Modelling Generalized Poisson Regression in the Number of Dengue Hemorrhagic Fever (DHF) in East Nusa Tenggara

Alan Prahutama, Dwi Ispriyanti, Budi Warsito*

Statistics Department, Faculty of Science and Mathematics, Diponegoro University, Semarang-Indonesia

Abstract. Regression analysis is an analysis used to model the relationship between the dependent variable (Y) and the independent variable (X). If the dependent variable is a discrete random variable, it is developed using the Poisson regression model. Poisson regression models require non-over-dispersion model assumptions. To deal with over-dispersion, a Generalized Poisson regression model was developed. Generalized Poisson regression (GPR) model is an extension of the Poisson regression model. In this study a GPR model is applied to model the number of dengue hemorrhagic fever (DHF) sufferers in East Nusa Tenggara Province in 2018. The independent variables used include percentage of poor population (X_1), population density (X_2), percentage of proper sanitation (X_3), percentage of decent homes (X_4), number of doctors (X_5), percentage of access to improved drinking water (X_6), average length of schooling (X_7), human development index (X_8). In the resulting model, Poisson regression experiences multicollinearity and overdispersion occurs. To overcome multicollinearity, variable selection is performed. Based on the measurement of the goodness of the model using AIC, the GPR model provides better accuracy than Poisson regression to model DHF in East Nusa Tenggara which is 218.5.

Keywords. Poisson regression; Generalized Poisson, Dengue Hemorrhagic Fever (DHF), East Nusa Tenggara

1. INTRODUCTION

Regression analysis is a method used to analyze the relationship between two variables, namely the dependent variable with the independent variable. In general, regression analysis is used to analyze data with the dependent variable in the form of continuous random variables. However, there are also data analyzed that the dependent variable is a discrete random variable. One regression model that can be used to analyze the relationship between the dependent variable in the form of discrete data with an independent variable is the Poisson regression model.

* Corresponding author: alan.prahutama@gmail.com

One of the assumptions that must be met in the Poisson regression model is the variance of the dependent variable whose value is equal to the average. This is in accordance with the characteristics of the Poisson distribution ie the value of variance is the same as the average value. But in the analysis of discrete data with Poisson regression models sometimes violations of assumptions can occur, where the variance value is greater than the mean value called overdispersion or the variance is smaller than the mean value called underdispersion. According to Cameron and Trivedi (1999) [1] overdispersion has the same effect as a violation of homokedasticity in the linear regression model. Homocedasticity is one of the assumptions that must be met in the classic linear regression model, where the value of the variance of residuals is constant. While the variance and the mean in Poisson regression that experiences dispersion (overdispersion / underdispersion) are

$$Var(Y) = \phi \mu \tag{1}$$

where *constant* ϕ is parameter of dispersion/ *scale parameter*. When it was overdisprepsion, then the value of ϕ is greater than one.

According to Wang and Famoye (1997) [2] one of the development methods used to overcome overdispersion is the Generalized Poisson regression model. The concept of forming this regression model is based on the generalized Poisson distribution which can explain a number of discrete data that maintain the overdispersion and underdispersion properties. Generalized Poisson regression is part of the Generalized Linear Model (GLM) that does not require the dependent variable to have a Normal distribution and does not require the homogeneity / constancy of the variance to test its hypothesis. One of the researches related to Generalized Poisson regression model is Mahama et. al (2020) [3] modeling technology adoption in increasing soybean production using the Generalized Poisson regression model. The dependent variable used is the number of technologies adopted, while significant independent variables include age, education, level of visits, and mass media via radio.

Fig 1. The Number of Dengue Hemorrhagic Fever (DHF) in East Nusa Tenggara in 2015-2019

Figure 1 shows the number of DHF sufferers in NTT from 2015 to 2019. Based on the figure, there was a significant increase in number from 2017 to 2019. In addition, in 2020, the number of DHF sufferers has almost doubled compared to 2019 which has reached the 4000s. This is certainly the concern of the NTT provincial government to deal with the problem of DHF. Not only intervention from the provincial government alone, the people of NTT must be aware of clean and healthy behavior. This is because dengue hemorrhagic fever is highly influenced by environmental, climate and weather factors, as well as the individual's

lifestyle. Therefore in this paper, the number of DHF sufferers in NTT is modeled in 2018 with the independent variables being the percentage of poor population, population density, percentage of proper sanitation, percentage of decent housing, number of doctors, percentage of access to safe drinking water, average length of school, human development index.

