JISTech (Journal of Islamic Science and Technology)

JISTech, 10(1), 107-115, January – June 2025

ISSN: 2528-5718

http://jurnal.uinsu.ac.id/index.php/jistech

Zero Inflated Negative Binomial Regression In Malaria Cases In North Sumatera

Riska Aulia Dalimunthe[1], Ismail Husein[2]

[1,2] *Program Studi Matematika, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia*

ABSTRACT

Malaria cases in North Sumatra Province continue to be a public health concern, with regional incidence rates varying. This investigation was designed to assess the variables that contribute to the prevalence of malaria in the province by employing the Zero Inflated Negative Binomial (ZINB) regression model.. The response variable is the number of malaria cases, while the predictor variables are the number of impoverished individuals, population density, percentage of households with proper sanitation, number of healthy homes, and rainfall. For the year 2022, secondary data was acquired from the North Sumatra Provincial Health Office. Excess zero, overdispersion, and multicollinearity tests were conducted prior to the implementation of the ZINB model. The study results indicated that the ZINB model was more suitable than the Poisson and Negative Binomial models. The data indicates that the following variables have a substantial impact on malaria prevalence: an increase in the number of individuals living in poverty (X_1) by 7.8%, an increase in population density (X_2) by 1.8%, an increase in the percentage of households with adequate sanitation facilities (X_3) by 5.9%, and an increase in the percentage of rainfall (X_5) by 3.3%.

Kata Kunci: Malaria, overdispersion, excess zero, North Sumatra, Zero Inflated Negative Binomial
Email Address: [1] riskaaulia1302@gmail.com

DOI: http://dx.doi.org/10.30829/jistech.v10i1.25658
Received 10 February 2025; Received in revised form 10 April 2025; Accepted 15 June 2025

Introduction

Malaria is an infectious disease that is highly lethal on a global scale. It is caused by the Plasmodium parasite and is transmitted to humans through the bite of the female Anopheles mosquito [1]. Malaria remains classified as a severe infectious disease that originated from the Plasmodium parasite infection and is transmitted through the bite of the female Anopheles mosquito. This classification remains in effect to this day. The World Health Organization 2023 reports that developing countries experience over 240 million new cases and 600,000 fatalities annually. Indonesia aspires to eradicate malaria by 2030 [2].

The high incidence of malaria in North Sumatra has posed challenges for health authorities, particularly in regions with limited access to healthcare and sanitation. Statistical modeling of malaria data is crucial for understanding the factors driving the disease's spread and for guiding intervention strategies. Common regression models, such as Poisson regression, often fail to adequately handle overdispersion and the prevalence of zero cases in areas with low malaria incidence.

Malaria cases in North Sumatra reached 5,133 in 2022, with an uneven distribution across districts and localities [3]. Due to the prevalence of numerous zero values (areas with no cases) and overdispersion (variance greater than the mean), this presents difficulties in data modeling. The Poisson model, which is frequently employed, is difficult to satisfy due to its mean-variance assumption. Conversely, the Negative Binomial (NB) model accounts for overdispersion but fails to account for excess zeros. As a result, the Zero Inflated Negative Binomial (ZINB) model is a viable alternative, as it integrates a logistic component (zero inflation) with the NB model to forecast structural zero incidence and actual cases.

A suitable approach to address these concerns has been identified as the Zero-Inflated Negative Binomial (ZINB) regression model, which is a combination of two processes: a logistic regression for the zero counts and a negative binomial regression for the count data. The effectiveness of the ZINB model in modeling malaria incidence and identifying key determinants of the disease's distribution is evaluated in this study by applying it to malaria data from North Sumatra.

The effectiveness of the ZINB model in modeling data in the health sector has been demonstrated in previous studies. The research conducted by Salsabila [5] The study's findings indicated that the ZINB regression model, when applied with the MLE approach, generates parameter estimates that are more susceptible to sample size, resulting in superior performance for larger sample sizes. High variability and the presence of excess zeros are reflected in data with a high proportion of zeros. Consequently, the ZINB regression model can generate more precise and stable parameter estimates than data with a smaller proportion of zeros. As a result, the ZINB regression model is effective for data with a high proportion of zeros, as it is consistent with the data distribution's characteristics, particularly in the context of infant mortality caused by pneumonia.

