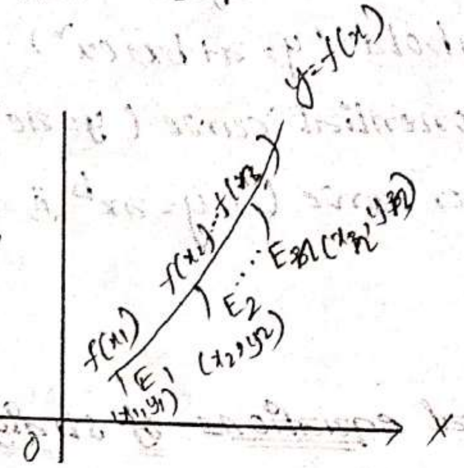
UNIT - II : correlation and curve fitting

Curve fitting

To find the equation of curve $y=f(x)$ which fit the given data is called curve fitting

Method of least squares:



Let $(x_1,y_1), (x_2,y_2), (x_3,y_3), \ldots, (x_n,y_n)$ be a set of n data points in a scatter diagram and $f(x_1), f(x_2), f(x_3), \ldots, f(x_n)$ be the expected values of y and $y_1, y_2, y_3, \ldots, y_n$ be the observed values of y.

The difference between observed values and expected values is known as error. These are denoted by $E_1, E_2, E_3, \ldots, E_n$ respectively

By using the legandry's least square criteria such that the curve having minimum sum of squares of the errors is the best fitting curve.

i.e $\sum_{i=1}^{n} E_i^2 = E_1^2 + E_2^2 + E_3^2 + \cdots + E_n^2$ is minimum

This method is called method of least squares

Types of curves:

There are mainly four types of curves. They are
1. straight line ($y = a+bx$)
2. parabola ($y = a+bx+cx^2$)
3. Exponential curve ($y = ae^{bx}$)
4. power curve (i. $y = ax^b$ ii. $y = ab^x$)

1. Straight line:

Derive the normal equations of straight line:

Let equation of straight line be
$$y = a+bx \longrightarrow ①$$

Let $f(x_1), f(x_2), f(x_3), \cdots, f(x_n)$ be the expected values $(a+bx)$ and $y_1, y_2, y_3, \cdots, y_n$ be the observed values of y corresponding to $x_1, x_2, x_3, \cdots, x_n$.

Suppose the above st-line ① passing through the points $(x_1,y_1), (x_2,y_2), \cdots, (x_n,y_n)$ then we have $y_i = a+bx_i, \forall i = 1,2,\cdots,n$

Let 'E_i' be the error between the expected values $(a+bx)$ and observed values (y_i) then we have
$$E_i = y_i - (a+bx_i) \longrightarrow ②$$

To find the values of a and b in eq (2), by using method of least squares we minimize the error

i.e $\sum_{i=1}^{n} E_i^2$ is minimum

consider $S = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - (a+bx_i))^2 \rightarrow (3)$

($\because$ from (2))

To minimise S, it means that differentiating eq (3) partially w.r.t a and b and equating to 0, we have

i.e $\frac{\partial S}{\partial a} = 0$

$$\sum_{i=1}^{n} \frac{\partial}{\partial a} \left[y_i - (a+bx_i) \right]^2 = 0$$

$$\sum_{i=1}^{n} 2 \left[y_i - (a+bx_i) \right](-1) = 0$$

$$\sum_{i=1}^{n} (y_i - (a+bx_i)) = 0$$

$$\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} - b\sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} y_i - an - b\sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i$$

$\frac{\partial S}{\partial b} = 0$

$$\frac{\partial}{\partial b} \left[\sum_{i=1}^{n} (y_i - (a+bx_i))^2 \right] = 0$$

$$\sum_{i=1}^{n} 2 \cdot (y_i - (a+bx_i)) (-x_i) = 0$$

$$\sum_{i=1}^{n} (y_i - (a+bx_i)) (x_i) = 0$$

$$\sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2$$

$$\boxed{\begin{aligned} \therefore \sum_{i=1}^{n} y_i &= an + b\sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i &= a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \end{aligned}}$$

The above two equations are called the normal eqs of straight line and solve those eqns, we get a and b values

Problems:

1. Fit a least square straight line to the following data

x	1	2	3	4	5
y	16	19	23	26	30

2.

x	1	2	3	4	5
y	5	12	26	60	90

3.

x	0	1	2	3	4
y	2.1	3.5	5.4	7.3	8.2

4.

x	1	3	5	7	9
y	1.5	2.8	4.0	4.7	6.0

5.

x	0	1	2	3	4
y	1.0	1.6	1.4	2.5	2.8

1. Let the equation of straight line be $y = a + bx \rightarrow (1)$

the normal equations of eq (1) are

$$\left.\begin{aligned} an + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \end{aligned}\right\} \rightarrow (2)$$

Here $n = 5$

the values of Σx, Σy, Σx^2 and Σxy are calculated as follows

x	y	x^2	xy
1	16	1	16
2	19	4	38
3	23	9	69
4	26	16	104
5	30	25	150
15	114	55	377

$\Sigma x = 15 \quad \Sigma y = 114 \quad \Sigma x^2 = 55 \quad \Sigma xy = 377$

substituting these Σx, Σy, Σx^2, Σxy values in (2), we have

$$a(5) + b(15) = 114 \rightarrow (3)$$

$$a(15) + b(55) = 377 \rightarrow (4)$$

$$5a + 15b = 114$$

$$15a + 55b = 377$$

Scanned with CamScanner

solving (3) and (4), we get a and b values

Now

$3 \times (3) \Rightarrow 15a + 45b = 342$

$(4) \Rightarrow 15a + 55b = 377$

$\quad (-) \quad (-) \quad\quad (-)$

$-10b = -35$

$b = 3.5$

from (3) $\Rightarrow$ $a = \frac{1}{5}(114 - 15b)$

$= \frac{1}{5}(114 - 52.5)$

$= \frac{1}{5}(61.5)$

$= 12.3$

Hence the required st. line is $y = 12.3 + (3.5)x$

2. let the equation of st. line be $y = a + bx \longrightarrow (1)$

The normal equations of eq(1) are

$$\left. \begin{aligned} an + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \end{aligned} \right\} (2)$$

The values of Σx, Σy, Σx^2, Σxy are calculated as follows

Here $n = 5$

x	y	x^2	xy
1	5	1	5
2	12	4	24
3	26	9	78
4	60	16	240
5	90	25	450
15	193	55	797

$\Sigma x = 15 \quad \Sigma y = 193 \quad \Sigma x^2 = 55 \quad \Sigma xy = 797$

sub $\Sigma x, \Sigma y, \Sigma x^2, \Sigma xy$ in eq ②, we have

$$a(5) + 15b = 193 \longrightarrow ③$$
$$15a + 55b = 797 \longrightarrow ④$$

solving ③ & ④ we get

$$a = -26.3 \quad , b = 21.8$$

Hence the required st line is $y = -26.3 + (21.8)x$

3. Let the equation of st-line is $y = a + bx \longrightarrow ①$

The normal equations of eq① are

$$\left.\begin{aligned} an + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \end{aligned}\right\} \longrightarrow ②$$

Here $n = 5$

The values of $\Sigma x, \Sigma y, \Sigma x^2, \Sigma xy$ are calculated

x	y	x^2	xy
0	2.1	0	0
1	3.5	1	3.5
2	5.4	4	10.8
3	7.3	9	21.9
4	8.2	16	32.8
10	26.5	30	69

$\Sigma x = 10 \quad \Sigma y = 26.5 \quad \Sigma x^2 = 30 \quad \Sigma xy = 69$

sub these values in (2), we have

$$5a + 10b = 26.5 \longrightarrow (3)$$
$$10a + 30b = 69 \longrightarrow (4)$$

solving (3) & (4), we get

$$a = 2.1 \quad b = 1.6$$

Hence the required st. line is $y = 2.1 + (1.6)x$

4. Let the eqn of st line is $y = a + bx \longrightarrow (1)$

The normal eqns are

$$\left.\begin{aligned} an + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \end{aligned}\right\} \longrightarrow (2)$$

Here $n = 5$

The values of Σx, Σy, Σx^2, Σxy are calculated

x	y	x^2	xy
1	1.5	1	1.5
3	2.8	9	8.4
5	4.0	25	20
7	4.7	49	32.9
9	6.0	81	54
$\Sigma x = 25$	$\Sigma y = 19$	$\Sigma x^2 = 165$	$\Sigma xy = 116.8$

sub these values in ②, we have

$$5a + 25b = 19 \rightarrow ③$$
$$25a + 165b = 116.8 \rightarrow ④$$

solve ③ & ④, we get

$$a = 1.075 \quad b = 0.545$$

Hence the required st. line is $y = 1.075 + (0.545)x$

5. Let the eqn of st. line is $y = a + bx \rightarrow ①$

The normal eqns are

$$\left. \begin{aligned} an + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \end{aligned} \right\} \rightarrow ②$$

Here $n = 5$

The values of Σx, Σy, Σx^2, Σxy are calculated

x	y	x^2	xy
0	1.0	0	0
1	1.6	1	1.6
2	1.4	4	2.8
3	2.5	9	7.5
4	2.8	16	11.2
$\Sigma x = 10$	$\Sigma y = 9.3$	$\Sigma x^2 = 30$	$\Sigma xy = 23.1$

sub these values in (2), we have

$$5a + 10b = 9.3 \longrightarrow (3)$$
$$10a + 30b = 23.1 \longrightarrow (4)$$

solve (3) & (4) we get

$$a = 0.96, \quad b = 0.45$$

Hence the required st. line is $y = 0.96 + (0.45)x$.

Method-II : Second degree parabola

Derive the normal eqns of second degree parabola

Let the normal eq of a st. line be $y = a + bx + cx^2 \rightarrow (1)$

let $f(x_1), f(x_2), f(x_3), \ldots, f(x_n)$ be the expected values and $y_1, y_2, y_3, \ldots, y_n$ be the observed values of y corresponding to $x_1, x_2, x_3, \ldots, x_n$.

Suppose that the above eq ① passing through the points $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$, then we have $y_i = a+bx_i+cx_i^2, \ \forall i=1,2,3,\cdots n$.

