Analysis of factors affecting the level of poverty in Central Sulawesi

L Handayani1*, Rais1 and F H A Putera2

1 Statistics Department, Tadulako University, Palu, Indonesia
2 Aquaculture Department, Tadulako University, Palu, Indonesia

*Email: lilies.stath@gmail.com

Abstract. Poverty is a problem faced both nationally and globally. Central Sulawesi Province is one of the provinces in Indonesia with an economic growth rate that is above the national rate, which is inversely proportional to the amount of poverty that occurs. This study aims to find a model using ordinal logistic regression analysis to predict predictor variables that affect poverty and rank the influencing predictor variables using dominant analysis. Based on 11 predictor variables that are thought to be closely related to the characteristics of poverty with the approach of poor households in Central Sulawesi province, ordinal logistic regression analysis concludes that seven predictor variables that affect it are $X_1$ (floor area), $X_4$ (roof), $X_5$ (latrine facilities), $X_6$ (drinking water), $X_9$ (credit), $X_{10}$ (assets), and $X_{11}$ (aid rice). The importance of the seven predictor variables can be ranked using dominant analysis. The dominant analysis concludes that the highest importance in influencing poverty in Central Sulawesi province is the variable $X_{11}$ (aid rice) followed by variables $X_6$ (drinking water), $X_5$ (latrine facilities), $X_1$ (floor area), $X_4$ (roof), $X_{10}$ (assets), and $X_9$ (credit).

1. Introduction

Indonesia is an archipelago with many provinces where each province has different characteristics of poverty. In March 2018, the number of poor people (residents with per capita expenditure per month below the poverty line) in Central Sulawesi province was 14.01%, this value was reduced compared to September 2017, which was 14.22%. However, the poverty rate is still far above the national poverty rate in Indonesia which reaches 9.82% [1]. The inconsistency between the increase in the rate of economic growth and the number of poor people in the province of Central Sulawesi is partly the result of the poverty reduction program by the Central Sulawesi provincial government that is not well-targeted. So it is necessary to have research on the factors that influence the level of poverty in Central Sulawesi province.

The statistical analysis used for the study of the relationship between response variables and predictor variables is regression analysis. If the response variable is on an interval or ratio scale, linear regression is used. If the response variable is categorical (nominal or ordinal) with continuous and categorical predictor variables, logistic regression is used [2].

This research uses a response variable, namely the level of poverty in the form of categorical data with an ordinal measurement scale where the poverty level is categorized as very poor, poor, moderate, and rich so that the appropriate analysis is logistic regression analysis. This research will examine inferential procedures for determining predictor importance in logistic regression models using the dominant analysis procedure to determine the importance ranking of the predictor variables that affect...
the response variable. Dominant analysis defines the importance of each predictor variable included in all possible models, so that the ranking of each predictor variable can be seen [3].

2. Materials and methods
The data used in this study is secondary data obtained from the Central Statistics Agency of Central Sulawesi Province. Where as many as 12 variables were selected which are thought to be closely related to the poverty level (Y) of a household, namely $X_1$ (floor area), $X_2$ (floor type), $X_3$ (wall), $X_4$ (roof), $X_5$ (latrine facilities), $X_6$ (water drinking), $X_7$ (lighting), $X_8$ (literacy) $X_9$ (credit), $X_{10}$ (assets), and $X_{11}$ (aid rice).

The method in this study uses ordinal logistic regression modeling, which is a statistical analysis method that describes the relationship between response variables using an ordinal scale with three or more categories with predictor variables which can be categorical or continuous data and consist of two or more variables. The model used in logistic regression analysis in general is [4]:

$$Logit(\pi_i) = \ln \left( \frac{\pi_i}{1-\pi_i} \right) = \theta + \beta x_i$$

(1)

To estimate the parameters of the ordinal logistic regression model, the maximum likelihood estimation (MLE) method can be used. The MLE method is an estimation method that maximizes the likelihood function. The basic idea of the MLE method is to find the parameter values that give the greatest likelihood to get the observed data as an estimator.

Initial ordinal logistic regression modeling is done by entering all the variables into the model then a series of tests is carried out to see the goodness of the resulting model and the significance of the included predictor variables. The goodness of fit test is used to test the accuracy of the sample regression function in estimating the actual value. The method often used for goodness of fit categorical data is deviance. The deviance statistic compares, in terms of likelihood, the model being fit with the saturated model [5]. Meanwhile, to test the significance of the model used the likelihood ratio test and wald test. The wald test compares the parameter estimate and its hypothetical value, and the likelihood ratio test calculates the difference between the log-likelihood function estimated at those two points [6].

