

Analiza_rezultata_istraživanja

Bruno Blažeka, Lukrecija Puljić, Mateo Stjepanović

29 svibnja 2017

```
AnketaPodaci <- read.csv("AnketaPodaci.csv")

require(modeest)

## Loading required package: modeest

##
## This is package 'modeest' written by P. PONCET.
## For a complete list of functions, use 'library(help = "modeest")' or
## 'help.start()'.

require(plotrix)

## Loading required package: plotrix
```

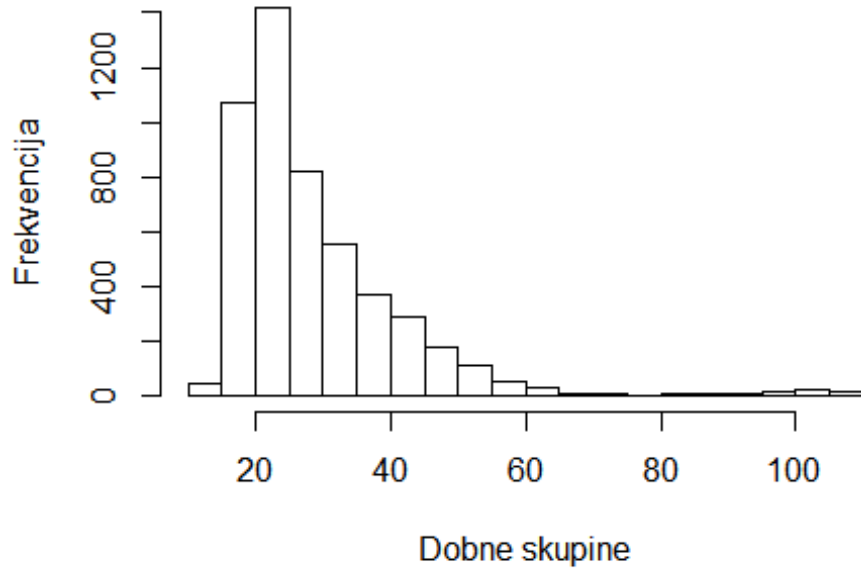
Prikaz ispitanika te rezultata ankete prema dobi ispitanika

```
art.sred.dob = mean(AnketaPodaci$age)
medijan.dob = median(AnketaPodaci$age)
mod.dob = mfv(AnketaPodaci$age)

var.dob = var(AnketaPodaci$age)
sd.dob = sd(AnketaPodaci$age)
IQR.dob = IQR(AnketaPodaci$age)
range.dom = max(AnketaPodaci$age) - min(AnketaPodaci$age)

# Histogram (15 razreda)
hist(AnketaPodaci$age, breaks = 15, main = "Dob histogram", xlab = "Dobne
skupine", ylab = "Frekvencija")
```

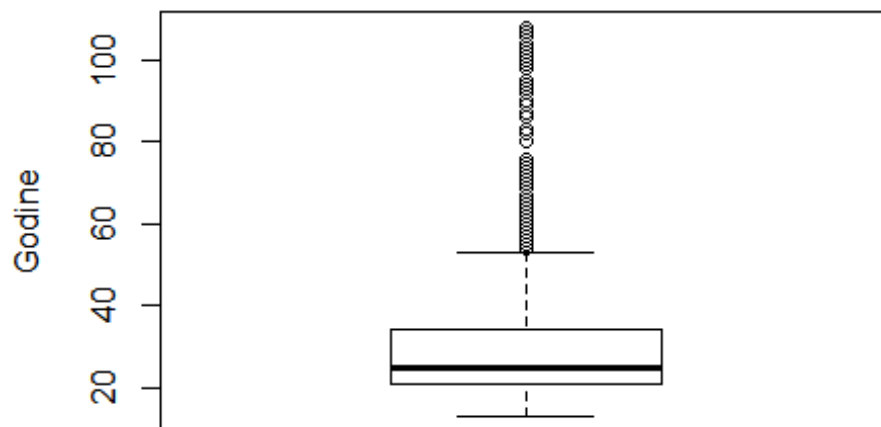
Dob histogram



Box and whiskers

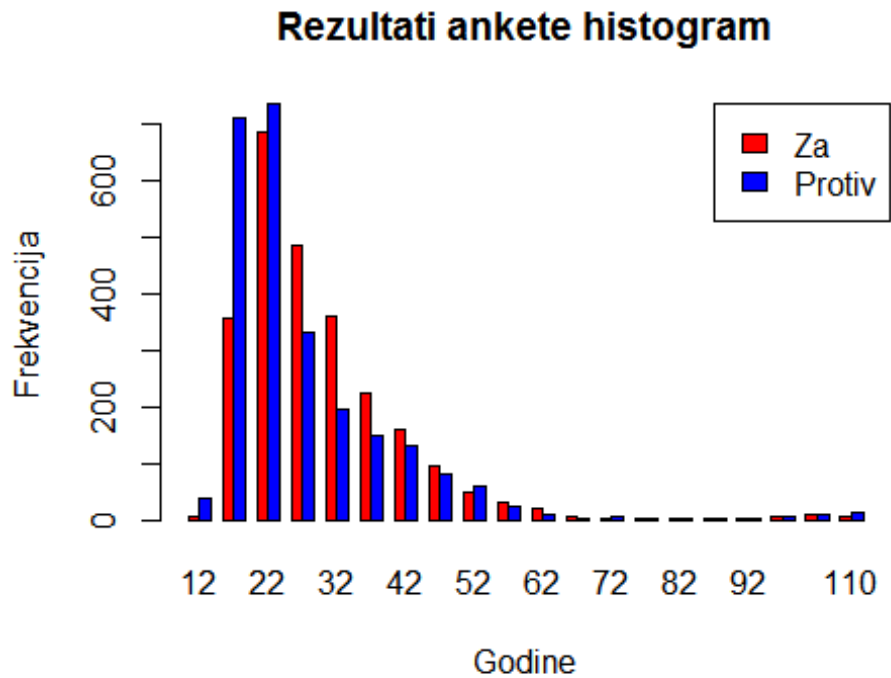
```
boxplot(AnketaPodaci$age, main="Godine Boxplot",ylab="Godine")
```

Godine Boxplot



```
# Histogram rezultati ankete i dob
```

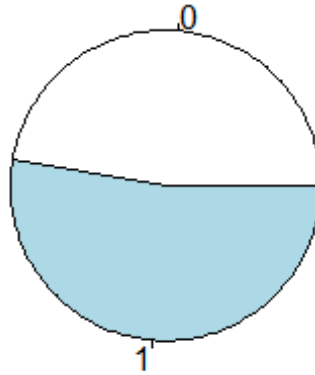
```
filtrirano.za = AnketaPodaci$age[AnketaPodaci$vote == 1]  
filtrirano.protiv = AnketaPodaci$age[AnketaPodaci$vote == 0]  
list1 = list(filtrirano.za, filtrirano.protiv)  
multhist(list1, breaks = 15, col=c("red","blue"), main="Rezultati ankete  
histogram",xlab="Godine",ylab="Frekvencija")  
legend("topright", c("Za", "Protiv"), fill=c("red", "blue"))
```



```
# piechart - glasanje odrasli
```