Let E_i be the error between observed values y_i and expected values $a+bx_i+cx_i^2$, then we have

$$E_i = y_i-(a+bx_i+cx_i^2) \longrightarrow ②$$

To find the values of a,b & c by using the method of least squares we minimize the error. i.e $\sum_{i=1}^{n} E_i^2$ is minimum

Consider $S = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} [y_i-(a+bx_i+cx_i^2)] \rightarrow ③$

To minimise S, it means that differentiating partially w.r.t a, b & c and equating to '0'

i.e $\frac{\partial S}{\partial a}=0$

differentiate

$$2\sum_{i=1}^{n}[y_i-(a+bx_i+cx_i^2)](-1) = 0$$

$$\sum_{i=1}^{n}[y_i-(a+bx_i+cx_i^2)](1)=0$$

$$\sum_{i=1}^{n}y_i - a\sum_{i=1}^{n} - b\sum_{i=1}^{n}x_i - c\sum_{i=1}^{n}x_i^2=0$$

i.e $an+b\sum_{i=1}^{n}x_i + c\sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}y_i$

Now $\frac{\partial S}{\partial b}=0$

$$2\sum_{i=1}^{n}\left[y_i-(a+bx_i+cx_i^2)\right](-x_i)=0$$

$$\sum_{i=1}^{n}\left[y_i-(a+bx_i+cx_i^2)\right]x_i=0$$

$$\sum_{i=1}^{n}x_iy_i-a\sum_{i=1}^{n}x_i-b\sum_{i=1}^{n}x_i^2-c\sum_{i=1}^{n}x_i^3=0$$

$$\sum_{i=1}^{n}x_iy_i=a\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}x_i^2+c\sum_{i=1}^{n}x_i^3$$

Now $\frac{\partial S}{\partial c}=0$

$$2\sum_{i=1}^{n}\left[y_i-(a+bx_i+cx_i^2)\right](-x_i^2)=0$$

$$\sum_{i=1}^{n}\left[y_i-(a+bx_i+cx_i^2)\right]x_i^2=0$$

$$\sum_{i=1}^{n}x_i^2y_i-a\sum_{i=1}^{n}x_i^2-b\sum_{i=1}^{n}x_i^3-c\sum_{i=1}^{n}x_i^4=0$$

$$\sum_{i=1}^{n}x_i^2y_i=a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i^3+c\sum_{i=1}^{n}x_i^4$$

$$\therefore\quad an+b\sum_{i=1}^{n}x_i+c\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}y_i$$

$$a\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}x_i^2+c\sum_{i=1}^{n}x_i^3=\sum_{i=1}^{n}x_iy_i$$

$$a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i^3+c\sum_{i=1}^{n}x_i^4=\sum_{i=1}^{n}x_i^2y_i$$

The above three eqns are called the normal equations of second degree parabola and solve these equations, we get a, b and c values.

Problems

1. Fit a least square second degree parabola to the following data

1.

x	0	1	2	3	4
y	1.0	1.4	1.6	2.5	3.8

2.

x	1	2	3	4	5
y	5	12	26	30	40

3.

x	0	1	2	3	4
y	2.1	3.5	5.4	7.3	8.2

4.

x	1	3	5	7	9
y	1.5	2.8	4.0	4.7	6.0

5.

x	0	1	2	3	4	5	6
y	1	3	5	9	15	23	33

1 Sol: Let the eq of second degree parabola be

$$y = a + bx_i + cx_i^2 \longrightarrow (1)$$

The normal eqns of eq (1) are

$$\left.\begin{aligned} an + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 &= \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 &= \sum_{i=1}^{n} x_i^2 y_i \end{aligned}\right\} \rightarrow (2)$$

Here $n=5$

The values of Σx, Σy, Σx^2, Σx^3, Σx^4, Σxy and Σx^2y are calculated as follows:

x	y	x^2	x^3	x^4	xy	x^2y
0	1.0	0	0	0	0	0
1	1.4	1	1	1	1.4	1.4
2	1.6	4	8	16	3.2	6.4
3	2.5	9	27	81	7.5	22.5
4	3.8	16	64	256	15.2	60.8
10	10.3	30	100	354	27.3	91.1

Substitute $\Sigma x=10$ $\Sigma x^2=30$ $\Sigma x^4=354$ $\Sigma x^2y=91.1$
$\Sigma y=10.3$ $\Sigma x^3=100$ $\Sigma xy=27.3$

in ② we have,

$$5a+10b+30c=10.3 \longrightarrow ③$$
$$10a+30b+100c=27.3 \longrightarrow ④$$
$$30a+100b+354c=91.1 \longrightarrow ⑤$$

solve ③, ④ & ⑤ we have

$a=1.0771$

$b=-0.0442$

$c=0.1785$

Hence the required par. line is $y=(1.0771)x+$
$b(-0.0442)x_i+$
$c(0.1785)x_i^2$

2. Let the eq of second degree parabola be

$$y = a + bx + cx^2 \rightarrow (1)$$

The normal eqns of eq (1) are

$$\left.\begin{aligned} an + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 &= \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 &= \sum_{i=1}^{n} x_i^2 y_i \end{aligned}\right\} (2)$$

Here $n = 5$.

The values of Σx, Σy, Σx^2, Σx^3, Σx^4, Σxy and $\Sigma x^2 y$ are calculated as follows

x	y	x^2	x^3	x^4	xy	x^2y
1	5	1	1	1	5	5
2	12	4	8	16	24	48
3	26	9	27	81	78	234
4	30	16	64	256	120	480
5	40	25	125	625	200	1000
15	113	55	225	979	427	1767

Sub $\Sigma x = 15$ $\Sigma x^2 = 55$ $\Sigma x^4 = 979$ $\Sigma x^2 y = 1767$

$\Sigma y = 113$ $\Sigma x^3 = 225$ $\Sigma xy = 427$ in (2)

$$5a + 15b + 55c = 113 \longrightarrow (3)$$

$$15a + 55b + 225c = 427 \longrightarrow (4)$$

$$55b + 225b + 979c = 1767 \longrightarrow (5)$$

By solving (3), (4) & (5) we get

$a = -5.8$, $b = 10.5142$, $c = -0.2857$

$\therefore$ eq of second degree parabola is $y = (-5.8) + (10.5142)x + (-0.2857)x^2$

3. Let the eq of second degree parabola is

$$y = a + bx + cx^2 \longrightarrow (1)$$

The normal equations of eq (1) be

$$\left.\begin{aligned} an + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 &= \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 &= \sum_{i=1}^{n} x_i^2 y_i \end{aligned}\right\} \longrightarrow (2)$$

Here $n = 5$

The values of $\Sigma x, \Sigma y, \Sigma x^2, \Sigma y^2, \Sigma x^3, \Sigma x^4, \Sigma xy, \Sigma x^2 y$ are calculated as follows

x	y	x^2	x^3	x^4	xy	x^2y
0	2.1	0	0	0	0	0
1	3.5	1	1	1	3.5	3.5
2	5.4	4	8	16	10.8	21.6
3	7.3	9	27	81	21.9	65.7
4	8.2	16	64	256	32.8	131.2
10	26.5	30	100	354	69	222

$\Sigma x = 10$, $\Sigma y = 26.5$, $\Sigma x^2 = 30$, $\Sigma x^3 = 100$, $\Sigma x^4 = 354$

$\Sigma xy = 69$, $\Sigma x^2 y = 222$

sub those in ②, we get

$$5a + 10b + 30c = 26.5 \longrightarrow ③$$
$$10a + 30b + 100c = 69 \longrightarrow ④$$
$$30a + 100b + 354c = 222 \longrightarrow ⑤$$

By solving ③, ④ & ⑤, we get

$$a = 1.9571,\ b = 1.8857,\ c = -0.0714$$

∴ eq of second degree parabola is $y = (1.9571) + (1.8857)x + (-0.0714)x^2$

4. Let the eq of second degree parabola is

$$y = a + bx + cx^2 \longrightarrow ①$$

The normal eq of eq① are

$$\left.\begin{aligned} an + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 &= \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 &= \sum_{i=1}^{n} x_i^2 y_i \end{aligned}\right\} \rightarrow ②$$

Here $n = 5$

The values of Σx, Σy, Σx^2, Σx^3, Σx^4, Σxy, $\Sigma x^2 y$ are calculated as follows.

Scanned with CamScanner

x	y	x^2	x^3	x^4	xy	x^2y
1	1.5	1	1	1	1.5	1.5
3	2.8	9	27	81	8.4	25.2
5	4.0	25	125	625	20	100
7	4.7	49	343	2401	32.9	230.3
9	6.0	81	729	6561	54	486
25	19	165	1225	9669	116.8	843

$\therefore \Sigma x = 25$, $\Sigma x^2 = 165$, $\Sigma x^4 = 9669$ $\Sigma x^2 y = 843$

$\Sigma y = 19$: $\Sigma x^3 = 1225$ $\Sigma xy = 116.8$ sub in (2)

$$5a + 25b + 165c = 19 \longrightarrow (3)$$

$$25a + 165b + 1225c = 116.8 \longrightarrow (4)$$

$$165b + 1225b + 9669c = 843 \longrightarrow (5)$$

solve (3), (4) & (5), we get

$a = 0.9232$ $b = 0.6342$ $c = -8.9285 \times 10^{-3}$

$c = -0.0089$

$\therefore$ The equation of second degree parabola is

$$y = 0.9232 + 0.6342x - 0.0089x^2$$

5. Let the eq of second degree parabola is

$$y = a + bx + cx^2 \longrightarrow (1)$$

The normal eqns of eq (1) are

$$an + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i \quad \rightarrow (2)$$

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i$$

Here $n=7$

The values of $\Sigma x, \Sigma y, \Sigma x^2, \Sigma x^3, \Sigma x^4, \Sigma xy, \Sigma x^2 y$ are calculated as follows.

