Predict the ability of students to conduct preliminary analysis using reverse and inverse regression

Helma
Mathematics Department, FMIPA, Universitas Negeri Padang, Indonesia
helma667@yahoo.co.id

Abstract. Real analysis is one of the courses to help the students gain experience in the type of critical thought. Students must understand the definitions and theorems to solve problems and perform an analysis process of thinking, so that student analysis develops to the given problem. The research that has been done shows that students' ability to perform the analysis can be improved. But, not to be examined whether the ability to perform the preliminary analysis can be seen from the learning outcomes and how its results reflect the student's ability to perform it. The problem is "Does the student's ability to perform an analysis of the given problem can be predicted based on the learning outcomes that obtained?". In this research used flow proof in performing a preliminary analysis. After theoretically analyzed the comparison between reverse regression and inverse regression, then established models and prediction interval of the students' ability to perform the analysis. The result of this research is the ability of students to conduct an analysis can be predicted from their learning outcomes.

1. Introduction
Mathematics is the science of organized structures. There are concepts, definitions, formulas, laws, postulates, statements, theorems and others in it [1]. If it is associated with Bloom's taxonomy, understanding the mathematical process includes calculating, formulating, symbolizing, interpreting and extrapolating.

There are three types of understanding [1], namely 1) translation, for example changing the matter in the form of words into symbols, and calculating the length of the third side of a right triangle if the length of the other two sides is known, calculating the logarithm of a number with a base 10, 2) interpretation, for example to mean an equation, explaining the difference between the proposition and axiom, 3) extrapolating, for example, making estimates or propensities from diagram.

The ability of mathematical understanding consists of four stages [2], that are 1) mechanical understanding, namely remembering and applying formulas routinely and performing simple calculations, 2) inductive understanding, namely applying formulas or concepts in simple or similar cases, 3) rational understanding, namely proving the truth of a formula or a theorem, 4) intuitive understanding, namely estimating the truth with certainty (without hesitation) before carrying out further analysis.

One of the subjects that is needed by students who are candidates for mathematics and mathematics education to form and train students' rational and intuitive understanding is Real Analysis. The competence that must be achieved after studying this course [3] is "to help the reader gain experience in the type of critical thought that is used in this deductive process".
Students must understand the definitions and theorems to solve problems in Real Analysis and carry out a process of thinking analysis which is often referred to as preliminary analysis. After the preliminary analysis is done, all students construct proof based on the preliminary analysis.

They can develop an analysis of the problems given based on their understanding abilities. Students can define definitions or theorems, interpret and explain reason and communicate their understandings with their own language, make connections between definitions or theorems, and use them to solve mathematical problems.

If students understand and are able to interpret definitions and theorems in Real Analysis, then they analyze a statement and use it to solve a problem. Students need to find and describe every piece of information that they have, and use the information quickly. The complexity of student thinking will be higher.

Research that has been done [4] shows that the ability of students to conduct analysis can be improved. This is seen from the development of the results of learning outcomes from the preliminary analysis conducted by students. The ability of students to carry out the analysis is obtained from the ability of students to conduct preliminary analysis before solving Real Analysis problems. However, can the ability to conduct preliminary analysis be seen from the results of the learning outcomes and how much its results reflect the ability of students to conduct preliminary analysis?

Questions about the relationship of ability in conducting preliminary analysis with learning outcomes arise because its can also be influenced by the students' habits of memorizing the proof and the proof step. As a result, the learning outcomes are not a whole can be used as a measure in determining the ability of students in analyzing.

Information about the ability of students conducting analyze is important to known, especially for courses that make Real Analysis a prerequisite. So, it can be predicted how the ability of students to analyze the material to be given. Lecturers can estimate a method or a way to provide maximum lectures, both on the development of students' ability to analyze and study the material. This will also result in the development of analysis in the next lecture.

On the problems that have been raised, the phenomenon is the ability of the analysis will affect the final exam results. Then the independent variable is the ability of the analysis and the dependent variable is the final exam result. Because that will be predicted independent variables based on the dependent variable, it can be used inverse regression. Inverse regression [5] looked at the inverse regression equation, that is

\[ y = \beta_0 + \beta_1 x + \varepsilon \]  \hspace{1cm} (1)

where \( y \) is response variable, \( \beta_0 \) is intercept, \( \beta_1 \) is slope, \( x \) is regressor variable, and \( \varepsilon \) is error.

In theory and associated with the rules of inverse regression experiments it is most appropriate to model this problem, but its use is not easy [6]. Users prefer to use reverse regression.

