CORRELATION

The statistical tool for measuring the degree of relationship between the two variables *i.e.*, a change in one variable results a positive or negative change in the other and also a greater change in one variable results in corresponding greater or smaller change in other variable is known as correlation.

• Co-efficient of Correlation:

The extent or degree of relationship between the two variables is measured in terms of another parameter called co-efficient of correlation.

It is denoted by 'r' *i.e.*, $-1 \le r \le 1$.

• Properties of Co-efficient of Correlation.

I. It is a measure of the closeness between the two variables.

II. It lies between -1 and $+1$ *i.e.*, $-1 \le r \le 1$.

III. The correlation is perfect and positive if $r = 1$ and it is perfect and negative if $r = -1$.

IV. If, $r = 0$ then there is no correlation between the two variables and said to be independent.

TYPES OF CORRELATION:

(a) Perfect Positive Correlation:

I. Here two variables denoted by letter X and Y are directly proportional and fully correlated with each other.

II. The correlation co-efficient (r) = +1 *i.e.*, both the variables increases and decreases in the same proportion.

III. Perfect correlation is not usually found in nature but approaching to that extent.

Example: Height and weight, age and weight up to a certain age.

IV. The graph forms a straight line originating from the lower ends of X and Y axes.

(b) Perfect Negative Correlation:

I. In this two variables (X & Y) are inversely proportional to each other *i.e.*, when one rises, the other falls in the same proportion.

II. Here correlation co-efficient $r = -1$.

III. It is not usually available in nature but some approaching to that extent.

Examples:

(*i*) Mean weekly temperature and intensity of cold in winter.

(*ii*) Pressure and volume of gas at a particular temperature.

IV. The graph will contain all the observations on a straight line starting from either of the extreme ends because one variable rises and other falls in a fixed proportion or $r = -1$.

