COMPARISON OF POISSON REGRESSION AND GENERALIZED POISSON REGRESSION IN MODELING THE NUMBER OF INFANT MORTALITY IN WEST JAVA 2022

Toha Saifudin[1*], Fatiha Nadia Salsabila[2], Mubadi'ul Fitriyani[3], Azizatul Kholidiyah[4], Nina Auliyah[5], Fildzah Tri Januar Ariani[6], Suliyanto[7]

[1,2,3,4,5,6,7]*Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga*
Jl. Dr. Ir. H. Soekarno, Mulyorejo, Surabaya, Jawa Timur, 60115, Indonesia

*Corresponding author's e-mail: * tohasaifudin@fst.unair.ac.id*

ABSTRACT

In line with the Sustainable Development Goals (SDGs), the Infant Mortality Rate (AKB) is a very important health indicator, especially in neonatal and perinatal care. West Java Province consistently ranks third nationally in terms of infant mortality in 2020 and 2021. This study analyzes the factors influencing infant mortality in West Java in 2022 using secondary data from the 2022 West Java Provincial Health Profile. The response variable is the number of infant deaths, while the predictor variables include the percentage of K-4 coverage (X1), high-risk pregnancy (X2), family with PHBS (X3), exclusive breastfeeding (X4), and complete immunization coverage (X5). Given the count-based nature of the data, Poisson regression was used, which assumes equidispersion where the variance is equal to the mean. However, the analysis found overdispersion, where the variance significantly exceeds the mean, making Poisson regression unsuitable. To address this, Generalized Poisson Regression (GPR) was applied, as GPR introduces a dispersion parameter that accounts for overdispersion, thus better fitting the data. The initial Poisson regression results showed that X1, X2, X4, and X5 significantly influenced infant mortality, while the GPR model showed that only X2 and X3 were significant factors, with a dispersion parameter of -3.116. The GPR model shows that every additional one high-risk pregnancy increases the infant mortality rate by 1.00006, while an increase of one unit of clean and healthy living practices reduces the mortality rate by 2.66%. Model evaluation using AIC, BIC, and RMSE confirmed that the GPR model better described the relationship between predictor variables and infant mortality rates compared to Poisson regression. These findings emphasize the need to use GPR to model cases with overdispersion in count data, so as to provide more reliable information for policy and intervention strategies.

1. INTRODUCTION

Infant Mortality Rate (IMR) is one of the critical indicators in health service efforts, especially in neonatal and perinatal care. This matter is in line with the third point in the Sustainable Development Goals (SDGs): ensuring healthy lives and improving the well-being of people of all ages, where one of the targets is to reduce the neonatal mortality rate (AKN) to at least 12 per 1000 live births and the infant mortality rate (IMR) to 12 per 1000 live births by 2030 [1]. Regarding the Infant Mortality Rate, Indonesia is one of the highest-ranked countries in Southeast Asia [2].

On a national scale, based on data from Indonesia's Health Profile for two consecutive years, West Java Province ranks third highest in the number of infant deaths [3]. Therefore, local governments strive to suppress the reduction in infant mortality by analyzing the factors that influence it. Infant mortality is generally caused by congenital factors obtained from parents during pregnancy, such as genetics and nutritional status of pregnant women [4]. In addition to internal factors, infant mortality can also be triggered by various external factors. Several external factors affect infant mortality, including visits by health worker during pregnancy, lack of health services, and lack of knowledge and application of Clean and Healthy Living Behavior (PHBS) in the surrounding environment [5]. Thus, the analysis of the Infant Mortality Rate (IMR) and the factors that influence it can be helpful as a benchmark for the level of health development in a region or country and the quality of life of its people. Thus, analyzing the Infant Mortality Rate (IMR) and its influencing factors can serve as a benchmark for the level of health development in a region or country and the quality of life of its people.

Poisson regression is one of the statistical analyses that can be used on discrete data (count), such as the number of infant deaths. According to McCullagh and Nelder in [6], in Poisson Regression, response data is assumed to be Poisson distributed by satisfying equidispersion characteristics, namely the similarity of mean values and population variance. However, finding such a qualified population is rare because the variance value is greater than the mean (overdispersion) or the variance value is smaller than the mean (underdispersion), so the conclusions are invalid. Several alternatives can be done if Poisson regression does not meet the equidispersion assumption, one of which is Generalized Poisson Regression (GPR). According to Wang and Famoye in [7], GPR is a regression analysis that can be an alternative when there are symptoms of underdispersion or overdispersion in Poisson regression with the addition of dispersion parameter, namely θ. There are three conditions that can occur in GPR regarding the parameter θ. If the value of $\theta = 0$, it means that the variance is equal to the mean, $Var(Y_i|x_i) = E(Y_i|x_i)$, so the addition of dispersion parameters is not necessary or the GPR model is similar to the conventional model in Poisson regression. The second condition, when the value of $\theta > 0$, means that the variance is greater than the mean, $Var(Y_i|x_i) > E(Y_i|x_i)$, so it is called an overdispersion condition. The last condition is underdispersion where the value of $\theta < 0$, meaning that the variance is smaller than the mean, $Var(Y_i|x_i) < E(Y_i|x_i)$.

In a case study of the number of infant deaths in East Java in 2014, the results of [5] research stated that the Negative Binomial Regression model is the best model with neonatal mortality predictor variables having smaller AIC and BIC values than the complete model. The study's weakness is that neonatal mortality is a constituent component of infant mortality, or infant mortality is the sum of neonatal and post-neonatal deaths, so these two variables should not be regressed to infant mortality. In addition to East Java, an analysis of infant deaths was carried out in Bandung City in 2019 by [8]. In the study, it was found that the factors that significantly influenced the number of infant deaths were the percentage of low birth weight (BBLR), the percentage of neonatal visits three times, and the percentage of babies who were exclusively breastfed. The study was conducted using Zero-Inflated Poisson regression to overcome overdispersion. However, Zero-Inflated Poisson regression is more appropriate for data containing many zero values. Categorical data analysis has also been carried out by [9]. In the study, it was found that BBLR, Neonatal Infection, Asphyxia, Congenital abnormalities, and hypothermia are associated with the incidence of infant mortality at Batara Siang Hospital, Pangkep Regency in 2020 using a cross-sectional approach and analyzed using the Chi-square test. The disadvantage of testing categorically is the limitation in information, resulting in the loss of information about data variance.

