BACHELOR OF COMPUTER APPLICATIONS

SEMESTER 3

DCA2101

COMPUTER ORIENTED NUMERICAL METHODS

Unit 7

Curve Fitting

Table of Contents

1. INTRODUCTION

Very often in engineering and science, we obtain experimentally a number of corresponding values of two variables x and y. For example, the relationship between stress and strain on a metal strip, relationship between time and temperature rise in heating a given volume of water, height, and weight of students, etc., can be tabulated by suitable experiments. It will be necessary to find a mathematical relation between them by constructing a function $y = f(x)$. This relation can then be used to predict the dependent variable $y(x)$ for the specified value of x.

The process of finding the equation of the curve of best fit which may be suitable for predicting the unknown values is known as curve fitting.

1.1 Objectives

At the end of this unit the student should be able to:

- *Learn the linear law to fit a numerical data*
- *Learn various methods of curve fitting*
- *Analyze the advantages of different methods*
- *Apply the techniques to write an Algorithm for large data.*

Methods

In this unit, we discuss the following standard methods for curve fitting.

i) Graphic method
ii) Method of group averages
iii) Method of moments
iv) Method of least squares

2. GRAPHICAL METHOD

When the curve representing the given data is a linear law y = mx + c, we follow the steps.

Step 1: Plot the given points on the graph paper taking a suitable scale.

Step 2: Draw the straight line of best fit such that the points are evenly distributed about the line.

Step 3: Taking two suitable points (x_1, y_1) and (x_2, y_2) on the line, calculate m the slope of the line, and c its intercept on the y-axis.

Step 4: When the points representing observed values do not approximate to a straight line, a smooth curve is drawn through them. From the shape of the graph, we try to infer the law of the curve and then reduce it to the form y = mx + c.

2.1 Laws reducible to the linear Law:

We list below some of the laws in common use, indicating the way these can be reduced to the linear form by suitable substitutions.

i) Law of the form $y = mx^n + c$.

Take $x^n = X$ and $y = Y$, then the above law becomes Y = mX + c.

ii) Law of the form $y = ax^n$.

Taking logarithms of both sides, we get

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

Putting $\log_{10}x = X$ and $\log_{10}y = Y$, it reduces to the form, Y = nX + c,

where $c = \log_{10}a$.

iii) When $y = ax^n + b \log x$.

Write: $\frac{y}{\log x} = a\frac{x^n}{\log x} + b$ and take $\frac{x^n}{\log x} = X$ and $\frac{y}{\log x} = Y$, then we get a linear law:

Y = aX + b.

iv) When the law is $y = ae^{bx}$.

Taking logarithm, it becomes

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

Put $x = X$ and $\log_{10}y = Y$, it takes the form

$Y = mX + c$ where $m = b \log_{10}e$ and $c = \log_{10}a$.

Note: On appling the Natural logarithm we get

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

$\ln y = \ln a + bx$ and it takes the form of Y=mX+c

Where Y=lnln y, m=b and c= $\ln a$

v) When the law is xy = ax + by.

Dividing by x, we get $y = b\,\frac{y}{x} + a$.

Putting $=\frac{y}{x}$ X and y = Y, it reduces to the form Y = bX +a.

Example

Convert the following equations into a linear form.

(i) $y = \frac{x}{a+bx}$, (ii) $y = \frac{ax+b}{x}$, (iii) y = a + bxy, (iv) xy = ax + b.

Solution:

i) $y = \frac{x}{a+bx} \Rightarrow \frac{1}{y} = \frac{a+bx}{x} = \frac{a}{x} + b$. Take $Y = \frac{1}{y}$ and $X = \frac{1}{x}$. Then the equation becomes Y = aX + b.

ii) $y = \frac{ax+b}{x} \Rightarrow y = a + \frac{b}{x}$. Take $X = \frac{1}{x}$ and Y = y. Then the given equation can be written as Y = a + bX.

iii) y = a + bxy.

Dividing by xy we get $\frac{1}{x} = \frac{a}{xy} + b$. Take $Y = \frac{1}{x}$ and $\frac{1}{xy}$ =X. Then the given equation becomes Y = aX + b.

iv) $xy = ax + b$.

