Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.29471
Page: 768-779

Negative Binomial Regression Modeling of Stunting Cases in Central Java: Handling Overdispersion in Count Data

[1]Ulfah Farida
Department of Mathematics Education, Sebelas Maret University, Surakarta, Indonesia

[2]Laila Fitriana
Department of Mathematics Education, Sebelas Maret University, Surakarta, Indonesia

Article Info

Article history:

Accepted: 03 August 2026

Keywords:

AICc;
Count Data;
Negative Binomial Regression;
Poisson Regression;
Stunting.

ABSTRACT

This study aims to identify factors associated with the number of stunting cases across districts and cities in Central Java Province using count data regression models. Secondary data from 35 districts/cities in 2025 were analyzed within the generalized linear modeling framework. Poisson regression was initially applied, followed by exhaustive subset selection using the corrected Akaike Information Criterion (AICc) to obtain a parsimonious model. Model adequacy was evaluated through goodness-of-fit assessment, information criteria, and overdispersion diagnostics. The results indicated severe overdispersion in the Poisson model, suggesting that the equidispersion assumption was violated and that the model was unsuitable as the primary inferential framework. Consequently, a Negative Binomial regression model was estimated to accommodate the extra-Poisson variation. Model comparison demonstrated that the Negative Binomial model provided substantially better statistical performance, as reflected by lower AIC and AICc values, a dispersion statistic closer to one, and satisfactory diagnostic results. Based on the Negative Binomial model, exclusive breastfeeding coverage was the only variable that remained statistically significant, with an incidence rate ratio (IRR) of 1.0258 (95% CI: 1.0090–1.0426). Meanwhile, the dependency ratio and doctor ratio exhibited marginal significance. These findings highlight the importance of selecting appropriate count data models when overdispersion is present and suggest that stunting cases across Central Java are associated with complex nutritional, demographic, and healthcare-related conditions.

Corresponding Author:

Laila Fitriana,
Department of Mathematics Education
Sebelas Maret University
Email: lailafitriana_fkip@staff.uns.ac.id

1. INTRODUCTION

Stunting remains a serious public health problem because it affects physical growth, cognitive development, productivity, and the quality of human resources in the future [1][2]. In Central Java, stunting cases remain an important public health concern because variations in maternal conditions, socioeconomic factors, and access to healthcare services may contribute to differences in stunting prevalence across districts and cities [3][4]. At the regional level, variations in demographic and health characteristics across districts/cities result in different patterns in the number of stunting cases, so a statistical approach is needed to model the case data accurately [5].

The responsive variable analyzed in this study was the number of reported stunting cases in each district/city. Although rate based measure are commonly used in epidemiological studies, information on the population at risk at the district/city level was not available in the dataset. Therefore, the analysis focused on modeling the observed counts of stunting cases across regions. The results should be interpreted as explaining regional variation in the number of reported cases rather than individual level risk or prevalence.

In the context of statistics, the number of stunting cases at the district/city level constitutes count data. Poisson regression is commonly used as an initial modeling framework for count responses because it relates the expected number of events to a set of explanatory variables through a log link function [6][7][8][9]. However, count data frequently exhibit overdispersion, which may require alternative modeling approaches such as Negative Binomial regression to obtain reliable statistical inference [10][11].

Previous research has shown that low birth weight (BBLR) and exclusive breastfeeding are biological factors that have a significant effect on the incidence of stunting [12][13]. In addition, socioeconomic factors such as the percentage of the population living in poverty and the burden of dependency also affect households' ability to meet children's nutritional needs [14][15]. In terms of health services, the ratio of doctors is an important indicator because it relates to access to growth and development monitoring, nutrition education, and early intervention for children [16].

However, most previous studies have focused on using a full model with all predictor variables, without systematically selecting all possible combinations of variables [17]. In fact, the use of too many predictors with a limited number of observations can make the model overly complex or produce a saturated model, resulting in poor fit and less informative [18][19][20]. In addition, research specifically examining the selection of the best Poisson model across all predictor combinations using the Akaike Information Criterion (AIC) in stunting cases in Central Java is still very limited [21].

Based on these conditions, this study offers novelty through an exhaustive subset selection strategy that evaluates all possible combinations of predictor variables using AICc within a count data modeling framework. The approach aims to identify a parsimonious model while subsequently assessing the adequacy of alternative count data specifications when overdispersion is detected [22][23]. This approach is expected to produce a simpler, more efficient, and easier-to-interpret model, while still providing a good explanation of the number of stunting cases between districts/cities.

Therefore, this study aims to identify factors associated with the number of stunting cases in Central Java Province in 2025 using count data regression modeling. Candidate predictor combinations were first evaluated using Poisson regression and AICc based model selection to identify a parsimonious specification. Subquently, model adequacy was assessed through overdispersion diagnostics, and Negative Binomial regression was employed as the primary inferential model when overdispersion was detected [24].

2. RESEARCH METHOD

2.1 Variables and Data Sources

This study employed a quantitative observational approach, utilizing secondary data from the official publication of the Central Statistics Agency of Central Java Province and the Health Profile of Central Java Province in 2025. The units of analysis consisted of 35 districts/cities in Central Java Province. The use of regional aggregate data was considered appropriate because the objective of the study was to examine variations in the number of stunting cases across districts/cities and to identify regional factors associated with these differences [25].

The response variable in this study was the number of stunting cases (Y). Because information on the population of children at risk was not consistently available for all districts/cities in the study dataset, the analysis focused on modeling the observed number of stunting cases rather than incidence rates. Consequently, the results should be interpreted as associations with regional differences in stunting case counts and not as estimates of individual level risk.

