JOURNAL OF CULTURAL ANALYSIS AND SOCIAL CHANGE

Journal of Cultural Analysis and Social Change,
2025, 10(2), 868-877
ISSN: 2589-1316

An Empirical Study on Unmet Need: A Statistical Inference Framework for Truncated Spline Nonparametric Binary Logistic Regression (TSNLBR)

Afiqah Saffa Suriaslan[1], I Nyoman Budiantara[2*], Vita Ratnasari[3], Muhammad Zulfadhli[4], Risdiana Chandra Dhewy[5]

[1,2,3,4,5] *Departement of Statistics, Institut Teknologi Sepuluh Nopember, Kampus ITS-Sukolilo, Surabaya 60111, Indonesia*

***Corresponding Author:** i_nyoman_b@statistika.its.ac.id

Citation: Suriaslan, A. S., Budiantara, I. N., Ratnasari, V., Zulfadhli, M., and Dhewy, R. C. (2025). An Empirical Study on Unmet Need: A Statistical Inference Framework for Truncated Spline Nonparametric Binary Logistic Regression (TSNLBR), *Journal of Cultural Analysis and Social Change*, 10(2), 868-877. https://doi.org/10.64753/jcasc.v10i2.1704

Published: November 13, 2025

ABSTRACT

Nonparametric regression offers a flexible approach to uncover complex relationships without relying on rigid functional form assumptions. Among the available techniques, truncated spline regression serves as a powerful tool for approximating nonlinear effects. This study introduces the Truncated Spline Nonparametric Binary Logistic Regression (TSNBLR) model, specifically designed to accommodate binary response data, with a particular emphasis on developing a rigorous hypothesis testing framework. Model parameters are estimated using the Maximum Likelihood Estimation (MLE) method, while inference and evaluation are conducted through the Likelihood Ratio Test (LRT) and the Wald test. The proposed methodology is applied to unmet need data from East Java, Indonesia, where the response variable reflects the achievement of family planning targets. Empirical findings demonstrate that the truncated spline regression model not only provides a superior fit but also achieves higher classification accuracy compared to conventional Binary Logistic Regression (BLR). These results shows the effectiveness of TSNBLR in capturing nonlinear structures and enhancing the reliability of hypothesis testing in binary response modeling.

Keywords: Nonparametric Regression, Truncated Spline, Hypothesis Test, Categorical Data

INTRODUCTION

Regression analysis is one of the most widely used statistical methods in various studies [1]. Recent methodological advances have encouraged the adoption of more flexible modeling approaches, including nonparametric regression frameworks. Nonparametric regression is flexible in identifying relationships between response and predictor variables without assuming a specific functional form. Various smoothing techniques have been developed to implement nonparametric regression models, including Spline estimators [2], Fourier Series estimators [3], Wavelet estimators [4], Kernel estimators [5] , and Local Polynomial Regression [6]. Each method is suitable for specific data structures and research contexts. Among nonparametric techniques, spline estimators are widely applied, especially when response and predictor variables exhibit patterns that change at specific intervals [7]. A number of studies have shown that nonparametric spline estimators are capable of modeling nonlinear relationships, particularly when knot points are determined optimally [8], [9], [10]. Since the pattern approach was introduced by Whittaker in 1923 and continued by Schoenberg in 1942, spline estimators are now applied in semiparametric regression [11], [12], [13]. Further developments include biresponse[14], and its performance framework has been extended to mixed estimators [15], [16], [17].

Despite their advantages in nonlinear effects, most applications of truncated spline-based estimators focus on quantitative response variables. In practice, however, many empirical problems require models that are appropriate for binary outcomes. Therefore, if this method is applied to estimate models with binary response data, it can produce bias and result in models that do not adequately represent the problem. As it has developed, the ability of nonparametric estimators to estimate models with binary response data for estimation purposes has been examined using the fourier series approach [18]. Recently, truncated splines have been proposed as a development of the method for binary responses [19], [20]. The application of truncated splines has emphasized the flexibility of estimation. However, it has not yet addressed the inferential framework needed for hypothesis testing. This limitation hinders its usefulness in application contexts, where determining which predictors have a statistically significant effect on binary outcomes is crucial for evidence-based decision making. In other words, previous research has not extended this approach to hypothesis testing to identify the significance of influencing factors.

This study develops a nonparametric regression modeling theory, named Truncated Spline Nonparametric Binary Logistic Regression (TSNBLR), with a primary focus on expanding its application to hypothesis testing. The model obtained was then applied to data on the status of unmet needs in East Java, Indonesia. According to the Central Statistics Agency (BPS), East Java is one of the provinces with the largest population in Indonesia. One of the key drivers of population growth is the birth rate, making its reduction an important policy objective. A major factor contributing to elevated birth rates is the relatively high proportion of unmet need for family planning. According to a representative from the National Population and Family Planning Agency (BKKBN), data from the Family Planning Information System (SIGA-YAN) in 2023 show that unmet need in East Java remains at 12.97%, exceeding the target of 11.74%. Tackling unmet need is essential, as it highlights deficiencies in family planning services and has significant implications not only for maternal and child health but also for achieving broader demographic objectives. This figure highlights the challenge for the government to ensure equitable improvements in the well-being of the population across all regions of Indonesia. Therefore, a thorough analysis and the involvement of all stakeholders are necessary in addressing this issue.

