

# Binary Logistic Regression Analysis of Variables that Influence Poverty in Central Java

## Research Article

Sendi Nugraha Nurdiansah\*, Laelatul Khikmah 

*Department of Statistics, Akademi Ilmu Statistika Muhammadiyah Semarang, Semarang 50185, Indonesia*

\**sandinugraha61@gmail.com (corresponden author)*  
*laelatul.khikmah@aismuh.ac.id*

**Article history:**  
Received: 1 Jan 2020  
Accepted: 18 Feb 2020  
Available online: 31 Mar 2020

## ABSTRACT

The phenomenon of poverty is a serious problem faced by almost every country in the world. The Central Java regional government incorporated poverty issues into the Regional Medium-Term Development Plan (RPJMD) because Central Java has a high number of poor people. This was done as an effort by the Central Java government to reduce poverty. The main objective in this research is how to find factors influence poverty in Central Java with the dichotomous categorical response variable, binary logistic regression analysis was used. The results showed that based on the analysis conducted did not obtain a logistic regression equation model because there were no significant parameters because there were no variables that had a sig value  $<0.05$ , so there are no variables that affect the level of poverty in Central Java.

**Keywords** : poverty, central java, binary logistic regression.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## 1. INTRODUCTION

The World Bank states that "Poverty is pronounced deprivation in well-being" (Poverty is a state of loss of well-being). The phenomenon of poverty is a serious problem faced by almost every country in the world. This is because poverty can affect various aspects of people's lives (Naranjo, 2012). According to (Group, 2016a), one of the causes of poverty is due to lack of income and assets to meet basic needs such as food, clothing, housing, health and education levels that are acceptable. In addition, poverty occurs because of the powerlessness of society to get out of the problems it faces. Therefore, community empowerment and improvement of community welfare in the regions are very important as the most important part of various policy strategies implemented by the regions (Aneta, 2010).

Indonesia is one of the countries that has a high level of poverty in the world (Group, 2016b). The number of poor people according to the archipelago in Indonesia with the following details: Java 14,789.43 thousand inhabitants, Sumatra 6,227.25 thousand inhabitants, Sulawesi 260.37 thousand inhabitants, Kalimantan 137.65 thousand inhabitants, Nusa Tenggara 71.64 thousand inhabitants, Papua 66.11 thousand inhabitants, and Bali 34.29 thousand inhabitants (Poruschi & Ambrey, 2019). Thus, Java is the island with the highest poverty population. Central Java is one of the provinces in Java which has a high number of poor people and high social inequality. Judging from the Provincial Gini Ratio data in 2002-2018 in the Central Statistics Agency, the level of inequality in Central Java has increased by 1.3% from the previous figure of only 0.365 in September 2017 to 0.378 in March 2018 which is a change in inequality the highest compared to Yogyakarta Special Region (BPS, 2019).

This was done as an effort by the Central Java government to reduce poverty. Therefore it is necessary to conduct research to find out the factors that most influence poverty in order to assist the government in developing the RPJMD.

Previous research on poverty has been carried out by (Nirwana & Istiawan, 2019) with the title "Mapping Regencies / Cities in Central Java Province Based on Poverty Levels Using the K-Harmonic Means Algorithm", using variables of population, female household heads, non-school children, disabled individuals, individuals who have chronic diseases, unemployment, unprotected drinking water sources, lighting sources other than electricity, kerosene and wood cooking fuels, defecation facilities (BAB) are not available (Nirwana & Istiawan, 2019). The results of research conducted there are 2 clusters of poverty, namely poor and not poor (Nirwana & Istiawan, 2019). The results of this cluster will be used as a category in the response variable in this study. Where poor as a success category and not poor as a failure category. Research using binary logistic regression was conducted by (Ramadhani, Faiz dan Zain, 2014) to

Identify Factors Affecting the Status of Poor Family Rice Acceptance (Raskin) in Gunung Anyar District. The results showed that the variables that influence the status of rice reception for poor families (Raskin) are gender, family members, population status, recent education, frequency of eating, floor area of residential buildings, chapel facilities, expenditure for food, asset ownership and frequency buy new clothes.

