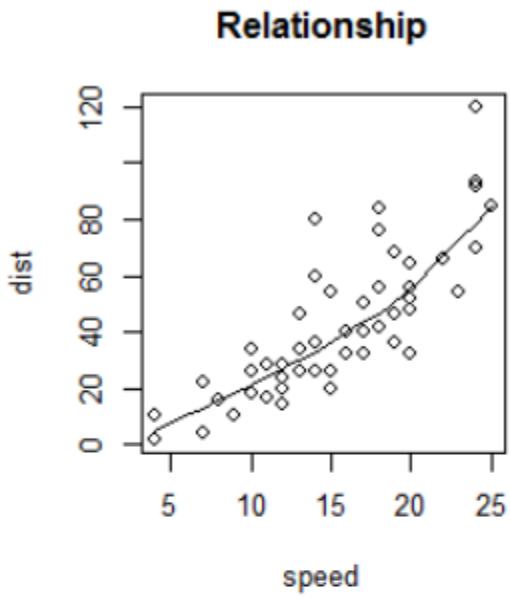
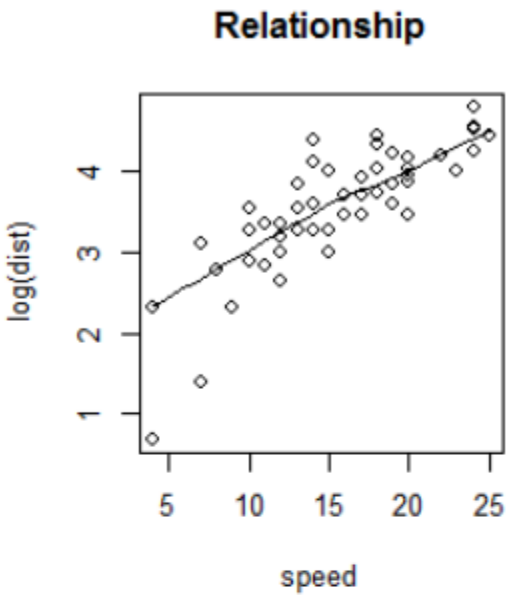
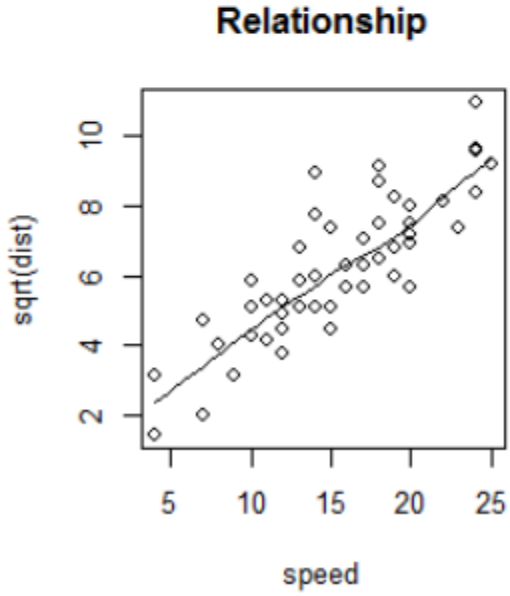
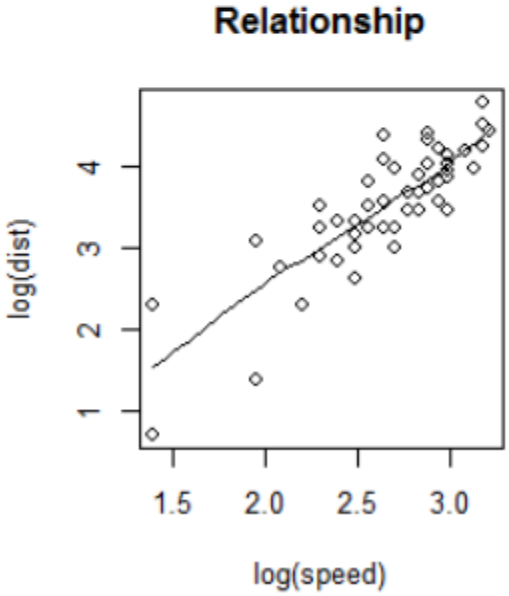
Razvoj novog modela - transformacija sedd i dist - Rattle









Prvi model : dist~speed

Summary of the Linear Regression model (built using lm):
Residuals:

Min	1Q	Median	3Q	Max
-18.47	-6.48	-2.44	4.46	45.79

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-21.097	6.776	-3.11	0.0038 **
speed	3.971	0.418	9.49	5.9e-11 ***

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

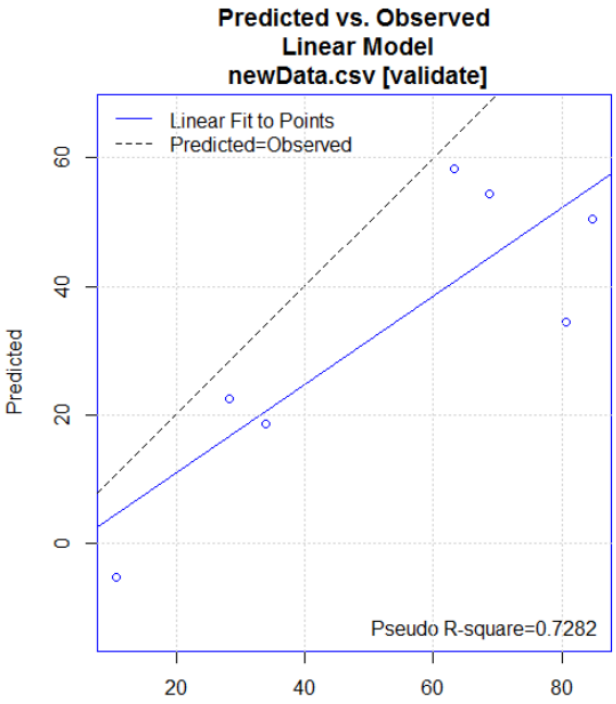
Residual standard error: 12.9 on 33 degrees of freedom
Multiple R-squared: 0.732, Adjusted R-squared: 0.724
F-statistic: 90.1 on 1 and 33 DF, p-value: 5.86e-11

