Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 17(1) 2024: 45-56
http://ejournal.undip.ac.id/index.php/media_statistika

CONWAY-MAXWELL POISSON REGRESSION MODELING OF INFANT MORTALITY IN SOUTH SULAWESI

Oktaviana[1], Wahidah Sanusi[2], Hisyam Ihsan[2], Aswi Aswi[2], Sukarna Sukarna[2], Serifat Adedamola Folorunso[3]

[1] Mathematics Study Program, Postgraduate Program, Makassar State University, Makassar, Indonesia
[2] FMIPA, Makassar State University, Makassar, Indonesia
[3] Department of Statistics, University of Ibadan, Nigeria

e-mail: *wahidah.sanusi@unm.ac.id*

DOI: 10.14710/medstat.17.1.45-56

Article Info:
Received: 3 September 2023
Accepted: 21 September 2024
Available Online: 14 October 2024

Keywords:
Poisson Regression; Overdispersion;Conway-Maxwell Poisson; Infant Mortality Rate
.

Abstract: Overdispersion is a common problem in count data that can lead to inaccurate parameter estimates in Poisson regression models. Quasi-Poisson and negative binomial regressions are often used to address overdispersion but have limitations, especially with small samples. The Conway-Maxwell Poisson (CMP) regression model, an extension of the Poisson distribution, effectively addresses both overdispersion and underdispersion, even with limited data, due to additional parameters that better control data dispersion. The Infant Mortality Rate (IMR) is a critical public health indicator, reflecting healthcare quality and broader social, economic, and environmental factors. Accurate IMR estimation is essential for evaluating health policies. This study aims to (1) identify overdispersion in IMR data from South Sulawesi, (2) model IMR using CMP regression, and (3) identify factors influencing IMR. The dataset includes IMR, Low Birth Weight (LBW), diarrhea, asphyxia, pneumonia, and exclusive breastfeeding. Analysis showed significant overdispersion with a ratio of 4.639, making CMP the optimal model with an AIC of 186.845. Significant factors identified were LBW, asphyxia, pneumonia, and exclusive breastfeeding. These findings advance statistical methodologies for count data analysis and offer a more accurate approach to evaluating public health policies, supporting efforts to reduce infant mortality in South Sulawesi Province.

1. INTRODUCTION

Poisson regression is a statistical method used to model discrete count data, such as the number of events occurring within a specific time or space interval, where the dependent variable follows a Poisson distribution (Amin et al., 2022; Aswi et al., 2022; Sundari & Sihombing, 2021; Winata, 2023). A key assumption of Poisson regression is equidispersion, meaning that the variance equals the mean (Ambarwati et al., 2020; Rahayu, 2021; Rahmadeni & Sari, 2018). However, this assumption is frequently violated in practice, particularly in public health data, leading to overdispersion (where the variance exceeds the mean) or underdispersion (where the variance is less than the mean) (Jao et al., 2022; Koerniawan et al., 2020; Ulfa et al., 2021). Overdispersion can arise due to factors such as outliers, model misspecification, correlation among individual responses, or population clustering (Afri, 2017). Applying Poisson regression to overdispersed data can result in underestimated standard errors, invalid confidence intervals, and erroneous statistical inferences (Kamalja & Wagh, 2021; Rahayu, 2021). These issues highlight the necessity for more robust methods to model discrete data with varying dispersion patterns. Consequently, this study emphasizes addressing the limitations of Poisson regression, particularly in the context of overdispersed data, to achieve more accurate parameter estimates and valid conclusions.

The central issue addressed in this study is overdispersion in the Infant Mortality Rate (IMR) data for South Sulawesi, where the rate remains above the national average. IMR is a critical health indicator reflecting the quality of maternal and child health services in a region (Jao et al., 2022). In Indonesia, the IMR stands at 30 per 1,000 live births, making it the fifth highest globally (Alfahmi, 2023). In 2020, the IMR in South Sulawesi was reported by BPS to be 17 per 1,000 live births, which is higher than the national average of 11 per 1,000 live births. This elevated rate underscores the need for more effective interventions to reduce IMR in the region. This study aims to analyze the factors influencing IMR in South Sulawesi using Conway-Maxwell Poisson (CMP) regression, a model that offers greater flexibility in handling overdispersion compared to traditional regression models. The application of CMP regression is anticipated to provide a more nuanced understanding of the factors contributing to the high IMR in South Sulawesi. The findings from this study are expected to support the development of more effective health policies to reduce IMR in the region.

