BAREKENG: Journal of Mathematics and Its Application
December 2022 Volume 16 Issue 4 Page 1197–1206
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss4pp1197-1206

INVERSE GAUSSIAN REGRESSION MODELING AND ITS APPLICATION IN NEONATAL MORTALITY CASES IN INDONESIA

M. Fathurahman*

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University
Jl. Barong Tongkok No. 4 Kampus Gunung Kelua, Samarinda, 75123, Indonesia

*Corresponding author's e-mail: * fathur@fmipa.unmul.ac.id*

Abstract. *Inverse Gaussian Regression (IGR) is a suitable model for modeling positively skewed response data, which follows the inverse Gaussian distribution. The IGR model was formed from the Generalized Linear Models (GLM). This study aims to model the IGR with applied to model the factors influencing the infant mortality cases of provinces in Indonesia. Estimation of the IGR model parameters was employed by the Maximum Likelihood Estimation (MLE) and Fisher scoring methods. The Likelihood Ratio Test (LRT) and Wald test were used for hypothesis testing of significance parameters. The IGR model was applied to the infant mortality cases of provinces in Indonesia in 2020. The data for modeling infant mortality cases using IGR were obtained from the Ministry of Health of the Republic of Indonesia and the Central Bureau of Statistics. The result shows that the factors influencing the infant mortality cases of provinces in Indonesia based on the IGR model were: the percentage of pregnant women who received blood-boosting tablets, the percentage of low birth weight, the percentage of complete neonatal visits (KN3), the percentage of toddlers who received early initiation of breastfeeding, the percentage of toddlers who are exclusively breastfeeding, the percentage of toddlers who received complete primary immunization, the percentage of households with access to adequate drinking water, and the percentage of households with access to appropriate sanitation.*

Keywords: *Fisher scoring, IGR, GLM, LRT, MLE, neonatal mortality, positively skewed data, Wald test.*

Article info:

Submitted: 2nd June 2022 *Accepted: 12th October 2022*

How to cite this article:
M. Fathurahman, "INVERSE GAUSSIAN REGRESSION MODELING AND ITS APPLICATION IN NEONATAL MORTALITY CASES IN INDONESIA", *BAREKENG: J. Math. & App.*, vol. 16, iss. 4, pp. 1197-1206, Dec., 2022.

1. INTRODUCTION

Linear Regression (LR) is commonly used in regression modeling. The LR model applies only if the dependent variable follows the normal distribution. In the other condition, this assumption is not met when the dependent variable data is positively skewed [**1**]. The Inverse Gaussian Regression (IGR) model is applicable when the dependent variable has an inverse Gaussian (IG) distribution. The IGR model is similar to the gamma regression model except for more considerable skewness and high pitch [**2**].

Recently, research on IGR is still minimal and continues to be developed. A standardized version of the empirical moment generating function's logarithm to construct plots for assessing the appropriateness of the IG distribution was used [**3**]. [**4**] proposed a unified procedure for simultaneously parameter estimation and variable selection for joint mean and dispersion models of the IG distribution. Hypothesis testing of fit for IG distributions was developed [**5**], [**6**] and applied in R software [**7**]. [**8**] designed Generalized Linear Models (GLM)-based control charts when the dependent variable follows the inverse Gaussian distribution. Handling the multicollinearity cases in the IGR model was employed by [**9**], [**10**], [**11**], and [**12**].

The IGR in this research applies to the public health data, namely modeling neonatal mortality cases in Indonesia in 2020. Infant mortality is still Indonesia's most significant health problem, where 63% of deaths are caused during the neonatal period. The trend of neonatal mortality rate in Indonesia from 2017 until 2018 has decreased from 15 to 13 deaths per 1,000 live births, but this is not following the sustainable development goals target [**13**], [**14**], [**15**]. Neonatal mortality is death in the first month of life (0–28 days of age). This first month is the most crucial period for the child's survival because the risk of neonatal mortality is up to 15 times greater in countries with the highest mortality compared to countries with low mortality [**16**].

Most of the previous research discussed handling multicollinearity problems in the IGR model using simulation and actual data. This research is limited to discussing the IGR model without multicollinearity problems and its application to actual data. The objectives of this study are: to obtain an IGR model using the Maximum Likelihood Estimation (MLE) and Fisher scoring methods; to test the significance of the parameters of the IGR model using the Likelihood Ratio Test (LRT) and Wald test methods; and to obtain the factors that significant effect on neonatal mortality cases in Indonesia, in 2020 based on the IGR model.

