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
```

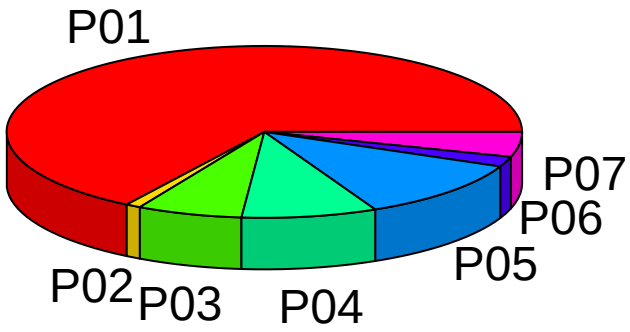
Podaci su učitani iz csv datoteke sa svim navedenim značajkama.

```
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 po grupama")
```

Raspodjela po grupama




```
boxplot(book$student_bezNagrade, book$cijela.grupa_bezNagrade, book$samo.prijatelji_bezNagrade,
        book$student_Nagrada, book$cijela.grupa_Nagrada, book$samo.prijatelji_Nagrada)
```

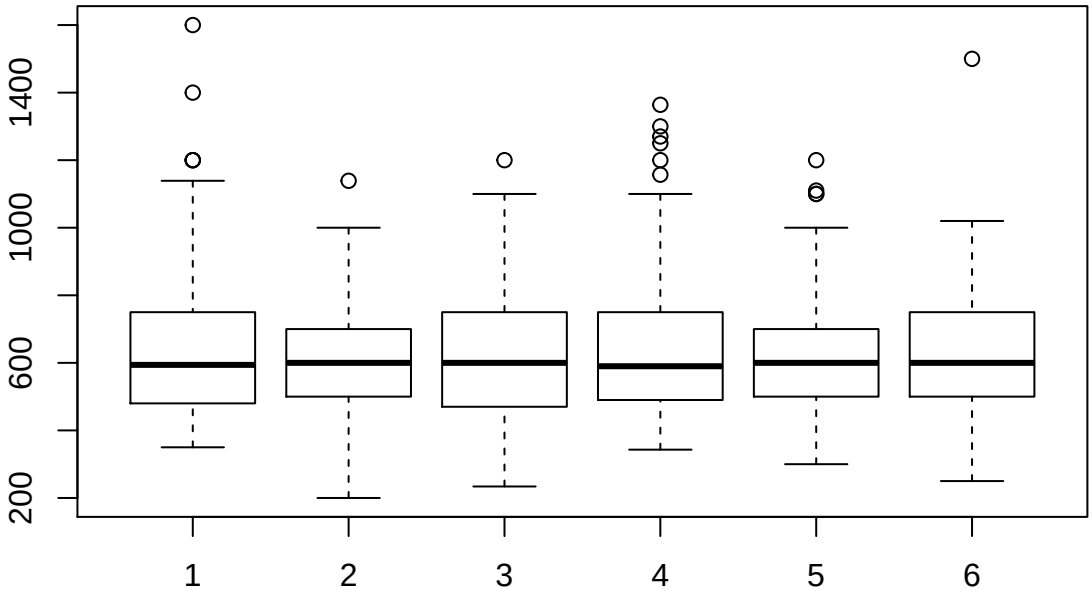


Table 1:

Statistic	N	Mean	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

```
require(stargazer)
```

```
## Loading required package: stargazer
```

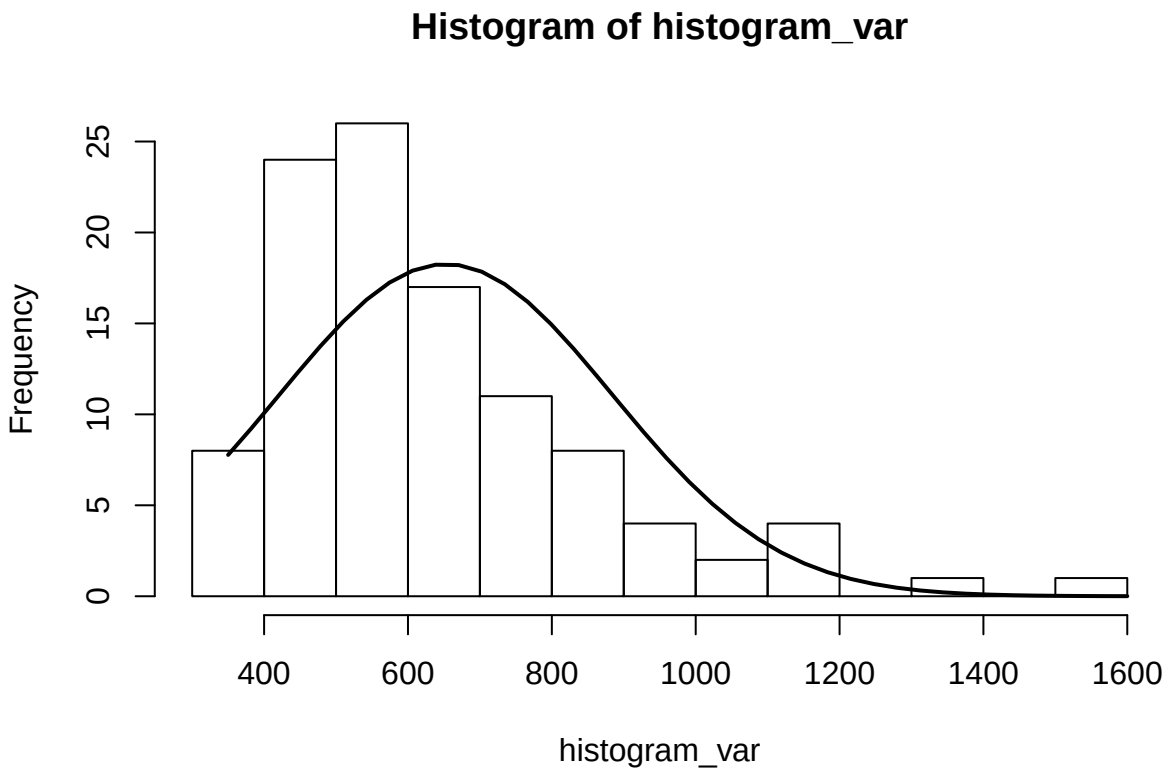
```
##
## Please cite as:

##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.1. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(book)
```

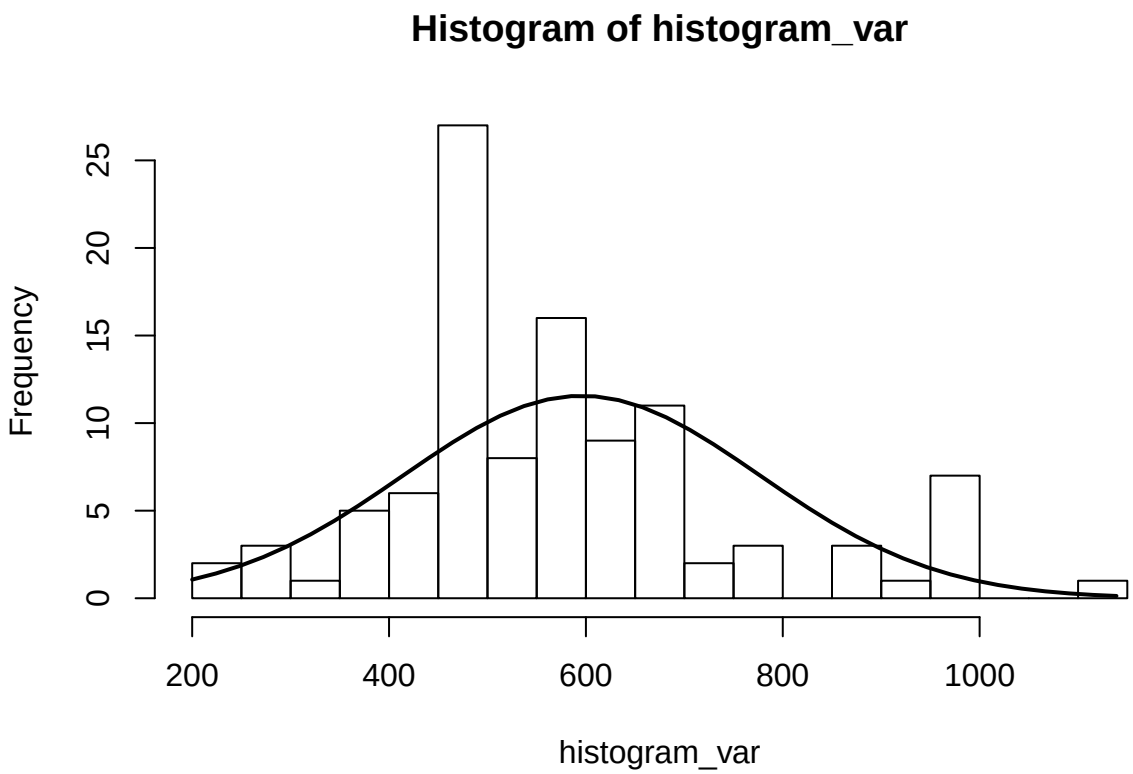
```
##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvar
## % Date and time: Wed, May 23, 2018 - 20:31:40
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Statistic & \multicolumn{1}{c}{N} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multi
## \hline \\[-1.8ex]
## 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 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

Mjere centralne tendencije

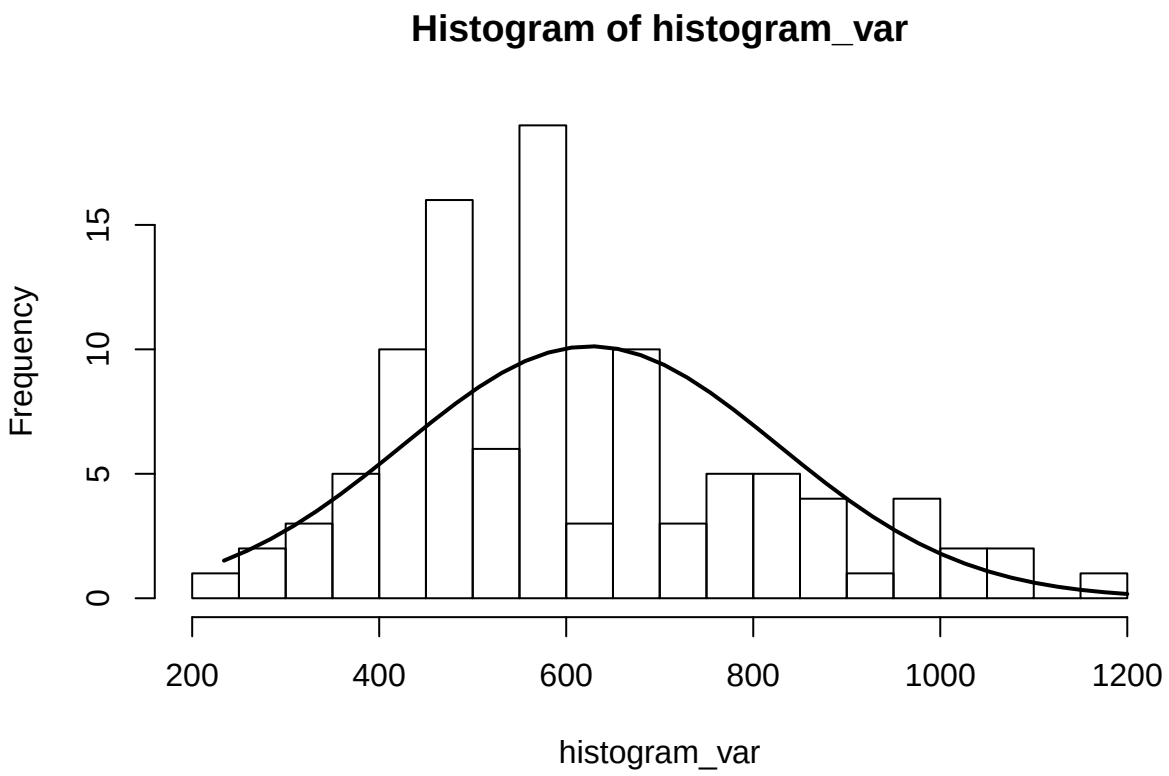
```
meanSBN = mean(book$student_bezNagrade[!is.na(book$student_bezNagrade)])
meanGBN = mean(book$cijela.grupa_bezNagrade[!is.na(book$cijela.grupa_bezNagrade)])
meanPBN = mean(book$samo.prijatelji_bezNagrade[!is.na(book$samo.prijatelji_bezNagrade)])
meanSN = mean(book$student_Nagrada[!is.na(book$student_Nagrada)])
meanGN = mean(book$cijela.grupa_Nagrada[!is.na(book$cijela.grupa_Nagrada)])
meanPN = mean(book$samo.prijatelji_Nagrada[!is.na(book$samo.prijatelji_Nagrada)])

medSBN = median(book$student_bezNagrade[!is.na(book$student_bezNagrade)])
medGBN = median(book$cijela.grupa_bezNagrade[!is.na(book$cijela.grupa_bezNagrade)])
medPBN = median(book$samo.prijatelji_bezNagrade[!is.na(book$samo.prijatelji_bezNagrade)])
medSN = median(book$student_Nagrada[!is.na(book$student_Nagrada)])
medGN = median(book$cijela.grupa_Nagrada[!is.na(book$cijela.grupa_Nagrada)])
medPN = median(book$samo.prijatelji_Nagrada[!is.na(book$samo.prijatelji_Nagrada)])

histogram_var = book$student_bezNagrade
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



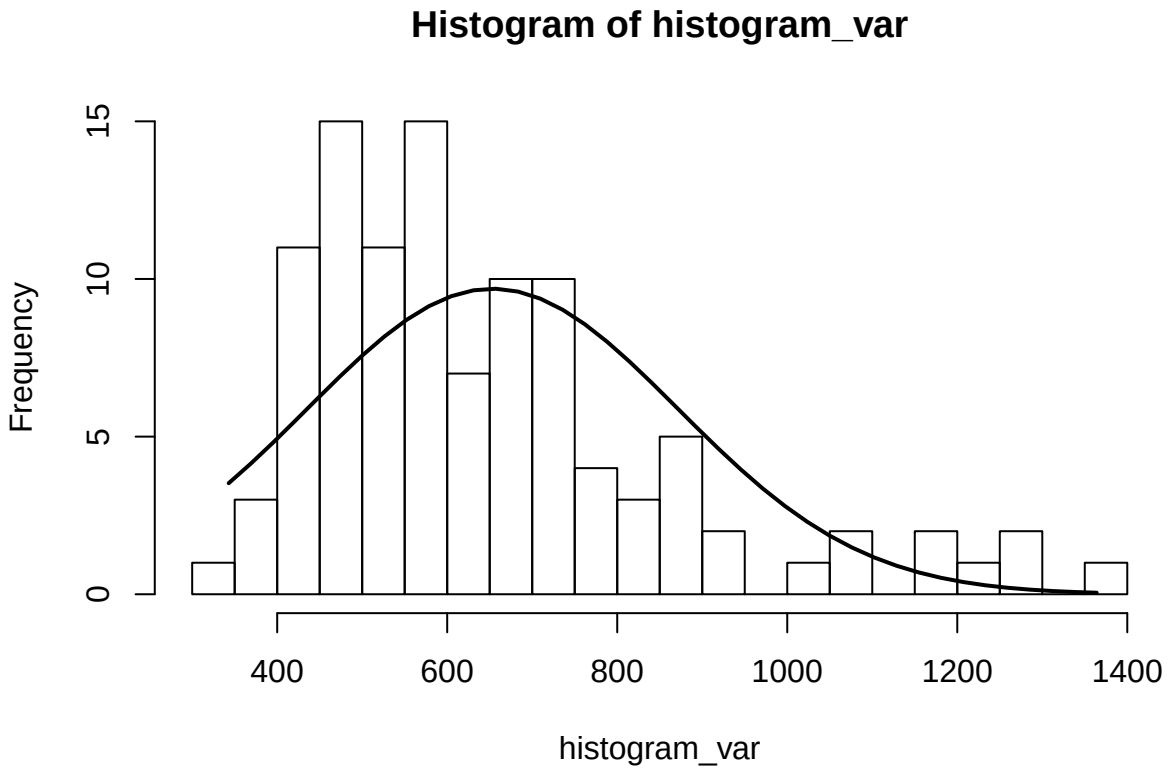

