Hacettepe Journal of Mathematics and Statistics
Volume 36 (2) (2007), 169 – 180

DATA ENVELOPMENT ANALYSIS APPROACH TO TWO-GROUP CLASSIFICATION PROBLEMS AND AN EXPERIMENTAL COMPARISON WITH SOME CLASSIFICATION MODELS

Hasan Bal* and H. Hasan Örkcu*

Received 22 : 08 : 2006 : Accepted 24 : 06 : 2007

Abstract

Discriminant Analysis is a method for determining group classifications for a set of similar units or observations. A number of new efficient mathematical programming approaches have been developed as an alternative to examining classification problems using statistical models. In this study two new mathematical programming approaches are developed for the minimization of the sum of the deviations and the concept of relative efficiency for Data Envelopment Analysis when solving the two group classification problem. The efficiency and practicability of the suggested approaches are supported with a simulation study involving three different distributions and different cases for the units in the groups.

Keywords: Discriminant analysis, Data envelopment analysis, Mathematical programming, CCR-DA, BCC-DA, LOO hit rate.

2000 AMS Classification: 62 H 30.

*Gazi University, Faculty of Arts and Science, Department of Statistics Teknikokullar, 06500 Ankara, Turkey.
E-mail: (H. H. Örkcu) hhorkcu@gazi.edu.tr (H. Bal) hasanbal@gazi.edu.tr

1. Introduction

Discriminant Analysis (DA) involves the classification of samples into two or more given groups according to certain known classification attributes. Discriminant Analysis is used for two purposes: to classify observations into one of a number of mutually exclusive groups, and to explain differences between these groups based one or more observable attributes. Linear discriminant analysis is a widely used research tool in the social sciences, in business areas such as finance, marketing, and accounting, and in other areas involving taxonomical and classification analyses such as biology [11]. In general, the two-group scenario is most widely used in real applications. Typical applications are credit scoring, targeting consumer groups for sales, etc. Linear discriminant analysis, developed by Fisher, is the classical method for the classification task. It is theoretically optimal for situations where the underlying populations are multivariate normal and where all the different groups have equal covariance structures. With multivariate populations having unequal covariance structures, quadratic discriminant analysis can be used [1]. However, when the normality assumption is not guaranteed, these well-known DA approaches are usually unable to provide good or even satisfactory classification results. In order to overcome such difficulties, mathematical programming approaches have been developed. These approaches have been found to outperform the statistical procedures in many applications, and have received a great deal of attention. In the study of classification problems many efficient mathematical programming techniques have been developed as an alternative to statistical models, see Bajgier and Hill[2], Fred and Glover[6,7,8], Glover[10], Koehler and Erenguc[11], Lam, Choo and Moy[12], Lam and Moy[13,14,15], Ragsdale and Stam[16], etc.

In this study, two new mathematical programming approaches are developed for the minimization of the sum of the deviations and for the relative efficiency concept of Data Envelopment Analysis in solving two group classification problems. The practicability of the suggested approaches are supported with a simulation study.

The remainder of this article is organized as follows: Mathematical programming approaches to two-group Discriminant Analysis are briefly discussed in section 2. Data Envelopment Analysis and the proposed CCR-DA and BCC-DA methods are discussed in section 3. Performance evaluation methods for classification models are discussed in section 4. Section 5 reports and discusses the results of the simulation. Section 6 concludes the study.

2. Mathematical Programming Approaches to Two-Group Discriminant Analysis

The first method in Discriminant Analysis had a statistical basis and was developed by Fisher. Fisher developed the discriminant function by maximizing the ratio of the among-group sum of squares to the within-group sum of squares. Then mathematical programming techniques, namely linear programming, were applied to Discriminant Analysis by Fred and Glover [6,7]. Fred and Glover suggested a classification model which is based on minimizing the sum of deviations. Bajgier and Hill [2] provide an experimental comparison of statistical discriminant analysis and linear programming approaches to classification problems. Following these studies, many linear programming approaches based on classification criteria such as minimizing the sum of the deviations, minimizing the sum of the misclassified objects and maximizing the distance of the intra groups have been developed. As with Fisher's original method, linear programming models are useful when the classification is to be independent of the distribution [11].