Research on addressing overdispersion and underdispersion in Poisson regression has explored various models, such as Zero-Inflated Poisson (ZIP) regression (Putri et al., 2022) and Hardle Poisson (Aswi et al., 2022). CMP regression are recognized as effective approaches for handling overdispersion (Afri, 2017). Models such as ZIP, Quasi, CMP (Hayati et al., 2018), and Poisson-Tweedie (Nasution et al., 2022) demonstrate that CMP regression is particularly adept at managing overdispersion. Comparisons of CMP regression with Negative Binomial and Generalized Poisson models have shown CMP to be superior in handling overdispersion (Hayati et al., 2018). Further comparisons between COM-Poisson regression and Poisson regression have also highlighted the superiority of COM-Poisson in addressing overdispersion (Radam & Hameed, 2023).

The CMP regression model extends the Poisson regression framework to better handle both overdispersion and underdispersion in discrete data (Afri, 2017; Hayati et al., 2018). It incorporates two main parameters: the regression parameter (β) and the dispersion parameter (ν), which provides increased flexibility compared to other models like Generalized Poisson and Negative Binomial (Sellers & Premeaux, 2021). CMP regression facilitates more accurate modeling in scenarios where variance does not equal the mean, thus offering an advantage over traditional regression models. Previous studies have indicated that CMP regression yields lower deviance, Akaike Information Criteria (AIC), and Bayesian Information Criteria (BIC) values compared to Generalized Poisson and Negative Binomial models, making it a preferred choice for handling overdispersion (Afri, 2017; Hayati et al., 2018; Nasution et al., 2022). Given its interpretive advantages and flexibility across diverse discrete data conditions, CMP regression is highly recommended for analyzing factors affecting the dependent variable, especially in cases involving complex data variability (Afri, 2017; Hayati et al., 2018; Radam & Hameed, 2023).

This research is undertaken to address the limitations of Poisson regression in managing overdispersion within the IMR data for South Sulawesi, which exceeds the national average. Due to its limitations in handling overdispersion, Poisson regression may lead to biased parameter estimates and inaccurate conclusions (Kamalja & Wagh, 2021; Rahayu, 2021). Therefore, this study will employ the CMP regression model, which has been shown to be more adaptable and effective in addressing various forms of data dispersion, including overdispersion and underdispersion (Afri, 2017; Hayati et al., 2018; Nasution et al., 2022). By utilizing the CMP model, the study aims to achieve more accurate parameter estimates and provide a comprehensive understanding of the factors influencing IMR in South Sulawesi. This research is expected to contribute not only to advancements in statistical analysis methods but also to the formulation of targeted health policies aimed at reducing IMR through evidence-based strategies (Jao et al., 2022; Prahutama et al., 2017; Yasril et al., 2022).

2. LITERATURE REVIEW

2.1. Poisson Regression

Poisson regression is a generalised linear model (GLM) to model the functional relationship between the dependent variable, which is numeric data, and the independent variables, which can be numeric or categorical data (Amin et al., 2022). The Poisson regression model (Arum et al., 2022) is defined using a function of the dependent mean (μ_i) which can be written as follows.

$$log\, \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} = \boldsymbol{X}^T \boldsymbol{\beta} \quad (1)$$

The connecting function used in the Poisson regression model is the log connecting function, because the function can be ensured that all values are positive. The connecting function is a function used to connect μ_i with linear predictors. Estimation of Poisson regression model parameters (Tendriyawati et al., 2023) can be done using the Maximum Likelihood Estimation (MLE) method with the probability function (log-likelihood) used is as follows.

$$ln\, L(\boldsymbol{\beta}|y_i) = \sum[y_i \boldsymbol{x}_i^T \boldsymbol{\beta} - e^{\boldsymbol{x}_i^T \boldsymbol{\beta}} - ln\,(y!)], \qquad i = 1, 2, 3, \ldots \quad (2)$$

In Poisson regression modeling, the response variable (Y) is assumed to follow a Poisson distribution and must satisfy the characteristics of this distribution. The goodness-of-fit for Poisson regression is assessed using the Chi-Square test if the p-value is less than the significance level (α), typically set at 0.1, the data is considered to be Poisson distributed. Additionally, the Poisson regression model requires that the assumption of equidispersion be met, which stipulates that the variance of the dependent variable should equal its mean (Ambarwati et al., 2020).

$$E(Y) = Var(Y) = \lambda \quad (3)$$

2.2. Multicollinearity

The multicollinearity test is used to detect the degree of correlation between independent variables in the regression model (Fox, 2016). The occurrence of multicollinearity can be seen from the tolerance value. Another test that can be used to see multicollinearity is Variance Inflation Factor (VIF). The VIF formula is as follows.

$$VIF_k = \frac{1}{1-R_k^2}, \quad k = 1, 2, 3, \ldots \quad (4)$$

R_k is the Pearson correlation coefficient for the *k*-th independent variable, which measures the correlation between the *k*-th independent variable and the other independent variables in the model. $1 - R_k^2$ is the denominator in the VIF formula, indicating the proportion of variance in the *k*-th independent variable that is not explained by its relationship with the other independent variables. k denotes a specific variable among all considered independent variables.

