World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Geographically weighted regression in malnourished toddlers with adaptive kernel bi-square weighting

Shafa Salsabila * and Desie Rahmawati

Department of Epidemiology, Biostatistics, Population Studies and Health Promotion, Faculty of Public Health, Airlangga University. Surabaya, East Java, Indonesia.

World Journal of Advanced Research and Reviews, 2023, 18(01), 1042–1047

Publication history: Received on 10 March 2023; revised on 19 April 2023; accepted on 22 April 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.18.1.0704

Abstract

Spatial data modeling cannot use linear regression due to the presence of unfulfilled assumptions, namely heteroscedasticity. One of the modeling that can be done is Geographically Weighted Regression (GWR) with weighting, one of which is Adaptive Kernel Bi-square. This study aims to prove that the GWR Adaptive Kernel Bi-square model can analyze and interpret factors that have a significant effect on cases of malnourished toddlers in East Java Province. This study used GWR Adaptive Kernel Bi-square analysis using secondary data on the Health Profile of East Java Province in 38 districts/cities in 2018. The results of GWR modeling with the Adaptive Kernel Bi-square weighting function, namely the value of the coefficient in each region is different and the difference in predictor variables has a significant effect. The GWR Adaptive Kernel Bi-square model can analyze and interpret factors that have a significant effect on cases of malnourished toddlers in East Java Province, so this study can be a reference in analyzing influential factors.

Keywords: Geographically weighted regression; Adaptive; Kernel; Bi-Square; Malnourished toddlers; East Java

1. Introduction

Regression is a measurement analysis with a finite dependent variable and one / more independent variables that form a relationship / function model [1]. Regression analysis has assumptions that must be met, one of which is heteroscedasticity. Heteroscedasticity is an error that is not constant. The heteroscedasticity problem results in inefficient variance and confidence intervals and other measures being incorrect [2]. Heteroscedasticity is obtained from various data including spatial data. Spatial data is data with geographic location information [3]. If heteroscedasticity is in spatial data which is a spatial effect of the relationship between observation locations, it is called spatial heterogeneity.

Spatial heterogeneity is defined as the diversity of systems in space and time. Spatial heterogeneity is indicated by differences in model parameters at each location [4]. The difference is due to the different characteristics in each location. Different environments and geographies between locations will affect the relationship between different events. Spatial data also has spatial dependence that describes the relationship of different parameter structures between regions [5].

Modeling spatial data cannot use linear regression due to the presence of unfulfilled assumptions, namely spatial heterogeneity. Therefore, an analysis is needed that can overcome these spatial effects. There are several spatial analysis modeling in connecting response variables and predictor variables that have spatial effects.

* Corresponding author: Shafa Salsabila

One of the modeling that can be done is using Geographically Weighted Regression (GWR) where the approach used is the point area [5]. GWR requires weighting which is an important part because it is used as a value for each location. The location assessed will have a different effect calculated from the close distance.

Weighting can be done by various methods. One of the weighting methods in the GWR model is the Bi-square Kernel function. The Bi-square kernel function uses coordinate points and distances that are continuous [6].

The GWR Adaptive Kernel Bi-square model will be applied to the prevalence of malnourished toddlers in East Java Province in 2018. Previous research stated that the prevalence of malnourished toddlers in East Java Province in 2018 contained spatial heterogeneity, so a GWR approach was needed in modeling factors affecting cases of malnourished toddlers in East Java Province [7].

Poor nutrition can be interpreted as a person's below-average nutritional condition [8]. Poor nutrition can cause a person's immunity to decrease and be at risk of being susceptible to disease so that the activities carried out are not perfect. The bad picture of an area is often seen from the fulfillment of the nutrition of its population.

Poor nutrition usually occurs in toddlers. Toddlers are a very rapid growth age so that nutritional needs in toddlers also increase [9]. Unmet needs during toddlerhood will affect the growth and development of toddlers and if left unchecked will have an impact on the future quality of toddlers and the quality of workers in the region. Poor nutrition at the age of toddlers is the worst form of malnutrition.

In East Java, malnutrition cases amounted to 3.3% with malnutrition cases in Indonesia amounting to 3.9%. This shows that cases in East Java are below the national percentage but the figure is still relatively high. This is the focus of the entire community in overcoming the problem, one of which is by analyzing the factors that cause malnutrition, especially in East Java.