Furthermore, remodelling was carried out by removing predictor variables which had no significant effect on the response variable. The resulting model is then tested just like the initial model. After that, checking the accuracy of the model using akaike's information criterion (AIC). AIC is a measuring tool to get the best model by looking at the minimum AIC value [7].

After obtaining the best logistic regression model, then proceed with the dominant analysis to determine the level of importance of the predictor variables contained in the logistic regression model. The basic principle of dominant analysis is by looking at the size of the contribution of a predictor variable to the response variable. This contribution can be seen through the $R^2$ value when the variable is entered into a model. $R^2$ is a measure of the percentage of total variation in the response variable that is accounted for by the predictor variable [8].

3. Results and discussion
3.1. Ordinal logistic regression modeling
The equation of the initial prediction model ordinal logistic regression with 11 predictor variables for the poverty variable below is a table that shows each parameter estimate obtained in the initial prediction model.

The results above explain that the significant parameter is the coefficient of the variables $X_1$, $X_4$, $X_5$, $X_6$, $X_9$, $X_{10}$, and $X_{11}$ because these variables have a $p$-value < 0.05. This means that the predictor variable has a significant effect on the response variable, namely the level of poverty. Meanwhile, the insignificant parameters are the coefficients of the variables $X_2$, $X_3$, $X_7$, and $X_8$. By obtaining information that the factors $X_2$ (floor type), $X_3$ (walls), $X_7$ (lighting), and $X_8$ (literacy) have no effect on...
poverty, it is necessary to improve the model by forming the next model by removing the insignificant predictor variable.

**Table 1. Initial model parameter estimation**

| Parameter | Estimasi | Wald Test | p-value          |
|-----------|----------|-----------|-----------------|
| Intercept 1 | -4.7887  | 27.3829   | < 0.0001        |
| Intercept 2 | -3.4534  | 14.4827   | 0.0001          |
| Intercept 3 | -0.4461  | 0.2457    | 0.6201          |
| X1        | -0.00850 | 35.4519   | < 0.0001        |
| X2        | -0.0610  | 0.0626    | 0.8024          |
| X3        | 0.0328   | 0.1846    | 0.6674          |
| X4        | 0.1016   | 4.1252    | 0.0422          |
| X5        | 0.6879   | 28.8742   | < 0.0001        |
| X6        | 0.1084   | 30.7951   | < 0.0001        |
| X7        | -0.0556  | 0.1302    | 0.7183          |
| X8        | 0.1985   | 0.6497    | 0.4202          |
| X9        | 0.6676   | 5.1932    | 0.0227          |
| X10       | -0.3531  | 7.1161    | 0.0076          |
| X11       | -0.7945  | 41.9378   | < 0.0001        |

The test used to test the feasibility of this initial model is the deviance test which is based on the likelihood ratio criteria with the hypothesis H₀: the model is suitable for use and H₁: the model is not suitable for use. It obtained a deviance value of 219.26978 with a p-value of 1.000 and α used of 0.05. The test criterion is reject H₀ if the p-value is < α (0.05). Because the p-value obtained is 1.00 > 0.05, H₀ is accepted or in other words the deviance test states that the initial model is feasible to use.

Furthermore, ordinal logistic regression modeling was carried out with 7 predictor variables that had a significant effect on the poverty variable. The following shows the parameter estimation results in the final estimate model.

**Table 2. Final model parameter estimation**

| Parameter | Estimasi | Wald Test | p-value          |
|-----------|----------|-----------|-----------------|
| Intercept 1 | -47225.0 | 42.9881   | < 0.0001        |
| Intercept 2 | -33889.0 | 22.7336   | < 0.0001        |
| Intercept 3 | -0.3836  | 0.2992    | 0.5844          |
| X3        | -0.00866 | 37.9833   | < 0.0001        |
| X4        | 0.1063   | 4.7842    | 0.0287          |
| X5        | 0.6989   | 32.6206   | < 0.0001        |
| X6        | 0.1083   | 32.1790   | < 0.0001        |
| X9        | 0.6666   | 5.2045    | 0.0225          |
| X10       | -0.3573  | 7.3026    | 0.0069          |
| X11       | -0.7991  | 42.5740   | < 0.0001        |
Based on the above results, it is found that the seven parameters are partially significant because the value of each parameter has a p-value < 0.05. From the model formed, the predictor variables that affect the response variable (poverty level) are \( X_1 \) (floor area), \( X_4 \) (roof), \( X_5 \) (latrine facilities), \( X_6 \) (drinking water), \( X_9 \) (credit), \( X_{10} \) (assets), and \( X_{11} \) (aid rice). Meanwhile, the deviance value for this model is 1739.8994 with a p-value of 1,000. Because the p-value obtained is greater than 0.05, it can be concluded that the final model formed is feasible to use.

