Co-relation & Regression

→ **Correlation**: Deals with assoc. betⁿ 2 or more variables

→ The deg. of relⁿ b/w variables under consideration is measured through correlation analysis

→ Measure of correlatⁿ: "Correlatⁿ coeff" or "Correlatⁿ index"

summarizes in one figure the dirⁿ & degree of correlation

① **Pos. correlation**:

If both variables vary in same dirⁿ. (Ht. & wt., Stock price & Prod. in comp.

② **Negative Correlation**:

If both variables vary in opp. dirⁿ.

Correlation coeff. (r): ~~-ve~~ $-1 \leq r \leq 1$

$r = 1$ → Perfect +ve corr.

$r = -1$ → — " — -ve — " —

$r = 0$ → no corr.

Values near +1 : Strong +ve corr.
— 〃 — −1 : — 〃 — −ve corr.

$$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} \qquad E(X) = \frac{\Sigma x}{n}$$

$$= \frac{cov(u,v)}{\sigma_x \cdot \sigma_y} \qquad \text{mean}$$

$$cov.(x,y) = E(XY) - E(X) \cdot E(Y)$$

$$= \frac{\Sigma xy}{n} - \frac{\Sigma x}{n} \cdot \frac{\Sigma y}{n}$$

$$\sigma_x = \sqrt{(E(X^2)) - (E(X))^2}$$

$$\sigma_y = \sqrt{E(Y^2) - (E(Y))^2}$$

Q1 Cal. correlat^n coeff. for the foll. data.

x	y	x^2	y^2	xy
9	15	81	225	135
8	16	64	256	128
7	14	49	196	98
6	13	36	169	78
5	11	25	121	55
4	12	16	144	48
3	10	9	100	30
2	8	4	64	16
1	9	1	81	9
45	108	285	1356	597

$\Sigma xy = 597$
$\Sigma x = 45$
$\Sigma y = 108$
$\Sigma x^2 = 285$
$\Sigma y^2 = 1356$

$n = 9$

$$\text{cov}(x,y) = \frac{\Sigma xy}{n} - \frac{\Sigma x}{n} \cdot \frac{\Sigma y}{n}$$

$$= \frac{597}{9} - \frac{\cancel{45}^{5}}{\cancel{9}} \cdot \frac{108}{9}$$

$$= 66.33 - 60$$

$$= 6.33$$

$$\sigma x = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{285}{9} - \left(\frac{\cancel{45}^{5}}{\cancel{9}}\right)^2} = \sqrt{31.666 - 25}$$

$$\boxed{\sigma x = 2.581}$$

$$\sigma y = \sqrt{\frac{1356}{9} - \left(\frac{108}{9}\right)^2} = \sqrt{150.666 - 144}$$

$$= \sqrt{6.666}$$

$$\boxed{\sigma y = 2.581}$$

$$\therefore r = \frac{6.33}{(2.581)^2} = 0.95$$

If q. has big values
x

	x	y	u = x-68	v = y-69	u^2	v^2	uv
	65	67					
	66	68					
	67	65					
	67	58					
	68	72					
	69	72					
	70	69					
	72	71					

Cal. the same way as we did earlier

Rank Correlation

$$\boxed{\rho = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}}$$

Q. The marks obt. in maths & Phy. are given below find rank correlation

Order of values of x or Rank of each var.

Maths marks (x)	Phy. marks (y)	R_x	R_y	$d^2 = (R_x - R_y)^2$
10	30	9	9	0
15	42	5	3	4
12	45	8	2	36
17	46	3	1	4
13	33	7	8	1
16	34	4	7	9
24	40	1	4	9
14	35	6	6	0
82	39	2	5	9
				$\Sigma d^2 = 72$

$\cancel{\rho = 1 - 6\,(7}$

$$\rho = 1 - \frac{6\,\Sigma d^2}{n(n^2-1)}$$

$$\rho = 1 - \frac{\cancel{6}^{\,2} \times \cancel{72}^{\,24}}{\cancel{9}(81-1)} = 1 - \frac{\cancel{2} \times \cancel{24}6}{\cancel{80}\cancel{4}0\,10}$$

$$\boxed{\rho = 0.4}$$

Q.2 Ten competitors in a beauty contest got mks by 3 judges in the foll. order :

1st	1	6	5	10	3	2	4	9	7	8
2nd	3	5	8	4	7	10	2	1	6	9
3rd	6	4	9	8	1	2	3	10	5	7

Use rank corr. coeff. to decide which pair of judges have nearest approach to common test of beauty.

