Unveiling Factors and Refining Predictions: A Geographically Weighted Regression Analysis of Poverty in Indonesia

Angelica

*Statistics Department*

*School of Computer Science*

*Bina Nusantara University*Jakarta, Indonesia 11530  
angelica008@binus.ac.id

Elvano Jethro Mogi Pardede

*Statistics Department*  
*School of Computer Science*

*Bina Nusantara University*Jakarta, Indonesia 11530  
elvano.pardede@binus.ac.id

Amsal Anugrah

*Statistics Department*

*School of Computer Science*

*Bina Nusantara University*

Jakarta, Indonesia 11530  
amsal.anugrah@binus.ac.id

***Abstract*— Poverty is still an important issue in Indonesia, so a comprehensive analysis is needed to identify influencing factors and improve prediction models. This study uses Geographically Weighted Regression (GWR) to assess poverty rates in 34 provinces in Indonesia, considering indicators such as education completion rate, sanitation, drinking water, health facilities, home ownership, and electricity supply with Ordinary Least Square (OLS) as the reference. This study aims to overcome the spatial variation problem of the OLS method by using GWR, which goes beyond the limitations of OLS and Weighted Least Square (WLS) regression models. Early findings show that education completion rate, adequate sanitation, and drinking water significantly affect the poverty rate. Comparison with Ordinary Least Square (OLS) and Weighted Least Square (WLS), shows the superiority of GWR in terms of explanatory power (R-squared), at 0.8195 and model fit (AIC), at 530.1814. The GWR model, using a fixed Gaussian Kernel Function Bandwidth, outperformed the other models, providing valuable insights for policy makers to formulate targeted strategies in poverty reduction efforts in Indonesia.**

***Keywords—* regression, GWR, poverty level, WLS**

# Introduction

Poverty is still one of the most concerning issues in Indonesia. Although there have been many ways to decrease the poverty rate in Indonesia, especially in some isolated areas, there are still economic inequalities that are worrisome. Poverty is not only measured by wealth but also as a challenge to provide a decent life to the community, the Government of Indonesia must take responsibility as a form of effort to overcome economic inequality in Indonesia. Based on data from [1], Indonesia holds the 91st position out of 207 countries in the global ranking of the poorest nations, as determined by the Gross National Income (GNI) per capita for each country. In this case, it means that Indonesia still has to intensify the reduction of the poverty rate.

Poverty levels can be measured based on several indicators, one of which is proper sanitation [2]. The higher the percentage of households that have access to proper sanitation, the lower the poverty rate in the community. With access to proper sanitation, it can improve the health of the local community so that people can move productively. With increased productivity, people can move out of the poverty zone [2]. According to the Central Bureau of Statistics in 2023 [3], the poverty rate in Indonesia in March 2023 reached up to 9.36%. This percentage has experienced a modest decrease of 0.21% from September 2022.

The government must take various actions to handle the poverty problem that occurs in Indonesia. Therefore, the factors that influence the poverty rate must be discussed first so that the government can strategize the real actions that must be carried out. It is hoped that the government can earnestly implement effective measures to substantially impact the poverty rate in Indonesia. Based on research conducted by Adhitya, et al [2] using multiple linear regression with a panel data approach to analyze the poverty rate in Indonesia, 3 variables affect the poverty rate. As a result, this study will employ statistical methods to examine the factors influencing the poverty rate in the country using MLR and GWR.

One frequently employed statistical technique for analysis is linear regression. This model examines the linear association among the response variable and the independent variable, aiming to predict the dependent value when the independent variable is not constant [4]. Linear regression analysis can produce parameter estimation values with the Ordinary Least Square (OLS) method which has 3 residual assumptions that must be met, namely independent, identical, and normally distributed. However, differences in environmental conditions in a location or region can violate the assumptions of homoscedasticity (identical) and non-autocorrelation, it can occur due to spatial effects [5].

Geographically Weighted Regression (GWR) is one of the statistical method solutions that can overcome spatial problems. Geographically Weighted Regression (GWR) is an extended method of multiple linear regression that obtains parameter estimates concerning the point of location (local). The GWR method is often found in various fields. There is research that discusses the use of the GWR method and the linear regression method [5]. The results indicated that the GWR method yielded a more effective model compared to the global regression model. In addition, other studies discuss the use of the GWR method in the spread of tuberculosis disease [6]. The results show that the GWR method produces higher R-squared values and lower AIC values compared to the multiple linear regression model.