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	1	0
1	3	1	1	1	3	3
2	5	4	8	16	10	20
3	9	9	27	81	27	81
4	15	16	64	256	60	240
5	23	25	125	625	115	575
6	33	36	216	1296	198	1188
21	89	91	441	2275	414	2107

$\Sigma x = 21 \quad \Sigma x^2 = 91 \quad \Sigma x^4 = 2275 \quad \Sigma x^2 y = 2107$

$\Sigma y = 89 \quad \Sigma x^3 = 441 \quad \Sigma xy = 414$ sub in (2)

$7a + 21b + 91c = 89 \longrightarrow (3)$

$21a + 91b + 441c = 414 \longrightarrow (4)$

$91a + 441b + 2275c = 2107 \rightarrow (5)$

solve (3), (4) & (5) we get

$a = 1.0119$, $b = 0.3928$, $c = 0.8095$

$\therefore$ The eq of second degree parabola is

$\therefore \quad y = 1.0119 + 0.3928x + 0.8095x^2$

Method-3: Exponential curve

Derive the normal eq of exponential curve

Let the eq of exponential curve is

$$y = ae^{bx} \longrightarrow ①$$

taking ln or $\log_e$ in ①, we have

i.e $\quad \ln y = \ln(ae^{bx})$

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

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

$$y = A + bx(1)$$

$$Y = A + bx \longrightarrow ②$$

eq ② represents a straight line in x and Y

The normal eqns of eq ② are

$$An + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} Y_i$$

$$A\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i Y_i$$

where $Y = \ln y$, $A = \ln a$, $a = \text{antilog} A$ (press: shift + ln)

The above two eqns are called the normal eqns of ① and solve these eqns, we get A, b

where $a = \text{antilog} A$

Problems : Fit a exponential curve from the following data

1.

x	0	1	2	3	4
y	2.1	3.5	5.4	7.3	8.2

2.

x	1	3	5	7	9
y	1.5	2.8	4.0	4.7	6.0

3.

x	1	2	3	4	5
y	5	12	26	60	90

4.

x	0	1	2	3	4
y	1.0	1.4	1.6	2.5	3.8

5.

x	0	1	2	3	4	5
y	1	3	5	9	15	23

1. let the eq of exponential curve be $y = ae^{bx} \rightarrow (1)$

The normal equations of eq (1) are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i Y_i \end{aligned}\right\} \rightarrow (2)$$

where $n=5$, $Y = \ln y$, $A = \ln a$, $a = \text{antilog } A$

The values of Σx, ΣY, Σx^2 and ΣxY are calculated as follows

x	y	x^2	$Y=\ln y$	xY
0	2.1	0	0.7419	0
1	3.5	1	1.2527	1.2527
2	5.4	4	1.6863	3.3726
3	7.3	9	1.9878	5.9634
4	8.2	16	2.1041	8.4164
10	26.5	30	7.7728	19.0051

sub these values in eq ② we have

$$5A + b10 = 7.7728 \longrightarrow ③$$

$$10b + 30b = 19.0051 \longrightarrow ④$$

solve ③ & ④ we have

$$A = 0.8626 \quad , \quad b = 0.3459$$

$a = \text{antilog}A$ $\qquad$ $Y = \ln y = \ln 26.5$

$a = 2.3693$ $\qquad$ $= 3.2771$

Hence the required exponential curve is

$$y = (2.3693)\, e^{(0.3459)x}$$

2. Let the eqn of exponential curve be $y = ae^{bx} \longrightarrow ①$

The normal eqns of eq① are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i Y_i \end{aligned}\right\} \longrightarrow ②$$

where $n=5$, $a = \text{antilog}A$, $A = \ln a$

The values of Σx, Σy, Σx^2, ΣY, ΣxY are calculated

x	y	x^2	$Y = \ln y$	xY
1	1.5	1	0.4054	0.4054
3	2.8	9	1.0296	3.0888
5	4.0	25	1.3862	6.931
7	4.7	49	1.5475	10.8325
9	6.0	81	1.7917	16.1253
25	19	165	6.1604	37.383

$\Sigma x = 25$, $\Sigma y = 19$ $\Sigma x^2 = 165$ $\Sigma Y = 6.1604$ $\Sigma xY = 37.383$

sub these in ②, we have

$$5A + 25b = \cancel{19}\ 6.1604 \rightarrow ③$$

$$25A + 165b = 37.383 \rightarrow ④$$

solve ③ & ④, we get

$$A = 0.4094, \quad b = 0.1645$$

$a = \text{antilog } A$

$a = 1.5059$

Hence the required exponential curve is

$$y = (1.5059)e^{(0.1645)x}$$

3. let the eqn of exponential curve be $y = ae^{bx} \rightarrow ①$

The normal eqns. of eq① are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i Y_i \end{aligned}\right\} \rightarrow ②$$

where $n=5$

The values of $\Sigma x, \Sigma y, \Sigma x^2, \Sigma Y, \Sigma xY$ are

x	y	x^2	$Y=\ln y$	xY
1	5	1	1.6094	1.6094
2	12	4	2.4849	4.9698
3	26	9	3.2580	9.774
4	60	16	4.0943	16.3772
5	90	25	4.4998	22.499
15	193	55	15.9464	55.2294

$\Sigma x = 15 \quad \Sigma y = 193 \quad \Sigma x^2 = 55 \quad \Sigma Y = 15.9464$

$\Sigma xY = 55.2294$ sub these values in ②, we have

$$5A + 15b = 15.9464 \rightarrow ③$$

$$15A + 55b = 55.2294 \rightarrow ④$$

solve ③ & ④, we have

$$A = 0.9722, \quad b = 0.7390$$

$a = \text{anti}\log A$

$= 2.6437$

Hence the required exponential curve is

$$y = (2.6437)\, e^{(0.7390)x}$$

4. let the eqn of exponential curve be $y = ae^{bx} \rightarrow ①$

The normal eqns of eq① are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ A\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \end{aligned}\right\} \rightarrow (2)$$

where $n = 5$ $Y = \ln y$, $a = \text{antilog} A$.

The values of $\Sigma x, \Sigma y, \Sigma x^2, \Sigma Y, \Sigma xY$ are calculated

x	y	x^2	$Y = \ln y$	xY
0	1.0	0	0	0
1	1.4	1	0.3364	0.3364
2	1.6	4	0.4700	0.94
3	2.5	9	0.9162	2.7486
4	3.8	16	1.3350	5.34
10	10.3	30	3.0576	9.365

$\Sigma x = 10$ $\Sigma x^2 = 30$ $\Sigma xY = 9.365$

$\Sigma y = 10.3$ $\Sigma Y = 3.0576$ sub in (2), we have

$$5A + 10b = 3.0576 \longrightarrow (3)$$

$$10A + 30b = 9.365 \longrightarrow (4)$$

solve (3) & (4), we get

$$A = -0.0384 \quad b = 0.3249$$

$a = \text{antilog} A$

$a = 0.9623$

Hence the required exponential curve is

$$y = (0.9623)e^{(0.3249)x}$$

5. Let the eq of exponential curve is $y = ae^{bx} \rightarrow (1)$

The normal eqn of eq (1) are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i Y_i \end{aligned}\right\} \rightarrow (2)$$

The values of $\Sigma x, \Sigma y, \Sigma x^2, \Sigma Y, \Sigma xY$ are calculated

x	y	x^2	$Y = \ln y$	xY
0	1	0	0	0
1	3	1	1.0986	1.0986
2	5	4	1.6094	3.2188
3	9	9	2.1972	6.5916
4	15	16	2.7080	10.832
5	23	25	3.1354	15.677
15	56	55	10.7486	37.418

$\Sigma x = 15$, $\Sigma y = 56$, $\Sigma x^2 = 55$, $\Sigma Y = 10.7486$, $\Sigma xY = 37.418$ sub in (2)

where $n = 6$

$$6A + 15b = 10.7486 \rightarrow (3)$$

$$15A + 55b = 37.418 \rightarrow (4)$$

solve (3) & (4), we get A, b

$$A = 0.2847 \qquad b = 0.6026$$

$a = \text{antilog } A$

$a = 1.3293$

Hence the required exponential eqns are

$$y = (1.3293)e^{(0.6026)x}$$

Method-4: Powercurve

1. Derive the normal eqns of powercurve ($y=ax^b$).

Let the eqn of power curve be $y=ax^b \rightarrow (1)$

Taking log on b.s in (1), we have,

$$\log y = \log ax^b$$

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

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

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

where $X=\log x$, $Y=\log y$, $A=\log a$

∴ eq (2) represents the st. line in X and Y

The normal eqns of eq (2) are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} X_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 &= \sum_{i=1}^{n} X_iY_i \end{aligned}\right\} \rightarrow (3)$$

The above two eqns are called the normal eqns (1) and solve these eqn's, we get A, b

where $a=$ antilog A (shift + log)

problems:

1.

x	1	2	3	4	5	6
y	4	5.7	6.9	8	8.9	9.8

Fit a power curve is of the form $y=ax^b$ from the below data

2.

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

3.

x	0	1	2	3	4	5
y	4	3	4.243	5.196	6	6.708

4.

x	0	1	2	3	4	5	6
y	10	21	35	59	92	200	400

5.

x	0	2	4
y	5.1	11	31.1

1. let the eqn of powercurve be $y = ax^b \rightarrow (1)$

The normal eqns of eq(1) are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} X_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 &= \sum_{i=1}^{n} X_i Y_i \end{aligned}\right\} \rightarrow (2)$$

The values of ΣX_i, ΣX_i^2, ΣY_i, $\Sigma X_i Y_i$ are calculated as follows

$X = \log x$, $Y = \log y$, $A = \log a$

x	y	$X = \log x$	X^2	Y	XY
1	4	0	0	0.6020	0
2	5.7	0.3010	0.0906	0.7558	0.2274
3	6.9	0.4771	0.2276	0.8388	0.4001
4	8	0.6020	0.3624	0.9030	0.5436
5	8.9	0.6989	0.4884	0.9493	0.6634
6	9.8	0.7781	0.6054	0.9912	0.7712
				5.0401	2.6057

sub these in eq ② we have

$$6A + 2.8571b = 5.0481 \rightarrow ③$$
$$2.8571A + 1.7744b = 2.6057 \rightarrow ④$$

solve ③ & ④ . we have

$A = 0.6033 \quad b = 0.4969$

$a = \text{antilog } A$

$a = 4.0114$

Hence the required power curve is

$$y = 4.0114\, x^{(0.4969)}$$

2. let the eqn of power curve be $y = ax^b \rightarrow ①$

The normal eqns of eq ① are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} X_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 &= \sum_{i=1}^{n} X_i Y_i \end{aligned}\right\} \rightarrow ②$$

