Available online at http://iorajournal.org/indx.php/orics /index

Operations Research: International Conference Series

e-ISSN: 2722-0974
p-ISSN: 2723-1739

Vol. 4, No. 4, pp. 126-140, 2023

Analysis of Risk Factors for Dengue Hemorrhagic Fever in Riau Province using Negative Binomial Regression

Aisyah Azhari Rangkuti[1*], Haposan Sirait[2]

[1,2]*Statistics Study Program, Faculty of Mathematics and Natural Sciences, Riau University, Riau, Indonesia*

**Corresponding author email: aisyah.azhari1519@student.unri.ac.id,haposansirait96@gmail.com*

Abstract

Dengue Hemorrhagic Fever (DHF) is a serious threat in Riau province, Indonesia. To better understand and control the spread of dengue fever, this research aims to analyze the factors that cause dengue fever. This study aims to identify significant risk factors that influence the spread of dengue fever in Riau Province. The Negative Binomial Regression Method was used to identify factors associated with the increase in dengue fever cases in Riau. The variables evaluated include population density of the Aedes aegypti vector, level of environmental cleanliness, prevention practices, and socio-economic factors. In addition, the best model was selected to overcome overdispersion in the data. The results of the analysis show that factors such as population density of the Aedes aegypti vector, environmental cleanliness, and the level of public understanding about dengue prevention practices have a significant influence on the spread of dengue fever in Riau. The best model used to overcome overdispersion in the 2021 dengue fever case data in Riau is Negative Binomial Regression. This research provides a deeper understanding of the factors causing dengue fever in Riau and selects an appropriate statistical model for analyzing data that experiences overdispersion. Negative Binomial Regression proved to be more appropriate in overcoming the problem of overdispersion in the data. These results can be used as a basis for designing more effective dengue prevention and control strategies and provide guidance for more targeted interventions in fighting dengue fever in this region.

Keywords: Dengue Hemorrhagic Fever, negative binomial regression, Aedes aegypti.

1. Introduction

Dengue Hemorrhagic Fever (DHF) is an infectious disease caused by the dengue virus and transmitted through the bite of the Aedes aegypti and Aedes albopictus mosquitoes. DHF is a disease caused by the Dengue virus which is transmitted by a mosquito vector. According to Trapsilowati (2007), the DHF virus has four different types, namely DEN-1, DEN-2, DEN-3, and DEN-4, and all of them have different genetic characteristics. According to Guerdan (2010), people affected by dengue fever usually experience high fever, pain in the joints, muscles and bones, headaches and pain behind the eyes. Dengue fever has become a serious public health problem in many tropical and sub-tropical countries, including Indonesia. According to Masyeni et al. (2021), Several factors are thought to influence the spread and incidence of dengue fever, including population density, environmental cleanliness, rainfall, temperature and air humidity. A better understanding of the risk factors for dengue can help in the development of effective prevention and control strategies.

Dengue Hemorrhagic Fever (DHF) is a disease that is a major health problem in various countries, including Riau Province, Indonesia. In 2021, the Riau Provincial Government recorded that the number of cases of Dengue Hemorrhagic Fever (DHF) had decreased drastically. This can be seen from the accumulated number of dengue fever cases in Riau Province over the last 3 years. Based on data received from the Riau Provincial Health Service, throughout 2019, a total of 4,139 dengue fever cases occurred in Riau Province from 12 districts/cities. According to the Head of the Riau Provincial Health Service, Mimi Yuliani Nazir, the number of dengue fever cases in 2020 has decreased by almost half, compared to the dengue cases that occurred in 2019, namely a total of 2,948 dengue fever cases throughout 2020. In 2021 the number dropped drastically to 472 cases of dengue fever in Riau as of August 2021.

morbidity rate in Riau Province is discrete and non-negative data. The analysis that is suitable for this type of data is Poisson Regression. The Poisson Regression Model is derived from the Poisson distribution. In the Poisson Regression model, the assumption that must be met is equidispersion. The equidispersion assumption is rare for discrete data types. Discrete data often experiences cases of overdispersion (high deviation), namely the mean and variance values are not the same or in other words the variance value is greater than the mean value (Utami 2013). If you continue to use Poisson Regression, it will result in the estimates of the regression coefficient parameters remaining consistent but inefficient. This has an impact on the standard error value being underestimated, so that the conclusion becomes invalid. Several studies related to solving overdispersion problems include the generalized Poisson regression model (Mahfudhotin, 2020) and negative binomial regression (Sauddin et al., 2020). According to research (Prahutama et al., 2017) one of the best choices is to use negative binomial regression. Therefore, the appropriate regression analysis to overcome overdispersion in Poisson Regression is Negative Binomial Regression.

Negative Binomial Regression is one way to model calculated data with equidispersion or overdispersion conditions. In estimating the parameters of the negative Binomial regression model, the maximum likelihood estimation (MLE) method can be used (Keswari et al., 2014). The aim of this research is to model dengue cases in Riau province in 2021 and also to determine the significant factors that influence dengue fever in Riau province using Negative Binomial Regression.

2. Literature Review

Previous research as a study is very important for writers to know the relationship between research conducted previously and research conducted currently and to avoid publication. This is useful for showing that the research carried out has important meaning so that the contribution of research to the development of science can be known. From the research, a recapitulation of results is obtained as in Table 1 below:

Table 1: Recapitulation of results

No.	Researcher Name	Researcher Variables	Research methods	Research result
1	Sartika (2012)	Number of DHF Cases Total Rainfall Number of fumigation operations Number of PHBS Households	Poisson Regression Negative Binomial Regression	When using Poisson Regression, it shows that there is a violation of assumptions, namely overdispersion. To overcome this incident, an alternative is used, namely Negative Binomial Regression. The results of the research are factors that influence the number of dengue fever sufferers, namely the population factor and the amount of rainfall per day.
2	Grace (2014)	Total Population Density Total Rainfall Number of DHF Cases Health Facilities and Personnel Area Height	Sequential Regression Panel Data Regression	Examining the factors that influence the spread of dengue fever in Bogor City using sequential regression and panel data regression approaches

Based on the results of several studies above, researchers obtained several indications of variables that influence dengue fever, namely dengue fever disease rate per 100,000 population, total population density, altitude of the area, number of health workers, number of health facilities, amount of rainfall, percentage of households that have Access to Decent Sanitation , and Percentage of Poor Population in Riau Province in 2021, so that the factors that researchers will use in conducting research are based on indications from these variables. Research on the factors that influence dengue fever in Riau Province uses Negative Binomial regression analysis. To the knowledge of researchers, research analyzing factors that influence dengue fever in Riau Province has never been carried out.

