LINE OF REGRESSION :

The line of regression is the line of best fit and is obtained by the principle of least square.

The regression line of y on x is

$$y - \bar{y} = b_{yx} [x - \bar{x}]$$

The regression co.eff of y on x

$$b_{yx} = \frac{\Sigma [x - \bar{x}] [y - \bar{y}]}{\Sigma [x - \bar{x}]^2}$$

The regression line of x on y is

$$x - \bar{x} = b_{xy} [y - \bar{y}]$$

The regression coeff of x on y

$$b_{xy} = \frac{\Sigma [x - \bar{x}] [y - \bar{y}]}{\Sigma [y - \bar{y}]^2}$$

The correlation coeff ρ in terms of the regression

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

NOTE :

The correclation coeff value always lies b/t -1 & +1.

The regression coeff can also

solved from

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

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

The marks in statistics and economics of 10 students are given below find the two regression lines, corelation coffecient and the most likely marks in statistics when mark in economics is 30.

Marks in Statistics	25	28	35	32	31	36	29	38	34	32
Marks in Economics	43	46	49	41	36	32	31	30	33	39

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

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

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

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

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	~~1225~~	-35
28	46	-4	8	16	64	~~1024~~	-32
35	49	3	11	9	121	~~1089~~	33
32	41	0	3	0	9	~~0~~	0
31	36	-1	-2	1	4	~~4~~	2
36	32	4	-6	16	36	~~576~~	-24
29	31	-3	-7	9	49	~~441~~	21
38	30	6	-8	36	64	~~2304~~	-48
34	33	2	-5	4	25	~~100~~	-10
32	39	0	1	0	1	~~0~~	0

$\bar{x} = 32$ $\quad \bar{y} = 38$

$$b_{xy} = \frac{\cancel{6763}}{\cancel{140}} = \cancel{48.3}$$

$$b_{xy} = \frac{-93}{398} = -0.23$$

$$b_{yx} = \frac{-93}{140} = -0.66.$$

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

$$x - 32 = (-0.23)(y - 38)$$

$$x = -0.23y + 8.74 + 32$$

$$\boxed{x = 40.74 - 0.23y}$$

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

$$y - 38 = (-0.66)(x - 32)$$

$$y = -0.66x + 21.12 + 38$$

$$\boxed{y = 59.12 - 0.66x}$$

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

$$\rho = \pm \sqrt{-0.23 \times -0.66}$$

$$\rho = \pm \sqrt{0.15}$$

$$\boxed{\rho = \pm 0.39}$$

Since b_{xy} & b_{yx} are negative in sign $\rho = -0.39$.

iii) The marks in statistics when mark in eco is 30 given by substituting $y = 30$ in regression line x on y

$$x = (-0.23)(y - 38) + 32$$

$$= -0.28y + 874 + 32$$

$$= 40.74 - 0.237y$$

$$= 40.74 - 6.90$$

$$\boxed{x = 33.84}$$

Height of dad x	150	152	155	157	160	161	164	166
Height of son y	154	156	158	159	160	162	161	164

$\bar{x} = 158.13$ $\bar{y} = 159.25$.

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
150	154	-8.13	-5.25	66.10	27.56	42.68
152	156	-6.13	-3.25	37.58	10.56	19.92
155	158	-3.13	-1.25	9.80	1.56	3.91
157	159	-1.13	-0.25	1.28	0.06	0.28
160	160	1.87	0.75	3.50	0.56	1.40
161	162	2.87	2.75	8.24	7.56	7.89
164	161	5.87	1.75	34.46	3.06	10.27
166	164	7.87	4.75	61.94	22.56	37.38.
				222.90	73.48	123.78

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

$$= \frac{123.78}{222.90}$$

$$= 0.56$$

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

$$= \frac{123.78}{73.48}$$

$$= 1.68$$

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

$$x - 158.13 = (0.56)(y - 159.25)$$

$$x - 158.13 = 0.56y - 89.18$$

$$x = 0.56y - 89.18 + 158.13$$

$$\boxed{x = 0.56y + 68.95}$$

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

$$y - 159.25 = 1.68 (x - 158.13)$$

$$y - 159.25 = 1.68x - 265.66$$

$$\boxed{y = 1.68x - 106.41}$$

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

$$= \pm \sqrt{0.56 \times 1.68}$$

$$= \pm \sqrt{0.94}$$

$$\boxed{\rho = \pm 0.970}$$

Since b_{xy} and b_{yx} are both positive in sign $\rho = 0.970$

17 Obtain the two regression equation and Estimate the yield of crops when the rainfall is 22 cm and the rainfall when yeild is 500 kg.

y (yeild) x (rainfall)

mean	508.4	26.7
SD	36.8	4.6

Correlation cof = 0.52.

