Eksakta : Berkala Ilmiah Bidang MIPA
VOLUME 22 NO 03 2021, pp 200-210
ISSN : Print 1411-3724 — Online 2549-7464
DOI : https://doi.org/10.24036//eksakta/vol22-iss2/272

Eksakta
Berkala Ilmiah Bidang MIPA
http://www.eksakta.ppj.unp.ac.id/index.php/eksakta

Article

Infant Mortality Case: An Application of Negative Binomial Regression in order to Overcome Overdispersion in Poisson Regression

Fadhilah Fitri[1]*, Fitri Mudia Sari[1], Nurul Fiskia Gamayanti[2], Iut Tri Utami[3]

[1]Department of Statistics, Faculty of Mathematics and Natural Science (FMIPA), Universitas Negeri Padang, Indonesia
[2]Department of Mathematics, Faculty of Mathematics and Natural Science (FMIPA), Universitas Tadulako, Indonesia
[3]Department of Statistics, Faculty of Mathematics and Natural Science (FMIPA), Universitas Diponegoro, Indonesia

Article Info

Article history :

Received May 5, 2021
Revised September 10, 2021
Accepted September 10, 2021
Published September 30, 2021

Keywords :

Overdispersion, poisson regression, negative binomials

Abstract. Infant mortality is an indicator to determine the degree of public health. Infant mortality is death that occurs in the period from birth to before the age of one. The high rate of infant mortality indicates that the quality of public health services is not optimal. The number of infant deaths is an example of count data that follows a Poisson distribution, so it can be analyzed using Poisson Regression. The assumption that must be met when using this method is the equidispersion or variance of the response variable is equal to mean. However, this condition rarely occurs because usually the counted data has a greater variance than the mean or it is called overdispersion. One way to solve this problem is to use the Negative Binomial Regression method. The data used in this study is the case of infant mortality in the city of Padang. First, we model the data using Poisson Regression, then we check the assumption, if there is overdispersion, we handle it by modeling the data with Negative Binomial Regression. The results showed that the equidispersion assumption could not be met so that the data was modeled with Negative Binomial Regression. The model obtained is $\mu_i = \exp(-6{,}1384 + 0{,}2489X_1 - 0{,}0055X_2 + 0{,}0301X_3 + 0{,}1038X_4)$ $+0{,}0554X_5 - 0{,}0005X_6)$. Based on AIC we can also conclude that Negative Binomial Regression is the best method to model this data.

Corresponding Author :
Fadhilah Fitri
Department of Statistics, Faculty of Mathematics and Natural Science (FMIPA),
Universitas Negeri Padang, Indonesia
Email : fadhilahfitri@fmipa.unp.ac.id

1. Introduction

Health is one of the areas that the government focuses on in order to improve the welfare of society. Various attempts were made to improve the standard of living therefore a high degree of public health was achieved. Infant mortality is one indicator to determine the degree of public health and quality of life of a population[1], [2], [3]. The cause of infant mortality not only due to medical aspect but also non-medical aspect such as access to healthcare including antenatal checkups, low economy level, human error, etc [4], [5], [6]. The Infant Mortality Rate (IMR) for West Sumatra Province based on the 2010 Population Census is higher than the national IMR. The IMR for West Sumatra in 2006 was 30 per 1000 live births, while the IMR for Indonesia was 26 per 1000 live births [7]. Padang as the provincial capital contributed the highest infant mortality rate in 2016 compared to other cities and districts in West Sumatra. The following is a graph of infant mortality cases in Padang in 2012 – 2017 [8].

