UNIT 5 FITTING OF CURVES

Structure

5.1 INTRODUCTION

All the methods that you have learnt in Block 1 of this course were based on the uni-variate distributions, i.e. all measures analysed only single variable. But many times we have data for two or more variables and our interest is to know the best functional relationship between variables that given data describes. In this unit, you will learn how to fit the various functions such as straight line, parabola of the second degree, power curve and exponential curves. Curve fitting has theoretical importance in regression and correlation analysis while practically it is used to present the relationship by simple algebraic expression. All these methods can be used to estimate the values of the dependent variable for the specific value of the independent variable.

Section 5.2 gives the basic idea of the principle of least squares and procedure of fitting any curve for any given set of data. Section 5.3 explains the fitting of straight line while Sections 5.4 and 5.5 give the fitting of second degree parabola and power curve respectively. Fitting of exponential curves is described in Sections 5.6 and 5.7. Fitting of all functions considered in this unit are explained with examples also.

Objectives

After reading this unit, you would be able to

- describe the principle of least squares;
- describe the procedure of fitting any curve or functional relationship;
- define and calculate the residuals;
- fit a straight line for the given data;
- fit the second degree parabola for given data;
- fit a power curve for the given data; and
- fit a exponential curves $Y = ab^x$ and $Y = ae^{bx}$.

5.2 PRINCIPLE OF LEAST SQUARES

Let Y and X be the dependent and independent variables respectively and we have a set of values $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, i.e. observations are taken from n individuals on X and Y. We are interested in studying the function $Y = f(X)$. If Y_i is the estimated value of Y obtained by the function and y_i is the observed value of Y at x_i then we can define residual.

The difference between y_i and Y_i i.e. the difference between observed value and estimated value is called error of the estimate or residual for y_i.

Principle of least squares consists in minimizing the sum of squares of the residuals, i.e. according to principle of least squares

$$U = \sum_{i=1}^{n} (y_i - Y_i)^2 \text{ should be minimum.}$$

Let us consider a curve (function) of the type

$$Y = a + bX + cX^2 + ... + tX^k \qquad ... (1)$$

where, Y is dependent variable, X is independent variable and a, b, c,…,t are unknown constants. Suppose we have $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ values of two variables (X, Y) i.e. data for variables X and Y. These variables may be height and weight, sales and profit, rainfall and production of any crop, etc. In all these examples, first variables, i.e. height, sales and rainfall seem to be independent variables, while second variables, i.e. weight, profit and production of crop seem to be dependent variables.

With the given values $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, curve (function) given in equation (1) produces set of n equations

$$\left.\begin{aligned} y_1 &= a + bx_1 + cx_1^2 + ... + tx_1^k \\ y_2 &= a + bx_2 + cx_2^2 + ... + tx_2^k \\ &. \\ &. \\ &. \\ y_n &= a + bx_n + cx_n^2 + ... + tx_n^k \end{aligned}\right\} \qquad ... (2)$$

Our problem is to determine the constants a, b, c,…, t such that it represents the curve of best fit given by equation (1) of degree k.

If n = k+1, i.e. number of equations and number of unknown constants are equal, there is no problem in determining the unknown constants and error can be made absolutely zero. But more often n > k+1 i.e. number of equations is greater than the number of unknown constants and it is impossible to do away with all errors i.e. these equations cannot be solved exactly which satisfy set of equations (2).

Therefore, we try to determine the values of a, b, c,…, t which satisfy set of equations (2) as nearly as possible.

Substituting $x_1, x_2, ..., x_n$ for X in equation (1) we have

$$\left.\begin{array}{l} Y_1 = a + bx_1 + cx_1{}^2 + \ldots + tx_1{}^k \\ Y_2 = a + bx_2 + cx_2{}^2 + \ldots + tx_2{}^k \\ . \\ . \\ . \\ Y_n = a + bx_n + cx_n{}^2 + \ldots + tx_n{}^k \end{array}\right\} \quad \ldots (3)$$

The quantities $Y_1, Y_2, \ldots, Y_n$ are called expected or estimated values of $y_1, y_2, \ldots, y_n$ (given values of Y) for the given values of $x_1, x_2, \ldots, x_n$. Here $y_1, y_2, \ldots, y_n$ are the observed values of Y.

