THE EGYPTIAN STATISTICAL JOURNAL
ISSR, CAIRO UNIV., VOL., 42, NO.1, 1998

On the Performance of Some Methods for Handling Missing Values in Classification Analysis.

by

D. Adel M. Zaher *D.* Ahmed H. Haroun Mahmoud A. Mahmoud

Abstract

Classification Analysis is concerned with the problem of classifying a subject to one of several distinct groups on the basis of a set of measurements. For example, in business, a bank loan officer wishes to classify loan applicants to low risk credit customers or high risk credit customers on the basis of a set of variables.

The presence of missing values in a data set used for building a classification rule is a serious problem that may face the investigator when applying the classification analysis (CA) to a practical situation. Many procedures have been developed to handle the missing values when applying (CA). The default method of handling missing values in (CA) used by many statistical packages (for example, SAS, Minitab and SPSS) is to omit all units containing missing values. Thus, considerable information may be lost due to the reduction of the sample size.

Many studies dealt with this problem in case of two multivariate populations with equal covariance matrices while a few studies treated it in case of two multivariate populations with unequal covariance matrices. The present study deals with the problem of classification analysis with missing values in case of two or three multivariate normal populations with equal and unequal covariance matrices through a simulation study. Three rules of classification and five methods of handling missing values are considered. The objective of this study is to compare the different methods of handling missing values with respect to their ability in obtaining a "good" classification rule. In this study, two patterns of missing values are considered and the mechanism that lead to the presence of missing values is assumed to be missing at random (MAR).

Seven factors are taken into consideration. The impact of each factor on the methods of handling missing values is studied. A Minitab macro was designed to run the necessary calculations.

1.Introduction.

The problem considered in (CA) is classifying a subject to one of several distinct groups on the basis of a set of measurements associated with the subject. When we want to build a classification rule to classify a subject, a serious problem may face us. This problem is the presence of missing values in the data set which is used to build the classification rule. This may occur in many practical situations, for example, some respondents may refuse to answer a particular question on a survey questionnaire. Also, the researcher may forget to record an answer or a measurement.

Many statisticians investigated the performance of the methods used for handling missing values in multivariate data analysis. Most of their studies were devoted to multivariate regression while a few statisticians compared this performance in (CA) . A review of related literature is presented in Section (2).

Many rules have been developed to solve the classification problems. The rules of classification used in this study are presented in Section (3). Section (4) is devoted to display the procedures of handling missing values suitable for multivariate data analysis. Methodology and tools are explained in Section (5). In this section, we will discuss the methods of choosing parameters, generating samples, and drawing units to be classified. The results of the simulation study are presented in Section (6) while summary and conclusions are given in Section (7).

2. Review of Related Literature.

In (1968), Jackson compared three methods of handling missing values using real data. The methods used in her study were, the complete case method, the mean substitution method and the regression estimation method. In the first method, the units containing missing values are discarded and the remaining units used in building the classification rule. In the second and the third methods, the missing values are replaced by estimated values and the classification rule is designed by the completed data set. Jackson used the Fisher's discriminant function as the classification rule. She pointed out that eliminating units with missing data may result in considerable bias if missing data are not randomly distributed.

In (1972), Chan and Dunn used simulation techniques to investigate the performance of seven methods of handling missing values for small samples . They assumed that all pairs of variables are equally correlated. Methods used in their study were:

Method A : Only complete data vectors, are used.

Method B :All available sample values are used to calculate means and covariances.
Method BA : If method B yields a negative definite covariance matrix, use method A, otherwise use method B.
Method C :The mean substitution method.
Method D :The regression estimation method.
Method DS :A modification of method D.
Method E : The principal component method.

Chan and Dunn compared the performance of these methods on the basis of the expected probability of correct classification calculated from Fisher's discriminant function. They found that method C and method E were in general superior to the other methods .

In (1976), Chan and Dunn studied the performance of seven methods of handling missing values using Fisher's discriminant function . They did not restrict themselves to equal correlation matrices and used methods A, B, C, D, E and two additional methods for handling missing values. The additional methods were:

Method D*: A modification of the regression method D .
Method E* : A modification of the principal component method E.

They found that method D* is almost always better than methods C , E , B , E* and performs approximately the same as method D . Method A is better for small values of k ,where k is the number of variables.

Little (1978), suggested two consistent regression methods for handling missing values in (CA) . The two methods are modifications of method D mentioned previously .

In (1992), Little collected all procedures of handling missing values suitable for the multivariate data analysis. He grouped these procedures in sex classes. In his study, the treating of missing values using each class of procedures is discussed when applying multivariate regression analysis.

Twedt and Gill (1992), examined the performance of four methods of handling missing values in case of equal and unequal covariance matrices. The methods used for handling missing values in their study were:

Method NR: Only complete units are used (non-replacement method).
Method MR: The mean substitution method.
Method PC : The principal component method.
Method EM: The expectation maximization algorithm.

The rules of classification used in their study are Fisher's discriminant function and the quadratic discriminant function. They found that methods MR, PC and EM (replacement methods) are better than the non-replacement method NR, and that the differences among these three replacement methods are slight.

All the previous studies were conducted under the condition that the mechanism that leads to missing values is "missing at random (MAR)".

3.Classification Analysis (CA).

In many practical situations, a problem occurs in which an observation has to be assigned to one of several populations. For example, a firm manager wishes to classify workers as skilled labor or unskilled labor on the basis of a set of related attributes of these workers. Also in medicine, a psychotherapist wants to predict whether an individual is more or less likely to be depressed on the basis of readily available information about this individual. The common attribute of the previous examples is that one needs to classify a subject to one of several distinct groups.

Thus, the problem considered in (CA) is the following : given that a subject is known to come from one of g distinct groups (populations), we wish to assign the subject to one of these groups on the basis of a set of measurements associated with the subject. The vector of measurements associated with a subject is usually called its profile. Many rules have been developed to solve the classification problem. This section presents an overview of the classification rules used in this study.

A good classification rule should result in a small number of misclassifications. In other words, the probability of correct classification should be large. The use of the probability of correct classification as a way of judging the performance of any classification rule will be discussed in this section.

3.1 Rules of Classification Used in This Study.

Let $\underline{X}$ be a vector of p measurements associated with a subject, i.e. $\underline{X}$ is the subject's profile. In order to classify $\underline{X}$ into one of g populations, Π1, Π2, ..., Πk, many classification rules were developed to do so. The rules used in this study are presented in the following subsections.