The values are $X = \log x,\ Y = \log y,\ A = \log a$

x	y	$X = \log x$	X^2	$Y = \log y$	XY
1	151	0	0	2.1789	0
2	100	0.3010	0.0906	2	0.602
3	61	0.4771	0.2276	1.7853	0.8517
4	50	0.6020	0.3624	1.6989	1.0227
5	20	0.6989	0.4884	1.3010	0.9092
6	8	0.7781	0.6054	0.9030	0.7026
		2.8571	1.7724	9.8671	4.0882

sub these in ② we get

$$6A + b\,2.8571 = 9.8671 \longrightarrow ③$$

$$2.8571A + 1.7744b = 4.0882 \longrightarrow ④$$

solve ③ & ④ we get :

$A = 2.3467$, $b = -1.4746$

$a = \text{antilog}\, A$

$a = 222.1774$

let the required power curve is

$$y = (222.1774)\, x^{(-1.4746)}$$

3\. let the powercurve be $y = ax^b \rightarrow ①$

The normal eqns of eq ① are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} X_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 &= \sum_{i=1}^{n} X_iY_i \end{aligned}\right\} \rightarrow ②$$

x	y	$X=\log x$	X^2	$Y=\log y$	XY
0	4	1	1	0.6020	0.6020
1	3	0	0	0.4771	0
2	4.243	0.3010	0.0906	0.6276	0.1889
3	5.196	0.4771	0.2276	0.7156	0.3414
4	6	0.6020	0.3624	0.7781	0.4684
5	6.708	0.6989	0.4884	0.8265	0.5776
15	29.147	3.079	2.169	4.0269	2.1783

Sub the values in ②, we get,

$$6A + b \cdot 0.379b = 4.0269 \rightarrow ③$$

$$0.379A + 2.169b = 2.1783 \rightarrow ④$$

solve ③ & ④, we get

$A = 0.6144$, $b = 0.8969$

$a = \text{antilog} A$

$a = 4.1152$

Hence the required power curve is

$$y = 4.1152\, x^{0.8969}$$

4. let the power curve be $y = ax^b \rightarrow$ ①

The normal eqns of eq ① are

$$\left.\begin{aligned} An + b\sum_{i=1}^{n} X_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 &= \sum_{i=1}^{n} X_i Y_i \end{aligned}\right\} \rightarrow ②$$

where $n = 7$

x	y	$X = \log x$	X^2	$Y = \log y$	XY
0	10	1	1	1	1
1	21	0	0	1.3222	0
2	35	0.3010	0.0906	1.5440	0.4647
3	59	0.4771	0.2276	1.7708	0.8448
4	92	0.6020	0.3624	1.9637	1.1819
5	200	0.6987	0.4884	2.3010	1.6077
6	400	0.7781	0.6054	2.6020	2.0246
21	817	3.8569	2.7744	12.5037	7.1237

sub these in (2), we get

$$7A + 3.8569 b = 12.5037 \longrightarrow (3)$$

$$3.8569A + 2.7744 = 7.1237 \longrightarrow (4)$$

solve (3) & (4), we get

$A = 1.5873$ $\quad b \doteq 0.3609$

$a = \text{antilog } A$

$a = 38.6633$

Hence the required power curve is

$$y = 38.6633\, x^{0.3609}$$

5. Let the power curve be $y = ax^b \longrightarrow (1)$

The normal eqns of eq (1) are

$$An + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$A\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \longrightarrow (2)$$

~~solve (3) & (2)~~ where $n=3$

x	y	$X = \log x$	X^2	$Y = \log y$	XY
0	5.1	1	1	0.7075	0.7075
2	11	0.3010	0.0906	1.0413	0.3134
4	31.1	0.6020	0.3624	1.4927	0.8986
6	47.2	1.903	1.453	3.2415	1.9195

sub values in (2), we get

$$3A + 1.903 b = 3.2415 \longrightarrow (3)$$

$$1.903A + 1.453b = 1.9195 \longrightarrow (4)$$

solve ③ & ④ we get

$A = 1.4331 \qquad b = -0.5559$

$a = \text{antilog}\, A$

$a = 27.1081$

Hence the required power curve is

$$y = (27.1081)\, x^{(-0.5559)}$$

II

Q. Derive the normal eqns of a power curve $y = ab^x$

Let the eqn of powercurve be $y = ab^x \rightarrow ①$

Taking log on b.s

$$\log y = \log ab^x$$

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

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

$$Y = A + Bx \longrightarrow ②$$

where $Y = \log y \quad A = \log a, \ B = \log b$

i.e eq② represents the st.line in x and Y

The normal eqns of eq② are

$$An + B \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} Y_i$$

$$A \sum_{i=1}^{n} x_i + B \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i Y_i$$

The above two eqns are called the normal eqns of eq ① and solve these equations, we get A, B

where $a = \text{antilog } A$ (shift + log)
$b = \text{antilog } B$ "

Problems

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

1.

x	1	2	3	4	5	6
y	4	5.7	6.9	8	8.9	9.8

2.

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

3.

x	0	1	2	3	4	5
y	4	3	4.243	5.196	6	6.708

4.

x	0	1	2	3	4	5	6
y	10	21	35	59	92	200	400

5.

x	0	2	4
y	5.1	11	31.1

1. let the eqn of power curve be $y = ab^x \rightarrow (1)$

The normal eqns of eq (1) are

$$\left.\begin{array}{l} An + B\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} x_i + B\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i Y_i \end{array}\right\} \rightarrow (2)$$

where $n = 6$ $\quad Y = \log y$

The values of Σx, Σx^2, ΣY, ΣxY are calculated as follows

x	y	x^2	$Y = \log y$	xY
1	4	1	0.6020	0 1
2	5.7	4	0.7558	1.5116
3	6.9	9	0.8388	2.5164
4	8	16	0.9030	3.612
5	8.9	25	0.9493	4.7465
6	9.8	36	0.9912	5.9472
21		91	5.0401	18.9357

sub in (2), we have

$$6A + 21B = 5.0401 \rightarrow (3)$$
$$21A + 91B = 18.9357 \rightarrow (4)$$

Solve

$A = 0.5809 \quad B = 0.0740$

$a = 3.8097 \quad b = 1.1857$

Hence the required power curve is

$$y = 3.8097\,(1.1857)^x$$

2. let the eqn of power curve is $y = ax^b \rightarrow (1)$

The normal eqns are of (1) is

$$\left.\begin{array}{l} An + B\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ A\sum_{i=1}^{n} x_i + B\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{array}\right\} \rightarrow (2)$$

where $n = 6$

x	y	x^2	$y = \log y$	xy
1	151	1	2.1789	2.1789
2	100	4	2	4
3	61	9	1.7853	5.3559
4	50	16	1.6989	6.7956
5	20	25	1.3010	6.505
6	8	36	0.9030	5.418
Σx 21		Σx^2 91	Σy 9.8671	Σxy 30.2534

Now sub these in

$$6A + B21 = 9.8671 \longrightarrow (3)$$

$$21A + 91B = 30.2534 \longrightarrow (4)$$

solve (4) & (3)

$A = 2.5008 \qquad B = -0.2446$

$a = 316.8108 \qquad b = 0.5693$

$\therefore$ The eqn of power curve is

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

3. let the eqn of power curve be $y = ax^b \rightarrow (1)$

the normal eqns of eq (1) are

$$\left.\begin{aligned} An + B\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} x_i + B\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i Y_i \end{aligned}\right\} \rightarrow (2)$$

where $n = 6$

x	y	x^2	$Y = \log y$	xY
0	4	0	0.6020	0
1	3	1	0.4771	0.4771
2	4.243	4	0.6276	1.2552
3	5.196	9	0.7156	2.1468
4	6	16	0.7781	3.1124
5	6.708	25	0.8265	4.1325
15	29.147	$\Sigma x^2 = 55$	$\Sigma Y = 4.0269$	$\Sigma xY = 11.124$

sub in (2), we have

$6A + 15B = 4.0269 \rightarrow (3)$

$15A + 55B = 11.124 \rightarrow (4)$

solve (3) & (4) we have

$A = 0.5201 \qquad B = 0.0603$

$a = 3.3120 \qquad b = 1.1489$

$\therefore$ the eqn of power curve is

$$y = 3.3120\,(1.1489)^x$$

4. let the eqn of power curve be $y = ab^x \rightarrow (1)$

the normal eqn of eq (1) are

$$\left.\begin{aligned} An + B\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} Y_i \\ A\sum_{i=1}^{n} x_i + B\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i Y_i \end{aligned}\right\} \rightarrow (2)$$

where $n = 7$

the values of Σx, Σx, Σy, ΣY are calculated as

x	y	x^2	$Y = \log y$	xY
0	10	0	1	0
1	21	1	1.3222	1.3222
2	35	4	1.5440	3.088
3	59	9	1.7708	5.3124
4	92	16	1.9637	7.8548
5	200	25	2.3013	11.5065
6	400	36	2.6020	15.612
21		$\Sigma x^2 = 91$	$\Sigma Y = 12.5037$	44.6959

sub in (2) we get

$$7A + 21B = 12.5037 \rightarrow (3)$$

$$21A + 91B = 44.6959 \rightarrow (4)$$

solve (3) & (4), we get

$A = 1.0164$ $B = 0.2566$

$a = 10.3848$ $b = 1.8055$

$\therefore$ The eqn $y = (10.3848)(1.8055)^x$

5. let the eqn of power curve be $y = ab^x \rightarrow (1)$

the normal eqn of eq (1) are

$$An + B\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} Y_i$$

$$A\sum_{i=1}^{n} x_i + B\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i Y_i$$

where $n = 3$

The values of $\Sigma x, \Sigma y, \Sigma Y$, are calculated as

x	y	x^2	$Y = \log y$	xY
0	5.1	0	0.7075	0
2	11	4	1.0413	2.0826
4	31.1	16	1.4927	5.9708
6		20	3.2415	8.0534

~~wh~~ sub in (2), we have

$$3A + 6B = 3.2415 \longrightarrow (3)$$

$$6A + 20B = 8.0534 \longrightarrow (4)$$

solve (3) & (4), we get

$A = 0.6879$ $\quad B = 0.1963$

$a = 4.8741$ $\quad b = 1.5714$

$\therefore$ The eqn of power curve be

$$y = 4.8741(1.5714)^x$$

Correlation:

Correlation is a statistical analysis which measures the degree or extend to two variable fluctuate with reference to each other. One variable is called the subject and the other variable is called relative.

