Jurnal Varian
Vol. 8, No. 2, April 2025, pp 199–208
e-ISSN: 2581-2017

Exploring Crime Problems from A Statistical Point of View with Negative Binomial Regression

Andrea Tri Rian Dani[1], M. Fathurahman[1], Ludia Ni'matuzzahroh[2], Regita Putri Permata[3], Fachrian Bimantoro Putra[1]
[1]Universitas Mulawarman, Samarinda, Indonesia
[2]Politeknik Perkapalan Negeri Surabaya, Surabaya, Indonesia
[3]Universitas Telkom, Surabaya, Indonesia

Article Info

Article history:
Received : 09-04-2024
Revised : 03-08-2025
Accepted : 07-23-2025

Keywords:
Criminality;
Negative Binomial Regression;
Overdispersion.

ABSTRACT

Criminality is a complex issue in Indonesia that holds great importance for the government, law enforcement agencies, and society. The underlying causes of Indonesia's crime problem are complex and impacted by various circumstances. This research aims to model the crime problem in Indonesia and identify the factors that influence it. The method used in this research is Negative Binomial Regression. The results of the study indicate that the negative binomial regression model can be effectively used to model criminal problems, as the variance value is significantly greater than the average. Based on the parameter significance test results, both simultaneously and partially, the open unemployment rate, Gini ratio, average years of schooling, and prevalence of inadequate food consumption significantly affect the crime rate, with an Akaike's Information Criterion Corrected (AICc) value of 698,098. These findings suggest that addressing economic inequality, unemployment, education, and food security could help reduce crime in Indonesia. Policies aimed at improving job opportunities, reducing income disparity, and enhancing education and food security are crucial in mitigating crime. This study offers policymakers and law enforcement agencies valuable insights, providing a foundation for more targeted and effective crime prevention strategies. Future research could employ the robust Poisson Inverse Gaussian Regression method to avoid the overdispersion problem.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i2.4445

Corresponding Author:
Andrea Tri Rian Dani,
Department of Statistics, Universitas Mulawarman,
Email: andreatririandani@fmipa.unmul.ac.id

How to Cite:
Dani, A. T. R., Fathurahman, M., Ni'matuzzahroh, L., Permata, R. P., & Putra, F. B. (2025). Exploring Crime Problems from A Statistical Point of View with Negative Binomial Regression. *Jurnal Varian*, *8*(2), 199-208.

A. INTRODUCTION

Criminality is a complex problem in Indonesia that is of particular concern to the government, law enforcement officials, and society (Nurferyanto & Takahashi, 2024). The crime rate in Indonesia has fluctuated every year (Sugiharti et al., 2023). According to publications from the Central Bureau of Statistics, the number of crimes in Indonesia that were reported in 2020 was 247218 cases; in 2021, as many as 239481 cases; in 2022, as many as 372965 cases, where the number of criminal cases in 2022 has increased quite drastically compared to the previous year, namely an increase of 55.73%. Then, in 2023, there were 372,897 cases, a slight decrease from the previous year. The types of criminal cases in Indonesia are very diverse, such as theft, robbery, persecution, and murder, and more complex criminal cases, such as corruption, money laundering, and human trafficking (Badan Pusat Statistik, 2024).

The root causes of criminality in Indonesia are diverse and influenced by various factors; therefore, a comprehensive approach is necessary to address these causes effectively. Before overcoming these problems, it is necessary to further and deeply identify the factors that affect the crime rate in Indonesia so that the policies taken by the government are right on target and able to handle and reduce the crime rate in Indonesia. In identifying the factors that influence crime, we can utilize one of the statistical methods, specifically regression analysis. Regression analysis can model the relationship between one or more predictor variables and response variables (Coshall & Hardle, 1993; Dani et al., 2021; Suparti et al., 2020). The response variables used in most studies are continuous so that this study will develop regression analysis with discrete response variables (Widyaningsih et al., 2021). If the response variable in regression analysis is discrete, several alternative methods can be used, such as developing regression analysis, namely Poisson and negative binomial regression (Fathurahman, 2022).