The purpose of this investigation is to employ the Zero-Inflated Negative Binomial regression model to analyze malaria case data in North Sumatra Province. We anticipate that this method will offer a more thorough comprehension of the spatial distribution and determinants that influence malaria incidence.In order to analyze malaria case data in North Sumatra Province, this study will employ the Zero-Inflated Negative Binomial regression model. The spatial distribution and determinants that influence malaria incidence are anticipated to be more comprehensively understood through this approach.

The purpose of this investigation is to employ the ZINB model to investigate the factors that influence the incidence of malaria in North Sumatra Province in 2022.

Research Methodology

The study, which focused on 2022, was carried out using secondary data from the North Sumatra Province Health Office. The dataset contains important socioeconomic and environmental factors as well as the number of malaria cases reported in different districts and municipalities. To evaluate geographical differences in malaria incidence, the data were combined at the district level.

Secondary data was employed in this quantitative investigation from the North Sumatra Provincial Health Office in 2022. Statistical data was employed to determine the number of malaria cases in 33 regencies and localities in North Sumatra province, as well as the factors that contributed to these cases. This investigation employed five variables: rainfall (X5), population density (X2), the percentage of households with adequate sanitation (X3), the number of impoverished individuals (X1), and the number of healthy homes (X4).

The data source for this quantitative study is the Health Department of the Province of North Sumatra in 2022. The statistical data utilized in this study includes the number of malaria cases in 33 cities and towns in the province of Sumatra, as well as the factors and variables that contribute to the number of cases. In this study, the variables that are employed are the number of male students (X1), the number of female students (X2), the proportion of households with adequate sanitation (X3), the number of healthy households (X4), and the number of homeless students (X5).

Zero Inflated Negative Binomial Regression (ZINB)

Based on the Poisson Gamma Mixture distribution, the ZINB regression model is constructed. An independent random variable with i = 1,2,...,n, the response random variable can exist in two states: the zero state and the negative binomial state [6]:

With a specific probability, the zero state generates zero observations. However, the negative binomial state is characterized by a negative binomial distribution with a mean of 0 and a specific probability of occurrence. In this two-state process with variables, the two-component mixture distribution is generated. The probability density function in question is as follows (Nuraeni et al., 2022):

$$f(y \mid \alpha, \beta) = \frac{\Gamma(y+a)}{y!\Gamma(\alpha)} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(1-\frac{1}{1+\beta}\right)^{y} \quad y = 0,1,2,...$$

The ZINB regression model is divided into two sections, each of which is dependent on the variables:

a. Discrete data model for μ_i is

$$\ln(\mu_i) = x_i^T \beta, \mu_i > 0, \ i = 1,2,...,n,$$

where $\boldsymbol{\beta}$ is a parameter in the estimated arithmetic regression model of the ZINB regression model.

b. Zero-Inflation Model for p_i

$$\log it\left(p_i\right) = \ln\left(\frac{p_i}{1-p_i}\right) = x_i^T \gamma, \quad 0 \le p_i \le 1, \ i = 1,2,...,n,$$

Test Statistics for Model Parameters

Model parameter testing is implemented in a partial and concurrent manner to ascertain the extent to which predictor variables affect the response variable. Simultaneous testing investigates the extent to which predictor factors affect the response variable. Simultaneously, the likelihood ratio test is employed to evaluate the subsequent hypotheses [7]:.

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

$$H_1 : \text{ One degree of freedom is the minimum, with } j = 1, 2, , k$$

The likelihood ratio test was implemented in this investigation, and the relevant test statistics are as follows:

$$G = -2\ln\left[\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right] = 2\ln L(\hat{\Omega}) - 2\ln L(\hat{\omega})$$

The probability value for the fundamental model, which lacks independent variables, is $L(\hat{\omega})$, and for the complete model with independent variables, is $L(\hat{\Omega})$.

The resulting test statistic must be assessed to ascertain whether the null hypothesis is rejected. The Chi-Square distribution and the test statistic are both relevant to this $X^2_{(a,v)}$. The value will be influenced by the number of parameters in the model. The values are shown in the Chi-Square table $X^2_{(a,v)}$. A partial Wald test is employed to quantify the impact of each independent variable on the dependent variable. The Wald test is employed to evaluate partial hypotheses.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

With the estimated value $\text{j} = 1, 2, \cdots, k$, value with the Wald test statistic as follows [7] :

$$\text{W}_{\text{count}} = \frac{\beta_j}{SE(\beta_j)}$$

Selecting the Best Model

Determining the optimal model through the application of the Akaike Information Criterion (AIC). A model that optimally explains the data with the adequate number of parameters is identified using the AIC. Evaluation of the model's fit to the data is facilitated by AIC. This research was conducted with the objective of identifying the most appropriate model for the combined Poisson regression model [8].

