Using Discriminatory Analysis to Determine the Factors Affected Eye

Haidar Raed Talib [1], Ahmed Razaq Abd*[2] and Ayad Habeeb Shimal [3]

[1] Department of Statics – Faculty of Administration and Economics – Sumer University – Iraq
[2] Department of Statics – Faculty of Administration and Economics – Wasit University – Iraq
[3] Department of Statics – Faculty of Administration and Economics – Diyala University – Iraq

*ahmedrazzaq@uowasit.edu.iq

Received: 25 September 2018 Accepted: 7 March 2019

Abstract

Discriminatory analysis is one of the most widely used statistical methods in the field of data analysis and classification, which can be used as linear models in the field of data classification. One of the basic conditions that must be available in the qualitative analysis is that the homogeneity of the matrix of variance and the common covariance and the test of the normal distribution of the data. We also note that the analysis of distinction is outlier of the same; Including the test (Kolmogorov Smirnov) and this test exists in the statistical program SPSS. The important objective of using the differential analysis method is to test the significance of differences between groups for independent variables as well as to differentiate the independent variables that share as much differences as possible between the dependent variable groups.

Keywords: Discriminatory analysis, Classification, Discriminate, Homogeneity of the matrix of variance–covariance, Discriminatory functions, Quadratic Discriminant Function, Linear discrimination function for two groups, Mixture Discriminant function.

استعمال التحليل التمييزي لتحديد اهم العوامل التي تؤثر على امراض العين

حيدر رائد طالب[1]، احمد رزاق عبد[2] و اياد حبيب شمال[3]

[1]قسم الإحصاء - كلية الإدارة والاقتصاد - جامعة سومر

[2]قسم الإحصاء - كلية الإدارة والاقتصاد - جامعة واسط

[3]قسم الإحصاء - كلية الإدارة والاقتصاد - جامعة ديالى

الخلاصة

يعتبر التحليل التمييزي من أكثر الطرق الإحصائية استخدامًا في مجال تحليل وتصنيف البيانات، والتي يمكن استخدامها كنماذج خطية في مجال تصنيف البيانات. أحد الشروط الأساسية التي يجب توافرها في التحليل التميزي هو أن تجانس مصفوفة التباين والتباين المشترك واختبار التوزيع الطبيعي للبيانات، نلاحظ أيضا أن تحليل التمييز في حد ذاته يعتمد على اختبار (Kolmogorov سميرنوف) وهذا الاختبار موجود في البرنامج الإحصائي SPSS. الهدف المهم من استخدام طريقة التحليل التغاير هو اختبار أهمية الاختلافات بين المجموعات للمتغيرات المستقلة وكذلك للتمييز بين المتغيرات المستقلة التي تشترك في أكبر قدر ممكن من الاختلافات بين مجموعات المتغيرات التابعة.

الكلمات المفتاحية: التحليل التمييزي، التصنيف، التميز، التجانس لمصفوفة التباين - التغاير، الدوال التمييزية، دالة التمييز التربيعية، وظيفة التمييز الخطي لمجموعتين، وظيفة تمييز المخلوط.

Introduction

Distinctive analysis is an important technique in multivariate statistical analysis where a set of variables is used to isolation two or more groups by means of discriminate functions.

These functions may be linear or quadratic. That is a linear combination of independent variables. This is based on the prediction or determination of independent variables that contribute significantly to the distinction between two groups or more and in a new single classification of one of the groups under study.

For example, if we have a sample of words in the classification (that is, groups are not defined), we need to separate the vocabulary into several groups on the basis of measured variables, often called cluster, so that the same group is homogeneous and different from the other totals. Specific measures and observational characteristics that must to some extent be consistent with the characteristics of the society to which they will be more closely aligned with any other society. In recent years, the use of taxonomy, particularly in the medical.

Practical importance

The importance of research is necessary and requires the application of scientific and statistical study on the health parts suffered by society, especially eye diseases need to conduct research and statistical studies to help the competent health personnel to identify the most important factors affecting the nature of eye diseases. It also helps governments to provide good health possibilities, which is a scientific contribution to solve the shortcomings experienced by health institutions.

Discriminatory Analysis

Discriminate analysis is one of the most important methods of multivariate analysis. It examines the possibility of distinguishing between two groups or more. Which are similar in many distinguishing after obtaining a sample of observations for each group and depending on a number of variables about of observations. We can distinctive function that will help distinguish some new observations of the correct group (set) as it is used in multiple range for example in medicine; agriculture; education; sociology, geography and other such problems called discrimination (Discrimination) or classification (Classification) is simple but different between them can be explained as [1, 2].

Discrimination

The differential analysis is used in the case of a sample of observations dating to two or more group (groups are defined in advance). A mathematical function or formula is required based on several variables around these observations. Through this function or mathematical formula, we can return any new observations to their correct group, with the lowest error possible [3,1].