Some assumptions must be met in Poisson regression analysis. One of these is the equidispersion assumption, which states that the variance and mean values of the response variable must be the same. In simpler terms, this means that the data should be evenly distributed around the mean. However, overdispersion circumstances are frequently encountered for discrete response variables, where the variance and mean values are not equal, or the variation value exceeds the mean. Based on this, negative binomial regression is appropriate method for dealing with overdispersion than Poisson regression. If Poisson regression analysis is used when there is overdispersion, the estimated model parameters will be underestimated. This is demonstrated when the model parameter significance test results indicate that the predictor variable influences the response variable while having no significant effect (İyit & Sevim, 2023). Based on current reality, criminological data rarely display the same variance and mean values, making negative binomial regression research increasingly popular for studying crime. Negative binomial regression analysis can describe the relationship between the response variable and the predictor variable under study (Jhwueng & Wu, 2023; Stoklosa et al., 2022). Negative binomial regression is a non-linear regression model based on the Poisson-gamma mixture distribution (Keswari et al., 2014). Several studies have used negative binomial regression analysis, including research conducted by Fitrial and Fatikhurrizqi in 2020 modeled the number of COVID-19 cases in Indonesia using a Poisson regression and negative binomial regression approach (Fitrial & Fatikhurrizqi, 2021). Sauddin et al. (2020) also modeled the number of maternal deaths in South Sulawesi Province using negative binomial regression. Other studies that serve as references for this study include Alem et al. (2024); Kang et al. (2021); Khattak et al. (2024); Ramadan et al. (2024); Stoklosa et al. (2022). The difference between this research and previous research is the application of negative regression to the crime problem in Indonesia. The response variable in this study is discrete and exhibits an overdispersion problem; hence, negative binomial regression is one of the methods utilized to address this issue.

Based on the description above, this study identifies the factors that affect the criminality rate in Indonesia in 2023 using negative binomial regression analysis and selects the best model using the Akaike's Information Criterion Corrected (AICc) value. This study aims to identify factors that significantly affect the criminality rate in Indonesia, providing input for the government and law enforcement to prevent and address criminality in the country.

B. RESEARCH METHOD

The criminality rate in Indonesia is the subject of this study, with data sourced from the Republic of Indonesia's Central Statistics Agency (BPS) publication. In this study, the population is the crime rate in each province of Indonesia, and the sample is the crime rate in each province of Indonesia in 2023. Table 1 shows the factors used in this investigation.

Table 1. Research variables

Notation	Variable	Operational Definition	Data Type
y	Criminality Rate	A crime rate is a number that indicates the frequency of crimes in a specific area and during a specified time frame.	Discrete
x_1	Population Density	Population density refers to the number of people residing in a square kilometer of land.	Continue
x_2	Open Unemployment Rate	The percentage of unemployed persons to the whole work force is the Open Unemployment Rate.	Continue
x_3	Gini Ratio	The gini coefficient is a measurement tool to assess the level of income inequality or economic welfare in a region.	Continue

Notation	Variable	Operational Definition	Data Type
x_4	Average Years of Schooling	The average number of years of schooling (AYS) is a statistic displaying the formal education intake of the population aged 15 years and above.	Continue
x_5	Prevalence of Inadequate Food Consumption	Prevalence of Food Consumption Inadequacy is a condition in which a person regularly consumes an insufficient amount of food to meet the energy needs for a normal, active, and healthy life.	Continue

The steps in conducting data analysis are written as follows:

1. Exploring data for each predictor variable and response variable.
2. Detect multicollinearity between predictor variables using the Variance Inflation Factor (VIF) criterion in Equation (1).

$$VIF = \frac{1}{1-R_j^2} \tag{1}$$

Where R_j^2 is the coefficient of determination between variable x_j and other independent variables, with $j = 1, 2, \ldots, p$. The Variance Inflation Factor (VIF) is a measure used to detect multicollinearity. VIF value greater than 10 indicates a multicollinearity problem. In negative binomial regression modeling, it is expected that there will be no multicollinearity problem.