The causes of malnutrition in each region can differ depending on the characteristics of the region [10]. District areas show higher cases of malnutrition compared to urban areas. These differences can cause data containing spatial heterogeneity, so this study will prove the Geographically Weighted Regression (GWR) Adaptive Kernel Bi-square model can analyze and interpret factors that have a significant effect on cases of malnutrition toddlers in East Java Province.

2. Material and methods

The data used in this study is secondary data on the Health Profile of East Java Province with 38 district/city observation units in 2018 [11]. The variables in this study used response variables and predictor variables. The response variable used, namely the prevalence of cases of malnourished toddlers in East Java Province (Y). While the predictor variables used, including the percentage of Low Birth Weight (BBLR) (X1), the percentage of infants exclusively breastfed (X2), the percentage of pregnant women receiving Fe tablets (X3), the percentage of toddler health services (X4), the percentage of complete basic immunization (X5), the percentage of toddlers receiving vitamin A (X6), the ratio of active posyandu (X7) and the percentage of the population with access to adequate drinking water (X8).

The analysis in this study uses GWR with the Adaptive fixed Kernel Bi-square weighting function so that data on the coordinate points of the distance of districts / cities is needed. The center of measurement of coordinate points is the office of the regent/mayor in each region. The software used is the Quantum Geographic Information System (QGIS) version 3.10 to create a thematic map of malnutrition toddler cases in East Java Province and the factors that influence it and Rstudio software version 4.0.2 to analyze the effect of predictor variables on response variables using the GWR method.

3. Results and discussion

3.1. Linearity

The relationship pattern of response variables and predictors can be assessed from the linearity test. The relationship pattern of response variables and predictor variables showed that the variables of BBLR percentage (X1), percentage of exclusive breastfeeding (X2), percentage of vitamin A (X6) and ratio of active posyandu (X7) had a positive relationship with cases of malnourished toddlers. The variables of immunization percentage (X5) and percentage of adequate drinking water (X8) have a negative relationship with cases of malnourished toddlers. The variables of percentage of Fe tablets (X3) and percentage of under-five health services (X4) tend to have a nonlinear relationship.

In addition, the pattern of relationships can be judged by the value of the correlation coefficient. The correlation coefficient can determine the relationship of the predictor variable to the response variable.

Table 1 The value of the correlation coefficient between variables

Variabel	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	0.270							
X_2	0.435	0.262						
X_3	-0.056	-0.153	0.108					
X_4	-0.014	0.121	0.226	0.554				
X_5	-0.150	-0.090	0.153	0.476	0.594			
X_6	0.133	0.212	0.295	0.335	0.286	0.393		
X_7	0.409	0.343	0.249	-0.003	0.241	0.135	0.403	
X_8	-0.170	-0.296	0.128	0.254	0.199	0.009	0.177	0.188

Table 1 shows variables X1 have a weak relationship with response variables, variables X3, X4, X5, X6 and X8, have a very weak relationship with response variables, and variables X2 and X7 have a sufficient relationship with response variables.

3.2. Assumption Test

There are several conditions in spatial regression modeling that must be met, namely multicollinearity, normality, and spatial heterogeneity (12). Correlation between variables has a big influence in spatial regression modeling because if there are interrelated variables, the influence of each variable is difficult to distinguish. Multicollinearity is defined as a condition where the absolute value of the response variable is smaller than the value of the predictor variable.

Multicollinearity can be determined from the value of the Variance Inflation Factor (VIF). A VIF value of less than 5 indicates no correlation between predictor variables. The data shows the smallest VIF value of 1.204, which is variable X2 and the largest VIF value is 2.219 as variable X4. This means that the data meets the requirements of multicollinearity because the data used does not contain multicollinearity, so modeling can continue using the GWR method.

The next requirement for spatial regression modeling is a normality test that can be assessed by the Kolmogorov-Smirnov test with Lilliefors correction. Based on the error normality assumption test, it is found that the p value is 0.3908 with α (0.05), so that H0 is accepted which means the error is normally distributed.

It is necessary to test the diversity of characteristics or spatial heterogeneity of each unit of location observation. Testing can use the Breusch-Pagan (BP) test. The results of the spatial heterogeneity test obtained a p value = 0.01008 using α (0.05) so that H0 was rejected which means that there is diversity in the data in each observed observation unit, meaning that each district / city has its own characteristics.

