Proceeding of International Conference on Green Technology
Vol.8, No.1, October 2017, pp. 385-389
p-ISSN: 2580-7080 – e-ISSN: 2580-7099

ESTIMATION PARAMETER GEOGRAPHICALLY WEIGHTED ZERO INFLATED POISSON REGRESSION (GWZIPR) WITH FIXED BISQUARE KERNEL

Adeliana[1], Sri Harini[2]
[1]UIN Maulana Malik Ibrahim Malang

Article Info

Article history:

Received Jul 12th, 2017
Revised Aug 20th, 2017
Accepted Oct 26th, 2017

Keyword:

Geographically Weighted Zero Inflated Poisson Regression, *Maximum Likelihood Estimation*, Tetanus.

ABSTRACT

Poisson regression is a form of regression analysis used to model data in the form of count (number). Poisson regression method requires the existence of equidispersi that is the condition where the mean and variance of the response variable is equal. But sometimes there is an overdispersion phenomenon in the data modeled with the poisson distribution. Overdispersion means that data has a variance greater than the mean. Overdispersion shows that there is a population heterogeneity. Consequently, the estimation of the parameters on the data under such conditions becomes imprecise. One method to overcome the overdispersion is the Zero Inflated Poisson regression. Then the development of the ZIP regression that has taken into account the spatial factor is called Geographically Weighted Zero Inflated Poisson Regression (GWZIPR). The parameter estimation of the GWZIPR model is carried out by the Maximum Likelihood Estimation (MLE) method and completed using the Expectation-Maximization (EM) algorithm. Weighting function used is fixed bisquare kernel. This study describes the factors that influence the disease Tetanus Neonatorum in all districts / cities in East Java Province.

Corresponding Author:

Adeliana
UIN Maulana Malik Ibrahim Malang
Email: elliana063@gmail.com

1. INTRODUCTION

Regression analysis is a popular statistical method for assessing the relationship between response variables (Y) with predictor variables (X). There are several kinds of regression analysis. Linear regression analysis is a regression analysis used to model the relationship between response variables and predictors of scale and interval ratio and normal distribution, while poisson regression is one regression analysis that can describe the relationship between response variables (Y) where the response variable is distributed poisson with predictor variable (X). Poisson regression model is a standard model for discrete data and includes models in linear regression models (Cameron dan Trivedi, 1998). Poisson model is widely used in various fields including public health, epidemiology, sociology, psychology, engineering, agriculture, and others.

Jansakul dan Hinde (2001) dalam Andres (2011) states that one of the causes of overdispersion is that more observations are zero. In poisson regression, the number of zero values in the observation result will exceed the predicted value (inflation). To overcome this, many methods are developed. One method for analyzing observations with zero values is more with the Zero-Inflated Poisson Regression model.

Poisson regression is a form of regression analysis used to model data in the form of count (amount). Poisson regression assumes that random variables are distributed poisson.

Poisson regression model according to Myers (1990) is $y_i = \mu_i + \varepsilon_i$. Poisson regression model is a Generalized Linear Model with connecting function g and $g = \sum_{k=0}^{p} x_{ik}\beta_k$. The average poisson is a non-negative value so from it used the log function as a liaison as follows:

$$g = \ln(\mu_i) = \sum_{k=0}^{p} x_{ik}\beta_k$$

So the poisson regression model is

$$\ln(\mu_i) = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ip}\beta_p$$

$$\mu_i = \exp(\boldsymbol{x_i\beta}) \qquad (1)$$

and:

i : $1,2,\ldots,n$; $n =$ the number of observation locations

k : $1,2,\ldots,p$; $p =$ the number of independent variables

y_i : dependent variable on the location of the observation to-i

μ_i : the average dependent variable at the location of observations to-*i* which is expectations of Y and $Y \sim \text{poisson}(\mu)$

x_i : vector independent variable at the location of observation to -i

β : vector of poisson regression parameters of size $(k+1)\times 1$

(Myers, 1990)

2.2 *Zero Inflated Poisson* Regression

One of the causes of overdispersion is that more observations are zero than those estimated for the poisson regression model. One of the proposed methods of analysis for more zero-point observations than the estimates is the Zero-Inflated Poisson (ZIP) regression model (Jansakul dan Hinde, 2001).

