Example 11·1. *Obtain the equations of two lines of regression for the following data. Also obtain the estimate of X for Y = 70.*

X :	65	66	67	67	68	69	70	72
Y :	67	68	65	68	72	72	69	71

Solution. Let $U = X - 68$ and $V = Y - 69$, then

$\bar{U} = 0, \bar{V} = 0, \sigma_U^2 = 4{\cdot}5, \sigma_V^2 = 5{\cdot}5$, Cov $(U, V) = 3$ and $r\ (U, V) = 0{\cdot}6$ (*c.f.* page 10·6)

Since correlation coefficient is independent of change of origin, we get

$$r = r(X, Y) = r(U, V) = 0{\cdot}6$$

We know that if $U = \dfrac{X-\alpha}{h},\ V = \dfrac{Y-b}{k}$, then

$$\bar{X} = a + h\bar{U}\ ,\ \bar{Y} = b + k\bar{V}\ ,\ \sigma_X = h\sigma_U \quad \text{and} \quad \sigma_Y = k\sigma_Y$$

Here we are given : $h = k = 1, a = 68$ and $b = 69$.

Thus $\quad \bar{X} = 68 + 0 = 68, \bar{Y} = 69 + 0 = 69$

$$\sigma_X = \sigma_U = \sqrt{4{\cdot}5} = 2{\cdot}12 \quad \text{and} \quad \sigma_Y = \sigma_V = \sqrt{5{\cdot}5} = 2{\cdot}35$$

Equation of line of regression of Y on X is : $Y - \bar{Y} = r\dfrac{\sigma_Y}{\sigma_X}(X - \bar{X})$

i.e., $\quad Y = 69 + 0{\cdot}6 \times \dfrac{2{\cdot}35}{2{\cdot}12}(X - 68) \quad \Rightarrow \quad Y = 0{\cdot}665\, X + 23{\cdot}78$

Equation of line of regression of X on Y is :

$$X - \bar{X} = r\,\frac{\sigma_X}{\sigma_Y}\,(Y - \bar{Y})$$

$$\Rightarrow \qquad X = 68 + 0{\cdot}6 \times \frac{2{\cdot}12}{2{\cdot}35}(Y - 69) \quad \Rightarrow \quad X = 0{\cdot}54Y + 30{\cdot}74$$

To estimate X for given Y, we use the line of regression of X on Y. If $Y = 70$, estimated value of X is given by : $\hat{X} = 0{\cdot}54 \times 70 + 30{\cdot}74 = 68{\cdot}54$,

where $\hat{X}$ is estimate of X.

Example 11·2. *In a partially destroyed laboratory, record of an analysis of correlation data, the following results only are legible :*

Variance of $X = 9$. *Regression equations :* $8X - 10Y + 66 = 0$, $40X - 18Y = 214$.

What are : (i) the mean values X and Y, (ii) the correlation coefficient between X and Y, and (iii) the standard deviation of Y ?

Solution. (*i*) Since both lines of regression pass through the point $(\bar{X}, \bar{Y})$, we have : $8\bar{X} - 10\bar{Y} + 66 = 0$, and $40\bar{X} - 18\bar{Y} = 214$. Solving, we get $\bar{X} = 13$, $\bar{Y} = 17$.

(*ii*) Let $8X - 10Y + 66 = 0$ and $40X - 18Y = 214$ be the lines of regression of Y on X and X on Y respectively. These equations can be put in the form :

$$Y = \frac{8}{10}X + \frac{66}{10} \qquad \text{and} \qquad X = \frac{18}{40} + \frac{214}{40}$$

$$\therefore \qquad b_{YX} = \text{Regression coefficient of } Y \text{ on } X = \frac{8}{10} = \frac{4}{5}$$

$$\text{and} \qquad b_{XY} = \text{Regression coefficient of } X \text{ on } Y = \frac{18}{40} = \frac{9}{20}$$

$$\text{Hence} \qquad r^2 = b_{YX} \cdot b_{XY} = \frac{4}{5} \cdot \frac{9}{20} = \frac{9}{25} \qquad \therefore\ r = \pm\frac{3}{5} = \pm 0{\cdot}6$$

But since both the regression coefficients are positive, we take $r = +0{\cdot}6$.

$$\text{(iii) We have} \qquad b_{YX} = r \cdot \frac{\sigma_Y}{\sigma_X} \quad \Rightarrow \quad \frac{4}{5} = \frac{3}{5} \times \frac{\sigma_Y}{3}$$

Hence $\qquad \sigma_Y = 4$.

