Chapter 3 Part I Linear Regression (Problems_8)

Kartheek Raj

12/21/2019

Problem 8: This question involves the use of simple linear regression on the Auto data set

Required Packages :ISLR

Answer

Loading the **Auto** data from ISLR package

```
require(ISLR)
```

```
## Loading required package: ISLR
```

```
library(ISLR)
auto<-data.frame(Auto)
head(auto)
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70      1
## 2  15         8          350        165   3693         11.5   70      1
## 3  18         8          318        150   3436         11.0   70      1
## 4  16         8          304        150   3433         12.0   70      1
## 5  17         8          302        140   3449         10.5   70      1
## 6  15         8          429        198   4341         10.0   70      1
##                       name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```

(a) Use the lm() function to perform a simple linear regression with mpg as the response and horsepower as the predictor. Use the summary() function to print the results. Comment on the output.

Answer

```
autolm<-lm(mpg~horsepower,data=auto)
summary(autolm)
```

```
##
## Call:
## lm(formula = mpg ~ horsepower, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5710  -3.2592  -0.3435   2.7630  16.9240
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 39.935861   0.717499   55.66   <2e-16 ***
## horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.906 on 390 degrees of freedom
## Multiple R-squared:  0.6059, Adjusted R-squared:  0.6049
## F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

(i) Is there a relationship between the predictor and the response?

Answer: There is relationship between the predictor and the response as evidence poised to P-value.

(ii) How strong is the relationship between the predictor and the response?

Answer: Adjusted R-squared value is 0.60 means 60% can explain variance in **mpg** by **horsepower**.

(iii) Is the relationship between the predictor and the response positive or negative?

Answer: Negative correlation indicated by horsepower coefficient is -0.157845 i.e m=negative slope.

(iv) What is the predicted mpg associated with a horsepower of 98? What are the associated 95% confidence and prediction intervals?

Answer

```
predict(autolm,data.frame(horsepower=98),interval="confidence")
```

```
##        fit      lwr      upr
## 1 24.46708 23.97308 24.96108
```

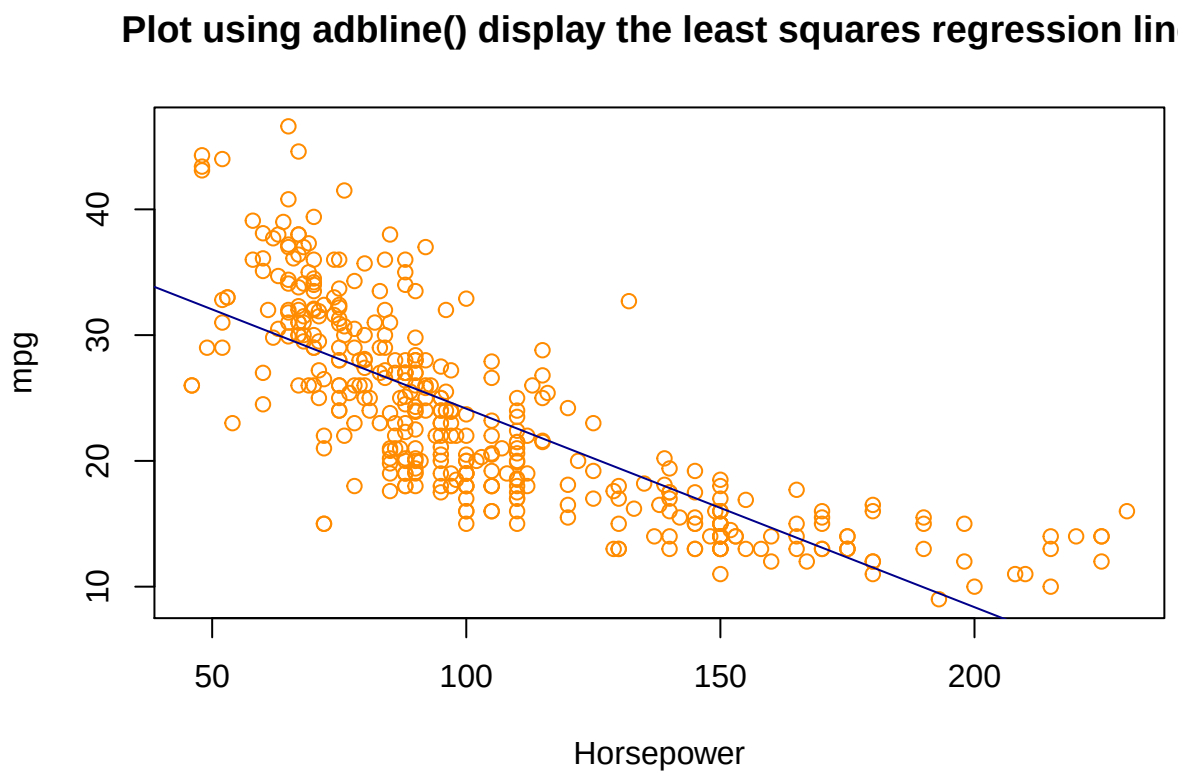
```
predict(autolm,data.frame(horsepower=98),interval="prediction")
```

```
##        fit     lwr      upr
## 1 24.46708 14.8094 34.12476
```