Dividing by x we get $y = a + \frac{b}{x}$. Take $X = \frac{1}{x}$ and $Y = y$. Then the given equation becomes

$Y = a + bX$.

Example: R is the resistance to motion of a train at speed V, find a law of the type $R = a + bV^2$ connecting R and V, using the following data:

V (km/hr)	10	20	30	40	50
R (kg/ton)	8	10	15	21	30

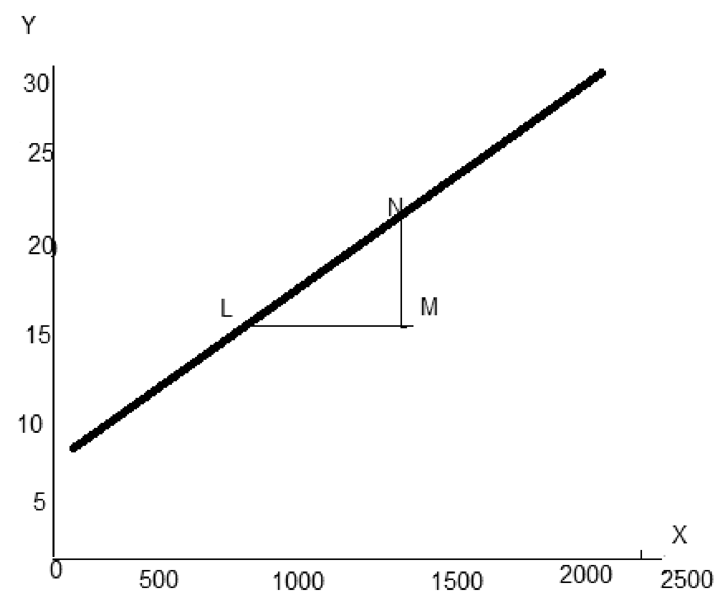
Solution: Given law is $R = a + bV^2$. Take $V^2 = x$ and $R = y$.

Then the linear law is $y = a + bx$.

The table of values of x and y are as follows:

x	100	400	900	1600	2500
y	8	10	15	21	30

Plot these points. Draw the straight line of best fit through these points (given in fig.)



Slope of this line b = $\dfrac{MN}{LM} = \dfrac{21-15}{1600-900} = \dfrac{6}{700} = 0.0085(\text{approx.})$

Since the point L(900, 15) lies on y = a + bx, we have

15 = a + 0.0085 $\times$ 900. Therefore a = 7.35.

Self-Assessment Questions - 1

1. The curve y = mx + c is ________

3. METHOD OF GROUP AVERAGES

Suppose we have n sets of observations in an experiment, such as (x_1, y_1), (x_2, y_2), …, (x_n, y_n).

Assume that a straight line $y = mx + c$ ……………(1)

fits this data.

Since the equation contains two constants m and c, we require two equations in m and c to determine them.

From the given data, when $x = x_1$, the observed value of $y = y_1$.

But from the assumed law, the expected value of $y_1 = mx_1+c$.

Now we define the residual e as $e_1 = y_1 - (mx_1+c)$.

Similarly, for the other sets of values, we construct the residuals as

$e_2 = y_2 - (mx_2+c)$, …, $e_n = y_n - (mx_n+c)$.

These expressions clearly indicate that some of the residuals may be positive and some of them may be negative.

However, the method of group averages states that the sum of all the residuals is zero. That is,

$$\sum_{i=1}^{n} e_i = 0. \qquad \text{……………(2)}$$

Now we divide the whole data into two groups and assume that (2) is true for each group.

Let us consider the first j observations constitutes one group and the rest as the second group, such that each group contains an approximately equal number of observations. Therefore using (2) for each group, we have

$$\sum_{i=1}^{j} e_i = [y_1 - (mx_1 + c)] + [y_2 - (mx_2 + c)] + \ldots + [y_j - (mx_j + c)] = 0. \ldots \quad \text{….(3)}$$

$$\sum_{i=j+1}^{n} e_i = [y_{j+1} - (mx_{j+1} + c)] + [y_{j+2} - (mx_{j+2} + c)] + \ldots + [y_n - (mx_n + c)] = 0.$$

……………(4)

