**Lab 1**

**Simple Linear Regression**

Name: Jesto Peter

Lab1: Simple Linear Regression

Date: 19/11/2022

Class: 2MSTAT

* **Introduction**

Simple linear regression is used to model the relationship between two continuous variables. Often, the objective is to predict the value of an output variable (or response) based on the value of an input (or predictor) variable.

We can use regression, and the results of regression modeling, to determine which variables have an effect on the response or help explain the response.

For this particular Lab Session, we will be focusing on the dataset involving 2 variables i.e Blood Pressure and Weight

* **Objective**

In this Lab Session, we use R programming and try to import a dataset of values of 26 individuals, containing the data of their weights and blood pressure. We try to come to conclusion about the relationship between Weight and Blood Pressure of individuals and try to find if they are related or not and how they are related.

We have to:

1. Obtain the scatter plot and interpret it.

2. Find a regression line connecting the variables systolic pressure and

weight. Interpret the plot, intercept term, and regression coefficients. What

do you infer from the sign of the regression coefficient?

3. Obtain the fitted values. Does the sum of fitted values is equal to the sum

of observed values?

* **Procedure**

1. **Plotting the Scatter Plot**

data<-read.csv(file.choose())  
head(data) #reading the csv file which has the data and printing only the top 5 values

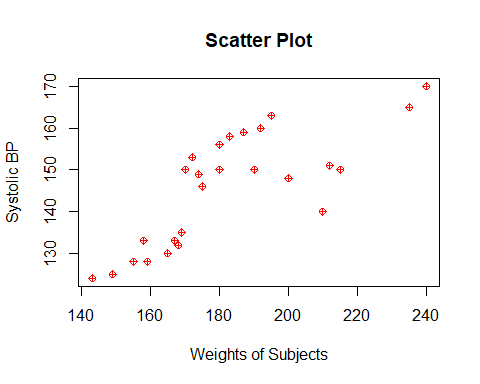
## ï..Subject Weight Systolic.BP  
## 1 1 165 130  
## 2 2 167 133  
## 3 3 180 150  
## 4 4 155 128  
## 5 5 212 151

***#Here we use attach() function to convert the columns in the data set to lists***

attach(data)   
x=Weight #assigning the Weight column to variable x  
y=Systolic.BP #assigning the Systolic.BP column to variable Y

***#Using plot() function to plot the graph***

#Using plot function for plotting the scatter plot  
plot(x, y, main = "Scatter Plot",xlab = "Weights of Subjects",   
 ylab = "Systolic BP",pch = 10,col='red')



**Interpretation**

Here the independent variable(Regressor) is the Weight and the dependent variable is the Blood Pressure of the Individuals. We see and can tell from the scatter plot that as the Weight of individuals increases, the Blood Pressure also increases in the positive direction.