MATERIAL AND METHODS

Nonparametric Binary Logistig Regression (TSNBLR)

Given $x_{1i},\ x_{2i},\ \dots,\ x_{pi}$; $i = 1, 2, \dots, n$, are as many as p predictor variables. Response variable (Y_i) is bernoulli distributed, with a probability distribution function [21]:

$$P(Y_i = y_i) = \pi(\boldsymbol{x_i})^{y_i}\left(1-\pi(\boldsymbol{x_i})\right)^{1-y_i};\ y_i = 0,1\ ;\ i = 1,2,\dots,n \tag{1}$$

Within logistic regression, the logit function provides a mechanism to convert the nonlinear association between predictor variables and probabilities into a linear form. The probability distribution is then transformed into the natural logarithm, formulated as:

$$\ln P(Y_i = y_i) = y_i \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) + \ln(1-\pi(\boldsymbol{x_i})) \tag{2}$$

When written in exponential form, the equation forms an exponential family distribution function. Where the logit function is defined as:

$$\ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = f(x_{1i},\ \dots, x_{pi}) \tag{3}$$

Then a logit transformation is performed to obtain a logistic regression model, is expressed as:

$$f(x_{1i},\ \dots, x_{pi}) = \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)$$

$$\ln\left(\exp f(x_{1i},\ \dots, x_{pi})\right) = \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)$$

$$\exp f(x_{1i},\ \dots, x_{pi}) = \frac{\pi(x_i)}{1-\pi(x_i)}$$

$$\exp f(x_{1i},\ \dots, x_{pi}) = \pi(\boldsymbol{x_i}) + \exp(z)\,\pi(\boldsymbol{x_i})$$

$$\pi(\boldsymbol{x_i}) = \frac{\exp f(x_{1i},\ \dots, x_{pi})}{1+\exp f(x_{1i},\ \dots, x_{pi})}$$

The function $f(x_{1i}, x_{2i,\ \dots,}\ x_{3i})$, a truncated spline function with knot points $K_{1j}, K_{2j,\ \dots,}\ K_{3j}$, where j is 1, 2, …, p, is employed. This yields the logit equation defined as [19]:

$$\pi(\boldsymbol{x_i}) = \frac{\exp f(\gamma_0 + \sum_{j=1}^{p}\gamma_{j1}x_{ji} + \sum_{j=1}^{p}\sum_{u=1}^{r}\gamma_{j(1+u)}(x_{ji} - K_{ju})_+)}{1+\exp f(\gamma_0 + \sum_{j=1}^{p}\gamma_{j1}x_{ji} + \sum_{j=1}^{p}\sum_{u=1}^{r}\gamma_{j(1+u)}(x_{ji} - K_{ju})_+)} \tag{4}$$

where, γ_0, γ_{jk}, and $\gamma_{j(m+u)}$, $j = 1,2,\dots,p$, $k = 1,2,\dots,m$, $u = 1,2,\dots,r$ are the model parameters in the truncated spline function. Truncated function for spline is represented by the following equation:

$$\left(x_{ji}-K_{ju}\right)_+ = \begin{cases}\left(x_{ji}-K_{ju}\right) & ,x_{ji} \geq K_{ju}\\ 0 & ,x_{ji} < K_{ju}\end{cases} \quad (5)$$

The model's logit function is represented in the following matrix form:

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{11}{}^{m} & \cdots & (x_{11}-K_{1r})_+^m & \cdots & x_{p1} & \cdots & \left(x_{p1}-K_{pr}\right)_+^m \\ 1 & x_{12} & \cdots & x_{12}{}^{m} & \cdots & (x_{12}-K_{1r})_+^m & \cdots & x_{p2} & \cdots & \left(x_{p2}-K_{pr}\right)_+^m \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{1n}{}^{m} & \cdots & (x_{1n}-K_{1r})_+^m & \cdots & x_{pn} & \cdots & \left(x_{pn}-K_{pr}\right)_+^m \end{bmatrix}\begin{bmatrix}\gamma_0\\ \gamma_{11}\\ \vdots\\ \gamma_{1m}\\ \vdots\\ \gamma_{1(m+r)}\\ \vdots\\ \gamma_{p1}\\ \vdots\\ \gamma_{p(m+r)}\end{bmatrix}$$

Maximum Likelihood Estimation (MLE)

MLE represents a statistical procedure that finds parameter estimates by optimizing the likelihood function [22]. Parameter estimation obtained is $\boldsymbol{\gamma}$, where

$\boldsymbol{\gamma} = (\gamma_0 \quad \gamma_{11} \quad \gamma_{12} \quad \dots \quad \gamma_{1(\mathrm{m+r})} \ \vdots \ \cdots \ \vdots \ \gamma_{\mathrm{p}1} \quad \cdots \quad \gamma_{\mathrm{p(m+r)}})'$

The form using MLE method is given by:

$$l(\boldsymbol{\gamma}) = \prod_{i=1}^{n} P(Y_i = y_i) = \pi(\boldsymbol{x_i})^{\sum_{i=1}^{n} y_i}(1-\pi(\boldsymbol{x_i}))^{n-\sum_{i=1}^{n} y_i} \quad (6)$$

The function is easier to maximize in the form of $\ln l(\boldsymbol{\gamma})$.

$$\ln\,[l(\boldsymbol{\gamma})] = L(\boldsymbol{\gamma}) = \ln\ \pi(\boldsymbol{x_i})^{\sum_{i=1}^{n} y_i}(1-\pi(\boldsymbol{x_i}))^{n-\sum_{i=1}^{n} y_i}$$
$$= \sum_{i=1}^{n}\left\{y_i\left(f(x_{1i}, \ \dots\ , x_{pi})\right) - \ln\left[1+\exp(f(x_{1i},\ \dots\ , x_{pi}))\right]\right\} \quad (7)$$

