PROC. INTERNAT. CONF. SCI. ENGIN.
Volume 4, February 2021
Page 240-254

ISSN 1504607797
E-ISSN 1505707533

Estimation of Zero-Inflated Negative Binomial Regression Parameters Using the Maximum Likelihood Method (Case Study: Factors Affecting Infant Mortality in Wonogiri in 2015)

Moh. Irfan Agus Saputro[1], Mohammad Farhan Qudratullah[2]
[1]Master Student of Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Gadjah Mada University.
Bulaksumur, Caturtunggal, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281, Indonesia. Tel/Fax. (0274)513339.
[2]Mathematics Departement, Faculty of Science and Technology, UIN Sunan Kalijaga
Jl. Marsda Adisucipto No 1 Yogyakarta 55281, Indonesia. Tel. +62-274-540971, Fax. +62-274-519739.
[1]Email: moh.irfan.agus@mail.ugm.ac.id

Abstract. The relationship between the response variable (Y) with one or several predictor variables (X) can be determined by using regression analysis. In a simple linear regression model there is an assumption that the response variable follows the normal distribution, but in reality it is often found that the response variable does not follow the normal distribution. If a response variable has a Poisson distribution then it can be analyzed with a Poisson regression model. There is an assumption that must be fulfilled in poisson regression that is equidispersion (the variance value must be equal to the average value), so this model is not suitable for use in data that is overdispersed (the variance value is greater than the average value). Poisson regression is a general model used to analyze discrete data where discrete data is often found to be of zero value with an excessive proportion of response variables (zero inflation). An alternative model for dealing with overdispersion and zero inflation is the Zero Inflated Negative Binomial (ZINB) model. This research aims to estimate the Zero-Inflated Negative Binomial (ZINB) regression parameters using the maximum likelihood method. The zero inflated negative binomial model (ZINB) was applied to the case of infant mortality in Wonogiri Regency in 2015. The results showed an independent variable that affected the infant mortality rate was the percentage of pregnant women (X_4) with an AIC value of 113.1961.

Keyword: Overdispersion, Poisson Regression, Zero Inflated Negative Binomial (ZINB), Zero Inflation.

Abbreviations: AIC (Akaike Infoemation Criterion), ZIP (Zero Inflated Poisson), ZINB (Zero-Inflated Negative Binomial), ZIGP (Zero Inflated Generalized Poisson), MLE (Maximum Likelihood Estimation), EM (Expectation Maximization).

Running title: Estimation of Zero-Inflated Negative Binomial Regression Parameters Using the Maximum Likelihood Method.

INTRODUCTION

Poisson regression has several conditions that must be met in its analysis, one of which is the state of equidispersion, namely the mean and variance of the response variable is the same or it can be said that the variance value of the response variable given must be equal to the mean value. Sometimes in the analysis of the Poisson regression model there is a violation of these requirements, namely overdispersion or underdispersion in discrete data. (Sekarmini, *et al*., 2013).

When the data with discrete type, the variance value is greater than the mean, it is called overdispersion. Meanwhile, if the mean value is greater than the variant it is called underdispersion. Handling models that can be used to solve the overdispersion problem in discrete-type response data include the Negative Binomial regression model, the Quasi-Likelihood regression model and the Generalized Poisson regression model (Ariawan, *et al*., 2012).

One of the causes of overdispersion is the number of excess zeros in the response variable (excess zeros), although in principle this can be estimated using Poisson regression. So that the handling of models that can be used to solve the overdispersion problem due to excess zeros on the discrete type response data includes the Zero Inflated Poisson (ZIP) regression model, the Zero-Inflated Negative Binomial (ZINB) regression model, the Zero Inflated Generalized Poisson (ZIGP) regression model and the Hurdle regression model (Ariawan, *et al*., 2012).

The method most often used in estimating parameters is the Maximum Likelihood Estomator (MLE) method. The MLE method is the best method for obtaining a point estimate. The role of this method has a good estimation in estimating until a convergent final result is obtained. (Kurniawan, 2017).

According to Sharma and Landge (2013), traffic accidents can use the Zero-Inflated Negative Binomial. The model performance test used is the Akaike Information Criterion (AIC). AIC can be used for model selection between Zero-Inflated Poisson and Zero-Inflated Negative Binomial based on the smallest AIC value. AIC aims to make it easier to determine the best model.

In a countdata modeling (calculated data) there are many zeros in the response variable (zero inflatoin), so the data is handled with Zero-Inflated Poisson (ZIP). However, if there is zero inflation and an overdispersion occurs, then Zero-Inflated Poisson (ZIP) is no longer appropriate. The appropriate model for such data is to use Zero-Inflated Generalized Poisson (Famoye and Singh, 2006). An overdispersion condition can be interpreted as a condition in a Poisson distribution where the variance (variant) is greater than the average (mean). In its development, there is another alternative method for modeling cases with zero observations and

overdispersion, namely the Zero-Inflated Negative Binomial (ZINB) method. The Zero Inflated Negative Binomial (ZINB) regression model is the best model compared to other models. This can be seen from the AIC value of the Zero Inflated Negative Binomial (ZINB) regression is smaller than the AIC value in other regressions.In this paper, the problems discussed are the use of the Zero-Inflated Negative Binomial (ZINB) method to overcome overdispersion in Poisson regression, parameter estimation, model suitability analysis and the significance of the ZINB coefficient. Its application in cases of infant mortality in Wonogiri Regency, Central Java Province.

MATERIALS AND METHODS

The author describes the research methods used in this chapter, starting from the data collection method, the data used in the case study, the analysis path carried out by the author, and the flowchart as a summary of the analysis chart.

1. Research Methods

A step or method or method or procedure used by researchers to obtain information on data is called a research method. In this study, researchers used the literature study method in obtaining relevant references to research on the estimation of the zero-inflated negative binomial (ZINB) parameter with the maximum likelihood method and used the case study method to determine the application of the zero-inflated negative binomial (ZINB) analysis. These references are obtained from books or by utilizing information technology, of course, by choosing reliable and accountable references such as mathematical journals, agencies that provide statistical data, and those related to research.

2. Method of collecting data

The data collection method is a method the author uses to obtain research data. The author in this study obtained data through secondary data, namely data obtained indirectly. Researchers took data from the Wonogiri District Health Profile in 2015. The non-participant observation method was used by the author to find observational data where the author did not participate directly in data collection activities and the author only processed data collection activities and the author only processed and analyzed the available data. The observation units in this study were 25 sub-districts in Wonogiri Regency, but in his observations in one subdistrict there were 1 to 2 health centers. So that the observed data becomes 34 health centers.

3. Research variable

There are two variables that will be used in the study, namely the predictor variable and the response variable. More details for the variables used will be detailed in the table as follows:

Table 3.1 Dependent Variables and Independent Variables.

Variable	Definition	Abbreviation	Information
Y	Infant Mortality Rate.	AKB	The number of infant deaths within one year.
X_1	The percentage of low birth weight Babies.	BBLR	Baby weight less than 2500 grams.
X_2	Percentage of deliveries in an area within a certain period of time assisted by professional health workers.	PSLN	Health workers such as midwifery specialists, general practitioners, midwives, midwife assistants, midwife nurses.
X_3	Health workers such as midwifery specialists, general practitioners, midwives, midwife assistants, midwife nurses.	PHBS	household that behaves clean and healthy.
X_4	Percentage of high risk pregnant Women.	RIST	Risti pregnant women are pregnant women with conditions deviating from normal which directly cause pain and death for both the mother and her baby.

4. Data Processing Tools

The software used in this research is SPSS software and R-Studio software.

5. Research Steps and Flowchart

The steps in the research will be described below as follows:

A. Zero Inflated Negative Binomial (ZINB) Regression Model Estimation

1) Knowing the probability function of the Zero Inflated Negative Binomial (ZINB) regression model.

2) Determine the likelihood function of the Zero Inflated Negative Binomial (ZINB) regression model based on the known probability function.
3) Develop algorithms for the parameter estimation process of the Zero Inflated Negative Binomial (ZINB) regression model based on known likelihood functions. The parameter estimation of the Zero Inflated Negative Binomial (ZINB) regression model was carried out using the Maximum Likelihood Estimation (MLE) method and solved using the Expectation Maximization (EM) method.

