MODULE-4 CURVE FITTING

Fitting of straight line $y = a + bx$ and $y = ax + b$

$y = a + bx$

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

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

$y = ax + b$

$\Sigma y = a\Sigma x + nb$

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

Fitting of parabola $y = a + bx + cx^2$ and $y = ax^2 + bx + c$

$y = a + bx + cx^2$

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

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

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

$y = ax^2 + bx + c$

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

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

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

Fitting of curves of the form $y = ab^x$, $y = ax^b$, $y = ae^{bx}$

$y = ab^x$

log on BS

$\log_e y = \log_e(ab^x)$

$\log_e y = \log_e a + \log_e b^x$

$\log_e y = \log_e a + x\log_e b$

$Y = A + BX \rightarrow (1)$

$Y = \log_e y \quad A = \log_e a$

$B = \log_e b \quad X = x$

$\Sigma Y = nA + B\Sigma X \rightarrow (2)$

$\Sigma XY = A\Sigma X + B\Sigma X^2 \rightarrow (3)$

$\log_e a = A$ & $a = e^A$

$\log_e b = B$ & $b = e^B$

$y = ab^x$

$y = ae^{bx}$

log on BS

$\log_e y = \log_e ae^{bx}$

$\log_e y = \log_e a + \log_e e^{bx}$

$\log_e y = \log_e a + bx\log_e e$

$\log_e y = \log_e a + bx$

$Y = A + bx$

$Y = \log_e y \quad A = \log_e a$ & $a = e^A$

$\Sigma Y = nA + b\Sigma x$

$\Sigma xY = A\Sigma x + b\Sigma x^2$

Scanned with CamScanner

$y = ax^b$

log on BS

$\log_e y = \log_e ax^b$

$\log_e y = \log_e a + b \log_e x$

$Y = A + bX$

$Y = \log_e y \quad A = \log_e a \quad X = \log_e x$

$\Sigma Y = nA + b\Sigma X$

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

Co-efficient of corelation and equation of lines of regression

Co-relation co-efficient $\quad r = \dfrac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$

$\bar{x} = \dfrac{\Sigma x}{n} \quad \bar{y} = \dfrac{\Sigma y}{n} \quad \bar{z} = \dfrac{\Sigma z}{n}$ where $z = x - y$

$\sigma_x^2 = \dfrac{\Sigma x^2}{n} - (\bar{x})^2 \quad \sigma_y^2 = \dfrac{\Sigma y^2}{n} - (\bar{y})^2 \quad \sigma_z^2 = \dfrac{\Sigma z^2}{n} - (\bar{z})^2$

$r = \dfrac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$

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

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

Regression and co-efficient of corelation

Find $\bar{x}\ \bar{y}$

$X = x - \bar{x}$, $Y = y - \bar{y}$, XY, X^2, Y^2

$Y = \dfrac{\Sigma XY}{\Sigma X^2} \cdot X$

$y - \bar{y} = \dfrac{\Sigma XY}{\Sigma X^2}(x - \bar{x})$

$X = \dfrac{\Sigma XY}{\Sigma Y^2} \cdot Y$

$x - \bar{x} = \dfrac{\Sigma XY}{\Sigma Y^2}(y - \bar{y})$

$r = \pm\sqrt{(\text{co-eff of } x)(\text{co-eff of } y)}$

$r = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}}$

In this topic we discuss the method of finding a specific relation $y = f(x)$ for the data to satisfy as accurately as possible and such as equation is called the best fitting equation or the a curve of best fit.

The method is called as the method of least squares described as follows.

Suppose $y = f(x)$ is an approximate relation that fits into a given data (x_i, y_i) $i = 1, 2, 3 \ldots n$ then y_i's are called the observed values and $y_i = f(x_i)$ are called the expected values. Their difference $R_i = y_i - Y_i$ are called the residual or estimate errors.

<u>Fitting of a straight line $y = a + bx$</u>

consider a set of n given values (x, y) for fitting the straight line $y = a + bx$ where a and b are parameters to be determine.

Normal equations for fitting the straight line $y = a + bx$

$$y = a + bx \rightarrow (1)$$

$$\Sigma y = na + b\Sigma x \rightarrow (2) \qquad \sum_1^n a = a + a + a \ldots = \underline{na}$$

x^{ly} x in (1)

$$xy = xa + bx^2$$

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

$\therefore$ The normal equations of $y = a + bx$ are

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

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

<u>NOTE</u> :- Normal equations for $y = ax + b$ are

$$\Sigma y = a\Sigma x + nb$$

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

<u>working procedure for problems</u>

1) we first write the normal equations appropriate to curve of fit

2) we prepare the relevant table and find the value of summation present in normal equations we substitute these values to arrive at

a system of equations in unknown parameters

3) we find the parameters by solving and substitute in given equations

problems

1) Fit a straight line $y = a + bx$ in the least square sence for the data.

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

Normal equations for $y = a + bx$ are

$$\Sigma y = na + b\Sigma x \rightarrow ①$$

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

$\boxed{n=8}$

x	y	xy	x^2
1	1	1	1
3	2	6	9
4	4	16	16
6	4	24	36
8	5	40	64
9	7	63	81
11	8	88	121
14	9	126	196
56	40	364	524

Normal equations of ① and ② becomes

$$40 = 8a + 56b$$

$$364 = 56a + 524b$$

$\boxed{a = 0.545}$ $\boxed{b = 0.636}$

put a and b in $y = a + bx$

$$y = 0.545 + 0.636x$$

2) Fit a straight line $y = a + bx$ for the data

x	0	1	2	3	4	5	6
y	2	1	3	2	4	3	5

Normal equations for $y = a + bx$ are

$$\Sigma y = na + b\Sigma x \rightarrow ①$$

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

$\boxed{n=7}$

x	y	xy	x^2
0	2	0	0
1	1	1	1
2	3	6	4
3	2	6	9
4	4	16	16
5	3	15	25
6	5	30	36
21	20	74	91

Normal equations ① and ② becomes

$$20 = 7a + 21b$$

$$74 = 21a + 91b$$

$\boxed{a = 1.357}$ $\boxed{b = 0.5}$

put a and b in $y = a + bx$

$$y = 1.357 + 0.5x$$

3) Find the equation of best fitting straight line $y = ax+b$ for the following data

x	5	10	15	20	25
y	16	19	23	26	30

Normal equations for $y = ax+b$ are

$$\Sigma y = a\Sigma x + nb \rightarrow ①$$

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

$\boxed{n=5}$

$\left\{ \sum_1^n b = nb \right.$

$y = ax+b \rightarrow ①$

$\Sigma y = a\Sigma x + nb$

x^{ly} x in ①

$xy = ax^2 + bx$

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

x	y	xy	x^2
5	16	80	25
10	19	190	100
15	23	345	225
20	26	520	400
25	30	750	625
75	114	1885	1375

