ZERO-INFLATED NEGATIVE BINOMIAL MODELING IN INFANT DEATH CASE DUE TO PNEUMONIA IN EAST JAVA PROVINCE

Cindy Cahyaning Astuti [1*], Agnes Ona Bliti Puka[2], Akbar Wiguna[3]

[1,3]*Information Technology of Education, Faculty of Psychology and Educational Sciences, University of Muhammadiyah Sidoarjo,*
Majapahit Street 666B Sidoarjo, Indonesia

[2]*Mathematics Education Department, Faculty of Teacher Training and Education, Larantuka Institute of Teacher Training and Technology,*
Ki Hajar Dewantara Street Waibalun Larantuka, Indonesia,

[12] *Doctoral Student of Statistics Department, Faculty of Science and Data Analytics,*
Sepuluh Nopember Institute of Technology,
Teknik Mesin Street 175 Sukolilo Surabaya, Indonesia

Corresponding author's e-mail: *cindy.cahyaning@umsida.ac.id

ABSTRACT

Pneumonia is an acute infectious disease of the respiratory tract and an infection caused by a virus, bacteria or fungus that attacks the lung tissue. Several cases of pneumonia have resulted in deaths that occurred in toddlers aged 12-59 months. Based on official health in profile data, East Java's health in 2021 has a zero number of deaths under five aged 12-59 months due to pneumonia. Modeling data with many response variables is zero and there is overdispersion can be done using Zero Inflated Negative Binomial (ZINB) regression. This study aims to model the number of infant deaths aged 12-59 months due to pneumonia in East Java Province based on seven factors that are considered to influence the number of deaths in infants due to pneumonia. From this model, it can be seen that the factors that significantly influence the death of infants aged 12-59 months due to pneumonia in East Java Province using Zero Inflated Negative Binomial (ZINB) regression. The results of testing the parameters of the ZINB regression model show that the predictor variables that have a partial significant effect on the negative binomial model in East Java are the percentage of infants who received complete basic immunization, the percentage of coverage of under-five health services, the percentage of under-five children with malnutrition, the percentage of LBW (low birth weight babies). Selection of the best model is obtained by using the Bayesian Information Criterion (BIC) of 101,587.

1. INTRODUCTION

Modeling count data experienced development fast on forming connections between variables that are linear. If the characteristics used do not meet the linear model conditions, then need to be developed on existing linear models. Regression analysis is one example of the usual classic linear model used in searching for connections Among linear variables. On the classic linear model, there is an assumption variable response characteristic is continuous and the distribution of the residuals follows the normal distribution, but in fact, often found condition variable response characteristics are categorical and do not follow the normal distribution. According to [1], to overcome this there is development in the classical linear model, which is the Generalized Linear Model (GLM). GLM assumes variable response follows spread family exponential, which has a nature more general as well as can be applied to the condition of the response variable characteristic categorical. On GLM use a link function that relates the average of the response variables to the linear combination of the predictor variables in the model. Describing count data characteristics of several incidents on one period of time some vary greatly in value and were spread from zero until not until. If at any modeling count data a lot there is valuable observation zero on the variable response (zero inflation) so could be overcome by using the Zero Inflate Poisson Regression Model (ZIP) [2]. However, if there is a lot of data valuable observation zero and happened overdispersion then the Zero Inflated Poisson (ZIP) regression model is no longer appropriate to use. Condition overdispersion could be defined as distribution conditions Poisson where variety is bigger than average. According to [3], overdispersion on Poisson regression could result in the standard error of the resulting regression parameter estimates tending to be more low than should be, so produce conclusions that do not fit the data. If at any modeling count data there is valuable observation zero on the variable response and it happened overdispersion, then the model can be used is the Zero Inflate d Generalized regression model Poisson [4].

In its development there is another alternative to modeling case by lot valuable observation zero and happened overdispersion besides using the Zero Inflate d Generalized regression model Poisson (ZIGP), the model is Zero Inflated Negative Binomial (ZINB) regression. According to [5], the Zero Inflated Negative Binomial regression model is the model formed from a spread mixture of Poisson Gamma. The Zero Inflated Negative Binomial (ZINB) regression model can be used as another alternative in modeling cases by lot valuable observation zero and happened overdispersion because this model does not require variety must be the same as the average, besides that the Zero Inflated Negative Binomial (ZINB) regression model also has a dispersion parameter that is useful for describing variation from the data, which is usually denoted by κ (kappa). ZIP regression modeling and ZINB regression have mostly been done by [6] using the ZIP regression model by applying the model to the data collected from a studies Quality Control.