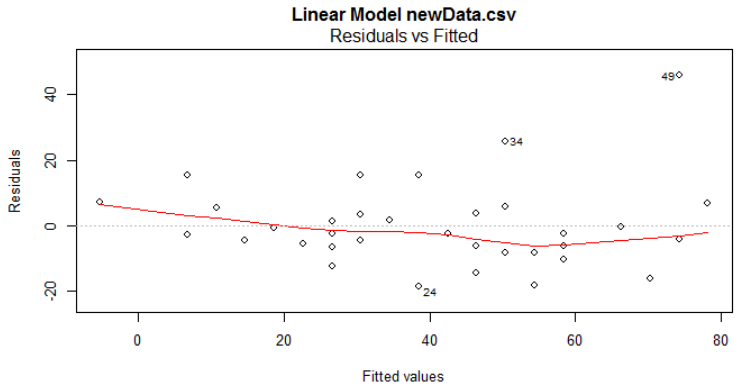
==== ANOVA ====
Analysis of Variance Table

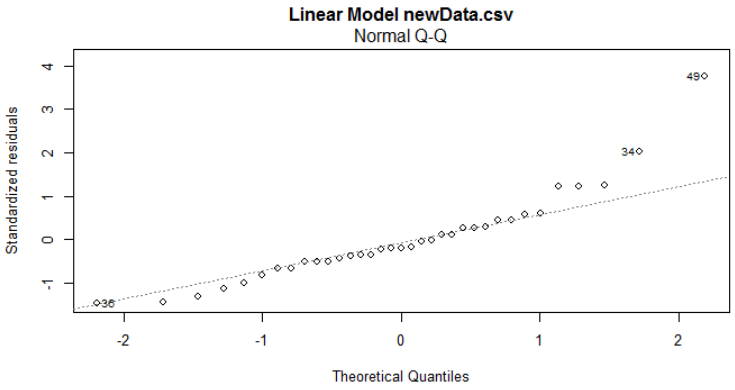
Response: dist

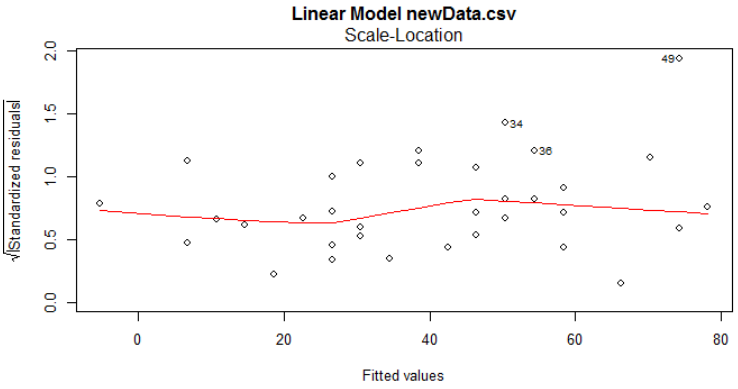
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
speed	1	14915	14915	90.1	5.9e-11 ***
Residuals	33	5462	166		

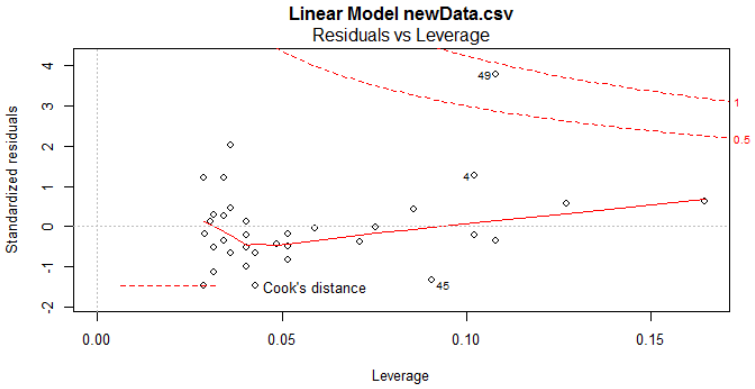
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] "\n"
Time taken: 0.02 secs
Rattle timestamp: 2018-02-01 01:20:26 admin











Sqrt(dist)~speed

Summary of the Linear Regression model (built using lm):

Residuals:

Min	1Q	Median	3Q	Max
-1.409	-0.597	-0.106	0.499	2.034

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.8150	0.4816	1.69	0.1	
speed	0.3377	0.0297	11.36	6.1e-13	***

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

Residual standard error: 0.914 on 33 degrees of freedom
Multiple R-squared: 0.796, Adjusted R-squared: 0.79
F-statistic: 129 on 1 and 33 DF, p-value: 6.05e-13

