

Discriminatory Analysis For The Classification of Urban Poverty Levels by Province in Indonesia 2023

Muhammad Zulfadhli¹, Nur Kholifah², Achmad Bintang Herida Kasuma³, Bokhoree Lamdoh⁴,
Chakrit Wisetmuen⁵

¹Departement of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

²Department of Statistics, Muğla Sıtkı Koçman University, Muğla 48000, Türkiye

³Department of Business Administration, Muğla Sıtkı Koçman University, Muğla 48000, Türkiye

⁴Department of Elementary Education, Albukhary International University, Kedah 05200, Malaysia

⁵Department of Public Administration, Bangkokthonburi University, Bangkok 10170, Thailand

Article Info

Article history:

Received august 25, 2026

Revised september 15, 2026

Accepted september 31, 2026

Keywords:

Classification

Discriminant Analysis

Poverty

ABSTRACT

Poverty remains a significant challenge in Indonesia, particularly in urban areas that face substantial economic and social pressures. The poverty line, which reflects the minimum needs for a decent life, is used as a benchmark to determine poverty status. To measure poverty, three main indicators are commonly employed: the poverty rate (P_0), poverty depth (P_1), and poverty severity (P_2). These indicators are interrelated and provide insight into community welfare. To better understand the factors influencing urban poverty across provinces in Indonesia, a discriminant analysis was conducted using the poverty depth index, poverty severity index, and Gini ratio as the primary variables. The analysis aimed to classify urban poverty levels and identify the most significant factors affecting them. The results indicate that the data on urban poverty rates by province in 2023 met the assumptions of multivariate normal distribution, homogeneity, and independence. The analysis revealed significant differences in the poverty depth and severity indices among the groups, while the Gini ratio did not show substantial differences. A single discriminant function was sufficient to separate the groups, with a canonical correlation of 0.729, explaining 67.89% of the data variability. Among the indicators, the Poverty Depth Index emerged as the most significant variable, while the Gini Ratio contributed the least. The centroid values for the low (-0.429) and high (2.486) poverty groups demonstrate the discriminant function's effectiveness in distinguishing between groups. The model achieved an overall classification accuracy of 91.2%, highlighting its strong performance in categorizing provinces based on urban poverty status.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Muhammad Zulfadhli
Departement of Statistics
Institut Teknologi Sepuluh Nopember
Surabaya 60111, Indonesia
Email: muhammadzulfadhli23@gmail.com

1. INTRODUCTION

Poverty is an economic condition in which a person is unable to meet basic needs and other needs based on his or her level of spending. A person is considered poor if his or her average monthly expenditure is below the poverty line. The Poverty Line describes the minimum amount of expenditure in rupiah that a person needs to meet his basic living needs for one month, including food and non-food needs [1].

Poverty measurement includes three main indicators, namely the poverty rate (P_0), the depth of poverty (P_1), and the severity of poverty (P_2). The reduction in the poverty rate (P_0) is not always followed

by a decrease in the depth of poverty (P1) and the severity of poverty (P2). Therefore, it is important to monitor changes in P1 and P2 over time to analyze whether the reduction in poverty really reflects an improvement in the welfare of the poor [2].

One of the important aspects in supporting poverty alleviation strategies is the availability of accurate poverty data. With this data, the government can make the right decisions regarding poverty alleviation efforts. In addition, this data also allows the government to compare poverty rates from time to time. Therefore, it is important for the government, both at the regional and central levels, to understand the classification of poverty levels in Indonesia. For this reason, a method that is able to classify poverty levels in various provinces is needed. One of the methods that can be used for this purpose is discriminatory analysis [3].

Discriminant analysis is a multivariate statistical technique used to separate or classify a group of objects into predetermined categories [4]. For the purpose of object recognition (observation), this method seeks to find a discriminant function whose numerical value can distinguish objects based on known characteristics. Meanwhile, for classification purposes, this method groups objects (observations) into two or more classes [5].

Based on this background, this study aims to classify the poverty rate in urban areas based on provinces in Indonesia using discriminatory analysis. It is hoped that the results of this analysis can be a reference for the government in formulating effective policies to reduce poverty levels in each region.

2. METHOD

The data used to conduct the discriminatory analysis is the data on the classification of poverty levels in urban areas by province in Indonesia in 2023 and it includes secondary data obtained through the BPS website. The poverty level classification used is the percentage of poverty, depth, severity, Gini ratio, and poverty category. The data was accessed on Monday, November 25, 2024 in the city of Surabaya, East Java.

The poverty level classification variables were analyzed using discriminant analysis. The analysis steps in this practicum are as follows: (1) Analyzing the characteristics of the data, (2) Conducting a multivariate normal distribution assumption test, a homogeneity assumption test, and a *Bartlett test*, (3) Conducting a discriminant analysis which includes the average difference test, *Eigen Value*, *Wilks Lambda*, discriminant and model functions, and classification accuracy, (4) Interpreting the results of the analysis, and (5) Drawing conclusions and suggestions.

