

**Name:** Vikram Surya R

**Reg No:** 19BCE1490

## **Essentials of Data Analytics**

### **Tasks for Week-1: Linear Regression**

Understand the following operations/functions on random dataset and perform similar operations on mtcars and 'data.csv' dataset based on given instructions.

**Aim:** To develop linear regression model for the given data using R programming and to verify the null hypothesis

#### **Algorithm:**

- Import the library 'dplyr'.
- Generate the data to be analyzed.
- Take a sample data using sample\_n() or sample\_frac() function and store it.
- Define the x and y variables.
- Plot a scatter plot for x and y variables and label them accordingly.
- Using cor.test(x,y) we can find the correlation.
- Using lm() create a linear regression model.
- Using the abline() function, draw the lm model curve in the plot.
- Using summary() function, find the summary of the lm model created.

#### **Statistic:**

##### **Case 1: mtcars**

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.2252	0.3606	14.488	2.12e-09 ***
x	-0.4886	0.1131	-4.321	0.00083 ***

Residual standard error: 1.292 on 13 degrees of freedom

Multiple R-squared: 0.2175

Adjusted R-squared: 0.1573

F-statistic: 3.77

p-value: 0.0008297

### **Case 2:data**

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	97.85009	47.15133	2.075	0.0433 *
x	0.06337	0.27567	0.23	0.8192

Residual standard error: 31.47 on 48 degrees of freedom

Multiple R-squared: 0.0011

Adjusted R-squared: 0.01971

F-statistic: 0.05284

p-value: 0.8192

## **Inference**

### **Case 1:**

the p value of b1 and b0 are of less than 0.05 that means the variables of model is significant and the overall p value id less than 0.05 so the model is significant.

### **Case 2:**

the p value of b1 and b0 are of more than 0.05 that means the variables of model is not significant and the overall p value id less than 0.05 so the model is not significant.

## **Program:**

### **Case 1:**

```
data1<-mtcars
data1
library(dplyr)
trail=sample_n(data1,15)
x=trail$wt
y=trail$drat
plot(x,y,main='Scatter plot',xlab='wt',ylab='mpg')
cor.test(x,y)
lmodel=lm(y~x)
abline(lmodel,col='red')
summary(lmodel)
```

### **Case 2:**

```
c <- read.csv("C:/Users/VIKRAM SURYA/Desktop/EDA_LAB/data.csv",header
= TRUE,sep = ",")
library(dplyr)
trail=sample_n(c,50)
y=trail$Weight
x=trail$Height
plot(x,y,main='Scatter plot',xlab='Height',ylab='Weight')
cor.test(x,y)
lmodel=lm(y~x)
abline(lmodel,col='red')
summary(lmodel)
```