Mathematical Journal of Modelling and Forecasting
Vol. 1, No. 1, Juni 2023, pp. 24-31

Geographically Weighted Regression Analysis on Cases of Malnutrition Under Five in the West Sumatra

Fifi Sandriani[1] and Helma[2]

[1,2] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Indonesia
[*] fifisandriani.fans2@gmail.com

Abstract. Malnutrition is a condition experienced by toddlers due to low nutrition or nutritional needs that have not been met. Efforts to improve health status by improving the nutritional status of toddlers. This research aims to clarify that the Geographically Weighted Regression model is the best compared to the multiple linear regression model. The data was obtained from the West Sumatra Provincial Health Office for 2020. The dependent variable used is the percentage of cases of malnutrition, and there are several independent variables, namely the percentage of children under five who were given vitamin A, the percentage of babies who were exclusively breastfed, and babies born with low birth weight. The results showed that the GWR model could explain the diversity of cases of malnutrition in children under five by 99% with a total squared error of 0.002 compared with the multiple linear regression model, which can explain the diversity of cases of malnutrition among children under five by 43% with total squared errors of 0.585. It is concluded that the GWR model is the best.

Keywords: Geographically Weighted Regression, Malnutrition, West Sumatra

1 Introduction

Malnutrition is a condition experienced by toddlers due to malnutrition or nutritional needs that have not been met. Children need total nutritional intake at five so their bodies and brains can develop adequately [1]. Malnutrition can lead to susceptibility to infectious diseases [2]. In general, two main factors cause malnutrition: direct and indirect [3]. The factors that directly cause it to occur due to limited availability of food and the presence of infectious diseases, while indirect causes occur due to low food availability at the family level, the mother's parenting style for children, and low access to environmental health services, as well as poor clean and healthy living habits [4]. In West Sumatra Province, there were 411 cases of malnutrition under five children handled in 2016 and decreased in 2017 to 361 cases, in 2018 it increased to 376 cases [5]. According to the Indonesian Ministry of Health, several steps are a solution to the problem of malnutrition, such as; maximizing the quality of handling cases of malnutrition in hospitals/public health center, weighing toddlers every month at integrated healthcare center, providing complementary food for breastfeeding to toddlers, increasing knowledge and skills of mothers to provide additional nutrition (vitamin A pills) to toddlers[6].

Multiple linear regression analysis is needed to see the causes of under-five malnutrition in West Sumatra Province and to know the regression model to determine the relationship between cases of under-five malnutrition and the factors that influence it [7]. Because the cases to be studied vary from area to area, an analysis that produces a model based on the region is needed to measure the relationship between the ratios of different variables at one point of observation and another [8].

Geographically Weighted Regression is a statistical method that can be applied to determine risk factors [9] spatially. The Geographically Weighted Regression model produces a local model and only applies to each observed location [8]. The element of the weighting matrix is used to see the weight of the influence between the proximity of locations, meaning that the closer a location is, the greater the effect of the weight [10].

2 Research Methods

The data used is secondary data obtained based on data from the West Sumatra Provincial Health Office [11]. Several theories will be used in analyzing Geographically Weighted Regression in cases of malnutrition under five in the province of West Sumatra, namely:

2.1 Multiple Linear Regression

Multiple linear regression analysis was applied to predict future situations by measuring several independent variables (X) over the dependent variable (Y). The form of the multiple linear regression equation is stated as follows [12]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon_i \quad (1)$$

With :

y : variabel dependen

β_0 : Constant

$\beta_1 \beta_2 \beta_k$: regression coefficient value

$x_1, x_{2,..,} x_k$: variabel independen

ε : error

2.2 Multicollinearity

The absence of multicollinearity is a condition that the independent variables must meet. The way to detect multicollinearity is by calculating the Variance Inflation Factor value with the following equation [13]:

$$VIF = \frac{1}{1 - R^2} \quad (2)$$

Where if the VIF value is below 10 and has a tolerance value greater than 0.10, it can be said that the multicollinearity-free regression model.

