

Negative Binomial Regression Modeling of HIV/AIDS Cases in West Java Province

BQ. Nurkhalishah K¹, Aswi Aswi^{2*}, Wahidah Sanusi³, Nur Salsabila Wero⁴

^{1,4}Postgraduate Student in Statistics, Faculty of Mathematics and Natural Sciences, Makassar State University, Makassar, Indonesia

²Department of Statistics, Faculty of Mathematics and Natural Sciences, Makassar State University, Makassar, Indonesia

³Department of Mathematics, Faculty of Mathematics and Natural Sciences, Makassar State University, Makassar, Indonesia

*Corresponding author: aswi@unm.ac.id

*Submission date: 06 July 2026, Revision: 20 August 2026, Accepted: 14 September 2026

ABSTRACT

Human Immunodeficiency Virus (HIV)/acquired immunodeficiency syndrome (AIDS) remains a public health problem in Indonesia, including West Java Province, which has a relatively high number of cases with varying characteristics across regions. This study aimed to model the factors associated with the number of HIV/AIDS cases in West Java Province using negative binomial regression. A quantitative approach was employed using secondary data from the 2024 West Java Provincial Health Profile and the Central Statistics Agency. The response variable was the number of HIV/AIDS cases, while the predictor variables included population density, the percentage of people living in poverty, and the open unemployment rate. The analysis involved descriptive statistics, multicollinearity assessment using the variance inflation factor, Poisson and negative binomial regression modeling, assessment of overdispersion, and model selection based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The results showed that the data exhibited overdispersion, indicating that negative binomial regression was more appropriate than Poisson regression. The negative binomial regression model including the open unemployment rate was selected as the best-fitting model, with the lowest AIC and BIC values. The open unemployment rate was significantly associated with the number of HIV/AIDS cases, with an incidence rate ratio (IRR) of 1.341, indicating an approximately 34.1% increase in the expected number of cases for a one-unit increase in the open unemployment rate.

KEYWORDS

HIV/AIDS, Negative Binomial Regression, Overdispersion

1. INTRODUCTION

Human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) remains a significant public health concern in many countries, including Indonesia [1]. HIV infection progressively weakens the immune system, thereby increasing individuals' susceptibility to a wide range of opportunistic infections and other HIV-related complications [2]. If left untreated, HIV infection may progress to acquired immunodeficiency syndrome (AIDS), a severe condition marked by progressive impairment of immune function [3]. Despite ongoing efforts to implement HIV prevention and treatment programs, the burden of HIV/AIDS in Indonesia remains substantial, posing a continuing challenge to the attainment of national health development targets [4].

West Java Province is among the most populous regions in Indonesia and is characterized by a high level of population mobility. This situation has the potential to increase the risk of the spread of infectious diseases, including HIV/AIDS. High population density, urbanization, socio-economic mobility, and variations in demographic characteristics between regions contribute to the variability in the number of HIV/AIDS cases in West Java. This variation in case numbers requires a more

in-depth study so that local governments can formulate more effective and targeted HIV/AIDS prevention and control policies. Therefore, a statistical approach is needed that can identify factors related to the number of HIV/AIDS cases to provide a more comprehensive understanding of disease spread patterns [5].

In health data analysis, the number of HIV/AIDS cases is generally count data, which is often modeled using Poisson regression [6]. However, one of the important assumptions in Poisson regression is that the mean and variance must be equal (equidispersion) [7]. In reality, data on the number of disease cases often show a variance greater than the mean (overdispersion) [8]. If overdispersion is ignored, the obtained parameter estimates will be inefficient and can produce inaccurate conclusions. Therefore, alternative methods are needed that are more appropriate to address these data characteristics [4].

Negative binomial regression is a widely employed statistical method for addressing overdispersion in count data [9]. This method expands on the Poisson regression model by adding a dispersion parameter to accommodate greater data variability [10]. Several studies have been conducted on modeling HIV/AIDS cases using a regression approach for census data. One study modeled the number of HIV/AIDS cases in Lampung Province using negative binomial regression and found that the model was superior to Poisson Regression because it was able to address the problem of overdispersion in the data [4]. Furthermore, a previous study employed a spline-based negative binomial regression model to analyze HIV cases in Indonesia, demonstrating greater flexibility in capturing the underlying relationship between the response variable and its explanatory variables [3]. Finally, previous research indicates that negative binomial regression is more suitable than Poisson regression for modeling the number of HIV-related deaths in Indonesia, as it provides more reliable estimates when the data exhibit overdispersion [2].

