$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{406}{8} = 50.75 \; ; \; \overline{X} = \frac{\Sigma X}{N} = \frac{405}{8} = 50.625$$

$$r\frac{\sigma_y}{\sigma_x} = \frac{N\Sigma d_x d_y - (\Sigma d_x)(\Sigma d_y)}{N\Sigma d_x^2 - (\Sigma d_x)^2} = \frac{(8)(1336) - (5)(6)}{(8)(1737) - (5)^2} = \frac{10688 - 30}{13896 - 25} = 0.768$$

$$Y - 50.75 = 0.768\,(X - 50.625)$$

$$Y - 50.75 = 0.768\,X - 38.88 \quad \text{or} \quad Y = 11.87 + 0.768\,X$$

$$Y_{49} = 11.87 + 0.768\,(49) = 11.87 + 37.632 = 49.502$$

Thus the expected blood pressure of a person who is 49 years old shall be 49.5.

Illustration 8. In a correlation study the following values are obtained :

	X	Y
Mean	65	67
Standard Deviation	2·5	3·5
Coefficient of Correlation	0·8	

Find the two regression equations that are associated with the above values.

(B.Com., Kashmir Univ., 2006; MBA, HPU, 2007; M.Com., Madurai Kamaraj Univ., 2009)

Solution. The two regression equations are :

Regression Equation of X on Y : $X - \overline{X} = r\dfrac{\sigma_x}{\sigma_y}(Y - \overline{Y})$

$$\overline{X} = 65, \; r = 0.8, \; \sigma_x = 2.5, \; \sigma_y = 3.5, \; \overline{Y} = 67$$

Substituting the values : $X - 65 = 0.8\,\dfrac{2.5}{3.5}(Y - 67)$

$$X - 65 = 0.5714\,(Y - 67)$$

$$X - 65 = 0.5714\,Y - 38.28 \quad \text{or} \quad X = 26.72 + 0.5714\,Y$$

Regression Equation of Y on X : $Y - \overline{Y} = r\dfrac{\sigma_y}{\sigma_x}(X - \overline{X})$

$$Y - 67 = 0.8\,\frac{3.5}{2.5}(X - 65)$$

$$Y - 67 = 1.12\,(X - 65)$$

$$Y - 67 = 1.12\,X - 72.8 \quad \text{or} \quad Y = -5.8 + 1.12X.$$

Illustration 9. In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible :

Variance of $X = 9$

Regression equations : $8X - 10Y + 66 = 0$

$40X - 18Y = 214$

Find on the basis of the above information

(*i*) The mean values of X and Y

(*ii*) Coefficient of correlation between X and Y, and

(*iii*) Standard deviation of Y.

(M.Com., Vikram Univ., 2003; M.Com., Mysore Univ., 2004; M.Com., Garhwal Univ., 2012; BBA, GGSIP Univ., 2012)

Solution.

(*i*) *Finding the Mean values of X* and *Y* : $8X - 10Y = -66$ ...(*i*)

$40X - 18Y = 214$ ...(*ii*)

Multiplying equation (*i*) by 5 $\quad 40X - 50Y = -330$

$$\begin{array}{r} 40X - 18Y = 214 \\ - \quad + \quad - \\ \hline -32Y = -544 \end{array}$$

$$Y = 17 \quad \text{or} \quad \overline{Y} = 17$$

Substituting the value of Y in eq. (*i*) ; $8X - 10 \times 17 = -66$

$$8X = -66 + 170$$

$$8X = 104 \; \therefore \; X = 13 \quad \text{or} \quad \overline{X} = 13$$

(*ii*) For finding out the correlation coefficient, we will have to find out the regression coefficient. Since we do not know which of the two regression equations is the equation of X on Y, we make an assumption. Let us take eq. (*i*) as the regression equation of X on Y

Solution. Let intelligence test score be denoted by X and weekly sales by Y.

CALCULATION OF REGRESSION EQUATIONS

X	$(X-60)$ x	x^2	Y	$(Y-50)$ y	y^2	xy
50	−10	100	30	−20	400	+200
60	0	0	60	+10	100	0
50	−10	100	40	−10	100	+100
60	0	0	50	0	0	0
80	+20	400	60	+10	100	+200
50	−10	100	30	−20	400	+200
80	+20	400	70	+20	400	+400
40	−20	400	50	0	0	0
70	+10	100	60	+10	100	+100
$\Sigma X = 540$	$\Sigma x = 0$	$\Sigma x^2 = 1600$	$\Sigma Y = 450$	$\Sigma y = 0$	$\Sigma y^2 = 1600$	$\Sigma xy = 1200$

Regression equation of Y on X : $Y - \overline{Y} = r\dfrac{\sigma_y}{\sigma_x}(X - \overline{X})$

$$r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{\Sigma x^2} = \frac{1200}{1600} = 0{\cdot}75$$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{540}{9} = 60,\ \overline{Y} = \frac{\Sigma Y}{N} = \frac{450}{9} = 50$$

$$Y - 50 = 0{\cdot}75\,(X - 60)$$

$$Y - 50 = 0{\cdot}75\,X - 45 \quad \text{or} \quad Y = 5 + 0{\cdot}75\,X$$

Expected weekly sales when intelligence test score of a salesman is 65

$$Y = 5 + 0.75\,X. \qquad \text{Putting } X = 65$$

$$Y = 0.75 \times (65) + 5 = 48.75 + 5 = 53.75$$

Illustration 7. The following table shows the ages (X) and blood pressure (Y) of 8 persons.

