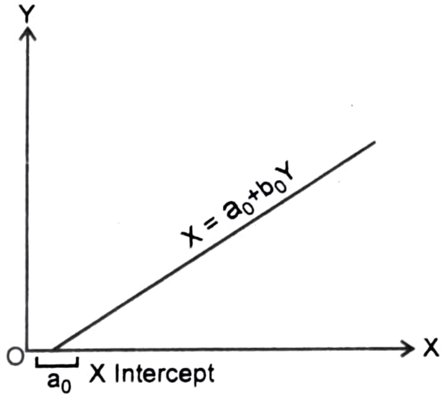
$$\Sigma XY = a_0 \Sigma Y + b_0 \Sigma Y^2$$

Solving these equations, we get the following value of a_0 and b_0.

$$b_0 = b_{xy} = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2}$$

$$a_0 = \overline{X} - b_0 \overline{Y}$$



Finally, the calculated value of a_0 and b_0 are put in the equation $X = a_0 + b_0 Y$. The regression equation of X on Y will be used to estimate the value of X when the value of Y is given.

Note: a_0 is the X-intercept, which indicates the minimum value of X for Y = 0 and **b_0 is the slope of the line or called regression coefficient of X on Y.**

The following examples makes the above said method more clear:

Example 1. Calculate the regression equation of X on Y from the following data by the method of least square:

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

Solution: **Calculation of Regression Equation**

X	X^2	**Y**	Y^2	**XY**
1	1	2	4	2
2	4	5	25	10
3	9	3	9	9
4	16	8	64	32
5	25	7	49	35
$N = 5,\ \Sigma X = 15$	$\Sigma X^2 = 55$	$\Sigma Y = 25$	$\Sigma Y^2 = 151$	$\Sigma XY = 88$

Regression Equation of X on Y is

$$X = a + bY$$

X = a + b Y

The two normal equations are:

$$\Sigma X = Na + b\,\Sigma Y$$

$$\Sigma XY = a\,\Sigma Y + b\,\Sigma Y^2$$

ΣX = Na + bΣY

ΣXY = aΣY + (ΣY)²

Substituting the values, we get

$$15 = 5a + 25b \quad ...(i)$$

$$88 = 25a + 151b \quad ...(ii)$$

Multiplying (i) by 5 and subtracting it from (ii)

$$88 = 25a + 151b$$

$$\underline{\underset{-}{75} = \underset{-}{25a} + \underset{-}{125b}}$$

$$13 = 26b$$

$$\therefore \quad b = \frac{13}{26} = 0.5$$

Putting the value of b in equation (i)

$$15 = 5a + 25 \times 0.5$$
$$15 = 5a + 12.5$$
$$5a = 2.5$$
$$a = 0.50$$

$\therefore$ $\boldsymbol{X = 0.5 + 0.5Y}$

Aliter: The value of a and b can also be obtained by using the following formula:

$$b_{xy} = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2} \qquad a = \bar{X} - b\bar{Y}$$

Substituting the values, we get

$$b_{xy} = \frac{5 \times 88 - (15)(25)}{5 \times 151 - (25)^2} = \frac{440 - 375}{755 - 625} = \frac{65}{130} = \frac{1}{2} = 0.5$$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{15}{5} = 3, \bar{Y} = \frac{\Sigma Y}{N} = \frac{25}{5} = 5$$

$$\therefore \quad a = \bar{X} - b\bar{Y} = 3 - \frac{1}{2} \times 5 = 3 - 2.5 = 0.5$$

$$\therefore \quad X = 0.5 + 0.5Y$$

Example 2. Obtain the regression equation of Y on X by the least square method for the following data:

X:	1	2	3	4	5
Y:	9	9	10	12	11

Also estimate the value of Y when X = 10.

Solution: **Calculation of Regression Equation of Y on X**

X	**Y**	**XY**	**X^2**
1	9	9	1
2	9	18	4
3	10	30	9
4	12	48	16
5	11	55	25
$N = 5$, $\Sigma X = 15$	$\Sigma Y = 51$	$\Sigma XY = 160$	$\Sigma X^2 = 55$

Regression Equation of Y on X is

$$\boxed{Y = a + bX}$$

The two normal equations are

$$\Sigma Y = Na + b\,\Sigma X$$

$$\Sigma XY = a\,\Sigma X + b\,\Sigma X^2$$

Substituting the values, we get

$$51 = 5a + 15b \quad ...(i)$$

$$160 = 15a + 55b \quad ...(ii)$$

Multiplying (*i*) by 3 and subtracting it from (*ii*)

$$160 = 15a + 55b$$

$$153 = 15a + 45b$$

$$- \quad - \quad -$$

$$7 = 10b$$

$$\therefore \quad b = \frac{7}{10} = 0.7$$

Putting the value of b in equation (*i*)

$$51 = 5a + 15(0.7) = 5a + 10.5$$

$$5a = 40.5$$

$$a = 8.1$$

Hence, the required regression equation of Y on X is given by

$$Y = 8.1 + 0.7X$$

Estimation for Y

For $X = 10$, $Y = 8.1 + 0.7(10) = 15.1$

Example 3. Given the following data:

$N = 8, \Sigma X = 21, \Sigma X^2 = 99, \Sigma Y = 4, \Sigma Y^2 = 68, \Sigma XY = 36$

Using the values, find

(*i*) Regression equation of Y on X.

(*ii*) Regression equation of X on Y.

