Discriminant Analysis In View Of Statistical And Operations Research Techniques

Prof. Enayat I. Hafez
Professor of Operations Research, Department of Statistics,Faculty of Commerce, Al-Azhar University (Girls' Branch)

Dr. Saida M. Abdel-Nabi
Associate Professor of Statistics
Faculty of Commerce, Al-Azhar University (Girls' Branch)

Dr. Eman M. Abdel-Fatah
Associate Professor of Statistics
Faculty of Commerce, Al-Azhar University (Girls' Branch)
Cairo, Egypt

Asmaa Salama Ahmed Zeidan
Assistant Lecturer, Department of Statistics
Faculty of Commerce, Al-Azhar University (Girls' Branch)
Cairo, Egypt

***Abstract*—The techniques of discriminant analysis were largely studied. The literature on this subject is very abundant. The origin of discriminant analysis is fairly old, and its development reflects the importance of these methods. The purpose of this paper is to perform an experimental comparison between parametric methods [statistical approaches such as Fisher's linear discriminant function (FLDF), and logistic regression (LR)] and non-parametric methods [mathematical programming methods such as linear programming (LP) models, and goal programming (GP) models] for resolving the classification problem. These methods are applied on the real-life data that are provided from Human Development Report (2010) [11]. The results of comparison demonstrate that; the best classification, easiest and less time consuming methods in application are linear programming models, followed by the statistical methods (FLDF, LR).**

Keywords—Discriminant analysis; Fisher's linear discriminant function; Logistic regression; Linear programming; Goal programming.

I. INTRODUCTION

Discriminant analysis is considered one of the most important statistical techniques that concerned with separating distinct sets of observations and with assigning new observation to one of two (or more) distinct groups on the basis of the measurements of some variables with a low error rate. The functions that are used to do this assignment may be identical to the ones used in the multivariate analysis of variance (MANOVA) procedures [14]. The classification problem arises when it is not possible to assign an individual directly with one of several populations. In this case, the classification problem has two uses; prediction and description [1]. Discriminant analysis is capable of handling either two groups or multiple groups. When two classifications are involved, the technique is referred to as two-group discriminant analysis. When three or more classifications are identified, the technique is referred to as multiple discriminant analysis [9]. In this paper, we will deal only with the two-group discriminant analysis.

Discriminant analysis has been successfully used for many fields such as; medicine, education [14], geology [13], personnel management, community, industry [22], routine banking, and plant taxonomy [17].

II. DISCRIMINANT ANALYSIS TECHNIQUES

A. *Introduction*

The methods of discriminant analysis were largely studied. The aim of these methods is to create rules for separating distinct groups as much as possible and for assigning an observation of unknown origin to one of two (or more) distinct groups, which minimize the total probability of misclassification or the average cost of misclassification. In this section a presentation of four models of discriminant analysis; Fisher's linear discriminant function (FLDF), logistic regression (LR), linear programming (LP) models, and goal programming (GP) models, are introduced.

B. *Fisher's linear discriminant function*

Fisher's linear discriminant function (FLDF) is considered the most commonly statistical method that is used for discrimination, classification and prediction under usual statistical assumptions such as; multivariate normality of the independent variables, equality of variance and covariance matrices and relative equality of groups sample sizes.

Fisher 1936 has suggested using a linear combination of the observations (x) to create (y)'s as they give simple enough functions of the (x) to be handled easily. Fisher's idea was to transform the multivariate observations (x) to univariate observations (y) such that the (y)'s derived from populations (Π_1) and (Π_2) were separated as much as possible. He used this method to classify two species of iris based on four measurements; sepal length, sepal width, petal length, and petal width; and the classification was excellent.