Consider the two-group classification problem with k attributes. Let x be the $k \times n$ matrix of attribute scores of a sample of size n drawn from the groups G_1 and G_2. If $w_1, w_2, \ldots, w_k$ are the attribute weights, then the classification score of any object is defined as $S_i = \sum_{j=1}^{k} x_{ij} w_j$. The assignment of an object to a group depends on the value of its classification score. The Minimization Sums of Deviations Model (MSD) can be formulated as follows [6]:

$$\text{(2.1)} \qquad \min \sum_{i=1}^{n} d_i$$

$$\sum_{j=1}^{k} w_j x_{ij} + d_i \geq c, \;\; i \in \mathrm{G}_1,$$

$$\sum_{j=1}^{k} w_j x_{ij} - d_i \leq c, \;\; i \in \mathrm{G}_2,$$

where $d_i \geq 0,\; i = 1, 2, \ldots, n),\; w_j,\; j = 1, 2, \ldots, k$, and c are unrestricted variables. The inequality $d_i > 0$ indicates an occurrence of misclassification on the i^{th} observation. A normalization constraint is needed to avoid trivial solutions (that is, an all zero solution). Solving this model gives us the values of w_j and c, from which we can obtain the classification score, $S_i = \sum_{j=1}^{k} x_{ij} w_j$, of any object. An object will be classified into G_1 if its classification score is greater than or equal to c, otherwise into G_2.

This model, and many of the other existing linear programming models, determine the attribute weights and cut-off values taking place here at the same time. Lam, Choo and Moy [12] divide the process of their model into two steps: the first constitutes the determination of attribute weights, and the second determines the cut-off values for the classification. Their model (in the first step) makes use of an objective function minimizing the sum of deviations from the group mean classification scores. The model of Lam, Choo and Moy [12] (LPMEAN) can be formulated as follows:

LPMEAN 1

$$\text{(2.2)} \qquad \min \sum_{i=1}^{n} d_i$$

$$\sum_{j=1}^{k} w_j \left(x_{ij} - \mu_{1j}\right) + d_i \geq 0, \;\; i \in \mathrm{G}_1,$$

$$\sum_{j=1}^{k} w_j \left(x_{ij} - \mu_{2j}\right) - d_i \leq 0, \;\; i \in \mathrm{G}_2,$$

$$\sum_{j=1}^{k} w_j \left(\mu_{1j} - \mu_{2j}\right) \geq 1,$$

where $d_i \geq 0,\; i = 1, 2, \ldots, n,\; w_j,\; j = 1, 2, \ldots, k$, are unrestricted variables, and μ_{1j} is the mean of the j^{th} variable in group G_1 and μ_{2j} is the mean of the j^{th} variable in group G_2. With the aid of this model, $w_j,\; j = 1, 2, \ldots, k$, the attribute weights are found, and then the object scores are obtained. In this model, the weights are obtained by making the object scores close to the mean score of the group in which they take place. Then the object scores are used in the following model, and the classification is made:

LPMEAN 2

$$\text{(2.3)} \qquad \min \sum_{i=1}^{n} h_i,$$

$$S_i + h_i \geq c,\ i \in \mathrm{G}_1,$$
$$S_i - h_i \leq c,\ i \in \mathrm{G}_2,$$

where $h_i \geq 0,\ i = 1, 2, \ldots, n$, and c is an unrestricted variable. As seen, the classification is made in two independent steps.

Rasgdale and Stam [16] proposed a two-stage linear programming model (RS) for solving classification problems. The model discussed currently works in this way:

(RS)

$$\min \sum_{i=1}^{n} d_i \tag{2.4}$$
$$\sum_{j=1}^{k} w_j x_{ij} + d_i \geq c_1,\ i \in \mathrm{G}_1,$$
$$\sum_{j=1}^{k} w_j x_{ij} - d_i \leq c_2,\ i \in \mathrm{G}_2,$$

where $d_i \geq 0,\ i = 1, \ldots, n$, the $w_j,\ j = 1, 2, \ldots, k)$ are unrestricted variables and c_1, c_2 are predetermined variables such that $c_1 > c_2$. It is important to note that when a classification score of an observation is between c_1 and c_2, the observation is in the classification gap. In the second stage, a linear programming problem that minimizes the sum of the deviations related with observations whose classification scores fall in the classification gap in the first stage, is solved properly.