The predictor variables used consisted of five variables, namely the percentage of babies with low birth weight (X1), the percentage of babies receiving exclusive breastfeeding (X2), the percentage of dependency burden (X3), the percentage of poor population (X4), and the ratio of doctors (X5). The doctor ratio variable represents physician density per 1000 population. Prior to modeling, this variable was multiplied by 100 to reduce scale differences among predictors and improve numerical stability during parameter estimation. This transformation only affects the measurement scale and does not change the substantive relationship between variables in the regression model.

The selection of these variables was based on epidemiological theories of stunting and previous empirical evidence indicating thath biological, socioeconomic, demographic, and healthcare access factors contribute to regional variation in stunting outcomes [26]. Because all variables were measured at the district/city level, the findings should be interpreted as ecological associations between regional characteristics and stunting case counts rather than casual relationship at the individual level.

Table 1. Variables in Research

Variable	Description	Measurement Unit
Y	Number of stunting cases	Number of cases
X1	Percentage of BBLR babies	Percent (%)
X2	Percentage of babies get exclusive breastfeeding	Percent (%)
X3	Percentage of dependency load numbers	Percent (%)
X4	Percentage of poor population	Percent (%)
X5	Doctor ratio	Physicians' density per 1000 population (scaled × 100)

2.2 Poisson Regression Model

The response variable in this study is the number of stunting cases across districts/cities in Central Java Province, which constitutes count data. Therefore, Poisson regression was employed as an initial count data modeling framework and as the basis for the model selection procedure. The analysis focuses on the observed number of reported stunting cases at the district/city level rather than estimating individual-level prevalence rates. Because information on the population of children at risk was not consistently available for all districts/cities in dataset, an offset term was not incorporated into the model.

If the number of stunting cases in the (i)-th district/city is denoted by (Y_i), then the response variable is assumed to follow a Possion distribution, which can be expressed as follow:

$$Y_i \sim Poisson\ (\lambda_i) \tag{1}$$

The probability distribution function of the Poisson model is expressed as:

$$P(Y_i = y_1) = \frac{e^{-\lambda} \lambda_i^{y_i}}{y_i},\ \ y_i = 0,1,2,\ldots \tag{2}$$

The Poisson distribution has the property that the mean and variance are equal, which can be expressed as follows:

$$E(Y_i) = Var(Y_i) = \lambda_i \tag{3}$$

This assumption in known as equidispersion and represents main characteristic of the Poisson regression model. When the variance substantially exceeds the mean, the equidispersion assumption may be violated, resulting in overdispersion and potentially unreliable statistical inference. In such situations, alternative count data models, including Negative Binomial regression, may provide a more appropriate representation of the data [27].

2.3 Negative Binomial Regression Model

The initial Poisson regression analysis indicated substantial overdispersion, suggesting that the variance of the response variable exceeded its mean. Under such conditions, the standard Poisson regression model may underestimate standard errors and produce unreliable statistical inference. Therefore, a Negative Binomial regression model was subsequently employed because it extends the Poisson model by incorporating an additional dispersion parameter to account for extra Poisson variability [28].

If Y_i denotes the number of stunting cases in the i-th district/city, the Negative Binomial model assumes:

$$Y_i \sim NB(\mu_i, \theta) \tag{4}$$

where μ_i represents the expected number of stunting cases and θ denotes the dispersion parameter. The mean structure is modeled using the log-link function:

$$log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} \tag{5}$$

with:
μ_i: Expected number of stunting cases in district/city-i
β_0: Intercept
β_j: Regression Coefficient
X_j: The j-th predictor variable
Unlike the Poisson model, the Negative Binomial model allows the variance to exceed the mean through the following variance following variance function:

$$Var(Y_i) = \mu_i + \frac{\mu_i^2}{\theta} \tag{6}$$

where larger values of θ indicate lower levels of overdispersion. Parameter estimation was performed using Maximum Likelihood Rstimation (MLE). The Negative Binomial model was subsequently evaluated using residual diagnostics, dispersion statistics, and information criteria to determine its suitability for modeling stunting count data.

2.4 Model Selection and Goodness of Fit

To identify a parsimonious count data model, this study evaluated all 31 possible combinations of predictor variables, ranging from one to five predictors. The initial model selection process was conducted using Poisson regression because it represents the standard modeling framework for count data. Model comparison was based on the Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc). Because the dataset consisted of only 35 observations, AICc was additionslly considered to reduce the potential bias of AIC under small sample conditions. Models with samller AIC and AICc values were considered to provide a better balance between model fit and model complexity [29].

$$AIC = -2\log L \ + 2k \tag{7}$$

where (L) denotes the likelihood value and (k) represents the number of estimated parameters in the model.
The corrected Akaike Information Criterion (AICc) was calculated as:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \tag{8}$$

where (n) denotes the sample size and (k) represents the number of estimated parameters.
In addition to information criteria, model adequacy was assessed using the Pearson Chi-Square dispersion statistics:

$$\phi = \frac{X^2}{df} \tag{9}$$

where X^2 denotes the Pearson Chi-Square statistic and (df) represents the residual degrees of freedom. Values of ϕ close to one indicat approximate equidispersion, whereas values substantially greater than one indicate overdispersion.

When substantial overdispersion was detected in the selected Poisson model, the model was not used as the primary basis for statistical inference. Instead, a Negative Binomial regression model was estimated because it accommodates extra Poisson variation through an additional dispersion parameter. The Poisson and Negative Binomial models were subsequently cmpared using AIC, AICc, residual deviance, and dispersion statistics. The model demonstrating better statistical adequacy and more reliable inference under the observed data conditions was selected as the final model for interpretation and discussion.