On simplifying (3) and (4) we get

$\left[y_1 + y_2 + \ldots + y_j\right] - jc - m\left[x_1 + x_2 + \ldots + x_j\right] = 0$ and

$\left[y_{j+1} + y_{j+2} + \ldots + y_n\right] - (n-j)c - m\left[x_{j+1} + x_{j+2} + \ldots + x_n\right] = 0$

These can be written as

$$\frac{\left[y_1 + y_2 + \ldots + y_j\right]}{j} = \frac{m\left[x_1 + x_2 + \ldots + x_j\right]}{j} + c \quad \ldots\ldots\ldots\ldots(5)$$

$$\frac{\left[y_{j+1} + y_{j+2} + \ldots + y_n\right]}{n-j} = \frac{m\left[x_{j+1} + x_{j+2} + \ldots + x_n\right]}{n-j} + c \quad \ldots\ldots\ldots\ (6).$$

Equations (5) and (6) are sufficient to determine the unknown constants m and c.

There is no unique way of dividing the observations into two groups and hence different choices will give different values for m and c.

Example:

Use the method of group averages and find a curve of the form y = mxn, that fits the following data:

X	10	20	30	40	50	60	70	80
Y	1.06	1.33	1.52	1.68	1.81	1.91	2.01	2.11

Solution: The required curve is of the form $y = mx^n$

Taking logarithm on both sides, we get

$\log_{10} y = \log_{10} m + n \log_{10} x$

Let $\log_{10} y = Y$, $\log_{10} m = c$ and $\log_{10} x = X$, then the equation becomes

$Y = nX + c$.

Now we take logarithms of the given pairs of data and divide them into two groups, such that the first group contains the first four values and the remaining constitutes the second group, as follows.

Group-I

x	y	$Y = \log_{10} y$	$X = \log_{10} x$
10	1.06	0.0253	1.0000
20	1.33	0.1239	1.3010
30	1.52	0.1818	1.4771
40	1.68	0.2253	1.6021
		ΣY = 0.5563	**ΣX = 5.3802**

Group-II

x	y	$Y = \log_{10} y$	$X = \log_{10} x$
50	1.81	0.2577	1.6990
60	1.91	0.2810	1.7782
70	2.01	0.3032	1.8451
80	2.11	0.3243	1.9031
		ΣY = 1.1662	**ΣX = 7.2254**

Using the method of group averages, we determine the constants n and c from (5) and (6) as follows.

$$\frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4) = \frac{n}{4}(X_1 + X_2 + X_3 + X_4) + c$$

$$\frac{1}{4}(Y_5 + Y_6 + Y_7 + Y_8) = \frac{n}{4}(X_5 + X_6 + X_7 + X_8) + c$$

Substituting the values from the above tables, we get

$$\frac{0.5563}{4} = n\frac{5.3802}{4} + c \text{ and } \frac{1.1662}{4} = n\frac{7.2254}{4} + c$$

$\Rightarrow 4c + 5.3802n - 0.5563 = 0$ and

$4c + 7.2254n - 1.1662 = 0.$

Solving we get $c = -0.3055$ and $n = 0.3305$.

Therefore $m = \text{antilog } c = 0.4949$.

Hence the required curve is $y = (0.4949)x^{0.3305}$.

4. METHOD OF LEAST SQUARES

By the above methods, the relationship between the variables are not well defined and the values are not accurate. Therefore, it will be meaningless to try to pass the curve through every point. The best strategy would be to construct a single curve that would represent the general trend of the data, without necessarily passing through the individual points.

The method of finding a specific relation $y = f(x)$ for the data to satisfy as accurately as possible and such an equation is called the "*the best fitting equation*" or the "*curve of best fit*". The method is called the *principle of least squares and is* described below:

Let the set of data points be (x_i, y_i), i = 1, 2, 3 …,n and let the curve given by $y = f(x)$ be fitted to the data. At $x = x_i$, the experimental value of the ordinate is yi and the corresponding value on the fitting curve is $f(x_i)$. If e_i is the error of approximation at $x = x_i$, then we have

$$e_i = y_i - f(x_i), \quad i = 1, 2, 3 \ldots, n$$

Clearly some of the errors $e_1, e_2, \ldots e_n$ will be positive and others negative. Thus to give equal weightage to each error, we square each of these and form their sum, that is

$$E = [y1 - f(x1)]^2 + [y_2 - f(x_2)]^2 + \ldots\ldots + [y_n - f(x_n)]^2$$

$$= e_1^2 + e_2^2 + \ldots + e_n^2 .$$

The method of least squares consists in minimizing E, that is, the sum of the squares of the errors.

