BAREKENG: Jurnal Ilmu Matematika dan Terapan
June 2022 Volume 16 Issue 2 Page 471–482
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss2pp471–482

# NEGATIVE BINOMIAL REGRESSION AND GENERALIZED POISSON REGRESSION MODELS ON THE NUMBER OF TRAFFIC ACCIDENTS IN CENTRAL JAVA

**M. Al Haris[1], Prizka Rismawati Arum[2*]**

*[1,2] Statistics Department, Faculty of Mathematics and Natural Sciences, Universitas Muhammadiyah Semarang Kedungmundu St., No. 18, Semarang City, 50273, Indonesia*

*Corresponding author's e-mail: [2*] prizka.rismawatiarum@unimus.ac.id*

***Abstract.*** *Traffic accidents that always increase along with the increasing population growth and the number of vehicles impact the national economy. The number of traffic accidents is a count data that a Poisson distribution can approximate. The Poisson regression model often found violations of the overdispersion assumption by modeling the factors that affect the number of traffic accidents. Alternative models proposed to overcome the emergence of overdispersion in the Poisson regression model are the Generalized Poisson Regression and Negative Binomial Regression Models. Based on the analysis results, it was found that the overdispersion assumption violates the Poisson regression model, and the Generalized Poisson regression model is the best because it has the smallest AIC value of 485.50. Factors that significantly affect the number of traffic accidents in Central Java Province are the percentage of adolescents and the percentage of accidents occurring in the road area of the district/city.*

***Keywords:*** *Generalized Poisson Regression, Negative Binomial Regression, Overdispersion, Poisson Regression.*

***Article info:***

*Submitted: 21st January 2022* *Accepted: 28th March 2022*

***How to cite this article:***

M. A. Haris and P. R. Arum, "NEGATIVE BINOMIAL REGRESSION AND GENERALIZED POISSON REGRESSION MODELS ON THE NUMBER OF TRAFFIC ACCIDENTS IN CENTRAL JAVA", *BAREKENG: J. Il. Mat. & Ter.*, vol. 16, iss. 2, pp. 471-482, June, 2022.

# 1. INTRODUCTION

Land transportation is one of the important and decisive sectors in supporting the successful implementation of development in Indonesia [1]. The existence of good transportation will affect economic distribution and economic growth. However, along with the increase in population, the demand for vehicles increases. It impacts many transportation problems, one of which is traffic accidents [2], [3]. Central Java Province has the second-highest number of traffic accidents after East Java Province. Statistical data shows the number of accidents in Central Java in 2018 was 19,016 cases with an estimated material loss of Rp.14,138,632,000. This figure has increased compared to 2017 data with the number of accidents, 17,552 cases, and material losses of Rp. 12,351,991,000 [4], [5]. The high number of traffic accidents in Central Java Province and the estimated amount of material losses caused imply that supervision and prevention are very important to reducing traffic accidents. One approach that can be taken is to identify the influencing factors and how much influence they have. The occurrence of traffic accidents is caused by several factors, including human factors, vehicles, road conditions, and weather [2], [6], [7].

The number of traffic accidents is a count data that can be approximated by a Poisson distribution [8]. Research problems that arise in modeling using Poisson regression often find violations of assumptions in Poisson regression; namely, the variance value is greater than the mean value (overdispersion) [9]. Overdispersion in the Poisson regression model resulted in the standard error for the estimated parameters being biased downwards, causing errors in concluding to determining the factors that influence the number of accidents. Handling the problem of overdispersion in the Poisson model can be overcome by alternative models, namely the Negative Binomial Regression and Generalized Poisson Regression models. [10]–[12].

Many previous studies on overdispersion have been carried out, including Prahutama, Ispriyanti, and Warsito [12]. They analyzed the factors that influence the number of patients with Dengue Hemorrhagic Fever (DHF) in East Nusa Tenggara using the Generalized Poisson Regression method. This study resulted in the Generalized Poisson Regression model being better than the Poisson regression model, which experienced overdispersion problems. Melliana et al. [10] compared Negative Binomial Regression and Generalized Poisson Regression to deal with the problem of overdispersion. The results show that the Generalized Poisson Regression model is better for modeling cervical cancer data that violates the overdispersion assumption. This research is focused on modeling the number of traffic accidents in Central Java with the factors that influence it with the Generalized Poisson Regression and Negative Binomial Regression models to overcome the overdispersion problem that appears in the Poisson regression model.

# 2. RESEARCH METHOD

This section will explain about Poisson Regression model, overdispersion, Negative Binomial Regression model, and Generalized Poisson Regression model. Furthermore, the research data and the steps taken in the research are presented.

## 2.1 Poisson Regression

Poisson regression model is usually used to model counting data which has a small chance of occurrence with the occurrence depending on a certain time interval or a certain area. The probability mass function of the Poisson distribution is as follows: [13]:

$$P(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!}; \qquad y = 0,1,2,\ldots \text{ and } \mu > 0 \tag{1}$$

$\mu$ is the average number of events in a given interval. The expected value and variance of the Poisson distribution are:

$$E[y] = Var\,[y] = \mu$$

The Poisson regression model can be written as follows:

$$y_i = \mu_i + \varepsilon_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) + \varepsilon_i \quad i = 1,2,\ldots,n \tag{2}$$

Poisson regression model is a nonlinear model, so the process of estimating the regression coefficient parameters uses iteration with the Newton-Raphson method. [12], [14], [15].