Given: $\bar{x} = 26.7$ $\bar{y} = 50.8.4$

$\sigma_x = 4.6$ $\sigma_y = 36.8$

$$x - \bar{x} = b_{xy} [y - \bar{y}]$$

$$b_{xy} = \rho \cdot \frac{\sigma_x}{\sigma_y} = 0.52 \times \frac{4.6}{36.8}$$

$$\boxed{b_{xy} = 0.07.}$$

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

$$= 0.52 \times \frac{36.8}{4.6}$$

$$\boxed{b_{yx} = 4.16.}$$

$$x - 26.7 = 0.07 (y - 508.4)$$

$$x = 0.07y - 35.59 + 26.7$$

$$\boxed{x = 0.07y - 8.89}$$

$$y - \bar{y} = b_{yx} [x - \bar{x}]$$

$$y - 508.4 = 4.16 [x - 26.7]$$

$$y = 4.16x - 111.07 + 508.4$$

$$\boxed{y = 4.16x + 397.33}$$

When $x = 22$

$$y = 4.16 \times 22 + 397.33$$

$$\boxed{y = 488.84}$$

When $y = 600kg$

$$x = 0.07 \times 600 - 8.89$$

$$\boxed{x = 33.11}$$

If the line of regression concerning two variables x & y is given by $y = 32 - x$ and $x = 13 - 0.25y$. Obtain the value of the mean and correlation coeffient.

Since the regression lines intersect at $\bar{x}, \bar{y}$, by solving these lines we get the means of x and y.

$$y = 32 - 13 + 0.25y$$

$$y = 19 + 0.25y$$

$$y - 0.25y = 19$$

$$0.75y = 19$$

$$y = \frac{19}{0.75} = 25.33$$

$$\boxed{y = 25.33}$$

$$x = 13 - 0.25(32 - x)$$

$$= 13 + 0.25x - 8$$

$$x = 5 + 0.25x$$

$$x - 0.25x = 5$$

$$0.75x = 5$$

$$\boxed{x = 6.67}$$

$\bar{x} = 6.67$ $\quad \bar{y} = 25.33$

$y = 32 - x \quad y - \bar{y} = b_{yx}(x - \bar{x})$

$x = 13 - 0.25y \quad x - \bar{x} = b_{xy}(y - \bar{y})$

From the regression line x on y

and y on x

$$\boxed{b_{yx} = -1}$$

$$\boxed{b_{xy} = -0.25}$$

$$\therefore \rho = \pm\sqrt{b_{xy} \cdot b_{yx}}$$

$$= \pm\sqrt{-1 \times -0.25}$$

$$= \pm\sqrt{0.25}$$

$$= \pm 0.50$$

$$\boxed{\rho = -0.50}$$

Give that the varience of x is 9

The regression equation are $8x - 10y + 66 = 0$ and $40x - 18y = 214$. Find

i) Avg value of x and y

ii) Correlation coff

iii) Std deviation of y.

Given: $8x - 10y = -66$ $\quad Var(x) = 9$

$40x - 18y = 214$ $\quad \sigma_x = 9$

$\times 5,$ $\cancel{40x} - 50y = -330$

$\underset{(-)}{\cancel{40x}} \underset{(+)}{-} 18y = \underset{-}{}214$

$-32y = -544$

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

$\boxed{y = 17}$

$8x - 170 = -66$

$8x = -66 + 170$

$8x = 104$

$x = \frac{104}{8}$

$\boxed{x = 13}$

$\therefore \boxed{\bar{x} = 13}$ $\boxed{\bar{y} = 17}$

$10y = 66 + 8x$

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

$y = 0.80x + 6.6$

$\therefore \boxed{b_{yx} = 0.8}$

$$40x - 18y = 214$$

$$40x = 18y + 214$$

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

$$x = 0.45y + 5.35$$

$$\boxed{b_{xy} = 0.45}$$

$$\rho = \pm\sqrt{0.80 \times 0.45}$$

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

$$= \pm 0.60$$

$$\therefore \boxed{\rho = 0.60}$$

$$\sigma_x = \sqrt{9} = 3$$

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

$$0.8 = 0.6 \frac{\sigma_y}{3}$$

$$0.6\sigma_y = 2.40$$

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

i] Average value $\bar{x} = 13$, $\bar{y} = 17$

ii] Correlation coefficient $\rho = 0.6$

iii] Standard Deviation of $\sigma_y = 4$

The following results were obtained from records of age (x) and systolic blood pressure (y) of a group of 10 women:

	x	y
Mean	58	142
Varience	130	165

$$\Sigma (x-\bar{x})(y-\bar{y}) = 1220$$

$$\sigma_x = \sqrt{130} = 11.40$$

$$\sigma_y = \sqrt{165} = 12.85$$

Average $x = 53$ $y = 142$

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

$$x = \frac{\Sigma (x-\bar{x})^2}{n} = 130$$

$n = 10$

$$\Sigma(x-\bar{x})^2 = 130 \times 10$$

$$\Sigma(x-\bar{x})^2 = 1300$$

$$b_{yx} = \frac{1220}{1300}$$

$$\boxed{b_{yx} = 0.94}$$

$$y-\bar{y} = 0.94(x-\bar{x})$$

$$y-142 = 0.94(x-53)$$

$$y-142 = 0.94x - 49.82$$

$$y = 0.94x + 92.18$$

$$x = 45$$

$$y = 42.30 + 92.18$$

$$\boxed{y = 134.48}$$

The estimated systolic blood pressure of woman whose age is 45 is 134.48.

