

The Implementation of ZIP Regression and ZINB Regression to Overcome Overdispersion of Death Cases Due to Filariasis in Indonesia

Azizah^{1*}, Dewi Novita Sari²
e-mail* : azizah.fmipa@um.ac.id

^{1,2}*Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Malang, Indonesia*

ABSTRACT

Filariasis occurs as a result of parasitic infection by a mosquito vector. Filariasis affects the lymph and causes blood loss in the body. Filariasis is uncured. Therefore, prevention measures are needed to reduce the number of deaths. Filariasis generally does not cause death directly and rarely occurs. However, this results in an excess number of zeros in the data, which causes overdispersion. The data used were obtained from the book Indonesia Health Profile 2021, which covers the research area of Indonesia. The ZIP and ZINB regression methods were used to model the factors that affect mortality due to filariasis. Based on research conducted using ZINB regression, the results showed that the percentage of households with good sanitation has a significant effect on deaths caused by filariasis.

Keywords: overdispersion, excess zero, Poisson regression, ZIP, ZINB

INTRODUCTION

Filariasis known as elephantiasis disease is an infectious disease caused by infection with 3 types of filarial worms *Brugia malayi*, *Brugia timori*, and *Wuchereria bancrofti* (Rahmi et al., 2022). Filariasis is transmitted through the bites of *Mansonia*, *Anophelex*, *Culex*, and *Armigeres* (Sa'diyah et al., 2022). In Indonesia, filariasis is caused by the *Wuchereria bancrofti* worm in Sumatra, Java, Kalimantan, Sulawesi, Nusa Tenggara, Maluku and Papua. Another worm is *Brugia malayi* which is spread in Sumatra, Kalimantan, Sulawesi, and several islands in Maluku. *Brugia timori* is found in Flores, Alor, Rote, Timor, and Sumba and is an agricultural area (Rahmi et al., 2022).

Filariasis attacks the lymph nodes of sufferers and causes swelling of the legs, breasts, arms, and genital organs in humans (Ministry of Health of the Republic of Indonesia, 2021). Filariasis is a disease that cannot be cured but can be

prevented. Prevention can be done by knowing the factors that cause filariasis. Whether or not there are proper sanitation facilities and healthy latrines can affect the number of cases of filariasis (Sa'diyah et al., 2022). The administration of mass prevention drugs (POPM) for filariasis in cities/regencies can also be carried out to prevent filariasis transmission.

Death from filariasis is something that rarely occurs because, basically, filariasis does not cause death directly (Maulida Layli et al., 2020). In a statistical review, the number of cases of death due to filariasis is a type of count with non-negative data that can be modeled using Poisson regression. Poisson regression is a method used to analyze countable and non-negative data types that express the number of events that rarely occur within a certain time span (Taufan et al., 2012). In this case, the count and non-negative data type variables are the response variables.

Poisson regression is a generalized linear model (GLM) in which the response variables follow the Poisson distribution (Arum et al., 2022). According to McCullagh and Nelder (McCullagh & Nelder, 1989), the generalized linear model (GLM) is an extension of the classical linear model (Zahro et al., 2018). In the classical linear model, it is assumed that the response variables are normally distributed, whereas, in the GLM, the response variables are exponentially distributed. One of the GLM models is Poisson regression. In Poisson regression, the response variables are not normally distributed but instead have a Poisson distribution that belongs to the exponential family.

Poisson regression analysis explains the equidispersion assumption that the mean and variance have the same value (Pramono et al., 2014). However, overdispersion often occurs. Overdispersion is a condition where the variance value is greater than the average value (Dwi Putri et al., 2022). The cause of overdispersion is a large number of observations with a zero value in the response variable. Rare cases of death due to filariasis also cause many observation values to be zero and are called excess zero. The excess of zero can be seen from the proportion of the response variable that is zero more than other discrete values. In this case, Poisson regression is not ideal to use. One method that can be used to overcome overdispersion due to excess zero is Zero-Inflated (ZI) regression. According to (Lambert, 1992), the Zero-Inflated regression model is a mixed model between models for modeling count and binary data (Fávero et al., 2021).