2. LITERATURE REVIEW

2.1. Poisson Regression

Regression analysis is a statistical tool that utilizes the relationship between two or more quantitative variables so that one variable can be predicted from other [4]. The right model for discrete data (count) is the Poisson regression model which is a nonlinear regression model (Cameron and Trivedi, 1998). Poisson regression is an application of the Generalized Linear Model (GLM) which describes the relationship between the dependent variable Y discrete poisson-distributed data with the independent variable X. The matrix form of the dependent variable, independent variables and poisson regression parameters are written as follows

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} ; \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

with y_i is dependent variable for observation i-th ($i=1,2,3, \dots, n$) and x_{ij} is independent variable ($j=1,2,3, \dots, k$), with n the number of observation and k the number of independent variable. According to Cameron dan Trivedi (1998) poisson regression model can be written as:

$$\begin{aligned} \mu_i &= \exp(x_i^T \beta) \quad (1) \\ x_i^T &= [1 \quad x_{1i} \quad x_{2i} \quad \cdots \quad x_{ki}] \\ \beta^T &= [\beta_0 \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_k] \end{aligned}$$

so

$$\exp(x_i^T \beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki})$$

with μ_i is the average number of events that occur in a certain time period in observation i-th and β is parameter of Poisson regression

The method used to estimate the parameters in Poisson regression is the Maximum Likelihood Estimation (MLE) method, namely by maximizing the likelihood function [5]. Estimating parameters for poisson regression are as follows

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \\ L(\beta) &= \prod_{i=1}^n \frac{e^{-\sum_{i=1}^n e^{(x_i^T \beta)}} (e^{\sum_{i=1}^n y_i x_i^T \beta})}{\prod_{i=1}^n y_i!} \quad (2) \end{aligned}$$

Furthermore, the likelihood function in equation (2) is changed in the form of natural logarithms (ln-likelihood) to equation (3)

$$\begin{aligned} \ln L(\beta) &= \ln \left(\frac{e^{-\sum_{i=1}^n e^{(x_i^T \beta)}} (e^{\sum_{i=1}^n y_i x_i^T \beta})}{\prod_{i=1}^n y_i!} \right) \quad (3) \\ \ln L(\beta) &= -\sum_{i=1}^n e^{(x_i^T \beta)} + \sum_{i=1}^n y_i x_i^T \beta - \sum_{i=1}^n \ln(y_i!) \quad (4) \end{aligned}$$

Then the equation (4) derivative to β

$$\frac{\partial \ln L(\beta)}{\partial \beta} = -\sum_{i=1}^n \mathbf{x}_i^T e^{(\mathbf{x}_i^T \beta)} + \sum_{i=1}^n y_i \mathbf{x}_i \quad (5)$$

Equation (5) is equated with zero to get the parameter estimator. However, this method does not get explicit results, so that the alternative to solving equation (5) is the Newton-Raphson iteration method. The purpose of the numerical iteration method is to maximize the ln-likelihood function [5].

2.1.1. The Testing of Parameter Poisson Regression Model Simultaneously

A good model is a model that has a small residual value. In poisson regression deviance is used to determine the best model. Finding deviance values from the model is obtained by the initial step of simultaneous testing (overall). Simultaneous testing the parameters of the Poisson regression model are used to see the suitability of the resulting model.

The concurrent testing hypothesis is as follows

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

$$H_1 = \text{at least } \beta_j \neq 0, j = 1, 2, \dots, k$$

The test statistic used was obtained from the Myers Maximum Likelihood Ratio Test (MLRT) method.

$$D(\beta) = 2 \left(-\sum_{i=1}^n e^{\mathbf{x}_i^T \hat{\beta}} + \sum_{i=1}^n y_i \mathbf{x}_i^T \hat{\beta} + n e^{\hat{\beta}_0} - \hat{\beta}_0 \sum_{i=1}^n y_i \right) \quad (6)$$

Devians is approaches from χ^2 distribution with the large sample and it has degree of freedom v . For criteria of testing is reject H_0 for level significant α for $D(\beta) > \chi^2_{(v, \alpha)}$. Deviance values will be smaller if the parameters in the model increases [6]. The smaller the devians, the smaller the error rate produced, so the model becomes more precise.

