Unit-5

Correlation + Regression

- Univariate distribution (1^{st} unit)
- **Bivariate distribution**

  $X \rightarrow$ Independent r.v.

  $Y \rightarrow$ Dependent r.v.

+ve correlation $\begin{cases} X\uparrow & Y\uparrow \\ X\downarrow & Y\downarrow \end{cases}$

-ve correlation $\begin{cases} X\uparrow & Y\downarrow \\ X\downarrow & Y\uparrow \end{cases}$

* Correlation $\Rightarrow$ lineary related

$r \Rightarrow (-1 \text{ to } 1)$

$r = 0 \Rightarrow$ uncorrelated

$r = -1 \Rightarrow$ perfect negatively correlated

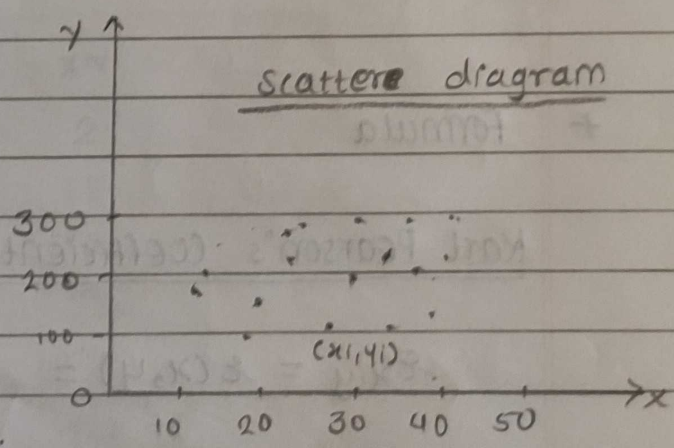
$r = 1 \Rightarrow$ perfect positively correlated

$r = (0 \text{ to } 1) \Rightarrow$ positively correlated

$r = (-1 \text{ to } 0) \Rightarrow$ negatively correlated.

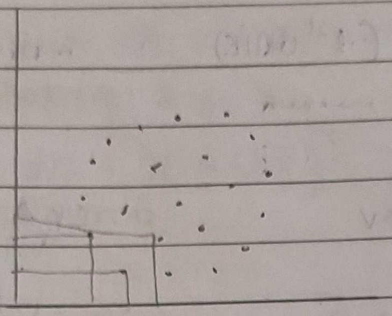
Q.

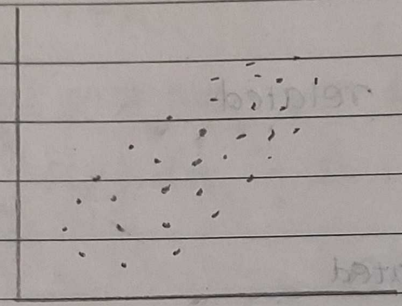
Price X	Demand Y
17	300
19	250
20	200
15	190

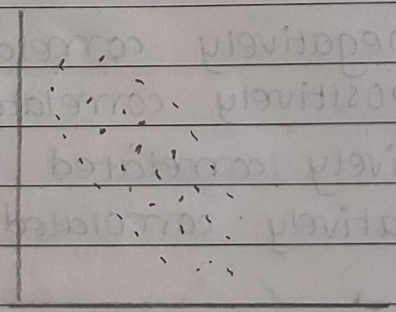


scattere diagram

n = no. of observations.







