6|4|2021

UNIT - II

Curve Fitting And Correction:

* Curve fitting:-

Suppose that a data is given in two variables x & y, the problem of finding an analytical expression $y = f(x)$ which fits the given data is called curve fitting.

→ Let $(x_1, y_1)(x_2, y_2) \ldots (x_n, y_n)$ be the given set of values where $y_1, y_2 \ldots y_n$ are called observed values.

→ At $x = x_1$ then $y = f(x_1)$,. $x = x_n$ then $y = f(x_n)$ where $f(x_1), f(x_2) \ldots f(x_n)$ are called expected values

→ The difference between observed and expected values are called errors (or) residuals. It is defined as

$$E_i = y_i - f(x_i) \quad \text{for } i = 1 \text{ to } n.$$

→ The sum of squares of residuals is minimum is called method of least squares (or) least squares method (or) best fitting curve

i.e $E = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} [y_i - f(x_i)]^2$ is minimum

* Least Squares Approximation:-

→ Linear least square Approximation:-

• straight line :-

→ Let $y = a + bx \rightarrow ①$ be the straight line. This straight line passes through the points $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$

i.e. (x_i, y_i) for $i = 1, 2, 3 \ldots n$

eq ① ⇒ $y_i = a + bx_i$ for $i = 1, 2, 3 \ldots n$

→ The difference between observed & expected values is defined as $E_i = y_i - (a + bx_i)$ for $i = 1$ to n

→ The sum of squares of residuals is $E = \sum_{i=1}^{n} E_i^2$

i.e. $E = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$

→ By method of least squares E is mininum.

for minimum of E $\frac{\partial E}{\partial a} = 0$ & $\frac{\partial E}{\partial b} = 0$

$-2\Sigma[y_i - (a + bx_i)] = 0$ & $-2x_i\Sigma[y_i - (a + bx_i)] = 0$

$\Sigma y_i = na + b\Sigma x_i$ & $\Sigma x_i y_i = a\Sigma x_i + b\Sigma x_i^2$

∴ $\Sigma y = na + b\Sigma x$ & $\Sigma xy = a\Sigma x + b\Sigma x^2 \rightarrow ②$

∴ eq ② is called normal equation of eq ①.

→ Solving equation ②, substitute a, b values in eq ① which is the best fitting curve to the given data.

. Parabola:-

Let $y = a + bx + cx^2 \rightarrow (1)$ be the given curve. This equation passing through the points $(x_1, y_1)(x_2, y_2) \ldots (x_n, y_n)$

i.e (x_i, y_i) for $i = 1$ to n

eq (1) $\Rightarrow y_i = a + bx_i + cx_i^2$ for $i = 1$ to n

The difference between observed & expected values is called residuals. It is defined as

$$E_i = y_i - (a + bx_i + cx_i^2) \text{ for } i = 1, 2, \ldots n$$

→ By method of least squares the sum of residuals is minimum

$$E = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} [y_i - (a + bx_i + cx_i^2)]^2 \quad \text{for } i = 1, 2, \ldots n$$

→ for minimum of E $\frac{\partial E}{\partial a} = 0$ & $\frac{\partial E}{\partial b} = 0$ & $\frac{\partial E}{\partial c} = 0$

i.e $\Sigma y = na + b\Sigma x_i + c\Sigma x_i^2$

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$ — Normal equations of eq (1).

1/4/2021

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

· Degree 3 polynomial:-

Let $y = a + bx + cx^2 + dx^3 \rightarrow (1)$ be the given curve then Normal equations of eq (1) is

$$\left.\begin{aligned} \text{i.e } \Sigma y &= na + b\Sigma x_i + c\Sigma x_i^2 + d\Sigma x_i^3 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 + d\Sigma x^4 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 + d\Sigma x^5 \\ \Sigma x^3 y &= a\Sigma x^3 + b\Sigma x^4 + c\Sigma x^5 + d\Sigma x^6 \end{aligned}\right\} (2)$$

Solving eq (2) and substitute a, b, c, d in eq (1) which is the best fitting curve to the given data

→ Non linear least square Approximation:

· Exponential curve:-

We have two exponential curves. They are $y = ae^{bx}$, $y = ab^x$

i. $y = e^{bx} a$

Let $y = ae^{bx}$ be the given curve.

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

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

$$Y = A + bx \rightarrow (2)$$

eq (2) is a straight line in x, Y

Normal equations of equation (2)

$$\left.\begin{aligned} \Sigma Y &= nA + b\Sigma x \\ \Sigma xY &= A\Sigma x + b\Sigma x^2 \end{aligned}\right\} (3)$$

Solving eq (3) we get A, b $a = e^A$

ii. $y = ab^x$

Let $y = ab^x$ be the given curve

$$\log_{10} y = \log_{10} a + x\log_{10} b$$

$$Y = A + xB \rightarrow (2)$$

eq (2) is a st. line in x, Y

Normal equations of eq (2) is

$$\left.\begin{aligned} \Sigma Y &= nA + B\Sigma x \\ \Sigma xY &= A\Sigma x + B\Sigma x^2 \end{aligned}\right\} (3)$$

Solving eq (3) we get A, B, $a = 10^A$, $b = 10^B$

Scanned with CamScanner

· Power curve :-

Let $y = ax^b$ be the given power curve.

$$\log_{10} y = \log_{10} a + b \log_{10} x$$

$$Y = A + bX \rightarrow (2)$$

eq(2) is st line in X, Y.

Normal equations of eq (2) is

$$\left.\begin{aligned} \Sigma Y &= nA + b\Sigma X \\ \Sigma XY &= A\Sigma X + b\Sigma X^2 \end{aligned}\right\} (3)$$

Solve eq (3) we get A, b

$a = 10^A$

Problems :-

* By method of least squares find the straight lines that fits the following data

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

Sol:- Given

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

Let straight line equation is $y = a + bx \rightarrow (1)$

normal equations of eq(1) $\left.\begin{aligned} \Sigma y &= na + b\Sigma x \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} (2)$ where $n = 5$

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
$\Sigma = 15$	204	748	55

$(2) \Rightarrow$

$$\begin{aligned} 15a + 45b &= \cancel{204}\ 612 \\ 15a + 55b &= 748 \\ \hline -10b &= -136 \\ b &= 13.6 \end{aligned}$$

$5a + 15(13.6) = 204$

$5a = 204 - 204$

$a = 0$

eq (1) $\Rightarrow$ $y = a + bx$

$\boxed{y = 13.6x} \rightarrow$ required best fitting curve of eq(1).

