Journal of Engineering and Applied Sciences 14 (8): 2477-2484, 2019
ISSN: 1816-949X

Mathematical Programming Approach for Variable Selection in Discriminant Analysis: An Environmental Application

[1]Aiman Majid Nassar and [2]Anwer Ahmed Saleh
[1]College of Applied Sciences-Heet,
[2]Collage of Basic Education/Haditha, University of Anbar, Anbar, Iraq

Abstract: The environmental impact assessment of projects is based on several variables. It is desirable to select the most important variables needed for classifying any project in only one of three classes according to severity of possible environmental impact. In this study, the selection of variables in discriminant analysis between more than two groups using the Mathematical programming approach is applied to Egypt environmental impact assessment survey 2000 data to select the most important variables for classifying new projects in the true classes. The results are compared with those of the stepwise method. The comparison shows that according to the percent of correct classification, the Mathematical programming model is better than the stepwise method.

Key words: Discriminant analysis, mixed integer programming, variable selection, mathematical programming, environmental impact assessment, environmental

INTRODUCTION

In light of the growing global concern about environmental problems and the importance of achieving sound management of the natural resources within the framework of sustainable development, environmental problems have to gain increasing attention, especially, regarding their impact on the global level (Palerm, 2000). The Environmental Protection Law in Egypt (Law # 4 in 1994) and its executive regulations state that new establishments or projects as well as expansion of existing establishments must be subject to an Environmental Impact Assessment (EIA) before a permit is issued.

EIA is a systematic process which provides a framework for gathering and documenting information and view regarding the environmental consequences of activities, so that, the importance of the effects and scope for enhancing, modifying or mitigating them can properly be evaluated.

One of the definitions of EIA is that "It is a planning aid concerned with identifying, predicting and assessing impacts arising from proposed activities such as policies, programs plans and development projects which may affect the environment" (EEAA., 2002).

According to this definition, the purpose of EIA for development projects is to collect and consider environmental information about the likely positive and negative impacts of the project and to report those impacts to decision makers in advance of project authorization (Weston, 2000).

The Egyptian Environmental Affairs Agency (EEAA) uses the list approach. It is a system for the management of EIA to classify the projects into three classes (A-C), reflecting different levels of EIA according to severity of possible Environmental Impact (EI), depending on the following principles:

- Type of activity performed by the project
- Extent of natural resources exploitation
- Location of establishment or project
- Type of energy used to operate the project

The classifying lists of projects into three classes represent a guide for the EIA. They are examples rather than exhaustive lists and the classification may be adjusted by the EEAA in accordance with available updated information about new projects (EEAA., 2002). This application tries to set up a model that can be used for carrying on the classification of new projects (EEAA., 2002). This application tries to set up a model that can be used for carrying on the classification of new projects easily and more accurately.

The data for this study is taken from the survey of the Knowledge, Attitude and Practices (KAP) of industry towards environment investigation which is a baseline

Corresponding Author: Aiman Majid Nassar, College of Applied Sciences-Heet, University of Anbar, Anbar, Iraq

survey conducted by El-Zanaty and associates in collaboration with Greencom (Kemprecos *et al*., 2000). This survey was undertaken to better understand and quantify the knowledge, attitudes and practices of industry regarding the environment. A sample of 1250 projects was randomly selected to represent every type of manufacturing industry in Egypt. These projects were grouped into eight categories from 6 Governorates (Cairo, Alexandria, Giza, Sharkia, Ismailia and Dakahlia).

Background information including project location (X_1), the main products of the projects (X_2) and the number of employees (X_3).

The level of using technology which is measured by the following questions: does energy represent a major cost of production (X_4). When purchasing equipment do you specify energy efficiency requirements (X_5). Do you purchase products with energy efficiency labels (X_6). The ownership of sources of pollution which are.

Air emission: From process smoke stacks (X_7) from generator or boiler smoke stack (X_8) or from fugitive emissions (X_9). Waste water from process discharges (X_{10}). Solid waste: empty packaging (X_{11}) off-spec product (X_{12}), scarp (X_{13}) and sludge (X_{14}). Since, the EIA of projects is based on several variables, it is desirable to select the most important ones needed for classifying any project in only one of the three classes (A-C). Class A contains all projects which have the low bad Environmental Impact (EI) class B contains all projects which have the medium bad EI and class C contains all projects which have the high bad EI.

