

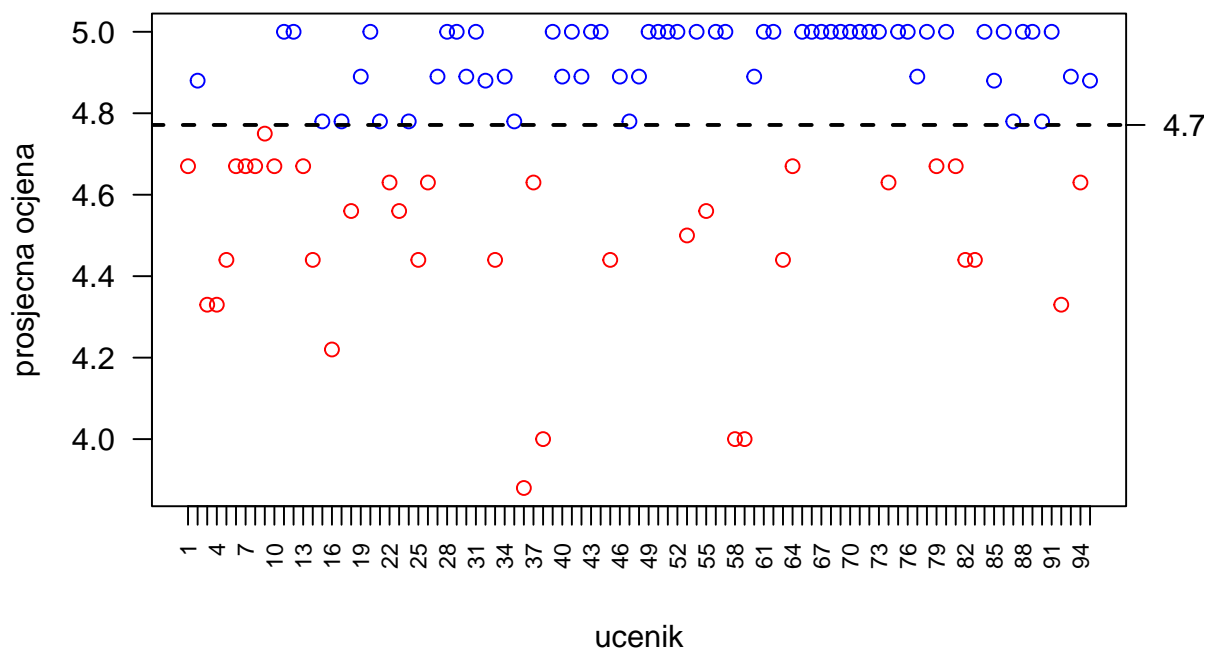
Istraživanje: Brzina čitanja učenika četvrtih razreda osnovne škole

