Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:1
https://doi.org/10.28919/cmbn/9642
ISSN: 2052-2541

MODELING HIV AND AIDS DATA IN TRENGGALEK AND PONOROGO WITH A BIVARIATE ZERO-INFLATED POISSON APPROACH

BAMBANG WIDJANARKO OTOK[1], SEKARSARI UTAMI WIJAYA[1,2,*], IRMA HARLIANINGTYAS[1,3]
[1]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Department of Logistics Engineering, Universitas Internasional Semen Indonesia, Gresik, Indonesia
[3]Department of Agricultural, Politeknik Negeri Jember, Indonesia

Abstract: The presence of excess zeros and overdispersion in count data often leads to biased parameter estimates when analyzed using the standard Poisson regression model. This study aims to model the number of HIV and AIDS cases in Trenggalek and Ponorogo Regencies using the Bivariate Zero-Inflated Poisson (BZIP) regression approach. The BZIP model accommodates correlated count responses as well as excess zeros, commonly found in epidemiological data. Two response variables, the number of HIV cases (Y_1) and the number of AIDS cases (Y_2), were analyzed against five explanatory variables: the percentage of the population aged 25–29 years (X_1), low education level (X_2), condom use among couples of reproductive ages (X_3), participation in health education programs (X_4), and community health insurance coverage (X_5). Parameter estimation was performed using the Expectation-Maximization (EM) algorithm. The results show that health education significantly increases the likelihood that an area has no HIV cases, while health insurance significantly reduces the number of AIDS cases. Moreover, individuals aged 25–29 years were identified as the group most at risk for AIDS. The model also confirmed strong overdispersion and zero inflation, supporting the use of BZIP as a more appropriate model than the standard bivariate Poisson regression. Additionally, the BZIP model achieved the best performance, indicated by the lowest AIC value compared to previous models.

Keywords: EM algorithm; bivariate zero-inflated Poisson regression; HIV/AIDS; Poisson distribution; zero-inflated Poisson distribution.

2020 AMS Subject Classification: 62J12.

---

[*]Corresponding author
E-mail address: sekarsari.wijaya@uisi.ac.id
Received October 15, 2025

1. INTRODUCTION

The Poisson distribution is a probability distribution that represents the number of successes occurring within a given interval of time, such that the possible values of the response variable are non-negative integers. In Poisson regression, the expected value (mean) is equal to the variance (a condition known as equidispersion). The Poisson model is often used because of its simplicity and its natural relationship to counting processes; however, this equidispersion assumption is frequently violated in real-world applications. Overdispersion occurs when the data variance is much greater than the mean. High variance is often caused by population heterogeneity, clustering, or the presence of many zero values (extra zeros) [1]. If overdispersion is not properly addressed, parameter estimates and confidence intervals may become biased and misleading. Modeling approaches used to overcome the problem of overdispersion involve combining the Poisson distribution with other discrete or continuous distributions, known as mixed Poisson distributions. Several mixed Poisson distributions that have been developed include the Zero-Inflated Poisson (ZIP), Generalized Poisson, Negative Binomial Poisson, and Poisson Inverse Gaussian distributions [2].

The zero-inflated regression model is a highly useful tool for analyzing count data that exhibits an excess of zeros. This model explains the presence of excess zeros by combining a degenerate distribution at zero with a standard count regression model (such as Poisson, binomial, or negative binomial for a single response variable, and bivariate Poisson or bivariate negative binomial for multiple response variables) [3]. A degenerate distribution is a type of probability distribution in which all the probability mass is concentrated at a single point. This distribution is used to represent the portion of data that is structurally zero-zero that do not occur due to random variation, but because the event in question cannot occur. Count data often contain excess zeros, and ignoring this condition may lead to inaccurate analytical results. The combination of a zero-inflated regression model and the Poisson model can be used to address excess zeros, resulting in the Zero-Inflated Poisson (ZIP) regression model. The ZIP model consists of two components (states): the Poisson state, which represents nonzero count data in the response variable, and the logit state, which represents zero count data in the response variable [4]. Many studies have applied and developed the ZIP model in various fields, such as manufacturing, epidemiology, public health, and others. The modeling approach used to address the problem of overdispersion involves combining the Poisson distribution with other discrete or continuous distributions, known as mixed Poisson distributions. If there are two interrelated response variables (for example, two types of injuries or two types of events), a bivariate model is required to simultaneously model the correlation between the responses and handle the presence of extra zeros. The Bivariate Zero-Inflated Poisson (BZIP) model allows for covariate modeling across both margins while capturing the sources of correlation arising from the Poisson component and/or the zero-inflation proportion. The multivariate ZIP model was first introduced by [5] to analyze events in manufacturing processes that involve several types of rare defects. Several other important BZIP studies include [6] introduced the BZIP model to analyze occupational injuries and demonstrated an improvement in model fit compared to univariate models; [7] developed numerical tests and score tests to detect zero inflation in the BZIP model; [8] applied a Bayesian (MCMC) approach to BZIP for blood donor data; and more recent methodological studies continue to emerge (e.g., handling missing covariates, MLE/Bayesian estimation).

Human Immunodeficiency Virus (HIV) and Acquired Immuno-Deficiency Syndrome (AIDS) are serious health conditions that must be well understood by the public. HIV is a virus that weakens the immune system, while AIDS is a condition in which the immune system becomes severely compromised due to HIV infection. Indonesia currently ranks 14th in the world in terms of the number of people living with HIV (PLHIV) and 9th for new HIV infections. It is estimated that there are approximately 564,000 PLHIV in Indonesia, yet only about 63% are aware of their status. According to data from the Indonesian Ministry of Health for the period January–March 2025, there were 15,382 HIV-AIDS cases reported, consisting of 4,850 AIDS cases and 10,532 HIV cases. Indonesian Ministry of Health reported that the majority of PLHIV identified during this period belonged to four key groups: Men who have sex with men (MSM): 4,716 cases; General population: 3,931 cases; Tuberculosis (TB) patients: 2,152 cases; Clients of sex workers: 1,206 cases. East Java was the province with the highest number of newly detected HIV cases during January–March 2025 [9].

In Indonesia, the spread of HIV and AIDS initially occurred predominantly among commercial sex workers (CSWs) and homosexual groups. The practice of having multiple sexual partners among CSWs became the main factor in the transmission of the virus, which was later passed on to their clients and subsequently to housewives within their communities. Transmission can also occur from an infected mother to her newborn child. In recent years, the pattern of HIV and AIDS transmission in Indonesia has been dominated by heterosexual contact and the use of narcotics, psychotropics, and addictive substances (NAPZA) through shared needles, where injecting drug users have the potential to transmit the virus to their partners. The widespread transmission of HIV and AIDS has had serious implications for national development, as most people living with the disease are within the productive age group. The negative impacts are not limited to health aspects but also extend to social and economic conditions within society [10].

