Eksakta : Berkala Ilmiah Bidang MIPA
VOLUME 24 NO 02 2023, pp 259-270
ISSN : Print 1411-3724 — Online 2549-7464
DOI : https://doi.org/10.24036/eksakta/vol24-iss02/363

Eksakta
Berkala Ilmiah Bidang MIPA
http://www.eksakta.ppj.unp.ac.id/index.php/eksakta

Article

Factors Affecting the Number of Infant Morality Cases in West Java for the 2019-2020 Period using Generalized Poisson Regression (GPR)

Article Info

Article history :

Received November 03, 2022
Revised May 15, 2023
Accepted May 23, 2023
Published June 30, 2023
(*In-Press*)

Keywords :

Equidispersion, generalized poisson regression, newton-raphson iteration

Kartika Dewi[1], Nurul Gusriani[1], Kankan Parmikanti[1]

[1]Department of Mathematics, Faculty of Mathematics and Natural Science (FMIPA), Universitas Padjadjaran, Bandung, Indonesia

Abstract. The number of infant mortality cases is data in the form of counts which is modeled by Poisson regression. There is an assumption that needs to be met, namely equidispersion. Equidispersion is a condition in which the mean and variance of the variables are the same, but in practice this assumption is often not met. There are two possible events, namely overdispersion and underdispersion. The Generalized Poisson Regression (GPR) model is one solution to solve this problem. In estimating the GPR parameter, the Maximum Likelihood Estimation (MLE) method is used, but the derivation of the log-likelihood function does not always produce explicit results, so the Newton-Raphson iteration method is used. Poisson regression analysis conducted on the number of infant mortality cases in West Java showed that the model had overdispersion as seen from the value of the dispersion parameter which was more than zero, so the GPR model was used. Parameter significance test was carried out on three factors, namely the poverty gap index (X_1), the percentage of low birth weight infants (X_2), and the percentage of exclusive breastfeeding for infants (X_3) the results obtained that all factors affected the number of infant mortality cases in West Java.

Corresponding Author :
Kartika Dewi
Department of Mathematics, Faculty of Mathematics and Natural Science (FMIPA), Universitas Padjadjaran, Bandung, Indonesia
Email : kartika18002@unpad.ac.id

1. Introduction

Infant mortality is an important indicator to see the degree of health in a community because the body of a newborn is very sensitive to the environment, so an increase in infant mortality can indicate a problem in the environment. There are many factors that can influence infant mortality. These factors can come from environmental conditions, maternal health, or congenital conditions such as low birth weight (LBW), respiratory infections, and a combination of neonatal disorders (babies less than 28 days old).

In Indonesia, the number of infant mortality is relatively high, namely 24 deaths per 1000 births. Reporting from Databoks.katadata.co.id (2021), in Indonesia in 2020, from a total of 28 thousand babies who died, there were 20 thousand babies (71.97%) died in the age range 0 to 28 days and as many as 5 thousand Infants (19.13%) died in the age range of 29 days to 11 months. On the other hand, one of the provinces that contributes the most infant mortality is West Java (Pikiran-rakyat.com, 2016).

To reduce the number of infant mortality, preventive measures are needed. Steps that can be taken by mothers is to give breast milk to the baby. From Sehatnegeriku.kemkes.go.id (2017), breastfeeding can reduce infant mortality due to infection by 88%. Another way to reduce infant mortality is to analyze the factors that influence it. The analysis that can be used to see the relationship between variables is regression analysis. The number of infant mortality cases is data in the form of count data which is definitely positive. The event can occur at a certain time with a small probability of happening. By looking at this, the number of infant mortality cases can be said to have a Poisson distribution. As the name implies, to model data with a Poisson distribution, Poisson regression is used.

In Poisson regression there is an assumption that needs to be met, namely equidispersion. Equidispersion is a condition where the mean and variance of the dependent variable are the same, but in practice this assumption is often not met [1-3]. The variance and mean values are often different which is a violation of poisson regression. These violations include cases of overdispersion and underdispersion. Overdispersion is an event where the value of the variance is greater than the mean value, while underdispersion is the opposite where the value of the variance is lower. The Generalized Poisson Regression (GPR) model is one solution to solve this problem. This is because the GPR model takes into account the dispersion factor in the model. The GPR model is also more flexible for data whose dispersion type is unknown [4-6].

There have been many studies on the application of GPR. Modeling the data of the American-Egyptian center which is the number of diseases using GPR while used GPR to overcome overdispersion in maternal mortality data [7-8]. From the research of [9], it was found that the Akaike's Information Criterion (AIC) value of the GPR model is smaller than the Poisson regression model so that the GPR model is better in modeling the data.

In finding the estimated value of the GPR model parameters, the Maximum Likelihood Estimation (MLE) method is usually used, but according to [10-11], the derivation of the likelihood function does not always produce explicit results, so we use additional methods to get a convergent parameter estimator. Research conducted by [10-11] used the Iteratively Weighted Least Square (IWLS) method as an addition to estimating the parameters, but in this study the MLE method was used with additional Newton-Raphson iterations to obtain the estimated parameters of the GPR model.

The formulas contained in the Newton-Raphson iteration method cannot be directly applied to the data, so the process requires formula derivation. The derivation of the formula will be carried out in forming a gradient vector and a Hessian matrix where the equation is obtained through the MLE method in the previous stage. The gradient vector and the Hessian matrix that are formed will then be broken down in such a way so they can be directly applied to the data, only after that the Newton-Raphson iteration method can be used to find the estimated parameters of the GPR model.

