

JEDNOSTAVNA LINEARNA REGRESIJA

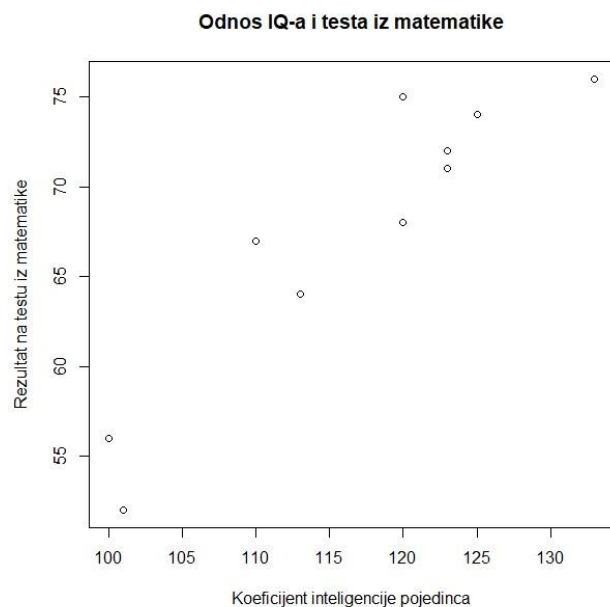
```
> TestIQ=data.frame(  
+ IQ=c(120,101,120,123,125,100,133,123,110,113),  
+ Test=c(68,52,75,71,74,56,76,72,67,64))
```

```
> TestIQ  
  IQ Test  
1 120  68  
2 101  52  
3 120  75  
4 123  71  
5 125  74  
6 100  56  
7 133  76  
8 123  72  
9 110  67  
10 113  64
```

#Podaci se odnose na IQ pojedinca i na ostvareni broj poena koje je pojedinac postigao na testu iz matematike

#Test predstavlja zavisnu promenjivu,a IQ nezavisnu(objasnjavajucu)promenjivu. Ispitivao sam jednostavnu linearnu regresiju izmedju ove dve velicine

```
> plot(Test~IQ,data=TestIQ,xlab="Koeficijent inteligencije pojedinca",ylab="Rezultat na testu iz matematike",main="Odnos IQ-a i testa iz matematike")
```



#Pomocu grafika mozemo zakljuciti da postoji linearna veza izmedju varijabli

```
> cov(TestIQ$Test,TestIQ$IQ)
```

[1] 80.11111

#Pomocu kovarijanse mozemo da zakljucimo da je veza izmedju promenljivih jaka i da se dve promenljive krecu u istom pravcu jer je $cov > 0$

```
> summary(TestIQ)
```

IQ	Test
Min. :100.0	Min. :52.00
1st Qu.:110.8	1st Qu.:64.75
Median :120.0	Median :69.50
Mean :116.8	Mean :67.50
3rd Qu.:123.0	3rd Qu.:73.50
Max. :133.0	Max. :76.00

#Posto su obe varijable kvantitativne vidimo tabelu koja prikazuje njihovu minimalnu i maksimalnu vrednost kao i medijanu, aritmeticku sredinu i prvi i treci kvartil

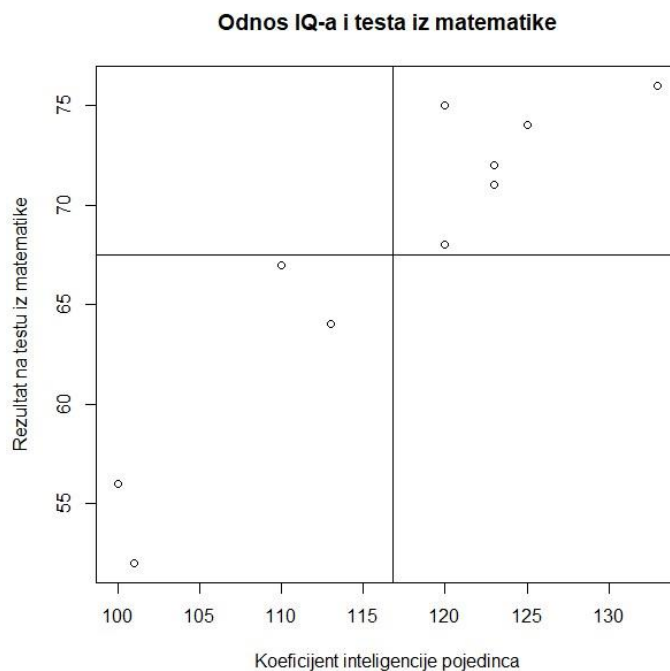
```
> TestSredina=mean(TestIQ$Test)
```

```
> IQSredina=mean(TestIQ$IQ)
```

```
> plot(Test~IQ,data=TestIQ,xlab="Koeficijent inteligencije pojedinca",ylab="Rezultat na testu iz matematike",main="Odnos IQ-a i testa iz matematike")
```