In previous studies, such as in East Java in 2014 [5], Negative Binomial Regression was found to be the best model for analyzing neonatal mortality rates due to its lower AIC and BIC values. However, the study's weakness lies in the fact that neonatal mortality is a constituent component of infant mortality, so these two variables should not be regressed together. Another analysis of infant deaths was conducted in Bandung City in 2019 by [8], which found that low birth weight (BBLR), the percentage of three neonatal

visits, and the percentage of exclusive breastfeeding significantly influenced infant mortality. That study used Zero-Inflated Poisson regression to overcome overdispersion, but this approach is better suited for data with many zero values. Categorical data analysis was also conducted by [9], which identified factors such as BBLR, neonatal infection, asphyxia, congenital abnormalities, and hypothermia as associated with infant mortality at Batara Siang Hospital, Pangkep Regency in 2020 using the Chi-square test. However, categorical testing has limitations, as it loses variance information.

Unlike previous studies focusing on different regions or using outdated data, this research introduces several novelties. First, the study analyzes recent infant mortality data from 2022 in West Java, which represents the latest public health challenges in one of Indonesia's most populous provinces [10]. Secondly, this study adopts the Generalized Poisson Regression model as a primary approach to address overdispersion, which has been identified as a common limitation in previous studies that used conventional Poisson or Negative Binomial models. Moreover, unlike some earlier works that included neonatal mortality as a predictor variable, this study recognizes neonatal mortality as a component of infant mortality, thereby avoiding multicollinearity issues and providing a clearer understanding of external and demographic factors influencing infant mortality. Additionally, there are no zero values in the data, rendering zero-inflated Poisson regression unnecessary.

This study uses the Poisson regression method to model the number of infant deaths as discrete data. However, when overdispersion occurs which indicates a mismatch in equity, this problem can be overcome by using the generalized Poisson regression method. Generalized Poisson Regression was developed by Famoye based on the Generalized Poisson distribution introduced by Consul and Jain. The method is able to handle discrete (count) data that experience overdispersion, underdispersion, and equidispersion [11]. Research by Joe and Zhu in [11] explained that the generalized Poisson regression model offers a flexible alternative to standard Poisson regression by incorporating additional dispersion parameters to account for overdispersion in count data. Joe and Zhu's research shows that the generalized Poisson distribution is a mixture of Poisson distributions, which highlights its distinctive probabilistic characteristics. In our study, the infant mortality data in West Java Province in 2022 showed signs of overdispersion, making the generalized Poisson regression model a more appropriate and accurate choice for our analysis. By using this model, we aim to provide more reliable estimates and insights into the factors affecting infant mortality, which can be a valuable reference for policy makers in crafting effective interventions.

2. RESEARCH METHODS

2.1 Infant Mortality

Infant mortality is a term used to describe infant mortality from one day to year [12]. The incidence of infant mortality at less than four weeks of age, which is generally influenced by endogenous factors from birth, is called neonatal death. Based on age, neonatal mortality is divided into two, namely early neonatal death (0-6 days) and late neonatal death (7-28 days). If infant death occurs at the age of more than 28 days and has been influenced by exogenous factors, it is called post-neonatal mortality [13]. Factors that influence infant mortality are K4 coverage, high-risk pregnancies, PHBS households, exclusive breastfeeding, and complete primary immunization.

2.2 Poisson Distribution

According to Walpole and Myers in [14], experiments that produce values for the numerical random variable Y, i.e., the number of successes over a specific time interval or in a particular area, are called Poisson experiments. The length of the time interval can be minutes, days, weeks, or years. The area in question can be a piece of line, a space, a content, or a part of the object.

A random variable Y is considered Poisson-distributed if it has the following probability function in [14] denoted by **Equation (1)**.

$$f(y, \mu) = \frac{e^{-\mu} \mu^y}{y!} = \frac{\exp(-\mu) \exp(y \ln \mu)}{y!} \tag{1}$$

Where μ is the average number of successes that occur in a given time interval or area where $y = 0, 1, 2, \ldots$; $\mu > 0$ and $E(Y) = Var(Y) = \mu$.

2.3 Poisson Regression

Poisson regression is a non-linear analysis of the Poisson distribution used in discrete data types. Poisson regression assumes that the response variable Y has a Poisson distribution, no symptoms of multicollinearity, and equidispersion [15]. Poisson regression can be expressed as follows in [16].

$$\eta_i = \ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \tag{2}$$

Where η_i is the link function, μ_i is mean value of the effect of the i-th response, β_0 is a constant, β_k is regression coefficient of the-k predictor variable, and x_{ik} is value of the j-th predictor variable of the i-th intensity.

2.3.1 Parameter Estimation of Poisson Regression

The method used to assess the parameters of the Poisson regression model is Maximum Likelihood Estimation (MLE). MLE is performed by maximizing the likelihood function [17].

$$\ln L(\boldsymbol{\beta}) = \ln\left(\prod_{i=1}^{n} \frac{exp\,(-\mu_i)\mu_i{}^{y_i}}{y_i!}\right) \tag{3}$$

$$\ln L(\boldsymbol{\beta}) = -\sum_{i=1}^{n} e^{x_i^T \beta} + \sum_{i=1}^{n} y_i x_i^T \beta - \sum_{i=1}^{n} \ln(y_i!) \tag{4}$$

With y_i is the i-th observation of the response variable Y. While $\boldsymbol{\beta}$ is a Poinnson regression model which is a function of x_i as a predictor variable and $\boldsymbol{\beta}$ as a regression parameter to be estimated.

The equation above is still implicit, so it is necessary to iterate Newton Raphson to perform numerical iterations with the following optimization algorithm.