For each observation Y_i which are mutually free $i = 1, 2, \ldots, n$ and

$$Y_i \sim \begin{cases} 0 & \text{with probability } \omega_i \\ \text{poisson}(\mu_{i,}) & \text{with probability } (1-\omega_i) \end{cases} \qquad (2)$$

then:

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\mu_i}, y_i = 0 \\ (1-\omega_i)\dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, y_i = 1,2,\ldots; 0 \le \omega_i \le 1 \end{cases} \qquad (3)$$

Lambert (1992) suggest a combined model for μ and ω is:

$$\ln(\mu_i) = \boldsymbol{X_i^T\beta} \text{ and } \text{logit}(\omega) = \ln\left(\frac{\omega_i}{1-\omega_i}\right) = \boldsymbol{X_i^T\gamma} \qquad (4)$$

$\boldsymbol{X_i^T}$ is the predictor variable matrix, β and γ is the parameter vector to be assessed, and ω is the probability of zero-valued observations.

(Lambert, 1992)

2.3 Tetanus Neonatorum (TN)

Tetanus Neonatorum (TN) is a disease caused by Clostridium Tetani in infants (age <28 days) that can cause death. Usually the toxin is produced by the vegetative form of the organism at the site of the occurrence of further injury transported and fixed in the central nervous system. Tetanus is usually acute and causes spastic paralytic caused by tetanospasmin. The spores of Clostridium Tetani germs enter through the sole entrance to the newborn's umbilical body. The event may occur at the time of the baby's umbilical cord when the baby is born. Handling Tetanus Neonatorum is not easy, so the most important is prevention efforts through the delivery of hygienic delivery and immunization Tetanus Toxoid (TT) pregnant women and cord care.

(Dinas Kesehatan, 2014)

2. RESEARCH METHOD

Data Sources and Research Variables

The data used in the study is secondary data from East Java Provincial Health Office on the number of cases of Tetanus Neonatorum disease in East Java Year 2014. This study also uses geographic coordinates of coordinates of latitude and longitude in each district / city in East Java. There are 38 districts / cities in East Java consisting of 29 districts and 9 cities. Dependent variable in this research is the number of cases of Tetanus Neonatorum disease in East Java Year 2014 (y) and independent variables which include: Percentage of immunization coverage of TT2+ to number of pregnant women (x_1), percentage of maternity assisted by health personnel (x_2), Percentage of visit coverage complete neonatal to infant number (x_3), percentage of neonatal complication handling to the number of pregnant women (x_4).

Step Analysis

1. Conducting descriptive data analysis as a preliminary description to know the state of the case of Tetanus Neonatorum disease in East Java Year 2014.Melakukan pengujian asumsi data.
2. Analyze data by using GWZIPR model estimation with fixed bisquare kernel fixing function using SPSS, GeoDa, and GWR4 software.
3. Create a thematic map of the number of cases of Tetanus Neonatorum disease in East Java and the factors that influence it in East Java Province based on estimation results using ArcGis software.
4. Conclusions.