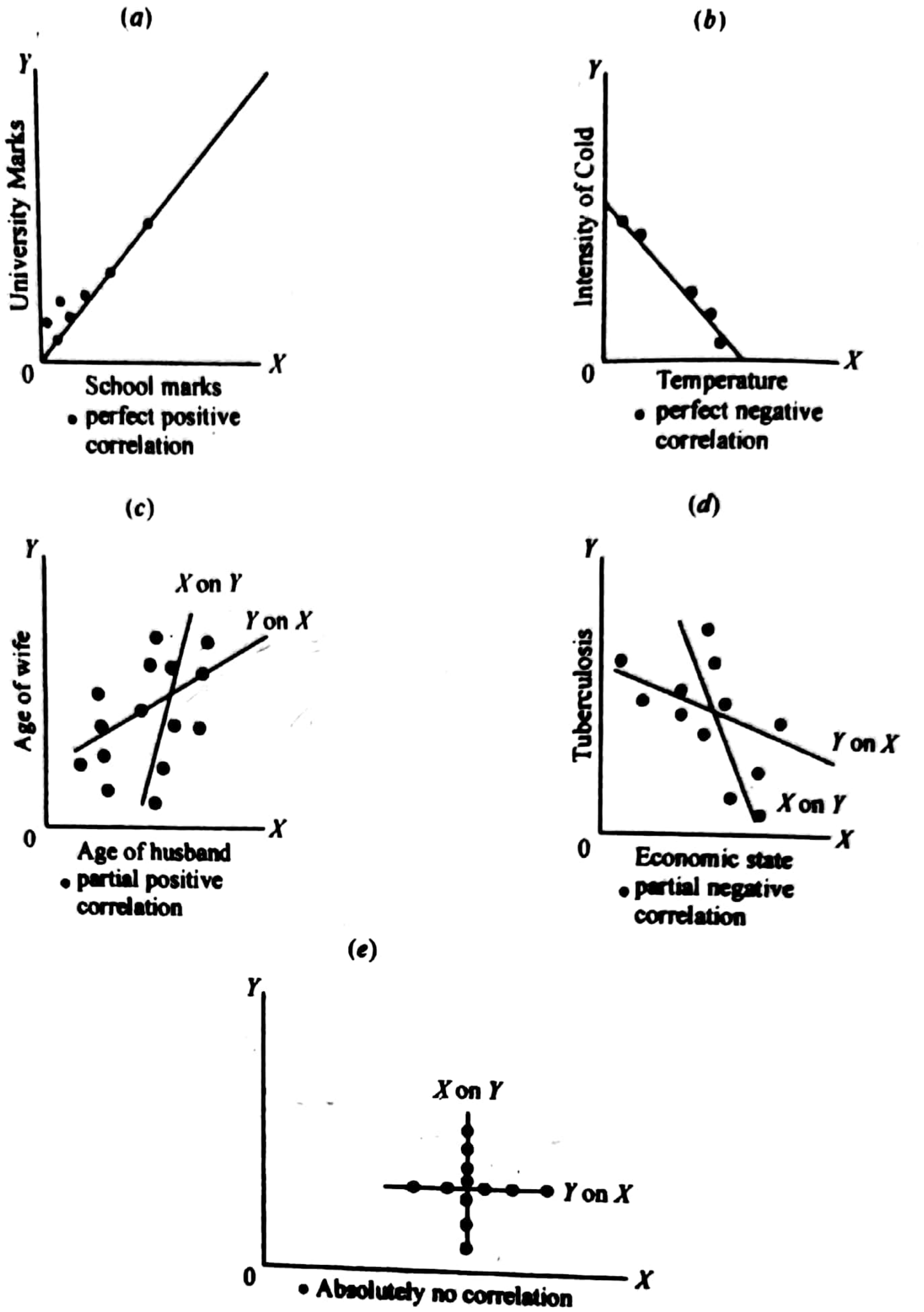

Fig. 13.1 The diagrams are taken from hypothetical numbers to show differnet types of correlation.

(c) Moderately (partial) Positive Correlation:

I. Here the non zero values of correlation co-efficient (r) lies between 0 and +1 *i.e.*, $0 < r < 1$.

Examples:

(*i*) Infant mortality rate and overcrowding.

(*ii*) Age of husband and age of wife.

(*iii*) Temperature and pulse rate.

II. Here scatter diagram will be, around an imaginary mean line rising from lower extreme values of both variables.

(d) Moderately (partial) Negative Correlation:

I. In this case, the non zero values of correlation co-efficient lies between –1 and 0 *i.e.*, $-1 < r < 0$.

Examples:

(*i*) Income and infant mortality rate.

(*ii*) Age and vital capacity in adult.

II. In this case the scatter diagram will be of the same type but the mean imaginary line will rise from extreme values of one variable.

(e) Absolutely No Correlation:

I. Here the value of correlation coefficient is zero, indicating that no linear relationship exists between the two variables.

II. There is no mean or imaginary line indicating trend of correlation.

III. Here the variable '*X*' is completely independent of variable '*Y*'.

Example: Height and pulse rate.

IV. Here the points are scattered, so that no imaginary line can be drawn, the correlation will be zero.

(f) Perfect and imperfect Correlation:

I. In perfect correlation, the dots lie exactly on a straight line.

II. Here the changes in the corresponding values of the two variables are proportional directly or inversely. The degree of imperfect correlation lies between perfect correlation and no or zero correlation.

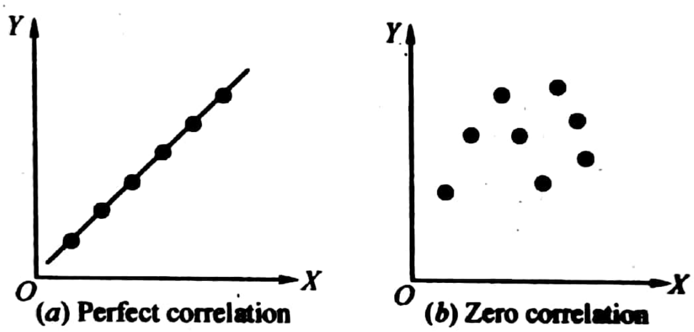

Fig. 13.2 (*a* & *b*) Showing (*a*) Perfect correlator & (*b*) Zero correlation.

Simple and multiple Correlation:

I. It deals with the study of two variables.

II. The relationship between this two variable is called simple correlation.

Example: Trunk length and wing length in a sample of cockroach.

In multiple correlations we study more than two variables simultaneously.

Example: Oxygen consumption and the tracheal ventilation & atmospheric oxygen tension.

Partial and Total Correlation:

The study of two variables partialling out (excluding) some other variables is called partial correlation.

Example: Oxygen consumption and tracheal veritilation excluding partialling out atmospheric oxygen tension.

In total correlation all the facts are taken into account.

Liner and non linear correlation:

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

Example: Body weight and gill weight in a sample of fishes.

σ_y = standard ...