2.1. Dengue Hemorrhagic Fever (DHF)

Dengue fever is an acute febrile disease caused by the dengue virus and spread through the Aedes aegypti mosquito which has been infected with the dengue virus. The incubation period for the Dengue virus in humans (intrinsic incubation) ranges from 3 to 14 days before symptoms appear, clinical symptoms appear on average on the fourth to seventh day, while the extrinsic incubation period (in the mosquito's body) lasts around 8-10 days.

3. Research Methods

3.1. Research Population and Sample

The population studied in this research were all residents of Riau Province who suffered from dengue fever. The sample used in this research is part of the tally of dengue fever cases in Riau in 2021 obtained from the Central Statistics Agency of Riau Province.

3.2. Data Types and Sources

The type of data used is secondary data in annual form for 2021. The secondary data source in this research was taken from the Riau Province Central Statistics Agency.

3.3. How to Collect Data

This research uses two data collection methods, namely:

(i) Literature review
This research collects data and theories that are relevant to literature and other library materials such as articles, journals, books and previous research.

(ii) Documentary Studies

The type of data used in this research is secondary data collected from the 2021 Central Statistics Agency annual report.

3.4. Variables and Operational Definitions of Variables

Table 2: contains explanations and operational definitions of each research variable:

Table 1: Operational Definition of Variables

Variable	Operational V variable definition
Dengue morbidity rate (Y)	The number of dengue fever cases occurring in Riau Province in 2021 is divided by the population in the area, then the result is multiplied by 100,000 to give the morbidity rate per 100,000 population.
Population density (X_1)	The amount of population density per each sub-district in Riau Province in 20-21 is calculated using the sum formula The population of a region (district) is divided by the area of that region (district) with units of measurement $\frac{Soul}{km^2}$
Area height (X_2)	The height of each sub-district in Riau Province is measured from above sea level with the unit of measurement being m above sea level.
Amount Power Health (X_3)	The number of health workers (General Practitioners, Village Midwives) in Riau Province per sub-district during 2021.
Amount Bulk Rain (X_5)	Rainfall is the thickness of rainwater that collects in a place on an area of 1 m2 , a flat surface, does not evaporate and does not flow. Rainfall per year is the amount of rain during a year. The amount of rainfall is usually expressed in millimeters (mm).
Percentage of Households that Have Access to Adequate Sanitation (X_5)	The percentage of households in Riau Province that have access to adequate sanitation facilities is calculated as the number of households with access to adequate sanitation divided by total households, then multiplied by 100 to get the percentage.
Percentage of Poor Population (X_6)	The percentage of poor people in Riau Province in 2021 is calculated as the number of poor people divided by the total population, then multiplied by 100 to get the percentage.

3.5. Data analysis method

Data analysis in this research used R 4.2.3 and SPSS 24.0 software, which includes descriptive analysis and negative binomial regression analysis. The output results from descriptive analysis, negative binomial regression analysis will be analyzed so that the description of dengue sufferers, the level of incidence of dengue fever and conclusions regarding the factors that influence dengue fever in Riau Province are known.

3.6. Data Analysis Process

The research steps are visualized in a diagram via Figure 1.

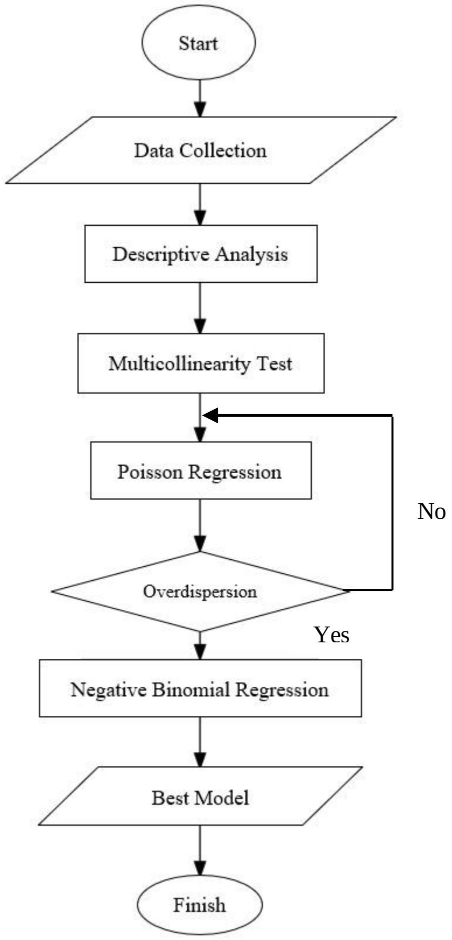

Figure 1: Data Analysis Flowchart

Data collection is looking for the data needed to carry out the analysis. The data in this research is data on factors that cause dengue fever in the form of data on the dengue morbidity rate per 100,000 population, total population density, altitude of the area, number of health workers, number of health facilities, total rainfall, Percentage of Households that Have Access to Adequate Sanitation , and Percentage of Poor Population in Riau in 2021. The observed data was then subjected to Poisson regression analysis to determine whether the data experienced overdispersion or not. Then the assumptions are tested from the observed data. If the data experiences equidispersion/underdispersion, the Overall Test and Partial Test will be immediately carried out . However, if the data experiences overdispersion then the data can be subjected to Negative Binomial regression analysis. Then an overall model test and partial test are carried out to find the best model from the analyzed data.

4. Results and Discussion

4.1. Descriptive Data

Descriptive analysis of the variables used is the first step in the research. The main benefit of descriptive analysis itself is that it provides an explanation of data. Researchers used the DHF morbidity rate in Riau in 2021 as the response variable (Y), then used seven predictor variables: population density of Riau Province (X_1), Regional Altitude (X_2), Number of Health Workers (X_3), Amount of Rainfall (X_4) , Percentage of Households that Have Access to Adequate Sanitation (X_5), and Percentage of Poor Population (X_6). In this research, secondary data was used from the Riau Province Central Statistics Agency and the Riau Province Health Service. The Table 3 shows descriptive statistics to show the characteristics of each variable.

