Analysis of Diphtery Disease Mortality in Indonesia Using Zero Inflated Conway Maxwell Poisson Regression Model

Eka Hermia[a*], Nurtiti Sunusi[a**], Erna Tri Herdiani[a***]

[*] *Corresponding Author: ekahermia05@gmail.com*
[**] *Corresponding Author nurtitisunusi@unhas.ac.id*
[***] *ernatriherdiani@unhas.ac.id*
[a]*Departement of Statistics, Hasanuddin University, Makassar, 90245, Indonesia*

Abstract

Diphtheria is an upper respiratory tract disease caused by the bacteria Corynebacterium Diphtheriae. Diphtheria is a dangerous infectious disease, because there were 23 deaths from 529 cases in 2019. The data used in this final project research is secondary data taken from the Health Office in 2019. The number of deaths of diphtheria patients is calculated data and the assumption is distributed as a Poisson distribution. . The number of zero data of 73.5 percent indicates an overdispersion of the response variables. One of the conditions of Poisson Regression is that it requires equality between the mean and variance, this model is not appropriate to be applied to the dispersion below and above. The Zero Inflated Conway Maxwel Poisson (ZICMP) regression model can be an alternative because it is flexible for under- and over-dispersion data and can solve the problem of excess zero values in the response variable. In this study, ZICMP was applied to mortality cases in diphtheria patients in Indonesia. This study aims to examine the probability function and develop an algorithm to estimate ZICMP parameters and apply the ZICMP model to mortality cases in diphtheria patients. The Maximum Likelihood Estimation (MLE) method is used to estimate the parameters in ZICMP and the probability function is maximized using the Expectation-Maximization (EM) algorithm.

Keywords: Overdispersion; Zero Inflation; Zero Inflated Conway Maxwel Poisson (ZICMP); Diphtheria.

1. Introduction

Diphtheria is an infectious infection caused by Corynebacterium diptheria (Huwae, 2012). Diphtheria is characterized by sore throat, fever, and the formation of the lining of the tonsils and throat. This disease spreads through direct physical contact, or through breathing in the air containing secretions from patients who cough or sneeze (MOH, 2019). The Indonesian Ministry of Health noted that the number of diphtheria cases in 2019 spread to almost all regions in Indonesia. The number of diphtheria cases in 2019 was 529 cases, the number of deaths was 23 cases, with a CFR of 4.35%. The number of diphtheria cases in 2019 decreased significantly compared to 2018 (1,386 cases). The number of deaths due to diphtheria also decreased compared to the previous year (29 cases) (Depkes, 2019). One of the efforts to reduce the number of deaths of diphtheria patients is to continue examine the factors that cause it. The relationship between risk factors can be analyzed using regression analysis. If the response variable used in this study is the number of mortality of patients with diphtheria in the form of discrete data, then one of the regression models used is the Poisson regression model.

Data analysis using Poisson regression must meet assumptions such as the variance and mean values of the response variables are the same or equidisperse. In reality, this assumption is not fully fulfilled, because discrete type data often experience cases of overdispersion (high deviation) and underdispersion. In fact, the calculated data does not only experience overdispersion or underdispersion, but can also experience excess zero. Excess zero is a condition when the percentage of zero values in the data is greater than other values. Count data containing zero values can be estimated using Poisson regression. However, for data with a very large number of zero values (excess zero) it requires a certain method to overcome it. If Poisson regression is still used, then the parameter estimate is not suitable for the estimation of excess zero (Nadhiroh, 2009). This has led to the development of statistical methods to overcome these problems. The proportion of excess zeros in the response variable can cause data overdispersion.

Another option is being developed for alternative for modeling cases with numerous observations. that are zero and overdispersion occurs, namely the Zero Inflated Conway Maxwell Poisson regression. Sellers and Raim (2016) introduced the Zero Inflated Conway Maxwell Poisson model, which is basically a combination of the Poisson distribution in the ZIP model with the Conway-Maxwell-Poisson (CMP) distribution.

