

Factors that Affect Poverty Areas in North Sumatera Using Discriminant Analysis

D H Nasution*, P Bangun, H R Sitepu

Department of Mathematics, Faculty of Mathematics and Natural Sciences,
University of Sumatera Utara, Medan 20155, Indonesia

*dewiharninasution@gmail.com, pengarapen@usu.ac.id, henry1@usu.ac.id

Abstract. In Indonesia, especially North Sumatera, the problem of poverty is one of the fundamental problems that become the focus of government both central and local government. Although the poverty rate decreased but the fact is there are many people who are poor. Poverty happens covers several aspects such as education, health, demographics, and also structural and cultural. This research will discuss about several factors such as population density, Unemployment Rate, GDP per capita ADHK, ADHB GDP per capita, economic growth and life expectancy that affect poverty in Indonesia. To determine the factors that most influence and differentiate the level of poverty of the Regency/City North Sumatra used discriminant analysis method. Discriminant analysis is one multivariate analysis technique are used to classify the data into a group based on the dependent variable and independent variable. Using discriminant analysis, it is evident that the factor affecting poverty is Unemployment Rate.

1. Introduction

Poverty is a problem that concerns economic growth and its relation to income distribution [4]. The problem of poverty is one of the fundamental problems that it becomes central government attention both central government and local government. The size of the poor is strongly influenced by the lines poverty, because the poor are people who have an average per capita spending per month below the poverty line.

The phenomenon experienced in North Sumatera is the number and percentage the many poor people. Even official data of poverty shows poverty at a higher level. Countermeasures program poverty of local government is seen at the village level, but these programs often do not meet the expectations of government officials regions and citizens due to weak implementation and supervision mode.

Discriminant analysis is one of the methods used in a multivariate analysis with dependency method (where the relationship between variables can already be distinguished where the dependent variable and which variable free). Basically discriminant analysis can be used for know the characteristic variables that distinguish population groups that exists, but it can also be used as a grouping criterion which is first known clearly grouping. Discriminant analysis aims to recognize which factors can distinguish two or more groups. These distinguishing factors will form a distinctive function (called a discriminant function).



2. Research Method

In his research suggests that essentially discriminant analysis (discriminant analysis) can be used to know the characteristic variables that differentiate the existing population group, can also be used as grouping criteria [1]. Technically, discriminant analysis is similar to regression analysis, because both have the dependent variable and the independent variable in the model. It's just in simple regression analysis and multiple, variable bound in the form of data metrics and the independent variables can be either metric or nonmetric, while the discriminant analysis, the type of data for the dependent variable must be a category (using ordinal or nominal scales) and the independent variables are metric data (interval and ratio) [2].

2.1. Purpose of Discriminant Analysis

Since the multivariate form of discriminant analysis is bound, the dependent variable is the variable underlying the discriminant analysis, with the aim of:

- a. Want to know if there is a clear distinction between groups on the dependent variable, or it can be said whether there is a difference between member of group 1 with member of group 2.
- b. If there is a difference, which independent variables are on the discriminant function which makes such a difference.
- c. Create a function or discriminant model, which is basically similar with regression equation.
- d. Classification of objects (in SPSS terminology is called line), whether an object (can be a person's name, the name of the plant, objects or other) belonging to group 1 or group 2, or others.

2.2. Assumptions in Discriminant Analysis

In addition to the basic and discriminant goals, it turns out there are assumptions that must be fulfilled prior to discriminant analysis, ie distributed free variables normal multivariate (multivariates normal distribution) and variance in each group are equal (equal variances).

2.3. Discriminant Analysis Process

In general the basic process of discriminant analysis is:

- a. Splitting the variables into dependent variables and independent variables.
- b. Determine the method to create a discriminant function. Where on the principle there are two basic methods for it:
 - 1) Simultaneous Estimation, method by entering all variables together then process discriminant.
 - 2) Stepwise Estimation, method by entering one by one variable into the discriminant model. In this method, of course there are variables that remain in the model, and there are possibilities one or more independent variables are removed from model.
- c. Conducting testing the significance of the discriminant function that has been formed, using Wilks Lambda, Pilai, F test and others.
- d. Understand the discriminant function that has been formed.
- e. Conducting the classification accuracy of the discriminant function, including knowing the accuracy of individual classification with casewise diagnostics.

2.4. Discriminant Analysis Model

The discriminant analysis model is an equation that shows a linear combination of various independent variables, namely [3]:

$$D_i = b_0 + b_1X_{i1} + b_2X_{i2} + b_3X_{i3} + \dots + b_pX_{ip}$$

Were D is an independent variable and D_i is value of discriminant score from observation / i-responden ($i = 1, 2, \dots, n$), b_j is discriminant coefficient of free variable or j-attribute ($j = 0, 1, 2, \dots, p$) and X_{ij} is j-independent variable (attribute) of the i-respondent.

3. Result and Discussion

This research uses quantitative research method, that is with collecting secondary data in accordance with the variables research that has been determined.

3.1. Types and Data Sources

The type of data used in this study is secondary data and data sources used are demographic, social and economic data area of North Sumatera in 2012 which has been published by the Agency Statistics Center.