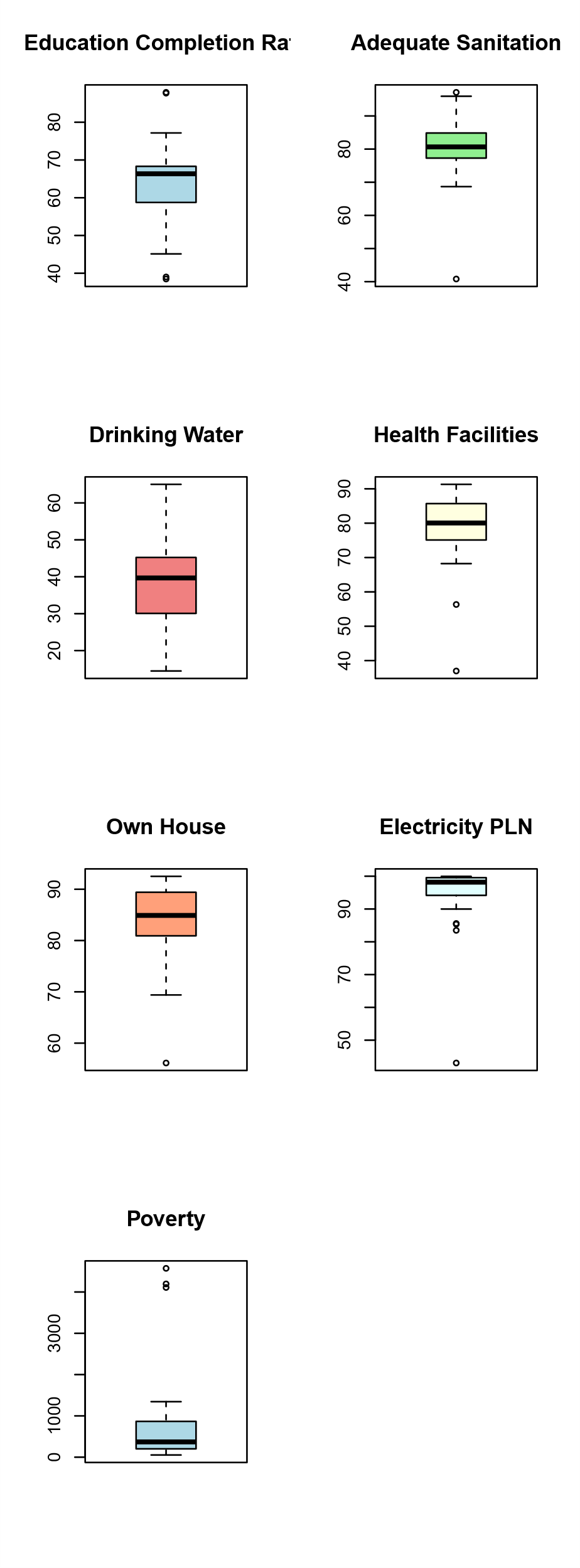
The purpose of this research is twofold: firstly, to identify the factors influencing the poverty levels in Indonesia, and secondly, to construct a model using the Geographically Weighted Regression (GWR) method to analyze the extent of poverty in the country. The study aims to delve into the multifaceted aspects contributing to poverty and, through the application of GWR, to develop a nuanced model that considers spatial variations in the Indonesian context. By accomplishing these goals, this research attempts to offer a valuable perspective that can help in reformulating successful policies and strategies to alleviate poverty in Indonesia. The investigation is specifically focused on understanding the diverse factors impacting poverty and utilizing advanced statistical methods to enhance the precision of poverty modeling, ultimately aiming to provide recommendations that will aid in government action to alleviate poverty and improve the well-being of the Indonesian population.