To see the best model between the initial model and the final model, it can be seen using the minimum Akaike Information Criterion (AIC) value. In this research, the AIC value for the final prediction model was obtained at 2877.969 where this model has a smaller AIC value than the initial prediction model, which is 2884.986. This means that the final prediction model that was formed using 7 predictor variables was better than the initial prediction model that used 11 predictor variables. The ordinal logistic regression model obtained for cases of poverty level in Central Sulawesi province using the best model is as follows.

\[
\hat{\text{logit}}[P(Y \leq 1|X)] = -4.7225 - 0.00866(X_1) + 0.1063(X_4) + 0.6989(X_5) + 0.1083(X_6) + 0.6666(X_9) - 0.3573(X_{10}) - 0.7991(X_{11})
\]

\[
\hat{\text{logit}}[P(Y \leq 2|X)] = -3.3889 - 0.00866(X_1) + 0.1063(X_4) + 0.6989(X_5) + 0.1083(X_6) + 0.6666(X_9) - 0.3573(X_{10}) - 0.7991(X_{11})
\]

\[
\hat{\text{logit}}[P(Y \leq 3|X)] = -0.3836 - 0.00866(X_1) + 0.1063(X_4) + 0.6989(X_5) + 0.1083(X_6) + 0.6666(X_9) - 0.3573(X_{10}) - 0.7991(X_{11})
\]

Based on the ordinal logistic regression model, the odds ratio is obtained for the interpretation of the model as follows.

| Table 3. Odds ratio value for each variable |
|--------------------------------------------|
| Variabel | Odds Ratio |
|----------|------------|
| \( X_1 \) | 0.991 |
| \( X_4 \) | 1.112 |
| \( X_5 \) | 2.011 |
| \( X_6 \) | 1.114 |
| \( X_9 \) | 1.948 |
| \( X_{10} \) | 0.700 |
| \( X_{11} \) | 0.450 |

Based on the table above, it can be interpreted as follows:

a. Odds ratio \( X_1 \) (floor area) of 0.991 means that for every 1 meter increase in floor area, the ratio of poverty rates in Central Sulawesi province from very poor to poor or poor to moderate and moderate to rich decreased by 0.991 times. In other words, households with a floor area larger than 1 meter have a lower chance of being rich than households with a smaller floor area.

b. Odds ratio \( X_4 \) (roof) of 1.112 means that the change in the widest type of roof of the house occupied will result in a comparison of the poverty rate in Central Sulawesi province from very poor to poor or poor to moderate and moderate to rich, an increase of 1.112 times.
c. Odds ratio $X_5$ (defecation facilities) of 2.011 means that a change in the type of defecation facility from the house occupied will result in a comparison of the poverty rate in Central Sulawesi province from very poor to poor or poor to moderate and moderate to rich, an increase of 2.011 times.

d. Odds ratio $X_6$ (drinking water) of 1.114 means that the change in the type of main drinking water source from the inhabited house will result in a comparison of the poverty rate in Central Sulawesi province from very poor to poor or poor to moderate and moderate to rich, increasing by 1.114 times.

e. Odds ratio $X_9$ (credit) of 1.948 means that whether or not household members have received people's business credit from yes to no will result in a comparison of the poverty level in Central Sulawesi province from very poor to poor or poor to moderate and moderate to rich, an increase of 1.948 times.

f. Odds ratio $X_{10}$ (assets) of 0.700 means that the change in the status of residential ownership from self-owned to contract / lease and from contract / lease to lease-free and from free to lease to the office and from service to other will result in a comparison of poverty levels in the province of Central Sulawesi. from very poor to poor or poor to moderately and moderately to rich decreased by 0.700 times.

g. Odds ratio $X_{11}$ (raskin) of 0.450 means that with the change in having received Raskin in the last 6 months from yes to no, it will result in the ratio of poverty levels in Central Sulawesi province from very poor to poor or poor to moderate and moderate to rich decreasing by 0.450 times.