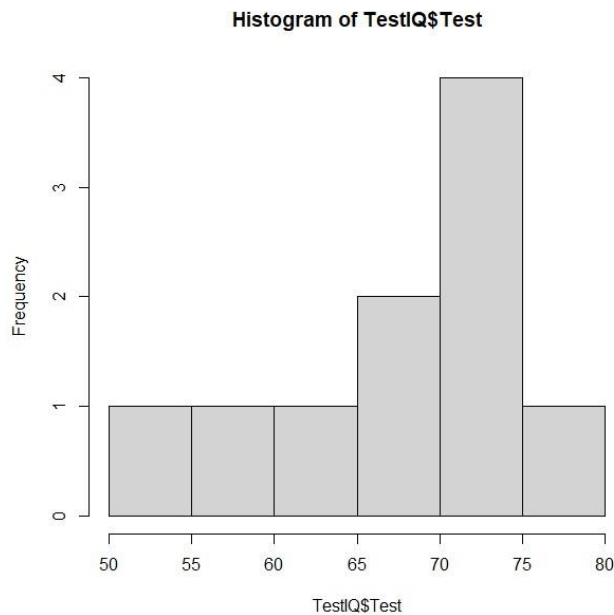
```
> abline(h=TestSredina)
```

```
> abline(v=IQSredina)
```



#Na grafiku su predstavljene aritmeticke sredine dve varijable, prava linearne regresije ce proci kroz njihov presek