Classification

The classification can be used to classify a sample of the observations (when the groups are not defined). It is possible to separate the observations into several groups based on the measured variables called cluster, so that the group is homogeneous and different from the other groups.

[4]. There are many functions of distinction or classification. The linear distinguish function is one of the simplest formulations of the discrimination function, which is presented by the expert Fisher. It is used in the case of equality of the matrix variance - covariance of all groups, and group expresses multivariate normal sets and a quadratic discriminant function [5].
Which are used in the case of the equality of the matrix variance -covariance differences of group as well as the logistic regression function used in the case of continuous, discrete or mixed variables and other models as well as non-parametric and robust methods. Each of these methods is evaluated by its advantage in predicting and classifying the new observations of one of the studied groups classification error possible [4,5].

Conditions of discriminatory analysis

After identifying the dependent variables and independent variables, it is necessary to ensure that the conditions for the analysis are met. Some conditions are checked either prior or before the analysis start. The discriminator gives results that we used to ascertain determent conditions. A number of conditions can be disregarded before the discriminatory analysis if the size of the sample is large, the conditions to be achieved before the discriminatory analysis can be abstract as follows [6,7]:

a. Sample size

There are several opinions about what the appropriate sample size is for the analysis First. suppose that the smallest groups comprise more than the number of independent variables, including that the total sample size should be at least ten times the most variables.
Second, explain: This method gives false results if the number of independent variables is maximal the number of group cases. The sample size is large for all analyzes of the analysis. The sample selection must be random and the degree of any observant sample dependent is independent [2,8].

b. Normal distribution test

Discriminatory analysis is easily affected outliers and they find that data distribution is distanced from normal distribution. In this study, used Kolmogorov Smirnov test, and Graphical methods such as histogram can also be used to prove normal distribution conditions, there are no outlier's values that are used in the Mahalanopisg test. The SPSS Statistical Program [2,6].

c. homogeneity of the matrix of variance –covariance:

In order to determine the homogeneity of two group, can be used Boxes -M. The result of the test is not significant to achieve the condition accept the null hypothesis that there is homogeneity, thus a significant level of 0.001 is used. There is also a log Dominants test. This test is also found in the statistical program SPSS [7,8].

The basic points of discriminatory analysis

A. One of the multivariate statistical methods used to study the range of sets overlap.

B. Depends on the basis of classification of observations between the sets under research and study.

C. Two or more sets that share one another shall be dealt with in a set of characteristics but separate from each other quantitatively [4,9].

There are also some important points in the method of discriminatory analysis:

- This is used for the method of discriminate between two or more groups depending on the comparison of independent variable values and can be used for any number of dependent variables and sets

-This is an extension method for multiple regression analysis and is therefore an advanced point for the use of statistical methods in analysis

-This method is more sensitive and more efficient than the logistic regression system depending on the error classification average [4,9].

Objective of using the discriminatory analysis method

1. Combine the best independent variables necessary to distinguish between variable group.
2. Examining the significance of differences between groups for independent variables.
3. Determine the independent variables that contribute as much as possible differences between the groups of the dependent variable [5,10].

Discriminatory functions

There are many discriminatory functions

a. Linear discriminating function of two groups

The function of discrimination is a model that can be explanation the randomly selected sample and placed in two different groups. By this function we can test observation and determine its return to any group, but if the following assumption [5, 6, 8, 9, 10 and 11].

1. The vector of independent variables should be a multivariate normal distribution
2. Equations variance studied sets (Matrices of variance-covariance), acceptance of the hypothesis of nullity when testing the hypothesis

   $\mathrm{H}_0 = \Sigma_1 = \Sigma_2 = \cdots = \Sigma_g$

   $\mathrm{H}_1 \neq \Sigma_1 \neq \Sigma_2 \neq \cdots \neq \Sigma_g$
3. The vectors of the means are different in each set. The application of robust estimator requires programs that provide us with discriminating discriminative functions robust, Thus Minimize the effect outlier on the estimates, estimating the probability of the classification error in a method robust gives robust results compared to the known robust methods in most research. Thus, the linear discriminating function of two groups is a linear structure based on several measures or observant of a selected sample this function can use to classify the index into the general sets and at the lowest possible classification. Assuming that we have two groups to be compared, under the condition that the two groups have a similar variance covariance and Different mean μ_1, μ_2 Assuming that we have a random sample of n1 size drawn from the first collection, a second random sample of n2 size drawn from the second collection, and that each vector contains p of the variables. Therefore, the discriminant function is a linear structure comprising (p) Works to maximize the difference between the two groups, and this function is [6, 8 and 9]:

$$Z = a_1x_1 + a_2x_2 + \cdots + a_px_p \qquad \textbf{(1)}$$

$$Z = \underline{a}' \underline{x}$$

This is the linear function of Fisher.

$\underline{a}$: Vector form parameters and used in the classification Process.

$\underline{x}$: Vector variables.