The Zero-Inflated model consists of ZIP (zero-inflated poisson) and the negative binomial zero-inflated. (ZINB). In ZIP regression, the distribution used is a combination of the Poisson distribution and the Bernoulli distribution (Fávero et

al., 2021). In ZIP regression, according to (Jansakul & Hinde, J, 2002) the occurrence of zero values occurs in two ways: the state of zero with probability ω_i and the Poisson state with Probability $1-\omega_i$, and the poisson distribution with an average μ_i . (Kusuma et al., 2013). While Zero-Inflated Negative Poisson (ZINB) regression, according (Ginzel et al., 2020) is a form of a mixed model of the Poisson and Gamma distributions (Anggreainy et al., 2022). In ZINB regression, there are two states, namely a zero state in which the observation is zero with the probability ω_i and a negative binomial state with a probability $1-\omega_i$ with its negative Binomial distribution with the average μ_i .

ZIP and ZINB regression methods have often been used in previous studies. For example, research (Dwi Putri et al., 2022) uses the ZIP regression method. In this study, the response variable was the number of infant mortality cases in Bandung. Because infant mortality is a rare event, there are too many zeros in the observations, so it can be modeled using the ZIP regression method. In the study by Anggreainy et al. (2022), which used the ZINB regression method, the response variable was the death rate due to COVID-19. In 2018, Rahayu et al. (2018) conducted research on the incidence of diphtheria in Indonesia using the ZIP and ZINB regression methods. In this study, the response variable used was the number of diphtheria cases, which is an enumeration data type that follows a Poisson distribution. The results showed that the ZIP regression method was still unable to overcome cases of overdispersion in the data; therefore, the ZINB regression method was used to overcome overdispersion (Rahayu et al., 2018).

Based on this problem, this study aimed to apply ZIP and ZINB regression to data on cases of death due to filariasis with an excess of zero to determine the

factors that have a significant effect. Knowing these factors is expected to assist related parties in efficiently implementing filariasis elimination programs. This research is limited to cases of death from filariasis in Indonesia in 2021 and the use of Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) regression methods.

METHODS

The data used in this study is secondary data from the 2021 Indonesia Health Profile book, which can be accessed via (<https://www.kemkes.go.id/downloads/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-2021.pdf>). This is a quantitative study because it examines the variables that affect the number of deaths from filariasis. In this study, the data were analyzed using the R Studio software with the MASS, pscl, and lmtest libraries. The coverage area of this research is Indonesia, which has 34 provinces. Based on previous research (Sa'diyah et al., 2022) (Amaliana et al., 2018), the variables used in this study were as follows:

- a. Number of death cases from chronic filariasis (Y)
- b. Total number of chronic cases of filariasis (X1)
- c. Number of districts/cities that have succeeded in reducing Microfilariae <1% (X2)
- d. Number of districts/cities that are still administering filariasis Mass Prevention Drugs (POPM) (X3)
- e. Population density per km² (X4)
- f. Percentage of households having access to proper sanitation (X5)
- g. Percentage of households having access to healthy latrines (X6)
- h. Percentage of households having access to proper drinking water (X7)

The analytical steps used in analyzing the factors that influence death from filariasis in Indonesia using the ZIP regression and ZINB regression methods are as follows:

- a. Description of research variables.
- b. Test the relationship between independent variables by looking at the VIF value to see multicollinearity.
- c. Excess zeros testing on the response variable Y.
- d. Poisson regression modeling
 1. Estimating the parameters of the Poisson regression model.
 2. Testing the suitability of the Poisson regression model with the Likelihood Ratio Test (LRT).
 3. Testing the significance of each Poisson regression parameter with the Wald test.
 4. Determine the Poisson regression model.
- e. Performing the overdispersion test on the Poisson regression model. If there is overdispersion then proceed with the ZIP regression method.
- f. Zero-Inflated Poisson (ZIP) regression modeling
 1. Estimating the parameters of the ZIP regression model.
 2. Testing the suitability of the ZIP regression model with the Likelihood Ratio Test (LRT).
 3. Testing the significance of each ZIP regression parameter with the Wald test.
 4. Determine the ZIP regression model.
- g. Testing overdispersion in the ZIP regression model. If there is still overdispersion then proceed with the ZINB regression method.
- h. Zero-Inflated Negative Binomial (ZINB) regression modeling
 1. Estimating the parameters of the ZINB regression model.
 2. Testing the suitability of the ZINB regression model with the Likelihood Ratio Test (LRT).

