JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI e-ISSN: 2614-8811
Published by Departement of Mathematics, Hasanuddin University, Indonesia
https://journal.unhas.ac.id/index.php/jmsk/index

Vol. 21, No. 3, May 2025, pp. 698-707
DOI: 10.20956/j.v21i3.42579

Log Normal Regression and its Application

Ni Luh Sri Diantini[1], Calvin Riswandi[1]

[1]Statistics Department Matana University
***Email**: [1]niluh.sridiantini@matanauniversity.ac.id,*
[1]calvin.riswandi@student.matanauniversity.ac.id
Received: 19 December 2024, revised: 17 March 2025, accepted: 24 March 2025

Abstract

The log-normal distribution represents a type of continuous probability distribution that is characterized by a positive skew, which signifies a long tail on the right. A log-normal distribution describes a statistical distribution of values that have been logarithmically transformed from a related normal distribution. In situations where predictor variables affect positive outcomes, log-normal regression becomes significant. This research will construct a regression model that utilizes a continuous response variable following a log-normal distribution, known as log-normal regression (LNR). The purpose of developing the LNR model is to overcome the assumptions in classical regression that are often unmet, such as normality, since not all data can full fill these assumptions. Therefore, need an alternative method that does not require the normality assumption. One method that can be employed is a regression with a specific distribution approach, such as the log-normal distribution, known as log-normal regression. The LNR model will be developed through the maximum likelihood estimation (MLE) approach to achieve parameter estimation using a numerical method based on the Newton-Raphson iteration. Following this, hypothesis testing will be performed using the maximum likelihood ratio test (MLRT) and a partial test that employs the Wald test. The ultimate objective of this research is to illustrate how to apply the proposed LNR model to real data. LNR model will be applied to analysis the number of poor people in Indonesia, examining the factors that contribute to this issue. The results obtained in this study that variables human development index, unemployment rate, percentage of gross regional domestic product, and minimum wage in provinces influencing significance the number of poor people in Indonesia.

Keywords: Log normal regression, maximum likelihood estimation, maximum likelihood ratio test, number of poor people.

1. INTRODUCTION

The log-normal distribution represents a category of continuous probability distributions characterized by a positive skewness, thin shoulders, indicating a long tail on the right side [9] [10]. This distribution is founded on the statistical distribution of logarithmic values that come from a related normal distribution. The log-normal distribution is especially applicable for cases where the variable in question must exceed zero, since the logarithm of a number is only meaningful when that number is positive. There are several justifications for utilizing log-normal distributions alongside normal distributions. Generally, the logarithm signifies the power to which a base number must be raised to yield the random variable depicted along a normally distributed curve. Across all log-normal distribution graphs, the logarithm of random variables is plotted from a normal distribution curve [5].

One statistical method that can be utilized for analysis and estimating the relationships between a response variable and one or more predictor variables is regression analysis. The regression model operates under the assumption that the errors in the model must adhere to normality; therefore, the response variable should ideally be normally distributed. However, in reality, not all datasets follow a normal distribution, prompting the need for alternative distributions for the response variable, one of which is the lognormal distribution. Regression model that incorporates a continuous response variable following a lognormal distribution, along with its predictor variables, is referred to as lognormal regression (LNR) [6]. The LNR model, developed by [7] [19], employs the maximum likelihood estimation (MLE) method to derive parameter estimates and subsequently uses the Wald test for hypothesis testing.

Research related to log normal regression was also conducted by [14] using log normal regression to predicting household per capita expenditure in Poso Regency. The result of these study show that log-normal regression outperforms and predicts more accurately than the traditional method. Diantini et al., (2023) [4] were developed log-normal regression with three parameters and Budinirmala et al., (2024) [3] were developed bivariate log-normal regression. The results of these two studies showed that the MLE method was used to estimate parameters and the maximum likelihood ratio test method was used to conduct hypothesis testing. However, in that study the model obtained had not been applied to real data.