* Formula

<u>Karl Pearson's Coefficient of correlation</u>

$$r_{x,y} = r(x,y) = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y}$$

$$\mathrm{cov}(x,y) = \frac{1}{n}\Sigma(x_i-\bar{x})(y_i-\bar{y}) \Rightarrow \frac{1}{n}\Sigma x_i y_i - \bar{x}\bar{y}$$

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma(x_i-\bar{x})^2} = \sqrt{\frac{1}{n}\Sigma x_i^2 - \bar{x}^2}$$

$$\sigma_y = \sqrt{\frac{1}{n}\Sigma(y_i-\bar{y})^2} = \sqrt{\frac{1}{n}\Sigma y_i^2 - \bar{y}^2}$$

* Effect of change of origin

NOTE: coefficient of correlation is independent of change of origin

$$\boxed{r_{xy} = r_{uv}}$$

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

$$cov(u,v) = \frac{1}{n} \Sigma u_i v_i - \bar{u}\bar{v}$$

$$\sigma_u = \sqrt{\frac{1}{n} \Sigma u_i^2 - \bar{u}^2}$$

$$\sigma_v = \sqrt{\frac{1}{n} \Sigma v_i^2 - \bar{v}^2}$$

Q. Calculate coefficient of correlation bet^n x + y.

	x	y	x^2	y^2	xy
	1	2	1	4	2
	3	6	9	36	18
	4	8	16	72	32
	5	10	25	100	50
	7	14	49	196	98
	8	16	81	256	128
	10	20	100	400	200
Total	38	76	264	1056	528

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{38}{7} = 5.428$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{76}{7} = 10.857$$

$$\text{cov}(x,y) = \frac{1}{n}\Sigma xy - \bar{x}\bar{y}$$

$$= \frac{1}{7} \times (528) - (5.428 \times 10.857)$$

$$= 16.496$$

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma x_i^2 - \bar{x}^2}$$

$$= \sqrt{\frac{1}{7}(264) - (5.428)^2}$$

$$= 2.872$$

$$\sigma_y = \sqrt{\frac{1}{n}\Sigma y_i^2 - \bar{y}^2}$$

$$= \sqrt{\frac{1}{7}(1056) - (10.587)^2}$$

$$= 5.742$$

$$r(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$$

$$= \frac{16.496}{(2.872)(5.742)}$$

$$= 1$$

i.e. x + y are perfectly positively correlated.

Date
Page

Q. calculate coeff of corr^n between height of fathers + their sons.

X	Y	$u_i = x_i - 67$	$v_i = y_i - 72$	u^2	v^2	uv
65	67	-2	-5	4	25	10
66	68	-1	-4	1	16	4
67	65	0	-7	0	49	0
$\boxed{67}$	68	0	-4	0	16	0
68	$\boxed{72}$	1	0	1	0	0
69	72	2	0	4	0	0
70	69	3	-3	9	9	-9
72	71	5	-1	25	1	-5
Total		8	-24	44	116	0

$$\bar{U} = \frac{1}{n}\Sigma u_i = \frac{8}{8} = 1$$

$$\bar{V} = \frac{1}{n}\Sigma v_i = \frac{-24}{8} = -3$$

$$\text{cov}(u,v) = \frac{1}{n}\Sigma uv - \bar{U}\bar{V}$$

$$= \frac{1}{8} \times 0 - (1)(-3)$$

$$= 3$$

$$\therefore \sigma_u = \sqrt{\frac{1}{n}\Sigma u_i^2 - \bar{u}_i^2} = \sqrt{\frac{1}{8} \times 44 - (1)^2} = 2.121$$

$$\therefore \sigma_v = \sqrt{\frac{1}{n}\Sigma v_i^2 - \bar{v}_i^2} = \sqrt{\frac{1}{8} \times 116 - (-3)^2} = 2.345$$

$$r(u,v) = \frac{\mathrm{cov}(u,v)}{\sigma_u \sigma_v}$$

$$= \frac{3}{2.121 \times 2.345}$$

$$= 0.6031$$

$\because r(u,v) = 0.6031$

$\therefore$ X & Y are positively correlated.