Working Procedure:

I. Denote one series by X and the other series by Y.

II. Calculate $\bar{X}$ (mean) and $\bar{Y}$ mean of the x and y series respectively.

III. Take the deviations of the observations in X series from $\bar{X}$ and written as $\sigma_x = X - \bar{X}$. Tak the deviations of the observations in Y series from $\bar{Y}$ and written it under the column heade by $\sigma_y = Y - \bar{Y}$.

IV. Square the deviations and written them under the columns headed by dx^2 and dy^2

V. Multiply the respective dx and dy and write it under the column headed by $dx\,dy$.

VI. Apply the following formula to calculate r or r_{xy} (the coefficient of correlation)

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}} \quad \text{or } r = \frac{\sum dxdy}{n\sigma_x \times \sigma_y}$$

n = number of observation

σ_x = S.D. of X

σ_y = S.D. of Y

Significance of r:

(*i*) The computed r can be transformed into t score for interpretation by using standard error (S of r.

(*ii*) The degrees of freedom df of computed t should be taken as $n - 2$.

$$r = Sr = \sqrt{\frac{t - r^2}{n-2}} \qquad t = \frac{r}{Sr} \qquad df = n - 2$$

(*iii*) The computed r is considered significant at or below the level of significance whose critica t either equal or lower than computed $t\ (p \le \alpha)$.

Example 1. Find the coefficient correlation between the heights of fathers and daughter both from the following Ganguli family members.

Height of father (in cm)	64	65	66	67	68	69	70
Height of daughters (in cm)	66	67	68	69	70	71	72

Solution: Let the height of father denoted by X & height of daughters by Y, then

$$\bar{X} = \frac{64 + 65 + 66 + 67 + 68 + 69 + 70}{7} = \frac{469}{7} = 67$$

$$\bar{Y} = \frac{66 + 67 + 68 + 69 + 70 + 71 + 72}{7} = \frac{483}{7} = 69$$

Height of father (X)	dx X − 67	dx² (X − 67)²	Height of daughter (Y)	dy Y − 69	dy² (Y − 69)²	dxdy
64	64 − 67 = −3	9	66	66 − 69 = −3	9	9
65	65 − 67 = −2	4	67	67 − 69 = −2	4	4
66	66 − 67 = −1	1	68	68 − 69 = −1	1	1
67	67 − 67 = 0	0	69	69 − 69 = 0	0	0
68	68 − 67 = +1	1	70	70 − 69 = +1	1	1
69	69 − 67 = +2	4	71	71 − 69 = +2	4	4
70	70 − 67 = +3	9	72	72 − 69 = +3	9	9
$\bar{x} = 67$	$\sum dx = 0$	$\sum dx^2 = 28$	$\bar{y} = 69$	$\sum dy = 0$	$\sum dy^2 = 28$	$\sum dx \sum dy = 28$

Now

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}}$$

$$= \frac{28}{\sqrt{28 \times 28}}$$

$$= \frac{28}{28} = 1$$

Example 2. Calculate the correlation co-efficient between X and Y from the following data.

X	**5**	**9**	**13**	**17**	**21**
Y	**12**	**20**	**25**	**33**	**35**

Solution: Let us prepare the table.

X	dx X − 13	dx² (X − 13)²	Y	dy Y − 25	dy² (Y − 25)²	dxdy
05	5 − 13 = − 8	64	12	12 − 25 = − 13	169	104
09	9 − 13 = − 4	16	20	20 − 25 = − 5	25	20
13	13 − 13 = 0	0	25	25 − 25 = 0	0	0
17	17 − 13 = + 4	16	33	33 − 25 = + 8	64	32
21	21 − 13 = + 8	64	35	35 − 25 = + 10	100	80
65 $\bar{X} = \frac{65}{5} = 13$	$\sum dx = 0$	$\sum dx^2 = 160$	125 $\bar{Y} = \frac{125}{5} = 25$	$\sum dy = 0$	$\sum dy^2 = 358$	$\sum dxdy = 236$

Now

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}}$$

$$= \frac{236}{\sqrt{160 \times 358}} = \frac{236}{\sqrt{57280}}$$

$$= \frac{236}{239.33} = 0.986 \, (approx)$$

Example 3. Calculate correlation co-efficient between X & Y for the following data.

X	1	2	3	4	5	6	7	8	9
Y	10	11	12	14	13	15	16	17	18

Solution: Let us prepare the table.