Let us define a quantity U, the sum of squares of errors i.e.

$$U = \sum_{i=1}^{n} (y_i - Y_i)^2$$

$$U = \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2 - \cdots - tx_i^k)^2 \qquad \ldots(4)$$

According to the principle of least squares the constant a, b,…, t are chosen in such a way that the sum of squares of residuals is minimum.

According to principle of maxima and minima (theorem of differential calculus), the extreme value (maximum or minimum) of the function U are obtained by

$$\frac{\partial U}{\partial a} = 0 = \frac{\partial U}{\partial b} = \frac{\partial U}{\partial c} = \cdots = \frac{\partial U}{\partial t}$$

(provided that the partial derivatives exist).

Let us take

$$\frac{\partial U}{\partial a} = 0 \Rightarrow \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - Y_i)^2 = 0$$

$$\Rightarrow \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2 - \cdots - tx_i^k)^2 = 0$$

$$\Rightarrow 2\sum (y_i - a - bx_i - cx_i^2 - \cdots - tx_i^k)(-1) = 0$$

$$\Rightarrow \sum y_i - na - b\sum x_i - c\sum x_i^2 - \cdots - t\sum x_i^k = 0$$

$$\Rightarrow \sum y_i = na + b\sum x_i + c\sum x_i^2 + \cdots + t\sum x_i^k \qquad \ldots(5)$$

$$\frac{\partial U}{\partial b} = 0 = \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - Y_i)^2$$

$$\Rightarrow \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2 - \cdots - tx_i^k)^2 = 0$$

$$\Rightarrow 2\sum (y_i - a - bx_i - cx_i^2 - \cdots - tx_i^k)(-x_i) = 0$$

$$\Rightarrow \sum x_i (y_i - a - bx_i - cx_i^2 - \cdots - tx_i^k) = 0$$
$$\Rightarrow \sum x_i y_i - a\sum x_i - b\sum x_i^2 - c\sum x_i^3 - \cdots - t\sum x_i^{k+1} = 0$$
$$\Rightarrow \sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3 + \cdots + t\sum x_i^{k+1}$$

Therefore, by the conditions $\frac{\partial U}{\partial a} = 0 = \frac{\partial U}{\partial b} = \frac{\partial U}{\partial c} = \cdots = \frac{\partial U}{\partial t}$, ultimately we get the following (k+1) equations

$$\left.\begin{aligned}
&\sum y_i = na + b\sum x_i + c\sum x_i^2 + \cdots + t\sum x_i^k = 0\\
&\sum y_i x_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3 + \cdots + t\sum x_i^{k+1} = 0\\
&.\\
&.\\
&.\\
&\sum y_i x_i^k = a\sum x_i^k + b\sum x_i^{k+1} + c\sum x_i^{k+2} + \cdots + t\sum x_i^{2k} = 0
\end{aligned}\right\} \quad \ldots (6)$$

where, summation extended to i from 1 to n.

In simple way equation (6) can be expressed as

$$\left.\begin{aligned}
&\sum y = na + b\sum x + \cdots + t\sum x^k\\
&\sum xy = a\sum x + b\sum x^2 + \cdots + t\sum x^{k+1}\\
&\sum x^2 y = a\sum x^2 + b\sum x^3 + \cdots + t\sum x^{k+2}\\
&.\\
&.\\
&.\\
&\sum x^k y = a\sum x^k + b\sum x^{k+1} + \cdots + t\sum x^{2k}
\end{aligned}\right\} \quad \ldots (7)$$

These equations are known as normal equations for the curve in equation (1). These equations are solved as simultaneous equations and give the value of (k+1) constants a, b, c, …, t. Substitution of these values in second order partial derivatives gives positive value of the function. Positive value of the function indicates that the values of a, b, c,…, t obtained by solving the set of equations (6), minimize U which is sum of squares of residuals . With these values of a, b, c,…, t , curve in equation (1) is the curve of best fit.