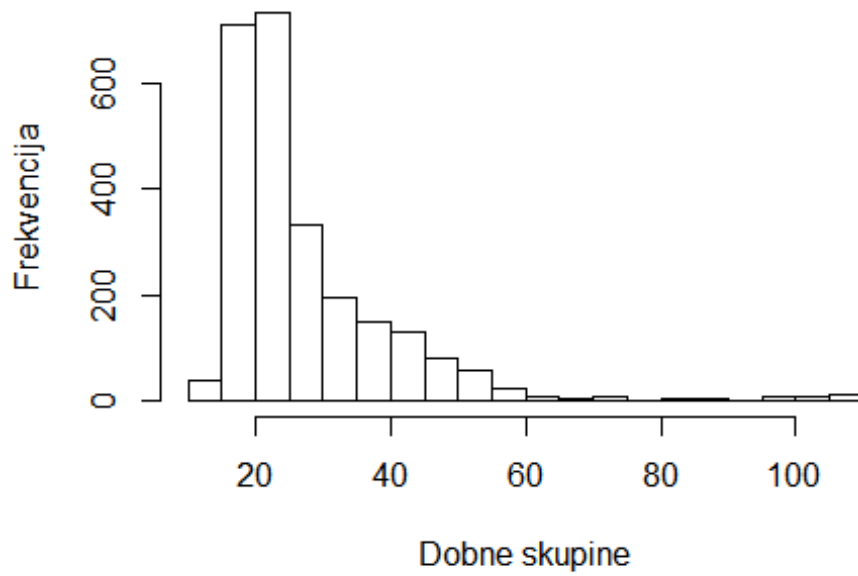
```
odrasli = AnketaPodaci[AnketaPodaci$age > 18,]  
pie(table(odrasli$vote),main="Glasanje starijih od 18 godina")
```

Glasanje starijih od 18 godina



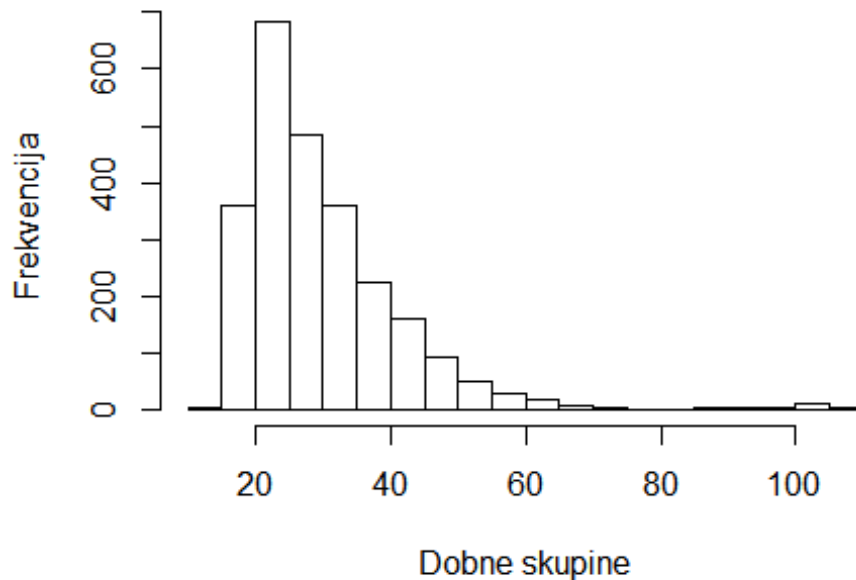
```
# Histogram dobi i glasanja za 0
za0 = AnketaPodaci[AnketaPodaci$vote == '0',]
hist(za0$age, breaks = 15, main = "Dob glasali PROTIV histogram", xlab =
"Dobne skupine", ylab = "Frekvencija")
```

Dob glasali PROTIV histogram



```
#Histogram dobi i glasanja za 1
za1 = AnketaPodaci[AnketaPodaci$vote == '1',]
hist(za1$age, breaks = 15, main = "Dob glasali ZA histogram", xlab = "Dobne
skupine", ylab = "Frekvencija")
```

Dob glasali ZA histogram



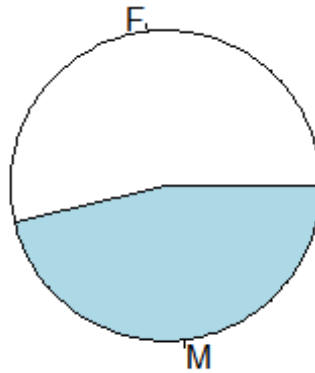
Na histogramu dobi se vidi izrazita desna zakrivljenost, što nam govori da će box plot za dob imati jako puno stršećih vrijednosti. S obzirom da je desna zakrivljenost a x os grafa raste s lijeva na desno vidimo da je veliki broj osoba između 18 i 40 godina pristupilo istraživanju, što nam može reći da je cijelo istraživanje namijenjeno mladim ljudima.

Na histogramu usporedbe rezultata i godina pristupnika vidimo da se nakon 22 godine života mišljenje ljudi mijenja o danoj temi. To bi trebalo ispitati preko testa te utvrditi dali je ta promjena u mišljenju značajna.

Prikaz ispitanika te rezultata ankete prema spolu

```
pie(table(AnketaPodaci$gender),main="Raspodjela muških i ženskih osoba u  
uzorku")
```

Raspodjela muških i ženskih osoba u uzorku



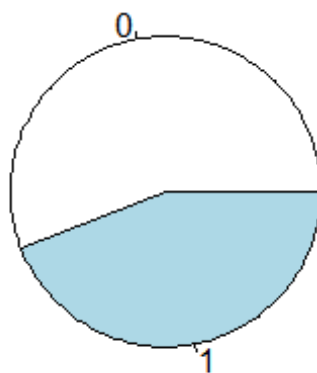
```
n = length(AnketaPodaci$gender)
m = length(which(AnketaPodaci$gender == "M"))
f = length(which(AnketaPodaci$gender == "F"))

m.udio = m / n
f.udio = f / n

# piechart - glasanje žene

zene = AnketaPodaci[AnketaPodaci$gender == 'F',]
pie(table(zene$vote),main="Glasanje žena")
```

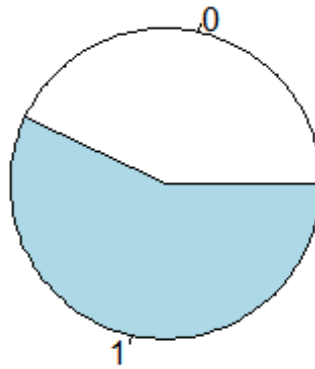
Glasanje žena



```
# piechart - glasanje muškarci
```

```
muski = AnketaPodaci[AnketaPodaci$gender == 'M',]  
pie(table(muski$vote),main="Glasanje muškaraca")
```


Glasanje muškaraca



```
# test o dvije proporcije spol, glasalo za
n = c(nrow(zene), nrow(muski))
x = c(length(which(zene$vote == "1")),
      length(which(muski$vote == "1")))
prop.test(x, n, alternative = "t", correct = "FALSE")

##
## 2-sample test for equality of proportions without continuity
## correction
##
## data:  x out of n
## X-squared = 84.469, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.1579101 -0.1027958
## sample estimates:
##      prop 1      prop 2
## 0.4397770 0.5701299

# test o dvije proporcije spol, glasalo protiv

n = c(nrow(zene), nrow(muski))
x = c(length(which(zene$vote == "0")),
      length(which(muski$vote == "0")))
prop.test(x, n, alternative = "t", correct = "FALSE")
```

```
##
## 2-sample test for equality of proportions without continuity
## correction
##
## data:  x out of n
## X-squared = 84.469, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1027958 0.1579101
## sample estimates:
##      prop 1      prop 2
## 0.5602230 0.4298701
```

