BAREKENG: Jurnal Ilmu Matematika dan Terapan
March 2022 Volume 16 Issue 1 Page 171-180
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss1pp171-180

3-PARAMETER GAMMA REGRESSION MODEL FOR ANALYZING HUMAN DEVELOPMENT INDEX OF CENTRAL JAVA PROVINCE

Hasbi Yasin[1], Syarifah Inayati[2], Setiawan[3*]

[1,2,3] Department of Statistics, Institut Teknologi Sepuluh Nopember
Jl. Teknik Mesin No.175, Keputih, Sukolilo, Surabaya, Jawa Timur 60115, Indonesia
[1] Department of Statistics, Universitas Diponegoro
Jl. Prof. Soedharto, S.H., Tembalang, Semarang, Jawa Tengah 50275, Indonesia
[2] Department of Statistics, Universitas Negeri Yogyakarta,
Jl. Colombo No 1, Karangmalang, Yogyakarta 55281, Indonesia

Corresponding author e-mail: [3] setiawan@statistika.its.ac.id*

Abstract. *The number and quality of the population are one of the determining factors for the success of national development. The quality of the population of a region can be seen from Human Development Index (HDI) achieved by a region. The HDI is based on three basic dimensions: a long and healthy life, knowledge, and a decent standard of living. This study aimed to determine the factors influencing HDI in Central Java Province in 2018-2020. The data used tend to follow the 3-Parameter Gamma distribution, which implies the HDI is modelled with 3-Parameter Gamma regression. 3-Parameter Gamma Regression is a regression that explains the relationship among one or more predictor variables with response variables that follow the 3-Parameter Gamma distribution. This research also includes the preparation of algorithms and computations in modelling 3-parameter Gamma regression. The estimation of model parameters was carried out using Maximum Likelihood Estimation (MLE) and Berndt Hall Hausman (BHHH) methods. HDI modelling with 3-Parameter Gamma regression produces a coefficient of determination of 61.58%. The results show that increasing HDI can be done by increasing the Pure Participation Rate (APM) for SMP/MTs, the ratio of SMP/MTs students, population density, Labour Force Participation Rate (TPAK), the percentage of households (RT) with access to water, drinking water, and the percentage of households (RT) that have their toilet facilities, as well as by reducing the student-teacher ratio of Junior High School (SMP)/Islamic Junior High School (MTs) and the Open Unemployment Rate (TPT).*

Keywords: *BHHH, Human Development Index (HDI), 3-Parameter Gamma Regression.*

Article info:

Submitted: 18th December 2021 *Accepted: 9th February 2022*

How to cite this article:

H. Yasin, S. Inayati and Setiawan, "3-PARAMETER GAMMA REGRESSION MODEL FOR ANALYZING HUMAN DEVELOPMENT INDEX OF CENTRAL JAVA PROVINCE", *BAREKENG: J. Il. Mat. & Ter.*, vol. 16, iss. 1, pp. 171-180, Mar. 2022.

1. INTRODUCTION

Java Island is the island with the most population in Indonesia. 52.33 percent of Indonesia's population occupies Java Island, totaling 269.60 million people in 2020. As many as 12.88% live in Central Java Province. An area with a high population must be a concern of the government, especially regarding the quality of its human resources. Human resources are one of the determining factors for the success of national development through the number and quality of the population. The quality of the population of a region can be seen from the Human Development Index (HDI) achieved by a region. HDI is a single composite indicator that measures three main dimensions of human development, which are considered capable of reflecting the basic capabilities of the population. The three basic abilities are the length of life as an indicator of achievement of health status, level of education, and access to resources needed to achieve a decent standard of living [1].

The average HDI in 2018-2020 based on provinces in Java Island from the highest to the lowest is DKI Jakarta (80.67), D.I. Yogyakarta (79.83), Banten (72.28), West Java (71.81), Central Java (71.57), and East Java (71.33). In this case, Central Java is only superior to East Java and is in the "high" category. Furthermore, research data also shows that the achievement of HDI values during the period 2018 to 2020 always increases from year to year, which is 0.86 percent in 2018 to 0.85 percent in 2019, while the increase in 2020 is only 0. 2 percent. According to investor.id in BPS (2020) [1], this is due to the Covid-19, which began to hit Indonesia in early March 2020 in such a way that it caused a decline in people's income. Although there is always an increase in HDI achievement from year to year, the HDI value is likely to decrease depending on the indicators' changes.