Based on the formula presented in Equation (4), the VIF value can be determined if the correlation coefficient (R) is provided, as shown in Table 1. From the simulation results of the R and VIF values, it is evident that an independent variable is considered to have no multicollinearity if its VIF value is less than 3 (VIF < 3).

Table 1. Simulation of R and VIF Values

No	R	VIF	No	R	VIF
1	0.50	1.33	14	0.87	4.11
2	0.55	1.43	15	0.88	4.43
3	0.60	1.56	16	0.89	4.81
4	0.65	1.73	17	0.90	5.26
5	0.70	1.96	18	0.91	5.82
6	0.75	2.29	19	0.92	6.51
7	0.80	2.78	20	0.93	7.40
8	**0.81**	**2.91**	21	0.94	8.59
9	**0.82**	**3.05**	22	0.95	10.26
10	0.83	3.21	23	0.96	12.76
11	0.84	3.40	24	0.97	16.92
12	0.85	3.60	25	0.98	25.25
13	0.86	3.84	26	0.99	50.25

2.3. Overdispersion Test

Overdispersion refers to a condition where the variance of the dependent variable exceeds its mean, which can impair the accuracy of parameter estimates in Poisson regression by causing the standard errors of these estimates to be underestimated (Rahayu, 2020; Sellers & Premeaux, 2021). This condition can arise from several factors, including an excess of zero values in the dependent variable data (Dewanti et al., 2016; Rahayu, 2021), correlations among dependent observations, clustering within the population, and the omission of relevant variables, all of which can introduce bias into the estimation of the effect of independent variables (Afri, 2017). Overdispersion is typically detected when the dispersion parameter $0 \leq v < 1$ in the CMP distribution (λ, v) conforms to the specified equation (Sellers & Premeaux, 2021):

$$E(Y) \approx \lambda^{\frac{1}{v}} - \frac{v-1}{2v} \tag{5}$$

$$Var(Y) \approx \frac{1}{v}\lambda^{\frac{1}{v}} \tag{6}$$

Furthermore, overdispersion can be assessed through Pearson's chi-square and the dispersion ratio deviance, where a dispersion ratio greater than 1 indicates the presence of overdispersion.

2.4. Conway-Maxwell Poisson (CMP) Regression

The CMP regression model has the flexibility to model data with various levels of dispersion or in this case when there is overdispersion and underdispersion (Afri, 2017; Hayati et al., 2018; Nasution et al., 2022). If the dependent variable (Y) follows the CMP (λ, v) distribution then the probability mass function is as follows:

$$P(Y;\lambda,v) = \frac{\lambda^y}{(y!)^v Z(\lambda{:}v)}, \quad y = 0,1,2,\ldots \tag{7}$$

where λ is the averaging parameter, and ν is an dispersion parameter that controls the dispersion (or variability) in the distribution, y the count variable, which can take non-negative integer values (0, 1, 2, ...), $Z(\lambda: \nu)$ normalizing constant or partition function for the CMP distribution, ensuring the total probability sums to 1.

The variance equation in Equation 6 can address different dispersion conditions: it accommodates equidispersion when $(v = 1)$, overdispersion when $(0 \leq v < 1)$ and the underdispersion when $(v > 1)$ (Sellers & Premeaux, 2021). The normalization constants for the CMP distribution are as follows:

$$Z(\lambda: \nu) = \frac{exp\left(\nu \lambda^{\frac{1}{\nu}}\right)}{\lambda^{\frac{\nu-1}{2\nu}} (2\pi)^{\frac{\nu-1}{2\sqrt{\nu}}}} \tag{8}$$

For the CMP model, the commonly used link function is log-link, which is the same as the Poisson model. Using log-link, the CMP model can be written as:

$$log\, \lambda_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} = \boldsymbol{X}^T \boldsymbol{\beta} \tag{9}$$

where λ_i = The mean parameter of the CMP distribution for the *i*-th observation, predicted by the regression model; x_i = Predictor vector for the *i*-th observation; β = Vector of regression coefficients.

In the CMP regression model, parameter estimation can be done by the maximum likelihood method. The likelihood function of the CMP distribution function is as follows:

$$L(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{n} \frac{\exp\,(\boldsymbol{X_i \beta})^{y_i} \exp\,(\boldsymbol{X_i \beta})^{\frac{\nu-1}{2\nu}} (2\pi)^{\frac{\nu-1}{2\sqrt{\nu}}}}{(y_i!)^{\nu} exp\left(\nu (\boldsymbol{X_i \beta})^{\frac{1}{\nu}}\right)} \tag{10}$$

where $L(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X})$ = Likelihood function of the CMP distribution given parameters β, data y, and predictors *X*; $\boldsymbol{\beta}$ = Vector of regression coefficients; y_i = Observed count for the *i*-th observation; $\boldsymbol{X_i}$ = Vector of predictors for the i-th observation; $y_i!$ = Factorial of $y_i$raised to the power of v; 2π = Constant factor involving pi.

