IJSA (2022), Vol. 6, No. 2, 170-179
DOI: https://doi.org/10.29244/ijsa.v6i2p170-179

Indonesian Journal of Statistics and Its Applications
Available at https://ijsa.stats.id
e-ISSN: 2599-0802

Comparison of Negative Binomial Regression Model and Geographically Weighted Poisson Regression on Infant Mortality Rate in South Sulawesi Province*

Perbandingan Model Regresi Binomial Negatif dan *Geographically Weighted Poisson Regression* Pada Angka Kematian Bayi di Provinsi Sulawesi Selatan

Siswanto[1‡], Edy Saputra R[2], Nurtiti Sunusi[1] and Nirwan Ilyas[1]

[1]Department of Statistics, Hasanuddin University, Indonesia
[2]Department of Mathematics, Hasanuddin University, Indonesia
[‡]corresponding author: siswanto@unhas.ac.id

Abstract

The number of infant mortality cases is an important indicator to assess the quality of a country's public health. A number of studies argue that the case of infant mortality has a close relation to the living area condition and the social status of the parents. Indirectly, the quality of life of babies in a country will impact the nation's quality of life in general. Therefore, many efforts are required to reduce the infant mortality in Indonesia. One of the steps that could be done to overcome this issue is to analyze the causative factors. The statistical method that has been developed for data analysis taking into account current spatial factors is the Geographically Weighted Poisson Regression (GWPR) with a weighted Bisquare kernel function. Based on the partial estimation with the GWPR model, there are seven groups based on significant variables that affect the number of infant deaths in South Sulawesi Province. Of the seven groups formed, the first group is the Selayar Islands where all variables have a significant effect. This needs to be a concern for the South Sulawesi provincial government to improve facilities and infrastructure in the Selayar Islands, of course the location which is very far from the city center can affect access to drug reception, medical personnel and so on. Based on the results of the analysis of the factors that affect the number of infant deaths in South Sulawesi Province using a negative binomial regression approach and GWPR with a bisquare kernel weighting, it can be concluded that the GWPR model used is the best for analyzing the number of infant deaths in South Sulawesi Province because it has an AIC value. The smallest is 167.668.

Keywords: geographically weighted poisson regression, mortality, negative binomial regression, spatial.

* Received: Jul 2021; Reviewed: Apr 2022; Published: Aug 2022

1. Introduction

Infant mortality rate is the chance of a baby dying before the age of one year, which is stated by per 1000 live births (LB) (Kemenkes RI, 2019). The number of cases of infant death is closely related to the environmental conditions of residence and the social status of the baby's parents. The quality of life of children will indirectly be the determining factor of the quality of life of the nation in the future. Therefore, various efforts are needed to reduce or reduce infant mortality in Indonesia. Infant mortality in 2012 was 32/1000 LB with neonatal deaths of 19/1000. There was a decrease in 2017 to 24/1000 LB and 15/1000 LB neonatal data based on the report of the Indonesian Demographic and Health Survey. So, it can be concluded that infant mortality in Indonesia has decreased (Lengkong *et al*., 2020).

South Sulawesi is one of the provinces in Indonesia that managed to reduce the infant mortality rate (Dinkes Sulsel, 2019). In 2017 the number of infant deaths in South Sulawesi was 1059 cases; then in 2018 it dropped to 1037 cases, and by the end of 2019 the number of infant deaths again dropped to 916 cases. The decrease in the number of cases shows success in efforts to reduce or suppress the number of infant deaths (Dinkes Sulsel, 2019).

One way to reduce or suppress infant mortality is to know the factors that contribute to infant mortality. The relationship between the factors that cause the number of newborn deaths can be done using a statistical approach that involves the relationship between bound variables and free variables using Poisson regression model. The problem that arises in Poisson regression modeling is the assumption of equidisperception that must be met, that is, the value of the dependent variable has an average equal to the variety. The fundamental fact is that it is very difficult to find similar conditions, so the assumption is often violated, namely overdispersion and underdispersion (Wang & Famoye, 1997). If there is excessive dispersion in the data, then Poisson regression is less accurately used for analysis, since the impact on the standard deviation presumption value of the regression coefficient is too weak. Negative binomial regression is one way to overcome the problem of underdispersion /overdispersion (Cameron & Trivedi, 2013).

