora
str(FlightDelays)                               #vraca strukturu podataka
names(FlightDelays)                             #vraca nazive promenljivih iz tabele
head(FlightDelays)                              #prvih 6 redova
head(FlightDelays,n=10)                         #prvih 10 redova
head(FlightDelays,n=-10)                        #svi redovi osim poslednjih 10
tail(FlightDelays)                              #poslednjih 6 redova
FlightDelays[1:10,1:3]                          #prvih 10 redova, prve 3 promenljive
```

3. Izdvojiti iz kolone Delay vrednosti koje se odnose na ponedeljak i nacrtati *Box plot* za Delay po danima.

```
delay<-FlightDelays$Delay
indeks<-FlightDelays$Day==c("Mon")
delaymon<-delay[indeks]
boxplot(Delay~Day,data=FlightDelays)
```

# Vežbe 8

## Tačkaste ocene parametara

### Metod momenta

Izjednačavamo uzoračke momente sa momentima obeležja.

### Metod maksimalne verodostojnsti

Funkcija verodostojnosti

$$L = L(x_1, x_2, \dots, x_n, \theta) = \begin{cases} \varphi(x_1, \theta)\varphi(x_2, \theta)\dots\varphi(x_n, \theta), & \text{neprekidno obeležje,} \\ p(x_1, \theta)p(x_2, \theta)\dots p(x_n, \theta), & \text{diskretno obeležje.} \end{cases}$$

Tražimo $\theta$ za koje $L$ dostiže maksimum.

1. Meren je prečnik sfere (u *cm*) pet puta i dobijen je uzorak:

$$6.33, 6.37, 6.36, 6.32, 6.37.$$

Odrediti centrirane ocene matematičkog očekivanja i disperzije obeležja.

```
x<-c(6.33,6.37,6.36,6.32,6.37)
m<-mean(x)
s<-sd(x)^2
```

2. Anketiranjem 100 vozača automobila iz Novog Sada dobijene su prosečne dnevne potrošnje benzina (u litrima).

| potrošnja | 2 | 4 | 5 | 6 | 8 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|
| broj vozača | 5 | 10 | 10 | 12 | 18 | 12 | 8 | 10 | 9 | 6 |

Odrediti centrirane ocene matematičkog očekivanja i disperzije potrošnje benzina jednog vozača.

```
xi<-c(2,4,5,6,8,10,11,12,13,14)
fi <-c(5,10,10,12,18,12,8,10,9,6)
x<-rep(xi,fi)
m<-mean(x)
s<-sd(x)^2
```

3. Obeležje $X$ date populacije ima raspodelu $X : \begin{pmatrix} -2 & 0 & 7 \\ \frac{\theta}{5} & \frac{\theta}{5} & 1 - \frac{2\theta}{5} \end{pmatrix}$.

   a) Metodom momenta naći ocenu nepoznatog parametra $\theta$ na osnovu uzorka

   $$(0,-2,7,-2).$$

   b) Na osnovu istog uzorka naći ocenu nepoznatog parametra $\theta$ metodom maksimalne verodostojnosti.

4. Reakcija oka na jednu vrstu nadražaja ima eksponencijalnu raspodelu $\mathcal{E}(a)$. Eksperiment je izvršen 10 puta i dobijeni su rezultati (izraženi u nano sekundama):

   $$1.41, 1.28, 2.49, 0.95, 0.26, 3.83, 1.56, 3.87, 0.83, 3.37.$$

   Metodom momenta i metodom maksimalne verodostojnosti naći ocenu nepoznatog parametra $a$.

deskstat.R