However, these studies still have limitations because they were conducted in different regions, used national or provincial data coverage, and mostly focused on developing methods or modeling the number of HIV patient deaths. To date, studies that specifically model the number of HIV/AIDS cases in West Java Province using negative binomial regression, taking into account the characteristics of data overdispersion and demographic factors such as population density, percentage of poor people, and unemployment rates, are still limited. In fact, West Java Province is one of the provinces with a high population and mobility, thus having characteristics of HIV/AIDS distribution that are different from other regions. Therefore, this study aims to model the number of HIV/AIDS cases in West Java Province using negative binomial regression to obtain a more appropriate model and identify factors that influence the number of HIV/AIDS cases as a basis for developing more effective control policies.

In this study, three demographic and socio-economic variables were selected as predictor variables, namely population density, poverty rate, and unemployment rate. These variables were selected because they represent important demographic and socio-economic characteristics that may be associated with variations in HIV/AIDS cases across regions. Population density reflects the concentration and mobility of the population, while poverty and unemployment represent socio-economic conditions that may be associated with access to healthcare, social vulnerability, and the risk of HIV/AIDS transmission. Previous studies have been conducted in different regions and have mostly focused on developing methods or modeling HIV-related outcomes, leaving limited research specifically examining HIV/AIDS cases in West Java Province using negative binomial regression while considering the overdispersion characteristics of the data and relevant demographic and socio-economic factors. Therefore, this study aims to model the number of HIV/AIDS cases in West Java Province using negative binomial regression to obtain an appropriate model and identify factors associated with variations in HIV/AIDS cases as a basis for developing more effective control policies.

2. LITERATURE REVIEW

2.1 Poisson Regression

Within the framework of Generalized Linear Models (GLMs), Poisson regression is commonly employed to analyze count responses that record the frequency of events occurring over a given time interval or within a defined area [6, 11]. The Poisson model assumes that the response follows a Poisson distribution in which its conditional mean coincides with its conditional variance, satisfying the equidispersion property [12, 13]. The assumed probability distribution serves as the basis for obtaining the model parameter estimates using the Maximum Likelihood Estimation (MLE) method [14]. Poisson regression is widely applied in the field of public health to model the number of disease events because it is able to explain the relationship between the response variable and the predictor variable through the log link function. The probability function of the Poisson distribution

is stated as follows [15].

$$P(Y_i = y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad (1)$$

with

$$E(Y_i) = \text{Var}(Y_i) = \mu_i \quad (2)$$

The relationship between the response variable and the set of predictor variables can be represented mathematically as follows:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} \quad (3)$$

where β_0 is the intercept and β_j are the regression coefficients estimated using the MLE method. Despite its widespread application in health data analysis, this model provides reliable results only when the equidispersion assumption, in which the variance equals the mean, is satisfied [12].

2.2 Multicollinearity Test

Multicollinearity test refers to a condition in which strong linear relationships exist among the independent variables in a regression model [16]. This condition may lead to unstable parameter estimates, inflated standard errors, and reduced accuracy in interpreting the individual effects of the independent variables [17]. Therefore, multicollinearity testing should be conducted prior to model estimation to ensure that the resulting model is more reliable. One of the most commonly used measures for detecting multicollinearity is the Variance Inflation Factor (VIF), which is defined as follows [18]:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (4)$$

where R_j^2 is the coefficient of determination obtained by regressing the j -th independent variable on the remaining independent variables. In general, a VIF value of less than 10 indicates that the model does not suffer from serious multicollinearity, whereas a VIF value of 10 or greater suggests the presence of multicollinearity that should be carefully addressed in the analysis [17].

