Uvod u $\mathcal{R}$[1]

Nikola Ajzenhamer

24. decembar 2015.

Sadržaj

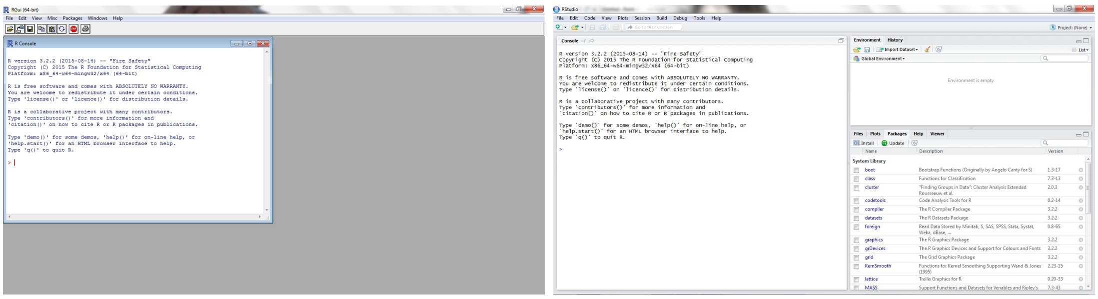
[1]Materijal je sačinjen od beleški sa časova vežbi kursa *Verovatnoća i statistika* na trećoj godini smera Informatika Matematičkog fakulteta i na osnovu `https://cran.r-project.org/doc/manuals/R-intro.pdf`.

1 Dobavljanje potrebnih resursa

Instalacija $\mathcal{R}$-a se vrši u dva koraka:

1. Potrebno je dobaviti odgovarajuci $\mathcal{R}$ instalacioni paket sa `https://cran.r-project.org/index.html` i pratiti detalje instalacije za odgovarajući sistem. Postoje verzije za Linux, Mac OS X i Windows, koje su skoro potpuno identične. Pretpostavimo da smo instalirali $\mathcal{R}$ na Windows sistemu. Nakon instalacije, u odabranom instalacionom folderu mogu se pronaći verzije 32-bitne i 64-bitne `Rgui.exe` datoteke koja se sasvim dovoljna za dalji rad.

2. (Opcioni korak) Ukoliko želimo da budemo više produktivni u $\mathcal{R}$-u, dostupan je i besplatan set integrisanih alatki pod nazivom *RStudio* koji se može dobaviti sa adrese `https://www.rstudio.com/products/rstudio/download/`. *RStudio* je dostupan na svim trima platformama, takođe.



Slika 1. Poređenje GUI-ja između dva navedena okruženja za rad u $\mathcal{R}$-u

2 Pomoć oko funkcija, komande, case sensitivity, i sl.

Ukoliko Vam je u bilo kom trenutku potrebna pomoć oko neke funkcije, recimo, ako ste zaboravili redosled argumenata funkcije `choose`, pomoć u vidu dokumentacije možete potražiti kucanjem `help(choose)` u konzolu. U zavisnosti od okruženja u kojem radite, prikazaće Vam se prozor sa detaljnim prikazom vezano za funkciju koju ste naveli kao argument funkcije `help()`.

Case sensitivity je prisutna u $\mathcal{R}$-u, te su imena `primer` i `Primer` dva različita objekta. Svi alfanumerički simboli su omogućeni za korišćenje pri davanju naziva[2]. Osim ovih karaktera dozvoljeno je i korišćenje tačke (`'.'`) i podvlake (`'_'`). Međutim, i ovde postoji izuzetak da ime može počinjati `'.'` ili slovom, a ako počinje `'.'`, onda drugi karakter ne sme biti broj.

Komande se odvajaju dvotačkom (`';'`) ili novim redom. Elementarne komande se grupišu pomoću otvorenih i zatvorenih vitičastih zagrada (`'{}'`) i čine blok izraz. Komentari se mogu stavljati gotovo svuda[3] i počinju znakom `'#'`.

[2] Osim ukoliko se ne radi o prenosivom $\mathcal{R}$ kodu (što uključuje i $\mathcal{R}$ pakete) gde je neophodno koristiti isključivo karaktere iz opsega [A-Za-z0-9].

[3] Osim u stringovima, kao i u listi argumenata u definiciji funkcije

3 Manipulacije objektima

3.1 Vektori i dodele

$\mathcal{R}$ operiše nad pojedinim strukturama podataka. Najosnovnija struktura podataka jeste numerički vektor, tj. niz brojeva. Primer ovakvog vektora, nazovimo ga `niz`, može se dobiti nekom od sledećih komandi:

> `niz <- c(1,2,3,4,5)` ili `niz = c(1,2,3,4,5)`,

> `niz = 1:6` ili `niz <- 1:6`,

> `niz = seq(from=1, to=5, by=1)` ili `niz <- seq(from=1, to=5, by=1)`

i predstavlja celobrojni niz brojeva od 1 do 5. Primetimo da je (skoro) svejedno da li koristimo `=` ili `<-` kao operator dodele.

Naravno da vektor ne mora biti celobrojnog tipa. Da smo imali, recimo, element 4.25 umesto 4 u našem nizu, onda bi se ceo niz pretvorio u vektor realnih brojeva, implicitno. Što se tiče indeksiranja, prvi član niza je numerisan jedinicom. Pristup nekom elementu vektora se vrši putem uglastih zagrada. U prethodnom nizu, element na poziciji 3, tj. element `3`, možemo dobiti pozivom `niz[3]`. Ukoliko želimo da pristupimo više članova odjednom, recimo da želimo da izdvojimo podniz od 2 do 4, to možemo uraditi nekom od sledeće dve komande:

> `niz[c(2,3,4)]` ili

> `niz[(niz >= 2) & (niz < 5)]`,

pri čemu niz koji dobijamo prilikom korišćenja prve komande nazivamo indeksni niz, a `niz >= 2` predstavlja vektor čiji su elementi logičkog tipa, tj. na i-tom mestu se nalazi `TRUE`[4] ako je uslov zadovoljen, a `FALSE`[5] inače.

Kad su u pitanju operacije sa vektorima, moguće je izvršavati standardne operacije sa vektorima, kao što su `+`, `-`, `*`, `/` i `^` za stepenovanje. Takođe je omogućeno korišćenje funkcija poput `log`, `exp`, `sin`, `cos`, `tan`, `sqrt`, ali i druge sa standardnim značenjem. Funkcije `max` i `min` pronalaze najveći, odnosno, najmanji element u vektoru, `range` za vrednost ima vektor od dva elementa koji zapravo predstavlja vektor dobijen komandom `c(min(niz), max(niz))`, `length` vraća broj elemenata u vektoru, `sum` vraća zbir, a `prod` proizvod elemenata vektora. Funkcija `mean` vraća aritmetičku sredinu, a `median` medijanu vektora.

Napomenuli smo da vektori ne moraju da imaju celobrojne vrednosti. Štaviše, oni mogu da sadrže i, recimo, stringove ili logičke tipove elemenata. Primer vektora stringova bi bio:

> `triReci = c("prva rec", " druga rec", "treca rec")`.

3.2 Višedimenzioni nizovi

Višedimenzione nizove možemo veoma jednostavno napraviti korišćenjem funkcije `array`. Njen prototip izgleda:

- `array(data = NA, dim = length(data), dimnames = NULL)`.

Objasnimo ovo na sledećem primeru: Neka je dat vektor `x = 1:5`, tj. vektor celih brojeva od 1 do 5. Želimo da napravimo matricu dimenzija 4×3 koju ćemo ispuniti vrednostima vektora `x`. Naravno,

[4] Može se pisati i `T`, ali ne i `True` ili `true`
[5] Može se pisati i `F`, ali ne i `False` ili `false`

možemo već da pretpostavimo da će se elementi vektora `x` ponavljati u našoj matrici. Ta matrica će, nazovimo je `M`, nakon komande `M = array(x, dim = c(4,3))` izgledati ovako:

$$M = \begin{pmatrix} 1 & 5 & 4 \\ 2 & 1 & 5 \\ 3 & 2 & 1 \\ 4 & 3 & 2 \end{pmatrix}.$$

Primetimo dve stvari. Prvo, višedimenzioni nizovi se popunjavaju po kolonama, a ne po vrstama. Drugo, nismo naveli treći argument iz prototipa funkcije, i dobili smo rezultat. Taj treći argument nije obavezan jer predstavlja *listu* imena dimenzija višedimenzionog niza. Često nam to i neće biti neophodno.

3.3 Obeležja i faktori

Objektima možemo pridružiti obeležja koja mogu biti:

- **numerička**, na primer, težina, visina itd.
- **kategorička**, na primer, pol, verska pripadnost, boja očiju itd.

S obzirom na činjenicu da su kategorička obeležja najčešće kodirana pri unosu podataka, u tu svrhu se mogu koristiti objekti poznati kao faktori. **Faktor** (*Factor*) je vektorski objekat koji se koristi da specifikuje klasifikaciju elemenata drugih vektora iste dužine. Recimo da nam je zadat niz od 15 ispitanika hipotetičke ankete. Svaki element ovog niza je ili 1 ili 2 ako se radi o osobi muškog, odnosno, ženskog pola:

```
> pol <- c(1,2,1,1,1,1,2,2,1,2,1,2,2,2,1).
```

Ukoliko bismo napravili faktor `fPol` niza `pol` na sledeći način:

```
> fPol <- factor(pol),
```

onda bismo dobili faktor koji, osim što sadrži informacije o elementima niza `pol`, sadrži i **nivoe** (*levels*), odnosno, spisak svih različitih obeležja. Pomoću komande

```
> levels(fPol) <- c("muski", ženski")
```

možemo zadati imena nivoima, kao i automatski konvertovati elemente faktora u imena nivoa. Rezultat primene prethodne komande daje sledeći izlaz:

```
[1] muski zenski muski muski muski muski zenski zenski muski zenski muski
[12] zenski zenski zenski muski
Levels: muski zenski
```

Postoji još jedna korisna funkcija `summary()`, koja će, ukoliko za argument uzme faktor, ispisati broj elemenata na svakom nivou. Dakle, primenom komande

```
> summary(fPol)
```

dobija se izlaz

```
muski zenski
    8      7
```

što se intenzivno koristi u praksi. Funkcija `summary()` može za argumente uzimati i neke druge objekte i prikazivati različite izlaze. Autor ohrabruje čitaoca da isproba komandu `summary(pol)`, tj. primenu funkcije na vektor i uveri se u njenu korisnost.

3.4 Liste

Lista (*list*) u programskom jeziku $\mathcal{R}$ označava objekat koji se sastoji od više numerisanih kolekcija koje se nazivaju **komponente**. Komponente u listi ne moraju (i često nema ni potrebe) da budu istog tipa. Sledeći jednostavan primer ilustruje konstrukciju liste:

> `lPorodica <- list(muz="Milovan", zena="Ana", brDece=3, godineDece=c(4,7,9)).`

Komponentama ove liste možemo pristupiti pomoću `lPorodica[[1]]`, `lPorodica[[2]]`, `lPorodica[[3]]` and `lPorodica[[4]]`. Dalje, pošto je objekat `lPorodica[[4]]` tipa vektor, onda prvom članu tog vektora možemo pristupiti pomoću `lPorodica[[4]][1]`.

Korišćenjem funkcije `length()` možemo dobiti broj komponenti liste, te se korišćenjem

> `length(lPorodica)`

dobija izlaz:
`[1] 4.`

Ukoliko se komponente imenuju, onda se svakoj od njih može pristupiti putem identifikatora i simbola `'$'` na sledeći način:

ime.Liste$ime.Komponente.

Primetićemo da, ukoliko u $\mathcal{R}$ terminalu otkucamo ime liste, tj. `lPorodica`, i pritisnemo `Enter`, za izlaz dobijamo:

```
$muz
[1] "Milovan"

$zena
[1] "Ana"

$brDece
[1] 3

$godineDece
[1] 4 7 9
```

što nam daje uvid u našu listu, ali i pokazuje na koji način možemo da pristupimo svakoj komponenti liste (putem prethodno navedene *$ime.Komponente* notacije, što je i očekivano).

3.4.1 Konstrukcija, izmenjivanje i nadovezivanje listi

Nove liste se mogu konstruisati korišćenjem već postojećih objekata korišćenjem funkcije `list()`. Naredba

> `Lst <- list(name_1=object_1, ..., name_n=object_n)`

konstruiše listu `Lst` od n objekata `object_1, ..., object_n` i daje im odgovarajuća imena koja mogu biti slobodno odabrana. Ako se imena izostave, komponente će samo biti numerisane. Bitno je napomenuti da se objekti *kopiraju* pri pravljenju listi, te originalni objekti neće biti izmenjeni pri radu sa listom.

Recimo da imamo listu `Lista` koja ima četiri (4) komponente. Dodavanje nove komponente (recimo, neke već postojeće matrice `Mat`) se lako vrši pomoću komande

> `Lista[5] <- list(matrix=Mat).`

Recimo da imamo tri nezavisne liste `lista.A, lista.B` i `lista.C`. Ukoliko želimo da od ove tri liste napravimo jednu (recimo `lista.ABC`) koja će imati sve njihove komponente, to možemo postići sledećom komandom:

```
> lista.ABC <- c(lista.A, lista.B, lista.C),
```

dakle, korišćenjem funkcije `c()`.

3.4.2 Baze podataka

Baze podataka (*Data frame*) su posebna vrsta listi koje imaju klasu `data.frame`. **Restrikcije baze podataka** su funkcije restrikcije koje izvode bazu podataka iz listi, i to su sledeće restrikcije:

- Komponente moraju biti vektori (numerički, karakterski ili logički), faktori, numeričke matrice, liste ili druge baze podataka.
- Matrice, liste i baze podataka obezbeđuju novoj bazi podataka toliko novih promenljivih koliko imaju kolona, elemenata, odnosno, promenljivih, redom.
- Numerički i logički vektori i faktori se unose u bazu podataka takvi kakvi su, a karakterski vektori se prema podrazumevanom[6] čuvaju kao faktori čiji nivoi su sve različite vrednosti koje se nalaze u tom vektoru.
- Vektorske strukture koje se javljaju kao promenljive u bazi podataka moraju da imaju *istu dužinu*, a matrične strukture moraju da imaju *istu dužinu redova.*

Objekti koji zadovoljavaju restrikcije baze podataka mogu se koristiti pri formiraju baze podataka korišćenjem funkcije

```
> BPstudenti <- data.frame(indeksi=vIndeksi, imena=vImena, prezimena=vPrezimena,
  godine_upisa=vGodineUpisa).
```

Lista čije se komponente prilagođavaju restrikcijama baza podataka mogu biti *nasiljene* (*coerce*) u bazu podataka korišćenjem funkcije `as.data.frame()`.

Pristup je omogućen pomoću *$notacije*, na primer, pomoću

```
> BPstudenti$indeksi
```

možemo pristupiti vektoru indeksa studenata.

Nekada nam je potrebno da izdvojimo podatke koji zadovoljavaju neki određeni uslov, tačije, da izdvojimo **podbazu** čiji elementi zadovoljavaju neki uslov. Za to nam služi funkcija `subset()`. Recimo da hoćemo da izdvojimo sve studente ǩoji su se upisali posle 2015. godine. To možemo uraditi pomoću sledeće komande:

```
> BPrez = subset(BPstudenti, godine_upisa > 2015).
```

[6]Konverzija kolona karaktera u faktore se prevazilazi (*override*) uključivanjem `stringAsFactors` argumenta pri pozivu funkcije `data.frame()`.