(*iii*) Most approximate value of Y for X = 10

(*iv*) Most approximate value of X for Y = 2.5

Solution: **(*i*) Regression Equation of Y on X**

$$Y = a + bX$$

$$byx = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2} = \frac{8 \times 36 - (21)(4)}{8 \times 99 - (21)^2} = 0.581$$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{21}{8} = 2.625, \quad \overline{Y} = \frac{\Sigma Y}{N} = \frac{4}{8} = 0.5$$

$$\therefore \quad a = \overline{Y} - b\overline{X} = 0.5 - (0.581)(2.625) = -1.025$$

$$\therefore \quad Y = -1.025 + 0.581X$$

(*ii*) Regression Equation of X on Y

$$X = a_0 + b_0 Y$$

$$b_0 = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2} = \frac{8 \times 36 - (21)(4)}{8 \times 68 - (4)^2} = 0.386$$

$$a_0 = \overline{X} - b_0 \overline{Y} = 2.625 - (0.386)(0.5) = 2.432$$

$$\therefore \quad X = 2.432 + 0.386Y$$

(*iii*) Prediction for Y

When $X = 10, Y = -1.025 + 0.581(10) = 4.785$

(*iv*) Prediction for X

When $Y = 2.5, X = 2.432 + 0.386(2.5) = 3.397$

EXERCISE 2.1

1. Obtain the line of regression of Y on X by least square method for the following data:

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

Also obtain an estimate of Y when $X = 2$. [**Ans.** $Y = 1.3 + 0.9X$; 3.1]

2. Find the regression of Y on X and X on Y by the least square method for the following data:

X:	1	2	3
Y:	2	4	5

Also find coefficient of correlation. [**Ans.** Y= 0.667+1.5X; X=–0.357+ 0.643Y; $r = 0.982$]

3. Compute the appropriate regression for the following data:

X (Independent variable):	1	3	4	8	9	11	14
Y (Dependent variable):	1	2	4	5	7	8	9

[**Ans.** $Y = 0.63X + 0.64$]

4. Obtain the two lines of regression from the following data:
$N = 3, \Sigma X = 6, \Sigma X^2 = 14, \Sigma Y = 15, \Sigma Y^2 = 77, \Sigma XY = 31$

[**Ans.** $Y = 0.5X + 4$, $X = 0.5Y - 0.5$]

5. Given: $\Sigma X = 15, \Sigma Y = 110, \Sigma XY = 400, \Sigma X^2 = 250, \Sigma Y^2 = 3200, N = 10$
Find the following:
(*i*) Regression coefficient of Y on X and the Y-intercept.
(*ii*) X-intercept, and the regression coefficient of X on Y.
(*iii*) Most approximate value of Y for X = 5.
(*iv*) Most approximate value of X for Y = 25.
[**Ans.** (*i*) $b = 1.033$, $a = 9.451$, (*ii*) $a = 0.201$, $b = 0.118$, (*iii*) $Y = 14.616, X = 3.151$]

▶ (2) Regression Equations using Regression Coefficients

Regression equations can also be computed with the help of regression coefficients. For this, we will have to find out $\overline{X}$, $\overline{Y}$, byx and bxy from the given data. Regression equations can be computed from the regression coefficients by any of the following methods:

(1) Using the actual values of X and Y series.

(2) Using deviations from Actual Means.

(3) Using deviations from Assumed Means.

(4) Using r, σ_x, σ_y and $\overline{X}$, $\overline{Y}$.

(1) Using the Actual Values of X and Y Series

In this method, actual values of X and Y are used to determine regression equations. With regard to regression coefficients, regression equations are put in the following way:

Regression Equation of Y on X

$$Y - \overline{Y} = byx\,(X - \overline{X})$$

or $$Y = \overline{Y} + byx\,(X - \overline{X})$$

Here, $\overline{X}$ = Arithmetic mean of X series $= \dfrac{\Sigma X}{N}$

$\overline{Y}$ = Arithmetic mean of Y series $= \dfrac{\Sigma Y}{N}$

byx = Regression coefficient of Y on X

Using actual values, the value of byx can be calculated as:

$$byx = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2} \quad \text{or} \quad byx = \frac{\Sigma XY / N - \overline{X}.\overline{Y}}{\sigma_x^2}$$

Note: This formula is based on the normal equations, yet its use avoids the solution of normal equations.

Regression Equation of X on Y

Stand. error

$S_{yx} = \sigma_y \sqrt{1-r^2}$

$$X - \overline{X} = bxy\,(Y - \overline{Y})$$

or $$X = \overline{X} + bxy(Y - \overline{Y})$$

$S_{xy} = \sigma_x \sqrt{1-r^2}$

Where bxy= Regression coefficient of X on Y.

