Application of Binary Logistic Regression in Modelling Women's Participation in Improving the Welfare of Fishermen Families

Kariyam[1,*], Fatma Hanani[1]

[1]Departement of Statistics, Universitas Islam Indonesia
*Corresponding author. Email: kariyam@uii.ac.id

ABSTRACT
Indonesia is an archipelagic country with enormous marine potential. Development in the marine sector is expected to encourage recovery and improvement of the people's economy. This is reasonable because Indonesia has large fisheries resources which include aquaculture, seaweed cultivation, freshwater aquaculture, or marine biotechnology. However, these resources have not been optimized to improve the welfare of fishing communities. This situation urged some fishermen's wives to work in order to improve family welfare. This research discusses about application of binary logistic regression in modelling women's participation in improving the welfare of fishermen families. Variables that are predicted to influence the participation of fishermen wives in improving family welfare are age, education, number of members and family economic burden, average family income, asset ownership, and positive husband's attitude towards wife. The sample of this study was 159 fishing households on the southern coast of Indonesia. Based on the binary logistic regression approach, specifically the logit model and the probit model, the results of the analysis show that the number of family dependent, asset ownership, and positive husband attitudes influence the participation of women in improving the welfare of fishermen families. Based on the accuracy value and the Pseudo R-squared, the logit model and the probit model statistically show the same results for factors related to women's participation in improving the fisheries household economy.
Keywords: *binary logistic regression, probit, logit, women's participation, accuracy, pseudo R.*

1. INTRODUCTION

Fishing communities are a group of people who live in coastal areas, and economic resources depend on the potential of the sea. Diverse marine wealth is one of the strong natural resources supporting Indonesia's exports. However, the great potential of the sea has not been followed by the welfare of the fishing community, even fishermen contribute about 26% of Indonesia's poor population. [1]
Poverty is often a trade mark for fishermen, which is supported by facts such as slum settlement conditions, low income and education levels, their vulnerability to social and political changes, and their powerlessness towards incoming investor intervention. In general, fishermen are considered as a layer of society with low welfare. But of course not all fishermen can be said to be on the poverty line, for example fisherman boat owners. In the socio-economic structure, working fishermen/laborers are still considered as one of the layers of society with a low level of welfare. Fishermen work is a high-risk job so that it is more dominated by men and usually other family members cannot helpfully. Fishermen work depends on weather conditions, and the average working day in a month is twenty days at sea, while ten days are relatively unemployed. As a wife, the role of women or housewives becomes important so that family needs are met even with uncertain financial conditions.
This paper discusses the application of binary logistic regression to model and obtain factors that influence the participation of coastal fishing women in the household economy.

2. MATERIALS AND METHODS

The population in this study were all fisherman households in Gunungkidul Regency as many as 917 households [2]. In Panggang District, there are 110 households, in Saptosari District there are 212 households, in Tanjungsari District there are 233 households, in Tepus District there are 102 households and in Girisubo District there are 260 households. The number of samples taken in this study was taken randomly and calculated using the Slovin formula with an error tolerance of 0.075, $n = \frac{N}{1+N(d)^2}$ where n: number of samples; N: total population; and d: error tolerance which obtained a sample of 149 fisherman households [3]. Sampling was carried out by interviewing fisherman households in five sub-districts with a proportional number, namely 18 Sub-districts of Panggang, 34 of Saptosari, 38 of Tanjungsari, 17 of Tepus, and 42 of Girisubo. [4].

The dependent variable used in this paper is the participation of the fishermen's wife in the economic (Y), and fourteen independent variables. The variable names and their operational definitions are presented in Table 1. These variables were selected and developed based on previous studies relating to the factors of married women for work. Some variables are the result of the development of the previous variables. The variable is the number of college children (X_7); this variable is the development of the number of school children (X_6). Other variables, namely land ownership (X_{11}), house ownership (X_{12}), and ship ownership (X_{13}) are the development of the asset ownership variable. Reference for some variables used in this research is sourced from the results of previous studies, as shown in Table 1.