2. RESEARCH METHODS

2.1 Data and Research Variables

The data for research variables in this research is secondary data obtained from the Ministry of Health of the Republic of Indonesia **[17]**. The research variables are displayed in Table 1.

Table 1. Research variables

Symbol	Variable	Variable Type
Y	The neonatal mortality cases	Discrete
X_1	The percentage of antenatal care (K4)	Continue
X_2	The percentage of pregnant women who received blood-boosting tablets	Continue
X_3	The percentage of low birth weight	Continue
X_4	The percentage of complete neonatal visits (KN3)	Continue
X_5	The percentage of toddlers who received early initiation of breastfeeding	Continue
X_6	The percentage of toddlers who are exclusively breastfeeding	Continue
X_7	The percentage of toddlers who received complete primary immunization	Continue
X_8	The percentage of households with access to adequate drinking water	Continue
X_9	The percentage of households with access to appropriate sanitation	Continue
X_{10}	The percentage of districts/cities that implement the policy of the healthy living community movement (GERMAS)	Continue

***Data source**: Ministry of Health of the Republic of Indonesia*.

2.2 Inverse Gaussian Regression Model

Inverse Gaussian Regression (IGR) is a regression model developed from GLM [**18**]. The dependent variable (Y) of the IGR model is a random variable with an Inverse Gaussian (IG) distribution with the probability density function defined as follows [**9**]:

$$P(Y = y|\mu, \zeta) = f(y, \mu, \zeta) = \frac{1}{\sqrt{2\pi y^3 \zeta}} \exp\left[-\frac{1}{2y}\left(\frac{y-\mu}{\mu\sqrt{\zeta}}\right)\right], y > 0, \mu > 0, \zeta > 0 \quad (1)$$

where y is the value of random variable Y, μ is the average of Y, and ζ is the dispersion parameter.

If there are m independent variables, namely X_1, X_2, …, X_m, then the IGR model can be written as follows:

$$\mu = \exp(\boldsymbol{x}_i^T \boldsymbol{\xi}), i = 1,2, \dots, n \quad (2)$$

where $\boldsymbol{x}_i^T = [1 \quad X_{1i} \quad X_{2i} \quad \cdots \quad X_{mi}]$ is a vector of independent variables for the i-th observation, and $\boldsymbol{\xi} = [\xi_0 \quad \xi_1 \quad \xi_2 \quad \cdots \quad \xi_m]^T$ is a parameter vector.

2.3 Estimation of the IGR Model Parameters

The estimated parameters of the IGR model in Equation (2) can be obtained using the MLE method [**2**], [**9**]. The parameter estimation of the IGR model using the MLE method begins with determining the likelihood function and the log-likelihood function as follows:

$$\mathcal{L}(\boldsymbol{\xi}) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi y_i^3 \zeta}} \exp\left[-\frac{1}{2y_i}\left(\frac{y_i - \exp(\boldsymbol{x}_i^T \boldsymbol{\xi})}{\exp(\boldsymbol{x}_i^T \boldsymbol{\xi})\sqrt{\zeta}}\right)^2\right] \right\} \quad (3)$$

$$\ell(\boldsymbol{\xi}) = \log \mathcal{L}(\boldsymbol{\xi}) = \prod_{i=1}^{n} \frac{1}{2} \left\{ \log(2\pi y_i^3 \zeta) \frac{1}{\zeta} \exp\left[-\frac{y_i}{(\exp(\boldsymbol{x}_i^T \boldsymbol{\xi}))^2} - \frac{2}{\exp(\boldsymbol{x}_i^T \boldsymbol{\xi})} + \frac{1}{y_i}\right] \right\} \quad (4)$$

It furthermore, maximizes the log-likelihood function by determining the first partial derivative of the log-likelihood function with respect to the estimated parameter and then equating it with zero,

$$\frac{\partial \ell(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \sum_{i=1}^{n} \left\{ -\frac{1}{\zeta}\left[\frac{\boldsymbol{x}_i^T}{\exp(\boldsymbol{x}_i^T \boldsymbol{\xi})}\right]\left[1 + \frac{y_i}{(\exp(\boldsymbol{x}_i^T \boldsymbol{\xi}))^2}\right] \right\} = \mathbf{0}. \quad (5)$$

