Week 5

Curve Fitting by Least Squares

In many branches of applied mathematics and engineering sciences we come across experiments and problems, which involve two variables. For example, it is known that the speed ***v*** of a ship varies with the horsepower ***p*** of an engine according to the formula $p = a + bv^3$. Here a and b are the constants to be determined. For this purpose we take several sets of readings of speeds and the corresponding horsepowers. The problem is to find the best values for a and b using the observed values of v and p. Thus, the general problem is to find a suitable relation or law that may exist between the variables x and y from a given set of observed values $(x_{i,}y_i)$, i=1,2,..........n, Such a relation connecting x and y is known as empirical law. For above example, $x = v$ and $y = p$.

The process of finding the equation of the curve of best fit, which may be most suitable for predicting the unknown values, is known as curve fitting. Therefore, curve fitting means an exact relationship between two variables by algebraic equations.

Some simple functions commonly used to fit data are:

- straight line or linear curve: $Y = a + bX$
- parabola of second degree or quadratic curve: $Y = a + bX + cX^2$
- parabola of third degree or cubic curve: $Y = a + bX + cX^2 + dX^3$
- exponential curve: $Y = ab^X$ or $Y = ae^{bX}$
- power curve: $Y = aX^b$
- hyperbola $\frac{1}{Y} = a + bX$

  and so on

There are following methods for fitting a curve.

I. Graphic method
II. Method of group averages
III. Method of moments
IV. Principle of least square.

Out of above four methods, we will only discuss and study here principle of least square.

Least Square Method

Least Square Method (LSM) is a mathematical procedure for finding the curve of best fit to a given set of data points.

The method of least squares helps us to find the values of unknowns a and b in such a way that the following two conditions are satisfied:

- The sum of the residual (deviations) of observed values of Y and corresponding expected (estimated) values of Y will be zero. $\sum(Y - \hat{Y}) = 0$
- The sum of the squares of the residual (deviations) of observed values of Y and corresponding expected values ($\hat{Y}$) should be at least $\sum(Y - \hat{Y})^2$.

  Note: Residual is the difference between observed and estimated values of dependent variable

Method of Least Squares can be used for establishing linear as well as non-linear relationships.

Fitting of a Straight Line:

A straight line can be fitted to the given data by the method of least squares. The equation of a straight line or least square line is $Y=a+bX$ where a and b are constants or unknowns.
To compute the values of these constants we need as many equations as the number of constants in the equation. These equations are called normal equations. In a straight line there are two constants a and b so we require two normal equations.

Normal Equation for '*a*' $\sum Y = na + b\sum X$………………………. (1)

Normal Equation for '*b*' $\sum XY = a\sum X + b\sum X^2$…………………………**(2)**

The direct formula of finding a and b is written as

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} \quad \text{and} \quad a = \bar{Y}\text{-b}\bar{X}$$

<u>Example:</u> Find the least square line *y*=*a*+*bx* for the data:

x	–2	–1	0	1	2
y	1	2	3	3	4

Solution:

X	Y	X^2	XY
–2	1	4	–2
–1	2	1	–2
0	3	0	0
1	3	1	3
2	4	4	8
$\sum X$=0	$\sum Y$=13	$\sum X^2$ =10	$\sum X$Y=7

Using above calculated values in eq. 1 & 12

$13 = 5a + b.(0)$..................(3)

$7 = a.(0) + 10b$...................(4)

solving 3 & 4 simultaneously

from eq. 3 we get, $13 = 5a$ or $a = \frac{13}{5}$=2.6

from eq. 4 we get, $7 = 10b$ or b= $\frac{7}{10}$=0.7

The required of least square line is : y=2.6+(0.7)x

Example:

Fit a straight line to the following set of data points:

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

Ans: straight line is y = 1.6 + 1.2x

Example:

By the method of least squares, find a straight line that best fits the following data points.

x	0	1	2	3	4
y	1.0	2.9	4.8	6.7	8.6

Solution: Let line of best fit be given by $y = ax + b$ ... ①

Normal equations are given by:

$$\sum y = a\sum x + nb \quad \text{... ②}$$

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

Calculating $\sum x$, $\sum y$, $\sum xy$ and $\sum x^2$

x	y	xy	x^2
0	1.0	0	0
1	2.9	2.9	1
2	4.8	9.6	4
3	6.7	20.1	9
4	8.6	34.4	16
$\sum x = 10$	$\sum y = 24$	$\sum xy = 67.0$	$\sum x^2 = 30$

