

Unit - 6

Curve Fitting and Regression

- 1) Curve Fitting by the method of Least Squares.
- 2) Fitting of Straight Lines.
- 3) Fitting of Second degree curves or parabolas.
- 4) More general curves.
- 5) Types of Regression.

$\Rightarrow$ We can know value of unknown value.

Types of Regression

$\Rightarrow$

- 1) Linear Regression
- 2) Multiple Regression

$\Rightarrow$ Curve fitting by the method of Least Squares.

{Let the set of data point for (x_i, y_i) for $i = 1, 2, \dots$

Suppose the curve $y = f(x)$ is fitted with this data.

Let the observed value at $x = x_i$ is y_i and the corresponding value on the curve is $f(x_i)$.

Let e_i is the error of approximation at $x = x_i$, then we have $e_i = y_i - f(x_i)$.

$$\text{Let } S = [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \dots + [y_n - f(x_n)]^2$$

$$\text{where } S = e_1^2 + e_2^2 + \dots + e_n^2$$

The method of Least Squares consists of minimizing S (reducing the error) }

Fitting a Straight Line

Let $y = a_0 + a_1 x$ to be fitted with the given data where a_0, a_1 are constants.

The Corresponding Normal equations are,

$$\sum y_i = a_0 \sum 1 + a_1 \sum x_i$$

$$\sum y_i = a_0 N + a_1 \sum x_i \quad \text{--- (1)}$$

$$\text{and } \sum x_i y_i = a_0 \sum x_i + a_1 \sum x_i^2 \quad \text{--- (2)}$$

By solving above 2 normal equations we get

a_0, a_1 and hence we get the required straight line.

where,

N = no. of data points.

- i) By the method of Least squares, Find the straight line that best fits the following data and estimates the value of y corresponding to $x=7$.

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

Here $N=5$

Let the straight line to be fitted with the given data is $y = a_0 + a_1 x$ --- (*) where a_0, a_1 are constants.

The corresponding Normal equations are,

$$\sum y_i = a_0 N + a_1 \sum x_i \quad \text{--- (1)}$$

$$\text{and } \sum x_i y_i = a_0 \sum x_i + a_1 \sum x_i^2 \quad \text{--- (2)}$$

x_i	y_i	$x_i y_i$	x_i^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
$\sum x_i =$ 15	$\sum y_i =$ 204	$\sum x_i y_i =$ 748	$\sum x_i^2 =$ 55

Here $N=5$,

$$\textcircled{1}, \textcircled{2} \Rightarrow 204 = 5 \times a_0 + 15a_1 \quad \text{--- (1)}$$

$$748 = 15a_0 + 55a_1 \quad \text{--- (2)}$$

Solving (1) and (2) we get

$$\boxed{a_0 = 0} \quad \boxed{a_1 = 13.6}$$

$$\textcircled{*} \Rightarrow y = 0 + 13.6x$$

$$\boxed{y = 13.6x}$$

which is the required straight line.

$$\text{when } x=7, y = 13.6 \times 7$$

$$\boxed{y = 95.2}$$

2) Fit a straight line $y = a + bx$ to the following data.

x	-5	-3	-1	1	2	4
$f(x)$	0.4	-0.1	-0.3	0.3	0.1	0.4

The given straight line is $y = a + bx$ --- $\textcircled{*}$
where a, b are constants.

The Corresponding Normal equations are

$$\sum y_i = aN + b \sum x_i \quad \text{--- (1)}$$

$$\text{and } \sum x_i y_i = a \sum x_i + b \sum x_i^2 \quad \text{--- (2)}$$

x_i	y_i	$x_i y_i$	x_i^2
-5	0.4	-2	25
-3	-0.1	0.3	9
-1	-0.3	0.3	1
1	0.3	0.3	1
2	0.1	0.2	4
4	0.4	1.6	16

$$\sum x_i = -2 \quad \sum y_i = 0.2 \quad \sum x_i y_i = 0.1 \quad \sum x_i^2 = 56$$

$$N = 6$$

$$\textcircled{1}, \textcircled{2} \Rightarrow 0.2 = 6a - 2b \quad \text{--- (1)}$$

$$0.1 = -2a + 56b \quad \text{--- (2)}$$

Solving $\textcircled{1}$ & $\textcircled{2}$ we get,

$$\boxed{a = 0.037} \quad , \quad \boxed{b = 0.003}$$

$$\textcircled{*} \Rightarrow \boxed{y = 0.037 + 0.003x}$$

which is the required straight line.

- 3) The Temperature T (in $^{\circ}\text{C}$) and lengths L (in mm) of a heated rod are given below. If $L = a_0 + a_1 T_1$. Find the best values for a_0 and a_1 .

T	20	30	40	50	60	70
L	800.3	800.4	800.6	800.7	800.9	801.0

here $N = 6$

The given straight line is $L = a_0 + a_1 T_1$ --- $\textcircled{*}$

where a_0, a_1 are constants.

The Corresponding Normal Equations are,

$$\sum y_i = a_0 N + a_1 \sum x_i \quad \text{--- (1)}$$

$$\sum x_i y_i = a_0 \sum x_i + a_1 \sum x_i^2 \quad \text{--- (2)}$$

x_i	y_i	$x_i y_i$	x_i^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	801.0	56070	4900
270	4803.9	216201	13900

$$\textcircled{1}, \textcircled{2} \Rightarrow 4803.9 = 6a_0 + 270a_1 \quad \text{--- (i)}$$

$$216201 = 270a_0 + 13900a_1 \quad \text{--- (ii)}$$

solving (i) and (ii) we get,

$$a_0 = 799.99$$

$$a_1 = 0.0146$$

4) By the method of Least Squares, Find the best fitting straight line to the data given below.