* Rank correlation

Spearman's Rank correlation coefficient

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

n = No. of observations

d = diff. in Rank of X & Rank of Y

- Repeated Rank

$$\rho = 1 - \frac{6\left[\Sigma d_i^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \cdots\right]}{n(n^2-1)}$$

m_i = No. of times rank is repeated

Q. 10 students got the following % of marks in mathematics & physics. Find the coeff. of Rank correlation.

Sol^n:

X Marks in Mathematics	Y Marks in Physics	Rank of X	Rank of Y	di	di^2
8	84	10	3	7	49
36	51	7	8	-1	1
98	91	1	1	0	0
25	60	9	6	3	9
[illegible]	68		4	0	0
75	62	4	5	-2	4
82	86	3	2	0	0
92	58	2	7	-1	1
62	35	6	10	-5	25
65	49	5	9	-1	1
35		8			
					90

$\therefore \quad \rho = 1 - \dfrac{6 \Sigma di^2}{n(n^2-1)}$

$= 1 - \dfrac{6(90)}{10(10^2-1)}$

$= 1 - \dfrac{540}{10 \times 99}$

$= 1 - \dfrac{6}{11}$

$= \dfrac{5}{11} = 0.\overline{45}$

≈ 0.455

Date
Page

Q. From the following table, calculate Rank correlation coefficient:

X	Y	Rank of X	Rank of Y	di	di^2
48	13	3	5.5	-2.5	6.25
33	13	5	5.5	-0.5	0.25
40	24	4	1	3	9.00
9	6	10	8.5	1.5	2.25
16	15	8	4	4	16.00
16	4	8	10	-2	4.00
65	20	1	2	-1	1.00
24	9	6	7	-1	1.00
16	6	8	8.5	-0.5	0.25
57	19	2	3	-1	1.00
					41

Here, $m_1 = 3$ [$\because$ 16 is repeated 3 times]
$m_2 = 2$ [$\because$ 13 is repeated 2 times]
$m_3 = 2$ [$\because$ 6 is repeated 2 times]

$$\therefore \rho = 1 - \frac{6\left[\Sigma di^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \frac{m_3(m_3^2-1)}{12}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[41 + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12}\right]}{10(10^2-1)}$$

$$= 1 - \frac{6\left[41 + \frac{24}{12} + \frac{6}{12} + \frac{6}{12}\right]}{10 \times 99}$$

$$= 0.733$$

Q. 10 competitors in a beauty contest are ranked by three judges in the following order.

x_i 1st Judge	y_i 2nd Judge	z_i 3rd Judge	$d_i = x_i - y_i$	d_i^2	$d_2 = y_i - z_i$	d_2^2	$d_3 = x_i - z_i$	d_3^2
1	3	6	-2	4	-3	9	-5	25
6	5	4	1	1	1	1	2	4
5	8	9	-3	9	-1	1	-4	16
10	4	8	6	36	-4	16	2	4
3	7	1	-4	16	6	36	2	4
2	10	2	-8	64	8	64	0	0
4	2	3	2	4	-1	1	1	1
9	1	10	8	64	-9	81	-1	1
7	6	5	1	1	1	1	2	4
8	9	7	-1	1	2	4	1	1
				200		214		60

Use Rank correlation coefficient to determine which pair of Judges has the nearest approach to common interest in beauty.

$$\rho_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(200)}{10(99)} = 1 - 1.212 = -0.212$$

$$\rho_{yz} = 1 - \frac{6\sum d_2^2}{n(n^2-1)} = 1 - \frac{6(214)}{10(99)} = 1 - 1.296 = -0.296$$

$$\rho_{xz} = 1 - \frac{6\sum d_3^2}{n(n^2-1)} = 1 - \frac{6(60)}{10(99)} = 1 - 0.364 = 0.636$$

$\because$ ρ_{xz} is maximum

$\therefore$ 1st & 3rd Judge have common interest in beauty.

Date
Page

Q. From the following data of marks obtained by 10 students in accountancy and statistics. Calculate Spearman's Rank correlation coefficient.