To estimate the parameters, the Maximum Likelihood Estimation (MLE) method can be used in regression models whose data follows a certain distribution. The MLE method is done by maximizing the likelihood function. However, in general, the maximum likelihood function cannot be solved analytically. Therefore, if an implicit or non-linear form is obtained, it can be solved using the Newton-Raphson (NR) algorithm, Fisher Scoring, or Expectation Maximization (EM) to get the numerical solution.

Therefore, in this study, we will examine the Zero Inflated Conway Maxwell Poisson (ZICMP) regression model using the MLE method, to maximize the likelihood function using the EM algorithm.

Based on the description above, this study aims to: (1) To examine the estimation of the parameters of the Zero Inflated Conway Maxwell Poisson regression model using the Maximum Likelihood Estimation method. (2) Determine the factors that influence the death of Diphtheria patients in Indonesia in 2019 using the Zero Inflated Conway Maxwell Poisson (ZICMP) regression model

2. Material

This section will explain the research data, sampling techniques, and research variables.

2. 1 Conway Maxwell Poisson

The Conway-Maxwell-Poisson distribution (CMP) was first introduced by Conway and Maxwell in 1962 in the context of linear systems, which later became known as a generalization of the Poisson distribution including Bernoulli and Geometrics as special and flexible cases for controlling underdispersion and overdispersion. The COM-Poisson distribution developed by Conway and Maxwell with the same parameter λ as the Poisson parameter and the dispersion parameter (ϕ). random variable Y ~ COM-Poisson (λ , v) and parameters $\lambda > 0$ and $v \geq 0$, the form of the COM-Poisson probability function is defined as follows (Shmueli, 2005):

$$P(Y = y|\lambda,) = \frac{\lambda^y}{(y_i!)^{\phi} Z(\lambda_i,\phi)}, \text{for } y = 0,1,2, \dots \text{n } \lambda > 0, \phi \geq 0$$

2.2 Zero Inflated Conway Maxwell Poisson

Sellers and Raim (2016) introduced the ZICMP model, which essentially replaces the Poisson distribution in the ZIP model with the Conway-Maxwell-Poisson distribution. The advantage of ZICMP is that it can handle not only overdispersion but also underdispersion in a count. ZICMP is very flexible because it allows dispersion in all directions. The probability function of Conway Maxwell Poisson's zero-inflated regression model (ZICMP) can be expressed as equation (Sim, 2018).

$$P(Y_i = y_i) = \begin{cases} \pi_i + \dfrac{(1-\pi_i)}{Z(\lambda,\phi)}, & \text{for } y_i = 0 \\ (1-\pi_i)\dfrac{\lambda^y}{(y_i!)^\phi Z(\lambda_i,\phi)}, & \text{for } y_i > 0 \end{cases}$$

3. Method

3.1 Data Sources and Research Variables

The case data used in this study is data on the mortality rate of diphtheria patients in 2019. The data used in this study is secondary data obtained from the official website of the Republic of Indonesia Health Service, https://pusdatin.kemkes.go.id. Completely the research variables can be seen on table 1.

Table 1. List of Variable

Variable	Variable Name
Endogenous	Y_1 = the mortality rate of patients diphtheria in each province
Exogenous	X_1 = Percentage of Poverty (%) X_2 = Percentage of Malnutrition (%) X_3 = Percentage of Population Accessing Clean Water (%) X_4 = Percentage of Food Management Places (Tpm) Meet Health Requirements (%), X_5 = Percentage of Public Places that meet Health requirements (%), X_6 = Number of Hospitals

3.2 Step of Analysis

The research steps of the ZICMP regression model using the Maximum Likelihood Estimation (MLE) method and the Expectation-Maximization (EM) algorithm on the mortality rate data of Diphtheria patients in 2019 are as follows:

1. Perform a Poisson distribution fit test
2. Perform Overdispertion Test
3. Perform Multicollinearity testing
4. Estimating Zero Inflated Conway Maxwell Poisson (ZICMP) regression model parameters using the MLE method with the EM algorithm
5. Applying the Zero Inflated Conway Maxwell Poisson (ZICMP) regression model in the case of mortality of Diphtheria patients in Indonesia in 2019 with predictor variables are the factors that are considered to have an effect on Diphtheria cases