Substituting values of $\sum x, \sum y, \sum xy$ and $\sum x^2$ in ② and ③

$$\Rightarrow 24 = 10a + 5b \qquad \text{... ④}$$

and $$67 = 30a + 10b \qquad \text{... ⑤}$$

Solving ④and ⑤, we get $a = 1.9$ and $b = 1$

Substituting in ①, line of best fit is $y = 1.9x + 1$

Example: Fit a straight line to following data

x	0	1	2	3	4
y	1.0	1.8	3.3	4.5	6.3

Solution: Let line of best fit be given by $y = ax + b$ ... ①

Normal equations are given by:

$$\sum y = a \sum x + nb \qquad \text{... ②}$$

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

Calculating $\sum x, \sum y, \sum xy$ and $\sum x^2$

x	y	xy	x^2
0	1.0	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
$\sum x = 10$	$\sum y = 16.9$	$\sum xy = 47.1$	$\sum x^2 = 30$

Substituting values of $\sum x, \sum y, \sum xy$ and $\sum x^2$ in ② and ③

$$\Rightarrow 16.9 = 10a + 5b \qquad \text{... ④}$$

and $$47.1 = 30a + 10b \qquad \text{... ⑤}$$

Solving ④and ⑤, we get $a = \frac{133}{100} = 1.33$ and $b = \frac{18}{25} = 0.72$

Substituting in ①, line of best fit is $y = 1.33x + 0.72$

Example : Fit a straight line to the given data regarding x as the independent variable.

x	1	2	3	4	5	6
y	1200	900	600	200	110	50

Sol. Let the straight line obtained from the given data by $y = a + bx$ ...(1)

Then the normal equations are $\sum y = na + b\sum x$ ...(2)

$$\sum xy = a\sum x + b\sum x^2 \quad ...(3)$$

x	y	x^2	xy
1	1200	1	1200
2	900	4	1800
3	600	9	1800
4	200	16	800
5	110	25	550
6	50	36	300
$\sum x = 21$	$\sum y = 3060$	$\sum x^2 = 91$	$\sum xy = 6450$

Putting all values in the equations (2) and (3), we get

$$3060 = 6a + 21b$$

$$6450 = 21a + 91b$$

Solving these equations, we get

$$a = 1361.97 \quad \text{and} \quad b = -243.42$$

Hence the fitted equation is $y = 1361.97 - 243.42x$. **Ans.**

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

Sol. Here we from the following table:

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
$\Sigma x = 546$	$\Sigma y = 548$	$\Sigma xy = 37422$	$\Sigma x^2 = 37314$

Let the equation of straight line to be fitted be

$$y = a + bx \qquad \text{...(1)}$$

And the normal equations are

$$\Sigma y = an + b\Sigma x \qquad \text{...(2)}$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \qquad \text{...(3)}$$

$$\Rightarrow \qquad 8a + 546b = 548$$

$$546a + 37314b = 37422$$

Solving these equations, we get

$$a = 39.5454,\ b = 0.4242$$

Hence from (1)

$$y = 39.5454 + 0.4242x. \quad \textbf{Ans.}$$

Example 3 If F is the force required to lift a load W, by means of a pulley, fit a linear expression $F = a + bW$ against the following data:

W	50	70	100	120
F	12	15	21	25

Solution: Line for best fit is given as $\quad F = a + bW \qquad$ ... ①

Normal equations are given by:

$$\sum F = na + b\sum W \qquad \text{... ②}$$

and $$\sum WF = a\sum W + b\sum W^2 \qquad \text{... ③}$$

$$\sum F = na + b\sum W \qquad \text{... ②}$$

and $$\sum WF = a\sum W + b\sum W^2 \qquad \text{... ③}$$

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

Substituting values of $\sum W$, $\sum F$, $\sum WF$ and $\sum W^2$ in ② and ③

$$\Rightarrow 73 = 4a + 340b \qquad \text{... ④}$$

and $$6750 = 340a + 31800b \qquad \text{... ⑤}$$

Solving ④and ⑤, we get $a = 2.2759$ and $b = 0.1879$

Substituting in ①, line of best fit is $F = 2.2759 + 0.1879W$

<u>Fitting a second degree parabola:</u> $\boldsymbol{Y = a + bX + cX^2}$

The three normal equations are

$$\sum y \quad = a\sum x^2 + b\sum x + nc$$

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

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

Example 4 Fit a parabola $y = ax^2 + bx + c$ to the given data

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

Solution: Let the parabola of best fit be given by $y = ax^2 + bx + c$ ...①

Normal equations are given by:

$$\sum y \quad = a\sum x^2 + b\sum x + nc \qquad \text{...②}$$

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

$$\sum x^2y = a\sum x^4 + b\sum x^3 + c\sum x^2 \qquad \text{...④}$$

x	y	xy	x^2	x^2y	x^3	x^4
10	14	140	100	1400	1000	10000
12	17	204	144	2448	1728	20736
15	23	345	225	5175	3375	50625
23	25	575	529	13225	12167	279841
20	21	420	400	8400	8000	160000
$\sum x =$ **80**	$\sum y =$ 100	$\sum xy =$ 1684	$\sum x^2 =$ 1398	$\sum x^2y =$ 30684	$\sum x^3 =$ 26270	$\sum x^4 =$ 521202

Substituting values of $\sum x, \sum y, \sum xy$ and $\sum x^2$ in ② and ③ and ④

$$\Rightarrow 100 = 1398a + 80b + 5c \qquad \text{... ⑤}$$

$$1684 = 26270a + 1398b + 80c \qquad \text{... ⑥}$$

$$30648 = 521202a + 26270b + 1398c \qquad \text{... ⑦}$$

Solving ⑤ ⑥ and ⑦ , we get $a = -0.07$, $b = 3.01$, $c = -8.73$

Substituting in ①, parabola of best fit is $y = -0.07\,x^2 + 3.01x - 8.73$

Example 5 Fit a 2[nd] parabola to the given data

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

Solution: Let the parabola of best fit be given by $y = ax^2 + bx + c$ ... ①

x	y	xy	x^2	x^2y	x^3	x^4
1	1	1	1	1	1	1
3	2	6	9	18	27	81
4	4	16	16	64	64	256
6	4	24	36	144	216	1296
8	5	40	64	320	512	4096
9	7	63	81	567	729	6561
11	8	88	121	968	1331	14641
14	9	126	196	1764	2744	38416
$\sum x =$ 56	$\sum y =$ 40	$\sum xy =$ 364	$\sum x^2 =$ 524	$\sum x^2y =$ 3846	$\sum x^3 =$ 5624	$\sum x^4 =$ 65348

Normal equations are given by:

$$\sum y = a\sum x^2 + b\sum x + nc \qquad \text{...②}$$

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

$$\sum x^2y = a\sum x^4 + b\sum x^3 + c\sum x^2 \qquad \text{...④}$$

Substituting values of $\sum x, \sum y, \sum xy$ and $\sum x^2$ in ② and ③ and ④

$$\Rightarrow 40 = 524a + 56b + 8c \qquad \text{... ⑤}$$

$$364 = 5624a + 524b + 56c \qquad \text{... ⑥}$$

$$3846 = 65348a + 5624b + 524c \qquad \text{... ⑦}$$

Solving ⑤ ⑥ and ⑦ , we get

$$a = \frac{103}{11229} = 0.009 \,,\ b = \frac{8672}{11229} = 0.77 \,,\ c = \frac{4375}{22458} = 0.195$$

Substituting in ①, parabola of best fit is $y = 0.009\, x^2 + 0.77x + 0.195$

Example Find the least square polynomial approximation of degree two to the data.

x	0	1	2	3	4
y	-4	-1	4	11	20

Sol. Let the equation of the polynomial be $y = a + bx + cx^2$ ...(1)

x	y	xy	x^2	x^2y	x^3	x^4
0	-4	0	0	0	0	0
1	-1	-1	1	-1	1	1
2	4	8	4	16	8	16
3	11	33	9	99	27	81
4	20	80	16	320	64	256
$\sum x = 10$	$\sum y = 30$	$\sum xy = 120$	$\sum x^2 = 30$	$\sum x^2y = 434$	$\sum x^3 = 100$	$\sum x^4 = 354$

The normal equations are,

$$\sum y = na + b\sum x + c\sum x^2 \quad ...(2)$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 \quad ...(3)$$

$$\sum x^2y = a\sum x^2 + b\sum x^3 + c\sum x^4 \quad ...(4)$$

Here $n = 5$, $\sum x = 10$, $\sum y = 30$, $\sum xy = 120$, $\sum x^2 = 30$, $\sum x^2y = 434$, $\sum x^3 = 100$,

$\sum x^4 = 354$.