Research on HDI, among others, was conducted by Melliana and Zain in 2013, which examined the factors that affect HDI in East Java province using Regression Panels [2]. Then, Yanthi and Budiantara, 2016 also examined the factors that affect HDI in the province of Central Java using the Spline Nonparametric Regression method [3]. Zakaria also conducted research related to HDI in 2018, which resulted in factors that affect HDI in Central Java province using Panel Regression [4]. One of the limitations of these studies is that they are still based on the assumption that HDI data must be normally distributed.

Linear regression models generally assume that the response variable follows a normal distribution. However, empirically, this assumption cannot always be met. It might happen because the data distribution is asymmetrical, or it can be tailed thinner or thicker than the normal distribution. Gamma distribution is one of the data distributions whose relaxation can capture the asymmetric pattern or the thickness of the tail of a distribution. Gamma distribution is a flexible and adaptive distribution so that it can overcome asymmetric data patterns. Therefore, the statistical inference of the regression model parameters will give better results. Thus, a more efficient approach can be obtained, which can be a solution when the response variable does not meet the assumption of data normality. The regression method with the response variable following the gamma distribution is called gamma regression [5-6].

The gamma distribution is a generalized form of the exponential distribution, first introduced in the 18th century by the Swiss mathematician Leonard Euler. There are several types of Gamma distributions. The two-parameter gamma distribution is commonly used. This distribution has form parameters and scale parameters. Then, the gamma distribution is generalized to the Generalized Gamma distribution [7-8]. Along with the development of science, then known as the 3-parameter Gamma Distribution, namely by adding a location parameter as a data threshold [9]. Application of the 3-Parameter Gamma Distribution will be more suitable if applied to data with a specific threshold value or a minimum value that is not equal to zero. The achievement of the Regency/City HDI in Central Java cannot possibly be zero in the sense that it has a particular minimum value. Then, the data distribution indicated that it was not normally distributed and tended to deviate to the right. Therefore, there is a strong indication that the HDI data in Central Java follows a 3-parameter Gamma distribution. Based on this description, this research determined the factors that influence the HDI in Central Java by using a 3-Parameter Gamma Regression approach. This research also includes the preparation of algorithms and computations in the modeling of 3-parameter Gamma regression.

2. RESEARCH METHOD

2.1 Data Source

This study used secondary data from 2018 to 2020 sourced from the Central Java Province BPS (Central Bureau of Statistics). The unit of observation was the Regency/City in Central Java Province.

2.2 Research Variables

The research variables used in this study were variables thought to have an effect on the HDI response variable as presented in Table 1.

Table 1. Research Variable

Variable	Meaning
Y	Human Development Index (IPM)
X_1	Pure Participation Rate (APM) SMP/MTs
X_2	Student-Teacher Ratio SMP/MTs
X_3	Student-School Ratio SMP/MTs
X_4	Population density
X_5	Open Unemployment Rate (TPT)
X_6	Labor Force Participation Rate (TPAK)
X_7	Percentage of Households (RT) with access to proper drinking water
X_8	Percentage of households (RT) that have their own toilet facilities

Data source: Central Java Province BPS

Based on Table 1, the variables included in the knowledge/educational dimension are the APM for SMP/MTs, the student-teacher ratio for SMP/MTs, and the ratio of students for SMP/MTs. The definition of APM for SMP/MTs is the percentage of the number of children in SMP/MTs. They are currently in school to the total number of children in the school-age group of 13 to 15 years. The SMP/MTs student-teacher ratio is between the number of SMP/MTs students and the number of SMP/MTs teachers. The ratio of SMP/MTs students is the ratio of SMP/MTs students to the number of SMP/MTs schools. Variables included in the dimensions of a decent standard of living are population density, open unemployment rate, and labor force participation rate. Population density figures show the average number of residents per 1 square kilometer. The open unemployment rate is the percentage of the number of unemployed to the total labor force. TPT is the percentage of the population aged 15 years and over in the workforce. TPAK is the percentage of the population aged 15 years and over in the workforce. The variables included in the health dimension are the percentage of households (RT) with access to proper drinking water and the percentage of households (RT) that have toilet facilities.

