Sainmatika: Jurnal Ilmiah Matematika dan Ilmu Pengetahuan Alam
Volume 21 No. 2, December 2024 DOI 10.31851/sainmatika.v21i2.15416
https://jurnal.univpgri-palembang.ac.id/index.php/sainmatika

Comparison of Ordinal Logistic Regression and Geographically Weighted Ordinal Logistic Regression (GWOLR) in Predicting Stunting Prevalence among Indonesian Toddlers

Silfiana Lis Setyowati[1,2*], Indahwati[1], Anwar Fitrianto[1], Erfiani[1], Muftih Alwi Aliu[1]
*e-mail: silfianalissetyowati@gmail.com

[1]*Departement of Statistics and Science Data, School of Data Science, Mathematics and Informatics IPB University*
[2]*Ministry of Higher Education, Science, and Technology*

ABSTRACT

Ordinal logistic regression is a type of logistic regression used for response variables with an ordinal scale, containing two or more categories with levels between them. This method is an extension of logistic regression where the observed response variable is ordinal with a clear order. It addresses spatial effects that can cause variance heterogeneity and improve parameter estimation accuracy compared to logistic regression. Geographically Weighted Regression (GWR) is a statistical analysis technique designed to account for spatial heterogeneity. GWOLR is an extension of OLS and GWR models that incorporates spatial elements into regression with categorical variables. This study compares the effectiveness of OLR and GWOLR in analyzing stunting prevalence in toddlers. Comparing OLR and GWOLR can help assess the spatial impact on stunting prevalence. This analysis could reveal that certain regions have a higher tendency for stunting prevalence, while others might have lower tendencies, thus helping in understanding regional disparities. Toddler height is a key indicator of health and nutrition in early growth. The prevalence of stunting for toddlers, according to WHO, is categorized into four levels: low, moderate, high, and very high. The Ordinal Logistic Regression model is better suited for modeling toddler stunting prevalence in Indonesia than the GWORL model. The Ordinal Logistic Regression model and the GWOLR both have a classification accuracy of 85.7%, but the OLR model has a lower AIC value. The GWOLR model is not suitable for analyzing stunting prevalence among Indonesian toddlers due to the lack of spatial variability in the data. The Breusch-Pagan test results indicate that there is no spatial heterogeneity in the data on stunting prevalence among Indonesian toddlers, as the p-value is less than the significance level of 0.05. The prevalence of undernourished toddlers is the main factor influencing stunting among Indonesian toddlers.

Keywords: Ordinal Logistic Regression, Geographically Weighted Ordinal Logistic Regression, Stunting

INTRODUCTION

The link between the stunting case and regression models, such as Ordinal Logistic Regression (OLR) and Geographically Weighted Ordinal Logistic Regression (GWOLR), is their ability to examine the factors influencing stunting prevalence in toddlers. OLR and GWOLR are tools that can help analyze the factors contributing to stunting in toddlers. Stunting, a condition where a child's height is below the expected range for their age, is influenced by socio-economic, environmental, and biological factors. OLR can be used to analyze stunting prevalence categorized as low, moderate, high, and very high, based on predictors like undernutrition prevalence, without considering spatial effects.

Meanwhile, GWOLR extends this analysis by incorporating spatial effects, assuming that stunting rates may vary across different geographic areas due to spatial heterogeneity.

Categorical response variables are commonly found in research across various fields like social sciences, economics, and health. If the response variable in classical linear regression is categorical, classical linear regression is not suitable. In classical regression analysis, it is necessary for the residual distribution to be normal. Categorical response variables can lead to non-normal residual distributions. Logistic regression is used to address this issue. (Jalarno & Ispriyanti,2012), residuals in logistic regression do not require the assumption of normality. Logistic regression provides estimates that can be directly interpreted in the form of odds ratios, offering an intuitive understanding of the strength of the relationship between predictor variables and the likelihood of each category in the response variable. Therefore, logistic regression not only fills the gap left by linear regression but also offers a rich and flexible method that can be used in these challenging analysis scenarios.

Ordinal logistic regression is a type of logistic regression used for response variables with an ordinal scale, containing two or more categories with levels between them. This method is an extension of logistic regression for ordinal response variables with a defined order. The coefficients in ordinal logistic regression are interpreted as log-odds to determine the likelihood that an event will be categorized in or above a certain category compared to the category below it. The odds ratio from the model indicates the strength and direction of the relationship between independent variables and the likelihood of the dependent variable.

