ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 5, Issue 1, April 2025, pp. 1-8

Modeling the Prevalence of Stunting in Indonesia Using Quantile Regression

Farida Nur Hayati [a,1,*], Diana Nurlaily [a,2], Primadian Hasanah [a,3]

[a] Department of Mathematics and Technology, Information, Institut Teknologi Kalimantan, Kampus ITK Karang Joang Balikpapan 76127, Indonesia
[1] farida.nur.h93@gmail.com*; [2] diana.nurlaily@lecturer.itk.ac.id; [3] primadina@lecturer.itk.ac.id
* Corresponding author

ARTICLE INFO

Keywords
Intervention
Prevalence
Quantile Regression
Risk Factors
Stunting

ABSTRACT

Stunting is a condition where a child's height is under the average height of their age. Stunting will have an impact on the quality of human resources. The 2022 Indonesian Nutrition Status Survey reported that the prevalence of stunting in Indonesia reached 21.6%. This number decreased compared to the previous year. However, it remains below the government's planned target of 14%. Therefore, appropriate methods are needed to model and identify the factors with the most significant impact on the data for each region studied. This research modeled the stunting problem using quantile regression. Quantile regression has several advantages, including the fact that it can be used on data with an inhomogeneous distribution and is not affected by outliers. The results showed that variables that had a significant effect on the prevalence of stunting using 0.95 quantile regression included babies receiving exclusive breast milk, percentage of family planning participants, percentage of households with access to adequate sanitation, low birth weight (LBW) babies, and percentage of toddlers who have Maternal and Child Health (MCH) books. It is hoped that this research can be utilized to carry out appropriate interventions to reduce the prevalence of stunting that occurs in Indonesia.

1. Introduction

Stunting is the failure of growth and development due to chronic malnutrition in children under five years (toddlers), especially in the first 1,000 days of life [1]. It is a condition where a child's height is under the average height of their age. Stunting will impact the quality of human resources. The problem of stunting not only affects physical condition but also impairs children's health and thinking abilities. It will also impact their ability to learn as well as cause intellectual disabilities and the emergence of other chronic diseases. Therefore, this problem must be resolved immediately. Based on the 2023 UNICEF report, one in five children under five were stunted [2]. From these results, Indonesia has achieved good results in overcoming stunting. However, continuous efforts remain necessitated to ensure that children in Indonesia get optimal nutrition [3]. Stunting can be caused by several factors: malnutrition, inadequate access to clean water and sanitation, lack of

p-ISSN 2798-253X
e-ISSN 2798-3153

nutritional knowledge, and poor maternal health. Given the long-term impacts of stunting, it is critical to implement nutritional interventions, improve hygiene, and promote education to prevent stunting. The 2022 Survei Status Gizi Indonesia (SSGI) reported that the prevalence of stunting in Indonesia reached 21.6% [4]; this number decreased compared to the previous year. However, it remains fall short of the government's established target of 14% [5]. Based on this report, provinces with high prevalence of stunting were East Nusa Tenggara, West Sulawesi, and Papua. This disparity shows that the handling of nutritional problems remains uneven. Therefore, an appropriate method is necessitated to model and identify factors with greatest impact on the data of each region studied.

An attractive alternative beyond mean modeling is quantile regression [6]. The quantile regression approach is an improved variant of the least squares method [7] and can be used when there is heteroscedasticity. The presence of heteroscedasticity will lower the quality of the resulting model. Several prior studies on stunting cases have shown that stunting is influenced by multiple significant factors, including administrative area, place of residence, head of household, wealth status, access to media, mother's age of birth, mother's education, height, body mass index, pregnancy check-ups, and place of birth [8]. Another study analyzed stunting prevalence using quantile regression and concluded that breastfeeding, better economic status, average or above average birth weight of children, and milk consumption are factors that can reduce stunting, overweight, and underweight in Pakistan [9]. Another study modeled stunting toddlers, focusing on quantile regression in chronically severely stunted toddlers in Ghana [10]. The results of the analysis indicated that maternal and household level factors identified as significant predictors of severely stunted toddlers include maternal age and education, maternal national health insurance status, household, wealth status, and number of toddlers in the household.