```
histogram_var = book$cijela.grupa_bezNagrade
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



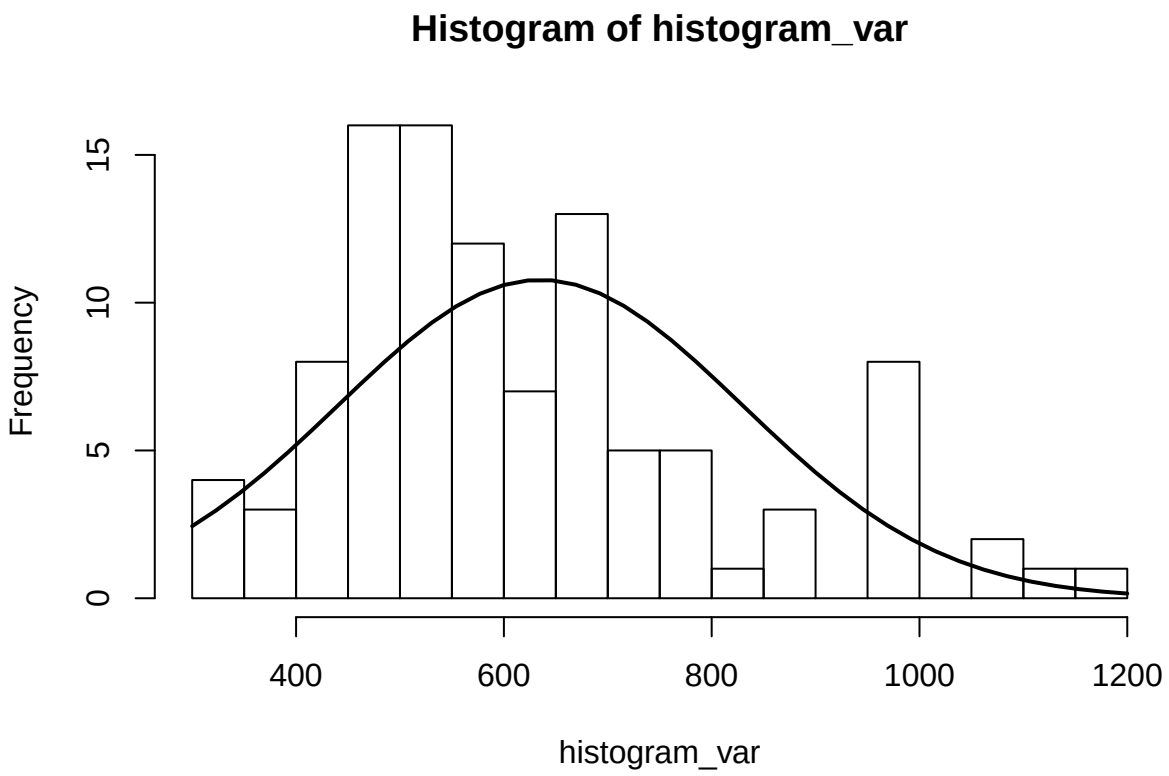

```
histogram_var = book$samo.prijatelji_bezNagrade
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```




```
histogram_var = book$student_Nagrada
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



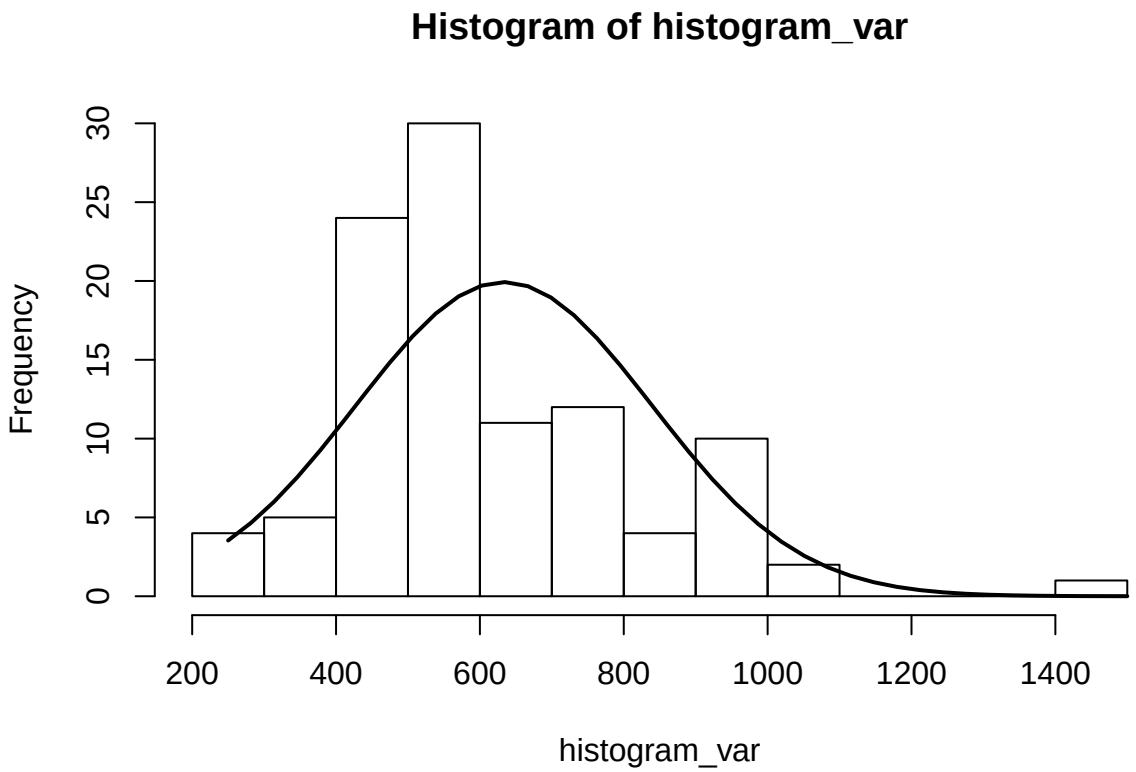

```
histogram_var = book$cijela.grupa_Nagrada
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



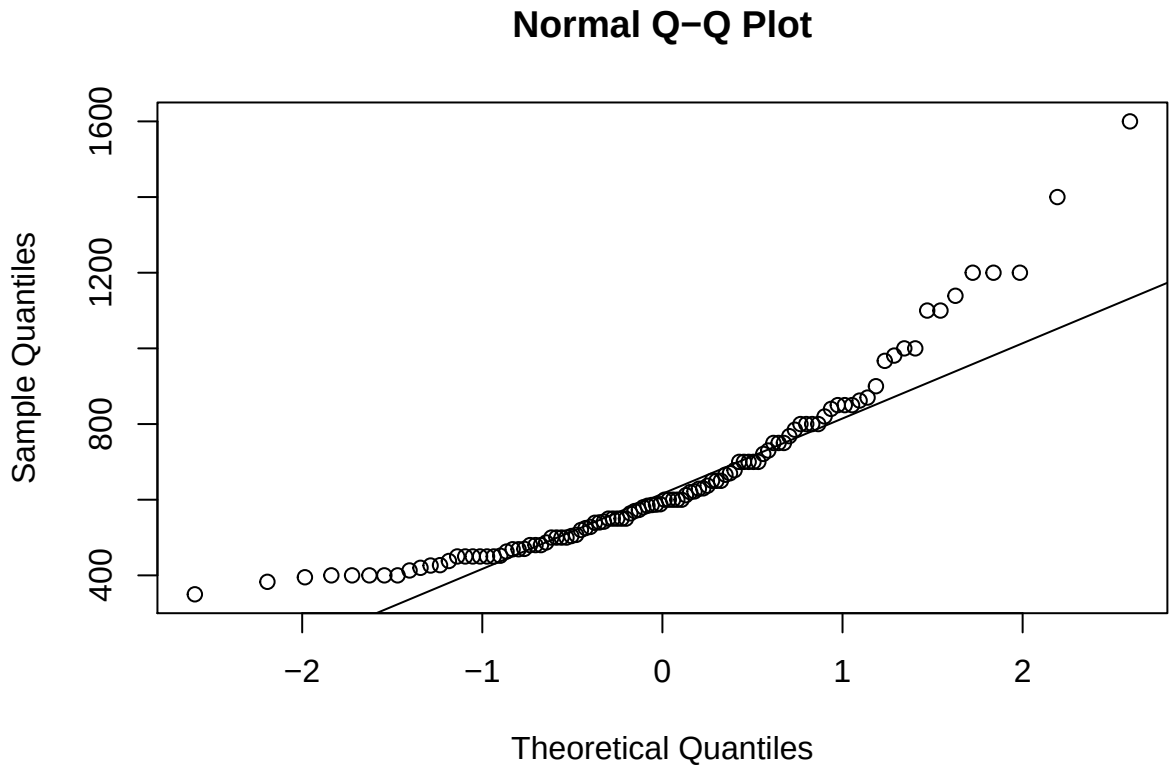

```
histogram_var = book$samo.prijatelji_Nagrada
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
min(histogram_var)
```

```
## [1] 250
```

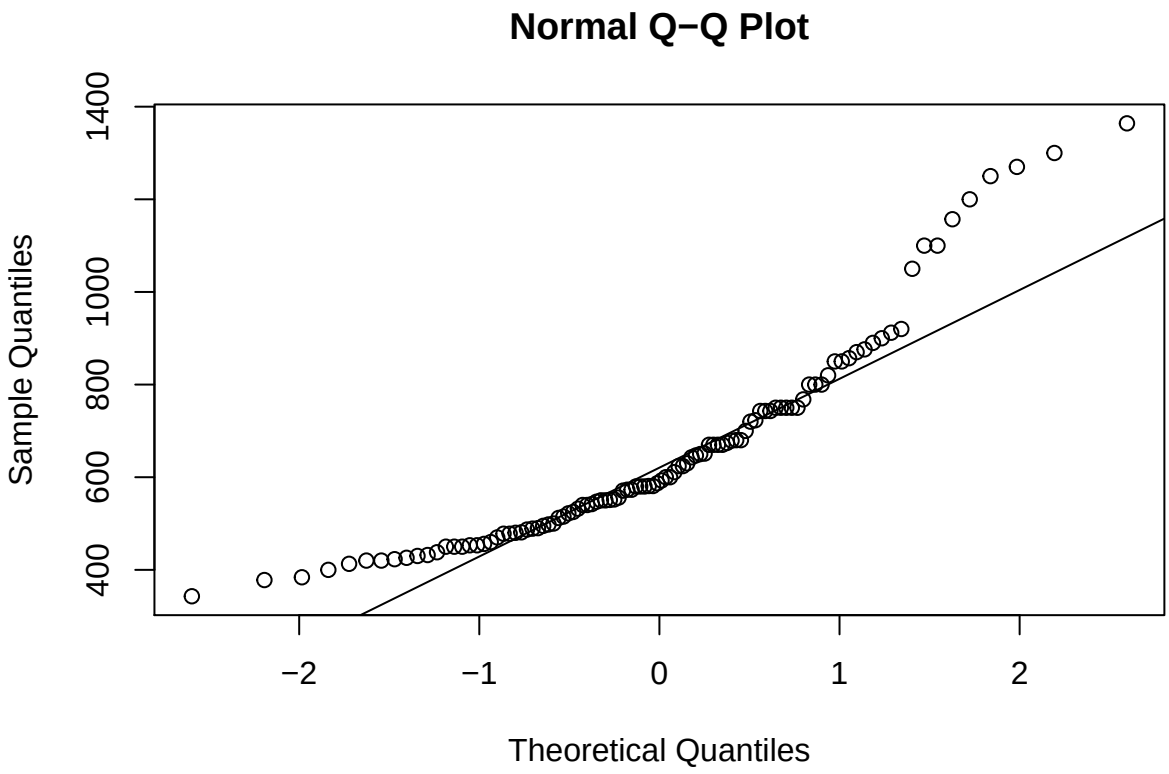
```
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