2.3 Simultaneous Test of Linear Regression Models

To clarify whether the regression model obtained affects the relationship between the dependent variable and the independent variable, then do a simultaneous test with the hypothesis for the following simultaneous test:

H_0: $\beta_0 = \beta_1 = \beta_2 = \cdots = \beta_P = 0$

H_0: there is at least one $\beta_i \neq 0, i = 1,2, \dots , p$

With test statistics:

$$F_{hitung} = \frac{MSR}{MSE} \tag{3}$$

The basis for decision-making reject H_0 if $F_{hitung} > F_{\alpha(P;N-P-1)}$

2.4 Partial Test

A partial test determines that the parameter values significantly affect the independent variables. The hypothesis is stated as follows:
$H_0: \beta_m = 0$ where $m = 0,1,2,\dots,k$
$H_0: \beta_m \neq 0$ where $m = 0,1,2,\dots,k$
With test statistics:

$$t_{hit} = \frac{\hat{\beta}_m}{SE(\hat{\beta}_m)} \tag{4}$$

Where:
$\hat{\beta}_m$: parameter guess the value $\hat{\beta}_m$ obtained from the old method
MSE : *Mean Square Error* from the regression model.

2.5 Spatial Heterogeneity

Spatial heterogeneity is used when the data in one place is not the same geographically as other things. As a result of the emergence of spatial heterogeneity, the regression parameters vary spatially and obtain different regression parameters from each observed location [13].

The test statistic used is Breusch Pagan on the following hypothesis:
$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (there is no heterogeneity between regions)
H_1: there is at least one $\sigma_1^2 \neq \sigma^2$ (there is heterogeneity between regions)
The statistical test is expressed in the following equation:

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T f \tag{5}$$

Where:
BP : *Breusch-Pagan* test values
e_i : error for the i-th observation
$\hat{y}$: the average of the dependent variable from all observed locations
Z : X matrix of size $n \times (p+1)$
σ^2 : variety of $e_i, i = 1,2,\dots,n$
Vector element f is $f_i = \left(\frac{e_i^2}{\sigma^2} - 1\right)$
Reject H_0 if $BP > x_p^2$, then there is heteroscedasticity between observation locations.

2.6 Geographically Weighted Regression Models

Geographically Weighted Regression (GWR) is the development of multiple linear regression analysis, and then the parameters are calculated at each observed location so that each observed location has a different parameter value. The GWR model equation is stated as follows [8]:

$$y_i = \beta_0(u_i, v_i) + \sum_k^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \;; i = 1,2, \dots, n \tag{6}$$

With:

y_i : dependent variable vector for location i
x_{ik} : vector of independent variable k for observation location i, k=1,2,…, n
$\beta_0(u_i, v_i)$: GWR model intercept
$\beta_k(u_i, v_i)$: the regression coefficient of the independent variable k for the observed location i
(u_i, v_i) : coordinate point latitude, geographic longitude for the observation location i
ε : error vector for location i.

2.7 Spatial Weighting

Spatial weighting is obtained from the information distance or the distance between the location to be observed and the other observed locations [14]. The GWR weighting can be determined via a kernel function. The kernel function that is imposed is the fixed kernel function, and the equation is stated as follows:

$$w_{ij} = exp\left[-{}^1\!/_2\left(\frac{d_{ij}}{b}\right)^2\right] \tag{7}$$

The value of the constant b is the bandwidth which is the radius where the circle point is still considered to affect the i and d_{ij} Locations are the Euclidean distance between the i and j th observation locations.

2.8 Bandwidth Value Selection

The choice of bandwidth value is essential because the model's accuracy affects the data. The optimum bandwidth value is determined using the Cross Validation or CV method, which is stated in the following equation [15]:

$$CV = \sum_{i=1}^{n} (y_i - \hat{y}_{\neq 1}(b))^2 \tag{8}$$

Where :
$\hat{y}_{\neq 1}(b)$: estimated value y_i
N : sample totals

2.9 Selection Of The Best Models

To determine the magnitude of the opportunity from the model formed based on the data is the goal of selecting the best model while comparing the AIC value, one of the criteria for selecting the best model. The calculation of the AIC value is written in the following equation:

$$AIC = 2N \log_e(\hat{\sigma}) + n \log_e(2\pi) + n \left\{\frac{n + tr(S)}{N - 2 - tr(S\_}\right\} \tag{9}$$

Where $\hat{\sigma}$ Is the estimated value of the residual standard deviation, and S is the hat matrix. To see the best model, you can look at the highest. $R^2$and the smallest SSE and AIC values.