2.3 Overdispersion Test

Equidispersion is a fundamental assumption in Poisson regression, under which the conditional mean and variance of the response variable are assumed to coincide. However, in health-related count data, the variance often exceeds the mean, a condition known as overdispersion [19]. If this condition is ignored, the estimated standard errors may be underestimated, leading to potentially misleading conclusions in parameter significance testing. Overdispersion is indicated when the deviance or Pearson chi-square statistic, relative to its corresponding degrees of freedom, exceeds the level expected under the Poisson assumption [20]. A ratio greater than one indicates the presence of overdispersion, suggesting that the Poisson regression model is no longer appropriate. Under such circumstances, an alternative model capable of accommodating a variance larger than the mean is required [21].

2.4 Binomial Negative Regression

The Negative Binomial regression model extends the Poisson framework by allowing for greater variability in count data, particularly when overdispersion is present [22]. This model adds a dispersion parameter so that the variance is no longer equal to the mean. With this parameter, the Negative Binomial model is able to provide more accurate parameter estimates and standard errors than Poisson regression when the data is overdispersion. The expected value and variance of the Negative Binomial distribution are given by [14]:

$$E(Y_i) = \mu_i \quad (5)$$

and

$$\text{Var}(Y_i) = \mu_i + \alpha \mu_i^2 \quad (6)$$

where α is the dispersion parameter. If $\alpha = 0$, the Negative Binomial model reverts to a Poisson model. The relationship between the response variable and the predictor variable remains the same using the log link function, namely [14]:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} \quad (7)$$

3. METHODOLOGY

This study adopts a quantitative approach to investigate the factors associated with the number of HIV/AIDS cases in West Java Province using Generalized Linear Models (GLM). The analysis employs two count-data regression methods, namely Poisson regression and negative binomial regression. The study utilizes secondary data sourced from the West Java Provincial Health Profile and the 2024 publications of the Central Statistics Agency (BPS) of West Java Province [1, 23]. The units of analysis in this study are 27 districts/cities in West Java Province that have complete data according to the research variables. The variables in this study are given in **Table 1**.

Table 1. Research Variables

Variable	Description	Unit
Y	Number of HIV/AIDS Cases	Cases
X_1	Population Density	Persons/km ²
X_2	Poverty Rate	Percent (%)
X_3	Open Unemployment Rate	Percent (%)

The steps taken in data processing to obtain the best model for negative binomial regression analysis as an alternative to addressing overdispersion in Poisson regression on HIV/AIDS case data in West Java Province are as follows:

1. Describe the characteristics of the number of HIV/AIDS cases, population density, percentage of poor people, and the open unemployment rate using descriptive statistics.
2. Evaluate multicollinearity among the independent variables using the VIF as the assessment criterion.
3. Conduct modeling using Poisson regression.
4. Detect overdispersion in the Poisson regression model by comparing the deviance value to the degrees of freedom (deviance/degrees of freedom).
5. If overdispersion is present, modeling using negative binomial regression is performed with the following steps:
 - (a) Estimate the parameters of the negative binomial regression model.
 - (b) Assess the statistical significance of the parameters in the negative binomial regression model through simultaneous and partial significance tests.
 - (c) Select the best model based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).
 - (d) Interpret the estimated negative binomial regression coefficients to assess the relationships between population density, poverty rate, and open unemployment rate and the number of HIV/AIDS cases in West Java Province.

4. RESULT & DISCUSSION

4.1 Descriptive Statistics

The descriptive analysis summarizes the distribution and characteristics of the study variables. In the model, HIV/AIDS case counts serve as the response variable, whereas population density, poverty rate, and unemployment rate are specified as explanatory variables. A summary of the descriptive statistics is presented in **Table 2**.

As presented in **Table 2**, the number of HIV/AIDS cases (Y) across the 27 districts and municipalities in West Java ranged from 13 to 373 cases, with an average of 113.9 cases. The variance (8370.564) was substantially larger than the mean, indicating considerable variability in the distribution of HIV/AIDS cases across regions. This finding suggests substantial heterogeneity in the number of HIV/AIDS cases among districts and municipalities and provides preliminary evidence of overdispersion. The population density variable (X_1) ranged from 385 to 15176 persons/km², with a mean of 3910.9 persons/km² and a variance of

Table 2. Descriptive Statistics of the Study Variables

Variable	Min	Max	Mean	Variance
Y	13.0	373.0	113.9	8370.564
X_1	385.0	15176.0	3910.9	21791309
X_2	2.340	11.930	8.014	7.014903
X_3	1.580	8.970	6.525	2.912172

21791309. The large variance indicates substantial differences in population density between districts and municipalities in West Java, suggesting that population density may contribute to regional differences in HIV/AIDS incidence.

