

Multivariable Discriminant Analysis; Application of a Three Dimensional Case on Students Measurements

Usooro Anthony E.

Department of Mathematics and Statistics, Akwa Ibom State University, Mkpato Enin, Akwa Ibom State, Nigeria

Abstract A three dimensional discriminant analysis was the research interest. Memberships in three different groups were studied. Three linear discriminant models were estimated from the pair-wise analysis. Misclassifications were identified in each group and associated probabilities of misclassification obtained.

Keywords Discriminant Analysis, Classification, Probability of Misclassification

1. Introduction

A discriminant analysis is a multivariate statistical method whose objective is to have a critical study about population and seek to differentiate between two or more groups within the population with respect to certain characteristics [2]. In most cases, observations or individuals under careful examination may have some properties that restrict their fundamental membership to two predefined groups or populations of which a group will have similar properties amongst its members, but distinct from another group or population. In spite of the common similarities, there may be some differentials based on certain features or characteristics, which could make some members of a particular group or population different from other members within the group [4].

The aim of discriminant analysis is to establish a parametric based procedure, known as a linear discriminant function with which individuals' membership can be assessed [10]. Classification or discriminant analysis has long been a subject of extensive studies in chemometrics. So far, a number of chemometrics methods have been established for discriminant function modeling. Among those are: (Bayesian) Quadratic discriminant analysis (QDA), (Bayesian) Linear discriminant analysis (LDA), Fisher linear discriminant analysis (FLDA), Discriminant partial least squares (DPLS), Soft independent modeling of class analogies (SIMCA) and Artificial neural networks (ANN) have been widely used for practical purposes [3].

Discriminant analysis has very wide areas of application in sciences, business, education and economic studies. [8] Carried out a research to distinguish relevant indices from

the vast range of indices that determine the economic situation or that influence the evaluation process and enable forecasting of its further state. The indices included gross domestic product, individual house hold consumption, public expenditures, gross fixed capital formation, export of goods and services, import of goods and services, and average monthly gross salary in real terms. Their research work covered the period, 1991-2002, and divided into two disjoint classes, namely (i) years of the upturn of polish economy and (ii) years of economic downturn. Discriminant analysis was used to distinguish and rank those indicators that are responsible in the most significant way for the differences between the objects and simultaneously for division into classes. Their work identified barometers for present economic and future economic situations. [9] applied discriminant analysis to re-classify students on the basis of their academic performances. The two groups of students used in the analysis were Business and Accountancy students in an institution. He used five courses offered to the two groups of students. Out 114 students in Accountancy, 37 were discriminated to Business Department, while retaining 77 students in Accountancy. Out of 202 students in Business Department, 78 were discriminated to Accountancy, while 124 were retained based on their performances. [5] applied discriminant analysis on the classification of implemented foreign assisted projects in Nigeria. Their study was to build a model that could be used to classify implemented foreign projects into one of the two categories – sustainable and unsustainable groups. The discrimination was between two groups, and the analysis was carried out using Mahalanobis Distance and Stepwise methods. [7] adopted principal component method to discriminate between two groups (sexes) based on the measurements of their sewing trousers. The parameters used were the measurements taken from both males and females. [1] considered different classification methods that can take care of unequal cost ratio of misclassification. That is appropriate method to be adopted

* Corresponding author:

toskila2@yahoo.com (Usooro Anthony E.)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

when the equal cost of misclassification is violated. From their work, Maximum Linear Discriminant Function (MLDF), Fisher's Linear Discriminant Function (FLDF) and Based Discriminant Function (BDF) were considered. Their investigation revealed that Maximum Linear Discriminant Function (MLDF) was observed as robust linear classifier.

This study is motivated by the need to apply discriminant analysis to a three dimensional case, involving measurements of student's shirts. The three dimensions imply three groups classification.

2. Method of Analysis

2.1. Data Collection Procedure and Variable Description

Secondary School Students were classified into three groups. Group I comprises JSS1 and JSS2 students, Group II comprises JSS3 and SSS1 and Group III comprises SSS2 and SSS3. Measurements were taken from a sample of 32

students in each group. The measuring parameters for the sewing shirts included: chest (X_1), length(X_2), back(X_3), neck(X_4), sleeve(X_5) and wrist(X_6).