Based on Equation (5), the result of the first partial derivative of the log-likelihood function with respect to the estimated parameters produces a not closed-form function. Therefore, a numerical approach is needed to obtain the ML estimator of the IGR model parameters. One numerical approach is the Fisher scoring method [**19**]. This method requires the second partial derivative of the log-likelihood function with respect to the estimated parameters as follows:

$$\frac{\partial^2 \ell(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T} = \sum_{i=1}^{n} \left\{ -\frac{1}{\zeta}\left[\frac{\boldsymbol{x}_i \boldsymbol{x}_i^T}{\exp(\boldsymbol{x}_i^T \boldsymbol{\xi})}\right]\left[1 + \frac{y_i}{(\exp(\boldsymbol{x}_i^T \boldsymbol{\xi}))^2} - \frac{2}{(\exp(\boldsymbol{x}_i^T \boldsymbol{\xi}))^3}\right] \right\}. \quad (6)$$

The formula used to obtain the ML parameter estimator of the IGR model using the Fisher scoring method is [**19**]:

$$\hat{\boldsymbol{\xi}}^{(q+1)} = \hat{\boldsymbol{\xi}}^{(q)} + \left[\boldsymbol{I}(\hat{\boldsymbol{\xi}}^{(q)})\right]^{-1} \boldsymbol{g}(\hat{\boldsymbol{\xi}}^{(q)}), q = 1,2, \dots \quad (7)$$

where $\hat{\boldsymbol{\xi}}$ is the ML estimator of the IGR model parameters, namely $\hat{\boldsymbol{\xi}} = [\hat{\xi}_0 \quad \hat{\xi}_1 \quad \hat{\xi}_2 \quad \cdots \quad \hat{\xi}_m]^T$. $\left[\boldsymbol{I}(\hat{\boldsymbol{\xi}}^{(q)})\right]^{-1}$ is the inverse of the Fisher information matrix defined by

$$\left[\boldsymbol{I}(\hat{\boldsymbol{\xi}}^{(q)})\right]^{-1} = -E\left[\frac{\partial^2 \ell(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T}\right] \quad (8)$$

where $E\left[\frac{\partial^2 \ell(\xi)}{\partial \xi \partial \xi^T}\right]$ is the expected value of the second partial derivative of the log likelihood function of the estimated parameter and $\frac{\partial^2 \ell(\xi)}{\partial \xi \partial \xi^T}$ as in Equation (6). $\boldsymbol{g}(\hat{\boldsymbol{\xi}})$ is a gradient vector whose elements are the first partial derivative of the likelihood function with respect to the estimated parameter, $\boldsymbol{g}(\hat{\boldsymbol{\xi}}) = \left[\frac{\partial \ell(\xi)}{\partial \xi}\right]^T$ with $\frac{\partial \ell(\xi)}{\partial \xi}$, as in Equation (5). The Fisher scoring iteration process to get the ML estimator of the IGR model parameters in Equation (7) stops if the convergence condition is met, namely $\left\|\hat{\boldsymbol{\xi}}^{(q+1)} - \hat{\boldsymbol{\xi}}^{(q)}\right\| \leq \delta$, where δ is a number very small positive. The ML estimator of the IGR model parameters obtained is $\hat{\boldsymbol{\xi}}^{(q+1)}$ at the last iteration. Meanwhile, the estimate for the variance-covariance matrix of parameter is $\boldsymbol{\xi}$, namely $Cov(\hat{\boldsymbol{\xi}}) = \left[\boldsymbol{I}(\hat{\boldsymbol{\xi}})\right]^{-1}$ [**19**].

2.4 Hypothesis Testing of the IGR Model

After assessing the parameters of the IGR model, it is followed by testing the parameter hypotheses. This test includes a simultaneous test and a partial test. The hypothesis of the simultaneous test is formulated as follows:

$$\begin{aligned} &H_0: \xi_1 = \xi_2 = \cdots = \xi_m = 0 \\ &H_1: \text{at least one of } \xi_l \neq 0, l = 1,2,\ldots,m. \end{aligned} \tag{9}$$

The test statistics for the simultaneous test can be obtained by the LRT method [**20**]. The first step to getting test statistics using the LRT method is to determine the set of model parameters under the null hypothesis (H_0), denoted by $\omega = \{\xi_0\}$. The next step, form the likelihood function and the log-likelihood as follows:

$$\mathcal{L}(\omega) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi y_i^3 \zeta}} \exp\left[-\frac{1}{2y_i} \left(\frac{y_i - \exp(\xi_0)}{\exp(\xi_0)\sqrt{\zeta}} \right)^2 \right] \right\}. \tag{10}$$

$$\ell(\omega) = \log[\mathcal{L}(\omega)] = \sum_{i=1}^{n} \frac{1}{2} \left\{ \log\left(2\pi y_i^3 \zeta\right) + \frac{1}{\zeta} \left[\frac{y_i}{(\exp(\xi_0))^2} - \frac{2}{\exp(\xi_0)} + \frac{1}{y_i} \right] \right\}. \tag{11}$$

