



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Discrimination and Classification Levels of Students as Per their Major Specialization

Ali Mahmud Atewi, Sadoon Abdullah Al-Obaidy and Iryna Volodymyrivna Komashynska
Department of Mathematics and Statistics, Faculty of Science,
Al-Hussein Bin Talal University, P.O. Box (20), Ma'an-Jordan

Abstract: This study is considered as an attempt for employing the discriminated analysis method and classification for the purpose of achieving the assessment of a discriminant function through which we can discover the reasons of the actual difference between two groups of students-graduates of mathematics and statistics section at Al-Hussein Bin Talal University in accordance with their major specialization. Fisher's Linear Discriminant Function was used as a tool for statistical analysis. It was estimated on the basis of a sample of graduated students consisting (28) male and female students who were classified into two groups based upon their accumulative averages which reflect their general specialization. The study results showed high significant differences between both groups based on the T^2 test, this is ensure the existence of real differences among the students mental levels and their mental and innovative abilities for discriminating those who are distinguished and provide them with good care. Finally, the study results showed that the adopted classification rule led to the classification of about 90% of the students within the actual group to which they belong and the classification of about 10% of the students in a group which they do not belong to. These results assert the efficiency of the discriminant function which we have reached and the possibility of its use for the purpose of discriminating and classifying the students (unknown or affiliation) into the correct group in future.

Key words: Fisher's linear discriminant function, multivariate, analysis, hotelling's T^2 test

INTRODUCTION

The discriminant analysis and classification are considered as multivariate analysis concerning the discrimination of different groups of observations and classifying the new observations into predefined groups. We resort to such method, when it is difficult to understand the causal relation of the difference among the different groups in a sufficient degree.

This study aims to use the discriminant analysis method and classification for achieving an estimation of discriminant function which can be used for separating or discriminating two groups of students-graduates of Mathematics and Statistics Department at Al. Hussein Bin Talal University, who were already classified on the basis of their accumulative averages which reflect their general specialization for the purpose of reclassifying them according to their major specialization, through the selection 5 basic study subjects. It also aims to identify the arrangement scale of the relative importance of these five subjects and to reach a rule which can be used for future discrimination and classification of a new students in one group of both specified groups.

Study methodology: This study aims to the Fisher's Linear Discriminant Function as a tool for statistical analysis.

Study sample: The researchers selected a random sample of students-graduates of mathematics and statistics department consisting of 28 male and female students who were classified into two groups, based upon their accumulative averages which reflect their general specialization. The first group included 10 students whose averages were Very good and Excellent the second group included 18 male and female students whose averages were Good. Five basis study subjects were selected, reflecting the major specialization down. These subjects were: Topology (1), Real Analysis (1), Partial Differential Equations, Mathematics Statistics (1) and Abstract Algebra (1).

Study hypothesis

- The study assumes the existence of significant difference between both groups of students so as the discriminant function which we aspire to achieve will

be useful for discrimination and classification. We can realize this hypothesis by using Hotelling's T² test or Mahalanobis D² test.

- The study assumes that major specialization subject are different in respect with their relative importance. This can be assured by specifying the relative importance scales of these subjects.
- The study assumes that the adopted classification rule leads to classification errors which can be realized by using the Actual Error Rate (AER) or the Appeared Error Rate (APER).

Study plan: This study has been assessed on three main researches: theoretical section, applied section and abstract and conclusions.

THEORETICAL SECTION

Through this study, we'll discuss the theoretical bases of discrimination functions in case of two population existing. Certainly, our concern will focus on: Fisher's Linear Discriminant Function due to its theoretical formation, its estimation method and the test of significance of the separation between two populations in addition to the methods of evaluating the performance of classification functions.

Discrimination and classification of two populations: Fisher's Method: Suppose π_1 and π_2 are the two groups which are required to be discriminated and classified on the bases of scales taken from p of random variables:

$$X' = [X_1, X_2, \dots, X_n] \tag{1}$$

Suppose that X₁ represents the population of X values of the group π_1 with mean vector μ_1 and covariance matrix \sum_1 and suppose that X₂ represents the population of X values of the group π_2 with the mean vector μ_2 and covariance matrix \sum_2 . We now assume that f₁(x) and f₂(x) are multivariate normal densities, the first with mean vector π_1 and covariance matrix \sum_1 and the second with mean vector π_2 and covariance matrix \sum_2 . And also assume that $\sum_1 = \sum_2 = \sum$.