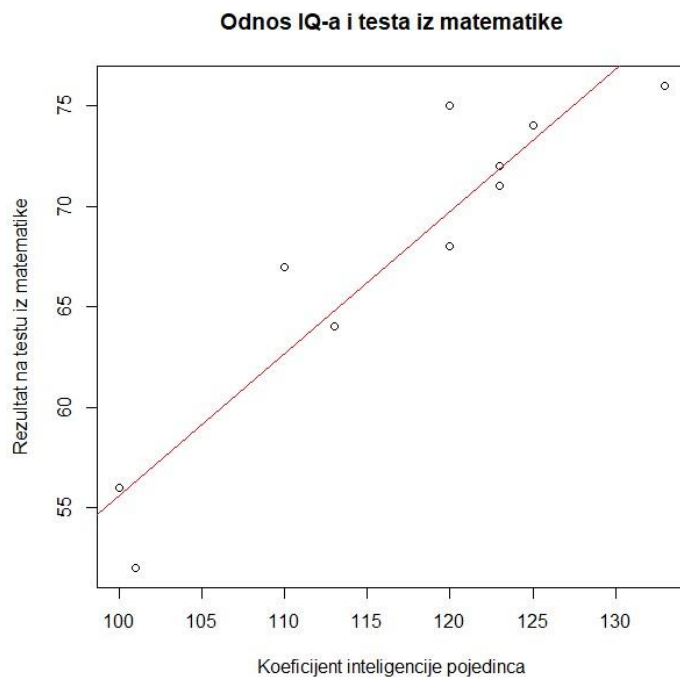
```
> hist(TestIQ$Test)
```



#Pomocu histograma mozemo proveriti da li zavisna promenjiva ima normalnu raspodelu, repovi su tezi od srednjeg dela raspodele pa mozemo da zakljucimo da zavisna promenjiva ima priblizno normalnu raspodelu

```
> IQTestlin=lm(Test~IQ,data=TestIQ)
#Kreiramo linearni model da bi mogli izvorsiti jednostavnu analizu linearne regresije
> coefficients(IQTestlin)
(Intercept)      IQ
-15.0939584  0.7071401
```

```
#Pomocu ove funkcije mozemo zapisati jednacinu linearne regresije y=-
15.0939584+0.7071401x
> plot(Test~IQ,data=TestIQ,xlab="Koeficijent inteligencije pojedinca",ylab="Rezultat na testu iz
matematike",main="Odnos IQ-a i testa iz matematike")
> abline(IQTestlin, col="red")
```



#Is crtavamo pravu linearne regresije na grafikonu

```
> summary(IQTestlin) Call:
lm(formula = Test ~ IQ, data =
TestIQ)
```

Residuals:

Min	1Q	Median	3Q	Max
4.3272	-1.5432	-0.3486	0.6211	5.2372

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-15.0940	11.5121	-1.311	0.226
IQ	0.7071	0.0982	7.201	9.23e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.136 on 8 degrees of freedom

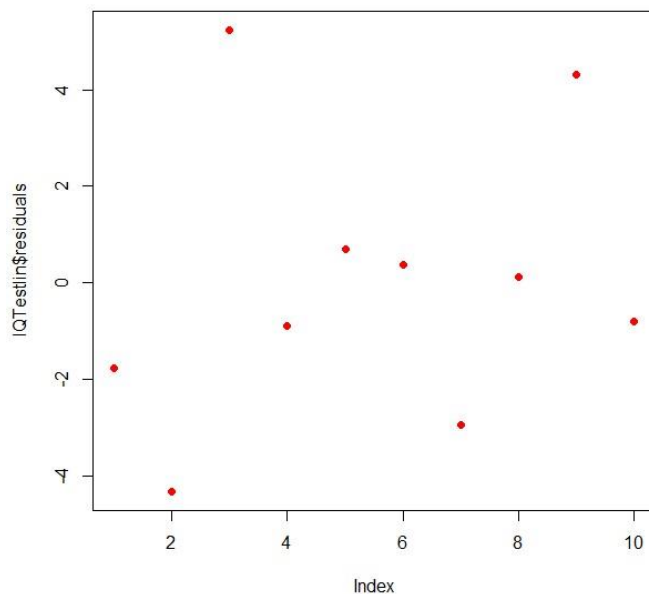
Multiple R-squared: 0.8664, Adjusted R-squared: 0.8496 F-statistic: 51.86 on 1 and 8 DF, p-value: 9.233e-05

#U tabeli reziduala je prikazan minimalni i maksimalni rezidual, medijana reziduala kao i vrednosti prvog i treceg kvartila.

#U dobrim modelima medijana treba da bude blizu nule, a minimum treba da bude sto veci a maksimum sto manji