To achieve this objective, the data set of all the 1250 projects with 14 variables is firstly used to find the classes using the k-means cluster method (Afifi and Clark, 1984). Ion this analysis, the Mathematical Programming (MP) cluster methods are not used because the capacity of GAMS Software is limited (Brooke *et al*., 2001). The results of applying the k-means cluster method using three clusters are 617 projects are grouped into class A; 496 into class B and 137 projects into class C.

Secondly, a training sample of 50 projects was proportionally and randomly drawn to reprsent the three classes (A-C). This sample was used to obtain a discriminant function which could be used to classify a new project into one of the three classes (A-C). After that the study selected the most important variables in the discriminant function using both the stepwise reductions method and the MP Model. The rest of the projects (1200) are used as a holdout sample.

MATERIALS AND METHODS

Variable selection in DA

Classical methods: The classical methods for variable selection in DA depend on the basic assumptions: normality, homogeneity of variance and covariance matrices and the linear independency of samples (Costanza and Afifi, 1979). They can be grouped into three categories, namely: stepwise, canonical variate and all-subset methods. The stepwise method has commonly been used in the literature as bench mark for comparison and will thus be used here.

The stepwise method is the most popular procedure used in computer packages. It's basic idea is to examine the variables one-at-a time to see if they could be selected for the DA or not. Several variants of stepwise methods are available such as forward stepwise and backward stepwise. However, the default settings usually result in forward stepwise whose process starts with a model without any variables and then attempts to add variables to the model one-at-a time. When entering a new variable, it checks to see if a previously entered variable can be removed or not. The process continues until no more variables can be entered or removed (Afifi and Clark, 1984).

If r variables are already selected in the equation and the variable X_{r-1} is to be examined to determine if it increase the separation provided by X_1, X_2, …, X_r and analysis of covariance is obtained, treating X_{r+1} as the response and X_1, X_2, …, X_r as covariates. Let the adjusted within-group and among-groups sums of squares be: $e_{r+1.123,\ldots,r}$ and $h_{r+1.123,\ldots,r}$, respectively. Then X_{r+1} provides signifcicatn additional information at level α if the partial F statistic (defined in Eq. 1) exceeds the critical value F:

$$F_{r+1.12m,\ldots,r} = (n_E - r)h_{r=1.123,\ldots,r}/(n_H) \times e_{r=1.123,\ldots,r} \quad (1)$$

Where:

n_H = m-1 denotes the number of groups

n_E = $\sum_{k=1}^{m} n_k - m; n_k$ denotes the sample size from a group k, (k = 1, 2, …, m)

If forward or backward stepwise is used it will be affected by the variable arrangement given within the data. A combination of forward and backward approaches was suggests by Hawkins (1976) and showed that for F tests the probability of concluding that there is no separation is $(1-\alpha_1)$ $(1-\alpha_2)$, ... $(1-\alpha_r)$. This is also a lower bound on the probability of successfully eliminating all of the redundant variables among those testes (McKay and Campbell, 1982a).

Disadvantages:

- The actual significance levels of individual tests are unknown
- Subsets of variables aren't compared with the original variable set
- The tests aren't independent and it is difficult to judge the magnitude of simultaneous significance for sequence of tests (McKay and Campbell, 1982b)

MP Model: The advantages of the MP approach over the classical methods for the selection of variables in DA include the following:

- The MP Model is free from underlying parametric assumptions
- It can handle varied objectives
- Some MP methods lend themselves to sensitivity analysis

El-Hefnawy (1999) introduced the general form for DA using the MP approach to discriminate between more than two groups which tries to find a hyper plane that best separates the groups. The general variable selection model in DA using the MP approach takes the following form minimize (Eq. 2):

$$Z = \sum_{i=1}^{n} d_{i12} + \sum_{i=1}^{n} d_{i12} + \sum_{h=1}^{2}\sum_{h=1}^{2}\sum_{k=2}^{m-1}\sum_{i=1}^{n} d_{ikh} \quad (2)$$

