Mathematical Journal of Modelling and Forecasting
Vol. 2, No. 2, December 2024, pp. 20-29

Modelling the Number of Stunting Cases in Indonesia in 2022 Using Negative Binomial Regression to Address Overdispersion

Cinta Rizki Oktarina[1*] and Reza Pahlepi[1]

[1] Department of Statistics, Universitas Bengkulu, Indonesia
[*]cintarizki199@gmail.com

Abstract. This study models the incidence of stunting in toddlers in Indonesia in 2022 using negative binomial regression to address the overdispersion issue often present in count data. The Poisson regression model, typically used for count data, showed less accurate results due to the variance exceeding the mean, indicating overdispersion. By adopting a negative binomial regression approach, this study accommodates higher variability in the data, leading to more accurate estimates. The results reveal that the percentage of pneumonia cases and low birth weight are significant factors in stunting incidence. In contrast, other variables, such as complete basic immunization and poverty levels, are insignificant in the final model. The final negative binomial model yielded a lower AIC value than the initial model, indicating an improved model fit, with an R-squared (Nagelkerke's R^2) of 50.50%. This study offers enhanced insights into the factors influencing stunting, supporting more targeted health policy decisions to reduce stunting rates in Indonesia.

Keywords: Stunting, Negative Binomial Regression, Overdispersion

1 Introduction

Regression analysis is a statistical technique used to explore and model relationships between variables [1]. One type of data suited for regression modeling is count data. Poisson regression is a common approach for modeling count data, assuming a Poisson distribution where the mean and variance are equal [2]. However, when variance exceeds the mean (overdispersion), the Poisson model may be unsuitable, leading to underestimated standard errors and potentially invalid conclusions [3]. Overdispersion can often arise due to positive correlation among response variables, necessitating alternative methods like Negative Binomial regression.

Negative Binomial regression introduces an additional parameter to handle better overdispersion, which assumes the Poisson mean is itself a random variable following a gamma distribution, resulting in the Negative Binomial distribution [4]. This method is widely used in various fields, such as epidemiology and social sciences, due to its flexibility in handling overdispersed and zero-inflated data [5]. Furthermore, Negative Binomial regression is more robust in skewed or asymmetrical data distributions and effectively addresses outliers through a longer-tailed distribution, yielding more accurate estimates compared to Poisson regression [6].

Studies utilizing Negative Binomial regression include [7], which addressed overdispersion in dengue fever data, and [8], which applied Negative Binomial regression to examine factors influencing infant and maternal mortality. This study aims to model stunting incidence factors in Indonesia in 2022, as existing research has not applied Negative Binomial regression to this case.

According to the Indonesian Ministry of Health (2022), stunting is a growth disorder due to chronic malnutrition and long-term infections, marked by slower development compared to peers. Despite annual declines in stunting prevalence (from 27.7% in 2019 to 24.4% in 2021), Indonesia's rates remain above WHO's recommended 20% threshold. The government aims to lower this rate to 19% by 2024. Stunting has both short-term impacts, such as increased susceptibility to illness, and long-term effects, including impaired cognitive development and lower reproductive health.

Numerous interventions have been introduced to address stunting, including the "Isi Piringku" campaign by Indonesia's Ministry of Health, which promotes balanced meals. Additionally, community-based health services can help improve maternal and child health, potentially through health workers or trained volunteers. Despite these efforts, studies such as [9] and [10] have yet to apply Negative Binomial regression to stunting data in Indonesia. Therefore, this study employs Negative Binomial regression to model stunting cases in Indonesia in 2022, offering a robust approach to address overdispersion in count data for more accurate analysis. Based on data from the 2022 National Nutrition Status Survey (SSGI), the stunting prevalence in Indonesia was 21.6%, a decline from 24.4% in the previous year. This study addresses the gap in existing research by employing the method and selecting relevant variables: pneumonia prevalence (X_1), immunization coverage (X_2), low birth weight (X_3), and poverty (X_4). These variables capture critical aspects of health and socio-economic conditions, forming a comprehensive framework to analyze stunting determinants in Indonesia.