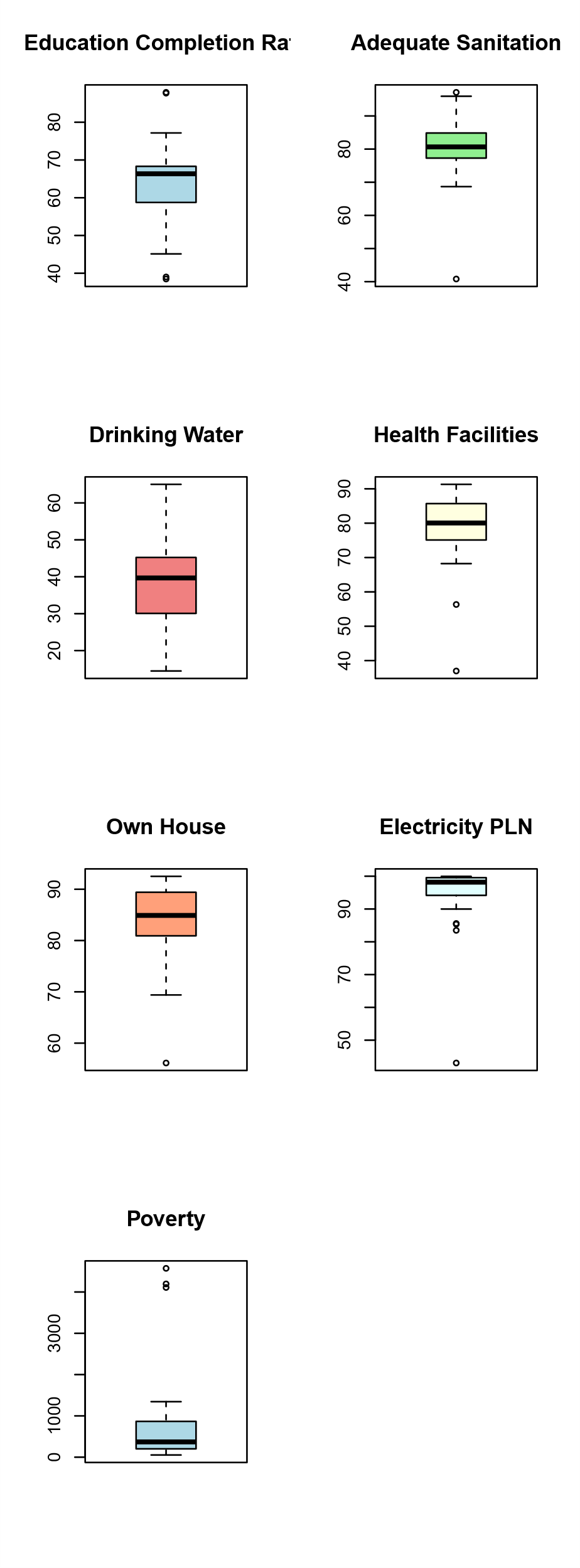
# Methodology

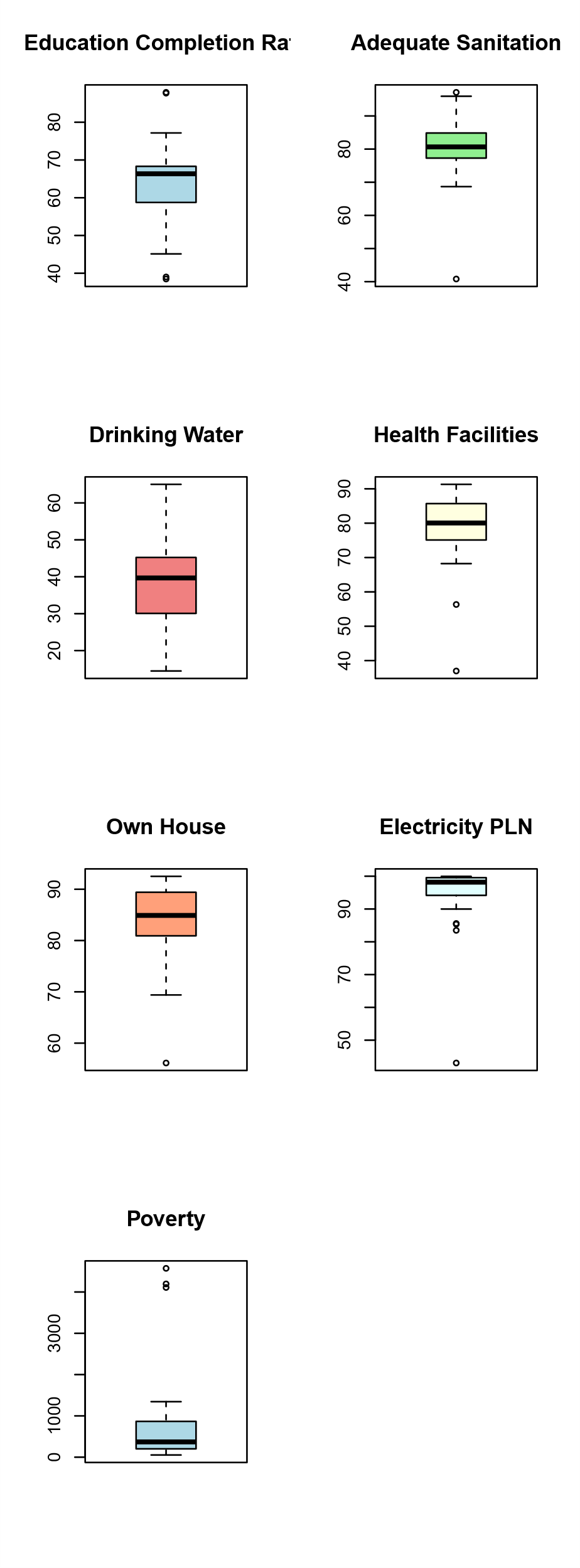
1. *Dataset*

This study utilized secondary data pertaining to the poverty status in Indonesia in the year 2022, sourced from BPS and accessible through [7]. The study encompasses observations across 34 provinces in Indonesia. The response variable is poverty and the predictor variables are education completion rate, adequate sanitation, drinking water, health facilities, own house, and electricity supply. "Educational Completion Rate" refers to the level of education completed by each household, usually up to senior high school (SMA) level or its equivalent. It reflects the percentage of the population in the household that has completed secondary education. "Adequate Sanitation" relates to the condition where each household has access to adequate sanitation facilities. This includes efforts to ensure that sanitation conditions in the household meet established standards to support health and hygiene. "Drinking Water" describes the condition where every household has access to a water source that meets the eligibility criteria for drinking water. It indicates efforts to ensure the availability of safe and potable drinking water for households. "Health Facilities" covers the condition where each household has access to adequate health facilities. This indicates the existence of health facilities that household residents can access to fulfill their health needs. "Own House" indicates that the household has its place of residence. This reflects the level of stability of the household's residence. "Electricity PLN" indicates that each household can access electricity from the State Electricity Company (PLN). This reflects the availability of electrical power for daily household needs.

