0302301 - STATISTICS FOR DATA ANALYSIS

UNIT	MODULE	WEIGHTAGE
1.	STATISTICS: OVERVIEW	20%
2.	MEASURE OF DISPERSION	20%
3.	CORRELATION AND REGRESSION	20%
4.	FUNDAMENTALS OF PROBABILITY	20%
5.	STATISTICAL ANALYSIS USING R PROGRAMMING	20%

0302301 - STATISTICS FOR DATA ANALYSIS

Text Book: Business statistics by Padmalochan hazarika

Related Programming Tool: R

UNIT - 3 Correlation and Regression

- Correlation Analysis
  - Introduction
  - Types of correlation
  - Positive, negative and zero correlation
  - Linear and Non - linear correlation
  - Simple, Multiple and Partial correlation
  - Karl Pearson method for measuring correlation
- Regression Analysis
  - Introduction
  - Method of least square
  - Regression lines
  - The regression equation Y on X
  - The regression equation X on Y
  - Regression coefficient and its properties (without proof)

Correlation Analysis

<u>Introduction</u>

- Relationship between two variables.
- Two variables are said to be correlated when the value of one variable changes with the change in change in the value of the other variable.
- If two variable are such that when one changes the other also changes then we say that there is correlation between them.
- Example: <u>**Price falls & demand high / Price rise & demand falls.**</u>
- Thus, there is a correlation between two variables <u>***price and demand***</u>.

Correlation Analysis

Types of Correlation

- Positive, Negative and Zero Correlation
- Linear and Nonlinear (curvilinear) Correlation
- Simple, Multiple and Partial Correlation

Methods of measuring Correlation

- Diagrammatic Method
- Mathematical Method

Diagrammatic Method

Correlation Analysis

<u>Positive Correlation</u>

&

<u>Perfect Positive Correlation</u>

Correlation Analysis

<u>Positive Correlation</u>

1. Positive Correlation: If two related variables are such that when one increases (decreases), other also increases (decreases) then we can say that there is positive correlation between the two variables.
2. Positive correlation is said to exist between two variables if they tend to increase or decrease together.

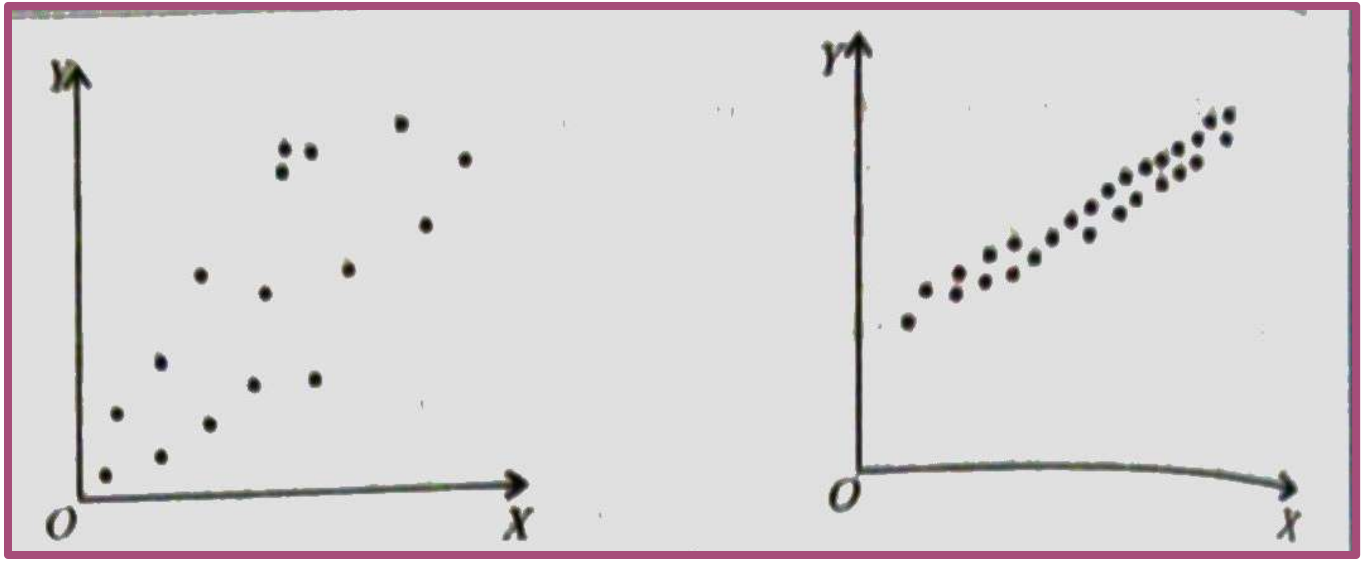
<u>Examples:</u>

- Income and expenditure of a family
- Amount of rainfall (upto a level) and yield of study
- Price and supply of a commodity
- Temperature and sale of ice-cream on different days of a month in summer.

Correlation Analysis

Positive Correlation

1. LOWER LEFT TO UPPER RIGHT, then it is positive correlation .



Correlation Analysis

Perfect Positive Correlation

When the changes in two related variables are exactly proportional and are in the same direction then we say that there is ***perfect positive correlation***.

Variable X:	10	20	30	40	50
Variable Y:	2	4	6	8	10

X increases by 10 units throughout and Y increases by 2 units throughout

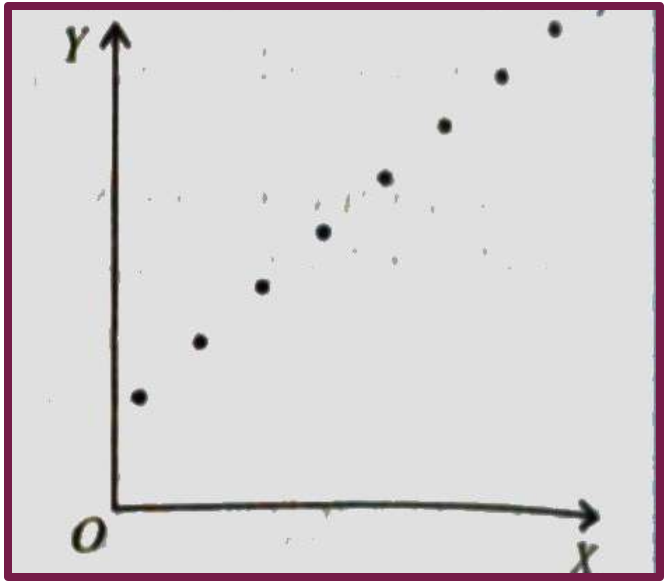
Variable U:	40	35	30	25	20	15	10
Variable V:	14	12	10	8	6	4	2

U decreases by 5 units throughout and V decreases by 2 units throughout

Correlation Analysis

Perfect Positive Correlation

- If all the points in scatter diagram fall on a line with positive slope then it implies that there is ***Perfect Positive Correlation*** between two variables.