After the diversity test, a test is needed that can see the influence of observation units that are close to each other. The test that can be used is the Moran's I test. Moran's I test is obtained at 0.00415 with a p value of 0.2663 which means there is no influence between the intensity units to be tested

3.3. Case Modeling of Malnourished Toddlers with Geographically Weighted Regression

The GWR model is a model with parameter estimation carried out at each location of the observation point. The first step in analyzing the GWR is to calculate the euclidean distance. Then calculate the spatial weighting. Weighting is an important component whose value can represent data from the observation unit. There are several methods in weighting, one of which is Adaptive Kernel Bi-square weighting. The Adaptive Kernel Bi-square weighter will weight the value zero if the distance between locations is greater than or equal to the bandwidth. While distances that are less than bandwidth will be given a weight value close to one each close distance between locations.

The results were obtained with different bandwidth values for each region with a bandwidth value of 3.0772, namely Banyuwangi Regency and the smallest bandwidth value of 1.2927, namely Malang City. Furthermore, create a weighting matrix for the estimation of each location, so that the model owned by each district / city location in East Java Province is different.

Table 2 GWR Model Parameter Estimation Value with Bi-square Adaptive Kernel Weighting Function

Parameter	Minimum	Median	Maksimum
β_0 (Intercept)	-6.3406	0.5382	3.5919
β_1 (BBLR)	-0.4651	-0.0791	0.2303
β_2 (Exclusive breastfeeding)	0.0332	0.0949	0.1326
β_3 (Fe tablets)	-0.0673	-0.0233	0.1286
β_4 (Toddler Health Services)	-0.0271	0.0137	0.0584
β_5 (Immunization)	-0.1313	-0.0632	-0.0239
β_6 (Vitamin A)	- 0.0223	0.0166	0.0482
β_7 (Posyandu)	0.0761	0.2115	0.3266
β_8 (Drinking water)	-0.0777	-0.0418	0.0002

Based on Table 2, it is known that the minimum estimated value of β1 is -0.4651 and the maximum estimated value is 0.2303 which means that the estimated value of parameter β1 is between -0.4651 to 0.2303. Likewise for the next parameter. The next step is to test the significance of the parameter partially by comparing t count and t table to find out the predictor variables that have a significant effect in each district / city location. Furthermore, grouping is carried out based on the results of significant parameters as follows:

Table 3 Regional Grouping Based on Significant Predictor Variables in GWR Model with Bi-square Adaptive Kernel Weighting Function

Group	District/City	Significant Variables
1	Malang Regency, Probolinggo Regency, Pasuruan Regency, Sidoarjo Regency, Mojokerto Regency, Jombang Regency, Bojonegoro Regency, Tuban Regency, Lamongan Regency, Gresik Regency, Bangkalan Regency, Sampang Regency, Pamekasan Regency, Sumenep Regency, Kediri Regency, Kediri City, Malang City, Probolinggo City, Pasuruan City, City	No significant predictor variables
2	Ponorogo Regency, Tulungagung Regency, Blitar Regency, Nganjuk Regency, Blitar City	X_2
3	Pacitan Regency, Trenggalek Regency, Madiun Regency, Magetan Regency, Ngawi Regency, Madiun City.	X_2, X_7
4	Situbondo Regency	X_7, X_8
5	Lumajang Regency, Jember Regency, Banyuwangi Regency, Bondowoso Regency	X_2, X_7, X_8

Based on the results of the grouping table 3, 5 groups with significant variables are different in each group. The first group is a group where there are no significant influential variables, namely Malang Regency, Probolinggo Regency, Pasuruan Regency, Sidoarjo Regency, Mojokerto City, Jombang Regency, Bojonegoro Regency, Tuban Regency, Lamongan Regency, Gresik Regency, Bangkalan Regency, Sampang Regency, Pamekasan Regency, Sumenep Regency, Kediri City, Kediri City, Malang City, Probolinggo City, Pasuruan City, Mojokerto City, Surabaya City, and Batu City. The second group consists of districts/cities that have one significant variable, namely X2. Regencies / cities included in the second group, including Ponorogo Regency, Tulungagung Regency, Blitar Regency, Nganjuk Regency, and Blitar City. The third group has 2 significant variables, namely X2 and X7. Districts/cities with 2 significant variables, namely Pacitan Regency, Trenggalek Regency, Madiun Regency, Magetan Regency, Ngawi Regency, and Madiun City. The fourth group also has 2 significant variables X7 and X8, namely Situbondo Regency. The fifth group has significant variables X2, X7, X8, namely Lumajang Regency, Jember Regency, Banyuwangi Regency, and Bondowoso Regency.