Based on [2], several variables that affect the poverty rate are education, health facilities, proper sanitation, and the average number of household members. Therefore, this study uses the variables of education completion rate, health facilities, adequate sanitation, and drinking water. As a result of the government action to alleviate poverty through the housing renovation program [8], this dataset includes the own house variable as one of the variables that may affect poverty. Based on the value of the IKM indicator, poor households have the highest problem with electrification from State Electricity Company (PLN)/non-PLN electricity, so the Electricity PLN variable is included in the dataset as one of the variables that may affect poverty levels.







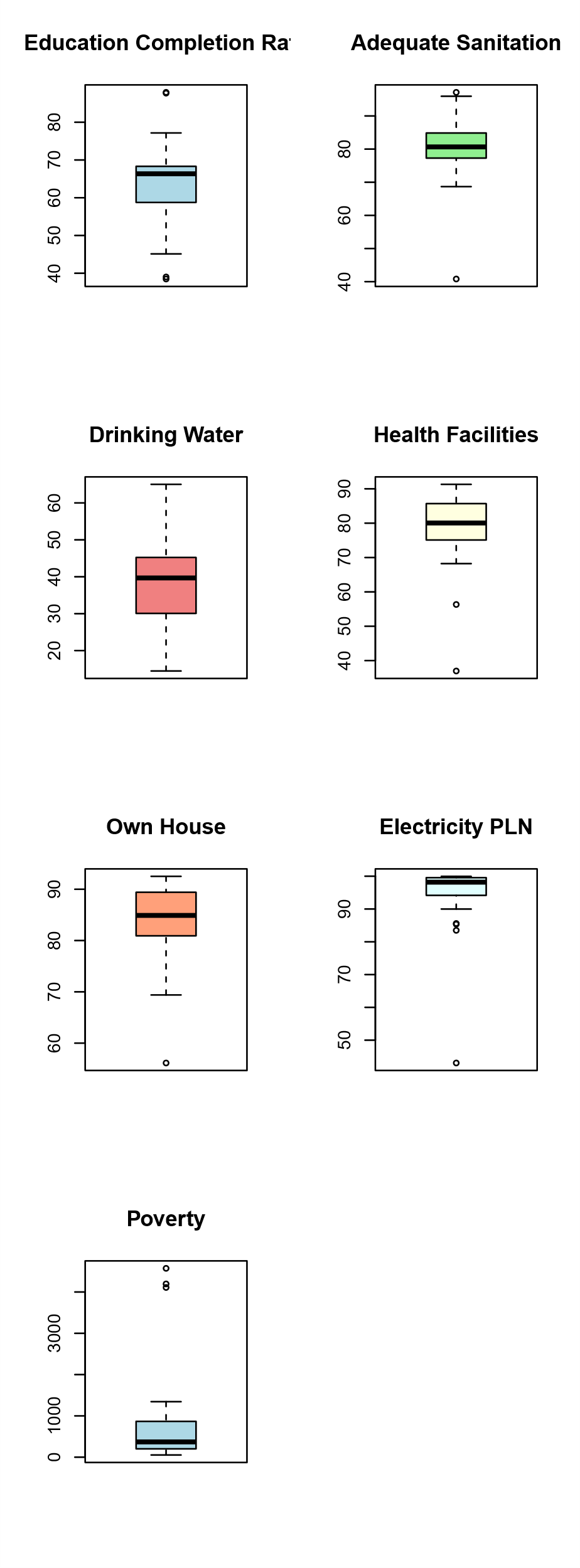


Figure 1. Boxplot of Every Variables in the Dataset

Numerical variables will be described using boxplots to identify outliers in each variable. From figure 1, it can be concluded that in each variable has outlier except variable drinking water such as in education completion rate, there are outliers above 80 and below 45.

1. *Multiple Regression Analysis*

Multiple Linear Regression (MLR) involves techniques in statistics that go beyond simple linear regression by capturing the relations between the response variable and multiple predictor variables. In MLR, the predictors and response variables are assumed to have a linear relationship. The model is expressed through a linear equation, where the coefficients associated with each predictor are determined by minimizing the total squared differential between the observed and predicted values. The intercept term () indicates the dependent variable's baseline value, while the coefficients () measure the alteration in the mean of the dependent variable when there is a unit change in the respective predictor while keeping the other predictors constant. The formula for the multiple regression model is as follow [9]:

(1)

1. *Ordinary Least Square*

Ordinary Least Squares (OLS) is a commonly applied statistical approach to obtain parameter values in linear regression analysis. The primary goal of OLS is to identify the coefficient values that closely approximate the residuals sum of squares among the predicted and observed values for the dependent variable. In the context of a simple linear regression model. OLS's purpose is to determine the line of best fit by the minimization of the sum of squared differences among the predicted and actual values. This process extends to a multidimensional space in the case of multivariate linear regression models with multiple predictors, seeking coefficients that minimize the residual sum of squares. The coefficient estimates obtained from OLS in the linear model are then used to make predictions and understand the power and the direction of the relationship among the dependent and independent variables. Despite its widespread use, OLS requires specific requirements for valid and unbiased parameter estimations, such as linearity, error independence, homoscedasticity, and residual normality [10]. The linear regression equation in its general form is expressed as follows.