The linear combination is denoted by: $y = a'x$ (1)

A fixed linear combination of the (x)'s takes the values $(y_{11}, y_{12}, \dots, y_{1n_1})$ for the observations from population (Π_1) and the values $(y_{21}, y_{22}, \dots, y_{2n_2})$ for the observations from population (Π_2). The objective is to select the linear combination of the (x) that maximize the ratio of squared distance between sample means of (y) to its variance as:

$$max \frac{(\bar{y}_1-\bar{y}_2)^2}{s_y^2} = \frac{(a'\mu_1-a'\mu_2)^2}{a'\Sigma a} = \phi \quad (2)$$

By differentiating (ϕ) with respect to (a), we get :

$$a = \Sigma^{-1}(\mu_1 - \mu_2) \quad (3)$$

Then, $$y = a'x = (\mu_1 - \mu_2)'\Sigma^{-1} x \quad (4)$$

In practice, it's rarely to know the parameters of the population (μ_1, μ_2, Σ), so they can be estimated by $\bar{x}_1, \bar{x}_2, S_{pooled}$ as follows:

$$y = a'x = (\bar{x}_1 - \bar{x}_2)' s_{pooled}^{-1} x \quad (5)$$

Using samples of size (n_1) from population (Π_1) and (n_2) from population (Π_2) to estimate the parameters. From the data matrices, the sample mean vectors and covariance matrices are determined by:

$$\bar{x}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1j} \quad , \quad \bar{x}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j}$$

$$s_1 = \frac{1}{n_1-1}\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)(x_{1j} - \bar{x}_1)'$$

$$s_2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)(x_{2j} - \bar{x}_2)'$$

(6) Since it is assumed that the two populations have the same covariance matrix (Σ), the sample covariance matrices (s_1, s_2) are combined (pooled) to derive a single, unbiased estimate of (Σ) as follows:

$$S_{pooled} = \frac{(n_1-1)s_1 + (n_2-1)s_2}{n_1+n_2-2} \quad (7)$$

Then, the allocation rule based on Fisher's linear discriminant function is given by:

-Allocate (x_0) to population (Π_1) if :

$$\hat{y}_0 = (\bar{x}_1 - \bar{x}_2)' s_{pooled}^{-1} x_0 \geq \hat{m} \quad \text{or} \quad (\hat{y}_0 - \hat{m} \geq 0) \quad (8)$$

-Allocate (x_0) to population (Π_2) if :

$$\hat{y}_0 < \hat{m} \quad \text{or} \quad (\hat{y}_0 - \hat{m} < 0) \quad (9)$$

Where $(\hat{m})$ is the midpoint of the interval between $(\bar{y}_1)$ and $(\bar{y}_2)$ as follows:

$$\hat{m} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' s_{pooled}^{-1} (\bar{x}_1 + \bar{x}_2) \quad (10)$$

[14, 12, 1, 19]

C. Logistic Discriminant Model (LGD)

Logistic discriminant model, also known as logistic regression (LR) or logit analysis, is a specialized form of regression that is formulated to predict and explain a binary (two-group) dependent variable rather than a metric dependent measure [9].

Logistic regression analyzes the relationship between multiple independent variables and a single dependent variable. It requires a nonmetric, dichotomous (binary) dependent variable; categorical variable with two categories. Like a dummy variable, this is coded 0/1 and indicates if a condition is or is not present, or if an event did or did not occur. Logistic regression (LR) employs binomial probability theory in which there are only two values, the probability (p) is one rather than zero, to predict that the event/ person belongs to one group rather than the other. There are main uses of logistic regression; the first is the prediction of group membership and the other one is that logistic regression provides knowledge of the relationships and strengths among the variables [20].

Logistic regression model has been widely used in social research, medical research, biological research, food science, design, customer behaviors, market segmentation, and bankruptcy prediction [23].

Logistic regression model (LR) does not necessarily require the assumptions of Fisher's linear discriminant analysis (FLDA). The independent variables in LR need not be normally distributed, nor linearly related, nor of equal variance within each group. However, the dependent variable in LR must be a dichotomy (two categories). If the populations are normal with identical variance-covariance matrices, Fisher's linear discriminant analysis (FLDA) method is preferred to logistic regression model for the discrimination problem [18].

