



Regression

Regression is a mathematical measure of the avg. relationship b/w two or more variables.

Lines of Regression:

1] The line of regression of y on x :

$$y - \bar{y} = b_{yx} (x - \bar{x})$$

$$\text{where } b_{yx} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad \text{or} \quad b_{yx} = r \frac{\sigma_y}{\sigma_x}$$

2] The line of regression of x on y :

$$x - \bar{x} = b_{xy} (y - \bar{y})$$

$$\text{where } b_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (y - \bar{y})^2} \quad \text{or} \quad b_{xy} = r \frac{\sigma_x}{\sigma_y}$$

Regression co-efficient:

i). Regression co-efficient of y on x

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

ii) Regression co-efficient of x on y

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

Correlation coefficient:

$$r = \pm \sqrt{b_{xy} \cdot b_{yx}}$$

Angle b/w two lines of regression

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



Q] From the following data, find

- i). two regression eqns.
- ii). The co-efficient of correlation b/w 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

Soln.

$$\text{Here } \bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32$$

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



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

Coimbatore-641035

DEPARTMENT OF MATHEMATICS

Regression



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
$\Sigma x =$ 320	$\Sigma y =$ 380	$\Sigma(x - \bar{x}) =$ 0	$\Sigma(y - \bar{y}) =$ 0	$\Sigma(x - \bar{x})^2 =$ 140	$\Sigma(y - \bar{y})^2 =$ 398	$\Sigma(x - \bar{x})(y - \bar{y}) =$ -93

$$b_{yx} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2} = \frac{-93}{140} = -0.664$$

$$b_{xy} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(y - \bar{y})^2} = \frac{-93}{398} = -0.2336$$

i) Eqn. of line of regression of y on x is,

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 38 = (-0.664)(x - 32)$$

$$= -0.664x + 21.248$$

$$y = -0.664x + 21.248 + 38$$

$$y = -0.664x + 59.248$$

ii) Eqn. of line of regression of x on y is,

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 32 = (-0.2336)(y - 38)$$



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

Coimbatore-641035

DEPARTMENT OF MATHEMATICS

Regression



$$x = 32 - 0.2336y + 8.8768$$

$$x = -0.2336y + 40.8768$$

$$\begin{aligned} \text{ii). } r &= \pm \sqrt{b_{xy} \cdot b_{yx}} = \pm \sqrt{(-0.2336)(-0.664)} \\ &= \pm \sqrt{0.15478} \end{aligned}$$

$$r = \pm 0.3934$$

Q7. Two lines of regression are,

$8x - 10y + 66 = 0$; $40x - 18y - 214 = 0$. The variance of x is

Q. Find

i). The mean values of x and y .

ii). The correlation coefficient b/w x and y .

Soln.

$$\text{Given } 8x - 10y + 66 = 0$$

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

Since both the lines of regression pass

through (\bar{x}, \bar{y}) .

$$8\bar{x} - 10\bar{y} + 66 = 0 \rightarrow (1)$$

$$40\bar{x} - 18\bar{y} - 214 = 0 \rightarrow (2)$$

$$(1) \times 5 \Rightarrow 40\bar{x} - 50\bar{y} + 330 = 0$$

$$(2) \Rightarrow 40\bar{x} - 18\bar{y} - 214 = 0$$

$$\begin{array}{r} (-) \quad \quad (+) \quad \quad (+) \\ \hline \end{array}$$

$$-32\bar{y} + 544 = 0$$

$$\bar{y} = \frac{544}{32}$$

$$\bar{y} = 17$$

Sub $\bar{y} = 17$ in (1),

$$8\bar{x} - 10(17) = -66$$



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

Coimbatore-641035

DEPARTMENT OF MATHEMATICS

Regression



$$8\bar{x} = -66 + 170 = 104$$

$$\bar{x} = 13$$

mean values of x and y are

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

i) from $8x - 10y + 66 = 0$

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

$$y = \frac{8x + 66}{10}$$

$$y = \frac{8}{10}x + \frac{66}{10}, \text{ which is the line of regression of } y \text{ on } x.$$

$$\therefore b_{yx} = \frac{8}{10}$$

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

$$40x = 18y + 214$$

$$x = \frac{18}{40}y + \frac{214}{40}, \text{ which is the line of regression of } x \text{ on } y.$$

$$\therefore b_{xy} = \frac{18}{40}$$

\therefore correlation coefficient :

$$r = \pm \sqrt{b_{xy} \cdot b_{yx}}$$

$$= \pm \sqrt{\frac{18}{40} \left(\frac{8}{10} \right)}$$

$$= \pm \sqrt{0.36}$$

$$r = \pm 0.6$$

$$r = 0.6 \times 1$$

Since both the regression coefficients are +ve, r must be +ve