Module-4
Part – 4
STATISTICAL METHODS

Mean (Arithmetic mean): If $x_1, x_2, x_3, \ldots, x_n$ are set of n values of a variate, then the mean is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\Sigma x_i}{n}.$$

In a frequency distribution, if $x_1, x_2, x_3, \ldots, x_n$ be the mid-values of the class-intervals having frequencies $f_1, f_2, f_3, \ldots, f_n$ respectively, we have

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{f_1 + f_2 + f_3 + \ldots + f_n} = \frac{\Sigma x_i f_i}{\Sigma f_i}$$

Variance: If a variate x take the values $x_1, x_2, x_3, \ldots, x_n$ then the variance V is defined as follows:

$$V = \frac{\Sigma\left(x - \bar{x}\right)^2}{n}$$

Also for a grouped data,

$$V = \frac{\Sigma f\left(x - \bar{x}\right)^2}{\Sigma f}$$

Standard deviation (S.D):

$$\sigma = \sqrt{V} \text{ or } \sigma^2 = V$$

$$\Rightarrow \qquad \sigma = \sqrt{\frac{\Sigma\left(x-\bar{x}\right)^2}{n}} = \sqrt{\frac{\Sigma X^2}{n}} \text{ where } X = x - \bar{x}$$

$$\text{or} \qquad \sigma^2 = \frac{\Sigma\left(x-\bar{x}\right)^2}{n} = \frac{\Sigma X^2}{n}$$

Alternative formula for σ^2

$$\sigma^2 = \frac{\Sigma x^2}{n} - \left(\bar{x}\right)^2$$

Correlation

- **The two variables** ***x and y are related in such a way that an increase in*** **one is accompanied by an increase or decrease in the other is called co-variation.**
- **Co-variation of two independent magnitudes is known as correlation.**

Correlation

- **Correlation is Positive when the values increase together and Correlation is Negative when one value decreases as the other increases.**

Coefficient of correlation

The numerical measure of correlation between two variables *x and y is* known as **Karl Pearson's coefficient of correlation** or **simply coefficient of correlation** and is denoted by ***r*** and is defined by

$$r = \frac{\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)}{n\sigma_x\sigma_y}$$

$$r = \frac{\Sigma XY}{n\sigma_x\sigma_y}$$

Where X = deviation from the mean = $x-\bar{x}$, Y = deviation from the mean = $y-\bar{y}$, n = number of values of the two variables, σ_x = S.D. of $x-series$, σ_y = S.D. of $y-series$.

Substituting the value of σ_x and σ_y in the above formula, we get

$$r=\frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

Another form of the above formula is

$$r=\frac{n\Sigma xy-\Sigma x\Sigma y}{\sqrt{\left\{n\Sigma x^2-(\Sigma x)^2\right\}\times\left\{n\Sigma y^2-(\Sigma y)^2\right\}}}$$

Property of coefficient of correlation

The coefficient of correlation numerically does not exceed '1',

i.e., $-1\le r\le 1$.

Formula for correlation coefficient

The formula to compute coefficient of correlation is

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}.$$

Here,

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

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

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

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

$$\sigma_{x-y}^2 = \frac{\Sigma (x-y)^2}{n} - \left(\overline{x-y}\right)^2$$

$$\overline{(x-y)} = \frac{\Sigma (x-y)}{n}$$

Regression

It is an estimation of one independent variable in terms of the other.

Example: **The best fitting straight line of the form $y = a + bx$ *is called the regression of* y *on* x *and* $x = a + by$ *is called the regression of* x *on* y.**

Equation of the Regression lines

The regression line of *y on x* is

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

The regression line of *x on y* is

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

Here $r\frac{\sigma_y}{\sigma_x}$ and $r\frac{\sigma_x}{\sigma_y}$ are the regression coefficients. Their product is r^2

Here $r\dfrac{\sigma_y}{\sigma_x}$ and $r\dfrac{\sigma_x}{\sigma_y}$ are the regression coefficients. Their product is r^2

Thus 'r' is the geometric mean of regression coefficients,

$$\text{i.e., } r = \pm\sqrt{(\text{coefficient of } x)\times(\text{coefficient of } y)}$$

Note: If both the coefficients are positive then we have to consider *r value as* positive and if both coefficients are negative then we have to consider *r value as* negative.

