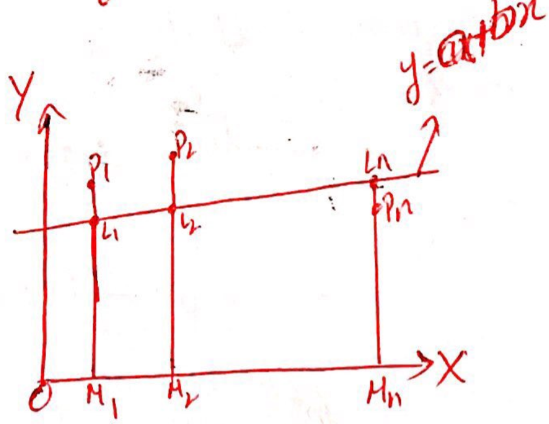
Curve Fitting

The process of finding the equation of the curve which best fits the given values & is most suitable for predicting the unknown values is known as curve fitting.

There are many methods of curve fitting.
(i) Graphical Method & (ii) Method of Grouped Averages
(iii) Method of Moments & (iv) Method of Least squares.

Method of Grouped Averages :-



Let the straight line

$$y = a + bx \quad \text{---} \quad (1)$$

fit the set of n observations (x_i, y_i) ; $i = 1, 2, 3, \ldots, n$.

Now from figure, when $x = x_1$, the observed value of $\boxed{y = y_1 = P_1 M_1}$ & expected value from (1) is $\boxed{y = a + bx_1 = L_1 M_1}$

Then $d_1 = y_1 - (a + bx_1) = P_1M_1 - L_1M_1 = P_1L_1$
which is called the residual or error at x_1

Ily, the residuals of remaining observations are
$$d_i = y_i - (a + bx_i) = P_iL_i \; , \; i = 1, 2, 3, \ldots, n.$$
From figure residuals may be +ve & others are -ve.

The method of group averages is based on the assum[ption] that the sum of the residuals is zero.

i.e. $\boxed{\sum_{i=1}^{n} d_i = 0}$

For finding values of a & b, we divide the given data into two groups:

(i) The Ist group containing k observations $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$

(ii) The 2nd group containing the remaining $(n-k)$ observations i.e $(x_{k+1}, y_{k+1}), (x_{k+2}, y_{k+2}), \ldots, (x_n, y_n)$.

Assuming that sum of the residuals in each group is zero, we get:

$$\sum_{i=1}^{k} d_i = \sum_{i=1}^{k} [y_i - (a + b x_i)] = 0 \quad \& \quad \text{—} (2)$$

$$\sum_{i=k+1}^{n} d_i = \sum_{i=k+1}^{n} [y_i - (a + b x_i)] = 0 \quad \text{—} (3)$$

On simplification, we get:

$$\frac{1}{k}\sum_{i=1}^{k} y_i = a + b \frac{1}{k}\sum_{i=1}^{k} x_i \Rightarrow \boxed{\bar{y}_1 = a + b\bar{x}_1} \quad \text{—} (3)$$

IIly $\boxed{\bar{y}_2 = a + b\bar{x}_2}$ — (4)

where $\bar{y}_2 = \frac{1}{n-k}\sum_{i=k+1}^{n} y_i$, $\bar{x}_2 = \frac{1}{n-k}\sum_{i=k+1}^{n} x_i$.

$\bar{x}_1$'s, $\bar{y}_1$'s & $\bar{x}_2$'s, $\bar{y}_2$'s are the averages of Ist & 2nd group resp.

<u>Remark</u>: Since the grouping can be done in diffn ways, the values of a & b are not unique.

obtain ... metho[d] ... data

Scanned with CamScanner

(2)

Obtain an eq^n of the form $y = a+bx$ using the method of group averages, from the following data:

x: 0 5 10 15 20 25

y: 12 15 17 22 24 30

Sol^n Let us divide the given data into two groups.

Group I	
x	y
0	12
5	15
10	17
$\Sigma x = 15$	$\Sigma y = 44$

Group II	
x	y
15	22
20	24
25	30
$\Sigma x = 60$	$\Sigma y = 76$

$\therefore$ The averages of group I are $\bar{x}_1 = \frac{15}{3} = 5$, $\bar{y}_1 = \frac{44}{3} = 14.667$

" " " " II are $\bar{x}_2 = \frac{60}{3} = 20$, $\bar{y}_2 = \frac{76}{3} = 25.333$.