2.2. Vector Means and Covariance Matrix

A. The vector means for the three groups are:

$$\bar{X} = \begin{matrix} \text{Group 1} & \text{Group 2} & \text{Group 3} \\ \begin{pmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \\ \bar{X}_{14} \\ \bar{X}_{15} \\ \bar{X}_{16} \end{pmatrix} & \begin{pmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \\ \bar{X}_{24} \\ \bar{X}_{25} \\ \bar{X}_{26} \end{pmatrix} & \begin{pmatrix} \bar{X}_{31} \\ \bar{X}_{32} \\ \bar{X}_{33} \\ \bar{X}_{34} \\ \bar{X}_{35} \\ \bar{X}_{36} \end{pmatrix} \end{matrix}$$

B. The covariance matrix:

$$S = \begin{pmatrix} v(x_1) & & & & & \\ cov(x_2x_1) & v(x_2) & & & & \\ cov(x_3x_1) & cov(x_3x_2) & v(x_3) & & & \\ cov(x_4x_1) & cov(x_4x_2) & cov(x_4x_3) & v(x_4) & & \\ cov(x_5x_1) & cov(x_5x_2) & cov(x_5x_3) & cov(x_5x_4) & v(x_5) & \\ cov(x_6x_1) & cov(x_6x_2) & cov(x_6x_3) & cov(x_6x_4) & cov(x_6x_5) & v(x_6) \end{pmatrix}$$

The above covariance matrix is obtained from each of the groups. The measurements are obtained from the three groups, and the variances and covariance are assumed not equal. Therefore, the pooled covariance matrix becomes,

$$S = \frac{S_1 + S_2 + S_3}{n_1 + n_2 + n_3 - 3}$$

From the above matrix, inverse S^{-1} is obtained. Inverse of a matrix is the transpose of the cofactor divided by the determinant.

C. Linear Discriminant Function:

There are many classification methods earlier mentioned in this paper. [6] applied Fisher's linear discriminant function for reclassification of samples from three varieties of sorghum, using length of flag leaf, width of flag leaf, length of stalk and girth of culm as parameters. Here, we apply the Fisher's linear discriminant function with six measuring parameters on three groups of students. The function is given by

$$F_{mn} = (\bar{X}_i - \bar{X}_j)^T S^{-1} X, i \neq j, m = 1, \dots, 3 \text{ groups, } n = 1, \dots, 32 \text{ individuals.}$$

The discriminant rule is given by

$$\frac{1}{2}(\bar{X}_2 - \bar{X}_1)^T S^{-1}(\bar{X}_1 + \bar{X}_2) \text{ for groups 1 and 2}$$

$$\frac{1}{2}(\bar{X}_3 - \bar{X}_1)^T S^{-1}(\bar{X}_1 + \bar{X}_3), \text{ for groups 1 and 3}$$

$$\frac{1}{2}(\bar{X}_3 - \bar{X}_2)^T S^{-1}(\bar{X}_2 + \bar{X}_3), \text{ for groups 2 and 3.}$$

3. Data Analysis

3.1. Mean Vector and Covariance Analysis

$$\bar{X} = \begin{pmatrix} \text{Group 1} \\ 33.906 \\ 23.844 \\ 14.281 \\ 13.219 \\ 17.062 \\ 8.000 \end{pmatrix} \quad \begin{pmatrix} \text{Group 2} \\ 35.406 \\ 25.906 \\ 16.125 \\ 15.062 \\ 19.562 \\ 9.000 \end{pmatrix} \quad \begin{pmatrix} \text{Group 3} \\ 37.312 \\ 27.031 \\ 18.469 \\ 15.500 \\ 20.937 \\ 9.500 \end{pmatrix}$$

$$(\bar{X}_2 - \bar{X}_1) = \begin{pmatrix} 2.000 \\ 2.062 \\ 1.844 \\ 1.843 \\ 2.500 \\ 1.000 \end{pmatrix}, \quad (\bar{X}_3 - \bar{X}_1) = \begin{pmatrix} 3.406 \\ 3.187 \\ 4.188 \\ 2.281 \\ 3.875 \\ 1.500 \end{pmatrix}, \quad (\bar{X}_3 - \bar{X}_2) = \begin{pmatrix} 1.406 \\ 1.125 \\ 2.344 \\ 0.438 \\ 1.375 \\ 0.500 \end{pmatrix}$$