To find out what variables affect poverty in Central Java with the dichotomous category of response variables, binary logistic regression analysis was used. The binary logistic regression model is one of the logistic regression models used to analyze the effect of one response variable and several explanatory variables, with the response variable being a dichotomous categorical data consisting of two categories, namely category 1 to declare a success event and category 0 to declare a failure event, and the explanatory variables are categorical, numeric, or a combination of categorical and numeric data.

The researcher plans to conduct research with the aim of how to model poverty in Central Java with the binary logistic regression approach and find out what variables affect poverty in Central Java with the binary logistic regression approach.

## 2. THE MATERIAL AND METHOD

### 2.1 Data Source

The data used in this study are secondary data obtained from the study (Nirwana & Istiawan, 2019). This data is in the form of quantitative data, which are indicators of the welfare status of the Integrated Database (BDT) of Central Java in 2015 consisting of ten variables. Fig. 1, shows the flow diagram of the proposed method.

### 2.2 Variable's Used

The response variable (Y) from this study is poverty, where the response variable category 1 = poor and 0 = not poor. In this study, the variables used consisted of response and explanatory variables. The explanatory variables used in this study are the number of residents, female household heads, out of school children, disabled individuals, individuals with chronic diseases, unemployment, unprotected drinking water sources, lighting sources other than electricity, kerosene and wood cooking fuel, defecation facility (BAB) is not available.

### 2.3 Binary Logistics Regression

Binary logistic regression is a statistical modeling technique where the probability of categorical dependent variables is related to the independent variables that are numeric or categorical on the scale. Suppose the dependent variable has an M category (Agresti, 2003). One value

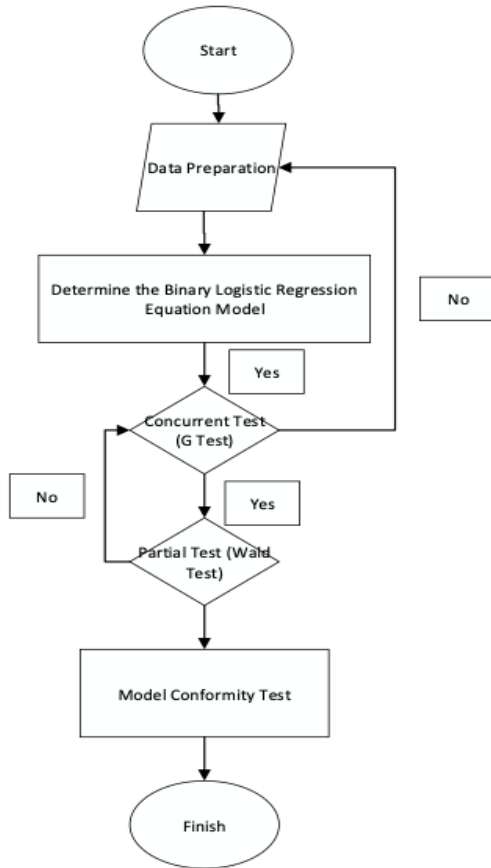


Fig. 1. Flow diagram of the proposed method.

(failure event) of the bound variable is designated as a category reference. The probability of membership in a category is compared to the probability of membership in a category reference (Agresti, 2003). In general the probability of logistic regression is shown in the equation.

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (1)$$

Where,  $\pi(x)$  is the probability of success category on the dependent variable, and is the estimating parameter for the independent variable.

### 3. RESULT AND DISCUSSION

#### 3.1 Simultaneous Significance Test Results

Simultaneous test is performed to test whether the explanatory variables together affect the response variable. The test used to test the significance of the model simultaneously uses the Likelihood Ratio Test obtained by comparing the Log Likelihood function using all explanatory variables with the Log Likelihood function without explanatory variables. Log-likelihood value with explanatory variables and without explanatory variables is obtained through iteration of processed results. The values obtained are as follows:

Based on (David & Stanley, 2000), the test statistics used for the likelihood ratio test are as follows:

$$G = -2 \ln \left[ \frac{l_0}{l_1} \right] = 0 \quad (2)$$

Table 1. Binary logistic regression log-likelihood value.

