

# Verovatnoća i statistika

Softversko inženjerstvo i informacione tehnologije

Inženjerstvo informacionih sistema

školska 2022/23

# Pokretanje R-a

1. Napraviti svoj folder sa oznakom XXNNYY, gde je XX oznaka smeru (SV/IT), NN broj indeksa (7→07) i YY godina upisa (2021→21)
2. Kopirati ikonu R-a u svoj folder
3. Podesiti Properties kopirane ikone: u Start In uneti punu putanju do svog foldera (na primer c:\Users\Desktop\IT0021)
4. Pokrenuti R
5. Isprobati u komandnom promptu

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

Save Workspace Image: Yes

(biće sačuvano u fajlu .RData, istorija komandi u .Rhistory)

6. Ponovo pokrenuti R

```
objects()
x           #treba da je x = 0.841471
```

# Normalna raspodela

1. Slučajna promenljiva  $Z$  ima normalnu  $\mathcal{N}(0,1)$  raspodelu. Izračunati verovatnoću

a)  $P(Z < 1.645) = \phi(1.645)$

`pnorm(1.645)`

b)  $P(Z > -1.645) = 1 - \phi(-1.645)$

`1-pnorm(-1.645)`

c)  $P(|Z| < 1.645) = 2\phi(1.645) - 1$

`2*pnorm(1.645)-1`

d)  $P(|Z| > 1.96) = 2 - 2\phi(1.96)$

`2-2*pnorm(1.96)`

2. Slučajna promenljiva  $X$  ima normalnu  $\mathcal{N}(1.5, 2.5)$  raspodelu. Izračunati verovatnoću

a)  $P(X < 1.645) = \Phi\left(\frac{1.645 - 1.5}{2.5}\right)$

`pnorm(1.645, 1.5, 2.5)`      `#pnorm((1.645-1.5)/2.5)`

b)  $P(X > -1.645) = 1 - \Phi\left(\frac{-1.645 - 1.5}{2.5}\right)$

`1-pnorm(-1.645, 1.5, 2.5)`

c)  $P(|X| < 1.645) = \Phi\left(\frac{1.645 - 1.5}{2.5}\right) - \Phi\left(\frac{-1.645 - 1.5}{2.5}\right)$

`pnorm(1.645, 1.5, 2.5) - pnorm(-1.645, 1.5, 2.5)`

d)  $P(|X| > 1.96) = 1 - \Phi\left(\frac{1.96 - 1.5}{2.5}\right) + \Phi\left(\frac{-1.96 - 1.5}{2.5}\right)$

`1-pnorm(1.96, 1.5, 2.5) + pnorm(-1.96, 1.5, 2.5)`

3. Slučajna promenljiva  $Z$  ima normalnu  $\mathcal{N}(0,1)$  raspodelu. Naći  $z$  za koje je

a)  $P(Z < z) = 0.9$

**qnorm(0.9)**

b)  $P(Z > z) = 0.95$

$\Rightarrow P(Z \leq z) = 0.05$

**qnorm(0.05)**

c)  $P(|Z| < z) = 0.95$

$P(|Z| < z) = 2\phi(z) - 1 \Rightarrow \phi(z) = 0.975$

**qnorm(0.975)**

4. Slučajna promenljiva  $X$  ima normalnu  $\mathcal{N}(1.5, 2.5)$  raspodelu. Naći  $x$  za koje je

a)  $P(X < x) = 0.9$

**qnorm(0.9)\*2.5+1.5**

b)  $P(X > x) = 0.95$

**qnorm(0.05)\*2.5+1.5**

5. Slučajna promenljiva  $Z$  ima normalnu  $\mathcal{N}(0,1)$  raspodelu. Neka je  $X = Z^2$ .

a) Izračunati verovatnoću  $P(X < 2.25)$

$$\Rightarrow X : \chi_1^2$$

`pchisq(2.25,1)`

b) Odrediti  $x$  za koje je  $P(X > x) = 0.95$ .

$$P(X \leq x) = 0.05$$

`qchisq(0.05,1)`

**Podsetnik:** Neka su  $Z_1, Z_2, \dots, Z_n$  nezavisne slučajne promenljive za koje važi  $Z_i : \mathcal{N}(0,1)$ , za svako  $i = 1, 2, \dots, n$ . Tada slučajna promenljiva

$$X = \sum_{i=1}^n Z_i^2$$

ima Pirsonovu  $\chi_n^2$  raspodelu.

# Granične teoreme

5. Novčić se baca 40 puta.

a) Odrediti tačnu verovatnoću da će biti barem 22 grba.

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

c) Koristeći aproksimaciju Poasonovom raspodelom oceniti verovatnoću da će biti barem 22 grba.