Correlation Analysis

Negative Correlation

&

Perfect Negative Correlation

Correlation Analysis

Negative Correlation

1. Negative Correlation: If two related variables are such that when one increases (decreases), other also decreases (increases) then we can say that there is negative correlation between the two variables.
2. Negative correlation is said to exist between two variables if they move in opposite direction.

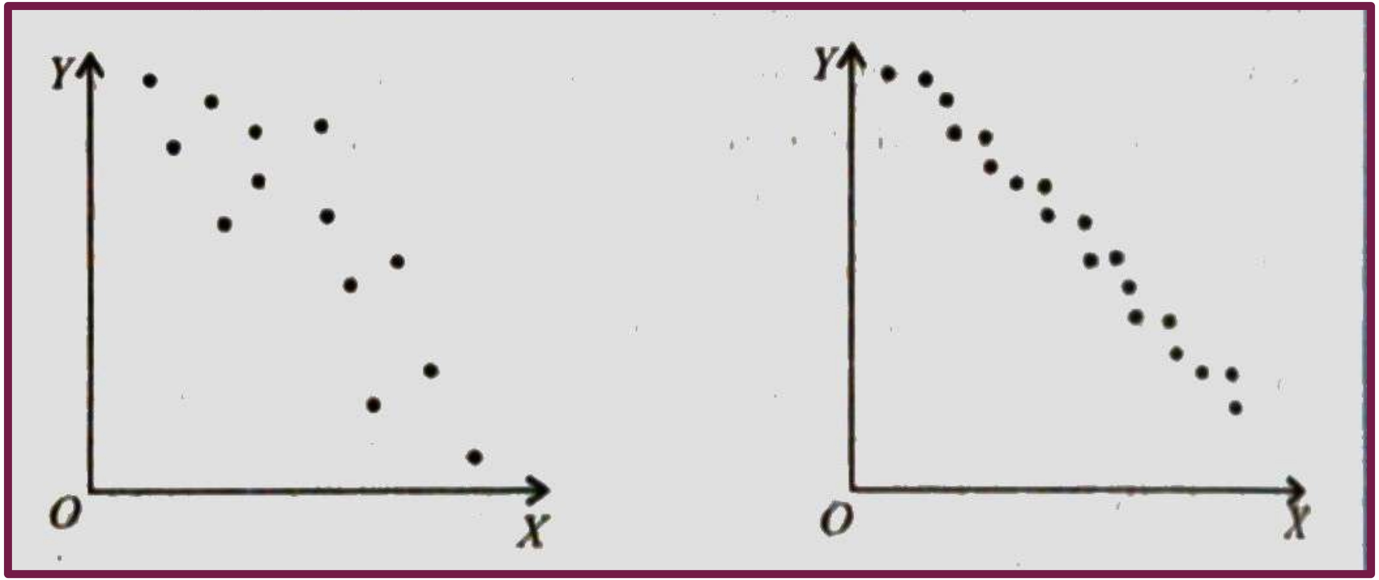
Examples:

- Price of a commodity and demand for it.
- Sale of woolen garments and the day temperature.
- Number of workers and time required to complete work.

Correlation Analysis

<u>Negative Correlation</u>

1. UPPER LEFT TO LOWER RIGHT, then it is negative correlation .



Correlation Analysis

Perfect Negative Correlation

When the changes in two related variables are exactly proportional and but are in the reverse direction then we say that there is ***perfect negative correlation***.

Variable X:	60	50	40	30	20
Variable Y:	2	4	6	8	10

X decreases by 10 units throughout and Y increases by 2 units throughout

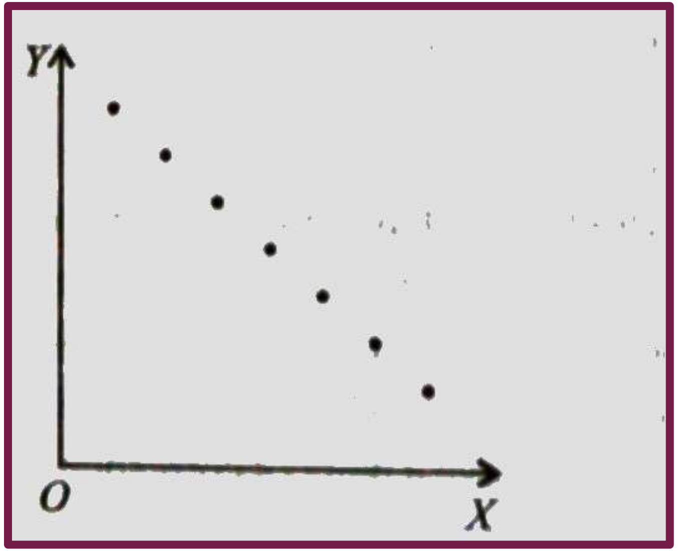
Variable U:	0	1	2	3
Variable V:	2	-3	-8	-13

U increases by 1 units throughout and V decreases by 5 units throughout

Correlation Analysis

Perfect Negative Correlation

- If all the points in scatter diagram fall on a line with negative slope then it implies that there is ***Perfect Negative Correlation*** between two variables.



Correlation Analysis

Zero Correlation

OR

NO Correlation

Correlation Analysis

Zero Correlation

Two variables are said to have zero correlation or no correlation between them if they tend to change with no connection with other.

Variable said to be uncorrelated.

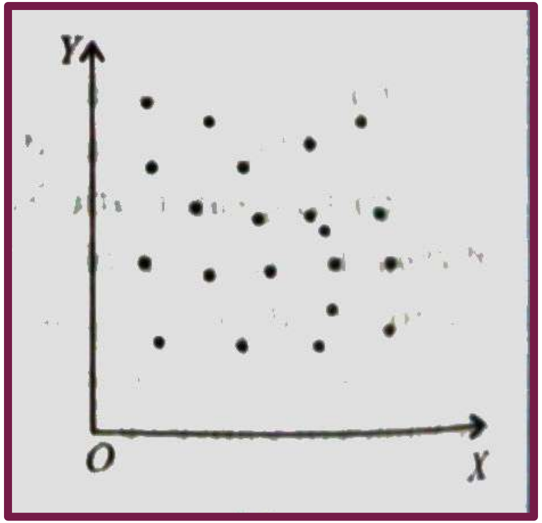
Example:

- Height of students and marks obtained by them.
- Price of rice and demand for salt.

Correlation Analysis

Zero or NO Correlation

- If there appears to be no obvious pattern of the points of the scatter diagram then it indicates that there is either no correlation or very weak correlation between the two variables.
- The points are dispersed over the surface of XY plane.