In this research, the LNR model will be developed using the MLE method to estimate the parameters. However, a closed-form solution is not achievable. Hence, a numerical approach is required to find the parameter estimates. The numerical optimization method applied in this research is Newton Raphson. Following that, hypothesis testing will be conducted using the maximum likelihood ratio test (MLRT). The final goal of this research is to demonstrate the application of the proposed LNR model to real-world data. The case study in this research focuses on factors influence the number of poor people on the provinces in Indonesia.

Previous studies have discussed several factors that influence poverty and the number of poor people. Sambuaga et al., (2024) [15] were conducted the research to analyze the factors that influence the number of poor people in North Sulawesi Province using the multiple linear regression method. Furthermore, Praja et al., (2023) [13] were conducted the research in DKI Jakarta using the panel data regression method with a fixed effect model approach. Based on previous studies, several factors that influence the number of poor people had been discussed, but have not considered the minimum wage for each province, city, or regency. Thus, in this study, one independent variable will be added that may influence the number of poor people, namely the provincial minimum wage (UMP). Furthermore, previous studies used methods that must meet classical assumptions, but in realization not all data are able to meet these assumptions. Therefore, this study uses a method with a distribution approach from the data used.

Based on the details provided, this study aims to: (i) develop the LNR model by using MLE method, (ii) assess the parameters by using MLRT method, and (iii) analysis the significance of the model applied to data on the number of poor people on the provinces in Indonesia. The predictor variables consist of the human development index (HDI), unemployment rate (TPT), percentage of gross regional domestic product (GDRP), and provincial minimum wage (UMP).

2. LOG NORMAL REGRESSION

Suppose n sample random that follows lognormal distribution and the corresponding predictor variables $(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{j})$. In this part, we examine how the LNR model is constructed, along with its parameter estimation and hypothesis testing. Based on [12] [10], a random variable Y is said to follow a log-normal distribution with characterized by the parameters, a location parameter (μ) and a squared scale parameter (σ^2), represented as $Y \sim LN(\mu, \sigma^2)$, with the probability density function outlined in Equation. (2.1)

$$f(y) = \begin{cases} \frac{1}{y\sqrt{2\pi}\sigma} \exp\left\{-\frac{[\ln(y) - \mu]^2}{2\sigma^2}\right\}, -\infty \leq \mu \leq \infty, \sigma > 0 \\ 0, \text{ otherwise} \end{cases} \quad (2.1)$$

Then, the statistics as follows [9] [17]

$$\mu = E(Y) = exp\left(\mu + \frac{\sigma^2}{2}\right), \qquad Var(Y) = \exp(2\mu + \sigma^2)\left[\exp(\sigma^2) - 1\right]$$

Here is regression model with response variable follows the lognormal distribution $Y \sim LN(\mu, \sigma^2)$ and several predictors $x_1, x_2, \dots, x_j$ [6],

$$\mu_i = \ln\left(\mathbf{x_i^T}\boldsymbol{\beta}\right) - \frac{\sigma^2}{2}, i = 1,2, \dots, n \quad (2.2)$$

where $\mathbf{x_i}$ represents the vector of predictor variables, while $\boldsymbol{\beta}$ denotes the parameter vector associated with the LNR model for the response variable.

$$\text{with } \mathbf{x_i} = [1 \quad x_{i1} \quad \cdots \quad x_{ik}]^{\boldsymbol{T}}_{(k+1)\times 1}, \qquad \boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \cdots \quad \beta_k]^{\boldsymbol{T}}_{(k+1)\times 1}$$

from the equation (2.2) was obtained $\mu_i = \ln\left(\mathbf{x_i^T}\boldsymbol{\beta}\right) - \frac{\sigma^2}{2}$, then μ is substituted into Equation (2.1). Therefore, the probability density function (PDF) of LNR model as follows

$$f(y_i) = \begin{cases} \frac{1}{y_i\sqrt{2\pi}\sigma} \exp\left(-\frac{\left[\ln y_i - \left(\ln\left(\mathbf{x_i^T}\boldsymbol{\beta}\right) - \frac{\sigma^2}{2}\right)\right]^2}{2\sigma^2}\right), y > 0 \\ 0, \text{otherwise} \end{cases} \quad (2.3)$$