Poisson regression is not enough to describe the local conditions of each region (Elyna *et al*., 2012). Therefore, statistical modeling is needed that takes into account regional or spatial factors (Nakaya *et al*., 2005). Social, economic, cultural and geographical conditions must differ from region to region. Therefore, it stands to reason that the dependent variable observed in this case, which is the number of infant mortality rates, will certainly vary between regions. There is a diversity in the mortality of infants from one region to another, so a statistical method was developed to analyze the data by considering spatial factors; namely Geographically Weighted Poisson Regression (GWPR), by using the bisquare kernel function weighted. With this method will be obtained significant variables for each region of South Sulawesi Province. The dependent variables studied are discrete random variables that distribute Poisson with regard to spatial factors, then the relationship between dependent variables and independent variables can be confirmed by that method, against factors that have a significant influence on spatial factors infant mortality in each regency/city of South Sulawesi Province.

2. Research Metodology

2.1 Data

This research uses secondary data obtained from the 2018 South Sulawesi Provincial Health Office Profile, which contains 2017 data (Dinkes, 2018). The observation used in this study is the Cities and Regencies in South Sulawesi Province, which includes 3 Cities and 21 Regencies. The dependent variables and the independent variables used are found in Table 1.

Table 1: Dependent and independent variables used

Variable	Variable Name
Y	Number of babies who died
X_1	Number of babies born with low body weight
X_2	Number of babies receiving vitamin A supplementation
X_3	The number of babies given exclusive breast milk
X_4	Number of babies fully immunized
X_5	Number of mothers giving birth with the help of medical personnel

2.2 Analysis Methods

1. Collect data from various available sources.
2. Performs Poisson distributed data assumption tests. To find out the bound variables distributed Poisson, use the Kolmogorov-Smirnov test.
3. Establish the Poisson regression model in determining infant mortality in South Sulawesi Province with the following steps:
   a. Perform modelling in dependent variable with independent variable
   The model equation for Poisson regression is
   $g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$, dengan $i = 1,2,\dots,n$
   By $g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$ being a connecting function or link function (Agresti, 2018).
   b. Testing overdispersed assumptions
   Overdispersion testing can be done using deviance values, obtained from the division of Pearson Chi-Square values with the value of degrees of freedom. If the value $\phi > 1$ then indicated overdispersion (Ulfa *et al*., 2020).
4. Establish a negative binomial regression with the following equation,
   $g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$, dengan $i = 1,2,\dots,n$
   By $g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$ being a connecting function or link function. The presumption of parameters in the model can use the maximum likelihood method, to obtain the approximate value of its regression parameters using Newton Raphson iteration method (Irawati, 2012).
5. Perform GWPR modelling in determining the number of infant deaths with the following stages:
   a. Determine euclidean distances between observations locations as a function of geographical location,
   b. Determine the optimum bandwidth based on the smallest CV value,
   c. Use the selected kernel function to create a spatial weighted,

   d. Testing the similarity of the GWPR model,
   e. Testing the GWPR model fit,
   f. Testing the parameters of the GWPR model partial,
6. Grouping based on the same significant variables in the GWPR model as the map.
7. Select the best model with the smallest *Akaike Information Criterion* (AIC) value from the negative binomial regression model with the GWPR model

3. Results and Discussion

3.1 Data Exploration

Based on Table 2, the average number of infant deaths in South Sulawesi Province is 44.13 in each Regency/City. Bantaeng Regency is the regency with the lowest number of infant deaths with 1 person in 2017. Jeneponto Regency had the highest number of infant deaths, with 91 deaths in the same year.