To assess the significance of parameters in the CMP model, a statistical test is conducted using p-values. This test evaluates whether the observed results significantly differ from the expected results under the null hypothesis. A parameter is deemed significant if the p-value is less than the chosen significance level, typically set at 0.1 (10%).

2.5. Previous research and novelty of the CMP Regression model

The CMP regression model extends the traditional Poisson regression framework to effectively address both overdispersion and underdispersion in discrete data. Studies have shown that CMP outperforms models such as the Generalized Poisson and Negative Binomial, providing lower values of deviance, AIC, and BIC, particularly in the presence of overdispersed data (Afri, 2017; Hayati et al., 2018; Radam & Hameed, 2023). Unlike Poisson regression, which assumes equidispersion, CMP introduces both a regression parameter (β) and a dispersion parameter (ν), offering greater flexibility in modeling count data.

The novelty of this research lies in the application of the CMP model to analyze infant mortality rate (IMR) data in South Sulawesi, a context in which it has not yet been applied. This study aims to fill this gap by evaluating the CMP model's effectiveness in handling overdispersion in IMR data from this region and identifying key factors influencing IMR. Ultimately, this research seeks to enhance the understanding of CMP in health data contexts and provide valuable insights for reducing IMR in South Sulawesi.

3. MATERIAL AND METHOD

3.1. Data Source

The data used in this study were the number of Infant Mortality Rate (IMR) in each district/city, the number of Low Birth Weight (LBW), the number of infants with exclusive breastfeeding, asphyxia, diarrhea and pneumonia in infants in 24 districts/cities of South Sulawesi Province in 2022. Data were obtained directly from the South Sulawesi Provincial Health Office.

3.2. Research Variables

The description of the research variables is described in Table 2.

Table 2. Description of the Research Variables

Variables	Operational Definitions
IMR (Y)	The number of IMR in each district/city in South Sulawesi.
LBW (X_1)	Number of babies born with low birth weight in each district/city in South Sulawesi.
Asphyxia (X_2)	Number of asphyxia babies per district/city in South Sulawesi.
Pneumonia (X_3)	Number of infants with acute lung infection, namely pneumonia, per district/city in South Sulawesi.
Diarrhea (X_4)	Number of infants with diarrhea per district/city in South Sulawesi.
Exclusive breastfeeding (X_5)	Number of exclusively breastfed infants per district/city in South Sulawesi.

3.3. Research Procedure

The CMP regression model was used to model IMR data in South Sulawesi with overdispersion conditions. The steps taken are as follows: (a) Exploration of IMR data through description and distribution test; (b) Building the IMR data model formulation using Poisson regression; (c) Identifying the assumption of equidispersion in the Poisson regression model; (d) If overdispersion occurs, select a model as a solution to handling overdispersion, where in this research the CMP regression model is used; (e) Perform multicollinearity test using VIF; (f) Estimating parameters with the MLE method to maximize the Likelihood function of CMP then modeling the data using the CMP regression model; (g) Assess parameter significance using p-values and select the best model based on the smallest AIC and parameter standard errors; (h) Interpretation of the results of the best CMP regression model.

4. RESULTS AND DISCUSSION

This study investigates secondary data on IMR, with IMR as the dependent variable (Y), and examines the impact of various predictor factors including (LBW), asphyxia, diarrhea, pneumonia, and exclusive breastfeeding. The analysis is conducted across 24 districts and cities in South Sulawesi Province for the year 2022.

Table 3. Statistics Descriptive

Variables	Minimum	Maximum	Mean	Variance	Standard Deviation
Y	8.00	167.00	46.54	1288.17	35.89
X_1	2.00	66.00	15.71	209.52	14.47
X_2	1.00	27.00	9.13	35.24	5.94
X_3	0.00	6.00	0.88	1.85	1.36
X_4	0.00	9.00	0.88	3.59	1.90
X_5	2.35	3.68	3.18	0.09	0.31

Table 3 presents the average IMR as 46.54 deaths per 1,000 live births, with the highest rate recorded at 167 in Makassar and the lowest at 8 in Bantaeng. The average value for variable X_1 (LBW) is 15.71, while variable X_2 (asphyxia) has an average of 9.13. The data for variables X_3 (pneumonia) and X_4 (diarrhea) exhibit identical mean and minimum values. Additionally, variable X_5 (exclusive breastfeeding) has an average of 3.18.

