ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

ELIMINATION OF PARAMETERS AND PRINCIPLE OF LEAST SQUARES: FITTING OF LINEAR CURVE TO AVERAGE MINIMUM TEMPERATURE DATA IN THE CONTEXT OF ASSAM

Atwar Rahman*, Dhritikesh Chakrabarty
* Department of Statistics, Pub Kamrup College, Baihata Chariali, Kamrup, Assam, India
Department of Statistics, Handique Girls' College, Guwahati-1, Assam, India

ABSTRACT

The principle of least squares, innovated by the French mathematician Legendre, when applied to observed data in order to fit a mathematical curve yields normal equations. The parameters involved in the curve are estimated by solving the normal equations. The number of normal equations becomes larger when the number of parameters associated to the curve becomes larger. In this situation, the solution of the normal equations for estimating the parameters becomes more complicated. For this reason, one more convenient method has been search for computing the estimates of the parameters. The method has been developed by the stepwise application of the principle of least squares. The method innovated here consists of the elimination of parameters first and then the minimization of the sum of squares of the errors. In this paper, the method has been described with reference to the estimation of parameters of a linear curve based on observed data on monthly average minimum temperature at Guwahati.

KEYWORDS: Linear curve, Least squares principle, Stepwise application, Monthly average temperature..

INTRODUCTION

The method of least squares, which is indispensible and is widely used method of curve fitting to numerical data, was first discovered by the French mathematician Legendre in 1805 [Mansfield (1877), Paris (1805)]. The first proof of this method was given by the renowned statistician Adrian (1808) followed by its second proof given by the German Astronomer Gauss [Hamburg (1809)]. Apart from this two proofs as many as eleven proofs were developed at different times by a number of mathematicians viz. Laplace (1810), Ivory (1825), Hagen (1837), Bassel (1838), Donkim (1844), John Herscel (1850), Crofton(1870) etc.. Though none of the thirteen proofs is perfectly satisfactory but yet it has given new dimension in setting the subject in a new light. In the method of least squares, the parameters of a curve are estimated by solving the normal equations of the curve obtained by the principle of least squares. However, for a curve of higher degree polynomial, the estimation of parameters by solving the normal equations carries a complicated calculation as the number of normal equations becomes large which leads to think of searching for some simpler method of estimation of parameter. In this study, an attempt has been made to discuss the method in the case of fitting of linear curve to observed data when the values of the independent variable are unequal intervals.

ESTIMATION OF PARAMETERS

Let the theoretical relationship between the dependent variable Y and the independent variable X be

$Y = aX + b$ (2.1)

where 'a' and 'b' are the two parameters.

Let Y_1, Y_2 ……, Y_n be n observations on Y corresponding to the observations X_1, X_2, ………, X_n of X.

The objective here is to fit the curve given by (2.1) to the observed data on X and Y.

Since the n pairs of observations

(X_1 , Y_1) , (X_2 , Y_2) , ……… (X_n, Y_n)

may not lie on the curve (2.1), they satisfy the model

IN LINEAR CURVE:

$y_i = ax_i + b + \xi_i$, $(i = 1 , 2 , …, n)$ (2.2)
n

ESTIMATION OF PARAMETER: *BY STEPWISE APPLICATION OF PRINCIPLES OFLEAST SQUARE AND BY SOLVING NORMAL EQUATION.*

From (2.2), one can obtain the following n sets of $(n – 1)$ equations in each set:

Set -1:

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

$$\left(\frac{y_1 - y_2}{x_1 - x_2}\right) = a + \left(\frac{\xi_1 - \xi_2}{x_1 - x_2}\right)$$
$$\left(\frac{y_1 - y_3}{x_1 - x_3}\right) = a + \left(\frac{\xi_1 - \xi_3}{x_1 - x_3}\right)$$
$$\left(\frac{y_1 - y_4}{x_1 - x_4}\right) = a + \left(\frac{\xi_1 - \xi_4}{x_1 - x_4}\right)$$

-- ----------------------------

$$\left(\frac{y_1 - y_n}{x_1 - x_n}\right) = a + \left(\frac{\xi_1 - \xi_n}{x_1 - x_n}\right)$$