Table 3: Descriptive Statistics

Variable	Average	Variance	Min	Max
morbidity rate (Y)	12.25	95.29	2.00	38.00
Population Density (X_1)	200.03	188685.96	31.29	1573.04
Area Height (X_2)	18.25	787.84	2.00	91.00
Number of Health Workers (X_3)	1164.58	2024584.99	345 .00	5624 .00
Amount of Rainfall (X_4)	210.09	2499.94	134.00	298.60
Percentage of Households that Have Access to Adequate Sanitation (X_5)	82.42	237.13	49.64	97.03
Percentage of Poor Population (X_6)	8.34	35.03	2.83	25.68

The number of cases of dengue fever in Riau in 2021 was recorded at 147 cases. The distribution of the highest dengue morbidity rate in 2021 for each district/city in Riau occurred in the city of Pekanbaru with a dengue morbidity rate of 38 people. Meanwhile, the lowest dengue morbidity rate in 2021 occurred in Pelalawan district with a dengue morbidity rate of 2 people. The observation units of the research are 10 districts and 2 cities in Riau. The Figure 2 following is a description of the first independent variable.

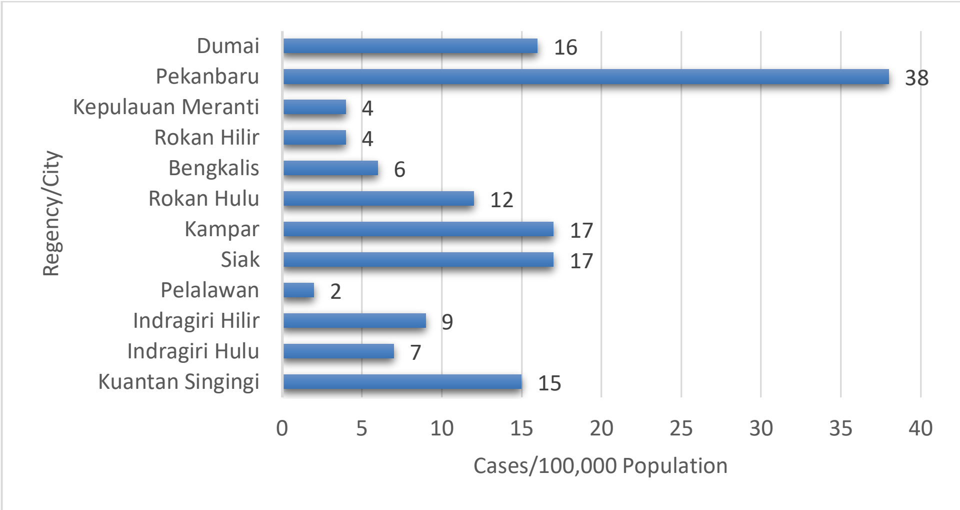

Figure 2: Districts/Cities with dengue fever sufferers based on dengue morbidity rates

It can be seen that the districts/cities with dengue morbidity rates highest in Riau Province in 20 21 based on Figure 2. The highest number of dengue fever cases is in Pekanbaru City, with 38 people.

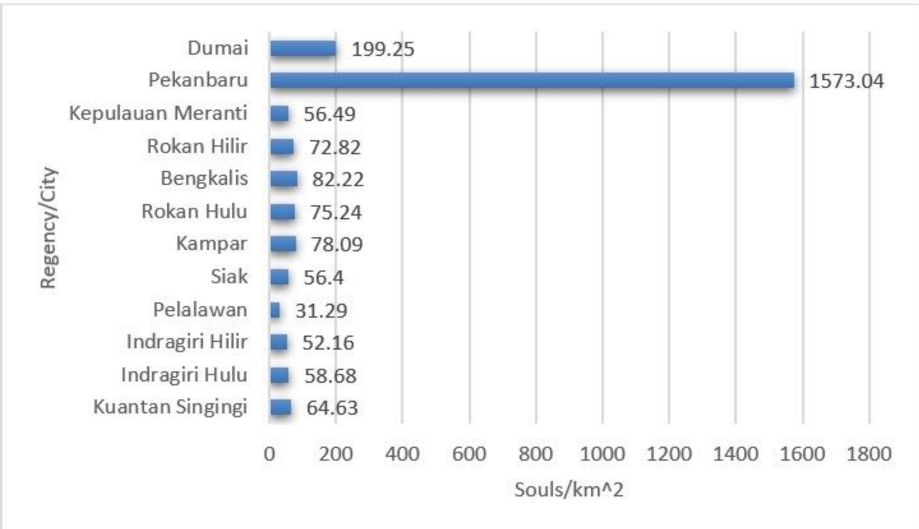

Figure 3: Regency/City with the highest number of dengue cases based on population density

It can be seen that the districts/cities with population density most in Riau Province in 20 21 based on Figure 3. Total population density Most of them are in Pekanbaru City as many as 1573.04 people / km^2 .

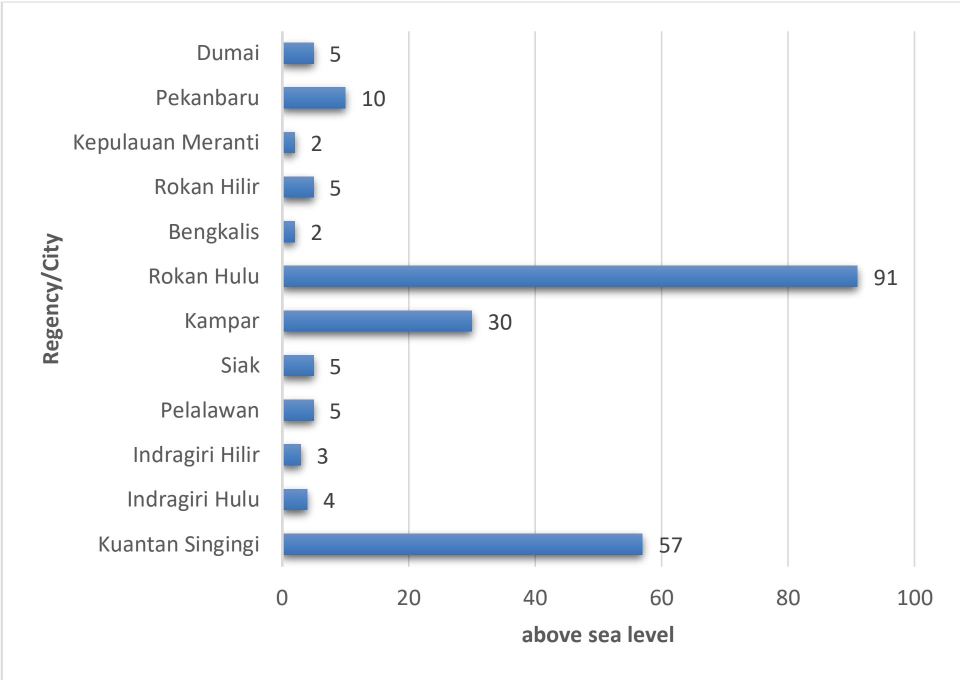

Figure 4: Districts/Cities with the most dengue fever sufferers based on regional altitude

It can be seen that the district/city with the highest regional altitude is in Riau Province in 20 21 based on Figure 4. It is located in Rokan Hulu Regency around 91 meters above sea level .