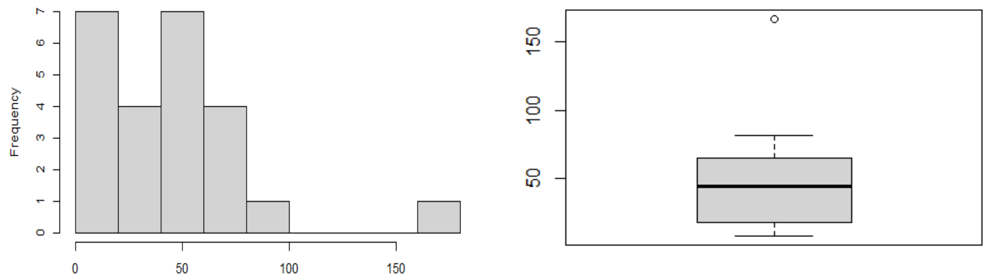

Figure 1. Histogram and Boxplot of IMR Data

The histogram and boxplot of the IMR data initially indicated deviations from a normal distribution and the presence of an outlier in Makassar city. Asymmetry, or skewness, signifies deviations from a normal distribution, which can impact the validity of statistical tests and the generalizability of results. Outliers, such as the one in Makassar, can disproportionately influence statistical measures, leading to skewed interpretations and potentially misleading conclusions. To address the influence of this outlier, a re-analysis was performed excluding data from Makassar city.

Table 4. Statistics Descriptive Data Without Makassar City

Variables	Minimum	Maximum	Mean	Variance	Standard Deviation
Y	8.00	82.00	41.30	658.49	25.66
X_1	2.00	30.00	13.52	99.07	9.95
X_2	1.00	27.00	8.78	33.90	5.82
X_3	0.00	6.00	0.82	1.87	1.37
X_4	0.00	9.00	0.91	3.71	1.92
X_5	2.35	3.68	3.18	0.09	0.31

The results of the analysis, excluding data from Makassar city, are presented in Table 4. These results reveal that variable X_5 (exclusive breastfeeding) retains similar characteristics, whereas variables X_1 (LBW), X_2 (asphyxia), X_3 (pneumonia), dan X_4 (diarrhea) exhibit differences. For the IMR data without Makassar city, the maximum rate is 82 deaths per 1,000 live births in Sinjai, and the minimum remains at 8 in Bantaeng. The histogram and boxplot in Figure 2 demonstrate that, in the absence of Makassar city data, the IMR data no longer displays outliers and aligns with a Poisson distribution. This adjustment ensures that the analysis reflects a more representative distribution of the data, thereby enhancing the reliability of the findings regarding the impact of predictor factors on IMR.

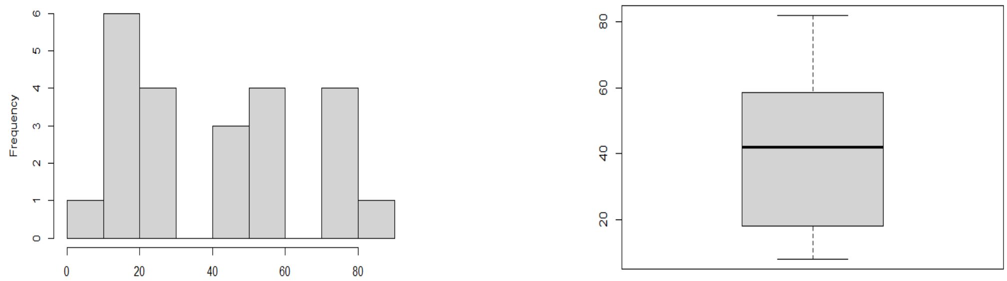

Figure 2. Histogram and Boxplot of IMR Data Without Makassar City

To build a Poisson regression model, previously the Poisson distribution test was carried out on dependent variable. Based on the results of descriptive analysis, the dependent variable follows the Poisson distribution. Data is said to be Poisson distributed if the p-value > α with a significance level (α) of 0.1. From the test results, it is found that the data is Poisson distributed. Then the parameter estimation of the Poisson regression model is carried out using the R studio software. The results of the Poisson regression model parameter estimation are shown in Table 5.

Table 5. Parameter Estimation of Poisson Regression Model

Variables	Parameter	Estimation	Std. Error	p-value
	Intercept (β_0)	1.314218	0.481427	0.006337
X_1	BBLR (β_1)	0.030768	0.005575	3.40e-08
X_2	Asphyxia (β_2)	0.028866	0.007058	4.32e-05
X_3	Pneumonia (β_3)	0.083957	0.024966	0.000772
X_4	Diarrhea (β_4)	0.018389	0.015081	0.222698
X_5	Exclusive breastfeeding (β_5)	0.469224	0.159022	0.003171

Based on the parameter estimation results presented in Table 5, it is evident that one parameter, specifically β_4, is insignificant, as indicated by a p-value greater than the significance level α. In contrast, the remaining parameters are statistically significant, each with p-value < α. The Poisson regression model can be expressed as follows:

$$log\, \mu = 1.314218 + 0.030768X_1 + 0.028866X_2 + 0.083957X_3 \\ +0.018389X_4 + 0.469224X_5$$