Using actual values, the value of bxy can be calculated as:

$$bxy = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2} \quad \text{or} \quad bxy = \frac{\Sigma XY / N - \overline{X}.\overline{Y}}{\sigma_y^2} = \frac{Cov\,(X_1Y)}{\sigma_y^2}$$

The following examples make this method more clear:

Example 4. Calculate the regression equations of X on Y and Y on X from the following data:

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

Solution:

Calculation of Regression Equations

X	X^2	Y	Y^2	XY
1	1	2	4	2
2	4	5	25	10
3	9	3	9	9
4	16	8	64	32
5	25	7	49	35
N=5, ΣX=15	ΣX^2=55	ΣY=25	ΣY^2=151	ΣXY=88

$$\overline{X} = \frac{\Sigma X}{N} = \frac{15}{5} = 3, \quad \overline{Y} = \frac{\Sigma Y}{N} = \frac{25}{5} = 5$$

Regression Coefficient of Y on X (by*x*):

$$byx = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{5 \times 88 - (15)(25)}{5 \times 55 - (15)^2} = \frac{440 - 375}{275 - 225} = \frac{65}{50} = 1.3$$

Regression Coefficient of X on Y (bxy):

$$bxy = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2}$$

$$= \frac{5 \times 88 - (15)(25)}{5 \times 151 - (25)^2} = \frac{440 - 375}{755 - 625} = \frac{65}{130} = +0.5$$

Regression Equation of Y on X

$$Y - \overline{Y} = byx(X - \overline{X})$$

Substituting the values,

$$Y - 5 = 1.3(X - 3)$$
$$Y - 5 = 1.3X - 3.9$$
$$Y = 1.3X - 3.9 + 5$$
$$Y = 1.3X + 1.1$$

Regression Equation of X on Y

$$X - \overline{X} = bxy(Y - \overline{Y})$$
$$X - 3 = +0.5(Y - 5)$$
$$X - 3 = 0.5Y - 2.5$$
$$X = 0.5Y + 0.5$$

Example 5. Calculate the two regression equations from the following data:

$$\Sigma X = 30, \Sigma Y = 23, \Sigma XY = 168, \Sigma X^2 = 224, \Sigma Y^2 = 175, N = 7$$

Hence or otherwise find Karl Pearson's coefficient of correlation.

30

Solution:
$$\overline{X} = \frac{\Sigma X}{N} = \frac{30}{7} = 4.286$$

$$\overline{Y} == \frac{\Sigma Y}{N} = \frac{23}{7} = 3.286$$

$$byx = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2} = \frac{7 \times 168 - (30)(23)}{7 \times 224 - (30)^2} = 0.728$$

$$bxy = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2} = \frac{7 \times 168 - (30)(23)}{7 \times 175 - (23)^2} = 0.698$$

Regression Equation of Y on X

$$Y - \overline{Y} = byx\,(X - \overline{X})$$
$$Y - 3.286 = 0.728\,(X - 4.286)$$
$$Y - 3.286 = 0.728X - 3.120$$
$$\therefore \quad Y = 0.728X + 0.166$$

Regression Equation of X on Y

$$X - \overline{X} = bxy\,(Y - \overline{Y})$$
$$X - 4.286 = 0.698(Y - 3.286)$$
$$X - 4.286 = 0.698Y - 2.294$$
$$\therefore \quad X = 0.698Y + 1.992$$

Karl Pearson's Coefficient of Correlation

$$r = \sqrt{byx \cdot bxy}$$
$$r = \sqrt{0.728 \times 0.698} = 0.712$$

IMPORTANT TYPICAL EXAMPLES

Example 6. In order to find the correlation coefficient between the two variables X and Y from 12 pairs of observations, the following calculations were made:

$\Sigma X = 30, \Sigma X^2 = 670, \Sigma Y = 5, \Sigma Y^2 = 285, \Sigma XY = 344$

On subsequent verifications, it was discovered that the pair (X = 11, Y = 4) was copied wrongly, the correct values being (X = 10, Y = 14). After making necessary corrections, find:

(*i*) the two regression coefficients.

(*ii*) the two regresssion equations.

(*iii*) the correlation coefficient.

Solution: Corrected ΣX = 30 + Correct value – Incorrect value

= 30 + 10 – 11 = 29

Corrected ΣY = 5 + 14 – 4 = 15

$$\text{Corrected } \Sigma X^2 = 670 + (\text{Correct value})^2 – (\text{Incorrect value})^2$$
$$= 670 + 10^2 - 11^2 = 649$$

$$\text{Corrected } \Sigma Y^2 = 285 + 14^2 - 4^2 = 465$$

$$\text{Corrected } \Sigma XY = 344 + (10)\,(14) - (11)\,(4) = 440$$

$$\overline{X} = \frac{Corrected\ \Sigma X}{N} = \frac{29}{12} = 2.416$$

$$\overline{Y} = \frac{Corrected\ \Sigma Y}{N} = \frac{15}{12} = 1.25$$

(*i*) Regression Coefficients

$$byx = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2} = \frac{12 \times 440 - 29 \times 15}{12 \times 649 - (29)^2}$$

$$= \frac{5280 - 435}{7788 - 841} = \frac{4845}{6947} = +0.697$$

$$bxy = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2} = \frac{12 \times 440 - 29 \times 15}{12 \times 465 - (15)^2}$$

$$= \frac{5280 - 435}{5580 - 225} = \frac{4845}{5355} = 0.904$$

(*ii*) Two Regression Equations

X on Y

$$X - \overline{X} = bxy\,(Y - \overline{Y})$$
$$X - 2.416 = 0.904\,(Y - 1.25)$$
$$X - 2.416 = 0.904Y - 1.13$$
$$X = 0.904Y + 1.286$$

Y on X

$$Y - \overline{Y} = byx\,(X - \overline{X})$$
$$Y - 1.25 = 0.697\,(X - 2.416)$$
$$Y - 1.25 = 0.697X - 1.683$$
$$Y = 0.697X - 0.433$$

(*iii*) Correlation coefficient

$$r = \sqrt{byx\,.\,bxy} = \sqrt{(0.697)(0.904)} = +0.793$$

Example 7. Given that

$\overline{X} = 15,\ \overline{Y} = 12, \Sigma XY = 1500, \sigma_x = 6.4, \sigma_y = 9.0, N = 10$

Compute: (a) Two regression Coefficients

(b) Correlation coefficient between X and Y.