Normal equations ① and ② becomes

$$114 = 75a + 5b$$

$$1885 = 1375a + 75b$$

$\boxed{a = 0.7}$ $\boxed{b = 12.3}$

put a and b in $y = ax+b$

$$y = 0.7x + 12.3$$

4) Find the equation of the best fitting straight line for the data.

x	0	1	2	3	4	5
y	9	8	24	28	26	20

Scanned with CamScanner

we shall fit the straight line $y = ax + b$ for the given data

The normal equations are

$\Sigma y = a\Sigma x + nb \rightarrow ①$

$\Sigma xy = a\Sigma x^2 + b\Sigma x \rightarrow ②$ $\boxed{n=6}$

x	y	xy	x^2
0	9	0	0
1	8	8	1
2	24	48	4
3	28	84	9
4	26	104	16
5	20	100	25
15	115	344	55

Normal equations ① and ② becomes

$115 = 15a + 6b$

$344 = 55a + 15b$

$\boxed{a = 3.228}$ $\boxed{b = 11.09}$

put a and b in $y = ax + b$

$y = 3.228x + 11.09$

5) fit a straight line for the data

x	50	70	100	120
y	12	15	21	25

we shall fit the straight line $y = ax + b$ for the given data.

The normal equations are

$\Sigma y = a\Sigma x + nb \rightarrow ①$

$\Sigma xy = a\Sigma x^2 + b\Sigma x \rightarrow ②$ $\boxed{n=4}$

x	y	xy	x^2
50	12	600	2500
70	15	1050	4900
100	21	2100	10000
120	25	3000	14400
340	73	6750	31800

Normal equations ① and ② becomes

$73 = 340a + 4b$

$6700 = 31800a + 340b$

$\boxed{a = 0.188}$ $\boxed{b = 2.23}$

put a and b in $y = ax + b$

$y = 0.188x + 2.23$

6) A simply supported beam carries a concentrated load p at its midpoint corresponding to various values of p the maximum deflecti-on y is measured and is given in the following table

P	100	120	140	160	180	200
y	0.45	0.55	0.60	0.70	0.80	0.85

Scanned with CamScanner

Find the law of the form $y = a + bp$ and hence estimate y when $p = 150$.

$y = a + bp$

Normal equations are

$\Sigma y = na + b\Sigma p \rightarrow ①$

$\Sigma py = a\Sigma p + b\Sigma p^2 \rightarrow ②$

$\boxed{n = 6}$

p	y	py	p^2
100	0.45	45	10000
120	0.55	66	14400
140	0.60	84	19600
160	0.70	112	25600
180	0.80	144	32400
200	0.85	170	40000
900	3.95	621	142000

Normal equations ① and ① becomes

$3.95 = 6a + 900b$

$621 = 900a + 142000b$

$\boxed{a = 0.048}$ $\boxed{b = 0.004}$

put a and b in $y = a + bp$

$y = 0.048 + 0.004p$.

when $p = 150$

$\boxed{y = 0.65}$

Fitting of a second degree parabola $y = a + bx + cx^2$

consider a set of n given values (x, y) for fitting the curve of $y = a + bx + cx^2$ where a, b and c are parameters to be determined

$y = a + bx + cx^2 \rightarrow ①$

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

x by x in eqn ①

$xy = xa + bx^2 + cx^3 \rightarrow ②$

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

x by x in eqn ②

$x^2y = ax^2 + bx^3 + cx^4$

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

Normal equations are

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

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

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

Problems

1) Fit a parabola of second degree $y = a + bx + cx^2$ for the data

x	0	1	2	3	4
y	1	1.8	1.3	2.5	2.3

Normal equations for $y = a + bx + cx^2$ are

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

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

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

$\boxed{n = 5}$

x	y	xy	x^2	x^2y	x^3	x^4
0	1	0	0	0	0	0
1	1.8	1.8	1	1.8	1	1
2	1.3	2.6	4	5.2	8	16
3	2.5	7.5	9	22.5	27	81
4	2.3	9.2	16	36.8	64	256
10	8.9	21.1	30	66.3	100	354

Normal equations become

$$5a + 10b + 30c = 8.9$$

$$10a + 30b + 100c = 21.1$$

$$30a + 100b + 354c = 66.3$$

$\boxed{a = 1.07}$ $\boxed{b = 0.415}$ $\boxed{c = -0.021}$

put a, b and c in $y = a + bx + cx^2$

$$y = 1.07 + 0.415x - 0.021x^2$$

2) Fit a parabola $y = a + bx + cx^2$ by the method of least squares for the data.

x	2	4	6	8	10
y	3.07	12.85	31.47	57.38	91.29

x	y	x^2	x^3	x^4	xy	x^2y
2	3.07	4	8	16	6.14	12.28
4	12.85	16	64	256	51.4	205.6
6	31.47	36	216	1296	188.82	1132.9
8	57.38	64	512	4096	459.04	3672.3
10	91.29	100	1000	10000	912.9	9129
30	196.06	220	1800	15664	1618.3	14152.1

Normal equations for $y = a + bx + cx^2$ are

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

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

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

$\boxed{n = 5}$

Normal equations become

$$5a + 30b + 220c = 196.06, \quad 30a + 220b + 1800c = 1618.3$$

$$220a + 1800b + 15664c = 14152.13$$

$\boxed{a = 0.696}$ $\boxed{b = -0.85}$ $\boxed{c = 0.99}$

put a, b and c in $y = a + bx + cx^2$

$$y = a + bx + cx^2$$

$$y = 0.696 - 0.85x + 0.99x^2$$

3) Fit a second degree parabola $y = A + Bx + Cx^2$ in the least square sense for the following data and hence find y at $x = 6$.

x	1	2	3	4	5
y	10	12	13	16	19

The normal equations of $y = A + Bx + Cx^2$ are

$$\Sigma y = nA + B\Sigma x + C\Sigma x^2$$

$$\Sigma xy = A\Sigma x + B\Sigma x^2 + C\Sigma x^3$$

$$\Sigma x^2 y = A\Sigma x^2 + B\Sigma x^3 + C\Sigma x^4$$

$\boxed{n = 5}$

x	y	x^2	x^3	x^4	xy	x^2y
1	10	1	1	1	10	10
2	12	4	8	16	24	48
3	13	9	27	81	39	117
4	16	16	64	256	64	256
5	19	25	125	625	95	475
15	70	55	225	979	232	906