X	dx $X-5$	dx^2 $(X-5)^2$	Y	dy $Y-14$	dy^2 $(Y-14)^2$	$dxdy$
1	$1-5=-4$	16	10	$10-14=-4$	16	16
2	$2-5=-3$	9	11	$11-14=-3$	9	9
3	$3-5=-2$	4	12	$12-14=-2$	4	4
4	$4-5=-1$	1	14	$14-14=\ 0$	0	0
5	$5-5=\ 0$	0	13	$13-14=-1$	1	0
6	$6-5=+1$	1	15	$15-14=+1$	1	1
7	$7-5=+2$	4	16	$16-14=+2$	4	4
8	$8-5=+3$	9	17	$17-14=+3$	9	9
9	$9-5=+4$	16	18	$18-14=+4$	16	16
$\sum x = 45$ $\bar{x} = 5$	$\sum dx = 0$	$\sum dx^2 = 60$	$\sum y = 126$ $\bar{y} = 14$	$\sum dy = 0$	$\sum y^2 = 60$	$\sum dxdy = 59$

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}}$$

$$= \frac{59}{\sqrt{60 \times 60}} = \frac{59}{60} = .98\,(approx)$$

- **Short Cut Method or Assume Mean Method:**

I. When the terms of series 'x' and 'y' are big and calculation of $\bar{X}$ (mean) and $\bar{Y}$ (mean) becomes difficult.

II. The means of X and Y are not integers.

The following formula is applied.

$$r_{xy} = \frac{\sum dxdy - \left(\frac{\sum dx \sum dy}{n}\right)}{\sqrt{\sum dx^2 - \frac{\left(\sum dx\right)^2}{n}} \times \sqrt{\sum dy^2 - \frac{\left(\sum dy\right)^2}{n}}}$$

Where $dx = x - a$

$dy = y - b$

a = assumed mean of x series.

b = assumed mean of y series.

n = number of observation of x or y.

- **Working Procedure:**

I. Take any term 'a' (Preferably the middle one) of X series as assumed mean and any term 'b' (*preferably middle one*) as assumed mean of Y.

II. Take deviations of the observation in 'X' series from 'a' *i.e.,* $dx = X - a$.

Take deviations of the observation in 'Y' series from 'b' *i.e.,* $dy = Y - b$.

Example 3. Calculate correlation co-efficient between X & Y for the following data.

X	1	2	3	4	5	6	7	8	9
Y	10	11	12	14	13	15	16	17	18

Solution: Let us prepare the table.

X	dx $X - 5$	dx^2 $(X - 5)^2$	Y	dy $Y - 14$	dy^2 $(Y - 14)^2$	$dxdy$
1	$1 - 5 = -4$	16	10	$10 - 14 = -4$	16	16
2	$2 - 5 = -3$	9	11	$11 - 14 = -3$	9	9
3	$3 - 5 = -2$	4	12	$12 - 14 = -2$	4	4
4	$4 - 5 = -1$	1	14	$14 - 14 = 0$	0	0
5	$5 - 5 = 0$	0	13	$13 - 14 = -1$	1	0
6	$6 - 5 = +1$	1	15	$15 - 14 = +1$	1	1
7	$7 - 5 = +2$	4	16	$16 - 14 = +2$	4	4
8	$8 - 5 = +3$	9	17	$17 - 14 = +3$	9	9
9	$9 - 5 = +4$	16	18	$18 - 14 = +4$	16	16
$\Sigma x = 45$ $\bar{x} = 5$	$\Sigma dx = 0$	Σdx^2 $= 60$	$\Sigma y = 126$ $\bar{y} = 14$	$\Sigma dy = 0$	Σy^2 $= 60$	$\Sigma dxdy$ $= 59$

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}}$$

$$= \frac{59}{\sqrt{60 \times 60}} = \frac{59}{60} = .98\,(approx)$$

- **Short Cut Method or Assume Mean Method:**

I. When the terms of series 'x' and 'y' are big and calculation of $\bar{X}$ (mean) and $\bar{Y}$ (mean) becomes difficult.

II. The means of X and Y are not integers.

The following formula is applied.

$$r_{xy} = \frac{\sum dxdy - \left(\frac{\sum dx \sum dy}{n}\right)}{\sqrt{\sum dx^2 - \frac{\left(\sum dx\right)^2}{n}} \times \sqrt{\sum dy^2 - \frac{\left(\sum dy\right)^2}{n}}}$$

Where $dx = x - a$

$dy = y - b$

a = assumed mean of x series.

b = assumed mean of y series.

n = number of observation of x or y.