Logistic regression does not accommodate spatial effects that can trigger heterogeneity in variance, resulting in less accurate parameter estimation. One statistical analysis technique developed to accommodate the influence of heterogeneity caused by spatial effects is the Geographically Weighted Regression (GWR) model. Furthermore, the development of this methodology includes Geographically Weighted Ordinal Logistic Regression (GWOLR), which extends the GWR framework to handle data where the response variable is ordinal. Geographically Weighted Ordinal Logistic Regression combines ordinal logistic regression with GWR to account for both the categorical order in the response variable and spatial variation in the influence of independent variables. GWOLR offers precise estimates of the local relationships between independent variables and response categories in the dependent variable.

Toddler height is a key indicator of health and nutrition in early growth. Genetic factors, nutritional intake, and environmental conditions play crucial roles in determining a child's physical growth. Another important factor to consider is the child's history of infectious diseases. Infections in early childhood, like diarrhea, measles, or respiratory illnesses, can disrupt normal growth.

A study using data from Nigeria Demographic and Health Surveys (2008-2018) found that inadequate housing, sanitation, and water sources were linked to childhood illnesses, impacting nutrition and growth, including height(Iyanda et al., 2021). Other research has shown a link between the duration of diarrhea in toddlers and stunting, as well as the relationship between acute respiratory infections and child height (Iyanda et al., 2021). Researching the link between toddler height and disease history is crucial in Indonesia due to the high rates of infectious diseases and various health challenges influenced by socioeconomic and environmental factors.

A high prevalence of stunting indicates that children are not receiving adequate nutrition during their critical growth period, which is the first 1000 days of life, encompassing pregnancy until the age of two. Given the serious long-term impacts of stunting, research and efforts to reduce the prevalence of stunting become a crucial priority, especially with the Indonesian government's target in *Rencana Pembangunan Jangka Menengah Nasional* (RPJMN) 2020-2024 to reduce the prevalence of stunting to 14% by 2024.

Stunting data in Indonesia highlights significant challenges in achieving the set targets. In 2018, the stunting prevalence among children under five was 31.4%. By 2022, it had only decreased to 21.6%, falling short of the 2024 target of 14%. The high rate of stunting is concerning as it indicates chronic undernutrition, impacting physical growth and cognitive development, which can result in long-term health problems. These effects not only harm individual health but also impede economic growth and national development.

Several studies have explored logistic regression models and the GWOLR model. Budiman and Cahyani (2022) used ordinal logistic regression to analyze 2020 Human Development Index (HDI) data with a classification accuracy of 76.3%. Pradita et al. (2015) applied GWOLR to study factors influencing HDI in East Java, achieving a classification accuracy of 86.84%. This suggests that spatial factors play a significant role in HDI and can inform strategic development policies.

Rahmadeni and Puspita (2023) studied the use of ordinal logistic regression to analyze toddler nutrition at the Limapuluh Pekanbaru Health Center. The model was deemed unsuitable for capturing the complexity of factors influencing nutritional status in this population. In contrast, Muhtadi utilized Geographically Weighted Ordinal Logistic Regression (GWOLR) to address stunting in West Kalimantan, achieving a 64.29% classification accuracy. Muhtadi's research shed light on the spatial distribution of stunting, guiding targeted interventions. This study aims to compare the performance of ordinal logistic regression and GWOLR models on Indonesian toddler stunting prevalence data. Comparing OLR and GWOLR is crucial to determine the most effective model. OLR identifies factors influencing stunting, while GWOLR considers spatial effects, helping us assess spatial influences on stunting cases. This understanding helps the government to tailor interventions more effectively according to the specific needs of each region, thereby enhancing the precision of policies and resource allocation to reduce stunting rates.

MATERIAL AND METHOD

The data used in this study is secondary data on the prevalence of toddler height for the year 2018, obtained from the BPS-Statistics Indonesia. Surveys on the prevalence of stunting in Indonesia are conducted every five years. The latest data on stunting prevalence available on the BPS website is from 2018.