4 Funkcije

Osim korišćenja nekih od standardnih funkcija sa kojima smo se do sada susretali, $\mathcal{R}$ omogućava i programeru da piše svoje funkcije, tj. da pravi objekte tipa *function*. **Funkcija** se definiše na sledeći način:

```
> name <- function(arg_1, arg_2, ...)  { expression }
```

pri čemu `expression` zapravo predstavlja jednu $\mathcal{R}$ naredbu ili grupnu naredbu, koja koristi argumente `arg_i` da bi izračunala vrednost. Vrednost izraza `expression` je ujedno i povratna vrednost funkcije. Poziv funkcije se obavlja pomoću naredbe

```
> name(param_1, param_2, ...)
```

i može se pojaviti na svim mestima na kojima je poziv funkcije dopušten. Na samom kraju funkcije se može staviti i naredba `return(value)`, gde je vrednost `value` zapravo vrednost izraza `expression`. Ukoliko se ne stavi `return()`, onda se vrednost poslednjeg reda funkcije smatra povratnom vrednošću funkcije. Navedimo jedan jednostavan primer funkcije:

```
# Funkcija "popust" izračunava popust na artikle na sledeći način:
# > Na artikle cene 1000 - 1999 popust je 10%, i
# > Na artikle cene veće od 1999 popust je 20%.

popust <- function(cene){
    cena <- cene
    n <- length(cene)
    for(i in 1:n){
        if(cene[i] > 1999)
            cena[i] <- 0.8 * cene[i]
        else if(cene[i] >= 1000)
            cena[i] <- 0.9 * cene[i]
        else
            cena[i] = cene[i]
    }
    return(cena)
}
```

5 Grafičko predstavljanje podataka

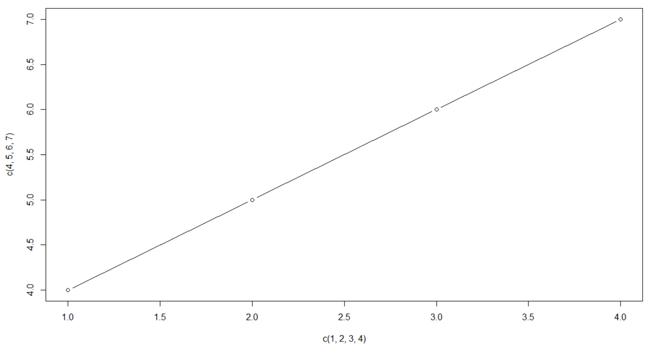
Za najosnovnije grafičko predstavljanje pomoću Dekartovog koordinatnog sistema koristi se funkcija `plot()`. Ona omogućava iscrtavanje $\mathcal{R}$ objekata na različite načine. Definicija ove funkcije je:

```
> plot(x, y, ...)
```

gde `x` predstavlja koordinate na x-osi (alternativno, pojedinačne strukture, funkcija ili bilo koji $\mathcal{R}$ objekat sa dostupnim `plot` metodom), `y` predstavlja koordinate na y-osi (što je opciono ako je `x` odgovarajuća struktura), a `...` razni argumenti pri čemu ćemo navesti samo sledeće:

- `type` – tip nacrta koji će biti prikazan. Među dostupnima su:
  - `"p"` za tačke
  - `"l"` za linije
  - `"b"` za oboje (tačke i linije)
  - `"c"` za `"b"` prikaz bez tačaka
  - `"h"` za histogram
  - `"s"` i `"S"` za stepeničaste prikaze
- `main` – naslov nacrta
- `sub` – podnaslov nacrta
- `xlab` i `ylab` – naslov za x-osu i y-osu
- `pch` – oblik tačaka
- `lwd` – debljina linije
- `xlim` i `ylim` – granica za x-osu i y-osu. Koristi se na sledeći način:

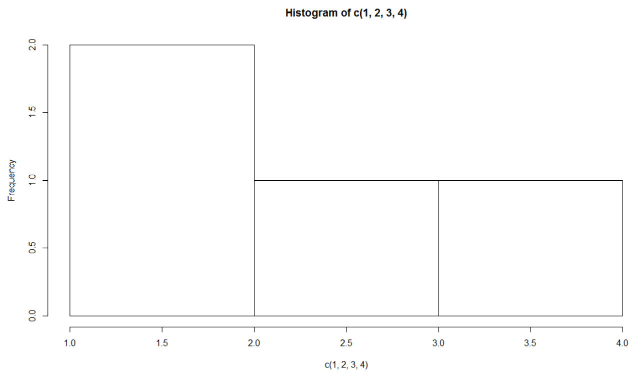
```
> xlim <- c(a, b)
  # a i b su koordinate granica
  # slicno za granice y-ose
```



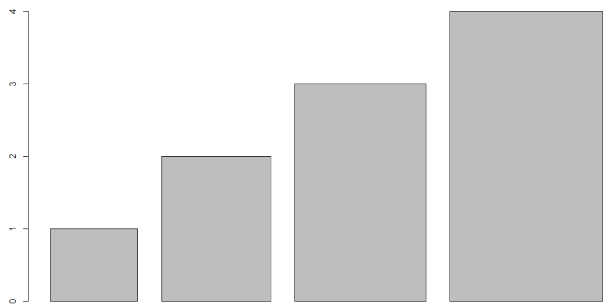
Slika 2. Primer jednostavnog iscrtavanja pomoću funkcije `plot()`

Osim navedene funkcije `plot()`, postoje i druge brojne funkcije koje iscrtavaju podatke na različite grafičke načine. Navedimo neke od njih:

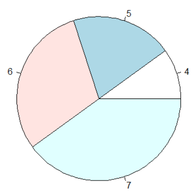
- `hist()` – iscrtavanje histograma
- `barplot()` – iscrtavanje slično histogramu, ali sa razmacima između kolona
- `pie()` – iscrtavanje pita-dijagrama



Slika 3. Primer jednostavnog iscrtavanja pomoću funkcije `hist()`



Slika 4. Primer jednostavnog iscrtavanja pomoću funkcije `barplot()`



Slika 5. Primer jednostavnog iscrtavanja pomoću funkcije `pie()`

6 Raspodele

Jedan pogodna upotreba $\mathcal{R}$-a jeste obezveđivanja skupa statističkih tabela. Za svaku raspodelu može se izračunati:

- Vrednost funkcije gustine ili verovatnoće u tački – `d`-funkcije,
- Vrednost funkcije raspodele u tački – `p`-funkcije,
- Kvantili – `q`-funkcije, i
- Pseudoslučajni brojevi – `r`-funkcije.

Sledi tabela raspodela koje se koriste u $\mathcal{R}$-u:

Distribution	$\mathcal{R}$ **name**	**additional arguments**
beta	beta	shape1, shape2, ncp
binomial	binom	size, prob
Cauchy	cauchy	location, scale
chi-squared	chisq	df, ncp
exponential	exp	rate
F	f	df1, df2, ncp
gamma	gamma	shape, scale
geometric	geom	prob
hypergeometric	hyper	m, n, k
log-normal	lnorm	meanlog, sdlog
logistic	logis	location, scale
negative	binomial	nbinom size, prob
normal	norm	mean, sd
Poisson	pois	lambda
signed	rank	signrank n
Student's t	t	df, ncp
uniform	unif	min, max
Weibull	weibull	shape, scale
Wilcoxon	wilcox	m, n

Navedimo sada primere nekih od mnogo postojećih funkcija za rad sa raspodelama:

```
> runif(20) # Dobija se vektor od 20 pseudoslučajnih brojeva iz uniformne raspodele
  # intervala U[0,1]
```

```
> dbinom(4, size <- 20, prob <- 0.2) # Dobija se verovatnoća da se od 20 nezavisnih
  # eksperimenata (pri čemu je verovatnoća realizacije 0.2) realizuju 4 (pri binomnoj
  # raspodeli)
```

```
> pnorm(0.3, mean <- 0, sd <- 1) # Dobija se vrednost funkcije raspodele sl.  veličine
  # koja ima N(0,1) (tj.  normalnu) raspodelu u tački x = 0.3
```

```
> qchisq(p <- 0.95, df <- 4) # Dobija se vrednost x u kome je vrednost funkcije raspodele
  # 0.95
```

7 Zadaci i rešenja

7.1 Verovatnoća

II NEDELJA

1. Eksperiment glasi: Baca se kockica N puta, treba odrediti frekvenciju bacanja šestice ako je kocka homogena, tj. $P(i) = \frac{1}{6}$, $i = \overline{1,6}$. Napraviti funkciju koja iscrtava histogram od n ponovljenih eksperimenata.

```
broj6<-function(N) {
br = 0
    for(i in 1:N) {
        x = runif(1) #runif je funkcija koja vraca slucajan broj izmedju 0 i 1
        if(x < 1/6) br = br + 1
    }
    return (br/N) #Ovo vracamo ako hocemo verovatnocu
                  #Vracamo return (br/N - 1/6) ako zelimo frekvenciju oko 0
}

iscrtajHistogram<-function(N, n) {
    odstupanja = c()
    for(i in 1:n)
    odstupanja = c(odstupanja, broj6(N))
    hist(odstupanja)
}
```

2. Napraviti funkciju koja vraća jedan broj od 1 do 6 tako da su svi jednako verovatni.

```
brojNaKocki<-function(n = 1) {
    x = runif(1)
    return(floor(6*x) + 1)
}
```

3. (a) Neka igrač ima početni kapital A i on baca kockicu. Ako dobije 1, 3 ili 5, njegov kapital se uveća za 1 dinar, a u suprotnom izgubi 1 dinar. Za N ponovljenih bacanja, nacrtati trajektoriju njegovog kapitala u zavisnosti od trenutnog bacanja.

```
kapital<-function(a = 20, n = 50) {
    trajektorija = c(a)
    trenutno = a
    for(i in 1:n) {
        x = runif(1)
        if(x < 1/2)
            trenutno = trenutno + 1
        else
            trenutno = trenutno - 1
        trajektorija = c(trajektorija, trenutno)
    }
    return (trajektorija)
}

trajekt = kapital()
plot(0, 50, trajekt)
trajekt = kapital(n = 100
plot(0, 100, trajekt, type = "l")
lines(0, 100, trajekt, type = "l", col = "red")
```

(b) Oceniti verovatnoću da je igrač u plusu nakon n partija?

```
ocenaVerovatnoce<-function(N = 1000,  n = 50, a = 0) {
    brojac = 0
    for(i in 1:N) {
        trajektorija = kapital(a, n)
        if(trajektorija[n+1] >= 0)
        brojac = brojac + 1
    }
    return (brojac / N)
}

ocenaVerovatnoceNaKraju<-function(N=10000, n = 50, a = 0) {
    brojacPlus = 0
    brojacMinus = 0
    for(i in 1:N) {
        trajektorija = kapital(a, n)
        if(trajektorija[n+1] > 0)
        brojacPlus = brojacPlus + 1
        else if(trajektorija[n+1] < 0)
        brojacMinus = brojacMinus + 1
    }
    sprintf("ocena verovatnoce da je na kraju u plusu je %f, a u minusu je %f ",
        brojacPlus/N,brojacMinus/N)
}
```

4. Neka porodica pravi decu sve dok ne dobije muško dete. Odrediti prosečan broj dece u porodici ako se testira 100 000 porodica. Pretpostavka je da je $P(\text{muško dete}) = P(\text{žensko dete}) = \frac{1}{2}$.

```
brDece<-function(Nmax = 10, p = 1/2) {
    OK = TRUE
    brojac = 0
    while(OK && (brojac < Nmax + 1)) {
        x = runif(1)
        if(x < p)
    OK = FALSE
        brojac = brojac + 1
    }
    return (brojac)
}

ocenaBrDece<-function(N, p = 1/2, Nmax = 10) {
    br = 0
    for(i in 1:N)
        br = br + brDece(Nmax, p)
    return (br / N)
}
```

5. Tri kockice za igru se bacaju n puta odjednom. Oceniti verovatnoću da se bar jednom u zavisnosti od n dogode sve tri šestice.

```
#vraca 1 ako se pojave sve 3 sestice, inace 0
nBacanja<-function(n) {
    OK = TRUE
    i = 1
    while((i <= n) && OK){
```

```
        triKockice = runif(3) # ako su svi manji od 1/6 dobili smo 3 sestice
        if(triKockice[1] < 1/6 && triKockice[2] < 1/6 && triKockice[3] < 1/6)
        OK = FALSE
        i = i + 1
    }
    if(OK)
        return(0)
    else
        return(1)
}

# N je broj ponavljanja eksperimenta
triSesticeUN<-function(n, N) {
    brojac=0
    for(i in 1:N)
        brojac = brojac + nBacanja(n)
    return(brojac / N)
}
```

III NEDELJA

6. Ako imamo skup elementarnih događaja $\Omega = \{a, b, c\}$ i njihove verovatnoće $P(a) = 1/2,\ P(b) = 1/3,\ P(c) = 1/6$, odrediti verovatnoću svih mogućih događaja.

   $P(\{a\}) = 1/2,$
   $P(\{b\}) = 1/3,$
   $P(\{c\}) = 1/6,$
   $P(\{a, b\}) = P(a) + P(b) = 1/2 + 1/3 = 5/6,$
   $P(\{a, c\}) = P(a) + P(c) = 1/2 + 1/6 = 4/6,$
   $P(\{b, c\}) = P(b) + P(c) = 1/3 + 1/6 = 3/6,$
   $P(\{a, b, c\}) = P(a) + P(b) + P(c) = 1/2 + 1/3 + 1/6 = 1.$

7. Ako je $P(AB) = 1/4,\ P(A^c) = 1/3$ i $P(B) = 1/2$, odrediti $P(A \cup B)$.

   $A = (A \backslash B) \cup (AB)$
   $B = (B \backslash A) \cup (AB)$
   $A \cup B = (A \backslash B) \cup (AB) \cup (B \backslash A)$
   $P(A) = 1 - P(A^c) = 1 - 1/3 = 2/3$
   $\Longrightarrow P(A) + P(B) = P(A \backslash B) + P(AB) + P(B \backslash A) + P(AB) = P(A \cup B) + P(AB)$
   $\Longrightarrow P(A \cup B) = P(A) + P(B) - P(AB) = 2/3 + 1/2 - 1/4 = 11/12.$

8. Označimo sa A, B i C ocene koje mogu da se dobiju na jednom predmetu. Predmet polažu John i Mary. Verovatnoća da John dobije ocenu B je 0,3. Verovatnoća da Mary dobije ocenu B je 0,4. Verovatnoća da niko ne dobije ocenu A, ali da bar neko dobije ocenu B je 0,1. Pronaći verovatnoću da bar neko dobije ocenu B, ali da niko ne dobije ocenu C.

   Neka je $\Omega = \{A, B, C\}$. Posmatraćemo uređene parove oblika (J, M) iz skupa $\Omega \times \Omega$, pri čemu prva koordinata označava ocenu koju John dobija, a druga koordinata ocenu koju Mary dobija. Treba primetiti da važe sledeće ekvivalencije:

   $$P(John\ gets\ B) = P((B, A)) + P((B, B)) + P((B, C)) = 0,3 \quad (1)$$
   $$P(Mary\ gets\ B) = P((A, B)) + P((B, B)) + P((C, B)) = 0,4 \quad (2)$$
   $$P(No\ one\ gets\ A;\ at\ least\ someone\ gets\ B) = P((B, B)) + P((B, C)) + P((C, B)) = 0,1 \quad (3)$$

   Iz $(1) + (2) - (3)$ sledi:

   $P((B, B)) + P((B, A)) + P((A, B)) = 0,6$, što je tražena verovatnoća.

9. Student treba da odabere dva predmeta od ponuđenih tri: umetnost, francuski jezik i matematika. Verovatnoća da odabere umetnost je $5/8$. Verovatnoća da odabere francuski jezik je $5/8$. Verovatnoća da odabere umetnost i francuski jezik je $1/4$. Odrediti verovatnoću da odabere matematiku.

   Slično prethodnom zadatku, mogući odabiri su $(U, F),\ (U, M)$ i (F, M). Stoga,

   $P(U, F) + P(U, M) = 5/8$
   $P(U, F) + P(F, M) = 5/8$
   $P(U, F) = 1/4$

   $$\Longrightarrow P(U, M) = 5/8 - P(U, F) = 5/8 - 1/4 = 3/8 \quad (4)$$
   $$P(F, M) = 5/8 - P(U, F) = 5/8 - 1/4 = 3/8 \quad (5)$$

   Iz $(4) + (5)$ sledi:

   $P(U, M) + P(F, M) = 6/8.$

`Teorija>` **Slučajna veličina**

Svaka funkcija $X : \Omega \longrightarrow \mathbf{R}$ naziva se *slučajna veličina*. Primer slučajne veličine je funkcija koja ishodima skupa $\Omega = \{1, 2, 3\}$ dodeljuje $-1, 1$ i 2, redom. Ukoliko su verovatnoće ovih veličina, na primer, $(1/2, 1/3, 1/6)$, onda se $X : \begin{pmatrix} -1 & 1 & 2 \\ 1/2 & 1/3 & 1/6 \end{pmatrix}$ naziva *zakon raspodele*. Dajmo još jedan primer. Neka je $Y : \{a \mapsto 1, b \mapsto 5, c \mapsto 5\}$ sa verovatnoćama $(1/3, 2/3)$, onda je zakon raspodele u ovom primeru $Y : \begin{pmatrix} 1 & 5 \\ 1/3 & 2/3 \end{pmatrix}$.

10. Neka je $X : \begin{pmatrix} -1 & 0 & 1 & 2 \\ 1/5 & 1/5 & 2/5 & 1/5 \end{pmatrix}$ i neka su $Y = X + 3$ i $Z = X^2$. Odrediti zakon raspodele za Y i Z.

Pošto je Y linearno zavisna od X, onda je

$$Y : \begin{pmatrix} -1+3 & 0+3 & 1+3 & 2+3 \\ 1/5 & 1/5 & 2/5 & 1/5 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 4 & 5 \\ 1/5 & 1/5 & 2/5 & 1/5 \end{pmatrix},$$

a kako je $Z = X^2$, i $(-1)^2 = (1)^2$, to je

$$Z : \begin{pmatrix} 0^2 & (-1)^2 = 1^2 & 4 \\ 1/5 & 1/5 + 2/5 & 1/5 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 4 \\ 1/5 & 3/5 & 1/5 \end{pmatrix}.$$

`Teorija>` **Uniformna raspodela**

Ako X uzima vrednosti $x_1, ..., x_n$ sa verovatnoćama $1/n$, onda kažemo da X ima **uniformnu raspodelu**.

Primer: Bacanje kockice, gde je X broj na kockici. Tada je:

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

Binomna raspodela

Pretpostavimo da ponavljamo n puta neki eksperiment pod istim uslovima. Označimo sa P verovatnoću uspeha pri tim ponavljanjima. Zanima nas broj uspešnih eksperimenata. U ovom slučaju je $X \in \{0, 1, 2, ..., n\}$, jer može da se desi da se nijednom nije desio uspeh, da se desio jedanput, dvaput, ..., ili da se svaki put desio uspeh. Da je broj uspešnih eksperimenata tačno k možemo da napišemo $P\{X = k\}$, za šta se ispostavlja da važi: $P\{X = k\} = \binom{n}{k} P^k (1-P)^{n-k}$. Tada kažemo da X ima **binomnu raspodelu**.

Primer: Ispitujemo broj uspešnosti 2 eksperimenta od 4 ponavljanja. Sa $+$ obeležavamo da je eksperiment uspeo, a sa $-$ da nije. Tada imamo sledeće mogućnosti:

$$\begin{array}{ccc}
++-- & \mapsto & P \cdot P \cdot (1-P) \cdot (1-P) \\
+-+- & \mapsto & P \cdot (1-P) \cdot P \cdot (1-P) \\
+--+ & \mapsto & P \cdot (1-P) \cdot (1-P) \cdot P \\
-++- & \mapsto & (1-P) \cdot P \cdot P \cdot (1-P) \\
--++ & \mapsto & (1-P) \cdot (1-P) \cdot P \cdot P \\
-+-+ & \mapsto & (1-P) \cdot P \cdot (1-P) \cdot P
\end{array}$$

Vidimo da je broj ovakvih mogućnosti jednak $6 \cdot P^2 \cdot (1-P)^2$, što je jednako $\binom{4}{2} \cdot P^2 \cdot (1-P)^{4-2}$.

11. Četiri puta se gađa meta sa verovatnoćom uspeha $2/3$. Odrediti verovatnoću da je bilo dva ili tri pogotka.