The poverty rate (X_2) varied from 2.34% to 11.93%, with a mean of 8.014% and a variance of 7.0149, indicating moderate variation in poverty levels across the study areas. Meanwhile, the open unemployment rate (X_3) ranged from 1.58% to 8.97%, with a mean of 6.525% and a variance of 2.9122, reflecting a relatively more homogeneous distribution compared with the other explanatory variables. Overall, the descriptive statistics indicate considerable regional variation in the characteristics of the study variables, particularly for the number of HIV/AIDS cases and population density. Furthermore, the response variable exhibited a variance that was substantially greater than its mean, suggesting that the equidispersion assumption of the Poisson regression model may not hold. Therefore, formal overdispersion testing is necessary before determining the most appropriate regression model for the analysis.

4.2 Multicollinearity Test

Multicollinearity testing is performed to determine whether there is a strong linear relationship between predictor variables in the model [17]. The presence of multicollinearity can cause unstable parameter estimates and increase standard errors, thus reducing model interpretation accuracy. In this study, multicollinearity detection was performed using the Variance Inflation Factor (VIF). A model is considered free of multicollinearity if each predictor variable has a VIF value < 10 [16]. The VIF values for each predictor variable in this study are presented in **Table 3**.

Table 3. VIF's Value of the Predictor Variables

Variable	VIF
X_1 (Population Density)	2.015916
X_2 (Poverty Rate)	1.819341
X_3 (Open Unemployment Rate)	1.232306

As shown in **Table 3**, the population density variable (X_1) had a VIF value of 2.015916, while the poverty rate (X_2) and open unemployment rate (X_3) had VIF values of 1.819341 and 1.232306, respectively. None of the predictor variables produced a VIF value above 10, indicating that the model was not affected by substantial multicollinearity. The absence of substantial multicollinearity indicates that the explanatory variables contain relatively distinct information. This condition allows their individual associations with the number of HIV/AIDS cases in West Java Province to be assessed without substantial distortion from linear dependence among predictors. These findings are consistent with previous studies reporting that VIF values below 10 indicate the absence of serious multicollinearity, allowing all predictor variables to be retained in the regression model [24]. Similarly, previous studies have emphasized that assessing multicollinearity using the VIF is an essential preliminary step before fitting Poisson or Negative Binomial regression models, as it helps ensure stable parameter estimation and facilitates reliable interpretation of the regression coefficients [2].

4.3 Poisson Regression Model

The Poisson regression model was initially employed to examine the effects of population density, poverty rate, and open unemployment rate on the number of HIV/AIDS cases in West Java Province. The parameter estimates of the Poisson regression model are presented in **Table 4**.

Table 4. Parameter Estimates of the Poisson Regression Model

Variable	Estimate	z-value	p-value
Intercept	3.808000	27.346	<0.0001
X_1 (Population Density)	-0.00001213	-2.412	0.0159
X_2 (Poverty Rate)	-0.068360	-7.249	<0.0001
X_3 (Open Unemployment Rate)	0.220500	15.162	<0.0001

Based on the estimation results, the resulting Poisson regression model can be expressed as follows:

$$\log(\hat{\mu}) = 3.808 - 0.00001213X_1 - 0.06836X_2 + 0.2205X_3$$

or equivalently,

$$\hat{\mu} = \exp(3.808 - 0.00001213X_1 - 0.06836X_2 + 0.2205X_3)$$

where $\hat{\mu}$ denotes the expected number of HIV/AIDS cases, X_1 represents population density, X_2 denotes the poverty rate, and X_3 represents the open unemployment rate. As shown in **Table 4**, all independent variables were statistically significant at the 5% significance level ($p\text{-value} < 0.05$). This indicates that each predictor variable has a statistically significant effect on the number of HIV/AIDS cases in West Java Province. Population density (X_1) had a negative coefficient of -0.00001213 ($p = 0.0159$), indicating that, after controlling for the other variables, higher population density was associated with a slight decrease in the expected number of HIV/AIDS cases. Likewise, the poverty rate (X_2) exhibited a negative coefficient of -0.06836 ($p < 0.0001$), suggesting that an increase in the poverty rate was associated with a lower expected number of HIV/AIDS cases. In contrast, the open unemployment rate (X_3) had a positive coefficient of 0.2205 ($p < 0.0001$), implying that higher unemployment was associated with an increase in the expected number of HIV/AIDS cases. Overall, these results indicate that all three explanatory variables contributed significantly to explaining the variation in HIV/AIDS cases under the Poisson regression model.

