BAREKENG: Journal of Mathematics and Its Applications
March 2023 Volume 17 Issue 1 Page 0417–0426
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0417-0426

OVERDISPERSION HANDLING IN POISSON REGRESSION MODEL BY APPLYING NEGATIVE BINOMIAL REGRESSION

Yesan Tiara [1*], Muhammad Nur Aidi[2], Erfiani[3], Rika Rachmawati[4]

[1,2,3] Department of Statistics, IPB University
Meranti Wing 22 Level 4 Dramaga, Bogor, West Java, 16680, Indonesia

Corresponding author's e-mail: [1] yesantiara@gmail.com*

Article History:

Received: 9th November 2022
Revised: 28th January 2023
Accepted: 11th February 2023

Keywords:

Poisson regression;
Overdispersion;
Negative binomial regression;
Anemia;
Women of Childbearing Age (WCA)

ABSTRACT

Statistical analysis that can be used if the response variable is quantified data is Poisson regression, assuming that the assumption must be met equidispersion, where the average response variable is the same as the standard deviation value. A negative binomial regression can overcome an unfulfilled equidispersion assumption where the mean is greater than the standard deviation value (overdispersion). This method is more flexible because it does not require that the variance be equal to the mean. The case studies used in this research are cases of anemia in women of childbearing age (WCA) in 33 provinces of Indonesia. This study aims to apply the Poisson regression method and negative binomial in the case data of anemia in WCA to prove the model's goodness and find the factors that influence anemia in WCA. This data was obtained from biomedical sample data for Riset Kesehatan Dasar (Riskesdas) and data obtained from the website *of the Badan Pusat Statistik (BPS) in 2013. By applying these two methods, the result is that negative binomial regression is the best model in modeling WCA cases with anemia in Indonesia because it has the smallest AIC value of 221.72; however, the difference is not too far from the AIC in the Poisson regression model, which is 221.83. It can also be supported that Poisson regression is unsuitable for the analysis because of the case of overdispersion. With a significance level of 10%, the number of WCA affected by malaria per 100 population influences cases of WCA anemia. At the same time, other independent variables have no effect.*

How to cite this article:

Y. Tiara, M. N. Aidi, Erfiani and R. Rachmawati., "OVERDISPERSION HANDLING IN POISSON REGRESSION MODEL BY APPLYING NEGATIVE BINOMIAL REGRESSION," *BAREKENG: J. Math. & App.*, vol. 17, iss. 1, pp. 0417-0426, March 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

One example of a condition is when the response variable data is count data by following the Poisson distribution, then the appropriate statistical method is used to analyze the Poisson regression. However, Poisson regression has an assumption that must be met before being used; namely, the average of the response variables must be equal to the variance, which is known as equidispersion [1]. However, in reality, this is rare because the data for the count usually has a variance greater than the mean or it is called overdispersion [2]. Overdispersion can cause the deviation value of the model to be very large and can cause the resulting Poisson regression model to be less appropriate to use [3]. The Negative Binomial regression method can be used to overcome overdispersion in Poisson regression. The advantage of this model is that it is more flexible because the assumptions of the mean and variance do not have to be the same [4].

Previous studies have shown that negative binomial regression is an effective method to overcome overdispersion in Poisson regression. Research conducted by Nadhifan Humam Fitrial and Akhmad Fatikhurrizqi in 2020 regarding the number of COVID-19 cases in Indonesia stated that Negative Binomial regression is better to use than Poisson regression based on the smallest AIC value [5]. Research by Luhung Mustika Budiharti in 2021 shows that the results of the Poisson regression modeling indicate an overdispersion problem, so the negative binomial regression model is the best model with the variable percentage of the male population influencing the number of lepers in West Java in 2019 [6]. In this study, there were several differences from previous studies, where the selected case study is related to the causative factors that influence cases of anemia in Women of Childbearing Age (WCA). In addition, there are also advantages, namely using additional variables, which are the results of trial and error, to determine the optimum θ to overcome the excess variance in Poisson regression.

The poisson regression method and negative binomial regression will be applied to cases of anemia in WCA in Indonesia, which will be the topic of this study. Anemia is a condition where the hemoglobin level in the blood decreases below the normal value for individual measurements. Anemia is one of the nutritional problems in Indonesia and is most commonly found in women of childbearing age and children. The problem of anemia is still quite serious. It has been going on for a long time (chronic), but this has received less attention because it is considered normal and does not show significant symptoms as a disturbance in activities in daily life, even though anemia is the second highest cause of disability in the world and become a serious public health problem throughout the world because of its effect on the quality of human resources and reflects the value of the socio-economic welfare of the community which will have a harmful impact [7].