The data on the number of HIV and AIDS cases in Trenggalek and Ponorogo Regencies are count data, representing the number of occurrences within a specific time period. This type of data generally follows a Poisson distribution because it is discrete, non-negative, and represents the number of rare events within a given unit of time or area [4]. Furthermore, HIV and AIDS cases in both regencies are interrelated, as an increase in the number of HIV cases tends to be followed by an increase in AIDS cases. Therefore, the modeling process involves two response variables (bivariate modeling). Consequently, the appropriate distribution to describe the relationship between these two variables is the bivariate Poisson distribution, which is capable of capturing the positive correlation between two types of related events [11]. The use of two separate regression models to estimate jointly related events will result in inconsistent and inefficient estimators [12]. However, the data on the number of HIV and AIDS cases in Trenggalek and Ponorogo Regencies show a high proportion of zero values, indicating that many subdistricts have no HIV or AIDS cases at all. This condition reflects the presence of excess zeros, which violates the equidispersion assumption in the bivariate Poisson regression model. As a result, the bivariate Poisson regression model becomes inadequate, leading to biased parameter estimates and poor model accuracy. To address this issue, the appropriate modeling approach is the Bivariate Zero-Inflated Poisson (BZIP) regression. This study proposes the use of the BZIP regression model to determine the significance of factors that may influence the number of HIV and AIDS cases.

2. PRELIMINARIES

2.1 Bivariate Poisson Distribution

The discrete random variables Y_1 and Y_2 follow a bivariate Poisson distribution, denoted as $(Y_1, Y_2) \sim BP(\lambda_1, \lambda_2, \lambda_3)$ if their joint probability mass function (pmf) for Y_1 and Y_2 is given as follows.

$$P(Y_1 = y_1, Y_2 = y_2) = \begin{cases} \phi + (1-\phi)\mathrm{e}^{-(\lambda_1+\lambda_2+\lambda_3)}, (y_1, y_2) = (0,0) \\ (1-\phi)\mathrm{e}^{-(\lambda_1+\lambda_2+\lambda_3)} \sum_{j=0}^{min(y_1,y_2)} \frac{\lambda_1^{y_1-j}\lambda_2^{y_2-j}\lambda_3^{j}}{(y_1-j)!\,(y_2-j)!\,j!}, (y_1, y_2) \neq (0,0) \end{cases} \quad (1)$$

And $0 < \phi < 1$ represents the additional proportion in the zero-zero cell. Furthermore, the correlation coefficient between Y_1 and Y_2 in equation (1) given by :

$$\mathrm{Corr}(Y_1, Y_2) = \frac{\lambda_3 + \phi(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}{\sqrt{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)[1+\phi(\lambda_1+\lambda_3)][1+\phi(\lambda_2+\lambda_3)]}}. \quad (2)$$

From equation (2) there are two sources that can generate correlation between Y_1 and Y_2. The first is λ_3, which represents the pure covariance parameter of the bivariate Poisson (BP) distribution. The second is ϕ, which represents the additional (inflated) proportion in the zero–zero cell. Therefore, $Y_1$and Y_2 cannot be independent unless $\phi = 0$ and $\lambda_3 = 0$.

2.2 Bivariate Zero Inflated Poisson Distribution

The bivariate Poisson model was proposed by [13] dan and presented by Johnson and Kotz (1969) in [14]. This model is used to model two correlated count variables. Let there be a probability space $(\Omega, \mathcal{C}, \mathbb{P})$. Consider three random variables U_1, U_2 and U_0 each following independent Poisson distributions with parameters λ_1, λ_2, and λ_0 respectively. Then the random variables

$$Y_1 = U_1 + U_0 \text{ dan } Y_2 = U_2 + U_0 \quad (3)$$

jointly follow a bivariate Poisson distribution, denoted $BP(\lambda_1, \lambda_2, \lambda_0)$. Equation (4) is the probability mass function of the Bivariate Zero-Inflated Poisson (BZIP) distribution.

$$P(Y_k = y_k) = \begin{cases} \phi + (1-\phi)\mathrm{e}^{-(\lambda_k+\lambda_0)}, y_k = 0. k = 1,2 \\ (1-\phi)\mathrm{e}^{-(\lambda_k+\lambda_0)} \frac{(\lambda_k+\lambda_0)^{y_k}}{y_k!}, y_k \neq 0, k = 1,2 \end{cases} \quad (4)$$

The moments of the BZIP distribution can be shown as follows:

$$\begin{aligned} E(Y_k) &= (1-\phi)(\lambda_k+\lambda_0) \\ \mathrm{Var}(Y_k) &= E(Y_k)[1+\phi(\lambda_k+\lambda_0)] \\ E(Y_1Y_2) &= (1-\phi)[(\lambda_1+\lambda_0)(\lambda_2+\lambda_0)+\lambda_0] \\ \mathrm{Cov}(Y_1, Y_2) &= (1-\phi)[\lambda_0+\phi(\lambda_1+\lambda_0)(\lambda_2+\lambda_0)] \end{aligned} \quad (5)$$

2.3 The Model Regression of BZIP

The regression model aims to (a) identify significant risk factors that influence the occurrence of zero inflation, and (b) determine the extent to which interventions or other confounders affect the mean number of events. For purpose (a), the Poisson component is held fixed (constant) with parameters $\beta_0, \beta_1,$ and β_2. The zero-inflation parameter ϕ is modelled as a logistic function of the covariates Z:

$$\log\left(\frac{\phi_i}{1-\phi_i}\right) = \xi_i = \boldsymbol{Z}_i^T \alpha \qquad (6)$$

The Poisson mean component is defined as:

$$\log(\lambda_{k,i}) = \eta_{k,i} = \beta_k + \log(t_i) \qquad (7)$$

with $k = 0,1,2;$ $i = 1, \dots, n;$ t_i represents the exposure term (time or risk measure).

For purpose (b), ϕ is considered fixed. However, the Poisson (mean) component is assumed to be related to the covariates X, as follows:

$$\log\left(\frac{\phi_i}{1-\phi_i}\right) = \xi_i = \alpha$$

$$\log(\lambda_{k,i}) = \eta_{k,i} = \boldsymbol{X}_{k,i}^T \beta_k + \log(t_i)$$

$$\log(\lambda_{k,i}) = \eta_{k,i} = \beta_0 + \log(t_i)$$

where X_1adn X_2 are design matrices that include the risk factors, β_1 and β_2 are vectors of regression coefficients, and β_0 is an unknown constant representing the common mean component. Similar to the univariate ZIP model, specifying the full BZIP model with identical covariates in all Poisson and logistic components may lead to parameter identifiability issues. Therefore, the complete model should be divided into two submodels: one for analyzing zero inflation and another for analyzing the mean count of events.