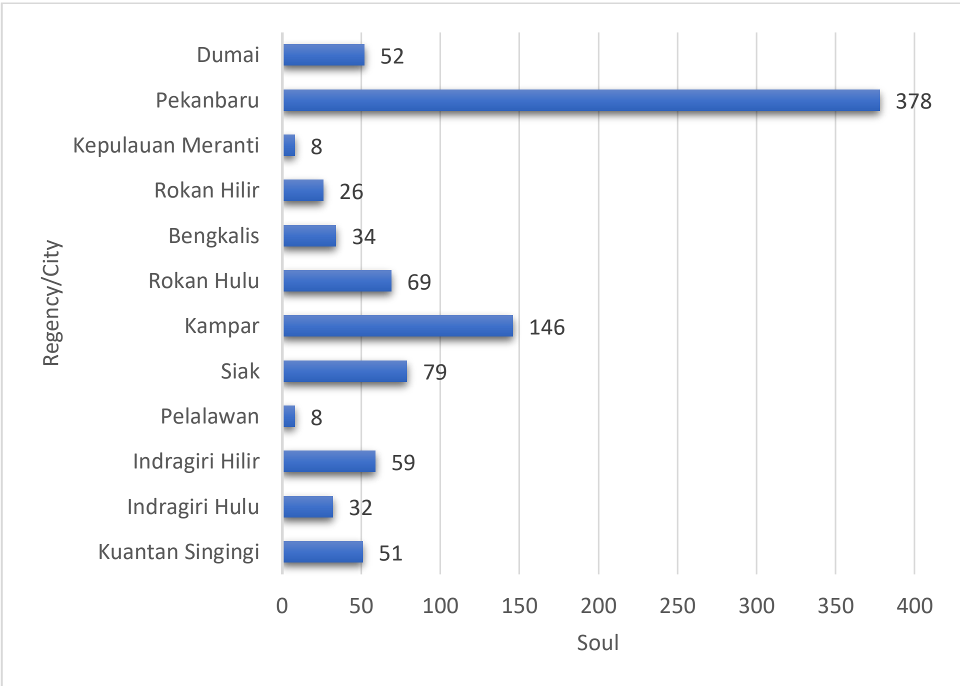

Figure 4: Regency/City with the most dengue fever sufferers based on the number of health workers

It can be seen that the districts/cities with the largest number of health workers are in Riau Province in 20 21 based on Figure 5. There are around 378 people in Pekanbaru City.

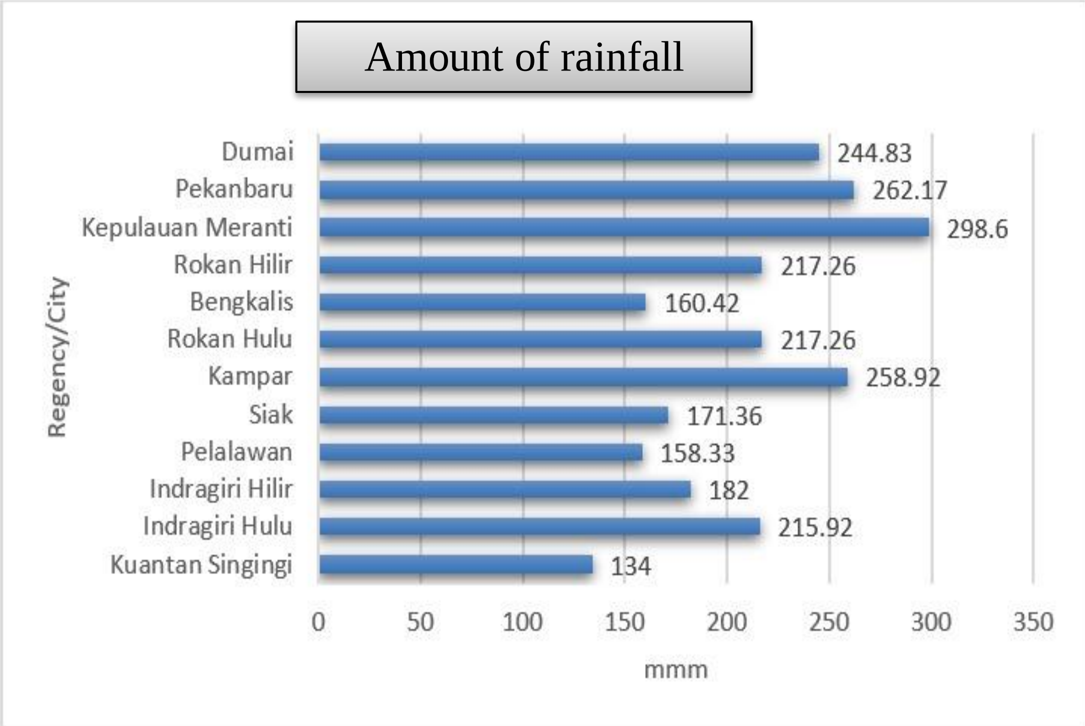

Figure 6: Districts/Cities with dengue fever sufferers based on the amount of rainfall

It can be seen that the districts/cities with the highest amount of rainfall are in Riau Province in 2021 based on Figure 6. In the Meranti Islands Regency, it is around 298.6 mm.