3.1.1 The Generalized Distance Rule (GDR).

The generalized distance rule (GDR), which was developed by Mahalanobis (1936), classifies the subject with profile $\underline{X}$ to the kth group (Π_k ; k=1,2,..,g) if the square distance between $\underline{X}$ and the vector of means for group k is less than the square distance between $\underline{X}$ and the vector of means for any other group. The squared distance between $\underline{X}$ and the kth group is defined as :

$$(D_k)^2 = (\underline{X} - \mu_k)' \Sigma_k^{-1} (\underline{X} - \mu_k) \qquad k=1,2,.....,g, \qquad (1).$$

where μ_k is the vector of means in group Π_k and Σ_k is the covariance matrix in the same group .

However, in practice, μ_k and Σ_k are usually unknown and then their sample estimates are used.

Another important issue related to the concept of the generalized distance is the squared distance between two populations, which is called the Mahalanobis distance between two populations. The Mahalanobis distance between population (i) and population (j) is defined as:

$$(\Delta_{ij})^2 = (\mu_i - \mu_j)'\Sigma_p^{-1} (\mu_i - \mu_j), \quad i,j=1,2,.....,g, \quad i\neq j, \qquad (2).$$

where Σ_p is the pooled covariance matrix.

The Mahalanobis distance $(\Delta_{ij})^2$ between population (i) and population (j) is a statistical measure for the separation between these populations. A large value of $(\Delta_{ij})^2$ means a large separation between the two populations and, consequently, a large ability of a classification rule to correctly classify subjects into populations.

3.1.2 The Quadratic Discriminant Function Rule (QDF).

The quadratic discriminant function (QDF) rule, which was introduced by Welch (1939), is sometimes called the maximum-likelihood classification rule. According to this rule, we allocate $\underline{X}$ to population Π_k if the likelihood function of $\underline{X}$ coming from Π_k (k=1,2,...,g) is greater than the likelihood function of it coming from any other population.

If we assume that in the kth population, $\underline{X}$ has a p-variate normal distribution with vector of means μ_k and covariance matrix Σ_k, k=1,2,..........,g, then the (QDF) rule becomes : allocate $\underline{X}$ to Π_k if $(D'_k)^2 = \min [(D'_1)^2, (D'_2)^2,....., (D'_g)^2]$, where

$$(D'_i)^2 = (D_i)^2 + \ln |\Sigma_i|, \qquad i=1,2,....,g. \qquad (3).$$

If the parameters are unknown, then their samples estimates are used.

3.1.3 The Linear Discriminant Function Rule (LDF).

The linear discriminant function rule(LDF) is a special case of the (QDF) rule if the distributions of $\underline{X}$ in the populations are multivariate normal with equal covariance matrices. If we assume that for the kth population Π_k, $\underline{X}$ has a p-variate normal distribution with mean vector μ_k and the covariance matrix Σ, then the LDF rule is as follows: allocate $\underline{X}$ to Π_k if $(D''_k)^2 = \min[(D''_1)^2, (D''_2)^2, \ldots, (D''_g)^2]$, where

$$(D''_i)^2 = (\underline{X} - \mu_i)' \Sigma^{-1} (\underline{X} - \mu_i), \quad i = 1,2,\ldots,g. \qquad (4).$$

3.2 Estimating the Probability of Correct Classification.

One important way of judging the performance of any classification rule is to calculate its ability in correctly classifying subjects. This ability is measured in terms of an important concept called " the probability of correct classification ". Often, we depend on sample statistics in building the classification rule. In this case, a way of estimating the probability of correct classification is needed. The apparent probability of correct classification (APCC), which is sometimes called the "hit rate" is the simplest way of estimating this probability. To evaluate the (APCC), suppose that g samples are generated from g populations and the classification rule is built using the statistics calculated from these samples. Also, suppose that from each population, n_i (i=1,2,.,g) observations are drawn. These observations are classified using the classification rule which was built from the g samples. If the number of observations correctly classified in population Π_i is m_i then, the apparent probability of correct classification is given by:

$$\mathrm{APCC} = \sum_{i=1}^{g} m_i \Big/ \sum_{i=1}^{g} n_i . \qquad (5).$$

For more details, the reader is refereed to Lachanbruch (1975), Stevens (1986, Chapter 7) and Richard and Wichern (1992, Chapter 11).

4. Procedure of Handling Missing Values

Many multivariate statistical techniques, Such as factor analysis, regression analysis and classification analysis, are based on calculating the sample statistics; mean vector and sample covariance matrix of the variables. When some observations in the data set are unavailable, the question of how to estimate these statistics is very important. When we talk about the classification analysis, this question becomes; how to estimate these statistics in order to build a good classification rule, i.e., to build a classification rule that has a high probability of correct classification.

In real-life situations, there are many reasons that can lead to the presence of missing values in the data set. For example; respondents in a household survey may refuse to report income. Another example is the negligence in recording a particular item on a survey questionnaire. The nature of the reason that leads to the presence of missing values is usually called "the mechanism of missing values". The style of the spread of missing values in a data set is called " the pattern of missing values". In this study the mechanism that leads to missing values is assumed to be missing at random (MAR). Also, two patterns of missing values (the univariate and the bivariate pattern of missing values) are chosen. For more details about the mechanisms and the patterns of missing values, the reader refereed to Little (1992).

Many procedures have been proposed to handle the missing values in the data set when applying (CA). The procedures of handling missing values used in this study are presented in the following subsections.

4.1 The Complete-Case (CC) Procedure.

The default method of handling missing values in a multivariate statistical analysis by many statistical packages, such as SAS, Minitab and SPSS is the (CC) procedure. According to this procedure, units containing missing values are discarded and the required statistics (mean vector and the covariance matrix) are obtained from the complete cases only. This procedure is sometimes called the "listwise deletion procedure".

4.2 The Estimation Procedure.

The most important procedure of handling missing values is to substitute them by their estimated values obtained by using the available data. A classification rule is then built using the statistics estimated from the completed data set (observed and estimated).

Many methods have been developed to estimate the missing values in multivariate statistical analysis. A brief introduction to the methods of estimation used in this study is presented in the following subsections.

4.2.1 The Mean Substitution Method (MS).

According to the mean substitution method (MS) which was proposed by Wilks (1932), we calculate means from all available sample values and substitute them for the missing values. This method yields a biased estimate of the covariance matrix and ,in general , it is not recommended in applications.

4.2.2 The Principal Component Estimation Method (PC).

In (1959), Dear proposed an alternative method to treat the problem of missing values which is suitable for the multivariate data analysis. He based the estimation of the missing values on a principal component analysis. The idea of Dear's method is to convert the original data matrix X into the standardized matrix Y, where $y_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_{ii}}}$ for the observed values and $y_{ij} = 0$ for the missing values, i=1,2,...,p, j=1,2,....,n , p is the number of variables and n is the sample size. Then, the coefficients of the first principal component of Y are obtained; these coefficients may be denoted by q , where q =($q_1, q_2, .., q_p$) is the eigenvector of unit length associated with the largest eigenvalue of the product matrix $Y^{\backslash}Y$. Then, if x_{ij} is missing, the value of y_{ij} is replaced by the value of $(a_j q_i)$, where

$$a_j = \sum_{i=1}^{p} q_i y_{ij} \quad , \qquad i=1,2,...,p, \qquad j=1,2,...,n \qquad (4.1).$$

After that, the matrix Y is transformed back to the original matrix X , where $x_{ij} = y_{ij}\sqrt{s_{ii}} + \bar{x}_i$. Several simulation studies have concluded that the (PC) method, in general, is better than the (MS) method.