Ovaj zadatak se radi pomoću binomne raspodele. Prikazaćemo koje formule bi trebalo koristiti pri rešavanju ovakvih zadataka, ali i kako nam $\mathcal{R}$ može pomoći da brže izračunamo traženu verovatnoću:

$$\begin{aligned} P\{X=2\}+P\{X=3\} &= \binom{4}{2}\cdot(2/3)^2\cdot(1/3)^{4-2}+\binom{4}{3}\cdot(2/3)^3\cdot(1/3)^{4-3}= \\ &= dbinom(2,4,2/3)+dbinom(3,4,2/3)\approx \\ &\approx 0,69. \end{aligned}$$

12. Koš se gađa dvadeset puta. Ako je verovatnoća uspeha $2/3$, odrediti verovatnoću:
    (a) da koš nije pogođen nijednom, i
    (b) da je koš pogođen barem dva puta.

(a)

$$\begin{aligned} P(nijednom) &= (1-P)^{20}=(1-2/3)^{20}\approx 0,0003,\ ili \\ &= dbinom(0,20,2/3)\approx 0,0003. \end{aligned}$$

(b)

$$\begin{aligned} P(barem\ 2) &= \sum_{k=2}^{20} P(bilo\ je\ k\ pogotka)= \\ &= 1-P(nijednom)-P(jedan\ pogodak)= \\ &= 1-dbinom(0,20,1/3)-P(1,20,1/3)\approx 0,99. \end{aligned}$$

`Teorija>` **Geometrijska raspodela**

Bacamo novčić dok ne padne glava. X je broj bacanja. Dakle, ako sa P označimo da je palo pismo, a sa G da je pala glava, onda je $\Omega=\{G,PG,PPG,...\}$ i $X\in\{1,2,3,...\}$. U ovom primeru kažemo da X ima **geometrijsku raspodelu** sa parametrom $1/2$. Ukoliko nas, pak, zanima broj bacanja *do uspeha*, označimo to sa Y, onda je $Y\in\{0,1,2,...\}$ i važi veza $Y=X-1$. Tada je verovatnoća da je bilo k bacanja novčića jednaka $(1/2)^{k-1}\cdot 1/2=(1/2)^k$.

Ako verovatnoću uspeha označimo sa P, onda je zakon geometrijske raspodele $P\{X=k\}=(1-P)^{k-1}\cdot P$.

Utvrdimo da li je ovaj zakon dobro definisan. Treba da važi $\sum P(w_i)=1$:

$$\begin{aligned} \sum P(w_i) &= \sum_{i=1}^{\infty}(1-P)^i\cdot P=P(1+(1-P)+(1-P)^2+...)= \\ &= P\cdot\frac{1}{1-(1-P)}=P\cdot\frac{1}{P}=1, \end{aligned}$$

te je zakon dobro definisan.

Primer: Posmatramo $Y\in\{0,1,2,...\}$. Označimo sa N da je eksperiment bio neuspešan, a sa U da je bio uspešan. Tada je $P\{Y=k\}=(1-P)^k\cdot P$.

Napomena: U $\mathcal{R}$-u je za funkciju $dgeom()$ implementirano Y, tako da, ako se traži ukupan broj bacanja, onda mora da se doda 1 na odgovarajuću vrednost.

13. Kockica za igru se baca do pojave petice. Odrediti verovatnoću:

    (a) da je broj bacanja između 2 i 4,
    (b) da je bilo najviše 3 bacanja,
    (c) da je bilo više od 3 bacanja,
    (d) da je bio neparan broj bacanja, i
    (e) da je bio paran broj bacanja.

    Poznato nam je da je verovatnoća da padne bilo koji broj (pa samim tim i petica) jednaka $P = 1/6$.

    (a)

$$\begin{aligned} P\{2 \leq X \leq 4\} &= P\{X = 2\} + P\{X = 3\} + P\{X = 4\} = \\ &= P\{Y = 1\} + P\{Y = 2\} + P\{Y = 3\} = \\ &= dgeom(1, 1/6) + dgeom(2, 1/6) + dgeom(3, 1/6) \approx \\ &\approx 0,35. \end{aligned}$$

    (b)

$$\begin{aligned} P\{X \leq 3\} &= P\{X = 1\} + P\{X = 2\} + P\{X = 3\} = \\ &= P\{Y = 0\} + P\{Y = 1\} + P\{Y = 2\} = \\ &= dgeom(0, 1/6) + dgeom(1, 1/6) + dgeom(2, 1/6) \approx \\ &\approx 0,42. \end{aligned}$$

    (c)

$$1 - P\{X \leq 3\} \approx 1 - 0,42 \approx 0,58.$$

    (d)

$$\begin{aligned} \sum_{k=0}^{\infty} P\{2k + 1 \ bacanja\} &= \sum_{k=0}^{\infty} (1 - P)^{2k} \cdot P = \\ &= P \cdot (1 + (1 - P)^2 + (1 - P)^4 + ...) = \\ &= P \cdot \frac{1}{1 - (1 - P)^2} = \\ &= \frac{1/6}{1 - 35/36} = \\ &= 6/11. \end{aligned}$$

    (e)

$$P\{paran \ broj \ bacanja\} = 1 - 6/11 = 5/11.$$

14. Igrači A i B izvlače po jednu kartu iz špila, a zatim je vraćaju natrag. Pobednik je onaj koji prvi izvuče damu. Izvlačenje počinje igrač A.

    (a) Odrediti verovatnoću da je A pobednik.
    (b) Odrediti verovatnoću da je B pobednik.

    (a) $P(dama\ je\ izvučena) = 4/52 = 1/13$. Da je igrač A pobedio znači da je dama izvučena u $1, 3, 5, ...$ izvlačenju, pa na osnovu računa iz prethodnog zadatka sledi:
    $P(A\ je\ pobedio) = \dfrac{P}{1 - (1 - P)^2} = \dfrac{1/13}{1 - 144/169} = 13/25 = 0,52.$

(b) Primetimo prvo da važi: $P(A\ je\ pobedio) + P(B\ je\ pobedio) + P(niko\ nije\ pobedio) = 1$, te ne možemo odmah da napišemo $P(B\ je\ pobedio) = 1 - P(A\ je\ pobedio)$. Međutim, ispostavlja se da važi sledeće:
$P(niko\ nije\ pobedio) = \lim_{n\to\infty} (12/13)^n = 0$, te u ovom slučaju ipak važi: $P(B\ je\ pobedio) = 1 - P(A\ je\ pobedio)$. Sledi:
$P(B\ je\ pobedio) = 1 - 13/25 = 12/25 = 0,48$. Napomenimo još jednom da je neophodno bilo eksplicitno izračunati da je verovatnoća dodađaja da niko nije pobedio jednaka nuli.

IV NEDELJA

15. Oceniti vrednost broja π ponavljanjem sledećeg eksperimenta: Slučajno se bira tačka unutar jediničnog kvadrata i registruje da li je tačka unutar upisanog kruga u dati kvadrat.

$$P = \frac{\textit{Broj uspešnih eksperimenata}}{\textit{Broj ukupnih eksperimenata}} = \frac{\Pi(kruga)}{\Pi(kvadrata)} = \frac{\frac{1}{4}\cdot\pi}{1} = \frac{\pi}{4} \Longrightarrow \pi = 4\cdot P.$$

Označimo broj ukupnih eksperimenata sa N. Simulirajmo ovaj eksperiment u $\mathcal{R}$-u:

```
oceniPI<-function(N) {
    s <- 0
    for(i in 1:N) {
        z <- runif(2)
        #z[1] je x koordinata, a z[2] je y koordinata
        if((z[1] - 1/2)^2 + (z[2] - 1/2)^2 < 1/4)
            s <- s + 1
    }
    return(4*s/N)
}
```

Ovaj kôd možemo dodatno optimizovati, bez upotrebe `for` petlje na sledeći način:

```
oceniPI<-function(N) {
    #izmenili smo koordinatni pocetak u centar kvadrata, ali moze i bez toga
    x <- runif(N, min = -1, max = 1)
    y <- runif(N, min = -1, max = 1)
    tacke <- ifelse((x^2 + y^2) <= 1, 1, 0)
    brojTacaka <- sum(tacke)
    return(4*brojTacaka/N)
}
```

16. Oceniti vrednost broja π računanjem integrala $\int\limits_0^1 \sin\pi x dx$.

Verovatnoća događaja da se slučajno izabrana tačka iz intervala $[0,1]\times[0,1]$ nalazi u oblasti grafika ispod integrala $\int\limits_0^1 \sin\pi x dx$ je $\hat{P} = \frac{P}{1} = P$, gde je $P = \int\limits_0^1 \sin\pi x dx = \int\limits_0^\pi \frac{\sin y dy}{\pi} = \frac{\cos y}{\pi}\Big|_0^\pi = \frac{2}{\pi}$.

```
oceniPI2<-function(N) {
    x = runif(N, min = 0, max = 1)
    y = runif(N, min = 0, max = 1)
    tacke <- ifelse(y <= sin(pi*x), 1, 0)
    brojTacaka <- sum(tacke)
    return(2/(brojTacaka/N))
}
```

17. Oceniti vrednost broja π koristeći eksperiment Bufonove igle.

Bacamo iglu dužine 1 na sto sa linijama razdaljine 1. Brojimo koliko puta je igla presekla linije. U odnosu na najbližu liniju gledamo udaljenost od centra igle, što označavamo sa d, i oštar ugao pod kojim je igla pala, što označavamo θ. Važi da je $d \in [0, \frac{1}{2}]$ i $\theta \in [0, \frac{\pi}{2}]$. Može se pokazati da je uslov da igla seče liniju: $d < \frac{1}{2}\sin\theta$.

```
oceniBufon <- function(N) {
    d = runif(N, 0, 1/2)
```

```
    teta = runif(N, 0, pi/2)
    uslov <- ifelse(d <= sin(teta), 1, 0)
    return(sum(uslov)/N)
}
```

`Teorija>` **Funkcija raspodele i funkcija gustine**

Funkcija

$$F(x) = P\{X \leq x\}$$

naziva se **funkcija raspodele**. Važno je zapamtiti da je ona monotono rastuća, kao i da je $\lim\limits_{x\to\infty} F(x) = 1$ i $\lim\limits_{x\to-\infty} F(x) = 0$. Važi sledeće:

Neka je X apsolutno-neprekidna slučajna veličina koja ima **funkciju gustine** $f(x)$. Tada je

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt,$$

i važi $f(t) \geq 0$ i $\int\limits_{-\infty}^{\infty} f(t)dt = 1$.

Primer: X ima uniformnu raspodelu u intervalu $[a, b]$ ako

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & inače. \end{cases}$$

Verovatnoća da se tačka nađe u intervalu $[a_1, b_1] \subseteq [a, b]$ je:

$$P\{a_1 < X \leq b_1\} = F(b_1) - F(a_1) = \int\limits_{a_1}^{b_1} \frac{1}{b-a} dx = \frac{b_1 - a_1}{b-a},$$

što se podudara sa onim što smo intuitivno zaključivali, tj.

$$P = \frac{X \ se \ nalazi \ u \ [a_1, b_1]}{X \ se \ nalazi \ u \ [a, b]} = \frac{b_1 - a_1}{b-a}.$$

18. Pretpostavimo da je X nalazi u intervalu $[0, 1]$ uniformno raspoređenom. Neka je $Y = \sqrt{X}$. Naći funkciju raspodele i funkciju gustine slučajne veličine Y, tj. naći $F_Y(y)$ i $f_Y(y)$.

$$F_Y(y) = P\{Y \leq y\} = P\{\sqrt{X} \leq y\} = \begin{cases} 0, & y < 0 \\ P\{X \leq y^2\} = F_X(y^2) \overset{(J1)}{=} y^2, & y \in [0, 1] \\ 1, & y > 1 \end{cases}$$

Možda nije najjasnije zbog čega važi jednakost obeležena sa $(J1)$. Izvedimo ovo u opštem slučaju:

Neka je X iz uniformno raspoređenog intervala $[0, 1]$. Tada je:

$$F_X(x) = \int\limits_{-\infty}^{x} f(t)dt = \begin{cases} 0, & x < a \\ \int\limits_{a}^{x} \frac{1}{b-a} dt = \dfrac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}.$$

19. Pretpostavimo da smo odabrali X iz intervala $[2, 10]$ sa funkcijom gustine $f(x) = cx$, gde je c konstanta. Pronaći:

    (a) vrednost konstante c,

(b) $P[E]$, gde je $E = [a_1, b_1]$ podinterval od $[2, 10]$, i

(c) $P\{x > 5\}, P\{x < 7\}$ i $P\{x^2 - 12x + 35 > 0\}$.

(a)

$$\int_2^{10} cxdx = 1 \Longrightarrow \frac{c}{2}(100 - 4) = 1 \Longrightarrow c = \frac{2}{96} = \frac{1}{48}.$$

(b)

$$P([a_1, b_1]) = \int_{a_1}^{b_1} \frac{xdx}{48} = \frac{1}{96}(b_1^2 - a_1^2).$$

(c)

$$P\{x > 5\} = P\{(5, 10]\} = \frac{1}{96}(100 - 25) = 75/96$$
$$P\{x < 7\} = P\{[2, 7)\} = \frac{1}{96}(49 - 4) = 45/96$$
$$P\{x^2 - 12x + 35 > 0\} = P\{(x - 5)(x - 7) > 0\} = (*)$$

1° $(x - 5 > 0 \Longleftrightarrow x > 5) \wedge (x - 7 > 0 \Longleftrightarrow x > 7) \Longrightarrow x > 7$
2° $(x - 5 < 0 \Longleftrightarrow x < 5) \wedge (x - 7 < 0 \Longleftrightarrow x < 7) \Longrightarrow x < 5$

$$(*) = P\{x > 7\} + P\{x < 5\} = 1 - 45/96 + 1 - 75/96 = 0,75.$$

V NEDELJA

`Teorija>` Neka imamo brojeve B i C koji se dobijaju iz intervala $[0,1]$ na slučajan način (uniformno). Verovatnoću izbora ovih brojeva možemo posmatrati kao verovatnoću izbora tačke (B,C) iz kvadrata $K=[0,1]\times[0,1]$ u BOC Dekartovom koordinatnom sistemu, a ta je verovatnoća jednaka površini kvadrata K, tj. $\Pi(K)$.

20. Izračunati sledeće verovatnoće:

   (a) $P\{BC<\frac{1}{2}\}=P\{C<\frac{1}{2B}\}=\Pi([0,\frac{1}{2}]\times[0,1])+\Pi(,,\mathit{oblast\ ispod\ grafika\ funkcije}\ C=\frac{1}{2B},\ \mathit{za}\ B\in[\frac{1}{2},1]")=\frac{1}{2}+\int\limits_{\frac{1}{2}}^{1}\frac{1}{2B}dB=\frac{1}{2}+\frac{1}{2}\ln B\Big|_{\frac{1}{2}}^{1}=\frac{1}{2}-\frac{1}{2}\ln 2\approx 0,84.$

   (b) Neka je $A_1=|B-C|<\frac{1}{2}$. Tada je $P\{A_1\}=P\{-\frac{1}{2}<B-C<\frac{1}{2}\}=P\{B-\frac{1}{2}<C<B+\frac{1}{2}\}=1-2\cdot\frac{\frac{1}{2}\cdot\frac{1}{2}}{2}=1-\frac{1}{4}=\frac{3}{4}$.

   (c) Neka je $A_2=B<\frac{1}{2},C>\frac{1}{2}$. Tada je $P\{A_2\}=\Pi(,,\mathit{drugi\ kvadrant\ kvadrata}\ K")=\frac{1}{4}$.

   (d) $P\{A_1A_2\}=\Pi(,,\mathit{presek\ oblasti}\ A_1\ \mathit{i}\ A_2")=\frac{1}{8}$.
   Primetimo sledeće: Kako važi $P\{A_1A_2\}=\frac{1}{8}\neq\frac{3}{16}=\frac{3}{4}\cdot\frac{1}{4}=P\{A_1\}\cdot P\{A_2\}$, to događaji A_1 i A_2 nisu nezavisni.

   (e) $P\{A_1|A_2\}=\frac{P\{A_1A_2\}}{P\{A_2\}}=\frac{\frac{1}{8}}{\frac{1}{4}}=\frac{1}{2}$.

   (f) $P\{max(B,C)<\frac{1}{2}\}=\Pi(,,\mathit{treci\ kvadrant\ kvadrata}\ K")=\frac{1}{4}$.

   (g) $P\{(B-\frac{1}{2})^2+(C-\frac{1}{2})^2<\frac{1}{4}\}=\frac{\Pi(\mathit{kruga\ sa\ centrom\ u}\ (\frac{1}{2},\frac{1}{2})\ \mathit{i\ poluprecnika}\ \frac{1}{2})}{\Pi(K)}=\frac{\frac{\pi}{4}}{1}=\frac{\pi}{4}$.

21. Igramo igru pokera: U špilu postoje 52 karte: 4 znaka po 13 vrednosti. Izvlačimo 5 karata. Posmatrajmo događaje:

   (a) A – izvukli smo boju (izraz koji govori da su karte u istom znaku, ali nisu 5 uzastopnih vrednosti), i

   (b) B – izvukli smo dva para (dva puta po ista vrednost i jedna karta različite vrednosti od prethodnih 4).

   Izračunati verovatnoće događaja A i B.

   (a) Neka je A_0 događaj da smo izvukli 5 karata iste boje. Tada je $n(A_0)=\binom{13}{5}\cdot 4$. Neka je A_1 događaj da su karte u istom znaku i da idu redom. Tada je $n(A_1)=4\cdot 10$ (ima ih 10 jer ne može red da počne nekom od karata J, Q ili K). Očigledno je $A_1\subset A_0$, a mi tražimo $A=A_0\backslash A_1$. Odavde je $n(A)=n(A_0)-n(A_1)$, pa je $P\{A\}=\frac{n(A)}{n}=\frac{\binom{13}{5}\cdot 4-4\cdot 10}{\binom{52}{5}}\approx 0,02$.