Roll no	Marks in Accountancy	Marks in Statistics	Rank of Accountancy	Rank of statistics	di	di^2
1	20	52	10	8	2	4
2	25	50	8	9.5	-1.5	2.25
3	60	55	4	7	-3	9
4	45	50	6	9.5	-3.5	12.25
5	80	60	1	6	-5	25
6	25	70	8	4	4	16
7	55	72	5	3	2	4
8	65	78	3	2	1	1
9	25	80	8	1	7	49
10	75	63	2	5	-3	9
						131.5

Here $m_1 = 3$ [$\because$ 25 is repeated 3 times]

$m_2 = 2$ [$\because$ 50 is repeated 2 times]

$$\therefore \quad S = 1 - \frac{6\left[\Sigma di^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[131.5 + \frac{3(8)}{12} + \frac{2(3)}{12}\right]}{10(10^2-1)}$$

$$= 1 - \frac{6(131.5+2+0.5)}{990}$$

$$= 1 - \frac{6(134)}{990}$$

$$= 1 - 0.812 = 0.188$$

Date
Page

⇒ covariance $(-\infty, \infty)$

+ve — +ve LR — $x\uparrow\ y\uparrow$, $x\downarrow\ y\downarrow$

-ve — -ve LR — $x\downarrow\ y\uparrow$, $x\uparrow\ y\downarrow$

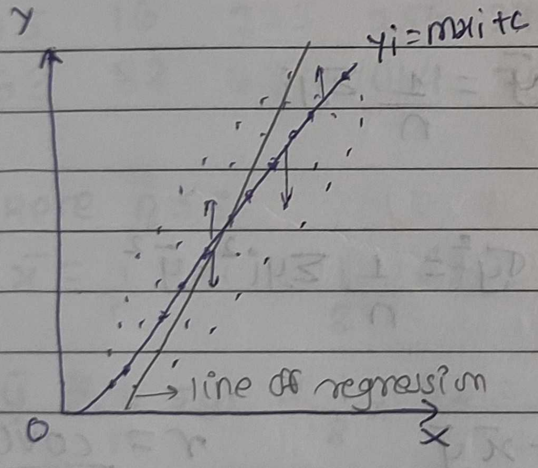
⇒ correlation

(-1 to 1) → Rank correlation

(-1 to 0)

(0 to 1)

⇒ Regression



Error

$\varepsilon_i = (y_i - y_i)$

$= (mx_i + c - y_i)$

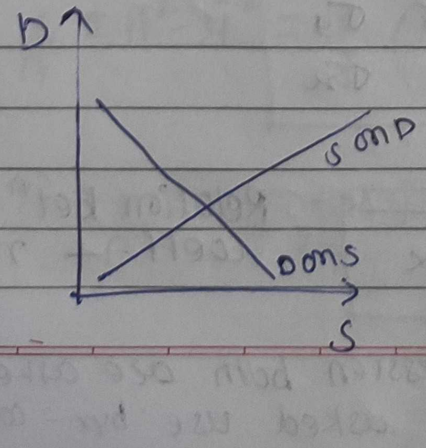
Total error = 0

$\Sigma \varepsilon_i^2$

$f(x, y) = \Sigma \varepsilon_i^2$

$= \Sigma (mx_i + c - y_i)^2$

supply	Demand
x	Y
S↑	D↓
D↑	S↑



$y = 3x + 7 \Rightarrow y$ on x
$x = 0.66y - 2.3 \Rightarrow x$ on y

Date
Page

* Regression

1] Line of regression X on Y

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

b_{xy} = regression coefficient of X on Y

2] Line of regression Y on X

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

b_{yx} = regression coefficient of Y on X.

$$\bar{x} = \frac{1}{n}\Sigma x_i \qquad \bar{y} = \frac{1}{n}\Sigma y_i$$

$$\sigma_x^2 = \frac{1}{n}\Sigma x_i^2 - \bar{x}^2 \qquad \sigma_y^2 = \frac{1}{n}\Sigma y_i^2 - \bar{y}^2$$

$$\mathrm{cov}(X,Y) = \frac{1}{n}\Sigma x_i y_i - \bar{x}\bar{y} \qquad r = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y}$$

$$\boxed{b_{xy} = \frac{\mathrm{cov}(X,Y)}{\sigma_y^2} = r\frac{\sigma_x}{\sigma_y}}$$

$$\boxed{b_{yx} = \frac{\mathrm{cov}(X,Y)}{\sigma_x^2} = r\frac{\sigma_y}{\sigma_x}}$$

$$\boxed{r = \sqrt{b_{xy} \cdot b_{yx}}} \Rightarrow$$ Relation betn correlation coeff. + regression coeff.