Normal equations become

$$5A + 15B + 55C = 70$$

$$15A + 55B + 225C = 232$$

$$55A + 225B + 979C = 906$$

$\boxed{A = 9.4}$ $\boxed{B = 0.48}$ $\boxed{C = 0.28}$

put A, B and C in $y = A + Bx + Cx^2$

$$y = 9.4 + 0.48x + 0.28x^2$$

when $x = 6$

$$y = 9.4 + 0.48 \times 6 + 0.28 \times 6^2$$

$$\boxed{y = 22.36}$$

4) Fit a second degree parabola $y = Ax^2 + Bx + C$ in the least square sense for the following data and hence find y at $x = 6$

x	1	2	3	4	5
y	10	12	13	16	19

Scanned with CamScanner

Normal equations of $y = Ax^2 + Bx + C \rightarrow ①$ are

$$\Sigma y = A\Sigma x^2 + B\Sigma x + nC$$

x^{ly} x in eqn ①

$$xy = Ax^3 + Bx^2 + Cx \rightarrow ②$$

$$\Sigma xy = A\Sigma x^3 + B\Sigma x^2 + C\Sigma x$$

x^{ly} x in eqn ②

$$x^2y = Ax^4 + Bx^3 + Cx^2 \rightarrow ③$$

$$\Sigma x^2 y = A\Sigma x^4 + B\Sigma x^3 + C\Sigma x^2$$

$\therefore$ Normal equations are

$$A\Sigma x^2 + B\Sigma x + nC = \Sigma y$$

$$A\Sigma x^3 + B\Sigma x^2 + C\Sigma x = \Sigma xy \qquad \boxed{n=5}$$

$$A\Sigma x^4 + B\Sigma x^3 + C\Sigma x^2 = \Sigma x^2 y$$

x	y	x^2	x^3	x^4	xy	x^2y
1	10	1	1	1	10	10
2	12	4	8	16	24	48
3	13	9	27	81	39	117
4	16	16	64	256	64	256
5	19	25	125	625	95	475
15	70	55	225	979	232	906

Normal equations becomes

$$55A + 15B + 5C = 70$$

$$225A + 55B + 15C = 232$$

$$979A + 225B + 55C = 906$$

$\boxed{A = 0.28}$ $\boxed{B = 0.48}$ $\boxed{C = 9.4}$

put A, B and C in $y = Ax^2 + Bx + C$

$$y = 0.28x^2 + 0.48x + 9.4$$

when $x = 6$

$$y = 0.28 \times 36 + 0.48 \times 6 + 9.4$$

$$y = \underline{22.36}$$

---

5) Fit a parabola for the data

x	1	2	3	4	5	6	7	8	9
y	2	6	7	8	10	11	11	10	9

we shall fit the parabola $y = a + bx + cx^2$ for the given data. The normal equations are

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

$\boxed{n=9}$

x	y	x^2	x^3	x^4	xy	x^2y
1	2	1	1	1	2	2
2	6	4	8	16	12	24
3	7	9	27	81	21	63
4	8	16	64	256	32	128
5	10	25	125	625	50	250
6	11	36	216	1296	66	396
7	11	49	343	2401	77	539
8	10	64	512	4096	80	640
9	9	81	729	6561	81	729
45	74	285	2025	15333	421	2771

Normal equations becomes

$$9a + 45b + 285c = 74$$
$$45a + 285b + 2025c = 421$$
$$285a + 2025b + 15333c = 2771$$

$\boxed{a = -0.928}$ $\boxed{b = 3.52}$ $\boxed{c = -0.26}$

put a, b and c in $y = a + bx + cx^2$

$$y = -0.928 + 3.52x + [-0.26x^2]$$
$$y = -0.928 + 3.52x - 0.26x^2$$

6) Find the best values of a, b, c if the equation $y = a + bx + cx^2$ is to fit most closely to the following observations

x	-2	-1	0	1	2
y	-3.15	-1.39	0.62	2.88	5.378

Normal equations of $y = a + bx + cx^2$

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

$\boxed{n=5}$

x	y	x^2	x^3	x^4	xy	x^2y
-2	-3.15	4	-8	16	6.3	-12.6
-1	-1.39	1	-1	1	1.39	-1.39
0	0.62	0	0	0	0	0
1	2.88	1	1	1	2.88	2.88
2	5.378	4	8	16	10.756	21.512
0	4.338	10	0	34	21.326	10.402

Normal equations becomes

$$5a + 0b + 10c = 4.338$$
$$0a + 10b + 0c = 21.326$$
$$10a + 0b + 34c = 10.402$$

$\boxed{a = 0.621}$ $\boxed{b = 2.132}$ $\boxed{c = 0.123}$

put a, b and c in $y = a + bx + cx^2$

$$y = 0.621 + 2.132x + 0.123x^2$$

Fitting of curves of the form $y = ab^x$, $y = ax^b$, $y = ae^{bx}$

Consider $y = ab^x$

Take log on both sides [to base e]

$$\log_e y = \log_e (ab^x)$$

$$\log_e y = \log_e a + \log_e b^x$$

$$\log_e y = \log_e a + x \log_e b$$

$$Y = A + BX \rightarrow (1)$$

where $Y = \log_e y$ $A = \log_e a$ $B = \log_e b$ $X = x$

Normal equations becomes

$$\Sigma Y = nA + B\Sigma X \rightarrow (2)$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2 \rightarrow (3)$$

Solving (2) and (3) we obtain A and B we have

$\log_e a = A$ $\log_e b = B$

$a = e^A$ $b = e^B$

Substitute the values of a and b in $y = ab^x$

---

problems

1) Fit a curve of the form $y = ab^x$ in the least square sense for the following data

x	0	2	4	5	7	10
y	100	120	256	390	710	1600

Consider $y = ab^x$

apply log on both sides

$$\log_e y = \log_e (ab^x)$$

$$\log_e y = \log_e a + \log_e b^x$$

$$\log_e y = \log_e a + x \log_e b$$

$$Y = A + BX$$

where $Y = \log_e y$

$A = \log_e a$ & $a = e^A$ $B = \log_e b$ & $b = e^B$ $X = x$

Scanned with CamScanner

Normal equations of $Y = A + BX$

$$\Sigma Y = nA + B\Sigma X$$

$\boxed{n = 6}$

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

$x = X$	y	$Y = \log_e y$	X^2	XY
0	100	4.6052	0	0
2	120	4.7875	4	9.575
4	256	5.5452	16	22.1808
5	390	5.9661	25	29.830
7	710	6.5653	49	45.957
10	1600	7.3778	100	73.778
28		34.8471	194	181.321