3. Testing the significance of each ZINB regression parameter with the Wald test.
4. Determine the ZINB regression model.
- i. Testing overdispersion in the ZINB regression model.
- j. Interpret the regression model obtained.
- k. Make conclusions from the results of the analysis.

RESULTS AND DISCUSSION

Description of Research Variable

In 2021, there will be 617 cases of death due to filariasis in Indonesia, spread across 34 provinces. The most cases were found in East Nusa Tenggara Province, with 280 deaths. There are 14 provinces where no cases of death due to filariasis were found.

Table 1. Descriptive Statistics

Variable	n	Min	Maks	Mean
Y	34	0,00	280,00	18,15
X1	34	0,00	3629,00	275,10
X2	34	0,00	17,00	5,59
X3	34	0,00	11,00	0,94
X4	34	10,00	17031,00	772,82
X5	34	40,81	97,12	80,97
X6	34	3,70	100,00	83,64
X7	34	64,92	99,86	86,68

Identification of Multicollinearity

Multicollinearity between the predictor variables can be seen from the VIF value. VIF value ≥ 10 indicates the presence of multicollinearity. Table 2 shows the VIF value of each predictor variable. This value is ≤ 10 , which means that there is no multicollinearity in all predictor variables.

Table 2. VIF of Predictor Variables

Predictor Variables	VIF
X1	4,49
X2	1,76
X3	3,36
X4	1,26
X5	3,78
X6	1,22
X7	1,49

Excess Zeros Testing

Excess zeros are conditions where the proportion of zero observed values in the response variable is greater than other discrete data (Samson et al., 2022). The existence of excess zeros can cause cases of overdispersion.

Table 2. The proportion of Observation for Every Variable

Value of Proportion	Frequency	Percentage
0	14	41,19
1	5	14,71
2	1	2,94
3	2	5,88
5	1	2,94
6	1	2,94
9	2	5,88
21	1	2,94
31	1	2,94
33	1	2,94
36	1	2,94
46	1	2,94
64	2	5,88
280	1	2,94
Total	34	100,0

Based on Table 3, it can be seen that the proportion of observations that have a zero value is more than the other discrete values. Therefore, it can be concluded that the response variable has excess zeros.

Poisson Regression Modeling

The first step to determining the Poisson regression model is to estimate its parameters. In this study, data were analyzed using R Studio software with the MASS library to obtain parameter values from the model. Following are the parameter estimates for the Poisson regression model based on the R Studio software output.

Table 3. Estimation of Parameters

Parameter	Estimation Value	SE	P-value
β_0	-2,007e+00	9,552e-01	0,0356
β_1	1,170e-04	1,726e-04	0,4978
β_2	1,752e-01	1,743e-02	< 2e-16
β_3	-7,663e-02	5,050e-02	0,1291
β_4	-1,965e-05	4,856e-05	0,6856
β_5	-1,101e-01	1,351e-02	3,53e-16
β_6	1,067e-01	1,279e-02	< 2e-16
β_7	3,265e-02	8,923e-03	0,0003

Parameter estimation is calculated using R studio software and the results are as shown in the table above. From the table, the parameters that have a p-value <0.05 are parameters β_2 , β_5 , β_6 , and β_7 . The next step is simultaneous testing using LRT and partial testing using the Wald test with a significance level of 5%.

Likelihood Ratio Test (LRT)

Simultaneous testing can be carried out using the LRT. The hypothesis used in LRT is as follows:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

H_1 : There are the least $\beta_i \neq 0$ for $i = 1, 2, 3, 4, 5, 6, 7$

Reject H_0 if $G_{score} > \chi^2_{(0,05;7)}$

Based on the calculation using R Studio software, the results are obtained, $G_{score} = 1362,8$. Because $G_{score} > \chi^2_{(0,05;7)} = 14,0671$ then reject H_0 , it means that there is at least one predictor variable that has a significant effect on the response variable.