3. Data Envelopment Analysis and Classification Models

Data Envelopment Analysis (DEA), developed by Charnes, Cooper and Rhodes, provides a non parametric methodology for evaluating the efficiency of each of a set of comparable decision making units (DMUs), relative to one another [3]. An important feature of DEA is its capability to provide efficiency scores, while taking account of both multiple inputs and multiple outputs. Charnes, Cooper and Rhodes who developed Farrel's [5] idea extended the single *output / input* ratio measure of efficiency to the multiple *output / input* measure of efficiency. Then the relative efficiency of any DMU is calculated by forming the ratio of a weighted sum of outputs to a weighted sum of inputs, where the weights for both outputs and inputs are to be selected so that the the resulting efficiency measure of each DMU satisfies the constraint that no DMU can have a relative efficiency score grater than unity.

In DEA there are many models which can be used to measure efficiency, these models being derived from the ratio models in which the efficiency is measured as the ratio of the weighted sum of the outputs to the weighted sum of the inputs [3]. Considering n units, each of which has m inputs denoted by $x_{ij},\ i = 1, 2, \ldots, m)$ and s outputs denoted by $y_{rj},\ r = 1, 2, \ldots, s$, the mathematical programming problem in ratio form can be given as follows:

$$\max \sum_{r=1}^{s} u_r y_{ro} \Bigg/ \sum_{i=1}^{m} v_i x_{io} \tag{3.1}$$
$$\sum_{r=1}^{s} u_r y_{rj} \Bigg/ \sum_{i=1}^{m} v_i x_{ij} \leq 1,\ j = 1, 2, \ldots, n$$
$$u_r, v_i \geq 0.$$

If the sum of the weighted inputs of DMU whose efficiency is to be measured is made equal to 1 (i.e $\sum_{i=1}^{m} v_i x_{io} = 1$), then the basic efficiency model, known as CCR, is obtained as follows:

$$
\begin{aligned}
(3.2)\qquad & \max\ w_o = \sum_{r=1}^{s} u_r y_{ro} \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0,\ j = 1, 2, \dots, n \\
& \sum_{i=1}^{m} v_i x_{io} = 1 \\
& u_r, v_i \ge 0,\ i = 1, 2, \dots, m,\ r = 1, 2, \dots, s.
\end{aligned}
$$

Likewise, the BCC model can be derived in the same way [4]:

$$
\begin{aligned}
(3.3)\qquad & \max\ w_o = \sum_{r=1}^{s} \mu_r y_{ro} + u_0 \\
& \sum_{i=1}^{m} v_i x_{io} = 1 \\
& \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + u_0 \le 0,\ j = 1, 2, \dots, n \\
& \mu,\ v \ge 0,\ i = 1, 2, \dots, m,\ r = 1, 2, \dots, s,\ u_o \text{unrestricted.}
\end{aligned}
$$

While the efficiency of DMUs has been measured by these models, it is necessary to solve the model n-times for each DMU. The optimum value of the objective function gives the efficiency score of the DMU of interest in the model. A different set of weights are selected $u_r,\ v_i$ for each DMU. Therefore, the set of optimum weights identifies a hyperplane for each DMU. Any DMU whose efficiency score is equal to one is defined as efficient, otherwise inefficient.

It is necessary to separate the variables as input and output in DEA. This separation depends on their effects related to DMU. Instead of input and output variables, Retzlaff-Roberts [17] prefered to use the concept of positive effect variables and negative effect variables on units. Any factor whose increase (while the others are held constant) leads to a unit considered better or more efficient is defined as *positive*. On the contrary, any factor whose decrease (while the others are held constant) leads to a unit considered better or more efficient is defined as *negative*.