2.1. Data Source

The data used to conduct the discriminatory analysis is the data on the classification of poverty levels in urban areas by province in Indonesia in 2023 and it includes secondary data obtained through the BPS website. The poverty level classification used is the percentage of poverty, depth, severity, Gini ratio, and poverty category. The data was accessed on Monday, November 25, 2024 in the city of Surabaya, East Java.

2.2. Research Variables

The research variables used in this study are shown in Table 1.

Table 1. Research Variables

Variable	Remarks	Data Scale	Categories
Y	Poverty Status	Ordinal	1 = Low 2 = Medium 3 = High
X1	Poverty Depth Index	Ratio	
X2	Poverty Severity Index	Ratio	
X3	Gini Ratio	Ratio	

2.3. Data Structure

The data structure utilized in this multivariate analysis involves n observations across the selected predictors and classification groups, structured as presented in Table 2.

Table 2. Data Structure

Observation (i)	Grouping Variable (Y)	Predictor Variables		
1	Y_1	X_{11}	X_{21}	X_{31}
2	Y_2	X_{12}	X_{22}	X_{32}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	Y_n	X_{1n}	X_{2n}	X_{3n}

2.4. Assumption Testing

Test the assumption of the multivariate normal distribution of poverty level classification in Urban by Province in Indonesia in 2023 using *the distance based* on the Pearson correlation value to find out whether the data is normally distributed multivariate or not which will be explained as follows.

Hipotesis :

H_0 : The classification of urban poverty levels by province in Indonesia in 2023 is normally multivariate.

H_1 : The classification of poverty levels in Urban Areas by Province in Indonesia in 2023 is not a normal multivariate distribution.

2.5. Descriptive Statistics

Descriptive statistics discusses the way data is collected, processing the observed numbers obtained, summarizing and presenting, describing and analyzing all data without drawing conclusions. The presentation of data in descriptive statistics is usually by making tabulations of the presentation in the form of graphs, diagrams, or by presenting the characteristics of the size of the concentration and the size of the spread. This is done to obtain information that is more interesting, useful and easier to understand. So descriptive statistics are statistics that discuss the collection, processing, presentation, and calculation of values from a data depicted in a table or diagram and does not concern the puller [6].

2.5.1 Average

How to calculate average, namely by adding up all data frequencies and dividing by the amount of data. *Red* This is very close to quantitative data, usually the value of this quantitative data with the symbol that is $x_1, x_2 \dots$ to x_n (if the data contains n values). For the symbol n , it is usually used to express a size in a sample, namely the amount of data, which will be studied, which is then used as a sample. N is also used to express the size of the population, which is the number of members as a whole in the data. Average or *red* is the ratio of the total value of observations to the number of observations [7].

The single data is as follows.

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} \quad (1)$$

Description:

x_i = The value of the variable to $-i$

i = 1,2,3, ... n

n = The amount of data of the value x

The group data is as follows [7].

$$\bar{x} = \sum_{i=1}^n \frac{x_i f_i}{n} \quad (2)$$

Description:

x_i = Midpoint of the i -value

f_i = Frequency of class i

n = The amount of data of the value x

2.5.2 Median

Median is one of the measures of data traceability, i.e. if a group of data is sorted from smallest to largest or largest to smallest, the value of the observation is right in the middle if the amount of data is odd, or the average of the two observations in the middle if the number of observations is even [8].

Formula *median* For a single data with n odd numbers is as follows.

$$Me = \frac{1}{2} x_{n+1} \quad (3)$$

Formula *median* For a single data with n even numbers is as follows.

$$Me = \frac{1}{2} x_{\frac{n}{2}} + x_{\frac{n}{2}} \quad (4)$$

Description:

Me = Median

x = Data value

n = Amount of data

Formula *median* for the group data is as follows.

$$Me = Tb \left[\frac{\frac{n}{2} - F}{f} \right] p \quad (5)$$

Description:

Tb = Bottom edge of the class *median*
 F = Cumulative frequency before class *median*
 f = Class frequency *median*
 n = Lots of data
 p = Class length

2.5.3 Variance

Variance is the sum of the squares of all the deviations of individual values from the mean of the group. *Variance* can describe how a quantitative data is scattered [9]. *Variance* (S_2) a single data is expressed as follows.

$$S_2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad (6)$$

The formula has an explanation, namely i is the first observation and x_1 is the first observation, then x_2 is the second observation and so on. $\bar{x}$ is the average and n is the number of observations. As for the number of observations. $\bar{x}$ *variance* (S_2) a group data is stated as follows [9].

$$S_2 = \frac{\sum_{i=1}^n f_1 (x_1 - \bar{x})^2}{\sum_{i=1}^n f_1} \quad (7)$$

Description:

f_1 = Frequency of class i
 x_1 = Middle grade of the i class
 i = 1,2,3,... n
 $\bar{x}$ = Average
 n = Amount of data

2.5.4 Standard Deviation

Standard deviation is the sum of the variation or spread in a range of values. *Standard deviation* low indicates that the value tends to be close to average. From this data, *Standard Deviation* indicates that the value of the data is getting farther or farther away from the average [10]. The formula used to calculate the value *Standard Deviation* shown as follows.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (8)$$

Description:

s = *Standard Deviation*
 $\bar{x}$ = average
 x_i = value to- x_i
 i = 1,2,3,... n
 n = a lot of data

2.6 Assumptions of Discriminative Analysis

Discriminant analysis has assumptions, the mean vector between different groups, the observation of multivariate normal distribution, and the covariance matrix between the same groups. The following is an explanation related to the assumptions that must be met in the discrimination analysis, namely the assumption of the multivariate normal distribution, the independent correlation test, and the homogeneity of the variance and covariance matrix [11].