3.2. Types and Data Sources

The sample in this study is all districts / cities in North Sumatera which amounted to 33 districts / cities. The variable is bound to the study this is the Percentage of Poverty (%), the dependent variable is divided into 2 (two) groups, namely:

- a. Group I, low poverty rates are areas with percentages poor population $< 15\%$.
- b. Group II, high poverty rates are areas with percentages the poor $\geq 15\%$.

while the independent variables are:

1. X_1 : Population Density (Soul / Km^2)
2. X_2 : Open unemployment rate (%)
3. X_3 : Per Capita PDRB Based on
Constant Price (Thousand Rupiah)
4. X_4 : Per Capita PDRB Based on
Current Prices (Thousand Rupiah)
5. X_5 : Economic Growth (%)
6. X_6 : Life Expectancy (AHH) (%)

Before performing discriminant analysis, it must be met first some assumptions, ie data must be multivariate normal distribution and the variance in each group is the same.

3.3. Test Data Normality

Calculations using SPSS program, can be seen from the test *Kolmogorv Smirnov* that the value of *Asymp.Sig. (2-tailed)* is 0.972 more large from 0.05 so it can be concluded that the data being tested is distributed normal.

3.4. Average Equality Test

Calculations using the SPSS program, judging from the results of the value significant that the figure is significant for the Open Unemployment variable (TPT) of 0.034 and PDRB Based on Constant Price (PDRB ADHK) 0.036 below 0.05 which means that there are inter-group differences, or a regions are said to have low and high dependent poverty rates on the Unemployment and Open Unemployment Rate variables Constant.

3.5. Matrix Covariance Similarity Test

Calculations using the SPSS program, viewed from the output of *log determinant* table that the number of *log determinant* for the level category low poverty (2,900) and high poverty (2,123) are no different many, so the *covariance matrices group* would be relatively similar to both groups. This means that the data has met the assumption of discriminant analysis, so the discriminant analysis process can proceed.

Table 1. Log Determinant

Y	Rank	Log Determinant
0	1	2.900
1	1	2.123
Pooled within-groups	1	2.790

3.6. Discriminant Analysis Process

Calculations using the SPSS program, based on Table *Variables in the Analysis* variables of the selected into the discriminant function and significant is TPT. Or it could be said to affect high or high the low level of poverty in the District / City of Sumatera Utara. Next will be analyzed based on the selected variables using the SPSS program:

3.6.1. Significant Test of Discriminant. Functions Formed

Table 2. Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.862	4.517	1	.034

Viewed table *Wilks' Lambda Chi-Square figure* of 4.517 with the 0.034 significance level well below 0.05 then sufficient evidence for reject H_0 , with a high significance level indicating the difference which is evident between the two regional groups of the Regency / City low poverty rates and high poverty rates.

3.6.2. Test the Power of Discriminant. Function Relationships

Table 3. Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.160 ^a	100.0	100.0	.371

Based on the Eigen Value table above, *Canonical Correlation* measures closeness relationship between *discriminant score* and group. The *Canonical Correlation* number of 0.371 indicates the closeness which is enough, with the size of the association scale between 0 to 1.

3.6.3. Discriminant Function

Table 4. The Canonical Coefficient Of Discriminant Function

	Function
	1
TPT	.248
(Constant)	-1.448

Based on the canonical coefficient table above the discriminant function, then obtained discriminant function:

$$D = -1,448 + 0,248 TPT$$

$$D = -1,448 + 0,248 X_2$$

3.6.4. Feasibility Assessment of Discriminant Functions

From the *Classification Results*, obtained the prediction accuracy of the model is 72.7%, so it is evident that the discriminant function has high precision prediction, because in general the accuracy in over 50% is considered valid. So the discriminant function can be used to predict factors affecting poverty district of North Sumatera. Based on the discriminant function that has been formed and tested its feasibility it can be determined the composition of group members (region low poverty rates and high poverty rates).

Table 5. Main Opportunities for Groups

P_Kemiskinan	Prior	Cases Used in Analysis	
		Unweighted	Weighted
0	0.500	26	26.000
1	0.500	7	7.000
Total	1.000	33	33.000

The main Opportunity Table for the above Group shows the composition to 33 objects, which by discriminant model yields 26 objects in groups of low poverty levels and 7 objects in groups high poverty rates, as well as *Casewise Statistics* tables the exact area predictions there are as many as 24 regions and the wrong classification there are 9 regions.

4. Conclusion

Based on the results of research on "Influencing Factors Poverty of North Sumatera Region Using Discriminant Analysis ", obtained the factors that most affect the level of poverty in The North Sumatera is open unemployment and its discriminant functions are: $D = -1,448 + 0,248 X_2$. The discriminant (function) model obtained has precision classifies case of 72.7%. Because above 50%, the precision of the model is considered high so that the above model (function) can be used for classify a case of factors affecting poverty district of North Sumatera.

References

- [1] Andriani, Yuniarti. 2010. Analisis Diskriminan untuk Mengetahui Faktor yang Mempengaruhi Pilihan Program Studi Matematika di FMIPA dan FKIP Universitas Sriwijaya. Penelitian Sains. 14403:1-2.
- [2] Santoso, Singgih. 2010. Statistik Multivariat: Konsep dan Aplikasi dengan SPSS. PT Elex Media Komputindo. Jakarta.
- [3] Supranto, J. 2010. Analisis Multivariat Arti dan Interpretasi. PT Rineka Cipta. Jakarta.
- [4] Todaro Michael P. 1995. Ekonomi untuk Negara Berkembang. Edisi Cetakan Pertama. Bumi Aksara. Jakarta.