==== ANOVA ====

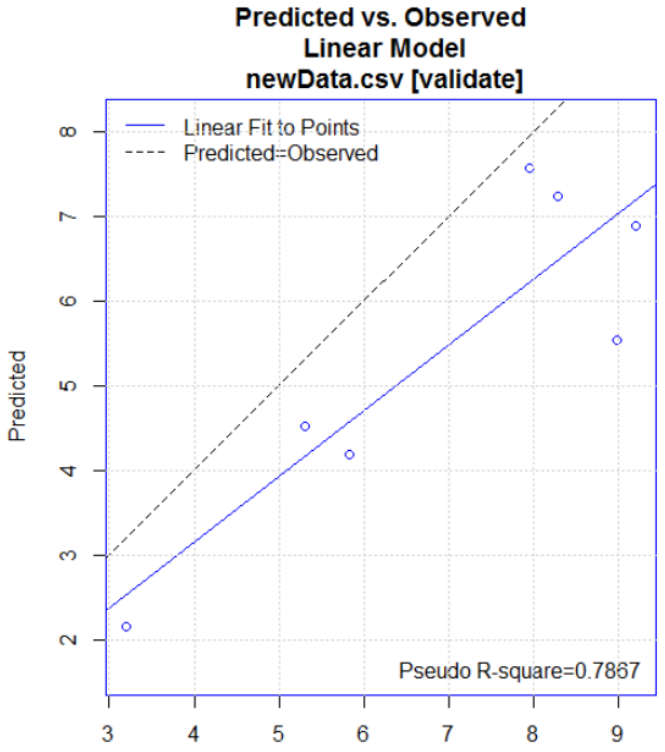
Analysis of Variance Table

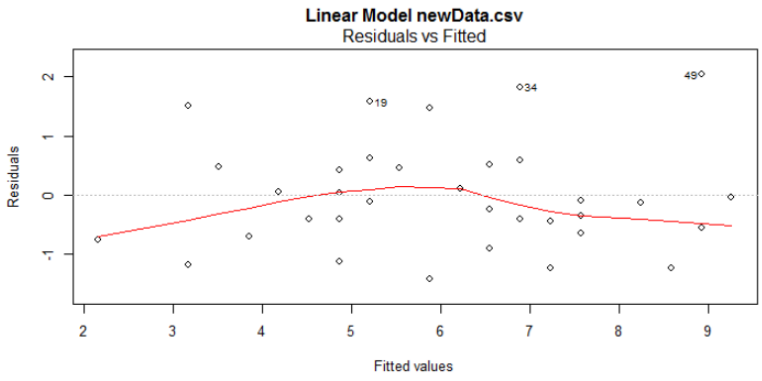
Response: sqrtDist

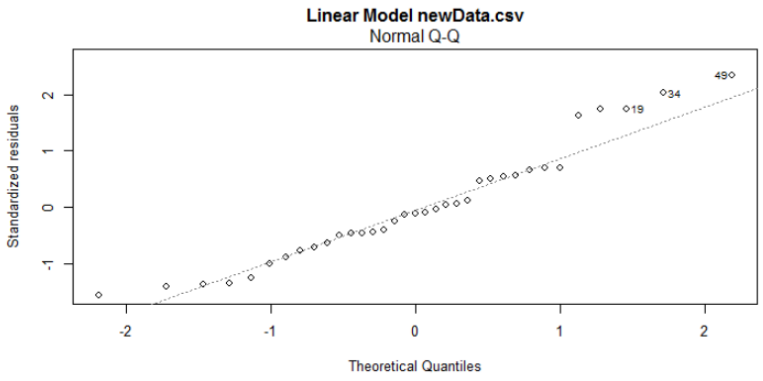
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
speed	1	107.9	107.9	129	6.1e-13	***
Residuals	33	27.6	0.8			

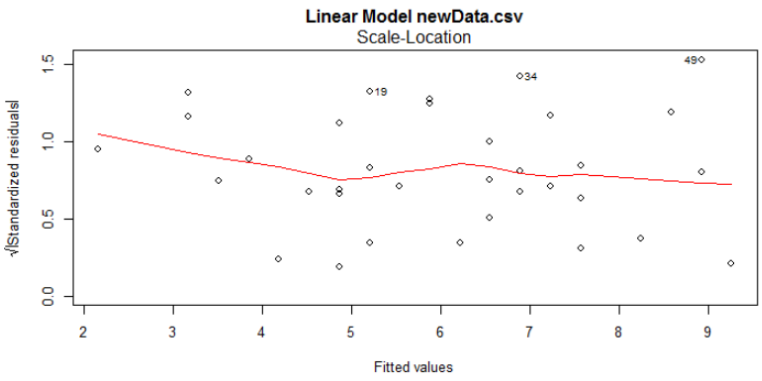
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] "\n"
Time taken: 0.02 secs

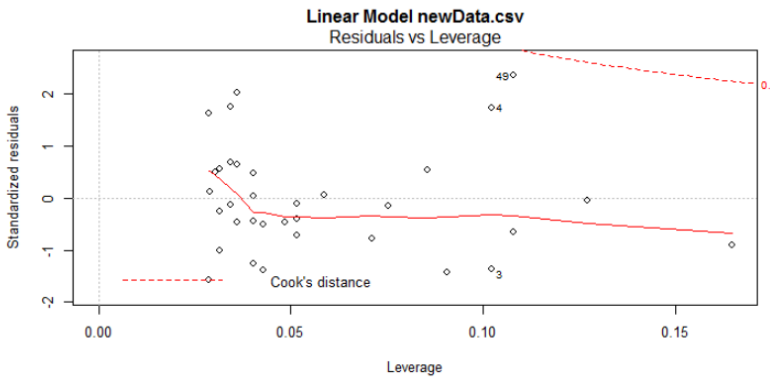
Rattle timestamp: 2018-02-01 00:51:46 admin











log(dist)~log(speed)

Summary of the Linear Regression model (built using lm):
Residuals:

Min	1Q	Median	3Q	Max
-0.748	-0.167	-0.034	0.194	0.957

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.433	0.400	-3.59	0.0011	**
logSpeed	1.833	0.148	12.35	6.4e-14	***

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

Residual standard error: 0.351 on 33 degrees of freedom
Multiple R-squared: 0.822, Adjusted R-squared: 0.817
F-statistic: 152 on 1 and 33 DF, p-value: 6.44e-14