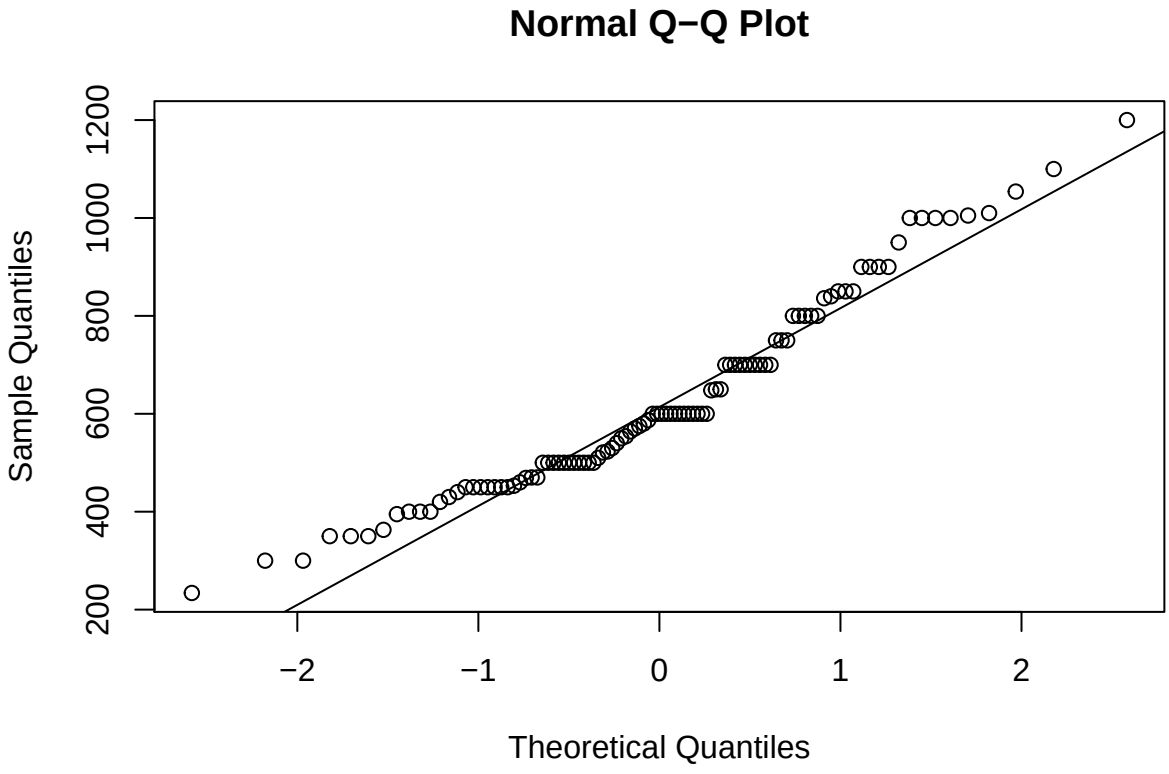

```
qqnorm(book$student_bezNagrade)
qqline(book$student_bezNagrade)
```



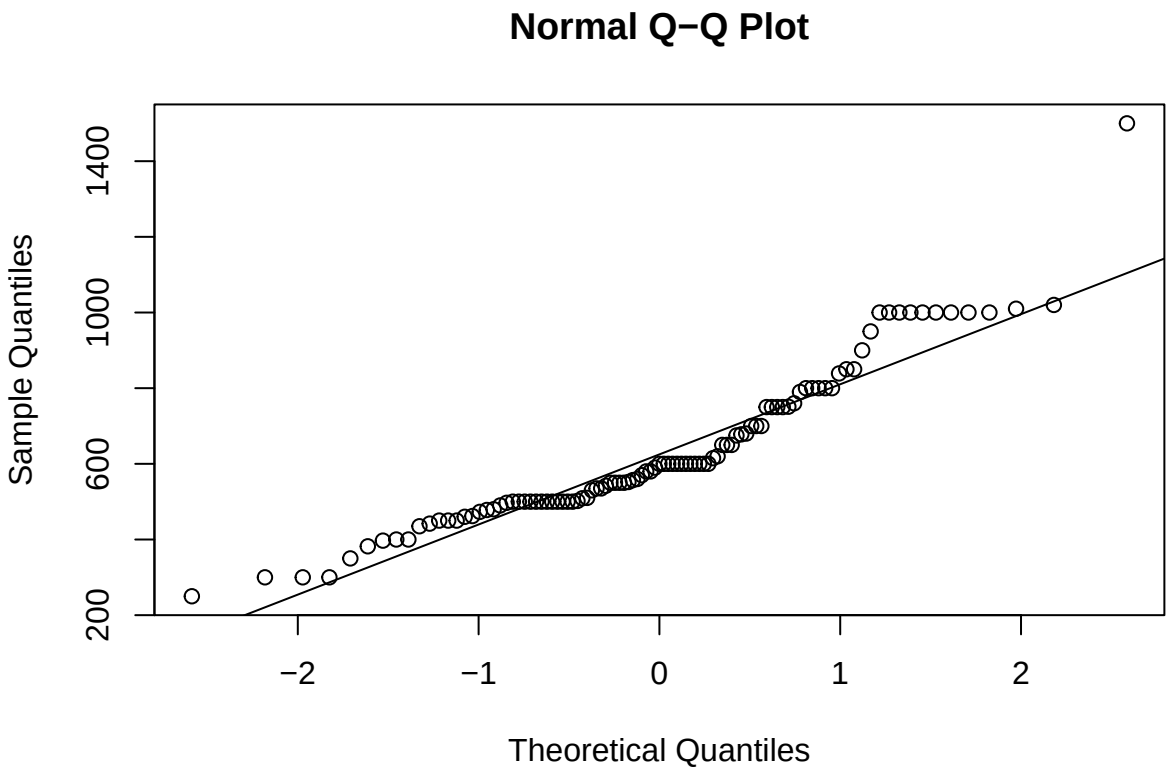

```
qqnorm(book$student_Nagrada)
qqline(book$student_Nagrada)
```



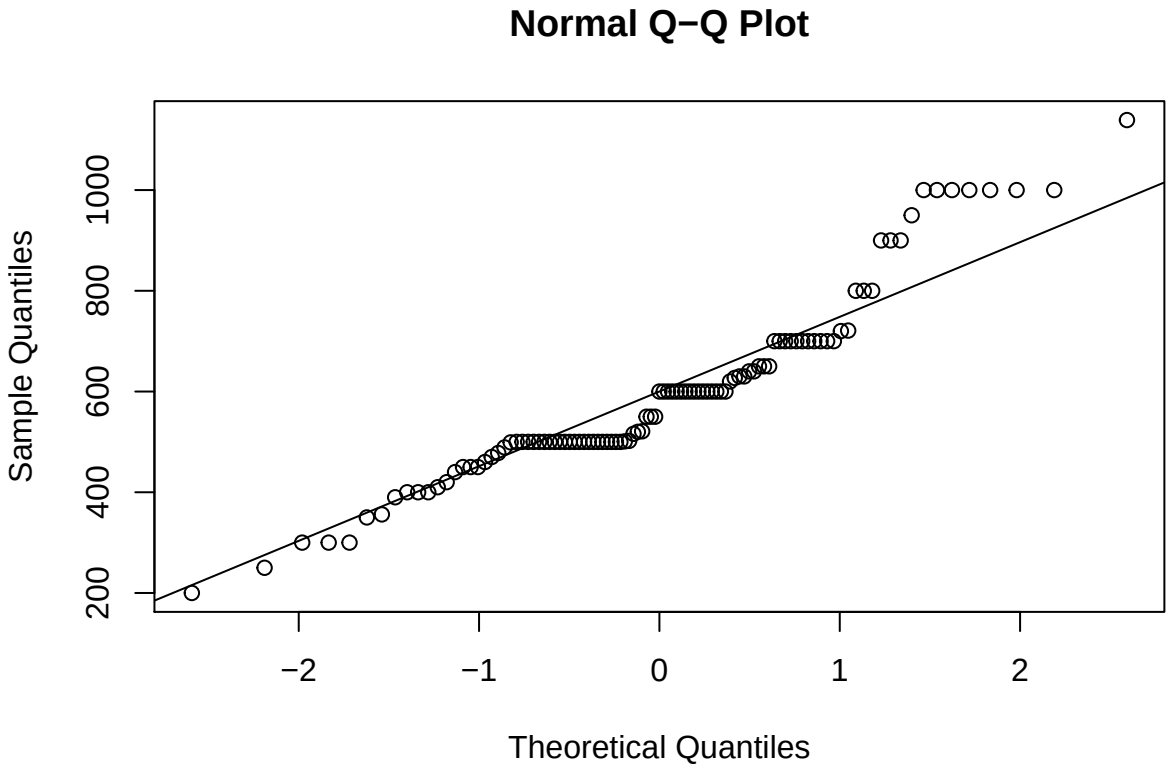

```
qqnorm(book$samo.prijatelji_bezNagrade)
qqline(book$samo.prijatelji_bezNagrade)
```



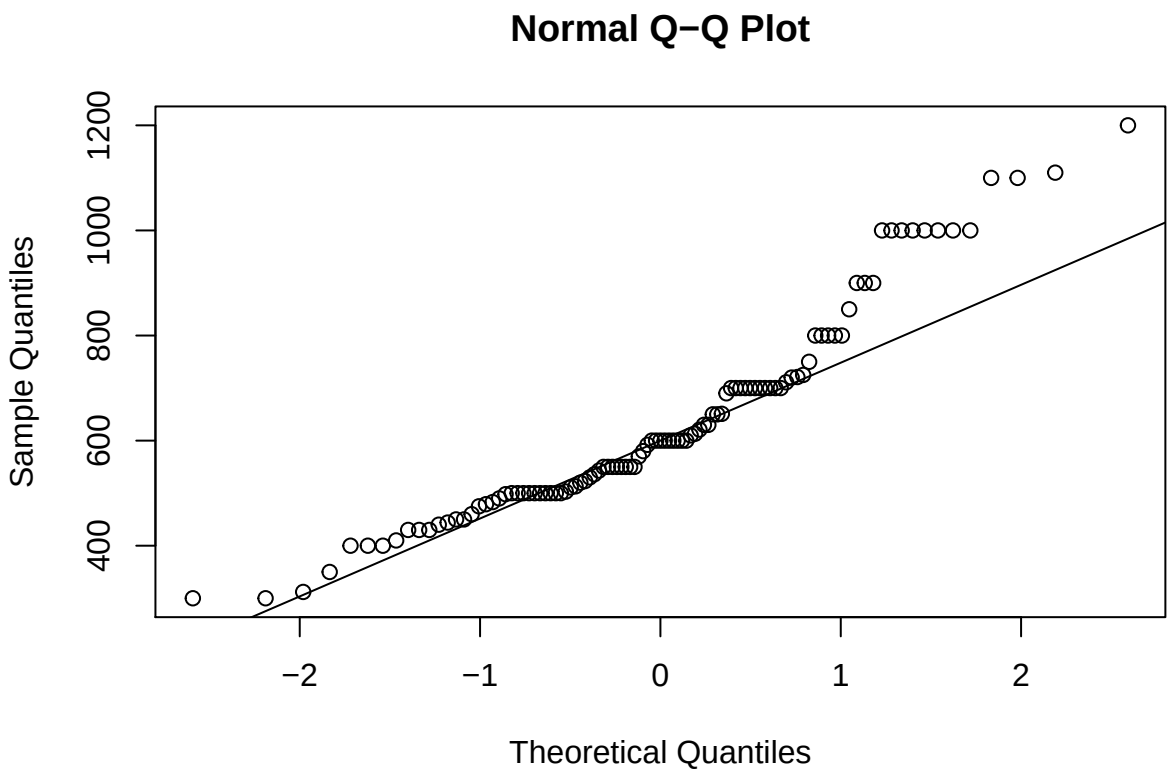

```
qqnorm(book$samo.prijatelji_Nagrada)
qqline(book$samo.prijatelji_Nagrada)
```