In earlier discussions, we covered the topic of parameter estimation for LNR using maximum likelihood estimation (MLE). The likelihood function was developed from the equation provided (2.3) is:

$$L(\boldsymbol{\beta}, \sigma) = \prod_{i=1}^{n} \frac{1}{y_i\sqrt{2\pi}\sigma} \exp\left(-\frac{\left[\ln(y_i) - \left(\ln\left(\mathbf{x_i^T}\boldsymbol{\beta}\right) - \frac{\sigma^2}{2}\right)\right]^2}{2\sigma^2}\right) \quad (2.4)$$

The ln likelihood function from equation (2.4) is

$$\ell(\boldsymbol{\beta}, \sigma) = \ln\left[\prod_{i=1}^{n} \frac{1}{y_i\sqrt{2\pi}\sigma} \exp\left(-\frac{\left[\ln(y_i) - \left(\ln\left(\mathbf{x_i^T}\boldsymbol{\beta}\right) - \frac{\sigma^2}{2}\right)\right]^2}{2\sigma^2}\right)\right]$$

$$\ell(\boldsymbol{\beta},\sigma) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \sum_{i=1}^{n}\ln(y_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\ln(y_i) - \ln\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right) + \frac{\sigma^2}{2}\right]^2 \quad (2.5)$$

The initial derivatives of the logarithm of the likelihood function for every parameter are as follows.;

$$\frac{\partial\ell(\boldsymbol{\beta},\sigma)}{\partial\boldsymbol{\beta}} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}\left(\ln(y_i) - \ln\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right) + \frac{\sigma^2}{2}\right)$$

$$\frac{\partial\ell(\boldsymbol{\beta},\sigma)}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} + \sum_{i=1}^{n}\left(\ln(y_i) - \ln\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right)^2 - \frac{n\sigma}{4}$$

To identify a maximum likelihood (ML), find the solution, all the previously mentioned derivatives must be set to zero, leading to a system of equations. Nonetheless, a closed-form solution for this system cannot be achieved. Therefore, a numerical method is required to determine the solution, and one such technique that can be employed is the Newton-Raphson algorithm, outlined below.

- Step 1. Establish the initial value for $\widehat{\boldsymbol{\theta}}_{(0)} = \left[\widehat{\sigma}_{(0)} \quad \widehat{\boldsymbol{\beta}}_{(0)}\right]^{\mathrm{T}}$, where $\widehat{\sigma}_{(0)} > 0$ and $\widehat{\boldsymbol{\beta}}_{(0)}$ are derived from the linear regression estimation.
- Step 2. Construct the gradient vector $\mathbf{g}\left(\widehat{\boldsymbol{\theta}}\right) = \left[\frac{\partial\ell(\boldsymbol{\beta},\sigma)}{\partial\boldsymbol{\beta}} \quad \frac{\partial\ell(\boldsymbol{\beta},\sigma)}{\partial\sigma}\right]$ which is a vector containing the initial derivatives of the likelihood function in natural logarithm form for all of the estimated parameters. Subsequently, the Hessian matrix $\mathbf{H}\left(\widehat{\boldsymbol{\theta}}\right)$ is formed with each element representing the second derivatives of the natural logarithm likelihood function for the estimated parameters.
$$\mathbf{H}\left(\widehat{\boldsymbol{\theta}}_{(\mathrm{m})}\right) = \begin{bmatrix} \frac{\partial^2\ell(\boldsymbol{\beta},\sigma)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} & \frac{\partial^2\ell(\boldsymbol{\beta},\sigma)}{\partial\boldsymbol{\beta}\partial\sigma} \\ \text{simetris} & \frac{\partial^2\ell(\boldsymbol{\beta},\sigma)}{\partial\sigma^2} \end{bmatrix}$$
The elements of second partial derivative of the $\ell(\boldsymbol{\beta},\sigma)$ function is according to;
$$\frac{\partial^2\ell(\boldsymbol{\beta},\sigma)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_i}{\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)^2}\left(1 + \ln(y_i) - \ln\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right) + \frac{\sigma^2}{2}\right)$$
$$\frac{\partial^2\ell(\boldsymbol{\beta},\sigma)}{\partial\sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}\left(\ln(y_i) - \ln\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right)^2 - \frac{n}{4}$$
$$\frac{\partial^2\ell(\boldsymbol{\beta},\sigma)}{\partial\boldsymbol{\beta}\partial\sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}\frac{\mathbf{x}_i^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}\left(\ln(y_i) - \ln\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right)$$
- Step 3. Begin the Newton-Raphson iteration utilizing the formula below [11].
$$\widehat{\boldsymbol{\theta}}_{(\mathrm{m}+1)} = \widehat{\boldsymbol{\theta}}_{(\mathrm{m})} - \mathbf{H}^{-1}\left(\widehat{\boldsymbol{\theta}}_{(\mathrm{m})}\right)\mathbf{g}\left(\widehat{\boldsymbol{\theta}}_{(\mathrm{m})}\right)$$
with, $\mathrm{m} = 1,2,\dots,\mathrm{m}^*$
- Step 4. The iteration stops at the $\mathrm{m}^* - th$ iteration when $\left\|\widehat{\boldsymbol{\theta}}_{(\mathrm{m}+1)} - \widehat{\boldsymbol{\theta}}_{(\mathrm{m})}\right\| \le \varepsilon$. i.e ε is the very small tolerance positive value. During convergence, the final iteration yields an estimated value for every parameter.