4.2.3 The Regression Estimation Method (RG).

A more promising method for estimating the missing values is to replace them by their estimates using an estimated linear regression equation. This method was proposed by Buck (1960). The basic idea in this method is that the estimated regression equation is obtained by regressing the variable with missing values on all the remaining ones using the complete data set.

In general, the (RG) method yields reasonable estimates of means especially when the multivariate normality assumption satisfied. On the other hand, this method yields a biased estimate for the covariance matrix. But this bias is less than the bias of the (MS) method.

In (1976), Chan and Dunn modified Buck's method. Instead of using only the complete cases in performing the regression function, Chan and Dunn proposed that all missing values are firstly replaced by the variables means and then the regression function is obtained using all data (observed and estimated). Chan and Dunn used the (RG) method and its modification in a simulation study. They found that the modified method is slightly better than Buck's method. In this study we used the modified method.

4.2.4 The Expectation Maximization Algorithm (EM).

The (EM) algorithm is an iterative method This algorithm was proposed by Dempster, Laird and Rubin in (1977). The iteration of this algorithm consists of two steps; the M step and the E step. In the M step, the maximum likelihood estimation of θ (the parameters vector) is performed depending on the complete cases. In the E step, assuming that the current estimated parameters are true, the expected value of the missing data given the observed data and the current parameters is computed.

Hence, the M step is to find $\hat{\theta}$ (the maximum likelihood estimator of θ) and the E step is to find $E(X_{missing} / X_{observed}, \hat{\theta})$. The calculation cycles from one step to the other until the current estimates do not differ appreciably from the one obtained in the previous iteration. One can summarize the steps of this algorithm as follows:

1- The parameters of the underlying distribution are estimated from the available data.

2- The missing values are replaced by estimated values depending on the observed data and the current estimated parameters.

3- The parameters are reestimated.

4- The missing values are reestimated, and so forth iterating until stability is achieved.

For more details, Richard and Wichern (1992) illustrated the computational aspects of this algorithm when the multivariate normality assumption holds through a numerical example (pp.204).

5. Methodology and Tools

Using a simulation technique, the performance of the five methods of handling missing values, mentioned before, in terms of their ability in building a good classification rule is examined. Furthermore, the three rules of classification are used to classify subjects into groups. The probabilities of correct classification are estimated for the three rules of classification before and after the presence of missing values using the APCC method which was described in Section (3.2). Besides the five methods of handling missing values and the three rules of classification, seven factors are taken into consideration. These factors are the sample size (n), the number of variables (p), the number of groups (g), the Mahalanobis distance (Δ^2), the covariance matrices (equal or unequal), the pattern of missing values (the univariate or the bivariate pattern) and the percentage of missing values (m%).

The samples are generated from populations assumed to be multivariate normal with equal and unequal covariance matrices. Subjects (units) are classified into either two or three populations (groups). In this section, the methods of choosing parameters, generating samples and drawing units to be classified will be presented.

5.1 The Two Populations Case.

In this case, two situations are considered ; equal covariance matrices and unequal covariance matrices. Methods of choosing parameters and drawing samples in each situation are presented in the following subsections .

5.1.1 Equal Covariance Matrices.

The Mahalanobis distance ,which is given as:

$$\Delta^2 \; (\mu_2 - \mu_1)' \, \Sigma_p^{-1} \, (\mu_2 - \mu_1) \quad , \qquad (6).$$

where μ_i is the vector of means in population (i) (i=1,2) and Σ_p is the pooled covariance matrix, is a statistical measure for the distance between two populations. In this study, the values 1, 4 and 16 are chosen for Δ^2 .

The steps of choosing parameters can be summarized as follows :

1) The number of variables (p) is determined , where p= 2 , 3 , 4 and 5 .

2) Without loss of generality , the covariance matrix Σ is taken to be the correlation matrix , i.e. $\Sigma = [\rho_{ij}]$, i, j =1,2,.........,p , where $\rho_{ij}=1$ for (i=j) and ρ_{ij} is a value generated from uniform (-1,1) for (i ≠j).

3) Three values are chosen for the Mahalanobis distance(Δ^2); Δ^2=1,4 and 16 .

4) Without loss of generality μ_1 is taken to be $\underline{0}$.Thus the Mahalanobis distance will take the form :

$$\Delta^2 = \mu_2' \, \Sigma_p^{-1} \, \mu_2 \qquad (7).$$

5) μ_2 is chosen to satisfy equation (7) .

After choosing the parameters using the previous steps, two samples are randomly generated from normal ($\underline{0}$, Σ) and normal (μ_2 , Σ) with equal sample size using the statistical package Statgraph (Ver. 7). We choose the values 20, 40 and 100 for the sample size . From each population 500 units are randomly drawn and classified into the two populations according to the classification rule obtained from the sample data. The estimated probability of correct classification is obtained for each classification rule.

5.1.2 Unequal Covariance Matrices.

The difference between this case and the previous one is that, the covariance matrices Σ_1 and Σ_2 are selected in an unequal fashion. These covariance matrices are selected as Σ in Section (5.1.1). Assuming that $\frac{N_1}{N_2} \approx 1$, where N_i (i=1,2) is the size of population (i), then the pooled covariance matrix Σ_p will approximately equal to($\Sigma_1 + \Sigma_2$)/2 and Equation (7) is changed to

$$\Delta^2 = \mu_2^\backslash \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} \mu_2 \qquad (8)$$

Then μ_2 is selected to satisfy Equation (5.3). After choosing the parameters we proceed as in Section (5.1.1).

5.2 The Three Populations Case.

Two situations are also considered in this case; the equal covariance matrices and the unequal covariance matrices situations.

5.2.1 Equal Covariance Matrices.

The Mahalanobis distance between two populations in this case takes the form :

$$\Delta^2_{ij} = (\mu_j - \mu_i)^\backslash \Sigma_p^{-1} (\mu_j - \mu_i), \quad i = 1, 2 \quad j = 2, 3 \text{ and } i \neq j \qquad (9)$$

The steps of choosing parameters in this case are as follows :

1) After choosing the number of variables (p), the covariance matrix Σ is selected as in Section (5.1.1).

