Identification of Factors Affecting the Mathematics Learning Difficulties of Students in Bandung Using a Linear Regression Model

Rahmi Wiganda Elastika
Master Study Program of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, INDONESIA.
elastikarahmi@gmail.com

Sukono, Stanley Pandu Dewanto
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, INDONESIA.
sukono@unpad.ac.id; stanleypd@bdg.centrin.net.id

Abdul Talib Bon
Department of Production and Operations, University Tun Hussein Onn Malaysia, MALAYSIA
talibon@gmail.com

Abstract

Learning difficulties experienced by students are basically a reflection of their learning efforts. In general, the less learning effort you have, the greater the learning difficulties you will experience. Many factors affect learning difficulties, including interests, motivation, parental support, parental income levels, and so on. This paper aims to analyze the difficulties in learning mathematics among students in Bandung. Factors that are thought to influence mathematics learning difficulties include interest, motivation, parental support, parents' income levels, and the number of dependent children of the parents. The population in this study is students from various universities in the city of Bandung. Data were collected by random sampling techniques, using a questionnaire to 100 respondents. The data analysis technique used descriptive statistical analysis and inferential statistical analysis. The results of inferential statistical analysis show that interest (X1), learning motivation (X2), parental support (X3), parents' income level (X4), and the number of dependent children (X5) together have an effect on mathematics learning difficulties. Student (Y) in the city of Bandung, which is indicated by the F-count value of 11.275 is greater than the F-statistic value of 2.30. The pattern of influence between these six variables is expressed by the multiple regression equation.

Keywords:
Learning difficulties, influencing factors, sampling techniques, inferential statistics, linear regression.

1. Introduction

Mathematics is the study of quantity, structure, space and change. Mathematicians assemble and use various patterns, then use them to formulate new conjectures, and construct truth through a method of rigorous deduction derived from axioms and corresponding definitions (Frieder et al., 2014; Juan et al., 2018). There is a debate whether mathematical objects such as numbers and points already exist in the universe, or have they been discovered and created by humans. Through the use of logical reasoning and abstraction, mathematics develops from counting, calculating, measuring, and systematically examining the shapes and movements of physical objects (Knifong et al., 1976; Ken, 1980). Practical mathematics is manifested in human activities since written records exist. Rigorous mathematical arguments first appear in Greek mathematics, especially in the work of Euclid. The importance of learning mathematics cannot be separated from its role in various aspects of life. In addition, by studying mathematics, a person is accustomed to thinking systematically, scientifically, using logic, critically, and can increase his or her creativity (Jenny et al., 2018; McLaren, 2015). Research shows that there are obstacles to
learning mathematics which can be categorized into internal and external factors, including lazy students to learn, an inaccurate view of mathematics, unfavorable classroom conditions and bad influence from the surrounding environment (Brenda et al., 2013; Elastika et al., 2019).

Research on mathematics learning difficulties, among others, by Puspitasari et al. (2017), conducted research that aims to determine student barriers in learning linear program courses. Analysis to determine student errors in solving simplek method questions. This error analysis refers to the types of errors put forward by the researcher. This research method is descriptive qualitative. The study population was all mathematics education students of STKIP Siliwangi who took linear program courses. The solutions to overcome these obstacles are through structured exercises or drill questions and provide various questions as material for practice. Lecturers try to use a different method every meeting in class so that students don't get bored easily and practice thinking solutions to linear program problems.

Nugraheni (2017), conducted research with the aim of uncovering things that were considered learning difficulties in the mechanics material. This research is a qualitative descriptive study that seeks to describe the types of student difficulties in solving mechanical problems. The subjects in this study were students of the third semester of Science Education at IKIP Veteran Central Java who took Mechanics courses. In this study, researchers collected information through diagnostic tests and interviews with students. The instruments used in this study were diagnostic tests and interview guides. The location of student learning difficulties is seen based on the mistakes made in completing the diagnostic test. The results of the analysis show that the learning difficulty in studying mechanics lies in the basic mathematical abilities, namely differential and integral. Thus, teachers need to develop learning strategies so that students' thinking skills in solving problems increase. Based on the description of the problem above, this study intends to carry out identification of factors affecting the mathematics learning difficulties of students in Bandung. The analysis was performed using a linear regression model. The research objective is to measure the factors that significantly affect the difficulty of learning mathematics, so that the learning system policy can be changed according to the information of this study.