3. HYPOTHESIS TESTING

The hypothesis for simultaneous testing of parameters is stated as follows:

$H_0 = \beta_1 = \beta_2 = \cdots = \beta_k$

$H_1 = \text{at least one } \beta_j \neq 0\,;\, j = 1,2,3,\ldots,k$

The set parameter under the population is denoted as $\Omega = \{\sigma, \beta_1, \beta_2, \ldots, \beta_k\}$. In contrast, the set parameter under the null hypothesis is represented by $\omega = \{\sigma, \beta_0\}$. The likelihood ratio test (LRT) method is used to obtain the critical region of hypothesis testing by comparing the ratio of the maximum value of the likelihood function under H_0 ($L(\omega)$), with the maximum value of the likelihood function under the population ($L(\Omega)$).

The likelihood ratio test (LRT) approach is utilized to determine the critical region for hypothesis testing by evaluating the ratio of the highest value of the likelihood function under the null hypothesis ($L(\omega)$) to the highest value of the likelihood function under the alternative hypothesis ($L(\Omega)$).

$$\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} < \Lambda_0, \qquad 0 < \Lambda_0 \leq 1$$

The $L(\Omega)$ was developed in equation (4). The $L(\omega)$ is as follows

$$L(\omega) = \prod_{i=1}^{n} \frac{1}{y_i\sqrt{2\pi}\sigma_\omega} \exp\left(-\frac{\left[\ln(y_i) - \left(\ln(\beta_{0\omega}) - \frac{\sigma_\omega^2}{2}\right)\right]^2}{2\sigma_\omega^2}\right) \tag{3.1}$$

The same steps as before, then obtained ln likelihood function under H_0 is

$$\ell(\omega) = \ln\left[\prod_{i=1}^{n} \frac{1}{y_i\sqrt{2\pi}\sigma_\omega} \exp\left(-\frac{\left[\ln(y_i) - \left(\ln(\beta_{0\omega}) - \frac{\sigma_\omega^2}{2}\right)\right]^2}{2\sigma_\omega^2}\right)\right]$$

$$\ell(\omega) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma_\omega - \sum_{i=1}^{n}\ln(y_i) - \frac{1}{2\sigma_\omega^2}\sum_{i=1}^{n}\left[\ln(y_i) - \ln(\beta_{0\omega}) + \frac{\sigma_\omega^2}{2}\right]^2 \tag{3.2}$$

The initial partial derivative of the logarithm of the likelihood function under the null hypothesis as follows.;

$$\frac{\partial \ell(\omega)}{\partial \beta_{\omega 0}} = \frac{1}{\sigma_\omega^2 \beta_{0\omega}} \sum_{i=1}^{n}\left(\ln(y_i) - \ln(\beta_{0\omega}) + \frac{\sigma_\omega^2}{2}\right)$$

$$\frac{\partial \ell(\omega)}{\partial \sigma_\omega} = -\frac{n}{\sigma_\omega} + \frac{1}{\sigma_\omega^3}\sum_{i=1}^{n}(\ln(y_i) - \ln(\beta_{0\omega}))^2 - \frac{n\sigma_\omega}{4}$$

The first partial derivative shown does not have a closed form, and the maximum likelihood estimator lacks analytical solutions. Therefore, by following the same procedure earlier a numerical analytical technique needs to apply for estimating parameters under H_0.