2.1.2. The Testing of Parameter Poisson Regression Model Partially

The parameters generated from the assessment process may not all have a significant influence on the dependent variable. Partial parameter testing is used to determine the parameters that have a significant influence on the model. The hypothesis used in the partial test are:

$$H_0: \beta_j = 0 \text{ (no significant)}$$

$$H_1: \beta_j \neq 0 \text{ dengan } j=1, 2, \dots, k \text{ (significant)}$$

The statistics testing as follow as:

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \quad (7)$$

where $se(\hat{\beta}_j)$ is standard error of estimation parameter of β . Criteria of rejection is reject H_0 for $|t| > t_{(n-k-1, \frac{\alpha}{2})}$

2.2. Over-dispersion

Poisson regression has an assumption that is equidispersion where the dependent variable used has the same mean and variance of value. One violation of that assumption is overdispersion where the value of variance is greater than the mean value. Overdispersion causes the alleged child of the regression parameters to be consistent but inefficient. If the

equidispersion assumptions are not met then the impact on the standard error value is under estimate, so the conclusions obtained are invalid [7].

According to McCullagh and Nelder (1998) [6] to detect overdispersion in Poisson regression, Pearson Chi-Square can be divided into degrees of freedom. The formula for Pearson Chi-Square statistical tests is as follows:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\text{var}(\mu_i)} \text{ with ratio of dispersion } \alpha = \frac{\chi^2}{n-k-1} \quad (8)$$

2.3. Generalized Poisson Regression

Generalized Poisson regression (GPR) model is model that used to count data to overcome overdispersion or underdispersion. GPR model have μ parameter and additional parameter α . The distribution of GPR model can be written as [8]:

$$f(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1 + \alpha \mu_i} \right)^{y_i} \frac{(1 + \alpha \mu_i)^{y_i - 1}}{y_i!} \exp \left(\frac{-\mu_i(1 + \alpha y_i)}{1 + \alpha \mu_i} \right) \quad (9)$$

Means and variance of Generalized Poisson distribution are $E(Y) = \mu$ and $\text{var}(Y) = \mu(1 + \alpha\mu)^2$. If the value of α is zero then the model formed is Poisson regression. If the value of $\alpha > 0$ then it called overdispersion, while the value of $\alpha < 0$ then it called underdispersion. The method that used for estimate parameter of GPR model is Maximum Likelihood Estimation (MLE) with the combination of Newton-Raphson iteration. The likelihood of GPR model can be written as:

$$L(\alpha, \beta) = \prod_{i=1}^n \left\{ \left(\frac{\mu_i}{1 + \alpha \mu_i} \right)^{y_i} \frac{(1 + \alpha \mu_i)^{y_i - 1}}{y_i!} \exp \left(\frac{-\mu_i(1 + \alpha y_i)}{1 + \alpha \mu_i} \right) \right\} \quad (10)$$

Furthermore, the likelihood function in equation (10) is changed in the form of natural logarithms (ln-likelihood) to equation (11)

$$\ln L(\alpha, \beta) = \sum_{i=1}^n \ln \left[\left(\frac{\mu_i}{1 + \alpha \mu_i} \right)^{y_i} \frac{(1 + \alpha \mu_i)^{y_i - 1}}{y_i!} \exp \left(\frac{-\mu_i(1 + \alpha y_i)}{1 + \alpha \mu_i} \right) \right] \quad (11)$$

if $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, then

$$\ln L(\alpha, \beta) = \sum_{i=1}^n \left[y_i (\ln e^{\mathbf{x}_i^T \boldsymbol{\beta}} - \ln(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})) + (y_i - 1) \ln(1 + \alpha y_i) - \ln y_i! - \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}} (1 + \alpha y_i)}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right]$$