Log-likelihood value with explanatory variables	Log-likelihood value without explanatory variables
43.574	43.574

Based on the G test it was found that simultaneously the response variable had no effect on the explanatory variable, so partial testing could not be carried out either. So it is necessary to check the data again. Data checking is done by looking back at the relationship between explanatory variables with correlation analysis.

#### 3.2 Correlation Analysis

Table 2. The results of correlation analysis.

Explanatory variable	Correlation value	p-value	Conclusion
X1-X2	0,699	0,000	There is a correlation
X1-X3	0,891	0,000	There is a correlation
X1-X4	0,882	0,000	There is a correlation
X1-X5	0,798	0,000	There is a correlation
X1-X6	0,966	0,000	There is a correlation
X1-X7	0,612	0,000	There is a correlation
X1-X8	0,715	0,000	There is a correlation
X1-X9	0,716	0,000	There is a correlation
X1-X10	0,863	0,000	There is a correlation
X2-X3	0,507	0,002	There is a correlation
X2-X4	0,861	0,000	There is a correlation
X2-X5	0,865	0,000	There is a correlation
X2-X6	0,700	0,000	There is a correlation
X2-X7	0,543	0,001	There is a correlation
X2-X8	0,375	0,027	There is a correlation
X2-X9	0,700	0,000	There is a correlation
X2-X10	0,470	0,004	There is a correlation

Based on Table 2, it can be seen that the..... correlation analysis results is <0.05 which means there is a correlation between explanatory variables. Can also be seen that the resulting correlation value above 0.5, which means the correlation between explanatory variables is very strong. This allows the occurrence of multicoinerity, namely linear correlation between all independent variables. So that it allows modeling in logistic regression to be insignificant.

#### 4. CONCLUSION

The binary logistic regression equation model cannot be used to describe poverty in Central Java, because simultaneous testing and partial tests are not significant. It is possible for multicollinearity between explanatory variables. So it is necessary to handle multicollinearity first, for example with the principal component analysis approach.

#### REFERENCES

- Agresti, A. (2003). *Categorical Data Analysis* (2th ed.). John Wiley & Sons.
- Aneta, A. (2010). Implementasi Kebijakan Program Penanggulangan Kemiskinan Perkotaan (P2KP) di Kota Gorontalo. *Jurnal Administrasi Publik*, 1(1), 54–65.
- BPS. (2019). Data dan Informasi kemiskinan kabupaten/kota tahun 2018. *Badan Pusat Statistik*.
- David, W. H., & Stanley, L. (2000). *Applied logistic regression*. Willey.
- Group, W. B. (2016a). International Bank for Reconstruction. *International Centre For Settlement*, 91–92.
- Group, W. B. (2016b). World Bank Group. Poverty and shared prosperity 2016: Taking on inequality. *World Bank Publications*.
- Naranjo, S. (2012). Enabling food sovereignty and a prosperous future for peasants by understanding the factors that marginalise peasants and lead to poverty and hunger. *Agriculture and Human Values*, 29(2), 231–246. <https://doi.org/10.1007/s10460-011-9345-7>
- Nirwana, R. S., & Istiawan, D. (2019). Pemetaan Kabupaten / Kota Di Provinsi Jawa Tengah K-Harmonic Means District / City Mapping In Central Java Province Based On Poverty. *The 9th University Research Colloquium (Urecol)*, 9(5), 236–243.
- Poruschi, L., & Ambrey, C. L. (2019). Energy justice, the built environment, and solar photovoltaic (PV) energy transitions in urban Australia: A dynamic panel data analysis. *Energy Research and Social Science*, 48(December 2017), 22–32. <https://doi.org/10.1016/j.erss.2018.09.008>
- Ramadhani, Faiz dan Zain, I. (2014). Analisis Regresi Logistik Biner Untuk Mengidentifikasi Faktor-Faktor Yang Mempengaruhi Status Penerimaan Beras Keluarga Miskin (Raskin) di Kecamatan Gunung Anyar. *Digital Library ITS*, 1–7.