Normal equations becomes

$$6A + 28B = 34.8471$$

$$28A + 194B = 181.3214$$

By solving we get

$\boxed{A = 4.4298}$ $\boxed{B = 0.2953}$

But $a = e^A = e^{4.4298}$ $a = \underline{83.914}$

$b = e^B = e^{0.2953}$ $b = \underline{1.343}$

Then required curve is $y = \underline{(83.914)(1.3435)^x}$

---

2) Fit a curve of the form $y = ab^x$ for the data hence find the estimation for y when $x = 8$

x	1	2	3	4	5	6	7
y	87	97	113	129	202	195	193

consider $y = ab^x$

Apply log on both sides

$$\log_e y = \log_e(ab^x)$$

$$\log_e y = \log_e a + \log_e b^x$$

$$\log_e y = \log_e a + x\log_e b$$

$$Y = A + BX$$

where $Y = \log_e y$ $A = \log_e a$ $B = \log_e b$ $X = x$

Normal equation of $Y = A + BX$ are

$$\Sigma Y = nA + B\Sigma X$$

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

$\boxed{n = 7}$

$x=X$	y	$Y=\log_e y$	XY	X^2
1	87	4.47	4.47	1
2	97	4.57	9.14	4
3	113	4.73	14.19	9
4	129	4.86	19.44	16
5	202	5.31	26.55	25
6	195	5.27	31.62	36
7	193	5.26	36.82	49
28		34.47	142.23	140

Normal equations becomes

$$7A + 28B = 34.47$$

$$28A + 140B = 142.23$$

$\boxed{A = 4.303}$ $\boxed{B = 0.1554}$

$A = \log_e a$ $\quad$ $B = \log_e b$

$a = e^A$ $\quad$ $b = e^B$

$a = e^{4.303}$ $\quad$ $b = e^{0.1554}$

$a = 73.92$ $\quad$ $b = 1.1681$

put a and b in $y = ab^x$

$$y = (73.92)(1.1681)^x$$

when $x = 8$

$$y = (73.92)(1.1681)^8$$

$$\boxed{y = 256.18}$$

3) At constant temperature the pressure p and volume v of a gas are connected by relation pv^γ = constant. Find the best fitting equation of this form to the following data and estimate v when p=4

p(kg.sq cm)	0.5	1.0	1.5	2.0	2.5	3.0
v (c.c)	1620	1000	750	620	520	460

consider $pv^\gamma = k$ where k is constant take log on BS

$$\log_e pv^\gamma = \log_e k$$

$$\log_e p + \gamma \log_e v = \log_e k$$

$$\log_e p = \log_e k - \gamma \log_e v$$

let us take $\log_e p = y$ $\quad$ $\log_e v = x$

$\log_e k = a$ $\quad$ $-\gamma = b$

so that we have

$y = a + bx$ which is a straight line

The associated normal equations are

$$\Sigma y = na + b\Sigma x \rightarrow ①$$

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

$\boxed{n=6}$

P	V	$x = \log_e V$	$y = \log_e P$	xy	x^2
0.5	1620	7.39	-0.69	-5.099	54.612
1.0	1000	6.91	0	0	47.748
1.5	750	6.62	0.40	2.648	43.824
2.0	620	6.43	0.69	4.436	41.344
2.5	520	6.25	0.92	5.75	39.062
3.0	460	6.13	1.1	6.743	37.576
		39.73	2.42	14.478	264.163

Normal equations ① and ② Becomes

$$6a + 39.73b = 2.42$$

$$39.73a + 264.1689b = 14.47$$

$\boxed{a = 9.7934}$ $\boxed{b = -1.42}$

But $\log_e K = a$ & $K = e^a$

$\boxed{K = 17915}$

$-\gamma = b \Rightarrow b = 1.42$

The curve of fit $pv^{\gamma} = k$

$$pv^{1.42} = 17915$$

when $p = 4$, $4v^{1.42} = 17915$

$$v^{1.42} = \frac{17915}{4} \text{ \& } v^{1.42} = 4478$$

$$v = (4478)^{1/1.42} = 372.59 \simeq 373$$

when $\boxed{p=4}$ $\boxed{v=373}$

---

4) Fit a curve of the form $y = ae^{bx}$ for the data

x	0	2	4
y	8.12	10	31.82

consider $y = ae^{bx}$

Apply log on both sides

$$\log_e y = \log_e [ae^{bx}]$$

$$\log_e y = \log_e a + bx \log_e e$$

$\boxed{\log_e e = 1}$

$$\log_e y = \log_e a + bx$$

$$Y = A + bx$$

where $Y = \log_e y$ $\quad A = \log_e a$ & $a = e^A$

Normal equations are

x	y	$Y = \log_e y$	xy	x^2
0	8.12	2.09	0	0
2	10	2.30	4.60	4
4	31.82	3.46	13.8	16
6		7.85	18.4	20

$$\Sigma Y = nA + b\Sigma x$$

$$\Sigma xY = A\Sigma x + b\Sigma x^2 \quad \boxed{n=3}$$

Normal equations Becomes

$$3A + 6B = 7.85$$

$$6A + 20B = 18.44$$

$$\boxed{A = 1.932} \quad \boxed{B = 0.3425}$$

But $A = \log_e a \Rightarrow a = e^A = e^{1.932} = 6.9033$

put a and b in $y = ae^{bx}$

$$y = 6.9033\, e^{0.3425x}$$

5) Find the equation of the best fitting curve in the form $y = ae^{bx}$ for the data

x	5	6	7	8	9	10
y	133	55	23	7	2	2

consider $y = ae^{bx}$

apply Log on both sides

$$\log_e y = \log_e [ae^{bx}] \quad \boxed{\log_e e = 1}$$

$$\log_e y = \log_e a + bx \log_e e$$

$$\log_e y = \log_e a + bx$$

$$Y = A + bx$$

where $Y = \log_e y$

$$A = \log_e a \Rightarrow a = e^A$$

Normal equations Becomes

$$\Sigma Y = nA + b\Sigma x$$

$$\Sigma xY = A\Sigma x + b\Sigma x^2$$

$$\boxed{n = 6}$$

x	y	$Y = \log_e y$	xY	x^2
5	133	4.89	24.45	25
6	55	4.007	24.042	36
7	23	3.135	21.945	49
8	7	1.946	15.568	64
9	2	0.693	6.237	81
10	2	0.693	6.93	100
45		15.364	99.172	355