Table 1. Variables and their Operational Definitions

Variable Name	Definition	Category
Fishermen's wife in the economic (Y),	Wife's employment status	0: the wife doesn't work 1: the wife works
Husband's age (X_1)	husband's age is the age of the respondent's husband	-
Wife's age (X_2)	the age of the wife is the age of the respondent	-
Husband's education (X_3)	The last level of education attained by the husband	0: No school 1: Elementary School/ equivalent 2: Junior School/equivalent 3: Senior high school/ equivalent 4: Diploma 5: Bachelor
Wife's education (X_4)	The last level of education taken by the respondent	
Number of family members (X_5)	The number of all family members or dependents in one house.	-
Number of school children (X_6)	The number of all children attending kindergarten, elementary, junior and senior high schools who are dependent	-
Number of college children (X_7)	The number of all dependent children taking college	-
Average husband's income per month (X_8)	Average husband's income in one month	-
Average family expenses per month (X_9)	Average family expenses in one month	-
Other working family members (X_{10})	Family members other than husband and wife who have the status of working and helping the family economy.	0: There are no other family members working 1: there are other family members working
Land ownership (X_{11})	Land ownership status	0: not own 1: own
Homeownership (X_{12})	Homeownership status	
Boat ownership (X_{13})	Ship ownership status	
Positive husband's attitude (X_{14})	A form of positive support from the husband for the wife's decision to work.	0: the husband does not support the decision to work 1: the husband supports the decision to work

Table 2. Reference for Some Variables of Research

No.	Variable	Reference Number									
		[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]
1.	Husband's age (X_1)					√					
2.	Wife's age (X_2)	√	√	√	√	√		√	√	√	
3.	Husband's education (X_3)	√	√		√	√					
4.	Wife's education (X_4)	√	√	√	√	√	√	√	√	√	
5.	Number of family members (X_5)	√	√	√	√	√	√	√		√	
6.	Number of school children (X_6, X_7)					√					
7.	Average husband's income (X_8)	√	√			√	√		√	√	√
8.	Average monthly family expenses (X_9)		√			√					
9.	Other working family members (X_{10})	√									
10.	Asset ownership (X_{11}-X_{13})	√			√	√					
11.	Husband's support (X14)		√								

2.1. Binary Logistic Regression

Binary logistic regression is a statistical method used to model response variables consisting of two categories (denoted by Y) based on one or more predictor variables that are categorized or continuous (notated X). General logistic regression models such as equation (1). [15]

$$\pi(x) = \frac{\exp(g(x))}{1+\exp(g(x))} \quad (1)$$

2.2. Logit Model

The logit transformation in equation (1) is such as equation (2)

$$g(x) = \ln\left\{\frac{\pi(x)}{1-\pi(x)}\right\} \quad (2)$$

With $g(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$, and p is the number of predictor variables. The function of $g(x)$ s a logit estimator that acts as a linear function of the explanatory variable. Because the connecting function used is the logit connecting function, the distribution of opportunities used is called logistic distribution. [16]

The method used to estimate parameters is Maximum Likelihood Estimation (MLE), with likelihood functions such as equation (3)

$$L(\beta) = \prod_{i=1}^{n} f(y_i|\pi_i) \quad (3)$$

Overall test of parameters with L_0 shows likelihood without predictor variables, and L_1 the likelihood with predictor variables, then Likelihood Ratio Test is as in equation (4)

$$G = -2ln\left[\frac{L_0}{L_1}\right] \quad (4)$$

Partial test to determine the effect of each independent variable on the dependent variable, is done using Wald test statistics that are defined as equation (5)

$$W = \frac{\widehat{\beta_J}}{SE(\widehat{\beta_J})} \quad (5)$$

2.3. Probit Model

In general, the probit model can be written such as equation (6)

$$P_i = F(Z_i) = F(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i) \quad (6)$$

F is a cumulative opportunity function and X_{ij} is a stochastic independent variable. Estimation of the parameter using the MLE method with the assumption that the observations are mutually independent, then the likelihood function can be written, such as equation (7). [17]

$$L = \prod_{i=1}^{N} [1 - F(x_i'\beta)]^{1-yi} [F(x_i'\beta)^{yi} \quad (7)$$

2.4. Akaike's Information Criterion (AIC)

The smallest AIC shows the criteria of the best regression model, which is calculated based on *ln L(P*)*, which is MLE with $p*$ predictor variables, such as equation (8). [18]

