SciencePG
Science Publishing Group

Annotations on the Relationship Among Discriminant Functions

Awogbemi Clement Adeyeye

Department of Statistics, National Mathematical Centre, Abuja, Nigeria

Email address:
awogbemiadeyeye@yahoo.com

To cite this article:
Awogbemi Clement Adeyeye. Annotations on the Relationship Among Discriminant Functions. *Pure and Applied Mathematics Journal*. Vol. 9, No. 6, 2020, pp. 124-128. doi: 10.11648/j.pamj.20200906.14

Abstract: Different forms of discriminant functions and the essence of their appearances were considered in this study. Various forms of classification problems were also considered, and in each of the cases mentioned, classification from simple functions of the observational vector rather than complicated regions in the higher-dimensional space of the original vector were made. Ever since the emergence of the Linear Discriminant Function (LDF) by Fisher, several other classification statistics have emerged and violation of condition of equal variance covariance matrix for Linear Discriminant Function (LDF) results to Quadratic Discriminant Function (QDF). While the Best Linear Discriminant Function (BLDF) is referred to Best Sample Discriminant Function (BSDF) when the parameters are estimated from a sample and also optimal in the same sense as Quadratic Discriminant Function (QDF), Rao statistic is best for discriminating between options that are close each other. The relationships among the classification statistics examined were established: Among the methods of classification statistics considered, Anderson's (W) and Rao's (R) statistics are equivalent when the two sample sizes n_1 and n_2 are equal, and when a constant is equal to 1, W, R and John-Kudo's (Z) classification statistics are asymptotically comparable. A linear relationship is also established between W and Z classification.

Keywords: Discriminant Functions, Classification Statistics, Classification Problems, Covariance Matrix, Probability of Misclassification

1. Introduction

Discriminant analysis is a statistical method used for classification of objects into mutually exclusive and exhaustive groups on the basis of a set of independent variables. The method handles two or multiple group problems. It also derives linear combinations of the independent variables that discriminate between the a priori defined groups, such that the error rates misclassification are minimized as much as possible [20]. Thus, discriminant analysis finds a mean of classifying objects into groups with accuracy and also determines the dimensions on which the group differ [10].

Suppose $Y \sim N_{1 \times N}(\mu, \Sigma)$, where Σ is positive definite. Then the probability density function (pdf) of Y expressed as

$$f(y/\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}}} \frac{1}{\left|\Sigma^{\frac{1}{2}}\right|} e^{\frac{1}{2}(y-\mu)\Sigma^{-1}(y-\mu)^1}.$$

Given that the individual $f_k^{'s}$ are multivariate normal densities with different means but common covariance matrix, so that $X_i/Y_i = m \sim N_{1xp}(\mu_m, \Sigma)$. The probability density function of Y becomes

$$f_k(X/\mu_m, \Sigma) = \frac{c}{\left|\Sigma^{\frac{1}{2}}\right|} e^{\frac{1}{2}(y-\mu_m)\Sigma^{-1}(y-\mu_m)^1} = \frac{c}{\left|\Sigma^{\frac{1}{2}}\right|} e^{-\frac{1}{2}x\Sigma^{-1}x^1} e^{x\Sigma^{-1}\mu_m^1 - \frac{1}{2}\mu_m\Sigma^{-1}\mu_m^1}.$$

The factors that depend on m are ignored since individual

interest is on the highest pdf of π_m. Thus, for a given X, m is

chosen to maximize $\pi_m \, e^{x\Sigma^{-1}\mu_m^1 - \frac{1}{2}\mu_m\Sigma^{-1}\mu_m^1}$ by taking its log to have the discriminant function:

$$g_m(x) = x\Sigma^{-1}\mu_m^1 - \frac{1}{2}\mu_m\Sigma^{-1}\mu_m^1 + \log\left(\pi_m\right).$$

Several reasons have been given for the emergence of different types of discriminant functions, and notably among them are: contraventions of assumptions employed for the cradle of Fishers Linear Discriminant Function (FLDF); efforts to reduce as much as possible, the derivatives of errors of misclassification; efforts to get permissible methods that minimize probabilities of misclassification and also as an underline issue, testing of hypothesis.