```
deskstat<-function(x,mi=NULL){
  #medijana
  Me<-median(x)
  print(paste0("Medijana:", Me))
  #modus
  tx<-table(x)
  Mo<-as.numeric(names(tx)[which.max(tx)])
  print(paste0("Modus:", Mo))
  #aritmeticka sredina
  xn<-mean(x)
  print(paste0("Aritmeticka sredina:",xn))
  #standardna devijacija
  sn<-sqrt(sd(x)^2*(length(x)-1)/length(x))
  print(paste0("Standardna devijacija:",sn))
  # koeficijent spljostenosti i koeficijent asimetrije
  library(e1071)
  mi4<-moment(x, order=4, center=TRUE)
  mi3<-moment(x, order=3, center=TRUE)
  mi2<-moment(x, order=2, center=TRUE)
  Ks<-mi4/mi2^2
  Ka<-mi3/mi2^(3/2)
  print(paste0("Koeficijent spljostenosti:",Ks))
  print(paste0("Koeficijent asimetrije:",Ka))
```

```
  #uzoracka funkcija raspodele
  plot.ecdf(x)
  #histogram
  k=1
  if (is.null(mi)){
    h<-hist(x)
  }
  else{
    h<-hist(x,breaks=mi,probability=TRUE,labels=TRUE)
    k=3
  }
  library(agricolae)
  polygon.freq(h,frequency=k)
  print(summary(x))
  #Q-Q plot
  qqnorm(x)
  qqline(x)
  #Boxplot
  boxplot(x)
}
```

5. Učitati iz fajla TXBirths2004.csv kolonu Weight samo za dečake i napraviti deskriptivnu statistiku

```
podaci<-read.csv("TXBirths2004.csv")
indeks<-podaci$Gender==c("Male")
weightMale<-podaci$Weight[indeks]

source("deskstat.R")
deskstat(weightMale)
```

6. Napraviti deskriptivnu statistiku za kolonu AveKW iz fajla Turbine.csv za mesec jun.

```
podaci<-read.csv("Turbine.csv")
indeks<-substr(podaci$Date2010,1,3)==c("Jun")
podaciJun<-podaci$AveKW[indeks]

source("deskstat.R")
deskstat(podaciJun)
```

7. Napraviti tablice Gausove i Studentove raspodele.

```
#tablica Normalne raspodele
p<-pnorm(seq(from=0, by=0.01, length=350))
dim(p)<-c(10,35)
p<-t(p)
View(p)

#tablica Studentove raspodele
x<-c(1:30,40,60,120,Inf)
p<-c(.75,.9,.975,.99,.995,.9995)
qt<-t(outer(p,x,"qt"))
View(qt)
```

# Vežbe 9

## Intervali poverenja i testiranje hipoteza

Posmatra se uzorak od 100 slučajnih brojeva iz intervala $(0,1)$.

1. Naći 90% interval poverenja za procenat brojeva koji su veći od 0.75.

```
#interval poverenja za nepoznatu verovatnocu
set.seed(12345)
x<-runif(100)
K<-length(which(x>.75))
n<-length(x)
z<-qnorm((.9+1)/2)
a<-n^2+z^2*n
b<--2*K*n-z^2*n
c<-K^2
d<-b^2-4*a*c
x1<-(-b-sqrt(d))/2/a
x2<-(-b+sqrt(d))/2/a
```

Interval poverenja $I = (0.1948498, 0.3377947)$

2. Testirati hipotezu da je procenat brojeva koji su veći od 0.75 jednak 0.3 sa pragom značajnosti 5%.

   a) Koristeći interval poverenja rađen na predavanjima.

```
# testiranje hipoteze o nepoznatoj verovatnoci
set.seed(12345)
x<-runif(100)
K<-length(which(x>.75))
n<-length(x)
z<-qnorm((.95+1)/2)
a<-n^2+z^2*n
b<--2*K*n-z^2*n
c<-K^2
d<-b^2-4*a*c
x1<-(-b-sqrt(d))/2/a
x2<-(-b+sqrt(d))/2/a
```

$H_0(p = 0.3)$ protiv $H_1(p \neq 0.3)$

$p_0 = 0.3 \in I = (0.184047, 0.3537099) \Rightarrow$ hipoteza $H_0$ se prihvata

b) Koristeći interval poverenja iz zbirke.