Reverse regression [7] in principle is to exchange the role of independent variables into dependent variables, and vice versa. If using reverse regression, it must fulfill the assumption that \( y \) is a certain value instead of a random variable and also \( x \) must fulfill the assumption that \( x \) is a random variable. But, experimentally these assumptions are not fulfilled.

Both methods can produce estimates that are very different from unknown \( x \) values, as well as different prediction intervals [6]. To solve the problems that have been raised, namely predicting the ability of students to analyze the problems given, then inverse regression and reverse regression are used. The formulation of the problem in this research is "Does the student's ability to perform an analysis of the given problem can be predicted based on the learning outcomes that obtained?"

2. Research methods
This research is applied research. In this study theories are collected, examined, modified, and used to get answers to the problem. The steps taken to answer the problem are as follows.

First step is collecting data about the ability of students in conducting preliminary analysis. The process of obtaining the data are 1) do Real Analysis learning, in this learning the students are trained...
to conduct preliminary analysis using flow proof analysis, 2) collect the flow proof analysis scores of each meeting, 3) calculate the average scores of flow proof analysis, 4) collect the scores of learning outcomes.

Second step is examine theoretically the comparison between reverse regression and inverse regression, and to model reverse regression and inverse regression. And then, make an interval estimate for each model of the student's ability to conduct preliminary analyzes and an interpretation to provide recommendations on the student's ability to conduct a preliminary analysis.

3. Results and Discussions
During the study, students were trained using flow proof analysis in conducting preliminary analysis. It aims to train students' ability in analyzing the given problems. Based on the observations, we obtained the value of students 'ability in conducting preliminary analysis.

The test results obtained by the students are influenced by their ability to perform preliminary analysis. The results of the two tests are calculated on average.

Data on the students' ability in conducting preliminary analysis and the average learning outcomes scores obtained by students are drawn up in plot. This plot can be used to examine data trends. The plot of the data can be seen in Figure 1.

![Figure 1. Plot of analysis and learning outcomes scores.](image)

Diagram shows that the ability to do preliminary analysis is higher than the test score of learning outcomes. This means that the students' ability to analyze the problems given is higher than the learning outcomes scores they obtained. In accordance with the objectives of the Real Analysis lecture, there is a change in the ability of students to analyze the problem given, but students have not been able to write the analysis in the form of a solution.

3.1. Inverse Regression
To see the relationship between the scores of the ability to do preliminary analysis (A) and learning outcomes (LO), it is described in the form scatter diagram. The scatter diagram between analysis and learning outcomes score can be seen in Figure 2.

There is a linear relationship between the ability to do preliminary analysis with the learning outcomes scores obtained by students in scatter diagram. Furthermore, there is a significant constant addition to the relationship. This is indicated by a linear regression line that is not close to point O. So, it can be concluded that simple linear regression can be used for the analysis of subsequent relationships.

The regression model from scores is \( LO = 14.5 + 0.611 \cdot A \). In this case, \( \hat{\beta}_0 = 14.468 \) with P-value is 0.001 and \( \hat{\beta}_1 = 0.61077 \) with P-value is 0.000. This means that the parameter \( \beta_0 \) has an influence on the
model. For parameter $\beta_1$, it can be said that there is a linear relationship between the scores of analysis and the learning outcomes. Variation of learning outcomes scores, 57.8% is caused by analysis scores.

![Scatterplot of learning outcomes vs analysis](image)

Figure 2. Scatter diagram of learning outcomes and analysis scores.

Basic regression assumptions must be investigated before the model is finally adopted for use. After investigated, all assumptions have been valid. For this reason, the model can be used for further analysis. Based on the equation (1), for the analysis and learning outcomes scores we find

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \bar{x} + \frac{\bar{y} - \hat{\beta}_1}{\hat{\beta}_1} = 43.25122 + \frac{LO - 40.88415}{0.61077}$$

(2)

The estimation interval of analysis scores at confidence coefficient less than 95% with used Graybill’s Method [8] is

$$43.25122 + \frac{(0.61077)(LO - 40.88415)}{0.3446} - \frac{(13.7124)(2.02)}{41} \left\{ \frac{0.3446}{41} + \frac{(LO - 40.88415)^2}{26952.86} \right\} \leq A \leq 43.25122 + \frac{(0.61077)(LO - 40.88415)}{0.3446} + \frac{(13.7124)(2.02)}{41} \left\{ \frac{0.3446}{41} + \frac{(LO - 40.88415)^2}{26952.86} \right\}$$

(3)

For example, learning outcomes score $y_0$ is 70. Estimation of analysis score is 90.92. The estimation interval of analysis score at confidence coefficient less than 95% is $42.3206 \leq A \leq 141.9986$. Because maximum score is 100, then $42.3206 \leq A \leq 100$.