Besides using ZIP development namely Zero Inflated Conway Maxwell Poisson (ZICMP) regression for modeling the number of Dead sufferer diphtheria in each province in Indonesia, and use marginalized Zero Inflate d Generalized Poisson (MZIGP) to model amount root of 270 shoots Apple cultivar sharp. Several studies related to Zero-Inflated Negative Binomial regression were carried out by [5] applying Zero-Inflated Negative Binomial regression to neonatal tetanus data in the province of Java east with a conclusion that the Poisson regression does not fit used in the data because experience overdispersion in the Poisson model, so the Zero Inflated Negative Binomial is more suitable used. Then, research by [7] in modeling influencing factors time needed to detect patient falls in the facility health services in Korea. In addition to [8] using ZINB in modeling cases of severe pneumonia in Gorontalo. This research uses a variable amount of incidence of severe pneumonia (Y), Percentage breast milk exclusive (X1) Percentage of toddlers who get vitamin A (X2), Percentage birth weight low (X3), and Percentage immunization base complete (X4). The research results are there several influential variables significant namely X1, X2, and X3.

World Health Organization (WHO) China Country Office reports unknown cases of mysterious Pneumonia the etiology in Wuhan China on December 31, 2019, hereinafter developing COVID-19 and its cases spread in almost all countries in the world. Based on official data health in profile health East Java Province reported exists cases of COVID-19 on March 18, 2020. Since it was first reported, cases of COVID-19 then slowly increase amount cases and spread all over districts/cities in East Java Province, even in East Java Province occupied the province the number of cases confirmed the highest in Indonesia surpassed province others at the start of the pandemic occur. Based on related research relevance and the occurrence of pneumonia in East Java Province researchers interested in conducting research with the same method on pneumonia cases in East Java Province using different research variables from previous research and applying the Bayesian Information Criterion (BIC) in selecting the best model. This study aims to model cases of

pneumonia in toddlers in East Java Province and also to find out factors significant influencing cases of pneumonia in toddlers in East Java Province using Zero Inflated Negative Binomial regression.

2. RESEARCH METHODS

2.1 Data Sources

In research using secondary data sourced from East Java Province Health Profile 2022 published by the East Java Provincial Health Office. The units of observation in this study were 38 regencies /cities in East Java province which included 29 regencies and 9 cities. Amount Dead in babies ages 12-59 months consequence of disease Pneumonia in East Java Province in 2021 is 31 toddlers.

2.2 Research Variables

The variables used in this study consist of response variables and predictor variables. Variable response (Y) used in this study Total Dead in babies ages 12-59 months consequence disease Pneumonia in every Regencies / City in East Java Province, meanwhile variable predictors (X) used are as many as 7 variables. Here are the variables Responses and predictors used in this study:

Y : Total Infant Mortality Aged 12-59 Months Consequence Disease Pneumonia in East Java Province
X_1 : Percentage of Babies Getting Complete Basic Immunization
X_2 : Percentage Families with Facilities Proper Sanitation
X_3 : Percentage Complete Neonatal Visit
X_4 : Percentage Scope Toddler Health Services
X_5 : Percentage Giving Vitamin A to Toddlers 12-59 Months
X_6 : Percentage of Malnourished Toddlers
X_7 : Percentage of LBW (Low Birth Weight)

The data structure used in this study is shown in **Table 1**. As following:

Tabel 1. Research Data Structure

Regency/City	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	y_1	$x_{1.1}$	$x_{2.1}$	$x_{3.1}$	$x_{4.1}$	$x_{5.1}$	$x_{6.1}$	$x_{7.1}$
2	y_2	$x_{1.2}$	$x_{2.2}$	$x_{3.2}$	$x_{4.2}$	$x_{5.2}$	$x_{6.2}$	$x_{7.2}$
3	y_3	$x_{1.3}$	$x_{2.3}$	$x_{3.3}$	$x_{4.3}$	$x_{5.3}$	$x_{6.3}$	$x_{7.3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
38	y_{38}	$x_{1.38}$	$x_{2.38}$	$x_{3.38}$	$x_{4.38}$	$x_{5.38}$	$x_{6.38}$	$x_{7.38}$