6. Performing parameter significance test simultaneously and partially one-on-one

4. Result

4.1 ZICMP Regression Model Parameter Estimation

The estimation of the Zero Inflated Conway Maxwell Poisson (ZICMP) regression parameter in this study uses an iterative technique, the Expectation-Maximization (EM) algorithm. Before that, it can be seen that the Maximum Likelihood Estimation (MLE) from the ZINB pdf in Equation 1 is as follows:

$$P(Y_i = y_i) = \begin{cases} \pi_i + \frac{(1-\pi_i)}{Z(\lambda_i, \upsilon)}, & \text{for } y_i = 0 \\ (1-\pi_i)\frac{\lambda^{y_i}}{(y_i!)^{\upsilon} Z(\lambda_i, \upsilon)}, & \text{for } y_i > 0 \end{cases} \quad (1)$$

With $0 \leq \pi_i \leq 1, v_i \geq 0, \upsilon$ are the dispersion parameters and $Z(\lambda_i, \upsilon) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu}}$ normalize the distribution.

Lambert [9] suggest a combined model for λ_i and π_i is:

$$\ln(\lambda_i) = X_i^T \beta \quad (2)$$

$$\lambda_i = \exp(X_i^T \beta)$$

And

$$logit\ (\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = X_i^T \gamma \quad (3)$$

The results of Equation (2) and Equation (3) are substituted into equation (3.1) to get the following Equation (4).

$$P(Y_i = y_i) = \begin{cases} \dfrac{\exp(X_i^T \gamma)}{1 + \exp(X_i^T \gamma)} + \dfrac{1}{1 + \exp(X_i^T \gamma)}\left(\dfrac{1}{Z(\lambda_i, \upsilon)}\right), & \text{for } y_i = 0 \\ \dfrac{1}{1 + \exp(X_i^T \gamma)} \dfrac{[\exp(X_i^T \beta)]^{y_i}}{(y_i!)^{\nu}\ Z(\lambda_i, \upsilon)}, & \text{for } y_i > 0 \end{cases} \quad (4)$$

$\boldsymbol{X}$ is a matrix of size $n \times (p+1)$ that contains predictor variables related to the probability at zero state and CMP state, while $\boldsymbol{v}, \boldsymbol{\gamma}$ *and* $\boldsymbol{\beta}$ are a vector of size $(p+1) \times 1$ from the regression parameters to be estimated. The form of the likelihood function of the Zero Inflated Conway Maxwell Poisson (ZICMP) regression model can be written as equations (5) and (6).

for $y_i = 0$

$$L(\boldsymbol{v}, \boldsymbol{\beta}, \boldsymbol{\gamma}; y_i) = \prod_{i=1}^{n} \left[\frac{\exp(\boldsymbol{X}_i^T \boldsymbol{\gamma})}{1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\gamma})} + \frac{1}{1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\gamma})}\left(\frac{1}{Z(\lambda_i, \upsilon)}\right)\right] \quad (5)$$

for $y_i > 0$

$$L(\boldsymbol{v}, \boldsymbol{\beta}, \boldsymbol{\gamma}; y_i) = \prod_{i=1}^{n} \left[\frac{1}{1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\gamma})} \frac{\left[\exp(\boldsymbol{X}_i^T \boldsymbol{\beta})\right]^{y_i}}{(y_i!)^v \, Z(\lambda_i, v)} \right] \tag{6}$$

Thus, the form of the total ln-likelihood function is is as follows.

$$\begin{aligned} \ln L(v, \beta, \gamma; y_i) = & \sum_{i=1}^{n} \ln\left[exp(X_i^T \gamma) + \left(Z(\lambda_i, v)\right)\right] - \sum_{i=1}^{n} \ln[1 + \exp(X_i^T \gamma)] - \sum_{i=1}^{n} \ln[1 + \exp(X_i^T \gamma)] \\ & + \sum_{i=1}^{n} (y_i \, X_i^T \beta) - \sum_{i=1}^{n} [v \ln(y_i!)] - \sum_{i=1}^{n} lnZ(\lambda_i, v) \end{aligned} \tag{7}$$

The probability function of the Zero Inflated Conway Maxwell Poisson (ZICMP) regression model consists of two conditions, $y_i = 0$ and $y_i > 0$ and it is known that the response variable y_i in this study also consists of two conditions, namely zero state and COMPoisson state. So, to describe the condition y_i in detail, it will be redefined variable y_i with a latent variable Z_i.