* Fit a straight line to the following data

x	6	7	7	8	8	8	9	9	10
y	5	5	4	5	4	3	4	3	3

Sol:- Let $y = a + bx \rightarrow (1)$ be the straight line.

normal equations of eq(1) $\left.\begin{aligned} \Sigma y &= na + b\Sigma x \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} (2)$ where $n = 9$

x	6	7	7	8	8	8	9	9	10
y	5	5	4	5	4	3	4	3	3

x	y	xy	x^2
6	5	30	36
7	5	35	49
7	4	28	49
8	5	40	64
8	4	32	64
8	3	24	64
9	4	36	81
9	3	27	81
10	3	30	100
Σ = 72	36	282	588

eq(2) ⟹ $9a + b\,72 = 36$

$\times 8$: $72a + b\,588 = 282$

$72a + 576b = 288$

$12b = -6$

$\boxed{b = -1/2}$

$a + 8b = 4$

$a - 4 = 4$

$\boxed{a = 8}$

$\therefore$ eq(1) ⟹ $y = 8 - \frac{1}{2}x$

$2y = 16 - x$ is the best fitting curve of eq(1).

9/4/2021

* Fit a straight line to the following data

x	0	1	2	3	4
y	1	1.8	3.3	4.5	6.3

Sol: Let $y = a + bx \rightarrow (1)$ where $n = 5$

Normal eq. of (1) is $\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$ } eq(2)

x	y	xy	x^2
0	1	0	0
1	1.8	1.8	1
2	3.3	6.6	4
3	4.5	13.5	9
4	6.3	25.2	16
10	16.9	47.1	30

$16.9 = 5a + b(10)$ $\quad$ $47.1 = a(10) + b(30)$

$33.8 = 10a + 20b$

$47.1 = 10a + 30b$

$-13.3 = -10b$

$b = 1.33$

$16.9 = 5a + 13.3$

$5a = 16.9 - 13.3$

$5a = 3.6$

$a = \frac{3.6}{5} = 0.72$.

7.2
36
80

Substitute a, b values in eq(1)

$y = 0.72 + x(1.33)$ → This is required best fit curve.

Scanned with CamScanner

* A chemical company wishing to study the effect of extraction time on the efficiency of an operation, obtained the data in the following table from a fitting curve st line

time (x min)	27	45	41	19	3	39	19	49	15	31
efficiency y	57	64	80	46	62	72	52	77	57	68

Sol: Let $y = a + bx \to ①$ where $n = 10$

Normal eq of ① is $\Sigma y = na + b\Sigma x$

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

x	y	xy	x^2
27	57	1539	729
45	64	2880	2025
41	80	3280	1681
19	46	874	361
3	62	186	9
39	72	2808	1521
19	52	988	361
49	77	3773	2401
15	57	855	225
31	68	2108	961
288	635	19291	10274

$635 = 10a + 288b$

$19291 = 288a + 10274b$

$182880 = 2880a + 82944b$

$192910 = 2880a + 102740b$

$10030 = 19796b$

$b = \frac{10030}{19796} = 0.50$

$a = 48.925$

$\therefore$ substitute a, b values in eq ①.

$y = 48.925 + 0.50(x) \to$ required best fitting curve

* The temperature T & length L of a heated rod are given below if $L = a_0 + a_1 T$, find the values of a_0, a_1 using least squares method.

T	20	30	40	50	60	70
L	800.3	800.4	800.6	800.7	800.9	801

Sol: Given $L = a_0 + a_1 T \to ①$ where $n = 6$

Normal eq. of ① is $\Sigma L = na_0 + a_1 \Sigma T$

$\Sigma LT = a_0 \Sigma T + a_1 \Sigma T^2$

T	L	TL	T^2
20	800.3	16006	400
30	800.4	24012	900
40	800.6	32024	1600
50	800.7	40035	2500
60	800.9	48054	3600
70	801	56070	4900
270	4803.9	216201	13900

$4803.9 = 6a_0 + a_1 270$

$216201 = a_0 270 + a_1 13900$

by solving above equation

$a_1 = 0.0145 \quad a_0 = 799.9$

$\therefore$ eq① $\Rightarrow 799.9 + 0.0145T = L \to$ is the required best fitting curve

* Fitting a second degree polynomial to the following data by method of least squares.

x	10	12	15	23	20
y	14	17	23	25	21

Sol:- Let $y = a + bx + cx^2 \rightarrow (1)$ be second degree polynomial

The normal equations of eq(1) is

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\} (2)$$

x	y	xy	x^2y	x^2	x^3	x^4
10	14	140	1400	100	1000	10000
12	17	204	2448	144	1728	20736
15	23	345	5175	225	3375	50625
23	25	575	13225	529	12167	279841
20	21	420	8400	400	8000	160000

$\Sigma x = 80$ $\quad \Sigma xy = 1684$ $\quad \Sigma x^4 = 521202.$

$\Sigma x^2 = 1398$ $\quad \Sigma y = 100$

$\Sigma x^3 = 26270$ $\quad \Sigma x^2 y = 30648$

eq (2) $\Rightarrow$ $100 = 5(a) + (b)80 + (c)1398$

eq: $1684 = 80(a) + 1398(b) + 26270(c)$

$30648 = 1398(a) + 26270(b) + 521202(c)$

$a = -8.727$

$b = 3.0099$

$c = -0.069$

substitute a, b, c values in eq (1)

$\therefore y = -8.727 + (x)3.0099 + x^2(-0.069) \rightarrow$ best fitting curve of eq(1).

* Fit a parabola of the form $y = a + bx + cx^2$ to the following data by method of least squares

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

Sol: Let $y = a + bx + cx^2 \rightarrow (1)$ be second degree polynomial

The normal equations of eq(1) is

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\} \text{eq}(2)$$

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	51.40	205.60
6	31.47	36	216	1296	188.82 ~~865.60~~	1132.92
8	57.38	64	512	4096	459.04	3672.32
10	91.29	100	1000	10000	912.9	9129

$\Sigma x = 30$

$\Sigma y = 196.06$

$\Sigma x^2 = 220$

$\Sigma x^3 = 1800$

$\Sigma x^4 = 15664$

$\Sigma x^2 y = 1618.3$

$\Sigma x^2 y = 14152.12$

Scanned with CamScanner

$5a + 30b + 220c = 196.06$

$30a + 220b + 1800c = 1618.3$

$220a + 1800b + 15664c = 14152.12$

$\therefore a = 0.696, \quad b = -0.855, \quad c = 0.991$

$\therefore$ substitute a, b, c values in eq①

$y = 0.696 - (0.855)x + (0.991)x^2 \rightarrow$ best fitting curve of given parabola

* Fit a polynomial of second degree to the following data

x	0	1	2
y	1	6	17

Sol: Let $y = a + bx + cx^2 \rightarrow$ ① be second degree polynomial

The normal equations of eq① is

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\} \text{eq②.}$$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	6	1	1	1	6	6
2	17	4	8	16	34	68
3	24	5	9	17	40	74

eq① ⇒ $24 = 3a + b3 + 5c$ → $16 = 2b + 4c$

$40 = 3a + 5b + 9c$ $\quad a = 1;\ b = 2;\ 3$

$74 = 5a + 9b + 17c$

$y = 1 + 2x + 3x^2 \rightarrow$ best fitting curve of given see eq①.