```
p<-K/n
q<-1-p
y1<-p-z*sqrt(p*q/(n-1))
y2<-p+z*sqrt(p*q/(n-1))
```

c) Koristeći `binom.test` iz R-a.

```
>binom.test(K,n,.3,conf.level=.95)

  Exact binomial test

data:  K and n
number of successes = 26, number of trials = 100, p-value = 0.4451
alternative hypothesis: true probability of success is not equal to 0.3
95 percent confidence interval:
 0.1773944 0.3573121
sample estimates:
probability of success
                  0.26
```

Učitati fajl prijemni.csv.

3. Naći 90% interval poverenja za uspeh iz srednje škole.

```
#interval poverenja za ocekivanje m, sigma nepoznato
podaci<-read.csv("prijemni.csv")
uspehSkola<-podaci$skola
n<-length(uspehSkola)
xn<-mean(uspehSkola)
s<-sd(uspehSkola)
t<-qt((1+.9)/2,n-1)
x1<-xn-t*s/sqrt(n)
x2<-xn+t*s/sqrt(n)
```

Interval poverenja $I = (28.1033, 30.99209)$

4. Testirati hipotezu da je srednja vrednost uspeha u srednjoj školi jednaka 32 ($\alpha = 0.05$).

```
# testiranje  hipoteze  o  ocekivanju  m, sigma nepoznato
podaci<-read.csv("prijemni.csv")
uspehSkola<-podaci$skola
n<-length(uspehSkola)
xn<-mean(uspehSkola)
s<-sd(uspehSkola)
t<-qt((1+.95)/2,n-1)
x1<-xn-t*s/sqrt(n)
x2<-xn+t*s/sqrt(n)
```

$H_0(m = 32)$ protiv $H_1(m \neq 32)$

$m_0 = 32 \notin I = (27.81335, 31.28203) \Rightarrow$ hipoteza $H_0$ se odbacuje

5. Naći 95% interval poverenja za varijansu uspeha iz srednje škole.

```
#interval poverenja za varijansu (disperziju)
podaci<-read.csv("prijemni.csv")
uspehSkola<-podaci$skola
n<-length(uspehSkola)
s<-sd(uspehSkola)^2*(n-1)/n
y1<-qchisq((1+.95)/2,n-1)
y2<-qchisq((1-.95)/2,n-1)
x1<-n*s/y1
x2<-n*s/y2
```

Interval poverenja $I = (19.11832, 47.54454)$

6. Testirati hipotezu o jednakosti srednje vrednosti uspeha na prijemnom kod muških i ženskih kandidata ($\alpha = 0.05$).

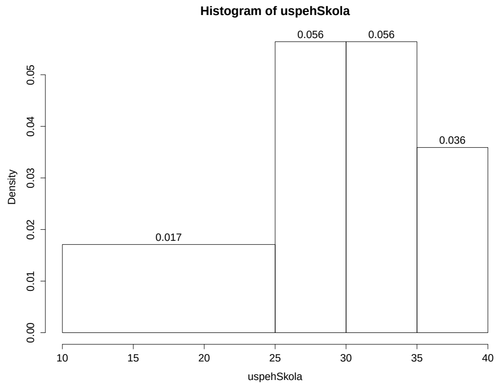
```
# testiranje hipoteze o jednakosti srednjih vrednosti dva obelezja
podaci<-read.csv("prijemni.csv")
prijemniM<-podaci$prijemni[podaci$pol=="m"]
prijemniZ<-podaci$prijemni[podaci$pol=="z"]
n1<-length(prijemniM)
n2<-length(prijemniZ)
xn1<-mean(prijemniM)
xn2<-mean(prijemniZ)
s1<-sd(prijemniM)^2*(n1-1)/n1
s2<-sd(prijemniZ)^2*(n2-1)/n2
t<-qt((1+.95)/2,n1+n2-2)
t0<-(xn1-xn2)/sqrt((n1*s1+n2*s2)/(n1+n2-2))/sqrt(1/n1+1/n2)
```

Nulta hipoteza $H_0(m_1 = m_2)$

$t_0 = -1.739699 \in I = (-2.026192, 2.026192) \Rightarrow$ hipoteza $H_0$ se prihvata

7. Nacrtati histogram za uspeh iz srednje škole ako su deobne tačke: 10, 25, 30, 35, 40.