A similar analytical strategy has been reported in earlier research on HIV incidence in West Java, which initially applied the Poisson regression framework to investigate potential determinants of newly reported cases, where several explanatory variables were found to have statistically significant effects on HIV incidence [6]. Similarly, a study on modeling AIDS cases in East Nusa Tenggara Province demonstrated that Poisson regression is an effective initial modeling approach for identifying factors influencing count-based disease outcomes before evaluating the underlying model assumptions, particularly the equidispersion assumption [24].

4.4 Overdispersion Test

The Poisson regression model is based on an equidispersion assumption, under which the conditional variance of the response variable corresponds to its conditional mean [25]. When the variance exceeds the mean, the data are said to exhibit overdispersion. This condition may lead to biased standard error estimates, resulting in unreliable statistical inference and inaccurate parameter significance tests [26]. Therefore, testing for overdispersion is necessary to determine whether the Poisson regression model remains appropriate or whether an alternative modeling approach is required [19]. Overdispersion can be assessed by evaluating the ratio of the deviance statistic or Pearson's chi-square statistic to its corresponding degrees of freedom. A ratio substantially greater than one suggests that the observed variability exceeds the level assumed under the Poisson model, indicating that Negative Binomial regression may provide a more appropriate alternative [27].

Table 5. Results of the Overdispersion Test

Deviance	Degrees of Freedom (Df)	Deviance/Df
1265.903	23	55.04

As shown in **Table 5**, the deviance was 1265.903 with 23 degrees of freedom, yielding a deviance to degrees of freedom ratio of 55.04. Since the ratio is considerably greater than one, the HIV/AIDS case data in West Java Province demonstrate a

clear presence of overdispersion. This result suggests that the equidispersion assumption required by the Poisson regression model is not satisfied, indicating that the Poisson model may be inadequate for capturing the observed variability in the count data.

These findings are consistent with previous studies reporting that health-related count data frequently exhibit variances substantially larger than their corresponding means, making the Poisson regression model less efficient for parameter estimation [28]. Furthermore, previous research has shown that, under overdispersion, the Negative Binomial regression model generally provides a better fit because it incorporates an additional dispersion parameter that allows the variance to exceed the mean [24]. Therefore, the subsequent analysis was performed using the Negative Binomial regression model to obtain a more appropriate model for explaining the number of HIV/AIDS cases in West Java Province.

4.5 Binomial Negative Regression

After overdispersion was detected in the Poisson regression model, the next step was to fit a Negative Binomial regression model. The Negative Binomial model was selected to account for excess variability in count data, particularly when the observed variance is larger than the corresponding mean. By incorporating a dispersion parameter, the model provides greater flexibility in representing overdispersed data and obtaining reliable coefficient estimates [2]. The estimated parameters and the results of the partial significance tests are presented in **Table 6**.

Table 6. Parameter Estimates of the Negative Binomial Regression Model

Parameter	Estimate	z-value	p-value
Intercept	2.933000	3.565	0.000364
X_1 (Population Density)	0.000000084	0.002	0.998255
X_2 (Poverty Rate)	-0.026670	-0.413	0.679815
X_3 (Open Unemployment Rate)	0.293700	3.510	0.000449

Based on the overdispersion test of the Poisson regression model, the variance of the response variable was found to exceed its mean, indicating that the equidispersion assumption was violated. Therefore, the analysis proceeded with the Negative Binomial regression model, which accounts for overdispersion by incorporating an additional dispersion parameter. This approach provides more efficient parameter estimation and more reliable statistical inference than the Poisson regression model under overdispersed conditions.