Subject to:

$$\sum_{j=1}^{p} X_{1ij}(\bar{a}_j - \bar{a}_j) - d_{i12} \le u_1; i = 1, 2, ..., n_1 \quad (3)$$

$$\sum_{j=1}^{p} X_{ij}(\bar{a}_j - \bar{a}_j) - d_{i21} \le u_1; i = 1, 2, ..., n_2 \quad (4)$$

$$\sum_{j=1}^{p} X_{2ij}(\bar{a}_j - \bar{a}_j) - d_{i22} \le u_2; i = 1, 2, ..., n_2 \quad (5)$$

$$\sum_{j=1}^{p} X_{3ij}(\bar{a}_j - \bar{a}_j) - d_{i31} \le u_2; i = 1, 2, ..., n_2 \quad (6)$$

$$\sum_{j=1}^{p} X_{3ij}(\bar{a}_j - \bar{a}_j) - d_{i32} \le u_3; i = 1, 2, ..., n_2 \quad (7)$$

$$\sum_{j=1}^{p} X_{m\text{-}iij}(\bar{a}_j - \bar{a}_j) - d_{im\text{-}11} \le u_{m-2}; i = 1, 2, ..., n_{m-1} \quad (8)$$

$$\sum_{j=1}^{p} X_{m\text{-}iij}(\bar{a}_j - \bar{a}_j) - d_{im\text{-}12} \le u_{m-1}; i = 1, 2, ..., n_{m-1} \quad (9)$$

$$\sum_{j=1}^{p} X_{iij}(\bar{a}_j - \bar{a}_j) - d_{im1} \le u_{m-1}; i = 1, 2, ..., n_m \quad (10)$$

$$\sum_{j=1}^{p} (\bar{a}_j - \bar{a}_j) = S \quad (11)$$

$$u_k - u_{k-1} \ge S_1; k = 2, 3, ..., m\text{-}1 \quad (12)$$

$$\bar{a}_j - \in \delta_j \ge 0; j = 1, 2, ..., p \quad (13)$$

$$\bar{a}_j - \in \delta_j \le 0; j = 1, 2, ..., p \quad (14)$$

$$\bar{a}_j - \in \gamma_j \ge 0; j = 1, 2, ..., p \quad (15)$$

$$\bar{a}_j - \in \gamma_j \ge 0; j = 1, 2, ..., p \quad (16)$$

$$\delta_j + \gamma_j \le 1; j = 1, 2, ..., p \quad (17)$$

$$\sum_{j=1}^{p} (\delta_j + \gamma_j) = r \quad (18)$$

where, X_{mij}= denotes the values of variable j on the observation I from group m.u_1, u_2, ..., u_{m-1} are decision variables, unrestricted in sign representing class boundaries. S is a positive constant.

S is a non-negative decision variable a_j^+, a_j^- are non-negative decision variables representing the coefficients of the discriminant function j = 1, 2, …, p δ_j and γ_j are decision variables each equal zero or one; j = 1, 2 , …, p. D_{ikh} are a non-negative decision variables, where i = 1, 2, …, n_k; k = 1, 2, …, m; h = 1, 2 which represents exterior deviations.

R represents the required number of selected variables according to the decision maker's choice it does not represent one of the decision variables.

The suggested model includes $n_1+n_m+2\sum_{k=2}^{m-1} n_k+5P+m$ constraints and $\sum_{k=2}^{m-1} n_k+4P+m+1$ decision variables of which 2p are binary variables and it can be solved by using any MP Software. This model is normalized for invariance under origin shifts by using constraints (Eq. 3-11). Constraint (Eq. 12) prevent the overlapping problem and constraints (Eq. 13-17) are required for the definitions of δ_j and γ_j while (Eq. 18) ensures that r variables have non-zero values out of the P variables.

