International Journal of Multidisciplinary Research and Growth Evaluation.

Modeling the risk of Nutritional Status Wasting with Nonparametric Ordinal Logistic Approach Based on Spline

Muhammad Syahrie Khamdani [1], Umi Habibah [2], Nur Chamidah [3*], Lailatul Muniroh [4]
[1-4] Magister of Management Study Program, Faculty of Economics and Business, Universitas Islam Bandung, Indonesia

* Corresponding Author: **Nur Chamidah**

Article Info

ISSN (online): 2582-7138
Volume: 06
Issue: 01
January-February 2025
Received: 29-11-2024
Accepted: 30-12-2024
Page No: 1053-1065

Abstract
The nutritional status of toddlers is an important indicator in assessing public health quality, especially among children. Nutrition problems in toddlers are still a challenge for the Indonesian nation, including the high prevalence of stunting, underweight, and wasting. These various nutritional problems can be one of the factors that inhibit Indonesia's chances of becoming a developed country. Kediri Regency is one of the regions in East Java, Indonesia that has a high rate of malnutrition status. The nutritional status of toddlers in Kediri Regency which has a prevalence rate lower than the average prevalence of East Java is in stunting incidence. Meanwhile, the prevalence of other malnutrition events, such as wasting, underweight, and overweight is still above the average prevalence rate in East Java. This study was conducted aims to model and predict the nutritional status of toddlers in Kediri Regency, especially Ngasem District so that later appropriate policies can be made for the control of malnutrition problems in Kediri Regency using ordinal logistic regression approach. The developed model shows good performance in predicting nutritional status with classification accuracy of 74.41% on insample data and 73.81% on outsample data. In addition, the Macro AUC value of 71.56% shows that this model has a good ability to distinguish between different classes of nutritional status, with a classification error of 28.44%. The nonparametric ordinal logistic regression model developed in this study showed good results in the classification of the nutritional status of children under five.

Keywords: Nutritional Status, Malnutrition, Children, Ordinal Logistics Regression

Introduction

Toddlers' nutritional status is a crucial metric for evaluating the quality of public health, especially among children. Until now, the nutritional problem of toddlers is still a big concern in Indonesia, especially in rural areas and areas with limited access to health. From Indonesian Ministry of Health, nutrition problems in toddlers are still a challenge for the Indonesian nation, including the high prevalence of stunting, underweight, and wasting. These various nutritional problems can be one of the factors that inhibit Indonesia's chances of becoming a developed country.

UNICEF (2021) [1] in its press release, stated that "… children under the age of two are most vulnerable to all forms of malnutrition …". Rahmi H.G (2017) [2] developed malnutrition problems in toddlers that are not treated immediately can hinder growth and development, raise the chance of disease and death and affect toddlers' future lives in the long run. Various internal factors of toddler growth, such as age, weight, and height, can significantly affect the nutritional status of toddlers. The age of toddlers is when growth and development occur very rapidly. The first 1000 days of human life are a "golden period" known as the Window of Opportunity because children will grow optimally (Evania Yafie, 2018) [3]. In addition to age, height and weight factors are also closely related to determining nutritional status because they are related to the nutrients consumed by a person so they affect physical growth, metabolism, and overall body function.

Based on information from Zierle-Ghosh & Jan (2022) [4], Body Mass Index (BMI) is used to classify people into four general categories based on their height and weight: underweight, normal weight, overweight, and obesity. BMI calculations can help identify nutritional problems in toddlers early so that appropriate nutritional interventions can be carried out to support optimal growth. Thus, BMI measurement is an important tool in efforts to prevent and handle nutritional problems.

One of the East Javan regions with a high rate of malnutrition is Kediri Regency, using data obtained from Indonesian Ministry of Health, in his book entitled "Survei Kesehatan Indonesia (SKI) 2023" [5], the nutritional status of toddlers in Kediri Regency which has a prevalence rate lower than the average prevalence of East Java is in stunting incidence, which is 16.8, this value is smaller than the average of East Java province of 17.7. Meanwhile, the prevalence of other malnutrition events, such as wasting, underweight, and overweight is still above the average prevalence rate in East Java. The prevalence of wasting in Kediri Regency is 7.9%, higher than the provincial average of 6.8%. The prevalence of underweight was also high, at 14.2%, surpassing the East Java average of 13.3%. Likewise, overweight, whose prevalence reaches 5.5%, exceeds the provincial average of only 4.3%. This fact shows that although efforts to reduce stunting rates have been successful, health policies in Kediri Regency in dealing with other malnutrition problems have not been optimal.

There are previous studies that have discussed the nutritional status of toddlers. For example, a study conducted by Siregar *et al.* (2019) [6], this study was conducted in Padang City which showed that the factors that affect the incidence of stunting in toddlers are parental knowledge, exclusive breastfeeding, diarrhea, and parental income. This study emphasizes more on external factors and only discusses the incidence of malnutrition in the form of stunting. Another study conducted by Rahmadeni (2023) [7], found that most toddlers have good nutritional status, and that maternal education and weight have the biggest effects on toddlers' nutritional status. In this study, it discusses more about external and internal factors that affect the nutritional status of toddlers. Based on the facts about the nutritional status of toddlers in Kediri Regency and several previous studies on the nutritional status of toddlers in several locations in Indonesia, it is necessary to conduct further studies on the nutritional status of toddlers in Kediri Regency and what internal factors influence. This study was conducted aims to model and predict the nutritional status of toddlers in Kediri Regency, especially Ngasem District so that later appropriate policies can be made for the control of malnutrition problems in Kediri Regency. One statistical technique that describes the association between a response variable (Y) and multiple predictor variables (X) in which the response variable has more than two categories and the measurement scale is ordinal logistic regression is used in this study (Hosmer *et al*., 2013) [8]. In this study, nutritional status became a response variable with those categorized ordinally, namely normal, risk of malnutrition, and wasting. Ordinal logistic regression was chosen in this study because it allows us to model the relationship between various predictor variables and response variables that are categorized as ordinal.

Materials

1. Nutritional Status

Nutritional status is a depiction of food security needed by the body (Anggraeni *et al*., 2021) [9]. One way to determine nutritional status is by anthropometric measurements, which is to calculate the weight index according to height and weight index according to body length. From this anthropometric calculation, a person's nutrition can be classified into obese, overweight, normal, or underweight.

2. Malnutrition

Malnutrition is a condition if the body experiences a deficiency or excess of nutrients, although it is often used to describe a condition of malnutrition (Muhammad Iqbal & Sartono, 2018) [10]. According to Suryani *et al.* [11] his study's findings, inadequate nutrition during toddlerhood can result in stunted growth, make children lazy to perform energy-producing tasks, disrupt their immune systems, making them more vulnerable to infectious diseases, prevent their brains from developing to their full potential, and alter their behavior by making them agitated, prone to crying, and influencing their continued apathetic behavior.

3. Wasting

Wasting is a condition in which toddlers suffer from malnutrition with a diagnosis that is established based on a height per weight assessment. This indicates a deficit or lack of weight proportion when compared to the height of toddlers. (Anggraeni *et al*., 2021) [9].

4. Multinomial Distribution

The multinomial distribution is an extended version of the binomial distribution. Let $n_{\kappa i}$ denote the number of observations of $Y = k$ values appearing at $x_i = (x_{i1}, x_{i2}, \dots, x_{ij})$, for $k = 1,2,\dots,q-1$ then $[Y_i, i = 1,2,\dots,n]$ is an independent random variable with multinomial distribution. Suppose $Y_i = (y_{1i}, y_{2i}, \dots, y_{qi})$ then $Y_i \sim multinomial(n_i, \pi_i)$; $n_i = (n_{1i}, n_{2i}, \dots, n_{qi})$,$\pi_i = (\pi_{1i}, \pi_{2i}, \dots, \pi_{qi})$ and $\pi_{1i} + \pi_{2i} + \cdots + \pi_{qi} = 1$. The probability function for the multinomial distribution is as following equation (1):

$$p(n_1, n_2, \dots, n_{q-1}) = \left(\frac{n!}{n_1!\, n_2! \dots n_q!}\right) {\pi_1}^{n_3} {\pi_2}^{n_2} \dots {\pi_q}^{n_q} \quad (1)$$

(Agresti, 2007) [12].