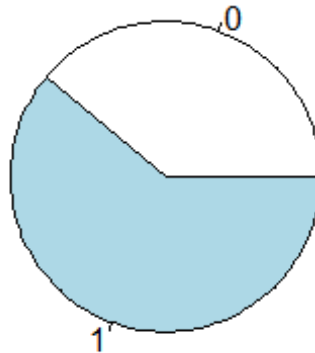
Udio muškaraca i žena je zadovoljavajući da bi nam ovaj uzorak bio značajan po toj osnovi. Iako je broj muškaraca neznatno manji, iz grafa je vidljivo da jako veći broj muškaraca glasa ZA, za razliku od žena koje su većinski PROTIV. To isčitavanje je ispitano testom o jednakosti proporcija dva uzorka koje se provelo na uzorcima muškaraca i žena koji su glasali za i protiv. Iz rezultata testa koji se odnosi na glasanje ZA, vidimo da proporcije spadaju u interval pouzdanosti, te su stoga one jednake. Zaključujemo da je jednak omjer žena i muškaraca glasao ZA. Ista stvar se može zaključiti iz rezultata za glasanje Protiv.

Prikaz ispitanika te rezultata ankete prema lokaciji

```
# piechart - glasanje Zagreb

zgb = AnketaPodaci[AnketaPodaci$locality == "Zagreb, Croatia",]
pie(table(zgb$vote), main="Rezultati u Zagrebu")
```

Rezultati u Zagrebu



glasaci za u 5 najmnogobrojnijih gradova

```
tail(sort(table(za1$locality)),5)
```

```
##
## Split, Croatia Osijek, Croatia Rijeka, Croatia      unkown
##           85           108           116           341
## Zagreb, Croatia
##           1025
```

glasaci protiv u 5 najmnogobrojnijih gradova

```
tail(sort(table(za0$locality)),5)
```

```
##
## Rijeka, Croatia Osijek, Croatia Split, Croatia      unkown
##           57           87           139           414
## Zagreb, Croatia
##           653
```

glasacice u 5 najmnogobrojnijih gradova

```
tail(sort(table(zene$locality)),5)
```

```
##
## Rijeka, Croatia Osijek, Croatia Split, Croatia      unkown
##           89           101           124           361
```

```
## Zagreb, Croatia
##                857

# glasaci u 5 najmnogobrojnijih gradova

tail(sort(table(muski$locality)),5)

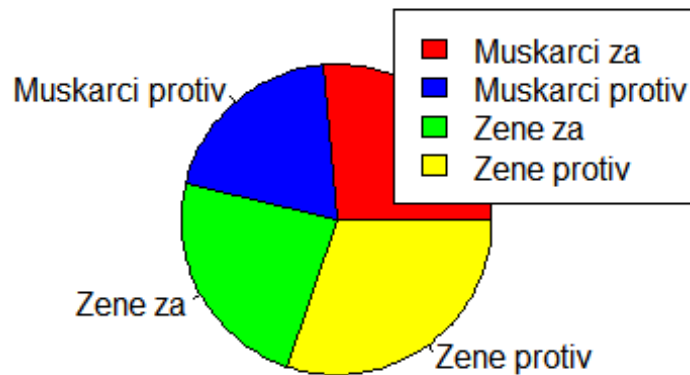
##
## Rijeka, Croatia Osijek, Croatia Split, Croatia      unkown
##                84                94                100                394
## Zagreb, Croatia
##                821
```

Rezultati ankete i spol

```
n.m = length(which(AnketaPodaci$gender == "M"))
n.m.za = length(which(AnketaPodaci$gender == "M" & AnketaPodaci$vote == 1))
n.m.protiv = n.m - n.m.za
n.m.udio.za = n.m / n.m.za
n.m.udio.protiv = 1 - n.m.udio.za

n.f = length(which(AnketaPodaci$gender == "F"))
n.f.za = length(which(AnketaPodaci$gender == "F" & AnketaPodaci$vote == 1))
n.f.protiv = n.f - n.f.za
n.f.udio.za = n.f / n.f.za
n.f.udio.protiv = 1 - n.f.udio.za

pie(c(n.m.za, n.m.protiv, n.f.za, n.f.protiv), labels = c("Muskarci za",
"Muskarci protiv", "Zene za", "Zene protiv"), col = c("Red", "Blue", "Green",
"Yellow"))
legend("topright", c("Muskarci za", "Muskarci protiv", "Zene za", "Zene
protiv"), fill=c("red", "blue", "green", "yellow"))
```



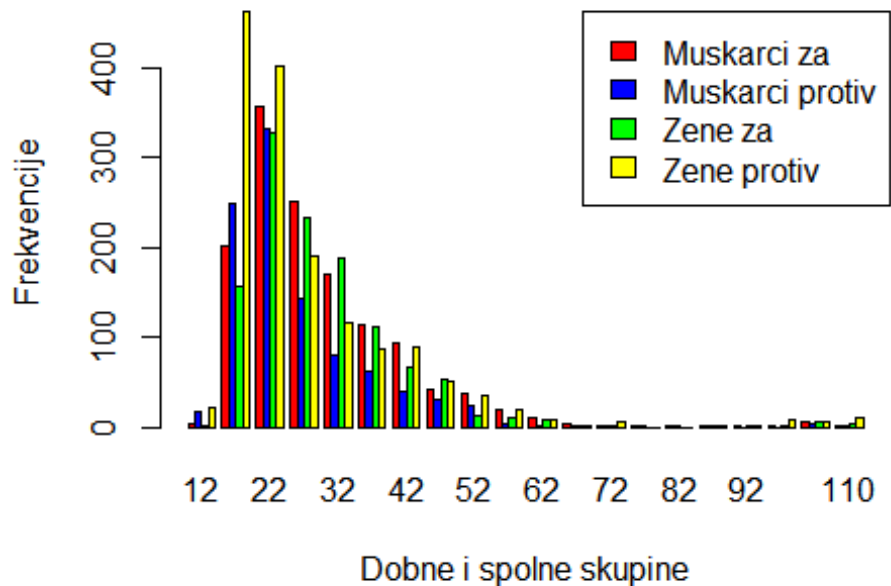
Histogram dob, spol i ishod glasanja

```

za.male = AnketaPodaci$age[AnketaPodaci$gender == "M" & AnketaPodaci$vote == 1]
protiv.male = AnketaPodaci$age[AnketaPodaci$gender == "M" & AnketaPodaci$vote == 0]
za.female = AnketaPodaci$age[AnketaPodaci$gender == "F" & AnketaPodaci$vote == 1]
protiv.female = AnketaPodaci$age[AnketaPodaci$gender == "F" & AnketaPodaci$vote == 0]
list.age.gender = list(za.male, protiv.male, za.female, protiv.female)
multhist(list.age.gender, breaks = 15, main = "Rezultati ankete po dobi i spolu", xlab = "Dobne i spolne skupine", ylab = "Frekvencije", col=c("Red", "Blue", "Green", "Yellow"))
legend("topright", c("Muskarci za", "Muskarci protiv", "Zene za", "Zene protiv"), fill=c("red", "blue", "green", "yellow"))

```

Rezultati ankete po dobi i spolu



Iz danog

histograma vidimo da su muškarci protiv do 22. godine a poslije im se mišljenje mijenja, dok je kod žena obrnuta situacija. One su protiv do 22. godine a poslije većinom za.

