O.S. Balogun, B.A. Oyejola, T.J. Akingbade / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue5, September- October 2012, pp.936-938

Use Of Discriminant Analysis In Classification Of Drug Peddlers And Non-Drug Peddlers In Kwara State.

[*]O.S. Balogun, [**]B.A. Oyejola and [***]T.J. Akingbade

[*]Department of Statistics and Operations Research, Modibbo Adama University of Technology, P.M.B. 2076, Yola, Adamawa State, Nigeria

[**]Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria

[***]Department of Mathematical Sciences, Kogi State University, Anyigba, Kogi State. Nigeria

Abstract

National Drug and Enforcement Agency has been using oral evidence in the past rather than a statistical tool to classify drug offenders into Drug Peddlers and Non-Drug Peddlers.The aim of this research is to construct a Discriminant Function that can be used to classify persons for drug related offences into two groups namely: Drug Peddlers and Non-Drug Peddlers. The following variables were used: Type of Exhibit, Age, Weight of Exhibit and Gender.A discriminant function was obtained and used for classifying drug offenders into groups. The result shows that Type of Exhibit, Age, Weight of Exhibit and Gender contribute to the discriminant function. The misclassification rate obtained was 28.2%.

Keywords: Drug Peddlers, Non-Peddlers, Discriminant Function, Classification.

Introduction

Discriminant analysis has had its earliest and most widespread educational research applications in the areas of vocational and careers development. Because education prepares people for a variety of positions in the occupational structures prevalent in their societies, an important class of education research studies is concerned with testing of theories about the causes of occupational placements and/or the estimation of production equations for allocating positions or anticipating such allocation. Discriminant analysis is a descriptive procedure of separation in which linear functions of the variables are used to describe or elucidate the differences between the two or more groups. That is, the aim of this analysis includes identifying the relative contribution of say, p variables to separation of groups and finding the optimal plane on which the points can be projected to best illustrate the configuration of the groups (Rencher, 2002).

The classification of objects to groups is usually thought of as partition of the objects into subsets in which the members are more similar. Classifying individuals into groups such that there is a relative homogeneity between the groups and heterogeneity between the groups is a problem which has been considered for many years (Ganesalingam, 1989).

The National Drug Law Enforcement Agency (NDLEA) in Nigeria is not an exemption. Some drugs are prohibited from the open market because of their side effect and abuse by public. The drug could be in tablet or liquid form such as Marijuana, Heroin, and Cocaine etc. To restrict the use government placed ban on them and offenders are penalized.

A socially undesirable class, including prostitutes, thieves and hoodlums had been known to use the forbidden drugs. The use also leads to violence among the users and also stimulates sexual assaults on the female folks. According to (Odedeji, 1992), those in peddling are ignorant of the problems caused by the drugs. Persons or individuals arrested for the related offences are taken to court and convicted based on the oral evidences supplied and amount of substances caught with them. It is taken as given that a peddler deserves a stiffer penalty than users. The reason being that dealing with prohibited drugs could be drastically reduced if those peddling face stiff penalties. These individuals are very difficult into peddlers and non-peddlers on the basis of possession and dealing and other variables.

The current effort enables us use a scientific method in classifying drug related offenders.

The Discriminant Model

The elements of the discriminant models are given as

$$Z = a + W_1X_1 + W_2X_2 + ... + W_kX_k$$

Where

Z = discriminant score

a = discriminant constant

W_k = discriminant weight or coefficient

X_k = an independent variable or predictors variable

Discriminant analysis uses ordinary least squares to estimate the values of the parameters ‘a’ and W_k that minimize the within Group Sum of Squares. Discriminant Analysis involves deriving linear combination of the independent variables that will discriminate between the prior defined groups in such a way that the misclassification error rates are minimized (Dillion and Goldstein, 1984). The function is given as

$D(X) = b'\underline{X}$

where $b' = S^{-1}(\bar{\underline{X}}_1 - \bar{\underline{X}}_2)'$

$X = (\bar{\underline{X}}_1 - \bar{\underline{X}}_2)$

Where, $D(X)$ is a $1 \times n$ vector of discriminant scores, b' is a $1 \times p$ vector of discriminant weights, and X is a $p \times n$ matrix containing the values of the n individuals on the p independent variables. S^{-1} the inverse of the pooled sample variance – covariance matrix of the independent variable. The Mahalanobis generalized distance D^2 Statistic is used to determine whether the between group differences in mean score profiles are statistically significant. Large values of D^2 would lead us to believe that the groups are sufficiently spread in terms of mean separation (Dillion and Goldstein, 1984). It is given as:

$D^2 = (\bar{\underline{X}}_1 - \bar{\underline{X}}_2)' S^{-1} (\bar{\underline{X}}_1 - \bar{\underline{X}}_2)$

The test can be constructed by forming:

$$Z = \frac{n_1 n_2 (n_1 + n_2 - p - 1) D^2}{n_1 + n_2 (n_1 + n_2 - 2) p}$$

Data Analysis and Result

The analysis is done to compute the discriminant weights, to examine the associated significance and assumption tests based on linear combination of the predictor variables and also to classify each case into one of the two groups it closely resembles. The variables used in this analysis are Dependent variable: group 1, group 2 and Independent variables: Type of Exhibit (X_1), Age (X_2), Weight of Exhibit (X_3) and Gender (X_4).