This study therefore considers different forms of classification statistics, instances of classification problems and correlations existing among the discriminant functions.

2. Examples of Classification Problems

(i) A large international Air Carrier collects data on employees in three different job classification (customer service personnel, mechanic and dispatches). The Director of Human Resources may wish to know if these three job classifications appeal to different personality types.[5]

(ii) A number of variables are measured at five weather stations. Based on these variables, we may wish to predict the ceiling at a particular airport in 2 hours. The ceiling categories are closed, low instrument, high instrument, low open and high open.

(ii) In a brand-switching system, one may wish to detect fast and slow consumers of newly introduced product on the basis of consumers' characteristics such as education, income, family size and amount of previous brand-switching.

(iv) A nutritionist may desire to classify different classes of food into distinct category of food nutrients such as carbohydrates, fats and oil, vitamins, proteins, minerals, etc., on the basis of measurements of amount of different nutrients in the food.

(v) Astronomers have been cataloguing distant objects in the sky using long exposure CCD images. The objects need to be labeled as star, galaxy, nebula, etc. The data is highly noisy, and the images are very faint. The cataloguing can take decades to complete. Can an automated cataloguing process be designed to improve its effectiveness and efficiency?

(vi) In a hospital, a patient is admitted with a diagnosis of myocardial infarction, and systolic blood pressure, heart rate, stroke index and mean arterial pressure are obtained by the Doctor. Is it possible to predict whether the patient will survive? On the basis of these measurements, can we compute a probability of survival for the patient? [15, 7].

(vii) In an anthropological study, an Archeologist obtained a jawbone excavated from a burial ground as having belonged to a male or female. Can an assignment be made on the basis of measurements such as circumference and volume made on the jawbones from the two sets of people? [14]

(viii) A Geologist obtained the mean, variance, skewness and kurtosis of the size of particles deposited in a beach. How can these statistics be used to determine if the beach is wave-laid or Aeolian in origin? Are there differences in particle size distribution?

(ix) An emergency room in a hospital measures a number of variables such as blood pressure, age, etc of newly admitted patients. A decision has to be taken whether to put the patient in an Intensive Care Unit (ICU). Due to the high cost of ICU, those patients who may survive a month or more are given higher priority. The problem is to predict high risk patients and discriminate them from low-risk patient [5].

(x) A credit card company receives hundreds of thousands of applications for new cards. The application contains information regarding several different attributes, such as annual salary, any outstanding debts, age, etc. The problem is to categorize applications into those who have good credit, bad credit, or fall into a gray area [5].

(xi) African or "killer bees" cannot be distinguished visually from ordinary domestic honey bees. What kind of variables based on chromatograph peaks can be used to readily identify them? [1].

(xii) A meteorologist wants to predict cloud ceiling at time, t_1, on the basis of physical measurements acquired at time, t_0, t_1, where $t_0 < t_1$. In this case, it is assumed that historic data are readily available to assist in determining an assignment rule [20].

3. Methods of Classification Statistics

Different types of discriminant functions and some of their properties are appraised in this study:

3.1. Linear Discriminant Function (LDF)

The Linear Discriminant Function (LDF) is a statistical procedure constructed as

$$C\left(\underline{X},; \underline{\mu}_1, \underline{\mu}_2; \Sigma\right) = \left(\underline{\mu}_1 - \underline{\mu}_2\right)^1 \Sigma^{-1} \underline{X} \qquad (1)$$

It assigns p-dimensional observation vector, $\underline{X}$, into one of the two populations, π_i (i = 1, 2), and it is employed as an assignment rule when the following assumptions are satisfied:

The density functions of observations from populations π_1 and π_2 are multivariate normal; $\left(\pi_i \sim N_p(\underline{\mu}_i, \Sigma), i = 1, 2\right)$; the variance-covariance matrix (Σ_1) in population, π_1 is the same as (Σ_2) in population π_2; the prior probabilities of observations coming from populations π_1 and π_2 are known; the parameters of the density functions are also known.