1. Determine the initial estimated value of a parameter $\left(\widehat{\boldsymbol{\beta}}_{(0)}\right)$ obtained by the Ordinary Least Square (OLS) method, namely:

$$\widehat{\boldsymbol{\beta}}_{(0)} = \left(X^T X\right)^{-1} X^T Y \tag{5}$$

   With

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k} \end{bmatrix} \text{ and } Y = [Y_1 \quad Y_2 \quad \cdots \quad Y_n]^T$$

2. Forms a gradient vector $\boldsymbol{g}$,

$$\boldsymbol{g}(\boldsymbol{\beta}) = -\sum_{i=1}^{n} x_i.\mu_i + \sum_{i=1}^{n} y_i x_i = \sum_{i=1}^{n} x_i(y_i - \mu_i) \tag{6}$$

So, the vector $\boldsymbol{g}^T\left(\boldsymbol{\beta}_{(m)}\right)_{(k+1)\times 1}$ is as follows.

$$\left[\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_0} \quad \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_1} \quad \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_2} \quad \cdots \quad \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_k}\right]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}(m)} \tag{7}$$

Where k is the number of estimated parameters.

$$H(\boldsymbol{\beta}) = -\sum_{i=1}^{n} x_i.x_i^T.\mu_i = -\sum_{i=1}^{n} \mu_i.x_i.x_i^T \tag{8}$$

$$H\big(\boldsymbol{\beta}_{(m)}\big)_{(k+1)\times(k+1)} = \begin{bmatrix} \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_0^2} & \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_0 \beta_1} & \cdots & \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_0 \partial \beta_k} \\ \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_0 \partial \beta_1} & \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_1^2} & \cdots & \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_0 \partial \beta_k} & \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_k} & \cdots & \dfrac{\partial^2 \ln L(\beta)}{\partial \beta_k^2} \end{bmatrix}_{\beta=\beta_{(m)}}$$

$$Var(\widehat{\boldsymbol{\beta}}) = -E\big[\boldsymbol{H}^{-1}(\widehat{\boldsymbol{\beta}})\big] \tag{9}$$

3. Input the value of $\widehat{\boldsymbol{\beta}}_{(0)}$ into vector elements $\boldsymbol{g}$ and matrix $\boldsymbol{H}$, so the vector that obtained is $\boldsymbol{g}\big(\widehat{\boldsymbol{\beta}}_{(0)}\big)$ and matrix $\boldsymbol{H}\big(\widehat{\boldsymbol{\beta}}_{(0)}\big)$.

4. Start from the value $\boldsymbol{m} = \boldsymbol{0}$ iterating through the equation: $\widehat{\boldsymbol{\beta}}_{(m+1)} = \widehat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{H}^{-1}\big(\boldsymbol{\beta}_{(m)}\big)\boldsymbol{g}\big(\boldsymbol{\beta}_{(m)}\big)$. The value of $\widehat{\boldsymbol{\beta}}_{(m)}$ is a group of parameter assessors that converge on the m-iteration.

5. The convergence of parameter estimators can be seen when the value $\big\|\widehat{\boldsymbol{\beta}}_{(m+1)} = \widehat{\boldsymbol{\beta}}_{(m)}\big\| \leq \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon}$ is a very small number. So, there is almost no difference between $\widehat{\boldsymbol{\beta}}_{(m+1)}$ and $\widehat{\boldsymbol{\beta}}_{(m)}$. If no convergent parameter estimation has been obtained, then proceed again the next step until iteration to $\boldsymbol{m} = \boldsymbol{m} + \boldsymbol{1}$.

2.3.2 Parameter Testing of Poisson Regression

To determine test statistics in testing Poisson regression model parameters is to use the Maximum Likelihood Ratio Test (MLRT) method with a hypothesis. $\boldsymbol{L}(\widehat{\boldsymbol{\Omega}})$ is the likelihood value for the complete model involving predictor variables and $\boldsymbol{L}(\widehat{\boldsymbol{\omega}})$ is the likelihood value for a null model, which is a simple model without involving predictor variables [17].

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_j = 0$
$H_1 :$ at least there is one $\beta_j \neq 0$; $j = 1, 2, \dots, k$

Test statistics for Poisson regression model qualification are as follows.

$$D(\widehat{\boldsymbol{\beta}}) = -2\ln\left[\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right] = 2\left[\ln\left(L(\widehat{\Omega})\right) - \ln\left(L(\widehat{\omega})\right)\right] \tag{10}$$

So, the decision to rejects $\boldsymbol{H_0}$ if $\boldsymbol{D}(\widehat{\boldsymbol{\beta}}) > \chi_{v;\alpha}^2$ where v is the number of model parameters under the population minus the number of parameters below $\boldsymbol{H_0}$. Poisson regression model parameters that have been generated from estimates need to be tested on individual Poisson regression model parameters with the following hypothesis.

$H_0 : \beta_1 = 0$ (the effect of the j-th variable is not significant)
$H_1 : \beta_j \neq 0$ (the effect of the j-th variable is significant)

The test statistics that used is:

$$z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{11}$$

where $se(\hat{\beta}_j)$ is the standard error or error rate of parameter β_j. $se(\hat{\beta}_j)$ is obtained from the $(j + 1)$th diagonal element of $\big[-\boldsymbol{H}^{-1}(\widehat{\boldsymbol{\beta}})\big]$. It is decided to reject H_0 jika $|z| > z_{\frac{\alpha}{2}}$ where a is the significance level.

2.3.3 Equidispersion

According to Darnah in [18] overdispersion is a condition where the response variable data shows a variance value greater than the average value or can be written $\boldsymbol{Var(Y)} > \boldsymbol{E(Y)}$. Overdispersion or underdispersion will result in the deviant's value of the model being very large so that the resulting model is less precise. [19] explained that when the Poisson model is applied to overdispersion data, it causes the standard error to be underestimated, so that some explanatory variables become insignificant.

There are two ways to detect overdispersion, namely with residual deviance and Pearson chi-square which is formulated as follows.

1. Deviance Residual

$$\phi_1 = \frac{D^2}{db}; \; D^2 = 2\sum_{i=1}^{n}\left\{y_i ln\left(\frac{y_i}{\widehat{\mu_i}}\right) - (y_i - \widehat{\mu_i})\right\} \tag{12}$$

with $db = n - k - 1$ where k is the number of parameters, n is the number of observations, and D^2 is the deviance residual values.

2. Pearson-Chi Square

$$\phi_2 = \frac{\chi^2}{db}; \; \chi^2 = \sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{var(y_i)} \tag{13}$$

with $db = n - k - 1$ where k is the number of parameters, n is the number of observations, and χ^2 is the Pearson chi-square value.

If ϕ_1 and ϕ_2 is more than 1, then there is overdispersion in the data.