The estimator $\widehat{\boldsymbol{\gamma}}$ can be determined by setting the derivative equation to zero. However, the resulting equation is implicit, requiring the use of numerical iteration methods to solve it using the fisher scoring method. Thus, the estimator $\widehat{\boldsymbol{\gamma}}$ is given by

$\widehat{\boldsymbol{\gamma}} = \left(\hat{\gamma}_0 \quad \hat{\gamma}_{11} \quad \hat{\gamma}_{12} \quad \dots \quad \hat{\gamma}_{1(\mathrm{m+r})} \ \vdots \ \cdots \ \vdots \ \hat{\gamma}_{\mathrm{p}1} \quad \cdots \quad \hat{\gamma}_{\mathrm{p(m+r)}}\right)'$

RESULTS

Simultaneous Hypothesis Testing For TSNBLR

The simultaneous test evaluates the significance of the full set of model parameters γ in the model. The hypothesis form, parameter set, and statistic test are given as follows.

Hypothesis:

$H_0: \gamma_0 = \gamma_{11} = \cdots = \gamma_{p1} = \cdots = \gamma_{p(m+r)} = 0$

H_1: at least one $\gamma_{jk} \neq 0$ or $\gamma_{j(m+u)} \neq 0 \ , j = 1,2,\dots,p,\ k = 1,2,\dots,(m+u),\ u = 1,2,\dots,r$

Parameter set:

Under H_0: $\omega = \{\gamma_{\omega 0}\}$

Under population: $\boldsymbol{\Omega} = \left\{\gamma_0,\ \gamma_{11},\cdots,\gamma_{1(m+r)},\gamma_{p1},\cdots,\gamma_{pm},\cdots,\gamma_{p(m+r)}\right\}$

The test statistic used is the Likelihood Ratio Test (LRT), is formulated as:

$$G^2 = -2\ln\Lambda = -2\ln\left(\frac{\max_{\omega} l(\omega)}{\max_{\omega} l(\boldsymbol{\Omega})}\right)$$
$$= 2[\ln l(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) - \ln l(\boldsymbol{\gamma}_\omega)] - 2[\ln l(\widehat{\boldsymbol{\gamma}}_\omega) - \ln l(\boldsymbol{\gamma}_\omega)] \quad (8)$$

The ln-likelihood function $\ln l(\boldsymbol{\gamma}_\omega)$ (8) is approximated by a second-order Taylor series expansion around the MLE under the full model $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}$

$$\ln l(\boldsymbol{\gamma}_\omega) \approx \ln l(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) + g(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})(\boldsymbol{\gamma}_\omega - \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) - \frac{1}{2}(\boldsymbol{\gamma}_\omega - \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})^T[\mathrm{I}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})](\boldsymbol{\gamma}_\omega - \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) \quad (9)$$

With $\mathbf{I}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})$ is the fisher information matrix evaluated at $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}$.

$$g(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) = \frac{\partial \ln l(\boldsymbol{\gamma}_{\boldsymbol{\Omega}})}{\partial \boldsymbol{\gamma}_{\boldsymbol{\Omega}}}\Big|_{\boldsymbol{\gamma}_{\boldsymbol{\Omega}}=\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}} = \mathbf{0} \qquad \mathbf{I}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) = -\frac{\partial^2 \ln l(\boldsymbol{\gamma}_{\boldsymbol{\Omega}})}{\partial \boldsymbol{\gamma}_{\boldsymbol{\Omega}} \partial \boldsymbol{\gamma}_{\boldsymbol{\Omega}}{}^T}\Big|_{\boldsymbol{\gamma}_{\boldsymbol{\Omega}}=\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}}$$

Because $g(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})$ is zero in (9), it can be expressed as follows:

$$2[\ln l(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) - \ln l(\boldsymbol{\gamma}_\omega)] \approx (\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})](\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega) \quad (10)$$

The ln-likelihood function $\ln l(\boldsymbol{\gamma}_\omega)$ expanding around $\widehat{\boldsymbol{\gamma}}_\omega$ gives:

$$\ln l(\boldsymbol{\gamma}_\omega) \approx \ln l(\hat{\boldsymbol{\gamma}}_\omega) + g(\hat{\boldsymbol{\gamma}}_\Omega)(\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_\omega) - \frac{1}{2}(\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)](\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_\omega) \quad (11)$$

As $g(\hat{\boldsymbol{\gamma}}_\Omega)$ is zero in (11), it may be defined as follows:

$$2[\ln l(\hat{\boldsymbol{\gamma}}_\omega) - \ln l(\boldsymbol{\gamma}_\omega)] \approx (\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)](\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega) \quad (12)$$

Based on (10) and (12), the LRT statistic becomes:

$$G^2 \approx (\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\boldsymbol{\gamma}_\Omega)](\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega) - (\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)](\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega) \quad (13)$$

