METHOD OF LEAST SQUARES IN REVERSE ORDER: FITTING OF LINEAR CURVE TO AVERAGE MAXIMUM TEMPERATURE DATA AT GUWAHAT AND TEZPUR

***Atwar Rahman**
Department of Statistics
Pub Kamrup College
Baihata Chariali
Kamrup, Assam, India

and

Dhritikesh Chakrabarty
Department of Statistics
Handique Girls' College
Guwahati-1, Assam, India

ABSTRACT

In 2015, Rahman and Chakrabarty have developed one method by stepwise application of the principle of least squares of estimating parameters associated to a linear curve based on the principle of elimination of parameter(s) first and then minimization of sum of squares of errors in place of the principle of least squares namely minimization of sum of squares of errors first and then elimination of parameter(s), where all values of the ratio of the difference of each pair of observed values of the dependent variable to the difference of each pair of observed values of the independent variable have been used in estimating parameters involved in the curve. In this paper, discussion has been made on how parameters can be estimated using the independent values from among the values of the said .ratio. One numerical application of the method has also been discussed in fitting of linear regression of monthly mean maximum temperature on the monthly average length of day at Guwahati and also at Tezpur.

Key Words

Linear curve, Least squares principle, Monthly mean maximum temperature, Monthly average length, Stepwise application

1. 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.

2: ESTIMATION OF PARAMETERS IN LINEAR CURVE:

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

$$Y = a + b X \qquad (2.1)$$

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

Let $Y_1, Y_2 \ldots\ldots, Y_n$ be n observations on Y corresponding to the observations $X_1, X_2, \ldots\ldots\ldots, X_n$ 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

$(X_1, Y_1), (X_2, Y_2), \ldots\ldots\ldots\ldots\ldots, (X_n, Y_n)$ may not lie on the curve (2.1), they satisfy the model

$$y_i = a + b x_i + \xi_i, \qquad (i = 1, 2 \ldots n) \qquad (2.2)$$

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

From (2.2), the following $\frac{n(n+1)}{2}$ independent equations can be obtained.

$$\left(\frac{y_1 - y_2}{x_1 - x_2}\right) = b + \left(\frac{\xi_1 - \xi_2}{x_1 - x_2}\right)$$

$$\left(\frac{y_1 - y_3}{x_1 - x_3}\right) = b + \left(\frac{\xi_1 - \xi_3}{x_1 - x_3}\right)$$

$$\left(\frac{y_1 - y_4}{x_1 - x_4}\right) = b + \left(\frac{\xi_1 - \xi_4}{x_1 - x_4}\right)$$

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

$$\left(\frac{y_2 - y_3}{x_2 - x_3}\right) = b + \left(\frac{\xi_2 - \xi_3}{x_2 - x_3}\right)$$

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

……

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

Now, writing $$Z_{ij} = \left(\frac{y_i - y_j}{x_i - xj}\right)$$

and $$e_{ij} = \left(\frac{\xi_i - \xi_j}{x_i - x_j}\right)$$

We have

$$Z_{ij} = b + e_{ij} \qquad i = 1,2,---n, \quad j = 1,2,---n, \quad i \angle j$$

Now minimizing $$S = \sum_{i-1}^{n}\sum_{j=1}^{n}\xi_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(Z_{ij} - b\right)^2$$

$i \angle j$ $i \angle j$

With respect to *'b'*, the LSE of *'b'* can be as

$$\hat{b} = \frac{2}{n(n+1)}\sum_{i=1}^{n}\sum_{j=1}^{n}Z_{ij} \qquad (2.3)$$

$$i \angle j$$

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

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

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

Let the linear equation be

$$Y_i = a + b X_i \qquad (i = 1, 2. \ldots\ldots n)$$

Where Y_i = Average of mean maximum Temperature.

X_i = Length of the day.

3: NUMERICAL PROBLEM: ON MAXIMUM TEMPERATURE

Ex: 3.1: Average of mean maximum 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	23.536	26.226	29.972	30.883	31.363	31.768	31.995	32.470	31.720	30.383	27.731	24.724

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

4.873

5.024 5.139

3.572 3.094 1.174

2.885 2.377 0.971 0.732

2.697 2.217 1.014 0.889 1.195

2.901 2.440 1.237 1.295 3.113 -1.669

3.773 3.438 2.298 5.103 -3.209 -1.026 -0.867

4.996 5.059 4.896 -1.998 -0.332 0.034 0.215 1.027

7.861 13.031 -1.002 0.422 0.532 0.635 0.788 1.394 1.743

20.767 -4.300 2.077 1.699 1.446 1.417 1.571 2.188 2.278 3.964

-7.665 2.124 3.654 2.784 2.315 2.196 2.368 3.070 3.902 5.516 8.423

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

$$\hat{b}_{(STW)} = \frac{2}{n(n+1)} \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ij}$$