NOTE: ▸ If correlation + regression both are asked use $r\frac{\sigma_y}{\sigma_x}$ + b_{yx}
▸ If only regression is asked use $b_{yx} = \mathrm{cov}(X,Y)/\sigma_x^2$

Q. Obtain Regression types from the following data. Also Estimate I] Y when X = 6
II] X when Y = 20

	X	Y	X^2	Y^2	XY
1	2	2	4	4	4
2	3	5	9	25	15
3	5	8	25	64	40
4	7	10	49	100	70
5	9	12	81	144	108
6	10	14	100	196	140
7	12	15	144	225	180
8	15	16	225	256	240
	63	82	637	1014	797

Here n = 8

$\bar{x} = \frac{1}{n} \Sigma x_i = \frac{1}{8} \times 63 = 7.875$

$\bar{y} = \frac{1}{n} \Sigma y_i = \frac{1}{8} \times 82 = 10.25$

$\sigma_x^2 = \frac{1}{n} \Sigma x_i^2 - \bar{x}^2 = \frac{1}{8}(637) - (7.87)^2$

$= 17.68$

$\sigma_y^2 = \frac{1}{n} \Sigma y_i^2 - \bar{y}^2 = \frac{1}{8}(1014) - (10.25)^2 = 21.68$

$COV(X, Y) = \frac{1}{n} \Sigma x_i y_i - \bar{x}\bar{y} = \frac{1}{8}(797) - (7.87)(10.25)$

$= 18.95$

$$b_{yx} = \frac{\text{cov}(x,y)}{\sigma_x^2} = \frac{18.92}{17.68} = 1.07$$

$$b_{xy} = \frac{\text{cov}(x,y)}{\sigma_y^2} = \frac{18.92}{21.68} = 0.87$$

$\therefore$ Line of Regression Y on X

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

$y - 10.25 = 1.07(x - 7.87)$

$y = 1.07x - 8.0999 + 10.25$

$\boxed{y = 1.07x + 2.16}$

$\therefore$ Line of Regression X on Y

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

$x - 7.87 = 0.87(y - 10.25)$

$x = 0.87y - 8.9175 + 7.87$

$\cancel{x = 0.87y - 2.384}$

$\boxed{x = 0.87y - 1.0475}$

I] Estimate Y when $x = 6$

$y = 1.07x + 2.16$

$= 1.07(6) + 2.16$

$\boxed{y = 8.58}$

II] Estimate X when $Y = 20$

$x = 0.87y - 1.0475$

$x = 0.87(20) - 1.0475$

$\boxed{x = 16.36}$

Q. Obtain regression types for the following data

x	y	x^2	y^2	xy
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
[illegible]	[illegible]			
30	40			

Q. The table below gives the respective heights x and y of a sample of 10 fathers & their sons.

1] Find regression line of y on x
2] Find regression line of x on y
3] Estimate son's Height if father's height is 65 inches.
4] Estimate father's Height if son's height is 60 inches.
5] Compute correlation coeff. betn x & y
6] Find angle betn regression lines.