5. Nonparametric Regression

To figure out the connection between the predictor and the response that does not have the assumption that the relationship has a certain shape, nonparametric regression will be used. When the data regression curve's form or the structure of the association between variables are uncertain, nonparametric regression will be used. Suppose an observation is obtained in the form of paired data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ which follow the following model:

$$y_i = g(x_i) + \varepsilon_{ij} \; i = 1,2,..,n \quad (2)$$

In the i-th observation, y_i is the response variable, x_i is a predictor variable, ε_i is a random error that is presumed to be independent with zero mean and variance. The regression function to be estimated is represented by σ^2 and g. It is assumed that the nonparametric regression function is only smooth or contained in a certain function space so that the nonparametric regression model has high flexibility (Eubank, 1999; Wang, 2011) [13, 14].

6. Nonparametric Ordinal Logistic Regression

The development of the ordinal logistic regression model with a non-parametric approach resulting nonparametric ordinal logistic regression model. According (Hastie & Tibshirani, 1987) [15], a model that can estimate nonparametric models by replacing the linear form into an additive form is the Generalized Additive Models (GAM) so that the additive nonparametric ordinal logistic regression model is

$$g(\gamma_k(x_i)) = \theta_k + \sum_{j=1}^{p} g_j(x_{ij}), k = 1,2,\dots,q-1; i = 1,2,\dots,n \tag{3}$$

with g_j is the j-th predictor variable's unpredictable regression function, x_{ij} is the j-th predictor variable's i-th observation. predictor variable and p indicates how many predictor variables there are. The g function is the link function for the cumulative logit model, which is $ln\left(\frac{\gamma_k(x_l)}{1-\gamma_k(x_l)}\right)$, so that $\gamma_k(x_i)$ can be obtained as follows:

$$ln\left(\frac{\gamma_k(x_i)}{1-\gamma_k(x_i)}\right) = \theta_k + \sum_{j=1}^{p} g_j(x_{ij})$$

$$\frac{\gamma_k(x_i)}{1-\gamma_k(x_i)} = \exp\left(\theta_k + \sum_{j=1}^{p} g_j(x_{ij})\right)$$

$$\gamma_k(x_i) = \exp\left(\theta_k + \sum_{j=1}^{p} g_j(x_{ij})\right) - \gamma_k(x_i)\exp\left(\theta_k + \sum_{j=1}^{p} g_j(x_{ij})\right)$$

$$\gamma_k(x_i)[1 + \exp(\theta_k + \sum_{j=1}^{p} g_j(x_{ij}))] = \exp(\theta_k + \sum_{j=1}^{p} g_j(x_{ij}))$$

$$\gamma_k(x_i) = \frac{\exp\left(\theta_k + \sum_{j=1}^{p} g_j(x_{ij})\right)}{1+\exp\left(\theta_k + \sum_{j=1}^{p} g_j(x_{ij})\right)}, k = 1,2,\dots,q-1; i = 1,2,\dots,n \tag{4}$$

(S. Meyers & Guarino, 2006) [16].

7. Least Squared Spline Estimator

The Least Square Spline Estimator The least square spline estimator is one of the estimation techniques in nonparametric regression with a smooth segmented or truncated polynomial model that uses a truncated basis function. Least square spline is a polynomial model that has higher flexibility than ordinary polynomial models, causing least square spline regression to adjust more effectively to the local characteristics of a data function, in other words, it can produce a regression function that fits the data. Least square spline in nonparametric regression has the ability to estimate data behavior that tends to be different at different intervals (Eubank, 1999) [13]. In general, the function f in the space of least square spline of degree p with knots. $\tau_1, \tau_2, \dots, \tau_k$ is an arbitrary function that can be expressed as the following equation:

$$g(x_i) = \sum_{j=0}^{p} \beta_j x^j + \sum_{j=1}^{k} \beta_{j+p}(x-\tau_j)_+^p \tag{5}$$

with

$$(x-\tau_{j-p})_+^p = \begin{cases} (x-\tau_{j-p})^p & ,x \geq \tau_{j-p} \\ 0 & ,x < \tau_{j-p} \end{cases} \tag{6}$$

with $i = 1,2,\dots,n, j = 1,2,\dots,k, \tau_j$ is knot node, β are the nonparametric model parameters (real constants), p is the degree of polynomial order, k is the number of knots, and $\lambda = (\tau_1, \tau_2, \dots, \tau_k)$. By taking n paired samples (x_i, y_i), based on equation (4) can be written as follows:

$$g(x_n) = \beta_0(x_n)^0 + \beta_1(x_n)^1 + \dots + \beta_p(x_n)^p + \beta_{(p+1)}(x_n-\tau_1)_+^p + \dots + \beta_{(p+k)}(x_n-\tau_k)_+^p \tag{7}$$

The least square spline function in equation (7) can be expressed in matrix form as follows.

$$\begin{pmatrix} g(x_1) \\ g(x_2) \\ g(x_3) \\ \cdot \\ \cdot \\ g(x_n) \end{pmatrix} = \begin{pmatrix} 1 & x_1 x_1^2 x_1^3 & \dots & x_1^p (x_1-\tau_1)_+^p & \dots & (x_1-\tau_k)_+^p \\ 1 & x_2 x_2^2 x_2^3 & \dots & x_2^p (x_2-\tau_1)_+^p & \dots & (x_2-\tau_k)_+^p \\ 1 & x_3 x_3^2 x_3^3 & \dots & x_3^p (x_3-\tau_1)_+^p & \dots & (x_3-\tau_k)_+^p \\ & & & \cdot & & \\ & & & \cdot & & \\ & & & \cdot & & \\ 1 & x_n x_n^2 x_n^3 & \dots & x_n^p (x_n-\tau_1)_+^p & \dots & (x_n-\tau_k)_+^p \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{(p+k)} \end{pmatrix} \quad (8)$$

for more than one predictor variable or multiple predictors, the following model is obtained

$$y_i = \sum_{j=1}^{s} g_j(x_{ij}) + \varepsilon_i \quad (9)$$

For the extension of the equation (5) with a predictor variable of more than one called least square splinemultipredictor can be expressed as follows:

$$g_s(x_{is}) = \beta_{0s} + \sum_{h=1}^{p_s} \beta_{sh} x_{is}^h + \sum_{m=1}^{k_s} \beta_{(p_s+m)s} (x_{is} - \tau_{ms})_+^{p_s} \quad (10)$$

Based on equation (9) and the model in equation (10), the least square splinemultipredictor function can be simplified as follows.

$$y_i = \sum_{j=1}^{s} \left[\beta_{0j} + \sum_{h=1}^{p_j} \beta_{jh}\, x_{ij}^h + \sum_{m=1}^{k_j} \beta_{(p_j+m)j}\, (x_{ij} - \tau_{mj})_+^{p_j} \right] + \varepsilon_i \quad (11)$$

8. Selection of Number and Location of Knot Nodes

The least square spline is a series of polynomials consisting of different segments, which are joined at knots, which are transitional nodes where the function changes its behavior. The choice of the number and location of knots is very important in this model. One way to select the optimal knot nodes is by Generalized Cross Validation (GCV), which is used to minimize the model error. The best knot node is obtained when the GCV value is the smallest, which also indicates an optimal least square spline model.

$$GCV(p, \tau_1, \tau_2, \dots, \tau_k) = \frac{MSE(p, \tau_1, \tau_2, \dots, \tau_k)}{\left(\frac{1}{n} t\tau\left(\mathbf{I} - \mathbf{A}(p, \tau_1, \tau_2, \dots, \tau_k)\right)\right)^2} \quad (12)$$

$$MSE(p, \tau_1, \tau_2, \dots, \tau_k) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{g}(x_i)\right)^2 \quad (13)$$

where p is the order of the polynomial, τ is the knot node and $\mathbf{A}(\boldsymbol{\tau})$ s obtained by the relationship of $\hat{y} = \mathbf{A}(\tau)\mathbf{y}$.
The Quantiles are location measures that indicate the position of part of the data relative to the entire dataset. For a continuous random variable, the concept of quantile can be defined as the value ξ_q for the random variable X such that

$$\Pr(X \le \xi_q) = F(\xi_q) = q \quad (14)$$

where q represents the probability that the value of X is less than or equal to ξ_q. The quantile can be denoted by the symbol ξ_q. Example: The quantile of order $\frac{1}{2}$ is the median of the distribution, thus $\Pr(X \le \xi_{0.5}) = \frac{1}{2}$. To determine the position of quantiles or knot nodes in data, one method involves using an index to sample quantiles in data as knot nodes, defined as

$$\tau_j = X_{\left(\left[\frac{j(n+1)}{k+1}\right]\right)} \text{ with } j = 1,2, \dots, k \quad (15)$$

where $X_{(j)}$ represents the order statistic of the j-th position. This method suggests intervals such as $j = \lfloor n/5 \rfloor$ to $j = \lfloor n/3 \rfloor$ (Wu & Zhang, 2006) [17].