3. Detecting cases of overdispersion based on mean and variance values in Equations (2) and (3).

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{2}$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y}) \tag{3}$$

where $\bar{y}$ is the mean, σ^2 is the variance. Detection of overdispersion can be done by examining the mean, which is smaller than the variance of the response variable data.

4. Performing negative binomial regression modeling, the parameter estimation method is Maximum Likelihood Estimation (MLE) with the Fisher Scoring iteration. The negative binomial probability function is written in Equation (4).

$$P(y = y_i|\lambda;\zeta) = \frac{\Gamma\left(y_i + \frac{1}{\zeta}\right)}{\Gamma(y_i+1)\,\Gamma\left(\frac{1}{\zeta}\right)} \left(\frac{1}{1+\zeta\lambda}\right)^{1/\zeta} \left(\frac{\zeta\lambda}{1+\zeta\lambda}\right)^{y_i} \tag{4}$$

for $i = 1, 2, \ldots, n$ and $\zeta \geq 0$ with $\lambda = E(y_i)$ is an average form y_i similar with mean, ζ is dispersion parameter, y_i is response variable. Using the natural logarithm as a linking function, the negative binomial regression model namely $\ln(\lambda_i) = \mathbf{x_i^T}\gamma$ in Equation (5)

$$\lambda_i = \exp\left(\mathbf{x_i^T}\gamma\right) + \varepsilon_i \tag{5}$$

using Maximum Likelihood Estimation (MLE), the Likelihood function from Equation (6).

$$L(\gamma,\lambda) = \prod_{i=1}^{n}\left(\left(\prod_{j=1}^{y_i-1}\frac{j}{\lambda}\right)\frac{1}{y_i!}\left(\frac{1}{1+\lambda\exp\left(\mathbf{x_i^T}\gamma\right)}\right)^{1/\lambda}\left(\frac{\lambda\exp\left(\mathbf{x_i^T}\gamma\right)}{1+\lambda\exp\left(\mathbf{x_i^T}\gamma\right)}\right)^{y_i}\right) \tag{6}$$

If $\ell(\gamma,\lambda) = \ln L(\gamma,\lambda)$, then it is known that the ln-likelihood function in Equation (7),

$$\ell(\gamma,\lambda) = \sum_{i=1}^{n}\left(\left(\sum_{j=1}^{y-1}\ln\left(j+\frac{1}{\lambda}\right)\right) - \ln(y_i!) + y_i\ln\left(\lambda\exp\left(\mathbf{x_i^T}\gamma\right)\right) - \left(y_i+\frac{1}{\lambda}\right)\ln\left(1+\lambda\frac{1}{\lambda}\right)\right) \tag{7}$$

the partial derivative of function $\ell(\gamma,\lambda)$ is in the form of a function that is not explicit, so it requires the help of iteration. The iteration used in this study is the Fisher-Scoring algorithm, as outlined in Equation (8).

$$\hat{\gamma}_{m+1} = \hat{\gamma}_{(m)} - \mathbf{I}^{-1}\left(\hat{\gamma}_{(m)}\right)\mathbf{g}^T\left(\hat{\gamma}_{(m)}\right) \tag{8}$$

The iteration is terminated when it reaches a convergent state , where is a very small positive real number.

5. Evaluating the negative binomial regression model's parameters for significance both simultaneously and partially
Hypothesis formulation from simultaneous testing
$H_0 \ : \ \gamma_1 = \gamma_2 = \ldots = \gamma_j = 0$
$H_1 \ : \ $ at least there is one $\gamma_j \neq 0, \ j = 1, 2, \ldots, p$
The test statistic used is G in Equation (9).