* Fit a second degree parabola to the following data

x	0	1	2	3	4
y	1	1.8	1.3	2.5	6.3

sol:- Let $y = a + bx + cx^2 \rightarrow$ ① be second degree polynomial $n = 5$

The normal equation of eq ① is

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\} ②$$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	1.8	1	1	1	1.8	1.8
2	1.3	4	8	16	2.6	5.2
3	2.5	9	27	81	7.5	22.5
4	6.3	16	64	256	25.2	100.8

$\Sigma x = 10$ $\quad \Sigma y = 12.9$

$\Sigma x^2 = 30$ $\quad \Sigma xy = 37.1$

$\Sigma x^3 = 100$ $\quad \Sigma x^2 y = 130.3$

$\Sigma x^4 = 354$

eq → ② ⇒ $12.9 = 5a + 10b + 30c$

$37.1 = 10a + 30b + 100c$

$130.3 = 30a + 100b + 354c$

by solving eq ② we get

$a = 1.42 \quad c = 0.55$

$b = -1.07$

substitute a, b, c values in eq①

$y = 1.42 + (-1.07)x + (0.55)x^2$

* Fit a curve $y = a+bx+cx^2$ to the following data

x	1	2	3	4	5	6	7
y	2.3	5.2	9.7	16.5	29.4	35.5	54.4

Sol:- Given curve is $y = a+bx+cx^2 \rightarrow$ ① $n = 7$

The normal equations of eq ① are

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\} \text{eq ②}$$

x	y	x^2	x^3	x^4	xy	x^2y
1	2.3	1	1	1	2.3	2.3
2	5.2	4	8	16	10.4	20.8
3	9.7	9	27	81	29.1	87.3
4	16.5	16	64	256	66.0	264.0
5	29.4	25	125	625	147.0	735.0
6	35.5	36	216	1296	213.0	1278.0
7	54.4	49	343	2401	380.8	2665.6

$\Sigma x = 28$ $\qquad$ $\Sigma y = 153$

$\Sigma x^2 = 140$ $\qquad$ $\Sigma xy = 848.6$

$\Sigma x^3 = 784$ $\qquad$ $\Sigma x^2 y = 5053.$

$\Sigma x^4 = 4676$

$$\text{eq②} \Rightarrow \left.\begin{aligned} 153 &= 7a + 28b + 140c \\ 848.6 &= 28a + 140b + 784c \\ 5053 &= 140a + 784b + 4676c \end{aligned}\right\} \text{eq③}$$

By solving eq ③

$a = 2.37$ $b = -1.09$

$c = 1.192$

substitute given a, b, c values in equation ①

$$y = 2.37 + (-1.09)x + 1.192(x^2)$$

* Fit a parabola to the following data

x	10	15	20	25	30	35
y	35.3	32.4	29.2	26.1	23.2	20.5

Sol:- Given Let $y = a+bx+cx^2 \rightarrow$ ① be the equation of parabola

The normal equations of eq① are $n = 6$

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\} \text{eq ②}$$

Scanned with CamScanner

x	y	x^2	x^3	x^4	xy	x^2y
10	35.3	100	1000	10000	353	3530
15	32.4	225	3375	30625	486	7290
20	29.2	400	8000	160000	584	11680
25	26.1	625	15625	390625	652.5	16312.5
30	23.2	900	27000	810000	696.0	20880.0
35	20.5	1225	42875	1500625	717.5	25112.5

$\Sigma x = 135$ $\qquad$ $\Sigma y = 166.7$

$\Sigma x^2 = 3475$ $\qquad$ $\Sigma xy = 3489$

$\Sigma x^3 = 97,875$ $\qquad$ $\Sigma x^2 y = 84,805$

$\Sigma x^4 = 2921875$

$$eq(1) \Rightarrow 166.7 = 6(a) + 135(b) + 3475(c)$$
$$3489 = 135(a) + 3475(b) + 97875(c)$$
$$84805 = 3475(a) + 97875(b) + 2921875$$

By solving eq(3)

$a = 35.23 \quad b = 0.002$

$c = -0.012$

subditute a, b, c values in eq(1)

$$y = 35.23 + (0.002)x + (-0.012)x^2$$

* Fit a straight line $y = a_0 + a_1 x$ by method of least squares following data.

x	0	2	5	7
y	-1	5	12	20

sol Given the straight line is $y = a_0 + a_1 x \rightarrow (1)$ $\quad n = 4$

The normal equations of eq(1) is

$$\left.\begin{aligned} \Sigma y &= n a_0 + a_1 \Sigma x \\ \Sigma xy &= a_0 \Sigma x + a_1 \Sigma x^2 \end{aligned}\right\} eq(1)$$

x	y	x^2	xy
0	-1	0	0
2	5	4	10
5	12	25	60
7	20	49	140

$\Sigma x = 14 \qquad \Sigma y = 36$

$\Sigma x^2 = 78 \qquad \Sigma xy = 210$

$$eq(1) \Rightarrow 36 = 4a_0 + 14a_1$$
$$210 = 14a_0 + 78a_1$$

By solving $a_0 = -1.137 \quad a_1 = 2.89$

substitute a_0, a_1 values in eq(1)

$$y = -1.137 + 2.89(x)$$

* Fit a straight line $y = a + bx$ by method of least square to following

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Sol: Given the straight line eq $y = a + bx \rightarrow (1)$ $n = 6$

normal equations of eq(1) is $\Sigma y = na + b\Sigma x$

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

x	y	xy	x^2
0	12	0	0
5	15	75	25
10	17	170	100
15	22	330	225
20	24	480	400
25	30	750	625

$\Sigma x = 75$ $\Sigma y = 120$

$\Sigma x^2 = 1375$ $\Sigma xy = 1805$

eq(1) $\Rightarrow 120 = 6a + b75$

$1805 = 75a + 1375b$

$a = 11.28$ $b = 0.697$

substitute a, b values in eq (1)

$$y = 11.28 + (0.697)x$$

13/9/2021

* Fit the curve $y = ae^{bx}$ to the following data by method of least squares

x	1	5	7	9	12
y	10	15	12	15	21

Sol: Given the curve is $y = ae^{bx} \rightarrow (1)$

The eq form of eq(1) is $Y = A + bx \rightarrow (1)$ where $Y = \log_e y$, $A = \log_e$

The normal equations of eq(1) is

$\left.\begin{array}{l}\Sigma Y = nA + b\Sigma x \\ \Sigma xY = A\Sigma x + b\Sigma x^2\end{array}\right\}$ eq(2) $n = 5$

x	y	Y	xY	x^2
1	10	2.3025	2.3025	1
5	15	2.7080	13.54	25
7	12	2.4849	17.3943	49
9	15	2.7080	24.372	81
12	21	3.0445	36.534	144
$\Sigma = 34$		13.2479	94.1428	300

The eq(2) $\Rightarrow 13.2479 = 5A + b(34)$

$94.1428 = 34A + 300b$

$A = ~~2.09737~~$ 2.2482

$b = ~~0.22996~~$ 0.058

$a = e^{A}$

$= 9.4682\ \ 9.4682\mathbf{7}$

substitute a, b values in eq (1)

$$y = 9.4682\mathbf{7}\, e^{(0.055)x}$$

* Using method of least squares find the constants a, b such that $y = ae^{bx}$

x	0	0.5	1	1.5	2	2.5
y	0.1	0.45	2.15	9.15	40.35	180.75

Sol: Given curve is $y = ae^{bx} \rightarrow (1)$

The eq. form of eq (1) is $Y = A + bx \rightarrow (2)$ where $Y = \log_e y$, $A = \log_e a$

