## REGRESSION :

### REGRESSION :

Regression is a mathematical measure of the average relationship between two or more variables interms of the original limits of the data.

### LINES OF REGRESSION :

(i) Equation of line x on y is

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

where regression coefficient, $b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}$

(ii) Equation of line y on x is

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

where regression coefficient, $b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

Correlation Coefficient $(P_{xy}) = \pm\sqrt{b_{xy} \cdot b_{yx}}$

NOTE: $\bar{x} = \frac{\Sigma x}{n}$ ; $\bar{y} = \frac{\Sigma y}{n}$

PROBLEMS :

1) From the following data, find

(i) the two regression equation.

(ii) Coefficient of correlation between marks in Economics & statistics.

(iii) Most likely marks in statistics when marks in economics are 30.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| X (Eco) : | 25 | 28 | 35 | 32 | 31 | 36 | 29 | 38 | 34 | 32 |
| Y (Stat) : | 43 | 46 | 49 | 41 | 36 | 32 | 31 | 30 | 33 | 39 |

Soln:

| $x$ | $y$ | $x-\bar{x}$ | $y-\bar{y}$ | $(x-\bar{x})^2$ | $(y-\bar{y})^2$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|---|
| 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 | 3 |
| 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 |

$\bar{x} : \frac{\Sigma x}{n} = 32$ $\bar{y} : \frac{\Sigma y}{n} = 38$

1 (i) Equation of line x on y:

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

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

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

$$\Rightarrow x + 0.2336y = 40.8768, \text{ Equation of line } x.$$

(ii) Equation of line y on x:

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

$$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = -\frac{93}{140} = -0.6642$$

$$y - 38 = -0.6642(x - 32)$$

$$0.6642x + y = 59.2544, \text{ Equation of line } y.$$

(2) Correlation Coefficient:

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

$$= \pm\sqrt{-0.2336 \times -0.6642}$$

$$= \sqrt{0.1551}$$

$$= 0.3938$$

(3) $y$ on $x = 30$

Equation of line $y$ on $x$ is $0.6642\,x + y = 59.2544$

$0.6642\,(30) + y = 59.2544$

$y = 39.3284$

8) Two lines of regressions are $8x - 10y + 66 = 0$, $40x - 18y - 214 = 0$. The var of $x$ is 9. Find

(i) mean value of $x$ and $y$

(ii) Correlation coefficient between $x$ & $y$

(iii) $E(x^2)$.

Soln:

(i)

$$8x - 10y + 66 = 0$$
$$40x - 18y - 214 = 0$$

$$40x - 50y = -330$$
$$-)\ 40x - 18y = 214$$

$$-32y = -544$$

$$y = 17 \quad (1)\ E(y) = 17$$

$$8x - 10(17) = -66$$

$$8x = 104$$

$$x = 13 \quad (2)\ E(x) = 13$$

(ii) Correlation coefficient:

For x:

$$40x - 18y = 214$$

$$40x = 18y + 214$$

$$x = 18/40\,y + 214/40 \Rightarrow b_{xy} = 18/40$$ (coeff of y does)

For y:

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

$$10y = 8x + 66$$

$$y = 8/10\,x + 66/10$$

$$\Rightarrow b_{yx} = 8/10$$

$\therefore$ Correlation coefficient $= \rho_{xy} = \pm\sqrt{b_{xy} \cdot b_{yx}}$

$$= \sqrt{18/40 \cdot 8/10} = 0.6$$

v) To find $E(x^2)$:

WKT $E(x) = 13$

Given $Var(x) = 9$

$$\Rightarrow E(x^2) - (E(x))^2 = 9$$

$$E(x^2) - (13)^2 = 9$$

$$E(x^2) = 178$$

3) If the equations of the two lines of regression of $y$ on $x$ and $x$ on $y$ are respectively, $7x-16y+9=0$; $5y-4x-3=0$, calculate the co-efficient of correlation, $\bar{x}$ & $\bar{y}$.

Soln:

$$7x-16y = -9$$
$$-4x+5y = 3$$

$$28x-64y = -36$$
$$-28x+35y = 21$$
$$-29y = -15$$
$$y = \frac{15}{29} \Rightarrow \bar{y} = \frac{15}{29}$$

put $y = \frac{15}{29}$

$$7x - 16\left(\frac{15}{29}\right) = -9$$
$$7x = -\frac{21}{29}$$
$$x = \frac{-3}{29} \Rightarrow \bar{x} = \frac{-3}{29}$$

$$7x - 16y = -9$$
$$7x = 16y - 9$$
$$x = \frac{16}{7}y - \frac{9}{7}$$
$$\Rightarrow b_{xy} = \frac{16}{7}$$

$-4x + 5y = 3$

$5y = 4x + 3$

$y = \frac{4}{5}x + \frac{3}{5}$

$\Rightarrow b_{yx} = \frac{4}{5}$

$\therefore$ Correlation coefficient : $\rho_{xy} = \pm\sqrt{b_{xy} \cdot b_{yx}}$

$= \sqrt{\frac{16}{7} \cdot \frac{4}{5}}$

$= 1.3522$

Now, $-4x + 5y = 3$

$4x = 5y - 3$

$x = \frac{5}{4}y - \frac{3}{4}$

$\Rightarrow b_{xy} = \frac{5}{4}$

$7x - 16y = -9$

$16y = 7x + 9$

$y = \frac{7}{16}x + \frac{9}{16}$

$\Rightarrow b_{yx} = \frac{7}{16}$

$\therefore$ Correlation coefficient : $\rho_{xy} = \pm\sqrt{b_{xy} \cdot b_{yx}}$

$= \sqrt{\frac{5}{4} \cdot \frac{7}{16}}$

$= 0.7395$