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

with L is equivalent to There are numerous parameters in the model that are estimated, including the maximum likelihood function and k. The best model in the data set is the one with the lowest AIC [9].

Research Procedures

The subsequent procedures were implemented during this investigation:

1. Gathering data from the North Sumatra Provincial Health Office regarding the number of malaria cases.
2. Analyzing the data that has been collected.
3. Determining the multicollinearity of predictor variables using R VIF > 10.
4. The null hypothesis of the response variable Y or the number of malaria cases was tested through hypothesis testing in this research.
5. Validating the overdispersion assumption using the Poisson regression model's output.
6. Utilizing the Generalized Linear Model (GLM) in R 4.3.3 software to construct a Poisson regression model. Applying significance tests to the Poisson regression model and parameter estimates.
7. Proceed with zero-inflated negative binomial regression modeling if the data demonstrates excess zeros and overdispersion.
8. Parametric model test statistics for significance testing.
9. The best model was chosen from the Zero Inflation Negative Binomial regression model and the Zero Inflation Negative Binomial regression model with significant variables based on the AIC value, which is the criterion for the best model.
10. State the significant factors that influence the results of the analysis and then draw conclusions.
11. Finished

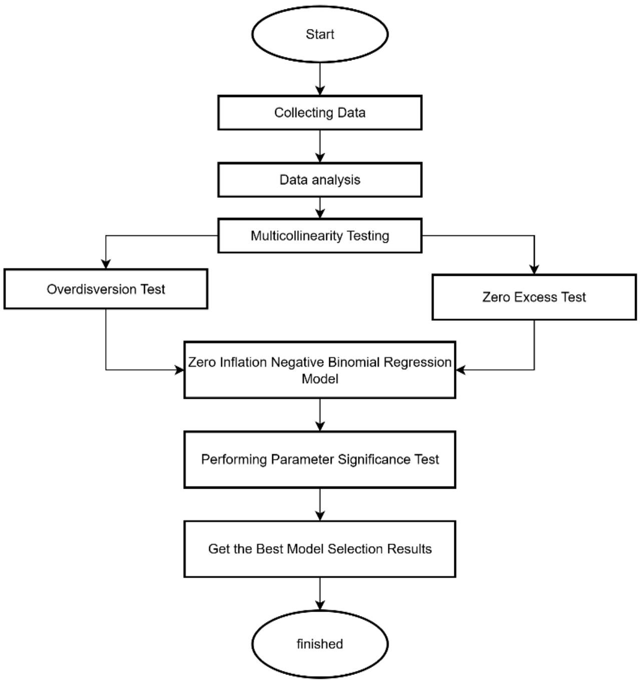

Figure 1. Flowchart

Results

The North Sumatra Provincial Health Office provided the secondary data utilized in this investigation. The statistical data utilized in this study comprises the number of malaria cases in 33 regencies and localities in North Sumatra Province in 2022, as well as the factors that contributed to these cases. The variables employed in this investigation include the number of impoverished individuals (X1), population density (X2), the proportion of households with adequate sanitation (X3), the quantity of healthy residences (X4), and rainfall (X5).

Table 1. Descriptive Statistics

Variabel	Mean	Varian	Min	Max
The quantity of malaria infections (Y)	93,939	48893,8	0	793
Age of the female population $(X1)$	38,430	1194,59	4,52	187,74
Population density in its entirety $(X2)$	8,817	3,125	3,706	8,865
The percentage of households that have access to adequate sanitation facilities $(X3)$	74,184	479,959	26,50	93,01
The quantity of healthful residences $(X4)$	32,144	2915,00	0	310,92
Percentage of rainfall $(X5)$	218,925	6223,26	117,39	407,77

The North Sumatra Province recorded 3,100 malaria cases in 2022. With 793 instances, Asahan Regency had the highest number of cases among the regencies and cities of North Sumatra. In contrast, Medan City and other regions experienced no malaria cases during the year. The high number of cases in Asahan Regency is suspected to be associated with limited access to adequate sanitation, high rainfall intensity, and high poverty rates, which, in turn, affect public health and environmental conditions, according to data from the North Sumatra Provincial Health Office in 2022.