The maximization is when:

$$\alpha = S^{-1}(Z_1 - Z_2) \quad (2)$$

It is necessary to find the vector that produce the Q maximize value as possible.

$$Q = \frac{[\bar{Z}_1 - \bar{Z}_2]^2}{S_Z^2} = \frac{[\bar{Z}_1 - \bar{Z}_2]^2}{\sum_{i1}^{n1}[Z_{i1} - \bar{Z}_1]^2 + \sum_{i2}^{n2}[Z_{i2} - \bar{Z}_2]^2} \quad (3)$$

Q : Square Mahallonobis Distance

$\bar{Z}_1$: Median observations values of the first group

$\bar{Z}_2$ Median observations of the second group

Z_{i1} The value observations i in the first set

Z_{i2} The value observations i in the second set

Then find $\bar{Z}_1$ $\bar{Z}_2$

$$\bar{Z}_1 = \underline{a}'\underline{\bar{x}}_1 = a_1\bar{x}_{11} + a_2\bar{x}_{21} + \cdots + a_p\bar{x}_{p1}$$

$$\bar{Z}_2 = \underline{a}'\underline{\bar{x}}_2 = a_1\bar{x}_{12} + a_2\bar{x}_{22} + \cdots + a_p\bar{x}_{p2} \quad (4)$$

$$[\bar{Z}_1 - \bar{Z}_2]^2 = [\underline{a}'\left(\underline{\bar{x}}_1 - \underline{\bar{x}}_2\right)]^2 = \underline{a}'\left(\underline{\bar{x}}_1 - \underline{\bar{x}}_2\right)'\left(\underline{\bar{x}}_1 - \underline{\bar{x}}_2\right)\underline{a} \quad (5)$$

$$\sum_{i=1}^{n1}\left[Z_{i1} - \bar{Z}_1\right]^2 = \underline{a}'[\sum_{i=1}^{n1}(Z_{i1} - \bar{Z}_1)'(Z_{i1} - \bar{Z}_1)]\underline{a} = \underline{a}'(n_1 - 1)S_1\underline{a} \quad (6)$$

$$\sum_{i=1}^{n2}\left[Z_{i2} - \bar{Z}_2\right]^2 = \underline{a}'[\sum_{i=1}^{n2}(Z_{i2} - \bar{Z}_2)'(Z_{i2} - \bar{Z}_2)]\underline{a} = \underline{a}'(n_2 - 1)S_2\underline{a} \quad (7)$$

S_1 , S_2 Matrix of variance and estimated variance within each of the first and second sets. Find the differences within the groups.

$$\sum_{i=1}^{n1}\left[Z_{i1} - \bar{Z}_1\right]^2 + \sum_{i=1}^{n2}\left[Z_{i2} - \bar{Z}_2\right]^2 = \underline{a}'(n_1 - 1)S_1\underline{a} + \underline{a}'(n_2 - 1)S_2\underline{a} = (n_1 + n_2 - 2)\underline{a}'S\underline{a} \quad (8)$$

Assuming that S is the variance –covariance matrix within the two sample

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)}$$

$$S(n_1 + n_2 - 2) = (n_1 - 1)S_1 + (n_2 - 1)S_2$$

find the value of Q, we compensate in equation (2-3)

$$Q = \left[\frac{\underline{a}'(\underline{\bar{x}}_1-\underline{\bar{x}}_2)'(\underline{\bar{x}}_1-\underline{\bar{x}}_2)\underline{a}}{(n_1+n_2-2)\underline{a}'S\underline{a}}\right] = \frac{1}{n_1+n_2-2} * \frac{\underline{a}'(\underline{\bar{x}}_1-\underline{\bar{x}}_2)'(\underline{\bar{x}}_1-\underline{\bar{x}}_2)\underline{a}}{\underline{a}'S\underline{a}} = \frac{1}{n_1+n_2-2}\, Q^* \quad (9)$$

Is the ration to be maximize: Q^{*}:

The condition can therefore be settled $\underline{a}'S\underline{a} = 1$.

That is, the variance within the groups is constant and we maximize the variance between the groups $\underline{a}'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)\underline{a}$ to be used by the Lagrange multiplier in the function and it becomes as follows:

$$\mathcal{L}(\underline{a},\lambda) = \underline{a}'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)\underline{a} - \lambda(\underline{a}'S\underline{a} - 1) \quad (10)$$

$$\frac{\partial\mathcal{L}}{\partial\underline{a}} = \underline{a}'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right) = \lambda S\underline{a} \quad (11)$$

After the calculations, we find value.

$$\lambda = \frac{\underline{a}'(\underline{\bar{x}}_1-\underline{\bar{x}}_2)'(\underline{\bar{x}}_1-\underline{\bar{x}}_2)\underline{a}}{\underline{a}'S\underline{a}} \quad (12)$$

$$\lambda = \left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)' S^{-1}\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right) = D^2 \quad (13)$$