(2)

Although it uses the same equation as MLR, the coefficient is established by diminishing the total of squared differences among the actual value and the predicted value. The equation as follows.

(3)

An important assumption in the Ordinary Least Squares (OLS) regression technique is that all variables considered are continuous, underscoring the importance of theregression coefficients true values [11]. This suggests that the anticipated value of the dependent variable should match the results obtained from the specified element [11]. The calculation of regression coefficients of OLS estimator can be achieved using the subsequent formula.

(4)

Where X consists of the independent variables, denotes the transpose operation on matrix X, and Y represents the dependent variable matrix..

1. *Weighted Least Square Regression*

Weighted Least Squares (WLS) Regression is an improvement on the Ordinary Least Squares (OLS) approach that addresses heteroscedasticity, which occurs when error variability is not consistent across all data. Unlike OLS, WLS takes into account the possibility that certain observations have higher or lower degrees of uncertainty. This is accomplished by giving each observation a different weight based on the estimated variance of the errors associated with that observation. During the estimation stage, minimization of the weighted sum of squared residuals occurs, with weights usually obtained from the error matrix estimated from the covariance and variance matrices. This method yields more efficient parameter estimations, especially when working with datasets with varied amounts of error variation. When OLS assumptions concerning homoscedasticity are broken, WLS provides a more accurate and reliable regression analysis by taking into account the variable degrees of uncertainty associated with distinct observations in the dataset. [12]. The linear regression equation in its general form is expressed as follows:

(5)

Although it uses the same equation as MLR, the coefficient is established by diminishing the total of squared differences among the actual value and the predicted value. Hence assigning wi into the equation.

(6)

1. *Residual Assumptions*

The errors in (MLR Equation) are assumed to be in 3 conditions: normality assumption, homoscedasticity assumption, and independent-error assumption [13].

1. Normality Assumption

The normality test is conducted to assess whether the errors follow a normal distribution (the assumption) or not [14]. The test uses the Kolmogorov-Smirnov normality test in R. The hypothesis for this test is as follows.

The residuals are distributed normally.

The residuals are not distributed normally.

1. Homoscedasticity Assumption

A spatial heterogeneity test was conducted to ascertain whether distinct characteristics or unique features were present at each observation location (homoscedasticity assumption) [14]. The test uses the Breusch Pagan Test in R. The hypothesis for this test is as follows.

Errors are distributed with equal variance (homoscedastic)

Errors are not distributed with equal variance (heteroscedastic)

1. Independent-Error Assumption

The goal of the autocorrelation test is to establish whether there is a correlation among adjacent members or observational data (independent-error assumption). The test uses the Durbin-Watson Test in R. The hypothesis for this test is as follows.

The residuals are non-autocorrelated

The residuals are autocorrelated

1. *Multicolinearity Test*

The examination of the assumption of multicollinearity is performed before conducting analysis to confirm the existence of a linear correlation among independent variables in the models. [15]. If the VIF value is below ten (VIF < 10), this implies that the independent variables are indeed free from multicollinearity.

1. *Geographically Weighted Regression*

Spatial heterogeneity is a fundamental assumption embedded in spatial data. The Geographically Weighted Regression (GWR) model is well-suited for modeling datasets characterized by spatial heterogeneity, offering advantages over multiple regression methods [16]. Geographically Weighted Regression (GWR) represents a spatial analysis technique derived from multiple linear regression designed to account for spatial heterogeneity effects within the data. The GWR (Geographically Weighted Regression) model utilizes various weighting functions, including fixed and adaptive kernels such as Bi-Square, Exponential, Gaussian, and Tricube [14]. The general form of GWR is as follows [17]

(7)

Where yi stands for the response variable at the i-th position (i = 1,2,3,..., n), x\_ik denotes the k-th independent variable at the i location. The regression coefficient is represented by β, and it is assumed that the residuals, denoted as ε, are identical, independent, and follow a normal distribution with consistent variances [17]. Since the assumption of homoscedasticity is not satisfied, GWR modeling is performed with several steps as follows:

1. Create a multiple regression analysis model
2. Test residual assumptions
3. Create a geographically weighted regression model
4. Declare the longitude and latitude of each province
5. Choose the most optimal bandwidth value
6. Perform model evaluation

The fundamental requirement in implementing Geographically Weighted Regression (GWR) lies in determining the optimal bandwidth and kernels. The process involves calculating the most suitable bandwidth for each independent variable, considering factors such as the selection between fixed or adaptive kernels and the method by which neighboring points are weighted [18]. In performing the model, it is necessary to estimate the parameters. The determination of parameters in the GWR model is described as follows. [19].