That is:

$$X_i \sim N_p(\mu_i, \sum), \quad i = 1, 2$$

As

$$\begin{aligned} \mu_i &= E(X \setminus \pi_i), & i &= 1, 2 \\ \sum &= E(X - \mu_i)(X - \mu_i)', & i &= 1, 2 \end{aligned}$$

Fisher's idea was to transform the multivariate observations X to univariate observations Y, such that the Y derived from populations π_1 and π_2 were separated as much as possible.

We suppose that μ_{1Y} refers to the average of values Y which we get from X values belonging to π_1 and μ_{2Y} refers to the average of values Y which we get from X values belonging to π_2 . That is:

$$\mu_{1Y} = E(Y \setminus \pi_1), \quad \mu_{2Y} = E(Y \setminus \pi_2)$$

Fisher proposed taking linear combinations from X for getting Y values from both populations π_1 and π_2 . Krzanowski (1977) selected the following Linear Combination:

$$Y = \alpha' X \tag{2}$$

(1 × 1) (1 × p) (p × 1)

Which maximizes the distance square between μ_{1Y} and μ_{2Y} ascribed to Y variance as follows:

$$\frac{(\mu_{1Y} - \mu_{2Y})^2}{\sigma_Y^2} \tag{3}$$

As:

$$\mu_{1Y} = E(Y \setminus \pi_1) = \alpha' \mu_1 \tag{4}$$

$$\mu_{2Y} = E(Y \setminus \pi_2) = \alpha' \mu_2 \tag{5}$$

$$\sigma_Y^2 = \text{var}(\alpha' X) = \alpha' \text{cov}(X) \alpha = \alpha' \sum \alpha \tag{6}$$

By compensating (4-6), in (3), we get

$$\frac{(\mu_{1Y} - \mu_{2Y})^2}{\sigma_Y^2} = \frac{\alpha'(\mu_1 - \mu_2)(\mu_1 - \mu_2)' \alpha}{\alpha' \sum \alpha} = \frac{(\alpha' \delta)^2}{\alpha' \sum \alpha} \tag{7}$$

where is

$$\delta = \mu_1 - \mu_2$$

By selecting:

$$\alpha' = c \sum^{-1} \delta = c \sum^{-1} (\mu_1 - \mu_2) \tag{8}$$

According to Johnson and Wichrn (1998) for any value c ≠ 0. selecting c = 1, we get the following linear combination:

$$Y = \alpha' X = (\mu_1 - \mu_2)' \sum^{-1} X = \delta' \sum^{-1} X \tag{9}$$

The Eq. 9 is known as: Fisher's Linear Discriminant Function.

The maximum of the ratio in Eq. 7 is:

$$\text{Max}_{\alpha} = \delta' \sum^{-1} \delta \tag{10}$$

From (Hand, 1981) it follows that the Eq. 9 can be used to classify new observations.

If $Y_0 = (\mu_1 - \mu_2)' \sum_{i=1}^{n_1} X_{0i}$ $S_{pooled} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$ (16)

is the value of the discriminant function at a new observation X_0 and

$$Mp = \frac{1}{2}(\mu_y + Y_{2y}) = \frac{1}{2}\alpha'(\mu_1 + \mu_2) = \frac{1}{2}(\mu_1 - \mu_2)' \sum (\mu_1 + \mu_2) \quad (11)$$

is the midpoint between both averages of the two populations of one variable, Whereas, it can be proved that:

$$\begin{aligned} E(Y_0 | \pi_1) - Mp &\geq 0 \\ E(Y_0 | \pi_2) - Mp &< 0 \end{aligned}$$

Accordingly, the classification rule is as follows: Allocate X_0 to π_1 if:

$$Y_0 = (\mu_1 - \mu_2)' \sum_{i=1}^{n_1} X_{0i} \geq Mp \quad (12)$$

And allocate X_0 to π_2 if:

$$Y_0 = (\mu_1 - \mu_2)' \sum_{i=1}^{n_1} X_{0i} < Mp \quad (13)$$

Estimation of Fisher's linear discriminant function:

Suppose that we have n_1 observations of the multivariate random variable: $X' = [X_1, X_2, \dots, X_n]$ from π_1 and μ_2 measurements of this quantity from π_2 , with $n_1 \neq n_2 \geq p$. Then the respective data matrices are:

$$\begin{aligned} X_1 &= [X_{11}, X_{12}, \dots, X_{1n_1}] \\ X_2 &= [X_{21}, X_{22}, \dots, X_{2n_2}] \end{aligned} \quad (14)$$

From these data matrices, the sample mean vectors and covariance matrices are determined by:

$$\begin{aligned} \bar{X}_1 &= \frac{1}{n_1} \sum_{j=1}^{n_1} X_{1j}, S_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)(X_{1j} - \bar{X}_1)' \\ \bar{X}_2 &= \frac{1}{n_2} \sum_{j=1}^{n_2} X_{2j}, S_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)(X_{2j} - \bar{X}_2)' \end{aligned} \quad (15)$$

As a result of the assumption $\sum_1 = \sum_2 = \sum$, the sample covariance matrices S_1 and S_2 are combined (pooled) to derive a single unbiased estimator of \sum , this estimator is:

And

$$\begin{aligned} \text{cov}(\bar{X}_1 - \bar{X}_2) &= \text{cov}(\bar{X}_1) + \text{cov}(\bar{X}_2) \\ &= \frac{1}{n_1} \sum + \frac{1}{n_2} \sum = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \sum \end{aligned} \quad (17)$$

where S_{pooled} is an unbiased estimation of \sum , and

$$\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_{pooled} \quad (18)$$

Is the estimator of $\text{cov}(\bar{X}_1 - \bar{X}_2)$.

By putting the sample estimators \bar{X}_1, \bar{X}_2 and S_{pooled} in place of μ_1, μ_2 and \sum in Eq. 9, we can get the sample estimator of Fisher's linear discriminant function and from (Morisson, 1976) it follows that

$$y = \hat{\alpha}'X = (\bar{X}_1 - \bar{X}_2)' S_{pooled}^{-1} X \quad (19)$$

The linear combination in (19) maximizes the ratio:

$$\frac{\text{squared distance between means of } Y}{\text{sample variance of } Y} = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{S_Y^2} = \frac{(\hat{\alpha}'d)^2}{\hat{\alpha}'S_{pooled}\hat{\alpha}} \quad (20)$$

As $d = \bar{X}_1 - \bar{X}_2$

The maximum of the ratio (20) is:

$$\begin{aligned} \text{Max}_{\hat{\alpha}} \frac{(\hat{\alpha}'d)^2}{\hat{\alpha}'S_{pooled}\hat{\alpha}} &= d' S_{pooled}^{-1} d = (\bar{X}_1 - \bar{X}_2)' \\ S_{pooled}^{-1} (\bar{X}_1 - \bar{X}_2) &= D^2 \end{aligned} \quad (21)$$

where D^2 is the sample squared distance between the two means.

Note that S_y^2 in (20) may be calculate as:

$$S_Y^2 = \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^2 + \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^2}{n_1 + n_2 - 2}$$

With $Y_{1j} = \hat{\alpha}' \bar{X}_1$ and $Y_{2j} = \hat{\alpha}' \bar{X}_2$

Since D^2 can be used, in certain situations, to test whether the population means μ_1 and μ_2 differ

significantly. Consequently, a test for differences in mean vectors can be viewed as a test for the significance of the separation that can be achieved.

The midpoint MP between the two univariate means $\bar{Y}_1 = \hat{\alpha}' \bar{X}_1$ and $\bar{Y}_2 = \hat{\alpha}' \bar{X}_2$ can be obtained from the following formula:

$$\hat{MP} = \frac{1}{2}(\bar{Y}_1 + \bar{Y}_2) = \frac{1}{2}(\bar{X}_1 + \bar{X}_2)' S_{pooled}^{-1} (\bar{X}_1 + \bar{X}_2) \quad (22)$$