x	y	z	R_x	R_y	R_z	d_{xy}^2	d_{yz}^2	d_{xz}^2
1	3	6	10	8	5	4	9	25
6	5	4	5	6	7	1	1	4
5	8	9	6	3	2	9	1	16
10	4	8	1	7	3	36	16	4
3	7	1	8	4	10	16	36	4
2	10	2	9	1	9	64	64	0
4	2	3	7	9	8	4	1	1
9	1	10	2	10	1	64	81	1
7	6	5	4	5	6	1	1	4
8	7	7	3	2	4	1	4	1

$$\rho_{xy} = 1 - \frac{6\sum d_{xy}^2}{n(n^2-1)} = \cancel{1-\frac{6(60)}{10(99)}} \; 1 - \frac{6(200)}{10\times 99}$$

$$= -0.212$$

$$\rho_{y2} = 1 - \frac{6\sum d^2_{y2}}{n(n^2-1)} = 1 - \frac{6(216)}{10\times 99} = -0.297$$

$$\rho_{x2} = 1 - \frac{6\sum d^2_{x2}}{n(n^2-1)} = 1 - \frac{6(64)}{10\times 99} = 0.636$$

As we can see ρ_{x2} is +ve, hence

1st & 3rd judge have the nearest approach.

~~When~~

Rep. ranks (For repeated ranks)

$$\boxed{\rho = 1 - \frac{6(\sum d^2 + F)}{n(n^2-1)}} \qquad \boxed{F = \frac{m(m^2-1)}{12}}$$

m = no. of repetitions

While writing ranks of rep. no's
Add the ranks & get average.

Q Obtain the rank ~~co~~ corr. coeff.

	x	y	R_x	R_y	$d^2=(R_x-R_y)^2$
1	68	62	4	5	1
2	64	58	6	7	1
3	75	68	2.5	3.5	1
4	50	45	9	10	1
5	64	07	6	1	25
6	80	60	1	6	25
7	75	68	2.5	3.5	1
8	40	48	10	9	1
9	55	50	8	8	0
10	64	74	6	2	16

$\Sigma d^2 = 72$

$$\rho = 1 - \frac{6(\Sigma d^2 + F)}{n(n^2-1)}, \quad F = f_{75} + f_{64} + f_{68}$$

$$= \frac{\cancel{2}(4-1)}{\cancel{12}\,6} + \frac{\cancel{3}(9-1)}{\cancel{12}\,4} + \frac{\cancel{2}(4-1)}{\cancel{12}\,6}$$

$$\rho = 1 - \frac{\overset{2}{\cancel{6}}(72+3)}{10(\underset{33}{\cancel{99}})}$$

$$= \frac{1}{2} + 2 + \frac{1}{2}$$

$$\rho = 1 - \frac{2 \times 75}{10 \times 33}$$

$$= 3$$

$$\boxed{\rho = 0.545}$$

Regressions:

If the curve of regression is a st. line then it is said to be line of regression

If the curve of regression is not a st. line then that regression is said to be curvilinear or non-linear regression

Regression of line y on x

① $\boxed{y-\bar{y} = b_{yx}(x-\bar{x})}$ $\boxed{\text{Slope}: b_{yx}}$

reg. coeff. y on x

both lines intersect at $\bar{x}, \bar{y}$

$\boxed{b_{yx} = r\frac{\sigma_y}{\sigma_x}}$ $x \to$ indep. $y \to$ dep.

② Regression of line x on y

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

$\boxed{b_{xy} = r\frac{\sigma_x}{\sigma_y}}$ ~~x → d~~ $y \to$ indep. $x \to$ dep.

$\boxed{\text{Slope} = \frac{1}{b_{xy}}}$

Q.1) Cal. the coeff. of correlation & obtain the line of regression for the foll. data.

x	y	$u = x-5$	$v = y-12$	u^2	v^2	uv
1	9	-4	-3	16	9	12
2	8	-3	-4	9	16	12
3	10	-2	-2	4	4	4
4	12	-1	0	1	0	0
5	11	0	-1	0	1	0
6	13	1	1	1	1	1
7	14	2	2	4	4	4
8	16	3	4	9	16	12
9	15	4	3	16	9	12
45	108	$\Sigma u = 0$	$\Sigma v = 0$	60	60	57

$$\bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5 \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{108}{9} = 12$$

$$r_{uv} = \frac{\text{cov}(u,v)}{\sigma_u \, \sigma_v}$$

~~cov(x,~~

$\text{cov}(u,v) = 6.33$

$$\sigma_u = \sqrt{\frac{\Sigma u^2}{n} - \left(\frac{\Sigma u}{n}\right)^2} = 2.58$$

$\sigma_v = 2.58$

$$r_{uv} = \frac{6.33}{(2.581)^2} = 0.95$$

$\underline{x \text{ on } y}$

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

As we have ans.