```
podaci<-read.csv("prijemni.csv")
uspehSkola<-podaci$skola
hist(uspehSkola,c(10,25,30,35,40),probability=TRUE,labels=TRUE)
```



8. Testirati hipotezu da je uspeh u srednjoj školi raspoređen po normalnoj raspodeli ako su grupe za $\chi^2$-test: 10-25, 25-30, 30-35, 35-40 ($\alpha = 0.05$).

```
podaci<-read.csv("prijemni.csv")
uspehSkola<-podaci$skola
mi<-c(10,25,30,35,40)
table(cut(uspehSkola,mi,include.lowest=TRUE,right=TRUE))
fi <-c(10,11,11,7)
n<-length(uspehSkola)
k<-length(fi) #broj intervala
s<-2 #broj parametara koje treba oceniti
xn<-mean(uspehSkola)
sn<-sd(uspehSkola)

#funkcija raspodele
Fx<-function(x){
  return(pnorm((x-xn)/sn)          #pnorm(x,xn,sn)
}
# teorijske verovatnoce
p<-numeric(length(fi))
p[1]<-Fx(mi[2])-Fx(-Inf)
p[2]<-Fx(mi[3])-Fx(mi[2])
p[3]<-Fx(mi[4])-Fx(mi[3])
p[4]<-Fx(Inf)-Fx(mi[4])
```

```
#realizovana vrednost statistike
y0<-sum((fi-n*p)^2/n/p)
#najveca dozvoljena vrednost statistike
y<-qchisq(1-.05,k-1-s)
```

$y_0 = 1.296183 < y = 3.841459 \Rightarrow$ hipoteza $H_0$ se prihvata

# Vežbe 10

1. Testom Kolmogorov-Smirnov testirati hipotezu o saglasnosti datog uzorka sa normalnom raspodelom $\mathcal{N}(10,4)$ ($\alpha = 0.05$).

| $0-5$ | $5-15$ | $15-20$ | $20-30$ | $30-40$ |
|---|---|---|---|---|
| 3 | 7 | 10 | 15 | 5 |

```
fi <-c(3,7,10,15,5)
mi<-c(0,5,15,20,30,40)
m0<-mi[1:length(mi)-1]
m1<-mi[2:length(mi)]
xi<-(m0+m1)/2
f<-cumsum(fi)
n<-sum(fi)
Fn<-f/n   #empirijska funkcija  raspodele
Fx<-pnorm(xi,10,4)   #data funkcija  raspodele
#realizovana vrednost  statistike
sqrt(n)*max(abs(Fn-Fx))
```

$2.970034 > 1.36 = \lambda_{0.95} \Rightarrow$ hipoteza $H_0$ se odbacuje

2. Ispitati nezavisnost obeležja $X$ i $Y$ čije su realizovane vrednosti uzorka:

| | | |
|---|---|---|
| 5 | 17 | 9 |
| 5 | 7 | 19 |