4.1 Fitting a straight line

Let $y = a + bx$ be the straight line to be fitted to the given data (x_i, y_i) i = 1, 2, …n by the method of least squares.

We have to find the constants a and b. For any x_i, the expected value of y_i (that is, the value calculated from the equation) is $a + b\,x_i$ and the observed value of y is y_i.

Hence error e_i = observed value – expected value

$$e_i = y_i - (a + bx_i)$$

Let E be the sum of the squares of the errors, that is $E = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$

Clearly E is a function of the parameters a and b.

For E to be minimum, the conditions are

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

Partially differentiating E with respect to a,

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

Equating this to zero, we get

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

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

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

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

where $\sum_{i=1}^{n} a = a + a + a + \ldots + a = na$

Now, differentiating partially E with respect to b,

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

Equating this to zero, we get

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} ax_i + b\sum_{i=1}^{n} x_i^2 \quad -(2)$$

Equations (1) and (2) are two simultaneous linear equations from which a and b can be solved. Thus we get the equation of the line best fitting the data as $y = a + bx$. The equations (1) and (2) are called normal equations.

Dropping off the suffixes, they can be written as

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

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

Example

Find the equation of the best fitting straight line for the data

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

Let $y = a + bx$ be the required best fit straight line.

Its normal equations are given by

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

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

There are 8 tabulated data, therefore n = 8

x	**y**	**x y**	**x^2**
1	1	1	1
3	2	6	9
4	4	16	16
6	4	24	36
8	5	40	64
9	7	63	81
11	8	88	121
14	9	126	196
$\Sigma x = 56$	**$\Sigma y = 40$**	**$\Sigma xy = 364$**	**$\Sigma x^2 = 524$**

The normal equations becomes

$40 = 8a + 56b$ – (i)

$364 = 56a + 524b$ – (ii)

(ii) – (i)×7, we get

$132\,b = 84$

$b = 0.64$

Substituting the value of b in (i) we get $a = 0.52$

The equation $y = a + bx$ becomes

$y = 0.52 + 0.64\,x$.

Example:

Using the method of least squares, find the straight line y = ax + b that fits the following data:

X	0.5	1.0	1.5	2.0	2.5	3.0
Y	15	17	19	14	10	7

Solution: The given straight line fit by y = ax + b.

The normal equations of least square fit are

$\sum y = nb + a\sum x$

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

From the given data, we get

X	y	x y	x^2
0.5	15	7.5	0.25
1.0	17	17.0	1.00
1.5	19	28.5	2.25
2.0	14	28.0	4.00
2.5	10	25.0	6.25
3.0	7	21.0	9.00
$\Sigma x = 10.5$	$\Sigma y = 82$	$\Sigma xy = 127$	$\Sigma x^2 = 22.75$

Substituting these summations into normal equations we obtain

$22.75\,a + 10.5\,b = 127$

$10.5\,a + 6b = 82$

Solving we get

a = -3.7714, b = 20.2667

Therefore the required fit for the given data is y = -3.7714x + 20.2667.

4.2 Fitting a Parabola $y = a + bx + cx^2$

Consider a set of n pairs of the given values (x, y) for fitting the curve

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

The error = $y - (a + bx + cx^2)$ the difference between the observed and estimated values of y. We have to find the parameters a, b, c such that the sum of the squares of the residuals is the least.

Let $E = \sum_{i=1}^{n} [y_i - (a + bx_i + cx_i^2]^2$

Treating E is a function of three variables a, b, c.

For E to be a minimum, we have

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

That is

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

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

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

Simplifying the above equations we have

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n} x_i y = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3$$

$$\sum_{i=1}^{n} x_i y = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3$$

or

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

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

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

These equations are called *normal equations* and on solving we get the values of a, b, c so that $y = a + bx + cx^2$ is the best fitting parabola of second degree in the least square method.