After obtaining the log-likelihood function, determine the maximum value of the log-likelihood function under the null hypothesis

$$L_0 = \max \ell(\omega) = \sum_{i=1}^{n} \frac{1}{2} \left\{ \log\left(2\pi y_i^3 \hat{\zeta}\right) + \frac{1}{\hat{\zeta}} \left[\frac{y_i}{\left(\exp\left(\hat{\xi}_0\right)\right)^2} - \frac{2}{\exp\left(\hat{\xi}_0\right)} + \frac{1}{y_i} \right] \right\}. \tag{12}$$

where $\hat{\zeta}$ and $\hat{\xi}_0$ are obtained from the Fisher scoring iteration in Equation (7).

Furthermore, determining the set of model parameters under the population, denoted by $\Omega = \{\xi_0, \xi_1, \xi_2, \ldots, \xi_m\}$, and the likelihood function under the population is:

$$\mathcal{L}(\Omega) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi y_i^3 \zeta}} \exp\left[-\frac{1}{2y_i} \left(\frac{y_i - \exp\left(\boldsymbol{x}_i^T \boldsymbol{\xi}\right)}{\exp\left(\boldsymbol{x}_i^T \boldsymbol{\xi}\right)\sqrt{\zeta}} \right)^2 \right] \right\}. \tag{13}$$

Based on Equation (13), the log-likelihood function under the population is formed as follows:

$$\ell(\Omega) = \log[\mathcal{L}(\Omega)] = \sum_{i=1}^{n} \frac{1}{2} \left\{ \log\left(2\pi y_i^3 \zeta\right) + \frac{1}{\zeta} \left[\frac{y_i}{\left(\exp\left(\boldsymbol{x}_i^T \boldsymbol{\xi}\right)\right)^2} - \frac{2}{\exp\left(\boldsymbol{x}_i^T \boldsymbol{\xi}\right)} + \frac{1}{y_i} \right] \right\}. \tag{14}$$

Then determine the maximum value of the log-likelihood function under the population,

$$L_1 = \max \ell(\Omega) = \sum_{i=1}^{n} \frac{1}{2} \left\{ \log\left(2\pi y_i^3 \hat{\zeta}\right) + \frac{1}{\hat{\zeta}} \left[\frac{y_i}{\left(\exp\left(\boldsymbol{x}_i^T \hat{\boldsymbol{\xi}}\right)\right)^2} - \frac{2}{\exp\left(\boldsymbol{x}_i^T \hat{\boldsymbol{\xi}}\right)} + \frac{1}{y_i} \right] \right\}. \tag{15}$$

Following Equation (12) and Equation (15), the statistical test is defined as follows:

$$G = -2(L_0 - L_1). \tag{16}$$

The test statistic in Equation (16) follows the chi-square distribution [**20**]. Therefore, the null hypothesis in Equation (9) is rejected when the G statistic value is greater than the $\chi^2_{(\alpha;v)}$ value (i.e., $G > \chi^2_{(\alpha;v)}$) or p-value is less than the significance level (α), where $v = m$ is the degrees of freedom.

In addition to the simultaneous test, a partial test was carried out using the following hypotheses:

$$H_0: \xi_l = 0$$
$$H_1: \xi_l \neq 0, l = 1,2, \dots, m. \tag{17}$$

The test statistic used to test the hypothesis in Equation (17) is the Wald test statistic which is formulated as follows [**20**]:

$$W = \frac{\hat{\xi}_l}{\widehat{SE}(\hat{\xi}_l)} \tag{18}$$

where $\hat{\xi}_l$ and $\widehat{SE}(\hat{\xi}_l)$ are estimators of ML parameters and standard error of estimators of ML parameters of the IGR model are obtained by the Fisher scoring iteration in Equation (7).

The test statistic in Equation (18) has a normal distribution [**20**]. Thus, the null hypothesis in Equation (18) is rejected when the $|W|$ statistic value greater than the $Z_{\alpha/2}$ value (i.e., $|W| > Z_{\alpha/2}$) or p-value is less than α.