Both DA and DEA can be considered as methods to calculate the performance when measuring the DMUs using linear programming. In order to distinguish the members of a group, each method uses some factor weights. In DA it is initially known which group each unit is a member of, and a set of factor weights and "threshold" value are sought which produce the best possible separation of two groups. The resulting weights and threshold value, which define a hyperplane that attempts to separate the two groups, can then be used to predict membership for subsequent observations. While DEA is not generally thought of as a two-group classification technique, it does classify units into two groups, namely the inefficient and DEA-efficient units. In DEA it is not initially known to which group units belong, but the threshold that separates the two groups is known, an efficiency score below 1 categorizes a unit as inefficient and the others are deemed DEA-efficient. A set of factor weights, which also form a hyperplane, is sought for each unit which attempts to classify that unit into DEA efficient groups [17].

Consider a Discriminant Analysis problem, whose classification attributes consist of m input attributes and s output attributes. We denote by x_{ij}, $i = 1, 2, \ldots, m$ and y_{rj}, $r = 1, 2, \ldots, s$ the values of the i^{th} input and r^{th} output attributes of the j^{th} sample. One of the advantages of DEA is its capability for dealing with multiple inputs and multiple outputs. When the efficiency of the multiple inputs and multiple outputs is used as the classification criterion, it is naturally hoped that there will exist an efficiency cut of value 1 (threshold, boundary value) that separates two groups.

$$
\begin{aligned}
(3.4) \qquad & \max \sum_{r=1}^{s} u_r y_{ro} \Big/ \sum_{i=1}^{m} v_i x_{io} \\
& \sum_{r=1}^{s} u_r y_{rj} \Big/ \sum_{i=1}^{m} v_i x_{ij} \geq 1, \ j \in G_1 \\
& \sum_{r=1}^{s} u_r y_{rj} \Big/ \sum_{i=1}^{m} v_i x_{ij} \leq 1, \ j \in G_2 \\
& u_r, v_i \geq 0, \ j = 1, 2, \ldots, n
\end{aligned}
$$

If the sum of the weighted inputs of DMU whose efficiency is to be measured is made equal to 1 (i.e. $\sum_{i=1}^{m} v_i x_{io} = 1$), then the model (3.5) is obtained as follows:

$$
\begin{aligned}
(3.5) \qquad & \max \sum_{r=1}^{s} u_r y_{ro} \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \geq 0, \ j \in G_1 \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \ j \in G_2 \\
& \sum_{i=1}^{m} v_i x_{io} = 1, \ j = 1, 2, \ldots, n \\
& u_r, v_i \geq 0, \ i = 1, 2, \ldots, m, \ r = 1, 2, \ldots, s.
\end{aligned}
$$

Considering the deviation variable d_j which is a measure of the misclassification ratio, we introduce the external deviation variables and its minimum sum criterion into the above model, which results in a multi objective linear programming model shown as follows:

$$
\begin{aligned}
(3.6) \qquad & \max \sum_{r=1}^{s} u_r y_{ro} \\
& \min \sum_{j=1}^{n} d_j \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - d_j \geq 0, \ j \in G_1 \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + d_j \leq 0, \ j \in G_2 \\
& \sum_{i=1}^{m} v_i x_{io} = 1, \ j = 1, 2, \ldots, n \\
& u_r, v_i, d_j \geq 0, \ i = 1, 2, \ldots, m, \ r = 1, 2, \ldots, s.
\end{aligned}
$$

The model (3.6) is a classification model based on both minimizing the sum of the external deviations and maximizing the sum of the weighted outputs for the related DMUs. This model is a multi objective linear programming model. There are several solution methods for solving a multiple objective linear programming model. We used the basic weighted approach in order to solve this multi objective linear programming model.

$$
(3.7) \qquad \max \left(\sum_{r=1}^{s} u_r y_{ro} - M \sum_{j=1}^{\mathrm{n}} d_j \right)
$$

$$
\begin{aligned}
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - d_j \geq 0, \ j \in G_1 \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + d_j \leq 0, \ j \in G_2 \\
&\sum_{i=1}^{m} v_i x_{io} = 1, \ j = 1, 2, \ldots, n \\
&u_r, v_i, d_j \geq 0, \ i = 1, 2, \ldots, m, \ r = 1, 2, \ldots, s,
\end{aligned}
$$

where M is a large number. We call this model CCR-DA.

