Factors Affecting Academic Achievement Students' of Mathematics Department by Using Probit Regression Analysis

N Amalita\textsuperscript{1*}, A Arnellis\textsuperscript{1} and D Fitria\textsuperscript{1}

\textsuperscript{1}Mathematics Department, Universitas Negeri Padang, Padang, Indonesia

*corresponding author: nongmat@fmipa.unp.ac.id

Abstract. One of the students' academic achievement can be seen GPA, the higher the index value, the better student achievement success rate of students who obtained. Many factors affect student achievement index both internal factors and external factors. This paper described at factors that affect student achievement index of Department of Mathematics UNP who enrolled 2016 by using probit regression analysis. The results of this study are students SBMPTN, does not work while in college, and was not active in organizations on campus have the opportunity to get a better grade.

1. Introduction

Academic performance in college is a picture of the student success rate in a given period. One of the students' academic results can be seen at GPA achieved by students every semester. Based on the regulation of the minister of research, technology and Higher Education of the Republic of Indonesia [1] No. 4/2015 on standards of higher education states that assessment of learning outcomes in each semester with GPA stated. To obtain a high GPA index takes hard work and perseverance in the learning process, but it is also supported by environmental factors. The higher the value of GPA students mean success rate is the better student. Based on information obtained from some of the students there are many factors that affect the GPA, including the motivation to learn from within itself, the full support of the family in education, influence of friends and environment, complete facilities in the learning process, and the economy.

According to Kuhl [2], factors which will affect the success of a college student learning can be divided into two stages, namely pra college experience that includes enrollment, academic preparation includes academic achievement at the level of previous education, intelligence and college readiness, family support and co motivation to learn, as well as demographic characteristics (eg, gender, race, and socioeconomic conditions), and lecturing experience that includes student behavior, the condition of the institution, and student engagement. According Suryabrata [3], which states that the factor that affect the learning process and results there are two factors that are external and factors internal. Thus, a lot of factors that affected the students’ GPA. It is necessary for an analysis of the factors that significantly affect student achievement index. It is necessary to set up a model of what factors that can influence student achievement academic

One of the statistical analyze that can be used to determine the factors that affect student achievement index is probit regression analysis. Linear regression model used is the analysis of the data describing
the relationship between the dependent variable with one or several independent variables. In generally regression analysis is used to analyze the data with quantitative data response variable. But in everyday life are often encountered cases with a variable response in the form of qualitative variables (categorical) by taking two or more possible values. To resolve this case can be used logit and probit models.

Probit regression model is non-linear regression model that describes the relationship dependent variable \(Y\) for which data was qualitative scale with two possible binary data that is successful or unsuccessful. Independent variable \(X\) may be in the form of quantitative data with nominal or ordinal scale or a combination of both.

- According to Gujarati [4] probit regression model is non-linear regression model using binary or dummy variables as the independent variable and error assumes normal distribution \(N(0, \sigma^2)\). Suppose that there are \(n\) observations, \(Y_i\) is the dependent variable equal to 1 which states "successful" and 0 are declared "failed" and \(X_i\) is a independent variable that affects \(Y_i\). If \(P_i\) stated the success probability, then the regression model probitnya are: [5]

\[
Y_i^* = \text{Probit} (P_i) = \Phi^{-1}(P_i) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon_i
\]

dan diasumsikan \(Y_i = \begin{cases} 1 & \text{jika } Y_i^* > 0 \\ 0 & \text{jika } Y_i^* \leq 0 \end{cases} \)

Where :
- \(Y_i^*\) = normally distributed dependent variable
- \(\Phi^{-1}\) = Inverse of the standard normal cumulative distribution function
- \(\beta_0\) = intercept unknown parameter
- \(\beta_i\) = coefficient parameter
- \(x_i\) = the independent variable
- \(\varepsilon_i\) = error of normal distribution

