

# A New Method to Estimate Parameters in the Simple Regression Linear Equation

Agung Prabowo<sup>1,\*</sup>, Agus Sugandha<sup>1</sup>, Agustini Tripena<sup>1</sup>, Mustafa Mamat<sup>2</sup>, Sukono<sup>3</sup>, Ruly Budiono<sup>4</sup>

<sup>1</sup>Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Jenderal Soedirman, Indonesia

<sup>2</sup>Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Malaysia

<sup>3</sup>Department of Mathematics, Faculty of Mathematics and Natural Science, Padjadjaran University, Indonesia

<sup>4</sup>Department of Biology, Faculty of Mathematics and Natural Science, Padjadjaran University, Indonesia

*Received November 20, 2019; Revised December 25, 2019; Accepted December 30, 2019*

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

**Abstract** Linear regression is widely used in various fields. Research on linear regression uses the OLS and ML method in estimating its parameters. OLS and ML method require many assumptions to complete. It is frequently found there is an unconditional assumption that both methods are not successfully used. This paper proposes a new method which does not require any assumption with a condition. The new method is called SAM (Simple Averaging Method) to estimate parameters in the simple linear regression model. The method may be used without fulfilling assumptions in the regression model. Three new theorems are formulated to simplify the estimation of parameters in the simple linear regression model with SAM. By using the same data, the simple linear regression model parameter estimation is conducted using SAM. The result shows that the obtained regression parameter is not quite far different. However, to measure the accuracy of both methods, a comparison of errors made by each method is conducted using Root Mean Square Error (RMSE) and Mean Averaged Error (MAE). By comparing the values of RMSE and MAE for both methods, SAM method may be used to estimate parameters in the regression equation. The advantage of SAM is free from all assumptions required by regression, such as error normality assumption while the data should be from the normal distribution.

**Keywords** Linear Regression, Parameter, Ordinary Least Square Method, Simple Average Method

## 1. Introduction

Linear regression is widely used in various fields. Research on linear regression uses the OLS method in estimating its parameters. The OLS method requires many

assumptions that need to be met. In this paper, a new method is offered that does not require the fulfillment of any assumptions. In terms of calculations, the new method is very practical and does not require high-level mathematics, making it suitable for use by researchers in various fields of research.

The relationship between two or more variables can be analyzed through the association and regression relationships. The association relationship is tenuous or weak (Sembiring, 2003) because in the association only the value of  $x$  and  $y$  is paired without knowing the form of the relationship between the two variables.

Two variables can be independent of one another or completely interdependent. In the event that the two variables are mutually independent, the correlation is zero. If the two variables are completely interdependent and the relationship is linear (both are called collinear), then the absolute price is one correlation.

In research generally used mathematical models or models which are a simplification and abstraction of the actual natural state. The naturally conditions under study are generally very complicated and the ability to examine them as a whole is also very limited. For this reason, simplifications are needed according to our ability.

According to Sembiring (2003), the use of model is to understand, explain, control, and predict the observed system behaviors. In predicting activities and not predicting, there is a special meaning of interpolation and extrapolation.

The created model illustrates a causal relationship (cause and effect) between two or more variables that the model shows a functional relationship between the constructed variables.

Predictions about the relationship between variables in the observed system are used to formulate the behavior of the system in various situations. Thus, the model created is a theory about the workings of the system under study.

Furthermore, the formulation of the relationship is stated in the form of a hypothesis and tested based on statistical data collected. This approach is induction as opposed to an axiom (deduction).

The model is a functional relationship between the variables that build it so that the model used will be in the form of function and regression to be a powerful tool in its formation. There are two types of data used namely laboratory data (there are controls so that they no longer describe the natural state by manipulating variables that interfere with the way it is made unchanged so that it has no effect and consequently the effect of the investigated variables can be cleaner observed). The second type of data is field data describing natural conditions and containing the influence of many variables that work together very complicated.

The created model illustrates a causal relationship (cause and effect) between two or more variables that the model shows a functional relationship between the constructed variables.