2 Theoretical Basic

2.1 Poisson Regression

Poisson Regression is a regression method used for data where the response variable is not normally distributed and is discrete, specifically following a Poisson distribution. For predictor variables, data can be either discrete or continuous. Poisson Regression is a non-linear regression model used for count data where the response variable follows a Poisson distribution [11]. The Poisson distribution is a probability distribution that describes the number of successes in a random experiment. In Poisson regression, it is assumed that the response variable follows a Poisson distribution and that there is no multicollinearity among predictor variables. This model also must meet the equidispersion assumption, meaning the mean of the response variable should equal its variance, with a dispersion parameter value of $\phi = 1$. If the response variable follows a Poisson distribution with a parameter μ, the probability function of the Poisson distribution can be expressed as follows:

$$f(y,\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}. \quad (1)$$

y: Number of events (observations).

λ: Average incidence (Poisson parameter).

2.2 Overdispersion

According to [12], overdispersion in Poisson regression occurs when the variance of the response variable is greater than its mean. Some causes of overdispersion include:

1. Correlation among observations.
2. Violation of the Poisson distribution assumption, where $Var(Y) > E(Y)$.
3. Excess zeros in the data.
4. Presence of outliers in the data.

Overdispersion causes the model's deviance to become very large, making the resulting model less accurate. The deviance value can be calculated using the following formula:

$$D = 2\sum_{i=1}^{n}\left(y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right). \quad (2)$$

One way to address this is by replacing the Poisson distribution assumption with a more flexible distribution. Overdispersion can be handled using negative binomial regression. The negative binomial regression model assumes that the response variable follows a negative binomial distribution. One advantage of negative binomial regression is its ability to be used in both equidispersion and overdispersion conditions [13]. Additionally, because the response variable in negative binomial regression is assumed to follow a negative binomial distribution, this model does not require the variance to be equal to the mean.

2.3 Negative Binomial Regression

As an application of GLM, the negative binomial distribution has three components: the random component, the systematic component, and the link function [14]. To form a regression model, the mean and variance are given as $E(y) = \mu$ and $Var\ (y) = \mu(1 + \theta\mu)$, with the probability mass function of the negative binomial as follows:

$$f(y; \mu, \theta) = \frac{\Gamma\left(y+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y!}\left(\frac{1}{1+\theta\mu}\right)^{\frac{1}{\theta}}\left(\frac{\theta\mu}{1+\theta\mu}\right)^{y}. \quad (3)$$

When $\theta = 0$, the negative binomial distribution has a variance $Var\ (y) = \mu$. The negative binomial distribution will approximate a Poisson distribution, which assumes that the mean and variance are the same, namely $E(y) = Var(y) = \mu$. The exponential family distribution function of the negative binomial distribution [14] is as follows:

$$f(y; \mu, \theta) = \exp\left\{y\ \ln\left(\frac{\theta\mu}{1+\theta\mu}\right) + \frac{1}{\theta}\ln\left(\frac{1}{1+\theta\mu}\right) + \ln\left(\frac{\Gamma\left(y+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)y!}\right)\right\}. \quad (4)$$

The contribution of predictor variables in the negative binomial regression model is expressed as a linear combination of the parameter η with the regression parameters to be estimated, namely[15]:

$$\eta_i = \beta_0 + \sum_{k=1}^{p}\beta_k x_{ik}$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}. \quad (5)$$

where $\boldsymbol{\eta}$ is an $(n \times 1)$vector of observations, $\mathbf{X}$ is an $(n \times c)$ matrix of predictor variables and $\boldsymbol{\beta}$ is a $(c \times 1)$ matrix of regression coefficients with $c = p + 1$. The response variable Y is discrete and positive. Therefore, to transform the value of η_i (a real number) to the appropriate range of the response$Y$, a link function g(.) is needed [14], which is:

$g(\mu_i) = \ln \mu_i$ with $\mu_i = \exp\left(\boldsymbol{x_i^T\beta}\right)$ then