3. RESULTS AND DISCUSSION

Modeling GWZIPR with fixed bisquare kernel

GWZIPR modeling for each district / city is obtained from partial testing. The test statistic used is the test statistic W with a decline decision H_0 if $W > Z_{0.025}$. Value $Z_{0.025} = 1.96$ then all predictor variables are significant in fixed bisquare kernel fixers. GWZIPR model with fixed bisquare kernel weighting on the number of cases of Tetanus Neonatorum disease in Kediri are:

$$\log\left(\hat{\mu}_{30}\right) = -23.53096742 - 0.00835287X_1 - 0.293322729X_2 + 0.512848748X_3 - 0.0126754X_4$$

and

$$\text{logit}\left(\hat{\omega}_{30}\right) = -51.94824996 + 0.221429638X_1 - 3.869690283X_2 + 4.541672316X_3 - 0.526613051X_4$$

The significant predictor variables in Kediri on log models are $X_1, X_2, X_3, X_4,$ and X_5. The log model explains that any 1% change in the percentage of complete neonatal visit coverage to the number of infants increases the average log of Tetanus Neonatorum cases by 0.512848748. This is not as expected, because it is variable X_3 which is expected to reduce the number of cases can actually increase the number of cases, although relatively small. Any change of 1% percentage of TT2 + immunization coverage to the number of pregnant women decreases the average log of Tetanus Neonatorum case 0.00835287.

Logit model explains that there is no chance of Tetanus Neonatorum case in every regency / city in East Java $\left(y_i = 0\right)$ influenced by four predictor variable that is percentage of immunization coverage of TT2 + to number of pregnant women $\left(X_1\right)$, the percentage of delivery mothers is assisted by health personnel $\left(X_2\right)$, percentage coverage of complete neonatal visit to infant number $\left(X_3\right)$, percentage of neonatal complications handling the

Proceeding of International Conference on Green Technology
Vol.8, No.1, October 2017, pp. 385-389
p-ISSN: 2580-7080 – e-ISSN: 2580-7099

number of pregnant women (X_4). Logit model that has the meaning of the risk of non-occurrence of cases of Tetanus Neonatorum disease in Kediri City 0.221429638 if the coverage of TT2 + immunization against the number of pregnant women in Kediri and neighboring districts increased by 1%. Likewise with the understanding of other variables that the maternal maternal variable helped by health personnel (X_2), coverage of complete neonatal visits to the number of infants (X_3), and treatment of neonatal complications on the number of pregnant women (X_4).

The result of the GWZIPR model with fixed bisquare kernel weighting in Kediri City still has a variable that the result is not in accordance with the theory. Based on the overall significant variables ie complete coverage of neonatal visit factors incompatible with the theory.

One of the factors that can reduce the number of cases of Tetanus Neonatorum disease is the complete coverage of neonatal visits. The coverage of neonatal visits was the coverage of neonates who received standardized service at least three times ie 1 time at 6-48 hours, 1 time on day 3 - 7th day and 1 time on day 8 - day 28 after birth in a work area for a certain period of time. Neonatal visits aim to improve neonatal access to basic health care, knowing as early as possible if there are abnormalities in the baby or having problems. The results of this study were not found in accordance with expectations, due to the GWZIPR model that is the coverage factor of neonatal visits to log the average number of cases of Tetanus Neonatorum disease.

Based on the Health Profile of East Java Province in 2014, coverage of complete neonatal visit coverage has not reached the target. Efforts to improve coverage coverage of this indicator are facilitated both in terms of management of maternal and child health programs (KIA) as well as recording and reporting, improving clinical skills of field officers and involving multi-stakeholders in the implementation of the program. Districts / cities that have not reached the target are expected to perform quality neonatal services by starting mapping as well as startup of pregnant women and perform quality AnteNatal Care (ANC) services.