Mahalunobis distance, a measure of distances between group centers D^2

$$where: \underline{a}'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right) = \left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)D^2 \quad (14)$$

$$\therefore \underline{a}'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)'\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right) = \lambda S\underline{a}$$

$$\therefore \lambda S\underline{a} = \left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)D^2$$

$$\underline{a} = S^{-1}\left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right) \quad (15)$$

The linear discriminating function (Fischer) is in the form

$$Z = \underline{a}'\underline{x} = \left(\underline{\bar{x}}_1-\underline{\bar{x}}_2\right)' S^{-1}\underline{x}$$

DOI: https://dx.doi.org/10.24237/djps.15.03.491C
P-ISSN: 2222-8373
E-ISSN: 2518-9255

After this, the significance of the linear discriminating function is tested. After the extraction of the coefficients, attention is focused on the robust of the linear discriminating function and it distinguish. This can be determined by the analysis of variance,

Sum squares between groups equals

$$Between\ S\,.S = \left(\frac{n_1 n_2}{n_1 + n_2}\right)[D^2]^2$$

K. The degree of freedom

Sum squares within groups

Within s. s $=D^2$

The degree of freedom (n_I+n_2-k-1)

Mean squares between groups

$$Mean\ S.S = \frac{\left(\frac{n_1 n_2}{n_1+n_2}\right)[D^2]^2/K}{D^2/(n_1+n_2-K-1)} \qquad \textbf{(15)}$$

The latter ratio follows the distribution of F with freedom degree K and (n1 + n2-k-1). When this is significant, this reference that the function is correct and that the variables used give the best distinction between the two groups.

To extract the linear discrimination function in the classification of observations belonging to unknown sets, we are the first step is to find the median values of the distinguishing function of the two samples as follows:

$$\overline{Z}_i = \left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)' S^{-1}\underline{\overline{x}}_i \quad i = 1,2$$

Then find the average mean linear discrimination function

$$\overline{Z}_1 = \left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)' S^{-1}\underline{\overline{x}}_1$$

$$\overline{Z}_2 = \left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)' S^{-1}\underline{\overline{x}}_2$$

$$C = \frac{\overline{Z}_1+\overline{Z}_2}{2} = \frac{\left(\underline{\overline{x}}_1-\underline{\overline{x}}_2\right)' S^{-1}\left(\underline{\overline{x}}_1-\underline{\overline{x}}_2\right)}{2} \qquad \textbf{(16)}$$

C: Cut Point

The second step is to form a classification base classification for the first sets

$$\overline{Z}_1 = \left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)' S^{-1}\underline{\overline{x}}_1 > \frac{\left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)' S^{-1}\left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)}{2} \quad \textbf{(17)}$$

And to the second sets:

$$\overline{Z}_2 = \left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)' S^{-1}\underline{\overline{x}}_2 < \frac{\left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)' S^{-1}\left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)}{2} \quad \textbf{(18)}$$

Another distinguishing function of Rao is used in the case of two or more groups, since the Fisher function is only for two groups, and Figure (1) explains the classification steps. [5][6][9][10][11]

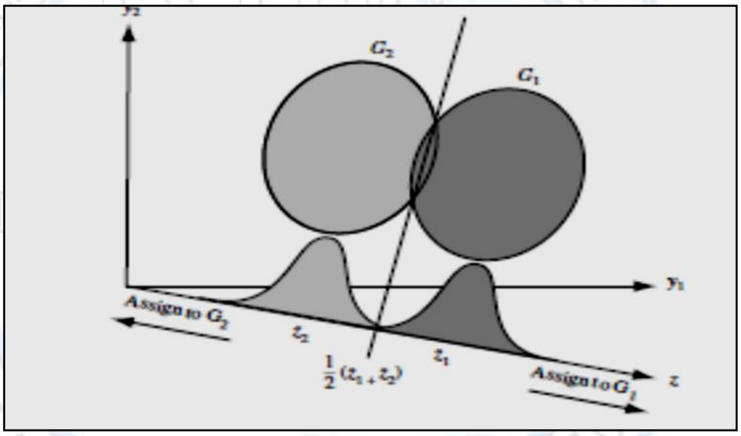

Figure 1: Classification steps

b. Quadratic Discriminant Function

This function is used in the case of data in which the conditions of linear discrimination assume that the independent variables (X's) are normal distributed by a variance-covariance matrix. If the matrix of variance covariance of groups is not equal, the R is as follows [11]:

$$R = \frac{f_1(x_i)}{f_2(x_j)} \neq 1 \quad \textbf{(19)}$$

Where (X) is observant first group if (1) < (R), and the second group if (1) (R By taking the normal logarithm of the two sides we get the following equation [5,9].