Solution: Given: $\overline{X} = 15,\ \overline{Y} = 12, \Sigma XY = 1500, \sigma_x = 6.4, \sigma_y = 9.0, N = 10$

Regression Coefficient of Y on X

$$byx = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2}$$

The values of N and ΣXY are given and the values of $\Sigma X^2, \Sigma Y^2, \Sigma X$ and ΣY are to be calculated as follows:

$$\overline{X} = \frac{\Sigma X}{N} \Rightarrow \Sigma X = N.\overline{X} = 10 \times 15 = 150$$

$$\overline{Y} = \frac{\Sigma Y}{N} \Rightarrow \Sigma Y = N.\overline{Y} = 10 \times 12 = 120$$

$$\Sigma X^2 = N[\sigma_x^2 + (\overline{X})^2] = 10[6.4^2 + 15^2] = 2659.6$$

$$\Sigma Y^2 = N[\sigma_y^2 + (\overline{Y})^2] = 10[9^2 + 12^2] = 2250$$

Now,
$$byx = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2} = \frac{10 \times 1500 - (150)(120)}{10 \times 2659.6 - (150)^2}$$

$$= \frac{15000 - 18000}{26596 - 22500} = \frac{-3000}{4096} = -0.73$$

Aliter: byx can also be calculated as follows:

$$byx = \frac{\frac{\Sigma XY}{N} - \overline{X}.\overline{Y}}{\sigma_x^2} = \frac{\frac{1500}{10} - (15)(12)}{(6.4)^2}$$

$$= \frac{150 - 180}{40.96} = \frac{-30}{40.96} = -0.73$$

Regression Coefficient of X on Y

$$bxy = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2} = \frac{10 \times 1500 - (150)(120)}{10 \times 2250 - (120)^2}$$

$$= \frac{15000 - 18000}{22500 - 14400} = \frac{-3000}{8100} = -0.37$$

Aliter: bxy can also be calculated as follows:

$$bxy = \frac{\frac{\Sigma XY}{N} - \overline{X}.\overline{Y}}{\sigma_y^2} = \frac{\frac{1500}{10} - (15)(12)}{(9)^2}$$

$$= \frac{150 - 180}{81} = \frac{-30}{81} = -0.37$$

Coefficient of Correlation

$$r = \pm\sqrt{byx.bxy} = -\sqrt{(-0.73) \times (-0.37)} = -0.519$$

Example 8. Find out the regression coefficients of Y on X and X on Y from the following data: $\Sigma X = 50, \overline{X} = 5, \Sigma Y = 60, \overline{Y} = 6, \Sigma XY = 350$, Variance of X = 4, Variance of Y = 9.

Solution: We know that: $\overline{X}=\frac{\Sigma X}{N} \Rightarrow 5=\frac{50}{N} \Rightarrow N=10$

$$byx=\frac{N.\Sigma XY-\Sigma X.\Sigma Y}{N.\Sigma X^2-(\Sigma X)^2} \text{ or } \frac{\Sigma XY/N-\overline{X}.\overline{Y}}{\sigma_x^2} \text{ or } \frac{Cov(X_1Y)}{\sigma_x^2}$$

$$\therefore \quad byx=\frac{\frac{350}{10}-(5)(6)}{4}=\frac{35-30}{4}=\frac{5}{4}$$

$$=1.25$$

$$bxy=\frac{N.\Sigma XY-\Sigma X.\Sigma Y}{N.\Sigma Y^2-(\Sigma Y)^2} \text{ or } \frac{\Sigma XY/N-\overline{X}.\overline{Y}}{\sigma_y^2} \text{ or }=\frac{Cov(X_1Y)}{\sigma_y^2}$$

$$\therefore \quad bxy=\frac{\frac{350}{10}-(5)(6)}{9}=\frac{35-30}{9}=\frac{5}{9}$$

EXERCISE 2.2

1. Given the following bivariate data:

X:	–1	5	3	2	1	1	7	3
Y:	–6	1	0	0	1	2	1	5

(*i*) Fit a regression line of Y on X and predict Y if X=10.

(*ii*) Fit a regression line of X on Y and predict X if Y=2.5

[**Ans.** $Y=-1.025+0.581X$; $X=2.432+0.386Y$; $Y_{10}=4.785$; $X_{2.5}=3.397$]

2. By using the following data, find the regression equation of Y on X and compute the value of Y when X = 10.

$\overline{X}=5.5$, $\overline{Y}=4.0$, $\Sigma X^2=385$, $\Sigma Y^2=192$, $\Sigma(X+Y)^2=947$ and $N=10$

[**Ans.** $Y=-0.42X+6.31$, $Y_{10}=2.11$]

3. Given that:

$\Sigma X=250, \Sigma Y=300, \sigma_x=5, \sigma_y=10, \Sigma XY=7900, N=10$

Compute : (*i*) Two regression coefficients,

(*ii*) Correlation coefficient between X and Y.

(*iii*) Most approximate value of Y when X = 55 and X when Y = 40.