2.4 Expectation Maximization Algorithm

To obtain the maximum likelihood estimates (MLE) of the model parameters, the Expectation–Maximization (EM) algorithm as described in [15] is used. Let the complete data be denoted by U_1, U_2, U_0, V, where V is a latent variable indicating whether an observation originates from the latent zero class ($V = 1$) or the nonzero class ($V = 0$), meanwhile, U_0 is also unobserved and represents the shared component between $Y_1$and Y_2. Although the complete-data log-likelihood function is mathematically complex, it can be expressed as a linear function of U_0 and V. Since U_0 and V are conditionally independent, the E-step in the EM algorithm is performed by replacing U_0 and V with their conditional expected values based on the current parameter estimates. The M-step is then conducted by partitioning the log-likelihood function into four orthogonal components, allowing parameter estimation to be performed separately through logistic regression (for the zero-inflated component) and three weighted Poisson regressions (for the bivariate Poisson components). The asymptotic standard deviations of the regression coefficients are obtained using the method described in [15]. The details of the estimation procedure are provided as follows.

$$l_C = l_\xi + l_{\eta_1} + l_{\eta_2} + l_{\eta_0}$$

constitute the complete data that form (build) the log-likelihood.

$$l_\xi = \sum_{i=1}^{n} v_i \xi_i - log[1 + \exp(\xi_i)]$$

$$l_{\eta_k} = \sum_{i=1}^{n} (1 - v_i)\left[(y_{k,i} - u_{0,1})\eta_{k,i} - \exp(\eta_{k,i}) - \log\left((y_{k,i} - u_{0,i})!\right)\right], k = 1,2$$

$$l_{\eta_0} = \sum_{i=1}^{n} (1 - v_i)\left[u_{0,i}, \eta_{0,1} - \exp(\eta_{k,i}) - \log(u_{0,i}!)\right]$$

$$v_i^{(m)} = \begin{cases} \dfrac{1}{1 + \exp\left(-\xi_i^{(m)} - exp\left(\eta_{1,i}^{(m)}\right) - exp\left(\eta_{2,i}^{(m)}\right) - exp\left(\eta_{0,i}^{(m)}\right)\right)}, jika\ (y_{1,i,}y_{2,i}) = (0,0) \\ 0, jika\ (y_{1,i,}y_{2,i}) \neq (0,0) \end{cases}$$

and

$$u_{0,i}^{(m)} = E(U_{0,i}|y_{1,i}, y_{2,i})\ dimana\ P\ (U_{0,i} = j|y_{1,i}, y_{2,i})$$

$$= \frac{P(Y_{1,i} = y_{1,i}, y_{2,i,}U_{0,i} = j)}{\sum_{r=0}^{\min(y_{1,i,}y_{2,i})} P(Y_{1,i} = y_{1,i}, y_{2,i,}, U_{0,i} = r)};\ j = 0, \dots, \min(y_{1,i,}, y_{2,i})$$

where

$$P(Y_{1,i} = y_{1,i}, Y_{2,i,} = y_{2,i,}, U_{0,i} = j) = \frac{\lambda_1^{y_{1,i-j}} \lambda_2^{y_{2,i-j}} \lambda_0^{y_{0,i-j}}}{(y_{1,i-j} - j)!\,(y_{2,i-j} - j)!\,j!} \exp(-\lambda_1 - \lambda_2 - \lambda_0)$$

The asymptotic variance of the parameters is obtained from the observed information matrix through the EM estimation procedure. The observed information matrix can be expressed as the conditional expectation of the gradient vector of the complete-data log-likelihood function, evaluated at the maximum likelihood estimates (MLE)[16]. Let $\Psi = (\alpha^{\mathrm{T}}, {\beta_1}^{\mathrm{T}}, {\beta_2}^{\mathrm{T}}, \beta_0)^{\mathrm{T}}$ it is known that

$$l_C = \sum_{i=1}^{n} l_C\left(U_{1,i}, U_{2,i}, U_{0,i}; V_i, \psi\right) = \sum_{i=1}^{n} l_C\left(Y_{1,i}, Y_{2,i}, U_{0,i}; V_i, \psi\right)$$

which indicates that

$$S_i(\psi;\ y_{1,i,}y_{2,i}) = E\left[\frac{\partial l_C(U_{1,i}, U_{2,i}, U_{0,i}; V_i, \psi)}{\partial\psi}\middle| y_{1,i,}y_{2,i}\right] = \frac{\partial l_C(y_{1,i}, y_{2,i}, \hat{u}_{0,i}; \hat{v}_i, \psi)}{\partial\psi}$$

where $\hat{v}_i$ and $\hat{u}_{0,i}$ denote the corresponding final estimates obtained in the E-step. Hence, the observed information matrix is given by

$$I_{obs} = \sum_{i=1}^{n} S_i(\psi;\ y_{1,i,}y_{2,i}) S_i(\hat{\psi};\ y_{1,i,}y_{2,i})^T$$

The asymptotic standard error of the parameter estimates $\hat{\psi}$ the BZIP regression model is obtained from the inverse of the observed information matrix, denoted as I_{obs}^{-1}.

2.5 Case Study

The data used in this study consist of HIV and AIDS case records obtained from the Health Offices of Trenggalek and Ponorogo Regencies for the year 2012. There is a total of 35 subdistricts, comprising 14 subdistricts in Trenggalek Regency and 21 subdistricts in Ponorogo Regency. This study aims to model the effects of health awareness and healthcare services on the tendency of individuals to contract HIV/AIDS. The cases of people living with HIV and those with AIDS are interdependent, as the occurrence of AIDS is highly dependent on HIV status. The data on the number of new HIV and AIDS cases in Trenggalek and Ponorogo Regencies in 2012 follow a Poisson distribution. However, the data exhibits a very high level of overdispersion due to the presence of many zero values 57% of subdistricts have zero HIV cases and 34% have zero AIDS cases. Therefore, the appropriate modeling approach used is the Bivariate Zero-Inflated Poisson (BZIP) regression. There are two response variables, $Y_1$and Y_2, and five explanatory variables, X_1, X_2, X_3, X_4,and X_5, used in this study. Specifically:

- Y_1: Number of HIV cases
- Y_2: Number of AIDS cases
- X_1: Percentage of population aged 25–29 years
- X_2: Percentage of population with low education level
- X_3: Percentage of couples of reproductive age using condoms
- X_4: Percentage of population participating in health education programs
- X_5: Percentage of population covered by the Community Health Insurance (Jamkesmas) program

2.6 Data Analysis

The stages of data analysis (figure 1) conducted in this study are as follows:

1. Data Collection: Collect data on HIV and AIDS cases, along with supporting variables, age of patients (X_1), education level (X_2), percentage of couples of reproductive age using condoms (X_3), percentage of the population participating in health education programs (X_4), and percentage of the population covered by the Community Health Insurance (Jamkesmas) program (X_5), as recorded by the Health Offices of Trenggalek and Ponorogo Regencies.
2. Descriptive Analysis: Perform descriptive statistics to determine the standard deviation, skewness, minimum and maximum values, and the percentage of zero occurrences for each subdistrict in Trenggalek and Ponorogo.
3. Correlation Test of Response Variables: Examine the correlation between the response variables, namely the number of HIV and AIDS cases (Y_1 and Y_2) in Trenggalek and Ponorogo.
4. Dispersion Test: Evaluate the dispersion of the response variables (Y_1 and Y_2) to assess whether the Poisson assumption of equidispersion holds.
5. Distribution Testing: Test the distribution of the response variables to determine whether the HIV and AIDS case data follow the Poisson or Zero-Inflated Poisson distribution.
6. Multicollinearity Test: Assess the multicollinearity among the explanatory variables (X_1, X_2, X_3, X_4, X_5).
7. Parameter Estimation: Estimate the model parameters using the Bivariate Zero-Inflated Poisson (BZIP) regression model through the Expectation-Maximization (EM) algorithm approach.
8. Model Selection: Select the best model based on the Akaike Information Criterion (AIC) as the indicator of model fit.
9. Model Construction: Develop and formulate the BZIP regression equations.
10. Model Interpretation: Interpret the regression model to explain the relationship between the explanatory variables and the number of HIV and AIDS cases.