The created model may be in the form of function and regression to become an ultimate weapon in its creation. Linear regression is frequently used in various fields. One model to relate the independent variable with the dependent variable is the simple linear regression model. The use of this model may be found in any field. The research in simple linear regression using OLS (Ordinary Least Square) or ML (Maximum Likelihood) method is conducted to estimate the parameters. Some researchers including Chioma (2009), Cliff dan Billy (2017); Prabitha and Archana (2015); Jain *et al.* (2016); Gustiarisanie, Rahardjo and Ernawati (2016; Permai and Tanty (2018; and Singh and Masuku (2013) use OLS and ML method to estimate parameters in the linear regression.

OLS and ML method require many assumptions to complete. It is frequently found that there is an unconditional assumption that that both methods are not successfully used. This paper proposes a new method which does not require any assumption with a condition. From the calculation point of view, this new method is considered highly practical and does not require higher mathematic level that it is appropriate for the researchers to conduct researchers in various fields.

## 2. Materials and Methods

At the initial stage, it is explained that the simple linear regression and its parameters estimation using OLS method and assumptions should be well fulfilled that the method may be used. There are 12 data sample pairs used to obtain the simple linear regression equation using OLS method.

Furthermore, it is explained related to SAM (Simple Averaging Method) developed by Cliff and Billy (2017) to estimate parameters in the simple linear regression model.

By using the same data, the simple linear regression model parameter estimation is conducted using SAM. The result shows that the obtained regression parameter is not quite far different. However, to measure the accuracy of both methods, a comparison of errors made by each method is conducted using RMSE (Root Mean Square Error) and MAE (Mean Averaged Error). There is no significant different in measuring the errors that SAM may become an alternative method in simple linear regression parameter estimation.

In the next stage, the method offered by Cliff and Billy (2017) is simplified in the form of a mathematical equation. Thus, three new theorems are constructed in order to estimate the simple linear regression parameters using SAM. The proofs for those theorems are given. The obtained three theorems may be used to estimate the simple linear regression parameters free from any assumption.

## 3. Results

The simplest linear regression model is a straight line. In this case there is a free variable, name  $x$  and dependent variable that depends on  $x$ , named  $y$ . The dependent variable  $y$  is sometimes called the response variable. The independent variable  $x$  is called the predictor variable.

Examples of problems that can be solved by making a simple linear regression model are a student's achievement in mathematics (variable  $x$ ) which is determined by the length of time a teacher has received education (variable  $y$ ). In this model it is assumed that the length of education of a teacher will affect the mathematics achievement of students. The hypothesis is that the longer the teacher's education, the better student achievement will be. As a result, the linear regression function graph will simply be an upward straight line. To obtain the equation and graph of the regression function, statistical data collection is then performed.

For this example, for example  $x$  stating the education factor with  $x = 0$  means never going to school and  $x = 16$  means having attended 16 years of school (graduating from undergraduate). Suppose student performance  $y$  is measured from 0 - 10 and a simple linear regression model is used.

The simple linear regression model is

$$\hat{y} = a + bx \quad (1)$$

with  $a, b$  regression coefficient which values are determined from the sample data pairs  $(x_i, y_i)$   $i = 1, 2, \dots, n$  and  $\hat{y}$  is prediction or estimation from  $y$ .

The occurring deviation values of data  $y_i$  from the average of  $\bar{y}$  result in errors called  $e_i$ . Therefore,

$$e_i = y_i - \bar{y} = y_i - a - bx_i \text{ such that}$$

$$y_i = a + bx_i + e_i; i = 1, 2, \dots, n. \tag{2}$$

The errors  $e_i$  contain random errors. The calculation of regression coefficient  $a$  and  $b$  with the OLS (Ordinary Least Square) method which requires the fulfillment of some assumptions in the data samples, that is, the data samples are considered from the normally distributed population, which means the random variable  $Y_i$  is distributed by  $N(\mu_i, \sigma^2)$  for each  $i$  and random error  $e_i$  is assumed distributing  $N(0, \sigma^2)$ .