5.3 FITTING OF STRAIGHT LINE

Section 5.2 described the procedure of fitting of any curve using principle of least squares. In this Section, we are fitting straight line for the given set of points (x_i, y_i) i = 1, 2, …, n, using principle of least squares and adopting the procedure given in Section 5.2.

Let $Y = a + bX$ … (8)

be an equation of straight line and we have a set of n points (x_i, y_i); i = 1, 2, …, n. Here, the problem is to determine a and b so that the straight line $Y = a + bX$ is the line of the best fit. With given n points (x_i, y_i), let the straight line be $y_i = a + bx_i$ where, y_i is the observed value of variable Y and $a + bx_i = Y_i$ is the estimated value of Y obtained by the straight line in equation (8). According to the principle of least squares, a and b are to be determined so that the sum of squares of residuals is minimum, i.e.

$$U = \sum_{i=1}^{n} (y_i - Y_i)^2$$

$$U = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \quad \text{… (9)}$$

is minimum. From the principle of maxima and minima we take partial derivatives of U with respect to a and b and equating to zero, i.e.

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

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

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

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

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

$$\Rightarrow \sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \quad \text{… (10)}$$

and

$$\Rightarrow \frac{\partial U}{\partial b} = 0$$

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

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

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

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

$$\Rightarrow \sum_{i=1}^{n} y_i x_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \quad \text{… (11)}$$

Equations (10) and (11) are known as normal equations for straight line in equation (8) to determine a and b. In simple form, equations (10) and (11) can be written as

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

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

Values of a and b are obtained by solving equations (10) and (11). With these value of a and b, straight line Y = a + bX is the line of best fit to the given set of points (x_i, y_i) where i =1, 2,…, n.

Let us discuss a problem of fitting of straight line numerically for the given set of data.

Example 1: Fit a straight line to the following data.

x	1	6	11	16	20	26
y	13	16	17	23	24	31

Solution: Let the straight line be $Y = a + bX$ and to obtain a and b for this straight line the normal equations are

$$\sum y = na + b\sum x \quad \text{and}$$

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

Here, there is need of $\sum y$, $\sum x$, $\sum xy$ and $\sum x^2$ which are obtained by the following table

x	**y**	**x^2**	**xy**
1	13	1	13
6	16	36	96
11	17	121	187
16	23	256	368
20	24	400	480
26	31	676	806
$\sum x = 80$	$\sum y = 124$	$\sum x^2 = 1490$	$\sum xy = 1950$

Substituting the values of $\sum y$, $\sum x$, $\sum xy$ and $\sum x^2$ in the normal equations, we get

$$124 = 6a + 80b \qquad \dots (12)$$

$$1950 = 80a + 1490b \qquad \dots (13)$$

Now we solve equations (12) and (13).

Multiplying equation (12) by 80 and equation (13) by 6, i.e.

$$124 = 6a + 80b \quad]\times 80$$

and

$$1950 = 80a + 1490b \quad]\times 6$$

we get,

$$9920 = 480\ a + 6400\ b \qquad \ldots (14)$$

$$11700 = 480\ a + 8940\ b \qquad \ldots (15)$$

Subtracting (14) from (15), we obtain

$$1780 = 2540\ b$$

$$\Rightarrow b = 1780/2540 = 0.7008$$

Substituting the value of b in equation (12), we get

$$124 = 6\ a + 80 \times 0.7008$$

$$124 = 6\ a + 56.064$$

$$67.936 = 6a$$

$$\Rightarrow a = 11.3227$$

with these values of a and b the line of best fit is Y = 11.3227 + 0.7008 X.

Now let us do one problem for fitting of straight line.