## 2.2 Generalized Poisson Regression

The Generalized Poisson distribution is a development of the Poisson distribution which is useful for modeling count data that has overdispersion or underdispersion. It is defined that the response variable Y which has a Generalized Poisson distribution has the following probability distribution function [16][13][15]:

$$P(y;\mu,\phi)=\left(\frac{\mu}{1+\phi\mu}\right)^{y}\frac{(1+\phi y)^{y-1}}{y!}\exp\left[\frac{-\mu(1+\phi y)}{1+\phi\mu}\right];` \quad y=0,1,2,\ldots \tag{3}$$

with the average value of $\mu$ and variety $\mu(1+\phi\mu)^2$.

The equidispersion condition occurs when the value of $\phi=0$, so that the Generalized Poisson distribution returns to the form of a Poisson distribution as follows:

$$\begin{aligned} P(y;\mu,\phi) &= \left(\frac{\mu}{1+0\mu}\right)^{y}\frac{(1+0y)^{y-1}}{y!}\exp\left[\frac{-\mu(1+0y)}{1+0\mu}\right]; \\ &= \frac{\mu^y}{y!}\exp[-\mu]\,; \end{aligned} \tag{4}$$

The Generalized Poisson regression model can be written as follows:

$$log(\mu_i)=\boldsymbol{x}_i'\boldsymbol{\beta}=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_kx_{ik} \tag{5}$$

Where, $x_i$: dimensional vector $k-1$

$\beta$: $k$-dimensional vector of the regression parameter

Generalized Poisson regression can be used to overcome overdispersion or underdispersion. The condition of $Var\,(y)<E(y)$ in Poisson regression, if $\phi>0$ indicates overdispersion, and if $\phi<0$ indicates underdispersion. Assumption of dispersion parameters and coefficients of Generalized Poisson regression was carried out using the maximum likelihood assumption method. The probability function is obtained by multiplying the generalized Poisson probability function. The probability natural logarithm function of the Generalized Poisson distribution is as follows:

$$l(y|\beta,\phi)=\sum_{i=1}^{n}y_i ln\left(\frac{\mu_i}{1+\phi\mu_i}\right)+(y_i-1)ln(1+\phi y_i)-\frac{\mu_i(1+\phi y_i)}{1+\phi\mu_i}-ln(y!) \tag{6}$$

Maximizing the function $l(y|\beta,\phi)$ can be done by finding the derivative of each parameter and then equating it with zero [6][12][13].

## 2.3 Negative Binomial Regression

The Negative binomial distribution can be explained as an approach to an experiment that wants to see the magnitude of the probability of $r$ successes after previously appearing some failed events. Suppose $r$ is the number of successful events, and the random variable $Y$ is the number of failed events before the *r-th* success event, the form of a Negative Binomial distribution with a probability of success is *p,* and a probability of failure of $(1-p)$ is as follows [11]:

$$P(Y=y|r,p)=\binom{y-1}{r-1}p^r(1-p)^{y-r};\quad Y=r,r+1,r+2,\ldots \tag{7}$$

with the expected value and variance is:

$$E[y]=r\frac{(1-p)}{p}\text{ dan } Var[y]=r\frac{(1-p)}{p^2} \tag{8}$$

The Negative Binomial Regression model is one approach to solve the overdispersion problem based on the Poisson-Gamma mixture model [14]. Assuming that there is $\delta$ Gamma-spreading variable with a mean of 1 and $\phi$ variance in the mean of a Poisson distribution, for example, *m* is a source of unobserved variance. Then, the mean value of a mixed Poisson-Gamma distribution is:

$$E(y_i)=\widetilde{\mu_i}=exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}+m_i\right)=exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)exp(m_i)=\mu_i\delta_i \tag{9}$$

with $\mu_i=exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)$ is the mean of the Poisson model. The probability function for the mixed Poisson-Gamma distribution can be written as: [8]:

$$f(y_i|x_i,\beta,\delta_i) = \frac{e^{-(\mu_{i,}\delta)}(\mu_{i,}\delta_i)^{y_i}}{y_i!} \tag{10}$$

The variable $\delta_i$ spreads Gamma with parameters $\alpha$ and $\beta$. The Gamma probability function is:

$$g(\delta_i) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \delta_i^{\alpha-1} e^{-\delta_i/\beta} \tag{11}$$

with the expected value $E(\delta_i) = \alpha\beta$, and assumption of $E(\delta_i) = 1$ then $\alpha = 1/\beta$. If the parameter $\alpha = 1/\phi$, the Gamma probability function becomes:

$$g(\delta_i) = \frac{1}{\phi^{\phi^{-1}} \Gamma(\phi^{-1})} \delta_i^{\phi^{-1}-1} e^{-\delta_i/\phi} \tag{12}$$

The mixed Poisson-Gamma distribution can be obtained by integrating the variable $\delta_i$ into the Poisson probability function as follows:

$$f(y_i|x_i,\beta,\phi) = \int_0^{\propto} f(y_i|\phi,\delta_i)g(\delta_i)d\delta_i = \frac{\Gamma(y_i+\phi^{-1})}{\Gamma(\phi^{-1})y_i!}\left(\frac{\phi\mu_i}{1+\phi\mu_i}\right)^{y_i}\left(\frac{1}{1+\phi\mu_i}\right)^{\phi^{-1}} \tag{13}$$

with

$$\phi > 0,\ E[y_i] = \mu_i \qquad \text{and} \qquad Var[y_i] = \mu_i + \phi\mu_i^2. \tag{14}$$