Based on equation (3.2) and (2.5) obtained the odds ratio as follows

$$\Lambda = \frac{L(\widehat{\omega})}{L(\widehat{\Omega})} = \frac{\prod_{i=1}^{n} \frac{1}{y_i\sqrt{2\pi}\widehat{\sigma}_\omega} \exp\left(-\frac{\left[\ln(y_i) - \left(\ln(\widehat{\beta}_{0\omega}) - \frac{\widehat{\sigma}_\omega^2}{2}\right)\right]^2}{2\widehat{\sigma}_\omega^2}\right)}{\prod_{i=1}^{n} \frac{1}{y_i\sqrt{2\pi}\widehat{\sigma}} \exp\left(-\frac{\left[\ln(y_i) - \left(\ln(\mathbf{x}_i^T\widehat{\boldsymbol{\beta}}) - \frac{\widehat{\sigma}^2}{2}\right)\right]^2}{2\widehat{\sigma}^2}\right)} < \Lambda_0 \quad (3.3)$$

than the statistic uji of the MLRT method is

$$\begin{aligned} G^2 &= -\ln\Lambda^2 = -\ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right)^2 \\ &= 2\left[\ln L(\widehat{\Omega}) - \ln L(\widehat{\omega})\right] \end{aligned} \quad (3.4)$$

with, G^2 is approach the chi square distribution by the large sample. Then the critical region of rejected H_0 is $G^2 > \chi^2_{\alpha;df}$, $df = n(\Omega) - n(\omega)$, $n(\Omega)$ represents the quantity of parameters within the population, while (ω) denotes the quantity of parameters within the null hypothesis.

Next steps are the partial hypothesis testing. The hypothesis for partial testing of parameters is stated as follows:

$H_0 = \beta_j = 0$

$H_1 = \beta_j \neq 0\ ;\ j = 1,2,3,\ldots,k$

According to [3], the statistic uji is stated in equation

$$Z = \frac{\widehat{\beta}_j}{SE(\widehat{\beta}_j)},$$

with, $SE(\widehat{\beta}_j) = \sqrt{\widehat{\text{var}}(\widehat{\beta}_j)}$. the $\sqrt{\widehat{\text{var}}(\widehat{\beta}_j)}$ is the diagonal entries that relate to the $-\mathbf{H}^{-1}(\widehat{\boldsymbol{\theta}})$ matrix. The rejected of the null hypothesis if $Z > Z_{\alpha/2}$.

4. DATA AND METHOD

The parameters estimation and hypothesis testing for the LNR model were carried out in accordance with standard procedures. The LNR model was characterized using the probability density function presented in Equation (2.5) for n observations, which facilitated the development of the likelihood and natural logarithm likelihood functions. The first derivative of the natural logarithm likelihood function was computed for each parameter and subsequently equated to zero. Since the solutions lacked closed forms, numerical optimization became necessary, prompting the use of the Newton-Raphson method in this research.

The parameters were estimated utilizing Maximum Likelihood Estimation (MLE). A simultaneous hypothesis test was performed to assess the significance of the LNR using the Maximum Likelihood Ratio Test (MLRT). Furthermore, the Wald test was employed for a partial evaluation of individual parameter significance within the LNR. The developed LNR model, with its parameter estimation and hypothesis testing, was implemented on real data as part of this research.

This research utilized secondary data sourced from Statistics Indonesia. The response variable considered was the number of poor people, alongside four predictor variables: the human development index, the unemployment rate, percentage of the gross regional domestic product, and province minimum wage. The data were gathered for 34 provinces in Indonesia, at the year of 2023.