E 1) Fit a straight line to the following data:

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

5.4 FITTING OF SECOND DEGREE PARABOLA

Let $Y = a + bX + cX^2$ … (16)

be the second degree parabola and we have a set of n points (x_i, y_i); i =1, 2,…, n. Here the problem is to determine a, b and c so that the equation of second degree parabola given in equation (16) is the best fit equation of parabola. Let with given n points (x_i, y_i) the second degree parabola be

$$y_i = a + bx_i + cx_i^2 \qquad \ldots (17)$$

Let $Y_i = a + bx_i + cx_i^2$ be the estimated value of Y. Then according to the principle of least squares, we have to determine a, b and c so that the sum of squares of residuals is minimum, i.e.

$$U = \sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2)^2 \qquad \ldots (18)$$

is minimum. Using principle of maxima and minima, we take partial derivatives of U with respect to a, b and c and equating to zero, i.e.

$$\frac{\partial U}{\partial a} = 0 = \frac{\partial U}{\partial b} = \frac{\partial U}{\partial c} \qquad \ldots (19)$$

Now

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

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

$$\Rightarrow \sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x^2 \qquad \ldots (20)$$

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

$$\Rightarrow \sum_{i=1}^{n} y_i x_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 \qquad \ldots (21)$$

Similarly, $\frac{\partial U}{\partial c} = 0$ provides

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

$$\Rightarrow \sum_{i=1}^{n} y_i x_i^2 = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 \qquad \ldots (22)$$

Equations (20), (21) and (22) are known as normal equations for estimating a, b and c which can be written as

$$\sum y = na + b\sum x + c\sum x^2$$

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

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

Values of a, b and c are obtained by solving equations (20), (21) and (22). With these values of a, b and c, the second degree parabola $Y = a + bX + cX^2$ is the best fit.

Now we solve a problem of fitting a second degree parabola.

Example 2: Fit a second degree parabola for the following data:

x	0	1	2	3	4
y	1	3	4	5	6

Solution: Let $Y = a + bX + cX^2$ be the second degree parabola and we have to determine a, b and c. Normal equations for second degree parabola are

$$\sum y = na + b\sum x + c\sum x^2,$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3, \text{ and}$$

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

To solve above normal equations, we need $\sum y$, $\sum x$, $\sum xy$, $\sum x^2 y$, $\sum x^2$, $\sum x^3$ and $\sum x^4$ which are obtained from following table:

x	y	xy	x^2	x^2y	x^3	x^4
0	1	0	0	0	0	0
1	3	3	1	3	1	1
2	4	8	4	16	8	16
3	5	15	9	45	27	81
4	6	24	16	96	64	256
$\sum x = 10$	$\sum y = 19$	$\sum xy = 50$	$\sum x^2 = 30$	$\sum x^2y = 160$	$\sum x^3 = 100$	$\sum x^4 = 354$

Substituting the values of $\sum y, \sum x, \sum xy, \sum x^2y, \sum x^2, \sum x^3$ and $\sum x^4$ in above normal equations, we have

$$19 = 5a + 10b + 30c \qquad \ldots (23)$$

$$50 = 10a + 30b + 100c \qquad \ldots (24)$$

$$160 = 30a + 100b + 354c \qquad \ldots (25)$$

Now, we solve equations (23), (24) and (25).

Multiplying equation (23) by 2, we get

$$38 = 10a + 20b + 60c \qquad \ldots (26)$$

Subtracting equation (26) from equation (24)

$$50 = 10a + 30b + 100c$$

$$38 = 10a + 20b + 60c$$

$$12 = 10b + 40c \qquad \ldots (27)$$

Multiplying equation (24) by 3, we get

$$150 = 30a + 90b + 300c \qquad \ldots (28)$$

Subtracting equation (28) from equation (25), we get

$$160 = 30a + 100b + 354c$$

$$150 = 30a + 90b + 300c$$

$$10 = 10b + 54c \qquad \ldots (29)$$

Now we solve equation (27) and (29)

Subtracting equation (27) from equation (29), we get

$$10 = 10b + 54c$$

$$12 = 10b + 40c$$

$$-2 = 14c$$

$$c = -2/14$$

$$c = -0.1429$$

Substituting the value of c in equation (29), we get

$$10 = 10\text{ b} + 54\times(-0.1429)$$

$$10 = 10\text{ b} - 7.7166$$

$$17.7166 = 10\text{ b}$$

$$\text{b} = 1.7717$$

Substituting the value of b and c in equation (23), we get

$$19 = 5\text{a} + 10\times(1.7717) + (-0.1429\times 30)$$

$$19 = 5\text{a} + 17.717 - 4.287$$

$$\text{a} = 1.114$$

Thus, the second degree of parabola of best fit is

$$Y = 1.114 + 1.7717\,X - 0.1429\,X^2 \qquad \ldots (30)$$

Now let us solve a problem.