2.3 Poisson Regression

Poisson regression in general used to analyze count data (count data). According to **[9]**, Poisson regression there is assuming Y~ Poisson (μ), this means variable response assumed spread Poisson with parameter μ. The Poisson regression model is obtained from spread Poisson defines the parameter μ as the variable covariate, with y_i being i - the observation from the variable response. Poisson regression is then used for modeling something relative event that seldom occurs in certain units. In general, the Poisson regression equation can be stated as **Equation (1)**.

$$\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}, \ i = 1,2,...,n \ \text{ and } j = 1,2,..., p \tag{1}$$

where:
p : amount variable predictor
n : amount observation
β̂ : the estimated poisson regression model

2.4 Overdispersion

Poisson regression analysis is an included regression analysis part from Generalized Linear Model (GLM). Poisson regression is used for data with variables response that follows spread Poisson (Y~ Poisson). The assumption important in this analysis is the variety must be equal to the so-called average equidispersion. However, in several studies this condition was not met, often found count data (count data) that has a variety

greater than the average is called overdispersion.Inspection overdispersion can be done using score deviance. Variety of spread poisson equals the mean ($\sigma^2 = \mu$). Overdispersion detected uses the score deviance divided by degrees of freedom to have a score greater than 1, whereas underdispersion detected by value deviance divided by degrees of freedom to have a score not enough of 1. According to **[10]**, value Deviance could state as **Equation (2)**.

$$D = 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) \tag{2}$$

where:

n : amount observation

y_i : variable the i response with i = 1,2,..., n

$\hat{\mu}_i$: variable average response y which is influenced by the value variable predictor on the i - the observation

2.5 Zero Inflation

Excessive zeros in variables response (zero inflation) are often found on Poisson regression analysis for either discrete data or count data. If a score of zero has an important meaning in research, the data can not be removed but must be included in the analysis process. Some studies can find conditions with too many zeros on the variable more response of 50 percent. According to **[11]** a magnitude valuable proportion of data zero could affect accuracy from inference. In addition, Poisson regression also becomes no longer appropriate to model the actual data.

2.6 Multicollinearity

Multicollinearity show there is a relationship between some or all variable that describes the regression model. There are two types multicollinearity that is multicollinearity perfect and imperfect multicollinearity. On multicollinearity perfect there is a linear relationship between variable predictor where one variable predictor is a linear function of variable another predictor, meanwhile imperfect multicollinearity occurs if there is an imperfect linear relationship between predictor **[12]**. Variable detection of multicollinearity can be done using the score Variance Inflation Factor (VIF). For regression with more than 2 variables, the equation to calculate the VIF value can be stated as **Equation (3)**.

$$VIF_j = \frac{1}{1 - R_j^2} \tag{3}$$

where:

R_j^2 : coefficient determination from auxiliary regression

2.7 Negative Binomial Regression

Negative Binomial Regression is one of the regression models which is applied from GLM. As application from GLM then spread Negative Binomials have a third component that is the random, component systematic, and functional link. In the Negative Binomial regression model, variables response y_i assumed have spread Negative Binomials generated from spread Poisson Gamma mix. The function probability of the Negative Binomial regression model could be stated as **Equation (4)**.

$$P(Y_i = y_i) = \frac{\Gamma\left(y_i + \frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y_i!}\left(\frac{1}{1 + \kappa\mu_i}\right)^{\frac{1}{\kappa}}\left(\frac{\kappa\mu_i}{1 + \kappa\mu_i}\right)^{y_i} \text{ with } i = 0,1,2,....,n \tag{4}$$

When $\kappa \rightarrow 0$ at the moment so spread Negative Binomials have variety $V[Y] \rightarrow \mu$. Spread Negative Binomials will approach something spread Poisson assuming the same mean and variance that is $E[Y] = V[Y] = \mu$. In the Negative Binomial regression model, y_i is a variable of the form count data. According to **[13]**, negative binomial regression model in general use function logarithm or log link in **Equation (5)**, namely:

$$\ln \mu_i = X_i^T \beta$$
$$\mu_i = \exp(X_i^T \beta) \tag{5}$$

2.8 Zero Inflated Negative Binomial (ZINB) Regression

The Zero Inflated Negative Binomial (ZINB) regression model is the model that was created from a spread mixture of Poisson Gamma. According to [14], this model can be used for modeling count data or discrete data with many scores zero on the variable response (zero inflation) and occurs overdispersion. It is a variable random with $i= 1,2,...n$ so the score from variable response happened in two circumstances. circumstances first called zero states and yield only observation worth zero, meanwhile circumstances second called negative binomial state that has spread Negative Binomials. ZINB regression consisted of two-component namely the negative binomial state model $\hat{\mu}_i$ in **Equation (6)** and the model for zero inflation $\hat{\pi}_i$ in **Equation (7)**, namely:

a) Model untuk parameter μ_i

$$\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}, i = 1,2,...,n \text{ and } j = 1,2,..., p \tag{6}$$

b) Model untuk parameter $\hat{\pi}_i$

$$\text{logit } \hat{\pi}_i = \hat{\gamma}_0 + \sum_{j=1}^{p} \hat{\gamma}_j x_{ij}, \ i = 1,2,...,n \text{ and } j = 1,2,..., p \tag{7}$$

where :
p : amount variable predictor
n : amount observation
$\hat{\beta}$: ZINB regression model parameters are estimated
$\hat{\gamma}$: ZINB regression model parameters are estimated

Estimation of the parameters of the ZINB regression model was carried out using the method Maximum Likelihood Estimation (MLE) and to maximize the function used EM (Expectation Maximization) algorithm. Function the probability of a ZINB regression model consisted of two conditions that are $y_i=0$ and $y_i>0$ and done is known that variable response y_i in this study also consists of two conditions that are zero state and negative binomial state. To describe condition y_i in detail than will be defined the return variable y_i with a latent variable Z_i as in **Equation (8)**:

$$Z_i = \begin{cases} 1 \text{ , if } y_i \text{ come from } zero\ state \\ 0, \text{ if } y_i > 0 \text{ come from } negative\ binomial\ state \end{cases} \tag{8}$$

The problem with this definition is in the circumstances negative binomial state, Z_i could value 0 or 1. Problem the could be resolved using the EM algorithm. The EM algorithm is one alternative method iteratively to maximize function likelihood on the data it contains outcome latent variable definition new variable like variable Z_i in **Equation (8)**. The EM algorithm consists of two Steps Step expectations and stage maximization. The expectation stage is the Step calculation expectation from the ln likelihood function, next Step maximization is the Step calculation to find maximizing parameter estimation ln function likelihood outcome from the Step expectation before.

2.9 Testing Model Significance

This test aims to find out the influence of variable predictors on variable response. Testing model significance is done simultaneously and partially.

a) Simultaneous Test
According to [15], simultaneous testing of model parameters is carried out using the G test statistic. The G test statistic is a ratio test possibility maximum (likelihood ratio test) used to test role variable predictors in the model together. Equality from G test statistics can be stated as **Equation (9)**.

$$G = -2\left[L_0 - L_p\right] \qquad (9)$$

where:

L_o : ln likelihood model without variable predictor (intercept model)

L_p : ln likelihood model with variable p predictor (full model)

G test statistics follow the distribution of χ^2 with degrees of freedom p. hypothesis zero rejected if the G test statistic is greater than $\chi^2_{p(\alpha)}$.

b) Partial Test

Testing partial model parameters include testing partial parameters β and γ. Partial regression parameter testing was carried out using t-test statistics, According to [13], equality from t-test statistics can be stated as **Equation (10)**.

$$t_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}, \text{ for } \beta \text{ parameter}$$

$$\qquad (10)$$

$$t_j = \frac{\gamma_j}{se(\gamma_j)}, \text{ for } \gamma \text{ parameter}$$

hypothesis zero is rejected if the value $|t|$ greater or equal to $t_{(\alpha/2;n-1)}$. Where α is the level used. On t-test statistics there is $se(\beta_j)$ and $se(\gamma_j)$ that is the standard error of parameter estimation β_j and γ_j so-called matrix variant covariance from each parameter.

2.10 Goodness of the Model

According to [15], the BIC (Akaike Information Criterion) value can be used for selecting the best model. The BIC value is calculated based on the score maximum likelihood and the number of parameters in the regression model that is formed. The equation to calculate BIC value is stated as the following **Equation (11)**.

$$\text{BIC} = -2\ln \, (maximum \, likelihood) + p\log(n) \qquad (11)$$

The best regression model is a regression model that has the smallest BIC value.