B. Application of the Zero Inflation Negative Binomial Regression Model (ZINB).
1) Applying the Zero Inflated Negative Binomial (ZINB) regression model to the case of infant mortality in Wonogiri Regency in 2015 with independent variables are the factors that are considered to influence the infant mortality rate.
2) Checks whether the dependent variable is overdispersed.
3) Checks the proportion of zero values in the dependent variable.
4) Checking multicollinearity on the independent variable using the VIF value.
5) Testing the significance of the regression model parameters. Testing is carried out simultaneously and partially. The test statistic used for the simultaneous test is the G statistic and for the partial test the Z test is used.
6) Interpret the zero-inflation negative binomial (ZINB) regression model that is formed. Zero Inflated Negative Binomial (ZINB) Regression Model Estimation.

RESULTS AND DISCUSSION

a. Binomial Zero-Inflated Negative Regression Model (ZINB).

According to Hilbe (2007) Zero-Inflated Negative Binomial (ZINB) regression is a model formed from the distribution of the poisson gamma mixture. The poisson gamma mixture distribution is formed if a poisson distribution (μ) where μ is the value of a random variable with a gamma distribution, a Poisson gamma mixture distribution is called the negative binomial distribution. According to Hilbe (2011) the opportunity density function is

$$f(y|\alpha,\beta) = \frac{\Gamma(y+a)}{y!\,\Gamma(a)}\left(\frac{1}{1+\beta}\right)^{a}\left(1-\frac{1}{1+\beta}\right)^{y} \quad (4.1)$$

With $y = 0,1,2,3, \ldots$

With mean and variance negative binomial distribution:

$$E[Y] = \alpha\beta \text{ and } V[Y] = \alpha\beta + \alpha\beta^2 \quad (4.2)$$

To form a regression model for the negative binomial distribution, the parameter values of the poisson gamma mixed distribution are expressed in terms of and so that the mean and variance are obtained in the form:

$$E(Y) = \mu \text{ and } V(Y) = \mu + k\mu^2 \quad (4.3)$$

Proof

$$E(Y) = \alpha\beta = \mu$$
$$V(Y) = \alpha\beta + \alpha\beta^2$$
$$= \mu + \mu\beta$$
$$= \mu + \mu\frac{\mu}{\alpha}$$
$$= \mu + \mu^2 k$$

Then the probability mass function becomes:

$$f(y|\alpha,\beta) = \frac{\Gamma\left(y+\frac{1}{k}\right)}{y!\,\Gamma\left(\frac{1}{k}\right)}\left(\frac{1}{1+k\mu}\right)^{\frac{1}{k}}\left(\frac{k\mu}{1+k\mu}\right)^{y} \text{ where } y = 0,1,2,3,\ldots \quad (4.4)$$

The distribution function in the equation (4.4) is called the negative binomial probability density function with the mean $E[Y] = \mu$ and variance $V[Y] = \mu + k\mu^2$, k is called the dispersion parameter. If the k parameter is constant, it can be shown that the negative binomial distribution function in equation (4.4) belongs to the exponential family as follows:

$$f(y|\mu,k) = \exp\left\{y\,In\left(\frac{k\mu}{1+k\mu}\right) + \frac{1}{k}\,In\left(\frac{1}{1+k\mu}\right)\,In\left(\frac{\Gamma\left(y+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)y!}\right)\right\} \quad (4.5)$$

From equation (4.5) in accordance with equation (2.29) it can be concluded that the negative binomial distribution is an exponential family with

$$\theta = In\left(\frac{k\mu}{1+k\mu}\right) b\theta = -\frac{1}{k}In\left(\frac{1}{1+k\mu}\right) c(y) = In\left(\frac{\Gamma(y+1\backslash k)}{\Gamma\left(\frac{1}{k}\right)y!}\right) \quad (4.6)$$

Moment $k \to 0$ then the negative binomial distribution has a variance $V[Y] \to \mu$. This allows the negative binomile distribution to approximate a Poisson distribution assuming the same mean and variance is $E[Y] = V[Y] = \mu$ (Agresti, 2002).

If Y_i is a discrete Independent (bound) random variable (count data) with $i = 1,2,3, \ldots n$, the value of 0 is thought to appear in two states. The first is in a zero state, which

is a situation that occurs with probability pi and results in only observations of value 0. Second is in a negative binomial state, which is a state that occurs with probability (1-pi) and has a negative binomial distribution (distribution) with the mean μ, with $0 \leq pi \leq 1$. The process of these two states with the Y_i variable gives a two-component mixed distribution and the probability function is obtained as follows:

$$P = (Y_i = y_i) = \begin{cases} p_i + (1\text{-}p_i)\left(\frac{1}{1+k\mu i}\right)^{\frac{1}{k}}, \text{for } y_i = 0 \\ (1\text{-}p_i)\frac{\Gamma\left(y_i+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}\left(\frac{1}{1+k\mu i}\right)^{\frac{1}{k}}\left(\frac{k\mu i}{1+k\mu i}\right)^{yi} \text{ for } yi > 0 \end{cases} \quad (4.7)$$

With $i = 1,2,3,\dots,n; 0 \leq pi \leq 1, \mu_i \geq 0, k$ is the parameter spread by $\frac{1}{k} > 0$ and Γ(.) is gamma function. The means and their variances are defined $E(Y_i) = (1\text{-}pi)\mu_i$ and $Var(Y_i) = (1\text{-}pi)\mu_i(1 + \mu_i k + p_i\mu_i)$. It is assumed that the parameters μ_i and p_i each depend on the x_i and z_i variables, so according to Garay and Hashimoto (2011) the model from ZINB regression is divided into two model components, namely:

a. Discrete data model for μ_i is

$$In(\mu_i) = x_i^T\beta, \quad \mu_i \geq 0, i = 1,2,3,\dots,n \quad (4.8)$$

$x_i$is a variable matrix containing different sets of experimental factors related to the probability of the mean Negative Binomial at the Negative Binomial State.

b. Zero-inflated model for p_i is

$$\text{logit}(p_i) = In\left(\frac{p_i}{1\text{-}p_i}\right) = Z_i^T\gamma, \quad 0 \leq p_i \leq 1, i = 1,2,3,\dots,n \quad (4.9)$$

with:

p : number of predictor variables

n : number of observations

β : ZINB regression model's estimated parameters

γ : ZINB regression model's estimated parameters

$z_i$is a variable matrix containing different sets of associated experimental factors and zero state.

The effect of each covariate matrix x_i and z_i to μ_i and p_i bias equal or not equal to μ_i and p_i then the matrix $x_i = z_i$, so that models (4.8) and (4.9) become:

1. Discrete data model for μ_i is

$$In(\mu_i) = x_i^T\beta, \text{or } \mu_i = e^{x_i^T\beta} \quad \mu_i \geq 1, i = 1,2,3,\dots \quad (4.10)$$

2. Zero-inflated model for p_i is

$$\text{logit}(p_i) = In\left(\frac{p_i}{1\text{-}p_i}\right) = x_i^T\gamma, \text{or } p_i = \left(\frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}\right) 0 \leq p_i \leq 1, i = 1,2,3,\dots,n \quad (4.11)$$

$z_i$is a variable matrix that contains different sets of experimental factors related to the probability of the Negative Binomial Mean at the Negative Binomial State, while β and γ are the regression parameters to be estimated.

A. Estimation of Binomial Negative Zero-Inflated Regression Parameters (ZINB)

In this paper, the parameter estimation used for the Zero-Inflated Negative Binomial (ZINB) regression is the Maximum Likelihood Estimation (MLE) method with the EM (Expectation Maximizationi) and Newton Rhapson Algorithm procedures. This method is generally used to estimate the parameters of a model with known density functions.