$\bar{u} = \bar{x} - 5 \qquad \bar{x} = \bar{u} + 5$

$\bar{v} = \bar{y} - 12 \qquad \bar{y} = \bar{v} + 12$

$= \underline{x \text{ on } y}$

$$x - (\bar{u} + 5) = r_{uv} \frac{\sigma u}{\sigma v}(y - (\bar{v} + 12))$$

$$x - 5 = 0.95(y - 12)$$

$$\boxed{x = 0.95y - 6.4}$$

$\underline{y \text{ on } x}$

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

$$= y - (\bar{v} + 12) = \cancel{b_y} \, r_{uv} \frac{\sigma v}{\sigma u}(x - (\bar{u} + 5))$$

$$= y - 12 = 0.95(x - 5)$$

$$\boxed{y = 0.95x + 7.25}$$

Q1) In a partially destroyed laboratory record of an analysis of a corr. data the following results only are legible

$Var(x) = 9$, Regression eqn

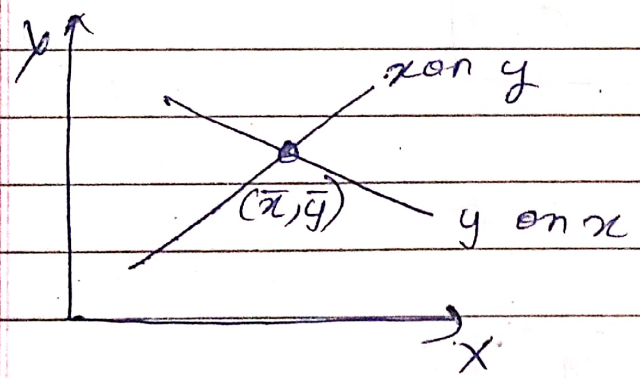
$\cancel{8x - 10y}$ $8x - 10y + 66 = 0$
$40x - 18y = 214$

(1) find the mean of x & y $(\bar{x}, \bar{y})$

(2) Coeff. of corr. (r)

(3) S.D. of y (σ_y)

Remember



$8\bar{x} - 10\bar{y} + 66 = 0$
$40\bar{x} - 18\bar{y} - 214 = 0$

On solv.
$\bar{x} = 13$
$\bar{y} = 17$

For x on y | y on x

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

$b_{xy} = r \times \frac{\sigma_x}{\sigma_y}$ | $b_{yx} = r \times \frac{\sigma_y}{\sigma_x}$

$$b_{xy} \times b_{yx} = r^2$$

$$\boxed{r = \pm\sqrt{b_{xy} \times b_{yx}}}$$

1. b_{xy} & $b_{yx} \geq 0$

   $r = +ve$

2. b_{yx} & $b_{xy} \leq 0$

   $r = -ve$

Now you have to do treal & error method

$8x-10y+66=0$

Try x on y

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

$b_{xy} = \frac{10}{8}$

$40x-18y=214$

Try y on x

$18y = 40x-214$

$y = \frac{40x}{18} - \frac{214}{16}$

$b_{yx} = \frac{40}{18}$

Page No.
Date.

$$\sqrt{\frac{10}{8}\times\frac{40}{18}} > 1$$

$$\cancel{\sqrt{\frac{10}{84}\times\frac{4623}{18}} > 1}$$

$$r = \sqrt{b_{xy}\cdot b_{yx}} = \cancel{\sqrt{\frac{18}{8}\times\frac{405}{18}}}$$

