UNIT-5 » CURVE FITTING BY NUMERICAL METHOD

❖ INTRODUCTION

- ✓ In particular statistics, we come across many situations where we often require to find a relationship between two or more variables. For example, weight and height of a person, demand and supply, expenditure depends on income, etc. These relation, in general, may be expresses by polynomial or they may have exponential or logarithmic relationship. In order to determine such relationship, first it is requiring to collect the data showing corresponding values of the variables under consideration.
- ✓ Suppose $(x_1, y_1), (x_2, y_2), \dots \dots, (x_n, y_n)$ be the data showing corresponding values of the variables x and y under consideration. If we plot the above data points on a rectangular coordinate system, then the set of points so plotted form a scatter diagram.
- ✓ From this diagram, it is sometimes possible to visualize a smooth curve approximating the data. Such a curve is called an approximating curve.
- ✓ In particular, if the data approximate well to a straight line, we say that a linear relationship exists between the variables. It is quite possible that the relationship of the form $y = f(x)$ between two variables x and y, giving the approximating curve and which fit the given data of x and y, is called curve fitting.

❖ CURVE FITTING

- ✓ Curve fitting is the process of finding the 'best-fit' curve for a given set of data. It is the representation of the relationship between two variables by means of an algebraic equation.

❖ THE METHOD OF LEAST SQUARE

- ✓ The method of least squares assumes that the best-fit curve of a given type is the curve that has the minimum sum of the square of the deviation (least square error) from a given set of data.
- ✓ Suppose that the data points are $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, where x is independent and y is dependent variable. Let the fitting curve f(x) has the following deviations (or errors or residuals) from each data points

$$d_1 = y_1 - f(x_1), d_2 = y_2 - f(x_2), \dots, d_n = y_n - f(x_n)$$

- Clearly, some of the deviations 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,

$$D = {d_1}^2 + {d_2}^2 + \cdots + {d_n}^2$$

- Now, according to the method of least squares, the best fitting curve has the property that

$$D = {d_1}^2 + {d_2}^2 + \cdots + {d_n}^2 = \sum_{i=1}^{n} {d_i}^2 = \sum_{i=1}^{n} [y_i - f(x_i)]^2 = \text{a minimum.}$$

❖ FITTING A STRAIGHT LINE $y = a + bx$ (LINEAR APPROXIMATION)

- Suppose the equation of a straight line of the form $y = a + bx$ is to be fitted to the n-data points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), n \geq 2$, Where a is y-intercept and b is its slope.
- For the general point (x_i, y_i), the vertical distance of this point from the line $y = a + bx$ is the deviation d_i, then $d_i = y_i - f(x_i) = y_i - a - bx_i$.
- Applying method of least squares, the values of a and b are so determined as to minimize

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

- This will be minimum,

$$\frac{\partial D}{\partial a} = 0 \Longrightarrow -2\sum_{i=1}^{n} (y_i - a - bx_i) = 0 \;\text{ and }\; \frac{\partial D}{\partial b} = 0 \Longrightarrow -2\sum_{i=1}^{n} x_i(y_i - a - bx_i) = 0$$

- Simplifying and expanding the above equations, we have

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} 1 + b\sum_{i=1}^{n} x_i \;\text{ and }\; \sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} {x_i}^2$$

- Which implies

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i \;\text{ and }\; \sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} {x_i}^2$$

- We obtain following normal equations for the best fitting straight line $y = a + bx$.

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

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

✓ The normal equations for the best fitting straight line $y = ax + b$ is

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

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

METHOD – 1: FITTING A STRAIGHT LINE

C	1	By the method of least square, find the straight line that best fits the following data.	

x	1	2	3	4	5
y	14	27	40	55	68

Answer: $\mathbf{y = 13.6x}$

C	2	Fit a straight line for following data. Also, find y when x = 2.8.	

x	2	5	6	9	11
y	2	4	6	9	10

Answer: $\mathbf{y = -0.0244 + 0.9431x.\ y(2.8) = 2.6163}$

H	3	Fit a straight line $y = a + bx$ to the following data.	

x	−2	−1	0	1	2
y	1	2	3	3	4

Answer: $\mathbf{y = 2.7907 + 0.4767x}$

C	4	Fit a straight line $y = ax + b$ for following data.	

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

Answer: $\mathbf{y = -0.5x + 8}$

H	5	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	68

Answer: $\mathbf{y = 46.9394 + 0.3232x}$

H 6 Fit a straight line to the following data regarding x as a dependent variable.

x	2.4	3.1	3.5	4.2	5	6
y	1	2	3	4	6	8

Answer: $\mathbf{y = -3.9703 + 1.9761x}$

H 7 The weight of a calf taken at weekly intervals are given below. Fit a straight line using method of least squares.

