

Analiza rezultata istraživanja

Josip Jukić, Josip Matak, Josip Mrđen

Lipanj, 2018

Opis problema

Provedeno je istraživanje o broju stranica neke knjige u sklopu jednog kolegija na fakultetu. Stotinjak studenata, podijeljenih u grupe odgovarala je na tri pitanja prije i poslije no što im je rečeno da će za točan odgovor dobiti određenu nagradu. Sljedeća tri pitanja postavljena su studentima :

- Koliko stranica ima pokazana knjiga ?
- Kako će broj stranica procijeniti cijela grupa ?
- Kako će broj stranica procijeniti moji prijatelji ?

Opis podataka

Skup se sastoji od 106 podataka sa sljedećim opisom značajki :

- **Grupa.Pred** - grupa u kojoj ispitanik sluša predavanja
- **student__bezNagrade** - studentova procjena broja stranica prije saznanja o nagradi
- **cijela.Grupa__bezNagrade** - studentova procjena mišljenja cijele grupe ispitanika prije saznanja o nagradi
- **samo.Prijatelji__bezNagrade** - studentova procjena mišljenja njegovih prijatelja prije saznanja o nagradi
- **student__Nagrada** - studentova procjena broja stranica nakon saznanja o nagradi
- **cijela.Grupa__Nagrada** - studentova procjena mišljenja cijele grupe ispitanika nakon saznanja o nagradi
- **samo.Prijatelji__Nagrada** - studentova procjena mišljenja njegovih prijatelja nakon saznanja o nagradi
- **MI** - broj bodova postignut na međuispitu
- **spol** - spol studenta

Zadan je i točan broj stranica knjige koja je bila procjenjivana.

```
CORRECT_PAGE_NO = 1171
```

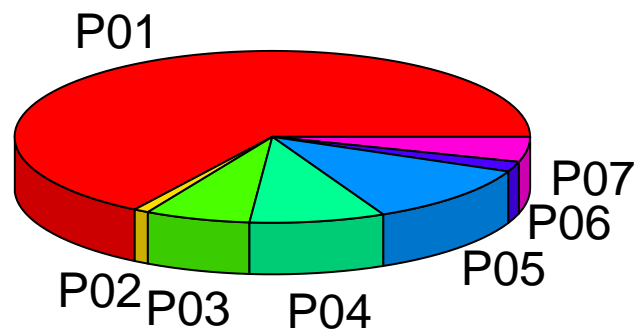
Podaci su učitani iz csv datoteke sa svim navedenim značajkama.

```
book = read.csv('knjiga.csv', header = TRUE)
```

Raspored intervjuiranih studenata po grupama:

```
library(plotrix)
pie3D(table(book$Grupa.Pred), labels=levels(book$Grupa.Pred),
      main="Raspodjela ispitanika po grupama")
```

Raspodjela ispitanika po grupama

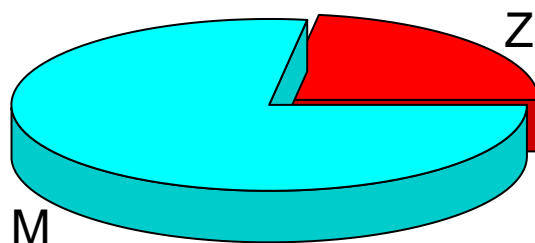


Prethodni grafikon pokazuje da anketa nije provedena ravnomjerno po grupama. Najveći broj ispitanika dolazi iz grupe P01.

Zatim je isti postupak proveden na udio muškaraca i žena koji su odgovorili na anketna pitanja.

```
pie3D(table(book$spol), labels=c("Z", "M"),
      main="Raspodjela ispitanika po spolu", explode = 0.03)
```

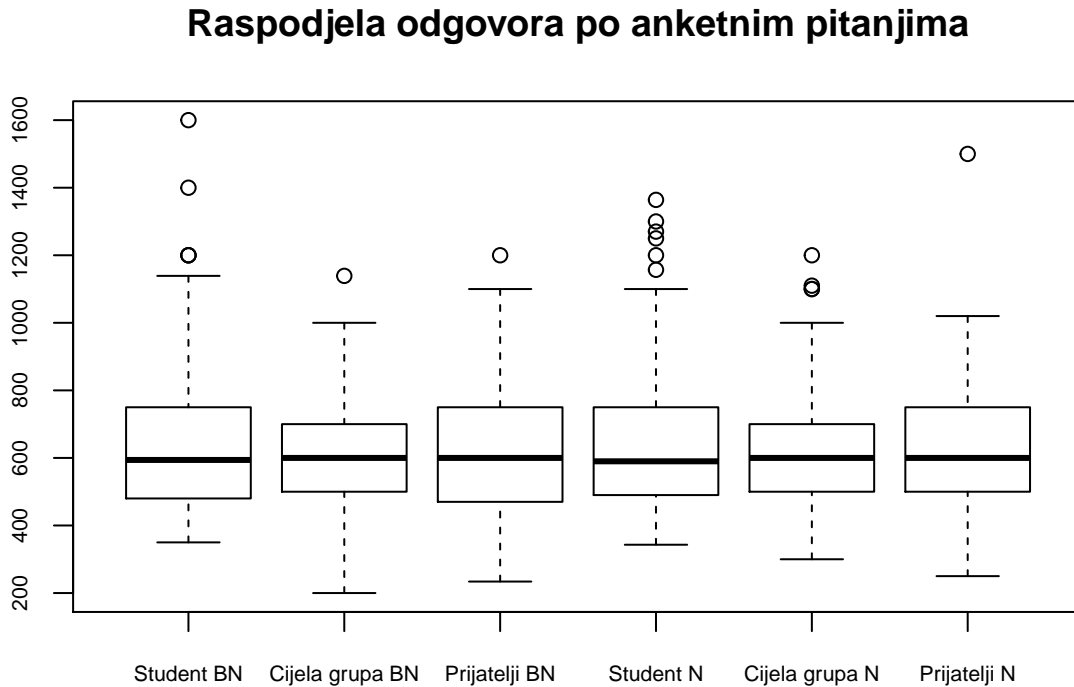
Raspodjela ispitanika po spolu



Iz prikazanog grafa proizlazi zaključak da je anketa provedena na nekom tehničkom fakultetu uzevši u obzir udio muškaraca i žena koji su na nju odgovorili.

U nastavku je prikazana raspodjela odgovora u pravokutnom dijagramu za svaki od anketnih pitanja.

```
boxplot(book$student_bezNagrade, book$cijela.grupa_bezNagrade,
        book$samo.prijatelji_bezNagrade, book$student_Nagrada,
        book$cijela.grupa_Nagrada, book$samo.prijatelji_Nagrada,
        main="Raspodjela odgovora po anketnim pitanjima",
        names=c("Student BN", "Cijela grupa BN", "Prijatelji BN",
                "Student N", "Cijela grupa N", "Prijatelji N"), cex.axis=0.7)
```



Na svako anketno pitanje dani su vrlo slični odgovori, stoga je raspored pravokutnih dijagrama gotovo identičan. Nisu vidljiva neka veća odstupanja od kvartilnih rangova, a tamo gdje su zabilježena, uglavnom se radi o preciznijem odgovoru na pitanje o broju stranica.

Mjere centralne tendencije

U tablici se nalaze najvažnije mjere centralne tendencije. Za dobivene podatke izračunati su srednja vrijednost, standardna devijacija, minimalno i maksimalno zabilježen podatak za svaku kategoriju.

```
require(stargazer)
stargazer(book)
```

Table 1:

Statistic	N	Sr.vrij.	St. dev.	Min	Max
student_bezNagrade	106	652.698	231.561	350	1,600
cijela.grupa_bezNagrade	105	595.610	181.193	200	1,139
samo.prijatelji_bezNagrade	102	625.990	200.998	234	1,200
student_Nagrada	106	653.368	218.169	343	1,364
cijela.grupa_Nagrada	105	635.086	194.420	300	1,200
samo.prijatelji_Nagrada	103	633.359	206.171	250	1,500
MI	106	18.297	4.917	6.500	28.000