2) Assuming that $\mu_1 = \underline{0}$, then the Mahalanobis distance between population (1) and population (2) will take the form:

$$\Delta^2_{12} = \mu_2^\backslash \Sigma_p^{-1} \mu_2 \quad , \qquad (10).$$

and between population (1) and population (3) will take the form :

$$\Delta^2_{13} = \mu_3^\backslash \Sigma_p^{-1} \mu_3 \qquad (11).$$

3) Δ^2_{12} and Δ^2_{13} are taken to be equal , i.e. , $\Delta^2_{12} = \Delta^2_{13} = 1,4$ and 16.

4) μ_2 is selected to satisfy equation (10).

5) μ_3 is selected to satisfy equation (11).

After choosing the parameters, a sample of size n (n=20 , 40,100) is generated from each of the three populations: normal ($\underline{0}$, Σ), normal (μ_2 , Σ) and normal (μ_3, Σ). 500 units are randomly drawn from each population and classified using the three rules of classification.

5.2.2 Unequal Covariance Matrices.

The covariance matrices in this case are chosen in an unequal fashion. Each matrix of them is selected as Σ in Section (5.1.1).

Assuming that $\frac{N_1}{N_2} \approx 1$ and $\frac{N_1}{N_3} \approx 1$, where N_i (i=1,2,3) is the size of population (i), then the pooled covariance matrix Σ_p in Equation(10) will approximately equal to($\Sigma_1 + \Sigma_2$)/2 and Equation (10) is changed to

$$\Delta^2_{12} = \mu_2{}' \left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1} \mu_2 \,, \qquad (12).$$

while Σ_p in Equation(11) will approximately equal to$(\Sigma_1+\Sigma_3)/2$ and Equation (11) is changed to

$$\Delta^2_{12} = \mu_3 \left(\frac{\Sigma_1+\Sigma_3}{2}\right)^{-1} \mu_3 \,. \qquad (13).$$

Then μ_2 and μ_3 are chosen to satisfy Equation (12) and Equation (13) respectively. After choosing the parameters, we proceed as in Section (5.2.1).

5.3 Estimating the Probability of Correct Classification in the Complete Data Set.

In the complete data case, five factors are considered. These factors and their levels are as follows:

1- The Mahalanobis distance Δ^2 ; $\Delta^2 =$ 1, 4 or 16.

2- The number of variables (p) ; p = 2, 3, 4 or 5.

3- The number of populations (groups) g ; g= 2 or 3.

4- The sample size (n); n = 20, 40 or 100.

5- The covariance matrices; equal or unequal covariance matrices.

Thus, there are 144 combinations of these factor's levels that are considered in this case. For each combination, the three rules of classification are applied and the probability of correct classification is estimated. A Minitab macro was designed for performing the necessary calculations.

5.4 Estimating the Probability of Correct Classification in the Incomplete Data Set.

After calculating the estimated probabilities of correct classification in the complete data set, m% of the sample size are randomly deleted from each sample. In this case two factors are considered besides the five factors mentioned in Section (5.3). These factors are the pattern of missing values and the percentage of missing values. Two patterns of missing values are considered in this study; the univariate missing values and the bivariate missing values. In the first pattern, we randomly select one variable from each sample and then randomly delete m% of its observed values. In the second pattern, we randomly select two variables from each sample and then randomly delete m% of the observed values of these two variables.

The first pattern is applied to the 144 combinations mentioned previously, while the second pattern is applied to the combinations where number of variables is 3, 4 or 5. For each pattern, m% of the sample size are randomly deleted where m= 5, 10, 20 and 30.

Therefore, in the case of incomplete data set, 576 combinations of the levels of the factors Δ^2, p, g, n, m and the covariance matrices (equal or not) are considered in the first pattern, while 432 combinations of these factors' levels are considered in the second pattern. This means that we have 1008 combinations in the case of incomplete data set.

For each combination, the methods of handling missing values is applied and the probability of correct classification is estimated for each method using the three rules of classification. To guarantee randomization, each combination in the two patterns is proceeded five times. For each time, we randomly delete m% of the observed values and then the methods of handling missing values are applied and the probability of correct classification is estimated using the three rules of classification. The average of the estimated probabilities of correct classification of the five times is obtained.

6. Results of the Simulation.

In this Section, the results of both the complete data set and the incomplete data set are presented. The results obtained in the complete data set give an evidence of the accuracy of this simulation while the results obtained in the incomplete data set give us a view about the performance of the methods of handling missing values. Also, the results clarify the impact of each factor considered in this study on the methods of handling missing values.

6.1 Results of the Complete Data Set.

As we mentioned in Section (5), 144 combinations are considered in this case. For each combination the estimated probability of correct classification is obtained using the three rules of classification. The results of all combinations considered are displayed in Table (1).

To study the effect of each factor individually, the average of the estimated probabilities of correct classification according to the levels of each factor are obtained. Table (2) presents the averages of the probabilities of correct classification according to the value of Δ^2. It is clear that the probabilities of correct classification increase significantly as Δ^2 increases. This result was expected. In fact, it may be regarded as a check on the accuracy of the simulation. The reason for this is that a large value of Δ^2 indicates a sufficient separation between populations; and this in turn should increase the ability of a classification rule to correctly classify subjects.

The averages of the probabilities of correct classification according to the other factors are also obtained and the following results are deduced :

As g increases from 2 to 3, the probabilities of correct classification decrease rapidly, especially for small values of Δ^2. This provided another check on the accuracy of the simulation as it is expected that the probability of correct classification decreases with the number of groups.

The probabilities of correct classification tend to increase as n increases, especially for small values of Δ^2. This is a logical result as it is known that more information are obtained with the increasing of the sample size and this leads to an increasing ability in correctly classifying subjects.

The probabilities of correct classification increase significantly as p increases when $\Delta^2 = 1$. These probabilities slightly increase as p increases if $\Delta^2 = 4$, while no impact is shown for the increasing of p when $\Delta^2 = 16$. The reason of this result is that a small value of Δ^2 indicates a large overlapping between populations and hence more information are needed to get a higher probability of correct classification. These information may be obtained as the number of variables increases. On the other hand, a large value of Δ^2 indicates, as we mentioned previously, a large separation between populations and consequently a large ability to correctly classify subjects.

The LDF rule is better than the other rules when the covariance matrices are equal while the QDF rule is the best when these matrices are not equal. Also, the probabilities of correct classification in case of equal covariance matrices are smaller than the probabilities in the unequal covariance case, especially for small values of Δ^2 . The last result can be explained by the fact that the overlap between populations in the first case (equal covariance matrices) is more than the overlap between populations in the second one.

The probabilities of correct classification for the given classification rules are slightly different for large values of Δ^2 while they are significantly different for small values of Δ^2.