Gordana Ljubas, Jan Ljubas

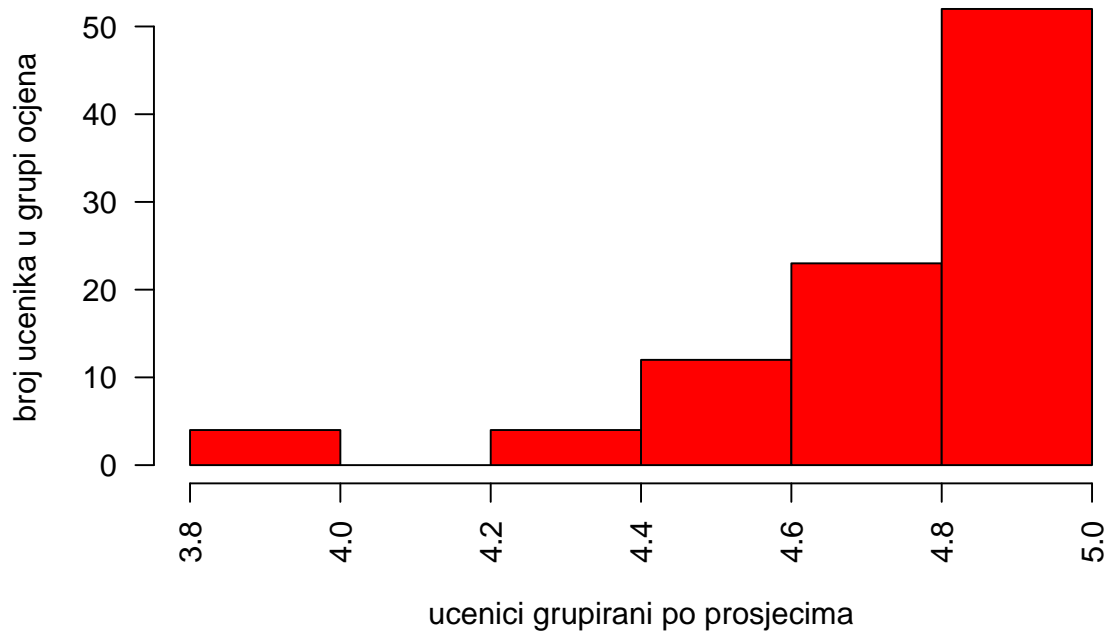
2023-2-20

Prikaz distribucije ocjena svih učenika

Prosječne ocjene svih 3. razreda

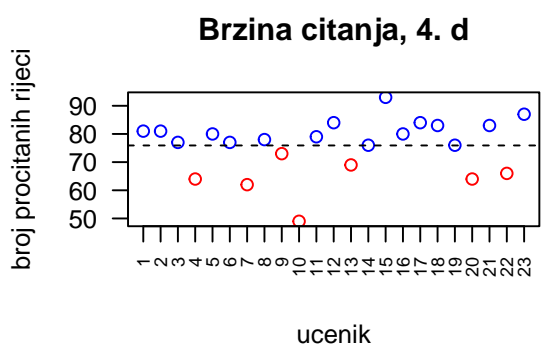
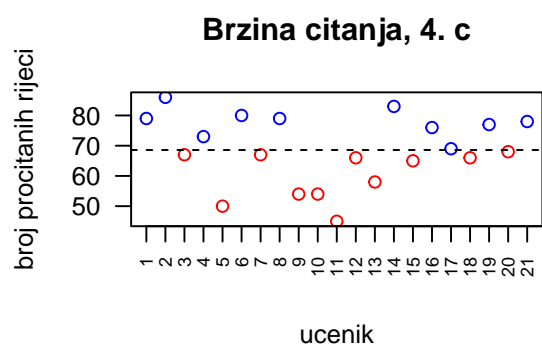
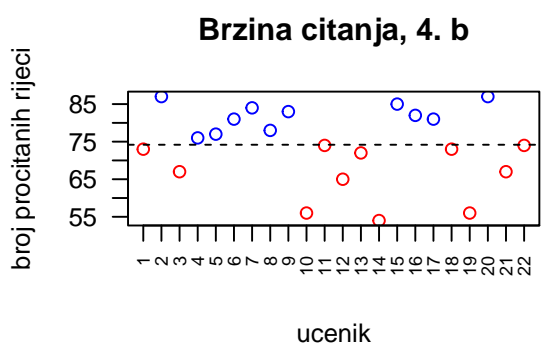
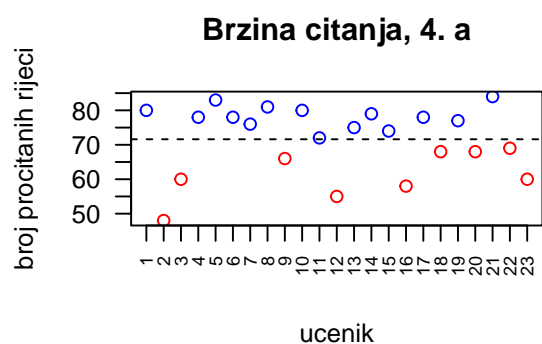


