Chapter 8

UNIT-V

APPLIED STATISTICS : CURVE FITTING (METHOD OF LEAST SQUARES)

§ 8.1. CURVE FITTING

Suppose we have collected some data and it is desired to find out the form of universe of which the observed values are regarded as a sample, *i.e.*, in other words we find (if possible) a functional relationship between the observed values. It is called *fitting of a curve* to a given data.

Definition. *Curve fitting is a general problem to find the equation of an approximate curve fitting to the given data.*

Suppose we are given a set $\{(x_i, y_i), i = 1, 2, \ldots, n\}$ of n pairs of values where x is independent variable and y is dependent variable; then by curve fitting we mean to find a functional relation of the form $y = f(x)$ suggested by the given data.

Curve fitting is very important to theoretical as well as to practical statistics. In theoretical statistics, it is useful since the lines of regression can be regarded as fitting of linear curves to the given bivariate values. In practical statistics we are required to find a functional relationship between two variables by simple algebraic expressions, for example: polynomials, exponential and logarithmic functions.

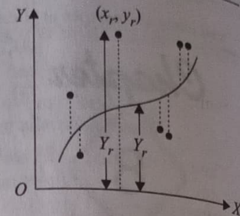
§ 8.2. THE METHOD OF LEAST SQUARES

Suppose we are required to fit a nth degree curve

$$y = a + bx + cx^2 + \ldots kx^n \qquad \ldots(1)$$

to the given values (x_i, y_i), $i = 1, 2, \ldots, m$ of two variables x and y. Now we are required to find $(n + 1)$ constants namely $a, b, c, \ldots, k$ so that (1) represents the curve of best fit of that degree.

If $m = n + 1$, we get $(n + 1)$ equations by substituting the values of (x_i, y_i) in (1) and a unique solution of the values $a, b, c, \ldots, k$ is possible.



If $m > n + 1$, then no unique solution is possible and then we shall try to find those values of $a, b, c, \ldots, k$ which may give the best fit; for such cases the method of least squares is used.

Let $Y_1, Y_2, \ldots, Y_m$ be the expected values of y corresponding to the values $x_1, x_2, \ldots, x_m$ of x. Suppose that $y_1, y_2, \ldots, y_m$ are observed values of y corresponding to the values $x_1, x_2, \ldots, x_m$ of x.

Usually, the values of $Y_1, Y_2, \ldots, Y_m$ are different as the observed points do not necessarily lie on the approximating curve.

Now substituting values $x_1, x_2, \ldots, x_m$ of x in (1) successively, we have

$$\left.\begin{aligned} Y_1 &= a + bx_1 + cx_1^2 + \ldots + kx_1^n \\ Y_2 &= a + bx_2 + cx_2^2 + \ldots + kx_2^n \\ &\ldots \quad \ldots \quad \ldots \quad \ldots \\ Y_m &= a + bx_m + cx_m^2 + \ldots + kx_m^n \end{aligned}\right\}. \qquad \ldots(2)$$

The difference of expected values $(Y_r, r = 1, 2, \ldots, m)$ and observed values $(y_r, r = 1, 2, \ldots, m)$ of y are called **residuals.** Thus we write

$$R_r = y_r - Y_r, \quad r = 1, 2, \ldots, m$$

where R_r is called rth residual.

Now consider a quantity U defined by

$$U = R_1^2 + R_2^2 + \ldots + R_m^2 = \sum_{r=1}^{m} R_r^2. \qquad \ldots(3)$$

Here U is called the **sum of squares of residuals.** The method of least squares asserts that **U should be minimum** for the best fitting of curve.

A curve, which has the properties of best fitting curve in the sense of least square of given data, is called a least square curve.

Now from (3), we have

$$U = R_1^2 + R_2^2 + \ldots + R_m^2 = \sum_{r=1}^{m} R_r^2 = \sum_{r=1}^{m} (y_r - Y_r)^2$$

$$= \sum_{r=1}^{m} (y_r - a - bx_r - cx_r^2 \ldots - kx_r^n)^2$$

or $\qquad U = U(a, b, c, \ldots, k)$ (say).

As stated above the principle of least squares asserts that the constants $a, b, c, \ldots, k$ are determined so that U is minimum.

Now the necessary conditions for U to be maximum or minimum are given by

$$\frac{\partial U}{\partial a} = 0, \quad \frac{\partial U}{\partial b} = 0, \quad \frac{\partial U}{\partial c} = 0, \quad \ldots, \quad \frac{\partial U}{\partial k} = 0.$$

Now simplifying these relations, we get

$$\frac{\partial U}{\partial a} = 0$$

$$\Rightarrow \quad \Sigma 2(y_r - a - bx_r - cx_r^2 \ldots - kx_r^n).(-1) = 0$$

$$\Rightarrow \quad \Sigma y_r - \Sigma a - b\Sigma x_r - c\Sigma x_r^2 - \ldots - k\Sigma x_r^n = 0$$

Prof. Lokesh Sharma

$$\begin{aligned} \Rightarrow \quad & \Sigma y = ma + b\Sigma x + \ldots + k\Sigma x^n \\ \therefore \quad & \Sigma xy = a\Sigma x + b\Sigma x^2 + \ldots + k\Sigma x^{n+1} \\ \text{Similarly} \quad & \Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + \ldots + k\Sigma x^{n+2} \\ & \Sigma x^n y = a\Sigma x^n + b\Sigma x^{n+1} + \ldots k\Sigma x^{2n} \end{aligned} \Bigg\}. \qquad \ldots(4)$$

The $(n + 1)$ equations given by (4) are called normal equations. By solving these equations as simultaneous equations, the values of $(n + 1)$ unknowns (*i.e.*, constants) $a, b, c, \ldots, k$ are determined. It can be verified that when second order partial derivatives $\frac{\partial^2 U}{\partial a^2}, \ldots, \frac{\partial^2 U}{\partial a\, \partial b}, \ldots$, etc. are calculated and above values of $a, b, c, \ldots, k$ are substituted, we get +ve value of the function and so U is minimum.

§ 8.3. PARTICULAR CASES

When $n = 1$ or fitting of straight line.

When $n = 1$, the curve to be fitted is a straight line. Let its equation be $y = a + bx$ and its normal equations are as follows :

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

When $n = 2$ or fitting of second degree parabola. When $n = 2$, the curve to be fitted is a second degree parabola

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

and its normal equations are given by

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

The values of $\Sigma x, \Sigma y, \Sigma xy, \ldots$, etc. are calculated by means of table (as we have done in calculating S.D.) and then the values of $a, b, c, \ldots$ are determined.

In the following art, we shall discuss the method of *fitting a straight line* and *fitting a second degree parabola independently.*

§ 8.4. FITTING A STRAIGHT LINE

Suppose, we are required to a fit a straight line

$$y = a + bx \qquad \ldots(1)$$

to the given set of values (*i.e.*, observations) (x_i, y_i), $i = 1, 2, \ldots, m$ of two variables x and y. Now we are to determine 2 constants namely a and b so that (1) represents the straight line of best fit.

Let $Y_1, Y_2, \ldots, Y_m$ be the *expected values* of y corresponding to the values $x_1, x_2, \ldots, x_m$ of x, so that

$$\left.\begin{aligned} Y_1 &= a + bx_1, \\ Y_2 &= a + bx_2, \\ &\ldots \ldots \ldots \\ Y_m &= a + bx_m \end{aligned}\right\} \qquad \ldots(2)$$

Since $y_1, y_2, \ldots, y_m$ are the observed values of y, so that residual at $x = x_r$ i.e., rth residual R_r is

$$R_r = y_r - Y_r,\ r = 1, 2, \ldots, m.$$

Now consider a quantity U-defined by

$$U = \sum_{r=1}^{m} R_r^2 = \sum_{r=1}^{m} (y_r - Y_r)^2 = \sum_{r=1}^{m} (y_r - a - bx_r)^2.$$

Now by the principle at least squares, U should be minimum, the conditions for which are :

$$\frac{\partial U}{\partial a} = 0 \text{ and } \frac{\partial U}{\partial b} = 0$$

i.e., $$\sum_{r=1}^{m} 2(y_r - a - bx_r)(-1) = 0$$

and $$\sum_{r=1}^{m} 2(y_r - a - bx_r)(-x_r) = 0$$

i.e., $$\Sigma y = ma + b\Sigma x \quad \ldots(3)$$

and $$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \ldots(4)$$

The equations (3) and (4) are the **normal equations** and by solving these equations the values of a and b are determined. By substituting these values of a and b in (1), we get the required line of best fit.