However, the unknown parameters are estimated from the samples as

$$C\left(\underline{X};\underline{\mu}_1,\underline{\mu}_2;S\right)=\left(\overline{X}_1-\overline{X}_2\right)^1 S^{-1}\underline{X}, \qquad (2)$$

where $\overline{X}_1,\overline{X}_2, S$ are estimates of $\underline{\mu}_1,\underline{\mu}_2,\Sigma$ respectively.

Suppose the assumptions specified above are satisfied, then the Linear Discriminant Function (LDF) provides optimal assignment rule in that it cannot be improved upon and the errors of misclassification are minimized. However, when some or all the assumptions are violated, it would be of interest to researchers to determine the effects of the violation

on the procedures using LDF. Based on [8, 19, 18] established that LDF optimal properties for two group classification of the populations are multivariate normally distributed.

3.2. Quadratic Discriminant Function (QDF)

When the assumptions of equal covariance matrices from two populations are violated, QDF arises and the derivation is established using likelihood ratio rule. If the parameters are known, the classification statistic is expressed as

$$D\left(\underline{X};\underline{\mu}_1,\underline{\mu}_2;\Sigma_1,\Sigma_2\right)=\frac{1}{2}\frac{|\Sigma_2|}{|\Sigma_1|}-\frac{1}{2}\left[\left(\underline{X}-\underline{\mu}_1\right)^1\Sigma_1^{-1}\left(\underline{X}-\underline{\mu}_1\right)+\left(\underline{X}-\underline{\mu}_2\right)^1\Sigma_2^{-1}\left(\underline{X}-\underline{\mu}_2\right)\right], \qquad (3)$$

where $\underline{X}$ is a vector of observations, $\underline{\mu}_1,\underline{\mu}_2$ are the mean vectors and Σ_1, Σ_2 are different covariance matrices from populations π_1 and π_2 respectively.

When the parameters are estimated from the samples,

$$C\left(\underline{X};\underline{\mu}_1,\underline{\mu}_2;S_1,S_2\right)=\frac{1}{2}\frac{|S_2|}{|S_1|}-\frac{1}{2}\left[\left(\underline{X}-\overline{X}_1\right)^1 S_1^{-1}\left(\underline{X}-\overline{X}_1\right)+\left(\underline{X}-\overline{X}_2\right)^1 S_2^{-1}\left(\underline{X}-\overline{X}_2\right)\right], \qquad (4)$$

The sturdiness of QDF was studied by [9] when $\Sigma=1$ and $\Sigma=m$ (m is a constant) and also by [13] in respect of non normality. The statistic (QDF) is optimal when the population parameters are known and $\Sigma_1=\Sigma_2$.

3.3. Anderson's Statistic (W)

The statistic arose from Anderson's derivation in respect of two population parameters that are multivariate normally distributed with different means and constant covariance matrix [3, 2]. When the population parameters are known, the W is defined as

$$C\left(\underline{X};\underline{\mu}_1,\underline{\mu}_2;\Sigma\right)=\left\{\left\{X-\frac{1}{2}\left(\underline{\mu}_1+\underline{\mu}_2\right)\right\}^1\Sigma^{-1}\left(\underline{\mu}_1-\underline{\mu}_2\right)\right\} \qquad (5)$$

and when the parameters are estimated from the samples,

$$C\left(\underline{X};\underline{\mu}_1,\underline{\mu}_2;S\right)=\left\{\left\{X-\frac{1}{2}\left(\overline{X}_1+\overline{X}_2\right)\right\}^1 S^{-1}\left(\overline{X}_1-\overline{X}_2\right)\right\} \qquad (6)$$

The statistic (W) is different from LDF by a constant, $\frac{1}{2}\left(\underline{\mu}_1+\underline{\mu}_2\right)^1\Sigma^{-1}\left(\underline{\mu}_1-\underline{\mu}_2\right)$ and $\frac{1}{2}\left(\underline{\mu}_1+\underline{\mu}_2\right)\Sigma^{-1}\left(\underline{\mu}_1-\underline{\mu}_2\right),$ when parameters are estimated from the samples.