2.5 Software and Analysis Procedure

All statistical analyses were performed using IBM SPSS Statistics *26* and R software version 4.4.1. IBM SPSS Statistics 26 was used for preliminary data analysis, estimation of the initial Poisson regression models, significance testing, and goodness of fit assessment. The exhaustive subset selection procedure, model comparison, and Negative Binomial regression analyses were conducted using R version 4.4.1.

Several R packages were utilized throughout the analysis. The stats package was used for Poisson regression modeling, prediction, confidence interval estimation, and information criterion calculations, while the MASS package was employed for Negative Binomial regression estimation. Data manipulation, model comparison, and visualization procedures were implemented using standard R functions and additional supporting packages.

The analysis procedure consisted of descriptive statistical analysis, estimation of Poisson regression models, exhaustive subset selection based on AIC and AICc, assessment of goodness of fit, A and evaluation of overdispersion using Pearson Chi-Square statistics. When substantial overdispersion was detected, a Negative Binomial regression model was subsequently estimated. Model comparison was then performed using AIC, AICc, residual deviance, and dispersion statistics to identify the most appropriate model for statistical inference and interpretation.

In addition, graphical model evaluation was conducted through visual inspection of observed versus predicted values and residual diagnostic plots. These visualizations were used to assess predictive performance and to identify potential indications of model misspecification or unexplained variability in the data.

The final interpretation of predictor effects was based on the model that demonstrated better statistical adequacy and more reliable inference the observed data conditions. When overdispersion was present, greater emphasis was placed on the Negative Binomial regression model because it explicity accommodates extra Poisson variation through an additional dispersion parameter, thereby reducing the risk of underestimated standard errors and inflated statistical significance.

3. RESULT AND ANALYSIS

3.1 Descriptive Statistics of Stunting Cases

A descriptive analysis was conducted to examine the distribution of stunting cases across districts/cities in Central Java Province in 2025. This preliminary analysis was intended to identify regional variations in stunting burden before the statistical modeling process was performed. Considering the substantial variation observed across districts/cities, further count data modeling was conducted to investigate factors associated with regional differences in stunting cases.

Table 2. Descriptive Statistics of Research Variables

Variable	Mean	Variance
Number of stunting cases	5368,5714	12225702,61
Percentage of BBLR babies	6,3086	2,927
Percentage of babies get exclusive breastfeeding	69,9089	175,501
Percentage of dependency load numbers	44,3760	2,867
Percentage of poor population	9,1871	7,962
Doctor ratio	27,3086	265,816

Based on Table 2, the number of stunting cases exhibits substantial variation across districts/cities, as indicated by the relatively large variance compared with the average number of cases. This finding suggests that the distribution of stunting cases is not homogeneous across regions and may reflect differences in demographic conditions, socioeconomic characteristics, and access to healthcare services. Similar regional heterogeneity is commonly observed in public health data, where disease or nutritional outcomes often vary considerably between geographic areas [30].

The predictor variables also demonstrate variability among districts/cities. The percentage of BBLR infants, exclusive breastfeeding coverage, dependency burden ratio, poverty rate, and doctor ratio all differ across regions, indicating the presence of diverse health and socioeconomic conditions throughout Central Java Province. Such variation provides a basis for investigating whether these factors are associated with differences in the number of stunting cases across districts/cities.

Considering the substantial variation observed in both the response and predictor variables, further count-data modeling was conducted to identify factors associated with regional differences in stunting cases. Poisson regression was initially employed as the standard approach for count data, followed by additional model evaluation procedures to assess its suitability for the observed data characteristics.

3.2 Preliminary Poisson Regression Analysis

Poisson regression was initially employed because the response variable, namely the number of stunting cases, represents count data. Under the Poisson framework, the response variable is assumed to follow a Poisson distribution, and the relationship between the expected number of stunting cases and the predictor variables is modeled through a log-link function.

$$Y_i \sim Poisson\ (\lambda_i) \tag{10}$$

with probability function:

$$P(Y_i = y_1) = \frac{e^{-\lambda}\lambda_i^{y_i}}{y_i} \tag{11}$$

The relationship between the independent variable and the expected value is expressed in the log link function:

$$log(\lambda_i) = \beta_0 + \sum_{j=1}^{5} \beta_j X_j \tag{12}$$

Based on the results of the estimate, the model was obtained:

$$log(\lambda_i) = 4{,}436 + 0{,}034X_1 + 0{,}015X_2 + 0{,}060X_3 + 0{,}056X_4 - 0{,}014X_5 \tag{13}$$

Table 3. Initial Poisson Regression Parameter Estimates

Variable	B	Wald Chi-Square	Sig.
Percentage of BBLR babies	0,034	319,676	0,000
Percentage of babies get exclusive breastfeeding	0,015	4421,073	0,000
Percentage of dependency load numbers	0,060	843.844	0,000
Percentage of poor population	0,056	2605,863	0,000
Doctor ratio	-0,014	3323,361	0,000

The initial Poisson regression results indicate statistically significant coefficients for all predictor variables. Positive coefficients were obtained for the percentage of low-birth-weight infants, exclusive breastfeeding coverage, dependency burden ratio, and poverty rate, whereas the doctor ratio exhibited a negative coefficient. These coefficient estimates suggest potential associations between the explanatory variables and the number of stunting cases across districts/cities.