Angle between the lines of regression

The angle between the lines of regression is

$$\tan\theta = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left(\frac{r^2-1}{r}\right)$$

Proof:

If m_1 and m_2 are the slopes of two lines then the angle between the lines is given by

$$\tan\theta = \left|\frac{m_1-m_2}{1+m_1m_2}\right| \qquad \text{----- (1)}$$

The lines of regression are

$$y-\bar{y}=r\frac{\sigma_y}{\sigma_x}\left(x-\bar{x}\right) \text{ and } x-\bar{x}=r\frac{\sigma_x}{\sigma_y}\left(y-\bar{y}\right)$$

$$\Rightarrow \quad y-\bar{y}=r\frac{\sigma_y}{\sigma_x}\left(x-\bar{x}\right) \text{ and } y-\bar{y}=\frac{\sigma_y}{r\sigma_x}\left(x-\bar{x}\right)$$

The slopes of regression lines are $m_1=\frac{r\sigma_y}{\sigma_x}$ and $m_2=\frac{\sigma_y}{r\sigma_x}$

Using these in (1), we get

$$\tan\theta=\left|\frac{\left(\frac{r\sigma_y}{\sigma_x}\right)-\left(\frac{\sigma_y}{r\sigma_x}\right)}{1+\left(\frac{r\sigma_y}{\sigma_x}\right)\left(\frac{\sigma_y}{r\sigma_x}\right)}\right| =\left|\frac{\frac{\sigma_y}{\sigma_x}\left(r-\frac{1}{r}\right)}{1+\frac{\sigma_y^2}{\sigma_x^2}}\right| =\left|\frac{\frac{\sigma_y}{\sigma_x}\left(\frac{r^2-1}{r}\right)}{\left(\frac{\sigma_x^2+\sigma_y^2}{\sigma_x^2}\right)}\right|$$

$$= \left| \frac{\dfrac{\sigma_y}{\sigma_x}\left(\dfrac{r^2-1}{r}\right)}{\left(\dfrac{\sigma_x^2+\sigma_y^2}{\sigma_x^2}\right)} \right| \quad = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left(\frac{r^2-1}{r}\right)$$

$\therefore$ The angle between the lines of regression is

$$\tan\theta = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left(\frac{r^2-1}{r}\right)$$

Its significance when $r=0$ and $r=\pm 1$

Case – 1: If $r = 0$ then $\tan\theta = \infty$

$$\Rightarrow \qquad \theta = \frac{\pi}{2}$$

$\therefore$ The lines are perpendicular

Case – 2: If $r = \pm 1$ then $\tan\theta = 0$

$$\Rightarrow \qquad \theta = 0$$

$\therefore$ The lines are parallel

Example 1: Find the coefficient of correlation and the lines of regression for the data

x	1	2	3	4	5
y	2	5	3	8	7

We have, $$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$$

We prepare a relevant table as follows

	x	y	$x-y$	x^2	y^2	$(x-y)^2$
	1	2	–1	1	4	1
	2	5	–3	4	25	9
	3	3	0	9	9	0
	4	8	–4	16	64	16
	5	7	–2	25	49	4
Σ	15	25	–10	55	151	30

Here, $n = 5$, $\Sigma x = 15$, $\Sigma y = 25$, $\Sigma(x-y) = -10$, $\Sigma x^2 = 55$, $\Sigma y^2 = 151$ and $\Sigma(x-y)^2 = 30$.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5} = 3 \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{25}{5} = 5$$

$$\overline{(x-y)} = \frac{\Sigma(x-y)}{n} = -\frac{10}{5} = -2$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - \left(\bar{x}\right)^2 = \frac{55}{5} - (3)^2 = 2$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - \left(\bar{y}\right)^2 = \frac{151}{5} - (5)^2 = 5.2$$

$$\sigma_{x-y}^2 = \frac{\Sigma(x-y)^2}{n} - \left(\overline{x-y}\right)^2 = \frac{30}{5} - (-2)^2 = 2$$

Substituting these values in r, we get

$$r = \frac{2+5.2-2}{2\left(\sqrt{2}\right)\left(\sqrt{5.2}\right)} = 0.81$$

Also, we have the equations of the regression lines are

$$y-\bar{y} = r\frac{\sigma_y}{\sigma_x}\left(x-\bar{x}\right) \text{ and } x-\bar{x} = r\frac{\sigma_x}{\sigma_y}\left(y-\bar{y}\right)$$

$$y-5 = (0.81)\frac{\sqrt{5.2}}{\sqrt{2}}(x-3) \text{ and } x-3 = (0.81)\frac{\sqrt{2}}{\sqrt{5.2}}(y-5)$$

$$y-5 = 1.306(x-3) \text{ and } x-3 = 0.502(y-5)$$

$$y = 1.306x+1.082 \text{ and } x = 0.502y+0.49$$

These are the lines of regression.