In this model if a given variable g; (g = 1, 2, …, p) is required to be included in the discriminant function, the associated constraint in Eq. 17 is replaced by:

$$\delta_g + \gamma_g = 1 \quad (19)$$

Table 1: Selection of variables using the MP Model

No. of selection variables	Variables selected by the MP Model	Correct classification (%)	Boundaries	
			U_1	U_2
1	X_6	52	0.071	0.571
2	X_3, X_6	94	0.778	1.278
3	X_3, X_6, X_8	98	0.146	0.646
4	X_3, X_6, X_7, X_{11}	96	2.01	3.307
5	$X_3, X_6, X_4, X_7, X_{11}$	96	0.219	0.719
6	$X_3, X_4, X_6, X_8, X_{11}, X_{12}$	98	0.442	0.942
7	$X_3, X_4, X_6, X_{10}, X_{11}, X_{12}, X_{14}$	94	0.510	1.010
8	$X_1, X_3, X_4, X_6, X_8, X_{10}, X_{11}, X_{13}$	96	0.638	1.138
9	$X_3, X_4, X_5, X_6, X_8, X_{10}, X_{11}, X_{13}, X_{14}$	94	0.196	0.696
10	$X_1, X_2, X_3, X_4, X_6, X_8, X_{10}, X_{11}, X_{13}, X_{14}$	96	0.186	0.686
11	$X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{10}, X_{11}, X_{12}, X_{14}$	98	0.711	1.211
12	$X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}$	98	0.219	0.719
13	$X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}$	96	0.269	0.769
14	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}$	96	1.007	1.507

This model is published by Mohamed *et al*. It is a general Minimizing Sum Deviations (MSD) Model for the selection of variables in DA. The special case obtained by putting m = 2, represents the second MSD Model by Glen (1999).

The solution procedure starts by solving the model at r = 1. The corresponding discriminant function is obtained and used to discriminate the projects into the three classes (A-C). The overall percent of correct classification is obtained. The process is repeated for r = 2 ,..., K. The model with the highest percent of correct classification and lest number of variables is selected for application.

The application: To apply the described MP Model for this set of real data with the three-classes discriminant problem (m = 3), the values of ϵ and S_1 are chosen to be 0.01 and 0.5, respectively.

Table 1 gives in addition to the selected variables, the percentages of correct classification of projects in the training sample which corresponds to every choses subset of the environmental variables for EIA and the corresponding boundaries according to which a project is classified. If the project discriminant score is less than or equal to U_1, the project is classified into class A. It is considered as belonging to class B if its discriminant score is greater than U_1 and less than or equal to U_2 and is considered in class C if its discriminant score is greater than U_2. The value of the corresponding objective function (Z) which aims to minimize the sum of exterior deviations is in general equal to zero.

It is concluded from Table 1 that for example the variables selected at the second step (when, r = 2) are X_3 and X_6 and that the correct classification percentage is 94%. It increases to 98% when the subset of selected variables of size 3, 6, 11 and 12. This represents the maximum correct classification percentage and is reduced to 96% when the number of selected variables increases to 13 and 14 variables. This result demonstrates that the classification performance does not necessarily improve as the number of variables is increased. This result is the same as that reached by Glen (1999). Moreover, it is noticed that the smaller subsets of variables do not have to be subsets of larger ones.

RESULTS AND DISCUSSION

The results of applying the stepwise method are given listed as follows: the initial F's-to-enter are 937.5 for X_6 and 83.8 for X_3. Thus, the variable X_6 is the first variable to be included in the discriminant function followed by X_3 and no other variable can enter or be removed, i.e., the variables and X_{14} are excluded from the discriminant function. The printout of the SPSS Program gives the two sets of unstandardized discriminant function coefficients for the classes of projects. Therefore, the two functions to discriminate between the three classes (A-C) are. First discriminant function from stepwise:

$$D_1 = -3.533+0.323X_3+2.457X_6$$

Second discriminant function from stepwise:

$$D_2 = -4.349+2.26X_3-0.068X_6$$

Table 2: Classification results from stepwise method and the MP Model with 2 variables on the training sample