For getting the estimation parameter β and α then it was derivative to ln likelihood function of β and α . The first derivative from ln likelihood function as follow as:

$$\frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} = \sum_{i=1}^n \left\{ \frac{-e^{\mathbf{x}_i^T \boldsymbol{\beta}} y_i}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} + \frac{y_i(y_i - 1)}{1 + \alpha y_i} - \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}} (y_i - e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right\}$$

Because it has not been able to provide a solution then another solution is used by the numerical iteration method called the Newton-Rhapson method. The purpose of the numerical iteration method is to maximize the ln-likelihood function. In the Newton-Rhapson method the first and second derivatives of the ln-likelihood function are required. For estimating the α and β with this method, it was needed prior estimation from α and β . According Wang dan Famoye (1997) the prior estimation of α getting the zero value or can used the Chi square divided to degree of freedom.

2.3.1. The Testing of Parameter Generalized Poisson Regression Model Simultaneously

The method that can be used to determine test statistics in parameter testing is the Maximum Likelihood Ratio Test (MLRT) method. The hypothesis for the simultaneous test as follows as:

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

$$H_1 = \text{at least there is } \beta_j \neq 0, j = 1, 2, \dots, k$$

The test statistic used is obtained from the Maximum Likelihood Ratio Test (MLRT) method using the deviance value.

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

$$D(\beta) = 2 \left(\sum_{i=1}^n y_i (\ln e^{x_i^T \hat{\beta}} - \ln(1 + \alpha e^{x_i^T \hat{\beta}})) - \sum_{i=1}^n \frac{e^{x_i^T \hat{\beta}} (1 + \alpha y_i)}{1 + \alpha e^{x_i^T \hat{\beta}}} \right) - \left(\sum_{i=1}^n y_i (\ln e^{\hat{\beta}_0} - \ln(1 + \alpha e^{\hat{\beta}_0})) - \sum_{i=1}^n \frac{e^{\hat{\beta}_0} (1 + \alpha y_i)}{1 + \alpha e^{\hat{\beta}_0}} \right)$$

For criteria of testing is reject H_0 for $D(\beta) > \chi^2_{(v, \alpha)}$.

2.3.2. The Testing of Parameter Generalized Poisson Regression Model Partially

Partial parameter testing is used to determine the parameters that have a significant influence on the model. The hypothesis used in the partial test as:

$$H_0: \beta_j = 0 \text{ (not significant)}$$

$$H_1: \beta_j \neq 0 \text{ with } j=1, 2, \dots, k \text{ (significant)}$$

the statistics testing as follow as:

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{12}$$

Criteria of testing, reject H_0 for $|t| > t_{(n-k-1, \frac{\alpha}{2})}$.

3. METHODOLOGY OF RESEARCH

There are 22 regencies / cities in the province of East Nusa Tenggara. The data from Central Bureau of Statistics of East Nusa Tenggara Province. The dependent variable used is the number of DHF sufferers in NTT province in 2018. The independent variables used include the percentage of poor population (X_1), population density (X_2), percentage of proper sanitation (X_3), percentage of decent homes (X_4), number of doctors (X_5), percentage of access to improved drinking water (X_6), average length of schooling (X_7), human development index (X_8). The steps of modeling using the Generalized Poisson Regression model are as follows:

1. Modelling using Poisson regression.
2. Parameter testing of Poisson regression model simultaneously
3. Parameter testing of Poisson regression model partially
4. The testing of over-dispersion or under-dispersion
5. Modelling used Generalized Poisson regression model
6. Parameter testing of Generalized Poisson regression model simultaneously
7. Parameter testing of Generalized Poisson regression model partially

4. RESULTS AND DISCUSSIONS

Descriptive statistics of the data used are as follows in Table 1. Table 1 shows the descriptive statistics of the research used, based on the table the number of DHF sufferers in East Nusa Tenggara in 2018 averaged 72.68 people with the highest number of patients as many as 539 people, while the lowest was 0. The average percentage of poor people reached 22.04%. The average population density reached 227.14 people / km2, while the percentage of proper sanitation reached 49.28%. The percentage of livable homes reached 71.66%. The average length of school is 7.38 years.

