Journal
of Statistics

DISCRIMINANT ANALYSIS TO PREDICT WORKERS WANTED TO FIND OTHER JOB
(Study on National Labor Force Survey Data, 2017 City of Padang Panjang)

Muhammad Faiz El Haq[1], Pardomuan Robinson Sihombing[2]
[1,2]Badan Pusat Statistics Indonesia

e-mail: *robinson@bps.go.id*

ABSTRACT

This study aims to produce a model function that can predict whether a worker wants to find another job or not by looking at demographic characteristics such as age and job characteristics such as length of work in the main job, working hours during the past month, and income received. The variables observed were age, length of work in the primary position, working hours during the past month, and income. The dependent variable is the desire to find another job. The research sample used secondary data, namely the results of the 2017 National Labor Force Survey (SAKERNAS) conducted by the Central Statistics Agency of Padang Panjang City. Based on the analysis results, the discriminant function contains significant Age and income variables with a negative correlation for age and a positive correlation for income.

Keywords: *Desire To Find Another Job, Discriminant, Multivariate*

INTRODUCTION

Multivariate analysis is an object of study in statistics that studies behavior with the relationship of two or more variables, so it can be said that multivariate analysis is an extension of univariate or bivariate analysis. Currently, multivariate statistics are applied in almost every branch of science, both natural and social sciences. The techniques are preferred because they can model real system complexity even though they are challenging to implement (Suryanto,1990).

Discriminant analysis is one of the methods used in multivariate analysis with the dependency method (where the relationship between variables can be distinguished from which the dependent variable is the independent variable). Discriminant analysis is used in cases where the independent variable is matrix data (interval or ratio), and the dependent variable is nonmetric (nominal or ordinal) data (Johnson and Wichern, 2002).

Change is one of the characteristics of life. Every area of human life is touched by evolution, as is one's job. Changes in the field of work can be divided into two. First, changes that occur voluntarily or voluntarily, namely changes that occur due to the workers' initiative. Second, changes that occur involuntary or involuntary. Changes in work like this are not the individual's will, but because of things outside himself. However, before moving to a new job, many factors must be considered to have no regrets in the future.

Seeing the various effects of a job change, it is interesting to explore the factors determining job change. Several researchers have found that the factors that influence this change in employment include 1. demographic characteristics, such as age, gender, marital status, number of children, and education; 2. job characteristics, such as work experience, Income, and the existence of a trade union; and 3. Income

The formulation of the problem that the researcher wants to be based on the background that has been previously stated is that the researcher wants to predict the category of wanting another job from a worker based on age, length of work in the primary job, working hours during the past month, and income. Can condition the main factors that dominate a worker not wanting another job, so it is hoped that the worker can focus on existing work so that his work results will be maximized.

MATERIAL AND METHODS

1. Data Source

The objects selected in this study were Padang Panjang City residents who were selected as the 2017 SAKERNAS Sample of Padang Panjang City who was working. The amount of data processed was 203 individual data which were used as the object of analysis in this research.

2. Discriminant Analysis

Discrimination and classification are multivariate techniques with a focus on separating a group of objects into groups that have been defined/formed previously (Johnson, 2002). Discriminant analysis is similar to multiple linear regression (multivariate regression). The difference is, Discriminant analysis is used if the dependent variable is categorical (using a nominal or ordinal scale) and the independent variable uses a numerical scale (interval or ratio). Whereas in multiple regression the dependent variable must be numeric, and the independent variable can be numeric or non-numeric. Just like multiple regression, in Discriminant analysis the dependent variable is only one, while the independent variable is multiple. Because it has dependent and independent variables, Discriminant analysis can be classified as a dependent technique. (Annas S. and Irwan, 2015).

The discriminant analysis model is a linear combination of the independent variables, namely:

$$D = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k \qquad (1)$$

where

D = Discriminant Score

b = coefficients of Discrimination

X = Independent Variable

The stages of discriminant analysis can be described as follows (Jayanti, 2014).