$g(\mu_i) = \ln \exp\left(\boldsymbol{x_i^T\beta}\right)$

$g(\mu_i) = \left(\boldsymbol{x_i^T\beta}\right)$

Parameter estimation of negative binomial regression uses the maximum likelihood method with the Newton-Raphson procedure. This method requires the first and second derivatives of the likelihood function y_i to have a negative binomial distribution probability mass function as follows:

$$f(y_i|\mu_i, \theta) = \frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\theta\mu_i}\right)^{\frac{1}{\theta}}\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i}. \quad (6)$$

Since the functions are mutually independent, the log-likelihood function is

$$L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^{n}\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)\Gamma(y_i+1)}\left(\frac{1}{1+\theta\mu_i}\right)^{\frac{1}{\theta}}\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i}. \quad (7)$$

With $\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)} = \prod_{r=1}^{n}(r + \theta^{-1})$

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

$\ln\{L(\boldsymbol{\beta}, \theta)\} = \sum_{i=1}^{n}[\Lambda]$

$\Lambda = \left(\sum_{r=1}^{y_i-1}\ln(r + \theta^{-1})\right) - \ln(y_i!) + y_i \ln(\theta\mu_i) - (\theta^{-1} + y_i)\ln(1 + \theta\mu_i)$

The first derivative of the log-likelihood function with respect to the coefficient $\boldsymbol{\beta}$ is :

$$\frac{\partial L(\boldsymbol{\beta},\theta)}{\partial \beta_0} = \sum_{i=1}^{n}\left[(y_i + \theta^{-1})\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)\right] = \sum_{i=1}^{n}\left[\frac{y_i-\mu_i}{1+\theta\mu_i}\right] = 0$$

$$\vdots$$

$$\frac{\partial L(\boldsymbol{\beta},\theta)}{\partial \beta_p} = \sum_{i=1}^{n}\left[y_i x_{ip} - (y_i + \theta^{-1})\left(\frac{\theta\mu_i x_{ip}}{1+\theta\mu_i}\right)\right]$$

$$= \sum_{i=1}^{n}\left[\frac{(y_i-\mu_i)x_{ip}}{1+\theta\mu_i}\right]$$

$$= \sum_{i=1}^{n}\left[\frac{\mu_i}{1+\theta\mu_i}\frac{(y_i-\mu_i)x_{ip}}{\mu_i}\right] = 0$$

The steps for estimating parameters in negative binomial regression are as follows:

1. Determine the initial estimate of θ, for example, $\hat{\theta}_i = 0.1$.
2. Determine the maximum likelihood estimate of the parameter $\boldsymbol{\beta}$ using the Fisher scoring iteration procedure with the assumption $\theta = \hat{\theta}_1$:

$$\widehat{\boldsymbol{\beta}}_{i+1} = \widehat{\boldsymbol{\beta}}_i + \left(\mathbf{X^T W_j X}\right)^{-1}\mathbf{X^T W_j z_i}. \quad (8)$$

3. Use $\widehat{\boldsymbol{\beta}}$ to estimate the parameter θ using a single-variable Newton-Raphson iteration procedure; the iteration ends when $|\hat{\theta}_{i+1} - \hat{\theta}_i| \leq \varepsilon$:

$$\hat{\theta}_{i+1} = \hat{\theta}_i - \frac{f'(\theta_i)}{f''(\theta_i)}. \quad (9)$$

4. If $|\hat{\theta}_{i+1} - \hat{\theta}_i| \leq \varepsilon$, the iteration stops; if not, use the parameter $\theta = \hat{\theta}_1$ and return to step 2. Here, ε is a very small positive number.

The parameter testing for the negative binomial regression model is conducted using the Maximum Likelihood Ratio Test (MLRT), which involves calculating two likelihood functions to obtain the test statistic for simultaneous parameter testing. These likelihood functions are as follows:

$L(\widehat{\Omega})$ is the likelihood value involving all predictor variables.