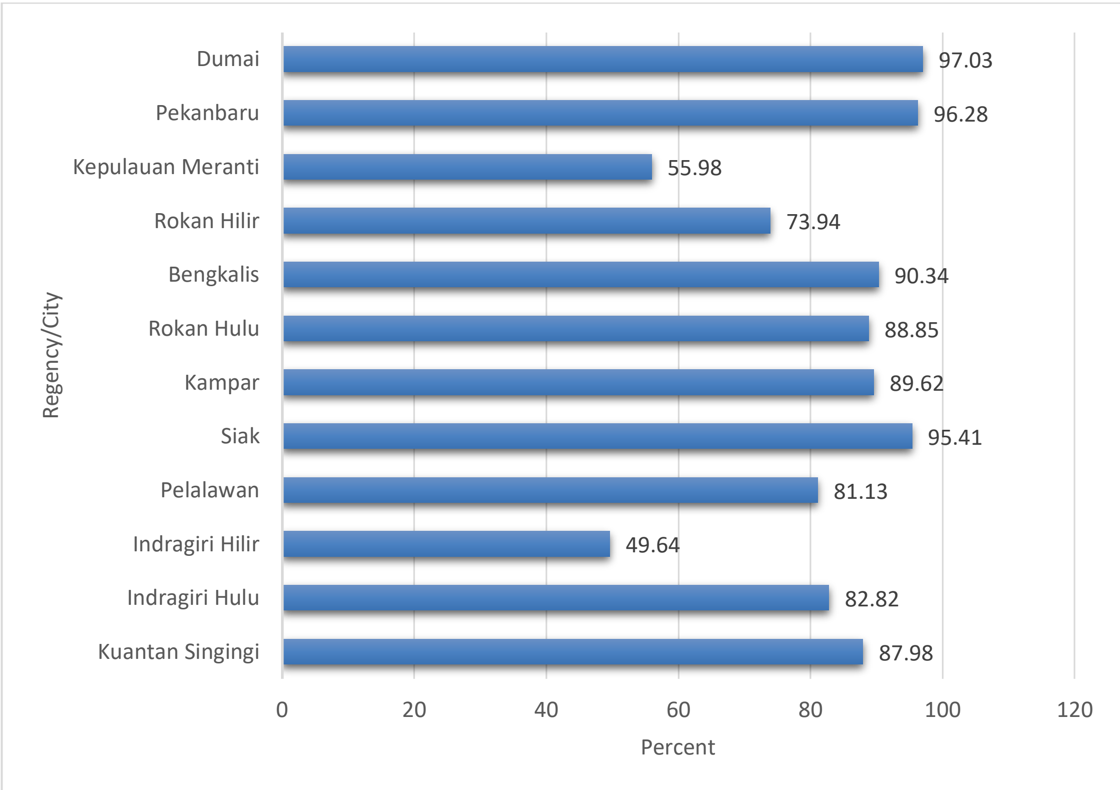

Figure 7: Districts/Cities with dengue fever sufferers based on the percentage of households that have access to adequate sanitation

It can be seen that the districts/cities with the highest percentage of households that have access to adequate sanitation are in Riau Province in 20 21 based on Figure 7. There are around 97.03 percent in Dumai City .

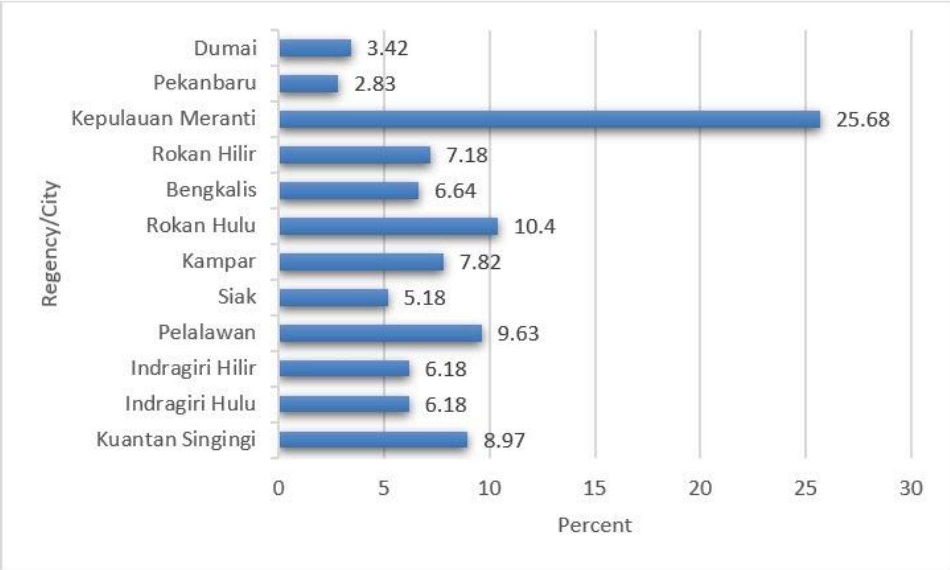

Figure 8: Districts/Cities with dengue fever sufferers based on the percentage of poor people

It can be seen that the districts/cities with the highest percentage of poor people are in Riau Province in 20 21 based on Figure 8. There are around 25.68 percent in Meranti Islands Regency .

4.2. Multicollinearity Test

One of the conditions that must be met in forming a regression model is that there are no cases of multicollinearity. The multicollinearity test is needed to determine whether there are independent variables that are similar to other independent variables in one model. Detection of multicollinearity cases can be seen in several ways , namely as follows.

If the Pearson correlation coefficient (r_{ij}) between the predictor variables is more than 0.95 then there is a correlation between these variables.

If the Variance Inflation Factor (VIF) value is greater than 10, it indicates that there is multicollinearity between the predictor variables. The VIF value is expressed as follows.

$$VIF = \frac{1}{1 - R^2} \quad (1)$$

with R 2 being the coefficient of determination, if the VIF value is more than 10 then it can be concluded that multicollinearity has occurred (Amaliana et al., 2018).

Table 4: Multicollinearity Test Results

Independent Variable	Tolerance	VIF
Population Density (X_1)	0.021	48,223
Area Height (X_2)	0.827	1,209
Number of Health Workers (X_3)	0.020	49,260
Amount of Rainfall (X_4)	0.527	1,898
Percentage of Households that Have Access to Adequate Sanitation (X_5)	0.357	2,798
Percentage of Poor Population (X_6)	0.363	2,753

Based on Table 3 above obtained the VIF value for each independent variable to test the multicollinearity assumption, so the hypothesis is as follows:

i. Hypothesis
   H 0 : There is no relationship between independent variables
   H 1 : There is a relationship between independent variables
ii. Significance Level α = 0.05
iii. Critical Area
   Reject H 0 if the VIF value is > 10
iv. Decision

Table 5: Decision Results of Multicollinearity Testing

Independent Variable	VIF	Criteria	Decision
Population Density (X_1)	48,223	10	Multicollinearity occurs
Area Height (X_2)	1,209	10	Multicollinearity does not occur
Number of Health Workers (X_3)	49,260	10	Multicollinearity occurs
Amount of Rainfall (X_4)	1,898	10	Multicollinearity does not occur
Percentage of Households that Have Access to Adequate Sanitation (X_5)	2,798	10	Multicollinearity does not occur
Percentage of Poor Population (X_6)	2,753	10	Multicollinearity does not occur

v. Decision
It can be seen that the VIF value of each independent variable is more than 10 and some is not more than 10. This shows that between the variables ((X_2, (X_4, (X_5, (X_6 there is no case of multicollinearity, so it is worth including in the formation of the regression model Poisson and Negative Binomial regression.