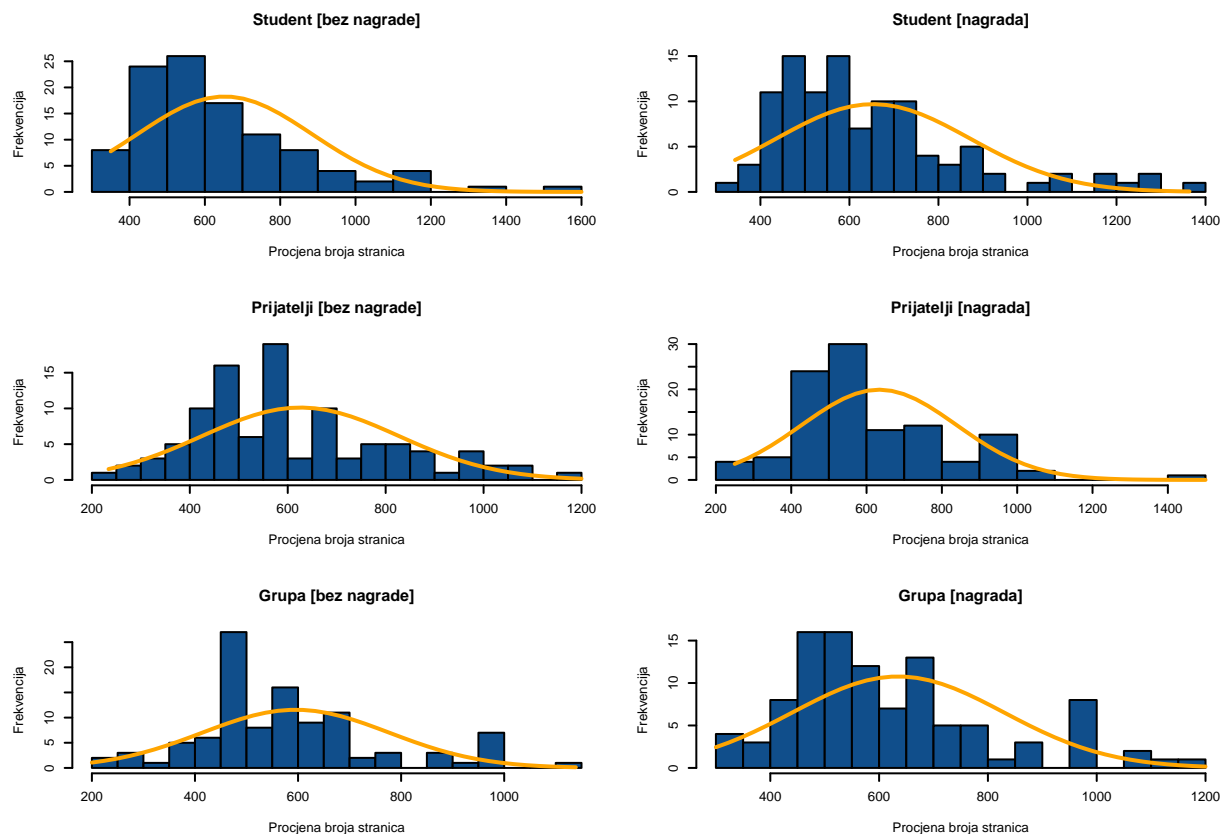
Normalnost varijabli

Analizirajući histograme i qq-grafove po pojedinim značajkama podataka, zaključili smo da se sve varijable mogu aproksimirati normalnom razdiobom. To nam omogućava raznoliku paletu testova koji pretpostavljaju normalnost podataka.

```
drawHistWithNorm <- function(variable, name,
                              xname = "Procjena broja stranica", yname = "Frekvencija",
                              histcol = "dodgerblue4", normcol = "orange",
                              numOfClasses = 16) {

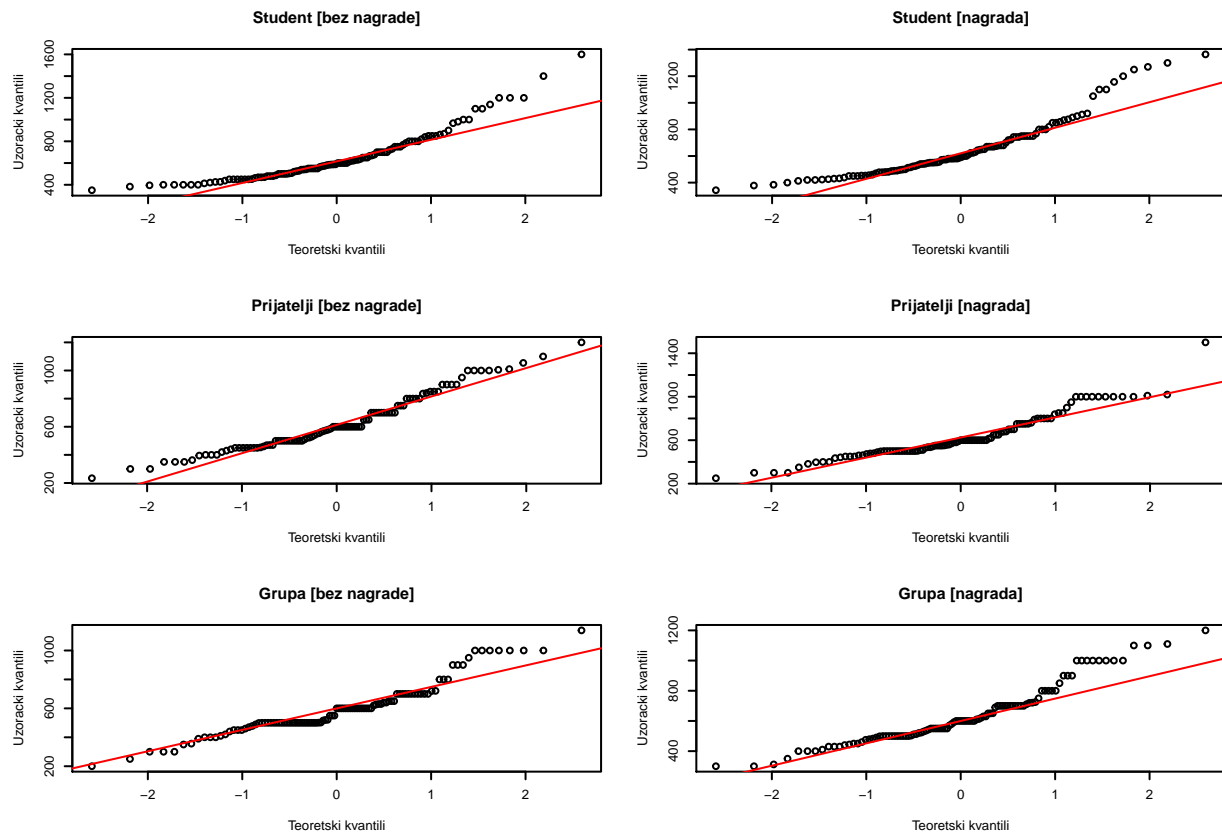
  histogram_var = variable[!is.na(variable)]
  h = hist(histogram_var
           , nclass = numOfClasses, col = histcol, main=name, xlab=xname, ylab=yname)
  xfit = seq(min(histogram_var), max(histogram_var), length = 40)
  yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
  yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
  lines(xfit, yfit, col = normcol, lwd = 2)
}
```

```
par(mfrow=c(3,2), cex=0.4)
drawHistWithNorm(book$student_bezNagrade, "Student [bez nagrade]")
drawHistWithNorm(book$student_Nagrada, "Student [nagrada]")
drawHistWithNorm(book$samo.prijatelji_bezNagrade, "Prijatelji [bez nagrade]")
drawHistWithNorm(book$samo.prijatelji_Nagrada, "Prijatelji [nagrada]")
drawHistWithNorm(book$cijela.grupa_bezNagrade, "Grupa [bez nagrade]")
drawHistWithNorm(book$cijela.grupa_Nagrada, "Grupa [nagrada]")
```