2.6.1 Multivariate Normal Distribution Test

Normality testing or commonly called normal distribution is used as a condition or assumption of testing on multivariate using independent testing *T-test*, Squirrel *paired T-test* [12] and various other tests, both multivariate and univariate analysis. The test of the assumption assumption of the normal distribution of the multivariate is carried out using the distance of mahalanobis, but in this practicum we use the statistics of the T-proportion test with the following steps.

Hypothesis:

H0: Multivariate Normal Distributed Dependent Variable

H1: Multivariate Normal Undistributed Dependent Variables

Significant level: 5% or $\alpha = 0.05$

Rejection area: Minus H_0 , if the $T_{\text{proportion}}$ is outside the $45\% \leq T_{\text{proportion}} \leq 55\%$.

Test Statistics [12]:

$$d_{j^2} = (x_j - \bar{x})S^{-1}(x_j - \bar{x}) \quad (9)$$

Description :

d_j^2 = Distance of the j-point

x_j = Observation of the j-d variable

S^{-1} = Covariance matrix

$\bar{x}$ = Average data

2.6.2 Independence Correlation Test (Bartlett Test)

Test *Bartlett* aims to find out if there is a relationship between variables in multivariate cases. If the variable $x_1, x_2, \dots, x_p$ are mutually free, so the correlation matrix between variables is the same as the identity matrix. To test the freedom between these variables, the *Bartlett* states the hypothesis as follows.

Hypothesis:

$H_0 : \rho = I$ (Correlation matrix is identical to identity matrix or between independent variables)

$H_1 : \rho \neq I$ (Correlation matrix is not identical to identity matrix or between dependent variables)

Significant Grade : α as per 0.05

Rejection District: H_0 is rejected if or P-value $< X^2_{hitung} > X^2_{\alpha, df}$

Test Stats [13]:

$$\chi^2 = (ln10)\{B - \sum_{i=1}^n (n-1)log s_i^2\} \quad (10)$$

$$B = (log s^2) \sum_{i=1}^n (n_i - 1) \quad (11)$$

$$S = \frac{\sum_{i=1}^n (n_i - 1)s^2}{\sum_{i=1}^n (n-1)} \quad (12)$$

Description:

n = amount of data

s_i^2 = variance of data to- i

s^2 = Data dissemination

B = Statistical Parameters

$ln10$ = Constant

$Log s_i^2$ = Logarithm Variance

i = 1,2,3,... n

2.6.3 Variance Homogeneity Test (Box's M Test)

Homogeneity test *variance* is a test used to find out whether the variance of several populations is the same or not. Two-way similarity test *variance* is used to test whether the distribution of the data is homogeneous or not, i.e. by comparing the two *variance* data. If two or more data groups have *variance* The same magnitude, the homogeneity test does not need to be carried out anymore because the data is already considered homogeneous. Homogeneity tests can be performed if the data groups are in normal distribution.

The homogeneity test is performed to show that the differences that occur in parametric statistical tests are actually due to differences between groups, not as a result of differences in groups [14].

Hypothesis:

H_0 : Homogeneous *covariance* matrix

H_1 : *Heterogeneous* covariance matrix

Significant Status : α

Rejection District: H_0 is subtracted if H_0 if $F > F_{1-\alpha(dbr, dbg)}$ or p-value $< \alpha$

Test Stats:

$$MC^{-1} = (n_k - 1)ln|s| - \sum_{k=1}^E (n_k - 1)ln|s_k| C^{-1} \quad (13)$$

Where,

$$C^{-1} = 1 - \frac{(2p^2 + 3p - 1)(k + 1)}{6(p + 1)kn} \quad (14)$$

Description:

n_k = Sample size k

s = Matrix *Kovarians variance* Combined

s_k = Matrix *Covarians* to the treatment of k

k = Number of groups

C^{-1} = A trans matrix

2.7 Discriminant Analysis

Discriminant analysis is a multivariate analysis applied to model the relationship between one categorical response variable and one or more quantitative predictor variables. This analysis will form a discriminant function to separate a new set of objects into a new object in a group, so that this analysis can be used as a classification method. Discriminant analysis is applied to model the relationship between one categorical response variable (nonmetric, nominal or ordinal, qualitative) and one or more quantitative predictor variables. The discriminant analysis model is shown in the following equation [15].

$$Y_K = b_{k1X1} + b_{k2X2} + \dots + b_{kpXp} \quad (15)$$

Remarks:

- Y_K = Discriminative value of k
 p = Number of independent variables
 β_k = Discriminating coefficient
 X = Independent (free) predictor or variable

2.7.1 Average Vector Similarity Test

The average vector similarity test was carried out to test the differences between groups formed from each independent variable [11].

Hypothesis:

H0: $m1 = m2 = \dots = mk$

H1: There are at least 2 different groups.