As shown in **Table 6**, only the open unemployment rate (X_3) had a statistically significant effect on the number of HIV/AIDS cases, with an estimated coefficient of 0.293700 and a p-value of 0.000449 (< 0.05). The positive coefficient indicates that higher unemployment rates are associated with an increase in the expected number of HIV/AIDS cases. In contrast, population density (X_1) ($p = 0.998255$) and the poverty rate (X_2) ($p = 0.679815$) were not statistically significant, as both p-values exceeded the 0.05 significance level.

The analysis indicates that differences in labor market conditions are more closely associated with the regional distribution of HIV/AIDS cases than variations in population density or poverty. These results also support the use of the Negative Binomial framework for count outcomes exhibiting overdispersion. Compared with the Poisson specification, this model provides greater flexibility in estimating the effects of explanatory variables when the equidispersion assumption does not hold [29]. Previous research has also shown that Negative Binomial regression can adequately account for excess variation in count outcomes. The inclusion of a dispersion component allows the model to accommodate overdispersion, while the significance of individual predictors can be evaluated to determine which variables contribute meaningfully to the response [19].

4.6 Selection of the Best-Fitting Model

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) derived from the Poisson and Negative Binomial regression models were then evaluated in order to select a model. The model that produced the lowest information criteria values was thought to suit the data better. The three candidate models were evaluated using AIC and BIC to determine the most suitable specification. Preference was given to the model that achieved the lowest values for both criteria, as this reflects a more favorable trade-off between goodness of fit and parsimony. The corresponding results are shown in **Table 7**.

Table 7. Comparison of AIC and BIC Values

Model	AIC	BIC
Poisson Regression	1442.7348	1447.9181
Negative Binomial Regression (Full Model)	305.2573	311.7365
Negative Binomial Regression (with unemployment rate)	301.5577	305.4452

The negative binomial regression model including the open unemployment rate produced the lowest AIC (301.5577) and BIC (305.4452) values among the competing models, as shown in **Table 7**. These results indicate that the negative binomial regression model including the open unemployment rate provides the best balance between goodness of fit and parsimony among the models considered. Accordingly, the final model used to explain the variation in HIV/AIDS cases across West Java Province was the negative binomial regression model including the open unemployment rate. These results further demonstrate that the negative binomial regression approach is more appropriate than the Poisson regression model for the present data because it effectively accounts for the overdispersion observed in the response variable. This finding is consistent with previous studies reporting that the negative binomial regression model provides superior performance for analyzing overdispersed count data [2]. Likewise, Bonamente [30] reported that the negative binomial model yields more accurate parameter estimates and more reliable statistical inference than the Poisson model when the variance exceeds the mean. The final negative binomial regression model including the open unemployment rate is expressed as follows:

$$\mu_i = \exp(2.933 + 0.2937X_3)$$

The regression coefficients were interpreted in terms of the incidence rate ratio (IRR), which was obtained by exponentiating each regression coefficient. The unemployment rate (X_3) had an IRR of 1.341, indicating that a one-unit increase in the unemployment rate was associated with a 1.341-fold increase, or approximately a 34.1% increase, in the expected number of HIV/AIDS cases, holding all other variables constant. In contrast, the population density (IRR = 1.000000084) and poverty rate (IRR = 0.9737) were not interpreted further because neither variable was statistically significant in the partial significance test ($p > 0.05$).

5. CONCLUSION

This study demonstrates that negative binomial regression is more appropriate than Poisson regression for modeling the number of HIV/AIDS cases in West Java Province because it effectively accounts for the overdispersion observed in the data. Based on comparisons of the AIC and BIC, the negative binomial regression model including the open unemployment rate was selected as the best-fitting model, with AIC and BIC values of 301.5577 and 305.4452, respectively. The results indicate that the open unemployment rate was significantly associated with the number of HIV/AIDS cases, with an incidence rate ratio (IRR) of 1.341. This indicates that a one-unit increase in the unemployment rate was associated with an approximately 34.1% increase in the expected number of HIV/AIDS cases, holding the other variables constant. In contrast, population density and the percentage of people living in poverty were not statistically significant.