Based on the description above, research was conducted on Poisson regression modeling and negative binomial regression, which was applied in cases of anemia in WCA in Indonesia to see what factors influence anemia in WCA.

2. RESEARCH METHODS

The data used in this study is data obtained from the 2013 Riset Kesehatan Dasar (Riskesdas) sample data which has been collected from biomedical data and data obtained from *the website* of the Badan Pusat Statistik (BPS) in 2013. In this study, there were 33 provinces in Indonesia as the location of observation along with the factors thought to influence cases of anemia in women of childbearing age (WCA). The number of WCA samples for each province is listed in **Table 1.**

Table 1. Total WCA Samples in Each Province

Province	Total	Province	Total	Province	Total
Aceh	115	Jawa Barat	1623	Kalimantan Timur	158
Sumatra Utara	425	Jawa Tengah	1468	Sulawesi Utara	109
Sumatra Barat	197	DI Yogyakarta	218	Sulawesi Tengah	152
Riau	120	Jawa Timur	1311	Ssulawesi Selatan	301
Jambi	122	Banten	668	Sulawesi Tenggara	20
Sumatra Selatan	373	Bali	190	Gorontalo	41
Bengkulu	50	Nusa Tenggara Barat	397	Sulawesi Barat	53

Province	Total	Province	Total	Province	Total
Lampung	346	Nusa Tenggara Timur	171	Maluku	83
Kepulauan Bangka Belitung	19	Kalimantan Barat	156	Maluku Utara	47
Kepulauan Riau	79	Kalimantan Tengah	34	Papua Barat	68
DKI Jakarta	73	Kalimantan Selatan	157	Papua	119

The research variables used are as follows:

Table 2. Research Variables

Variables	Information
Y_1	Number of WCA affected by anemia per 100 WCA in each province in Indonesia
X_1	Number of WCA living in rural areas per 100 WCA
X_2	Number of WCA affected by pneumonia per 100 WCA
X_3	Number of WCA affected by ARI per 100 WCA
X_4	Number of WCA affected by malaria per 100 WCA
X_5	Number of WCA affected by TB per 100 WCA
X_6	Number of WCA affected by hepatitis per 100 WCA
X_7	Percentage of districts/cities that have PHBS policies

The methods used in this study are Poisson regression and binomial regression negative to determine the factors that influence cases of anemia in WCA in Indonesia. This study uses software R version 4.1.3 to perform data analysis.

The stages carried out in this research are as follows:

1. Exploring data on the number of WCA affected by anemia per 100 WCA in each province.
2. Identifying multicollinearity by looking at the correlation criteria and VIF.
3. Perform Poisson regression analysis with the following steps:
   a. Estimating the Poisson regression model.
   b. Test the significance of the Poisson regression model parameters simultaneously and partially.
2. Identify overdispersion in Poisson regression.
3. Perform negative binomial regression analysis with the following steps:
   a. Estimating the negative binomial regression model.
   b. Test the significance of the negative binomial regression model parameters simultaneously and partially.
   c. Comparing the AIC value between the Poisson regression model and the negative binomial regression to select the best model used and the best model interpretation.

2.1 Multicollinearity

Multicollinearity is a condition with a strong correlation or relationship between the independent variables (Xi). The independent variables that are interrelated (collinear) are very similar, and it is difficult to separate their effects, so the result of this strong correlation is that it will produce an estimate of the regression coefficient that is biased, unstable, and inefficient. The presence of multicollinearity can be detected using the *Variance Inflation Factor* (VIF) value with the following formula [8]:

$$VIF(X_i) = \frac{1}{(1-R_i^2)} \quad (1)$$

where

$i \ = 1,2,3, \dots, n$

R_i^2 = i-th coefficient of determination (square of the correlation coefficient)

High multicollinearity can be indicated by the high value of VIF obtained. If the VIF value is not more than 10, the conclusion is that there are no cases of multicollinearity [9].

2.2 Poisson Regression

Poisson regression is a nonlinear regression model used for the dependent variable that follows the Poisson distribution and is a count data [10]. The distribution function is as follows:

$$\boldsymbol{f(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!} \qquad , y = 0, 1, 2, \ldots} \tag{2}$$

where $\boldsymbol{\mu}$ is the average random variable Y with a Poisson distribution with the average value and variance of the random variable Y more than zero. The equations for the Poisson regression model are as follows:

$$\boldsymbol{\mu_i = \exp\,(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}} \tag{3}$$

$\boldsymbol{\mu_i}$ is a Poisson regression model with $\boldsymbol{x_{pi}}$ as an independent variable and $\boldsymbol{\beta_p}$ as a parameter of the regression coefficient to be estimated.

