CURVE FITTING (NON-LINEAR REGRESSION)

Derivation of Non-Linear Regression Model of Y on X:

(Fitting of Second degree equation)

Suppose (x_i, y_i); i= 1,2,…,n., are n pairs of observations on variables X, Y. We assume that Y as dependent variable, which can be expressed in terms of X. The general second degree curve will be $y = a + bx + cx^2$. We find the constants a, b and c by using least square principle.

We assume the model $y = a + bx + cx^2 + e$ ... (1)

$$e = \left(y - a - bx - cx^2\right)^2$$

By using principal of least square

Symbolically we write $S = \sum_{i=1}^{n} e_i^2$ as sum of squares of errors. We find the points minima using calculus methods.

The solution of equation $\frac{\partial s}{\partial a} = 0$, $\frac{\partial s}{\partial b} = 0$ and $\frac{\partial s}{\partial c} = 0$ gives, extreme points.

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

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

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

$$0 = -2\sum\left(y_i - a - bx_i - cx_i^2\right)$$

$$0 = \sum y_i - \sum a - b\sum x_i - c\sum x_i^2$$

$$\sum y_i = \sum a + b\sum x_i + \mathrm{c}\sum \mathrm{x}_i^2$$

$$\sum y_i = na + b\sum x_i + \mathrm{c}\sum \mathrm{x}_i^2 \qquad \cdots(2)$$

Similarly, $\frac{\partial s}{\partial b} = \frac{\partial}{\partial b}\sum\left(\mathrm{y_i – a – bx_i - cx_i^2}\right)^2 = 0$ gives

$$\frac{\partial s}{\partial b} = \sum_{i=1}^{n} \frac{\partial}{\partial b}(\mathrm{y_i\text{-}a\text{-}bx_i\text{-}cx_i^2})^2$$

$$0 = 2\sum\left(\mathrm{y_i\text{-}a\text{-}bx_i\text{-}cx_i^2}\right)\left(\mathrm{\text{-}x_i}\right)$$

$$0 = \sum x_i y_i - \sum x_i a - b\sum x_i^2 - \mathrm{c}\sum \mathrm{x}_i^3$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + \mathrm{c}\sum \mathrm{x}_i^3 \qquad \cdots(3)$$

Similarly, $\frac{\partial s}{\partial c} = \frac{\partial}{\partial c}\sum\left(\mathrm{y_i – a – bx_i - cx_i^2}\right)^2 = 0$ gives

$$\frac{\partial s}{\partial c} = \sum_{i=1}^{n} \frac{\partial}{\partial c}(\mathrm{y_i\text{-}a\text{-}bx_i\text{-}cx_i^2})^2$$

$$0 = 2\sum\left(\mathrm{y_i\text{-}a\text{-}bx_i\text{-}cx_i^2}\right)\left(\mathrm{\text{-}x_i^2}\right)$$

$$0 = \sum \mathrm{x}_i^2\, y_i - \sum \mathrm{x}_i^2\, a - b\sum x_i^3 - \mathrm{c}\sum \mathrm{x}_i^4$$

$$\sum \mathrm{x}_i^2\, y_i = a\sum \mathrm{x}_i^2 + b\sum x_i^3 + \mathrm{c}\sum \mathrm{x}_i^4 \qquad \cdots(4)$$

Equations (2), (3) and (4) are called as normal equations. Solving these equations simultaneously, we get a, b and c . after putting the values of a, b and c in the second degree curve , we get predicted y.