Poverty is an incredibly complicated issue, making it difficult to address. It is associated with various challenges, including insufficient food and nutrition, educational shortcomings, crime, unemployment, prostitution, and difficulties stemming from low per capita income levels [1]. Poverty according to its causes is divided into two, namely poverty cultural and structural. Cultural poverty refers to poverty that arises from traditional or cultural elements within a specific community, which bind individuals or certain groups to their impoverished circumstances. This type of poverty can potentially be alleviated or even eliminated by challenging the factors that hinder individuals from improving their quality of life. Structural poverty, on the other hand, is a form of poverty that results from the powerlessness of individuals or specific groups against the prevailing system or an unjust social order, leaving them in a weak negotiating position with limited opportunities to grow and escape the cycle of poverty.

According to the Statistics Indonesia (BPS) [2], poor people are defined as those whose average monthly per capita spending falls below the poverty threshold. The BPS assesses poverty through the lens of the capability to full fill fundamental requirements, employing a basic needs approach. This perspective views poverty as the economic incapacity to satisfy both food and non-food necessities, evaluated from the standpoint of expenditure.

The Human Development Index (HDI) is a composite indicator to measure development achievements in the quality of human life. This index is derived from an average that evaluates accomplishments in three key aspects of human development: longevity and health, education, and a satisfactory standard of living. The unemployment rate represents the proportion of unemployed individuals relative to the entire workforce. An elevated unemployment rate suggests that a significant portion of the workforce is not engaged in employment opportunities. Gross Regional Domestic Product (GRDP) serves as a foundation for national or regional development strategies, particularly within the economic sector. Additionally, these GRDP statistics can be utilized to assess the outcomes of economic development efforts carried out by various stakeholders, including the central government, regional governments, and private entities [2].

5. RESULTS

Initially, important to assess the fit of the empirical data to a log-normal distribution. Inferential approach using Anderson Darling's (AD) goodness-of-fit test, since that test fit for testing right skewed distribution including log normal distribution [12] and AD test was found to be the powerful test compared to the Kolmogorov-Smirnov and Cramer von Mises tests for fit log normal distribution testing [18]. The null hypothesis asserts that the data follows a log-normal distribution, while the alternative hypothesis suggests that the data does not align with a log-normal distribution. For the response variable, the AD test yielded a test statistic of 0.386 and a p-value of 0.373. Since the p-value exceeds the significance threshold (α) of 5%, we can conclude that the response variable (Y) is consistent with a log-normal distribution. Anderson-Darling test for testing right skewed distribution

This study also examined the multicollinearity present among the predictor variables. The variance inflation factor (VIF) associated with each predictor variable is available in Table (5.1).

Table 5.1. Multicollinearity test

	X_1	X_2	X_3	X_4
VIF	1,36	1,45	1,43	1,04

The VIF values for every predictor variable are below 5, indicating that there is no multicollinearity among them and no needed corrective action [8] [16].

The following provides descriptive statistics regarding various variables for 34 provinces in Indonesia. According to Table 5.2, the mean number of individuals living in poverty in 2023 is 762 thousand, with the lowest figure in North Kalimantan at around 48 thousand and the highest in East Java at 4188 thousand. The average Human Development Index (HDI) is 72.62, with Papua recording the lowest HDI at approximately 62.25, while DKI Jakarta has the highest at 82.46. The unemployment rate averages 4.71%, with the lowest are found in West Sulawesi at 2.65% and the highest Banten at 7.74%. The percentage of gross regional domestic product stands at 2.91% with Gorontalo is the lowest at 0.25% and DKI Jakarta is the highest at 16.77%. The last one is the provincial minimum wage are stands at 2.92 billion rupiah, with the lowest is Central Java at 1.95 billion rupiah and DKI Jakarta highest at 4.90 billion rupiah.