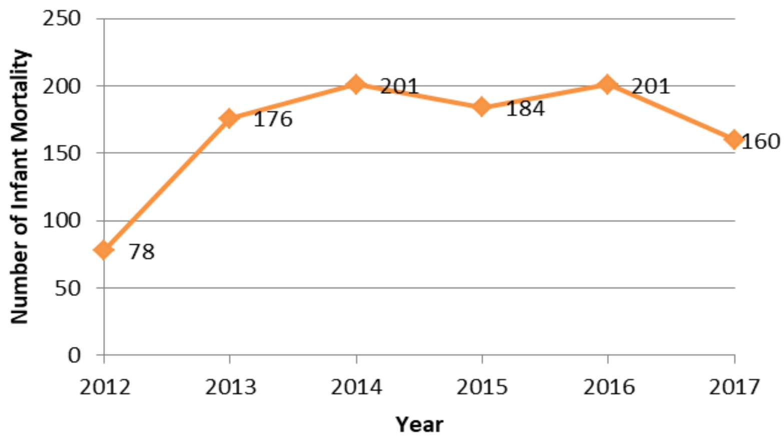

Fig 1. Number of Infant Mortality in Padang

In Figure 1, it can be seen that infant mortality in Padang has fluctuated. More effort is needed to reduce cases. Therefore, it is important to know what factors can cause death in infants and how much these factors contribute. One method that can be used is regression analysis. Regression will analyze the relationship between the response variables, both continuous and discrete, for example in the form of binary, nominal, ordinal, truncated or chopped data[9], [10].

Count data is part of the discrete response variable. This data is non-negative and states the number of events in an interval of time or space. When an event rarely occurs in a large sample space, it follows a Poisson distribution [11]. The number of infant deaths is an example of a count data that follows a Poisson distribution. Poisson regression is derived from the Poisson distribution, hence to analyze this data Poisson regression is used.

In Poisson regression analysis, there are assumptions that must be fulfilled. The assumption is that the variance of the response variable is the same as the mean [12]. In fact, this condition is very rare because usually the count data has a greater variance than the mean or is called overdispersion [11], [12]. This condition will result in inefficient parameter estimation. One way to solve this problem is to use the Negative Binomial Regression method. The Negative Binomial Regression is derived from the Negative Binomial distribution. Unlike Poisson distribution, it has additional parameters so that the variance can exceed the mean. As a result, the Negative Binomial Regression method can overcome the overdispersion problem in Poisson Regression. As for the multicollinearity assumption, it does not need to be checked because Negative Binomal Regression, which is part of the Generalized Linear Model, does not use Least Square but Generalized Inverse therefore it does not require a non-singular (**X'X**) matrix [13]. There have been several studies conducted using negative binomial in order to overcome the overdispersion in several different object such as utilization of antenatal care [14], cyberbullying [15], health behavior [16], actuarial pricing [17], etc. Meanwhile another alternatives for modeling infant mortality are Poisson Regression Approach Based on Local Linear Estimator [18], Panel Data Analysis [5], [19], Spatial Statistics Approach [20] and so on.

The purpose of this study was to model data on infant mortality cases, therefore this research is an applied research. First, we model the data using Poisson Regression because the data is count data [11]. Then we check the assumption, if there is overdispersion, we handle it by modeling the data with Negative Binomial Regression [12].

Poisson Regression

Poisson regression is a nonlinear regression to model random events where the probability of occurrence is relatively small in a certain time interval or at a certain place. The response variable in Poisson regression is discrete data that follows a Poisson distribution which is an exponential family, hence this regression is part of the Generalized Linear Model (GLM). GLM has three components: a random component, a systematic component and a link function [21]. The probability function for the Poisson distribution is:

$$f(y_i;\mu) = \frac{e^{-\mu}\mu^{y_i}}{y_i!} \quad \text{for } y_i = 0,1,2,\dots \text{ and } \mu > 0 \tag{1}$$

$$= \exp(-\mu + y_i \ln\mu - \ln y_i!) \tag{2}$$

Based on Equation (2), the link function is $\ln\mu$, then the relationship between the mean response variable and the linear combination of the predictor variables is:

$$\ln\mu_i = \mathbf{x}_i^T\boldsymbol{\beta} \text{ or } \mu_i = e^{\mathbf{x}_i^T\boldsymbol{\beta}} \tag{3}$$

Equation (3) is a Poisson regression model. The $\boldsymbol{\beta}$ parameter in Poisson regression was estimated using the maximum likelihood method where implicit and nonlinear equations were generated. The $\boldsymbol{\beta}$ parameter estimate is obtained by maximizing the function using the iterative method. The numerical iteration methods that can be used are Newton-Raphson or Fisher Scoring.

Poisson regression inherits from the Poisson distribution that the mean and variance are equal, otherwise known as equidispersion. Therefore this assumption must be fulfilled. If the variance value is less than the means value, it is called underdispersion, and when the variance value is greater than the mean, an overdispersion condition occurs. One solution to dealing with this overdipersion is Negative Binomial.