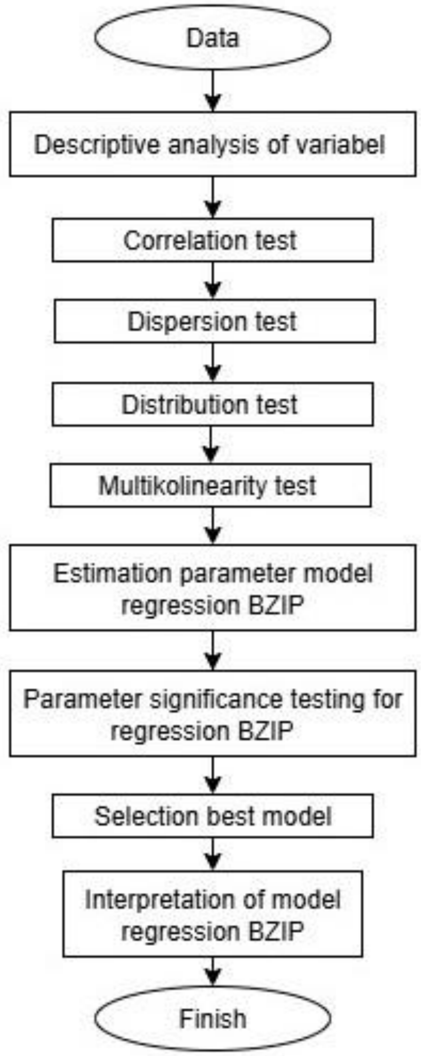

Figure 1 Flowchart of research

3. MAIN RESULTS

3.1 Description of Response and Explanatory Variables

Table 1 resents the descriptive statistics of the two response variables. The average number of new HIV cases per subdistrict in Trenggalek and Ponorogo Regencies is one person per year. The standard deviation of 1.278 indicates that the number of new HIV cases does not vary greatly across subdistricts. Furthermore, since the standard deviation is greater than the mean, the data exhibits overdispersion. The highest number of new HIV cases, totaling four individuals, occurred in Kampak and Bendungan subdistricts in Trenggalek Regency. Similarly, the average number of new AIDS cases per subdistrict in Trenggalek and Ponorogo Regencies is one person per year.

The standard deviation of 1.173 suggests that the number of new AIDS cases also does not differ substantially among subdistricts. As with the HIV data, the standard deviation being greater than the mean indicates the presence of overdispersion in the AIDS case data.

Table 1 Descriptive Statistics of HIV and AIDS Cases

Response Variable	Standard Deviation	Minimum	Maximum	Skewness	Percentage of Zeros
Number of HIV cases (Y1)	1.279	0	4	1.39	57%
Number of AIDS cases (Y2)	1.173	0	5	1.57	34%

Table 2 shows that the average percentage of the population aged 25–29 years in each subdistrict is 6.798%. The standard deviation of 0.513% indicates that the percentage of residents aged 25-29 does not vary greatly across subdistricts. Watulimo Subdistrict has the lowest percentage, at 5.771%, while Trenggalek Subdistrict has the highest, at 7.821%. The average percentage of the population with a low education level in each subdistrict is 14.784%. The standard deviation of 5.783% indicates that some subdistricts differ considerably from others. Pudak Subdistrict has the lowest percentage, at 6.491%, while Trenggalek Subdistrict has the highest percentage, at 34.954%.

Table 2 Descriptive Statistics Explanatory variables

Response Variable	Mean	Standard Deviation	Mini-mum	Maxi-mum
Percentage of population aged 25–29 years (X1)	6.798	0.513	5.711	7.821
Percentage of population with low education level (X2)	14.784	5.783	6.491	34.954
Percentage of couples of reproductive ages (PUS) using condoms (X3)	2.894	2.607	0.153	13.313
Percentage of population participating in health education programs (X4)	0.671	0.459	0.041	2.512
Percentage of population covered by the Community Health Insurance (X5)	40.510	11.000	16.660	60.980

The average percentage of couples of reproductive age (PUS) using condoms in each subdistrict is 2.894%. The standard deviation of 2.607% indicates that there are considerable differences among subdistricts. Ngebel Subdistrict has the lowest percentage, at 0.513%, while Trenggalek Subdistrict has the highest percentage, at 13.313%. The average percentage of the population participating in health education programs in each subdistrict is 0.671%. The standard deviation of 0.459% shows that the percentages do not vary greatly between subdistricts, with the lowest value of 0.041% in Gandusari Subdistrict and the highest value of 2.512% in Pudak Subdistrict. The average percentage of the population covered by the Community Health Insurance

(Jamkesmas) program in each subdistrict is 40.51%. The standard deviation of 11% indicates a considerable difference across subdistricts. Siman Subdistrict has the lowest coverage, at 16.66%, while Pudak Subdistrict has the highest, at 60.98%.

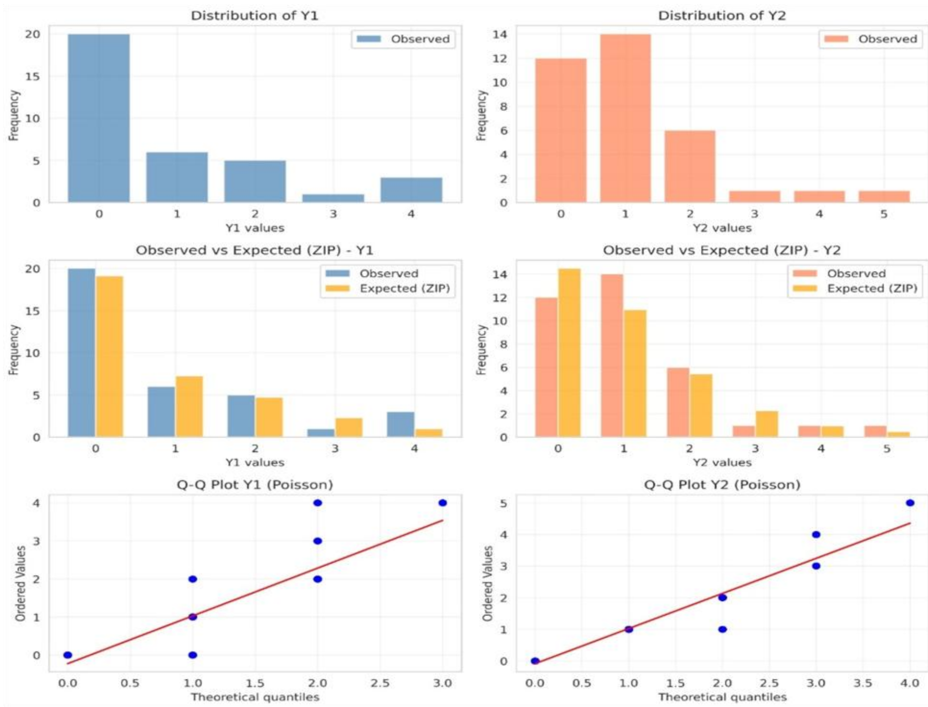

Figure 2 Graph of the distribution of Y_1 and Y_2 (top), observed vs expected Y_1 and Y_2 (middle), and Q-Q plot Y_1 dan Y_2 (bottom)