In order to maximize the likelihood function on data containing latent variables as a result of the definition of new variables, such as variable Z_i on the equation, another iterative method is available. This method is called the EM algorithm. Determine the distribution of the latent variable Z_i before moving on to the expectation stage, which is represented by equation (11).

$$P(Z_i = z_i) = \begin{cases} \pi_i, & If \ z_i = 1 \\ 1 - \pi_i, & If \ z_i = 0 \end{cases} \tag{8}$$

When $Z_i = 1$ then the opportunity for Z_i will be equal with opportunity y_i in the zero state condition that is equal to π_i, meanwhile when $Z_i = 0$ then the opportunity for Z_i will be equal with the opportunity y_i in the COMpoisson state, which is $(1 - \pi_i)$. Furthermore, a joint distribution is formed between y_i and Z_i which is as equation (9).

$$P(y_i, z_i) = \begin{cases} (z_i) \frac{\exp(\boldsymbol{X}_i^T \boldsymbol{\gamma})}{1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\gamma})} + (1 - z_i) \frac{1}{1 + exp(X_i^T \gamma)} \left(\frac{1}{Z(\lambda_i, v)} \right), & \text{for } y_i = 0 \\ (1 - z_i) \frac{1}{1 + exp(X_i^T \gamma)} \frac{\left[exp(X_i^T \beta)\right]^{y_i}}{(y_i!)^v \, Z(\lambda_i, v)}, & \text{for } y_i > 0 \end{cases} \tag{9}$$

While the likelihood function is as follows:

for $y_i = 0$

$$l(\boldsymbol{v}, \boldsymbol{\beta}, \boldsymbol{\gamma} | y_i, z_i) = \prod_{i=1}^{n} \left[(z_i) \frac{\exp(\boldsymbol{X}_i^T \boldsymbol{\gamma})}{1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\gamma})} + (1 - z_i) \frac{1}{1 + exp(X_i^T \gamma)} \left(\frac{1}{Z(\lambda_i, v)} \right) \right] \tag{10}$$

for $y_i > 0$

$$l(\boldsymbol{v}, \boldsymbol{\beta}, \boldsymbol{\gamma} | y_i, z_i) = \prod_{i=1}^{n} \left[(1 - z_i) \frac{1}{1 + exp(X_i^T \gamma)} \frac{[exp(X_i^T \beta)]^{y_i}}{(y_i!)^v \, Z(\lambda_i, v)} \right] \tag{11}$$

With the total ln-likelihood can be given by $l_t = l(\boldsymbol{v}, \boldsymbol{\beta}, \boldsymbol{\gamma} | y_i, z_i)_{y_i = 0} + l(\boldsymbol{v}, \boldsymbol{\beta}, \boldsymbol{\gamma} | y_i, z_i)_{y_i > 0}$:

$$L_t = \sum_{i=1}^{n} (z_i X_i^T \gamma \ ) - \ln\left(1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\gamma}) \ \right) + \sum_{i=1}^{n} (1 - z_i) [(y_i \, X_i^T \beta) - (v \ln(y!)) - (\ln Z(\lambda_i, v))] \tag{12}$$

To get the maximum value of the likelihood function, the first and second derivative of the ln-likelihood function for $\boldsymbol{v}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is carried out as follows:

$$\frac{\partial l(\boldsymbol{v},\boldsymbol{\beta}|y_i,z_i)}{\partial \boldsymbol{v}} = \sum_{i=1}^{n}(1-z_i)\left[-\ln(y_i!) - \frac{1}{Z_i}\left(\frac{\partial Z_i}{\partial \boldsymbol{v}}\right)\right]$$

$$\frac{\partial^2 l(\boldsymbol{v},\boldsymbol{\beta}|y_i,z_i)}{\partial \boldsymbol{v}\partial \boldsymbol{v}} = \sum_{i=1}^{n}(1-z_i)\left[\frac{1}{Z_i^2}\left(\frac{\partial Z_i}{\partial \boldsymbol{v}}\right)^2 - \frac{1}{Z_i}\left(\frac{\partial^2 Z_i}{\partial \boldsymbol{v}^2}\right)\right]$$