Example: Fit a parabola of second degree $y = a + bx + cx^2$ for the data:

X	0	1	2	3	4
Y	1	1.8	1.3	2.5	2.3

Solution: The normal equations for $y = a + bx + cx^2$ are

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

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

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

here n = 5.

The relevant table is given below:

x	y	x^2	x^3	x^4	xy	$x^2 y$
0	1	0	0	0	0	0
1	1.8	1	1	1	1.8	1.8
2	1.3	4	8	16	2.6	5.2
3	2.5	9	27	81	7.5	22.5
4	2.3	16	64	256	9.2	36.8
$\Sigma x = 10$	$\Sigma y = 8.9$	$\Sigma x^2 = 30$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$	$\Sigma xy = 21.1$	$\Sigma x^2 y = 66.3$

The normal equations becomes

$5a + 10b + 30c = 8.9$ (1)

$10a + 30b + 100c = 21.1$ (2)

$30a + 100b + 354c = 66.3$ (3)

Solving these simultaneous equations, we get

$a = 1.078$

$b = 0.414$

$c = -0.021$

Hence the required parabola is

$y = 1.078 + 0.414x - 0.021x^2$

Example: If P is the pull required to lift a load W by means of a pulley block, find a linear law of the form $P = mW + C$ connecting P and W, using the following data :

P	12	15	21	25
W	50	70	100	120

Where P and W are taken in kilogram weight. Compute P when W =150 kg.

Solution: The normal equations for $P = mW + C$ are

$\sum P = 4C + m\sum W$

$\sum WP = C\sum W + m\sum W^2$

The values $\sum P$, $\sum W$, $\sum W^2$ and $\sum WP$ are calculated by means of the following table :

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

The normal equations become

$73 = 4C + 340m$

$6750 = 340\,C + 31800\,m$

Solving these equations, we get

$C = 2.2785$

$m = 0.1879$

Hence the line of best is

$P = 2.2785 + 0.1879\,W$

when $W = 150$ kgs.

$P = 2.2785 + 0.1879 \times 150 = 30.4635$ kg.

Note

For the sake of convenience and ease in calculations, it is sometimes advisable to change the origin and scale with the substitutions $X = \frac{x - A}{h}$ and $Y = \frac{y - B}{h}$, where A and B are the assumed means (or middle values) of x and y series respectively and h is the width of the interval.

Example

Fit a parabola $y = a + bx + cx^2$ by the method least squares for the data

X	2	4	6	8	10
Y	3.07	12.85	31.47	57.38	91.29

Solution: Let us choose $X = \frac{x - 6}{2}$ and $Y = y - 39$

Since $\bar{x} = \frac{30}{5} = 6 = A$ (say)

$\bar{y} = \frac{196.06}{5} \simeq 39 = B$ (say).

The normal equations for $Y = a + bX + cX^2$ are

$\sum Y = na + b\sum X + C\sum X^2$

$\sum XY = a\sum X + b\sum X^2 + C\sum X^3$

$\sum X^2 Y = a\sum X^2 \quad X2 \ + b\sum X^3 \ + C\sum X^4$

Here n = 5.

The relevant table is given below:

X	Y	$X = \frac{x-6}{2}$	Y = y −39	XY	X^2Y	X^2	X^3	X^4
2	3.07	−2	−35.93	71.86	−143.72	4	−8	16
4	12.85	−1	−26.15	26.15	−26.15	1	−1	1
6	31.47	0	−7.53	0	0	0	0	0
8	57.83	1	18.38	18.38	18.38	1	1	1
10	91.29	2	52.29	104.58	209.16	4	8	16
		$\Sigma X = 0$	$\Sigma Y = 1.06$	$\Sigma XY=220.97$	$\Sigma X^2Y=57.67$	$\Sigma X^2=10$	$\Sigma X^3 = 0$	$\Sigma X^4 = 34$

The normal equations becomes

$5a + 10c = 1.06$

$10b = 220.97$

$10a + 34c = 57.67$

Solving these simultaneous, equations, we get,

$a = -7.73, \ b = 22.097, \ c = 3.97$

The parabola $Y = a + bX + cX^2$ becomes

$Y = -7.73 + 22.097X + 3.97X2$

or

$$y - 39 = -7.73 + 22.09\left(\frac{x-6}{2}\right) + 3.97\left(\frac{x-6}{2}\right)^2$$

on simplification we obtain

$y = 0.7 - 0.86x + 0.9925x^2$

Example

Fit a second-degree parabola $y = a + bx + cx^2$ in the least square method for the following data and hence estimate y at $x = 6$.