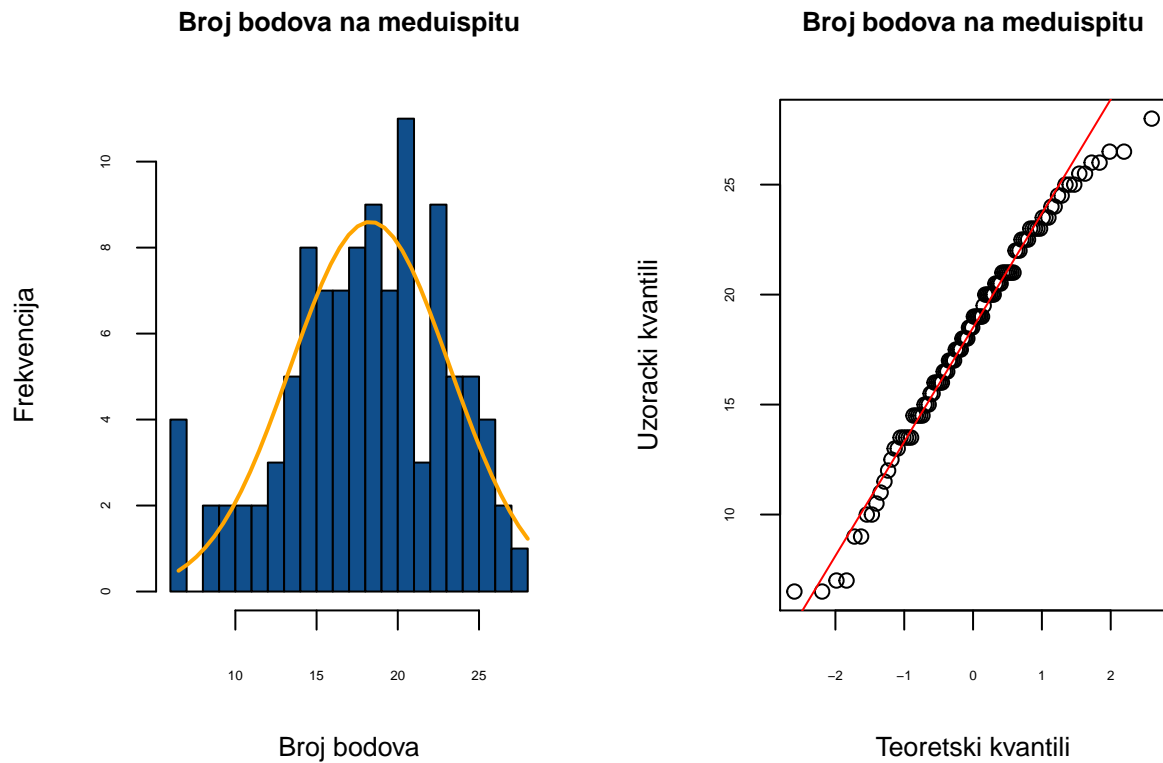
```
drawqqplot <- function(variable, name) {
  qqnorm(variable, main = name, xlab = "Teoretski kvantili", ylab = "Uzoracki kvantili")
  qqline(variable, col="red")
}
```

```
par(mfrow=c(3,2), cex = 0.4)
drawqqplot(book$student_bezNagrada, "Student [bez nagrade]")
drawqqplot(book$student_Nagrada, "Student [nagrada]")
drawqqplot(book$samo.prijatelji_bezNagrada, "Prijatelji [bez nagrade]")
drawqqplot(book$samo.prijatelji_Nagrada, "Prijatelji [nagrada]")
drawqqplot(book$cijela.grupa_bezNagrada, "Grupa [bez nagrade]")
drawqqplot(book$cijela.grupa_Nagrada, "Grupa [nagrada]")
```



Raspodjela broja bodova na međuispitu očekivano se ravna po normalnoj razdiobi.

```
par(mfrow=c(1,2), cex.axis = 0.4, cex.main = 0.8, cex.lab = 0.8)
drawHistWithNorm(book$MI, "Broj bodova na međuispitu", xname = "Broj bodova")
drawqqplot(book$MI, "Broj bodova na međuispitu")
```

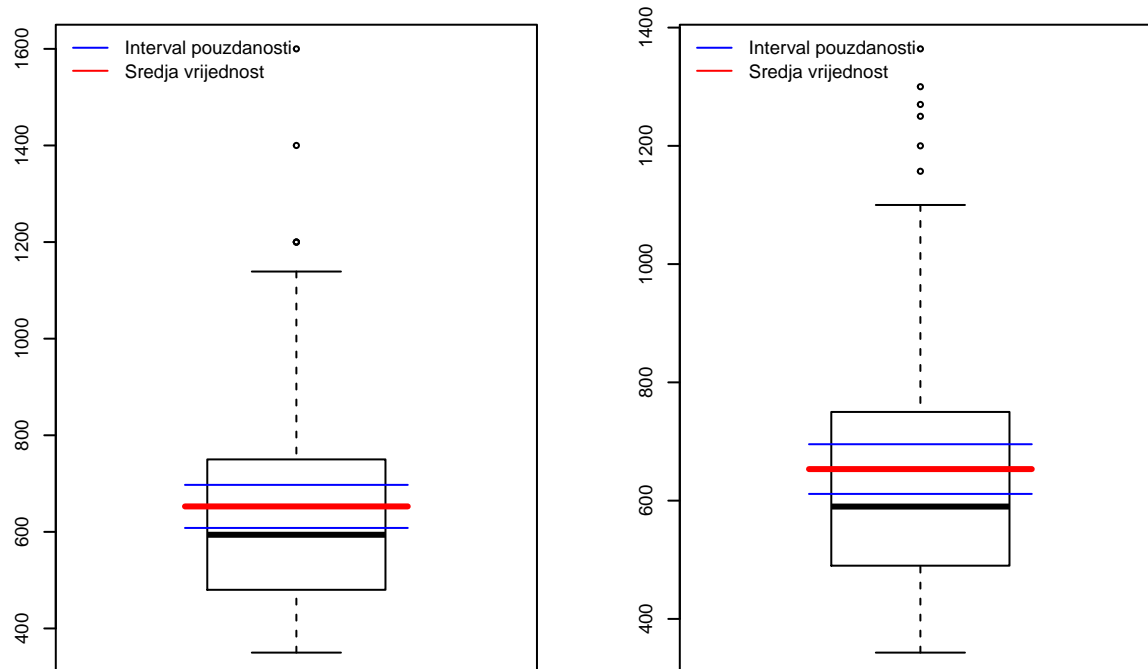


Intervali pouzdanosti

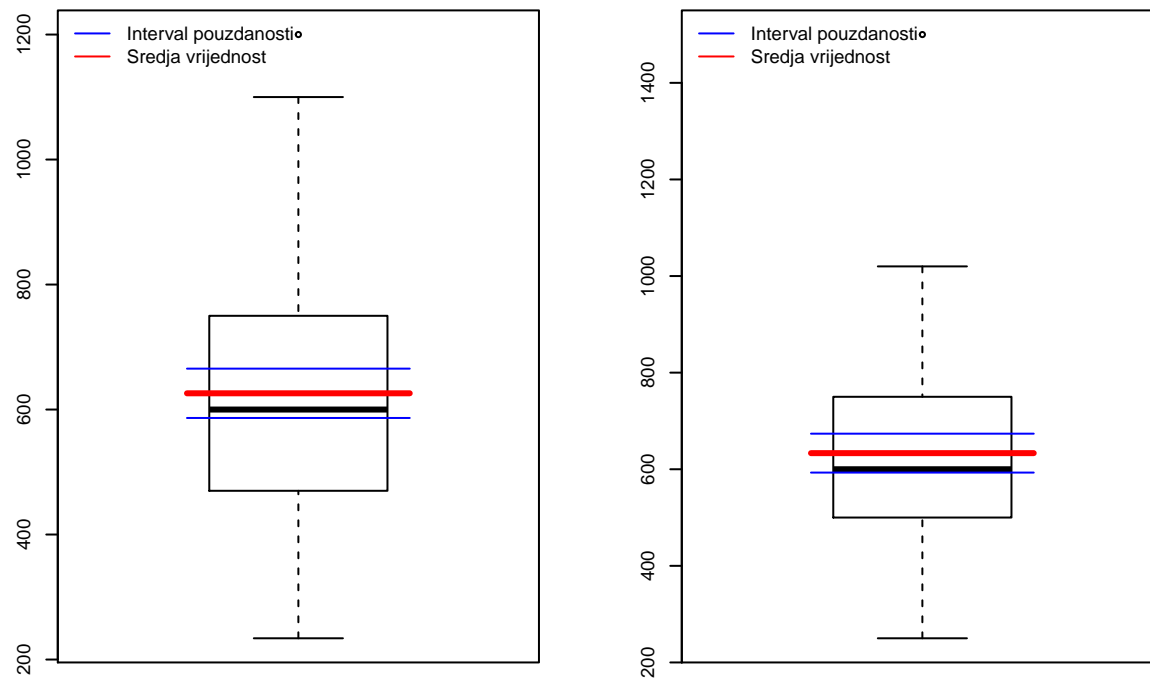
S obzirom na svako anketno pitanje, provedeni su 95%-tni intervali pouzdanosti, te su grafički uspoređeni rezultati dobiveni prije i nakon što je studentu rečeno da će za precizan odgovor dobiti nagradu.

```
drawBoxPlotWithConfidenceInterval <- function(variable, alpha = 0.05) {  
  t_conf_var = variable[!is.na(variable)]  
  x_mean = mean(variable, na.rm = T)  
  sd = sd(t_conf_var)  
  N = length(t_conf_var)  
  dof = N-1  
  t_alpha_h = qt(1-alpha/2, df = dof)  
  se = sd / sqrt(N)  
  left = x_mean - t_alpha_h * se  
  right = x_mean + t_alpha_h * se  
  boxplot(t_conf_var)  
  lines(c(0.75,1.25),c(left,left),col=4)  
  lines(c(0.75,1.25),c(x_mean,x_mean),col=2,lwd=3)  
  lines(c(0.75,1.25),c(right,right),col=4)  
  legend("topleft", c("Interval pouzdanosti", "Sredja vrijednost"),  
        lty=1,col = c(4, 2),bty ="n")  
}
```