Rejection area: minus H0 if $F_{hit} > F_{(\alpha; df1; df2)}$ or $P\text{-Value} < \alpha$

Test statistics:

$$F_{hit} = \left(\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} \right) \times t^2(\alpha) \quad (16)$$

Where:

$$t^2(\alpha) = \frac{n_1 n_2}{n_1 + n_2} \times (\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2) \quad (17)$$

Description:

- $N1$ = number of observations from group 1
 $N2$ = number of observations from group 2
 p = number of independent variables
 S^{-1} = inverse matrix of the combined variety
 $\bar{X}_1$ = group average 1
 $\bar{X}_2$ = average group 2

2.7.2 Eigenvalue

Eigenvalue is a measure of the value of *variance* a variable so that it can be constructed into a factor. The process to find the eigenvalue is to form a matrix *variance-Covarians*, then define the matrix *Varians-Covarians* with the identity matrix, i.e. the formula . It further forms the characteristic equation of the matrix with the formula. $Ax = \lambda Ix \det(A - \lambda I) = 0$

2.7.3 Wilk's Lambda Test

Wilks' Lambda is the ratio between the number of squares in a group and the total number of squares. Wilks' Lambda values range from 0 to 1. Values close to 1 indicate that the average of the groups is likely to be no different [17].

Hypothesis:

H0 : There is no significant difference between independent variables and dependent variable categories

H1 : There is a significant difference between independent variables and dependent variable categories

The critical region is minus H0 if χ^2 calculates $> \chi^2(\alpha; df)$ or $P\text{-Value} < \alpha$.

Test Statistics:

$$\Lambda = \frac{SS_w}{SS_k} \quad (18)$$

Where:

$$\frac{SS_w}{SS_k} = \frac{\sum_{i=1}^g \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{\sum_{i=1}^g \sum_{j=1}^n (x_{ij} - \bar{x})^2} \quad (19)$$

Description:

- Λ = Wilk's Lambda

SS_w = Sum of square within group
 SS_k = Sum of square total

2.7.4 Discriminating Function

The discriminant function is used to predict the group of a new object that is observed. This discriminant function is obtained from the unstandardized coefficient. This shows the partial contribution of each variable to the dyschemy function. The following is an equation of the discriminant function in general [17].

$$Y_k = bk_1X_1 + bk_2X_2 + \dots + bk_pX_p \quad (20)$$

$$Y_k = b'kX \quad (21)$$

In this case, Y_k is the discriminant value of k ; $k = 1, 2, \dots, s$; $s \leq \min(m-1, p)$; m is the number of observations; p is the sum of independent variables; b is the discriminant coefficient; and X is the independent variable.

The value of the discriminant function coefficient is obtained using the following equation.

$$b_k = \sqrt{W_{ii}a_m} \quad (22)$$

$$a = (W^{-1}B - \lambda I) = 0 \quad (23)$$

In this case, b_k is the standard discriminant coefficient of the i th variable in the discriminant function of m ; w_{ii} is the main diagonal of the covariance matrix in the group; a_{mi} is the non-standard discriminant coefficient of the i in the discriminant function of m ; B is the matrix of covariance between groups; W is the matrix of covariance in a group; λ is the eigenvalue; and I is the identity matrix.

2.8 Poverty Depth Index

The Poverty Gap Index or poverty depth index is defined as a measure of the gap in the average expenditure of each population that is classified as poor against the poverty line. This index can function as a comparison between regions, where the higher the value of an area, the more severe the poverty depth index [18].

2.9 Poverty Severity Index

The Poverty Severity Index (P2) is an index that provides information about the distribution of expenditure among the poor. Provides complementary information on the incidence of poverty. For example, there may be cases where some poor groups have a high incidence of poverty but a low poverty gap, while other groups have a low incidence of poverty but have a high poverty gap for the poor. The higher the index value, the higher the inequality of expenditure among the poor [19].

2.10 Gini Ratio

Gini Ratio is a measure of aggregate inequality that was first developed by an Italian statistician named Corrado Gini and published in 1912. One measure of inequality that is also often used is the Gini ratio with a scale of 0- 1. The higher the value *gini ratio* shows an increasing inequality. If the Gini ratio in an area increases, it means that the income inequality gap of the people in that area is widening [20][21][22].

2.11 Previous Research

Research related to modeling socio-economic determinants, health status, and regional development in Indonesia has demonstrated the importance of multivariable and nonparametric statistical approaches in capturing the complexity of regional data, such as unmet need status [23],[24], nonparametric binary logistic regression inference frameworks for the status of disadvantaged areas [25], and some studies using classification models such as [26] as well as modeling of open unemployment status and public health [27],[28],[29].

3. RESULTS AND DISCUSSION

3.1. Characteristics of Regency/City Grouping Data in Central Java in 2022 Based on Factors Affecting the Human Development Index

The characteristics of the data through the boxplot of factors that cause the Poverty Rate in Urban Areas are displayed in the following descriptive statistical table.