[**Ans.** $byx=1.6$, $bxy=0.4$, $r=0.8$, $Y_{30}=78$, $X_{40}=29$]

4. By using the following data, find correlation coefficient and regression equation of Y on X and estimated value of Y when X = 20

$N=10, \Sigma X=140, \Sigma Y=150, \Sigma(X-10)^2=180, \Sigma(Y-15)^2=215, \Sigma(X-10)(Y-15)=60$

[**Hint:** See Example 53 on Correlation]

[**Ans.** $r=0.915$, $Y=3X-27$, $Y_{20}=33$]

5. Following information was computed through a computer:

$\Sigma X = 125, \Sigma Y = 100, \Sigma X^2 = 650, \Sigma Y^2 = 460, \Sigma XY = 508, N = 25$

Later on it was discovered that two pairs of X and Y were miscopied as (6, 14) and (8, 6) instead of (8, 12) and (6, 8). Determine (*i*) the correct regression equations (*ii*) correct coefficient of correlation. [**Ans.** (*i*) $X = 0.556Y + 2.776$, $Y = 0.8X$, (*ii*) $r = 0.67$]

6. On each of 30 sets, two measurements are made. The following summaries are given :

$\Sigma X = 15, \Sigma Y = -6, \Sigma XY = 56, \Sigma X^2 = 61$ and $\Sigma Y^2 = 90$

Calculate the product moment correlation coefficient and the slope of regression line of Y on X. [**Hint:** See Example 52] [**Ans.** $r = 0.856$, $byx = 1.10$]

---

(2) Using Deviations taken from Actual Means

When the size of the values of X and Y is very large, then the method using actual values becomes very difficult to use. In such case, in place of actual values, deviations taken from arithmetic means ($\overline{X}$, $\overline{Y}$) are used to simplify the computation process. In such a case, regression equations are expressed as follows:

Regression Equation of Y on X

$$Y - \overline{Y} = byx\,(X - \overline{X})$$

or

$$Y = \overline{Y} + byx\,(X - \overline{X})$$

Here, $\overline{X}$ = Arithmetic mean of X

$\overline{Y}$ = Arithmetic mean of Y

byx = Regression coefficient of Y on X

Using deviations from actual means, the value of byx can be calculated as:

$$byx = \frac{\Sigma xy}{\Sigma x^2}$$

Where, $x = X - \overline{X}$; $y = Y - \overline{Y}$

Regression Equation of X on Y

$$X - \overline{X} = bxy\,(Y - \overline{Y})$$

or

$$X = \overline{X} + bxy(Y - \overline{Y})$$

Where, bxy = Regression coefficient of X on Y.

Using deviations from actual means, the value of bxycan be calculated as:

$$bxy = \frac{\Sigma xy}{\Sigma y^2}$$

Where, $x = X - \overline{X}$; $y = Y - \overline{Y}$

The following examples make this method more clear.

Example 9. Obtain the two regression equations from the following data:

X:	2	4	6	8	10	12
Y:	4	2	5	10	3	6

Solution:

X	$\bar{X}=7$ $(X-\bar{X})$ x	x^2	Y	$\bar{Y}=5$ $(Y-\bar{Y})$ y	y^2	xy
2	–5	25	4	–1	1	+5
4	–3	9	2	–3	9	+9
6	–1	1	5	0	0	0
8	+1	1	10	+5	25	+5
10	+3	9	3	–2	4	–6
12	+5	25	6	+1	1	+5
$\Sigma X=42$ N = 6	$\Sigma x=0$	$\Sigma x^2=70$	$\Sigma Y=30$	$\Sigma y=0$	$\Sigma y^2=40$	$\Sigma xy=18$

$$\bar{X}=\frac{\Sigma X}{N}=\frac{42}{6}=7;\quad \bar{Y}=\frac{\Sigma Y}{N}=\frac{30}{6}=5$$

Since, the actual means of X and Y are whole numbers, we should taken deviations from $\bar{X}$ and $\bar{Y}$ to simplify calculations:

$$byx=\frac{\Sigma xy}{\Sigma x^2}=\frac{18}{70}=0.257$$

$$bxy=\frac{\Sigma xy}{\Sigma y^2}=\frac{18}{40}=0.45$$

Regression Equation of Y on X

$$Y-\bar{Y}=byx\,(X-\bar{X})$$
$$Y-5=0.257\,(X-7)$$
$$Y-5=0.257X-1.799$$
$$\therefore\quad Y=0.257X+3.201$$

Regression Equation of X on Y

$$X-\bar{X}=bxy\,(Y-\bar{Y})$$
$$X-7=0.45(Y-5)$$
$$X-7=0.45Y-2.25$$
$$X=0.45Y-2.25+7$$
$$\therefore\quad X=0.45Y+4.75$$

Example 10. The following are the intermediate results of the two series X and Y:

$$\overline{X} = 90, \overline{Y} = 70, N = 10, \Sigma x^2 = 6360, \Sigma y^2 = 2860, \Sigma xy = 3900$$

(Where x and y are deviations from the respective means)

Find two regression equations.

Solution: **Regression Coefficient of Y on X**

$$byx = \frac{\Sigma xy}{\Sigma x^2} = \frac{3900}{6360} = 0.613$$

Regression Coefficient of X on Y

$$bxy = \frac{\Sigma xy}{\Sigma y^2} = \frac{3900}{2860} = 1.363$$

Regression Equation of X on Y

$$X - \overline{X} = bxy\,(Y - \overline{Y})$$

$$X - 90 = 1.363\,(Y - 70)$$

$$X - 90 = 1.363Y - 95.41$$

$$\therefore \quad X = 1.363Y - 5.41$$

Regression Equation of Y on X

$$Y - \overline{Y} = byx\,(X - \overline{X})$$

$$Y - 70 = 0.613\,(X - 90)$$

$$Y - 70 = 0.613X - 55.17$$

$$\therefore \quad Y = 0.613X + 14.83$$

Example 11. The following table gives the aptitude test scores and productivity indices of 10 workers at random:

Aptitude score	Productivity index
60	68
62	60
65	62
70	80
72	85
48	40
53	52
73	62
65	60
82	81

Estimate:

(*i*) the test score of a worker whose productivity index is 75.