Based on [23], the maximum likelihood estimator $\hat{\boldsymbol{\gamma}}_\Omega$ satisfies the asymptotic distribution:

$$(\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\Omega) \xrightarrow{d} N(\mathbf{0}, [\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)]^{-1})$$

where,

$$\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega) = \begin{bmatrix} [\mathrm{I}_{11}] & [\mathrm{I}_{12}] \\ [\mathrm{I}_{21}] & [\mathrm{I}_{22}] \end{bmatrix} \text{ and } [\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)]^{-1} = \begin{bmatrix} [\mathrm{I}^{11}] & [\mathrm{I}^{12}] \\ [\mathrm{I}^{21}] & [\mathrm{I}^{22}] \end{bmatrix}$$

The quadratic form $(\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)](\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega)$ in (13) can be decomposed into:

$$(\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)](\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega) = \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix}^T \begin{bmatrix} [\mathrm{I}_{11}] & [\mathrm{I}_{12}] \\ [\mathrm{I}_{21}] & [\mathrm{I}_{22}] \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix} \quad (14)$$

then,

$$(\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)](\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega) = \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix}^T \begin{bmatrix} [\mathrm{I}_{12}][\mathrm{I}_{22}]^{-1}[\mathrm{I}_{21}] & [\mathrm{I}_{12}] \\ [\mathrm{I}_{21}] & [\mathrm{I}_{22}] \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix} \quad (15)$$

thus obtained:

$$G^2 \approx (\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\boldsymbol{\gamma}_\Omega)](\hat{\boldsymbol{\gamma}}_\Omega - \boldsymbol{\gamma}_\omega) - (\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_\Omega)](\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega)$$

$$G^2 \approx \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix}^T \begin{bmatrix} [\mathrm{I}_{11}] & [\mathrm{I}_{12}] \\ [\mathrm{I}_{21}] & [\mathrm{I}_{22}] \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix}^T \begin{bmatrix} [\mathrm{I}_{12}][\mathrm{I}_{22}]^{-1}[\mathrm{I}_{21}] & [\mathrm{I}_{12}] \\ [\mathrm{I}_{21}] & [\mathrm{I}_{22}] \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix}$$

$$G^2 \approx \hat{\boldsymbol{\gamma}}_1{}^T([\mathrm{I}_{11}] - [\mathrm{I}_{12}][\mathrm{I}_{22}]^{-1}[\mathrm{I}_{21}])\hat{\boldsymbol{\gamma}}_1 \quad (16)$$

Equation (16) is simplified to

$$G^2 \approx \hat{\boldsymbol{\gamma}}_1{}^T[\mathrm{I}^{11}]^{-1}\boldsymbol{\gamma}_1 \quad (17)$$

where,

$$\hat{\boldsymbol{\gamma}}_1 \xrightarrow{d} N\left(\mathbf{0}, [\mathrm{I}^{11}]_{(p(m+r)\times p(m+r))}\right) \text{ and } [\mathrm{I}^{11}]^{-\frac{1}{2}}\hat{\boldsymbol{\gamma}}_1 \xrightarrow{d} N\left(\mathbf{0}, \mathrm{I}_{(p(m+r)\times p(m+r))}\right)$$

Assuming the null hypothesis H_0 holds, the LRT statistic asymptotically follows a chi-square distribution with $p(m+r)$ degrees of freedom.

$$G^2 \approx \left[[\mathrm{I}^{11}]^{-\frac{1}{2}}\hat{\boldsymbol{\gamma}}_1\right]^T \left[[\mathrm{I}^{11}]^{-\frac{1}{2}}\boldsymbol{\gamma}_1\right] \approx Z^TZ \xrightarrow{d} \chi^2_{p(m+r)}$$

where $Z = [\mathrm{I}^{11}]^{-\frac{1}{2}}\hat{\boldsymbol{\gamma}}_1 \xrightarrow{d} N\left(\mathbf{0}, \mathrm{I}_{(p(m+r))}\right)$

Thus it can be concluded $G^2 \xrightarrow{d} \chi^2_{p(m+r)}$

For simultaneous testing of model parameters, the rejection region of the null hypothesis H_0 at significance level α is defined according to the distribution of the test statistic, expressed as:

$$G^2 > \chi^2_{(\alpha,p(m+r))} \quad (18)$$

Partial Hypothesis Testing For TSNBLR

Partial tests are used to evaluate the significance of each parameter separately. The formulation of the hypothesis, parameter sets, and test statistics are as follows.

Hypothesis:

For the main effect parameters γ_{jk}, where $j = 1,2,\ldots,p, j = 1,2,\ldots,(m+u), u = 1,2,\ldots,r$

$H_0: \gamma_{jk} = 0$

$H_1: \gamma_{jk} \neq 0$

The test statistic used to evaluate a partial hypothesis, known as the Wald test, is given

By

$$Z = \frac{\hat{\gamma}_{jk}}{\sqrt{Var(\hat{\gamma}_{jk})}} \quad (19)$$

where $\hat{\gamma}_{jk}$ is the estimated coefficient of the j-th parameter, and $Var(\hat{\gamma}_{jk})$ is its estimated variance.