Multicollinearity Test

Multicollinearity can be employed to ascertain whether an independent variable in a model is correlated with other independent variables. The objective of this test is to ascertain whether there is a substantial correlation between the regression model and the predictor variables. Table 4.2 provides the values of the VIF for the predictor variables.

Table 2. VIF Values of Predictor Variables

Independent Variable	Variance Of Inflasion
$(X1)$	1,611172
$(X2)$	1,692676
$(X3)$	1,149156
$(X4)$	1,266118
$(X5)$	1,051650

The final predictor variable's VIF value exceeds 10, suggesting that none of the predictor variables are multicollinear. Consequently, this value may be implemented to construct a regression model.

The data was analyzed using the Poisson regression modeling model. Table 3. Details regarding the Rstudio application Null Deviance: 46994; df: 32; Residual Deviance: 41941; df: 27; AIC value: 42225 Thus, the Poisson Regression Model is attained in the following manner:

$$\ln \mu_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

$$\mu_i = \exp\ (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6)$$

$$\mu_i = \exp\ (6{,}175 - 4{,}433X10^{-3} - 3{,}363X10^{-4} + 1{,}013X10^{-2}$$
$$- 1{,}711X10^{-3} + 1{,}991X10^{-3})$$

Table 3. Estimated values of the Poisson regression model

P	Estimate	Standar Eror	Z value	Pr(>|Z|)
β_0	$6{,}175$	$2{,}176X10^{-2}$	227,35	$< 2e-16***$
β_1	$-4{,}433X10^{-3}$	$2{,}981X10^{-4}$	-14,87	$< 2e-16***$
β_2	$-3{,}363X10^{-4}$	$6{,}635X10^{-6}$	-50,68	$< 2e-16***$
β_3	$1{,}013X10^{-2}$	$2{,}763X10^{-4}$	36,67	$< 2e-16***$
β_4	$-1{,}711X10^{-3}$	$1{,}425X10^{-4}$	-12,01	$< 2e-16***$
β_5	$1{,}991X10^{-3}$	$7{,}884X10^{-5}$	25,26	$< 2e-16***$

Table 3, which displays the number of districts and cities in North Sumatra Province that were used as experimental units, demonstrates that the dependent variable is influenced by nearly all independent variables.

Excess Zero

The Excess Zero test is evident in the quantity of response values entered. As indicated by the excess zero test results, the proportion of data that contains excess zero values exceeds 50%, or 57.58%. Data from the 2022 malaria case survey in North Sumatra Province was employed, which exhibited overdispersion and excess zero values. Consequently, one potential measure to resolve this matter is to implement ZINB regression.

Overdispersi

To identify overdispersion in the data, the overdispersion test is implemented. Application of the subsequent formula is implemented during this examination:

$$\phi = \frac{nilaideviance}{df}$$

$$\phi = \frac{4929{,}3}{27}$$

$$\phi = 182{,}5663$$

Based on the aforementioned data, the deviation value is 182.5663 when divided by the degrees of freedom. Overdispersion in the Poisson regression model is defined as a deviation value that is greater than 1 when divided by the degrees of freedom. The Poisson regression model's performance is rendered worse when there is overdispersion due to its high error rate. Consequently, the causal components of malaria in North Sumatra Province in 2022 are best described using Zero Inflated Negative Binomial (ZINB) regression analysis.

Regresi Zero Inflated Negative Binomial (ZINB)

The model was estimated and constructed using the zero-inflated negative binomial regression parameter estimation model. The parameter estimation results of the ZINB regression model are as follows:

Table 5. Results of ZINB Regression Model Parameter Estimation

Variable	Regresi ZINB		
	Model *Count Regressionm*		
	Estimate	Std.Error	*P-value*
Intersep	11,305	2,302	9,06e-07***
X_1	0,075	0,026	0,004502**
X_2	0,018	0,004	0,000152***
X_3	-0,061	0,022	0,006124**
X_4	-0,035	0,006	6,81e-07***
X_5	-0,657	0,343	0,55936.
	Model *Zero Inflated*		
	Estimate	Std.Error	*p-value*
Intersep	117,844	233,545	0,614
X_1	-0,274	4,894	0,955
X_2	0,040	0,336	0,904
X_3	-1,169	2,701	0,665
X_4	0,258	1,327	0,846
X_5	-0,306	0,715	0,669