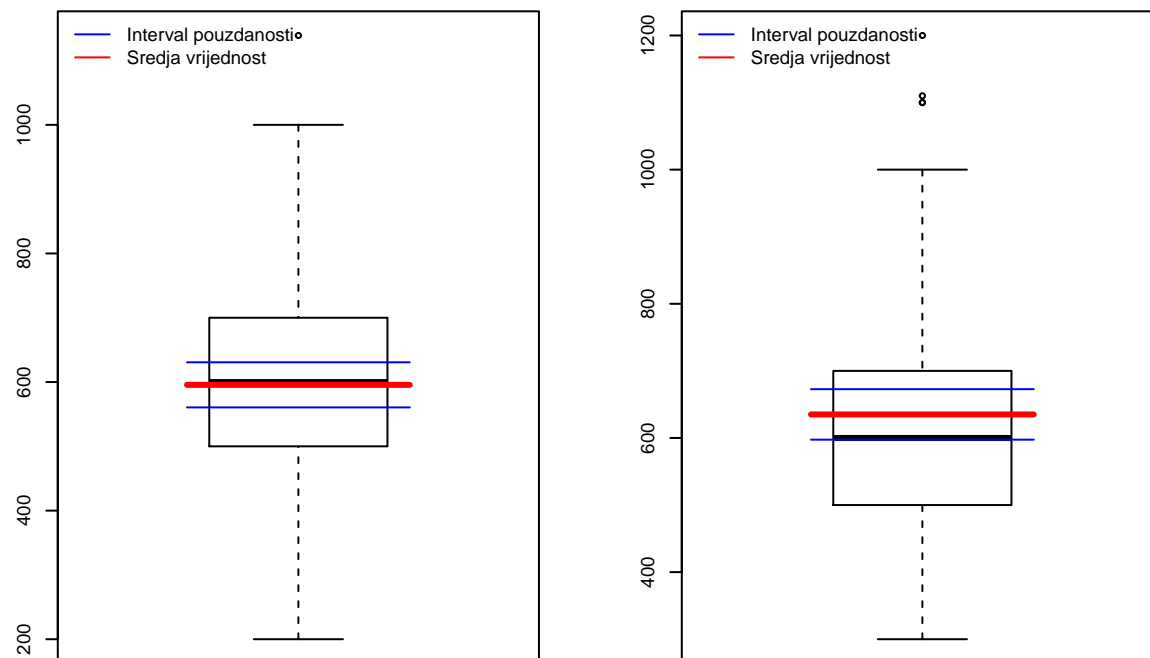
```
par(mfrow=c(1,2), cex= 0.6)  
drawBoxPlotWithConfidenceInterval(book$student_bezNagrade)  
drawBoxPlotWithConfidenceInterval(book$student_Nagrada)
```



```
par(mfrow=c(1,2), cex= 0.6)
drawBoxPlotWithConfidenceInterval(book$samo.prijatelji_bezNagrade)
drawBoxPlotWithConfidenceInterval(book$samo.prijatelji_Nagrada)
```



```
par(mfrow=c(1,2), cex= 0.6)
drawBoxPlotWithConfidenceInterval(book$cijela.grupa_bezNagrade)
drawBoxPlotWithConfidenceInterval(book$cijela.grupa_Nagrada)
```



Parovi značajki (s i bez nagrade) se gotovo ne razlikuju kod anketnog pitanja za odgovor studenta te za odgovor njegovih prijatelja, dok se u slučaju odgovora za srednju grupu interval pouzdanosti pomakao prema većem broju stranica.

Testiranje hipoteza

Budući da nam nije poznata varijanca značajki, za procjenu intervala pouzdanosti za srednju vrijednost koristili smo t-statistiku. To podrazumijeva korištenje statistike koja se ravna po Studentovoj razdiobi.

Test o uparenim podacima

Koristili smo t-statistiku za uparene podatke nad komplementarnim parovima značajki - bez nagrade i s nagradom. Za parove vlastitih odgovora te pretpostavljenih odgovora prijatelja, ne može se zaključiti ništa statistički značajno o različitosti srednjih vrijednosti. Kod para za pretpostavljeni odgovor grupe, s razinom značajnosti od 1% može se odbaciti hipoteza da su srednje vrijednosti odgovora jednake. U tom slučaju prihvaćamo alternativnu hipotezu da je srednja vrijednost odgovora za broj stranica uz nagradu veća.

```
t.test(book$student_Nagrada, book$student_bezNagrada,
       alternative = "t", conf.level = 0.99, paired = T)
```

```
##
## Paired t-test
##
## data: book$student_Nagrada and book$student_bezNagrada
## t = 0.056748, df = 105, p-value = 0.9549
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## -30.29585 31.63547
## sample estimates:
## mean of the differences
## 0.6698113
```

```
t.test(book$samo.prijatelji_Nagrada, book$samo.prijatelji_bezNagrada,
       alternative = "t", conf.level = 0.99, paired = T)
```

```
##
## Paired t-test
##
## data: book$samo.prijatelji_Nagrada and book$samo.prijatelji_bezNagrada
## t = 0.43843, df = 100, p-value = 0.662
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## -33.83880 47.40316
## sample estimates:
## mean of the differences
## 6.782178
```

```
t.test(book$cijela.grupa_Nagrada, book$cijela.grupa_bezNagrade,
       alternative = "g", conf.level = 0.99, paired = T)

##
## Paired t-test
##
## data:  book$cijela.grupa_Nagrada and book$cijela.grupa_bezNagrade
## t = 2.3635, df = 103, p-value = 0.009989
## alternative hypothesis: true difference in means is greater than 0
## 99 percent confidence interval:
##  0.006947849      Inf
## sample estimates:
## mean of the differences
##                37.93269
```