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

Here $N = 5$

Let the straight line to be fitted with the given data is $y = a_0 + a_1x$ --- (*)

where a_0, a_1 are constants.

The Corresponding Normal Equations are,

$$\sum y_i = a_0 N + a_1 \sum x_i \quad \text{--- (1)}$$

$$\sum x_i y_i = a_0 \sum x_i + a_1 \sum x_i^2 \quad \text{--- (2)}$$

x_i	y_i	$x_i y_i$	x_i^2
5	15	75	25
10	19	190	100
15	23	345	225
20	26	520	400
25	30	750	625
75	113	1880	1375

$$\textcircled{1}, \textcircled{2} \Rightarrow 113 = 5a_0 + 75a_1 \quad \text{--- (i)}$$

$$1880 = 75a_0 + 1375a_1 \quad \text{--- (ii)}$$

Solving (i) & (ii) we get,

$$\boxed{a_0 = 11.5}$$

$$\boxed{a_1 = 0.74}$$

$$\textcircled{*} \Rightarrow \boxed{y = 11.5 + 0.74x}$$

which is the required straight line

Second Degree Curves Or Parabolas

Suppose the curve to be fitted for the given data is second degree curve or parabola. Let y

$$\text{Let } y = a + bx;$$

$$y = a + bx + cx^2 \quad \text{--- } \textcircled{*}$$

where a, b, c are constants

The Corresponding Normal equations are:

$$\sum y_i = aN + b \sum x_i + c \sum x_i^2 \quad \text{--- } \textcircled{1}$$

$$\sum x_i y_i = a \sum x_i + b \sum x_i^2 + c \sum x_i^3 \quad \text{--- } \textcircled{2}$$

$$\sum x_i^2 y_i = a \sum x_i^2 + b \sum x_i^3 + c \sum x_i^4 \quad \text{--- } \textcircled{3}$$

Solving $\textcircled{1}, \textcircled{2}, \textcircled{3}$ we get a, b, c values and hence, the required second degree Normal curve or parabola.

1) Fit a second degree polynomial for the following data, by the method of least squares.

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

Here $N=5$

Let the second degree polynomial

$$y = a + bx + cx^2 \quad \text{--- (*)}$$

where a, b, c are constants

The Corresponding Normal eqns are ;

$$\sum y_i = aN + b \sum x_i + c \sum x_i^2 \quad \text{--- (1)}$$

$$\sum x_i y_i = a \sum x_i + b \sum x_i^2 + c \sum x_i^3 \quad \text{--- (2)}$$

$$\sum x_i^2 y_i = a \sum x_i^2 + b \sum x_i^3 + c \sum x_i^4 \quad \text{--- (3)}$$

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
10	14	100	1000	10000	140	1400
12	17	144	1728	20736	204	2448
15	23	225	3375	50625	345	5175
23	25	529	12167	279841	575	13225
20	21	400	8000	160000	420	8400
80	100	1398	26270	521202	1684	30648

①, ②, ③ $\Rightarrow$

$$100 = 5a + 80b + 1398c \quad \text{--- (i)}$$

$$1684 = 80a + 1398b + 26270c \quad \text{--- (ii)}$$

$$30648 = 1398a + 26270b + 521202c \quad \text{--- (iii)}$$

Solving (i)(ii)(iii) we get,

$$a = -8.7279$$

$$b = 3.0099$$

$$c = -0.0694$$

2)

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

here $N = 5$

Let the second degree polynomial

$$y = a + bx + cx^2 \quad \text{--- (*)}$$

where a, b, c are constants.

The Corresponding Normal Equations are,

$$\sum y_i = aN + b\sum x_i + c\sum x_i^2 \quad \text{--- (1)}$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3 \quad \text{--- (2)}$$

$$\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 \quad \text{--- (3)}$$

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
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
10	12.9	30	100	354	37.1	130.3

①, ②, ③ $\Rightarrow$

$$12.9 = 5a + 10b + 30c \quad \text{--- (i)}$$

$$37.1 = 10a + 30b + 100c \quad \text{--- (ii)}$$

$$130.3 = 30a + 100b + 354c \quad \text{--- (iii)}$$

Solving (i)(ii)(iii) we get,

$$a = 1.42$$

$$b = -1.07$$

$$c = 0.55$$

3)

x	0	1	2
y	1	6	17

Here $N=3$

Let the Second degree polynomial,

$$y = a + bx + cx^2 \quad \text{--- (*)}$$

where a, b, c are constants.

The Corresponding Normal Equations are,

$$\sum y_i = aN + b\sum x_i + c\sum x_i^2 \quad \text{--- (1)}$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3 \quad \text{--- (2)}$$

$$\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 \quad \text{--- (3)}$$

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
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

①, ②, ③ $\Rightarrow$

$$24 = 3a + 3b + 5c \quad \text{--- (i)}$$

$$40 = 3a + 5b + 9c \quad \text{--- (ii)}$$

$$74 = 5a + 9b + 17c \quad \text{--- (iii)}$$

Solving (i)(ii)(iii) we get

$$a = 1$$

$$b = 2$$

$$c = 3$$

4) Find the parabola of the form $y = ax^2 + bx + c$, passing through the points $(-1, 2)$, $(0, 1)$ and $(1, 4)$

Given form $y = ax^2 + bx + c$

where a, b, c are constants.

x	-1	0	1
y	2	1	4

Parabola passes through the points
 $(-1, 2)$, $(0, 1)$, $(1, 4)$.