2.11 Data Analysis

Stages data analysis was performed using help R software by modeling the amount case of Pneumonia in toddlers ages (12-59 months) in each regency/City in East Java Province as follows.

a. Check the proportion score of zero on the variable response.
b. Check overdispersion is carried out using score Deviance.
c. Applying the ZINB regression model to the sum case Pneumonia in toddlers ages (12-59 months) in each regency/City in East Java Province in 2022 with variables predictors are the factors that are considered to take effect to case Pneumonia in toddlers.
d. Testing the significance of the regression model parameters. Testing is done simultaneously and partially. The test statistic used for the simultaneous test is the G test statistic and the partial test used z test statistics.
e. Interpret the ZINB regression model formed.
f. Determine the level the goodness of the Zero Inflated Negative Binomial (ZINB) regression model formed.

3. RESULTS AND DISCUSSION

Modeling amount Dead in babies ages 12-59 months consequence disease Pneumonia in East Java Province was carried out using ZINB regression. Before modeling especially First, check the overdispersion of variables response, Zero Inflation check on variable response, check multicollinearity, and ZINB regression modeling. Here are a few stages of analysis that has been done.

3.1. Examination Overdispersion

According to [11], an inspection of Poisson regression overdispersion was performed using the score Deviance divided by degrees of freedom. Condition overdispersion detected use score Deviance is divided by degrees free to have a score greater than 1. Based on results inspection obtained score Deviance is divided by degrees free of 1.549. Score Deviance divided by degrees free have a value greater than 1, this indicates that variable response experience overdispersion

3.2. Inspection Zero Inflation Variable Response

Inspection zero inflation is done by counting the percentage of valuable observation zero on the variable response. Examination results zero inflation on variables response presented in **Table 2**. As following:

Table 2. Examination Results Zero Inflation on Variables Response

Amount Dead in Infants aged 12-59 Months Consequence Disease Pneumonia	Frequency	Percentage
0	22	57.9%
1	9	23.7%
2	2	5.3%
3	2	5.3%
4	3	7.8%

Examination results zero inflation variable the response in **Table 2** shows that occur zero inflation on variables response because percentage observation worth zero more than 50 percent that is by 57.9% percent.

3.3. Multicollinearity Test

Inspection multicollinearity was carried out to find out the connection between variable predictors that explain the regression model. According to [12], the value used as a reference for the inspection of multicollinearity is the value of VIF (Variance Inflation Factor). More VIF value of 5 is proof enough to detect multicollinearity. Examination results of multicollinearity are presented in **Table 3.**

Tabel 3. Examination Results Multicollinearity

Variable Predictor	VIF value	Conclusion
X_1	1.667	
X_2	1.614	None Multicollinearity Between
X_3	1.249	Variables predictor
X_4	1.359	
X_5	1.599	
X_6	1.244	
X_7	1.109	

Examination results of multicollinearity in **Table 3.** show that doesn't exist multicollinearity between variable predictors, because all variable predictors have a VIF value less than 5.

3.4. Zero Inflated Negative Binomial (ZINB) Regression Modeling

The Zero Inflated Negative Binomial (ZINB) regression model is a regression model that can be performed on many valuable observations zero on the variable response and it happened overdispersion. This model is a development of the Negative Binomial (NB) regression model for data with a lot of valuable observation zero on the variable response (zero inflation). The ZINB regression model was applied to cases Dead in babies ages 12-59 months consequence of the disease Pneumonia in every Regencies/City in East Java Province. Modeling case Dead in babies ages 12-59 months consequence disease Pneumonia in the ZINB regression model using eight predictor variables, then next reduced becomes four variable predictor. To know the level significance results parameter estimation in the ZINB regression model, tested simultaneous and partial significance. According to testing significance results parameter estimation in the ZINB regression model simultaneously uses G test statistics and testing partial significance use z test statistic.

Parameter estimation results of the ZINB regression model in cases Dead in babies ages 12-59 months consequence disease Pneumonia as well as score G test statistics and t-test statistics presented in full in **Table 4.**

Tabel 4. Parameter Estimation Results of the ZINB Regression Model Using 8 Predictor Variables