Table 5.2. Descriptive Statistics of Data

Variable	Mean	Std. Dev	Min	Max
Number of Poor People (Y)	762	1063	47.97	4186
Human Development Index (X_1)	72.62	3.86	62.25	82.46
Unemployment Rate (X_2)	4.71	1.41	2.65	7.74
Gross Regional Domestic Product (X_3)	2.91	4.13	0.25	16.77
Provincial Minimum Wage (X_4)	2.92	0.61	1.95	4.90

Model for Number of Poor People by LNR was conducted to identify the factors influencing it in Indonesian provinces. The parameter estimation results of the LNR model are shown in Tables 5.3. Subsequently, a simultaneous hypothesis test was performed using the test statistic G^2. The value obtained for the simultaneous test statistic was 505.13, since this value is greater than $\chi^2_{(0.05;4)} = 9.4877$, the null hypothesis has been rejected, suggesting that one or more independent variables have a significant effect on the response variable. Subsequently, a partial test was performed. The Z test statistic values can be found in Table 5.3. With a significance level of 5%, it was concluded that every predictor variable significantly influences the number of poor people.

Table 5. 3. Parameter Estimation Model for Number of Poor People

Variables	Parameters	Estimate	Std Error	**Z**	p − Value
Intercept	β_0	9.33×10^3	5.14×10^{-3}	1.81×10^6	0.000
HDI	β_1	-0.95×10^2	5.84×10^{-5}	-1.63×10^6	0.000
TPT	β_2	0.65×10^2	1.42×10^{-4}	4.58×10^5	0.000
GDRP	β_3	2.13×10^2	4.47×10^{-4}	4.77×10^5	0.000
UMP	β_4	-8.82×10^2	8.07×10^{-4}	-1.09×10^6	0.000

According to Table 5.3, the variables used in this study-namely HDI, TPT, GDRP, and UMP-have a significant effect on the number of poor people. Based on previous research, Sambuaga et al., (2024) [15] stated that TPT also affects the number of poor people in North Sulawesi Province, which is consistent with the results of this study. Furthermore, Praja et al., (2023) indicated that TPT and the population growth rate have a significant effect, but HDI does not significantly affect the percentage of poor people in DKI Jakarta Province. There are differences in the results obtained, where this study found that HDI has a significant effect on the number of poor people in Indonesia. The model formed is based on Table 5.3, as described below.

$$\hat{\mu} = 9330 - 95x_1 + 65x_2 + 213x_3 - 882x_4$$

According to the model presented, it can be interpreted that if human development index, unemployment rate, percentage of gross regional domestic product, and province minimum wage are constant, it will increase number of poor people by 9330 (thousand people). An increase of one unit of the human development index will decrease 95 (thousand individuals) in the number of poor people, assuming the other predictor variables are constant. For every one-unit increase in the unemployment rate, the number of poor individuals will rise by 65 (thousand individuals), assuming the other predictor variables remain the constant. Each one-unit increase in the percentage of gross regional domestic product, assuming all other predictor variables are constant, will result in an increase of 213 (thousand individuals) in the number of poor people. Furthermore, a rise of one unit in the province minimum wage will decrease in the number of poor people by 882 (thousand individuals), with all other variables kept constant.

6. CONCLUSIONS

The estimation of parameters for the LNR model through MLE yields an equation that lacks a closed form, necessitating a numerical approach to derive the estimator values for each parameter. In this research, the numerical method employed is Newton Raphson. LNR model has been applied to analysis the cases of poverty levels in Indonesia. The results of the model reveal that the significant factors influencing poverty levels include the human development index, unemployment rate, the percentage of gross regional domestic product, and the minimum wage in provinces. Based on these findings, it is recommended for future research that scholars consider utilizing the bivariate model of LNR, acquire the data for both response and predictor variables from a singular source, and examine alternative statistical tests to identify multicollinearity. Furthermore, it is also advised for subsequent studies to incorporate the LNR model with spatial effects to develop models relevant to each specific location.

7. REFERENCES

[1] Adji, A., Hidayat, T., Tuhiman, H., Kurniawati, S., dan Maulana, A., 2020. Pengukuran Garis Kemiskinan di Indonesia: Tinjauan Teoretis dan Usulan Perbaikan. Jakarta: Tim Nasional Percepatan Penanggulangan Kemiskinan.