The assumption of the Negative Binomial Regression coefficient parameter was carried out using the maximum likelihood assumption method. The probability function of the negative binomial distribution is:

$$L(\boldsymbol{\beta},\phi|y,\boldsymbol{x}) = \prod_{i=1}^{n}\left\{\frac{\Gamma(y_i+\phi^{-1})}{\Gamma(\phi^{-1})y_i!}\left(\frac{\phi\mu_i}{1+\phi\mu_i}\right)^{y_i}\left(\frac{1}{1+\phi\mu_i}\right)^{\phi^{-1}}\right\} \tag{15}$$

and the natural logarithm of the probability function is as follows:

$$lnL(\boldsymbol{\beta},\phi|y,\boldsymbol{x}) = \sum_{i=1}^{n}\left\{ln\left[\frac{\Gamma(y_i+\phi^{-1})}{\Gamma(\phi^{-1})\Gamma(y_i+1)}\right] - (y_i+\phi^{-1})ln(1+\phi\mu_i) + y_i ln(\phi\mu_i)\right\} \tag{16}$$

### 2.4 Best Model Selection

The selection of the best model among the Poisson Regression, Negative Binomial Regression, and Generalized Poisson Regression models was carried out with the Akaike's Information Criteriation (AIC) value. AIC is formulated as follows:

$$AIC = -2(\mathcal{L} - k)$$

where $\mathcal{L}$ is the log likelihood model, and $k$ is the number of parameters in the model. The best model is determined by the model that has the smallest AIC value [14], [17].

### 2.5 Data Source

The data used in this study was secondary data from the Central Java Central Statistics Agency in 2018 [5]. The research unit observed was each district/city in Central Java Province in 2018, with the number of traffic accidents in 2018 as the response variable.

### 2.6 Analysis Steps

The steps taken by researchers in this study were as follows:

1. Exploring data to determine the characteristics of the value of the response variable and the explanatory variable among districts/cities in Central Java.

2. Identifying and resolving violations of the multicollinearity assumption on the explanatory variables.
3. Checking the overdispersion violation on the Poisson regression model.
4. Modeling data with Negative Binomial Regression
5. Modeling data with Generalized Poisson regression
6. Selecting the best model based on the smallest AIC value
7. Interpreting the best model.

## 3. RESULT AND DISCUSSION

Central Java Province is located in the central part of the island of Java, which West Java Province borders in the west, the Indian Ocean and the Special Region of Yogyakarta Province in the south, and East Java Province and the Java Sea in the north. It is located between 5.40 to 8.30 south latitude 108.30 to 111.30 east longitude. Its area is 32,544.12 km2 which is administratively divided into 29 regencies and 6 cities, spread into 573 sub-districts and 8576 villages/kelurahan (urban villages). The largest area is the Cilacap district, while the smallest area is city [5].

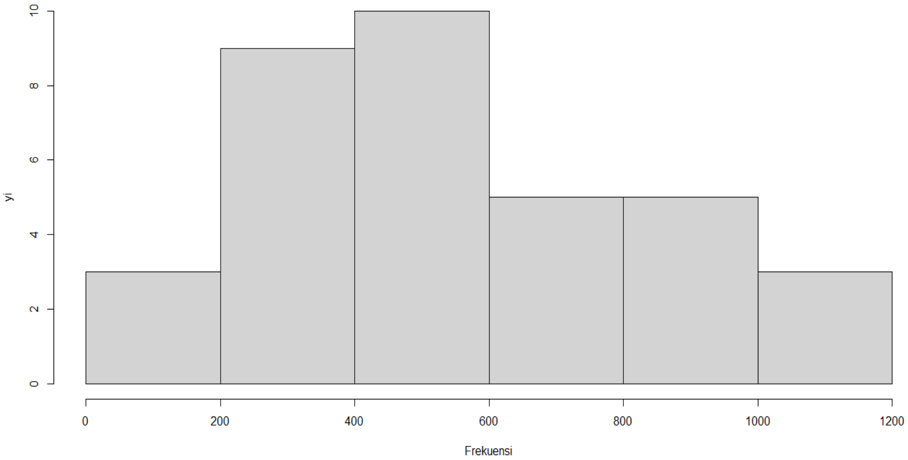


**Figure 1. Histogram Plot of Traffic Accident Rate Data in Central Java in 2018**

Data exploration is useful for studying the characteristics of the data to make it easier to determine the appropriate statistical analysis model. Figure 1 is a histogram plot of traffic accidents in Central Java Province in 2018. The plot shows an asymmetrical shape, so the number of traffic accidents in Central Java Province in 2018 shows a deviation from the normal distribution. The modeling of the factors that affect the number of traffic accidents in Central Java Province in 2018 is then approached with a Poisson distribution because the number of accidents is a census data which assumes that the probability of occurrence is very small or rarely occurs.

