pijmath
Volume 04, issue 01, pages 041-052, May 2025
Pattimura International Journal of Mathematics (PIJMath)
E-ISSN: 2830-6791
RESEARCH ARTICLE ◙ OPEN ACCESS
DOI: https://doi.org/10.30985/pijmathvol4iss1pp41-52

Factor Analysis on Poverty in Kalimantan Island with Geographically Weighted Negative Binomial Regression

Alvin Octavianus Halim[1*], Neva Satyahadewi[2], Preatin[3]

[1,2]Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Tanjungpura
Jl. Prof. Dr. H Hadari Nawawi, Pontianak, 78124, Indonesia

[3]BPS-Statistics of Kalimantan Barat Province
Jl. Sultan Syahrir, Pontianak, 78116, Indonesia

Corresponding author's e-mail: [1]*neva.satya@math.untan.ac.id

ABSTRACT

Article History

Received: March 11th, 2025
Revised: March 28th, 2025
Accepted: March 31st, 2025
Published: May 1st, 2025

Keywords

GWNBR;
Kernel Function;
Negative Binomial Regression;
Poisson Regression;
Poverty;

Poverty is one of the problems still faced by Indonesia. The problem of poverty is a development priority because poverty is a complex and multidimensional problem. Therefore, to reduce poverty, it is necessary to know the factors that influence the number of people living in poverty. The influencing factors in each region are different due to the effects of spatial heterogeneity between regions such as geographical, economic, and socio-cultural conditions. This research considers spatial factors by using the Geographically Weighted Negative Binomial Regression (GWNBR) method on poverty-based regions in Kalimantan Island. This research uses eleven independent variables. The weighting function used is the Adaptive gaussian kernel because the adaptive kernel can produce the number of weights that adjust to the distribution of observations. The stage starts with descriptive statistics and checking multicollinearity. Then proceed with the formation of Poisson Regression, because the data used is enumerated data. Then check for overdispersion. If overdispersion is detected where the variance is bigger than the mean, then Negative Binomial Regression is continued. After that, it is tested for the presence or absence of spatial heterogeneity. If there is, proceed to find the bandwidth and Euclidean distance. After that, the graphical weighting matrix is searched. Then proceed with GWNBR modeling. The results of the analysis show that there are seven significant variables, including the percentage of households with the main source of lighting is non-state electricity company (PLN), average monthly net income of informal workers, population density for every square kilometer, monthly per capita expense on food and non-food essentials, percentage of people who have a health complaint and do not treat it because there is no money and percentage of population 15 years and above who do not have a diploma. Based on the categories of significant variables, six groups were formed in 56 districts/cities in Kalimantan Island.

How to cite this article:
Halim A. O., Satyahadewi N., Preatin, "FACTOR ANALYSIS ON POVERTY IN KALIMANTAN ISLAND WITH GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION", *Pattimura Int. J. Math. (PIJMATH).*, vol. 04, iss. 01, pp. 041-052, May, 2025.

e-mail: pijmath.journal@mail.unpatti.ac.id
Homepage https://ojs3.unpatti.ac.id/index.php/pijmath

1. Introduction

Poverty is one of the problems that Indonesia still faces. Poverty is a development priority because poverty is a complex and multidimensional problem [1]. Many programs have been attempted by the government to create community welfare, in the form of poverty alleviation program [2]. For the ideals of the nation to be realized and the creation of a prosperous and just society, poverty alleviation efforts are needed. To support poverty reduction strategies, accurate data is required. If the data is available, then the government can make decisions that are carried out for the reduction [1]. The Central Bureau of Statistics (BPS) started measuring the number and percentage of people living in poverty back in 1984. Since 2003, they have consistently published data on how many people are poor and what percentage of the population that represents.

The Kalimantan Island has vast natural resources, but it suffers from an unequal distribution of wealth in the region. According to the Statistics of Indonesia (BPS), in March 2023, Kalimantan Island had the lowest percentage of poor people compared to other islands at 5.67% in Indonesia. The reason to choose Kalimantan because its relatively low poverty rate allows writer to analyze key success factors in poverty reduction. Additionally, writer familiarity with the region enables a deeper and more accurate contextual analysis. To measure poverty, BPS uses the ability to fill basic needs. Several interrelated factors cause poverty. For example, education is the main thing to achieve a more decent life and health which is the core of welfare. In addition, regional development inequality is caused by differences in the content of natural resources, which affects economic growth in each region. Differences in geographical conditions among regions contribute to the variation in the number of poverty cases observed in different locations [3]. Equitable Good development is expected to provide good services.

Kalimantan has a large area coverage, so it has different environments in each region. Spatial heterogeneity such as differences in geographical, socio-cultural, and economic conditions cause spatial diversity. The differences between regions also cause each region to have characteristics related to environmental conditions, behavior, and community knowledge [4]. The factors that cause poverty will be different in each region depending on the characteristics of the region. One analysis that can be done to find out what influences poverty is by using regression analysis. [5] research shows that Human Development Index and open unemployment rate have a significant effect on poverty in Kalimantan Island. Then [6] research also shows that the percentage of households with a main lighting being a non-state electricity company, average monthly net income of informal workers, population density for every square kilometer, monthly per capita expense on food and non-food essentials, the percentage of people who have a health complaint and do not treat it because there is no money and the percentage of the population aged 15 and above without a diploma have a significant effect on poverty in Kalimantan Island.