Set -2:

$$\left(\frac{y_2 - y_1}{x_2 - x_1}\right) = a + \left(\frac{\xi_2 - \xi_1}{x_2 - x_1}\right)$$
$$\left(\frac{y_2 - y_3}{x_2 - x_3}\right) = a + \left(\frac{\xi_2 - \xi_3}{x_2 - x_3}\right)$$

$$\left(\frac{y_2 - y_n}{x_2 - x_n}\right) = a + \left(\frac{\xi_2 - \xi_n}{x_2 - x_n}\right)$$

Set -*n*:

$$\left(\frac{y_n - y_1}{x_n - x_1}\right) = a + \left(\frac{\xi_n - \xi_1}{x_n - x_1}\right)$$
$$\left(\frac{y_n - y_2}{x_n - x_2}\right) = a + \left(\frac{\xi_n - \xi_2}{x_n - x_2}\right)$$

$$\left(\frac{y_n - y_{n-1}}{x_n - x_{n-1}}\right) = a + \left(\frac{\xi_n - \xi_{n-1}}{x_n - x_{n-1}}\right)$$

Now, Since $a_i = a + \zeta_i$, $i = 1,2,........n$ therefore, the sum of squares of errors is

$$S = \sum_{i=1}^{n} \zeta_i^{\,2} = \sum_{i=1}^{n} (a_i - a)^2$$

Differentiating S w.r.t 'a' and equating to zero we get

$$\therefore \frac{\delta S}{\delta a} = 0$$

Which yield

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

$$\Rightarrow \quad \sum_{i=1}^{n} a_i = na$$

$$\therefore \quad \hat{a} = \frac{\sum_{i=1}^{n} a_i}{n} \qquad (2.3)$$

Using the value of (2.3) in (2.1) we get the value of 'b'

$$\hat{b} = \bar{y} - \hat{a}\bar{x} \qquad (2.4)$$

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Here we have considered average of mean minimum and maximum temperature of five cities in the context of Assam as observed data to fit the following linear equation.

NUMERICAL EXAMPLE: *ON MINIMUM TEMPERATURE*

Let the linear equation be

$Y_i = aX_i + b$ (i = 1, 2. …….n)

Where Y_i = Average of mean minimum Temperature.

X_i = Length of the day.

Ex: 3.1: Average of mean minimum temperature of Guwahati:

X_i	**10.553**	**11.105**	**11.834**	**12.610**	**13.266**	**13.605**	**13.469**	**12.921**	**12.191**	**11.424**	**10.755**	**10.398**
Y_i	**10.815**	**14.849**	**16.087**	**20.488**	**22.793**	**25.045**	**25.660**	**25.635**	**24.650**	**22.058**	**16.982**	**12.226**

Solution: The matrix $(C_{ij})_{12x12}$ where $C_{ij} = \dfrac{(y_i - y_j)}{(x_i - x_j)}$ has been obtained as

0	7.308	4.116	4.703	4.415	4.663	5.091	6.258	8.446	12.908	30.530	-9.103
7.310	0	1.698	3.747	3.676	4.078	4.573	5.939	9.025	22.599	-6.094	3.710
4.116	1.698	0	5.671	4.683	5.058	5.855	8.784	23.986	-14.563	-0.829	2.689
4.703	3.747	5.671	0	3.514	4.580	6.021	16.550	-9.933	-1.324	1.890	3.735
4.415	3.676	4.683	3.514	0	6.643	14.123	-8.238	-1.727	0.399	2.314	3.684
4.662	4.078	5.058	4.580	6.643	0	-4.522	-0.863	0.279	1.370	2.829	3.997
5.091	4.573	5.855	6.021	14.123	-4.522	0	0.046	0.790	1.761	3.197	4.375
6.258	5.939	8.784	16.550	-8.238	-0.862	0.0466	0	1.349	2.389	3.995	5.315
8.446	9.025	23.986	-9.933	-1.727	0.279	0.790	1.349	0	3.379	5.340	6.929
12.908	22.599	-14.563	-1.324	0.399	1.370	1.761	2.389	3.379	0	7.587	9.583
30.530	-6.094	-0.829	1.890	2.314	2.829	3.197	3.995	5.340	7.587	0	13.322
-9.103	3.710	2.689	3.735	3.684	3.997	4.374	5.315	6.929	9.583	13.322	0