Sub $(-1, 2)$ in $y = ax^2 + bx + c$.

$$2 = a(-1)^2 + b(-1) + c$$

$$2 = a - b + c \quad \text{--- (i)}$$

Sub $(0, 1)$ in $y = ax^2 + bx + c$

$$1 = a(0) + b(0) + c$$

$$\boxed{c = 1} \quad \text{--- (ii)}$$

Sub $(1, 4)$ in $y = ax^2 + bx + c$

$$4 = a(1)^2 + b(1) + c$$

$$4 = a + b + c \quad \text{--- (iii)}$$

From (i) (iii), add

$$2 = a - b + c \quad \Rightarrow \quad 2 = a - b + 1$$

$$4 = a + b + c \quad \Rightarrow \quad 4 = a + b + 1$$

$$a - b = 1$$

$$a + b = 3$$

$$\hline 2a = 4$$

$$\boxed{a = 2}$$

Sub in (i)

$$2 = 2 - b + 1$$

$$2 - 3 = -b$$

$$-1 = -b$$

$$\boxed{b = 1}$$

Hence,

$$a = 2$$

$$b = 1$$

$$c = 1$$

More

Fitting of Exponential Curves.

i) Fitting of an Exponential curve $y = ae^{bx}$ — (*)
where a, b are constants.

Take Logarithm on both sides with base e

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

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

$$\log a^m = m \log a$$

$$\boxed{\log_e y = \log_e a + bx}$$

The above equⁿ can be written as,

$$\boxed{Y = A + Bx} \text{ — ①, which is a straight line.}$$

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

As Equⁿ ① is a straight line, The corresponding Normal equations are:

$$\Sigma Y = AN + B \Sigma x \text{ — ②}$$

$$\text{and } \Sigma XY = A \Sigma x + B \Sigma x^2 \text{ — ③}$$

Solving the above 2 equations we get A, B and
Hence from the relation

$$a = e^A \quad \boxed{a = e^A} \quad \boxed{b = B}$$

Finally we get the required Exponential Curve.

ii) Let the Exponential Curve $y = ab^x$ — (*) to be
log fitted with the given data.

Take Logarithm on both sides with base 10

$$\log_{10} y = \log_{10} (ab^x)$$

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

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

The above eqn can be written as,

$$Y = A + Bx \text{ --- (4), which is a straight line,}$$

where $Y = \log_{10} y$, $A = \log_{10} a$, $B = \log_{10} b$, $x = x$

As eqn (4) is a straight line, The corresponding Normal eqns are:

$$\sum Y = AN + B \sum x \text{ --- (5)}$$

$$\text{and } \sum xY = A \sum x + B \sum x^2 \text{ --- (6)}$$

Solving the above 2 eqns we get A, B and hence from the relation $A = \log_{10} a$, $B = \log_{10} b$, we get,

$$\boxed{a = 10^A}$$

$$\boxed{b = 10^B}$$

Finally we get the required Exponential Curve.

i) Determine the constants a and b by the method of Least square sum $y = ae^{bx}$

x	2	4	6	8	10
y	4.077	11.084	30.128	81.897	222.62

The given Exponential Curve is

$$y = ae^{bx} \text{ --- (*)}$$

Take log on both sides with base 'e'

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

The above eqn can be written as,

$$Y = A + Bx \text{ --- (1) which is a straight line}$$

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

The Corresponding Normal eqns of eqn (1) are,

$$\sum Y = AN + B \sum x \text{ --- (2)}$$

$$\text{and } \sum xY = A \sum x + B \sum x^2 \text{ --- (3)}$$

Here $N=5$

$x = X$	y	$Y = \log_e y$	XY	x^2
2	4.077	1.405	2.81	4
4	11.084	2.405	9.62	16
6	30.128	3.405	20.43	36
8	81.897	4.405	35.24	64
10	222.62	5.505	55.05	100

$$\sum x = 30 \quad \sum y = \quad \sum Y = 17.025 \quad \sum XY = 122.150 \quad \sum x^2 = 220$$

Sub in 2, 3 eqn

$$17.025 = 5A + 30B \quad \text{--- (i)}$$

$$122.150 = 30A + 220B \quad \text{--- (ii)}$$

From (i), (ii) we get

$$A = 0.405$$

$$B = 0.5$$

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

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

$$a = e^{0.405}$$

$$a = 1.499$$

$$b = 0.5$$

Hence, the required Exponential Curve is

$$y = 1.499 e^{0.5x}$$

$$a = 9.4725$$

$$b = 0.059$$

2)

x	1	2	7	9	12
y	10	15	12	15	21

The given Exponential Curve is,

$$y = ae^{bx} \quad \text{--- (*)}$$

Take log on both sides with base e,

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

The above eqn can be written as,

$$Y = A + Bx \text{ --- (1) which is a straight line.}$$

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

The Corresponding Normal eqns of eqn (1) are,

$$\sum Y = AN + B \sum x \text{ --- (2)}$$

$$\text{and } \sum XY = A \sum x + B \sum x^2 \text{ --- (3)}$$

Here $N = 5$.