2.5 Procedures of Data Analysis

The procedures of data analysis in this research are as follows:

1. Analyzing the descriptive statistics of research variables.
2. Fitting the distribution of the dependent variable.
3. Detecting the multicollinearity problem of independent variables.
4. Modeling the neonatal mortality cases using the IGR model.
5. Getting the factors that influence neonatal mortality cases.
6. Interpreting the IGR model of neonatal mortality cases.
7. Getting the conclusions.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistical Analysis of Research Variables

Modeling the neonatal mortality cases in Indonesia in 2020 using the IGR model begins with a descriptive statistical analysis of research variables. The results are presented in Table 1.

Table 1. Summary statistical values of research variables

Variable	Minimum	Maximum	Mean	Standard Deviation
Y	40	3,031	596	718
X_1	27.50	98.90	77.40	19.40
X_2	25.30	99.30	77.72	17.34
X_3	0.80	6.90	3.37	1.60
X_4	33.30	100	82.77	18.03
X_5	52.10	96.10	79.04	10.85
X_6	34.00	87.30	64.21	13.06
X_7	41.80	99.40	78.55	14.49
X_8	62.50	99.80	85.42	9.59
X_9	40.30	97.00	79.82	9.96
X_{10}	0.00	100.00	33.82	33.24

Table 1 shows that Indonesia's average neonatal mortality cases in 2020 were 596, with a standard deviation of 718. The highest and lowest, 3,031 and 40, were found in Central Java Province and North Sulawesi Province, respectively. The average antenatal care (K4) percentage was 77.4 percent, with a standard deviation of 19.4 percent. The highest and lowest, 98.9 percent and 27.5 percent, were found in DKI Jakarta Province and Papua Province, respectively. The average percentage of pregnant women who received blood-boosting tablets was 77.72 percent, with a standard deviation was 17.34 percent. The highest and lowest, 99.3 percent and 25.3 percent, were found in DKI Jakarta Province and Papua Province, respectively. The average percentage of low birth weight was 3.37 percent, with a standard deviation was 1.6 percent. The highest and lowest, 6.9 percent and 0.8 percent, were found in East Nusa Tenggara Province and Riau Province, respectively. The average number of complete neonatal visits (KN3) was 82.77 percent, with a standard deviation of 18.03 percent. The highest and lowest, 100 percent and 33.3 percent, were found in North Kalimantan Province and West Papua Province, respectively.

Meanwhile, the average percentage of toddlers who received early breastfeeding initiation was 79.04 percent, with a standard deviation of 10.85 percent. The highest and lowest, 96.1 percent and 52.1 percent, were found in DKI Jakarta Province and Maluku Province, respectively. Furthermore, the average percentage of breastfeeding toddlers was 64.21 percent, with a standard deviation of 13.06 percent. The highest and lowest, 87.3 percent and 34 percent, were found in West Nusa Tenggara Province and West Papua Province, respectively. The average percentage of toddlers receiving complete primary immunization was 78.55 percent, with a standard deviation of 14.49 percent. The highest and lowest, 99.4 percent and 41.8 percent, were found in Bali Province and Aceh Province, respectively.

Furthermore, the average percentage of households with access to adequate drinking water was 85.42 percent, with a standard deviation was 9.59 percent. The highest and lowest, 99.8 percent and 62.5 percent, were found in DKI Jakarta Province and Bengkulu Province, respectively. The average percentage of households with access to appropriate sanitation was 79.82 percent, with a standard deviation was 9.96 percent. The highest and lowest, 97 percent and 40.3 percent, were found in DI Yogyakarta Province and Papua Province, respectively. The average percentage of districts/cities implementing the healthy living community movement (GERMAS) policy was 33.82 percent, with a standard deviation of 33.24 percent. The highest and lowest, 100 percent and 0 percent, respectively. The highest was in West Java Province and South Kalimantan Province, whereas the lowest was in Papua, West Papua, Maluku, North Sulawesi, West Kalimantan, Riau, North Sumatera, and Banten Provinces.

3.2 Fitting the Distribution of Dependent Variable

This section discusses the detection and testing of data for the dependent variable with an IG distribution. This detection uses a density plot via the *ggplot2 package* [**21**] in R software, and the results are shown in Figure 1. The pattern of neonatal mortality cases data (Y) in Figure 1 is positively skewed. These results indicate that the data on neonatal mortality cases has an IG distribution.