(*ii*) the productivity index of a worker whose test score is 92.

Solution:

Calculation of Regression Equations

Aptitude Score X	$(\overline{X} = 65)$ x	x^2	Productivity index Y	$(\overline{Y} = 65)$ y	y^2	xy
60	–5	25	68	+ 3	9	– 15
62	–3	9	60	–5	25	+ 15
65	0	0	62	–3	9	0
70	+ 5	25	80	+ 15	225	+ 75
72	+ 7	49	85	+ 20	400	+ 140
48	–17	289	40	–25	625	+ 425
53	–12	144	52	–13	169	+ 156
73	+ 8	64	62	–3	9	– 24
65	0	0	60	–5	25	0
82	+ 17	289	81	+ 16	256	+ 272
$\Sigma X = 650$	$\Sigma x = 0$	$\Sigma x^2 = 894$	$\Sigma Y = 650$	$\Sigma y = 0$	$\Sigma y^2 = 1752$	$\Sigma xy = 1044$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{650}{10} = 65 \; : \; \overline{Y} = \frac{\Sigma Y}{N} = \frac{650}{10} = 65$$

Regression Equation of X on Y: $X - \overline{X} = bxy(Y - \overline{Y})$

$$bxy = \frac{\Sigma xy}{\Sigma y^2} = \frac{1044}{1752} = +\,0.596$$

$$X - 65 = 0.596\,(Y - 65)$$

$$X - 65 = 0.596\,Y - 38.74$$

or $$X = 26.26 + 0.596\,Y$$

For finding out the test score (X) of a person whose productivity index (Y) is 75, put Y = 75 in the above equation:

$$X_{75} = 26.26 + 0.596(75) = 26.26 + 44.7 = 70.96.$$

Regression Equation of Y on X: $Y - \overline{Y} = byx\,(X - \overline{X})$

$$byx = \frac{\Sigma xy}{\Sigma x^2} = \frac{1044}{894} = +\,1.168$$

$$Y - 65 = 1.168\,(X - 65)$$

$$Y - 65 = 1.168\,X - 75.92 \text{ or } Y = -10.92 + 1.168\,X$$

For finding out the productivity index (Y) of a worker whose test score (X) is 92, we put X = 92 in the above equation.

$$Y_{92} = -10.92 + 1.168(92)$$

$$= -10.92 + 107.456 = 96.536$$

IMPORTANT TYPICAL EXAMPLES

Example 12. The following table shows the number of motor registrations in a certain territory for a term of 5 years and the sale of motor tyres by a firm in that territory for the same period:

Year:	1	2	3	4	5
Motor registration:	600	630	720	750	800
No. of tyres sold:	1,250	1,100	1,300	1,350	1,500

Find the regression equation to estimate the sale of tyres when motor registration is known. Estimate the sale of tyres when registration is 850.

Solution: Let X denotes number of motor registrations and Y denotes the number of tyres sold by a firm.

To simplify the calculation, let

$$x = \frac{X - \bar{X}}{i_x} \qquad y = \frac{Y - \bar{Y}}{i_y}$$

X	$x = \frac{X - \bar{X}}{10}$	x^2	Y	$y = \frac{Y - \bar{Y}}{50}$	y^2	xy
600	– 10	100	1,250	– 1	1	+ 10
630	– 7	49	1,100	– 4	16	+ 28
720	2	4	1,300	0	0	0
750	5	25	1,350	+ 1	1	+ 5
800	10	100	1,500	+ 4	16	+ 40
$\Sigma X = 3500$, $N = 5$	$\Sigma x = 0$	$\Sigma x^2 = 278$	$\Sigma Y = 6500$, $N = 5$	$\Sigma y = 0$	$\Sigma y^2 = 34$	$\Sigma xy = 83$

$$\bar{X} = \frac{3500}{5} = 700, \quad \bar{Y} = \frac{6500}{5} = 1300$$

Here, we have the regression of Y on X.

$$byx = \frac{\Sigma xy}{\Sigma x^2} \times \frac{i_y}{i_x} = \frac{83}{278} \times \frac{50}{10} = 1.4928$$

Regression Equation of Y on X

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

$$Y - 1300 = 1.4928\,(X - 700)$$

$$Y - 1300 = 1.4928X - 1044.96$$

$$Y = 1.4928X + 255.04$$

The estimate of sale of tyres (Y) when registration $X = 850$ is given by

$$Y = 1.4928 \times 850 + 255.04$$

$$= 1268.88 + 255.04 = 1523.92 \approx 1524$$

since the number of tyres cannot be fractional.

Example 13. Calculate the correlation coefficient from the following results:

$$N = 10, \Sigma X = 350, \Sigma Y = 310$$

$$\Sigma (X-35)^2 = 162, \Sigma (Y-31)^2 = 222, \Sigma (X-35)(Y-31) = 92$$

Also find the regression line of Y on X.

