18/8/25

Unit - 2 Statistical Techniques

Curve fitting - Suppose we have data set

$x \rightarrow x_1, x_2, x_3, \ldots, x_n$
$y \rightarrow y_1, y_2, y_3, \ldots, y_n$

from the above given data set we have to fit these data to the given eqn of curve

for this we have to find a best fitted curve by using Method of least sq. This is called curve fitting.

Case-I Fitting of a straight line $y = a + bx$ to the given data set.

$y = a + bx$ ——— (I)

where a & b are called unknowns.
The Normal eqns of (I) are:

$\Sigma y = a\Sigma 1 + b\Sigma x$ ——— (II) $\times \Sigma$
$\Sigma y = an + b\Sigma x$
$\Sigma xy = a\Sigma x + b\Sigma x^2$ ——— (III) $\times x$

where $n \rightarrow$ Total no. of data.

On solving (II) & (III), we get the value of a & b. These values of a & b put in eqn of curve (I) we will get req. best fitted curve.

Q. Fit a st. line to the following data set by using a method of least squares.

$x \rightarrow$	1	2	3	4	5
$y \rightarrow$	14	27	40	55	68

Let the st. line is : $y = a + bx$ ——— (I)

Normal eqn of (I) are :

$\Sigma y = an + b\Sigma x$ ——— (II)
$\Sigma xy = a\Sigma x + b\Sigma x^2$ ——— (III)

x	y	xy	x^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
15	204	748	55

from (II) : $204 = 5a + 15b$ $\times 3$
$748 = 15a + 55b$

On solving. $612 = 15a + 45b$
$136 = 10b \Rightarrow b = 13.6$

$\boxed{a = 0 \quad b = 13.6}$

$204 = 5a + 15 \times 13.6$
$= 5a + 204$
$5a = 0$, $a = 0$

Q. Fit the given data set to the second degree parabola by using method of least square.

$y = a + bx + cx^2$ —— (1)

$y = ax^2 + bx + c$ $\quad$ ~~$\Sigma y = ax$~~

$\Sigma y = a\Sigma x^2 + b\Sigma x + cn$

$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x$

$\Sigma x^2 y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2$

x	y	x^2	x^3	xy	x^4	x^2y
1	1.8	1	1	1.8	1	1.8
2	5.1	4	8	10.2	16	20.4
3	8.9	9	27	26.7	81	80.1
4	14.1	16	64	56.4	256	225.6
5	19.8	25	125	99.0	625	495
15	49.7	55	225	194.1	979	822.9

$49.7 = 5a + 15b + 5c$ $\quad$ -0.291 $\quad$ $a = 0.4428$

$194.1 = 225a + 55b + 15c$ $\quad$ 1.64 $\quad$ $b = 1.843$

$822.9 = 979a + 225b + 55c$ $\quad$ 0.46 $\quad$ $c = -0.41$

$y = a + bx + cx^2$

$\Sigma y = an + b\Sigma x + c\Sigma x^2$

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$

$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$

Case-II Eqns look like polynomeal.

eg- $y = ax + \frac{b}{x}$ —— (1)

Normal eqn of (1):

$\Sigma xy = a\Sigma x^2 + bn$ —— (II)

$\Sigma \frac{y}{x} = an + b\Sigma \frac{1}{x^2}$ —— (III)

(2) $y = a\sqrt{x} + \frac{b}{\sqrt{x}}$ etc.

Q.1. Use the method of least squares to fit the curve $y = \frac{C_0}{x} + C_1\sqrt{x}$ to the following.

x	0.1	0.2	0.4	0.5	1	2
y	21	11	67	06	05	06

Normal eqn -

$\Sigma \frac{y}{x} = C_0 \Sigma \frac{1}{x^2} + C_1 \Sigma \frac{1}{\sqrt{x}}$ —— (I)

$\Sigma y\sqrt{x} = C_0 \Sigma \frac{1}{\sqrt{x}} + C_1 \Sigma x$ —— (II)

x	y	y/x	$1/x$	$1/x^2$	$1/\sqrt{x}$	$\sqrt{x}$	$y\sqrt{x}$
0.1	21	210	10	100	3.162	0.316	6.636
0.2	11	55	5	25	2.236	0.447	4.917
0.4	67	175	2.5	6.25	1.581	0.632	4.424
0.5	06	12	2	4	1.414	0.707	4.242
1	05	5	1	1	1	1	5
2	06	3	0.5	0.25	0.707	1.414	6
[illegible]	[illegible]	[illegible]	[illegible]	136.5	10.1	4.51	33.7