   (b) Vrednost $n(B)$ možemo izračunati na dva načina:

   - $$n(B)=\underbrace{\binom{13}{2}}_{\substack{\mathit{izvucemo}\\ \mathit{dva\ znaka}}}\ \underbrace{\binom{4}{2}}_{\substack{\mathit{par\ prvog}\\ \mathit{znaka}}}\ \underbrace{\binom{4}{2}}_{\substack{\mathit{par\ drugog}\\ \mathit{znaka}}}\ \underbrace{\binom{44}{1}}_{\substack{\mathit{peta}\\ \mathit{karta}}},\ \text{ili}$$
   - $$n(B)=\underbrace{\underbrace{\binom{13}{1}}_{\substack{\mathit{izvucemo}\\ \mathit{prvi\ znak}}}\ \underbrace{\binom{4}{2}}_{\substack{\mathit{biramo\ par}\\ \mathit{prvog\ znaka}}}\ \underbrace{\binom{12}{1}}_{\substack{\mathit{izvucemo}\\ \mathit{drugi\ znak}}}\ \underbrace{\binom{4}{2}}_{\substack{\mathit{biramo\ par}\\ \mathit{drugog\ znaka}}}\ \underbrace{\binom{44}{1}}_{\substack{\mathit{peta}\\ \mathit{karta}}}}_{\substack{\mathit{u\ ovom\ slucaju\ racunamo\ duplo,\ jer\ su,\ na\ primer,}\\ \mathit{kombinacije:}\\ 10\clubsuit\ 10\heartsuit\ Q\clubsuit\ Q\diamondsuit\ 1\diamondsuit,\ \mathit{i}\\ Q\clubsuit\ Q\diamondsuit\ 10\clubsuit\ 10\heartsuit\ 1\diamondsuit\\ \mathit{u\ ovom\ slucaju\ razlicite\ (iako\ ih\ mi\ sustinski\ ne\ razlikujemo)}\\ \mathit{pa\ sve\ moramo\ deliti\ 2.}}}\cdot\frac{1}{2}$$

Odavde sledi da je $P\{B\} = \frac{n(B)}{n}$.

22. Zadat je ispit sa 10 pitanja sa mogućim odgovorima DA/NE. Za prolazak na ispitu potrebno je 70% tačnih odgovora. Izračunati:

    (a) verovatnoću da student položi ispit ako nasumično zaokružuje;
    (b) verovatnoću da student ima više od dva tačna odgovora;
    (c) verovatnoće $P\{A\}$ i $P\{A^c\}$ ako ispit polaže 340 studenata i događaj A je da barem jedan student ima manje od dva tačna odgovora; i
    (d) verovatnoće $P\{B\}$ i $P\{B^c\}$ ako ispit polaže 340 studenata i događaj B je da student ima najviše 90% načnih odgovora.

    (a) Obeležimo broj tačno odgovorenih pitanja X. Tada je $X : B(10, \frac{1}{2})$. Tražimo: $P\{X \geq 7\} = P\{X = 7\} + P\{X = 8\} + P\{X = 9\} + P\{X = 10\} = \binom{10}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3 + \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^1 + \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0 = dbinom(7, 10, 1/2) + dbinom(8, 10, 1/2) + dbinom(9, 10, 1/2) + dbinom(10, 10, 1/2) \approx 0,17$.

    Ovo možemo uraditi i na drugi način:
    $P\{X \geq 7\} = 1 - P\{X \leq 6\} = 1 - F_X(6) = 1 - pbinom(6, 10, 1/2) \approx 0,17$.

    (b) $P\{X > 2\} = 1 - P\{X \leq 2\} = 1 - F_X(2) = 0,943$.

    (c) A^c – svi studenti imaju više od dva tačna odgovora. Sledi: $P\{A^c\} = (0,943)^{340}$, $P\{A\} = 1 - (0,943)^{340}$.

    (d) $P\{X \leq 9\} = 1 - \left(\frac{1}{2}\right)^{10} \approx 0,999$.
    Ako ispit polaže 340 studenata, sledi: $P\{B\} = (0,999)^{340}$, $P\{B^c\} = 1 - (0,999)^{340} \approx 0,28$.

23. Imamo tri kutije i šest identičnih kuglica.

    (a) Ako delimo kuglice tako da se u svakoj kutiji nađe bar jedna, kolika je verovatnoća da u jednoj kutiji budu dve kuglice, u drugoj da budu tri, a u trećoj da bude jedna?
    (b) Kolika je verovatnoća ako dozvolimo da u nekoj kutiji nema kuglica?
    (c) Posmatrajmo sada da imamo 10 kuglica koje su numerisane. Kolika je verovatnoća da u prvoj kutiji budu kuglice od 1 do 4, u drugoj kuglice 5 i 6, a u trećoj kuglice od 7 do 10?

    (a) Ovo se svodi na stavljanje 2 pregrade između 6 kuglica, ali tako da ne sme pregrada da se stavi na kraj, tj. ispred prve i iza poslednje kuglice. Ovakvih mogućnosti ima $\binom{5}{2} = 10$, te je tražena verovatnoća $\frac{1}{10}$.

    (b) Ovo se svodi na raspored 6 kuglica i pregrada, ali tako da pregrade mogu i da se poklope, pa je ukupan broj $\binom{6+3-1}{2} = \binom{8}{2} = 28$, pa je tražena verovatnoća $\frac{1}{28}$.

    (c) Za prvu kutiju ima 4! načina da se raspodele kuglice od 1 do 4. Slično, za drugu kutiju ima 2! i za treću kutiju ima 4!. Međutim, svaki od ovih rasporeda su jednaki jer je kutija *skup* kuglica. Ukupan broj rasporeda 10 kuglica u 3 kutije ima 10!, a nama je potreban broj rasporeda (nizova) koji se dobija kada broj svih rasporeda podelimo brojem onih koje smatrajmo istim, tj. $\dfrac{10!}{4! \cdot 2! \cdot 4!}$. Tražena verovatnoća je recipročna vrednost dobijenog broja.

VI NEDELJA

Teorija> Neka su $X_1, ..., X_n$ slučajne veličine. Tada vektor $(X_1, ..., X_n)$ nazivamo *slučajni vektor*. Funkcija raspodele je sledeća funkcija:

$$F_{X_1,...,X_n}(x_1, ..., x_n) = P\{X_1 \leq x_1, ..., X_n \leq x_n\}.$$

Slučajne veličine $X_1, ..., X_n$ su nezavisne akko za svaku k-torku $X_l, ..., X_{l+k-1}$ važi:

$$F_{X_l,...,X_{l+k-1}}(x_l, ..., x_{l+k-1}) = F_{X_l}(x_l) \cdot ... \cdot F_{X_{l+k-1}}(x_{l+k-1}),$$

gde je $l, ..., l+k-1 \in \{1, ..., n\}$ i svi su međusobno različiti.

S obzirom da obično razmatramo slučaj $n = 2$, onda možemo reći da za slučajne veličine X i Y važi da su nezavisne akko važi da je $F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$. Ako su X i Y apsolutno neprekidne slučajne veličine, onda se prethodna jednakost svodi na jednakost $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$. Pogledajmo sledeći primer:

Primer: Veličine X, Y slučajnog vektora (X, Y) su ravnomerno raspoređene na $[0, 1]^2$.
Dokaz: Neka su X, Y uniformne slučajne veličine na $[0, 1]$ i neka je $X \leq x$ i $Y \leq y$. Tada važi da je $f_{X,Y}(x, y) = \begin{cases} 1, & (x, y) \in [0, 1]^2 \\ 0, & inace. \end{cases}$. Izračunajmo $F_X(x), F_Y(y)$ i $F_{X,Y}(x, y)$ i uporedimo ih:
$F_X(x) = P\{X \leq x\} = \int_0^x \int_0^1 f(x, y)dxdy = \int_0^x \int_0^1 1dxdy = x$, jer je $f(x, y) = 1$ za ravnomernu raspodelu. Slično, važi da je $F_Y(y) = y$. Računamo $F_{X,Y}(x, y)$:
$F_{X,Y}(x, y) = P\{X \leq x, Y \leq y\} = \int_0^x \int_0^y f(x, y)dxdy = xy = F_X(x) \cdot F_Y(y)$,
što je ekvivalentno činjenici da su X, Y nezavisne slučajne veličine.

Pojam uslovne gustine uvodimo na sledeći način:

$$f_X(x|A) = \frac{f_X(x)}{P(A)},$$

gde je x na intervalu na kojem je $P(A) > 0$.

24. Ako je $X \sim U[0, 1]$, izračunati $P\{X > \frac{1}{2} | X < \frac{3}{4}\}$.

I način:

$$\begin{aligned} P\left\{X > \frac{1}{2} \mid X < \frac{3}{4}\right\} &= \int_{\frac{1}{2}}^{\frac{3}{4}} f_X\left(x|X < \frac{3}{4}\right) dx = \\ &= \int_{\frac{1}{2}}^{\frac{3}{4}} \frac{f_X(x)}{P\left\{X < \frac{3}{4}\right\}} dx = \\ &= \int_{\frac{1}{2}}^{\frac{3}{4}} \frac{1}{\frac{3}{4}} dx = \\ &= \frac{1}{4} \cdot \frac{4}{3} = \\ &= \frac{1}{3}. \end{aligned}$$

II način:

$$P\left\{X > \frac{1}{2} | X < \frac{3}{4}\right\} = \frac{P\left\{X > \frac{1}{2}, X < \frac{3}{4}\right\}}{P\left\{X < \frac{3}{4}\right\}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

25. Ako je $X \sim U[0,1]$, izračunati $P\{X > \frac{1}{2} | X^2 - X + \frac{2}{9} < 0\}$.

$$\begin{aligned}
P\{X > \frac{1}{2} | X^2 - X + \frac{2}{9} < 0\} &= \frac{P\{X > \frac{1}{2}, X^2 - X + \frac{2}{9} < 0\}}{P\{X^2 - X + \frac{2}{9} < 0\}} = \\
&= \frac{P\{X > \frac{1}{2} | (X - \frac{2}{3})(X - \frac{1}{3}) < 0\}}{P\{(X - \frac{2}{3})(X - \frac{1}{3}) < 0\}} = \\
&= \frac{P\{X > \frac{1}{2} | \frac{1}{3} < X < \frac{2}{3}\}}{P\{\frac{1}{3} < X < \frac{2}{3}\}} = \\
&= \frac{\frac{1}{6}}{\frac{1}{3}} = \\
&= \frac{1}{2}.
\end{aligned}$$

`Teorija>` **Eksponencijalna raspodela**

Ako neprekidna slučajna veličina X ima *eksponencijalnu raspodelu* sa parametrom λ, onda je njena funkcija gustine

$$f_X(x) = \lambda e^{-\lambda x}, \ \lambda \geq 0.$$

Ovu činjenicu obeležavamo $X : \varepsilon(\lambda)$.

26. Ako je $X : \varepsilon(\lambda)$, izračunati $P\{X > 10 | X > 1\}$.

$$\begin{aligned}
P\{X > 10 | X > 1\} &= \frac{P\{X > 10, X > 1\}}{P\{X > 1\}} = \\
&= \frac{P\{X > 10\}}{P\{X > 1\}} = \\
&= \frac{\int\limits_{10}^{\infty} \lambda e^{-\lambda x} dx}{\int\limits_{1}^{\infty} \lambda e^{-\lambda x} dx} = \\
&= \frac{e^{-10\lambda}}{e^{-\lambda}} = \\
&= e^{-9\lambda}.
\end{aligned}$$

27. Pretpostavimo da se veličine X i Y biraju slučajno i nezavisno iz intervala $[0,1]$. Odrediti verovatnoće sledećih događaja:

(a) $P\{max(X,Y) < \frac{1}{2} | X + Y \in (0,1)\}$, i

(b) $P\{X^2 + Y^2 < \frac{1}{4} | 0 < X + Y < 1\}$.

(a) Prvo izračunamo:

$$P\left\{max(X,Y) < \frac{1}{2}\right\} = \frac{1}{4},$$

$$P\left\{0 < X + Y < 1\right\} = P\left\{-X < Y < -X + 1\right\} = \frac{1}{2}, \ i$$

$$P\left\{max(X,Y) < \frac{1}{2}, 0 < X+Y < 1\right\} = \frac{1}{4}.$$

Zatim računamo:

$$P\left\{max(X,Y) < \frac{1}{2}\middle| X+Y \in (0,1)\right\} = \frac{P\{max(X,Y) < \frac{1}{2}, X+Y \in (0,1)\}}{P\{X+Y \in (0,1)\}} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

(b) Slično prethodnom primeru:

$$P\left\{X^2+Y^2 < \frac{1}{4}\middle| 0 < X+Y < 1\right\} = \frac{P\left\{X^2+Y^2 < \left(\frac{1}{2}\right)^2, -X < Y < -X+1\right\}}{P\{-X < Y < -X+1\}} = \frac{\frac{1}{4}\cdot\frac{1}{4}\pi}{\frac{1}{2}} = \frac{\pi}{8}.$$

28. Ako je $X\sim[0,\infty)$ i $X : \varepsilon(1)$, izračunati uslovnu gustinu $f_X(x|1 < x < 10)$.

$$\begin{aligned} f_X(x|1<x<10) &= \frac{f_X(x)}{P\{1<x<10\}} = \\ &= \frac{e^{-x}}{\int\limits_1^{10} e^{-x}dx} = \\ &= \frac{e^{-x}}{e^{-1}-e^{-10}} \approx \\ &\approx 2,71e^{-x},\ 1<x<10. \end{aligned}$$

29. Neka su $X,Y\sim U[0,1]$. Dati su sledeći događaji: $A = \{X > \frac{1}{3}\}, B = \{Y > \frac{2}{3}\}, C = \{X > Y\}$ i $D = \{X+Y < 1\}$. Odrediti koji od događaja A,B,C i D su nezavisni.

Prvo treba izračunati verovatnoće datih događaja. Neposredno se dolazi do sledećeg: $P\{A\} = \frac{2}{3}, P\{B\} = \frac{1}{3}, P\{C\} = \frac{1}{2}$ i $P\{D\} = \frac{1}{2}$. Računamo verovatnoće svih preseka (sa $\Pi(F)$ je označena površina figure F):

AB: $P\{AB\} = \Pi([\frac{1}{3},1]\times[\frac{2}{3},1]) = \frac{1}{3}\cdot\frac{2}{3} = P\{A\}\cdot P\{B\}$.
Sledi: A,B su nezavisne.

AC: $P\{AC\} = \Pi(T_{(0,\frac{1}{3}),(1,0),(1,1),(\frac{1}{3},\frac{1}{3})}) = \frac{\frac{1}{3}+1}{2}\cdot\frac{2}{3} \neq P\{A\}\cdot P\{C\}$, gde je $T_{(0,\frac{1}{3}),(0,1),(1,1),(\frac{1}{3},\frac{1}{3})}$ trapez sa temenima $(0,\frac{1}{3}),(0,1),(1,1),(\frac{1}{3},\frac{1}{3})$.
Sledi: A,C nisu nezavisne.

AD: $P\{AD\} = \Pi(\triangle_{(0,\frac{1}{3}),(1,0),(\frac{1}{3},\frac{2}{3})}) = \frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{2} \neq P\{A\}\cdot P\{D\}$, gde je $\triangle_{(0,\frac{1}{3}),(0,1),(\frac{1}{3},\frac{2}{3})}$ trougao sa temenima $(0,\frac{1}{3}),(0,1),(\frac{1}{3},\frac{2}{3})$.
Sledi: A,D nisu nezavisne.

BC: $P\{BC\} = \Pi(\triangle_{(\frac{2}{3},\frac{2}{3}),(1,\frac{2}{3}),(1,1)}) = \frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{2} \neq P\{B\}\cdot P\{C\}$.
Sledi: B,C nisu nezavisne.

BD: $P\{BD\} = \Pi(\triangle_{(0,\frac{2}{3}),(\frac{1}{3},\frac{2}{3}),(0,1)}) = \frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{2} \neq P\{B\}\cdot P\{D\}$.
Sledi: B,D nisu nezavisne.

CD: $P\{CD\} = \Pi(\triangle_{(0,0),(1,0),(\frac{1}{2},\frac{1}{2})}) = \frac{1}{4} = P\{C\}\cdot P\{D\}$.
Sledi: C,D su nezavisne.

VII NEDELJA

`Teorija>` **Matematičko očekivanje i disperzija**

Neka slučajna veličina X ima funkciju raspodele $F(x)$ i gustinu raspodele $f(x)$. Definišemo matematičko očekivanje i disperziju slučajne veličine na sledeći način:

$$Eg(x) = \int_{-\infty}^{+\infty} g(x)f(x)dx,$$

$$Dg(x) = E(g(x) - Eg(x))^2.$$

Napomenimo da važi: $Dg(x) = E(g(x))^2 - (Eg(x))^2$.

30. Neka slučajna veličina X ima pomerenu eksponencijalnu raspodelu za parametar ν. Izračunati matematičko očekivanje EX i disperziju DX.