(a) **Test for equality of means**

Table 1: Wilks' Lamda

Test of function(s)	Wilks' Lamda	Sig.
1	0.919	0.000

Hypothesis: $H_0 : \underline{\mu}_1 = \underline{\mu}_2$ vs $H_1 : Not\ H_1$

Test Statistic: $\lambda = \dfrac{|BSS|}{|BSS + WSS|}$

$\alpha = 0.05$

Decision: since $p - value(0.001) > 0.05$, Reject H_0.

Under the hypothesis $H_0 : \underline{\mu}_1 = \underline{\mu}_2$ and common variance covariance matrix the test statistic F is distributed as a $F - distribution$ with P and $n_1 + n_2 - p - 1\ df$ i.e.

$F \sim F_\alpha : (p, n_1 + n_2 - p - 1)$

Using the above relationship H_0 is rejected at the significance level α, if $F \geq F_\alpha (p, n_1 + n_2 - p - 1)$

Classification Rule

Assign an individual with realized score $\underline{X}$ on the p independent variables to G_1 if $D(X) \geq C$ otherwise, to G_2 if $D(X) < C$

Where

$C = (\bar{\underline{X}} - \bar{\underline{X}})' S^{-1} (\bar{\underline{X}} - \bar{\underline{X}})$ We assume that in each group observed scores on the p independent variables as multivariate normal with mean $\mu_i, i = 1, 2$ and variance – covariance matrix S^{-1}, and if we can further assume that the prior probabilities of group membership and costs of misclassification of an individual that actually belongs to group $1(2)$ into $2(1)$ are equal.

Conclusion: The vector of the 2 group means are not the same.

(b) **Test for equality of group means individual variable**

Table 2: $t - test$ for equality of mean

Hypothesis: $H_0 : t_1 = t_2$ vs $H_1 :$ Not H_0

Test Statistic: $t = \dfrac{\bar{X}_1 - \bar{X}_2}{S_P \sqrt{\dfrac{1}{N_1} + \dfrac{1}{N_2}}}$

Where $S_p = \sqrt{\dfrac{N_1 S_1^{\ 2} + N_2 S_2^{\ 2}}{N_1 + N_2 - 2}}$

Variable	t	df	$Sig.$
Age	0.2739	260	0.785
Weight	4.4161	260	< 0.001

(i) Age

Decision: Since $p - value\ (< 0.785) < 0.05.$ Reject H_0.

Conclusion: There is a difference in the mean ages for group 1 and group 2.

(ii) Weight of Exhibit

Decision: Since $p-value\ (<0.001)<0.05.$

Reject H_0.

Conclusion: There is a difference in the mean Weight of Exhibits for group 1 and group 2.

(c) **Table 3: Canonical Discriminant Function Coefficient**

Variable	Function
	1
Exhibit	–1.373
Age	–0.022
Weight	0.104
Gender	0.151
Constant	1.653

Hence,

$$D(X)=1.653-1.373X_1-0.022X_2+0.104X_3+0.154X_4$$

(d)Table 4: Classification of Results

	Predicted Group Membership		Total
Original count	1	2	
1	67	52	119
2	22	121	143
% 1	56.3	43.7	100
2	15.4	84.6	100

From the table 71.8% of the group cases were correctly classified while 28.2% were wrongly classified.

Conclusion

In general, we were able to construct a Discriminant score: (1) for detecting the variables (Type of Exhibit, Age, Weight and Gender) that allow the researcher to discriminate between Drug Peddlers and Non –Drug Peddlers, and (2) for classifying cases into different groups which is better than oral evidence. Also, we have shown that the group differs with regards to the mean of variables, and the variables to predict group membership.

References:

1. **Dillion, .W.R. and Goldstein, .S. (1984),** "Multivariate Analysis Method and Applications. John Wiley Sons, Inc. Canada.
2. **Fisher, R.A. (1936), "**The Use of Multivariate Measurements in Taxonomic Problems", Annals of Eugenics, 7, 179-188.
3. **Ganesalingam, .S. (1989),** "Classification and Mixture Approaches to Clustering via Maximum Likelihood".Applied Statistics, 38, NO. 3, Page 445-466.
4. **Odedeji, A.O. (1992),** Drugs in the Third World in Drugs and Society to Year 2000, Ed by Vamos And Corrivean, Page 116-119.
5. **Rencher, A.C. (2002),** Method of Multivariate Analysis, Second Edition. John Wiley and Sons, Inc.