Table 3. Data Characteristics

Poverty Status		Human Development Index Factors	
		Mean	Standard Deviation
Low	Poverty Depth Index	0,0098	0,00425
	Poverty Severity Index	0,0024	0,00138
	Gini Ratio	0,3513	0,04973
Medium	Poverty Depth Index	0,0190	0,00323
	Poverty Severity Index	0,0046	0,00102
	Gini Ratio	0,3932	0,03743
Height	Poverty Depth Index	0,112	0,00525
	Poverty Severity Index	0,0027	0,00154
	Gini Ratio	0,3574	0,04995

Based on Table 1. shows that in areas with low poverty status, the average poverty depth index is 0.0098 with a standard deviation of 0.00425, indicating a small level of poverty depth and low variation. The average poverty severity index is 0.0024 with a standard deviation of 0.00138, describing a very low poverty severity and a small spread. Meanwhile, the average Gini ratio is 0.3513 with a standard deviation of 0.04973, indicating a moderate level of income distribution inequality with considerable variation.

In areas with moderate poverty status, the average poverty depth index increased to 0.0190 with a standard deviation of 0.00323, indicating a greater depth of poverty than in low-poverty areas, but still relatively small with a low variation. The average poverty severity index was 0.0046 with a standard deviation of 0.00102, slightly higher than in low-poverty areas, with a small degree of distribution. The mean Gini ratio increased to 0.3932 with a standard deviation of 0.03743, reflecting higher income inequality than low-poverty areas, but with more controlled variation.

In areas with high poverty status, the average poverty depth index reached 0.112 with a standard deviation of 0.00525, indicating a much greater increase in poverty depth than in moderate and low poverty areas, with a small variation. The mean poverty severity index was 0.0027 with a standard deviation of 0.00154, indicating that despite high poverty status, the severity of poverty was not much different from that of low and medium poverty areas, with a slightly larger spread. The average Gini ratio in this region was 0.3574 with a standard deviation of 0.04995, describing a moderate level of income distribution inequality, with slightly lower values than in moderate poverty areas, but with a larger spread.

Overall, the data showed that the higher the poverty status, the more the average poverty depth index tended to increase, but the poverty severity index did not show a significant increase. Meanwhile, the Gini ratio tended to fluctuate, with the highest values in moderate poverty areas. Variation between regions tended to be small for the depth and severity of poverty indexes, while the Gini ratio varied greater. This characteristic provides an idea of the dynamics of human development and the distribution of welfare in regions with different poverty levels.

3.2. Assumption Test

A significant level (α) is set of 0.05 and the area of rejection is minus H_0 if $p\text{-value} > 0.05$, a Pearson correlation value of 0.745 was obtained and reinforced by a $p\text{-value}$ of 0.000 which was smaller than 0.05 so that it was decided to fail to subtract H_0 which means that the classification of poverty levels in Urban Areas by Province in Indonesia in 2023 is normally distributed multivariately.

The Independence Test of the classification of poverty levels in Urban Areas by Province in Indonesia in 2023 using the Bartlett test to find out whether the data meets the dependent assumptions or not which will be explained as follows.

Hipotesis :

$H_0: \rho = I$ (Classification of poverty levels in Urban areas by Province in Indonesia in 2023 independent)

$H_1: \rho \neq I$ (Classification of poverty levels in Urban areas by Province in Indonesia in 2023 dependent)

A significant level (α) is set of 0.05 and the area of rejection is minus H_0 if $\chi^2_{hitung} > \chi^2_{tabel}$ p-value < 0.05 , obtained χ^2_{hitung} by 91,913 which is greater than χ^2_{tabel} by 7,815 and strengthened by a p-value of 0.000 which is smaller than 0.05 so that it is decided to reject H_0 which means that the classification of poverty levels in Urban Areas by Province in Indonesia in 2023 is dependent.

Multicollinearity detection is used to find out whether there is a correlation between the predictor variables that indicate a person is smoking or not explained as follows.

Table 4. Multicollinearity Detection

Variabel	Poverty Depth Index	Poverty Severity Index	Gini Ratio
Poverty Depth Index	1	0,962	0,134
Poverty Severity Index	0,962	1	0,020
Gini Ratio	0,134	0,020	1

The table above shows that there is a high correlation value between the poverty severity index variable and the poverty depth index because the correlation value is close to 1, but a low correlation is obtained between the poverty depth and severity index and *the Gini ratio* so that it does not indicate the existence of multicollinearity between the predicate variables.

The variance homogeneity test is used to meet the assumption of discriminant analysis, namely a homogeneous covariance matrix using the *Box's M* as follows.

Hipotesis :

H_0 : (matrix variance covariance homogeneous) $\Sigma_1 = \Sigma_2 = \dots = \Sigma_i$

H_1 : there is at least one (heterogeneous covariance matrix) $\Sigma_i \neq 0$

Table 5. Box's M Test Result

F	$F_{(0,05; 6; 279,026)}$	P-value
0,156	2,131	0,988

A significant level (α) is set of 0.05 and the area of rejection is minus H_0 if $F > F_{(\alpha, df1, df2)}$ or p-value < 0.05 , obtained F of 0.156 which is greater than 2.131 and strengthened by a p-value of 0.988 which is greater than 0.05 so that it is decided to fail to subtract H_0 which means that the classification of urban poverty levels by province in Indonesia in 2023 is homogeneous.