Some assumptions / assumptions in data are needed in order to obtain a good regression equation with the least squares method:

1. Data considered to be from populations that are normally distributed with averages  $\mu_i$  and variances  $\sigma^2$ . This means, random variables  $Y_i$  are  $N(\mu_i, \sigma^2)$  distributed for each  $i$ . The symbol  $Y_i$  represents a random variable, whereas  $y_i$  it is only one of many values  $Y_i$  taken at random. Therefore for each distribution is normal, then

$$p(y_i|x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma}\right)^2\right),$$

$$-\infty < y_i < \infty.$$

The symbol  $p(y|x)$  represents the probability density function (pdf)  $y$  if  $x$ . known. Thus, one or more terokok will be taken from a population with normal distribution  $N(\mu_i, \sigma^2)$ . In population theory, Equation (1) can be written as

$$\hat{y} = a + bx \Leftrightarrow E(Y_i|x_i) = \alpha + \beta x_i$$

and Equation (2) becomes

$$y_i = \alpha + \beta x_i + \varepsilon_i, i = 1, 2, \dots, n.$$

In this case,  $\alpha, \beta$  it states the population parameters which are constants but whose magnitude is unknown, and are distributed random errors  $N(0, \sigma^2)$ .

2. Assume  $x_i$  that there is no distribution. In practice, the value  $x_i$  is often under the control of the

researcher, meaning that the value  $x_i$  is determined by the researcher.

As a result, calling the independent variable  $x_i$  less appropriate or mistaken because  $x_i$  it does not change freely. Sometimes,  $x_i$  is called a predictor variable while  $y_i$  is called a response variable. Because  $\varepsilon$  has a distribution while  $x$  not,  $Y$  has a normal distribution according to the distribution  $\varepsilon$  ie

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2)$$

The assumption that  $\text{Var}(\varepsilon_i)$  and  $\text{Var}(Y_i)$  are the same  $\sigma^2$  for each  $i$  means that the variance does not change in every data set. This assumption is important for the purpose of hypothesis testing which is solved by Test  $t$  and Test  $F$ .

Assumptions  $E(\varepsilon_i) = 0$  so that  $E(Y_i) = \alpha + \beta x_i$  for each  $i$  is not so important to be hypothesized because it can always be obtained by using translation.

### 3.1. Linear Regression Parameter Estimation with OLS Method

Equation (2) will be estimated by

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{3}$$

For example, there is a data pair  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ . The regression coefficient estimation  $\alpha$  and  $\beta$  with OLS method fulfills the minimum SSE (sum of square error), i.e.

$$SSE = \sum_{i=1}^n e_i = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

minimum

With OLS method, parameter  $\alpha$  estimated by  $a$  and parameter  $\beta$  is estiated by  $b$  with

$$a = \bar{y} - b\bar{x} = \frac{1}{n} \sum_{i=1}^n y_i - b \cdot \left(\frac{1}{n} \sum_{i=1}^n x_i\right) \tag{4}$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x}) \sum_{i=1}^n (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{5}$$

Example 1 (Cliff and Billy, 2017):

Data is given in Table 1 with the number of data pairs with  $n = 12$ .

**Table 1.** Data pair  $(x_i, y_i)$ 

| $i$   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $x_i$ | 6   | 3   | 6   | 9   | 3   | 9   | 6   | 3   | 9   | 6   | 3   | 9   |
| $y_i$ | 526 | 421 | 581 | 630 | 412 | 560 | 434 | 443 | 590 | 570 | 346 | 672 |

The solution with OLS method using equation (5) and (4) results in the values of  $b = 34.5833$  and  $a = 307.9167$  that the regression equation with OLS method is

$$\hat{y}_{i:(OLS)} = 307,9167 + 34,5833x_i ; \quad (6)$$

$$i = 1,2,\dots,12.$$

### 3.2. Linear Regression Parameter Estimation with Simple Averaging Method

Cliff and Billy (2017) have developed a new method in solving the simple and multiple linear regression equation called SAM (Simple Averaging Method). According to Cliff and Billy (2017), with SAM, parameter  $\alpha$  and  $\beta$  are calculated with

$$\hat{\beta}_i \approx \bar{\beta} = b = \frac{1}{n-1} \sum_{i=2}^n \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \quad (7)$$

$$\alpha \approx a = \frac{1}{n-1} \sum_{i=2}^n \left( y_i - \sum_{i=2}^n \bar{\beta} \cdot x_i \right) \quad (8)$$