Age (x)	1	2	3	4	5	6	7	8	9	10
Weight (y)	52.5	58.7	65	70.2	75.4	81.1	87.2	95.5	102.5	108.4

Answer: $\mathbf{y = 46.6094 + 6.1554x}$

H 8 If P is the pull required to lift a load W by means of a pulley block, find a linear approximation of the form $P = mW + c$ connecting P and W, using the following data.

P	13	18	23	27
W	51	75	102	119

Answer: $\mathbf{P = 0.2028W + 2.6580}$

T 9 By method of least squares, fit a linear relation of the form $P = a + bW$ to the following data, P is the pull required to lift a weight W. Also estimate P, when W is 150.

P	50	70	100	120
W	12	15	21	25

Answer: $\mathbf{P = 2.2759 + 0.1879W.\ P(150) = 30.4655}$

T 10 Fit a straight line for the given pairs of (x, y) which are (1, 5), (2, 7), (3, 9), (4, 10), (5, 11).

Answer: $\mathbf{y = 3.9 + 1.5x}$

❖ FITTING A PARABOLA

✓ Consider a set of n pairs of the given values (x, y) for fitting the curve $y = a + bx + cx^2$. The residual $R = y - (y = a + bx + cx^2)$ is the difference between the observed and estimated values of y. We have to find a, b, c such that the sum of the squares of the residuals is minimum. Let

$$S = \sum_{1}^{n} [y - (a + bx + cx^2)]^2 \quad \dots\dots \ (1)$$

- ✓ Differentiating S with respect to a, b, c and equating zero. We obtain following normal equations for the best fitting $y = a + bx + cx^2$ curve (parabola) of second degree.

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

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

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

- ✓ The normal equation for $y = ax^2 + bx + c$ are

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

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

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

METHOD – 2: FITTING A PARABOLA

C 1 Fit a second degree polynomial of y on x to the following data.

x	50	70	100	120
y	12	15	21	25

Answer: $\mathbf{y = 5.5259 + 0.1029x + 0.0005x^2}$

H 2 Fit a polynomial of degree two using least square method for the following experimental data. Also, estimate y(2.4).

x	1	2	3	4	5
y	5	12	26	60	97

Answer: $\mathbf{y = 10.4 - 11.0857x + 5.7143x^2.\ y(2.4) = 16.7086}$

C	3	Fit a relation of the form $R = a + b\,V + c\,V^2$ to the following data, where V is the velocity in km/hr. and R is the resistance in km/quintal. Estimate R when $V = 90$. V: 20, 40, 60, 80, 100, 120 R: 5.5, 9.1, 14.9, 22.8, 33.3, 46.0 **Answer**: $\mathbf{R = -4.35 + 0.0024V + 0.0029V^2. R(90) = 27.8183}$	
H	4	Fit a second degree parabola $y = ax^2 + bx + c$ to the following data. x: −1, 0, 1, 2, 3 y: 5, 6, 21, 50, 93 **Answer**: $\mathbf{y = 7x^2 + 8x + 6}$	
C	5	Fit a second degree parabola $y = ax^2 + bx + c$ to the following data. x: −3, −2, −1, 0, 1, 2, 3 y: 12, 4, 1, 2, 7, 15, 30 **Answer**: $\mathbf{y = 0.5453x^2 + 3.6714x + 7.7954}$	
T	6	Fit a parabola to the following observations. x: 1, 2, 3, 4, 5, 6 y: 3.13, 3.76, 6.94, 12.62, 20.86, 31.53 **Answer**: $\mathbf{y = 4.982 - 3.1199x + 1.2579x^2}$	
H	7	Fit a second degree parabola $y = a + bx + cx^2$ 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 **Answer**: $\mathbf{y = 1.0357 - 0.1929x + 0.2429\,x^2}$	
T	8	For 10 randomly selected observations, the following data were recorded. Observation Number: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 Overtime Hours (x): 1, 1, 2, 2, 3, 3, 4, 5, 6, 7 Additional units (y): 2, 7, 7, 10, 8, 12, 10, 14, 11, 14 Determine the coefficient of regression using the non-linear form $y = a + b_1x + b_2x^2$. **Answer**: $\mathbf{y = 1.8022 + 3.4823x - 0.2690x^2}$	

H	9	The following are the data on the drying time of a certain varnish and the amount of an additive that is intended to reduce the drying time?	W 2019 (7)

Amount of varnish additive(grams) "x"	0	1	2	3	4	5	6	7	8
Drying time(hr.) "y"	12	10.5	10	8	7	8	7.5	8.5	9

I. Fit a second degree polynomial by the method of least square.

II. Use the result to predict the drying time of the varnish when 6.5 gms of the additive is being used.

Answer: $\mathbf{y = 12.1848 - 1.8465x + 0.1829x^2. y(6.5) = 7.9099}$

❖ FITTING THE GENERAL CURVES

✓ $\mathbf{y = ae^{bx}}$ yields

- Taking Logarithm on both sides $\log y = \log a + bx$.
- Denoting $\log y = Y$ and $\log a = A$, then above equation becomes $Y = A + bx$.
- Find A, b & consequently a=Antilog A can be calculated.