From equation (4.10) and (4.11) it is obtained:

$$In(\mu_i) = x_i^T\beta$$

$$\mu_i = e^{x_i^T\beta}$$

$$In\left(\frac{p_i}{1\text{-}p_i}\right) = x_i^\gamma\gamma \quad (4.12)$$

$$p_i = \frac{e^{x_i^\gamma\gamma}}{1+e^{x_i^\gamma\gamma}} \quad (4.13)$$

$$(1\text{-}p_i) = \frac{1}{1+e^{x_i^\gamma\gamma}} \quad (4.14)$$

From equation (4.12), (4.13) and (4.14) are substituted for equation (4.7), it is obtained:

$$p(Y_i = y_i) = \begin{cases} \frac{e^{x_i^\gamma\gamma}}{1+e^{x_i^\gamma\gamma}} - \frac{1}{1+e^{x_i^\gamma\gamma}}\left(\frac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}, \text{for } y_i = 0 \\ \frac{1}{1+e^{x_i^\gamma\gamma}}\frac{\Gamma\left(y_i+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}\left(\frac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}\left(\frac{ke^{x_i^T\beta}}{1+ke^{x_i^T\beta}}\right)^{y_i}, \text{for } y_i = 1,2,3,\dots \end{cases} \quad (4.15)$$

Suppose a sample is taken $(y_1, x_1, z_1), \dots, (y_n, x_n, z_n)$ of n independent experiments. The likelihood function of equation (4.15) for parameters $\theta = \left(\frac{1}{k}, \beta^T, \gamma^T\right)^T$ is

$$L(\theta|y_i) = \begin{cases} \Pi \dfrac{e x_i^T\gamma + \left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}}{1+e^{x_i^T\gamma}}, \text{for } y_i = 0 \\ \Pi \dfrac{\Gamma\left(y_i + {}^1/_k\right)\left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}\left(\dfrac{ke^{x_i^T\beta}}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}}{1+e^{x_i^T\gamma}\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}, \text{for } y_i = 1,2,3,\ldots \end{cases} \quad (4.16)$$

So the log-likelihood function of equation (4.16) is:

$$\text{In } L(\theta|y_i) = \begin{cases} \sum_{i=1}^{n} \text{In}\left\{e^{x_T^i\gamma} + \left(\dfrac{1}{1+ke^{x_T^i\beta}}\right)^{\frac{1}{k}}\right\} - \sum_{i=1}^{n} \text{In}\left[1+e^{x_T^i\gamma}\right], \text{for } y_i = 0 \\ -\sum_{i=1}^{n} \text{In}\left[1+e^{x_T^i\gamma}\right] + \sum_{i=1}^{n} \text{In}\left[\Gamma\left(\frac{1}{k}+y_i\right)\right] - \sum_{i=1}^{n} \text{In}[\Gamma(y_i+1)] - \sum_{i=1}^{n} \text{In}\left[\Gamma\left(\frac{i}{k}\right)\right] \\ +y_i \sum_{i=1}^{n} \text{In}\left\{\left(\dfrac{e^{x_T^i\gamma}}{1+e^{x_T^i\gamma}}\right)^{y_i}\right\} - \left(\frac{1}{k}\sum_{i=1}^{n} \text{In}\left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}\right), \text{for } y_i = 1,2,3,\ldots \end{cases} \quad (4.17)$$

With $= 1,2,3,\ldots,n$. The estimate with the maximum likelihood ratio is calculated by maximizing its log-likelihood in equation (4.17). According to Ariawan, *et al*. (2012), the summation of the log-likelihood function in equation (4.17) is not linear, so that this likelihood function cannot be solved by ordinary numerical methods. So that the EM (Expectation Maximization) algorithm is used, which is one of the methods used to find the estimate of a parameter through the Maximum Likelihood Estimation (MLE) framework.

For example variables $y_i (i = 1,2,3,\ldots,n)$ relates to the indicator variable vector $W = (w_1, w_2, \ldots w_n)^T$ is:

$$wi = \begin{cases} 1, \text{if } y_i \text{ comes from zero state.} \\ 0, \text{if } y_i \text{ comes from the Negative Binomial state.} \end{cases}$$

With $= 1,2,3,\ldots,n$, if the value of the response variable $y_i = 1,2,3,\ldots$then the value $wi = 0$. Remember if the value of the response variable $y_i = 0$, then the value w_i maybe 0 might be 1. Hence, the value w_i considered lost. Opportunity from w_i can be stated:
$P(w_i = 1) = p_i$
$P(w_i = 0) = 1\text{-}P_I$

With $i = 1,2,3,\ldots,n$ so the distribution of the variable W is $w_i \sim$ *Binomial* $(1, p_i)$ have mean and variance $E(w_i) = p_i$ and $var(w_i) = p_i\,(1\text{-}p_i)$. The combined distribution between y_i dan w_i which is formed, namely:

$$f(w_i, y_i|p_i, \mu_i) = (p_i)^{w_i}(1\text{-}p_i)^{(1\text{-}w^i)}\left[\frac{\Gamma\left(y_i+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}\left(\frac{1}{1+k\mu_i}\right)^{\frac{1}{k}}\left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i}\right]^{1\text{-}w_i} \quad (4.18)$$

Substituting equations (4.12), (4.13), and (4.14) into equation (4.18) we get the log-likelihood equation:

$$\text{In } L(\beta, \gamma|y_i, w_i) = \sum_{i=1}^{n}\left\{w_i x_i^t\gamma\text{-In}\,[1+\exp(x_i^t\gamma)] + (1\text{-}w_i)\text{In}\left[g\left(y_i;\beta,\frac{1}{k}\right)\right]\right\} \quad (4.19)$$

Where

$g\left(y_i;\beta,\frac{1}{k}\right) = \frac{\Gamma\left(y_i+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}\left(\frac{1}{1+k\mu_i}\right)^{\frac{1}{k}}\left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i}$, and $\mu_i = e^{x_i^t\beta}$

with $i = 1,2,3,\ldots,n$. Equation (4.18) which will be maximized using the EM algorithm, with parameters β and γ can be estimated separately to be:
$\text{InL}\,(\beta, \gamma|y_i, w_i) = \text{In } L(\beta|y_i, w_i) + \ln L(\gamma|y_i, w_i)$
with:

$$\text{InL}(\gamma|y_i, w_i) = \sum_{i=1}^{n}\left[w_i x_i^T\gamma\text{-In}\left[1+\exp\left(x_i^T\gamma\right)\right]\right] \quad (4.20)$$

and

$$\text{In } L(\beta|y_i, w_i) = \sum_{i=1}^{n} (1\text{-}w_i)\left\{\frac{\Gamma\left(y_i + \frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}\left(\frac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}\left(\frac{ke^{x_i^T\beta}}{1+ke^{x_i^T\beta}}\right)^{y_i}\right\} \quad (4.21)$$

The EM algorithm is divided into two steps, namely:

1. Expectation Stage (E-step)

Overrides variables w_i with $w_i^{(m)}$ which is the expectation of w_i

$$w_i^{(m)} = E\left(w_i|y_i,\gamma^{(m)},\beta^{(m)}\right) = \begin{cases}\left(1+\left(e^{x_i^t\gamma^{(m)}}\right)\left[\frac{1}{1+k^{(m)}e^{x_i^t\beta^{(m)}}}\right]^{\frac{1}{k^{(m)}}}\right)^{-1}, & \text{jika } y_i = 0\\ 0, \text{jika } y_i = 1,2,3,\ldots\end{cases}$$

So that

$$Q\left(\beta,\gamma;\beta^{(m)},\gamma^{(m)}\right) = E_{\theta^{(k)}}\left\{\text{InL}(\beta,\gamma|y_i,w_i)|y_i,\beta^{(m)},\gamma^{(m)}\right\} = \sum_i^n \text{InL}\left(\gamma^{(m)}|y_i,w_i^{(m)}\right) + \sum_i^n \text{InL}\left(\beta^{(m)}|y_i,w_i^{(m)}\right)$$

Where equations (4.19) and (4.20) become

$$\text{InL}\left(\gamma^{(m)}|y_i,w_i^{(m)}\right) = \sum_{i=1}^{n}\left[w_i^{(m)}x_i^t\gamma\text{-In}\left(1+e^{x_i^t\gamma}\right)\right] \quad (4.22)$$

$$\text{InL}\left(\gamma^{(m)}|y_i,w_i^{(m)}\right) = \sum_{i=1}^{n}\left(1\text{-}w_i^{(m)}\right)\left\{\frac{\Gamma\left(\frac{1}{k}+y_i\right)}{\Gamma(y_i+1)\Gamma\left(\frac{1}{k}\right)}\left(\frac{e^{x_i^t\gamma}}{1+e^{x_i^t\gamma}}\right)^{y_i}\left(\frac{1}{1+ke^{x_i^t\beta}}\right)^{\frac{1}{k}}\right. \quad (4.23)$$

2. Maximization stage (M-step)

Maximizing β and γ in equations (4.22) and (4.23) by calculating $\beta^{(m+1)}$ and $\gamma^{(m+1)}$ with the Newton Raphson method. (Taufan, *et al*, 2012), as for the steps as follows:

a. Suppose $\beta^{(m)}$ and $\gamma^{(m)}$ is the maximum likelihood method of estimating $\hat{\beta}$ and $\hat{\gamma}$
b. Calculate $\beta^{(m+1)}$ and $\gamma^{(m+1)}$ in a way:
   $\beta^{(m+1)} = \beta^{(m)}\text{-}\left(H^{(m)}\right)^{-1}U^{(m)}$
   and
   $\gamma^{(m+1)} = \gamma^{(m)}\text{-}\left(H^{(m)}\right)^{-1}U^{(m)}$
   Where H is the second derivative of $\text{In } L\left(\beta^{(m)}|y_i,w_i^{(m)}\right.$ and $\text{In } L\left(\gamma^{(m)}|y_i,w_i^{(m)}\right)$, U is the first derivative of $\text{In } L\left(\beta^{(m)}|y_i,w_i^{(m)}\right)$ and $\text{In } L\left(\gamma^{(m)}|y_i,w_i^{(m)}\right.$).
c. Replace $\beta^{(m)}$and $\gamma^{(m)}$with β^{m+1} and γ^{m+1}in the next iteration, then return to the expectation stage (E-Step).