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

Obtain the regression equation of Y on X and find the expected blood pressure of a person who is 49 years old.

Solution. CALCULATION OF REGRESSION EQUATION OF Y ON X

X	$(X-50)$ d_x	d_x^2	Y	$(Y-50)$ d_y	d_y^2	$d_x d_y$
52	+2	4	62	+12	144	+24
63	+13	169	53	+3	9	+39
45	−5	25	51	+1	1	−5
36	−14	196	25	−25	625	+350
72	+22	484	79	+29	841	+638
65	+15	225	43	−7	49	−105
47	−3	9	60	+10	100	−30
25	−25	625	33	−17	289	+425
$\Sigma X = 405$	$\Sigma d_x = 5$	$\Sigma d_x^2 = 1737$	$\Sigma Y = 406$	$\Sigma d_y = 6$	$\Sigma d_y^2 = 2058$	$\Sigma d_x d_y = 1336$

$$Y - \overline{Y} = r\frac{\sigma_y}{\sigma_x}(X - \overline{X})$$

$$8X = -66 + 10Y$$

$$X = -\frac{66}{8} + \frac{10}{8}Y; \quad \text{or} \quad b_{xy} = \frac{10}{8} = 1.25$$

From eq. (*ii*) we can calculate b_{yx} $\quad 40X - 18Y = 214$

or, $\quad -18Y = 214 - 40X$

$$Y = -\frac{214}{18} + \frac{40}{18}X \quad \text{or} \quad b_{yx} = 2.22$$

Since both the regression coefficients are exceeding 1, our assumption is wrong. Hence the first equation is equation of Y on X.

From eq. (*i*)

$$-10Y = -8X - 66$$

$$Y = -\frac{8}{10}X + 6.6 \quad \text{or} \quad b_{yx} = \frac{8}{10} = 0.8$$

From eq. (*ii*) $\quad b_{xy} = \frac{18}{40} = 0.45$

$$r = \sqrt{0.8 \times 0.45} = \sqrt{0.36} = 0.6$$

(*iii*) S.D. of Y: $\quad \sigma_x = \sqrt{9} = 3\,;\; b_{xy} = r\frac{\sigma_x}{\sigma_y}$

$$0.45 = 0.6\frac{3}{\sigma_y} \quad \text{or} \quad 0.45\,\sigma_y = 1.8 \quad \text{or} \quad \sigma_y = \frac{1.8}{0.45} = 4$$

Hence standard deviation of Y is 4.

Illustration 10. For 50 students of a class the regression equation of marks in Statistics (X) on the marks in Accountancy (Y) is $3Y - 5X + 180 = 0$. The mean marks in Accountancy is 44 and variance of marks in Statistics is 9/16th of the variance of marks in Accountancy. Find the mean marks in Statistics and the coefficient of correlation between marks in the two subjects. *(M.Com., Jamia Millia Univ., 2006; B.Com., Delhi Univ., 2007)*

Solution. We are given

$$3Y - 5X + 180 = 0 \quad \text{or} \quad 3Y + 180 = 5X$$

X represents marks in Statistics and Y marks in Accountancy. When $Y = 44$, X will be given by

$$5X = (3)(44) + 180 = 0\,;\; 5X = 132 + 180 \text{ or } X = \frac{312}{5} = 62.4$$

Hence the mean marks in Statistics are 62.4

For calculating coefficient of correlation we know that

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

Regression coefficient of X on Y from the given equation is

$$5X = 3Y + 180 \quad \text{or} \quad X = 0.6Y + 36$$

$$b_{xy} = 0.6\,;\; r\frac{\sigma_x}{\sigma_y} = \frac{\sqrt{9}}{\sqrt{16}} \quad \text{given}$$

$$0.6 = r\frac{\sqrt{9}}{\sqrt{16}} \quad \text{or} \quad 0.6 = r\frac{3}{4}$$

Hence $\quad 3r = 2.4 \quad$ or $\quad r = 0.8.$

Illustration 11. You are given the following data :

	X	Y
Arithmetic Mean	36	85
Standard Deviation	11	8

Correlation coefficient between X and Y $\quad$ 0.66

(*i*) Find the two Regression Equations

(*ii*) Estimate the value of X when $Y = 75$.

