



EXPERIMENT NO 2

Aim: To write the implementation of linear regression.

Objective:- To understand the use of simple linear regression techniques by implementing user define dataset and importing dataset

Description:

Regression analysis is a very widely used statistical tool to establish a relationship model between two variables. One of these variables is called a predictor variable whose value is gathered through experiments. The other variable is called response variable whose value is derived from the predictor variable.

In Linear Regression these two variables are related through an equation, where the exponent (power) of both these variables is 1. Mathematically a linear relationship represents a straight line when plotted as a graph. A non-linear relationship where the exponent of any variable is not equal to 1 creates a curve.

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.

Procedure:

The steps to create the relationship is

1. Carry out the experiment of gathering a sample of observed values of height and corresponding weights.
2. Create a relationship model using the `lm()` functions in R.
3. Find the coefficients from the model created and create the mathematical equation using these. Get a summary of the relationship model to know the average error in prediction. Also called residuals.

To predict the weight of new persons, use the `predict()` function in R.



Program:

```
n = int(input("Enter size of data: "))
x = []
y = []
sum_x = 0
sum_y = 0
# Collect data points
for i in range(0, n):
    x.append(float(input("Enter the data for x: ")))
    y.append(float(input("Enter the data for y: ")))
    sum_x += x[i]
    sum_y += y[i]
# Calculate mean
x_bar = sum_x / n
y_bar = sum_y / n
print("Mean of x:", x_bar)
print("Mean of y:", y_bar)
sxx = 0
sxy = 0
syy = 0
# Calculate variances and covariance
for i in range(0, n):
    sxx += (x[i] - x_bar) ** 2
    syy += (y[i] - y_bar) ** 2
    sxy += (y[i] - y_bar) * (x[i] - x_bar)
print("Variance of x:", sxx)
print("Variance of y:", syy)
print("Covariance of x and y:", sxy)
# Calculate slope (b1) and intercept (b0)
b1 = sxy / sxx
b0 = y_bar - (b1 * x_bar)
# Specific value of x for prediction
x_new = float(input("Enter the value of x for prediction: "))
# Calculate predicted y for the specific value of x
```



Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

```
y_pred = b0 + b1 * x_new  
print("Predicted y for x =", x_new, ":", y_pred)
```

Output:

```
linear.py  
1 n = int(input("Enter size of data: "))  
2  
3 x = []  
4 y = []  
5 sum_x = 0  
6 sum_y = 0  
7  
8 # Collect data points  
9 for i in range(0, n):  
10     x.append(int(input("Enter the data for x: ")))  
11     y.append(int(input("Enter the data for y: ")))  
12  
13 # Calculate mean, variance, and covariance  
14 mean_x = sum_x / n  
15 mean_y = sum_y / n  
16 var_x = (sum_x**2 / n) - (sum_x**2 / n)  
17 var_y = (sum_y**2 / n) - (sum_y**2 / n)  
18 cov_xy = (sum_x * sum_y) / n - (sum_x * sum_y) / n  
19  
20 # Predict y for a new x value  
21 x_new = int(input("Enter the value of x for prediction: "))  
22 y_pred = b0 + b1 * x_new  
23 print("Predicted y for x =", x_new, ":", y_pred)
```

PS D:\Vartak college\SEM 6\DAV\Exp\Programs> python .\linear.py
Enter size of data: 5
Enter the data for x: 1
Enter the data for y: 2
Enter the data for x: 2
Enter the data for y: 3
Enter the data for x: 3
Enter the data for y: 4
Enter the data for x: 4
Enter the data for y: 5
Enter the data for x: 5
Enter the data for y: 6
Mean of x: 3.0
Mean of y: 4.0
Variance of x: 10.0
Variance of y: 10.0
Covariance of x and y: 10.0
Enter the value of x for prediction: 6
Predicted y for x = 6.0 : 7.0
PS D:\Vartak college\SEM 6\DAV\Exp\Programs>

Conclusion:

1. Function used for linear regression in R is

The function used for linear regression in R is `lm()`. This function is employed to create a linear regression model based on given data, where it estimates the coefficients (constants) of the linear equation relating the predictor variable (x) to the response variable (y).

2. Explain use of `summary()`.

The `summary()` function in R is utilized to obtain a summary of statistical information about the linear regression model. This includes details such as coefficients, standard errors, t-values, p-values, and the coefficient of determination (R-squared). It provides a comprehensive overview of the model's performance and the significance of each predictor variable in explaining the variation in the response variable.