The research units consist of 34 provinces in Indonesia. The minimum sample size for ordinal regression should ideally be more than 30 overall to ensure the stability of model estimates and reduce excessive variability in the results. In some cases, ordinal regression analysis can still yield valid results as long as the model assumptions are met. Hosmer and Lemeshow (2013) suggest that the minimum sample size for logistic regression analysis should not be less than 30 to minimize variability in the estimates and enhance the reliability of the model.

The response variable in this study is the prevalence of toddlers with short and very short height (stunting). The prevalence of toddlers with short and very short height is categorized into 4 groups (WHO, 2010): (1) low prevalence (less than 20%), (2) moderate prevalence (20%-29%), (3) high prevalence (30% - 39%), and (4) very high prevalence (more than 40%).

Infectious diseases in toddlers can affect their growth and development. Diarrhea and respiratory illnesses in early childhood can impact nutritional status and height (Jelenkovic et al., 2016). Maternal conditions during pregnancy are also crucial, as factors such as maternal nutrition and exposure to harmful substances can affect fetal growth. During pregnancy, the mother's nutrition and mental well-being play a crucial role in the child's long-term development and health (Naaz an Muneshwar, 2023). Finally, health behaviors and environmental conditions play a crucial role in child growth and development. The predictor variables considered in this study include toddler health, maternal conditions during pregnancy, health behaviors, and environmental factors, as outlined in Table 1.

The response variable scale is ordinal with categories:

1 =Low Prevalence of Stunting;
2 =Moderate Prevalence of Stunting;
3 =High Prevalence of Stunting; and
4 =Very High Prevalence of Stunting

while all predictor variables are measured on a ratio scale.

Table 1. Research Variables

Variable	**Definition**
Y	Prevalence of Toddlers with Short and Very Short Height (stunting). Categories into 4 groups: (1) low prevalence (less than 20%); (2) moderate prevalence (20%-29%); (3) high prevalence (30% - 39%); and (4) very high prevalence (more than 40%).
X1	Percentage of Infants Under 6 Months Who Receive Exclusive Breastfeeding
X2	Percentage of Ever-Married Women Aged 15-49 Years Who Gave Birth in a Health Facility
X3	Percentage of Smoking Among Population Aged ≥ 15 Years
X4	Percentage of Obese Toddlers (Weight/Height) in the Age Group 0-59 Months
X5	Prevalence of Malnourished Toddlers
X6	Percentage of Households with Access to Sustainable and Safe Drinking Water Services
X7	Percentage of Urban Slum Households
X8	Percentage of Households with Adequate Sanitation

The analysis procedure involves checking multicollinearity among predictor variables. Variables with Variance Inflation Factor (VIF) values exceeding 10 will be excluded from the model because high VIF values indicate significant multicollinearity and require special attention (Vatcheva et al., 2024).

The data is divided into two parts, with 80% for training data and 20% for testing data. Next, an ordinal logistic regression model is developed, and a simultaneous test is conducted to determine whether all predictor variables collectively influence the response variable. The individual effects of each predictor variable on the response variable can be observed by conducting partial tests using the Wald test statistic. The obtained model is then interpreted.

In the analysis using the GWOLR method, the longitude and latitude coordinates (u_i and v_i) of the provincial capitals in Indonesia are used to determine the weighting matrix by incorporating Euclidean distances and optimal bandwidth values, where $j=1,2,...,34$ so that each location will have a weighting of 34. Next, parameter estimation of the GWOLR model is performed, followed by simultaneous and partial tests using the Wald test statistic. The obtained model is then interpreted.

Classification accuracy is a common metric used to evaluate model performance in classification problems (James, 2013). In ordinal logistic regression, an accuracy above 70% is considered good. Liu (2015) cautions that higher accuracy does not always mean a better model, as it could suggest oversimplification. Therefore, in addition to classification accuracy, it is crucial to assess other metrics like the Akaike Information Criterion (AIC) for a comprehensive evaluation of model performance.

The classification accuracy and Akaike Information Criterion (AIC) values were calculated to compare the ordinal logistic regression and GWOLR models. The best model was chosen based on the highest accuracy and the lowest AIC value. Data analysis was performed using R software.