Stupčasti dijagram svih ocjena

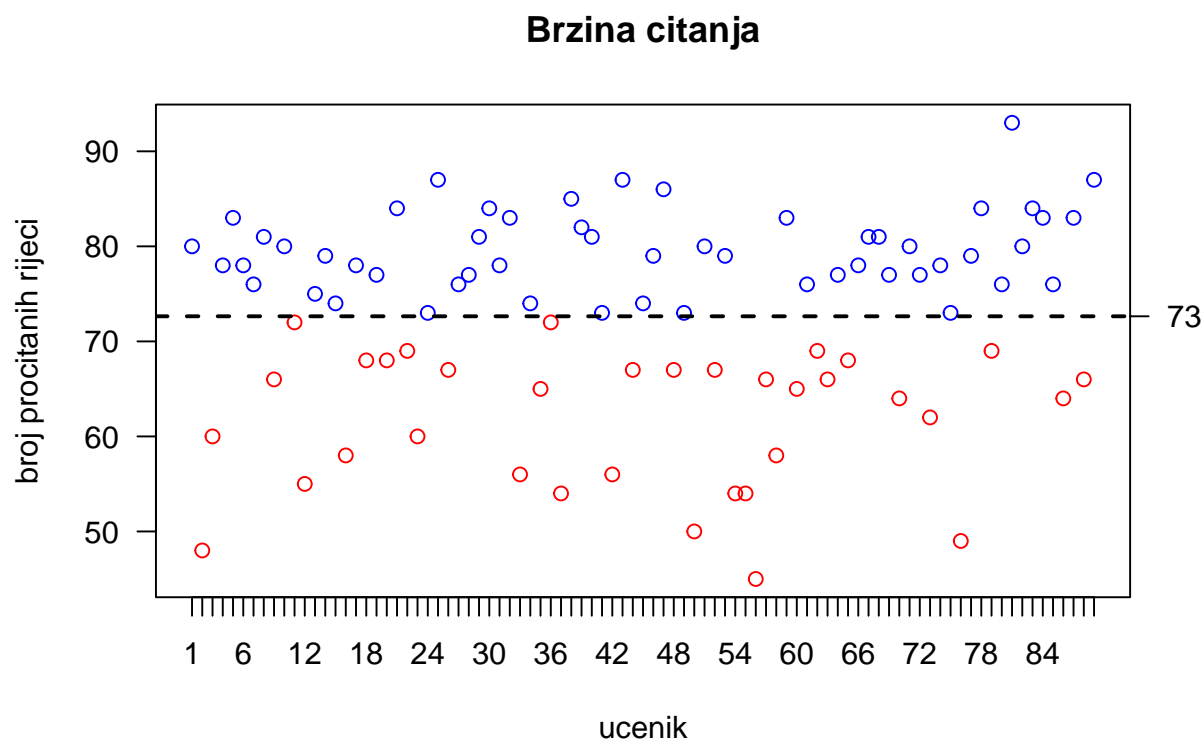


Brzina čitanja

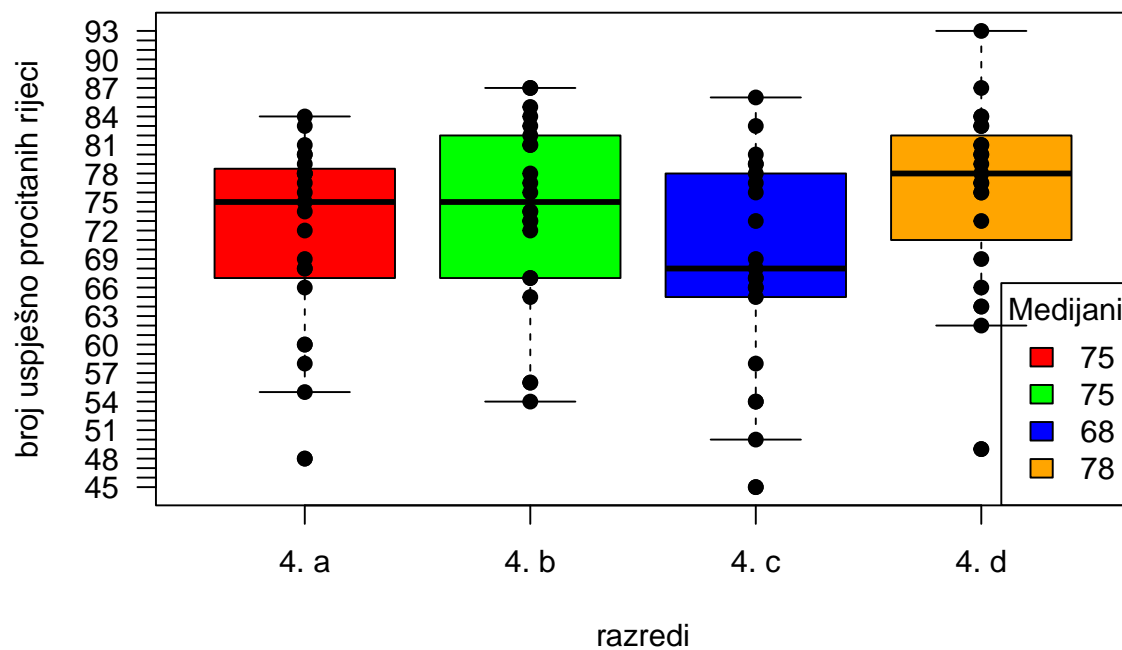
Brzine čitanja pojedinačnih razreda



Brzine čitanja svih razreda

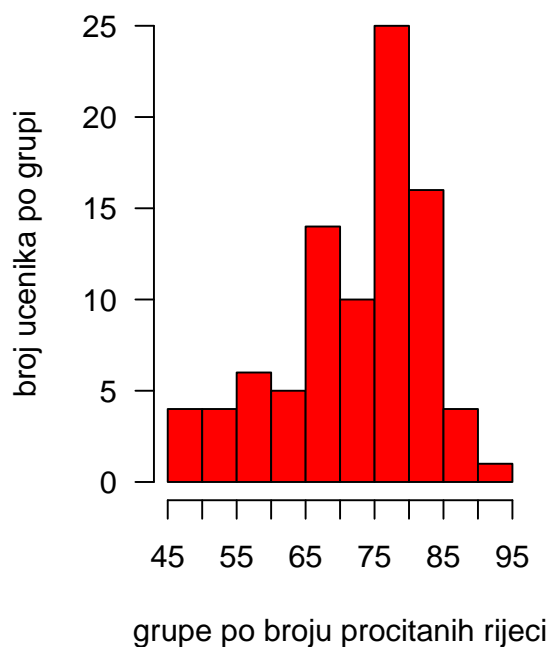


Dijagram kvantila brzine citanja

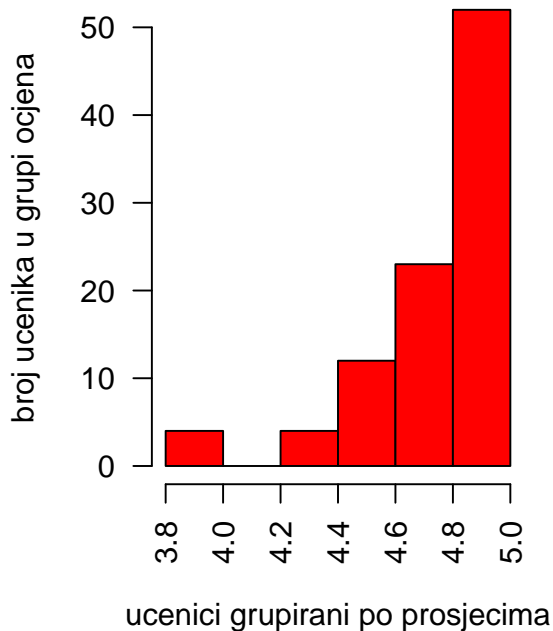


Usporedba distribucija brzine čitanja i ocjena

Stupcasti dijagram brzine citanj



Stupcasti dijagram svih ocjena



Dijagrami su vizualni indikator da postoji nesklad između distribucije prosjeka ocjena i distribucije brzine čitanja.

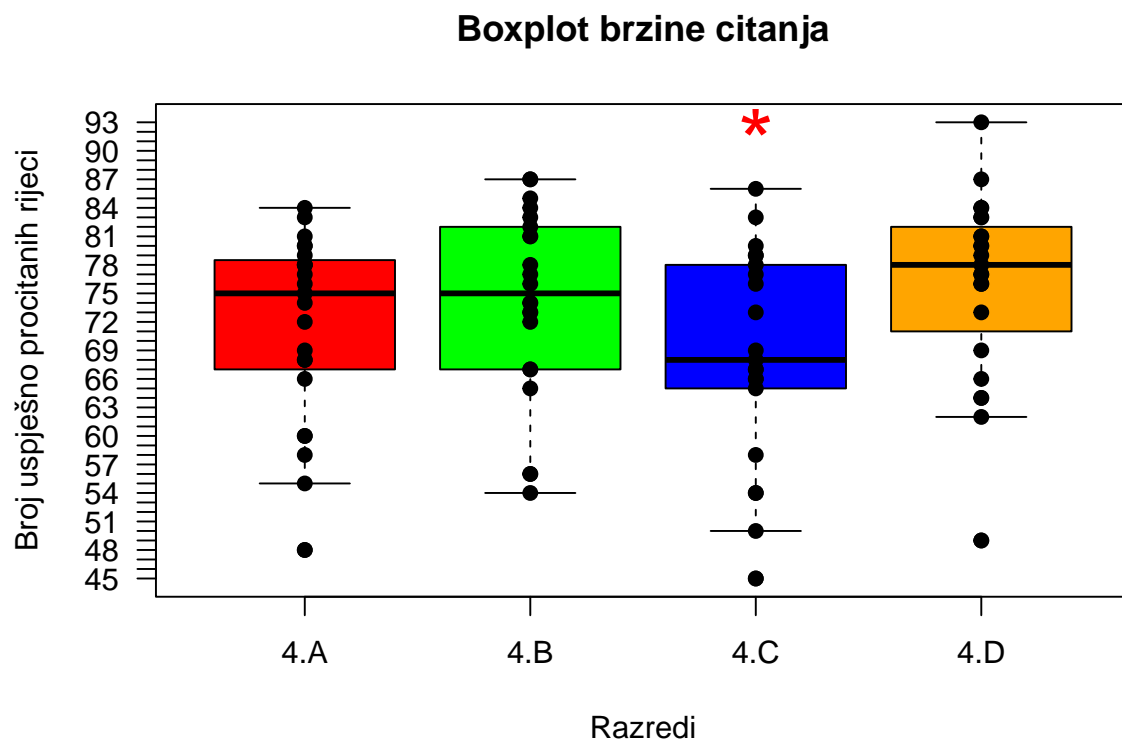
Zbog određivanja prikladnih vrsta statističkih testiranja, bitno je odrediti formiraju li podatci normalnu distribuciju.

Slijedi ponovni prikaz box-plotova i prikaza testa za normalnost.

```
##  
##  Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data:  svi_citanje$Broj.riječi  
## D = 0.14358, p-value = 0.0001065
```

Zaključak: Zbog malene P-vrijednosti Lillieforseovog testa, podatci vjerojatno nisu normalno distribuirani.

Za testiranje ovisnosti brzine čitanja o pripadnosti razredu, koristimo analizu varijance.



Crvena zvjezdica iznad trećeg stupca označava da on odstupa od ostalih u pogledu ovisnosti brzine čitanja (slabiji rezultat)

Rezultati neparametarskog testa:

```
kw_result
```

```
##  
##  Kruskal-Wallis rank sum test  
##  
## data:  data_kw by groups  
## Kruskal-Wallis chi-squared = 5.8965, df = 3, p-value = 0.1168
```

Na razini značajnosti $(1 - \alpha) = 0.88$ (drugim riječima, s 88% sigurnosti) odbacuje se hipoteza da NEMA statističke razlike između medijana brzine čitanja svakog razreda!

Zaključak: uspjeh učenika u pogledu brzine čitanja u nekoj mjeri ovisi i o razrednoj okolini.