The parameters obtained will be used to model the number of infant mortality cases as the dependent variable Y to the independent variable X, namely factors that are thought to affect the number of infant mortality cases such as the poverty gap index, the percentage of low birth weight babies (LBW), and the percentage of exclusive breastfeeding for infants. The research variables consisted of 27 districts/cities in West Java in the 2019-2020 period.

2. Materials and Methods

2.1. Materials

This study uses the number of infant mortality cases in West Java for the 2019-2020 period as the Y variable, with three independent variables X, namely the factors that are thought to have an effect such as the poverty gap index (X_1), the percentage of low birth weight babies (X_2), and the percentage of exclusive breastfeeding for infants (X_3) obtained from the official website of the West Java regional government (opendata.jabarprov.go.id) with 54 observations (n). The method used to estimate the number of infant mortality parameters based on the GPR model is the MLE method with additional Newton-Raphson iteration methods.

2.2 Multicollinearity

Before modeling the data, it is necessary to first see whether the data has multicollinearity, namely the linear relationship between independent variables in a regression model. To detect the presence of multicollinearity in the multiple linear regression model, the Variance Inflation Factor (VIF) value can be used. The VIF formula can be written as follows [12-14]:

$$VIF_j = \frac{1}{1-R_j^2}. \quad (1)$$

Multicollinearity occurs when $VIF > 10$ [15-16]. If multicollinearity occurs, then eliminate the variable with a high VIF value.

2.3 Newton-Raphson Iteration

The reduction of the likelihood function does not always produce an explicit value and is analytically difficult to do so numerical methods are used, namely the Newton-Raphson [12]. In general, the Newton-Raphson method is taken from a Taylor series of degree two around its parameter estimate with the following equation:

$$\widehat{\boldsymbol{\beta}}_{(r+1)} = \widehat{\boldsymbol{\beta}}_{(r)} - \mathbf{H}^{-1}\left(\widehat{\boldsymbol{\beta}}_{(r)}\right)\mathbf{g}\left(\widehat{\boldsymbol{\beta}}_{(r)}\right). \quad (2)$$

when **g** represents the gradient vector containing the first partial derivative of the likelihood function, **H** is the Hessian matrix, i.e. the symmetry matrix containing the second partial derivative of the likelihood function.

Newton-Raphson iteration is carried out until a convergent parameter estimate is obtained. The steps of the Newton-Raphson iteration are as follows [17]:

1. Determine the initial estimated value $\widehat{\boldsymbol{\beta}}_{(0)}$,
2. Form a gradient vector $\mathbf{g}(\widehat{\boldsymbol{\beta}}_{(r)})$ starting from $r = 0$,

$$\mathbf{g}(\widehat{\boldsymbol{\beta}}_{(r)})_{(k+1)\times 1} = \begin{bmatrix} \frac{\partial \ln L(y)}{\partial \beta_{0(r)}} \\ \frac{\partial \ln L(y)}{\partial \beta_{1(r)}} \\ \frac{\partial \ln L(y)}{\partial \beta_{2(r)}} \\ \vdots \\ \frac{\partial \ln L(y)}{\partial \beta_{k(r)}} \end{bmatrix}$$

where k is the number of parameters to be estimated and r is the iteration,

3. Form a Hessian matrix $\mathbf{H}^{-1}\left(\widehat{\boldsymbol{\beta}}_{(r)}\right)$ starting from $r = 0$, namely:

$$\left[\mathbf{H}\left(\widehat{\boldsymbol{\beta}}_{(r)}\right)\right]^{-1}_{(k+1)\times(k+1)} = \begin{bmatrix} \frac{\partial^2 \ln L(y)}{\partial \beta_0^2} & \frac{\partial^2 \ln L(y)}{\partial \beta_0\, \partial \beta_1} & \cdots & \frac{\partial^2 \ln L(y)}{\partial \beta_0\, \partial \beta_k} \\ \frac{\partial^2 \ln L(y)}{\partial \beta_1\, \partial \beta_0} & \frac{\partial^2 \ln L(y)}{\partial \beta_1^2} & \cdots & \frac{\partial^2 \ln L(y)}{\partial \beta_1\, \partial \beta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L(y)}{\partial \beta_k\, \partial \beta_0} & \frac{\partial^2 \ln L(y)}{\partial \beta_k\, \partial \beta_1} & \cdots & \frac{\partial^2 \ln L(y)}{\partial \beta_k^2} \end{bmatrix}^{-1}$$

4. Calculating the value of $\widehat{\boldsymbol{\beta}}_{(r+1)}$ through equation (2),
5. Iterate until you get a convergent value of $\widehat{\boldsymbol{\beta}}$.

The iteration stops when the value of has converged, namely when $\left\|\widehat{\boldsymbol{\beta}}_{(r+1)} - \widehat{\boldsymbol{\beta}}_{(r)}\right\| \leq \varepsilon$, where ε is a tolerance value in the form of a very small positive number (usually taken $\varepsilon = 10^{-5}$). The notation $\|.\|$ denotes the length (norm) of the vector, which is the distance between the two vectors being searched for.

