CURVE FITTING

curve fitting is to form an equation of the curve from a given data. There are two methods for ~~findin~~ fitting a curve

i) graphical method ii) Method of least squares.

<u>Graphical method</u> is simple here we have to draw a curve passing through the maximum number of points in practice it is difficult for forming of the points. The simple curve can be fitted to number of points is line for which $n=1$
we can also find the curves of the second degree, third degree.

<u>Method of least squares</u>:. let x_i, y_i, $i=1$ to n be n sets of observations of the related data and $y=a+bx$ be the straight line to be fitted let $y=f(x)$ be the relation suggested between x & y.
The difference between the expected value & observed values when $x=x_i$ the observed values is y_i the expected values is $f(x_i)$ error is $e_i = y_i - f(x_i)$.
Our Aim is to find the functional relation ship between the observed values & expected values. The difference between the obseved & expected values is known as residual (error). we want to minimize these residual (errors). Since these

difference may be positive in some cases & of square of these residual a minimum. This method is called the method of least squares.

Derive the normal equations of straight line.

let the straight line be $f(x) = y = a + bx \rightarrow (1)$

let the straight line (1) passes through the data points $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ i.e (x_i, y_i) $i = 1, 2, \dots n$.

$$y_i = a + bx_i \rightarrow (2)$$

The error E_i between the observed & expected value of $y = y_i$ is defined as.

$$E_i \text{ or } S_i = y_i - f(x_i) \quad , i = 1, 2, \dots n.$$

$$E_i = y_i - (a + bx_i).$$

sum of the squares of these errors is.

$$\sum E_i^2 \text{ or } S_i^2 = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2.$$

let $\sum E_i^2 = S$.

$$\therefore S = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2 = \text{least square error.}$$

Now For S to be minimum

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

$$\frac{\partial S}{\partial a} = 2 \sum [y_i - a - bx_i].(-1) = 0$$

$$\Rightarrow \sum y_i - \sum a - \sum bx_i = 0$$

$$\Rightarrow \sum y_i - a \sum 1 - b \sum x_i = 0.$$

$$\Rightarrow \sum y_i = na + b \sum x_i \quad \Rightarrow \boxed{\sum y = na + b \sum x} \rightarrow (1)$$

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

$2\,\Sigma[y_i - a - bx_i](-x_i] = 0$.

$\Rightarrow \Sigma x_i y_i - \Sigma a x_i - b\Sigma x_i^2 = 0$

$\Rightarrow \Sigma x_i y_i - a\Sigma x_i - b\Sigma x_i^2 = 0$.

$\Rightarrow \boxed{\Sigma xy = a\Sigma x + b\Sigma x^2} \rightarrow (2)$

equations (1) & (2) are normal equations of a straight line.

Non_linear least squares approximations:

(second degree curve)

let the equation of the Parabola to be fit be

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

let the Parabola (1) passes through the data points $(x_1, y_1), (x_2, y_2) \ldots\ldots (x_n, y_n)$ i.e (x_i, y_i), $i = 1, 2, \ldots n$

$\therefore f(x_i) = y_i = a + bx_i + cx_i^2$.

The error E_i between the observed & expected values of $y = y_i$ is defined as

E_i (or) $S_i = y_i - (a + bx_i + cx_i^2) \rightarrow (3)$ $i = 1, 2, 3, \ldots n$

The sum of the squares of these errors is

$S = \Sigma S_i^2 = \sum_{i=1}^{n} [y_i - a - bx_i - cx_i^2]^2 \rightarrow (4)$.

For S to be minimum, we have

$$\frac{\partial S}{\partial a} = 0, \quad \frac{\partial S}{\partial b} = 0, \quad \frac{\partial S}{\partial c} = 0.$$

$$\frac{\partial S}{\partial a} = 2 \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2)(-1) = 0.$$

$$\Sigma y_i = a \Sigma 1 + b \Sigma x_i + c \Sigma x_i^2.$$

$$\Sigma y = na + b \Sigma x + c \Sigma x^2 \rightarrow (5)$$

$$\frac{\partial S}{\partial b} = 2 \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2)(-x_i) = 0$$

$$\Rightarrow \Sigma x_i y_i = a \Sigma x_i + b \Sigma x_i^2 + c \Sigma x_i^3$$

$$\Rightarrow \Sigma xy = a \Sigma x + b \Sigma x^2 + c \Sigma x^3 \rightarrow (6).$$

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

$$\Sigma x^2 y = a \Sigma x^2 + b \Sigma x^3 + c \Sigma x^4 \rightarrow (7).$$

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

are the normal equations for fitting a parabola.

1) By the method of least squares find the straight line that best fits the following data.