The Maximum Likelihood Estimation (MLE) method is used to estimate the parameters of the Poisson regression model. The Function likelihood of the Poisson regression model is formulated as follows:

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

Testing the poisson regression model parameters is used to determine whether the independent variable has a significant effect on the dependent variable. Simultaneous significance test using Maximum Likelihood Ratio Test (MLRT) with the following hypothesis [1]:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$
$H_1: there\ is\ at\ least\ one\ \beta_k \neq 0, i = 1, 2, \ldots., n; k = 0,1,2, \ldots, p$
Test statistics:

$$D(\hat{\beta}) = -2ln\left(\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right) \tag{5}$$

The rejection area is rejected H_0 if $D(\hat{\beta}) > \chi^2_{p(a)}$ or p-value $< \alpha$, which means that at least one parameter in the Poisson regression has a significant effect on the response variable.

Furthermore, partial testing is used to determine which independent variables have an effect on the dependent variable, with the following hypothesis:

$H_0: \beta_k = 0$ (The effect of the *k-th* is not significant)
$H_1: \beta_k \neq 0, k = 1, 2, \ldots., p$ (The effect of the *k-th* is significant)
Test statistics:

$$Z_{hit} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \tag{7}$$

The rejection area is rejected H_0 if $|Z_{hit}| > z_{\frac{\alpha}{2}}$ or p-value $< \alpha$, which means the k-th independent variable affects the dependent variable.

2.3 Overdispersion

Overdispersion in Poisson regression occurs if the variance value is greater than the average value. As a result, if you continue to analyze data containing overdispersion- ties, the conclusions obtained will be invalid because the standard error value is underestimated. Overdispersion can be written as follows [11]:

$$Var(Y) > E(Y)$$

The deviance dispersion value Pearson Chi-Square in the Poisson regression model divided by the degree of freedom is a way to detect cases of overdispersion. If the comparison result is more than 1, then there is an overdispersion case in Poisson regression.

2.4 Negative Binomial Regression

Negative binomial regression is a mixed model of Poisson and Gamma, which is one solution for overcoming the problem of overdispersion in Poisson regression [12]. The negative binomial probability function can be stated as follows:

$$f(y,\mu,\theta) = \frac{\Gamma\left(y+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y!}\left(\frac{1}{1+\theta\mu}\right)^{1/\theta}\left(\frac{\theta\mu}{1+\theta\mu}\right)^{y} \quad ,y = 0,1,2,\ldots \tag{8}$$

The Negative Binomial regression model equation can be written as follows:

$$y_i = \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}\right) \tag{9}$$

The Maximum Likelihood Estimation (MLE) method with Newton Raphson iteration is also used in estimating negative binomial regression model parameters. The function likelihood of the Poisson regression model is formulated as follows:

$$L(\beta,\theta) = \prod_{i=1}^{n}\left(\prod_{r=0}^{y_i-1}(r+\theta^{-1})\right)\frac{1}{(y_i!)}\left(\frac{1}{1+\theta\mu_i}\right)^{1/\theta}\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i} \tag{10}$$

Simultaneously test the significance of the negative binomial regression model using the Maximum Likelihood Ratio Test (MLRT) with the following hypothesis [1]:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$
$H_1: there\ is\ at\ least\ one\ \beta_k \neq 0, i = 1,2,\ldots.,n; k = 0,1,2,\ldots,p$

Test statistics:

$$D(\hat{\beta}) = -2ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) \tag{11}$$

The rejection area is rejected H_0 if $D(\hat{\beta}) > \chi^2_{p(a)}$ or p-value $< \alpha$.

Furthermore, the partial test of the Negative Binomial regression model with the following hypothesis:

$H_0: \beta_k = 0$ (The effect of the *k-th* is not significant)
$H_1: \beta_k \neq 0, k = 1,2,\ldots.,p$ (The effect of the *k-th* is significant)
Test statistics:

$$W = \frac{\widehat{\beta}_k}{se(\widehat{\beta}_k)} \tag{12}$$

The rejection area is rejected H_0 if $W\ or\ t-value > t_{(n-k,\alpha/2)}$ or p-value $< \alpha$, which means the k-th independent variable has an effect on the dependent variable.

