(a) Linear regression:- One Independent Variable using **lm()** function; Interpret the output of Model Analysis, Compute Coefficient of Determination(r^2), Interpret results. (**Mandatory**)

- ***Introduction:-***

  The general mathematical equation for a linear regression is:-

  $$y = ax + b$$

  Following is the description of the parameters used:-

  - **y** is the **response** variable.
  - **x** is the **predictor** variable.
  - **a** and **b** are **constants** which are called the **coefficients**.

lm() Function:-

This function creates the relationship model between the predictor and the response variable.

- ***Syntax:-***

  The basic syntax for **lm()** function in linear regression is:-

  lm(formula = y ~ x, data)

  Following is the description of the parameters used:-

  - **formula** is a symbol presenting the relation between x and y.
  - **data** is the vector on which the formula will be applied.

predict() Function:-

The basic syntax for predict() in linear regression is:-

predict(object, newdata)

Following is the description of the parameters used:-

- **object** is the formula which is already created using the lm() function.
- **newdata** is the vector containing the new value for predictor variable.

Problem:- Develop the equation of the simple regression line to predict sales(y) from advertising(x) expenditures using the given data:-

Advertising(x):-	12.5	3.7	21.6	60.0	37.6	6.1	16.8	41.2
Sales(y):-	148	55	338	994	541	89	126	379

Determine the predicted value of **Sales(y) =** ? for **Advertising(x)** = **50**. Compute $\mathbf{r^2}$

Solution:-

```
x = c(12.5, 3.7, 21.6, 60, 37.6, 6.1, 16.8, 41.2)      #create a vector.
y = c(148, 55, 338, 994, 541, 89, 126, 379)            #create a vector.
y.lm = lm(y~x)
coeffs = coefficients(y.lm)
coeffs                    #we get the value of the coefficients of simple regression line
```

Output:-

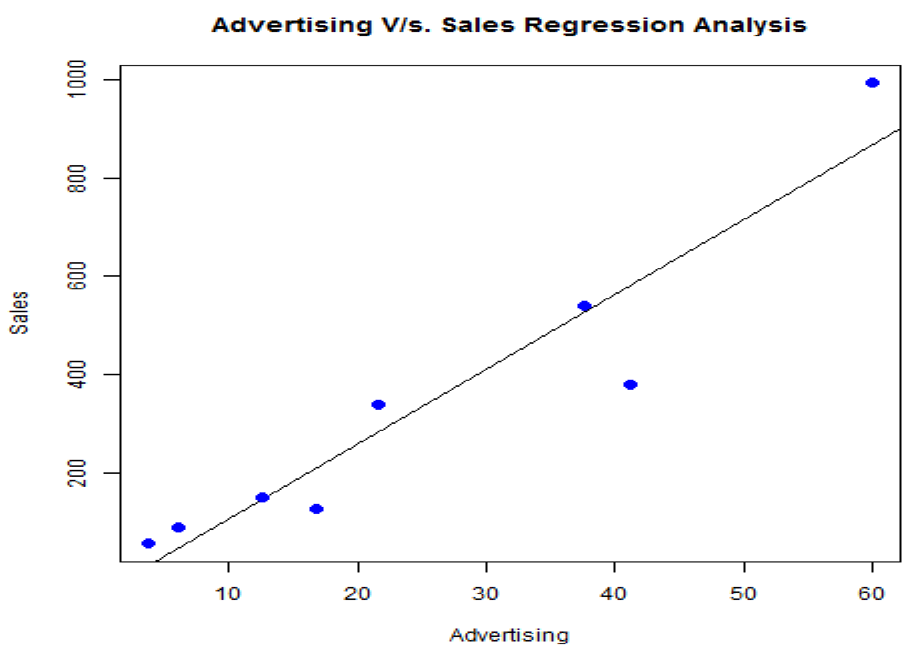
```
(Intercept)           x
  -46.29181    15.23977
```

newdata = data.frame(x = 50)
predict(y.lm, newdata) #To compute y = ?, when x = 50 is given predict() is used
Output:-
```
       1
715.6968
```