Putting all these values in (2), (3) and (4), we get

$$30 = 5a + 10b + 30c \quad ...(5)$$

$$120 = 10a + 30b + 100c \quad ...(6)$$

$$434 = 30a + 100b + 354c \quad ...(7)$$

On solving these equations, we get $a = -4$, $b = 2$, $c = 1$. Therefore required polynomial is $y = -4 + 2x + x^2$, **Ans.**

Example 5: Fit a second degree curve of regression of y on x to the following data:

x	1	2	3	4
y	6	11	18	27

Sol. We form the following table:

x	y	x^2	x^3	x^4	xy	x^2y
1	6	1	1	1	6	6
2	11	4	8	16	22	44
3	18	9	27	81	54	162
4	27	16	64	256	108	432
$\Sigma x = 10$	$\Sigma y = 62$	$\Sigma x^2 = 30$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$	$\Sigma xy = 190$	$\Sigma x^2y = 644$

The equation of second degree parabola is given by

$$y = a + bx + cx^2 \qquad \text{...(1)}$$

And the normal equations are

$$\Sigma y = an + b\Sigma x + c\Sigma x^2 \quad ...(2)$$

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

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

$$\Rightarrow \quad \left.\begin{aligned} 4a + 10b + 30c &= 62 \\ 10a + 30b + 100c &= 190 \\ 30a + 100b + 354c &= 644 \end{aligned}\right\} \Rightarrow a = 3, b = 2, c = 1$$

Hence $y = 3 + 2x + x^2$. **Ans.**

Example 6: By the method of least squares, find the straight line that best fits the following data:

x	1	2	3	4	5	6	7
y	14	27	40	55	68	77	85

Sol. The equation of line is

$$y = a + bx \quad ...(1)$$

The normal equations are $\Sigma y = an + b\Sigma x$ ...(2)

and

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad ...(3)$$

Now we from the following table:

x	y	xy	x^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
6	77	462	36
7	85	595	49
$\Sigma x = 28$	$\Sigma y = 356$	$\Sigma xy = 1805$	$\Sigma x^2 = 140$

$\therefore$ From equations (2) and (3), we get

$$7a + 28b = 356$$

$$28a + 140b = 1805$$

On solving these equations, we get

$$a = -3.5714$$

$$b = 13.6071$$

$$y = -3.5714 + 13.6071x. \quad \textbf{Ans.}$$

Example 8: Fit a second-degree parabola to the following data taking x as the independent variable.

x	1	2	3	4	5	6	7	8	9
y	2	6	7	8	10	11	11	10	9

Sol. The equation of second-degree parabola is given by $y = a + bx + cx^2$ and the normal equations are:

$$\left.\begin{aligned} \sum y &= na + b\sum x + c\sum x^2 \\ \sum xy &= a\sum x + b\sum x^2 + c\sum x^3 \\ \sum x^2 y &= a\sum x^2 + b\sum x^3 + c\sum x^4 \end{aligned}\right\} \quad \ldots(1)$$

Here $n = 9$. The various sums are appearing in the table as follows:

x	y	xy	x^2	x^2y	x^3	x^4
1	2	2	1	2	1	1
2	6	12	4	24	8	16
3	7	21	9	63	27	81
4	8	32	16	128	64	256
5	10	50	25	250	125	625
6	11	66	36	396	216	1296
7	11	77	49	539	343	2401
8	10	80	64	640	512	4096
9	9	81	81	729	729	6561
$\sum x = 45$	$\sum y = 74$	$\sum xy = 421$	$\sum x^2 = 284$	$\sum x^2y = 2771$	$\sum x^3 = 2025$	$\sum x^4 = 15333$

Putting these values of $\sum x, \sum y, \sum x^2, \sum xy, \sum x^2y, \sum x^3$ and $\sum x^4$ in equation (1) and solving the equations for a, b and c, we get

$$a = -0.923\,;\ b = 3.520\,;\ c = -0.267\,.$$

Hence the fitted equation is

$$y = -0.923 + 3.53x - 0.267x^2. \textbf{ Ans.}$$

Change of Scale

If the data values are equispaced (with height (h)) and quite large for computation, simplification may be done by origin shifting as given below:

- When number of observations (n) is odd, take the origin at middle value of the table; say (x_0) and substitute $u = \frac{x-x_0}{h}$
- y values if small; may be left unchanged; or we can shift them at average value of y data $v = y-y_0$
- When number of observations (n) is even, take the origin as mean of two middle values, with new height $\frac{h}{2}$ and substitute $u = \frac{x-x_0}{h/2}$

Example6 Fit a 2^{nd} degree parabola for the following data:

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

Solution: Since number of observations is odd and $h = 1$,

taking $x_0 = 1933, y_0 = 357, u = x - 1933, v = y - 357$

The equation $y = ax^2 + bx + c$ is transformed to $v = Au^2 + Bu + C$ ...①

Normal equations are

$$\sum v = A\sum u^2 + B\sum u + 9c \qquad ...②$$

$$\sum uv = A\sum u^3 + B\sum u^2 + c\sum u \qquad ...③$$

$$\sum u^2 v = A\sum u^4 + B\sum u^3 + c\sum u^2 \qquad ...④$$

Calculating $\sum u, \sum u^2, \sum u^3, \sum u^4, \sum v, \sum uv$ and $\sum u^2 v$

x	u	y	v	uv	u^2	u^2v	u^3	u^4
1929	-4	352	-5	20	16	-80	-64	256
1930	-3	356	-1	3	9	-9	-27	81
1931	-2	357	0	0	4	0	-8	16
1932	-1	358	1	-1	1	1	-1	1
1933	0	360	3	0	0	0	0	0
1934	1	361	4	4	1	4	1	1
1935	2	361	4	8	4	16	8	16
1936	3	360	3	9	9	27	27	81
1937	4	359	2	8	16	32	64	256
	$\sum u =$ 0		$\sum v =$ 11	$\sum uv =$ 51	$\sum u^2 =$ 60	$\sum u^2 v =$ -9	$\sum u^3 =$ 0	$\sum u^4 =$ 708

Substituting $\sum u, \sum u^2, \sum u^3, \sum u^4, \sum v, \sum uv$ and $\sum u^2 v$ in ② and ③ and ④

$$\Rightarrow 11 = 60A + 9C \qquad \text{...⑤}$$

$$51 = 60B \qquad \text{...⑥}$$

$$-9 = 708A + 60C \qquad \text{...⑦}$$

Solving ⑤ ⑥ and ⑦ , we get $A = \frac{-247}{924}$ and $B = \frac{17}{20}$, $C = \frac{694}{231}$

Substituting in ①, parabola of best fit is $v = \frac{-247}{924} x^2 + \frac{17}{20} x + \frac{694}{231}$

$$\Rightarrow y - 357 = \frac{-247}{924} (x - 1933)^2 + \frac{17}{20} (x - 1933) + \frac{694}{231}$$

$$\Rightarrow y = -0.267x^2 + 1034.29x - 1000106.41$$

Example 7 The weight of a calf taken at end of every month is given below. Fit a straight line using the method of least squares. Also compute monthly growth rate.

x	1	2	3	4	5	6	7	8	9	10
y	52.5	58.7	65.0	70.2	75.4	81.1	87.2	95.5	102.2	108.4

Solution: Here $n = 10$ is even, $\therefore$ taking origin at $\frac{5+6}{2} = 5.5$ and new height as

$$\frac{h}{2} = 0.5 \;\therefore u = \frac{x-5.5}{0.5} \text{ and let } v = y$$

Let line of best fit $y = ax + b$ be transformed to $v = Au + B$ ...①

Normal equations are given by $\sum v = a\sum u + nb$ ...②

and $\sum uv = a\sum u^2 + b\sum u$ ...③

x	$u = \frac{x-5.5}{0.5}$	$v = y$	uv	u^2
1	-9	52.5	-472.5	81
2	-7	58.7	-410.9	49
3	-5	65.0	-325.0	25
4	-3	70.2	-210.6	9
5	-1	75.4	-75.4	1
6	1	81.1	81.1	1
7	3	87.2	261.6	9
8	5	95.5	477.5	25
9	7	102.2	715.4	49
10	9	108.4	975.6	81
	$\sum u = 0$	$\sum v = 796.2$	$\sum uv = 1016.8$	$\sum u^2 = 330$

Substituting values of $\sum u, \sum v, \sum uv$ and $\sum u^2$ in ② and ③

$\Rightarrow 796.2 = 10B$ and $1016.8 = 330a$

$\therefore A = 3.081$ and $B = 79.62$

Substituting in ①, line of best fit is $v = 3.081u + 79.62$

$$\Rightarrow y = 3.081\left(\frac{x-5.5}{0.5}\right) + 79.62 \therefore \text{ Line of best fit is } y = 6.162x + 45.729$$

Average growth rate per month is given by: $\frac{dy}{dx} = 6.162$

Example 9: Show that the line of fit to the following data is given by $y = 0.7x + 11.2$

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

Sol. Here $n = 6$ (even)

Let $x_0 = 12.5,\ h = 5,\ y_0 = 20$ (say)