2.5 Anemia

Anemia is a condition in which the hemoglobin level in the blood decreases below the normal value for individual measurements. Anemia is a public health problem for around 1.62 billion people worldwide because people are often unaware that they suffer from anemia. Anemia remains a major challenge for human health [13]. The prevalence of anemia in developed countries is 9 percent, while in developing countries, the prevalence is 43 percent. Anemia contributes to more than 115,000 maternal deaths and 591,000 prenatal deaths globally per year [14].

Anemia is one of the nutritional problems in Indonesia and is most commonly found in women of childbearing age and children. Anemia ranks fourth out of 25 health problems women of childbearing age and children suffer. The prevalence of anemia in children is 47 percent and 42 percent in pregnant women, and 30 percent in non-pregnant women aged 15-49. According to the Riskesdas survey data, anemia cases in Indonesia increased from 2013 to 2018, from 37.1% to 48.9%. The World Health Organization (WHO) targets to reduce the prevalence of anemia in WCA by 50 percent by 2025 [15].

The anemia problem is still quite serious even though various efforts have been made to reduce it, among others, through iron supplementation through the administration of Fe tablets. Anemia conditions can increase the risk of maternal death during childbirth, give birth to babies with low birth weight, the fetus and mother are susceptible to infection, miscarriage, increase the risk of premature birth [16], as well as abortion and congenital disabilities [17]. The consequences for the population with chronic anemia will prolong the loss of productivity from impaired work capacity and cognitive impairment, increase susceptibility to infection, and provide an economic burden.

The high prevalence of anemia in the community has been going on for a long time (chronic), but this has received less attention because it is considered normal and does not show significant symptoms as a disturbance in activities in daily life. In fact, anemia is the second leading cause of disability worldwide and a serious public health problem worldwide because of its impact on the quality of human resources. It reflects the value of the community's socio-economic welfare, which will have a harmful impact. Therefore anemia requires serious attention from all parties related to health services at the forefront [7]. Thus, efforts are needed to prevent and overcome them. One way that can be done is to determine the factors that influence anemia, especially in women of childbearing age (WCA) in Indonesia.

WCA groups are susceptible to iron deficiency anemia due to several problems experienced by WCA, such as experiencing monthly menstruation, experiencing pregnancy, lack of dietary iron intake, and parasitic infections such as malaria and worms. The factors that cause anemia in the population involve a complex interaction of social, economic, ecological, and biological factors [18].

Regarding infectious diseases, tuberculosis, malaria, and helminthiasis are causes of anemia, especially in endemic areas [19], because they cause a decrease in hemoglobin levels due to excessive destruction of red blood cells and disruption of erythrocytes [20]. In addition, anemia is also associated with a history of other diseases, including pneumonia, cancer, kidney failure, Upper Respiratory Tract Infection (ARI), or other chronic conditions that increase the risk of anemia. Chronic disease can cause changes in the body's system to produce healthy red blood cells. This condition causes the production of red blood cells to be inhibited, red blood cells to die more quickly, or to fail altogether. In addition, it can also interfere with an appetite, ultimately reducing the level of nutritional consumption [21].

In addition, household socio-economic conditions are also associated with the incidence of anemia. Several studies have shown that poor geographic location can also influence the incidence of anemia [7]. The problem of anemia does not only occur in remote areas but also in rural and urban areas. There are differences in the prevalence of anemia and the risk of CED between women in rural and urban areas [22]. Based on the 2013 Riskesdas, the prevalence of pregnant women with anemia in rural areas was 37.8 percent, higher than in urban areas at 36.4 percent [23]. Factors that cause anemia, such as inadequate nutrition and poor sanitation, are indicators that must be met in a clean and healthy living behavior in the household [24].