RESULT AND DISCUSSION

The target prevalence of stunting set by the World Health Assembly (WHA) is to reduce the stunting rate by 40 percent from the prevalence in 2013, which was 22 percent, to 13.2 percent by 2025 (Acceleration of Poverty Reduction, 2019). Stunting is a major concern in the health sector in many developing countries, including Indonesia. According to the results of the Basic Health Research (Riskesdas) in 2018, the prevalence of stunting among toddlers in Indonesia is 30.8%. Figure 1 is classified as high prevalence according to WHO standards, as it falls below the 20% threshold for low prevalence. The Indonesian government aims to reduce stunting among toddlers to 14% by 2024 as outlined in the National Medium-Term Development Plan (Bappenas, 2020).

To achieve the national target reduction in stunting prevalence, each province in Indonesia must reduce its stunting prevalence. In 2018, Indonesia's stunting prevalence was categorized as low, moderate, high, or very high based on BPS data.

Figure 1 shows the distribution of stunting prevalence in Indonesia. Half of the total number of provinces, which is 17 provinces, fall into the category of high stunting prevalence, while 41%, or 14 provinces, fall into the category of moderate prevalence. There is one province with low stunting prevalence, which is DKI Jakarta, while there are 2 provinces that require special attention as they fall into the category of very high stunting prevalence, West Sulawesi and East Nusa Tenggara provinces.

Figure 1. Map of Indonesian Toddlers Stunting Prevalence

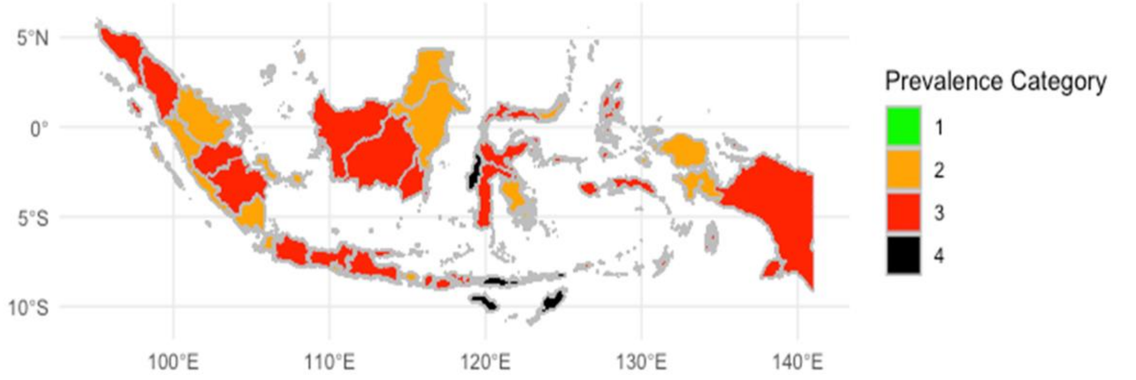

Table 2. Descriptive Statistic Predictor Variables

Variable	Min	Max	Mean
X1	25.69	64.28	45.43
X2	33.91	99.87	76.18
X3	25.80	36.56	31.78
X4	3.300	13.20	7.588
X5	13.00	29.50	19.21
X6	36.89	88.39	64.60
X7	2.750	17.00	9.086
X8	33.75	91.14	68.35

Table 2 displays the characteristics of each predictor variable. There is a notable disparity in exclusive breastfeeding rates among infants under six months, ranging from 25.69% to 64.28% with an average of 45.43%. Additionally, the percentage of births in health facilities (X2) varies widely from 33.91% to an average of 76.28%, indicating unequal access to healthcare services for women across Indonesian provinces.

The smoking rate among individuals aged 15 years and older in Indonesia is 31.78%, highlighting a significant prevalence of smoking that could affect public health. The obesity rates among toddlers (X5) range from 3.3% to 13.2%, averaging 7.588%, indicating varying nutritional health issues among provinces. The average prevalence of malnourished toddlers (X6) at 19.21% also reflects significant levels of undernutrition in various regions. From an environmental health perspective, the average percentage of households with access to sustainable and safe drinking water services (X7) is 64.6%, the percentage of urban slum households (X8) is 9.086%, and the percentage of households with adequate sanitation is 68.35%, highlighting the challenges and disparities in access to basic services across Indonesia.

