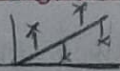
- **Regression:** Method of estimating value of one variable when value of other is known and they are correlated.

2 variables

- Lines of Regression for :-  y on x → x on y

1. Evaluation of data problems :-

y on x

equation: $y = a + bx$

~~Use normalised eq^n to~~

To find a & b, use normalised eq^n

$\Sigma y = aN + (\Sigma x) b$

$\Sigma xy = a \Sigma x + b \Sigma x^2$

x on y

$x = p + qy$

" p & q " "

$\Sigma x = pN + q \Sigma y$

$\Sigma xy = p \Sigma y + q \Sigma y^2$

2. Coefficients of Regression :-

given by b_{yx}

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

$$b_{yx} = \frac{r \sigma_x}{\sigma_x} = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2}$$

$d_x = \frac{x - a}{h}$ $\quad d_y = \frac{y - b}{k}$

$$b_{yx} = \frac{\Sigma d_x d_y - N \bar{d_x} \bar{d_y}}{\Sigma d_x^2 - N \bar{d_x}^2}$$

$$= \frac{\Sigma xy - N \bar{x} \bar{y}}{\Sigma x^2 - N \bar{x}^2}$$

given by b_{xy}

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

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

$d_x = \frac{x - a}{h}$ $\quad d_y = \frac{y - b}{h}$

$$b_{xy} = \frac{\Sigma d_x d_y - N (\bar{d_x})(\bar{d_y})}{\Sigma d_y^2 - N (\bar{d_y})^2}$$

$$= \frac{\Sigma xy - N \bar{x} \bar{y}}{\Sigma y^2 - N (\bar{y})^2}$$

Page No.
Date

- **Relation** : r is the geometric mean b/w b_{yx} and b_{xy}

1 $b_{yx} b_{xy} = r^2$

2 $\frac{b_{xy} + b_{yx}}{2} \geq r$

3 if $b_{xy} > 0$ then $b_{yx} > 0$
if $b_{xy} < 0$ then $b_{yx} < 0$
and vice-versa

4 if $b_{xy} > 1$ then $b_{yx} < 1$

5 if $r = \pm 1$ $\quad b_{xy} = 1/b_{yx}$

$0 \leq r^2 \leq 1$

$0 < r^2 \leq 1$

$b_{yx} b_{xy} \leq 1$

$b_{yx} \leq \frac{1}{b_{xy}}$

6 angle b/w lines of $x-y$

$(y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \xrightarrow{} m_1$; y on x

$(x-\bar{x}) = \frac{r\,\sigma_x}{\sigma_y}(y-\bar{y}) \longrightarrow (y-\bar{y}) = \frac{\sigma_y}{r\,\sigma_x}(x-\bar{x}) \xrightarrow{} m_2$

$\tan\theta = \frac{m_1 - m_2}{1+m_1 m_2} - \frac{1-r^2}{r}\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$

$r = 0 \quad \tan\theta = \infty \quad \theta = \pi/2$

$r = \pm 1 \quad \tan\theta = 0 \quad \theta = 0$

- **Questions** :

1 Find the eq^n of lines of regression and their correl^n coeff. for the following data. Estimate y for $x = 13$ and x for $y = 20$

x: 5 6 7 8 9 10 11

y: 11 14 14 15 12 17 18

	x	y	x^2	y^2	xy
	5	11	25	121	55
	6	14	36	196	84
	7	14	49	196	98
	8	15	64	225	120
	9	12	81	144	108
	10	17	100	289	170
	11	18	121	324	198
Σ	56	101	476	1495	833

$N = 7$ Normalisation :-

$y = a + bx$

$\Sigma y = aN + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$101 = a(7) + b(56)$

$833 = a(56) + b(476)$

$a = 7.28 \qquad b = 0.89$

$y = 7289 - x \times 0.89$

y on x :-

$b_{yx} = 0.89$

$x = p + qy$

$\Sigma x = pN + q\Sigma y$

$\Sigma xy = p\Sigma y + q\Sigma y^2$

$56 = p(7) + q(101)$

$833 = p(101) + q(1495)$

$q = 0.66 \qquad p = -1.564$

$x = 0.66y - 1.564$

x on y :-

$b_{xy} = 0.66$

$r = \sqrt{b_{xy}\, b_{yx}} = \sqrt{0.89 \times 0.66} = 0.766$

put $x = 13$ $\quad y = 7289 + 0.89 \times 13 = 7300.57$

put $y = 20$ $\quad x = 0.66 \times 20 - 1564 = -1550.8$

2 Find the angle b/w line of regression using data :-

$\sigma_x = 4 \qquad \sigma_y = 5 \qquad r = 0.6$

→ Let θ be angle b/w lines of regression

$$\tan\theta = \left(\frac{1 - r^2}{r}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right) = \left(\frac{1 - (0.6)^2}{0.6}\right)\left(\frac{4 \times 5}{16 + 25}\right)$$

$$= 1.06 \times 0.48 = 0.52$$

3 Given the following weight x and height y of 1000 people

$\bar{x} = 150$ lbs $\qquad \sigma_x = 20$ lbs

$\bar{y} = 68$ inches $\qquad \sigma_y = 2.5$ inches $\qquad$ y = height, x = weight

$r = 0.6$

John's weight is 200 lbs, find height

Smith is 5ft tall, find weight $\quad$ 60 inches

→ To find height using weight

$$(y - \bar{y}) = \frac{r\sigma_y}{\sigma_x}(x - \bar{x}) \rightarrow (y - 68) = 0.075\,(200 - 150)$$

$$y = 71.75$$

To find weight using height (regression x on y)

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

$(x-150) = 0.6 \times \frac{20}{2.5}(y-68)$

$x = 150 + 18 = 168$

4 Angle b/w lines of regression is $\tan^{-1} 3$, then find coefficient of correlation. Lines of regression: $6y = 5x + 90$, $15x = 8y + 130$, $\sigma_x^2 = 16$

Find $\bar{x}$ and $\bar{y}$, r, σ_y

→ Solve equation of lines of regression simultaneously

$5x - 6y + 90 = 0$

$15x - 8y - 130 = 0$

Assume $6y = 5x + 90$ as line of regression x on y

$5x = 6y - 90 \longrightarrow x = 6/5\, y - 90$

$b_{xy} = 6/5$

Assume $15x = 8y + 130$ as line of regression y on x

$8y = 15x - 130 \longrightarrow y = 15/8\, \cancel{y}\, x - 130$

$b_{yx} = 15/8$

$r = \sqrt{b_{xy} b_{yx}} = \sqrt{9/4} = 1.5$

Assume $6y = 5x + 90$ as line of regression y on x

$y = 5/6\, x + 90/6 \qquad b_{yx} = 5/6$

take $15x = 8y + 130$ as line of regression x on y

$x = 8/15\, y + 130/15 \qquad b_{xy} = 8/15 = 0.66$

Idk how tf to get $\bar{x}$ and $\bar{y}$ and σ_y
I ain't understanding shit
If you know, welp