2.4 Poisson Regression Model

Poisson regression uses the Generalized Linear Model (GLM) principle so that it can be used in observations, where the GLM model uses a connecting function, namely a regression model that connects the mean of the dependent variable Y with the independent variable X. In the Poisson regression model, the connecting function used is the log function so that the mean of the dependent variable will be in the form of an exponential function and also guarantee the value of the variable in it is non-negative. The Poisson regression connecting function is as follows [17]:

$$ln(\lambda_i) = \mathbf{x}_i'\boldsymbol{\beta}, \tag{3}$$

$$\lambda_i = exp(\mathbf{x}_i'\boldsymbol{\beta}) = exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}), \tag{4}$$

where $\mathbf{x}_i'$ is a vector of size $(k \times 1)$ containing the i-th row of the matrix $\mathbf{X}$ with the symbol $'$ representing the transpose, $\mathbf{X}$ is a matrix of size $(n \times (k+1))$ containing independent variables, and $\boldsymbol{\beta}$ is a vector of size $((k+1) \times 1)$ which contains the coefficients of the regression parameters.

The probability function of the Poisson distribution is

$$P(y_i; \beta) = \frac{[\lambda_i(x_i; \beta)]^{y_i} e^{-[\lambda_i(x_i;\beta)]}}{y_i!}, \tag{5}$$

then a Poisson regression model is formed with the log link function as follows::

$$\begin{aligned} y_i = \lambda_i + \varepsilon_i &= exp(\mathbf{x}_i'\boldsymbol{\beta}) + \varepsilon_i \\ &= exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}) + \varepsilon_i, \end{aligned} \tag{6}$$

or estimated model

$$\hat{y}_i = \hat{\lambda}_i = exp\left(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}\right) = exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}\right). \tag{7}$$

In the Poisson regression model there is an assumption that must be met, namely equidispersion, where the value of the variance of the dependent variable Y must be equal to its mean value or $Var(y_i|x_i) = E(y_i|x_i) = \lambda_i$ [18-19].

In estimating the Poisson regression parameters, the MLE approach will be used, the parameter estimation steps are as follows [20-21]:

1. Forming the likelihood function of the Poisson distribution as follows:

$$L(y; \beta) = \frac{\{\prod_{i=1}^{n}[\lambda_i]^{y_i}\} e^{-\sum_{i=1}^{n}[\lambda_i]}}{\prod_{i=1}^{n} y_i!}. \tag{8}$$

2. Forming the log-likelihood function from equation (8) as follows:

$$\ell_{poi} = ln\left(L(y; \beta)\right) = \sum_{i=1}^{n}[y_i\,(\mathbf{x}_i'\boldsymbol{\beta}) - exp(\mathbf{x}_i'\boldsymbol{\beta}) - ln(y_i!)]. \tag{9}$$

3. Deriving the log-likelihood function from equation (9) for each parameter and then equating it with zero [17]:

$$\frac{\partial \ell_{poi}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}[y_i x_i - x_i \, exp(\mathbf{x}_i'\boldsymbol{\beta})] = 0. \quad (10)$$

Because the function obtained using MLE is still implicit, the Newton-Raphson iteration method is used to obtain the estimated parameters of the Poisson regression model. The steps are as follows:

1. Determining the initial estimated value $\widehat{\boldsymbol{\beta}}_{(0)}$.
$$\widehat{\boldsymbol{\beta}}_{(0)} = [0 \quad 0 \quad \cdots \quad 0]^{-1}$$
2. Multiplying the matrix $\mathbf{X}$ by $\widehat{\boldsymbol{\beta}}_{(r)}$ starting from $r = 0$, we get
$$\mathbf{X}\widehat{\boldsymbol{\beta}}_{(r)} = [\mathbf{x}_1'\boldsymbol{\beta}_{(r)} \quad \mathbf{x}_2'\boldsymbol{\beta}_{(r)} \quad \cdots \quad \mathbf{x}_n'\boldsymbol{\beta}_{(r)}]^{-1}$$
3. Forming the vector $\mathbf{s}$ by exponentiating the second step, we get:
$$\mathbf{s} = exp\left(\mathbf{X}\widehat{\boldsymbol{\beta}}_{(r)}\right) = \left[exp\left(\mathbf{x}_1'\boldsymbol{\beta}_{(r)}\right) \quad exp\left(\mathbf{x}_2'\boldsymbol{\beta}_{(r)}\right) \quad \cdots \quad exp\left(\mathbf{x}_n'\boldsymbol{\beta}_{(r)}\right)\right]^{-1}$$
4. Forming a gradient vector $\mathbf{g}(\widehat{\boldsymbol{\beta}}_{(r)})$,
To form a gradient vector, first find the value of the first partial derivative of the Poisson regression log-likelihood function, then the following equation is obtained:
$$\mathbf{g}\left(\widehat{\boldsymbol{\beta}}_{(r)}\right) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{s}, \quad (11)$$
$\mathbf{y}$ is a vector of size $(n \times 1)$ which contains the dependent variable,
5. Forming an inverse Hessian matrix $\left[\mathbf{H}\left(\widehat{\boldsymbol{\beta}}_{(r)}\right)\right]^{-1}$,
To form the Hessian matrix, first find the value of the second partial derivative of the Poisson regression log-likelihood function, then obtain the following equation:
$$\left[\mathbf{H}\left(\widehat{\boldsymbol{\beta}}_{(r)}\right)\right]^{-1} = [-\mathbf{X}'\mathbf{C}\mathbf{X}]^{-1}, \quad (12)$$
where
$$\mathbf{C} = \begin{bmatrix} exp\left(\mathbf{x}_1'\boldsymbol{\beta}_{(r)}\right) & 0 & \cdots & 0 \\ 0 & exp\left(\mathbf{x}_2'\boldsymbol{\beta}_{(r)}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & exp\left(\mathbf{x}_n'\boldsymbol{\beta}_{(r)}\right) \end{bmatrix} = diag(\mathbf{s}). \quad (13)$$
6. Calculate $\widehat{\boldsymbol{\beta}}_{(r+1)}$ using equation (2),
7. Iterate until you get a convergent value of $\widehat{\boldsymbol{\beta}}$.