### 3.1 Multicollinearity Check

The factors used in the formation of the model must meet the assumption of no multicollinearity. The criteria that can be used to check for multicollinearity between explanatory variables are the correlation coefficient (Pearson Correlation) and Variance Inflation Factor (VIF). The relationship between the response variable and the explanatory variable can be seen from the resulting Pearson correlation coefficient. The higher the correlation value, the stronger the relationship between the two variables. The explanatory variable is also said to be independent if it has a VIF value of less than 10.

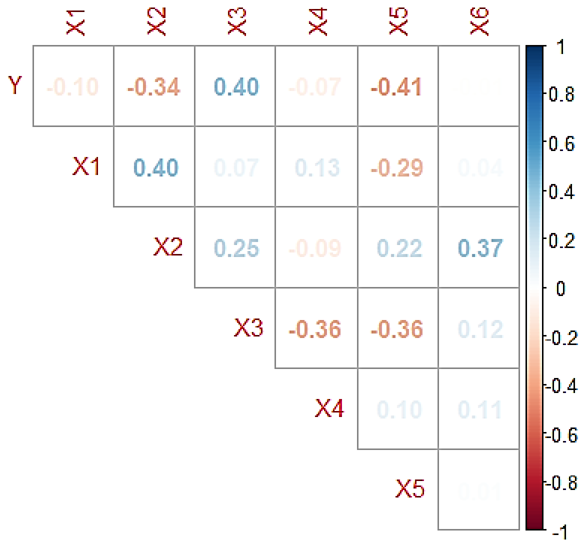


**Figure 2. Cross Correlation between Independent Variables Used in Modeling**

The cross-correlation between the variables shown in Figure 2 results in an overall correlation value of less than 0.5. It shows the conditions of freedom among the independent variables used in the modeling. This conclusion is also strengthened by the VIF value of all independent variables which is less than 10 as shown in Table 1 below. Therefore, all independent variables will be included in the modeling.

**Table 1. VIF Value of Explanatory Variables Used in Modeling**

| Independent Variable | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|---|---|---|---|---|---|---|
| VIF | 1.620 | 2.427 | 1.467 | 1.275 | 1.766 | 1.448 |

### 3.2 Poisson Regression Model

The number of traffic accidents in Central Java Province in 2018 is census data with a very small chance of occurrence against a certain time interval. Therefore, the appropriate regression model for estimating the factors that influence the number of traffic accidents in Central Java Province in 2018 is the Poisson regression model. The Poisson regression parameter assumption process was carried out using the Maximum Likelihood Assumption (MLE) method and Newton-Raphson iteration. The resulting Poisson Regression model is as follows:

$$\widehat{\mu_i} = exp(10.54 - 0.00000426X_1 - 0.114X_2 + 0.0185X_3 + 0.003547X_4 - 0.03371X_5 + 0.00809\, X_6)$$

Testing the goodness of the resulting Poisson regression model was carried out with the following hypothesis:

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

$H_1$: there is at least one $\beta_j \neq 0$, $\quad j = 1,2,3, \dots, k$.

The test criteria with a significance level (α=0.05) is rejecting $H_0$ if $D(\beta) > \chi^2_{(v,\alpha)}$. Based on the results of the analysis $D(\beta) = 3123.1 > \chi^2_{(28,0.05)} = 41.34$ was obtained so that $H_0$ is rejected. These results indicate that there is at least one explanatory variable that affects the number of traffic accidents in Central Java Province in 2018.

The next step was testing for each of the regression parameter coefficients. The hypothesis being tested is as follows:

$$H_0: \boldsymbol{\beta}_i = 0$$

$H_1$: $\boldsymbol{\beta}_i \neq 0$, with $i = 1,2,3, \dots, p$.

The rejection criterion is $H_0$ with a significant level ($\alpha = 0.05$) if $|Z| > Z_{\alpha/2} = 1.96$. The test results are presented in the following table:

**Table 2. The Test Results of the Poisson Regression Model Parameters Alleged**

| Parameter | Assumption | Standard Error | Z Value | *P-value* | Interpretation |
|---|---|---|---|---|---|
| (Intercept) | $1.054 \times 10^{1}$ | $4.230 \times 10^{-1}$ | 24.925 | **$< 2 \times 10^{-16}$** | Significant |
| $X_1$ | $-4.268 \times 10^{-6}$ | $3.996 \times 10^{-6}$ | -1.068 | $2.86 \times 10^{1}$ | Insignificant |
| $X_2$ | $-1.140 \times 10^{-1}$ | $5.494 \times 10^{-3}$ | -20.755 | **$< 2 \times 10^{-16}$** | Significant |
| $X_3$ | $1.850 \times 10^{-2}$ | $6.811 \times 10^{-4}$ | 27.166 | **$< 2 \times 10^{-16}$** | Significant |
| $X_4$ | $3.547 \times 10^{-3}$ | $9.518 \times 10^{-4}$ | 3.727 | **$1.94 \times 10^{-4}$** | Significant |
| $X_5$ | $-3.371 \times 10^{-2}$ | $4.637 \times 10^{-3}$ | -7.270 | **$3.58 \times 10^{-13}$** | Significant |
| $X_6$ | $8.093 \times 10^{-3}$ | $1.209 \times 10^{-3}$ | 6.696 | **$2.15 \times 10^{-11}$** | Significant |

Based on Table 2 above, of the six independent variables used in the modeling, it is concluded that only the $X_1$ variable (population density) has no significant effect on the variable number of traffic accidents. The results of the assumption of the Poisson regression model parameters above were studied further for the violation existence of the assumption of the mean value similarity with the model variance value (equidispersion).