Negative Binomial Distribution

The Negative Binomial Distribution is a distribution that has many approaches. There are several ways to approach the Negative Binomial distribution, including that it can be approached as a Bernoulli experimental sequence and the Poisson-Gamma mixture distribution [12]. The classic approach of the Negative Binomial distribution that is often used is the Negative Binomial distribution as a Bernoulli experiment sequence: the number of Bernoulli trials required to obtain *r* successes, where each repetition is independent, and the probability of success in each experiment is constant is p, while the probability of failure is 1 – p. Suppose that the random variable X states the number of experiments needed to get *r* successful, then X has a Negative Binomial distribution with the probability function [22] as follow:

$$\Pr(X = x; r, p) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r}, & x = r, r+1, r+2, \dots \\ 0 & , x \text{ other} \end{cases} \tag{4}$$

The probability function of the random variable X can be denoted in other terms. Suppose there are a number of y failures before the *r* th success, then *x* is the sum of *y* failures plus *r* successes or *x* = *y* + *r*. Thus, a new random variable Y will be formed, which states the number of failures before *r* success occurs with the variable transformation method where the transformation function is Y = X - *r*. Then the random variable Y has a probability function as follows:

$$\Pr(Y = y; r, p) = \begin{cases} \binom{y + r\text{-}1}{y} p^r (1\text{-}p)^y, & y = 0,1,2,\dots \\ 0 & , y \text{ other} \end{cases} \tag{5}$$

The moment generation function for the negative binomial distribution is:

$$M_y(t) = p^r(1 - qe^t)^{-r} \tag{6}$$

The negative binomial distribution has the following mean and variance:

$$E(Y) = \frac{r(1-p)}{p} \tag{7}$$

$$\mathrm{Var}(Y) = \frac{r(1\text{-}p)}{p} \tag{8}$$

The Negative Binomial Distribution formed from Poisson-Gamma [11], [12], [23] has the following probability function:

$$\Pr(Y = y) = \frac{\Gamma\left(y+\frac{1}{\alpha}\right)}{y!\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{1}{1+\alpha\mu}\right)^{\frac{1}{\alpha}}\left(1\text{-}\frac{1}{1+\alpha\mu}\right)^{y} \tag{9}$$

$$= \begin{pmatrix} y + \frac{1}{\alpha} - 1 \\ \frac{1}{\alpha} - 1 \end{pmatrix}\left(\frac{1}{1+\alpha\mu}\right)^{\frac{1}{\alpha}}\left(1 - \frac{1}{1+\alpha\mu}\right)^{y}$$

Equation (9) above has a similar form to the probability function form of the negative binomial distribution in equation (5) where $r = \frac{1}{\alpha}$ dan $p = \frac{1}{(\alpha\mu+1)}$.

The mean and variance of the negative binomial distribution are:

$$E(Y) = \mu \tag{10}$$

$$Var(Y) = \mu + \alpha\mu^2 \tag{11}$$

Negative Binomial Regression

The Negative Binomial Regression is derived from the Negative Binomial distribution. Unlike the Poisson Distribution, this distribution has an additional parameter hence the variance can exceed the mean [21]. Suppose we want to know the relationship between a response variable Y and *k* explanatory variables $X_1, X_2, \dots, X_k$. The regression model uses the relationship between the response variable Y and the explanatory variables $X_1, X_2, \dots, X_k$ as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \;\; ; \; i = 1,2,\dots,n \tag{12}$$

where $\beta_0, \beta_1, \dots, \beta_k$ represents the number of unknown parameters and ε_i states the error for the-*i*th observation and assumes that the expected value of ε_i is $(E(\varepsilon_i) = 0)$. If equation (12) above is expressed in vector form, it becomes:

$$y_i = \mathbf{X_i}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i \tag{13}$$

$$\text{where } \mathbf{X_i}^{\mathrm{T}} = [1, x_{1i}, \dots, x_{ki}] \text{ dan } \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