2.3 Gamma Regression 3 Parameter

2.3.1 Estimation of 3 Parameter Gamma Regression Parameters

The density function of the random variable Y with Gamma distribution (α,θ,γ) with *a* is the shape parameter, θ is the scale parameter, and γ is the location parameter that can be written as follows:

$$f(y|\alpha,\theta,\gamma)=\begin{cases}\dfrac{(y-\gamma)^{\alpha-1}e^{-\frac{(y-\gamma)}{\theta}}}{\theta^{\alpha}\Gamma(\alpha)} & ;y>\gamma;\alpha>0;\ \theta>0;\gamma>0\\ 0 & ;y \text{ another}\end{cases} \quad (1)$$

with: $\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt$ and $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$

The mean, variance and moment generating functions are respectively:

$E(Y)=\alpha\theta+\gamma;\ Var(Y)=\alpha\theta^2;$ and $M_Y(t)=\frac{e^{\gamma t}}{(1-\theta t)^{\alpha}}, t<\frac{1}{\theta}.$

Several gamma regression models can be arranged based on the parameters used. This study used the shape parameter as the basic model. The scale parameter was considered constant for each observation. The univariate gamma regression model for 3 parameters (α,θ,γ) with the "log" link function is as follows:

$$\mu=E(Y)=e^{x^T\boldsymbol{\beta}} \quad (2)$$

with : $x^T\boldsymbol{\beta}=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k$ where $X_1,X_2,\cdots,X_k$ is a predictor variable., or in matrix form it can be written as follows:

$$\mathbf{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ \mathbf{X}=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix},\text{ and } \boldsymbol{\beta}=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}.$$

Based on equation (2), assuming the scale parameter and location parameter (threshold) are fixed, the shape parameter equation is obtained as follows:

$$\alpha\theta+\gamma=e^{\mathbf{x}^T\boldsymbol{\beta}} \Rightarrow \alpha=\left(\frac{e^{\mathbf{x}^T\boldsymbol{\beta}}-\gamma}{\theta}\right) \tag{3}$$

By substituting equation (3) into the density equation in equation (1), the probability density function of the 3-parameter gamma regression is obtained as follows [10]–[12]:

$$f(y)=\frac{(y-\gamma)^{\left(\frac{e^{x^T\boldsymbol{\beta}}-\gamma}{\theta}-1\right)} e^{\left(-\frac{(y-\gamma)}{\theta}\right)}}{\theta^{\left(\frac{e^{x^T\boldsymbol{\beta}}-\gamma}{\theta}\right)}\Gamma\left(\frac{e^{x^T\boldsymbol{\beta}}-\gamma}{\theta}\right)};\theta>0,\gamma>0;y>\gamma \tag{4}$$

One method that can be used to estimate the parameters of the gamma regression model is the Maximum Likelihood Estimation (MLE) method. MLE is a method that maximizes the ln likelihood function to get the parameter estimator. The likelihood function of *n* random variables $y_1,y_2,\ldots,y_i,\ldots,y_n$ defined as a joint density function of n random variables. If $y_1,y_2,\ldots,y_i,\ldots,y_n$ is a random sample of the density function $f\left(y;\theta,\gamma,\boldsymbol{\beta}\right)$ then the ln likelihood function is obtained as follows.

$$L\left(\theta,\gamma,\boldsymbol{\beta}\right)=\prod_{i=1}^{n} f\left(y_i;\theta,\gamma,\boldsymbol{\beta}\right) \tag{5}$$

$$\ln L\left(\Omega\right)=\sum_{i=1}^{n}\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}-1\right)\ln\left(y_i-\gamma\right)-\sum_{i=1}^{n}\frac{\left(y_i-\gamma\right)}{\theta}-\sum_{i=1}^{n}\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)\ln\theta-\sum_{i=1}^{n}\ln\Gamma\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right) \tag{6}$$

Getting $\hat{\theta},\hat{\gamma}$ and $\hat{\boldsymbol{\beta}}$ was done by finding the first derivative of ln likelihood in equation (6) for each parameter, then equating it with zero. Apparently, a solution that is not closed-form is obtained, so numerical optimization is necessary. In this study, optimization using the *Berndt-Hall-Hall-Hausman* Algorithm (BHHH) [10-11], [13] was used. It was chosen because this algorithm does not require a second derivative to form a Hessian . matrix.