Figure 2 (top) shows the empirical (observed) distribution of Y_1 and Y_2. For Y_1 (HIV) most observations have a value of zero (around 20 cases), indicating the presence of zero inflation or areas with no HIV cases. The Y_2 (AIDS) data exhibit a right-skewed distribution, but the number of zeros is smaller than in Y_1, suggesting that AIDS cases are more widely distributed than HIV cases. Figure 2 (middle) presents the comparison between observed and expected values under the Zero-Inflated Poisson (ZIP) model. The graph demonstrates a strong agreement between observed and predicted frequencies for each variable. For Y_1, the ZIP model accurately captures the actual data pattern, especially for counts of zero and one, confirming that the model effectively accounts for excess zeros. For Y_2, the predicted ZIP distribution also closely follows the observed pattern, with only minor deviations for counts of one and two. Overall, these results indicate that the Bivariate Zero-Inflated Poisson (BZIP) model provides a good fit in describing the joint occurrence patterns of HIV and AIDS. Figure 2 (bottom) shows the Q–Q plots for Y_1 and Y_2 under the Poisson distribution assumption. The blue dots represent the ordered observed values, while the red line represents the theoretical quantiles. The points that closely follow the diagonal line suggest that the Poisson distribution assumption for the count process component is well satisfied. The slight deviations at the upper right tail indicate mild overdispersion, which has been successfully corrected by the zero-inflation component within the BZIP model.

3.2 Examination of Correlation Between Response Variables

In bivariate regression analysis, the response variables must exhibit a correlation. This study uses the number of new HIV cases (Y_1) and the number of new AIDS cases (Y_2) as the response variables. The correlation coefficient between the two response variables is 0.399, indicating a positive correlation, meaning that as the number of new HIV cases increases, the number of new AIDS cases also tends to increase. Conversely, when the number of new HIV cases decreases, the number of new AIDS cases also tends to decrease. The hypotheses tested are as follows:

H_0: There is no correlation between Y_1 and Y_2

H_1: There is correlation between Y_1 and Y_2

The t-value obtained is 2.5, which is greater than the critical value $t\alpha/2,33 = 2.034$ and the p-value is 0.018, which is smaller than the significance level α (0.05), Therefore, the decision is to reject H_0. In conclusion, there is a significant correlation between the number of new HIV and AIDS cases in Trenggalek and Ponorogo Regencies.

3.3 Overdispersion Diagnostic

Overdispersion testing is conducted to evaluate whether the count data follow the Poisson distribution assumption, which requires that the mean and variance are equal. A mismatch between the variance and mean indicates that the standard Poisson regression model may not be appropriate, and an alternative model such as the Zero-Inflated Poisson (ZIP) model should be considered.

Table 3 Overdispersion Diagnostic

Variable	Mean	Variance	Ratio
Y_1	0.8857	1.6336	1.8444
Y_2	1.0857	1.3748	1.2663

Table 3 As shown in Table 3, the mean of Y_1 is 0.8857 and the variance is 1.6336, producing a variance-to-mean ratio of 1.8444. This value is much greater than 1, and even exceeds the empirical threshold of 1.5, which is commonly used to detect significant overdispersion. Hence, Y_1 exhibits strong overdispersion. For Y_2 the mean is 1.0857 and the variance is 1.3748, resulting

in a variance-to-mean ratio of 1.2663, which is slightly above 1. This indicates mild overdispersion. Although not as pronounced as in Y_1, this result still suggests that the ZIP model (and later the BZIP model) is more suitable for handling the excess zeros and overdispersion present in the data.

3.4 Testing the Distribution of Response Variables

The hypothesis testing for the distribution of the response variables is divided into two complementary tests, each serving a distinct statistical purpose: (1) to determine whether the data exhibit excess zeros (zero inflation) that cannot be adequately explained by the standard Poisson distribution, and (2) to assess whether the data follow the Zero-Inflated Poisson (ZIP) distribution. If $\phi = 0$, then the joint probability density function of Y_1 and Y_2 becomes identical to that of the traditional bivariate Poisson (BP) regression model given by [17]. Therefore, the hypothesis test comparing the BP model with the Bivariate Zero-Inflated Poisson (BZIP) regression model can be formulated as follows:

$$\begin{aligned} &\mathrm{H}_0: \phi = 0 \ \text{(No zero-inflation)} \\ &\mathrm{H}_1: \phi > 0 \ \text{(Zero-inflation exists)} \end{aligned} \tag{8}$$

The alternative hypothesis in Equation (8) is one-sided, as the main interest in real data analysis lies in detecting the presence of zero inflation. To test this hypothesis, following [18], a reparameterization from ϕ to ψ, is performed as:

$$\psi = \frac{\phi}{1-\phi} \tag{9}$$

Thus, the hypothesis test in Equation (8) becomes equivalent to:

$$\begin{aligned} &\mathrm{H}_0: \psi = 0 \ \text{(No zero-inflation)} \\ &\mathrm{H}_1: \psi > 0 \ \text{(Zero-inflation exists)} \end{aligned} \tag{10}$$

To obtain the score test statistic for Equation (10), the log-likelihood function of the BZIP model based on n independent observations is derived as follows:

$$\begin{aligned} \log L = \sum_{i=1}^{n} l_i \ = \sum_{i=1}^{0} \Big[&-\log(1+\psi) + I_{(y_{1i}, y_{2i})=(0,0)} \log\big(\psi + e^{-\lambda_{1i}^* - \lambda_{2i}^* + \lambda_3}\big) - (1 \\ &- I_{(y_{1i}, y_{2i})=(0,0)})((\lambda_{1i}^* + \lambda_{2i}^* - \lambda_3) + \log(\gamma(y_{1i}, y_{2i})))\Big] \end{aligned} \tag{10}$$

where l_i is the logarithm of the probability function for the ke-i of (Y_{1i}, Y_{2i}), evaluated at the values $(Y_{1i}, Y_{2i}) = (y_{1i}, y_{2i})$, and $I_{(\cdot)}$ is an indicator function that equals 1 if the condition inside the parentheses is true, and 0 otherwise.

After testing for the presence of zero inflation, the next step is to perform a goodness-of-fit test to determine whether the data follow the Zero-Inflated Poisson (ZIP) distribution.

