8. Correlation Analysis

Correlation analysis is *the study of relationship between two or more variables.* *It is the* ***co-variation*** *of two variables.*

The measure of correlation is called ***correlation coefficient.***

The weight of a man depends on height. When height increases weight also increases. This is a correlation between weight and height.

The travelling time depends on speed. When the speed is low, the travelling time will be high. When the speed is high, the travelling time will be low. The relationship between time and speed is a correlation.

The correlation is mainly two types, namely ***positive correlation*** and ***negative correlation.***

If the two variables move together in the same direction, the correlation is called ***positive correlation.*** It is also called ***direct correlation.*** In positive correlation, there is an increase in the value of one variable is accompanied by an increase in the value of the other variable; or a decrease in the value of one variable is accompanied by a decrease in the other variable.

Eg. *Height* and *weight - Increase*

In ***negative correlation,*** *the two variables tend to move in opposite directions.* It is also called ***inverse correlation.*** Here, an increase in the value of one variable is accompanied by a decrease in the value of the other variable; or a decrease in the value of one variable is accompanied by an increase in the value of the other variable.

Eg. *Time* and *speed.*

The correlation is measured by a method called ***Pearson's coefficient of correlation*** devised by ***Karl Pearson.*** Pearson's coefficient of correlation explains the degree of relationship between two variables.

Pearson's coefficient of correlation is denoted by the symbol **γ**.

The formula for the calculation of Pearson's coefficient of correlation.

$$\boxed{\gamma = \frac{\Sigma\, dx\, dy}{\sqrt{\Sigma dx^2 \times \Sigma dy^2}}}$$

$dx = x - \bar{x}$ = deviation from $\bar{x}$

$dy = y - \bar{y}$ = deviation from $\bar{y}$

$\Sigma dxdy$ = Multiply the deviations of x and y and get the total

Σdx^2 = The sum of the square of the deviations of x

Σdy^2 = The sum of square of the deviations of y

The value of Pearson's coefficient of correlation always lies between +1 and -1.

When γ is +1, then there is perfect positive correlation between the two variables.

When γ is -1, then there is perfect negative correlation between the two variables.

When γ is 0, then there is no relationship between the two variables.

Theoretically, the correlation coefficient values lie inbetween +1 and -1. But actually the values lie between +0.8 and -0.5.

According to ***Prof. Boulding*** "*Wherever some definite connection exists between 2 or more groups, classes or series of data, there is said to be correlation*".

A very simple definition is given by ***A M. Tuttle.*** "*An analysis of the co-variation of two or more variables* is usually called *correlation.*

Uses

1. Correlation is very useful to study the relationship between variables.

2. Some variables show some kind of relationship; correlation analysis helps in measuring the degree of relationship between the variables.

3. The relationship between variables can be verified and tested for significance, with the help of the correlation analysis. The effect of correlation is to reduce the range of uncertainty of our prediction.

4. The co-efficient of correlation is a relative measure and we can compare the relationship between variables which are expressed in different units.

5. Sampling error can also be calculated.

6. Correlation is the basis for the concept of regression and ratio of variation.

Correlation and Causation

Correlation analysis deals with the association or co-variation between two or more variables and helps to determine the degree of relationship between two or more variables. But correlation does not indicate a cause and effect relationship between two variables. It explains only co-variation. The high degree of correlation between two variables may exist due to any one or a combination of the following reasons:

1. It is due to pure chance. For example, the relationship between *cars production* and the *children born* in a country.

Table 8.1: Degree of correlation.

Degree of Correlation	Positive	Negative
Perfect correlation	+1	-1
Very high degree of correlation	+0.9 or more	-0.9 or more
Fairly high degree of correlation	from + 0.75 to +0.9	from -0.75 to - 0.9
Moderate degree of correlation	from + 0.5 to +0.75	from - 0.5 to - 0.75
Low degree of correlation	from + 0.25 to +0.5	from - 0.25 to - 0.5
Very low degree of correlation	less than 0.25	less than - 0.25
No correlation	0	0

Types of Correlation

Correlation is classified into many types, but the important ones are:

1. *Positive and negative*
2. *Simple and multiple*
3. *Partial and total*
4. *Linear and non-linear*

1. Positive and Negative Correlation

Positive and negative correlation depend upon the direction of change of the variables. If two variables tend to move together in the same direction i.e., an increase in the value of one variable is accompanied by an increase in the value of the other variable; or a decrease in the value of one variable is accompanied by a decrease in the value of the other variable; then the correlation is called ***positive correlation*** or ***direct correlation.*** for example Height and weight, rainfall and yield of crops, age and weight of the baby are examples of positive correlation.