Table 1. Descriptive statistics of research variable

Variable	Mean	Minimum	Maximum	Variance
Y	72.68	0.00	539.00	15127.18
X ₁	22.04	9.61	34.85	52.89
X ₂	227.14	36.00	2351.00	227718.12
X ₃	49.28	10.71	81.36	565.87
X ₄	71.66	47.67	93.88	190.76
X ₅	65.77	9.00	469.00	8627.14
X ₆	80.85	45.30	98.95	215.30
X ₇	7.38	5.96	11.47	1.25
X ₈	63.23	55.79	78.84	18.56

4.1.Poisson Regression Model for The Number Dengue Herrmohagic Fever (DHF) in East Nusa Tenggara

Data on the number of patients with Dengue Hemorrhagic Fever (DHF) is a count data with a very small probability and the amount of data that has a certain time interval, so that the appropriate analysis is poisson regression modeling. To get the estimated parameters of the Poisson regression model the Maximum Likelihood Estimation (MLE) method is used with Newton-Raphson iteration. The resulting model is as follows:

$$\hat{\mu} = exp (-12.25 - 0.158X_1 + 0.000945X_2 - 0.0438X_3 + 0.0272X_4 - 0.007085X_5 - 0.0232X_6 - 1.52X_7 + 0.5202X_8)$$

4.1.1. Parameter Testing Simultaneously

The hypothesis for simultaneous testing is as follows
H₀ = β₁ = β₂ = ... = β_k = 0
H₁ = at least one of β_j ≠ 0, j = 1,2, ..., k
with significant level is α = 0.05 and criteria of testing is reject H₀ for D(β) > χ²_(v,α). Based on results, the value of D(β) = 1560.9 is greater than χ²_(13,0.05) = 22.36 so H₀ was rejected. So that, at least one of β_j ≠ 0. Therefore, this indicates that simultaneously the independent variable affected DHF sufferers in East Nusa Tenggara in 2018

4.1.2. Parameter Testing Partially

The next step is to partially test the parameters. The hypothesis of partial parameter testing is as follows
 $H_0 : \beta_j = 0$ (parameter is not significant), where $j = 1,2,3$
 $H_1 : \beta_j \neq 0$, (parameter is significant)
with significant level is $\alpha = 0.05$ for criteria of testing is reject H_0 for $|Z| > Z_{(\frac{\alpha}{2})} = 1.96$. The results as follow as:

Table 2. The results of significant parameter testing for Poisson regression

Parameter	Coefficient	Standard Error od Estimate parameter	Z-Value	Probability	Conclusion
Intercept	-12.25	1.373	-8.924	$< 2 \times 10^{(-16)}$	Significant
X ₁	-0.158	0.01137	-13.893	$< 2 \times 10^{(-16)}$	Significant
X ₂	0.000945	0.000328	2.879	0.00399	Significant
X ₃	-0.0438	0.002414	-18.134	$< 2 \times 10^{(-16)}$	Significant
X ₄	0.0272	0.00378	7.19	$6.47 \times 10^{(-13)}$	Significant
X ₅	-0.007085	0.00163	-4.351	$1.36 \times 10^{(-5)}$	Significant
X ₆	-0.0232	0.00539	-4.3	$1.71 \times 10^{(-5)}$	Significant
X ₇	-1.52	0.0992	-15.297	$< 2 \times 10^{(-16)}$	Significant
X ₈	0.5202	0.0308	16.889	$< 2 \times 10^{(-16)}$	Significant

Based on Table 2 it appears that all independent variables are significant for the model. In poisson regression obtained AIC value of 1677.2. Furthermore, multicollinearity testing will be performed on all independent variables. Obtained the Variance Influence Factors (VIF) values as follows:

Table 3. Multicollinearity testing

Variables	VIF value	Conclusion
X ₁	12.48	Multicollinearity
X ₂	104.548	Multicollinearity
X ₃	6.2	Non-Multicollinearity
X ₄	3.609	Non-Multicollinearity
X ₅	92.035	Multicollinearity
X ₆	7.656	Non-Multicollinearity
X ₇	42.039	Multicollinearity
X ₈	49.066	Multicollinearity

Multicollinearity occurs when the VIF value> 10. Based on Table 3, the variables that occur multicollinearity are variables X₁, X₂, X₅, X₇, and X₈. To handle multicollinearity assumptions, a variable selection process is carried out. The variable selection process is done by removing the variable that has the highest multicollinearity value, namely X₂. Furthermore, it is modeled again using Poisson regression, until we get a variable that does not contain multicollinearity. Obtained variables that do not contain multicollinearity such as X₁, X₃, X₄, X₅ and X₆. Poisson regression model is obtained as follows in Table 4.