Sljedeće, želimo odrediti postoji li statistički značajna korelacija između brzine čitanja i prosjeka ocjena. Naslućujemo kauzalnost, a pozitivnim rezultatima definirali bismo i asocijativnost između te dvije vrijednosti.

```
##
## Spearman's rank correlation rho
##
## data:  x_a[-22] and y_a
## S = 1714.8, p-value = 0.4865
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.152757

##
##      * * *
##
##

##
## Spearman's rank correlation rho
##
## data:  x_populacija and y_populacija
## S = 84237, p-value = 0.007212
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.2829657
```

Zaključak: Uz blage razlike ovisno o razredu, postoji velika korelacija između brzine čitanja i prosjeka ocjena!

Promatranje samo MAT, HJ i PiD

Želimo usporediti korelacije: 1. prosjeka svih ocjena s brzinom čitanja 2. prosjeka Matematike, Hrvatskog jezika i Prirode i društva s brzinom čitanja

Ako se pokaže da je 2. korelacija značajno veća, to nas vodi k zaključku da su ova 3 predmeta najznačajniji indikatori brzine čitanja učenika.

Opravdano je micati oko 5% stršćih vrijednosti ako želimo 95% vjerne i pouzdane statističke zaključke. To ne krši tzv. pravilo 2σ dozvoljenog odstupanja, često primjenjivano u statističkoj analizi u društvenim znanostima. Bitno je naglasiti da nam priroda mjerenja i promjenjivost rezultata ovih podataka također daje za pravo poduzimati takve mjere.

Ako zanemarimo blagu ovisnost ocjena o razredu, korelacija prosjeka ocjena 3 glavna predmeta i brzine čitanja je:

```
##
## Pearson's product-moment correlation
##
## data:  x_svi_1 and y_svi_1
## t = 5.0477, df = 79, p-value = 2.802e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3087761 0.6428619
## sample estimates:
##          cor
## 0.4938309
```

Korelacija prosjeka ocjena svih ocjena i broja riječi je:

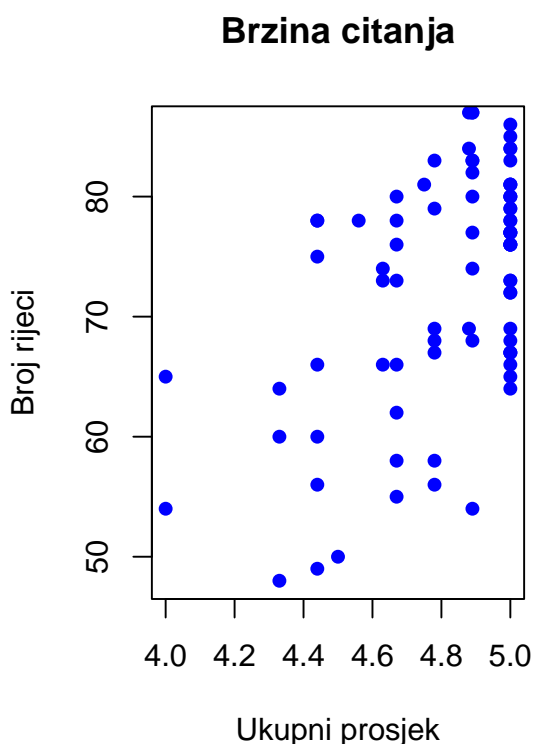
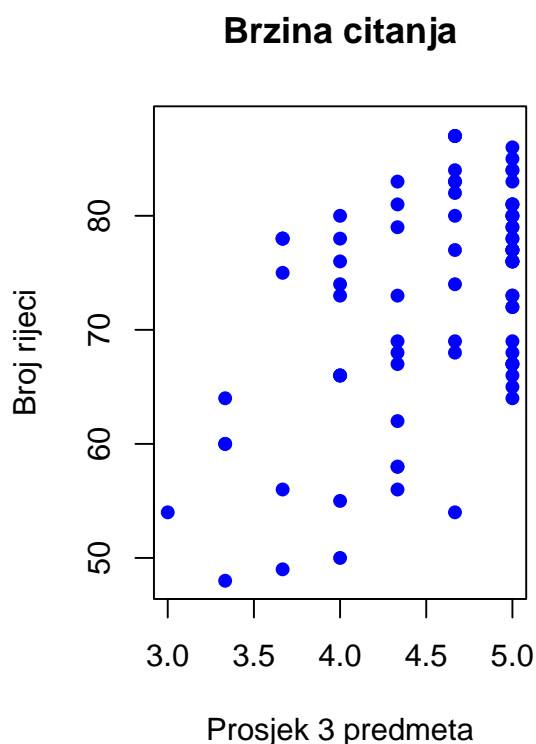
```
##
## Pearson's product-moment correlation
##
## data:  x_svi_2 and y_svi_2
## t = 5.4045, df = 79, p-value = 6.693e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3397363 0.6627147
## sample estimates:
##          cor
## 0.5195452
```

Ako zbog rezultata analize varijance želimo računati korelaciju samo za razredno-neovisne učenike, rezultat je sljedeći:

```
##
## Pearson's product-moment correlation
##
## data:  x and y
## t = 3.839, df = 59, p-value = 0.0003038
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2200122 0.6281658
## sample estimates:
##      cor
## 0.447064
```