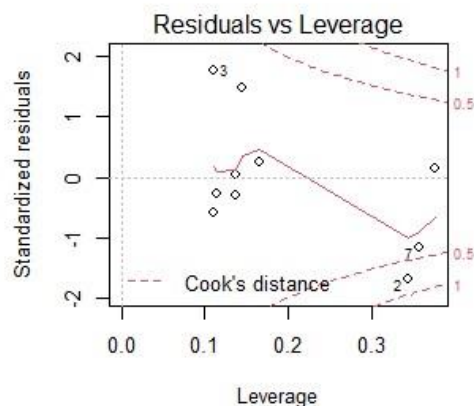
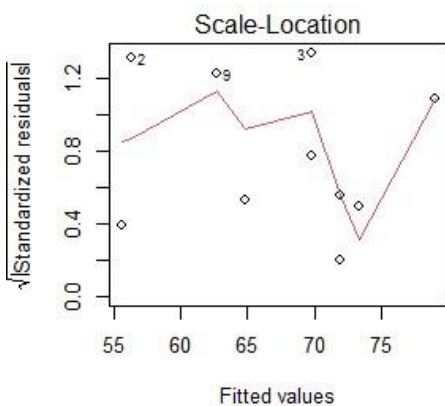
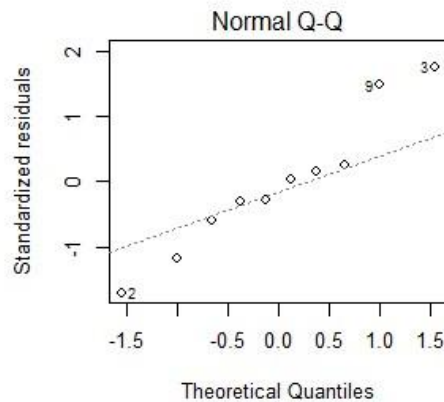
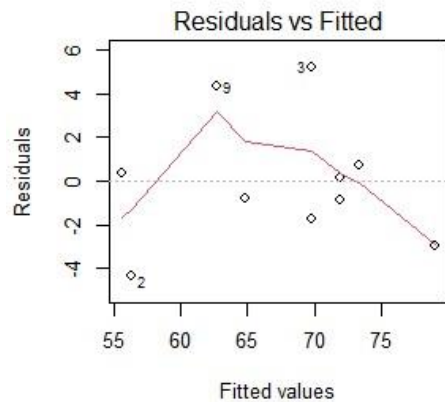
#U tabeli koeficijenata intercept predstavlja slobodan koeficijent, a IQ predstavlja koeficijent uz nezavisnu promenjivu u jednacini linearne regresije
 #p vrednost pokazuje da li se prihvata ili odbacuje nulta hipoteza da objasnjavajuca varijabla nije znacajna za model, a posto je p vrednost manja od 0,05 (iznosi 0,0000923) znaci da je IQ znacajna za model i nulta hipoteza se odbacuje
 #Sto je standardna greska blize nuli to je bolje, a standardna greska reziduala je 3,136
 #t vrednost treba da bude veka od 1,96, a u modelu iznosi 7,201
 #R-squared bi trebao da bude sto blize 1 za modele koji se dobro uklapaju u podatke, a veci od 0,70 je povoljan, a u modelu iznosi 0,8664 sto znaci da se model dobro uklapa u podatke
 #Adjusted R-squared bi trebao da bude sto blize 1, a u modelu iznosi 0.8496

```
> plot(IQTestlin$residuals, pch = 16, col = "red")
```



#Reziduali su rasporedjeni nasumicno sto je dobro i sto ukazuje na to da ne postoji skriveni obrazac koji linearni model ne razmatra

```
> par(mfrow=c(2,2))
> plot(IQTestlin)
```



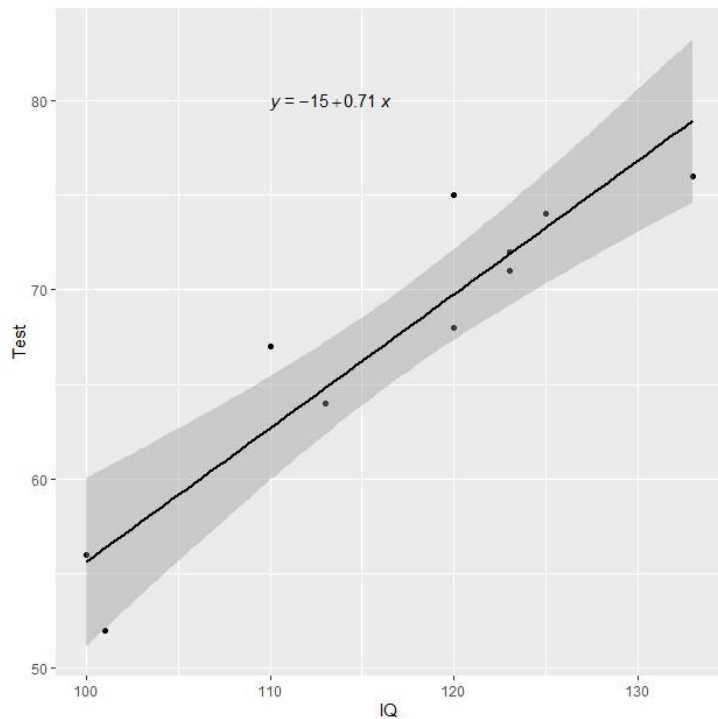
#Grafik Residuals vs Fitted daje informaciju o rezidualima. Što je model bolji funkcija prikazana na ovom grafiku bi trebala da se stabilizuje oko horizontalne linije koja sece y osu u tacki 0.