3.3. Discriminant Analysis

The results of the discriminatory analysis on the classification of poverty levels in Urban Areas by Province in Indonesia in 2023 are as follows. The difference in the average classification of poverty levels in Urban by Province in Indonesia in 2023 is explained as follows.

Table 6. Average Classification of Poverty Level

Poverty status	average		
	X1	X2	X3
Low	0,0098	0,0024	0,3513
moderate	0,019	0,0046	0,3932
height	0,011	0,0027	0,3574

The table shows that there is no difference in the mean depth index, severity index, and *Gini ratio* in poverty status reinforced by testing the following hypothesis.

Hipotesis :

H_0 : There is no difference in the average poverty depth index in poverty status.

H_1 : There is a difference in the average poverty depth index in poverty status.

H_0 : There is no difference between the average poverty severity index and poverty status.

H_1 : There is a difference in the average poverty severity index and poverty status.

H_0 : No difference in average *gini ratio* on poverty status.

H_1 : There is an average difference *gini ratio* at the level of poverty status.

A significant level (α) is set of 0.05 and the area of rejection is minus H_0 if $F > p\text{-}F_{(\alpha, df_1, df_2)} \text{value} < 0.05$.

Table 7. Signification Test Result

Variable	F count	F	P-value
Poverty Depth Index	21,2		0,000
Poverty Severity Index	11,426	4,149	0,002
<i>Gini Ratio</i>	3,205		0,083

Obtained F on the poverty depth index and poverty severity index of 21.2 and 11.426 respectively, which are greater than $F_{(0,05; 1; 32)}$ by 4,149 and p-values of 0.000 and 0.002 which are smaller than α 0.05 so that it was decided to reject H_0 which means that there is a difference in the average poverty depth index and poverty severity index in poverty status. Meanwhile, it is obtained F on *gini ratio* of 3,205 which is smaller than $F_{(0,05; 1; 32)}$ of 4.149 and a p-value of 0.083 which is greater than α 0.05 so that it was decided to fail to reject H_0 which means there is no difference in the average *gini ratio* on poverty status.

3.4. Discriminant Analysis Modeling

After all the assumptions of the discriminatory analysis have been met, the next is the modeling of the discriminatory analysis for the classification of poverty levels in Urban by Province in Indonesia in 2023 as follows.

Table 8. Modeling Discriminatory Analysis Results

Function	Self-esteem	% of Variance	Canonical Correlation
1	1,132	100	0,729

Based on the eigenvalue table, an eigenvalue of 1.132 indicates the magnitude of the variance explained by the discriminant function. This value indicates that the discriminant function has good discriminating power to separate groups. The first discriminant function explains 100% variance, which means that all relevant information in the model is contained in this function. The cumulative value also reaches 100%, confirming that one discriminant function is sufficient to model the relationship between the variables in this analysis. In addition, the canonical correlation value of 0.729 indicates a strong relationship between the discriminant function and the identified group. This means that with only one discriminant function, the model or the predictor variable is able to explain the response variable by 67.89%, while the remaining 32.11% is explained by other factors.

Table 9. Discriminatory Analysis Function

Test of Function	Wilks Lambda	Chi-Square	df	P-Value
1	0,469	23,095	3	0,001

Based on Wilks' Lambda table, Wilks' Lambda value of 0.469 indicates that the discriminant function is quite effective in explaining differences between groups. This value indicates that most of the variability in the data is successfully explained by the discriminant model. In addition, the significance value is very small of 0.001, which means that this result is significant at a 95% confidence level. Thus, it can be concluded that the discriminant function used in this analysis has a good ability to separate existing groups statistically and that the independent variables used make a significant contribution to differentiating groups.

Table 10. Significant Contribution

Variable	Function 1
Poverty Depth Index	3,144
Poverty Severity Index	-2,462
<i>Gini Ratio</i>	-0,075

Based on the *Standardized Canonical Discriminant Function Coefficients* table, the results obtained that the Poverty Depth Index variable has the largest coefficient, which is 3.144, which shows that this variable has the most significant influence in differentiating groups. On the other hand, the Poverty Severity Index variable has a coefficient of -2.462, which is also significant but with a negative direction, indicating that the lower value in this index is more related to one group than the other. Meanwhile, the Gini Ratio variable has the smallest coefficient, which is -0.075, so its contribution to the discriminant function is relatively smaller than the other two variables. Overall, these results show that the Poverty Depth Index is the most decisive variable in distinguishing groups, while the Gini Ratio plays a less significant role in separating groups.

Based on the table of *Canonical Discriminant Function Coefficients*, the discriminant functions formed are as follows.

$$\hat{D} = -2,933 + 759,668 \text{ Poverty Depth Index} - 1834,394 \text{ Poverty Severity Index} - 1,542 \text{ Gini Ratio}$$

This function shows that the Poverty Depth Index has a large positive coefficient, which is 759,668, which indicates that provinces with higher levels of depth of poverty tend to have greater discriminant function values, which are associated with higher rates of urban poverty. In contrast, the Poverty Severity Index has a large negative coefficient, which is -1834,394, indicating that an increase in the severity of poverty in a province is related to a lower value of the discriminant function, which may indicate the characteristics of provinces with lower levels of urban poverty. Meanwhile, the Gini Ratio, which has a small negative coefficient of -1,542, contributes less to differentiating provinces based on urban poverty levels.