$L(\widehat{\omega})$ is the likelihood value without involving all predictor variables.

The hypotheses are as follows:

$\mathrm{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1 : At least one $\beta_j \neq 0 ; j = 1,2,\cdots,p$

Test statistics:

$$D\left(\widehat{\boldsymbol{\beta}}\right) = -2\ln\Delta$$

$$= -2\ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right)$$

$$= 2\left(\ln L\left(\widehat{\Omega}\right) - \ln L(\widehat{\omega})\right). \quad (10)$$

Where $D\left(\widehat{\boldsymbol{\beta}}\right)$ is the deviance value of the Poisson regression model or the likelihood ratio. This test statistic follows a chi-square distribution with K degrees of freedom [16]. The rejection region for H_0 is if $D\left(\widehat{\boldsymbol{\beta}}\right) > \chi^2_{(\alpha;p)}$. From the results of the Poisson regression model formation, the parameter estimates may not necessarily have a significant effect on the model.

The hypotheses are as follows:

$\mathrm{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1 : At least one $\beta_j \neq 0 ; j = 1,2,\cdots,p$

Test statistics:

$$G = -2\ln\left(\frac{L_0}{L_1}\right) = -2(\ln L_0 - \ln L_1). \quad (11)$$

Therefore, partial parameter testing is needed to examine the significance of each parameter in the model. According to Hocking (1996), the test is as follows:

The hypotheses are as follows:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0; j = 1,2,\cdots,p$

Test statistics:

$$z = \frac{\widehat{\beta_j}}{se(\hat{\beta}_j)}.$$

2.4 Number of Stunting Cases

Stunting is a growth and development disorder in children due to chronic malnutrition and recurrent infections, marked by height or length below the standard set by the ministry responsible for health affairs. Stunting is caused by multidimensional factors and is not only due to poor nutrition experienced by pregnant women or young children. According to WHO, the characteristics of stunted children include delayed growth, late tooth eruption, a face that appears younger than their age, reduced focus, decreased memory ability, and limited eye contact.

3 Method

The research methodology section discusses the data sources and variables analyzed to achieve the study's objectives. This study uses secondary data from the 2022 Indonesian Health Profile, published by the Indonesian Ministry of Health in 2023, focusing on stunting rates across provinces in Indonesia. The response variable (Y) is the stunting rate, while the predictor variables (X) include the percentage of pneumonia (X_1), reflects the prevalence of respiratory infections that can hinder child growth. At the same time, complete basic immunization (X_2) indicates healthcare access, reducing disease risks linked to stunting. Low birth weight (X_3) represents poor prenatal conditions, often leading to growth impairment, and poverty (X_4) highlights socioeconomic challenges that limit access to nutrition, healthcare, and sanitation. These variables, sourced from health profiles and the Central Bureau of Statistics, are critical for understanding the determinants of stunting.

The data analysis techniques carried out in this study are as follows:

1. Determine the appropriate model (Poisson or negative binomial) based on overdispersion.
2. Estimate parameters using the Maximum Likelihood Estimation (MLE) method.
3. Test parameter significance with the Maximum Likelihood Ratio Test (MLRT).
4. Evaluate the model using deviance or chi-square.
5. Interpret the results to understand factors affecting stunting.
6. Conclude findings and policy implications.

4 Result and Discussion

4.1 Poisson Regression

The Poisson regression is used to model the average occurrence (λ) as an exponential function of a linear combination of predictor variables. This model is suitable for counting data and ensures that the predicted occurrences are always positive. Mathematically, the Poisson regression model is expressed as:

$$\ln(\lambda_i) = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

After modeling, the following results were obtained:

Table 4.1. Poisson regression parameter estimation

Parameter	Estimation Parameter	Standard Error	Z-value	P-value
β_0	9.636	0.005	2040.700	<0.001**
β_1	0.022	0.00002	906.200	<0.001**
β_2	0.025	0.00004	558.400	<0.001**
β_3	-0.259	0.0003	-936.700	<0.001**
β_4	0.025	0.0001	239.200	<0.001**

