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

and

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

$$Y = a + b X \tag{2.1}$$

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

$$y_i = a + b x_i + \xi_i,$$
 (*i* = 1, 2... *n*) (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.

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

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

------

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

\_\_\_\_\_

.....

\_\_\_\_\_

\_\_\_\_\_

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories
International Journal in Physical & Applied Sciences
<a href="http://www.ijmr.net.in">http://www.ijmr.net.in</a> email id- <a href="http://www.ijmr.net.in">irjmss@gmail.com</a> Page 26

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

 $Z_{ij} = \left(\frac{y_i - y_j}{x_i - x_j}\right)$ 

\_\_\_\_\_

\_\_\_\_\_

Now, writing

$$e_{ij} = \left(rac{\xi_i - \xi_j}{x_i - x_j}
ight)$$

We have

and

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

Now minimizing

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

i∠j i∠j

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

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

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

$$\widehat{a} = \overline{y} - b\overline{x} \tag{2.4}$$

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

 $Y_i = a + b X_i$  (i = 1, 2. .....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:

Х	10.55	11.10	11.83	12.61	13.26	13.60	13.46	12.92	12.19	11.42	10.75	10.39
i	3	5	4	0	6	5	9	1	1	4	5	8
Y	23.53	26.22	29.97	30.88	31.36	31.76	31.99	32.47	31.72	30.38	27.73	24.72
i	6	6	2	3	3	8	5	0	0	3	1	4

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

$\left( \right)$	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	
	<b>. . .</b>		2 2) 4		4h		<b></b>			h . l		

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

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

= 2.390530303

The equation (2.4) gives

 $\widehat{a}_{(STW)} = \overline{y} - \widehat{b}_{(STW)}\overline{x}$ 

= 0.685123063

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

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 (  $\widehat{y}$  ) by both the methods: Guwahati

Length of Day	Observed	Estimated	Estimated	Estimates of Errors	Estimates of Errors
(x)	Temperature (y)	Temperature	Temperature	$\left  \widehat{e}_{(NE)} \right  = \left  (y - \widehat{y}_{(NE)}) \right $	$\left  \widehat{e}_{(STW)} \right  = \left  (y - \widehat{y}_{(STW)}) \right $
		$\widehat{\mathcal{Y}}_{(STW)}$	$\widehat{\mathcal{Y}}_{(NE)}$		
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 =	Total= 352.771	Sum of Absolute Deviation	Sum of Absolute Deviation
		352.7709999		$\sum  \hat{e}_{(NE)}  = 15.0713973$	$\sum  \hat{e}_{(cTW)} $
		≅ <b>352.771</b>			$\sum_{i=1}^{n} (SIW)$
					= 15.12714053

# Table: 3.1(a)

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

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

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

Where

$$S^{2} = \frac{1}{(n_{1} + n_{2} - 2)} \left[ \sum_{i=1}^{12} (y_{i} - \overline{y})^{2} + (\widehat{y}_{i} - \widehat{\overline{y}}_{(STW)})^{2} \right]$$

And

 $n_1 = n_2 = 11$ 

Since

 $\overline{y} = \widehat{\overline{y}}_{(STW)} = 29.39758333$ 

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

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

Thus, the null hypothesis H<sub>0</sub> 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<sub>0</sub>, the test statistic is

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

Where

$$S^{2} = \frac{1}{(n_{1} + n_{2} - 2)} \left[ \sum_{i=1}^{12} (y_{i} - \overline{y})^{2} + (\widehat{y}_{i} - \widehat{\overline{y}}_{(NE)})^{2} \right]$$

And

 $n_1 = n_2 = 11$ 

Since

$$\overline{y} = \overline{\overline{y}}_{(NE)} = 29.39758333$$

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

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories International Journal in Physical & Applied Sciences <u>http://www.ijmr.net.in</u> email id- <u>irjmss@gmail.com</u> Page 31

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

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

Х	10.55	11.10	11.83	12.61	13.26	13.60	13.46	12.92	12.19	11.42	10.75	10.39
i	3	5	4	0	6	5	9	1	1	4	5	8
Y	23.55	25.97	29.39	30.01	30.89	31.86	31.90	32.19	31.58	30.68	28.19	24.71
i	9	6	7	5	5	7	0	7	7	0	5	6

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

(												$\overline{\ }$
[	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	

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories International Journal in Physical & Applied Sciences http://www.ijmr.net.in email id- irjmss@gmail.com

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

 $\widehat{a}_{(STW)} = \overline{y} - \widehat{b}\overline{x}$ 

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

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

Length of	Observed	Estimated	Estimated	Estimates of Errors	Estimates of Errors	
Day (x)	Temperature	Temperature	Temperature	$\left  \widehat{e}_{(NE)} \right  = \left  (y - \widehat{y}_{(NE)}) \right $	$\left  \widehat{e}_{(STW)} \right  = \left  (y - \widehat{y}_{(STW)}) \right $	
	(y)	$\widehat{\mathcal{Y}}_{(STW)}$	$\widehat{\mathcal{Y}}_{(NE)}$			
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 =	Total	Total	Total =	Sum of Absolute	Sum of Absolute	
144.131	=350.984	=350.984	350.984	Deviation	Deviation	
				$\sum \left  \widehat{e}_{(NE)} \right $	$\sum \left  \widehat{e}_{(STW)} \right  =$	
				= 15.68804603	16.12277017	

# Table: 3.2(a)

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

Absolute Mean Deviation (  $\overline{e}_{(NE)}$ ) = 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<sub>0</sub>: There is no significant difference between the values of observed temperature and estimated temperature.

Under the null hypothesis H<sub>0</sub>, the test statistic is

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

Where

$$S^{2} = \frac{1}{(n_{1} + n_{2} - 2)} \left[ \sum_{i=1}^{12} (y_{i} - \overline{y})^{2} + (\widehat{y}_{i} - \widehat{\overline{y}}_{(STW)})^{2} \right] \text{ and } n_{1} = n_{2} = 11$$

[

*S* = 3.196123506

Since

$$\overline{y} = \overline{\hat{y}}_{(STW)} = 29.248666667$$

$$|t|_{cal} = 0$$
 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<sub>0</sub>: There is no significant difference between the values of observed temperature and estimated temperature.

Under the null hypothesis H<sub>0</sub>, the test statistic is

$$t = \frac{(\bar{y} - \hat{\bar{y}}_{(NE)})}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
 with (n<sub>1</sub> + n<sub>2</sub> -2) d.f.

Where

 $S^{2} = \frac{1}{(n_{1} + n_{2} - 2)} \left[ \sum_{i=1}^{12} (y_{i} - \overline{y})^{2} + (\widehat{y}_{i} - \widehat{\overline{y}}_{(NE)})^{2} \right] \text{ and } n_{1} = n_{2} = 11 \sqrt{\frac{1}{2}}$ *S* = 3.067057039  $\overline{y} = \hat{\overline{y}}_{(NE)} = 29.248666667$  $|t|_{cal} = 0$  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	i)
---------------	----

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

#### Table – 4.2(a)

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

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

#### Table – 4.3(a)

Ex.	Comparison of 't' values of both the methods
No.	
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.

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.