$$\frac{\partial l(\boldsymbol{v},\boldsymbol{\beta}|y_i,z_i)}{\partial (\boldsymbol{\beta^T})} = \sum_{i=1}^{n}(1-z_i)\left[(y_i X_i^T) - \frac{1}{Z_i}\left(\frac{\partial Z_i}{\partial \boldsymbol{\beta}}\right)\right]$$

$$\frac{\partial^2 l(\boldsymbol{v},\boldsymbol{\beta}|y_i,z_i)}{\partial (\boldsymbol{\beta^T})\partial(\boldsymbol{\beta})} = \sum_{i=1}^{n}(1-z_i)\left[\frac{1}{Z_i^2}\left(\frac{\partial Z_i}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial Z_i}{\partial \boldsymbol{\beta^T}}\right) - \frac{1}{Z_i}\left(\frac{\partial^2 Z_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta^T}}\right)\right]$$

$$\frac{\partial^2 l(\boldsymbol{v},\boldsymbol{\beta}|y_i,z_i)}{\partial (\boldsymbol{v})\partial(\boldsymbol{\beta})} = \sum_{i=1}^{n}(1-z_i)\left(\frac{1}{Z_i^2}\left(\frac{\partial Z_i}{\partial \boldsymbol{v}}\right)\left(\frac{\partial Z_i}{\partial \boldsymbol{\beta}}\right) - \frac{1}{Z_i}\left(\frac{\partial^2 Z_i}{\partial \boldsymbol{v}\partial \boldsymbol{\beta}}\right)\right)$$

$$\frac{\partial l(\boldsymbol{\gamma}|y_i,z_i)}{\partial (\boldsymbol{\gamma^T})} = \sum_{i=1}^{n}(z_i X_i^T) - \sum_{i=1}^{n}\frac{X_i^T \exp(X_i^T\gamma)}{1+\exp(X_i^T\gamma)}$$

$$\frac{\partial^2 l(\boldsymbol{\gamma}|y_i,z_i)}{\partial (\boldsymbol{\gamma^T})\partial(\boldsymbol{\gamma})} = -\sum_{i=1}^{n}\frac{[X_i^T X_i][\exp(X_i^T\gamma)]^2}{[1+\exp(X_i^T\gamma)]^2} - \frac{[X_i^T X_i][\exp(X_i^T\gamma)]}{[1+\exp(X_i^T\gamma)]}$$

Expectation stage

Changing the variables z_i with $z_i^{(m)}$ with $m = 0,1,2,\ldots.$ which is the expectation of z_i as follows

$$z_i^{(m)} = E\left(z_i \middle| y_i, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(\boldsymbol{m})}\right)$$
$$= P(z_i|y_i, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(\boldsymbol{m})})$$

For $y_i = 0$

$$= \frac{1}{1+\exp\left(-\boldsymbol{X_i^T}\boldsymbol{\gamma}^{(m)} - \exp\left(\boldsymbol{X_i^T}\boldsymbol{\beta}^{(\boldsymbol{m})}\right)\right)} \quad (13)$$

For $y_i > 0$

$$z_i^{(m)} = P\left(z_i = \boldsymbol{0} \middle| y_i, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\beta}^{(\boldsymbol{m})}\right)$$
$$= 0$$

Maximization stage

1. Specifies the initial estimate for the parameter $\hat{\beta}^{(0)}$ and $\hat{\gamma}^{(0)}$ obtained by the OLS method.
2. Perform iterations starting from $m = 0$ by entering $\hat{\beta}^{(0)}$ dan $\hat{\gamma}^{(0)}$ to each gradient vector $\boldsymbol{g}$ and the Hessian matrix $\boldsymbol{H}$. Then iterate using the following equation

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} - \left[H\left(\hat{\beta}^{(m)}\right)\right]^{-1} g\left(\hat{\beta}^{(m)}\right)$$

