Meher Shrishti Nigam
20BRS1193

EDA LAB – 1
16 / 12 / 22

```
# Meher Shrishti Nigam
# 20BRS1193
# EDA Lab 1
# EDA-LAB-EXPERIMENT-1 (Date-16/12/2022)

options(prompt="MEHERSHRISHTI>", continue =" ")
library(ISLR)

# This question involves the use of simple linear regression on the Auto data set.
# Perform a simple linear regression with mpg as the response and horsepower as the predictor.
# Print the summery of the dataset. Comment on the output.

df <- Auto
df <- na.omit(df)
mpgY <- df$mpg
hpX <- df$horsepower

# (a) Is there a relationship between the predictor and the response?
relation <- lm(mpgY~hpX)
relation
summary(relation)
```

```
MEHERSHRISHTI>relation <- lm(mpgY~hpX)
MEHERSHRISHTI>relation

Call:
lm(formula = mpgY ~ hpX)

Coefficients:
(Intercept)          hpX
    39.9359      -0.1578
```

```
MEHERSHRISHTI>summary(relation)

Call:
lm(formula = mpgY ~ hpX)

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 ***
hpX         -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
```

correlation <- cor(hpX, mpgY, method = "pearson")
correlation

```
MEHERSHRISHTI>correlation <- cor(hpX, mpgY, method = "pearson")
MEHERSHRISHTI>correlation
[1] -0.7784268
```

Yes, as there is a moderately strong correlation of -0.7784268 between mpg and horsepower.
Thus, there is a relationship between the predictor and the response.

 # (b) How strong is the relationship between the predictor and the response?
correlation <- cor(hpX, mpgY, method = "pearson")
correlation

```
MEHERSHRISHTI>correlation <- cor(hpX, mpgY, method = "pearson")
MEHERSHRISHTI>correlation
[1] -0.7784268
```

An absolute correlation greater than 0.7 shows that there is a significant relationship between the predictor and the response.

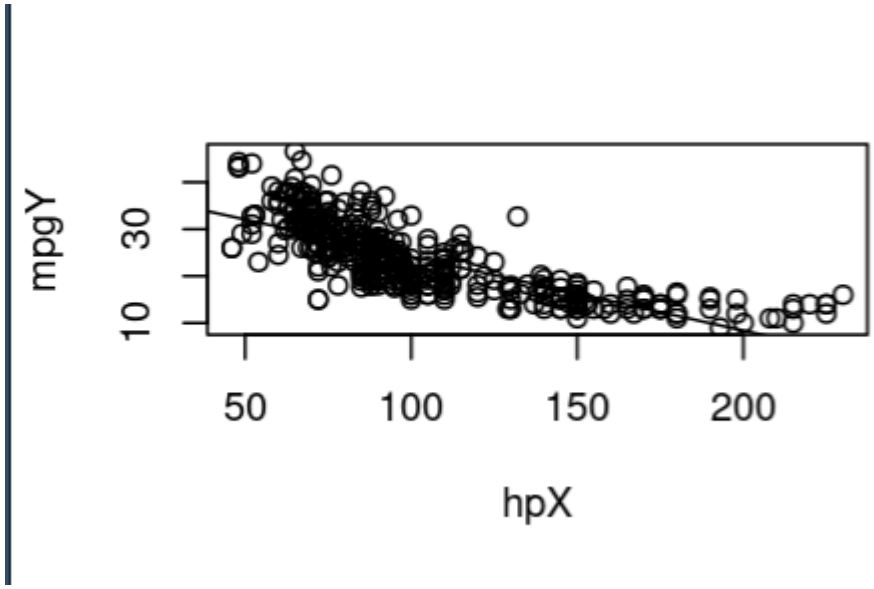
(c) Is the relationship between the predictor and the response positive or negative?
The relationship is negative, i.e, miles per gallon (mpg) increases when horsepower decreases, and vice versa.