(8)

1. *Metrics*

Evaluation metrics are designed to identify the number of accurately detected objects. Various evaluation metrics exist in regression analysis, such as AIC, AIC corrected, BIC, R-Squared, SSE, and adjusted R-Squared. The effectiveness of the model, when compared to alternative models, is considered better if it shows a higher R-squared value and a lower AIC value in modeling the data [20].

1. Akaike Information Criterion (AIC)

AIC is used to evaluate and compare the fit between different models. The formula of the AIC is as follows [21].

(9)

1. R-squared

R-squared serves as a measure to assess the adequacy of the overall fit for a particular model. The formula of the R-squared is as follows [22].

(10)

1. Adjusted R-squared

Adjusted R-squared closely resembles Multiple R-squared values, with the difference that this metric includes a correction for the total amount of variables [23]. This metric is used to effectively compare multiple models. The formula of the adjusted R-squared is as follows [22].

(11)

1. Bayesian Information Criterion (BIC)

BIC is close to AIC, but BIC customizes the penalty factor for complexity according to the sample size and the BIC formula is as follows [22].

(12)

1. AIC Corrected (AICc)

AICc is advised in the case where the dataset is limited in size or when the number of predictors is very large compared to the available sample size and the formula of the AIC corrected is as follows [22].

(13)

1. Residuals Sum of Squares (SSE)

The residual sum of squares (RSS) gauges the extent of variability in the residual errors within a regression model. A reduced Residual Sum of Squares (RSS) suggests a more optimal fit of the model to the dataset. The equation of the SSE is shown below [13].

(14)

# Results and Discussions.

In this research, the regression used to estimate is the simplest regression, which is linear regression. Given the presence of multiple independent variables, the estimation of the poverty rate in Indonesia will be carried out using a multiple linear regression equation. The variables used to estimate are all variables. The equation representing the regression model for the poverty rate in Indonesia is as follows.

(15)

The primary objective of Multiple Linear Regression modeling is to identify the factors influencing the poverty rate in Indonesia, excluding spatial considerations. This approach differs from the GWR model, as it operates on a global scale. The initial phase of the Multiple Linear Regression Analysis involves conducting tests for multicollinearity and assessing classical assumptions, which include independence, normal distribution, and the absence of identical patterns among variables [14].

1. *Normality Test*

The test uses the Kolmogorov-Smirnov normality test in R. The hypothesis for this test is as follows.

The residuals are distributed normally.

The residuals are not distributed normally.

The result indicates the p-value for the test is is 0.9792, since the significance level is 0.05 it fails to reject Thus, the errors are distributed normally.

1. *Homoscedasticity Test*

The test uses the Breusch Pagan Test in R. The hypothesis for this test is as follows.

Errors are distributed with equal variance (homoscedastic)

Errors are not distributed with equal variance (heteroscedastic)

The result indicates the p-value for the test is 0.01527 since the significance level is 0.05 so it is rejected . Therefore, the residuals do not show a homogeneous distribution of variance.

1. *Autocorrelation Test*

The test uses the Durbin-Watson Test in R. The hypothesis for this test is as follows.

The residuals are non-autocorrelated

The residuals are autocorrelated

The result indicates the p-value is 0.0638, since the significance level is 0.05 it fails to reject Thus, the residuals are non-autocorrelated.

1. *Multicollinearity Test*

The existences of multicollinearity is assessed through the VIF value and is considered non-existent if the VIF value for the independent variable is below 10.

Table 1. Multicollinearity Test Result

|  |  |
| --- | --- |
| **Variable Name** | **VIF** |
| Educational Completion Rate | 1.505833 |
| Adequate Sanitation | 2.407062 |
| Drinking Water | 1.858289 |
| Health Facilities | 2.629495 |
| Own house | 1.516493 |
| Electricity | 1.110844 |

From table 1, it shows that in the test there are no variables that have multicollinearity because the VIF value < 10.

Parameter testing is conducted to identify which independent variables have an impact on the dependent variable [5]. Consequently, both backward and forward elimination procedures are implemented to ascertain the significant variables. The results showed that the significant variables were Educational Completion Rate, Adequate Sanitation, and Drinking Water, so a new model was developed with these three variables. The model is shown in Eq. 10

(16)

1. *Weighted Least Square Regression Model*

The assumption of homoscedasticity is not satisfied so a WLS model is used to overcome the problem. The WLS model is as follows.

(17)

1. *Geographically Weighted Regression Model*

In this study, a fixed kernel weighting matrix, specifically a Gaussian kernel is used. The optimal bandwidth value is determined through the application of the Gaussian kernel. The result is shown in table 2.