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

2. Nezavisne slučajne promenljive  $X_1, X_2, \dots, X_{30}$  imaju istu uniformnu raspodelu  $\mathcal{U}(0,1)$ .

Simulacijom u R-u približno odrediti verovatnoću  $P\left(\sum_{i=1}^{30} X_i < 17\right)$ .

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

$p \approx 0.89$

# 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)                                     #tabela frekvencija
Mo<-as.numeric(names(tx)[which.max(tx)])         #ispisuje samo prvi modus
Mo<-as.numeric(names(tx)[tx==max(tx)])           #ispisuje sve moduse
```

b) Izračunati aritmetičku sredinu i standardnu devijaciju uzorka.

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



c) Odrediti koeficijent spljoštenosti i koeficijent asimetrije.

```
mi4<-mean((x-xn)^4)
```

```
mi3<-mean((x-xn)^3)
```

```
mi2<-mean((x-xn)^2)
```

```
Ks<-mi4/mi2^2
```

```
Ka<-mi3/mi2^(3/2)
```

```
#pomocu funkcija iz paketa e1071
```

```
library(e1071)
```

```
moment(x, order=5)
```

```
#obicaj moment reda 5
```

```
moment(x, order=5, center=TRUE)
```

```
#centralni moment reda 5
```

```
kurtosis(x,type=1)+3
```

```
#koeficijent spljostenosti
```

```
skewness(x,type=1)
```

```
#koeficijent asimetrije
```

Koeficijent spljoštenosti (*kurtosis*):  $K_s = \frac{\mu_4}{\mu_2^2} = 2.594 \rightarrow$  raspodela je spljoštena u odnosu na normalnu

Koeficijent asimetrije (*skewness*):  $K_a = \frac{\mu_3}{\mu_2^{3/2}} = 0.442 \rightarrow$  raspodela je asimetrična u desno

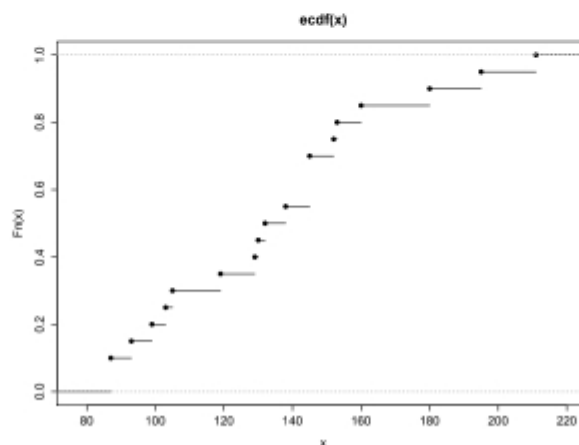
d) Nacrtati uzoračku funkciju raspodele.

Uzoračka (empirijska) funkcija raspodele  $F_n^* = \frac{N_x}{n}$

- $N_x$  broj elemenata uzorka koji su  $\leq x$
- $n$  obim realizovanog uzorka

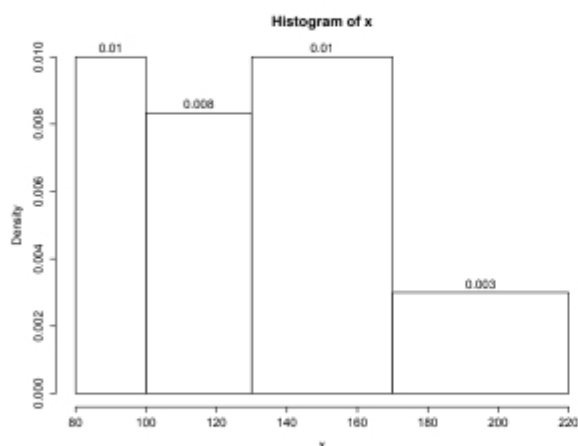
```
plot.ecdf(x)
table(x)/n
cumsum(table(x))/n
```

*#empirical cumulative distribution function  
#verovatnoce da se x nalazi u intervalu  
#vrednosti empirijske funkcije raspodele*



e) Nacrtati histogram i poligon.