Methods of correlation:

1. karlpearsons coefficient of correlation
2. Spearman's rank correlation coefficient

Coefficient of correlation:

correlation is a statistical technic used for analysing the behaviour of two or more variables. Statistical measures of correlation related to covariation between series but not off function.

1. karl pearsons coefficient of correlation:

karl pearson and statistician suggested a mathematical method for measuring the magnitude of linear relationship b/w two variables. This method is called karl pearson coefficient of correlation. It is denoted by 'r'. This method is most widely used.

There are several formulae to the calculate 'r'. They are:

i) $r = \dfrac{\text{covariance of } x \cdot y}{\sigma_x \sigma_y}$

ii) $r = \dfrac{\Sigma xy}{N \sigma_x \sigma_y}$

iii) $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$,

where x, y are random variables,

Here $x = x - \bar{x}$, $\bar{x}$ and $\bar{y}$ be the means of series
$y = y - \bar{y}$ x and y

Σx = standard deviation of series x
Σy = standard deviation of series y

N = no. of values in series

we will be use formula iii), most widely

properties of correlation coefficient:

1. The coefficient of correlation lies b/w -1 and 1 symbolycally return as $-1 \le r \le 1$ (or) $|r| \le 1$.

2. The coefficient of correlation is independent of the change of origin and scale of measurements.

3. If X and Y are random variables and a, b, c, d are any numbers such that $a_0 = 0$ and $c_0 = 0$ then
$$r(ax+b, cy+d) = \frac{ac}{|a \cdot c|} r(x, y)$$

4. If two independent variables are uncorrelated variables then i.e if X and Y are independent variables then
$r(x, y) = 0$

Problems:

1. calculate the coefficient of correlation to the following data

X	12	9	8	10	11	13	7
Y	14	8	6	9	11	12	3

Sol: Given that $N=7$

Computation of coefficient of correlation

X	Y	X^2	Y^2	XY
12	14	144	196	168
9	8	81	64	72
8	6	64	36	48
10	9	100	81	90
11	11	121	121	121
13	12	169	144	156
7	3	49	9	21
70	63	728	651	676

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

$$r = \frac{676}{\sqrt{728 \cdot 651}}$$

$$r = \frac{676}{\sqrt{473928}}$$

$$r = \frac{676}{688.4242}$$

$$r = 0.9819$$

Hence the coefficient of correlation $r = 0.9819$

2. Find the coefficient of correlation between the heights and weights given below

Height(inch):	57	59	62	63	64	65	55	58	57
weight (lbs):	113	117	126	126	130	129	111	116	112

Sol: Given N = 9

computation of coefficient of correlation

let x = height and y = weight

Height X	deviation of series $x = X - \bar{x}$ 60	weight Y	deviation of series $y = Y - \bar{y}$ 120	x^2	y^2	xy
57	-3	113	-7	9	49	21
59	-1	117	-3	1	9	3
62	2	126	6	4	36	12
63	3	126	6	9	36	18
64	4	130	10	16	100	40
65	5	129	9	25	81	45
55	-5	111	-9	25	81	45
58	-2	116	-4	4	16	8
57	-3	112	-8	9	64	24
$\Sigma x = 540$		$\Sigma y = 1080$	Σ	102	472	216

$\frac{540}{10} = 54 = \bar{x}$ 108

Now $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$

$= \frac{216}{\sqrt{102 \cdot 472}} = \frac{216}{\sqrt{48144}}$

$= \frac{216}{219.4174}$

$= 0.9844$

Hence the correlation is $r = 0.9844$

3. Psycological tests of intelligents and engineering ability were applied to 10 students. Here is a record of ungrouped data showing intelligents ratio (I.R) and engineering ratio (E.R). calculate the coefficient of correlation.

Students	A	B	C	D	E	F	G	H	I	J
I.R	105	104	102	101	100	99	98	96	93	92
E.R	101	103	100	98	95	96	104	92	97	94

4. Find the Karlpearsons coefficient of correlation from the following data.

Wages	100	101	102	102	100	99	97	98	96	95
cost of living	98	99	99	97	95	92	95	94	90	91

5. calculate the coefficient of correlation between age of cars and anual maintainence cost are given below.

Age of cars (years)	2	4	6	8	10	12	14
Annual maintainence (cost)	1600	1500	1800	1900	1700	2100	2000

3. Given that $N = 10$

computation of coefficient of correlation

I.R (x)	E.R (y)	deviation of series $X = x - \bar{x}$	deviation of series $Y = y - \bar{y}$	X^2	Y^2	XY
105	101	6	3	36	9	18
104	103	5	5	25	25	25
102	100	3	2	9	4	6
101	98	2	0	4	0	0
100	95	1	-3	1	9	-3
99	96	0	-2	0	4	-2
98	104	-1	6	1	36	-6
96	92	-3	-6	9	36	18
93	97	-6	-1	36	1	6
92	94	-7	-4	49	16	28
990	980			170	140	90

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

$$= \frac{90}{\sqrt{170 \cdot 140}}$$

$$= \frac{90}{\sqrt{23800}}$$

$$= \frac{90}{154.2724}$$

$$= 0.58338$$

Hence the correlation is $r = 0.5833$

4. Given that $N = 10$

computation of coefficient of correlation

wages (X)	cost of living (Y)	deviation of series $x = X - \bar{x}$	deviation of series $y = Y - \bar{y}$	x^2	y^2	xy
100	98	1	3	1	9	3
101	99	2	4	4	16	8
102	99	3	4	9	16	12
102	97	3	2	9	4	6
100	95	1	0	1	0	0
99	92	0	-3	0	9	0
97	95	-2	0	4	0	0
98	94	-1	-1	1	1	1
96	90	-3	-5	9	25	15
95	91	-4	-4	16	16	16
990	950			54	96	61

Given that $N = 10$

$$\text{Now } r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{61}{\sqrt{54 \cdot 96}}$$

$$= \frac{61}{\sqrt{5184}} = \frac{61}{72}$$

$$= 0.8472$$

Hence the correlation is $r = 0.8472$

5. Given that $N = 7$

computation of coefficient of correlation

Age of cars (x)	Annual maintanence (y)	deviation of series $x = X - \bar{x}$	deviation of series $y = Y - \bar{y}$	x^2	y^2	xy
2	1600	−6	−200	36	40000	1200
4	1500	−4	−300	16	90000	1200
6	1800	−2	0000	4	0	0
8	1900	0	100	0	10000	0
10	1700	2	−100	4	10000	−200
12	2100	4	300	16	30000	1200
14	2000	6	200	36	40000	1200
56	12600			112	220000	4600

Now $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$

$= \dfrac{4600}{\sqrt{112 \cdot 220000}}$

$= \dfrac{4600}{\sqrt{24640000}}$

$= \dfrac{4600}{4963.8694}$

$= 0.9266$

Hence the correlation is $r = 0.9266$

When deviations are taken from assumed mean:

When the actual mean is not a whole number but a fraction. when the series is large, the calculation by direct method solving a lot of time. To avoid such calculation by using assumed mean method.

i.e $$r = \frac{(\Sigma xy \cdot N) - (\Sigma x \cdot \Sigma y)}{\sqrt{\Sigma x^2 \cdot N - (\Sigma x)^2}\sqrt{(\Sigma y^2 \cdot N - (\Sigma y)^2)}}$$

where N = no. of items

x = deviations of the items of x series from assumed mean

y = deviations of the items of y series from assumed mean.

Σx = The total deviation of x series from assumed mean

Σy = The total deviation of y series from assumed mean.

Σxy = The total of product of the deviation of x and y series from assumed mean.

$\Sigma x \cdot \Sigma y$ = The total of the deviations of x and y series from assumed mean

Σx^2 = The total of square of the deviation of x series from assumed mean

Σy^2 = The total of square of deviation of y series from assumed mean.

1. Calculate the karlpearsons coefficient of correlation from the following data

x	28	41	40	38	35	33	40	32	36	33
y	23	34	33	34	30	26	28	31	36	38

Sol: Given that $N = 10$

computation of coefficient of correlation

X	deviation of x series $X = x - A_{35}$	X^2	y	deviation $Y = y - A$ (31) 3	Y^2	XY
28	-7	49	23	-8	64	56
41	6	36	34	3	9	18
40	5	25	33	2	4	10
38	3	9	34	3	9	9
35	0	0	30	-1	1	0
33	-2	4	26	-5	25	10
40	5	25	28	-3	9	-15
32	-3	9	31	0	0	0
36	1	1	36	5	25	5
33	-2	4	38	7	49	-14
356	6	162	313	3	195	79

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

Scanned with CamScanner

$$r = \frac{79 \times 10 - 6 \times 3}{\sqrt{162 \times 10 - 36}\sqrt{313 \times 10 - 9}}$$

$$r = \frac{790 - 18}{\sqrt{1620 - 36}\sqrt{3130 - 9}}$$

$$r = \frac{772}{39.7994 \times 55.8667}$$

$$= \frac{772}{1753.4293}$$

$$r = 0.4402$$

Hence the coefficient of correlation $r = 0.4402$

2. Find a coeffecient of correlation from the following data

fertilizers	15	18	20	24	30	35	40	50
Productivity	85	93	95	105	120	130	150	160

Given that $N = 8$

let x = fertilizers

y = productivity

x	$X = x - \bar{x}$	X^2	y	$Y = y - \bar{y}$	Y^2	XY
15	-14	196	85	-32	1024	448
18	-11	121	93	-24	576	264
20	-9	81	95	-22	484	198
24	-5	25	105	-12	144	60
30	1	1	120	3	9	3
35	6	36	130	13	169	78
40	11	121	150	33	1089	363
50	21	441	160	43	1849	903
232	0	1022	938	2	5344	2317

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

$$r = \frac{2317 \cdot 8 - 0}{\sqrt{1022 \cdot 8 - 0}\sqrt{5344 \cdot 8 - 4}}$$

$$r = \frac{18536}{90.4212 \times 206.75589}$$

$$r = \frac{18536}{18695.11568}$$

$$r = 0.9914$$

Hence the coefficient of correlation $r = 0.9914$

Spearman's rank correlation coefficient

The formula for spearman's rank correlation coefficient is given by

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

where N = no. of paired observations

ΣD^2 = total sum of squares of the differences of two ranks

ρ = rank correlation coefficient.