Suppose that it is assumed that the expected value for Y_i is $E(Y_i|X_{1i} = x_{1i}, X_{2i} = x_{2i}, \dots, X_{ki} = x_{ki}) = \mu_i$ and previously it has been assumed that the expected value for ε_i is 0, then it will be obtained:

$$\mu_i = E(Y_i|X_{1i} = x_{1i}, X_{2i} = x_{2i}, \dots, X_{ki} = x_{ki})$$

$$= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}$$

or if expressed in vector form will be

$$\mu_i = X_i^T\beta \quad (14)$$

In a negative binomial model, Y_i is a variable in the form of counted data therefore Y_i s a non-negative integer, so the expected value of Y_i cannot be negative either. Based on equation (14), this is contradictory because the value space for $\boldsymbol{X_i}^T\boldsymbol{\beta}$ is the real number in the interval of $(-\infty, \infty)$. This makes the regression model unable to be used to analyze the counted data. To overcome this conflicting situation, a link function is used that connects the fitted value (μ_i) with the linear predictor $\boldsymbol{X_i}^T\boldsymbol{\beta}$. As a member of the exponential family, the negative binomial has the canonical link function, $\left(\frac{\theta\mu}{1+\theta\mu}\right) = \boldsymbol{X_i}^T\boldsymbol{\beta}$, with invers $\mu = \frac{1}{\theta[\exp(-X_i^T\beta)-1]}$. The inverse form shows that the link function produces a fairly complicated shape. As a result, the interpretation of the regression model parameters will be more difficult. The negative binomial model generally uses a logarithmic link function:

$$ln\, \mu_i = \boldsymbol{X_i}^T\boldsymbol{\beta} \quad (15)$$

The Negative Binomial Model can use log links because $ln\, \mu_i$ and $\boldsymbol{X_i}^T\boldsymbol{\beta}$ will be defined in the $(0, \infty)$ interval and interpretation of the regression parameters will become easier. After obtaining the correct link function, it can be stated that the binomial regression model is negative for modeling the count data:

$$\ln[E(Y_i|X_1)] = \ln(\mu_i) = X_i^T\beta \text{ , ; } i=1,2,\dots,n \text{ ,} \quad (16)$$

Then it can be obtained:

$$\mu_i = \exp(\boldsymbol{X_i}^T\boldsymbol{\beta}) \quad (17)$$

2. Experimental Sections

The data used in this study are secondary data obtained from the publication of the Dinas Kesehatan Kota Padang "Profil Kesehatan Tahun 2019" [24] which contains information about the general description of the city of Padang, the situation of health status, the situation of health efforts, and the situation of health resources in 2019. The research variables used were the number of infant mortality cases (Y), the percentage of babies with low birth weight (X_1), the percentage of babies who were exclusively breastfed (X_2), the percentage of pregnant women who received blood booster tablets (X_3), the percentage of deliveries assisted by non-medical personnel (X_4), the percentage of infants who received complete basic immunization (X_5), and the percentage of infants who received vitamin A (X_6). Each line of observation is a subdistrict in Padang. The number of observations of subdistrict in Padang is 11 observations.

The steps in modeling infant mortality cases at the Public Health Center Padang are: first, we conduct descriptive analysis of the response variables and explanatory variables. Descriptive analysis is an important step in conducting statistical analysis. Descriptive analysis can stand on its own as a research product, such as when it identifies phenomena or patterns in data that have not previously been recognized [25]. Then, we plotted the correlations between the variables. We can refer to this step as exploratory analysis.

The next step is modeling with the Poisson Regression method. Model parameters will be estimated and AIC will be calculated so that a Poisson Regression model will be formed. After that we perform overdispersion testing. If the mean is equal to the variance, we can use the Poisson Regression model, but if this assumption is violated, we will handle overdispersion using Negative Binomial Regression.

When we decide to use Negative Binomial Regression, we estimate the model parameters first. Afterwards, perform regression parameter testing simultaneously and partially. Then interpret the resulting model. We can compare the AIC value of the Poisson regression model with the AIC of the negative binomial regression model, the smaller the better. The last step, making conclusions and suggestions. The complete steps given in the flowchart below