In (I): $302.5 = C_0(136.5) + C_1(10.1)$ — (III)

In (II): $33.703 = C_0(10.1) + C_1(4.2)$ — (IV)

(III) − (IV): × (2.4047)

$C_0(136.5) + C_1(10.1) = 302.5$

$C_0(24.288) + C_1(10.1) = 81.048$

$(112.212)\,C_0 = 221.452$

$C_0 = 1.974$

$C_1 = 3.278$

$\therefore$ Best fit curve $y = \frac{1.974}{x} + 3.278\sqrt{x}$

Q. Fit the given data set to the curve $xy = ax + b$

$x \to$	1	3	5	7	9	10
$y \to$	36	29	28	26	24	15

Normal eqn: $\Sigma x^2y = a\Sigma x^2 + b\Sigma y$ — (I)

$\Sigma xy = a\Sigma x + bn$ — (II)

x	y	xy	x^2	x^2y
1	36	36	1	36
3	29	87	9	261
5	28	140	25	700
7	26	182	49	1274
9	24	216	81	1944
10	15	150	100	1500
35	158	811	265	5715

In (I):

$5715 = (265)a + (158)b$ — (III)

In (II):

$811 = (35)a + 6b$ — (IV)

Case-III When the given curve is not a polynomial.

1) $y = ae^{bx}$, $y = ab^x$, $PV^Y = k$ (const.)

Take $y = ae^{bx}$ — (I)

Taking log both sides on base 10.

$\log_{10} y = \log_{10} ae^{bx}$

$= \log_{10} a + \log_{10} e^{bx}$

$= \log_{10} a + bx \log_{10} e$

Let $\log_{10} y = Y$

$\log_{10} a = A$

" $b\log_{10} e = B$

" $x = X$

Know, $\therefore Y = A + Bx$ (Eqn of st. line)

Normal eqn − $\Sigma Y = An + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

On solving, we get A, B.

$a = 10^A$

$b = \frac{B}{\log_{10} e}$

Case - When the format of the curve is not a polynomial.

Q.1. Fit the curve is in form $y = ae^{bx}$ to the following data set -

$x \rightarrow$	1	5	7	9	12
$y \rightarrow$	10	15	12	15	21

$y = ae^{bx}$ —— (I)

Taking log both sides on base 10.

$\log_{10} y = \log_{10} ae^{bx}$

$= \log_{10} a + \log_{10} e^{bx}$

$= \log_{10} a + bx \log_{10} e$

Let $\log_{10} y = Y$, $\log_{10} a = A$, $b \log_{10} e = B$, $x = X$

$Y = A + BX$ —— (II)

Normal eqn (II) are:

$\Sigma Y = An + B\Sigma x$ —— (III)

$\Sigma XY = A\Sigma X + B\Sigma x^2$ —— (IV)

$X=x$	y	$Y = \log_{10} y$	x^2	XY
1	10	1	1	1
5	15	1.1761	25	5.8805
7	12	1.0792	49	7.5544
9	15	1.1761	81	10.5849
12	21	1.3222	144	15.8664
34	73	5.7536	300	[illegible]

from (III) & (IV):

$5.7536 = 5A + 34B$ $\times 34$

$40.8862 = 34A + 300B$ $\times 5$

$195.6224 = 170A + 1156B$

$204.431 = 170A + 1500B$

$8.8086 = 344B$

$\boxed{B = 0.025606}$

$5.7536 = 5A + 0.8704$

$4.8832 = 5A$

$\boxed{A = 0.97664}$

$A = \log_{10} a$

$a = 10^A \Rightarrow a = 10^{0.9766}$

$a = 9.4755$

$B = b \log_{10} e \Rightarrow \left\{ b = \dfrac{B}{\log_{10} e} = B \log_e 10 \right\}$