3 Results And Discussion

West Sumatra is administratively divided into 12 regencies and seven cities, and the population in 2020 was 5,534,742 thousand people consisting of 2,786,360 thousand male residents and 2,748,112 thousand female residents. This study had one dependent variable, namely malnutrition under five, and three independent variables, namely toddlers who were given vitamin A, babies who were exclusively breastfed, and babies born with low birth weight.

3.1 Multiple Linear Regression Models

The regression model is considered free from multicollinearity if it has a VIF value for each independent variable less than 10. The following is the result of the VIF value for each variable.

Table 1. Multicollinearity Assumption Test

Variable	VIF	Conclusion
X_1	2,011	Multicollinearity Does Not Occur
X_2	1,970	Multicollinearity Does Not Occur
X_3	1,119	Multicollinearity Does Not Occur

The table displays a VIF value of no more than 10 for all independent variables, meaning there is no multicollinearity or the independent variables are unrelated.

The following is the parameter estimation of the multiple linear regression model in cases of malnourished children under five in West Sumatra.

Table 2. Multiple Linear Regression Model Parameters

Parameters	Estimate	*Standard Error*
Constant	1,617	0,360
X_1	0,002	0,006
X_2	-0,015	0,007
X_3	0,124	0,045

The form of the multiple linear regression model equation in the table above is stated as follows:

$$y = 1{,}617 + 0{,}002X_1 - 0{,}015X_2 + 0{,}124X_3$$

1.1 Hypothesis Test

Simultaneous tests using the F-test can determine the independent variable affecting the dependent variable by using analysis of variance (ANOVA)

Table 3. Anova Test

Source of Variance	Squared Sum	Df	Squared Average	F
Regresi	0,458	3	0,153	3,933
Residual	8,609	15	0,039	
Total	9,066	18		

Based on the table, the F_{hit} value is 3.933 by taking a significance level of α = 5%, the F_{tabel} The value is 3.29. This shows that the value of $F_{hitung} > F_{tabel}$, means that one independent variable influences cases of malnutrition under five in West Sumatra Province.

3.2 Regression Partial Test

The partial test hypothesis is stated as follows:
H_0: the independent variable does not affect cases of malnutrition under five in West Sumatra Province
H_1: the independent variable effect on cases of malnutrition under five in West Sumatra Province
Based on the Table 4 with a 5% confidence interval, it was found that there were independent variables that influenced cases of malnutrition under five in West Sumatra Province, namely babies who were exclusively breastfed (X_2) and babies born with low birth weight (X_3).

Table 4. Regression Partial Test

Variable	t_{hit}	t_{tabel}	Conclusion
Constant	4,145	2,131	Reject H_0
X_1	0,289	2,131	Accept H_0
X_2	-2,156	-2,131	Reject H_0
X_3	2,745	2,131	Reject H_0

3.3 Spatial Heterogeneity Test

Spatial heterogeneity is used in order to be able to test statistics through the Pagan Breusch test. The hypothesis used in the Breusch Pagan test is as follows:
$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (There is no heterogeneity between regions)
H_1: there is at least one $\sigma_1^2 \neq \sigma^2$ (There is no heterogeneity between regions)

Table 5. Spatial Heterogeneity Test

Value BP	DB
10,221	3

Decision making reject H_0 if $BP > x_p^2$ then there is heterogeneity between observation locations. Then the value of $BP = 10{,}221 > x_3^2 = 7{,}8147$ is obtained, meaning there is spatial heterogeneity between locations in the districts/cities in West Sumatra Province.

3.4 Model Geographically Weighted Regression

Determine the GWR model by calculating the Euclidean distance between observed locations from one observed location to another. Then select the optimum bandwidth value based on the distance that affects the observed area. The optimum bandwidth value is 0.098.

Then determine the weights using a fixed Gaussian kernel and estimate the parameter estimators of the Geographically Weighted Regression model for each observed area. GWR model estimates are presented in the Table 6.

Table 6. Estimation of the GWR Model

Parameters	Minimum	Maximum
Constant	-7,8989	6,8779
X_1	-0,0289	0,0556
X_2	-0,1035	0,0475
X_3	-0,2402	0,2767

The table shows that the parameter estimation estimates have negative and positive regression parameter coefficients in several districts/cities observed in West Sumatra Province.