Properties of rank correlation coefficient:

→ The values of row lies between −1 and +1

→ If $\rho = 1$, then there is a complete agrement in the order if they ranks and the direction of rank is same.

→ If $\rho = -1$ then there is a complete disagrement in the order of the ranks and they are oppisite directions.

Procedure for find rank correlation coefficient:

When the ranks are given:

→ compute the differences of two ranks and it is denoted by D.

→ Squaring D and get ΣD^2

→ obtained ρ and substituting the values in the given formula.

When the ranks are not given:

→ when the actual data given then we must give ranks. we can give ranks by taking the highest values as '1' or lowest value as '1', next to the highest value as 2, lowest value as 2 and follow the same procedure both variables, we get rank.

Problems

1. The following are the ranks obtained by 10 students in 2 subjects as statistics and mathematics. To what extent the knowledge of the students in two subjects is related?

statistics	1	2	3	4	5	6	7	8	9	10
Mathematics	2	4	1	5	3	9	7	10	6	8

Sol: Here $N=10$

x = Statistic

y = Mathematics

Computation of rank correlation coefficient

x	y	$D=x-y$	D^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4

By using spearman rank correlation coefficient

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

$$= 1 - \frac{6(40)}{10(100-1)}$$

$$= 1 - \frac{240}{10(99)}$$

$$= 1 - \frac{24\cancel{0}}{99\cancel{0}}$$

$$= 1 - 0.2424$$

$$\rho = 0.7576$$

$\therefore$ The rank correlation coefficient $= 0.7576$.

2. A random sample of five college students is selected and their grades in mathematics and statistics are given below.

Students	1	2	3	4	5
Mathematics	85	60	73	40	90
Statistics	93	75	65	50	80

calculate the spearmans rank correlation coefficient.

Sol: Given $N = 5$

X = Mathematics grades

Y = statistics grades

x = rank in mathematics

y = rank in statistics

computation of rank correlation coefficient

X	Y	x	y	$D=x-y$	D^2
85	93	2	1	1	1
60	75	4	3	1	1
73	65	3	4	-1	1
40	50	5	5	0	0
90	80	1	2	-1	1
					$\Sigma D^2 = 4$

By using spearman rank correlation coefficient

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

$$= 1 - \frac{6(4)}{5(25-1)}$$

$$= 1 - \frac{24}{5(24)}$$

$$= 1 - 0.2$$

$$= 0.8$$

Hence $\rho = 0.8$

3. Ten computators in a musical test were ranked by three judges A, B and C from the following data.

Ranks by A	1	6	5	10	3	2	4	9	7	8
Ranks by B	3	5	8	4	7	10	2	1	6	9
Ranks by C	6	4	9	8	1	2	3	10	5	7

Using Spearman's rank correlation coefficient, discuss which pair of judges has the nearest approach to common liking in music.

Sol: Given that $N = 10$

let x = Ranks by A

y = Ranks by B

z = Ranks by C

computation of rank correlation coefficient

x	y	z	$D_1 = x-y$	$D_2 = y-z$	$D_3 = z-x$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-3	5	4	9	25
6	5	4	1	1	-2	1	1	4
5	8	9	-3	-1	4	9	1	16
10	4	8	6	-4	-2	36	16	4
3	7	1	-4	6	-2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	+1	-1	4	1	1
9	1	10	8	-9	1	64	81	1
7	6	5	1	1	-2	1	1	4
8	9	7	-1	2	-1	1	4	1
						200	214	60

$$\rho_{1\,(x,y)} = 1 - \frac{6\Sigma D_1^2}{N(N^2-1)} = 1 - \frac{6(200)}{10(100-1)} = 1 - \frac{120\cancel{0}}{99\cancel{0}}$$

$= 1 - 1.2121$

$= -0.2121$

$$\rho_{2\,(y,z)} = 1 - \frac{6\,\Sigma D_2^2}{N(N^2-1)}$$

$$= 1 - \frac{6(214)}{10(100-1)}$$

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

$$= 1 - 1.2969$$

$$= -0.2969$$

$$\rho_{3\,(z,x)} = 1 - \frac{6\,\Sigma D_3^2}{N(N^2-1)}$$

$$= 1 - \frac{6(60)}{990}$$

$$= 1 - \frac{360}{990}$$

$$= 1 - 0.3636$$

$$= 0.6364$$

Since $\rho_3(z,x)$ is maximum, we conclude that the pair of judges A and C as the nearest approach to common likings in music.

Equal or Repeated ranks

If any two or more persons are equal in any classification. If there is more than one item, with the same value in the series then the spearmen's formula for calculating the rank correlation coefficient breaks down. In this case common tasks ranks are given to repeated items. The common ranks is the average of ranks which these items would have assumed if they were different from each other and the next item will get the rank next two ranks already assumed. We use the following formula.

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

where N = no. of paired observations.

ΣD^2 = Sum of squares of difference of two ranks.

Problems

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

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

Sol: Here $N=10$

let x = ranks in X

y = ranks in Y

computation of rank correlation coefficient. we do it by following table

X	Y	x = Ranks in X	y = Ranks in Y	$D = x-y$	D^2
48	13	3	5.5	−2.5	6.25
33	13	5	5.5	−0.5	0.25
40	24	4	1	3	9
9	6	10	8.5	1.5	2.25
16	15	8	4	4	16
16	4	8	10	−2	4
65	20	1	2	−1	1
24	9	6	7	−1	1
16	6	8	8.5	−0.5	0.25
57	19	2	3	−1	1
					$\Sigma D^2 = 41$

Since 16 is repeated 3 times, $m_1 = 3$

13 is repeated 2 times, $m_2 = 2$

6 is repeated 2 times, $m_3 = 2$

By using rank correlation coeffecient we have

$$\rho = 1-6\frac{\left[\Sigma D^2+\frac{1}{12}(m_1^3-m_1)+\frac{1}{12}(m_2^3-m_2)+\frac{1}{12}(m_3^3-m_3)\right]}{N(N^2-1)}$$

$$= 1-6\frac{\left[41+\frac{1}{12}(27-3)+\frac{1}{12}(8-2)+\frac{1}{12}(8-2)\right]}{10(100-1)}$$

$$= 1-6\left[\frac{41+2+0.5+0.5}{10\times 99}\right]$$

$$= 1-6\left[\frac{44}{990}\right]$$

$$= 1-0.2666$$

$$= 0.7334$$

2. obtain a rank correlation coefficient from 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: Here $N=10$

let x = ranks in x

y = ranks in y

computation of rank correlation coefficient

X	Y	x	y	D = x - 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

$\Sigma D^2 = 72$

Here 75 is repeated 2 times $m_1 = 2$
64 is repeated 3 times $m_2 = 3$
68 is repeated 2 times $m_3 = 2$

By using rank correlation coefficient, we have

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3)\right]}{N(N^2-1)}$$

$$= 1 - \frac{6\left[72 + \frac{1}{12}(8-2) + \frac{1}{12}(27-3) + \frac{1}{12}(8-2)\right]}{10 \times 99}$$

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

$$= 1 - 6\left[\frac{75}{990}\right]$$

$$= 1 - 0.4545$$

$$\rho = 0.5455$$

3. A sample of 12 fathers and there elders sons gave the following data. calculate the rank correlation coefficient.

fathers	65	63	67	64	68	62	70	66	68	67	69	71
sons	68	66	68	65	69	66	68	65	71	67	68	70

sol: let X = fathers

Y = eldersons

Here $N = 12$

let x = ranks in X (or)

y = ranks in Y.

computation of rank correction coefficient

X	Y	x	y	D=x-y	D^2
65	68	9	8.5	3.5	12.25
63	66	1	9.5	1.5	2.25
67	68	6.5	8.5		·1
64	65	10	11.5	-1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	-3.5	12.25
66	65	8	11.5	-3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	-1.5	2.25
69	68	3	5.5	-2.5	6.25
71	70	1	2	-1	

$\Sigma D^2 = 60.25$

Now Here

68 is repeated 2 times $m_1 = 2$

67 is reapeated 2 times $m_2 = 2$

68 is repeated 4 times $m_3 = 4$

66 is repeated 2 times $m_4 = 2$

Now rank correlation coefficient is

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{12}(m_3^3 - m_3) + \frac{1}{12}(m_4^3 - m_4)\right]}{N(N^2-1)}$$

$$= 1 - \frac{6\left[60.25 + \frac{1}{12}(8-2) + \frac{1}{12}(8-2) + \frac{1}{12}(64-4) + \frac{1}{12}(8-2)\right]}{12(143)}$$

$$= 1 - \frac{6\left[60.25 + 0.5 + 0.5 + 5 + 0.5\right]}{12 \times 143}$$

$$= 1 - \frac{6(66.75)}{1716}$$

$$= 1 - \frac{400.5}{1716}$$

$$= 1 - 0.2333$$

$$\rho = 0.7667$$

Regression:

Uses:

1. It is used to estimate the relation between two economic variables like as income and expenditure.
2. It is a highly valuable tool in economics and bussiness

3. It is most widely used for prediction purpose
4. We can calculate coefficient of correlation and coefficient of retermination with the help of their regression coefficient.
5. It is useful in statistical estimation of demand curves, suplier curves, production function, cost function, consumption function etc.

Comparision between correlation and regression:

The correlation coefficient is a measure of degree of covariability between two variables, while the regression establish a functional relation between independent variables, so that the former can be predicted for a given value of the later. In correlation, both the variables X and Y are random variables, where as in regression x is a random variable and Y is a fixed variable.

The coefficient of correlation is a relative measure where as regression coefficient is an absolute figure.

Regression lines:

A regression line is a straight line fitted to the data by the method of least squares. It indicates the best possible mean value of one variable, corresponding to the mean value of the other variable.

There are always 2 regression lines constructed for the relationship between two variables X and Y. Thus one regression line shows a regression of X upon Y and there other shows a regression of Y upon X

Regression Equations

Regression Equation is an algebric expression of straightline. It can be classified into regression equation, regression coefficient, individual observation and group discussion.