The QDF rule is always better than the GDR rule in case of unequal covariance matrices while they are approximately identical in case of equal covariance matrices. This result is regarded as another check of the accuracy of our study as it is known from Section(3) that the generalized distance takes the form:

$$(D_k)^2 = (\underline{X} - \mu_k)' \Sigma_k^{-1} (\underline{X} - \mu_k), \qquad k=1,2,.....,g, \qquad (14).$$

while the quadratic discriminant function takes the form:

$$(D'_k)^2 = (D_k)^2 + \ln |\Sigma_k| , \qquad k=1,2,....,g . \qquad (15).$$

Comparing the above equations, one observe that the QDF rule modifies the GDR rule by taking into account the extent to which the variables in each group are dispersed. Therefore, it is expected that the GDR rule and the QDF rule will be approximately identical if the covariance matrices are equal. If these matrices are not equal, the QDF rule will be better than the GDR rule.

For a small value of Δ^2, the probabilities of correct classification are very sensitive to the increasing of the sample size and the number of variables. They are also sensitive to the decreasing of the number of groups. This is because a small value of Δ^2 means a large overlap between populations. Hence, to increase the ability of a classification rule to correctly classify subjects, more information are needed. These information can be obtained with the increasing of the number of variables or the sample size. On the other hand, the decreasing of the number of groups reduces the overlap between these groups.

6.2 Results of the Incomplete Data Set.

In this case, 1008 combinations are considered as was mentioned in Section (5). Each combination is proceeded five times. At each time, the probability of correct classification is estimated using a rule of classification after applying a method of handling missing values. The average of the probabilities obtained from the five times is calculated. This is done for each method of handling missing values and each rule of classification. Therefore, 75 runs, using the Minitab macro were made for each combination; this is because we proceeded each combination five times using the three rules of classification and the five methods of handling missing values.

Based on the results obtained for all combinations considered in this study, the ranks (from worst to best with respect to the estimated probabilities of correct classification) of the different methods of handling missing values were obtained. The medians of the ranks for the different methods of handling missing values according to each rule of classification are given in Table (3). It is clear that the median of the EM method is always greater than (or equal to) the medians of the other methods while the median of the MS method is always less than the medians of the other methods.

Also, from Table (3) one can observe that the lowest medians of the MS and PC methods are found when the QDF rule is used. The greatest medians of the EM and CC methods are found when this rule is used while the greatest median of the RG method is found when the LDF rule is used.

To study the impact of each factor considered on the methods of handling missing values, the medians of the ranks according to each factor considered are obtained. As an example, Table (4) presents the medians according to the number of groups . According to the medians of the ranks obtained, one can deduced the following results:

The EM and RG methods often tend to be better than the other methods. The results of these tables emphasize that the performance of the CC method gets better when the number of groups increase or the number of variables increase or the sample size increase.

The most influential factors in the case of incomplete data set are the sample size, the percentage of missing values and the pattern of missing values. Based on the results obtained for these factors, we conclude that the CC method is the most efficient method when the sample size is large especially if pattern of bivariate missing values is presented, otherwise the EM algorithm is better than the CC method.

In small and moderate sample size, methods of estimation are often better than the CC method. However, in this case, the EM and RG methods are always better than the PC and MS methods. The PC method is always better than the MS method but is always worse than the EM and RG methods.

Another kind of results are presented in Tables (5-7). In these tables the averages of the probabilities of correct classification according to some factors are obtained , based on the results obtained for all combinations considered in this study. The averages of the probabilities of correct classification according to the remaining factors are also obtained.

Looking at the figures in Tables (5-7), we observe that the difference between the averages of the probabilities according to the different methods of handling missing values is often very small; whereas this difference is significant for the large values of Δ^2, especially when the proportion of missing values is large.

Another important feature of these tables is the increasing of the efficiency of the CC procedure with the value of Δ^2. This is a logical result; it can be explained by the fact that when the distance between populations is large then a small number of units are sufficient to build a good classification rule. Consequently the CC procedure is recommended for use in this case.

Inspection of Tables (6-7) shows a decrease of the averages of the probabilities of correct classification as the percentage of missing values increases. This is another logical and expected result. Also, These tables ensure that the RG method and the EM algorithm are often the best methods of estimation. Also, one can observe that In small and moderate sample size, methods of estimation are often better than the CC method while the last method is the most efficient method if the sample size is large.

Table(1):The Estimated Probabilities of Correct Classification for All CombinationsConsidered in the Complete Data Set.