So, the classification rule will be as follows:
allocate X_0 to π_1 if:

$$Y_0 - \hat{MP} \geq 0 \quad (23)$$

And allocate X_0 to π_2 if:

$$Y_0 - \hat{MP} < 0 \quad (24)$$

Testing the significance of the separation between two populations: Suppose the populations π_1 and π_2 are multivariate normal with a common covariance matrix Σ . From the work (Morisson, 1976), we know that to test the following statistical hypothesis:

$$H_0: \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 = \mu_1 \neq \mu_2$$

We can use the hotelling's T^2 -statistic as:

$$T^2 = (\bar{X}_1 - \bar{X}_2)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_{pooled}^{-1} \right]^{-1} (\bar{X}_1 - \bar{X}_2) \quad (25)$$

$$= \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)' S_{pooled}^{-1} (\bar{X}_1 - \bar{X}_2)$$

To prove that

$$T^2 = \frac{(n_1 + n_2 - 2)P}{(n_1 + n_2 - P - 1)} F_{P, n_1 + n_2 - P - 1} \quad (26)$$

Or:

$$F = \frac{(n_1 + n_2 - P - 1)}{(n_1 + n_2 - 2)P} T^2 \sim F_{P, n_1 + n_2 - P - 1} \quad (27)$$

where $F_{P, n_1 + n_2 - P - 1}^{(\alpha)}$ is the value of F with degrees of freedom of numerator P and degrees of freedom of denominator $n_1 + n_2 - P - 1$, where the area on its right is equal to α , we use the work (Anderson, 1984).

So, we reject the null hypothesis $H_0: \mu_1 = \mu_2$ at the level of significance α when:

$$F > F_{P, n_1 + n_2 - P - 1}^{(\alpha)}$$

Then we can conclude that the separation between the two populations π_1 and π_2 is significant.

Furthermore, D^2 can be used for testing the above hypothesis as follows:

Using (21), then Eq. 25 becomes:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2 \quad (28)$$

by substituting (28) in (27), we get:

$$F = \frac{(n_1 + n_2 - P - 1)}{(n_1 + n_2 - 2)P} \left(\frac{n_1 n_2}{n_1 + n_2} \right) D^2 \sim F_{P, n_1 + n_2 - P - 1} \quad (29)$$

Or:

$$D^2 = \left(\frac{(n_1 + n_2 - 2)P}{n_1 + n_2 - P - 1} \right) \left(\frac{n_1 + n_2}{n_1 n_2} \right) F_{P, n_1 + n_2 - P - 1} \quad (30)$$

Therefore, we reject $H_0: \mu_1 = \mu_2$ at the level of significance α when:

$$F > F_{P, n_1 + n_2 - P - 1}^{(\alpha)}$$

The Eq. 27 and 29 are equivalent. Either of them can be used for testing the statistical hypothesis above.

Evaluation of classification functions: When calculating the midpoint (MP) between the averages of π_1 and π_2 populations, incorrect classification may result, i.e., an observation comes from π_2 and is misclassified as π_1 or vice-versa. In both cases, we get on error in classification.

One of the important reason for judging the performance of any method of classification is the calculation of its error rate, i.e., calculation Misclassification probabilities, probable classification through knowing the probability distribution of π_1 and π_2 or unknown it.

In respect with this matter, we'll discuss two evaluating methods. They are:

Optimum Error Rate (OER): This rate is used for evaluating the performance of classification function when the statistical distribution of π_1 and π_2 is known.

Let $f_1(x)$ and $f_2(x)$ be the probability density functions with the $(P \times 1)$ vector random variable X for the populations π_1 and π_2 , respectively. Let Ω be the sample space-that is, the collection of all possible observations

X. let R_1 be that set of X values for which we classify objects as π_1 , let $R_2 = \Omega - R_1$ be the remaining X values for which we classify objects as π_2 .

Since every objects must be assigned to one and only one of the two populations, the sets R_1 and R_2 are mutually exclusive and exhaustive.

The conditional probability $P(2|1)$ of classifying an object as π_2 when, in fact, it is from π_1 is:

$$P(2|1) = P(X \in R_2 / \pi_1) = \int_{R_2 = \Omega - R_1} f_1(x) dx$$

Similarly, the conditional probability $P(2|1)$ of classifying an objects as π_1 when, it is really from π_2 is:

$$P(1|2) = P(X \in R_1 / \pi_2) = \int_{R_1} f_2(x) dx$$

Let P_1 be the Prior Probability of π_1 and P_2 be the prior probability of π_2 , where $P_1 + P_2 = 1$

According to (Huberty, 1984) the optimum error rate will be known

$$OER = P_1 \int_{R_2} f_1(x) dx + P_2 \int_{R_1} f_2(x) dx$$

Let us derive an expression for the OER when $P_1 = P_2 = 1/2$ and $f_1(x), f_2(x)$, are the multivariate normal densities.