Height of father x (inches)	Height of son y (inches)	$U = x-A$	$V = y-B$	U^2	V^2	UV
65	68	-3	0	9	0	0
63	66	-5	-2	25	4	10
67	$\boxed{68}$	-1	0	1	0	0
$\boxed{68}$ 64	65	-4	-3	16	9	12
64 $\boxed{68}$	69	0	1	0	1	0
70 62	66	-6	-2	36	4	12
66 70	68	2	0	4	0	0
68 66	65	-2	-3	4	9	6
67 68	71	0	3	0	9	0
67	67	-1	-1	1	1	1
		-20	-7	96	37	41

$\bar{U} = \frac{\Sigma u_i}{n} = \frac{-20}{10} = -2$

$\bar{V} = \frac{\Sigma V}{n} = \frac{-7}{10} = -0.7$

$\sigma_U^2 = \frac{1}{n} \Sigma U^2 - \bar{U}^2 = \frac{1}{10}(96) - (-2)^2 = 9.6 - 4 = 5.6$

$\sigma_V^2 = \frac{1}{n} \Sigma V^2 - \bar{V}^2 = \frac{1}{10}(37) - (-0.7)^2 = 3.7 - 0.49 = 3.21$

$$cov(U,V) = \frac{1}{n}\Sigma UV - \bar{U}\bar{V} = \frac{1}{10}(41) - (-2)(-0.7)$$
$$= 4.1 - 1.4$$
$$= \cancel{2.7}\ 2.7$$

$$b_{UV} = \frac{cov(U,V)}{\sigma_V^2} = \frac{2.7}{3.21} = 0.84$$

$$b_{VU} = \frac{cov(U,V)}{\sigma_U^2} = \frac{2.7}{5.6} = 0.48$$

5] $r_{UV} = \sqrt{b_{UV} \cdot b_{VU}} = \sqrt{(0.84)(0.48)} = 0.635$

$$\boxed{\begin{aligned} r_{xy} &= r_{uv} = 0.635 \\ b_{xy} &= b_{uv} = 0.84 \\ b_{yx} &= b_{vu} = 0.48 \end{aligned}}$$

$\bar{x} = A + \bar{U} = 68 + (-2) = 66$

$\bar{y} = B + \bar{V} = 68 + (-0.7) = 67.3$

1] Line of Regression Y on X

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

$y - 67.3 = 0.48(x - 66)$

$y - 67.3 = 0.48x - 31.68$

$y = 0.48x + 35.62$

~~y = 0.48~~

2] Line of Regression x on y.

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

$x - 66 = 0.84(y - 67.3)$

$x - 66 = 0.84y - 56.53$

$x = 0.84y + 9.47$

3] $x = 65$

$\therefore y = 0.48x + 35.62$

$y = 0.48(65) + 35.62$

$\boxed{y = 66.82}$

4] $y = 60$

$\therefore x = 0.84y + 9.47$

$= 0.84(60) + 9.47$

$\boxed{x = 59.87}$

6] $\tan\theta = \dfrac{1 - r^2}{|r|} \cdot \dfrac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$

$= \dfrac{1 - 0.403}{0.635} \times \dfrac{2.36 \times 1.79}{5.6 + 3.21}$

$= \dfrac{0.597 \times 4.224}{0.635 \times 8.81}$

$\tan\theta = 0.4507$

$\theta = \tan^{-1}(0.4507)$

$\boxed{\theta = 24.26^\circ}$

Date
Page

Q. The following are marks obtained by 10 students in statistics & economics

Marks in economics (X)	Marks in statistics (Y)	U $X-A$	V $Y-B$	U^2	V^2	UV
25	43	-7	11	49	121	-77
28	46	-4	14	16	196	-56
35	49	3	17	9	289	51
$\boxed{32}$	41	0	9	0	81	0
31	36	-1	4	1	16	-4
36	$\boxed{32}$	4	0	16	0	0
29	31	-3	-1	9	1	3
38	30	6	-2	36	4	-12
34	33	2	1	4	1	2
32	39	0	7	0	49	0
		0	60	140	758	-93