4.3. Establishment of a Poisson regression model

The Poisson regression model is a nonlinear regression model derived from the Poisson distribution which is an application of GLM which describes the relationship between the dependent variable and the independent variable, with the dependent variable in the form of discrete/count data with the assumption $E\ (y_i)$ = $Var\ (y_i)$ = (μ_i or called equidispersion (Agresti, 2002) . Poisson regression is a regression used to model the relationship between response variables and predictor variables where the distribution of the response variable follows the Poisson distribution (Aulele, 2012). Poisson regression is an analytical model that can be used to model response variables in the form of chopped data (Fitrial and Fatikhurrizqi 2021). The following is the natural logarithm mapping used to model Poisson Regression (Sma et al., 2012):

$$g(\mu_i) = \ln \mu_i = x_i'\beta \quad (2)$$

Meanwhile, the Poisson regression model can be written as follows (Sundari, 2012):

$$y_i = \exp(x_i'\beta) + \varepsilon_i \quad (3)$$

The results of parameter estimation for the Poisson regression model can be seen in Table 6. These results were obtained using R 4.2.3 software.

Table 6: Estimated Values of Poisson Regression Model Parameters

P	Estimate	Std. Error	z value	Pr(>|z|)
$\boldsymbol{\beta_0}$	0.668637	1.017131	0.657	0.51094
$\boldsymbol{\beta_2}$	0.007929	0.003698	2,144	0.03203*
$\boldsymbol{\beta_4}$	0.005951	0.002082	2,859	0.00425 **
$\boldsymbol{\beta_5}$	0.013393	0.009690	1,382	0.16692
$\boldsymbol{\beta_6}$	-0.103522	0.036160	-2,863	0.00420 **

Significant. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for poisson family taken to be 1)
Null deviance: 75.340 on 11 degrees of freedom
Residual deviance: 26,557 on 7 degrees of freedom
AIC: 85,577
Number of Fisher Scoring iterations: 5

The estimated form of the Poisson Regression model obtained is as follows:

$$\ln(\mu_i) = \beta_0 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

$$\mu_i = \exp(\beta_0 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6) \quad (4)$$
$$\mu_i = \exp(0.668637 + 0.007929X_2 + 0.005951X_4 + 0.013393X_5 - 0.103522X_6)$$

Based on Table 6, it can be seen that there are four independent variables that have a significant effect on the response at a real level of 5%, namely Height of the Area (X_2) , Number Bulk Rain (X_4), Percentage of Households that Have Access to Adequate Sanitation (X_5), and Percentage of Poor Population (X_6).

The model results show that the intercept value or dengue morbidity rate per 100,000 population in basic conditions (when all predictor variables are zero) is exp (0.668637) . In addition, a one unit increase in the height of the area (X_2) dengue fever pain. On the other hand, a one unit increase in the Percentage of Poor Population (X_6) is estimated to reduce the number of dengue fever cases.

4.4. Poisson Regression Model Overdispersion Test

Before proceeding to the Negative Binomial Regression stage, you must check whether the data is overdispersed. In the Overdispersion test results using the AER package from the Rstudio software, the p-value = 0.0006283 < α, where the α value is 0.1, so the research data experiences Overdispersion, which means the Poisson Regression model is not suitable for use, which will then be continued in the Negative Binomial step. Thus, to overcome cases of overdispersion, negative binomial regression can be used to model the DHF morbidity rate in Riau in 20-21 .

4.5. Establishment of a Negative Binomial Regression Model

One way to model calculated data with equidisperse conditions or overdispersion conditions, namely negative binomial regression (Keswari et al., 2014). The general form of Negative Binomial Regression is as follows (Utami, 2013):

$$y_i = \exp(X\beta) \quad (5)$$

Parameter estimation results for the Negative Binomial Regression model can be seen in Table 7. These results were obtained using R 4.2.3 software.

Table 7: Estimated Values of Negative Binomial Regression model parameters

P	Estimate	Std. Error	z value	Pr(>|z|)
$\boldsymbol{\beta_0}$	0.691853	1.293779	0.535	0.5928
$\boldsymbol{\beta_2}$	0.008345	0.005255	1.588	0.1123
$\boldsymbol{\beta_4}$	0.006039	0.003038	1.987	0.0469*
$\boldsymbol{\beta_5}$	0.011420	0.012849	0.889	0.3741
$\boldsymbol{\beta_6}$	-0.086578	0.040461	-2.140	0.0324*

Significant. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial (9.2786) family taken to be 1)
Null deviance: 32,350 on 11 degrees of freedom
Residual deviance: 12,535 on 7 degrees of freedom
AIC: 82,818
Number of Fisher Scoring iterations: 1
Theta: 9.28
Std. Err.: 7.25
2 x log-likelihood: -70.818

In the first Negative Binomial model, it involves all predictor variables, namely variables x 2 , x 4 , x 5 and x 6 . The first model obtained is as follows:

$$\ln(\mu_i) = \beta_0 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$
$$\mu_i = \exp(\beta_0 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6) \quad (6)$$
$$\mu_i = \exp(0.691853 + 0.008345X_2 + 0.006039X_4 + 0.011420X_5 - 0.086578X_6)$$

Based on Table 6, it can be seen that there are four independent variables that have a significant effect on the response at a real level of 5%, namely Height of the Area (X_2) , Number Bulk Rain (X_4), Percentage of Households that Have Access to Adequate Sanitation (X_5), and Percentage of Poor Population (X_6).

The model results show that the intercept value or dengue morbidity rate per 100,000 population in basic conditions (when all predictor variables are zero) is exp (0.691853) . In addition, a one unit increase in the height of the area (X_5) dengue fever pain. On the other hand, a one unit increase in the Percentage of Poor Population (X_6) is estimated to reduce the number of dengue fever cases.

4.6. Negative Binomial Model Significance Test

Parameter testing is carried out to determine the effect of predictor variables on the response variable. Parameter testing is carried out simultaneously and partially. The simultaneous test aims to see the influence of the predictor variables together on the response variable (Fitrial and Fatikhurrizqi 2021). Simultaneous testing is carried out using the likelihood ratio test with the following hypothesis. (Dhiya, 2020):

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1 : \textit{there is at least one } \beta_j \neq 0, \textit{ with } j = 1,2, \dots , k \quad (7)$$

This test uses the likelihood ratio test with the following test statistics (Dhiya 2020):

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \quad (8)$$

$L\ (\hat{\omega})$is the likelihood value for a simple model without involving independent variables and $L\ (\hat{\Omega})$ is the likelihood value for the complete model involving independent variables.