E2) Fit a second degree parabola to the following data:

x	2	4	6	8	10
y	1	2	3	4	5

5.5 FITTING OF A POWER CURVE $Y = aX^b$

Let $Y = aX^b$ … (31)

be a power curve where a and b are constants. We have a set of n points (x_i, y_i) $i = 1, 2, ..., n$. Here, the problem is to determine a and b such that the power curve $Y = aX^b$ is the curve of best fit.

Taking log both sides of equation (31), we get

$$\log Y = \log(aX^b)$$

$$\log Y = \log a + \log X^b$$

$$\Rightarrow \log Y = \log a + b\log X \qquad \ldots (32)$$

Let $\log Y = U$, $\log a = A$ and $\log X = V$

Then equation (32) becomes

$$U = A + bV \qquad \ldots (33)$$

Now equation (33) is the linear form of the power curve (31).

Adopting the procedure of fitting of straight line, the normal equations for straight line equation (33) are

$$\sum u = nA + b\sum v \qquad \ldots (34)$$

$$\sum uv = A\sum v + b\sum v^2 \qquad \ldots (35)$$

equations (34) and (35) can be solved for A and b.

After getting A, we get a = antilog (A)

With these a and b, power curve equation (31) is the best fit equation of the curve to the given set of points.

Note: Here we are using log at the base 10.

Example 3: Fit power curve $Y = aX^b$ for the following data:

x	6	2	10	5	8
y	9	11	12	8	7

Solution: Let the power curve be $Y = aX^b$ and normal equations for estimating a and b are

$$\sum u = nA + b\sum v$$

$$\sum uv = A\sum v + b\sum v^2$$

where,

u = log y, v = log x and A = log a

Note: Here we are using log at the base 10.

To find the values of a and b from the above normal equations, we require $\sum u$, $\sum v$, $\sum uv$ and $\sum v^2$ which are being calculated in the following table:

x	**y**	**u=log y**	**v= log x**	**uv**	**v^2**
6	9	0.9542	0.7782	0.7425	0.6055
2	11	1.0414	0.3010	0.3135	0.0906
10	12	1.0792	1.0000	1.0792	1.0000
5	8	0.9031	0.6990	0.6312	0.4886
8	7	0.8451	0.9031	0.7632	0.8156
31	47	4.8230	3.6813	3.5296	3.0003
		$\sum u = 4.8230$	$\sum v = 3.6813$	$\sum uv = 3.5296$	$\sum v^2 = 3.0003$

Substituting the values of $\sum u = 4.8230$, $\sum v = 3.6813$, $\sum uv = 3.5296$ and $\sum v^2 = 3.0003$ in above normal equations, we obtain

4.8230 = 5A + 3.6813 b … (36)

3.5296 = 3.6813A +3.0003b … (37)

Now, we solve the equation (36) and equation (37). Multiplying equation (36) by 3.6813 and equation (37) by 5, we have

17.7549 =18.4065A +13.5519 b … (38)

17.6480 =18.4065A +15.0015 b … (39)

Subtracting equation (39) from equation (38), we have

$0.1069 = -1.4496\, b$

$\Rightarrow b = -0.0737$

Substituting the value of b in equation (36), we get

A = 1.0216

Now a = antilog A = antilog (1.0216)

a = 10.5099

Thus, the power curve of the best fit is $Y = 10.5099\ X^{-0.0737}$.

Now let us solve a problem.

E 3) Fit a power curve $Y = aX^b$ to the following data:

x	5	6	9	8	11
y	2	5	8	11	15

5.6 FITTING OF THE EXPONENTIAL CURVE $Y = ab^X$

Let $Y = ab^X$ … (40)

be an exponential curve and we have a set of n points $(x_i, y_i)\ i = 1, 2, ..., n$. We have to determine a and b such that equation (40) is the curve of best fit.