The Case Considered	Sample Size	Rule of Classificat.	$\Delta^2=1$				$\Delta^2=4$				$\Delta^2=16$			
			p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Two Groups and Equal Covariance Matrices.	n=20	LDF	0.663	0.657	0.653	0.685	0.842	0.778	0.805	0.78	0.98	0.988	0.95	0.972
		GDR	0.622	0.643	0.603	0.607	0.84	0.753	0.797	0.713	0.977	0.985	0.917	0.96
		QDF	0.628	0.647	0.59	0.613	0.843	0.745	0.802	0.72	0.977	0.98	0.92	0.965
	n=40	LDF	0.642	0.662	0.672	0.647	0.842	0.812	0.832	0.807	0.983	0.985	0.975	0.972
		GDR	0.643	0.617	0.6	0.63	0.835	0.805	0.8	0.81	0.982	0.98	0.97	0.973
		QDF	0.643	0.635	0.655	0.63	0.838	0.805	0.795	0.8	0.982	0.985	0.97	0.973
	n=100	LDF	0.66	0.655	0.675	0.685	0.833	0.805	0.833	0.802	0.983	0.985	0.975	0.973
		GDR	0.663	0.642	0.665	0.668	0.833	0.805	0.818	0.805	0.983	0.987	0.972	0.97
		QDF	0.648	0.657	0.67	0.652	0.853	0.805	0.825	0.8	0.983	0.987	0.97	0.973
Two Groups and Unequal Covariance Matrices.	n=20	LDF	0.707	0.708	0.678	0.658	0.837	0.828	0.822	0.759	0.979	0.963	0.965	0.973
		GDR	0.671	0.72	0.834	0.84	0.841	0.861	0.895	0.856	0.967	0.98	0.986	0.988
		QDF	0.667	0.792	0.86	0.861	0.834	0.893	0.896	0.898	0.965	0.981	0.988	0.984
	n=40	LDF	0.732	0.76	0.742	0.666	0.85	0.838	0.838	0.817	0.979	0.972	0.972	0.978
		GDR	0.722	0.809	0.845	0.862	0.843	0.885	0.904	0.914	0.973	0.978	0.987	0.989
		QDF	0.719	0.826	0.865	0.874	0.842	0.906	0.908	0.922	0.974	0.98	0.985	0.99
	n=100	LDF	0.716	0.772	0.729	0.667	0.855	0.831	0.852	0.815	0.98	0.967	0.975	0.981
		GDR	0.717	0.799	0.871	0.857	0.858	0.899	0.916	0.926	0.981	0.986	0.986	0.989
		QDF	0.715	0.844	0.882	0.899	0.852	0.906	0.916	0.935	0.982	0.985	0.987	0.991
Three Groups and Equal Covariance Matrices.	n=20	LDF	0.569	0.497	0.485	0.542	0.748	0.709	0.734	0.73	0.967	0.967	0.875	0.969
		GDR	0.528	0.435	0.445	0.453	0.745	0.718	0.679	0.638	0.963	0.957	0.872	0.949
		QDF	0.538	0.46	0.453	0.468	0.739	0.714	0.677	0.69	0.963	0.961	0.877	0.955
	n=40	LDF	0.563	0.543	0.503	0.573	0.741	0.738	0.734	0.756	0.963	0.969	0.893	0.967
		GDR	0.554	0.531	0.501	0.549	0.743	0.737	0.75	0.736	0.959	0.967	0.882	0.954
		QDF	0.569	0.526	0.519	0.545	0.734	0.735	0.735	0.732	0.959	0.967	0.894	0.957
	n=100	LDF	0.564	0.545	0.528	0.577	0.751	0.744	0.753	0.77	0.963	0.969	0.902	0.969
		GDR	0.533	0.535	0.521	0.549	0.745	0.743	0.747	0.745	0.967	0.968	0.903	0.965
		QDF	0.536	0.541	0.518	0.55	0.738	0.743	0.76	0.765	[illegible]	0.968	0.897	0.964
Three Groups and Unequal Covariance Matrices.	n=20	LDF	0.524	0.505	0.534	0.543	0.775	0.599	0.77	0.771	0.833	[illegible]	0.925	0.927
		GDR	0.521	0.676	0.637	0.759	0.781	0.796	0.797	0.862	0.813	[illegible]	0.959	0.937
		QDF	0.528	0.667	0.763	0.773	0.791	0.796	0.845	0.871	[illegible]	0.977	0.963	0.941
	n=40	LDF	0.515	0.479	0.551	0.579	0.749	0.625	0.766	0.786	0.847	0.917	0.947	0.937
		GDR	0.518	0.71	[illegible]	0.77	0.736	0.81	0.807	[illegible]	0.832	0.976	0.961	0.976
		QDF	0.548	0.709	[illegible]	0.797	0.767	0.809	0.868	0.882	0.845	0.977	0.973	0.979
	n=100	LDF	0.539	0.5	0.561	0.57	0.789	0.65	0.796	0.774	0.848	0.907	0.947	0.947
		GDR	0.543	0.709	0.691	0.798	0.793	0.818	0.847	0.895	0.841	0.979	0.972	0.981
		QDF	0.562	0.72	0.81	0.827	[illegible]	0.821	0.878	0.901	0.847	0.981	0.978	0.986

Table(2):The Averages of the Estimated Probabilities of Correct Classification of theThree Rules for VariousValues of Δ^2.

Rule of Classification	$\Delta^2=1$	$\Delta^2=4$	$\Delta^2=16$
LDF	0.612	0.781	0.951
GDR	0.653	0.807	0.957
QDF	0.671	0.815	0.960

Table (3): Medians of the Ranks for the Different Methods of Handling Missing Values.

Rule of Classification.	MS	PC	RG	EM	CC
LDF	2	3	3.5	3.5	3
GDR	2	3	3	3.5	3
QDF	1.5	2	3	4	3.5

Table (4) . Medians of the Ranks for the Different Methods of Handling Missing Values and the Number of Groups.

No. of Groups	Rule of Classification	MS	PC	RG	EM	CC
g=2	LDF	2	3	3.5	3.5	3
	GDR	1.5	2.5	3.5	3.5	3
	QDF	1.5	2	3.5	4	3.5
g=3	LDF	2	3	3	3.5	3
	GDR	2	3	3	3	3
	QDF	1.5	2	3	3.5	4

Table (5) . The Averages of the Estimated Probabilities of Correct Classification of the Three Rules for Various Values of Δ^2 According to Each Method of Handling Missing Values.

Rule	$\Delta^2=1$					$\Delta^2=4$					$\Delta^2=16$				
	MS	PC	RG	EM	CC	MS	PC	RG	EM	CC	MS	PC	RG	EM	CC
LDF	.606	.607	.608	.609	.607	.769	.771	.771	.771	.770	.941	.943	.949	.951	.951
GDR	.653	.655	.656	.655	.655	.794	.796	.798	.797	.796	.945	.948	.955	.958	.958
QDF	.667	.670	.674	.674	.674	.801	.804	.807	.807	.806	.948	.951	.958	.960	.960

Table (6): TheAverages of the Estimated Probabilities of Correct Classification of the Three Rules for Various Values of Δ^2 , the SampleSize and the Proportion of Missing According to Each Method of Handling Missing Values in the Univariate Pattern.