Ako slučajna veličina X ima pomerenu eksponencijalnu raspodelu za parametar ν, onda je $f(x) = \lambda e^{-\lambda(x-\nu)}$, $x \geq \nu$. Tada je:

$$\begin{aligned} EX &= \int_{\nu}^{\infty} x\lambda e^{-\lambda(x-\nu)}dx = \\ &\langle x-\nu = y \Longrightarrow dx = dy \rangle \\ &= \int_0^{\infty} (\nu + y)e^{-\lambda y}\lambda dy = \\ &= \underbrace{\nu \int_0^{\infty} \underbrace{\lambda e^{-\lambda y}}_{\varepsilon(\lambda)} dy}_{1} + \int_0^{\infty} \underbrace{ye^{-\lambda y}}_{f(y)} \lambda dy = \\ &= \nu + \frac{1}{\lambda}. \end{aligned}$$

Ostaje da se objasni poslednja jednakost:

$$\begin{aligned}
\int_0^\infty ye^{-\lambda y}\lambda dy &= \int_0^\infty y\lambda e^{-\lambda y} dy = \\
&\langle z = \lambda y \Longrightarrow dz = \lambda dy \rangle \\
&= \int_0^\infty \frac{z}{\lambda} e^{-z} dz = \\
&= \frac{1}{\lambda} \int_0^\infty ze^{-z} dz = \\
&\begin{bmatrix} u = z & \Longrightarrow & du = dz \\ dv = e^{-z}dz & \Longrightarrow & v = -e^{-z} \end{bmatrix} \\
&= \frac{1}{\lambda}\left(\underbrace{-ze^{-z}\big|_0^\infty}_{0} + \underbrace{\int_0^\infty e^{-z}dz}_{1} \right) = \\
&= \frac{1}{\lambda}.
\end{aligned}$$

Računamo disperziju:

$$\begin{aligned}
DX &= E(X - EX) = \\
&= \int_\nu^\infty \left(x - \nu - \frac{1}{\lambda} \right)^2 \lambda e^{-\lambda(x-\nu)} dx = \\
&\langle x - \nu = y \Longrightarrow dx = dy \rangle \\
&= \int_0^\infty \left(y - \frac{1}{\lambda} \right)^2 \lambda e^{-\lambda y} dy = \\
&= Dy \\
&= \frac{1}{\lambda^2}.
\end{aligned}$$

Ostaje da se objasni poslednja jednakost. Ako $y : \varepsilon(\lambda)$, onda:

$$\begin{aligned}
Dy &= Ey^2 - (Ey)^2 = \\
&= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^2},
\end{aligned}$$

na osnovu:

$$
\begin{aligned}
Ey^2 &= \int\limits_0^\infty y^2 \lambda e^{-\lambda y} dy = \\
&\langle \lambda y = z \Longrightarrow \lambda dy = dz \rangle \\
&= \int\limits_0^\infty \frac{z^2}{\lambda^2} e^{-z} dz = \\
&= \frac{1}{\lambda^2} \int\limits_0^\infty z^2 e^{-z} dz = \\
&\begin{bmatrix} u = z^2 & \Longrightarrow & du = 2zdz \\ dv = e^{-z}dz & \Longrightarrow & v = -e^{-z} \end{bmatrix} \\
&= \frac{1}{\lambda^2} \left(-e^{-z} \cdot z^2 \Big|_0^\infty + 2 \int\limits_0^\infty z e^{-z} dz \right) = \\
&= \frac{1}{\lambda^2} 2 \int\limits_0^\infty z e^{-z} dz = \\
&= \frac{2}{\lambda^2}.
\end{aligned}
$$

31. Pretpostavimo da za slučajnu veličinu X znamo da ima funkciju gustine raspodele oblika $f(x) = cx^5$, za $x \in [0, 2]$.

(a) Odrediti konstantu c tako da $f(x)$ zaista bude funkcija gustine raspodele slučajne veličine X;

(b) Odrediti EX i DX.

(a) Treba da budu zadovoljeni sledeći uslovi:

- nenegativnost $\Longrightarrow$ c mora biti pozitivno
- $\int\limits_0^2 cx^5 dx = 1$:

$$
\begin{aligned}
\int\limits_0^2 cx^5 dx = 1 &\Longleftrightarrow c \cdot \frac{1}{6} x^6 \Big|_0^2 = 1 \\
&\Longleftrightarrow \frac{c}{6} \cdot 64 = 1 \\
&\Longleftrightarrow c = \frac{3}{32} > 0 \surd
\end{aligned}
$$

(b) Računamo EX:

$$
\begin{aligned}
EX &= \int\limits_0^2 x c x^5 dx = \\
&= \frac{c}{7} x^7 \Big|_0^2 = \\
&= \frac{c}{7} 2^7 \\
&= \frac{12}{7}.
\end{aligned}
$$

Slično, računamo EX^2:

$$EX^2 = \int_0^2 x^2 c x^5 dx =$$
$$= 3.$$

Konačno, računamo DX:

$$DX = 3 - \frac{144}{49} =$$
$$= \frac{3}{49}.$$

`Teorija>` Izračunajmo EX i DX slučajne veličine $X{\sim}U[a,b]$. Znamo da je $f(x) = \begin{cases} \frac{1}{b-a} & , \quad x \in [a,b] \\ 0 & , \quad inace. \end{cases}$

Računamo EX i EX^2:

$$EX = \int_a^b \frac{x}{b-a} dx =$$
$$= \frac{b+a}{2}.$$
$$EX^2 = \int_a^b \frac{x^2}{b-a} dx =$$
$$= \frac{b^2+ba+a^2}{3}.$$

Računamo DX:

$$DX = \frac{b^2+ba+a^2}{3} - \frac{b^2+2ba+a^2}{4} =$$
$$= \frac{(b-a)^2}{12}.$$

32. Naći EX, DX ako je funkcija gustine $f_X(x) = \frac{3}{8}(x+1)^2,\ |x| < 1.$

Kako je $|x| < 1 \Longleftrightarrow x \in (-1, 1)$, to je:

$$\begin{aligned} EX &= \frac{3}{8} \int\limits_{-1}^{1} (x+1)^2 \cdot x dx = \\ &= \frac{3}{8} \int\limits_{-1}^{1} (x^3 + 2x^2 + x) dx = \\ &= \frac{1}{2}. \end{aligned}$$

$$\begin{aligned} EX^2 &= \int\limits_{-1}^{1} x^2 \cdot \frac{3}{8} (x+1)^2 dx = \\ &= \frac{3}{8} \int\limits_{-1}^{1} (x^4 + 2x^3 + x^2) dx = \\ &= \frac{2}{5}. \end{aligned}$$

$$\begin{aligned} DX &= EX^2 - (EX)^2 = \\ &= \frac{2}{5} - \frac{1}{4} = \\ &= \frac{3}{20}. \end{aligned}$$

33. Neka je $X \sim \varepsilon(1)$ i $Y = [X]$. Naći zakon raspodele za Y.

Prvo primetimo da je $f_X(x) = e^{-x}, \ x \geq 0$ i da je Y diskretna slučajna veličina. Takođe, $y = k \Longleftrightarrow x \in [k, k+1)$. Dakle,

$$\begin{aligned} P\{y = k\} &= \int\limits_{k}^{k+1} e^{-x} dx = \\ &= -e^{-x}\Big|_{k}^{k+1} = \\ &= e^{-k} - e^{-(k+1)} = \\ &= \frac{1}{e^k} - \frac{1}{e^{k+1}}. \end{aligned}$$

`Teorija>` **Normalna raspodela**

Ako slučajna veličina X ima *normalnu raspodelu*, onda to zapisujemo $X \sim N(m, \sigma^2)$, gde je m tačka simetrije normalne raspodele, a σ^2 je parametar ,,širine" (što je σ manje, to je grafik raspodele uži i viši, a što je veće, to je grafik širi i niži). Funkcija gustine normalne raspodele je

$$f(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \ x \in \mathbb{R}.$$

Napomenimo da integral $\int f(x)dx$ ne možemo da izračunamo. Ipak u $\mathcal{R}$-u možemo simulirati $F(x)$ normalne raspodele i za to koristimo funkciju `pnorm(x, m, `σ`)`, gde `m` možemo dobiti funkcijom `mean()`, a σ funkcijom `sd()`. Podrazumevane vrednosti su $m = 0, \ \sigma = 1$.

Normalna raspodelu ima osobinu da je svaka normalna raspodela linearna transformacija druge normalne raspodele:

$$X \sim N(0,1) \Longrightarrow Y = \sigma X + m,\ Y \sim N(m, \sigma^2)$$
$$Y \sim N(m, \sigma^2) \Longrightarrow X = \frac{Y - m}{\sigma},\ X \sim N(0,1)$$

Takođe, za $X \sim N(0,1)$ funkcija gustine obeležava se $\varphi(x)$ i iznosi $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},\ x \in \mathbb{R}$, a funkcija raspodele $\Phi(x)$ i iznosi $\Phi(x) = \int\limits_{-\infty}^{x} \varphi(t)dt$. Zbog simetričnosti oko tačke $(0,0)$, važi $\Phi(-x) = 1 - \Phi(x)$.

Aproksimacija binomne raspodele

Ako su slučajne veličine $X_1, X_2, ..., X_n$ jednako raspodeljene i $DX_1 < \infty$, i ako posmatramo njihovu parcijalnu sumu $S_n = X_1 + ... + X_n$, onda važi

$$\frac{S_n - ES_n}{\sqrt{DS_n}} \overset{n \to \infty}{\sim} N(0,1).$$

Specijalno, za $S_n \sim B(n,p)$ važi $\frac{S_n - np}{\sqrt{np(1-p)}} \sim N(0,1)$.

Aproksimaciju radimo ako su ispunjeni sledeći uslovi:

- $n > 30$, i
- $np > 10$.

Za mnogo malo p, aproksimira se *Puasonovom raspodelom* $P(\lambda)$. Aproksimacija je $S_n \sim P(\lambda)$, za $\lambda = np$. U $\mathcal{R}$-u nam je na raspolaganju funkcija Puasonove raspodele `ppois(n, λ)`. Uslovi za ovu aproksimaciju su:

- $n > 10$, i
- $np < 10$.

34. Posmatrajmo eksperiment bacanja lopte u koš. Neka imamo ukupno 300 bacanja i neka je verovatnoća pogotka jednaka $\frac{1}{4}$. Izračunaj verovatnoću da je bilo najviše 100 pogodaka.

$$P\{S_n \leq 100\} \approx P\{S_n \leq 100.5\} = P\left\{ \underbrace{\frac{S_n - 75}{\sqrt{75 \cdot \frac{3}{4}}}}_{N(0,1)} \leq \frac{100.5 - 75}{\sqrt{75 \cdot \frac{3}{4}}} \right\} =$$
$$= \Phi\left(\frac{100.5 - 75}{\sqrt{75 \cdot \frac{3}{4}}} \right) = \Phi\left(\frac{50}{3}\right) =$$
$$= \texttt{pnorm}(3.4) = 0.999518.$$

35. Matematičar igra fudbal i veruje da je njegova verovatnoća da pogodi gol jednaka 0.02. Ako šutira ukupno 30 puta na gol, proceniti verovatnoću da je broj pogodaka najmanje 3, a najviše 15.

U ovom zadatku aproksimiraćemo S_n Puasonovom raspodelom jer:

- $n = 30 > 10$, i
- $np = 30 \cdot 0.02 = 0.6 < 10$.

Dakle, $S_n \sim P(0.6)$. Tada je:

$$\begin{aligned}P\{3 \leq S_n \leq 15\} &= P\{3 \leq X \leq 15\},\ X \sim P(0.6) = \\ &= P\{X \leq 15\} - P\{X \leq 2\} = \\ &= \texttt{ppois}(15, 0.6) - \texttt{ppois}(3, 0.6) = \\ &= 0.003358069.\end{aligned}$$

36. Novčić se baca 100 puta. Odrediti verovatnoću da je broj glava između 40 i 60.

Kako je $S_n \sim B(100, \frac{1}{2})$ i $np = 50$, $\sqrt{np(1-p)} = 5$, to je $\dfrac{S_n - 50}{5} \sim N(0, 1)$. Tada je:

$$\begin{aligned}P\{40 \leq S_n \leq 60\} &\approx P\{39.5 \leq S_n \leq 60.5\} = \\ &= P\left\{\frac{39.5 - 50}{5} \leq \frac{S_n - 50}{5} \leq \frac{60.5 - 50}{5}\right\} = \\ &= \Phi(2,1) - \Phi(-2,1) = \\ &= 2\Phi(2,1) - 1 = \\ &= 2 * \texttt{pnorm}(2.1) - 1 = \\ &= 0,96.\end{aligned}$$

37. Na stolu se nalaze dva novčića – jedan homogen i jedan nehomogen (verovatnoća da padne pismo je $\frac{1}{4}$). Slučajno se bira jedan novčić i baca dva puta. Ako su pala dva pisma, odrediti verovatnoću da je odabran nehomogen novčić.

$$\begin{aligned}P\{\textit{odabran je nehomogen} \mid \textit{pala su 2 pisma}\} &= \frac{P\{\textit{odabran je nehomogen, pala su 2 pisma}\}}{P\{\textit{pala su 2 pisma}\}} = \\ &= \frac{\frac{1}{2}\left(\frac{1}{4}\right)^2}{\frac{1}{2}\left(\frac{1}{4}\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)^2} = \\ &= \frac{\frac{1}{2} \cdot \frac{1}{16}}{\frac{1}{32} + \frac{1}{8}} = \\ &= \frac{1}{5}.\end{aligned}$$

38. U kutiji se nalazi n kuglica nepoznatih boja. Ubacujemo $m > 0$ belih kuglica. Sve pretpostavke o prvobitnom broju kuglica su jednako verovatne. Ako se iz kutije izvlači jedna kuglica, odrediti verovatnoću da je kuglica baš bele boje.

Neka je X slučajna veličina koja opisuje broj belih kuglica na početku, a Y ona koja opisuje broj nakon ubacivanja dodatnih kuglica. Zakon raspodele slučajne veličine X je $\begin{pmatrix} 0 & 1 & \ldots & n \\ \frac{1}{n+1} & \frac{1}{n+1} & \ldots & \frac{1}{n+1} \end{pmatrix}$, a slučajne veličine Y je $\begin{pmatrix} m & m+1 & \ldots & m+n \\ \frac{1}{n+1} & \frac{1}{n+1} & \ldots & \frac{1}{n+1} \end{pmatrix}$. Računamo verovatnoću da je izvučena bela kuglica:

$$
\begin{aligned}
P\{izvučena\ je\ bela\ kuglica\} &= \sum_{k=m}^{m+n} P\{izvučena\ je\ bela\ kuglica \mid Y=k\} \cdot P\{Y=k\} = \\
&= \sum_{k=m}^{n+m} \frac{k}{n+m} \cdot P\{Y=k\} = \\
&= \frac{1}{(n+1)(n+m)} \left(\sum_{k=1}^{n+m} k - \sum_{k=1}^{m-1} k \right) = \\
&= \frac{1}{(n+1)(n+m)} \cdot \left(\frac{(m+n)(m+n+1)}{2} - \frac{(m-1)\cdot m}{2} \right).
\end{aligned}
$$

39. Dva strelca gađaju cilj nezavisno jedan od drugog sa verovatnoćama uspeha 0.5 i 0.3.

(a) Slučajno se bira jedan strelac i beleži njegov broj pogodaka. Ako je cilj pogođen dva puta, odrediti verovatnoću da je odabran prvi strelac.

(b) Beleži se broj pogodaka cilja. Ako je cilj pogođen dva puta, odrediti verovatnoću da ga je pogodio prvi strelac.

(a) Tražena verovatnoća je:

$$
\begin{aligned}
P\{odabran\ prvi\ strelac \mid bila\ su\ dva\ pogotka\} &= \frac{P\{odabran\ prvi\ strelac,\ bila\ su\ dva\ pogotka\}}{P\{bila\ su\ dva\ pogotka\}} = \\
&= \frac{\frac{1}{2}\cdot(0.5)^2}{\frac{1}{2}(0.5)^2 + \frac{1}{2}(0.3)^2} = \\
&= \frac{25}{34}.
\end{aligned}
$$

(b) Da bismo izračunali traženu verovatnoću, treba da izračunamo verovatnoću da je cilj pogođen dva puta. Gledamo sve moguće slučajeve u odnosu na to koliko puta je koji strelac gađao cilj:

Prvi, dva puta	Drugi, dva puta	Prvi, jedanput i drugi, jedanput
$(0.5)^2\cdot(0.7)^2$	$(0.5)^2\cdot(0.3)^2$	$\binom{2}{1}(0.5)(0.5)\cdot\binom{2}{1}(0.3)(0.7)$

Neka je $p_1 = (0.5)^2\cdot(0.7)^2 + (0.5)^2\cdot(0.3)^2 + \binom{2}{1}(0.5)(0.5) * \binom{2}{1}(0.3)(0.7) = 0.355$. Tražena verovatnoća je:

$$
\begin{aligned}
P\{pogodio\ je\ prvi \mid p_1\} &= \frac{P\{pogodio\ je\ prvi,\ p_1\}}{p_1} = \\
&= \frac{(0.5)^2\cdot(0.7)^2}{p_1} = \\
&= \frac{0.1225}{0.355} = \\
&= 0.3450704.
\end{aligned}
$$

7.2 Statistika

VIII NEDELJA

Teorija> **Osnovni pojmovi**

U statistici, posmatramo nekakav skup objekata, na primer, skup studenata koji nazivamo **populacijom**. Posmatrajući populaciju, gledamo neku njenu osobinu, na primer, broj godina studija koju nazivamo **obeležje** kojeg obeležavamo X. U verovatnoći, obeležje smo nazivali *slučajnom veličinom*. Problem je taj da veoma često ne možemo u potpunosti da sagledamo celu populaciju. Obeležje kod različitih elemenata populacije može imati iste ili različite vrednosti. Zbog toga posmatramo **uzorak**, tj. podskup populacije. Ideja uzorka je da izmerimo vrednosti slučajne veličine na delu populacije; dobijena raspodela biće polazna tačka za određivanje raspodele populacije. Dakle, uzorak je n-torka slučajnih vrednosti $(X_1, X_2, ..., X_n)$, a realizacija uzorka je n-torka vrednosti $(x_1, x_2, ..., x_n)$. Ako su elementi uzorka nezavisni i sa istom raspodelom, onda govorimo o **prostoj** slučajnoj veličini.

Osobine uzorka:

- Uzoračka sredina:
$$\overline{x_n}$$
- Uzoračka disperzija:
$$\overline{S_n}^2 = \frac{1}{n}\sum_{k=0}^{n}(x_k - \overline{x_n})^2.$$
Ako $X : N(m, \sigma^2)$, onda $\overline{S_n}^2 \approx \sigma^2$.
Ipak, u praksi se često koristi popravljena uzoračka disperzija, koja iznosi
$$\widetilde{S_n}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x_n})^2.$$
- Standardno uzoračko odstupanje:
$$\sigma_n = \sqrt{\widetilde{S_n}^2}.$$
- Statistike poretka:
$$\begin{aligned} x_{(1)} &= min\{x_1, ..., x_n\}, \\ &\vdots \\ x_{(i)} &= i - ti\ po\ redu, \\ &\vdots \\ x_{(n)} &= max\{x_1, ..., x_n\}. \end{aligned}$$
- Uzoračka medijana: m_e. Ako je $X_1, X_2, ..., X_n$ uzorak, onda je
$$m_e = \frac{x_{(k)} + x_{(k+1)}}{2},\ za\ n = 2k\ \vee\ m_e = x_{(k+1)},\ za\ n = 2k + 1.$$
- Uzorački kvartili: q_1 i q_3. Ako je $X_1, X_2, ..., X_n$ uzorak, onda:
  > Ako je $n = 2k$, onda q_1 je uzoračka medijana niza $x_{(1)}, x_{(2)}, ..., x_{(k)}$ i q_3 je uzoračka medijana niza $x_{(k+1)}, x_{(k+2)}, ..., x_{(2k)}$, a

> Ako je $n = 2k+1$, onda q_1 je uzoračka medijana niza $x_{(1)}, x_{(2)}, ..., x_{(k+1)}$ i q_3 je uzoračka medijana niza $x_{(k+1)}, x_{(k+2)}, ..., x_{(2k+1)}$.