◈ § 8.5. FITTING A SECOND DEGREE PARABOLA

Suppose, we are required to fit a second degree parabola

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

to the given set of values (*i.e.*, observations) (x_i, y_i), $i = 1, 2, \ldots, m$ of two variables x and y. Now we are to determine 3 constants namely a, b and c so that (1) represents the second degree parabola of best fit.

Let $Y_1, Y_2, \ldots, Y_m$ are the *expected values* of y corresponding to the values $x_1, x_2, \ldots, x_m$ of x, so that

$$\left.\begin{aligned} Y_1 &= a + bx_1 + cx_1^2, \\ Y_2 &= a + bx_2 + cx_2^2, \\ &\ldots\ \ldots\ \ldots\ \ldots \\ Y_m &= a + bx_m + cx_n^2 \end{aligned}\right\} \quad \ldots(2)$$

Since $y_1, y_2, \ldots, y_m$ are the *observed values* of y, so that residual at x_r *i.e.*, rth residual R_r is

$$R_r = y_r - Y_r,\ r = 1, 2, \ldots, m.$$

Now consider a quantity U defined by

$$U = \sum_{r=1}^{m} R_r^2 = \sum_{r=1}^{m} (y_r - Y_r)^2 = \sum_{r=1}^{m} (y_r - a - bx_r - cx_r^2)^2.$$

Prof. Lokesh Sharma

Now by the principle of least squares, U should be minimum, the conditions for which are :

$$\frac{\partial U}{\partial a} = 0,\ \frac{\partial U}{\partial b} = 0 \text{ and } \frac{\partial U}{\partial c} = 0$$

i.e., $$\sum_{r=1}^{m} 2(y_r - a - bx_r - cx_r^2)(-1) = 0$$

$$\sum_{r=1}^{m} 2(y_r - a - bx_r - cx_r^2)(-x_r) = 0$$

and $$\sum_{r=1}^{m} 2(y_r - a - bx_r - cx_r^2)(-x_r^2) = 0$$

i.e., $$\Sigma y = ma + b\Sigma x + c\Sigma x^2 \quad \ldots(3)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^2 \quad \ldots(4)$$

and $$\Sigma xy = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad \ldots(5)$$

The values of $\Sigma x, \Sigma y, \Sigma x^2, \Sigma xy$ etc. are calculated by means of table for the given set of observations (x_i, y_i).

The equations (3), (4) and (5) are the **normal equations** and by solving these equations, the values of a, b and c are determined. By substituting these values of a, b and c in (1), we get the required second degree parabola of best fit.

ILLUSTRATIVE EXAMPLES

Example 1. ***Fit a straight line to the following data regarding x as the independent variable :***

x:	0	1	2	3	4
y:	1	1·8	3·3	4·5	6·3

Solution. Suppose a straight line to be fitted to the given data is as follows:

$$y = a + bx \quad \ldots(i)$$

then the normal equations are as :

$$\Sigma y = ma + b\Sigma x \quad \ldots(ii)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2. \quad \ldots(iii)$$

Now from the given data, we have

x	y	xy	x^2
0	1	0	0
1	1·8	1·8	1
2	3·3	6·6	4
3	4·5	13·5	9
4	6·3	25·2	16
Total $\Sigma x = 10$	$\Sigma y = 16·9$	$\Sigma xy = 47·1$	$\Sigma x^2 = 30$

Here $$m = 5.$$

Prof. Lokesh Sharma

Now substituting these values in the normal equations, we get

$$16\cdot 9 = 5a + 10b \quad \text{...(iv)}$$
$$47\cdot 1 = 10a + 30b \quad \text{...(v)}$$

Solving (iv) and (v), we have

$$a = 0\cdot 72 \quad \text{and} \quad b = 1\cdot 33.$$

Thus the required equation of the straight line is

$$y = 0\cdot 72 + 1\cdot 33x \quad \text{[putting for } a, b \text{ in (i)]}$$

Example 2. ***Fit a second degree parabola to the following data regarding x as an independent variable***

x:	0	1	2	3	4
y:	1	5	10	22	38

Solution. Let the equation of second degree parabola to be fitted to the given data be

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

then its normal equations are

$$\Sigma y = ma + b\Sigma x + c\Sigma x^2 \quad \text{...(ii)}$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad \text{...(iii)}$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad \text{...(iv)}$$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	5	1	1	1	5	5
2	10	4	8	16	20	40
3	22	9	27	81	66	198
4	38	16	64	256	152	608
Total $\Sigma x = 10$	$\Sigma y = 76$	$\Sigma x^2 = 30$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$	$\Sigma xy = 243$	$\Sigma x^2 y = 851$

Substituting these values in the normal equations, we get

$$76 = 5a + 10b + 30c \quad [\because m = 5] \quad \text{...(1)}$$
$$243 = 10a + 30b + 100c \quad \text{...(2)}$$
$$851 = 30a + 100b + 354c \quad \text{...(3)}$$

Solving the equations (1), (2) and (3), we have

$$a = 1\cdot 43, b = 0\cdot 24, c = 2\cdot 21 \quad \text{[from (i)]}$$

$\therefore$ The required equation of second degree parabola is,

$$y = 1\cdot 43 + 2\cdot 24x + 2\cdot 21x^2.$$

Example 3. ***Fit a straight line to the following data***

x:	71	68	73	69	67	65	66	67
y:	69	72	70	70	68	67	68	64

Solution. Suppose a straight line to be fitted to the given data is

$$y = a + bx. \quad \text{...(i)}$$

Then the normal equations are as :

$$\Sigma y = ma + b\,\Sigma x \quad \text{...(ii)}$$
$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2. \quad \text{...(iii)}$$

Now from the given data, we have

x	y	xy	x^2
71	69	4899	5041
68	72	4896	4624
73	70	5110	5329
69	70	4830	4761
67	68	4556	4489
65	67	4355	4225
66	68	4488	4356
67	64	4288	4489
$\Sigma x = 546$	$\Sigma y = 548$	$\Sigma xy = 37422$	$\Sigma x^2 = 37314$

Here $m = 8$. Now substituting these values in normal equations (ii) and (iii), we get

$$548 = 8a + 546b \quad \text{...(iv)}$$
$$37422 = 546a + 37314b \quad \text{...(v)}$$

Multiplying (iv) by 273, (v) by 4 and then subtracting, we get

$$273\times 548 - 4\times 37422 = (273\times 546 - 4\times 37314)\,b$$
$$\Rightarrow \quad -84 = -198b \Rightarrow b = 0\cdot 4242.$$

Put for b in (iv), $8a = 548 - 546\times 0\cdot 4242 \Rightarrow a = \dfrac{316\cdot 3868}{8} \Rightarrow a = 39\cdot 5484.$

Thus the required straight line of best fit is

$$y = 39\cdot 5484 + 0\cdot 4242\,x. \quad \textbf{Ans.}$$

Example 4. ***If F is the pull required to left a load W by means by pulley, fit a linear law F = a − bW connecting F and W against the following data :***

W :	50	70	100	120
F :	12	15	21	25

Solution. The linear law to be fitted is (given)

$$F = a - bW. \quad \text{...(i)}$$

The normal equations are as :

$$\Sigma F = ma - b\,\Sigma W \quad \text{...(ii)}$$
$$\Sigma WF = a\,\Sigma W - b\,\Sigma W^2. \quad \text{...(iii)}$$

Now from the given data, we have

W	F	WF	W^2
50	12	600	2500
70	15	1050	4900
100	21	2100	10000
120	25	3000	14400
$\Sigma W = 340$	$\Sigma F = 73$	$\Sigma WF = 6750$	$\Sigma W^2 = 31800$

Here $m = 4$. Now substituting these values in normal equations (ii) and (iii), we have

$$73 = 4a - 340b \quad \text{...(iv)}$$
$$6750 = 340a - 31800b. \quad \text{...(v)}$$

Solving (iv) and (v), $a = 2 \cdot 2785$, $b = -0 \cdot 1879$.