The estimation steps with MLE and the BHHH Algorithm are as follows:

1. Determine the Log Likelihood function from the Gamma regression equation $\ln L\left(\Omega\right)$ with equation (6)
2. Determine the initial estimated value of the model parameters $\Omega_{(0)}=\left(\theta_{(0)},\gamma_{(0)},\boldsymbol{\beta}_{(0)}\right)$. This study uses $\hat{\theta}_{(0)}=1$, $\hat{\gamma}_{(0)}=min(\boldsymbol{y})-1$, dan $\widehat{\boldsymbol{\beta}}_{(0)}$ is a regression parameter with the OLS method with a response ln(y) or $\widehat{\boldsymbol{\beta}}_{(0)}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\ ln(\boldsymbol{y})$. Then it is also determined and the tolerance limit for convergence is $\varepsilon>0$.
3. Creating the gradient vector $\boldsymbol{g}\left(\hat{\Omega}_{(m)}\right)$ with equation (7):

$$\mathbf{g}\left(\hat{\Omega}_{(m)}\right)=\begin{bmatrix}\sum_{i=1}^{n}\frac{1}{\theta}\left(\left(\frac{y_i-\gamma}{\theta}\right)+\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)\left\{-\ln\left(\frac{y_i-\gamma}{\theta}\right)-1+\Psi\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)\right\}\right)\\ \sum_{i=1}^{n}\left(\frac{1}{\theta}\left[1-\ln\left(\frac{y_i-\gamma}{\theta}\right)+\Psi\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)-\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}-1\right)\left(\frac{\theta}{y_i-\gamma}\right)\right]\right)\\ \sum_{i=1}^{n}\mathbf{x}_i\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{\theta}\right)\left\{\ln\left(\frac{y_i-\gamma}{\theta}\right)+\Psi\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)\right\}\end{bmatrix}_{\Omega=\hat{\Omega}_{(m)}} \tag{7}$$

4. Determine the first derivative of density ln with respect to parameter, $\mathbf{l}_i\left(\hat{\Omega}_{(m)}\right)$ for $i = 1, 2, \cdots, n$ by using the equation (8):

$$\mathbf{l}_i\left(\hat{\Omega}_{(m)}\right)=\begin{bmatrix}\frac{1}{\theta}\left(\left(\frac{y_i-\gamma}{\theta}\right)+\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)\left\{-\ln\left(\frac{y_i-\gamma}{\theta}\right)-1+\Psi\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)\right\}\right)\\ \frac{1}{\theta}\left[1-\ln\left(\frac{y_i-\gamma}{\theta}\right)+\Psi\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)-\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}-1\right)\left(\frac{\theta}{y_i-\gamma}\right)\right]\\ \mathbf{x}_i\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{\theta}\right)\left\{\ln\left(\frac{y_i-\gamma}{\theta}\right)+\Psi\left(\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}-\gamma}{\theta}\right)\right\}\end{bmatrix}_{\Omega=\hat{\Omega}_{(m)}} \tag{8}$$

5. Create a Hessian matrix

$$H\left(\hat{\Omega}_{(m)}\right) = -\sum_{i=1}^{n} l_i\left(\hat{\Omega}_{(m)}\right) l_i\left(\hat{\Omega}_{(m)}\right)^T$$

6. Perform iterations to obtain the estimated value of the parameter with the equation (9):

$$\hat{\Omega}_{(m+1)} = \hat{\Omega}_{(m)} - \boldsymbol{H}^{-1}\left(\hat{\Omega}_{(m)}\right)\boldsymbol{g}\left(\hat{\Omega}_{(m)}\right) \tag{9}$$

$\hat{\Omega}_{(m)}$ value is a set of parameter estimators in the m iteration.

7. The iteration stops if $\left\|\Omega_{(m+1)} - \Omega_{(m)}\right\| < \varepsilon$, or has reached the specified maximum iteration.

2.3.2 Hypothesis Testing Parameters of 3 Parameter Gamma Regression Model

Parameter testing was carried out to know the model's suitability and which variables had a significant effect. To test the significance of the parameters in the model, the likelihood ratio test and Z test were used. The Likelihood Ratio test was a test that compared the model that contained the independent variables and the model that did not contain the independent variables.

Hypothesis H$_0$: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

H$_1$: at least one $\beta_j \neq 0$, j = 1, 2, … , k

Test statistics:

$$G^2 = -2\ln\left(\frac{L(\omega)}{L(\Omega)}\right) = 2\ln L(\Omega) - 2\ln L(\omega) \tag{10}$$

with:

$\omega = (\theta, \gamma, \beta_0)$ is the set of model parameters below H$_0$ (Null Model)

$\Omega = (\theta, \gamma, \boldsymbol{\beta})$ is the set of model parameters below H$_1$ (Full Model)

Test Criteria: Reject H$_0$ if $G^2 > \chi^2_{(1-\alpha;k)}$ or sig < α.