If two variables tend to move together in opposite directions i.e., an increase in the value of one variable is accompanied by a decrease in the value of the other variable; or a decrease in the value of one variable is accompanied by an increase in the value of the other variable; then the correlation is called ***negative correlation*** or ***inverse correlation.*** for example *Yield of crops* and *price* are the examples of negative correlation.

2. Simple and Multiple Correlation

When we study two variables, the relationship is described as ***simple correlation;*** Eg. *Age and weight of the baby; height and weight of the students,* etc. But in a ***multiple correlation*** we study more than two variables simultaneously. Eg. *Diet, age and weight of the baby.*

3. Partial and Total Correlation

The study of two variables excluding some other variables is called ***partial correlation.*** For example, we study *age and weight of the baby eliminating the diet.*

In total correlation, all the facts are taken into account.

4. Linear and Non-linear Correlation

If the ratio of change between two variables is uniform, then there will be linear correlation between them. Consider the following:

X	5	10	15	20	25
Y	3	6	9	12	15

The ratio of change between the variables is the same

[X : 5 + 5 = 10; 10 + 5 = 15; 15 + 5 = 20; 20 + 5 = 25; Y : 3 + 3 = 6; 6 + 3 = 9; 9 + 3 = 12; 12 + 3 = 15]

If we plot these on a graph, we get a straight line like this

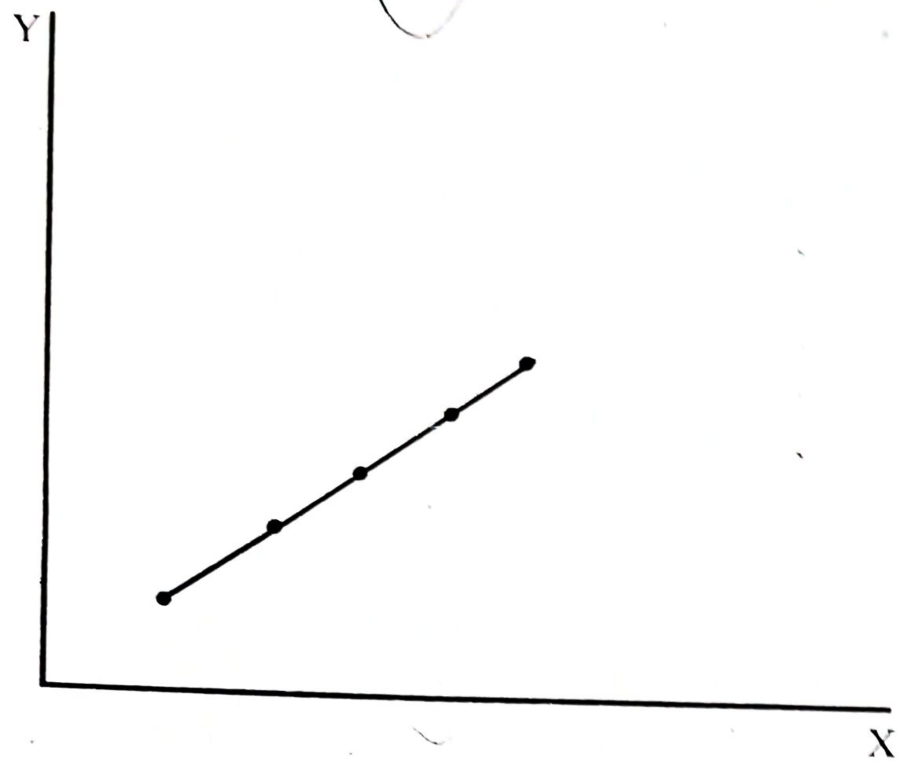

Fig.8.1: Linear correlation.