Depending on the Eq. 12 and 13, we can defined the regions R_1 and R_2 as follows:

$$R_1 : (\mu_1 - \mu_2)' \sum^{-1} X - \frac{1}{2} (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2) \geq 0$$

$$R_2 : (\mu_1 - \mu_2)' \sum^{-1} X - \frac{1}{2} (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2) < 0$$

We can express these two regions by using:

$$Y = (\mu_1 - \mu_2)' \sum^{-1} X = \alpha' X$$

And as follows:

$$R_1(Y) : Y \geq \frac{1}{2} (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2)$$

$$R_2(Y) : Y < \frac{1}{2} (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2)$$

But Y is a linear combination of normal random variables, so the probability densities of Y $f_1(y)$ and $f_2(y)$, are univariate normal with means and variance given by:

$$\mu_{1Y} = \alpha' \mu_1 = (\mu_1 - \mu_2)' \sum^{-1} \mu_1$$

$$\mu_{2Y} = \alpha' \mu_2 = (\mu_1 - \mu_2)' \sum^{-1} \mu_2$$

$$\sigma_Y^2 = \alpha' \Sigma \alpha = (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 - \mu_2) = \Delta^2$$

Now,

$$\begin{aligned} OER &= \frac{1}{2} P(2|1) + \frac{1}{2} P(1|2) \\ &= \frac{1}{2} P \left[Y < \frac{1}{2} (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2) \right] + \\ &\quad \frac{1}{2} P \left[Y \geq \frac{1}{2} (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2) \right] \\ &= \frac{1}{2} P \left[\frac{Y - \mu_{1Y}}{\sigma_Y} < \frac{\frac{1}{2} (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2) - (\mu_1 - \mu_2)' \sum^{-1} \mu_1}{\Delta} \right] \\ &\quad + \frac{1}{2} P \left[\frac{Y - \mu_{2Y}}{\sigma_Y} \geq \frac{\frac{1}{2} (\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2) - (\mu_1 - \mu_2)' \sum^{-1} \mu_2}{\Delta} \right] \\ &= \frac{1}{2} \Phi \left(\frac{-\Delta}{2} \right) + \frac{1}{2} \left[1 - \Phi \left(\frac{\Delta}{2} \right) \right] = \Phi \left(\frac{-\Delta}{2} \right) \end{aligned} \tag{31}$$

Where, Φ is the cumulative distribution function of a standard normal random variable.

In most practical situations, the population quantities and are unknown, so the Eq. 31 must be modified by replacing the population parameters by their sample counterparts. Therefore, the evaluated of the performance of the sample classification function is calculated through the Actual Error Rate (AER) as follows.

$$AER = \Phi \left(-\frac{\sqrt{D^2}}{2} \right) \tag{32}$$

Apparent Error Rate (APER): This rate is a measure of performance that does not depend on the form of the parent populations and that can be calculated for any classification procedure.

The (APER) is defined as the fraction of observations in the training sample that are misclassified by the sample classification function.

The apparent error rate can be calculated from the confusion matrix, which shows actual versus predicted group membership.

According to work (Johnson and Wichern, 1988) the confusion matrix has the form:

	Predicted membership		Actual membership
Total	π_1	π_2	
n_1	$n_{1m} = n_1 - n_{1c}$	n_{1c}	π_1
n_2	n_{2c}	$n_{2m} = n_2 - n_{2c}$	π_2

where:

n_1 represents the observations from π_1 , and n_2 the observations from π_2 ,

- n_{1c} = No. of π_1 items correctly classified as π_1 Items,
- n_{1m} = No. of π_1 items misclassified as π_2 items,
- n_{2c} = No. of π_2 items correctly classified as π_2 items,
- n_{2m} = number of π_2 items misclassified as π_1 Items,

The apparent error rate then is:

$$APER = \frac{n_{1m} + n_{2m}}{n_1 + n_2} \quad (33)$$

The APER is intuitively appealing and easy to calculate.