Normal equations becomes

$$6A + 45b = 15.364$$

$$45A + 355b = 99.172$$

$\boxed{A = 9.443}$ $\boxed{b = -0.92}$

$A = \log_e a$ & $a = e^A$ & $e^{9.443}$

$a = 12619$

put a and b in $y = ae^{bx}$

$$y = 12619\, e^{-0.92x}$$

6) Fit a least square geometric curve $y = ax^b$ from the following data

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

consider $y = ax^b$

Apply Log on Both sides

$$\log_e y = \log_e ax^b$$

$$\log_e y = \log_e a + \log_e x^b$$

$$\log_e y = \log_e a + b \log_e x$$

$$Y = A + BX$$

where $Y = \log_e y$ $A = \log_e a$ $X = \log_e x$

Normal equations are

$$\Sigma Y = nA + B\Sigma X$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$ $\boxed{n=5}$

$$5A + 4.787b = 6.1092$$

$$4.787A + 6.1993b = 9.0804$$

$\boxed{A = -0.693}$ $\boxed{b = 2}$

Scanned with CamScanner

x	y	$X = \log_e x$	$Y = \log_e y$	X^2	XY
1	0.5	0	− 0.6931	0	0
2	2	0.6931	0.6931	0.4804	0.4804
3	4.5	1.0986	1.5041	1.2069	1.6524
4	8	1.3863	2.0794	1.9218	2.8827
5	12.5	1.6094	2.5257	2.5902	4.0649
		4.7874	6.1092	6.1993	9.0804

$A = \log_e a \Rightarrow a = e^A$

$a = e^{-0.693}$

$\boxed{a = 0.5}$ $\boxed{b = 2}$

put a and b in $y = ax^b$

$y = (0.5)x^2$

Correlation : co-variation of two independent magnitudes is known as co-relation.

Co-relation co-efficient : The numerical measure of corelation between two variables x and y is known as pearsons co-efficient of correlation usually denoted by r.

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{xy}^2}{2\sigma_x \sigma_y}$$

Regression : It is an estimation of 1 independent variable in terms of the order.

The best fitting straight line of the form $y = ax+b$ is called regression line of y on x and $x = ay + b$ is called regression line of x on y

Working procedure to find the co-efficient of corelation and equation of lines of regression

1) Prepare the table showing the columns x, y, z and x^2, y^2, z^2 and finding the summation of x, y, z, x^2, y^2, z^2

2) Find $\bar{x} = \frac{\Sigma x}{n}$, $\bar{y} = \frac{\Sigma y}{n}$, $\bar{z} = \frac{\Sigma z}{n}$ where $z = x - y$

3) Find $\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2$, $\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2$ and

$$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2$$

4) $r = \dfrac{a_x^2 + a_y^2 - a_z^2}{2 a_x a_y}$

5) Find the lines of regression

$$y - \bar{y} = r \frac{a_y}{a_x} (x - \bar{x})$$

$$x - \bar{x} = r \frac{a_x}{a_y} (y - \bar{y})$$

NOTE : Find the lines of regression and co-efficient of corelation

1) Find $\bar{x}, \bar{y}$

2) prepare the table showing $X = x - \bar{x}$, $Y = y - \bar{y}$, XY, X^2, Y^2

3) Find ΣXY, ΣX^2, ΣY^2

4) Find regression lines i.e in the form of

a) $Y = \dfrac{\Sigma XY}{\Sigma X^2} \cdot X$

$y - \bar{y} = \dfrac{\Sigma XY}{\Sigma X^2} (x - \bar{x})$ regression line y on x

b) $X = \dfrac{\Sigma XY}{\Sigma Y^2} Y$

$x - \bar{x} = \dfrac{\Sigma XY}{\Sigma Y^2} (y - \bar{y})$ regression line x on y

5) Find the corelation co-efficient

$$r = \pm \sqrt{(\text{co-eff of } x)(\text{co-eff of } y)}$$

co-relation co-efficient can also be written as

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2} \sqrt{\Sigma Y^2}}$$

NOTE : co-efficient of co-relation numerically does not exceed unity i.e $-1 \le r \le 1$

Problems

1) Compute the co-efficient of correlation and the equation of the lines of regression for the data

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

$\boxed{n=7}$

x	y	$z=x-y$	x^2	y^2	z^2
1	9	-8	1	81	64
2	8	-6	4	64	36
3	10	-7	9	100	49
4	12	-8	16	144	64
5	11	-6	25	121	36
6	13	-7	36	169	49
7	14	-7	49	196	49
28	77	-49	140	875	347

$$\bar{x} = \frac{\Sigma x}{n} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{77}{7} = 11$$

$$\bar{z} = \frac{\Sigma z}{n} = \frac{-49}{7} = -7$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{140}{7} - (4)^2 = 20-16 = 4$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{875}{7} - (11)^2 = 125-121 = 4$$

$$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = \frac{347}{7} - (-7)^2 = 49.57-49 = 0.57$$

Substitute all these in

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x \sigma_y} = \frac{4+4-0.57}{2 \times \sqrt{4} \times \sqrt{4}} = \frac{7.43}{8} = 0.928 \simeq 0.93$$

$$\boxed{r = 0.93}$$

The lines of regression are given by

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

$$y - 11 = \frac{0.93 \times \cancel{2}}{\cancel{2}}(x-4)$$

$$y - 11 = 0.93x - 3.72$$

$$y = 0.93x - 3.72 + 11$$

$$y = 0.93x + 7.28$$

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

$$x - 4 = \frac{(0.93)\cancel{(2)}}{\cancel{(2)}}(y-11)$$

$$x - 4 = 0.93y - 10.23$$

$$x = 0.93y - 10.23 + 4$$

$$x = 0.93y - 6.23$$

These are lines of regression

2) Find the correlation co-efficient for two groups and equation of lines of regression

A	92	89	87	86	83	77	71	63	53	50
B	86	83	91	87	68	85	52	82	37	57

$\boxed{n=10}$

A = x	B = y	Z = x - y	x^2	y^2	z^2
92	86	6	8464	7396	36
89	83	6	7921	6889	36
87	91	-4	7569	8281	16
86	77	9	7396	5921	81
83	68	15	6889	4624	225
77	85	-8	5929	7225	64
71	52	19	5041	2704	361
63	82	-19	3969	6724	361
53	37	16	2809	1369	256
50	57	-7	2500	3249	49
751	718	33	58487	54390	1485

$$\bar{x} = \frac{\Sigma x}{n} = \frac{751}{10} = 75.1$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{718}{10} = 71.8$$

$$\bar{z} = \frac{\Sigma z}{n} = \frac{33}{10} = 3.3$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{58487}{10} - (75.1)^2$$

$$\sigma_x^2 = 208.69 \quad \sigma_x = 14.446$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{54390}{10} - (71.8)^2$$

$$\sigma_y^2 = 283.76 \quad \sigma_y = 16.845$$

$$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2$$

$$\sigma_z^2 = \frac{1485}{10} - (3.3)^2$$

$$\sigma_z^2 = 148.5 - 10.89$$

$$\sigma_z^2 = 137.61$$

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x \sigma_y} = \frac{208.69 + 283.76 - 137.61}{2 \times 14.446 \times 16.845}$$