2.4 Generalized Poisson Regression

Generalized Poisson Regression (GPR) is a development of Poisson regression that is used in overcoming overdispersion. The Generalized Poisson distribution function is as follows in [20].

$$f(y; \mu; \theta) = \left(\frac{\mu}{1 + \theta\mu}\right)^y \frac{(1 + \theta y)^{y-1}}{y!} \exp\left(\frac{-\mu(1 + \theta)}{1 + \theta\mu}\right) \tag{14}$$

The mean of the GPR model is $E(y) = \mu$ and the variance is

$$Var(y) = \mu(1 + \theta\mu)^2$$

If $\theta = 0$ then the GPR model will be an ordinary Poisson regression model, while if $\theta > 0$ then the GPR model represents count data containing overdispersion. The GPR model is shown in **Equation (15)** and **Equation (16)**.

$$\ln(\mu_i) = x_i^T\beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \tag{15}$$

$$\mu_i = \exp(x_i^T\beta), i = 1, 2, \dots, n \tag{16}$$

2.4.1 Parameter Estimation of Generalized Poisson Regression

Parameter estimation of the Generalized Poisson Regression model is done using the Maximum Likelihood Estimation method. The likelihood function for the GPR model is as follows.

$$L(\beta, \theta) = \prod_{i=1}^{n}\left\{\left(\frac{\mu_i}{1 + \theta\mu_i}\right)^{y_i} \frac{(1 + \theta y_i)^{y_i-1}}{y_i!} exp\left(\frac{-\mu_i(1 + \theta)}{1 + \theta\mu_i}\right)\right\} \tag{17}$$

$$L(\beta, \theta) = \prod_{i=1}^{n}\left(\frac{\mu_i}{1 + \theta\mu_i}\right)^{yi}\left(\prod_{i=1}^{n}\frac{(1 + \theta y_i)^{y_i-1}}{y_i!}\right)exp\left[\frac{-\mu_i(1 + \theta yi)}{1 + \theta\mu_i}\right] \tag{18}$$

Then, **Equation (18)** is converted into natural logarithm function as follows

$$\ln L(\beta, \theta) = \sum_{i=1}^{n}\{y_i \ln(\mu_i) - y_i \ln(1 + \theta\mu_i) + (y_i - 1)\ln(1 + \theta y_i) + \Delta\}$$

for $\Delta = -\ln(y_i!) - \frac{\mu_i(1 + \theta y_i)}{1 + \theta\mu_i}$ and $\mu_i = \exp(x_i^T\beta)$. So, it is obtained

$$\ln L(\boldsymbol{\beta}, \theta) = \sum_{i=1}^{n} \left\{ \begin{array}{c} y_i \ln\left(\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})\right) - \\ y_i \ln(1 + \theta \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})) - \ln(y_i!) + \Delta^* \end{array} \right\} \tag{19}$$

with

$$\Delta^* = (y_i - 1)\ln(1 + \theta y_i) - \frac{\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})(1 + \theta y_i)}{1 + \theta \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})}$$

Furthermore, the parameter estimate $\widehat{\boldsymbol{\beta}}$ is obtained by lowering the natural logarithm equation of the likelihood function to $\boldsymbol{\beta}^T$ while the parameter estimate θ is obtained by lowering the natural logarithm equation of the likelihood function to θ and equalized to zero. The decrease in the ln-likelihood function is carried out using the Newton-Raphson iteration method until a convergent parameter estimate is obtained.

2.4.2 Parameter Testing of Generalized Poisson Regression

In Generalized Poisson Regression, the parameter testing step is the Maximum Likelihood Ratio Test (MLRT) method with the following hypothesis [20].

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_j = 0$$
$$H_1 : \text{there exists at least one } \beta_j \neq 0 \; ; j = 1, 2, \dots, k$$

The test statistic used is $D(\widehat{\boldsymbol{\beta}})$ in **Equation (10)**. if $D(\widehat{\boldsymbol{\beta}}) > \chi^2_{v;\alpha}$ then the decision taken is to reject H_0 with v stating the number of parameters under population minus the number of model parameters under H_0. Rejection of H_0 means that there is at least one $\beta_j \neq 0$ which indicates X_j has a significant effect on the model. With this conclusion, partial testing will be carried out using the following hypothesis.

$$H_0 : \beta_1 = 0 \text{ (the effect of the } j\text{-th variable is not significant)}$$
$$H_1 : \beta_j \neq 0 \text{ (the effect of the } j\text{-th variable is significant)}$$

and

$$H_0 : \theta = 0 \text{ (there are no overdispersion case)}$$
$$H_0 : \theta \neq 0 \text{ (there is an overdispersion case)}$$

The test statistic used is the same as in **Equation (11)** with the H_0 rejection region if $|z| > z_{\frac{\alpha}{2}}$.

2.5 Multicollinearity

Multicollinearity occurs when two or more predictor variables in a regression model are correlated with moderate to high correlated with each other. The accuracy of the estimation of the regression coefficient decreases as the predictor is added to the model. In addition, test the hypothesis for $\beta_k = 0$ is possible to generate different conclusions depending on which predictor variables are present in the model [21]. There are two ways to detect multicollinearity, namely Variance Inflation Factor (VIF) and Pearson correlation coefficient value (r_{ij}) between predictor variables. Cases of multicollinearity can be identified through Variance Inflation Factor (VIF) with this formula.

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \; , j = 1, 2, \dots, p \tag{20}$$

where R_j^2 states the coefficient of determination between x_j and other predictor variables. If the VIF value is more than 10, it indicates the presence of symptoms of multicollinearity among the predictor variables. The Pearson correlation coefficient between predictor variables denoted by r_{ij}. If the value of the Pearson correlation coefficient between variables is more than 0.95, it can be overcome by removing the predictor variable that has a high correlation by re-progressing between the response variable and the remaining predictor variables.

2.6 Poisson Distribution Fit Test

The distribution fit test is a statistical hypothesis test used to determine whether $x_1, x_2, \dots, x_n$ is the value of a random sample $X_1, X_2, \dots, X_n$ derived from a distribution with a distribution function $F(.)$ [22]. The distribution fit test can be used to test hypotheses as follows:

$H_0 : x_1, x_2, \dots, x_n$ is the value of distributed random sample with a distribution function $F(.)$.
$H_1 : x_1, x_2, \dots, x_n$ is not the value of distributed random sample with a distribution function $F(.)$.