Notes: ** parameter is significant for 5% level

After modeling stunting data in toddlers using Poisson regression, the parameter estimates show that all predictor variables significantly impact stunting incidence (P-value < 0.001). The intercept ($\beta_0 = 9.636$) represents the log of the average stunting incidence when predictor variables are zero. The pneumonia percentage ($\beta_1 = 0.022$) increases the average stunting rate by $\exp(0.022) = 1.022$ for each 1 percent increase in pneumonia. Both the complete immunization rate ($\beta_2 = 0.025$) and poverty rate ($\beta_4 = 0.025$) increase stunting incidence by $\exp(0.025) = 1.025$. Conversely, the low birth weight rate (BBLR) ($\beta_3 = -0.259$) decreases stunting incidence by $\exp(0.259) = 1.295$. The model's AIC value is 5470694, with Nagelkerke's R^2 reaching 100%. The following is a visualization of the Poisson regression prediction results compared to the original value of the stunting number.

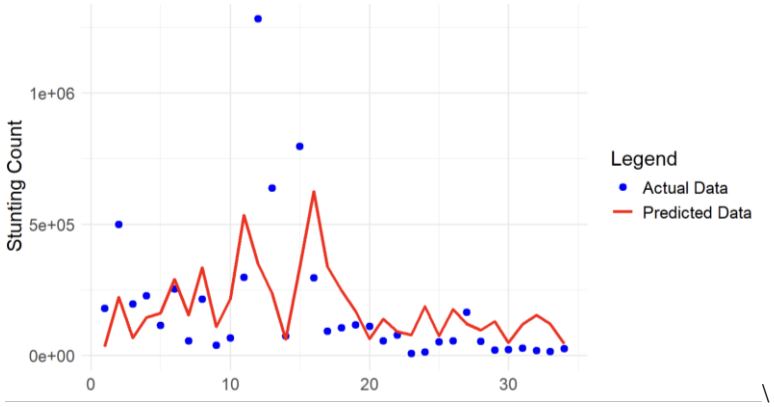

Fig 1. Poisson regression visualization.

4.2 Overdispersion

Overdispersion occurs when the variance of count data is significantly larger than its expected value, violating a core assumption of Poisson regression. This can lead to inaccurate estimates and misleading inferences. Therefore, detecting and addressing overdispersion is crucial for ensuring the model fits the data characteristics. Before switching to a negative binomial or alternative model, it is important to check the response variable's distribution to confirm the chosen model aligns with the data's distribution characteristics. Below is a further analysis of response variable distribution in the context of overdispersion.

Table 4.2. Checking the distribution of the response variable

Distribution	Probability
Lognormal	44%
Negatif Binomial	34%
Beta Binomial	9%

The distribution check for the response variable indicates that the most likely predicted distribution is lognormal (44%), followed by zero-inflated negative binomial (34%) and beta-binomial (9%). However, as lognormal is a continuous distribution, it is less suitable for count data, which is discrete. Thus, the negative binomial distribution, with a 34% probability, is more relevant for count data.

Table 4.3. Overdispersion test

Statistics	Value
Rasio Dispersion	379114.186
Pearson's Chi-Squared	12510768.126
P-value	<0.001**

The overdispersion test results in Table 2 indicate a serious issue with the Poisson regression model. A dispersion ratio of 379,114.186, much greater than 1, confirms significant overdispersion, as the data variance far exceeds its expectation. Pearson's Chi-square of 12,510,768.126 and a p-value below 0.001 confirm the model's poor fit. These findings suggest that the Poisson model is unsuitable, and an alternative, like negative binomial regression, is needed for more accurate estimation.