Taking log both sides of equation (40)

$$\log Y = \log a + \log b^X$$

$$\log Y = \log a + X \log b$$

Let, logY = U, log a = A and log b = B

Now, equation (40) comes in the linear form as

$$U = A + BX \qquad \ldots (41)$$

which is the equation of straight line. Normal equations for equation (41) can be obtained as

$$\sum u = nA + B\sum x \qquad \cdots (42)$$

$$\sum ux = A\sum x + B\sum x^2 \qquad \cdots (43)$$

By solving equation (42) and equation (43), we obtain A and B and finally

a = antilog A and b = antilog B.

With these a and b, the exponential curve $Y = ab^X$ is the curve of best fit for the given set of data.

Note: Here we are using log base 10.

Now let us solve a problem of fitting of exponential curve $Y = ab^X$.

Example 4: Fit the exponential curve $Y = ab^X$ from the following data.

x	2	4	6	8	10
y	1	3	6	12	24

Solution: Let the exponential curve be $Y = ab^X$ and normal equations for estimating a and b are

$$\sum u = nA + B\sum x$$

$$\sum ux = A\sum x + B\sum x^2$$

where, a = antilog(A) and b = antilog (B)

x	y	u=log y	ux	x^2
2	1	0.0000	0.0000	4
4	3	0.4771	1.9085	16
6	6	0.7782	4.6690	36
8	12	1.0792	8.6334	64
10	24	1.3802	13.8021	100
$\sum x = 30$	$\sum y = 46$	$\sum u = 3.7147$	$\sum ux = 29.0130$	$\sum x^2 = 220$

Substituting the values of $\sum x$, $\sum y$, $\sum u$, $\sum ux$ and $\sum x^2$ in the above normal equations, we get

$$3.7147 = 5A+30B \qquad \ldots (44)$$

$$29.0130 = 30\,A+220\,B \qquad \ldots (45)$$

Multiplying equation (44) by 6, we have

$$22.2882 = 30A+180B \qquad \ldots (46)$$

Subtracting equation (46) from equation (45), we have

$$29.0130 = 30A + 220B$$

$$22.2882 = 30A + 180B$$

$$6.7248 = 40\,B$$

$$B = 0.1681$$

Substituting the value of B in equation (44), we get

$$A = -\,0.26566$$

Thus, a = antilog A = antilog (− 0.26566) =1.8436 and

b = antilog B = antilog (0.1681) =1.4727

Thus, exponential curve of best fit is $Y = 1.8436(1.4727)^X$

Now let us solve a problem.

E 4) Fit an exponential curve of the type $Y = ab^X$ for the following data

x	1	2	3	4	5
y	8	15	33	65	130

5.7 FITTING OF EXPONENTIAL CURVE $Y = ae^{bX}$

Let $Y = ae^{bX}$ … (47)

be an exponential curve and we have a set of n points $(x_i, y_i)\ i = 1, 2, ..., n$. Here problem is to determine a and b such that equation (47) is the curve of best fit.

Taking log of both side of equation (47)

$$\log Y = \log a + X\, b \log e$$

Let log Y =U, log a = A and b log e = B

Now, equation (47) can be written as

$$U = A + BX \quad \text{… (48)}$$

(which is the equation of straight line)

Normal equations for equation (48) can be obtained as

$$\sum u = nA + B\sum x \quad \text{… (49)}$$

$$\sum ux = A\sum x + B\sum x^2 \quad \text{… (50)}$$

We can get A and B from these normal equations. Then

$$a = \text{antilog } A \text{ and } b = \frac{B}{\log e}$$

With these a and b, the exponential curve $Y = ae^{bX}$ is the best fit equation of the curve for the given set of data.

Note: Here also we are using log base 10.

Now let us solve a problem of fitting of exponential curve of type $Y = ae^{bX}$.