Proof:

The Wald test statistic is given by

$$\chi^2 \approx (\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_0)^T([\mathbf{I}_{11}] - [\mathbf{I}_{12}][\mathbf{I}_{22}]^{-1}[\mathbf{I}_{21}])(\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_0)$$
$$\approx (\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_0)^T[\mathbf{I}^{11}]^{-1}(\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_0)$$
$$\approx (\hat{\gamma}_{jk} - \boldsymbol{\gamma}_0)^T[\mathbf{I}^{11}]^{-1}(\hat{\gamma}_{jk} - \boldsymbol{\gamma}_0)$$
$$\chi^2 \approx \hat{\boldsymbol{\gamma}}_{jk}^2[\mathbf{I}^{11}]^{-1} \tag{20}$$

Considering the normal asymptotic properties:

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - E(\hat{\boldsymbol{\gamma}})) \xrightarrow{d} N(\mathbf{0}, [\mathbf{I}(\hat{\boldsymbol{\gamma}})]^{-\mathbf{1}})$$

where, according to [24]:

$$\mathbf{I}(\hat{\boldsymbol{\gamma}}) = -E(H(\hat{\boldsymbol{\gamma}})) \text{ dan } [\mathbf{I}(\hat{\boldsymbol{\gamma}})]^{-\mathbf{1}} = [-E(H(\hat{\boldsymbol{\gamma}}))]^{-\mathbf{1}}$$

The element $\mathbf{I}^{11} = Var(\hat{\gamma}_{jk})$ in (18) is the main diagonal element of the matrix $-[H(\hat{\boldsymbol{\gamma}})]^{-1}$

$$\chi^2 \approx \hat{\gamma}_{jk}^2[\mathbf{I}^{11}]^{-1} \approx \hat{\gamma}_{jk}^2[Var(\hat{\gamma}_{jk})]^{-1} \approx \frac{\hat{\gamma}_{jk}^2}{Var(\hat{\gamma}_{jk})} \xrightarrow{d} \chi^2_{(1)}$$

Form of (19) is equivalent to a standard normal test:

$$Z = \frac{\hat{\gamma}_{jk}}{\sqrt{Var(\hat{\gamma}_{jk})}} \xrightarrow{d} N(0,1)$$

In light of the statistical test and its associated distribution, the region used to reject the null hypothesis H_0 at the significance level α for the partial parameter test in the TSNBLR model is specified below.

$$|Z| > Z_{a/2} \tag{21}$$

$Z_{a/2}$ is the upper $a/2$ quantile of the standard normal distribution.

Data Application

In applying the TSNBLR, this research utilized data on unmet need target achievement in East Java, Indonesia. The dataset is sourced from offcical BKKBN publications, and the variable are described in Table 1.

Table 1. Variable Desciption

Variabel	Notation	Descipriton
Response	y	1 = unmet need target unachieved 0 = unmet need target achieved
Predictor	X_1	Proportion of household without primary education
	X_2	Proportion of couples of childbearing age (PUS) with two children
	X_3	Proportion of couples of childbearing age (PUS) utilizing services at primary health facilities

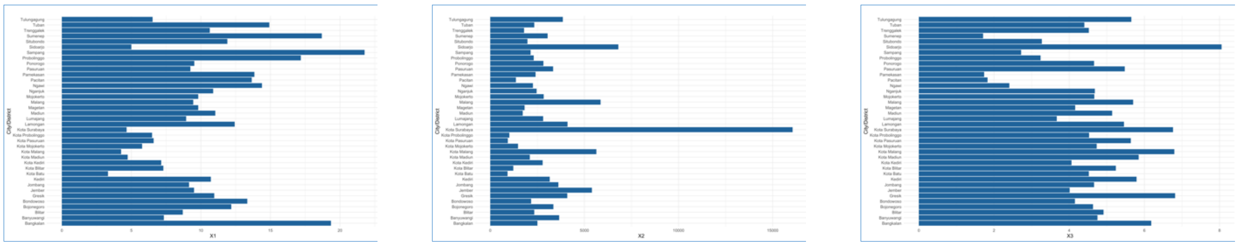
Based on unmet need data, there are 23 districts/cities where the unmet need target status has not been achieved and 15 districts/cities where the unmet need target status has been achieved.

Descriptive Analysis

Table 2. Desciptive Statistics of Research Variables

Variable	Category	Mean	Median	Minimum	Maximum
X_1	1	9.562	7.830	1.810	30.30
	0	7.755	4.770	1.610	19.800
X_2	1	41.98	43.46	33.34	46.48
	0	41.53	41.66	37.49	44.54
X_3	1	78.12	78.32	57.31	91.11
	0	70.65	74.49	46.91	87.91

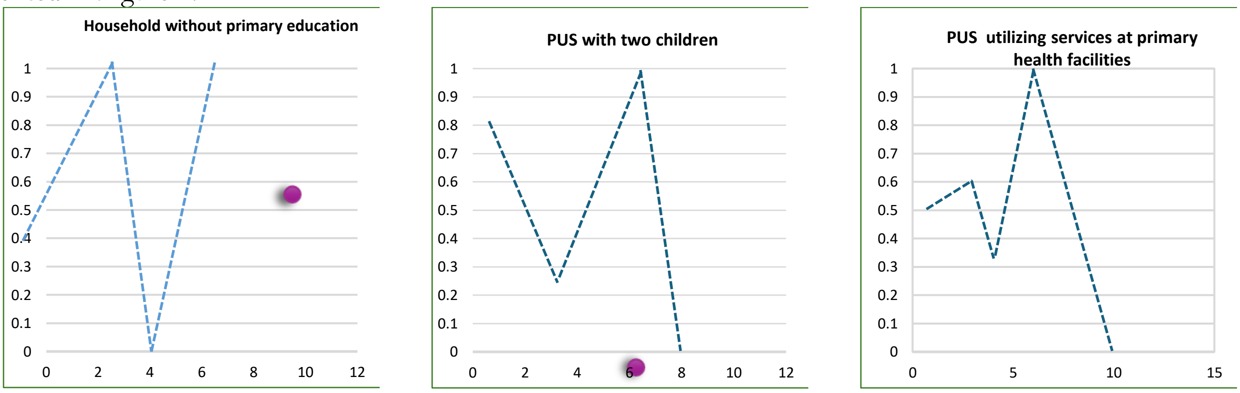
Table 2 summarized the descriptive statistics for the three predictor variable employed in the study, reflecting characteristics of 38 cities/districts in East Java. All independent variables are have no missing values, and show no indication of multicollinearity [25]. In addtion, the characteristics data for each variables shown thorugh in below:



(a) (b) (c)

Fig. 1. Characteristics of Each Predictor Variable

Fig. 1. (a) shown that Sampang is the highest, and Madiun is the lowest proportion of household without primary education. Fig. 1. (b) shows that Mojokerto is the highest, and Sampang is the lowest proportion of PUS with two children, and Fig. 1. (c) shows that Sampang is the highest, and Madiun is the lowest proportion of PUS utilizing services at primary health facilities. To explore possible nonlinear associations between each variable, scatterplots created of grouped data against the proportion of unmet need target unachieved [21]. These are presented in Figure 2.



(a) (b) (c)

Fig. 2. Scatterplot of Each Predictor Variable

As illustrared in Figure 2, the scatterplots indicated that each predictor shows shows that each predictor variable changes pattern at certain sub-intervals, with extremely increases and decreases. This nonlinear behavior suggests the suitability of applying a truncated spline approach. This case is suitable for using a truncated spline approach. Therefore, this study continues by comparing the Binary Logistic Regression (BLR) model with the TSNBLR model. Based on equation (4), the TSNBLR for achievement status data of target unmet need in East Java is as follows:

$$\pi(\boldsymbol{x_i}) = \frac{\exp\left(\gamma_0 + \gamma_{11}x_{1i} + \gamma_{21}x_{2i} + \gamma_{31}x_{3i} + \sum_{j=1}^{3}\sum_{u=1}^{r}\gamma_{j(1+u)}\left(x_{ji} - K_{ju}\right)_+\right)}{1 + \exp\left(\gamma_0 + \gamma_{11}x_{1i} + \gamma_{21}x_{2i} + \gamma_{31}x_{3i} + \sum_{j=1}^{3}\sum_{u=1}^{r}\gamma_{j(1+u)}\left(x_{ji} - K_{ju}\right)_+\right)}$$

The Optimal Knots Point

To determine the optimal mode, the analysis was carried out using up to three knot points. The knot point locations along with the corresponding AIC values are presented in Table 3. As shown in the table, the optimal knot points identified are 8.71, 12.27, and 19.37 for X_1, 41.55, and 43.20 for X_2, then 63.49, and 80.06 for X_3 is a model with an minimum AIC value of 30.98.

Table 3. Knot Point Candidate

Number of Knot Points	K_{1u}			K_{2u}			K_{3u}			AIC
	K_{11}	K_{12}	K_{13}	K_{21}	K_{22}	K_{23}	K_{31}	K_{32}	K_{33}	
1,1,1	22.93			34.92			63.49			46.40
1,2,2	22.93			39.91	41.55		63.49	74.54		47.72
3,1,3	8.71	12.27	19.37	41.55			69.01	74.54	80.06	31.72
⋮										
2,2,2	12.27	22.92		39.91	41.55		63.49	80.06		41.29

3,2,2	**8.71**	**12.27**	**19.37**	**41.55**	**43.20**		**63.49**	**80.06**		**30.98***
				⋮						
3,3,3	8.71	12.27	19.37	41.55	43.20	44.84	69.01	74.54	80.06	34.22

The TSNBLR model based on the optimal knot points is given as follows:

$$\pi(\boldsymbol{x}_i) = \frac{\exp(z)}{1 + \exp(z)}$$

Where

$$\begin{aligned} z = \gamma_0 + \gamma_{11}x_{1i} + \gamma_{12}(x_{1i} - 8.71)_+ + \gamma_{13}(x_{1i} - 12.27)_+ + \gamma_{14}(x_{3i} - 19.37)_+ + \gamma_{21}x_{2i} \\ + \gamma_{22}(x_{2i} - 41.55)_+ + \gamma_{23}(x_{2i} - 43.20)_+ + \gamma_{31}x_{3i} + \gamma_{32}(x_{3i} - 63.49)_+ \\ + \gamma_{33}(x_{3i} - 80.06)_+ \end{aligned}$$

Parameter Estimation Results of TSBLR

Model parameter estimates derived from the TSBLR method, applied to the unmet need target achievement data in East Java, are displayed in the subsequent table.

Table 4. Parameter Estimation Results TSBLR

Parameters	Estimations	z value	Pr(>|z|)	Desicion
γ_0	-10.970	-1.37	0.1706	fail to reject H_0
γ_{11}	-0.022	-0.075	0.9405	fail to reject H_0
γ_{12}	1.345	0.009	0.993	fail to reject H_0
γ_{13}	-4.256	-0.01	0.9922	fail to reject H_0
γ_{14}	7.090	0.009	0.9925	fail to reject H_0
γ_{21}	-0.650	1.213	0.2251	fail to reject H_0
γ_{22}	2.146	-1.33	0.1836	fail to reject H_0
γ_{23}	-1.032	1.251	0.211	fail to reject H_0
γ_{31}	0.634	1.734	0.0829	reject H_0
γ_{32}	-0.958	-1.778	0.0754	reject H_0
γ_{33}	0.755	0.047	0.9629	fail to reject H_0