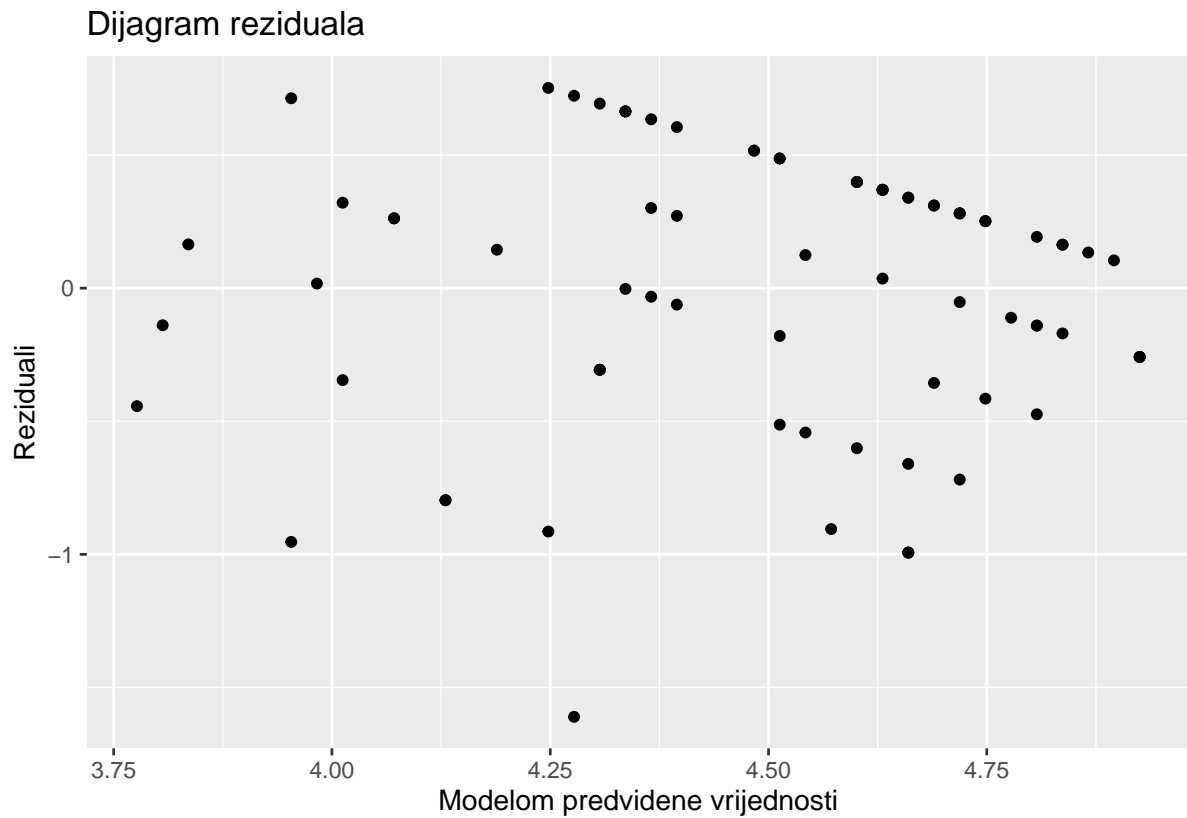
Nakon što smo izbacili stršeće vrijednosti, korelacije prema Pearsonovom koeficijentu vrlo su slične u svim navedenim slučajevima, oko $r = 0,5$. Takva vrijednost jasan je pokazatelj da za učenika koji brže čita u pravilu možemo očekivati i viši prosjek ocjena.

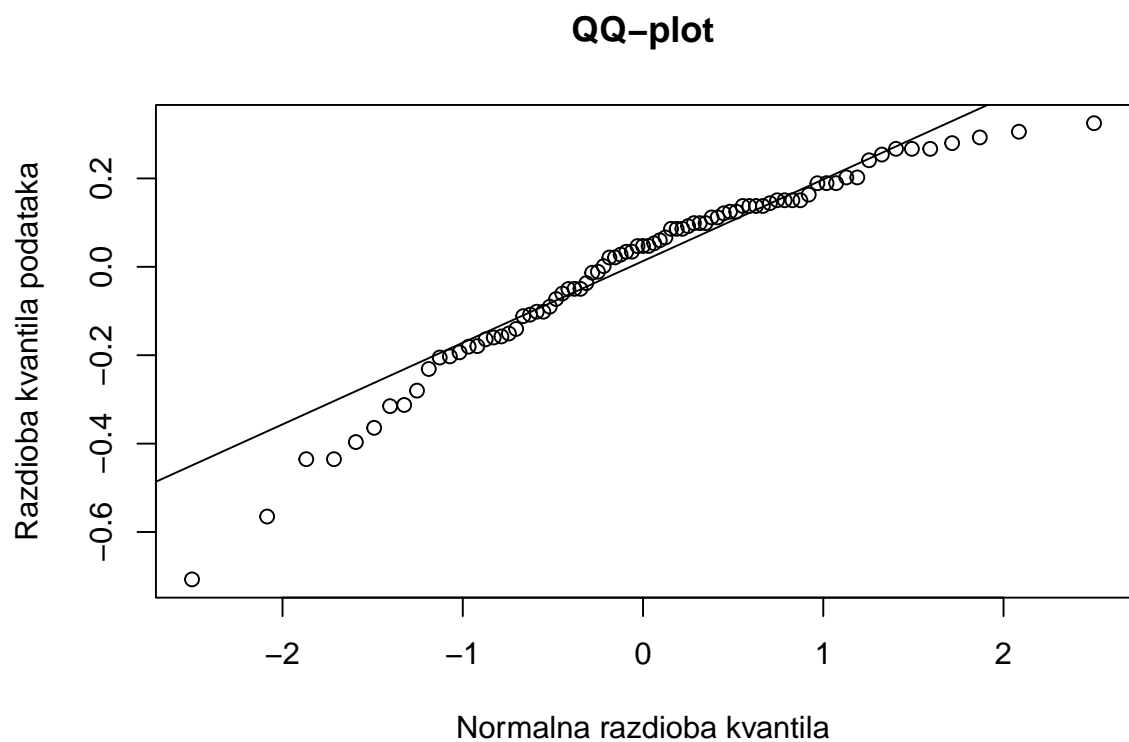
Usporedba korelacija:



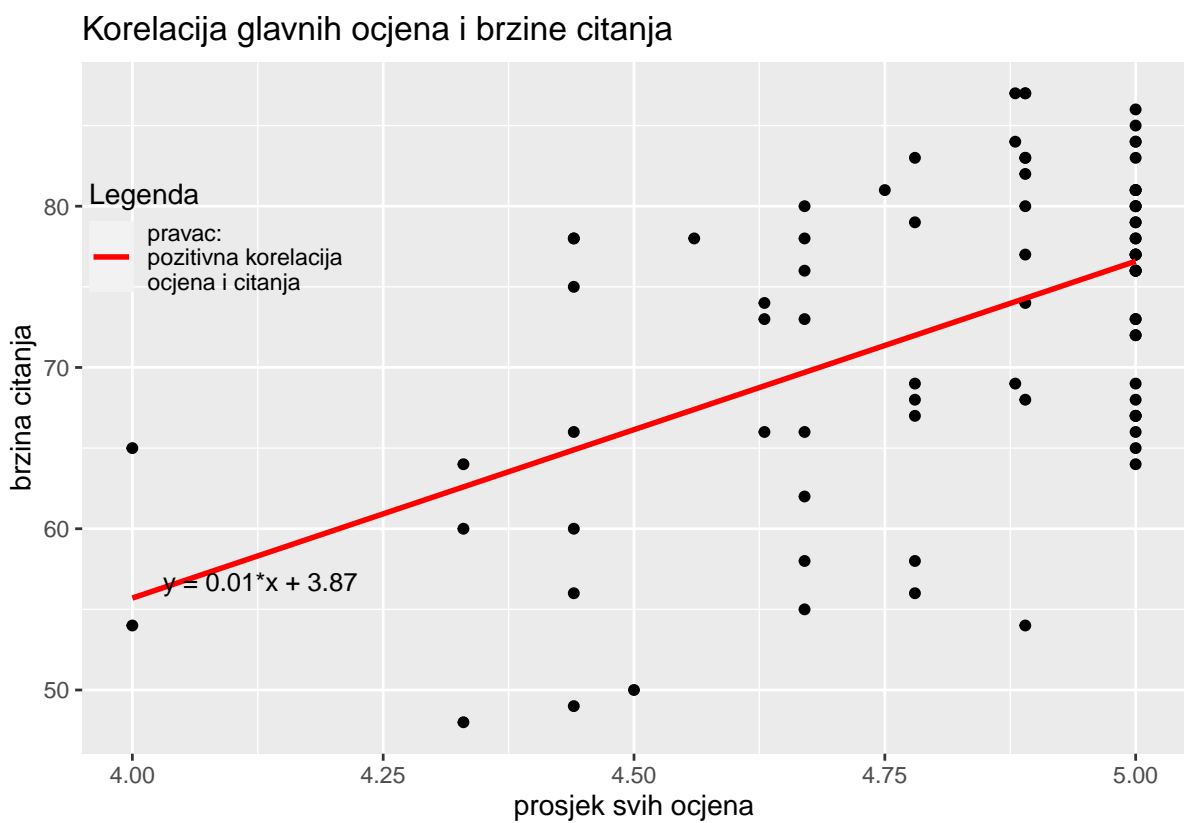
Linearna regresija

Pokušat ćemo međuovisnost čitanja i prosjeka ocjena modelirati linearnom regresijom. Prije toga trebamo provjeriti uvjete za linearnu regresiju, poput normalnosti reziduala. Slijede dijagram reziduala i QQ-plot:





Budući da se reziduali ne povećavaju porastom vrijednosti, kao niti da QQ-plot ne odstupa drastično od pravca, zaključujemo da smijemo izrađivati linearni model nad podatcima.