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

Normal equations are $\Sigma y = na + b\Sigma x$

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

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

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

$204 = 5a + 15b \rightarrow (1)$

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

$748 = 15a + 55b \rightarrow (2)$

By solving (1) & (2) equations, we get.

$a = 0, \quad b = 13.6.$

$y = a + bx$

$y = 13.6x$

) Certain experimental values of x & y are given below.

x	0	2	5	7
y	-1	5	12	20

If $y = a_0 + a_1 x$ find the approximation values of a_0 & a_1.

x	y	x^2	xy
0	-1	0	0
2	5	4	10
5	12	25	60
7	20	49	140
$\Sigma x = 14$	$\Sigma y = 36$	$\Sigma x^2 = 78$	$\Sigma xy = 210$

Normal equations are $\boxed{\begin{array}{l} \Sigma y = na + b\Sigma x \\ \Sigma xy = a\Sigma x + b\Sigma x^2 \end{array}}$

$$36 = 4a + 14b \rightarrow ①$$

$$210 = 14a + 78b \rightarrow ②$$

By solving ① & ② equations are $a = -1.1381$, $b = 2.8966$.

3) The temperatures T (in °c) & lengths L (in mm) of a heated rod are given below. If $L = a_0 + a_1 T$, find the best values for a_0 & a_1.

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

Normal equations are $\Sigma y = na + b\Sigma x$

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

x (T)	y (L)	x^2 (T^2)	xy (TL)
20	800.3	400	16006
30	800.4	900	24012
40	800.6	1600	32024
50	800.7	2500	40035
60	800.9	3600	48054
70	801.0	4900	56070
270	4803.9	113900	216201

$$4803.9 = 6a + 270b \rightarrow (1)$$

$$216201 = 270a + 13900b \rightarrow (2)$$

By solving (1) & (2), we get $a = 800$, $b = 0.0146$.

Note:-

i) If n, the number of time series values is odd, then the transformation is : $x = \dfrac{t - \text{middle value}}{\text{interval (h)}}$.

Thus, if we are given yearly figures for, say, 1990, 1991, 1992, ... 1996, i.e $n=7$, then.

$$x = \frac{t - \text{middle value}}{1} = t - 1993.$$

Putting $t = 1990, 1991, 1992, \ldots 1996$, we get, $x = -3, -2, -1, 0, 1, 2$
so that $\Sigma x = \Sigma x^3 = 0$.

ii) If n is even, then the transformation is

$$x = \frac{t - (\text{Arithmetic mean of two middle values})}{\frac{1}{2}(\text{interval})}.$$

Thus, if we are given the yearly values for, say,

1995, 1996, 1997, ... 2002, then.

$$x = \frac{t - \frac{1}{2}(1998+1999)}{\frac{1}{2}}$$

$$= 2(t - 1998.5) = 2t - 3997.$$

Putting $t = 1995, 1996, \ldots 2002$, we get.

$x = -7, -5, -3, -1, 1, 3, 5, 7$. so that $\Sigma x = \Sigma x^3 = 0$

The above transformations will always give. $\Sigma x = 0 = \Sigma x^3$.

& thus reduces the algebraic calculations for the solution of normal equations to a great extent.

1) Fit a linear trend to the following data by the least squares method. verify that $\Sigma(y - y_e) = 0$, where y_e is the corresponding trend value of y.

year :	1990	1992	1994	1996	1998
Production(in '000 units) :	18	21	23	27	16.

$n = 5$ i.e odd. Hence, we shift the origin to the middle of the time period, the year 1994.

let $x = t - 1994$.

let the line equation be $y = a + bx$. (origin 1994)

Year (t)	production ('000 units) (y)	$x = t - 1994$	x^2	xy	$y_e = 21 + 0.1x$	$y - y_e$
1990	18	-4	16	-72	20.6	-2.6
1992	21	-2	4	-42	20.8	0.2
1994	23	0	0	0	21.0	2.0
1996	27	2	4	54	21.2	5.8
1998	16	4	16	64	21.4	-5.4
	105	0	40	4		0

Normal equations are $\Sigma y = na + b\Sigma x$

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

$105 = 5a + 0$ $\qquad$ $4 = a \times 0 + b(40)$

$\Rightarrow a = \frac{105}{5} = 21$ $\qquad$ $\Rightarrow b = \frac{4}{40} = \frac{1}{10} = 0.1$.

The straight line trend equation is given by.

$y = a + bx = 21 + 0.1x$

5) The sales of a company in million of rupees for the years 1994-2001 are given below:

Year :	1994	1995	1996	1997	1998	1999	2000	2001
Sales :	550	560	555	585	540	525	545	585

i) Find the linear trend equation.

ii) Estimate the sales for the year 1993.