To partially test the significance of the parameters, the Z test was used. This test was conducted to determine which predictor variables had a significant effect on the model..

Hypothesis:

H_0: $\beta_j = 0$

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

Test statistics:

$$Z = \frac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)} \tag{11}$$

with $SE\left(\hat{\beta}_j\right)$ is the standard error of the parameter $\hat{\beta}_j$, $SE(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j)}$ and $\widehat{Var}(\hat{\beta}_j)$ is the main diagonal element $(j+2)$ matrix $-\left[\mathbf{H}^{-1}\left(\hat{\Omega}\right)\right]$.

Test Criteria: Reject H_0 if $|Z| > Z_{\alpha/2}$ or sig < α.

2.3.3 Evaluation Model

Several measures can be used to determine the size of the goodness of a model, including AIC and R^2. AIC is a method that can be used to select the best regression model found by Akaike. The AIC method is based on the Maximum Likelihood Estimation (MLE) method [14]. To calculate the AIC value, the following formula was used:

$$\text{AIC} = -2\ln L\left(\Omega\right) + 2k \tag{12}$$

where k is the number of parameters estimated in the model. Selection of the best model seen from the smallest AIC value.

Then, the coefficient of one way to measure the coefficient of determination R^2 is to approach one minus the ratio of two deviances. The numerator is the deviance of the selected model (Full Model), and the denominator is the deviance of the model that only contains the intercept (without predictor variables / Null Model) [15]. The value of R^2 can be calculated by:

$$R_D^2 = 1 - \left(\frac{D(y;\hat{\boldsymbol{\mu}})}{D(y;\bar{\boldsymbol{\mu}})}\right) = 1 - \left(\frac{\sum_{i=1}^{n}\left[-\log\left(y_i/\hat{\mu}_i\right) + \left((y_i - \hat{\mu}_i)/\hat{\mu}_i\right)\right]}{\sum_{i=1}^{n}\left[-\log\left(y_i/\bar{\mu}\right) + \left((y_i - \bar{\mu})/\bar{\mu}\right)\right]}\right) \tag{13}$$

While the value of R^2-Adjusted can be calculated by:

$$R_{D\,Adj}^2 = 1 - \left(\frac{(n-1)}{(n-k-1)}\frac{\sum_{i=1}^{n}\left[-\log\left(y_i/\hat{\mu}_i\right) + \left((y_i - \hat{\mu}_i)/\hat{\mu}_i\right)\right]}{\sum_{i=1}^{n}\left[-\log\left(y_i/\bar{\mu}\right) + \left((y_i - \bar{\mu})/\bar{\mu}\right)\right]}\right) \tag{14}$$

2.4 Data analysis steps

The steps of data analysis in this study were as follows:

1. Check the data distribution, especially the distribution of the response variable, namely the Central Java HDI data. Several distribution similarity test methods were used to estimate the appropriate distribution.
2. Check the correlation between predictor variables
3. Parameter estimation and hypothesis testing of 3-parameter gamma regression model
4. Model interpretation
5. Drawing conclusions

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics of Research Variables

Table 2 presents descriptive statistics for response and predictor variables, while Figure 1 presents histograms for Central Java HDI variables.

Table 2. Descriptive Statistics of Research Variables

Variable	Min	Median	Max	Average
$\boldsymbol{Y}$	65.67	71.35	83.19	72.23
$\boldsymbol{X_1}$	70.84	78.82	91.77	79.10
$\boldsymbol{X_2}$	7.31	15.12	37.18	15.57
$\boldsymbol{X_3}$	131.00	305.50	465.70	310.30
$\boldsymbol{X_4}$	479.00	1114.00	11762.00	2053.00
$\boldsymbol{X_5}$	2.18	4.57	9.83	5.08
$\boldsymbol{X_6}$	58.73	69.05	76.60	69.20
$\boldsymbol{X_7}$	61.91	92.83	100.00	89.57
$\boldsymbol{X_8}$	65.57	83.08	95.83	82.91