A significance test was undertaken after the ZINB regression model parameters were tested. This test was conducted using simultaneous and partial tests to determine the significance of the zero-inflated negative binomial regression parameters. Importance Test of Zero-Inflated Negative Binomial (ZINB)

Parameters for Regression In order to assess the simultaneous and partial impact of these parameters, significance tests are implemented. The exams are administered in the following manner:

a. Simultaneous Test

The ZINB regression was conducted with simultaneous analysis to ascertain whether the independent variables have a substantial impact on the dependent variable. The alternative hypothesis employed is as follows:

$H_0 : \beta_0 = \beta_1 = \cdots = \beta_6 = \gamma_0 = \gamma_1 = \cdots = \gamma_6$ (Independent factors have no influence on the incidence of malaria)

H_1 : there is at least one, with $\beta_j \neq 0$ atau $\gamma_j \neq 0; j = 0,1,2,\cdots,6$

Table 6. Simultaneous Test on ZINB regression

Likelihood Chi-Square Ratio	Degrees of Freedom	*p-value*
43,713	10	3.706e-06***

G = 43.713 > X2 (0.1; 10) = 15.987 was determined using the Rstudio software's output. Ultimately, the H0 hypothesis was denied by this ZINB regression model test, which suggests that at least one parameter had a substantial impact on the number of malaria cases in North Sumatra Province by 2022.

a. Partial Test

Parametric models are parameter models that are composed of the values of each variable.

1. Model for Counting Regression

$H_0 : \beta_j = 0$ (There is no correlation between the independent and dividend variables)

$H_1 : \beta_j \neq 0$ (There is a relationship between the independent and dependent variables)

Table 7 illustrates the importance of the discrete data model parameters.

Table 7. Wald Test Results on Count Regression Model

Parameter	W_{count}	$x^2_{0,1;1}$	Criteria	Conclusion
β_0	$W_0 = \left(\frac{11,305}{2,302}\right)^2 = 24,117$	$2,706$	Reject H_0	Variable X_0 is significant to the dependent variable
β_1	$W_1 = \left(\frac{0,075}{0,026}\right)^2 = 8,321$	$2,706$	Reject H_0	Variable X_1 is significant to the dependent variable
β_2	$W_2 = \left(\frac{0,018}{0,004}\right)^2 = 20,25$	$2,706$	Reject H_0	Variable X_2 is significant to the dependent variable
β_3	$W_3 = \left(\frac{-0,061}{0,022}\right)^2 = 7,688$	$2,706$	Reject H_0	Variable X_3 is significant to the dependent variable
β_4	$W_4 = \left(\frac{-0,035}{0,006}\right)^2 = 34,027$	$2,706$	Accept H_0	Variable X_4 is not significant on the dependent variable
β_5	$W_5 = \left(\frac{-0,675}{0,343}\right)^2 = 3,872$	$2,706$	Reject H_0	Variable X_5 is significant to the dependent variable

Therefore, the regression model count is influenced by the following variables in the Western Indonesian Province in 2022: the number of male population, the number of female population, the proportion of households with low sanitation levels, and the proportion of households with high sanitation levels.

2. Model *Zero Inflation*

$H_0 : \beta_j = 0$ (There is no correlation between the independent and dividend variables)

$H_1 : \beta_j \neq 0$ (Relationship exists between the independent and dividend variables)

The parameter significance of the Zero-Inflation model is indicated in Table 8 below.

Table 8. Wald Test Results on Zero Inflation Model

Parameter	W_{count}	$x^2_{0,1;1}$	Criteria	Conclusion
γ_0	$W_0 = \left(\frac{117,844}{233,545}\right)^2 = 0,254$	$2,706$	Accept H_0	Variable X_0 is not significant on the dependent variable
γ_1	$W_1 = \left(\frac{-0,274}{4,894}\right)^2 = 0,003$	$2,706$	Accept H_0	Variable X_1 is not significant on the dependent variable
γ_2	$W_2 = \left(\frac{0,040}{0,336}\right)^2 = 0,014$	$2,706$	Accept H_0	Variable X_2 is not significant on the dependent variable
γ_3	$W_3 = \left(\frac{-1,169}{2,701}\right)^2 = 0,187$	$2,706$	Accept H_0	Variable X_3 is not significant on the dependent variable
γ_4	$W_4 = \left(\frac{0,258}{1,327}\right)^2 = 0,037$	$2,706$	Accept H_0	Variable X_4 is not significant on the dependent variable
γ_5	$W_5 = \left(\frac{-0,306}{0,715}\right)^2 = 0,183$	$2,706$	Accept H_0	Variable X_5 is not significant on the dependent variable