Substituting the averages of x's & y's of the two groups in the reg^n line $y = a+bx$, we get

$$\bar{y}_1 = \bar{x}_1 b + a \Rightarrow \boxed{14.667 = a + 5b}$$

$$\bar{y}_2 = a + b\bar{x}_2 \Rightarrow \boxed{25.333 = a + 20b}$$

Solving the above two eq^{n}s, we get

$a = 11.1117$, $b = 0.7111$ (app.)

$\therefore$ The required line is $\boxed{y = 11.1117 + 0.7111x}$

Q27: The head of water H (ft) & the quantity of water Q (ft³) flowing per second are related by the law $Q = CH^n$. Find the best values of C & n by the method of group averages.

H: 1.2 1.4 1.6 1.8 2.0 2.4 2.6
Q: 4.2 6.1 8.5 11.5 14.9 23.5 27.1

Soln: Given eqn is $Q = CH^n$.
Taking log B.S, $\log_{10} Q = \log_{10} C + n \log_{10} H$
$\Rightarrow y = a + nx$ — (1)

Now ~~considering~~ dividing the data into two groups:

Group I	
$x = \log_{10} H$	$y = \log_{10} Q$
0.07918	0.62325
0.14163	0.78533
0.20412	0.92942
0.25527	1.06070
$\Sigma x = 0.6802$	$\Sigma y = 3.3987$

Group II	
$x = \log_{10} H$	$y = \log_{10} Q$
0.30103	1.17319
0.38021	1.37107
0.41497	1.43297
$\Sigma x = 1.09621$	$\Sigma y = 3.97723$

∴ The averages of group I are

~~$\bar{x}_1 = \frac{0.6802}{3.39787} = 0.17005$; $\bar{y}_1 = \frac{3.3987}{3.97723}$~~

$\bar{x}_1 = \frac{0.6802}{4} = 0.17005$; $\bar{y}_1 = \frac{3.3987}{4} = 0.84968$

group II: $\bar{x}_2 = \frac{1.09621}{3} = 0.3654$; $\bar{y}_2 = \frac{3.97723}{3} = 1.32574$.

Substituting the averages of x's, y's of two groups in (1) are;

$0.84968 = a + 0.17005n$

$1.32574 = a + 0.3654n$

Solving these eqn we get

$a = \cancel{2.43364593}$ 0.431308 ; $n = \cancel{9.33}$ 2.4608 (app.)

$\Rightarrow$ ~~C = antilog₁₀ a = 273.50124~~

$\boxed{C = 2.6996}$

Scanned with CamScanner

Problems involving three constants

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

Let (x_1, y_1) be a particular point on the curve, satisfying the data: $y_1 = a + bx_1 + cx_1^2$

$$\therefore \; y - y_1 = b(x - x_1) + c(x^2 - x_1^2)$$

$$\Rightarrow \boxed{\frac{y - y_1}{x - x_1} = b + c(x + x_1)} \text{ or } \boxed{Y = b + cX}$$

where $Y = \frac{y - y_1}{x - x_1}$, $X = x + x_1$

Ques Using the method of averages, fit a parabola $y = a + bx + cx^2$ to the following data:

x:	20	40	60	80	100	120
y:	5.5	9.1	14.9	22.8	33.3	46

Solⁿ Let the pt. (20, 5.5) lies on parabola

$y = a + bx + cx^2$ —— (1)

$5.5 = a + 20b + 400c$ —— (2)

(1)-(2) we get, $\frac{y - 5.5}{x - 20} = b + c(x + 20)$ —— (3)

$\Rightarrow \; Y = b + cX$

Now dividing the given data into two groups, we get

Group I	
$X = x + 20$	$Y = \frac{y - 5.5}{x - 20}$
—	—
60	0.18
80	0.235
$\Sigma X = 140$	$\Sigma Y = 0.415$

Group II	
$X = x + 20$	$Y = \frac{y - 5.5}{x - 20}$
100	0.288
120	0.348
140	0.405
$\Sigma X = 360$	$\Sigma Y = 1.041$

$\overline{X_1} = \frac{140}{2} = 70$, $\overline{Y_1} = 0.2075$; $\overline{X_2} = \frac{360}{3} = 120$, $\overline{Y_2} = 0.347$

$\overline{Y_1} = b + c\overline{X_1}$ i.e $0.2075 = b + 70c$

& $\overline{Y_2} = b + c\overline{X_2}$ i.e $0.347 = b + 120c$.