The solution with a simple average method may result in values  $b$  and  $a$  calculated with equation (7) and (8) respectively by  $b = 33.1667$  and  $a = 315.4545$  that the simple linear regression equation is

$$\hat{y}_{i:(SAM)} = 315,4545 + 33,1667x_i ; \quad (9)$$

$$i = 2,\dots,12.$$

By comparing the values of RMSE and MAE for both methods, Cliff and Billy (2017) conclude that SAM (Simple Averaging Method) may be used to solve the regression equation. The values of RMSE in OLS method and SAM are respectively 47.1367 and 47.2740. Meanwhile, the values of MAE in both OLS method and SAM are respectively 39.2362 and 39.3409. The advantage of SAM is free from all assumptions required by regression, such as error normality assumption while the data should be from the normal distribution.

### 3.3. Similar Formula in SAM

This part will present two equations which may be used

to estimate parameter  $\alpha$  using SAM. First is equation (10) and second is equation (11).

$$\alpha \approx a = \frac{1}{n-1} \left( \sum_{i=2}^n (y_i) - A \right) ;$$

$$\text{where } A = \frac{1}{n-1} \sum_{i=2}^n \beta_i \cdot x_i = \frac{\beta}{n-1} \sum_{i=2}^n x_i = \bar{\beta} \sum_{i=2}^n x_i \quad (10)$$

then

$$\alpha \approx a = \frac{1}{n-1} \left( \sum_{i=2}^n y_i - \bar{\beta} \cdot \sum_{i=2}^n x_i \right) \quad (11)$$

Equation (10) and (11) may be used as alternatives from equation (8). The following is the calculation example of parameter  $\alpha$  using equation (10) and (11). The data used is that shown in Table 1. In this case, parameter  $\beta$  is still calculated using equation (7) and parameter  $\alpha$  calculated using equation (10) presented in Table 2.

The obtained regression equation is

$$\hat{y}_{i:(SAM)} = 315,4545 + 33,1667x_i \quad (12)$$

$$i = 2,\dots,12.$$

Equation (12) is the same with the regression equation obtained using SAM from Cliff and Billy (2017). See equation (9). The calculation of simple linear regression equation with equation (11) is given to the readers to make.

### 3.4. SAM Simplification

The estimation for  $\alpha$  given by Cliff and Billy (2017) in equation (8) may be stated in a more attractive form, that is,

$$a = \bar{Y} - \bar{\beta} \cdot \bar{X}.$$

#### Theorem 1:

The estimation of parameter  $\alpha$  for SAM in equation (8) may be stated in the form of  $a = \bar{Y} - b \cdot \bar{X}$ .

#### Proof:

Will be shown that

**Table 2.** Calculation of Regression Parameters using SAM Method

| $i$ | $x_i$ | $y_i$ | $\beta_i = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$ | $\bar{\beta}$ | $\bar{\beta} \cdot x_i$ | $A = \frac{1}{n-1} \sum_{i=2}^n \beta_i \cdot x_i$ | $y_i - A$ | $\sum_{i=2}^n (y_i) - A$ | $\frac{1}{n-1} \left( \sum_{i=2}^n (y_i) - A \right)$ |
|-----|-------|-------|---|---------------|-------------------------|--|-----------|--------------------------|---|
| 1   | 6     | 526   | -   |               | -                       | -  | -         | -                        | -   |
| 2   | 3     | 421   | 35,0000   | 33,1667       | 99,5                    | 199  | 222       | 3470                     | 315,4545  |
| 3   | 6     | 581   | 53,3333   |               | 199,0                   |  | 382       |                          |   |
| 4   | 9     | 630   | 16,3333   |               | 298,5                   |  | 431       |                          |   |
| 5   | 3     | 412   | 36,3333   |               | 99,5                    |  | 213       |                          |   |
| 6   | 9     | 560   | 24,6667   |               | 298,5                   |  | 361       |                          |   |
| 7   | 6     | 434   | 42,0000   |               | 199,0                   |  | 235       |                          |   |
| 8   | 3     | 443   | -3,0000   |               | 99,5                    |  | 244       |                          |   |
| 9   | 9     | 590   | 24,50000  |               | 298,5                   |  | 391       |                          |   |
| 10  | 6     | 570   | 6,66667   |               | 199,0                   |  | 371       |                          |   |
| 11  | 3     | 346   | 74,6667   |               | 99,5                    |  | 147       |                          |   |
| 12  | 9     | 672   | 54,3333   |               | 298,5                   |  | 473       |                          |   |