$b = 0.0256 \times \ln 10$

$= 0.0256 \times 2.3026$

$= 0.0589$

$y = 9.4755\, e^{0.0589x}$

Q. Fit the curve $pv^{\gamma} = k$ to the data:

p (kg/cm²)	0.5	1	1.5	2	2.5	3
V (litres)	1620	1000	750	620	520	460

10m *

$$PV^{\gamma} = K$$
$$(V^{\gamma})^{1/\gamma} = \left(\frac{K}{P}\right)^{1/\gamma}$$

$$\begin{cases} y = f(x) \\ V = f(P) \end{cases}$$

$$V = \frac{K^{1/\gamma}}{P^{1/\gamma}}$$

$$\log_{10} V = \log_{10} K^{1/\gamma} - \log_{10} P^{1/\gamma}$$

$$\log_{10} V = \frac{1}{\gamma} \log_{10} K - \frac{1}{\gamma} \log_{10} P$$

$$y = a - bx \quad \text{--- (II)}$$

$$y = \log_{10} V \quad \Rightarrow \quad a = \frac{1}{\gamma} \log_{10} K$$

$$b = \frac{1}{\gamma}$$

$$PV^{1.4225} = 18462.48$$

$$x = \log_{10} P$$

$$\boxed{y = a - bx}$$

Normal eqn:

$$\Sigma y = an - b\Sigma x \quad \text{--- (II)}$$
$$\Sigma xy = a\Sigma x - b\Sigma x^2 \quad \text{--- (III)}$$

x	y
0.5	1620
1	1000
1.5	750
2	620
2.5	520
3	460

Method of least square

$$y = a + bx \quad \text{--- (1)}$$

$$C = \text{Total error} = (y_1 - a - bx_1)^2 + (y_2 - a - bx_2)^2 + \ldots + ($$

$$E = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \quad \text{--- (II)}$$

$$\frac{\partial E}{\partial a} = 0, \quad \frac{\partial E}{\partial b} = 0$$

$$\frac{\partial E}{\partial a} = \Sigma 2(y_i - a - bx_i)'(-1) = 0$$

$$0 = -2\Sigma(y_i - a - bx_i)$$
$$\Sigma(y_i - a - bx_i) = 0$$

$$\Sigma y_i - \Sigma a - \Sigma bx_i = 0$$
$$\Sigma y_i = \Sigma a + \Sigma bx_i$$
$$\boxed{\Sigma y = an + b\Sigma x}$$

$$\frac{\partial \varepsilon}{\partial b} = \Sigma 2(y_i - a - bx_i)(-x_i)$$

$$= -2\Sigma(x_iy_i - ax_i - bx_i^2)$$

$$0 = \Sigma x_iy_i - \Sigma a x_i - \Sigma b x_i^2$$

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

Co-relation – In the bivariant distribution, if the change in one variable affects the change in the other var. the var. are said to be co-related. The co-relⁿ gives us the degree of relationship among the variables.

If the two variables deviate in the same directⁿ i.e. if the increase (or decrease) in one var. results in a corresponding ↑ or ↓ in the var., then the co-relⁿ is said to be direct or +ve corrⁿ

+ve) +1

−ve) −1

eg– Income & expenditure. (+ve / dir. correlⁿ).

If the two var. deviate in opp. direcⁿ i.e. if the ↑ or ↓ in one var. results in corresp. ↓ or ↑ in the other var. then the corr. is said to be inverse or −ve corr.

Co.relⁿ is said to be perfect if the dev. in one var. is followed by corr. equal proportional in the other var.

DATE 21/8/25

Karl Pearson's coeff. of co-relⁿ –

$$r_{xy} = r = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

where $\sigma_x \rightarrow$ S.D. of x

$\sigma_y \rightarrow$ " " y

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

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

$$r_{xy} = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2} - \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

Q.1. Find the coeff of co-relⁿ for the foll. data Karl Pearson method.

x	y	xy	x^2	y^2
1	8	8	1	64
3	12	36	9	144
5	15	75	25	225
7	17	119	49	289
8	18	144	64	324
10	20	200	100	400
34	90	582	248	1446

$$r_{xy} = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{6 \times 582 - 34 \times 90}{\sqrt{6 \times 248 - (34)^2}\sqrt{6 \times 1446 - (90)^2}}$$

$$= \frac{3492 - 3060}{\sqrt{1488 - 1156}\sqrt{8676 - 8100}}$$

$-1 \leq r \leq +1$

$$= \frac{432}{\sqrt{332}\sqrt{576}}$$

$$= \frac{432}{18.22 \times 24} = \frac{432}{437.28}$$

$$= 0.98792$$

The Karl Pearson coeff. of corelⁿ gives the degree of linear association b/w two variables.