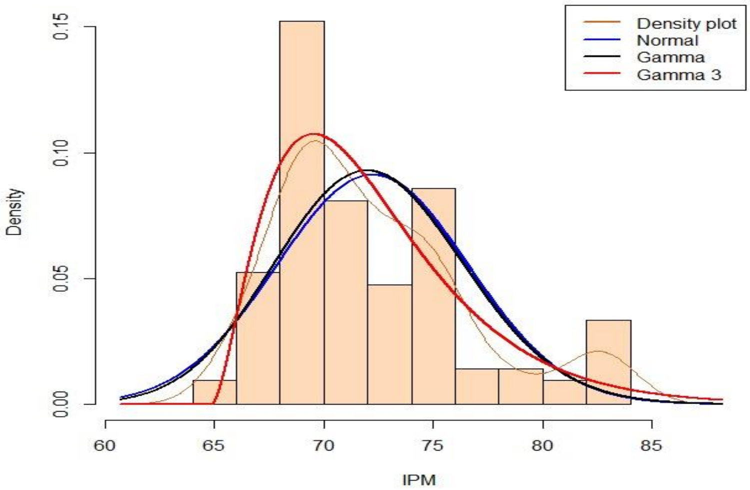

Figure 1. Histogram and plot of Central Java HDI variables

Based on Table 2, the average HDI in Central Java is 72.23, which is included in the high HDI category, while the maximum HDI value is 83.19. It is included in the very high HDI category, and the minimum HDI value is 65.67, which is still included in the medium category. Based on Figure 1, it can be seen that the Y response data is skewed to the right and tends to follow the 3-Parameter Gamma distribution. It can be seen that visually, the 3-Parameter Gamma Distribution is the most suitable compared to the Normal Distribution and Gamma Distribution (2 Parameters).

3.2. Response Variable Distribution Test (Y)

Testing the distribution of the HDI response variable was conducted to determine the distribution of the HDI variable. This test was done through a normality test and tested the distribution of Gamma 2-Parameter and Gamma 3-Parameter. This test was carried out using the Anderson Darling (AD) test. This test modifies the Kolmogorov-Smirnov (KS) test, which tests whether the sample comes from a population with a particular distribution. The critical value in the KS test does not depend on the particular distribution being tested. At the same time, AD utilizes a specific distribution in calculating the critical point [5]. The hypothesis used in this test is as follows:

$H_0: F_Y = F_Y^0$ (Distribution of data according to a certain distribution)

$H_0: F_Y \neq F_Y^0$ (The distribution of data does not match a certain distribution)

The test results are shown in Table 3 and Table 4. Based on Table 3, the HDI response variable does not follow a normal distribution because the p-value obtained is less than = 0.05. Based on Table 4, it failed to reject H_0 in the 2-Parameter Gamma distribution and 3-Parameter Gamma distribution indicated by the p-value greater than = 0.05. However, the p-value in the 3-Parameter Gamma distribution is much more significant than the p-value in the 2-Parameter Gamma distribution. Meanwhile, the HDI response variable does not follow a normal distribution because the p-value obtained is less than = 0.05. To find out the parameter values of the 3-Parameter Gamma distribution, the Maximum Likelihood Estimation (MLE) method can be used using the *FAdist* package in the R software. [16].

Table 3. Data Normality Test on Responses (Y)

Method	p-value	Conclusion
Shapiro-Wilk Test	0.00000	Not Normal
Jarque-Bera Test	0.00032	Not Normal
Kolmogrov-Smirnov Test	0.05173	Not Normal
Anderson-Darling Test	0.04299	Not Normal

Table 4. Goodness of Fit Test on Responses

Distribution	Test Statistics	P-Value
Normal	2.620	0.04299
Gamma 2-Parameter	2.284	0.06458
Gamma 3-Parameter	0.522	0.72410

3.3. Check the correlation between predictor variables

At this stage, identification of the presence or absence of cases of multicollinearity between predictors was carried out through the VIF value. If the VIF value was more than 10, it could be concluded that there were cases of multicollinearity. The complete VIF value is shown in Table 5 below.

Table 5. VIF of Predictor Variables

Predictor	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
VIF	1.1254	1.3117	2.0365	1.7750	1.4766	1.2165	1.6625	1.5697

Based on the VIF value in Table 5 above, it can be seen that all predictors have a VIF value of less than 10. This indicates that in the data, there are no cases of multicollinearity between predictor variables. Therefore, the predictor variables can be entered into the 3-Parameter Gamma regression analysis.