Logistic regression tries to find the best fitting model to describe the relationship between the dependent variable (response variable / outcome) and a set of independent (predictor / explanatory) variables in the form:

$$y_i = logit\ (p) = log\left[\frac{p}{1-p}\right] = a + b_1x_1 + b_2x_2 + \cdots + b_ix_i \qquad (11)$$

Where (p) is the probability of the dependent variable (outcome) which can only range from zero to one, $logit\ (p)$ is the logistic transformation of (p), (a) is the intercept term (the constant of the equation), (b_i) represents the coefficients of the predictor variables, and (x_i) is the independent variables [23].
Where also (p) and $(1-p)$ represent respectively, the probabilities of belonging to (G_1) and (G_2).

For any observation (i), the dependent variable takes the values:

$$y_i = \begin{cases} 1 & \text{if an event occurs} \\ 0 & \text{if an event doesn't occur} \end{cases} \qquad (12)$$

A maximum likelihood method, which maximizes the probability of getting the observed results given the fitted regression coefficients, is used. (p) can be calculated with the following formula as follows:

$$p = \frac{exp^{(a+b_1x_1+b_2x_2+\cdots+b_ix_i)}}{1+exp^{(a+b_1x_1+b_2x_2+\cdots+b_ix_i)}} \qquad (13)$$

Where (p) is the probability that a case is in a particular category [18].

D. *Linear programming models*

Linear programming (LP) model was first conceived by George B.Dantzig 1947. He published the ˝simplex" method for solving linear programs. It was developed further by Charnes, Cooper and Ferguson 1955. The success of linear programming (LP) returns to its flexibility in describing multitudes of real-life situations. Linear programming (LP) has been applied in many areas including military, industry, agriculture, transportation, economics, health systems and even behavioral and social sciences [21].

The discriminant analysis (D.A) based on linear programming models (LP) has received considerable attention from many researchers. Freed and Glover (1981) [4, 5] suggested linear programming (LP) weighting scheme as an alternative solution to the discriminant problem. They applied the model to the two-group classification problems. Bajgier and Hill (1982) [2] introduced an experimental comparison of three linear programming approaches (linear programming, linear goal programming and mixed-integer programming) and the Fisher´s procedure for the discriminant problems. Gochet, Stam, Srinivasan and Chen (1995) [7] introduced a nonparametric linear programming formulation for the general multi-group classification problem. Lam and Moy (1996) [15] also provided an improved linear programming model for the multi-group discriminant problem.

Linear programming (LP) technique can be effectively used to reduce or eliminate the complexities of conventional statistical approaches without sacrificing the essential power of the method.

Considering two groups (G_1, G_2) with (k) attributes where (x) is the $(k * n)$ matrix of attribute scores,(n) is the sample size, (w_1, w_2 ,…, w_k) are the attribute weights, and (x_{ij}) is the value of the variable (j), $(j = 1,2, \dots, k)$ for the observation (i), $(i = 1,2, \dots, n)$ of the two groups $(1,2)$. The assignment of an object into a group depends on the value of its classification score. There are many models of Linear programming (LP) technique to discriminant analysis and some of them are introduced.

1) *The MMD model*

This model was suggested by Freed and Glover (1981a) [4] to solve the classification problem. The MMD model means "maximize the minimum distance of any group member's score from the cutoff value (critical value/ breakpoint) (C). The critical value is arbitrary pre-chosen. Their formulation is as follows:

$$maximize \quad (d)$$

Subject to:

$$\sum_{j=1}^{k} w_j\, x_{ij} - d \geq c \qquad , i \in G_1 \qquad (14)$$

$$\sum_{j=1}^{k} w_j\, x_{ij} + d \leq c \qquad , i \in G_2 \qquad (15)$$

$$\sum_{j=1}^{k} w_j \qquad \text{(Normalization constraint)} \qquad (16)$$

Where (d) is the distance between groups; w_j $(j = 1, 2, \dots\dots, K)$ and (c) are unrestricted in sign. The normalization constraint is needed to avoid the trivial solution (that is, an all zero solution).