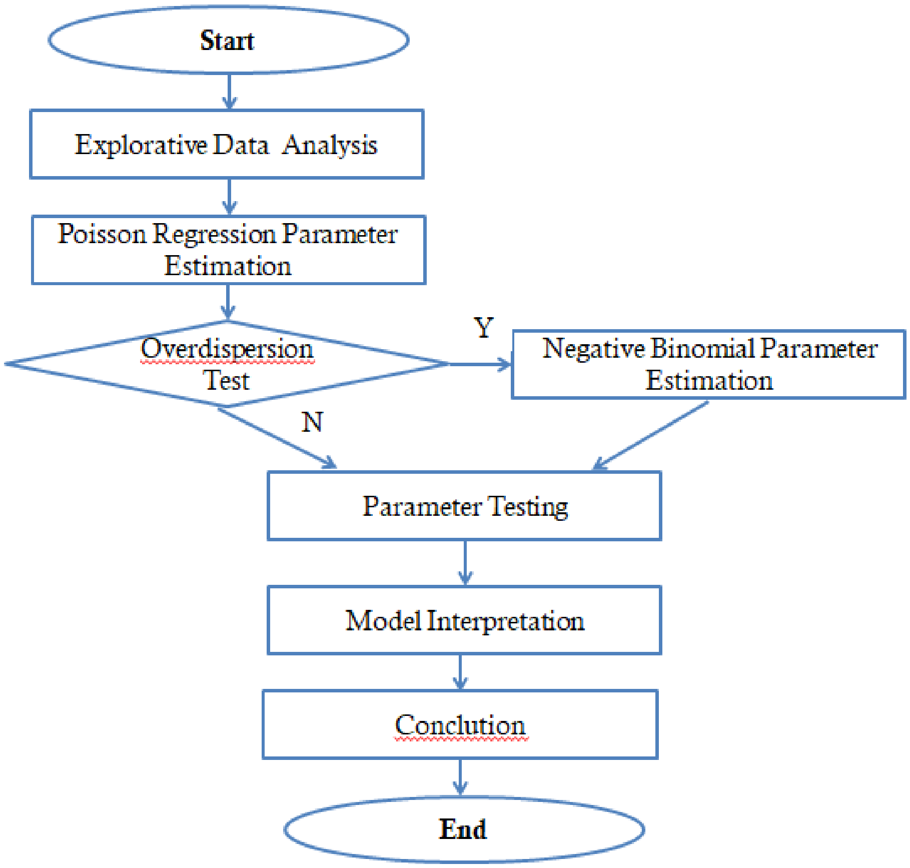

Fig 2. The Flowchart

3. Results and Discussion

3.1 Descriptive Analysis

As initial information to find out the characteristics and patterns of data that will be used for further analysis, it is necessary to look at the statistical descriptions for each of the variables used in this study. This descriptive analysis was carried out to see the characteristics of the data to be processed. The descriptions of the variables used in this study are presented in Table 1.From Table 1, it can be seen that the distribution form of each variable by looking at the comparison of the mean value and the second quartile. If the mean value is greater than the value of the second quartile, the spread will skew to the right. Meanwhile, if the mean value is smaller than the value of the second quartile, the distribution will skew to the left. And if the mean is equal to the second quartile then the data is spread normally.

Table 1. Descriptive Statistics of the Variables Used in the Study

Variable	Minimum	Quartile 1	Quartile 2	Quartile 3	Maximum	Mean
Y	2.00	5.50	8.00	10.50	17.00	8.04
X1	0.00	1.06	1.46	2.33	3.75	1.73
X2	52.02	72.06	83.33	87.60	100.00	79.65
X3	67.09	88.07	95.29	98.22	100.58	91.50
X4	1.36	6.14	8.84	16.47	36.71	12.45
X5	65.84	89.06	89.78	93.35	97.10	89.73
X6	41.93	57.42	84.03	90.36	96.04	75.26

Based on Table 1, it can be seen that several variables have skew distribution to the right, because the mean value is greater than the value of the second quartile, they are the variables Y, X_1, dan X_4. The X_2, X_3, X_5, and X_6 ariables have a distribution that skew to the left. This also shows that there are several health centers where the percentage of babies is exclusively breastfed (X_2), the percentage of pregnant women who receive blood booster tablets (X_3), the percentage of babies who receive complete basic immunization (X_5), and the percentage of babies who receive vitamin A (X_6) less than other *Puskesmas* in Padang. Therefore, the attention of the City Health Office is needed to make it equal.