$$r = 0.729$$

$$\boxed{r = 0.73}$$

The lines of regression are given by

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

$$y - 71.8 = \frac{0.73 \times 16.845}{14.446}(x - 75.1)$$

$$y - 71.8 = 0.85(x - 75.1)$$

$$y - 71.8 = 0.85x - 63.84$$

$$y = 0.85x - 63.84 + 71.8$$

$$y = 0.85x + 7.965$$

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

$$x - 75.1 = \frac{0.73 \times 14.446}{16.845}(y - 71.8)$$

$$x - 75.1 = 0.63(y - 71.8)$$

$$x - 75.1 = 0.63y - 45.234$$

$$x = 0.63y - 45.234 + 75.1$$

$$x = 0.63y + 29.866$$

3) Find the correlation co-efficient and equation of lines of a regression for the following values of x and y

x	1	2	3	4	5
y	2	5	3	8	7

$\boxed{n = 5}$

Scanned with CamScanner

x	y	z=x-y	x^2	y^2	z^2
1	2	-1	1	4	1
2	5	-3	4	25	9
3	3	0	9	9	0
4	8	-4	16	64	16
5	7	-2	25	49	4
15	25	-10	55	151	30

$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5} = 3$

$\bar{y} = \frac{\Sigma y}{n} = \frac{25}{5} = 5$

$\bar{z} = \frac{\Sigma z}{n} = \frac{-10}{5} = -2$

$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{55}{5} - (3)^2 = 11 - 9 = 2$ & $\sqrt{2}$

$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{151}{5} - (5)^2 = 30.2 - 25 = 5.2$ and $\sqrt{5.2}$

$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = \frac{30}{5} - (-2)^2 = 6 - 4 = 2$

$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x \sigma_y} = \frac{2 + 5.2 - 2}{2 \times \sqrt{2} \times \sqrt{5.2}} = 0.806$ $\boxed{r = 0.81}$

The lines of regression are given by

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

$y - 5 = \frac{0.81 \times \sqrt{5.2}}{\sqrt{2}}(x - 3)$

$y - 5 = 1.306(x - 3)$

$y - 5 = 1.306x - 3.918$

$y = 1.306x - 3.918 + 5$

$y = 1.306x + 1.082$

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

$x - 3 = \frac{0.81 \times \sqrt{2}}{\sqrt{5.2}}(y - 5)$

$x - 3 = 0.502(y - 5)$

$x - 3 = 0.502y - 2.51$

$x = 0.502y - 2.51 + 3$

$x = 0.502y + 0.49$

4) Obtain the lines of regression and hence find the co-efficient of correlation of the data

x	1	3	4	2	5	8	9	10	13	15
y	8	6	10	8	12	16	16	10	32	32

$\boxed{n = 10}$

$\bar{x} = \frac{\Sigma x}{n}$

$\bar{x} = \frac{70}{10}$

$\bar{x} = 7$

$\bar{y} = \frac{\Sigma y}{n}$

$\bar{y} = \frac{150}{10}$

$\bar{y} = 15$

Scanned with CamScanner

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	X^2	Y^2	XY
1	8	-6	-7	36	49	42
3	6	-4	-9	16	81	36
4	10	-3	-5	9	25	15
2	8	-5	-7	25	49	35
5	12	-2	-3	4	9	6
8	16	1	1	1	1	1
9	16	2	1	4	1	2
10	10	3	-5	9	25	-15
13	32	6	17	36	289	102
15	32	8	17	64	289	136
70	150			204	818	360

we have lines of regression in the form

$$Y = \frac{\Sigma XY}{\Sigma X^2} X$$

$$y-\bar{y} = \frac{360}{204}(x-\bar{x})$$

$$y-15 = 1.764(x-7)$$

$$y = 1.764x - 12.348 + 15$$

$$\underline{y = 1.764x + 2.652}$$

$$X = \frac{\Sigma XY}{\Sigma Y^2} Y$$

$$x-\bar{x} = \frac{360}{818}(y-\bar{y})$$

$$x-7 = 0.44(y-15)$$

$$x-7 = 0.44y - 6.6$$

$$\underline{x = 0.44y + 0.4}$$

$y = 1.764x + 2.652$ and

$x = 0.44y + 0.4$ are lines of regression.

co-efficient of corelation is

$$r = \sqrt{[\text{co-eff of } x][\text{co-eff of } y]}$$

$$r = \sqrt{1.764 \times 0.44}$$

$$\boxed{r = 0.88}$$

sign of r is positive since both the regression co-efficients are positive

5) Find the corelation co-efficient between x and y for the following data. Also obtain the regression lines

x	1	2	3	4	5	6	7	8	9	10
y	10	12	16	28	25	36	41	49	40	50

$\boxed{n=10}$

$$\bar{x} = \frac{\Sigma x}{n} \qquad \bar{y} = \frac{\Sigma y}{n} \qquad \bar{z} = \frac{\Sigma z}{n}$$

$$\bar{x} = \frac{55}{10} \qquad \bar{y} = \frac{307}{10} \qquad \bar{z} = -\frac{252}{10}$$

$$\bar{x} = 5.5 \qquad \bar{y} = 30.7 \qquad \bar{z} = -25.2$$

Scanned with CamScanner

x	y	z = x - y	x^2	y^2	z^2
1	10	-9	1	100	81
2	12	-10	4	144	100
3	16	-13	9	256	169
4	28	-24	16	784	576
5	25	-20	25	625	400
6	36	-30	36	1296	900
7	41	-34	49	1681	1156
8	49	-41	64	2401	1681
9	40	-31	81	1600	961
10	50	-40	100	2500	1600
55	307	-252	385	11387	7624

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{385}{10} - (5.5)^2$$

$$\sigma_x^2 = 38.5 - 30.25 = 8.25$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{11387}{10} - (30.7)^2$$

$$\sigma_y^2 = 1138.7 - 942.49 = 196.21$$

$$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = \frac{7624}{10} - (-25.2)^2$$

$$\sigma_z^2 = 762.4 - 635.04 = 127.36$$

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x \sigma_y} = \frac{8.25 \times 196.21 \times 127.3}{2 \times \sqrt{8.25} \times \sqrt{196.21}}$$

$$\boxed{r = 0.958}$$

The lines of regression are given by

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

$$y - 30.7 = \frac{0.958 \times \sqrt{196.21}}{\sqrt{8.25}}(x - 5.5)$$

$$y - 30.7 = 0.958 \times 4.876(x - 5.5)$$

$$y - 30.7 = 4.671(x - 5.5)$$

$$y = 4.671x - 25.691 + 30.7$$

$$y = 4.671x + 5.009$$

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

$$x - 5.5 = \frac{0.958 \times \sqrt{8.25}}{\sqrt{196.21}}(y - 30.7)$$

$$x - 5.5 = 0.1964(y - 30.7)$$

$$x = 0.1964y - 6.029 + 5.5$$

$$x = 0.1964y - 0.529$$

6) the following data lines gives the age of husband (x) and age of wife (y) in years. Form the two regression lines and calculate the age of husband corresponding to 16 years age of wife.