Sample Size	m%	Rule	Δ^2 =1					Δ^2 =4					Δ^2 =16				
			MS	PC	RG	EM	CC	MS	PC	RG	EM	CC	MS	PC	RG	EM	CC
n=20	5%	LDF	.598	.600	.600	.600	.598	.768	.768	.767	.767	.767	.942	.943	.944	.944	.944
		GDR	.624	.625	.622	.621	.620	.781	.783	.784	.783	.783	.941	.944	.948	.948	.948
		QDF	.641	.642	.642	.642	.641	.794	.795	.796	.795	.794	.946	.948	.951	.951	.950
	10%	LDF	.598	.599	.599	.598	.595	.769	.770	.769	.768	.769	.939	.941	.944	.944	.943
		GDR	.623	.622	.621	.618	.619	.782	.783	.783	.783	.779	.938	.941	.946	.947	.945
		QDF	.642	.640	.641	.640	.638	.792	.794	.796	.795	.787	.943	.945	.950	.950	.948
	20%	LDF	.594	.596	.597	.598	.589	.759	.766	.767	.765	.765	.931	.933	.940	.943	.943
		GDR	.618	.616	.618	.616	.612	.769	.773	.777	.776	.771	.926	.931	.942	.946	.944
		QDF	.636	.634	.638	.638	.633	.779	.783	.787	.788	.784	.932	.935	.945	.948	.946
	30%	LDF	.585	.589	.591	.597	.589	.762	.765	.766	.764	.759	.926	.929	.935	.940	.940
		GDR	.606	.608	.610	.611	.606	.764	.764	.772	.771	.760	.919	.924	.929	.938	.939
		QDF	.620	.622	.626	.628	.623	.776	.777	.784	.783	.776	.926	.929	.935	.941	.941
n=40	5%	LDF	.613	613	.615	.614	.613	.765	.767	.767	.768	.767	.952	.953	.954	.954	.954
		GDR	.660	.661	.661	.661	.661	.795	.794	.795	.795	.793	.958	.959	.959	.959	.958
		QDF	.674	.675	.678	.678	.677	.801	.801	.802	.803	.802	.961	.961	.962	.962	.961
	10%	LDF	.612	.613	.614	.613	.613	.765	.766	.767	.767	.766	.947	.949	.953	.953	.952
		GDR	.661	.662	.661	.661	.660	.792	.791	.792	.792	.792	.953	.954	.958	.959	.959
		QDF	.674	.674	.677	.676	.675	.798	.799	.800	.800	.801	.956	.957	.961	.961	.961
	20%	LDF	.610	.612	.612	.613	.612	.765	.766	.767	.768	.766	.943	.946	.952	.953	.952
		GDR	.656	.658	.659	.659	.658	.793	.791	.792	.790	.791	.949	.950	.956	.956	.957
		QDF	.667	.669	.673	.674	.673	.796	.797	.800	.799	.800	.952	.953	.959	.959	.960
	30%	LDF	.606	.607	.607	.609	.606	.763	.766	.766	.765	.767	.938	.942	.950	.952	.952
		GDR	.648	.648	.653	.654	.654	.783	.786	.789	.785	.786	.942	.944	.952	.953	.954
		QDF	.656	.659	.666	.669	.670	.789	.793	.798	.796	.799	.944	.946	.955	.956	.958
n=100	5%	LDF	.618	.620	.621	.621	.620	.790	.790	.791	.791	.791	.953	.953	.954	.954	.954
		GDR	.672	.672	.673	.673	.672	.824	.824	.824	.824	.824	.962	.963	.964	.964	.964
		QDF	.686	.688	.690	.690	.689	.829	.830	.831	.830	.830	.964	.964	.965	.965	.965
	10%	LDF	.617	.620	.620	.621	.619	.789	.790	.790	.790	.790	.947	.949	.954	.954	.954
		GDR	.672	.673	.673	.673	.671	.823	.823	.823	.823	.824	.955	.956	.964	.964	.964
		QDF	.683	.685	.688	.689	.689	.827	.827	.829	.829	.829	.958	.959	.965	.965	.965
	20%	LDF	.615	.617	.619	.621	.621	.787	.789	.790	.790	.790	943	.945	.953	.954	.954
		GDR	.667	.669	.670	.672	.671	.819	.822	.823	.822	.821	.951	.954	.963	.963	.963
		QDF	.676	.681	.686	.687	.689	.822	.824	.828	.827	.827	.953	.955	.964	.964	.965
	30%	LDF	.615	.618	619	.622	.619	.784	.789	.789	.789	.789	.942	.944	.952	.953	.954
		GDR	.664	.665	.668	.669	.668	.813	.818	.821	.820	.822	.948	.959	.960	.962	.963
		QDF	.672	.673	.682	.684	.686	.816	.822	.825	.827	.827	.950	.952	.961	.964	.965

Table (7): TheAverages of the Estimated Probabilities of Correct Classification of the Three Rules for Various Values of Δ^2 , the SampleSize and the Proportion of Missing According to Each Method of Handling Missing Values in the Bivariate Pattern.

Sample Size	m%	Rule	$\Delta^2=1$					$\Delta^2=4$					$\Delta^2=16$				
			MS	PC	RG	EM	CC	MS	PC	RG	EM	CC	MS	PC	RG	EM	CC
n=20	5%	LDF	.589	.590	.592	.592	.591	.758	.758	.757	.756	.756	.940	.942	.945	.946	.944
		GDR	.638	.638	.639	.637	.637	.776	.777	.776	.775	.774	.943	.947	.950	.954	.954
		QDF	.661	.663	.665	.662	.662	.791	.792	.792	.792	.792	.948	.951	.954	.956	.956
	10%	LDF	.592	.593	.595	.593	.593	.758	.759	.757	.756	.757	.935	.938	.944	.945	.944
		GDR	.630	.634	.632	.630	.630	.774	.777	.779	.774	.775	.934	.937	.950	.951	.951
		QDF	.652	.656	.658	.656	.657	.788	.791	.794	.791	.792	.942	.946	.953	.954	.955
	20%	LDF	.584	.586	.590	.591	.589	.756	.757	.757	.754	.754	.922	.927	.938	.943	.942
		GDR	.623	.628	.629	.626	.625	.761	.770	.771	.771	.766	.918	.925	.942	.946	.947
		QDF	.644	.647	.651	.648	.651	.774	.783	.785	.785	.783	.925	.931	.948	.948	.950
	30%	LDF	.578	.583	.585	.589	.583	.754	.755	.753	.750	.752	.918	.921	.934	.944	.941
		GDR	.617	.621	.623	.617	.617	.750	.759	.762	.756	.757	.910	.916	.930	.944	.943
		QDF	.631	.638	.641	.638	.640	.765	.772	.778	.772	.778	.915	.920	.937	.947	.947
n=40	5%	LDF	.611	.612	.613	.612	.612	.758	.758	.757	.759	.757	.951	.954	.956	.956	.956
		GDR	.677	.679	.678	.677	.678	.797	.795	.796	.796	.795	.961	.964	.966	.966	.967
		QDF	.691	.693	.695	.696	.696	.803	.804	.804	.805	.804	.964	.966	.968	.968	.969
	10%	LDF	.610	.612	.612	.613	.612	.762	.763	.763	.762	.762	.947	.950	.955	.956	.956
		GDR	.676	.678	.679	.678	.676	.798	.798	.797	.794	.794	.956	.959	.964	.966	.967
		QDF	.687	.692	.696	.696	.695	.804	.806	.806	.806	.807	.958	.961	.966	.968	.969
	20%	LDF	.609	.610	.610	.610	.608	.758	.760	.760	.759	.759	.940	.944	.955	.955	.955
		GDR	.667	.672	.675	.675	.674	.791	.794	.796	.791	.794	.951	.955	.964	.964	.965
		QDF	.677	.683	.690	.692	.693	.795	.799	.805	.804	.807	.953	.956	.966	.967	.967
	30%	LDF	.607	.606	.608	.608	.608	.754	.756	.757	.757	.755	.931	.934	.950	.952	.954
		GDR	.654	.661	.666	.663	.665	.779	.783	.789	.785	.785	.937	.942	.957	.960	.961
		QDF	.663	.671	.683	.682	.685	.784	.791	.799	.797	.800	939	.943	.961	.962	.964
n=100	5%	LDF	.618	.619	.620	.621	.621	.784	.784	.786	.785	.784	.953	.955	.957	.958	.958
		GDR	.692	.693	.693	.693	.692	.830	.830	.830	.830	.830	.965	.967	.970	.972	.971
		QDF	.710	.712	.712	.714	.715	.834	.835	.835	.837	.837	.967	.969	.971	.972	.972
	10%	LDF	.618	.620	.620	.621	.620	.784	.784	.785	.785	.785	.950	.953	.957	.958	.958
		GDR	.692	.691	.692	.691	.691	.827	.829	.829	.829	.829	.963	.966	.970	.972	.971
		QDF	.705	.710	.710	.712	.713	.832	.834	.834	.837	.837	.964	.967	.971	.973	.972
	20%	LDF	.616	.617	.619	.619	.619	.782	.783	.783	.783	.784	.943	.946	.958	.958	.958
		GDR	.686	.688	.689	.689	.688	.821	.824	.827	.827	.827	.951	.956	.970	.970	.971
		QDF	.696	.702	.707	.709	.712	.824	.829	.833	.834	.835	.953	.958	.971	.971	.972
	30%	LDF	.614	.615	.616	.619	.618	.778	.781	.781	.781	.782	.938	.940	.955	.957	.956
		GDR	.677	.681	.685	.685	.688	.812	.820	.824	824	.826	.945	.950	.967	.969	.969
		QDF	.686	.693	.701	.704	.710	.817	.824	.828	.828	.833	.947	.952	.968	.970	.971