Next, Figure 2 will show a correlation plot between variables. The color of blue shows a positive correlation between variables, while the red one shows a negative correlation. The smaller the correlation, the smaller the circle size and the faded the color will be. From Figure 2 it can be seen that some of the independent variables used in the study have a high correlation, so it can be concluded that there is multicollinearity between the independent variables used in the study. However this will not be a problem in Poisson Regression analysis.

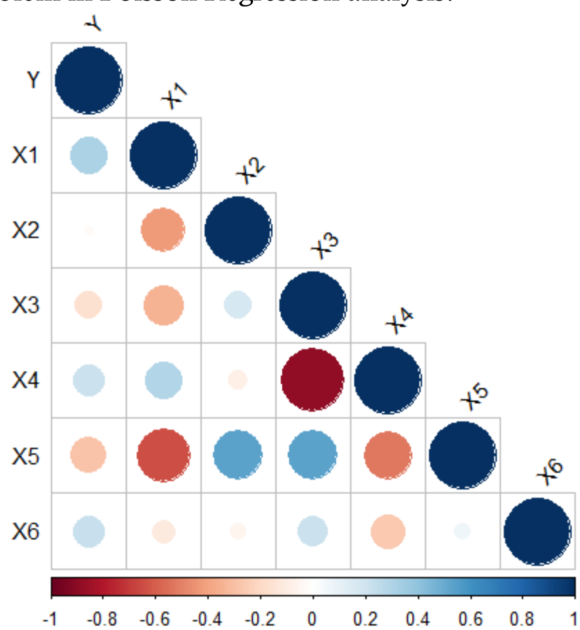

Fig 3. Correlation Between Variables Used in the Study

3.2 Poisson Regression Analysis

The following is an estimate of the Poisson regression model parameters.

Table 2. Poisson Regression Model Parameter Estimation

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-8.116	6.277	-1.293	0.196
X1	0.276	0.197	1.406	0.159
X2	-0.001	0.009	-0.115	0.909
X3	0.054	0.046	1.181	0.238

X4	0.121	0.048	2.506	0.012
X5	0.048	0.033	1.456	0.145
X6	-0.003	0.004	-0.675	0.499
Deviance = 33,640			DF = 4	
AIC = 97,212				

The Poisson regression model for the number of infant deaths using the six explanatory variables that have been selected shows that: at the 5% significance level, the variable that has a significant effect on infant mortality is the percentage of deliveries assisted by non-medical personnel (X_4). Five other variables still have an effect on infant mortality, but the effect of these five variables is not too significant. The Poisson regression model that is formed is:

$$\mu_i = \exp(-8{,}1164 + 0{,}2762X_1 - 0{,}0011X_2 + 0{,}0541X_3 + 0{,}1212X_4 + 0{,}0482X_5 - 0{,}0030X_6)$$

The next step is the overdispersion test. Overdispersion occurs when the variance value is greater than the mean, overdispersion can also be seen from the ratio between the devian residue and the degrees of freedom [26], [27]. The ratio value between the remaining deviance and the degrees of freedom is 8.41, this value is greater than 1. This indicates that the Poisson regression model is overdispersed, so the Poisson Regression model is not appropriate to use in modeling infant mortality in Padang. One method that can be used to overcome overdispersion cases in Poisson Regression is the Negative Binomial Regression Model.

3.3 Negative Binomial Regression Analysis

The initial step in negative binomial regression modeling is to determine the initial theta value. Based on the trial-error results, an initial theta of 7.9294 was obtained. Therefore, a negative binomial regression modeling was carried out with an initial theta of 7.9294. Table 3 shows the parameter estimates of the Negative Binomial Regression model.

Table 3. Negative Binomial Regression Model Parameter Estimation

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-6.1384	11.6395	-0.527	0.598
X1	0.2489	0.3307	0.753	0.452
X2	-0.0055	0.0168	-0.325	0.745
X3	0.0301	0.0825	0.365	0.715
X4	0.1038	0.0887	1.171	0.242
X5	0.0554	0.0563	0.984	0.325
X6	-0.0005	0.0080	-0.066	0.948
Deviance = 10,575			DF = 4	
AIC = 87,997				

