MATHEMATICAL PROGRAMMING APPROACHES TO THE THREE-GROUP CLASSIFICATION PROBLEM

DISSERTATION

Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Constantine Loucopoulos, Diploma, M.S.

Denton, Texas

August, 1993

MATHEMATICAL PROGRAMMING APPROACHES TO THE THREE-GROUP CLASSIFICATION PROBLEM

DISSERTATION

Presented to the Graduate Council of the University of North Texas in Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Constantine Loucopoulos, Diploma, M.S.

Denton, Texas

August, 1993

Loucopoulos, Constantine, Mathematical Programming Approaches to the Three-Group Classification Problem. Doctor of Philosophy (Management Science), August, 1993, 178 pp., 37 tables, 30 illustrations, bibliography, 38 titles.

In the last twelve years there has been considerable research interest in mathematical programming approaches to the statistical classification problem, primarily because they are not based on the assumptions of the parametric methods (Fisher's linear discriminant function, Smith's quadratic discriminant function) for optimality. This dissertation focuses on the development of mathematical programming models for the three-group classification problem and examines the computational efficiency and classificatory performance of proposed and existing models. The classificatory performance of these models is compared with that of Fisher's linear discriminant function and Smith's quadratic discriminant function. Additionally, this dissertation investigates theoretical characteristics of mathematical programming models for the classification problem with three or more groups.

A computationally efficient model for the three-group classification problem is developed. This model minimizes directly the number of misclassifications in the training sample. Furthermore, the classificatory performance of the

proposed model is enhanced by the introduction of a two-phase algorithm. The same algorithm can be used to improve the classificatory performance of any interval-based mathematical programming model for the classification problem with three or more groups. A modification to improve the computational efficiency of an existing model is also proposed. In addition, a multiple-group extension of a mathematical programming model for the two-group classification problem is introduced.

A simulation study on classificatory performance reveals that the proposed models yield lower misclassification rates than Fisher's linear discriminant function and Smith's quadratic discriminant function under certain data configurations. Data configurations, where the parametric methods outperform the proposed models, are also identified.

A number of theoretical characteristics of mathematical programming models for the classification problem are identified. These include conditions for the existence of feasible solutions, as well as conditions for the avoidance of degenerate solutions. Additionally, conditions are identified that guarantee the classificatory non-inferiority of one model over another in the training sample.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF DATA SETS x

LIST OF ILLUSTRATIONS xi

Chapter

I. INTRODUCTION 1

Overview of the Classification Problem
An Application Comparing Different Classification Rules
Purpose, Problem and Significance

II. LITERATURE REVIEW 8

Overview
Major Mathematical Programming Models for Two Groups
Mathematical Programming Models for More than Two Groups
Published Simulation Studies
Anomalies of Mathematical Programming Models for Two Groups

III. PROPOSED MODELS 32

The Modified GSFC Model
MIP3G: A Mixed Integer Programming Model for the Three-Group Problem
MSDMG: A Model for the Minimization of the Sum of Deviations in the Multiple-Group Problem
A Two-Goal Extension to the MIP3G Model
The Grid Algorithm

IV. COMPUTATIONAL EFFICIENCY AND CLASSIFICATORY PERFORMANCE 53

Computational Efficiency of the Modified GSFC Model Relative to the GSFC Model
Computational Efficiency of the Modified GSFC Model Relative to the MIP3G Model
Classificatory Performance

V. THEORETICAL CHARACTERISTICS OF MATHEMATICAL PROGRAMMING MODELS 74

Overview
Theorems, Lemmas and Corollaries

VI. CONCLUSIONS 93

Research Questions Addressed
Limitations of this Study
Proposals for Further Study
Contribution to Managerial Decision-Making

Appendix

A. TABLES 103

B. DATA SETS 141

C. ILLUSTRATIONS .144

REFERENCES 175

LIST OF TABLES

Table Page

1. Effect of varying M in the GSFC Model 104

2. Effect of varying M_1 in the modified GSFC model when data are generated from multivariate normal populations with collinear means and equal covariance structures 105

3. Effect of varying M_1 in the modified GSFC model when data are generated from multivariate normal populations with collinear means and unequal covariance structures 106

4. Effect of varying M_1 in the modified GSFC model when data are generated from multivariate normal populations with equidistant means and equal covariance structures 107

5. Effect of varying M_1 in the modified GSFC model when data are generated from multivariate normal populations with equidistant means and unequal covariance structures 108

6. Effect of varying M_2 in the modified GSFC model on the average number of misclassifications . . 109

7. Effect of varying M_2 in the modified GSFC model on the average CPU times 110

8. Effect of varying M_2 in the modified GSFC model on the average number of integer iterations . . 111

9. Effect of varying é in the modified GSFC model on the average number of misclassifications . . 112

10. Effect of varying é in the modified GSFC model on the average CPU times 113

11. Effect of varying é in the modified GSFC model on the average number of integer iterations . . 114

12. Effect of varying e in the modified GSFC model on the average number of misclassifications . . 115

13. Effect of varying e in the modified GSFC model on the average CPU times 116

14. Effect of varying e in the modified GSFC model on the average number of integer iterations . . 117

15. Average number of misclassifications yielded by the modified GSFC and the MIP3G model when data are generated from multivariate normal populations with collinear means and equal covariance structures 118

16. Average CPU times yielded by the modified GSFC and the MIP3G model for data generated from multivariate normal populations with collinear means and equal covariance structures 119

17. Average number of integer iterations yielded by the modified GSFC and the MIP3G model when data are generated from multivariate normal populations with collinear means and equal covariance structures 120

18. Average number of misclassifications yielded by the modified GSFC and the MIP3G model for data generated from multivariate normal populations with collinear means and unequal covariance structures 121

19. Average CPU times yielded by the modified GSFC and the MIP3G model for data generated from multivariate normal populations with collinear means and unequal covariance structures 122

20. Average number of integer iterations yielded by the modified GSFC and the MIP3G model for data generated from multivariate normal populations with collinear means and unequal covariance structures 123

21. Average number of misclassifications yielded by the modified GSFC and the MIP3G model when data are generated from multivariate normal populations with equidistant means and equal covariance structures 124

22. Average CPU times yielded by the modified GSFC and the MIP3G model for data generated from multivariate normal populations with equidistant means and equal covariance structures 125

23. Average number of integer iterations yielded by the modified GSFC and the MIP3G model when data are generated from multivariate normal populations with equidistant means and equal covariance structures 126

24. Average number of misclassifications yielded by the modified GSFC and the MIP3G model for data generated from multivariate normal populations with equidistant means and unequal covariance structures 127

25. Average CPU times yielded by the modified GSFC and the MIP3G model for data generated from multivariate normal populations with equidistant means and unequal covariance structures 128

26. Average number of integer iterations yielded by the modified GSFC and the MIP3G model for data generated from multivariate normal populations with equidistant means and unequal covariance structures 129

27. Percentage misclassification rates: Case 1 130

28. Percentage misclassification rates: Case 2 131

29. Percentage misclassification rates: Case 3 132

30. Percentage misclassification rates: Case 4 133

31. Percentage misclassification rates: Case 5 134

32. Percentage misclassification rates: Case 6 135

33. Percentage misclassification rates: Case 7 136

34. Percentage misclassification rates: Case 8 137

35. Percentage misclassification rates: Case 9 138

36. Percentage misclassification rates: Case 10 139

37. Percentage misclassification rates: Case 11 140

LIST OF DATA SETS

Data Set Page

1. MBA admissions data 142

2. Data set 2 143

LIST OF ILLUSTRATIONS

Illustration — Page

1.1 MBA admissions data 145

1.2 Observations misclassified by FLDF using MBA admissions data 146

1.3 Observations misclassified by SQDF using MBA admissions data 147

1.4 Observations misclassified by GSFC using MBA admissions data 148

2.1 Effect of magnitude of M in the GSFC on the average number of misclassifications 149

2.2 Effect of magnitude of M in the GSFC on the average CPU times 150

2.3 Effect of magnitude of M in the GSFC on the average number of integer iterations 151

3.1 Effect of magnitude of M_1 in the modified GSFC on the average number of misclassifications (normal populations with collinear means and equal variances) 152

3.2 Effect of magnitude of M_1 in the modified GSFC on the average CPU times (normal populations with collinear means and equal variances) . . . 153

3.3 Effect of magnitude of M_1 in the modified GSFC on the average number of integer iterations (normal populations with collinear means and equal variances) 154

4.1 Effect of magnitude of M_1 in the modified GSFC on the average number of misclassifications (normal populations with collinear means and unequal variances) 155

4.2 Effect of magnitude of M_1 in the modified GSFC on the average CPU times (normal populations with collinear means and unequal variances) . . 156

4.3 Effect of magnitude of M_I in the modified GSFC on the average number of integer iterations (normal populations with collinear means and unequal variances) 157

5.1 Effect of magnitude of M_I in the modified GSFC on the average number of misclassifications (normal populations with equidistant means and equal variances) 158

5.2 Effect of magnitude of M_I in the modified GSFC on the average CPU times (normal populations with equidistant means and equal variances) . . 159

5.3 Effect of magnitude of M_I in the modified GSFC on the average number of integer iterations (normal populations with equidistant means and equal variances) 160

6.1 Effect of magnitude of M_I in the modified GSFC on the average number of misclassifications (normal populations with equidistant means and unequal variances) 161

6.2 Effect of magnitude of M_I in the modified GSFC on the average CPU times (normal populations with equidistant means and unequal variances) 162

6.3 Effect of magnitude of M_I in the modified GSFC on the average number of integer iterations (normal populations with equidistant means and unequal variances) 163

7.1 Percentage Misclassification Rates: Case 1 164

7.2 Percentage Misclassification Rates: Case 2 165

7.3 Percentage Misclassification Rates: Case 3 166

7.4 Percentage Misclassification Rates: Case 4 167

7.5 Percentage Misclassification Rates: Case 5 168

7.6 Percentage Misclassification Rates: Case 6 169

7.7 Percentage Misclassification Rates: Case 7 170

7.8 Percentage Misclassification Rates: Case 8 171

7.9 Percentage Misclassification Rates: Case 9 172

7.10 Percentage Misclassification Rates: Case 10 173

7.11 Percentage Misclassification Rates: Case 11 174

CHAPTER I

INTRODUCTION

Overview of the Classification Problem

The problem of classifying an observation into one of two or more mutually exclusive groups is encountered in a variety of business situations. A loan officer may wish to assess if a loan applicant is a "good risk" (creditworthy) or a "bad risk" (not creditworthy) based on certain applicant characteristics. Similarly, a manufacturer of consumer products may wish to classify consumers in terms of the frequency of their use of rebates. Jolson, Wiener and Rosecky (1987) used discriminant analysis to classify consumers into three groups in terms of their rebate proneness. They used variables like effort/value relationship, brand requirements, shopping efficiency and price awareness, together with demographic characteristics to classify consumers as frequent users, light users or non-users of rebate offers.

Discriminant analysis helps the decision maker by classifying a new observation into one of two or more groups. The most commonly used discriminant analysis procedures have been Fisher's (1936) linear discriminant function and Smith's (1947) quadratic discriminant function. Fisher's linear discriminant function (FLDF) performs optimally when the distributions of the groups are multivariate normal and all covariance structures are equal. Smith's quadratic discriminant function (SQDF) assumes multivariate normality with unequal covariance structure for optimality.