2. Model Regresi Linier
This section discusses methods that include regression models, parameter estimation methods, and goodness of fit tests.

2.1 Simple Linear Regression Model
The actual simple linear regression model can be written as follows:

\[ Y = b_0 + b_1 X + e, \]  

where \( Y \) is the dependent variable (regression), \( X \) is the independent variable (regressor), \( b_0 \) is the intercept parameter (constant), \( b_1 \) is the coefficient parameter (slope), and \( e \) is the residual. Equation (1) can be estimated with the following equation:

\[ \hat{Y} = b_0 + b_1 X, \]

so this equation is referred to as a simple regression estimator (Sheryl et al., 2018; Williams et al., 2016).

2.2 General Multiple Linear Regression Model
Multiple linear regression models in general can be written as follows:

\[ Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k + e. \]  

Where \( Y \) is the dependent variable (regression), \( X \) is the independent variable (regressor), \( b_0 \) is the intercept parameter (constant), \( b_1, b_2, \ldots, b_k \) is the coefficient parameter (slope), and \( e \) residual. The multiple linear regression equation (2), has an estimator equation which is written as follows:

\[ \hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k. \]  

so this equation is called a multiple linear regression estimator (Paul et al., 2018; Beran et al., 2018).

2.3 Parameter Estimation Method
This section discusses the method of estimating the parameters of multiple linear regression with a matrix approach. Using the matrix equation approach, the multiple linear regression equation (3) can be written as follows:

\[ Y = XB + e. \]  

© IEOM Society International
Where:

\[
Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & X_{12} & X_{13} & \cdots & X_{1k} \\ 1 & X_{22} & X_{23} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n2} & X_{n3} & \cdots & X_{nk} \end{bmatrix}; \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}; \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix},
\]

Where the \( Y \) matrix is \((n \times 1)\), \( X \) matrix is \((n \times k)\), \( B \) matrix is \((1 \times k)\), and \( e \) matrix is \((1 \times k)\).

To determine the estimator value for the matrix parameter \( B \), it is carried out by the least squares method in which the number of squares of the residual must be minimized, namely:

\[
\text{Minimization} \sum e_i^2 = e^T e = (Y - XB)^T (Y - XB).
\]  

(5)

Where \( e^T = (Y - XB)^T \) transpose from \( e \). Because \( B^T X^T Y \) is a scalar, so it is the same as the transpose, which is \( Y^T X B \). For the minimization process, equation (5) is determined as follows:

\[
\frac{\partial \sum e_i^2}{\partial B} = -2X^T Y + 2X^T XB = 0.
\]  

(6)

From equation (6), the following parameter estimators can be obtained:

\[
B = (X^T X)^{-1} X^T Y.
\]  

(7)

Where \((X^T X)^{-1}\) is the inverse of \((X^T X)\) (Sukono et al., 2016).

This approach can be used if \((X^T X)\) has an inverse matrix, but if there is multicollinearity, the inverse matrix calculation is incorrect.

### 2.4 Goodness of Fit Test

The goodness of fit test aims to ensure that the model is able to represent the actual data. The goodness of fit test of the parameter estimator is carried out using the individual significance test, the simultaneous significance test, the residual normality assumption test, and the coefficient of determination test.

**Individual significance test**

The individual significance test aims to test each parameter of \( \beta_i \) \((i = 0,1,2,...,k)\), where \( \beta_i \in \{ b_0, b_1, b_2, ..., b_k \} \) from equation (3), in influencing the dependent variable. To test the \( \beta_i \) parameters, the established hypotheses are \( H_0: \beta_i = 0 \) and \( H_1: \beta_i \neq 0 \). The test is carried out using the \( t_{stat} \) statistic, with equations:

\[
t_{Stat} = \frac{\beta_i}{SE(\beta_i)},
\]

(8)

where \( SE(\beta_i) \) is the standard error of parameter \( \beta_i \).