2) The MSD model

This model was proposed by Freed and Glover (1981b) [5] and Bajgier and Hill (1982) [2] to solve the classification problem. The MSD model means "minimize the sum of deviations". The objective of this model is to minimize the total group overlap. This model can be formulated as follows:

$$min(\sum_{i=1}^{n} d_i)$$

Subject to:

$\sum_{j=1}^{k} w_j\, x_{ij} + d_i \geq c \qquad , i \in G_1$
(17)

$\sum_{j=1}^{k} w_j\, x_{ij} - d_i \leq c \qquad , i \in G_2$
(18)

$\sum_{j=1}^{k} w_j = 1$ (Normalization constraint)
(19)

$$d_i \geq 0 , (i = 1,2, \ldots, n)$$

Where (w_j , c) are unrestricted in sign. A normalization constraint is needed to avoid a trivial solution.

3) The LCM model

This model was suggested by Lam, Choo, Moy (1996) [LCM] [16]. They divided the process of their model into two steps; the first step determines the value of the attribute weights by minimizing the sum of deviations from the observation scores to their group mean classification scores, and the second one determines the cut-off value for the classification. The linear programming (LP) formulation for the first step can be stated as follows:

[LCM 1] $\quad min(\sum_{i=1}^{n} d_i)$

Subject to:

$\sum_{j=1}^{k} w_j\, (x_{ij} - u_{1j}) + d_i \geq 0 \qquad , i \in G_1$
(20)

$\sum_{j=1}^{k} w_j\, (x_{ij} - u_{2j}) - d_i \leq 0 \qquad , i \in G_2$
(21)

$\sum_{j=1}^{k} w_j\, (u_{1j} - u_{2j}) \geq 1$ (Normalization constraint)
(22)

$$d_i \geq 0 , (i = 1,2, \ldots, n)$$

Where $w_j\ (j = 1,2, \ldots, k)$ are unrestricted in sign, (u_{1j}) is the mean of (j^{th}) variable in group (G_1) and (u_{2j}) is the mean of (j^{th}) variable in group (G_2). The normalization constraint is needed to avoid a trivial solution.

After solving (LCM1), the weights (w_j) and the object scores are obtained. Then the object scores are used in the second step and the classification is made as follows:

[LCM 2] $\quad min(\sum_{i=1}^{n} h_i)$

Subject to:

$\sum_{j=1}^{k} w_j\, x_{ij} + h_i \geq c \qquad , i \in G_1$
(23)

$\sum_{j=1}^{k} w_j\, x_{ij} - h_i \leq c \qquad , i \in G_2$
(24)

$$h_i \geq 0 , (i = 1,2, \ldots, n)$$

Where (w_j , c) are unrestricted in sign, and (h_i) are the sum of deviations. This model can also be referred as linear programming based on the mean [3].

E. Goal programming models

The goal programming (GP) is an important technique for decision makers. It solves multi-objective decision- making problems by finding a set of satisfying solutions [3]. Most recently researches used goal programming (GP) and applied it in many functional areas, including academic planning, financial planning and hospital administration. There are many types of goal programming (GP) such as lexicographic goal programming (LGP) that also known as pre-emptive GP, non- pre-emptive GP, Weighted GP (WGP) and Minimax GP. The goal programming (GP) was first suggested by Charnes, cooper and Ferguson 1961 and developed further by Lee 1972 and Ignizio 1976.

The main factors that making goal programming (GP) has an advantage over Linear programming (LP) are the structure of goal programming (GP) and the use of the objective function. The objective function of Linear programming (LP) is measured in only one dimension such as maximizing total profit or minimizing total cost that is impossible in Linear programming technique having multiple objectives unless they can be measured in the same units. However, the goal programming can solve such these problems by using "simplex" method to find the optimum solution to a single-dimensional or multi-dimensional linear objective function subject to a set of linear constraints [10, 8]. Hence goal programming technique is used for solving multi-objective models and the principal idea is to convert the original multiple objectives into a single goal [21]. Goal programming technique has a useful advantage in minimizing the unwanted deviations between the achievement of goals and their aspiration levels [3].