Then the Poisson regression model formed is as follows

$$\hat{\mu} = \exp(6.7256 - 0.094X_1 - 0.033X_3 + 0.004X_4 + 0.00228X_5 - 0.0256X_6)$$

In the Poisson regression model, the variables X_2 , X_7 and X_8 are not included in the model, because these variables experience multicollinearity. The model produces an AIC value of 986.4, the value is smaller than that involving all variables. This shows that the selection of variables made provides a measure of the greater good of the model. The next step is overdispersion testing for the formed Poisson regression model

Table 4. The estimation of Poisson Regression model with non-multicollinearity’s variable

Variables	Coefficient of parameter	Standard Error od Estimate parameter	Z-Value	Probability	Conclusion
Intercept	6.7256	0.461	14.575	$< 2 \times 10^{(-16)}$	Significant
X_1	-0.094	0.008	-11.68	$< 2 \times 10^{(-16)}$	Significant
X_3	-0.033	0.00167	-19.828	$< 2 \times 10^{(-16)}$	Significant
X_4	0.004	0.00304	13.935	$< 2 \times 10^{(-16)}$	Significant
X_5	0.00228	0.00021	10.708	$< 2 \times 10^{(-16)}$	Significant
X_6	-0.0256	0.00367	-6.974	3.09×10^{-12}	Significant

4.1.3. Overdispersion testing for Poisson Regression Model

Analysis of the case of overdispersion is used to determine whether the Poisson regression model obtained meets the assumptions.

Table 5. The *Pearson Chi-Square* value for Poisson Regression Model

Criteria	Value	Degree of Freedom (DF)	Value/DF
<i>Pearson Chi-Square</i>	2934.6	22	133.4

Table 5 shows the *Pearson Chi-Square* value of the Poisson regression model is 2934.6. If the value is divided by the degree of freedom it will produce a dispersion value of 133.4 whose value is greater than 0, which indicates overdispersion. From this it can be concluded that the data on the number of DHF sufferers in East Nusa Tenggara in 2018 experienced overdispersion. To overcome this, Generalized Poisson regression modeling is used.

4.2.Generalized Poisson Regression (GPR) Model for Number of Dengue Herrmohagic Fever (DHF) in East Nusa Tenggara

In the case of the number of DHF sufferers in East Nusa Tenggara in 2018 experiencing overdispersion cases, this can be overcome by using Generalized Poisson Regression (GPR) modeling. To get the estimated parameters of the GPR model the Maximum Likelihood Estimation (MLE) method is used with Newton-Raphson iteration.

4.2.1. Parameter Testing Simultaneously

The hypothesis for simultaneous testing is as follows
 $H_0 = \beta_1 = \beta_2 = \dots = \beta_k = 0$
 $H_1 = \text{at least one of } \beta_j \neq 0, \quad j = 1, 2, \dots, k$
 with significant level is $\alpha = 0.05$ and criteria of testing is reject H_0 for $D(\beta) > \chi^2_{(v,\alpha)}$. Based on results, the value of $D(\beta) = 204.5$ is greater than $\chi^2_{(13,0.05)} = 22.36$ so H_0 was rejected. So that, at least one of $\beta_j \neq 0$. Therefore, this indicates that simultaneously the independent variable affected DHF sufferers in East Nusa Tenggara in 2018 use GPR model.