Then, $u = \frac{x - 12.5}{2.5}$ and $v = y - 20$, we get

x	y	u	v	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
		$\sum u = 0$	$\sum v = 0$	$\sum uv = 122$	$\sum u^2 = 70$

The normal equations are,

$$0 = 6a + 0 \cdot b \quad \Rightarrow \quad a = 0$$

$$122 = 0 \cdot a + 70b \quad \Rightarrow \quad b = 1.743$$

Thus line of fit is $v = 1.743u$.

or $$y - 20 = (1.743)\left(\frac{x - 12.50}{2.5}\right) = 0.69x - 8.715$$

or $$y = 0.7x + 11.285. \quad \textbf{Ans.}$$

Example 10: Fit a second-degree parabola to the following data by least squares method.

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

Sol. Taking $x_0 = 1933$, $y_0 = 357$ then $u = \dfrac{(x - x_0)}{h}$

Here $h = 1$

Taking $u = x - x_0$ and $v = y - y_0$, therefore, $u = x - 1933$ and $v = y - 357$

x	$u = x - 1933$	y	$v = y - 357$	uv	u^2	u^2v	u^3	u^4
1929	–4	352	–5	20	16	–80	–64	256
1930	–3	356	–1	3	9	–9	–27	81
1931	–2	357	0	0	4	0	–8	16
1932	–1	358	1	–1	1	1	–1	1
1933	0	360	3	0	0	0	0	0
1934	1	361	4	4	1	4	1	1
1935	2	361	4	8	4	16	8	16
1936	3	360	3	9	9	27	27	81
1937	4	359	2	8	16	32	64	256
Total	$\sum u = 0$		$\sum v = 11$	$\sum uv = 51$	$\sum u^2 = 60$	$\sum u^2v = -9$	$\sum u^3 = 0$	$\sum u^4 = 708$

Then the equation $y = a + bx + cx^2$ is transformed to $v = A + Bu + Cu^2$ ...(1)

Normal equations are:

$$\sum v = 9A + B\sum u + C\sum u^2 \quad \Rightarrow 11 = 9A + 60C$$

$$\sum uv = A\sum u + B\sum u^2 + C\sum u^3 \quad \Rightarrow B = 17/20$$

$$\sum u^2v = A\sum u^2 + B\sum u^3 + C\sum u^4 \quad \Rightarrow -9 = 60A + 708C$$

On solving these equations, we get $A = \frac{694}{231} = 3$, $B = \frac{17}{20} = 0.85$ and $C = -\frac{247}{924} = -0.27$

$$\therefore \quad v = 3 + 0.85u - 0.27u^2$$

$$\Rightarrow \quad y - 357 = 3 + 0.85(x - 1933) - 0.27(x - 1933)^2$$

$$\Rightarrow \quad y = -1010135.08 + 1044.69x - 0.27x^2.$$ **Ans.**

Example 11: Fit second degree parabola to the following:

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

Sol. Here $n = 5$ (odd) therefore $x_0 = 2$, $h = 1$, $y_0 = 0$ (say)

Now let $u = x - 2$, $v = y$ and the curve of fit be $v = a + bu + cu^2$.

x	y	u	v	uv	u^2	u^2v	u^3	u^4
0	1	–2	1	–2	4	4	–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		0	12.9	11.3	10	33.5	0	34

Hence the normal equations are,

$$\sum v = 5a + b\sum u + c\sum u^2$$

$$\sum uv = a\sum u + b\sum u^2 + c\sum u^3$$

$$\sum u^2v = a\sum u^2 + b\sum u^3 + c\sum u^4$$

On putting the values of $\sum u, \sum v$ etc. from the table in these, we get

$$12.9 = 5a + 10c,\ 11.3 = 10b,\ 33.5 = 10a + 34c.$$

On solving these equations, we get

$$a = 1.48,\ b = 1.13 \text{ and } c = 0.55$$

Therefore the required equation is $v = 1.48 + 1.13u + 0.55u^2$.