In the last twelve years a large number of articles on mathematical programming approaches to the discriminant analysis problem have appeared in the Management Science literature (Joachimsthaler and Stam, 1990; Erenguc and Koehler, 1990). Such models have attracted much attention because they do not require the parametric assumptions for optimality. At first, the proposed mathematical programming models were rather simplistic and were designed to handle only two-group classification problems (Freed and Glover, 1981a). However, later more complex and sophisticated models were proposed for the two-group problem (Glover, Keene and Duea, 1988), as well as models for the classification problem with three or more groups (Gehrlein, 1986).

The classificatory performance of mathematical programming models relative to that of the FLDF method and the SQDF method was investigated in a number of journal articles, and some of these models perform very competitively under certain data configurations. In the last few years, emphasis has been placed on the identification and remedy of anomalies that have plagued some of these mathematical programming models (Koehler, 1989a; Koehler, 1991).

This dissertation focuses on mathematical programming approaches for the three-group classification problem. In the following section, an example of a three-group classification problem is presented, together with the solutions given by the FLDF method, the SQDF method, as well as the general single function classification (GSFC) model. The GSFC model is one of the mathematical programming models that have appeared in the literature for classification problems with more than two groups (Gehrlein, 1986).

An Application Comparing Different Classification Rules

Johnson and Wichern (1992) presented an example on the use of discriminant analysis in order to classify MBA applicants into three categories (admitted, not admitted and borderline). GPA and GMAT scores were the two criteria used to determine the admissibility of each applicant. In the presented example, out of 85 applicants for admission to an MBA program, 31 were admitted, 28 were not admitted and the remaining 26 were considered as borderline cases as shown in Data Set 1 (Appendix B) and exhibited in Figure 1.1 (Appendix C).

If the FLDF method is used on this example, 7 out of the 85 applicants will be classified incorrectly as exhibited in Figure 1.2 (Appendix C). Specifically, four admitted applicants will be classified as borderline, two applicants who were not admitted will be classified as borderline as well, and finally a borderline applicant will be classified as admitted. If the SQDF method is used, then 3 out of the 85 applicants will be misclassified as exhibited in Figure 1.3 (Appendix C). Specifically, one admitted applicant will be classified as borderline and one applicant who was not admitted will be classified as borderline, whereas a borderline applicant will be classified as admitted. If one mathematical programming model, say the GSFC model, is used then only one of the 85 applicants will be misclassified as exhibited in Figure 1.4 (Appendix C). Specifically, one borderline applicant will be classified as admitted. It is interesting to note that the applicant who is misclassified by the GSFC model is also misclassified by both parametric procedures and the two additional applicants who are misclassified by the SQDF method are also misclassified by the FLDF method.

The GSFC model identifies weights to be assigned to the two criteria, say a_1 for the GPA and a_2 for the GMAT score. Thus, for every applicant a composite score $a_0 + a_1(GPA) + a_2(GMAT)$ is calculated where a_0 is a shifting constant. Applicant admissibility is determined by the identification of cutoff values for each of the three groups of applicants. In this example, the GSFC model generated the following values: $a_0 = 39.527052$, $a_1 = 11.835800$ and $a_2 = .024720$. Applicants with composite scores higher than 88.3837 are admitted, those with scores between 83.3633 and 88.3837 are considered borderline, while applicants with composite scores below 83.3633 are not admitted. Now suppose that a new application is received with GPA = 3.1 and GMAT = 470. Should this applicant be admitted? As $39.527052 + (3.1)(11.8358) + (470)(.02472) = 87.836$, the applicant is considered borderline. If his GPA had been 3.2, then his composite score would have been $39.527052 + (3.2)(11.8358) + (470)(.02472) = 89.020$ and he would have been admitted. Thus, this mathematical programming model identifies a rule that can be used in order to determine the admissibility of future applicants. A similar example on the use of mathematical programming models for the determination of admissibility of students in the MBA program at Simon Fraser University was presented in Choo and Wedley (1985).

Purpose, Problem and Significance

This dissertation proposes modifications to existing mathematical programming models for the three-group classification problem, introduces new models and compares the computational efficiency of different models. Furthermore, the classificatory performance of the proposed models is compared with that of existing models as well as that of the parametric methods (Fisher's linear discriminant function, Smith's quadratic discriminant function). Also certain theoretical characteristics and properties of mathematical programming models are identified.

The appeal of the mathematical programming approaches to the statistical classification problem stems from the absence of the assumptions made for parametric procedures. In real life problems, the assumptions of the FLDF and SQDF methods may be violated, and the degree of such violation affects the classificatory performance of these approaches. One drawback of the mathematical programming approaches to the classification problem, especially the mixed integer programming models, is their computational intensity. The goal of this dissertation is to develop computationally efficient models with high classificatory accuracy.

A number of anomalies have plagued several mathematical programming models for the two-group classification problem (Koehler, 1989a; Koehler, 1991). This dissertation investigates theoretical characteristics of existing and proposed mathematical programming models for the three-group classification problem and identifies conditions under which such anomalies cannot occur.

When the assumptions of the parametric methods are violated, the classificatory accuracy of these methods may be low (Johnson and Wichern, 1992). The proposed mathematical programming models aim at providing the decision maker with accurate classification instruments, useful when the assumptions of the parametric methods are violated.

CHAPTER II

LITERATURE REVIEW

Overview

A considerable number of articles have appeared in the Management Science literature over the last twelve years on mathematical programming approaches to the statistical classification problem (Joachimsthaler and Stam, 1990; Erenguc and Koehler, 1990). The majority of these articles have focused on the presentation of new mathematical programming models for solving the statistical classification problem, while other articles in this area have addressed issues of classificatory performance of these proposed models relative to that of existing methods. Several articles have dealt with certain anomalies that have plagued mathematical programming models for the two-group classification problem (Koehler, 1989a; Koehler, 1991). Of the articles proposing new models, most have presented linear programming or mixed-integer programming models with single or multiple objectives, while a small number of articles have focused on heuristics (Banks and Abad, 1991; Abad and Banks, 1992) and non-linear programming approaches (Stam and Joachimsthaler, 1989).

Major Mathematical Programming Models for Two Groups

Major interest in mathematical programming models for the statistical classification problem was triggered by Freed and Glover (1981a, 1981b). One of the first proposed models was the MMD (maximize the minimum deviation) model (Freed and Glover, 1981a). This model assigns a weight a_k to each attribute variable X_k ($k = 1, 2, \ldots, p$) and thus identifies a linear discriminant score of the form $\sum_{k=1}^{p} a_k x_{ik}$, where x_{ik} is the value of variable X_k for each observation i ($i = 1, 2, \ldots, n$ $k = 1, 2, \ldots, p$). The aim of the model is to identify weights a_k ($k = 1, 2, \ldots, p$) that will maximize the deviation between an arbitrary cutoff value c and the discriminant score that is closest to c. The MMD model is presented below. It should be noted that the symbol $\forall$ denotes "for every" in all the formulations in this dissertation.

Notation:

x_{ik} is the value of variable X_k for observation i

a_k is the weight assigned to variable X_k

d is the minimum deviation between a discriminant score and the cutoff value

c is an arbitrarily chosen cutoff value between the two groups

Formulation:

max d

s.t.

$$\sum_{k=1}^{p} a_k x_{ik} + d \le c \qquad \forall i \epsilon G_1$$

$$\sum_{k=1}^{p} a_k x_{ik} - d \ge c \qquad \forall i \epsilon G_2$$

d, a_k sign-unrestricted variables (k = 1, 2, ..., p)

In this model a positive value of d signifies the absence of any misclassified observations whereas a negative value of d implies that at least one observation is misclassified. It is possible to obtain max d=0 when there are no misclassifications and there exist observations $i \epsilon G_1$ and $j \epsilon G_2$ such that $\sum_{k=1}^{p} a_k x_{ik} = c = \sum_{k=1}^{p} a_k x_{jk}$.

In an attempt to improve upon the classificatory performance of the MMD model, Freed and Glover (1981a) also presented a variation of the above model by considering a variable d_i for each observation i which measures the deviation of the discriminant score $\sum_{k=1}^{p} a_k x_{ik}$ of observation i from an arbitrarily chosen cutoff value c. This model is known by the acronym MSD (maximize sum of deviations) and is presented below:

Notation:

x_{ik} is the value of variable X_k for observation i ($i = 1, 2, \ldots, n \quad k = 1, 2, \ldots, p$)

a_k is the weight assigned to variable X_k

d_i is the deviation of the discriminant score for observation i from the cutoff value c

c is an arbitrarily chosen cutoff value between the two groups

Formulation:

$$\max \sum_{i=1}^{n} d_i$$

s.t.

$$\sum_{k=1}^{p} a_k x_{ik} + d_i \le c \qquad \forall i \in G_1$$

$$\sum_{k=1}^{p} a_k x_{ik} - d_i \ge c \qquad \forall i \in G_2$$

a_k, d_i sign-unrestricted variables

($k = 1, 2, \ldots, p \quad i = 1, 2, \ldots, n$)

Several variations of the MMD model have appeared in the Management Science literature. Freed and Glover (1986a) proposed a model that minimizes the maximum deviation and is also referred to as MMD. Freed and Glover (1986b) also proposed a modification of the MMD model that minimizes the maximum deviation. The cutoff value c is treated as an unrestricted variable and the normalization $\sum_{k=1}^{p} a_k + c = N$ is introduced to eliminate the trivial solution $a_1=a_2=...=a_p=0$. It should be noted that N is a non-zero constant.

Bajgier and Hill (1982) introduced a model that minimizes the sum of misclassification deviations and maximizes the sum of the deviations of correctly classified observations from a cutoff value c. Their OSD (optimize the sum of distances) model is presented below:

Notation:

a_k is the weight assigned to variable X_k

x_{ik} is the value of variable k for observation i

$$d_i^+ = \begin{cases} \sum_{k=1}^{p} a_k x_{ik} - c & \text{if observation i is classified correctly} \\ 0 & \text{if observation i is misclassified} \end{cases}$$

$$d_i^- = \begin{cases} 0 & \text{if observation i is classified correctly} \\ \sum_{k=1}^{p} a_k x_{ik} - c & \text{if observation i is misclassified} \end{cases}$$

c is an arbitrary cutoff value(constant)

n is the number of observations

p is the number of variables

P_1 is the weight assigned to the goal minimizing the sum of misclassification deviations

P_2 is the weight assigned to the goal of maximizing the sum of the deviations of correctly classified observations from the cutoff value c.

Formulation:

$$\min P_1 \sum_{i=1}^{n} d_i^- - P_2 \sum_{i=1}^{n} d_i^+$$

s.t.

$$\sum_{k=1}^{p} a_k x_{ik} + d_i^+ - d_i^- = c \qquad \forall i \epsilon G_1$$

$$\sum_{k=1}^{p} a_k x_{ik} - d_i^+ + d_i^- = c \qquad \forall i \epsilon G_2$$

a_k sign-unrestricted variables $(k = 1, 2, \ldots, p)$

$d_i^+, d_i^- \geq 0 \quad (i = 1, 2, \ldots, n)$

Bajgier and Hill (1982) also proposed a mixed-integer programming model which is an extension to the OSD model. It includes the goal of minimizing the number of misclassifications in addition to the two goals of the OSD model.