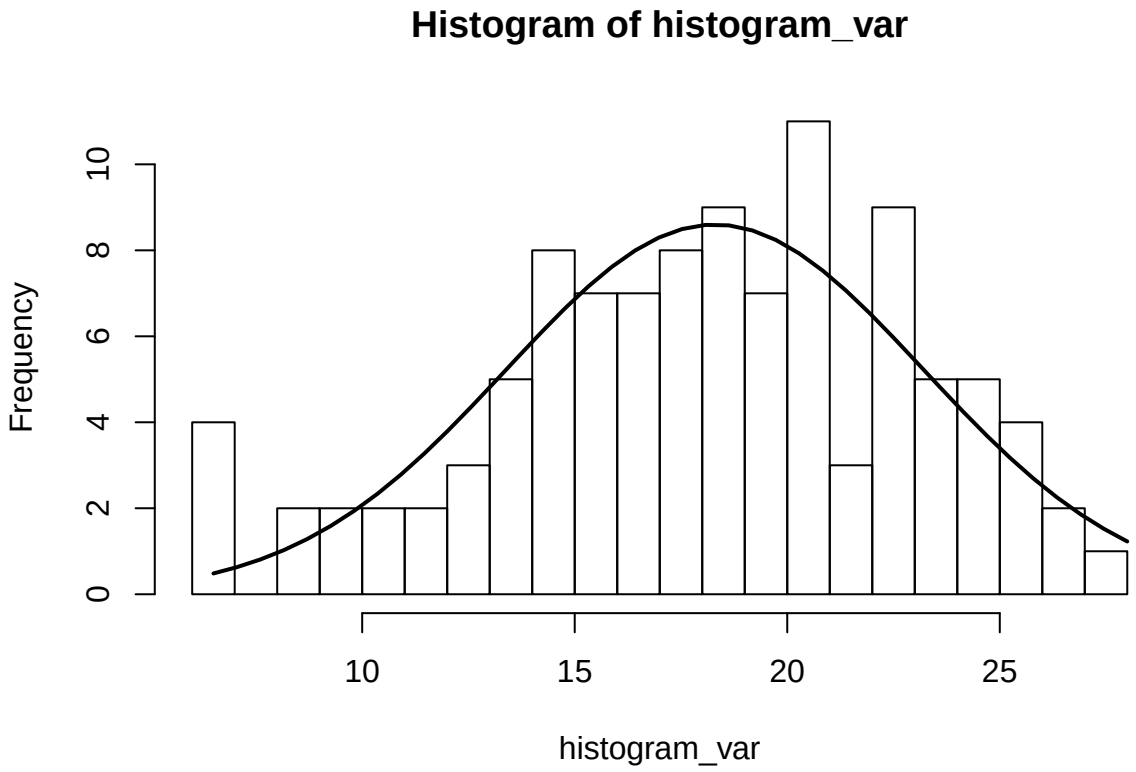

```
qqnorm(book$cijela.grupa_bezNagrade)
qqline(book$cijela.grupa_bezNagrade)
```



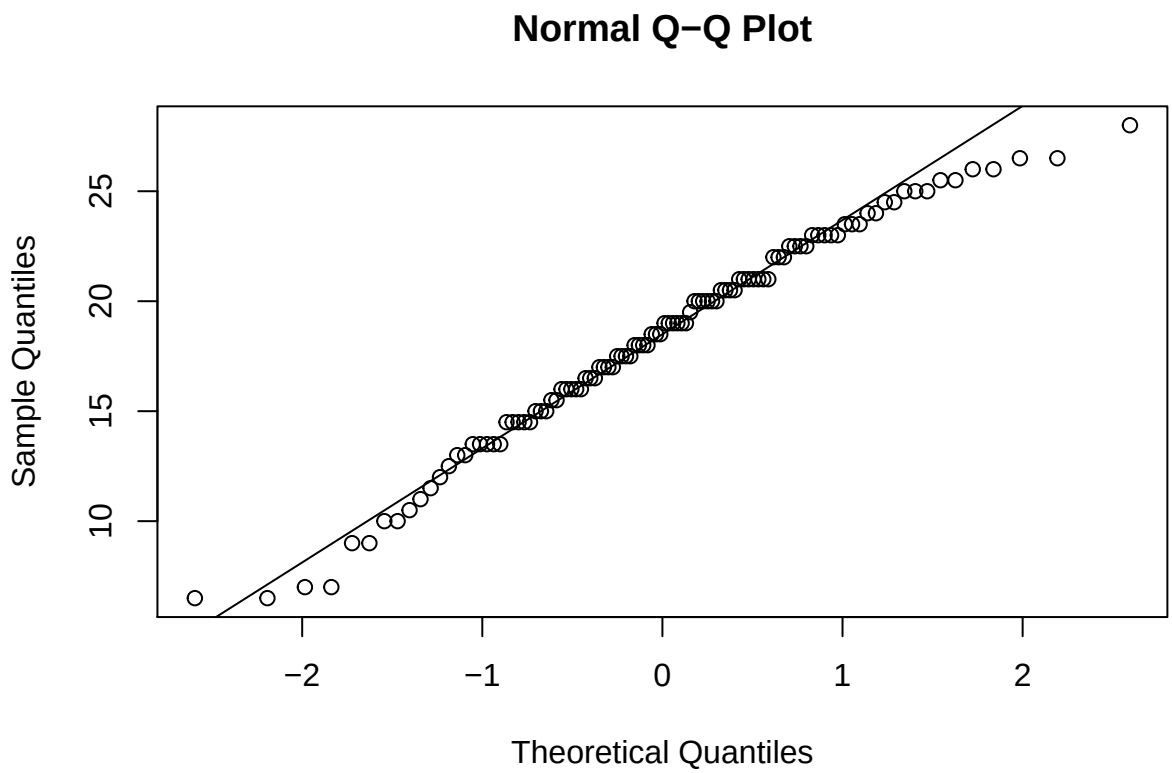

```
qqnorm(book$cijela.grupa_Nagrada)
qqline(book$cijela.grupa_Nagrada)
```




```
histogram_var = book$MI
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```




```
qqnorm(book$MI)
qqline(book$MI)
```

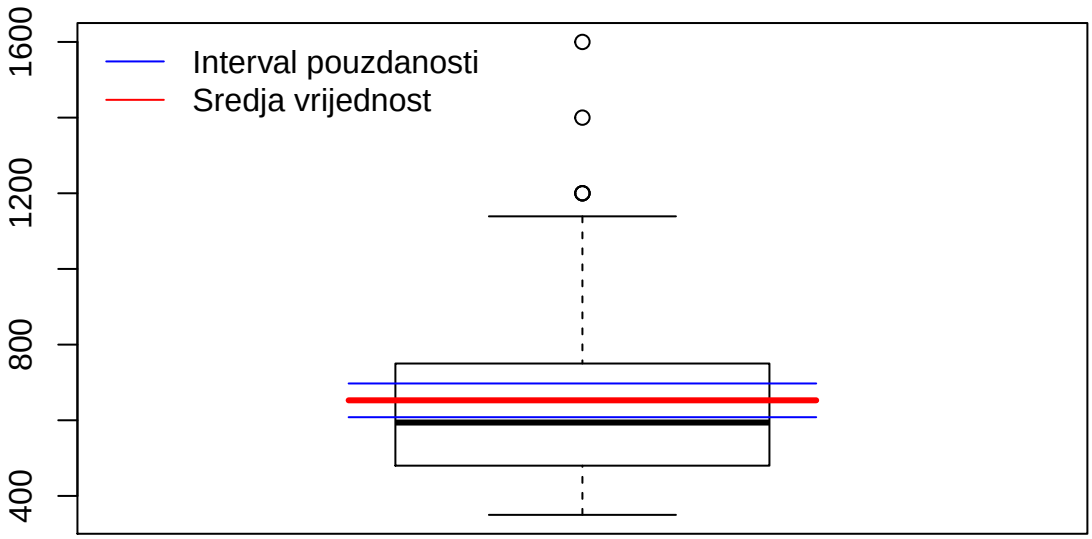


##Normalnost varijabli Analizirajući histograme i qq-grafove po pojedinim značajkama podataka, zaključili smo da se sve varijable mogu aproksimirati normalnom razdiobom. To nam omogućava raznoliku paletu testova koji pretpostavljaju normalnost podataka.

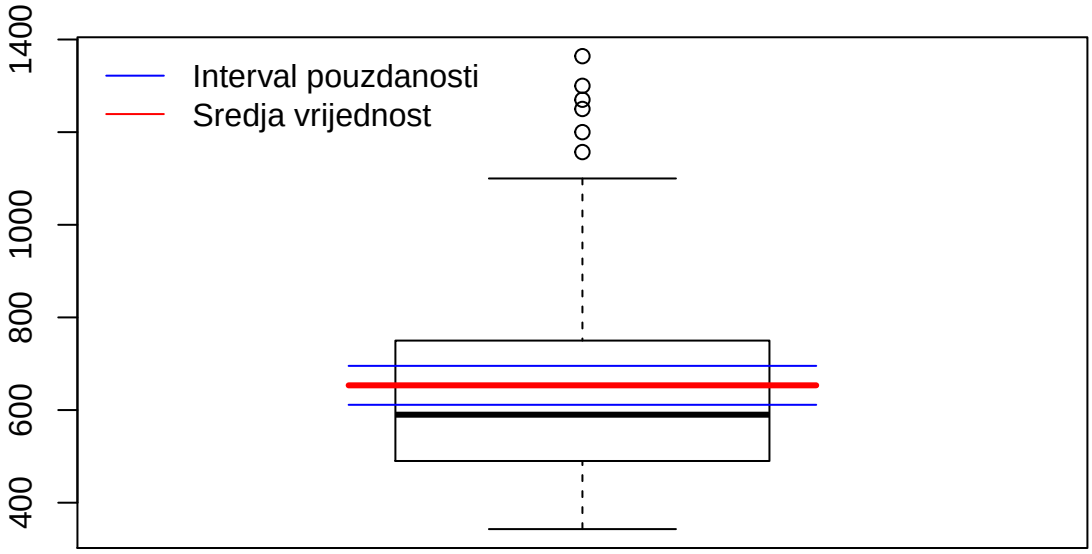
Testiranje hipoteza

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.

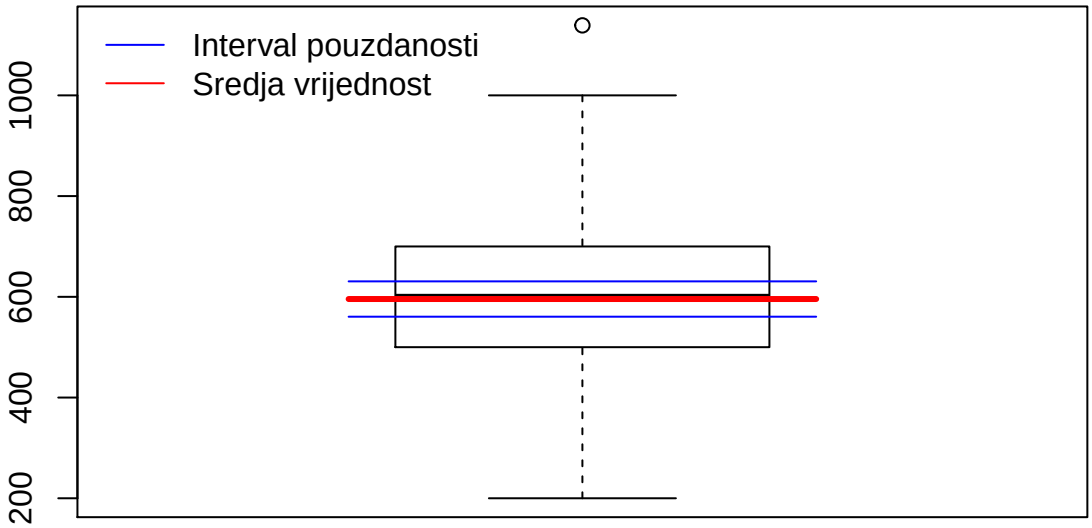
```
alpha = 0.05
t_conf_var = book$student_bezNagrade
t_conf_var = t_conf_var[!is.na(t_conf_var)]
x_mean = meanSBN
sd = sd(t_conf_var)
N = length(t_conf_var)
dof = N-1
t_alpha_h = qt(1-alpha/2, df = dof)
se = sd / sqrt(N)
left = x_mean - t_alpha_h * se
right = x_mean + t_alpha_h * se
boxplot(t_conf_var)
lines(c(0.75,1.25),c(left,left),col=4) # xstart - xend, ystart - yend, color
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")
```



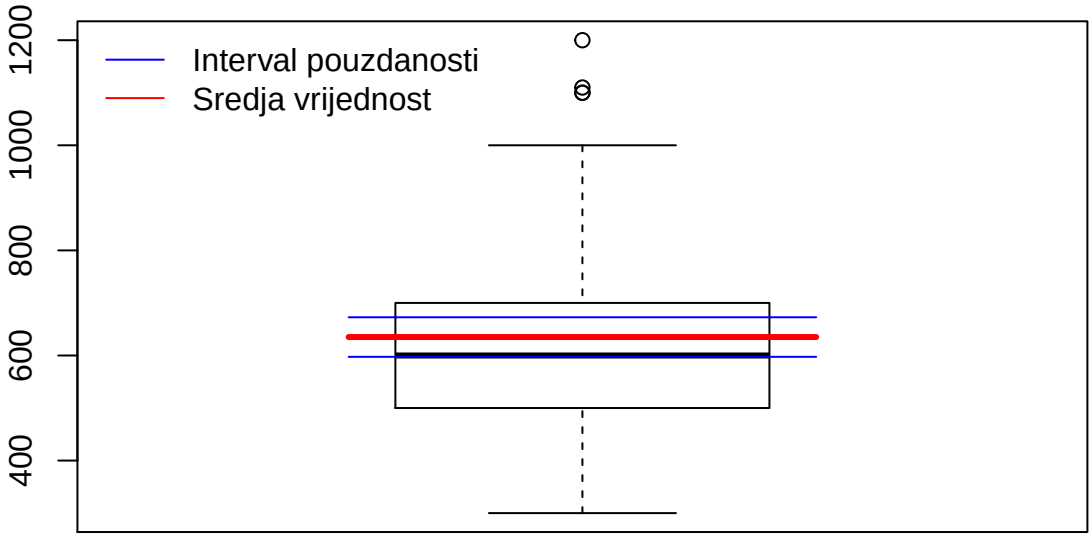