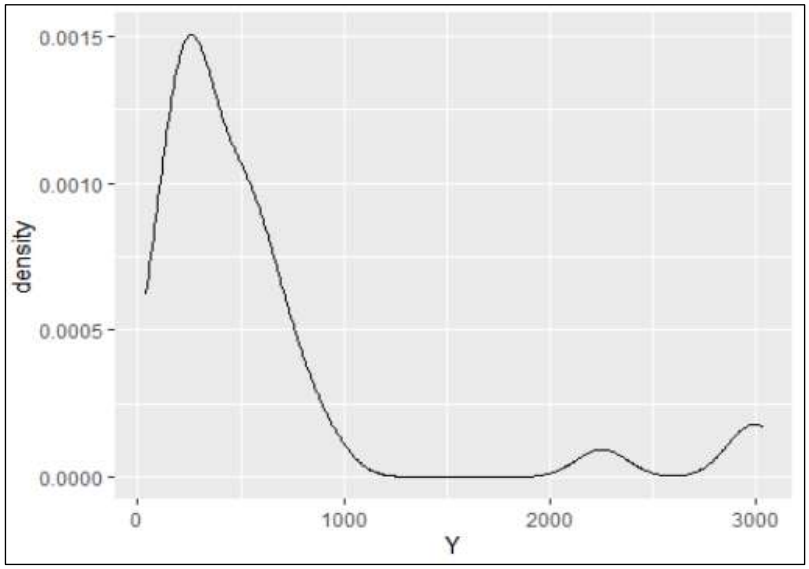

Figure 1. The density plot of the neonatal mortality cases in Indonesia, in 2020

Furthermore, hypothesis testing was to test the neonatal mortality cases (Y) following an IG distribution. Adopting [**5**], [**6**], and [**7**], the hypotheses used for this test are:

H_0: $Y_1 = Y_2 = \cdots = Y_n \sim IG(\mu, \zeta)$

(Y_i does follow an IG distribution)

H_1: Y_i does not follow an IG distribution, for $i = 1,2, \ldots, n$.

Based on the calculations using the *goft package* [**7**] in R software, the T_1 statistic value was 0.3534 less than the $Z_{\alpha/2}$ value of 1.6445, where the α value used in this research was 0.1. Meanwhile, the p-value was 0.7238 greater than α. Therefore, the null hypothesis is not rejected, concluding that the neonatal mortality cases follow an IG distribution. Thus, the data on neonatal mortality cases (Y) is feasible for the IGR model.

3.3 Detecting the Multicollinearity

This section discusses the detection of collinearity among the independent variables, namely multicollinearity. Multicollinearity detection uses the Variance Inflation Factor (VIF) [**22**]. The IGR model has a multicollinearity problem when the VIF value of independent variables is greater than 10. The VIF value of all independent variables is presented in Table 2.

Table 2. VIF value of independent variables

Variable	VIF Values
X_1	6.1214
X_2	6.7242
X_3	1.3500
X_4	6.4206
X_5	1.2866
X_6	1.9894
X_7	1.7120
X_8	1.5421
X_9	2.0869
X_{10}	1.4694

Table 2 shows that all covariates have a VIF value of less than 10. These results indicate that there is no multicollinearity. Therefore, all of independent variables are appropriate for the IGR model.

3.3 Modeling of Neonatal Mortality Cases Using IGR

The results of modeling the neonatal mortality cases in Indonesia in 2020 using the IGR model are shown in Table 3.

Table 3. Parameter estimates and the value of partial test statistic

Parameter	Estimation	Standard Error	W	p-value
ξ_0	3.6159	1.4601	2.4765	0.0211*
ξ_1	0.0096	0.0109	0.8807	0.3876
ξ_2	-0.0747	0.0121	-6.1736	2.6809×10^{-6}*
ξ_3	-0.0919	0.0507	-1.8126	0.0830*
ξ_4	0.0373	0.0086	4.3372	0.0002*
ξ_5	0.0327	0.0077	4.2468	0.0003*
ξ_6	0.0621	0.0106	5.8585	5.7005×10^{-6}*
ξ_7	0.0186	0.0070	2.6571	0.0141*
ξ_8	0.0479	0.0099	4.8384	6.9672×10^{-5}*
ξ_9	-0.0874	0.0140	-6.2429	2.2743×10^{-6}*
ξ_{10}	0.0004	0.0032	0.1250	0.9016

*) Significant at the level of significance, $\alpha = 0.1$.