1. Significance Test of Discriminant Function

Suppose there are independent variables that are not significant. In that case, these variables can be excluded from the analysis calculation then the analysis starts again from the significance test of the discriminant function.

2. Strength Test of Discriminant Function Relationship

3. Establishment of the discriminant function

4. Classification

The classification process can use a cutting score calculation. $Z_{cu} = \dfrac{n_1 Z_1 + n_2 Z_2}{n_1 + n_2}$ $\qquad (2)$

Prediction Accuracy $Akurasi = \dfrac{Amount\ of\ data\ predicted\ correctly}{Number\ of\ predictions\ made} \ x\ 100\%$ $\qquad (3)$

The stages of Discriminant Analysis are:
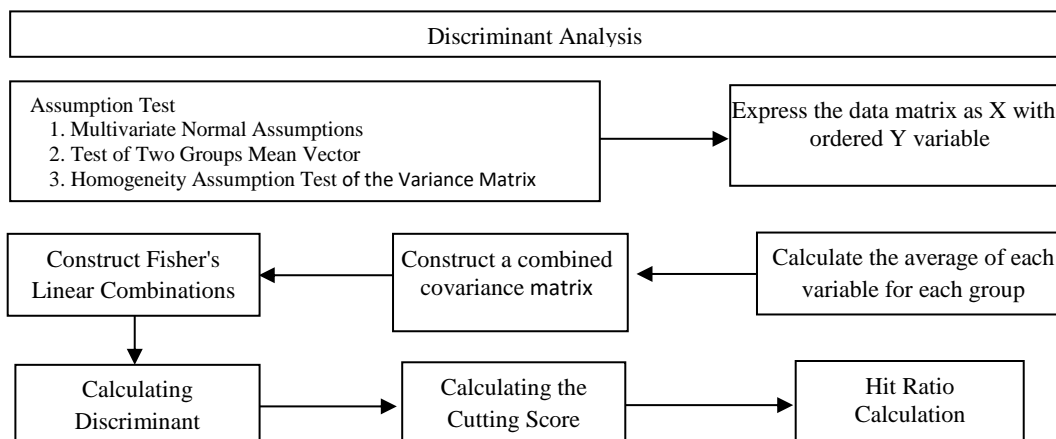

Figure 1. Stages of Discriminant Analysis

RESULTS AND DISCUSSION

1. Stages of Fulfilling Assumptions
   a. Multivariate Normality Test
      By using the R software, a scatter plot is obtained; in this case, the Mahalanobis distance with the chi-square value is as follows:
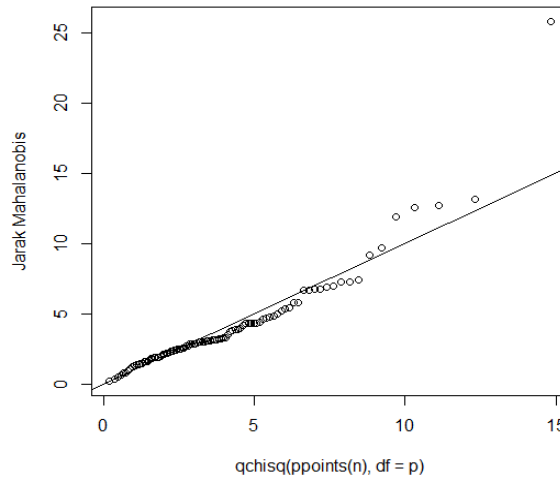

Figure 2. Q-Q Plot Desire to find another job

Because these scatter-plots tend to form a straight line and more than 50% of the Mahalanobis Distance value is less or equal to the chi-square value, the data is a multivariate normal distribution.