```
fij <-matrix(c(5,5,17,7,9,19),2,3)
#matrica kontingencije
#f<-matrix(c(5,5,10,17,7,24,9,19,28,31,31,62),3,4)
f<-matrix(nrow=nrow(fij)+1,ncol=ncol(fij)+1)
for (i in 1:nrow(f)){
  for (j in 1:ncol(f)){
    if (i<nrow(f) & j<ncol(f)){f[i,j]=fij[i,j]}
    else if ((i==1 | i==2) & j==ncol(f)){f[i,j]=sum(fij[i,])}
    else if (i==nrow(f) & (j==1 | j==2 | j==3)){f[i,j]=sum(fij[,j])}
    else if (i==nrow(f) & j==ncol(f)){f[i,j]=sum(f[1:i-1,j])}
  }
}
n<-sum(fij) #obim uzorka
#matrica sa teorijskim frekvencijama
ft<-matrix(NA,nrow(fij),ncol(fij))
for (i in 1:nrow(ft)){
  for (j in 1:ncol(ft)){
    ft[i,j]=f[i,ncol(f)]*f[nrow(f),j]/n
  }
}
```

```
#realizovana vrednost statistike
y0<-sum((ft-fij)^2/ft)
#najveca dozvoljena vrednost statistike
y<-qchisq(1-.05,(nrow(fij)-1)*(ncol(fij)-1))
```

$H_0$ : obeležja $X$ i $Y$ su nezavisna

$y_0 = 7.738095 > y = 5.991465 \Rightarrow$ hipoteza $H_0$ se odbacuje

3. $\chi^2$-testom ispitati saglasnost obeležja čiji je realizovani uzorak (0.12, 0.14, 0.25, 0.05, 0.02, 0.08, 0.03, 0.04, 0.51, 0.07, 0.42, 0.08, 0.33, 0.36, 0.06, 0.23) sa raspodelom:

$$\varphi(x) = \begin{cases} \theta x^{\theta-1}, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}, \theta > 0.$$

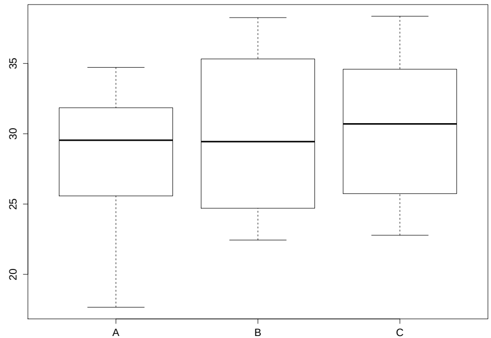
```
#uzorak
x<-c(0.12,0.14,0.25,0.05,0.02,0.08,0.03,0.04,0.51,0.07,0.42,0.08,0.33,0.36,0.06,0.23)
xs<-sort(x) #sortiran uzorak
fi <-c(6,5,5) #frekvencije po intervalima (sami konstruisemo intervale)
mi<-c(0,0.07,0.23,1)
n<-sum(fi) #obim uzorka
xn<-mean(x)
theta<-xn/(1-xn) #ocena parametra theta
k<-length(fi) #broj intervala
s<-1 #broj parametara
#funkcija raspodele
Fx <- function(arg){
  if (arg<0){fx=0}
  else if (0<=arg & arg<1){fx=arg^theta}
  else {fx=1}
  return(fx)
}
```

```
# teorijske verovatnoce
p<-numeric(k)
for (i in 1:k){
  p[i]<-Fx(mi[i+1])-Fx(mi[i])
}
#realizovana vrednost statistike
y0<-sum((fi-n*p)^2/n/p)
#najveca dozvoljena vrednost statistike
y<-qchisq(1-.05,k-1-s)
```

$y_0 = 3.39369 < y = 3.841459 \Rightarrow$ hipoteza $H_0$ se prihvata

4. Nacrtati *Box plot* uspeha iz srednje škole po grupama.

```
podaci<-read.csv("prijemni.csv")
boxplot(skola~grupa,data=podaci)
```



5. Testirati hipotezu o jednakosti srednje vrednosti uspeha iz srednje škole po grupama (ANOVA).