$$T = LnR = Ln f_1(x_i) - Ln f_2(x_j) \neq 1 \quad \textbf{(20)}$$

If there are (P) variables for each (x_1, x_2… x_p) of the two groups, the measure of discrimination is:

$$T = Lnf_1(x_1, x_2, \dots, x_p) - Lnf_2(x_1, x_2, \dots, x_p) \quad \textbf{(21)}$$

the parameters of the probabilistic function are not known, the measure of discrimination as follows:

$$g = T = -\frac{1}{2} Ln|S| - \frac{1}{2}(\underline{x} - \mu)' S^{-1}(\underline{x} - \mu) \quad \textbf{(22)}$$

$$g = T = -\frac{1}{2} Ln|S| - \frac{1}{2}\underline{x}' S^{-1}\underline{x} + \underline{\mu}' S^{-1}\underline{x} - \frac{1}{2}\underline{\mu}' S^{-1}\underline{\mu} \quad \textbf{(23)}$$

g; Discriminant faction (T) were estimated by estimating both the variances for method (M.L.E)

μ: : Mean

$\underline{X}$: Vector variables

The new observations will be belonging to the first group if (g> 0) and the second group if (g<=0)

C. Mixture Discriminant Function (MDF)

One of the methods that is an extension to QDF, LDF, is based on a mixture of multivariate normal distributions. The advantage of this method is to study relation between different groups, but it can also provide us with a new classification base for one of the groups under consideration. Is suitable when single samples are taken from several different sets and each observation has P of the scales that suppose that the normal distribution of the multivariate [6,10].

$$\rho_{in} = \frac{\hat{T}_n f_n(\frac{x_i}{\hat{\mu}_n}, \Sigma n)}{\sum_{J=1}^{G} \hat{T}_j f_j(\frac{x_i}{\hat{\mu}_j}, \Sigma j)} \quad \textbf{(24)}$$

n = 1, 2 j = 1, 2   i = 1, 2

$P_{in:}$ Is the probability value of the function (MDF)

$\hat{T}_n$: Is the probability that the observation is belonging to group n.

$\hat{\mu}_n$: Estimated mean of group n.

$\sum n$:Variance- covariance estimated of group n.

n: represents the number of groups

j: Number of cases

The mixed distinctive function is calculated in two steps:

The first step: (Probability Models for Cluster analysis)

$f_n(\frac{x_i}{\hat{\mu}_n}, \sum n)$

The probability density function is calculated through a mixture of probability distributions in which each complex represents a different set of samples$\underline{x} = x_1 x_2 \dots x_p$ observations

Let the density function of the xi of the set represent the parameters and let G the number of complexes in the mixture and the model of the group of groups using the Likelihood function

$$LC\ (\theta_1\ \dots\ \theta_G\ ;\ Y_1\ \dots\ Y_G \setminus X\ ) = \prod_{i=1}^{m} f_{y_i}\left(x_i \setminus \theta_{y_i}\right) \quad (27)$$

(yi) are discreet values called classifications

$$L_i\ (\theta_1\ \dots\ \theta_G\ ;\ T_G \setminus x\ ) = \prod_{i=1}^{m} \prod_{n=1}^{G} T_n f_n\ (x_i \setminus \theta_n) \quad (28)$$

y_i = n If xi belongs to the nth group, the probability of the mixture maximum, and important mainly with the case $f_n(x_i \setminus \theta_n)$where the normal model is multivariate. In this case, the parameters θ_n are composed of theμ_n and matrix variance –covariance$\sum_n$. The probability density function defines.

$$f_n\left(x_i \setminus \mu_n, \sum_n\right) = \frac{\exp[-\frac{1}{2}(x_i - m_n)' \sum_n^{-1}(x_i - m_n)}{(2\pi)^{\frac{p}{2}} |\sum_n|^{\frac{1}{2}}} \quad (29)$$

n=1, 2

Second step: The researcher found the parameters of the variance model and the mean through the algorithm [Expectation - Maximization] is an important method to find parameters of the mixed model and is in two steps [5,9]:

1. By (Expectation) through which the matrix W is found by win estimated by conditional probabilities

$$win = \begin{cases} 1\ if\ x_i\ \ belong\ to\ group\ n \\ 0\quad otherwise \end{cases}$$

$$E(win) = p(n \setminus x_i) = \frac{p(x_i \setminus n)}{p(x_i)} = \frac{p(x_i \setminus \mu_n) p(n)}{\sum_{j=1}^{n} p(x_i \setminus \mu_j) p(j)} \quad (28)$$

$$= \frac{\exp[-\frac{1}{2}(x_i - m_n)' \Sigma_n^{-1} (x_i - m_n)}{\sum_{j=1}^{G} \exp[-\frac{1}{2}(x_i - m_j)' \Sigma_j^{-1} (x_i - m_j)} \quad \textbf{(29)}$$

P (n) is deleted with P (i) when the previous probability of both groups is equal in the theory of the Bayes in the case of normal distribution.