The data used in this analysis is usually continuous, but there are several cases, especially in the dependent variable, which is discrete data. Discrete count data consists of nonnegative values representing the number of events within a specific time, space, or volume [7]. This type of data often appears in cases like the number of people living in poverty, making Poisson Regression a suitable approach. However, Poisson Regression assumes equidispersion, where the mean equals the variance. When overdispersion occurs (variance exceeds the mean), Negative Binomial Regression provides a suitable solution [8].

GWNBR is a geographically weighted regression method designed for count data with overdispersion, where variability exceeds the mean. Its key advantage is capturing spatial variations while addressing distribution imbalances, making it ideal for poverty analysis and socioeconomic studies. By allowing regression coefficients to vary across locations, GWNBR provides more accurate estimations. Unlike other methods, it effectively handles overdispersion, reducing bias in results. Thus, GWNBR is the best choice for spatial count data with uneven distribution [9].

2. Research Methods

The following are the research stages in GWNBR, starting with the research methodology as an introduction

2.1 Multicollinearity

Multicollinearity occurs when independent variables are highly correlated, leading to inaccurate estimations [10]. The Variance Inflation Factor (VIF) is used to detect multicollinearity and is calculated using the following equation [6]:

$$VIF_k = \frac{1}{1-R_k^2}, k = 0,1,2, \dots, n \tag{1}$$

VIF serves as an indicator of the degree to which multicollinearity inflates the variance of an estimator, where R_k^2 is the coefficient of determination. A VIF value below 10 indicates the absence of multicollinearity [10] If it is above 10, then erase the variable with highest number first and check again.

2.2 Poisson Regression

Poisson regression is a widely used model for count data, assuming equal mean and variance [11]. Poisson regression is a nonlinear model that assumes a Poisson-distributed dependent variable [12]. The Poisson distribution represents the count of events occurring within a fixed time interval or area [13]. If the variable (Y) follows a Poisson distribution, its probability function is given by [3]:

$$f(y,\mu) = \frac{e^{-\mu}\mu^{y}}{y!}, y = 0,1,2,\dots,n \tag{2}$$

with μ is the mean of the dependent variable which is Poisson distributed where the mean value and variance are bigger than zero. The Poisson Regression Equation can be written as follows [8]:

$$\mu_i = \exp\left(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}\right),\ i = 1,2,\dots,n \tag{3}$$

where μ_i represents the Poisson Regression equation with X_{ik} denotes the independent variable and β_k is the regression parameter to be estimated.

The parameters of the Poisson Regression model are estimated using the Maximum Likelihood Estimation (MLE) method. The likelihood function for Poisson Regression is formulated as follows [11]:

$$L(\beta) = \prod_{i=1}^{n} \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \tag{4}$$

In Poisson Regression, the parameter to be estimated is β_k. To obtain the estimated values, the likelihood function of the Poisson distribution is constructed, followed by parameter estimation using the Newton-Raphson iteration method [8].

The next step is parameter testing, conducted both simultaneously and partially. The simultaneous test evaluates the combined effect of independent variables on the dependent variable [13]. Simultaneous testing is carried out using the likelihood ratio test with the following hypothesis [10]:

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_k = 0$
H_1 : at least there is one $\beta_k \neq 0, k = 1,2,\dots,n$
The test statistic used is:

$$D(\hat{\beta}) = -2ln\left[\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right] = 2lnL(\widehat{\Omega}) - 2lnL(\widehat{\omega}) \tag{5}$$

where $D(\hat{\beta})$ represents the deviance value of the Poisson Regression model. $L(\widehat{\omega})$ is the likelihood value for a simple model without independent variables and $L(\widehat{\Omega})$ is the likelihood value for the complete model with independent variables. The significance test criterion for the Poisson Regression model is to reject H_0 if $D(\hat{\beta}) > \chi^2_{(k;\alpha)}$ [10].

Next, a partial test is conducted using the Wald test to evaluate the significance of each independent variable on the dependent variable. The hypotheses for the Wald test are as follows [13]:

H_0 : $\beta_k = 0$
H_1 : $\beta_k \neq 0, k = 1,2,\dots,n$
The test statistic used is:

$$Z_{count} = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \tag{6}$$

with $k = 1,2,\dots,n$ then $\hat{\beta}_k$ is the estimated value of β_k and $SE(\hat{\beta}_k)$ is the standard error of $\hat{\beta}_k$. The test criterion is to reject H_0 if $Z_{count} > Z_{\frac{\alpha}{2}}$ or $Z_{count} < -Z_{\frac{\alpha}{2}}$ where $Z_{\frac{\alpha}{2}}$ is obtained from the normal distribution table [10]. The significance level (α) used in this study is 10%. This choice increases the test's power, enhancing the likelihood of detecting a true effect.

2.3 Overdispersion

Poisson regression must fill the assumption of equidispersion where the variance and mean values must be the same. However, Poisson Regression violates the assumption of equidispersion, as overdispersion occurs when the variance exceeds the mean [11]. Overdispersion can cause the standard error value of the estimation results to be biased, resulting in inconsistent estimation of model parameters.

Overdispersion can be detected using the pearson chi-square dispersion value or deviance divided by the degrees of freedom ($\boldsymbol{n-k-1}$). If the result is bigger than one, overdispersion is present. if it is equal to one, the data exhibits equidispersion [12]. If overdispersion occurs, Negative Binomial Regression is used as an alternative [11].