### 3.3 Overdispersion Check

Violation of the assumption of mean and variance similarity (equidispersion) in the Poisson regression model can be in the form of underdispersion or overdispersion. Allegations of a violation of these assumptions can be analyzed by paying attention to the visualization of the average value plot of the model with the variance of the Poisson Regression regression model, which is shown in the following figure:

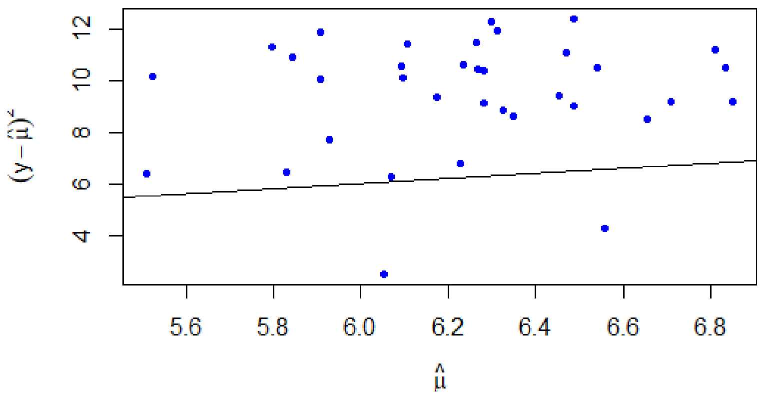


**Figure 3. Plot between the Mean Value of the Model and the Variance of the Poisson Regression Model**

Based on Figure 3 above, the distribution of points on the plot between the mean value and the variance of the Poisson Regression model spreads over a linear line (equidispersion). It indicates an overdispersion problem in the Poisson regression model. Further examination of assumption violations can be carried out by ratifying the Deviance and Pearson Chi-Square values with their degrees of freedom. Deviance and Pearson Chi-Square ratio values in Table 3 are 111.54 and 112.32, respectively. Deviance and Pearson Chi-Square ratio values that are more than 1 indicate an assumption of an overdispersion problem in the Poisson regression model.

**Table 3. Values of Chi-square Deviance and Pearson Ratios in the Poisson Regression Model**

| Criteria | Free degrees | Value | Free Grades/Degrees |
|---|---|---|---|
| *Deviance* | 28 | 3123.11 | 111.54 |
| Pearson Chi-Square | 28 | 3145.04 | 112.32 |

Poisson Regression modeling with data containing overdispersion causes the standard error to be too low so that the parameter assumption is biased downwards (underestimate). As a result, it will cause errors in concluding if the model is used. The problem of overdispersion in Poisson Regression will then be handled using Generalized Poisson Regression and Negative Binomial Regression models.

### 3.4 Negative Binomial Regression Model

The modeling of the number of traffic accidents in Central Java province in 2018 with six explanatory variables with the Negative Binomial Regression model is presented in table 5. Modeling with Negative Binomial Regression produces the Deviance and Pearson Chi-Square ratio values with the degrees of freedom being 1.292 and 1.268, respectively. The ratio of Deviance and Pearson Chi-Square values with degrees of freedom in the Negative Binomial Regression model is close to one and much smaller than in the Poisson regression model. It indicates that modeling the number of traffic accidents in Central Java province in 2018 with the Negative Binomial Regression model has able to overcome the problem of overdispersion in the Poisson regression model.

The estimated parameter values of the Poisson Regression and Negative Binomial Regression models are almost the same. It shows the consistency of the estimated values for the two models. The effect of the overdispersion problem of the Poisson regression model is shown in the standard deviation of the two models. The Poisson Regression model produces a smaller standard deviation than the Negative Binomial Regression model. The small standard deviation value in the Poisson Regression model causes the significance test of each explanatory variable to be downwardly biased.

**Table 4. Assumption of Poisson Regression Model and Negative Binomial Regression Model**

| | ***Poisson Regression* Model** | | | ***Negative Binomial Regression* Model** | | |
|---|---|---|---|---|---|---|
| **Parameter** | **Assumption** | **Standard Error** | ***P-value*** | **Assumption** | **Standard Error** | ***P-value*** |
| (Intercept) | $1.054 \times 10^{1}$ | $4.230 \times 10^{-1}$ | **$< 2 \times 10^{-16}$** | $1.066 \times 10^{1}$ | $4.364 \times 10^{0}$ | 0.0145 |
| $X_1$ | $-4.268 \times 10^{-6}$ | $3.996 \times 10^{-6}$ | $2.86 \times 10^{1}$ | $-9.254 \times 10^{-6}$ | $3.992 \times 10^{-5}$ | 0.8167 |
| $X_2$ | $-1.140 \times 10^{-1}$ | $5.494 \times 10^{-3}$ | **$< 2 \times 10^{-16}$** | $-1.289 \times 10^{-1}$ | $5.469 \times 10^{-2}$ | **0.0184** |
| $X_3$ | $1.850 \times 10^{-2}$ | $6.811 \times 10^{-4}$ | **$< 2 \times 10^{-16}$** | $1.905 \times 10^{-2}$ | $6.498 \times 10^{-3}$ | **0.0034** |
| $X_4$ | $3.547 \times 10^{-3}$ | $9.518 \times 10^{-4}$ | **$1.94 \times 10^{-4}$** | $9.453 \times 10^{-4}$ | $9.224 \times 10^{-3}$ | 0.9184 |
| $X_5$ | $-3.371 \times 10^{-2}$ | $4.637 \times 10^{-3}$ | **$3.58 \times 10^{-13}$** | $-2.986 \times 10^{-2}$ | $4.685 \times 10^{-2}$ | 0.5239 |
| $X_6$ | $8.093 \times 10^{-3}$ | $1.209 \times 10^{-3}$ | **$2.15 \times 10^{-11}$** | $1.015 \times 10^{-2}$ | $1.258 \times 10^{-2}$ | 0.4198 |
| Theta ($\phi$) | 0 | | | 5.12 | $1.20 \times 10^{0}$ | |
| *Deviance* | 3123.11 | | | 36.179 | | |
| Pearson $\chi^2$ | 3145.04 | | | 35.497 | | |
| Df | 28 | | | 28 | | |