Example 5: Fit an exponential curve of the type $Y = ae^{bX}$ from the following data.

x	1	2	4
y	5	10	30

Solution: To fit the exponential curve $Y = ae^{bX}$ normal equations are

$$\sum u = nA + B\sum x$$

$$\sum ux = A\sum x + B\sum x^2$$

x	y	u =log y	u x	x^2
1	5	0.6990	0.6990	1
2	10	1.0000	2.0000	4
4	30	1.4771	5.9085	16
$\sum x = 7$	$\sum y = 45$	$\sum u = 3.1761$	$\sum x^2 = 8.6075$	$\sum ux = 21$

Now the normal equations are

$$3A + 7B = 3.1761$$

$$7A + 21B = 8.6075$$

By solving these equations as simultaneous equations, we get

$$A = 0.4604 \text{ and } B = 0.2564$$

Then,

$$a = \text{antilog }(A) = \text{antilog }(0.4604) = 2.8867$$

$$b = \frac{B}{\log e} = \frac{0.2564}{\log(2.71828)} = \frac{0.2564}{0.43429} = 0.5904$$

Thus, the curve of best fit is $Y = 2.8867\ e^{0.5904\ x}$

5.8 SUMMARY

In this unit, we have discussed:

1. The purpose of fitting the curve;
2. Residual is the difference of observed value and estimated value;
3. The principle of least squares;
4. How to fit a straight line;
5. How to fit a second degree parabola;
6. How to fit a power curve; and
7. How to fit exponential curves.

5.9 SOLUTIONS / ANSWERS

E 1) Let the straight line be $Y = a + bX$ and to obtain a and b for this straight line the normal equations are

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

and

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

S. No.	x	y	x^2	xy

1	6	5	36	30
2	7	4	49	28
3	8	3	64	24
4	9	2	81	18
5	11	1	121	11
Total	41	15	351	111

Here,

$$\sum y = 15,\ \sum x = 41,\ \sum xy = 111 \text{ and } \sum x^2 = 351$$

Then normal equations are

$$15 = 5a + 41b \qquad \ldots (51)$$

$$111 = 41a + 351b \qquad \ldots (52)$$

Now, we solve above two normal equations.

Multiplying equation (51) by 41 and equation (52) by 5, we have

$$15 = 5a + 41b \quad]\times 41$$

and

$$111 = 41 + 351b \quad]\times 5$$

we obtain

$$615 = 205\,a + 1681b \qquad \ldots (53)$$

$$555 = 205\,a + 1755\,b \qquad \ldots (54)$$

Subtracting equation (53) from equation (54), we have

$$-60 = 74\,b$$

$$\Rightarrow b = -\ 60/74 = -0.8108$$

Substituting the value of b in equation (51), we get

$$15 = 5\,a + 41 \times (-0.8108)$$

$$15 = 5\,a - 33.2428$$

$$\Rightarrow a = 9.6486$$

with these of a and b, the line of best fit is $Y = 9.6486 - 0.8108\,X$.

E2) Let $Y = a + bX + cX^2$ be the second degree parabola and we have to determine a , b and c. Normal equations for second degree parabola are

$$\sum y = na + b\sum x + c\sum x^2$$

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

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

To solve above normal equations, we need

$$\sum y, \sum x, \sum xy, \sum x^2 y, \sum x^3 \text{ and } \sum x^4$$

which are obtained from following table

x	y	xy	x^2	x^2y	x^3	x^4
2	1	2	4	4	8	16
4	2	8	16	32	64	256
6	3	18	36	108	216	1296
8	4	32	64	256	512	4096
10	5	50	100	500	1000	10000
$\sum x = 30$	$\sum y = 15$	$\sum xy = 450$	$\sum x^2 = 220$	$\sum x^2 y = 900$	$\sum x^3 = 1800$	$\sum x^4 = 15664$

Here, $\sum x = 30, \sum y = 15, \sum xy = 450, \sum x^2 y = 900,$
$\sum x^2 = 220, \sum x^3 = 1800$ and $\sum x^4 = 15664$.