Odmak od stvarnog broja stranica

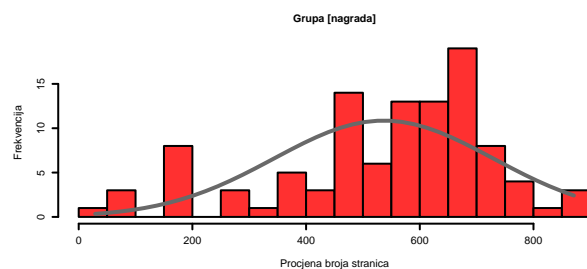
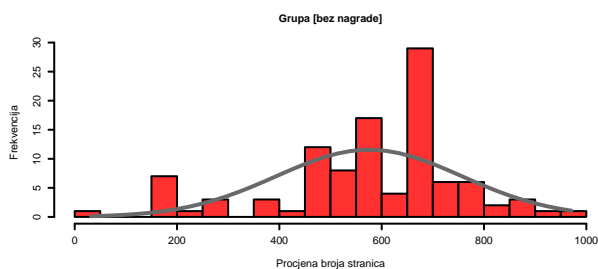
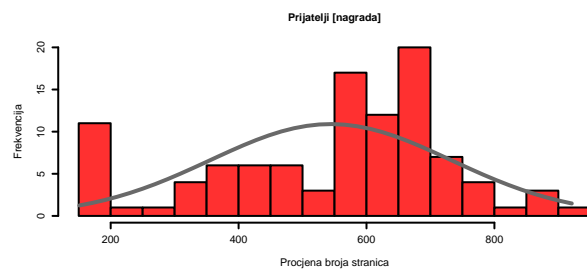
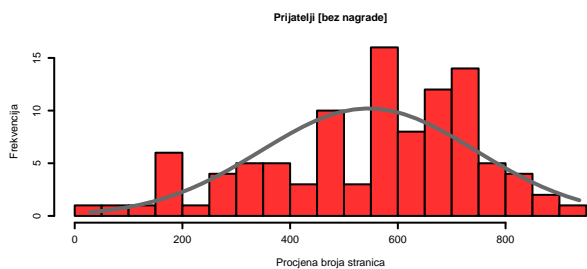
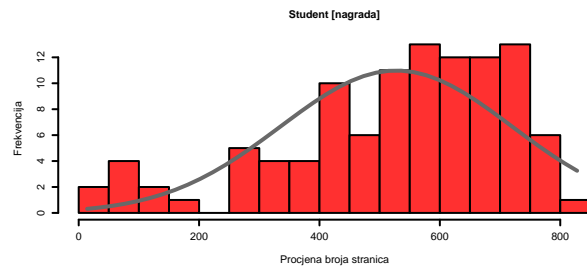
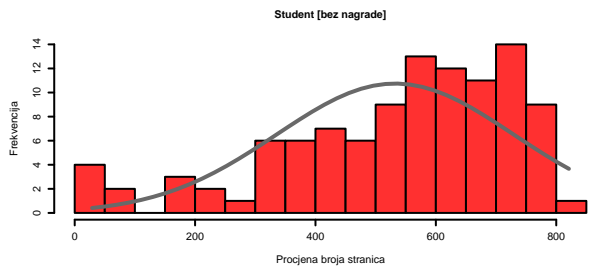
Podaci su transformirani tako da se za svaku značajku izračunao odmak od stvarne vrijednosti stranica knjige - 1171. Podaci se također ravnaju po normalnoj razdiobi budući da oduzimanje od konstante zadržava normalnost.

```
offset = book
offset$Grupa.Pred = NULL
offset$spol = NULL
offset$MI = NULL
offset = abs(CORRECT_PAGE_NO - offset)
```

```

par(mfrow=c(3,2), cex=0.3)
drawHistWithNorm(offset$student_bezNagrade, "Student [bez nagrade]"
, histcol = 'firebrick1', normcol = 'dimgray')
drawHistWithNorm(offset$student_Nagrada, "Student [nagrada]"
, histcol = 'firebrick1', normcol = 'dimgray')
drawHistWithNorm(offset$samo.prijatelji_bezNagrade, "Prijatelji [bez nagrade]"
, histcol = 'firebrick1', normcol = 'dimgray')
drawHistWithNorm(offset$samo.prijatelji_Nagrada, "Prijatelji [nagrada]"
, histcol = 'firebrick1', normcol = 'dimgray')
drawHistWithNorm(offset$cijela.grupa_bezNagrade, "Grupa [bez nagrade]"
, histcol = 'firebrick1', normcol = 'dimgray')
drawHistWithNorm(offset$cijela.grupa_Nagrada, "Grupa [nagrada]"
, histcol = 'firebrick1', normcol = 'dimgray')

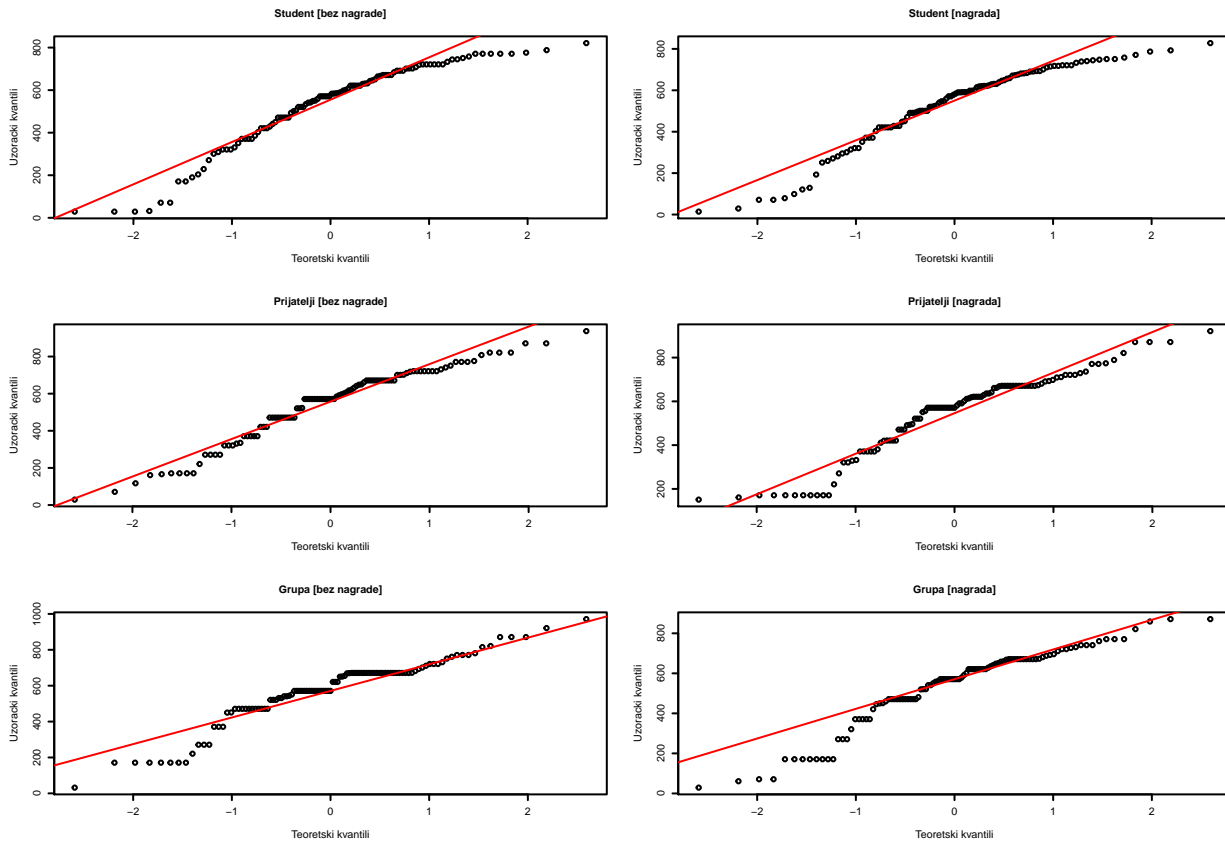
```



```

par(mfrow=c(3,2), cex = 0.3)
drawqqplot(offset$student_bezNagrade, "Student [bez nagrade]")
drawqqplot(offset$student_Nagrada, "Student [nagrada]")
drawqqplot(offset$samo.prijatelji_bezNagrade, "Prijatelji [bez nagrade]")
drawqqplot(offset$samo.prijatelji_Nagrada, "Prijatelji [nagrada]")
drawqqplot(offset$cijela.grupa_bezNagrade, "Grupa [bez nagrade]")
drawqqplot(offset$cijela.grupa_Nagrada, "Grupa [nagrada]")

```



Svi parovi značajki testirani su tako da je alternativna hipoteza ona za koju je odmak uz ponuđenu nagradu manji od odmaka bez nagrade. Uz nivo značajnosti od 5% osnovna hipoteza je odbačena samo za pretpostavku o mišljenju cijele grupe. Za druga 2 slučaja, nagrada ne utječe pretjerano na rezultate o broju stranica.

```
t.test(offset$student_Nagrada, offset$student_bezNagrada, alternative = "less", conf.level = 0.95, pair

##
## Paired t-test
##
## data: offset$student_Nagrada and offset$student_bezNagrada
## t = -0.45937, df = 105, p-value = 0.3235
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 12.39723
## sample estimates:
## mean of the differences
##      -4.745283

t.test(offset$samo.prijatelji_Nagrada, offset$samo.prijatelji_bezNagrada, alternative = "less", conf.le

##
## Paired t-test
##
## data: offset$samo.prijatelji_Nagrada and offset$samo.prijatelji_bezNagrada
## t = -0.062287, df = 100, p-value = 0.4752
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 21.59065
## sample estimates:
## mean of the differences
##      -0.8415842

t.test(offset$cijela.grupa_Nagrada, offset$cijela.grupa_bezNagrada, alternative = "less", conf.level =

##
## Paired t-test
##
## data: offset$cijela.grupa_Nagrada and offset$cijela.grupa_bezNagrada
## t = -2.3353, df = 103, p-value = 0.01073
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -10.81166
## sample estimates:
## mean of the differences
##      -37.375
```

Neparametarski testovi

Iz qqplotova vidljivo je da se ekstrahirani podaci ne ravnaju potpuno po normalnog razdiobi. Zbog te pretpostavke korištenje t-testa može biti statistički potpuno neispravno. Iako sa manjom snagom jednostavni neparametarski postupci poslužiti će za testiranje istih hipoteza bez pretpostavke normalnosti. Pokazatelj koji se testira ovaj puta je medijan, naspram srednje vrijednosti koja se testirala u t-testu.

```
wilcox.test(offset$student_bezNagrada, offset$student_Nagrada, paired = T)

##
## Wilcoxon signed rank test with continuity correction
##
## data: offset$student_bezNagrada and offset$student_Nagrada
## V = 2209.5, p-value = 0.6461
## alternative hypothesis: true location shift is not equal to 0

wilcox.test(offset$samo.prijatelji_bezNagrada, offset$samo.prijatelji_Nagrada, paired = T)

##
## Wilcoxon signed rank test with continuity correction
##
## data: offset$samo.prijatelji_bezNagrada and offset$samo.prijatelji_Nagrada
## V = 2020.5, p-value = 0.2941
## alternative hypothesis: true location shift is not equal to 0

wilcox.test(offset$cijela.grupa_Nagrada, offset$cijela.grupa_bezNagrada, paired = T, alternative = "l")

##
## Wilcoxon signed rank test with continuity correction
##
## data: offset$cijela.grupa_Nagrada and offset$cijela.grupa_bezNagrada
## V = 1071, p-value = 0.002744
## alternative hypothesis: true location shift is less than 0
```

Naposlijetku je moguće izvući isti zaključak, jedina signifikantna razlika krije se u podacima za procjenu pogotka cijele grupe. Uz razinu značajnosti od 1% moguće je odbaciti nultu hipotezu o jednakosti medijana za procjenu pogotka cijele grupe prije i nakon nagrade. Izgledno je kako ispitanici nakon što im je rečeno da postoji nagrada preciznije određuju konačni cilj, odnosno iako pokušavaju procijeniti što će ostali reći, zapravo se približavaju rješenju.