Second Method:

u = x-a ; x = a + u

or

$$\mathrm{u} = \frac{(\mathrm{x\text{-}a})}{\mathrm{h}}; \quad \mathrm{x = a + hu}$$

$y = a + bx + cx^2$ becomes $y = a' + b'u + c'u^2$

$\sum y_i = na + b\sum x_i + c\sum x_i^2$ becomes $\sum y_i = na' + b'\sum u_i + c'\sum u_i^2$

$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3$ becomes $\sum u_i y_i = a'\sum u_i + b'\sum u_i^2 + c'\sum u_i^3$

$\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4$ becomes $\sum u_i^2 y_i = a'\sum u_i^2 + b'\sum u_i^3 + c'\sum u_i^4$

$\sum u_i = 0$ and $\sum u_i^3 = 0$

$$a' = \frac{\left(\sum y_i - c'\sum u_i^2\right)}{n};\ b' = \frac{\left(\sum u_i y_i\right)}{\sum u_i^2} \quad \text{and} \quad c' = \frac{\left(n\sum u_i^2 y_i - \sum u_i^2 \times \sum u_i^4\right)}{n\sum u_i^4 - \left(\sum u_i^2\right)^2}$$

[i] Fitting of an Exponential Curve $y = ab^x$

In some situation growth of y is at larger rate with respect to x for exa. Suppose y; population and x ;year in this case instead of second degree ,an exponentian curve fits well The nature of curve is given by following figure,

Suppose $\{(x_i, y_i)\}$, i=1,2….n $\}$is a sample of n pairs on (x, y) we can write $y = ab^x$ in linear form (if a > 0, b > 0) by taking log on both side

$y = ab^x$

$\text{Log } y = \text{Log}\left(ab^x\right)$

$\text{Log } y = \text{Log } a + \text{Log } b^x$

$\text{Log } y = \text{Log } a + x \text{ Log } b$

Let, v = log y, A' = log a and B'=log b then we get,

$V = A' + B'x$

$V = A + BU$

Normal equations are

$\sum V = nA + B\sum U$ and $\sum UV = A\sum U + B\sum U^2$

Solving these normal equations for A and B , we get

a = antilog A and b = antilog B . With this value of A and B, curve is the best fit to the given set of n points.

[ii] Fitting of an Exponential Curve $y = ax^b$

$$y = ax^b$$

$$\text{Log } y = \text{Log}\left(ax^b\right)$$

$$\text{Log } y = \text{Log } a + \text{Log } x^b$$

$$\text{Log } y = \text{Log } a + b \text{ Log } x$$

Let, v = log y, A = log a and U =log x then we get,

$$V = A + bU$$

Normal equations are

$$\sum V = nA + b\sum U \quad \text{and} \quad \sum UV = A\sum U + b\sum U^2$$

Solving these normal equations for A and b.

Numerical Examples:

Example 1: The profit in lakhs of Rs. earned by company in x^{th} year is tabulated below. Fit a second degree curve $y = a + bx + cx^2$. Also, estimate profit in 7^{th} year.

Year (x)	**1**	**2**	**3**	**4**	**5**
Profit (y)	**24**	**27**	**32**	**38**	**45**

Solution: Take 3 as origin

X	**U**	**y**	**U^2**	**U^3**	**U^4**	**Uy**	**U^2 y**
1	-2	24	4	-8	16	-48	96
2	-1	27	1	-1	1	-27	27
3	0	32	0	0	0	0	0
4	1	38	1	1	1	38	38
5	2	45	4	8	16	90	180
Total	0	166	10	0	34	53	341

The equation $y = a + bx + cx^2$ has following normal equations.

$$\sum y_i = na + b\sum x_i + c\sum x_i^2$$
$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3$$
$$\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4$$

Example 2: The population of a state is given below. Fit the curve $y = ab^x$

Year	1951	1961	1971	1981	1991
Population	140	170	200	250	300

Solution:- We know that

$$y = ab^x$$
$$\text{Log } y = \text{Log}\left(ab^x\right)$$
$$\text{Log } y = \text{Log } a + \text{Log } b^x$$
$$\text{Log } y = \text{Log } a + x \text{ Log } b$$
$$V = A' + B'x$$
$$V = A + BU$$
$$U = \frac{(x-\bar{x})}{10} \quad \text{and} \quad \bar{x} = \frac{\sum x}{n} = \frac{9855}{5} = 1971$$

