**Experiment no 11:-**

**Aim:-** Analyzing Data - Implementation of  Regression Analysis .

**Theory:-**

# Linear Regression

• It is a statistical approach for modeling relationship between a dependent variable (responses) with a given set of independent variables(features).

# Simple Linear Regression

* Simple linear regression is an approach for predicting a response using a single feature.
* It is assumed that the two variables are linearly related. Hence, we try to find a linear function that predicts the response value(y) as accurately as possible as a function of the feature or independent variable(x).
* Let us consider a dataset where we have a value of response y for every feature x:
* For generality, we define:
* x as feature vector, i.e x = [x1, x2, …., xn],
* y as response vector, i.e y = [y1, y2, …., yn ]
* for n observations
* Now, the task is to find a line which fits best in above scatter plot so that we can predict the response for any new feature values. (i.e a value of x not present in dataset)
* This line is called regression line.
* The equation of regression line is represented as:
* h(𝑥𝑖)= β0 + β1 𝑥𝑖
* Here,
* h(𝑥𝑖) represents the predicted response value for ith observation.
* β0 and β1  are regression coefficients and represent y-intercept and slope of regression line respectively.

## The mathematical technique to fit a regression line involves estimating the values of β0 , β1

This has to be done in such a way that the sum of the squared errors for all the data points is minimum. This is known as the least squares criterion.

**Scenario:-**

Uni-super ltd in it’s 1994 annual report listed the number of register in each states and the membership in the super amuation scheme for each state as given in the first three columns of the tables the problem is to regress y on x .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **States** | **NO of institutions (X)** | **Membership(Y)** | **X**2 | **XY** |
| NSW | 17 | 5987 | 289 | 101779 |
| QLD | 11 | 5950 | 121 | 65450 |
| sa | 10 | 3588 | 100 | 35880 |
| tas | 3 | 1356 | 9 | 4068 |
| VIC | 41 | 14127 | 1681 | 579207 |
| WA | 9 | 4847 | 81 | 43623 |
| Others | 11 | 3893 | 121 | 42823 |
| Total | 102 | 39748 | 2402 | 872830 |
|  |  |  |  |  |
|  | SLR:- y = mx + c |  |  |  |
|  | Avg(Y) | 5678.285714 |  |  |
|  | Avg(X) | 14.57142857 |  |  |
|  | Sxy | sum of cross product deviation | 293644.8571 |  |
|  |  |  |  |  |
|  | SXX | sum of squared deviation X | 915.7142857 |  |
|  | b1 | SXY/SXX | 320.6730109 |  |
|  | b0 | Avg(Y)-b1\*Avg(X) | 1005.621841 |  |
|  |  |  |  |  |
|  |  | y= 1005 + 320.7x |  |  |