2.4 Negative Binomial Regression

Negative Binomial Regression is used to model count data that exhibit overdispersion [3]. This method assumes that the dependent variable follows a negative binomial distribution, which originates from a mixed Poisson-Gamma distribution. The probability function of the negative binomial distribution is expressed as follows [11]:

$$\mathbf{f(y,\mu,\theta)} = \frac{\Gamma\left(y+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y!}\left(\frac{1}{1+\theta\mu}\right)^{\frac{1}{\theta}}\left(\frac{\theta\mu}{1+\theta\mu}\right)^{y}, \mathbf{y = 0, 1, 2, \dots, n} \tag{7}$$

Negative Binomial Regression has a similar structure to Poisson Regression, as shown in Equation 3. The model parameters are estimated using the Maximum Likelihood Estimation (MLE) method. The likelihood function for Negative Binomial Regression is formulated as follows [14]:

$$\boldsymbol{L(\beta,\theta)} = \prod_{i=1}^{n}\left\{\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!}\left(\frac{1}{1+\theta\mu_i}\right)^{\frac{1}{\theta}}\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i}\right\} \tag{8}$$

After that, form the log function of the likelihood function in Equation 8 so that it becomes as follows [11]:

$$\boldsymbol{Ln\ L(\beta,\theta)} = \sum_{i=1}^{n}\left\{\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y_i!} + \theta^{-1}\left(\frac{1}{1+\theta\mu_i}\right)^{\frac{1}{\theta}} + y_i ln\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i}\right\} \tag{9}$$

Then Equation 9 is derived for each parameter $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ then equated to zero and solved by Fisher-Scoring iteration. The next process is parameter significance testing which consists of partial and simultaneous testing. Testing the significance of parameters in Negative Binomial Regression has the same method as Poisson Regression.

2.5 Spatial Heterogeneity

Spatial heterogeneity testing is conducted to identify differences in characteristics between observation points, which contribute to spatial variability in the data. This test helps determine whether a global model is sufficient or if a geographically weighted approach, such as GWNBR, is necessary to account for spatial heterogeneity [8]. The spatial heterogeneity test can be done with the Breusch-Pagan (BP) test. The following hypotheses are [15]:

H_0 : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (Variance between place is same)
H_1 : at least there is $\sigma_n^2 \neq \sigma^2$ (Variance between place is different)

The test statistic used is:

$$\boldsymbol{BP} = \left(\frac{1}{2}\right)\boldsymbol{f^T Z(Z^T Z)^{-1} Z^T f \sim \chi^2_{(k;\alpha)}} \tag{10}$$

where:

f : $(f_1, f_2, \dots, f_n)^T$ with $f_i = \frac{e_i^2}{\sigma^2} - 1$
$\boldsymbol{Z}$: matrix size n x $(k+1)$ contains normalized vectors (Z) for every observation, with k is the number of independent variables
e_i : $(y_i - \hat{y}_i)$, residual observation i
σ^2 : sample of y variance

The criteria for testing are to reject $\boldsymbol{H_0}$ if the value of $\boldsymbol{BP > \chi^2_{(k;\alpha)}}$ or *p-value* < α, this indicates the presence of spatial heterogeneity in the model, meaning that variance differs across locations.

2.6 Bandwidth and Optimum Weighting Function

Weighting is essential at each observation location because the variability of data differs across locations. In GWNBR, weighting plays a crucial role as it reflects the spatial relationship between observation points, ensuring that closer locations have a greater influence on parameter estimation than those farther away [16]. Weighting is determined using the bandwidth value and Euclidean distance. The bandwidth represents an area with a radius c_i from the center point, used to assign weights to each observation in the regression model at that location. This study employs spatial weights using an adaptive Gaussian kernel function. The adaptive Gaussian kernel is chosen because it generates weights that adjust to the distribution of observations [17]. Additionally, it is well-suited for datasets with dispersed and irregular spatial patterns [17]. One of the optimal weighting functions is the adaptive Gaussian kernel function, which assigns a unique bandwidth to each observation location. The weighting for the adaptive Gaussian kernel function is calculated using the following formula [18]:

$$\boldsymbol{w_{ij}} = \begin{cases} exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{c_i}\right)^2\right], \text{if } d_{ij} \le c_i \\ 0, \text{if } d_{ij} > c_i \end{cases} \quad (11)$$

where:

i, j : observation location $1,2, \dots, n$
c_i : optimum bandwidth for *i* location observation
d_{ij} : Euclidean distance between *i* observation location and *j* observation location

In GWNBR, Euclidean distance is used to calculate spatial weights between observation locations. These weights are then used in the kernel function to give bigger influence to data points that are closer to the location being analyzed. The use of Euclidean distance allows the model to capture local variation more accurately, as the influence of the data is determined based on its geographical proximity [17]. To get the d_{ij}, use the following equation:

$$\boldsymbol{d_{ij}} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \quad (12)$$

where:

u_i : Latitude value for observation location *i*
u_j : Latitude value for observation location *j*
v_i : Longitude value for observation location *i*
v_j : Longitude value for observation location *j*

Selecting the optimum bandwidth is crucial, as it balances the trade-off between model fit and data smoothness [16]. The selection of the optimum bandwidth is also important because it affects the accuracy of the model to the data. The optimum bandwidth value calculation uses the Cross Validation (CV) criterion using the following equation:

$$CV(c_i) = \sum_{i=1}^{n}\left(y_i - \hat{y}_{\neq i}(c_i)\right)^2 \quad (13)$$

where $\hat{y}_{\neq i}(c_i)$ Is the estimated value for y_i with the observations of the location (u_i, v_i) omitted from the estimation process.