Example 2: Find the coefficient of correlation and the lines of regression for the data

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

We have, $$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}$$

We prepare a relevant table as follows

	x	y	$x-y$	x^2	y^2	$(x-y)^2$
	1	9	−8	1	81	64
	2	8	−6	4	64	36
	3	10	−7	9	100	49
	4	12	−8	16	144	64
	5	11	−6	25	121	36
	6	13	−7	36	169	49
	7	14	−7	49	196	49
Σ	28	77	−49	140	875	347

Here, $n = 7$, $\Sigma x = 28$, $\Sigma y = 77$, $\Sigma(x - y) = -49$, $\Sigma x^2 = 140$, $\Sigma y^2 = 875$ and $\Sigma(x - y)^2 = 347$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{28}{7} = 4 \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{77}{7} = 11$$

$$\overline{(x-y)} = \frac{\Sigma(x-y)}{n} = -\frac{49}{7} = -7$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - \left(\bar{x}\right)^2 = \frac{140}{7} - (4)^2 = 4$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - \left(\bar{y}\right)^2 = \frac{875}{7} - (11)^2 = 4$$

$$\sigma_{x-y}^2 = \frac{\Sigma(x-y)^2}{n} - \left(\overline{x-y}\right)^2 = \frac{347}{7} - (-7)^2 = 0.57$$

Substituting these values in r, we get

$$r = \frac{4+4-0.57}{2(2)(2)} = 0.93$$

Also, we have the equations of the regression lines are

$$y-\bar{y} = r\frac{\sigma_y}{\sigma_x}\left(x-\bar{x}\right) \text{ and } x-\bar{x} = r\frac{\sigma_x}{\sigma_y}\left(y-\bar{y}\right)$$

$$y-11 = (0.93)\frac{2}{2}(x-4) \text{ and } x-4 = (0.93)\frac{2}{2}(y-11)$$

$$y-11 = 0.93\left(x-4\right) \text{ and } x-4 = 0.93\left(y-11\right)$$

$$y = 0.93x+7.28 \text{ and } x = 0.93y-6.23$$

These are the lines of regression.

Example 3: Find the coefficient of correlation and the lines of regression for the data

x	1	2	3	4	5	6	7	8	9	10
y	10	12	16	28	25	36	41	49	40	50

We have, $$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$$

We prepare a relevant table as follows

	x	y	$x-y$	x^2	y^2	$(x-y)^2$
	1	9	–9	1	100	81
	2	12	–10	4	144	100
	3	16	–13	9	256	169
	4	28	–24	16	784	576
	5	25	–20	25	625	400
	6	36	–30	36	1296	900
	7	41	–34	49	1681	1156
	8	49	–41	64	2401	1681
	9	40	–31	81	1600	961
	10	50	–40	100	2500	1600
Σ	55	307	–252	385	11387	7624

Here, $n = 10$, $\Sigma x = 55$, $\Sigma y = 307$, $\Sigma(x - y) = -252$, $\Sigma x^2 = 385$, $\Sigma y^2 = 11387$ and $\Sigma(x - y)^2 = 7624$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{55}{10} = 5.5 \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{307}{10} = 30.7$$

$$\overline{(x - y)} = \frac{\Sigma(x - y)}{n} = -\frac{252}{10} = -25.2$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - \left(\bar{x}\right)^2 = \frac{385}{10} - (5.5)^2 = 8.25$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - \left(\bar{y}\right)^2 = \frac{11387}{10} - (30.7)^2 = 196.21$$

$$\sigma_{x-y}^2 = \frac{\Sigma(x - y)^2}{n} - \left(\overline{x - y}\right)^2 = \frac{7624}{10} - (-25.2)^2 = 127.36$$

Substituting these values in r, we get

$$r = \frac{8.25 + 196.21 - 127.36}{2\left(\sqrt{8.25}\right)\left(\sqrt{196.21}\right)} = 0.96$$

Also, we have the equations of the regression lines are

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}\left(x - \bar{x}\right) \text{ and } x - \bar{x} = r\frac{\sigma_x}{\sigma_y}\left(y - \bar{y}\right)$$

$$y - 30.7 = (0.96)\frac{\sqrt{196.21}}{\sqrt{8.25}}(x - 5.5) \text{ and } x - 5.5 = (0.96)\frac{\sqrt{8.25}}{\sqrt{196.21}}(y - 30.7)$$

$$y = 4.686x + 4.927 \text{ and } x = 0.197y - 0.548$$

These are the lines of regression.