However, these results should be interpreted cautiously because the validity of statistical inference under the Poisson model depends on the equidispersion assumption, namely that the variance is approximately equal to the mean. When this assumption is violated, standard errors may be underestimated and significance tests may become unreliable [31].

Subsequent model evaluation revealed substantial overdispersion in the data, indicating that the observed variability considerably exceeded the level assumed by the standard Poisson distribution. Consequently, although the Poisson model remained useful for exploratory model selection and preliminary assessment of predictor relationships, it was not considered the final inferential model. A formal assessment of model adequacy and overdispersion is presented in the following section, followed by the evaluation of an alternative Negative Binomial specification

3.3 Assessment of Poisson Model Fit and Overdispersion

The adequacy of the initial Poisson regression model was evaluated using the Pearson Chi-Square statistic. A commonly used diagnostic for count-data models is the ratio between the Pearson Chi-Square statistic and the residual degrees of freedom. Under the Poisson assumption, this ratio is expected to be close to one, whereas substantially larger values indicate the presence of overdispersion [32].

Table 4. Poisson Model Fit Assessment

Statistics	Value	df	Value/df
Pearson Chi-Square	51434,097	29	1773,590

Based on Table 4, the Pearson Chi-Square statistic was 51,434.097 with 29 residual degrees of freedom, resulting in a dispersion ratio of 1773.590. This value is substantially greater than one and provides strong evidence of severe overdispersion in the stunting count data. The result indicates that the observed variability greatly exceeds the variability assumed under the standard Poisson distribution.

The presence of severe overdispersion suggests that the equidispersion assumption of the Poisson model was violated. Under such conditions, the standard Poisson regression model may underestimate standard errors, inflate statistical significance, and produce unreliable inferential conclusions. Consequently, although the Poisson model remained useful for exploratory model selection purposes, it was not considered sufficiently adequate as the final inferential model.

These findings motivated the evaluation of an alternative count-data specification capable of accommodating extra-Poisson variation. Therefore, a Negative Binomial regression model was subsequently estimated and compared with the Poisson model to obtain more reliable statistical inference and model interpretation.

3.4 Model Evaluation and Selection

Model evaluation and exploratory subset selection were conducted using the Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc). Because the dataset consisted of only 35 districts/cities, AICc was preferred over standard AIC to reduce potential small-sample bias during model comparison [33].

$$AIC = -2ln\,(L) + 2k \tag{14}$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \tag{15}$$

where L denotes the likelihood value, k is the number of estimated parameters, and n is the sample size. Models with smaller AICc values were considered to provide a better balance between goodness of fit and model complexity.

Table 5. AICc-Based Comparison of Candidate Poisson Models

Best Predictor Combination	Residual df	AIC	AICc
X_4	33	56601,51	56601,89
X_3, X_4	32	53278,63	53279,41
X_2, X_4, X_5	31	49027,00	49028,33
X_2, X_3, X_4, X_5	30	46450,48	46452,55
X_1, X_2, X_3, X_4, X_5	29	46131,91	46134,91

Based on Table 5, models containing larger numbers of predictor variables generally produced smaller AICc values. Among the evaluated Poisson candidate models, the five-predictor model yielded the lowest AICc value (46134.91), indicating the best relative fit within the Poisson modeling framework.

However, the model adequacy assessment presented in Section 3.3 revealed severe overdispersion, with a Pearson Chi-Square to residual degree-of-freedom ratio of 1773.59. This result indicates that the variability of the stunting count data substantially exceeded the assumptions of the standard Poisson distribution. Consequently, although the AICc-based subset selection remained useful for identifying potentially relevant predictor combinations, the resulting Poisson models were considered exploratory rather than inferential.

Therefore, the Poisson model selection results were retained primarily to guide variable assessment and model comparison, while subsequent statistical inference and interpretation were based on a Negative Binomial regression model, which is more appropriate for overdispersed count data.

3.5 Visualization of Negative Binomial Model Results

To further evaluate the adequacy of the Negative Binomial Regression model in explaining the number of stunting cases in Central Java Province, diagnostic visualizations were generated. These visualizations include a comparison between observed and predicted values, the distribution of Pearson residuals, and the relationship between residuals and fitted values.

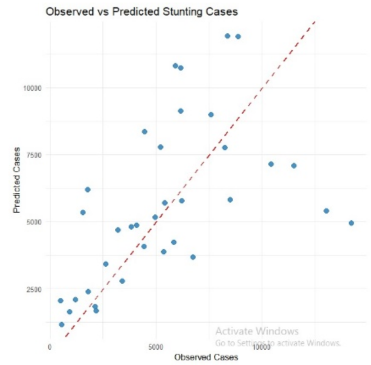

Figure 1. Observed versus Predicted Stunting Cases

Figure 1 compares the observed stunting cases with the values predicted by the Negative Binomial Regression model. The red dashed line represents the ideal condition where predicted values are equal to observed values. Most observations are located relatively close to this reference line, indicating that the model is capable of capturing the general pattern of variation in stunting cases across districts and cities in Central Java.

However, several observations are noticeably distant from the reference line. In particular, some regions with observed cases above 9,000 exhibit predicted values that are considerably lower than the actual values. This suggests that the model tends to underestimate stunting cases in a few high-incidence regions. Despite these deviations, the overall distribution of points around the diagonal line indicates a reasonably good agreement between observed and predicted values, supporting the suitability of the Negative Binomial model for the data.