Thus the required law of best fit is $F = 2 \cdot 2786 + 0 \cdot 1879\ W$. **Ans.**

Example 5. ***Employ the method of least squares to fit a parabola*** $y = a + bx + cx^2$ ***in the following data :***

$$(x, y) : (-1, 2), (0, 0), (0, 1), (1, 2).$$

Solution. The equation of parabola to be fitted is

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

Its normal equations are : $\Sigma y = ma + b\,\Sigma x + c\,\Sigma x^2$ ...(ii)

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3 \quad \text{...(iii)}$$
$$\Sigma x^2 y = a\,\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4 \quad \text{...(iv)}$$

From the given data, we have

x	y	x^2	x^3	x^4	xy	x^2y
−1	2	1	−1	1	−2	2
0	0	0	0	0	0	0
0	1	0	0	0	0	0
1	2	1	1	1	2	2
$\Sigma x = 0$	$\Sigma y = 5$	$\Sigma x^2 = 2$	$\Sigma x^3 = 0$	$\Sigma x^4 = 2$	$\Sigma xy = 0$	$\Sigma x^2 y = 4$

Here $m = 4$. Substituting these values in the normal equations (ii), (iii) and (iv), we get

$$\left.\begin{aligned} 5 &= 4a + 2c \\ 0 &= 2b \\ 4 &= 2a + 2c \end{aligned}\right\}. \text{ Solving, } a = 0 \cdot 5,\ b = 0,\ c = 1 \cdot 5.$$

Substituting values in (i), the required equation of parabola is

$$y = 0 \cdot 5 + 1 \cdot 5\,x^2.$$ **Ans.**

◈ § 8.6. CHANGE OF ORIGIN

When the values of x are equidistant, *i.e.*, are of equal intervals, then the normal equations can be simplified.

Let the values of the variate x be $x, x + h, x + 2h, \ldots$, *i.e.*, length of interval is h.

Now we consider two cases :

Case I. When m is odd, say $m = 2n + 1$.

In this case we take the origin at the middle of the values of x and h as the unit of measurement. Let a be the middle of the values of x; then $u = (x - a)/h$ assumes the values $-n, \ldots, -1, 0, 1, \ldots, n$ and

$$\Sigma u = 0 = \Sigma u^3 = \Sigma u^5 = \ldots$$

Prof. Lokesh Sharma

Case II. When m is even, say $m = 2n$.

In this case we take the origin at the mean of the middle pair of values of x and $\frac{1}{2}h$ as the unit of measurement. In this case, we get

$$\Sigma u = 0 = \Sigma u^3 = \Sigma u^5 = \ldots.$$

ILLUSTRATIVE EXAMPLES

Example 1. ***Fit a second degree parabola to the following data :***

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

When $x = 2$, find the difference between the real value of y and the values of y obtained from the fitted curve.

Solution. Here $m = 5$, which is odd and the values of x are equidistant. Therefore we take the origin for x-series at the middle value 2.

Now suppose $X = x - 2$ and $Y = y$; then equation of the parabola of 2nd degree to be fitted is of the form given by

$$Y = a + bX + cX^2 \quad \text{...(1)}$$

Thus its normal equations are :

$$\Sigma Y = 5a + b\Sigma X + c\Sigma X^2 \quad \text{...(2)}$$
$$\Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3 \quad \text{...(3)}$$
$$\Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 \quad \text{...(4)}$$

x	y	X	Y	XY	X^2	X^2Y	X^3	X^4
0	1	−2	1	−2	4	1	−8	16
1	1·8	−1	1·8	−1·8	1	1·8	−1	1
2	1·3	0	1·3	0	0	0	0	0
3	2·5	1	2·5	2·5	1	2·5	1	1
4	6·3	2	6·3	12·6	4	25·2	8	16
Total = 10	12·9	0	12·9	11·3	10	33·5	0	34

Substituting these values from the table in the normal equations, we get

$$12 \cdot 9 = 54 + 10c \quad \text{...(1)}$$
$$11 \cdot 3 = 10b \quad \text{...(2)}$$
$$33 \cdot 5 = 10a + 34c \quad \text{...(3)}$$

Solving equations (1), (2) and (3), we have

$$a = 1 \cdot 48,\ b = 1 \cdot 13,\ c = 0 \cdot 55.$$

∴ The required equation is [putting values in (1)]

$$Y = 1 \cdot 48 + 1 \cdot 13X + 0 \cdot 55X^2. \quad \text{...(4)}$$

Now changing the origin, *i.e.*, putting $X = x - 2$ and $Y = y$, in (4), we get

$$y = 1 \cdot 48 + 1 \cdot 13(x - 2) + 0 \cdot 55(x - 2)^2$$

or $$y = 1 \cdot 42 - 1 \cdot 07x + 0 \cdot 55x^2. \quad \text{...(5)}$$

Prof. Lokesh Sharma

Now putting $x = 2$, from (5), we have

$$(y)_{x=2} = 1 \cdot 42 - 1 \cdot 07(2) + 0 \cdot 55(2)^2$$
$$= 1 \cdot 42 - 2 \cdot 14 + 2 \cdot 20 = 1 \cdot 48.$$

The required difference $= (y)_{x=2} - y = 1 \cdot 48 - 1 \cdot 30 = 0 \cdot 18$

[$\because$ $y = 1 \cdot 30$ at $x = 2$ from the table]

Example 2. ***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

Solution. Here $m = 9$, *i.e.*, odd, therefore the origin in x series will be taken at the mid-value of x, *i.e.*, at 5. Let the origin in y-series be taken at 7. Now consider new variables X and Y defined by

$$X = x - 5,\ Y = y - 7.$$

Let the equation of the parabola (*i.e.*, curve) to fit be

$$Y = a + bX + cX^2. \qquad \text{...(i)}$$

Hence its normal equations are

$$\Sigma Y = 9a + b\Sigma X + c\Sigma X^2 \qquad \text{...(ii)}$$
$$\Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3 \qquad \text{...(iii)}$$
$$\Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 \qquad \text{...(iv)}$$

x	y	X	Y	XY	X^2	X^2Y	X^3	X^4
1	2	−4	−5	20	16	−80	−64	256
2	6	−3	−1	3	9	−9	−27	81
3	7	−2	0	0	4	0	−8	16
4	8	−1	1	−1	1	1	−1	1
5	10	0	3	0	0	0	0	0
6	11	1	4	4	1	4	1	1
7	11	2	4	8	4	16	8	16
8	10	3	3	9	9	27	27	81
9	9	4	2	8	16	32	64	256
Total = 45	74	0	11	51	60	−9	0	708

Substituting the values from the table, we have

$$11 = 9a + 60c \qquad \text{...(1)}$$
$$51 = 60b \qquad \text{...(2)}$$
$$-9 = 60a + 708c. \qquad \text{...(3)}$$

Solving these equations, we get

$$a = 3,\ b = 0 \cdot 85,\ c = -0 \cdot 27.$$

The equation of parabola (curve) to fit is [putting values in (i)]

$$Y = a + bX + cX^2$$

or $$Y = 3 + 0 \cdot 85X - 0 \cdot 27X^2$$

$$y - 7 = 3 + 0 \cdot 85(x - 5) - 0 \cdot 27(x - 5)^2$$

or $$y - 7 = 3 + 0 \cdot 85x - 4 \cdot 25 - 0 \cdot 27x^2 + 2 \cdot 7x - 9 \cdot 75$$

or $$y = -1 + 3 \cdot 55x - 0 \cdot 27x^2.$$

Example 3. ***Fit the straight line to the following data :***

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

Solution. Here $m = 6$, *i.e.*, even and the values of x are equally spaced, h (common difference in x-series) = 5. Therefore, to make the calculation easier, we take the unit of measurement $\frac{1}{2}h$, *i.e.*, $2 \cdot 5$; and the origin at the mean of two middle terms 10 and 15, *i.e.*, origin is taken at $\frac{1}{2}(10 + 15) = 12 \cdot 5$.