Based on the *Functions at Group Centroids* table, the centroid value shows the average discriminant function value for each poverty status group. The group with low poverty status has a centroid value of -0.429, while the group with high poverty status has a centroid value of 2.486. The considerable difference between these two values shows that the discriminant function is able to effectively distinguish the two groups. This centroid value is also used to determine the group membership of a province based on the discriminant function value it obtains. If the discriminant function value of a province is close to -0.429, then the province tends to be classified into the group with low poverty status. Conversely, if the discriminant function value is close to 2.486, then the province tends to be classified into the group with high poverty status.

3.5. Model Evaluation and Classification Accuracy

Model evaluation and classification results table are as follows

Table 11. Model Evaluation and Classification Results

Poverty Percentage		Predicted Group Membership		Total
		Low	Height	
Count	Low	26	3	29
	Height	0	5	5
Presses	Low	89,7	10,3	100
	Height	0	100	100

Based on the *Classification Results* table, classification results were obtained showing that out of 29 provinces with low poverty status, 26 provinces were correctly classified, while 3 provinces were incorrectly classified into the high poverty status group, resulting in an accuracy rate of 89.7% for this group. For the high poverty status group, all 5 provinces were correctly classified, providing an accuracy rate of 100%. Overall, the accuracy of the classification in the original data was 91.2%, indicating that the discriminator function was able to distinguish poverty status groups with excellent success rates.

4. CONCLUSION

The conclusions obtained from the classification of poverty levels in Urban by Province in Indonesia in 2023 are seen from the characteristics of the data showing that the index severity index variable has the lowest average and standard deviation in all poverty statuses and the Gini ratio variable has the highest average and standard deviation in all poverty statuses. The results of the assumption testing analysis on the poverty level classification data in Urban by Province in Indonesia in 2023 have met the assumptions of multivariate normal distribution, homogeneity assumptions, and independence assumptions. The results of the discriminant analysis show that the classification of poverty levels in Urban by Province in Indonesia in 2023 has a classification accuracy of 91.2. Through this study, it is recommended that the government pay attention to the high income inequality (*Gini ratio*) in urban areas. In addition, social protection programs need to be strengthened to reduce community vulnerability, especially in areas with high poverty severity.

ACKNOWLEDGMENTS

Thank you to all parties involved in the compilation of this article.

FUNDING INFORMATION

There is no funding in this article.

AUTHOR CONTRIBUTIONS STATEMENT

Specify the contributions of each author and the role of using limited assistance from Artificial Intelligence (AI) as follows:

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muhammad Zulfadhli	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Nur Kholifah		✓				✓		✓	✓	✓	✓	✓		
Achmad Bintang H.K.	✓		✓	✓			✓		✓	✓	✓		✓	✓
Bokhoree Lamdoh		✓				✓		✓	✓	✓	✓	✓		
Chakrit Wisetmuen	✓		✓	✓			✓		✓	✓	✓		✓	✓