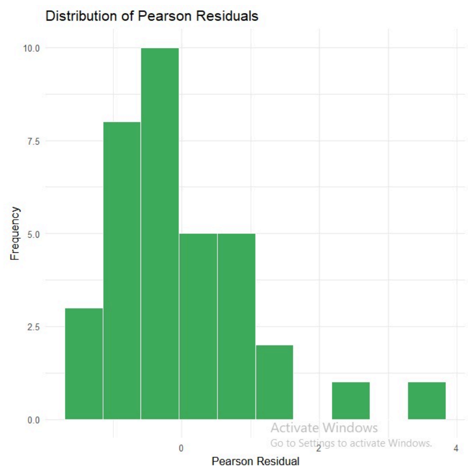

Figure 2. Distribution of Pearson Residuals

Figure 2 presents the distribution of Pearson residuals obtained from the Negative Binomial Regression model. The histogram shows that most residuals are concentrated around zero, with values generally ranging between -1 and 1. This pattern indicates that the prediction errors are relatively small for the majority of observations.

The residual distribution appears slightly right-skewed due to the presence of a few large positive residuals. These positive residuals correspond to observations where the actual number of stunting cases exceeds the predicted values. Nevertheless, the concentration of residuals near zero suggests that the model does not systematically overestimate or underestimate the response variable. Therefore, the residual distribution provides evidence that the Negative Binomial model adequately represents the stunting data.

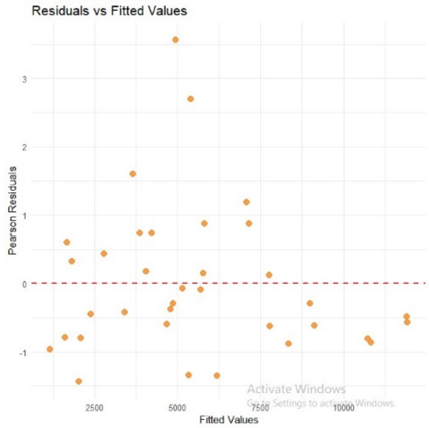

Figure 3. Residuals versus Fitted Values

Figure 3 illustrates the relationship between Pearson residuals and fitted values. The residuals are scattered around the horizontal reference line at zero without forming a clear trend or systematic pattern. This random distribution indicates that the model assumptions are reasonably satisfied and that the fitted values do not exhibit substantial bias across different levels of predicted stunting cases.

Most residuals lie within the interval from approximately -1.5 to 1.5, although a small number of observations display relatively large positive residuals exceeding 2.5. These observations indicate districts or cities where the model underestimates the actual number of stunting cases. Despite the presence of these outlying residuals, no funnel-shaped pattern or strong curvature is observed, suggesting that the variance structure of the Negative Binomial model is appropriate for handling the overdispersed count data.

Overall, the diagnostic visualizations demonstrate that the Negative Binomial Regression model provides a satisfactory fit to the stunting data. The observed and predicted values show reasonable agreement, the residuals are centered around zero, and no substantial systematic pattern is evident in the residual plots. These findings support the use of the Negative Binomial Regression model for identifying factors associated with stunting cases in Central Java Province.

3.6 Negative Binomial Regression Analysis

The Poisson regression model initially estimated in this study indicated severe overdispersion, as evidenced by the extremely large Pearson Chi-Square to residual degrees of freedom ratio. This result suggests that the variance of the stunting count data substantially exceeded its mean, violating the equidispersion assumption required by the Poisson distribution. Consequently, the Poisson model was not considered suitable as the primary inferential model.

To accommodate the observed overdispersion, a Negative Binomial regression model was estimated using the same explanatory variables, namely the percentage of BBLR babies, percentage of babies receiving exclusive breastfeeding, percentage of dependency load numbers, percentage of poor population, and doctor ratio. The Negative Binomial model was subsequently adopted as the primary inferential model because it incorporates an additional dispersion parameter that allows the variance to exceed the mean.

The response variable was modeled as the observed number of stunting cases at the district/city level in Central Java Province. The objective of the study was to examine regional variation in reported stunting counts rather than estimating individual-level risk or prevalence rates. Therefore, the analysis focused on aggregated count data across districts/cities. The estimation results of the Negative Binomial regression model are presented in Table 6.

Table 6. Negative Binomial Regression Estimation Results.

Variable	Estimate	Std. Error	z-value	p-value	IRR	95% CI Lower	95% Upper
Intercept	0,4830	3,2459	0,149	0,8817	1,6209	0,0012	2741,4424
Percentage of BBLR babies	-0,0341	0,0703	-0.485	0,6277	0,9665	0,8233	1,1329
Percentage of babies get exclusive breastfeeding	0,0255	0,0078	3,278	0,0011	1,0258	1,0090	1,0426
Percentage of dependency load numbers	0,1376	0,0787	1,748	0,0805	1,1475	0,9628	1,3633
Percentage of poor population	0,0696	0,0442	1.575	0,1152	1,0721	0,9877	1,1654
Doctor ratio	-0,0121	0,0068	-1,795	0,0726	0,9880	0,9742	1.0026

The estimation results indicate that the effects of stunting determinants varied across districts and cities in Central Java Province. Previous studies have reported that stunting is influenced by a combination of nutritional, demographic, socioeconomic, and healthcare-related factors whose effects may differ across regions [34]. Among the explanatory variables, the percentage of babies receiving exclusive breastfeeding remained statistically significant at the 5% significance level ($p = 0.0011$). In addition, the percentage of dependency load numbers and doctor ratio exhibited marginal significance at the 10% significance level. The remaining explanatory variables did not demonstrate statistically significant effects after overdispersion was accommodated through the Negative Binomial specification.

