JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 5, No. 2, October 2021, pp. 287-293

The Comparison Results of Logit and Probit Regression on Factors of Woman Criminal

Eva Khoirun Nisa[1], Any Muanalifah[2]
[1]Mathematics, Universitas Islam Negeri Walisongo, Indonesia
[2]Mathematics, University of Birmingham, England
[1]evakn@walisongo.ac.id, [2]a.muanalifah@bham.ac.uk

Article History:

Received : 18-02-2021
Revised : 29-04-2021
Accepted : 01-05-2021
Online : 26-10-2021

Keyword:

Logit;
Probit;
Woman Criminal

ABSTRACT

Logit and probit regression are statistical methods that have the same is to determine predictor variables affect the categorical response variable. From relevant studies, they have advantages and disadvantages on certain cases. Therefore, the logit and probit regression will be applied on factors of woman criminal considering women crime rate in Central Java is still relatively high. This study aims to compare the result of the two regression so that the best regression model is obtained to explain factors that influence woman criminal. The data type used is primary data obtained using a questionnaire and validated by non-empirical validation. The sample was taken using a quota sampling technique. The result showed that probit regression is the best model with the factors that influence woman criminal are the age of 17-25 years, 26-35 years, junior high school education level, married and ever married marital status.

https://doi.org/10.31764/jtam.v5i2.4150

A. INTRODUCTION

There have been many social studies use the statistical methods in processing categorical data. The social research more often uses closed questions on questionnaires with categorical answers (Emily, 2017). The function of logit regression to determine predictor variables affect the categorical response variable. The advantages of this model that it is well known for its simple form, the results are significant and robust (Antipov, 2019) and provide the probability of independent variables affect the dependent variable (Agresti, 2002). It can be said that identified practical advantages of logit over probit model were the simplicity of its structural form and interpretability of the results (Adekanmbi, 2017).

Logit regression has log link function (Agresti, 2002). It is difficult to transform log model to linear model. Therefore the probit regression is created by statistician in order to can be easily converted to a normal table. Probit regression is the same as a logit regression. The difference is the link function is approached with a standard normal distribution (Agresti, 2002) so that easier to interpret the model (Widhiarso, 2012). However, according to German

in (Klieštik et al., 2015) logit regression has simpler model than probit because it does not contain integral. They have advantages and disadvantages on certain cases.

Some previous studies only applied logit regression or probit. Ari and Aydin's research (Ari & Aydin, 2016) apply multinomial logit regression on the influence of factors on women victims of domestic violence. The response variable (Y) in this study is the type of domestic violence with three categories, namely physical violence, psychological violence, physical and psychological violence. The results of these studies indicate that education level of woman and husband's work sector were statistically significant in physical violence; the agnation with husband, education level of husband, frequency of seeing drunk husband and frequency of gambling of husband were statistically significant in psychological violence; the region, deceived by husband and common-law female for husband were statistically significant in physical and psychological violence. There is also a study on women as victims of domestic violence using a probit regression (Borrego & Raquel, 2017). (Williams, 2016) compared the logit and probit regression coefficients using the heterogeneity model. This research only examines statistical theory, there is no application to the data.

Meanwhile, data released by the Central Statistics Agency (BPS) in 2018 shows that the crime rate by women in Central Java is still high. Likewise, what happened in the Semarang women's prison, from 2017 to 2018 the decline was not significant, only 9%. It is very interesting to identify the cause because equality before the law but there are differences in the factors that cause men and women to commit crimes (Mieczyslaw, 2017). This is the background of this research in order to obtain the best model from the comparison of logit regression and probit. The causes of general and specific female crimes can be determined from the best model.

B. METHODS

The type of data used is primary data obtained using a questionnaire and validated by non-empirical validation. Non-empirical validation was chosen because the research was carried out in in Semarang women's prison which was limited by time, opportunity and data collection policies. The non-empirical validation does not require a trial sample (Sousa, 2014). Therefore the instrument in this study was validated by a team of experts in the fields of criminal law, criminology and psychology. The variables used in this study are as follows

Table 1. Variables

Variables	Categories
Y: woman criminal	0: specific, 1: general
X1: age	1: < 17 years old, 2: 17 - 25 years old, 3: 26 - 35 years old, 4: 36 - 45 years old, 5: > 45 years old
X2: education level	1: not in school, 2: primary school, 3: junior high school, 4: senior high school, 5: higher education
X3: marital status	1: single, 2: married, 3: ever married
X4: profession	1: unemployed or housewives, 2: informal workers, 3: entrepreneur, 4: state employees, civil servants
X5: invitation to	1: no, 2: yes

commit a crime	
X6: psychological disorders	1: no, 2: yes
X7: religiosity level	1: low, 2: medium, 3: high