Wald Test

After simultaneous testing, partial testing is carried out to find out the significant variables in the Poisson regression model. Partial testing uses the Wald test with the following hypotheses:

$$H_0: \beta_i = 0$$

$H_1: \beta_i \neq 0$, for $i = 1, 2, 3, 4, 5, 6, 7$

Tolak H_0 if $W_i > \chi^2_{(0,05;1)}$

Table 4. Test for Partial Parameter Significance

W_{score}	$\chi^2_{0,05;1}$	Conclusion
$W_1 = 0,46$	3,8415	Do not Reject H_0
$W_2 = 101,04$	3,8415	Reject H_0
$W_3 = 2,30$	3,8415	Do not Reject H_0
$W_4 = 0,16$	3,8415	Do not Reject H_0
$W_5 = 66,42$	3,8415	Reject H_0
$W_6 = 69,59$	3,8415	Reject H_0
$W_7 = 13,38$	3,8415	Reject H_0

Based on the partial parameter significance test in Table 5, it is known that there are four predictor variables that have a significant effect on the Poisson regression model. These factors are the number of districts/cities that have succeeded in reducing Microfilariae to <1% (X2), the percentage of households that have access to proper sanitation (X5), the percentage of households that have access to healthy latrines (X6), and the percentage of households that have access to healthy latrines (X6). percent of households had access to proper drinking water (X7). In summary, the parameter estimates of the Poisson regression model with significant variables were as follows:

Table 5. Estimation of Poisson Regression Model Parameters with Significant Variables

Parameter	Estimation	SE	P-value
β_0	-2,831036	0,654	1,49e-05
β_2	0,193075	0,011	< 2e-16
β_5	-0,107411	0,010	<2e-16
β_6	0,111687	0,011	< 2e-16
β_7	0,032782	0,008	4,53e-05



Based on the parameter estimates in Tables 6, the Poisson regression model is obtained as follows:

$$\mu_i = \exp(-2,831036 + 0,193075X_{2i} - 0,107411X_{5i} + 0,111687X_{6i} + 0,032782X_{7i})$$

Overdispersion Testing on the Poisson Regression Model

The stage after obtaining the Poisson regression model is to test the overdispersion. Test for overdispersion by dividing the Pearson's Chi Square value by the degrees of freedom. The hypothesis used is as follows.

$$H_0: \hat{\phi} \leq 1: \text{data are not overdispersed.}$$

$$H_1: \hat{\phi} > 1: \text{data are overdispersed.}$$

$$\hat{\phi} = \frac{\chi^2}{n - p - 1} = \frac{1362,8}{26} = 52,42 > 1$$

Because $\hat{\phi} > 1$, reject H_0 , it means overdispersion. Therefore, the Poisson regression method is no longer ideal for use.

Zero-Inflated Poisson (ZIP) Regression Modeling

In the ZIP regression, there are two models, namely, the log model, to determine the probability of the response variable having a value other than zero or a Poisson distribution. The second model, namely the logit model, was used to determine the probability of a zero-value response variable. The following is the parameter estimation for the ZIP regression model using R Studio software with the pscl library.

Table 6. Parameter Estimation for ZIP Regression

	Parameter	Estimation	SE	P - value
log	β_0	4,181	1,052	7,07e-05
	β_2	0,136	0,012	< 21-16
	β_5	-0,127	0,012	< 21-16
	β_6	0,099	0,010	< 21-16
	β_7	-0,008	0,009	0,394

logit	γ_0	7,408	5,183	0,1529
	γ_2	-0,250	0,130	0,0544
	γ_5	0,004	0,053	0,9349
	γ_6	0,004	0,028	0,8800
	γ_7	-0,084	0,056	0,1377

Based on Table 7, in the log model there are three parameters that have a p-value < 0.05 while in the logit model all parameters have a p-value > 0.05.

Likelihood Ratio Test (LRT)

The step after estimating the parameters is to test the suitability of the model or test the significance of the parameters simultaneously using LRT.. The hypothesis used in LRT is as follows:

$$H_0: \beta_i = \gamma_i = 0$$

$$H_1: \beta_i \neq 0 \text{ or } \gamma_i \neq 0 \text{ for } i = 2, 5, 6, 7$$

Using R Studio software, the results of the model suitability test using LRT were obtained with a significance level of 5%, namely $G_{score} = 935,67$. Because $G_{score} > \chi^2_{(0,05;8)} = 15,5073$ then reject H_0 , it means that there is at least one predictor variable that influences the response variable and the model is appropriate.

Wald Testing

The Wald test was used to partially test the significance of the parameters of the ZIP regression model. In the Wald test, the hypothesis used is as follows:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0, \text{ if } i = 2, 5, 6, 7$$

$$\text{Reject } H_0 \text{ if } W_i > \chi^2_{(0,05;1)}$$

Following are the results of the Wald test to see the significance of the parameters partially.