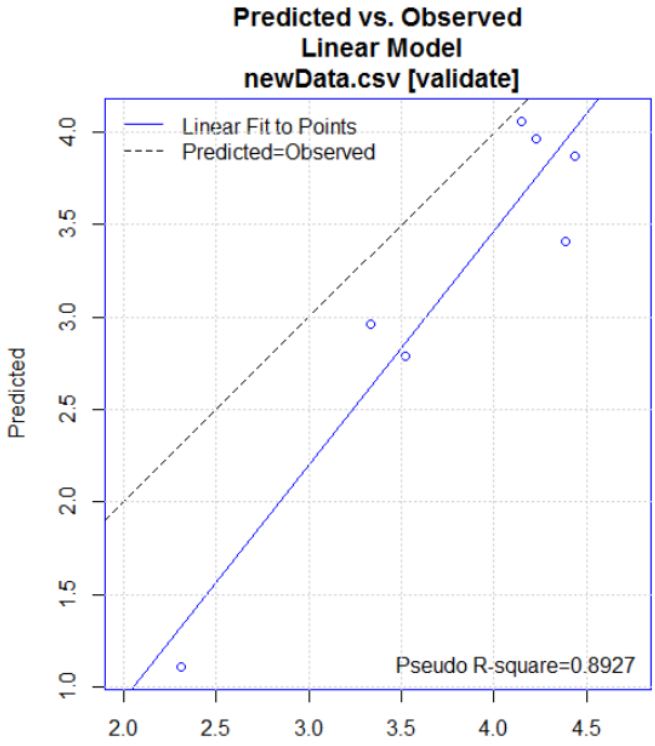
==== ANOVA ====

Analysis of Variance Table

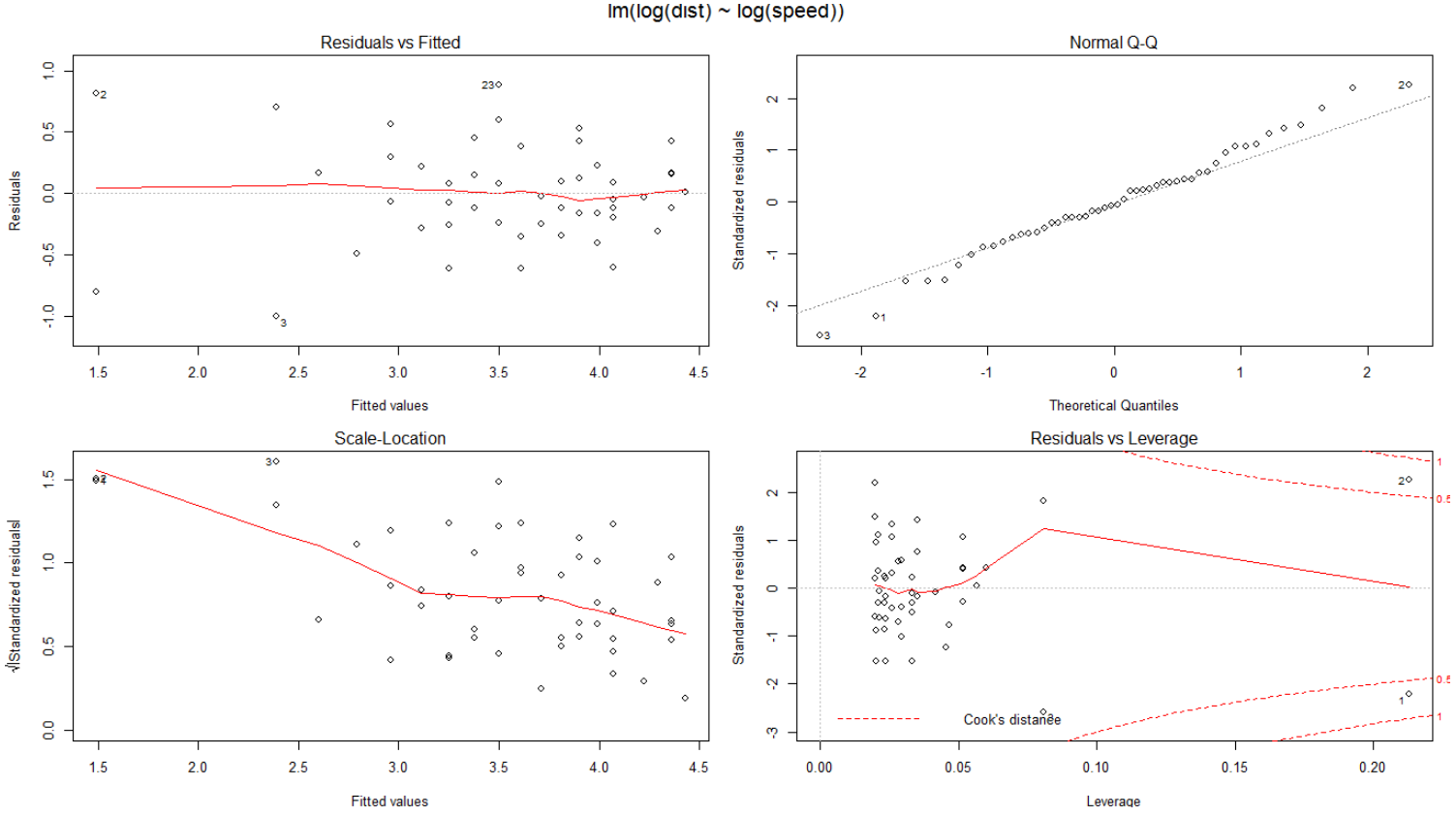
Response: logDist

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
logSpeed	1	18.74	18.74	152	6.4e-14	***
Residuals	33	4.06	0.12			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] "\n"
Time taken: 0.02 secs



Izvor : https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/cars.html



dist~speed^2

Summary of the Linear Regression model (built using lm):
Residuals:

Min	1Q	Median	3Q	Max
-20.48	-6.44	-1.96	5.56	39.41

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.714	4.001	1.43	0.16
powSpeed	0.130	0.013	10.00	1.6e-11 ***

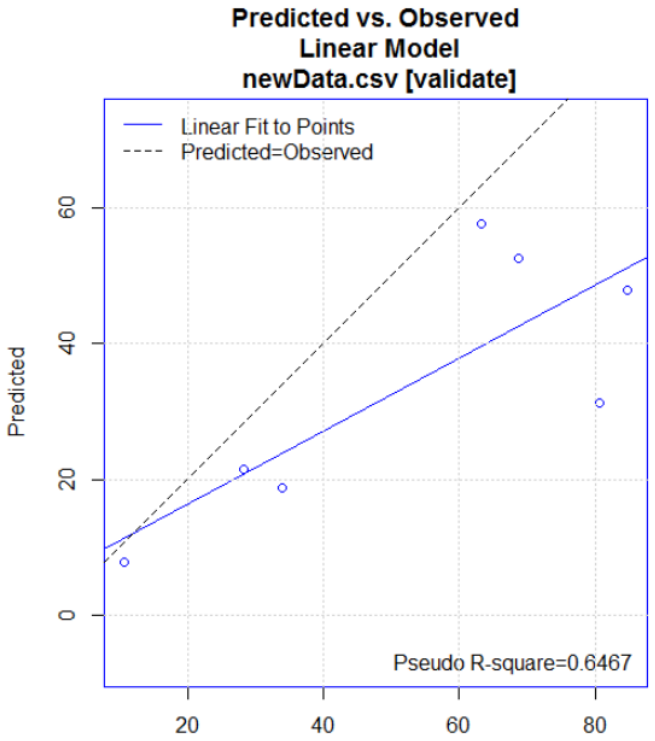
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