$$\begin{aligned} G \quad = \quad & -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) \\ & -2 \left(\ell \left(\hat{\Omega} \right) - \ell \left(\hat{\omega} \right) \right) \end{aligned} \tag{9}$$

with,

$$L(\hat{\Omega}) = \sum_{i=1}^{n} \left(\left(\sum_{j=1}^{y-1} \ln \left(j + \frac{1}{\hat{\lambda}} \right) \right) - \ln (y_i!) + y_i \ln \left(\hat{\lambda} \exp \left(\mathbf{x_i^T} \hat{\gamma} \right) \right) - \left(y_i + \frac{1}{\hat{\lambda}} \right) \ln \left(1 + \lambda \mathbf{x_i^T} \hat{\gamma} \right) \right)$$

$$L(\hat{\omega}) = \sum_{i=1}^{n} \left(\left(\sum_{j=1}^{y-1} \ln \left(j + \frac{1}{\hat{\lambda}} \right) \right) - \ln (y_i!) + y_i \ln \left(\hat{\lambda} \exp \left(x_i^T \gamma_0 \right) \right) - \left(y_i + \frac{1}{\hat{\lambda}} \right) \ln \left(1 + \lambda x_i^T \gamma_0 \right) \right)$$

The testing criteria is H_0 rejected when the value $G > \chi^2_{(\alpha,p)}$. Next, the hypothesis formulation from partial testing:
$H_0 \ : \ \gamma_j = 0$
$H_1 \ : \ \gamma_j \neq 0, \ j = 1, 2, \ldots, p$
The test statistic utilizes the Z-test statistic as presented in Equation (10).

$$Z = \frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)} \tag{10}$$

The testing criteria is H_0 rejected when the value $|Z| > Z_{\alpha/2}$.

6. Perform interpretation and draw conclusions.

C. RESULT AND DISCUSSION

1. Exploration Data

In this research, descriptive statistics are used in the form of bar charts for the response variable (y), specifically the Criminality Rate. By using the median value as the assessment limit, we can determine which provinces have numbers above the median (indicating danger) and those below the median. From this description, it can be used as an initial exploration of areas prone to crime.

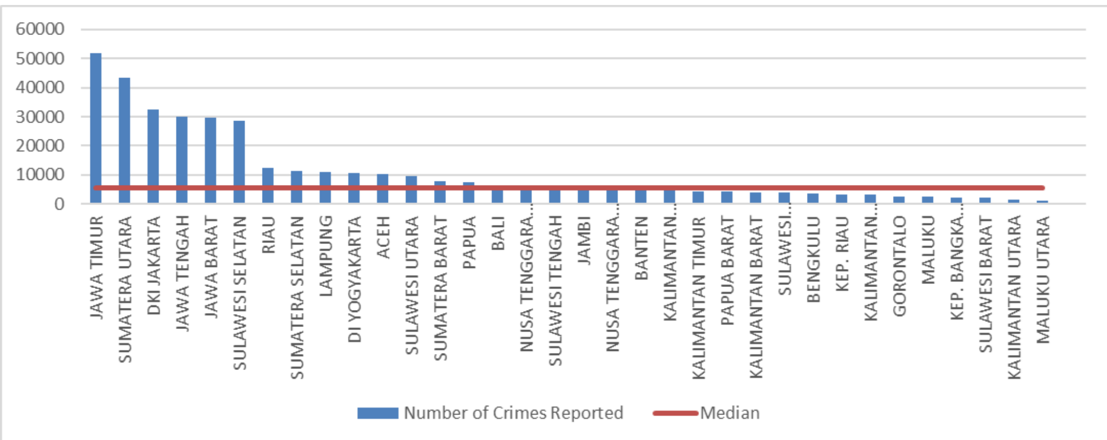

Figure 1. Bar Chart

Jawa Timur, Sumatera Utara, DKI Jakarta, Jawa Tengah, and several other provinces are known to have crime rates higher than the median value, as shown in Figure 1. Population density, economic issues, and ineffective law enforcement are just a few

DOI: https://doi.org/10.30812/varian.v8i2.4445

of the numerous factors contributing to high crime rates in certain regions. The goal of this study is to identify what causes and influences high crime rates, which utilizes statistical inference with negative binomial regression.

2. Multicollinearity Detection

The goal of multicollinearity detection is to determine if the predictor variables have a strong linear relationship. It is believed that multicollinearity will not be an issue in regression modeling. The researchers in this study looked for signs of multicollinearity using the VIF value. The results of the multicollinearity detection are shown in Figure 2 and Table 2.