Notation:

$$I_i = \begin{cases} 1 & \text{if observation i is misclassified} \\ 0 & \text{if observation i is correctly classified} \end{cases}$$

M is a constant that limits the maximum misclassification deviation

P_1 is the weight assigned to the goal of minimizing the number of misclassifications

P_2 is the weight assigned to the goal of minimizing the sum of misclassification deviations

P_3 is the weight assigned to the goal of maximizing the sum of the deviations of correctly classified observations from the cutoff value c

Formulation:

$$\min P_1\sum_{i=1}^{n} I_i + P_2\sum_{i=1}^{n} d_i^- - P_3\sum_{i=1}^{n} d_i^+$$

s.t.

$$\sum_{k=1}^{p} a_k x_{ik} + d_i^+ - d_i^- = c \qquad \forall i \epsilon G_1$$

$$\sum_{k=1}^{p} a_k x_{ik} - d_i^+ + d_i^- = c \qquad \forall i \epsilon G_2$$

$MI_i \geq d_i^-$ $(i = 1, 2, \ldots, n)$

a_k sign-unrestricted variables $(k = 1, 2, \ldots, p)$

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

Glover (1988) and Glover, Keene and Duea (1988) also proposed a model with several goals in the objective function just as Bajgier and Hill (1982) did. These goals are the minimization of the maximum exterior deviation, the minimization of the weighted sum of exterior deviations, the maximization of the minimum interior deviation and the maximization of the weighted sum of interior deviations. This model is known as the hybrid model.

For the two-group classification problem, several authors have presented mixed-integer programming models with the single objective of minimizing the number of misclassifications in the training sample. These models are modifications of the general single function classification model for three or more groups, presented in Gehrlein (1986). A two-group version of this model is presented below:

Notation:

$$I_i = \begin{cases} 1 & \text{if observation } i \text{ is misclassified} \\ 0 & \text{if observation } i \text{ is correctly classified} \end{cases}$$

M is a constant that limits the maximum misclassification deviation

ε is a small constant denoting half the width of the gap between the two groups

Formulation:

$$\min \sum_{i=1}^{n} I_i$$

s.t.

$$\sum_{k=1}^{p} a_k x_{ik} - MI_i \le c - \varepsilon \qquad \forall i \in G_1$$

$$\sum_{k=1}^{p} a_k x_{ik} + MI_i \ge c + \varepsilon \qquad \forall i \in G_2$$

The above model identifies a small gap of width 2ε in order to generate increased separation between the two groups.

Mathematical Programming Models for More than Two Groups

All the models presented in the above section can only be used for a classification problem with two groups. For the classification problem with more than two groups, Freed and Glover (1981b) proposed the decomposition of the m-group classification problem ($m \geq 3$) into ${}_mC_2$ two-group problems, where ${}_mC_2$ represents the number of combinations of two digits taken from m objects. Then, ${}_mC_2$ pairwise comparisons are performed using any of the above models to classify each observation. Obviously, the number of pairwise comparisons can be quite large for the multiple group case with many groups. Gehrlein (1986) proposed the first mathematical programming model (which was not a pairwise procedure) specifically designed for the classification problem with more than two groups. The general single function classification (GSFC) model projects the data onto a line which is partitioned into intervals, one for each group in such a way that the number of misclassifications is minimized.

Notation:

$$I_i = \begin{cases} 1 & \text{if observation i is misclassified} \\ 0 & \text{if observation i is correctly classified} \end{cases}$$

a_k is the weight assigned to variable X_k ($k = 1, 2, \ldots, p$)

a_0 is a shifting constant

$X_k^{(i)}$ is the value of variable k for observation i (i = 1, 2, ..., n k = 1, 2, ..., p)

U_h is the upper endpoint of the interval assigned to group G_h

L_h is the lower endpoint of the interval assigned to group G_h

e is the minimum gap between intervals

é is the minimum width of an interval assigned to a group

n is the number of observations

m is the number of groups

$$J_{gh} = \begin{cases} 1 & \text{if group } G_g \text{ precedes group } G_h \\ 0 & \text{otherwise} \end{cases}$$

M is a constant limiting the maximum deviation of a misclassified observation from its group, as well as the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval

Formulation:

$$\min \sum_{i=1}^{n} I_i$$

s.t.

$$\left.\begin{array}{l} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M I_i \le U_h \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M I_i \ge L_h \end{array}\right\} \quad \forall i \epsilon G_h \quad (h = 1, 2, \ldots, m)$$

$$U_h - L_h \ge é \qquad (h = 1, 2, \ldots, m)$$

$$\left.\begin{array}{r} L_g - U_h + MJ_{gh} \ge e \\ L_h - U_g + MJ_{hg} \ge e \\ J_{gh} + J_{hg} = 1 \end{array}\right\} \quad \forall G_g, \forall G_h \quad (g,h = 1, 2, \ldots, m \;\; g \ne h)$$

I_i binary variables $(i = 1, 2, \ldots, n)$

J_{gh} binary variables $(g,h = 1, 2, \ldots, m \;\; g \ne h)$

a_0, a_k sign-unrestricted variables $(k = 1, 2, \ldots, p)$

L_h, U_h sign-unrestricted variables $(h = 1, 2, \ldots, m)$

The above model identifies a linear discriminant score $a_0 + \sum_{k=1}^{p} a_k X_k^{(i)}$ for every observation $i \epsilon G_h$ $(h = 1, 2, \ldots, m)$ and then checks if such score falls within the interval $[L_h, U_h]$. This is accomplished by the first two constraints. If the discriminant score $a_0 + \sum_{k=1}^{p} a_k X_k^{(i)}$ for $i \epsilon G_h$ falls outside the interval $[L_h, U_h]$, then observation i is misclassified. The third constraint guarantees that each group will be assigned an interval of width é or more. The last three constraints guarantee that there is no overlap between group intervals.

Gehrlein (1986) also proposed the general multiple function classification (GMFC) model which is presented below.

Notation:

$$I_i = \begin{cases} 1 & \text{if observation } i \text{ is misclassified} \\ 0 & \text{if observation } i \text{ is correctly classified} \end{cases}$$

a_{gk} is the weight assigned to variable X_k in group G_g ($k = 1, 2, \ldots, p \quad g = 1, 2, \ldots, m$)

a_{g0} is a shifting constant for group G_g ($g = 1, 2, \ldots, m$)

n is the number of observations

m is the number of groups

Formulation:

$$\min \sum_{i=1}^{n} I_i$$

s.t.

$$a_{g0} + \sum_{k=1}^{p} a_{gk} X_k^{(i)} - a_{h0} - \sum_{k=1}^{p} a_{hk} X_k^{(i)} + M I_i \geq e \quad \begin{matrix} \forall i \in G_g \\ (g,h=1,2,\ldots,m \quad g \neq h) \end{matrix}$$

In the GMFC model the number of binary variables is n whereas the number of continuous variables is m(p+1). The number of constraints is n(m-1). In the GSFC model the number of binary variables is n+m(m-1) whereas the number of continuous variables is p+1+2m. The number of constraints is $\frac{3}{2}m(m-1) + 2n$. Thus, for a large number of groups and attribute variables, the GSFC model has fewer constraints and continuous variables. This may be desirable from a computational standpoint.

<u>Motivation for Research Question 1</u>

A number of parameters have to be assigned values in the GSFC model. These are the minimum length of each interval (é), the minimum gap size (e) and the maximum deviation (M) of a misclassified observation from the nearest endpoint of the interval assigned to its group. The parameter M also limits the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval to M_2 - e. Gehrlein (1986) used M = 100, é = 5 and e = .01 and pointed out the following:

> Using larger M or smaller e and é values resulted in scaling problems with the program used. Improper scaling refers to the situation where the range of values used for coefficients in a linear programming formulation is too large. The result of this problem is that greater rounding errors result. (Gehrlein 1986, p. 303)

Referring to the issue of selecting appropriate values for the minimum gap size and the maximum misclassification deviation in the two-group mixed-integer programming model, Erenguc and Koehler (1990) state:

> The choice of the values to be assigned to these parameters is both a theoretical and computational issue. (Erenguc and Koehler 1990, p. 221)

Stam and Jones (1990), referring to the maximum misclassification deviation in the two-group mixed-integer programming model, state:

> The parameter M is an arbitrary, sufficiently large real-valued constant. (Stam and Jones 1990, p. 245)

Introducing a two-goal extension to the two-group mixed-integer programming model, Rubin (1990a) used parameters M_1 and M_2 to denote the maximum deviations from their groups of misclassified observations that belong to group G_1 and group G_2 respectively. He noted that,

> Although any adequately large M_1 and M_2 would do, it is desirable to select them as small as it is possible... (Rubin 1990a, p. 257)

Koehler and Erenguc (1990) point out that the gap size (.5ε, according to their notation) has to be small relative to M and note:

> In practice we know of no a priori method for choosing ε and M ... We rely on the standard maxim in mixed integer programming to choose M large enough and ε small enough. (Koehler and Erenguc 1990, p. 71)

The parameter M is often encountered in linear programming problems when one of two mutually exclusive constraints must hold. The standard practice is to introduce a binary variable I and then the term MI (that is, M multiplied by I) is added to the right hand side of the first constraint, while the term M(1-I) is added to the right hand side of the second constraint.

Markland and Sweigart (1987) and Turban and Meredith (1988) recommend assigning an arbitrarily large positive value to M. In their discussion on how large M should be to force artificial variables out of the basis in a linear programming problem, Bazaraa and Jarvis (1977) state:

> Without solving the linear program it is difficult to determine how large M should be to ensure that the artificial variables are driven out of the basis, ... A large value of M will completely dominate other cost coefficients and may result in serious round-off error problems in a computer. (Bazaraa and Jarvis 1977, p. 163)

This uncertainty about the choice of parameter values raises the following research question:

Research Question 1:

How does the choice of parameter values affect the computational efficiency of the GSFC model for the three-group classification problem? What pattern in computational efficiency can be identified for various selections of distinct parameters for the maximum misclassification deviation and the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval?

Motivation for Research Question 2

In the presentation of their three-goal mixed integer programming model for the two-group classification problem, Bajgier and Hill (1982) noted:

> As would be expected for any zero-one formulation of this type, computational time was significant. The MIP formulation with 30 estimation sample cases required

> approximately 5 seconds per problem on a CDC CYBER 173. Computation time grows very rapidly with the number of zero-one variables. (Bajgier and Hill 1982, p. 610)

Koehler (1990) voiced concerns about the computational effort required in mixed-integer approaches where the number of binary variables is at least as large as the number of observations, noting that real-world problems usually have a large number of observations and a small number of attributes.

Since the GSFC model considers six group orderings for the three-group classification problem, it requires six binary variables for the identification of the different orderings. However, three of the possible six group orderings are mirror images of the remaining orderings. For example, the ordering (G_1, G_2, G_3) is the mirror image of the ordering (G_3, G_2, G_1). If three orderings were to be considered, then only two binary variables would be required for identification of the order of the three groups. Mixed-integer programming (MIP) models are unpredictable in terms of CPU times and number integer iterations required. However, reducing the number of binary variables may improve the computational efficiency of the model. Schrage (1986) points out:

> As the number of integer variables is increased the solution time may increase dramatically. (Schrage 1986, p. 186)

A further advantage would be that the optimization package will have to consider only three possible group orderings instead of six. This leads to the following research question:

Research Question 2:

As three of the possible six group orderings in the three-group GSFC model are mirror images of the remaining orderings, can this fact be used in the construction of a computationally more efficient model with fewer binary variables?