Usporedba srednjih vrijednosti predviđanja kod muškaraca i žena

Usporedba jednakosti srednjih vrijednosti kod ženskih i muških ispitanika. Potrebno je procijeniti postoji li statistički značajna razlika kojom bi zaključili da se kod procjene broja stranica knjige razlikuju ispitanici različitog spola

```
muski = which(book$spol == 'M')
zenski = which(book$spol == 'F')
muskiBezNagrade = book[muski,]$student_bezNagrade
zenskiBezNagrade = book[zenski,]$student_bezNagrade

t.test(muskiBezNagrade, zenskiBezNagrade, alternative = 't', var.equal = F)
```

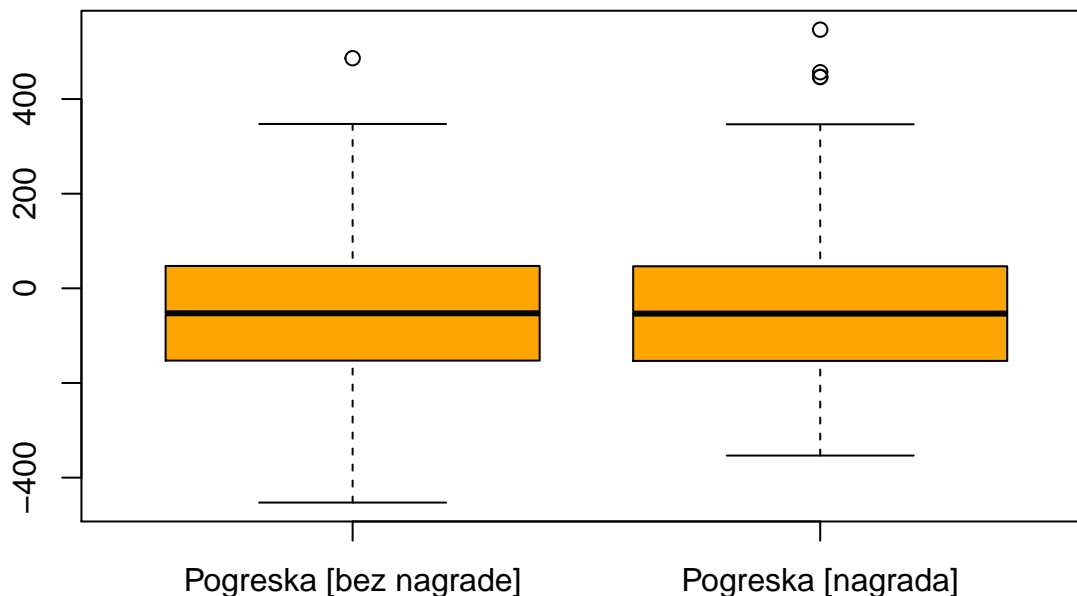
```
##
##  Welch Two Sample t-test
##
## data:  muskiBezNagrade and zenskiBezNagrade
## t = 1.5304, df = 44.616, p-value = 0.133
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -23.33933 170.89421
## sample estimates:
## mean of x mean of y
##  669.4024  595.6250
```

Uz razinu značajnosti od 13.3% imati ćemo dovoljno dokaza da se odbaci hipoteza o jednakosti srednjih vrijednosti procjena muške i ženske populacije.

Analiza procjene broja stranica za cijelu ispitnu grupu

Provedena je analiza rezultata kvalitete pogađanja broja stranica za cijelu grupu. S obzirom na to da je cilj te kategorije što bolje ocijeniti prosječan odgovor cijele grupe, bilo je potrebno transformirati podatke. Pronađena je srednja vrijednost ukupnih pogađanja kod vlastitih odgovora te je oduzeta od pojedinačnih pogađanja za grupu.

```
transformGuesses <- function(groupGuesses, studentGuesses) {  
  groupGuessesOffset =  
    groupGuesses[!is.na(groupGuesses)] - mean(studentGuesses[!is.na(studentGuesses)])  
  return(groupGuessesOffset)  
}  
  
guessesOffsetBN = transformGuesses(book$cijela.grupa_bezNagrade, book$student_bezNagrade)  
guessesOffsetN = transformGuesses(book$cijela.grupa_Nagrada, book$student_Nagrada)  
  
boxplot(guessesOffsetBN, guessesOffsetN, col = 'orange',  
        names = c("Pogreska [bez nagrade]", "Pogreska [nagrada]"))
```




```
t.test(guessesOffsetBN, mu=0, alternative = "t")

##
## One Sample t-test
##
## data: guessesOffsetBN
## t = -3.2285, df = 104, p-value = 0.001665
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -92.15389 -22.02329
## sample estimates:
## mean of x
## -57.08859
```

```
t.test(guessesOffsetN, mu=0, alternative = "t")

##
## One Sample t-test
##
## data: guessesOffsetN
## t = -0.96357, df = 104, p-value = 0.3375
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -55.90731 19.34289
## sample estimates:
## mean of x
## -18.28221
```

Linearna regresija

```
drawRegression <- function(var1, var2, name, xname, yname, col = "darkmagenta") {  
  plot(var1, var2, pch = 16, cex = 1.3, col = "darkmagenta",  
        main = name, xlab = xname, ylab = yname)  
  
  abline(lm(var1 ~ var2))  
}
```

```
guessesOffsetBN = abs(guessesOffsetBN)  
guessesOffsetN = abs(guessesOffsetN)
```

```
modelBN = lm(book$MI[complete.cases(book$cijela.grupa_bezNagrada)] ~ guessesOffsetBN)  
summary(modelBN)
```

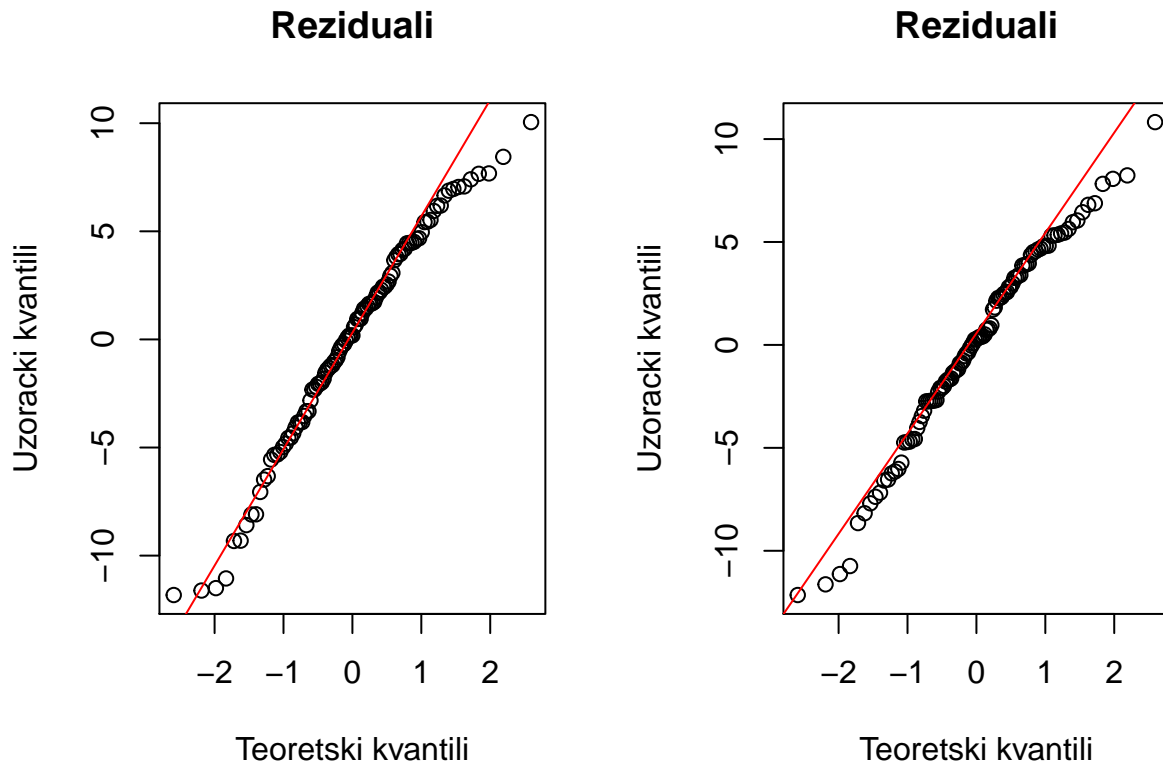
```
##  
## Call:  
## lm(formula = book$MI[complete.cases(book$cijela.grupa_bezNagrada)] ~  
##     guessesOffsetBN)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -11.8220  -3.3220   0.1833   3.9442  10.0507   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    17.915489   0.825017   21.71  <2e-16 ***  
## guessesOffsetBN  0.002662   0.004362    0.61   0.543      
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 4.948 on 103 degrees of freedom  
## Multiple R-squared:  0.003604,    Adjusted R-squared:  -0.006069   
## F-statistic: 0.3726 on 1 and 103 DF,  p-value: 0.5429
```

```
modelN = lm(book$MI[complete.cases(book$cijela.grupa_Nagrada)] ~ guessesOffsetN)  
summary(modelN)
```