print(summary(y.lm)) #It displays the output of Model Analysis.
Output:-
```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-202.59  -18.09   28.30   47.46  125.91

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -46.292     64.891  -0.713 0.502402
x             15.240      2.096   7.271 0.000344 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 108.8 on 6 degrees of freedom
Multiple R-squared:  0.8981,	Adjusted R-squared:  0.8811
F-statistic: 52.86 on 1 and 6 DF,  p-value: 0.0003445
```

summary(y.lm)$r.squared #It calculate the value of **r**^{2} separately/directly.
Output:- 0.8980643

png(file = "Linear Regression.png") #create an image of scatter diagram
plot(x, y, col="blue", main="Advertising V/s. Sales Regression Analysis", abline(lm(y~x)), cex = 1.3, pch=16, xlab="Advertising", ylab="Sales") #use the function of scatterplot
dev.off() #save the file
Output:-



(b) Linear regression:- Multiple Independent Variables using **lm()** function; Interpret the output of Model Analysis. (**Mandatory**)

- ***Introduction:-***

If we choose the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta_k}$ (k = 1, 2, ..., p) in the multiple linear regression model so as to minimize the sum of squares of the error term $\boldsymbol{\epsilon}$, we will have the so called estimated multiple regression equation. It allows us to compute fitted values of $\boldsymbol{y}$ based on a set of values of $\boldsymbol{x_k}$ (k = 1, 2, ..., p) .

$$\hat{y} = a + \sum_k b_k x_k$$

Following is the description of the parameters used:-

- **y** is the **response** variable.
- **a, b_1, b_2...b_n** are the **coefficients**.
- **x_1, x_2, ...x_n** are the **predictor** variables.

lm() Function:-

This function creates the relationship model between the predictor and the response variable.

Syntax:-

The basic syntax for **lm()** function in multiple regression is:-

lm(formula = y ~ x_1 + x_2 + x_3 ..., data)

Following is the description of the parameters used:-

- **formula** is a symbol presenting the relation between x and y.
- **data** is the vector on which the formula will be applied.

Problem:-

Use a computer to develop the equation of the regression model for the following data. Comment on the regression coefficients. Determine the predicted value of **y** for **x_1 = 33**, **x_2 = 29** and **x_3 = 13**.

y	x_1	x_2	x_3
114	21	6	5
94	43	25	8
87	56	42	25
98	19	27	9
101	29	20	12
85	34	45	21
94	40	33	14
107	32	14	11
119	16	4	7
93	18	31	16
108	27	12	10
117	31	3	8

Solution:-

y = c(114, 94, 87, 98, 101, 85, 94, 107, 119, 93, 108, 117) #create a vector.
x_1 = c(21, 43, 56, 19, 29, 34, 40, 32, 16, 18, 27, 31) #create a vector.
x_2 = c(6, 25, 42, 27, 20, 45, 33, 14, 4, 31, 12, 3) #create a vector.
x_3 = c(5, 8, 25, 9, 12, 21, 14, 11, 7, 16, 10, 8) #create a vector.
y.lm = lm(y ~ x_1 + x_2 + x_3)
coeffs = coefficients(y.lm)
coeffs #we get the value of the coefficients of multiple regression model
Output:-

```
(Intercept)          x1          x2          x3
118.55951024  -0.07940245  -0.88428115   0.37690982
```

newdata = data.frame(x_1 = 33, x_2 = 29, x_3 = 13)
predict(y.lm, newdata) #To compute y = ?, when x_1 = 33, x_2 = 29, x_3 = 13 is given predict() is used
Output:-

```
          1
     95.1949
```

print(summary(y.lm)) #It displays the output of Model Analysis.
Output:-

```
Call:
lm(formula = y ~ x1 + x2 + x3)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7481 -1.6934  0.5343  1.1214  2.6097 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 118.55951    1.85798  63.811 4.05e-12 ***
x1           -0.07940    0.06848  -1.159    0.280    
x2           -0.88428    0.08631 -10.245 7.08e-06 ***
x3            0.37691    0.21973   1.715    0.125    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.134 on 8 degrees of freedom
Multiple R-squared:  0.975, Adjusted R-squared:  0.9656 
F-statistic: 103.8 on 3 and 8 DF,  p-value: 9.582e-07
```

summary(y.lm)$r.squared #It calculate the value of $\mathbf{r^2}$ separately/directly.
Output:- `0.9749568`