The discriminant analysis (D.A) based on goal programming (GP) has received a considerable attention by many researchers. Bal, Örkcü, and Celebioğlu discussion (2006) [3] is more relevance to this topic. They developed two new mathematical approaches; goal programming (GP) model based on the mean (GP Mean) and Goal programming (GP) model based on the median (GP Median) in solving two- group classification problems and these models are perform well both in separating the groups and the group-membership predictions of new objects. These two models will be applied to proposed real data.

Goal programming (GP) model in discriminant analysis (D.A) has an advantage over Linear programming (LP) model in discriminant analysis (D.A). While (LP) model determines the attribute weights and cut-off value in two steps, (GP) model determines simultaneously all of these values in one step. So (GP) model is faster, more efficient and also practicable than (LP) model [3].

1) Discriminant analysis (D.A) using goal programming (GP) technique based on the mean

In the goal programming (GP) model based on the mean (GP Mean), the first priority is to minimize the sum of deviations of object classification scores from the group mean scores. The second priority is to minimize the sum of deviations between classification scores and cut-off value without degrading the solution of the first priority. This model was suggested by Bal, H. ¸ Örkcü, H.H. and Celebioğlu, H. (2006) [3]. The goal programming model based on the mean (GP mean) can be given as follows:

[GP mean] $\min a = \{\sum_{i=1}^{n} d_i , \sum_{i=1}^{n} h_i \}$

Subject to:

$\sum_{j=1}^{k} w_j (x_{ij} - u_{1j}) + d_i \geq 0 \quad , i \in G_1$ (25)

$\sum_{j=1}^{k} w_j (x_{ij} - u_{2j}) - d_i \leq 0 \quad , i \in G_2$ (26)

$\sum_{j=1}^{k} w_j (u_{1j} - u_{2j}) \geq 1$ (Normalization constraint) (27)

$\sum_{j=1}^{k} w_j x_{ij} + h_i \geq c \quad , i \in G_1$ (28)

$\sum_{j=1}^{k} w_j x_{ij} - h_i \leq c \quad , i \in G_2$ (29)

$$d_i , h_i \geq 0 , (i = 1,2, \dots , n)$$

Where (w_j , c) are unrestricted in sign, and (h_i) are the sum of deviations. An object will be classified into (G_1) if the classification score is greater than or equal to the cut-off value (c), otherwise the new object will be classified into the second group (G_2).

2) Discriminant analysis (D.A) using goal programming (GP) technique based on the median

This model was also suggested by Bal, H. ¸ Örkcü, H.H. and Celebioğlu, H. (2006). Goal programming (GP) model based on the minimization of deviations from the group median scores (GP Med) can be used similar to the foregoing model. By using the median instead of the mean, goal programming (GP) model based on the median (GP Med) can be defined as follows:

[GP Med] $\min a = \{\sum_{i=1}^{n} d_i , \sum_{i=1}^{n} h_i \}$

Subject to:

$\sum_{j=1}^{k} w_j (x_{ij} - med_{1j}) + d_i \geq 0 \quad , i \in G_1$ (30)

$\sum_{j=1}^{k} w_j (x_{ij} - med_{2j}) - d_i \leq 0 \quad , i \in G_2$ (31)

$\sum_{j=1}^{k} w_j (med_{1j} - med_{2j}) \geq 1$ (Normalization constraint) (32)

$\sum_{j=1}^{k} w_j x_{ij} + h_i \geq c \quad , i \in G_1$ (33)

$\sum_{j=1}^{k} w_j x_{ij} - h_i \leq c \quad , i \in G_2$ (34)

$d_i , h_i \geq 0 , (i = 1,2, \dots , n)$

Where (w_j , c) are unrestricted in sign, (h_i) are the sum of deviations, (med_{1j}) is the median of (j^{th}) variable in group (G_1) and (med_{2j}) is the median of (j^{th}) variable in group (G_2). The normalization constraint is needed to avoid a trivial solution.