3.2. Dominant analysis

The rank of each predictor variable that dominates the poverty variable can be determined by using the dominant analysis and with the measure of the fit the model used is McFadden's R2 to determine the contribution of the addition of each predictor variable. This ranking uses 7 predictor variables that have a significant effect on the ordinal logistic regression model. In this study, it was concluded that there were no predictor variables that were completely dominance because there were no predictor variables that had a greater contribution value than other predictor variables in each model. The predictor variable for generally dominance can be seen through the following average overall contribution value ($G_i$).

| $K$ | $X_1$ | $X_4$ | $X_5$ | $X_6$ | $X_9$ | $X_{10}$ | $X_{11}$ |
|-----|-------|-------|-------|-------|-------|----------|----------|
| 0   | 0.043 | 0.023 | 0.047 | 0.050 | 0.003 | 0.009    | 0.058    |
| 1   | 0.021 | 0.024 | 0.015 | 0.016 | 0.032 | 0.030    | 0.013    |
| 2   | 0.043 | 0.035 | 0.045 | 0.048 | 0.025 | 0.028    | 0.052    |
| 3   | 0.029 | 0.031 | 0.029 | 0.027 | 0.035 | 0.035    | 0.026    |
| 4   | 0.032 | 0.030 | 0.032 | 0.034 | 0.029 | 0.030    | 0.035    |
| 5   | 0.023 | 0.023 | 0.024 | 0.023 | 0.024 | 0.024    | 0.023    |
| 6   | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014    | 0.014    |

$G_i$ | 0.029 | 0.026 | 0.029 | 0.030 | 0.023 | 0.024    | 0.032    |

The table above shows that the variable $X_{11}$ (aid rice) has the largest overall average value of 0.032, so it can be concluded that in general the variable $X_{11}$ (aid rice) has the highest level of importance compared to other predictor variables. Therefore, the variable $X_{11}$ (aid rice) is in the first rank in influencing Y (poverty) followed by variables $X_6$ (drinking water), $X_5$ (latrine facilities), $X_1$ (floor area), $X_4$ (roof), $X_{10}$ (assets), and $X_9$ (credit).

4. Conclusion

Predictor variables that have a significant effect on poverty levels in Central Sulawesi province based on ordinal logistic regression analysis are $X_1$ (floor area), $X_4$ (roof), $X_5$ (latrine facilities), $X_6$ (drinking water), $X_9$ (credit), $X_{10}$ (assets), and $X_{11}$ (aid rice).
water), $X_9$ (credit), $X_{10}$ (assets), and $X_{11}$ (aid rice). The ranking of predictor variables that have the highest to the lowest level of importance in influencing the poverty level in Central Sulawesi province based on the results of the dominant analysis is $X_{11}$ (aid rice), $X_6$ (drinking water), $X_5$ (latrine facilities), $X_1$ (floor area), $X_4$ (roof), $X_{10}$ (assets), and $X_9$ (credits).

**Acknowledgement**

Thanks to Tadulako University through the Faculty of Mathematics and Natural Sciences for funding this research so that this research can be carried out smoothly. Thank you also to the Central Statistics Agency of Central Sulawesi Province who has collaborated in providing the required research data and the Statistics Laboratory which has helped in the process of analyzing research data.

**References**

[1] Badan Pusat Statistik (BPS) 2019 *Sulteng Dalam Angka Provinsi Sulawesi Tengah* (Palu: BPS Indonesia)

[2] Hosmer D W and Lemeshow S 2000 *Applied Logistic Regression Second Edition* (Canada: John Wiley and Sons, Inc)

[3] Azen R and Traxel N 2009 *Journal of Educational and Behavioral Statistics* 34 (3) 319 - 347

[4] Tampil Y A, Komalig H and Langi Y 2017 *D’Cartesian Jurnal Matematika dan Aplikasi* 6 (2) 56 - 62

[5] Bellocco R and Algeri S 2013 *The Stata Journal* 13 (2) 356 - 365

[6] Dobek A, Molinski K and Skotarczak E 2015 *Biometrical Letters* 52 (2) 95 - 104

[7] Motulsky H and Christopoulos A 2003 *Sitting Models to Biological Data Using Linear and Nonlinear Regression* (San Diego: Grabh Pad Software, Inc)

[8] Manurung H T and Haryanto A M 2015 *Diponegoro Journal of Accounting* 4 (4) 1 - 16