Prikaz ispitanika prema stupnju cvora u mreži

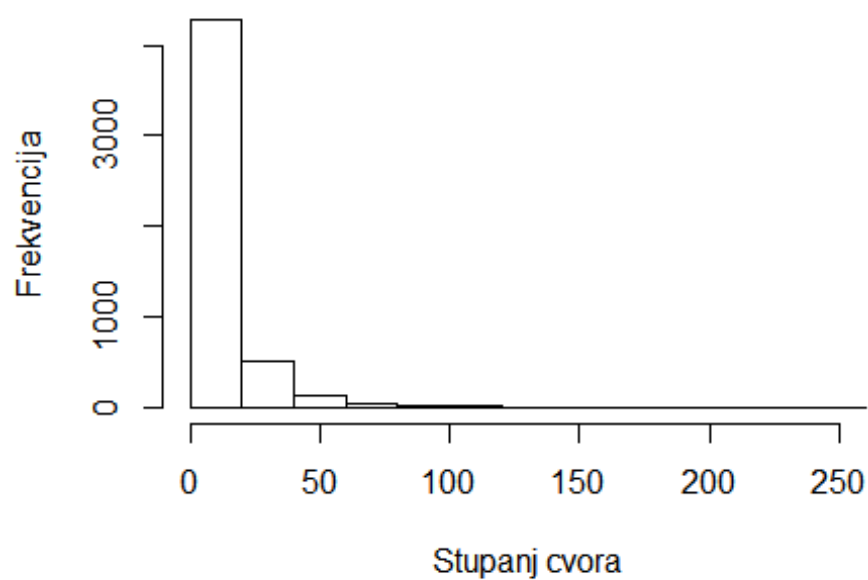
```
art.sred.degree = mean(AnketaPodaci$degree)
medijan.deegree = median(AnketaPodaci$degree)
mod.degree = mfv(AnketaPodaci$degree)

var.degree = var(AnketaPodaci$degree)
sd.degree = sd(AnketaPodaci$degree)
IDR.degree = IQR(AnketaPodaci$degree)
range.degree = max(AnketaPodaci$degree) - min(AnketaPodaci$degree)
```

Histogram stupanj cvora u mreži

```
hist(AnketaPodaci$degree, main = "Stupanj cvora u mreži", xlab = "Stupanj
cvora", ylab = "Frekvencija")
```

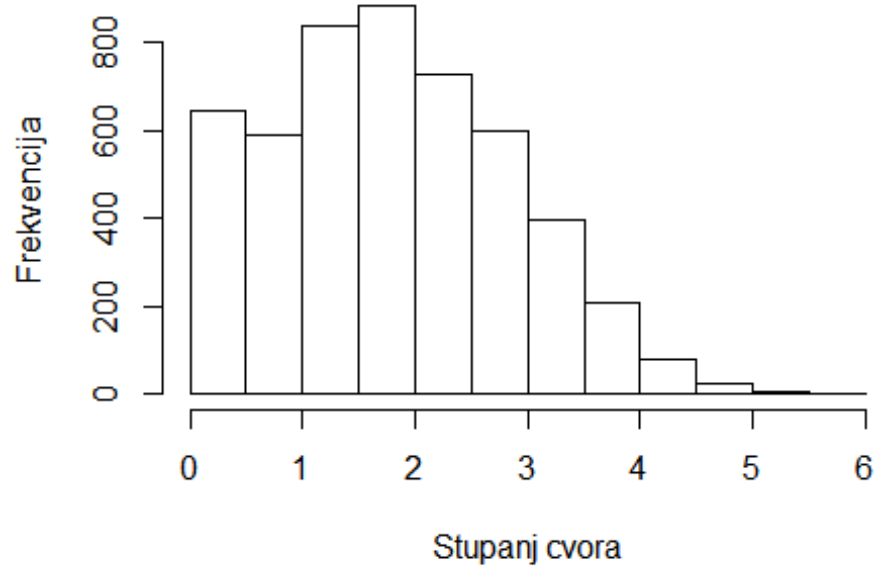
Stupanj cvora u mrezi



#Histogram stupanj cvora u mrezi log transformacija

```
hist(log(AnketaPodaci$degree), main = "Stupanj cvora u mrezi-logaritamska  
transformacija", xlab = "Stupanj cvora", ylab = "Frekvencija")
```

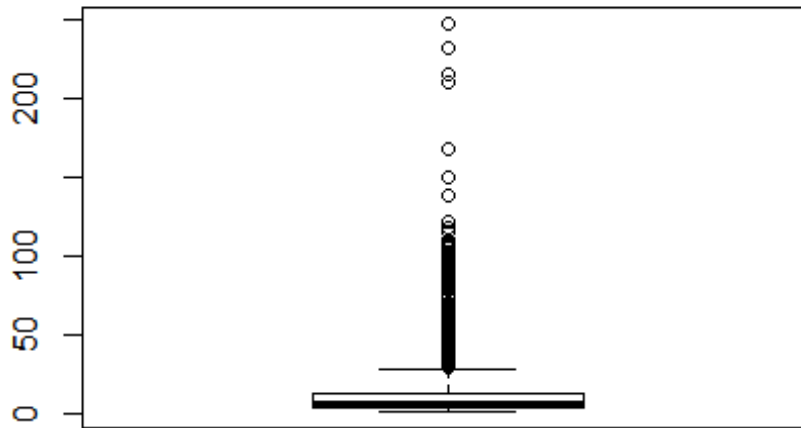
Stupanj cvora u mrezi-logaritamska transformacij



```
# Box and whiskers
```

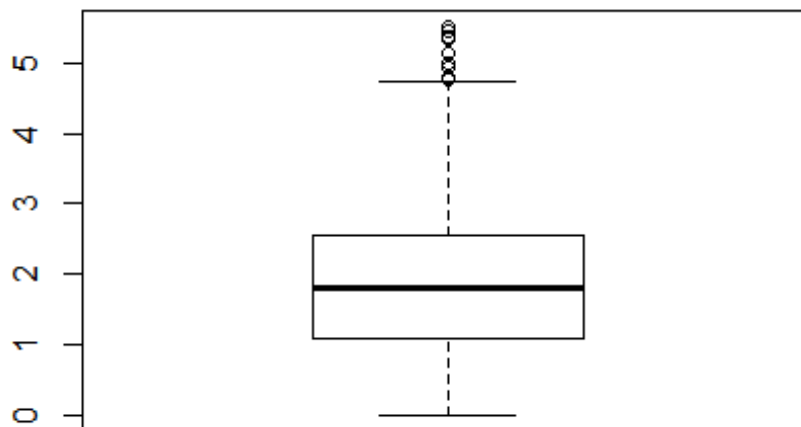
```
boxplot(AnketaPodaci$degree,main="Prikaz ispitanika prema stupnju čvora u mreži")
```