```
hist(x) #visine pravougaonika odgovaraju frekvencijama
mi<-c(80,100,130,170,220)
hist(x,breaks=mi,probability=TRUE,labels=TRUE) #histogram gustine
#poligon rucno doctamo
```



Kod histograma gustine ukupna površina svih pravougaonika iznosi 1.

Napomena: Naredbom **hist** se dobija histogram kod kog visine pravougaonika odgovaraju frekvencijama  $f_i$ , umesto vrednostima  $\frac{f_i}{h_i}$ , gde je  $h_i$  širina intervala. U okviru paketa **ggplot2** postoje funkcije pomoću kojih se može nacrtati histogram koji odgovara histogramu koji je definisan na predavanjima.

## Za radoznale

```
h1 <- hist(x,breaks=c(70,mi,230),probability=FALSE)
```

```
par(new=TRUE)
```

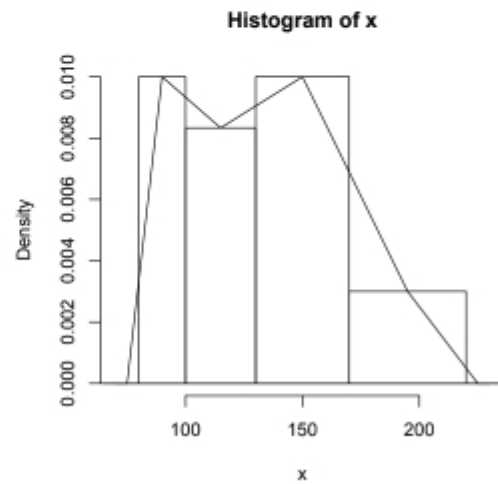
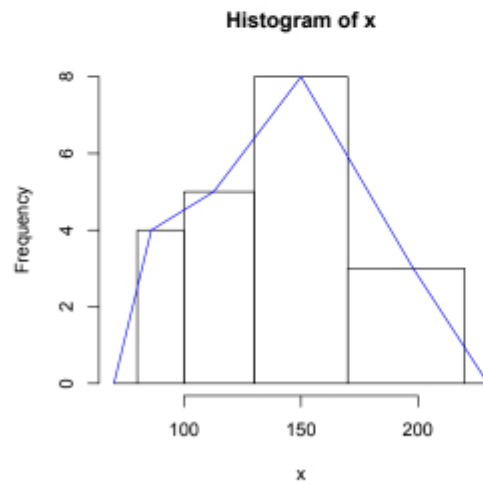
```
plot(h1$mids, h1$counts, type="l", col="blue", axes=FALSE, xlab="", ylab="")
```

```
library(agricolae)
```

```
h2 <- hist(x,breaks=c(70,mi,230),probability=TRUE)
```

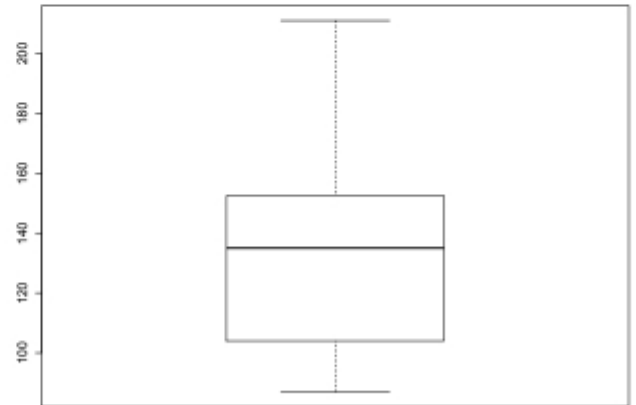
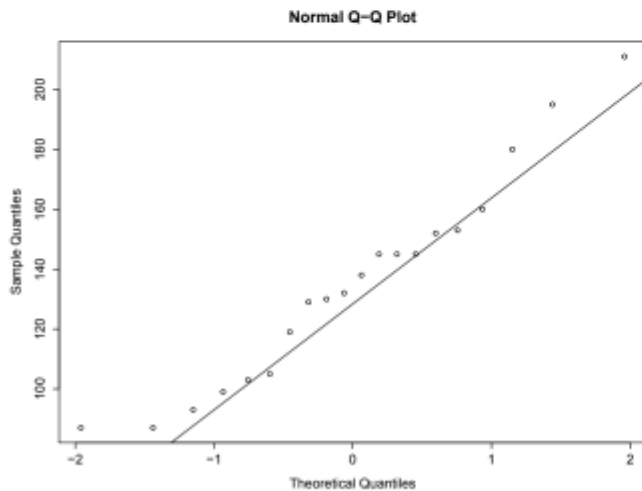
```
polygon.freq(h2,frequency = 3)
```

*#frequency=1 za histogram frekvencija*



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

<b>summary(x)</b>	#vraca vrednosti za min, Q1, Me=Q2, Xn, Q3, max
<b>IQR(x)</b>	#IQR=Q3-Q1
#kvantil proizvoljnog reda p za uzorak x mozemo dobiti naredbom quantile(x,p)	
<b>qx&lt;-qnorm((1:n)/n)</b>	#peske crtanje QQ plota
<b>plot(qx,sort(x))</b>	
<b>qqnorm(x)</b>	#pomocu ugradjenih funkcija
<b>qqline(x)</b>	
<b>boxplot(x)</b>	



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