Grafičko prikazivanje statističkih podataka

Ukoliko želimo da podatke delimo po kategorijama, onda nam je *histogram* pogodan za prikaz. Na primer:

katerogije (granice)	a_0		a_1		a_2	...	a_k		a_{k+1}
broj podataka		n_1		n_2		...		n_{k+1}	

Postoje raličiti histogrami. Označimo sa h_i visinu stuba histograma koji prebrojava elemente između granica a_{i-1} i a_i (taj broj jednak je n_i). Ako je $N = n_1 + n_2 + ... + n_{k+1}$ i

- $h_i = n_i$, onda govorimo o histogramu *apsolutnih veličina*,
- $h_i = n_i/N$, onda govorimo o histogramu *relativnih veličina*.

Ako je $d_i = a_i - a_{i-1}$ rastojanje između granica i $h_i = n_i/(N \cdot d_i)$, onda govorimo o histogramu *gustine*.

Za histogram je bitno odabrati pogodan broj kategorija. Pratimo sledeća dva pravila za biranje k:

- k nije manje od 5, ili
- $k = [\log_2 N] + 1$.

Rastojanja d_i mogu (ali i ne moraju) da budu jednaka za svako i. Ako su sva rastojanja ista, onda $d = d_i$ možemo računati

$$d \approx R/N,$$

gde je $R = x_{(n)} - x_{(1)}$ **uzorački raspon**.

1. Tvrdi se da je prosečna minimalna cena bezolovnog benzina u Americi bila 1.35$. U reklamne svrhe kompanija želi da pokaže kako je njihova cena niža. Da bi potkrepili svoju tvrdnju, statističari iz firme su sakupili sledeće podatke na osnovu slučajnog uzorka:

1.22 1.37 1.27 1.20 1.42 1.41 1.22 1.24
1.28 1.42 1.48 1.32 1.40 1.26 1.39 1.45
1.44 1.49 1.47 1.47 1.24 1.34 1.27 1.35
1.34 1.45 1.49 1.45 1.23 1.20 1.42 1.34
1.43 1.21 1.49 1.36 1.24 1.20 1.45
1.23 1.25 1.24 1.35 1.23 1.39 1.38
1.46 1.48 1.26 1.36 1.22 1.46 1.39
1.22 1.29 1.47 1.24 1.35 1.21 1.21

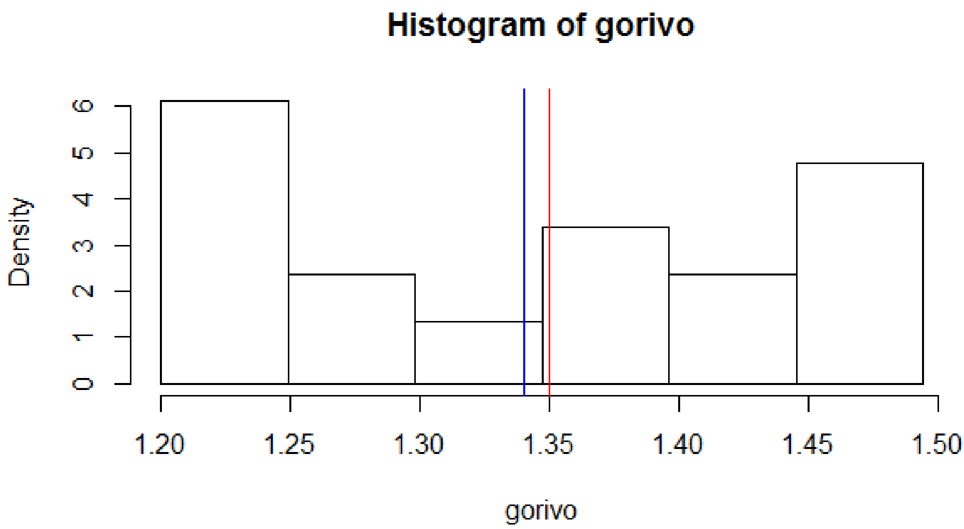
Napisati program u $\mathcal{R}$-u koji računa uzoračku sredinu i medijanu, ispituje da li je raspodela pomerena, računa kvantile, uzorački raspon, i iscrtava histogram nad zadatim podacima.

```
#Pravimo vektor cena
gorivo = c(1.22, 1.37, 1.27, 1.20, 1.42, 1.41, 1.22, 1.24,
           1.28, 1.42, 1.48, 1.32, 1.40, 1.26, 1.39, 1.45,
           1.44, 1.49, 1.47, 1.47, 1.24, 1.34, 1.27, 1.35,
           1.34, 1.45, 1.49, 1.45, 1.23, 1.20, 1.42, 1.34,
           1.43, 1.21, 1.49, 1.36, 1.24, 1.20, 1.45,
           1.23, 1.25, 1.24, 1.35, 1.23, 1.39, 1.38,
           1.46, 1.48, 1.26, 1.36, 1.22, 1.46, 1.39,
           1.22, 1.29, 1.47, 1.24, 1.35, 1.21, 1.21)
```

```
#Uzoracka sredina
mean(gorivo)
duzina = length(gorivo)

#Medijana
#1. Po definiciji:
gorivoSortirano = sort(gorivo)
medijanaG = (gorivoSortirano[duzina/2]+gorivoSortirano[duzina/2 + 1])/2

#2. Pomocu gotove formule
medijanaG2 = median(gorivo)

#Uzoracka disperzija (popravljena)
var(gorivo)

#Standardno odstupanje
sqrt(var(gorivo))

#Primetimo da velicina
mean(gorivo) - medijanaG2
#upada u interval (-stand.odstupanje, +stand.odstupanje)

#Kvartili

#1. Po definiciji
prviKvartil = (gorivoSortirano[duzina/4]+gorivoSortirano[duzina/4+1])/2
#treciKvartil = (gorivoSortirano[3*duzina/4]+gorivoSortirano[3*duzina/4 + 1])/2
treciKvartil = median(gorivoSortirano[duzina+1:duzina])

#2. Pomocu funkcije koja koristi razlicite algoritme
#Ako hocemo da koristimo algoritam sa casa, dodajemo type=1
quantile(gorivo, type=1)

#Minimum
min = gorivoSortirano[1]
#Maximum
max = gorivoSortirano[duzina]

#Uzoracki raspon
R = max - min

#Broj kategorija
#Treba nam 6 kategorija, jer je 2^5 = 32, pa gledamo 5 + 1 = k.
#Ovo mozemo dobiti koriscenjem formule
k = ceiling(log2(duzina)) + 1
d = R/k

#Posto se za d=R/k dobija 0.04833333, mozemo za d da uzmemo
d = 0.049

#Granice za histogram
granice = min + 0:k * d
#Paziti da granice pokriju sve elemente!
```

```
#Crtanje histograma. Postoje razliciti pozivi funkcije za iscrtavanje histograma

hist(gorivo)
#Broj kategorija koje R racuna se cesto ne poklapa sa nasim brojem

hist(gorivo, breaks = granice)
#Breaks predajemo niz podeoka ili broj kategorija, s tim da:
#Ako mu predamo niz, on usvoji nase granice, a
#ako mu predamo broj kategorija, on nekad usvoji, a nekad ga ignorise.
#Najbolje je slati niz podeoka.

hist(gorivo, breaks = granice, plot = FALSE)
#Ako zelimo samo da dobijemo podatke, bez crtanja grafika

hist(gorivo, breaks = granice, freq = FALSE)
#Histogram gustine

#Dodavanje medijane i uzoracke sredine na grafik
abline(v = median(gorivo), col = "red") #v je vertikalno, col je boja
abline(v = mean(gorivo), col = "blue")
```

Rezultat rada programa:



Teorija> Osim putem histograma, podatke možemo prikazivati i tzv. *stablo–lišće dijagramom*. Recimo, ako imamo podatke na dve decimale, onda su grane prve dve cifre, a treća cifra predstavlja lišće. Primer:

Za vektor $\langle 1.22, 1.37, 1.27, 1.20, 1.42, 1.41, 1.22\rangle$, stablo–lisće dijagram izgleda:

12 | 2024
13 | 7
14 | 21.

Ukoliko na nekoj grani ima previše lišća, onda možemo deliti grane, na primer, na pola da bismo imali bolji prikaz.

2. Napisati program u $\mathcal{R}$-u koji iscrtava stablo–lišće dijagram za vektor `gorivo` iz prethodnog zadatka. Zatim razrediti podatke skaliranjem grana na 0.2 i 0.5 i uporediti dobijene dijagrame.

```
#Standardan poziv
stem(gorivo)

#Mozemo mi sami da kazemo koliko ce se razbijati grane u odnosu na standardan poziv
stem(gorivo, scale=0.5)
stem(gorivo, scale=0.2)
```

`Teorija>` Navedimo još jedan grafički prikaz podataka, koji se naziva *box–plot dijagram*. Korišćenjem ovog dijagrama se mogu pronaći *autlajeri* (najverovatnije pogrešni podaci, tj. podaci koje treba proveriti). Da bismo našli autlajere, treba naći niz podataka:

$$\langle q_1,\ q_3,\ IQR,\ f_1,\ f_3,\ F_1,\ F_3\rangle\,,$$

gde su:

- q_1, q_3 uzorački kvartili,
- $IQR = q_3 - q_1$ interkvartijalno rastojanje,
- $f_1 = q_1 - \frac{3}{2}\cdot IQR$, $F_1 = q_1 - 3\cdot IQR$ i $f_3 = q_3 + \frac{3}{2}\cdot IQR$, $F_3 = q_3 + 3\cdot IQR$ takve tačke za koje važi sledeće:
  - $>$ podaci koji se nalaze u nekom od intervala (F_1, f_1) ili (f_3, F_3) se nazivaju *blagi autlajeri* i oni su na dijagramu označeni $\circ$,
  - $>$ podaci koji se nalaze u nekom od intervala $(-\infty, F_1)$ ili $(F_3, +\infty)$ se nazivaju *pravi autlajeri* i oni su na dijagramu označeni $\times$.

Ponekad se traže tačke a_1 i a_3 za koje važi:

- tačka a_1 je prvi element koji je veći od f_1,
- tačka a_3 je prvi element koji je manji od f_3.

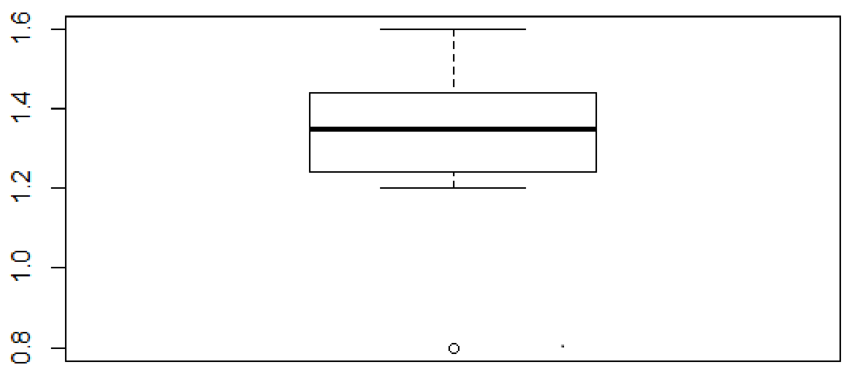
Ove tačke se često koriste umesto f_1 i f_3 pri analizi podataka.

3. Napisati program u $\mathcal{R}$-u koji iscrtava box–plot dijagram za vektor `gorivo` iz prvog zadatka uz dodata dva nova podatka: 1.6 i 0.8. Naći blage i prave autlajere.

```
#Pravimo novi niz sa dodatnim podacima
gorivo1 = c(gorivo, 1.6, 0.8)

#Na osnovu poziva
quantile(gorivo1, type=1)
#dobijamo podatke o kvantilima q1 i q3
q1 = 1.24
q3 = 1.44
IQR = q3 - q1
f1 = q1 - 3/2 * IQR
F1 = q1 - 3 * IQR
f3 = q3 + 3/2 * IQR
F3 = q3 + 3 * IQR

#Nalazenje tacke a1
```

```
#Potrebni su nam svi elementi koji su veci od tacke f1
gorivo1[gorivo1 >= f1]
sort(gorivo1[gorivo1 >= f1])
a1 = 1.20 #prvi u sortiranom nizu

#Nalazenje tacke a3
#Potrebni su nam svi elementi koji su manji od tacke f3
gorivo1[gorivo1 <= f3]
sort(gorivo1[gorivo1 <= f3])
a3 = 1.60

blagiAutlajeri = gorivo1[((gorivo1 < a1) & (gorivo1 > F1))
| ((gorivo1 > a3) & (gorivo1 < F3))]

#Vidimo da je 0.8 blagi autlajer

praviAutlajeri = gorivo1[gorivo < F1
| gorivo > F3]

boxplot(gorivo1)
```

Rezultat rada programa je box-plot dijagram



kao i vektori `blagiAutlajeri` i `praviAutlajeri`. Vidimo da se slažu podaci sa dijagrama i oni u vektorima, tj. da pravih autlajera nema, a da je tačka 0.8 blagi autlajer.

4. Prilikom proučavanja rasta dece, posmatra se obim glave deteta pri rođenju izraženo u cm. Dobijeni su sledeći podaci:

33.1 33.7 33.7 33.8 33.4
33.9 33.6 33.4 34.1 34.2
34.5 34.2 34.6 34.9 34.8
34.0 34.5 34.2 34.2 34.7
34.7 34.6 34.3 34.3 34.2
35.1 36.0 35.8 35.2 35.6
36.1 35.1 35.3 35.2

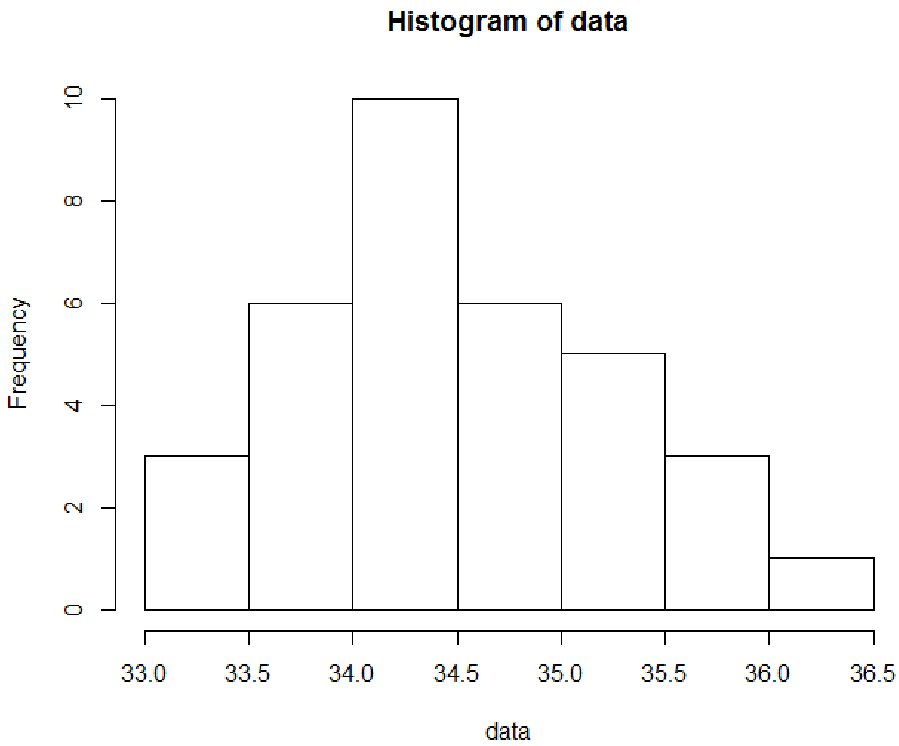
Napisati program u $\mathcal{R}$-u koji iscrtava stablo-lišće dijagram i histogram nad zadatim podacima i proverava postojanje autlajera.

```
data=c(33.1, 33.7, 33.7, 33.8, 33.4,
       33.9, 33.6, 33.4, 34.1, 34.2,
       34.5, 34.2, 34.6, 34.9, 34.8,
       34.0, 34.5, 34.2, 34.2, 34.7,
       34.7, 34.6, 34.3, 34.3, 34.2,
       35.1, 36.0, 35.8, 35.2, 35.6,
       36.1, 35.1, 35.3, 35.2)

stem(data)
hist(data)

dataSort=sort(data)

d=length(data)

q1=median(dataSort[1:d/2])
q3=median(dataSort[((d/2)+1):d])

Iqr=q3-q1
f1=q1-3/2*Iqr
f3=q3+3/2*Iqr
F1=q1-3*Iqr
F3=q3+3*Iqr

a1=min(data[data>=f1])
a3=max(data[data<=f3])

blagiAutlajeri=data[((data>F1) && (data<f1)) | ((data>f3) && (data<F3))]
praviAutlajeri=data[(data<F1) | (data>F3)]
```

Rezultat rada programa:

```
The decimal point is at the |
```

33|144
33|67789
34|012222233
34|55667789
35|11223
35|68
36|01.



Nema autlajera.

IX NEDELJA

`Teorija>` **Metod maksimalne verodostojnosti** za diskretne slučajne veličine

$$L(x,\theta) = P_\theta\{X_1 = x_1, ..., X_n = x_n\} = \prod_{k=1}^{n} P_\theta(X_k = x_k), \tag{6}$$

gde je $\theta = argmax\ L(x,\theta) = \log L(x,\theta)$.