To make a decision to reject H 0, it can be seen if the value of $G > x^2_{(\alpha,v}$, this is because the test statistic G follows the Chi-Square distribution. The v value is obtained from the number of parameters in the model. Mark$x^2_{(\alpha,v)}$ can be seen in the Chi-Square table.

Meanwhile, partial testing was carried out using the Wald test, with the following hypothesis (Dhiya 2020):

$$H_0 : \beta_j = 0 \quad (9)$$
$$H_1 : \beta_j \neq 0$$

with j = 1,2, …, k with Wald test statistics as follows (Dhiya 2020):

$$W_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (10)$$

with j = 1,2, …, k and $\hat{\beta}_j$is the estimator value of βj , as well $SE(\hat{\beta}_j)$as the standard error of $\hat{\beta}_j$.

The decision taken for the Wald test is to reject $H0$ if $p - value < \alpha$ or use the value W count with the decision to reject H_0 if the value W count > $Z_{\propto/2}$ or value W count $Z_{\propto/2}$, value $Z_{\propto/2}$ can be seen in the normal distribution table. The α value used in this research is 0.05.

4.6.1. Overall Test of the Negative Binomial Regression Model

The first step in parameter estimation is the overall test (model feasibility) of negative Binomial regression, with the following output:

Based on Table 5.9 above, it can be used to determine whether simultaneously the independent variables (Altitude of Area, Number of Bulk Rain, the percentage of households that have access to adequate sanitation, and the percentage of poor people influence the dependent variable (DHF morbidity rate) so the following hypothesis test is used:

1. Hypothesis
   H0: $\beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$ (There is no influence of independent variables on dengue cases
   H1: There is at least one j with $\beta_j \neq 0$ j = 1,2, …, p (There is an influence of the independent variable ondengue cases)
2. Significance Level
   α = 0.05
3. Critical Area
   Reject H_0 if $G \geq \chi^2_{a,db}$ where α is the real level and db is the degree of freedom or reject H_0 if Sig is smaller than the value of α.
4. Decision
   Based on the chi-squares table with a significance level of 0.05 and degrees of freedom of 4, the value obtained $\chi^2_{0.05,\ 4}$ = 9.4877

2 lnL ($\widehat{\Omega}$) = −35.4 09 (this value was obtained through the RStudio program on an incomplete model or a model without independent variables) and 2 lnL ($\widehat{\omega}$)= − 41.051 (this value was obtained through the RStudio program on a complete model or a model with independent variables), so we get:

$$G = -2\ \ln\left[\frac{L\ (\widehat{\omega})}{L\ (\widehat{\Omega})}\right] = 2\ lnL\ (\ \widehat{\Omega}) - 2\ lnL\ (\widehat{\omega}) \tag{11}$$
$$G = -\ 35.409 - (-\ 41.051)$$
$$G = 5.642$$

Value of G = 5.642 < Value $\chi^2_{0.05,\ 4}$ = 9.4877
Sig value = 0.02355 < α = 0.05 which means reject H0.

5. Decision
   With a confidence level of 95%, the decision to reject H0 is obtained. So this model can be used to describe the relationship between DHF morbidity rates and regional height, number Bulk Rain, Percentage of Households that Have Access to Adequate Sanitation, and Percentage of Poor Population.

4.6.2. Partial Test of the Negative Binomial Regression Model

After carrying out the overall test, the next step is a partial test, this test is used to determine whether partially the independent variables (Height of Area, Number of Bulk Rain, the percentage of households that have access to adequate sanitation, and the percentage of poor people) influence the dependent variable (DHF morbidity rate), so the following hypothesis test is used:

1. Hypothesis
   H0: β_j = 0 (There is no influence of the independent variable on the dependent variable)
   H1: There is at least one j with $\beta_j \neq 0$ (There is an influence of the independent variable on the dependent variable)
2. Significance Level
   α = 0.0 5
3. Critical Area
   Reject H_0 if the Wald test statistic $W \geq Z_{\propto/2}$ = and Sig < α = 0.0 5
4. Decision

Table 8: Partial Test of the Negative Binomial Regression Model

Variable	B	W value	$Z\ \alpha/2$	Decision	Information
Constant	0.691853	0.53475	1.96	Accept H_0	Not significant
X_2	0.008345	1,587 97	1.96	Accept H_0	Not significant
X_4	0.006039	1.98742	1.96	Reject H_0	Significant
X_5	0.011420	0.88882	1.96	Accept H_0	Not significant
X_6	-0.086578	-2.13977	1.96	Reject H_0	Significant

5. Decision
   Based on the results of the Partial Test in Table 8, it can be seen that the variables X_4 and $W < Z_{\propto/2}$ which means accept H_0, where the value of α is 0.05 or 5%. Apart from that, the variables X_4 and X_6 so that variables X_2 and X_5 are ignored and analyzed again to get a new model without variables X_4 and X_5.

4.6.3. Significant Negative Binomial Regression Model

Parameter estimation results for the Negative Binomial Regression model can be seen in Table 9. These results were obtained using R 4.2.3 software.

Table 9: Estimated Values of Negative Binomial Regression model parameters

P	Estimate	Std. Error	z value	Pr(>|z|)
$\boldsymbol{\beta_0}$	2.019920	0.753019	2,682	0.00731 **
$\boldsymbol{\beta_4}$	0.005310	0.003546	1,498	0.13425
$\boldsymbol{\beta_6}$	-0.089901	0.033357	-2,695	0.00704 **

Significant. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial (5.0373) family taken to be 1)
Null deviance: 22.314 on 11 degrees of freedom
Residual deviance: 12,494 on 9 degrees of freedom
AIC: 83,073
Number of Fisher Scoring iterations: 1
Theta: 5.04
Std. Err.: 3.11
2 x log-likelihood: -75.073

In the first Negative Binomial model, it involves all predictor variables, namely variable X_2 and X_6. The first model obtained is as follows:

$$\begin{gathered} \ln(\mu_i) = \beta_0 + \beta_4 X_4 + \beta_6 X_6 \\ \mu_i = \exp(\beta_0 + \beta_4 X_4 + \beta_6 X_6) \\ \mu_i = \exp(2.019920 + 0.005310 X_4 - 0.089901 X_6) \end{gathered} \tag{12}$$

The model results show that the intercept value or dengue morbidity rate per 100,000 population in basic conditions (when all predictor variables are zero) is exp (2.019920). In addition, a one unit increase in the amount of rainfall (X_4) is predicted to increase the number of dengue fever cases. On the other hand, a one unit increase in the Percentage of Poor Population (X_6) is estimated to reduce the number of dengue fever cases.