Correlation Analysis

Linear Correlation

&

Non-Linear Correlation

Correlation Analysis

<u>Linear Correlation</u>

If the points corresponding to the ordered pairs formed by taking the corresponding values of the two correlated variables cluster around a line then the correlation between the two variables is said to be linear.

<u>Non-Linear/curvilinear Correlation</u>

If the points cluster around a curve then the correlation between the two variables is said to be non-linear or curvilinear.

Correlation Analysis

Simple Correlation

&

Multiple Correlation

&

Partial Correlation

Correlation Analysis

Simple Correlation

The study relates to two variables only, e.g. income and saving.

Multiple Correlation

When one variable is related to a number of variables, the study of the relationship between one variable and on one side and all the other variables together on the other side is called multiple correlation.
E.g. the study of relationship between the production of a particular crop on one side and rainfall and use of fertilizer on the other side false under multiple correlation.

Partial Correlation

When one variable (x) is related to a group of variables y,z,t, then if we study the correlation between x and any one of the variables y,z,t... keeping all other variables constant then it will the case of partial correlation.

How to draw scatter diagram based on data?

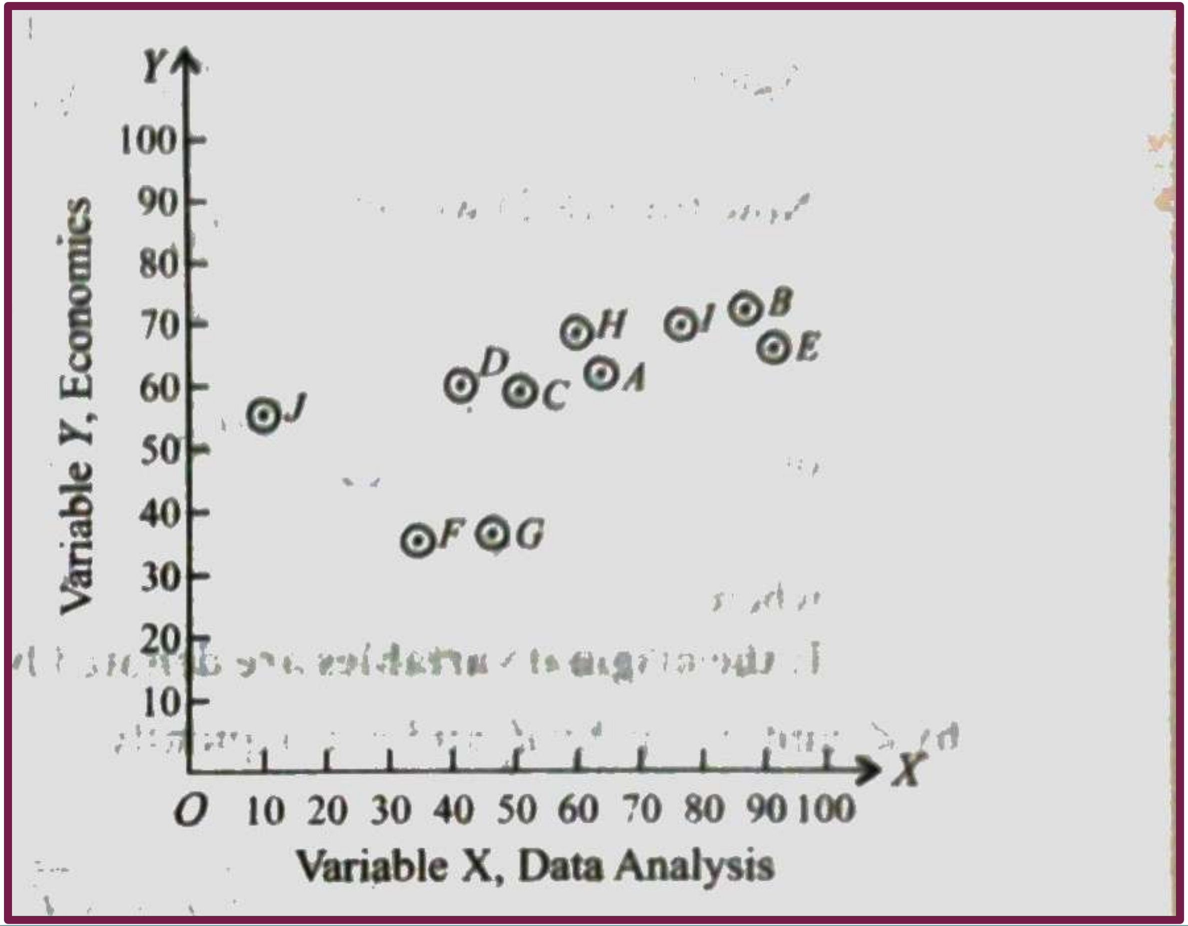
Illustration 1. The percentage examination scores of 10 students in Data Analysis and Economics were as follows. Draw a scatter diagram for the data and comment on the nature of correlation.

Student	:	A	B	C	D	E	F	G	H	I	J
Data Analysis	:	65	90	52	44	95	36	48	63	80	15
Economics	:	62	71	58	58	64	40	42	66	67	55

Solution : A scatter diagram will give a preliminary indication of whether linear correlation exists. We plot the ordered pairs (65,62), (90, 71), etc. as shown in fig. 8.1.2.

It appears from the scatter diagram that some form of positive correlation exists between the scores of the students in Data Analysis and Economics.