9. Maximum Likelihood Estimator

The maximum likelihood estimator method is a method that maximizes the likelihood function. The likelihood function is the joint probability function of the random sample $Y_1, Y_2, \dots, Y_n$ which is identically independent (iid) and is a function of the parameters θ. The likelihood function in nonparametric regression has the following equation:

$$L(\beta,\theta)=\prod_{i=1}^{n} f\,(y_i|x_i) \tag{16}$$

after getting the likelihood function, then calculate the log-likelihood function as follows:

$$\ell=\ln\{L(\beta,\theta)\}=\sum_{i=1}^{n}\ln\left(f(y_i|x_i)\right) \tag{17}$$

The estimator β̂ is obtained by lowering the log-likelihood function in equation (16) against β and then equating to 0 as in the following equation

$$\frac{\partial\ell(\beta,\theta)}{\partial\beta}=0 \tag{18}$$

While the estimator $\hat{\theta}$ s obtained by lowering the log-likelihood function in equation (16) to θ and then equalizing 0 as in the following equation:

$$\frac{\partial\ell(\beta,\theta)}{\partial\theta}=0 \tag{19}$$

After getting the first derivative, the second derivative of β and θ negative definite matrix is a guarantee that the result is maximum (Hogg *et al*., 2019) [18].

10. Newton Raphson for Non-linear Equation

Newton Raphson's method is a numerical iteration method used to solve equations that cannot be solved directly because they are not linear. This method uses a one point approach as the initial value (Agresti, 2007) [12]. For non-linear equation iteration is performed using the following equation (20) until a sufficiently precise value is reached

$$\varphi^{(i+1)}=\varphi^{(i)}-\left(\boldsymbol{H}^{(i)}\right)^{-1}r^{(i)} \tag{20}$$

Iteration end until fulfiled max $\left|\varphi^{(i+1)}-\varphi^{(i)}\right|<\delta$, where δ is precise value and

$$\boldsymbol{\varphi}=(\boldsymbol{\theta}^T,\boldsymbol{\beta}^T)^T=\left(\theta_1,\theta_2,\dots,\theta_{q-1},\beta_1,\beta_2,\dots,\beta_p\right)^T \tag{21}$$

$$r=\frac{\partial\ell}{\partial\boldsymbol{\varphi}}=\begin{pmatrix}\frac{\partial\ell}{\partial\theta_3}\\ \frac{\partial\ell}{\partial\theta_2}\\ \vdots\\ \frac{\partial\ell}{\partial\theta_{q-1}}\\ \frac{\partial\ell}{\partial\boldsymbol{\beta}}\end{pmatrix} \tag{22}$$

$$\boldsymbol{H}=\frac{\partial^2\ell}{\partial\boldsymbol{\varphi}\,\partial\boldsymbol{\varphi}'}=\begin{pmatrix}\frac{\partial^2\ell}{\partial\theta_1^2} & \frac{\partial^2\ell}{\partial\theta_1\,\partial\theta_2} & \cdots & \frac{\partial^2\ell}{\partial\theta_1\,\partial\theta_{q-1}} & \frac{\partial^2\ell}{\partial\theta_1\,\partial\theta'}\\ \frac{\partial^2\ell}{\partial\theta_2\,\partial\theta_1} & \frac{\partial^2\ell}{\partial\theta_2^2} & & \frac{\partial^2\ell}{\partial\theta_2\,\partial\theta_{q-1}} & \frac{\partial^2\ell}{\partial\theta_2\,\partial\theta'}\\ \vdots & & \ddots & & \vdots\\ \frac{\partial^2\ell}{\partial\theta_{q-1}\,\partial\theta_1} & \frac{\partial^2\ell}{\partial\theta_{q-1}\,\partial\theta_2} & & \frac{\partial^2\ell}{\partial\theta_{q-1}^2} & \frac{\partial^2\ell}{\partial\theta_{q-1}\,\partial\theta'}\\ \frac{\partial^2\ell}{\partial\beta\,\partial\theta_1} & \frac{\partial^2\ell}{\partial\beta\,\partial\theta_2} & \cdots & \frac{\partial^2\ell}{\partial\beta\,\partial\theta_{q-1}} & \frac{\partial^2\ell}{\partial\beta\,\partial\beta'}\end{pmatrix} \tag{23}$$

(Burden & Faires, 2010) [19].

11. Model Fit Test

Deviance is a statistical test used to test the suitability of the model by comparing the actual model to the estimated model. The ordinal logit model fit test using the deviance test statistic is carried out with the following hypothesis

H_0: Logistic regression model obtained is fit

H_1: Logistic regression model obtained is not fit

To test this hypothesis, the deviance test statistic is used, which is defined as follows:

$$D = -2[\ell(\hat{\pi}; y) - \ell(\pi; y)] \quad (24)$$

$$D = -2\sum_{i=1}^{n}\left[n_{1i}\ln\left(\frac{\hat{\pi}_1(x_i)}{y_{i1}}\right) + n_{2i}\ln\left(\frac{\hat{\pi}_2(x_i)}{y_{i2}}\right) + \cdots + n_{qi}\ln\left(\frac{\hat{\pi}_q(x_i)}{y_{iq}}\right)\right] \quad (25)$$

with $\hat{\pi}_{ik} = \hat{\pi}_k(x_i)$ is the probability of the i-th observation in the k-th category. The critical regions are H_0 rejected if $D > X^2_{\alpha,(q-1)(J-p-1)}$, where J is the number of combinations of levels of different predictor variables, p is the number of predictor variables and q is the number of categories (Hosmer *et al.*, 2013) [8].

Methods

The type of research used in this study is quantitative research. According to Sugiyono (2019) [20], quantitative research is a method rooted in the philosophy of positivism, used to investigate a population or a specific sample with data collected through certain instruments. The data in this study is primary data obtained from posyandu data under the auspices of the Ngasem Health Center, Kediri Regency from January 2023 to August 2024. In this study, the predictor variables used are:

X1: Body weight at birth

X2: Body height at birth

X3: Age

Y : Nutritional Status (1 = Normal, 2 = Risk od Malnutrition, 3 = Wasting)

The parameters for Nutritional are estimated using the Nonparametric Ordinal Logistic Approach based on Spline.

Results and Discussion

1. Descriptive Statistics and Characteristics on Toddler Nutrition Status Data at Posyandu Ngasem District, Kediri Regency

This study uses data obtained at the Posyandu Ngasem Kediri District with a total of 340 consisting of toddlers aged 0-60 months after birth recorded in August. From the toddler data obtained, a grouping was made in the form of classification 1 as a group of toddlers in the green area (the value of Z-score body weight/body height in the median or above), classification 2 as a group of toddlers in the yellow area (the value of Z-score body weight/body height is between -1 SD to -2 SD), and classification 3 as a group of toddlers in the red area (exceeding the limits of -2 SD). Furthermore, grouping writing is used according to each classification. The characteristics of nutritional status can be described through descriptive statistics for each predictor. Presented with a table that includes the mean, variance, min, median, and max can be seen as follows:

Table 1: Descriptive Statistics Table for Toddler Nutrition Status Data at Posyandu Ngasem Sub-District, Kediri Regency

Variable	Classification	Mean	Variant	Min	Median	Max
Body Weight At Birth (X_1)	1	3,29	0,207	1,7	3,3	4,8
	2	2,667	0,085	1,35	2,7	3,03
	3	2,607	0,108	1,69	2,6	3
Body Height At Birth (X_2)	1	49,88	0,599	41	50	54
	2	47,49	2,872	40	48	49
	3	47,24	2,618	42	48	49
Age (X_3)	1	28,22	267,749	0	26	59
	2	25,21	255,652	1	22	59
	3	30,42	362,142	0	34	59

Based on the Table 1, the average birth weight of toddlers classified as 1 is 3.29 kg, with a variance of 0.207, a maximum value of 4.8 kg, a median value of 3.3 kg, and a minimum value of 1.7 kg. Furthermore, the average birth weight of toddlers classified as 2 is 2.667 kg, with a variance of 0.085, a maximum value of 3.03 kg, a median value of 2.7 kg, and a minimum value of 1.35 kg. The average birth weight of toddlers classified as 3 is 2.607 kg, with a variance of 0.108, a maximum value of 3 kg, a median value of 2.6 kg, and a minimum value of 1.69 kg.