Estimating the parameters of the Negative Binomial Regression model based on Table 4 reveals that two of the six explanatory variables used have a significant effect on the response variable at the significant level ($\alpha = 0.05$). The two explanatory variables are the percentage of adolescent age ($X_2$) and the percentage of accidents occurring in the district/city road area ($X_3$).

### 3.5 Generalized Poisson Regression Model

Another alternative approach to deal with overdispersion problems that arise in the Poisson regression model is the Generalized Poisson regression model. The Generalized Poisson Regression model can be used to overcome the problem of overdispersion and underdispersion in the Poisson Regression model. The results of the assumption of the parameters of the Generalized Poisson Regression model are as follows:

**Table 5. Assumption of the Poisson Regression Model and the Generalized Poisson Regression Model**

| | Poisson Regression Model | | | Generalized *Poisson Regression* Model | | |
|---|---|---|---|---|---|---|
| **Parameter** | **Assumption** | **Standard Error** | ***P-value*** | **Assumption** | **Standard Error** | ***P-value*** |
| (Intercept) | $1.054 \times 10^{1}$ | $4.230 \times 10^{-1}$ | **$< 2 \times 10^{-16}$** | $7.450 \times 10^{0}$ | $3.656 \times 10^{0}$ | **0.0416** |
| $X_1$ | $-4.268 \times 10^{-6}$ | $3.996 \times 10^{-6}$ | $2.86 \times 10^{1}$ | $-6.060 \times 10^{-5}$ | $3.913 \times 10^{-5}$ | 0.1214 |
| $X_2$ | $-1.140 \times 10^{-1}$ | $5.494 \times 10^{-3}$ | **$< 2 \times 10^{-16}$** | $-1.246 \times 10^{-1}$ | $4.851 \times 10^{-2}$ | **0.0140** |
| $X_3$ | $1.850 \times 10^{-2}$ | $6.811 \times 10^{-4}$ | **$< 2 \times 10^{-16}$** | $1.763 \times 10^{-2}$ | $5.791 \times 10^{-3}$ | **0.0023** |
| $X_4$ | $3.547 \times 10^{-3}$ | $9.518 \times 10^{-4}$ | **$1.94 \times 10^{-4}$** | $6.633 \times 10^{-4}$ | $8.303 \times 10^{-3}$ | 0.9363 |
| $X_5$ | $-3.371 \times 10^{-2}$ | $4.637 \times 10^{-3}$ | **$3.58 \times 10^{-13}$** | $-1.928 \times 10^{-2}$ | $4.015 \times 10^{-2}$ | 0.6311 |
| $X_6$ | $8.093 \times 10^{-3}$ | $1.209 \times 10^{-3}$ | **$2.15 \times 10^{-11}$** | $9.887 \times 10^{-3}$ | $1.067 \times 10^{-2}$ | 0.3542 |
| Theta ($\phi$) | 0 | | | 2.19 | $1.550 \times 10^{-1}$ | **0.0000** |
| *Deviance* | 3123.11 | | | | | |
| Pearson $\chi^2$ | 3145.04 | | | 67.098 | | |
| Df | 28 | | | 62 | | |

Assumption of the dispersion of the Generalized Poisson Regression model on the number of traffic accidents in the province of Central Java with the six explanatory variables in Table 5 above produces a positive theta dispersion ($\phi$) value of 2.19. The positive value indicates the overdispersion parameter in the Generalized Poisson regression model. The assumption of the Generalized Poisson regression model parameters based on table 6 above gives the same conclusion as the previous Negative Binomial Regression model. The explanatory variables that have a significant effect on the response variable at the real level ($\alpha = 0.05$) are the percentage of adolescent age ($X_2$) and the percentage the accident that occurred in the district/city road area ($X_3$).