```
xi <- c(0.5,1.5,2.5,4,7.5,15)      #sredine intervala
fi <- c(15,11,7,7,6,4)             #frekvencije
x <- rep(xi,fi)

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

## Ideja za uzorke većeg obima

```
mi<-c(0,1,2,3,5,10,20)      #deobne tacke  
  
ml<-mi[1:length(mi)-1]      #leve granice intervala  
md<-mi[2:length(mi)]        #desne granice intervala  
  
xi<-(ml+md)/2                #sredine intervala  
fi<-c(15,11,7,7,6,4)  
x<-rep(xi,fi)
```

b) Odrediti modus i medijanu.

Kod intervalnog uzorka postoje posebne formule za računanje modusa i medijane:

$$M_o = m_{s-1} + h_s \frac{r_1}{r_1 + r_2} = 0 + 1 \cdot \frac{15}{15 + 4} = 0.7895$$

- $I_s = (m_{s-1}, m_s)$  modalni interval
- $h_s$  širina modalnog intervala
- $r_1 = f_s - f_{s-1}$
- $r_2 = f_s - f_{s+1}$

$$M_e = m_{l-1} + h_l \frac{\frac{n}{2} - n_{x_{l-1}}}{f_l} = 1 + 1 \cdot \frac{25 - 15}{11} = 0.90909$$

- $I_l = (m_{l-1}, m_l)$  medijalni interval
- $h_l$  širina medijalnog intervala
- $n_{x_{l-1}}$  kumulativna frekvencija intervala koji prethodi medijalnom
- $f_l$  frekvencija medijalnog intervala



c) Izračunati aritmetičku sredinu i standardnu devijaciju uzorka.

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

d) Odrediti koeficijent spljoštenosti i koeficijent asimetrije.

```
mi4<-mean((x-xn)^4)
mi3<-mean((x-xn)^3)
mi2<-mean((x-xn)^2)
Ks<-mi4/mi2^2
Ka<-mi3/mi2^(3/2)
```

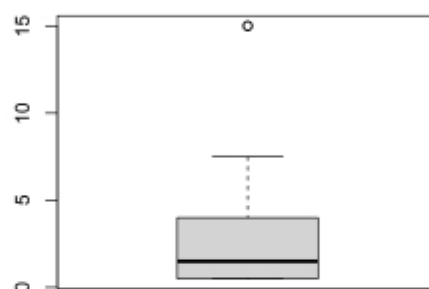
e) Naći kvartile i nacrtati *Q-Q plot*.

```
Q1<-quantile(x,.25)
Q2<-median(x)
Q3<-quantile(x,.75)
IQR(x)
qqnorm(x)
qqline(x)
```

f) Nacrtati *Box plot* i odrediti *outlier*-e.

```
boxplot(x)
```

```
x[x < Q1 - 1.5*IQR(x) | x > Q3 + 1.5*IQR(x)] # outlieri
```



## Domaći rad

Na slučajan način odabrati neke od prvih zadataka iz profesorove zbirke ("Zbirka rešenih zadataka iz verovatnoće i statistike") i kompletno ih rešiti peške\* i u R-u. (Poželjno je uraditi jedan zadatak sa običnim uzorkom, jedan sa intervalnim uzorkom i zadatak 144.)

\*Nacrtati histogram, boxplot i empirijsku funkciju raspodele bez upotrebe računara

## Domaći rad

U posebnom script fajlu napisati funkciju koja za prosleđen uzorak ispisuje deskriptivnu statistiku.

Na primer:

deskstat.R

```
deskstat <- function(x){  
  n <- length(x)  
  print(paste0("Obim_uzorka:", n))  
  
  ...          #deskriptivna statistika kao u prvom zadatku  
  
  boxplot(x)  
}
```

Pronaći kako da se funkciji doda opcioni vektor sa zadatim deobnim tačkama za crtanje histograma i implementirati to u napisanu funkciju.

## Manipulacija podacima

Pokretanje skript fajla iz R-a:

```
setwd("c:/Users/Desktop/IT0021")  
source("zadatak1.R")
```

Proveru putanje do radnog direktorijuma možemo izvršiti naredbom **getwd()**.