- **Working Procedure:**

I. Take any term 'a' (Preferably the middle one) of X series as assumed mean and any term 'b' (*preferably middle one*) as assumed mean of Y.

II. Take deviations of the observation in 'X' series from 'a' *i.e.*, $dx = X - a$.

Take deviations of the observation in 'Y' series from 'b' *i.e.*, $dy = Y - b$.

There are written under the columns dx and dy respectively.

III. Square the deviations and write them under the columns headed by dx^2 and dy^2.

IV. Multiply the respective 'dx' and 'dy' and write it under the column $dx.dy$.

V. Apply the following formula to calculate r or r_{xy} (correlation coefficient).

$$r = \frac{\sum dxdy - \left(\frac{\sum dx \sum dy}{n}\right)}{\sqrt{\sum dx^2 - \frac{\left(\sum dx\right)^2}{n}} \times \sqrt{\sum dy - \frac{\left(\sum dy\right)^2}{n}}}$$

Example: Calculate coefficient-correlation between X and Y for the following data.

X	1	3	4	5	7	8	10
Y	2	6	8	10	14	16	20

Solution: Let 5 be the assumed mean for the values of x and 14 be the assumed for the values of y.

X	dx X – 5	dx^2 $(X-5)^2$	Y	dy Y – 14	dy^2 $(Y-14)^2$	dxdy
1	1 – 5 = – 4	16	2	2 – 14 = – 12	144	48
3	3 – 5 = – 2	4	6	6 – 14 = – 8	64	16
4	4 – 5 = – 1	1	8	8 – 14 = – 6	36	6
5	5 – 5 = 0	0	10	10 – 14 = – 4	16	0
7	7 – 5 = + 2	4	14	14 – 14 = 0	0	0
8	8 – 5 = + 3	9	16	16 – 14 = + 2	4	6
10	10 – 5 = + 5	25	20	20 – 14 = + 6	36	30
	$\sum dx = 3$	$\sum dx^2 = 59$		$\sum dy = -30 + 8 = -22$	$\sum dy^2 = 300$	$\sum dxdy = 106$

$$r_{xy} = \frac{\sum dxdy - \left(\frac{\sum dx \sum dy}{n}\right)}{\sqrt{\sum dx^2 - \frac{\left(\sum dx\right)^2}{n}} \times \sqrt{\sum dy^2 - \frac{\left(\sum dy\right)^2}{n}}}$$

$$= \frac{106 - \left(\frac{3.-22}{7}\right)}{\sqrt{59 - \frac{(3)^2}{7}} \times \sqrt{300 - \frac{(-22)^2}{7}}} = \frac{106 + \frac{66}{7}}{\sqrt{59 - \frac{9}{7}} \times \sqrt{300 - \frac{484}{7}}}$$

$$= \frac{\frac{742 + 66}{7}}{\sqrt{\frac{413 - 9}{7}} \times \sqrt{\frac{2100 - 484}{7}}} = \frac{\frac{808}{7}}{\sqrt{\frac{404}{7} \times \frac{1616}{7}}} = \frac{\frac{808}{7}}{\frac{808}{7}} = 1$$

- **Covariance Method or Product Moment Method**

The most widely used mathematical method of measuring correlation is "Pearson co-efficient correlation" due to *Karl Pearson.*

r = Product moment correlation co-efficient.

$$= \frac{\text{covariance of } x \text{ and } y}{(S.D\, of\, X)(S.D\, of\, Y)} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N\sigma_x\sigma_y}$$

$$\left[\text{cov.}(X.Y) = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N}\right]$$

σ_X = S.D. of X

σ_Y = S.D. of Y

N = Number of observation

Example 1. Marks of 10 students in Mathematics and Statistics are given below.

Mathematics (X)	32	38	48	43	40	22	41	69	35	64
Statistics (Y)	30	31	38	43	33	11	27	76	40	59

Calculate correlation co-efficient between X & Y using product moment formula.

Solution: Let 43 be the assumed mean for the values of X & 40 be the assumed mean for the values of Y.