III. APPLICATION

In order to compare and evaluate the selected models (FLDF, LR, MMD, MSD, GP mean, and GP median); a real data set of forty countries with four variables is considered (see Appendix A). The data is taken from Human Development Report (HDR) of the year (2010) [11]. The forty countries are previously divided into two groups; twenty countries are considered very high human development countries and the other twenty are medium human development countries (G_1: 20 countries, G_2: 20 countries).

Note, the Human Development Report (HDR) of the year (2010) considered seventeen composite measures, under each of them number of variables between four and ten variables, to divide (169) countries into four groups. This research considered carefully only four variables for a total of forty countries from two groups, which may change the allocation of one (or more) country from one group to another. These techniques are used to show the power of the methods, but not to criticize the results of Human Development Report as it will be the reference for comparison.

The classification between the chosen two groups is made depending on the chosen four variables as follows:

X_1: The political freedom (Democracy). The values of this variable take scores (0-2); where (0) is nondemocratic, (1) is democratic with no alternation, and (2) is democratic.

X_2: The overall life satisfaction. It takes values from zero to ten; where (0) is the least satisfaction, and (10) is the most satisfaction.

X_3: Public expenditure on education (% of GDP), where GDP represents the gross domestic product.

X_4: Public expenditure on health (% of GDP).

For solving linear programming and goal programming models, Win QSB (Quantitative System for Business Plus) package has been employed. The SPSS program is used to solve Fisher's and Logistic methods.

IV. COMPUTATIONAL RESULTS

Computational discriminant analysis results for the previous models are summarized in the following tables:

Table (1): Classification results for developed and under developed countries using Fisher's linear discriminant function (FLDF):

Classification Results

		DEVELOP	Predicted Group Membership 1	2	Total
Original	Count	1	18	2	20
		2	2	18	20
	%	1	90.0	10.0	100.0
		2	10.0	90.0	100.0

(90% of original grouped countries are correctly classified).

Table (2): Classification results for developed and under developed countries using Logistic Regression (LR) model:

Classification Table[a]

Observed			Predicted COUNTRY 1	2	Percentage Correct
Step 1	COUNTRY	1	18	2	90.0
		2	2	18	90.0
	Overall Percentage				90.0

a. The cut value is .500

(90% of original grouped countries are correctly classified).

Table (3): Classification results for developed and under developed countries using MMD model:

Classification Table[a]

Observed			Predicted COUNTRY 1	2	Percentage Correct
Step 1	COUNTRY	1	20	0	100.0
		2	0	20	100.0
	Overall Percentage				100.0

a. (100% of original grouped countries are correctly classified).

Table (4): Classification results for developed and under developed countries using MSD model:

Classification Table[a]

Observed			Predicted COUNTRY 1	2	Percentage Correct
Step 1	COUNTRY	1	19	1	95.0
		2	2	18	90.0
	Overall Percentage				92.5

a. (92.5% of original grouped countries are correctly classified).

Table (5): Classification results for developed and under developed countries using GP mean model:

Classification Table[a]

Observed			Predicted COUNTRY 1	Predicted COUNTRY 2	Percentage Correct
Step 1	COUNTRY	1	19	1	95.0
		2	0	20	100.0
	Overall Percentage				97.5

a. (97.5% of original grouped countries are correctly classified).

Table (6): Classification results for developed and under developed countries using GP median model:

Classification Table[a]

Observed			Predicted COUNTRY 1	Predicted COUNTRY 2	Percentage Correct
Step 1	COUNTRY	1	20	0	100.0
		2	0	20	100.0
	Overall Percentage				100.0

a. (100% of original grouped countries are correctly classified).