Unfortunately, it tends to underestimate the AER and the problem does not disappear unless the sample sizes n_1 and n_2 are very large.

Identifying the relative importance of the variables of discriminant function: The coefficient vector

$\alpha = \sum_{i=1}^p (\mu_i - \mu_2)$ and the coefficient vector are

$\hat{\alpha} = S_{pooled}^{-1} (\bar{X}_1 - \bar{X}_2)$ not unique as any vector $C\alpha$ and $C\hat{\alpha}$

when ($C \neq 0$) will also maximize both ratios in (7) and in (11), respectively. Mostly, the vector $\hat{\alpha}$ can be transformed into a standard vector or multiplied by a constant amount for simplifying the concept and the illustration of its components. The work (Breiman *et al.*, 1984) gives the common method of use in this respect

$$\hat{\alpha}^* = \frac{\hat{\alpha}}{\sqrt{\hat{\alpha}'\hat{\alpha}}} = C\hat{\alpha} \quad C = (\hat{\alpha}'\hat{\alpha})^{-\frac{1}{2}} \quad (34)$$

So that $\hat{\alpha}^*$ has unit length.

It can be noted that the magnitudes of $\hat{\alpha}_1^*, \hat{\alpha}_2^*, \dots, \hat{\alpha}_p^*$ in (34) all lie in the interval [-1, 1] this will help in comparing the coefficients and in assessing the relative importance of the variables of the discriminant function.

APPLICATION SECTION

The organization of data: A random sample was selected, consisting of (28) students of the graduates of mathematics and statistics department at Al. Hussein Bin

Table 1: The achieved scores of major specialization subjects for the sample of students in both groups

	X_1	X_2	X_3	X_4	X_5	Group
1	90	96	93	88	88	1
2	89	99	91	95	76	1
3	84	80	90	70	74	1
4	71	64	93	73	84	1
5	84	81	80	89	90	1
6	76	84	98	90	89	1
7	79	63	77	90	79	1
8	71	70	73	84	87	1
9	57	60	83	69	80	1
10	50	84	70	72	62	1
11	64	53	81	65	62	2
12	56	56	80	68	54	2
13	56	76	82	57	76	2
14	61	54	92	66	90	2
15	65	53	80	72	70	2
16	59	74	79	60	71	2
17	50	59	75	61	81	2
18	57	70	80	65	94	2
19	53	75	87	70	84	2
20	61	87	66	76	66	2
21	50	56	82	55	85	2
22	51	50	82	60	71	2
23	50	55	74	78	75	2
24	56	50	76	63	64	2
25	50	72	67	66	68	2
26	50	64	91	78	67	2
27	61	54	67	50	70	2
28	54	53	64	68	60	2

X_1 = Topology (1), X_2 = Real Analysis(1), X_3 = Partial Differential Equation, X_4 = Mathematical Statistic (1), X_5 = Abstract Algebra (1)

Talal University for three academic courses. They are: the second course (2003/2004) and the summer course (2003/2004) and the first course (2004/2005). Their accumulative averages were: Good, very good, excellent.

On the basis of this sample, the data of this study were prepared on two stages: in the first stage, the students sample was distributed on two groups on the basis of their accumulative averages which reflect the student's general specialization. The first group included (10) students whose accumulative averages were very good and excellent. The second group included (18) students whose accumulative averages were good.

But in the next stage, (5) basic study materials were selected representing the major specialization of the graduate students. These materials are: X_1 = Topology (1), X_2 = Real analysis (1), X_3 = Partial Differential Equations, X_4 = Mathematical Statistics and X_5 = Abstract Algebra (1).

Table 1 shows the achieved scores of each student in the research sample and in the five materials which represent the variables of discriminant function which aims to estimate it.