Again substituting $u = x - 2$ and $v = y$, we get

$$y = 1.48 + 1.13(x-2) + 0.55(x-2)^2$$

$$y = 1.42 - 1.07x + 0.55x^2. \quad \textbf{Ans.}$$

Exercise

1. Fit a straight line $y = ax + b$ to the following data

x	0	1	3	6	8
y	1	3	2	5	4

2. Fit a straight line $y = a + bx$ to the following da

x	25	19	50	36	40	45	30
y	77	76	85	80	82	83	79

3. Fit a second degree parabola $y = ax^2 + bx + c$ to the following data

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

4. Fit a second degree parabola $y = a + bx + cx^2$ to the following data

x	2	4	6	8	10
y	3.07	12.85	31.47	57.38	91.29

Answers

1. $y = 0.38x + 1.6$
2. $y = 70.052 + 0.292x$
3. $y = 0.55x^2 - 1.07x + 1.42$
4. $y = 0.34 - 0.78x + 0.99x^2$

FITTING OF AN EXPONENTIAL CURVE

Suppose an exponential curve of the form

$$y = ae^{bx}$$

Taking logarithm on both the sides, we get

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

i.e., $$Y = A + Bx \quad \ldots(1)$$

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

The normal equations for (1) are,

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

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

On solving the above two equations, we get A and B

then $$a = \text{antilog } A,\ b = \frac{B}{\log_{10} e}$$

Example 12: 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

Sol. The curve to be fitted is $y = ae^{bx}$ or $Y = A + Bx$, where $Y = \log_{10} y$, $A = \log_{10} a$, and $B = b\log_{10} e$.

Therefore the normal equations are:

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

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

Substituting the values of $\sum x$, etc. and calculated by means of above table in the normal equations, we get

$$5.7536 = 5A + 34B$$

and
$$40.8862 = 34A + 300B$$

On solving these equations, we obtain,

$$A = 0.9766\,;\ B = 0.02561$$

Therefore $\quad a = \text{antilog}_{10} A = 9.4754\ ;\ b = \dfrac{B}{\log_{10} e} = 0.059$

Hence the required curve is $y = 9.4754e^{0.059x}$. **Ans.**

Example 13: For the data given below, find the equation to the best fitting exponential curve of the form $y = ae^{bx}$.

x	1	2	3	4	5	6
y	1.6	4.5	13.8	40.2	125	300

Sol. $y = ae^{bx}$

On taking log both the sides, $\log y = \log a + bx \log e$ which is of the form $Y = A + Bx$, where $Y = \log y$, $A = \log a$ and $B = b \log e$.

x	y	$Y = \log y$	x^2	xY
1	1.6	0.2041	1	0.2041
2	4.5	0.6532	4	1.3064
3	13.8	1.1399	9	3.4197
4	40.2	1.6042	16	6.4168
5	125	2.0969	25	10.4845
6	300	2.4771	36	14.8626
$\sum x = 21$		$\sum Y = 8.1754$	$\sum x^2 = 91$	$\sum xY = 36.6941$

Normal equations are: $\sum Y = 6A + B\sum x$

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

Therefore from these equations, we have

$$8.1754 = 6A + 21B$$

$$36.6941 = 21A + 91B$$

$\Rightarrow$ $$A = -0.2534, B = 0.4617$$

Therefore, $a = \text{antilog} A = \text{antilog}(-0.2534) = \text{antilog}(\bar{1}.7466) = 0.5580.$

and $$b = \frac{B}{\log e} = \frac{0.4617}{0.4343} = 1.0631$$

Hence required equation is $y = 0.5580e^{1.0631x}$. **Ans.**

Example 16: Fit an exponential curve of the form $y = ab^x$ to the following data:

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

Sol. $y = ab^x$ takes the form $Y = A + Bx$, where $Y = \log y$; $A = \log a$ and $B = \log b$. Hence the normal equations are given by

$$\sum Y = nA + B\sum x \text{ and } \sum xY = A\sum x + \sum x^2 .$$

x	y	$Y = \log y$	xY	x^2
1	1.0	0.0000	0.000	1
2	1.2	0.0792	0.1584	4
3	1.8	0.2553	0.7659	9
4	2.5	0.3979	1.5916	16
5	3.6	0.5563	2.7815	25
6	4.7	0.6721	4.0326	36
7	6.6	0.8195	5.7365	49
8	9.1	0.9590	7.6720	64
$\sum x = 36$	$\sum y = 30.5$	$\sum Y = 3.7393$	$\sum xY = 22.7385$	$\sum x^2 = 204$

Putting the values in the normal equations, we obtain

$$3.7393 = 8A + 36B \text{ and } 22.7385 = 36A + 204B$$

$$\Rightarrow \qquad B = 0.1407 \text{ and } A = 0.1656$$

$$\Rightarrow \qquad b = \text{antilog} B = 1.38 \text{ and } a = \text{antilog } A = 0.68.$$

Thus the required curve of best fit is $y = (0.68)(1.38)^x$. **Ans.**

Example 17: Fit a curve $y = ab^x$ to the following data:

x	2	3	4	5	6
y	144	172.8	207.4	248.8	298.5

Sol. Given equation $y = ab^x$ reduces to $Y = A + Bx$ where $Y = \log y$, $A = \log a$ and $B = \log b$.