To evaluate model adequacy, both the Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc) were considered. While AIC is widely used for comparing competing statistical models, AICc includes an additional correction for finite sample sizes and is therefore recommended when the number of observations is relatively limited compared with the number of estimated parameters.

The Negative Binomial regression model produced substantially smaller information criterion values than the Poisson regression model. The AIC value decreased from 46131.91 under the Poisson model to 654.01 under the Negative Binomial model, while the AICc value decreased from 46134.91 to 657.01. These reductions indicate a substantial improvement in model fit after accounting for overdispersion.

Furthermore, the dispersion statistic decreased from 1773.59 under the Poisson model to 1.37 under the Negative Binomial model. Since a dispersion value close to one generally indicates an adequate fit for count-data models, this result suggests that the overdispersion problem was largely resolved through the incorporation of the additional dispersion parameter.

The improvement in model performance is consistent with previous studies that have demonstrated the suitability of Negative Binomial regression for analyzing overdispersed count data. By allowing the variance to exceed the mean, the model provides greater flexibility in variance estimation and yields more reliable statistical inference than standard Poisson regression [35][36].

The statistical findings are further supported by the diagnostic visualizations presented in Section 3.5. The observed-versus-predicted plot showed reasonable agreement between observed and fitted values, while the residual analyses indicated no substantial systematic pattern. Together with the reduced dispersion statistic, these diagnostic results provide additional evidence that the Negative Binomial model adequately represents the variability present in the stunting data.

Overall, the Negative Binomial regression model demonstrated superior statistical performance compared with the Poisson regression model. The substantial reductions in AIC, AICc, and dispersion statistics indicate that the model effectively accommodated overdispersion and provided a more appropriate framework for analyzing district/city-level stunting counts in Central Java Province. Therefore, the subsequent interpretation and discussion of stunting determinants were based primarily on the Negative Binomial regression results

3.7 Regional Level Interpretation of the Negative Binomial Regression Results

Based on the Negative Binomial regression model presented in Table 6, the interpretation of stunting determinants was conducted using both the estimated coefficients and the incidence rate ratios (IRRs). Because the analysis was performed using district/city-level aggregated data, the findings should be interpreted as regional-level associations rather than individual-level causal effects.

Among the explanatory variables, the percentage of babies receiving exclusive breastfeeding was the only variable that remained statistically significant at the 5% significance level. The estimated coefficient was 0.0255, corresponding to an IRR of 1.0258 with a 95% confidence interval ranging from 1.0090 to 1.0426. This result indicates that a one-percentage-point increase in exclusive breastfeeding coverage was associated with an expected

2.58% increase in the reported number of stunting cases, holding the other variables constant. Since the confidence interval does not include one, the association is statistically significant. However, this finding should be interpreted cautiously because the analysis was conducted at the regional level. The observed association does not imply that exclusive breastfeeding increases the risk of stunting at the individual level. Instead, it may reflect the complexity of aggregated regional data and the influence of unobserved contextual factors not captured in the current model.

Previous studies have consistently highlighted the importance of exclusive breastfeeding in supporting child growth, improving nutritional status, and reducing the risk of adverse health outcomes during early childhood [37]. Therefore, the positive association identified in the present study should be interpreted as an ecological relationship observed across districts and cities rather than as evidence contradicting the established benefits of exclusive breastfeeding.

The percentage of dependency load numbers exhibited a positive coefficient with an IRR of 1.1475 and a 95% confidence interval of 0.9628–1.3633. Although the estimated effect suggests that higher dependency burdens were associated with higher numbers of stunting cases, the confidence interval includes one and the variable was only marginally significant. Consequently, the evidence is insufficient to conclude a statistically significant relationship at the 5% significance level. Nevertheless, the direction of the association is consistent with the notion that greater dependency burdens may place additional pressure on household and community resources that support child nutrition and healthcare.

Similarly, the doctor ratio showed a negative coefficient with an IRR of 0.9880 and a 95% confidence interval of 0.9742–1.0026. This result suggests that districts and cities with greater availability of medical personnel tended to report slightly lower numbers of stunting cases. Although the confidence interval narrowly includes one, the negative direction of the association is consistent with public health expectations, indicating that improved healthcare accessibility may contribute to better nutritional monitoring, maternal healthcare, and child growth surveillance.

The percentage of BBLR babies and the percentage of poor population did not demonstrate statistically significant effects after overdispersion was accommodated through the Negative Binomial specification. Their confidence intervals also included one, indicating substantial uncertainty regarding the magnitude of their effects within the current dataset. Nevertheless, previous studies have emphasized that stunting is a multidimensional public health issue influenced by demographic, nutritional, socioeconomic, environmental, and healthcare-related factors whose impacts may vary considerably across geographical regions [38]. Therefore, the absence of statistical significance in the present analysis should not be interpreted as evidence that these variables are unimportant determinants of stunting.

It is important to emphasize that the model was estimated using aggregated district/city-level counts of stunting cases. Consequently, the estimated relationships represent ecological associations and should not be generalized directly to individual children. Additional studies using household-level or longitudinal data would be necessary to identify causal mechanisms and evaluate individual-level risk factors more precisely.

Overall, the findings suggest that variations in stunting cases across Central Java Province are associated with complex interactions among nutritional, demographic, and healthcare-related conditions. These results support the need for integrated stunting reduction strategies involving nutrition improvement programs, maternal and child healthcare services, healthcare accessibility, and continuous growth monitoring. Such multidimensional interventions are likely to be more effective than policies focusing on a single determinant of stunting.