Subsequent analysis involved testing for overdispersion in the IMR data. Overdispersion testing can be conducted by comparing the variance to the mean of the dependent variable. Specifically, if the variance $Var(Y)$ of the dependent variable exceeds its mean $E(Y)$, overdispersion is present. As shown in Table 4, it can be seen that

$$Var(Y) = 658.49 > E(Y) = 41.30$$

This indicates that the IMR variable (Y) exhibits overdispersion. Additionally, overdispersion can be assessed using the Pearson chi-square test and the dispersion ratio obtained from the R Studio software analysis. A dispersion ratio > 1 signifies overdispersion. According to the results presented in Table 6, these findings confirm that the IMR variable (Y) is experiencing overdispersion.

Poisson regression is not suitable for use as a predictive model in this context due to its failure to meet the assumption of equidispersion. Consequently, the CMP regression model is employed for the subsequent analysis to model the IMR data in South Sulawesi.

Table 6. Overdispersion Test Results

Test	Estimation
Pearson's Chi-Squared	78.869
Dispersion ratio	4.639
p-value	<0.001

The CMP regression model extends the Poisson regression model by incorporating two parameters: the regression parameter (β) and the dispersion parameter (ν). This model is well-suited for handling IMR data exhibiting overdispersion. A key criterion for a robust regression model is the absence of multicollinearity among the independent variables (X). Multicollinearity was assessed for the factors influencing IMR events in South Sulawesi for the year 2022. Using the VIF test, as shown in Table 7, all independent variables (X) had VIF values < 3, indicating no significant multicollinearity among them. Therefore, the results of the multicollinearity assessment confirm that all independent variables are suitable for inclusion in the regression model.

Table 7. VIF Value Independent Variable

Variables	VIF
LBW (X_1)	2.89
Asphyxia (X_2)	2.10
Pneumonia (X_3)	1.09
Diarrhea (X_4)	1.32
Exclusive breastfeeding (X_5)	1.61

CMP regression modeling was conducted by including all five independent variables (X), given the absence of multicollinearity among them. The parameter estimates for the CMP regression model, aimed at analyzing the IMR data and its influencing factors, are presented in Table 8. Only those variables with a p-value < α=0.1, were included in the model, indicating their statistical significance and relevance to the analysis.

Tabel 8. Parameter Estimation of CMP Regression Model

Variabel	Parameter	Estimasi	Std. Error	p-value
	Intercept (β_0)	0.3447	0.2602	0.1851
X_1	BBLR (β_1)	0.0094	0.0043	0.0282
X_2	Asphyxia (β_2)	0.0089	0.0047	0.0583
X_3	Pneumonia (β_3)	0.0260	0.0158	0.0994
X_4	Diarrhea (β_4)	0.0056	0.0085	0.5114
X_5	Exclusive breastfeeding (β_5)	0.1522	0.0488	0.0018

Table 8 presents the results of the parameter estimation for the CMP regression model. It indicates that only the variable X_4, representing diarrhea, has a p-value greater than the significance level (α = 0.1), suggesting that it does not have a significant effect on the incidence of IMR in South Sulawesi. In contrast, the variables LBW (X_1), asphyxia (X_2), pneumonia (X_3), and exclusive breastfeeding (X_5) each have p-values < α, indicating that these variables significantly influence the incidence of IMR in South Sulawesi for the year 2022. The CMP regression model, with an AIC value of 186.845, demonstrating its suitability for modeling IMR data. CMP regression model that achieves the best fit, as indicated by the lowest AIC value from the parameter estimation, is as follows.

$$log\,\lambda = 0.3447\ + 0.0094\,X_1 + 0.0089\,X_2 + 0.0260X_3 + 0.0056\,X_4 + 0.1522X_5$$

Based on the CMP regression model, each additional infant with LBW (*X_1)* increases the incidence of IMR by exp (0.0094)=1.0094 times. Similarly, the incidence of IMR rises by exp(0.0089)=1.0089 times for each additional infant with asphyxia (X_2). For infants with pneumonia (X_3), the incidence of IMR increases by exp(0.0260)=1.0263 times.

Exclusive breastfeeding (X_5) is associated with a exp (0.1522)=1.1644 times higher incidence of IMR, indicating a significant association between exclusive breastfeeding and increased infant mortality. The association of exclusive breastfeeding with a higher incidence of IMR might seem surprising, as many studies emphasize the health benefits of exclusive breastfeeding. However, the findings of this study are consistent with research by Putri et al (2022). Several factors may explain this counterintuitive outcome. In regions with poor overall health conditions, mothers who breastfeed exclusively may face challenges such as malnutrition, inadequate hygiene, or limited access to healthcare, all of which can heighten infant mortality risk, irrespective of breastfeeding practices. Furthermore, infants with pre-existing health issues, such as low birth weight, may have a higher risk of mortality, regardless of whether they are exclusively breastfed. Mothers may attempt to exclusively breastfeed these vulnerable infants, but their underlying health conditions could still result in increased mortality.