The E-Step and M-Step stages were repeated until a convergent parameter assessment was obtained $\left|\beta^{(m)}\text{-}\beta^{(m+1)}\right| \le \varepsilon$ and $\left|\gamma^{(m)}\text{-}\gamma^{(m+1)}\right| \le \varepsilon$ usually ε is a very small positive number, eg $\varepsilon = 10^{-5}$.

B. ZINB Regression Parameter Testing

1. ZINB Regression Model Suitability Testing.

According to Zamzami & Ismail (2013) testing the suitability of the ZINB regression model using the G test.The function of this test is to test the role of the independent variable simultaneously (simultaneously) with the following test procedures:

Hypothesis:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = \gamma_1 = \gamma_2 = \cdots = \gamma_p = 0$

(Independent variables together have no influence on the dependent variable)

H_1: there is at least one $\beta_j \neq 0$ or $\gamma_j \neq 0$, with $j = 1,2,3,\ldots,p$

(the independent variables together have an influence on the dependent variable)

With β_j is the j-th parameter of the model $\text{In}(\mu_i) = x_i^T\beta$ with $i = 1,2,\ldots n$. γ_j with is the j-th parameter of the model $\text{logit}(p_i) = \text{In}\left(\frac{p_i}{1\text{-}p_i}\right) = x_i^T\gamma$ with $i = 1,2,3,\ldots,n$.

Test statistics:

$$G = \text{-2 In}\left[\frac{L_0}{L_1}\right] = \text{-2(In } L_0\text{-In}(L_1)$$

$$= \text{-2}\{[\text{In } L(\beta_0|y_i,w_i) + \text{In } L(\gamma_0|y_i,w_i)]\text{-}[\text{In } L(\beta|y_i,w_i) + \text{In } L(\gamma|y_i,w_i)]\}$$

$G \sim X_p^2$

With

L_0 is the likelihood without the independent variable.

L_1 is the likelihood with the independent variable
Test criteria:

If H_0 true, the G test statistic will follow the spread X^2 with degrees of freedom p and H_0 will be rejected H_0 at the significance level α if the value $G_{hitung} > X^2_{\alpha;2p}$ (Hosmer and Lemeshow, 1989).

2. Parameter Significance Testing.

According to Myers (2010) significance testing is divided into 2 models, namely:

1. Parameter Signification Test β.

   Testing the model parameter significance $In(\mu_i) = x_i^T\beta$ with $i = 1,2,3,\dots,n$ for the ZINB model. Hypothesis:

   $H_0: \beta_j \neq 0$

   (There is no significant effect between the independent variables on the dependent variable)

   $H_1: \beta_j \neq 0$

   (there is a significant influence between the independent variables on the dependent variable)

   For each $j = 1,2,\dots,p$

   Test statistics:

$$W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2$$

   Test criteria:

   Refuse H_0 at the level of significance α if $W_j > X^2_{\alpha:1}$ at the alpha level. Rejection H_0 indicates that explanatory X_j has an influence on the response variable Y at the level of significance.

2. Parameter Significance Testing γ.

   Testing the model parameter significance $In\left(\frac{p_i}{1-p_i}\right) = X_i^T\gamma$ with $i = 1,2,3,\dots,n$ for ZINB. Hypothesis:

   $H_0: \gamma_j = 0$

   (There is no significant effect between the independent variables on the dependent variable)

   $H_0: \gamma_j \neq 0$

   (there is a significant influence between the independent variables on the dependent variable)

   For each $j = 1,2,3,\dots,p$

   Test statistics:

$$W_j = \left(\frac{\gamma_j}{SE(\gamma_j)}\right)^2$$

   $W_j \sim X^2_1$

   Test criteria:

Refuse H_0 at the level of significance α if $W_j > X^2_{\alpha:1}$ at the alpha level. Rejection H_0 at indicates that explanatory X_j has an influence on the response variable Y at the level of significance.

C. Model Feasibility Test

According to Sharma and Landge (2013) Akaike Infoemation Criterion (AIC) is used to assess model performance. AIC is defined by

$$AIC = -2 \ln L(\hat{\theta}) + 2k$$

With $L(\hat{\theta})$ is the likelihood value and k is the number of parameters. The best model is to choose a model that has the smallest AIC value.

Case Study

In the case study chapter, the implementation of the Zero-Inflated Negative Binomial (ZINB) model will be discussed. The case study will be applied to infant mortality cases in Wonogiri Regency in 2015. This study will examine the relationship between the number of infant deaths (Y) with a low percentage of birth weight babies (X_1), percentage of deliveries in an area within a certain period of time assisted by a health professional (X_2), the percentage of households that have a clean and healthy lifestyle (X_3) and the proportion of high-risk pregnant women (X_4).

1. Characteristics of Infant Mortality Data in Wonogiri District.

Wonogiri Regency has an area of 182,236.02 Ha or 5.59% of the area of Central Java Province. Geographically it is located at latitude $7^0 32' - 8^0 15'$ to longitude $110^0 41' - 111^0 18'$. Wonogiri Regency is bordered by Karanganyar and Sukoharjo Regencies in the north, bordering Karanganyar Regency and Ponorogo Regency in the east, bordering Pacitan Regency and Indonesian Ocean in the South and bordering the Special Region of Yogyakarta and Klaten Regency in the west. Wonogiri Regency is divided into 25 districts with 251 villages and 43 sub-districts and 2,306 hamlets.

Infant mortality cases in Wonogiri Regency can be viewed from the percentage of infant mortality rates in Wonogiri Regency. In the following graph will be shown the presentation of infant mortality rates in Wonogiri Regency in 2011-2015 which originated from the Wonogiri District Health Profile data.

Trend of infant mortality rate (IMR)

Figure 5.1 Graph of Trend in Infant Mortality Rate. Source: District Health Profile Data. Wonogiri, 2011, 2012, 2013, 2014, and the District Health Office Health Effort Report. Wonogiri in 2015.

Based on the graph above, it can be observed that from 2011 to 2012 there has been a decrease in the percentage of infant mortality, but if we observe the following year, it can be seen that in 2013 the percentage of infant mortality rates has increased and is greater than in 2011. Although this has happened There was a decline in 2014, but it can be seen that the percentage of infant mortality rates has increased from 2014 to 2015 and in 2015 experienced the highest percentage of infant mortality rates since 2011.

2. **Descriptive Analysis of Research Variables.**

The dependent variable (dependent) in this study is the Infant Mortality Rate (Y) and the independent variable used is:

- BBLR X_1: The percentage of low birth weight, namely body weight babies less than 2500 grams.
- PSLN X_2: Percentage of deliveries in an area within a certain period of time assisted by professional health workers, such as: midwifery specialist, general practitioner, midwife, midwife assistant, midwife nurse.
- PHBS X_3: Percentage of households that have a clean and healthy lifestyle
- RIST X_4: Percentage of risti (high risk) pregnant women where risti pregnant women are pregnant women with a condition deviating from normal which directly causes pain and death for both the mother and her baby.

Percentage of risti (high risk) pregnant women where risti pregnant women are pregnant women with a condition deviating from normal which directly causes pain and death for both the mother and her baby:

Table 5.1 Descriptive Analysis of Research Data for Variable Y and Variable X.

No.	Variabel	N	Minimum	Maksimum	Rata-rata	Standard Deviasi
1	AKB (Y)	34	0	11	2,76	3,46
2	BBLR (X1)	34	1,16	14,88	5,38	2,848
3	PSLN (X2)	34	74,62	111, 28	89,44	6,69
4	PHBS (X3)	34	61,20	100	90,50	11,02
5	RIST (X4)	34	20,60	216,90	67,57	36,55

More clearly, the infant mortality rate will be presented in the following graphical form:

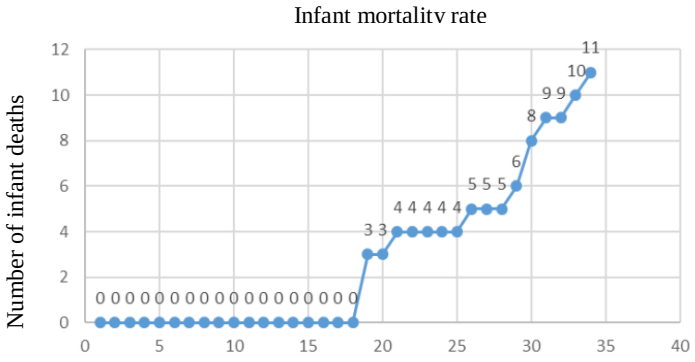

Figure 5.2 Graph of Infant Mortality Rate in 2015.