After obtaining the ML estimator of the IGR model parameters in Table 3, simultaneous and partial tests were carried out on the IGR model parameters. The hypotheses for the simultaneous test are:

H_0: $\xi_1 = \xi_2 = \cdots = \xi_{10} = 0$

H_1: at least one of $\xi_l \neq 0, l = 1,2,\ldots,10$.

The statistical test (G) value was 30.7230 and the p-value was 0.0007. Furthermore, the $\chi^2_{(\alpha,v)}$ value was 15.9872. Since the G value is greater than the $\chi^2_{(\alpha,v)}$ value and the p-value is less than α, then the null hypothesis (H_0) is rejected. So, it can be concluded that the percentage of antenatal care (K4), the percentage of pregnant women who received blood-boosting tablets, the percentage of low birth weight, the percentage of complete neonatal visits (KN3), the percentage of toddlers who received early initiation of breastfeeding, the percentage of toddlers who are exclusively breastfeeding, the percentage of toddlers who received complete primary immunization, the percentage of households with access to adequate drinking water, and the percentage of households with access to appropriate sanitation, and the percentage of districts/cities that implement the policy of the healthy living community movement (GERMAS) have a simultaneously significant effect on neonatal mortality cases in Indonesia, in 2020.

To obtain an independent variable that has a partially significant effect on the dependent variable, a partial test was carried out. The hypotheses for the partial test are:

H_0: $\xi_l = 0$

H_1: $\xi_l \neq 0, l = 1,2,\ldots,10$.

Based on Table 3, there are two parameters, ξ_1 and ξ_{10}, which have a W statistic value less than the $Z_{\alpha/2}$ value and a p-value greater than α, so it fails to reject. Therefore, it can be concluded that the percentage of antenatal care (K4) and the percentage of districts and cities that implement the policy of the healthy living community movement (GERMAS) have not a partially significant effect on neonatal mortality cases in Indonesia in 2020. Meanwhile, the percentage of pregnant women who received blood-boosting tablets, the percentage of low birth weight, the percentage of complete neonatal visits (KN3), the percentage of toddlers who received early initiation of breastfeeding, the percentage of toddlers who are exclusively breastfeeding, the percentage of toddlers who received complete primary immunization, the percentage of households with access to adequate drinking water, and the percentage of households with access to appropriate sanitation have a partially significant effect on the neonatal mortality cases in Indonesia, in 2020.

Finally, the IGR model for modeling the neonatal mortality cases in Indonesia in 2020 based on Equation (2) and Table 3 can be written as follows:

$$\hat{\mu} = \exp(3.6159 + 0.0096X_1 - 0.0747X_2 - 0.0919X_3 + 0.0373X_4 + 0.0327X_5 + 0.0621X_6 + 0.0186X_7 + 0.0479X_8 - 0.0874X_9 + 0.0004X_{10}). \quad (19)$$

To interpret the IGR model in Equation (19), as an example, one of the independent variables that significantly affect the dependent variable is the percentage of pregnant women who received blood-boosting tablets (X_2). If the percentage of pregnant women who received blood-boosting tablets (X_2) increases by 1%, then the average of neonatal mortality cases $(E(Y))$ will decrease by exp(-0.0747) or 0.928 times, where the other independent variables are fixed.

4. CONCLUSIONS

IGR is an accurate regression technique for modeling positively skewed data based on the GLM framework. The MLE and Fisher scoring methods can obtain the IGR model. Meanwhile, the LRT and Wald test methods can be used to obtain simultaneous and partial statistical tests, respectively. The statistical test of the simultaneous test follows the Chi-square distribution, whereas the statistical test of the partial test follows the normal distribution. The IGR model in this research was applied to model the factors significantly affecting neonatal mortality cases in Indonesia in 2020. Based on the IGR model, the factors that significantly affect the neonatal mortality cases in Indonesia in 2020 were: the percentage of pregnant women who received blood-boosting tablets, the percentage of low birth weight, the percentage of complete neonatal visits (KN3),

the percentage of toddlers who received early initiation of breastfeeding, the percentage of toddlers who are exclusively breastfeeding, the percentage of toddlers who received complete primary immunization, the percentage of households with access to adequate drinking water, and the percentage of households with access to appropriate sanitation. However, this research still needs to be continued by adding other factors thought to influence neonatal mortality cases as independent variables. In addition, the different numerical approaches for obtaining the ML estimator of the IGR model parameters, such as the Newton-Raphson or quasi-Newton methods, are recommended for future research.