The average birth height of toddlers classified as 1 is 49.88 cm, with a variance of 0.599, a maximum value of 54 cm, a median value of 50 cm, and a minimum value of 41 cm. In classification 2, the average birth height of children under five was 47.49 cm, with a variance of 2.872, a maximum value of 49 cm, a median value of 48 cm, and a minimum value of 40 cm. Classification 3 showed an average birth height of 47.24 cm, with a variance of 2.618, a maximum value of 49 cm, a median value of 48 cm, and a minimum value of 42 cm.

The average age of under-fives classified as 1 is 28.22 months, with a variance of 267.749, a maximum value of 59 months, a median value of 26 months, and a minimum value of 0 months. For classification 2, the average age of under-fives is 25.21 months, with a variance of 255.652, a maximum value of 59 months, a median value of 22 months, and a minimum value of 1 month. Classification 3 showed an average age of 30.42 months, with a variance of 362.142, a maximum value of 59 months, a median value of 34 months, and a minimum value of 0 months.

The characteristics of the nutritional status of toddlers according to BW based on age group in months can be described using a bar chart presented below in Figure 1 as follows:

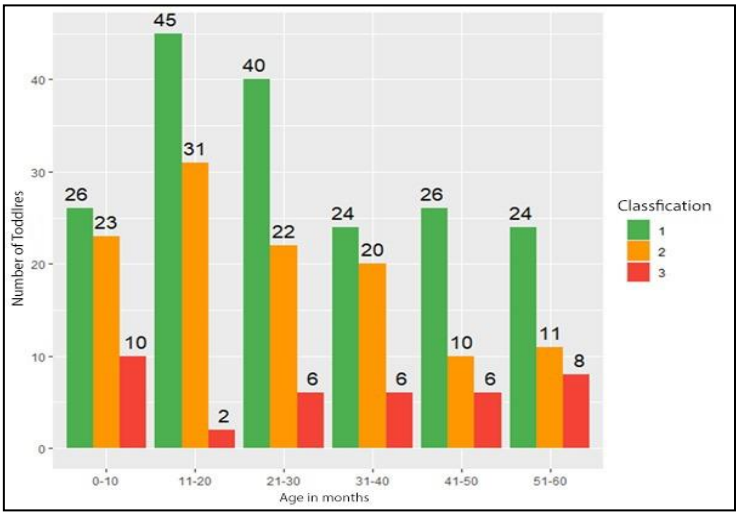

Fig 1: Nutritional Status of Toddlers According Age in Month

Based on the bar chart that has been presented in Figure 2 obtained information in the form of interpretation, namely, that in the group of toddlers aged 0-10 months there are 26 toddlers classified 1, there are 23 toddlers classified 2 and there are 10 toddlers classified 3. The group of toddlers aged 11 - 20 months there are 45 toddlers classified 1, there are 31 toddlers classified 2 and there are 2 toddlers classified 3. The group of toddlers aged 21-30 months there are 40 toddlers classified 1, there are 22 toddlers classified 2 and there are 6 toddlers classified 3. Group of toddlers aged 31 - 40 months there are 24 toddlers classified 1, there are 20 toddlers classified 2 and there are 6 toddlers classified 3. Group of toddlers aged 41 - 50 months there are 26 toddlers classified 1, there are 10 toddlers classified 2 and there are 6 toddlers classified 3. Group of toddlers aged 51-60 months there are 24 toddlers classified 1, there are 11 toddlers classified 2 and there are 8 toddlers classified 3.

From the bar chart, it can also be concluded that toddlers who are classified 1 are most in the age group 11-20 months with a total of 45 and the proportion is 57.69%. toddlers who are classified 1 are least in the age group 31-40 months and 51-60 months with a total of 24 for each and the proportion is 60% for the age group 31 - 40 months and 55.81% for the age group 51-60 months. Toddlers who were classified 2 were most in the age group 11-20 months with a total of 31 and the proportion was 39.74%. toddlers who were classified 2 were least in the age group 41-50 months with a total of 10 and the proportion was 23.8%. Toddlers who are classified 3 most in the age group 0-10 months with a total of 10 and the proportion is 16.94%. toddlers who are classified 3 least in the age group 11-20 months with a total of 2 and the proportion is 0.02%.

Furthermore, a bar chart is presented to determine the characteristics of the nutritional status of toddlers according to body weight/body height based on gender can be seen in the following Figure 2.

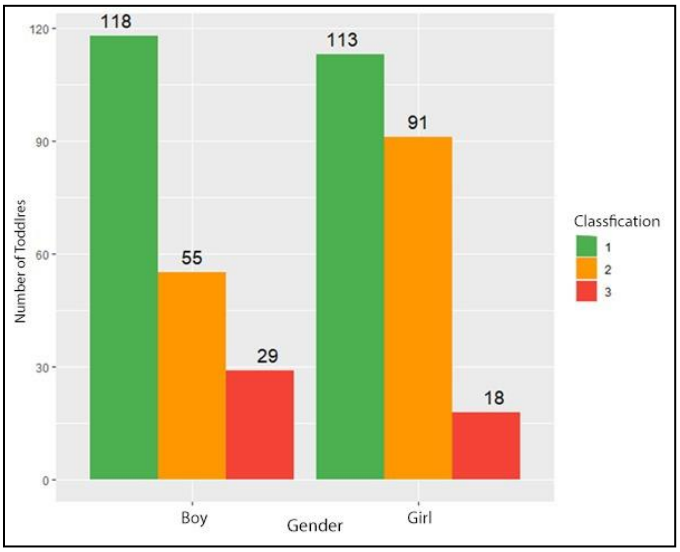

Fig 2: Nutritional Status of Toddlers According Gender

In the bar chart shown in Figure for male toddlers classified 1 amounted to 118 with a proportion of 58.41%, for male toddlers classified 2 amounted to 55 with a proportion of 27.22% and for male toddlers classified 3 amounted to 29 with a proportion of 14.35%. For female toddlers classified 1 amounted to 113 with a proportion of 50.9%, for female toddlers classified 2 amounted to 91 with a proportion of 40.99% and for female toddlers classified 3 amounted to 18 with a proportion of 8.1%. From the bar chart also for toddlers classified 1 and 3 most in the male sex group and toddlers classified 2 most in the female sex group.