3.4. 3-Parameter Gamma Regression Model

This model was chosen because the distribution of the response variables closely matches the 3-Parameter Gamma distribution. This study used a 3-Parameter Gamma regression model based on the shape model. It means that the scale and location parameters were considered fixed. The computational process for the 3-parameter gamma regression model was not carried out using the *Gammareg Package* because the package is used only for the 2-parameter gamma regression model [17]. This study developed a separate syntax according to the algorithm used. The estimation results of the model parameters are presented in Table 6. Based on the obtained HDI model, simultaneous parameter testing produces a value of G^2 = 170.428 with a p-value = 0. This means that with a significance level of 0.05, it can be concluded that H0 is rejected. It means that at least one variable predictor affects the response variable in the model. Then a partial test is needed to determine the variables that significantly affect the model using the Z test. In Table 6, based on the results of the Z test, a very small P-value is obtained from each parameter (sig. at alpha 1%), except for the variable X_2 (sig. at 5% alpha), so it can be concluded that all variables have a significant effect on the model. The HDI model with 3-Parameter Gamma regression gives an R^2 value of 61.58%. The predictor variables studied can explain response variability of 61.58%, while the remaining 38.42% is explained by other variables that have not been included in the model. The full size of the goodness of the model is presented in Table 7.

Table 6. 3-Parameter Gamma Regression Model Parameter Estimation

Parameter	Estimate	Std.Error	Z Value	Pr(>|Z|)	Odd
β_0	3.80278	0.01140	333.5192	0.00000	44.82544
β_1	0.00195	0.00008	25.06538	0.00000	1.00195
β_2	-0.00022	0.00009	-2.31532	0.02060	0.99978
β_3	0.00010	0.00001	13.1752	0.00000	1.00010
β_4	0.00002	0.00000	104.7906	0.00000	1.00002
β_5	-0.00616	0.00022	-28.26873	0.00000	0.99386
β_6	0.00073	0.00013	5.71554	0.00000	1.00073
β_7	0.00070	0.00005	13.62518	0.00000	1.00070
β_8	0.00218	0.00008	28.97753	0.00000	1.00218

Table 7. 3-Parameter Gamma Regression Model Estimation Test Statistics

Statistics	Output
Null Deviance	655.4406 on 104 degrees of freedom
Residual Deviance	485.0126 on 96 degrees of freedom
G^2	170.428
P-value	0
AIC	501.0126
MSE	7.548017
R^2	0.6158487
Adjusted-R^2	0.5838361

Based on Table 6, the 3-Parameter Gamma Regression model for Central Java Province HDI can be formed as follows:

$$\hat{y} = \exp(3.80278 + 0.00195X_1 - 0.00022X_2 + 0.0001X_3 + 0.00002X_4 - 0.00616X_5 + 0.00073X_6 + 0.0007X_7 + 0.00218X_8) \quad (15)$$

Based on the model (15), it can be interpreted that if the APM of SMP/MTs (X_1) increases by 1%, then the average HDI (Y) increases by exp(0.00195) or 1.0019545 times, with the other independent variables considered constant. If the student-teacher ratio of SMP/MTs (X_2) increases by 1, the average HDI (Y) will decrease by exp(-0.00022) or 0.9997829 times, with the other independent variables being held constant. If the ratio of SMP/MTs students (X_3) increases by 1, then the average HDI (Y) increases by exp(0.0001) or 1.0000969 times, with the other independent variables being held constant. If the population density (X_4) increases by 1 then the average HDI (Y) will increase by exp(0.00002) times or 1.0000173 times, with the other independent variables being held constant. If the TPT (X_5) increases by 1%, the average HDI (Y) will decrease by exp(-0.00616) or 0.9938631 times , with the other independent variables considered constant. If the LFPR (X_6) increases by 1%, the average HDI (Y) will increase by exp(0.00073) or 1.0007296 times with the other independent variables considered constant. If the percentage of households (RT) with access to safe drinking water (X_7) increases by 1%, the average HDI (Y) increases by exp(0.00070) or 1.0006978 times, with the other independent variables being held constant. Also, if the percentage of households (RT) that have their own toilet facilities (X_8) increases by 1%, then the average HDI (Y) increases by exp (0.00218) or 1.0021826 times, with the other independent variables considered constant.

Based on the results of the estimation of model parameters and interpretations that have been made, it can be concluded that to increase the HDI can be done by increasing the APM value of SMP/MTs (X_1), the ratio of SMP/MTs students (X_3), population density (X_4), TPAK (X_6), the percentage of households (RT) with access to safe drinking water (X_7), and the percentage of households (RT) that have their own toilet facilities (X_8) and reduce the student-teacher ratio of SMP/MTs (X_2) and TPT (X_5).