```
>podaci<-read.csv("prijemni.csv")
>anova(lm(skola~grupa,data=podaci))
Analysis of Variance Table

Response: skola
          Df  Sum Sq  Mean Sq  F value  Pr(>F)
grupa      2    26.6   13.300   0.4512  0.6404
Residuals 36  1061.2   29.476

#isto dobijamo i sa komandom
summary(aov(lm(skola~grupa,data=podaci)))
```

$H_0(m_1 = m_2 = m_3)$ protiv $H_1(\exists i,j, m_i \neq m_j)$

Tabela ANOVE: $\begin{array}{ccccc} Df_1 & SSTR & MSTR & F & \alpha^* \\ Df_2 & SSE & MSE & & \end{array}$

$Df_1, Df_2$ – broj stepeni slobode

$F$ – realizovana vrednost Fišerove test statistike sa $Df_1, Df_2$ stepeni slobode

$\alpha^*$ – $p$-vrednost

$\alpha^* = 0.6404 > \alpha = 0.05 \Rightarrow H_0$ se prihvata

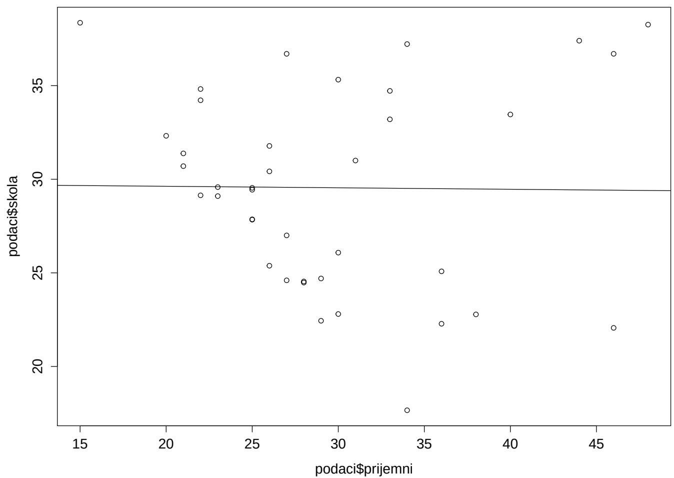
6. Naći koeficijent korelacije uspeha iz srednje škole u zavisnosti od uspeha na prijemnom. Prognozirati kom uspehu u srednjoj školi odgovara 35 bodova osvojenih na prijemnom.

```
>podaci<-read.csv("prijemni.csv")
>lm(skola~prijemni,data=podaci)

Call:
lm(formula = skola ~ prijemni, data = podaci)

Coefficients:
(Intercept)     prijemni
  29.784053    -0.008009
```

Jednačina linearne regresije

$$skola = -0.008009 * prijemni + 29.784053$$

```
plot(podaci$prijemni,podaci$skola)
abline(lm(skola~prijemni,data=podaci))
```



```
# koeficijent korelacije
cor(podaci$skola,podaci$prijemni)

#predikcija
lm(skola~prijemni,data=podaci)$coefficients%*%c(1,35)
```

Koeficijent korelacije $r = -0.01157393$
$|r| < 0.3 \Rightarrow$ uspeh iz srednje škole i uspeh na prijemnom nisu u korelacionoj vezi

Broju od 35 poena osvojenih na prijemnom odgovara 29.50375 poena iz srednje škole.

# Vežbe 11

Po jedan uzorak sa dve mašine za pakovanje deterdženta od 10 $kg$ je izmeren na preciznoj vagi.

- 9.81, 9.83, 10.43, 11.13, 9.70, 9.59, 10.88, 10.97, 9.35, 9.34, 9.41, 9.95, 11.03, 10.12, 9.33, 9.73, 10.17, 9.48, 10.89, 10.11, 10.30, 8.87, 9.51, 10.42, 10.02, 10.84, 9.96, 10.15, 10.64, 11.30
- 9.85, 9.30, 9.08, 8.07, 9.22, 9.55, 7.88, 7.84, 8.50, 11.95, 10.92, 9.78, 10.61, 9.49, 7.81, 8.90, 8.60, 8.50, 9.31, 9.97, 8.89, 8.87, 7.23, 7.82, 7.65, 9.11, 8.65, 6.30, 9.38, 8.31, 10.48, 10.56, 9.96, 8.84, 9.10, 11.07, 9.84, 9.75, 9.07, 9.09, 8.96, 8.11, 8.17, 9.73, 9.06, 8.40, 11.12, 9.38, 7.26, 8.69