Prikaz ispitanika prema stupnju čvora u mreži



```
boxplot(log(AnketaPodaci$degree),main="Prikaz ispitanika prema stupnju čvora  
u mreži - logaritamska transformacija")
```

itanika prema stupnju čvora u mreži - logaritamska t

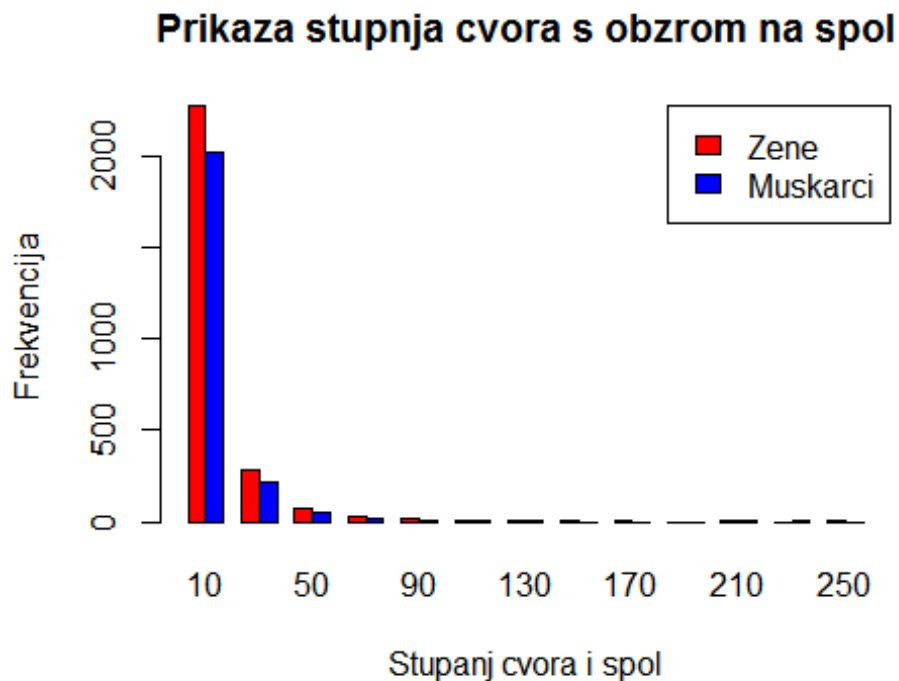


Na prvom histogramu možemo vidjeti izraženu pozitivnu zakrivljenost, koja je naposljetku uzrok

jako mnogo stršećih podataka. U tom slučaju najmjerodavnija mjera centralnosti je mod. Budući da je zakrivljenost desna, transformacija podataka se obavlja pomoću logaritamske funkcije.

Stupanj cvora i spol

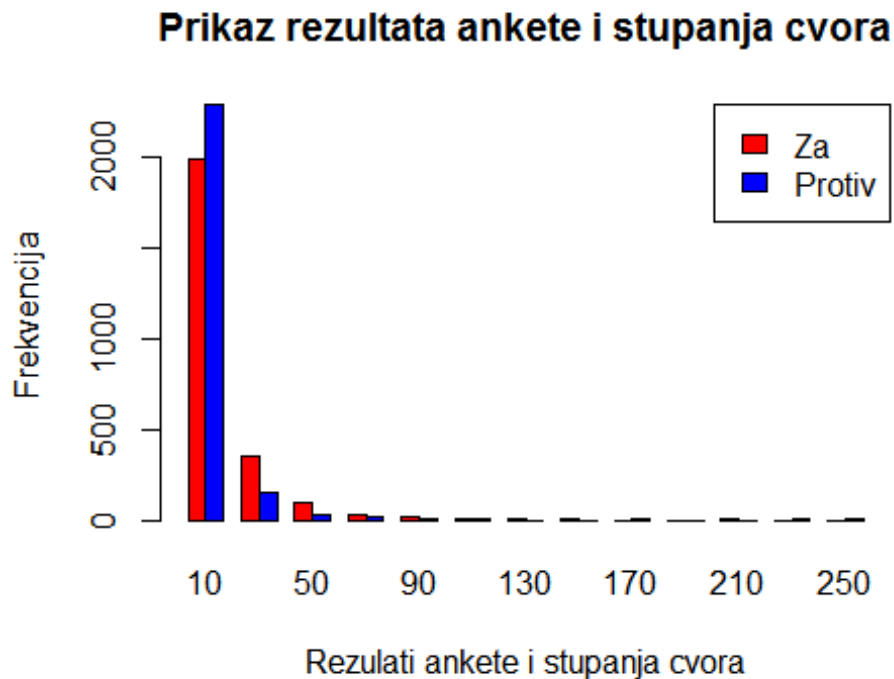
```
multhist(list(AnketaPodaci$degree[AnketaPodaci$gender == "F"],
AnketaPodaci$degree[AnketaPodaci$gender == "M"]), main = "Prikaza stupnja
cvora s obzrom na spol", xlab = "Stupanj cvora i spol", ylab = "Frekvencija",
col = c("Red", "Blue"))
legend("topright", c("Zene", "Muskarci"), fill=c("red", "blue"))
```



Rezultat ankete i

stupanj cvora

```
multhist(list(AnketaPodaci$degree[AnketaPodaci$vote == 1],
AnketaPodaci$degree[AnketaPodaci$vote == 0]), main = "Prikaz rezultata ankete
i stupanja cvora", xlab = "Rezultati ankete i stupanja cvora", ylab =
"Frekvencija", col=c("Red", "Blue"))
legend("topright", c("Za", "Protiv"), fill=c("red", "blue"))
```

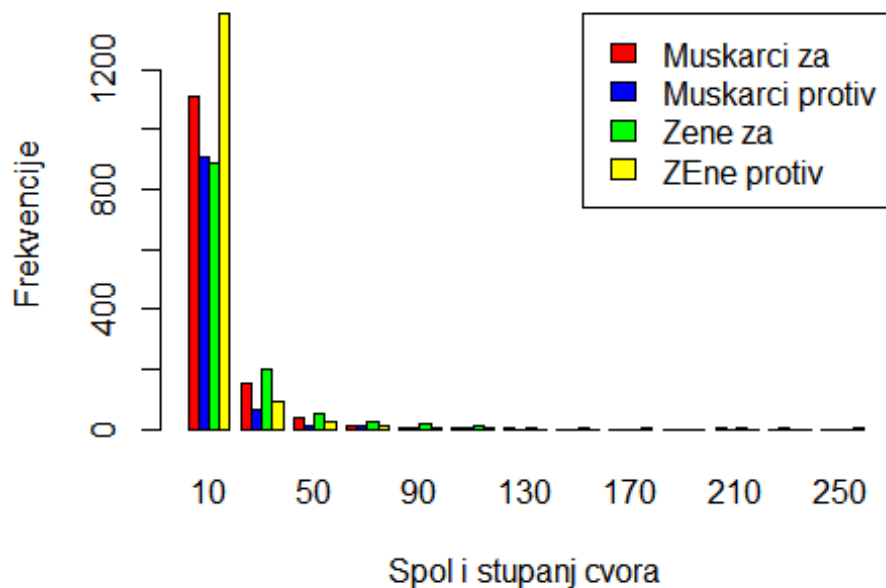


Rezultati ankete,
 spol i stupanj cvora

```

multhist(list(AnketaPodaci$degree[AnketaPodaci$vote == 1 &
AnketaPodaci$gender == "M"], AnketaPodaci$degree[AnketaPodaci$vote == 0 &
AnketaPodaci$gender == "M"], AnketaPodaci$degree[AnketaPodaci$vote == 1 &
AnketaPodaci$gender == "F"], AnketaPodaci$degree[AnketaPodaci$vote == 0 &
AnketaPodaci$gender == "F"]), col = c("red", "blue", "green", "yellow"), main
= "Rezultati ankete s obzorom na spol i stupanj cvora", xlab = "Spol i
stupanj cvora", ylab = "Frekvencije")
legend("topright", c("Muskarci za", "Muskarci protiv", "Zene za", "Zene
protiv"), fill=c("red", "blue", "green", "yellow"))
  