In a non - linear (curvilinear) correlation, the amount of change in one variable does not bear a constant ratio to the amount of change in the other variables. Consider the following:

X	5	10	15	20	25
Y	3	5	6	9	10

The ratio of change between the variable - X is the same [X : 5 + 5 = 10; 10 + 5 = 15; 15 + 5 = 20; 20 + 5 = 25]

But the ratio of change between the variable-Y is not the same [Y : 3 + 2 = 5; 5 + 1 = 6; 6+3=9; 9 + 1 = 10]

If we plot these on a graph, we get a curve like this

UNIT - IV

CORRELATION

Karl - Pearson's correlation coefficient,

$$\gamma = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

$$cov(x,y) = \frac{\Sigma xy}{n} - (\bar{x})(\bar{y})$$

where,

$$\bar{x} = \frac{\Sigma x}{n}$$

$$\bar{y} = \frac{\Sigma y}{n}$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2}$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - (\bar{y})^2}$$

1. Find correlation coefficient from the following data:-

x	10	12	18	24	23	27
y	13	18	12	25	30	10

soln:-

Karl - Pearson's correlation coefficient,

$$\gamma = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

x	y	x^2	y^2	xy
10	13	100	169	130
12	18	144	324	216
18	12	324	144	216
24	25	576	625	600
23	30	529	900	690
27	10	729	100	270
114	108	2402	2262	2122

$\Sigma x = 114$

$\Sigma y = 108$ $\quad -0.8698$

$\Sigma x^2 = 2402$

$\Sigma y^2 = 2262$

$\Sigma xy = 2122$

$n = 6$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{114}{6} = 19$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{108}{6} = 18$$

$$\text{cov}(x, y) = \frac{\Sigma xy}{n} - (\bar{x})(\bar{y})$$

$$= \frac{2122}{6} - (19)(18)$$

$$= 353.6667 - 342$$

$$\text{cov}(x, y) = 11.6667.$$

$$\sigma_x = \sqrt{\frac{2402}{6} - (19)^2}$$

$$= \sqrt{400.3333 - 361}$$

$$= \sqrt{39.3333}$$

$$\sigma_x = 6.2716$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - (\bar{y})^2}$$

$$= \sqrt{\frac{2262}{6} - (18)^2}$$

$$= \sqrt{377 - 324}$$

$$= \sqrt{53}$$

$$\sigma_y = 7.2801$$

$$r = \frac{\text{cov}(x,y)}{\sigma_x \ \sigma_y}$$

$$= \frac{11.6667}{(6.2716)(7.2801)}$$

$$= \frac{11.6667}{45.6579}$$

$$r = 0.2555$$

2. Find pearsons's co-efficient of correlation.

x	10	14	15	28	35	48
y	74	61	50	54	43	36

Soln:-

Karl - pearsons's correlation coefficient,

$$r = \frac{\text{cov}(x,y)}{\sigma_x \ \sigma_y}$$

x	y	x^2	y^2	xy
10	74	100	5476	740
14	61	196	3721	854
15	50	225	2500	750
28	54	784	2916	1512
35	43	1225	1849	1505
48	36	2304	1296	1728
150	318	4834	17758	7089

$\Sigma x = 150$

$\Sigma y = 318$

$\Sigma x^2 = 4834$

$\Sigma y^2 = 17758$

$\Sigma xy = 7089$

$n = 6$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{150}{6} = 25$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{318}{6} = 53$$

$$\mathrm{cov}(x, y) = \frac{\Sigma xy}{n} - (\bar{x})(\bar{y})$$

$$= \frac{7089}{6} - (25)(53)$$

$$= 1181.5 - 1325$$

$$= -143.5$$

$$\sigma_x = \sqrt{\frac{4834}{6} - (25)^2}$$

$$= \sqrt{805.6667 - 625}$$

$$= \sqrt{180.6667}$$

$$\sigma_x = 13.4412.$$

$$\sigma_y = \sqrt{\frac{17758}{6} - (53)^2}$$

$$= \sqrt{2959.6667 - 2809}$$

$$= \sqrt{150.6667}$$

$$\sigma_y = 12.2746$$

$$r = \frac{cov(x, y)}{\sigma_x \, \sigma_y}$$

$$= \frac{-143.5}{(13.4412)(12.2746)}$$

$$= \frac{-143.5}{164.9854}$$

$$r = -0.8698.$$

3. Find Pearson's co-efficient of correlation for

$n = 12,\quad \Sigma x^2 = 921,$

$\Sigma x = 23,\quad \Sigma y^2 = 734.$

$\Sigma y = 28,$

$\Sigma xy = -432,$

Spearman rank correlation coefficient:-

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

where, $d = R_X - R_Y$

1. Spearman's rank correlation coefficient for the following data:-

X	52	63	45	36	72	65	47	25
Y	62	53	51	25	79	43	60	33

Soln:-

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \qquad d = R_x - R_y$$

x	y	R_x	R_y	$d = R_x - R_y$	d^2
52	62	4	2	2	4
63	53	3	4	-1	1
45	51	6	5	1	1
36	25	7	8	-1	1
72	79	1	1	0	0
65	43	2	6	-4	16
47	60	5	3	2	4
25	33	8	7	1	1
					28