Example Find the lines of regression for the following data

Ages of cars (in years)	2	4	6	7	8	10	12
Annual Maintenance cost (in hundreds)	16	15	18	19	17	21	20

Hence estimate the maintenance cost if the age of a car is 9 years and the age of a car if the maintenance cost is Rs.1550/-

We prepare a relevant table as follows

	x	y	$x-y$	x^2	y^2	$(x-y)^2$
	2	16	–14	4	256	196
	4	15	–11	16	225	121
	6	18	–12	36	324	144
	7	19	–12	49	361	144
	8	17	–9	64	289	81
	10	21	–11	100	441	121
	12	20	–8	144	400	64
Σ	49	126	–77	413	2296	871

Here, $n = 7$, $\Sigma x = 49$, $\Sigma y = 126$, $\Sigma(x - y) = -77$, $\Sigma x^2 = 413$, $\Sigma y^2 = 2296$ and $\Sigma(x - y)^2 = 871$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{49}{7} = 7$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{126}{7} = 18$$

$$\overline{(x - y)} = \frac{\Sigma(x - y)}{n} = -\frac{77}{7} = -11$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{413}{7} - (7)^2 = 10$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{2296}{7} - (18)^2 = 4$$

$$\sigma_{x-y}^2 = \frac{\Sigma(x - y)^2}{n} - \left(\overline{x - y}\right)^2 = \frac{871}{7} - (-11)^2 = 3.429$$

Substituting these values in r, we get

$$r = \frac{10+4-3.429}{2\left(\sqrt{10}\right)\left(\sqrt{4}\right)} = 0.8357$$

Also, we have the equations of the regression lines are

$$y-\bar{y} = r\frac{\sigma_y}{\sigma_x}\left(x-\bar{x}\right) \text{ and } x-\bar{x} = r\frac{\sigma_x}{\sigma_y}\left(y-\bar{y}\right)$$

$$y-18 = (0.8357)\frac{2}{\sqrt{10}}(x-7) \text{ and } x-7 = (0.8357)\frac{\sqrt{10}}{2}(y-18)$$

$$y = 0.52857x + 14.3 \text{ and } x = 1.321429y - 16.786$$

These are the lines of regression.

Also,

(i) If the age of the car $x = 9$ then the maintenance cost y is given by the line of regression of y on x is

$$y = 0.52857\,(9) + 14.3 = 19.06 \text{ hundreds.}$$

$\therefore$ The maintenance cost = Rs.1906/-

(ii) If the maintenance cost y is Rs.1550/- i.e., 15.5 hundreds then the age of the car x is given by the line of regression of x on y is

$$x = 1.321429\,(15.5) - 16.786 = 3.7 \text{ years}$$

$\therefore$ Age of car = 3.7 years.

Example : In the following table are recorded data showing the test scores made by salesmen on an intelligence test and their weekly sales:

Salesmen	1	2	3	4	5	6	7	8	9	10
Test scores	40	70	50	60	80	50	90	40	60	60
Sales	2.5	6.0	4.5	5.0	4.5	2.0	5.5	3.0	4.5	3.0

Calculate the regression line of sales on test scores and estimate the most probable weekly sales volume if a salesman makes a score of 70.

Example In a partially destroyed laboratory record of correlation data, the following results only are available:
Variance of x is 9 and the regression equations are $4x - 5y + 33 = 0$ and $20x - 9y = 107$. Calculate (i) the mean values of x and y, (ii) standard deviation of y and (iii) the coefficient of correlation between x and y.