The criterion rejects the hypothesis \( H_0 \) if \( |t_{Stat}| > t_{(n-2,1-\alpha/2)} \), or \( Pr[t_{Stat}] < \alpha \), where \( t_{(n-2,1-\alpha/2)} \) is the critical value of the distribution-\( t \) at a significance level of \( 100(1-\alpha)\% \) and \( n \) number of data (Sukono et al., 2016).

**Simultaneous significance test**

Simultaneous significance test aims to test simultaneously the parameter \( \beta_i \) \((i = 0,1,2,...,n)\), where \( \beta_i \in \{ b_0, b_1, b_2, ..., b_k \} \) from equation (4), in influencing the dependent variable. The hypothesis is \( H_0: b_0 = b_1 = b_2 = ... = b_k = 0 \) and \( H_1: \exists b_0 \neq b_1 \neq b_2 \neq ... \neq b_k \neq 0 \). The test is carried out using statistical \( F \), with the equation:

\[
F_{stat} = \frac{\text{MSReg}}{\text{MSError}},
\]

(9)
where $MS_{\text{Reg}}$ mean square due to regression, and $MS_{\text{Error}}$ mean square due to residual variation.

The criterion rejects the hypothesis $H_0$ if $F_{\text{Stat}} > F_{(1,n-2,1-\alpha)}$, or $\Pr[F_{\text{Stat}}] \alpha$, where $F_{(1,n-2,1-\alpha)}$ is the critical value of the distribution- $F$ at a significance level of $100(1-\alpha)$% and $n$ number of data (Sukono et al., 2016).

**Residual normality test**

The normality test aims to see that the residual data spreads normally. The normality test can be performed using the Kolmogorov-Smirnov (KS) statistic. The hypothesis set is that $H_0$ residual data is normally distributed, and $H_1$ residual data is not normally distributed. The test is done by calculating the residual standard deviation using the equation:

$$S_{e_i} = \sqrt{\frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{n-1}}. \quad (10)$$

Then the $e_i$ value is transformed into $z_i$ with the $z_i = (e_i - \bar{e}) / S_{e_i}$ equation. The determination of the $P(z_i)$ probability value is carried out based on the standard normal distribution table. Meanwhile, the $S(z_i)$ probability is determined using the $S(z_i) = \text{randl}(z_i)/n$ equation. Next, the $|S(z_i) - P(z_i)|$ absolute difference value is calculated. Kolmogorov-Smirnov $KS_{\text{Stat}}$ statistics are calculated using the equation:

$$KS_{\text{Stat}} = \max |S(z_i) - P(z_i)|. \quad (11)$$

Next, determine the critical value of the $KS_{(\alpha,n-1)}$ statistic, at the level of significance $\alpha = 0.05$. The test criterion is reject $H_0$ if $KS_{\text{Stat}} > KS_{(\alpha,n-1)}$ (Sukono et al., 2014; 2016).

**The coefficient of determination**

Referring to Sidi et al. (2017; 2018), the coefficient of determination $R^2$ is a measure of how much diversity the independent variable has on the dependent variable, based on the strength of the relationship. Therefore, the coefficient of determination determines the level of ability or influence of the independent variable $X_i$ ($i = 1,2,...,k$) in influencing the dependent variable $Y$. The coefficient of determination $R^2$ is determined using the equation:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}. \quad (12)$$

The value of the coefficient of determination is between 0 and 1. A small determination value close to 0 indicates that the variation of the independent variable is very weak, and a value close to 1 indicates that the variation of the independent variable provides all the information needed to predict the dependent variable (Sukono et al., 2014; 2016).

3. Results and Discussion

3.1 Data Analyzed

The factors that are thought to affect the difficulty of learning mathematics and will be analysed include: student interest, student motivation, parental support, parental income levels, and the number of dependents of children from parents. The population in this study were students from various universities who live in Bandung, West Java, Indonesia. Data collection was carried out by random sampling technique, using a questionnaire to 100 student respondents. The data analysis technique was performed using descriptive statistical methods and inferential statistical methods.
Furthermore, for these 5 factors, descriptive statistical testing was carried out using the overall mean square analysis method. Based on the test results, it shows that student interest, student motivation, parental support, parental income levels, and the number of dependents of children from parents, which happened as a whole was good according to the respondent's perception. Thus, this data set can be used for analysis using a multiple linear regression analysis approach.