In the descriptive analysis it can be seen that the average number of infant deaths (JKB) in Wonogiri Regency in 2015 was 2.76 cases. The highest number of cases was 11 babies that occurred in Wonogiri Subdistrict, to be precise at Wonogiri I Health Center and the lowest number of cases was the number of cases of infant deaths, consisting of 18 public health center.

The following is a graph of the percentage of low birth weight, which is the baby's weight less than 2500 grams.

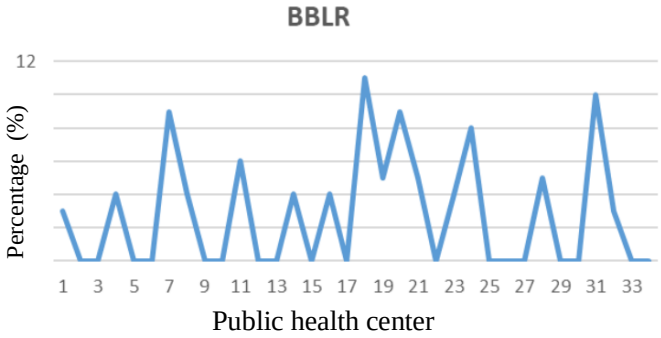

Figure 5.3 Graph of BBLR.

The average percentage of BBLR in Wonogiri Regency is 5.38 with the highest percentage in Purwantoro District at Purwantoro II Health Center and the lowest percentage is 1.16 in Jatisrono District, precisely at Jatisrono II Health Center.

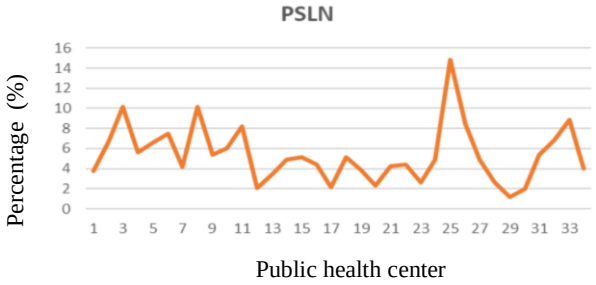

Figure 5.4 PSLN Graph.

The average percentage of PSLN in Wonogiri Regency is 89.44 with the highest percentage located in Tirtomoyo District with a percentage of 111.28 and the lowest is located in Baturetno District at Baturetno II Community Health Center with a percentage of 74.62.

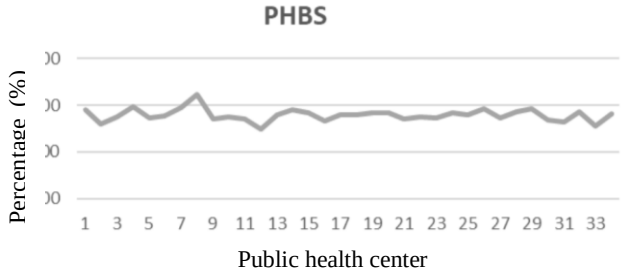

Figure 5.5 PHBS Graph.

The average percentage of PHBS in Wonogiri Regency is 90.5 with the highest percentage located in 9 sub-districts including Nguntoronadi District at Nguntoronadi II Health Center, Eromoko District at Eromoko II Health Center, Selogiri District, Sidoharjo District, Purwantoro District at Purwantoro I Health Center, Slogohimo District and Jatisrono Subdistrict is in Jatisrono II Puskesmas, Girimarto District and Karangtengah District. While the lowest percentage with a percentage of 61.2 is located in Batuwarno sub-district.

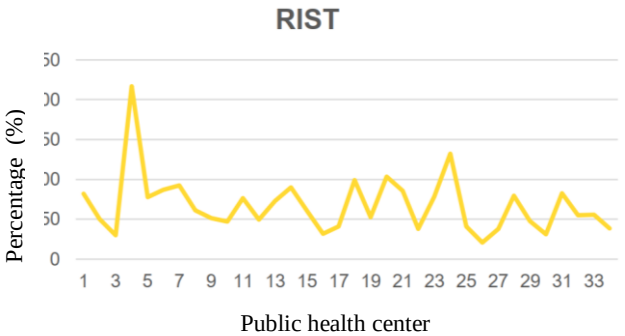

Figure 5.6 RIST Graph.

The percentage of RIST in Wonogiri Regency is 67.57 with the highest percentage of 216.90 located in Giriwoyo sub-district to be precise at Giriwoyo 1 puskesmas and the lowest percentage is located in Kismantoro sub-district with a percentage of 20.60.

3. Assumption Test.

a. Overdispersion Test.

The overdispersion test is a test to determine whether the response variable has overdispersion or not. The overdispersion test will use the AER package from the R-Studio software. This overdispersion test will use a significance level of 5% or 0.05.

Hypothesis testing is as follows:

H_0: The dependent variable does not experience overdispersion

H_1: The dependent variable has overdispersion

If the value is sig. (p-value) is smaller than α, with α valued at 0.05, then H_0 is rejected. Besides using the sig value. The response variable can be said to have overdispersion if the dispersion value is more than 0.

Table 5.2 Results of the overdispersion test using the R-Studio software.

p-value **(nilai sig.)**	**Dispersi Value**	**Information**
0,041	1,19	H_0 rejected

Based on Table 5.2, it is known that the p-value is 0.041 which means that it is smaller than α of 0.05, so H_0 is rejected, so it can be concluded that the response variable has overdispersion. This is also confirmed by the dispersion value (∅) of 1.19 which is greater than 0, this also indicates that the response variable has overdispersion.

b. Zero Inflation Check for Response Variables.

Zero inflation checks are carried out by calculating the percentage of observations that are zero on the response variable. The results of the zero inflation examination on the response variable are presented as follows:

Table 5.3 Zero Inflation examination results on the Response Variable.

Total Infant Mortality Rate	**Frequency of Total Infant Mortality Rate**	**Percentage**	**Cumulative percentage**
0	18	52,9	26,5
3	2	5,9	58,8
4	5	14,7	73,5
5	3	8,8	82,4
6	1	2,9	85,3
8	1	2,9	88,2
9	2	5,9	94,1
10	1	2,9	97,1
11	1	2,9	100

For more details, a bar chart will be displayed that explains the frequency of the number of infant mortality rates.

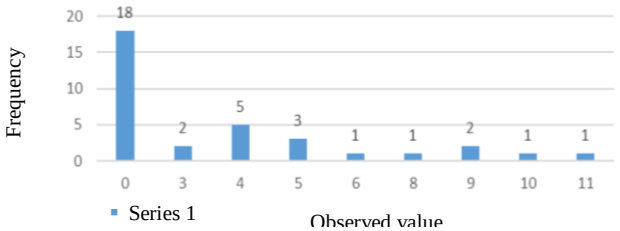

Figure 5.7 Frequency Graph of Infant Mortality Rate

According to Asuti, et al. (2015) zero inflation occurs when observations are zero more than 50%. Based on the diagram above shows that the very high frequency of observations that are zero is 18 times. Then the results of the zero-inflation examination on the response variable based on Table 5.3 show thatthere is zero inflation in the variable because the percentage of observations is zero more than 50%, namely 52.9% with a frequency of 18.

c. Multicollinearity Examination.

Multicollinearity is a condition that shows a strong correlation (correlation) between two or more independent variables in a regression model. If multicollinearity occurs, then a variable that has a strong correlation with other variables in the model, the power of the prediction is not reliable and stable.

Table 5.4 VIF Value of Predictor Variables.

Predictor Variable	VIF
x_1	1.288
x_2	1.177
x_3	1.323
x_4	1.199

In Table 5.4, the VIF value of each predictor is presented. In multicollinearity testing using a VIF value of more than 10. Because the VIF value of each predictor variable is presented in the table below 10, it is concluded that there is no multicollinearity case between variables. So that all predictor variables meet the non-multicollinearity assumption.