This study aimed to model the stunting problem using quantile regression. Quantile regression has several advantages, including being able to be used on data with a nonhomogeneous distribution and is not affected by outliers. Quantile regression demonstrates a greater robustness to outliers compared to ordinary least squares regression. This enhanced resilience arises from the fact that observations that are significantly distant from the mean may exert high leverage, potentially resulting in substantial bias in the estimates of the mean [11]. It is hoped that this study can be used to carry out appropriate interventions in reducing the prevalence of stunting in Indonesia. By gaining a deeper understanding of these factors, it is hoped that the government can implement suitable interventions to reduce the prevalence of stunting and enhance the health and quality of life for children in Indonesia.

2. Methods

2.1. Regression

Regression model is the application of linear model with a response variable and predictor variable. Multiple linear regression is utilized when analyzing the relationships between several predictor variables [12]. Regression analysis is one of the most used statistical methods. Its application can be found in many fields of science, including specific fields, such as medicine, biology, agriculture, economics, engineering, sociology, and geology. The purpose of regression analysis includes establishing a causal relationship between the response variable y and the regression; predicting y based on the values of $x_1, x_2, \dots, x_n$; and identifying variables that are more important than others to explain the response variable y [13]. The multiple linear regression model is written as a direct extension of the simple linear model, with the model as in (1) [14]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{1}$$

where y denotes dependent variable; x_j denotes independent variable; β_0 denotes intercept (value when all independent variables are 0); β_j denotes regression coefficient where j=1,2,..p; and ε denotes the error that is assumed to be normally distributed with a mean of 0 and a variance σ^2.

p-ISSN 2798-253X
e-ISSN 2798-3153

2.2. Quantile Regression

Quantile regression is a development of the regression method introduced by Koenker and Basset in 1978. This analysis measures the relationship between predictor variables and conditional quantiles of the response variable without assuming a particular distribution [15]. Quantile regression effectively utilizes the median estimator to minimize absolute errors, allowing for a more accurate estimation of the median function [16]. It can be used when the data used is not homogeneous (the variance of the dependent variable changes as the predictor variable changes). For example, there is data $\{y_1, y_2, \dots, y_n\}$ and τ is a cumulative function or quantile to τ then,

$$F_y(y) = F(y) = P(Y \le y) = \tau \tag{2}$$

In linear regression, $E(y|x_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$, while quantile regression $Q_y(y|x_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau$ which can be described with the following model:

$$y_i = \beta_{\tau,0} + \sum_{k=1}^{p} \beta_{\tau k} x_{ik} + \varepsilon_{\tau i}; \qquad i = 1,2, \dots n \tag{3}$$

where $\beta_{\tau 0}, \beta_{\tau 1}, \dots \beta_{\tau p}$ is an estimator at the τth quartile. In matrix form, this model can be seen in (4).

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_\tau + \boldsymbol{\varepsilon}_\tau \tag{4}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_{\tau 1} \\ \beta_{\tau 2} \\ \vdots \\ \beta_{\tau n} \end{bmatrix} + \begin{bmatrix} \varepsilon_{\tau 1} \\ \varepsilon_{\tau 2} \\ \vdots \\ \varepsilon_{\tau n} \end{bmatrix} \tag{5}$$

where y is the response variable vector, X is the predictor variable matrix, and β_τ is the parameter vector at the τth quantile [17].

3. Methodology

3.1. Research Variables

The variables used in this research were obtained from the secondary data from the health service. The data consisted of 1 dependent variable and 6 independent variables, with a total of 34 observations in 34 provinces. Table 1 shows the research variables used.