And

$$\hat{\gamma}^{(m+1)} = \hat{\gamma}^{(m)} - \left[H\left(\hat{\gamma}^{(m)}\right)\right]^{-1} g\left(\hat{\gamma}^{(m)}\right)$$

3. Repeat step 2 and step 3 over and over again. The iteration will stop when the parameters have converged, i.e. when $\left\|\hat{\beta}^{(m+1)}-\hat{\beta}^{(m)}\right\| \leq 10^{-3}$ dan $\left\|\hat{\gamma}^{(m+1)}-\hat{\gamma}^{(m)}\right\| \leq 10^{-3}$ then, take $\hat{\beta}^{(m+1)}$ as an estimate $\hat{\beta}^{(m)}$ and $\hat{\gamma}^{(m+1)}$ as an estimate $\hat{\gamma}^{(m)}$

4.2 Description Data

4.2.1 Poisson Distribution Fit Test

The Poisson distribution fit test is carried out using the Kolmogorov-Smirnov test with the following hypothesis:

Hypothesis

H_0 : Response variable with Poisson distribution

H_1 : Response variable is not Poisson distribution

Test Statistics :

$D_{count} = \max|F_n(Y) - F_0(Y)|$

The result is the value $D_{value} = 0.060 < D_{table} = 0.233$, this shows that the number of cases of mortality in diphtheria in Indonesia in 2019 has a Poisson distribution.

4.2.2 Dispersion Test

Poisson regression overdispersion examination is carried out using the Deviance value divided by degrees of freedom with the following hypothesis:

Hypothesis

H_0 : $v = 1$ (Response variable is not overdispersion)

H_1 : $v > 1$ (Response variable overdispersion)

Test Statistics :

$v = \frac{x^2}{db}$

The Deviance value is divided by the degree of freedom of 1.667 which is greater than 1, this indicates that the number of cases of mortality in diphtheria in Indonesia in 2019 is experiencing overdispersion.

4.2.3 Multicollinearity test

Multicollinearity examination was conducted to determine the relationship between the predictor variables that explained the regression model. Multicollinearity examination was detected using the VIF (Variance Inflation Factor) value with the following hypothesis:

H_0 : $VIF < 10$ (there is no multicollinearity between predictor variables)

H_1 : $VIF > 10$ (there is multicollinearity between predictor variables)

Test Statistics :

$VIF_j = \frac{1}{1-R_j^2}$ dengan $j = 1,2,\ldots,p$

Table 2 Multicollinearity Examination Results

Predictor Variable	Value VIF
X_1	1.581
X_2	1.647
X_3	1.551
X_4	1.633
X_5	1.476
X_6	1.351

Based on Table 1, it can be seen that the relationship between the independent variables does not have a strong relationship between the variables, so it can be concluded that there is no case of multicollinearity.

4.2.4 Zero Inflation Check for Response Variables.

Zero inflation check is done by calculating the percentage of zero observations on the response variable. The following will display a bar chart that describes the frequency of the number of diphtheria diseases mortality rates.

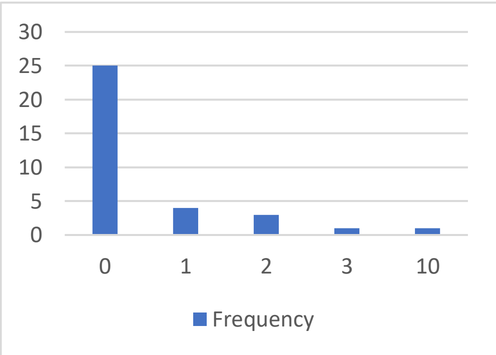

Figure 1 Frequency Graph of diphtery disease mortality rates

Based on the diagram above shows that the very high frequency of observations that are zero is 25 times. this show that there is zero inflation in the response variable because the percentage of observations is zero more than 50 percent, which is 73.53 percent.