Parameter	Estimation	Z count	(Pr >| z|)
$\hat{\beta}_0$	23.393	2.077	0.038*
$\hat{\beta}_1$	-0.066	-2.144	0.032*
$\hat{\beta}_2$	-0.055	-1.927	0.054
$\hat{\beta}_3$	-0.209	-1.879	0.060
$\hat{\beta}_4$	0.104	2.625	0.009*
$\hat{\beta}_5$	0.007	0.187	0.851
$\hat{\beta}_6$	0.319	2.657	0.008*
$\hat{\beta}_7$	-0.970	-2.638	0.008*
$\hat{\gamma}_0$	-7.046	-0,001	0,999
$\hat{\gamma}_1$	11.458	0.140	0.889
$\hat{\gamma}_2$	7.830	0.090	0.928
$\hat{\gamma}_3$	-15.907	-0.191	0.849
$\hat{\gamma}_4$	2.559	0.082	0.935
$\hat{\gamma}_5$	-6.847	-0.094	0.925
$\hat{\gamma}_6$	16.890	0.100	0.920
$\hat{\gamma}_7$	-23.102	-0.168	0.866
	Pr (>Chisq = 0.0089*)		

*) Significant with a degree significance of 5%

ZINB regression model equation that is formed is as follows the following:

The negative binomial state model for $\hat{\mu}$

$$\hat{\mu} = exp(23.393 - 0.066\,X_1 - 0.055\,X_2 - 0.209\,X_3 + 0.104\,X_4 + 0.007\,X_5 + 0.104\,X_6 - 0.970\,X_7)$$

Zero inflation model for $\hat{\pi}$

$$\hat{\pi} = \frac{exp(-7.046 + 11.458\,X_1 + 7.830\,X_2 - 15.907\,X_3\,2.559\,X_4 - 6.847\,X_5 + 16.890\,X_6 - 2.102\,X_7)}{1 + exp(-7.046 + 11.458\,X_1 + 7.830\,X_2 - 15.907\,X_3\,2.559\,X_4 - 6.847\,X_5 + 16.890\,X_6 - 2.102\,X_7)}$$

Test results in significance parameter estimation of the ZINB regression model simultaneously with the level significance by 5% based on Chi-Square test statistics. Pr (> Chisq = 0.0468*) indicates that simultaneously the variable predictors X_1, X_2, X_3, X_4, X_5, X_6, X_7 give influence significant to the variable response. For results significance parameter estimation of the ZINB regression model partially by level significance by 5% based on the z-test statistic. Based on Table 4. There are four variable predictors of parameter estimation in the negative binomial state model which has a p-value not enough of α (0.05). It shows that variable influential predictor partially significant in the negative binomial state model is the percentage baby get immunization base complete (X_1), the percentage scope service health (X_4), the Percentage of toddler nutrition poor (X_6) and the percentage of LBW (Low Birth Weight) (X_7). So that The ZINB regression model equation is formed based on level partial parameter significance as follows:

The negative binomial state model for $\hat{\mu}$

$$\hat{\mu} = exp(23.393 - 0.066\,X_1 + 0.104\,X_4 + 0.104\,X_6 - 0.970\,X_7).$$

The interpretation of the zero-inflation model for πis as follows.

1) Every 1 percent addition baby gets immunization base complete (X_1) then will lower opportunity amount Dead in babies ages 12-59 months consequence disease Pneumonia equal to exp(0.066)=1.068 times of amount Dead in babies ages 12-59 months consequence disease Recurrent pneumonia, if other variables are valuable constant.

2) Every 1 percent addition complete neonatal visits (X_4)) then will raise the opportunity amount Dead in babies ages 12-59 months consequence disease Pneumonia of exp(0.104)=1.109 times of amount Dead

in babies ages 12-59 months consequence disease Recurrent pneumonia, if other variables are valuable constant.

3) Every 1 percent addition toddler nutrition bad (X_6) then will raise the opportunity amount Dead in babies ages 12-59 months consequence disease Pneumonia of exp(0.008)=1.008 times of amount Dead in babies ages 12-59 months consequence disease Recurrent pneumonia, if other variables are valuable constant.

4) Every addition of 1 percent Low Birth Weight (LBW) (X_6) then will lower the opportunity amount Dead in babies ages 12-59 months consequence disease Pneumonia of exp (0.008)=1.008 times of amount Dead in babies ages 12-59 months consequence disease Recurrent pneumonia, if other variables are valuable constant.

3.5 Goodness of the Model

Size The goodness of the model used in this study is the Bayesian Information Criterion (BIC). The following are the BIC values for each model presented in **Table 5,** as following:

Tabel 5. BIC Values

Model	BIC
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	110,144

Based on the BIC value is known that the model with 4 predictor variables namely X_1, X_4, X_6, X_7 has the smallest BIC value. So that size is the goodness of the best model in this study using BIC.