2. Estimation of Ordinal Logistic Regression Model with Multipredictor Nonparametric Approach based on Least Square Spline Estimator

Suppose that $(x_i, y_i) = 1,2,\dots,n$ is paired data with $\boldsymbol{x_i} = (x_{1i}, x_{2i}, \dots, x_{si})^T$is the predictor variable of the i-th observation and y_i is the ordinal scale response variable with q categories. It is assumed that the i-th observation is multinomially distributed, and the multinomial distribution equation model given below.

$$f(y_i \mid \boldsymbol{x_i}) = \pi_1(\boldsymbol{x_i})^{n_{1i}}\pi_2(\boldsymbol{x_i})^{n_{2i}} \dots \pi_q(\boldsymbol{x_i})^{n_{qi}}$$

$$f(y_i \mid \boldsymbol{x_i}) = \left(\gamma_1(\boldsymbol{x_i})^{n_{1i}}[\gamma_2(\boldsymbol{x_i}) - \gamma_1(\boldsymbol{x_i})]^{n_{2i}} \dots \left[1 - \gamma_{q-1}(\boldsymbol{x_i})\right]^{n_{qi}}\right) \quad (26)$$

Using a log link function, the following formula determines the cumulative probability of the response variable y in ordinal logistic regression, which is γ_k dependent on the predictor variable $\boldsymbol{x_i}$.

$$\gamma_k = \frac{e^{\left(\theta_k + \sum_{j=1}^{j} g_j(x_{ij})\right)}}{1 + e^{\left(\theta_k + \sum_{j=1}^{s} g_j(x_{ij})\right)}}, k = 1,2,\dots,q-1 \text{ and } j = 1,2,\dots,s \quad (27)$$

A regression function with an unknown curve shape and an assumed smooth curve is $\sum_{j=1}^{s} \boldsymbol{g_j}(\boldsymbol{x_{ij}})$. Based on the previous equation, substituted with equation (1) for q = 3, the equation $f(y_i \mid \boldsymbol{x_i})$ is obtained as follows

$$f(y_i \mid \boldsymbol{x_i}) = \left(\frac{e^{\left(\theta_1 + \Sigma\, f_1 \boldsymbol{g_j}(x_{ij})\right)}}{1 + e^{\left(\theta_1 + \Sigma\ j=1} \boldsymbol{g_j}(\boldsymbol{x_{ij}})\right)}\right)^{n1i} \left(\frac{e^{\left(\theta_2 + \Sigma\, f_1 \boldsymbol{g}_j(x_{ij})\right)}}{1 + e^{\left(\theta_2 + \Sigma\ j=1} \boldsymbol{g}_j(\boldsymbol{x_{ij}})\right)} - \frac{e^{\left(\theta_1 + \Sigma\, f_1 \boldsymbol{g_j}(x_{ij})\right)}}{1 + e^{\left(\theta_1 + \Sigma\ j=1} \boldsymbol{g_j}(\boldsymbol{x_{ij}})\right)}\right)^{n2i}$$

$$\left(1 - \frac{e^{\left(\theta_2 + \Sigma\, f_1 \boldsymbol{g}_j(x_{ij})\right)}}{1 + e^{\left(\theta_2 + \Sigma\ j=1} \boldsymbol{g}_j(\boldsymbol{x_{ij}})\right)}\right)^{n3i} \quad (28)$$

The function $\sum_{j=1}^{s} \boldsymbol{g_j}(\boldsymbol{x_{ij}})$is estimated by a nonparametric regression approach using the multipredictor least square spline estimator, with $\boldsymbol{g}_j(\boldsymbol{x_{ij}})$ as follows:

$$\boldsymbol{g_j}(\boldsymbol{x_{ij}}) = \beta_{0j} + \sum_{h=1}^{p_j} \beta_{jh} x_{ij}^{\lambda} + \sum_{m=1}^{k_j} \beta_{(p_j+m)j} (\boldsymbol{x_{ij}} - \tau_{\text{mj}})_+^{p_j}; j = 1,2,\dots,s$$

with

$$(\boldsymbol{x_{ij}} - \tau_{mj})_+^{p_j} = \begin{cases} (\boldsymbol{x}_{ij} - \tau_{mj})^{p_j}, & x \geq \tau_{j-p} \\ 0 & , x < \tau_{j-p} \end{cases}$$

$$\boldsymbol{g_1}(\boldsymbol{x_{i1}}) = \beta_{01} + \sum_{k=1}^{p_1} \beta_{1n} x_{i1}^{h} + \sum_{m=1}^{k_1} \beta_{(p_1+m)1} (\boldsymbol{x_{i1}} - \tau_{m1})_+^{p_1}$$

$$\boldsymbol{g_2}(\boldsymbol{x_{i2}}) = \beta_{02} + \sum_{n=1}^{p_x} \beta_{2h} x_{i2}^{h} + \Sigma_{m=1}^{k_2} \beta_{(p_2+m)2} (\boldsymbol{x_{i2}} - \tau_{m2})_+^{p_2} \quad (29)$$

$$\boldsymbol{g_s}(\boldsymbol{x_{is}}) = \beta_{0s} + \sum_{h=1}^{p_s} \beta_{sh} x_{is}^{h} + \sum_{m=1}^{k_s} \beta_{(p_s+m)s} (\boldsymbol{x_{\text{ix}}} - \tau_{\text{ms}})_+^{p_s}$$

Based on the description of the least square spline function based on equation (29), the function $\sum_{j=1}^{s} \boldsymbol{g_j}(\boldsymbol{x_{ij}})$ is converted into vector notation form, so the multipredictor least square spline function can be written as follows.

$$\boldsymbol{g}(\boldsymbol{x_i}) = \boldsymbol{x_i \beta} \quad (30)$$

$$\boldsymbol{g}(\boldsymbol{x}_i) = [1 \quad \boldsymbol{x}_{1i}\boldsymbol{x}_{2i} \quad \cdots \quad \boldsymbol{x}_{st}] \begin{bmatrix} \sum_{j=1}^{p} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_z \\ \vdots \\ \boldsymbol{\beta}_s \end{bmatrix} \tag{31}$$

with

$$\begin{aligned} \boldsymbol{x}_{1t} &= \left[x_{1i}x_{1i}^1x_{1i}^2 \dots x_{1i}^{p_1}(x_{1i}-\tau_{11})_+^{p1} \dots \left(x_{1i}-\tau_{1i_1}\right)_+^{p_1}\right] \\ \boldsymbol{x}_{2i} &= \left[x_{2i}x_{2i}^1x_{2i}^2 \dots x_{2i}^{p2}(x_{2i}-\tau_{21})_+^{p2} \dots \left(x_{2i}-\tau_{2k_2}\right)_+^{p2}\right] \\ &\vdots \\ \boldsymbol{x}_{si} &= \left[x_{si}x_{xi}^1x_{si}^2 \dots x_{si}^{ps}(x_{si}-\tau_{s1})_+^{ps} \dots \left(x_{si}-\tau_{sk_d}\right)_+^{ps}\right] \text{ and} \\ \boldsymbol{\beta}_1 &= [\beta_{11}\beta_{12} \dots \beta_{1p_1}\beta_{1(p_1+1)} \cdots \beta_{1(p_1+k_1)}] \\ \boldsymbol{\beta}_2 &= [\beta_{21}\beta_{22} \quad \cdots \quad \beta_{2p_2}\beta_{2(p_2+1)} \quad \cdots \quad \beta_{2(p_2+k_2)}] \\ &\vdots \\ \boldsymbol{\beta}_x &= [\beta_{x1}\beta_{x2} \quad \cdots \quad \beta_{xp_s}\beta_{x(p_s+1)} \cdots \quad \beta_{x(p_y+k_5)}] \end{aligned} \tag{32}$$

So equation (5) can be substituted into equation (3) and can be written in another form as follows.

$$\begin{aligned} f(y_i \mid \boldsymbol{x}_i) &= \left(\frac{e^{(\theta_1+x_i\boldsymbol{\beta})}}{1+e^{(\theta_1+x_i\boldsymbol{\beta})}}\right)^{n_{1i}} \left(\frac{e^{(\theta_2+x_i\boldsymbol{\beta})}}{1+e^{(\theta_2+x_i\boldsymbol{\beta})}} - \frac{e^{(\theta_1+x_i\boldsymbol{\beta})}}{1+e^{(\theta_1+x_i\boldsymbol{\beta})}}\right)^{n2i} \left(1-\frac{e^{(\theta_2+x_i\boldsymbol{\beta})}}{1+e^{(\theta_2+x_i\boldsymbol{\beta})}}\right)^{n3i} \\ &= \left(\frac{e^{(\theta_1+x_i\boldsymbol{\beta})}}{1+e^{(\theta_1+x_i\boldsymbol{\beta})}}\right)^{n1i} \left(\frac{e^{(\theta_2+x_i\boldsymbol{\beta})}-e^{(\theta_1+x_i\boldsymbol{\beta})}}{(1+e^{(\theta_2+x_i\boldsymbol{\beta})})(1+e^{(\theta_1+x_i\boldsymbol{\beta})})}\right)^{n2i} \left(\frac{1}{1+e^{(\theta_2+x_i\boldsymbol{\beta})}}\right)^{n3i} \end{aligned} \tag{33}$$