4.3 The Application to Data the mortality rate of diphtheria patients in 2019

The Zero Inflated Conway Maxwell Poisson (ZICMP) regression model aims to improve the data that has overdispersion and excess zero in the dependent variable. The estimation results of the ZICMP regression parameter, the value of the G statistical test and the Wald and AIC test are as follows:

Table 3. Parameter Estimation of Maxwell Poisson Zero Inflated Conway Model (ZICMP)

Parameter	Estimation	Parameter	Estimation
β_0	2.1769	γ_0	4.2344
β_1	-0.0289	γ_1	-0.0744
β_2	0.1635	γ_2	-0.1669
β_3	-0.0110	γ_3	-0.0088
β_4	-0.0825	γ_4	-0.0428
β_5	0.0325	γ_5	0.0037
β_6	0.0069	γ_6	0.009
Test statistics G= 57.6			

The equation of the Zero Inflated Conway Maxwell Poisson (ZICMP) regression model that is formed is as follows.

a. Model *count* $\hat{\mu}$

$$\hat{\mu} = \exp\left(2.1769 - 0.0289X_1 + 0.1635X_2 - 0.0110X_3 - 0.083X_4 - 0.033X_5 + 0.007X_6\right)$$

b. Model zero inflation for $\hat{\pi}$

$$\hat{\pi} = \frac{exp\left(4.2344 - 0.074X_1 - 0.1669X_2 - 0.0088X_3 - 0.043X_4 + 0.0037X_5 - 0.009X_6\right)}{1 + exp\left(4.2344 - 0.074X_1 - 0.1669X_2 - 0.0088X_3 - 0.043X_4 + 0.0037X_5 - 0.009X_6\right)}$$

Based on Table 3, the simultaneous parameter significance test of the ZICMP regression model with G was found to have a value of 57.6. Because the value of $G = 57{,}6 > 22{,}36 = X^2_{(0{,}05;13)}$ which means that there is at least one independent variable that has a significant effect on the dependent variable.

Then the significance of each parameter of the ZICMP regression model will be tested using the Wald test for the model (μ). Obtained parameter $\beta_1, \beta_3, and\ \beta_6$ which is significant while the other parameters are not significant.

Based on these conclusions, the following model is obtained:

$$\hat{\mu} = \exp\left(7.675 - 0.346X_1 - 0.074X_3 + 0.015X_6\right)$$

For Model $(\hat{\pi})$, Testing the significance of the parameters individually used Wald test. Obtained only a significant γ_4 parameter.

so that the model can be formed as follows

$$\hat{\pi} = \frac{exp\left(-1.316 + 0.028X_4\right)}{1 + exp\left(-1.316 + 0.028X_4\right)}$$

The interpretation of the model formed from ZICMP is based on the value of $exp(\beta)$. And here is the best interpretation of the ZICMP regression model:

a. For Model *count* $\hat{\mu}$

Every 1 percent increase in poverty percentage (X_1) it will reduce the average number of cases of mortality in diphtheria patients by $exp(0.346) = 1{,}4134$ if other variables are constant. Every 1 percent addition Complete basic immunization coverage (X_3) it will reduce the average number of cases of mortality in diphtheria patients by $exp(0.074) = 1{,}076$ if other variables are constant. Every addition 1 Number of hospitals (X_6) it will increase the average number of cases of mortality of diphtheria patients by $exp(0.015) = 1{,}015$ if other variables are constant.

b. Model zero inflation for $\hat{\pi}$

Every 1 Percentage of Food Management Places (Tpm) Meet Health Requirements (X_4) it will reduce the average number of cases of mortality in diphtheria patients by $\frac{exp(0.028)}{1+\ exp(0.028)} = 0.5069$ if other variables are constant.

The next stage is the selection of the best model in the ZICMP regression by looking at the AIC (Akaike's Information Criterion) value. To select the best model, the AIC value for the Conway Maxwell Poisson regression model was compared with the AIC value for the ZICMP regression model. Prior to that, regression modeling was also carried out for the Conway Maxwell Poisson model on maternal and infant mortality data using the MLE method so that the AIC value was obtained as shown in Table 4

Table 4. AIC values of the CMP and ZICMP Regression models

Model Regression	Value AIC
Conway Maxwell Poisson	87.6
Zero Inflated Conway Maxwell Poisson	79.5

Table 4 shows that the ZICMP regression model has an AIC value that is smaller than the Conway Maxwell Poisson regression model so that the ZICMP regression model is better than the Conway Maxwell Poisson regression model.