✓ $\mathbf{y = ax^b}$ yields

- Taking Logarithm on both sides $\log y = \log a + b\log x$.
- Denoting $\log y = Y$, $\log a = A$ and $\log x = X$, we obtain $Y = A + bX$.
- Find A, b & consequently a=Antilog A can be calculated.

✓ $\mathbf{y = ab^x}$ yields

- Taking Logarithm on both sides $\log y = \log a + x\log b$.
- Denoting $\log y = Y$, $\log a = A$, $\log b = B$, we obtain $Y = A + Bx$.
- Find A, B & consequently $a = \text{Antilog } A$ and $b = \text{Antilog } B$ can be calculated.

✓ $\mathbf{y = a + bx^2}$

- Take the auxiliary equations

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

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

- Find the value of $\sum x^2, \sum x^2y, \sum x^4$.
- Find a, b.

✓ $\mathbf{y = ax^2 + \frac{b}{x}}$

✓ Take the auxiliary equations

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

$$\sum \frac{y}{x} = a\sum x + b\sum \frac{1}{x^2}$$

- Find the value of $\sum x, \sum x^2y, \sum x^4, \sum \frac{y}{x}, \sum \frac{1}{x^2}$.
- Find a, b.

✓ $\mathbf{pv^{\gamma} = C}$

- $v = \left(\frac{C}{p}\right)^{\frac{1}{\gamma}} \Rightarrow v = C^{\frac{1}{\gamma}} p^{-\frac{1}{\gamma}}$
- Take logarithm both the sides $\log v = \frac{1}{\gamma}\log C - \frac{1}{\gamma}\log p$.
- Denoting $\log v = Y$, $\frac{1}{\gamma}\log C = A, -\frac{1}{\gamma} = B$.
- Find A, B & consequently $\gamma = -\frac{1}{B}$ and $C = \text{Antilog}(\gamma A)$ can be calculated.

METHOD – 3: FITTING THE GENERAL CURVES

C	1	Fit a curve of the best fit of the type $y = ae^{bx}$ to the following data. <table><tr><td>x</td><td>1</td><td>5</td><td>7</td><td>9</td><td>12</td></tr><tr><td>y</td><td>10</td><td>15</td><td>12</td><td>15</td><td>21</td></tr></table>**Answer**: $\mathbf{y = 9.4751e^{0.0590x}}$	

H 2

Fit a curve of the best fit of the type $y = ae^{bx}$ to the following data.

x	1	2	3	4
y	1.65	2.7	4.5	7.35

Answer: $\mathbf{y = 1.0001e^{0.4993x}}$

C 3

Find the least square fit of the form $y = a_0 + a_1x^2$ to the following data.

x	−1	0	1	2
y	2	5	3	0

Answer: $\mathbf{y = 4.1667 - 1.1111x^2}$

H 4

Fit a curve of the best fit of the type $y = ax^b$ to the following data.

x	2	3	4	5
y	27.8	62.1	110	161

Answer: $\mathbf{y = 7.3799x^{1.9302}}$

C 5

Fit a curve of the best fit of the type $y = ax^b$ to the following data.

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

Answer: $\mathbf{y = 0.5\,x^2}$

C 6

Fit a curve of the best fit of the type $y = ab^x$ to the following data.

x	2	3	4	5	6
y	8.3	15.4	33.1	65.2	126.4

Answer: $\mathbf{y = 2.0494\,(1.9916)^x}$

H 7

Fit a curve of the best fit of the type $y = ab^x$ to the following data.

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

Answer: $\mathbf{y = 100.0262\,(1.1999)^x}$

H 8

Fit a curve of the best fit of the type $y = ab^x$ to the following data.

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

Answer: $\mathbf{y = 0.6823\,(1.3828)^x}$

T	9	Using least square method fit the curve $y = ax^2 + \frac{b}{x}$ to the following data. x: 1, 2, 3, 4 y: −1.51, 0.99, 8.88, 7.66 **Answer**: $\mathbf{y = 0.509x^2 - \frac{2.04}{x}}$	
T	10	The pressure P of the gas corresponding to various volume V is measured given by the following data, fit the data to the equation $PV^{\gamma} = C$. P: 50, 60, 70, 80, 90 V: 64.7, 51.3, 40.5, 25.9, 78 **Answer**: $\mathbf{PV^{3.0939} = 11340638.62}$	

Data for T 9:

x	1	2	3	4
y	−1.51	0.99	8.88	7.66

Data for T 10:

P	50	60	70	80	90
V	64.7	51.3	40.5	25.9	78

★ ★ ★ ★ ★ ★ ★ ★ ★ ★