2.7 Geographically Weighted Negative Binomial Regression

An effective approach for modeling count data with overdispersion while accounting for spatial heterogeneity is Geographically Weighted Negative Binomial Regression (GWNBR). GWNBR extends the Negative Binomial Regression model by incorporating spatial variations, resulting in location-specific parameters that vary across different geographic areas [19]. The GWNBR model can be formulated as follows [19]:

$$\mu_i = \exp\left(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)X_{1i} + \cdots + \beta_k(u_i, v_i)X_{ki}\right) \quad (14)$$

where:

x_{ki} : observation values independent variable *k* for every location (u_i, v_i)
$\beta_k(u_i, v_i)$: regression coefficient independent variable *k* for every location (u_i, v_i)
$\theta(u_i, v_i)$: dispersion parameter for every location (u_i, v_i)

Then the negative binomial distribution function for each location can be seen in Equation 15 below [19]:

$$f(y_i|x_i, \beta(u_i, v_i), \theta) = \frac{\Gamma\left(y_i + {}^1/_\theta\right)}{\Gamma\left({}^1/_\theta\right)y_i!}\left(\frac{1}{1+\theta\mu_i}\right)^{{}^1/_\theta}\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i} \quad (15)$$

with $i = 1,2, \dots, n$ and $\mu_i = exp\left(\boldsymbol{X_i}^T\beta(u_i, v_i)\right)$ and $\theta_i = \theta(u_i, v_i)$.

The parameters of the GWNBR model are estimated using the Maximum Likelihood Estimation (MLE) method. The likelihood equation for GWNBR is as follows [19]:

$$L(.) = \prod_{i=1}^{n}\frac{\Gamma\left(y_i + {}^1/_{\theta(u_i,v_i)}\right)}{\Gamma\left({}^1/_{\theta(u_i,v_i)}\right)y_i!}\left(\frac{1}{1+\theta(u_i,v_i)exp\left(X_i{}^T\beta(u_i,v_i)\right)}\right)^{{}^1/_{\theta(u_i,v_i)}}\left(\frac{\theta(u_i,v_i)exp\left(X_i{}^T\beta(u_i,v_i)\right)}{1+\theta(u_i,v_i)exp\left(X_i{}^T\beta(u_i,v_i)\right)}\right)^{y_i} \quad (16)$$

where $L(.) = L(\beta(u_i, v_i), \theta_i|y_i, x_i)$.

The next process is testing the significance of the GWNBR model parameters partially and simultaneously. The hypothesis for testing the significance of model parameters simultaneously is [9]:

H_0 : $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_k(u_i, v_i) = 0$
H_1 : at least one $\beta_k(u_i, v_i) \neq 0\,;\, j = 1,2, \dots, k$

The test statistic used is:

$$D\left(\hat{\beta}(u_i, v_i)\right) = 2\left(lnL\left(\hat{\hat{\Omega}}\right) - L(\widehat{\omega})\right) \tag{17}$$

The criteria for testing using the significance level α, where reject H_0 if $D\left(\hat{\beta}(u_i, v_i)\right) > \chi^2_{(k;\alpha)}$ or if the *p-value* < α. If the simultaneous parameter significance test is obtained to reject H_0, then proceed with a partial test with the following hypothesis [9]:

H_0 : $\beta_k(u_i, v_i) = 0$
H_1 : $\beta_k(u_i, v_i) \neq 0\,;\, k = 1,2, \dots, n$

The test statistic used is:

$$Z_{count} = \frac{\widehat{\beta}(u_i, v_i)}{SE\left(\widehat{\beta}(u_i, v_i)\right)} \tag{18}$$

The criteria for testing the significance of parameters partially using the significance level α, namely reject H_0 if $|Z_{count}| > Z_{\frac{\alpha}{2}}$ or if the *p-value* < α.

3. Results And Discussion

The data utilized in this study is secondary data obtained from BPS publications for each province in Kalimantan Island named “Provinsi Kalimantan Dalam Angka 2024”. The data used consists of the dependent variable (Y) which is the number of people living in poverty by district/city. Meanwhile, the independent variables consist of the percentage of households with the main source of lighting being a non-state electricity company (PLN) (X_1), the percentage of households with access to adequate and reliable drinking water services (X_2), the percentage of households with access to proper and sustainable sanitation services (X_3), the percentage of households that have no facility of toilet facilities (X_4), the Human Development Index (HDI) (X_5), open unemployment rate (X_6), average monthly net income of informal workers (X_7), population density for every square kilometer (X_8), monthly per capita expense on food and non-food essentials (X_9), the percentage of people who have a health complaint and do not treat it because there is no money (X_{10}) and the percentage of the population aged 15 and above without a diploma (X_{11}). The variables used are divided into three categories, these categories shown in **Table 1** below:

Table 1. Variables by Categories

Variables	Categories
X_1, X_2, X_3, X_4 and X_8	Housing/Environment
X_5 and X_{11}	Human
X_6, X_7, X_9 and X_{10}	Economy

Then continue to the multicollinearity check. Multicollinearity arises when independent variables in a regression model exhibit a high degree of correlation. It is typically evaluated using the Variance Inflation Factor (VIF), with a VIF value greater than 10 suggesting the presence of multicollinearity. This analysis explores multicollinearity among variables influencing the number of people living in poverty on Kalimantan Island.

Table 2. Variance Inflation Factor

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
VIF	2.24	2.09	2.30	1.85	6.68	2.58	3.74	2.11	3.39	1.51	3.36

Based on **Table 2**, the results indicate that all VIF values are below 10, suggesting no multicollinearity among the independent variables influencing the number of people living in poverty in Kalimantan Island. Because all independent variables do not contain elements of multicollinearity, we can proceed to the next stage.