By solving we get; $b = 0.0122$; $c = 0.00279$

Put these values in (3), we get

$$\frac{y-5.5}{x-20} = 0.0122 + 0.00279(x+20)$$

$$\Rightarrow \boxed{y = 0.00279x^2 + 0.0122x + 4.14}$$

(2) The curve of the form $y = ax^b + c$

Ques In magnetic arc at constant arc length the voltage V consumed by the arc is observed for values of the current i.

i:	0.5	1	2	4	8	12
V:	160	120	94	75	62	56

If V & i are connected by a relation of the form $V = ai^b + c$, find a, b & c & hence find curve.

Soln Now $V = ai^b + c \Rightarrow \boxed{V - c = ai^b}$ — *

Taking log B.S, we get $\log_{10}(V-c) = \log_{10}a + b\log_{10}i$

$$\Rightarrow \boxed{Y = A + bX}$$

$Y = \log_{10}(V-c)$, $A = \log_{10}a$, $X = \log_{10}i$

Take $i_1 = 1$, $i_2 = 2$ & $i_3 = 4$ which are in G.P i.e $\boxed{i_2^2 = i_1 i_3}$ to find c. Then $V_1 = 120$, $V_2 = 94$ & $V_3 = 75$

$$\therefore c = \frac{V_1V_3 - V_2^2}{V_1 + V_3 - 2V_2} = \frac{(120)(75) - (94)^2}{120 + 75 - 2(94)} = 23.4$$

$$\therefore V = ai^b + 23.4 \Rightarrow \boxed{V - 23.4 = ai^b} \quad \text{— (1)}$$

$$\Rightarrow \boxed{Y = A + bX} \quad \text{— (2)}$$

$\log_{10}(V-23.4)$, $A = \log_{10}a$ & $X = \log_{10}i$

Scanned with CamScanner

Divide the given data into two groups:

Group I	
$X = \log_{10} i$	$Y = \log_{10}(V - 23.4)$
-0.3010	2.1354
0.000	1.9850
0.3010	1.8488
$\Sigma X = 0$	$\Sigma Y = 5.9692$

Group II	
X	Y
0.6021	1.7126
0.9031	1.5866
1.0792	1.5132
$\Sigma X = 2.5844$	$\Sigma Y = 4.8124$

$\bar{X}_1 = 0$; $\bar{Y}_1 = 1.9897$; $\bar{X}_2 = 0.8615$, $\bar{Y}_2 = 1.6041$

Averages equations are :-

$\bar{Y}_1 = A + b\bar{X}_1 \Rightarrow 1.9897 = A + (0)b \Rightarrow \boxed{A = 1.9897}$

$\bar{Y}_2 = A + b\bar{X}_2 \Rightarrow 1.6041 = A + 0.8615b \Rightarrow \boxed{b = -0.4476}$

$\therefore \boxed{a = \text{antilog}_{10} A = 97.66}$, $\boxed{c = 23.4}$, $\boxed{b = -0.45}$

& The curve is $\boxed{V = 97.66\, i^{-0.45} + 23.4}$

Points to remember:

From (*) $V_1 - c = a i_1^b$, $V_2 - c = a i_2^b$, $V_3 - c = a i_3^b$

Now $i_2^2 = i_1 i_3$.

$\Rightarrow (V_1 - c)(V_3 - c) = a^2 (i_1 i_3)^b = (a i_2^b)^2 = (V_2 - c)^2$

After simplification, we get:

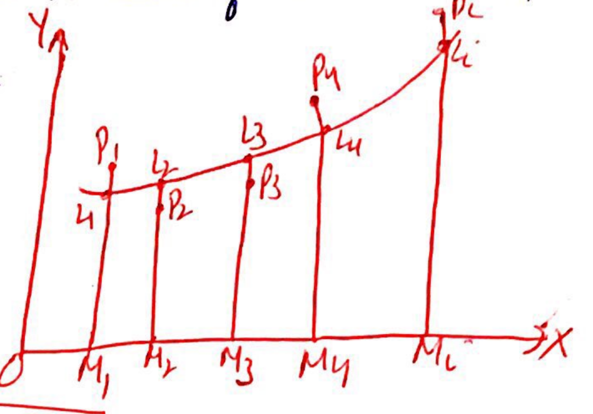
$$c = \frac{V_1 V_3 - V_2^2}{V_1 + V_3 - 2V_2}.$$

Principle Of Least Squares

It provides unique set of values to the constants.