Motivation for Research Question 3

Gehrlein (1986) used Fisher's (1936) iris data to compare the classificatory performance of the general single function classifications (GSFC) model and that of the general multiple function classification (GMFC) model. Both models misclassified one out of the 150 observations of the iris data. However, the classificatory performance of these models on a holdout sample was not tested, neither was their classificatory performance compared to that of the parametric methods (FLDF and SQDF). No article to date has performed a Monte Carlo simulation study to compare the classificatory performance of mathematical programming models for three or more groups with that of the FLDF or SQDF methods. This raises the following research question:

Research Question 3:

What is the classificatory performance of the mathematical programming models for the three-group classification problem relative to that of the FLDF or SQDF methods?

Motivation for Research Question 4

Gehrlein (1986) states below that the GMFC model will generally outperform the GSFC model on the training sample:

> While the example problem results in both the single and multiple function schemes producing the same number of improper classifications, the multiple function scheme must generally be expected to produce fewer improper classifications. (Gehrlein 1986, p. 303)

However, Gehrlein (1986) does not provide a proof that the GSFC model cannot outperform the GMFC model in the training sample. This raises the following research question:

Research Question 4:

Can it be proved that the number of misclassifications yielded by the GMFC model in the training sample will not exceed the number of misclassifications yielded by the GSFC model?

Motivation for Research Question 5

The GSFC model identifies three intervals for the three-group classification problem and assigns each of the groups to one of these intervals. If an observation falls outside the interval assigned to its group, then it is misclassified. If an interval-based mathematical programming model is used to solve a three-group classification problem (phase 1), and then the same model is used to resolve the same problem with the additional constraint that the sample covariance of the discriminant scores is zero (phase 2), then the data will be spread out in a different direction from the original one. This approach is similar to that of the second discriminant in parametric discriminant analysis (Johnson and Wichern, 1992). Using the interval cutoff values generated in phases 1 and 2 one can construct a grid consisting of 9 cells. Then each cell can be assigned to the three groups. Such approach will allow greater flexibility in the assignment of groups and may improve classificatory performance. This raises the following research question:

Research Question 5:

> Can a two-phase method be identified that can improve the classificatory performance of the mathematical programming models?

Published Simulation Studies

Freed and Glover (1986a) compared the classificatory performance of the MMD (minimize the maximum deviation), the MSID (maximize the sum of the interior deviations), the MSD (minimize the sum of the deviations) and the FLDF method. All three mathematical programming models included the normalization $c + \sum_{k=1}^{p} a_k = N$. The data were generated from normal populations and various degrees of group separation were considered. The holdout classificatory performance of the MSD model was superior to that of the other three models. The FLDF method outperformed the MMD and the MSID.

Markowski and Markowski (1987) compared the estimation sample classificatory performance of the MMD (maximize the minimum distance) model with that of the FLDF method in the presence of qualitative variables. In this study there were two normally distributed variables, one binary and one tri-valued variable. The FLDF method was found to be superior to the MMD for such data configuration.

The MSD model, the FLDF method, the SQDF method and the logistic discriminant model were considered in a study by Joachimsthaler and Stam (1988) using simulated data from normal and certain non-normal symmetric distributions. The study was limited to evaluating the classificatory performance on the generated training sample only and not on holdout samples. None of the four methods yielded significantly lower misclassifications rates.

Rubin (1990b) carried out a simulation study comparing the holdout classificatory performance of the FLDF method and the SQDF method with that of fourteen mathematical programming models for the two-group classification problem. The MSD, the OSD as well as several variants of the MMD and the hybrid model were included in this study which was restricted to normally distributed data. The SQDF method outperformed the FLDF method and all the mathematical programming models tested.

Stam and Joachimsthaler (1990) compared the classificatory performance of the MSD model, the MIP model presented in Koehler and Erenguc (1990), the FLDF method and the SQDF method. On a set of real data from non-normal populations, the MIP model outperformed the other three methods. A simulation study was carried out using data from normal populations with equal variances, continuous uniform populations with equal variances and discrete uniform populations with equal and unequal variances. The classificatory performance of the MIP model in the estimation training sample was superior to that of the other methods in the first three configurations. When data from discrete uniform populations with unequal variances were used, then the SQDF method outperformed the other methods in the training sample. In terms of holdout classificatory performance the MIP model outperformed the other methods when the data used were from discrete uniform distributions with equal variances. The MSD model outperformed the other methods for data from continuous uniform distributions with equal variances.

The MSD model, the MIP model, the FLDF method and the SQDF method were considered in a simulation study by Stam and Jones (1990). The data were generated from normal, continuous uniform and discrete uniform distributions with equal and unequal variances. The sample sizes ranged from 10 to 50 per group and only equal group sizes were considered. The holdout classificatory performance of the MIP model was superior to that of the other three methods for data from normal populations with unequal variances when $n_i \geq 40$ $(i = 1, 2)$. This was also true for data from discrete uniform populations with equal variances when $n_i \geq 30$ $(i = 1, 2)$. The MSD model outperformed the other methods when the data were generated from normal populations with unequal variances and $n_i \leq 30$ $(i = 1, 2)$, discrete uniform populations with equal variances and $n_i \leq 20$ $(i = 1, 2)$ or continuous uniform populations with equal variances for $n_i = 10$ or $n_i \geq 40$ $(i = 1, 2)$.

Motivation for Research Question 6

The classification performance of the MSD model for the two-group classification model was found to be competitive in all of the above simulation studies except the one by Rubin (1990b). However, no article has considered extending the MSD model so that it can be used in classification problems with more than two groups. This raises the following research question:

Research Question 6:

Can the MSD model be extended for use in classification problems with more than two groups and how will its classificatory performance compare with that of the parametric methods?

Anomalies of Mathematical Programming Models for Two Groups

Various anomalies have plagued several of the mathematical programming approaches for the two-group classification problem. Markowski and Markowski (1985) identified problems that can arise when the MMD (maximize the minimum deviation) or the OSD model are applied to certain types of data structures. They proved that both models can yield unacceptable solutions. An unacceptable solution is one where $a_1=a_2=\ldots=a_p=0$. The presence of observations in the interior of all four quadrants was found to be a sufficient (but not necessary) condition for the existence of an unacceptable solution when p=2. They also showed that if all observations are in the first quadrant, then an unacceptable solution cannot be obtained. Koehler (1989a) and Rubin (1989) identified necessary and sufficient conditions for the existence of unacceptable solutions in the MMD (maximize the minimum deviation) model. Koehler (1989a) also showed that if the MMD (maximize the minimum deviation) model yields an unacceptable solution for $c = c^*$, then the MMD (minimize the maximum deviations) model may yield an unacceptable solution for $c = -c^*$. He also showed that the inclusion of a normalization of the form

$\sum_{k=1}^{p} a_k + c = N$ in the MMD (minimize the maximum deviation) model does not eliminate unacceptable solutions. Furthermore, Koehler (1989a) proved that the MSD and MSID models can yield unacceptable solutions. Koehler (1989b) showed that the hybrid model is not immune to unacceptable solutions, while Koehler (1991) showed that the inclusion of a normalization in a model may lead to the elimination of potentially optimal solutions.

Motivation for Research Question 7

Because of the various anomalies that have plagued several of the mathematical programming models for the two-group classification problem, the following research question is raised:

Research Question 7:

What anomalies, if any, are present in the mathematical programming models for the three-group classification problem?

CHAPTER III

PROPOSED MODELS

The Modified GSFC Model

A modified version of the GSFC model will be presented in this section. First an explanation will be given as to the motivation of using this modified version of Gehrlein's model. This motivation can be understood by examining the role that the M parameter plays. The general single function classification (GSFC) model, presented in Chapter II, uses an arbitrarily large number M which is multiplied by the binary variables I_i ($i = 1, 2, \ldots, n$) and by the binary variables J_{gh} ($g = 1, 2, \ldots, m \quad h = 1, 2, \ldots, m \quad g \neq h$) in the appropriate constraints to allow for misclassified observations and to maintain certain requirements on the structure of the intervals. Gehrlein (1986) selected the value of 100 for M in an application of the GSFC model on a celebrated set of data.

The relationship of the size of M to the CPU times and the number of integer iterations can be illustrated by examining the results of the following simulation. Samples of size 25 were generated from each of three multivariate normal populations with the following specifications:

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \quad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \quad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \quad \Sigma_2 = I \quad \Sigma_3 = I$$

It should be noted that, in the above specifications, I denotes the 2x2 identity matrix. This configuration was selected because the assumptions for optimality for Fisher's linear discriminant function (FLDF) are satisfied and because the group means are collinear.

In this simulation, 25 replications of the data were taken and the parameter M was assigned the values 25, 50, 75, 100, 200 and 500. The package SAS/OR (version 6.07) was used on a Solbourne 5E/902 computer operating under UNIX. The average number of misclassifications generated by the GSFC model over the 25 runs, the average CPU times and the average number of integer iterations are given in Table 1 (Appendix A) and exhibited in Figure 2.1 through Figure 2.3 (Appendix C). The results presented in Table 1 show that there is a strictly monotonic increase in both the average CPU times and the average number of integer iterations as M increases. Clearly CPU times can be prohibitive for large values of M.

A modified version of the GSFC model is considered to make the model less computationally intensive. In the GSFC model, when the parameter M is multiplied by the binary variable I_i, it identifies the maximum deviation of a misclassified observation from the interval assigned to its group. When the parameter M is multiplied by the binary variable J_{gh}, it limits the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval to M - e. The proposed modification to the GSFC model involves the introduction of two parameters M_1 and M_2 to replace M. The parameter M_1 identifies the maximum deviation of a misclassified observation from its group. The parameter M_2 limits the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval to M_2 - e. Such modification allows more flexibility in the specification of parameter values. It will be shown in Chapter IV that this modified version of the GSFC model with M_1 and M_2 can be more computationally efficient than the original version of the GSFC model. The new model, referred to as the modified GSFC, is presented below:

Notation:

$$I_i = \begin{cases} 1 & \text{if observation } i \text{ is misclassified} \\ 0 & \text{if observation } i \text{ is correctly classified} \end{cases}$$

a_k is the weight assigned to variable X_k $(k = 1, 2, \ldots p)$

a_0 is a shifting constant

M_1 is the maximum deviation of a misclassified observation from its group

$X_k^{(i)}$ is the value of variable k for observation i ($i = 1, 2, \ldots, n \quad k = 1, 2, \ldots, p$)

U_h is the upper endpoint of the interval assigned to group G_h ($h = 1, 2, \ldots, m$)

L_h is the lower endpoint of the interval assigned to group G_h ($h = 1, 2, \ldots, m$)

e is the minimum gap between intervals

é is the minimum width of an interval assigned to a group

$$J_{gh} = \begin{cases} 1 & \text{if the interval associated with group } G_g \\ & \text{precedes the interval associated with group } G_h \\ 0 & \text{otherwise} \end{cases}$$

M_2 is a constant limiting the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval to $M_2 - e$

Formulation:

$$\min \sum_{i=1}^{n} I_i$$

s.t.