This study employs the CMP regression model to analyze IMR in South Sulawesi, effectively addressing the issue of overdispersion that traditional Poisson regression cannot resolve. The CMP model was selected for its superior capability in managing overdispersion compared to Poisson or negative binomial regression models. Prior research, including studies by Afri (2017), Fitri et al. (2021), and Radam & Hamed (2023), has validated the effectiveness of the CMP model in similar contexts. This study advances existing research by incorporating an analysis of additional factors such as LBW, asphyxia, pneumonia, and exclusive breastfeeding. It utilizes a more suitable CMP model and a broader set of variables, offering enhanced insights into the determinants of IMR and highlighting the critical roles of LBW and exclusive breastfeeding in mitigating IMR in South Sulawesi.

5. CONCLUSION

This study employed the CMP regression model to analyze the IMR in South Sulawesi, effectively addressing the issue of overdispersion that traditional Poisson regression cannot handle. The CMP model identified LBW, asphyxia, pneumonia, and exclusive breastfeeding as significant factors influencing IMR, while diarrhea was found to be non-significant. These results corroborate previous research and underscore the necessity of focusing on these key factors to mitigate IMR. Future research should continue to utilize the CMP model and investigate additional variables such as access to healthcare services and socio-economic factors.

ACKNOWLEDGMENT

The authors wish to acknowledge the Kemenristekdikti, Indonesia. This work was funded by Kemenristekdikti, contract number 2804/UN36.11/LP2M/2023.

REFERENCES

Afri, L. E. (2017). Perbandingan Regresi Binomial Negatif dan Regresi Conway-Maxwell-Poisson dalam Mengatasi Overdispersi pada Regresi Poisson. *Jurnal Gantang*, *2*(1), 79–87. https://doi.org/10.31629/jg.v2i1.66

Alfahmi, F. (2023). Hubungan Pola Makan dan Asupan Protein Ibu Hamil dengan Kejadian BBLR di Puskesmas Kadugede. *Jurnal Ilmu Kesehatan*, *2*(2), 13–26.

Ambarwati, P. C., Indahwati, & Aidi, M. N. (2020). Kajian Simulasi Overdispersi pada Regresi Poisson dan Binomial Negatif Terboboti Geografis untuk Data Balita Gizi Buruk. *Journal of Statistics and Its Applications*, *4*(3), 484–497. https://doi.org/https://doi.org/10.29244/ijsa.v4i3.684

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. https://doi.org/10.1080/03610918.2020.1775851

Arum, K. C., Ugwuowo, F. I., & Oranye, H. E. (2022). Robust modified jackknife ridge estimator for the Poisson regression model with multicollinearity and outliers. *Scientific African*, *17*, e01386. https://doi.org/10.1016/j.sciaf.2022.e01386

Aswi, A., Astuti, S. A., & Sudarmin, S. (2022). Evaluating the Performance of Zero-Inflated and Hurdle Poisson Models for Modeling Overdispersion in Count Data. *Inferensi*, *5*(1), 17. https://doi.org/10.12962/j27213862.v5i1.12422

Dewanti, N. P. P., Susilawati, M., & Srinadi, I. G. A. M. (2016). Perbandingan Regresi Zero Inflated Poisson (ZIP) dan Regresi Zero Inflated Negatif Binomial (ZINB) pada Data Overdispersi (Studi kasus: angka kematian ibu di Provinsi Bali). *E-Jurnal Matematika*, *5*(4), 133. https://doi.org/10.24843/mtk.2016.v05.i04.p132

Fitri, F., Sari, F. M., Gamayanti, N. F., & Utami, I. T. (2021). Infant Mortality Case: An Application of Negative Binomial Regression in order to Overcome Overdispersion in Poisson Regression. *Eksata: Berkala Ilmiah Bidang MIPA*, *22*(03), 200–210. https://doi.org/https://doi.org/10.24036//eksakta/vol22-iss2/272 Eksakta BerkalaiIlmiah Bidang MIPA http://www.eksakta.ppj.unp.ac.id/index.php/eksakta Article

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models* (3rd ed.). SAGE Publications Ltd.

