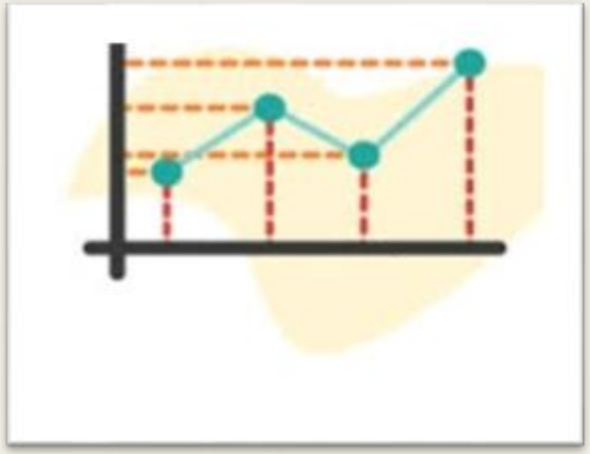
B. COM
SEMESTER II
CC 204
REGRESSION ANALYSIS.

PREPARED BY- PUJA KUMARI
ASSISTANT PROFESSOR
DEPARTMENT OF COMMERCE
PATNA WOMEN’S COLLEGE, PATNA.
EMAIL. ID- pujasngh26@gmail.com

Regression Analysis:



- The statistical tool with the help of which we are in a position to estimate (or predict) the unknown values of one variable from known values of another variable is called **Regression**.
- The dictionary meaning of the term 'regression' is the act of returning or going back.
- The line describing this tendency to regress or going back was called by Galton a '**Regression Line'.**
- Regression Analysis is the branch of statistical theory widely used in almost all the disciplines.
- It is the basic technique for measuring or estimating the relationship among economic variables that constitute the essence of economic theory and economic life.

Significance of Regression analysis :

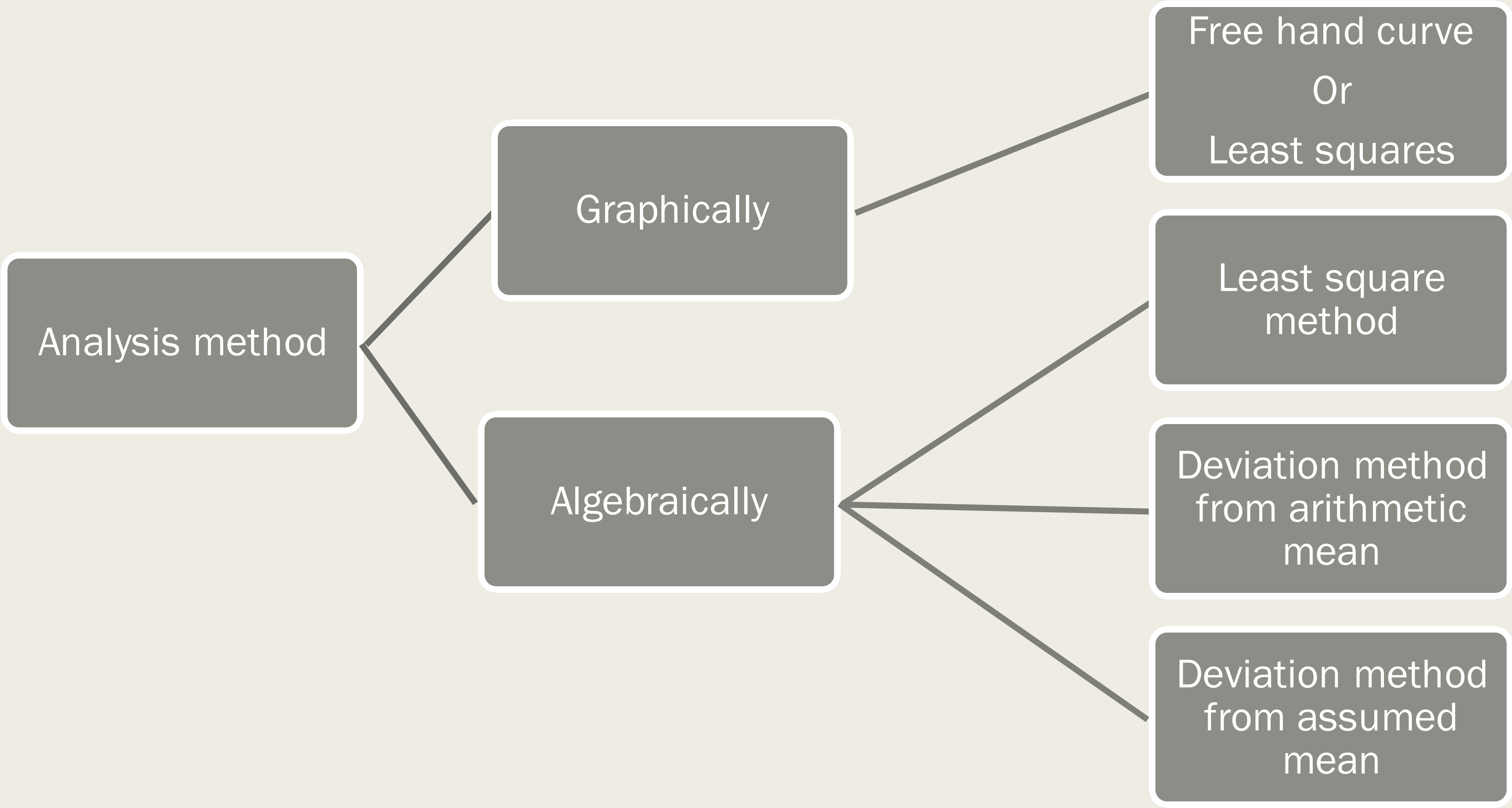
- It’s a statistical methodology that helps estimate the strength and direction of the relationship between two or more variables.
- Predict sales in the near and long term.
- It also helps us to understand inventory levels, supply and demand.