The ordinal logistic regression model in equation (6) can be written in another form as follows:

$$\begin{aligned} f(y_i \mid \boldsymbol{x}_i) = &\ \exp\left(\ln\left(\frac{e^{(\theta_1+x_i\boldsymbol{\beta})}}{1+e^{(\theta_1+x_i\boldsymbol{\beta})}}\right)^{n1i} \left(\frac{e^{(\theta_2+x_i\boldsymbol{\beta})}-e^{(\theta_1+x_i\boldsymbol{\beta})}}{(1+e^{(\theta_2+x_i\boldsymbol{\beta})})(1+e^{(\theta_1+x_i\boldsymbol{\beta})})}\right)^{n2i} \left(\frac{1}{1+e^{(\theta_2+x_i\boldsymbol{\beta})}}\right)^{n3i}\right) \\ = &\ \exp\left\{\left[n_{1i}\ln\left(\frac{e^{(\theta_1+x_i\boldsymbol{\beta})}}{1+e^{(\theta_1+x_i\boldsymbol{\beta})}}\right) + n_{2i}\ln\left[\frac{e^{(\theta_2+x_i\boldsymbol{\beta})}-e^{(\theta_1+x_i\boldsymbol{\beta})}}{(1+e^{(\theta_2+x_i\boldsymbol{\beta})})(1+e^{(\theta_1+x_i\boldsymbol{\beta})})}\right] + \right.\right. \\ &\left.\left. n_{3i}\ln\left(\frac{1}{1+e^{(\theta_2+x_i\boldsymbol{\beta})}}\right)\right]\right\} \\ = &\ \exp\left\{\left[n_{1i}\left(\ln\left(e^{(\theta_1+x_i\boldsymbol{\beta})}\right) - \ln\left(1+e^{(\theta_1+x_i\boldsymbol{\beta})}\right)\right) + n_{2i}\left(\left(\ln\left(e^{(\theta_2+x_i\boldsymbol{\beta})} - \right.\right.\right.\right.\right. \\ & \left.\left.e^{(\theta_1+x_i\boldsymbol{\beta})}\right) - \left(\ln\left(1+e^{(\theta_2+x_i\boldsymbol{\beta})}\right) + \ln\left(1+e^{(\theta_1+x_i\boldsymbol{\beta})}\right)\right)\right) + n_{3i}(\ln(1) - \\ & \left.\left.\ln\left(1+e^{(\theta_2+x_i\boldsymbol{\beta})}\right)\right]\right\} \end{aligned} \tag{34}$$

The method to estimate the parameters $\boldsymbol{\beta}, \theta_1$ and θ_2 in logistic regression is the Maximum Likelihood Estimator (MLE) method. The form of the multinomial distribution likelihood function with the least square spline approach is as follows.

$$L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^{n} f(y_i \mid x_i) \tag{35}$$

$$\begin{aligned} L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^{n} \Big[&\ \exp\left\{\left[n_{1i}\left(\ln\left(e^{(\theta_1+x_d\boldsymbol{\beta})}\right) - \ln\left(1+e^{(\theta_1+x_i\boldsymbol{\beta})}\right)\right) + n_{2i}\left(\left(\ln\left(e^{(\theta_2+x_1\boldsymbol{\beta})} - \right.\right.\right.\right.\right. \\ & \left.\left.e^{(\theta_1+x_i\rho)}\right) - \left(\ln\left(1+e^{(\theta_2+x_i\boldsymbol{\beta})}\right) + \ln\left(1+e^{(\theta_1+x_i\boldsymbol{\beta})}\right)\right)\right) + n_{3i}(\ln(1) - \\ & \left.\left.\ln\left(1+e^{(\theta_2+x_i\boldsymbol{\beta})}\right)\right)\right]\Big\}\Big] \end{aligned} \tag{36}$$

Based on equation (9), the log-likelihood function is obtained as follows.

$$\begin{aligned}\ell(\boldsymbol{\beta},\theta) = & \ln\left(\prod_{i=1}^{n}\left[\exp\left\{\left[n_{1i}\left(\ln\left(e^{(\theta_1+x_d\boldsymbol{\beta})}\right)-\ln\left(1+e^{(\theta_1+x_i\boldsymbol{\beta})}\right)\right)+n_{2i}\left(\left(\ln\left(e^{(\theta_2+x_1\boldsymbol{\beta})}-\right.\right.\right.\right.\right.\right. \\ & \left. e^{(\theta_1+x_i\rho)}\right)-\left(\ln\left(1+e^{(\theta_2+x_i\boldsymbol{\beta})}\right)+\ln\left(1+e^{(\theta_1+x_i\boldsymbol{\beta})}\right)\right)\Big)\Big)+n_{3i}(\ln(1)- \\ & \ln\left(1+e^{(\theta_2+x_i\boldsymbol{\beta})}\right))]\}]) \\ = & \ln\left(\exp\sum_{i=1}^{n}\left[n_{1i}\left(\ln\left(e^{(\theta_1+x_i\beta)}\right)-\ln\left(1+e^{(\theta_1+x_i\beta)}\right)\right)+n_{2i}\left(\left(n\left(e^{(\theta_2+x_i{}^T\beta)}-\right.\right.\right.\right.\right. \\ & \ln\left(1+e^{(\theta_2+x_i\beta)}\right))]\}] \\ = & \sum_{i=1}^{n}\left[n_{1i}\left(\ln\left(e^{(\theta_1+x_i\boldsymbol{\beta})}\right)-\ln\left(1+e^{(\theta_1+x_i\boldsymbol{\beta}')}\right)\right)+n_{2i}\left(\left(\ln\left(e^{\left(\theta_2+x_i\boldsymbol{\beta}\ \right)}-\right.\right.\right.\right. \\ & \left. e^{(\theta_1+x_i\boldsymbol{\beta})}\right)-\left(\ln\left(1+e^{\left(\theta_2+x_i\boldsymbol{\beta}\ \right)}\right)+\ln\left(1+e^{(\theta_1+x_i\boldsymbol{\beta})}\right)\right)\Bigg)\Bigg)+n_{3t}(\ln(1)- \\ & \ln\left(1+e^{(\theta_2+x_i\boldsymbol{\beta})}\right))]\end{aligned} \tag{37}$$

When the first derivative of the parameters $\boldsymbol{\beta},\theta_1$ and θ_2 is reached, the values of $\boldsymbol{\beta},\theta_1$ and $\theta_2$that make up the log-likelihood function in equation (37) will be at their maximum. The log-likelihood function's first derivative with regard to parameter $\boldsymbol{\beta}$ is as follows: first derivative of the log-likelihood function with respect to parameter $\boldsymbol{\beta}$ is as follows:

$$\begin{aligned}\frac{\partial\ell(\boldsymbol{\beta},\theta)}{\partial\boldsymbol{\beta}} = & \sum_{i=1}^{n}\left\{n_{1i}\left(x_i-\frac{x_i\exp(\theta_1+\boldsymbol{x_i\beta})}{1+\exp(\theta_1+\boldsymbol{x_i\beta})}\right)+n_{2i}\left(x_i-\frac{x_i\exp(\theta_1+\boldsymbol{x_i\beta})}{1+\exp(\theta_1+\boldsymbol{x_i\beta})}-\right.\right. \\ & \left.\left.\frac{x_i\exp(\theta_2+\boldsymbol{x_i\beta})}{1+\exp(\theta_2+\boldsymbol{x_i\beta})}\right)+n_{3i}\left(-\frac{x_i\exp(\theta_2+\boldsymbol{x_i\beta})}{1+\exp(\theta_2+\boldsymbol{x_i\beta})}\right)\right\}\end{aligned} \tag{38}$$

The first derivative of the log-likelihood function for the parameters θ_1 and θ_2 is as follows:

$$\begin{aligned}\frac{\partial R(\boldsymbol{\beta},\theta)}{\partial\theta_1} = & \sum_{i=1}^{n}\left\{\left(\frac{n_{1i}}{1+\exp(\theta_1+\boldsymbol{x_i\beta})}\right)+n_{2i}\left(-\frac{\exp(\theta_1)}{\exp(\theta_2)-\exp(\theta_1)}-\frac{\exp(\theta_1+\boldsymbol{x_i\beta})}{1+\exp(\theta_1+\boldsymbol{x_i\beta})}\right)\right\} \\ = & \sum_{i=1}^{n}\left\{n_{2i}\left(-\frac{\exp(\theta_2)}{\exp(\theta_2)-\exp(\theta_1)}-\frac{\exp(\theta_2+\boldsymbol{x_i\beta})}{1+\exp(\theta_2+\boldsymbol{x_i\beta})}\right)+n_{3i}\left(-\frac{\exp(\theta_2+\boldsymbol{x_i\beta})}{1+\exp(\theta_2+\boldsymbol{x_i\beta})}\right)\right\}\end{aligned} \tag{39}$$

A numerical method was required to obtain parameter estimates because the implicit derivative in the first derivative of the parameters $\boldsymbol{\beta},\theta_1$ and θ_2 produced an unfeasible solution. As a result, the Newton-Raphson iteration method a numerical technique is employed. This approach necessitates a Hessian matrix (H). In the OSS-R program, the "optim" syntax is used for the second derivative of the log-likelihood function for parameters $\boldsymbol{\beta},\theta_1$ and θ_2.