```
alpha = 0.05
t_conf_var = book$student_Nagrada
t_conf_var = t_conf_var[!is.na(t_conf_var)]
x_mean = meanSN
sd = sd(t_conf_var)
N = length(t_conf_var)
dof = N-1
t_alpha_h = qt(1-alpha/2, df = dof)
se = sd / sqrt(N)
left = x_mean - t_alpha_h * se
right = x_mean + t_alpha_h * se
boxplot(t_conf_var)
lines(c(0.75,1.25),c(left,left),col=4) # xstart - xend, ystart - yend, color
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")
```



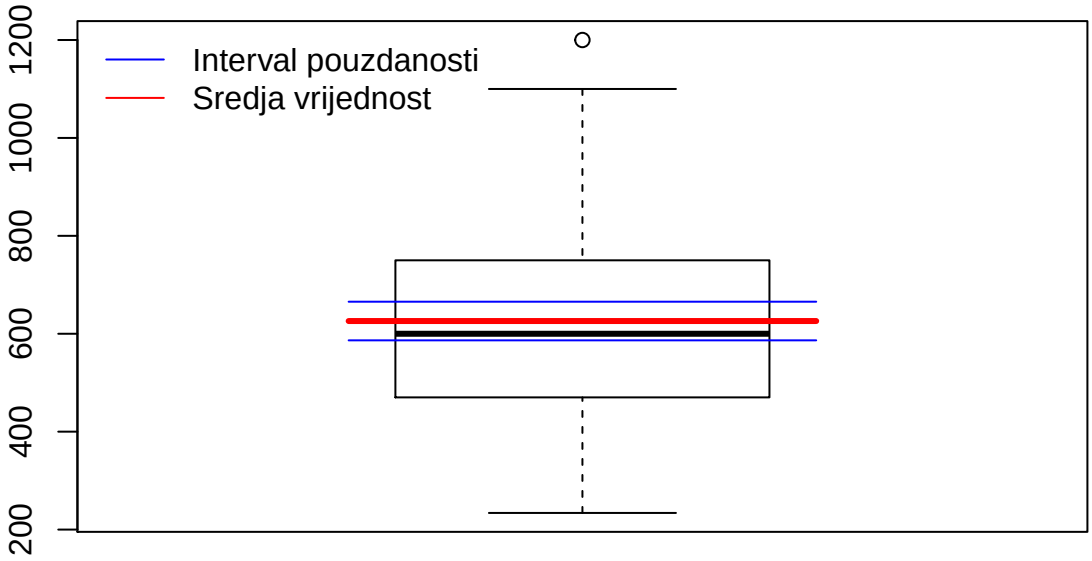

```
alpha = 0.05
t_conf_var = book$cijela.grupa_bezNagrade
t_conf_var = t_conf_var[!is.na(t_conf_var)]
x_mean = meanGBN
sd = sd(t_conf_var)
N = length(t_conf_var)
dof = N-1
t_alpha_h = qt(1-alpha/2, df = dof)
se = sd / sqrt(N)
left = x_mean - t_alpha_h * se
right = x_mean + t_alpha_h * se
boxplot(t_conf_var)
lines(c(0.75,1.25),c(left,left),col=4) # xstart - xend, ystart - yend, color
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")
```



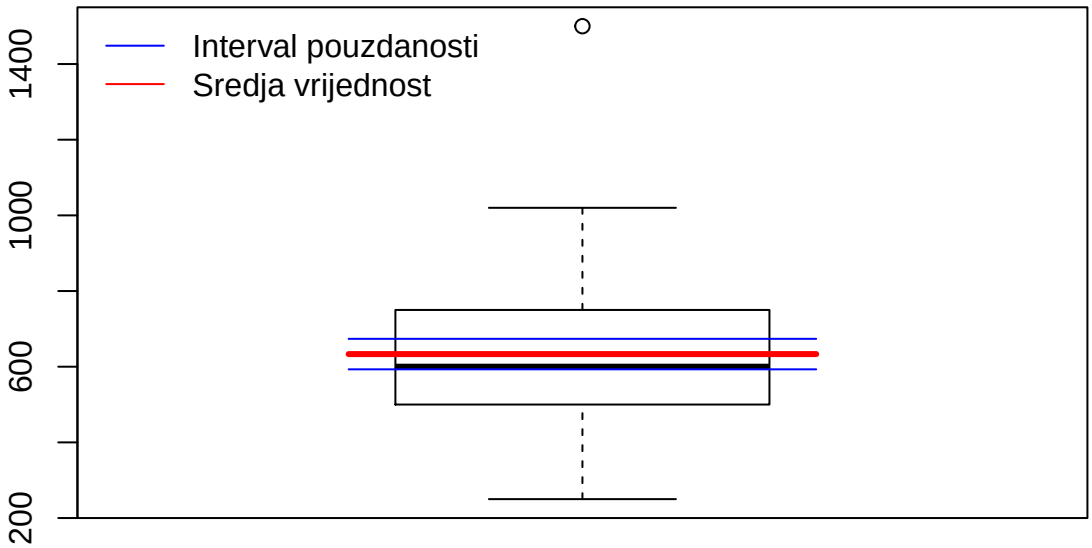
```
alpha = 0.05
t_conf_var = book$cijela.grupa_Nagrada
t_conf_var = t_conf_var[!is.na(t_conf_var)]
x_mean = meanGN
sd = sd(t_conf_var)
N = length(t_conf_var)
dof = N-1
t_alpha_h = qt(1-alpha/2, df = dof)
se = sd / sqrt(N)
left = x_mean - t_alpha_h * se
right = x_mean + t_alpha_h * se
boxplot(t_conf_var)
lines(c(0.75,1.25),c(left,left),col=4) # xstart - xend, ystart - yend, color
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")
```




```
alpha = 0.05
t_conf_var = book$samo.prijatelji_bezNagrade
t_conf_var = t_conf_var[!is.na(t_conf_var)]
x_mean = meanPBN
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) # xstart - xend, ystart - yend, color
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")
```




```
alpha = 0.05
t_conf_var = book$samo.prijatelji_Nagrada
t_conf_var = t_conf_var[!is.na(t_conf_var)]
x_mean = meanPN
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) # xstart - xend, ystart - yend, color
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")
```



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_bezNagrade, alternative = "two.sided", conf.level = 0.99, pai
```

```
##
##  Paired t-test
##
## data:  book$student_Nagrada and book$student_bezNagrade
## 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_bezNagrade, alternative = "two.sided", conf.l
```

```
##
##  Paired t-test
##
## data:  book$samo.prijatelji_Nagrada and book$samo.prijatelji_bezNagrade
## 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
```

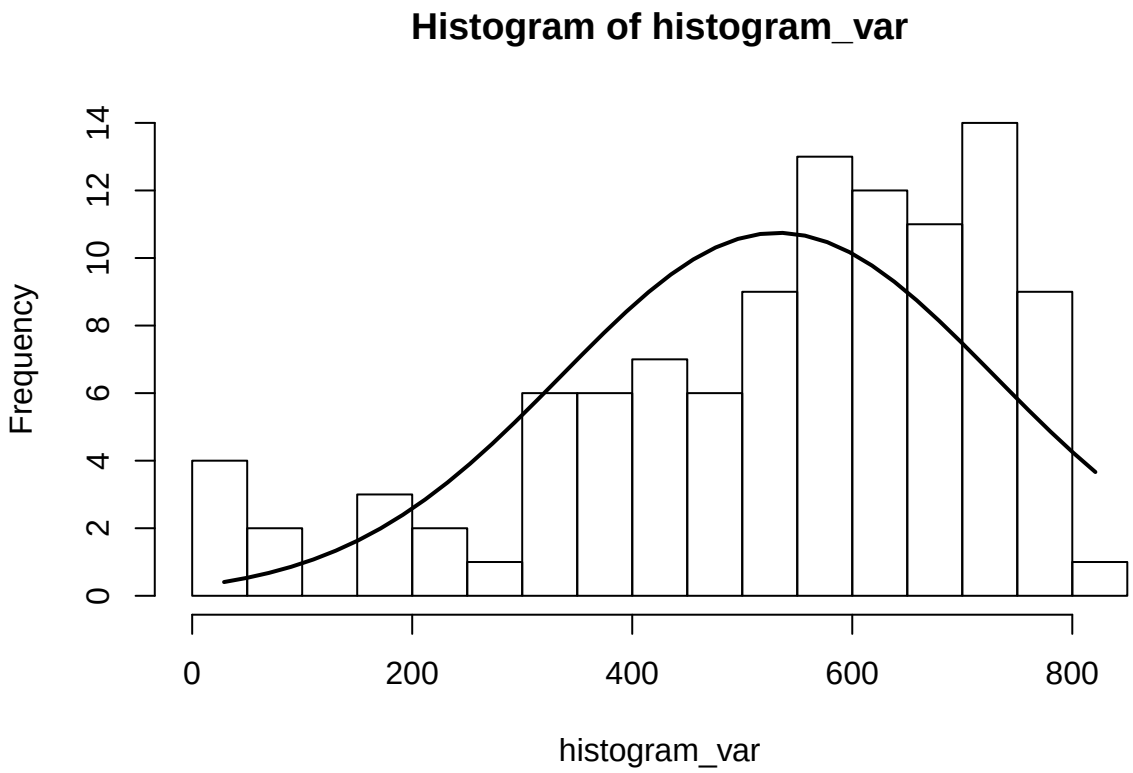
```
t.test(book$cijela.grupa_Nagrada, book$cijela.grupa_bezNagrade, alternative = "greater", conf.level = 0
```

```
##
##  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
```

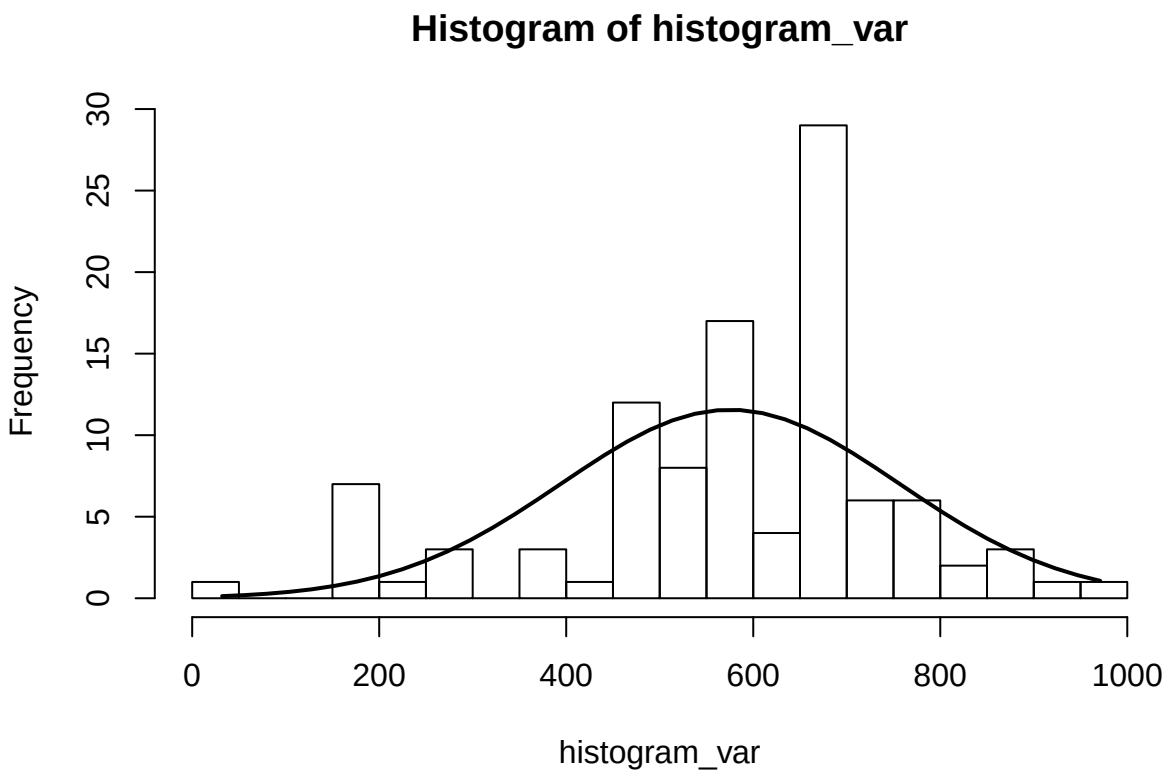
Odmak od stvarnog broja stranica

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.