### 3.6 Best Model Selection

The selection of the best model from several models that have been produced was made by taking into account the AIC value. The smaller the AIC value of a model, the better the model. The AIC values of the three models are presented in the following table:

**Table 6. AIC Value in Each Model**

| Model | AIC |
|---|---|
| *Poisson Regression* Model | 3416.30 |
| *Negative Binomial Regression* Model | 490.66 |
| *Generalized Poisson Regression* Model | **485.50** |

The best model was determined by considering the AIC value with the criteria that the model with the smallest AIC value is the best. Table 6 shows that the Generalized Poisson regression model has the smallest AIC value of 485.50 compared to other models. Therefore, the Generalized Poisson regression model is the best model to model the number of traffic accidents in Central Java province in 2018. Evaluation of the model's goodness can also be done by looking at the visualization of the actual data distribution plot with the estimated data for each model. Figure 4 below also shows the estimated value data of the Generalized Poisson regression model, which is closer to the actual value data compared to the data on the estimated value of the Negative Binomial Regression model.

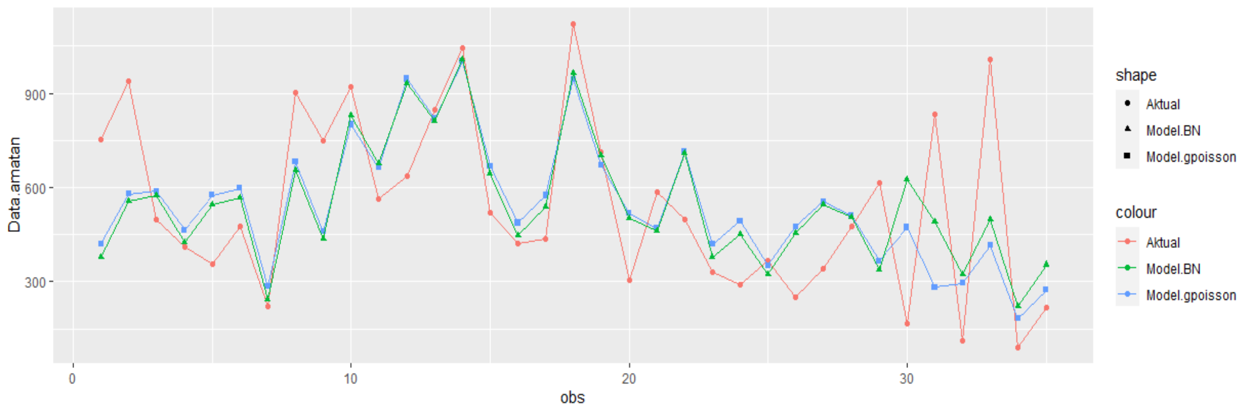


**Figure 4. Plot of Observed Values with Estimated Values of Each Model**

### 3.7 Interpretation

Modeling the factors that influence the number of traffic accidents in Central Java province in 2018 resulted in the Generalized Poisson regression model as the best model. The independent variables that significantly affect the level of significance ($\alpha$ = 0.05) are the percentage of adolescent age ($X_2$) and the percentage of accidents occurring in the district/city road area ($X_3$). The equation of the model is as follows:

$$\widehat{\mu_i} = exp(7.45 - 0.1246\ X_2 + 0.0176X_3)$$

The resulting Generalized Poisson Regression model shows that every 1% addition of adolescent age in the district/city ($X_2$) causes the expected value of traffic accidents in Central Java province to decrease by exp (0.1246) = 1.327, times assuming other variables are considered permanent. These results indicate that road users with vehicles in the districts/cities of Central Java province are still dominated by adults for activities and are prone to accidents.

The interpretation of the variable ($X_3$) with a regression coefficient of 0.0176 is that every 1% addition of the percentage of accidents occurring in the district/city road area causes the expected value of traffic accidents in Central Java province to increase by exp (0.0176) = 1.0178 times assuming other variables are held constant. Increased community activity in driving and the damaged condition of district/city roads in Central Java cause a higher chance of accident risk.

## 4. CONCLUSION

Modeling the number of traffic accidents in the province of Central Java in 2018 with the Poisson regression model experienced a violation of the overdispersion assumption. Handling overdispersion with Negative Binomial Regression and Generalized Poisson Regression models can overcome the problem of overdispersion that appears in the Poisson regression model. It is evidenced by the Deviance and Pearson Chi-Square ratio values in both models, which are close to one. Based on the smallest AIC value, the Generalized Poisson Regression model was chosen as the best model to model the number of traffic accidents in the Central Java province in 2018. Based on this model, it was found that the factors that significantly influence the real level are the percentage of adolescents and the percentage of accidents occurring in the district/city road area.

## ACKNOWLEDGEMENT

This research was funded by DIPA Universitas Muhammadiyah Semarang in the 2021 fiscal year through the Penelitian Dosen Pratama scheme with Contract Number: 063/UNIMUS.L/PT/PJ.INT/2021. Therefore, the authors and the research team give our gratitude to the relevant parties for their assistance in conducting this research.

## REFERENCES


[1] M. Kadarisman, A. Gunawan, and I. Ismiyati, “Implementasi Kebijakan Sistem Transportasi Darat dan Dampaknya terhadap Kesejahteraan Sosial di Jakarta,” ["Implementation of Land Transportation System Policies and Its Impact on Social Welfare in Jakarta,"] *J. Manaj. Transp. DAN LOGISTIK*, vol. 2, no. 1, pp. 59–78, 2017, doi: 10.25292/j.mtl.v2i1.129.