H0: The data follow the ZIP distribution
H1: The data do not follow the ZIP distribution (10)

Table 4 shows that for the variable Y_1, the score test for detecting zero inflation yields a Z-statistic of 7.9138 with a very small p-value (0.000001). This provides strong evidence to reject the null hypothesis at the 5% significance level, indicating the presence of significant zero inflation in Y_1, Similarly, for the variable Y_2, the Z-statistic is 2.2731 with a p-value of 0.0115, suggesting that although the zero inflation is not as strong as in Y_1, it is still statistically significant at the 5% level. Therefore, the standard Poisson model is inadequate to represent the data, and an alternative ZIP model is required to properly account for the large number of zero observations.

Table 4 Testing the Distribution of Response Variables

Respon Variable	Zero-Inflation		ZIP		
	Z-Statistic	P-value	Chi-Square	df	P-value
Y_1	7.9138	0.0000	3.3476	1	0.0673
Y_2	2.2731	0.0115	0.3421	1	0.5586

Furthermore, the results of the goodness-of-fit test indicate that the ZIP model provides an adequate fit to the data. For the variable Y_1, the Chi-square value of 3.3476 with a p-value of 0.0673 shows that there is insufficient evidence to reject the null hypothesis. This means that the ZIP model is considered adequately representative in describing the distribution of Y_1. For the variable Y_2, the Chi-square value is 0.3421 with a p-value of 0.5586, indicating no significant deviation between the model and the observed data. Therefore, the ZIP distribution is deemed to fit Y_2 very well. These results suggest that the use of the Bivariate Zero-Inflated Poisson (BZIP) model which extends the advantages of the ZIP model to the bivariate case is an appropriate approach for analyzing these two variables, given the evidence of significant zero inflation and satisfactory model fit.

The empirical (observed) and fitted frequency distributions of Y1 and Y2 under the BZIP model are presented in Figure 3. As expected, the zero–zero combination $(y_1, y_2) = (0,0)$ exhibits the highest frequency, representing subdistricts with no HIV or AIDS cases. The figure clearly shows that the BZIP distribution provides an adequate fit to the bivariate data and performs better overall. According to Wang et al. (2003), the BZIP model outperforms the bivariate Poisson, marginal

Poisson, and marginal ZIP distributions in terms of log-likelihood value, demonstrating its superior capability in handling correlated count data with excess zeros.

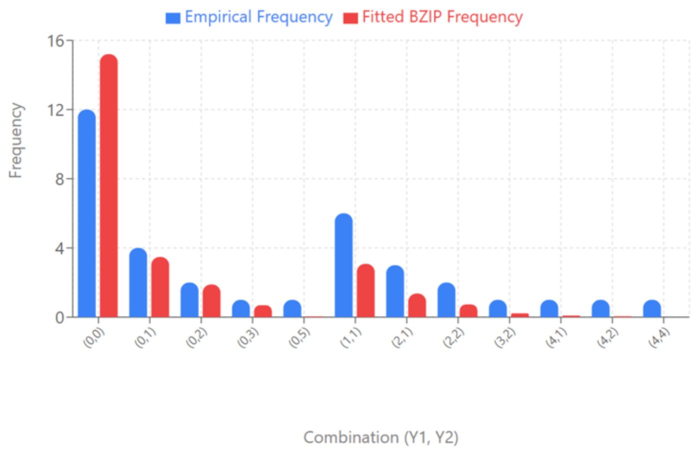

Figure 3 Empirical and Fitted Frequency Distribution of the BZIP Model

3.5 Multicollinierity Diagnostic

Multicollinearity testing is conducted to determine whether there is a correlation among the explanatory variables when considered simultaneously. One method to detect the presence of multicollinearity is by using the Variance Inflation Factor (VIF). If the VIF value exceeds 10, it indicates the presence of multicollinearity among the explanatory variables. Based on Table 5, the VIF values for each explanatory variable with respect to the others are less than 10, indicating that no multicollinearity exists among the explanatory variables. Therefore, all five explanatory variables can be appropriately used in the Bivariate Zero-Inflated Poisson (BZIP) regression model.

Table 5 VIF value of explanatory variables

Response Variable	Explanatory Variable	Rk2	VIF	Conclusion
X1	X2, X3, X4, X5	0.188	1.231	
X2	X1, X3, X4, X5	0.515	2.062	
X3	X1, X2, X4, X5	0.485	1.942	No multicollinearity
X4	X1, X2, X3, X5	0.062	1.066	
X5	X1, X2, X3, X4	0.172	1.208	

3.6 Modeling the Number of New HIV and AIDS Cases in Trenggalek and Ponorogo Regencies in 2012 Using the Bivariate Zero-Inflated Poisson (BZIP) Model

Table 6 Poisson parameter estimation for Y1

Parameter	Estimate	Standard Error	z-value	p-value
Intercept	-6.3896	3.4031	-1.8776	0.0604
X1	0.7022	0.4340	1.6181	0.1056
X2	0.0145	0.0359	0.4033	0.6867
X3	0.1751	0.1053	1.6634	0.0962
X4	0.0657	0.5328	0.1233	0.9019
X5	0.0295	0.0206	1.4320	0.1522

Based on Table 6, the estimation results for the Poisson component of the dependent variable representing the number of people living with HIV (Y_1) show that none of the explanatory variables have a statistically significant effect at the 5% significance level. However, several interesting tendencies can be observed. The coefficient for the percentage of the population aged 25–29 years (X is positive, at 0.702239, with a p-value of 0.1056, indicating a tendency that an increase in the proportion of individuals in this productive age group may lead to a higher number of HIV cases, although the effect is not statistically significant. The percentage of the population with a low education level (X_2) has a small and insignificant coefficient ($p = 0.6867$), suggesting that low education is not necessarily a major factor contributing to the increase in HIV cases within the study area. The percentage of couples of reproductive ages (PUS) using condoms (X_3) has a positive coefficient of 0.175126 with a p-value of 0.0962, which is close to the 10% significance threshold. This finding indicates that areas with higher condom use actually report more HIV cases, possibly due to greater awareness and better reporting resulting from more intensive prevention programs. The health education participation variable (X_4) and the community health insurance coverage variable (X_5) both show no significant effect on the number of HIV cases ($p > 0.05$).

Overall, the Poisson component for Y1 suggests that while none of the factors are statistically significant, the positive indication from the condom use variable (X_3) deserves further attention and investigation in future studies.

Table 7 Poisson parameter estimation for Y2

Parameter	Estimate	Standard Error	z-value	p-value
Intercept	1.5761	4.2264	0.3729	0.7092
X1	0.0267	0.5672	0.0470	0.9625
X2	-0.0094	0.0379	-0.2479	0.8042
X3	0.0709	0.0816	0.8690	0.3849
X4	0.0716	0.5812	0.1233	0.9019
X5	-0.0425	0.0182	-2.3382	0.0194

In Table 7 for the dependent variable representing the number of people living with AIDS (Y_2), most explanatory variables do not show statistically significant effects, except for one factor that exhibits a clear and meaningful influence. The community health insurance variable (X_5) has a negative coefficient of –0.042518 with a p-value of 0.0194, which is statistically significant at the 5% level. This finding indicates that an increase in the proportion of the population covered by community health insurance is associated with a decrease in the number of AIDS cases. In other words, the health insurance program likely provides better access to HIV/AIDS prevention and treatment services, thereby helping to reduce disease progression. Meanwhile, the variables representing the population aged 25–29 years (X_1), low education level (X_2), condom use among couples of reproductive ages (X_3), and participation in health education programs (X_4) do not have significant effects on AIDS cases ($p > 0.05$). Thus, these results suggest that community health insurance (X_5) plays an important role in controlling AIDS cases, while the other explanatory variables have not shown a statistically meaningful relationship in the Poisson component of the model.