The Negative Binomial Regression Model on the number of infant deaths using the six explanatory variables that have been selected shows that there are no explanatory variables that have a significant effect on infant mortality at the 5% significance level. The choice of this level of significance depends on the needs of the researcher. For example, research conducted by Pratama and Wulandari (2015) used a significance level of 20% [28]. Then the research conducted by Hajarisman (1998) who chose to use a significance level of 30% [29]. The size of this significance level will affect the confidence interval. The greater the selected significance level, the narrower the confidence interval will be. In this study, the significance level used was 25%. As the result, the variable that had a significant effect on infant mortality was the percentage of deliveries assisted by non-medical personnel (X_4). Five other variables are still included in the model, this is done because there is the possibility that a weak variable when analyzing a single variable, becomes an important variable if it is done simultaneously with other variables [29]. The Negative Binomial regression model that is formed is:

$$\mu_i = \exp(-6{,}1384 + 0{,}2489X_1 - 0{,}0055X_2 + 0{,}0301X_3 + 0{,}1038X_4 + 0{,}0554X_5 - 0{,}0005X_6)$$

Based on the variables that have a significant effect on infant mortality in Padang, it can be concluded that for each additional 1 percent of births assisted by non-medical personnel, it will increase the number of cases of infant mortality by $\exp(0{,}1038) = 1{,}1094 \approx 1$ cases, assuming other variables are constant.

4. Conclusion

The data on infant mortality in Padang cannot be modeled using Poisson Regression because of the violation of the overdispersion assumption. Therefore, it is handled by using Negative Binomial Regression which produces the following model:

$$\mu_i = \exp(-6{,}1384 + 0{,}2489X_1 - 0{,}0055X_2 + 0{,}0301X_3 + 0{,}1038X_4 + 0{,}0554X_5 - 0{,}0005X_6)$$

Based on the variables that have a significant effect on infant mortality in Padang, it can be concluded that for each additional 1 percent of births assisted by non-medical personnel, it will increase the number of cases of infant mortality by $\exp(0{,}1038) = 1{,}1094 \approx 1$ cases, assuming other variables are constant. Based on AIC we can also conclude that Negative Binomial Regression is the best method to model thus data. AIC of Negative Binomial Regression is 87,997 meanwhile AIC of Poisson Regression is 97,212

The suggestion based on this research is that the government should reduce the percentage of births assisted by non-medical personnel by, for example, by providing information to the community to reduce this because of the risks that may occur, etc.

References

[1] Sartorius, B. K., & Sartorius, K. (2014). Global infant mortality trends and attributable determinants–an ecological study using data from 192 countries for the period 1990–2011. *Population Health Metrics*, *12*(1), 1-15.

[2] Reidpath, D. D., & Allotey, P. (2003). Infant mortality rate as an indicator of population health. *Journal of Epidemiology & Community Health*, *57*(5), 344-346.

[3] Simeoni, S., Frova, L., & De Curtis, M. (2019). Inequalities in infant mortality in Italy. *Italian journal of pediatrics*, *45*(1), 1-7.

[4] Gonzalez, R. M., & Gilleskie, D. (2017). Infant mortality rate as a measure of a country's health: a robust method to improve reliability and comparability. *Demography*, *54*(2), 701-720.

[5] Kiross, G. T., Chojenta, C., Barker, D., & Loxton, D. (2020). The effects of health expenditure on infant mortality in sub-Saharan Africa: evidence from panel data analysis. *Health economics review*, *10*(1), 1-9.

[6] Vijay, J., & Patel, K. K. (2020). Risk factors of infant mortality in Bangladesh. *Clinical Epidemiology and Global Health*, *8*(1), 211-214.

[7] Rodriguez, A., Furquim, F., & DesJardins, S. L. (2018). Categorical and limited dependent variable modeling in higher education. In *Higher education: Handbook of theory and research* (pp. 295-370). Springer, Cham.

[8] Smith, E. K., Lacy, M. G., & Mayer, A. (2019). Performance simulations for categorical mediation: Analyzing khb estimates of mediation in ordinal regression models. *The Stata Journal*, *19*(4), 913-930.

[9] Azen, R., & Walker, C. M. (2021). *Categorical data analysis for the behavioral and social sciences*. Routledge.