Poslednja jednakost u (6) važi jer je reč o prostom slučajnom uzorku.

5. Neka slučajna veličina X ima Puasonovu raspodelu $P(\lambda)$. Znamo da je

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Tada je:

$$L(x,X) = \prod_{k=1}^{n} \frac{\lambda^{x_k} e^{-k}}{x_k!} = \frac{\lambda^{\sum_{k=1}^{n} x_k} e^{-n\lambda}}{\prod_{k=1}^{n} x_k!},\ i$$

$$l(x,\lambda) = \log L(x,\lambda) = \sum_{k=1}^{n} x_k \log\lambda - n\lambda - \log\prod_{k=1}^{n} x_k!$$

Dalje računamo stacionarne tačke prvog izvoda funkcije l:

$$\frac{\delta l}{\delta\lambda} = \frac{\sum_{k=1}^{n} x_k}{\lambda} - n = 0$$

odakle sledi

$$\hat{\lambda} = \frac{\sum_{k=1}^{n} x_k}{n}.$$

Iz činjenice

$$\frac{\delta^2 l}{\delta\lambda^2} = -\frac{\sum_{k=1}^{n} x_k}{n} < 0$$

sledi da je prvi izvod je pozitivan do tačke $\hat{\lambda}$, a nadalje je negativan, te funkcija l raste do tačke $\hat{\lambda}$, što znači da je $\hat{\lambda}$ tačka lokalnog maksimuma i ona predstavlja ocenu maksimalne verovatnoće.

Računamo $E\hat{\lambda}$ i $D\hat{\lambda}$:

$$\begin{aligned} E\hat{\lambda} &= \frac{E\left(\sum\limits_{k=1}^{n} x_k\right)}{n} = \\ &= \frac{nEX_1}{n} = \\ &= EX_1 = \\ &= \lambda, \end{aligned}$$

$$\begin{aligned} D\hat{\lambda} &= \frac{DX_1 + ... + DX_n}{n^2} = \\ &= \frac{nDX_1}{n_2} = \\ &= \frac{DX_1}{n} = \\ &= \frac{\lambda}{n}. \end{aligned}$$

6. Neka slučajna veličina X ima binomnu $B(N, p)$ raspodelu. Oceniti parametar p metodom maksimalne verodostojnosti.

$$\begin{aligned} L(x,p) &= \prod_{k=1}^{n} P\{X_k = x_k\} = \\ &= \prod_{k=1}^{n} \binom{N}{x_k} p^{x_k}(1-p)^{N-x_k} = \\ &= \prod_{k=1}^{n} \binom{N}{x_k} p^{\sum\limits_{k=1}^{n} x_k}(1-p)^{\sum\limits_{k=1}^{n}(N-x_k)} = \\ &= \prod_{k=1}^{n} \binom{N}{x_k} p^{\sum\limits_{k=1}^{n} x_k}(1-p)^{nN-\sum\limits_{k=1}^{n} x_k}, \end{aligned}$$

$$l(x,p) = \log \prod_{k=1}^{n} \binom{N}{x_k} + \sum_{k=1}^{n} x_k \log p + (nN - \sum_{k=1}^{n} x_k)\log(1-p),$$

$$\begin{aligned} \frac{\delta l}{\delta p} &= \frac{\sum\limits_{k=1}^{n} x_k}{p} = \\ &= \frac{nN - \sum\limits_{k=1}^{n} x_k}{1-p} = \\ &= \frac{\sum\limits_{k=1}^{n} x_k - \cancel{p\sum\limits_{k=1}^{n} x_k} - pnN + \cancel{p\sum\limits_{k=1}^{n} x_k}}{1-p} = \\ &= \frac{\sum\limits_{k=1}^{n} x_k - pnN}{1-p} = 0, \end{aligned}$$

odakle dobijamo ocenu za p

$$\hat{p} = \frac{\sum\limits_{k=1}^{n} x_k}{nN} = \frac{\overline{X}_n}{N}.$$

7. Neka slučajna veličina X ima binomnu $G(p)$ raspodelu. Oceniti parametar p metodom maksimalne verodostojnosti.

   Znamo da je

$$P\{X = k\} = (1-p)^{k-1} \cdot p, \;\; k = 1, 2, ...$$

   Dakle,

$$\begin{aligned} L(x,p) &= \prod_{k=1}^{n} (1-p)^{x_k - 1} p = \\ &= (1-p)^{\sum\limits_{k=1}^{n}(x_k - 1)} p^n = \\ &= (1-p)^{\sum\limits_{k=1}^{n} x_k - n} p^n = \end{aligned}$$

$$l(x,p) = (\sum_{k=1}^{n} x_k - n) \log(1-p) + n \log p,$$

$$\begin{aligned} \frac{\delta l}{\delta p} &= -\frac{\sum\limits_{k=1}^{n} x_k - n}{1-p} + \frac{n}{p} = \\ &= -\frac{p \sum\limits_{k=1}^{n} x_k \cancel{-np} + n \cancel{+np}}{p(1-p)} = 0, \end{aligned}$$

   odakle dobijamo ocenu za p

$$\hat{p} = \frac{n}{\sum\limits_{k=1}^{n} x_k}.$$

8. Neka slučajna veličina X ima Bernulijevu (indikatorsku) $B(1,p)$ raspodelu. Oceniti parametar p metodom maksimalne verodostojnosti.

   Zakon raspodele Bernulijeve raspodele glasi $X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$.

   Prvi način: $P\{X_k = x_k\} = p^{x_k}(1-p)^{1-x_k} \Longrightarrow \hat{p} = \overline{X}_n$.
   Drugi način: $P\{X_k = x_k\} = p^{\mathrm{I}\{x_k=1\}}(1-p)^{\mathrm{I}\{x_k=0\}} = p^{\mathrm{I}\{x_k=1\}}(1-p)^{1-\mathrm{I}\{x_k=1\}}$, gde je I oznaka za indikator. Računamo:

$$\begin{aligned} L(x,p) &= \prod_{k=1}^{n} p^{\mathrm{I}\{x_k=1\}}(1-p)^{1-\mathrm{I}\{x_k=1\}} = \\ &= p^{\sum \mathrm{I}\{x_k=1\}}(1-p)^{n-\sum \mathrm{I}\{x_k=1\}}. \end{aligned}$$

Pre nego što nastavimo, označimo $\sum \mathrm{I}\{x_k = 1\} = A$. Dakle,

$$l(x, p) = A \log p + (n - A)\log(1 - p),$$

$$\begin{aligned}\frac{\delta l}{\delta p} &= \frac{A}{p} - \frac{n - A}{1 - p} = \\ &= \frac{A - \cancel{Ap} - pn + \cancel{PA}}{p(1 - p)} = \\ &= \frac{A - pn}{p(1 - p)} = 0,\end{aligned}$$

odakle se dobija

$$\hat{p} = \frac{\sum \mathrm{I}\{x_k = 1\}}{n} = \overline{X}_n.$$

9. Neka slučajna veličina X ima zakon raspodele $\begin{pmatrix} y_1, & ..., & y_m \\ p_1, & ..., & p_m \end{pmatrix}$. Nad uzorkom $X_1, ..., X_n$, oceniti $p_1, ..., p_m$ metodom maksimalne verodostojnosti.

Koristeći ideju prethodnog zadatka i činjenice

$$p_1 + ... + p_n = 1,$$

dobijamo da je

$$P\{X = x_k\} = p_1^{\mathrm{I}\{x_k = y_1\}} \cdot p_2^{\mathrm{I}\{x_k = y_2\}} \cdot ... \cdot p_m^{\mathrm{I}\{x_k = y_m\}}, \ x_k \in \{y_1, ..., y_m\}$$

Računamo:

$$\begin{aligned}L(x, p) &= \prod_{k=1}^{n} p_1^{\mathrm{I}\{x_k = y_1\}} \cdot ... \cdot p_m^{\mathrm{I}\{x_k = y_m\}} = \\ &= p_1^{\sum\limits_{k=1}^{n} \mathrm{I}\{x_k = y_1\}} \cdot ... \cdot p_m^{\sum\limits_{k=1}^{n} \mathrm{I}\{x_k = y_m\}},\end{aligned}$$

$$l(x, p) = \sum_{k=1}^{n} \mathrm{I}\{x_k = y_1\} \log p_1 + ... + \sum_{k=1}^{n} \mathrm{I}\{x_k = y_m\} \log p_m,$$

na ovom mestu ćemo posmatrati Lagranžovu funkciju G:

$$G(x, p, \lambda) = l(x, p) - \lambda(p_1 + ... + p_m - 1).$$

Vidimo da je

$$\left(\frac{\delta G}{\delta p_i} = 0, \ i = 1, ..., m\right) \wedge \left(\frac{\delta G}{\delta \lambda} = 0\right),$$

pa iz

$$\frac{\delta G}{\delta p_i} = \frac{\sum\limits_{k=1}^{n} \mathrm{I}\{x_k = y_i\}}{p_i} - \lambda = 0$$

sledi

$$p_i = \frac{\sum\limits_{k=1}^{n} \mathrm{I}\{x_k = y_i\}}{\lambda}.$$

Takođe, na osnovu

$$\frac{\delta G}{\delta \lambda} = -(p_1 + ... + p_m - 1) = 0$$

sledi

$$\frac{\sum\limits_{k=1}^{n} |\{x_k = y_1\}}{\lambda} + ... + \frac{\sum\limits_{k=1}^{n} |\{x_k = y_m\}}{\lambda} = 1,$$

tj.

$$\frac{n}{\lambda} = 1 \Longrightarrow \lambda = n.$$

`Teorija>` **Metod maksimalne verodostojnosti** za neprekidne slučajne veličine

Razlika između ovog i prethodnog metoda jeste što se posmatra drugačija funkcija L, koja je sada definisana na sledeći način:

$$L(x, \theta) = f(x_1, ..., x_n, \theta) - \prod_{k=1}^{n} f(x_i, \theta),$$

gde je $x_1, ..., x_n$ realizovana vrednost uzorka $X_1, ..., X_n$.

10. Neka neprekidna slučajna veličina X ima gama $\gamma(\alpha, \beta)$ raspodelu i neka nam je α poznato. Metodom maksimalne verodostojnosti oceniti parametar β.

Funkcija gustine gama $\gamma(\alpha, \beta)$ raspodele glasi

$$f(x, \alpha, \beta) = \frac{x^{\alpha-1}\beta^{\alpha}e^{-\beta x}}{\Gamma(\alpha)},$$

gde je funkcija Γ data sa

$$\Gamma(m) = (m-1)!.$$

Napomenimo prvo da je

$$P(\alpha) = \int\limits_{0}^{\infty} x^{\alpha-1}e^{-x}dx = (\alpha-1)\Gamma(\alpha-1).$$

Računamo:

$$\begin{aligned} L(x, \alpha, \beta) &= \prod_{k=1}^{n} \frac{x_k^{\alpha-1}\beta^{\alpha}e^{-\beta x_k}}{\Gamma(\alpha)} = \\ &= \frac{\left(\prod\limits_{k=1}^{n} x_k\right)^{\alpha-1} \cdot \beta^{n\alpha} \cdot e^{-\beta \sum\limits_{k=1}^{n} x_k}}{(\Gamma(\alpha))^n}, \end{aligned}$$

$$l(x, \alpha, \beta) = (\alpha-1)\log \prod_{k=1}^{n} x_k + n\alpha \log \beta - \beta \sum_{k=1}^{n} x_k - n \log \Gamma(\alpha),$$

$$\frac{\delta l}{\delta \beta} = \frac{n\alpha}{\beta} - \sum_{k=1}^{n} x_k = 0,$$

odakle sledi

$$\hat{\beta} = \frac{n\alpha}{\sum\limits_{k=1}^{n} x_k}.$$

11. Neka neprekidna slučajna veličina X ima uniformnu $U[a,b]$ raspodelu. Oceniti parametre a i b metodom maksimalne verodostojnosti.

Znamo da je funkcija gustine slučajne veličine X

$$f(x,a,b) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & inače \end{cases} = \frac{1}{b-a} \mathbb{I}\{x \in [a,b]\}.$$

Računamo:

$$\begin{aligned} L(x,a,b) &= \prod_{k=1}^{n} \frac{1}{b-a} \mathbb{I}\{x \in [a,b]\} = \\ &= \frac{1}{(b-a)^n} \mathbb{I}\{(\forall k)\ (x_k \geq a,\ x_k \leq b)\} = \\ &= \frac{1}{(b-a)^n} \mathbb{I}\{max\{x_1,...,x_n\} \geq a,\ max\{x_1,...,x_n\} \leq b\}. \end{aligned}$$

Maksimum tražimo za one vrednosti a, b za koje je indikator različit od nule. To se dostiže za najmanje $b-a$ na ovom skupu, tj. za najmanje moguće b, a najveće moguće a tako da cela realizacija uzorka upada u interval $[a,b]$. Dakle, zaključujemo

$$\hat{b} = max\{x_1,...,x_n\},$$
$$\hat{a} = min\{x_1,...,x_n\}.$$

`Teorija>` **Metod maksimalne verodostojnosti** u $\mathcal{R}$-u

Da bismo simulirali ocenu raspodele na osnovu nekog uzorka u $\mathcal{R}$-u, koristićemo funkciju

$$\texttt{fitdistr}(\texttt{x}, \texttt{densfun}, ...),$$

sadržanu u paketu `MASS`, a koja za argumente prihvata:

- `x` – uzorak,
- `densfun` – ime funkcije raspodele ili konkretnu funkciju raspodele. Na raspolaganju su nam `beta, cauchy, chi-squared, exponential, f, gamma, geometric, log-normal, lognormal, logistic, negative binomial, normal, Poisson, t` i `weibull` raspodele.
- ...

Primer: Neka je zadat uzorak

42.01159, 34.78635, 62.04313, 47.42140, 54.62714, 62.12806, 39.54939, 42.96735, 37.35422, 31.75614, 51.30831, 55.48523, 51.07290, 41.24063, 48.54175, 40.26295, 71.12446, 43.97390, 37.96417, 51.58539, 50.18400, 49.89327, 47.27160, 53.77644, 54.69077, 41.95900, 39.91186, 63.69526, 50.91429, 46.59034, 70.67112, 45.22308, 37.07300, 53.04710, 64.39713, 43.10263, 38.28027, 53.97566, 48.69479, 58.42842, 44.43358, 47.25470, 48.56333, 82.63110, 57.26334, 52.05529, 34.21297, 33.89213, 55.71156, 61.26512, 40.85732, 41.49857, 37.14342, 51.40030, 59.33215, 37.39154, 43.83656, 43.03050, 41.32934, 52.49921,

koji je sačuvan u vektoru `uzorak`, a za koji imamo razloga da verujemo da pripada normalnoj raspodeli sa nepoznatim parametrima. Ukoliko u $\mathcal{R}$-u pozovemo funkciju

$$\texttt{fitdistr}(\texttt{uzorak}, "\texttt{normal}"),$$

kao rezultat dobijamo

```
    mean          sd
 48.7431093   10.0778995
( 1.3010512) ( 0.9199821).
```

Kako je gornji uzorak generisan pozivom funkcije

$$\texttt{uzorakN} < -\texttt{rnorm}(60, 50, 10),$$

vidimo da je funkcija `fitdistr()` dobro ocenila parametre normalne raspodele.

12. Napisati funkciju `oceni` u $\mathcal{R}$-u koja za argumente prima uzorak, ime raspodele i niz poznatih parametre. Funkcija `oceni` ima zadatak da simulira rad funkcije `fitdistr` koristeći ocene iz zadataka 5 – 11.

```
#x je niz x1,...,xm za tablicnu raspodelu

oceni<-function(uzorak,raspodela,parametri){
    n=length(uzorak)

    if (raspodela=='Puasonova')
        return(mean(uzorak))

    else if (raspodela=='Geometrijska') {
        if (mean(uzorak)==0)
            return("pogresna pretpostavka o raspodeli")
        else
```

```
            return(1/mean(uzorak))
    }

    else if (raspodela=='Binomna')
        return(mean(uzorak)/parametri[1]) #parametri[1]=N

    else if (raspodela=='Tablicna'){
        y = parametri
        m = length(y)
        frekvencija=rep(0,m)
        for(i in 1:m) {
            frekvencija[i] = length(uzorak[uzorak == y[i] ])/n
        }
        return(frekvencija)
    }

    else if(raspodela == "Gama"){
        alfa = parametri[1]
        return (n*alfa/sum(uzorak))
    }

    else if(raspodela == "Uniformna"){
        aocena = min(uzorak)
        bocena = max(uzorak)
        return (c(aocena, bocena))
    }

    else
        return("nema unete raspodele")
}
```

X NEDELJA

`Teorija>` **Testiranje statističkih hipoteza**

Ukoliko imamo neku hipotezu, nju možemo prihvatiti ili odbaciti. Neka imamo hipotezu H_0 i njenu alternativnu hipotezu H_1. Reći da odbacujemo hipotezu H_0 je isto što i prihvatiti hipotezu H_1. Koristićemo sledeće oznake:

- T – statistika kojom testiramo hipotezu (ima samo one nepoznate parametre koje ocenjujemo),
- W – kritična oblast,
- $\hat{T}$ – realizovana vrednost.

Važi:

$$\hat{T} \in W \Longrightarrow \textit{odbacujemo } H_0.$$

Pri testiranju hipoteze možemo napraviti dve vrste grešaka:

1) greška prve vrste – odbacujemo tačnu H_0 hipotezu, i
2) greška druge vrste – ne odbacujemo hipotezu H_0, ako je tačno H_1,

sa verovatnoćama tih greški:

1) $P\{\textit{greška prve vrste}\} = P_{H_0}\{T_n \in W\} = \alpha$ – nivo značajnosti testa, i
2) $P\{\textit{greška druge vrste}\} = P_{H_1}\{T_n \notin W\} = \beta$.

Takođe, uvodimo i *moć testa*, u oznaci M, koji je jednak $M = 1 - \beta$.