3.4. Best Linear Classification Statistic (BLCS)

The statistic was introduced by [6, 4] under the same condition with QDF. The statistic is expressed as:

$$C\left(\underline{X};\underline{\mu}_1,\underline{\mu}_2;\Sigma_1\Sigma_2\right)=\begin{cases}\underline{X}^1\underline{m}-\underline{m}^1\underline{\mu}_1-\underline{n}_1\underline{m}^1\Sigma_1\underline{m}\\ \underline{X}^1\underline{m}-\underline{m}^1\underline{\mu}_2-\underline{n}_2\underline{m}^1\Sigma_2\underline{m}\,,\end{cases} \qquad (7)$$

where $\underline{m}=\left(n_1\Sigma_1+n_2\Sigma_2\right)^{-1}\left(\mu_1-\mu_2\right)$ and n_1 and n_2 are scalars selected to reduce the probability of misclassification as much as possible.

When the parameters are estimated from the samples, BLCS is defined as:

$$C\left(\underline{X};\overline{X}_1,\overline{X}_2;S_1 S_2\right) = \begin{cases} \underline{X}^1 \underline{m} - \underline{m}^1 \overline{X}_1 - n_1 \underline{m}^1 S_1 \underline{m} \\ \underline{X}^1 \underline{m} - \underline{m}^1 \overline{X}_2 - n_2 \underline{m}^1 S_2 \underline{m} \end{cases}, \tag{8}$$

where $\underline{m} = \left(n_1 S_2 + n_2 S_2\right)\left(\overline{X}_2 - \overline{X}_1\right)$ and n_1 and n_2 minimize probability of misclassification. The BLCS is optimal when the conditions for QDF hold.

3.5. John-Kudo's Classification Statistic (Z)

Let $\underline{X}$ be a vector of observations whose distribution in two populations is multivariate normal, with the same covariance matrix and different estimated sample means. Then the statistic Z is defined as

$$C(\underline{X};\overline{X}_1,\overline{X}_2;n_1,n_2;S) = \frac{n_1}{n_1+1}\left(\underline{X}-\overline{X}_1\right)S^{-1}\left(\underline{X}-\overline{X}_1\right) - \alpha\frac{n_2}{n_2+1}\left(X-\overline{X}_2\right)S^{-1}\left(X-\overline{X}_2\right), \tag{9}$$

where n_1 and n_2 are the sample sizes, α is a constant and S is the estimated covariance matrix [11, 12].

3.6. Rao's Classification Statistics (R)

The Statistic (R) was derived by [16] under the condition that the distribution of a vector of observations $\underline{X}$ from two populations π_1 and π_2 is multivariate normal. He also assumed that samples n_1 and n_2 are from populations π_1 and π_2 respectively. The R statistic is expressed as

$$C(\underline{X};\overline{X}_1,\overline{X}_2;n_1,n_2;\Sigma) = \frac{c_1^2 - c_2^2}{d_2^2}U\Sigma^{-1}U + 2\frac{(c_1-c_2)}{d_1 d_2}U\Sigma^{-1}K^1, \tag{10}$$

where

$$c_1 = \frac{n_2}{n_1+n_2}, \quad c_2 = -\frac{n_1}{n_1+n_2}, \quad d_1 = \frac{1}{n_1}+\frac{1}{n_2^2}, \quad d_2 = 1+\frac{1}{n_1+n_2}, \quad U = \underline{X} - \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1+n_2}, \quad K = \overline{X}_1 + \overline{X}_2$$

4. Correlations Linking the Classification Statistics

The LDF results to QDF when the variance covariance matrices condition is violated. If the samples sizes n_1 and n_2 are the same, then Rao's (R) and Anderson's (W) statistics are comparable. When $\alpha = 1$, the statistics W, R and Z are asymptotically comparable. Suppose that $n_1 = n_2 = n, \alpha = 1$, then there is a linear relationship between Z and W statistics expressed as

$$Z = -\left(\frac{2n}{n+1}\right)^W \tag{11}$$

If the two sample sizes are the same, and the total sample size less by 2 are sufficiently large, then $\sum_{i=1}^{n_1} \frac{n_1}{2}n_1 + 1$ approximates to 1 (Siotani, 1975).