One test that can be used to test distribution fit is to use the chi-squared fit test which can be used for both discrete and continuous data. Test statistics to test chi-squared fit are:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (21)$$

Where O_i is the number of observations in the category i, E_i is calculated by npx for $x = 0, 1, 2, ...$ and $i = 1, 2, 3, ..., k$.

The critical value is calculated from the chi-squared distribution with degree of freedom $k - p - 1$, where k represents the number of categories while p is the number of parameter distribution. With the test criterion is to reject the null hypothesis if the chi-squared test statistics are greater than the quantile value of the chi-squared distribution at the real level of α and degree of freedom $k - p - 1$ or $\chi^2 \geq \chi^2_{(k-p-1)(1-\alpha)}$.

2.7 Regression Model Conformity

In order to select the best model in regression analysis, one of the methods to be used is Akaike's Information Criterion (AIC), Bayes Information Criterion (BIC) dan Root Mean Squared Error (RMSE). Here is the formula of AIC.

$$\text{AIC} = -2 \ln L(\hat{\theta}) + 2k \qquad (22)$$

with $L(\tilde{\theta})$ is Maximum Likelihood Function and k is the number of parameters.

The criteria for the merits of the model with the Bayesian or BIC approach are formulated as follows.

$$\text{BIC} = -2 \ln L(\hat{\theta}) + k \ln(n) \qquad (23)$$

where n denotes the sample size.

Root Mean Squared Error (RMSE) is one indicator of the model goodness in regression analysis. RMSE calculation is done by calculating the average difference between the predicted value or what is often called an error, sot that estimates of model accuracy can be known. The smaller the RMSE value is close to zero, the better the model is assumed because it can predict the true value more.

If several n-sized samples of a model with errors are known ϵ to be calculated as e_i $(i = 1, 2, ..., n)$ and assumed set ϵ is unbiased, so that RMSE is calculated by the formula as follows [23].

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^{2}} \qquad (24)$$

2.8 Data Analysis Method

This research is a descriptive study with a secondary data analysis approach obtained from the West Java Provincial Health Office through the 2022 West Java Provincial Health Profile report. This study focuses more on analyzing the factors that influence infant mortality. Data analysis was performed through descriptive statistics, Poisson regression analysis, and *Generalized Poisson Regression* (GPR) analysis to determine the relationship between predictor variables and infant mortality and determine the best model that explains the phenomenon. Here are the steps of analysis in this study.

1. Describe the factors that influence infant mortality using descriptive statistical analysis.

2. Obtain appropriate models for the number of infant mortality cases and factors affecting infant mortality by the Poisson regression method. Here are the steps on the Poisson regression method.

   a. Detect the presence of two or more predictor variables in a regression model correlated with each other.

   b. Stimulating the parameters of the Poisson regression model.

   c. Perform simultaneous and partial significance tests on Poisson regression model parameters.

   d. Perform equidispersion assumption tests on Poisson regression models.

   e. Obtained Poisson regression models as well as AIC, BIC, and RMSE values.

3. Getting a suitable model using Generalized Poisson Regression (GPR). Here are the steps of modelling with GPR.

   a. Stimulates the parameters of the Generalized Poisson Regression (GPR) model.

   b. Conduct simultaneous and partial significance tests on the parameters of the Generalized Poisson Regression (GPR) model.

4. Validate model conformance using AIC, BIC, and RMSE.

3. RESULTS AND DISCUSSION

3.1 Description of The Infant Deaths in West Java 2022

To see the characteristics of each variable, descriptive statistics will be displayed in the table as follows.

Table 1. Descriptive Statistics

Variable	Mean	StDev	Minimum	Median	Maximum
Number of Infant Mortality (Y)	109.6	70.2	23	93	300
K4-Coverage Percentage (X1)	97.22	7.25	87.7	97.7	125.3
Number of High-Risk Pregnant Women (X2)	6501	5641	116	4822	24422
Percentage of Families with PHBS (X3)	63.63	10.78	43.88	64.24	81.59
Percentage of Exclusively Breastfeeding (X4)	70.5	9.43	51.7	72.1	86.3
Immunization Percentage (X5)	103.26	11.05	85	99.7	132.5

Based on **Table 1**, it is found that the number of infant deaths in West Java in 2022 is an average of 109.6 infants for each district. The lowest number of deaths was 23 infant deaths in Bekasi Regency and the highest number was Garut Regency with 300 deaths. In addition, the difference in the number of high infant deaths is also due to a high variance of 4921.9. This high variance illustrates the uneven distribution of infant mortality in West Java and is concentrated in certain areas where the three districts with the highest number of infant deaths are in adjacent areas, namely in Garut Regency, Bandung Regency, and Sumedang Regency. While the lowest number was in Bekasi Regency with 23 infant deaths.

The highest variance value for the predictor variable that is thought to affect the number of infant deaths in West Java is variable X2, namely the number of pregnant women (high risk pregnant women) has an average of 6501 people for each district with the highest rates being in Bogor Regency, Bekasi Regency, and Bandung Regency. In addition, other variables are the percentage of K4 coverage for each district on average reaching 97.22%; the percentage of families with PHBS reached 63.63%; the percentage of exclusive breastfeeding reaches 70.5%; The percentage of immunization is 103.26%.

3.2 Poisson Regression

3.2.1 Poisson Distribution Assumption Test

Before Poisson regression analysis is carried out, it is necessary to test the suitability of the Poisson distribution on the response variable (Y), namely the number of infant deaths. The test was carried out using a null hypothesis, namely data on the number of infant deaths in West Java Province in 2022 distributed by Poisson. **Table 2** below is the result of calculating observation values and expectations for testing conformity with the Chi-Square test.

Table 2. Value of Expectation and Observation of Response Data

Number of Infant Mortality	Poisson Distribution Probability	Number of Observables	Number of Expectation	Chi-Square Contribution
≤ 110	0.540810	15	14.6019	0.01086
111	0.037499	0	1.0125	1.01246
$112 - 114$	0.106489	0	2.8752	2.87520
$115 - 117$	0.092253	3	2.4908	0.10408
≥ 118	0.222949	9	6.0196	1.47562
3 (60.00%) of total expectations less than 5				

The critical area of testing rejects H$_0$ if p $-$ value $< \alpha$. By using software, a p-value of the test is 0.14 so that p-value $< \alpha$ (5%). Therefore, the decision taken is to fail to reject H$_0$ or it can be concluded that the data on the number of infant deaths in West Java Province in 2022 is Poisson distributed so that Poisson regression analysis can be continued.