Remarks 1. It can be verified that the values of $\bar{X} = 13$ and $\bar{Y} = 17$ as obtained in part (*i*) satisfy both the regression equations. In numerical problems of this type, this check should invariably be applied to ascertain the correctness of the answer.

2. If we had assumed that $8X - 10Y + 66 = 0$, is the equation of the line of regression of X on Y and $40X - 18Y = 214$ is the equation of line of regression of Y on X, then we get respectively :

$$8X = 10Y - 66 \qquad \text{and} \qquad 18Y = 40X - 214$$

$$\Rightarrow \qquad X = \frac{10}{8}Y - \frac{66}{8} \qquad \text{and} \qquad Y = \frac{40}{18}X - \frac{214}{18}$$

$$\Rightarrow \qquad b_{XY} = \frac{10}{8} \qquad \text{and} \qquad b_{YX} = \frac{40}{18}$$

$$\therefore \qquad r^2 = b_{XY} \cdot b_{YX} = \frac{10}{8} \times \frac{40}{18} = 2{\cdot}78$$

But since r^2 always lies between 0 and 1, our supposition is wrong.

Example 11·3. *Find the most likely price in Mumbai corresponding to the price of Rs. 70 at Kolkata from the following :*

	Kolkata	*Mumbai*
Average price	65	67
Standard deviation	2·5	3·5

Correlation coefficient between the prices of commodities in the two cities is 0·8.

Solution. Let the prices (in Rupees) in Kalkata and Mumbai be denoted by X and Y respectively. Then we are given :

$$\bar{X} = 65, \bar{Y} = 67, \sigma_X = 2{\cdot}5, \sigma_Y = 3{\cdot}5 \text{ and } r = r(X, Y) = 0{\cdot}8. \text{ We want } Y \text{ for } X = 70.$$

Line of regression of Y on X is :

$$Y - \bar{Y} = r\,\frac{\sigma_Y}{\sigma_X}(X - \bar{X}) \quad \Rightarrow \quad Y = 67 + 0{\cdot}8 \times \frac{3{\cdot}5}{2{\cdot}5}(X - 65)$$

When $X = 70$, $\quad \hat{Y} = 67 + 0{\cdot}8 \times \frac{3{\cdot}5}{2{\cdot}5}((70 - 65) = 72{\cdot}6$

Hence, the most likely price in Mumbai corresponding to the price of Rs. 70 at Kolkata is Rs. 72·60.

Example 11·4. *Can $Y = 5 + 2{\cdot}8\,X$ and $X = 3 - 0{\cdot}5Y$ be the estimated regression equations of Y on X and X on Y respectively ? Explain your answer with suitable theoretical arguments.*

Solution . Line of regression of Y on X is : $Y = 5 + 2{\cdot}8\,X \quad \Rightarrow \quad b_{YX} = 2{\cdot}8 \quad \ldots (*)$

Line of regression of X on Y is : $X = 3 - 0{\cdot}5Y \quad \Rightarrow \quad b_{XY} = -0{\cdot}5 \quad \ldots (**)$

This is not possible, since each of the regression coefficients b_{YX} and b_{XY} must have the same sign, which is same as that of Cov (X, Y). If Cov (X, Y) is positive, then both the regression coefficients are positive and if Cov (X, Y) is negative, then both the regression coefficients are negative. Hence (*) and (**) cannot be the estimated regression equations of Y on X and X on Y respectively.