```

Rezultati ankete s obzorom na spol i stupanj cvor



Prikaz ispitanika

prema centralnost cvora

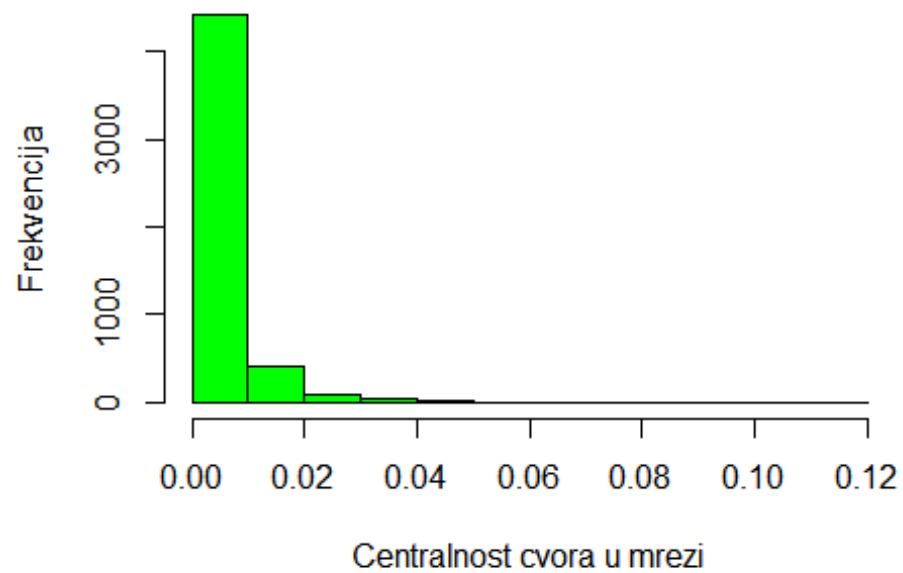
```
art.sred.center = mean(AnketaPodaci$katz_centrality)
medijan.center = median(AnketaPodaci$katz_centrality)
mod.center = mfv(AnketaPodaci$katz_centrality)

var.center = var(AnketaPodaci$katz_centrality)
sd.center = sd(AnketaPodaci$katz_centrality)
IDR.center = IQR(AnketaPodaci$katz_centrality)
range.center = max(AnketaPodaci$katz_centrality) -
min(AnketaPodaci$katz_centrality)

# Prikaz ispitanika prema centralnosti covra u mrezi

hist(AnketaPodaci$katz_centrality, breaks = 15, main = "Prikaz ispitanika
prema centralnosti cvora u mrezi", xlab = "Centralnost cvora u mrezi", ylab =
"Frekvencija", col= "Green")
```

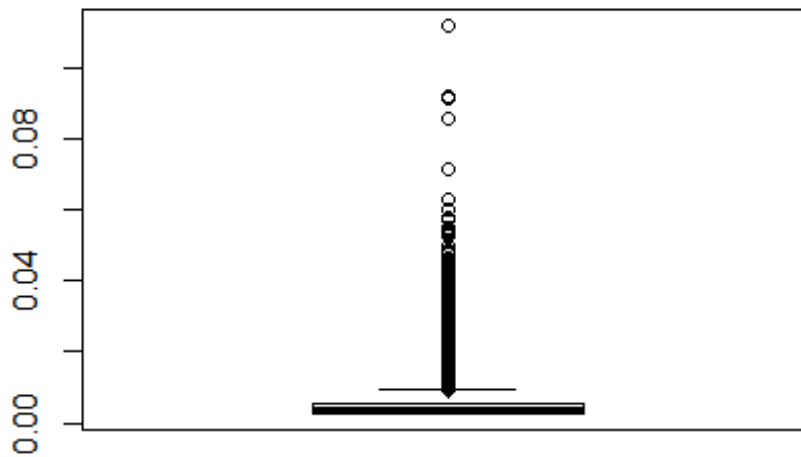
Prikaz ispitanika prema centralnosti cvora u mreži



```
# Box and whiskers
```

```
boxplot(AnketaPodaci$katz_centrality,main="Prikaz ispitanika prema  
centralnosti čvora")
```

Prikaz ispitanika prema centralnosti čvora

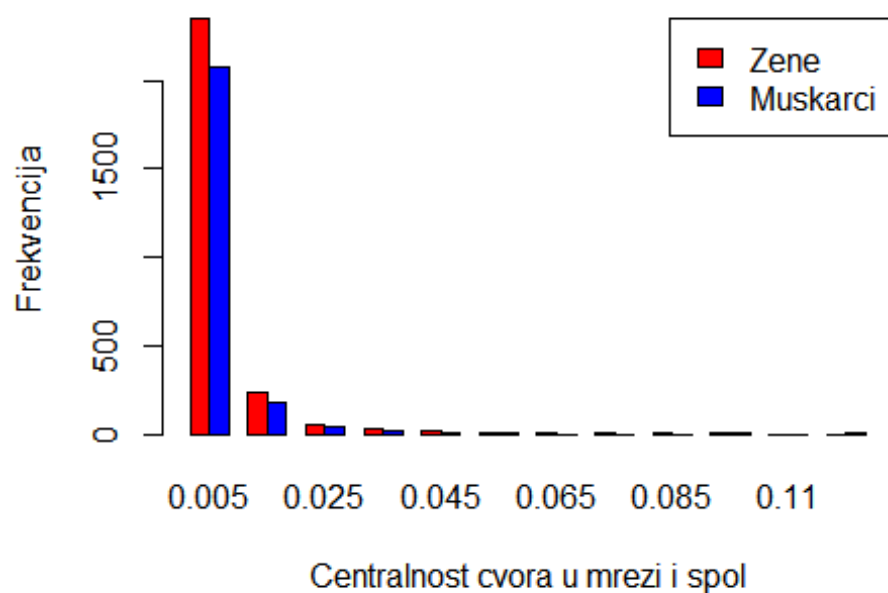


Ponovno je jako puno stršećih podataka te je histogram izrazito pozitivno nakrivljen, zbog toga je naprikladinija mjera centralnosti mod.

