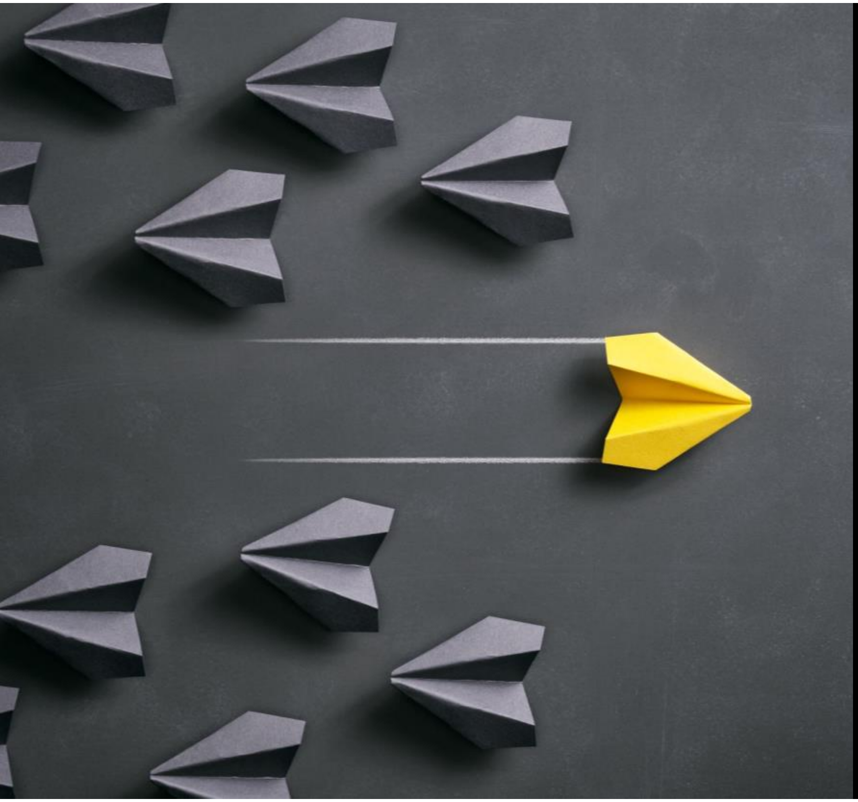


PRACTICAL 9
LINEAR AND MULTIPLE LINEAR REGRESSION

Dr. Sudeep Tanwar

LINEAR REGRESSION

In regression, one variable is considered independent (=predictor) variable (X) and the other the dependent (=outcome) variable Y.

Widely applied:

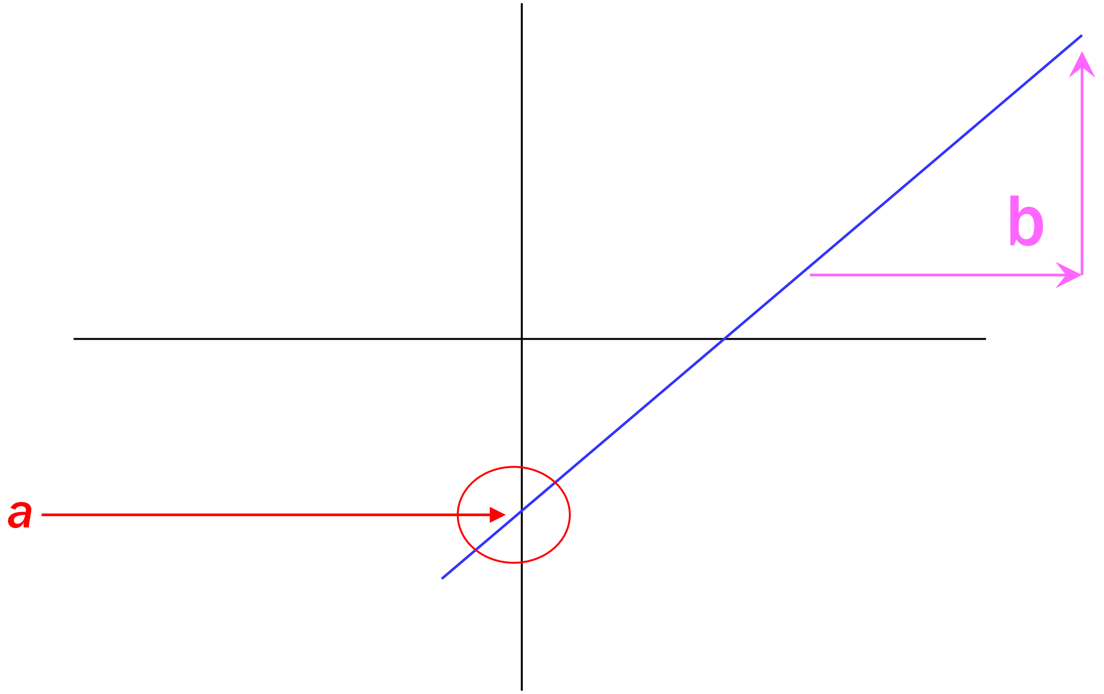
- to predict the outputs,
- forecasting the data,
- analyzing the time series, and
- finding the causal effect dependencies between the variables