4.3 Negative Binomial Regression

Negative binomial regression extends the Poisson model by adding a dispersion parameter, allowing the variance to exceed the mean, which provides more accurate estimates when overdispersion is present. This section discusses the theory of negative binomial regression, parameter estimation methods, and its application and interpretation in the context of stunting data in children. This model aims to offer better insights into factors affecting stunting and more reliable results than the Poisson model. The modeling results are as follows:

Table 4.4. Parameter estimation of initial negative binomial regression model

Parameter	Estimation Parameter	Standard Error	Z-value	P-value
β_0	13.449	1.4713	9.141	<0.001**
β_1	0.0222	0.0098	2.260	0.0238**
β_2	-0.0069	0.0132	-0.516	0.6058
β_3	-0.2567	0.1056	-3.378	0.0007**
β_4	-0.0172	0.0356	-0.484	0.6281

Notes: ** parameter is significant for 5% level

The negative binomial regression modeling results show that not all predictor variables significantly impact stunting incidence in children. The percentage of pneumonia (X_1) has a parameter estimate of 0.0222 with a p-value of 0.0238, indicating a significant effect at the 5% significance level. Conversely, the percentage of complete basic immunization (X_2), with a parameter estimate of -0.0069 and p-value of 0.6058, and the percentage of poor population (X_4), with a parameter estimate of -0.0172 and p-value of 0.6281, do not significantly affect stunting (p-value > 0.05). This suggests that X_2 and X_4 do not meaningfully contribute to predicting stunting. The lack of significance for X_2 and X_4 could be due to limited variability or an indirect relationship with stunting factors. It is recommended to remove X_2 and X_4 from the model and re-estimate to achieve more accurate and efficient results. Below is the visualization of initial negative binomial regression model predictions compared to actual stunting counts.

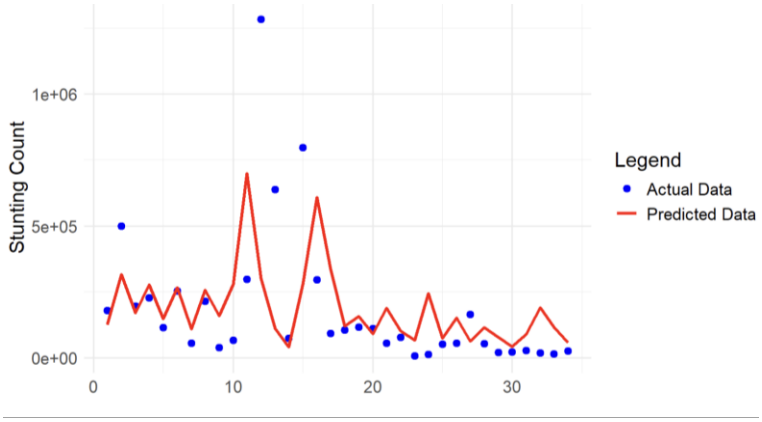

Fig 2. Binomial negative initial regression visualization.

Variables X_2 (complete immunization) and X_4 (poverty rate) may not significantly impact stunting due to low variability or indirect relationships with stunting. Immunization prevents disease but may not directly affect nutrition, while poverty influences stunting through complex factors like food access and sanitation. The negative binomial model, excluding X_2 and X_4, is presented below.

Table 4.5. Parameter estimation of the final negative binomial regression model

Parameter	Estimation Parameter	Standard Error	Z-value	P-value
β_0	12.364	0.4591	26.932	<0.001**
β_1	0.0276	0.0090	3.058	0.0022**
β_3	-0.3329	0.0979	-3.398	0.0006**

Notes: ** parameter is significant for 5% level

The final negative binomial regression model shows a highly significant intercept (β_0) estimate of 12.364 ($p < 0.001$). The percentage of pneumonia (X_1) has a parameter estimate of 0.0276 ($p = 0.0022$), indicating a significant effect on stunting at the 1% level, where each 1% increase in pneumonia percentage corresponds to a 1.0279 times increase in stunting cases. Additionally, the percentage of low birth weight (X_3) shows a parameter estimate of -0.3329 ($p = 0.0006$), signifying that a 1% increase in low birth weight correlates with a 0.7168 times decrease in stunting cases. Below is the visualization comparing final model predictions to actual stunting counts.