The next step is Poisson Regression modeling. Since the number of people living in poverty represents count data, Poisson Regression is used for analysis. The estimation results of the Poisson Regression model for the number of people living in poverty are presented in **Table 3** below:

Table 3. **Poisson Regression Modelling**

Variable	Estimation	Standard Error	$\lvert Z_{count} \rvert$	*p-value*
Intercept	2.45	1.42	1.71	0.08
X_1	-0.02	5.60×10^{-3}	-3.90	9.37×10^{-5}
X_2	-8.18×10^{-3}	3.34×10^{-3}	-2.44	0.01
X_3	9.04×10^{-3}	4.71×10^{-3}	1.91	0.05
X_4	-0.01	0.01	-0.83	0.40
X_5	-5.01×10^{-3}	0.02	-0.23	0.81
X_6	0.06	0.02	2.52	0.01
X_7	0.52	0.06	7.62	2.43×10^{-14}
X_8	2.53×10^{-4}	3.32×10^{-5}	7.63	2.33×10^{-14}
X_9	-0.60	0.19	-3.05	2.28×10^{-3}
X_{10}	0.25	0.03	7.13	1×10^{-12}
X_{11}	0.03	9.14×10^{-3}	4.01	6.06×10^{-5}
Deviance	284.87		DF	44

Simultaneous parameter significance testing at a 10% significance level resulted in a deviance value of 284.47, which exceeds the critical value of $\boldsymbol{\chi^2_{(11:0.1)}} = \mathbf{17.275}$. This indicates that the null hypothesis $\boldsymbol{H_0}$ is rejected, suggesting that at least one parameter is statistically significant to the model. Next, partial significance testing is performed to evaluate the individual effect of each independent variable on the dependent variable. Using a 10% significance level, the critical Z-value is $\boldsymbol{Z_{\left(\frac{0.1}{2}\right)}} = \mathbf{1.64}$. This value is compared with the absolute Z-value $\lvert \boldsymbol{Z_{count}} \rvert$ of each independent variable. Based on **Table 3**, two independent variables, $\boldsymbol{X_4}$ and $\boldsymbol{X_5}$ are found to be insignificant. The model formed as follows:

$$\mu_i = exp(2{,}45 - 0{,}02X_1 - 8{,}18 \times 10^{-3}X_2 + 9{,}04 \times 10^{-3}X_3 + 0{,}06X_6 + 0{,}52X_7 + 2{,}53 \times 10^{-4}X_8 - 0{,}60X_9 + 0{,}25X_{10} + 0{,}03X_{11}).$$

From this model we can interpretate:

1. For every 1% increase of the percentage of households with a main source of lighting being a non-state electricity company (PLN) will decrease the number of people living in poverty by $\exp(-0.02) = 1$ times.
2. For every 1% increase of the percentage of households with access to adequate and reliable drinking water services will decrease the number of people living in poverty by $\exp(-8.18 \times 10^{-3}) = 1$ times.
3. For every 1% increase of the percentage of households with access to proper and sustainable sanitation services will increase the number of people living in poverty by $\exp(9.04 \times 10^{-3}) = 1$ times.
4. For every 1% increase of open unemployment rate will increase the number of people living in poverty by $\exp(0.06) = 1$ times.
5. Every one million increase in the average monthly net income of informal workers will increase the number of people living in poverty by $\exp(0.52) = 1$ times.
6. Every 1 increase in population density for every square kilometer will increase the number of people living in poverty by $\exp(2.53 \times 10^{-4}) = 1$ times.
7. For every one million spent monthly per capita expense on food and non-food essentials will decrease the number of people living in poverty by $\exp(-0.60) = 1$ times.
8. Every 1% increase in the percentage of people who have a health complaint and do not treat it because there is no money will increase the number of people living in poverty by $\exp(0.25) = 1$ times.
9. Every 1% increase in the proportion of the population aged 15 and above without a diploma will increase the number of people living in poverty by $\exp(0.03) = 1$ times.

After Poisson regression modelling, continue to overdispersion check. Overdispersion can be identified by calculating the ratio of the deviance value to the degrees of freedom. In the Poisson Regression model, the deviance value is 284.47 with 44 degrees of freedom, resulting in a ratio of 6.46. Since this value exceeds one, it indicates the presence of overdispersion. Therefore, Poisson Regression is not an appropriate modeling choice, and Negative Binomial Regression will be used to address this issue.

Modeling the number of people living in poverty in Kalimantan Island using Poisson Regression revealed overdispersion. Therefore, Negative Binomial Regression was used for further analysis. The estimated parameters of the Negative Binomial Regression model are shown in **Table 4** below:

Table 4. **Negative Binomial Regression Modelling**

Variable	Estimation	Standard Error	$\lvert Z_{count} \rvert$	*p-value*
(Intercept)	2.05	0.92	2.21	0.02
X_1	-0.02	0.10	-2.02	0.04
X_2	-8.30×10^{-3}	7.21×10^{-3}	-1.15	0.24
X_3	9.88×10^{-3}	9.52×10^{-3}	1.03	0.29
X_6	-0.06	0.64	-1.05	0.29
X_7	0.53	0.15	3.48	4×10^{-4}
X_8	2.56×10^{-4}	8.09×10^{-5}	3.16	1.57×10^{-3}
X_9	-0.70	0.42	-1.66	0.09
X_{10}	0.22	0.09	2.39	0.02
X_{11}	0.04	0.01	2.65	0.07
Deviance	57.48		DF	46