= 2.390530303

The equation (2.4) gives

$$\hat{a}_{(STW)} = \bar{y} - \hat{b}_{(STW)} \bar{x}$$

= 0.685123063

In this example, the normal equations for estimating *a* & *b* are

352.771 = 12 a + 144.131 b

& 4271.610108 = 144.131 a + 1746.108159 b

Thus in this case,

$\hat{b}_{(NE)}$ = 2.306198622

$\hat{a}_{(NE)}$ = 1.698023869

Result:

$\hat{b}_{(NE)}$ **= 2.306198622** $\hat{b}_{(stw)}$ = **2.390530303**

$\hat{a}_{(NE)}$ **= 1.698023869** $\hat{a}_{(stw)}$ **= 0.685123063**

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

Table: 3.1(a)

Length of Day (x)	Observed Temperature (y)	Estimated Temperature $\hat{y}_{(STW)}$	Estimated Temperature $\hat{y}_{(NE)}$	Estimates of Errors $\lvert\hat{e}_{(NE)}\rvert = \lvert(y - \hat{y}_{(NE)})\rvert$	Estimates of Errors $\lvert\hat{e}_{(STW)}\rvert = \lvert(y - \hat{y}_{(STW)})\rvert$
10.553	23.536	25.91238935	26.03533793	2.49933793	2.37638935
11.105	26.226	27.23196208	27.30835957	1.08235957	1.00596208
11.834	29.972	28.97465867	28.98957836	0.98242164	0.99734133
12.610	30.883	30.82971018	30.77918849	0.10381151	0.05328982
13.266	31.363	32.39789806	32.29205479	0.92905479	1.03489806
13.605	31.768	33.20828784	33.07385612	1.30585612	1.44028784
13.469	31.995	32.88317571	32.76021311	0.76521311	0.88817571
12.921	32.470	31.57316511	31.49641626	0.97358374	0.89683489
12.191	31.720	29.82807799	29.81289127	1.90710873	1.89192201
11.424	30.383	27.99454124	28.04403693	2.33896307	2.38845876
10.755	27.731	26.39527647	26.50119005	1.22980995	1.33572353
10.398	24.724	25.54185715	25.67787714	0.95387714	0.81785715
Total = 144.131	**Total = 352.771**	**Total = 352.7709999 $\cong$ 352.771**	**Total= 352.771**	**Sum of Absolute Deviation $\sum\lvert\hat{e}_{(NE)}\rvert$ =15.0713973**	**Sum of Absolute Deviation $\sum\lvert\hat{e}_{(STW)}\rvert$ = 15.12714053**

Absolute Mean Deviation ($\bar{e}_{(STW)}$) = 1.26059504

Absolute Mean Deviation ($\bar{e}_{(NE)}$) = 1.255949775

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

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

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

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

And $$n_1 = n_2 = 11$$

Since $\bar{y} = \hat{\bar{y}}_{(STW)}$ = 29.39758333

$|t|_{cal}$ = 0 and $t_{(tab,5\%,20df)}$ = 1.725

$|t|_{cal} < t_{(tab,5\%,20df)}$

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 solution of 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

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

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

And $$n_1 = n_2 = 11$$

Since $\bar{y} = \hat{\bar{y}}_{(NE)}$ = 29.39758333

$|t|_{cal}$ = 0 and $t_{(tab,5\%,20df)}$ = 1.725

$$|t|_{cal} < t_{(tab,5\%,20df)}$$

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.

Ex: 3.2: Average of mean maximum temperature of :Tezpur

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	23.559	25.976	29.397	30.015	30.895	31.867	31.900	32.197	31.587	30.680	28.195	24.716

Solution: The matrix Z_{ij} where $Z_{ij\,j} = \dfrac{(y_i - y_j)}{(x_i - x_j)}$ has been obtained as

4.378

4.557 4.693

3.139 2.684 0.796

2.704 2.276 1.046 1.341

2.722 2.356 1.395 1.861 2.867

2.860 2.506 1.531 2.194 4.951 -0.243

3.648 3.426 2.576 7.016 -3.774 -0.483 -0.542

4.901 5.167 6.134 -3.752 -0.644 0.198 0.245 0.836

8.176 14.746 -3.129 -0.561 0.117 0.544 0.597 1.013 1.183

22.950 -6.340 1.114 0.981 1.075 1.288 1.365 1.847 2.362 3.714

-7.465 1.782 3.260 2.396 2.154 2.230 2.339 2.965 3.832 5.812 9.745

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

$$\hat{b}_{(STW)} = \frac{2}{n(n+1)} \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ij}$$

= 2.419060606

The equation (2.4) gives

$$\hat{a}_{(STW)} = \bar{y} - \hat{b}\bar{x}$$

= 0.193531319

The normal equations to estimate *'a'* and *'b'* become

350.984 = 12a + 144.131 b

4248.336627 = a144.131 + 1746.108159 b

By solving these equations we get

$\hat{a}_{(NE)}$ = 3.002012944 $\hat{b}_{(NE)}$ = 2.185233188

Result:

$\hat{a}_{(NE)}$ **= 3.002012944** $\hat{a}_{(stw)}$ **= 0.193531319**

$\hat{b}_{(NE)}$ **= 2.185233188** $\hat{b}_{(stw)}$ **= 2.419060606**

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

Table: 3.2(a)

Length of Day (x)	**Observed Temperature (y)**	**Estimated Temperature** $\hat{y}_{(STW)}$	**Estimated Temperature** $\hat{y}_{(NE)}$	**Estimates of Errors** $\left\|\hat{e}_{(NE)}\right\| = \left\|(y - \hat{y}_{(NE)})\right\|$	**Estimates of Errors** $\left\|\hat{e}_{(STW)}\right\| = \left\|(y - \hat{y}_{(STW)})\right\|$
10.553	23.559	25.72187789	26.06277878	2.50377878	2.16287789
11.105	25.976	27.05719935	27.26902750	1.29302750	1.08119935
11.834	29.397	28.82069453	28.86206249	0.53493751	0.57630547
12.610	30.015	30.69788556	30.55780344	0.54280344	0.68288556
13.266	30.895	32.28478932	31.99131642	1.09631642	1.38978932
13.605	31.867	33.10485086	32.73211047	0.86511047	1.23785086
13.469	31.900	32.77585862	32.43491875	.053491875	0.87585862
12.921	32.197	31.45021341	31.23741097	0.95958903	0.74678659
12.191	31.587	29.68429917	29.64219074	1.94480926	1.90270083
11.424	30.680	27.82887968	27.96611688	2.71388312	2.85112032
10.755	28.195	26.21052814	26.50419588	1.69080412	1.98447186
10.398	24.716	25.34692350	25.72406763	1.00806763	0.63092350
Total = 144.131	**Total =350.984**	**Total =350.984**	**Total = 350.984**	**Sum of Absolute Deviation** $\sum\left\|\hat{e}_{(NE)}\right\|$ **= 15.68804603**	**Sum of Absolute Deviation** $\sum\left\|\hat{e}_{(STW)}\right\|$ = **16.12277017**

Absolute Mean Deviation ($\bar{e}_{(STW)}$ **) = 1.343564181**

Absolute Mean Deviation ($\bar{e}_{(NE)}$ **) = 1.307337169**

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

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

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

Where $$S^2 = \frac{1}{(n_1 + n_2 - 2)}\left[\sum_{i=1}^{12}(y_i - \bar{y})^2 + (\hat{y}_i - \hat{\bar{y}}_{(STW)})^2\right] \quad \text{and} \quad n_1 = n_2 = 11$$

S = 3.196123506

Since $\bar{y} = \hat{\bar{y}}_{(STW)}$ = 29.24866667

$|t|_{cal}$ = 0 and $t_{(tab,5\%,20df)}$ = 1.725 [

$|t|_{cal} < t_{(tab,5\%,20df)}$

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 solution of 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

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

Where $S^2 = \frac{1}{(n_1+n_2-2)}\left[\sum_{i=1}^{12}(y_i-\bar{y})^2+(\hat{y}_i-\hat{\bar{y}}_{(NE)})^2\right]$ and $n_1 = n_2 = 11$\

S = 3.067057039

$\bar{y}$ = $\hat{\bar{y}}_{(NE)}$ = 29.24866667

$|t|_{cal}$ = 0 and $t_{(tab,5\%,20df)}$ = 1.725

$|t|_{cal}$ < $t_{(tab,5\%,20df)}$

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.

4. 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)) & (Table – 4.2(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 < $t_{(tab,5\%,20df)}$ =1.725	H_0 Accepted	Insignificant
2	t_{cal} = 0 < $t_{(tab,5\%,20df)}$ =1.725	H_0 Accepted	Insignificant

Table – 4.2(a)

Ex. No.	***Values of 't' in case of method of solution of NE***	***Hypothesis***	***Significance /Insignificance***
1	t_{cal} = 0 < $t_{(tab,5\%,20df)}$ = 1.725	H_0 Accepted	Insignificant
2	t_{cal} = 0 < $t_{(tab,5\%,20df)}$ = 1.725	H_0 Accepted	Insignificant

The following table (Table-4.3(a)) shows the comparison between the two methods of estimating parameters.

Table – 4.3(a)

Ex. No.	***Comparison of 't' values of both the methods***
1	$t_{(STW)} = t_{(NE)}$
2	$t_{(STW)} = t_{(NE)}$

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

5. 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. 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. 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.(2015): Elimination of parameters and principle of least squares: Fitting of linear curve to average minimum temperature data in the context of Assam", International Journal of Engineering Sciences & Research Technology, Vol. 4, No. 2, Feb, 2015.

13. Rahman A. and Chakrabarty D.(2015): Elimination of parameters and principle of least squares: Fitting of linear curve to average maximum temperature data in the context of Assam", Aryabhatta Journal of Mathematics & Informatics, Vol. 7, No. 1, Jan – June, 2015.