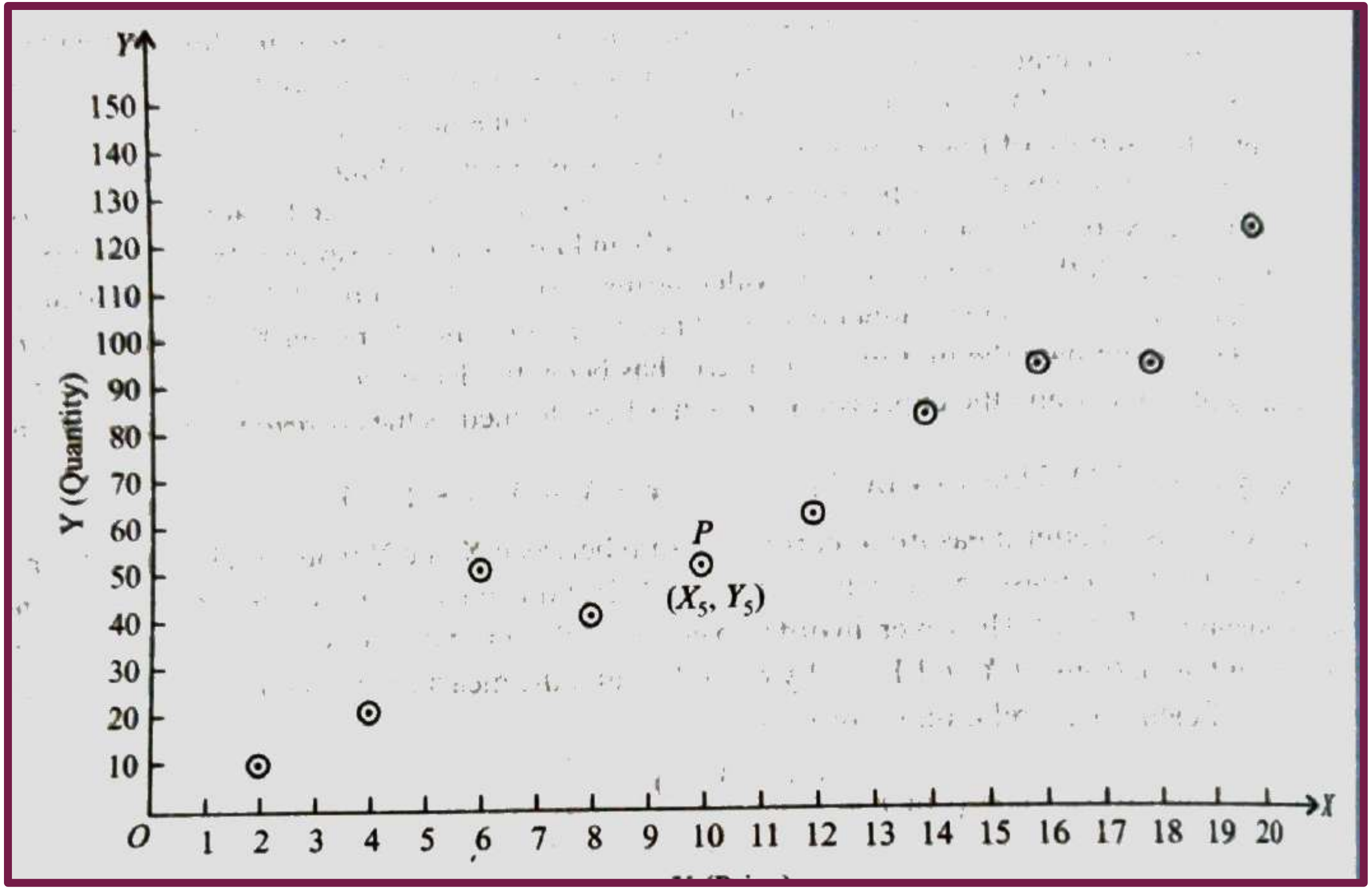
How to draw scatter diagram based on data?



How to draw scatter diagram based on data?

Time period :	1	2	3	4	5	6	7	8	9	10
Quantity supplied (Y) :	10	20	50	40	50	60	80	90	90	120
(in kgs)	(Y_1)	(Y_2)	(Y_3)	(Y_4)	(Y_5)	(Y_6)	(Y_7)	(Y_8)	(Y_9)	(Y_{10})
Price (X) :	2	4	6	8	10	12	14	16	18	20
(in rupees)	(X_1)	(X_2)	(X_3)	(X_4)	(X_5)	(X_6)	(X_7)	(X_8)	(X_9)	(X_{10})

How to draw scatter diagram based on data?



Mathematical Method

Karl Pearson's Method

Karl Pearson's Method-

Fig. 8.2

method of measuring correlation between two variables is associated with the concept of covariance between two variables. If two variables X and Y are such that when X takes the values $X_1, X_2, \ldots, X_n$, the corresponding values of Y are $Y_1, Y_2, \ldots, Y_n$ then the covariance between X and Y is denoted by Cov (X, Y) and is defined as below :

$$\text{Cov}(X,Y) = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{n} \quad \text{where } \bar{X} = \frac{\Sigma X}{n}, \bar{Y} = \frac{\Sigma Y}{n},$$

Karl Pearson's Method- To covariance of correlation

Covariance is a measure of association between two variables. Karl Pearson's measure of correlation between two variables is called the *coefficient of correlation* or *correlation coefficient*. It is also known as *product moment coefficient*. Karl Pearson's coefficient of correlation between two variables X and Y is denoted by the symbol r_{XY} (r having as subscripts the variables whose correlation it measures) or by simply r if the correlation is based on a sample of values of the population of X and Y. If, however, the correlation between X and Y is based on all the values of the population of X and Y then the correlation coefficient is represented by the symbol ρ_{XY} or ρ where ρ is a Greek letter pronounced as 'rho'. Here, of course, unless explicitly stated that the values of the variables whose correlation is to be studied are population values we shall use the symbol r_{XY} which is a sample estimate for ρ_{XY}. r_{XY} is defined as below :

$$r_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X).\,Var(Y)}}$$

or,

$$r_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \,.\, \sigma_Y} \qquad \text{...(8.1)}$$

$$\left[\because \sigma_X = \sqrt{Var(X)},\ \sigma_Y = \sqrt{Var(Y)}\right]$$

Putting the expression for Cov (X, Y) we can express (8.1) as below :

$$r_{XY} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{n\,\sigma_X\,\sigma_Y}$$

Karl Pearson's Method

Coefficient of correlation

$$r_{XY} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\,\Sigma(Y-\bar{Y})^2}}$$

OR

Now $$r_{XY} = \frac{\Sigma X'Y' - \frac{(\Sigma X')(\Sigma Y')}{n}}{\sqrt{\left\{\Sigma X'^2 - \frac{(\Sigma X')^2}{n}\right\}\left\{\Sigma Y'^2 - \frac{(\Sigma Y')^2}{n}\right\}}}$$, where $x' = x - A$ and $y' = y - B$

[Formula (8.4D), which is same as formula (8.4B)]

Using step deviation from A. M.

$$r_{XY} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$x = X - \bar{X}$, $y = Y - \bar{Y}$.