3. RESULTS AND DISCUSSION

3.1. Data Exploration of WCA Affected by Anemia in Indonesia

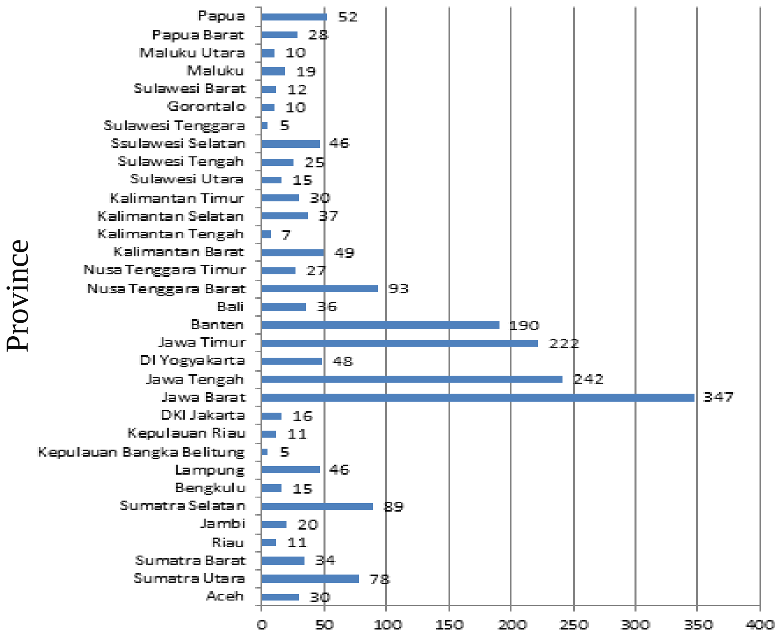

Figure 1. Number of Cases of Anemia in WCA per Province in Indonesia

Based on **Figure 1**, it can be seen that the province with the highest number of cases of WCA affected by anemia is Java. West Java Province has the highest number of cases of anemic WCA with a population of 347, meaning that the number of WCA with anemia per 100 WCA in this province is 21.38, with a total of 1623 WCA. Meanwhile, the Provinces of Southeast Sulawesi and the Bangka Belitung Islands had the least number of cases, namely 5 WCA each, meaning that there were 25 WCA who were anemic per 100 WCA in

Southeast Sulawesi Province with a total of 20 WCA. Bangka Belitung Islands Province has 26.32 anemic WCA per 100 WCA with a total of 19 WCA.

The number of WCA cases affected by anemia in each province in Indonesia has an average of 57.73 with a large standard deviation of 78.78 because there are significant differences between the number of WCA cases in each province in Indonesia.

3.2. Multicollinearity

The multicollinearity test is used to find out whether there is a strong relationship between the independent variables used. In the multicollinearity test, the value of VIF (Variance Inflation Factor) is less than 10, indicating that the assumption of not multicollinearity is met [7]. The VIF value of the variables used is as follows:

Table 3. ***VIF value.***

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7
VIF	1.44	1.46	1.64	1.80	2.00	1.52	1.30

From these results, it can be seen that the VIF value of each variable is less than 10, so it can be concluded that the assumption of not multicollinearity is met.

3.3. Poisson Regression Modeling

The number of WCA affected by anemia per 100 WCA in each province in Indonesia is count data so that it is assumed to have a Poisson distribution. The following are the results of parameter estimation using Poisson regression:

Table 4. Parameter Estimation of Poisson Regression

	Estimator	**Z Value**	**P Value**
Constant	3.362	17.880	$\mathbf{2 \times 10^{-16}}$
X_1	-0.002	-1.143	0.2529
X_2	0.005	0.174	0.8620
X_3	0.002	0.286	0.7746
X_4	0.017	2.568	0.0102
X_5	0.014	0.826	0.4089
X_6	-0.015	-0.171	0.8645
X_7	-0.006	-1.976	0.0481
Deviance = 44.044		DF = 25	
AIC = 221.83			

To find out whether all independent variables have an effect on the dependent variable, simultaneous testing is used (overall). The hypotheses used are:

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

$H_1: there\ is\ at\ least\ one\ \ \beta_k \neq 0, i = 1, 2, \ldots., n; k = 0,1,2, \ldots, p$

Based on the overall test with a significance level of 10%, the deviance value $D(\hat{\beta})$ =44.044 so that the value $\chi^2_{p(a)} = \chi^2_{7(0.1)} = 12.017$ is smaller than the deviance so H_0 is rejected, which means that there is at least one independent variable that has a significant effect on the dependent variable. So it is necessary to do a partial test to find out which independent variables have an effect on the dependent variable with the following hypothesis:

$H_0: \beta_k = 0$ (The effect of the *k-th* is not significant)

$H_1: \beta_k \neq 0, k = 1, 2, \ldots., p$ (The effect of the *k-th* is significant)

Based on the partial test, with a significance level of 10% obtained the value $|Z_{count}|$ value is greater than $Z_{\left(\frac{0.1}{2}\right)} = 1.28$ value or p-value is smaller than α (0.1) on variable X_4 dan X_7 which means the independent variable and has an influence on the dependent variable. While other independent variables have no effect on the dependent variable.