1. Obtain regression coeff. to estimate marks in statistics if marks in economics are 30

$$\bar{U} = \frac{\Sigma U}{n} = \frac{0}{10} = 0$$

$$\bar{V} = \frac{\Sigma V}{n} = \frac{60}{10} = 6$$

$$\sigma_x^2 = \sigma_U^2 = \frac{1}{n}\Sigma U^2 - \bar{U}^2 = \frac{1}{10} \times 140 - 0 = 14$$

$$\sigma_y^2 = \sigma_V^2 = \frac{1}{n}\Sigma V^2 - \bar{V}^2 = \frac{1}{10} \times 758 - 36 = 39.8$$

$$\text{cov}(U,V) = \frac{1}{n}\Sigma UV - \bar{U}\bar{V} = \frac{1}{10} \times -93 - 0 = -9.3$$

$$b_{uv} = \frac{cov(u,v)}{\sigma_v^2} = \frac{-9.3}{39.8} = -0.234$$

$$b_{vu} = \frac{cov(u,v)}{\sigma_u^2} = \frac{-9.3}{14} = -0.664$$

$$\bar{y} = B + \bar{v} = 32 + 6 = 38$$

$$\bar{x} = A + \bar{u} = 32 + 0 = 32$$

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

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

$$y - 38 = -0.664x + 21.25$$

$$y = -0.664x + 59.25$$

$$y = -0.664(30) + 59.25$$

$$y = 39.32$$

Date
Page

Q. The regression equations are

$8x - 10y + 66 = 0$

$40x - 18y = 214$

The value of variance of x (σ_x^2) is 9, find

1) The ~~valu~~ mean values of x and y.

11) Correlation coefficient betn x + y.

111) Standard deviation of Y.

Soln:

Given: $\sigma_x^2 = 9$

Let, $8x - 10y + 66 = 0$ be line of regression x on y.

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

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

Let, $40x - 18y = 214$ be line of regression Y on x.

$$Y = \frac{40}{18}x - \frac{214}{18}$$

$$\boxed{b_{yx} = 40/18}$$

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

$$= \sqrt{\frac{10}{8} \times \frac{40}{18}}$$

$$= \sqrt{\frac{400}{8 \times 18}} = 1.667$$

But, $-1 \leq r \leq 1$

$\therefore$ Let, $8x - 10y + 66 = 0$ be line of regression Y on X

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

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

Let, $40x - 18y = 214$ be line of regression X on Y.

$$X = \frac{18Y}{40} + \frac{214}{40}$$

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

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

$$= \sqrt{\frac{18}{40} \times \frac{8}{10}}$$

$$= \pm 0.6$$

Since the lines passes through $\bar{x}$ & $\bar{y}$

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

$40\bar{x} - 18\bar{y} = 214$

$\bar{x} = 13, \ \bar{y} = 17$

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

$\frac{18}{40} = 0.6 \frac{3}{\sigma_y}$

$\sigma_y = 1.8 \times \frac{40}{18}$

$\boxed{\sigma_y = 4}$

Q. Given $x-4y=5$ & $x-16y=-64$ are regression lines.

1) Find regression coeff. x and Y & Y on x

ii) correlation coefficient.

iii) $\bar{x}, \bar{y}$

Sol^n: i) $x-4y=5$

$x-16y=-64$

let $x-4y=5$ be line of regression x on y

$x = 4y+5$

$\boxed{b_{xy} = 4}$

let $x-16y=-64$ be line of regression Y on x

$y = \frac{x}{16} + \frac{64}{16}$

$\boxed{b_{yx} = 1/16}$

ii) $r = \sqrt{byx \cdot bxy}$

$= \sqrt{\frac{1}{16} \times 4}$

$= \sqrt{1/4}$

$= 1/2$

$= 0.5$

(ii) since, lines passes through $\bar{x}$ + $\bar{y}$

$\bar{x} - 4\bar{y} = 5$

$\bar{x} - 16\bar{y} = -64$

$$\boxed{\begin{aligned}\bar{x} &= 28 \\ \bar{y} &= 5.75\end{aligned}}$$

Q. Find lines of regression if $n=7$, $\Sigma x = 172$, $\Sigma Y = 174$, $\Sigma X^2 = 4910$, $\Sigma Y^2 = 4910$, $\Sigma XY = 4904$