4. Early Stage Zero-Inflated Negative Binomial Regression Modeling.

Binomial Zero-Inflated Negative Regression is a regression applied to overdispersed data. The data used for the application of Zero-Inflated Negative Binomial Regression is the number of infant deaths in WonogiriRegency in 2015. In accordance with the results of the tests that have been done, it is found that the data is overdispersed so that the assumption to do modeling with binomial negative zero-inflated regression (ZINB) has been fulfilled.

The Zero Inflation Negative Binomial (ZINB) regression model is applied to cases of infant mortality in Wonogiri Regency. Modeling cases of infant mortality using three variables used, namely percentage of LBW (x_i), percentage of PSLN (x_2), percentage of PHBS (x_3) and percentage of RIST (x_4). So that the Zero Inflation Negative Binomial (ZINB) regression model is formed, namely

- Discrete data model for $\hat{\mu}_i$

$$\hat{\mu}_i = \exp(\beta_0\text{-}\beta_1 X_i\text{-}\beta_2 X_2\text{-}\beta_3 X_3 + \beta_4 X_4)$$

- Zero inflation model for $\hat{p}_i$

$$\hat{p}_i = \exp\left(\frac{\gamma_0 + \gamma_1 X_1\text{-}\gamma_2 X_2\text{-}\gamma_3 X_3\text{-}\gamma_4 X_4}{1 + \gamma_0 + \gamma_1 X_1\text{-}\gamma_2 X_2\text{-}\gamma_3 X_3\text{-}\gamma_4 X_4}\right)$$

a. Estimation Results of ZINB Regression Model Parameters.

Zero inflation negative binomial regression model is estimated using the maximum likelihood method (maximum likelihood estimation). Then to find out the significance level of the parameter estimation results in the ZINB regression model by testing the significance of the ZINB parameter estimation results using the R-studio software with the pscl package so as to produce the following parameter estimates:

Table 5.5 Estimation Results for ZINB Parameters.

Parameters	Estimation
$\hat{\beta}_0$	3,66
$\hat{\beta}_1$	-0,0042
$\hat{\beta}_2$	-0,021
$\hat{\beta}_3$	-0,0017
$\hat{\beta}_4$	0,0026
$\hat{\gamma}_0$	20,40
$\hat{\gamma}_1$	0,008
$\hat{\gamma}_2$	-0,092
$\hat{\gamma}_3$	-0,073
$\hat{\gamma}_4$	-0,083

Simultaneous and partial testing. According to Hosmer and Lameshow (2000), testing the significance of the parameter estimation results in the ZINB regression model simultaneously uses the G test and partial significance testing uses the Z test statistic. Testing the significance of the parameter estimation results in the Zero-Inflation Negative Binomial (ZINB) model simultaneously (simultaneously) using the R-Studio software with the lmtest package with the syntax lrtest () while the partial test (individually) will also be tested with the RStudio software but using the pscl package with the zeroinfl syntax () (more clearly in the attachment). Simultaneous and partial testing as follows:

b. Simultaneous Parameter Testing (Simultaneously).

Testing the suitability of this model is testing the parameters simultaneously (simultaneously). Testing the significance of the parameter estimation results in the Zero-Inflation Negative Binomial (ZINB) model simultaneously (simultaneously) using the R-Studio software with the lmtest package with the syntax lrtest (). This hypothesis can be seen from the G statistical

value with the following hypothesis:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$

(Independent variables together have no influence on the dependent variable)

H_1: There is at least one $\beta_j \neq 0$ or $\gamma_j \neq 0$ with $j = 1,2,3,\dots,6$

(The independent variables together have an influence on the dependent variable)

The test was carried out with the α significance level of 5% or 0.05. If the p-avlue or Pr (> chisq) value is smaller than 0.05, H_0 is rejected. The results of simultaneous hypothesis testing will be explained in the table as follows:

Table 5.6 Simultaneous ZINB Regression Parameter Testing.

Log-Likelihood	Df	*chi-square* value	*Pr(>chisq)*	Information
-89,98	12,93		0,011	H_0 rejected

Based on Table 5.5, it shows if the Pr value (> chisq) is 0.011, which is less than the α value of 0.05. So it can be concluded that H_0 is rejected. This shows that simultaneously the variables X_1, X_2, X_3 and X_4 have a significant effect on the response variable Y.

c. Partial Parameter Testing (Individual).

The parameter significance test is a partial (individual) test which is used to look for predictor variables that have a significant effect on infant mortality in Wonogiri Regency in 2015. The significance test will use the R-Studio software using the pscl package with the syntax zeroinfl (). Testing the significance of this parameter is divided into two tests, namely:

1. Parameter β significance testing.

This test uses a model $In(\mu_i) = x_i^T\beta$ with $i = 1,2,3,\dots,n$. This test uses the following hypothesis:

- $H_0: \beta_j \neq 0$ for each $j = 1,2,\dots,p$

(There is no significant effect between the independent variables on the dependent variable)

- $H_1: \beta_j \neq 0$ for each $j = 1,2,\dots,p$

(There is a significant influence between the independent variables on the dependent variable)

Testing the significance of this parameter uses the α significance level of 5% with the test criteria that is rejecting H_0 if the p-value (Pr> | Z |) is less than the α significance level of 0.05.

2. Parameter γ significance testing.

Testing the parameter significance using the model $In\left(\frac{p_i}{1-p_i}\right) = X_i^T\gamma$ with $i = 1,2,3,\dots,n$. This test uses the following hypothesis:

- $H_0: \gamma_j \neq 0$ for each $j = 1,2,\dots,p$

(There is no significant effect between the independent variables on the dependent variable)

- $H_1: \gamma_j \neq 0$ for each $j = 1,2,\dots,p$

(There is a significant influence between the independent variables on the dependent variable)

Testing the significance of this parameter uses the α significance level of 5% with the test criteria that is rejecting H_0 if the p-value (Pr> | Z |) is less than the α significance level of 0.05. The results of the significance of the ZINB model parameters in infant mortality cases in Wonogiri Regency in 2015 are presented in the following table:

Table 5.7 The results of the significance test for the ZINB parameter estimate.

Parameter	Estimate	SE (Standard Error)	Z test	*(Pr>|Z|)*
$\hat{\beta}_0$	3,66	1,91	1,91	0,056
$\hat{\beta}_1$	-0,0042	0,06	-0,07	0,94
$\hat{\beta}_2$	-0,021	0,018	-0,017	0,24
$\hat{\beta}_3$	-0,0017	0,014	-0,129	0,90
$\hat{\beta}_4$	0,0026	0,0025	1,026	0,30
$\hat{\gamma}_0$	20,40	9,63	2,12	0,034*
$\hat{\gamma}_1$	0,008	0,23	0,04	0,97
$\hat{\gamma}_2$	-0,092	0,085	-1,090	0,27
$\hat{\gamma}_3$	-0,073	0,053	-1,37	0,17
$\hat{\gamma}_4$	-0,083	0,031	-2,67	0,0075*

C Value: 119,107

Based on Table 5.7, as a result of testing the significance of ZINB parameters partially, it is known that the predictor variable in the discrete parameter data estimation $\hat{\mu}_i$ has no significant value because the p-value (Pr> | Z |) is greater than 0.05. As for the predictor variable on the parameter estimate of zero inflation p_i, there is one variable that has a significant effect, namely the variable percentage of pregnant women with a p-value of 0.0075. So that the Zero Inflation Negative Binomial (ZINB) regression model equation is formed which is partially tested, namely: Discrete data model for $(\hat{\mu}_i)$.

On the data that has been processed and based on Table 5.7 it is known that for the predictor variables in the estimation of the discrete data parameter $\hat{\mu}_i$, none of the parameters are significant to the dependent variable. So that in the partial test for discrete data the model cannot be formed.

- Zero inflation model for $\hat{p}_i$

As for the predictor variable on the parameter estimate of zero inflation $\hat{p}_i$, there is one variable that has a significant effect, namely the variable percentage of pregnant women with a p-value of 0.0075.

$$\hat{p}_i = \exp\left(\frac{20{,}40\text{-}0{,}083X_4}{1 + 20{,}40 + \text{-}0{,}083X_4}\right)$$

With AIC value 113,1961.

5. Advanced Binomial Zero-Inflated Negative Regression Modeling.

Based on Table 5.7, it is known that there are several parameters that are not significant. So that a further test will be carried out to test the significance of the parameter estimation results in the Zero-Inflation Negative Binomial (ZINB) model but by removing some of the parameters that have the largest p-value so that a model with significant parameters can be obtained with the smallest AIC value. So by paying attention to the results of the p-value in each parameter in Table 5.7 it is known that the highest p-value lies in the variable X_1 parameter $\hat{\beta}_1$ which is 0.94 so that the significance test of the parameter estimation results in the Zero-Inflation Negative Binomial (ZINB) model will be carried out. By setting aside the predictor variable X_1 so that the following results are obtained.