Based on Table 3, the Poisson regression model is obtained as follows:

$\mu_i = \exp(3.362 - 0.002X_1 + 0.005X_2 + 0.002X_3 + 0.017X_4 + 0.014X_5 - 0.015X_6 - 0.006X_7)$

3.4. Overdispersion

The assumption that must be met in Poisson regression is that the average of the response variables must be equal to the variance, which is known as equidispersion [1]. However, in reality this rarely happens because generally enumerated data has a variance greater than the average or it is called overdispersion [2]. To detect cases of overdispersion, it is done by dividing the *deviance* in the Poisson regression model by the degree of freedom.

In this case, the deviance while the degree of freedom is 25 so that the ratio between the deviance degree of freedom is 1.7617. Because the value is more than 1, it can be concluded that the data on the number of WCA affected by anemia in Indonesia experience cases of overdispersion. In addition to looking at the ratio between the deviance and the degree of freedom, overdispersion detection can also be done by statistical testing. The statistical test that can be used to detect overdispersion is the overdispersion test that can use the package from the R software [25]. Hypothesis:

H_0: no overdispersion
H_1: occurs overdispersion
Test statistic:

```
        Overdispersion test

data:  overdispersi.anemia
z = 6.5329, p-value = 3.225e-11
alternative hypothesis: true alpha is greater than 0
sample estimates:
       alpha
6.296231e-57
```

Figure 2. Overdispersion Test

Reject H_0 because *p-value* $< \alpha$. So, with a confidence level of 90%, it can be concluded that there is overdispersion.

If there is an overdispersion case, the resulting parameter estimation will be biased and inefficient if Poisson regression is still used in the analysis. The way to overcome overdispersion in Poisson regression is to use the Negative Binomial regression method. The advantage of this model is that it is more flexible because the assumptions of the mean and variance do not have to be the same [4]. The first step in using negative binomial regression is to determine the initial θ value obtained through trial-error results so that the ratio of the deviance value to the degree of freedom is 1, which means that there is no case of overdispersion. The initial θ determination of aims to minimize the dispersion parameter so that it can overcome the overdispersion case.

Table 5. Trial-Error Initial θ

Initial θ	*Deviance*	DF	*Deviance*/ DF
1	1.9970	25	0.08
5	8.4047	25	0.34
10	14.058	25	0.56
20	21.230	25	0.85
28	24.877	25	0.99508
28.2	24.953	25	0.99812
28.3	24.991	25	0.99964
28.32	24.998	25	0.99992
28.321	24.999	25	0.9999
28.324	25	25	1

Based on the results of trial error, the initial θ value is 28.324 which has a ratio of deviance to the degree of freedom of 1. So that further negative binomial regression modeling was carried out with an initial θ of 28.324.

3.5. Negative Binomial Regression Modeling

The parameter estimation results of the Negative Binomial regression model using the Maximum Likelihood Estimation with the Newton Raphson procedure are as follows:

Table 6. Parameter Estimation of Negative Binomial Regression

	Estimator	T Value	P Value
Constant	3.343	13.648	$\mathbf{4.34\times 10^{-13}}$
X_1	-0.002	-0.866	0.3949
X_2	0.003	0.082	0.9350
X_3	0.002	0.218	0.8290
X_4	0.017	1.803	0.0834
X_5	0.014	0.635	0.5312
X_6	-0.018	-0.166	0.8695
X_7	-0.006	-1.421	0.1676
Deviance = 25		DF = 25	
AIC = 221.72			

Based on the output above, the negative binomial regression model is obtained, namely:
$\mu_i = \exp\,(3.343 - 0.002X_1 + 0.003X_2 + 0.002X_3 + 0.017X_4 + 0.014X_5 - 0.018X_6 - 0.006X_7)$

To determine whether all the independent variables jointly affect the case of anemia in women of childbearing age (WCA). WCA so is used overall test with the hypothesis, namely:
$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$
$H_1: there\ is\ at\ least\ one\ \beta_k \neq 0, i = 1, 2, \ldots, n; k = 0,1,2, \ldots, p$

Based on the overall test with a significance level of 10%, the deviance value is $D(\hat{\beta}) = 25$ so $\chi^2_{p(a)} = \chi^2_{7(0.1)} = 12.017$ is smaller than the deviance value so H_0 is rejected which means there is at least one independent variable that has a significant effect on the dependent variable. So, it is necessary to do a partial test to find out which independent variables have an effect on the dependent variable with the following hypothesis:

$H_0: \beta_k = 0$ (The effect of the *k-th* is not significant)
$H_1: \beta_k \neq 0, k = 1, 2, \ldots, p$ (The effect of the *k-th* is significant)

Based on the partial test results, with a significance level of 10 % obtained p-value is smaller than α (0.1) on variable X_4 so that it rejects which means the independent variable X_4 has an influence on the dependent variable. While other independent variables have no effect on the dependent variable.