```
uzorak1<-c(9.81, 9.83, 10.43, 11.13, 9.70, 9.59, 10.88, 10.97, 9.35, 9.34, 9.41, 9.95, 11.03, 10.12, 9.33,
9.73, 10.17, 9.48, 10.89, 10.11, 10.30, 8.87, 9.51, 10.42, 10.02, 10.84, 9.96, 10.15, 10.64, 11.30)

uzorak2<-c(9.85, 9.3, 9.08, 8.07, 9.22, 9.55, 7.88, 7.84, 8.5, 11.95, 10.92, 9.78, 10.61, 9.49, 7.81, 8.9,
8.6, 8.5, 9.31, 9.97, 8.89, 8.87, 7.23, 7.82, 7.65, 9.11, 8.65, 6.3, 9.38, 8.31, 10.48, 10.56, 9.96, 8.84,
9.10, 11.07, 9.84, 9.75, 9.07, 9.09, 8.96, 8.11, 8.17, 9.73, 9.06, 8.40, 11.12, 9.38, 7.26, 8.69)
```

a) Kolmogorov-Smirnov testom testirati hipotezu o saglasnosti prvog uzorka sa normalnom raspodelom $\mathcal{N}(10,0.81)$.

```
> ks.test(uzorak1,'pnorm',10,0.81)

   One-sample Kolmogorov-Smirnov test

data:  uzorak1
D = 0.17074, p-value = 0.3097
alternative hypothesis: two-sided
```

$\alpha^* = 0.3097 > \alpha = 0.05 \Rightarrow$ hipoteza $H_0$ se prihvata

Napomena:

Vrednost $D = 0.17074$ je realizovana vrednost statistike $D_n = \sup_x |F_n^*(x) - F(x)|$.

Kako je $\sqrt{n_1}D = \sqrt{30} \cdot 0.17074 = 0.9351815 < 1.36$ zaključujemo da se hipoteza $H_0$ prihvata.

b) $\chi^2$-testom testirati hipotezu o saglasnosti drugog uzorka sa normalnom raspodelom (deobne tačke: 8.0, 8.5, 9.0, 9.5, 10.0).

```
xn<-mean(uzorak2)
sn<-sd(uzorak2)
m<-c(8,8.5,9,9.5,10)
pm<-pnorm(m,xn,sn)
p<-c(pm,1)-c(0,pm)
mi<-c(floor(min(uzorak2)),m,ceiling(max(uzorak2)))
fi <-hist(uzorak2,mi)$counts

y0<-chisq.test(fi,p=p)$statistic
y<-qchisq(.95,length(fi)-1-2) #ocenili smo dva parametra

alpha<-1-pchisq(y0,length(fi)-1-2) #p-vrednost
```

$y_0 = 2.079539 < y = 7.814728 \Rightarrow$ hipoteza $H_0$ se prihvata

Napomena:

Pošto smo morali da ocenimo 2 parametra, $p$−vrednost koju vraća `binom.test` nije tačna (ova vrednost bi bila tačna da nije bilo nepoznatih parametara) i moramo peške izračunati $\alpha^*$.
$\alpha^* = 0.5560632 > \alpha = 0.05 \Rightarrow$ hipoteza $H_0$ se prihvata

c) Testirati hipotezu da je srednja vrednost prvog uzorka veća od srednje vrednosti drugog.