	Predicted group membership using MP with 2 variables				Predicted group membership using stepwise			
Classes	A	B	C	Total	A	B	C	Total
A	24	-	-	24	20	4	-	24
B	1	19	-	20	-	20	-	20
C	-	2	4	60	2	-	4	60
Total	25	21	4	50	22	24	4	50

Table 3: Classification results from the MP Model with 2 variables and stepwise on the holdout sample

	Predicted group membership using MP with 2 variables				Predicted group membership using stepwise			
Classes	A	B	C	Total	A	B	C	Total
A	570	23	-	593	431	102	60	593
B	51	424	1	476	-	470	60	476
C	11	46	74	131	51	200	78	131
Total	632	493	75	1200	482	574	144	1200

Table 4: Classification results from the MP Model with 3 variables on the training sample and the holdout samples

	Predicted group membership on training sample				Predicted group membership on holdout sample			
Classes	A	B	C	Total	A	B	C	Total
A	24	-	-	24	553	40	-	593
B	1	19	-	20	22	438	16	476
C	-	-	6	6	2	4	125	131
Total	25	19	6	50	577	482	141	1200

The classification rule to classify a new observation X_0 is defined as follows: classify an observation X_0 (X_{01}, X_{02} X_{0p}) in:

- Group 1 if $D_{10}>0$ and $D_{20}>0$
- Group 2 if $D_{10}<0$ and $(D_{10}-D_{20})>0$
- Group 3 if $D_{10}>0$ and $(D_{10}-D_{20})<0$

where, D_{i0}; i = 1, 2 represents the discriminant scores for the observation X_0, obtained by direct substitution with its variables value. The classification of projects into classes is based on the discriminant score of a project on the two discriminant functions simultaneously. The overall percentage of correctly classified projects in the training sample is 88%.

Now, to compare performance of the stepwise method and the MP Model, the subset of the variables X_3 and X_6 is used. The corresponding discriminant function using the MP Model is:

$$D_2 = 0.389X_3+0.111X_6 \text{ with } U_1 = 0.778 \text{ and } U_2 = 1.278$$

Table 2 gives the numbers of correctly classified projects according to the two methods for the training sample. It shows that the correct classification percentage (for the training sample) is 94% for the MP Model with 2 variables and 88% for the stepwise method.

Table 3 shows the classification results from the MP Model with two variables X_3 and X_6 and from the stepwise method on the holdout sample. It shows that the correct classifications percentage from the MP Model with 2 variables on the holdout sample is 89%, while the stepwise model gives 81.6% correctly classified.

Since, the aim of this study is to reduce the number of selected variables as much as possible and at the same time to increase the accuracy of the model, then selecting the three variables X_3, X_6, X_8, seems to be the best choice. The corresponding discriminant function is given by:

$$D_j = 0.073X_3+0.073X_6+0.01X_8 \text{ with } U_1 = 0.146 \text{ and } U_2 = 0.646$$

The classification results from training and holdout samples are given in the following Table 4. It can be seen from Table 4 that correct classification percentage using the MP Model with three variables in 93% for the holdout sample which is greater than the correct classification percentages of both the stepwise method and the MP model with two variables. The analysis suggests that projects can be classified into classes (A-C) according to the number of employees employed (X_3) whether the project purchases products with energy efficiency labels or not (X_6) and wither the air emissions is from generator or boiler

Table 5: Selection of variables using modified MP Model

No. of selection variables	Variables selected by modified MP Model	Correct classification (%)	Boundaries	
			U_1	U_2
3	X_1, X_2, X_3	84	0.155	0.455
4	X_1, X_2, X_3, X_4	94	0.176	0.676
5	$X_1, X_2, X_3, X_6, X_{11}$	98	0.117	0.417
6	$X_1, X_2, X_3, X_6, X_8, X_{13}$	96	0.882	1.382
7	$X_1, X_2, X_3, X_4, X_6, X_8, X_{10}$	96	0.455	0.955
8	$X_1, X_2, X_3, X_4, X_6, X_7, X_8, X_{13}$	94	2.535	3.751
9	$X_1, X_2, X_3, X_5, X_6, X_7, X_{10}, X_{12}, X_{14}$	94	0.669	1.169
10	$X_1, X_2, X_3, X_5, X_6, X_8, X_{10}, X_{11}, X_{13}, X_{14}$	96	0.232	0.732
11	$X_1, X_2, X_3, X_4, X_6, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}$	98	0.229	0.729
12	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{11}, X_{12}, X_{13}, X_{14}$	98	0.565	1.065
13	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}$	98	0.226	0.726
14	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}$	96	1.007	1.507