$$AIC(P^*) = \frac{-2lnL(P^*)}{n} + \frac{2p^*}{n} \quad (8)$$

2.5. Accurate Values

The accuracy of the model prediction is a value that indicates the level of accuracy of the model in predicting the classification of observations. The accuracy value can be calculated based on the Apparent Error Rate (APER) such as equation (9). APER value is a comparison of the number of observations that are miss-classification with the total of all observations. [19]

$$Accuration = 1 - \frac{total\_missclassification}{n} \quad (9)$$

2.6. Pseudo R^2

The Pseudo R^2 is a value used to evaluate the goodness of the logistic model. One of the Pseudo R^2 values developed by McFaddens is a comparison of the full log likelihood model with a model without predictors, such as equation (10). The smaller of Pseudo R^2 value indicated that the full model with predictors is better than models without predictor variables.

$$R^2 = 1 - \frac{ln\hat{L}(M_{full})}{ln\hat{L}(M_{intercept})} \quad (10)$$

3. RESULTS AND DISCUSSION

Based on 159 respondents, as many as 89 fishermen's wives worked, and 57.89% of them worked as traders. Wives of fishermen who work as farmers are 18.9%, 12.2% as traders and farmers, 5.6% as laborers, and the else are other jobs. The wife of a fisherman who works as a trader mostly sells sea catches from her husband. Variable of the husband positive attitude in supporting his wife, obtained that 87% of the husband of the fisherman supported the wife to work, while 13% of the husband did not support the wife to work. As many as 52% of fishermen with primary school education, 28% with a junior high school education, 13% with a senior high school education, and 7% did not attend school. This shows that the majority of fishermen have an awareness of the importance of education. The last education of the fishermen's wife is 2% of undergraduate, 9% graduated from high school, 33% graduated from junior high school, 53% graduated from elementary school, and 3% did not graduate from elementary school.

3.1. Model Logit

Application of binary logistic regression for the logit model, in the overall test using the likelihood ratio test, it was concluded that there is at least one independent variable that influences the participation of fishermen's wives. Furthermore, a partial test to determine the effect of each independent variable, as listed in Table 3. and a logit model as in Table 3.

Table 3. Partial Test and AIC of Logit Model

Predictor	Estimate	Std. Error	*p-value*	AIC_Logit
X_4 (S_1)	6.790e-01	2.819e+03	0.9998	209.79
X_{13} (Y)	-3.272e-02	3.894e-01	0.9330	206.44
X_9	-3.899e-08	2.095e-07	0.8524	204.45
X_2	-1.906e-02	4.416e-02	0.6660	202.48
X_3 (SMA)	-5.618e-01	1.082e+00	0.6035	200.67
X_{12} (Y)	-6.031e-01	7.643e-01	0.4300	196.49
X_1	2.364e-02	2.489e-02	0.3423	195.12
X_7	1.208e+00	1.124e+00	0.2828	194.03
X_5	-1.893-01	1.699e-01	0.2652	193.46
X_{10} (Y)	5.110e-01	3.929e-01	0.1934	192.72
X_8	2.160e-07	1.482e-07	0.1449	192.44
X_{11} (Y)	0.8358	0.4137	0.0433	192.74

Table 4. Odds Ratio Values of The Best Logit Model

	Estimate	Std. Error	*p-value*	*Odds ratio*
Intercept	-2.1582	0.8179	0.0083	0.12
X_6	-0.5492	0.2258	0.0150	0.58
X_{11} (Y)	0.8358	0.4137	0.0433	2.31
X_{14} (Y)	2.5835	0.7783	0.0009	13.24

Based on Table 3 and Table 4. by using a significance level of 5%, it was concluded that the number of school children (X_6), land ownership (X_{11}) and positive husband's attitude in supporting working wives (X_{14}), influences the motivation of working wives to participate in supporting family economy. The best logit model has an AIC value of 192.74 with a Pseudo R^2 value of 0.1531. Thus it can be concluded that 15.31% of the variable number of school children, land ownership and positive attitude of the husband in supporting working wives are able to explain the motivation of a fisherman wife who works to improve the family economy. The accuracy of the logit model in predicting the motivation of working wives is 71.07% or as many as 113 cases have been correctly predicted by the logit model. The best logit model as in equation (11).

$$\pi(x) = \frac{\exp(-2.1582 - 0.5492X_6 + 0.8358X_{11} + 2.5835X_{14})}{1 + \exp(-2.1582 - 0.5492X_6 + 0.8358X_{11} + 2.5835X_{14})} \quad (11)$$

Based on the odds ratio values in Table 3. It was concluded that each increase of one schoolchildren's tendency for fishermen's wives to work was 0.58 times. Ownership of land itself influences the tendency of fishermen's wives to work 2.31 times compared to the status of ownership of non-own land. The positive attitude of the husband influences the tendency of the fishermen's wife to work 13.24 times compared to the husband, who does not support his wife to work.