Table 3. VIF Values of Reseach Variables

Variable	VIF
X1	1.37
X2	3.29
X3	1.56
X4	3.66
X5	2.84
X6	2.20
X7	1.46
X8	3.11

The assumption that must be met in logistic regression is multicollinearity, which means there are no predictor variables that have high correlation with each other. This assumption is crucial because multicollinearity can lead to inaccurate coefficient estimates. Violating this assumption may result in high standard errors, unreliable statistical tests, and less interpretable models, which can lead to misleading conclusions and reduced predictive power. Table 3 displays the VIF values for each predictor variable. A VIF value greater than 10

p-ISSN 1829 586X
e-ISSN 2581-0170

indicates multicollinearity (Vatcheva et al., 2024). All predictor variables in this study have VIF values below 10, indicating no multicollinearity.

Ordinal Logistic Regression Model

The ordinal logistic model parameters were tested using both simultaneous and partial tests. The G test statistic, a likelihood ratio test, was used for overall parameter testing. Simultaneous testing in the model helps determine if predictor variables collectively influence the response variable. Hypothesis testing is conducted for:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_8 = 0$
H_1: at least there is one $\beta_k \neq 0$, $k = 1,2,\ldots,8$

The procedure of likelihood ratio test comparison can be used to evaluate the significance of the ordinal logistic regression model. The G test statistic is used to assess the overall influence of predictor variables in the model (Hosmer & Lemeshow, 2013). This test compares a full model with predictor variables to a model with only an intercept to assess if adding predictors significantly improves the model fit. The formula is:

$$G = -2ln\left[\frac{Likelihood\text{ Model A}}{Likelihood\text{ Model B}}\right]$$

Model A is the restricted or simpler model, which may only include a constant or a few basic predictor variables. This model has fewer parameters or lacks certain effects being tested. Model B has more parameters than Model A. Model A is a simplified version with fewer variables or effects, while Model B is a more comprehensive version with all variables and interactions included. Model A has limited parameters, while Model B has more parameters.

Table 4. Summary of Simultaneous Testing in Ordinal Logistic Regression Model

Model	df	Log-Likelihood	Chisq	P-value
Final	11	-13.78		
Intercept Only	3	-27.96	28.35	0.0004

Table 4 shows the results of simultaneous testing in the ordinal logistic regression model. The G test statistic value is 28.351 with a p-value of 0.0004, indicating significance at the 0.05 level. Thus, we reject the null hypothesis H_0. \The rejection of the null hypothesis (H0) suggests that the ordinal logistic regression model is suitable for the data, as at least one predictor variable is significant. The predictor variables have a significant influence on the prevalence of stunting among toddlers in Indonesia.

In logistic regression, a partial test is conducted using the Wald test, which compares coefficient estimates to their standard errors to determine if they are statistically different from zero (Bewick et al., 2005). The Wald test evaluates the significance of each coefficient in the model. The partial hypothesis for the ordinal logistic regression model is:

$H_0: \beta_k = 0$
$H_1: \beta_k \neq 0, k = 1,2,\ldots,8$

Table 5. Summary of Partial Testing in Ordinal Logistic Regression Model

Variable	Parameter Estimation	Wald	P-value
X1	0.080	1.431	0.152
X2	0.061	1.228	0.219
X3	-0.272	-1.061	0.288
X4	0.746	1.522	0.128
X5	0.992	2.544	0.010*
X6	-0.122	-1.580	0.112
X7	-0.024	-0.166	0.868
X8	-0.002	- 0.033	0.973
1|2	9.138	0.519	0.603

Variable	Parameter Estimation	Wald	P-value
2|3	14.943	0.835	0.403
3|4	23.388	1.242	0.214

*significant at 0.05

Thus, the ordinal logistic regression model obtained is as follows:

$$logit[P(Y_1 \leq 2)|x_i] = 9.138 + 0.080_{1i} + 0.061_{2i} - 0.272_{3i} + 0.746_{4i} + 0.992_{5i} - 0.122_{6i} - 0.024_{7i} - 0.002_{8i}$$

$$logit[P(Y_1 \leq 3)|x_i] = 14.943 + 0.080_{1i} + 0.061_{2i} - 0.272_{3i} + 0.746_{4i} + 0.992_{5i} - 0.122_{6i} - 0.024_{7i} - 0.002_{8i}$$

$$logit[P(Y_1 \leq 4)|x_i] = 23.388 + 0.080_{1i} + 0.061_{2i} - 0.272_{3i} + 0.746_{4i} + 0.992_{5i} - 0.122_{6i} - 0.024_{7i} - 0.002_{8i}$$

Table 5 shows the results of partial tests for estimating each predictor variable against the response variable. The variable influencing the prevalence categories of stunting is X5, which is the prevalence of malnourished toddlers by Province in Indonesia. The AIC value of the ordinal logistic regression model is 46.986.