2.5 Overdispersion and Underdispersion

After estimating the parameters of the Poisson regression model, it is then sought whether the model meets the equidispersion assumption. To test whether the data has overdispersion or underdispersion, it is necessary to look for the dispersion parameters, which are as follows [22-24]:

$$\delta = \frac{D}{n-k-1}, \quad (14)$$

where

$$Deviance: D = 2\sum_{i=1}^{n}\left\{y_i \, ln\left(\frac{y_i}{\lambda_i}\right) - (y_i - \lambda_i)\right\}, \quad (15)$$

and $\lambda_i = exp(\mathbf{x}_i'\boldsymbol{\beta})$.

The value of $\delta < 0$ indicates that the model is underdispersion, i.e. when the value of the variance is smaller than its mean value, on the other hand, when $\delta > 0$ indicates that the model is experiencing overdispersion, that is, when the value of the variance is greater than its mean value [12].

2.6 GPR Model

GPR is an alternative to the data model (count data) which contains both overdispersion and underdispersion. This is due to the addition of the dispersion parameter (θ) into the model. The probability distribution function can be written as [17]:

$$f(y) = \begin{cases} \dfrac{\theta(\theta + y\gamma)^{y-1}e^{-\theta-y\gamma}}{y!} & ; y = 0,1,2,\dots \\ 0 \quad ; \text{for } y > m, & \text{when } \gamma > 0 \end{cases}, \tag{16}$$

where $\theta > 0, \max\left[-1, -\frac{\theta}{m}\right] < \gamma \leq 1$, and $m \geq 4$ is the largest positive integer for which $\theta + m\gamma > 0$ when γ is negative.

In estimating GPR parameters, the MLE approach will be used. Based on the GPR probability function in equation (16), the steps for parameter estimation using the MLE method are as follows:

1. Forming the likelihood function of the GPR distribution as follows:
$$L(y;\beta,\delta) = \prod_{i=1}^{n}\left(\frac{\lambda_i}{1+\delta\lambda_i}\right)^{y_i}\prod_{i=1}^{n}\frac{(1+\delta y_i)^{y_i-1}}{y_i!}\exp\left(-\sum_{i=1}^{n}\frac{\lambda_i(1+\delta y_i)}{1+\delta\lambda_i}\right). \tag{17}$$
2. Forming the log-likelihood function from equation (17) as follows [9]:
$$\ell_{gpr} = ln\big(L(y;\beta,\delta)\big) = \sum_{i=1}^{n}\left[y_i\ ln\left(\frac{\lambda_i}{1+\delta\lambda_i}\right) + (y_i-1)\ ln(1+\delta y_i) - \frac{\lambda_i(1+\delta y_i)}{1+\delta\lambda_i} - ln(y_i!)\right]. \tag{18}$$
3. Deriving the log-likelihood function from equation (18) for each parameter and then equating it with zero [9]:
$$\frac{\partial \ell_{gpr}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\left[\frac{(y_i-\lambda_i)x_i}{(1+\delta\lambda_i)^2}\right] = 0, \tag{19}$$
and
$$\frac{\partial \ell_{gpr}}{\partial \delta} = \sum_{i=1}^{n}\left\{\left[\frac{-y_i\lambda_i}{1+\delta\lambda_i}\right] + \left[\frac{(y_i-1)y_i}{1+\delta y_i}\right] - \left[\frac{(y_i-\lambda_i)\lambda_i}{(1+\delta\lambda_i)^2}\right]\right\} = 0. \tag{20}$$

Because the function obtained using MLE is still implicit, the Newton-Raphson iteration method is used to obtain the estimated parameters of the GPR model. The steps are as follows:

1. Determining the initial estimated value $\widehat{\boldsymbol{\beta}}_{(0)}$.
$$\widehat{\boldsymbol{\beta}}_{(0)} = [0 \quad 0 \quad \cdots \quad 0]^{-1}$$
2. Multiplying the matrix $\mathbf{X}$ by $\widehat{\boldsymbol{\beta}}_{(r)}$ starting from $r = 0$, we get
$$\mathbf{X}\widehat{\boldsymbol{\beta}}_{(r)} = [\mathbf{x}_1'\boldsymbol{\beta}_{(r)} \quad \mathbf{x}_2'\boldsymbol{\beta}_{(r)} \quad \cdots \quad \mathbf{x}_n'\boldsymbol{\beta}_{(r)}]^{-1}$$
3. Forming the vector $\mathbf{s}$ by exponentiating the second step, we get:
$$\mathbf{s} = exp\big(\mathbf{X}\widehat{\boldsymbol{\beta}}_{(r)}\big) = \big[exp\big(\mathbf{x}_1'\boldsymbol{\beta}_{(r)}\big) \quad exp\big(\mathbf{x}_2'\boldsymbol{\beta}_{(r)}\big) \quad \cdots \quad exp\big(\mathbf{x}_n'\boldsymbol{\beta}_{(r)}\big)\big]^{-1}$$
4. Forming a gradient vector $\mathbf{g}(\widehat{\boldsymbol{\beta}}_{(r)})$,
To form a gradient vector, first find the value of the first partial derivative of the GPR log-likelihood function, then the following equation is obtained:
$$\mathbf{g}\big(\widehat{\boldsymbol{\beta}}_{(r)}\big) = \mathbf{X}'\mathbf{V}^2\mathbf{y} - \mathbf{X}'(\mathbf{V}^2\mathbf{s}), \tag{21}$$
where
$$\mathbf{V} = [\mathbf{I} + \delta\mathbf{C}], \tag{22}$$
and $\mathbf{I}$ is an Identity matrix.
5. Forming an inverse Hessian matrix $\left[\mathbf{H}\big(\widehat{\boldsymbol{\beta}}_{(r)}\big)\right]^{-1}$,
To form the Hessian matrix, first find the value of the second partial derivative of the GPR log-likelihood function, then obtain the following equation:
$$\left[\mathbf{H}\big(\widehat{\boldsymbol{\beta}}_{(r)}\big)\right]^{-1} = \big[(\mathbf{X}'\mathbf{C})\big(\mathbf{V}^3\mathbf{A}\big)\mathbf{X}\big]^{-1}, \tag{23}$$
where
$$\mathbf{A} = \delta\mathbf{C} - 2\delta\mathbf{D} - \mathbf{I}, \tag{24}$$