7. Summary and Conclusions

The presence of missing values in a data set is a serious problem that may face the investigator when building a classification rule to classify subjects into groups. In this study we used the Monte Carlo Simulation technique to investigate the performance of five methods of handling missing values when applying classification analysis. The probabilities of correct classification are estimated for three rules of classification; the LDF rule, the GDR rule and the QDF rule, before and after the presence of missing values. Two patterns of missing values were chosen; the univariate missing values pattern and the bivariate missing values pattern . The mechanism of missing values was assumed to be MAR. The samples were generated from populations assumed to be multivariate normal with equal and unequal covariance matrices. Cases of two and three populations were considered. Seven factors were taken into consideration. The impact of each factor on both methods of handling missing values and rules of classification was studied.

The results obtained emphasized the accuracy of the present simulation study . These results simplify the following conclusions. The LDF rule is better than the other rules when the populations are homogeneous (with equal covariance matrices). On the other hand, the QDF rule is recommended for use when the populations are non-homogeneous. In the case of small sample size, estimating missing values is better than to eliminate them. However, in this case the RG method and the EM algorithm are superior to the other methods of estimation, especially when the proportion of missing values is high. When this proportion is low, the differences among the methods of estimation are often slight and consequently any method of them will yield a good classification rule.

The CC procedure, which discards the units containing missing values is recommended if the sample size is large. However, if the proportion of missing values is low then the EM algorithm performs as well as the CC procedure in this case. In general, the CC procedure is the most efficient method if the available information are sufficient to build a good classification rule. This sufficiency is achieved in large sample size, large number of variables and large distance between populations (measured by the Mahalanobis distance Δ^2). On the other hand, if the available information are not sufficient, then methods of estimation will be better than the CC procedure.

Finally, the reader should be reminded that in this study the mechanism of missing values is assumed to be MAR. Under other mechanisms of missing values, the results may differ considerably from the results obtained here. It is remarkable that the present study can be extended to more missing values problems. It may be extended to other mechanisms and patterns of missing values. Also, it can be extended to other rules of classification. Extensions to non-normal populations and more than three populations are also possible. In this thesis, we assumed that $\frac{N_i}{N_j} \approx \backslash$ (i,j =1,2,..,g, i≠j). This assumption can be discarded in another extended study.

References

1. Afifi, A. A, and El Ashoff, R. M., 1966 , " Missing Observations Multivariate Statistics I. Review of the Literature," Journal of the American Statistical Association, vol. 61, pp. 595-604 .

2. Buck, S. F., 1960, " A Method of Estimation of Missing Values in Multivariate Data Suitable for Use with an Electronic Computer ," Journal of the Royal Statistical Society , Ser. B , vol.22, pp. 302-307 .

3. Chan, L. S., and Dunn , O. J., 1972, " The Treatment of Missing Values in Discriminant Analysis-1. The Sampling Experiment ," Journal of the American Statistical Association , vol. 67, pp. 473-477.

4. Chan, L. S., and Dunn, O. J., 1974, " A Note on the Asymptotic Aspect of the Treatment of Missing Values in Discriminant Analysis , " Journal of the American Statistical Association , vol.69, pp. 672-673 .

5. Chan, L. S., Gilman, J. A., and Dunn, O. J., 1976, " Alternative Approaches to Missing Values in Discriminant Analysis, " Journal of the American Statistical Association , vol. 71, pp. 842-844 .

6. Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977, " Maximum Likelihood from Incomplete Data via the EM-Algorithm ," Journal of the Royal Statistical Society , Ser . B , vol. 39, pp. 1-38 .

7. Fisher, R. A., 1936, " The Use of Multiple Measurements in Taxonomic Problems ," Annals . of Eugenics , vol. 7, pp. 179 .

8. Jackson, E.C., 1968, " Missing Values in Linear Multiple Discriminant Analysis," Biometrics , vol. 24, pp. 835-844 .

9. Hand, D.J., 1981, Discrimination and Classification, New York : John Wiley.

10.Lachanbruch, P.A., 1975, Discriminant Analysis, New York: Hafner Press.

11.Little, R. J. A., and Mickey , M. R. , 1968, " Estimation of Error Rates in Discriminant Analysis , " Technomitrics , vol. 10, pp. 1-11.

12.Little, R. J. A., 1978, "Consistent Regression Methods for Discriminant Analysis With Incomplete Data," Journal of the American Statistical Association, vol. 73, pp. 319-322 .

13.Little, R. J. A., and Rubin, D. B., 1987, Statistical Analysis with Missing Data, New York : John Wiley.

14.Little, R. J. A., 1992, " Regression With Missing X's: A Review, " Journal of the American Statistical Association, vol.87, pp. 1227-1237 .

15.Richard, A. J., and Wichern, W.D., 1992, Applied Multivariate Statistical Analysis, New Jersey : Prentice-Hall.

16.Seber, G.A.F., 1984, Multivariate Observations, New York : John Wiley.

17.Stevens, J., 1986, Applied Multivariate Statistics for the Social Sciences, New Jersey: Lowrence Erlbaum Associates.

18.Tatsuoka, M. M., Discriminant Analysis, 1970, The Study of Group Differences, Champaign, II Institute for Personality and Ability Testing.

19.Twedt, D. J., and Gill, D.S., 1992, " Comparison of Algorithms for Replacing Missing Data in Discriminant Analysis, " Communication Statistics - Theory and Methods , vol. 21, pp. 1567-1578.

20.Zaher, A.M., 1995, " On Classification Analysis and its Applications," The 7th Annual Conference on Statistical and Computer Modeling in Human and Social Science, Faculty of Economics and Political Science, Cairo University.