The next process is partial and simultaneous significance testing. Simultaneous testing with an alpha of 10% obtained a value of $\chi^2_{(11:0.1)} = 17.275$ which is smaller than the deviance value of the Negative Binomial Regression model which is 57.48. Then, it can be concluded that reject H_0 where there is at least one independent variable that affects the dependent variable. After that, continue the partial test. For the partial test with 10% alpha, the value of $Z_{\left(\frac{0.1}{2}\right)} = 1.64$ is obtained. This value is compared with each independent variable value $\lvert Z_{count} \rvert$. For variables that has $\lvert Z_{count} \rvert$ bigger than 1.64 are variables X_1,X_7, X_8, X_9, X_{10}, and X_{11}.The model formed as follows:

$$\mu_i = \exp\left(2{,}05 - 0{,}02X_1 + 0{,}53X_7 + 2{,}56 \times 10^{-4}X_8 - 0{,}70X_9 + 0{,}22X_{10} + 0{,}04X_{11}\right)$$

From this model we can conclude that:

1. For every 1% increase of the percentage of households with a main source of lighting being a non-state electricity company (PLN) will decrease the number of people living in poverty by $\exp(-0.02) = 1$ times.
2. Every one million increase in the average monthly net income of informal workers will increase the number of people living in poverty by $\exp(0.53) = 1$ times.
3. Every 1 increase in population density for every square kilometer will increase the number of people living in poverty by $\exp(2.56 \times 10^{-4}) = 1$ times.
4. For every one million spent monthly per capita expense on food and non-food essentials will decrease the number of people living in poverty by $\exp(-0.70) = 1$ times.
5. Every 1% increase in the percentage of people who have a health complaint and do not treat it because there is no money will increase the number of people living in poverty by $\exp(0.22) = 1$ times.
6. Every 1% increase in the proportion of the population aged 15 and above without a diploma will increase the number of people living in poverty by $\exp(0.04) = 1$ times.

The deviance value in Negative Binomial Regression is 57.48 with a degree of freedom of 46 so the deviance ratio value is 1.24. This value is smaller than the Poisson Regression model. Therefore, it can be said that Negative Binomial Regression is proven to be able to overcome the problem of overdispersion in Poisson Regression.

Next, the heterogeneity testing stage is carried out. Spatial heterogeneity is a condition where there are different characteristics between one observation point and another, which can be seen in the Breusch-Pagan test. Based on the results of spatial heterogeneity testing, a p-value of 0.03 is obtained which is smaller than the value of α = 0.1, which means that there are differences in characteristics or spatial heterogeneity between one observation location and another.

In testing the significance of GWNBR parameters, there are two types of significance testing, namely simultaneously and partially. Testing simultaneously using 10% alpha, the deviance value of the GWNBR model is bigger than $\chi^2_{(11:0.1)} = 17.275$, so it can be concluded that reject H_0, which means there is at least one GWNBR model parameter that has a significant effect on the model. Therefore, it is necessary to continue with partial testing, which uses a significance level of 10% so that $Z_{\left(\frac{0.1}{2}\right)} = 1.64$ is obtained. If the value of $\lvert Z_{count} \rvert > 1.64$ then rejecting H_0 indicates that the variable is significant to the model. The analysis results identified six groups of districts/cities based on significant variables. These groups, categorized according to significant variables, are presented in **Table 5** below:

Table 5. Groups by Significant Categories

Province	District/City	Significant Categories
West Kalimantan	Melawi	Human
	Sanggau, Kuburaya	Housing/Environment and Economy
	Sintang, Landak, Pontianak, Kapuas Hulu, Kayong Utara, Singkawang, Mempawah, Bengkayang, Ketapang, Sekadau, Sambas	Housing/Environment, Human, and Economy
East Kalimantan	Penajam Paser Utara, Balikpapan	Housing/Environment and Economy
	Paser	Human and Economy
	Kutai Barat, Kutai Timur, Kutai Kertanegara, Mahakam Ulu, Berau, Bontang, Samarinda	Housing/Environment, Human, and Economy
Central Kalimantan	Gunung Mas, Lamandau	Housing/Environment
	Seruyan, Barito Timur	Economy
	Sukamara, Murung Raya, Katingan, Kapuas, Palangkaraya	Housing/Environment and Economy
	Kotawaringin Timur, Kotawaringin Barat	Human and Economy
	Barito Selatan, Pulang Pisau, Barito Utara	Housing/Environment, Human, and Economy
North Kalimantan	Tarakan, Nunukan, Malinau	Housing/Environment and Economy
	Bulungan	Human and Economy
	Tana Tidung	Housing/Environment, Human, and Economy
South Kalimantan	Hulu Sungai Selatan, Banjarmasin, Tanah Laut, Barito Kuala, Kotabaru	Housing/Environment and Economy
	Tabalong	Human and Economy
	Hulu Sungai utara, Balangan, Hulu Sungai Tengah, Banjarbaru, Banjar, Tanah Bumbu, Tapin	Housing/Environment, Human, and Economy

Based on **Table 5**, it can be seen that there are six groups from significant categories. Those groups are Human, Economy, Housing/Environment, Housing/Environment and Economy, Human and Economy, Housing/Environment and Human and Economy.