Find the coeff. of co-relⁿ by KP method

x	y	xy	x^2	y^2
1	2	2	1	4
2	4	8	4	16
3	6	18	9	36
4	8	32	16	64
5	10	50	25	100
15	30	110	55	220

$(y = 2x)$

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{5 \times 110 - 15 \times 30}{\sqrt{5(55) - (15)^2}\sqrt{5(220) - (30)^2}}$$

$$= \frac{550 - 450}{\sqrt{275 - 225}\sqrt{1100 - 900}}$$

$$= \frac{100}{\sqrt{50}\sqrt{200}} = \frac{100}{5\times\sqrt{2}\times 10\times\sqrt{2}} = 1$$

Q. The following data regarding the height (y) & weight (x) of 100 students as-

$\sum x = 15000$, $\sum x^2 = 2272500$
$\sum y = 6800$, $\sum y^2 = 4630255$, $\sum xy = 1022250$
Find the coeff of co-relⁿ.

$$r_{xy} = \frac{100 \times 1022250 - 15000 \times 6800}{\sqrt{100(2272500) - (15000)^2}\sqrt{100(4363025\overline{5}) - (6800)^2}}$$

$$= \frac{102225\cancel{0}0 - 102000\cancel{0}00}{\sqrt{227250000 - 225000000}\sqrt{436302550\overline{0} - 46240000}}$$

$$= \frac{2250}{\sqrt{2250000}\sqrt{}}$$

$$= \frac{2250}{1500 \times \sqrt{62500}} = \frac{225\cancel{0}}{1500 \times 25\cancel{0}}$$

$$= \frac{225}{37500} = 0.6$$

Q-3.

x	y	xy	x^2	y^2	$y=x^2$
1	1	1	1	1	
2	4	8	4	16	
3	9	27	9	81	
4	16	64	16	256	
5	25	125	25	625	
15	55	225	55	979	

$$r = \frac{5 \times 225 - 15 \times 55}{\sqrt{5 \times 55 - (15)^2}\sqrt{5 \times 979 - (55)^2}}$$

$$= \frac{1125 - 825}{\sqrt{275 - 225}\sqrt{4895 - 3025}}$$

$$= \frac{300}{\sqrt{50}\sqrt{1870}}$$

$$= \frac{300}{7.07 \times 43.24} = \frac{300}{305.70}$$

$$= 0.9813$$

Spearman's rank co-relation coeff.-

Spearman's $\rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$

where, $D = R_1 - R_2$

$n \rightarrow$ Total no. of data.

Note- This for. is app. for no tied relⁿ.

This formula gives better result for monotonic relⁿ b/w the var.

Q-1. The marks secured by students in the selection test (x) and the proficiency test (y) are given. Find S.R.C. of co-relⁿ

R_1	x	y	R_2	$D = R_1 - R_2$	D^2
9	10			0	0
5	15	30	9	2	4
8	19	42	3	6	36
3	17	45	2	2	4
7	13	46	1	-1	1
4	16	33	8	-3	9
1	24	34	7	-3	9
6	14	40	4	0	0
2	22	35	6	-3	9
		395			72

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

$$= 1 - \frac{6(72)}{9(81-1)}$$

$$= 1 - \frac{\cancel{432}\ 432}{720}$$

$$= 1 - 0.6 = 0.4$$

Q-2 10 competitors in a beauty contest were ranked by three judges. Use the method of rank co-relⁿ to determine which pair of judges has the nearest approach to comm. taste in bea

R_1	First Judge	R_2	S.J	R_3 T.J.	R_{12}	D^2	R_{23}	D^2
10	1	8	3	6	-2	4	-3	9
5	6	6	5	4	1	1	1	1
6	5	3	8	9	-3	9	-1	1
1	10	7	4	8	6	36	-4	16
8	3	4	7	1	-4	16	6	36
9	2	1	10	2	-8	64	8	64
7	4	9	2	3	2	4	-1	1
2	9	10	1	10	8	64	-9	81
4	7	5	6	5	1	1	1	1
3	8	2	9	7	-1	1	2	4
						200		214

$$\rho_{12} = 1 - \frac{6\,\Sigma D^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 200}{10(100-1)} = 1 - \frac{120\cancel{0}}{99\cancel{0}}$$

$$= 1 - 1.21 = -0.212$$

R_{31}	
-5	25
2	4
-4	16
2	4
2	4
0	0
1	1
-1	1
2	4
1	1
	60

$$\rho_{23} = 1 - \frac{6 \times 214}{990}$$

$$= 1 - \frac{1284}{990}$$

$$= 1 - 1.29 = -0.297$$

$$\rho_{31} = 1 - \frac{6 \times 60}{990} = 1 - \frac{360}{990} = 1 - 0.363$$

$$= 0.63\cancel{7}$$

Spearman's coeff of corel^n.
(when there are tied ranks).