3.2.2 Multicollinearity Test

Multicollinearity examination is performed to determine the relationship between predictor variables. The criterion that can be used to identify the presence of cases of multicollinearity is the correlation between predictor variables. The variables that have a high correlation are the variables X_1 with X_5 (0.514); X_1 with X_2 (0.407); and X_2 with X_5 (0.271). Another criterion that can be used to identify cases of multicollinearity is the Variance Inflated Factor (VIF) value. The VIF values for each predictor variable can be seen in the following **Table 3.** From the results, it can be concluded that there are no cases of multicollinearity between predictor variables that have a VIF value of > 10. So, all these variables were used in modeling the number of infant mortality cases in West Java.

Table 3. VIF Value of Predictor Variables

Variable	VIF Value
x_1	1.43
x_2	1.33
x_3	1.03
x_4	1.07
x_5	1.29

3.2.3 Parameters Estimation of Poisson Regression

Data on the number of infant mortality cases are discrete data that follows the Poisson distribution. To determine the factors that affect the number of infant mortality cases in West Java, Poisson regression analysis was carried out. The estimation of Poisson regression parameters is shown in **Table 4** below.

Table 4. Estimation Parameter Model Poisson Regression

Parameter	Estimation	Standard Error
β_0	6.948	0.3764
β_1	-0.02801	0.003447
β_2	0.00002219	0.000003464
β_3	-0.01758	0.001832
β_4	-0.003644	0.002016
β_5	0.01593	0.002018

3.2.4 Parameters Significance Test of Poisson Regression

Furthermore, the significance of each parameter of the Poisson regression model was tested with the following hypothesis.

$H_0 : \beta_j = 0$ for $j = 1, 2, \dots, 5$
$H_1 : \beta_j \neq 0$ for $j = 1, 2, \dots, 5$
Critical region: Reject *H$_0$* if p-value $< \alpha$ (5%) or $|Z| > Z_{\frac{\alpha}{2}}$

Parameter significance testing was initially carried out on the five predictor variables, but there was an insignificant variable, namely X_4 with a p-value of 0.0707. These variables are then eliminated from the model and re-estimated for the remaining parameters as shown in **Table 5** below:

Table 5. Poisson Regression Model Partial Parameters Significance Test

Parameter	Estimation	Z-Score	P-Value
β_0	6.676	19.299	0.0000
β_1	-0.02830	-8.23871907	0.0000
β_2	0.00002344	6.9472436277	0.0000
β_3	-0.01787	-9.5961	0.0000
β_5	0.01645	7.8940	0.0000

3.2.5 Simultaneous Significance Test of Poisson Regression

In addition to testing parameters partially, parameter tests will be carried out simultaneously with the following hypotheses.

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_5 = 0$
$H_1 :$ At least there is one $\beta_j \neq 0$; $j = 1, 2, 3, 5$

Test statistics $D(\widehat{\boldsymbol{\beta}})$ are calculated by the following formula.

$$D(\widehat{\boldsymbol{\beta}}) = 2\left[In\left(L(\widehat{\Omega})\right) - In\left(L(\widehat{\omega})\right)\right]$$

$$D(\widehat{\boldsymbol{\beta}}) = 2\left[-463.2777 - (-630.9285)\right]$$

$$D(\widehat{\boldsymbol{\beta}}) = 2\left[167.6508\right]$$

$$D(\widehat{\boldsymbol{\beta}}) = 335.3016$$

For a significance level of 5% or 0.05 then the $(\alpha)\chi^2_{4;0,05}$ value is 0.710723 so the value is $D(\widehat{\boldsymbol{\beta}})$ much greater than $\chi^2_{4;0,05}$. Thus, the decision taken is to reject H_0 or it can be concluded that there is at least one predictor variable that has a significant effect on the Poisson regression model with selected variables.

3.2.6 Equidispersion Assumption Test

The assumption used in Poisson regression is that there is a similarity between the mean value and variance (equidispersion). Identification of cases of overdispersion was done with statistical deviance residual and Pearson chi-square. Here are the results of the calculation of both test statistics with the software.

Tabel 6. Poisson Regression Model Partial Parameter Significance Test

Goodness of fit	Statistics	Degree of Freedom
D^2	755.4273	22
X^2	774.3029	22

According to the **Equation (12)** and **Equation (13)**, the value of $\phi_1 = 34.3376$ and $\phi_2 = 35.19559$, while is obtained so that both are far above 1. Thus, it can be concluded that there is overdispersion in the estimation results of the Poisson regression model.

3.3 Generalized Poisson Regression

3.3.1 Parameters Estimation of Generalized Poisson Regression

After it is known that there is overdispersion in the Poisson regression model, the way to overcome it is to use Generalized Poisson Regression. The results of parameter estimation in the regression model are shown in **Table 7** as follows.

Table 7. Estimation Parameter Model Generalized Poisson Regression

Parameter	Estimation	Standard Error
β_0	6.257	1.753

Parameter	Estimation	Standard Error
θ	-3.213	0.1770
β_1	-0.02414	0.01728
β_2	0.00005489	0.00002234
β_3	-0.02193	0.009654
β_4	-0.006547	0.01140
β_5	0.02161	0.01094

3.3.2 Parameter Significance Test of Generalized Poisson Regression

Furthermore, the significance of each parameter of the Generalized Poisson regression model was tested with the following hypothesis.

$H_0 : \beta_j = 0 ; j = 1, 2, …, 5$
$H_1 : \beta_j \neq 0 ; j = 1, 2, …, 5$

and testing of the significance of dispersion parameters is carried out with the following hypotheses.

$H_0 : \theta = 0$
$H_0 : \theta \neq 0$

Critical region: Reject H_0 if p-value $< \alpha$ (5%) or $|Z| > Z_{\frac{\alpha}{2}}$

Parameter significance testing was initially carried out on the five predictor variables, but there was an insignificant variable, namely X_4 with a p-value of 0.0707 and X_5 with a p-value of 0.135263. These variables are then eliminated from the model and re-estimated for the remaining parameters as shown in **Table 8** below.