$x = X$	y	$Y = \log_e y$	XY	x^2
1	10	2.302	2.302	1
2	15	2.708	5.416	4
7	12	2.484	17.388	49
9	15	2.708	24.372	81
12	21	3.044	36.528	144
31		13.246	86.006	279

Sub in 2,3 eqns

$$13.246 = 5A + 31B \text{ --- (i)}$$

$$86.006 = 31A + 279B \text{ --- (ii)}$$

From (i) (ii) we get,

$$A = 2.372$$

$$B = 0.044$$

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

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

$$a = e^{2.372}$$

$$a = 10.718$$

$$b = 0.044$$

Hence the Required Exponential Curve is,

$$y = 10.718 e^{0.044x}$$

3) Obtain the relation of a form $y = ab^x$ for the following data by the method of Least Squares.

x	2	3	4	5	6
y	8.3	15.4	33.2	65.2	127.4

Given eqn is, $y = ab^x$ — (*), which is an exponential curve.

Take log on both sides with base 10.

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

The above eqn can be written as,

$$Y = A + Bx \text{ — (1), which is straight line.}$$

where $Y = \log_{10} y$, $A = \log_{10} a$, $B = \log_{10} b$, $x = x$

The Corresponding Normal eqns for eqn (1) are

$$\sum Y = AN + B \sum x \text{ — (2)}$$

$$\text{and } \sum XY = A \sum x + B \sum x^2 \text{ — (3)}$$

Here $N = 5$

$x = X$	y	$Y = \log_{10} y$	XY	X^2
2	8.3	0.919	1.838	4
3	15.4	1.187	3.561	9
4	33.2	1.521	6.084	16
5	65.2	1.814	9.07	25
6	127.4	2.105	12.63	36
20		7.546	33.183	40

Sub in 2.3 eqns

$$7.546 = 5A + 20B \quad \text{--- (i)}$$

$$33.183 = 20A + 40B \quad \text{--- (ii)}$$

Solving (i), (ii) we get

$$A = 0.309$$

$$B = 0.29$$

Since ; $A = \log_{10} a$, $B = \log_{10} b$

$$a = 10^A , b = 10^B$$

$$a = 10^{0.309} , b = 10^{0.29}$$

$$a = 2.03$$

$$b = 1.94$$

Hence, the required Exponential Curve is,

$$y = (2.03)(1.94)^x$$

A)

x	0	1	2	3	4	5	6	7
y	10	21	35	59	92	200	400	610

Given eqn is $y = ab^x$ — (*)

Take log on both sides with base 10.

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

The above eqn can be written as,

$$Y = A + BX \text{ — (1) which is straight line.}$$

where $Y = \log_{10} y$, $A = \log_{10} a$, $B = \log_{10} b$, $x = x$.

The Corresponding Normal eqn are:

$$\sum Y = AN + B \sum x \text{ — (2)}$$

$$\text{and } \sum XY = A \sum x + B \sum x^2 \text{ — (3)}$$

Here $N = 8$

$x = x$	y	$Y = \log_{10} y$	xy	x^2
0	10	1	0	0
1	21	1.322	1.322	1
2	35	1.544	3.088	4
3	59	1.770	5.31	9
4	92	1.963	7.852	16
5	200	2.301	11.505	25
6	400	2.602	15.612	36
7	610	2.785	19.495	49
$\Sigma 8$		15.287	64.184	140

From 2.3 eqns,

$$15.287 = 8A + 28B \quad \text{--- (i)}$$

$$64.184 = 28A + 140B \quad \text{--- (ii)}$$

From (i) & (ii)

$$A = 1.020$$

$$B = 0.254$$

Since $A = \log_{10} a$, $B = \log_{10} b$

$$a = 10^A, \quad b = 10^B$$

$$a = 10^{1.020}, \quad b = 10^{0.254}$$

$$a = 10.471$$

$$b = 1.794$$

Hence the required Exponential Curve is,

$$y = (10.471)(1.794)^x$$

Fitting of Power Curve

Let $y = ax^b$ — (*), be the power curve to be fitted with the given data.

where a, b are constants.

Take log on both sides with base 10.

$$\log_{10} y = \log_{10} (ax^b)$$

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

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

The above eqn can be written as

$$Y = A + Bx \quad \text{--- (1)}, \text{ which is a straight line.}$$

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

The corresponding Normal Equns for eqn ① are,

$$\sum Y = AN + B \sum X \quad \text{--- (2)}$$

$$\sum XY = A \sum X + B \sum X^2 \quad \text{--- (3)}$$

Solving the above 2 eqns we get A and B and hence we get a and b.