4. CONCLUSION

This study analyzed district/city-level stunting cases in Central Java Province using count data regression models. The initial Poisson regression analysis indicated severe overdispersion, suggesting that the equidispersion assumption was violated and that the model was not appropriate as the final inferential framework. Consequently, a Negative Binomial regression model was adopted because it provides greater flexibility for modeling overdispersed count data. The Negative Binomial model demonstrated substantially better performance, as indicated by lower AIC and AICc values, a dispersion statistic closer to one, and diagnostic results showing adequate agreement between observed and predicted values.

Based on the Negative Binomial regression model, exclusive breastfeeding coverage was the only variable that remained statistically significant at the 5% significance level, with an incidence rate ratio (IRR) of 1.0258 (95% CI: 1.0090–1.0426). Meanwhile, the dependency ratio and doctor ratio exhibited marginal significance and may contribute to regional variation in stunting cases. The findings highlight that appropriate model selection and careful assessment of overdispersion are essential for obtaining reliable statistical inference from count data. Furthermore, the results suggest that stunting prevention efforts should be supported by integrated nutrition, healthcare, and community-based interventions to address the complex factors associated with stunting across districts and cities in Central Java Province.

5. REFERENCES

[1] Y. D. Purwani, T. Yuwono, E. Rimawati, S. Handayani, S. Wiwoho, and D. C. Kharisma, "Determinants of Childhood Stunting in Central Java , Indonesia : The Role of Maternal Mental Health and Child Health History," 2025, https://doi.org/10.1177/30502225251401695.

[2] I. H. Sumon *et al.*, "Determinants of stunting among under-five children: Evidence from Cambodian Demographic and Health Survey 2021–2022," *Child: Care, Health and Development*, vol. 50, no. 4, p. e13291, 2025.

[3] H. F. A. Subratha and K. T. S. Agustia, "Determinants of stunting among toddlers: A systematic literature review," *WMJ (Warmadewa Medical Journal)*, vol. 9, no. 2, pp. 50–56, 2024.

[4] Priatmadani, H. U. Anjarweni, S. M. Putri, A. S. Pramana, R. Palupi, and Budiasih, "Determinan Prevalensi Balita Stunting di Indonesia Tahun 2021 Pendekatan Model Spasial," pp. 653–662, 2021.

[5] T. T. Tamir, B. Tekeba, E. G. Mekonen, A. F. Zegeye, and D. A. Gebrehana, "Spatial heterogeneity and predictors of stunting among under five children in Mozambique : a geographically weighted regression," no. December, pp. 1–17, 2024, https://doi.org/10.3389/fpubh.2024.1502018.

[6] S. Appelbaum, J. Stronski, and U. Konerding, "The Use of Double Poisson Regression for Count Data in Health and Life Science — A Narrative Review," pp. 1–12, 2025.

[7] J. Xu, Y. Lu, Y. Su, T. Liu, Y. Qi, and W. Xie, "Intrinsic Functional Partially Linear Poisson Regression Model for Count Data," pp. 1–19, 2024.

[8] M. Roccetti and G. Cacciapuoti, "Beyond the Gold Standard : Linear Regression and Poisson GLM Yield Identical Mortality Trends and Deaths Counts for COVID-19 in Italy : 2021 – 2025," pp. 2021–2025, 2025.

[9] M. Bonamente, "Linear regression for Poisson count data : a new semi-analytical method with applications to COVID- events".

[10] G. A. Fernandez and K. P. Vatcheva, "A comparison of statistical methods for modeling count data with an application to hospital length of stay," *BMC Medical Research Methodology*, pp. 1–21, 2022, https://doi.org/10.1186/s12874-022-01685-8.

[11] I. E. Okorie, E. Afuecheta, S. Nadarajah, A. Bright, and A. C. Akpanta, "OPEN A Poisson regression approach for assessing morbidity risk and determinants among under five children in Nigeria," *Scientific Reports*, pp. 1–16, 2024, https://doi.org/10.1038/s41598-024-72373-4.

[12] E. Nugroho, P. A. Wanti, C. W. Suci, B. B. Raharjo, and Najib, "Social Determinants of Stunting in Indonesia," *National Seminar on Official Statistics*, vol. 18, no. 4, pp. 546–555, 2023.

[13] A. D. V. Id *et al.*, "Predictors of infection , symptoms development , and mortality in people with SARS-CoV-2 living in retirement nursing homes," pp. 1–14, 2021, https://doi.org/10.1371/journal.pone.0248009.

[14] J. K. Id and S. K. Mohanty, "Poverty induced inequality in nutrition among children born during 2010 – 2021 in India," pp. 1–29, 2024, https://doi.org/10.1371/journal.pone.0313596.

[15] O. Dadras, C. Suwanbamrung, M. Jafari, and M. H. Stanikzai, "Prevalence of stunting and its correlates among children under 5 in Afghanistan : the potential impact of basic and full vaccination," pp. 1–10, 2024.

[16] S. Sukmawati, Y. Hermayanti, E. Fadlyana, I. Maulana, and H. S. Mediani, "Health cadres ' experiences in detecting and preventing childhood stunting in Indonesia : a qualitative study," 2025.

[17] K. F. Sellers and G. Shmueli, "A flexible regression model for count data," *Annals of Applied Statistics*, vol. 4, no. 2, pp. 943–961, 2010, https://doi.org/10.1214/09-AOAS306.

[18] J. Zhang, "Information criteria for model selection," no. August 2022, 2023, doi: 10.1002/wics.1607.