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i &\le U_h \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i &\ge L_h \end{aligned}\right\} \quad \begin{aligned} &\forall i \in G_h \\ &(h = 1, 2, \ldots, m) \end{aligned}$$

$$U_h - L_h \geq \acute{e} \qquad (h = 1, 2, \ldots, m)$$

$$\left.\begin{array}{r} L_g - U_h + M_2 J_{gh} \geq e \\ L_h - U_g + M_2 J_{hg} \geq e \\ J_{gh} + J_{hg} = 1 \end{array}\right\} \quad \begin{array}{c} \forall G_g,\ \forall G_h \\ (g,h = 1, 2, \ldots, m \quad g \neq h) \end{array}$$

I_i binary variables ($i = 1, 2, \ldots, n$)

J_{gh} binary variables ($g,h = 1, 2, \ldots, m$ $g \neq h$)

a_0, a_k sign-unrestricted variables ($k = 1, 2, \ldots, p$)

L_h, U_h sign-unrestricted variables ($h = 1, 2, \ldots, m$)

MIP3G: A Mixed Integer Programming Model for the Three-Group Problem

A new mixed integer programming formulation is presented in this section for the three-group discriminant problem and is motivated by the need to further reduce the computational effort used by the modified GSFC model. For the three-group discriminant problem, the GSFC model has to consider six different orderings of the groups assigned to those intervals. The model proposed in this section eliminates group orderings which are mirror images of other group orderings and can be generated by the multiplication of certain constraints by -1. For example, if the ordering (G_2, G_1, G_3) is considered, then the ordering (G_3, G_1, G_2) is eliminated from further consideration. This proposed model can be used only for three groups and is presented below. As in the case of the GSFC model, the objective of the MIP3G model is to minimize the number of misclassifications in the training sample.

Notation:

a_k is the weight assigned to variable X_k

$X_k^{(i)}$ is the value of variable k for observation i

a_0 is a shifting constant

é is the width of the middle interval

e is the width of the gap between adjacent intervals

M_1 is a constant that limits the maximum possible distance of a misclassified observation from its group

M_2 is a constant that limits the maximum possible distance of a correctly classified observation that belongs to either the leftmost or the rightmost group from -e or e + é respectively

$$I_i = \begin{cases} 1 & \text{if observation i is misclassified} \\ 0 & \text{if observation i is correctly classified} \end{cases}$$

$$K_2 = \begin{cases} 1 & \text{if group } G_2 \text{ is assigned to the interval on the left} \\ 0 & \text{if group } G_2 \text{ is assigned to the interval in the middle} \end{cases}$$

$$K_3 = \begin{cases} 1 & \text{if group } G_3 \text{ is assigned to the interval on the right} \\ 0 & \text{if group } G_3 \text{ is assigned to the interval in the middle} \end{cases}$$

Formulation:

$$\min \sum_{i=1}^{n} I_i$$

s.t.

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i - (e+é)K_2 + (M_2+e)K_3 \le M_2 \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i - (M_2+e)K_2 + (e+é)K_3 \ge é - M_2 \end{aligned}\right\} \quad \forall i \in G_1$$

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i + (e+é)K_2 \le é \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i + (M_2+e)K_2 \ge 0 \end{aligned}\right\} \quad \forall i \in G_2$$

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i - (M_2+e)K_3 \le é \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i - (e+é)K_3 \ge 0 \end{aligned}\right\} \quad \forall i \in G_3$$

$K_2 + K_3 \ge 1$

$I_1, I_2, \ldots, I_n, K_2, K_3$ binary variables

$a_0, a_1, \ldots, a_p$ sign-unrestricted variables

The above formulation identifies three intervals $[-M_2-e, -e]$, $[0, é]$ and $[e+é, e+é+M_2]$ and assigns one of the following group orderings to these three intervals: (G_1, G_2, G_3), (G_2, G_1, G_3) and (G_2, G_3, G_1). The remaining three orderings are just mirror images of these three orderings, as illustrated in the following example. Suppose that observation i is correctly assigned to the leftmost interval, observation j is correctly assigned to the middle interval and observation r is correctly assigned to the rightmost interval. The six constraints of this model are then equivalent to the following set of equations:

$$-M_2 - e \leq a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} \leq -e$$

$$0 \leq a_0 + \sum_{k=1}^{p} a_k X_k^{(j)} \leq é$$

$$\text{and} \quad e + é \leq a_0 + \sum_{k=1}^{p} a_k X_k^{(r)} \leq e+é+M_2$$

Multiplying each constraint by -1, we obtain:

$$e \leq -a_0 - \sum_{k=1}^{p} a_k X_k^{(i)} \leq M_2 + e$$

$$-é \leq -a_0 - \sum_{k=1}^{p} a_k X_k^{(j)} \leq 0$$

$$\text{and} \quad -e - é - M_2 \leq -a_0 - \sum_{k=1}^{p} a_k X_k^{(r)} \leq -e - é$$

Now if we add the term é (width of the middle interval) to every side of each inequality, we obtain:

$$e + é \leq é - a_0 - \sum_{k=1}^{p} a_k X_k^{(i)} \leq M_2 + e + é$$

$$0 \leq é - a_0 - \sum_{k=1}^{p} a_k X_k^{(j)} \leq é$$

$$\text{and} \quad -e - M_2 \leq é - a_0 - \sum_{k=1}^{p} a_k X_k^{(r)} \leq -e$$

Thus, observation i has moved from the leftmost interval to the rightmost interval, whereas observation r has moved from the rightmost interval to the leftmost interval. Observation j has remained in the middle interval. In the above formulation, G_1 has been chosen as the group that can be assigned to any of the three intervals. However, such choice was arbitrary and does not affect the optimal solution in any way.

The MIP3G model has four fewer binary variables than the GSFC or the modified GSFC models as it uses only two binary variables to maintain the requirements of the line partition. Furthermore, the MIP3G model has six fewer continuous variables and eleven fewer constraints than the other two models.

MSDMG: A Model for the Minimization of the Sum of Deviations in the Multiple-Group Problem

The GSFC model has as an objective the minimization of the number of misclassifications. If the objective is changed to the minimization of the sum of misclassification deviations, then each binary variable I_i ($i = 1, 2, \ldots, n$) will be replaced by a pair of continuous non-negative variables d_{iu} and $d_{i\ell}$. Thus the computational intensity of the model will be reduced substantially. This model is presented below and will be referred to as the MSDMG.

Notation:

a_k is the weight assigned to variable X_k

a_0 is a shifting constant

$X_k^{(i)}$ is the value of variable k for observation i

($i = 1, 2, \ldots, n \quad k = 1, 2, \ldots, p$)

L_h is the lower endpoint of the interval assigned to group G_h ($h = 1, 2, \ldots, m$)

U_h is the upper endpoint of the interval assigned to group G_h ($h = 1, 2, \ldots, m$)

e is the minimum gap between intervals

é is the minimum width of an interval assigned to a group

n is the number of observations

$$J_{gh} = \begin{cases} 1 & \text{if group } G_g \text{ precedes group } G_h \\ 0 & \text{otherwise} \end{cases}$$

d_{iu} is the deviation from U_h of an observation $i \epsilon G_h$ that is misclassified to the right of U_h ($h = 1, 2, \ldots, m$)

$d_{i\ell}$ is the deviation from L_h of an observation $i \epsilon G_h$ that is misclassified to the left of L_h ($h = 1, 2, \ldots, m$)

Formulation:

$$\min \sum_{i=1}^{n} (d_{iu} + d_{i\ell})$$

s.t.

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - d_{iu} \le U_h \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + d_{i\ell} \ge L_h \end{aligned}\right\} \quad \begin{aligned} &\forall i \epsilon G_h \\ &(h = 1, 2, \ldots, m) \end{aligned}$$

$$U_h - L_h \ge e \qquad h = 1, 2, \ldots, m$$

$$\left.\begin{aligned} L_h - U_g + MJ_{hg} \ge \acute{e} \\ L_g - U_h + MJ_{gh} \ge \acute{e} \\ J_{gh} + J_{hg} = 1 \end{aligned}\right\} \quad \begin{aligned} &\forall G_g, \forall G_h \\ &(g,h = 1, 2, \ldots, m \quad g \ne h) \end{aligned}$$

J_{hg} binary variables $(g,h = 1, 2, \ldots, m \quad g \ne h)$

a_k, L_h, U_h sign-unrestricted variables

$(k = 1, 2, \ldots, p \quad h = 1, 2, \ldots, m)$

$d_{iu}, d_{i\ell} \ge 0 \quad (i = 1, 2, \ldots, n)$

A Two-Goal Extension to the MIP3G Model

Both the modified GSFC and the MIP3G models directly minimize the number of misclassifications in the training sample. One characteristic of these models is that they generate alternate optimal solutions and each of these alternate optimal solutions may give a different holdout misclassification rate. According to Rubin (1990a),

> Having a secondary criterion is important, since there may be many solutions with the same (minimal) misclassification rate on the training data, not all of which will be equally efficacious on future observations. (Rubin 1990a, p. 257)

One of the aims of some classification models is to "spread out" the data as much as possible (Johnson and Wichern, 1992). This can be accomplished by the inclusion of a secondary goal maximizing the deviation between the projected mean of the group assigned to the leftmost interval and the projected mean of the group assigned to the rightmost interval. The MIP3G model is extended to include this secondary goal in the following formulation:

Notation:

a_k is the weight assigned to variable X_k

$X_k^{(i)}$ is the value of variable k for observation i

a_0 is a shifting constant

é is the width of the middle interval

e is the width of the gap between adjacent intervals

M_1 is a constant that limits the maximum possible deviation of a misclassified observation from the nearest endpoint of the interval assigned to its group

M_2 is a constant that limits the maximum possible deviation of a correctly classified observation that belongs to either the leftmost or the rightmost group from -e or e + é respectively

$$I_i = \begin{cases} 1 & \text{if observation } i \text{ is misclassified} \\ 0 & \text{if observation } i \text{ is correctly classified} \end{cases}$$

$$K_2 = \begin{cases} 1 & \text{if group } G_2 \text{ is assigned to the interval on the left} \\ 0 & \text{if group } G_2 \text{ is assigned to the interval in the middle} \end{cases}$$

$$K_3 = \begin{cases} 1 & \text{if group } G_3 \text{ is assigned to the interval on the right} \\ 0 & \text{if group } G_3 \text{ is assigned to the interval in the middle} \end{cases}$$

$\overline{X}_h$ is the projected mean of the group G_h $(h = 1, 2, 3)$

δ is the deviation between the projected mean of the group assigned to the leftmost interval and the projected mean of the group assigned to the rightmost interval

P_1 is the weight assigned to the goal of minimizing the number of misclassifications

P_2 is the weight assigned to the goal of maximizing the deviation between the projected mean of the group assigned to the leftmost interval and the projected mean of the group assigned to the rightmost interval

n_h is the number of observations in group G_h $(h = 1, 2, 3)$

Formulation:

$$\min P_1 \sum_{i=1}^{n} I_i - P_2 \delta$$

s.t.