The normal equations of eq (2) is

$$\left.\begin{aligned} \Sigma Y &= nA + b\Sigma x \\ \Sigma xY &= A\Sigma x + b\Sigma x^2 \end{aligned}\right\} \text{eq (3)}$$

$n = 6$

x	y	Y	xY	x^2
0	0.1	-2.3025	0	0
0.5	0.45	-0.7985	-0.3992	0.25
1	2.15	0.7654	0.7654	1
1.5	9.15	2.2137	3.3205	2.25
2	40.35	3.6975	7.395	4
2.5	180.75	5.19711	12.9927	6.25
$\Sigma = 7.5$	232.95	8.7727	24.0744	13.75

$8.7727 = 6A + b\, 7.5$ $\qquad A = -2.2831 \qquad a = 0.1019$

$24.0744 = 7.5A + b\, 13.75 \qquad b = 2.9962$

substitute a, b in eq (1)

$$y = 0.1019\, e^{(2.9962)x}$$

* Fit the curve $y = ae^{bx}$ to the following data by method of least squares

x	0	1	2	3	4	5	6	7	8
y	20	30	52	77	135	211	326	515	1052

Sol: Given curve is $y = ae^{bx} \rightarrow (1)$

The eq. form eq (1) is $Y = A + bx \rightarrow (2)$

The normal equations of eq (2) is

$$\Sigma Y = nA + b\Sigma x$$
$$\Sigma xY = A\Sigma x + b\Sigma x^2 \qquad n = 9$$

x	y	Y	xY	x^2
0	20	2.9957	0	0
1	30	3.4011	3.4011	1
2	52	3.9512	7.9024	4
3	77	4.3438	13.0314	9
4	135	4.9052	19.6208	16
5	211	5.3518	26.759	25
6	326	5.7868	34.7208	36
7	515	6.2441	43.7087	49
8	1052	6.9584	55.6672	64
$\Sigma =$ 36	2418	43.9381	204.8114	204

$$43.9381 = 9A + 36b \qquad A = 2.9447 \Rightarrow a = 19.0049$$

$$204.8114 = A\,36 + 204\,b \qquad b = 0.4843$$

substitute a, b values in eq ①

$$y = 19.0049\, e^{(0.4843)x}$$

* Fit the curve $y = ae^{bx}$ to the following data by method of least squares

x	77	100	185	239	285
y	2.4	3.4	7	11.1	19.6

Sol: Given the curve is $y = ae^{bx} \rightarrow ①$

The equation form of ① is $Y = A + bx \rightarrow ②$ where $Y = \log_e y$, $A = \log_e a$

The normal equations of eq② is

$$\left.\begin{aligned} \Sigma Y &= nA + b\Sigma x \\ \Sigma xY &= A\Sigma x + b\Sigma x^2 \end{aligned}\right\} eq③ \qquad n = 5$$

x	y	Y	xY	x^2
77	2.4	0.8754	67.4058	5929
100	3.4	1.2237	122.37	10000
185	7	1.9459	359.9915	34225
239	11.1	2.4069	575.2491	57121
285	19.6	2.9755	848.0175	81225
$\Sigma =$ 886		9.4274	1973.0339	188500

eq③ $\Rightarrow$ $9.4274 = 5A + 886(b)$ $\qquad A = 0.1838 \Rightarrow a = e^A$

$1973.0339 = 886(A) + 188500(b)$ $\qquad b = 0.0096$ $\qquad \Rightarrow 1.2017$

substitute a, b values in eq ①

$$y = 1.2017\, e^{(0.0096)x}$$

Fit the curve $y = ae^{bx}$ to the following data by method of least square

x	0	2	4
y	5.1	10	13.1

Sol: Given the curve $y = ae^{bx} \rightarrow (1)$

The equation form of (1) is $Y = A + bx \rightarrow (2)$ where $Y = \log_e y$, $A = \log_e a$

The normal equations of eq(2) are

$$\left.\begin{aligned} \Sigma Y &= nA + b\Sigma x \\ \Sigma xY &= A\Sigma x + b\Sigma x^2 \end{aligned}\right\} eq(3) \quad n = 3$$

x	y	Y	xY	x^2
0	5.1	1.6292	0	0
2	10	2.3025	4.605	4
4	13.1	2.5726	10.2904	16
$\Sigma = 6$		6.5043	14.8954	20

eq (3) $\Rightarrow$ $6.5043 = 3A + b(6)$

$14.8954 = 6A + b(20)$

$A = 1.6964 \Rightarrow a = e^A = 5.4542$

$b = 0.2358$

substitute a, b values in (1)

$$y = 5.4542\,(e^{(0.2551)x})$$

Fit the curve $y = ae^{bx}$ to the following data by method of least squares

x	0	1	2	3
y	1.05	2.10	3.85	8.3

Sol: Given the curve $y = e^{bx} \cdot a \rightarrow (1)$

The equation form of (1) is $Y = A + bx \rightarrow (2)$

The normal equations of eq(1) is

$$\left.\begin{aligned} \Sigma Y &= nA + b\Sigma x \\ \Sigma xY &= A\Sigma x + b\Sigma x^2 \end{aligned}\right\} eq(3) \quad n = 4$$

x	y	Y	xY	x^2
0	1.05	0.0487	0	0
1	2.10	0.7419	0.7419	1
2	3.85	1.3480	2.696	4
3	8.3	2.1162	6.3486	9
$\Sigma = 6$		4.2548	9.7865	14

eq (3) $\Rightarrow$ $4.2545 = 4A + 6b$

$9.7865 = 6A + 14b$

$A = 0.04241$; $a = e^A$

$b = 0.68016$; $a = 1.0433$

Substitute a, b values in (1)

$$y = 1.0433 \times e^{(0.6806)x}$$

* Fit the curve parabola to the following data
$f(-1)=-2,\ f(0)=1,\ f(1)=2,\ f(1)=4.$

So: Let $y=a+bx+cx^2$ is the parabola eq. $\to (1)$

The normal equations of eq(1) is

$$\left.\begin{aligned}\Sigma y &= na+b\Sigma x+c\Sigma x^2\\ \Sigma xy &= a\Sigma x+b\Sigma x^2+c\Sigma x^3\\ \Sigma x^2y &= a\Sigma x^2+b\Sigma x^3+c\Sigma x^4\end{aligned}\right\}\text{eq}(2) \quad n=4.$$

x	y	xy	x^2y	x^2	x^3	x^4
-1	-2	2	-2	1	-1	1
0	1	0	0	0	0	0
1	2	2	2	1	1	1
2	4	8	16	4	8	16
2	5	12	16	6	8	18.