REFERENCES

- [1] Dinas Kesehatan Jawa Barat, “Profil kesehatan jawa barat tahun 2024,” Dinas Kesehatan Jawa Barat, Jawa Barat, Tech. Rep., 2024.
- [2] M. H. Ziaulhaq and A. Affandi, “Analisis faktor sosial terhadap jumlah kematian pasien hiv di jawa barat: Regresi poisson dan binomial negatif,” *EQUIVA J. Math. Inf. Technol.*, vol. 2, no. 2, 2024, <https://journal.itk.ac.id/index.php/equiva/article/view/1250/698>.
- [3] A. Ramadan, N. Chamidah, and I. N. Budiantara, “Modelling the number of hiv cases in indonesia using negative binomial regression based on least square spline estimator,” *Commun. Math. Biol. Neurosci.*, vol. 2024, pp. 1–16, 2024.

- [4] C. T. Fatmala, M. Hayati, R. P. Sari, and M. Hudori, “Pemodelan jumlah kasus hiv/aids di provinsi lampung menggunakan regresi binomial negatif,” *JOMTA J. Math. Theory Appl.*, vol. 6, no. 2, pp. 168–177, 2024.
- [5] R. Miju and A. Prasetyo, “Geographically weighted negative binomial regression (gwnbr) modeling to analyze factors affecting the number of new hiv/aids cases on the island of java in 2019,” in *Seminar Nasional Official Statistics*, 2021, pp. 858–867.
- [6] S. Priyastiwi and I. D. Ratih, “Pemodelan kasus baru hiv di jawa barat menggunakan poisson inverse gaussian regression,” *J. Ris. dan Apl. Mat.*, vol. 9, no. 1, pp. 114–128, 2025.
- [7] Tendriyawati, G. N. A. Wibawa, and B. Abapihi, “Pemodelan regresi poisson terhadap faktor-faktor yang mempengaruhi terjadinya hipertensi di kota kendari,” *J. Mat. Komputasi dan Stat.*, vol. 3, no. 1, pp. 255–262, 2023.
- [8] C. R. Oktarina and R. Pahlepi, “Modelling the number of stunting cases in indonesia in 2022 using negative binomial regression to address overdispersion,” *Math. J. Model. Forecast.*, vol. 2, no. 2, pp. 20–29, 2024.
- [9] P. R. Arum, I. M. Nur, A. J. Syafiqoh, and H. R. Utami, “Permodelan jumlah kasus tuberkulosis di kabupaten purbalingga tahun 2022 menggunakan regresi binomial negatif,” *J. Data Insights*, vol. 1, no. 2, pp. 44–50, 2023.
- [10] R. I. Kelirey, I. M. Hutabarat, and I. Chairani, “Penerapan regresi binomial negatif dalam menganalisis faktor-faktor yang mempengaruhi kasus angka kematian ibu di indonesia tahun 2022,” *J. Sunan Gunung Djati State Islam. Univ.*, vol. 10, no. 1, pp. 98–108, 2025.
- [11] A. Aswi *et al.*, “Evaluasi performa model regresi poisson tweedie dan conway maxwell poisson dalam menangani masalah dispersi: Studi angka kematian ibu di provinsi sulawesi selatan,” *Indones. J. Fundakental Sci.*, vol. 11, no. 2, pp. 122–134, 2025.
- [12] N. I. Fitriyah, P. R. Arum, and R. Wasono, “Pemodelan hiv dan aids di provinsi jawa timur menggunakan metode regresi bivariat poisson invers gaussian (bpig),” in *Pros. Semin. Nas. UNIMUS*, vol. 7, 2024, pp. 226–235.
- [13] Irwan, M. I. Nawawi, and Rezeki, “Model kasus hiv dan aids di sulawesi selatan dengan regresi poisson bivariat,” *Teknosains Media Inf. dan Teknol.*, vol. 17, no. 3, pp. 270–276, 2023.
- [14] B. Louis, T. Priharyanto, and U. Alifah, “Pemodelan jumlah kekerasan terhadap perempuan di jawa timur dengan regresi poisson dan binomial negatif,” *JSN J. Sains Nat.*, vol. 2, no. 3, pp. 68–77, 2024.
- [15] R. Esra, D. A. Nohe, and M. Fathurahman, “Pemilihan model terbaik pada generalized poisson regression menggunakan akaike information criterion,” *Stat. J. Theor. Stat. Its Appl.*, vol. 23, no. 1, pp. 73–87, 2023.
- [16] P. Pendi, “Analisis regresi dengan metode komponen utama dalam mengatasi masalah multikolinearitas,” *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 10, no. 1, pp. 131–138, 2021.
- [17] D. N. Rahma, R. A. K. Dewi, and A. Nanda, “Uji multikolinearitas: Pengaruh pembangunan infrastruktur terhadap pengurangan pengangguran kab. sidoarjo,” *J. Edukasi Ekon.*, vol. 8, no. 1, pp. 24–35, 2023.
- [18] J. P. K. P, E. Suratno, M. Kadar, R. Naibaho, S. K. Hati, and V. Aryati, “Penerapan uji multikolinieritas dalam penelitian manajemen sumber daya manusia,” *J. Ilm. Manaj. dan Kewirausahaan, Jumanage*, vol. 1, no. 2, pp. 94–102, 2022, <https://ejournal.unama.ac.id/index.php/jumanage>.
- [19] M. M. Majore, D. T. Salaki, and J. D. Prang, “Penerapan regresi binomial negatif dalam mengatasi overdispersi regresi poisson pada kasus jumlah kematian ibu,” *J. Mat. dan Apl.*, vol. 9, no. 2, pp. 133–139, 2020.
- [20] W. Kusuma, C. F. Utomo, S. Tervia, and R. N. S. Setiawan, “Pemodelan kasus tuberkulosis (tb) di nusa tenggara barat menggunakan model regresi binomial negatif,” *J. Serunai Mat.*, vol. 14, no. 2, pp. 142–147, 2022.