Image Grouping on Log Models

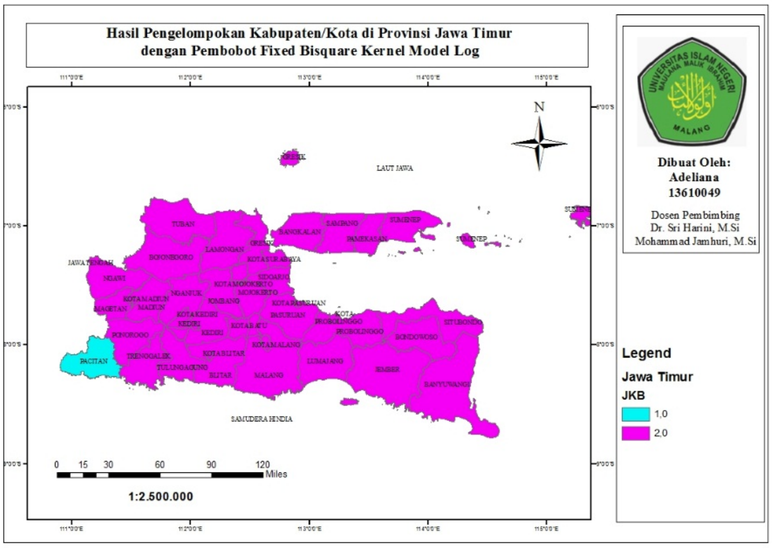
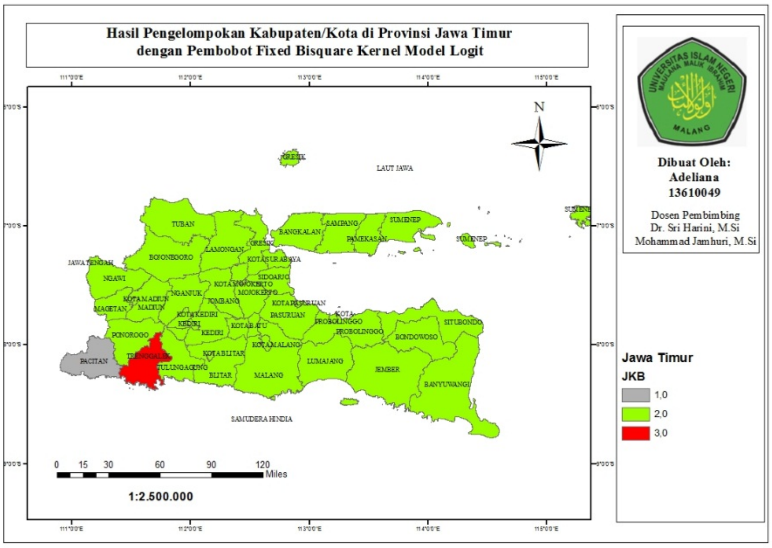


Image Grouping on Logit Models



CONCLUSION

Geographically Weighted Zero Inflated Poisson Regression (GWZIPR) model on the number of cases of Tetanus Neonatorum disease is

$$\log\left(\hat{\mu}_{30}\right) = -23.53096742 - 0.00835287X_1 - 0.293322729X_2 + 0.512848748X_3 - 0.0126754X_4$$

and

$$\text{logit}\left(\hat{\omega}_{30}\right) = -51.94824996 + 0.221429638X_1 - 3.869690283X_2 + 4.541672316X_3 - 0.526613051X_4$$

This research there are still some problems that there are variables that the results are less in accordance with the theory. Suggestions for further research should add other variables that are still related to the number of cases of Tetanus Neonatorum disease.

REFERENCES

Cameron, A C dan Trivedi, P K. 1998. *Regression Analysis of Count Data*. Cambridge University Press: Cambridge.

Dinas Kesehatan Provinsi Jawa Timur. 2014. *Profil Kesehatan Jawa Timur Tahun 2014*. Jawa Timur: Dinas Kesehatan Provinsi Jawa Timur.

Jansakul, N. Dan Hinde, J.P., 2001. *Score Test for Zero-Inflated Poisson Models*. Computational Statistics & Data Analysis 40,2002:75-96.

Lambert, D. 1992. *Zero Inflated Poisson Regression, With An Application To Detect In Manufacturing*. Journal Techno metrics, Feb 1992 Vol 32 no 1.

Myers,R.H. 1990. *Classical and Modern Regression with Application*. Boston: PWS-KENT Publishing Company