$$a = \frac{1}{n-1} \sum_{i=2}^n \left( y_i - \sum_{i=2}^n \bar{\beta} \cdot x_i \right) = \bar{Y} - b \cdot \bar{X}$$

$$\text{with } \bar{Y} = \frac{1}{n-1} \sum_{i=2}^n y_i \text{ and } \bar{X} = \frac{1}{n-1} \sum_{i=2}^n x_i$$

From equation (8),

$$\begin{aligned} \alpha \approx a &= \frac{1}{n-1} \sum_{i=2}^n \left( y_i - \sum_{i=2}^n \bar{\beta} \cdot x_i \right) \\ &= \frac{1}{n-1} \sum_{i=2}^n \left( y_i - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \right) \cdot x_i \right) \\ &= \frac{1}{n-1} \left( \left( y_2 - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \right) \cdot x_2 \right) + \right. \\ &\quad \left. \left( y_3 - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \right) \cdot x_3 \right) + \right. \\ &\quad \left. \dots + \left( y_n - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \right) \cdot x_n \right) \right) \\ &= \frac{1}{n-1} \left( (y_2 + y_3 + \dots + y_n) - \right. \\ &\quad \left. \left( \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \right) \cdot (x_2 + x_3 + \dots + x_n) \right) \right) \\ \alpha \approx a &= \frac{1}{n-1} \sum_{i=2}^n y_i - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \right) \cdot \left( \frac{1}{n-1} \sum_{i=2}^n x_i \right) \\ &= \bar{Y} - \bar{\beta} \cdot \bar{X} \end{aligned}$$

Since the estimation for  $\beta$  is

$$\hat{\beta} \approx \bar{\beta} = b = \frac{1}{n-1} \sum_{i=2}^n \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$$

the equation (8) may be written to

$$a = \bar{Y} - b \cdot \bar{X}$$

This result may cause the estimation of parameter  $\beta$  that is  $\hat{\beta} \approx \bar{\beta} = b$  first calculated. The estimator  $\alpha$  obtained using SAM is similar with the result of estimation  $\alpha$  using OLS method, that is, the average of both variables constructing its liner regression equation.

The average of both simple methods calculated to obtain estimation  $\hat{\beta}$  is the same that is, using equation (7). However, the calculation to obtain the estimation  $\alpha$  has a difference. In the first method, estimation  $\alpha$  is calculated with equation (9), while in table 2, it is calculated with equation (9).

Furthermore, it will be shown that equation (10) and (11) may be created into equation in Theorem 1 that equation (10) and (11) are identical with equation (8).

**Theorem 2:**

The estimation of parameter  $\alpha$  for SAM in equation (10) may be stated in the form  $a = \bar{Y} - b \cdot \bar{X}$ .

**Proof:**

Will be shown that  $a = \frac{1}{n-1} \left( \sum_{i=2}^n (y_i) - A \right)$ , with

$$A = \frac{1}{n-1} \sum_{i=2}^n \beta_i \cdot x_i \text{ may be changed into } a = \bar{Y} - \bar{\beta} \cdot \bar{X}$$

$$\text{with } \bar{Y} = \frac{1}{n-1} \sum_{i=2}^n y_i \quad \bar{\beta} = \frac{1}{n-1} \sum_{i=2}^n \beta_i \quad \text{and}$$