3.5 Selection of the Best Models

Based on data on cases of malnutrition under five in West Sumatra Province by comparing the R^2 value and the SSE value in the global regression with GWR, Table 7.

Table 7. Selection of the Best Models

Model	R^2	SSE
Regresi Global	43%	0,585
GWR	99%	0,002

The GWR model is stated to be better than the linear regression model because the GWR model can increase the amount of R^2 by 99% and decrease the number of SSE by 0.002. So it is concluded that the GWR model is the best.

4 Conclusion

Based on the data used, it contains spatial effects, so the GWR modeling is done by calculating the weight using a fixed Gaussian kernel and getting different results for each district/city in West Sumatra Province. The GWR model can also explain that the GWR model can explain the diversity of cases of malnutrition under five in the Province of West Sumatra by 99% with a total squared error of 0.002 compared to the multiple linear regression model, which can explain the diversity of cases of under-five malnutrition by 43% with the sum of the squares error of 0.585. Moreover, the factors affecting malnutrition sufferers for each district/city are exclusively breastfed babies born with low birth weight.

5 References

[1] B. & A. D. Y. Sutomo, *Menu Sehat Alami Untuk Batita dan Balita*. Agromdeia Pustaka, 2010.

[2] S. Notoatmodjo, “Prinsip-Prisnip Dasar Ilmu Kesehatan Masyarakat,” Jakarta: Rineka Cipta, 2003.

[3] Soekirman, *Ilmu Gizi dan Aplikasinya Untuk Keluarga dan Masyarakt*. Jakarta: Ditjen DIKTI-DEPDIKNAS, 2000.

[4] D. P. N. T. L. Damayanti, “Gizi Dalam Daur Kehidupan,” *Syria Stud.*, vol. 7, no. 1, pp. 37–72, 2017, [Online]. Available: https://www.researchgate.net/publication/269107473_What_is_governance/link/548173090cf22525dcb61443/download%0Ahttp://www.econ.upf.edu/~reynal/Civil wars_12December2010.pdf%0Ahttps://think-asia.org/handle/11540/8282%0Ahttps://www.jstor.org/stable/41857625.

[5] Badan Pusat Statistik Provinsi Sumatera Barat, “Profil Kesehatan Provinsi Sumatera Barat Tahun 2020,” *13520.2005*, p. 57, 2020.

[6] R. I. DepartemenKesahatan, “Rencana Aksi Nasional Pencegahan dan Penanggulangan Gizi Buruk 2005-2009,” *Kementeri. Kesehat. RI*, 2005.

[7] D. . ; E. A. P. Montgomery, *Introduction to Linear Regression Analysis*. New York: Jhon Wiley & Sons, 2006.

[8] A. S. B. C. C. M. Fotheringham, *Geographically Weighted Regression the analysis of spatially varying relationships*, vol. 1999, no. December. England: Jhon Wiley & Sons, LTD, 2002.

[9] H. Yasin, "Pemilihan Variabel Pada Model Geographically Weighted Regression," *Media Stat.*, vol. 4, no. 2, pp. 63–72, 2011, doi: 10.14710/medstat.4.2.63-72.

[10] R. & A. D. Rahmawati, “Analisis Geographically Weighted Regression (GWR) Dengan Pembobot Kernel Gaussian Untuk Data Kemiskinan,” *Pros. Semin. Nas. Stat. Univ. Diponegoro*, vol. 47, no. 4, pp. 124–134, 2021, doi: 10.31857/s013116462104007x.

[11] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta, 2012.

[12] R. K. Sembiring, *Analisis Regresi*. Bandung: ITB, 1995.

[13] L. Anselin, "Spatial Models in Econometric Research," *Oxford Res. Encycl. Econ. Financ.*, no. March 1998, doi: 10.1093/acre for/9780190625979.013.643.

[14] C.-L. Mei, "Geographically weighted regression technique for spatial data analysis," *Complex Stat. Chinese Acad. ...*, no. 1, 2004.

[15] N. Lutfiani and S. Mariani, “Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-square,” *UNNES J. Math.*, vol. 5, no. 1, pp. 82–91, 2017, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/ujmUJM8.