Finally we get the required power curve $y = ax^b$

i) Obtain the power curve $y = ax^b$ for the following data.

x	1	2	3	4	5	6
y	2.98	4.26	5.21	6.10	6.8	7.5

Sol. Given power curve is $y = ax^b$ --- (*)

Take log on both sides with base 10.

$$\log_{10} y = \log_{10} (ax^b)$$

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

The above eqn can be written as

$$Y = A + BX \quad \text{--- (1) which is a straight line.}$$

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

The corresponding Normal eqns for eqn ① are,

$$\sum Y = AN + B \sum X \quad \text{--- (2)}$$

$$\sum XY = A \sum X + B \sum X^2 \quad \text{--- (3)}$$

x	y	$X = \log_{10} x$	$Y = \log_{10} y$	XY	X^2
1	2.98	0	0.4742		
2	4.26	0.3010	0.6294		
3	5.21	0.4771	0.7168		
4	6.10	0.6020	0.7853		
5	6.8	0.6989	0.8325		
6	7.5	0.7781	0.8750		
		2.8571	4.3132	2.267	1.773

2, 3 eqns $\Rightarrow$

$$4.313 = 6A + 2.857B \quad \text{--- (i)}$$

$$2.267 = 2.857A + 1.773B \quad \text{--- (ii)}$$

Solving (i) & (ii) we get,

$$A = 0.472$$

$$B = 0.516$$

$$A = \log_{10} a, \quad B = b$$

$$a = 10^A, \quad b = B$$

$$a = 10^{0.472}$$

$$a = 2.964$$

$$b = 0.516$$

Hence, the power curve is

$$y = (2.964)x^{0.516}$$

9) Find the curve of best fit of the type $y = ae^{bx}$ to the following data by the method of least squares.

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

The given exponential curve is

$$y = ae^{bx} \quad \text{--- *}$$

Take log on b.s with base e .

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

The above eqn can be written as,

$$Y = A + BX \quad \text{--- ① which is a straight curve.}$$

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

The corresponding Normal eqns from eqn ① are,

$$\sum Y = AN + B \sum X \quad \text{--- ②}$$

$$\sum XY = A \sum X + B \sum X^2 \quad \text{--- ③}$$

Here $N = 5$.

$x = X$	y	$Y = \log_e y$	XY	X^2
1	10	2.30	2.30	1
5	15	2.70	13.5	25
7	12	2.48	17.36	49
9	15	2.70	24.3	81
12	21	3.04	36.48	144
34	73	13.22	93.94	300

Sub in 2,3 eqns $\Rightarrow$

$$13.22 = 5A + 34B \quad \text{--- (i)}$$

$$93.94 = 34A + 300B \quad \text{--- (ii)}$$

From (i) (ii) we get,

$$A = 2.244$$

$$B = 0.058$$

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

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

$$a = e^{2.244}$$

$$a = 9.43$$

$$b = 0.05$$

Hence the required Exponential Curve is

$$y = 9.43 e^{0.05x}$$

10) Fit a curve $y = ab^x$ by the method of least squares to the data given below.

x	0	1	2	3	4	5	6	7
y	10	21	35	59	92	200	400	610

Given eqn is $y = ab^x$ — *

Take log on b.s with base 10.

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

The above eqn can be written as,

$$Y = A + Bx \text{ — (1) which is straight line}$$

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

The Corresponding Normal eqns are,

$$\sum Y = AN + B \sum X \text{ — (2)}$$

$$\sum XY = A \sum X + B \sum X^2 \text{ — (3)}$$

Here $N = 8$

$x = X$	y	$Y = \log_{10} y$	XY	X^2
0	10	1	0	0
1	21	1.32	1.32	1
2	35	1.54	3.08	4
3	59	1.77	5.31	9
4	92	1.96	7.84	16
5	200	2.30	11.5	25
6	400	2.60	15.6	36
7	610	2.78	19.46	49
28	1427	15.27	64.11	140.

Sub in 2,3 eqns,

$$15.27 = 8A + 28B \quad \text{--- (i)}$$

$$64.11 = 28A + 140B \quad \text{--- (ii)}$$

Solving (i)(ii) we get,

$$A = 1.02$$

$$B = 0.25$$

Since,

$$A = \log_{10} a, \quad B = \log_{10} b$$

$$a = 10^A, \quad b = 10^B$$

$$a = 10^{1.02}, \quad b = 10^{0.25}$$

$$a = 10.47$$

$$b = 1.77$$

Hence the required Exponential Curve is;

$$y = (10.47)(1.77)^x$$

11) Fit a power curve of the form $y = ax^b$ to the table.

x	2	4	6	8	10
y	0.973	3.839	8.641	15.987	23.794

Given power curve is

$$y = ax^b \quad \text{--- *}$$

Take log on b.s with base 10.

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

The above eqn can be written as,

$$Y = A + BX \quad \text{--- ① which is straight line.}$$

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

The Corresponding Normal eqns are,

$$\sum Y = AN + B \sum X \quad \text{--- ②}$$

$$\sum XY = A \sum X + B \sum X^2 \quad \text{--- ③}$$

x	y	$X = \log_{10} x$	$Y = \log_{10} y$	XY	X^2
2	0.973	0.301	-0.011	-0.003	0.090
4	3.839	0.602	0.584	0.351	0.362
6	8.641	0.778	0.936	0.728	0.605
8	15.987	0.903	1.203	1.086	0.815
10	23.794	1	1.376	1.376	1
30	53.234	3.584	4.088	3.538	2.872

Sub in 2,3 eqns $\Rightarrow$

$$4.088 = 5A + 3.584B \quad \text{--- (i)}$$

$$3.538 = 3.584A + 2.872B \quad \text{--- (ii)}$$

From (i)(ii) we get,

$$A = -0.620$$

$$B = 2.005$$

$$A = \log_{10} a, B = b$$

$$a = 10^A, b = B$$

$$a = 10^{-0.620}$$

$$a = 0.239$$

$$b = 2.005$$

Hence, the power curve is,

$$y = (0.239)x^{2.005}$$

12) Fit a Second degree curve $y = a + bx + cx^2$ to the following data.