In **Figure 1**, groups of districts/cities categorized according significant variables is shown as follows:

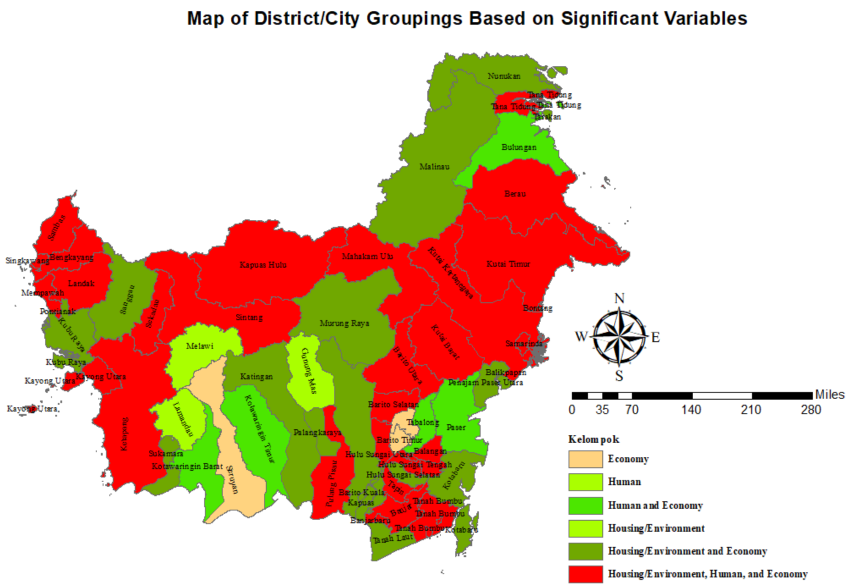

Figure 1 Map of Groups by Significant Variables

For example, partial testing at the research location in the research data, Pontianak City, can be seen in **Table 6** below:

Table 6 Parameter Estimation of GWNBR Model of Pontianak City

Parameter	Coefficient	$\lvert Z_{count} \rvert$
(Intercept)	9.11	13.91
X_1	-0.03	-2.26
X_7	0.49	5.13
X_8	1.82×10^{-4}	1.90
X_9	-0.79	-0.67
X_{10}	0.17	-5.97
X_{11}	0.05	67.53

The model parameters are significant if the $|Z_{count}|$ value is bigger than 1.64. Based on the $|Z_{count}|$ value above, the variables that are significant to the number of people living in poverty in Kalimantan Island are X_1, X_7, X_8, X_{10}, and X_{11}. The model formed is as follows:

$$\mu_1 = exp(9.11 - 0.03X_1 + 0.49X_7 + 1{,}82 \times 10^{-4}X_8 + 0.17X_{10} + 0.05X_{11})$$

Based on the model, it can be interpreted that:

1. Every 1% increase in the percentage of households with a main source of lighting being a non-state electricity company (PLN) will decrease the number of people living in poverty by $\exp(-0.03) = 1$. From Indonesia Official Statistic (BPS), This is because majority of the percentage of non-state electricity company comes from solar power, which is more efficient and economical than PLN electricity, thus reducing poverty.
2. Every one million increase in the average monthly net income of informal workers will increase the number of people living in poverty by $\exp(0.49) = 1$ times. This happens because based on the Official Statistical News at BPS Indonesia, in March 2023 there was an increase in the Gini ratio in urban and rural areas.
3. Every 1 increase in population density for every square kilometer will increase the number of people living in poverty by $\exp(1.82 \times 10^{-4}) = 1$times. This shows that the more densely populated an area is, the more poverty it will cause.
4. Every 1% increase in the percentage of people who have a health complaint and do not treat it because there is no money will increase the number of people living in poverty by $\exp(0.17) = 1$ times. This shows that every person who does not seek medical treatment because there is no money will increase the number of people living in poverty.
5. Every 1% increase in the proportion of the population aged 15 and above without a diploma will increase the number of people living in poverty by $\exp(0.05) = 1$ times. This shows that every increase in the number of people who do not have a diploma will increase the number of people living in poverty.

4. Conclusions

Based on the results of the analysis and discussion in this study, it is concluded that the number of people living in poverty in Kalimantan Island happen at all city/district. The model of Poisson regression has overdispersion so it is overcome by using Negative Binomial regression. Based on modeling using Negative Binomial regression, there are seven significant variables. These variables are the percentage of households with the main source of lighting being a non-state electricity company (PLN) that negative significant. Based on BPS, majority in Kalimantan Island with non-state electricity are from solar panel. Then the average monthly net income of informal workers that are positive significant. Based on BPS, rate of gini ratio in 2023 increase, cause income inequality. Then population density for every square kilometer. Then monthly per capita expense on food and non-food essentials. Increasing purchasing power enables people to meet basic needs and improve their quality of life, thereby reducing the number of poor individuals. Then the percentage of people who have a health complaint and do not treat it because there is no money and the percentage of the population aged 15 and above without a diploma. From the GWNBR model, with the Adaptive gaussian kernel weighting function, the results of clustering regions based on poverty show that there are six district/city groups formed based on significant variables.

References

[1] Atno Parluhutan Sinaga, A. (2022). POVERTY PERSPECTIVES AND REDUCTION STRATEGIES IN INDONESIA. *Three Seas Economic Journal*, *3*(3), 1–9. https://doi.org/10.30525/2661-5150/2022-3-1

[2] Hermawan, I., Mulya Firdausy, C., Rizqy Rambe, K., Zuhdi, F., Erwidodo, Dewi Nugraheni, R., Malisan, J., Isnasari, Y., Marpaung, E., & Milawati Asshagab, S. (2024). Road traffic facilities, traffic accidents, and poverty: Lesson learned from Indonesia. *Transportation Research Interdisciplinary Perspectives*, *28*(July), 101273.