(b) Plot the response and the predictor. Use the abline() function to display the least squares regression line.

Answer

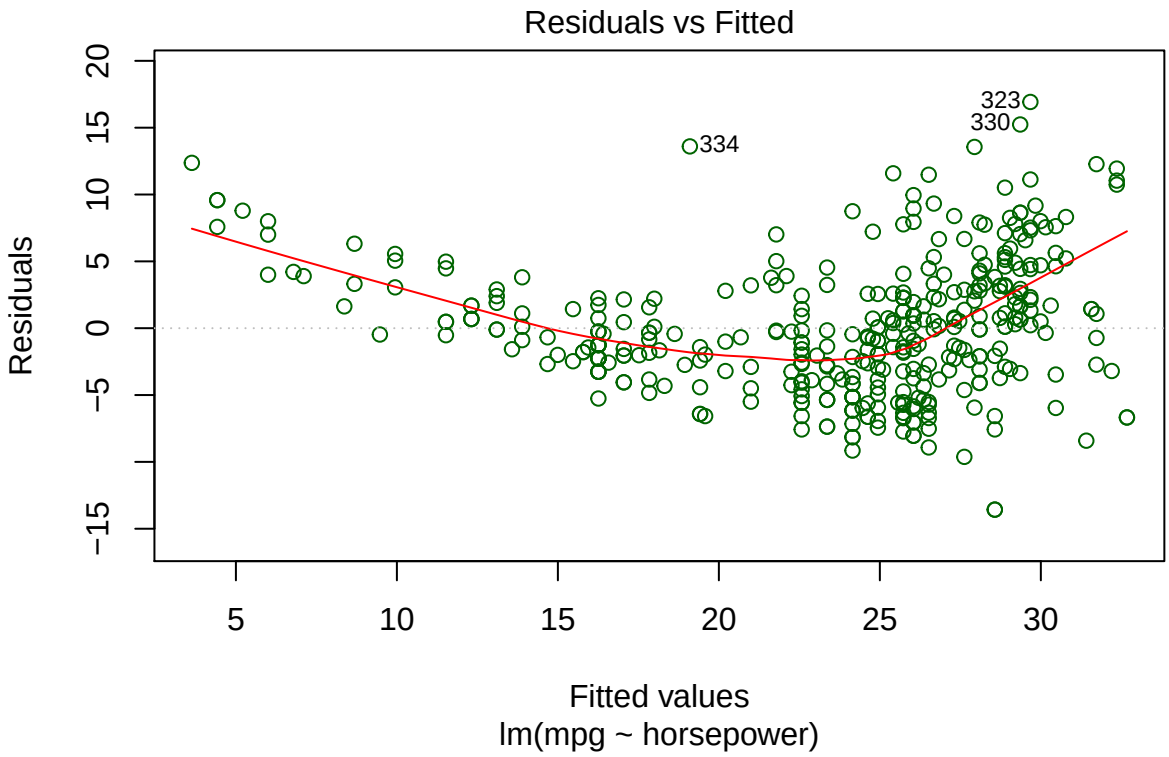
```
plot(auto$horsepower,auto$mpg,main="Plot using adbline() display the least squares regression line",xlab
abline(autolm,col="darkblue")
```