Čuvanje promenljivih

<b>load</b> ("RData")	<i>#ucitavanje promenljivih iz R formata</i>
<b>save.image</b> ("novo.RData")	<i>#cuvanje Work space u fajl novo.RData</i>
<b>rm</b> (list=ls())	<i>#oslobadjanje svih promenljivih</i>
<b>load</b> ("novo.RData")	
<b>save</b> (xn,sn, <b>file</b> ="xnsn.rda")	<i>#cuvanje promenljivih xn i sn u fajl xnsn.rda</i>

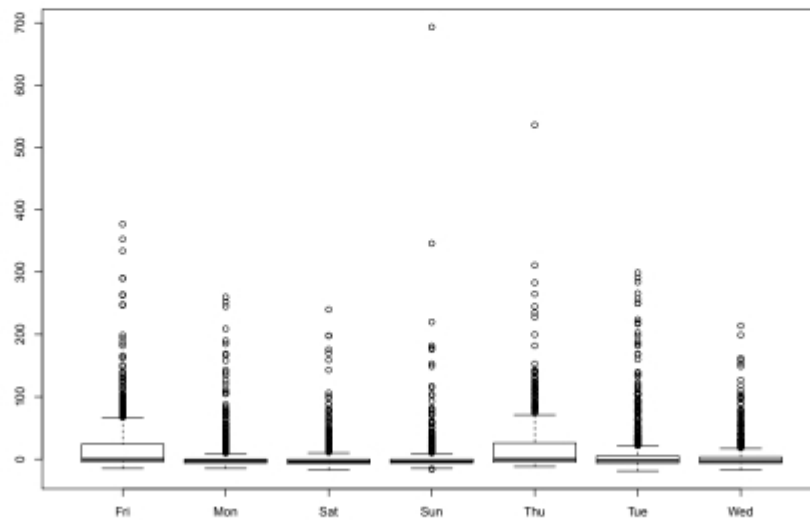
## Učitavanje podataka u R-u

<code>FlightDelays &lt;- read.csv("FlightDelays.csv")</code>	<i>#ucitavanje csv tabele</i>
<code>summary(FlightDelays)</code>	<i>#osnovna deskriptivna statistika</i>
<code>FlightDelays &lt;- edit(FlightDelays)</code>	<i>#otvaranje data editora</i>
<code>str(FlightDelays)</code>	<i>#vraca strukturu podataka</i>
<code>names(FlightDelays)</code>	<i>#vraca nazive promenljivih iz tabele</i>
<code>head(FlightDelays)</code>	<i>#prvih 6 redova</i>
<code>head(FlightDelays,n=10)</code>	<i>#prvih 10 redova</i>
<code>head(FlightDelays,n=-10)</code>	<i>#svi redovi osim poslednjih 10</i>
<code>tail(FlightDelays)</code>	<i>#poslednjih 6 redova</i>
<code>FlightDelays[1:10,1:3]</code>	<i>#prvih 10 redova, prve 3 kolone</i>
<code>FlightDelays[1:10,c(1,3,7)]</code>	<i>#prvih 10 redova, kolone 1, 3 i 7</i>

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

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

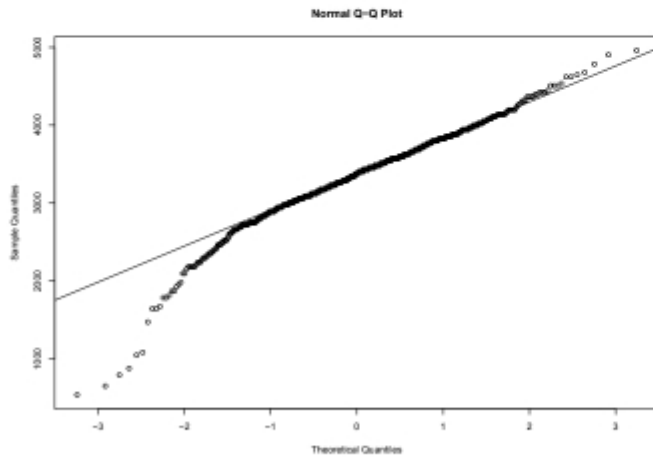
#sa \$ se pristupa podelementu  
#u vektor indeks upisuje TRUE ako je ponedjeljak  
#izdvaja vrednosti za koje je indeks TRUE



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

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

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



Na osnovu dobijenog *Q-Q plot*-a zaključujemo da postoji mogućnost da posmatrani uzorak ima normalnu raspodelu. Međutim, ovo ne možemo tvrditi sa sigurnošću. Zato je neophodno sprovesti odgovarajuće statističke testove koji će dati odgovor na ovo pitanje.

5. Napraviti deskriptivnu statistiku za kolonu AveKW iz fajla Turbine.csv za mesece april i jun.

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

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

## Domaći rad

Pronaći kako se u R-u zapisuju logički operatori  $\neq$ ,  $\wedge$ ,  $\neg$ .