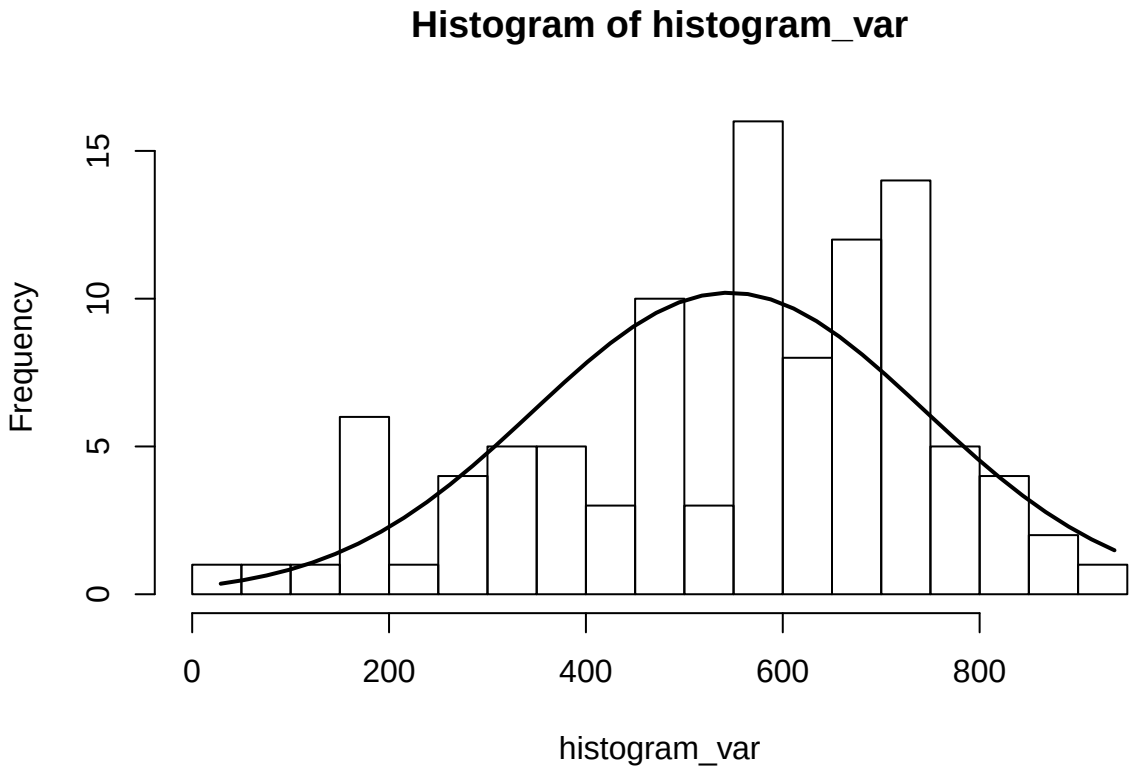
```
offset = book
offset$Grupa.Pred = NULL
offset$spol = NULL
offset$MI = NULL
offset = abs(1171 - offset)

histogram_var = offset$student_bezNagrade
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



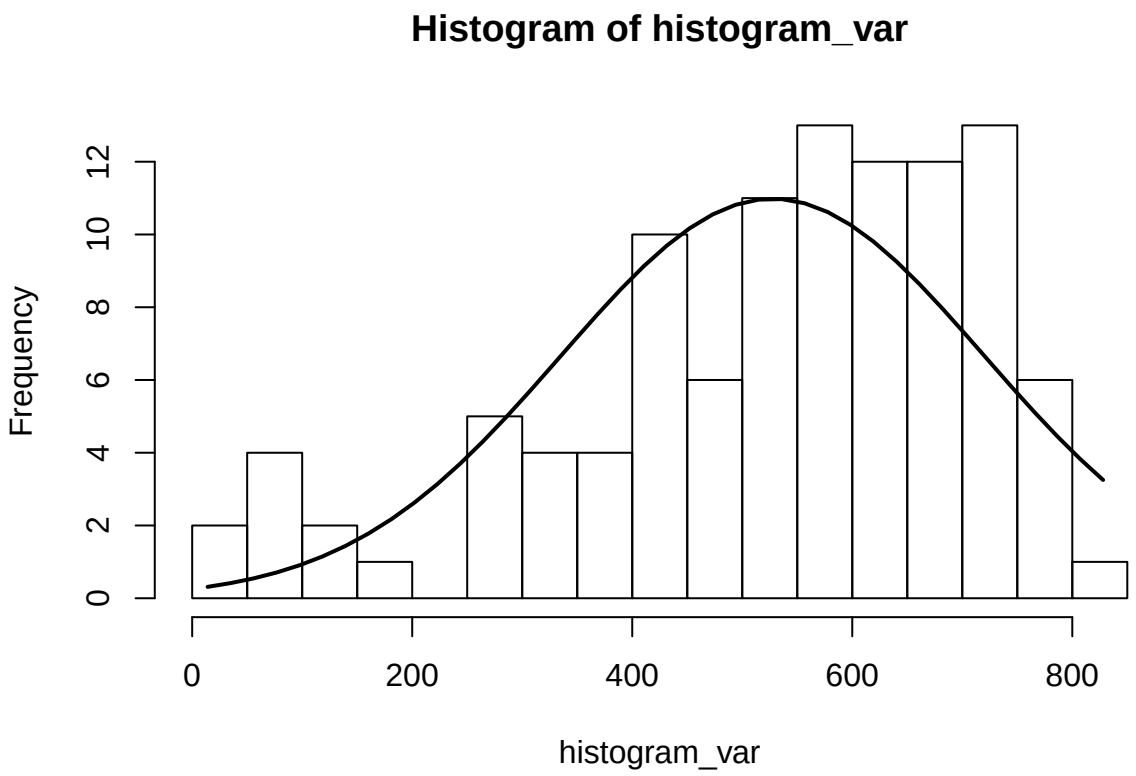

```
histogram_var = offset$cijela.grupa_bezNagrade
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



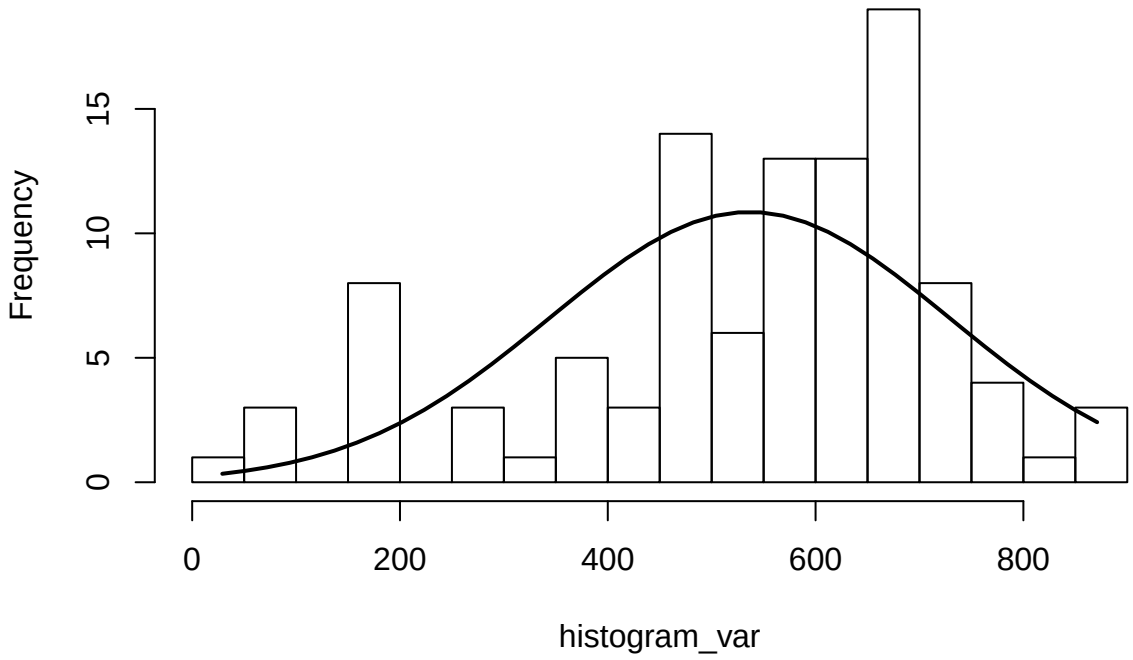

```
histogram_var = offset$samo.prijatelji_bezNagrade
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```




```
histogram_var = offset$student_Nagrada
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



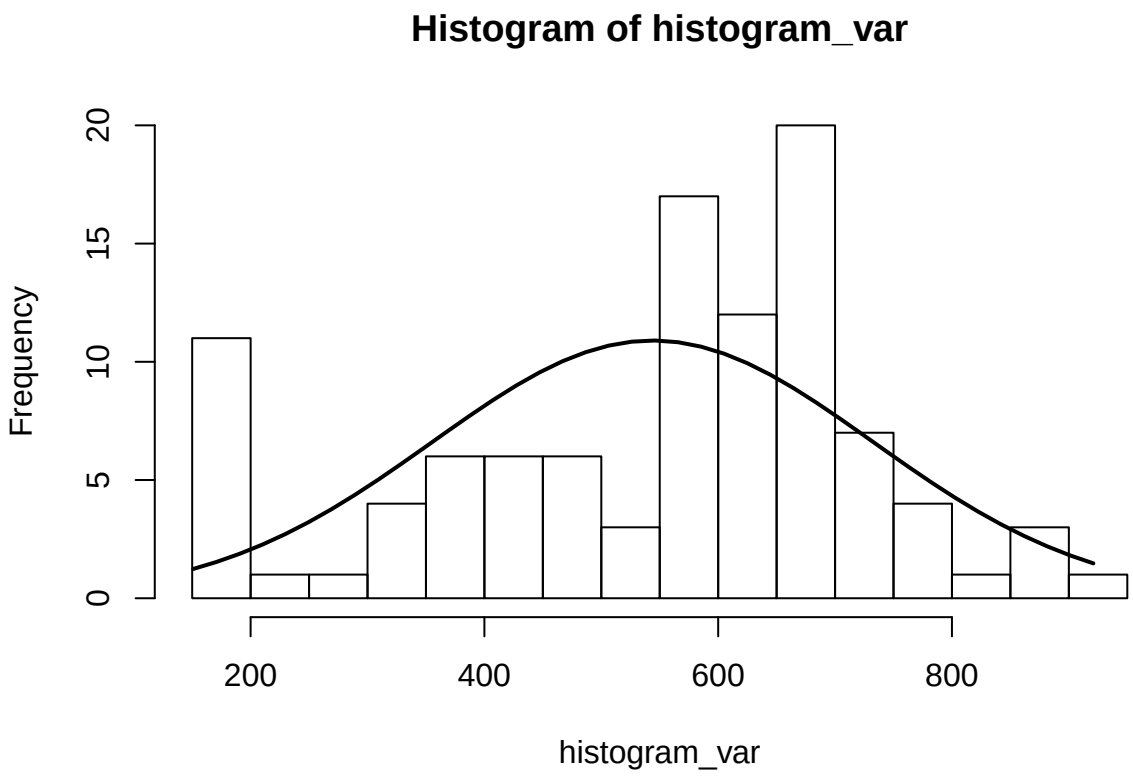

```
histogram_var = offset$cijela.grupa_Nagrada
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



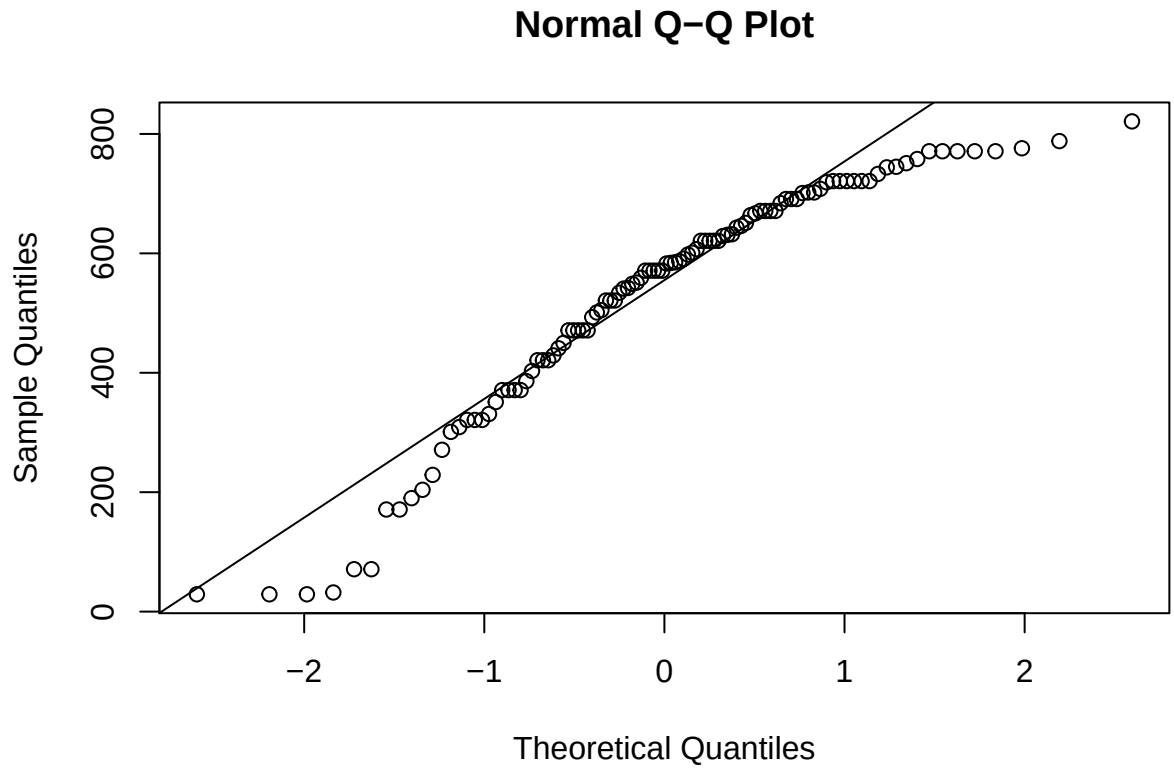

