ISL Assignment 1-Q1

Q-1. For this question, use of simple linear regression on the 'Auto' data set.

(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 to include in your submission. Comment on the output.

Answer:

Step 1: To execute simple linear regression, the given data "Auto-rev " has been loaded into R studio using below line commands.

```
> library(readr)
> Auto_rev <- read_csv("~/R/Auto-rev.csv")
Parsed with column specification:
cols(
mpg = col_double(),
cylinders = col_integer(),
displacement = col_double(),
horsepower = col_integer(),
weight = col_integer(),
acceleration = col_double(),
year = col_integer(),
origin = col_integer(),
name = col_character()
)
> View(Auto_rev)
> names(Auto_rev)

        [1] "mpg"           "cylinders"    "displacement" "horsepower"
"weight"       "acceleration"
[7] "year"         "origin"       "name"
```

Step 2:

To execute simple linear regression by using lm() function with mpg as the response and horsepower as the predictor.

```
> Auto_rev.mod1=lm(mpg~horsepower,data=Auto_rev)
> Auto_rev.mod1

Call:
lm(formula = mpg ~ horsepower, data = Auto_rev)

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

Step 3: Executed *summary*() function to print the results

```
> summary(Auto_rev.mod1)

Call:
lm(formula = mpg ~ horsepower, data = Auto_rev)

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
```

Comment on the Output:

Hence, we can analyze that the p-value is very small, so we can infer that the above data model has relationship among predictor and response by excluding null hypothesis. And by seeing the R-square 0.6059 value we can analyze that there is a strong relationship (R^2 >=0.5 is fairly a good correlation).

i. Is there a relationship between the predictor and the response?

Answer: A relationship between predicator and response can be resolute by considering the p-value. If the p-value is very small, then we can infer that there is a relationship between the predictor and response (x and y)-Alternative hypothesis. Such that we can reject the null hypothesis.

```
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

In this scenario based on the output received the p value is very small. So, we can neglect the null hypothesis and infer that there is a relationship between the mpg(response) and horsepower(predictor).

ii. How strong is the relationship between the predictor and the response?

Answer:

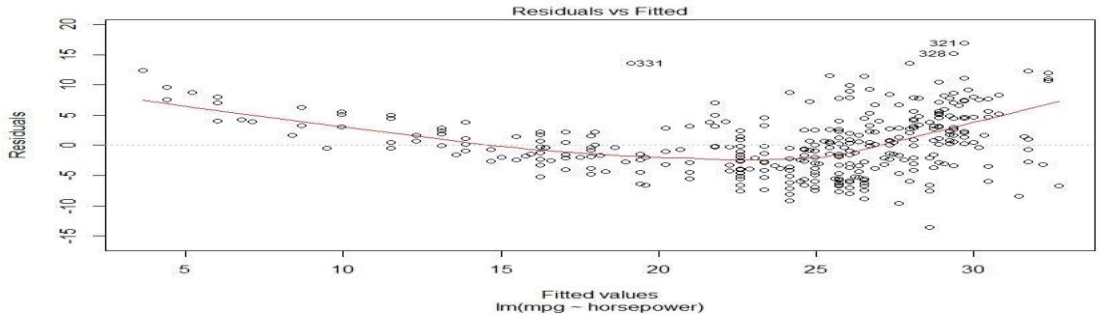
Mean:

```
> mean(Auto_rev$mpg)
[1] 23.44592
```

Residual Standard Error is 4.906 from the output of the lm() function.
Percentage Error: 20.924%

From the lm()function we can see that the value R-squared: 0.6059 which is 60.59% of the variability between mpg and horsepower. More the variance that is shown up by the regression model. So the data points will fall to the fitted regression line and the stronger the relationship.

Thus, the R^2 >=0.5 is fairly a good correlation. Below is the graph which shows the fitted regression.



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

Answer:

```
horsepower   -0.157845
```

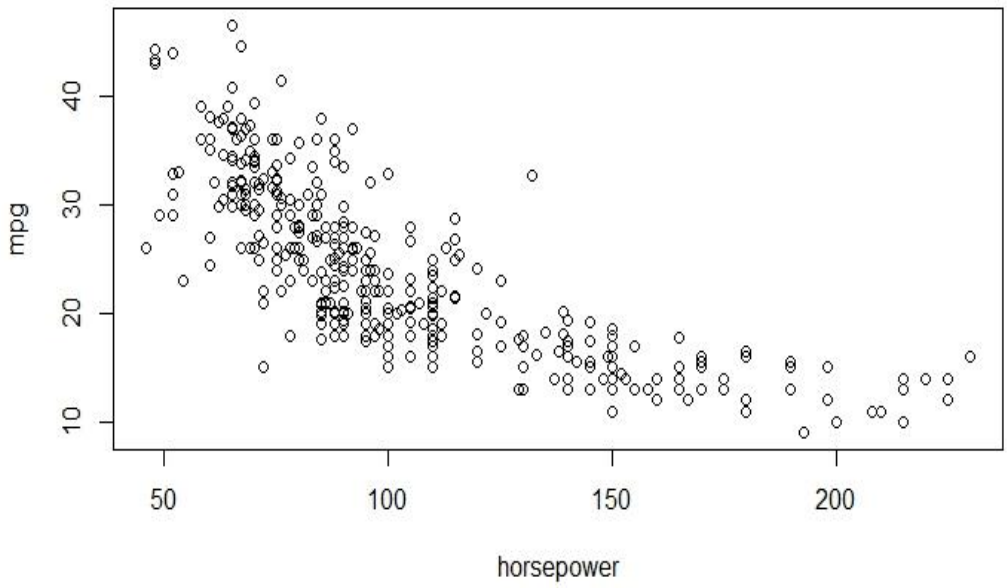
Coefficient of the predictor is negative from the simple linear regression function. If coefficient of the predictor is negative, then the relationship between is also a negative. Thus, in this case the relationship between the predictor and the response is negative as horsepower -0.157845.

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

```
> predict(Auto_rev.mod1,data.frame(horsepower=98),interval="confidence")
       fit      lwr      upr
1 24.46708 23.97308 24.96108
> predict(Auto_rev.mod1,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. Include this graph.

```
plot (mpg~horsepower,data=Auto_rev)
```



Least squares regression line:

```
plot (mpg~horsepower,data=Auto_rev)
> abline(Auto_rev.mod1,col="red")
```