## Grupisanje podataka

```
x<-c(1,2,4,5,1,9,9,1)
```

```
y<-c(rep("a",4),rep("b",3),"c")
```

*#losa ideja :*

```
Dc<-cbind(x,y)
```

*#"spajanje" vektora u matricu po kolonama/vrstama*

```
Dr<-rbind(x,y)
```

*#korekcija tipova podataka*

*#bolja ideja :*

```
D<-data.frame(x,y)
```

*#data frame moze da sadrzi kolone razlicitih tipova, za razliku od matrica*

```
D<-data.frame(Brojevi=x,Slova=y)
```

```
D$Brojevi[7]=2
```

*#izmena vrednosti*

```
D$Brojevi==2
```

*#upit o vrednosti*

```
D[D$Brojevi==2, ]
```

*#selekcija vrsta po kriterijumu*

*#alternativno:*

```
subset(D,D$Brojevi==2)
```

```
D$Brojevi[D$Slova=="a"] = D$Brojevi[D$Slova=="a"] * 2
```

## Domaći rad

1. Pronaći kako sve možemo upotrebiti funkciju **rep**.
2. Pronaći šta radi funkcija **seq**.
3. Šta radi naredba **par(new=TRUE)** sa slajda 7? Pronaći kako da se pomoću funkcije **par** na ekranu prikaže više grafika istovremeno.
4. Šta vraćaju naredbe **h\$mids** i **h\$counts** za prosleđen histogram **h**?
5. Proveriti šta radi blok komandi (i uporediti rezultat sa modusom/modusima):  

```
y<-tabulate(match(x,unique(x)));  
unique(x)[y==max(y)]
```
6. Pronaći razliku između vrednosti **NA**, **NaN**, **Null** i **Inf**. Šta vraća funkcija **is.na**?

# Deskriptivna statistika - nastavak

1. Posmatraju se dva uzorka obima 100. Prvi uzorak čine brojevi sa  $\mathcal{N}(0,1)$  raspodelom (zaokruženi na 2 decimale), a u drugom uzorku su brojevi sa  $\mathcal{P}(0.5)$  raspodelom. (Postaviti da "seme" za oba uzorka bude 1111.)

```
set.seed(1111)
```

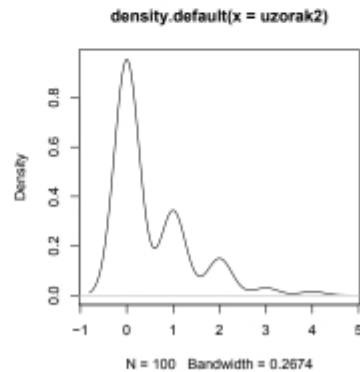
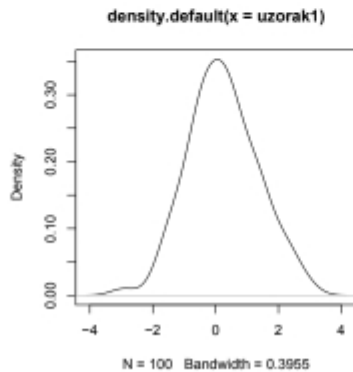
```
uzorak1 <- round(rnorm(100), digits=2)
```

```
uzorak2 <- rpois(100, lambda=.5)
```

```
plot(density(uzorak1))
```

*#grafik funkcije gustine*

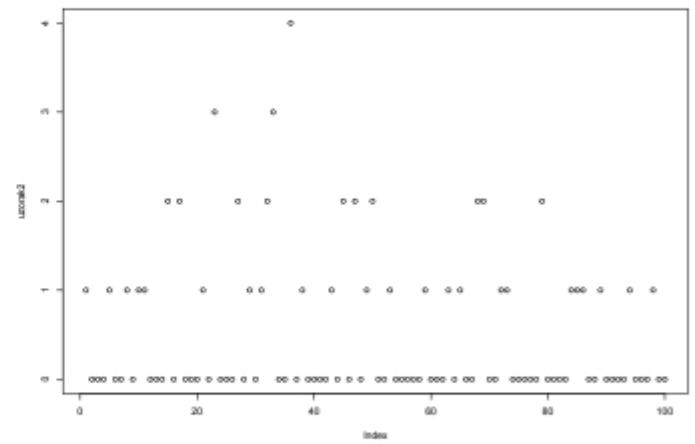
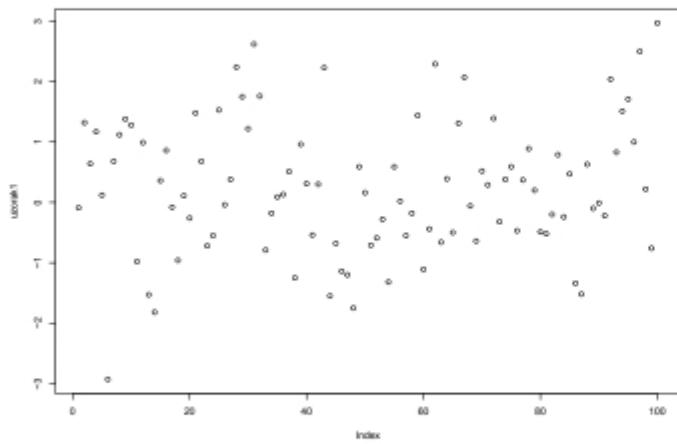
```
plot(density(uzorak2))
```