Statistical analysis: By using the data given in table (1) and the mathematical formula given in (15) and (16), a computer calculation yields the summary statistics.

$$\bar{X}_1 = \begin{bmatrix} 75.1 \\ 78.1 \\ 84.8 \\ 82.0 \\ 80.9 \end{bmatrix}, \bar{X}_2 = \begin{bmatrix} 55.7778 \\ 61.7222 \\ 78.0556 \\ 65.4444 \\ 72.6667 \end{bmatrix}, n_1 = 10, n_2 = 18$$

$$S_{\text{pooled}} = \begin{bmatrix} 78.000 & 31.338 & 23.132 & 28.453 & 14.645 \\ 31.338 & 144.174 & 9.403 & 33.278 & 10.901 \\ 32.132 & 9.403 & 74.867 & 9.444 & 35.351 \\ 28.453 & 33.278 & 9.444 & 72.632 & 3.487 \\ 14.645 & 10.901 & 35.351 & 3.487 & 100.690 \end{bmatrix}$$

$$S_{\text{pooled}}^{-1} = \begin{bmatrix} 0.0168 & -0.0022 & -0.0040 & -0.0050 & -0.0006 \\ -0.0022 & 0.0081 & 0.0003 & -0.0029 & -0.0006 \\ -0.0040 & 0.0003 & 0.0172 & -0.0005 & -0.0055 \\ -0.0050 & -0.00029 & -0.0005 & 0.0171 & 0.0006 \\ -0.0006 & -0.0006 & -0.0055 & 0.0006 & 0.0120 \end{bmatrix}$$

The above estimates represent the prerequisites required for all our further accounts.

Testing the significance of the separation between the two groups: To test whether the population means μ_1 and μ_2 differ significantly, Hotelling's T^2 -statistic can be used through the test of the following statistical hypothesis:

$$H_0: \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 = \mu_1 \neq \mu_2$$

Now, from (25) $T = 43.0132$ and from (27), $F = 7.2792$.

We see that $F > F_{5,22}(0.01)$, therefore we reject H_0 at 10% level of significance and then we conclude that the separation between the two groups under study is significant. This gives a logic justification for the importance of the discriminant function for discriminating or classifying the students in one group of both identified groups.

Calculating of the discriminant function: Fisher's (sample) linear discriminant function; Eq. 19:

$$Y = \hat{\alpha}' X = (\bar{X}_1 - \bar{X}_2)'^{-1} S_{\text{pooled}} X$$

Was estimated and the our estimated equation is:

$$y = 0.1739X_1 + 0.0402X_2 - 0.0105X_3 + 0.1403X_4 + 0.051X_5$$

The analysis of variance table for the discriminant analysis can be performed to test the following hypothesis:

$H_0 = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ Vs. H_1 At least one of the five α 's is nonzero and as follows:

F	MS	SS	d.f	SOV
7.2791	2.2138	$SS_{(B)} = 11.0692$	5	Between x's
	0.3041	$SS_{(W)} = 6.6909$	22	Within x's
		17.7601	27	Total

Where

$$SS_{(B)} = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)} [D^2]^2 = 11.0692$$

$$SS_{(D)} = D^2 = 6.6909$$

since $F > F_{5,22}(0.01)$, we reject H_0 at the 10% level of significance. This means that the discriminant function has a big advantage in discriminating or classifying the students under study.

Identifying the relative importance of the variables of discriminant function: The Importance of each variable of discriminant function can be calculated in accordance with the Eq. 34.

$$\hat{\alpha}_i^* = \frac{\hat{\alpha}_i}{\sqrt{\hat{\alpha}'\hat{\alpha}}}$$

By applying this formula, we find that:

$$\hat{\alpha}_1^* = 0.74668$$

$$\hat{\alpha}_2^* = 0.17279$$

$$\hat{\alpha}_3^* = -0.04490$$

$$\hat{\alpha}_4^* = 0.60229$$

$$\hat{\alpha}_5^* = 0.21875$$

On the basis of the above calculations, we observe that Topology (1) subject will be on top importance among the major specialization subject followed by the Mathematical Statistics (1) subject and then the Abstract Algebra (1) subject and the Real Analysis and the Partial Differential Equations.