Solution:

$$\overline{X} = \frac{\Sigma X}{N} = \frac{350}{10} = 35$$

$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{310}{10} = 31$$

Thus, the given deviations $(X-35)$ and $(Y-31)$ are from actual means $(\overline{X} = 35, \overline{Y} = 31)$.

Thus, $\Sigma(X-35)^2 = 162$ or $\Sigma x^2 = 162$ where, $x = X - \overline{X}$

$\Sigma(Y-31)^2 = 222$, or $\Sigma y^2 = 222$ $y = Y - \overline{Y}$

$\Sigma (X-35)(Y-31) = 92$ *or* $\Sigma xy = 92$

Coefficient of Correlation

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = \frac{92}{\sqrt{162}\sqrt{222}} = +0.485$$

Regression Equation of Y on X

$$Y - \overline{Y} = byx(X - \overline{X})$$

$$byx = \frac{\Sigma xy}{\Sigma x^2} = \frac{92}{162} = 0.568$$

$$\therefore \quad Y - 31 = 0.568(X - 35)$$

$$Y - 31 = 0.568X - 19.88$$

$$Y = 0.568X - 19.88 + 31$$

$$Y = 0.568X + 11.12$$

Graphing Regression Lines

It is quite easy to graph the regression lines once they have been computed. The procedure adopted is as follows:

(*i*) **Regression line of X on Y.** The regression line of X on Y can be drawn with the help of regression equation of X on Y, *i.e.*,

$$X = a + bY$$

If we put the respective values of Y in the above regression equation, we will find the estimated values of X. If we plot estimated values of X with the actual values of Y on the graph, we can draw regression line of X on Y.

(*ii*) **Regression line of Y on X.** The regression line of Y on X can be drawn with the help of regression equation of Y on X, *i.e.*,

$$Y = a + bX$$

If we put the respective values of X in the above equation, we will find the estimated values of Y. If we plot estimated values of Y with the actual values of X on the graph, we can draw regression line of Y on X.

The following example illustrate the graphing of regression lines.

Example 14. From the following data:

(*i*) Obtain the two regression equations.

(*ii*) Draw up the two regression lines on the graph paper with the help of two regression equations.

X:	1	2	3
Y:	5	4	6

Solution:

Calculation of Regression Equation

X	$\bar{X}=2$ x	x^2	Y	$\bar{Y}=5$ y	y^2	xy
1	−1	1	5	0	0	0
2	0	0	4	−1	1	0
3	+1	1	6	+1	1	+1
$\Sigma X = 6$ $N = 3$	$\Sigma x = 0$	$\Sigma x^2 = 2$	$\Sigma Y = 15$ $N = 3$	$\Sigma y = 0$	$\Sigma y^2 = 2$	$\Sigma xy = +1$

$$\therefore \quad \bar{X} = \frac{\Sigma X}{N} = \frac{6}{3} = 2; \qquad \bar{Y} = \frac{15}{3} = 5$$

$$byx = \frac{\Sigma\, xy}{\Sigma\, x^2} = \frac{1}{2} \qquad bxy = \frac{\Sigma\, xy}{\Sigma\, y^2} = \frac{1}{2}$$

Regression Equation of Y on X

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

$$Y - 5 = \frac{1}{2}(X - 2)$$

$$Y - 5 = \frac{1}{2}X - 1$$

$$\therefore \quad Y = \frac{1}{2}X + 4$$

Regression Equation of X on Y

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

$$X - 2 = \frac{1}{2}(Y - 5)$$

$$X - 2 = \frac{1}{2}Y - \frac{5}{2}$$

$$\therefore \quad X = \frac{1}{2}Y - \frac{1}{2}$$

(*ii*) **Regression Lines:** In order to draw up the two regression lines on the graph, we shall have to plot the given values of X and the computed values of Y and the given values of Y and the computed values of X

Computed Values of Y

Regression equation of Y on X

$$Y = \frac{1}{2}X + 4$$

Computed Values of X

Regression equation of X on Y

$$X = \frac{1}{2}Y - \frac{1}{2}$$

when $X=1,\ Y=\frac{1}{2}(1)+4=4.5$

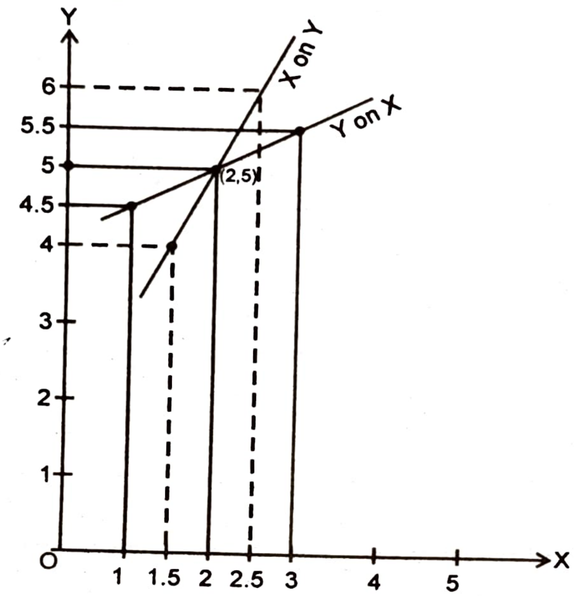
when $X=2,\ Y=\frac{1}{2}(2)+4=5.0$

when $X=3,\ Y=\frac{1}{2}(3)+4=5.5$

when $Y=5,\ X=\frac{1}{2}(5)-\frac{1}{2}=2$

when $Y=4,\ X=\frac{1}{2}(4)-\frac{1}{2}=1.5$

when $Y=6,\ X=\frac{1}{2}(6)-\frac{1}{2}=2.5$



Example 15. Compute the appropriate regression equation for the following data:

X (Independent variable)	Y (Dependent variable)
2	18
4	12
5	10
6	8
8	7
11	5

Solution: The appropriate regression equation will be Y on X

X	$\overline{X}=6$ x	x^2	Y	$\overline{Y}=10$ y	y^2	xy
2	-4	16	18	8	64	-32
4	-2	4	12	2	4	-4
5	-1	1	10	0	0	0
6	0	0	8	-2	4	0
8	$+2$	4	7	-3	9	-6
11	$+5$	25	5	-5	25	-25
$\Sigma X=36$	$\Sigma x=0$	$\Sigma x^2=50$	$\Sigma Y=60$	$\Sigma y=0$	$\Sigma y^2=106$	$\Sigma xy=-67$

$$\overline{X} = \frac{36}{6} = 6; \qquad \overline{Y} = \frac{60}{6} = 10$$

$$byx = \frac{\Sigma xy}{\Sigma x^2}$$

$$= \frac{-67}{50} = -1.34$$

Regression Equation of Y on X

$$Y - \overline{Y} = byx\,(X - \overline{X})$$

$$Y - 10 = -1.34\,(X - 6)$$

$$\therefore \quad Y - 10 = -1.34X + 8.04$$

$$Y = -1.34X + 18.04$$

EXERCISE 2.3

1. For the following data, set up regression equation and estimate sales for an advertisement expenditure of Rs. 75 lakh.

Sales (Rs. crore):	14	16	18	20	24	30	32
Adv. expenditure (Rs. lakh):	52	62	65	70	76	80	78

[**Hint:** Let X denote sales) [**Ans.** $X = 0.621Y - 20.85, X_{75}, = 25.725$]

2. Find the correlation coefficient and the equations of regression lines for the following values of X and Y:

X:	11	7	2	5	8	6	10
Y:	7	5	3	2	6	4	8

[**Ans.** $r = 0.884$, $X = 0.75 + 1.25Y$, $Y = 0.625 + 0.625X$]

3. The following data relate to marketing expenditure and the corresponding sales:

Expenditure (X) (Rs. lac):	10	12	15	20	23
Sales (Y) (Rs. crore) :	14	17	23	21	35

Estimate the marketing expenditure to obtain a sales target of Rs. 40 crore.

[**Ans.** $X = 0.59Y + 3.02; X_{40} = 26.62$]

4. The following are the intermediate results of the two series X and Y

$\overline{X} = 65, \overline{Y} = 65, N = 10, \Sigma x^2 = 894, \Sigma y^2 = 1752, \Sigma xy = 1044$

(Where x and y are deviations from the respective means)

Find two regression equations. Also estimate Y when $X = 92$ and X when $Y = 75$.

[**Ans.** $Y = 1.168X - 10.92$, $Y_{92} = 96.536$; $X = 0.596Y + 26.26$, $X_{75} = 70.96$]

5. An investigation in to the demand for Television sets in 7 towns has resulted in the following data:

Population ('000) (X):	11	14	14	17	17	21	25
No. of T.V. sets demanded (Y):	15	27	27	30	34	38	46

Calculate the regression equation of Y on X and estimate the demand for T.V. sets for a town with a population of 30 thousand. **[Ans.** $Y = -3 + 2X$; $Y_{30} = 57$]

6. A departmental store gives in-service training to its salesmen which is followed by a test. It is considering whether it should terminate the service of any salesman who does not do well in the test. The following data give the test scores and sales made by nine salesmen during a certain period:

Test scores:	14	19	24	21	26	22	15	20	19
Sales ('00 Rs.):	31	36	48	37	50	45	33	41	39

Calculate the coefficient of correlation between the test scores and the sales. Does it indicate that the termination of services of low test scores is justified ? If the firm wants a minimum sales volume of Rs. 3,000, what is the minimum test score that will ensure continuation of service ? Also estimate the most probable sales volume of a salesman making a score of 28.
[**Hint :** See Example 57] **[Ans.** $r = 0.9471$, justified, $X = 14.422 \approx 14$, $Y = 5286.64$]

7. The following table gives the marks in Economics and Statistics of 10 students selected at random:

Marks in Economics:	25	28	35	32	31	36	29	38	34	32
Marks in Statistics:	43	46	49	41	36	32	31	30	33	39

Find (*i*) The two regression equations.
(*ii*) The coefficient of correlation between marks in Economics and Statistics.
(*iii*) The most likely marks in statistics when marks in economics are 30.
[Ans. (*i*) $X = -0.2337Y + 40.8806$, $Y = -0.6643X + 59.2576$
(*ii*) $r = -0.394$, (*iii*) 39.3286, or 39 marks]

8. The profits (Y) of a company in the Xth year of its life were as follows:

Years of life (X):	1	2	3	4	5
Profits (Y) (in lakh of Rs.):	1250	1400	1650	1950	2300

Estimate the profit of a company in the 6th year. **[Ans.** $Y = 265X + 915$, $Y_6 = $ Rs.2505 lakh]

9. From the following data:
(*i*) Obtain the two regression equations.
(*ii*) Draw up two regression lines on the graph paper.

X:	65	66	67	68	69	70	71
Y:	67	68	64	70	70	69	68

[Ans. $X = 0.462Y + 36.72$, $Y = 0.353X + 44$]