3.2. Reverse Regression

Reverse regression in principle views the regression problem in reverse, namely exchanging the role of the independent variable into the dependent variable, and vice versa. The solution with reverse regression is to make regression $x$ over $y$, that is

$$x = \gamma + \eta y + \epsilon'$$

(4)

So, in this case the variable $y$ is considered as a regressor and the variable $x$ is a response variable. Inverse estimator of $x_0$ is obtained by forming a simple linear regression model of the analysis scores over the learning outcomes. To see the relationship that occurs between them, it is expressed in the form of scatter diagram. The diagram between learning outcomes score and analysis scores can be seen in Figure 3.
In scatter diagram, it can be seen that there is a linear relationship between the learning outcomes and the analysis scores. So, it can be concluded that simple linear regression can be used for the analysis of subsequent relations.

The simple linear regression model is $A = 4.54 + 0.947 \hat{LO}$. In this case, $\hat{\gamma} = 4.543$ with $P$-value is 0.448 and $\hat{\eta}$ is 0.9468 with $P$-value is 0.000. This means that the parameter $\gamma$ has no effect on the analysis scores and there is a linear relationship between the learning outcomes and analysis scores with a 5% error level. This model 57.8% can illustrate the varying scores.

Before using the regression model that has been obtained, all basic assumptions must be investigated. If all basic assumptions have been valid, then the model can be used. The estimation interval [10] of analysis scores at a 95% confidence level is

$$
(4.54 + 0.947 \hat{LO}) - (2.02) \sqrt{291.5 \left(1 + \frac{1}{41} + \frac{(\hat{LO} - 40.88415)^2}{\sum_{i=1}^{n} (y_i - 40.88415)^2}\right)} \leq A \leq
$$

$$
(4.54 + 0.947 \hat{LO}) + (2.02) \sqrt{291.5 \left(1 + \frac{1}{41} + \frac{(\hat{LO} - 40.88415)^2}{\sum_{i=1}^{n} (y_i - 40.88415)^2}\right)}
$$

For example, learning outcomes score $y_0$ is 70. Estimation of analysis score is 70.83. The estimation interval of analysis score at a 95% confidence level is $35.10 \leq A \leq 106.5565$. Because maximum score is 100, then $35.10 \leq A \leq 100$.

4. Conclusion
The result of this research is the ability of students to conduct an analysis can be predicted from their learning outcomes. The model of reverse regression and inverse regression can be used to predict the ability of students to analyze the problem given. Both of these results have advantages and disadvantages of each.

References
[1] Ruseffendi H E T 2006 Pengantar kepada membantu guru mengembangkan kompetensinya
dalam pengajaran matematika untuk meningkatkan CBSA (Bandung: Penerbit Tarsito)

[2] Soemarmo U 2008 Berpikir matematik, apa, mengapa, dan bagaimana cara mempelajarinya. Paper in lecture Analisis Hasil Penelitian Internasional

[3] Bartle R G and Sherbert D R 2011 *Introduction to real analysis fourth edition* (Singapore: John Wiley & Sons Inc)

[4] Helma and Subhan M 2011 *Upaya meningkatkan kemampuan mahasiswa dalam menyelesaikan masalah pada perkuliahan Real Analysis I dengan menggunakan mind mapping dan dipresentasikan dalam Bahasa Inggris* (Indonesia: FMIPA UNP)

[5] Montgomery D C, Peck E A and Vining G G 2006 *Introduction to linear regression analysis*. (New York: Jhon Wiley & Sons Inc)

[6] Parker P A, Geoffrey V G, Wilson S R, Szarka J L and Johnson N G 2010 *The prediction properties of inverse and reverse regression for the simple linear calibration problem* Journal of Quality Technology vol 42 p 332–347

[7] Krutchkoff R G 1967 *Classical and inverse regression methods of calibration* Technometrics vol 9 (America: American Statistical Association) p 425–439

[8] Thonnard M 2006 *Confidence intervals in inverse regression* (Endhoven: Technische Universiteit Endhoven Department of Mathematics and Computer Science)

[9] Graybill F A and Iyer H K 1962 *Regression analysis concepts and applications* Jhon W. Turkey vol. 33 (Turkey: Jhon W. Turkey)

[10] Seber G A F and Lee A J 2003 *Linear regression analysis second edition* (New York: Jhon Wiley & Sons Inc)