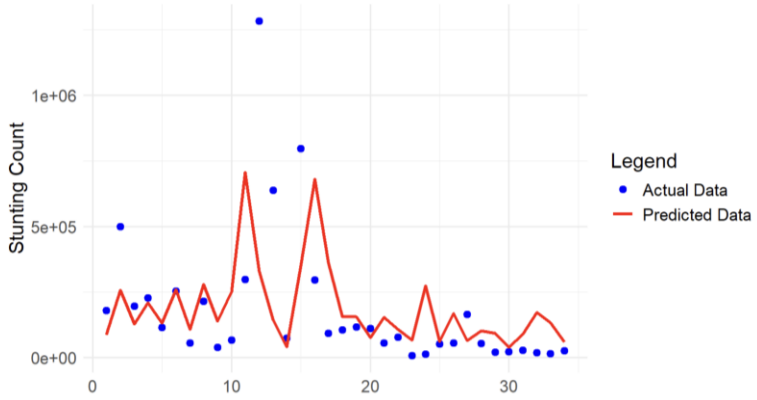

Fig 3. Binomial negative final regression visualization.

Fig 3 compares the actual stunting cases (blue dots) with the predicted cases from the final negative binomial model (red line). The blue dots represent actual stunting counts for each data point (ID), while the red line shows model predictions. The final model better aligns with the data trend, capturing higher peaks in stunting cases more accurately than previous models. This improved fit demonstrates that the final negative binomial model effectively addresses overdispersion, providing predictions that closely match actual stunting incidence and better handle data variability and significant outliers.

4.4 Best Model

This evaluation is based on several model performance indicators, including log-likelihood, AIC, BIC, and Nagelkerke's R^2. The primary goal is to identify a model that fits the data well and provides accurate and valid estimates. Selecting the best model offers more precise insights into the factors affecting stunting incidence and establishes a strong foundation for more effective child health policy decisions.

Tabel 4.6. Best model

Regression Model	Log-Likelihood	AIC	BIC	Nagelkerke's R^2
Poisson	-2735342	5470694	5470701	100%
Initial Negative Binomial	-439.3922	899.9426	890.7845	44.7%
Final Negative Binomial	**-438.3123**	**884.6246**	**890.7300**	**50.50%**

The final negative binomial regression model demonstrates improved performance with a log-likelihood of -438.3123, AIC of 884.6246, BIC of 890.7300, and an increased Nagelkerke's R^2 of 50.5%. This model provides a more accurate and reliable fit for modeling stunting incidence compared to previous models, effectively addressing overdispersion and offering better insight into the data. The final model is as follows:

$$\hat{y}_i = \exp(12.364 + 0.0276X_1 - 0.3329X_3)$$

Based on the parameter estimates of the final negative binomial regression model, the intercept (β_0) is estimated at 12.364 with a P-value < 0.001, indicating high significance. The pneumonia percentage (X_1) has a parameter estimate of 0.0276, meaning that each 1% increase in pneumonia is significantly associated with a 2.79% increase in stunting incidence ($\exp(0.0276) = 1.0279$). Additionally, the low birth weight percentage (BBLR) (X_3) has a parameter estimate of -0.3329, where each 1% increase in BBLR is significantly related to a 28.32% decrease in stunting incidence ($\exp(-0.3329) = 0.7168$).

5 Conclusion

The initial Poisson regression model indicated all variables significantly impacted stunting (p-value < 0.001), but high AIC (5470694) and Nagelkerke's R^2 (100%) suggested overdispersion. Switching to negative binomial regression, the final model excluded non-significant variables (Complete Basic Immunization and Poverty Percentage) and showed improved fit (AIC = 884.6246, Nagelkerke's R^2 = 50.50%). Pneumonia and low birth weight percentages remained significant predictors of stunting.

References

[1] D. C. Montgomery, E. A. Peck, and G. . Vinning, *Linear Regression Analysis*. John Wiley & Sons, Inc, 2021. doi: 10.2307/1268395.