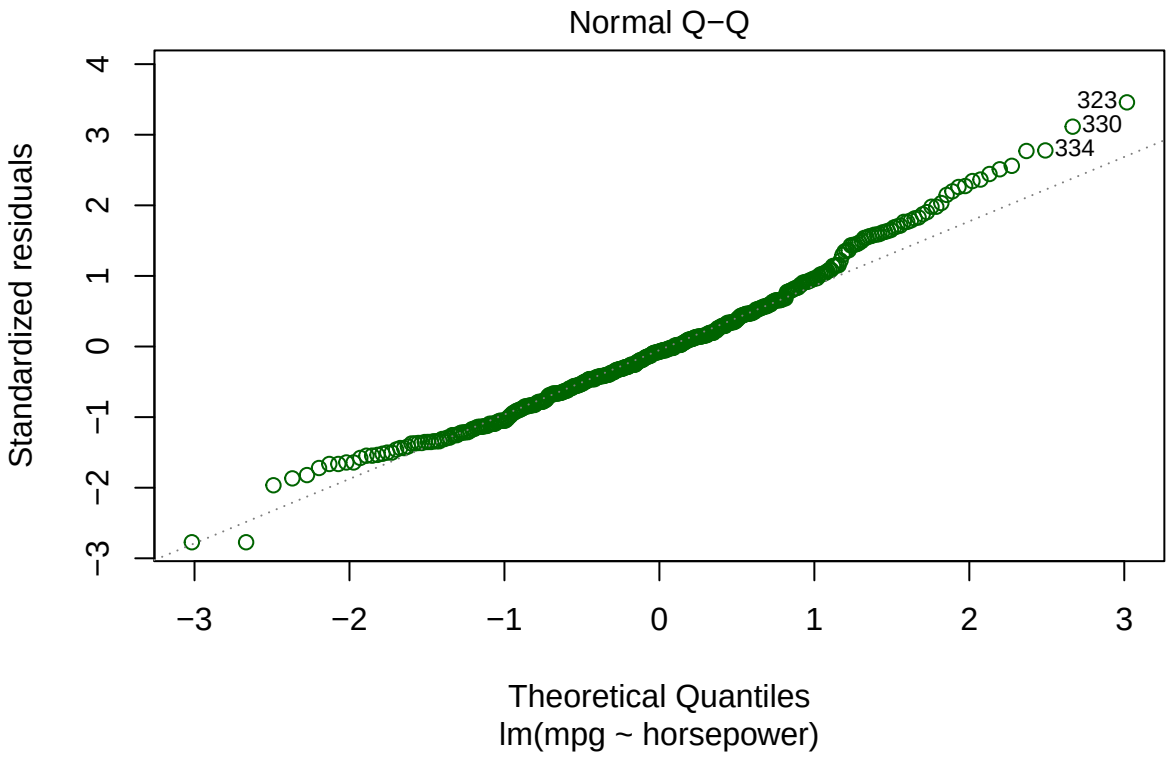


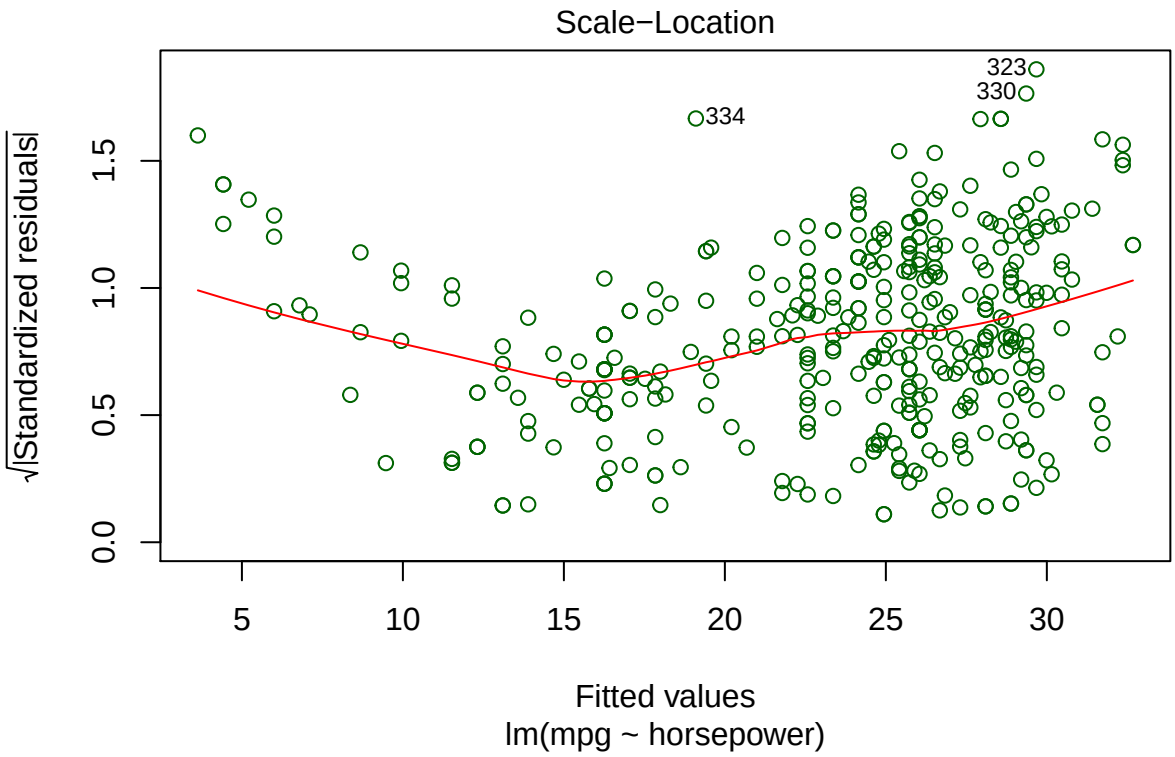
(c) Use the plot() function to produce diagnostic plots of the least squares regression fit. Comment on any problems you see with the fit.