Table 2. Optimum Bandwidth using Gaussian Kernel

|  |  |
| --- | --- |
| **Bandwidth** | **CV** |
| 1840.836 | 29867244 |
| 2975.562 | 34110724 |
| 1139.537 | 26115205 |
| 706.1106 | 27503451 |
| 1135.29 | 26096314 |
| 1012.664 | 25682611 |
| 895.5707 | 25709698 |
| 961.8084 | 25624520 |
| 959.6948 | 25624156 |
| 956.0382 | 25623950 |
| 956.5197 | 25623950 |
| 956.5329 | 25623950 |
| 956.5314 | 25623950 |
| 956.5315 | 25623950 |
| 956.5314 | 25623950 |
| 956.5314 | 25623950 |

From table 6, the optimum bandwidth is 956.5314 corresponding to the lowest cross-validation (CV) value of 25623950. The results of Geographically Weighted Regression (GWR) are substantially influenced by the choice of bandwidth [15]. Once the optimal bandwidth is determined, a model is constructed using this bandwidth value. The model parameter estimation is described as follows.

Table 3. Comparison of OLS and GWR Model

|  |  |  |  |
| --- | --- | --- | --- |
|  | OLS Model | GWR Model | |
|  | -841.76 | Min | 3224.9155 |
| Max | 3550.4575 |
|  | 36.83 | Min | -16.7440 |
| Max | -8.3409 |
|  | 46.47 | Min | 2.0299 |
| Max | 7.7157 |
|  | -72.88 | Min | -9.2864 |
| Max | -4.2650 |
|  | -13.36 | Min | -25.3306 |
| Max | -24.5775 |
|  | -20.98 | Min | 612.2092 |
| Max | 2489.9970 |
|  | 11.55 | Min | -12.1333 |
| Max | -9.3417 |

1. *Metrics Evaluation*

In this research, we use AIC and R-Squared for the metrics evaluation. The result is as follows.

Table 4. Regression Models Comparison

|  |  |  |  |
| --- | --- | --- | --- |
| Model | R-squared | Adjusted R-squared | AIC |
| OLS\* | 0.6034 | 0.5152 | 559.5954 |
| OLS\*\* | 0.5729 | 0.5302 | 556.1101 |
| WLS\* | 0.4312 | 0.3049 | 533.8177 |
| GWR\* | 0.8195 | - | 530.1814 |
| GWR\*\* | 0.7403 | - | 537.701 |

|  |  |  |  |
| --- | --- | --- | --- |
| Model | BIC | AICc | SSE |
| OLS\* | 571.8062 | 565.3554 | 17480249 |
| OLS\*\* | 563.7419 | 558.253 | 18822173 |
| WLS\* | 546.0286 | 539.5777 | 28.2851 |
| GWR\* | - | 569.1375 | 7956444 |
| GWR\*\* | - | 556.7167 | 11443877 |

\* Independent Variables: education\_completion\_rate + adequate\_sanitation + drinking\_water + healty\_facilities + own\_house + electricity\_supply (6 variables)

\*\* Independent Variables: education\_completion\_rate + adequate\_sanitation + drinking\_water (3 variables)

# Conclusion

The most effective model to examine the poverty rate in Indonesia is the GWR model with a fixed bandwidth Gaussian Kernel Function. This model reached an R-squared value of 81.95%, which indicates its ability to describe 81.95% of the variability in the data. The GWR model has the lowest AIC of 530.1814 compared to other models. Based on the overall metrics, the GWR model with 6 variables is the best model even though it does not have the smallest BIC, AICc, and SSE values. The significant variables that affected the poverty level in Indonesia are Educational Completion Rate, Adequate Sanitation, and Drinking Water.

Looking ahead, there are promising opportunities for future research and refinement of poverty analysis in Indonesia. Future research could explore additional variables over time that could improve the accuracy of model predictions. In addition, investigating the structural dynamics of poverty can provide valuable insights into evolving socio-economic challenges. Furthermore, the integration of advanced machine learning techniques or ensemble models can contribute to a more comprehensive understanding of the various determinants of poverty.

# References

[1] “Poorest Asian Countries 2023,” World Population Review. Accessed: Dec. 13, 2023. [Online]. Available: https://worldpopulationreview.com/country-rankings/poorest-asian-countries

[2] B. Adhitya, A. Prabawa, and H. Kencana, “Analisis Pengaruh Pendidikan, Kesehatan, Sanitasi dan Rata-Rata Jumlah Anggota Keluarga Per Rumah Tangga terhadap Kemiskinan di Indonesia,” *Ekonomis: Journal of Economics and Business*, vol. 6, no. 1, p. 288, Mar. 2022, doi: 10.33087/ekonomis.v6i1.501.

[3] “Profil Kemiskinan di Indonesia Maret 2023,” Jul. 2023. Accessed: Dec. 13, 2023. [Online]. Available: https://www.bps.go.id/id/pressrelease/2023/07/17/2016/profil-kemiskinan-di-indonesia-maret-2023.html

[4] C. Siregar, A. S. Sembiring, and H. K. Siburian, “Perancangan Aplikasi Prediksi Penjualan Laptop dengan Menerapkan Metode Regresi Linier,” *Jurnal Pelita Informatika*, vol. 7, no. 2, pp. 179–184, Oct. 2018.

[5] Ulfie Safitri and L. Amaliana, “Model Geographically Weighted Regression dengan Fungsi Pembobot Adaptive dan Fixed Kernel pada Kasus Kematian Ibu di Jawa Timur,” *Jurnal Statistika dan Aplikasinya*, vol. 5, no. 2, pp. 208–220, Dec. 2021, doi: 10.21009/JSA.05209.