#Posto su crvene linije koje predstavljaju srednju vrednost reziduala centrirane oko nule i imaju horizontalni pravac znaci da u podacima nema vecih odstupanja ili pristrasnosti koje bi linearnu regresiju ucinile nevazecom

#Kod Normal Q-Q dijagrama stvarni reziduali ne odstupaju previse od teorijskih reziduala prikazani isprekidanom linijom, da su reziduali normalno # distribuirani, svi bi lezali na isprekidanoj liniji koja je prikazana na grafiku

#Mozemo da zakljucimo da ovaj model ispunjava pretpostavku homoskedasticnosti

```
> install.packages("tidyverse")
> library(ggplot2)
> install.packages("ggpubr")
> library(ggpubr)
> IQTest.graph<-
ggplot(TestIQ,aes(x=IQ,y=Test))+geom_point()+geom_smooth(method="lm",col
="black")+stat_regline_equation(label.x = 110, label.y = 80)
> IQTest.graph
```



#Grafik sa podacima prikazani tackama i sa linijom linearne regresije i sa standardnom greskom koja je prikazana svetlosivom trakom koja okružuje liniju i sa jednačinom linearne regresije ($y = -15 + 0.71x$)

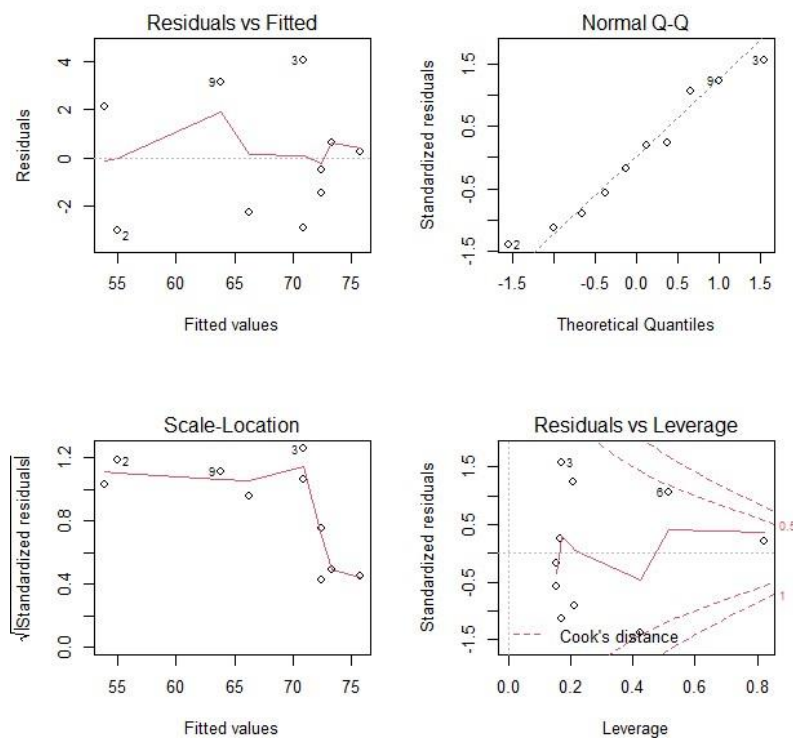
#Postoji znacajna veza izmedju IQ-a i rezultata testa iz matematike ($p < 0,001$)

```
> IQTestlin2=lm(Test~IQ+I(IQ^2),data=TestIQ)
```

#Primena transformacije dodavanjem nezavisnog kvadratnog clana

```
> par(mfrow=c(2,2))
```

```
> plot(IQTestlin2)
```