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i - (e+\acute{e})K_2 + (M_2+e)K_3 &\leq M_2 \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i - (M_2+e)K_2 + (e+\acute{e})K_3 &\geq \acute{e} - M_2 \end{aligned}\right\} \quad \forall i \in G_1$$

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i + (e+\acute{e})K_2 &\leq \acute{e} \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i + (M_2+e)K_2 &\geq 0 \end{aligned}\right\} \quad \forall i \in G_2$$

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i - (M_2+e)K_3 &\leq \acute{e} \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i - (e+\acute{e})K_3 &\geq 0 \end{aligned}\right\} \quad \forall i \in G_3$$

$$\overline{X}_1 + M_2 K_3 - \overline{X}_2 + M_2(1-K_2) \geq \delta$$

$$\overline{X}_3 + M_2(1-K_3) - \overline{X}_1 + M_2 K_2 \geq \delta$$

$$\overline{X}_3 + M_2(1-K_3) - \overline{X}_2 + M_2(1-K_2) \geq \delta$$

δ is a non-negative variable

$$\text{and} \quad \overline{X}_h = a_0 + \frac{\sum_{k\neq 0} \sum_{i\in G_h} a_k X_k^{(i)}}{n_h} \qquad (h = 1, 2, 3)$$

The last three constraints guarantee that the deviation between the projected mean of the group assigned to the leftmost interval and the projected mean of the group assigned to the rightmost interval will be greater than or equal to δ. Suppose that $K_2 = 0$ and $K_3 = 1$, i.e. the ordering of the three groups is (G_1, G_2, G_3). Then:

$$\overline{X}_1 - \overline{X}_2 + M_2 \geq \delta$$

$$\overline{X}_3 - \overline{X}_1 \geq \delta$$

$$\overline{X}_3 - \overline{X}_2 + M_2 \geq \delta$$

and therefore, the deviation between the projected mean of the group assigned to the leftmost interval (G_1) and the projected mean of the group assigned to the rightmost interval (G_3) is at least δ.

To understand what value of P_1 and P_2 will make the first goal preemptive over the second goal, consider the following. The deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval is $(M_2 + e) + (e + \acute{e} + M_2) = 2M_2 + 2e + \acute{e}$. Furthermore, the maximum deviation of a misclassified observation from its group is M_1. Thus, if $P_1 > P_2(2M_1 + 2M_2 + 2e + \acute{e})$, the first goal is preemptive over the second goal. This follows since the first goal consists of binary variables.

The Grid Algorithm

The grid algorithm is a proposed two-phase procedure. In the first phase a three-group problem is handled using either the MIP3G or the modified GSFC model. In the second phase the problem is solved again using the same model but with the additional constraint that the sample covariance of the discriminant scores obtained in the two phases is equal to zero. A pair of scores is thus obtained for each observation and plotted with the discriminant score obtained for each observation being on the vertical axis. Horizontal lines are then drawn at levels corresponding to the midpoints of the gaps obtained in the first phase. Vertical lines are also drawn at levels corresponding to the midpoints of the gaps obtained in the second phase. Thus we get nine cells containing a total of n observations. Then each cell will be assigned to a group. As will be seen in the formal presentation of the algorithm, a cell will be assigned to a specific group, if the number of observations from that group exceeds the number of observations from each of the other two groups. A rule is also proposed to handle cases of ties in the number of observations correctly classified by each group. The steps of the Grid algorithm are given below:

Step 1

Solve the three-group classification problem using either the MIP3G or the modified GSFC. Let a_1^*, a_2^*, ..., a_p^* be the optimal solution values of a_1, a_2, ..., a_p and let

$$\ell_1 = \begin{bmatrix} a_1^* \\ a_2^* \\ \cdot \\ \cdot \\ \cdot \\ a_p^* \end{bmatrix}$$

It should be noted that ℓ_1 consists of the values of the weights of the variables X_k (k = 1, 2, ..., p) and does not include a_0 (shifting constant).

Step 2

Let X_i be the column vector of observation i, that is

$$X_i = \begin{bmatrix} X_1^{(i)} \\ X_2^{(i)} \\ \cdot \\ \cdot \\ \cdot \\ X_p^{(i)} \end{bmatrix}$$

Let $$\overline{X} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \cdot \\ \cdot \\ \cdot \\ \overline{X}_p \end{bmatrix} \qquad \text{where } \overline{X}_k = \frac{\sum_{i \in G} X_k^{(i)}}{n} \qquad \begin{matrix} (k=1, 2, \ldots, p) \\ (G=G_1 \cup G_2 \cup G_3) \\ (n=n_1 + n_2 + n_3) \end{matrix}$$

Set $\ell_1^T(X_i - \overline{X}) = g_i \qquad \forall i \epsilon G$

Step 3

Resolve the above problem with the following additional constraint $\ell_2^T \sum g_i(X_i - \overline{X}) = 0 \quad \forall i \epsilon G$. This constraint means that the sample covariance $cov(\ell_1^T X_i, \ell_2^T X_i) = 0$, where

$$\ell_2 = \begin{bmatrix} \tilde{a}_1^* \\ \tilde{a}_2^* \\ \cdot \\ \cdot \\ \cdot \\ \tilde{a}_p^* \end{bmatrix}$$

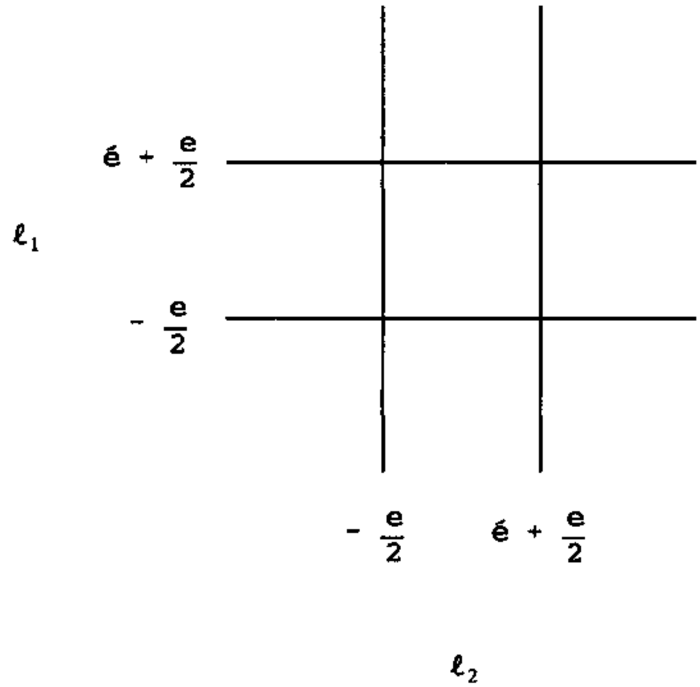
and $\tilde{a}_1^*$, $\tilde{a}_2^*$, ..., $\tilde{a}_p^*$ are the optimal solution values of a_1, a_2, ... , and a_p respectively, when the problem is resolved. Thus the second discriminant will tend to spread the data in a different direction from the direction that the first discriminant spread the data.

Step 4

Calculate $\ell_1^T X_i$ and $\ell_2^T X_i$ and plot $(\ell_1^T X_i, \ell_2^T X_i)$ $\forall i \epsilon G$.

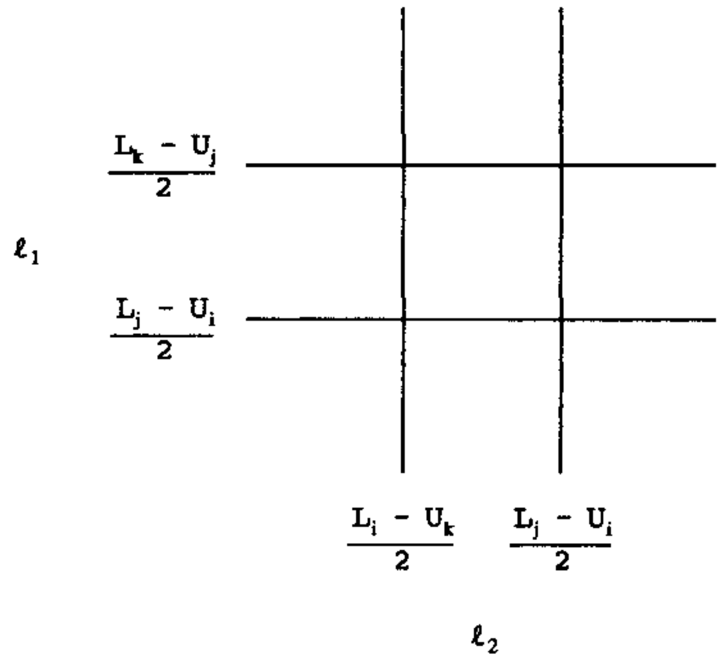
Step 5

Draw horizontal and vertical lines at levels $-\frac{e}{2}$ and $\acute{e} + \frac{e}{2}$ as the below diagram indicates if the MIP3G model is used.



If the modified GSFC model is used instead, and, say for example, the first discriminant group ordering is (G_i, G_j, G_k) for some i, j, k = 1, 2, 3 and $i \neq j$, $j \neq k$, and $i \neq k$, while the second discriminant ordering is (G_k, G_i, G_j) for some i, j, k = 1, 2, 3 and $i \neq j$, $j \neq k$, and $i \neq k$, then horizontal lines are drawn at levels $\frac{L_j - U_i}{2}$ and $\frac{L_k - U_j}{2}$. Similarly, vertical lines are drawn at levels $\frac{L_i - U_k}{2}$ and $\frac{L_j - U_i}{2}$.

Thus, the following grid is generated:



Step 6

Let $C_i^{(j)}$ (i = 1, 2, 3 and j = 1, 2, ..., 9) be the number of observations from G_i that fall into cell j.

Let $C_{max}^{(j)} = \max_{i=1,2,3} [C_i^{(j)}]$ (j = 1, 2, ..., 9).

Let $C^{(j)} = \sum_{i=1}^{3} C_i^{(j)}$ (j = 1, 2, ..., 9).

Then the total number of misclassified observations is

$$\sum_{j=1}^{9} [C^{(j)} - C_{max}^{(j)}].$$

The above procedure assigns the 9 cells to the three groups. Group G_i will be assigned to the cells in which the number of observations from that group exceeds the number of observations from each of the other two groups. So the total number of observations from the other two groups in that cell will give the number of misclassifications for the cell. In case of a tie, that is $C_{max}^{(j)} = C_i^{(j)} = C_k^{(j)}$ for $i,k = 1, 2, 3$, and $i \neq k$, the assignment of a group to a cell by the first discriminant prevails ($j = 1, 2, ..., 9$). Similarly, if a cell does not have any observations, the assignment by the first discriminant will be used for holdout testing.

The classificatory performance of the Grid algorithm will, at worst, be identical to that of the first discriminant in the training sample. The Grid algorithm was not used in the simulation study in Chapter IV for comparison of the classificatory performance of the different models when the data were generated from uniform and normal populations with collinear means. This is because, in several runs, the model failed to yield a solution when the data were generated from normal or continuous uniform populations with collinear means. The same problem was also observed in some preliminary simulation runs using the Grid algorithm, when the number of misclassifications, yielded in Step 1, was zero. Intuitively one can reason that for collinear data from uniform or normal populations the MIP3G method should be optimal and that the Grid algorithm should not be able to provide additional classificatory power with the second discriminant.