Table 2. VIF value

x_1	x_2	x_3	x_4	x_5
1.528	1.518	1.178	1.712	1.103

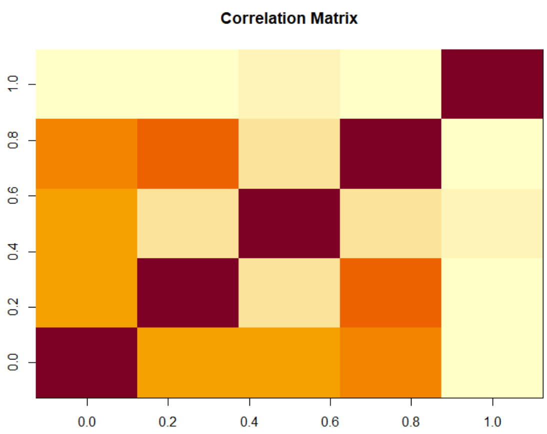

Figure 2. Correlation Matrix

Figure 2 shows that the stronger the connection between the predictor variables, the darker the correlation matrix. Suppose all the predictor variables in Table 2 have VIF values less than 10. In that case, there is no multicollinearity issue with the predictor variables, supported by the results in Figure 2.

3. Overdispersion Problem Detection

Detection of overdispersion problems in crime rate data in Indonesia using mean and variance measures. Overdispersion is a condition where the variance value is greater than the mean value $(Var\,(y_i) > E\,(y_i))$. Descriptive statistics for the response variable (y_i) are shown in Table 3.

Table 3. Descriptive Statistics

Mean	**Variance**	**Minimum**	**Maximum**
10967.560	161244632	1220	51905

According to Table 3, the crime rate data in Indonesia can be modeled using negative binomial regression, as it addresses the overdispersion problem. Negative binomial regression is an alternative model that can represent discrete data in situations of overdispersion.

4. Modeling Criminality Rate Using Negative Binomial Regression

The general model of negative binomial regression is shown in Equation 12, represent the crime rate μ_i as an exponential function of several explanatory variables. In this model, the crime rate is influenced by factors such as population density, open

DOI: https://doi.org/10.30812/varian.v8i2.4445

unemployment rate, Gini ratio, average years of schooling, and prevalence of inadequate food consumption, which are captured by the coefficients γ_1, γ_2, γ_3, γ_4, and γ_5.

$$\mu_i = \exp\left(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \gamma_4 x_{4i} + \gamma_5 x_{5i}\right) + e_i \tag{11}$$

To estimate the model's parameters, we employed Maximum Likelihood Estimation (MLE) with Fisher-Scoring iteration, a commonly used method for estimating parameters in count data models, such as the negative binomial regression. The results of the parameter estimation are presented in Table 4, where the coefficients and standard errors of each parameter are provided.

Table 4. Parameter estimation results

Parameter Estimation	Coefficient	Standard Error
$\hat{\gamma}_0$	9.180	2.131
$\hat{\gamma}_1$	3.436×10^{-5}	5.776×10^{-5}
$\hat{\gamma}_2$	3.312×10^{-1}	9.789×10^{-2}
$\hat{\gamma}_3$	7.234	3.031
$\hat{\gamma}_4$	-4.005×10^{-1}	2.032×10^{-1}
$\hat{\gamma}_5$	4.459×10^{-2}	1.710×10^{-2}

Following Table 4, a parameter significance test was conducted to determine whether the parameters significantly contribute to explaining the variation in the crime rate. Both simultaneous and partial significance tests were applied to evaluate the statistical significance of each parameter. These tests help in identifying which variables have a meaningful relationship with the crime rate, guiding further policy and intervention strategies. The hypothesis for the simultaneous test of the negative binomial regression model is presented below, and the G-test statistics are displayed in Table 5.

Hypothesis:

H_0 : $\gamma_1 = \gamma_2 = \cdots = \gamma_5 = 0$
(predictor variables simultaneously do not affect the criminality rate in Indonesia)

H_1 : there is at least one $\gamma_j \neq 0,\ j = 1, 2, \ldots, 5$
(predictor variables simultaneously affect the criminality rate in Indonesia.)

