Experiment 3

Aim :- Implement R Program for Multiple Linear Regression.

Objective:- To understand the use of Multiple linear regression techniques by implementing a predefined dataset of R Studio.

Description-

Multiple linear regression is the extension of linear regression in the relationship between more than two variables. In simple linear regression, we have one predictor and one response variable. But in multiple regressions, we have more than one predictor variable and one response variable.

There is the following general mathematical equation for multiple regression -

$$y=b_0+b_1*x_1+b_2*x_2+b_3*x_3+\cdots b_n*x_n$$

Here,

- y is a response variable.
- bO, bl, b2...bn are the coefficients.
- xl, x2, ...xn are the predictor variables.

In R, we create the regression model with the help of the **Im()** function. The model will determine the value of the coefficients with the help of the input data. We can predict the value of the response variable for the set of predictor variables using these coefficients.

There is the following syntax of lm() function in multiple regressio

1. lm(yx l+x2+x3, data)

Before proceeding further, we first create our data for multiple regression. We will use the "mtcars" dataset present in the R environment. The main task of the model is to create the relationship between the "mpg" as a response variable with "wt", "disp" and "hp" as predictor variables.

For this purpose, we will create a subset of these variables from the "mtcars" dataset.

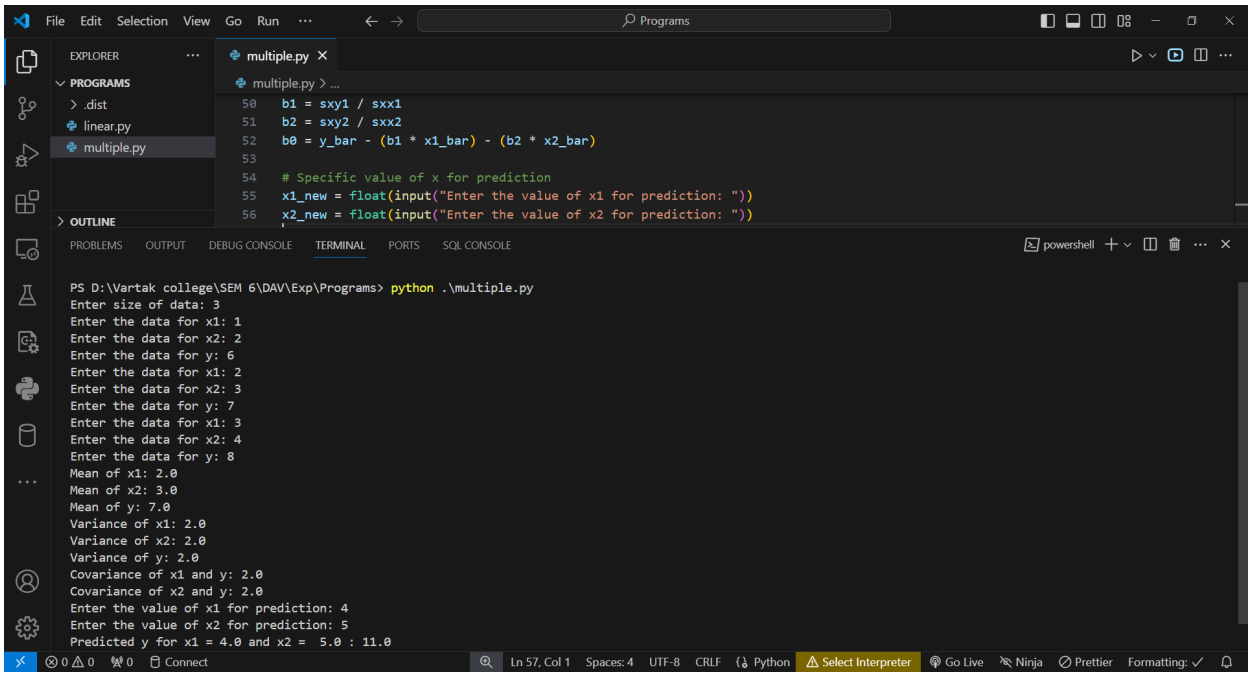
1. data<-mtcars[,c("mpg","wt","disp","hp")]
2. print(head(input)

Program:

```
n = int(input("Enter size of data: "))
x1 = []
x2 = []
y = []
sum_x1 = 0
sum_x2 = 0
sum_y = 0
# Collect data points
for i in range(0, n):
    x1.append(float(input("Enter the data for x1: ")))
    x2.append(float(input("Enter the data for x2: ")))
    y.append(float(input("Enter the data for y: ")))
    sum_x1 += x1[i]
    sum_x2 += x2[i]
    sum_y += y[i]
# Calculate mean
x1_bar = sum_x1 / n
x2_bar = sum_x2 / n
y_bar = sum_y / n
print("Mean of x1:", x1_bar)
print("Mean of x2:", x2_bar)
print("Mean of y:", y_bar)
sxx1 = 0
sxx2 = 0
sxy1 = 0
```

```
sxy2 = 0
syy = 0
# Calculate variances and covariance
for i in range(0, n):
    sxx1 += (x1[i] - x1_bar) ** 2
    sxx2 += (x2[i] - x2_bar) ** 2
    syy += (y[i] - y_bar) ** 2
    sxy1 += (y[i] - y_bar) * (x1[i] - x1_bar)
    sxy2 += (y[i] - y_bar) * (x2[i] - x2_bar)
print("Variance of x1:", sxx1)
print("Variance of x2:", sxx2)
print("Variance of y:", syy)
print("Covariance of x1 and y:", sxy1)
print("Covariance of x2 and y:", sxy2)
# Calculate slope (b1) and intercept (b0)
b1 = sxy1 / sxx1
b2 = sxy2 / sxx2
b0 = y_bar - (b1 * x1_bar) - (b2 * x2_bar)
# Specific value of x for prediction
x1_new = float(input("Enter the value of x1 for prediction: "))
x2_new = float(input("Enter the value of x2 for prediction: "))
# Calculate predicted y for the specific value of x
y_pred = b0 + (b1 * x1_new) + (b2 * x2_new)
print("Predicted y for x1 =", x1_new, "and x2 = ", x2_new, ":", y_pred)
```

Output:



Conclusion-

1. **Equation for multiple linear regression is**

   Equation for multiple linear regression is represented as:

   $$y=b0+b1\cdot x1+b2\cdot x2+\ldots+bn\cdot xn$$

   where:

   - y is the response variable.
   - b0,b1,b2,…,bn are the coefficients.
   - x1,x2,…,xn are the predictor variables.

2. **When there is only one dependent variable and multiple independent variable then this types of regression is known as** Multiple Regression.

3. **How to check inbuilt dataset in R studio?**

   To check an inbuilt dataset in R Studio, you can use the following command:

   ```
   data()
   ```

   This command lists the available datasets in the R environment, which you can then explore further using commands like head() to view the first few rows of a dataset.