SIMPLE LINEAR REGRESSION

- Only one independent variable, X
- Relationship between X and Y is described by a linear function
- Changes in Y are assumed to be caused by changes in X

SIMPLE LINEAR REGRESSION

$$y = a + bx$$



SIMPLE LINEAR REGRESSION

a and *b* are given by the following formulas:

$$a\,(intercept) = \frac{\sum y \sum x^2 - \sum x \sum xy}{(\sum x^2) - (\sum x)^2}$$

$$b\,(slope) = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

Where,

x and *y* are two variables on the regression line.

b = Slope of the line.

a = *y*-intercept of the line.

x = Values of the first data set.

y = Values of the second data set.

Question: Find linear regression equation for the following two sets of data:

x	2	4	6	8
y	3	7	5	10

Solution:

Construct the following table:

x	y	x^2	xy
2	3	4	6
4	7	16	28
6	5	36	30
8	10	64	80
$\sum x = 20$	$\sum y = 25$	$\sum x^2 = 120$	$\sum xy = 144$

EXAMPLE

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$b = \frac{4 \times 144 - 20 \times 25}{4 \times 120 - 400}$$

b = 0.95

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{25 \times 120 - 20 \times 144}{4(120) - 400}$$

a = 1.5

Linear regression is given by:

y = a + bx

y = 1.5 + 0.95 x

CODE

```
for i in range(10):
        x.append(random.randint(1,100));
        y.append(random.randint(1,100));
x.sort();
y.sort();
n=len(x);
E_y=sum(y);
E_x=sum(x);
E_xy=sum([i*j for i,j in zip(x,y)]);
E_x^2=sum([i*i for i in x]);
E_y^2=sum([i*i for i in y]);
intercept=(E_y*E_x^2-E_x*E_xy)/(E_x^2-E_x**2);
slope=(n*E_xy-E_x*E_y)/(n*E_x^2-E_x**2);
```

```
x_val=list(range(80));
y_val=[intercept+slope*i for i in x_val];
print(\"intercept={}, slope={}\".format(intercept,slope));
plt.plot(x,y);
plt.plot(x_val,y_val);
plt.xlabel(\"X-axis\");
plt.ylabel(\"Y-axis\");
plt.title(\"Linear Regression\");
plt.show();
```

MULTIPLE LIN REGRESSION

In linear regression, there is only one independent and dependent variable involved. But, in the case of multiple regression, there will be a set of independent variables that helps us to explain better or predict the dependent variable y.

The multiple regression equation is given by

$$Y = a + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k + e$$

MULTIPLE LIN REGRESSION

Multiple linear regression is used to estimate the relationship between **two or more independent variables** and **one dependent variable**.

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

$$b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

and

$$b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$