Table (7): Summary of the results of the selected models (FLDF, LR, MMD, MSD, GP mean, GP median) for developed and under developed countries:

Method	Group (1) No. of correct classifi-cations	Group (1) No. of misclassifi-cations	Group (2) No. of correct classifi-cations	Group (2) No. of misclassifi-cations	Total Number of correct classifi-cations	Hit-ratio %
FLDF	18	2	18	2	36	90
LR	18	2	18	2	36	90
MMD	20	0	20	0	40	100
MSD	19	1	18	2	37	92.5
GP Mean	19	1	20	0	39	97.5
GP Median	20	0	20	0	40	100

Hit-ratio is the ratio of correctly classified objects to the total number of objects to be classified.

Considering the results of the hit-ratio, it is clear that the MMD method and the GP median method are the best as they classify the countries correctly to their groups with 100 % hit ratio. The next best method is GP mean with hit-ratio 97.5 %, then MSD method with hit-ratio 92.5 %. The two statistical methods (FLDF, LR) give equal hit-ratio of 90 %.

It is also important to notice the misclassified events (countries) as; FLDF and LR methods misclassify the countries(11, 19) in group one to group two and also misclassify the countries (21, 32) in group two to group one. So, the two statistical methods give identical results.

Another view to the misclassified events (countries) in statistical methods; in the results of LP models (MMD, MSD) and GP mean model, it is found that these events either misclassified or can be called border ones as the left hand side (L.H.S) equal zero for each of them, and according to the rules of LP and GP models; if the L.H.S is greater than (or equal to) zero or less than (or equal to) zero, the event stays in its group. So, it's clear that the ones that have L.H.S equal zero can be called border (boundary) ones. This argument shows that the statistical methods are also good ones and are sensitive to the border events.

Although the logistic regression (LR) method is more suitable to this problem as the data does not meet Fisher's assumptions of normality and equal variance-covariance matrices within each group where applying its method in such case may not give exact results, the results of FLDF were perfect matching the LR method results.

REFERENCES

[1] A.A. Afifi, and V. Clark, "Computer-aided Multivariate Analysis", Wadsworth, Inc., 1984.

[2] S.M. Bajgier, and A.V. Hill, "An experimental comparison of statistical and linear programming approaches to the discriminant problem", Decision Sciences, Vol.13, Issue 4, pp. 604-618, 1982.

[3] H. Bal, H.H. Örkcü, and S. Celebioğlu, "An experimental comparison of the new goal programming and the linear programming approaches in the two-group discriminant problems", Computers and Industrial Engineering, Vol.50, pp. 296-311, 2006.

[4] N. Freed, and F. Glover, "A linear programming approach to the discriminant problem", Decision Sciences, Vol.12, pp.68–74, 1981a.

[5] N. Freed, and F. Glover, "Simple but powerful goal programming models for discriminant problems", European Journal of Operational Research, Vol.7, pp.44–60, 1981b.

[6] N. Freed, and F. Glover, "Evaluating alternative linear programming models to solve the two-group discriminant problem", Decision Sciences, Vol.17, Issue 2, pp.151–162, 1986.

[7] W. Gochet, A. Stam, V. Srinivasan, and S. Chen, "Multi-group discriminant analysis using linear programming", Journal of Operations Research, Vol.45, No.2, pp.213-225, 1995.