Take $$u = \frac{x - 12 \cdot 5}{2 \cdot 5} \text{ and } v = y - 20.$$

Let the equation of straight line of fit be

$$v = a + bu. \qquad \text{...(i)}$$

Hence its normal equations are

$$\Sigma v = ma + b\Sigma u \qquad \text{...(ii)}$$
$$\Sigma uv = a\Sigma u + b\Sigma u^2 \qquad \text{...(iii)}$$

x	y	$u = \frac{x - 12 \cdot 5}{2 \cdot 5}$	$v = y - 20$	uv	u^2
0	12	−5	−8	40	25
5	15	−3	−5	15	9
10	17	−1	−3	3	1
15	22	1	2	2	1
20	24	3	4	12	9
25	30	5	10	50	25
Total		0	0	122	70

Here $m = 6$ [$\because$ m is number of terms]

Substituting these values in the equations (ii) and (iii), we have

$$0 = 6a + b(0) \text{ and } 122 = a(0) + b(70)$$

or $$a = 0 \text{ and } b = 1 \cdot 743.$$

$\therefore$ The required equation of st. line is [put values in (i)]

$$v = 0 + (1 \cdot 743)\,u$$

or $$y - 20 = (1 \cdot 743)\left(\frac{x - 12 \cdot 5}{2 \cdot 5}\right)$$

or $$y - 20 = \frac{1 \cdot 743}{2 \cdot 5}x - (1 \cdot 743) \times 5$$

or $$y - 20 = 0 \cdot 7x - 8 \cdot 715$$

or $$y = 0 \cdot 7x + 11 \cdot 285.$$

Prof. Lokesh Sharma

Example 4. ***Fit a parabolic curve of regression of y on x to the 7 pairs of values (or fit a second degree parabola to the following)***

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

Solution. Here $m = 7$ and so take the origin at 2·5; the mid-terms value of x varies.

Suppose $$X = \frac{x - 2\cdot 5}{0\cdot 5} = 2x - 5 \text{ and } Y = y.$$

Let the equation of parabolic curve to fit be
$$Y = a + bX + cX^2 \quad \text{...(i)}$$

Hence its normal equations are
$$\Sigma Y = ma + b\Sigma X + c\Sigma X^2 \quad \text{...(ii)}$$
$$\Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3 \quad \text{...(iii)}$$
$$\Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 \quad \text{...(iv)}$$

x	y	X = 2x − 5	Y = y	X^2	XY	X^2Y	X^3	X^4
1·0	1·1	−3	1·1	9	−3·3	9·9	−27	81
1·5	1·3	−2	1·3	4	−2·6	5·2	−8	16
2·0	1·6	−1	1·6	1	−1·6	1·6	−1	1
2·5	2·0	0	2·0	0	0	0	0	0
3·0	2·7	1	2·7	1	2·7	2·7	1	1
3·5	3·4	2	3·4	4	6·8	13·6	8	16
4·0	4·1	3	4·1	9	12·3	36·9	27	81
Total		0	16·2	28	14·3	69·9	0	196

Substituting these values in the equations (ii), (iii) and (iv), we have
$$16\cdot 2 = 7a + 28c \quad \text{...(1)}$$
$$14\cdot 3 = 28b \quad \text{...(2)}$$
$$69\cdot 9 = 28a + 196c. \quad \text{...(3)}$$

Solving equations (1), (2) and (3), we get
$$a = 2\cdot 07,\ b = 0\cdot 511,\ c = 0\cdot 061$$

∴ The required equation of parabolic curve of fit is given by [putting values in (i)]
$$Y = 2\cdot 07 + 0\cdot 511X + 0\cdot 061X^2$$
or $$y = 2\cdot 07 + 0\cdot 511(2x - 5) + 0\cdot 061(2x - 5)^2$$
or $$y = 2\cdot 07 + 1\cdot 022x - 2\cdot 555 + 0\cdot 061(4x^2 - 20x + 25)$$
or $$y = 1\cdot 04 - 0\cdot 193x + 0\cdot 243x^2.$$

Example 5. ***Fit a second degree parabola to the following data :***

x:	1	2	3	4	5
y:	1090	1220	1390	1625	1915

Solution. Introducing two new variables u and v connected by changing the origin $u = x - 3$ and $v = (1/5)(y - 1220)$.

Let the equation of parabolic curve of fit be
$$v = a + bu + cu^2 \quad \text{...(i)}$$

Hence its normal equations are
$$\Sigma v = ma + b\Sigma u + c\Sigma u^2 \quad \text{...(ii)}$$
$$\Sigma uv = a\Sigma u + b\Sigma u^2 + c\Sigma u^3 \quad \text{...(iii)}$$
$$\Sigma u^2 v = a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4. \quad \text{...(iv)}$$

The calculations of Σuv, Σu^2, etc. are shown in the following table :

x	y	u = x − 3	$v = \frac{y - 1220}{5}$	u^2	u^3	u^4	uv	u^2v
1	1090	−2	−26	4	−8	16	52	−104
2	1220	−1	0	1	−1	1	0	0
3	1390	0	34	0	0	0	0	0
4	1625	1	81	1	1	1	81	81
5	1915	2	139	4	8	16	278	556
Total	7240	0	228	10	0	34	411	533

Substituting these values, the normal equations are
$$228 = 5a + b(0) + c(10) \quad \text{...(1)}$$
$$411 = a(0) + b(10) + c(0) \quad \text{...(2)}$$
$$533 = a(0) + b(0) + c(34) \quad \text{...(3)}$$

From (2), we get $b = 41\cdot 1$.

Solving (1) and (3), we get
$$c = 5\cdot 5,\quad a = 34\cdot 6.$$

∴The equation of parabolic curve of fit is [from (i)]
$$v = 34\cdot 6 + 41\cdot 1u + 5\cdot 5u^2$$
or $$\frac{y - 1220}{5} = 34\cdot 6 + 41\cdot 1(x - 3) + 5\cdot 5(x - 3)^2$$
or $$y - 1220 = 173 + 205\cdot 5(x - 3) + 27\cdot 5(x - 3)^2$$
or $$y = 1024 + 40\cdot 5x + 27\cdot 5x^2.$$

Example 6. ***Fit a second degree parabola to the following data :***

x:	1	2	3	4	5
y:	25	28	33	39	46

Solution. Introducing two new variables u and v connected by changing the origin $u = x - 3$, $v = y - 33$. Let the equation of parabolic curve of fit be

$$v = a + bu + cu^2 \qquad \text{...(i)}$$

Hence its normal equations are

$$\Sigma v = ma + b\Sigma u + c\Sigma u^2 \qquad \text{...(ii)}$$

$$\Sigma uv = a\Sigma u + b\Sigma u^2 + c\Sigma u^3 \qquad \text{...(iii)}$$

$$\Sigma u^2 v = a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4. \qquad \text{...(iv)}$$

The calculations of Σuv, Σu^2 etc. are shown in the following table :

x	y	u = x − 3	v = y − 33	u^2	uv	u^2v	u^3	u^4
1	25	− 2	− 8	4	16	− 32	− 8	16
2	28	− 1	− 5	1	5	− 5	− 1	1
3	33	0	0	0	0	0	0	0
4	39	1	6	1	6	6	1	1
5	46	2	13	4	26	52	8	16
Total		0	6	10	53	21	0	34

Substituting these values, the normal equations are

$$6 = 5a + 10c \qquad \text{...(1)}$$

$$53 = 10b \qquad \text{...(2)}$$

$$21 = 10a + 34c. \qquad \text{...(3)}$$

Solving equations (1), (2) and (3), we have

$$b = 5 \cdot 3,\ c = 0 \cdot 643,\ a = -0 \cdot 086.$$

Hence the required equation of parabolic curve of fit is

$$v = -0 \cdot 086 + 5 \cdot 3u + 0 \cdot 643u^2$$

$$y - 33 = 0 \cdot 087 + 5 \cdot 3(x - 3) + 0 \cdot 643(x - 3)^2$$

$$y = 22 \cdot 8 + 1 \cdot 442x + 0 \cdot 643x^2.$$

Example 7. ***The weights of a calf taken at weekly intervals are given below. Fit a straight line using the method of least squares, and calculate the average rate of growth per week.***

Age :	1	2	3	4	5	6	7	8	9	10
Weight :	52·5	58·7	65·0	70·2	75·4	81·1	87·2	95·5	102·2	108·4

Solution. Here $m = 10$, *i.e.*, even and $h = 1$; thus we take the origin at the middle value of mid-pair, *i.e.*, $\frac{5+6}{2} = 5 \cdot 5$, and $\frac{1}{2}$ *i.e.*, $0 \cdot 5$ as scale of measurement.

Prof. Lokesh Sharma

Now let u be $\frac{x - 5 \cdot 5}{0 \cdot 5}$ and v be y.