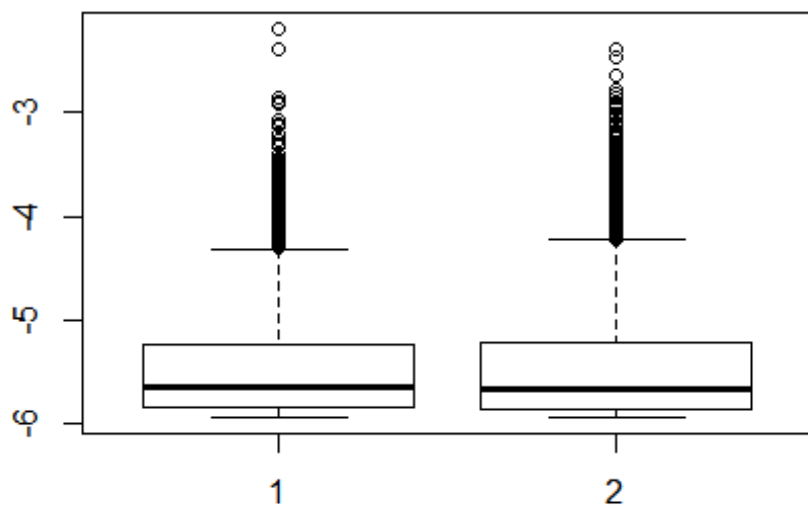
Centralnost cvora u mrezi i spol

```
multhist(list(AnketaPodaci$katz_centrality[AnketaPodaci$gender == "F"],
AnketaPodaci$katz_centrality[AnketaPodaci$gender == "M"]), main = "Prikaza
centralnosti cvora u mrezi s obzrom na spol", xlab = "Centralnost cvora u
mrezi i spol", ylab = "Frekvencija", col = c("Red", "Blue"))
legend("topright", c("Žene", "Muškarci"), fill=c("red", "blue") )
```

Prikaza centralnosti cvora u mrezi s obzrom na sp



```
male = AnketaPodaci$katz_centrality[AnketaPodaci$gender == "M"]  
female = AnketaPodaci$katz_centrality[AnketaPodaci$gender == "F"]  
boxplot(log(male), log(female))
```



```

var.test(male,female)

##
## F test to compare two variances
##
## data: male and female
## F = 0.71601, num df = 2309, denom df = 2689, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.6619263 0.7746987
## sample estimates:
## ratio of variances
## 0.7160106

t.test(male,female,alt="two.sided",var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data: male and female
## t = -2.5238, df = 4996.9, p-value = 0.01164
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.0008309770 -0.0001043986
## sample estimates:
## mean of x mean of y
## 0.005481368 0.005949056

```

Iz početnog histograma vidimo da je jaka desna zakrivljenost, stoga smo za boxplot odlučili koristiti logaritamsku transformaciju. Iz boxplota se može vidjeti da ne postoji razlika u očekivanjima, ipak je potrebno provesti test da vidimo je li razlika koja postoji značajna. Proveli smo t-test te zaključili da su očekivanja uistinu jednaka. Testom varijance smo zaključili da odbacujemo početnu hipotezu o jednakosti varijanci.

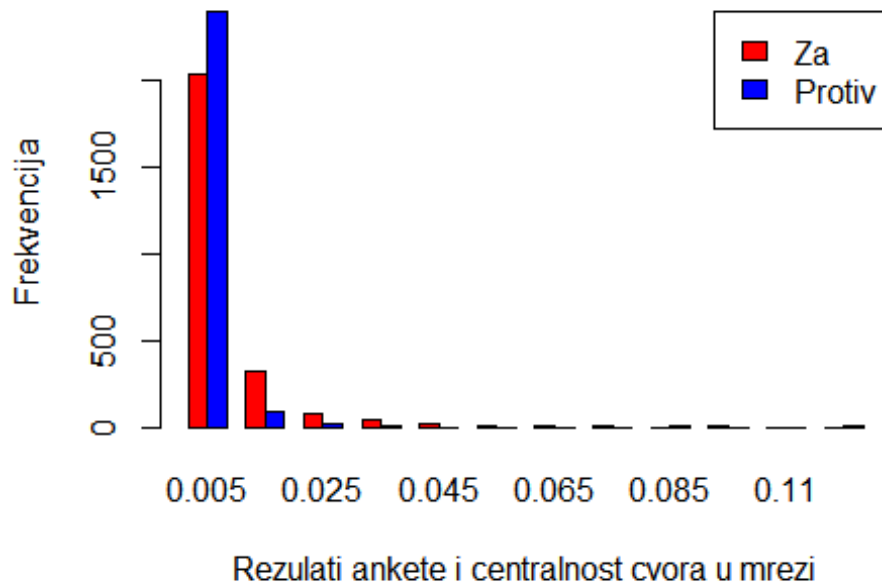
Rezultat ankete i centralnost cvora u mreži

```

multhist(list(AnketaPodaci$katz_centrality[AnketaPodaci$vote == 1],
AnketaPodaci$katz_centrality[AnketaPodaci$vote == 0]), main = "Prikaz
rezultata ankete i centralnost cvora u mreži", xlab = "Rezultati ankete i
centralnost cvora u mreži", ylab = "Frekvencija", col=c("Red", "Blue"))
legend("topright", c("Za", "Protiv"), fill=c("red", "blue"))

```


Prikaz rezultata ankete i centralnost cvora u mrezi

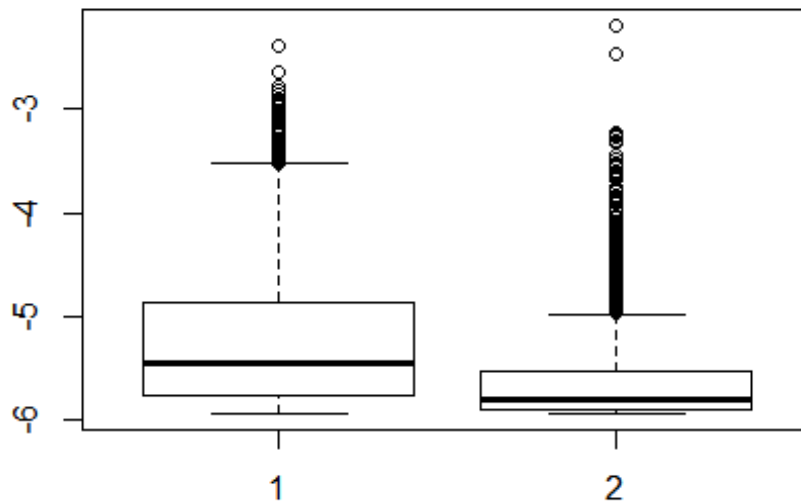


```
glasaliZa = AnketaPodaci$katz_centrality[AnketaPodaci$vote == 1]
glasaliProtiv = AnketaPodaci$katz_centrality[AnketaPodaci$vote == 0]

var.test(glasaliZa,glasaliProtiv)

##
## F test to compare two variances
##
## data:  glasaliZa and glasaliProtiv
## F = 3.5314, num df = 2499, denom df = 2499, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  3.264972 3.819484
## sample estimates:
## ratio of variances
##           3.531361

boxplot(log(glasaliZa),log(glasaliProtiv))
```



```
t.test(glasaliZa,glasaliProtiv,alternative = "greater",var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data:  glasaliZa and glasaliProtiv
## t = 16.79, df = 3809.3, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.002758896      Inf
## sample estimates:
##  mean of x   mean of y
## 0.007262284 0.004203684
```