After solving the above model, the classification score ($\sum_{r=1}^{s} u_r y_{ro}$) for each unit will be obtained. These scores are compared with 1, if greater than 1 we assign the unit to the first group, otherwise to the second group.

Likewise, applying the same process to the BCC model gives:

$$
(3.8) \qquad \max \left(\sum_{r=1}^{s} u_r y_{ro} + u_0 - M \sum_{j=1}^{\mathrm{n}} d_j \right)
$$

$$
\begin{aligned}
&\sum_{r=1}^{s} u_r y_{rj} + u_0 - \sum_{i=1}^{m} v_i x_{ij} - d_j \geq 0, \ j \in G_1 \\
&\sum_{r=1}^{s} u_r y_{rj} + u_0 - \sum_{i=1}^{m} v_i x_{ij} + d_j \leq 0, \ j \in G_2 \\
&\sum_{i=1}^{m} v_i x_{io} = 1, \ j = 1, 2, \ldots, n \\
&u_0 \text{ unrestricted}, u_r, v_i, d_j \geq 0, \ i = 1, 2, \ldots, m, \ r = 1, 2, \ldots, s.
\end{aligned}
$$

We call this model BCC-DA. Here the relative efficiency of a unit under classification is defined as $\sum_{r=1}^{s} u_r y_{ro} + u_0$. When it is greater than or equal to 1, the unit should be classified into the first group, otherwise into the second group.

4. Performance Evaluation for Classification Models

The performance of classification models is generally assessed in terms of hit rates, i.e. the proportion of observations classified correctly. Although the apparent hit rate, which is obtained by using the model to classify the observations used in its derivation, is widely used in this context, this measure is positively biased. The hit rate in a holdout sample, i.e. a sample of observations of known class membership that is separate from the training sample, can be used to overcome this difficulty. Since the holdout sample hit rate is training/holdout sample specific, the average holdout hit rate for a number of training/holdout samples should be determined, although the average holdout hit rate may be affected by sampling bias. The LOO (*leave-one-out)* is another type of performance indicator used in classification problems [9]. A reasonable estimate of future classification performance can be obtained from the LOO. In order to calculate the LOO hit ratio for each model each observation is omitted in turn from the whole sample and the remaining observations used to generate a classification function which is then used to classify the omitted observations. The LOO examines whether the omitted observation is correctly classified by the estimated classification function. In this study, the LOO hit rate is used.

5. Simulation Experiment

We carried out a Monte Carlo simulation study with 100 repetitions to compare the performance of the methods FLDF (Fisher's Linear Discriminant Function), MSD, LP-MEAN, RS, CCR-DA and BCC-DA for the two-group classification problem. All of the results for these methods were obtained using the program MATLAB 7.0. and WINQSB. The distributions considered in this study and their parameters are as follows:

Multivariate Normal $G_1 \sim (6\,6\,6)$ $G_2 \sim (5\,5\,5)$ $\Sigma = \mathrm{I}$.

Exponential $G_1 \sim (2)$ $G_2 \sim (1)$.

Uniform $G_1 \sim (20\,40)$ $G_2 \sim (15\,30)$.

It is necessary to separate the variables (factors) as input and output in the CCR-DA and BCC-DA classification models. In order to use the same variables in the models considered (FLDF, MSD, LPMEAN, RS, CCR-DA, BCC-DA), all of the variables in the CCR-DA and BCC-DA models are taken as output variables. Also, the input variable is taken as one ("1") in these models. Similar applications in DEA can be found Charnes *et al* [4].

For the two-groups, the unit number in the first group is n_1, and that in the second group is n_2, the characteristics for the different values of n_1 and n_2 are taken randomly from the distributions given above (as $n_1 = 50$, $n_2 = 50$; $n_1 = 60$, $n_2 = 40$; $n_1 = 30$, $n_2 = 70$; $n_1 = 80$, $n_2 = 20$). For example, $n_1 = 50$ units for the first group and $n_2 = 50$ units for the second group are selected from the related distributions for the case of $n_1 = 50$ and $n_2 = 50$. For the case $n_1 = 60$ and $n_2 = 40$, $n_1 = 60$ units for the first group and $n_2 = 40$ units for the second group are chosen from the related distributions. The selection of unit process for the other cases is done in a similar way.