3.2. Model Probit

As in the binary logistic regression with logit model, the overall test of the probit model concluded that there was at least one independent variable that affected the participation of fishermen's wives. Partial test as shown in Table 5. and probit model as shown in Table 6.

Table 5. Partial Test and AIC of Probit Model

	Estimate	Std. Error	*p-value*	AIC_Probit
X_4 (S_1)	3.279e-01	4.251e+02	0.9994	209.53
X_{13} (Y)	-4.101e-02	2.347e-01	0.8613	206.35
X_9	-4.323e-08	1.212e-07	0.7213	204.38
X_2	-1.312e-02	2.632e-02	0.6181	202.50
X_3 (SMA)	-3.636e-01	6.016e-01	0.5456	200.75
X_{12} (Y)	-3.282e-01	4.614e-01	0.4769	196.34
X_1	1.404e-02	1.493e-02	0.3473	194.83
X_7	7.268e-01	6.279e-01	0.2471	193.69
X_5	-1.184e-01	1.017e-01	0.2444	193.18
X_{10} (Y)	2.968e-01	2.338e-01	0.2042	192.55
X_8	1.308e-07	8.488e-08	0.1232	192.18
X_{11} (Y)	1.5538	0.2516	0.03925	192.49

Table 6. Odds Ratio Values of The Best Probit Model

	Estimate	Std. Error	*p-value*	*Odds ratio*
Intercept	-1.3033	0.4446	0.0034	0.27
X_6	-0.3334	0.1354	0.0138	0.72
X_{11} (Y)	0.5188	0.2516	0.0393	1.68
X_{14} (Y)	1.5538	0.4178	0.0002	4.73

Table 5. and Table 6. provide the same conclusions as the logit model. The best probit model has an AIC value of 192.49 with a Pseudo R^2 value of 0.1543. This means that 15.43% of the variable number of school children, land ownership and positive attitude of the husband in supporting working wives are able to explain the motivation of a fisherman wife who works to improve the family economy. The accuracy of the probit model in predicting the motivation of working wives is 71.07%, the same achievement as the logit model. The best probit model is as in equation (12).

$$Pi = F(Z) = F(-1{,}3033 - 0{,}3334X_6 + 0{,}5188X_{11} + 1{,}5538X_{14}) \quad (12)$$

The odds ratio value in Table 6. for the dependent variable of school children is 0.72; this means increasing the number of school children causes the tendency of the fishermen's wife to work by 0.72 times. Ownership of land itself affects the tendency of fishermen's wives to work by 1.68 times compared to the ownership status of non-own land. The positive attitude of the husband influences the tendency of the fishermen's wife to work by 4.73 times compared to the husband who does not support his wife to work.

To interpret the probit model equation (12) is slightly different from the logit model. In the probit model it is necessary to use the cumulative standard normal distribution table. For example, one of the data of fisherwomen with conditions that do not have dependents of school children, own land ownership and the husband supports the wife's decision to work, then the following calculations can be used.

Z = - 1.3033 - 0.3334 X6 + 0.5188 X11 +1.5538 X14
= -1.3033 - 0.3334 (0) + 0.5188 (1) + 1.5538 (1)
= 0.7693

The results above were fulfilled to be 0.77 then the value was looked for in the cumulative standard normal distribution table and the result was 0.7794. Thus, the opportunity for fisherwomen to work status is 77.94% and 22.06% not work.

3.3. *Comparison of Logit Models and Probit Models*

Some criteria that can be used to determine the best model are the AIC value, the accuracy value of prediction, and the value of R^2 Pseudo, with a summary of results as in Table 7. The logit model and the probit model give the same conclusion.