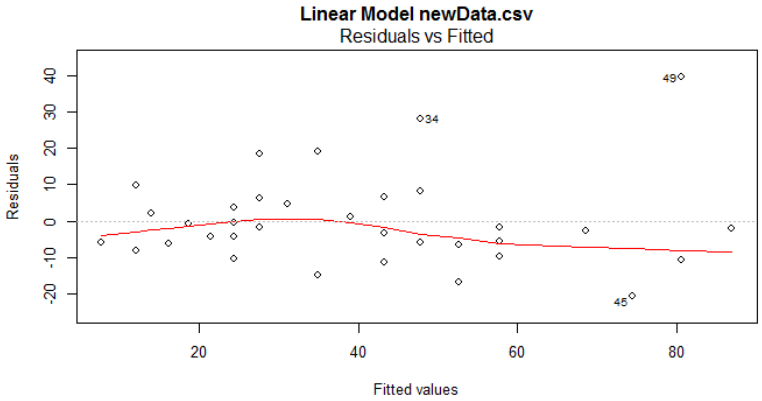
Residual standard error: 12.4 on 33 degrees of freedom
Multiple R-squared: 0.752, Adjusted R-squared: 0.744
F-statistic: 100 on 1 and 33 DF, p-value: 1.61e-11

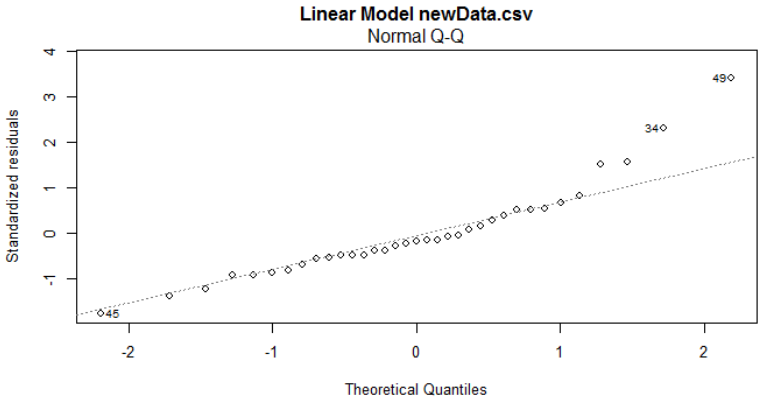
==== ANOVA ====
Analysis of Variance Table
Response: dist

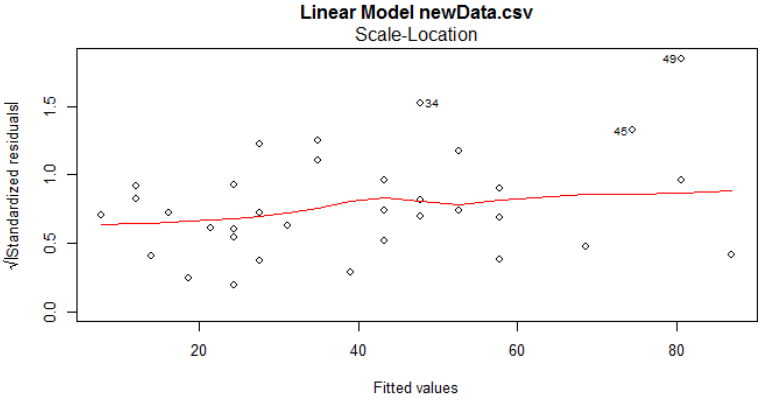
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
powSpeed	1	15323	15323	100	1.6e-11 ***
Residuals	33	5054	153		

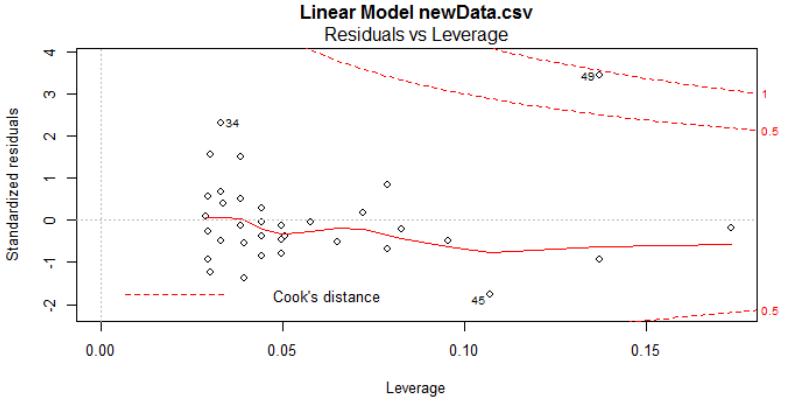
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] "\n"
Time taken: 0.03 secs
Rattle timestamp: 2018-02-01 02:27:16 admin