Again let the equation of curve of fit be

$$v = a + bu \qquad \text{...(i)}$$

Hence its normal equations are

$$\Sigma v = ma + b\Sigma u \qquad \text{...(ii)}$$

$$\Sigma uv = a\Sigma u + b\Sigma u^2 \qquad \text{...(iii)}$$

x	v = y	$u = \frac{x - 5 \cdot 5}{0 \cdot 5}$	uv	u^2
1	52·5	− 9	− 472·5	81
2	58·7	− 7	− 410·9	49
3	65·0	− 5	− 325·0	25
4	70·2	− 3	− 210·6	9
5	75·4	− 1	− 75·4	1
6	81·1	1	81·2	1
7	87·2	3	261·6	9
8	95·5	5	477·5	25
9	102·2	7	715·4	49
10	108·4	9	975·6	81
Total	796·2	0	1016·8	330

Substituting the values, we get from (ii) and (iii)

$$796 \cdot 2 = 10a \quad \text{and} \quad 1016 \cdot 8 = b(330)$$

or $$a = 79 \cdot 62 \quad \text{and} \quad b = 3 \cdot 081.$$

∴ The required equation of curve of fit is

$$v = (79 \cdot 6) + (3 \cdot 081)\,u \qquad \text{[from (i)]}$$

or $$y = 79 \cdot 62 + 3 \cdot 081\left(\frac{x - 5 \cdot 5}{0 \cdot 5}\right)$$

$$y = 6 \cdot 162x + 45 \cdot 729$$

and the average rate of growth per week is 6·162 units.

Example 8. ***Fit a second degree parabola in the following data :***

x:	0·0	1·0	2·0	3·0	4·0
y:	1·0	4·0	10·0	17·0	30·0

Solution. Here $m = 5$, *i.e.*, odd, therefore the origin in x-series will be taken at the mid-value of x *i.e.*, at 2. Let the origin in y-series be taken at 10. Now consider the new variables X and Y defined by

$$X = x - 2,\ Y = y - 10.$$

Let the equation of the parabola to fit be

$$Y = a + bX + cX^2. \qquad \text{...(i)}$$

$\therefore$ The normal equations are

$$\Sigma Y = 5a + b\,\Sigma X + c\,\Sigma X^2 \qquad \text{...(ii)}$$

$$\Sigma XY = a\,\Sigma X + b\,\Sigma X^2 + c\,\Sigma X^3 \qquad \text{...(iii)}$$

$$\Sigma X^2Y = a\,\Sigma X^2 + b\,\Sigma X^3 + c\,\Sigma X^4 \qquad \text{...(iv)}$$

x	y	$X = x-2$	$Y = y-10$	XY	X^2	X^2Y	X^2Y	X^4
0·0	1·0	−2	−9	18	4	−36	−8	16
1·0	4·0	−1	−6	6	1	−6	−1	1
2·0	10·0	0	0	0	0	0	0	0
3·0	17·0	1	7	7	1	7	1	1
4·0	30·0	2	20	40	4	80	8	16
Total		$\Sigma X = 0$	$\Sigma Y = 12$	$\Sigma XY = 71$	$\Sigma X^2 = 10$	$\Sigma X^2Y = 45$	$\Sigma X^3 = 0$	$\Sigma X^4 = 34$

Substituting values in normal equations (ii), (iii), (iv), we get

$$12 = 5a + 10c \qquad \text{...(1)}$$

$$71 = 10b \qquad \text{...(2)}$$

$$45 = 10a + 34c. \qquad \text{...(3)}$$

Solving (1), (2) and (3), we get

$$a = -0\cdot 6,\ b = 7\cdot 1,\ c = 1\cdot 5.$$

The equation of parabola to fit is [putting values in (i)]

$$y - 10 = -0\cdot 6 + 7\cdot 1\,(x-2) + 1\cdot 5\,(x-2)^2$$

$$\Rightarrow \quad y = 10 - 0\cdot 6 + 7\cdot 1\,x - 14\cdot 2 + 1\cdot 5\,x^2 - 6x + 6$$

$$\Rightarrow \quad y = 1\cdot 2 + 1\cdot 1\,x + 1\cdot 5\,x^2. \qquad \textbf{Ans.}$$

Example 9. ***Fit a second degree curve of regression of y on x to the following data :***

x:	1·0	2·0	3·0	4·0
y:	6·0	11·0	18·0	27·0

Solution. Here $m = 4$, *i.e.*, even, and values of x are equally spaced, h (common difference in x-series) = 1. Therefore, to make calculation easier, we take the unit of measurement $\frac{1}{2}h$ *i.e.*, 0·5, and the origin at the mean of two middle terms 2 and 3 *i.e.*, origin is taken at $\frac{1}{2}(2+3) = 2\cdot 5$.

Taking $\quad u = \dfrac{x - 2\cdot 5}{0\cdot 5}$ and $v = y - 10$.

Prof. Lokesh Sharma

Let the equation of second degree curve (parabola) of fit be

$$v = a + bu + cu^2. \qquad \text{...(i)}$$

Its normal equations are

$$\Sigma v = 4a + b\,\Sigma u + c\,\Sigma u^2 \qquad \text{...(ii)}$$

$$\Sigma uv = a\,\Sigma u + b\,\Sigma u^2 + c\,\Sigma u^3 \qquad \text{...(iii)}$$

$$\Sigma u^2 v = a\,\Sigma u^2 + b\,\Sigma u^3 + c\,\Sigma u^4 \qquad \text{...(iv)}$$

x	y	$u = \frac{x-2\cdot 5}{0\cdot 5}$	$v = y - 10$	u^2	uv	u^2v	u^3	u^4
1	6	−3	−4	9	12	−36	−27	81
2	11	−1	1	1	−1	1	−1	1
3	18	1	8	1	8	8	1	1
4	27	3	17	9	51	153	27	81
Total		$\Sigma u = 0$	$\Sigma v = 22$	$\Sigma u^2 = 20$	$\Sigma uv = 70$	$\Sigma u^2 v = 126$	$\Sigma u^3 = 0$	$\Sigma u^4 = 164$

Substituting values in normal equations, we get

$$22 = 4a + 20c \qquad \text{...(1)}$$

$$70 = 20b \qquad \text{...(2)}$$

$$126 = 20a + 164c. \qquad \text{...(3)}$$

Solving (1), (2), (3); we get

$$a = 4\cdot 25,\ b = 3\cdot 5,\ c = 0\cdot 25.$$

The equation of parabola to fit is [putting values in (i)]

$$y - 10 = 4\cdot 25 + 3\cdot 5\left(\frac{x-2\cdot 5}{0\cdot 5}\right) + 0\cdot 25\left(\frac{x-2\cdot 5}{0\cdot 5}\right)^2$$

or $$y = 3 + 2x + x^2. \qquad \textbf{Ans.}$$

Example 10. ***Fit a second degree parabola in the following data by least square method :***

x:	1929	1930	1931	1932	1933	1934	1935	1936	1937
y:	352	356	357	358	360	361	361	360	359

Solution. Here $m = 9$ *i.e.*, odd, therefore the origin in x-series will taken at mid-value of x *i.e.*, at 1933. Also $h = 1$. Let the origin in y-series be taken at 358. Then

$$u = \frac{x - 1933}{1} = x - 1933 \text{ and } v = y - 358.$$

Let the equation of parabola to fit be

$$v = a + bu + cu^2. \qquad \text{...(i)}$$

$\therefore$ The normal equations are :

$$\Sigma v = 9a + b\,\Sigma u + c\,\Sigma u^2 \qquad \text{...(ii)}$$

$$\Sigma uv = a\,\Sigma u + b\,\Sigma u^2 + c\,\Sigma u^3 \qquad \text{...(iii)}$$

$$\Sigma u^2 v = a\,\Sigma u^2 + b\,\Sigma u^3 + c\,\Sigma u^4 \qquad \text{...(iv)}$$

x	y	u = x − 1933	v = y − 358	u^2	uv	u^2v	u^3	u^4
1929	352	− 4	− 6	16	24	− 96	− 64	256
1930	356	− 3	− 2	9	6	− 18	− 27	81
1931	357	− 2	− 1	4	2	− 4	− 8	16
1932	358	− 1	0	1	0	0	− 1	1
1933	360	0	2	0	0	0	0	0
1934	361	1	3	1	3	3	1	1
1935	361	2	3	4	6	12	8	16
1936	360	3	2	9	6	18	27	81
1937	359	4	1	16	4	16	64	256
Total		$\Sigma u = 0$	$\Sigma v = 22$	$\Sigma u^2 = 60$	$\Sigma uv = 51$	$\Sigma u^2v = -69$	$\Sigma u^3 = 0$	$\Sigma u^4 = 708$