Table 8 Estimation Parameter Zero-Inflation for Y_1

Parameter	Estimate	Standard Error	z-value	p-value
Intercept	-306.9160	216.9058	-1.4150	0.1571
X1	57.5542	71.2383	0.8079	0.4191
X2	-5.1381	33.3300	-0.1542	0.8775
X3	-92.4985	196.0869	-0.4717	0.6371
X4	209.6302	46.4640	4.5117	0.0000
X5	0.0591	14.2102	0.0042	0.9967

The estimation results presented in Table 8 illustrate how independent factors influence the likelihood of zero inflation, that is, areas with no HIV cases at all. The findings show that health education participation (X_4) is the only variable that has a statistically significant effect on the zero-inflation component, with a positive coefficient of 209.6302 and a p-value < 0.001. This indicates that the higher the percentage of the population participating in health education programs, the greater the probability that a region will have zero HIV cases. In other words, health education has proven to be effective in reducing the emergence of new HIV cases by promoting public awareness and preventive behavior. Meanwhile, other variables population aged 25-29 years (X_1), low education level (X_2), condom use (X_3), and community health insurance coverage (X_5) do not show significant effects ($p > 0.05$). Therefore, it can be concluded that health education plays a major role in explaining the probability of HIV-free regions, whereas other demographic and socioeconomic factors do not make a statistically significant contribution to zero inflation in Y_1.

Table 9 Estimation Parameter Zero-Inflation for Y_2

Parameter	Estimate	Standard Error	z-value	p-value
Intercept	5621.2511	341.5423	16.4584	0.0000
X1	-702.4752	89.4648	-7.8520	0.0000
X2	6.7695	65.1372	0.1039	0.9172
X3	-171.4635	288.4713	-0.5944	0.5523
X4	-777.5620	447.1320	-1.7390	0.0820
X5	-9.6843	11.9313	-0.8117	0.4170

Based on Table 9, the estimation results for the zero-inflation component of the dependent variable representing the number of people living with AIDS (Y_2) indicate that the population aged 25-29 years (X_1) is a significant factor affecting the probability of excess zeros. The coefficient value of –702.4752 with $p < 0.001$ shows that as the proportion of people aged 25-29 increases, the likelihood that a region will have zero AIDS cases decreases. This finding suggests that the 25-29 age group is particularly vulnerable to HIV/AIDS transmission. Additionally, the health education variable (X_4) shows a p-value of 0.0820, which is close to the 10% significance level, indicating that health education programs may also reduce the likelihood of zero inflation. In other words, they may help promote case detection and reporting in areas that previously had no recorded AIDS cases. Other variables, such as low education level (X_2), condom use (X_3), and community health insurance coverage (X_5) do not have a significant effect on the zero-inflation component.

In summary, these results indicate that the 25-29 age group plays a dominant role in determining the distribution of AIDS cases, while health education activities potentially influence case detection through increased awareness and reporting. Based on the estimation results of both the Poisson and Zero-Inflation components, the final model is obtained as follows:

$$\log(\lambda_1) = -6.3896 + 0.7022X_1 + 0.0145X_2 + 0.1751X_3 + 0.0657X_4 + 0.0295X_5$$

$$\log(\lambda_2) = 1.5761 + 0.0267X_1 - 0.0094X_2 + 0.0709X_3 + 0.0716X_4 - 0.0425X_5$$

$$\log\left(\frac{\pi_1}{1-\pi_1}\right) = -306.9160 + 57.5542X_1 - 5.1381X_2 - 92.4985X_3 + 209.6302X_4 + 0.0591X_5$$

$$\log\left(\frac{\pi_2}{1-\pi_2}\right) = 5621.2511 - 702.4752X_1 + 6.7695X_2 - 171.4635X_3 - 777.5620X_4 - 9.6843X_5$$

Based on the obtained model, the interpretation of the Bivariate Zero-Inflated Poisson (BZIP) model parameters is as follows:

(a) Poisson Component for Y_1 (HIV Cases)

No explanatory variable is statistically significant at the 5% level. However, the condom use variable (X_3) has a p-value = 0.0962 (approaching significance) with a positive coefficient of 0.1751. This indicates that an increase in condom use tends to coincide with an increase in the number of reported HIV cases. This finding may be interpreted as showing that regions with higher awareness of condom use also tend to have better reporting systems, resulting in more HIV cases being identified rather than actual increases in transmission.

(b) (Poisson Component for Y_2 (AIDS Cases)

The community health insurance variable (X_5) has a significant negative effect ($p = 0.0194$). This means that as the proportion of the population covered by health insurance increases, the number of AIDS cases decreases. Access to health services facilitates early detection and treatment, preventing the progression from HIV infection to AIDS.

(c) Zero-Inflation Component for Y_1 (HIV Cases)

The health education participation variable (X_4) is highly significant ($p < 0.001$) with a positive coefficient of 209.6302. This suggests that the higher the level of public participation in health education activities, the greater the probability that a region will have no HIV cases (increased zero inflation). In other words, health education programs are effective in preventing new HIV infections through improved public awareness.

(d) Zero-Inflation Component for Y_2 (AIDS Cases)

The population aged 25-29 years (X_1) variable shows a significant negative effect ($p < 0.001$) with a coefficient of –702.4752, indicating that as the proportion of people aged 25-29 increases, the likelihood that a region will be free of AIDS cases decreases. This age group represents the productive population, which is generally more socially active and therefore at a higher risk of HIV/AIDS transmission. Additionally, the health education variable (X_4) has a p-value of 0.0820, which approaches the 10% significance level, indicating that education programs may help reduce AIDS risk by improving awareness and encouraging early detection and reporting of cases.

After thirty iterations using the Expectation–Maximization (EM) algorithm, the Bivariate Zero-Inflated Poisson (BZIP) model produced an Akaike Information Criterion (AIC) value of 180.3489, indicating that the resulting model is highly efficient. This efficiency reflects the model's strong ability to accommodate the large number of zero values (excess zeros) present in the HIV and AIDS data. Consequently, the BZIP model demonstrates a good level of fit to the observed counts of individuals living with HIV and AIDS. For comparison (Table 10), several previous studies reported higher AIC values, indicating less optimal model performance. [19] applied BZIP regression using the Newton–Raphson numerical method and obtained an AIC value of 340.6977. [20] implemented the Geographically Weighted BZIP Regression (GWBZIPR) model, also with the Newton–Raphson method and achieved an AIC of 320.3074. Another study, referred to as [10], reported an AIC of 910.2177, which is substantially higher, indicating poor performance in handling overdispersion and zero inflation. [21] employed the BZIP Inverse Gaussian (BZIPIG) Regression with the BHHH algorithm, yielding an AIC of 317.96. Among all these models, the BZIP model in the present study produced the lowest AIC (180.3489), demonstrating the best performance. This result confirms that the proposed BZIP model efficiently captures the underlying data structure, achieving a balance between predictive accuracy and model simplicity, making it a highly effective approach for modelling bivariate count data with excess zeros.