We know that the regression line passes through $(\bar{x}, \bar{y})$

$$\therefore \quad 4\bar{x} - 5\bar{y} = -33$$

$$20\bar{x} - 9\bar{y} = 107$$

Solving these equations, we get

$$\bar{x} = 13 \quad \text{and} \quad \bar{y} = 17$$

We rewrite the given regression equations as

$$y = \frac{4}{5}x + \frac{33}{5} \quad \text{and} \quad x = \frac{9}{20}y + \frac{107}{20} \qquad \text{------ } (1)$$

We have, the coefficient of correlation between x and y is

$$r = \pm\sqrt{(\text{coefficient of } x)(\text{coefficient of } y)}$$

$$r = \pm\sqrt{\left(\frac{4}{5}\right)\left(\frac{9}{20}\right)} = \pm\frac{3}{5}$$

$$r = \frac{3}{5} \quad (\because \text{ both the coefficients are positive})$$

Given that, $\sigma_x^2 = 9 \Rightarrow \sigma_x = 3$

We have, $$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}\left(x - \bar{x}\right) \qquad \text{----- (2)}$$

Comparing (2) with (1),

we have $$r\frac{\sigma_y}{\sigma_x} = \frac{4}{5}$$

we have $r\dfrac{\sigma_y}{\sigma_x} = \dfrac{4}{5}$

$$\sigma_y = \frac{4}{5} \times \frac{\sigma_x}{r}$$

$$= \frac{4}{5} \times \frac{3}{(3/5)} = 4$$

Example **Compute $\bar{x}, \bar{y}$ and r from the equations of the regression lines $2x+3y+1=0$ and $x+6y=4$.**

We know that the regression line passes through $(\bar{x}, \bar{y})$

$$\therefore \quad 2\bar{x}+3\bar{y}=-1$$

$$\bar{x}+6\bar{y}=4$$

Solving these equations, we get

$$\bar{x}=-2 \text{ and } \bar{y}=1$$

We rewrite the given regression equations as

$$y=-\frac{2}{3}x-\frac{1}{3} \quad \text{and} \quad x=-6y+4$$

We have, the coefficient of correlation between x and y is

$$r=\pm\sqrt{(\text{coefficient of } x)(\text{coefficient of } y)}$$

$$r = \pm\sqrt{\left(-\frac{2}{3}\right)(-6)} \quad = \pm\sqrt{4} = \pm 2$$

which is not possible because $-1 \le r \le 1$.

Hence, we rewrite the given regression equations in other form as

$$x = -\frac{3}{2}y - \frac{1}{2} \quad \text{and} \quad y = -\frac{1}{6}x + \frac{2}{3}$$

We have, the coefficient of correlation between x and y is

$$r = \pm\sqrt{(\text{coefficient of } x)(\text{coefficient of } y)}$$

$$r = \pm\sqrt{\left(-\frac{3}{2}\right)\left(-\frac{1}{6}\right)} = \pm\frac{1}{2}$$

$$r = -\frac{1}{2} \quad (\because \text{both the coefficients are negative})$$

Example If $2x - 3y = 0$ and $3x - 2y = 5$ are the lines of regression of the variables x and y. Find the following:

(i) Mean of x and y

(ii) Coefficient of correlation between x and y.

(iii) Angle between the lines of regression.

(iv) Standard deviation of x when variance of y is 2.

We know that the regression line passes through $(\bar{x}, \bar{y})$

$$\therefore \quad 2\bar{x} - 3\bar{y} = 0$$

$$3\bar{x} - 2\bar{y} = 5$$

Solving these equations, we get

$$\bar{x} = 3 \text{ and } \bar{y} = 2$$

We rewrite the given regression equations as

$$y = \frac{2}{3}x \quad \text{and} \quad x = \frac{2}{3}y + 5 \qquad \text{-----} (1)$$

We have, the coefficient of correlation between x and y is

$$r = \pm\sqrt{(\text{coefficient of } x)(\text{coefficient of } y)}$$

$$r = \pm\sqrt{\left(\frac{2}{3}\right)\left(\frac{2}{3}\right)} = \pm\frac{2}{3}$$

$$r = \frac{2}{3} \quad (\because \text{both the coefficients are positive})$$

The angle between the lines of regression is $\tan\theta = \left|\frac{m_1 - m_2}{1 + m_1 m_2}\right|$

Now, $m_1 = -\dfrac{\text{coefficient of } x}{\text{coefficient of } y} = -\dfrac{2}{-3} = \dfrac{2}{3}$ and

$$m_2 = -\frac{\text{coefficient of } x}{\text{coefficient of } y} = -\frac{3}{-2} = \frac{3}{2}$$

$$\therefore \quad \tan\theta = \left|\frac{(2/3)-(3/2)}{1+(2/3)(3/2)}\right| = \left|\frac{-(5/6)}{2}\right| = \frac{5}{12}$$

$$\Rightarrow \qquad \theta = \tan^{-1}\left(\frac{5}{12}\right)$$

Given that, $\sigma_y^2 = 2 \quad \Rightarrow \sigma_y = \sqrt{2}$

We have, $y - \bar{y} = r\frac{\sigma_y}{\sigma_x}\left(x - \bar{x}\right)$ ----- (2)