Table 6. Parameter Estimattion Ordinal Logistic Regression Model Y and X5

Variable	Parameter Estimation	Standard Error	t-value
X5	0.398	0.137	2.915
1|2	3.305	2.347	1.408
2|3	6.997	2.450	2.855
3|4	12.176	3.588	3.393

The partial test results show that only variable X5 is significant, therefore a re-modeling is conducted. Table 6 shows the parameter estimation results of the ordinal logistic regression model between the prevalence of stunting in toddlers in Indonesia and the prevalence of malnutrition in toddlers. Thus, the new ordinal logistic regression model is:

$$logit[P(Y_1 \leq 2)|x_i] = 3.305 + 0.398_{5i}$$
$$logit[P(Y_1 \leq 3)|x_i] = 6.997 + 0.398_{5i}$$
$$logit[P(Y_1 \leq 4)|x_i] = 12.176 + 0.398_{5i}$$

The AIC value of the model is 46.696. To determine the influence of variable X5, the odds ratio is calculated. The odds ratio for variable X5 is 1.402. This means that toddlers who suffer from malnutrition have 1.402 times higher odds of being at risk of stunting compared to those who do not suffer from it. Therefore, it can be concluded that the higher the prevalence of malnutrition among toddlers, the higher the prevalence of stunting among toddlers will be.

Model *Geographically Weighted Ordinal Logistic Regression*

The Breusch-Pagan test is used to detect spatial effects in the data. A heteroscedasticity test is conducted to check the consistency of error variance in the model (Iffah et al., 2023). The hypothesis for testing spatial heterogeneity is:

$H_0: \sigma_i^2 = \sigma^2$
H_1: at least there is one$\sigma_i^2 \neq \sigma^2$
$i = 1,2, \dots, 34$

The statistical value of the BP test obtained is 9.064 with a p-value of 0.336. The p-value obtained is<α (0.05), which means there is no spatial heterogeneity in the data of the Indonesian toddlers stunting prevalence.

The first step in modeling children's height prevalence with the GWOLR model involves identifying geographical locations based on the latitude and longitude of provincial capital offices. The Euclidean distance between locations i and j is calculated, and the optimal bandwidth is determined using the Cross-Validation (CV) method, resulting in a bandwidth of 10.144. The study uses adaptive Gaussian kernel weighting as the weighting function. The weights assigned to each province are used to estimate the parameters of the GWOLR model. The model is then simultaneously tested by evaluating the hypothesis:

$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_8 = 0$
H_1: at least there is one $\beta_k(u_i, v_i) \neq 0$, $k = 1,2, \ldots 8$

Table 7. Summary of Simultaneous Testing in GWOLR

Model	df	Log-Likelihood	Chisq	P-value
Final	11	-10.88		
Intercept Only	3	-21.99	22.20	0.004

Table 7 summarizes the simultaneous test of the GWOLR model, with a test statistic value of 22.205 and a p-value of 0.004. The results indicate that the predictor variables in the model have a statistically significant impact on stunting prevalence at a 95% confidence level.

The GWOLR model parameters are partially tested by evaluating the hypothesis:

$H_0: \beta_k(u_i, v_i) = 0$
$H_1: \beta_k(u_i, v_i) \neq 0, k = 1,2, \ldots ,8$

Each province in Indonesia has its own unique model. For instance, in East Java Province, the p-value test value for each parameter can be observed when testing the parameter β_kin.

Table 8. Summary of Partial Testing in GWOLR on East Java Province

Variable	Parameter estimation	Wald	P-value
X1	0.011	1.392	0.191
X2	0.004	0.745	0.471
X3	-0.032	-0.925	0.374
X4	0.085	1.391	0.191
X5	0.112	3.796	0.003*
X6	-0.013	-1.268	0.231
X7	0.005	0.255	0.803
X8	-0.004	-0.411	0.688
1|2	5.207	0.323	0.800
2|3	11.592	0.704	0.609
3|4	-0.034	1.252	0.429

*significant at α=0.05

Table 8 shows that the prevalence of malnourished toddlers is the key factor influencing stunting rates in East Java. While each province has unique models, the overall analysis reveals that only variable X5 significantly impacts stunting prevalence. The Breusch-Pagan test confirms that there is no spatial variation in stunting rates among Indonesian toddlers.