and

$$\mathbf{D} = \begin{bmatrix} y_1 & 0 & \cdots & 0 \\ 0 & y_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_n \end{bmatrix} = diag(\mathbf{y}), \quad (25)$$

6. Calculate $\widehat{\boldsymbol{\beta}}_{(r+1)}$ using equation (2),
7. Iterate until you get a convergent value of $\widehat{\boldsymbol{\beta}}$.

2.7 Simultaneous Test and Partial Test

Simultaneous significance testing for parameter estimation of the GPR model uses the likelihood ratio test with the following hypothesis:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ (there is no independent variable that has an effect of Y variable)

H_1: at least one $\beta_k \neq 0$ (there is at least one independent variable that has an effect of Y variable)

The statistics used are G test statistics, which are as follows [12]:

$$G = -2\, ln\left(\frac{h_0}{h_1}\right) = 2[ln(h_1) - ln(h_0), \quad (26)$$

where h_1 is the log-likelihood value of the model containing all independent variables and h_0 is the log-likelihood value of the model without independent variables.

For GPR, we get

$$G_{gpr} = 2\sum_{i=1}^{n}\left[y_i\ ln\left(\frac{exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\delta\ exp(\mathbf{x}_i'\boldsymbol{\beta})}\right) - \frac{exp(\mathbf{x}_i'\boldsymbol{\beta})(1+\delta y_i)}{1+\delta\ exp(\mathbf{x}_i'\boldsymbol{\beta})}\right] - \left[y_i\ ln\left(\frac{exp(\beta_0)}{1+\delta\ exp(\beta_0)}\right) - \frac{exp(\beta_0)(1+\delta y_i)}{1+\delta\ exp(\beta_0)}\right]. \quad (27)$$

The decision-making criteria is to reject H_0 if the value of $G_{gpr} > \chi^2_{(\alpha,db)}$ where $\chi^2_{(\alpha,db)}$ is the value of the Chi-Square table at the level of accuracy α with $db = n - k - 1$.

Partial testing using the Wald test [12], the hypothesis used is as follows:

H_0: $\beta_j = 0$ (the j-th independent variable has no effect on the dependent variable)

H_1: $\beta_j \neq 0$ (the j-th independent variable has effect on the dependent variable)

The test statistics used are as follows [13-15]:

$$W_j = \frac{\widehat{\beta}_j^2}{Var(\widehat{\beta}_j)} \sim \chi^2_{(1)}, \quad (28)$$

with

$$Var(\hat{\beta}_j) = \left|\frac{\partial^2 ln(L(y;\beta,\delta))}{\partial\widehat{\beta}_j^2}\right|^{-1}, \quad (29)$$

where

$\hat{\beta}_j$: j-th parameter estimate

The test criteria is to reject H_0 if $W_j > \chi^2_{(\alpha,1)}$ where $\chi^2_{(\alpha,1)}$ is the value of *Chi-Square* table at the level of accuracy α with degrees of freedom 1. Or reject H_0 jika $p - value > \alpha$.

3. Results and Discussion

3.1 Multicollinearity Test Results

Before looking for the Poisson parameter estimation, first look for whether there is multicollinearity in the data or there is a relationship between the independent variables. Multicollinearity detection using the R Studio program produced value of each variable, namely, $X_1 = 1{,}30$; $X_2 = 1{,}63$; and $X_3 = 1{,}33$. All of these values are less than 10, so it can be concluded that there is no multicollinearity in the data. Because there is no multicollinearity in the data, the data can be used to model the number of infant mortality cases in West Java.

3.2 Parameter Estimation Results of Poisson Regression Model

Calculations to get the estimated value of the Poisson regression model parameters are carried out using the Newton-Raphson iteration method with a tolerance value of 10^{-5}. The iteration results are as follows:

1st iteration (r=0)

1. Determining the initial estimated value $\widehat{\boldsymbol{\beta}}_{(0)}$, the initial estimated value is taken to be 0 so that we get $\hat{\beta}_{0(0)} = \hat{\beta}_{1(0)} = \hat{\beta}_{2(0)} = \hat{\beta}_{3(0)} = 0$
2. Multiplying the matrix $\mathbf{X}$ by $\widehat{\boldsymbol{\beta}}_{(0)}$ then change into exponential form, we get

$$\mathbf{s} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

3. Forming a gradient vector $\mathbf{g}(\widehat{\boldsymbol{\beta}}_{(0)})$ using equation (11), we get

$$\mathbf{g}(\widehat{\boldsymbol{\beta}}_{(0)}) = \begin{bmatrix} 1{,}15 \\ 2{,}53 \\ 4{,}57 \\ 30{,}18 \end{bmatrix}$$