The determination of predictor variables is based on the results of several relevant studies such as research by Mili et al. (2015) that basically a woman commits a crime due to three factors, biological, psychological and social. The population in this study were all female prisoners in Semarang women's prison. The sample was taken using a quota sampling technique because the data collection policy was carried out by prison officials. Quota sampling might be an acceptable alternative when it is impossible to achieve high response rate in probability sampling (Yang & Banamah, 2014). n is sample size to be taken, n_0 the recommended sample size and the N population size. According to (Cochran, 1977) if the population size is too small, the sample size can be taken

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}} \tag{1}$$

The analysis begins with determining the estimator of the logit and probit regression models then testing the parameters simultaneously and partially so that all parameters are significant. Finally, the best model is obtained with the smallest Akaike Information Criteria (AIC). AIC was chosen because is calculated by obtaining the maximum value of the likelihood function for the model (Pham, 2019). The purpose of AIC is to balance both the fit of the model to the data with any information lost by including more variables (Snipes & Taylor, 2014).

C. RESULT AND DISCUSSION

1. Logit Regression on Woman Criminal

Estimating parameters in logit regression uses Maximum Likelihood Estimation (MLE) method because it produces an unbiased estimator (Bain & Engelhardt, 1992). The combination of MLE and AIC is appropriate to select the statistical model which fits the data (MIURA, 2011). The results of the parameter estimation with are as follows

$$\sum_{i=1}^{n}\left(y_i \sum_{j=1}^{k} X_{ij}\right) = \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}}} X_{ij} \tag{2}$$

It is nonlinear model so it is continued with Newton Raphson. The convergence rate of Newton method is fast as compared to other methods (Dedieu, 2015). The estimator of logistic regression model as below.

$$\begin{aligned} g(x) = {} & 5,42 - 2,443x_{1(1)} - 2,328x_{1(2)} - 0,831x_{1(3)} + 23,477x_{2(1)} + 1,985x_{2(2)} \\ & + 4,441x_{2(3)} + 0,199x_{2(4)} - 0,402x_{3(1)} - 0,187x_{3(2)} - 1,627x_{4(1)} \\ & + 1,113x_{4(2)} - 3,44x_{4(3)} - 0,496x_{5(1)} - 0,238x_{6(1)} - 0,252x_{7} \end{aligned} \tag{3}$$

From the results of testing the parameters simultaneously with the likelihood ratio (G) (Hosmer & Lemeshow, 2000) $G > \chi^2_{2,0.05}$ $(69,442 > 12,592)$ is obtained so that there are seven significant parameters in the model. Furthermore, partially using the Wald test, some parameters are not significant so that the estimation and testing of the parameters are carried out again until all the parameters were significant. The model is as follows

$$g(x) = 0,133 - 1,513x_{1(2)} + 0,76x_{2(1)} - 1,402x_{4(1)} \quad (4)$$

From equation (4) can be formed into the probability function of specific woman criminal π_0 (x) and general woman criminal π_1 (x) :

$$\pi_0(x) = \frac{1}{1+e^{0,133-1,513x_{1(2)}-1,402x_{4(1)}}}$$

$$\pi_1(x) = \frac{e^{0,133-1,513x_{1(2)}+0,76x_{2(1)}-1,402x_{4(1)}}}{1+e^{0,133-1,513x_{1(2)}+0,76x_{2(1)}-1,402x_{4(1)}}} \quad (5)$$

The equestion (5) states that woman criminal are affected by 17-25 years old, elementary education level, and unemployed or housewives.

2. Probit Regression on Woman Criminal

As with logit regression, the equestion of the probit regression parameter estimator is not closed form is as follows.

$$\sum_{i|y_i=0}^{N} \frac{-\Phi\left(\beta' x_i\right)}{1-\Phi\left(\beta' x_i\right)} x_i + \sum_{i|y_i=0}^{N} \frac{\Phi\left(\beta' x_i\right)}{\Phi\left(\beta' x_i\right)} x_i = 0 \quad (6)$$

Furthermore, this estimation is continued with Newton Raphson's iteration. This iteration obtains an unbiased estimator and close to the normal distribution (Greene, 2008).

$$\begin{aligned}\mathbf{Z} = {} & -2,902+1,361\ X_{1(2)}+1,277\ X_{1(3)}+0,385X_{1(4)}-8,261X_{2(1)} \\ & -1,196X_{2(2)}-2,531X_{2(3)}-0,176X_{2(4)}+0,234X_{3(1)}+0,096X_{3(2)}+0,845X_{4(1)} \\ & -0,662X_{4(2)}+1,861X_{4(3)}+0,259X_{5(1)}+0,102X_{6(1)}+0,141X_7 \end{aligned} \quad (7)$$