Table 7. Comparison of Measure The Best Model

Criteria	**Model_Logit**	**Model_Probit**
AIC	192.74	192.49
Accuration	71.07%	71.07%
R^2 Pseudo	15.31%	15.43%

4. CONCLUSION

The application of binary logistic regression with logit and probit models produces the same accuracy value of 71.07%. The R^2 Pseudo value for the two models is not different, namely for the logit model by 15.31% and the probit model by 15.43%. These results indicate that the logit model and the probit model produce the same conclusions, and both models can be used well to predict the factors that influence a fisherman's wife's work. Both models produce the conclusion that the number of school children, land ownership status, and positive attitude of the husband in supporting working wives influence the tendency of fishermen wives to work in supporting family welfare.

ACKNOWLEDGMENT

Thank you to all the coastal communities at Gunungkidul Yogyakarta Indonesia, who have agreed to be respondents. Thank you also said to the Department of Statistics of Universitas Islam Indonesia for supporting financial for the publication of this scientific work.

REFERENCES

[1] BPS. (2018). Statistik Sumber Daya Laut dan Perikanan 2018. Jakarta: Badan Pusat Statistik Indonesia

[2] BPS. (2018). Kabupaten Gunungkidul Dalam Angka 2018. Gunungkidul: Badan Pusat Statistik Kabupaten Gunungkidul Yogyakarta Indonesia.

[3] Riduwan. (2013). Skala Pengukuran Variabel-Variabel Penelitian. Bandung: Alfabeta.

[4] Sugiyono. (2015). Statistika untuk Penelitian. Bandung: Alfabeta.

[5] Azid, T. (2010). Labor Force Participation of Married Women in Punjab (Pakistan). International Journal of Social Economics. 37(8): 592-612.

[6] Bibi, A. (2012). Determinants of Married Women Labor Force Participation in Wah Cantt: A Descriptive Analysis. Academic Research International. 2(1): 599-622.

[7] Babalola, S.J. 2013. An Empirical Analysis of Labour Force Participation of Married Women in Adamawa State Nigeria. Journal of Emerging Trends in Economics and Management Science. 4(1): 1-7.

[8] Gayawan, E. (2015). Spatial Analysis of Women Employment Status in Nigeria. CBN Journal of Applied Statistics. 6(2): 1-17.

[9] Harlianingtyas, I. (2013). Pemodelan Partisipasi Wanita dalam Kegiatan Ekonomi Rumah Tangga Nelayan di Pesisir Timur Surabaya. Jurnal Sains dan Seni Pomits. 2(1).

[10] Havist, M., Kasman, K, dan Firdaus, Sy. (2014). Faktor-Faktor yang Mempengaruhi Keputusan Perempuan Berstatus Menikah untuk Bekerja. Jurnal Fakultas Ekonomi. 5(3).

[11] Listyandra, K. (2016). Kontribusi Wanita Nelayan Dalam Upaya Pemenuhan Kebutuhan Ekonomi Keluarga Nelayan di Muara Angke Kecamatan Penjaringan Jakarta Utara. Jurnal Perikanan Kelautan. 7(2): 80-90.

[12] Rizkia, F.N. (2017). Peran Perempuan dalam Meningkatkan Perekonomian Keluarga Melalui Program P2WKSS di Sumber Gamol, Balecatur, Gamping, Sleman. Yogyakarta: Jurnal Pendidikan IPS. 4(6): 406-418.

[13] Safariah. (2018). Faktor-Faktor yang Mempengaruhi Keputusan Perempuan Berstatus

Menikah untuk Bekerja. Jurnal Pendidikan Geografi. 5 (2): 19-24.

[14] Thamma-Apiroam, R. (2016). Factors Influencing the Labor Force Participation of Married Women in The United States. Asian Social Science. 12(3): 24-30.

[15] Hosmer, D.W., Lemeshow, S. (2000). Applied Logistic Regression. New York: Wiley

[16] McCullagh, P., Nelder, J.A. (1989). Generalized Linear Models 2nd Edition. London: Chapman & Hall.

[17] Agresti, A. (1990). Categorical Data Analysis. New York: John Wiley & Sons.

[18] Fathurahman, M. (2009). Pemilihan Model Regresi Terbaik Menggunakan Metode Akaike's Information Criterion dan Schwarz Information Criterion. Jurnal Informatika Mulawarman. 4(3): 37-41.

[19] Johnson, R.A., Wichern, D.W. (2007). Applied Multivariate Statistical Analysis, Sixth Edition, New Jersey: Prentice Hall.