For the purposes of analyzing multiple linear regression models, the data needs to be tested for normality before further use. Testing for normality in the analysis of multiple linear regression models aims to ensure that the data distribution is normally distributed. Because data is good for analysis in multiple linear regression models, it is data that is patterned following a normal distribution. Data normality testing is carried out referring to equations (14) and (15) with the help of SPSS version 17.0 software. The results show that the data has been normally distributed, so that this data set can be used in estimating the parameters of multiple linear regression models.

3.2 Mengestimasi Parameter Model Regresi Linier Berganda

In this sub-section, the estimation of multiple linear regression parameters is carried out in order to obtain a vector estimator of the coefficient parameter \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k) \) which gives the smallest value in equation (9). The estimation process is carried out using the least squares approach, with the help of SPSS version 17.0 software. The estimation results of the parameter coefficients, standard error values, and ratio values, of each parameter coefficient estimator are presented in Table 1.

<table>
<thead>
<tr>
<th>Coefficient Parameter of Variables ((X_i))</th>
<th>Estimator of Parameter ((\hat{\beta}_i))</th>
<th>Error Standard (SE(\hat{\beta}_i))</th>
<th>Ratio ((\hat{Z})) (\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)})</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.2531</td>
<td>0.02412</td>
<td>51.95273632</td>
<td>Significance</td>
</tr>
<tr>
<td>Interest ((X_1))</td>
<td>0.0132</td>
<td>0.0013</td>
<td>10.15384615</td>
<td>Significance</td>
</tr>
<tr>
<td>Motivation ((X_2))</td>
<td>0.0163</td>
<td>0.0027</td>
<td>6.037037037</td>
<td>Significance</td>
</tr>
<tr>
<td>Parental Support ((X_3))</td>
<td>0.0051</td>
<td>0.0021</td>
<td>2.428571429</td>
<td>Significance</td>
</tr>
<tr>
<td>Parents’ Income ((X_4))</td>
<td>0.0074</td>
<td>0.0036</td>
<td>2.055555556</td>
<td>Significance</td>
</tr>
<tr>
<td>Number of dependents of parents ((X_5))</td>
<td>0.0018</td>
<td>0.0007</td>
<td>2.571428571</td>
<td>Significance</td>
</tr>
</tbody>
</table>

The individual significance testing of the estimator of the estimated coefficient parameters is carried out in order to analyze the significance level of each coefficient parameter of the multiple linear regression model obtained from the research results. In this individual significance test, the hypothesis is as follows:

\[ H_0: \text{There is no difference between each parameter estimator of the coefficient of the multiple linear regression model with the observations obtained;} \]

\[ H_1: \text{There is a difference between each parameter estimator of the coefficient of the multiple linear regression model and the observations obtained.} \]

The statistical test of individual significance can be done by referring to equation (12). The results of individual significance testing of each parameter coefficient estimator are given in Table 1, Ratio column. When the level of significance \( \alpha = 0.05 \) is set, referring to the decision criteria in equation (12), from the distribution table \(-t\) the value of \( t_{stat} = -1.986086317 \) is obtained. Therefore each coefficient parameter estimator produces \( |t_{stat}| = |t(n-2, \frac{1}{2}\alpha)| \), it is clear that the hypothesis \( H_0 \) is rejected, which means that "each parameter estimator coefficient of the multiple linear regression model is significant".