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

where,

$m_1 \to$ repeated no. of times data either in x or y.

$m_2 \to$ repeated no. of times second data either in x or y.

Q-2 Find the coeff. of rank co-rel^n from foll. data.

x	y	R_1	R_2	D	D^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
					72

$m_1 = 2$, $m_2 = 3$, $m_3 = 2$

$$\rho = 1 - \frac{6\left[72 + \frac{2(4-1)}{12} + \frac{3(9-1)}{12} + \frac{2(4-}{12}\right.}{10(100-1)}$$

$$= 1 - \frac{6\,[72 + 0.5 + 2 + 0.5]}{990} = 1 - \frac{450}{990}$$

$$= 1 - 0.454 = 0.545$$

Note - If we apply change of origin & scale in the data set, then Karl Pearson's coeff. of co-rel^n does'nt change.

$$r_{xy} = r_{uv} = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

where $u = \frac{x-a}{h}$, $v = \frac{y-b}{h}$

Q- 10 students got the foll. % marks in the eco & stats. Calculate the coef. of co-rel^n by shifting of or. & ch. of scale (Pearson).

0.784.

Roll.	Eco(x)	Stat(y)	$u = x-65$	$v = y-66$	uv	u^2	v^2
1	78	84	13	18	234	169	324
2	36	51	-29	-15	435	841	225
3	98	91	33	25	825	1089	625
4	25	60	-40	-6	240	1600	36
5	75	68	10	2	20	100	4
6	82	62	17	-4	68	289	16
7	90	86	25	20	500	625	400
8	62	58	-3	-8	-24	9	64
9	65	53	0	-13	0	0	169
10	39	47	-26	-19	494	676	361
			0	0	2704	5398	229

$\bar{x} = 65$

$\bar{y} = 66$

$$r_{uv} = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

$$= \frac{10(2704) - 0}{\sqrt{10(5398) - (0)}\sqrt{10(2294) - 0}} = \frac{27040}{}$$

$$= \frac{27040}{\sqrt{53980}\sqrt{22940}}$$

$$= \frac{27040}{232.3 \times 149.1} = \frac{27040}{34,635}$$

$$= 0.78$$

$r_{uv} = 0.789$

Regression Analysis In this, we obtain the relationship b/w the dependent var. (o/p) and the one or more independent var. (i/p).

The basic differences b/w the curve fitting & the regression analysis is that we can take any var. independent or dependent in the regression var.

$x \longrightarrow$ Investment $\longrightarrow$ data(x)

$y \longrightarrow$ Profit $\longrightarrow$ y (output)

Regression lines - Let we have data set.

$x \rightarrow x_1, x_2, x_3, \ldots, x_n$

$y \rightarrow y_1, y_2, y_3, \ldots, y_n$

Case-I. (Regression line y on x) $x \rightarrow$ In, $y \rightarrow$ dep.

Here we take x as independent var. -

$$y = a + bx \quad \text{--- (1)}$$

Apply method of least sq. (Norm. eq^n)

$y = a + bx$

$\sum y = a\cancel{m}n + b\sum x$ —— (II)

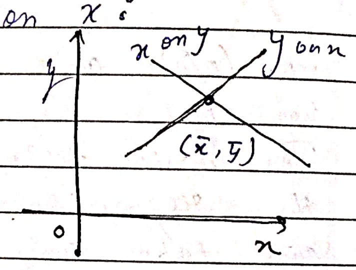
$\sum xy = a\sum x + b\sum x^2$ —— (III)

On solving (II) & (III): we get a & b,
Hence we get req. reg. line.

$\boxed{y = a + bx}$

OR

Regression line y on x:



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

where

b_{yx} = regression coeff. y on x.

where b_{yx} is given by-

$$b_{yx} = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\ \cancel{\sqrt{n\sum y^2 - (\sum y)^2}}}$$

Case-II (Regression line x on y).

Let regression line x on y.

$x = c + dy$ —— (I)

Apply method of least sq.-

Normal eqn-

$\sum x = cn + d\sum y$ —— (II)

$\sum xy = c\sum y + d\sum y^2$ —— (III)

On solving, we get c & d,
& put in (I), we get-

required regression line x on y.