Table 5. G test statistics

G	p-value	$\chi^2_{(0.05,\ 5)}$
20.432	0.001	11.071

From Table 5, the results of the G test statistics obtained G value of 20.432 is greater than $\chi^2_{(0.05,\ 5)}$ of 11.071 and or p-value < 0.05 so that the decision H_0 is rejected. The predictor variables simultaneously affect the crime rate in Indonesia. Furthermore, partial testing aims to determine which predictor variables have a substantial impact on the response variable. The test statistic used is the Z-test statistic with a significance level of 0.05, and the decision is shown in Table 6. The partial test hypothesis:

Hypothesis:

H_0 : $\gamma_j = 0$
(predictor variable partially do not affect the criminality rate in Indonesia)

H_1 : $\gamma_j \neq 0,\ j = 1, 2, \ldots, 5$
(predictor variable partially affect the criminality rate in Indonesia)

Table 6. Z Test Statistics

Parameter	$\lvert Z \rvert$value	$Z_{\alpha/2}$	p-value	Decision
$\hat{\gamma}_0$	4.308	1.96	1.65×10^{-5}	
$\hat{\gamma}_1$	0.595	1.96	0.552	Failed to Reject the null hypothesis
$\hat{\gamma}_2$	3.383	1.96	0.001	Reject the null hypothesis
$\hat{\gamma}_3$	2.387	1.96	0.017	Reject the null hypothesis
$\hat{\gamma}_4$	1.971	1.96	0.049	Reject the null hypothesis
$\hat{\gamma}_5$	2.607	1.96	0.009	Reject the null hypothesis

It is known that there are four predictor variables with values $|Z| > Z_{\alpha/2}(1.96)$ which indicates H_0 is rejected, meaning the predictor variable is significant. The significant predictor variables are open unemployment rate (x_2), Gini ratio (x_3), average years of schooling (x_4), and prevalence of inadequate food consumption (x_5). The negative binomial regression model formed has an AICc value of 698.098. Furthermore, based on Table 4, the negative binomial regression model can be written in Equation 13.

$$\mu_i = \exp\left(9.180 + 3.436 \times 10^{-5}x_{1i} + 3.312 \times 10^{-1}x_{2i} + 7.234x_{3i} - 4.005 \times 10^{-1}x_{4i} + 4.459 \times 10^{-2}x_{5i}\right) \tag{12}$$

Based on the negative binomial regression model, it can be seen that: **(1)** If (x_1) population density increases by one unit while other factors remain constant, the average criminality rate (y) will increase by $\exp(3.436 \times 10^{-5})$ or 1.000 times. **(2)** For every one unit increase in the (x_2) open unemployment rate while other factors remain constant, the average criminality rate (y) will increase by $\exp(3.312 \times 10^{-1})$ or 1.393 times. **(3)** For every one unit increase in the Gini ratio (x_3) while other factors remain constant, the average criminality rate (y) will increase by $\exp(7.234)$ or 1385.754 times. **(4)** For every one unit increase in the average years of schooling (x_4) while other factors remain constant, the average criminality rate (y) will decrease by $\exp(-4.005 \times 10^{-1})$ or 0.667 times. **(5)** For every one unit increase in the prevalence of insufficient food consumption (x_5) while other factors remain constant, the average criminality rate (y) will increase by $\exp(4.459 \times 10^{-2})$ or 1.046 times.