```
> t.test(uzorak1,uzorak2,alternative='greater')

   Welch Two Sample t-test

data:  uzorak1 and uzorak2
t = 5.3151, df = 77.935, p-value = 4.929e-07
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.7067711       Inf
sample estimates:
mean of x mean of y
 10.10867  9.07960
```

$H_0(m_1 = m_2)$ protiv $H_1(m_1 > m_2)$

$\alpha^* = 4.929e - 07 < \alpha = 0.05 \Rightarrow$ hipoteza $H_0$ se odbacuje, tj. prihvata se $H_1$

d) Testirati hipotezu da je srednja vrednost prvog uzorka veća od srednje vrednosti drugog, pod pretpostavkom da imaju jednake varijanse.

```
> t.test(uzorak1,uzorak2,alternative='greater',var.equal=TRUE)

  Two Sample t-test

data:  uzorak1 and uzorak2
t = 4.6821, df = 78, p-value = 5.892e-06
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.6631998       Inf
sample estimates:
mean of x mean of y
 10.10867  9.07960
```

$H_0(m_1 = m_2)$ protiv $H_1(m_1 > m_2)$

$\alpha^* = 5.892e - 06 < \alpha = 0.05 \Rightarrow$ hipoteza $H_0$ se odbacuje, tj. prihvata se $H_1$

e) Testirati hipotezu da je standardna devijacija prvog uzorka jednaka 1.

```
n<-length(uzorak1)
s<-sd(uzorak1)^2*(n-1)/n
y1<-qchisq((1+.95)/2,n-1)
y2<-qchisq((1-.95)/2,n-1)
x1<-n*s/y1
x2<-n*s/y2
```

Nulta hipoteza $H_0(\sigma = 1)$

$\sigma_0 = 1 \notin I = (0.2552004, 0.7271325) \Rightarrow$ hipoteza $H_0$ se odbacuje

f) Testirati hipotezu da su standardne devijacije kod oba uzorka jednake.

```
> var.test(uzorak1,uzorak2)

  F test to compare two variances

data:  uzorak1 and uzorak2
F = 0.33427, num df = 29, denom df = 49, p-value = 0.002289
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1776679 0.6653103
sample estimates:
ratio of variances
         0.3342673
```

Nulta hipoteza $H_0(\sigma_1^2 = \sigma_2^2)$

$\alpha^* = 0.002289 < \alpha = 0.05 \Rightarrow$ hipoteza $H_0$ se odbacuje

# Vežbe 9, zadatak 2. a)

```
>prop.test(K,n,.3,conf.level=.95,correct=FALSE)

	1-sample proportions test without continuity correction

data:  K out of n, null probability 0.3
X-squared = 0.7619, df = 1, p-value = 0.3827
alternative hypothesis: true p is not equal to 0.3
95 percent confidence interval:
 0.1840470 0.3537099
sample estimates:
   p
0.26
```

$H_0(p = 0.3)$ protiv $H_1(p \neq 0.3)$

$\alpha^* = 0.3827 > \alpha = 0.05 \Rightarrow$ hipoteza $H_0$ se prihvata

Vežbe 9, zadatak 4.

```
>t.test(uspehSkola,mu=32,conf.level = .95)

   One Sample t-test

data:   uspehSkola
t  = -2.8624, df = 38, p-value = 0.006803
alternative  hypothesis:  true  mean is not equal to 32
95 percent confidence  interval :
 27.81335 31.28203
sample estimates:
mean of x
 29.54769
```

$H_0(m = 32)$ protiv $H_1(m \neq 32)$

$\alpha^* = 0.006803 < \alpha = 0.05 \Rightarrow$ hipoteza $H_0$ se odbacuje

## Vežbe 10, zadatak 2.

```
>chisq.test(matrix(c(5,5,17,7,9,19),2,3))

   Pearson's Chi-squared test

data: fij
X-squared=7.7381, df=2, p-value=0.02088
```

$H_0$ : obeležja $X$ i $Y$ su nezavisna

$\alpha^* = 0.02088 < \alpha = 0.05 \Rightarrow$ hipoteza $H_0$ se odbacuje