Thus, the conclusion Accept H_0 It can be inferred that the number of malaria sufferers in North Sumatra Province in 2022 is influenced by the number of poor people, the percentage of households with adequate sanitation, the number of healthy houses, and the percentage of rainfall.

Selecting the Best Model

Optimization was implemented in order to ascertain the optimal model for the ZINB regression method. The zero-inflation negative binomial (ZINB) regression model and the significant ZINB regression model were subjected to this test using the AIC, as detailed below:

Table 9. AIC Value

Model	AIC Value
Regresi ZINB	187,5
Regresi Poisson	5018

Based on Table 9, the zero-inflated negative binomial (ZINB) regression model was the most appropriate model. This is because the AIC of the ZINB regression model for this variable was lower than the AIC of the Poisson regression model, which was determined at 5018. Therefore, the better ZINB regression model was used, with an AIC of 187.5.

Based on this study, malaria prevalence in the Western Province of Indonesia is influenced by several statistically significant variables in the ZINB regression model: the number of visitors, the number of rooms, room cleanliness, and the cleanliness of the premises. Each of these variables has a significant impact on the number of malaria cases in the United States, as shown by the estimated results for each parameter in the model below:

Table 10. The ZINB Regression Model's parameter estimates are substantial.

Variable	Model Count Regression	Model Zero Inflated
Intersep	11,305	117,844
X_1	0,075	-0,274
X_2	0,018	0,040
X_3	-0,061	-1,169
X_5	-0,034	-0,306

Model Interpretation

The parameter estimation results of the zero inflation negative binomial (ZINB) model are determined by the results of the simultaneous and partial parameter tests. As indicated below, the regression calculation of the model and the zero inflation model are as follows:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_5 X_{i5}$$

$$\ln(\mu_i) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_5 X_{i5})$$

$$\ln(\mu_i) = \exp(11,305 + 0,075 X_{i1} + 0,018 X_{i2} - 0,061 X_{i3} - 0,034 X_{i5})$$

$$\log it(p_i) = \frac{\exp(117,844 - 0,274 x_{i1} + 0,040 x_{i2} - 1,169 x_{i3} - 0,306 x_{i5})}{1 + \exp(117,844 - 0,274 x_{i1} + 0,040 x_{i2} - 1,169 x_{i3} - 0,306 x_{i5})}$$

Using negative binomial regression analysis with zero inflation, variables that are statistically significant in influencing malaria cases in North Sumatra Province in 2022 were identified. An increase of one unit in the number of impoverished individuals (X_1) can result in a 7.8% increase in malaria cases. For every one unit increase in population density (X_2), the number of malaria cases can increase by 1.8%. Every 1% increase in the proportion of households with adequate sanitation (X_3) can result in a 5.9% decrease in the number of malaria cases. Every 1% increase in rainfall (X_5) can reduce malaria incidence by 3.3%.

Conclusion

This study shows how well the Zero-Inflated Negative Binomial (ZINB) regression model analyzes North Sumatra malaria data. The model produced more accurate and trustworthy predictions by effectively accounting for overdispersion and excessive zeros. Rainfall, population density, and poverty rate are important factors that affect the incidence of malaria in the area. Access to sanitary facilities and a healthy living environment are also important factors. The results highlight the necessity of focused interventions based on socioeconomic and environmental factors, offering important insights for malaria control strategies.

Test of Parameters The model's results indicate that there are variables that have a substantial impact on malaria. For example, a one-person increase in the number of impoverished individuals (X_1) can result in a 7.8% increase in malaria cases. Malaria cases can be increased by 1.8% for each unit increase in population density (X_2). Malaria cases can be reduced by 5.9% for each 1 percent increase in the percentage of households with adequate sanitation conditions (X_3). Malaria cases can be reduced by 3.3% for each 1 percent increase in the percentage of rainfall (X_5).