Table 8. **Generalized Poisson Regression Model Partial Parameter Significance Test**

Parameter	Estimation	Z-Score	*P-value*
β_0	5.942	3.578	0.00000
θ	-3.116	-17.94931	0.00000
β_2	0.00005981	3.095903	0.00869
β_3	-0.02622	-2.661364	0.01020

Based on **Table 8**, then the decision is to reject H_0 or it can be concluded that the dispersion parameter has an effect in the model. This can be interpreted that the addition of dispersion parameters to Generalized Poisson regression needs to be done to overcome the occurrence of overdispersion in the Poisson regression model ($\theta \neq 0$).

3.3.3 Simultaneous Significance Test of Generalized Poisson Regression

The next stage is testing the significance of the model with test statistics $D(\widehat{\boldsymbol{\beta}})$ as well as the following test hypotheses.

$H_0 : \beta_2 = \beta_3 = 0$
$H_1 : \beta_2 \neq \beta_3$

Test statistics $D(\widehat{\boldsymbol{\beta}})$ are calculated by the following formula.

$$D(\widehat{\boldsymbol{\beta}}) = 2\left[In\left(L(\widehat{\Omega})\right) - In(L(\widehat{\omega}))\right]$$

$$D(\widehat{\boldsymbol{\beta}}) = 2\left[-143.2607 - (-147.8982)\right]$$

$$D(\widehat{\boldsymbol{\beta}}) = 2\left[4.6375\right]$$

$$D(\widehat{\boldsymbol{\beta}}) = 9.275$$

For a significance level of 5% or 0.05 then the value $(\alpha)\chi^2_{2;0,05}$ is 0.102587 so the value is $D(\widehat{\boldsymbol{\beta}})$ greater than $\chi^2_{2;0,05}$. Thus, the decision taken is to reject *H0* or it can be concluded that there is at least one predictor variable that has a significant effect on the *Generalized Poisson* regression model with selected variables.

3.3.4 Model Interpretation of Generalized Poisson Regression

The increase or decrease in the number of infant deaths in West Java in 2022 depends on the coefficient of each variable. So, the Generalized Poisson Regression model is as follows.

$$\hat{\mu} = \exp(5.942 + 0.00005981X_2 - 0.02622X_3) \tag{25}$$

Based on this model on **Equation (25)**, it can be interpreted that when the number of pregnant women increases by one unit, the average number of infant deaths in West Java in 2022 will increase by $\exp(0.00005981) = 1.00006$ times provided that other variables are constant. This is in line with the fact that high-risk pregnancies can lead to an increased chance of infant death due to complications. If the percentage of families with PHBS increases by one unit, the average number of infant deaths in West Java in 2022 decreases by $\exp(0.02622) = 1.026567$ times or 2.6567% provided that other variables are constant. These results are in accordance with the theory, namely if more families run a clean and healthy lifestyle, it is expected to reduce the risk of exposure to diseases in infants who tend to be vulnerable and result in death. So, it can be concluded in the results of the Generalized Poisson Regression model that the number of infant deaths in West Java in 2022 will increase by one case for every increase in each predictor variable that has a significant effect of one unit assuming the value of other variables is fixed.

3.4 Regression Model Conformity

Criteria on model merit can use AIC, BIC and RMSE values. With the help of software, here are the results of a comparison of the virtues of Poisson regression model and Generalized Poisson Regression.

Table 9. Comparison of Poisson Regression and Generalized Poisson Regression

Model	AIC	BIC	RMSE
Poisson Regression	936.555	943.035	60.058
Generalized Poisson Regression	294.521	299.705	0.742

This study assessed the performance of Poisson Regression and Generalized Poisson Regression models using criteria such as AIC, BIC, and RMSE values to determine their suitability for modelling infant mortality data in West Java in 2022. As presented in **Table 9**, the Generalized Posson Regression model demonstrated substantially lower values of AIC (294.521), BIC (299.705), and RMSE (0.742) compared to the Poisson model (AIC: 936.555, BIC: 943.035, RMSE: 60.058). These results suggest that the Generalized Poisson Regression model fits the data more accurately and provides more efficient estimates and better predictive accuracy.

A review of these results indicates significant improvements in model performance by adopting the Generalized Poisson Regression approach. Lower AIC and BIC values typically signify a model with a better fit that maintains simplicity. At the same time, a reduced RMSE points to fewer errors on average, which is crucial for predictive accuracy and reliability in public health.

These findings align with those of **[17]**, who found that Generalized Poisson Regression outperforms ordinary Poisson models in handling overdispersion in count data, such as the number of infant mortalities. Overall, the superior performance of the Generalized Poisson Regression model in this study underscores its potential for enhancing the accuracy and reliability of infant mortality predictions in demographic studies. Future research could explore the integration of additional predictive variables and apply these findings to other regions and different demographic datasets to validate the model's generalizability.

4. CONCLUSIONS

Based on the discussion above, it can be concluded that the average number of infant deaths in West Java in 2022 reached 109.6 infants per district. The Poisson regression model indicates that the number of infant deaths will increase by one case for every one unit increase in each significant predictor variable (X1, X2, X3, X4, X5), assuming other variables remain constant. However, the presence of overdispersion in the model estimates suggests that Poisson regression may not fully explain the variability in the data. To address this issue, a Generalized Poisson Regression model, which includes additional dispersion parameters, is used. This model provides a more accurate picture of the data distribution, indicating that a one-unit increase in the

number of pregnant women will result in a 1.00006-fold increase in the average number of infant deaths, while a one-unit increase in the percentage of families practicing clean and healthy living behaviors (PHBS) will reduce the average number of infant deaths by 1.026567 or 2.66%, assuming other variables are constant. The model fit was evaluated using AIC, BIC, and RMSE, with the results showing that the Generalized Poisson Regression model provides a better fit for the infant mortality data in West Java, especially in handling overdispersion that cannot be addressed well by the standard Poisson regression model.

REFERENCES

[1] Kementerian Perencanaan Pembangunan Nasional/Badan Perencanaan Pembangunan Nasional, "Pilar Pembangunan Sosial Edisi II," Kedeputian Bidang Kemaritiman dan Sumber Daya Alam, Kementerian Perencanaan Pembangunan Nasional/Badan Perencanaan Pembangunan Nasional, Jakarta, 2020.