Table 1. The Output of the Variance Matrix Homogeneity Test

	DF	F value	Pr (>F)
Desire to Find Another Job based on Age	1	3,5705	0,06025
Desire to Find Another Job based on Length of Work in Main Job	1	16,34	$7{,}537\ e^{-05}$
Desire to Find Another Job based on hours of work a month ago	1	0,0214	0,8839
Desire to Find Another Job based on Income	1	28,448	$2{,}579e^{-07}$

Based on the output, it is found that those that meet the assumption of variance homogeneity are only age and working hours a month ago, while income and length of work in the main job do not.

   b. Test Vector mean of the two groups
      1) Hypothesis
         H0: The average of group 1 is the same as the average of group 2
         H1: Group 1 average is the same as group 2 average

         Test Criteria
            If | t count | > | t critical |, then H0 is rejected and H1 is accepted, and vice versa
            If the p-value $<\alpha$ then H0 is rejected and H1 is accepted, and vice versa

      2) Test Statistics
         By using Software R, the following output is obtained.

Table 2. Average Vector Test Output of the Two Groups

	t	df	p-value	Mean of x	Mean of y
Desire to Find Another Job based on Age	8,0706	201	6,277e^{-14}	43,94	32,89
Desire to Find Another Job based on Length of Work in Main Job	7,6042	176,39	1,638e^{-12}	167,79	64,86408
Desire to Find Another Job based on hours of work a month ago	-0,63348	201	0,5271	174,92	181,2816
Desire to Find Another Job based on Income	7,3103	124,82	2,803e^{-11}	3.281.975	1.476.049

Based on this output, it can be concluded that the desire to find another job based on age, length of work in the main job, and income has a different average between the two groups, while the average working hours a month ago was not different between the two groups. So that the X variable that will be analyzed next is only age, length of work in the main job, and income. The existing assumption test stages have been met, so the data can be used for the analysis process

2. Analysis Stages
a. Establishment of Discriminant Functions
By using Package lda from Software R, the following output is obtained:

Table 3. Output Discriminant Functions

		Want to find another job	Don't want to find another job
Prior probabilities of group		0,4876847	0,5123153
Group means	(desire to find another job $ age)	44,04040	32,90385
	(desire to find another job $ income)	1 526 384	3 164 591
			LD1
Coefficients of linear discriminants	(desire to find another job $ age)		-8,181740e^{-02}
	(desire to find another job $ income)		3,676873e^{-07}

Based on the output figure above, the Discriminant Functions produced in the Discriminant Analysis are

$$Y = -0,0818X_1 + 0,000000367X_4 \qquad (4)$$

From this model, it can be seen that the coefficient X1 (age) has a negative contribution. This is in line with Suriastini's (2006) research which states that increasing age will reduce the intention to seek new ones. This is because age is increasing, so the need is increasing so that it will reduce the intention to change jobs. In other hand, X4 (income) has a positive contribution to the desire to find another job. This is in line with Siregar's research (2015) which states that there is a relationship related to income and the desire to change jobs. If there is information about a higher income in other places, a person tends to want to change jobs for a better life.

b. Classification

By using Package KlaR from Software R, the following output is obtained

Partition Plot

app. error rate: 0.187

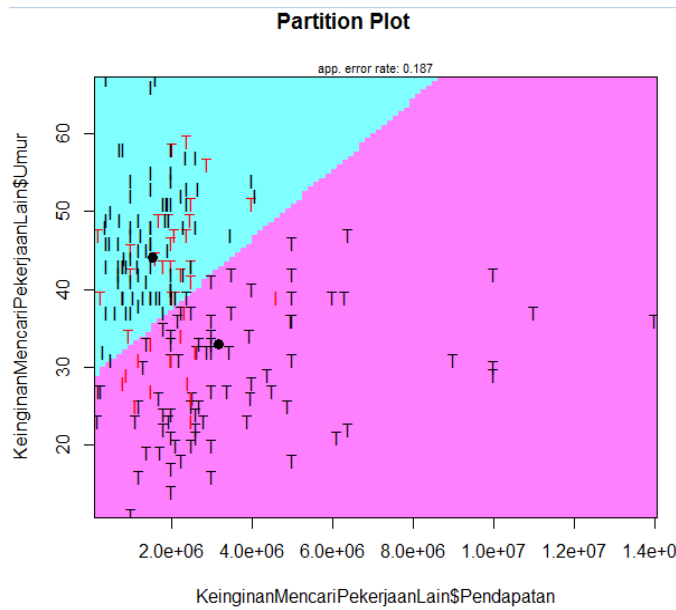KeinginanMencariPekerjaanLain$Pendapatan

Figure 3. Partition Plot

c. Classification Accuracy

The results of classification using the Discriminant Function using R software produce the Prediction Accuracy Table between the Number of Original Categories and the Prediction Categories as follows:

Table 4. Table of Prediction Accuracy Using Classification

	Prediction		
		Want to find another job	Don't want to find another job
Original	Want to find another job	79	20
	Don't want to find another job	18	86

The table above shows that of the total Wanted category there were 99 observations, with a discriminant function classification 79 observations were predicted to Want, and 20 were predicted Not Want, meaning that there were 20 misclassifications of observations. Meanwhile, from the total category of Do Not Want, there are 104 observations, with the discriminant function classification 86 observations are predicted to be Not Wanted, and 18 are predicted to Want means that there are misclassification of 18 observations. So that the level of accuracy of the discriminant function can be expressed by

$$accuracy = \frac{79+86}{203} x \ 100\% = 81,28\% \tag{5}$$

CONCLUSION

Based on the results and discussion, it can be concluded that the significant variables to predict the desire of workers to find another job are age and income variables with the function of the model below which shows a negative correlation between age and income variables and the desire to find another job.

The discriminant function $Y = -0,0818X_1 + 0,000000367X_4$ can be used by related parties to predict workers' desire to find another job (or in other words to see the level of worker satisfaction with their current job with a classification prediction accuracy of 81.28%.

For further research, the researcher suggests adding the X variable, because the two X variables are still not representative enough to describe the difference between workers who want other jobs and not.

REFERENCES

Annas S. and Irwan. 2015. *Application of Discriminant Analysis in Grouping Poor Villages in Wajo District* Jurnal SCIENTIFIC PINISI, Vol. 1 No.1 October 2015

Badan Pusat Statistik, 2017. Pedoman Pendataan Survei Angkatan Kerja Nasional Tahun 2017. Jakarta

Dadang Sukandar, Nani Sufiani Suhanda, Leily Amalia, and Khairunisa. 2008. Discriminant Analysis to Determine Poverty Lines. Journal of Nutrition and Food, Vol 3 No 2 Pages 94-100.

Jayanti, Ririn Dwi and Noeryanti . 2014. Application of the K-Nearest Neighbor Method and Discriminant Analysis for Credit Risk Analysis at Savings and Loans Cooperatives at Kopinkra Sumber Rejeki.Yogyakarta: SNAST.

Johnson, Richard Arnold. 2002. Applied Multivariate Statistical Analysis 5th Edition. New Jersey: Prentice Hall

Prana Ugiana Gio, Dasapta Erwin Irawan, Studying Statistics with R: USU Press

Rusli, Syamsul Bahri Gafar, Jasruddin, and M. Arifin Ahmad. 2015. Testing of Multivariate Normal and Homoscedasticity Matrix of Covariance Variants on Learning Achievement and Graduated Credit for Class 2014 UNM Students

Siregar, Isni Annisaa.2015 .Faktor-faktor Yang Berhubungan Dengan Keinginan Pindah Kerja (Turn Over Intention) Perawat di Rumah Sakit Sehat Terpadu DD Tahun 2014. SKIPSI, Jakarta: UIN Syaraif Hidayatul

Suriastini, IW. 2006.Hubungan Umur dengan Perubahan Pekerjaan: Suatu Aplikasi Metaanalisis Dalam Ukuran Sampel.Jurnal Populasi vol 1 no 17

Suryanto. 1990. Multivariate Statistical Methods. Jakarta; P2LPTK.

Yuli Andriani, Dian Cahyawati, and Vivin Gusmaryanita. 2011. Discriminant Analysis to Determine the Factors Affecting the Choice of Mathematics Study Programs in FMIPA and FKIP Sriwijaya University. Journal of Science Research, Vol. 11 No. 4 (A) 14403 October 2011.