$r > 1$ ~~by~~ but $-1 \le r \le +1$

∴ We need to reverse x on y & y on x

$8x - 10y + 66 = 0$

Try y on x

$\cancel{8x - 10y = 66}$

$\cancel{x = \frac{10}{8}y - 66}$

$\cancel{b_{xy}}$

$\cancel{b_{yx}}$

$10y = 8x + 66$

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

$$\boxed{b_{yx} = \frac{8}{10} = \frac{4}{5}}$$

$40x = 18y + 214$

Try x on y

$x = \frac{18}{40}y + \frac{214}{40}$

$\cancel{b_{xy} = \frac{9}{23}}$ $b_{xy} = \frac{18}{40}$

$r = \sqrt{b_{xy}\, b_{yx}}$

$$r = \sqrt{\frac{4}{5}\times\frac{18}{40}}$$

$$\boxed{r = \frac{3}{5} = 0.6}$$

$\Rightarrow$ $\sigma_x^2 = 9$ $\sigma_x = 3$ find σ_y

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

$$\cancel{\frac{18}{8} = \frac{0.6 \times 3}{\sigma_y}} \qquad \frac{18}{40} = \frac{0.6 \times 3}{\sigma_y}$$

$$\cancel{\sigma_y = \frac{1.8 \times 8}{10}} \qquad \sigma_y = \frac{40 \times \cancel{0.6}^{0.1} \times \cancel{3}}{\cancel{18}_{6}}$$

$$\boxed{\sigma_y = 4}$$

Q Lines of regression:

$$2x + 3y = 10 \ \& \ 4x + 5y = 18$$

Decide which one is x on y & which one is y on x.

Given $x = 5$, find y
$y = 2$, find x

$2x + 3y = 10$	$4x + 5y = 18$
x on y	y on x
$x = \frac{-3}{2} y + 5$	$y = \frac{-4}{5} x + \frac{18}{5}$
$b_{xy} = \frac{-3}{2}$	$b_{yx} = \frac{-4}{5}$

$$r = \pm\sqrt{\left(\frac{-3}{2}\right)\left(\frac{-4}{5}\right)} = \sqrt{\frac{12}{10}} = \sqrt{1.2} > 0$$

Not possible

$2x + 3y = 10$	$4x + 5y = 18$
y on x	x on y
$y = -\frac{2}{3}x + \frac{10}{3}$	$x = -\frac{5}{4}y + \frac{18}{4}$
$b_{yx} = -\frac{2}{3}$	$b_{xy} = -\frac{5}{4}$

$$r = -\sqrt{\left(\frac{-2}{3}\right)\left(\frac{-5}{4}\right)} = -\sqrt{\frac{10}{12}}$$ ✓

Note: When x is given
& y asked choose
y on x eqn & vice-versa

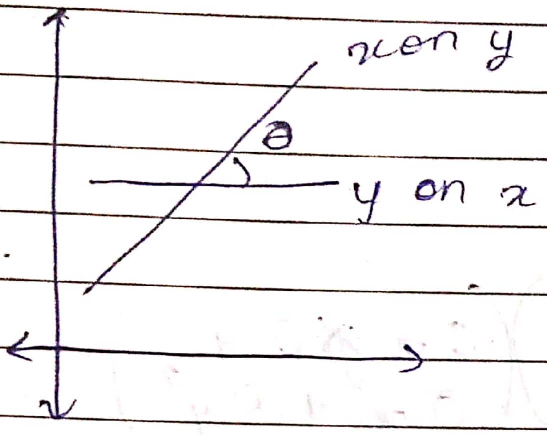
At $x = 5$	At $y = 2$
$y = \frac{-10}{3} + \frac{10}{3}$	$x = \frac{-5}{2} + \frac{9}{2}$
$y = 0$	$x = 2$

Q Show that ~~that~~ tan θ, the acute ∠ betⁿ 2 lines ~~is~~ of reg. is given by

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

Interpret the case when $r=0$ & $r=\pm 1$

Soln:



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

$$\cancel{x-\bar{x} = r\frac{\sigma_x}{\sigma}}$$

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

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

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

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

$$\boxed{m_1 = \frac{\sigma_y}{r.\sigma_x}}$$

$$\boxed{m_2 = r\frac{\sigma_y}{\sigma_x}}$$

$$\tan\theta = \pm\left(\frac{m_1-m_2}{1+m_1m_2}\right)$$

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

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

(1) when $r=0$

$$\tan\theta = \frac{(1-0)}{0}\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right) = \infty$$

$$\tan\theta = \infty$$

$$\boxed{\theta = \pi/2}$$

(2) when $r=\pm 1$

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

$\therefore \tan\theta = 0$

$\therefore \theta = 0$ or π

$r=1$; $r=-1$