Analysis of factors affecting the undergraduate student quit study

V Eminita' and R Widiyasari
Program Studi Pendidikan Matematika, Universitas Muhammadiyah Jakarta, Jl. KH. Ahmad Dahlan Cirendeu, Tanggerang Selatan 15419, Indonesia

*phiartea.emn@gmail.com

Abstract. Many undergraduate students on Faculty of Education University of Muhammadiyah Jakarta who are beginning to fall in their motivation to study and then quit their study. Based on data in 2010 there were 20.22% students quit the study. Many factors that affect students quit their study. This research is expected to contribute to overcoming the problem. The purpose of this study is to determine what factors affect the student quit studying. This research uses Poisson or Binomial Negative regression analysis to know these factors. Both regression models used are equally the fit model to be used, because based on the result of over dispersion study did not occur. Based on the AIC and Log-Likelihood criteria it is concluded that negative binomial model is more fit than Poisson regression models. Wald test results for each parameter in the Negative Binomial regression are only three variables that affect students quit the study. The three parameters are the percentage of Jakarta students (X2), the percentage of male students (X3), the average student age (X4). The three variables have a positive effect, which means that increasing value in X2, X3, or X4 will increase the number of the students who quit studying.

1. Introduction
Universitas Muhammadiyah Jakarta (UMJ) is one of the private universities in Jakarta that upholds the values of Islam. This university consists of nine faculties that are increasingly favored by students of SMA / SMK who will continue their study. This is evident from the increasing number of student who enters each year. However, in recent years, many students have begun dropping their motivation to study and then quit their studies, especially in the faculty of education (FIP) UMJ. Base on academic data in 2010, there were 20.22% of students who quit studying. Many factors make students quit studying, so this needs to be a concern for university and faculty in making policy.

The number of students that quit study is a case of count data, so the exact analysis for this case is Poisson or Binomial Negative Regression. In recent years, Poisson regression models have been used to model count response variable affected by one or more covariates [1]. The Poisson regression model is basically a regression model that meets the classical assumptions with only one exception, that is the dependent variable assumes Poisson distribution [2]. Poisson-distributed dependent variables, this assumption implies that the conditional variance of the dependent variable, being equal to the mean (Var[Y] = E[Y]) [3]. Over dispersion often occurs in Poisson Regression (PR), this is because Var[Y] > E[Y]. Overdispersion in count data may be due to unobserved heterogeneity which leads to the widely-used negative binomial regression (NBR) model [4,5]. The evidence of over dispersion in the data is The ratio value between Devians and degrees of freedom is not close to 1 [6]. Aike Information Criteria
(AIC) and log-likelihood were used in the selection of appropriate regression model, the small value of AIC and log-likelihood of the model accepted as a good model. Akaike's entropic information criterion, which is known as AIC, has had a fundamental impact on statistical model evaluation problems [7]. The log-likelihood function can then be maximized by changing the value of the parameter and seeing which values give the maximum of the log-likelihood (Note that the log-likelihood is often negative), so the maximum is the value closest to zero [8].

Based on the above description, the authors want to describe the characteristics of students who quit the study and examine what factors affect the students undergraduate FIP UMJ quit study at each course of the year 2010, 2011, 2012, 2013, and 2015 use PR and NBR model. This research is expected to be useful for faculty to university to solve this problem.

2. Method
The research uses the quantitative approach which population is 5 program study undergraduate FIP UMJ, they are Primary Teacher Education (PGSD), Early Childhood Education (PAUD), Indonesian Language and Literature Education (PBSI), English Language Education (PBI), and Mathematics Education (MTK) in the year 2010-2015. Data obtained from the academic section of FIP UMJ. The sample data used in this study is the data of each undergraduate FIP UMJ Degree Program in 2010, 2011, 2012, 2013 and 2015 for PGSD, PAUD, PBSI, PBI, and MTK study program and 4 years for PBI study so that the sample size used is \( n = 24 \).

The dependent variable (\( Y \)) used in this study is the number of incidents of undergraduate FIP UMJ students who quit study in each study program. While the independent variables (\( X \)) are shown in Table 1, they are:

| Variables | Definition |
|-----------|------------|
| X1        | Percentage of Students coming from public schools in each course |
| X2        | Percentage of Students from DKI Jakarta |
| X3        | Percentage of Male Students in each course. |
| X4        | The average of Students in each Course |
| X5        | Percentage of Students that just graduated from high school |

This research will be outlined in 3 stages. First, the data exploration of each \( Y \) and \( X \) with descriptive statistical analysis. Second, identify the PR and NBR models. The second stage is done by exploring \( Y \) that is making a histogram of \( Y \) data to know the indication of Poisson distribution conditions. Then perform the chi-square test on \( Y \) variable to identify the \( Y \) variable is Poisson or Negative Binomial distribution.