[2] BPS Indonesia, S. I., 2023. Catalog : 1101001. *Statistik Indonesia 2023*, *1101001*, 790. https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html

[3] Budinirmala, K., Purhadi, & Choiruddin, A., 2024. Parameter Estimation and Hypothesis Testing on Bivariate Log-Normal Regression Models. *KnE Life Sciences*, *2024*, 177–185. https://doi.org/10.18502/kls.v8i1.15546

[4] Diantini, N. L. S., Purhadi, & Choiruddin, A., 2023. Parameter estimation and hypothesis testing on three parameters log normal regression. *The 8Th International Conference and Workshop on Basic and Applied Science (Icowobas) 2021*, *2554*, 030024. https://doi.org/10.1063/5.0104443

[5] Diwakar, R., 2017. An evaluation of normal versus lognormal distribution in data description and empirical analysis. *Practical Assessment, Research and Evaluation*, *22*(13), 1–15.

[6] Gustavsson, S., Fagerberg, B., Sallsten, G., & Andersson, E. M. (2014). Regression models for log-normal data: Comparing different methods for quantifying the association between abdominal adiposity and biomarkers of inflammation and Insulin Resistance. *International Journal of Environmental Research and Public Health*, *11*(4), 3521–3539. https://doi.org/10.3390/ijerph110403521

[7] Gustavsson, S. M., Johannesson, S., Sallsten, G., & Andersson, E. M., 2012. Linear Maximum Likelihood Regression Analysis for Untransformed Log-Normally Distributed Data. *Open Journal of Statistics*, *02*(04), 389–400. https://doi.org/10.4236/ojs.2012.24047

[8] Ibrahim, S., Al-Kassab, M. M., & Al-Awjar, M. Q., 2023. Investigating Multicollinearity in Factors Affecting Number of Born Children in Iraq. *Springer Proceedings in Mathematics and Statistics*, *418*(2), 81–91. https://doi.org/10.1007/978-981-99-0447-1_7

[9] Krishnamoorthy, K., 2006. *Handbook of statistical distributions with applications*. Chapman and Hall/CRC.

[10] Moral, J., & Rubia, D., 2024. *Lognormal distribution for social researchers : A probability classic*. *16*(June), 10–25. https://doi.org/10.5897/IJPC2024.0702

[11] Nadhim, A. I., & Al-Jilawi, A. S., 2022. The bridge between Newton's method and Newton-Raphson's method in numerical optimization. *International Journal of Health Sciences*, *6*(April), 5249–5267. https://doi.org/10.53730/ijhs.v6ns1.6042

[12] Neamvonk, J., & Phuenaree, B., 2022. Assessment of Anderson-Darling and their Modified Tests for right skewed distribution. *International Journal of Mathematics and Computer Science*, *17*(3), 1327–1339.

[13] Praja, R. B., Muchtar, M., & Sihombing, P. R., 2023. Analisis Pengaruh Indeks Pembangunan Manusia, Laju Pertumbuhan Penduduk, dan Tingkat Pengangguran Terbuka terhadap Kemiskinan di DKI Jakarta. *Ecoplan*, *6*(1), 78–86. https://doi.org/10.20527/ecoplan.v6i2.656

[14] Roosyidah, N. A. N., Choiruddin, A., & Zain, I., 2024. Predicting household per capita expenditure by using log-normal regression in Poso Regency Central Sulawesi. In *Procedia Computer Science* (Vol. 234, pp. 512–519). https://doi.org/10.1016/j.procs.2024.03.034

[15] Sambuaga, O. R., Kalangi, J. B., & Siwu, H. F. D., 2024. Analisis Faktor-Faktor Yang Mempengaruhi Jumlah Penduduk Miskin Provinsi Sulawesi Utara. *Jurnal Berkala Ilmiah Efisiensi*, *24*(7), 1-14.

[16] Shrestha, N., 2020. Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

[17] Siegrist, K., 2017. *Probability, mathematical statistics, stochastic processes*.

[18] Wang, C., & Zhu, H., 2024. Tests of fit for the power function lognormal distribution. *PLoS ONE*, *19*(2 February), 1–18. https://doi.org/10.1371/journal.pone.0298309

[19] Zuanetti, D. A., Diniz, C., & Leite, J. G., 2006. a Lognormal Model for Insurance Claims Data. *REVSTAT-Statistical Journal*, *4*(2), 131–142.