Karl Pearson's Method

2. Correlation coefficient or coefficient of correlation means Karl Pearson's coefficient of correlation.

Illustration 2. Find the coefficient of correlation between X and Y by taking deviations from actual means :

X :	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Y :	**4**	**5**	**6**	**8**	**9**	**7**	**10**

Solution:

Table 8.1: Calculation of Coefficient of Correlation

X	Y	$x = X - \bar{X} - 5$	$y = Y - \bar{Y} - 7$	x^2	y^2	xy
2	4	−3	−3	9	9	9
3	5	−2	−2	4	4	4
4	6	−1	−1	1	1	1
5	8	0	1	0	1	0
6	9	1	2	1	4	2
7	7	2	0	4	0	0
8	10	3	3	9	9	9
$\Sigma X = 35$	$\Sigma Y = 49$	$\Sigma x = 0$	$\Sigma y = 0$	$\Sigma x^2 = 28$	$\Sigma y^2 = 28$	$\Sigma xy = 25$

Here

$$\bar{X} = \frac{\Sigma X}{n} = \frac{35}{7} = 5;\ \bar{Y} = \frac{\Sigma Y}{n} = \frac{49}{7} = 7$$

$$r_{XY} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$$= \frac{25}{\sqrt{28 \times 28}} = \frac{25}{28} = \mathbf{0.89}$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{35}{7} = 5;\ \bar{Y} = \frac{\Sigma Y}{n} = \frac{49}{7} = 7$$

$$r_{XY} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$$= \frac{25}{\sqrt{28 \times 28}} = \frac{25}{28} = \mathbf{0.89}$$

Karl Pearson’s Method

Illustration 3. From the following data compute coefficient of correlation between X and Y:

	X- series	*Y- series*
Arithmetic mean	25	18
Sums of Square of deviations from arithmetic mean	136	138

Summation of products of deviations of X and Y- series from their respective means = 122; no. of pairs of values = 15.

Solution: The appropriate form of Karl Pearson’s coefficient of correlation formula in this case will be the following :

$$r_{XY} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2 . \Sigma(Y - \bar{Y})^2}}$$

[This is the same formula as applied in Illustration 2 above.]

$$= \frac{122}{\sqrt{136 \times 138}} = \frac{122}{\sqrt{18768}}$$

$$= \frac{122}{137} = \mathbf{0.89.}$$

Note : Although the values of $\bar{X}, \bar{Y}$ and n are given, we do not need these values to find r_{XY}

Karl Pearson's Method

Illustration 4. The following are the marks obtained by twelve students in Accountancy and Business Statistics :

Marks in Accountancy :	50	54	56	59	60	62	61	65	67	71	71	74
Marks in B. Statistics :	22	25	34	28	26	30	32	30	28	34	36	40

Obtain Karl Pearson's coefficient of correlation.

Solution: Let us denote marks in Accountancy by X and marks in Business Statistics by Y. Here we shall use the assumed mean method. [*Since in assumed mean method* $\bar{X}$ *and* $\bar{Y}$ *are not needed hence there is no need of finding* ΣX *and* ΣY.] Let the asumed mean A of X be 62 and let the assumed mean B of Y be 30.

Table 8.2: Calculation of Coefficient of Correlation

X	Y	$X' = X - 62$	$Y' = Y - 30$	X'^2	Y'^2	$X'Y'$
50	22	−12	−8	144	64	96
54	25	−8	−5	64	25	40
56	34	−6	4	36	16	−24
59	28	−3	−2	9	4	6
60	26	−2	−4	4	16	8
62	30	0	0	0	0	0
61	32	−1	2	1	4	−2
65	30	3	0	9	0	0
67	28	5	−2	25	4	−10
71	34	9	4	81	16	36
71	36	9	6	81	36	54
74	40	12	10	144	100	120
		$\Sigma X' = 6$	$\Sigma Y' = 5$	$\Sigma X'^2 = 598$	$\Sigma Y'^2 = 285$	$\Sigma X'Y' = 324$

Now

$$r_{XY} = \frac{\Sigma X'Y' - \frac{(\Sigma X')(\Sigma Y')}{n}}{\sqrt{\left\{\Sigma X'^2 - \frac{(\Sigma X')^2}{n}\right\}\left\{\Sigma Y'^2 - \frac{(\Sigma Y')^2}{n}\right\}}}, \text{ where } x' = x - A \text{ and } y' = y - B$$

[Formula (8.4D), which is same as formula (8.4B)

$$= \frac{324 - \frac{6 \times 5}{12}}{\sqrt{\left\{598 - \frac{6^2}{12}\right\}\left\{285 - \frac{5^2}{12}\right\}}} = \frac{324 - 2.5}{\sqrt{595 \times 282.92}} = \frac{321.5}{410.27} = \mathbf{0.78}.$$

Karl Pearson's Method

Example 1. Find the coefficient of correlation from the following data.

X :	1	2	3	4	5	6	7
Y :	6	8	11	9	12	10	14

Solution :

Table 8.8: Calculation of Coefficient of Correlation

X	Y	$x = X - \bar{X}$ $\bar{X} = 4$	$y = Y - \bar{Y}$ $\bar{Y} = 10$	x^2	y^2	xy
1	6	−3	−4	9	16	12
2	8	−2	−2	4	4	4
3	11	−1	1	1	1	−1
4	9	0	−1	0	1	0
5	12	1	2	1	4	2
6	10	2	0	4	0	0
7	14	3	4	9	16	12
$\Sigma X = 28$	$\Sigma X = 70$			$\Sigma x^2 = 28$	$\Sigma y^2 = 42$	$\Sigma xy = 29$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{28}{7} = 4 \text{ and } \bar{Y} = \frac{\Sigma Y}{n} = \frac{70}{7} = 10$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2} \times \sqrt{\Sigma y^2}} = \frac{29}{\sqrt{28} \times \sqrt{42}} = \frac{29}{5.9 \times 6.48}$$

$$= \frac{29}{34.28} = 0.85 \text{ (approx)}$$

$$r = \frac{29}{\sqrt{28} \times \sqrt{42}} = \frac{29}{5.29 \times 6.48} = \frac{29}{34.28} = 0.85$$

Regression Analysis

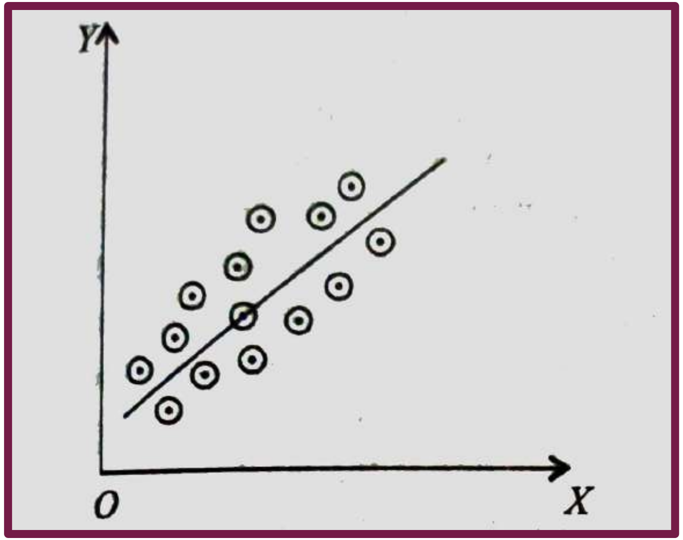
- <u>Introdiction</u>
- Regression is the measure of the average relationship between two or more variables in terms of the original units of the data.
- The variable corresponding to whose value, the average value of the other variable is estimated is called the independent variable.
- Independent is also called the *explanatary* or *exogenous* variable.
- The variable whose average value is estimated is called the dependent variable.
- Dependent variable is also called *explained* or *endogenous* variable.
- The average value is also termed as the most probable value or the most likely value.