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(28)}{8(8^2-1)}$$

$$= 1 - \frac{168}{8(63)}$$

$$= 1 - \frac{168}{504}$$

$$= 1 - 0.3333$$

$$\rho = 0.6667.$$

2. Calculate rank correlation coefficient.

X	8	7	6	3	2	1	5	4
Y	7	5	4	1	3	2	6	8

Soln:-

$$P = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

R_x	R_y	$d = R_x - R_y$	d^2
8	7	1	1
7	5	2	4
6	4	2	4
3	1	2	4
2	3	-1	1
1	2	-1	1
5	6	-1	1
4	8	-4	16
			32

$$P = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(32)}{8(8^2-1)}$$

$$= 1 - \frac{192}{8(63)}$$

$$= 1 - \frac{192}{504} = 1 - 0.3810$$

$$P = 0.619$$

3. Ten competitors in musical test were ranked by three judges A, B and C in the following data :-

A	1	6	5	10	3	2	4	9	7 8
B	3	5	8	4	7	10	2	1	6 9
C	6	4	9	8	×1	2	3	10	5 7

Using rank correlation method, discuss which pair of judges has the nearest approach to common ~~links~~ taste in music.

Soln :-

$$P_{AB} = 1 - \frac{6\Sigma d_{AB}^2}{n(n^2-1)}$$

$$P_{BC} = 1 - \frac{6\Sigma d_{BC}^2}{n(n^2-1)}$$

$$P_{AC} = 1 - \frac{6\Sigma d_{AC}^2}{n(n^2-1)}$$

R_A	R_B	R_C	d_{AB}	d_{AB}^2	d_{BC}	d_{BC}^2	d_{AC}	d_{AC}^2
1	3	6	−2	4	−3	9	−5	25
6	5	4	1	1	1	1	2	4
5	8	9	−3	9	−1	1	−4	16
10	4	8	6	36	−4	16	2	4
3	7	1	−4	16	6	36	2	4
2	10	2	−8	64	8	64	0	0
4	2	3	2	4	−1	1	1	1
9	1	10	8	64	−9	81	−1	1
7	6	5	1	1	1	1	2	4
8	9	7	−1	1	2	4	1	1
				200		214		60

$$P_{AB} = 1 - \frac{6 \Sigma d_{AB}^2}{n(n^2-1)}$$

$$= 1 - \frac{6(200)}{10(100-1)}$$

$$= 1 - \frac{1200}{990}$$

$$= 1 - 1.2121$$

$$= -0.2121.$$

$$P_{BC} = 1 - \frac{6(214)}{10(100-1)}$$

$$= 1 - \frac{1284}{990}$$

$$= 1 - 1.2970$$

$$= -0.2970.$$

$$P_{CA} = 1 - \frac{6(60)}{10(99)}$$

$$= 1 - \frac{360}{990}$$

$$= 1 - 0.3636$$

$$= 0.6364$$

P_{AC} is nearest to $+1$.

the pair of judges A and C has nearest approach to common taste in music.

Repeated Rank correlation coefficient :-

Calculate spearman rank correlation coefficient.

X 68 64 75 50 64 80 75 40 55 64

Y 62 58 68 45 81 60 68 48 50 70

Soln:-

$$\rho = 1 - \frac{6\left[\Sigma d^2 + \frac{m(m^2-1)}{12} + \cdots\right]}{n(n^2-1)}$$

X	Y	R_X	R_Y	$d = R_X - R_Y$	d^2
68	62	4	5	-1	1
64	58	6	8	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	7	0	0
64	70	6	2	4	16
					72

$$\rho = 1 - \frac{6\left[\Sigma d^2 + \frac{m(m^2-1)}{12} + \frac{m(m^2-1)}{12} + \frac{m(m^2-1)}{12}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[72 + \frac{5(9-1)}{12} + \frac{2(4-1)}{12} + \frac{2(4-1)}{12}\right]}{10(100-1)}$$

$$= 1 - \frac{6[72 + 2 + 0.5 + 0.5]}{990}$$

$$= 1 - \frac{6[75]}{990}$$

$$= 1 - \frac{450}{990}$$

$$= 1 - 0.4545$$

$$\rho = 0.5455$$

Calculate product moment correlation from the following bivariate frequency table:-