Table 5.8 Testing ZINB Regression Parameters Simultaneously Without X_1.

Log-Likelihood	Df	*chi-square value*	*Pr(>chisq)*	Information
-60,194	-6	23,375	0,00071	H_0 rejected

Based on Table 5.8, it is shown that the Pr value (> chisq) is 0.00071 which is less than the α value of 0.05. So it can be concluded that H_0 is rejected. This shows that simultaneously the variables X_2, X_3 and X_4 have a significant effect on the response variable Y.

Table 5.9 ZINB Parameter Significance Test Partially Without Variables X_1.

Parameter	Estimation	SE *(Standard Error)*	Z test	*(Pr>|Z|)*
$\hat{\beta}_0$	3,68	1,86	1,98	0,0475*
$\hat{\beta}_2$	-0,022	0,017	-1,248	0,212
$\hat{\beta}_3$	-0,0019	0,013	-0,145	0,88
$\hat{\beta}_4$	0,0027	0,0026	1,041	0,30
$\hat{\gamma}_0$	20,47	9,38	2,18	0,029*
$\hat{\gamma}_2$	-0,093	0,085	-1,085	0,28
$\hat{\gamma}_3$	-0,074	0,053	-1,406	0,16
$\hat{\gamma}_4$	-0,083	0,031	-2,68	0,00743 *

AIC Value: 115,1136

Based on Table 5.9, the results obtained from testing the parameter estimates in the Zero-Inflation Negative Binomial (ZINB) model by removing the X_1 variable. The significant parameter is at $\hat{\gamma}_4$ with a p-value of 0.00743 with the AIC value obtained is 115.1136. So that the model formed is as follows:

- Discrete data models for $\hat{\mu}_i$

$$\hat{\mu}_i = \exp(3{,}68\text{-}0{,}022X_2\text{-}0{,}0019X_3 + 0{,}0027X_4)$$

- Zero inflation model for $\hat{p}_i$

$$\hat{p}_i = \exp\left(\frac{20{,}47\text{-}0{,}093X_2\text{-}0{,}074X_3\text{-}0{,}083X_4}{1 + 20{,}47\text{-}0{,}093X_2\text{-}0{,}074X_3\text{-}0{,}083X_4}\right)$$

Because there are still insignificant parameters, a further test will be carried out again to test the significance of the parameter estimation results in the Zero-Inflation Negative Binomial (ZINB) model by removing some of the parameters that have the largest p-value so that a model with a significant parameter value can be obtained. The smallest AIC. So by paying attention to the results of the p-value in each parameter in Table 5.9, it is known that the highest p-value lies in the predictor variable X_3 parameter $\hat{\beta}_3$ which is 0.88 so that the significance test of the parameter estimation results in the Binomial Zero-Inflation Negative Model will be carried out (ZINB) by setting aside the predictor variables X_1 and X_3 as follows.

Table 5.10 Testing ZINB Regression Parameters Simultaneously Without X_1 and X_3.

Log-Likelihood	Df	*chi-square value*	*Pr(>chisq)*	Information
-60,194	-4	20,964	0,00032	H_0 rejected

Based on Table 5.10, it is shown that the Pr value (> chisq) is 0.00032, which is less than the α value of 0.05. So it can be concluded that H_0 is rejected. This shows that simultaneously the variables X_2 and X_4 have a significant effect on the response variable Y.

Table 5.11 ZINB Parameter Significance Test without Variables X_1 and X_3.

Parameter	Estimation	SE *(Standard Error)*	Z test	*(Pr>|Z|)*
$\hat{\beta}_0$	3,53	1,51	2,33	0,0197*
$\hat{\beta}_2$	-0,022	0,017	-1,27	0,2037
$\hat{\beta}_4$	0,0027	0,0026	1,038	0,99
$\hat{\gamma}_0$	13,993	7,46	1,9	0,061
$\hat{\gamma}_2$	-0,103	0,080	-1,281	0,20005
$\hat{\gamma}_4$	-0,071	0,025	-2.869	0,00412*

[C Value: 113,4242

Based on Table 5.11, the results obtained from testing the parameter estimates in the Zero-Inflation Negative Binomial (ZINB) model by removing the X_1 and X_3 variables. The significant parameter is at $\hat{\gamma}_4$ with a p-value of 0.00412 with the AIC value obtained is 113.4242. So that the model formed is as follows.

- Discrete data models for $\hat{\mu}_i$

$$\hat{\mu}_i = \exp(3{,}53\text{-}0{,}022X_2\text{-}0{,}0027X_4)$$

- Zero inflation model for $\hat{p}_i$

$$\hat{p}_i = \exp\left(\frac{13{,}993\text{-}0{,}103X_2\text{-}0{,}071X_4}{1 + 13{,}993\text{-}0{,}103X_2\text{-}0{,}071X_4}\right)$$

Because there are still insignificant parameters, a further test will be carried out again to test the significance of the parameter estimation results in the Zero-Inflation Negative Binomial (ZINB) model by removing some of the parameters that have the largest p-value so that a model with a significant parameter value can be obtained. The smallest AIC. So by paying attention to the results of the p-value in each parameter

in Table 5.11, it is known that the highest p-value lies in the variable X_2 parameter $\hat{\beta}_2$, namely 0.2 so that the significance test of the parameter estimation results in the Zero-Inflation Negative Binomial (ZINB) model will be carried out. By setting aside the predictor variables X_1, X_2 and X_3 as follows:

Table 5.12 Testing of ZINB Regression Parameters Without X_1, X_2 and X_3.

Log-Likelihood	Df	*chi-square value*	*Pr(>chisq)*	Information
-60,194	-2	17,192	0,00018	H_0 rejected

Based on Table 5.12, it is shown that the Pr value (> chisq) is 0.00018, which is less than the α value of 0.05. So it can be concluded that H_0 is rejected. This shows that simultaneously the variable X_3 has a significant effect on the response variable Y.

Table 5.13 ZINB Parameter Significance Test without Variables X_1, X_2 and X_3.

Parameter	Estimation	SE *(Standard Error)*	Z test	*(Pr>|Z|)*
$\hat{\beta}_0$	1,6	0,26	6,086	1.16e-09*
$\hat{\beta}_4$	0,0018	0,0026	0,691	0,49
$\hat{\gamma}_0$	4,82	1,60	3,004	0,0027*
$\hat{\gamma}_4$	-0,073	0,0244	-3,010	0,00262*

.C value: 113,1961

Based on Table 5.13, the results obtained from testing the parameter estimates in the Zero-Inflation Negative Binomial (ZINB) model by removing variables X_1, X_2 and X_3. The significant parameter is at $\hat{\gamma}_4$ with a p-value of 0.00262 with the AIC value obtained is 113.1961. So that the model formed is as follows:

- Discrete data model for $\hat{\mu}_i$

$$\hat{\mu}_i = \exp(1{,}6 + 0{,}0018X_4)$$

- Zero inflation model for $\hat{p}_i$

$$\hat{p}_i = \exp\left(\frac{4{,}82\text{-}0{,}073X_4}{1 + 4{,}82\text{-}0{,}073X_4}\right)$$

The advanced test that has been carried out can be summarized in the table which is presented as follows:

Table 5.14 Summary of Advanced Test.

Variables from the model	Simultaneous Test	Significant Parameters	AIC
X_1, X_2, X_3, X_4	Significant	$\hat{\gamma}_4$	119,107
X_2, X_3, X_4	Significant	$\hat{\gamma}_4$	115,1136
X_2, X_4	Significant	$\hat{\gamma}_4$	113,4242
X_4	Significant	$\hat{\gamma}_4$	113,1961

6. Best Model Selection and Discussion.

Selection of the best model to use is to use the syntax AIC () in the R-studio software to find out the AIC value in a particular model that is formed where the model is significant in simultaneous testing. After testing the parameters simultaneously and partially, both from the formation of the model from the initial stage to the final stage, it can be seen in Table 5.14 so that from several models that have been tested by removing several variables with the highest p-value, it is obtained that all the models formed are significant means that all models formed by the independent variable have a significant effect on the dependent variable. However, in the various models that are formed the best model will be selected by taking into account the AIC value obtained. The smaller the AIC value, the better the model.