3. Optimum Knot Count and Nodes Selection

Determining the number of knots and the ideal knot nodes is the first stage in estimating a logistic regression model using a nonparametric method based on the least square spline estimator. The purpose of this study is to categorize toddlers' nutritional status at Ngasem Health Center in Kediri Regency according to their weight/height Z-score. Three predictor variables that were deemed significant in determining the number of knots and optimal knot nodes according to the Generalized Cross Validation (GCV) criteria were used in the analysis. By choosing the minimum GCV value, the number of knots and knot nodes are chosen. The following table shows the minimum GCV value, the number of knots, and the ideal knot nodes.

Table 2: Optimum Knot Obtained

Variable	Number of Knots	Knot Nodes	Minimum GCV
Body Weight At Birth	3	2,435 2,93 3,395	0,2675078
Body Height At Birth	2	45,333 49,667	0,2677094
Age	1	29,5	0,4731475

4. Estimation Results of Nonparametric Ordinal Logistic Regression Model

After obtaining the optimum smoothing parameters of each predictor variable, it can then be done by iterating the initial value with the Newton Raphson method to obtain an estimate of the nonparametric logistic regression model based on the least squared spline estimator. From the output of the ordinal regression model estimation program with a nonparametric approach based on the least square spline estimator on the toddler data of the Ngasem health center in Kediri District, $\boldsymbol{\beta}$, θ_1 and θ_2 are obtained as follows.

$$\begin{aligned} \hat{\theta}_1 &= [-16{,}67341402] \\ \hat{\theta}_2 &= [-13{,}13673401] \\ \hat{\beta}_0 &= [-11{,}56843469] \\ \hat{\boldsymbol{\beta}}_1 &= [-1{,}54664975\ 7{,}34720729\ -0{,}20954320\ -7{,}96872797] \\ \hat{\boldsymbol{\beta}}_2 &= [\,0{,}63334465\ -0{,}81016696\ 3{,}22449459] \\ \hat{\boldsymbol{\beta}}_3 &= [0{,}01072471\ -0{,}02437810] \end{aligned} \quad (40)$$

For each predictor variable with n observations, the form of the least square spline estimator of the initial value function $\hat{\boldsymbol{g}}_j(\boldsymbol{x}_{ij})$ can be described as follows based on the values of the parameters $\boldsymbol{\beta}$, θ_1 and θ_2.

The 1st predictor model, namely birth body weight on the nutritional status of toddlers by holding other variables constant, can be written as the following equation.

$$\begin{aligned} \hat{g}_1(x_{i1}) = \; & 1{,}54664975x_{i1} + 7{,}34720729(x_{i1}-2{,}435)_+ - 0{,}20954320(x_{i1}-2{,}93)_+ \\ & -7{,}96872797(x_{i1}-3{,}395)_+ \end{aligned} \quad (40)$$

can be parsed into

$$\hat{g}_1(x_{i1}) = \begin{cases} -1{,}54664975x_{i1} & x_{i1} < 2{,}435 \\ -3{,}766092 + 7{,}34720729x_{i1} & 2{,}435 < x_{i1} < 2{,}93 \\ 17{,}76123 - 0{,}20954320x_{i1} & 2{,}93 < x_{i1} < 3{,}395 \\ 17{,}04983 - 7{,}96872797x_{i1} & x_{i1} > 3{,}395 \end{cases} \quad (41)$$

The 2nd predictor model, namely body height at birth on the nutritional status of toddlers by holding other variables constant, can be written as the following equation.

$$\hat{g}_2(x_{i2}) = 0{,}63334465x_{i2} - 0{,}81016696(x_{i2}-45{,}333)_+ + 3{,}22449459(x_{i2}-49{,}667)_+ \quad (42)$$

can be parsed into

$$\hat{g}_1(x_{i2}) = \begin{cases} 0{,}63334465x_{i12} & x_{i2} < 45{,}333 \\ 28{,}7116 - 0{,}81016696x_{i12} & 45{,}333 < x_{i2} < 49{,}667 \\ -11{,}52696 + 3{,}22449459x_{i12} & x_{i2} > 49{,}667 \end{cases} \quad (43)$$

The 3rd predictor model is age on the nutritional status of toddlers by holding other variables constant, so the following equation can be written.

$$\hat{g}_3(x_{i3}) = 0{,}01072471x_{i3} - 0{,}02437810(x_{i3}-29{,}5)_+ \quad (44)$$

can be parsed into

$$\hat{g}_3(x_{i3}) = \begin{cases} 0{,}63334465x_{i13} & x_{i3} < 29{,}5 \\ 18{,}68367 - 0{,}02437810x_{i13} & x_{i3} > 29{,}5 \end{cases} \quad (45)$$

From those equations that have been obtained, the following equation of odds can be obtained.

$$\begin{aligned} \hat{\pi}_1(\boldsymbol{x}_i) &= \frac{\exp(-16{,}67341402 - 1{,}54664975x_{i1} + 7{,}34720729(x_{i1}-2{,}435)_+ \dots + 0{,}01072471x_{i3} - 0{,}02437810(x_{i3}-29{,}5)_+)}{1+\exp(-16{,}67341402 - 1{,}54664975x_{i1} + 7{,}34720729(x_{i1}-2{,}435)_+ \dots + 0{,}01072471x_{i3} - 0{,}02437810(x_{i3}-29{,}5)_+)} \\ \hat{\pi}_2(\boldsymbol{x}_i) &= \frac{\exp(-13{,}13673401 - 1{,}54664975x_{i1} + 7{,}34720729(x_{i1}-2{,}435)_+ \dots + 0{,}01072471x_{i3} - 0{,}02437810(x_{i3}-29{,}5)_+)}{1+\exp(-13{,}13673401 - 1{,}54664975x_{i1} + 7{,}34720729(x_{i1}-2{,}435)_+ \dots + 0{,}01072471x_{i3} - 0{,}02437810(x_{i3}-29{,}5)_+)} \\ \hat{\pi}_2(\boldsymbol{x}_i) &= -\frac{\exp(-16{,}67341402 - 1{,}54664975x_{i1} + 7{,}34720729(x_{i1}-2{,}435)_+ \dots + 0{,}01072471x_{i3} - 0{,}02437810(x_{i3}-29{,}5)_+)}{1+\exp(-16{,}67341402 - 1{,}54664975x_{i1} + 7{,}34720729(x_{i1}-2{,}435)_+ \dots + 0{,}01072471x_{i3} - 0{,}02437810(x_{i3}-29{,}5)_+)} \\ \hat{\pi}_3(\boldsymbol{x}_i) &= \frac{1}{1+\exp(-13{,}13673401 - 1{,}54664975x_{i1} + 7{,}34720729(x_{i1}-2{,}435)_+ \dots + 0{,}01072471x_{i3} - 0{,}02437810(x_{i3}-29{,}5)_+)} \end{aligned} \quad (46)$$

After obtaining the nonparametric logistic regression model, the next step is to test the suitability of the model by calculating the deviance value. The results of the deviance statistical test are presented in the following Table 3.