The third stage of this research is to predict the model to find out what factors that affect the students quit the study. This stage uses the PR and NBR analysis to estimate the model coefficients and test the parameters of the regression coefficient parameters with the Wald test. Then test the goodness of fit model obtained by the Pearson Chi-squared test on both models to determine whether there is over dispersion or not. Furthermore, comparing PR and NBR based on exploratory and test results on \( Y \) variables, over dispersion testing, and model evaluation of \( Y \) estimators using AIC and log-likelihood. Lastly, interpret the output analysis results from the best-selected model.

3. Result and discussion

3.1. Data exploration
Table 2 shows that the number of undergraduate FIP UMJ students from 2010 to 2015 (except 2014) is decreasing, while the number of students who quit study is still stable at around 5% to 20%, but this
number is still worrisome for the FIP UMJ because of the decreasing number of students each year. In 2010, the highest percentage of students who quit studying, which was 20.22%. This should be a concern for undergraduate FIP UMJ to avoid happening in the following years.

Table 2. Percentage of quit studying.

| Year | Number of Students | Number of Quit Studying | Percentage of Quit Studying |
|------|--------------------|-------------------------|-----------------------------|
| 2010 | 544                | 110                     | 20.22%                      |
| 2011 | 549                | 88                      | 16.03%                      |
| 2012 | 320                | 18                      | 5.63%                       |
| 2013 | 326                | 33                      | 10.12%                      |
| 2015 | 228                | 23                      | 10.09%                      |

Figure 1 shows the percentage of students coming from public schools. The average percentage of students who exceed 50% occurs in 2012, whereas in 2013 the average student coming from the country declined, which is below 40%. PBSI and PBI studies in 2010 do not have students from public schools.

Figure 2 is the percentage of students coming from Jakarta for each study program from 2010-2015. PBSI and PBI studies in 2010 have not been established, so there is no number of students. Figure 2 also shows that students who come from Jakarta more choose PAUD study compared to other studies.
Figure 3 shows the percentage of male students in each study. In general, the percentage of male students did not exceed 50% and the highest percentage was achieved by PBSI in 2010. In PAUD study did not even have a single male student or 0%.

![Figure 3. Percentage male students.](image3)

Figure 4 is a picture of the average age of FIP UMJ students in each study program. Figure 4 explains that for the four study programs, PAUD, PGSD, MTK, PBSI, and PBI have an average age below 20 years. While the PAUD study on average age over 25 years.

![Figure 4. The average age of FIP UMJ students.](image4)

Figure 5 is the percentage of students who immediately continue their studies immediately after graduating from high school. Figure 5 shows that the PGSD, MTK, PBSI, and PBI study programs have a percentage above 50%, whereas only the PAUD study program has a percentage below 50%. This is evident from the characteristics of students, the average already working and married, but each year the percentage is increasing. This means that every year many students who have just graduated from high school continue their study to PAUD study program.

![Figure 5. The percentage of students who immediately continue their studies immediately after graduating from high school.](image5)
3.2. Poisson regression

The Poisson regression model of the number of students who quit study using five variables is shown in Table 3. Table 3 explains that all variables affect the number of students who quit the study. The increasing percentage of students coming from public schools (X1), the percentage of students coming from Jakarta (X2), the percentage of male students (X3), and the average age of the students (X4) will increase the number of students who quit the study from undergraduate FIP UMJ. While only the percentage change of students who graduated from high school immediately continue their study in FIP UMJ (X5) which will decrease the number of students who quit studying at FIP UMJ.

Table 3. The estimated value of poisson regression model parameter with 5 variables.

| Parameter | Estimated Value | Standard error | Wi   | conclusion |
|-----------|-----------------|----------------|------|------------|
| Intercept | -0.25           | 0.45           | -0.56| accept Ho  |
| X1        | 2.54            | 0.54           | 4.67 | reject Ho  |
| X2        | 3.54            | 0.52           | 6.79 | reject Ho  |
| X3        | 0.11            | 0.03           | 3.96 | reject Ho  |
| X4        | 0.07            | 0.01           | 5.69 | reject Ho  |
| X5        | -1.75           | 0.42           | -4.21| reject Ho  |

The plot of the relationship between the estimated value and the residual of the Poisson model is shown in Figure 6. Figure 6 shows that the data diversity tends to shrink around zero. Although there are some large ones. Data patterns tend not to be random or form a particular pattern (systematic). The ratio value between Devians and degrees of freedom is close to 1, this means that the Poisson regression model is not over dispersed [6]. So it is worth using.