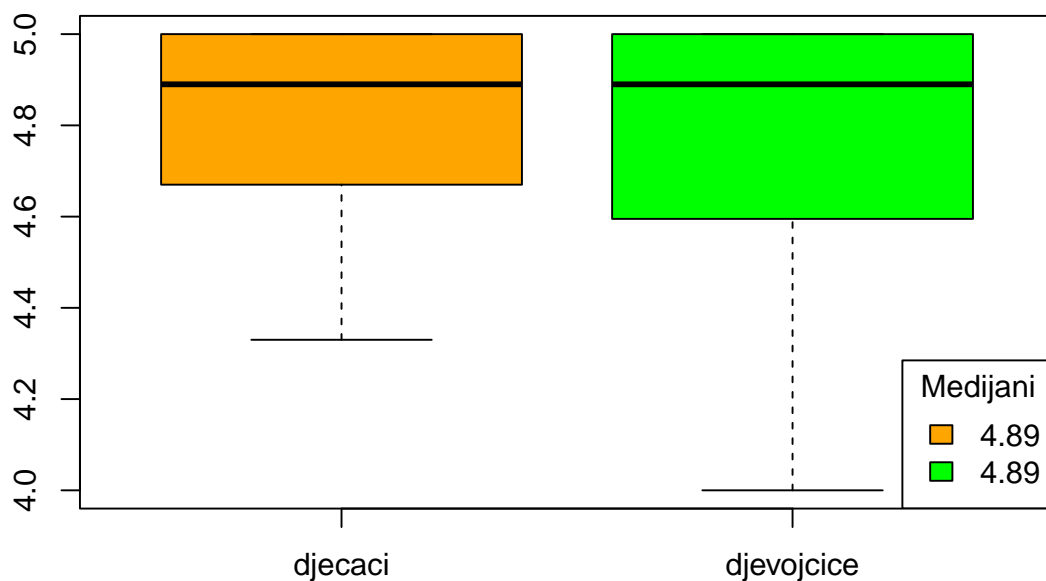
Razlika među dječacima i djevojčicama

Želimo testirati postoji li značajna razlika u ocjenama i u brzini čitanja između dječaka i djevojčica.

Test vezan za ocjene:

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: m_ocjene$`Srednja ocjena` and f_ocjene$`Srednja ocjena`  
## W = 986.5, p-value = 0.4624  
## alternative hypothesis: true location shift is greater than 0
```

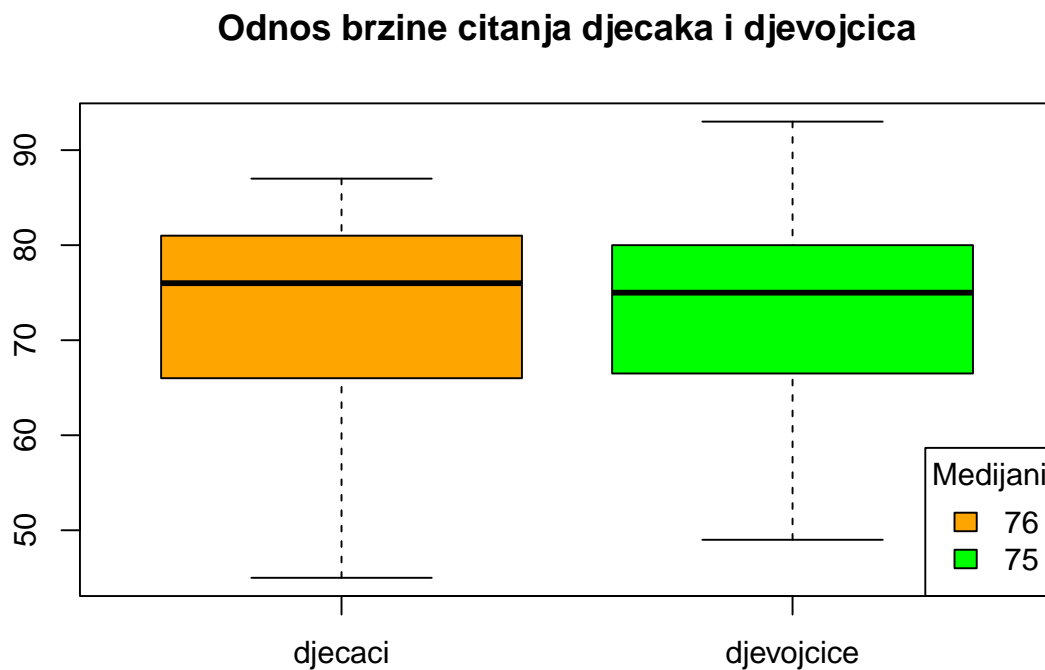
Odnos ocjena dječaka i djevojčica



Jako blaga razlika u prosjeku ocjena našem uzorku postoji, ali nedovoljno velika da se može zaključiti išta na ukupnoj populaciji učenika!

Test vezan za brzinu čitanja:

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: m_citanje$`Broj riječi` and f_citanje$`Broj riječi`  
## W = 987, p-value = 0.9242  
## alternative hypothesis: true location shift is not equal to 0
```



Zaključak: Statistički nema gotovo nikakve razlike u brzini čitanja dječaka i djevojčica!