[19] O. Efthimiou, M. Seo, K. Chalkou, T. Debray, M. Egger, and G. Salanti, "Developing clinical prediction models : a step-by-step guide," pp. 1–15, https://doi.org/10.1136/bmj-2023-078276.

[20] K. Yu and T. Tao, "Consistent Model Selection Procedure for Random Coefficient," no. 1, pp. 1–29, 2023.

[21] H. I. Ayue, N. Nurdiana, V. Wardhani, A. B. Astuti, H. Prayitno, and A. D. Laksono, "Exploring the multifactorial predictors of stunting in children under five : A systematic review of the literature , 2015 – 2024," 2025, https://doi.org/10.1177/22799036251403945.

[22] C. Sutherland, D. Hare, P. J. Johnson, D. W. Linden, R. A. Montgomery, and E. Droge, "Practical advice on variable selection and reporting using Akaike information criterion," no. Dic, 2023.

[23] S. L. Sclove, "Using Model Selection Criteria to Choose the Number of Principal Components," *Journal of Statistical Theory and Applications*, vol. 20, no. 3, pp. 450–461, 2021, https://doi.org/10.1007/s44199-021-00002-4.

[24] S. Sen and A. S. Cohen, "An Evaluation of Fit Indices Used in Model Selection of Dichotomous Mixture IRT Models," 2024, https://doi.org/10.1177/00131644231180529.

[25] Y. Chen *et al.*, "Predicting Arm Nonuse in Individuals with Good Arm Motor Function after Stroke Rehabilitation : A Machine Learning Study," pp. 1–12, 2023.

[26] B. Faso *et al.*, "Process of developing models of maternal nutrition interventions integrated into antenatal care services," no. December 2021, pp. 1–18, 2022, https://doi.org/10.1111/mcn.13379.

[27] O. I. Guneri and B. Durmus, "Models fors Overdispersion Count Data with Generalized Distribution: An Application to Parasites Intensity," *Journal of New Theory*, vol. 35, no. 35, pp. 48–61, 2021.

[28] Y. Tiara, M. N. Aidi, and R. Rachmawati, "Overdispersion Handling in Poisson Regression Model by Applying Negative Binomial Regression," vol. 17, no. 1, pp. 417–426, 2023.

[29] R. U. T. Esra, D. A. Nohe, and M. Fathurahman, "Pemilihan Model Terbaik pada Generalized Poisson Regression Menggunakan Akaike Information Criterion," vol. 23, no. 1, pp. 73–87, 2023.

[30] Z. Wang and Q. Li, "Spatial heterogeneity in public health count data: Implications for modeling disease incidence," *International Journal of Environmental Research and Public Health*, vol. 20, no. 5, p. 4123, 2023, https://doi.org/10.3390/ijerph20054123.

[31] W. Lee, J. Kim, and D. Lee, "Revisiting the analysis pipeline for overdispersed Poisson and," vol. 50, no. 7, pp. 1455–1476, 2023.

[32] Y. Hagiwara and Y. Matsuyama, "Goodness-of-fit tests for modified Poisson regression possibly producing fitted values exceeding one in binary outcome analysis," *Statistical Methods in Medical Research*, vol. 33, no. 7, pp. 1185–1196, 2024, https://doi.org/10.1177/09622802241254220.

[33] J. T. Behan, C. T. Hodgdon, and Y. Chen, "Scale-dependent assumptions influence habitat suitability estimates for the American lobster (Homarus americanus): Implications for a changing Gulf of Maine," *Ecological Modelling*, vol. 470, no. May, p. 110030, 2022, https://doi.org/10.1016/j.ecolmodel.2022.110030.

[34] D. Kurniawati, S. P. Hastono, I. Safika, and W. Wahyuningsih, "Geographically Weighted Regression Model of Stunting Determinants in Indonesia," vol. 10, pp. 140–152, 2025.

[35] Y. Liu, J. Zhou, Z. Chen, and X. Zhang, "A general averaging method for count data with overdispersion and/or excess zeros in biomedicine," *Statistical Methods in Medical Research*, vol. 32, no. 5, pp. 904–926, 2023, https://doi.org/10.1177/09622802231159213.

[36] B. Sidumo, E. Sonono, and I. Takaidza, "Count Regression and Machine Learning Techniques for Zero-Inflated Overdispersed Count Data : Application to Ecological Data," *Annals of Data Science*, vol. 11, no. 3, pp. 803–817, 2024, https://doi.org/10.1007/s40745-023-00464-6.

[37] D. Susanti, A. W. Doni, R. A. Yuda, and M. Yanti, "Determining the Role of Exclusive Breastfeeding and Nutritional Status in Stunting Prevention : A Literature Review Kemenkes Padang , Padang City , West Sumatra , Indonesia Mohammad Husni Thamrin , Jakarta , Indonesia 3 Sekolah Tinggi Ilmu Kesehatan Alifah Padang , Padang City , West Sumatra , Indonesia 2 Universitas 1 Poltekkes (Correspondence author email , dewisusanti.bd@gmail.com)," vol. 17, no. 3, pp. 1014–1025, 2023.

[38] T. Astika, E. Permatasari, Y. Chadirin, F. Elvira, and B. A. Putri, "The association of sociodemographic , nutrition , and sanitation on stunting in children under five in rural area of West Java Province in Indonesia," 2023, https://doi.org/10.1177/22799036231197169.

ACKNOWLEDGEMENT

The authors are very grateful to The Institute for Research and Community Service (LPPM) Universitas Sebelas Maret Surakarta under grant Penguatan Kapasitas Grup Riset (PKGR-UNS) B 2026.