```
histogram_var = offset$samo.prijatelji_Nagrada
histogram_var = histogram_var[!is.na(histogram_var)]
h = hist(histogram_var, nclass = 17)
min(histogram_var)
```

```
## [1] 151
```

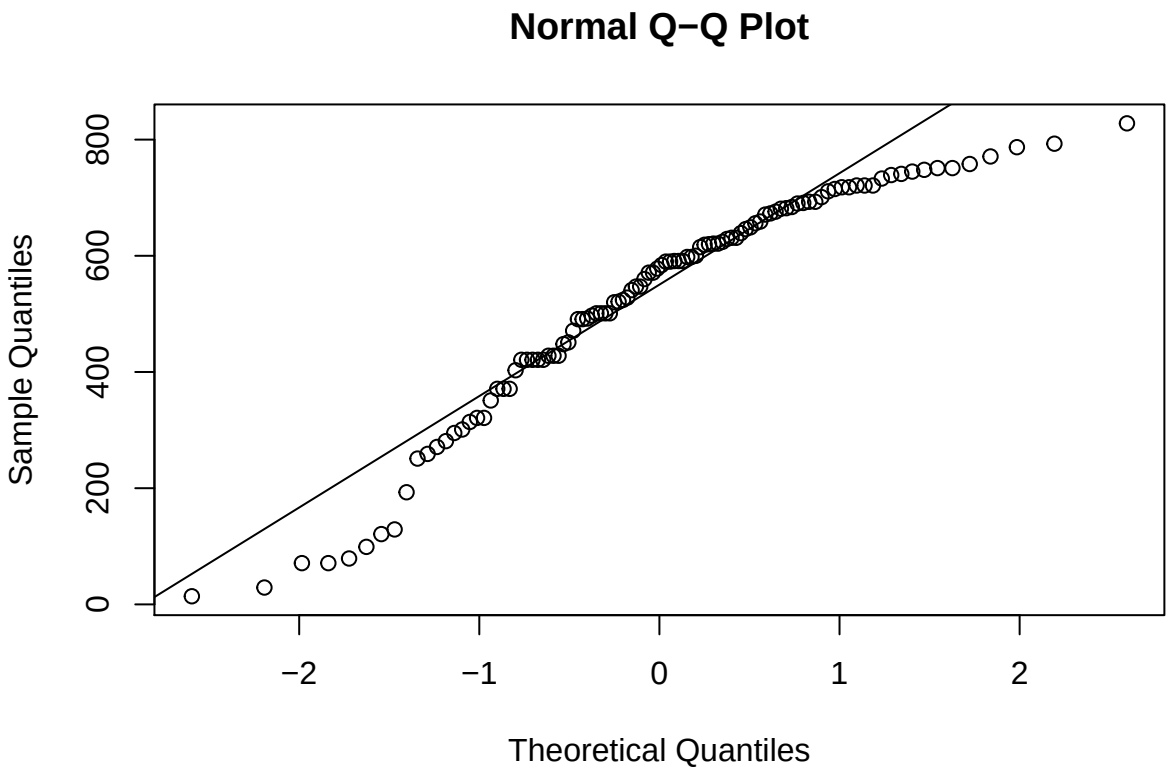
```
xfit = seq(min(histogram_var), max(histogram_var), length = 40)
yfit = dnorm(xfit, mean = mean(histogram_var), sd = sd(histogram_var))
yfit = yfit * diff(h$mids[1:2]) * length(histogram_var)
lines(xfit, yfit, col = "black", lwd = 2)
```



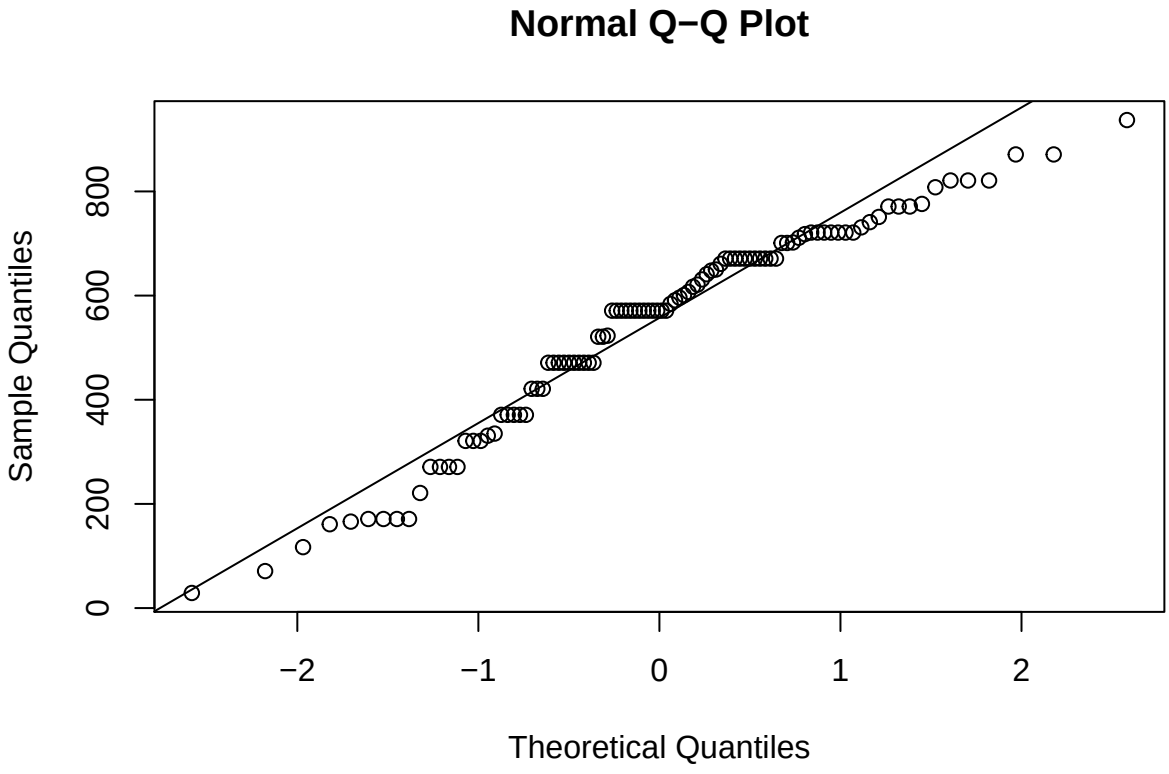

```
qqnorm(offset$student_bezNagrade)
qqline(offset$student_bezNagrade)
```



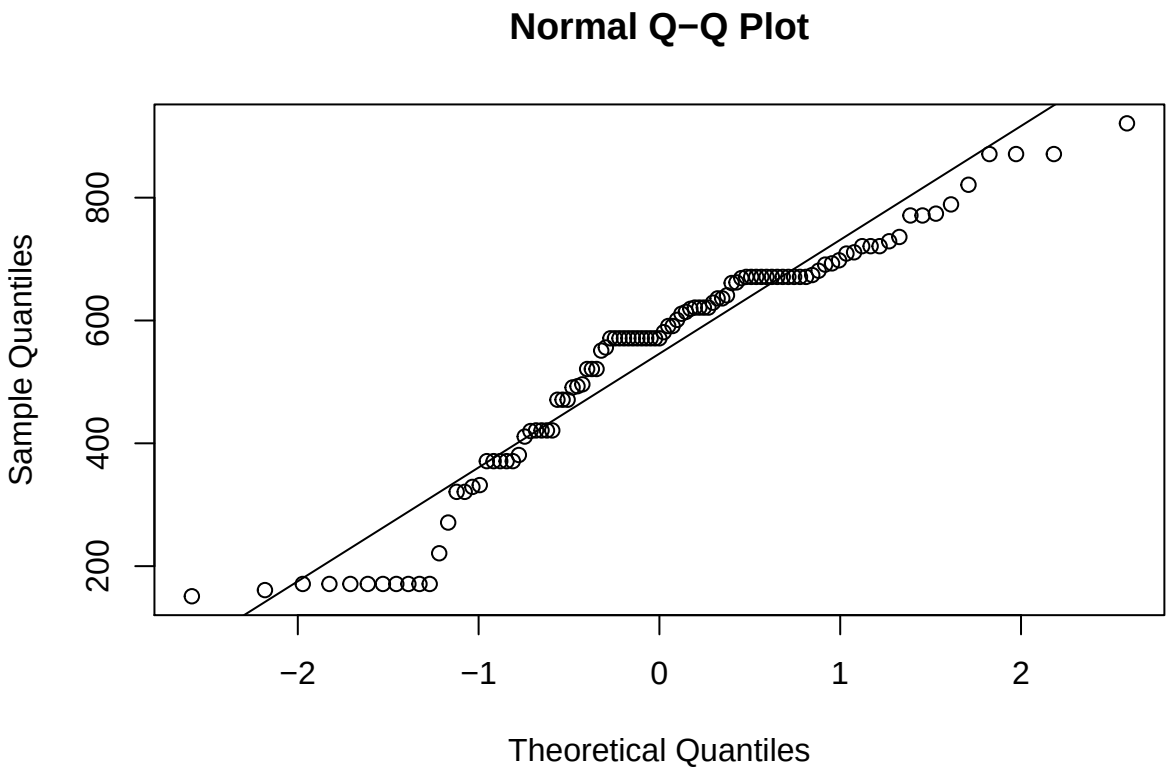

```
qqnorm(offset$student_Nagrada)
qqline(offset$student_Nagrada)
```



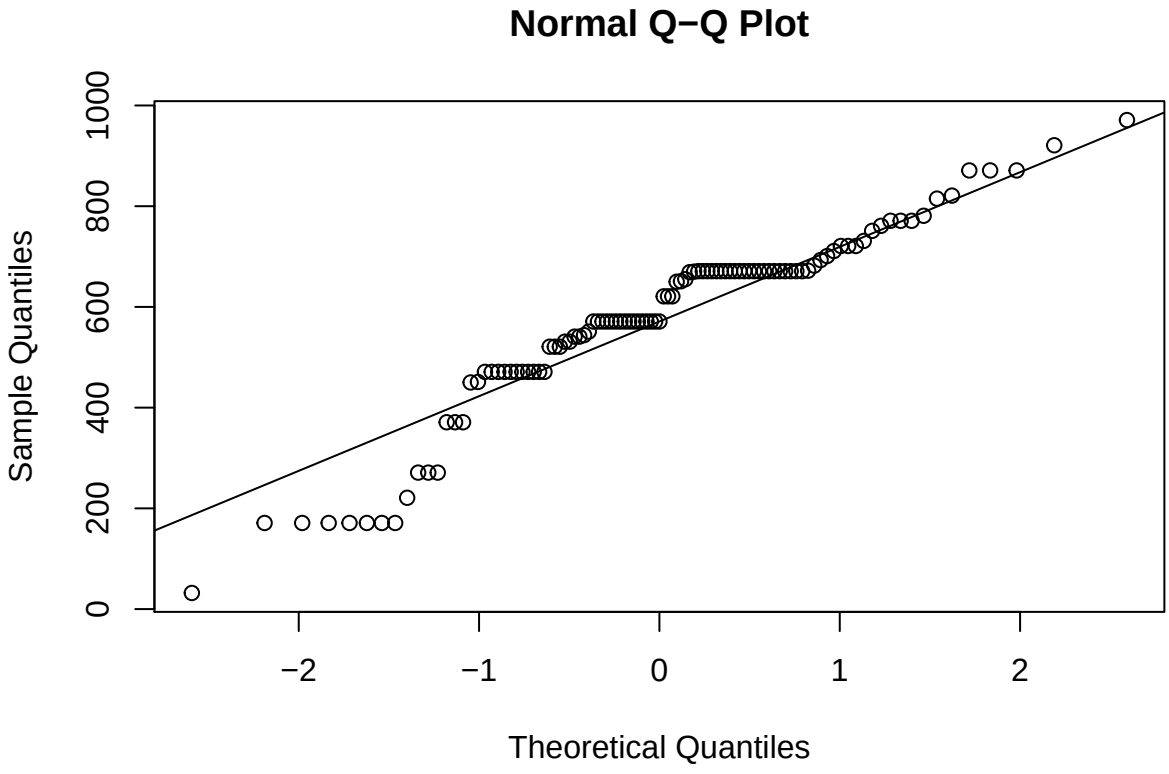