Table 6: Classification results from MP model modified MP Model and stepwise on the holdout sample

	MP Model with 3 variables			Modified MP with 5 variables			Stepwise with 2 variables			
Classes	A	B	C	A	B	C	A	B	C	Total
A	553	40	-	536	550	2	431	102	60	593
B	22	438	16	30	442	4	-	470	6	476
C	2	4	125	-	29	102	51	2	78	131
Total	577	482	141	566	526	108	482	475	144	1200

smoke stack or not (X_8) using only these three variables 93% of the projects will be correctly classified.

The modified MP Model: The variable selection methods applied do not take the nature of the variable into consideration. Environmental theoretical studies suggest that some variables are important and cannot be ignored in any environmental study. Inclusion of such variables in the model is necessary. This presumably will lead to a lower correct classification percentage.

One of the advantages of the MP Model is that it can force a certain variable, g (or variables) to be selected by replacing its associated constrains in (Eq. 19) with the following constraint:

$$\delta_g + \gamma_g = 1$$

This property is used here to force the variables X_1-X_3, (Project location, the main products of the project and the number of employees, respectively) to be included in the selection, since, they were recommended by environmental experts in EEAA and therefore, the best discriminant function should include them. By modifying the suggested MP Model with the previous constraint, it can be called the "Modified MP Model". By applying the "Modified MP Model" and using the number of selected variables ($r = 3, 4, \ldots, 14$) with $\varepsilon = 0.01$, $S_1 = 0.3$, the following table is obtained.

It can be seen from Table 5 that the best choice has five selected variables, three of them are determined before applying the model and two variables X_6 and X_{11} are chosen by the model. The following discriminant function is obtained.The modified MP discriminant function with 5 variables:

$$D_5 = 0.021X_1 + 0.021X_2 + 0.043X_3 + 0.043X_6 + 0.01X_{11}$$

With the boundaries: $U_1 = 0.117$, U_2 0.417. The variables in this function are: X_1 represents the project location it takes "zero" if the project location lies in an industrial city and "one" if the project location is in a sensitive city.

X_2 represent the main products this project produces where "one" indicates black industry (chemicals, mining and building materials and metal industries); "zero" indicates gray industry (textiles and leather, paper and publishing and engineering equipment) and "negative one" indicates white industry (food and beverages and wood and furniture).

X_3 represents the number of employees in the project; it takes the value "one" if the number of employees is 10-49 the value "two" if the number of employees is 50.99 and the value "three" if the number

of employees is 100 or more. X_6 is a binary variable that takes "zero" if the project purchases products with energy efficiency labels and "one" if not. X_{11} is a binary variable that takes "one" if the project has empty packaging and "zero" if not.

The ability of the model to correctly allocate a new project is examined and compared. Table 6 shows the classification results from both the stepwise method and the MP Models on the holdout sample.

Results obtained on the holdout sample, displayed in Table 6, indicate that the performances of the models are similar to those obtained on the training sample where the MP Model with 3 variables (r = 3) gives 93% correctly classified and the modified MP Model with 5 variables (r =5) gives 90% correctly classified. This means that the MP Models in general give better results compared to the stepwise method.

CONCLUSION

One of the main goals of the Egyptian Environmental Affairs Agency (EEAA) is to classify projects into three classes (A-C) according to severity of possible environmental impact. Classification of the projects is based on a number of variables. The importance of correctly selecting these variables cannot be overemphasized.