Substituting values in normal equations (ii), (iii), (iv), we get

$$2 = 9a + 60c \quad ...(1)$$

$$51 = 60b \quad ...(2)$$

$$-69 = 60a + 708c. \quad ...(3)$$

Solving (1), (2) and (3), we get

$$a = 2 \cdot 004,\ b = 0 \cdot 850,\ c = -0 \cdot 267.$$

Hence the parabola to fit is [putting values in (i)]

$$y - 358 = 2 \cdot 004 + 0 \cdot 84\,(x - 1933) - 0 \cdot 267\,(x - 1933)^2$$

$$\Rightarrow \quad y = -99925 \cdot 609 + 1033 \cdot 072x - 0 \cdot 267x^2. \quad \textbf{Ans.}$$

EXERCISE 8 (A)

1. (a) Fit the straight line $y = a + bx$ to the following data :

x:	1	2	3	4	5
y:	4	2	3	5	7

(b) Fit a straight line to the following data regarding y as a dependent variable :

x:	1	2	3	4	5
y:	5	7	9	10	11

2. Show that the line of fit to the following data is given by $y = -0 \cdot 5x + 8$:

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

3. Taking 1913 as origin for x-series, fit a straight line to the following data showing the production of a commodity in different years in Punjab :

Year	x:	1911	1912	1913	1914	1915
Production (1000 tons)	y:	10	12	8	10	14

4. Fit a straight line to the following data showing the production of wheat in different years in any village of Punjab :

Year :	1982	1983	1984	1985	1986	1987	1988
Production (1000 tons) :	14	17	15	18	20	19	23

Prof. Lokesh Sharma

5. The profit £ 100y, of certain company in the xth year of its life are given by :

x:	1	2	3	4	5
y:	25	28	33	39	46

Fit a second degree parabola to the above data.

6. Fit a second degree parabola to the following data :

x:	− 4	− 3	− 2	− 1	0	1	2	3	4
y:	101	104	107	107	108	110	109	110	108

7. The profit of a certain company in the xth year of its life are given by :

x:	1	2	3	4	5
y:	1250	1400	1650	1950	2300

Taking $u = x - 3$ and $v = \dfrac{y - 1650}{50}$, show that the parabola of the second degree of v on u is

$$v + 0 \cdot 086 = 5 \cdot 30u + 0 \cdot 643u^2,$$

and deduce that the parabola of second degree of y on x is $y = 1140 + 72 \cdot 14x + 32 \cdot 14x^2$.

8. A manurial experiment on paddy gave the following results :

Dose of measure (in 1bs)	x:	0	200	300	400
Yield per acre (in 1bs)	y:	1544	1898	2133	2327

fit a parabola of the second degree to represent the data.

9. Show that the best fitting linear function for the points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ may be expressed in the form

$$\begin{vmatrix} x & y & 1 \\ \Sigma x & \Sigma y & n \\ \Sigma x^2 & \Sigma y^2 & \Sigma x \end{vmatrix} = 0.$$

Show also that the line passes through the mean point $(\bar{x}, \bar{y})$.

10. (a) Write the method to fit a straight line by the method of least squares.

(b) Write a brief note to fit the parabola by the method of least squares.

11. What do you mean by normal equations ? Explain the method to determine them.

12. Obtain through the principle of least squares the normal equations for fitting a second degree equation.

13. Fit the straight line $y = a + bx$ to the following data regarding x as independent variable

x:	1	2	3	4	6	8
y:	2·4	3	3·6	4	5	6

14. Fit the straight line to the following data :

x:	1	2	3	4	5
y:	2·0	3·0	5·0	6·5	8·2

15. The results of measurements of electric resistance R of a copper bar at various temperatures t°C are listed below :

t:	19	25	30	36	40	45	50
R:	76	77	79	80	82	83	85

Find a relation $R = \alpha + bt$ when α and b are constants to be determined by you.

16. The following table gives the results of the measurements of train resistance; V is the velocity in miles per hour, R is the resistance in pounds per ton :

V :	20	40	60	80	100	120
R :	5.5	9.1	14.9	22.8	33.3	46.0

If R is related to V by the relation $R = a + bV + cV^2$; find a, b and c.

17. Fit a second degree parabola to the following data :

x :	20	40	60	80	100	120
y :	5.5	9.1	14.9	22.8	33.3	46.0

[Hint. Same as Q. 16]

ANSWERS

1. (b) $y = 3{\cdot}9 + 1{\cdot}5x$ 3. $y = 10{\cdot}8 + 0{\cdot}6x$ 4. $y = 18 + 1{\cdot}3x$
5. $v = -0{\cdot}086 + 5{\cdot}3u + 0{\cdot}643u^2$ 6. $y = 109 + 0{\cdot}883x - 2457x^2$
8. $y = 1547{\cdot}9 + 378{\cdot}4x - 40x^2$ 13. $y = 1{\cdot}976 + 0{\cdot}506x$
15. $R = 70{\cdot}052 + 0{\cdot}292\,t$ 16. $R = 3{\cdot}48 - 0{\cdot}002\,V + 0{\cdot}0029\,V^2$.
17. $y = 3{\cdot}48 - 0{\cdot}002x + 0{\cdot}0029x^2$.

§ 8.7. MOST PLAUSIBLE VALUE

Let $x, y, z, \ldots, k$ be n unknowns and let there be a number of linear equations in $x, y, z, \ldots$, given by

$$\left.\begin{aligned} a_{11}x + a_{12}y + a_{13}z + \ldots + a_{1n}k &= B_1 \\ a_{21}x + a_{22}y + a_{23}z + \ldots + a_{2n}k &= B_2 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots & \ldots \\ a_{m1}x + a_{m2}y + a_{m3}z + \ldots + a_{mn}k &= B_m \end{aligned}\right\} \qquad \ldots(1)$$

where a's and B's are constant.

If $m = n$, *i.e.*, a number of linear equations is equal to the number of unknowns, then, in general, there exists a unique set of values satisfying the given system of equations. But if $m > n$, *i.e.*, number of equations is greater than the number of unknowns, then there may exist no such solution. Therefore, in such cases we try to find those values of $x, y, z, \ldots, k$ which will satisfy the given system of equations as nearly as possible.

The principle of least squares asserts that these values are those which shall make U a minimum, where U is given by

$$U = \sum_{r=1}^{\infty} (a_{r_1} x + a_{r_2} y + \ldots + a_{r_n} k - B_r)^2 \qquad \ldots(2)$$

Now applying the conditions of minimum of U, *i.e.*,

$$\frac{\partial U}{\partial x} = \frac{\partial U}{\partial y} = \ldots = \frac{\partial U}{\partial k} = 0 \qquad \ldots(3)$$

we get n equations, called the normal equations. These equations can be solved as simultaneous equations for $x, y, z, \ldots, k$. These values are called the **best** or **most plausible values.** If we substitute these values in 2nd order partial derivatives, we see that the expressions are +ve. Hence U is minimum.

Prof. Lokesh Sharma

ILLUSTRATIVE EXAMPLES

Example 1. ***Form normal equations and hence find the most plausible values of x and y from the following :***

$$x + y = 3{\cdot}01, \quad 2x - y = 0{\cdot}03,$$
$$x + 3y = 7{\cdot}03, \quad 3x + y = 4{\cdot}97.$$

Solution. Suppose

$$U = (x + y - 3{\cdot}01)^2 + (2x - y - 0{\cdot}03)^2 + (x + 3y - 7{\cdot}03)^2 + (3x + y - 4{\cdot}97)^2. \qquad \ldots(1)$$

These values of x and y which make U minimum are called most plausible values. They will be given by the following equations :

$$\frac{\partial U}{\partial x} = 0 = \frac{\partial U}{\partial y}.$$

Differentiating (1) partially w.r.t. x, we have

$$(x + y - 3{\cdot}01) + (2x - y - 0{\cdot}03)\,2 + (x + 3y - 7{\cdot}03) + (3x + y - 4{\cdot}97)\,3 = 0$$

i.e.,
$$15x + 5y = 25. \qquad \ldots(2)$$

Differentiating (1) partially with respect to y, we have

$$(x + y - 3{\cdot}01) - (2x - y - 0{\cdot}03) + 3(x + 3y - 7{\cdot}03) + (3x + y - 4{\cdot}97) = 0$$

or
$$5x + 12y = 29{\cdot}04. \qquad \ldots(3)$$

Solving (2) and (3), we have

$$x = 0{\cdot}999, \quad y = 2{\cdot}004.$$

These values are the most plausible values of x and y and so integral values of x and y are 1 and 2 respectively.