The problem of crime rates indicates an overdispersion problem where the classical regression model cannot be used. Moreover, the crime rate is a discrete variable; therefore, the negative binomial regression model, which incorporates a dispersion parameter, is one alternative that can be used. The findings of this study indicate that four variables have a significant influence on crime. This is also supported by research by Fatmala et al. (2024), which explains that negative binomial regression is an alternative that can be used to overcome the overdispersion problem. Negative binomial regression has a smaller AIC value, making it more effective for use in the modeling process. Other research by Manurung et al. (2024) showed that negative binomial regression is an alternative modeling that can be used when overdispersion problems occur and is better than Poisson regression. In this research, the application of the negative binomial regression model to crime problems is novel, as it has not been used before. Based on the significant predictors, the government could consider to: implement policies to reduce unemployment rates, such as job creation programs, skills training initiatives, and support for small and medium-sized enterprises; invest in infrastructure development to stimulate job growth; Improve Education Access and Quality: increase access to quality education at all levels and provide scholarships and financial aid to support students; implement programs to improve food security, such as food assistance programs, school meal programs, and support for local food production; the government should regularly evaluate the impact of its policies on the outcome variable. This will help to identify which policies are most effective and inform future policy decisions, and; continuous monitoring and analysis of relevant data are crucial for understanding the underlying factors affecting the outcome variable and for informing effective policy interventions.

D. CONCLUSION AND SUGGESTION

The best negative binomial regression model is one of the alternatives to model the crime rate in Indonesia where there is an overdispersion problem. It is known that the significant predictor variables are open unemployment rate (x_2), Gini ratio (x_3), average years of schooling (x_4), and prevalence of inadequate food consumption (x_5). Goodness criteria using AICc with a value of 698,098. The findings suggest that addressing issues such as unemployment, income inequality, education, and food security could be essential in reducing crime. Policymakers should focus on creating programs that improve employment opportunities, reduce income inequality, enhance educational access, and improve food security to mitigate crime in Indonesia. Suggestions for further research include utilizing the robust Poisson Inverse Gaussian Regression method to model data that exhibits overdispersion or underdispersion issues.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) and Universitas Mulawarman for their partial funding and support of this work.

DECLARATIONS

AUTHOR CONTIBUTION

Andrea Tri Rian Dani: Conceptualization, Methodology, Validity Test, Writing-Preparation of the First Draft, and Supervision. M. Fathurahman: Methodology, Validity Test Fachrian Bimantoro Putra: Data Curation, Analysis, Visualization. Ludia Ni'matuzzahroh: Analysis, Editing Draft and Visualization. Regita Putri Permata: Analysis, Validity Test, Supervision.

FUNDING STATEMENT

This work was supported in part by the funding of the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) and Universitas Mulawarman.

COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Alem, A. Z., Tegegne, B. A., Aragaw, F. M., Teklu, R. E., & Baykeda, T. A. (2024). Multilevel negative binomial analysis of factors associated with numbers of antenatal care contacts in low and middle income countries: Findings from 59 nationally representative datasets (M. Nair, Ed.). *PLOS ONE*, *19*(4), e0301542. https://doi.org/10.1371/journal.pone.0301542

Badan Pusat Statistik. (2024). *Statistik Kriminalitas Indonesia 2023*.

Coshall, J., & Hardle, W. (1993). Applied Nonparametric Regression. *The Statistician*, *42*(2), 195. https://doi.org/10.2307/2348990

Dani, A. T. R., Ni'matuzzahroh, L., Ratnasari, V., & Budiantara, I. N. (2021). Pemodelan Regresi Nonparametrik Spline Truncated pada Data Longitudinal. *Inferensi*, *4*(1), 47. https://doi.org/10.12962/j27213862.v4i1.8737

Fathurahman, M. (2022). Regresi Binomial Negatif untuk Memodelkan Kematian Bayi di Kalimantan Timur. *Eksponensial*, *13*(1), 79. https://doi.org/10.30872/eksponensial.v13i1.888

Fatmala, C. T., Hayati, M., Permatasari, R., Hudori, M., & Dalimunthe, D. Y. (2024). Pemodelan Jumlah Kasus HIV/AIDS di Provinsi Lampung Menggunakan Regresi Binomial Negatif. *Journal of Mathematics: Theory and Applications*, *6*(2), 168–177. https://doi.org/10.31605/jomta.v6i2.4069

Fitrial, N. H., & Fatikhurrizqi, A. (2021). Pemodelan Jumlah Kasus COVID-19 di Indonesia Dengan Pendekatan Regresi Poisson dan Regresi Binomial Negatif. *Seminar Nasional Official Statistics*, *2020*(1), 65–72. https://doi.org/10.34123/semnasoffstat.v2020i1.465