[8] P.K. Gupta, and D.S. Hira, "Problems in Operations Research: Principles and Solutions", P.K. Gupta & D.S. Hira, 1991.
[9] J.F. Hair, W.C. Black, R.E. Anderson, and R.L. Tatham, "Multivariate Data Analysis", 5th Ed. New Jersey, Prentice Hall, Inc., 1998.
[10] F.S. Hillier, and G.J. Lieberman, "Introduction to Operations Research". McGraw-Hill, 1995.
[11] Human Development Report, "The real wealth of nations: Pathways to human development", United Nations Development Programme (UNDP), New York, USA, 2010.
[12] R.A. Johnson, and D.W. Wichern, "Applied Multivariate Statistical Analysis", 5th Ed., New Jersey, Prentice Hall, Inc., 1998.
[13] S.K. Kachigan, "Multivariate Statistical Analysis": A conceptual Introduction, Radius Press, New York, 1991.
[14] P.A. Lachenbruch, "Discriminant Analysis". Hafner Press, New York, Macmillan Publishing Co., Inc., 1975. [15] K.F. Lam, and J.W. Moy, "Improved linear programming formulations for the multi-group discriminant problem", Journal of the Operational Research Society, Vol.47, Issue 12, p.1526, 1996.
[16] K.F. Lam, E.U. Choo, and J.W. Moy, "Minimizing deviations from the group mean: A new linear programming approach for the two-group classification problem", European Journal of Operational Research, vol.88, pp.358-367, 1996.
[17] D.F. Morrison, "Multivariate Statistical Methods", Mc Graw-Hill Book Company, New York, 1967.
[18] M. Pohar, M. Blas, and S. Turk, "Comparison of logistic regression and linear discriminant analysis": A simulation study, Metodološki Zvezki, Vol.1, No.1, pp.143-161, 2004.
[19] A.C. Rencher, "Methods of Multivariate Analysis", New York, John Wiley & Sons, Inc., 1995.
[20] S.A. Sweet, and K. Grace-Martin, "Data Analysis with SPSS": A First Course in Applied Statistics. 2ndEd. Boston, MA: Allyn & Bacon, 2003.
[21] H.A. Taha, "Operations Research": An Introduction, Macmillan Publishing Company, division of Macmillan, Inc., 1992.
[22] N.H. Timm, "Applied Multivariate Analysis", Springer-Verlag, New York, Inc., 2002.
[23] S. Youssef, and A. Rebai, "Comparison between statistical approaches and linear programming for resolving classification problem", International Mathematical Forum, Vol.2, No. 63, pp.3125-3141, 2007.

Appendix (A)

Raw data of developed and under developed countries

	Group	country	X_1	X_2	X_3	X_4
Group (1)	1	Norway	2	8.1	6.7	7.5
	2	Australia	2	7.9	4.7	6.0
	3	New Zealand	2	7.8	6.2	7.1
	4	United States	2	7.9	5.5	7.1
	5	Ireland	2	8.1	4.9	6.1
	6	Nether lands	2	7.8	5.5	7.3
	7	Canada	2	8.0	4.9	7.1
	8	Sweden	2	7.9	6.7	7.4
	9	Germany	2	7.2	4.4	8.0
	10	Japan	2	6.8	3.4	6.5
	11	Korea	2	6.3	4.2	3.5
	12	Switzer land	2	8.0	5.3	6.4
	13	France	2	7.1	5.6	8.7
	14	Finland	2	8.0	5.9	6.1
	15	Iceland	2	7.8	7.5	7.7
	16	Belgium	2	7.3	6.1	7.0
	17	Denmark	2	8.2	7.9	8.2
	18	Greece	2	6.8	4.0	5.8
	19	Singapore	1	6.7	2.8	1.0
	20	Qatar	0	6.7	3.3	2.9

Group (2)	21	Dominican	2	7.6	2.2	1.9
	22	China	0	6.4	1.9	1.9
	23	El Salvador	2	6.7	3.6	3.6
	24	Thailand	2	6.3	4.9	2.7
	25	Bolivia	2	6.5	6.3	3.4
	26	Paraguay	2	6.9	4.0	2.4
	27	Philippinin	2	5.5	2.6	1.3
	28	Botswana	1	4.7	8.1	4.3
	29	Moldova	2	5.7	8.2	5.2
	30	Mongolia	2	5.7	5.1	3.5
	31	Egypt	1	5.8	3.8	2.4
	32	Guyana	1	6.5	6.1	7.2
	33	Namibia	1	5.2	6.5	3.2
	34	Indonesia	2	5.7	3.5	1.2
	35	Kyrgyzstan	2	5.0	6.6	3.5
	36	South Afric	1	5.0	5.1	3.6
	37	Syrian	0	5.9	4.9	1.6
	38	Tajikistan	1	5.1	3.5	1.1
	39	Viet Nam	0	5.4	5.3	2.8
	40	Morocco	0	5.8	5.7	1.7