Because of an error on this probit regression equation was the normal distribution, and based on obtained equation probability are as follows: \(Y_i = 1\)

\[
P_i = P(Y_i^* > 0) = P (\beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon_i > 0 ) = P (\varepsilon_i > - [ \beta_0 + \sum_{i=1}^{n} \beta_i x_i ] ) = 1 - F(- \beta_0 - \sum_{i=1}^{n} \beta_i x_i )
\]

Where \(F\) indicated the function of the cumulative probability. \(P_i\) represent an opportunity as measured by the area under the standard normal curve of \(-\infty\) until \(Z_i\) \((Y_i^*\) on probit regression model). If there is no certain integral and normal distribution used is raw, then to see the value of the probability to use the standard normal distribution table. On the table has been calculated value:

\[
P(-\infty \leq Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt
\]

Having obtained the regression model, then we can suppose that \(P_i\) probability by using the cumulative normal distribution table.

Estimation of parameters in the probit regression can use the maximum likelihood method. Which is one good estimation methods used to estimate the regression parameter, especially for a large enough sample size, such as the data used in the probit model. Estimation of this method is done by maximizing the likelihood function. Furthermore, the testing models and parameters were done. According probit regression model, the parameters resulting from the estimation parameters do not necessarily have a significant effect on model. Therefore, be tested against each parameter

- Concurrent test
According [6] Concurrent test is done to check the significance coefficient β in whole or simulatnt. Significance testing was conducted to examine whether the significant independent variables in the models
Hypothesis testing is:
\[ H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0 \]
\[ H_1: \text{at least one, } \beta_j \neq 0; j = 1, 2, \ldots, p \]

Statistical tests conducted is a test G or likelihood ratio test items, namely:
\[ G = 2 \sum \left[ y_i \ln(\hat{\pi}_i) + (1-y_i)\ln(1-\hat{\pi}_i) \right] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \]
where:
- \( y_i \) is valued by 1
- \( n_1 \) = number of \( y_i \) is valued by 1
- \( n_0 \) = number of \( y_i \) is valued by 0
- \( n \) = many \( y_i \)

If the test statistic G follows \( \chi^2 \) distribution, the test is done by comparing the value of the test statistic G and a value table \( \chi^2 \) with degrees of freedom \( v \) (number of parameters) at the significant level of \( \alpha \). \( H_0 \) is rejected, if the value \( G > \chi^2(v, \alpha) \) or \( p\)-value < \( \alpha \). This means, on the model of there is at least one parameter estimators are not equal to zero.

- Partial test
  Partial test was performed to test the significant of the partial β coefficient by comparing estimasi β by the estimates of standard error.
  By hypothesis:
  \[ H_0 : \beta_j = 0 \]
  \[ H_1 : \beta_j \neq 0; j = 1, 2, \ldots, p \]
  Wald test statistic [6]
  \[ W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \]
  With:
  \( \hat{\beta}_j \) = dugaan koefisien β ke j
  \( SE(\hat{\beta}_j) \) = standar error dari koefisien \( \hat{\beta}_j \)

Wald test is a test that aims to determine whether a significant regression parameters or not. Statistic of W test follows the standard normal distribution, then the test is done by comparing the Wald test statistic with a standard normal distribution at significance level \( \alpha \). \( H_0 \) rejected if the value of \( |W| > Z_{\alpha/2} \) or \( p\)-value < \( \alpha \).

- Selection of the Best Model
  Selection is done so that the best model obtained probit model is the appropriate model. Selection of the best models using backward elimination method is by reducing one by one independent variable is not significant in order to obtain all the independent variables were significant. Criteria for selecting the best model is based on the value of deviance. The best model is the model that have smallest Deviance value [6]. Deviance value is formulated as follows:
  \[ D = -2 \sum_{i=1}^{n} \left[ y_i \ln\left( \frac{\hat{\pi}_i}{y_i} \right) + (1-y_i)\ln\left( \frac{1-\hat{\pi}_i}{1-y_i} \right) \right] \]

2. Method
This type of research is applied research. Data was taken from the Department of Mathematics student enrollment in academic year 2016 as many as 257 students. Data collection techniques are using by questionnaire from students of the Department of Mathematics Joined Year 2016. Variables that are used consist of a dependent variable and independent variables Independent variable (X) may be in the
form of quantitative data with nominal or ordinal scale. Data analysis techniques use probit regression model. Probit regression model is non-linear regression model that describes the relationship dependent variable (Y) for which data was qualitative scale.