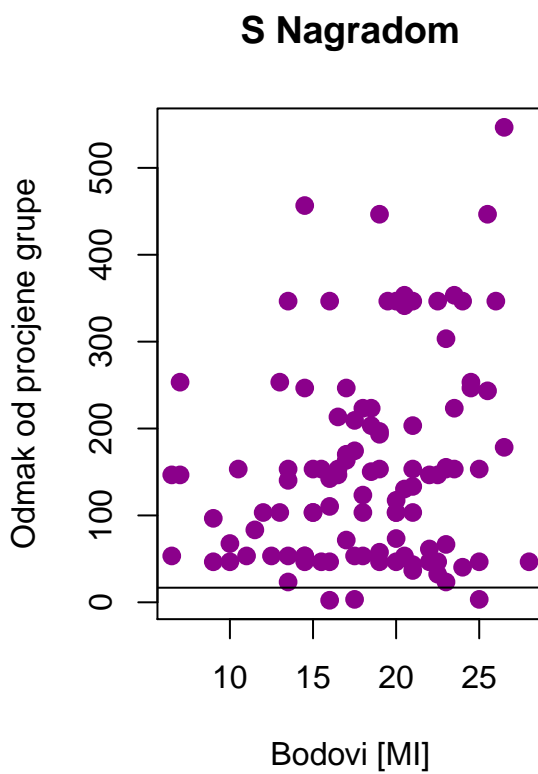
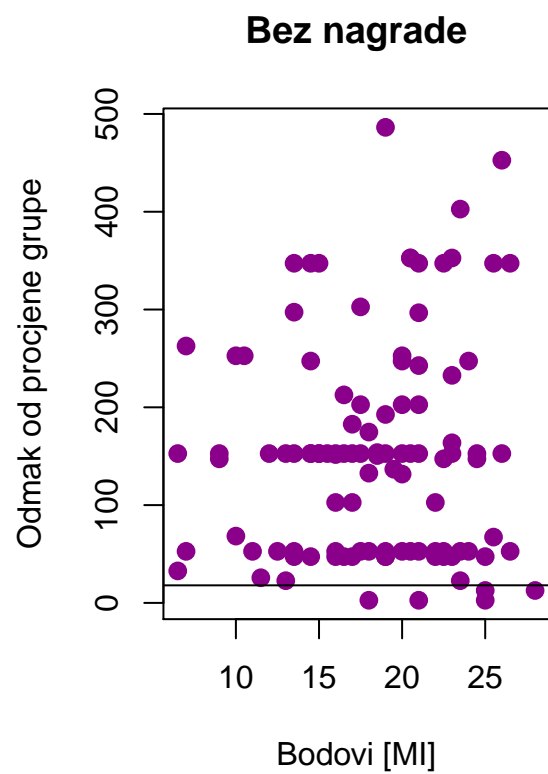
```
##  
## Call:  
## lm(formula = book$MI[complete.cases(book$cijela.grupa_Nagrada)] ~  
##     guessesOffsetN)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -12.1492  -2.7183   0.2817   3.8541  10.8229   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    16.73222   0.78511   21.312  <2e-16 ***  
## guessesOffsetN  0.00954   0.00404    2.362  0.0201 *      
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##
```

```
## Residual standard error: 4.778 on 103 degrees of freedom
## Multiple R-squared:  0.05136,    Adjusted R-squared:  0.04215
## F-statistic: 5.577 on 1 and 103 DF,  p-value: 0.02008
```

```
par(mfrow=c(1,2))
drawqqplot(modelBN$residuals, name="Reziduali")
drawqqplot(modelN$residuals, name="Reziduali")
```



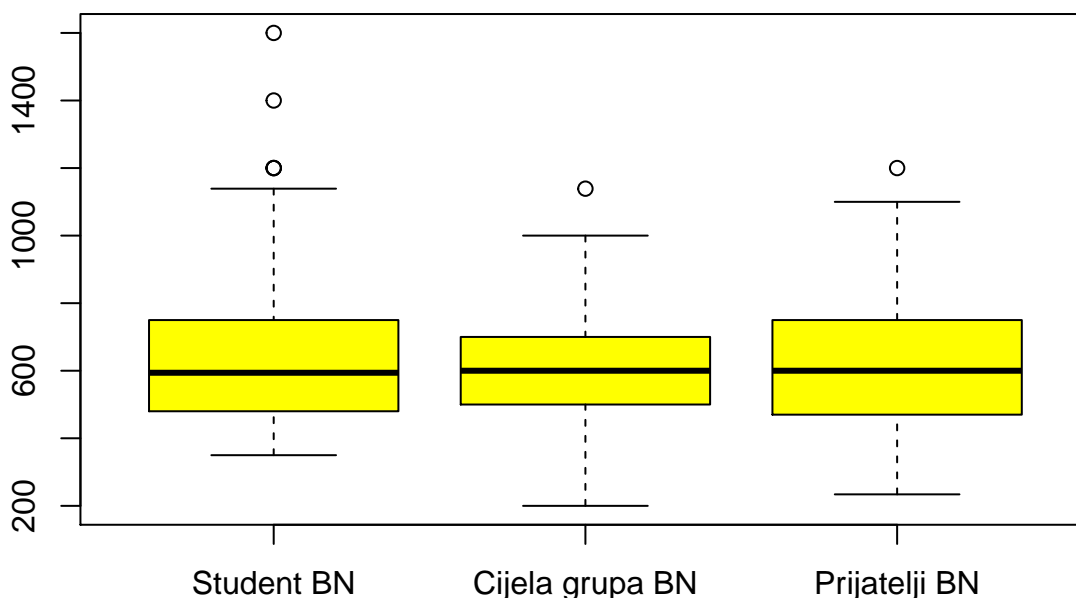
```
par(mfrow=c(1,2))
drawRegression(book$MI[complete.cases(book$cijela.grupa_bezNagrada)], guessesOffsetBN,
  xname="Bodovi [MI]", yname="Odmak od procjene grupe", name="Bez nagrade")
drawRegression(book$MI[complete.cases(book$cijela.grupa_Nagrada)], guessesOffsetN,
  xname="Bodovi [MI]", yname="Odmak od procjene grupe", name="S Nagradom")
```



ANOVA

Proveli smo analizu varijanci u 2 slučaja - za situaciju s i bez nagrade. Htjeli smo testirati jesu li srednje vrijednosti iste u slučaju kad student govori za sebe, cijelu grupu i svoje prijatelje. ANOVA za uporabu ima preduvjete da su svi uzorci normalno distribuirani (pokazano prije) te da su sve varijance jednake. Jednakost varijanci provodi se Bartlettovim testom.