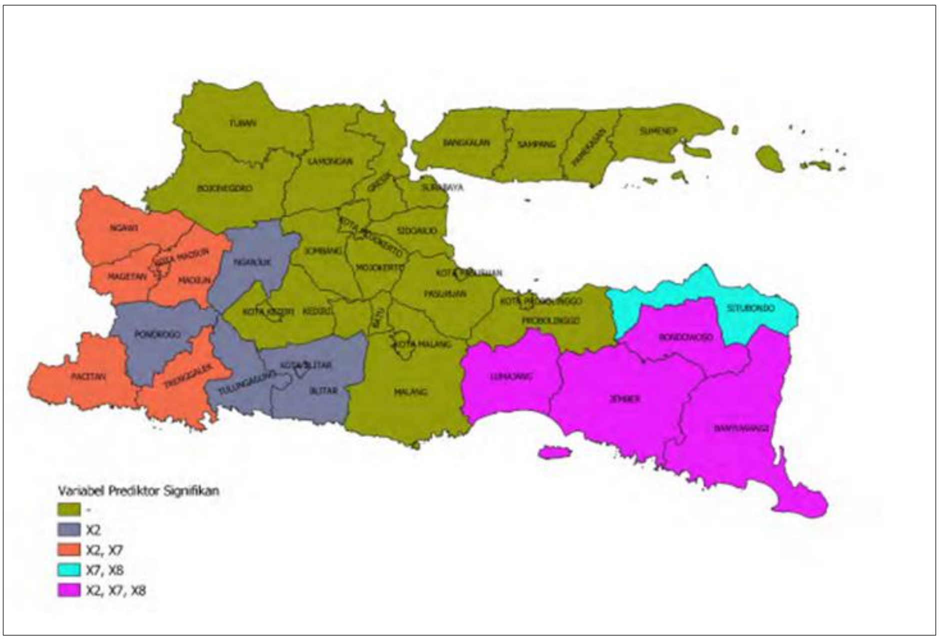

Figure 1 Map of Regional Groups Based on Significant Predictor Variables in GWR Model with Bi-square Adaptive Kernel Weighting Function

Table 4 GWR Modeling with Adaptive Kernel Bi-square Weighting Function in Three Districts/Cities in East Java Province

District/City	**β_2**	**$t_{count}\ \beta_2$**	**β_7**	**$T_{count}\ \beta_7$**	**β_8**	**$t_{count}\ \beta_8$**	**R^2**
Jember Regency	0.1227	2.4275	0.3017	2.3119	-0.0759	-2.4851	81.24%
Magetan Regency	0.1121	2.3244	0.3265	2.2279	-0.0443	-1.0921	47.97%
Mojokerto City	0.0469	0.8815	0.1202	0.8305	-0.0097	-0.2434	40.69%

Based on table 4, 3 districts/cities were obtained as examples of GWR modeling using Adaptive Kernel Bi-square weighting. The 3 districts / cities produced different coefficient values in each region and differences in predictor variables that had a significant effect. Jember Regency has a percentage of access to adequate drinking water (X8) which has a significant effect on cases of malnutrition toddlers (t count>t table), but in Magetan Regency and Mojokerto City does not have a significant effect.

The R2 value for Jember Regency is 81.24%, which means that the influence of all predictor variables on cases of malnutrition toddlers is 81.24%. The effect of variables X2 and X7 is positive, which means that if variables X2 and X7 are increasing, cases of malnourished toddlers in the region also increase. While the effect of variable X8 is negative, which means that if variable X8 increases, cases of malnourished toddlers in the region decrease.