[2] Kementerian Perencanaan Pembangunan Nasional/Badan Perencanaan Pembangunan Nasional, "Pilar Pembangunan Sosial Edisi II," Kedeputian Kemaritiman dan Sumber Daya Kementerian Perencanaan Pembangunan Nasional/Badan Perencanaan Pembangunan Nasional, Jakarta, 2021.

[3] Dinas Kesehatan Provinsi Jawa Barat, "Profil Kesehatan Provinsi Jawa Barat Tahun 2021," Dinas Kesehatan Provinsi Jawa Barat, Bandung, 2021.

[4] World Health Organization, "Congenital Disorders," 27 February 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/birth-defects. [Accessed 21 November 2023].

[5] V. A. Rahma, "Analisis Regresi Binomial Negatif Untuk Memodelkan Jumlah Kematian Bayi di Jawa Timur Tahun 2014," 2016.

[6] A. Arisandi, E. T. Herdiani and S. Sahriman, "Aplikasi Generalized Poisson Regression dalam Mengatasi Overdispersi pada Data Jumlah Penderita Demam Berdarah Dengue," *Statistika,* vol. 18, no. 2, pp. 123-130, November 2018.

[7] D. Y. Kurniawan, Studi Banding Metode Generalized Poisson Regression dan Regresi Binomial Negatif untuk Mengatasi Overdispersi pada Data Diskrit (Studi Kasus Jumlah Kasus Baru TB di Provinsi Jawa Timur Tahun 2013), Surabaya: ADLN Perpustakaan Universitas Airlangga, 2015.

[8] A. D. Putri, D. Devianto and F. Yanuar, "PEMODELAN JUMLAH KEMATIAN BAYI DI KOTA BANDUNG DENGAN MENGGUNAKAN REGRESI ZERO-INFLATED POISSON," *Jurnal Matematika UNAND,* vol. Vol. 11 No. 1, pp. 12-24, Januari 2022.

[9] N. Y. Irkan, R. A. Ahri and S. , "Analisis Faktor yang Berhubungan dengan Kejadian Kematian Bayi," *JOURNAL OF MUSLIM COMMUNITY HEALTH (JMCH),* vol. III, pp. 24-32, 2022.

[10] Badan Pusat Statistik Provinsi Sulawesi Utara, "Jumlah Penduduk Menurut Provinsi di Indonesia (Ribu Jiwa), 2024," Badan Pusat Statistik Provinsi Sulawesi Utara, Manado, 2024.

[11] B. Durmus and O. I. Guneri, "An Application of The Generalized Poisson Model for Over Dispersion Data on The Number of Strikes Between 1984 and 2017," *Alphanumeric Journal,* vol. 8, no. 1, pp. 249-260, 2020.

[12] J. X. M.D, S. L. M. B. S, K. D, K. M. A and E. Arias, Ph.D., "Mortality in the United States, 2021," *National Center For Health Statistics,* Desember 2022.

[13] A. P. Rachmadiani, M. A. Shodikin and C. Komariah, "Faktor-Faktor Risiko Kematian Bayi Usia 0-28 Hari di RSD dr. Soebandi Kabupaten Jember," *Journal Of Agromedicine and Medical Sciences,* vol. Vol. 4 No. 2, 2018.

[14] L. M. N, D. Yuniarti and M. N. Hayati, "Penerapan Generalized Poisson Regression I Untuk Mengatasi Overdispersi Pada Regresi Poisson," *Jurnal EKSPONENSIAL,* Vols. Volume 7, Nomor 1, Mei 2016.

[15] S. Ardifasalma and U. Azmi, "Pemodelan Kasus Covid-19 di Jawa Timur Menggunakan Metode Generalized Poisson Regression dan Negative Binomial Regression," *JURNAL SAINS DAN SENI ITS,* Vols. Vol. 11, No. 6, pp. 2337-3520, 2022.

[16] P. R. Chaniago, D. Devianto and I. R. HG, "ANALISIS FAKTOR RISIKO ANGKA KEMATIAN IBU DENGAN PENDEKATAN REGRESI POISSON," *Jurnal Matematika UNAND,* vol. VII No. 2, pp. 126-131, 2018.

[17] A. D. Chaniago and S. P. Wulandari, "Pemodelan Generalized Poisson Regression (GPR) dan Negative Binomial Regression (NBR) untuk Mengatasi Overdispersi pada Jumlah Kematian Bayi di Kabupaten Probolinggo," *JURNAL SAINS DAN SENI ITS,* vol. 11, p. 8, 27 07 2022.

[18] R. P. Hendikawati and A. Agoestanto, "PEMODELAN GENERALIZED POISSON REGRESSION (GPR) UNTUK MENGATASI PELANGGARAN EQUIDISPERSI PADA REGRESI POISSON KASUS CAMPAK DI KOTA SEMARANG TAHUN 2013," *UNNES Journal Of Matjematics,* 2016.

[19] I. and D. P. Sari, "PEMODELAN REGRESI POISSON, BINOMIAL NEGATIF, DAN PADA KASUS KECELAKAAN KENDARAAN BERMOTOR," in *Seminar Nasional Matematika dan Pendidikan Matematika*, 2013.

[20] N. I. Islamiyah, "Pemodelan Generalized Poisson Regression (GPR) pada Faktor-Faktor yang Mempengaruhi Kasus Pneumonia pada Balita di Provinsi Sulawesi Selatan Tahun 2018," Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Alauddin Makassar, Makassar, 2020.

[21] M. Sriningsih, D. Hatidja and J. D. Prang, "Penanganan Multikolinieritas dengan Menggunakan Analisis Regresi Komponen Utama pada Kasus Impor Beras di Provinsi Sulut," *Jurnal Ilmiah Sains ,* vol. 18, no. 1, pp. 18-24, 2018.

[22] Y. Fahrizal and A. K. Mutaqin, "Pemodelan Distribusi Poisson-Amarendrapada Data Frekuensi Klaim Asuransi Kendaraan Bermotor di Indonesia," *Bandung Conference Series: Statistics,* vol. Vol. 3. No. 1, 2023.

[23] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature," *Geosci. Model Dev,* pp. 1247-1250, June 2014.