INTRODUCTION TO STATISTICAL LEARNING

QUESTION-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.

i. Is there a relationship between the predictor and the response?
ii. How strong is the relationship between the predictor and the response? iii. Is the relationship between the predictor and the response positive or negative? iv. What is the predicted **mpg** associated with a **horsepower** of 98? What are the associated 95% confidence and prediction intervals?

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

Ans:

Loading the data

```
R Console

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> ISLAs1<-read.csv("C:\\Users\\sudheesha\\Desktop\\ISL 1\\Auto-rev.csv")
> ISLAs1
    mpg cylinders displacement horsepower weight acceleration year origin
1  18.0         8        307.0        130   3504         12.0   70      1
2  15.0         8        350.0        165   3693         11.5   70      1
3  18.0         8        318.0        150   3436         11.0   70      1
4  16.0         8        304.0        150   3433         12.0   70      1
5  17.0         8        302.0        140   3449         10.5   70      1
6  15.0         8        429.0        198   4341         10.0   70      1
7  14.0         8        454.0        220   4354          9.0   70      1
8  14.0         8        440.0        215   4312          8.5   70      1
9  14.0         8        455.0        225   4425         10.0   70      1
10 15.0         8        390.0        190   3850          8.5   70      1
11 15.0         8        383.0        170   3563         10.0   70      1
12 14.0         8        340.0        160   3609          8.0   70      1
13 15.0         8        400.0        150   3761          9.5   70      1
14 14.0         8        455.0        225   3086         10.0   70      1
15 24.0         4        113.0         95   2372         15.0   70      3
16 22.0         6        198.0         95   2833         15.5   70      1
17 18.0         6        199.0         97   2774         15.5   70      1
18 21.0         6        200.0         85   2587         16.0   70      1
19 27.0         4         97.0         88   2130         14.5   70      3
20 26.0         4         97.0         46   1835         20.5   70      2
21 25.0         4        110.0         87   2672         17.5   70      2
22 24.0         4        107.0         90   2430         14.5   70      2
23 25.0         4        104.0         95   2375         17.5   70      2
24 26.0         4        121.0        113   2234         12.5   70      2
25 21.0         6        199.0         90   2648         15.0   70      1
26 10.0         8        360.0        215   4615         14.0   70      1
```

a. Using lm function, we performed a linear regression in which mpg is the response and horse power is the predictor. The data obtained from this is stored in lmdata variable and the output can be obtained by executing the following statement summary(lmdata);

```
> lm.fit<-lm(mpg~horsepower,data=ISLAs1)
> summary(lm.fit)

Call:
lm(formula = mpg ~ horsepower, data = ISLAs1)

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. Yes , It's clearly evident that there is relationship between mpg and horsepower. We can say that by the p-value (<2.2e-16) corresponding to the F-statistic value.

ii. From the above data, It is clear that R-square = 0.6059. That means almost 60.59% variation in mpg can be explained by using horsepower.

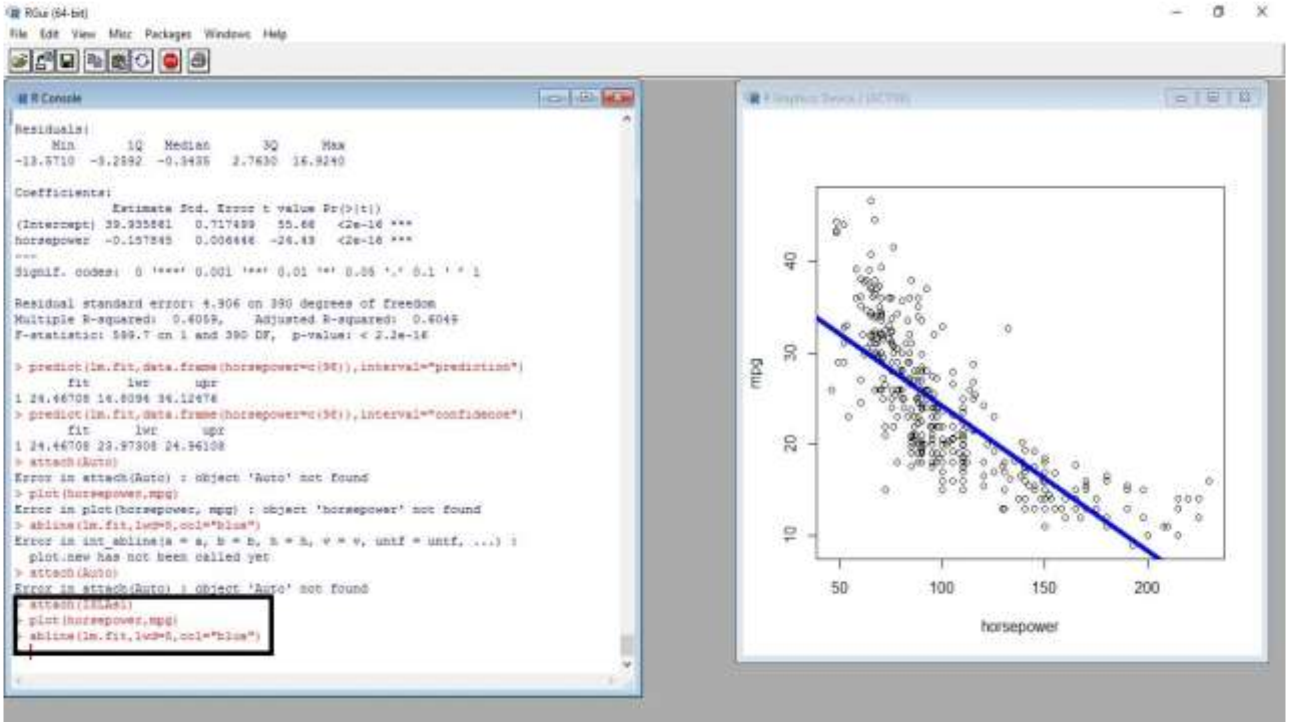
iii. You can see that, the co-efficient of horsepower is -0.157845 from the screenshot attached above, which shows that the relationship between predictor and horsepower is **negative**.

iv. We can have a glance at the responses for the question, in the screenshot below.

```
> predict(lm.fit,data.frame(horsepower=c(98)),interval="prediction")
       fit     lwr      upr
1 24.46708 14.8094 34.12476
> predict(lm.fit,data.frame(horsepower=c(98)),interval="confidence")
       fit      lwr      upr
1 24.46708 23.97308 24.96108
```

b. Below screenshot is for the response and the predictor using the *abline()* function to display the least squares regression line.

Firstly, we have attached the data [attach(assignement1)] and then plotted the graph for horsepower vs mpg [plot(horsepower~mpg)]


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
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
> predict(lm.fit,data.frame(horsepower=c(98)),interval="prediction")
fit lwr upr
1 24.46708 14.8094 34.12476
> predict(lm.fit,data.frame(horsepower=c(98)),interval="confidence")
fit lwr upr
1 24.46708 23.97308 24.96108
> attach(Auto)
Error in attach(Auto) : object 'Auto' not found
> plot(horsepower,mpg)
Error in plot(horsepower, mpg) : object 'horsepower' not found
> abline(lm.fit,lwd=5,col="blue")
Error in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...) :
plot.new has not been called yet
> attach(Auto)
Error in attach(Auto) : object 'Auto' not found
> attach(ISLR)
> plot(horsepower,mpg)
> abline(lm.fit,lwd=5,col="blue")
mpg
horsepower
10
20
30
40
50
100
150
200