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

Regression

The statistical method which help us to estimate the unknown value of one variable from the known value of its related variable is called Regression.

Note:

1) There are 2 Regression lines constructed for the given data

Regression line shows the Regression of y on x and the other x on y .

2) There are 2 methods to find Regression lines

i) Method of Least squares

ii) Mean deviation method.

Regression lines using

Method of Least squares.

i) Regression line of y on x .

$$y = a + bx$$

The corresponding Normal Equns are,

$$\sum y = aN + b \sum x$$

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

ii) Regression line of x on y .

$$x = a + by$$

The Corresponding Normal eqns are,

$$\sum x = aN + b \sum y$$

$$\sum xy = a \sum y + b \sum y^2$$

Here N = no. of data points / observations.

i) From the following data obtain 2 regression lines.

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

Here $N=5$.

i) Regression line of y on x:

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

The Corresponding Normal eqns are,

$$\sum y = aN + b\sum x \quad \text{--- (1)}$$

$$\sum xy = a\sum x + b\sum x^2 \quad \text{--- (2)}$$

x	y	x^2	y^2	xy
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
30	40	220	340	214

1.2 eqns $\Rightarrow$

$$40 = 5a + 30b \quad \text{--- (i)}$$

$$214 = 30a + 220b \quad \text{--- (ii)}$$

Solving (i), (ii) we get

$$a = 11.9$$

$$b = -0.65$$

Hence

$$y = (11.9) - 0.65x$$

ii) Regression line of x on y .

$$x = a + by \quad \text{--- (*)}$$

The corresponding eqns are,

$$\sum x = aN + b \sum y \quad \text{--- (3)}$$

$$\sum xy = a \sum y + b \sum y^2 \quad \text{--- (4)}$$

3, 4 eqns $\Rightarrow$

$$30 = 5a + 40b \quad \text{--- (i)}$$

$$214 = 40a + 340b \quad \text{--- (ii)}$$

From (i)(ii)

$$a = 16.4$$

$$b = -1.3$$

Hence

$$x = 16.4 - 1.3y$$

2) Determine the Regression Equations.

x	2	4	6	8	10	12	14
y	4	2	5	10	4	11	12

Here $N = 7$.

i) Regression line of y on x :

$$y = a + bx \quad \text{--- *}$$

The corresponding eqns are,

$$\sum y = aN + b \sum x \quad \text{--- (1)}$$

$$\sum xy = a \sum x + b \sum x^2 \quad \text{--- (2)}$$

x	y	xy	x^2	y^2
2	4	8	4	16
4	2	8	16	4
6	5	30	36	25
8	10	80	64	100
10	4	40	100	16
12	11	132	144	121
14	12	168	196	144
56	48	466	560	426

eqns 1,2 $\Rightarrow$

$$48 = 7a + 56b \quad \text{--- (i)}$$

$$466 = 56a + 560b \quad \text{--- (ii)}$$

Solving (i)(ii) we get,

$$\boxed{a=1} \quad \boxed{b=0.73}$$

Hence,

$$\boxed{y = 1 + 0.73x}$$

ii) Regression line of x on y .

$$x = a + by \quad \text{--- *}$$

The corresponding eqns are,

$$\Sigma x = aN + b \Sigma y \quad \text{--- (3)}$$

$$\Sigma xy = a \Sigma y + b \Sigma y^2 \quad \text{--- (4)}$$

3,4 eqns $\Rightarrow$

$$56 = 7N + 48b \quad \text{--- (i)}$$

$$466 = 48a + 426b \quad \text{--- (ii)}$$

Solving (i)(ii) we get,

$$a = 2.19$$

$$b = 0.84$$

Hence, $x = 2.19 + 0.84y$

3)

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

Here $N = 7$

i) Regression line of y on x :

$$y = a + bx \quad \text{--- *}$$

The corresponding N.E are,

$$\sum y = aN + b\sum x \quad \text{--- (1)}$$

$$\sum xy = a\sum x + b\sum x^2 \quad \text{--- (2)}$$

x	y	xy	x^2	y^2
10	10	100	100	100
12	22	264	144	484
13	24	312	169	576
16	27	432	256	729
17	29	493	289	841
20	33	660	400	1089
25	37	925	625	1369
113	182	3186	1983	5188

eqn (1) & (2) $\Rightarrow$

$$182 = 7a + 113b \quad \text{--- (i)}$$

$$3186 = 113a + 1983b \quad \text{--- (ii)}$$

Solving (i) (ii) we get,

$$a = 0.798$$

$$b = 1.561$$

Hence,

$$y = 0.798 + 1.561x$$

ii) Regression line of x on y :

$$x = a + by \quad \text{--- *}$$

The corresponding eqns are,

$$\sum x = aN + b \sum y \quad \text{--- (3)}$$

$$\sum xy = a \sum y + b \sum y^2 \quad \text{--- (4)}$$

3, 4 eqns $\Rightarrow$

$$113 = 7a + 182b \quad \text{--- (i)}$$

$$3186 = 182a + 5188b \quad \text{--- (ii)}$$

Solving (i) (ii) we get,

$$a = 2.002$$

$$b = 0.543$$

$$x = 2.002 + 0.543y$$

Regression equations by mean deviation method:

Regression. Equ^{ns} of

y on x

x on y.

$$1) y - \bar{y} = b_{yx}(x - \bar{x})$$

b_{yx} = Regression coefficient
of y on x

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

- 2) x is independent variable
y is dependent variable

- 3) Used to calculate y when
x is given.