∵ the no of pairs is even, 8, we shift the origin to the time which is the arithmetic mean of the two middle times 1997 & 1998.

$$x = \frac{t - \left(\frac{1997+1998}{2}\right)}{\frac{1}{2}(\text{interval})} = 2(t-1997.5) = 2t - 3995.$$

Year (t)	Sales (y)	$x=2(t-1997.5)$	xy	x^2	$y_c = a+bx$ (Trend values)
1994	550	-7	-3850	49	$555.63 - 7\times 0.21 = 554.16$
1995	560	-5	-2800	25	554.58
1996	555	-3	-1665	9	555
1997	585	-1	-585	1	555.42
1998	540	1	540	1	555.84
1999	525	3	1575	9	556.26
2000	545	5	2725	25	556.68
2001	585	7	4095	49	557.1
	4445	0	35	168	

The normal equations are $\Sigma y = na + b\Sigma x$

$4445 = 8a + 0$

$\Rightarrow a = \frac{4445}{8} = 555.63$

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

$35 = 0 + 168b \Rightarrow b = \frac{35}{168} = 0.21$

straight line equation $y = a + bx$

$y = 555.63 + 0.21x$

ii) The estimated sales for 1993 are

$$y = 555.63 + 0.21x.$$

$$x = 2(t - 1997.5) = 2(1993 - 1997.5)$$
$$= -9.$$

$$\therefore y = 555.63 + 0.21(-9).$$
$$= 553.74 \text{ million Rs.}$$

6) Fit a second degree Polynomial to the following data by the method of least squares.

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

let the required Polynomial equation is $y = a + bx + cx^2$

Normal equations are $\sum y = na + b\sum x + c\sum x^2$
$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$
$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4.$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	1.8	1	1	1	1.8	1.8
2	1.3	4	8	16	2.6	5.2
3	2.5	9	27	81	7.5	22.5
4	6.3	16	64	256	25.2	100.8
Total 10	12.9	30	100	354	37.1	130.3

$12.9 = 5a + 10b + 30c \rightarrow ①$

$37.1 = 10a + 30b + 100c \rightarrow ②$

$130.3 = 30a + 100b + 354c \rightarrow ③$

By solving above equations, we get

$a = 1.42$, $b = -1.07$, $c = 0.55$.

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

$\Rightarrow y = 1.42 - 1.07x + 0.55x^2$.

7) Fit a second degree parabola to the following data.

x	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y	1.1	1.3	1.6	2.0	2.7	3.4	4.1

$a = 1.04$, $b = -0.198$ $c = 0.244$.

$y = a + bx + cx^2 = 1.04 - 0.198x + 0.244x^2$

8) For 10 randomly selected observations, the following data were recorded:

observation No.	1	2	3	4	5	6	7	8	9	10
overtime hrs (x)	1	1	2	2	3	3	4	5	6	7
Additional units (y)	2	7	7	10	8	12	10	14	11	14.

Determine the regression equation using the non-linear form $y = a + bx + cx^2$

S.No	x	y	x^2	x^3	x^4	xy	x^2y
1	1	2	1	1	1	2	2
2	1	7	1	1	1	7	7
3	2	7	4	8	16	14	28
4	2	10	4	8	16	20	40
5	3	8	9	27	81	24	72
6	3	12	9	27	81	36	108
7	4	10	16	64	256	40	160
8	5	14	25	125	625	70	350
9	6	11	36	216	1296	66	396
10	7	14	49	343	2401	98	686
	34	95	154	820	4774	377	1849

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

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

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

$95 = 10a + 34b + 154c$

$377 = 34a + 154b + 820c$

$1849 = 154a + 820b + 4774c$

By solving above equations, we get

$a = 1.80$, $b = 3.48$, $c = -0.27$

Regression equation is $y = 1.80 + 3.48x - 0.27x^2$.

1) Find the least squares fit of the form $y = a_0 + a_1 x^2$ to the following data.

x	-1	0	1	2
y	2	5	3	0

Let $x^2 = X$ then $Y = a_0 + a_1 X$.

Normal equations are $\Sigma Y = n a_0 + a_1 \Sigma X$.

$$\Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2$$

x	Y	$x^2 = X$	XY	X^2
-1	2	1	2	1
0	5	0	0	0
1	3	1	3	1
2	0	4	0	16
Σ	10	6	5	18

$$10 = 4a_0 + 6a_1$$

$$5 = 6a_0 + 18a_1$$

By solving these equations, we get

$a_0 = 4.167$, $a_1 = -1.111$

$$Y = a_0 + a_1 X = 4.167 - 1.111 x^2$$

Exponential curve.

i) $y = ae^{bx}$ ii) $y = ab^{x}$

Exponential curve is given by $y = ae^{bx}$.