Hayati, M., Sadik, K., & Kurnia, A. (2018). Conwey-Maxwell Poisson Distribution: Approach for Over- and-Under-Dispersed Count Data Modelling. *IOP Conference Series: Earth and Environmental Science*, *187*(1). https://doi.org/10.1088/1755-1315/187/1/012039

Jao, N., Islamiyati, A., & Sunusi, N. (2022). Pemodelan Regresi Nonparametrik Spline Poisson Pada Tingkat Kematian Bayi di Sulawesi Selatan. *Estimasi*, *3*(1), 14–23. https://doi.org/10.20956/ejsa.vi.11997

Kamalja, K. K., & Wagh, Y. S. (2021). Estimation in zero-inflated Generalized Poisson distribution. *Journal of Data Science*. https://doi.org/10.6339/jds.201801_16(1).0010

Koerniawan, V., Sunusi, N., & Raupong, R. (2020). Estimasi Parameter Model Poisson Hidden Markov Pada Data Banyaknya Kedatangan Klaim Asuransi Jiwa. *Estimasi: Journal of Statistics and Its Application*, *1*(2), 65–73. https://doi.org/10.20956/ejsa.v1i2.9302

Nasution, A. R., Sadik, K., & Rizki, A. (2022). Perbandingan Kinerja Regresi Conway-Maxwell-Poisson dan Poisson-Tweedie dalam Mengatasi Overdispersi Melalui Data Simulasi. *Xplore: Journal of Statistics*, *11*(3), 215–225.

https://doi.org/https://doi.org/10.29244/xplore.v11i3.1018

Prahutama, A., Sudarno, Suparti, & Mukid, M. A. (2017). Analisis Faktor-Faktor Yang Mempengaruhi Angka Kematian Bayi Di Jawa Tengah Menggunakan Regresi Generelized Poisson Dan Binomial Negatif. *Statistika*, *5*(2), 1–6.

Putri, A. D., Devianto, D., & Yanuar, F. (2022). Pemodelan Jumlah Kematian Bayi di Kota Bandung dengan Menggunakan Regresi Zero-Inflated Poisson. *Jurnal Matematika UNAND*, *11*(1), 12–24. https://doi.org/https://doi.org/10.25077/jmu.11.1.12-24.2022

Radam, S. A., & Hameed, L. M. A. (2023). Comparison of Poisson Regression and Conway Maxwell Poisson Models Using Simulation. *Journal of Economics and Administrative Sciences (JEAS)*, *29*(136), 81–99. https://www.jeasiq.uobaghdad.edu.iq/index.php/JEASIQ/article/view/2609/2186

Rahayu, A. (2020). Model-model Regresi untuk Mengatasi Masalah Overdispersi pada Regresi Poisson. *Journal Peqguruan: Conference Series*, *1*(April), 1–5. https://doi.org/http://dx.doi.org/10.35329/jp.v2i1.1866

Rahayu, A. (2021). Model-Model Regresi untuk Mengatasi Masalah Overdipersi pada Regresi Poisson. *Journal Peqguruang: Conference Series*. https://doi.org/10.35329/jp.v2i1.1866

Rahmadeni, & Sari, N. (2018). Solusi Overdispersi Menggunakan Generalized Poisson Regression (Studi Kasus : Penderita HIV di Provinsi Riau). *Jurnal Sains Matematika Dan Statistika*, *4*(2), 28–36.

Sellers, K. F., & Premeaux, B. (2021). Conway–Maxwell–Poisson regression models for dispersed count data. *Wiley Interdisciplinary Reviews: Computational Statistics*, *13*(6), 1–13. https://doi.org/10.1002/wics.1533

Sundari, M., & Sihombing, P. R. (2021). Penanaganan Overdispersi Pada Regresi Poisson (Studi Kasus : Pengaruh Faktor Iklim Terhadap Jumlah Penderita Penyakit Demam Berdarah di Kota Bogor). *Lebesgue: Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *2*(1), 1–9. https://doi.org/10.46306/lb.v2i1

Tendriyawati, Wibawa, gusti ngurah adhi, & Abapihi, B. (2023). *Pemodelan Regresi Poisson Terhadap Faktor-faktor yang Mempengaruhi Terjadinya Hipertensi di Kota Kendari*. *3*(April), 255–262.

Ulfa, Y. A., Soleh, A. M., & Sartono, B. (2021). Handling of Overdispersion in the Poisson Regression Model with Negative Binomial for the Number of New Cases of Leprosy in Java. *Indonesian Journal of Statistics and Its Applications*, *5*(1), 1–13. https://doi.org/10.29244/ijsa.v5i1p1-13

Winata, H. M. (2023). Mengatasi Overdispersi dengan Regresi Binomial Negatif pada Angka Kematian Ibu di Kota Bandung. *Jurnal Gaussian*, *11*(4), 616–622. https://doi.org/10.14710/j.gauss.11.4.616-622

Yasril, A. I., Yuhelmi, & Safitri, Y. (2022). Penerapan Analisis Jalur (Path Analysis) pada Faktor yang Mempengaruhi Angka Kematian Bayi di Sumatera Barat. *Jurnal Endurance: Kajian Ilmiah Problema Kesehatan*, *6*(2), 236–249. https://doi.org/10.22216/jen.v6i2.189