3.6. Best Model Selection

Selection of the best model can be seen using the AIC value in the Poisson regression model and negative binomial regression. Poisson regression has an AIC value of 221.83, while negative binomial regression has an AIC value of 221.72. So it can be concluded that negative binomial regression is the best model in modeling the number of WCA affected by anemia per 100 WCA in Indonesia because it has the smallest AIC value even though the difference is not too far from the AIC in the Poisson regression model. It can also be supported that Poisson regression is not suitable to be used in the analysis because of the case of overdispersion. The Negative Binomial regression model is as follows:

$$\mu_i = \exp\,(3.343 - 0.002X_1 + 0.003X_2 + 0.002X_3 + 0.017X_4 + 0.014X_5 - 0.018X_6 - 0.006X_7)$$

Based on the significant variables from the model formed in Indonesia, it can be concluded that for every 1 WCA infected with malaria per 100 WCA, the number of WCA who are anemic per 100 WCA becomes 1.02 times than before, assuming other variables are constant. Regardless of the significant variables, the variable is that many WCA lives in rural areas, are exposed to hepatitis, and districts/cities that have PHBS policies have a negative regression coefficient. This means that for every increase in 1 WCA living in rural areas and 1 WCA exposed to hepatitis per 100 WCA, as well as 1% districts/cities with PHBS, the number of WCA who are anemic per 100 WCA will be 0.998, 0.982, and 0.994 times, respectively, from the previous with assuming other variables are constant. While the variables pneumonia, ARI, and TB have a positive regression coefficient, meaning that for every 1 WCA who has pneumonia, ARI, and TB increases per 100 WCA, the number of WCA who are anemic per 100 WCA becomes 1.003, 1.002, 1.014 times from the previous assuming other variables are constant.

4. CONCLUSIONS

Based on the results of the study, it is concluded that negative binomial regression is more effective in overcoming the problem of overdispersion in Poisson regression. This is evidenced by the results of applying of both methods to case studies of anemia in WCA in Indonesia, where the AIC value in the negative binomial regression is smaller than the Poisson regression, which is 221.72 compared to 221.83. In the case study, with a significance level of 10%, the number of WCA affected by malaria per 100 WCA affects cases of anemia in WCA. While other independent variables have no effect. Pneumonia, ARI, malaria, and tuberculosis have positive regression coefficients, meaning that an increase of 1 WCA affected by pneumonia, ARI, TBS, and malaria per 100 WCA means the number of WCA who are anemic per 100 WCA becomes 1.003, 1.002, 1.02, and 1.014 times from previously assuming other variables are constant.