<b>cor</b> (uzorak1,uzorak2)	<i>#korelacija</i>
<b>cov</b> (uzorak1,uzorak2)	<i>#kovarijansa</i>
<b>mean</b> (uzorak1,trim=0.2)	<i>#odseca po 20% vrednosti sa obe strane pre izracunavanja</i>
<b>table</b> (uzorak2)	
<b>prop.table</b> (table(uzorak2))	<i>#relativne frekvencije (2. nacin)</i>
<b>quantile</b> (uzorak1, probs= <b>seq</b> (0,1,0.05))	<i>#kvantili za uzorak1</i>
<b>scale</b> (uzorak1)	<i>#z-score za uzorak1</i>
(uzorak1 - <b>mean</b> (uzorak1))/ <b>sd</b> (uzorak1)	<i>#peske</i>
kvantili_p <- <b>qchisq</b> ( <b>ppoints</b> (100), <b>df</b> =3)	<i>#kvantili Pirsonove raspodele</i>
<b>qqplot</b> (kvantili_p, <b>sort</b> (uzorak1))	
<b>abline</b> (a=0,b=1)	<i>#prava y=x</i>
kvantili_n <- <b>qnorm</b> ( <b>ppoints</b> (100),5,5)	<i>#kvantili N(5,5) raspodele</i>
<b>qqplot</b> (kvantili_n, <b>sort</b> (uzorak1))	
<b>abline</b> (a=0,b=1)	

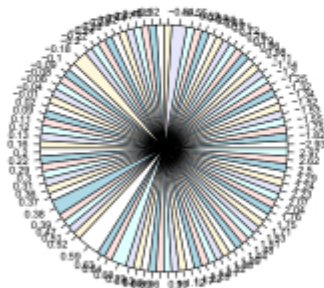
## Dijagram rasipanja (*scatterplot*)

```
plot(uzorak1)
plot(uzorak1,pch=3)          # scatterplot sa +
plot(uzorak2)
```



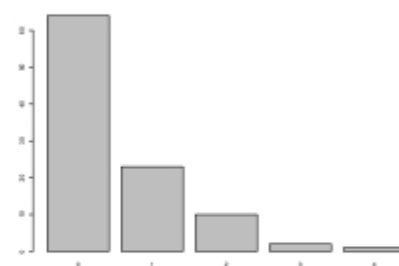
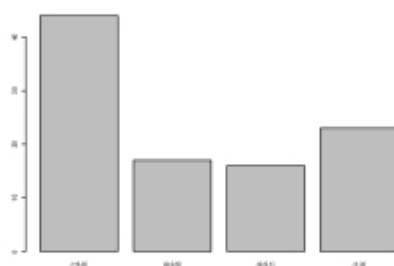
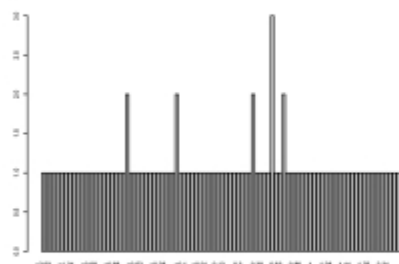
## Pita (pie chart)

```
pie(table(uzorak1))  
x <- cut(uzorak1, breaks = c(-5, 0, .5, 1, 3))  
table(x)  
pie(table(x))  
pie(table(uzorak2))
```



## Stubičasti dijagram (*bar chart*)

```
barplot(table(uzorak1))  
barplot(table(x))  
barplot(table(uzorak2))  
#barplot(table(uzorak2),col="cyan")
```



## Domaći rad

Koja je razlika između histograma i barplota?