$$(\bar{X}_1 + \bar{X}_2) = \begin{pmatrix} 69.812 \\ 49.750 \\ 30.406 \\ 28.281 \\ 36.625 \\ 17.000 \end{pmatrix}, \quad (\bar{X}_1 + \bar{X}_3) = \begin{pmatrix} 71.218 \\ 50.875 \\ 32.750 \\ 28.719 \\ 37.999 \\ 17.500 \end{pmatrix}, \quad (\bar{X}_2 + \bar{X}_3) = \begin{pmatrix} 73.218 \\ 52.937 \\ 35.594 \\ 30.562 \\ 40.499 \\ 18.500 \end{pmatrix}$$

$$S = \begin{pmatrix} 1.5732 & & & & & \\ 1.4939 & 1.9990 & & & & \\ 1.8123 & 2.2937 & 3.6122 & & & \\ 0.9661 & 1.2665 & 1.8095 & 1.0897 & & \\ 1.5538 & 1.9254 & 2.6573 & 1.4456 & 2.6411 & \\ 0.7744 & 0.9402 & 1.2849 & 0.6989 & 1.1391 & 0.6452 \end{pmatrix}$$

$$S^{-1} = \begin{pmatrix} 2.3357 & & & & & \\ -1.3541 & 3.2364 & & & & \\ -0.1720 & -0.3115 & 2.0870 & & & \\ 0.4293 & -1.2892 & -2.0586 & 6.7598 & & \\ -0.1655 & -0.3376 & -0.4142 & -0.5838 & 2.2306 & \\ -0.6869 & -0.4779 & -0.5346 & -0.8289 & -1.7905 & 8.1946 \end{pmatrix}$$

3.2. Linear Discriminant Function

A. for groups 1 and 2, the estimated model is

$$F_{12} = 1.2962X_1 - 0.3070X_2 - 2.5020X_3 + 4.5739X_4 + 0.9191X_5 - 1.1543X_6$$

The discriminant figure is obtained as

$$\frac{1}{2}(1.2962, -0.3070, -2.5020, 4.5739, 0.9191, -1.1543) \begin{pmatrix} 69.812 \\ 49.750 \\ 30.406 \\ 28.281 \\ 36.625 \\ 17.000 \end{pmatrix} = 71.2673.$$

B. for groups 1 and 3, the estimated model is

$$F_{13} = 2.3017X_1 - 0.5679X_2 + 0.059X_3 + 0.6453X_4 + 1.2519X_5 - 2.6386X_6$$

The discriminant figure is obtained as

$$\frac{1}{2}(2.3017, -0.5679, +0.059, 0.6453, 1.2519, -2.6386) \begin{pmatrix} 71.218 \\ 50.875 \\ 32.750 \\ 28.719 \\ 37.999 \\ 17.500 \end{pmatrix} = 78.4453.$$

C. for groups 2 and 3, the estimated model is

$$F_{23} = 1.8115X_1 - 0.2609X_2 + 2.5612X_3 - 3.9286X_4 + 0.3327X_5 - 1.4842X_6$$

The discriminant figure is obtained as

$$\frac{1}{2}(1.8115, -0.2609, +2.5612, -3.9286, 0.3327, -1.4842) \begin{pmatrix} 73.218 \\ 52.937 \\ 35.594 \\ 30.562 \\ 40.499 \\ 18.500 \end{pmatrix} = 38.0966.$$

3.3. Discriminant Rule

From the function $F_{mn} = (X_i - X_j)'S^{-1}X$, the measurements are substituted to obtain discriminant figure for each pair of groups. The discrimination rules are:

From Group 1:

Retain Y^{th} individual in group1 if Y^{th} score from model $F_{12} \leq 71.2673$

Assign Y^{th} individual to group2 if Y^{th} score from model $F_{12} > 71.2673$

Assign Y^{th} individual to group3 if Y^{th} score from model $F_{13} > 78.4453$

From Group 2:

Assign Y^{th} individual to group1 if Y^{th} score from model $F_{12} \leq 71.2673$

Retain Y^{th} individual in group2 if Y^{th} score from model $F_{12} > 71.2673$

Assign Y^{th} individual to group3 if Y^{th} score from model $F_{23} > 38.0966$

From Group 3:

Assign Y^{th} individual to group1 if Y^{th} score from model $F_{13} \leq 78.4453$

Assign Y^{th} individual to group2 if Y^{th} score from model $F_{23} \leq 38.0966$

Retain Y^{th} individual in group3 if Y^{th} score from model $F_{23} > 38.0966$

3.4. Confusion Matrix

This table contains both correct and misclassified values from which the probabilities of misclassification are calculated.

The probability of misclassifying Y^{th} individual into group 2 or 3 given that the individual is from group 1 is $P(G_2, G_3/G_1) = 0/32 + 2/32 = 0.0625$.

The probability of misclassifying Y^{th} individual into group 1 or 3 given that the individual is from group 2 is $P(G_1, G_3/G_2) = 3/32 + 0/32 = 0.09375$.

The probability of misclassifying Y^{th} individual into group 1 or 2 given that the individual is from group 3 is $P(G_1, G_2/G_3) = 3/32 + 7/32 = 0.3125$.

The total probability of misclassification is $2/32 + 3/32 + 3/32 + 7/32 = 15/32 = 0.4688$.

Table 1. Classification Table

	G_1	G_2	G_3	TOTAL
G_1	30	3	3	36
G_2	0	29	7	36
G_3	2	0	22	24
TOTAL	32	32	32	96

4. Conclusions

The application of discriminant analysis in this paper was motivated by the need to study membership of individuals in three heterogeneous groups and discriminate their membership based on their measurements. It is an indisputable fact that every unit of a population possesses some characteristics similar to or different from other members of the group in which its membership is established. The need for discrimination arises if there are some individuals in a group whose characteristics are significantly different from general characteristics of members forming the greater proportion of membership in the group. In the three dimensional case carried out in this work, some individuals, based on their measurements have exceeded the next superior group to where they should belong. The process of the analysis was pair-wise, and this ended in estimating three models for discrimination.

REFERENCES

- [1] Ariyo Oludare S. and Adebajji A.O. (2011): Robust Linear Classifier for Unequal Cost Ratios of Misclassification. *CBN Journal of Applied Statistics*. Vol.2 No.1, pp51-59.
- [2] Htiberty, C. J. (1994): Applied discriminant analysis. John Wiley and Sons, New York.
- [3] Jiang Jian-Hui, Tsnkova Roumiana and Ozaki Yukihiro (2001): Principal Discriminant Variate Method for Classification of Multicollinear Data: Principle and Application. *Analytical Sciences*, Vol.17 Supplement pp i471-i474.
- [4] Lubischew, A. A. (1962): On the use of discriminant functions in taxonomy. *Biometrika*, 18: 455-477.
- [5] Nworuh, G. E. and Anyiam K. E. (2010): Application of Discriminant Analysis on the Classification of Implemented Foreign Assisted Projects in Nigeria. *Journal of Nigerian Statistical Association* Vol.22, pp 95-105.
- [6] Onyeagu Sidney I. (2003): A first Course in Multivariate Statistical Analysis. Mega Concept.
- [7] Osuji, G. A., Ogbonna U. V., Nnanatu, C. C., Afuecheta E. O., Nwosu, D. F. (2010): Multivariate Analysis of Tailors Measurement in Sewing Trousers. *Conference of Nigerian Statistical Association*, pp 355-360.
- [8] Toanna A. and Urszula E. (2004): Application of Discriminant Analysis on Economic Situation. *Journal of physical Sciences*. Vol.3 no.2 pp 57-63.
- [9] Usoro, A. E. (2006): Discriminant Analysis and its Application to the Classification of Students on the Basis of their Academic Performances. *Journal of Research in Physical Sciences* Vol.2, No.3 pp53-55.
- [10] Welch, B. L. (1939): Not on discriminant function. *Biometrika*, 31: 218-220.