3.3. Binomial Negative Regression

Negative Binomial regression model for students who quit study from FIP UMJ is presented in Table 4. Table 4 shows that only three variables affect the number of students who quit the study. The increasing percentage of students coming from Jakarta (X2), the percentage of male students (X3), and the average age of the students (X4) will increase the number of students who quit the study from FIP UMJ. While the other two variables of the percentage of students coming from state senior high school (X1), the percentage of freshmen who graduated from high school (X5) will immediately continue their study at FIP UMJ which will decrease the number of students who quit studying at FIP UMJ.
Table 4. The estimated value of negative binomial regression parameters with 5 variables.

| Parameter | Estimator value | Standard error | Gi    | Conclusion |
|-----------|-----------------|----------------|-------|------------|
| Intercept | -0.81           | 1.19           | -0.68 | accept Ho  |
| X1        | -0.47           | 1.78           | -0.27 | accept Ho  |
| X2        | 4.79            | 1.66           | 2.89  | Reject Ho  |
| X3        | 0.15            | 0.08           | 2.01  | Reject Ho  |
| X4        | 0.11            | 0.04           | 2.78  | Reject Ho  |
| X5        | -0.35           | 1.55           | -0.23 | Accept Ho  |

The plot of the relation between the residual and the estimated value of the negative binomial regression is presented in Figure 7. Figure 7 shows the same thing with the Poisson regression model, i.e., the data diversity tends to narrow around zero. Although there are some large but not very large ones like the Poisson regression model. Data patterns tend not to be random or form a particular pattern (systematic). The ratio value between Devians and degrees of freedom is close to 1, this means that the regression model is not over dispersed [6], so it is worth using. So there needs to be another comparison used to compare the two regression models which are more feasible to use.

Figure 7. Figure of The relationship between the estimated value with the residual of the binomial negative model.

3.4. Interpretation of poisson regression and binomial negative model

Based on previous studies indicating that both models are equally worthy of use. So there needs to be the further in-depth review of both models. Figure 8 shows the histogram of the variable distribution of Y. Figure 8 explains that the distribution of Y variable resembles the characteristic of the Poisson and Binomial Negative distribution whose longer tail extends to the right [3].

Figure 8. Histogram of variable Y.
A review of the two models is presented in Table 5 below. Table 5 shows that the negative binomial model is superior to the Poisson model. This is shown by smaller AIC values and larger Log-likelihood values than the Poisson model. So it can be concluded that the model used to explain the number of students who quit the study of FIP UMJ is a negative binomial model [7,8]. The model that only three variables only affect the number of students quit the study.

**Table 5.** Fitted value of poisson regression and binomial negative regression model.

| Model         | AIC    | Log-Likelihood |
|---------------|--------|----------------|
| Binomial Negatif | 168.12 | -77.06         |
| Poisson       | 385.34 | -186.67        |

### 4. Conclusion

The data characteristic of FIP UMJ students who quit study in each study program in 2010, 2011, 2012, 2013, and 2015 has the distribution that resembles Poisson and Binomial Negative distribution. This is also shown from the results of over dispersion studies on both regression models. Based on the in-depth study of both models using AIC and log-likelihood criteria, the Negative Binomial regression model is a viable regression model used to determine which factors affect the UMM FIP students who quit the study. The result of analysis by using binomial negative regression model are only three variables that affect the students of FIP UMJ who quit the study, that is the percentage of students who come from Jakarta (X2), the percentage of male students (X3), and the average student age (X4). The three variables have a positive effect, which means that increasing percentage in X2, X3, or X4 average value will increase the number of FIP UMJ students who quit studying.

### Acknowledgments

The authors would like to thank the Directorate of Research and Community Service for funding all research and publication of the results of this research. Head of FIP UMJ as a research partner in inter-university research that has helped the author in the implementation of research and refinement of research writing results.

### References

[1] Famoye F, Wuluh J and Sing K 2004 On the Generalized Poisson Regression Model with an Application to Accident Data *J. Data Sci.* 2 287–95

[2] Cupal M, Deev O and Linnertova D 2014 The Poisson Regression Analysis for Ocurrence of Floods *Proceeding 2nd Global Conference on Business, Economics, Management and Tourism* (Prague: Elsevier) pp 1499–502

[3] Cameron A C and Trivedi P K 2007 *Essentials of Count Data Regression*

[4] Moksony F and Rita H 2014 The Use of Poisson Regression in the Sociological Study of Suicide *Corvinus J. Sociol. Soc. Policy* 5 97–114

[5] Jansakul N 2005 Zero-Inflated Negative Binomial via R *In Proceeding 20th International Workshop on Statistical Modelling* (Sydney, Australia) pp 277–84

[6] Alison P and Waterman R 2002 Fixed-Effects Negative Binomial Regression Models *Sociol. Methodol.* 32 247–66

[7] Bozdogan 1987 Model Selection And Akaike’s Information Criterion (Aic): The General Theory And Its Analytical Extensions *Psychometrika* 52 345–70

[8] Schwarz C 2011 *Sampling, Regression, Experimental Design and Analysis for Environmental Scientists, Biologists, and Resource Managers* (The big JMP book 1668)