x	36	23	27	28	28	29	30	31	33	35
y	29	18	20	22	27	21	29	27	29	28

$\boxed{n = 10}$

x	y	$z = x - y$	x^2	y^2	z^2
36	29	7	1296	841	49
23	18	5	529	324	25
27	20	7	729	400	49
28	22	6	784	484	36
28	27	1	784	729	1
29	21	8	841	441	64
30	29	1	900	841	1
31	27	4	961	729	16
33	29	4	1089	841	16
35	28	7	1225	784	49
300	250	50	9138	6414	306

$$\bar{x} = \frac{\Sigma x}{n} = \frac{300}{10} = 30$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{250}{10} = 25$$

$$\bar{z} = \frac{\Sigma z}{n} = \frac{50}{10} = 5$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{9138}{10} - (30)^2$$

$$\sigma_x^2 = 913.8 - 900 = 13.8 = 3.71$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{6414}{10} - (25)^2$$

$$\sigma_y^2 = 641.4 - 625 = 16.4 = 4.04$$

$$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = \frac{306}{10} - (5)^2$$

$$\sigma_z^2 = 30.6 - 25 = 5.6 = 2.366$$

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x \sigma_y}$$

$$r = \frac{13.8 + 16.4 - 5.6}{2 \times 3.71 \times 4.04} \qquad \boxed{r = 0.82}$$

The Regression Lines are given by

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

$$y - 25 = \frac{0.82 \times 4.04}{3.71}(x - 30)$$

$$y - 25 = 0.8939(x - 30)$$

$$y = 0.8939x - 26.817 + 25$$

$$y = 0.8939x - 1.817$$

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

$$x - 30 = \frac{0.82 \times 3.71}{4.04}(y - 25)$$

$$x - 30 = 0.7521(y - 25)$$

$$x = 0.7521y - 18.8049 + 30$$

$$x = 0.7521y + 11.19$$

when $y = 16$, $x = ?$

$$x = 0.7521y + 11.19$$

$$x = 0.7521 \times 16 + 11.19$$

$$x = 23.22 \approx 23$$

$$\boxed{x = 23}$$

Husbands age is 23 years corresponding to wife age of 16 years.

7) Find the co-efficient of corelation for the following data.

x	10	14	18	22	26	30
y	18	12	24	6	30	36

$\boxed{n=6}$

x	y	$X=x-\bar{x}$	$Y=y-\bar{y}$	X^2	Y^2	XY
10	18	−10	−3	100	9	30
14	12	−6	−9	36	81	54
18	24	−2	3	4	9	−6
22	6	2	−15	4	225	−30
26	30	6	9	36	81	54
30	36	10	15	100	225	150
120	126			280	630	252

$\bar{x} = \frac{\Sigma x}{n} = \frac{120}{6} = 20$

$\bar{y} = \frac{\Sigma y}{n} = \frac{126}{6} = 21$

$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}} = \frac{252}{\sqrt{280}\sqrt{630}}$

$\boxed{r = 0.6}$

OR

x	y	$z=x-y$	x^2	y^2	z^2
10	18	−8	100	324	64
14	12	2	196	144	4
18	24	−6	324	576	36
22	6	16	484	36	256
26	30	−4	676	900	16
30	36	−6	900	1296	36
120	126	−6	2680	3276	412

$\bar{x} = \frac{\Sigma x}{n} = \frac{120}{6} = 20$

$\bar{y} = \frac{\Sigma y}{n} = \frac{126}{6} = 21$

$\bar{z} = \frac{\Sigma z}{n} = \frac{-6}{6} = 1$

$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{2680}{6} - (20)^2$

$\sigma_x^2 = 46.66 = 6.83$

$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{3276}{6} - (21)^2$

$\sigma_y^2 = 105 = 10.24$

$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2$

$\sigma_z^2 = \frac{412}{6} - (1)^2 = 67.66$

$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x \sigma_y} = \frac{46.66 + 105 - 67.66}{2 \times 6.83 \times 10.24}$

$\boxed{r = 0.6}$

8) obtain the lines of regression and hence find the co-efficient of corelation for the data.

x	10	14	18	22	26	30
y	18	12	24	6	30	36

$\boxed{n=6}$

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	X^2	Y^2	XY
10	18	-10	-3	100	9	30
14	12	-6	-9	36	81	54
18	24	-2	3	4	9	-6
22	6	2	-15	4	225	-30
26	30	6	9	36	81	54
30	36	10	15	100	225	150
120	126			280	630	252

$$\bar{x} = \frac{\Sigma x}{n} = \frac{120}{6} = 20$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{126}{6} = 21$$

we have lines of regression

$$Y = \frac{\Sigma XY}{\Sigma X^2} X$$

$$y - \bar{y} = \frac{252}{280}(x - \bar{x})$$

$$y - 21 = 0.9(x - 20)$$

$$y - 21 = 0.9x - 18$$

$$y = 0.9x - 18 + 21$$

$$y = 0.9x + 3$$

$$X = \frac{\Sigma XY}{\Sigma Y^2} Y$$

$$x - \bar{x} = \frac{252}{630}(y - \bar{y})$$

$$x - 20 = 0.4(y - 21)$$

$$x = 0.4y - 8.4 + 20$$

$$\boxed{x = 0.4y + 11.6}$$

$y = 0.9x + 3$ and $x = 0.4y + 11.6$ are lines of regression

co-relation co-efficient is

$$r = \sqrt{(\text{co-eff of } x)(\text{co-eff of } y)} = \sqrt{0.9 \times 0.4}$$

$$\boxed{r = 0.6}$$

$r = +0.6$ sign of r is positive since both regression co-effici-ents are positive

---

Q) Find the corelation co-efficient by obtaining the lines of regression for the above data

x	17	18	19	19	20	20	21	22	21	23
y	12	16	14	11	15	19	22	15	16	20

$\boxed{n = 10}$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{200}{10} = 20$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{160}{10} = 16$$