The data for this study is taken from the Egypt Environmental Impact Assessment Survey 2000 data. A sample of 1250 projects was randomly selected to represent every type of manufacturing industry in Egypt. The data collected included 14 environmental variables. This study applies a MP Model to select the variables that can best discriminate between the three groups that is the fewest variables with the highest correct classification percentage. Out of the 14 variables the model selected three variables, namely, the number of employees in the project whether the project purchases products with energy efficiency labels or not and whether the air emissions is from generator or boiler smoke stack or not. The analysis suggests that using only these three variables, 93% of the projects are expected be correctly classified.

During the selection of variables the nature of the variables and their importance from the environmental point of view are not taken into consideration. This might lead to specification bias. Inclusion of these variables in the model becomes necessary even if they were not originally selected by the model. In this study these variables include the project location, the main products of the project and the number of employees.

After forcing these variables to be included in the selected variables the model adds the variables: whether or not the project purchase products with energy efficiency labels and if the project has empty packaging or not. These represent the main variables that affect the classification of any project as corresponding to its Environmental Impact (EI) after taking the environmental theoretical point of view into account.

SUGGESTIONS

The suggested MP Model proved most effective in correctly selecting the variables leading to a better classification of projects compared to the stepwise method which is commonly applied in variables selection problems. The MP Model has other advantage over the stepwise method. For example it is able to classify any new project, depending on certain variables, just by substituting for values of variables in one suitable discriminant function and comparing the resultant value with the computed boundaries. The stepwise method requires two discriminant functions. In addition, the objective function in the MP Model, can be modified to include priority or weights for any class if the decision maker chooses to do so.

Furthermore, in the MP Model, the discriminant function with r-variables is derived directly from the solution to the MP Model using GAMS Software. It can also be seen that the classification performance does not necessarily improve as the number of variables is increased and it is not necessary that the elements of the smaller subset should be found in the larger one.

The MP Model is recommended to be used by the Egyptian Environmental Affairs Agency to classify new projects into their classes, so, corrective actions can be taken as early as possible to avoid possible undesired environmental problems.

REFERENCES

Afifi, A.A. and V. Clark, 1984. Computer-Aided Multivariate Analysis. Lifetime Learning Publications, Belmont, CA USA.

Brooke, A., D. Kendrick and A. Meeraus, 2001. GAMS: A Uswers Guide. The Scientific Press, Redwood City, Californ, USA.,.

Costanza, M.C. and A.A. Afifi, 1979. Comparison of stopping rules in forward stepwise discriminant analysis. J. Am. Statist. Assoc., 74: 777-785.

EEAA., 2002. Guidelines for Egyptian environmental impact assessment. Egyptian Environmental Affairs Agency, Cairo, Egypt.

El-Hefnawy, A.E., 1999. Mathematical programming approach to discriminant analysis with application in demography. Ph.D Thesis, Cairo University, Giza, Egypt.

Glen, J.J., 1999. Integer programming methods for normalisation and variable selection in mathematical programming discriminant analysis models. J. Oper. Res. Soc., 50: 1043-1053.

Hawkins, D.M., 1976. The subset problem in multivariate analysis of variance. J. Royal Stat. Soc. Ser. Methodol., 38: 132-139.

Kemprecos, L., O. Hernandez, F. El-Zanaty and R. Hamed, 2000. Environmental management systems, energy efficiency and waste management: A cross-sectoral KAP study of Egyptian firms. GreenCom, Egyptian Environmental Policy Program and The Egyptian Environmental Affairs Agency, Cairo, Egypt. https://webcache.googleusercontent.com/search?q=cache:TXWJmSZHCusJ:https://rmportal.net/library/content/usaid-greencom/greencom-reports/environmen

McKay, R.J. and N.A. Campbell, 1982a. Variable selection techniques in discriminant analysis: I. Description. Br. J. Math. Stat. Psychol., 35: 1-29.

McKay, R.J. and N.A. Campbell, 1982b. Variable selection techniques in discriminant analysis: II. Allocation. Br. J. Math. Stat. Psychol., 35: 30-41.

Palerm, J.R., 2000. An empirical-theoretical analysis framework for public participation in environmental impact assessment. J. Environ. Plann. Manage., 43: 581-600.

Weston, J., 2000. EIA, decision-making theory and screening and scoping in UK practice. J. Environ. Plann. Manage., 43: 185-203.