	X	dx X − 43	dx^2 $(X-43)^2$	Y	dy Y − 40	dy^2 $(Y-40)^2$	dxdy
1	32	32 − 43 = − 11	121	30	30 − 40 = − 10	100	110
2	38	38 − 43 = − 5	25	31	31 − 40 = − 9	81	45
3	48	48 − 43 = + 5	25	38	38 − 40 = − 2	04	−10
4	43	43 − 43 = 0	0	43	43 − 40 = + 3	09	0
5	40	40 − 43 = − 3	9	33	33 − 40 = − 7	49	21
6	22	22 − 43 = − 21	441	11	11 − 40 = − 29	841	609
7	41	41 − 43 = − 2	4	27	27 − 40 = − 13	169	26
8	69	69 − 43 = + 26	676	76	76 − 40 = + 36	1296	936
9	35	35 − 43 = − 8	64	40	40 − 40 = 0	0	0
10	64	64 − 43 = + 21	441	59	59 − 40 = + 19	361	399
		$\sum dx = 2$	$\sum dx^2 = 1806$		$\sum dy = -12$	$\sum dy^2 = 2910$	2136

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N\sigma_x\sigma_y}. \text{ It may be written as}$$

$$r = \frac{n\sum dxdy - (\sum dx)(\sum dy)}{\sqrt{n\sum dx^2 - (\sum dx)^2} \times \sqrt{n\sum dy^2 - (\sum dy)^2}} \quad \textit{(Proved)}$$

Here $\sum dx = 2$ $\quad \sum dy = -12$ $\quad n = 10$

$\sum dx^2 = 1806$ $\quad \sum dy^2 = 2910$ $\quad \sum dxdy = 2136$

$$r = \frac{10.2136 - 2.(-12)}{\sqrt{10.1806 - (2)^2} \times \sqrt{10.2910 - (-12)^2}}$$

$$= \frac{10.2136 + 24}{\sqrt{10.1806 - 4} \times \sqrt{10.2910 - 144}}$$

$$= \frac{21360 + 24}{\sqrt{18060 - 4}\sqrt{29100 - 144}} = \frac{21384}{\sqrt{18056}\sqrt{28956}}$$

$$= \frac{21384}{134.3 \times 170.16} = \frac{21384}{22852} = 0.935\,(approx)$$

Example 2. Calculate Pearson's coefficient of correlation from the following data and interpret the result.

A	104	111	104	114	118	117	105	108	106	100	104	105
B	57	55	47	45	45	50	64	63	66	62	69	61

Solution:

A(X)	$X - \bar{X} = dx$ $\bar{X} = 100$	dx^2	B(Y).	$Y - \bar{Y} = dy$ $\bar{Y} = 50$	dy^2	$dx \times dy$
104	104 – 100 = 4	16	57	57 – 50 = 7	49	4 × 7 = 28
111	111 – 100 = 11	121	55	55 – 50 = 5	25	11 × 5 = 55
104	104 – 100 = 4	16	47	47 – 50 = –3	09	4 × –3 = –12
114	114 – 100 = 14	196	45	45 – 50 = –5	25	14 × –5 = –70
118	118 – 100 = 18	324	45	45 – 50 = –5	25	18 × –5 = –90
117	117 – 100 = 17	289	50	50 – 50 = 0	00	17 × 0 = 0
105	105 – 100 = 5	25	64	64 – 50 = 14	196	5 × 14 = 70
108	108 – 100 = 8	64	63	63 – 50 = 13	169	6 × 16 = 96
106	106 – 100 = 6	36	66	66 – 50 = 16	256	0 × 12 = 0
100	100 – 100 = 0	00	62	62 – 50 = 12	144	4 × 19 = 76
104	104 – 100 = 4	16	69	69 – 50 = 19	361	5 × 11 = 55
105	105 – 100 = 5	25	61	61 – 50 = 11	121	
	$\sum dx = 96$	$\sum dx^2 = 1128$		$\sum dy = 84$	$\sum dy^2 = 1380$	$\sum dxdy = 312$

$$r_{xy} = \frac{\sum dxdy - \left(\frac{\sum dx \sum dy}{n}\right)}{\sqrt{\sum dx^2 - \frac{(\sum dx)^2}{n}} \times \sqrt{\sum dy^2 - \frac{(\sum dy)^2}{n}}}$$

$\sum dx = 96$ $\sum dx^2 = 1128$
$\sum dy = 84$ $\sum dy^2 = 1380$
$\sum dxdy = 312$

$$r_{xy} = \frac{312 - \left(\frac{96 \times 84}{12}\right)}{\sqrt{1128 - \frac{(96)^2}{12}} \times \sqrt{1380 - \frac{(84)^2}{12}}} = \frac{312 - \frac{8064}{12}}{\sqrt{1128 - 768} \times \sqrt{1380 - 588}}$$

$$r_{xy} = \frac{312 - 672}{\sqrt{360} \times \sqrt{792}} = \frac{-360}{\sqrt{285120}} = \frac{-360}{533.96} = 0.6742 = 0.67\,(approx)$$

The result *i.e.,* the value of correlation coefficient indicates that there is negative correlation between the two variables.