The normal equations are,

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

$$\sum x \log y = \log a \sum x + \log b \sum x^2$$

The calculations of $\sum x$, $\sum \log y$, $\sum x^2$ and $\sum x \log y$ are substitute in the following tabular form.

x	y	x^2	$\log y$	$x \log y$
2	144	4	2.1584	4.3168
3	172.8	9	2.2375	6.7125
4	207.4	16	2.3168	9.2672
5	248.8	25	2.3959	11.9795
6	298.5	36	2.4749	14.8494
20		90	11.5835	47.1254

Putting these values in the normal equations, we have

$$11.5835 = 5 \log a + 20 \log b$$

$$47.1254 = 20 \log a + 90 \log b.$$

Solving these equations and taking antilog, we have a = 100, b = 1.2 approximate. Therefore equation of the curve is $y = 100(1.2)^x$. **Ans.**

Example 20: For the data given below, find the equation to the best fitting exponential curve of the form $y = ae^{bx}$.

x	1	2	3	4	5	6
y	1.6	4.5	13.8	40.2	125	300

Sol. Given $y = ae^{bx}$, taking log we get $\log y = \log a + bx\log_{10} e$ which is of the $Y = A + Bx$,

where $$Y = \log y,\ A = \log a \text{ and } B = \log_{10} e.$$

Put the values in the following tabular form, also transfer the origin of x series to 3, so that $u = x - 3$.

x	y	$\log y = Y$	$x - 3 = u$	uY	u^2
1	1.6	0.204	−2	−0.408	4
2	4.5	0.653	−1	−0.653	1
3	13.8	1.140	0	0	0
4	40.2	1.604	1	1.604	1
5	125.0	2.094	2	4.194	4
6	300	2.477	3	7.431	9
Total		8.175	3	12.168	19

In case $Y = A + B.u$, then normal equations are given by

$$\sum Y = nA + B\sum u \quad \Rightarrow \quad 8.175 = 6A + 3B \qquad ...(1)$$

$$\sum uY = A\sum u + B\sum u^2 \quad \Rightarrow \quad 12.168 = 3A + 19B \qquad ...(2)$$

Solving (1) and (2), we get

$$A = 1.13 \text{ and } B = 0.46$$

Thus equation is $Y = 1.13 + 0.46u$, *i.e.* $Y = 1.13 + 0.46(x-3)$

$$Y = 0.46x - 0.25$$

Which gives $\log a = -0.25$ *i.e.* $\text{antilog}(-0.25) = \text{antilog}(\bar{1}.75) = 0.557$

$$b = \frac{B}{\log_{10} e} = \frac{.46}{0.4343} = 1.06$$

Hence, the required equation of the curve is $y = (0.557)e^{1.06x}$. **Ans.**

PROBLEM SET 5.1

1. Fit a straight line to the given data regarding x as the independent variable:

x	1	2	3	4	6	8
y	2.4	3.1	3.5	4.2	5.0	6.0

[**Ans.** $y = 2.0253 + 0.502x$]

2. Fit a straight line $y = a + bx$ to the following data by the method of least square:

x	0	1	3	6	8
y	1	3	2	5	4

[**Ans.** $1.6 + 0.38x$]

3. Find the least square approximation of the form $y = a + bx^2$ for the following data:

x	0	0.1	0.2	0.3	0.4	0.5
y	1	1.01	0.99	0.85	0.81	0.75

[**Ans.** $y = 1.0032 - 1.1081x^2$]

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

x	0.0	1.0	2.0
y	1.0	6.0	17.0

[**Ans.** $y = 1 + 2x + 3x^2$]

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

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

[**Ans.** $y = 1.04 - 0.193x + 0.243x^2$]

6. Fit a second degree parabola to the following data by the least square method:

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

[**Ans.** $y = 27.5x^2 + 40.5x + 1024$]

7. Fit a parabola $y = a + bx + cx^2$ to the following data:

x	2	4	6	8	10
y	3.07	12.85	31.47	57.38	91.29

[**Ans.** $y = 0.34 - 0.78x + 0.99x^2$]

8. 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

[**Ans.** $y = 1.49989e^{0.50001x}$]

9. Fit a least square geometric curve $y = ax^b$ to the following data:

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

[**Ans.** $y = 0.5012x^{1.9977}$]