Next is the simultaneous significance test, to the estimator of the coefficient parameter it is necessary to test the significance of the overall influence of the predictor variable on the response variable \( Y \). In this test the hypothesis set is \( H_0: \hat{\beta}_0 = \hat{\beta}_1 = ... = \hat{\beta}_k = 0 \), against the alternative hypothesis \( H_1: \exists \hat{\beta}_0 \neq \hat{\beta}_1 \neq ... \neq \hat{\beta}_k \neq 0 \) \((k = 0,1, ..., 5)\). The testing process is carried out using a test referring to equation (13). The test results are also presented at the bottom of Table 1, where it is found that \( F_{count} = 11.72226249994748 \). When the significance level is set at \( \alpha = \)
0.05, from the distribution table the value \( F_{\text{statistic}} = 2.30 \), so it is clear that \( F_{\text{count}} > F_{\text{statistic}} \). Therefore, the hypothesis \( H_0 \) is rejected, which means that "the simultaneous estimator of the coefficient parameter of the multiple linear regression model is significant".

Next is the residual normality assumption test, intended to ensure that the residuals follow a normal distribution. Testing is done by referring to equations (14) and (15). Based on the test results using SPSS version 17.0 software, it shows that the residual \( \varepsilon \sim N(0.00003, 5.0624) \).

The next step is to measure the strength of the correlation between the predictor variables and the response variables. This measurement can be done based on the magnitude of the \( R^2 \) statistical value, which can be calculated using equation (16). In this study, the linear regression equation estimator resulted from \( R^2 = 0.8987691 \). This shows that the predictor variables: student interest (\( X_1 \)), student motivation (\( X_2 \)), parental support (\( X_3 \)), income level of parents (\( X_4 \)), and the number of dependents of children from parents (\( X_5 \)), have correlations very strong with the response variable \( Y \). It also means that the predictor variables are 0.8987691 which can explain the response variable, and 0.1012309 is explained by other variables. Based on the results of the estimation analysis presented in Table 1 and referring to equation (6), the multiple linear regression model has the following equation:

\[
Y = 1.2531 + 0.0132X_1 + 0.0163X_2 + 0.0051X_3 + 0.0074X_4 + 0.0018X_5 + \varepsilon ,
\]

and the estimator of the multiple linear regression model has the following equation:

\[
\hat{Y} = 1.2531 + 0.0132X_1 + 0.0163X_2 + 0.0051X_3 + 0.0074X_4 + 0.0018X_5 .
\]

The multiple linear regression estimator equation (18) represents how and how much each predictor variable influences the response variable or learning difficulties for students in the city of Bandung.

4. Kesimpulan
This paper has identified the factors that affect the learning difficulty of students in the city of Bandung. Based on the data and research findings on the factors that cause learning difficulties experienced by students in constructivist learning mathematics courses in the city of Bandung, it can be concluded that the factors include: student interest (\( X_1 \)), student motivation (\( X_2 \)), parental support (\( X_3 \)), The level of parents’ income (\( X_4 \)), and the number of dependents of the parents (\( X_5 \)), significantly affect the difficulty of learning mathematics for students in Bandung. This effect can also be explained by a very strong correlation with the \( Y \) response variable, namely 0.8987691, and 0.1012309 explained by other unknown variables.

Acknowledgments
Acknowledgments are conveyed to the Director of General of Higher Education of the Republic of Indonesia (Deputy for Strengthening and Development of the Ministry of Research and Technology/National Research and Innovation Agency), and Chancellor, Director of the Directorate of Research, Community Engagement, and Innovation, and the Dean of the Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, who have provided the Master with the Contract Number: 1827/UN6.3.1/LT/2020. Thesis Research Grant. This grant is intended to support the implementation of research and publication of master students.

References
Ashley Castleberry.,Amanda Nolen.,Thematic analysis of qualitative research data: Is it as easy as it sounds?,Currents in Pharmac Teaching and Learning, Volume 10, Issue 6, pp. 807-815, 2018.


Williams B., and Davis S., Maths anxiety and medication dosage calculation errors. *Nurse Education Pract.*, pp. 139-146. 2016.

**Biographies**

**Rahmi Wiganda Elastika** is a Master Student in Mathematics at Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, in research field of mathematics education.
Sukono is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. Currently as Chair of the Research Collaboration Community (RCC), the field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Stanley Pandu Dewanto is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. The field of applied mathematics, with a field of concentration of mathematics modeling.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He’s bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.