In Table 1, the different values of n_1 and n_2, the three different choices of distribution (multivariate Normal, exponential, uniform), and the average LOO hit rate (correct classification) numbers obtained from the 100 repetitions of the sample are presented for the six different classification methods.

Upon examination of Table 1, it is seen that the CCR-DA model and especially the BCC-DA model usually have a higher LOO hit rate than the other four methods.

There are three main hypotheses we wish to test, and these hypotheses were repeated for the three distributions and four different sample cases taken from the groups ($n_1 = n_2 = 50$; $n_1 = 60, n_2 = 40$; $n_1 = 30, n_2 = 70$ and $n_1 = 80, n_2 = 20$). The hypotheses for the CCR-DA model are:

H_0 **:** There is no difference between the average LOO hit rate of CCR-DA and the average LOO hit rate of the i^{th} method,

H_1 **:** The average LOO hit rate of CCR-DA is higher than average LOO hit rate of the i^{th} method.

Here i = FLDF, MSD, LPMEAN, RS. In a similar way, the hypotheses for the BCC-DA model are:

H_0 **:** There is no difference between the average LOO hit rate of BCC-DA and the average LOO hit rate of the i^{th} method,

H_1 **:** The average LOO hit rate of BCC-DA is higher than average LOO hit rate of the i^{th} method.

Table 1. Average LOO hit rate (correct classification) of the six approaches.

		METHOD					
n_1, n_2	Distribution	***FLDF***	***MSD***	***LPMEAN***	***RS***	***CCR-DA***	***BCC-DA***
$n_1 = 50$	***Mult. Norm.***	85.01 (5.16)*	81.02 (5.21)	85.98 (4.68)	82.10 (4.96)	86.22 (5.03)	86.88 (4.81)
$n_2 = 50$	***Exponential***	75.63 (5.27)	76.33 (4.87)	78.13 (5.04)	77.06 (4.32)	79.87 (5.07)	80.09 (4.68)
	Uniform	83.62 (4.72)	80.03 (5.26)	85.13 (4.68)	81.26 (5.11)	85.06 (4.51)	86.91 (4.92)
$n_1 = 60$	***Mult. Norm.***	86.61 (4.18)	84.12 (4.89)	86.78 (4.68)	85.51 (4.42)	87.91 (5.17)	88.54 (4.79)
$n_2 = 40$	***Exponential***	76.15 (4.87)	75.81 (5.21)	77.02 (5.09)	76.63 (4.71)	78.84 (4.96)	79.47 (4.84)
	Uniform	87.35 (4.84)	83.11 (5.01)	88.96 (4.50)	86.79 (5.11)	89.61 (4.72)	90.91 (4.87)
$n_1 = 30$	***Mult. Norm.***	87.17 (4.82)	83.61 (5.09)	87.92 (4.22)	84.63 (5.29)	88.17 (4.52)	88.89 (4.38)
$n_2 = 70$	***Exponential***	77.22 (5.15)	77.18 (5.27)	79.42 (4.96)	79.17 (5.03)	81.52 (4.26)	81.87 (5.09)
	Uniform	87.15 (4.57)	81.87 (5.11)	89.46 (4.91)	84.28 (4.46)	89.28 (4.62)	90.81 (4.69)
$n_1 = 80$	***Mult. Norm.***	85.56 (5.25)	82.48 (4.70)	88.11 (5.23)	83.96 (4.42)	90.03 (5.09)	90.57 (4.71)
$n_2 = 20$	***Exponential***	77.18 (4.57)	78.38 (4.78)	78.93 (5.17)	78.88 (5.12)	80.92 (4.83)	81.82 (4.46)
	Uniform	86.58 (4.83)	82.38 (5.21)	87.59 (4.53)	85.43 (4.92)	88.26 (5.09)	89.94 (4.78)

* The values in parentheses are standard deviations for the LOO hit rate (correct classification number).