The standard form of the regression equation is equal to

$$Y = a + bx$$

where a, b are called constants; a is indicated the value of Y and when $x = 0$; it is called Y-intercept b indicates the value of the slope of the regression coefficient of Y on x. Thus if we compute know the value of a and b we can easily compute the value of y for any given value of x. The values of a and b are found with the help of the following normal equations.

Regression Equation of Y on x:

The regression equation of Y on x is

$$Y = a + bx$$

The normal equations of above equations are

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

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

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

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

Regression equation of X on Y:

The regression equation of X on Y is

$$X = a + bY$$

The normal equations of above equations are

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

$$a\Sigma Y + b\Sigma Y^2 = \Sigma XY$$

Problems

1. Determine a regression equation of a st-line of y on x, from the following data.

x	10	12	13	16	17	20	25
y	10	22	24	27	29	33	37

Sol: The regression equation of y on x is

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

The normal equations of eq (1) are

$$\left.\begin{aligned} aN + b\Sigma x &= \Sigma y \\ a\Sigma x + b\Sigma x^2 &= \Sigma xy \end{aligned}\right\} \longrightarrow (2)$$

Here $N = 7$

The values of $\Sigma x, \Sigma y, \Sigma x^2, \Sigma xy$ can be calculated as follows:

x	y	x^2	xy
10	10	100	100
12	22	144	264
13	24	169	312
16	27	256	432
17	29	289	493
20	33	400	660
25	37	625	925
$\Sigma x = 113$	$\Sigma y = 182$	$\Sigma y^2 = 1983$	$\Sigma xy = 3186$

Now sub $\Sigma x, \Sigma y, \Sigma x^2, \Sigma xy$ in (2), we get

$$7a + 113b = 182 \longrightarrow (3)$$

$$113a + 1983b = 3186 \longrightarrow (4)$$

solving (3) & (4) we get

$a = 0.7985$
$b = 1.5611$

∴ The required regression eq of st-lines of y on x is

$$y = a + bx$$
$$y = 0.7985 + 1.5611x$$

Deviation arithmetic mean of x on y and y on x:

This method is easier and simpler then the previous to find the values of A & B. we can find the deviations of x & y series from the respective means.
The Regression eq'n of x on y is

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

Where, $\bar{x}$ = mean of x series and
$\bar{y}$ = mean of y series

Regression coefficient of x on y is

$$b_{xy} = r \frac{\sigma x}{\sigma y} = \frac{\sigma xy}{\Sigma y^2}$$

The Regression coefficient of y on x is

$$y - \bar{y} = r \frac{\sigma y}{\sigma x} \cdot (x - \bar{x})$$

where, $\bar{x}$ = mean of x series and
$\bar{y}$ = mean of y series

The Regression coefficient of y on x is

$$b_{yx} = r \frac{\sigma y}{\sigma x} = \frac{\Sigma xy}{\Sigma x^2}$$

Thus, $r^2 = b_{xy} \cdot b_{yx}$

Problems

1\. A panel of two judges P & Q graded 7 dramatic performance by independently according marks as follows.

performance	1	2	3	4	5	6	7
marks by P	46	42	44	40	43	41	45
marks by Q	40	38	36	35	39	37	41

The 8 performance, which judges Q could not attend, was awarded 37 moves by Judge P.
If judge Q had also be present, how many marks would be expected to have been awarded by him to the 8 performance.

Sol: let x = marks by P
y = marks by Q

Computation of Regression coefficient.

Performance	x	$X = x-\bar{x}$	X^2	y	$Y = y-\bar{y}$	Y^2	XY
1	46	3	9	40	2	4	6
2	42	-1	1	38	0	0	0
3	44	1	1	36	-2	4	-2
4	40	-3	9	35	-3	9	9
5	43	0	0	39	1	1	0
6	41	-2	4	37	-1	1	2
7	45	2	4	41	3	9	6
	301	0	28	266	0	28	21

Now, the regression coefficient are

$$Y - \bar{Y} = r \frac{\sigma Y}{\sigma X} (X - \bar{X}) \quad \text{(or)}$$

$$Y - \bar{Y} = \frac{\Sigma XY}{\Sigma X^2} (X - \bar{X})$$

$$Y - 38 = \frac{21}{28} (X - 43)$$

$$Y - 38 = \frac{21}{28} (X - 43)$$

$$Y - 38 = 0.75 (X - 43)$$

$$Y = 0.75X - 0.75(43) + 38$$

$$Y = 0.75X - 32.25 + 38$$

$$Y = 0.75X - 5.75$$

But given judge be awarded 37 marks

i.e $X = 37$

$$\therefore Y = (0.75)37 + 5.75$$

$$Y = 33.5 \text{ marks}$$

Hence, if B to be present, he would be have 33.5 marks to the 8^{th} performance.

2. Find the most likely production corresponds to a rainfall 40 from the following data:

	Rainfall (x)	Production (y)
Average	30	500 kgs
S.D	5	100 kgs

correlation coefficient $r = 0.8$

Sol: Given, $\bar{X} = 30$, $\bar{Y} = 500$

$\Sigma X = 5$, $\Sigma Y = 100$ and also given

$r = 0.8$

The regression eq'n of Y on X is

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

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

$$Y-500 = 16(X-30)$$

$$Y-500 = 16X-480$$

$$Y = 16X-480+500$$

$$Y = 16X+20$$

Here, $X=40$

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

$$Y = 60$$

3. From a sample of "200 of observation the following quantities were calculated.

$\Sigma X = 11.34$, $\Sigma Y = 20.78$, $\Sigma X^2 = 12.16$, $\Sigma Y^2 = 84.96$, $\Sigma XY = 22.13$

From the above data, find how to compare the coefficient of the eq'n $Y = a+bX$

Sol: Given, $N=200$

$\Sigma X = 11.34$, $\Sigma Y = 20.78$, $\Sigma X^2 = 12.16$, $\Sigma Y^2 = 84.96$, $\Sigma XY = 22.13$ and also given that $Y = a+bX \longrightarrow (1)$

The normal eqn's of eq (1) are

$$\left.\begin{array}{l} aN + b\Sigma X = \Sigma Y \\ a\Sigma X + b\Sigma X^2 = \Sigma XY \end{array}\right\} \longrightarrow (2)$$

sub values in (2), we get

$$200a + (11.34)b = 20.78$$

$$(11.34)a + (12.16)b = 22.13$$

$$a = 7.5132 \times 10^{-4} = 0.0007$$

$$b = 1.8192$$

Hence, the compare of coefficient of eq'n

$y = (0.0007) + (1.8192)b.$

Angle between two regressive lines

Let the regressive eqn's of x on y and y on x is given by $(x-\bar{x}) = r.\frac{\sigma x}{\sigma y}(y-\bar{y})$

$$(y-\bar{y}) = r\cdot\frac{\sigma y}{\sigma x}(x-\bar{x})$$

Slope of the line $m_1 = \frac{1}{r}\cdot\frac{\sigma y}{\sigma x}$,

slope of the line $m_2 = r\cdot\frac{\sigma y}{\sigma x}$

The Angle b/w to regression lines is

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

$$= \frac{\frac{1}{r}\cdot\frac{\sigma y}{\sigma x} - r\cdot\frac{\sigma y}{\sigma x}}{1 + \left(\frac{1}{r}\cdot\frac{\sigma y}{\sigma x}\right)\left(r\cdot\frac{\sigma y}{\sigma x}\right)}$$

$$= \frac{\frac{\sigma y}{\sigma x}\left(\frac{1}{r} - r\right)}{1 + \left(\frac{\sigma y}{\sigma x}\right)^2}$$

$$= \frac{\frac{\sigma y}{\sigma x}\left(\frac{1-r^2}{r}\right)}{1 + \left(\frac{\sigma y}{\sigma x}\right)^2}$$

Note:

1. If $r=0$, then $\tan\theta = \infty \Rightarrow \theta = \pi/2$

2. If $r=\pm 1$ then $\tan\theta = 0 \Rightarrow \theta = 0$ or π

1. If θ is the angle between two regression lines and standard deviation of lines is twice the standard deviation of lines and $r = 0.25$, find $\tan\theta$

Sol: Given that, $r = 0.25$ and $\sigma y = 2 \times \sigma x$

The angle b/w two regression line is

$$\tan\theta = \frac{\frac{\sigma y}{\sigma x}\left(\frac{1-r^2}{r}\right)}{1+\left(\frac{\sigma y}{\sigma x}\right)^2}$$

$$= \frac{2 \cdot \frac{\sigma x}{\sigma y}\left(\frac{1-(0.25)^2}{0.25}\right)}{1+\left(2 \frac{\sigma x}{\sigma x}\right)^2}$$

$$= \frac{2(3.75)}{5}$$

$$\therefore \tan\theta = 1.5$$

2. If $\sigma x = \sigma y = \sigma$ and the angle b/w the regression lines is $\tan^{-1}\left(\frac{4}{3}\right)$. Find r

Sol: Given that, $\sigma x = \sigma y = 6$,

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

The angle b/w two regression lines is

$$\tan\theta = \frac{\frac{\sigma y}{\sigma x}\left(\frac{1-r^2}{r}\right)}{1+\left(\frac{\sigma y}{\sigma x}\right)^2}$$

$$= \frac{\frac{\cancel{\sigma}}{\cancel{\sigma}}\left(\frac{1-r^2}{r}\right)}{1+\left(\frac{\cancel{\sigma}}{\cancel{\sigma}}\right)^2}$$

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

$$\tan\theta = \frac{1-r^2}{2r}$$

i.e $\theta = \tan^{-1}\left(\frac{1-r^2}{2r}\right)$

but given, $\theta = \tan^{-1}\left(\frac{4}{3}\right)$

$$\tan^{-1}\left(\frac{4}{3}\right) = \tan^{-1}\left(\frac{1-r^2}{2r}\right)$$

$$\frac{4}{3} = \frac{1-r^2}{2r}$$

$$8r = 3-3r^2$$

$$3r^2+8r-3=0$$

$$3r^2+9r-r-3=0$$

$$3r(r+3)-1(r+3)=0$$

$$(3r-1)(r+3)=0$$

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

$\because -1 < 1$ so, we don't take $r=-3$

Hence, $r=\frac{1}{3}$

Multiple Regression:

In the study of simple regression b/w two variables x & y, we find the relative movement of x series for a unit movement of y-series and the relative movement of y-series for a unit movement of x-series. we get two regression equations one of x on y and the other of y on x.