[2] K. A. Yulianingsih, K. G. Gde, and L. P. Suciptawati, "Penerapan Regresi Poisson Untuk Mengetahui Faktor-Faktor Yang Memengaruhi Jumlah Siswa SMA/SMK Yang Tidak Lulus UN Di Bali," *e-Jurnal Mat.*, vol. 1, no. Agustus, pp. 59–63, 2012.

[3] M. Kroll, "Non-parametric Poisson regression from independent and weakly dependent observations by model selection," *J. Stat. Plan. Inference*, vol. 199, no. xxxx, pp. 249–270, 2019, doi: 10.1016/j.jspi.2018.07.003.

[4] G. Bouche, B. Lepage, V. Migeot, and P. Ingrand, "Intérêt de la détection et de la prise en compte d'une surdispersion dans un modèle de Poisson : illustration à partir d'un exemple," *Rev. Epidemiol. Sante Publique*, vol. 57, no. 4, pp. 285–296, 2009, doi: 10.1016/j.respe.2009.02.209.

[5] F. Suryadi, S. Jonathan, K. Jonatan, and M. Ohyver, "Handling overdispersion in poisson regression using negative binomial regression for poverty case in west java," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 517–523, 2022, doi: 10.1016/j.procs.2022.12.164.

[6] E. Allain and T. Brenac, "odGles de Poisson , binomial kgatif et quasi binomial negatif (1 Bj . Xj," 2001.

[7] T. W. Utami, "Analisis Regresi Binomial Negatif Untuk Mengatasi Overdispersion Regresi Poisson Pada Kasus Demam Berdarah Dengue," *J. Stat.*, vol. 1, no. 2, pp. 59–65, 2013.

[8] W. Fitriyanti and U. Kurniawan, "Regresi Negatif Binomial Bivariat untuk Mengatasi Overdispersi Regresi Poisson Bivariat.," *Statistika*, vol. 7, no. 1, pp. 39–46., 2019.

[9] M. F. Lais, A. Atti, R. M. Pangaribuan, and R. D. Guntur, "Model Generalized Poisson Regression (GPR) Pada Kasus Stunting Di Provinsi Nusa Tenggara Timur," *J. Difer.*, vol. 5, no. 2, pp. 68–75, 2023, doi: 10.35508/jd.v5i2.11562.

[10] F. Zubedi, F. A. Oroh, and M. A. Aliu, "Pemodelan Stunting dan Gizi Kurang di Kabupaten Bone Bolango menggunakan Regresi Poisson Generalized," *JMPM J. Mat. dan Pendidik. Mat.*, vol. 6, no. 2, pp. 113–128, 2021, doi: 10.26594/jmpm.v6i2.2507.

[11] A. Agresti, *An Introduction to Categorical Data Analysis*, 2nd ed., vol. 28, no. 11. Florida: A JOHN

WILEY & SONS, INC, 2009. doi: 10.1002/sim.3564.

[12] J. M. Hilbe, *Negative Binomial Regression*, 2nd ed. Inggris: Cambridge University Press, 2011. [Online]. Available: www.cambridge.org

[13] J. Stoklosa, R. V. Blakey, and F. K. C. Hui, "An Overview of Modern Applications of Negative Binomial Modelling in Ecology and Biodiversity," *Diversity*, vol. 14, no. 5, pp. 1–25, 2022, doi: 10.3390/d14050320.

[14] W. Greene, "Functional forms for the negative binomial model for count data," *Econ. Lett.*, vol. 99, no. 3, pp. 585–590, 2008, doi: 10.1016/j.econlet.2007.10.015.

[15] D. C. Jhwueng and C. Y. Wu, "A Novel Phylogenetic Negative Binomial Regression Model for Count-Dependent Variables," *Biology (Basel).*, vol. 12, no. 8, pp. 1–29, 2023, doi: 10.3390/biology12081148.

[16] R. R. Hocking, "Methods and Applications of Linear Models," *Technometrics*, vol. 39, no. 3, p. 332, 1996, doi: 10.2307/1271138.