- 4) Slope of the line is
 $m_1 = b_{yx}$

$$1) x - \bar{x} = b_{xy}(y - \bar{y})$$

b_{xy} = Regression coefficient
of x on y

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

- 2) y is independent variable
x is dependent variable

- 3) Used to calculate x
when y is given.

- 4) Slope of the line is
 $m_2 = \frac{1}{b_{xy}}$

Note:

- 1) The 2 Regression lines intersect at the point $(\bar{x}, \bar{y})$
where, $\bar{x}$ = Mean of x series, $\bar{y}$ = Mean of y series
- 2) Another formula to find Regression coefficients are

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

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$$

where, $r =$

$$b_{yx} \times b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} \times r \cdot \frac{\sigma_y}{\sigma_x}$$

$$r^2 = b_{yx} \times b_{xy}$$

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

i) Find both the Regression lines x on y and y on x for the following data

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

Here $n=5$

i) Regression line of x on y :

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

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6+2+10+4+8}{5} \Rightarrow 6$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{9+11+5+8+7}{5} \Rightarrow 8$$

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

x	y	xy	x^2	y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49

$$\sum x = 30$$

$$\sum x^2 = 220$$

$$\sum xy = 214$$

$$\sum y = 40$$

$$\sum y^2 = 340$$

$$b_{xy} = \frac{(5 \times 214) - (30 \times 40)}{(5 \times 340) - (40)^2}$$

$$b_{xy} = -1.3$$

Sub in * $\Rightarrow$

$$x - 6 = -1.3(y - 8)$$

$$x = -1.3y + 16.4$$

which is the Required Regression line of x on y .

ii) Regression equn of y on x :

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

$$\text{where, } b_{yx} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$= \frac{(5 \times 214) - (30 \times 40)}{(5 \times 220) - (30)^2}$$

$$b_{yx} = -0.65$$

Sub in ** $\Rightarrow$

$$y - 8 = -0.65(x - 6)$$

$$y = -0.65x + 11.9$$

which is required Regression line of y on x .

Properties of Regression coefficients:

- 1) The Coefficient of Correlation r , is geometric mean of coefficient of Regression i.e.,

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

where $(-1 \leq r \leq 1)$

- 2) If one of the Regression coefficient is greater than 1, then other should be less than 1.
- 3) Sign of both Regression coefficients should be same i.e., either both b_{xy} and b_{yx} are positive or Negative.
- 4) Both the Regression lines intersect at the point $(\bar{x}, \bar{y})$.

2) From the following data

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

- i) Obtain 2 Regression lines
ii) Calculate Coefficient of Correlation
iii) Estimate the value y for $x = 6.2$

Sol. Here $N = 9$

i) a) Regression line of x on y :

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

where,

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

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3+\dots+9}{9} \Rightarrow 5$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{9+8+10+\dots+15}{9} \Rightarrow 12$$

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

x	y	xy	x ²	y ²
1	9	9	1	81
2	8	16	4	64
3	10	30	9	100
4	12	48	16	144
5	11	55	25	121
6	13	78	36	169
7	14	98	49	196
8	16	128	64	256
9	15	135	81	225
$\sum x = 45$	$\sum y = 108$	$\sum xy = 597$	$\sum x^2 = 285$	$\sum y^2 = 1356$

$$b_{xy} = \frac{(9 \times 597) - (45 \times 108)}{(9 \times 1356) - (108)^2}$$

$$b_{xy} = 0.95$$

Sub in * $\Rightarrow$

$$x - 5 = 0.95(y - 12)$$

$$\boxed{x = 0.95y - 6.4}$$

which is the Required Regression line of x on y.

b) Regression lines of y on x :

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

$$\text{where, } b_{yx} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$= \frac{(9 \times 597) - (45 \times 108)}{(9 \times 285) - (45)^2}$$

$$b_{yx} = 0.95$$

Sub in ** $\Rightarrow$

$$y - 12 = 0.95(x - 5)$$

$$\boxed{y = 0.95x + 7.25}$$

Which is the Required Regression line of y on x .

ii)

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

$$= \sqrt{0.95 \times 0.95}$$

$$r = 0.95$$

iii) $x = 6.2$, $y = ?$

Regression eqn of y on x is,

$$y = 0.95x + 7.2$$

$$y = 0.95(6.2) + 7.2$$

$$\boxed{y = 13.09}$$

3) Find the Correlation between x and y when the lines of regression are $2x - 9y + 16 = 0$ and $x - 2y + 1 = 0$

Sol.

Let the given 2 regression lines be

$$x \text{ on } y \text{ is } 2x - 9y + 16 = 0$$

$$y \text{ on } x \text{ is } x - 2y + 1 = 0$$

$$2x = 9y - 16, \quad 2y = x + 1$$

$$x = \frac{9}{2}y - 8 \quad \text{--- (i)}$$

$$y = \frac{1}{2}x + \frac{1}{2} \quad \text{--- (ii)}$$

$$b_{xy} = \frac{9}{2}, \quad b_{yx} = \frac{1}{2}$$

$$r = \sqrt{b_{xy} \times b_{yx}} = \frac{3}{2} > 1$$

$$\therefore (-1 \leq r \leq 1)$$

$\Rightarrow$ Our choice of regression lines is not correct.

