



DEPARTMENT OF MATHEMATICS

34

REGRESSION :

Regression is the estimation or prediction of unknown values of one variable from the known values of another variable. It actually measures the nature and extent to which the variables are correlated.

Lines of Regression :

If the variables in a bivariate distribution are related we will find that the points in the scattered diagram will cluster around some curve called the curve of regression. If the curve is a straight line, it is called the line of regression.

Equations of lines of regression :

The lines of regression of X on Y is ,

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

The lines of regression of Y on X is ,

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

Regression Coefficients:

Regression Coefficient of X on Y is ,

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

Scanned by CamScanner



(36)

PROBLEMS :

- ① From the following data, find
- (i) the two regression lines
 - (ii) the Coefficient of Correlation between the marks in Economics and Statistics
 - (iii) the most likely marks in Statistics when marks in Economics are 30

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

Solution :

X	Y	$x - \bar{x}$ $= x - 32$	$y - \bar{y}$ $= y - 38$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
320	380	0	0	140	398	-93

Scanned by CamScanner



DEPARTMENT OF MATHEMATICS

$$\bar{X} = \frac{\sum X}{n} = \frac{320}{10} = 32$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{380}{10} = 38$$

(i) Coefficient of regression of Y on X is,

$$b_{YX} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} = \frac{-93}{140} = -0.6643$$

Coefficient of regression of X on Y is,

$$b_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (Y - \bar{Y})^2} = \frac{-93}{398} = -0.2337$$

Equation of the line of regression of X on Y is,

$$X - \bar{X} = b_{XY} (Y - \bar{Y})$$

$$X - 32 = -0.2337 (Y - 38)$$

$$X = -0.2337Y + 0.2337 \times 38 + 32$$

$$X = -0.2337Y + 40.8806$$

Equation of the line of regression of Y on X is,

$$Y - \bar{Y} = b_{YX} (X - \bar{X})$$

$$Y - 38 = -0.6643 (X - 32)$$

$$Y = -0.6643X + 0.6643 \times 32 + 38$$

$$Y = -0.6643X + 59.2576$$

(ii) Coefficient of Correlation,

$$r^2 = b_{YX} \times b_{XY}$$

$$= (-0.6643) (-0.2337)$$

Scanned by CamScanner



DEPARTMENT OF MATHEMATICS

$$r = \frac{1}{3}(x+y), \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 2.$$

$$r^2 = 0.155a$$

$$r = \pm \sqrt{0.155a}$$

$$r = \pm 0.394$$

(37)

(40)

(iii) To find the most likely marks in Statistics (y) when marks in Economics (x) are 30. We use the line of regression of y on x .

$$y = -0.6643x + 59.2575$$

Put $x = 30$,

$$y = -0.6643(30) + 59.2575$$

$$y = 39.3286$$

$$y \approx 39$$

(2) The two lines of regression are

$$8x - 10y + 66 = 0$$

$$40x - 18y - 214 = 0$$

The variance of x is 9. Find (i) the mean values of x and y (ii) Correlation Coefficient between x and y .

Solution:

$$\text{Given: } 8x - 10y = -66$$

$$40x - 18y = 214$$

Since both the lines of regression passes through the mean values \bar{x} and \bar{y} ,

$$8\bar{x} - 10\bar{y} = -66 \quad \rightarrow \textcircled{1}$$

$$40\bar{x} - 18\bar{y} = 214 \quad \rightarrow \textcircled{2}$$

$$\frac{\sigma_y^2}{\sigma_x^2} = \frac{16}{16} \Rightarrow \frac{\sigma_y}{\sigma_x} = \frac{4}{4}$$

Scanned by CamScanner