The equation (2.3) which gives the estimate of the parameter '*a*' as shown below

$$\widehat{a} = \frac{\sum_{i=1}^{n} a_i}{n}$$

= 4.368164802

The equation (2.4) gives the estimates of '*b*'

$$\widehat{b} = \overline{y} - \widehat{a}\overline{x}$$

= -32.69166344

Now, the normal equations to estimate *'a' and 'b'* are as follows

237.288 = 144.131 a + 12 b

2909.975716 = 1746.108159 a + 144.131.b

Hence, in this case

$$\widehat{b}_{(NE)} = -28.33260446$$

$\widehat{a}_{(NE)}$ = 4.005240049 $\widehat{a}_{(stw)}$ = **4.368164802**

Result: $\widehat{b}_{(NE)}$ = **-28.33260446** $\widehat{b}_{(stw)}$

$\widehat{a}_{(NE)}$ = **4.005240049** = **-32.69166344**

Estimated value of temperature ($\widehat{y}$) by both the methods: Guwahati

Table: 3.1(a)

Length of Day (x)	**Observed Temperature (y)**	**Estimated Temperature** $\widehat{y}_{(STW)}$	**Estimated Temperature** $\widehat{y}_{(NE)}$	**Estimates of Error** $\lvert\widehat{e}_{(STW)}\rvert = \lvert(y - \widehat{y}_{(STW)})\rvert$	**Estimates of Error** $\lvert\widehat{e}_{(NE)}\rvert = \lvert(y - \widehat{y}_{(NE)})\rvert$
10.553	10.815	13.40557972	13.93469378	2.59057972	3.11969378
11.105	14.849	15.81680669	16.14558628	0.96780669	1.29658628
11.834	16.087	19.00119883	19.06540628	2.91419883	2.97840628
12.610	20.488	22.39089471	22.17347256	1.90289471	1.68547256
13.266	22.793	25.25641082	24.80091003	2.46341082	2.00791003
13.605	25.045	26.73721869	26.15868641	1.69221869	1.11368641
13.469	25.660	26.14314828	25.61397376	0.48314828	0.04602624
12.921	25.635	23.74939397	23.41910221	1.88560603	2.21589779
12.191	24.650	20.56063366	20.49527698	4.08936634	4.15472302
11.424	22.058	17.21025126	17.42325786	4.84774874	4.63474214
10.755	16.982	14.28794901	14.74375227	2.69405099	2.23824773
10.398	12.226	12.72851436	13.31388157	0.50251436	1.08788157
Total = 144.131	**Total = 237.288**	**Total = 237.288**	**Total = 237.288**	**Sum of Absolute Deviation** $\sum\lvert\widehat{e}_{(STW)}\rvert$ = **26.96351241**	**Sum of Absolute Deviation** $\sum\lvert\widehat{e}_{(NE)}\rvert$ = **26.57927383**

Absolute Mean Deviation ($\bar{e}_{(STW)}$ **) = 2.246959368**

Absolute Mean Deviation ($\bar{e}_{(NE)}$ **) = 2.214939486**

H_0: There is no significant difference between the values of observed temperature and estimated temperature.

Under the null hypothesis H_0, the test statistic is

$$t = \frac{(\bar{y} - \bar{\widehat{y}}_{(STW)})}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with } (n_1 + n_2 - 2) \text{ d.f.}$$

Where

$$S^2 = \frac{1}{(n_1 + n_2 - 2)}\left[\sum_{i=1}^{12}(y_i - \bar{y})^2 + (\widehat{y}_i - \bar{\widehat{y}}_{(STW)})^2\right]$$

And $n_1 = n_2 = 12$

Since $\bar{y} = \bar{\widehat{y}}_{(STW)} = 19.774$

$\therefore \; \lvert t\rvert_{cal} = 0$ and $t_{(tab, 5\%, 22 d.f)} = 1.717$

$\lvert t\rvert_{cal} < t_{(tab, 5\%, 22 d.f)}$

1. Test of significance for estimated temperature obtained by Stepwise Application of Principles of Least Squares (stw).

The null hypothesis to be tested is

Thus, the null hypothesis H_0 is accepted.