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be the n sets of observations

Let $y = f(x)$ —— (1) be relation b/w x & y



when $x = x_i$, observed value of y_i

$\boxed{y_i = P_i M_i}$

& $\boxed{\text{Expected value} = L_i M_i = f(x_i)}$

∴ Error at $x = x_i$ is $\boxed{e_i = y_i - f(x_i)}$.

From figure, some errors may be +ve or may be −ve.

Thus by giving equal weightage to each error consider

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

The minimum of E results the best fitting curve to the data.

① Fitting a straight line $y = ax + b$.

Let (x_i, y_i) be n sets of observations; $i = 1, 2, \ldots, n$.

The residuals at $x = x_i$ is given by:

$$e_i = y_i - f(x_i) = y_i - (ax_i + b), \quad i = 1, 2, \ldots, n$$

$$\therefore E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2$$

By the principle of least squares, E is minimum.

$$\therefore \frac{\partial E}{\partial a} = 0 \quad \& \quad \frac{\partial E}{\partial b} = 0$$

$$\therefore \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) = 0 \quad \& \quad \sum_{i=1}^{n} 2[y_i - ax_i - b](-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i x_i - ax_i^2 - bx_i) = 0 \quad \& \quad \sum_{i=1}^{n} (y_i - ax_i - b) = 0$$

$$\Rightarrow \boxed{\sum_{i=1}^{n} y_i x_i = a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i} \quad \& \quad \boxed{\sum_{i=1}^{n} y_i = a \sum_{i=1}^{n} x_i + nb}$$

By dropping the suffix i, we get:

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

$$\& \quad \boxed{\Sigma y = a \Sigma x + nb}$$

These are called normal eqn's for straight line.

② Fitting a parabola $y = ax^2 + bx + c$.

The residuals at $x = x_i$ for n observations is given by:- $e_i = y_i - f(x_i) = y_i - (ax_i^2 + bx_i + c)$

$$\therefore E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)]^2$$

Now E ~~should~~ is minimum by the principle of least squares!

$$\therefore \frac{\partial E}{\partial a} = 0, \quad \frac{\partial E}{\partial b} = 0 \quad \& \quad \frac{\partial E}{\partial c} = 0$$

ie $\sum_{i=1}^{n} 2[y_i - (ax_i^2 + bx_i + c)](-x_i^2) = 0$

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

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

On simplification & dropping suffix i, we get.

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

These are called normal eqn's for parabola.

3) Fitting an exponential curve : $\boxed{y = ae^{bx}}$

Taking $\log_{10}$ B.S, we get

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

$\Rightarrow \boxed{Y = A + Bx}$

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

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

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

4) Fitting the curve $y = ax^b$

Taking log B-S, we get

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

$$\boxed{Y = A + bX}$$

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

$\therefore$ Normal equations are :

$$\boxed{\begin{array}{l} \Sigma Y = nA + b\Sigma X \\ \Sigma XY = A\Sigma X + b\Sigma X^2 \end{array}}$$

Ques 17 : By the method of least squares, find the straight line that best fits the given data :

x : 1 2 3 4 5

y : 14 27 40 55 68

Soln Let the straight line of best be

$$y = ax + b \quad \text{—— (1)}$$

$\therefore$ Normal equations are : $\Sigma xy = a\Sigma x^2 + b\Sigma x$ — (2)

$\Sigma y = a\Sigma x + 5b$ — (3)

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

Put the values of Σxy, Σy, Σx, Σx^2 in eqn's (2) & (3), we get

$\therefore$ The equations (2) & (3) becomes

$$748 = 55a + 15b$$
$$204 = 15a + 5b$$

Solving these, we have $\boxed{a = 13.6,\ b = 0}$

Putting these values in eqn (1), we get

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

Q27 Fit a parabola $y = ax^2 + bx + c$ in least square to the given data

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

Soln The normal equations to the curve $y = ax^2 + bx + c$ are:

$$\Sigma y = a\Sigma x^2 + b\Sigma x + 5c$$
$$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x$$
$$\Sigma x^2 y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2$$

The values of Σy, Σxy, $\Sigma x^2 y$, Σx -- are calculated by means of the following table:

x	y	x^2	x^3	x^4	xy	x^2y
10	14	100	1000	10,000	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
$\Sigma x = 80$	$\Sigma y = 100$	$\Sigma x^2 = 1398$	$\Sigma x^3 = 26270$	$\Sigma x^4 = 521202$	$\Sigma xy = 1684$	$\Sigma x^2y = 30648$