Table 2: Descriptive statistics of infant mortality in South Sulawesi Province

Variable	Average	Standard Deviation	Minimum	Maximum
Y	44.13	23.29	1	91
X1	225.25	154.42	46	846
X2	3867.75	4617.47	167	24094
X3	2830.96	3690.35	268	15960
X4	5889.17	5344.77	0	26715
X5	6231.75	5023.57	1884	25946

3.2 Poisson Distribution Match Test and Poisson Regression Model

The Kolmogorov-Smirnov test is used to check Poisson's dependent variables with the hypothesis:

H_0 : Number of newborn deaths in 2017 distributed Poisson

H_1 : Number of newborn deaths in 2017 did not distribute Poisson

The statistical test value obtained is $D_{count} = 0.308 < D_{table} = 0.323$ which means not to refuse H_0. That is, the variable number of newborn deaths in South Sulawesi Province is distributed Poisson so that it can be analyzed with Poisson regression.

The formation of global regression is the best regression model among the number of newborn deaths and influential factors. The initial models of Poisson regression are as follows:

$$\eta_i = 3.240 + 0.002011X_1 - 0.0001855X_2 + 0.00007798X_3 - 0.00002184X_4 + 0.0001084X_5$$

After obtaining the approximate value of the Poisson regression model parameter, it is continued by testing the parameters simultaneously and partially. Processing result obtained a deviation value generated by 210 with a significant level of 0.01 so that the value $\chi^2_{(0.01;5)}$ of 15,086. Thus, the value of deviance = 210 $> \chi^2_{(0.01;5)} = 15.086$, which means it can be concluded that there is at least one independent variable that affects the dependent variable.

Table 3: The alleged value of Poisson regression model parameters

Parameters	Estimated Value	Default Error	Z_{count}	P_{value}
β_0	3.240	7.636×10^{-2}	42.427	$< 2 \times 10^{-16}$*
β_1	2.011×10^{-3}	4.621×10^{-4}	4.352	1.350×10^{-5}*
β_2	-1.855×10^{-4}	2.085×10^{-5}	-8.897	$< 2 \times 10^{-16}$*
β_3	7.798×10^{-5}	1.700×10^{-5}	4.586	4.52×10^{-6}*
β_4	-2.184×10^{-5}	2.255×10^{-5}	-0.968	0.333
β_5	1.084×10^{-4}	2.750×10^{-5}	3.942	8.090×10^{-5}*

Description: *Significant at a real level of 1%

Based on Table 3, there are four variables that have a significant influence on the number of newborn deaths in South Sulawesi Province, namely X_1, X_2, X_3 and X_5. The interpretation of the Poisson regression model uses the values of the odd ratio for each coefficient which is spelled out as follows:

a. $\hat{\beta}_1 = 2.011 \times 10^{-3}$, meaning that for every addition of one baby born with a low body weight, the average case of newborn death will 0.201% rise assuming other variables remain.

b. $\hat{\beta}_2 = -1.855 \times 10^{-4}$, meaning that for every addition of one baby who gets vitamin A supplementation, the average of newborn deaths will 0.018% drop assuming other variables remain.

c. $\hat{\beta}_3 = 7.798 \times 10^{-5}$, meaning that for every addition of one baby who gets vitamin A supplementation, the average number of newborn deaths will 0.018% drop assuming other variables remain.

d. $\hat{\beta}_5 = 1.084 \times 10^{-4}$, meaning that for every addition of one mother who gives birth with the help of medical personnel, the average case of newborn death will 0.010% rise assuming other variables remain.

3.3 Overdispersion Test

To find out if a regression model is overdispersed or not is done under the assumption of equidispersion. Overdispersion on the data can be known from the value ϕ, that is, the value of deviance divided by degrees of freedom. Based on Table 4, the test statistical value means the $\phi = 11.667 > 1$ which means the dependent variable is overdispersed.