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	X^2	Y^2	XY
17	12	-3	-4	9	16	12
18	16	-2	0	4	0	0
19	14	-1	-2	1	4	2
19	11	-1	-5	1	25	5
20	15	0	-1	0	1	0
20	19	0	3	0	9	0
21	22	1	6	1	36	6
22	15	2	-1	4	1	-2
21	16	1	0	1	0	0
23	20	3	4	9	16	12
200	160			30	108	35

$$Y = \frac{\Sigma XY}{\Sigma X^2} X$$

$$y-\bar{y} = \frac{35}{30}(x-\bar{x})$$

$$y-16 = 1.16(x-20)$$

$$y-16 = 1.16x - 23.33$$

$$y = 1.16x - 7.33$$

$$X = \frac{\Sigma XY}{\Sigma Y^2} Y$$

$$x-\bar{x} = \frac{35}{108}(y-\bar{y})$$

$$x-20 = 0.32(y-16)$$

$$x = 0.32y - 5.12 + 20$$

$$x = 0.32y + 14.88$$

$y = 1.16x - 7.33$ and $x = 0.32y + 14.88$ are lines of regression

co-relation co-efficient is

$$r = \sqrt{[\text{co-eff } x][\text{co-eff } y]}$$

$$r = \sqrt{(1.16)(0.32)}$$

$$\boxed{r = 0.609}$$

$r = +0.609$ sign r is positive since both regression co-efficients are positive

11)

10) Given

	x-series	y-series
Mean	18	100
SD	14	20

and $r = 0.8$ write down the equation of lines of regression and hence find the most probable value of y where $x = 70$

By data $\bar{x} = 18$ and $\bar{y} = 100$

$\sigma_x = 14$ and $\sigma_y = 20$

we have the equation of regression lines are given by

$$y-\bar{y} = r\frac{a_y}{a_x}(x-\bar{x})$$

$$y-100 = \frac{0.8\times 20}{14}(x-18)$$

$$y = 1.14x + 79.48$$

$$x-\bar{x} = r\frac{a_x}{a_y}(y-\bar{y})$$

$$x-18 = \frac{0.8\times 14}{20}(y-100)$$

$$x = 0.56y - 38$$

when $x=70$ we obtain from first equation

$$y = 1.14x + 79.48$$

$$y = 1.14\times 70 + 79.48$$

$$\boxed{y = 159.28}$$

ii) Show that if θ is the angle between the lines of regression then $\tan\theta = \frac{a_x a_y}{a_x^2 + a_y^2}\left[\frac{1-r^2}{r}\right]$

WKT if θ is acute the angle between the lines $y = m_1x_1 + c_1$ and $y = m_2x_2 + c_2$ is given by

$$\tan\theta = \frac{m_2 - m_1}{1+m_1m_2} \rightarrow *$$

we have lines of regression

$$y-\bar{y} = r\frac{a_y}{a_x}(x-\bar{x}) \rightarrow 1 \text{ and}$$

$$x-\bar{x} = r\frac{a_x}{a_y}(y-\bar{y})$$

$$y-\bar{y} = \frac{a_y}{ra_x}(x-\bar{x}) \rightarrow 2$$

Slopes of 1 and 2 are respectively given by

$$m_1 = r\frac{a_y}{a_x} \quad m_2 = \frac{a_y}{ra_x}$$

Substitute these in the formula for $\tan\theta$ * we have

$$\tan\theta = \frac{\frac{a_y}{ra_x} - \frac{ra_y}{a_x}}{1 + r\frac{a_y}{a_x}\frac{a_y}{ra_x}} = \frac{\frac{a_y}{a_x}\left[\frac{1}{r} - r\right]}{1 + a_y^2/a_x^2}$$

Scanned with CamScanner

$$\tan\theta = \frac{\frac{\sigma_y}{\cancel{\sigma_x}}\left[\frac{1-r^2}{r}\right]}{\frac{\sigma_x^2+\sigma_y^2}{\cancel{\sigma_x^2}}}$$

$$\tan\theta = \frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\left[\frac{1-r^2}{r}\right]$$

12) In a bivarite distribution $\sigma_x = \sigma_y$ and the angle between the regression lines is $\tan^{-1}(3)$ find the correlation co-efficient

If θ is angle between the lines of regression we have

$$\tan\theta = \frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\left[\frac{1-r^2}{r}\right] \rightarrow ①$$

By data $\theta = \tan^{-1}(3)$ or $\tan\theta = 3$ and $\sigma_x = \sigma_y$

eq^n ① ⇒ $3 = \frac{\sigma_x^2}{2\cancel{\sigma_x}}\left[\frac{1-r^2}{r}\right]$

$$3 = \frac{1-r^2}{2r}$$

$$6r = 1-r^2$$

$$r^2+6r-1=0$$

$$r = \frac{-6\pm\sqrt{40}}{2} = \frac{-6\pm 2\sqrt{10}}{2}$$

$$r = -3\pm\sqrt{10}$$

$r = -3 \pm \sqrt{10} = 0.1623$

$r = -3 - \sqrt{10} = -6.1623$

consider $\boxed{r = 0.1623}$ since r doesnot exceed unity

13) If $x = 4y+5$ and $y = kx+4$ are the regression lines of x on y and y on x respectively. prove that $0 \le k \le 1/4$

WKT co-efficient of corelation

$$r = \sqrt{[\text{co-eff of } y][\text{co-eff of } x]}$$

Scanned with CamScanner

$r = \sqrt{4(K)}$

$r = \sqrt{4K}$ squaring on both sides

$r^2 = 4K$

WKT $0 \le r^2 \le 1$

$0 \le 4K \le 1$

$0 \le K \le 1/4$

---

14) $8x - 10y + 66 = 0$ and $40x - 18y = 214$ are the two regression lines. Find the mean of x's, y's and the corelation co-efficient

Find σ_y if $\sigma_x = 3$

WKT regression line passes through $\bar{x}$ and $\bar{y}$

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

$40\bar{x} - 18\bar{y} - 214 = 0$ & $40\bar{x} - 18\bar{y} = 214$

By solving we get $\bar{x} = 13$ and $\bar{y} = 17$ we shall now rewrite the equation of the regression line to find the regression co-efficients

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

$10y = 8x + 66$ or $y = 0.8x + 6.6 \rightarrow ①$

$40x = 18y + 214$ $x = 0.45y + 5.35 \rightarrow ②$

From ① $r\frac{\sigma_y}{\sigma_x} = 0.8$; $r\frac{\sigma_x}{\sigma_y} = 0.45$

co-relation co-efficient $r = \sqrt{0.8 \times 0.45}$ $\boxed{r = 0.6}$

Thus $r = 0.6$ since both the regression co-efficients are positive also $\sigma_x = 3$ by data and we have

$r\frac{\sigma_y}{\sigma_x} = 0.8$ & $r\frac{\sigma_y}{3} = 0.8$ & $0.6\sigma_y = 2.4$ $\boxed{\sigma_y = 4}$

$r\frac{\sigma_x}{\sigma_y} = 0.45$ & $\frac{0.6 \times 3}{0.45} = \sigma_y$ $\boxed{\sigma_y = 4}$