4.7. Best Model

Akaike's Information Criterion was introduced in 1973 by Akaike as an approach to unbiased estimation of a modeling result. The Akaike's Information Criterion (AIC) method is a method that can be used to select the best regression model based on the Maximum Likelihood Estimation method. The best regression model is the regression model that has the smallest AIC value. The AIC value can be denoted as follows (Ramadhani et al., 2018):

$$AIC = -2\ln L(\beta) + 2k \tag{13}$$

Where: $L(\beta)$: The likelihood function of the parameter being estimated
k : Number of Parameters estimated

Table 10: Results of Akaike's Information Criterion

Model	AIC
Poisson Regression	85,577
Negative Binomial Regression	82,818
Negative Binomial Regression with significant variables (X 4 and X 6)	83,073

Based on Table 10, it can be said that the best model is the Negative Binomial Regression Model. This is because the AIC value of the Negative Binomial Regression Model is smaller than the AIC value produced by the Poisson Regression model, namely 82.818 < 85.577.

5. Conclusion

The regression analysis used is Negative Binomial Regression because there is an assumption of overdispersion in the data on DHF morbidity rates in Riau Province for 20-21 and after being analyzed using software RStudio found that the significant factors influencing dengue fever in Riau Province in 20 21 were the number Bulk Rain (X_4), and Percentage of Poor Population (X_6). The negative binomial regression model is the best model with the smallest AIC value of 82.818

References

Alan Agresti (2009). *An introduction to categorical data analysis (2nd edn).*, John Wiley & Sons, Inc., Hoboken, New Jersey. https://doi.org/10.1002/sim.3564.

Amaliana, L., Sa'adah, U., and Wardhani, N. W. S. (2018). Performance of Zero-Inflation Proportion in Zero-Inflated Negative Binomial Regression (Case Study: Tetanus Neonatorum Data in East Java). *E-Jurnal Matematika, 7,* 41. https://doi.org/10.24843/mtk.2018.v07.i01.p183.

Aulele, S. N. (2012). Modeling the Number of Infant Deaths in Maluku Province in 2010 Using Poisson Regression. *BAREKENG: Jurnal Ilmu Matematika dan Terapan, 6,* 23–27. https://doi.org/10.30598/barekengvol6iss2pp23-27.

Cameron, A. & Trivedi, P., (1990). Regression-Based Test For Overdispersion In The Poisson Model. *Journal of Econometrics, 46*(1), pp. 347-346.

Dhiya, A. Y. (2020). *Modeling Stroke and Diabetes Mellitus Patients in Padang City with Bivariate Binary Logistic Regression Model.* IX, 270–277.

Fitrial, N. H., and Fatikhurrizqi, A. (2021). Modeling the Covid-19 Morbidity Rate in Indonesia with Poisson Regression and Negative Binomial Regression Approaches. National Seminar on Official Statistics, *2020*, 65–72. https://doi.org/10.34123/semnasoffstat.v2020i1.465.

Guerdan, R. (2010). Symptoms and Signs of Dengue Fever. Journal of Dengue Medicine, *5*(2), 56-62.

Hu, Shuhua. 2007. *Akaike Information Criterion*. North Carolina State University.

Keswari, N. M. R., Sumarjaya, I. W., and Suciptawaty, N. L. P. (2014). Comparison of Negative Binomial Regression and Generalized Poisson Regression in Overcoming Overdispersion (Case Study: Number of Workers of Tile Making Business in Br. Dukuh, Pejaten Village), *E-Jurnal Matematika, 3,* 107. https://doi.org/10.24843/mtk.2014.v03.i03.p072.

Mahfudhotin, M. (2020). Regresi Generalized Poisson Untuk Memodelkan Jumlah Penderita Gizi Buruk Pada Balita di Surabaya. *Jambura Journal of Probability and Statistics, 1,* 47–56. https://doi.org/10.34312/jjps.v1i1.6876.

Mahfudhotin, U. (2020). Regresi Generalized Poisson untuk Mengatasi Overdispersi. *Jurnal Biometrika, 12*(2), 165-178.

Masyeni, S., Haque, U., & Syafruddin, D. (2021). The Relationship Between Climate Factors and Dengue Hemorrhagic Fever Incidence in Jakarta, Indonesia. Global Health Action, 14(1), 56-78.

Masyeni, S., Haque, U., & Syafruddin, D. (2021). The Relationship Between Climate Factors and Dengue Hemorrhagic Fever Incidence in Jakarta, Indonesia. *Global Health Action, 14*(1).

Nazir, M.Y. (2021). *Decrease in DHF Cases in Riau Province. Riau Sehat News, August 12, 2021. Accessed from* https://rskriausehat.com

Prahutama, A., Ningrum, D.K., & Purhadi (2017). Negative Binomial Regression Analysis of DHF Cases. Saintika Medika: Scientific Journal of Health, *13*(2), 132-142.

Prahutama, A., Sudarno, Suparti, and Mukid, M. A. (2017). Analysis of Factors Affecting Infant Mortality Rates in Central Java Using Generelized Poisson and Negative Binomial Regression. *Statistika, 5,* 1–6.

Ramadhani, N., Yanuar, F., and Yozza, H. (2018). Application of Generalized Poisson Regression in Modeling Maternal Mortality Cases in West Sumatra in 2015. *Jurnal Matematika UNAND, 7,* 112. https://doi.org/10.25077/jmu.7.2.112-117.2018.

Rochmad. (2013). Application of Newton-Raphson Method to Approach the Solution of Non Linear Equations. *Jurnal MIPA, 36,* 193–200.

Sauddin, A., Auliah, N. I., and Alwi, W. (2020). Modeling the Number of Maternal Deaths in South Sulawesi Province Using Negative Binomial Regression, Journal of MSA (Mathematics and Statistics and Its Applications), *8*, 42

Sauddin, S., Singh, Y., & Tripathi, S. (2020). Regression model for overdispersed count data in epidemiology. *Journal of Applied Quantitative Methods. 14*(2), 104-117.

Tobing, T., 2011. *Modeling Dengue Fever Cases in East Java with Poisson and Negative Binomial Models.* Bogor: Institut Pertanian Bogor.

Trapsilowati, W. (2007). *Virology of Dengue Fever. Pharmaceutical Science Magazine*, 4(1), 12-26.

Utami, T. W. (2013). Negative binomial regression analysis to overcome overdispersion of Poisson regression in dengue hemorrhagic fever cases, *Jurnal Statistika Universitas Muhammadiyah Semarang*, *1*, 0–6

Utami, W.T. (2013). *Overdispersion in Poisson Regression. Scientific*. Bulletin of Statistics, 16(2), 49-56.