Example 2. ***Find the most plausible values of x, y and z from the following equations :***

$$x - y + 2z = 3, \quad 3x + 2y - 5z = 5,$$
$$4x + y + 4z = 21, \quad -x + 3y + 2z = 14.$$

Solution. Suppose

$$U = (x - y + 2z - 3)^2 + (3x + 2y - 5z - 5)^2 + (4x + y + 4z - 21)^2 + (-x + 3y + 3z - 14)^2.$$

These values of x, y and z which make U minimum are called most plausible values. They will be given by

$$\partial U / \partial x = 0 = \partial U / \partial y = \partial U / \partial z.$$

Differentiating (1) partially with respect to x, y and z respectively, we have

$$(x - y + 2z - 3) + 3(3x + 2y - 5z - 5) + 4(4x + y + 4z - 21) - (-x + 3y + 3z - 14) = 0 \qquad \ldots(1)$$

or
$$27x + 6y = 88$$

$$-(x - y + 2z - 3) + 2(3x + 2y - 5z - 5) + (4x + y + 4z - 21) + 3(-x + 3y + 3z - 14) = 0 \qquad \ldots(2)$$

or
$$6x + 15y + z = 70$$

Example 4. ***The pressure and volume of a gas are related by the equation $pv^a = b$, where a and b are constants. Fit this equation to the following set of data :***

$p\,(kg/cm^2)$	0·5	1·0	1·5	2·0	2·5	3·0
v (litres)	1·62	1·00	0·75	0·62	0·52	0·46

Solution. Given equation $pv^a = b \Rightarrow v = \left(\frac{b}{p}\right)^{1/a}$.

Taking log base 10 of both sides, we get

$$\log_{10} v = \frac{1}{a}\log_{10} b - \frac{1}{a}\log_{10} p.$$

This is of the form, $Y = A + BX$ ...(1)

where $Y = \log_{10} v,\ X = \log_{10} p,\ A = \frac{1}{a}\log_{10} b,\ B = -\frac{1}{a}$.

Equation (1) represents a straight line, therefore, normal equations are

$$\Sigma Y = 6A + B\,\Sigma X \qquad [\because\ m = 6] \quad ...(2)$$

$$\Sigma XY = A\,\Sigma X + B\,\Sigma X^2 \quad ...(3)$$

p	v	$X = \log_{10} p$	$Y = \log_{10} v$	XY	X^2
0·5	1·62	−0·30103	0·20952	−0·06307	0·09062
1·0	1·00	0	0	0	0
1·5	0·75	0·17609	−0·12494	−0·02200	0·03101
2·0	0·62	0·30103	−0·20761	−0·06250	0·09062
2·5	0·52	0·39794	−0·28400	−0·11301	0·15836
3·0	0·46	0·47712	−0·33724	−0·16090	0·22764
		$\Sigma X = 1·05115$	$\Sigma Y = -0·74427$	$\Sigma XY = -0·42148$	$\Sigma X^2 = 0·59825$

Substituting values in normal equations (2) and (3), we get

$$-0·74427 = 6A + 1·05115\,B \quad ...(4)$$

$$-0·42148 = 1·05115\,A + 0·59825\,B. \quad ...(5)$$

Eliminating A,

$$B = \frac{6 \times 0·42148 - 0·74427 \times 1·05115}{1·05115 \times 1·05115 - 6 \times 0·59825} = \frac{1·74654}{-2·48458} = -0·70295.$$

Putting for B in (4), we get

$$A = \frac{1·05115 \times 0·70295 - 0·74427}{6} = -0·00089.$$

$\therefore \quad a = -\frac{1}{B} = \frac{1}{0·70295} = 1·42258$

and $\quad \log_{10} b = aA = -0·00127 = \bar{1}·99773$

$$b = \text{Antilog}\,(\bar{1}·99873) = 0·997080.$$

Hence the required curve is

$$pv^{1·42} = 0·99 \qquad \text{[upto two decimal places]}$$ **Ans.**

Example 5. ***The pressure of the gas corresponding to various volumes V is measured, given by the following data :***

$V\,(cm^3)$	50	60	70	90	100
$P\,(kg/cm^2)$	64·7	51·3	40·5	25·9	78

Fit the data to the equation $PV^{\gamma} = C$.

Solution. Here the pressure P is dependent variable while V is independent variable.

The given equation is $PV^{\gamma} = C$. ...(1)

Taking log base 10 of both sides of (1), we get

$$\log_{10} P = \log_{10} C - \gamma \log_{10} V.$$

This is of the form, $y = a + bx$ ...(2)

where $y = \log_{10} P,\ x = \log_{10} V,\ a = \log_{10} C,\ b = -\gamma$.

Equation (2) represents a straight line, therefore, normal equations are

$$\Sigma y = 5a + b\,\Sigma x \qquad [\because\ m = 5] \quad ...(3)$$

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 \quad ...(4)$$

V	P	$x = \log_{10} V$	$y = \log_{10} P$	xy	x^2
50	64·7	1·69897	1·81090	3·0766	2·88650
60	51·3	1·77815	1·71012	3·04085	3·16182
70	40·5	1·84510	1·60746	2·96592	3·40439
90	25·9	1·95424	1·41330	2·76193	3·81905
100	78	2	1·89209	3·78418	4
		$\Sigma x = 9·27646$	$\Sigma y = 8·43387$	$\Sigma xy = 15·62954$	$\Sigma x^2 = 17·27176$

Putting values in (3) and (4), the normal equations are :

$$8·43387 = 5a + 9·27646\,b \quad ...(5)$$

$$15·62954 = 9·27646\,a + 17·27176\,b. \quad ...(6)$$

Eliminating a,

$$b = \frac{9·27646 \times 8·43387 - 5 \times 15·62954}{(9·27646)^2 - 5 \times 17·27176} = \frac{0·08876}{-0·30609} = -0·28997.$$

Putting for b in (5), we get

$$a = 2·22475.$$

$\therefore \quad \gamma = -b = 0·28997$

and $\quad C = \text{Antilog}\ a = \text{Antilog}\ 2·22475 = 167·78379.$

Hence the required curve [from (1)] is $PV^{0·28997} = 167·78379$. **Ans.**

Example 6. ***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

Solution. Given $\quad y = ae^{bx}$. ...(1)

Taking log base 10 of both sides of (1)

$$\log_{10} y = \log_{10} a + bx \log_{10} e.$$

This is of the form, $Y = A + BX$ ...(2)

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

Equation (2) represents a straight line, so its normal equations are

$$\Sigma Y = 5A + B\,\Sigma X \qquad [\because m = 5] \qquad ...(3)$$

and

$$\Sigma XY = A\,\Sigma X + B\,\Sigma X^2 \qquad ...(4)$$

$x = X$	y	$Y = \log_{10} y$	XY	X^2
1	10	1	1	1
5	15	1·17609	5·88045	25
7	12	1·07918	7·55426	49
9	15	1·17609	10·58481	81
12	21	1·32222	15·86664	144
$\Sigma X = 34$		$\Sigma Y = 5·75358$	$\Sigma XY = 40·88616$	$\Sigma X^2 = 300$

Substituting values in (3) and (4), the normal equations are

$$5·75358 = 5A + 34B \qquad ...(5)$$

$$40·88616 = 34A + 300B. \qquad ...(6)$$

Eliminating B, $\quad A = \dfrac{34 \times 5·75358 - 5 \times 40·88616}{34 \times 34 - 5 \times 300} = \dfrac{8·80908}{344} = 0·02561.$

Putting for B in (5), $A = 0·97657.$

Now $a = \text{Antilog } A = 9·47480,\ b = \dfrac{B}{\log_{10} e} = 0·05897.$

Hence the required curve [from (1)] is $y = 9·47480\, e^{0·05897\,x}$. **Ans.**

Example 7. ***By the method of least squares, find the curve*** $y = \dfrac{a}{x} + b\sqrt{x}$ ***that best fits the following data :***

x:	0·1	0·2	0·4	0·5	1	2
y:	21	11	7	6	5	6

Solution. Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be n given points, then, residual R_r for rth point (x_r, y_r) is given by

$$R_r = \left(y_r - \frac{a}{x_r} - b\sqrt{x_r}\right).$$

Let

$$U = \sum_{r=1}^{n} R_r^2 = \sum_{r=1}^{n}\left(y_r - \frac{a}{x_r} - b\sqrt{x_r}\right)^2.$$

By the principle of least squares, a and b are determined such that U is minimum.