İyit, N., & Sevim, F. (2023). A novel statistical modeling of air pollution and the COVID-19 pandemic mortality data by Poisson, geometric, and negative binomial regression models with fixed and random effects. *Open Chemistry*, *21*(1), 1–15. https://doi.org/10.1515/chem-2023-0364

Jhwueng, D.-C., & Wu, C.-Y. (2023). A Novel Phylogenetic Negative Binomial Regression Model for Count-Dependent Variables. *Biology*, *12*(8), 1148. https://doi.org/10.3390/biology12081148

Kang, K. I., Kang, K., & Kim, C. (2021). Risk Factors Influencing Cyberbullying Perpetration among Middle School Students in Korea: Analysis Using the Zero-Inflated Negative Binomial Regression Model. *International Journal of Environmental Research and Public Health*, *18*(5), 2224. https://doi.org/10.3390/ijerph18052224

Keswari, N. M. R., Sumarjaya, I. W., & Suciptawati, N. L. P. (2014). Perbandingan Regresi Binomial Negatif dan Regresi Generalisasi Poisson dalam Mengatasi Overdispersi (Studi Kasus: Jumlah Tenaga Kerja Usaha Pencetak Genteng di BR. Dukuh, Desa Pejaten). *E-Jurnal Matematika*, *3*(3), 107. https://doi.org/10.24843/mtk.2014.v03.i03.p072

Khattak, M. W., De Backer, H., De Winne, P., Brijs, T., & Pirdavani, A. (2024). Comparative Evaluation of Crash Hotspot Identification Methods: Empirical Bayes vs. Potential for Safety Improvement Using Variants of Negative Binomial Models. *Sustainability*, *16*(4), 1537. https://doi.org/10.3390/su16041537

Manurung, B. L., Priharyanto, T., & Alifah, U. (2024). Pemodelan Jumlah Kekerasan Terhadap Perempuan di Jawa Timur dengan Regresi Poisson dan Binomial Negatif. *JSN : Jurnal Sains Natural*, *2*(3), 68–77. https://doi.org/10.35746/jsn.v2i3.547

Nurferyanto, D., & Takahashi, Y. (2024). Establishing Boundaries to Combat Tax Crimes in Indonesia. *Laws*, *13*(3), 29. https://doi.org/10.3390/laws13030029

Ramadan, A., Chamidah, N., & Budiantara, I. N. (2024). Modelling the number of HIV cases in Indonesia using negative binomial regression based on least square spline estimator. *Communications in Mathematical Biology and Neuroscience*, *2024*, 1–16. https://doi.org/10.28919/cmbn/8582

Sauddin, A., Auliah, N. I., & Alwi, W. (2020). Pemodelan Jumlah Kematian Ibu di Provinsi Sulawesi Selatan Menggunakan Regresi Binomial Negatif. *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, *8*(2), 42. https://doi.org/10.24252/msa.v8i2.17409

Stoklosa, J., Blakey, R. V., & Hui, F. K. C. (2022). An Overview of Modern Applications of Negative Binomial Modelling in Ecology and Biodiversity. *Diversity*, *14*(5), 320. https://doi.org/10.3390/d14050320

Sugiharti, L., Purwono, R., Esquivias, M. A., & Rohmawati, H. (2023). The Nexus between Crime Rates, Poverty, and Income Inequality: A Case Study of Indonesia. *Economies*, *11*(2), 62. https://doi.org/10.3390/economies11020062

Suparti, S., Warsito, B., Santoso, R., Yasin, H., Caraka, R. E., & Sudargo, S. (2020). Biresponses Kernel Nonparametric Regression: Inflation and Economic Growth. *International Journal of Criminology and Sociology*, *10*, 465–471. https://doi.org/10.6000/1929-4409.2021.10.54

Widyaningsih, Y., Arum, G. P., & Prawira, K. (2021). Aplikasi K-Fold Cross Validation dalam Penentuan Model Regresi Binomial Negatif Terbaik. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *15*(2), 315–322. https://doi.org/10.30598/barekengvol15iss2pp315-322

[This page intentionally left blank.]