Regression:

Regression analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data.

In regression analysis, there are two types of variables: independent variable, dependent variable. The variable whose value is influenced or changed is called dependent variable, and the variable that influences is called independent variable.

There are two types of linear regression -

i) Simple linear regression (1dependent +1 independent)
ii) Multiple linear regression (1dependent +more than 1 independent)

Simple Linear Regression Model:

A simple linear regression model of a dependent variable y on an independent variable x is given by –

$$y_i = \alpha + \beta x_i + \varepsilon \ \ , \ i = 1, 2, . \ . \ \ ., n \ ,$$

Where,

y_i = Dependent variable

x_i = Independent variable, $\alpha = Intercept$, β = Slope/ Regression coefficient,

$$\varepsilon = Error\ term$$

Assumptions:

The assumptions of regression line $Y = \alpha + \beta X + \varepsilon$

i. The X are non- random.
ii. The distribution of error term ε is normal with mean 0 and variance σ^2.
iii. The $y_1, y_2, . \ . \ ., y_n$ are independent.
iv. Y is linear in X.
v. Y is normally distributed with mean μ and variance σ^2.

OLS estimate of α and β :

When y depends on x then-

$$\beta_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}} \quad , \quad i = 1, 2,3, \dots n$$

$n = sample\ size$

$$\alpha = \bar{y} - \beta_1 \bar{x}$$

$$= \frac{\sum y_i}{n} - \beta_1 \frac{\sum x_i}{n}$$

When x depends on y then-

$$\beta = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum y_i{}^2 - \frac{(\sum y_i)^2}{n}} \quad , \quad i = 1, 2,3, \dots n$$

$n = sample\ size$

$$\alpha = \bar{x} - \beta \bar{y}$$

$$= \frac{\sum x_i}{n} - \beta \frac{\sum y_i}{n}$$

Interpretation:

$\boldsymbol{\alpha}$ **:** The average value of dependent variable is α when the independent variable is zero.

$\boldsymbol{\beta}$ **:** For one unit increase in independent variable, the dependent variable will change on average β unit.

Coefficient of determination:

It determines the proportion of total variation in the dependent variable explained by the independent variable. In regression analysis, coefficient of determination is denoted by $\boldsymbol{R^2}$.

$$R^2 = 1 - \frac{Unexplained\ variation\ of\ y}{Total\ \ variation\ of\ y}$$

$$= 1 - \frac{Error\ sum\ of\ square}{Total\ sum\ of\ square}$$

Interpretation:

Proportion/percentage of variation in dependent variable can be explained by the independent variable.

Example:

The following data represents the demand and supply of 8 items.

Demand	Supply
52	82
55	90
59	95
62	106
63	120
66	132
70	140
71	135

a) Fit a regression line when – i) Demand depends on Supply ii) Supply depends on Demand
b) Estimate the demand for the supply of 82.
c) Estimate the supply for the demand of 52.

Solution:

Calculation Table

	Demand	Supply	$(Demand)^2$	$(Supply)^2$	Demand* Supply
	52	82	2704	6724	4264
	55	90	3025	8100	4950
	59	95	3481	9025	5605
	62	106	3844	11236	6572
	63	120	3969	14400	7560
	66	132	4356	17424	8712
	70	140	4900	19600	9800
	71	135	5041	18225	9585
Total	498	900		104734	57048

a) i. Demand depends on supply:

Demand = Dependent variable = y

Supply = Independent variable = x

Then,

$$\beta_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}}$$

Here,

$\sum x_i = 900, \sum y_i = 498, \sum x_i{}^2 = 104734, \sum x_i y_i = 57048$

$$\beta_1 = \frac{57048 - \frac{498 \times 900}{8}}{104734 - \frac{900^2}{8}}$$

$$= 0.2936$$

Interpretation: If the value of supply is increased by one unit, there will be an average increase in demand by 0.2936 unit.

$$\boldsymbol{\alpha} = \frac{\sum \boldsymbol{y_i}}{\boldsymbol{n}} - \beta_2 \frac{\sum \boldsymbol{x_i}}{\boldsymbol{n}}$$

$$= \frac{\mathbf{498}}{\mathbf{8}} - 0.2936\frac{900}{8}$$

$$= 29.22$$

Interpretation:

The average supply is 29.22 when the demand is zero.

The fitted regression line is-

Y = 29.22 + 0.2936X

a) ii. Supply depends on Demand:

Supply = Dependent variable = y

Demand = Independent variable = x

Then,

$$\beta_2 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum {x_i}^2 - \frac{(\sum x_i)^2}{n}}$$

Here,

$\sum x_i = 498, \sum y_i = 900, \sum {x_i}^2 = 31320, \sum x_i y_i = 57048$

$$\beta_2 = \frac{57048 - \frac{498 \times 900}{8}}{31320 - \frac{498^2}{8}}$$

$$= 3.20$$

Interpretation: If the value of demand is increased by one unit, there will be an average increase in supply by 3.20 unit.

$$\boldsymbol{\alpha} = \frac{\sum \boldsymbol{y_i}}{\boldsymbol{n}} - \beta_2 \frac{\sum \boldsymbol{x_i}}{\boldsymbol{n}}$$

$$= \frac{\mathbf{900}}{\mathbf{8}} - 3.20 \times \frac{498}{8}$$

$$= -86.7$$

Interpretation:

The average demand is -86.7 when the supply is zero.

The fitted regression line is-

Y = -86.7 + 3.20 X

b. Estimation demand for the supply of 82

Demand =Dependent variable = y

Supply = Independent variable = x = 82

Y = 29.22 + 0.2936 X

= 29.22 +0.2936× 82

= 53.2952

c. Estimation supply for the demand of 52

Supply =Dependent variable = y

Demand = Independent variable = x = 52

Y = -86.7 + 3.20 X

$= -86.7 + 3.20 \times 52$

= 79. 7