Based on simultaneous and partial parameter testing, there are several insignificant parameter so it is followed by estimating and testing the parameters until all parameters are significant. The probability function for woman criminal of specific crimes P(Y=0) and general P (Y=1) are as follows

$$\begin{aligned} P(Y=0) &= 1-\Phi(Z) \\ &= 1-\Phi\left(0,222-1,394\ X_{1(2)}-1,320X_{1(3)}-1,512X_{2(3)}+1,326X_{3(2)}+2,588X_{3(3)}\right) \\ P(Y=1) &= \Phi(Z) \\ &= \Phi\left(0,222-1,394\ X_{1(2)}-1,320X_{1(3)}-1,512X_{2(3)}+1,326X_{3(2)}+2,588X_{3(3)}\right) \end{aligned} \quad (8)$$

Based on these results, the factors that influence woman criminal are age of 17-25 years, 26-35 years, junior high school education level, married and ever married marital

status. This result is the same as the logit regression. There is one different factor in the logit regression, namely profession, while on probit, namely marital status.

3. The Comparison Results of Logit And Probit Regression on Woman Criminal

The logit and probit regression model have been estimated in the previous stage. The following table shows the factors that influence woman criminal.

Table 2. Influencing Factors of Logit and Probit Regression

Model	Influencing Factor	Category
Logit regression	age educational level profession	17-25 years old elementary school unemployed or housewives.
Probit regression	age educational level marital status	17-25 years old and 26-35 years old junior high school married and ever married

Table 2 shows there is the same of factor influence the woman criminal of logit and probit regression. The age and educational level predominantly affect woman criminal. They are in accordance with criminological and psychological theories that state biological and social factors that affect the woman criminal. The age represents biological factors and education level represents social factors. When we look at factor of profession and marital status from two regression, there are relevant to the results of previous research. The profession and marital status belong to social factors. Therefore, both models are appropriate in explaining the case of woman criminal.

The best model of this study is obtained by comparing the AIC value of the two regression models. The model with the smallest AIC value is the best model.

Table 3. Determination of The Best Model with AIC

Model	AIC
Logit regression	125,540
Probit regression	100,878

Based on Table 3 it can be seen that the probit regression is the best model because it has smaller AIC than logit regression model. Finally, we can state that factors affect the woman criminal are age, educational level, and marital status. The women aged 17-25 years are the age where psychologically immature, unstable. For women aged 26-35 years, they are leading the life of the household for the first time. There is a decrease in the rate of crimes committed throughout the age of marriage (Monsbakken et al., 2012). Many new things that force them to accept it. Therefore, marital status also influences a woman to commit a crime. Familial conflict forces women to marginal position in a society then female are engaging more with crime (Islam et al., 2015). Besides that, the result of this study shows that low education affects women to commit criminal acts. Compared to men, the majority of women who commit crimes have low education (Hjalmarsson & Lochner, 2012). There is relationship between education and marital status. The effects of education on crime for women related to changes in marital opportunities and family formation (Cano-Urbina & Lochner, 2019).

D. CONCLUSION AND SUGGESTIONS

Logit regression and probit are appropriate in determining the influencing factors on woman criminal. The factors resulting from the two models are relatively the same. However, the best model for the case of woman criminal is probit regression because it has the smallest AIC. The probit regression for women in specific and general crimes is as follow

$$\begin{aligned} P(Y=0) &= 1-\Phi(Z) \\ &= 1-\Phi\left(0,222-1,394\,X_{1(2)}-1,320X_{1(3)}-1,512X_{2(3)}+1,326X_{3(2)}+2,588X_{3(3)}\right) \\ P(Y=1) &= \Phi(Z) \\ &= \Phi\left(0,222-1,394\,X_{1(2)}-1,320X_{1(3)}-1,512X_{2(3)}+1,326X_{3(2)}+2,588X_{3(3)}\right) \end{aligned}$$

By using probit regression, factors that influence woman criminal are age, education level and marital status. These factors are in accordance with the criminological theory that the main causes of women committing criminal acts are biological, psychological and social (Mili et al., 2015). Age represents biological factors, education level represents social factors, marital status represents social and psychological factors. The result of this research only the influencing factors on woman criminal, does not come to criminological analysis. For example, predicting the female crime rate in the next years so that the analysis is more complete and clear.