Comparing (2) with (1),

we have $r\frac{\sigma_y}{\sigma_x} = \frac{2}{3}$

we have $r\dfrac{\sigma_y}{\sigma_x} = \dfrac{2}{3}$

$$\left(\frac{2}{3}\right)\frac{\sqrt{2}}{\sigma_x} = \frac{2}{3}$$

$$\sigma_x = \sqrt{2}$$

Example While calculating correlation coefficient between two variables x and y from 25 pairs of observations, the following results were obtained: $n = 25, \Sigma x = 125, \Sigma y = 100, \Sigma x^2 = 650, \Sigma y^2 = 460, \Sigma xy = 508$. Later it was discovered at the time of checking that the pairs of values

x	y
8	12
6	8

were copied down as

x	y
6	14
8	6

Obtain the correct value of correlation coefficient.

To get the correct results, we subtract the incorrect values and add the corresponding correct values.

$\therefore$ The correct results are

$$n = 25,$$
$$\Sigma x = 125 - 6 - 8 + 8 + 6 = 125,$$
$$\Sigma y = 100 - 14 - 6 + 12 + 8 = 100,$$
$$\Sigma x^2 = 650 - 6^2 - 8^2 + 8^2 + 6^2 = 650,$$
$$\Sigma y^2 = 460 - 14^2 - 6^2 + 12^2 + 8^2 = 436,$$
$$\Sigma xy = 508 - 6\times 14 - 8\times 6 + 8\times 12 + 6\times 8 = 520$$

We have, $$r = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{\left\{n\Sigma x^2 - (\Sigma x)^2\right\}\times\left\{n\Sigma y^2 - (\Sigma y)^2\right\}}}$$

$$\Rightarrow \quad r = \frac{25\times 520 - 125\times 100}{\sqrt{\left\{25\times 650 - (125)^2\right\}\times\left\{25\times 436 - (100)^2\right\}}} = \frac{20}{\sqrt{25\times 36}} = \frac{2}{3}$$

EXERCISE

1. Find two lines of regression and coefficient of correlation for the data
   $n = 18, \Sigma x = 12, \Sigma y = 18, \ \Sigma x^2 = 6, \ \Sigma y^2 = 96, \ \Sigma xy = 48.$

2. Find the coefficient of correlation and the lines of regression for the data

x	2	4	6	8	10
y	5	7	9	8	11

3. Find the coefficient of correlation from the following data

x	78	89	97	69	59	79	68	57
y	125	137	156	112	107	138	123	108

4. Find the coefficient of correlation for the data

x	21	23	30	54	57	58	72	78	87	90
y	60	71	72	83	110	84	100	92	113	135

5. If the coefficient of correlation between two variables x and y is 0.5 and the acute angle between their lines of regression is $\tan^{-1}\left(\frac{3}{8}\right)$, show that

$\sigma_x = \frac{1}{2}\sigma_y$. **(VTU 2004)**

6. Two random variables have the regression lines with equations $3x+2y=26$ and $6x+y=31$. Find the mean values and the correlation coefficient between x and y.
7. The regression equations of two variables x and y are $x=0.7y+5.2$ and $y=0.3x+2.8$. Find the means of the variables and the coefficient of correlation between them.
8. In a partially destroyed laboratory data, only the equations giving the two lines of regression of y on x and x on y are available and are respectively, $7x-16y+9=0$ and $5y-4x-3=0$. Calculate the coefficient of correlation and mean values of x and y.
9. The two regression equations of the variables x and y are $x=19.13-0.87y$ and $y=11.64-0.50x$. Find the mean values of x and y and the correlation coefficient between x and y. **(VTU 2004)**
10. Psychological tests of intelligence and of engineering ability were applied to 10 students. Here is a record of ungrouped data showing intelligence ratio (I.R) and engineering ratio (E.R). Calculate the coefficient of correlation.

Student	A	B	C	D	E	F	G	H	I	J
I.R	105	104	102	101	100	99	98	96	93	92
E.R	101	103	100	98	95	96	104	92	97	94