$$\bar{X} = \frac{1}{n-1} \sum_{i=2}^n x_i$$

$$1. \text{ Write } \beta_i = \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \quad ; \text{ for } i = 2, \dots, n$$

$$2. \text{ Calculate the average of } \beta_i \text{ , i.e. } \bar{\beta} = \frac{1}{n-1} \sum_{i=2}^n \beta_i$$

$$3. \text{ Write } \bar{\beta} \cdot x_i \quad ; \text{ for } i = 2, \dots, n$$

$$4. \text{ Calculate sum of } \bar{\beta} \cdot x_i \text{ and write as } A, \text{ i.e.}$$

$$A = \sum_{i=2}^n \bar{\beta} \cdot x_i = \bar{\beta} \cdot \sum_{i=2}^n x_i$$

$$5. \text{ Write } y_i - A \quad ; \text{ for } i = 2, \dots, n$$

$$6. \text{ Calculate } \left( \sum_{i=2}^n y_i \right) - A$$

$$\left( \sum_{i=2}^n y_i \right) - A = (y_2 + y_3 + \dots + y_n) - A$$

$$= (y_2 + y_3 + \dots + y_n) - \left( \bar{\beta} \cdot \sum_{i=2}^n x_i \right)$$

$$= (y_2 + y_3 + \dots + y_n) - \left( \bar{\beta} \cdot (x_2 + x_3 + \dots + x_n) \right)$$

$$7. \text{ Calculate}$$

$$\begin{aligned} \frac{1}{n-1} \left( \left( \sum_{i=2}^n y_i \right) - A \right) &= \frac{1}{n-1} \left( \left( \sum_{i=2}^n y_i \right) - \left( \bar{\beta} \cdot \sum_{i=2}^n x_i \right) \right) \\ &= \frac{1}{n-1} \left( \left( \sum_{i=2}^n y_i \right) - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \cdot \sum_{i=2}^n x_i \right) \right) \\ &= \frac{1}{n-1} \left( \sum_{i=2}^n y_i \right) - \left( \frac{1}{n-1} \right)^2 \sum_{i=2}^n \beta_i \cdot \sum_{i=2}^n x_i \\ &= \frac{1}{n-1} \left( \sum_{i=2}^n y_i \right) - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \right) \cdot \left( \frac{1}{n-1} \sum_{i=2}^n x_i \right) \\ &= \bar{Y} - \bar{\beta} \cdot \bar{X} \end{aligned}$$

So, it is proven that

$$a = \frac{1}{n-1} \left( \sum_{i=2}^n (y_i) - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \cdot x_i \right) \right) \text{ may be stated}$$

in the form  $a = \bar{Y} - b \cdot \bar{X}$  with  $\bar{Y} = \frac{1}{n-1} \sum_{i=2}^n y_i$  and

$$\bar{X} = \frac{1}{n-1} \sum_{i=2}^n x_i$$

### Theorem 3:

The estimation of parameter  $\alpha$  for SAM in equation

(11) may be stated in the form  $a = \bar{Y} - b \cdot \bar{X}$ .

### Proof:

To show that equation (9) is identical with equation (8), adequately by showing that the equation may be stated in the form of equation in Theorem 1. From equation (11), it is proven that

$$\begin{aligned} a &= \frac{1}{n-1} \left( \sum_{i=2}^n y_i - \bar{\beta} \cdot \sum_{i=2}^n x_i \right) \\ &= \frac{1}{n-1} \left( \sum_{i=2}^n y_i - \frac{1}{n-1} \sum_{i=2}^n \beta_i \cdot \sum_{i=2}^n x_i \right) \\ &= \frac{1}{n-1} \sum_{i=2}^n y_i - \left( \frac{1}{n-1} \sum_{i=2}^n \beta_i \right) \cdot \left( \frac{1}{n-1} \sum_{i=2}^n x_i \right) \\ &= \bar{Y} - \bar{\beta} \cdot \bar{X} \end{aligned}$$