Table 4: Overdispersion Test

Deviance	Degrees of Freedom	ϕ
210	18	11.667

3.4 Negative Binomial Regression Model

Based on Table 4, it was obtained that the Poisson regression model was overdispersed. Then, the next step is to handle by forming a negative binomial regression model, with the result of estimating parameters, namely,

$$\eta_i = 3.140 + 0.002775X_1 - 0.0002046X_2 + 0.00009759X_3 - 0.00003007X_4 + 0.0001065X_5$$

To check the match of the negative binomial regression model is used deviance test, as for the test stages as follows:

a) Hypothesis
$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$
H_1: At least one of the $\beta_j \neq 0, j = 1, 2\ 3$
b) The significance level used $\alpha = 1\%$
c) Statistics test
$D = 2[L(y_i; y_i) - L(\mu_i; y_i)]$
$$D = 2\sum_{i=1}^{n}\left\{y_i \ln\left(\frac{y_i}{\mu_i}\right) - \left(\frac{1}{k} + y_i\right)\ln\left(\frac{1 + ky_i}{1 + k\mu_i}\right)\right\} = 27.280$$
d) Test criteria
Reject H_0 if test statistics $D > \chi^2_{\alpha;p}$
e) Decision
Reject H_0 because of the value $D = 27.280 >> \chi^2_{0.01;5} = 15.086$
f) Conclusion
Repulsion H_0, which means negative binomial regression model can be used as a model.

Table 5: The alleged value of the negative binomial regression model parameter

Parameters	Estimated Value	Default Error	Z_{count}	P_{value}
β_0	3.140	2.409×10^{-1}	13.033	$< 2 \times 10^{-16}$*
β_1	2.775×10^{-3}	1.633×10^{-3}	1.699	0.089
β_2	-2.046×10^{-4}	7.952×10^{-5}	−2.574	0.01*
β_3	9.759×10^{-5}	6.821×10^{-5}	1.431	0.152
β_4	-3.007×10^{-5}	7.460×10^{-5}	−0.403	0.687
β_5	1.065×10^{-4}	9.285×10^{-5}	1.147	0.251

Description: *Significant at a real level of 1%

Based on Table 5, only one variable has a significant influence on the number of newborn deaths in South Sulawesi Province, namely X_2. With the alleged value $\hat{\beta}_2 = -2.046 \times 10^{-4}$, meaning for each additional one baby who gets vitamin A supplementation, then the average case of newborn death will go down 0.020% with other variables assumed to remain.

3.5 Geographically Weighted Poisson Regression (GWPR)

GWPR is a development model of Poisson regression that pays attention to weighted location. In building the GWPR model, the first step that was done was to determine the geographical location of each regency/city observed. Furthermore, is to choose an optimal bandwidth. The value of optimal bandwidth obtained is that 2.801, with each regency/city location has a different bandwidth value. The next step is to get the weighted matrix using the kernel function used in the GWPR model usage process.

Available kernel functions include Gaussian kernel and bisquare. This study used the bisquare kernel because this weighting process adjusts based on proximity between sample points (Caraka, 2017). After searching for bandwidth values and weighted functions from each location, the next step is to estimate the parameters of the GWPR model. Hypothesis testing is performed by testing the similarity of negative binomial regression models and GWPR to test the significance of geographic factors.

Table 6: Model similarity test

Model	Deviance	Degrees of Freedom (Df)	Deviance/Db	F_{count}
Negative Binomial Regression	27.280	18	1.515	0.185
GWPR	89.070	10.865	8.198	

Based on the value F_{count} of Table 6, it can be concluded $0.185 < F_{0.01;18;11} = 0.241$, then $F_{count} < F_{table}$, so it can be concluded not to reject H_0, meaning that there is no difference between the negative binomial regression model and the GWPR model.

3.6 GWPR Model Parameter Partial Test

To find out the independent variables that affect the number of newborn deaths at each location, a partial test was conducted and compares the values with t_{count} and $t_{table} = 3.197$. Reject H_0 if $t_{count} > 3.197$, then it can be concluded that the parameters to- j at location i have a significant effect on the model. Table 7 shows regencies/cities in South Sulawesi Province grouped into seven groups according to significant variable similarities.