- [21] N. Nabilah and A. Sofro, “Negative binomial modeling of district-level diarrhea incidence in west java province,” *CAUCHY – J. Mat. Murni dan Apl.*, vol. 11, no. 2, pp. 1451–1471, 2026.
- [22] A. Fajri, N. Rahmi, P. A. Maharani, and M. I. Amal, “Modeling tuberculosis in children under five using poisson and negative binomial regression,” *Desimal J. Mat.*, vol. 7, no. 2, pp. 311–320, 2024.
- [23] Badan Pusat Statistik, *Provinsi Jawa Barat dalam Angka*. Jawa Barat: Badan Pusat Statistik, 2024.
- [24] J. F. S. Berek and R. D. Guntur, “Pemodelan generalized poisson regression (gpr) terhadap jumlah kasus penyakit tuberculosis di provinsi nusa tenggara timur,” *J. Sains Mat. dan Stat.*, vol. 11, no. 1, p. 127, 2025.
- [25] G. P. Silva, H. A. Laureano, R. R. Petterle, P. J. R. Júnior, and W. H. Bonat, “Multivariate generalized linear mixed models for count data,” *Austrian J. Stat.*, vol. 53, no. 1, pp. 44–69, 2024.
- [26] S. Appelbaum, J. Stronski, U. Konerding, and T. Ostermann, “The use of double poisson regression for count data in health and life science—a narrative review,” *Stats*, vol. 8, no. 4, pp. 0–12, 2025.
- [27] A. Ramadan, N. Chamidah, I. N. Budiantara, N. R. A. A. Siregar, and D. Aydin, “Modeling pulmonary tuberculosis case based on hiv and aids cases in indonesia using negative binomial regression least square spline,” *Stat. Optim. Inf. Comput.*, vol. 16, no. 1, pp. 247–259, 2026.
- [28] N. Nisa *et al.*, “Comparison of poisson and negative binomial regression models in identifying factors influencing covid-19 deaths in indonesia,” *Int. J. Quant. Res. Model.*, vol. 6, no. 4, pp. 437–442, 2026.
- [29] A. A. Hida, R. R. Robby, and R. Akbarita, “Penanganan overdispersi pada faktor-faktor yang mempengaruhi stunting di kabupaten blitar menggunakan regresi binomial negatif,” in *Pros. SENKIM Semin. Nas. Karya Ilm. Multidisiplin*, vol. 2, no. 1, 2022, pp. 87–94.
- [30] M. Bonamente, “Systematic errors in the maximum-likelihood regression of poisson count data: Introducing the overdispersed χ^2 distribution,” *Mon. Not. R. Astron. Soc.*, vol. 522, no. 2, pp. 1987–2001, 2023.