5. Conclusion

The problem of classification of observations has been the focus of research since the introduction of Fisher's Linear Discriminant Function (FLDF). Contraventions on the assumptions of LDF have resulted to appearances of QDF and other discriminant functions. Correlations existing among LDF, QDF and some other forms of classification statistics are well established in this study. For each of the classification problems considered, a classification from a simple function of observational vector has been made, rather than complicated regions in the higher-dimensional space of the original vector.

References

[1] Alvin, C. R. (2002). Methods of Multivariate Analysis. New York: John Wiley & Sons.

[2] Anderson, T. W. (1973). Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions (LDF): Discriminant Analysis and Applications. (T. Cacoullos, Ed.), pp. 17-35, Academic Press, New-York.

[3] Anderson, T. W. (2003). An Introduction to Multivariate Statistical Methods. New York: New ed. John Wiley & Sons.

[4] Anderson, T. W. and Bahadur, R. R. (1962). Classification into Two Multivariate Normal Distributions with Different Covariance Matrices. Annals of Mathematical Statistics, 33, pp. 420-431.

[5]   Awogbemi, C. A. (2019). Errors of Misclassification Associated with Edgeworth Series Distribution, Unpublished Ph. D. Dissertation, Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria.

[6]   Clumies-Ross, C. W. and Riffenburgh, R. H. (1960). Geometry and Linear Discrimination. Biometrika, 47, pp. 185-189.

[7]   Ekezie, D. D. and Onyeagu, S. I. (2013). Comparison of Seven Asymptotic Expansions for the Sample Linear Discriminant Function. Canadian Journal of Computations in Mathematics, Natural Sciences, Engineering and Medicine, 4 (1), pp. 93-115.

[8]   Fisher, F. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, 7, pp. 179-188.

[9]   Gilbert, E. S. (1969). The Effect of Unequal Variance-Covariance Matrices on Fisher's Linear Discriminant Function. Biometrics, 25, pp. 424-427.

[10]   John, I. M. (2015). Multivariate Statistics, Department of Statistics, University of Illinois, Urbana Champaign. URL: http://stat.istics.net/Multivariate.

[11]   John, S. (1965): Corrections to: "On Classification by Statistics, R and Z". Ann. Inst. Math., 7, pp. 113.

[12]   Kudo, A. (1959). The Classificatory Problem viewed as a Two-Decision Problem. Unpublished Masters' Thesis of the Faculty of Science, Kyushu University, 13, pp. 96-125.

[13]   Lachenbruch, P. A., Sneeringer, C. and Revo, L. T. (1973). Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-normality. Journal of Communication Statistics, 1, pp. 39-57.

[14]   Morrison, D. F. (2003). Multivariate Statistical Methods. London: McGraw-Hill Publishing.

[15]   Onyeagu, S. I. (2003). A First Course in Multivariate Statistical Analysis. Awka: Mega Concept.

[16]   Rao, C. R. (1954). A General Theory of Discrimination when the Information about Alternative Population Distribution is Based on Samples. Annals of Mathematical Statistics, 25, pp. 651- 670.

[17]   Siotani, M. (1975). Comparison of Two Procedures in Discriminant Analysis based on Anderson's W-Criterion and John-Kudo's Z-Criterion. Tech. Report No. 67, University Of Manitoba, Department of Statistics.

[18]   Wald, A. (1944). On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups. Ann. Math. Stat. 15, pp. 145-162.

[19]   Welch, B. L. (1939). Notes on Discriminant Functions. Biometrika, 31, pp. 218-220.

[20]   William, R. D. and Matthew, G. (1984). Multivariate Analysis, Methods and Applications. New York: John Wiley & Sons Inc.