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

$$Y = A + bx.$$

Then the problem reduces to finding a least square straight line.

$$\boxed{\begin{aligned} \Sigma Y &= nA + b\Sigma x \\ \Sigma xY &= A\Sigma x + b\Sigma x^2 \end{aligned}}$$

10) Using the method of least squares, find the constant a & b such that $y = ae^{bx}$ fits the following data.

x	0	0.5	1.0	1.5	2.0	2.5
y	0.10	0.45	2.15	9.15	40.35	180.75.

x	y	$Y = \log_e y$	xY	x^2
0	0.1	-2.302	0	0
0.5	0.45	-0.798	-0.399	0.25
1.0	2.15	0.765	0.765	1
1.5	9.15	2.213	3.319	2.25
2.0	40.35	3.697	7.394	4
2.5	180.75	5.197	12.992	6.25.
7.5		8.772	24.071	13.75.

Normal equation are $\Sigma Y = nA + b\Sigma x$, $6A + 7.5b = 8.772$

$\Sigma xY = A\Sigma x + b\Sigma x^2$ $7.5A + 13.75b = 25.189$

$A = -2.28$ $\qquad$ $b = \cancel{3.4448}\ 2.99$

$A = \log_e a \Rightarrow a = e^A = e^{-2.28} = 0.101$

$y = ae^{bx} = 0.101\, e^{2.99x}$.

1) Fit an exponential curve of the form $y = ab^x$ to the following data:

x :	1	2	3	4	5	6	7	8
y :	1.0	1.2	1.8	2.5	3.6	4.7	6.6	9.1

x	y	$Y = \log_{10} y$	xY	x^2
1	1.0	0	1	1
2	1.2	0.0792	0.1584	4
3	1.8	0.2553	0.7659	9
4	2.5	0.3979	1.5916	16
5	3.6	0.5563	2.7815	25
6	4.7	0.6721	4.0326	36
7	6.6	0.8195	5.7365	49
8	9.1	0.9590	7.6720	64
36	30.5	3.7393	22.7385	204

$y = ab^x$

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

$\Rightarrow Y = A + xB$.

Normal equations are $\quad \Sigma Y = nA + B\Sigma x$.

$\qquad \Sigma xY = A\Sigma x + B\Sigma x^2$.

$3.7393 = 8A + 36B \rightarrow (1)$

$22.7385 = 36A + 204B \rightarrow (2)$

By solving above equations we get

$B = 0.1408 \qquad A = -0.1662.$

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

$a = 0.6820, \quad b = 1.3829.$

$y = ab^x = 0.682(1.38)^x.$

12) Predict the mean radiation dose at an altitude of 3000 feet by fitting an exponential curve to the given data:

Altitude(x) :	50	450	780	1200	4400	4800	5300
Dose of radiation :	28	30	32	36	51	58	69.

Exponential curve is $y = ab^x$

$$y = ab^x$$

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

$$\Rightarrow Y = A + xB.$$

$\therefore$ Normal equations are $\Sigma Y = nA + B\Sigma x.$

$\Sigma xY = A\Sigma x + B\Sigma x^2.$

x	y	$Y = \log_{10} y$	xY	x^2
50	28	1.4471	72.357	2500
450	30	1.4471	664.704	202500
780	32	1.5051	1174.017	608400
1200	36	1.5563	1867.563	1440000
4400	51	1.7075	7513.308	19360000
4800	58	1.7634	8464.454	23040000
5300	69	1.8388	9745.899	28090000
16980		11.2955	29502.305	72743400

$$11.295 = 7A + 16980B$$

$$29502.305 = 16980A + 72743400B.$$

By solving above equations we get

$A = 1.4521 \qquad B = 0.0000666$

$$Y = 1.4521 + 0.0000666x.$$

13) obtain a relation of the 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	33.1	65.2	127.4.

x	y	$Y = \log_{10} y$	xY	x^2
2	8.3	0.9191	1.8382	4
3	15.4	1.1875	3.5625	9
4	33.1	1.5198	6.0792	16
5	65.2	1.8142	9.0710	25
6	127.4	2.1052	12.631	36
20		7.5455	33.1819	90.

$$\boxed{\begin{aligned} y &= ab^x \\ \log_{10} y &= \log_{10} a + x \log_{10} b. \\ Y &= A + xB. \end{aligned}}$$

Normal equations are
$$\Sigma Y = nA + B\Sigma x.$$
$$\Sigma xY = A\Sigma x + B\Sigma x^2$$

$$7.5455 = 5A + 20B$$
$$33.1819 = 20A + 90B$$

$$A = 0.31, \quad B = 0.3$$

$A = \log_{10} a$, $\quad a = 10^A = 10^{0.31} = 2.04$

$B = \log_{10} b$, $\quad b = 10^B = 10^{0.3} = 1.995$.

$y = ab^x = 2.04\,(1.995)^x$.