[2] N. K. E. Y. Utari, I. G. A. M. Srinadi, and M. Susilawati, “Model Geographically Weighted Regression (GWR) Faktor-faktor yang Memengaruhi Kecelakaan Lalu Lintas di Provinsi Bali,” ["Geographically Weighted Regression (GWR) Model of Factors Influencing Traffic Accidents in Bali Province,"] *E-Jurnal Mat.*, vol. 8, no. 2, pp. 140–147, 2019, doi: 10.24843/mtk.2019.v08.i02.p245.

[3] Y. Wilandari, S. Sugito, and C. Silvia, “Analisis Kecelakaan Lalu Lintas di Kota Semarang Menggunakan Model Log Linier,” ["Analysis of Traffic Accidents in the City of Semarang Using the Linear Log Model,"] *MEDIA Stat.*, vol. 9, no. 1, pp. 51–61, 2016, doi: 10.14710/medstat.9.1.51-61.

[4] BPS, *Statistik TRansportasi Darat 2019 Land Transportation Statistics 2019*, vol. 53, no. 9. 2020.

[5] BPS Provinsi Jawa Tengah, *Provinsi Jawa Tengah Dalam Angka 2019*. 2019.

[6] P. L. A. Dewi and I. Zain, “Pemodelan Faktor Penyebab Kecelakaan Lalu Lintas Berdasarkan Metode Geographically Weighted Regression di Jawa Timur,” ["Modeling the Causes of Traffic Accidents Based on the Geographically Weighted Regression Method in East Java," *Jurmnal Sains dan Seni*, vol. 5, no. 1, p. D 58-D 63, 2016.

[7] S. P. Agustianto, S. Martha, and N. Satyahadewi, “Pemodelan Faktor Penyebab Kecelakaan Lalu Lintas di Kalimantan Barat dengan Metode Geographically Weighted Regression (GWR),” ["Modeling the Causes of Traffic Accidents in West Kalimantan with the Geographically Weighted Regression (GWR) Method,"] vol. 07, no. 4, pp. 303 – 310, 2018.

[8] M. S. Fuad, A. Kudus, and S. Sunendiari, “Pemodelan Data Kecelakaan Lalu Lintas Menggunakan Metode Regresi Poisson dan Regresi Binomial Negatif,” ["Traffic Accident Data Modeling Using Poisson Regression Method and Negative Binomial Regression,"] in *SPeSIA-UNISBA* , 2018, pp. 26–33.

[9] Y. Widyaningsih, G. P. Arum, and K. Prawira, “Aplikasi K-Fold Cross Validation dalam Penentuan Model Regresi Binomial Negatif Terbaik,” ["K-Fold Cross Validation Application in Determining the Best Negative Binomial Regression Model,"] *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 2, pp. 315–322, Jun. 2021, doi: 10.30598/barekengvol15iss2pp315-322.

[10] A. Melliana, Y. Setyorini, H. Eko, S. Rosi, and Purhadi, “The Comparison Of Generalized Poisson Regression and Negative Binomial Reression Methods In Overcoming Overdispersion,” *Int. J. Sci. Technol. Res.*, vol. 2, no. 8, pp. 255–258, 2013.

[11] R. Tongaril Simarmata and D. Ispriyanti, “Penanganan Overdispersi pada Model Regresi Poisson Menggunakan Model Regresi Binomial Negatif,” ["Handling Overdispersion on the Poisson Regression Model Using the Negative Binomial Regression Model,"] *Media Stat.*, vol. 4, no. 2, pp. 95–104, 2011.

[12] A. Prahutama, D. Ispriyanti, and B. Warsito, “Modelling Generalized Poisson Regression in the Number of Dengue Hemorrhagic Fever (DHF) in East Nusa Tenggara,” in *E3S Web of Conferences*, 2020, vol. 202, doi: 10.1051/e3sconf/202020212017.

[13] B. Yadav *et al.*, “Can Generalized Poisson model replace any other count data models? An evaluation,” *Clin. Epidemiol. Glob. Heal.*, vol. 11, 2021, doi: 10.1016/j.cegh.2021.100774.

[14] J. M. Hilbe, *Negative binomial regression, second edition*. 2011.

[15] J. U. Ibeji, T. Zewotir, D. North, and L. Amusa, “Modelling fertility levels in Nigeria using Generalized Poisson regression-based approach,” *Sci. African*, vol. 9, 2020, doi: 10.1016/j.sciaf.2020.e00494.

[16] S. Bae, F. Famoye, J. T. Wulu, A. A. Bartolucci, and K. P. Singh, “A rich family of generalized Poisson regression models with applications,” in *Mathematics and Computers in Simulation*, 2005, vol. 69, no. 1-2 SPEC. ISS., doi: 10.1016/j.matcom.2005.02.026.

[17] N. M. R. Kwswari, I. W. Sumarjaya, and N. L. P. Suciptawati, “Perbandingan Regresi Binomial Negatif dan Regresi Generalisasi Poisson dalam Mengatasi Overdispersi (Studi Kasus: Jumlah Tenaga Kerja Usaha Pencetak Genteng di Br. Dukuh, Desa Pejaten),” ["Comparison of Negative Binomial Regression and Poisson's Generalized Regression in Overcoming Overdispersion (Case Study: Number of Manpower in Tile Printing Business in Br. Dukuh, Pejaten Village),"] *E-Jurnal Mat.*, vol. 3, no. 3, pp. 107–115, 2014, doi: 10.24843/mtk.2014.v03.i03.p072.