X	1	2	3	4	5
Y	10	12	13	16	19

Solution: Let us choose $X = x - 3$. We have $\bar{y} = \frac{\Sigma y}{n} = \frac{70}{5} = 14$

and let $Y = y - \bar{y} = y - 14$.

The normal equations for $Y = a + bX + cX^2$ are

$\sum Y = na + b\sum X + C\sum X^2$

$\sum XY = a\sum X + b\sum X^2 + C\sum X^3$

$\sum X^2 Y = a\sum X^2 \;\; X2 + b\sum X^3 + C\sum X^4$

The relevant table is as below:

X	Y	X = x – 3	Y = y –14	XY	X^2Y	X^2	X^3	X^4
1	10	–2	– 4	8	– 16	4	– 8	16
2	12	– 1	– 2	2	– 2	1	– 1	1
3	13	0	– 1	0	0	0	0	0
4	16	1	2	2	2	1	1	1
5	19	2	5	10	20	4	8	16
		$\Sigma X = 0$	$\Sigma Y = 0$	$\Sigma XY = 22$	$\Sigma X^2Y = 4$	$\Sigma X^2 = 10$	$\Sigma X^3 = 0$	$\Sigma X^3 = 34$

The normal equations becomes

$0 = 5a + 0\,b + 10\,c$

$22 = 0a + 10\,b + 0\,c$

$4 = 10\,a + 0b + 34\,c$

That is $5a + 10\,c = 0$

$10b = 22$

$10a + 34\,c = 4$

Solving, we get $a = -0.58$, $b = 2.2$ and $c = 0.29$

The parabola of fit $Y = a + bX + cX^2$ becomes

$\Rightarrow y - 14 = -0.58 + 2.2\,(x - 3) + 0.29\,(x - 3)^2$

$\Rightarrow y = 14 - 0.58 + 2.2x - 6.6 + 0.29(x^2 - 6x + 9)$

$\Rightarrow y = 9.43 + 0.46x + 0.29x^2$ is the required equation of the parabola

Put $x = 6$ we get

$y_{\text{at } x=6} = 22.63$

Problem: Applying the method of least squares find an equation of the form y = ax + bx2 that fits the following data:

X	1	2	3	4	5	6
Y	2.6	5.4	8.7	12.1	16.0	20.2

Solution: The required curve fit is $y = ax + bx^2$, which can be written as $\frac{y}{x} = a + bx$.

Let $\frac{y}{x} = Y$, then the curve to be fitted is Y = a + bx.

The corresponding data when rewritten takes the form

x	1	2	3	4	5	6
y	2.6	2.7	2.9	3.025	3.2	3.367

The corresponding normal equation are

$$b\sum x_i^2 + a\sum x_i = \sum x_i Y_i$$
$$b\sum x_i + na = \sum Y_i$$

From the given data, we have

x	Y	xY	x^2
1	2.6	2.6	1
2	2.7	5.4	4
3	2.9	8.7	9
4	3.025	12.1	16
5	3.2	16.0	25
6	3.367	20.2	36
$\sum x_i = 21$	$\sum Y_i = 17.792$	$\sum x_i Y_i = 65.0$	$\sum x_i^2 = 91$

Substituting these summations into normal equations, we obtain

$91 b + 21 a = 65$ ……………………(1)

$21 b + 6a = 17.792$ ……………… (2)

Solving (1) and (2), we get $b = 0.15589$, $a = 2.41973$.

Hence, the required equation of fit for the given data is

$Y = 0.15589 x + 2.41973$

That is $y = 0.15589 x^2 + 2.41973 x$.

4.3 Fitting a Curve of the Form $y = ax^b$:

Data from experimental observations need not be linear. Sometimes we may be interested to fit a curve of the form $y = ax^b$.