Substituting these values in above normal equations, we have

15 = 5a + 30b + 220c … (55)

110 = 30a + 220b + 1800c … (56)

900 = 220a + 1800b b + 15664c … (57)

Now we solve equations (55), (56) and (57).By multiply equation (55) by 6

90 = 30a + 180b + 1320c … (58)

Subtracting equation (58) from equation (56), we get

110 = 30a + 220b +1800c

90 = 30a +180b +1320c

----

20 = 40b + 480c … (59)

Multiplying equation (55) by 44, we have

660 = 220a +1320b + 9680c … (60)

Subtracting equation (59) from equation (57), we get

900 = 220a +1800b +15664c

660 = 220a +1320b + 9680c

----

240 = 480b + 5984c … (61)

Now we solve equation (59) and equation (61)

Multiplying equation (59) by 12, we get

240 = 480b + 5760c … (62)

Subtracting equation (61) from equation (62)

$$240 = 480b + 5760c$$
$$240 = 480b + 5984c$$

$$0 = -224c$$
$$\Rightarrow c = 0$$

Substituting the value of c in equation (62), we get

$$240 = 480b + 5760 \times 0 \Rightarrow b = \frac{240}{480} = 0.5$$

Putting the values of b and c in equation (55), we get

$$15 = 5a + 30 \times (0.5) + 220 \times (0)$$
$$15 = 5a + 15$$
$$\Rightarrow a = 0$$

Thus, the second degree parabola of best fit is

$$Y = 0 + 0.5\,X + 0\,X^2 \Rightarrow Y = 0.5X$$

E 3) Let the power curve be $Y = aX^b$ and normal equations for estimating a and b are

$$\sum u = nA + b\sum v$$
$$\sum uv = A\sum v + b\sum v^2$$

where,

$$U = \log Y,\ V = \log X \text{ and } A = \log a$$

To find the values of a and b from the above normal equations we require $\sum u$ $\sum v$, $\sum uv$ and $\sum v^2$ which are being calculated in the following table:

x	y	u=log y	v=log x	uv	v
5	2	0.3010	0.6990	0.2104	0.4886
6	5	0.6990	0.7782	0.5439	0.6055
9	8	0.9031	0.9542	0.8618	0.9106
8	11	1.0414	0.9031	0.9405	0.8156
11	15	1.1761	1.0414	1.2248	1.0845
		$\sum u = 4.1206$	$\sum v = 4.3759$	$\sum uv = 3.7814$	$\sum v^2 = 3.9048$

Substituting the values $\sum u = 4.1206$, $\sum v = 4.3759$, $\sum uv$ =3.7814 and $\sum v^2 = 3.9048$ in above normal equations, we have

$4.1206 = 5A + 4.3759\,b$ … (63)

$3.7184 = 4.3759A + 3.9048\,b$ … (64)

Now we solve equation (63) and equation (64).

Multiplying equation (63) by 4.3759 and equation (64) by 5, we have

$18.0303 = 21.8795A + 19.1485\,b$ … (65)

$18.5920 = 21.8795A + 19.5240\,b$ … (66)

Subtracting equation (65) from equation (66), we have

$0.5619 = 0.3755b$

$\Rightarrow b = 1.4964$

Substituting the value of b in equation (63), we obtain

$A = -0.4571$

Now a = antilog (A) = antilog (−0.4571)

$a = 0.3491$

Thus, the power curve of the best fit is $Y = 0.3491\ X^{1.4964}$

E 4) Let the exponential curve be $Y = ab^X$ and normal equations for estimating a and be are

$$\sum u = nA + B\sum x$$

$$\sum ux = A\sum x + B\sum x^2$$

where, a = antilog(A) and b = antilog (B)

To find the values of a and b from the above normal equations we require $\sum u$ $\sum x$, $\sum ux$ and $\sum x^2$ which are being calculated in the following table:

x	**y**	**u = log y**	**ux**	$\mathbf{x^2}$
1	8	0.9031	0.9031	1
2	15	1.1762	2.3522	4
3	33	1.5185	4.5555	9
4	65	1.8129	7.2517	16
5	130	2.1139	10.5697	25
$\sum x = 15$		$\sum u = 7.5246$	$\sum ux = 25.6322$	$\sum x^2 = 55$

Substituting the values $\sum x$, $\sum u$, $\sum ux$ and $\sum x^2$ in above normal equations, we have

$7.5246 = 5A + 15\,B$

$$25.6322 = 15\ A + 55\ B$$

After solving these equations we get B = − 0.3058

and A = 2.4224

Consequently, a = antilog (2.4224) = 264.5087 and

b = antilog (B) = antilog (− 0.3058) = 0.4494

Thus, the exponential curve of best fit is $Y = 264.5087(0.4494)^X$