2. Maximization of mixed model parameters are estimated

$$p(S \backslash \mu_n, win\ for\ n = 1,2, \dots, N\ and\ i = 1,2, \dots m) = \prod_{i=1}^{m} \sum_{n=1}^{N} \frac{win}{\sqrt{2\pi \Sigma_n}} exp \sum_{n=1}^{N} \text{win} \left[-\frac{1}{2}(x_i - m_n)' \Sigma_n^{-1} (x_i - m_n) \right] \quad \textbf{(30)}$$

$$= \prod_{i=1}^{m} \frac{win}{\sqrt{2\pi \Sigma_n}} exp \sum_{n=1}^{N} win[-\frac{1}{2}(x_i - m_n)' \Sigma_n^{-1} (x_i - m_n)] \quad \textbf{(31)}$$

$$= \sum_{i=1}^{m} \sum_{n=1}^{N} Win\ (x_i - \mu_n)'(x_i - \mu_n) \quad \textbf{(32)}$$

$L_n(\mu_n$, Win for n = 1, ….. N and i=1, …., m)

Random variable depends on to the distribution that generates S of the data, in which case we must find its expected value as long

Ln (μ_n , Win for n=1, …. N and i=1, …,.m)

Is a linear function for Win?

$E[Ln(\mu_n, win\ for\ n = 1,2,3, \dots, N\ and\ i = 1,2,3, \dots m)]$

$$\text{E} = \sum_{i=1}^{m} \sum_{n=1}^{N} \text{Win}\ (x_i - \mu_n)'(x_i - \mu_n) \quad \textbf{(33)}$$

$$= \sum_{i=1}^{m} \sum_{n=1}^{N} \text{E}(\text{Win})(x_i - \mu_n)'(x_i - \mu_n) \quad \textbf{(34)}$$

$$\frac{\partial}{\partial \mu_n} E[Ln(\mu_n, win\ for\ n = 1,2,3, \dots, N\ and\ i = 1,2,3, \dots m)]$$

$$0 = -2 \sum_{i=1}^{m} E(win)(x_i - \mu_n) \quad \textbf{(35)}$$

$$\mu_n = \frac{\sum_{i=1}^{m} E\ (win) x_i}{\sum_{i=1}^{m} E\ (win)} \quad \textbf{(34)}$$

In the same way, by applying the expectation-maximization algorithm we can find the variance

$$\Sigma_n = \frac{\sum_{i=1}^{m} E(win)[(x_i - \mu_n)'(x_i - \mu_n)}{\sum_{i=1}^{m} E(win)} \quad \textbf{(35)}$$

n = 1, 2

$\hat{T}_n$Calculate through

$$\hat{T}_n = \frac{\sum_{i=1}^{m} E\,(win)}{n_1 + n_2} \quad \textbf{(36)}$$

Find the value of difference d1 between the probability values of groups d = Pi1-Pi2 and the new observations will be classified as belonging to the first group if the value of d1> o and the second group if the value of d1 <o is smaller than zero and random otherwise [8,10].

Preface

In the study, the researcher is interested in studying the importance of the human eye and its role in the human body. The eye is the most sensitive members of the other members of the body anatomical.

The individual's daily behavior and behavior may cause damage to the eye, but the greatest risk is eye inflammation. Care must be taken because the eye is the organ that can diagnose diseases before they feel from the veins and arteries. Eye infections in this study are classified into three categories (corneal injury, macular degeneration, external sting).

Experiments and Results

1. Description of the research sample

A random sample of (105), (50), and (45) patients represent three groups of patients with ophthalmitis were withdrawn from Ibn al - Haytham Eye Hospital in Baghdad. The first group represent patients with corneal disease and group 2 patients with macular degeneration and the third group patients are infected with external sting the three variables were considered as a dependent variable and were (0), (1) and (2). The independent variables were 7 variables (sex, age, eye pressure, blood glucose level, blood urea ratio, vision level, sight.)

2. Statistical analysis of the discriminatory function

a. Test the discriminatory function

Before determining the discriminatory functions of the three groups of ophthalmic diseases, it is necessary to conduct some tests that meet the normal distribution condition, the significance of the discriminating function and the homogeneity of data between the three groups

Given that the sample size skip 30 observation, data can be considered to be distributed close to normal distribution according to the central limit theory.

1. Test the significance of the linear discriminating function

When the distinction between the three groups or more is to be tested and the composition of statistically significant discriminative functions is statistically significant, the significance of the differences between the averages of the groups studied is determined by the following hypothesis.

$\boldsymbol{H_0: \mu_1 = \mu_2}$

$\boldsymbol{H_1: \mu_1 \neq \mu_2}$

There is different measure to test the hypothesis above and use the test tester (walks') as shown

$$V = \frac{|w|}{|w+B|} = \frac{|w|}{|T|} \qquad \textbf{(39)}$$

W: Matrix of variance –covariance in groups

T: Matrix of variance and total variance of groups

B: Matrix of variance –covariance between groups

The equation (39) is approximately distribution (x^2) with a free degree (k-1) and a significant level (α).