Table 7. Partial Significance Test of ZIP Regression Model Parameters

	W_{score}	$\chi^2_{0,05;1}$	Conclusion
β_2	122,59	3,8415	Reject H_0
β_5	102,94	3,8415	Reject H_0
β_6	89,66	3,8415	Reject H_0
β_7	0,73	3,8415	Do Not Reject H_0
γ_2	3,70	3,8415	Do Not Reject H_0
γ_5	0,01	3,8415	Do Not Reject H_0
γ_6	0,02	3,8415	Do Not Reject H_0
γ_7	2,21	3,8415	Do Not Reject H_0

Based on Table 8, it is known that the predictor variables that have a significant effect on the response variable are X2, X5, and X6. Parameter estimation with significant variables as follows.

Table 8. Parameter Estimation of the ZIP Regression Model with Significant Variables

	Parameter	Estimation	SE	P-value
Log	β_0	3,619268	0,826	1,19e-05
	β_2	0,138650	0,012	< 21-16
	β_5	-0,12628	0,012	< 21-16
	β_6	0,096151	0,010	< 21-16
Logit	γ_0	2,55399	3,839	0,5060
	γ_2	-0,21742	0,117	0,0639
	γ_5	-0,03286	0,047	0,4827
	γ_6	0,00999	0,027	0,7083

Table 9 shows there are three parameters for the log model that reject H_0 namely β_2, β_5 and β_6 while in the logit model, all parameters accept H_0 . Because all parameters in the logit model accept H_0 , only the log model is obtained as follows:

$$\ln(\mu) = 3,619268 + 0,138650X_2 - 0,12628X_5 + 0,096151X_6$$

Overdispersion Testing on the ZIP Regression Model

The hypothesis used is as follows.

$H_0: \hat{\phi} \leq 1$: There is no overdispersion in the data.

$H_1: \hat{\phi} > 1$: There is overdispersion in the data

$$\hat{\phi} = \frac{X^2}{n - p - 1} = \frac{932,56}{29} = 32,157 > 1$$

From the calculations, it is found still overdispersion, because $\hat{\phi} > 1$. It can be interpreted that a regression model using ZIP still can't overcome the overdispersion case. Then, to overcome overdispersion, the Zero-Inflated Negative Binomial (ZINB) regression method is used.

Zero-Inflated Negative Binomial (ZINB) Regression Modeling

Because the ZIP regression still cannot overcome overdispersion, modeling is carried out using the ZINB regression method. Parameter estimation for the ZINB regression model with R Studio software using the pscl library is as follows.

Table 9. Estimation of Parameter Using ZINB Regression

	Parameter	Estimation	SE	P-value
Log	β_0	15,77	5,99	0,0085
	β_2	0,06	0,08	0,4957
	β_5	-0,19	0,07	0,0047
	β_6	0,02	0,01	0,1505
Logit	γ_0	20,357	27,25	0,455
	γ_2	-0,530	0,45	0,242
	γ_5	-0,244	0,31	0,441
	γ_6	-0,006	0,05	0,908

In the R Studio output from Table 10, in the log model, only the β_5 parameter has a p-value < 0.05, while in

the logit model, there are no parameters with a p-value < 0.05. The next step after estimating the parameters of the ZINB regression model, is to perform a model suitability test or a simultaneous test using LRT.

Likelihood Ratio Test (LRT)

The hypothesis used in LRT is as follows.

$H_0: \beta_i = \gamma_i = 0$

$H_1: \beta_i \neq 0$ atau $\gamma_i \neq 0$ for $i = 2, 5, 6, 7$

The results of the model suitability test using the LRT with a significance level of 5%, namely $G_{score} = 18,295$. Conclusion reject H_0 because $G_{score} > \chi^2_{(0,05;8)} = 15,5073$, it means that there is at least one predictor variable that influences the response variable and the model is appropriate.

Wald Test

After the simultaneous test was carried out, a partial test was carried out using the Wald test. The hypothesis used in the Wald test is as follows.