Table 1. Research Variables

	Variable
y	Prevalence of stunting in toddlers (%)
X1	Babies receive exclusive breast milk (%)
X2	percentage of family planning participants (%)
X3	Percentage of households with access to adequate sanitation (%)
X4	Low birth weight (LBW) babies (%)
X5	Newborns receive early initiation of breastfeeding (%)
X6	Percentage of toddlers having Maternal and Child Health (MCH) books (%)

3.2. Steps

This research aimed to analyze the factors influencing the prevalence of stunting. The first step involved collecting relevant data and conducting a literature review to identify key determinants of stunting. Once the data was gathered, preprocessing was carried out, including summarizing the dataset and handling missing values to ensure data quality. Descriptive statistical analysis was then performed to understand the characteristics of the data, followed by a multicollinearity check using the variance inflation factor (VIF) to detect potential correlations among independent variables. Quantile regression was applied to model the prevalence of stunting, allowing for a more comprehensive understanding of how different factors impact various distribution quantiles. Finally, the results were interpreted, and conclusions were drawn, providing insights into effective interventions and policy recommendations to reduce stunting prevalence.

p-ISSN 2798-253X
e-ISSN 2798-3153

4. Results and Discussion

4.1. Characteristics of Stunting Prevalence in Indonesia

The prevalence of stunting in Indonesia persists above 20%, surpassing the World Health Organization's target. The five provinces with the highest rates of stunting are East Nusa Tenggara, West Sulawesi, Papua, West Nusa Tenggara, and Aceh, along with West Papua. In contrast, the provinces with the lowest prevalence of stunting are Bali, Jakarta, the Riau Islands, Lampung, and Yogyakarta. This can be seen from the bar chart in Fig. 1.

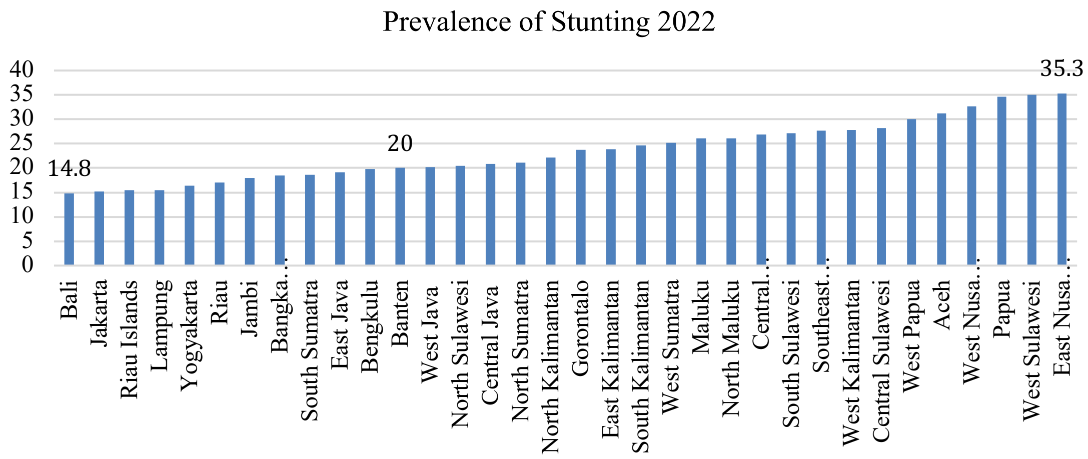

Fig. 1 Characteristics of stunting data.

As previously stated, quantile regression is applicable when the data used is not homogeneous and not symmetrical. Fig. 1 indicates asymmetrical data distribution; hence, the prevalence data need to be analyzed using quantile regression to determine the factors that significantly affect the response variables in each quantile. The histogram of the data is presented in Fig. 2