Linear and Non-Linear Regression

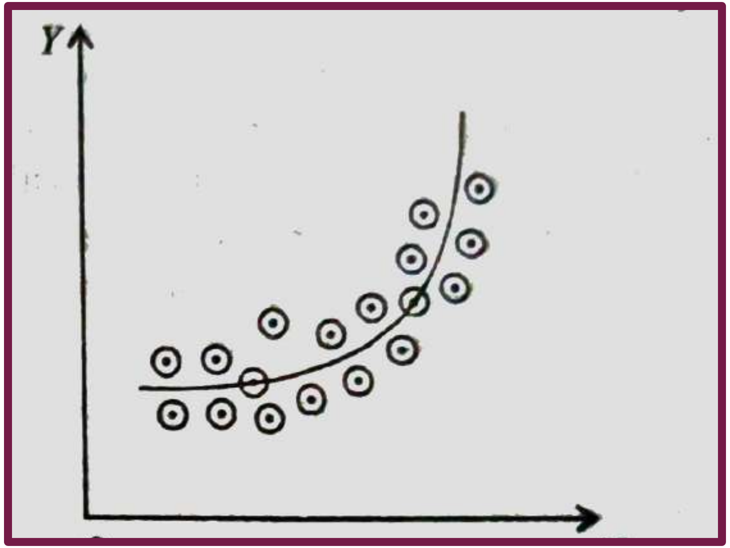
- <u>Non-Linear Regression</u>
    - If the scatter diagram clusters around a curve then we say that there is non-linear and curvilinear regression between the two variables.

- <u>Linear Regression</u>
    - If the scatter diagram of two correlated variables clusters around a line then we can say that there is linear regression between two variables.

Linear and Non-Linear Regression



Linear Regression



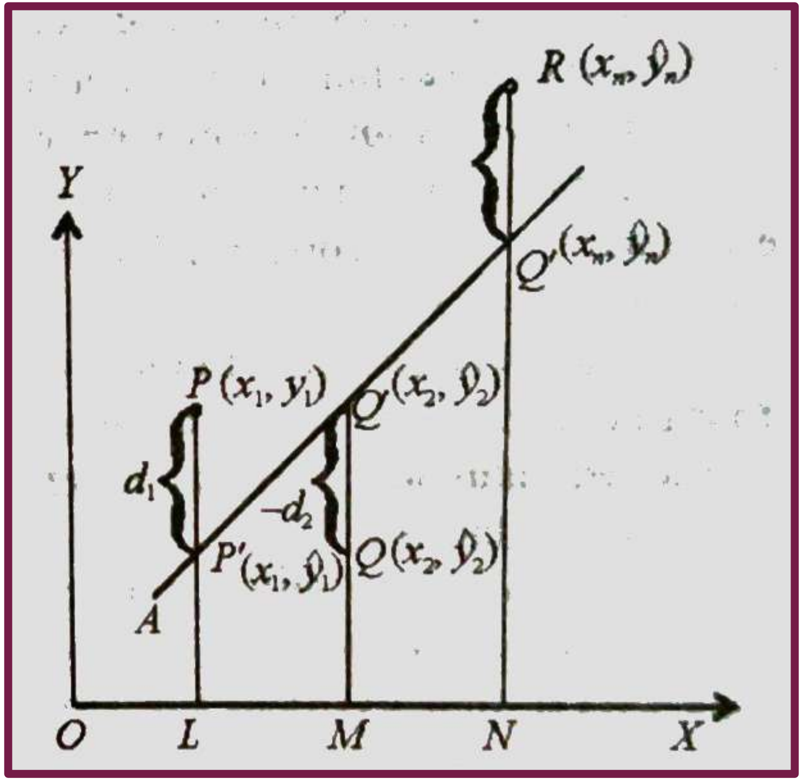
Non-Linear Regression

Method of Least Square

- Let x and y two correlated variables such that corresponding to the values $x_1, x_2, \ldots\ldots x_n$ of x value of y are $y_1, y_2, \ldots\ldots y_n$ respectively.
- Now if we draw the line AB through the points $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ then the estimated values of y corresponding to the values $x_1, x_2, \ldots\ldots x_n$ of x are y_1^, y_2^....y_n^ respectively.
- Thus estimated deviation...
- $d_1 = (y_1 - y_1$^ $=$ LP - LP'), $d_2 = (y_2 - y_2$^ $=$ MQ - MQ'),.....$d_n = (y_n - y_n$^ $=$ NR - NR').

Method of Least Square

- Thus estimated deviation…
- $d_1 = (y_1 - y_1\hat{} = LP - LP')$, $d_2 = (y_2 - y_2\hat{} = MQ - MQ')$,…..$d_n = (y_n - y_n\hat{} = NR - NR')$.



Method of Least Square

- Thus estimated deviation...
- $d_1 = (y_1 - y_1\hat{} = LP - LP')$, $d_2 = (y_2 - y_2\hat{} = MQ - MQ')$,.....$d_n = (y_n - y_n\hat{} = NR - NR')$.
- Deviation will be positive if estimated value is less than the actual value.
- Deviation will be negative if estimated value is greater than the actual value.
- Deviation will be zero if estimated value is equals to the actual value.
- The sums of squares of these deviations = d_1^2, d_2^2.....d_n^2.
- The position of AB line will change according to the new deviations.