5. Conclusion

The Zero Inflated Conway Maxwell Poisson regression model obtained on mortality data for Diphtheria patients in 2019 is as follows:

a. Model *count* $\hat{\mu}$

$$\hat{\mu} = \exp\,(7.675 - 0.346X_1 - 0.074X_3 + 0.015X_6\,)$$

b. Model zero inflation for $\hat{\pi}$

$$\hat{\pi} = \frac{exp\,(-1.316 + 0.028X_4\,)}{1 +\ exp\,(-1.316 + 0.028X_4)}$$

The interpretation of the model formed from ZICMP is based on the value of $exp(\beta)$. And here is the best interpretation of the ZICMP regression model:

a. For Model *count* $\hat{\mu}$

Every 1 percent increase in poverty percentage (X_1) it will reduce the average number of cases mortality in diphtheria patients by $exp(0.346) = 1{,}4134$ if other variables are constant. Every 1 percent addition Complete basic immunization coverage (X_3) it will reduce the average number of cases of mortality in diphtheria patients by $exp(0.074) = 1{,}076$ if other variables are constant. Every addition 1 Number of hospitals (X_6) it will increase the average number of cases of mortality of diphtheria patients by $exp(0.015) = 1{,}015$ if other variables are constant.

b. Model zero inflation for $\hat{\pi}$

Every 1 Percentage of Food Management Places (Tpm) Meet Health Requirements (X_4) it will reduce the average number of cases of mortality in diphtheria patients by $\frac{exp(0.028)}{1+\ exp(0.028)} = 0.5069$ if other variables are constant.

References

Department of health. 2018. Indonesian Health Profile. Jakarta : Ministry of Health of the Republic of Indonesia

Department of health. 2019. Indonesia Health Profile. Jakarta : Ministry of Health of the Republic of Indonesia

Huwae IR, Shodikin MA. Clinical appearance of diphtheria in children in Saiful Anwar General Hospital, Malang, East Java, Indonesia. International Journal of Infectious Diseases. 2012;16

Hinde, J., & Demetrio, C. G. B. 2007). Overdispersion: Models and Estimation. Exeter: Department of Laver Building.

Izza, N., and Soenarnatalia. 2015. Analysis of Diphtheria Disease Spatial Data in East Java Province in 2010 and 2011. Health System Research Bulletin.

Ministry of Health of the Republic of Indonesia. 2018. Guidelines for the Prevention and Control of Diphtheria.

Kurniasih, Rini. 2015. Aggregated Claim Model with Erlang Distributed Claim Size and COM-Poisson Distributed Number of Claims. Academia Accelerating The World's Research.

MacGill, Mark. 2018. Everything You Need to Know About Diphtheria. Medical News Today.

Mustika, Diva A., Nooraeni, Rani. 2019. Study of Spatial Effects in Diphtheria Cases Using Geographically Weighted Negative Binomial Regression (Gwnbr) Method. Indonesian Journal of Statistics and Its Applications, Vol 3 No 1 (2019), 91 – 104.

Myers, R.H., Douglas C. Montgomery, G. Geoffrey Vining, & Timothy J, Robinson. 2010. Generalized Linear Models with Applications in Engineering and The Sciences. Second edition. New Jersey: John Wiley and Sons

Nadhiroh, I. M. 2009. Zero Inflated Negative Binomial Models in Small Area Estimation. Bogor: IPB

Data and information Center. 2021. Indonesia Health Profile 2020.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., & Boatwright, P. 2005. A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(1), 127–142. doi:10.1111/j.1467-9876.2005.00474.x

Sellers, Kimberly F., Raim, Andrew. 2016. A flexible Zero Inflated model to address data dispersion. Computational Statistics and Data Analysis 99 (2016) 68–80.

Sim, Shin Z., Gupta, Ramesh C., Ong, Seng H. 2018. Zero Inflated Conway-Maxwell Poisson Distribution to Analyze Discrete Data. The International Journal of Biostatistics. 2018; 20160070

WHO. 2017. Review of the epidemiology of diphtheria 2000-2016. World Health Organization. Geneva.

World Health Organization. 2018. Diphtheria.