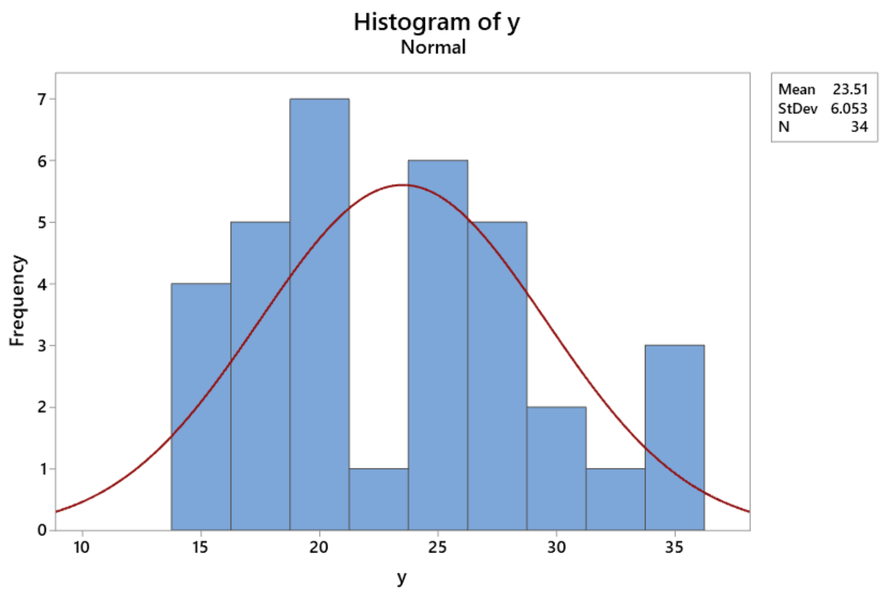

Fig. 2 Data histogram of stunting.

4.2. Linear Regression Analysis

Multicollinearity check is one of the assumptions of multiple linear regression. VIF values greater than 10 indicate that there is multicollinearity in the data used. Multicollinearity occurs when

p-ISSN 2798-253X
e-ISSN 2798-3153

there is a relationship between predictor variables. Table 2 presents the multicollinearity check results.

Table 2. Multicollinearity Check

Variable	VIF
X_1	1.891879
X_2	2.024969
X_3	1.749988
X_4	1.043289
X_5	1.133597
X_6	1.358513

The analysis results (Table 2) showed that the VIF value of all variables used was not more than 10, indicating that there is no high relationship between the predictor variables used. Furthermore, multiple linear regression analysis was carried out, with the model equation obtained as follows:

$$\hat{y} = 33.59 + 0.038\, X_1 - 0.108\, X_2 - 0.197\, X_3 + 1.383\ X_4 + 0.119\ X_5 - \ 0.088\, X_6$$

The simultaneous test of multiple linear regression analysis resulted in an F-value of 6.39 and a p-value of 0.000. With an error rate of 10%, it is concluded that there is at least one variable that affects the prevalence of stunting. A partial test (Table 3) was then carried out to find out which variables have a significant influence on the prevalence of stunting.

Table 3. Partial Test

	T-Value	P-Value
Constant	2.83	0.009
X1	0.64	0.526
X2	-1.24	0.225
X3	-1.94	0.063
X4	3.01	0.006
X5	1.17	0.252
X6	-2.06	0.049

Based on the results of the partial test, the variables with significant influence on the prevalence of stunting with α of 10% were variables percentage of households with access to adequate sanitation (X3), LBW babies (X4), and percentage of toddlers having MCH books (X6). From the analysis carried out, the residuals met the assumptions of independence and normal distribution but did not meet the assumptions of identicality.

4.1. Modeling Stunting Prevalence Rates Using Quantile Regression

In quantile regression analysis, parameter estimation is carried out to construct a model for each quantile, thereby resulting in an overview of how much influence the predictor variable has on the response variable at each quantile. The estimation of the quantile regression model is presented in Table 4.

Table 4. Model Parameter Estimation

Parameter	Quantile				
	0,05	*0,25*	*0,50*	*0,75*	*0,95*
$\hat{\beta}_0$	43.46292	40.06713	39.55595	48.52556	61.79815
$\hat{\beta}_1$	-0.07551	-0.04182	-0.01262	0.11429	0.17468
$\hat{\beta}_2$	0.01265	0.02900	-0.09950	-0.14603	-0.23669
$\hat{\beta}_3$	-0.25351	-0.27541	-0,21715	-0.35621	-0.45109
$\hat{\beta}_4$	1.22104	1.27573	1,22249	1.66190	2.66573
$\hat{\beta}_5$	-0.05841	-0.01951	0.08488	0.10768	0.02148
$\hat{\beta}_6$	-0.00085	-0.00412	-0.07330	-0.10404	-0.07413

The model of quantile regression analysis can be written in the following equation:

p-ISSN 2798-253X
e-ISSN 2798-3153

Quantile 0,05

$\hat{y}_{0.05} = 43.46 - 0.076\, X_1 + 0.013\, X_2 - 0.25\, X_3 + 1.22\ X_4 - 0.058\ X_5 -\ 0.00085\, X_6$