[6] L. O. Asril, M. I. Rizki, A. Nugraha, and T. A. Taqiyyuddin, “Geographically Weighted Regression Pada Penyebaran Penyakit Tuberkulosis Pada Anak di Kota Bandung,” *Seminar Nasional Statistika X*, vol. 10, Dec. 2021, [Online]. Available: http://prosiding.statistics.unpad.ac.id

[7] “Badan Pusat Statistik,” Badan Pusat Statistik.

[8] J. Saputra, “Upaya Pemerintah Mengurangi Kemiskinan Dengan Program Bedah Rumah di Kota Jambi,” *ILTIZAM Journal of Shariah Economics Research*, vol. 6, no. 2, pp. 241–249, Dec. 2022, doi: 10.30631/iltizam.v6i2.1524.

[9] M. Tranmer, J. Murphy, M. Elliot, and M. Pampaka, “Multiple Linear Regression (2nd Edition),” 2020. [Online]. Available: https://hummedia.manchester.ac.uk/institutes/cmist/a

[10] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 6th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., 2021.

[11] G. A. Halim, P. Agustin, E. Adiwijayanto, and M. Ohyver, “Estimation of cost of living in a particular city using multiple regression analysis and correction of residual assumptions through appropriate methods,” *Procedia Comput Sci*, vol. 216, pp. 613–619, 2023, doi: 10.1016/j.procs.2022.12.176.

[12] S. Weisberg, *Applied Linear Regression*, 4th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., 2013.

[13] S. Chatterjee and A. S. Hadi, *Regression Analysis by Example*, 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., 2012.

[14] R. Putra, S. Wahyuning Tyas, and M. G. Fadhlurrahman, “Geographically Weighted Regression with The Best Kernel Function on Open Unemployment Rate Data in East Java Province,” *Enthusiastic : International Journal of Applied Statistics and Data Science*, pp. 26–36, Apr. 2022, doi: 10.20885/enthusiastic.vol2.iss1.art4.

[15] N. Sunusi and A. Subarkah, “GEOGRAPHICALLY WEIGHTED REGRESSION WITH DIFFERENT KERNELS: APPLICATION TO MODEL POVERTY,” *Indonesian Journal of Applied Research (IJAR)*, vol. 4, no. 1, pp. 27–41, Apr. 2023, doi: 10.30997/ijar.v4i1.283.

[16] D. Ardianti, H. Pramoedyo, and N. Nurjannah, “Distance weight of GWR-Kriging model for stunting cases in East Java,” *J Phys Conf Ser*, vol. 1968, no. 1, p. 012028, Jul. 2021, doi: 10.1088/1742-6596/1968/1/012028.

[17] J. P. Sinaga, A. Hendriadi, M. Firdaus, A. Fauzi, and I. W. Arsanti, “Analysis of Geographically Weighted Regression (GWR) on Retail Prices of Medium Rice and Red Chili in Java,” *Jurnal Manajemen dan Agribisnis*, Jul. 2021, doi: 10.17358/jma.18.2.144.

[18] Z. Shabrina, B. Buyuklieva, and M. K. M. Ng, “Short‐Term Rental Platform in the Urban Tourism Context: A Geographically Weighted Regression (GWR) and a Multiscale GWR (MGWR) Approaches,” *Geogr Anal*, vol. 53, no. 4, pp. 686–707, Oct. 2021, doi: 10.1111/gean.12259.

[19] Suritman, Raupong, and A. Kalondeng, “Pemodelan Mixed Geographically Weighted Regressionyang  Mengandung Multikolinearitas dengan Regresi Ridge,” *ESTIMASI: Journal of Statistics and Its Application*, vol. 4, no. 1, pp. 100–113, Jan. 2023.

[20] M. B. Nirwana and D. Wulandari, “Comparison of Simple and Segmented Linear Regression Models on the Effect of Sea Depth toward the Sea Temperature,” *Enthusiastic : International Journal of Applied Statistics and Data Science*, vol. 1, no. 2, pp. 68–75, Dec. 2021, doi: 10.20885/enthusiastic.vol1.iss2.art3.

[21] F. A. M. Mediavilla, “A Comparison of the Coefficient of Predictive Power, the Coefficient of Determination and AIC for Linear Regression,” 2008. [Online]. Available: https://api.semanticscholar.org/CorpusID:3740510

[22] C. Lindsey and S. Sheather, “Variable Selection in Linear Regression,” *The Stata Journal: Promoting communications on statistics and Stata*, vol. 10, no. 4, pp. 650–669, Dec. 2010, doi: 10.1177/1536867X1101000407.

[23] M. Keer, H. Lohiya, and S. Chouhan, “Goodness of Fit for Linear Regression using R squared and Adjusted R-Squared,” *International Journal of Research Publication and Reviews Journal homepage: www.ijrpr.com*, vol. 4, pp. 2431–2439, 2023, [Online]. Available: www.ijrpr.com