CHAPTER IV

COMPUTATIONAL EFFICIENCY AND CLASSIFICATORY PERFORMANCE

Computational Efficiency of the Modified GSFC Model Relative to the GSFC Model

A simulation study was conducted to assess the computational efficiency of the modified GSFC model in comparison to the Gehrlein's GSFC model for the three group classification problem. There are numerous configurations and distributions that could be used in performing this simulation study. However, as multivariate normal populations with equal or unequal covariance structures have been used in many Monte Carlo simulation studies (Rubin, 1990b; Joachimsthaler and Stam, 1988; Freed and Glover, 1986a), data were generated from three multivariate normal populations using the following two configurations:

Configuration 1

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

Configuration 2

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

In both configurations, μ_h denotes the mean of group

G_h (h = 1, 2, 3). It should be noted that in configuration 1 the means of the three groups lie on a straight line, whereas, in configuration 2 the means of the three groups fall on the vertices of an equilateral triangle. The following two covariance structures were considered with I representing the 2x2 identity matrix:

Covariance Structure 1:

$\Sigma_1 = I \quad \Sigma_2 = I \qquad \Sigma_3 = I$

Covariance Structure 2:

$\Sigma_1 = I \quad \Sigma_2 = 1.44I \quad \Sigma_3 = .64I$

Sample sizes were $n_1=n_2=n_3=25$ and the generated data for each of the three groups were replicated 25 times to estimate the computational characteristics of the modified GSFC model under certain parameter settings. The average number of misclassifications, the average CPU times and the average number of integer iterations for configuration 1 with covariance structure 1 are reported in Table 2 (Appendix A) and exhibited in Figures 3.1 through 3.3 (Appendix C). The same computational characteristics for configuration 1 with covariance structure 2, configuration 2 with covariance structure 1 and configuration 2 with covariance structure 2 are reported in Tables 3 through 5 of Appendix A respectively and exhibited in Figures 4.1 through 6.3 (Appendix C). These computational characteristics were generated by setting the values of M_2, é and e equal to 100, 5 and .01 respectively. The parameter M_1 was assigned the values 3, 5, 10, 15, 25, 50, 75, 100 and 1000.

For each of the four combinations of configurations and covariance structures, Tables 2 through 5 (Appendix A) show that the average CPU times and the average number of integer iterations monotonically increase as M_1 increases. Furthermore, the computational efficiency of the modified GSFC model is by far superior to that of Gehrlein's GSFC model. In 25 replications of the data using configuration 1 with covariance structure 1, the average CPU time for the GSFC model with M = 100 (ie. M_1 = 100 and M_2 = 100) is 580.00 seconds. For the modified GSFC model with M_1 = 10 and M_2 = 100 the average CPU time is 115.12 seconds. For the GSFC model with M = 100, the average number of integer iterations is 4980.28 whereas for the modified GSFC model, with M_1 = 10 and M_2 = 100, the average number of integer iterations is 1024.60. It should be noted that the average number of misclassifications generated by the modified GSFC model with M_1 = 10 and M_2 = 100 is identical to that of Gehrlein's GSFC model with M = 100 as seen in Table 2 (Appendix A) and exhibited in Figure 3.1 (Appendix C).

In 25 replications of the data using configuration 1 with covariance structure 2, the average CPU time for the GSFC model with M = 100 (ie. M_1 = 100 and M_2 = 100) is 683.16 seconds and the average number of integer iterations is

6040.40 as seen in Table 3 (Appendix A). For the modified GSFC model with $M_1 = 10$ and $M_2 = 100$ the average CPU time is 142.28 seconds and the average number of integer iterations is 1248.88. Again, for the above parameter values, both models yielded identical numbers of misclassifications.

For configuration 2 with covariance structure 1, the average CPU time for the GSFC model with $M = 100$ (ie. $M_1 = 100$ and $M_2 = 100$) is 393.72 seconds and the average number of integer iterations is 4650.48. For the modified GSFC model with $M_1 = 10$ and $M_2 = 100$ the average CPU time is 195.52 seconds and the average number of integer iterations is 2135.28. The average number of misclassifications is 9.04 (ie. 12.05%) for both models as seen in Table 4 (Appendix A).

For configuration 2 with covariance structure 2 the average CPU time for the GSFC model with $M = 100$ is 346.40 seconds and the average number of integer iterations is 3925.72. These results are presented in Table 5 (Appendix A). The average CPU time for the modified GSFC model with $M_1 = 10$ and $M_2 = 100$ is 162.04 seconds and the average number of integer iterations is 1869.36. Again, for the above parameter values, both models yielded identical numbers of misclassifications.

The computational characteristics presented in Tables 2 through 5 (Appendix A) show that the computational efficiency of the modified GSFC model increases as the value of M_1 is decreased. However, if the value of M_1 is very small, the minimum number of misclassifications in the training sample may not be obtained. Thus, for example, when the data are generated using configuration 1 with covariance structure 1, the average number of misclassifications for $M_1 = 3$ is 13.40. For $M_1 = 5$, the average number of misclassifications reduces to 10.92, whereas for $M_1 \geq 10$ the average number of misclassifications is 10.60.

The effect of varying the value of parameter M_2 on the average number of misclassifications, average CPU times and average number of integer iterations was investigated in a simulation study using samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with configuration 2 and covariance structure 2. The experiment was replicated 25 times and M_2 was assigned the values 20, 40, 75, 100, 250 and 400. As before, the parameter M_1 was assigned the values 3, 5, 10, 15, 25, 50, 75, 100 and 1000. The values of the other parameters were $\acute{e} = 5$ and $e = .01$. The results of this experiment are presented in Tables 6 through 8 (Appendix A). There are several interesting observations that can be made from these tables. For example, the average number of misclassifications was not affected by the magnitude of M_2 for $M_2 \geq 40$. However, the average number of misclassifications for $M_2 = 20$ exceeded the corresponding averages for $M_2 \geq 40$ regardless of the value assigned to the parameter M_1. Furthermore for $M_2 \geq 40$, the average CPU times and the average number of integer iterations were not affected by the magnitude of M_2. For $M_2 = 20$ the average CPU times and average number of integer iterations were higher than those for $M_2 \geq 40$.

Data generated from normal populations with configuration 2 and covariance structure 2 were also used to assess the effect of the magnitude of é, the minimum width of an interval, on the average number of misclassifications, the average CPU times and the average number of integer iterations. The parameter é was assigned the values 5, 2.5, 1 and .5. As before, the parameter M_1 was assigned the values 3, 5, 10, 15, 25, 50, 75 and 100. The values of the other parameters were $M_2 = 100$ and $e = .01$. These results are presented in Tables 9 through 11 (Appendix A).

Several important observations should be noted from the results in Tables 9 through 11. For example, when $M_1 = 3$, the average number of misclassifications for é = 5 and é = 2.5 exceeded the corresponding averages for é = 1 and é = .5. When $M_1 = 5$, the average number of misclassifications for é = 5 exceeded the corresponding averages for 2.5, é = 1 and é = .5. When $M_1 \geq 10$, the magnitude of é did not affect the average number of misclassifications. Another observation to note from Tables 9 through 11 is that the average CPU times were a decreasing function of é for $M_1 \leq 100$. For $M_1 = 1000$ the average CPU times appeared to be unaffected by the magnitude of é. The average number of integer iterations decreased as the value of é was increased. This was true for any value of M_1, however, the decreases in the average number of integer iterations were more striking for small values of M_1.

The same data used for assessing the role of é were also used to assess the effect of the magnitude of e, which is the minimum gap size, on the average number of misclassifications, average CPU times and average number of integer iterations. The parameter e was assigned the values 1, .5, .1, .01, .001 and 0. As before the parameter M_1 was assigned the values 3, 5, 10, 15, 25, 50, 75 and 100. The values of the other parameters were $M_2 = 100$ and é = 5. These results are presented in Tables 12 through 14 (Appendix A). It should be noted that for e = 1 and $M_1 = 3$ there was no feasible solution in 5 out of the 25 runs. For é = 5, the feasibility conditions identified in Theorem 1 (Chapter V) were not satisfied. The minimum number of misclassifications was not attained for e = 1, e = .5 or e = .1 even for very large values of M_1. There were very small differences in the average CPU times for e = .01,

$e = .001$ and $e = 0$. The same is also true for the average number of integer iterations.

Computational Efficiency of the Modified GSFC Model Relative to the MIP3G Model

The same two configurations and covariance structures that were used to assess the computational efficiency of the modified GSFC model for various parameter settings, were also used to assess its computational efficiency relative to that of the MIP3G for various values of M_1. In this study, $\acute{e}$ and e are selected to be equal to 5 and .01, respectively, in both models. The parameter M_2 is assigned the value of 100 in the modified GSFC model and 47.485 in the MIP3G model. The parameter M_2 is assigned this value in the MIP3G because it identifies the width of the leftmost and the rightmost interval. Thus, the deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval is $2M_2 + \acute{e} + 2e$. In the three-group modified GSFC model, this deviation is equal to $M_2 - e$. Therefore, $99.99 = 2M_2 + 5.02$ and thus $M_2 = 47.485$.

In the simulation study to assess the computational efficiency of the MIP3G relative to the modified GSFC, the sizes of the samples were $n_1=n_2=n_3=25$ and the parameter M_1 was assigned the values 3, 5, 10, 15, 25, 50, 75, 100 and 1000. The reported averages were computed from 25 replications of the generated data. The average number of misclassifications yielded by the two models for the different values assigned to the parameter M_1 when the data are generated using configuration 1 with covariance structure 1 are reported in Table 15 (Appendix A). This table shows that the average number of misclassifications is not affected by the magnitude of M_1 for $M_1 \geq 10$ in both models. However, the modified GSFC model yields a slightly lower average number of misclassifications that the MIP3G.

The average CPU times and average number of integer iterations for configuration 1 with covariance structure 1 are reported in Table 16 and Table 17 (Appendix A), respectively. The results in these tables show that the monotonic increase in the average CPU times and the average number of integer iterations, as M_1 increases, is a characteristic of the MIP3G model as well. However, the average CPU times and the average number of integer iterations for the MIP3G model are substantially lower than those of the modified GSFC model. In particular for $M_1 \geq 75$, the average CPU times and the average number of integer iterations for the modified GSFC model were about double those of the MIP3G model.

For configuration 1 with covariance structure 2, the average number of misclassifications, the average CPU times and average number of integer iterations are reported in Table 18, Table 19 and Table 20 (Appendix A), respectively.

The choice of the model used did not affect the average number of misclassifications and both models did not yield the minimum average number of misclassifications for $M_1 \leq 5$. Again, the average CPU times and the average number of integer iterations for the MIP3G model were substantially lower than the same measures for the modified GSFC model for every value of M_1. For $M_1 \geq 50$ the average CPU times and the average number of integer iterations for the modified GSFC model were more than double those of the MIP3G model.

For configuration 2 with covariance structure 1, the average number of misclassifications, the average CPU times and the average number of integer iterations are reported in Table 21, Table 22 and Table 23 (Appendix A), respectively. Again the number of misclassifications was not affected by the choice of the model used. For $M_1 \leq 5$, the two models did not yield the minimum average number of misclassifications. The average CPU times and the average number of integer iterations for the MIP3G model were less than half those for the modified GSFC model regardless of the value assigned to the parameter M_1.