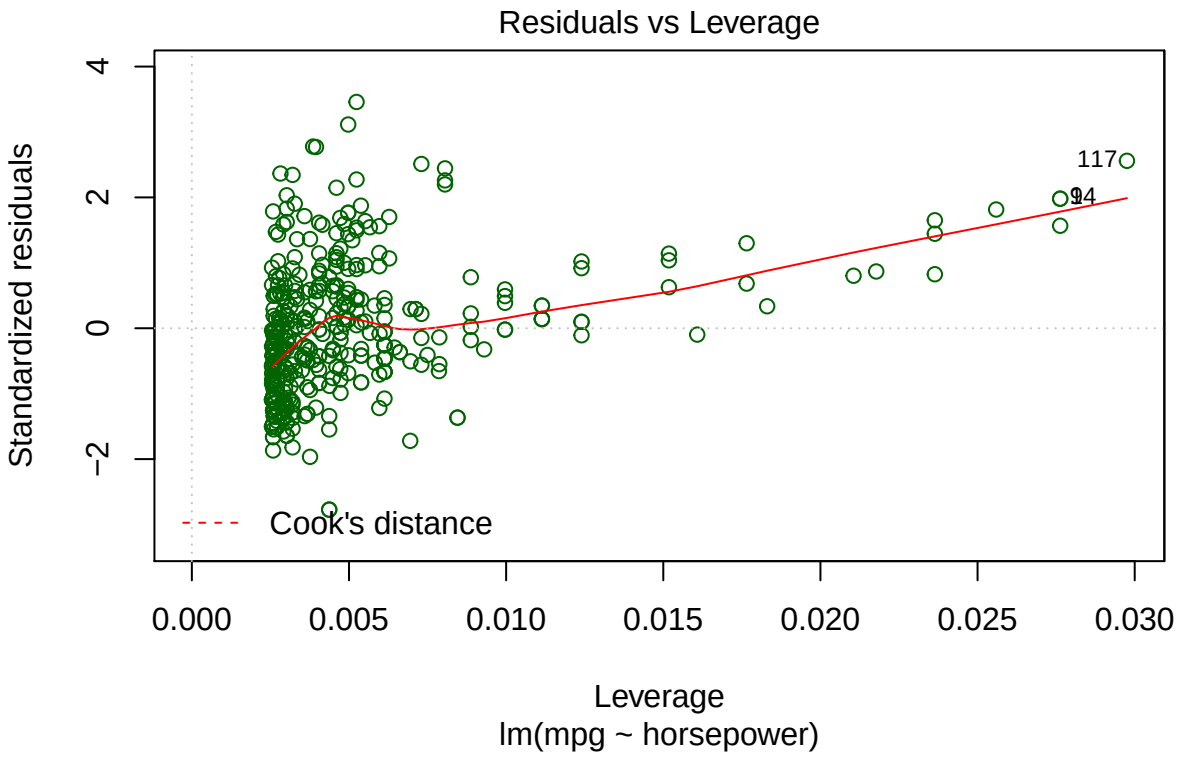
Answer

```
plot(autolm,col="darkgreen")
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ horsepower)
323
330
334
20
15
10
5
0
−5
−15
5
10
15
20
25
30


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ horsepower)
323
330
334
−3
−2
−1
0
1
2
3
4


Scale−Location
323
330
334
√|Standardized residuals|
0.0
0.5
1.0
1.5
5
10
15
20
25
30
Fitted values
lm(mpg ~ horsepower)


Residuals vs Leverage
Standardized residuals
4
2
0
−2
117
94
Cook's distance
0.000
0.005
0.010
0.015
0.020
0.025
0.030
Leverage
lm(mpg ~ horsepower)