Sol^n:

Given:

$n = 7$	$\Sigma x^2 = 4910$
$\Sigma x = 172$	$\Sigma Y^2 = 4910$
$\Sigma Y = 174$	$\Sigma XY = 4904$

$\bar{X} = \frac{\Sigma X}{n} = \frac{172}{7} = 24.57$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{174}{7} = 24.85$

$$\sigma_x^2 = \frac{1}{n}\Sigma x^2 - \bar{x}^2 = \frac{1}{7} \times 4910 - (24.57)^2 = 97.74$$

$$\sigma_y^2 = \frac{1}{n}\Sigma y^2 - \bar{y}^2 = \frac{1}{7} \times 4910 - (24.85)^2 = 83.91$$

$$\operatorname{cov}(x,y) = \frac{1}{n}\Sigma xy - \bar{x}\bar{y} = \frac{1}{7} \times 4904 - (24.57)(24.85)$$

$$= 90;$$

$$b_{xy} = \frac{\operatorname{cov}(x,y)}{\sigma_y^2} = \frac{90}{83.91} = 1.072$$

$$b_{yx} = \frac{\operatorname{cov}(x,y)}{\sigma_x^2} = \frac{90}{97.74} = 0.921$$

Line of Regression y on x

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

$$y - 24.85 = 0.921(x - 24.57)$$

$$y - 24.85 = 0.921x - 22.63$$

$$\boxed{y = 0.921x + 2.22}$$

Line of Regression x on y

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

$$x - 24.57 = 1.072(y - 24.85)$$

$$x - 24.57 = 1.072y - 26.64$$

$$\boxed{x = 1.072y - 2.07}$$

Q. Find line of Regression

$n = 10$

$\Sigma(X-32) = 0$

$\Sigma(Y-36) = -10$

$\Sigma(X-32)^2 = 140$

$\Sigma(Y-36)^2 = 1158$

$\Sigma(X-32)(Y-36) = -93$

$\bar{U} = \dfrac{\Sigma U}{n} = \dfrac{0}{10} = 0$

$\bar{V} = \dfrac{\Sigma V}{n} = \dfrac{-10}{10} = -1$

$\sigma_x^2 = \sigma_U^2 = \dfrac{1}{n}\Sigma U^2 - \bar{U}^2 = \dfrac{140}{10} - 0 = 14$

$\sigma_y^2 = \sigma_V^2 = \dfrac{1}{n}\Sigma V^2 - \bar{V}^2 = \dfrac{1158}{10} - 1 = \dfrac{1148}{10} = 114.8$

$\mathrm{COV}(X,Y) = \dfrac{1}{n}\Sigma XY - \bar{X}\bar{Y}$

$\mathrm{COV}(U,V) = \dfrac{1}{n}\Sigma UV - \bar{U}\bar{V} = \dfrac{1}{10} \times -93 + 1 = \dfrac{-93}{10} + 0$

$= -9.3$

~~bxy =~~

$b_{UV} = \dfrac{\mathrm{COV}(U,V)}{\sigma_V^2} = \dfrac{-9.3}{114.8} = -0.081$

$b_{yx} = b_{VU} = \dfrac{\mathrm{COV}(U,V)}{\sigma_U^2} = \dfrac{-9.3}{14} =$ ~~50.592~~ -0.664

$\bar{X} = A + \bar{U} = 32 + 0 = 32$

$\bar{Y} = B + \bar{V} = 36 + (-1) = 35$

Line of Regression Y on X

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

$Y-35 = -0.664(X-32)$

$Y = -0.664X + 21.25 + 35$

$$\boxed{Y = -0.664X + 56.25}$$

Line of Regression X on Y

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

$X-32 = -0.081(Y-35)$

$X = -0.081Y + 2.835 + 32$

$$\boxed{X = -0.081Y + 34.84}$$