Table 2 lists the results of the hypotheses tests which claim the average LOO hit rate of models CCR-DA and BCC-DA is greater than the average LOO hit rate of the models FLDF, MSD, LPMEAN and RS.

In Table 1, for example, the values 75.63 and 5.27 pertaining to the FLDF model for the case $n_1 = 50$, $n_2 = 50$ and the exponential distribution are the average LOO hit rate and the standard deviation of the model over 100 random samples. Similarly the values 79.87 and 5.07 of the model CCR-DA are also the average LOO hit rate and the standard deviation over 100 samples, respectively. Using this knowledge, we perform the following hypothesis test.

H_0 **:** There is no difference between the average LOO hit rate of CCR-DA and the average LOO hit rate of the FLDF method,

H_1 **:** The average LOO hit rate of CCR-DA is higher than average LOO hit rate of the FLDF method.

After testing these hypothesis, the value 4.10 in Table 2 shows the value of the test statistic t for the hypothesis that the average LOO hit rate of the model CCR-DA is greater than the average LOO hit rate of the method FLDF for the exponential distribution. In Table 2, the letter "a" shows that the calculated $p-$ value corresponding to 4.10 is less than 0.05. A decision on the outcome of the hypothesis can be given according to the value 4.10 of the test statistic, or the value of p. The fact that the value of p is less than 0.05 implies that the hypothesis H_0 is rejected with a significance level of 0.05. As a result, the average LOO hit rate of CCR-DA is found to be statistically greater than the average LOO hit rate of FLDF.

A similar result holds for a comparison of the methods BCC-DA and FLDF. The calculated value of 4.48 shows that for the cases $n_1 = 50$, $n_2 = 50$ and the exponential distribution, the average LOO hit rate of the model BCC-DA is greater statistically than the average LOO hit rate of the method FLDF. Furthermore, the model BCC-DA is superior to CCR-DA for the FLDF method (the calculated value is 4.10 for CCR-DA, and 4.48 for BCC-DA). The superiority of the model BCC-DA over the model CCR-DA is clearly seen in all cases.

Table 2. Hypothesis test results (t-values of paired t-tests) of CCR-DA and BCC-DA against (FLDF, MSD, LPMEAN, RS)

	CCR-DA				**BCC-DA**			
	$n_1 = 50$ $n_2 = 50$	$n_1 = 60$ $n_2 = 40$	$n_1 = 30$ $n_2 = 70$	$n_1 = 80$ $n_2 = 20$	$n_1 = 50$ $n_2 = 50$	$n_1 = 60$ $n_2 = 40$	$n_1 = 30$ $n_2 = 70$	$n_1 = 80$ $n_2 = 20$
Test 1	**FLDF**							
M. Nor.	1.18	1.24	1.01	3.43a	1.87a	1.98a	1.71a	4.14a
Exp.	4.10a	2.67a	4.07a	3.05a	4.48a	3.36a	4.12a	4.05a
Uni.	1.56	2.33a	2.01a	1.21	3.44a	3.52a	3.46a	2.57a
Test 2	**MSD**							
M. Nor.	5.07a	3.24a	4.21a	6.07a	5.84a	4.13a	4.97a	6.94a
Exp.	3.57a	2.91a	3.79a	2.03a	3.93a	3.47a	3.97a	2.93a
Uni.	5.13a	6.54a	6.65a	4.55a	6.81a	7.62a	8.01a	5.97a
Test 3	**LPMEAN**							
M. Nor.	0.29	1.10	0.27	1.50	0.94	1.81a	1.02	2.22a
Exp.	1.80a	1.79a	2.12a	1.62	2.01a	2.42a	2.20a	2.47a
Uni.	-0.07	0.61	-0.16	0.53	1.97a	2.06a	1.20	1.88a
Test 4	**RS**							
M. Nor.	4.22a	2.44a	3.19a	4.89a	4.89a	3.19a	3.87a	5.69a
Exp.	3.11a	2.26a	2.23a	1.67a	3.36a	2.90a	2.41a	2.54a
Uni.	4.04a	2.88a	5.12a	2.23a	5.75a	4.09a	6.52a	3.52a

[a] H_0 reject (p value < 0.05).