Year	$U = \frac{(x-1971)}{10}$	**Y**	**V= log y**	**U^2**	**UV**
1951	**-2**	**140**	**2.1461**	**4**	**-4.2922**
1961	**-1**	**170**	**2.2304**	**1**	**-2.2304**
1971	**0**	**200**	**2.3010**	**0**	**0**
1981	**1**	**250**	**2.3979**	**1**	**2.3979**
1991	**2**	**300**	**2.4771**	**4**	**4.9542**
9855	**0**		**11.5525**	**10**	**0.8295**

As usual the normal equation obtained using least square principle will be

$\sum V = nA + B\sum U \quad \Rightarrow 11.5525 = 5A + 0B$

$\Rightarrow 11.5525 = 5A \quad \Rightarrow 2.3105 = A$

$\sum UV = A\sum U + B\sum U^2 \quad \Rightarrow 0.8295 = 0A + 10B$

$\Rightarrow 0.8295 = 10B \quad \Rightarrow 0.08295 = B$

$V = 2.3105 + 0.08295\ U$

$a = \text{antilog } A \quad \Rightarrow a = \text{antilog } 2.3105 \quad \Rightarrow a = 204.41$

$b = \text{antilog } B \quad \Rightarrow b = \text{antilog } 0.08295 \quad \Rightarrow b = 1.2105$

$y = ab^x$

$y = 204.41 \times 1.2105^U$

Example 3: Fit $y = ax^b$ to the following data. Also, find R^2 using MS-Excel commands

X	2	3	4	5	6
Y	4	23	50	78	175

Solution:-

X	Y	$U = \log_e x$	$V = \log_e y$	U^2	UV
2	4	0.6931	3.1355	0.4805	2.1734
3	23	1.0986	3.9120	1.2069	4.2978
4	50	1.3863	4.3567	1.9218	6.0397
5	78	1.6094	4.8122	2.5903	7.7449
6	175	1.7918	5.1648	3.2104	9.2541
20	330	6.5793	21.3812	9.4099	29.5098

The normal equation will be

$5A + 6.5793b = 21.3812 \quad \ldots(1)$

$6.5793A + 9.4099b = 29.5098 \quad \ldots(2)$

Solving (1) and (2), we get

b=1.8275, A=1.8716

a=Antilog A $\Rightarrow a = 6.4984$

$\therefore$ The equation is $y = 6.4984x^{1.8275}$

$y = 6.4984x^{1.8275}$

[A] THEORY QUESTIONS:

[1] Explain the procedure of fitting second degree curve $Y = a + bX + cX^2$ for bivariate data.

[2] Explain the procedure of fitting the curve $Y = ab^x$ for a bivariate data.

[3] Explain the procedure of fitting the curve $Y = ax^b$; $a > 0, b > 0$ for a bivariate data.

[B] Numerical Problems:

[1] Fit a curve of the type $Y = b^x$ for the following data using least squares principal:

X	1	2	3	4
Y	3.5	12	45	150.5

Answer: $Y = (3.5107)^x$

[2] Fit a curve of the type $Y = ab^x$ for the following data using least squares principal:

X	2	3	4	5	6
Y	144	172.8	207.4	248.8	298.5

Answer: $Y = 100.026(1.42)^x$

[3] Fit a curve of the type $Y = ax^b$ for the following data using least squares principal:

X	2	3	4	5	6
Y	144	172.8	207.4	248.8	298.5

Answer: $Y = 87.4849(x)^{0.6558}$

[4] Fit a curve of the type $Y = a + bX + cX^2$ for the following data using least squares principal:

Year	2015	2016	2017	2018	2019
Index of Jute export prices	185	169	191	203	275

Answer: $Y = 48195126.59 - 47810.1414X + 11.8571X^2$