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 dplyr library
* Store the mtcars or data.csv data into variables data1 and data2
* Take a sample of 15 data using sample\_n() function and store it in train
* The x variable of mtcars is wt and the y variable of mtcars is mpg
* The x variable of data.csv is height and the y variable of data.csv is weight
* Plot a scatter plot for x and y variable using plot function
* Using cor.test we get the correlation between x and y variables
* Create a linear regression model using lm() function for mtcars and data.csv
* Draw the lm model curve in scatter plot the drew before using abline()
* Find the summary of the lm model created using summary() function

**Statistic:**

**Case 1-mtcars data:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 40.5789 | 2.8699 | 14.139 | 2.86e-09 \*\*\* |
| x | -6.5926 | 0.9145 | -7.209 | 6.86e-06 \*\*\* |

**Residual standard error: 2.201**

**R2: 0.7999** **Adj-R2: 0.7845**

**F-statistic:  51.97**  **p-value: 6.859e-06**

**Case 2-data.csv data:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 83.0858 | 49.8122 | 1.668 | 0.102 |
| x | 0.1407 | 0.2885 | 0.488 | 0.628 |

**Residual standard error: 32.08**

**R2: 0.004931** **Adj-R2: -0.0158**

**F-statistic:  0.2379**  **p-value: 0.628**

**Inference**:

**Case 1-mtcars:**

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-data.csv:**

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:**

rm(list=ls())

# linear regression for mtcars

library(dplyr)

data1<-mtcars

train=sample\_n(data1,15)

train

x=train$wt

y=train$mpg

plot(x,y,main="Scatter plot for train mtcars data")

cor.test(x,y)

lmodel<- lm(y~x)

abline(lmodel,col='red')

summary(lmodel)

print("by this inference the p value of b1 and b0 are of less than 0.05 that means the model is significant")

# linear regression for data.csv

setwd("C:/Abhi notes/class3-2/eda/lab")

data2<-read.csv("data.csv")

data2

train=sample\_n(data2,50)

x=train$Height

y=train$Weight

plot(x,y,main="Scatter plot for train data.csv data")

cor.test(x,y)

lmodel<- lm(y~x)

abline(lmodel,col='red')

summary(lmodel)

print("by this inference the p value of b1 and b0 are of more than 0.05 that means the model is not significant")