Quantile 0,25

$\hat{y}_{0.25} = 40.07 - 0.042\, X_1 + 0.029\, X_2 - 0.275\, X_3 + 1.275\ X_4 - 0.019\ X_5 -\ 0.00412\, X_6$

Quantile 0,5

$\hat{y}_{0.5} = 39.56 - 0.013\, X_1 - 0.099\, X_2 - 0.217\, X_3 + 1.22\ X_4 + 0.08488\, X_5 -\ 0.0733\, X_6$

Quantile 0,75

$\hat{y}_{0.75} = 48.53 + 0.114\, X_1 - 0.146\, X_2 - 0.35621\, X_3 + 1.662\ X_4 + 0.108\ X_5 -\ 0.104\, X_6$

Quantile 0,95

$\hat{y}_{0.95} = 61.79 + 0.175\, X_1 - 0.237\, X_2 - 0.451\, X_3 + 2.66\ X_4 + 0.02148\ X_5 -\ 0.07413\, X_6$

From the model $\hat{y}_{0,05}$, it can be interpreted that for data below quantile 0.05, an increase of X1 by one unit will result in the decrease of the stunting prevalence below quantile 0.05 by 0.076. Assuming other variables are constant, an increase of one unit in X2 will result in an increase in the stunting prevalence below quantile 0.05 by 0.013. Assuming other variables are constant.

Table 5. Significance of Model Parameters

Parameter	Quantile				
	0.05	***0.25***	***0.50***	***0.75***	***0.95***
$\hat{\beta}_0$	0.000*	0.002*	0.047	0.292	0.000
$\hat{\beta}_1$	0.000*	0.475	0.896	0.129	0.000*
$\hat{\beta}_2$	0.577	0.733	0.482	0.182	0.000*
$\hat{\beta}_3$	0.000*	0.009*	0.193	0.008*	0,000*
$\hat{\beta}_4$	0.000*	0.008*	0.109	0.007*	0.000*
$\hat{\beta}_5$	0.035*	0.845	0.608	0.397	0.693
$\hat{\beta}_6$	0.939	0.921	0.292	0.056*	0.003*

Table 5 shows that variables influencing the prevalence of stunting between quantiles are different. From the results, at quantile 0.05, the variables with a significant effect included X1, X3, X4, X5. While at quantile 0.95, the variables with a significant effect were X1, X2, X3, X4, X6. In the quantile model, the value $\boldsymbol{R'(\tau)}$ of the goodness of the model was obtained, which can be seen from the value of each quantile.

Table 6. Model Goodness

	Quantile				
	0.05	***0.25***	***0.50***	***0.75***	***0.95***
$\mathbf{R^1(\tau)}$	48.84%	41.55%	42.19%	45.15%	54.46%

Table 6 displays the calculation results of the parameters obtained. The results showed that the value of each quotient was different. The highest value was found in the quantile of 0.95, with a value of 54.46%. This indicates that the predictor variables used were able to explain the prevalence of stunting located below 0.95 by 54.46%. Meanwhile, the lowest value of goods was in the 0.25 quantile, with a value of 41.55%. This suggests that the predictor variable used was able to explain the variability of the response variable by 41.55%. The relationship between predictor variables and responses in each quotient is presented in Fig. 1.