X \ Y	1	3	5
-1	1	1	4
0	3	7	1
2	6	2	0

Soln:-

$$r = \frac{N\Sigma fxy - (\Sigma fx)(\Sigma fy)}{\sqrt{N\Sigma fx^2 - (\Sigma fx)^2}\sqrt{N\Sigma fy^2 - (\Sigma fy)^2}}$$

9. Regression Analysis

Regression is *the measures of the average relationship between two or more variables in terms of the original units of the data.*

Literally regression means ***going back*** or ***stepping back.***

(The estimation of regression is called ***regression analysis.*** In regression analysis, two variables are involved. One variable is called ***dependent variable*** and the other is called ***independent variable.***

The yield of rice and rainfall are related. Yield of rice is a dependent variable and rainfall is an independent variable.)

In regression analysis, the value of dependent variable can be calculated from the value of independent variable.

Thus in regression analysis the value of an unknown variable can be calculated from the value of a known variable.

One variable is represented as ***X*** and the other variable is represented as ***Y.***

The graphic representation of regression is called ***regression line.***

There are two regression lines. They are,

Regression line of ***X*** on ***Y.***

Regression line of ***Y*** on ***X.***

Definitions

1. *"Regression is the measure of the average relationship between two or more variables in terms of the original units of the data"*- ***Blair.***

2. *"Regression analysis attempts to establish the nature of the relationship between variables, that is, to study the functional relationship between the variables and thereby provide a mechanism for predicting or forecasting".* ***-Ya-Lun-Chow.***

3. *"It is often more important to find out what the relation actually is, in order to estimate or predict one variable (the dependent variable) and statistical technique appropriate in such a case is called Regression analysis".* ***-Walis*** and ***Robert.***

4. *"One of the most frequently used techniques in economics and business research, to find a relation between two or more variables that are related casually, is regression analysis".* ***-Taro Yamane.***

In the regression analysis, the independent variable is also known as the ***"regressor"*** or ***"predictor"*** *or* ***"explanator"*** *and the dependent variable is known as* ***"regressed"*** *or* ***"explained"*** *variable.*

Types of Regression Analysis

There are 3 types of regression analysis. They are:

a. Simple and multiple
b. Linear and non-linear
c. Total and partial.

. Simple and Multiple

The regression analysis confined to the study of only two variables at a time is termed as ***simple regression.*** On the otherhand, the regression analysis for studying more than two variables at a time is known as ***multiple regression.***

Regression:-

The regression equation on x on y is,

$$(x-\bar{x}) = \boxed{r\frac{\sigma_x}{\sigma_y}}(y-\bar{y})$$

$b_{xy} \to$ regression co-efficient

The regression equation for y on x is,

$$(y-\bar{y}) = \boxed{r.\frac{\sigma_y}{\sigma_x}}(x-\bar{x})$$

$b_{yx} \to$ regression co-efficient

Find two regression equations and also estimate y when x = 20.

X	10	12	13	12	16	15
Y	40	38	43	45	37	46

Soln:-

The regression equation of x on y,

$$(x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

The regression equation of y on x,

$$(y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

$$r = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$$

$$\text{cov}(x,y) = \frac{\Sigma xy}{n} - (\bar{x})(\bar{y})$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2}$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - (\bar{y})^2}$$

$$\bar{x} = \frac{\Sigma x}{n}, \quad \bar{y} = \frac{\Sigma y}{n}.$$

x	y	x^2	y^2	xy
10	40	100	1600	400
12	38	144	1444	456
13	43	169	1849	559
12	45	144	2025	540
16	37	256	1369	592
15	43	225	1849	645
78	246	1038	10136	3192

$$\bar{x} = \frac{\Sigma x}{n} = \frac{78}{6} = 13$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{246}{6} = 41.$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2}$$

$$= \sqrt{\frac{1038}{6} - (13)^2}$$

$$= \sqrt{173 - 169}$$

$$= \sqrt{4}$$

$$= 2$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - (\bar{y})^2}$$

$$= \sqrt{\frac{10136}{6} - 1681}$$

$$= \sqrt{1689.3333 - 1681}$$

$= \sqrt{8.3333}$

$= 2.8867$.