Table 10 AIC Comparison Among Competing Models

Model Type	Estimation Method	AIC Value
Bivariate Zero-Inflated Poisson (BZIP) Regression	**EM Algorithm**	**180.3489**
Bivariate Zero-Inflated Poisson (BZIP) Regression	Newton–Raphson	340.6977
Geographically Weighted BZIP (GWBZIP) Regression	Newton–Raphson	320.3074
Bivariate Poisson Inverse Gaussian (BPIG) Regression	Newton–Raphson	910.2177
BZIP Inverse Gaussian (BZIPIG) Regression	BHHH Algorithm	317.9600

3.7 Conclusions and suggestions

Based on the analysis using the Bivariate Zero-Inflated Poisson (BZIP) model, it can be concluded that the factors influencing the number of people living with HIV and AIDS exhibit different characteristics. The health education variable (X_4) was found to have a significant positive effect in increasing the probability of regions being free of HIV cases, while the community health insurance variable (X_5) showed a significant negative effect on the number of AIDS cases, indicating that access to healthcare services plays an essential role in suppressing disease progression. In addition, the 25-29 age group (X_1) was identified as the most vulnerable group to the spread of AIDS. Overall, the findings of this study highlight that enhancing health education and expanding community health insurance coverage are effective strategies for controlling HIV and AIDS at the regional level. It is recommended that health education programs be strengthened and continuously expanded to raise public awareness about HIV prevention, particularly among the productive age group (25-29 years), which has the highest risk of infection. Furthermore, the coverage of community health insurance should be increased to ensure that individuals living with HIV have adequate access to treatment and counseling services, thereby preventing the progression to AIDS. The government and relevant institutions are also encouraged to implement integrated policies that combine education, healthcare services, and social support, in order to create a more effective and sustainable approach to HIV and AIDS control in the community. Finally, the BZIP model applied in this study demonstrates the best performance, as evidenced by the lowest AIC value compared to all previously tested models, confirming its superiority in modeling bivariate count data with excess zeros.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] Y. Altinisik, Addressing Overdispersion and Zero-Inflation for Clustered Count Data via New Multilevel Heterogenous Hurdle Models, J. Appl. Stat. 50 (2022), 408-433. https://doi.org/10.1080/02664763.2022.2096875.

[2] H. A. Rahayuning And Purhadi, Pemodelan Jumlah Kematian Ibu Nifas Di Karesidenan Pekalongan Provinsi Jawa Tengah Tahun 2017 Menggunakan Regresi Zero-Inflated Poisson Inverse Gaussian, (2020), Vol. 3, No. September, Pp. 99–106.

[3] K.J. Geoffroy Kouakou, O. Hili, J. Dupuy, Estimation of Zero-Inflated Bivariate Poisson Regression with Missing Covariates, Commun. Stat. - Theory Methods 53 (2023), 7216-7243. https://doi.org/10.1080/03610926.2023.2262637.

[4] R. Azwarini, P. Purhadi, Pemodelan Jumlah Kasus Kusta Di Kabupaten Mojokerto Dan Kabupaten Jombang Tahun 2019 Menggunakan Regresi Zero-Inflated Poisson Inverse Gaussian, Inferensi 4 (2021), 85. https://doi.org/10.12962/j27213862.v4i2.10667.

[5] C. Li, J. Lu, J. Park, K. Kim, P.A. Brinkley, et al., Multivariate Zero-Inflated Poisson Models and Their Applications, Technometrics 41 (1999), 29-38. https://doi.org/10.1080/00401706.1999.10485593.

[6] K. Wang, A.H. Lee, K.K. Yau, P.J. Carrivick, A Bivariate Zero-Inflated Poisson Regression Model to Analyze Occupational Injuries, Accid. Anal. Prev. 35 (2003), 625-629. https://doi.org/10.1016/s0001-4575(02)00036-2.

[7] J. Lee, B.C. Jung, S.H. Jin, Tests for Zero Inflation in a Bivariate Zero‐inflated Poisson Model, Stat. Neerlandica 63 (2009), 400-417. https://doi.org/10.1111/j.1467-9574.2009.00430.x.

[8] T. Mohammadi, S. Kheiri, M. Sedehi, Analysis of Blood Transfusion Data Using Bivariate Zero-Inflated Poisson Model: A Bayesian Approach, Comput. Math. Methods Med. 2016 (2016), 7878325. https://doi.org/10.1155/2016/7878325.

[9] A. Kautsar, Sebaran Kasus HIV di Indonesia 2025, Ini Pentingnya Edukasi dan Deteksi Dini, 2025. https://health.detik.com/berita-detikhealth/d-8034016/sebaran-kasus-hiv-di-indonesia-2025-ini-pentingnya-edukasi-dan-deteksi-dini.

[10] S. U. Wijaya, Pendugaan Parameter dan Pengujian Hipotesis pada Bivariate Poisson Inverse Gaussian Regression, Master Thesis, Institut Teknologi Sepuluh Nopember, 2017.

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

[12] P. Zhang, D. Pitt, X. Wu, A New Multivariate Zero-Inflated Hurdle Model with Applications in Automobile Insurance, ASTIN Bull. 52 (2022), 393-416. https://doi.org/10.1017/asb.2021.39.

[13] P. Holgate, Estimation for the Bivariate Poisson Distribution, Biometrika 51 (1964), 241. https://doi.org/10.2307/2334210.

[14] Q. Aini, Penaksiran Parameter dan Pengujian Hipotesis Geographically Weighted Bivariate Zero Inflated Generalized Poisson Regression, Thesis, Institut Teknologi Sepuluh Nopember, 2020.

[15] S.K. Ng, T. Krishnan, G.J. McLachlan, The EM Algorithm, Springer, Berlin, Heidelberg, 2012. https://doi.org/https://doi.org/10.1007/978-3-642-21551-3_6.

[16] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, John Wiley & Sons, Inc., 1998. https://doi.org/10.1002/9780470191613.

[17] S. Kocherlakota, K. Kocherlakota, Bivariate Discrete Distributions, CRC Press, 2017. https://doi.org/10.1201/9781315138480.

[18] J. van den Broek, A Score Test for Zero Inflation in a Poisson Distribution, Biometrics 51 (1995), 738. https://doi.org/10.2307/2532959.

[19] R.L. Umami, Estimation Parameters And Hypotheses Testing Of Bivariate Zero-Inflated Poisson Regression, Thesis, Institut Teknologi Sepuluh Nopember, 2015.

[20] J, Pangulimang, Parameter Estimation and Hypothesis Testing of Geographically Weighted Bivariate Zero-Inflated Poisson Regression, Thesis, Institut Teknologi Sepuluh Nopember, 2016.

[21] R. Noviyana, Bivariate Zero-Inflated Poisson Inverse Gaussian Regression Model and Its Application, Thesis, Institut Teknologi Sepuluh Nopember, 2021.