p-ISSN 2798-253X
e-ISSN 2798-3153

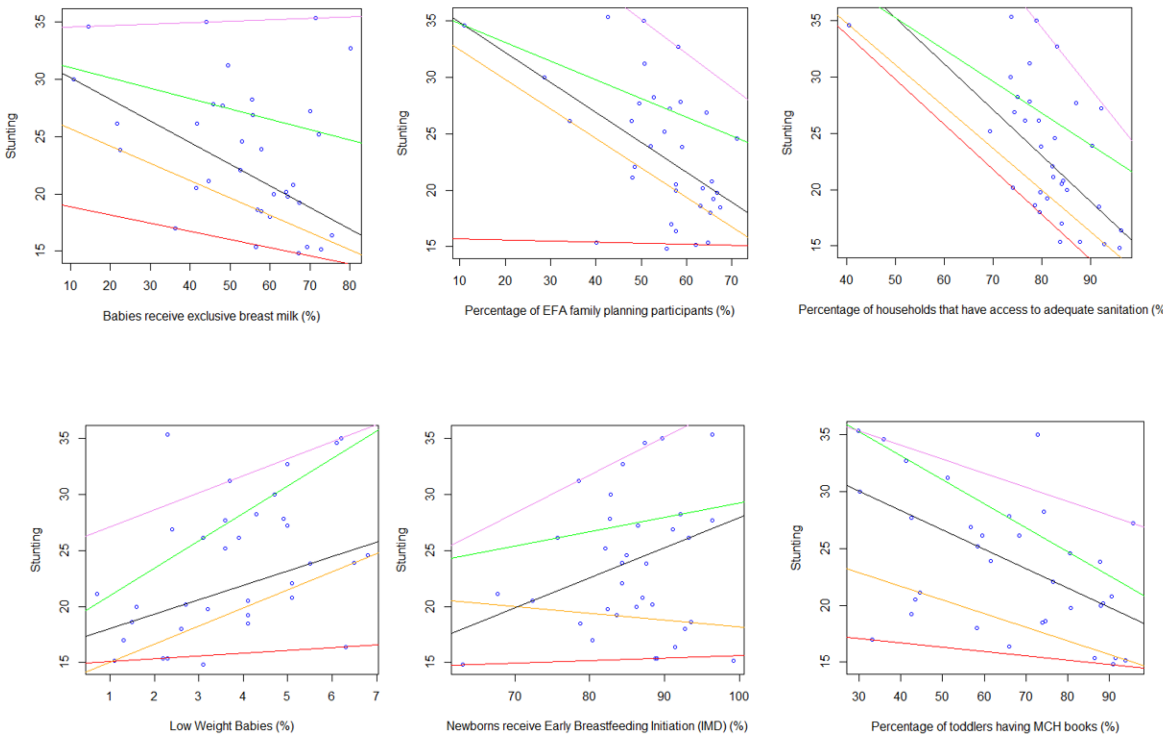

Fig. 1 Regression line based on variables.

The variable X1 exhibited a consistent pattern at quantiles 0.05, 0.25, 0.5, and 0.75, while at quantile 0.95, the regression line tended to slope. Variable X1 significantly influenced the prevalence of stunting at the quantiles 0.05 and 0.95. The results also exhibited an almost similar regression pattern for variable X2; however, it demonstrated a sloping pattern at quartile 0.05. At X2, the influence stunting prevalence was found in quartile 0.95. The variables X3, X4, and X6 showed similar patterns at each quantile. Meanwhile, for the variable X5, there was an influence on the prevalence of stunting in the 0.05 quartile.

5. Conclusion

The quantile regression analyzed used several quantiles, including 0.05, 0.25, 0.5, 0.75 and 0.95. The results of the analysis showed that there were significant differences in estimated values and variables in the regression models of several quantile regressions analyzed. Based on the goodness of the model, it is known that the highest value is in the 0.95 quantile with a value of 54.46%. Based on regression analysis at quantile 0.95, the variables that significantly influence the prevalence of stunting include Babies receive exclusive breast milk, percentage of family planning participants, Percentage of households with access to adequate sanitation, LBW babies, percentage of toddlers having MCH books. This study has several limitations, including a relatively low goodness-of-fit value and the inability to capture temporal dynamics. To address these limitations, future research is encouraged to utilize panel data, which would allow for the analysis of changes over time through longitudinal quantile regression techniques. Additionally, incorporating spatial analysis could provide valuable insights into geographic disparities in stunting, thereby enhancing the effectiveness of targeted policy interventions.

References

[1] F. Agushybana, A. Pratiwi, P.L. Kurnia, N. Nandini, J. Santoso, and A. Setyo, “Reducing Stunting Prevalence: Causes, Impacts, and Strategies,” in *BIO Web Conf.*, Nov. 2002, Art. no 00009, doi: 10.1051/bioconf/20225400009.