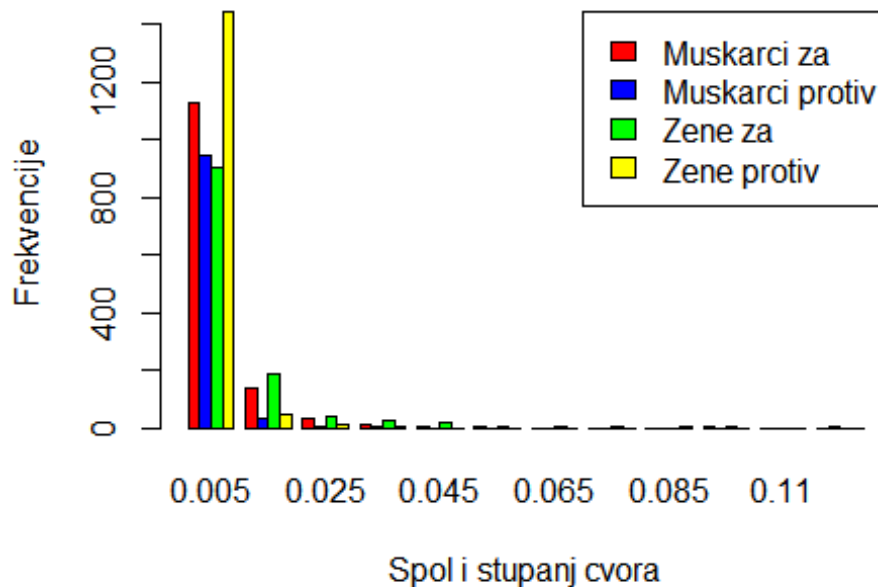
Ponovno zbog izrazite desne zakrivljenosti provodimo logaritamsku transformaciju nad podacima, te provodimo test jednakosti varijanci i zaključujemo da odbacujemo početnu hipotezu o jednakosti. Nakon t-testa dobivamo rezultat koji je u skladu s očekivanjima, tj. očekivanja uzoraka se razlikuju.

Rezultati ankete, spol i centralnost cvora u mreži

```
multhist(list(AnketaPodaci$katz_centrality[AnketaPodaci$vote == 1 &
AnketaPodaci$gender == "M"], AnketaPodaci$katz_centrality[AnketaPodaci$vote
== 0 & AnketaPodaci$gender == "M"],
AnketaPodaci$katz_centrality[AnketaPodaci$vote == 1 & AnketaPodaci$gender ==
"F"], AnketaPodaci$katz_centrality[AnketaPodaci$vote == 0 &
AnketaPodaci$gender == "F"]), col = c("red", "blue", "green", "yellow"), main
```

```
= "Rezultati ankete s obzorom na spol i centralnost cvora u mrezi", xlab =
"Spol i stupanj cvora", ylab = "Frekvencije")
legend("topright", c("Muskarci za", "Muskarci protiv", "Zene za", "Zene
protiv"), fill=c("red", "blue", "green", "yellow"))
```

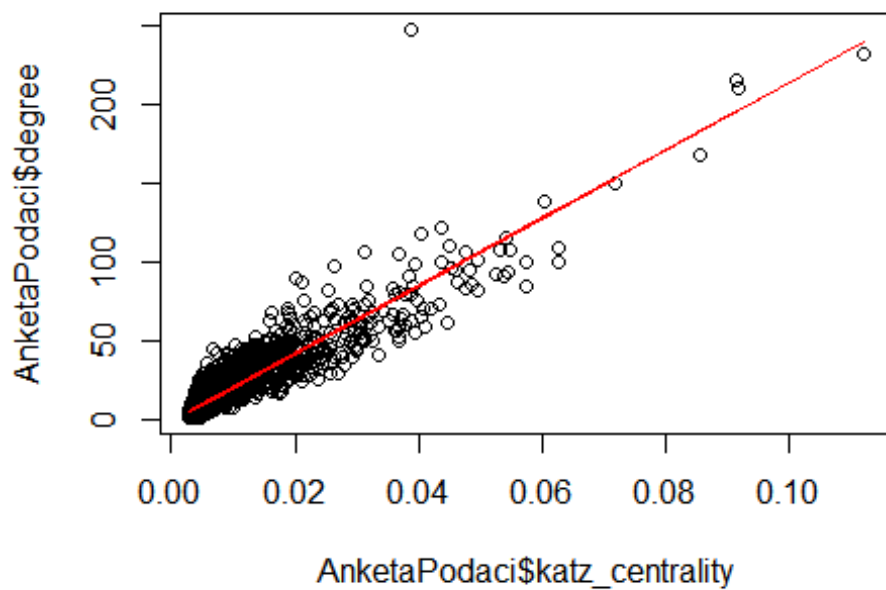
Rezultati ankete s obzorom na spol i centralnost cvora



```
fit = lm(degree~katz centrality, AnketaPodaci)
fit$coefficients

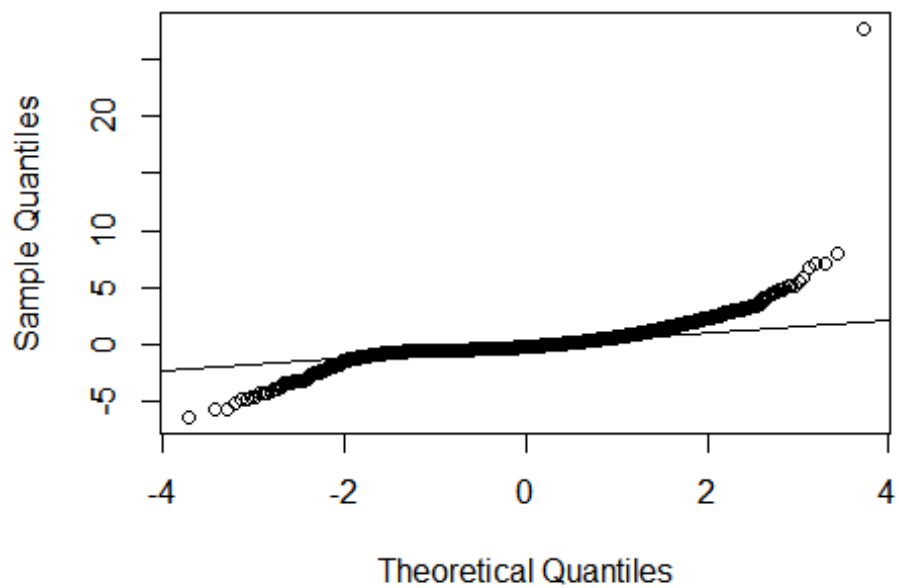
##      (Intercept) katz centrality
##      -1.4313      2151.5323

plot(AnketaPodaci$katz centrality, AnketaPodaci$degree)
lines(AnketaPodaci$katz centrality, fit$fitted.values, col='red')
```



```
qqnorm(rstandard(fit))
qqline(rstandard(fit))
```

Normal Q-Q Plot



```
cor(AnketaPodaci$katz_centrality, AnketaPodaci$degree)
```

```
## [1] 0.922016
```

Koeficijent korelacije je jako visok, što govori da su ove dvije varijable povezane. Što se moglo i zaključiti iz grafa. Regresijom predviđamo stupanj na temelju centralnosti, te možemo vidjeti da porastom stupnja raste i centralnost. Izračunom se dobije koeficijent regresije 2151 što je zapravo nagib pravca regresije. Te je naposljetku potrebno ispitati normalnost reziduala. Ona se može ispitati preko grafa, mi smo se odlučili na qq-plot, koji pokazuje zadovoljavajuću normalnost reziduala.