REFERENCES

[1] A. J. Dobson and A. Barnett, *An Introduction to Generalized Linear Models*. Boca Raton: Chapman & Hall/CRC Press, 2018.
[2] D. Rahmayanti and S. W. Rizki, "Penanganan Overdispersi dengan Model Binomial Negatif pada Data Klaim Asuransi Kendaraan Bermotor Roda Empat," *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 7, no. 2, pp. 55–62, 2018.
[3] M. Majore, D. T. Salaki, and J. D. Prang, "Penerapan Regresi Binomial Negatif dalam Mengatasi Overdispersi Regresi Poisson pada Kasus Jumlah Kematian Ibu," *D'CARTESIAN*, vol. 9, no. 2, pp. 133–139, 2021.
[4] D. Yuli and D. Indriani, "Pemodelan Binomial Negatif Untuk Mengatasi Overdispersi Data Diskrit pada Kasus TB di Provinsi Jawa Timur," *Biometrika dan Kependud.*, vol. 4, no. 2, pp. 134–142, 2015.
[5] N. H. Fitrial and A. Fatikhurrizqi, "Modelling The Number Of COVID-19 Cases In Indonesia Using Poisson Regression And Binomial Regression Approach," *Semin. Nas. Off. Stat. 2020*, no. 1, pp. 65–72, 2020.
[6] Luhung Mustika Budiharti and S. Sunendiari, "Pemodelan dan Pemetaan Jumlah Penderita Kusta di Jawa Barat dengan Regresi Binomial Negatif dan Flexibly Shaped Spatial Scan Statistic," *J. Ris. Stat.*, vol. 1, no. 2, pp. 99–106, 2021, doi: 10.29313/jrs.v1i2.409.
[7] M. IB, *Ilmu Kebinanan, Penyakit Kandungan, dan Keluarga Berencana*. Jakarta: EGC, 2012.
[8] M. Ridwan and S. Sunendiari, "Mendeteksi dan Mengatasi Multikolinieritas pada Data Penelitian Diabetes Melitus Wanita Suku Indian Tahun 2018," *Statistika*, vol. 7, no. 1, pp. 64–70, 2021.
[9] Rahmadeni and D. Anggreni, "Analisis Jumlah Tenaga Kerja Teehadap Jumlah Pasien RSUD Arifin Achmad Pekanbaru Menggunakan Metode Regresi Gulud," *Sains, Teknol. dan Ind.*, vol. 12, no. 1, pp. 48–57, 2014.
[10] R. G. Walpole, R. H. Myers, and K. Ye, *Probability and Statistics*. United States of America: Pearson Education, 2012.
[11] S. A. Pardo and M. A. Pardo, *Statistical Methods for Field and Laboratory Studies in Behavioral Ecology*. New York: Chapman & Hall/CRC Press, 2018.
[12] A. A. Johnson, M. Q. Ott, and M. Dogucu, *An Introduction to Applied Bayesian Modelling*. New York: Chapman & Hall/CRC Press, 2022.
[13] WHO, *The World Health Report. Reducing Risks, Promoting Healthy Life*. Geneva: World Health Organization, 2022.
[14] Y. Balarajan, U. Ramakrishnan, E. Ozaltin, A. H. Shankar, and Subramanian, "Anemia in Low-Income and Middle-Income Countries," *Lancet*, vol. 278, no. 9809, pp. 2123–2135, 2012.
[15] WHO, *WHA Global Nutrition Targets 2025: Anemia Policy Brief*. Geneva: World Health Organization, 2014.
[16] G. Masukume, A. Khashan, L. Kenny, P. Baker, and G. Nelson, "Risk Factors and Birth Outcomes of Anemia in Early Pregnancy in A Nulliparous Cohort," *PLoS One*, vol. 10, no. 4, pp. 1–15, 2015.
[17] A. W. Sumantri, "Hubungan Pengetahuan dan Pendidikan Ibu Hamil dengan Anemia di RT 10 RW 8 Wilayah Kerja Puskesmas Kemalaraja Baturaja," *Kesehat. Saelmakers Perdana*, vol. 4, no. 1, pp. 51–56, 2018.
[18] M. I. Akbar *et al.*, "Pelaksanaan Program Keping Emas pada Ibu Hamil Kekurangan Energi Kalori dan Anemia di Desa Kronjo Tahun 2019," *Bid. Ilmu Kesehat.*, vol. 10, no. 1, pp. 108–121, 2020.
[19] H. Sari, Y. Yarmaliza, and Z. Zakiyuddin, "Faktor-Faktor yang Mempengaruhi Kejadian Anemia pada Ibu Hamil di Wilayah Kerja Puskesmas Samadua Kecamatan Samadua Kabupaten Aceh Selatan," *Mhs. Kesehat. Masy.*, vol. 2, no. 1, pp. 178–202, 2018.
[20] E. Agustin, E. Resnhaleksmana, and I. W. Getas, "Hubungan Kepadatan Parasit Terhadap Kadar Hemoglobin pada Penderita Malaria Asimtomatik di Gunung Sari," *Anal. Med. Biosains*, vol. 9, no. 1, pp. 60–65, 2022.
[21] A. M. Maulani, "Faktor-Faktor yang Berhubungan dengan Kejadian Anemia pada Remaja Putri di SMPN 1 Kalibawang Kabupaten Kulon Progo Tahun 2020," Poltekkes Kemenkes Yogyakarta, 2020.
[22] I. Oktaviani, D. Rahmawati, and Y. N. R. Kana, "Prevalensi dan Faktor Risiko Anemia pada Anak di Negara Maju," *D'CARTESIAN*, vol. 16, no. 4, pp. 218–226, 2021.
[23] R. K. Dasar, *Riset Kesehatan Dasar (Riskesdas)*. Indonesia: Riset Kesehatan Dasar, 2013.
[24] E. Purwaningsih, W. Widyawati, and A. Akhmadi, "Perilaku Hidup Bersih dan Sehat pada Ibu Hamil dengan Anemia di Yogyakarta," *Kesehat. Samodra Ilmu*, vol. 9, no. 1, pp. 29–43, 2018.
[25] A. Y. Herindrawati, I. N. Latra, and Purhadi, "Pemodelan Regresi Poisson Inverse Gaussian Studi Kasus Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2015," *Sains dan Seni ITS*, vol. 6, no. 1, pp. 143–148, 2017.