Uses of Regression analysis in companies:

Companies might use regression analysis to understand, for example:

- Why customer service calls dropped in the past year or even the past month.
- Predict what sales will look like in the next six month.
- Whether to chose one marketing promotion over another.
- Whether to expand the business or create and market a new product.

Methods of Regression analysis:



Algebraic Methods:

1. Least square method

- For X on Y

  The regression equation :

  $$X = a + bY$$

  where, X = Dependent variable

  Y = Independent variable

  Normal equation to find value of a & b :-

  $$\sum X = n a + b \sum Y$$ (n= number of pair of observation given)

  $$\sum XY = a \sum Y + \sum y^2$$

Algebraic Method Continued:

- For Y on X

The regression equation :

$$Y = a + b X$$

where, Y= Dependent variable

X = Independent variable

Normal equation to find value of a & b :-

$$\sum Y = n a + b \sum X$$ (n= number of pair of observation given)

$$\sum XY = a \sum X + \sum X^2$$

- **Example** : From the given data obtain the two regression equations using method of least squares.

X	0	1	2	3	4	5	6
Y	3	2	1	4	5	6	2

- Solution:

X	Y	XY	X^2	Y^2
0	3	0	0	9
1	2	2	1	4
2	1	2	4	1
3	4	12	9	16
4	5	20	16	25
5	6	30	25	36
6	2	12	36	4
$\sum X=$	$\sum Y$	$\sum XY = 78$	$\sum X^2 = 91$	$\sum Y^2 = 95$

1) For Y on X

Normal equations are :

$$\sum Y = n\,a + b \sum X$$

$$\sum XY = a \sum X + b \sum X2$$

So, $23 = 7a + 21b$ 1

$78 = 21a + 91b$ 2

From solving eq.1 and eq.2, we get ;

$a = 2.32$ and $b = 0.32$

Now, Regression equation for Y on X is

$$Y = a + bX$$

So, $Y = 2.32 + 0.32$

2) For X on Y

Normal equations are :

$$\sum X = n\,a + b \sum Y$$

$$\sum XY = a \sum Y + b \sum Y2$$

So, $21 = 7a + 23b$ 1

$78 = 23a + 95b$ 2

From solving eq.1 and eq.2, we get ;

$a = 1.47$ and $b = 0.46$

Now, Regression equation for X on Y is

$$X = a + bY$$

So, $X = 1.47 + 0.46\,Y$

DEVIATION FROM ARITHMETIC MEAN METHOD:

- The calculation by the least squares method are quit cumbersome when the values of X and Y are large. So the work can be simplified by using this method.

Steps:-

- Obtain $\bar{X}$ (mean of x – series)
- Obtain $\bar{Y}$ (mean of Y – series)
- Obtain b_{XY} (Regression coefficient of X on Y)
- Regression Coefficient equation of Y on X

  $(Y - \bar{Y}) = b_{YX} (X - \bar{X})$
- Regression coefficient equation of X on Y

  $(X - \bar{X}) = b_{XY} (Y - \bar{Y})$

- Where

$$b_{xy} = \frac{\sum xy}{\sum x^2}$$

$$and$$

$$b_{yx} = \frac{\sum xy}{\sum y^2}$$

- **Example:** From the following data, Find the regression equations

X	6	2	10	4	8
Y	9	11	5	8	7

Solution:

$\bar{X} = \sum X/n = 30/5 = 6$

$\bar{Y} = \sum Y/n = 40/5 = 8$

X	y	$x= X - \bar{X}$	$y= Y - \bar{Y}$	x^2	y^2	Xy
6	9	0	1	0	1	0
2	11	-4	3	16	9	-12
10	5	4	-3	16	9	-12
4	8	-2	0	4	0	0
8	7	2	-1	4	1	-1
$\sum X=30$	$\sum Y=40$	$\sum x=0$	$\sum y=0$	$\sum x^2= 40$	$\sum y^2= 20$	$\sum xy= -25$