Table 5 GWR Model with Bi-square Adaptive Kernel Weighting Function in 4 City Districts in East Java Province

Kab/Kota	**Model**	**R^2**
Blitar Regency	$Y = 0.1105\ X_2$	57.74%
Magetan Regency	$Y = 0.1121\ X_2 + 0.3265\ X_7$	47.97%
Situbondo Regency	$Y = 0.2776\ X_7 - 0.0686\ X_8$	79.77%
Jember Regency	$Y = 0.1227\ X_2 + 0.3017\ X_7 - 0.0759\ X_8$	81.24%

Based on the results of the analysis conducted in this study, Geographically Weighted Regression (GWR) modeling with the Adaptive Kernel Bi-square weighting function can analyze and interpret factors in cases of malnourished toddlers in East Java Province by grouping the intensity according to influential variables. We get 5 groupings based on influential predictor variables. Significant groupings of influential regions are influenced by geographical proximity to each other. The characteristics of closely spaced areas are almost similar and tend to be the same.

4. Conclusion

Based on the results of the analysis conducted in this study, Geographically Weighted Regression (GWR) modeling with the Adaptive Kernel Bi-square weighting function can analyze and interpret factors in cases of malnourished toddlers in East Java Province by grouping the intensity according to influential variables. We get 5 groupings based on influential predictor variables. Significant groupings of influential regions are influenced by geographical proximity to each other. The characteristics of closely spaced areas are almost similar and tend to be the same.

Compliance with ethical standards

Acknowledgments

The authors acknowledge the reviewer's positive suggestions on this paper.

Disclosure of conflict of interest

No potential conflict of interest was reported by the authors.

References

[1] Saidah, Yanuar F, Devianto D. Analisis Regresi Kuantil. J Mat UNAND [Internet]. 2016;5(1):103–7. Available from: 10.25077/JMU.5.1.103-107.2016

[2] R AS, Hadijati M, Switrayni NW. Analisis Masalah Heteroskedastisitas Menggunakan Generalized Least Square dalam Analisis Regresi. Eig Math J [Internet]. 2019;61–72. Available from: doi:10.29303/emj.v1i2.43

[3] Supuwiningsih nyoman N, Januhari Utami Nym Ni, Suniantara Putu Ketut I, Hanief Shofwan. Integritas Data Spasial dan Non Spasial Sistem Informasi Geografis. Media Sains Indonesia; 2022.

[4] Anselin L. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht; 1988.

[5] Taryono APN, Ispriyanti D, Prahutama A. Analisis Faktor-Faktor yang Mempengaruhi Penyebaran Penyakit Demam Berdarah Dengue (Dbd) di Provinsi Jawa Tengah dengan Metode Spatial Autoregressive Model dan Spatial Durbin Model. Indones J Appl Stat [Internet]. 2018;1(1):1. Available from: 10.13057/ijas.v1i1.24026

[6] Lutfiani N, Mariani S, Sugiman. Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-square. UNNES J Math [Internet]. 2017;5(1):82–91. Available from: 10.15294/UJM.V8I1.17103

[7] Novika Pratnyaningrum, Hasbi Yasin AH. Pemodelan persentase balita gizi buruk di Jawa Tengah dengan pendekatan Geographically Weighted Regression Principal Components Analysis (GWRPCA). 2015; Available from: http://eprints.undip.ac.id/47169/1/Novika_Pratnya.pdf

[8] Saputra W, Nurrizka RH. Demographic Factors and the Risk of Malnutrition and Nutrition for Less at Three Different Communities in West Sumatra. Makara J Heal Res [Internet]. 2013;16(2). Available from: 10.7454/msk.v16i2.1636

[9] Indrayani N, Khadijah S. Hubungan Karakteristik Ibu Dengan Tumbuh Kembang Balita Periode Emas Usia 12-60 Bulan. J Kebidanan Indones [Internet]. 2020;11(2):37. Available from: 10.36419/jkebin.v11i2.371

[10] Anggarini R, Logistik AMR. Pemodelan Faktor-Faktor Yang Berpengaruh Terhadap Prevalensi Balita Kurang Gizi Di Provinsi Jawa Timur Dengan Pendekatan Geographically Weighted Logistic Regression (GWLR). 2012;1(1). Available from: 10.12962/j23373520.v1i1.1960

[11] Dinkes Jatim. Profil Kesehatan Jawa Timur 2018. Dinas Kesehat Provinsi Jawa Timur. 2018;100.

[12] Dede M. Pengaruh Kondisi Lingkungan Terhadap Kerawanan Kejahatan Di Kawasan Perkotaan [Internet]. Vol. 151. Universitas Pendidikan Indonesia; 2018. p. 10–7. Available from: 10.31227/osf.io/ewg4j