This can be liberalized by taking logarithms on both sides, we get

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

Let $\log_{10}y = \log_{10} a + b \log_{10}x = X$. Then the above equations reduces to $Y = A + bX$, which is linear in Y and X.

Example

Using the method of least squares, find a relation of the form

$y = ax^b$, that fits the data,

x	2	3	4	5
y	27.8	62.1	110	161

Solution: Let the equation of fit be,

$$y = ax^b$$

Taking logarithm on both sides, we get

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

which is of the form $Y = A + bX$, where $\log10y = Y$, $\log_{10} a = A$, $\log_{10}x = X$.

The corresponding normal equations are

$\sum Y = nA + b\sum X$

$\sum XY = A\sum x + b\sum X^2$

From the data

X	Y	XY	X^2
0.3010	1.4440	0.4346	0.0906
0.4771	1.7931	0.8555	0.2276
0.6021	2.0414	1.2291	0.3625
0.6990	2.2068	1.5426	0.4886
$\Sigma X = 2.0792$	$\Sigma Y = 7.4853$	$\Sigma XY = 4.0618$	$\Sigma X^2 = 1.1693$

Substituting these in normal equations we get

1. $1693b + 2.0792A = 4.0618$
2. $2.0792b + 4A = 7.4853$

Solving these equations we obtain $b = 1.9311$, $A = 0.8678$.

Therefore $a = \text{antilog } 0.8678 = 7.375$.

Hence the required fit for the given data is $y = 7.375x^{1.9311}$.

4.4 Fitting an Exponential Curve

Suppose we have n set of observations (x_i, y_i), $i = 1, 2, \ldots, n$ in an experiment and we wish to fit an exponential curve of the form,

$y = ae^{bx}$

to these observations. At first, we take logarithms on both sides to get

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

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

Then it is equivalent to fitting a straight line. The procedure is given in the following example.

Example

Using the principle of least squares, fit an equation of the form y = aebx to the data

x	1	2	3	4
y	1.65	2.70	4.50	7.35

Solution: Let $y = ae^{bx}$ is the desired fit to the given data. Taking logarithms both sides, we have $\log_{10} y = \log_{10} a + bx \log_{10} e$.

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

Then the above equation assumes the form

Y = Bx + A.

The least square normal equations are

$$B\sum x_i^{\,2} + A\sum x_i = \sum x_i Y_i$$

$$B\sum x_i + 4A = \sum Y_i$$

From the given data, we have

x	y	$Y = \log_{10} y$	x^2	xY
1	1.65	0.2175	1	0.2175
2	2.70	0.4314	4	0.8628
3	4.50	0.6532	9	1.9596
4	7.35	0.8663	16	3.4652
$\sum x_i = 10$	$\sum y_i = 16.2$	$\sum Y_i = \log_{10} y$	$\sum x_i^2 = 30$	$\sum x_i Y_i = 6.5051$

Substituting these summations into normal equation

30 B + 10 A = 6.5051

10 B + 4A = 2.1684.

Solving these equations, we obtain B = 0.2168, A = 0.

Now $B = b \log_{10} e = 0.4343b = 0.2168$.

Therefore, b = 0.4992, a =1.

Hence, the required equation of fit is $y = e^{0.4992x}$.

5. METHOD OF MOMENTS

Consider n set of observations (x_i, y_i), i = 1, 2, …, n in an experiment such that x_i's are equally spaced. That is, $x_2 - x_1 = x_3 - x_2 = \ldots = x_n - x_{n-1} = \Delta x$.

For such set of observations, we shall define the observed moments as follows.

$\mu 1$ = The first moment = $\sum y_i \Delta x = \Delta x \Delta \sum y_i$

μ 2 = The second moment = $\sum x_i y_i \Delta x = \Delta x \sum x_i y_i$

μ 3 = The third moment = = $\sum x_{i^2} y_i \, \Delta x = \sum x_{i^2} y_i$

and so on.