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

- [1] Afriyanti, et al., "Analisis Diskriminan pada Indikator IPM Kabupaten/Kota di Provinsi Sulawesi Selatan," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 2020.
- [2] S. Eskani, "Pengaruh Karbon Aktif dari Batok Kelapa untuk Menurunkan Zat Besi dan Mangan dalam Air Sumur," *Jurnal Ilmiah Teknik Lingkungan*, 2017.
- [3] Á. García del Castillo-López and M. P. Domínguez, "Employees' Emotional Intelligence and Job Satisfaction: The Mediating Role of Work Climate and Job Stress," *Administrative Sciences*, 2024.
- [4] M. S. Ginting, Hubungan Pilihan Mahasiswa Unja Dalam Pembayaran Uang Kuliah Tunggal Berdasarkan Jenis Bank Dengan Tabel Kontingensi Uji Statistik Chi-Square, Jambi: Universitas Jambi, 2023.
- [5] R. Hoppock, "Job satisfaction of psychologists," *J. Appl. Psychol.*, vol. 21, pp. 300-303, 1937.
- [6] T. Hidayati, *Statistika Dasar*, Purwokerto: CV. Pena Persada, 2019.
- [7] S. M. Nisak Ruwah Ibnatur Husnul, *Statistika Deskriptif*, Pamulang - Tangerang Selatan: Unpam Press, 2020.
- [8] R. Walpole and M. Myers, *Probability & Statistics for Engineers & Scientists*, London: Pearson Education, 2016.
- [9] M. Maswar, "Analisis Statistik Deskriptif Nilai UAS Ekonometrika Mahasiswa dengan Program SPSS 23 & Eviews 8.1," *Jurnal Pendidikan Islam Indonesia*, pp. 273–292, 2017.
- [10] Indwiarti, R. P. H. G. et al., *Pengantar Statistika Eksplorasi Data dengan MS.Excel dan R*, Yogyakarta: KBM Indonesia, 2021.
- [11] N. R. Asnidar, "Penerapan Analisis Diskriminan Kuadratik Robust Pada Klasifikasi Desa," *ESTIMASI: Journal of Statistics and Its Application*, vol. 5, no. 2, pp. 196–240, 2024.
- [12] D. N. Putri, *Rancangan Penelitian Bidang Teknologi Pangan*, Malang: UMMPress, 2020.
- [13] N. E. C. Qonitatin Nafisah, "Analisis Cluster Average Linkage Berdasarkan Faktor-Faktor Kemiskinan di Provinsi Jawa Timur," *Zeta – Math Journal*, pp. 31–36, 2017.
- [14] Usmadi, "PENGUJIAN PERSYARATAN ANALISIS (Uji Homogenitas dan Uji Normalitas)," *Inovasi Pendidikan*, vol. 7, no. 1, 2020.
- [15] Barnabas Anthony Philbert Marbun, "Penentuan Kategori Kelulusan Mahasiswa Menggunakan Metode Analisis Diskriminan," *Zeta-Math Journal*, vol. 9, no. 1, pp. 31–40, 2024.
- [16] S. Nia Shantian Lumbantobing, "Confirmatory Factor Analysis to Measure the Satisfaction Level of Air Soda Tourists in North Tapanuli," *Formosa Journal of Science and Technology (FJST)*, vol. 2, no. 2, pp. 475–492, 2023.
- [17] Rizka Widya Berliana et al., "Analisis Diskriminan Pada Indikator yang Memengaruhi Indeks Ketahanan Pangan Menurut Kabupaten/Kota di Indonesia Tahun 2022," *Seminar Nasional Official Statistics*, pp. 11–20, 2023.
- [18] M. T. Ashshiddiqi et al., "Strategi Program Indeks Pembangunan Manusia (IPM) di Desa Pakuaon Kecamatan Sukaresmi Kabupaten Cianjur," *Jurnal Inovasi Penelitian (JIP)*, vol. 2, no. 4, pp. 1153–1162, 2021.
- [19] S. Sitompul, "Peran Tenaga Kerja Formal dan Upah Minimum Dalam Mengatasi Kemiskinan Pulau Jawa Tahun 2015-2022," *e-journal UAJY*, 2023.
- [20] Grace Wulandari, "Pemodelan Persentase Penduduk Miskin Di Indonesia Menggunakan Regresi Probit dan Regresi Logistik," *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*, pp. 354–368, 2022.
- [21] A. U. Yogi Ramadhan, "Pendugaan Persentase Penduduk Miskin Ekstrem Menggunakan Small Area Estimation dengan Partitioning Around Medoids Clustering," *Politeknik Statistika STIS*, pp. 557–568, 2022.
- [22] T. A. N. A. Ayu Noviatamara, "Analisis Pengaruh Pertumbuhan Ekonomi dan Tingkat Pengangguran Terbuka di Daerah Istimewa Yogyakarta," *Jurnal REP (Riset Ekonomi Pembangunan)*, pp. 53–60, 2019.
- [23] M. Zulfadhli, I. N. Budiantara, V. Ratnasari, and A. S. Suriaslan, "Fourier series nonparametric regression estimator for modeling status of unmet need in East Java Province in 2023," *Commun. Math. Biol. Neurosci.*, vol. 2025, Art. ID 2025.

- [24] A. S. Suriaslan, I. N. Budiantara, V. Ratnasari, M. Zulfadhli, and R. C. Dhewy, "An Empirical Study on Unmet Need: A Statistical Inference Framework for Truncated Spline Nonparametric Binary Logistic Regression (TSNLBR)," *J. Cultural Anal. Social Change*, pp. 868–877, 2025.
- [25] M. Zulfadhli, I. N. Budiantara, V. Ratnasari, A. S. Suriaslan, and R. C. Dhewy, "A statistical inference framework for FSNBLR: Modeling underdeveloped regional status in Eastern Indonesia," *MethodsX*, p. 103746, 2025.
- [26] M. Zulfadhli, I. N. Budiantara, V. Ratnasari, and A. S. Suriaslan, "Semiparametric Regression Estimator of Fourier Series For Categorical Data," *IAENG Int. J. Appl. Math.*, vol. 55, no. 7, pp. 2157–2164, 2025.
- [27] B. W. Otok, M. Zulfadhli, R. D. Pangesti, M. I. Kurniawan, A. E. P. Haryanto, and I. Kurniawan, "Comparison of Binary Probit Regression and Fourier Series Nonparametric Logistic Regression in Modeling Diabetes Status at Hajj General Hospital Surabaya," *Barekeng*, vol. 20, no. 1, pp. 255–270, 2026.
- [28] A. M. Achmad, M. Zulfadhli, and N. Nurfaudzan, "Binary Logistic Regression on the Status of Unemployment Rate in West Java Province in 2023," *Likelihood: J. Stat. Its Appl.*, vol. 1, no. 1, pp. 30–34, 2025.
- [29] M. Zulfadhli, I. N. Budiantara, and V. Ratnasari, "Nonparametric regression estimator of multivariable Fourier series for categorical data," *MethodsX*, vol. 13, p. 102983, 2024.