1. **Plotting and Interpreting the Regression Line**

**#We use lm() function to find the Coefficients**

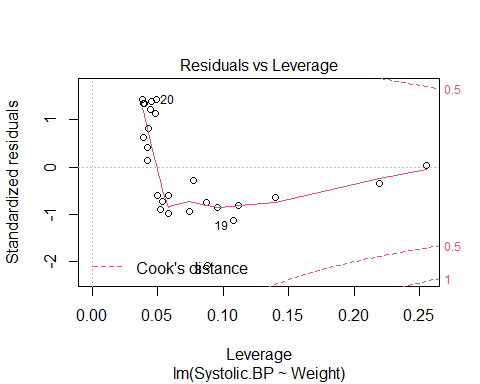
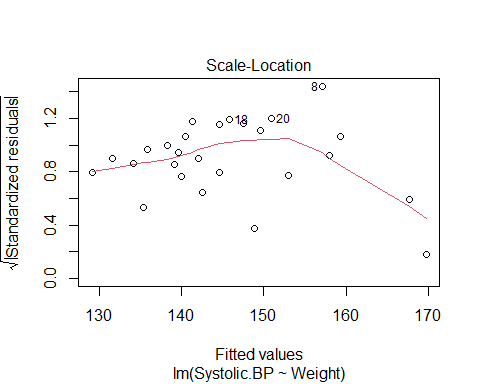
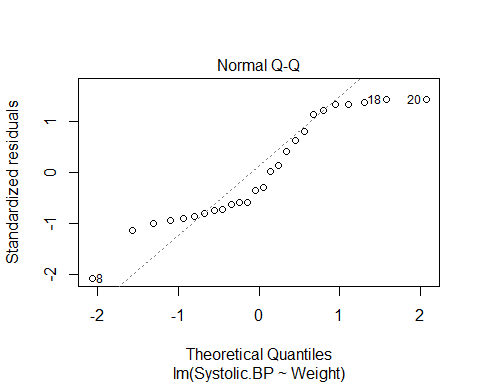
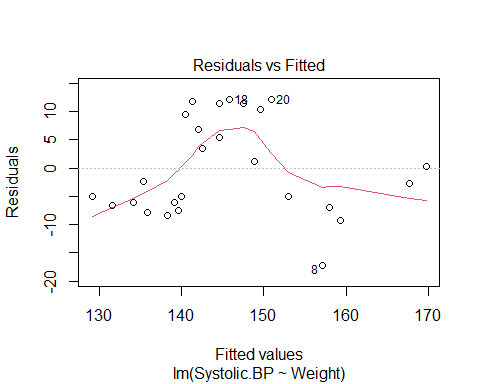
#Finding the Coefficients  
slr1=lm(Systolic.BP~Weight)  
slr1

##   
## Call:  
## lm(formula = Systolic.BP ~ Weight)  
##   
## Coefficients:  
## (Intercept) Weight   
## 69.1044 0.4194

**#We see that the Regression Coefficient is 0.4194 and the Intercept is 69.1044**

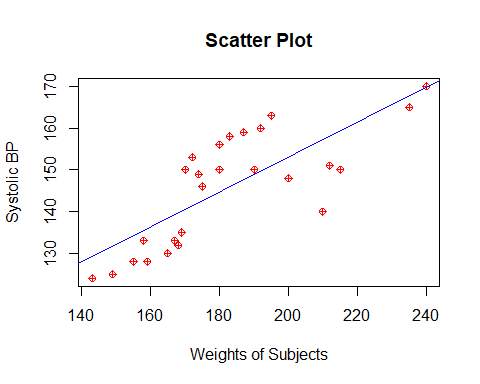
#Plotting the linear model obtained using lm()

plot(slr1)



**#Plotting the regression line using abline() function**

plot(x, y, main = "Scatter Plot",xlab = "Weights of Subjects",   
 ylab = "Systolic BP",pch = 10,col='red')  
abline(slr1,col='blue')



**Interpretation**

We see that the Coefficients are: 69.1044(Intercept) and 0.4194(Regression Coefficient). We have the regression coefficient as a positive value. A positive coefficient indicates that as the value of the independent variable increases, the mean of the dependent variable also tends to increase.

Hence we can say that there is a positive correlation between the Weight and Blood Pressure of individuals.

1. **Obtain the fitted values. Seeing whether the sum of fitted values is equal to the sum of observed values?**

*#We use fitted.values() function*

a=fitted.values(slr1)

*#Sum of fitted values*  
s1=sum(a)  
s1

## [1] 3786

*#Sum of Observed Values*

s2=sum(Systolic.BP)  
s2

## [1] 3786

**Interpretation**

We see that the sum of the Fitted Values equals the sum of the Observed values.

* **Conclusion**

We see from the scatter plot and the positive regression coefficient that the independent variable(Regressor) i.e Weight and the dependent variable i.e Blood Pressure of the Individuals have a positive Correlation. Hence the fitted model is showing a positively correlated We see and can tell from the plots, that as the Weight of individuals increases, the Blood Pressure also increases in the positive direction.

The Regression Coefficient obtained is 0.4194 and the Intercept is 69.1044

Hence the regression equation obtained is y=0.4194(x) + 69.1044

We also see that the sum of the fitted values equals the sum of the observed values of the data i.e 3786.