Substitutes these values in normal equations.

$$100 = 1398a + 80b + 5c$$
$$1684 = 26270a + 1398b + 80c$$
$$30648 = 521202a + 26270b + 1398c$$

On solving, we get:-
$a = -0.07, \quad b = 3.03, \quad c = -8.89$

$\therefore$ The required eqn is $\boxed{y = -0.07x^2 + 3.03x - 8.89}$

Ques37 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

Soln Given curve $y = ae^{bx}$
Taking log B.S, we get
$\log y = \log a + bx \log_{10} e$
$\Rightarrow Y = A + Bx$ —(1) $\quad Y = \log y, \quad A = \log a$
$B = b \log_{10} e$

The normal equations of (1) is given by :-
$$\Sigma Y = 5A + B\Sigma x$$
$$\Sigma xY = A\Sigma x + B\Sigma x^2$$

x	y	$Y = \log_{10} y$	x^2	xY
1	10	1.0000	1	1
5	15	1.1761	25	5.8805
7	12	1.0792	49	7.5544
9	15	1.1761	81	10.5849
12	21	1.3222	144	15.8664
$\Sigma x = 34$		$\Sigma Y = 5.7536$	$\Sigma x^2 = 300$	$\Sigma xY = 40.8862$

Substitut
we get

(4)

Substitutes these values in the normal equations, we get:-

$$5.7536 = 5A + 34B$$
$$40.8862 = 34A + 300B$$

On solving, we get:

$A = 0.9766$; $B = 0.02561$

$\therefore$ $a = \text{antilog} A = 9.4754$

$b = \frac{B}{\log_{10} e} = 0.059$

$\therefore$ Req. curve is $y = 9.4754\, e^{0.059x}$

Ques 47 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

Sol^n The curve to be fitted is $y = ab^x$

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

$\Rightarrow Y = A + Bx$; $Y = \log y$, $A = \log a$, $B = \log b$.

$\therefore$ The normal equations are

$$\Sigma Y = 5A + B\Sigma x$$
$$\Sigma xY = A\Sigma x + B\Sigma x^2$$

x	y	$Y = \log y$	x^2	xY
2	8.3	0.9191	4	1.8382
3	15.4	1.1872	9	3.5616
4	33.1	1.5198	16	6.0792
5	65.2	1.8142	25	9.0710
6	127.4	2.1052	36	12.6312
$\Sigma x = 20$		$\Sigma Y = 7.5455$	$\Sigma x^2 = 90$	$\Sigma xY = 33.1812$

Substituting the required values from table in the normal equations, we get!

$$7.5455 = 5A + 20B$$
$$33.1812 = 20A + 90B$$

On solving, we get! $A = 0.31$ & $B = 0.3$

$\therefore$ $a = \text{antilog} A = 2.04$ & $b = \text{antilog} B = 1.995$

Hence the required curve is

$$\boxed{y = 2.04(1.995)^x}$$

Ques 5] Fit a curve of form $y = ae^{bx}$ to the data!

x	0	2	4
y	5.102	10	31.62

Sol^n Given curve is $y = ae^{bx}$.

$\Rightarrow \log y = \log a + bx \log e$

$\Rightarrow Y = A + Bx$; $Y = \log y$, $A = \log a$, $B = b \log e$

$B = 0.4343b$

$\therefore$ The normal equations are!

$$\Sigma Y = 3A + B\Sigma x$$
$$\Sigma xY = A\Sigma x + B\Sigma x^2$$

x	y	$Y = \log y$	x^2	xY
0	5.102	0.70774	0	0
2	10	1	4	2
4	31.62	1.49996	16	5.9998
$\Sigma x = 6$		$\Sigma Y = 3.2077$	20	7.9998

Substituting the req. values from table in the normal equations, we get!

$$3.2077 = 3A + 6B$$
$$7.9998 = 6A + 20B$$

On Solving, we get:-

$B = 0.19805$, $\quad A = 0.6731$

$\Rightarrow a = \text{antilog} A = \text{antilog}(0.6731) = 4.642$

$\Rightarrow b = \dfrac{B}{0.4343} = \dfrac{0.19805}{0.4343} = 0.456$ (app.)

$\therefore$ Required curve is $\boxed{y = 4.642\, e^{0.456}}$