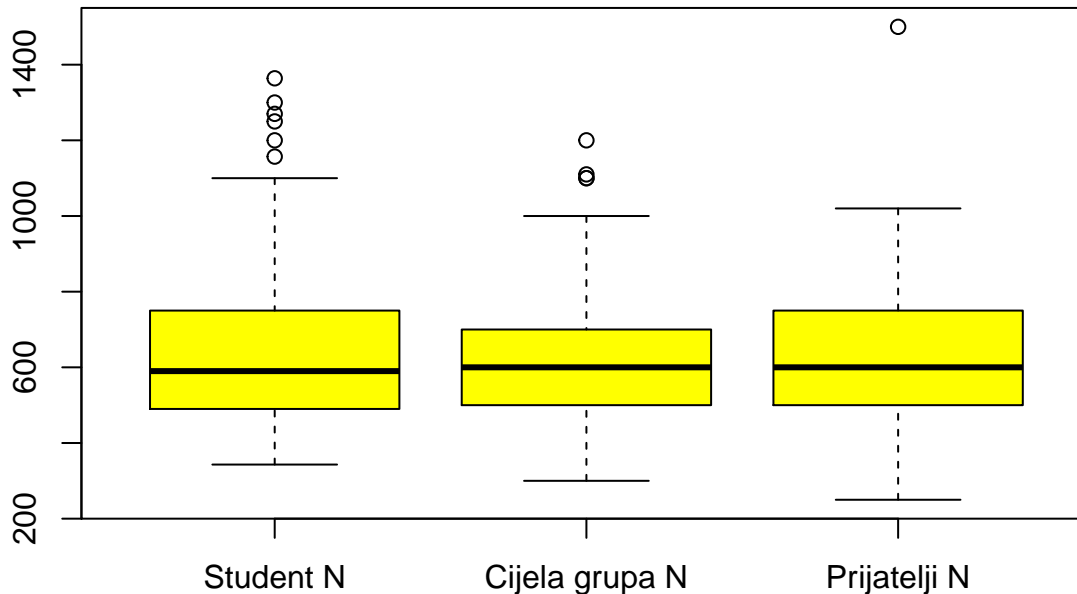
```
anovaTestingData <- function(a, b, c, aname, bname, cname) {  
  a = a[!is.na(a)]  
  b = b[!is.na(b)]  
  c = c[!is.na(c)]  
  boxplot(a, b, c, names = c(aname, bname, cname), col = "yellow")  
  values = c()  
  values = append(values, a)  
  values = append(values, b)  
  values = append(values, c)  
  labels = rep(c('a', 'b', 'c'), c(length(a), length(b), length(c)))  
  return(data.frame(values, labels))  
}  
  
anovaModelBN = anovaTestingData(book$student_bezNagrade,  
                                book$cijela.grupa_bezNagrade,  
                                book$samo.prijatelji_bezNagrade,  
                                "Student BN", "Cijela grupa BN", "Prijatelji BN")
```



```
bartlett.test(anovaModelBN$values~anovaModelBN$labels)  
  
##  
## Bartlett test of homogeneity of variances  
##  
## data: anovaModelBN$values by anovaModelBN$labels  
## Bartlett's K-squared = 6.3523, df = 2, p-value = 0.04175
```

U prvom testu ANOVU nije bilo moguće provesti budući da pretpostavka o jednakosti varijanci nije bila zadovoljena. P-vrijednost s Bartlettovim testom iznosila je 4.175%, a to nije bilo dovoljno da pretpostavimo da su varijance jednake uz razinu značajnosti od 5%.

```
anovaModelN = anovaTestingData(book$student_Nagrada,
                                book$cijela.grupa_Nagrada,
                                book$samo.prijatelji_Nagrada,
                                "Student N", "Cijela grupa N", "Prijatelji N")
```



```
bartlett.test(anovaModelN$values~anovaModelN$labels)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: anovaModelN$values by anovaModelN$labels
## Bartlett's K-squared = 1.3793, df = 2, p-value = 0.5017
```

```
anova(lm(anovaModelN$values~anovaModelN$labels))
```

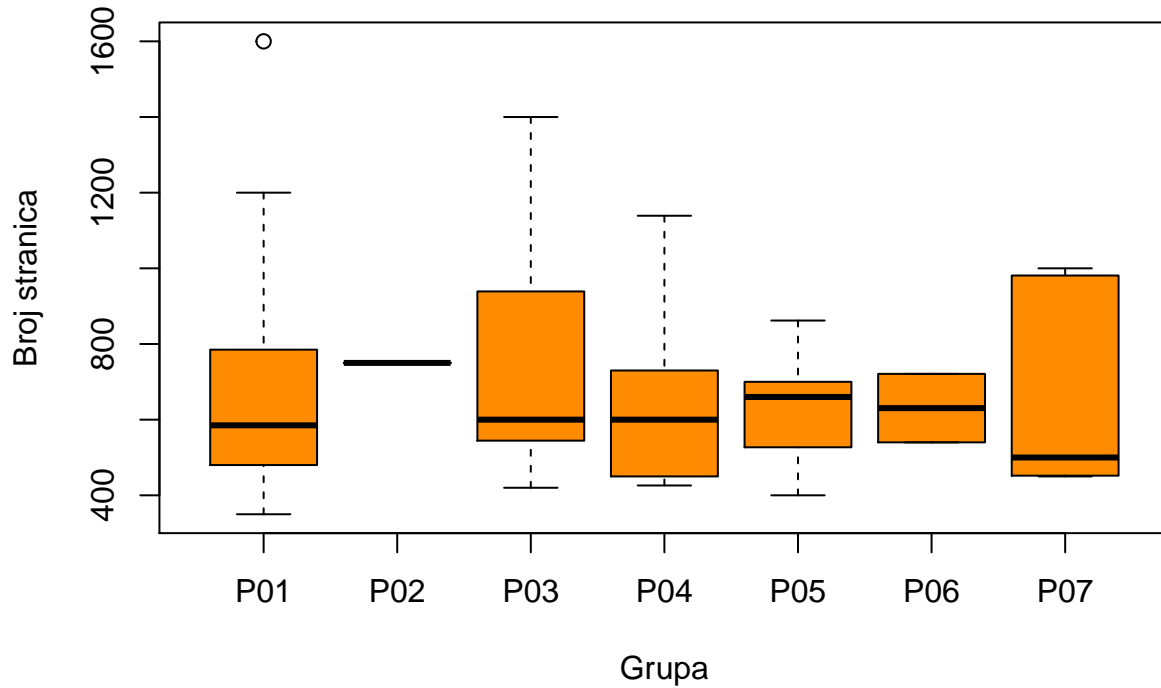
```
## Analysis of Variance Table
##
## Response: anovaModelN$values
##              Df    Sum Sq Mean Sq F value Pr(>F)
## anovaModelN$labels  2     25870    12935  0.3033  0.7386
## Residuals          311 13264541    42651
```

U drugom slučaju, p-vrijednost Bartlettovog testa iznosila je 50.17% te se ANOVA mogla provesti. P-vrijednost je iznosila 73.86% u korist hipoteze da su srednje vrijednosti uzoraka iste što se može uvidjeti iz boxplota tih uzoraka.

Sljedeće što je ispitano je ANOVA za odgovor studenta bez nagrade po grupama. Uz p-vrijednost koja podržava H0 hipotezu, nije bilo dovoljno dokaza da se utvrdi razlika između odgovora o broju stranica.

```
boxplot(book$student_bezNagrade ~ book$Grupa.Pred, col="darkorange", main="Odgovori studenata po grupama")
```

Odgovori studenata po grupama



```
library(plyr)
grps = count(book$Grupa.Pred)
grps = grps[grps$freq >= 2,]
filtered2OrMore = book[book$Grupa.Pred %in% grps$x,]
bartlett.test(filtered2OrMore$student_bezNagrade ~ filtered2OrMore$Grupa.Pred)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: filtered2OrMore$student_bezNagrade by filtered2OrMore$Grupa.Pred
## Bartlett's K-squared = 7.9217, df = 5, p-value = 0.1606
```

```
model = lm(filtered2OrMore$student_bezNagrade ~ filtered2OrMore$Grupa.Pred)
anova(model)
```

```
## Analysis of Variance Table
##
## Response: filtered2OrMore$student_bezNagrade
##              Df Sum Sq Mean Sq F value Pr(>F)
## filtered2OrMore$Grupa.Pred  5  111820    22364  0.4019  0.8465
## Residuals                99 5508778    55644
```

Neparametarski test

Kruskal - Wallisov test je kao neparametarski test potvrdio da nema razlika između grupa u odgovorima.

```
kruskal.test(student_bezNagrade ~ Grupa.Pred, data = book)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: student_bezNagrade by Grupa.Pred  
## Kruskal-Wallis chi-squared = 1.4695, df = 6, p-value = 0.9615
```

Ovisnost procijenjene vjerojatnosti prolaska na MI-u o veličini skupa za učenje

Iskorištena je metoda logističke regresije kako bi se procijenila vjerojatnost ostvarivanja bodovnog praga od 15 bodova (50%) na ispitu. Pomoću stroja potpornih vektora izvršena je klasifikacija studenata temeljena na svim popisanim značajkama u skupu podataka. U nastavku je prikazan graf koji ilustrira ovisnost veličine skupa za učenje o sigurnosti procjene. Na temelju brojnih oscilacija u iscrtanoj krivulji, mogu se primijetiti naznake kaotičnosti podataka. Kada bi postojala neka zakonitost, dobivena krivulja bila bi rastuća bez velikih odstupanja.

```
probabilityPlot = read.csv("ucenje.csv", header = T)  
plot(probabilityPlot, xlab="Velicina uzorka", ylab="Aproksimirana vjerojatnost prolaska",  
      pch=20)  
lines(probabilityPlot, col="blue")
```