4. Forming an inverse Hessian matrix $\left[\mathbf{H}(\widehat{\boldsymbol{\beta}}_{(0)})\right]^{-1}$ using equation (12), we get

$$\left[\mathbf{H}(\widehat{\boldsymbol{\beta}}_{(0)})\right]^{-1} = \begin{bmatrix} -0{,}63 & 0{,}13 & -0{,}00 & 0{,}01 \\ 0{,}14 & -0{,}20 & 0{,}03 & 0{,}00 \\ -0{,}00 & 0{,}03 & -0{,}03 & 0{,}00 \\ 0{,}01 & 0{,}00 & 0{,}00 & -0{,}00 \end{bmatrix}$$

5. Calculate $\widehat{\boldsymbol{\beta}}_{(1)}$ using equation (2), we get

$$\widehat{\boldsymbol{\beta}}_{(1)} = \begin{bmatrix} 0{,}19 \\ 0{,}23 \\ 0{,}04 \\ -0{,}01 \end{bmatrix}$$

We get $\left\|\widehat{\boldsymbol{\beta}}_{(1)} - \widehat{\boldsymbol{\beta}}_{(0)}\right\| = \mathbf{0,30}$; where the value is still greater than $\mathbf{10^{-5}}$, so the calculation continues to the next iteration.

The iteration is continued until a convergent value of is obtained. The estimation of the Poisson regression model parameters using R Studio software obtained a convergent value in the 4th iteration with the results, $\hat{\beta}_0 = 0{,}22$; $\hat{\beta}_1 = 0{,}22$; $\hat{\beta}_2 = 0{,}03$; and $\hat{\beta}_3 = -0{,}01$; with $\left\|\widehat{\boldsymbol{\beta}}_{(4)} - \widehat{\boldsymbol{\beta}}_{(3)}\right\| = 6{,}97 \times 10^{-7} < \varepsilon = 10^{-5}$, so that the Poisson regression model $(\hat{\lambda})$ is as follows:

$$\hat{\lambda} = exp(0{,}22 + 0{,}22x_1 + 0{,}03x_2 - 0{,}01x_3). \qquad (30)$$

After obtaining the Poisson regression model in equation (30), then look for whether the model meets the assumption of equidispersion or not. The test was carried out by finding the dispersion parameter (δ) using R Studio software and we get the value of the dispersion parameter of the Poisson regression model is $\delta = 35.26 > 0$, so the model is declared to have overdispersion. Because the model violates the equidispersion assumption, it is necessary to use GPR to model the data.

3.3 Parameter Estimation Results of GPR Model

Calculations to get the estimated value of the GPR model parameters are carried out using the Newton-Raphson iteration method with a tolerance value of 10^{-5}. The iteration results are as follows:

1st iteration (r=0)

1. Determining the initial estimated value $\widehat{\boldsymbol{\beta}}_{(0)}$, the initial estimated value is taken to be 0 so that we get $\hat{\beta}_{0(0)} = \hat{\beta}_{1(0)} = \hat{\beta}_{2(0)} = \hat{\beta}_{3(0)} = 0$
2. Multiplying the matrix $\mathbf{X}$ by $\widehat{\boldsymbol{\beta}}_{(0)}$ then change into exponential form, we get

$$\mathbf{s} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

3. Forming a gradient vector $\mathbf{g}(\widehat{\boldsymbol{\beta}}_{(0)})$ using equation (21), we get

$$\mathbf{g}(\widehat{\boldsymbol{\beta}}_{(0)}) = \begin{bmatrix} 1{,}15 \\ 2{,}53 \\ 4{,}57 \\ 30{,}18 \end{bmatrix}$$

4. Forming an inverse Hessian matrix $\left[\mathbf{H}(\widehat{\boldsymbol{\beta}}_{(0)})\right]^{-1}$ using equation (23), we get

$$\left[\mathbf{H}(\widehat{\boldsymbol{\beta}}_{(0)})\right]^{-1} = \begin{bmatrix} -0{,}63 & 0{,}13 & -0{,}00 & 0{,}01 \\ 0{,}14 & -0{,}20 & 0{,}03 & 0{,}00 \\ -0{,}00 & 0{,}03 & -0{,}03 & 0{,}00 \\ 0{,}01 & 0{,}00 & 0{,}00 & -0{,}00 \end{bmatrix}$$

5. Calculate $\widehat{\boldsymbol{\beta}}_{(1)}$ using equation (2), we get

$$\widehat{\boldsymbol{\beta}}_{(1)} = \begin{bmatrix} 0{,}19 \\ 0{,}23 \\ 0{,}04 \\ -0{,}01 \end{bmatrix}$$

We get $\left\|\widehat{\boldsymbol{\beta}}_{(1)} - \widehat{\boldsymbol{\beta}}_{(0)}\right\| = \mathbf{0,30}$; where the value is still greater than $\mathbf{10^{-5}}$, so the calculation continues to the next iteration.