When the model CCR-DA is compared with the FLDF, MSD, LPMEAN and RS methods, the model CCR-DA is superior statistically in 34 of 48 different hypothesis tests. In the remaining 14 hypothesis tests, there is no difference statistically between the four methods. While the method of CCR-DA is superior to the methods MSD and RS in all the situations considered, it is superior to the methods FLDF and LPMEAN for the exponential distribution. It is found that the model BCC-DA is superior statistically than the methods FLDF, MSD, LPMEAN and RS in 45 of 48 different hypotheses test. While the method CCR-DA is superior to the methods FLDF, MSD and RS in all the cases considered, for the values $n_1 = 50$, $n_2 = 50$ and the multivariate Normal distribution, and $n_1 = 30$, $n_2 = 70$ and the multivariate Normal and exponential distributions, there is no statistical difference between LPMEAN and BCC-DA.

As a result of the simulation studies, we deduce that the model of CCR-DA, and especially the model of BCC-DA, can be used efficiently in the two-group classification problems.

6. Conclusion

In this study, two new mathematical programming models CCR-DA and BCC-DA are developed to solve two-group classification problems by using the relative efficiency concept model and the model minimizing the sum of deviations. When these suggested approaches are applied to simulation data, the results show that these are practicable and efficient. Especially, the model BCC-DA is superior to all other classification models, in the majority of experiments.

References

[1] Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 1984)

[2] Bajgier, S. M. and Hill, A. V. *An experimental comparison of statistical and linear programming approaches to the discriminant problem*, Decision Sciences **13**, 604–618, 1982.

[3] Charnes A, Cooper W. W. and Rhodes E. *Measuring the efficiency of decision making units*, European Journal of Operational Research **2**, 429–444, 1978.

[4] Cooper, W. W., Seiford, L. M. and Tone, K. *Data Envelopment Analysis* (Kluwer Academic Publishers, Boston, 2000).

[5] Farrell M. J. *The measurement of productivity efficiency*, Journal of Royal Statistical Society, Series A, CXX, 253–287, 1957.

[6] Fred, N. and Glover, F. *Simple but powerful goal programming formulations for the statistical discriminant problem*, European Journal of Operational Research **7**, 44–60, 1981.

[7] Fred, N. and Glover, F. *A linear programming approach to the discriminant problem*, Decision Sciences **12**, 68–74, 1981.

[8] Fred, N. and Glover, F. *Evaluating alternative linear programming models to solve the two-group discriminant problem*, Decision Sciences **17**, 589–585, 1986.

[9] Glen, J. J. *A comparison of standard and two-stage mathematical programming discriminant analysis methods*, European Journal of Operational Research **171**, 496–515, 2006.

[10] Glover, F. *Improved linear programming models for discriminant analysis*, Decision Sciences **21**, 771–785, 1990.

[11] Koehler, G. J. and Erenguc, S. S. *Minimizing misclassifications in linear discriminant analysis*, Decision Sciences **21**, 63–85, 1990.

[12] Lam, KF., Choo, E. U. and Moy, J. W. *Minimizing deviations from the group mean: A new linear programming approach for The two-group classification problem*, European Journal of Operational Research **88**, 358–367, 1996.

[13] Lam, KF. and Moy, J. W. *Improved linear programming formulations for the multi-group discriminant problem*, Journal of Operational Research Society **47**, 1526–1529, 1996.

[14] Lam, KF. and Moy, J. W. *An experimental comparison of some recently developed linear programming approaches to the discriminant problem*, Computers and Operations Research **24** (7), 593–599, 1997.

[15] Lam, KF. and Moy, J. W. *Combining discriminant methods in solving classification problems in two-group discriminant analysis*, European Journal of Operational Research **138**, 294–301, 2002.

[16] Ragsdale, C. T. and Stam, A. *Mathematical programming formulations for the discriminant problem: An old dog does new tricks*, Decision Sciences **22**, 296–307, 1991.

[17] Retzlaff-Roberts, D. L. *Relating discriminant analysis and data envelopment analysis to one another*, European Journal of Operational Research **23**, 311–322, 1996.