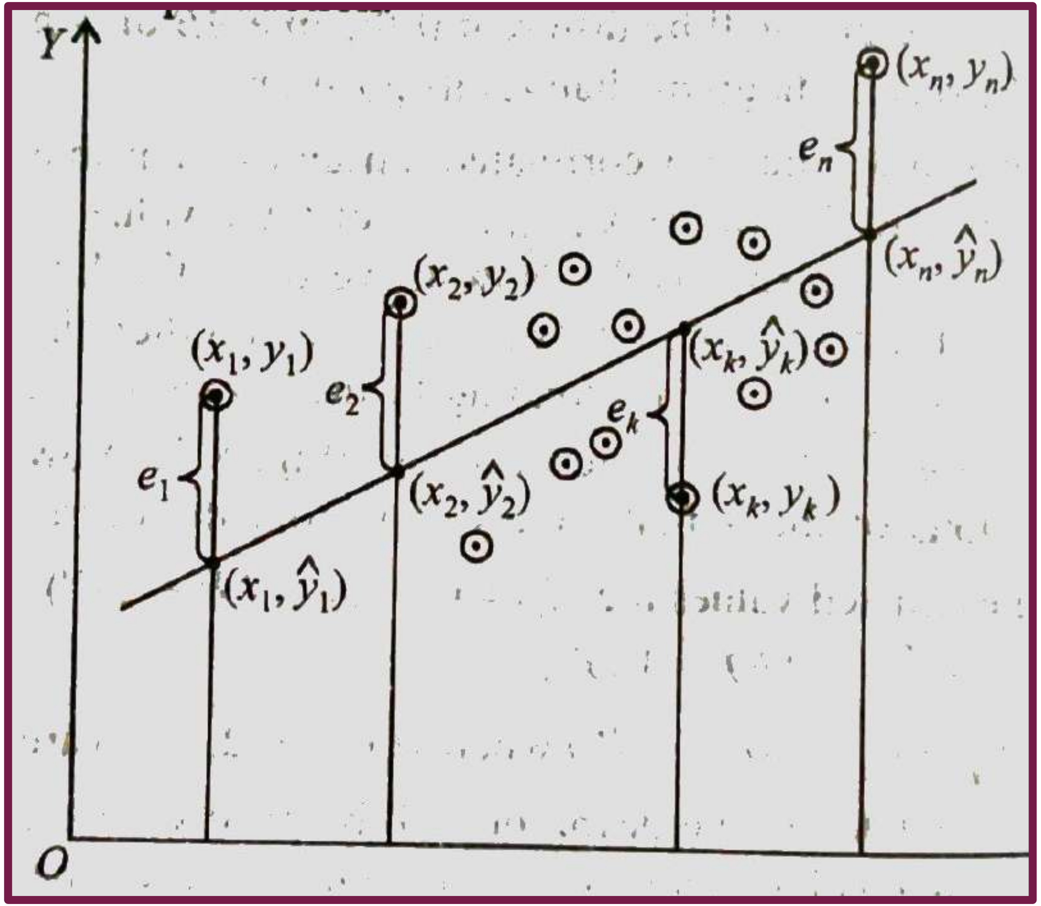
Method of Least Square

- The sums of the deviations about this line is the least’ means that the sum of squares of the error about this line is least.
- Hence, this line is called the least squares line.
- This method of determined the line of best fit or the least squares line is known as the **Method of Least Squares.**

Regression Lines

- The regression line of y on x is that line from which we get the best estimated value of y corresponding to a given value of x.
- The regression line of x and y is that line from which we get the best estimated value of x corresponding to a given value of y.
- If there exists either perfect positive correlation or perfect negative correlation between the two variables x and y then two lines of regression coincide.
- The mathematical expression of a regression line is called a regression equation.

Regression Lines



The Regression Equation Y on X

$$\underline{\mathbf{Y = a + bx}}$$

The Regression Equation X on Y

$$\underline{\mathbf{X = a + by}}$$

Equation for "b"

When value of two variables are small

$$b = \frac{\Sigma xy - n\bar{x}\,\bar{y}}{\Sigma x^2 - n\bar{x}^2}$$

Example:

Example 1. *From the following bi-variate distribution find the regression equation of y (price) on x (demand) :*

Demand (x) :	10	8	5	4	2	1
Price (y) :	4	6	5	7	8	9

Predict price when demand is 7 units.

Solution : To predict price (y) corresponding to a given value of demand we need the regression equation of y and x. Let the regression equation of y on x be $y = a + bx$...(1)

Table 8.24: Estimation of regression equation

Demand x	*Price* y	x^2	y^2	xy
10	4	100	16	40
8	6	64	36	48
5	5	25	25	25
4	7	16	49	28
2	8	4	64	16
1	9	1	81	9
$\Sigma x = 30$	$\Sigma y = 39$	$\Sigma x^2 = 210$	$\Sigma y^2 = 271$	$\Sigma xy = 166$

Here $n = 6$, $\bar{x} = \frac{\Sigma x}{n} = \frac{30}{6} = 5$; $\bar{y} = \frac{\Sigma y}{n} = \frac{39}{6} = 6.5$

Applying formula 8.3.6,

$$b = \frac{\Sigma xy - n\bar{x}\,\bar{y}}{\Sigma x^2 - n\bar{x}^2} = \frac{166 - 6 \times 5 \times 6.5}{210 - 6 \times 5 \times 5} = -\frac{29}{60} = -0.48$$

Again, $y = a + bx \Rightarrow \bar{y} = a + b\bar{x} \Rightarrow a = \bar{y} - b\bar{x}$.

$$a = \bar{y} - b\bar{x} = 6.5 - (-0.48) \times 5 = 8.9$$

Now putting the values of a and b in (1) we get the regression equation of y on x as

$$\hat{y} = \mathbf{8.92 - 0.48}\,x$$

$\hat{y}$ (y_e or y_c) denotes the estimated value of y.

When demand $x = 7$ units, the predicted price y will be :

$$y = 8.92 - 0.48 \times 7 = 8.92 - 3.36 = \mathbf{5.54\ units}$$

Assumed mean method for Regression

When value of two variables are large, then assumed mean method is applied.

The regression equation of y on x is :

$$y - \bar{y} = b_{yx}\,(x - \bar{x})$$

Where

$$b_{yx} = \frac{\Sigma x'y' - \dfrac{(\Sigma x')\,(\Sigma y')}{n}}{\Sigma x'^2 \;\dfrac{(\Sigma x')^2}{n}}$$

Example

Alternative method: Assumed mean method :

To predict priice (y) corresponding to a given value of demand (x) we need the regression equation of y on x.

Table 8.25: Estimation of regression equation

x	y	$x' = x - 5$	$y' = y - 7$	x'^2	y'^2	$x'y'$
10	4	5	–3	25	9	–15
8	6	3	–1	9	1	–3
5	5	0	–2	0	4	0
4	7	–1	0	1	0	0
2	8	–3	1	9	1	–3
1	9	–4	2	16	4	–8
$\Sigma x = 30$	$\Sigma y = 39$	$\Sigma x' = 0$	$\Sigma y' = -3$	$\Sigma x'^2 = 60$	$\Sigma y'^2 = 19$	$\Sigma x'y' = -29$

Now $\bar{x} = \frac{\Sigma x}{n} = \frac{30}{6} = 5$; $\bar{y} = \frac{\Sigma y}{x} = \frac{39}{6} = 6.5$

The regression equation of y on x is :

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

Where $$b_{yx} = \frac{\Sigma x'y' - \frac{(\Sigma x')(\Sigma y')}{n}}{\Sigma x'^2 \frac{(\Sigma x')^2}{n}}$$ [Assu

$$= \frac{-29 - \frac{0 \times (-3)}{6}}{60 - \frac{0^2}{6}} = -0.48$$

Now (1) $\Rightarrow y - 6.5 = -0.48(x - 5)$

$\Rightarrow y = -0.48x + 2.4 + 6.5$

$\Rightarrow \mathbf{y = -0.48x + 8.9}$

When demand $x = 7$ units, the predicted price y will be :

$y = -0.48 \times 7 + 8.9 = -3.36 + 8.9 =$ **5.54** units

An Important Note : The first method should be applied ...