The differences between the mean of the three groups of ophthalmic diseases tested were by its equation (39). The results of this test were shown in table 1.

Table 1: The significant test of the discriminatory function

Test functions	Wilks Lamda	Chi-Square	df	Sig.
1 through (2)	0.856	45.681	14	0.000
2	0.968	9.745	6	0.013

The results showed that there are significant differences between the three means. This means that there are significant differences between the average of patients with corneal disease and patients with degenerative disease and those with external disease. This means that the discriminatory functions distinguish between the groups, observation to one of the three groups.

2. To test the homogeneity of the differences between the three groups (corneal injury, macular degeneration, external sting)

To apply the linear discrimination function first, equalize the variance -covariance matrices. Second, the use of the Bartlett test is one of the tests used to sureness the heterogeneity requirement through which the hypothesis of this test is tested:

$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 \quad V.S \quad H_1: \Sigma_1 \neq \Sigma_2 \neq \Sigma_3$

The equation of his statistics is the tests

$$\mu = \left(\sum_{i=1}^{k} V_i\right) \ln|S| - \sum_{i=1}^{k}\left(V_i \ln|S_p|\right) \quad (40)$$

$$S_p = \frac{1}{\sum_{i=1}^{k} V_i} \sum_{i=1}^{k} V_i S_i \quad (41)$$

S_p :Estimated variance –covariance Matrix (pooled covariance).

S_i :Sample variance ($i = 1, \dots\dots, k$).

$V_i = n_i - 1$:The degree of freedom of the sample i.

k :Number of a sets.

Box (1949) has shown that if multiplying μ in constant $C^{(-1)}$ equal

$$C^{-1} = 1 - \frac{2p^2+3P-1}{6(p+1)(k-1)}\left[\sum_{i=1}^{k}\frac{1}{V_i} - \frac{1}{\sum_{i=1}^{k} V_i}\right] \quad (42)$$

p: Number of independent variables.

a scale is obtained that approximates the distribution of x^2 and the degree of freedom of 1/2 (k-1) p (p + 1) where

$$\mu c^{-1} \sim x^2\left[\frac{1}{2}(k-1)p(p+1)\right] \quad (43)$$

According to this test, the results were obtained in table 2.

Table 2: Test homogeneity of the differences between the three groups

Box's M	76.460
F Approx.	1.309
Df1	56
Sig	0.060

In table 2, the results showed that the value of P-Value> 0.05 reference acceptance of the null hypothesis. This reference the homogeneity of the differences between the three groups.

b. Test the significance of the variables in the discriminatory function

The significance of the variables was tested to determine the importance of each variable i and its effect on the formation of the linear discriminate function the results were as in table 3.

Table 3: significance test of discriminatory function variables

Variables	Wills' Lamda	F	df1	df2	Sig
(x_1)gender	1.000	0.016	2	197	0.984
(x_2)age	0.944	8.870	2	197	0.000
(x_3)eye pressure	0.975	3.874	2	197	0.022
(x_4)Diabetes level disease	0.997	7.741	2	197	0.001
(x_5)Urea ratio	0.996	0.564	2	197	0.570
(x_6)Level of vision	0.962	5.799	2	197	0.003
(x_7)optic nerve	0.982	2.653	2	197	0.072

Table 3 shows, the variables of the study (eye pressure, age, Level of vision, Diabetes level disease) have an important effect on the formation and construction of the linear discriminating function, and both gender, urea, and optic nerve status have no importance in the function.

3. Estimation of linear discriminatory functions

The linear discriminative functions of the three groups of ophthalmic patients were estimated after estimating the mean of the three sample samples and the matrix of variance covariance using the maximum likelihood method as in table 3.

Table 4: Estimated Linear Distinguished Functions

Variables	Function1	Function2	Function3
Constant	-20.206	-20.057	-22.005
(x_1)gender	1.092	1.288	1.112
(x_2)age	5.109	4.365	5.069
(x_3)eye pressure	3.092	2.698	3.296
(x_4)Diabetes level disease	2.728	2.816	2.953
(x_5)Urea ratio	1.721	1.679	1.375
(x_6)Level of vision	2.060	2.582	2.444
(x_7)optic nerve	1.557	2.022	2.066

Linear discriminating functions can be written using a table 4.

$$d_0^*(\underline{x}) = 1.09X_{01} + 5.10X_{02} + 3.09X_{03} + 2.72X_{04} + 1.72X_{05} + 2.06X_{06} + 1.55X_{07} - 20.20 \quad \textbf{(44)}$$

$$d_1^*(\underline{x}) = 1.28X_{11} + 4.36X_{12} + 2.69X_{13} + 2.81X_{14} + 1.67X_{15} + 2.58X_{16} + 4.02X_{17} - 20.05 \quad \textbf{(45)}$$

$$d_2^*(\underline{x}) = 1.11X_{21} + 5.06X_{22} + 3.29X_{23} + 2.95X_{24} + 1.37X_{25} + 2.44X_{26} + 4.06X_{27} - 20.005 \quad \textbf{(46)}$$

The values of the equations, the explained variables, and the linear functions discriminating, through the three functions of linear discriminate, the discriminating importance of the variable of age, variable eye pressure, variable Diabetes level, variable vision level, variable urea ratio. The development of the optic nerve and finally the gender variable in the diagnosis of the three diseases of eye diseases in all functions of linear discriminate.