Comparison of Ordinal Logistic Regression and GWOLR Model

To evaluate the accuracy of a classification model, the confusion matrix is a useful tool. It shows the frequency of correct and incorrect classifications of behaviors. Performance indicators like sensitivity and specificity are commonly used to measure classification accuracy (Ruuska et al., 2018).

Table 9. Confussion Matrix for Testing Data of Ordinal Logisstic Regression Model

Prediction	Reference		
	2	3	4
2	3	0	0
3	0	3	1
4	0	0	0

p-ISSN 1829 586X
e-ISSN 2581-0170

Table 10. Sensitivity and Specifity Ordinal Logistic Regression Model

Category	Sensitivity	Specifity
2	1	1
3	1	0.75
4	0	1

The confusion matrix for the ordinal logistic regression model is presented in Table 9, and Table 10 displays the sensitivity and specificity values. In Table 9, prediction 4 includes two provinces, West Sulawesi and East Nusa Tenggara, which were not in the testing data set, resulting in a value of "0" in the confusion matrix for this category.

The sensitivity values indicate that all data in categories 2 and 3 were predicted correctly, while no data in category 4 was predicted accurately. The low sensitivity for category 4 is due to the small number of data points in that category. The specificity values show that data not in classes 2 and 4 were correctly classified. The model accurately identified 75% of data not in category 3. Category 1 was not included in the confusion matrix as it only consisted of one province, DKI Jakarta, which was not part of the testing data for the ordinal logistic regression model.

Table 11. Confussion Matrix for Testing Data of GWOLR Model

Prediction	Reference			
	1	2	3	4
1	0	1	0	0
2	1	2	0	0
3	0	0	2	0
4	0	0	0	0

The confusion matrix for the GWOLR model can be seen in Table 11, and the sensitivity and specificity values in Table 12.

Table 12. Sensitivity and Specifity GWOLR Model

Category	Sensitivity	Specifity
1	0.000	0.667
2	0.833	0.750
3	1	1
4	1	1

The model accurately classified all data in categories 3 and 4, and 83% of the data in category 2. However, it did not correctly classify any data in category 1, as there was only one data point in that category (DKI Jakarta out of 34 provinces in Indonesia). The model successfully identified all data not belonging to categories 3 and 4 as such. Specifically, 75% of the data not in category 3 and 67% of the data not in category 1 were correctly classified by the model.

Table 13 compares the modeling of stunting prevalence in Indonesian toddlers using ordinal logistic regression and GWORL in terms of classification accuracy and AIC values.

Table 13. Comparison of Models Based on Classification Accuracy and AIC

Model	Classification Accuracy	AIC
Ordinal Logistic Regression	0.857	46.696
GWOLR	0.857	52.040

Both the ordinal logistic regression model and the GWOLR provide the same classification accuracy value of 85.7%. This indicates that the model successfully predicts or classifies the data correctly in 85.7% of all cases or data tested, in other words, 85.7% of the model predictions match the actual results or categories, reflecting the model accuracy in identifying or classifying the data accurately. However, in terms of the AIC value, the ordinal logistic regression model has a lower value. This indicates that the ordinal logistic regression model is more suitable for modeling the prevalence of stunting for toddlers in Indonesia.

CONCLUSION

The ordinal logistic regression model is better for modeling the prevalence of stunting for Indonesian toddlers compared to the GWORL model based on the obtained AIC value. Both OLR model and GWOLR model achieve a classification accuracy of 85.7%. However, the OLR model has a lower AIC value, indicating a better fit to the data compared to the GWOLR model.

The GWOLR model is not ideal for analyzing Indonesian toddlers' stunting prevalence data due to the lack of spatial variation. The Breusch-Pagan test can be used to detect spatial effects early on. In this case, the test results show a p-value below α (0.05), indicating no spatial heterogeneity in stunting prevalence among Indonesian toddlers. The prevalence of malnourished toddlers is a significant factor affecting stunting prevalence in Indonesia.