eq(2) $\Rightarrow$ $5=4a+b2+c(6)$
$12=2a+6b+8c$
$16=6a+8b+18c$

$a=0.55$
$b=2.15$
$c=-0.25$

Substitute a,b,c values in eq(1)

$$y=0.55+2.15(x)-(0.25)(x^2)$$

* Fit the parabola to the following data & passing through the points $(-1,2)\ (0,1)\ (1,4)$

So Let the eq. of parabola is $y=a+bx+cx^2 \to (1)$

The normal equations of eq(1) is

$$\left.\begin{aligned}\Sigma y &= na+b\Sigma x+c\Sigma x^2\\ \Sigma xy &= a\Sigma x+b\Sigma x^2+c\Sigma x^3\\ \Sigma x^2y &= a\Sigma x^2+b\Sigma x^3+c\Sigma x^4\end{aligned}\right\}\text{eq}(3) \quad n=3$$

x	y	xy	x^2y	x^2	x^3	x^4
-1	2	-2	+2	1	-1	1
0	1	0	0	0	0	0
1	4	4	4	1	1	1
0	7	2	6	2	0	2

eq(3) $\Rightarrow$ $7=3a+2c$
$2=2b \Rightarrow \boxed{b=1} \Rightarrow \boxed{a=1}\quad \boxed{c=2}$
$6=2a+2c$

substitute a,b,c in eq(1)

$$y=1+x+2x^2$$

Scanned with CamScanner

* Fit a straight line for the following data by method of least squares

x	-5	-3	-1	0	1	2	4
y	+0.4	-0.1	-0.2	-0.3	-0.3	0.1	0.4

sol:- Let the eq. of st. line is $y = a + bx \rightarrow ①$

The normal equations of eq① are

$$\Sigma y = na + b\Sigma x \quad \text{eq②}$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad n = 7$$

x	y	xy	x^2
-5	-0.4	+2	25
-3	-0.1	0.3	9
-1	-0.2	+0.2	1
0	-0.3	0	0
1	-0.3	-0.3	1
2	0.1	0.2	4
4	0.4	1.6	16
-2	-0.8	4	56

eq② ⇒ $-0.1 = 7a - 2b$

$4 = 2a + 56b$

$a = 2 - 28b$

$0 = 7(2 - 28b) - 2b$

$0 = -196b - 2b + 14$

$+198b = +14$

$b = \frac{14}{198} = 0.0707$

$a = 2 - 28b$

$= -0.0202$

substitute a, b in eq①

$y = -0.0202 + (0.0707)x$

15/4/2021

* Find a, b so that $y = ab^x$ to the following data

x	1	2	3	4	5	6
y	151	100	61	50	20	8

Sol:- Given curve $y = ab^x \rightarrow ①$

The eq form of ① is $Y = A + xB \rightarrow ②$

The normal equation of eq② is

$$\Sigma Y = nA + B\Sigma x \quad \text{eq③} \quad n = 6$$
$$\Sigma xY = A\Sigma x + B\Sigma x^2$$

x	y	xY	x^2	Y
1	151	2.1789	1	2.1789
2	100	4	4	2
3	61	5.3559	9	1.7853
4	50	6.7956	16	1.6989
5	20	6.505	25	1.3010
6	8	5.418	36	0.9030
21		30.2534	91	9.8671

Scanned with CamScanner

$eq (3) \Rightarrow 9.5671 = 6A + B\,21$

$30.2534 = 21A + B\,91$

$A = 2.5008 \quad , \quad B = -0.2446$

$a = 10^A \qquad\qquad b = 0.5693$

$= 316.8108$ substitute a, b in eq (1)

$y = 316.8108\,(0.5693)^x$

* Fit a curve of the form $y = ab^x$ to the following data

x	0.2	0.3	0.4	0.5	0.6	0.7
y	3.16	2.31	1.75	1.34	1.00	0.74

So: Given curve $y = ab^x \to (1)$

The eq. form of (1) is $Y = A + Bx \to (2)$

Normal equations of (2) is

$$\left.\begin{aligned} \Sigma Y &= nA + B\Sigma x \\ \Sigma xY &= A\Sigma x + B\Sigma x^2 \end{aligned}\right\} eq (3) \quad n = 6$$

x	y	Y	xY	x^2
0.2	3.16	0.4996	0.0999	0.04
0.3	2.38	0.3765	0.1129	0.09
0.4	1.75	0.2430	0.0972	0.16
0.5	1.34	0.1271	0.6355	0.25
0.6	1.00	0	0	0.36
0.7	0.74	-0.1307	-0.0914	0.49
0.27		1.1155	0.8541	1.39

$1.1155 = 6A + 0.27\,Bx$

$0.8541 = 0.27A + 1.39B$

$A = 0.1596 \Rightarrow a = 10^A = 1.4441$

$B = 0.5834 \Rightarrow b = 10^B = 3.8317$

substituted a, b in (1)

$y = 1.4441\,(3.8317)^x$

Fit a curve of the form $y = ax^b$ to the following data

x	1	2	3	4	5	6
y	2.98	4.26	5.21	6.10	6.80	7.50

Sol: Given curve $y = ax^b \rightarrow (1)$

The eq. form of (1) is $Y = A + b\Sigma X \rightarrow (2)$

Normal equation of (2) is

$$\left.\begin{array}{l}\Sigma Y = nA + b\Sigma X \\ \Sigma XY = A\Sigma X + b\Sigma X^2\end{array}\right\} eq (3) \quad n = 6$$

x	y	X	Y	XY	X^2
1	2.98	0	0.4742	0	0
2	4.26	0.3010	0.6294	0.1894	0.0906
3	5.21	0.4771	0.7168	0.3419	0.2276
4	6.10	0.6020	0.7853	0.4727	0.3624
5	6.80	0.6989	0.8325	0.5818	0.4884
6	7.50	0.7781	0.8750	0.6808	0.6054
		2.8571	4.3132	2.2666	1.7744

$eq(3) \Rightarrow 4.3132 = 6A + b(2.8571)$

$2.2666 = (2.8571)A + b(1.7744)$

$A = 0.4741, \quad b = 0.5139$

$a = 10^A$

$= 2.9792$

Substitute a, b in eq (1)

$y = (2.9792)(x)^{0.5139}$

Fit a curve of the form $V = at^b$ to the following data experimental data

V ft/mit	350	400	500	600
t (min)	61	26	7	2.6

Sol: Given curve $V = at^b \rightarrow (1)$

The eq form of (1) is $Y = \qquad \rightarrow (2)$

Normal equation of eq(2)

$$\left.\begin{array}{l}\Sigma T = nA + b\Sigma v \\ \Sigma vT = A\Sigma v + b\Sigma v^2\end{array}\right\} eq (3) \quad n = 4$$

$V(y)$	$t(x)$	$y(Y)$	$T(X)$	XY	X^2
350	61	2.5440	1.7853	4.5418	3.1872
400	26	2.6020	1.4149	3.6815	2.0019
500	7	2.6989	0.8450	2.2805	0.7140
600	2.6	2.7781	0.4149	1.1526	0.1721
		10.6230	4.4601	11.6564	6.0752

$$10.6230 = 4A + 4.4601(b)$$

$$11.6564 = 4.4601(A) + 6.0752(b)$$

$$A = 2.846v \;;\; b = -0.1710$$

$$a = 702.1016$$

Scanned with CamScanner

Correlation:

Def:-

Correlation is a statistical technique, which is used for analysing the behaviour of 2 or more variables.

correlation coefficients:-

Correlation coefficient is denoted by r, where r value lies between -1 to $+1$

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

where $X = x - \bar{x}$ ∴ $\bar{x}$ is mean of x

$Y = y - \bar{y}$ ∴ $\bar{y}$ is mean of y

Problems:-

1. Find coefficient of correlation to the following data

wage (x):	100	101	102	102	100	99	97	98	96	95
cost of living (y):	98	99	99	97	95	92	95	94	90	91

Sol:

x	y	X	Y	XY	X^2	Y^2
100	98	1	3	3	1	9
101	99	2	4	8	4	16
102	99	3	4	12	9	16
102	97	3	2	6	9	4
100	95	1	0	0	1	0
99	92	0	-3	0	0	9
97	95	-2	0	0	4	0
98	94	-1	-1	1	1	1
96	90	-3	-5	15	9	25
95	91	-4	-4	16	16	16

$\bar{x} = 99$ $\quad \bar{y} = 95$ $\quad \Sigma XY = 61$ $\quad \Sigma Y^2 = 96$

$X = x - \bar{x}$ $\quad Y = y - \bar{y}$ $\quad \Sigma X^2 = 54$

∴ we know that $r = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}} = \frac{61}{\sqrt{54}\sqrt{96}}$

Scanned with CamScanner

$$= \frac{61}{(7.3485)(9.798)}$$

$$= 0.8472$$

$$\boxed{\therefore r = 0.8472}$$

2 Find coefficient of correlation by using the given data

Height: (x) 57 59 62 63 64 65 55 58 57

weight: (y) 113 117 126 126 130 129 111 116 112

Sol: $\bar{x} = 60$ $\bar{y} = 120$

$X = x - \bar{x}$ $Y = y - \bar{y}$

x	y	X	Y	XY	Y^2	X^2
57	113	-3	-7	21	49	9
59	117	-1	-3	3	9	1
62	126	2	6	12	36	4
63	126	3	6	18	36	9
64	130	4	10	40	100	16
65	129	5	9	45	81	25
55	111	-5	-9	45	81	25
58	116	-2	-4	8	16	4
57	112	-3	-8	24	64	9

$\Sigma XY = 216$ $\Sigma X^2 = 102$ $\Sigma Y^2 = 472$

w.k.T

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

$$= \frac{216}{\sqrt{102}\sqrt{472}}$$

$$= \frac{216}{10.0995 \times 21.7250}$$

$$\boxed{r = 0.9844}$$

Scanned with CamScanner

eviations are taken from a assumed mean :- ($\bar{x}$ is fractional)
(ϑ)
Karl pearson's coefficient :

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}} \sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}}$$

where $X = x - A$
$Y = y - A$ ∴ A is assumed mean
N = No. of given values.

Problems :-

1. Calculate karl pearson's coefficient :-

x	39	45	46	38	35	38	46	32	36	38
y	28	34	38	34	36	26	28	29	25	36

$\Sigma x = 392$, $N = 10$, $\Sigma y = 314$

$\bar{x} = \frac{\Sigma x}{N} = 39.2$; $\bar{y} = \frac{\Sigma y}{N} = 31.4$

∴ $A = 39$ $A = 31$

$\boxed{∴ X = x - 39}$ $\boxed{Y = y - 31}$

x	y	X	Y	X^2	Y^2	XY
38	28	-1	-3	1	9	3
45	34	+6	3	36	9	18
46	38	7	7	49	49	49
38	34	-1	3	1	9	-3
35	36	-4	5	16	25	-20
38	26	-1	-5	1	25	5
46	28	+7	-3	49	9	-21
32	29	-7	-2	49	4	14
36	25	-3	-6	9	36	18
38	36	-1	5	1	25	-5

$\Sigma X = 7$; $\Sigma Y = 7$; $\Sigma XY = 58$ $\Sigma X^2 = 212$ $\Sigma Y^2 = 200$

W.K.T.

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}} \sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}}$$

now $\Sigma x^2 - \frac{(\Sigma x)^2}{N} = 212 - \frac{(7)^2}{10}$

$= 212 - \frac{49}{10}$

$= 212 - 4.9$

$= 207.1$

$\Sigma y^2 - \frac{(\Sigma y)^2}{N} = 200 - \frac{(4)^2}{10}$

$= 200 - 1.6$

$= 198.4$

$\Sigma xy - \frac{\Sigma x \Sigma y}{N} = 58 - \frac{7 \times 4}{10}$

$= 58 - 2.8$

$= 55.2$

Now $r = \frac{55.2}{\sqrt{207.1}\sqrt{198.4}}$

$= 0.2723$

2. Find coefficient of correlation

Father Heights (inches) (x)	65	66	67	67	68	69	71	73
son Heights (y)	67	68	64	68	72	70	69	70

Sol: $\Sigma x = 546$ $N = 8$ $\Sigma y = 548$

$\bar{x} = \frac{\Sigma x}{N} = 68.25$ $\bar{y} = \frac{\Sigma y}{N} = 68.5$

$A = 68$ $A = 68$

$X = x - 68$ $Y = y - 68$

x	y	X	Y	X^2	Y^2	XY
65	67	-3	-1	9	1	3
66	68	-2	0	4	0	0
67	64	-1	-4	1	16	4
67	68	-1	0	1	0	0
68	72	0	4	0	16	0
69	70	1	2	1	4	2
71	69	3	1	9	1	3
73	70	5	2	25	4	10

$\Sigma X = 2$, $\Sigma Y = 4$, $\Sigma X^2 = 50$, $\Sigma Y^2 = 42$ $\Sigma XY = 22$

Now $\Sigma x^2 - \frac{(\Sigma x)^2}{N} = 50 - \frac{(2)^2}{8}$

$$= 50 - \frac{\cancel{4}^{1}}{\cancel{8}_{2}} = 50 - 0.5$$

$$= 49.5$$

$$\Sigma Y^2 - \frac{(\Sigma Y)^2}{N} = 42 - \frac{\cancel{16}^{2}}{\cancel{8}_{1}} = 42 - 2$$

$$= 40$$

$$\Sigma XY - \frac{\Sigma X \Sigma Y}{N} = 22 - \frac{\cancel{2 \times 4}}{\cancel{8}}$$

$$= 22$$

Now ~~r ≠~~   22

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}} \sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}}$$

$$= \frac{22}{\sqrt{49.5}\sqrt{40}}$$

$$= \frac{22}{(7.0356)(6.3245)}$$

$$= \frac{22}{44.4966}$$

$$\boxed{r = 0.4944}$$

3. Find correlation coefficient to the following data

Age of cars	2	4	6	7	8	10	12
Annual maintanence	1600	1500	1800	1900	1700	2100	2000