3. Result and Discussion

Data was taken from the Department of Mathematics student enrollment in academic year 2016 as many as 257 students with details such as the following table:

**Table 1. Number of student enrollment in academic year 2016 from the Department of Mathematics**

| Study Program         | Amount |
|-----------------------|--------|
| Mathematics Education | 122    |
| Mathematics           | 75     |
| Statistics D3         | 60     |
| Amount                | 257    |

Variables that are used consist of a dependent variable and independent variables [7]. Each variable is categorized as follows:

**Table 2. Dependent and Independent Variables**

| Types of Variables | Variable | Information | Category          |
|--------------------|----------|-------------|-------------------|
| Dependent          | Y        | GPA         | 0: ≥ 3.00         |
|                    |          |             | 1: <3.00          |
| Independent        | X1       | Gender      | 0: Male           |
|                    |          |             | 1: Female         |
|                    | X2       | The average value of the UN | 0: ≥ 70 |
|                    |          |             | 1: <70            |
|                    | X3       | Entrance    | 0: SNMPTN         |
|                    |          |             | 1: SBMPTN         |
|                    | X4       | School from | 0: State          |
|                    |          |             | 1: Private        |
|                    | X5       | Program selection | 0: First |
|                    |          |             | 1: Second         |
|                    |          |             | 2: Third          |
|                    | X6       | Study While Working | 0: Yes |
|                    |          |             | 1: No             |
|                    | X7       | Receive Scholarships for college | 0: Yes |
|                    |          |             | 1: No             |
|                    | X8       | Organizations Active Campus | 0: Yes |
|                    |          |             | 1: No             |

The first step is to form a probit regression model that is formed by including all independent variables are

\[
\Phi^{-1}(P_i) = -0.979 + 0.036 X_1 + 0.204 X_2 + 0.765 X_3 - 0.012 X_4 + 0.156 X_5 + 1.075 X_6 - 0.087 X_7 + 0.438 X_8
\]

Further testing the significance of the model by including all the independent variables to test the significance test G. The output of the model can be seen in the following table:
Table 3. Significance Tests Probit Regression Model

| Log Likelihood | G  | df | p-value |
|----------------|----|----|---------|
| -143,698        | 66,821 | 8 | 0,000 |

p-value = 0,000 < α = 0,05, berarti hipotesis $H_0$ ditolak, it means that at least there is one independent variable which affects the dependent variable.

The parameter significant of testing by using the Wald test. Hypotheses used are:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$, untuk: $j = 1,2, ..., 8$

The results of significance test using the Wald test parameters can be seen in the following table:

Table 4. Probit Regression Parameter Significance Tests

| Variables                      | Z   | p-value |
|--------------------------------|-----|---------|
| Constants                      | -3,46 | 0,001 |
| Gender ($X_1$)                 | 0,15 | 0,879 |
| The average value of the UN ($X_2$) | 1,08 | 0,282 |
| Entrance ($X_3$)               | 2,72 | 0,007 |
| School From ($X_4$)            | -0,12 | 0,906 |
| Program selection ($X_5$)      | 1,33 | 0,185 |
| Study While Working ($X_6$)    | 5,83 | 0,000 |
| Receive Scholarships ($X_7$)   | -0,45 | 0,654 |
| Active Campus Organization ($X_8$) | 2,47 | 0,014 |

Based on Table 4, the variables of entrance, college while working and active campus organization is a significant variable or variables that affect achievement index, because the p-value less than the value of $\alpha = 0,05$, while the other variables do not have a significant effect on academic achievement. Thus, it is necessary to do the selection of the best model by reducing one by one independent variable which has a value of $\alpha$ greater than 0.05. Variables that are reduced in a row is the school from, gender, receiving scholarships, average score UN and studied program options. After the reduction, the obtained three independent variables were significant as in the following table:

Table 5. Results of Probit Regression Model Reduction

| Variables                            | Coefficient | p-value |
|--------------------------------------|-------------|---------|
| Constants                            | -0,882      | 0,000   |
| Entrance ($X_3$)                     | 0,772       | 0,005   |
| Study While Working ($X_6$)          | 1,087       | 0,000   |
| Active Campus Organization ($X_8$)   | 0,435       | 0,011   |

Furthermore, the significance test after reduced probit regression models with G test that can be seen in the following table:

Table 6. Significance Test by Regression Model Probit Having Reduced

| Log Likelihood | G  | df | p-value |
|----------------|----|----|---------|
| -145,302       | 63,613 | 3 | 0,000   |
Based on table 6, p-value of probit regression models reduced after 0.000 less than the value of \( \alpha = 0.05 \). This means that the independent variable in the regression model of probit t affect the dependent variable.

The criteria be selecting the best models is based on the Deviance value. The best model that has a value that is the result of the fifth reduction of the smallest deviance value is 0.103. So that is obtained the best probit regression models is

\[
\Phi^{-1}(P_i) = -0.882 + 0.772 X_3 + 1.087 X_6 + 0.435 X_8
\]

Based on the best probit regression models can be interpreted that the independent variable Entrance \( (X_3) \), College while working \( (X_6) \), and Organizations Active Campus \( (X_8) \) significantly affect students' academic achievement.

Then calculated the probability value of models can be obtained by calculating the value of probit beforehand and the results are shown in table normal distribution can be seen in the following table:

| No | X3 | X6 | X8 | Yi* | Pi (%) |
|----|----|----|----|-----|--------|
| 1  | 0  | 0  | 0  | -0.882 | 18.94  |
| 2  | 0  | 0  | 1  | -0.447 | 32.64  |
| 3  | 0  | 1  | 0  | 0.205  | 58.32  |
| 4  | 0  | 1  | 1  | 0.64   | 73.89  |
| 5  | 1  | 0  | 0  | -0.11  | 45.62  |
| 6  | 1  | 0  | 1  | 0.325  | 62.93  |
| 7  | 1  | 1  | 0  | 0.977  | 83.65  |
| 8  | 1  | 1  | 1  | 1.412  | 92.07  |

Based on Table 7 above shows that with the greatest opportunity to have a good academic record are students with entrance of SBMPTN, did not work during college and do not follow the student organization with a chance of 92.07%.

4. Conclusion
Probit regression model that describes the factors that affect students' academic achievement Department of Mathematics UNP who enrolled 2016 are:

\[
\Phi^{-1}(P_i) = -0.882 + 0.772 X_3 + 1.087 X_6 + 0.435 X_8
\]

Which is Entrance \( (X_3) \), college while working \( (X_6) \), and organizations active campus \( (X_8) \) significantly affect students' academic achievement. The best opportunity students have a good academic record are students with entrance of SBMPTN, did not work during college and do not follow the student organizations.

References
[1] Greene W 2007 Econometrics Analysis (Printice Hall: New Jersey)
[2] Gujarati 2004 Basic Econometric 4th Edition The McGraw-Hill Companies
[3] Hosmer DW and Lemeshow S 2000 Applied Logistic Regression, Second Edition (New York: John Wiley & Sons, Inc)
[4] Kuth et.al 2006 What Matters to Student Success; A REvier of Literature Commissioned Report for the National Symposium on Postsecondary Student Success
[5] Merdekawati G 2018 Penerapan Metode Chaid Untuk Mengklasifikasikan Faktor-faktor yang Mempengaruhi Prestasi Akademik Jurusan Matematika Tahun Masuk 2016 FMIPA UNP
[6] Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia (PERMENRISTEKDIKTI) Nomor 44 Tahun 2015 Tentang Standar Pendidikan Tinggi