Accordingly, it can be concluded that the difference between observed temperature and the corresponding estimated temperature is insignificant.

2. Test of significance for estimated temperature obtained by solving normal equations (NE).

The null hypothesis to be tested is

H_0: There is no significant difference between the values of observed temperature and estimated temperature.

Under the null hypothesis H_0, the test statistic is

$$t = \frac{(\bar{y} - \bar{\widehat{y}}_{(NE)})}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with } (n_1 + n_2 - 2) \text{ d.f.}$$

Where

$$S^2 = \frac{1}{(n_1 + n_2 - 2)}\left[\sum_{i=1}^{12}(y_i - \bar{y})^2 + (\widehat{y}_i - \bar{\widehat{y}}_{(NE)})^2\right]$$

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

And $n_1 = n_2 = 12$

Here, $\bar{y} = \hat{\bar{y}}_{(NE)} = 19.774$

$\therefore \quad |t|_{cal} = 0$ and $t_{(tab,5\%,22d.f)} = 1.717$

$|t|_{cal} < t_{(tab,5\%,22d.f)}$

Thus, the null hypothesis H_0 is accepted.

Accordingly, it can be concluded that the difference between observed temperature and the corresponding estimated temperature is insignificant.

CONCLUSION

The method, developed here, is based on the principle of the elimination of parameters first and then the minimization of the sum of squares of the errors whiles the ordinary least squares is based on the principle of the minimization of the sum of squares of the errors first and then elimination of parameters.

It is to be noted that the number of steps of computations in estimating the parameters by the method introduced here is less than the number of steps in estimation of parameters by the solutions of the normal equations. This implies that the error that occurs due to approximation in computation is less in the former than in the later.

We, therefore, may conclude that stepwise application of principles of least squares method is a simpler method of obtaining least square estimates of parameters of linear curve than the method of solving the normal equations.

The following tables **(Table-4.1(a))** show the values of *t* for the testing the significance of difference between the observed temperature and estimated temperature by both the method and comparison of their't' values.

Table: 4.1(a)

Ex. No.	***Values of 't' in case of method of STW***	***Hypothesis***	***Significance /Insignificance***
1	$t_{cal} = 0 < t_{(tab,5\%.22d.f)} = 1.717$	H_0 Accepted	Insignificant
	Values of 't' in case of method of solution of NE	***Hypothesis***	***Significance /Insignificance***
2	$t_{cal} = 0 < t_{(tab,5\%.22d.f)} = 1.717$	H_0 Accepted	Insignificant
Comparison of 't' values of both the methods			
$t_{(STW)} = t_{(NE)}$			

From the above table, It is found that both the method are almost equal in estimating parameters associated with a linear equation in case of unequal interval of the independent variable. ·

In this study, attempt has been made for the case of linear curve only. Other types of the curves are yet to be dealt

REFERENCE

1. Adrian, A. (1808): The Analyst. No.-IV. Pq. 93-109
2. Bassel (1838): On untersuchungen ueber die Wahrscheinlichkeit der Beobachtungsfehler.
3. Crofton (1870): On the proof of the law of error of observations. Phil. Trans, London, pp 175-188
4. Donkin (1844): An essay on the theory of the combination of observations. Joua. Math., XV: 297-322
5. Gauss, C.F. (1809): Theory of Motion of Heavenly bodies. Humburg. Pp 205-224
6. Hagen (1837): Grandzuge der Wahrscheinlichkeitsrechnung, Berlin.
7. Herschel, J. (1850): Quetelet On Probabilities. Edinburgh Review, Simpler XCII: 1-57.
8. Ivory (1825): On the method of least squares, Phil. Mag., LXV: 3-10
9. Koppelmann.R.(1971).The calculus of operation, History of Exact Sc., 8: 152-242
10. Laplace (1810): Theoretic Analytique des prob. Ch-IV.
11. Merriman.M (1877): The Analyst. Vol.-IV, NO.-2
12. Rahman A. and Chakrabarty D.(2009): " Linear Curve: A Simpler Method of Obtaining Least Squares Estimates of the Parameters", Int. J. Agricult. Stat. Sci., 5(2), 415 – 42