Table 7: Grouping by significant variables

Regency/City	Significant Variables
Selayar Islands	X_1, X_2, X_3, X_4, X_5
Bantaeng, Bulukumba, Gowa, Makassar, Maros Sinjai, Takalar	X_1, X_2, X_5
Jeneponto	X_1, X_2, X_3, X_5
Barru, Bone, Pangkep, Soppeng, Wajo	X_2, X_5
Sidrap, Pinrang, Enrekang, Pare-Pare	X_2
Luwu, Tana Toraja, North Toraja, Palopo	X_3
North Luwu, East Luwu	X_3, X_5

Table 7 shows that there are seven groups formed, the first consisting of Selayar Islands Regency with significant variables for infant mortality are all variables, namely the number of babies born with low body weight (X_1), the number of babies who get vitamin A supplementation (X_2), the number of babies who are breastfed exclusively (X_3), the baby is fully immunized (X_4) and the mother who gives birth with the help of medical personnel (X_5). This needs to be a concern for the Provincial Government of South Sulawesi for the improvement of facilities and infrastructure in the Selayar Islands, of course, a location very far from the city center can affect access to drug admissions, medical personnel and so on.

Regencies/cities in the second group are Bantaeng, Bulukumba, Gowa, Makassar, Maros Sinjai, Takalar with significant variables, namely the number of babies born with low body weight (X_1), the number of babies who get supplementation of vitamin A (X_2) and the number of mothers who give birth with the help of medical personnel(X_5). Jeneponto Regency is an area that belongs to group three with a significant total of variables there are four, namely the number of babies born with low body weight(X_1), the number of babies who get vitamin A supplementation(X_2), the number of babies who are exclusively breastfed (X_3) and the number of mothers who give birth with the help of medical personnel(X_5). The regency is also one of the regencies that entered the disadvantaged area, the South Sulawesi provincial government must also pay more attention in terms of health facilities and infrastructure.

Regencies included in group four are Pangkep, Barru, Bone, Soppeng, Wajo with significant variables are the number of babies who get vitamin A supplementation (X_2) and the number of mothers who give birth with the help of medical personnel (X_5). Regencies/cities that are included in the group of five, namely Sidrap, Pinrang, Enrekang, Pare-Pare with significant variables are the number of babies who get vitamin A supplementation, of course this is very interesting because the number of infant deaths is influenced by vitamin A consumption, it could be due to the behavior of people who are less educated about vitamins that must be consumed or the habits of people who are afraid to consume drugs because it will affect the condition of the baby in the womb.

Regencies/cities that are included in the group of six are Luwu, Tana Toraja, North Toraja, Palopo with significant variables about babies who are exclusively breastfed (X_3). The last group of seven, namely North Luwu Regency, East Luwu, with significant variables, is the number of babies who are exclusively breastfed (X_3) and the number of mothers who give birth with the help of medical personnel (X_5).

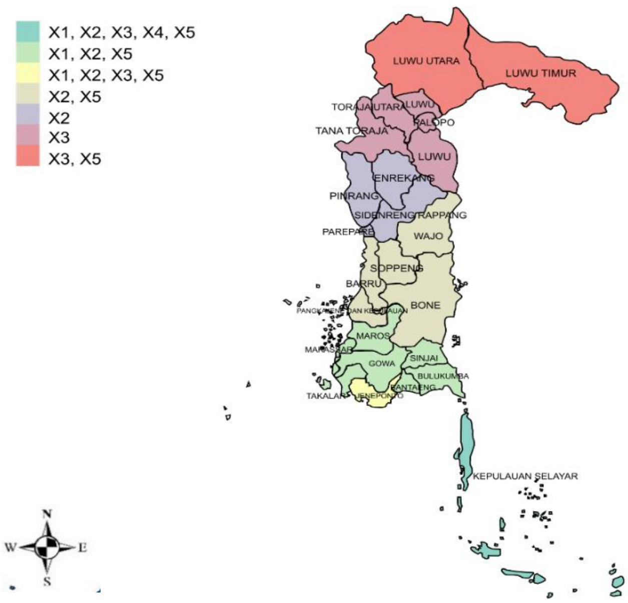

Figure 1: Mapping regency/city groups based on significant variables

Based on Figure 1, a significant variable occurs in an adjacent location. The interesting thing about the picture is that the areas that include the disadvantaged areas, namely Selayar Islands and Jeneponto Regencies, each form into one group with the most number of significant variables. Of course, this can be an input to the

South Sulawesi Provincial Government to pay attention to areas that are left behind in terms of health facilities and infrastructure.