```
qqnorm(offset$samo.prijatelji_bezNagrade)
qqline(offset$samo.prijatelji_bezNagrade)
```



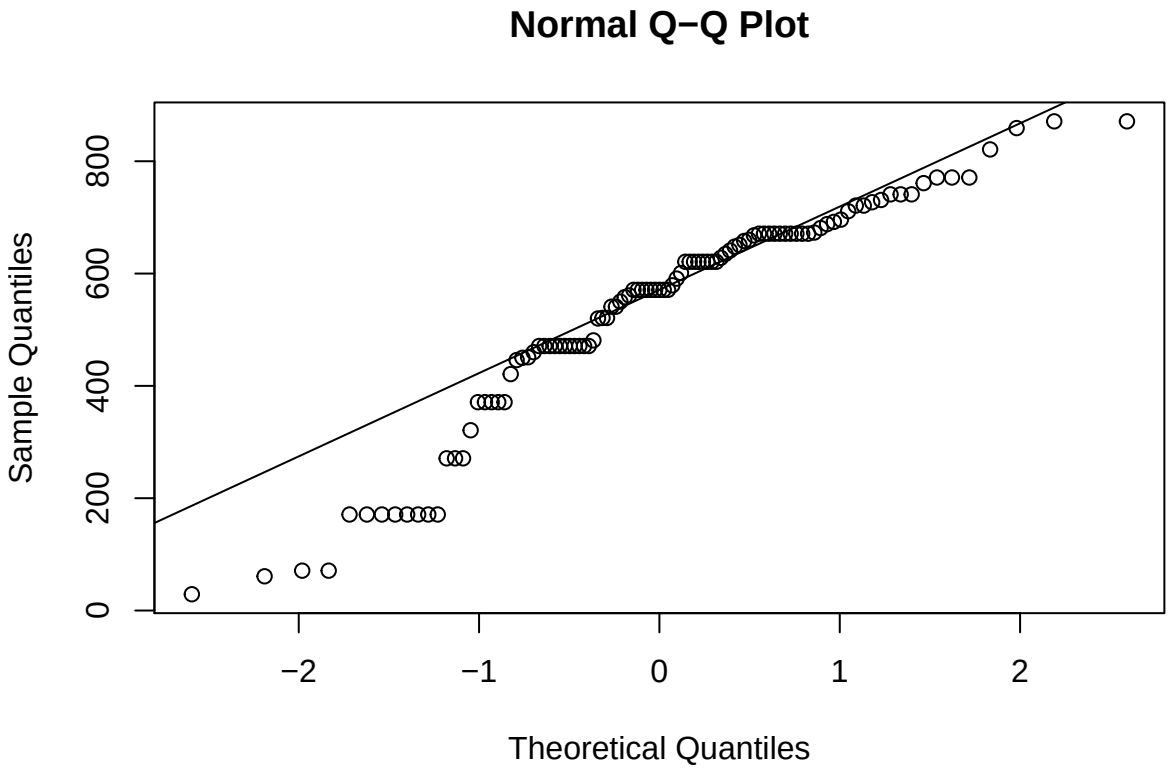

```
qqnorm(offset$samo.prijatelji_Nagrada)
qqline(offset$samo.prijatelji_Nagrada)
```




```
qqnorm(offset$cijela.grupa_bezNagrade)
qqline(offset$cijela.grupa_bezNagrade)
```




```
qqnorm(offset$cijela.grupa_Nagrada)
qqline(offset$cijela.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_bezNagrade, alternative = "less", conf.level = 0.95, pair
```

```
##
##  Paired t-test
##
## data:  offset$student_Nagrada and offset$student_bezNagrade
## 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_bezNagrade, alternative = "less", conf.le
```

```
##
##  Paired t-test
##
## data:  offset$samo.prijatelji_Nagrada and offset$samo.prijatelji_bezNagrade
## 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
```

```r
t.test(offset$cijela.grupa_Nagrada, offset$cijela.grupa_bezNagrade, alternative = "less", conf.level = 0
```

```
##
##  Paired t-test
##
## data:  offset$cijela.grupa_Nagrada and offset$cijela.grupa_bezNagrade
## 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
```

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

```r
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
```

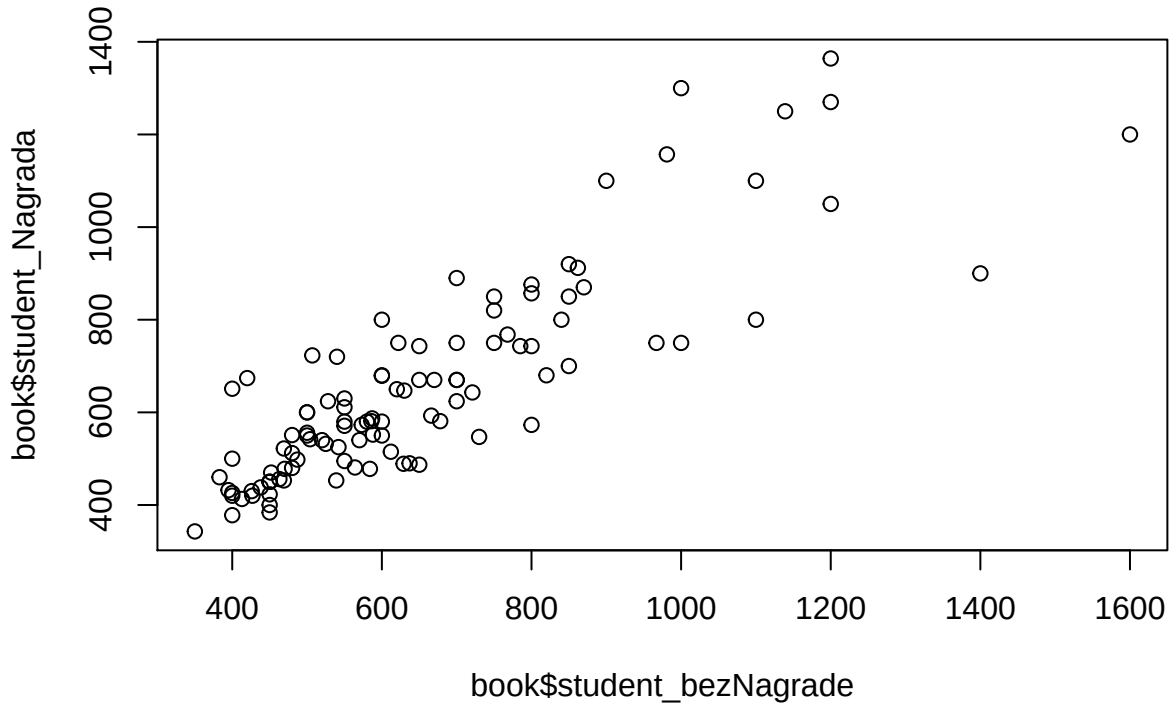
Uz razinu značajnosti od 13.3% imati ćemo dovoljno dokaza da se odbaci hipoteza o jednakosti srednjih vrijednosti procjena muške i ženske populacije.

```r
fit = lm(MI~student_bezNagrade + student_Nagrada + cijela.grupa_bezNagrade + cijela.grupa_Nagrada + samo

summary(fit)
```

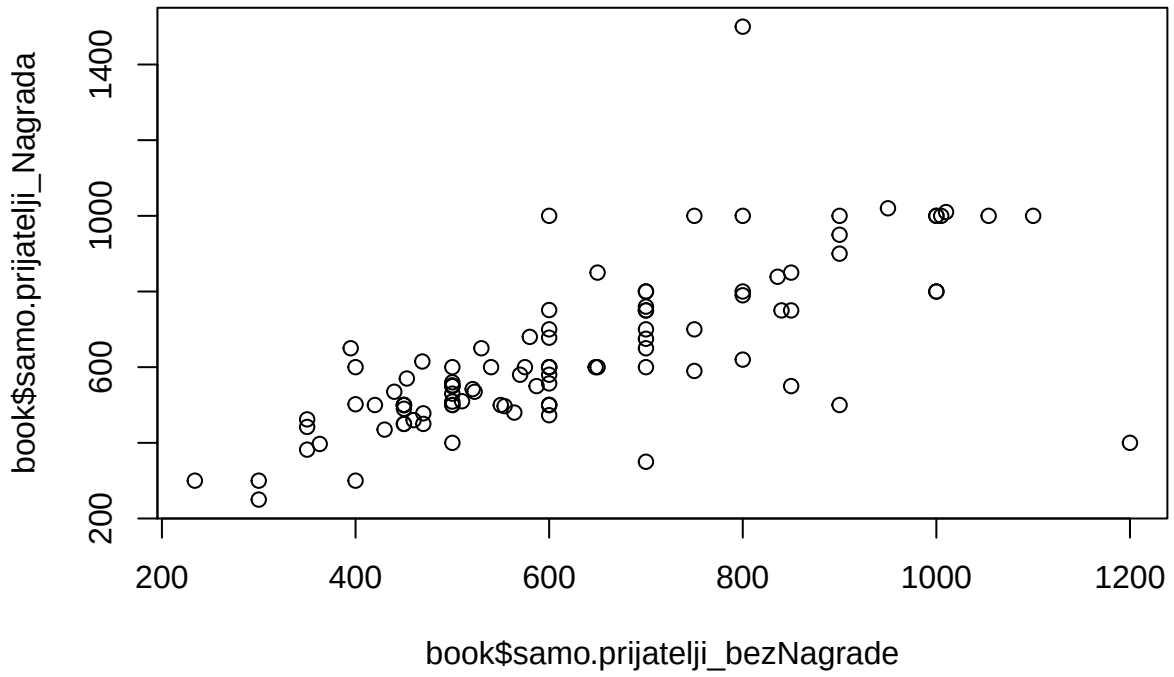
```
##
## Call:
## lm(formula = MI ~ student_bezNagrade + student_Nagrada + cijela.grupa_bezNagrade +
##     cijela.grupa_Nagrada + samo.prijatelji_bezNagrade + samo.prijatelji_Nagrada,
##     data = book)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.7244  -2.8040   0.5379   2.9606   8.7188
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                17.145746   2.039667   8.406 4.77e-13 ***
## student_bezNagrade         -0.003536   0.004674  -0.757    0.451
## student_Nagrada             0.004620   0.005414   0.853    0.396
## cijela.grupa_bezNagrade     0.002096   0.003710   0.565    0.573
## cijela.grupa_Nagrada       -0.003783   0.004599  -0.823    0.413
## samo.prijatelji_bezNagrade -0.001905   0.004307  -0.442    0.659
## samo.prijatelji_Nagrada     0.004610   0.005247   0.879    0.382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.898 on 93 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.02554,    Adjusted R-squared:  -0.03733
## F-statistic: 0.4062 on 6 and 93 DF,  p-value: 0.8732
```

```
fit = lm(MI~student_bezNagrade, data = book)
fit = lm(student_Nagrada~student_bezNagrade, data = book)

plot(book$student_bezNagrade, book$student_Nagrada)
```

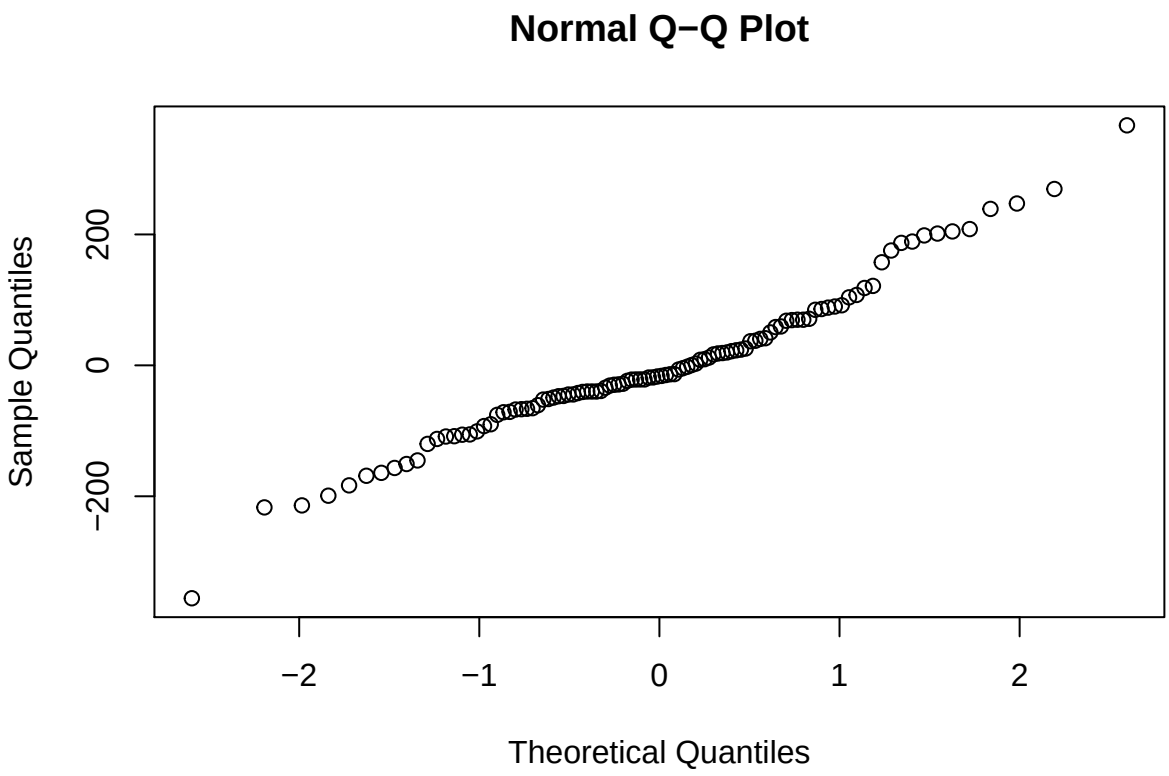



```
plot(book$samo.prijatelji_bezNagrade, book$samo.prijatelji_Nagrada)
```




```
plot(book$cijela.grupa_bezNagrade, book$cijela.grupa_Nagrada)
```


```
qqnorm(fit$residuals)
```




```
fit = lm(MI~student_bezNagrade, data = book)
fit = lm(MI~student_bezNagrade, data = book)
fit = lm(MI~student_bezNagrade, data = book)
fit = lm(MI~student_bezNagrade, data = book)
fit = lm(MI~student_bezNagrade, data = book)
```