Example 3. Ten students of Zoology of Serampore college obtained marks in Taxonomy and Animal physiology are given below. Calculate Karl Pearson's correlation coefficient between the marks in Taxonomy and Animal physiology.

Marks in Taxonomy	10	25	13	25	22	11	12	25	21	20
Names in Animal Physiology	12	22	16	15	18	18	17	23	24	17

Solution:

Marks in Taxonomy (X)	Assume mean 18 $X - \bar{X} = dx$	dx^2	Marks in Animal Phsiology (Y)	Assume mean 18 $Y - \bar{Y} = dy$	dy^2	dx dy
10	10 – 18 = – 8	64	12	12 – 18 = – 6	36	– 8 × 6 = 48
25	25 – 18 = + 7	49	22	22 – 18 = + 4	16	7 × 4 = 28
13	13 – 18 = – 5	25	16	16 – 18 = – 2	04	– 5 × – 2 = 10
25	25 – 18 = + 7	49	15	15 – 18 = – 3	09	7 × – 3 = – 21
22	22 – 18 = + 4	16	18	18 – 18 = 0	00	4 × 0 = 0
11	11 – 18 = – 7	49	18	18 – 18 = 0	00	–7 × 0 = 0
12	12 – 18 = – 6	36	17	17 – 18 = – 1	01	–6 × –1 = 6
25	25 – 18 = + 7	49	23	23 – 18 = + 5	25	7 × 5 = 35
21	21 – 18 = + 3	09	24	24 – 18 = + 6	36	3 × 6 = 18
20	20 – 18 = + 2	04	17	17 – 18 = – 1	01	2 × –1 = –2
184	$\sum dx = 4$	350	182	$\sum dy = 2$	$\sum dy^2 = 128$	$\sum dxdy = 122$

$$r_{xy} = \frac{\sum dxdy - \left(\frac{\sum dx \sum dy}{n}\right)}{\sqrt{\sum dx^2 - \frac{\left(\sum dx\right)^2}{n}} \times \sqrt{\sum dy^2 - \frac{\left(\sum dy\right)^2}{n}}}$$

$\sum dx = 4$ $\sum dx^2 = 350$

$\sum dy = 2$ $\sum dy^2 = 128$

$\sum dxdy = 122$ $n = 10$

$$= \frac{122 - \left(\frac{4 \times 2}{10}\right)}{\sqrt{350 - \frac{(4)^2}{10}} \times \sqrt{128 - \frac{(2)^2}{10}}} = \frac{122 - 0.8}{\sqrt{350 - 1.6} \times \sqrt{128 - 0.4}}$$

$$= \frac{121.2}{\sqrt{340.4} \times \sqrt{127.6}} = \frac{121.2}{\sqrt{44455.84}} = \frac{121.2}{210.8455}$$

$$= \frac{121.2}{210.85} = 0.5748$$

Example 4. Calculate Karl Pearson's co-efficient of correlation from the following data.

Serial No. of Student:	1	2	3	4	5	6	7	8	9	10
Marks in Zoology:	15	18	21	24	27	30	36	39	42	48
Marks in Chemistry:	25	25	27	27	31	33	35	41	41	45

Solution:

Sl No.	Marks in Zoology (X)	$X - \bar{X} = dx$	dx^2	Marks in Chemistry (Y)	$Y - \bar{Y} = dy$	dy^2	$dx.dy$
1	15	15 – 30 = – 15	225	25	25 – 33 = – 8	64	120
2	18	18 – 30 = – 12	144	25	25 – 33 = – 8	64	96
3	21	21 – 30 = – 09	81	27	27 – 33 = – 6	36	54
4	24	24 – 30 = – 06	36	27	27 – 33 = – 6	36	36
5	27	27 – 30 = – 03	09	31	31 – 33 = – 2	04	06
6	30	30 – 30 = 00	00	33	33 – 33 = 00	00	00
7	36	36 – 30 = + 06	36	35	35 – 33 = + 02	04	12
8	39	39 – 30 = + 09	81	41	41 – 33 = + 08	64	72
9	42	42 – 30 = + 12	144	41	41 – 33 = + 08	64	96
10	48	48 – 30 = + 18	324	45	45 – 33 = + 12	144	216
	$\sum x = 300$		$\sum dx^2 = 1080$	330		$\sum dy^2 = 480$	708