REFERENCES

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression.. *BioMed Central, 9(1), 112-112*. https://doi.org/10.1186/cc3045

BPS-Satistics Indonesia. (2024). Produk-Table Dinamis. Retrieved Mei 12, 2024, from https://www.bps.go.id/id/query-builder

Budiman, M.A., & Cahyani, N. (2022). Pemodelan Regresi Logistik Ordinal Pada Indeks Pembangunan Manusia (IPM) Di Jawa Timur Tahun 2020. *Jurnal Statistika dan Komputasi (STATKOM), I(2), 64-73*. https://doi.org/10.32665/statkom.v1i2.1169

Hosmer D. W., Lemeshow S., & Sturdivant R. X. (2013). *Applied Logistic Regression*. New Jersey: John Wiley & Sons Inc.

Iffah, A., Suliyanto, S., Sediono, S., Saifudin, T., Ana, E., & Amelia, D. (2023). Poverty Modeling in Indonesia: a Spatial Regression Analysis. *Economics Development Analysis Journal*, *12*(4), 441-457.https://doi.org/10.15294/edaj.v12i4.66027

Iyanda, A E., Adaralegbe, A., Miranker, M., Lasode, M K., & Lu, Y. (2021). Housing conditions as predictors of common childhood illness: Evidence from Nigeria Demographic and Health Surveys, 2008–2018. *SAGE Publishing, 25(4), 659-676*.https://doi.org/10.1177/1367493520987049

Jalarno, D., & Ispriyanti, D. (2012). Penentuan Model Regresi Terpotong Atas Dengan Metode MaksimumLikehood. *Media Statistika, 1*(2), 53-62. https://doi.org/10.14710/medstat.1.2.53-62

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. New York: Springer.

Jelenkovic, A., Sund, R., Hur, Y M., Yokoyama, Y., Hjelmborg, J., Möller, S., Honda, C., Magnusson, P K E., Pedersen, N L., Ooki, S., Aaltonen, S., Stazi, M A., Fagnani, C., D'Ippolito, C., Freitas, D., Maia, J., Ji, F., Feng, N., Pang, Z., &Sørensen, T I A. (2016). Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts. *Nature Portfolio, 6(1)*. https://doi.org/10.1038/srep28496

Ministry of National Development Planning/ National Development Planning Agency (Bappenas). (2020).

Rencana Pembangunan Jangka Menengah Nasional 2020-2024. Jakarta: Kementerian PPN/BAPPENAS.

Liu, X. (2015). *Applied Ordinal Logistic Regression Using Stata*. California: Sage Publications.

Ministry of Health. *Buku Saku Hasil Survei Status Gizi Indonesia (SSGI) 2022*. Jakarta: Badan Kebijakan Pembangunan Kesehatan

Naaz, A.,&Muneshwar, K.N. (2023). How Maternal Nutritional and Mental Health Affects Child Health During Pregnancy: A Narrative Review. Cureus, 15(11):e48763.https://doi.org/10.7759/cureus.48763

Pradita, R.N., Yasin, H., & Safitri, D. (2015). Pemodelan Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Kabupaten/ Kota Di Jawa Timur Menggunakan Geographically Weighted Ordinal Logistic Regression. *Jurnal Gaussian, 4(3), 639-650*. http://ejournal-s1.undip.ac.id/index.php/gaussian

Rahmadeni & Puspita, S. (2023). Aplikasi Regresi Logistik Ordinal dalam Pemodelan Status Gizi Balita (Studi Kasus: PuskesmasLimapuluh Di Kota Pekanbaru). Jurnal Sains Matematika dan Statistika, 9(2), 21-30. https://dx.doi.org/10.24014/jsms.v9i2.19080

Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., & Mononen, J. (2018).Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behav Processes, 148:56-62*. doi: https://doi.org/10.1016/j.beproc.2018.01.004

The National Team for the Acceleration of Poverty Reduction. (2019). *Strategi Nasional Percepatan Pencegahan Anak Kerdil (stunting) Periode 2018-2024*. Jakarta: Sekretariat Wakil Presiden Republik Indonesia.

Vatcheva, K.P., Lee, M., McCormick, J.B., &Rahbar, M.H. (2016) Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale). 6(2):227*.https://doi.org/10.4172/2161-1165.1000227.