From equation (11), it is proven that

$a = \frac{1}{n-1} \left( \sum_{i=2}^n y_i - \bar{\beta} \cdot \sum_{i=2}^n x_i \right)$  may be stated in the form

$$a = \bar{Y} - b \cdot \bar{X} \quad \text{with} \quad \bar{Y} = \frac{1}{n-1} \sum_{i=2}^n y_i \quad \text{and}$$

$$\bar{X} = \frac{1}{n-1} \sum_{i=2}^n x_i$$

## 4. Conclusion

Cliff and Billy (2017) have given a simple linear regression parameter estimation method in the equation (7) and (8) known as SAM. The method may be used to obtain a simple linear regression equation with a high accuracy. For the example given, the RMSE values in OLS and SAM method are respectively 47.1367 and 47.2740, while the MAE values are respectively 39.2362 and 39.3409. Therefore, SAM may become one alternative for OLS method. The advantage of SAM is not requiring any condition related to its distribution of errors.

The research conclusions are:

1. The estimation of parameter  $\alpha$  in SAM from Cliff and Billy may be stated in the form  $a = \bar{Y} - b \cdot \bar{X}$  (Theorem 1).

2. The simple linear regression model parameter is  $\hat{y} = a + bx$  calculated using theorem 1 with

$$b = \frac{1}{n-1} \sum_{i=2}^n \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \quad , \quad \bar{Y} = \frac{1}{n-1} \sum_{i=2}^n y_i \quad \text{and}$$

$$\bar{X} = \frac{1}{n-1} \sum_{i=2}^n x_i \quad . \text{ In this case, coefficient } b$$

should be initially calculated. Theorem 2 and 3 prove

that the estimation of parameter  $\alpha$  in equation (9) is identical with equation (11) and (12).

## Acknowledgements

The authors would like to thank Jenderal Soedirman University (UNSOED) and the Ministry of Research, Technology and High Education of Republic of Indonesia. This work was supported by BLU UNSOED No. P/419/UN23/14/PN/2019, date on March, 29 2019, via Improvement Competency Research.

---

## REFERENCES

- [1] Chioma, N. J. (2009). Causal Relationship between Gross Domestic Product and Personal Consumption Expenditure of Nigeria. *African Journal of Mathematics and Computer Science Research*. Vol. 2, No. 8, pp. 179-183.
- [2] Cliff, K. R. and Billy, K. M. (2017). Estimation of the Parameters of Linear Regression System using the Simple Averaging Method. *Global Journal of Pure and Applied Mathematics*. Vol. 13, No. 11, pp. 7749-7758.
- [3] Gustiarisanie, A., Rahardjo, M. F. and Ernawati, Y. (2016). Hubungan Panjang-Bobot dan Faktor Kondisi Ikan Lidah Cynoglossus cynoglossus, Hamilton1882 (Pisces: Cynoglosiade) di Teluk Pabean Indramayu, Jawa Barat (Length-Weight Relationship and Condition Factor of Tonguesole Cynoglossus cynoglossus, Hamilton1882 (Pisces: Cynoglosiade) in Pabean Bay, Indramayu, West Java. *Jurnal Iktiologi Indonesia*. Vol. 16, No. 3, pp. 337-344.
- [4] Jain, S., Chourse, S., Dubey, S., Jain, S., Kamakoty, J., and Jain, D (2016). Regression Analysis – Its Formulation and Execution in Dentistry. *Journal of Applied Dental and Medical Science*. Vol. 2, No. 1, pp. 199-208.
- [5] Permai, S. D. and Tanty, H. (2018). Linear Regression Model using Bayesian Approach for Energy Performance of Residential Building. *Procedia Computer Science*. Vol. 135, pp 671-677.
- [6] Prabitha, J. and Archana, S. (2015). Application of Regression Analysis in Numeroustimes. *International Journal of Science, Enginering and Technology Research*. Vol. 4, Issue 4, pp. 1002-1005.
- [7] Sembiring, R. K. (2003). *Analisis Regresi*. Bandung: Penerbit ITB.
- [8] Singh, A. S. and Masuku, M. B. (2013). Application of Modeling and Statistical Regression Techniques in Research. *Research Journal of Mathematical and Statistical Sciences*. Vol. 1, No. 16, pp. 14-20.