https://doi.org/10.1016/j.trip.2024.101273
[3] Jasmadi, I., Widodo, D. A., & Oktaviana, P. P. (2016). Pemodelan dan Pemetaan Kasus Jumlah Penduduk Miskin di Provinsi Jambi pada Tahun 2014 dengan Menggunakan Geographically Weighted Negative Binomial Regression. *Jurnal Sains dan Seni ITS*, *5*(2), 350–354. https://doi.org/http://dx.doi.org/10.12962/j23373520.v5i2.16704
[4] Wiratama, B. F., Kurniawan, R., Mulyanto, Isnaeni, M. A., Sumargo, B., & Gio, P. U. (2023). Measuring the physical infrastructure development as poverty reduction program in Kalimantan, Indonesia. *Cities*, *141*, 104515. https://doi.org/10.1016/j.cities.2023.104515
[5] Halida, A., Salam, N., & Lestia, A. S. (2021). Modeling of poverty in Kalimantan with nonparametric spline regression approach. *Journal of Physics: Conference Series*, *2106*(1), 012008. https://doi.org/10.1088/1742-6596/2106/1/012008
[6] Winata, H. M. (2023). Mengatasi Overdispersi Dengan Regresi Binomial Negatif Pada Angka Kematian Ibu Di Kota Bandung. *Jurnal Gaussian*, *11*(4), 616–622. https://doi.org/10.14710/j.gauss.11.4.616-622
[7] Delvia, N., Mustafid, M., & Yasin, H. (2021). Geographically Weighted Negative Binomial Regression Untuk Menangani Overdispersi Pada Jumlah Penduduk Miskin. *Jurnal Gaussian*, *10*(4), 532–543. https://doi.org/10.14710/j.gauss.v10i4.33106
[8] Fitriani, R., & Gede Nyoman Mindra Jaya, I. (2020). Spatial modeling of confirmed COVID-19 pandemic in East Java Province by geographically weighted negative binomial regression. *Communications in Mathematical Biology and Neuroscience*, *2020*, 1–17. https://doi.org/10.28919/cmbn/4874
[9] Zubedi, F., Aliu, M. A., Rahim, Y., & Oroh, F. A. (2021). Analisis Faktor-Faktor Yang Mempengaruhi Stunting Pada Balita Di Kota Gorontalo Menggunakan Regresi Binomial Negatif. *Jambura Journal of Probability and Statistics*, *2*(1), 48–55. https://doi.org/10.34312/jjps.v2i1.10284
[10] Octavianus Halim, A., & Imro'ah, N. (2024). Negative Binomial Regression in Overcoming Overdispersion Poverty Data in Kalimantan. *Jurnal Forum Analisis Statistik (FORMASI)*, *4*(1), 1–10. https://doi.org/10.57059/formasi.v4i1.67
[11] Selvia, E., Imro'ah, N., & Andani, W. (2023). Pemodelan Jumlah Siswa Putus Sekolah Menggunakan Metode Geographically Weighted Negative Binomial Regression. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, *12*(4), 335–344.
[12] Priambodo, B. W. Y., & Irhamah, I. (2019). Pemetaan Jumlah Property Crime di Provinsi Jawa Timur Menggunakan Metode Geographically Weighted Negative Binomial Regression (GWNBR) dan Geographically Weighted Poisson Regression (GWPR). *Inferensi*, *2*(2), 53. https://doi.org/10.12962/j27213862.v2i2.6818
[13] Fitrial, N. H., & Fatikhurrizqi, A. (2020). *Pemodelan Jumlah Kasus COVID-19 di Indonesia dengan Pendekatan Regresi Poisson dan Regresi Binomial Negatif*. *1*. https://doi.org/https://doi.org/10.34123/semnasoffstat.v2020i1.465
[14] Suryadi, F., Jonathan, S., Jonatan, K., & Ohyver, M. (2023). Handling overdispersion in poisson regression using negative binomial regression for poverty case in west java. *Procedia Computer Science*, *216*, 517–523. https://doi.org/10.1016/j.procs.2022.12.164
[15] Santi, V. M., & Rahayuningsih, Y. (2023). Negative Binomial Regression in Overcoming Overdispersion in Extreme Poverty Data in Indonesia. *Pattimura International Journal of Mathematics (PIJMath)*, *2*(2), 43–52. https://doi.org/10.30598/pijmathvol2iss2pp43-52
[16] Fadilah, F. W. R., Handajani, S. S., Zukhronah, E., & Pratiwi, H. (2019). Geographically weighted negative binomial regression model to analysis of factors that influence on maternal mortality in Central Java Province. *AIP Conference Proceedings*, *2202*, 1–6. https://doi.org/10.1063/1.5141718
[17] Wahyuni, L. (2021). *Analisis Faktor-Faktor Risiko yang Mempengaruhi Jumlah Kasus Tuberkulosis (TBC) di Provinsi Jawa Barat Tahun 2018 Menggunakan Pendekatan Geographically Weighted Negative Binomial Resgression (GWNBR)*. https://repository.uinjkt.ac.id/dspace/handle/123456789/56263%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/56263/1/LELY WAHYUNI-FST.pdf
[18] Fauwziyah, F., Astutik, S., & Pramoedyo, H. (2022). Geographically Weighted Negative Binomial Regression Modeling using Adaptive Kernel on the Number of Maternal Deaths during Childbirth. *Mathematics and Statistics*, *10*(5), 1133–1139. https://doi.org/10.13189/ms.2022.100525
[19] Zhang, L., Cheng, J., & Jin, C. (2019). Spatial interaction modeling of OD flow data: Comparing geographically weighted negative binomial regression (GWNBR) and OLS (GWOLSR). *ISPRS International Journal of Geo-Information*, *8*(5). https://doi.org/10.3390/ijgi8050220
[20] Suryani, I., Yasin, H., & Kartikasari, P. (2021). Pemodelan Jumlah Kasus Demam Berdarah Dengue (Dbd) Di Jawa Tengah Dengan Geographically Weighted Negative Binomial Regression (Gwnbr). *Jurnal Gaussian*, *10*(1), 136–148. https://doi.org/10.14710/j.gauss.v10i1.29400