Let us consider the Regression lines

$$x \text{ on } y \text{ is } x - 2y + 1 = 0$$

$$y \text{ on } x \text{ is } 2x - 9y + 16 = 0$$

$$x = 2y - 1, \quad 9y = 2x + 16$$

$$x = 2y - 1 \quad \text{--- (iii)}$$

$$y = \frac{2}{9}x + \frac{16}{9} \quad \text{--- (iv)}$$

$$b_{xy} = 2, \quad b_{yx} = \frac{2}{9}$$

$$r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{2 \times \frac{2}{9}} \Rightarrow \frac{2}{3} < 1$$

$$\boxed{r = \frac{2}{3}}$$

$$(\because -1 \leq r \leq 1)$$

$$\boxed{r = 0.66}$$

- 5) The two regression equations of the variable x and y are
 $x = 19.13 - 0.87y$ and $y = 11.64 - 0.50x$. Find
- Mean of x
 - Mean of y
 - The coefficient of correlation b/w x & y .

Sol. We know that 2 regression lines intersect at the point $(\bar{x}, \bar{y})$

$$\bar{x} = 19.13 - 0.87\bar{y}$$

$$\bar{y} = 11.64 - 0.50\bar{x}$$

$$\bar{x} + 0.87\bar{y} = 19.13$$

$$\bar{y} + 0.50\bar{x} = 11.64$$

Solving, we get

$$\boxed{\bar{x} = 15.934}$$

$$\boxed{\bar{y} = 3.672}$$

Note: The Regression that are discussed so far is Linear Regression.

Multiple Regression (or) Multiple Linear Regression:

We know that the production of rice (y) depends not only on the amount of rainfall (x_1), but also on the Fertilizers (x_2), pesticides (x_3), Quality of seeds (x_4), Quality of soil (x_5), etc.

In a Multiple Linear Regression, the dependent variable y is function of more than 1 independent variable.

Multiple Linear Regression is a Linear Relationship between y and $x_1, x_2, x_3, x_4, x_5, \dots$

$$\text{i.e., } y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots$$

Consider a Multiple Linear Regression,

$$y = a_0 + a_1x_1 + a_2x_2 \quad \text{---} \quad *$$

where a_0, a_1, a_2 are constants

The Corresponding Normal Equations are,

$$\sum y = a_0 N + a_1 \sum x_1 + a_2 \sum x_2 \quad \text{---} \quad (1)$$

$$\sum x_1 y = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 \quad \text{---} \quad (2)$$

$$\sum x_2 y = a_0 \sum x_2 + a_1 \sum x_2 x_1 + a_2 \sum x_2^2 \quad \text{---} \quad (3)$$

Solving the above 3 equations we get a_0, a_1, a_2 and hence the required multiple Linear Equation
Here $N = \text{no. of observations / No. of data points.}$

1) Find the Least Square Regression equation of y on x_1 and x_2 from Following data.

y	3	5	6	8	12	14
x_1	16	10	7	4	3	2
x_2	90	72	54	42	30	12

Sol.

Here $N = 6$.

Let the Multiple Linear Regression of y on x_1, x_2 is given by.

$$y = a_0 + a_1 x_1 + a_2 x_2 \quad \text{---} *$$

The Corresponding Normal Equns are,

$$\sum y = a_0 N + a_1 \sum x_1 + a_2 \sum x_2 \quad \text{---} (1)$$

$$\sum x_1 y = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 \quad \text{---} (2)$$

$$\sum x_2 y = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 \quad \text{---} (3)$$

y	x_1	x_2	$x_1 x_2$	x_1^2	x_2^2	$x_1 y$	$x_2 y$
3	16	90	1440	256	8100	48	270
5	10	72	720	100	5184	50	360
6	7	54	378	49	2916	42	324
8	4	42	168	16	1764	32	336
12	3	30	90	9	900	36	360
14	2	12	24	4	144	28	168
$\sum y =$ 48	$\sum x_1 =$ 42	$\sum x_2 =$ 300	$\sum x_1 x_2 =$ 2820	$\sum x_1^2 =$ 434	$\sum x_2^2 =$ 19008	$\sum x_1 y =$ 236	$\sum x_2 y =$ 1818

equⁿ ①, ②, ③ $\Rightarrow$

$$48 = 6a_0 + 42a_1 + 300a_2$$

$$236 = 42a_0 + 434a_1 + 2820a_2$$

$$1818 = 300a_0 + 2820a_1 + 19008a_2$$

$$a_0 = 16.111$$

$$a_1 = 0.427$$

$$a_2 = -0.222$$

The Required Multiple Regression equⁿ is ,

$$y = 16.111a_0 + 0.427x_1 - 0.222x_2$$

2) The Following data on the no. of twists required to break a certain kind of forged alloy bar and the percent of 2 alloy elements present in the metal.

No. of twists (y)	41	49	69	65	40	50	58	57	31	36	44	57	19	31	33	43
% of element A = x_1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
B = x_2	5	5	5	5	10	10	10	10	15	15	15	15	20	20	20	20

Fit a Least Square Regression plane and use its equⁿ to estimate the no. of twists required to break one of the bars when $x_1 = 2.5$ and $x_2 = 12$.