[10] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression*

analysis. John Wiley & Sons.
[11] Hayes, A. F., & Montoya, A. K. (2017). A tutorial on testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures*, *11*(1), 1-30.
[12] Iqbal, W., Tang, Y. M., Chau, K. Y., Irfan, M., & Mohsin, M. (2021). Nexus between air pollution and NCOV-2019 in China: application of negative binomial regression analysis. *Process Safety and Environmental Protection*, *150*, 557-565.
[13] Ardiles, L. G., Tadano, Y. S., Costa, S., Urbina, V., Capucim, M. N., da Silva, I., ... & Martins, L. D. (2018). Negative Binomial regression model for analysis of the relationship between hospitalization and air pollution. *Atmospheric Pollution Research*, *9*(2), 333-341.
[14] Islam, M. A., Kabir, M. R., & Talukder, A. (2020). Triggering factors associated with the utilization of antenatal care visits in Bangladesh: An application of negative binomial regression model. *Clinical Epidemiology and Global Health*, *8*(4), 1297-1301.
[15] Cho, S., Lee, H., Peguero, A. A., & Park, S. M. (2019). Social-ecological correlates of cyberbullying victimization and perpetration among African American youth: Negative binomial and zero-inflated negative binomial analyses. *Children and Youth Services Review*, *101*, 50-60.
[16] Green, J. A. (2021). Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychology and Behavioral Medicine*, *9*(1), 436-455.
[17] Lee, S. C. (2020). Delta boosting implementation of negative binomial regression in actuarial pricing. *Risks*, *8*(1), 19.
[18] Utoyo, M. I., & Chamidah, N. (2019, March). Modeling of Maternal Mortality and Infant Mortality Cases in East Kalimantan using Poisson Regression Approach Based on Local Linear Estimator. In *IOP Conference Series: Earth and Environmental Science* (Vol. 243, No. 1, p. 012023). IOP Publishing.1315/243/1/012023.
[19] Dutta, U. P., Gupta, H., Sarkar, A. K., & Sengupta, P. P. (2020). Some determinants of infant mortality rate in SAARC countries: an empirical assessment through panel data analysis. *Child Indicators Research*, *13*, 2093-2116.
[20] Singh, M. P., Bharti, A., Singh, N. K., & Singh, R. D. (2018). Spatial Scan Study for Mortality Under Age 5 year in the EAG States and Assam.
[21] Hancock, J. T., & Khoshgoftaar, T. M. (2020). Survey on categorical data for neural networks. *Journal of Big Data*, *7*(1), 1-41.
[22] Kabir, R. H., & Lee, K. (2020, July). Receding-horizon ergodic exploration planning using optimal transport theory. In *2020 American Control Conference (ACC)* (pp. 1447-1452). IEEE.
[23] Braccini, M., Denham, A., O'Neill, M. F., & Lai, E. (2021). Spatial and temporal patterns in catch rates from multispecies shark fisheries in Western Australia. *Ocean & Coastal Management*, *213*, 105883.
[24] Jestel, C., Surmann, H., Stenzel, J., Urbann, O., & Brehler, M. (2021, February). Obtaining Robust Control and Navigation Policies for Multi-robot Navigation via Deep Reinforcement Learning. In *2021 7th International Conference on Automation, Robotics and Applications (ICARA)* (pp. 48-54). IEEE.
[25] Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive Analysis in Education: A Guide for Researchers. NCEE 2017-4023. *National Center for Education Evaluation and Regional Assistance*.
[26] Hastie, T. J., & Pregibon, D. (2017). Generalized linear models. In *Statistical models in S* (pp. 195-247). Routledge.
[27] Utami, I. U., Setiawan, I., & Daniaty, D. (2021). Modelling the number of HIV/AIDS in Central Sulawesi. In *Journal of Physics: Conference Series* (Vol. 1763, No. 1, p. 012046). IOP

Publishing.

[28] Pratama, W. (2015). *Pemetaan dan pemodelan jumlah kasus penyakit tuberculosis (TBC) di provinsi Jawa Barat dengan pendekatan geographically weighted negative binomial regression (GWNBR)* (Doctoral dissertation, Institut Technology Sepuluh Nopember).

[29] Martin, B. D., Witten, D., & Willis, A. D. (2020). Modeling microbial abundances and dysbiosis with beta-binomial regression. *The annals of applied statistics*, *14*(1), 94.