$$\bar{X} = \frac{\sum x}{N} = \frac{300}{10} = 30 \quad \bar{Y} = \frac{\sum y}{N} = \frac{330}{10} = 33 \quad \sum dxdy = 708$$

$$\therefore \quad r(xy) = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}} = \frac{708}{\sqrt{1080 \times 480}}$$

$$= \frac{708}{\sqrt{9 \times 12 \times 48 \times 100}} = \frac{708}{\sqrt{9 \times 3 \times 4 \times 3 \times 16 \times 10^2}}$$

$$= \frac{708}{\sqrt{9^2 \times 8^2 \times 10^2}} = \frac{708}{9 \times 8 \times 10} = \frac{708}{720} = 0.9833 = 0.98$$

Example 5. The rainfall and the output of wheat per acre for a farm under Kalyani agricultural institute was as follows.

Rain fall (in cms):	40	20	32	35	40	45	43	30	25	50
Wheat production (in quintals):	120	120	145	150	100	120	120	135	130	140

Find the correlation coefficient between the rainfall and wheat production.

Solution:

Rain fall (X)	$X - \bar{X} = dx$	dx^2	Wheat production (Y)	$Y - \bar{Y} = dy$	dy^2	$dx.dy$
40	40 – 36 = + 04	16	120	120 – 130 = – 10	100	– 40
20	20 – 36 = – 16	256	120	120 – 130 = – 10	100	160
32	32 – 36 = – 04	16	145	145 – 130 = + 15	225	– 60
35	35 – 36 = – 01	01	150	150 – 130 = + 20	400	– 20
40	40 – 36 = + 04	16	100	100 – 130 = – 30	900	– 120
45	45 – 36 = + 09	81	120	120 – 130 = – 10	100	– 90
43	43 – 36 = + 07	49	120	120 – 130 = – 10	100	– 70
30	30 – 36 = – 06	36	155	155 – 130 = + 25	625	– 150
25	25 – 36 = – 11	121	130	130 – 130 = 00	00	00
50	50 – 36 = + 14	196	140	140 – 130 = + 10	100	140
$\sum x = 360$		$\sum dx^2 = 788$	$\sum y = 1300$		$\sum dy^2 = 2650$	–250

$$\bar{X} = \frac{360}{10} = 36 \quad \bar{Y} = \frac{1300}{10} = 130$$

$$r(xy) = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}} = \frac{-250}{\sqrt{788 \times 2650}}$$

$$= \frac{-250}{\sqrt{2088200}} = \frac{-250}{1445} = -0.173$$

Example 6. Find out the Karl Pearson's coefficient of correlation of the following data.

Runs of Team A:	14	19	21	26	22	15	20	19	24
Runs of Team B:	31	36	37	50	45	33	41	39	48

Solution:

Team A Runs (X)	$X - \bar{X} = dx$	dx^2	Team B Runs (Y)	$Y - \bar{Y} = dy$	dy^2	dx.dy
14	14 – 20 = – 06	36	31	31 – 40 = – 09	81	54
19	19 – 20 = – 01	01	36	36 – 40 = – 04	16	04
21	21 – 20 = + 01	01	37	37 – 40 = – 03	09	- 03
26	26 – 20 = + 06	36	50	50 – 40 = + 10	100	60
22	22 – 20 = + 02	04	45	45 – 40 = + 05	25	10
15	15 – 20 = – 05	25	33	33 – 40 = – 07	49	35
20	20 – 20 = 00	00	41	41 – 40 = + 01	01	00
19	19 – 20 = – 01	01	39	39 – 40 = – 01	01	01
24	24 – 20 = + 04	16	48	48 – 40 = + 08	64	32
$\Sigma x = 180$		$\Sigma dx^2 = 120$	$\Sigma y = 360$		$\Sigma dy^2 = 346$	$\Sigma dxdy = 193$

$$\sum X = 180 \qquad \sum y = 360$$

$$\bar{X} = \frac{180}{9} = 20 \quad \bar{Y} = \frac{360}{9} = 40 \quad n = 9$$

$$\sum dx^2 = 120 \quad \sum dy^2 = 346 \quad \sum dxdy = 193$$

$$r_{xy} = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}} = \frac{193}{\sqrt{120 \times 346}}$$

$$= \frac{193}{\sqrt{41520}} = \frac{193}{203.76} = 0.947 = 0.95$$

... the brothers and ...