These moments are known as moments of the observed values of y. We shall also define the expected moments of the computed values of y, assuming y = f(x) be an equation of the curve fitting the given data, as

$\gamma 1$ = The first moment = $\int y \, dx = \int f(x)dx$

γ 2 = The second moment = $\int xy \, dx = \int xf(x)dx$

γ 3 = The third moment = $\int x^2 y \, dx = \int x^2 f(x)dx$.

The method of moments states that the moments of the observed y's are respectively equal to the expected moments. That is

$\mu 1 = \gamma_1$ $\mu 2 = \gamma_2$, $\mu_3 = \gamma_3$, and so on.

The observed moments can be easily calculated from the given data, while the moments of the expected y's can be computed as follows.

$$\gamma_1 = \int y \, dx = \int_{x_1 - (\frac{\Delta x}{2})}^{x_n + (\frac{\Delta x}{2})} f(x)dx$$

$$\gamma_2, \; = \int xy \, dx = \int_{x_1 - (\frac{\Delta x}{2})}^{x_n + (\frac{\Delta x}{2})} xf(x)dx$$

$$\gamma_{3,} = \int x^2 y\, dx = \int_{x_1-(\frac{\Delta x}{2})}^{x_n+(\frac{\Delta x}{2})} x^2 f(x)dx$$

Thus, from the principle of the method of moments, we have the following set of equations:

$$\Delta x \sum y_i = \int_{x_1-(\frac{\Delta x}{2})}^{x_n+(\frac{\Delta x}{2})} f(x)dx$$

$$\Delta x \sum x_i y_i = \int_{x_1-(\frac{\Delta x}{2})}^{x_n+(\frac{\Delta x}{2})} xf(x)dx$$

$$\Delta x = \sum x_{i^2} y_i = \int_{x_1-(\frac{\Delta x}{2})}^{x_n+(\frac{\Delta x}{2})} x^2 f(x)dx$$

Given the form of f(x), we can determine the unknown constants in f(x).

Example

Fit straight line of the form y = ax + b, to the following data by the method of moments

X	2	3	4	5
Y	27	40	55	68

Solution: We have to determine two constants for the required curve fit. We compute the first –two moments.

$$\mu 1 = \Delta x \sum_{i=1}^{4} y_i = 1(27 + 40 + 55 + 68) = 190$$

$$\mu 2 = \Delta x \sum_{i=1}^{4} x_i y_i = 1(54 + 120 + 220 + 340) = 734.$$

We compute the expected moments for the given y. Therefore

$\gamma_1 = \int_{1.5}^{5.5} (ax+b)dx = \left[a\frac{x^2}{2} + bx \right]_{1.5}^{5.5} = 14a + 4b$

$\gamma_2 = \int xy\, dx = \int_{1.5}^{5.5} x(ax+b)dx = \left[a\frac{x^3}{3} + b\frac{x^2}{2} \right]_{1.5}^{5.5} = \frac{163}{3}a + 14b$

Using the method of moments, we have

$14a + 4b = 190$, $\frac{163}{3}a + 14b = 734$.

Solving these equations, we get a = 12.9375, b = 2.2188.

Therefore the required straight line fit is y = 12.9375 x + 2.2188.

6. SUMMARY

Approximating curve is the graph of data obtained through measurement or observation. In this unit, we discussed various types of fitting the curve from the numerical data. Based on this mathematical equation, predictions can be made in many statistical investigations. The least squares method is the best curve fitting method and is easily implemented on computers than other methods like the method of moments, method of group averages, and graphical method.

7. TERMINAL QUESTIONS

1. Given the following data

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

   Find the straight line and the parabola of best fit.

2. Fit parabola of second degree $y = a + bx + cx^2$ for the data

X	0	1	2	3	4
Y	1	1.8	1.3	2.5	2.3

3. Using the method of group averages, find a curve of the form

   $y = a + bx^2$, that fits the following data

x	20	30	35	40	45	50
Y	10.0	11.0	11.8	12.4	13.5	14.4

8. ANSWERS

Self Assessment Questions

1. Linear Equation is called a slope-intercept form of the equation.

Terminal Questions

1. ($y = 9.1x - 3$, $y = 1.4 + 0.3x + 2.2x^2$)
2. ($y = 1.078 + 0.414x - 0.021x^2$)
3. ($y = 9.1799 + 2.083 \times 10^{-3}x^2$)