[2] UNICEF Indonesia, “Laporan Tahunan 2023,” 2003. [Online]. Available: https://www.unicef.org/indonesia/media/21331/file/UNICEF%20Laporan%20Tahunan%202023.pdf.pd f

[3] UNICEF Indonesia, “Laporan Tahunan Indonesia 2022,” 2022. [Online]. Available: https://www.unicef.org/indonesia/media/17686/file/Laporan%20Tahunan%20UNICEF%20Indonesia%202022%20(Single%20page).pdf

[4] Kementerian Kesehatan Republik Indonesia, “Profil Kesehatan Indonesia 2022,” 2023. [Online]. Available: https://kemkes.go.id/id/profil-kesehatan-indonesia-2022

[5] Badan Pusat Statistik, “Laporan Indeks Khusus Penanganan Stunting Kabupaten/Kota 2021-2022,” 2023. [Online]. Available: https://www.bps.go.id/id/publication/2023/12/19/37ee0674d0e89152b2425e9f/laporan-indeks-khusus-penanganan-stunting-2021-2022.html

[6] J. Mazucheli, B. Alves, A.F. Menezes, and V. Leiva, “An overview on parametric quantile regression models and their computational implementation with applications to biomedical problems including COVID-19 data,” *Comput. Methods Progr. Biomed.*, vol. 221, Jun. 2022, Art. no 106816, doi: 10.1016/j.cmpb.2022.106816.

[7] V.K. Patidar, R. Wadhvani, S. Shukla, M. Gupta, and M. Gyanchandani, “Quantile regression comprehensive in machine learning: A review,” in *2023 IEEE Int. Stud. Conf. Elect. Electron. Comput. Sci. (SCEECS)*, 2024, pp. 1–6, doi: 10.1109/SCEECS57921.2023.10063026.

[8] M.H. Chowdhury, M.F. Aktar, M.A. Islam, and N.M. Khan, “Factors associated with stunting status among under-5 years children in Bangladesh: Quantile regression modelling approach,” *Child. Youth Serv. Rev.*, vol. 155, Dec. 2023, Art. no 107199, doi: 10.1016/j.childyouth.2023.107199

[9] M. Siddiqa, A. Zubair, A. Kamal, M. Ijaz, and T. Abushal, “Prevalence and associated factors of stunting, wasting and underweight of children below five using quintile regression analysis (PDHS 2017–2018),” *Sci. Rep.*, vol. 12, no. 1, Nov. 2022, Art. no 20326, doi: 10.1038/s41598-022-24063-2.

[10] J.M.K. Aheto, “Simultaneous quantile regression and determinants of under-five severe chronic malnutrition in Ghana,” *BMC Public Health*, vol. 20, May 2020, Art. no 644, doi: 10.1186/s12889-020-08782-7.

[11] S.J. Staffa, D.M. Kohane, and D. Zurakowski, “Quantile regression and its applications: A primer for anesthesiologists,” *Anesth. Analg.*, vol. 128, no. 4, pp. 820–830, 2019, doi: 10.1213/ANE.0000000000004017.

[12] Y. Li, X. Li, M. Guo, C. Chen, P. Ni, and Z. Huang, “Regression analysis and its application to oil and gas exploration: A case study of hydrocarbon loss recovery and porosity prediction, China,” *Energy Geosci.*, vol. 5, no. 4, Oct. 2024, Art. no 100333, doi: 10.1016/j.engeos.2024.100333.

[13] X. Yan and X. G. Su, *Linear Regression Analysis*. Singapore, Singapore: World Scientific Publishing, 2009.

[14] R.J. Freund, W.J. Wilson, and P. Sa, *Regression Analysis: Statistical Modeling of a Response Variable*. Cambridge, MA, USA: Academic Press, 2006.

[15] E. Waldmann, “Quantile regression: A short story on how and why,” *Stat. Model.*, vol. 18, no. 3–4, pp. 213-218, 2018, doi: 10.1177/1471082X18759142.

[16] M. Maiti, “OLS versus quantile regression in extreme distributions,” *Contad. Adm.*, vol. 64, no. 2, pp. 1–11, Apr.–Jun. 2019, doi: 10.22201/fca.24488410e.2018.1702.

[17] R. Koenker, *Quantiles Regression*. Cambridge, England: Cambridge University Press, 2005.