$$\operatorname{cov}(x, y) = \frac{\Sigma xy}{n} - (\bar{x})(\bar{y})$$

$$= \frac{3192}{6} - (13)(41)$$

$$= 532 - 533$$

$$= -1.$$

Mode 2

Reg 2

Lin 1

x , y M+

Shift 2 →

Shift $\boxed{2}$ → $r =$

$$r = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y}$$

$$= \frac{-1}{5.7734}$$

$$= -0.1732.$$

The required equation of x on y

$$(x - \bar{x}) = r \cdot \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$(x - 13) = (-0.1732) \frac{2}{2.8867} (y - 41)$$

$$(x - 13) = -0.1200(y - 41)$$

$$x - 13 = -0.1200y + 4.92$$

$$x = -0.1200y + 4.92 + 13$$

$$x = -0.1200y + 17.92.$$

$$(y-\bar{y}) = r \cdot \frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

$$(y-41) = (-0.1732)\frac{2.8867}{2}(x-13)$$

$$(y-41) = -0.2500(x-13)$$

$$y-41 = -0.2500x + 3.25$$

$$y = -0.2500x + 3.25 + 41$$

$$y = -0.2500x + 44.25$$

when $x = 20$

$$y = -0.2500(20) + 44.25$$

$$y = -5 + 44.25$$

$$y = 39.25.$$

$9x - 3y = 165$ and $3x - 4y = 40$ are the regression equations of x (expenditure on advertisement in Rs. Lakhs) and y (sales in ₹. Lakhs). Variance of $x = 25$ (σ_x^2). Find (i) $\bar{x}$ and $\bar{y}$, (ii) r, (iii) S.D of Y, (iv) the probable change in sales when the expenditure on advertisement is ₹. 24,00,000 (Lakhs).

Soln:-

(i) Take (1) $\Rightarrow$ $9\bar{x} - 3\bar{y} = 165 \rightarrow (1)$

(2) $\times 3$ $\Rightarrow$ $9\bar{x} - 12\bar{y} = 120$

$$9\bar{y} = 45$$

$$\bar{y} = \frac{45}{9} = 5.$$

Sub $\bar{y} = 5$ in equ (1)

$$9\bar{x} - 3(5) = 165$$

$$9\bar{x} = 165 + 15$$

$$9\bar{x} = 180$$

$$\bar{x} = \frac{180}{9} = 20.$$

(ii) The regression equation x on y

$$9x - 3y = 165$$

$$9x = 165 + 3y$$

$$x = \frac{3}{9}y + \frac{165}{9}$$

$$b_{xy} = \frac{3}{9}.$$

The regression equation y on x

$$3x - 4y = 40$$

$$-4y = -3x + 40$$

$$y = \frac{+3}{+4}x - \frac{40}{4}$$

$$b_{yx} = \frac{3}{4}.$$

$$r = \pm\sqrt{b_{xy}\ b_{yx}}$$

$$= \pm\sqrt{\frac{3}{9} \times \frac{3}{4}}$$

$$= \pm\sqrt{1/4}$$

$$r = 0.5$$

(iii) $V(x) = 25$. (i.e) $\sigma x^2 = 25$

$$\sigma x = \sqrt{25} = 5$$

$$b_{xy} = r \cdot \frac{\sigma x}{\sigma y}$$

$$0.3333 = (0.5) \frac{5}{\sigma y}$$

$$\sigma y = \frac{(0.5)(5)}{0.3333}$$

$$\sigma y = 7.5008.$$

(iv) $x = 24$ lakhs.

$$Y = \frac{3}{4}(x) - \frac{40}{4}$$

$$= \frac{3}{4}(24) - 10$$

$Y = 8$. Lakhs.

From the following data. Find the

	X	Y
sum of the squares of deviations from the arithmetic mean	8250	724
Sum of the product of deviation of x and Y from their respective mean	2350	
No. of pairs of value	10	

Soln :-

Sum of the squares of deviations from the arithmetic mean.

$\Sigma(x-\bar{x})^2$.

$\bar{x}=0$

$\Sigma x^2 = 8250$

$\Sigma(y-\bar{y})^2$ $\bar{y}=0$

$\Sigma y^2 = 724$

$\Sigma(x-\bar{x})(y-\bar{y})$

$\Sigma xy = 2350$.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$ (For sum of deviations.)

$$= \frac{2350}{\sqrt{8250}\sqrt{724}}$$

$$= \frac{2350}{90.8295 \times 26.9072}$$

$$= \frac{2350}{2443.9675}$$

$= 0.9616$.