4.2.2. Parameter Testing Partially

Based on the results of simultaneous testing which shows that the independent variable influences the number of DHF sufferers in NTT in 2018 using the GPR model. The next step is to partially test the parameters. The hypothesis of partial parameter testing is as follows
 $H_0 : \beta_j = 0$ (parameter is not significant), where $j = 1, 2, 3$
 $H_1 : \beta_j \neq 0$, (parameter is significant)
 with significant level is $\alpha = 0.05$ for criteria of testing is reject H_0 for $|t| > t_{(n-k-1, \frac{\alpha}{2})} = 2.11$ or probability less than significant level. The results as follow as:

Table 6. The results of significant parameter testing for Generalized Poisson regression

Variables	Coefficient of parameter	Standard Error od Estimate parameter	t-Value	Probability	Conclusion
Intercept	-22.274	5.52919	-4.02844	0.000686133	Significant
X ₁	-0.5654	0.09505	-5.94845	6.42303 x 10 ⁽⁻⁶⁾	Significant
X ₃	0.02393	0.02258	1.059787	0.222483137	Not-Significant
X ₄	-0.8163	0.12772	-6.39132	2.25633 x 10 ⁽⁻⁶⁾	Significant
X ₅	0.6012	0.0866	6.942263	6.32871 x 10 ⁽⁻⁷⁾	Significant
X ₆	0.9602	0.1465	6.554266	1.54359 x 10 ⁽⁻⁶⁾	Significant

Table 6 shows a partial test of Generalized Poisson Regression models. Based on Table 6, all variables are significant except X3. It shows that the percentage of households with proper sanitation does not affect the generalized Poisson regression model. However, when compared to the Poisson regression model with the GPR model, with the same number of variables X1, X3, X4, X5 and X6, the GPR model has a smaller AIC than the Poisson regression model of 218.5.

For the interpretation of the GPR model, it cannot be done partially. For example, the X5 variable is the number of doctors, in theory the greater number of doctors should be able to reduce the number of DHF sufferers in NTT. This shows that the parameter coefficient should be negative. However, the model obtained from the GPR model and Poisson regression shows that the parameter coefficient is positive. Therefore, the interpretation of the GPR model or Poisson regression is done together by involving variables X1, X3, X4, X5 and X6. In the GPR model, the X3 variable is not significant, so it can be assumed that the value is zero. The GPR model formed is as follows:

$$\hat{\mu} = \exp (-22.274 - 0.5654X_1 + 0.02393X_3 - 0.8163X_4 + 0.6012X_5 + 0.9602X_6)$$

5. CONCLUSION

Based on data on the number of DHF sufferers in East Nusa Tenggara, involved unevenly distributed, therefore it is necessary to do a mapping analysis of the spread of dengue fever in East Nusa Tenggara. In Poisson and Generalized Poisson regression (GPR) modeling that has been done does not provide information on the spread of DHF by region, so it is necessary to do further analysis related to generalized Poisson modeling by region. In this modeling, handling Poisson regression model assumptions provides a better measure of model goodness, so in Poisson regression analysis, assumption testing is needed. For the interpretation of Poisson and GPR regression models, all variables are involved.

Acknowledgement

This research was supported by the Institute of Research and Community Service's Diponegoro University for scheme the development of research and its application with Number of Contract 233-103 / UN7.6.1 / PP / 2020.

REFERENCES

1. A.C. Cameron and K.P. Trivedi, *Regression Analysis of Count Data*, (USA: Cambridge University Press, 1998)
2. W. Wang and F. Famoye, *J. of Population Economics* 10, 3 (1997).
3. J.A. Mahama, F.N. Awuni, S.B. Mabe, Azumah, *Heliyon* 6 (2020).
4. D. N. Gujarati, *Basic Econometrics (4th ed)*, (New York: The McGraw Hill Companies, 2004)
5. M. V. Myers, *Generalized Linear Model with Applications in Engineering and Sciences, 2th Edition* (New Jersey: John Wiley & Sons, 2010).
6. P. McCullagh, and J.A. Nelder. 1989. *Generalized Linear Models*. Second Edition. (New York: Chapman and Hall. 1989)
7. Cupal, M., Deev O., Linnertova, D., *Procedia Economics and Finance* 23 (2015)
8. F. Famoye, *Restricted Theor.&Meth* 22, 5 (1993)