#Mozemo da zakljucimo da je ova transformacija poboljsala model jer se funkcija na grafiku Residual vs Fitted priblizila isprekidanoj liniji sto znaci da je model poboljsan
 #Kod Normal Q-Q grafika svi reziduali se nalaze blize isprekidane linije i nema vecih odstupanja tako da se moze zakljuciti da se povecala normalnost raspodele

```
> coefficients(IQTestlin2)
(Intercept)      IQ      I(IQ^2)
-206.31654503  4.06000960 -0.01458016
```

#Mozemo da vidimo da je jednacina krive linearne regresije $y = -206.31655 + 4.06x - 0.01458x^2$

```
> summary(IQTestlin2) Call: lm(formula = Test
~ IQ + I(IQ^2), data = TestIQ)
```

Residuals:

Min	1Q	Median	3Q	Max
3.0123	-2.0883	-0.1189	1.7454	4.0696

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.063e+02	1.165e+02	-1.771	0.1199
IQ	4.060e+00	2.036e+00	1.994	0.0864
I(IQ^2)	-1.458e-02	8.847e-03	-1.648	0.1433

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.845 on 7 degrees of freedom

Multiple R-squared: 0.9037, Adjusted R-squared: 0.8762 F-statistic: 32.85 on 2 and 7 DF, p-value: 0.000277

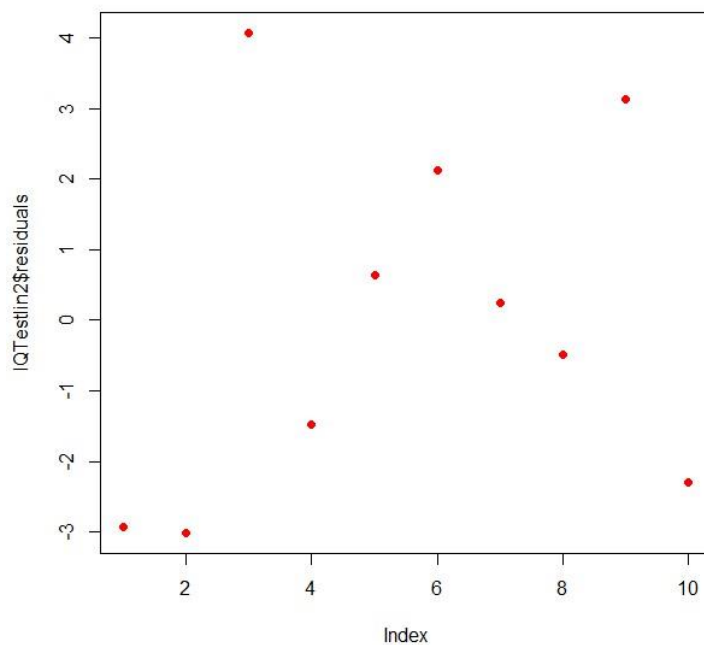
#Kod reziduala medijana se priblizila nuli, minimum se povecao i priblizio nuli a maksimum se smanjio sto ukazuje na to da se model poboljsao

#Smanjila se standardna greska reziduala sa 3,136 na 2,845

#Multiple R-squared i Adjusted R-squared su se povecali i priblizili jedinici

#Mozemo da zakljucimo da je model poboljsan

```
> plot(IQTestlin2$residuals, pch = 16, col = "red")
```



#Reziduali su rasporedjeni malo neravnomjernije sa ovom transformacijom sto je dobro i sto ukazuje na to da ne postoji skriveni obrazac koji linearni model ne razmatra

```
> AIC(IQTestlin)
```

```
[1] 55.00325
```

```
> AIC(IQTestlin2)
```

```
[1] 53.72452
```

#Dobili smo poboljsanje jer se AIC smanjio sa 55,00325 na 53,72452, ali poboljsanje nije znacajno

```
> anova(IQTestlin,IQTestlin2)
```

Analysis of Variance Table

Model 1: Test ~ IQ

Model 2: Test ~ IQ + I(IQ^2)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	8	78.652				
2	7	56.665	1	21.987	2.7161	0.1433