## Faktori

```
gender<-factor(c("male", "female", "female", "male"))
typeof(gender)
attributes(gender)
as.character(gender)

f.pomocni<-c(rep("a",4),rep("b",6),rep("c",2))
f<-rep(f.pomocni,length=100)           #f ima 100 elemenata
f<-as.factor(f)                       #nivoi faktora: a, b, c

y<-round(rnorm(100),digits=2)
z<-rpois(100,lambda=.5)

D<-data.frame(f,y,z)

D$y[D$f=="a"]
D[D$f=='a', ]                         #navodnici ili apostrofi
D[D$f=="a" & D$z==0, ]

write.csv(D, file="frame.csv") #upisivanje u csv fajl
#write.csv(D, file="frame.csv",row.names=FALSE)
```



## Unos podataka i NA

<code>x &lt;- scan()</code>	<i>#pojedinačni ručni unos</i>
<code>x &lt;- c(1,2,3,5,7,9)</code>	
<code>data.entry(x)</code>	<i>#pojedinačna izmena</i>
<code>x &lt;- c(x,NA)</code>	
<code>mean(x)</code>	<i>#javlja gresku</i>
<code>mean(x,na.rm=T)</code>	<i>#pravilno</i>
<code>mean(x[!is.na(x)])</code>	<i>#pravilno</i>
<code>mean(na.omit(data.frame(x)\$x))</code>	<i>#izbegavati posto brise cele vrste</i>

2. Napraviti tablicu Gausove raspodele (zbirka, str. 169).

```
p<-pnorm(seq(from=0, by=0.01, length=350))
dim(p)<-c(10,35)
p<-t(p) #transponovanje
View(round(p,digits=4))

#2. nacin
P<-pnorm(seq(from=0, by=0.01, length=350))
P<-matrix(P,35,10,byrow=TRUE)
View(round(P,digits=4))
```

3. Napraviti tablicu Studentove raspodele (zbirka, str. 170).

```
df<-c(1:30,40,60,120,Inf)
p<-c(.75,.9,.95,.975,.99,.995,.9995)
q<-t(outer(p,df,"qt"))
View(round(q,digits=3))
```

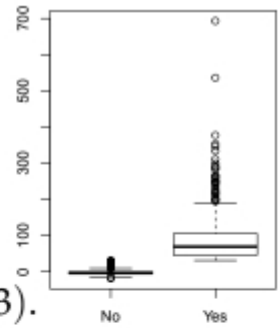
### **Domaći rad**

Napraviti tablicu Pirsonove  $\chi^2$  raspodele (zbirka, str. 171).

## Domaći rad

1. Na slučajan način je izabrano 150 prirodnih brojeva manjih od 1000.
  - a) Formirati raspodelu frekvencija sa 10 klasa iste širine i rezultate predstaviti tabelarno.
  - b) Odrediti modalni i medijalni interval iz a).
  - c) Odrediti aritmetičku sredinu i uzoračku disperziju uzorka.
  - d) Odrediti koeficijent asimetrije (skewness) i sedmi centralni momenat uzorka.
  - e) Izračunati realizovanu vrednost empirijske funkcije raspodele  $f_n^*(615.5)$ .
  - f) Nacrtati histogram i *pie chart* koristeći intervale iz a).
2. Učitati fajl FlightDelays.csv.
  - a) Nacrtati *Box plot* za dužinu leta (FlightLength) po mesecu.
  - b) Odrediti koliko su u proseku kasnili letovi (Delay) za Denver (Destination: DEN).
  - c) Da li je standardna devijacija dužine leta veća kod aviona koji su poleteli između 4 i 8 časova (DepartTime: 4-8am) ili kod onih koji su poleteli između 16 i 20 časova (DepartTime: 4-8pm)?

3. Na osnovu *Box plot*-a uzorka zaključiti da li postoje statistički značajne razlike između srednjih vrednosti elemenata uzorka u ove dve grupe. Odgovor obrazložiti.



4. Naći treći decil za raspodele:  $\mathcal{B}(10, \frac{3}{4})$ ,  $\mathcal{G}(\frac{1}{10})$ ,  $\mathcal{E}(0.1)$ ,  $t_{10}$ ,  $\chi^2_{10}$ ,  $\mathcal{N}(15, 3)$ .
5. Izračunati vrednost funkcije raspodele u tački  $x = 7.5$  ako  $X$  ima  $\mathcal{B}(10, \frac{3}{4})$ ,  $\mathcal{G}(\frac{1}{10})$ ,  $\mathcal{E}(0.1)$ ,  $t_{10}$ ,  $\chi^2_{10}$ ,  $\mathcal{N}(15, 3)$  raspodelu.
6. Izračunati  $\Gamma(-\frac{3}{2})$ .
7. Pronaći šta vraćaju naredbe **dnorm** i **dbinom**.
8. Instalirati paket **ggplot2** i nacrtati nekoliko deskriptivnih grafika koristeći funkcije iz ovog paketa.