(B.Com Garhwal Univ. 2010; B.Com., Bharthiar Univ., 2011)

Solution:

(*i*) *Regression Equation of X on Y* : $X - \overline{X} = r\frac{\sigma_x}{\sigma_y}(Y - \overline{Y})$

$$\overline{X} = 36,\ r = 0.66, \sigma_x = 11,\ \sigma_y = 8,\ \overline{Y} = 85$$

$$X - 36 = 0.66\frac{11}{8}(Y - 85)$$

$$X - 36 = 0.9075(Y - 85)$$

$$X = 0.9075\ Y - 77.1375 + 36 \quad \text{or} \quad X = -41.1375 + 0.9075\ Y$$

Regression Equation of Y on X : $Y - \overline{Y} = r\frac{\sigma_y}{\sigma_x}(X - \overline{X})$

$$Y - 85 = 0.66\frac{8}{11}(X - 36)$$

$$Y - 85 = 0.48(X - 36)$$

$$Y - 85 = 0.48\ X - 17.28 \quad \text{or} \quad Y = 67.72 + 0.48\ X$$

(*ii*) From the regression equation of X on Y, we can find out the estimated value of X when $Y = 75$; $X = 0.9075\ (75) - 41.1375$

$$= 68.0625 - 41.1375 = 26.925 \quad \text{or} \quad Y_{75} = 26.925.$$

Illustration 12. For certain X and Y series which are correlated, the two lines of regression are :

$$5\ X - 6\ Y + 90 = 0$$

$$15\ X - 8\ Y - 130 = 0$$

Find the mean of the two series and the correlation coefficient. *(BBM., Mysore Univ. 2007)*

Solution : (*i*) Finding mean of the two series :

$$5\ X - 6\ Y = -90 \quad \ldots(i)$$

$$15\ X - 8\ Y = 130 \quad \ldots(ii)$$

Multiplying eq. (*i*) by 3,

$$15\ X - 18\ Y = -270$$

$$15\ X - 8\ Y = 130$$

$$-\quad +\quad -$$

$$-10\ Y = -400$$

$$Y = 40 \quad \text{or} \quad \overline{Y} = 40$$

Putting the value of Y in eq. (*i*), $5\ X - 6\ (40) = -90$

$$5\ X = -90 + 240$$

$$5\ X = 150 \quad \text{or} \quad X = 30 \quad \text{or} \quad \overline{X} = 30$$

(*ii*) *Finding correlation coefficient.* Let us assume that eq. (*i*) is the regression equation of X on Y;

$$5\ X = 6\ Y - 90$$

$$X = \frac{6}{5}\ Y - 18 \quad \text{or} \quad b_{xy} = \frac{6}{5}$$

Taking eq. (*ii*) as the eq. of Y on X; $-8\ Y = -15\ X + 130$

$$8\ Y = 15\ X - 130$$

$$Y = \frac{15}{8}\ X - \frac{130}{8} \quad \text{or} \quad b_{yx} = \frac{15}{8}$$

Since both the regression coefficients are exceeding one, our assumption is wrong.
Hence eq. (*i*) is the regression eq. of Y on X

$\therefore$

$$-6\ Y = -5\ X - 90 \quad \text{or} \quad 6\ Y = 5\ X + 90$$

or

$$Y = \frac{5}{6}\ X + 15 \quad \text{or} \quad b_{yx} = \frac{5}{6}$$

Eq. (*ii*) is the regression eq. of X on Y; $15\ X = 130 + 8\ Y$

$$X = \frac{130}{15} + \frac{8}{15}\ Y;\ b_{xy} = \frac{8}{15}$$

$$r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{\frac{8}{15} \times \frac{5}{6}} = 0.666$$

Illustration 21. The following data give the experience of machine operators and their performance ratings as given by the number of good parts turned out per 100 pieces :

Operator	:	1	2	3	4	5	6	7	8
Experience *(X)*	:	16	12	18	4	3	10	5	12
Performance Ratings *(Y)*	:	87	88	89	68	78	80	75	83

Calculate the regression lines of performance ratings on experience and estimate the probable performance if an operator has 7 years experience. *(MBA, Kumaun Univ., 2012)*

Solution : Let performance ratings be denoted by Y and experience by X. We have to calculate the regression line of Y on X.

CALCULATING REGRESSION LINE OF Y ON X

Experience X	$(X-\overline{X})$ $\overline{X}=10$ x	x^2	*Performance Ratings* Y	$(Y-\overline{Y})$ $\overline{Y}=81$ y	y^2	xy
16	+6	36	87	+6	36	+36
12	+2	4	88	+7	49	+14
18	+8	64	89	+8	64	+64
4	−6	36	68	−13	169	+78
3	−7	49	78	−3	9	+21
10	0	0	80	−1	1	0
5	−5	25	75	−6	36	+30
12	+2	4	83	+2	4	+4
$\Sigma X=80$	$\Sigma x=0$	$\Sigma x^2=218$	$\Sigma Y=648$	$\Sigma y=0$	$\Sigma y^2=368$	$\Sigma xy=247$

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

$$b_{yx}=\frac{\Sigma xy}{\Sigma x^2}=\frac{247}{218}=1.133\ ;\ \overline{Y}=\frac{648}{8}=81\ ;\ \overline{X}=\frac{80}{8}=10$$

$\therefore$ $Y-81=1.133\,(X-10)=1.133\,X-11.33$ or $Y=69.67+1.133\,X$

when $X=7$, Y will be $Y=69.67+1.133\,(7)=69.67+7.931=77.601$.

Thus the probable performance of an operator who has 7 years experience = 77.601 or 78 good parts out of 100.