$H_0: \beta_i = 0$

$H_1: \beta_i \neq 0$, for $i = 2, 5, 6$

Reject H_0 if $W_i > \chi^2_{(0,05;1)}$

Table 10. Partial Significance Test of ZINB Regression Model Parameters

	W_{score}	$\chi^2_{0,05;1}$	Conclusion
β_2	0,4641	3,8415	Do Not Reject H_0
β_5	7,9755	3,8415	Reject H_0
β_6	0,0207	3,8415	Do Not Reject H_0
γ_2	1,36881	3,8415	Do Not Reject H_0
γ_5	0,5933	3,8415	Do Not Reject H_0
γ_6	0,01332	3,8415	Do Not Reject H_0

The Wald test results show that only the β_5 parameter is significant or the X_5 variable has a significant effect on the ZINB regression model. Parameter estimation for significant variables as follows.

Table 11. Parameter Estimation of the ZINB Regression Model with Significant Variables

	Parameter	Est.	SE	P-value
log	β_0	19,693	4,169	2,32e-06
	β_5	-0,214	0,051	2,79e-05
logit	γ_0	23,02	48,15	0,633
	γ_5	-0,423	0,90	0,633

There are no significant parameters in the logit model causing no model to be formed. Therefore, the model obtained for the ZINB regression is as follows: $\ln(\mu) = 19,69328 - 0,21427X_5$

Overdispersion Testing on the ZINB Regression Model

The hypothesis used for this test is as follows.

$H_0: \hat{\phi} \leq 1$: data are not overdispersed.

$H_1: \hat{\phi} > 1$: data are overdispersed.

$$\hat{\phi} = \frac{X^2}{n - p - 1} = \frac{24,102}{29} = 0,8311 < 1$$

Based on the calculation, it obtained that $\hat{\phi} < 1$ which means there is no case of overdispersion. This indicates that the ZINB regression method is better at handling data overdispersion than the Poisson regression and ZIP regression methods.

ZINB Regression Model Interpretation

Based on the analysis that has been done, the ZINB regression model is obtained as follows:

$$\ln(\mu) = 19,69328 - 0,21427X_5$$

Where X_5 represents the percentage of households that have access to proper sanitation by the province in Indonesia by 2021. The model is interpreted as follows for the count data model, every 1% change in households that have access to proper

sanitation by the province in Indonesia in 2021 will result in a decrease in cases of death from filariasis by $\exp(0.21427) \approx 0.807$.

CONCLUSION

Based on this study, Poisson regression is not ideal for data with overdispersion because there are more zeros. Zero-inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) regression methods can be used to overcome cases of overdispersion due to excess zero. The analysis that has provided evidence that the ZIP regression is still unable to overcome overdispersion then proceeded with using the ZINB regression method, which shows that the ZINB regression is able to overcome overdispersion in this study. The ZINB regression model obtained is as follows:

$$\ln(\mu) = 19,69328 - 0,21427X_5$$

The model can interpret that the factor that has a significant influence on cases of death from filariasis in Indonesia is the percentage of households with access to inadequate sanitation. For the count or $\ln(\mu)$ data model, every 1% change in households that have access to proper sanitation by the province in Indonesia causes a decrease in cases of death from filariasis by 0.807 times. Meanwhile, for the zero-inflation or logit (ω) model, no model was formed because there were no significant parameters in the logit model.

Suggestions for further research include the use of other methods, such as zero-inflated generalized Poisson (ZIGP) regression, to overcome the problem of overdispersion.

REFERENCES

Amaliana, L., Fernandes, A. A. R., & Solimun. (2018). Comparison of

Two Weighting Functions in Geographically Weighted Zero-Inflated Poisson Regression on Filariasis Data. *Journal of Physics: Conference Series*, 1097(1). <https://doi.org/10.1088/1742-6596/1097/1/012070>

Anggreainy, M. S., Illyasu, A. M., Musyaffa, H., & Kansil, F. H. (2022). Analysis of Factors Influencing the COVID-19 Mortality Rate in Indonesia using Zero Inflated Negative Binomial Model. *International Journal of Advanced Computer Science and Applications*, 13(4), 728–734. <https://doi.org/10.14569/IJACSA.2022.0130485>

Arum, K. C., Ugwuowo, F. I., & Oranye, H. E. (2022). Robust modified jackknife ridge estimator for the Poisson regression model with multicollinearity and outliers. *Scientific African*, 17, e01386. <https://doi.org/10.1016/j.sciaf.2022.e01386>

Dwi Putri, A., Devianto, D., & Yanuar, F. (2022). Pemodelan Jumlah Kematian Bayi di Kota Bandung dengan Menggunakan Regresi Zero-Inflated Poisson. *Jurnal Matematika UNAND*, 11(1), 12–24.