Table 3: Criteria for Parametric and Nonparametric Model Fit

Criteria for Model Fit	Logistic Regression	
	Parametric	**Nonparametric**
Deviance	473,5353	455,5851
P-Value	1,00	1,00

The logistic regression model's fit criteria based on the deviance value were tested using the following hypothesis.
H_0: Logistic regression model obtained is fit.
H_1: Logistic regression model obtained is not fit.

Based on the Table 3. It can be seen with a significance level (α) of 0.05 that the deviance test statistical value of the nonparametric logistic regression model is 455.5851 and the p-value is 1.00. That way the p-value > α, then the decision can be made to accept H_0. So it is concluded that the nonparametric logistic regression model is suitable. In addition, the deviance test statistic value of logistic regression with a nonparametric approach has a smaller value than the deviance test statistic value of logistic regression with a parametric approach. So, it can also be concluded that logistic regression with a nonparametric approach provides better results than logistic regression with a parametric approach.

5. Classification Accuracy Value for Insample and Outsample Data

The classification accuracy value on insample data is obtained with the confusion matrix presented in the following table:

Table 4: Performance of In-Sample Data Using Confusion Matrix

		Prediction		
	Classification	**1**	**2**	**3**
Observartion	**1**	141	1	2
	2	41	109	33
	3	3	7	3

Based on Table 4, the classification accuracy value for in-sample data can be calculated as follows.

$$Accuracy = \frac{141 + 109 + 3}{340} = 0{,}7441 \quad (47)$$

The accuracy value for the insample data, as determined by the calculation in the equation above, is 74.41%. This indicates that the estimated ordinal logistic regression model using a nonparametric approach based on the obtained least square spline estimator is valid for classifying the nutritional status of toddlers in the sample data. Furthermore, the stability of classification accuracy or model accuracy is carried out by calculating the Press'Q value which is compared to the Chi-Square value with a free degree of 1. The following hypothesis is used for the Press'Q test.
H_0: Model classification results are inconsistent or unstable
H_1: Model classification results are consistent or stable

The Press'Q value must be greater than the Chi-Square value with a free degree of 1 in order to reject H_0 and conclude that the model classification results are stable or consistent. In this study, 340 observations were used as sample data; 253 of these observations were correctly classified, and three groups were used. The Press'Q value is calculated as follows:

$$Press'Q = \frac{\left(340 - \left(3 \times (141 + 109 + 3)\right)\right)^2}{340(3-1)} = 248{,}17 \quad (48)$$

The Press'Q value is 248.17, while the Chi-square table value of degree 1 with a significance level (α) of 0.05 is 3.841459. Based on the calculation results, it is known that the Press'Q value > Chi-square table value, then a decision can be made to reject H_0 and it can be concluded that the model on insample data is stable or consistent.
While, the classification accuracy value on the outsample data is obtained with the confusion matrix presented in the following table:

Table 5: Performance of Out-Sample Data Using Confusion Matrix

		Prediction		
	Classification	**1**	**2**	**3**
Observation	**1**	36	0	0
	2	10	26	9
	3	0	3	0

Based on Table 5, the classification accuracy value can be calculated for the outsample data as follows.

$$Accuracy = \frac{36 + 26 + 0}{84} = 0{,}7381 \quad (49)$$

Based on the calculation in the equation above, the accuracy value for outsample data is 73.81%, so it can be seen that the estimated ordinal logistic regression model with a nonparametric approach based on the least square spline estimator obtained is valid for calculating the classification of nutritional status of toddlers in outsample data. On the other hand the sensitivity, specificity and AUC value for each classification of outsample data is as follows:

Table 6: The Sensitivity, Specificity and AUC Value for Each Classification

Classification	Sensitivity	Specificity	AUC
1	0.7826	1.0000	0.8913
2	0.8966	0.6545	0.7755
3	0.000	0.9600	0.48

Based on Table 6 the model performed well in classification 1, moderately well in classification 2, but very poorly in classification 3. And furthermore, it is obtained macro AUC as follows

$$Macro\ AUC = \frac{1}{3}\sum_{i=1}^{3} AUC_i = 0.7156 \quad (50)$$

Based on the calculations obtained, the macro AUC value is 71.56% so that it gets a 'good' predicate. This indicates that the ordinal logistic regression model can avoid misclassification by 28.44%.

Conclusion

This study successfully built a nonparametric ordinal logistic regression model with a least squares spline estimator to classify the nutritional status of toddlers in the Ngasem Health Center area, Kediri. the nonparametric logistic regression model shows that this method is superior compared to the parametric approach in terms of model fit based on lower deviation values. The developed model shows good performance in predicting nutritional status with classification accuracy of 74.41% on insample data and 73.81% on outsample data. In addition, the Macro AUC value of 71.56% shows that this model has a good ability to distinguish between different classes of nutritional status, with a classification error of 28.44%. The nonparametric ordinal logistic regression model developed in this study showed good results in the classification of the nutritional status of children under five. For future research, it is recommended that this model be further developed by considering additional variables.

References

1. UNICEF. Young children’s diets show no improvement in last decade, ‘could get much worse’ under COVID-19 – UNICEF. [Press Release]. 2021 Sep 22. Available from: https://www.unicef.org/indonesia/press-releases/young-childrens-diets-show-no-improvement-last-decade-could-get-much-worse-under
2. Rahmi HG, I. Telaah faktor-faktor yang mempengaruhi status gizi balita di Kota Padang berdasarkan berat badan per tinggi badan menggunakan metode CART. Eksakta: Berkala Ilmiah Bidang MIPA. 2017;18(2):86–99. doi:10.24036/eksakta/vol18-iss02/59
3. Yafie E. Parents’ understanding in children development needs at the first 1000 days of birth. Int Conf Educ Technol (ICET). 2018;285:36–43.
4. Zierle-Ghosh A, Jan A. Physiology, body mass index - StatPearls - NCBI Bookshelf. In: NCBI Bookshelf. National Library of Medicine, National Institutes of Health; 2022.
5. Indonesian Ministry of Health. Survei Kesehatan Indonesia (SKI). 2023. Available from: https://www.badankebijakan.kemkes.go.id/hasil-ski-2023
6. Siregar AM, HG IR, Yozza H. Determinan kejadian stunting pada balita di Kota Padang dengan menggunakan analisis regresi logistik ordinal. Jurnal Matematika UNAND. 2019;8(1):9. doi:10.25077/jmu.8.1.9-16.2019
7. Rahmadeni R. Pemodelan status gizi balita menggunakan regresi logistik ordinal (Studi kasus: Puskesmas Limapuluh di Kota Pekanbaru). Jurnal Sains Matematika dan Statistika. 2023;9(2):21. doi:10.24014/jsms.v9i2.19080
8. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. Wiley; 2013. doi:10.1002/9781118548387
9. Anggraeni LD, Toby YR, Rasmada S. Analisis asupan zat gizi terhadap status gizi balita. Faletehan Health Journal. 2021;8(2):92–101. doi:10.33746/fhj.v8i02.191
10. Iqbal M, Sartono. Penilaian status gizi: ABCD. Jakarta: Salemba Medika; 2018.
11. Suryani L, Payung S, Pekanbaru N. Faktor yang mempengaruhi status gizi balita di wilayah kerja Puskesmas Payung Sekaki Pekanbaru. J Midwifery Sci. 2017;1(2):2549–2543.
12. Agresti A. An introduction to categorical data analysis. 2nd ed. Wiley; 2007.
13. Eubank RL. Nonparametric regression and spline smoothing. CRC Press; 1999. doi:10.1201/9781482273144
14. Wang Y. Smoothing splines: Methods and applications. CRC Press; 2011.
15. Hastie T, Tibshirani R. Non-parametric logistic and proportional odds regression. Appl Stat. 1987;36(3):260. doi:10.2307/2347785
16. Meyers SG, Guarino A. Applied multivariate research. SAGE Publications; 2006.
17. Wu H, Zhang JT. Nonparametric regression methods for longitudinal data analysis: Mixed-effects modeling approaches. Wiley; 2006. doi:10.1002/0470009675
18. Hogg RV, McKean JW, Craig AT. Introduction to mathematical statistics. 8th ed. Boston: Pearson; 2019.
19. Burden RL, Faires JD. Numerical analysis. 9th ed. Michelle Julet, Ed.; 2010.
20. Sugiyono. Metode penelitian kuantitatif, dan R&D. Bandung: Alphabet; 2019.