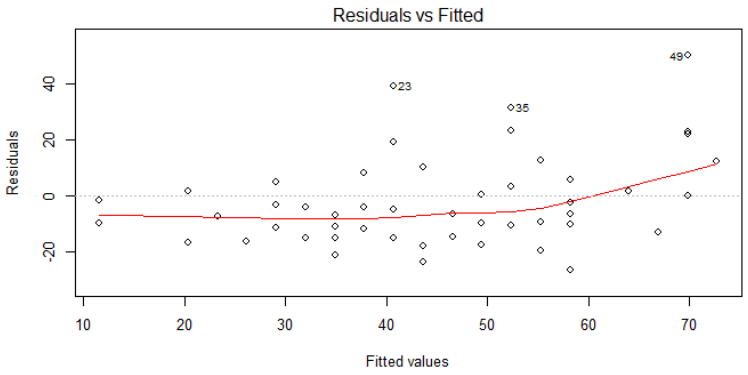


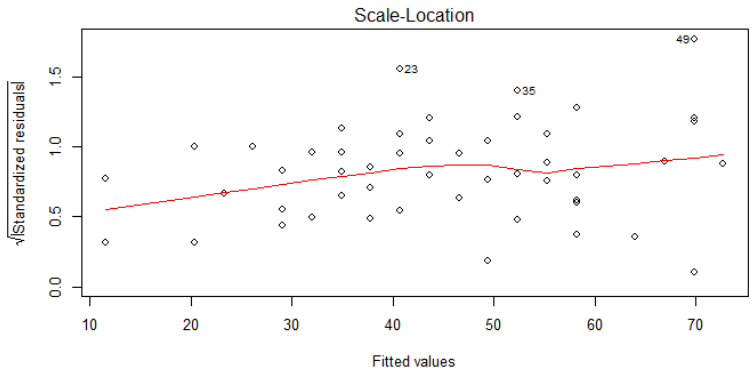


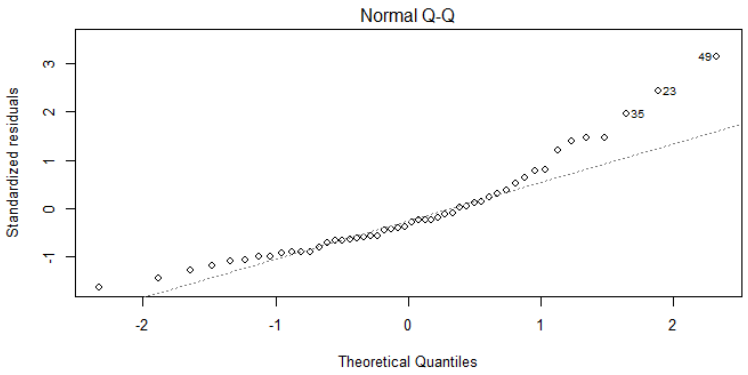
- Razvoj modela linearne regresije u programskom okruženju R

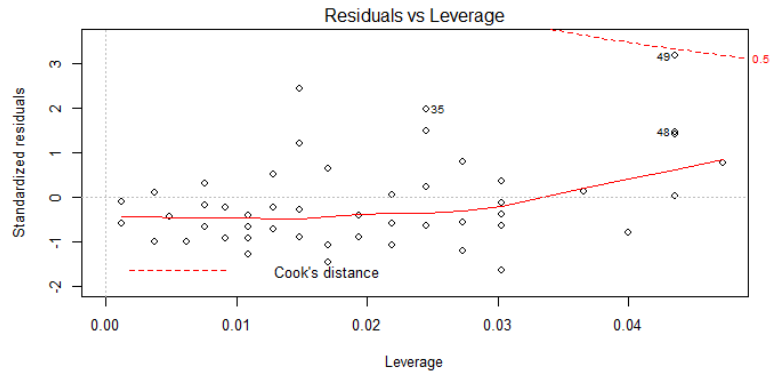
R sintaksa	Model	Komentar
Y ~ A	$Y = \beta_0 + \beta_1 \cdot A$	
Y ~ -1 + A	$Y = \beta_1 \cdot A$	
Y ~ A + I(A^2)	$Y = \beta_0 + \beta_1 \cdot A + \beta_2 \cdot A^2$	

```
> y = lm(formula = dist ~ (-1+speed), data=cars)
>
>   summary(y)

Call:
lm(formula = dist ~ (-1 + speed), data = cars)

Residuals:
   Min     1Q Median     3Q    Max
-26.18 -12.64  -5.46   4.59  50.18

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
speed    2.909      0.141    20.6   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.3 on 49 degrees of freedom
Multiple R-squared:  0.896,    Adjusted R-squared:  0.894
F-statistic:  423 on 1 and 49 DF,  p-value: <2e-16

>   anova(y)
Analysis of Variance Table

Response: dist
          Df Sum Sq Mean Sq F value Pr(>F)
speed      1 111949  111949     423 <2e-16 ***
Residuals 49  12954     264
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

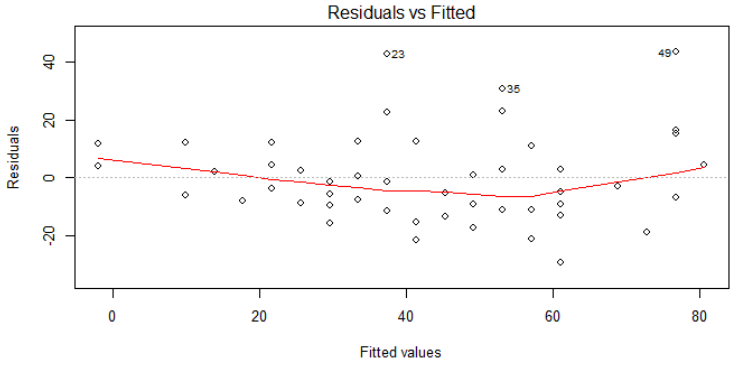


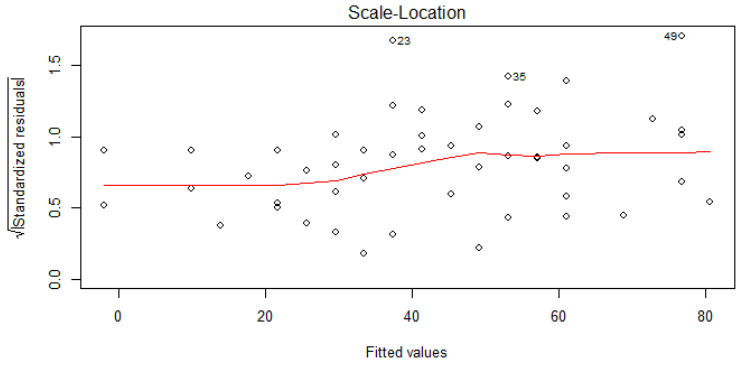


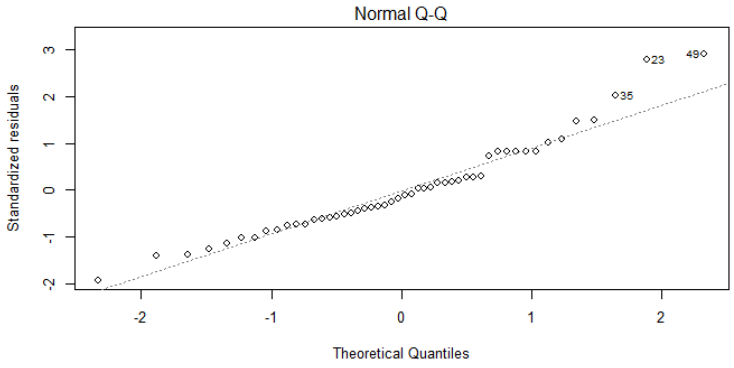


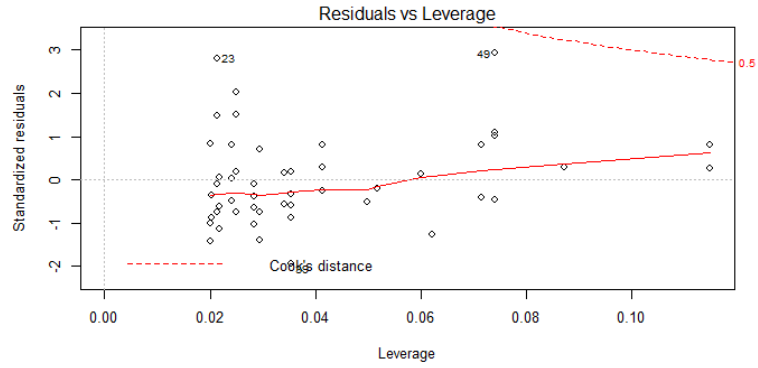