OR

Regression line x on y:

$x - \bar{x} = \underbrace{b_{xy}}_{\text{slope}}(y - \bar{y})$

Note - $b_{yx} = \dfrac{r\sigma_y}{\sigma_x}$

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

$b_{yx} \rightarrow$ reg. coeff. x on y

$r \rightarrow$ (pearson) coef of co-rel.

$\sigma_y \rightarrow$ S.D of y

$\sigma_x \rightarrow$ S.D of x

$$b_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum y^2 - (\sum y)^2}}$$

(I.P.)

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

Here b_{xy}, b_{yx}, r have same sign.

1. Obtain the both regression lines for the foll. data set

x	y	xy	x^2	y^2
1.	3	3	1	9
2.	7	14	4	49
3.	10	30	9	100
4.	12	48	16	144
5	14	70	25	196
6.	17	102	36	289
7	20	140	49	400
8	24	192	64	576
36	107	599	204	1763

$y = a + bx$ — (I)

Normal eqn are –

$\Sigma y = an + b\Sigma x$ — (II)

$\Sigma xy = a\Sigma x + b\Sigma x^2$ — (III)

$107 = 8a + 36b \quad \times 9$

$599 = 36a + 204b \quad \times 2$

$a = 0.785$

$b = 2.797$

$963 = 72a + 324b$

$1198 = 72a + 408b$

$235 = 84b \Rightarrow b = 2.797$

$a = 0.785$

Regr. line y on x is

$y = 0.785 + 2.797x$

Now let reg. line x ony be

$x = c + dy$

Normal eqn are –

$\Sigma x = cn + d\Sigma y$ — (IV)

$\Sigma xy = c\Sigma y + d\Sigma y^2$ — (V)

$36 = 8c + 107d$

$599 = 107c + 1763d$

$c = -0.235$

$d = 0.354$

$c = -0.235$

$d = 0.354$

$x =$

Properties of coeff of regression.

1) The co-reln coeff is the geometric mean of the two regression coeff b_{xy} & b_{yx}

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

II) If one of the reg. coef is greater than unity, the other reg. coeff. must be less than unity

III) The arithmetic mean of the two reg. coeff. is greater than the co-reln coeff

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

IV) The reg. coeff. are independent of the change of origin or change of scale

V) The co-reln coeff and the two reg. coeff have same sign.

Q – If the two coeff. of reg. are 0.4 & 1.6 then find the coeff. of co-reln.

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

$= \sqrt{0.4 \times 1.6}$ $\quad \because b_{xy}, b_{yx}$ & r must be same sign

$r = \sqrt{0.64}$

$= 0.8$

Note-7 When the regression lines are in the format as:

y on x $\rightarrow$ $a_1x + b_1y = c_1$ $\quad (a_1x + b_1y - c_1 = 0)$

x on y $\rightarrow$ $a_2x + b_2y = c_2$ $\quad (a_2x + b_2y - c_2 = 0)$

I- When the two regr. lines are in the format as.

$c_1 y = ax + b$ — (1) $\longrightarrow$ y on x

$c_2 x = cy + d$ — (2) $\longrightarrow$ x on y

Q-3 ★ In a attack of WTC destroyed lab record of an analysis of the co-relⁿ & regr. data, the following results are left only,

Variance of $x = 9$

Regression lines are - $8x - 10y + 66 = 0$ (y on x)

$40x - 18y = 214$ (x on y)

Determine i) The mean value of data x and y

ii) The coeff. of co-relⁿ are.

iii) Std. deviation of y

$x = 13$

$y = 17$

$8x - 10y + 66 = 0$

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

$x = 13, y = 17$

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

ii) from (1);

$8x - 10y + 66 = 0$

$10y = 8x + 66$

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

Hence $b_{yx} = \frac{8}{10} = 0.8$

from (2):

$40x - 18y = 214$

$40x = 18y + 214$

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

Hence,

$b_{xy} = \frac{18}{40} = 0.45.$

But

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

$= \pm\sqrt{0.8 \times 0.45} = 0.6$

$r = \pm 0.6$

iii) Given-

$\sigma_x^2 = 9$

$\sigma_x = \pm 3$

We know that

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

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

$\sigma_y = \frac{\cancel{8}^{4} \times 3}{\cancel{6}_{2}}$

$\sigma_y = 6$

Non linear regression curve -

We calculate both regression curve x on y and y on x by the method of least squares of a given data.