Based on the estimation results in Table 4, the TSBLR model is given as follows:

$$\pi(\boldsymbol{x}_i) = \frac{\exp(z)}{1 + \exp(z)}$$

Where

$$\begin{aligned} z = -10.970 - 0.022x_{1i} + 1.345(x_{1i} - 8.71)_+ - 4.256(x_{1i} - 12.27)_+ + 7.090(x_{3i} - 19.37)_+ \\ - 0.650x_{2i} + 2.146(x_{2i} - 41.55)_+ - 1.032(x_{2i} - 43.20)_+ + 0.634x_{3i} \\ - 0.958(x_{3i} - 63.49)_+ + 0.755(x_{3i} - 80.06)_+ \end{aligned}$$

Parameter evaluation in the model was carried out through hypothesis tests. Based on Table 4, the results indicated that only γ_{31} and γ_{32} are statistically significant. The variable PUS utilizing services at primary health facilities is statistically significant in explaining the probability of achieving the unmet need target. This implies that the higher the PUS utilizing services at primary health facilities, the more likely it is for a region to meet the unmet need target. Other predictors did not show statistically significant effects on the response variable.

The refined model using only statistically significant parameters is:

$$\pi(\boldsymbol{x}_i) = \frac{\exp(-10.970 - 0.022x_{1i} + 1.345(x_{1i} - 8.71)_+)}{1 + \exp(-10.970 - 0.022x_{1i} + 1.345(x_{1i} - 8.71)_+)}$$

Parameter Estimation Results of BLR

The BLR model for achievement status data of target unmet need in East Java is as following table:

Table 5. Parameter Estimation Results Parametric

Parameters	Estimations	z value	Pr(>|z|)	Desicion
β_0	-8.421	-1.365	0.1723	fail to reject H_0
β_1	-0.025	-0.322	0.7475	fail to reject H_0
β_2	-0.081	0.602	0.5474	fail to reject H_0
β_3	0.076	1.712	0.0868	Reject H_0

Based on the estimation results in Table 5, the BLR model is given as follows:

$$\pi(\boldsymbol{x_i}) = \frac{\exp(-10.970 - 0.022x_{1i} + 1.345(x_{1i} - 8.71)_+)}{1 + \exp(-10.970 - 0.022x_{1i} + 1.345(x_{1i} - 8.71)_+)}$$

Based on Table 5, the results indicated that only β_3 (PUS utilizing services at primary health facilities) is statistically significant in explaining the probability of achieving the unmet need target. The refined model using only statistically significant parameters is:

$$\pi(\boldsymbol{x_i}) = \frac{\exp\,(0.076x_{3i})}{1 + \exp\,(0.076x_{3i})}$$

Comparing of TSNBLR and BLR

The regression model chosen is the model with the smallest deviance value. Based on the deviance statistical test, the results obtained are as follows:

Table 6. Comparison of Deviance Values

Model	Deviance Values
TSNSBLR	29.430
BLR	46.123

The deviance value for TSBLR (29.430) is smaller than the deviance value for the BLR (46.123). So that the TSBLR model is a better model for achievemet status data of unmet need target in East Java, Indonesia. To assess the performance of regression models for binary response data, classification criteria are employed.

Table 7. Comparison of Evaluation Criteria Value

Criteria	**Model**	
	TSNBLR	**BLR**
Accuracy	81.58%	68.42%
Sensitivity	91.30%	86.96%
Specificity	66.76%	40%
Precision	80.77%	68.97%

The accuracy value represents the proportion of correct classifications relative to the total number of observations. As shown in Table 7, the TSNBLR model achieves higher accuracy (81.58%) compared to BLR (68.42%), indicating that the nonparametric approach provides more reliable overall predictions. Sensitivity, which reflects the model's capability to correctly detect positive cases, is also superior in TSNBLR (91.30%) relative to BLR (86.96%), suggesting greater effectiveness in identifying positive outcomes. Specificity, describing the ability to correctly classify negative cases, is considerably better in TSNBLR (66.76%) than in BLR (40%), implying that the nonparametric model reduces false positives more effectively. Precision values indicates the proportion of predicted positives that are truly positive, is 80.77% for TSNBLR compared with 68.97% for BLR, demonstrating that the nonparametric model provides more reliable positive predictions. Overall, the TSNBLR performed better than the BLR. This nonparametric model showed advantages in accuracy, sensitivity, specificity and precision, making it a more appropriate choice for the analysis of data with similar pattern data characteristics.

DISCUSSION

Theoretically, the development of TSNBLR is motivated by the need to flexibly capture nonlinear effects of relationships between variable. In this study, parameter estimation of the proposed model was carried out using the Maximum Likelihood Estimation (MLE) method. The hypothesis testing framework was also established, consisting of simultaneous testing using the Likelihood Ratio Test (LRT) and partial testing using the Wald Test. This theoretical framework allows for rigorous evaluation of both the overall model fit and the contribution of each predictor. The empirical results show that the truncated spline model provides significant improvement compared to the BLR. While the logistic regression relies solely on parametric assumptions, the truncated spline approach enabling the model to accommodate more complex structures in the data. Specifically, the PUS utilizing services at primary health facilities was found to be significant in determining the probability of achieving the unmet need target. This emphasizes the essential role of health service utilization in reducing unmet need. Furthermore, the classification accuracy of the TSNBLR was higher than that of the BLR, indicating its superiority in predictive performance. These findings highlight the practical advantage of the truncated spline approach, as it not only provides theoretical contributions in extending binary response models with hypothesis testing, but also demonstrates empirical relevance in capturing important policy-related determinants of unmet need.