```
> y = lm(formula = dist ~ (1+speed), data=cars)
> summary(y)

Call:
lm(formula = dist ~ (1 + speed), data = cars)

Residuals:
    Min      1Q  Median      3Q     Max
-29.07  -9.53  -2.27    9.21  43.20

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -17.579      6.758   -2.60    0.012 *
speed          3.932      0.416    9.46  1.5e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.4 on 48 degrees of freedom
Multiple R-squared:  0.651,    Adjusted R-squared:  0.644
F-statistic: 89.6 on 1 and 48 DF,  p-value: 1.49e-12

>   anova(y)
Analysis of Variance Table

Response: dist
          Df Sum Sq Mean Sq F value  Pr(>F)
speed      1  21185   21185    89.6 1.5e-12 ***
Residuals 48  11354     237
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>   layout(matrix(1:4,2,2))
>   plot(y)
```

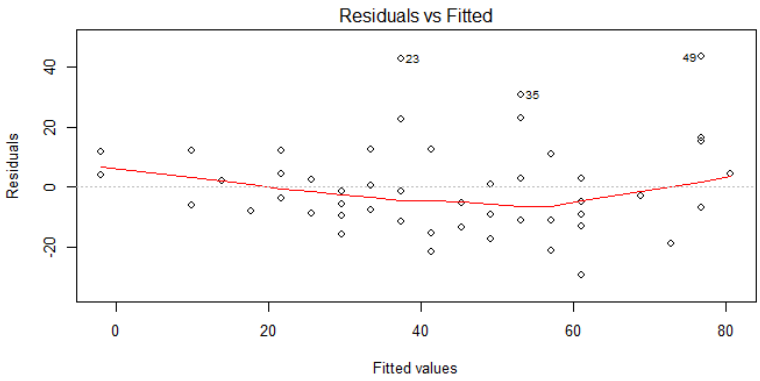


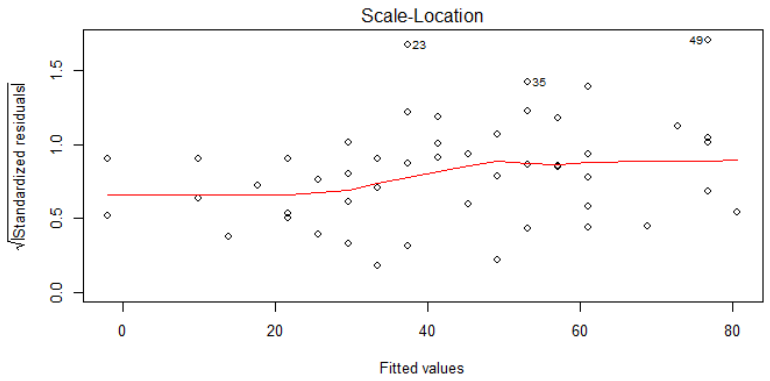


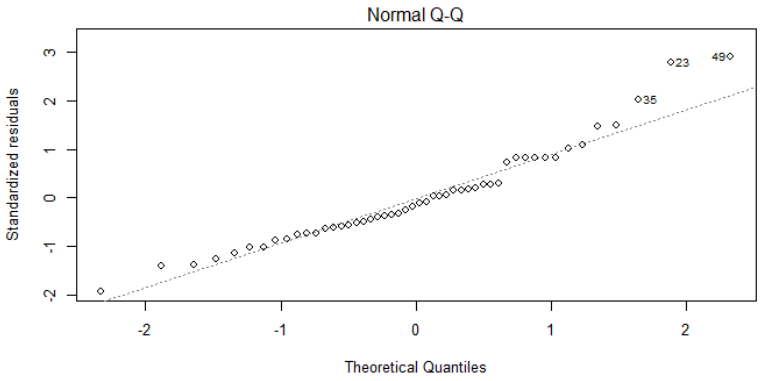


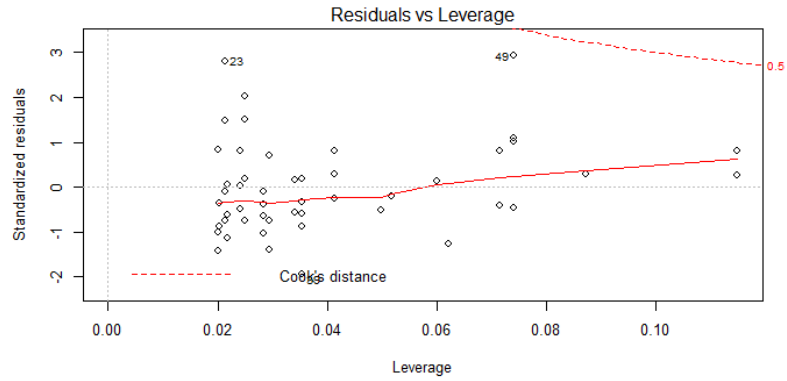