Estimating the midpoint between both groups: The midpoint between both groups can be estimated by using the Eq. 22 and accordingly, we can get the following:

$$\hat{MP} = 27.5968$$

Therefore, the classification rule will be as follows:

Table 2: The actual classification and predicted classification in accordance with fisher's function of students sample in both groups

Case No.	Actual group	Predicated group
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	2
10	1	2
11	2	2
12	2	2
13	2	2
14	2	2
15	2	2
16	2	2
17	2	2
18	2	2
19	2	2
20	2	2
21	2	2
22	2	2
23	2	2
24	2	2
25	2	2
26	2	2
27	2	2
28	2	2

allocate X_o to π_1 if:
 $Y_o = \hat{\alpha}' X_o \geq 27.5968$
 Allocate X_o to π_2 if:
 $Y_o = \hat{\alpha}' X_o < 27.5968$

Table 2 displays the results of the predicted classification in accordance with the above classification rule.

Evaluating of classification function: On the basis of the previous classification rule and in following up the values of the Table 2, we observe a wrong classification within the first group as two students belonging to the second group were classified wrongly in the first group.

For the purpose of evaluating the performance of the adopted classification method, we'll calculate both averages APER and AER.

Apparent Error Rate (APER): For calculating this rate, we'll first calculate the confusion matrix as follows:

Total	Predicted membership		Actual membership
	π_2	π_1	
10	$n_{1c} = 8$	$n_{1m} = 2$	π_1
18	$n_{2m} = 0$	$n_{2c} = 18$	π_2

By using the Eq. 33 we find that:

$$APER = \frac{2}{28} = 0.0714$$

Actual Error Rate (AER): This rate is calculated in accordance with the formula No. (32) in which we find that:

$$ARE = \Phi\left(-\frac{\sqrt{D^2}}{2}\right) = \Phi(-1.29) = 0.0985$$

The value of the error rate in both cases is close and does not exceed (10%). This value means that the optimal classification rule in use, resulted in placing about (10%) of the students in a population which they do not belong to.

CONCLUSIONS

This study is considered as an attempt of using the discriminant analysis method and classification for the purpose of achieving a discriminant function, through which we can discover the reasons of the actual difference between two groups of graduate students from mathematics and statistics sections at Al-Hussein Bin Talal University. Then their classification will be repeated on the basis of their major specialization.

In this study, Fisher's linear discriminant Function was used as a tool for statistical analysis. Then, it will be estimated on the basis of a sample consisting of (28) male and female students who were classified into two groups as per their accumulative averages which reflect their general specialization. The first group included (10) students. Their averages were very good and excellent. Whereas the second group included (18) male and female students of Good averages. Five basic academic subjects were selected, representing the major specialization of students and their achieved scores were written down. These subjects are: Topology (1), Real Analysis (1), Partial Differential Equations, Mathematical Statistics (1) and Abstract Algebra (1).

The most important results and conclusions which this study achieved are:

- The study argued the high significant differences of big morale function between both groups of students on the basis of (T^2) test. These big differences assert the existence of real differences among the scientific levels of students and their mental and innovative abilities. Such thing requires the discrimination of those who are distinguished of them.

- The study showed and on the basis of linear discriminant function which had been estimated that two students were re-classified into the second group as per their major specialization after they were within the first group on the basis of their general specialization.
- The study showed that Topology (1) subject is on top of the major specialization subjects in respect of its relative importance, followed by the Mathematical Statistics(1) subject and Abstract Algebra (1)subject, the Real Analysis subject and finally the Partial Differential Equations subject.
- The classification rule used in this study is marked by its high performance which led to classifying about 90% of students in the actual group which they physically belong to. It also classifies about 10% of students in a group other than the group which they belong to. This denotes the power of discriminant function which we achieved and the possibility of using it for discrimination and future classification of the students (of unknown affiliation) into the correct group.

REFERENCES

- Anderson, T.W., 1984. An Introduction to Multivariate Statistical Methods: 2nd Edn., New York, John Wiley.
- Breiman, L.J., R. Friedman, Olshen and C. Stone, 1984. classification and regression trees. Belmont, Ac, Wadshorth, Inc.
- Hand, D.J., 1981. Discrimination and Classification , New York, John Wiley.
- Huberty, C.J., 1984. Applied Discriminant Analysis, John Wiley.
- Johnson, R.A. and D.A. Wichern, 1998. Applied Multivariate Statistical Analysis, Prentic Hall, Inc.
- Krzanowski, W.J., 1977. The performance of fisher's linear discriminant function under non-optimal conditions. Technometrics, 19: 191-200.
- Morisson, D.F., 1976. Multivariate Statistical Methods: Mc Gray-Hill, Inc.