The multiple regression analysis represents when extension of two variables to 3 or more variable

In this, instead of one dependent variable and two or more independent variables are used to estimate the values of dependent variable.

In this analysis, the effect of 2 or more independent variables on one dependent variables is studied.

Regression Equations:

A regression equation is an equation for estimating a dependent variable say X_1 from the independent variables X_2, X_3 is called a regression eq'n of X_1 on X_2 & X_3.

The procedure for studying multiple regression is similar to the one we have for simple regression, with the difference that the another variables are added in the regression eq'n.

If there are 3 variables X_1, X_2, X_3, the multiple regression as says following form.

$$X_1 = a_{1.2.3} + b_{12.3} X_2 + b_{13.2} X_3$$

In the above eq'n there are 3 constants are $a_{1.23}$, $b_{12.3}$ & $b_{13.2}$

The subscript after the indicates the variables which are held constant.

If X_2 & X_3 on '0' then $X = a_{1.2}$

In this above eq'n, $b_{12.3}$ indicates the slope of a regression line of X_1, X_2 when X_3 is held constant.

Similarly, $b_{13.2}$ indicates the slope of the regression eqn's of X_1 on X_3 when X_2 is held constant.

Interpretation of constants

1. $a_{1.23}$

The constant $a_{1.23}$ is they intersept made by the regression plane. It gives the value of dependent variable when all the independent variables are zero.

2. $b_{12\cdot3}$:

The constant $b_{12\cdot3}$ measures the amount by which a unit change in x_2 is expected to effect x, when x_3 is held constant.

3. $b_{13\cdot2}$:

The constant $b_{13\cdot2}$ measures the amount by which a unit change in x_3 is expected to effect x_1 when x_2 is kept constant.

Because of the fact x_1 various partial give to the variations in x_2 and partially give to the variations in x_3.

we call $b_{12\cdot3}$ & $b_{13\cdot2}$

Normal eq'ns for multiple regression equations:

W.k.T in simple regression there is a regression line approximating a set of n data points (x & y).

Similarly, there exists a least square regression line fitting a set of N data points (x_1, x_2 & x_3)

a. The regression eq'n on x_1 on x_2 x_3 is

$$x_1 = a_{1\cdot23} + b_{12\cdot3} x_2 + b_{13\cdot2} x_3 \longrightarrow (1)$$

The normal eqn's of eq (1) are

$$Na_{1\cdot23} + b_{12\cdot3}\sum_{i=1}^{n} x_2 + b_{13\cdot2}\sum_{i=1}^{n} x_3 = \sum_{i=1}^{n} x_1$$

$$a_{1\cdot23}\sum_{i=1}^{n} x_2 + b_{12\cdot3}\sum_{i=1}^{n} x_2^2 + b_{13\cdot2}\sum_{i=1}^{n} x_2 x_3 = \sum_{i=1}^{n} x_1 x_2$$

$$a_{1\cdot23}\sum_{i=1}^{n} x_3 + b_{12\cdot3}\sum_{i=1}^{n} x_2 x_3 + b_{13\cdot2}\sum_{i=1}^{n} x_3^2 = \sum_{i=1}^{n} x_1 x_3$$

The regression eq'n of x_2 on x_1 & x_3 is

$$x_2 = a_{2.13} + b_{21.3} x_1 + b_{23.1} x_3 \longrightarrow (2)$$

The normal eqn's of eq (2) are

$$N a_{2.13} + b_{21.3} \sum_{i=1}^{n} x_1 + b_{23.1} \sum_{i=1}^{n} x_3 = \sum_{i=1}^{n} x_2$$

$$a_{2.13} \sum_{i=1}^{n} x_1 + b_{21.3} \sum_{i=1}^{n} x_1^2 + b_{23.1} \sum_{i=1}^{n} x_1 x_3 = \sum_{i=1}^{n} x_1 x_2$$

$$a_{2.13} \sum_{i=1}^{n} x_3 + b_{21.3} \sum_{i=1}^{n} x_1 x_3 + b_{23.1} \sum_{i=1}^{n} x_3^2 = \sum_{i=1}^{n} x_2 x_3$$

The regression eqn's of x_3 on x_1 & x_2 is

$$x_3 = a_{3.12} + b_{31.2} + b_{32.1} x^2 \longrightarrow (3)$$

The normal eqns of eq (3) are

$$N a_{3.12} + b_{31.2} \sum_{i=1}^{n} x_1 + \sum_{i=1}^{n} x_2 = \sum_{i=1}^{n} x_3$$

$$a_{3.12} \sum_{i=1}^{n} x_1 + b_{31.2} \sum_{i=1}^{n} x_1^2 + b_{32.1} \sum_{i=1}^{n} x_1 x_2 = \sum_{i=1}^{n} x_1 x_3$$

$$a_{3.12} \sum_{i=1}^{n} x_2 + b_{31.2} \sum_{i=1}^{n} x_1 x_2 + b_{32.1} \sum_{i=1}^{n} x_2^2 = \sum_{i=1}^{n} x_1 x_2$$

In generalizing the regression eqn for more than 3 variables is

$$x_1 = a_{1.234} + b_{12.34} x_2 + b_{13.24} x_3 + b_{14.23} x_4$$

Problems

1. Find the multiple regression eq'n of x_1 on x_2 & x_3 from the following data:

x_1	2	4	6	8
x_2	3	5	7	9
x_3	4	6	8	10

2. Find the multiple regression eqn of x_1 on x_2 & x_3 from the following data

x_1	11	17	26	28	31	35	41	49	63	69
x_2	2	4	5	6	7	8	10	11	13	14
x_3	2	3	4	5	6	7	9	10	11	13

Sol: Here N=4.

∴ The regression eqn of x_1 on x_2 & x_3 are

$$x_1 = a_{1.23} + b_{12.3}x_2 + b_{13.2}x_3 \longrightarrow (1)$$

The normal eq'ns of eq(1) are

$$Na_{1.23} + b_{12.3}\sum_{i=1}^{n} x_1 + b_{13.2}x_3 \sum_{i=1}^{n} = \sum_{i=1}^{n} x_1$$

$$a_{1.23}\sum_{i=1}^{n} x_2 + b_{12.3}\sum_{i=1}^{n} x_2^2 + b_{13.2}\sum_{i=1}^{n} x_2x_3 = \sum_{i=1}^{n} x_ix_2 \longrightarrow (2)$$

$$a_{1.13}\sum_{i=1}^{n} x_3 + b_{12.3}\sum_{i=1}^{n} x_2x_3 + b_{13.2}\sum_{i=1}^{n} x_3^2 = \sum_{i=1}^{n} x_1x_3$$

The values of $\Sigma x_1, \Sigma x_2, \Sigma x_3, \Sigma x_1^2, \Sigma x_3^2, \Sigma x_1x_2; \Sigma x_1x_3, \Sigma x_3x_2$ are calculated as follows:

x_1	x_2	x_3	x_2^2	x_3^2	x_1x_2	x_2x_3	x_1x_3
2	3	4	9	16	6	12	8
4	5	6	25	36	20	30	24
6	7	8	49	64	42	56	48
8	9	10	81	100	72	90	80
20	24	28	164	216	140	188	160

From (2) sub all these value

$$4a + 24b + 28b = 20 \longrightarrow (3)$$

$$24a + 164b + 188b = 140 \longrightarrow (4)$$

$$28a + 188b + 216b = 160 \longrightarrow (5)$$

solving ③, ④ & ⑤ we get

$a_{1.23} = 0$, $b_{12.3} = 2$, $b_{13.2} = -1$

∴ Hence the required regression eqn of x_1 on x_2 & x_3

$$x_1 = 0 + 2x_2 - x_3$$

$$x_1 = 2x_2 - x_3$$

2(10): Here $N = 10$

Since the regression equation of x_1 on x_2 & x_3 is

$$x_1 = a_{1.23} + b_{12.3}x_2 + c_{13.2}x_3 \longrightarrow ①$$

The normal equations of eq ① are

$$\left.\begin{aligned} N a_{1.23} + b_{12.3}\sum_{i=1}^{n} x_2 + b_{13.2}\sum_{i=1}^{n} x_3 &= \sum_{i=1}^{n} x_1 \\ a_{1.2.3}\sum_{i=1}^{n} x_2 + b_{12.3}\sum_{i=1}^{n} x_2^2 + b_{13.2}\sum_{i=1}^{n} x_2 x_3 &= \sum_{i=1}^{n} x_1 x_2 \\ a_{1.23}\sum_{i=1}^{n} x_3 + b_{12.3}\sum_{i=1}^{n} x_2 x_3 + b_{13.2}\sum_{i=1}^{n} x_3^2 &= \sum_{i=1}^{n} x_1 x_3 \end{aligned}\right\} \text{--- ②}$$

The values $\Sigma x_1, \Sigma x_2, \Sigma x_3, \Sigma x_2^2, \Sigma x_3^2, \Sigma x_1 x_2, \Sigma x_2 x_3, \Sigma x_1 x_3$ can be calculated as follows

X_1	X_2	X_3	X_2^2	X_3^2	X_1X_2	X_2X_3	X_1X_3
11	2	2	4	4	22	4	22
17	4	3	16	9	68	12	51
26	5	4	25	16	130	20	104
28	6	5	36	25	168	30	140
31	7	6	49	36	217	42	186
35	8	7	64	49	280	56	245
41	10	9	100	81	410	90	369
49	11	10	121	100	539	110	490
63	13	11	169	121	819	143	693
69	14	13	196	169	966	182	897
370	80	70	780	610	3619	689	3197

Now substitute these values in ② we get

$10a_{1.23} + 80\, b_{12.3} + 70\, b_{13.2} = 370 \longrightarrow$ ③

$80a_{1.23} + 780\, b_{12.3} + 689\, b_{13.2} = 3619 \longrightarrow$ ④

$70\, a_{1.23} + 689\, b_{12.3} + 610\, b_{13.2} = 3197 \longrightarrow$ ⑤

solving ③, ④ & ⑤ we get

$a_{1.23} = -0.7295$

$b_{12.3} = 4.8867$

$b_{13.2} = -0.1949.$

Now sub these in ①, we get

$x_1 = -0.7295 + 4.8867\, x_2 (-0.1949)\, x_3$