When mean is positive

Note : Whenever both $\bar{x}$ and $\bar{y}$ are integers then we can easily estimate b_{yx} and b_{xy} by using the following formulae:

$$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{\Sigma d_x d_y}{\Sigma d_x^2}, \quad b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} = \frac{\Sigma d_x d_y}{\Sigma d_y^2}$$

where $d_x = x - \bar{x}, d_y = y - \bar{y}$.

Example

Example 2. Obtain the lines of regression for the follwoing data :

x :	120	90	80	150
y :	40	36	40	44

Solution: The following table has been formed.

Table 8.26: Estimation of Regression Lines

x	$d_x = x - \bar{x}$ $= x - 110$	d_x^2	y	$d_y = y - \bar{y}$ $= y - 40$	d_y^2	$d_x d_y$
120	10	100	40	0	0	0
90	−20	400	36	−4	16	80
80	−30	900	40	0	0	0
150	40	1600	44	4	16	160
$\Sigma x = 400$ 440		$\Sigma d_x^2 = 3000$	$\Sigma y = 160$		$\Sigma d_y^2 = 32$	$\Sigma d_x d_y = 240$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{440}{4} = 110, \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{160}{4} = 40$$

Now, $$b_{yx} = \frac{\Sigma d_x d_y}{\Sigma d_x^2} = \frac{240}{3000} = 0.08$$

$$b_{xy} = \frac{\Sigma d_x d_y}{\Sigma d_y^2} = \frac{240}{32} = 7.5$$

$$(d_x = x - \bar{x},\ d_y = y - \bar{y})$$

Now, the regression equation of y on x is :

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$\Rightarrow y - 40 = 0.08(x - 110)$$

$$\Rightarrow y = 0.08x - 8.8 + 40$$

$$\Rightarrow \mathbf{y = 0.08x + 31.2}$$

Again, the regression equation of x on y is :

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$\Rightarrow x - 110 = 7.5(y - 40)$$

$$\Rightarrow x = 7.5y - 300 + 110$$

$$\Rightarrow \mathbf{x = 7.5y - 190}$$

Example

3.81 lakh acres.

Example 8. ***In the estimation of regression equations of two variables X and Y the following results were obtained.***

$\bar{X} = 90$, $\bar{Y} = 70$, $n = 10$, $\Sigma x^2 = 6360$, $\Sigma y^2 = 2860$, $\Sigma xy = 3900$ ***where*** $x = X - \bar{X}$, $y = Y - \bar{Y}$ ***i.e., x and y are the deviations of X and Y from their respective means. Obtain the two regression equations.***

Solution : The regression coefficient of Y or X is:

$$b_{YX} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(X-X)^2} = \frac{\Sigma xy}{\Sigma x^2} = \frac{3900}{6360} = 0.6132$$

The regression coefficient of X on Y is:

$$b_{XY} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(Y-Y)^2} = \frac{\Sigma xy}{\Sigma y^2} = \frac{3900}{2860} = 1.3636$$

The regression equation on Y on X is:

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

$$\Rightarrow Y - 70 = 0.6132(X - 90)$$

$$\Rightarrow Y = 0.6132\,X - 55.188 + 70$$

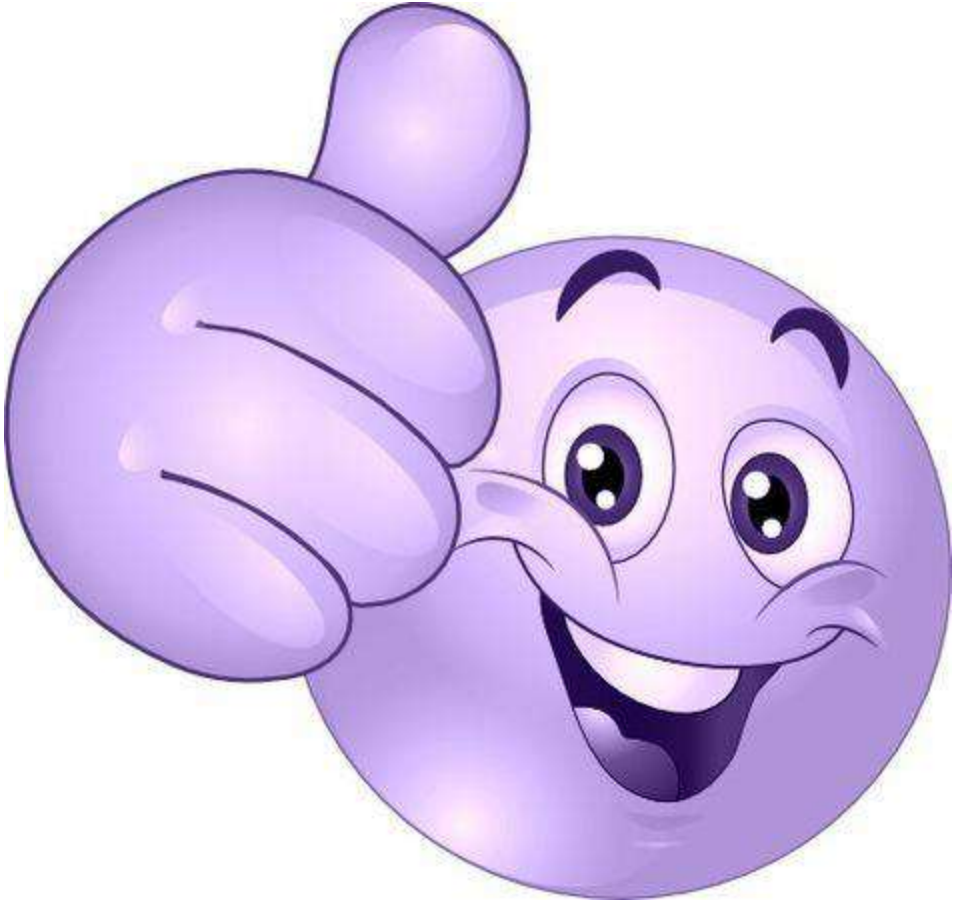
$$\Rightarrow \mathbf{Y = 0.6132\,X + 14.812}$$

The regression equation of X on Y is:

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

$$\Rightarrow X = 1.3636\,Y - 95.452 + 90 \quad (\because\ \bar{X} = 90,\ \bar{Y} = 70)$$

$$\Rightarrow \mathbf{X = 1.3636\,Y - 5.452}$$



THANK

YOU…