The iteration is continued until a convergent value of is obtained. The estimation of the GPR model parameters using R Studio software obtained a convergent value in the 3rd iteration with the results, $\hat{\beta}_0 = 0{,}20$; $\hat{\beta}_1 = 0{,}23$; $\hat{\beta}_2 = 0{,}04$; and $\hat{\beta}_3 = -0{,}01$; with $\left\|\widehat{\boldsymbol{\beta}}_{(4)} - \widehat{\boldsymbol{\beta}}_{(3)}\right\| = 3{,}27 \times 10^{-6} < \varepsilon = 10^{-5}$, so that The GPR provisional model formed is as follows:

$$\hat{\lambda} = exp(0{,}20 + 0{,}23x_1 + 0{,}04x_2 - 0{,}01x_3). \qquad (31)$$

The next step is to test the significance of the parameters simultaneously using the Likelihood Ratio Test and partially using the Wald Test to find out which factors affect the number of infant mortality cases in West Java.

3.4 Simultaneous Test Results of GPR Model

Based on the GPR model obtained through equation (31), the test statistic value obtained is $G_{gpr} = 22033{,}22$. The value of the Chi-Square table with an accuracy level of $\alpha = 0{,}05$ and degrees of freedom 50 is 67,50; so the value of $G_{gpr} > \chi^2_{(0{,}05;50)}$. Therefore, H_0 is rejected, which means that there is at least one significant parameter/at least one independent variable that contributes to the number of infant mortality cases in West Java, or in other words the data can be modeled with GPR.

3.5 Partial Test Results of GPR Model

The calculation is based on the model that has been obtained in equation (31) using the Wald test value in equation (28) and the value of $Var(\hat{\beta}_j)$ is calculated through equation (29). The value obtained is then compared with the value of the Chi-Square table at an accuracy level of $\alpha = 0{,}05$ and a degree of freedom 1, which is 3,84. The results of the Wald test for each parameter are as follows:

1. Poverty gap index factor (X_1)
   The test statistic value $W_1 = 25{,}54 > 3{,}84$; so H_0 is rejected, meaning that the influence of the poverty gap index contributes significantly to the number of infant mortality cases in West Java.
2. Factors for low birth weight babies (X_2)
   The statistical value of the test $W_2 = 4{,}89 > 3{,}84$; so H_0 was rejected, meaning that the effect of low birth weight babies contributed significantly to the number of infant mortality cases in West Java.
3. Factors of exclusive breastfeeding in infants (X_3)
   The statistical value of the test $W_3 = 40{,}63 > 3{,}84$; so H_0 was rejected, meaning that the effect of exclusive breastfeeding on infants contributed significantly to the number of infant mortality cases in West Java.

3.6 The Final Model of GPR on the Number of Infant Mortality Cases in West Java for the 2019-2020 Period

Based on the results of the simultaneous test and partial test, it is found that all independent variables contribute to changes in the dependent variable, so that the final GPR model can be made as follows:

$$\hat{\lambda} = exp(0{,}20 + 0{,}23x_1 + 0{,}04x_2 - 0{,}01x_3). \quad (32)$$

The model in equation (32) can be interpreted as follows:

1. If the poverty gap index (X_1) increases by one unit, the number of infant mortality cases will increase by $exp(0{,}23)$ or 1,26 times.
2. If low birth weight babies (X_2) increase by one unit, then the number of infant mortality cases will increase by $exp(0{,}04)$ or 1,04 times.
3. If exclusive breastfeeding for infants (X_3) increases by one unit, then the number of infant mortality cases will decrease by $exp(-0{,}01)$ or 0,99 times.

After analyzing the data on the number of infant mortality cases in West Java for the 2019-2020 period, the results showed that infant mortality cases were influenced by the number of infant mortality cases such as the poverty gap index, the percentage of low birth weight babies (LBW), and the percentage of exclusive breastfeeding for infants. In further research, other models can be sought that can overcome similar cases and there are other problems such as outliers in the data, for example the Poisson's robust hurdle regression model [25-27].

4. Conclusion

The Poisson regression model of the data on the number of infant mortality cases in West Java in the 2019-2020 period violates the equidispersion assumption because the dispersion parameter value $\delta > 0$, then the GPR model is used to model the data. Estimation of GPR model parameters using the MLE method and continued with Newton-Raphson iteration. The resulting parameters were then tested simultaneously using the Likelihood Ratio Test (LRT) and partially using the Wald Test to see the significance of the parameters that contributed to the dependent variable. Factors that influence the number of infant mortality cases in West Java for the 2019-2020 period are the poverty gap index, the percentage of low birth weight infants (LBW), and the percentage of exclusive breastfeeding for infants.

References

[1] Rashwan, N. A., & Kamel, M. M. (2011). Using generalized Poisson log linear regression models in analyzing two-way contingency tables. *Applied Mathematical Sciences*, *5*(5), 213-222.

[2] Purhadi, Sutikno, Berliana, S. M., & Setiawan, D. I. (2021). Geographically weighted bivariate generalized Poisson regression: application to infant and maternal mortality data. *Letters in Spatial and Resource Sciences*, *14*, 79-99.