(1) Regression equation for X on Y is

$$(X-\bar{X}) = b_{xy}(Y-\bar{Y})$$

$$b_{xy} = \frac{\sum xy}{\sum y^2} = \frac{-25}{20} = -1.25$$

$$X-6 = -1.25(Y-8)$$

$$X-6 = -1.25Y+10$$

$$X = -1.25Y+16$$

2) Regression equation for Y on X

$$(Y-\bar{Y}) = b_{yx}(X-\bar{X})$$

$$b_{yx} = \frac{\sum xy}{\sum x^2} = \frac{-25}{40} = -0.63$$

$$Y-8 = -0.63(X-8)$$

$$Y-8 = -0.63X+3.78$$

$$Y = -0.63+11.78$$

Deviation from assumed mean method:

As the name suggest, here we assume any mean and then solve the problem on basis of that.

Regression Coefficient equation of Y on X

$$(Y - \overline{Y}) = b_{YX} \; (X - \overline{X})$$

Regression coefficient equation of X on Y

$$(X - \overline{X}) = b_{XY} \; (Y - \overline{Y})$$

but here, formula for b_{XY} and b_{YX} will be different from arithmetic mean method.

$$b_{xy} = \frac{n\sum d_x d_y - \sum d_x \sum d_y}{n\sum d_y{}^2 - \sum d_y{}^2}$$

$$b_{yx} = \frac{n\sum d_x d_y - \sum d_x \sum d_y}{n\sum d_x{}^2 - \sum d_x{}^2}$$

- Example: From the following data, Find the regression equations by assumed mean method.

X	1	1	2	4	5
Y	2	3	1	2	4

Solution : Assume 2 as the mean for x series and 2 as mean for y series. (we can assume different mean for each. Its not necessary to pick same value)

X	y	Dev. From assumed mean 2		Dev. From assumed mean 2		
1	2	-1	1	0	0	0
1	3	-1	1	1	1	-1
2	1	0	0	-1	1	0
4	2	2	4	0	0	0
5	4	3	9	2	4	6
∑X=13	∑Y=12	∑ =3				

- $\bar{X} = (\sum x)/(n)$ and $\bar{Y} = (\sum y)/(n)$

  Regression coefficient of X on Y:

$$b_{xy} = \frac{n\sum d_x d_y - \sum d_x \sum d_y}{n\sum d_y^{\,2} - \sum d_y^{\,2}}$$

$$b_{xy} = \frac{5(5) - (3)(2)}{(5)(6) - 6} = 0.79$$

Regression equation of X on Y :

$$(X - \bar{X}) = b_{xy}\,(Y - \bar{Y})$$

$$X - 2.6 = 0.79(Y - 2.4)$$

$$X = 0.79Y - 1.89 + 2.6$$

$$X = 0.79Y + 0.71$$

Regression coefficient of Y on X :

$$b_{yx} = \frac{n\sum d_x d_y - \sum d_x \sum d_y}{n\sum d_x^{\,2} - \sum d_x^{\,2}}$$

$$b_{yx} = \frac{5(5) - (3)(2)}{5(15) - 15} = \frac{19}{60} = 0.31$$

- Regression equation of Y on X:

$$(Y - \bar{Y}) = b_{YX} \ (X - \bar{X})$$

$$Y - 2.4 = 0.31 \ (X - 2.6)$$

$$Y = 0.31X + 1.58$$

Regression Coefficient:

- Regression coefficients are estimates of the unknown population parameters and describe the relationship between a predictor variable and the response. In linear regression, coefficients are the values that multiply the predictor values. Suppose you have the following regression equation: $y = 3X + 5$. In this equation, +3 is the coefficient, X is the predictor, and +5 is the constant.
- The sign of each coefficient indicates the direction of the relationship between a predictor variable and the response variable.
- A positive sign indicates that as the predictor variable increases, the response variable also increases.
- A negative sign indicates that as the predictor variable increases, the response variable decreases.
- The coefficient value represents the mean change in the response given a one unit change in the predictor. For example, if a coefficient is +3, the mean response value increases by 3 for every one unit change in the predictor.

Properties Of Regression Coefficient:

The coefficient of correlation is the geometric mean of the two regression coefficients.

- If one of the regression coefficient is greater than unity, the other must be less than unity.
- Both the regression coefficients will have the same sign, i.e., they will be either positive or negative.
- The coefficient of correlation will have the same sign as that of regression coefficients.
- The average value of the two regression coefficients would be greater than the value of correlation.
- Regression coefficients are independent of change of origin but not scale.

THANK YOU