```
> y = lm(formula = dist ~ (speed+speed^2), data=cars)
> summary(y)

Call:
lm(formula = dist ~ (speed + speed^2), data = cars)

Residuals:
    Min      1Q  Median      3Q     Max
-29.07  -9.53  -2.27    9.21  43.20

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -17.579      6.758   -2.60    0.012 *
speed          3.932      0.416    9.46  1.5e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.4 on 48 degrees of freedom
Multiple R-squared:  0.651,    Adjusted R-squared:  0.644
F-statistic: 89.6 on 1 and 48 DF,  p-value: 1.49e-12

>   anova(y)
Analysis of Variance Table

Response: dist
          Df Sum Sq Mean Sq F value  Pr(>F)
speed      1  21185   21185    89.6 1.5e-12 ***
Residuals 48  11354     237
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>   layout(matrix(1:4,2,2))
>   plot(y)
```



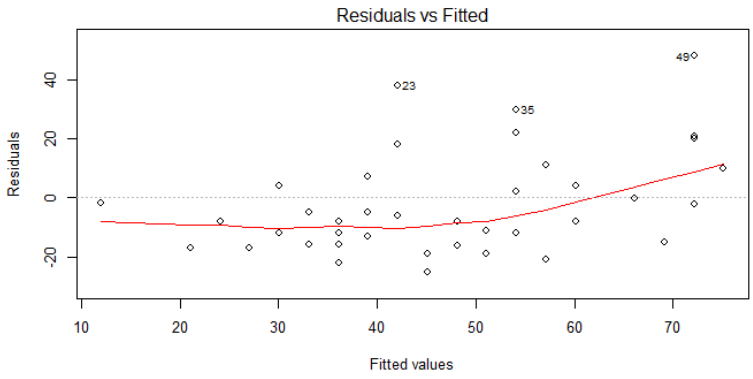
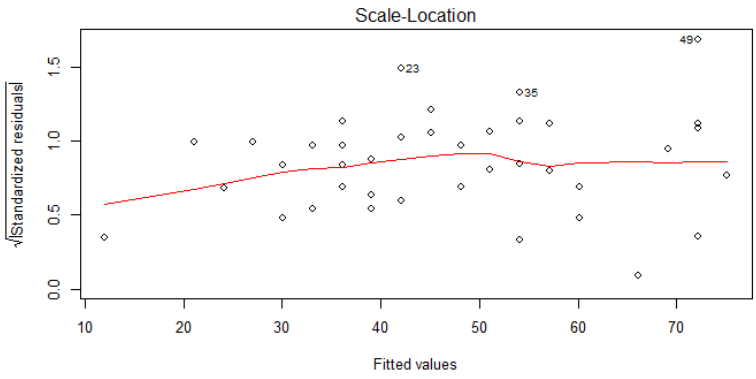
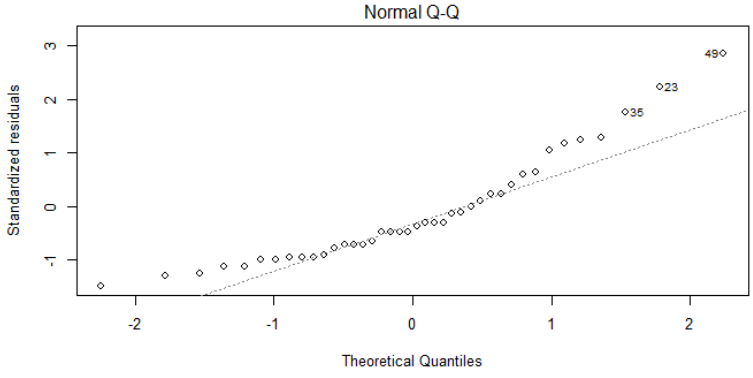
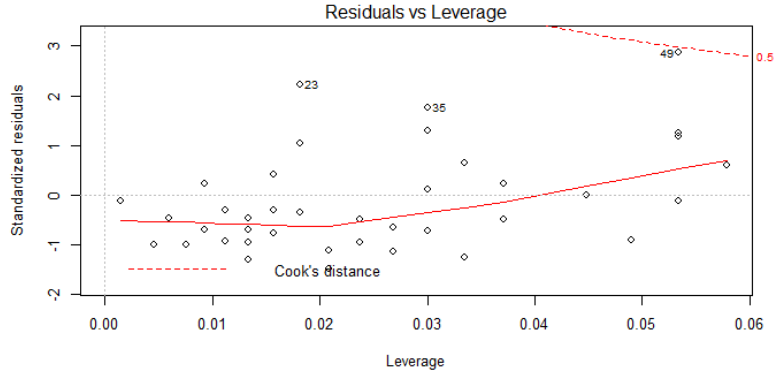






y = lm(formula = dist ~ (-1+speed), data=cars) *dobar na cijeloj bazi ^^*

```
set.seed(100)

trainingRowIndex = sample(1:nrow(cars), 0.8*nrow(cars))

trainingdata = cars[trainingRowIndex, ]

testData = cars[-trainingRowIndex, ]
```

```
> y = lm(formula = dist ~ (-1+speed), data=trainingdata)
> summary(y)

Call:
lm(formula = dist ~ (-1 + speed), data = trainingdata)

Residuals:
   Min     1Q Median     3Q    Max
-25.09 -15.37  -7.07   4.68  47.85

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
speed    3.006      0.166    18.1   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.2 on 39 degrees of freedom
Multiple R-squared:  0.894,    Adjusted R-squared:  0.891
F-statistic:  330 on 1 and 39 DF,  p-value: <2e-16
> anova(y)
Analysis of Variance Table

Response: dist
          Df Sum Sq Mean Sq F value Pr(>F)
speed      1  97524   97524     330 <2e-16 ***
Residuals 39  11543     296
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> layout(matrix(1:4,2,2))
>   plot(y)
```

dobar na train podacima