REFERENCES

[1] S. Kinat, M. Amin and T. Mahmood, "GLM-based control charts for the inverse Gaussian response variable," *Qual. Reliab. Eng. Int.*, vol. 36, no. 2, pp. 765-783, March 2020.

[2] M. N. Akram, M. Amin and M. Qasim, "A new Liu-type estimator for the inverse Gaussian regression model," *J. Stat. Comput. Simul.*, vol. 90, no. 7, pp. 1153-1172, January 2020.

[3] I. A. Koutrouvelis and A. Karagrigoriou, "Cumulant plots and goodness-of-fit tests for the inverse Gaussian distribution," *J. Stat. Comput. Simul.*, vol. 82, no. 3, pp. 343-358, 2012.

[4] L. Wu and H. Li, "Variable selection for joint mean and dispersion models of the inverse Gaussian distribution," *Metrika*, vol. 6, pp. 795-808, 2012.

[5] J. A. Villasenor and E. Gonzalez-Estrada, "Tests of fit inverse Gaussian distribution," *Stat. Probab. Lett.*, vol. 105, pp. 189-194, 2015.

[6] J. A. Villasenor, E. Gonzalez-Estrada and A. Ochoa, "On testing the inverse Gaussian hypothesis," *Sankhya: Indian J. Stat., doi: 10.1007/s13571-0170148-8*, pp. 1-14, 2017.

[7] E. Gonzalez-Estrada and J. A. Villasenor, "An R package for testing goodness of fit: goft," *J. Stat. Comput. Simul.*, vol. 88, no. 4, pp. 726-751, 2018.

[8] S. Kinat, M. Amin and T. Mahmood, "GLM-based control charts for the inverse Gaussian response variable," *Qual. Reliab. Eng. Int., doi: 10.1002/qre.2603*, pp. 1-19, 2019.

[9] Z. Y. Algamal, "Performance of ridge estimator in inverse Gaussian regression model," *Commun. Stat. Theor. Method.*, vol. 48, no. 15, pp. 3836-3849, 2019.

[10] R. E. Shamany, N. Z. Alobaidi and Z. Y. Algamal, "A new two-parameter estimator for the inverse Gaussian regression model with application in chemometrics," *Electron. J. Appl. Stat. Anal.*, vol. 12, no. 2, pp. 453-464, 2019.

[11] Y. M. Bulut and M. Isilar, "Two parameter Ridge estimator in the inverse Gaussian regression model," *Hacet. J. Math. Stat.*, vol. 50, no. 3, pp. 895-910, 2021.

[12] A. Lukman, Z. Y. Algamal, B. M. G. Kibria and K. Ayinde, "The KL estimator for the inverse Gaussian regression model," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 13, pp. 1-9, 2021.

[13] The Ministry of Health of the Republic of Indonesia, Demographic Survey and Basic Health Research (RISKESDAS), Jakarta: The Ministry of Health of the Republic of Indonesia, 2018.

[14] WHO, World Health Statistics 2020: Monitoring Health for the Sustainable Development Goals, Geneva: WHO, 2020.

[15] F. Rumiaty and A. C. Adisasmita, "Determinants of neonatal mortality based on the 2017 Indonesian Demographic and Health Surveys (IGHS)," *Indonesian J. Public. Health.*, vol. 16, no. 3, pp. 363-374, 2021.

[16] UNICEF, Levels & Trends in Child Mortality 2018, New York: UNICEF, 2018.

[17] The Ministry of Health of the Republic of Indonesia, The Indonesian Profile of Health 2020, Jakarta: The Ministry of Health of the Republic of Indonesia, 2021.

[18] A. Agresti, Foundations of Linear and Generalized Linear Models, New Jersey: John Wiley & Sons, 2015.

[19] M. Fathurahman, "Pemodelan indeks pembangunan kesehatan masyarakat kabupaten/kota di Pulau Kalimantan menggunakan pendekatan regresi probit [Modeling the public health development index of districts/cities in Kalimantan Island using probit regression approach]," *J. Varian*, vol. 2, no. 2, pp. 47-54, 2019.

[20] P. K. Dunn and G. K. Smyth, Generalized Linear Models with Examples in R, New York: Springer, 2018.

[21] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, 2nd ed., Switzerland: Springer, 2016.

[22] W. H. Greene, Econometric Analysis, 7th ed., New York: Pearson Education, 2012.