[3] Gao, G., Wang, H., & Wüthrich, M. V. (2022). Boosting Poisson regression models with telematics car driving data. *Machine Learning*, 1-30.
[4] Famoye, F., Wulu, J. T., & Singh, K. P. (2004). On the generalized Poisson regression model with an application to accident data. *Journal of Data Science*, *2*(3), 287-295.
[5] Lukman, A. F., Adewuyi, E., Månsson, K., & Kibria, B. M. (2021). A new estimator for the multicollinear Poisson regression model: simulation and application. *Scientific Reports*, *11*(1), 1-11.
[6] Amin, M., Akram, M. N., & Amanullah, M. (2022). On the James-Stein estimator for the Poisson regression model. *Communications in Statistics-Simulation and Computation*, *51*(10), 5596-5608.
[7] Motta, V. (2019). Estimating Poisson pseudo-maximum-likelihood rather than log-linear model of a log-transformed dependent variable. *RAUSP Management Journal*, *54*, 508-518.
[8] Ogallo, W., Wanyana, I., Tadesse, G. A., Wanjiru, C., Akinwande, V., Kabwama, S., ... & Walcott-Bryant, A. (2023). Quantifying the impact of COVID-19 on essential health services: a comparison of interrupted time series analysis using Prophet and Poisson regression models. *Journal of the American Medical Informatics Association*, *30*(4), 634-642.
[9] Zubedi, F., Aliu, M. A., Rahim, Y., & Oroh, F. A. (2021). Analisis Faktor-Faktor Yang Mempengaruhi Stunting Pada Balita Di Kota Gorontalo Menggunakan Regresi Binomial Negatif. *JAMBURA Journal of probability and statistics*, *2*(1), 48-55.
[10] Aulele, S. N., Lewaherilla, N., & Matdoan, M. Y. (2022). Pendekatan Geographically Weighted Poisson Regression Dengan Pembobot Fungsi Kernel Gauss Untuk Menganalisis Jumlah Kematian Bayi Di Provinsi Maluku. *Jurnal Aplikasi Statistika & Komputasi Statistik*, *14*(2), 67-80.
[11] Jao, N., Islamiyati, A., & Sunusi, N. (2022). Pemodelan Regresi Nonparametrik Spline Poisson pada Tingkat Kematian Bayi di Sulawesi Selatan. *Estimasi: Journal of Statistics and Its Application*, 14-22.
[12] Majore, M., Salaki, D. T., & Prang, J. D. (2020). Penerapan Regresi Binomial Negatif Dalam Mengatasi Overdispersi Regresi Poisson Pada Kasus Jumlah Kematian Ibu. *d'CARTESIAN: Jurnal Matematika dan Aplikasi*, 133-139.
[13] Aminullah, A. A. H., & Purhadi, P. (2020). Pemodelan untuk Jumlah Kasus Kematian Bayi dan Ibu di Jawa Timur Menggunakan Bivariate Generalized Poisson Regression. *Jurnal Sains dan Seni ITS*, *8*(2), D72-D78.
[14] Setyawan, Y., Suryowati, K., & Octaviana, D. (2022). Application of Negative Binomial Regression Analysis to Overcome the Overdispersion of Poisson Regression Model for Malnutrition Cases in Indonesia. *Parameter: Journal of Statistics*, *2*(2), 1-9.
[15] Wasilaine, T. L., Talakua, M. W., & Lesnussa, Y. A. (2014). Model Regresi Ridge Untuk Mengatasi Model Regresi Linier Berganda Yang Mengandung Multikolinieritas. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *8*(1), 31-37.
[16] Herawati, N., Nisa, K., Setiawan, E., Nusyirwan, N., & Tiryono, T. (2018). Regularized multiple regression methods to deal with severe multicollinearity. *International Journal of Statistics and Applications*, *8*(4), 167-172.
[17] Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Science & Business Media.
[18] Sundari, I. (2012). Regresi poisson dan penerapannya untuk memodelkan hubungan usia dan perilaku merokok terhadap jumlah kematian penderita penyakit kanker paru-paru. *Jurnal Matematika UNAND*, *1*(1), 71-76.
[19] Bauer, T., Göhlmann, S., & Sinning, M. (2007). Gender differences in smoking behavior. *Health Economics*, *16*(9), 895-909.

[20] Cahyandari, R. (2014). Pengujian Overdispersi pada Model Regresi Poisson (Studi Kasus: Laka Lantas Mobil Penumpang di Provinsi Jawa Barat). *Statistika*, *14*(2), 69-76.
[21] Saputro, D. R. S., Susanti, A., & Pratiwi, N. B. I. (2021). The handling of overdispersion on Poisson regression model with the generalized Poisson regression model. In *AIP Conference Proceedings* (Vol. 2326, No. 1, p. 020026). AIP Publishing LLC.
[22] Ruliana, R., Hendikawati, P., & Agoestanto, A. (2016). Pemodelan Generalized Poisson Regression (GPR) untuk Mengatasi Pelanggaran Equidispersi pada Regresi Poisson Kasus Campak di Kota Semarang Tahun 2013. *Unnes Journal of Mathematics*, *5*(1), 39-46.
[23] Sandjadirja, L. M., Aidi, M. N., & Rizki, A. (2019). Penanganan Overdispersi pada Regresi Poisson dengan Regresi Binomial Negatif pada Kasus Kemiskinan di Indonesia. *Xplore: Journal of Statistics*, *8*(1).
[24] Schober, P., & Vetter, T. R. (2021). Count data in medical research: Poisson regression and negative binomial regression. *Anesthesia & Analgesia*, *132*(5), 1378-1379.
[25] Cantoni, E., & Zedini, A. (2011). A robust version of the hurdlemodel. *Journal of Statistical Planning and Inference*, *141*(3), 1214-1223.
[26] Feng, C. X. (2021). A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of statistical distributions and applications*, *8*(1), 8.
[27] D'Este, M., Ganga, A., Elia, M., Lovreglio, R., Giannico, V., Spano, G., ... & Sanesi, G. (2020). Modeling fire ignition probability and frequency using Hurdle models: A cross-regional study in Southern Europe. *Ecological Processes*, *9*, 1-14.