4. CONCLUSION

The 3-Parameter Gamma regression model is proposed to obtain the factors that affect the HDI of Central Java Province. The response variable (HDI) follows a 3-Parameter Gamma distribution (without having to under the assumption of a normal distribution). The 3-Parameter Gamma Distribution application will be more suitable if applied to data with a particular threshold value or a minimum value that is not equal to zero. Eight variables have a significant effect on HDI, including the APM for SMP/MTs, student-teacher ratio for SMP/MTs, the ratio of students for SMP/MTs, population density, TPT, TPAK, Percentage of Households (RT) with access to safe drinking water, and Percentage of Households (RT) that have toilet facilities. As a suggestion, for further research, it is necessary to include a spatial element because it is suspected that there is a spatial effect in the HDI model.

REFERENCES

[1] BPS, *Statistik Indonesia dalam Infografis 2020 [Indonesia Statistics in Infographics 2020]*. Jakarta: Badan Pusat Statistik, 2020.
[2] A. Melliana and I. Zain, "Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten / Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel," ["Statistical Analysis of Factors Affecting the Human Development Index in Regencies / Cities of East Java Province by Using Panel Regression,"] vol. 2, no. 2, pp. 237–242, 2013.
[3] N. P. D. Yanthi and I. N. Budiantara, "Pemodelan Faktor - Faktor yang Mempengaruhi Indeks Pembangunan Manusia Menggunakan Regresi Nonparametrik Spline di Jawa Tengah," ["Modeling Factors Affecting Human Development Index Using Nonparametric Spline Regression in Central Java,"] *J. Stat. dan Seni ITS*, vol. 5, no. 2, pp. 157–162, 2016.
[4] R. Zakaria, "Pengaruh Tingkat Jumlah Penduduk, Pengangguran, Kemiskinan, Pertumbuhan Ekonomi, dan Belanja Modal Terhadap Indeks Pembangunan Manusia di Provinsi Jawa Tengah Tahun 2010-2016." ["The Influence of Population Level, Unemployment, Poverty, Economic Growth, and Capital Expenditure on the Human Development Index in Central Java Province in 2010-2016."] 2018.
[5] D. E. Putri, Purhadi, and D. D. Prastyo, "Parameter estimation and hypothesis testing on geographically weighted gamma regression," *J. Phys. Conf. Ser.*, vol. 893, p. 12025, 2017.
[6] A. S. Nasution, Purhadi, and Sutikno, "Estimasi Parameter Dan Pengujian Hipotesis Pada Model Regresi Gamma (Studi Kasus: Pemodelan Pencemaran Sungai Di Surabaya)," ["Parameter Estimation and Hypothesis Testing in the Gamma Regression Model (Case Study: River Pollution Modeling in Surabaya)," *Paid. J. Pendidik.*, vol. 2, no. 2, pp. 17–26, 2017.
[7] R. Shanker and K. K. Shukla, "On modeling of lifetime data using three-parameter generalized lindley and generalized gamma distributions," vol. 4, no. 7, pp. 283–288, 2016.
[8] E. W. Stacy, "A Generalization of the Gamma Distribution," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1187–1192, 1962.
[9] A. M. Mathai and P. G. Moschopoulos, "A Form of Multivariate Gamma Distribution," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 97–106, 1992.
[10] A. Rahayu, Purhadi, Sutikno, and D. D. Prastyo, "Multivariate gamma regression: Parameter estimation, hypothesis testing, and its application," *Symmetry (Basel).*, vol. 12, no. 5, 2020.
[11] G. H. Wenur, Purhadi, and A. Suharsono, "Three-parameter bivariate gamma regression model for analyzing under-five mortality rate and maternal mortality rate," *J. Phys. Conf. Ser.*, vol. 1538, no. 1, 2020.
[12] Purhadi, A. Rahayu, and G. H. Wenur, "Geographically weighted three-parameters bivariate gamma regression and its application," *Symmetry (Basel).*, vol. 13, no. 2, pp. 1–17, 2021.
[13] A. C. Cameron and P. K. Trivedi, *Microeconomics Methods and Applications*. Cambridge: Cambridge University Press, 2005.
[14] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.
[15] M. Mittlböck and H. Heinzl, "MEASURES OF EXPLAINED VARIATION IN GAMMA REGRESSION MODELS," *Commun. Stat. - Simul. Comput.*, vol. 31, no. 1, pp. 61–73, Mar. 2002.
[16] F. Aucoin, "Package ' FAdist .'" 2020.
[17] M. Corrales and E. Cepeda-Cuervo, "Package ' Gammareg .'" 2020.