CONCLUSSION

This study contributes to the development of statistical modeling by extending statistic method for binary response into a truncated spline framework. Theoretical results show that the TSNBLR model provides a more flexible representation of predictor–response relationships, as it is able to capture varying patterns across different subintervals. Parameter estimation was carried out through the MLE method, and hypothesis testing was performed using both the LRT and the Wald test, ensuring rigorous inference within the proposed framework.

Applied to the case of unmet need in East Java, Indonesia, the TSNBLR demonstrated superior performance compared to the BLR. In particular, the proportion of couples of childbearing age (PUS) utilizing services at primary health facilities was identified as significant factor influencing the probability of achieving the unmet need target. Furthermore, the model achieved higher classification accuracy, confirming its empirical advantage.

Overall, the findings highlight both theoretical and practical contributions. Theoretically, the TSNBLR model advances the methodology for handling binary response data with nonlinear relationships across subintervals. Practically, it offers a more reliable basis for policy formulation in health and family planning programs, supporting progress toward the Sustainable Development Goals (SDGs).

REFERENCES

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis. New Jersey: Wiley.

Eubank, L. R. (1999). Nonparametric Regression and Spline Smoothing (2nd ed.). Texas: CRC Press.

Bilodeau, M. (1992). Fourier Smoother and Additive Models. Canada: The Canadian Journal of Statistics.

Li, L. (2003). Non-linear wavelet-based density estimators under random censorship. Journal of Statistical Planning and Inference, 117(1), 35–58.

Wang, Z., Xu, H., Xia, L., Zou, Z., & Soares, C. G. (2020). Kernel-based support vector regression for nonparametric modeling of ship maneuvering motion. Ocean Engineering, 216, 107994.

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. Journal of the American Statistical Association, 115(531), 1449–1455.

Wahba, G. (1990). Spline Models for Observational Data. Madison: SIAM.

Budiantara, N. I., Ratna, M., Zain, I., & Wibowo, W. (2012). Modeling the percentage of poor people in Indonesia using spline nonparametric regression approach. International Journal of Basic & Applied Sciences IJBAS-IJENS, 12(06), 119–124.

Adibmanesha, L., & Rashidiniab, J. (2021). Sinc and B-spline scaling functions for time-fractional convection-diffusion equations. Journal of King Saud University - Science, 33(2), 101343.

Iqbal, M., et al. (2024). A modified basis of cubic B-spline with free parameter for linear second order boundary value problems: Application to engineering problems. Journal of King Saud University - Science, 36(9), 103397.

Zhang, Y., Hua, L., & Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. Scandinavian Journal of Statistics, 37(2), 338–354.

Hazelton, M. L., & Turlach, B. A. (2011). Semiparametric regression with shape-constrained penalized splines. Computational Statistics & Data Analysis, 55(10), 2871–2879.

Lestari, B., Chamidah, N., Budiantara, I. N., & Aydin, D. (2023). Determining confidence interval and asymptotic distribution for parameters of multiresponse semiparametric regression model using smoothing spline estimator. Journal of King Saud University - Science, 35(5), 102664.

Islamiyati, A., Kalondeng, A., Sunusi, N., Zakir, M., & Amir, A. K. (2022). Biresponse nonparametric regression model in principal component analysis with truncated spline estimator. Journal of King Saud University - Science, 34(3), 101892.

Octavanny, M. A. D., Budiantara, I. N., Kuswanto, H., & Rahmawati, D. P. (2020). Nonparametric regression model for longitudinal data with mixed truncated spline and Fourier series. Abstract and Applied Analysis, 2020, 4710745.

Laome, L., Budiantara, I. N., & Ratnasari, V. (2023). Estimation curve of mixed spline truncated and Fourier series estimator for geographically weighted nonparametric regression. Mathematics, 11(1), 152.

Sriliana, I., Budiantara, I. N., & Ratnasari, V. (2022). A truncated spline and local linear mixed estimator in nonparametric regression for longitudinal data and its application. Symmetry, 14(12), 2687.

Zulfadhli, M., Budiantara, I. N., & Ratnasari, V. (2024). Fourier smoother and additive models. MethodsX, 13, 102983.

Suriaslan, A. S., Budiantara, I. N., & Ratnasari, V. (2025). Nonparametric regression estimation using multivariable truncated splines for binary response data. MethodsX, 14, 103084.

Suriaslan, A. S., Budiantara, I., & Ratnasari, V. (2025). Truncated spline regression for binary response: A comparative study of nonparametric and semiparametric approaches. Communications in Mathematical Biology and Neuroscience, 51, 9209.

Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression. New York: Wiley.

Hogg, R. V., & Craig, A. T. (1995). Introduction to Mathematical Statistics (4th ed.). New Jersey: Prentice Hall.

Pawitan, Y. (2001). In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford: Clarendon Press.

Robert, V. H., McKean, J., & Craig, A. T. (2019). Introduction to Mathematical Statistics: Probability and Distributions (7th ed.). Pearson.

Hamie, H., Hoayek, A., El-Ghoul, B., & Khalifeh, M. (2022). Application of non-parametric statistical methods to predict pumpability of geopolymers for well cementing. Journal of Petroleum Science and Engineering, 212, 110333.