Fávero, L. P., de Freitas Souza, R., Belfiore, P., Corrêa, H. L., & Haddad, M. F. C. (2021). Count Data Regression Analysis: Concepts, Overdispersion Detection, Zero-inflation Identification, and Applications with R. *Practical Assessment, Research and Evaluation*, 26, 1–22. <https://doi.org/10.7275/44nn-cj68>

Ginzel, M. D. C., Saefuddin, A., & Erfiani. (2020). Determination of Extracurricular Activities for High School Students Based on Multiple Intelligence. *Journal of Statistics*, 9.

Jansakul, N., & Hinde, J. P. (2002). Score Tests for Zero-Inflated Models.



- Computational Statistics and Data Analysis*, 40, 75–96.
- Kementerian Kesehatan Republik Indonesia. (2021). Profil Kesehatan Indonesia Tahun 2021. In *Pusdatin.Kemkes.Go.Id.*
- Kusuma, W., Komalasari, D., & Hadijati, M. (2013). *Model Regresi Zero Inflated Poisson Pada Data Overdispersion*. 3(2).
- Lambert, D. (1992). Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *American Statistical Association and the American Society for Quality Control*, Vol. 34, N.
- Luciani Rahayu, R., Asrof, A., Rustiana, S., Puspitasari, W., & Suparman, Y. (2018). Perbandingan Regresi Zero Inflated Negatif Binomial dan Regresi Hurdle Negatif Binomial pada Data Overdispersi (Studi Kasus: Kejadian Difteri di Indonesia). *Jurnal Sains Matematika Dan Statistika*, 4(1).
- Maulida Layli, F., Sakundarno Adi, M., Dian Saraswati, L., & Ginandjar, P. (2020). Gambaran Faktor-Faktor Kepatuhan Minum Obat Dalam Pelaksanaan POPM di Kabupaten Semarang (Studi di Wilayah Kerja Puskesmas Gedangan Kecamatan Tuntang). *Jurnal Kesehatan Masyarakat*, 8(75), 147–154. <https://doi.org/10.1016/j.jnc.2020.12.5798><https://doi.org/10.1016/j.smr.2020.02.002><http://www.ncbi.nlm.nih.gov/pubmed/810049><http://doi.wiley.com/10.1002/anie.197505391><http://www.sciencedirect.com/science/article/pii/B9780857090409500205>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Model*. London: Chapman and Hill.
- Pramono, M. S., Maryani, H., & Wulandari, S. P. (2014). Analysis of Filariasis Through Zero Inflated Poisson (ZIP) Regression Approach. *Buletin Penelitian Sistem Kesehatan*, 17(1). <https://doi.org/10.22435/bpsk.v17i1Jan.3573>
- Rahmi, I. R., Sutiningsih, D., Hestiningsih, R., & Saraswati, L. D. (2022). Faktor-Faktor yang Berhubungan dengan Kasus Filariasis di Indonesia: Sistematis Review. *Jurnal Epidemiologi Kesehatan Komunitas*, 7(2), 501–521. <https://doi.org/10.14710/jekk.v7i2.11515>
- Sa'diyah, N. K., Astuti, A. B., & Mitakda, M. B. T. (2022). Simulation Study of Bayesian Hurdle Poisson Regression on the Number of Deaths from Chronic Filariasis in Indonesia. *Mathematics and Statistics*, 10(3), 603–609. <https://doi.org/10.13189/ms.2022.100316>
- Setiaji, et al. (2021). Profil Kesehatan Indonesia. Retrieved May 1, 2022, from <https://www.kemkes.go.id/downloads/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-2021.pdf>
- Taufan, M., Suparti, & Rusgiyono, A. (2012). *Analisis Faktor-Faktor yang Mempengaruhi Banyaknya Klaim Asuransi Kendaraan Bermotor Menggunakan Model Regresi Zero-Inflated Poisson (Studi Kasus di PT. Asuransi Sinar Mas Cabang Semarang Tahun 2010)*. 5, 49–61.
- Zahro, J., Caraka, R. E., & Herliansyah, R. (2018). *Aplikasi Generalized Linear Model pada R* (Issue July).