The normal equations are given by

$$\frac{\partial U}{\partial a} = 0 \Rightarrow \sum_{r=1}^{n} 2\left(y_r - \frac{a}{x_r} - b\sqrt{x_r}\right)\left(-\frac{1}{x_r}\right) = 0$$

$$\Rightarrow \sum_{r=1}^{n} \frac{y_r}{x_r} = a \sum_{r=1}^{n} \frac{1}{x_r^2} + b \sum_{r=1}^{n} \frac{1}{\sqrt{x_r}}$$

$$\Rightarrow \Sigma \frac{y}{x} = a\,\Sigma \frac{1}{x^2} + b\,\Sigma \frac{1}{\sqrt{x}} \qquad ...(1)$$

[Omitting the suffix r]

and

$$\frac{\partial U}{\partial b} = 0 \Rightarrow \sum_{r=1}^{n} 2\left(y_r - \frac{a}{x_r} - b\sqrt{x_r}\right)(-\sqrt{x_r}) = 0$$

$$\Rightarrow \Sigma\, y\sqrt{x} = a\,\Sigma \frac{1}{\sqrt{x}} + b\,\Sigma x. \qquad ...(2)$$

Thus (1) and (2) are normal equations.

x	y	$\frac{1}{\sqrt{x}}$	$\frac{1}{x^2}$	$\frac{y}{x}$	$y\sqrt{x}$
0·1	21	3·16228	100	210	6·64078
0·2	11	2·23607	25	55	4·91935
0·4	7	1·58114	6·25	17·5	4·42719
0·5	6	1·41421	4	12	4·24264
1·0	5	1	1	5	5
2·0	6	0·70711	0·25	3	8·48528
$\Sigma x = 4·2$		$\Sigma 1/\sqrt{x} = 10·10081$	$\Sigma 1/x^2 = 136·5$	$\Sigma y/x = 302·5$	$\Sigma y\sqrt{x} = 33·71524$

Substituting values in (1) and (2), the normal equations are :

$$302·5 = 136·5\,a + 10·10081\,b \qquad ...(3)$$

$$33·71524 = 10·10081a + 4·2\,b. \qquad ...(4)$$

Eliminating b,

$$b = \frac{4·2 \times 302·5 - 33·71524 \times 10·10081}{4·2 \times 136·5 - (10·10081)^2} = \frac{929·94877}{471·27364} = 1·97327.$$

Putting for a in (4), $b = 3·28178$

Hence the required curve is $y = \dfrac{1·97327}{x} + 3·28178\sqrt{x}.$ **Ans.**

Example 8. ***A person runs the same race track for five consecutive days and is timed as follows :***

Day (x) :	1	2	3	4	5
Time (y) :	15·3	15·1	15	14·5	14

Make a least square fit to the above data using a function $a + \dfrac{b}{x} + \dfrac{c}{x^2}$.

Solution. Let $\quad y = a + \dfrac{b}{x} + \dfrac{c}{x^2}.$

Let $(x_r, y_r), r = 1, 2, \ldots, n$ be n given points.

By the principle of least squares the values of a, b and c are such that

$$U = \sum_{r=1}^{n}\left(y_r - a - \frac{b}{x_r} - \frac{c}{x_r^2}\right)^2 \text{ is minimum.}$$

The normal equations are

$$\frac{\partial U}{\partial a} = 0 \Rightarrow \sum_{r=1}^{n} 2\left(y_r - a - \frac{b}{x_r} - \frac{c}{x_r^2}\right)(-1) = 0$$

$$\Rightarrow \Sigma y = na + b\,\Sigma\frac{1}{x} + c\,\Sigma\frac{1}{x^2}. \qquad \text{...(1)}$$

[Neglecting the suffix r]

Similarly, $\dfrac{\partial U}{\partial b} = 0$ and $\dfrac{\partial U}{\partial C} = 0$ give

$$\Sigma\frac{y}{x} = a\,\Sigma\frac{1}{x} + b\,\Sigma\frac{1}{x^2} + c\,\Sigma\frac{1}{x^3} \qquad \text{...(2)}$$

and

$$\Sigma\frac{y}{x^2} = a\,\Sigma\frac{1}{x} + b\Sigma\frac{1}{x^3} + c\,\Sigma\frac{1}{x^4} \qquad \text{...(3)}$$

x	y	$\frac{1}{x}$	$\frac{1}{x^2}$	$\frac{1}{x^3}$	$\frac{1}{x^4}$	$\frac{y}{x}$	$\frac{y}{x^2}$
1	15·3	1	1	1	1	15·3000	15·3
2	15·1	0·5000	0·2500	0·1250	0·0625	7·5500	3·7750
3	15	0·3333	0·1111	0·0370	0·0123	5·0000	1·6667
4	14·5	0·2500	0·0625	0·0156	0·0039	3·6250	0·9063
5	14	0·2000	0·0400	0·0080	0·0016	2·8000	0·5600
$\Sigma x = 15$	$\Sigma y = 73·9$	$\Sigma 1/x = 2·2833$	$\Sigma 1/x^2 = 1·4636$	$\Sigma 1/x^3 = 1·1856$	$\Sigma 1/x^4 = 1·0803$	$\Sigma y/x = 34·2750$	$\Sigma y/x^2 = 22·2080$

Substituting values in (1), (2), (3), the normal equations are :

$$73·9 = 5a + 2·2833b + 1·4636c \qquad [\because n = 5] \qquad \text{...(4)}$$

$$34·2750 = 2·2833a + 1·4636b + 1·1856c \qquad \text{...(5)}$$

$$22·2080 = 1·4636a + 1·1856b + 1·0803\,c. \qquad \text{...(6)}$$

Multiplying (4) by 2·2833 and (5) by 5 and subtracting, we get

$$2·6391 = 2·1045b + 2·5862c. \qquad \text{...(7)}$$

Multiplying (4) by 1·4636 and (6) by 5 and subtracting, we get

$$2·8800 = 2·5862\,b + 3·2594\,c. \qquad \text{...(8)}$$

Eliminating c between (7) and (8), $\quad b = \dfrac{1·1536}{0·1710} = 6·7462.$

Put for b in (7), $\quad c = -4·4692.$

Putting values of b and c in (4), $\quad a = 13·0075.$

Hence the required curve is $y = 13·0075 + \dfrac{6·7462}{x} - \dfrac{4·4692}{x^2}$. **Ans.**

Prof. Lokesh Sharma

EXERCISE 8 (C)

1. Fit the curve $y = ae^{bx}$ to the following data, e being Napierian base 2·71828 :

x:	0	2	3
y:	5·012	10	31·62

2. Fit the curve $y = ae^{bx}$ to the following data :

x:	1	2	3	4	5	6	7	8
y:	15·3	20·5	27·7	36·6	49·1	65·6	87·8	117·6

3. Fit the curve $y = ax^b$ to the following data :

x:	1	2	3	4	5	6
y:	2·98	4·25	5·21	6·10	6·80	7·60

4. Use the method of least squares to determine a and b in the formula $y = ax + bx^2$ for the following observations :

x:	1	2	3	4	5
y:	1·8	5·1	8·9	14·1	19·8

5. The population of a state at ten yearly intervals is given below :

Year	1881	1891	1901	1911	1921	1931	1941	1951
Population	3·9	5·3	7·3	9·6	12·9	17·1	23·2	30·9

Fit the curve of the form $y = ab^x$ to these data and estimate the population for 1961.

6. Determine the constants a and b by the method of least squares such that $y = ae^{bx}$ fits the following data :

x:	2	4	6	8	10
y:	4·077	11·084	30·128	81·897	222·62

7. Fit the curve $pv^{\gamma} = k$ to the following data :

p (kg/cm^2)	0·5	1	1·5	2	2·5	3
v (litres)	1620	1000	750	620	520	560

8. Fit a curve of the type $xy = ax + b$ to the following data :

x:	1	3	5	7	9	10
y:	36	29	28	26	24	15

ANSWERS

1. $y = 4·642e^{0·46x}$
2. $y = 11·58e^{-0·2898x}$
3. $y = 2·978x^{-5·444}$
4. $a = 1·521$, $b = 0·49$, $y = 5·006$ when $x = 2$.
5. $a = 11·08$, $b = 1·158$, population 42·37 in 10 lakhs.
6. $a = 1·49989$, $b = 0·50001$
7. $pv^{1·42252} = 18462·48$
8. $xy = 16·18x + 40·78.$