Based on Table 5.14, it is known that the smallest AIC value is obtained in a model that has an AIC value of 113.1961 with the predictor variable the percentage of pregnant women with high risk (X_4) in the parameter $\hat{\gamma}_4$ which is significant to the model. So that the Zero Inflation Negative Binomial (ZINB) regression model equation is formed, namely:

- Discrete data models for $\hat{\mu}_i$

$$\hat{\mu}_i = \exp(1{,}6 + 0{,}0018X_4)$$

- Zero inflation model for $\hat{p}_i$

$$\hat{p}_i = \exp\left(\frac{4{,}82\text{-}0{,}073X_4}{1 + 4{,}82\text{-}0{,}073X_4}\right)$$

The interpretation of the Zero Inflation Negative Binomial (ZINB) regression model is as follows:

a. Discrete data model for $\hat{\mu}_i$
   - A constant of 1.6 means that if RIST (X_4) is zero then the number of infant deaths is exp (1.6) = 4.95. This is because infant mortality is influenced by independent variable factors other than the model.
   - Each additional 1% of pregnant women is risti (X_4), it will increase the average number of infant deaths by exp (0.0026) = 1.0026 times the average number of cases of infant mortality in the first place, if other variables are not included in the model.

b. Zero inflation model for $\hat{p}_i$
   - A constant of 4.82 means that if RIST X_4 is zero then the number of infant deaths is exp (4.82) = 123.96. This is because infant mortality is influenced by independent variable factors other than the model.
   - Each addition of 1% of pregnant women is risti X_4, it will reduce the average number of infant deaths by exp (0.083) = 1.086 times the average number of cases of initial infant mortality, if other variables are not included in the model.

CONCLUSIONS

Based on the analysis and discussion of cases of infant mortality in Wonogiri Regency in 2015 using binomial negative zero-inflated regression, it was concluded that the dependent variable in this study was

the number of infant deaths (Y) and there were 4 independent variables, namely the percentage of low birth weight (X_1) , the percentage of deliveries in an area within a certain period of time assisted by a health professional (X_2), the percentage of households who have a clean and healthy lifestyle (X_3) and the percentage of pregnant women high risk (X_4).

a. Estimation of Binomial Negative Zero-Inflated Regression Parameters (ZINB).

The first step in estimating the zero inflation negative binomial (ZINB) regression is to form the likelihood function. The function obtained is:

$$L(\theta|y_i) = \begin{cases} \Pi \dfrac{e^{x_i^T\gamma} + \left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}}{1+e^{x_i^T\gamma}}, \text{for } y_i = 0 \\ \Pi \dfrac{\Gamma\left(y_i + {}^1/_k\right)\left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}\left(\dfrac{ke^{x_i^T\beta}}{1+ke^{x_i^T\beta}}\right)^{y_i}}{1+e^{x_i^T\gamma}\Gamma\left(\dfrac{1}{k}\right)\Gamma(y_i+1)}, \text{for } y_i = 1,2,3,\ldots \end{cases}$$

Next is to find the log-likelihood function. The log-likelihood function is obtained as follows:

$$\text{In } L(\theta|y_i) = \begin{cases} \sum_{i=1}^{n} \text{In}\left\{e^{x_i^T\gamma} + \left(\dfrac{1}{1+ke^{x_T^i\beta}}\right)^{\frac{1}{k}}\right\} - \sum_{i=1}^{n} \text{In}\left[1+e^{x_T^i\gamma}\right], \text{for } y_i = 0 \\ -\sum_{i=1}^{n} \text{In}\left[1+e^{x_T^i\gamma}\right] + \sum_{i=1}^{n} \text{In}\left[\Gamma\left(\dfrac{1}{k}+y_i\right)\right] - \sum_{i=1}^{n} \text{In}\left[\Gamma(y_i+1)\right] - \sum_{i=1}^{n} \text{In}\left[\Gamma\left(\dfrac{1}{k}\right)\right] \\ \quad +y_i \sum_{i=1}^{n} \text{In}\left\{\left(\dfrac{e^{x_T^i\gamma}}{1+ke^{x_T^i\gamma}}\right)^{y_i}\right\} - \left(\dfrac{1}{k}\sum_{i=1}^{n} \text{In}\left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{\frac{1}{k}}\right), \text{for } y_i = 1,2,3,\ldots \end{cases}$$

The estimate with the maximum likelihood ratio is calculated by maximizing its log-likelihood. Since the log-likelihood function is not linear, it cannot be solved by ordinary numerical methods. So the EM (Expectation Maximization) algorithm is used.

The EM algorithm is divided into 2 stages, namely the expectation stage (E-step) and the maximization stage (M-step) or also known as the Newton Raphson method. The results of the expectation stage (E-step) are:

- E-step results for parameters β

$$\ln L\left(\gamma^{(m)}\middle|y_i, w_i^{(m)}\right) = \sum\nolimits_{i=1}^{n}\left(1\text{-}w_i^{(m)}\right)\left\{\frac{\Gamma\left(\frac{1}{k}+y_i\right)}{\Gamma(y_i+1)\Gamma\left(\frac{1}{k}\right)}\left(\frac{e^{x_i^t\gamma}}{1+e^{x_i^t\gamma}}\right)^{y_i}\left(\frac{1}{1+ke^{x_i^t\beta}}\right)^{\frac{1}{k}}\right.$$

- E-step results for parameters γ

$$\ln L\left(\gamma^{(m)}\middle|y_i, w_i^{(m)}\right) = \sum_{i=1}^{n}\left[w_i^{(m)}x_i^t\gamma\text{-In}\left(1+e^{x_i^t\gamma}\right)\right]$$

The result of the maximization stage (M-step) or also known as the Newton Raphson method, the parameter estimation formula is obtained as follows:

- M-step results for parameters β
  $\beta^{(m+1)} = \beta^{(m)}\text{-}\left(H^{(m)}\right)^{-1}U^{(m)}$
- M-step results for parameters γ
  $\gamma^{(m+1)} = \gamma^{(m)}\text{-}\left(H^{(m)}\right)^{-1}U^{(m)}$

b. Application of Binomial Zero-Inflated Negative Regression Parameters (ZINB).

Based on the tests that have been carried out, the zero-inflated negative binomial regression model is obtained (ZINB which has the smallest AIC value of 113.1961 in cases of infant mortality in Wonogiri Regency in 2015 as follows:

- Discrete data models for $\hat{\mu}_i$
  $\hat{\mu}_i = \exp(1{,}6 + 0{,}0018X_4)$
- Zero inflation model for $\hat{p}_i$
  $\hat{p}_i = \exp\left(\dfrac{4{,}82\text{-}0{,}073X_4}{1+4{,}82\text{-}0{,}073X_4}\right)$

The interpretation of the Zero Inflation Negative Binomial (ZINB) regression model is as follows:

1. Discrete data models for $\hat{\mu}_i$
   - A constant of 1.6 means that if RIST (X_4) is zero then the number of infant deaths is exp (1.6) = 4.95. This is because infant mortality is influenced by independent variable factors other than the model.
   - Each additional 1% of pregnant women is risti (X_4), it will increase the average number of infant deaths by exp (0.0026) = 1.0026 times the average number of cases of infant mortality in the first place, if other variables are not included in the model.
2. Zero inflation model for $\hat{p}_i$
   - A constant of 4.82 means that if RIST (X_4) is zero then the number of infant deaths is exp (4.82) =123.96. This is because infant mortality is influenced by independent variable factors other than the model.
   - Each addition of 1% of pregnant women is risti (X_4), it will reduce the average number of infant deaths by exp (0.083) = 1.086 times the average number of cases of initial infant mortality, if other variables are not included in the model.

REFERENCES

Agresti, A. 2002. Categorical Data Analysis (2thed.). New York: Jhon Wiley & Sons, Inc.

Ariawan, B., Suparti and Sudarno. 2012. Zero-Inflated Negative Binomial (ZINB) Regression Modeling for Discrete Response Data with Excess Zeros. *Gaussian Journal* 1(1):55-64.

Garay, A.M., E.M. Hashimoto, E.M.M. Ortega, & V.C. Lachos. 2011. On Estimation and Influence Diagnostics for Zero -Inflated Negative Binomial Regression Models. Computational Statistics and Data Analysis, 55:1304-1318.

Kurniawan, Ilham. 2017. Model Regresi Poisson Terbaik Menggunakan Zero-Inflated Poisson (ZIP) dan ZeroInflated Negative Binomial (ZINB). [Essay]. Faculty of Mathematics and Natural Sciences, Semarang State University.

Sekarmini, Ni Made, I Komang Gde Sukarsa, I Gusti Ayu Made Srinadi. 2013. Penerapan Regresi Zeri-Inflated Negative Binomial (ZINB) untuk Penduga Kematian Anak Balita. *E-Jurnal Matematika* 2(4): 11-16.