3. Classification accept to the probability equation

In order to classify any observation based on the estimated discriminatory functions in (44) (45) (46) we replace the values of the observation variables in the given equations, $d_0^*(\underline{x}) > d_1^*(\underline{x}), d_2^*(\underline{x})$ The first observation is classified for the group (0) if$d_1^*(\underline{x}) > d_0^*(\underline{x}), d_2^*(\underline{x})$ They are classified as the second group (1) or $d_2^*(\underline{x}) > d_0^*(\underline{x}), d_1^*(\underline{x})$ They are classified as the third group.(2)

There is a type of error known as a misclassification, which is the probability of classifying an observant random variable into the first group while really returning to the second or third group and vice versa. There are two types of errors, the first known as the apparent error and the second the ration real error. In this study, the first type of error was depending for all three sample observations. The following table represents the classification to the three linear discrimination functions.

Table 5: The three linear discrimination functions of classification of observations

Classification				
correct r classification ratio	patient returns to the third group (2)	patient returning to the second group (1)	patient returning to the first group (0)	case
49.6 %	24	23	46	patient returning to the first group (0)
50.8 %	13	49	24	patient returning to the second group (1)
47.0 %	14	4	3	patient returns to the third group (2)
49.6 %	27.0 %	35.4 %	37.6 %	total classification ratio

The results of table 5 showed that the probability of the correct classification of a patient is from the first group (0) and was 49.6%, while the probability of error of classification was 50.8%. The results the probability of correct classification of a patient dependent to the second group (1) was 50.8% (49.2%), while the probability of correct classification of a patient dependent to the third group (2) was 47% and the probability of error of classification 53%. The total correct classification ration 49.6% while the total error ration 51.4%

Conclusions

1. After the test of the significance of the discrimination for the three groups, the value of the wilks of the first group is zero and robust discrimination of the function, While the value of the wilks of the second and third group is off from zero reference that the discrimination power of the function is not robust.
2. The results of the function of linear discrimination that some of the variables of the study have an effect and importance in the construction and composition of the function of linear discrimination and found that a set of independent variables have an effect on the dependent variable where the results have been seen previously.
3. The linear discrimination function gave the maximum error rate of 50.8% for the second group (1) according equations in table 5.

Recommendations

1. We recommend using other (squared-linear-mixed) functions in the classification of multi-response data.
2. It is preferable to compare the probability of the classification line according to the probability of response.
3. Conduct similar studies periodically to estimate the effect of plans on the incidence of eye disease.

References

1. T. H. Al-Baldawi, Analytical Comparison between Models of Logistics and Models of Discriminatory Functions. Ph.D. Thesis, University of Baghdad, Baghdad, Iraq, (1996).
2. I. M. Ibrahim, Comparison between Linear Regression Model and Distinguished Function Model, M.Sc. Thesis, University of Mustansiriya, Baghdad, Iraq, (2011).
3. S. S. Mohammed, using a descriptive analysis to determine the most important factors affecting the deposition and repetition of students at all levels, M.Sc. Thesis, University of Sulaymaniyah, Sulaymaniyah, Iraq, (2003).
4. G. D. Al-Thabhawee, A Comparison between Discriminant Analysis and logistic regression on the classification of cancer Patients, M.Sc. Thesis, University of Kufa, Najaf, Iraq, (2012).
5. S.O. Fawzy, The use of the discriminatory function in qualitative control with the application of births of children infants, M.Sc. Thesis, University of Mosul, Mosul, Iraq, (2004).
6. D. H. AL-Yasin, Using Certain Methods of robust Discrimination to representation Leukemia, M.Sc. Thesis, University of Mustansiriya, Baghdad, Iraq, (2004).
7. A. Al-karkhi, A. Ease, Journal of Sustainable Development ,1(2),102 (2008).
8. H. Bernstein, P. Calvin, K. Gary, Applied multivariate analysis,1st ed. (Springer, New York,1988), pp.247 -250.
9. C. Alvin, Methods of Multivariate Analysis, 2nd ed. (John Wiley and Sons Publication, Brigham,2002), pp. 270- 298.
10. A.F. Alkarkhi, A. M. Easa, Journal of sustainable development, 1(2),102-106(2008).
11. N. D. Abdel-Karim, Using Statistical Discriminatory Methods for Diagnosing Some Heart Diseases, M.Sc. Thesis, University of Mustansiriya, Baghdad, Iraq, (2006).