Marks obtained by 12 students in the college test (x) and the university text (y) are the follows:

x	41	45	50	68	47	77	90	100	80	100	40	43
y	60	63	60	48	85	56	53	91	74	98	65	43

What is your estimate of the marks a student could have obtained in the university test if he obtained 60 in the college test but was ill at the time of the university test?

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
41	60	-24.08	-6.33	579.85	40.07	152.43
45	63	-20.08	-3.33	403.21	11.09	66.87
50	60	-15.08	-6.33	227.41	40.07	95.46
68	48	2.92	-18.33	8.53	335.97	-53.52
47	85	-18.08	18.67	326.89	348.57	-337.53
77	56	11.92	-10.33	142.09	106.71	-123.13
90	53	24.92	-13.33	621.01	177.69	-332.18
100	91	34.92	24.67	1219.41	608.61	861.48
80	74	14.92	7.67	222.61	58.83	114.41
100	98	34.92	31.67	1219.41	1002.99	1105.92
40	65	-25.08	-1.33	629.01	1.77	33.36
43	43	-22.08	-23.33	487.53	544.29	515.13

$$\bar{x} = \frac{781}{12} \qquad \bar{y} = \frac{79.6}{12}$$

$$\boxed{\bar{x} = 65.08} \qquad \boxed{\bar{y} = 66.33}$$

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

$$b_{yx} = \frac{2098.81}{6086.12}$$

$$\boxed{b_{yx} = 0.34}$$

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

$$= \frac{2098.81}{3176.68}$$

$$\boxed{b_{xy} = 0.66}$$

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

$$x - 65.08 = 0.66(y - 66.33)$$

$$x - 65.08 = 0.66y - 43.78$$

$$\boxed{x = 0.66y + 21.30}$$

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

$$y - 66.33 = 0.34(x - 65.08)$$

$$y - 66.33 = 0.34x - \cancel{4316.76}\ 22.13$$

$$y = 0.34x - 22.13 + 66.33$$

$$\boxed{y = 0.34x + 42.20}$$

when $x = 60$

$$y = 0.34 \times 60 + 44.20$$

$$y = 20.40 + 44.20$$

$$\boxed{y = 64.60 \text{ (approx) or } 65.6}$$

8.7.25

MARGINAL AND CONDITIONAL FREQUENCY DISTRIBUTION :

1. Find the marginal and conditional distribution of x and y from the following table.

x | y	111 - 120	121 - 130	131 - 140	141 - 150	151 - 160
58 - 61	1				
62 - 65	1	1	1		
66 - 69			3	2	3
70 - 73		1	2	1	

x | y					
58 - 61	1				
62 - 65	1	1	1		
66 - 69			3	2	3
70 - 73		1	2	1	
	2	2	6	3	3

Total frequency = 16

x	Marginal distribution f(x)
58 - 61	1
62 - 65	3
66 - 69	8
70 - 73	4

y	Marginal Distribution f(y)
111 - 120	2
121 - 130	2
131 - 140	6
141 - 150	3
151 - 160	3

CONDITIONAL DISTRIBUTION :

14 univariate frequency distribution which is derived from the bivariate frequency distribution for a specified value or class interval of the other variable.

x	Conditional distribution
58 - 61	0
62 - 65	1
66 - 69	0
70 - 73	1

The following data represents the points scored in a match by 2 players x & y at the end of 20 games.

(10, 12), (7, 11), (7, 9), (15, 19), (17, 21), (12, 8), (16, 10), (14, 14) (22, 18), (16, 7), (15, 16), (22, 20), (19, 15) (7, 18), (11, 11), (12, 18), (10, 10), (5, 13), (11, 7), (10, 10) Taking class intervals (5 to 9, 10 to 14, 15 to 19... for both x & y. construct a bivariate frequency table & also find the marginal distribution x & y, conditional frequency distribution for y given x > 15.

x/y	5 - 9	10 - 14	15 - 19	20 - 24
5 - 9	0	2	1	0
10 - 14	1	5	0	0
15 - 19	1	3	3	1
20 - 24	0	2	1	0

Total frequency = 20

x	frequency
5 - 9	3
10 - 14	6
15 - 19	8
20 - 24	

y	frequency
5 - 9	2
10 - 14	12
15 - 19	5
20 - 24	1

Conditional

x/y	5-9	10-14	15-19	20-24
15-19	1	3	3	1
20-24	0	2	1	0
Total	1	5	4	1

y interval	frequency	probability
5-9	1	1/11
10-14	5	5/11
15-19	4	4/11
20-24	1	1/11