# (d) What is the predicted mpg associated with a horsepower of 98? What are the associated 95 % confidence and prediction intervals?
testData <- data.frame(hpX = 98)
confidenceMpg <- predict(relation, testData, interval = "confidence")
confidenceMpg
predictedMpg <- predict(relation, testData, interval = "prediction")
predictedMpg

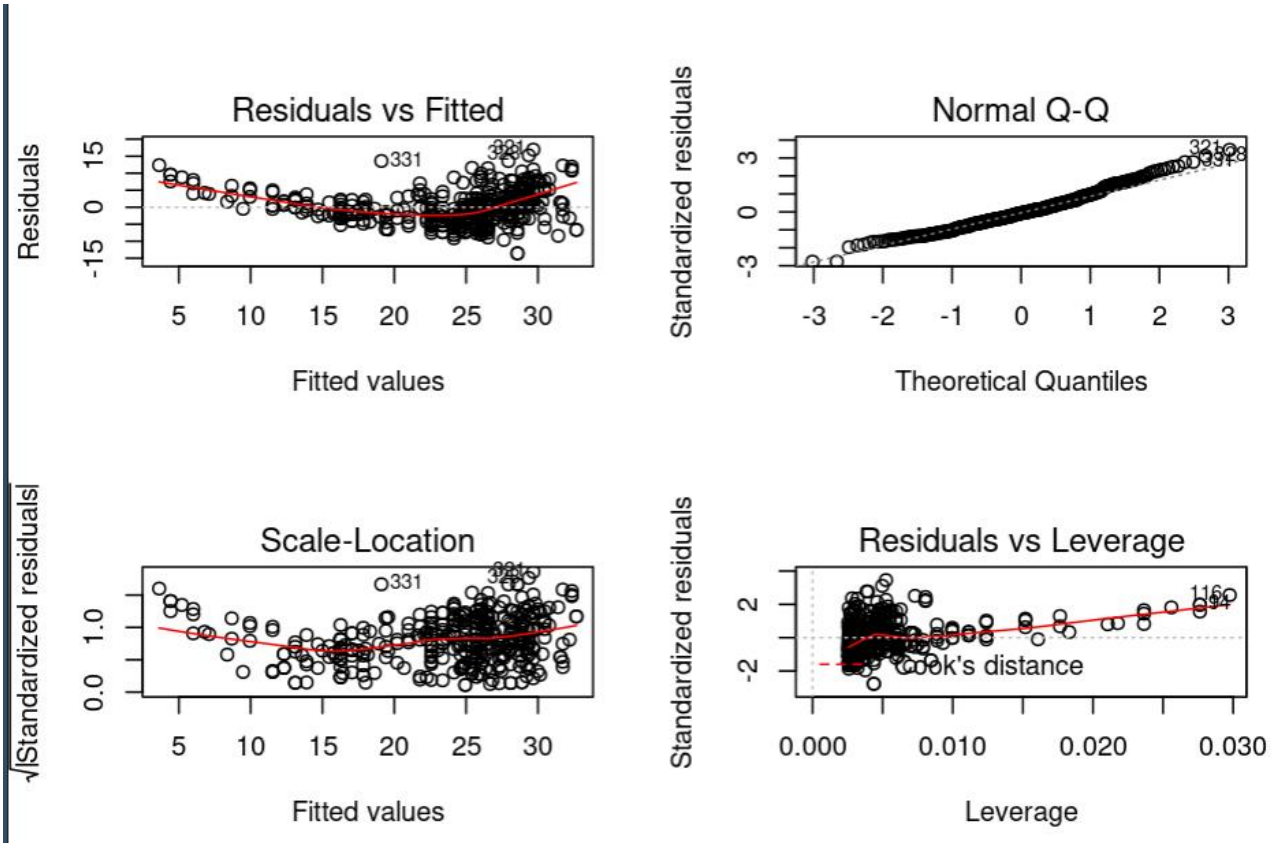
```
MEHERSHRISHTI>testData <- data.frame(hpX = 98)
MEHERSHRISHTI>confidenceMpg <- predict(relation, testData, interval = "confidence")
MEHERSHRISHTI>confidenceMpg
       fit      lwr      upr
1 24.46708 23.97308 24.96108
MEHERSHRISHTI>predictedMpg <- predict(relation, testData, interval = "prediction")
MEHERSHRISHTI>predictedMpg
       fit     lwr      upr
1 24.46708 14.8094 34.12476
MEHERSHRISHTI>
```

Thus, this shows that we can predict the mpg of a car with horsepower of 98 to be between 14.81 to 34.12 with 95% confidence.

 # (e) Plot the response and the predictor. Display the least squares regression line.
plot(hpX, mpgY)
abline(relation)



 # (f) Use the plot() function to produce diagnostic plots of the least squares regression fit. Comment on any problems you see with the fit.
par(mfrow=c(2,2))
plot(relation)



We observe that there is non-linearity in the data from the plot of residuals v. fitted values.
Thus, fitting a linear equation is not the best solution for this problem.
A few outliers can be noticed in the plot of standardized residuals versus leverage, as they are higher than 2 or lower than -2.