Iako postoje različite vrste testova, mi ćemo se pozabaviti onima u kojim neka raspodela ima nepoznate neke od parametara. Testovi koji testiraju takve raspodele nazivaju se *parametarski.*

Primer: Neka je $X \sim N(m, \sigma^2)$. Pretpostavimo da nam je σ^2 poznato i da testiramo hipotezu $m = m_0$, za $m_0 \in \mathbb{R}$. Alternativna hipoteza H_1 može biti $m \neq m_o$, $m > m_0$ ili $m < m_0$. Ovaj test možemo uraditi na dva načina:

1) Ako važi H_0, onda

$$T_n = \frac{\overline{X}_n - m_0}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1).$$

U zavisnosti od H_1, skup W je neki od sledećih

$$W = \{|T_n| > c\},\ W = \{T_n > c\}\ \vee\ W = \{T_n < c\}.$$

Neka je $H_1 : m \neq m_0$. Tada je

$$\alpha = P_{H_0}\{|T_n| > c\} = P_{H_0}\{T_n > c\} + P_{H_0}\{T_n < -c\},$$

iz čega se dobija

$$\Phi(c) = 1 - \frac{\alpha}{2},$$

tj.

$$c = \Phi^{-1}(1 - \frac{\alpha}{2}) = \texttt{qnorm(1 - alpha/2)}.$$

2) Pomoću p-vrednosti testa. Definišemo p-vrednost testa kao najmanji nivo značajnosti testa za koji ćemo sa datim uzorkom odbaciti H_0 hipotezu. Dakle,

$$p < \alpha \Longrightarrow \textit{odbacujemo } H_0.$$

Napomenimo da je α unapred zadato, tj. dobija se iz sprovedenog statističkog istraživanja, ali je najčešće jednako 0.05, što ćemo i mi koristiti.

U našem konkretnom primeru,

$$P = P_{H_0}\{|T_n| > |\hat{T_n}|\}.$$

Za ovo koristimo funkciju

$$\texttt{z.test}(\texttt{x}, \texttt{y} = \texttt{NULL}, \texttt{alternative} = \texttt{"two.sided"}, \texttt{mu} = 0, \texttt{sigma.x} = \texttt{NULL},$$
$$\texttt{sigma.y} = \texttt{NULL}, \texttt{conf.level} = 0.95),$$

koja se nalazi u paketu `BSDA`.

13. Pretpostavimo da imamo uzorak

[1]	1.51983920	3.74121307	5.56845489	2.82948995	-0.81478350	1.16580998
[7]	4.74196836	4.54240095	1.93830403	1.51566087	6.38791915	-0.03172442
[13]	2.21676188	2.98737103	6.55425403	-3.60479067	4.77348583	6.44761216
[19]	0.47893794	-1.48994008	1.54555880	3.08107501	-5.32621046	3.83494286
[25]	8.25839836	2.25472335	1.15410725	-0.91442943	2.70109672	1.20129951
[31]	9.83914404	4.86904120	2.12230132	1.55712246	2.55188607	-0.90137700
[37]	-1.34612660	2.87061217	4.35577137	3.01561994	-0.32023556	2.30803158
[43]	-1.60016500	8.37069195	-2.77429117	-3.64248338	3.33717942	2.83589193
[49]	0.02129508	5.78776686	-1.28942128	0.88250603	-1.54122125	2.20595511
[55]	6.27081708	1.87644834	3.94336967	1.43024362	5.44544720	-0.65566701,

iz normalne $N(m, 10)$ raspodele zadat u promenljivoj `uzorak`. Testirati koristeći p-vrednost testa u $\mathcal{R}$-u hipotezu da je $m = 2$, protiv alternativne hipoteze da je $m > 2$.

Pozivom funkcije

$$\texttt{z.test}(\texttt{x} = \texttt{uzorak}, \texttt{sigma.x} = \texttt{sqrt}(10), \texttt{mu} = 2, \texttt{alternative} = \texttt{"greater"}),$$

dobijamo rezultat

```
One-sample z-Test

data:  uzorak
z = 0.45254, p-value = 0.3254
alternative hypothesis: true mean is greater than 2
95 percent confidence interval:
1.513241       NA
sample estimates:
mean of x
2.184749
```

Vidimo da je p-vrednost testa jednaka 0.3254, što je svakako veće od $\alpha = 0.05$. Dakle, prihvatamo hipotezu H_0, tj. $m = 2$. Kako je navedeni uzorak generisan funkcijom `uzorak <- rnorm(60, 2, sd=sqrt(10))`, zaista vidimo da je test prihvatio ispravnu hipotezu.

14. Prodavac smatra da, ako je prosečna zarada 2$ po osobi, ostvaruje se dobar profit i da treba otvoriti prodavnicu. Profit ima normalnu $N(m, (1.7)^2)$ raspodelu. Prodavac je uzeo uzorak od 25 slučajno odabranih osoba i dobio prosečnu vrednost profita 2.842$. Ako je $\alpha = 0.05$, da li se na osnovu dobijenog rezultata može zaključiti da treba otvoriti prodavnicu?

   Vidimo da je $H_0 : m = 2$ i $H_1 : m > 2$.
   Ovaj zadatak ćemo uraditi na dva načina:

   1) Tražimo T_n, W i p-vrednost preko formula:

$$T_n = \frac{\overline{X}_n - 2}{\frac{1,7}{5}} = \frac{2,842 - 2}{\frac{1,7}{5}} = 2,470588$$
$$W = \{T_n > c\}$$
$$P\{T_n > \hat{T_n}\} \Longrightarrow \texttt{p\_value} = 1 - \texttt{pnorm}(2.470588) = 0.006744556$$

   Kako važi $\texttt{p\_value} < \alpha$, zaključujemo da odbacujemo hipotezu H_0, tj. treba otvoriti prodavnicu.

   2) Tražimo vrednost c na sledeći način:

$$\begin{aligned} P\{T_n > c\} &= 0.05 \\ &\Longrightarrow \Phi(c) = 0.95 \\ &\Longrightarrow c = \Phi^{-1}(0.95) = \texttt{qnorm}(0.95) = 1.64454. \end{aligned}$$

   Kako važi $T_n > c$ $(2,74 > 1,64)$, zaključujemo da odbacujemo hipotezu H_0, tj. treba otvoriti prodavnicu.

`Teorija>` Šta ako je σ^2 nepoznati parametar? U tom slučaju, za statistiku T uzimamo veličinu

$$V = \frac{\overline{X}_n - m_0}{\sqrt{\widetilde{S}_n^2}}\sqrt{n},$$

koja ima Studentovu t raspodelu sa parametrom $n-1$. To zapisujemo $V \sim t_{n-1}$. Napomenimo da, kako je ova raspodela simetrična oko tačke (0,0), to se za $n > 30$ ona aproksimira normalnom $N(0,1)$ raspodelom. U ovom slučaju, koristićemo funkciju

$$\texttt{t.test}(\texttt{x}, \texttt{mu}, \texttt{alternative}, ...)$$

iz paketa `PairedData`.

15. Posmatrajmo uzorak generisan u zadatku 13. Takođe, ovoga puta ćemo posmatrati da nam nije poznat ni parametar σ^2. Koristeći p-vrednost testa, testirati hipotezu $H_0 : m = 2$ nasuprot hipoteze $H_1 : m \neq 2$.

   Pozivom funkcije `t.test` na sledeći način:

$$\texttt{t.test}(\texttt{x} = \texttt{uzorak}, \texttt{mu} = 2, \texttt{alternative} = \texttt{"two.sided"})$$

   dobijamo rezultat

```
One Sample t-test

data:  uzorak
t = 0.46389, df = 59, p-value = 0.6444
alternative hypothesis: true mean is not equal to 2
95 percent confidence interval:
1.387836 2.981663
sample estimates:
mean of x
2.184749
```

Kako je $p > \alpha$, zaključujemo da prihvatamo hipotezu H_0.

16. Prosečno vreme izvršavanja nekog programa na računaru je $45s$. Hoćemo da kupimo bolji računar i pustili smo taj program na 30 drugih računara i dobili smo prosečno srednje vreme $44.5s$ i standardno odstupanje 2. Da li na osnovu dobijenih zaključaka treba kupiti novi računar?

    Jasno je da je $H_0 : m = 45$ i $H_1 : n < 45$. Takođe, $n = 30$, $\overline{X}_n 44,5$ i $\widetilde{S}_n = 2$. Računamo:

$$T_n = \frac{\overline{X}_n - m_0}{\sqrt{\widetilde{S}_n^2}}\sqrt{n} = \frac{44,5 - 45}{2}\sqrt{30} = -1,369306$$
$$W = \{T_n < c\}$$
$$p = P\{T_n < \hat{T_n}\} = \texttt{pt}(-1.369306, \texttt{df} = 29) = 0.09070778.$$

    Kako je $p > \alpha$, zaključujemo da prihvatamo hipotezu H_0.

`Teorija>` Do sada smo testirali hipoteze u kojima je figurisala jednakost parametra m. Pozabavimo se sada sledećim slučajem:
Neka je X slučajna veličina iz normalne $N(m, \sigma^2)$ raspodele. Ukoliko želimo da testiramo hipotezu $H_0 : \sigma^2 = \sigma_0^2$, za $\sigma_o^2 \in \mathbb{R}$, nasuprot neke od hipoteza $H_1 : \sigma^2 \neq \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o^2$ ili $H_1 : \sigma^2 < \sigma_o^2$, onda posmatramo statistiku

$$T_n = \frac{(n-1)\widetilde{S}_n^2}{\sigma^2},$$

za koju se ispostavlja da ima χ^2 raspodelu sa parametrom $n - 1$. To zapisujemo $T_n \sim \chi^2_{n-1}$.
Kako ova raspodela nije simetrična, to nam nije dovoljno jedno c, već moramo razlikovati c_1 i c_2, tako da za, na primer, $H_1 : \sigma^2 \neq \sigma_o^2$, skup W jednak je $W = \{T_n < c_1\} \cup \{T_n > c_2\}$. Odavde sledi:

$$c_1 = F^{-1}(\frac{\alpha}{2}),$$
$$c_2 = F^{-1}(1 - \frac{\alpha}{2}),$$
$$p = min(2P\{T_n < \hat{T_n}\},\ 2P\{T_n > \hat{T_n}\}).$$

Za $H_1 : \sigma^2 > \sigma_o^2$, dobija se $W = \{T_n > c\}$, $p = P\{T_n > \hat{T_n}\}$ i $c = F^{-1}(1 - \alpha)$ i
Za $H_1 : \sigma^2 < \sigma_o^2$, dobija se $W = \{T_n < c\}$, $p = P\{T_n < \hat{T_n}\}$ i $c = F^{-1}(\alpha)$.

17. Proizvode se teniske loptice i smatra se da je pritisak unutar loptica ima normalnu $N(28, 0.25)$ raspodelu. Uvodi se nova procedura u proizvodnji i gledamo da li ta procedura smanjuje disperziju. Na osnovu uzorka 25 slučajno odabranih loptica dobijeno je da je uzoračka disperzija 0.497. Da li na osnovu ovog uzorka možemo da smatramo da su loptice bolje?

    Ovaj zadatak ćemo uraditi na tri načina:

    1) Vidimo da je $H_0 : \sigma^2 = 0.25$ nasuprot $H_1 : \sigma^2 < 0.25$. U $\mathcal{R}$-u računamo $T_n = \frac{(n-1)\widetilde{S}_n^2}{\sigma^2}$ i p na sledeći način:

```
t_realizovano = 24*0.1497/0.25
p = pchisq(t_realizovano, df=24)
```

    i time dobijamo da je vrednost p jednaka 0.06218258, te hipotezu H_0 prihvatamo.

    2) U $\mathcal{R}$-u računamo c na prethodno opisan način:

```
c = qchisq(0.05, df=24)
```

    i iz rezultata za c (13.84843) i prethodno sračunatog T_n (14.3712) zaključujemo $T_n > c$, tj. prihvatamo hipotezu H_0.

`Teorija>` Testiranje hipoteza koje se odnose na disperziju možemo izvršiti pomoću funkcije:

$$\texttt{var.test(x, y = NULL,ratio = 1, alternative = c("two.sided",}$$
$$\texttt{"less","greater"), paired = FALSE, conf.level = 0.95, ...),}$$

koja se nalazi u paketu `PairedData`.

3) U našem slučaju, pozivom

$$\texttt{var.test(uzorak, ratio = 10, alternative = "two.sided")}$$

dobijamo rezultat

```
One-sample variance test

data:  x
X-squared = 56.148, df = 59, p-value = 0.8374
alternative hypothesis: true variance is not equal to 10
95 percent confidence interval:
6.837513 14.156645
sample estimates:
variance
9.516591
```

iz kojeg zaključujemo $p > \alpha$, tj. prihvatamo H_0.

`Teorija>` Neka slučajna veličina X ima Bernulijevu $B(p)$ (ili $B(1,p)$) raspodelu. Testiranje hipoteze $H_0 : p = p_0$, za $p_0 \in [0,1]$ protiv alternativne hipoteze $H_1 : p \neq p_0$, $H_1 : p < p_0$ ili $H_1 : p > p_0$, možemo vršiti statistikom

$$Z_n = \frac{\overline{X}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}},$$

za koju se ispostavlja da ima normalnu $N(0,1)$ raspodelu. U zavisnosti od H_1, skup W može biti $W = \{|Z_n| > c\}$, $W = \{Z_n < c\}$ ili $W = \{Z_n > c\}$, a p-vrednost testa je za, na primer, $H_1 : p \neq p_0$ jednaka $p = P\{|Z| > |\hat{Z}|\}$.

U $\mathcal{R}$-u nam je na raspolaganju funkcija

$$\texttt{binom.test(x, n,p = 0.5, alternative = c("two.sided",}$$
$$\texttt{"less","greater"), conf.level = 0.95).}$$

18. Smatramo da ako barem 30% kancelarija koristi naš softver, onda naša firma dobro posluje. U suprotnom, trebalo bi uložiti u reklamu. Uzet je uzorak od 500 kancelarija da bi se utvrdilo da li treba ulagati u marketing. Dobijeno je da 118 kancelarija koristi softver.

    (a) Da li je potrebno, na osnovu dobijenog istraživanja, uložiti u reklamu?

    (b) Napisati funkciju `softver(n)` u $\mathcal{R}$-u koja generiše istraživanje nad n kancelarija i prikazati rezultate testiranja da se softver koristi sa verovatnoćom 0.3.

    Rešenje:

    (a) Ovaj deo zadatka ćemo uraditi na tri načina:

1) Pozivom funkcije

$$\texttt{binom.test}(118, 500, \texttt{p} = 0.3, \texttt{alternative} = \text{"}\texttt{less}\text{"})$$

dobijamo rezultat

```
Exact binomial test

data:  118 and 500
number of successes = 118, number of trials = 500, p-value = 0.0008528
alternative hypothesis: true probability of success is less than 0.3
95 percent confidence interval:
0.0000000 0.2693567
sample estimates:
probability of success
0.236
```

iz kojeg zaključujemo $p < \alpha$, tj. treba uložiti u reklamu.

2) Računamo:

$$\hat{Z} = \frac{\frac{118}{500} - 0.3}{\sqrt{\frac{(0.3)(0.7)}{500}}} = -3.12288$$

$$p = P\{Z < \hat{Z}\} = \texttt{pnorm}(-3.12288) = 0.000895453$$

iz čega dobijamo $p < \alpha$.

3) Za $\alpha = 0.05$ računamo $c = \Phi^{-1}(0.05) = \texttt{qnorm}(0.05) = -1.644854$ odakle zaključujemo da je $T_n < c$, pa je zaključak jednak prethodnim.

(b) Tražena funkcija je:

```
softver<-function(n){
    uzorakbinom = rbinom(n, size=1, prob=0.3)
    brojuspeha = sum(uzorakbinom)
    binom.test(brojuspeha, n, p=0.3, alternative="less")
}
```

`Teorija>` **Testiranja hipoteza pomoću dva uzorka**[7]

U praksi se najčešće koriste sledeći testovi. Svi oni se nalaze u osnovnom paketu koji se automatski dostavlja uz $\mathcal{R}$.

t.test Ovaj test se koristi za testiranje hipoteze $H_0 : m_1 = m_2$ nasuprot neke od hipoteza $H_1 : m_1 \neq m_2$, $H_1 : m_1 < m_2$ ili $H_1 : m_1 > m_2$. Funkcija se poziva na sledeći način:

```
t.test(x, y, alternative = "two.sided",
             paired = FALSE, var.equal = TRUE).
```

Dakle, prosleđujemo joj dve realizacije uzorka, `x` i `y`, zatim alternativnu hipotezu, što može biti `"two.sided"`, `"greater"` ili `"less"`; `paired` služi da napravi razliku između toga da li je istraživanje vršeno nad istim uzorkom (`FALSE` ako ne postoji uparivanje, tj. ako su uzorci različiti, a `TRUE` ako su jednaki) i `var.equal = TRUE` kaže da su disperzije jednake.

[7] Ovaj deo je profesor pričao na predavanju i rekao je da će doći kao poslednji zadatak na kolokvijumu, iako nije obrađeno na vežbama.

$\boxed{\textbf{var.test}}$ Ovaj test se koristi za testiranje hipoteze $H_0 : \sigma_1^2 = \sigma_2^2$ nasuprot neke od hipoteza $H_1 : \sigma_1^2 \neq \sigma_2^2$, $H_1 : \sigma_1^2 < \sigma_2^2$ ili $H_1 : \sigma_1^2 > \sigma_2^2$. Funkcija se poziva na sledeći način:

```
var.test(x, y, alternative = "two.sided").
```

Dakle, prosleđujemo joj dve realizacije uzorka, `x` i `y`, pa zatim alternativnu hipotezu, što može biti `"two.sided", "greater"` ili `"less"`.

$\boxed{\textbf{prop.test}}$ Ovaj test se koristi za testiranje hipoteze $H_0 : p_1 = p_2$ nasuprot neke od hipoteza $H_1 : p_1 \neq p_2$, $H_1 : p_1 < p_2$ ili $H_1 : p_1 > p_2$.