3.7 Best Model Selection

Criteria for choosing the best model using the AIC value. A model is said to be good if it has a small AIC value. Based on Table 8, comparing negative binomial regression models and GWPR, it was obtained that the GWPR model is best at analyzing the number of newborn deaths in South Sulawesi Province. This is because the AIC value of the GWPR model is smaller at 167,668 compared to the AIC value of the negative binomial regression model of 226,730.

Table 8: Comparison of model conformity

Model	AIC
Negative Binomial Regression	226.730
GWPR	167.668

4. Conclusion

Model GWPR is divided into 7 regency/city groups based on the same significant variables. Group 1 is Selayar Islands with all significant variables. Group 2 is Bulukumba, Bantaeng, Takalar, Gowa, Sinjai, Maros, and Makassar with significant variables X_1, X_2, X_5. Group 3 is Jeneponto with significant variables X_1, X_2, X_3, X_5. Group 4 is Pangkep, Barru, Bone, Soppeng, and Wajo with significant variables X_2, X_5. Group 5 is Sidrap, Pinrang, Enrekang, and Pare-Pare with significant variables X_2. Group 6 is Luwu, Tana Toraja, North Toraja, and Palopo with significant variables X_3. Group 7 is Northern Luwu, and Eastern Luwu with significant variables X_3, X_5. The GWPR Model used is the best to analyze the number of newborn deaths in South Sulawesi Province because it has the smallest AIC value of 167.668.

Acknowledgement. The author expressed his gratitude to Hasanuddin University for conducting this research through the Beginner Lecturer Research Internal Grant for the 2021 fiscal year with contract 915/UN4.22/PT.01.03/2021.

Bibliography

Agresti, A. (2018). *An Introduction to Categorical Data Analysis.* New York: John Wiley & Sons.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). New York: Cambridge university press.

Caraka, R. E., & Yasin, H. (2017). *Geographically Weighted Regression (GWR) Sebuah Pendekatan Regresi Geografis.* Yogyakarta: Mobius.

[Dinkes] Dinas Kesehatan Provinsi Sulawesi Selatan. (2018). *Profil Kesehatan Provinsi Sulawesi Selatan Tahun 2017*. Makassar: Pusat Data dan Informasi.

[Dinkes] Dinas Kesehatan Provinsi Sulawesi Selatan. (2019). *Laporan Kinerja (LKJ) Organisasi Perangkat Daerah Tahun 2019.* Makassar: Dinkes Sulsel.

Elyna, M. A. K., Srinadi, I. G. A. M., & Susilawati, M. (2012). Pemodelan Angka Kematian Bayi dengan Pendekatan Geographically Weighted Poisson Regression di Provinsi Bali. *E-Jurnal Matematika*, 1(1): 94-98.

Irawati, B. (2012). Perbandingan Analisis Generalized Poisson Regression (GPR) dan Regresi Binomial Negatif untuk Mengatasi Overdispersi Studi Kasus: Pemodelan Jumlah Kasus Kanker Serviks di Jawa Timur. *Jurnal Matematika*, 2(2): 13-24.

[Kemenkes RI] Kementerian Kesehatan Republik Indonesia. (2019). *Profil Kesehatan Indonesia Tahun 2018.* Jakarta: Kementerian Kesehatan RI.

Lengkong, G. T., Langi, F. L., & Posangi, J. (2020). Faktor-Faktor yang Berhubungan dengan Kematian Bayi di Indonesia. *Kesmas*, 9(4): 41-47.

Ulfa, Y. A., Soleh, A. M., & Sartono, B. (2020). Handling of Overdispersion in the Poisson Regression Model with Negative Binomial for the Number of New Cases of Leprosy in Java: Penanganan Overdispersi pada Model Regresi Poisson dengan Binomial Negatif untuk Jumlah Kasus Baru Kusta di Jawa. *Indonesian Journal of Statistics and Its Applications,* 5(1): 1-13.

Nakaya, T., Fotheringham, A. S., Brudson, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in medicine*, 24(17): 2695-2717.

Wang, W., & Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression. *Journal of population economics*, 10(3), 273-283.