4. CONCLUSIONS

Based on the results modeling amount Dead in babies ages 12-59 months consequence disease Pneumonia in East Java Province was carried out using known ZINB regression predictor variables X_1, X_2, X_3, X_4, X_5, X_6, simultaneously give influence significant to variable response, As for the results significance parameter estimation of the ZINB regression model partially by level significance by 5% based on the z test statistic there are four variable predictors of parameter estimation in a negative binomial state model that has a p-value not enough of α (0.05). That is a variable percentage baby get immunization base complete (X_1), the percentage of scope service health (X_4), and the percentage of LBW (Low Birth Weight) (X_7).

REFERENCES

[1]   J. Fox, *Applied Regression Analysis and Generalized Linear Models*, vol. 3rd Edition. Sage Publication, Inc, 2016.
[2]   H. Permadi and S. D. Maulidah, "Zero Inflated Poisson Regression Analysis on IMB Ownership in Sidoarjo Regency 2019," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, May 2021. doi: 10.1088/1742-6596/1872/1/012030.
[3]   H. Campbell, "The Consequences of Checking for Zero-Inflation and Overdispersion in The Analysis of Count Data," *Methods Ecol Evol*, vol. 12, no. 4, pp. 665–680, Apr. 2021, doi: 10.1111/2041-210X.13559.
[4]   F. Famoye and J. S. Preisser, "Marginalized zero-inflated generalized Poisson regression," *J Appl Stat*, vol. 45, no. 7, pp. 1247–1259, May 2018, doi: 10.1080/02664763.2017.1364717.
[5]   P. Saengthong, W. Bodhisuwan, and A. Thongteeraparp, "The Zero Inflated Negative Binomial Crack distribution: Some Properties and Parameter Estimation," 2015. [Online]. Available: http://www.sjst.psu.ac.th
[6]   C. C. Astuti and A. D. Mulyanto, "Estimation Parameters And Modelling Zero Inflated Negative Binomial," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 4, no. 3, pp. 115–119, Nov. 2016, doi: 10.18860/ca.v4i3.3656.
[7]   K. J. Hong and J. Kim, "Risk Factors Preventing Immediate Fall Detection: A Study Using Zero-Inflated Negative Binomial Regression," *Asian Nurs Res (Korean Soc Nurs Sci)*, vol. 15, no. 4, pp. 272–277, Oct. 2021, doi: 10.1016/j.anr.2021.09.001.
[8]   N. Achmad, M. R. F. Payu, and Y. Rahim, "Pemodelan Pneumonia Berat Menggunakan Regresi Zero Inflated Negative Binomial di Gorontalo," *Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 10, no. 1, pp. 45–53, May 2022, doi: 10.34312/euler.v10i1.13990.
[9]   J. Kruppa and L. Hothorn, "A Comparison Study on Modeling of Clustered and Overdispersed Count Data for Multiple Comparisons," *Journal of Applied Statistics*. Taylor and Francis Ltd., pp. 1–13, 2020. doi: 10.1080/02664763.2020.1788518.
[10] Y. Tiara, M. N. Aidi, E. Erfiani, and R. Rachmawati, "Overdispersion Handling In Poisson Regression Model By Applying Negative Binomial Regression," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 1, pp. 0417–0426, Apr. 2023, doi: 10.30598/barekengvol17iss1pp0417-0426.
[11] J. Stachurski, *A Primer in Econometric Theory*. London, England: MIT Press, 2017.

[12] T. Kyriazos and M. Poga, "Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions," *Open J Stat*, vol. 13, no. 03, pp. 404–424, 2023, doi: 10.4236/ojs.2023.133020.

[13] H. Zamani and N. Ismail, "Functional Form for the Zero Inflated Generalized Poisson Regression Model," *Commun Stat Theory Methods*, vol. 43, no. 3, pp. 515–529, Feb. 2014, doi: 10.1080/03610926.2012.665553.

[14] Y. Gençtürk and A. Yiğiter, "Modelling Claim Number Using a New Mixture Model: Negative Binomial Gamma Distribution," *J Stat Comput Simul*, vol. 86, no. 10, pp. 1829–1839, Jul. 2016, doi: 10.1080/00949655.2015.1085987.

[15] P. Roback and J. Legler, "Beyond Multiple Linear Regression: Applied Generalized Linear Models and Multilevel Models in R," 2021.