$\Sigma x = 49$; $N = 7$; $\Sigma Y = 12600$

$\bar{x} = \frac{\Sigma x}{N} = \frac{49}{7}$; $\bar{y} = \frac{\Sigma Y}{N}$

$= 7$ $= \frac{12600}{7}$

$\therefore A = 7$ $= 1800$ $A = 1800$

$X = x - 7 \quad ; \quad Y = y - 1800$

x	y	X	Y	X^2	Y^2	XY
2	1600	-5	-200	25	40000	1000
4	1500	-3	-300	9	90000	900
6	1800	-1	0	1	0	0
7	1900	0	100	0	10000	0
8	1700	1	-100	1	10000	-100
10	2100	3	300	9	90000	900
12	2000	5	200	25	40000	1000

$\Sigma X = 0 \; ; \; \Sigma Y = 0 \; ; \; \Sigma X^2 = 70 \; ; \; \Sigma Y^2 = 280{,}000 \; ; \; \Sigma XY = 3700$

Now $\Sigma X^2 - \frac{(\Sigma X)^2}{N} = 70 - \frac{0}{7} = 70 \quad ; \quad \Sigma Y^2 - \frac{(\Sigma Y)^2}{N} = 280000 - \frac{0}{7} = 280000$

$$\Sigma XY - \frac{\Sigma X . \Sigma Y}{N} = 3700 - \frac{0}{N} = 3700$$

Now

$$r = \frac{\Sigma XY - \frac{\Sigma X . \Sigma Y}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}}$$

$$= \frac{3700}{\sqrt{70}\sqrt{280000}}$$

$$= \frac{3700}{(8.3666)(529.1502)}$$

$$= \frac{3700}{4427.1880}$$

$$\boxed{r = 0.83574}$$

Rank correlation coefficient

Rank correlation coefficient is denoted by

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

ΣD^2 = Sum of squares of differences of two ranks

N = no of given values

Problems

1. Calculate Rank correlation coefficient to the following data

statistics (x)	1	2	3	4	5	6	7	8	9	10
maths (y)	2	4	1	5	3	9	7	10	6	8

Sol: $D : (x-y)$

D : -1 -2 2 -1 2 -3 0 -2 3 2

D^2 : 1 4 4 1 4 9 0 4 9 4

$\Sigma D^2 = 40 \qquad N = 10$

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

$$= 1 - \frac{6 \times 40}{10(10^2-1)}$$

$$= 1 - \frac{6 \times 4}{99}$$

$$= 1 - 0.2424$$

$$\boxed{\rho = 0.7576}$$

2. Find the rank of correlation coefficient to the following ranks

(x, y)
(1,1) (2,10) (3,3) (4,4) (5,5) (6,7) (7,2) (8,6) (16,13) (15,16) (14,12) (13,14) (9,8) (10,11) (11,15) (12,9).

Sol:

$D = x - y$

D : 0 -8 0 0 0 -1 5 2 -3 -1 2 -1 1 -1 -4 3

D^2 : 0 64 0 0 0 1 25 4 9 1 4 1 1 1 16 9

$\Sigma D^2 = 136 \qquad N = 16$

Scanned with CamScanner

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

$$= 1 - \frac{6 \times \cancel{136}\ 8.5}{\cancel{16}(16^2-1)}$$

$$= 1 - \frac{51}{255}$$

$$= 1 - 0.2$$

$$\boxed{P = 0.8}$$

3 Find rank correlation coefficient to the following data

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

Sol:

$D = x - y$

D : -3 1 -2 1 -4 -1 3 2 1 8

D^2 : 9 1 4 1 16 1 9 4 1 64

$\Sigma D^2 = 110 \qquad n = 10$

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

$$= 1 - \frac{\cancel{6}^{2} \times \cancel{11}\cancel{0}}{1\cancel{0}\,(\cancel{99})_{\cancel{9}\,3}}$$

$$= 1 - 2/3$$

$$= 1 - 0.6666$$

$$\boxed{P = 0.3333}$$

* Equal (or) Repeated Ranks:

$$\rho = 1 - 6\left[\frac{\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \dots}{n(n^2-1)}\right]$$

m = the no. of items where ranks are repeated

Problems'

1. From the following data calculate the rank correlation coefficient after making adjustment for tied ranks.

X	48	33	40	9	16	16	65	24	16	57
Y	13	13	24	6	15	4	20	9	6	19

Sol: First we have to assign ranks to the variables.

X	Rank (X)	Y	Rank (Y)	D [(R(x)) - R(y))]	D^2
48	8	13	5.5	2.5	6.21
33	6	13.	5.5	0.5	0.25
40	7	24	10	-3	9
9	1	6	2.5	-1.5	2.25
16	3	15	7	4	16
16	3	4	1	2	4
65	10	20	9	1	1
24	5	9	4	1	1
16	3	6	2.5	1	1
57	9	19	8	0.5	0.25
				1	1

$\therefore \Sigma D^2 = 41 \quad n = N = 10$

$\therefore$ 16 is repeated 3 times in x items hence $m = 3$.

$\therefore$ Since 13 & 6 are repeated twice in Y. items, hence $m = 2$

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right]}{N(N^2 - 1)}$$

$$= 1 - \frac{6\left[41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{10(99)}$$

$$\boxed{\rho = 0.733}$$

* obtain the rank correlation coefficient for the following data

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

Sol: In X - series 75 occurs two times.

so rank $= \frac{2+3}{2} = 2.5$; $m = 2$

→ Next value 68, so rank $= 4$

→ 64 occurs 3 times so rank $= \frac{5+6+7}{3} = 6$; $m = 3$

In Y- series

68 occurs twice so rank $= \frac{3+4}{2} = 3.5, m = 2.$

$$P = 1 - 6\left\{\frac{\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m)}{N(N^2-1)}\right]$$

$\cancel{4X} \times \quad 6\{\cancel{72}$

X	Y	R(x)	R(Y)	D = R(x) − R(y)	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

$\Sigma D^2 = 72.$

$$P = 1 - 6\left[\frac{72 + \frac{1}{12}(3^3-3) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2)}{10(10^2-1)}\right]$$

$$= 1 - \frac{6\left[72 + \frac{5}{2} + \frac{1}{2}\right]}{10(10^2-1)}$$

$$= 1 - \frac{6(75)}{10(99)}$$

$$= 1 - 0.4545$$

$$\boxed{P = 0.5455}$$

Scanned with CamScanner

* Regression -

Def :-

The statistical method or technique which help us to estimate an unknown value of one variable by using the known value of related variable is called regression.