References

[1] Abdullah Husein, Rusmin Saragih, & Husnul Khair. (2024). Diagnosa Penyakit Malaria Menggunakan Metode Case Base Reasoning (CBR) (Studi Kasus: RSUD Djoelham Kota Binjai). *Modem : Jurnal Informatika Dan Sains Teknologi.*, 2(4), 01–09. https://doi.org/10.62951/modem.v2i4.211

[2] Achmad, N., Payu, M. R. F., & Rahim, Y. (2022). Pemodelan Pneumonia Berat Menggunakan Regresi Zero Inflated Negative Binomial di Gorontalo. *Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, *10*(1), 45–53. https://doi.org/10.34312/euler.v10i1.13990

[3] APRILIA, A. D., & Sofro, A. (2023). Regresi Zero Inflated Poisson Untuk Pemodelan Angka Positif Penyakit Malaria Di Jawa Timur. *MATHunesa: Jurnal Ilmiah Matematika*, *11*(2), 139–146. https://doi.org/10.26740/mathunesa.v11n2.p139-146

[4] Bernad Julvian Zebua, Jetslin Simbolon, & Selviani Damayanti Sipayung. (2024). Tingkat Pengetahuan Mahasiswa Prodi Manajemen Informasi Kesehatan tentang Penyakit Malaria. *SEHATMAS: Jurnal Ilmiah Kesehatan Masyarakat*, *3*(1), 171–176. https://doi.org/10.55123/sehatmas.v3i1.2868

[5] 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. https://doi.org/10.25077/jmu.11.1.12-24.2022

[6] Hasibuan, H. M., & Husein, I. (2025). *Journal of Computer Networks , Architecture and High Performance Computing Regression Modeling of Zero-Inflated Negative Binomia (ZINB) in Pneum onia Cases in Toddlers in North Sumatra Province Journal of Computer Networks , Architecture and High Perfor*. 7(1).

[7] Kusuma, B., Tampubolon, E. U. T., & Alfarisi, S. (2025). Analisis Tingkat Kriminalitas di Jawa Tengah dengan Pendekatan Distribusi Poisson dan Binomial Negatif. *AKSIOMA : Jurnal Sains Ekonomi Dan Edukasi*, 2(1), 314–334. https://doi.org/10.62335/58brp592

[8] Nuraeni, A., Martha, S., & Aprizkiyandari, S. (2022). *Penerapan Regresi Zero-Inflated Negative Binomial (Zinb) Pada Data Kecelakaan Lalu Lintas Di Kota Pontianak*. *11*(1), 89–96.

[9] Nurhidayah, F., & Muthiah Nur Angriany, A. (2025). Pendekatan Zero-Inflated Poisson Inverse Gaussian dalam Pemodelan Kasus Malaria di Puskesmas Kota Makassar. *ESTIMASI: Journal of Statistics and Its Application*, *6*(1), 101–115. https://doi.org/10.70561/ejsa.v6i1.43164

[10] Purnama, D. I. (2021). Comparison of Zero Inflated Poisson (ZIP) Regression, Zero Inflated Negative Binomial Regression (ZINB) and Binomial Negative Hurdle Regression (HNB) to Model Daily Cigarette Consumption Data for Adult Population in Indonesia. *Jurnal Matematika, Statistika Dan Komputasi*, *17*(3), 357–369. https://doi.org/10.20956/j.v17i3.12278

[11] Rasyidin, M. F., Anggraini, D., Muttaqin, H., & Selatan, K. (2023). *RAGAM: Journal of Statistics and Its Application Volume 02 Nomor 02, Desember 2023*. *02*(2017), 1–15.

[12] Ruliansyah, A., & Pradani, F. Y. (2020). Perilaku-Perilaku Sosial Penyebab Peningkatan Risiko Penularan Malaria di Pangandaran. *Buletin Penelitian Sistem Kesehatan*, *23*(2), 115–125. https://doi.org/10.22435/hsr.v23i2.2797

[13] Salsabila, S. W., & Efendi, A. (2025). *Zero Inflated Negative Binomial (ZINB) Regression : Application to the Pneumonia Study and Simulation under Several Scenarios*. *10*(1), 457–468.

[14] Sugiarta, C. K., Sumampouw, J. O., & Pinontoan, O. R. (2024). Annual Parasite Incidence Malaria di Kota Bitung tahun 2021-2023. *Sam Ratulangi Journal*

[15] WHO, "Malaria," World Health Organization, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/malaria