E.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support from BOPTN LP2M UIN Walisongo Semarang.

REFERENCES

Adekanmbi, B. (2017). Comparison of Probit and Logit Models for Binary Response Variable with Applications to Birth Data in South-Western, Nigeria. *American Journal of Mathematics and Statistics*, *7*(5), 199–208. https://doi.org/10.5923/j.ajms.20170705.02

Agresti, A. (2002). *An Introduction to Categorical Data Analysis* (Second Edi). John Wiley & Sons, Inc Publication.

Antipov, E. (2019). *Applying CHAID for logistic regression diagnostics and classification accuracy improvement*. *January*. https://doi.org/10.2139/ssrn.1412208

Ari, E., & Aydin, N. (2016). *Examination By Multinomial Logistic Regression Model Of The Factors Affecting The Types Of Domestic Violence Against Women : A Case Of Turkey*. *5*(11).

Bain, L., & Engelhardt. (1992). *Introduction to Probability and Mathematical Statistics*. Duxbury Press.

Borrego, C. A., & Raquel, C. (2017). Employment and the risk of domestic violence: does the breadwinner's gender matter? *Applied Economics Journal*, *49*(50).

Cano-Urbina, J., & Lochner, L. (2019). The effect of education and school quality on female crime. *Journal of Human Capital*, *13*(2), 188–235. https://doi.org/10.1086/702927

Cochran, W. (1977). *Sampling Techniques*. John Wiley and Sons Inc.

Dedieu, J.-P. (2015). Newton-Raphson Method. *Encyclopedia of Applied and Computational Mathematics*, *6*(7), 1023–1028. https://doi.org/10.1007/978-3-540-70529-1_374

Emily, M. (2017). *Contributions to biostatistics : categorical data analysis , data modeling and statistical inference To cite this version : Contributions to biostatistics : categorical data analysis , data modeling and statistical inference Mathieu Emily*.

Greene, W. H. (2008). *Econometrics Analysis* (Fourth edi). Prentice Hall.

Hjalmarsson, R., & Lochner, L. (2012). The impact of education on crime: International evidence. *CESifo DICE Report*, *10*(2), 49–55.

Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression*. A John Wiley & Sons, Inc Publication.

Islam, M. J., Banarjee, S., & Khatun, N. (2015). Theories of Female Criminality: A criminological analysis. *International Journal of Criminology and Sociological Theory*, *7*(1), e39737–e39737.

Klieštik, T., Ko, K., & Mišanková, M. (2015). *Logit and Probit Model used For Prediction of Financial Health of Company*. *23*(October 2014), 850–855. https://doi.org/10.1016/S2212-5671(15)00485-2

Mieczyslaw, H. (2017). The principle of equality before the law and real property sales. *Scientific Journal WSFiP*, *1*, 146–174. https://doi.org/10.19192/wsfip.sj1.2017.10

Mili, P. M. K., Perumal, R., & Cherian, N. S. (2015). Female Criminality in India: Prevalence, Causes and Preventive Measures. *International Journal of Criminal Justice Sciences*, *10*(1), 65–75.

Miura, K. (2011). An Introduction to Maximum Likelihood Estimation and Information Geometry. *Interdisciplinary Information Sciences*, *17*(3), 155–174. https://doi.org/10.4036/iis.2011.155

Monsbakken, C., Lyngstad, T., & Skardhamar, T. (2012). Crime and the transition to marriage. The roles of gender and partner's criminal involvement. *27 S.*, *678*.

Pham, H. (2019). A new criterion for model selection. *Mathematics*, *7*(12), 1–12. https://doi.org/10.3390/MATH7121215

Snipes, M., & Taylor, D. C. (2014). Model selection and Akaike Information Criteria: An example from wine ratings and prices. *Wine Economics and Policy*, *3*(1), 3–9. https://doi.org/10.1016/j.wep.2014.03.001

Sousa, D. (2014). Validation in Qualitative Research: General Aspects and Specificities of the Descriptive Phenomenological Method. *Qualitative Research in Psychology*, *11*(2), 211–227. https://doi.org/10.1080/14780887.2013.853855

Widhiarso, W. (2012). Berkenalan dengan Regresi Probit. *Psikologi Universitas Gajah Mada*, 1–5.

Williams, R. (2016). Sociological Methods & Coefficients Across Groups. *Sociological Methods and Research*, *37*(May 2009), 531–559. https://doi.org/10.1177/0049124109335735

Yang, K., & Banamah, A. (2014). Quota sampling as an alternative to probability sampling? An experimental study. *Sociological Research Online*, *19*(1). https://doi.org/10.5153/sro.3199