→ The standard form of regression line or regression equation is $Y = a + bX \rightarrow (1)$ It is called regression eq. of Y on X

→ The regression Equation of X on Y is

$$X = a + bY \rightarrow (2)$$

→ Regression equation of X on Y is

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

where $\bar{x}$ is mean of X

$\bar{y}$ is mean of Y

r is correlation coefficient

σ_x is S.D on X

σ_y is S.D on Y

→ Regression equation of Y on X is

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

→ Regression coefficient of X on Y

$$b_{xy} = \frac{\Sigma XY}{\Sigma Y^2} = r \cdot \frac{\sigma_x}{\sigma_y} \rightarrow (3)$$

→ Regression coefficient of Y on X

$$b_{yx} = \frac{\Sigma XY}{\Sigma X^2} = r \cdot \frac{\sigma_y}{\sigma_x} \rightarrow (4)$$

From (3) & (4)

$$r^2 = b_{xy} * b_{yx}$$

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

Scanned with CamScanner

Problems:

1. By using method of least squares find the relation $Y = a + bx$ and find Y where $x = 10$

X	0	1	2	3	4
Y	1	1.8	3.3	4.5	6.3

Sol: Given

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

Normal equations of eq (1) is

$$\left.\begin{aligned} \Sigma Y &= an + b\Sigma x \\ \Sigma XY &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} eq (2)$$

x	Y	xY	x^2
0	1	0	0
1	1.8	1.8	1
2	3.3	6.6	4
3	4.5	13.5	9
4	6.3	25.2	16
			30

$\Sigma x = 10$; $\Sigma Y = 16.9$; $\Sigma xY = 47.1$ $\Sigma x^2 = 30$

eq (2) => $5a + 10b = 16.9$

$10a + 30b = 47.1$

$a = 0.72$; $b = 1.33$

eq(1) => $Y = 0.72 + (1.33)x$

where $x = 10$

$Y = 0.72 + (1.33 \times 10)$

$= 0.72 + 13.3$

$\boxed{Y = 14.02}$

2. In the following table 's' is weight of potassium bromide which will dissolve in 100 gr ml of water at $T°c$. Fit a curve of the form $s = mT + b$ by the method of least squares use this relation to estimate s when $T = 50°c$

T	0	20	40	60	80
S	54	65	75	85	96

Sol: Given form is

$$S = mT + b \rightarrow (1)$$

The normal equations of eq(1) is

$$\left.\begin{aligned} \Sigma S &= m\Sigma T + nb \\ \Sigma S &= m\Sigma T^2 + b\Sigma T \end{aligned}\right\} eq(2)$$

T	S	TS	T^2
0	54	0	0
20	65	1300	400
40	75	3000	1600
60	85	5100	3600
80	96	7680	6400

$\Sigma T = 200$; $\Sigma T^2 = 12000$; $\Sigma TS = 17080$; $n = 5$

$\Sigma S = 375$

eq(2) ⇒ $375 = m(200) + 5b$

$17080 = m(12000) + 200b$

$\therefore m = 0.52$, $b = 54.2$

$S = mT + b$

$= (0.52)T + 54.2$

$\therefore$ where $T = 50^\circ c$

$S = (0.52)50 + 54.2$

$$\boxed{S = 80.2}$$

3 Find the most likely production corresponding to a rainfall '40' from the following data

	average	S.D	coef. correction.
Rain fall (x)	$\bar{x}$ 30	σ_x 5	$r = 0.8$.
production (y)	$\bar{y}$ 500kgs	" y 100kgs	

Sol: Given that $\bar{X} = 30$ $\sigma_x = 5$ $r = 0.8$.

$\bar{Y} = 500$ $\sigma_y = 100$

Regression equation of Y on x

$$Y - \bar{Y} = r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

$$(Y - 500) = (0.8)\frac{100}{5}(X - 30)$$

$$Y - 500 = 16x - 480$$

$$16x - Y = -20$$

$$Y = 16x + 20$$

where $x = 40$

$$Y = 16(40) + 20$$

$$= 660$$

* Angle b/w 2 regression lines:-

The angle b/w 2 regression lines

i.e Regression of x on Y & Regression of Y on x

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

(or),

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

Problems:-

1. Find the angle θ between 2 regression lines, standard deviation of Y is twice the standard deviation of x, $r = 0.25$

Sol. Given $\sigma_y = 2\sigma_x$, $r = 0.25$

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

$$= \frac{2\sigma_x}{\sigma_x}\left(\frac{1-(0.25)^2}{(0.25)}\right)$$

$$= 2\left[\frac{1-0.0625}{0.25}\right]$$

$$= 2\left[\frac{0.9375}{0.25}\right]$$

$$\tan\theta = 7.5$$

$$\theta = \tan^{-1}(7.5)$$

$$= 82.4054$$

2. If $\sigma_x = \sigma_y = \sigma$ & angle b/w regression lines is $\tan^{-1}(4/3)$ find r.

Sol:- we know that

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

$$\tan\theta = \frac{\cancel{\sigma}}{\cancel{\sigma}}\left(\frac{1}{r} - r\right)$$

$$\theta = \tan^{-1}\left(\frac{1}{r} - r\right)$$

$\therefore$ Given $\theta = \tan^{-1}(4/3)$

$$\therefore \tan^{-1}(4/3) = \tan^{-1}\left(\frac{1}{r} - r\right)$$

$$\tan\left[\tan^{-1}(4/3)\right] = \frac{1}{r} - r$$

$$\frac{4}{3} = \frac{1}{r} - r$$

$$4r = 3 - 3r^2$$

$$3r^2 + 4r - 3 = 0$$

$$\cancel{3r^2 + r = 0}$$

$$r = \frac{-4 \pm \sqrt{16 + 4(3\times 3)}}{2(3)}$$

$$= \frac{-4 \pm \sqrt{16 + 36}}{6}$$

$$= \frac{-4 \pm \sqrt{52}}{6}$$

$$= \frac{-4 \pm 7.2111}{6}$$

$$r = \left(\frac{3.2111}{6} \text{ (or) }, \frac{-11.2111}{6}\right)$$

$$r = (0.5352 \text{ (or) }, -1.8685)$$

$\therefore$ Since 'r' lies between -1 to 1.

$-1.8685 \notin (-1$ to $1)$

$\therefore r = 0.5352$.

3. The tangent of the angle between 2 regression lines is 0.6 and $\sigma_x = \frac{\sigma_y}{2}$ find correlation coefficient r.

Sol: Given that

$$\tan\theta = 0.6 \quad \& \quad \sigma_x = \frac{\sigma_y}{2}$$

w.k.T

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

$$0.6 = \frac{\sigma_y}{\sigma_y/2}\left(\frac{1}{r} - r\right)$$

$$0.6 = 2\left(\frac{1}{r} - r\right)$$

$$1.2 = \frac{1 - r^2}{r}$$

$$r^2 + 1.2r - 1 = 0$$

$$r = \frac{-(1.2) \pm \sqrt{1.44 + 4}}{2}$$

$$= \frac{-(1.2) \pm \sqrt{5.44}}{2}$$

$$= \frac{-(1.2) \pm 2.3323}{2}$$

$$= \left(\frac{1.1323}{2} \ \&, \ \frac{-3.5323}{2}\right)$$

$$= (0.5661 \ \& \ -1.76615)$$

we know that

r lies b/w -1 to 1

but $-1.76615 \notin -1$ to 1

$$\boxed{\therefore r = 0.5661}$$