For configuration 2 with covariance structure 2, the average number of misclassifications, the average CPU times and the average number of integer iterations are reported in Table 24, Table 25 and Table 26 (Appendix A), respectively. Both models yielded non-optimal average number of misclassifications for $M_1 \leq 5$. For $M_1 \geq 10$, the modified GSFC model yielded a slightly lower average number of misclassifications than the MIP3G model. Again, the average CPU times and the average number of integer iterations for the MIP3G model were less than half those for the modified GSFC model regardless of the value assigned to the parameter M_1.

As mentioned above, the MIP3G model is computationally more efficient than the modified GSFC model. Theorems 4 and 5 in Chapter V will establish conditions for the parameter settings which characterize the classificatory equivalency of the two models on the training sample. Because of their classificatory equivalence in the training sample and their difference in computational efficiency, the modified GSFC model was not considered in the simulation study for comparison of classificatory performance.

Classificatory Performance

The classificatory performance of six methods in the training sample and the holdout sample was evaluated in this study. The following methods were included:

1. Gehrlein's general multiple function classification (GMFC) model
2. Two-goal MIP3G
3. Grid algorithm
4. MSDMG

5. Fisher's linear discriminant function (FLDF)

6. Smith's quadratic discriminant function (SQDF)

The normal distribution was included in this simulation study because it had been used in a number of published simulation studies on the classificatory performance of mathematical programming models (Rubin, 1990b; Joachimsthaler and Stam, 1988; Freed and Glover, 1986). Four distinct cases were identified: collinear means with equal or unequal covariance structures and equilateral (equidistant) means with equal or unequal covariance structures. As outliers are a common characteristic of data sets and the contamination fraction (p) is usually in the range from .001 to .10 (Hampel, 1974), a case with normal populations and outlier observations was considered. The uniform distribution was also included because it was used in published simulation studies (Stam and Joachimsthaler, 1990; Stam and Jones, 1990). Four distinct cases were identified: collinear means with equal or unequal covariance structures and equilateral (equidistant) means with equal or unequal covariance structures. As two groups may be clearly defined but the third group may have a combinations of shapes, a case was identified with two normal groups and one group consisting of three non-overlapping uniform distributions. The last case considered was motivated by the L-distribution (Bradley, 1982).

Case 1: Multivariate normal populations with collinear means and equal variances

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \quad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \quad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \quad \Sigma_2 = I \quad \Sigma_3 = I$$

Case 2: Multivariate normal populations with collinear means and unequal variances

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = .64I \quad \Sigma_2 = .81I \quad \Sigma_3 = 1.21I$$

Case 3: Multivariate normal populations with equilateral (equidistant) means and equal variances

$$\mu_1 = \begin{bmatrix}0\\173.205\end{bmatrix} \quad \mu_2 = \begin{bmatrix}-100\\0\end{bmatrix} \quad \mu_3 = \begin{bmatrix}100\\0\end{bmatrix}$$

$$\Sigma_1 = 5625I \qquad \Sigma_2 = 5625I \quad \Sigma_3 = 5625I$$

Case 4: Multivariate normal populations with equilateral (equidistant) means and unequal variances

$$\mu_1 = \begin{bmatrix}0\\173.205\end{bmatrix} \quad \mu_2 = \begin{bmatrix}-100\\0\end{bmatrix} \quad \mu_3 = \begin{bmatrix}100\\0\end{bmatrix}$$

$$\Sigma_1 = 6400I \qquad \Sigma_2 = 5625I \quad \Sigma_3 = 4900I$$

Case 5: Multivariate normal with a proportion p=.10 of outlier observations in the leftmost and rightmost population where all five means are collinear

$$\mu_1 = \begin{bmatrix}13\\13\end{bmatrix} \quad \mu_2 = \begin{bmatrix}15\\15\end{bmatrix} \quad \mu_3 = \begin{bmatrix}17\\17\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

$$\mu_1^{(0)} = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_3^{(0)} = \begin{bmatrix}30\\30\end{bmatrix}$$

$$\Sigma_1^{(0)} = 9I \qquad \Sigma_3^{(0)} = 9I$$

where $\mu_i^{(0)}$ and $\Sigma_i^{(0)}$ denote the mean and covariance matrix of the outlier of population i.

Case 6: Continuous uniform populations with collinear means and equal variances

U_1 (x, y; 0, 10, 0, 10)

U_2 (x, y; 6, 16, 0, 10)

U_3 (x, y; 14, 24, 0, 10)

where U_i (x, y; a, b, c, d) denotes a population i with variable X having a uniform distribution over the interval [a,b], variable Y having a uniform distribution over the interval [c,d], and variables X and Y are independently distributed.

Case 7: Continuous uniform populations with collinear means and unequal variances

U_1 (x, y; 0, 10, 0, 10)

U_2 (x, y; 8, 16, -4, 14)

U_3 (x, y; 14, 19, -6, 16)

Case 8: Continuous uniform populations with equilateral (equidistant) means and equal variances

U_1 (x, y; 0, 10, 0, 10)

U_2 (x, y; 8, 18, 0, 10)

U_3 (x, y; 4, 14, 6.928, 16.928)

Case 9: Continuous uniform populations with equilateral (equidistant) means and unequal variances

U_1 (x, y; 0, 10, 0, 10)

U_2 (x, y; 6, 24, 4, 6)

U_3 (x, y; 8, 12, 6.66, 20.66)

Case 10:

Group 1: Multivariate normal population with

$$\mu_1 = \begin{bmatrix} 3.25 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_1 = 1.2\,I$$

Group 2: Multivariate normal population with

$$\mu_2 = \begin{bmatrix} 9.75 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_2 = 1.2\,I$$

Group 3: Consists of three uniform populations

U_1 (x, y; 2.3, 11.5, 3.5, 4)

U_2 (x, y; 2.3, 11.5, -4, -3.5)

U_3 (x, y; -.5, .5, -2.5, 2.5)

where the proportions of observations in U_1, U_2, and U_3 are .275, .275 and .45, respectively.

Case 11: Consists of three rectangles (one for each group) in the shape of an H. The density function for the horizontal rectangle, Group 1, is uniform in the Y direction over the interval [-.5,.5] and U-shaped in the X direction as specified below:

$$f(x) = \begin{cases} \frac{3x^2}{128} & -4 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

The density function for the left vertical rectangle, Group 2, is uniform in the X direction over the interval [-4,-3] and U-shaped in the Y direction as specified below:

$$f(y) = \begin{cases} \frac{3y^2}{128} & -4 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

The density function for the right vertical rectangle, Group 3, is uniform in the X direction over the interval [3,4] and U-shaped in the Y direction as specified below:

$$f(y) = \begin{cases} \frac{3y^2}{128} & -4 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Samples of size $n_1=n_2=n_3=25$ were used to estimate classificatory performance in the training sample. Samples of size $n_1=n_2=n_3=30$ were used to estimate classificatory performance in the holdout sample. There were 25 replications for each configuration. Replications were limited to this number because of the computational intensity of the MIP procedures.

According to the results of the simulation study, the GMFC model has the lowest misclassification rates in the training sample among the methods tested, when the data are generated from normal or uniform populations. This is true regardless of the configurations or the covariance structure used, as it can be seen in Tables 27 through 36 (Appendix A). But, the GMFC model does not always have the lowest misclassification rate in the holdout samples.

When the data are generated from multivariate normal populations with collinear means and equal or unequal variances, then the FLDF method has the lowest misclassification rate in the holdout sample. It should be noted that the difference between the misclassification rate of the FLDF method and that of the SQDF method is very small. It is also noteworthy that the misclassification rates of the GMFC, the MIP3G and the MSDMG in the holdout sample are not much higher than those of the parametric methods, as it can be seen in Table 27 and Table 28 (Appendix A) and exhibited in Figures 7.1 and 7.2 (Appendix C). When the data are generated from multivariate normal populations with equilateral (equidistant) means and equal variances, then the FLDF method and the SQDF method clearly outperform the three proposed models (MIP3G, Grid and MSDMG). The simulation study also shows that the misclassification rate of the GMFC model in the holdout sample is not much higher than those of the parametric methods, as can be seen in Table 29 (Appendix A) and exhibited in Figure 7.3 (Appendix C). It should be noted that the Grid algorithm reduces the misclassification rate of the MIP3G model by more than 34 percent in the holdout sample.

According to the results of the simulation study, when the data are generated from multivariate normal populations with equilateral (equidistant) means and unequal variances, then the differences in misclassification rates between the FLDF method, the SQDF method and the GMFC model are very small. These three methods clearly outperform the MIP3G and the MSDMG models, as well as the Grid algorithm. It is noteworthy that the Grid algorithm reduces the misclassification rate of the MIP3G model by more than 31 percent in the holdout sample. These results are shown in Table 30 (Appendix A) and exhibited in Figure 7.4 of Appendix C.

When the data include outlier observations, then the necessary assumptions for the optimality of the parametric methods are violated. In order to evaluate the classificatory performance of the different methods in the presence of outlier observations, data are generated from multivariate normal populations with a proportion p=.10 of outlier observations in the leftmost and the rightmost populations. The outliers of the leftmost population fall to the left of it and are normally distributed. The outliers of the rightmost population fall to the right of it and are also normally distributed. All five means are collinear and the variances of the three populations are equal. The variances of the outliers are equal to each other and are nine times bigger than the variances of the three populations. The MSDMG model has the lowest holdout misclassification rate in this case. The SQDF method clearly outperforms the FLDF method, but still the misclassifications rate of the SQDF method is more than 50 percent higher than that of the MSDMG model. The GMFC model and the MIP3G model also outperform the parametric methods as shown in Table 31 (Appendix A) and exhibited in Figure 7.5 (Appendix C).

When the data are generated from continuous uniform distributions with collinear means and equal variances, the SQDF method has the lowest misclassification rate in the holdout sample. Among the mathematical programming models, the GMFC has the lowest misclassification rate, as shown in Table 32 (Appendix A) and exhibited in Figure 7.6 (Appendix C).

According to the simulation study, when the data are generated from continuous uniform distributions with collinear means and unequal variances, the SQDF method has the lowest misclassification rate. However, it should be noted that the differences in misclassification rates among the methods tested are very small. These results are presented in Table 33 (Appendix A) and exhibited in Figure 7.7 (Appendix C). When the data are generated from continuous uniform distributions with equilateral (equidistant) means and equal variances, then the SQDF has the lowest misclassification rate in the holdout sample. The misclassification rates of the MIP3G, the Grid and the MSDMG models are substantially higher than those of the parametric methods. It should be noted that the Grid algorithm reduced the misclassification rate of the MIP3G model by more than 33 percent as shown in Table 34 (Appendix A) and exhibited in Figure 7.8 (Appendix C).

According to the simulation study, when the data are generated from continuous uniform distributions with equilateral (equidistant) means and unequal variances, then the SQDF method outperforms all other methods in the holdout sample. The GMFC model has a lower misclassification rate than that of the FLDF method, while the three proposed models (MIP3G, Grid and MSDMG) have substantially higher misclassification rates (Table 35, Appendix A). The same results are also exhibited in Figure 7.9 (Appendix C).

When the data are generated as specified in Case 10, then the Grid algorithm has the lowest misclassification rate (17.69%) in the holdout sample, while the SQDF method has the second lowest misclassification rate (18.80%). The MSDMG model and the FLDF method have the highest misclassification rates, 52.98% and 40.71%, respectively, as shown in Table 36 (Appendix A) and exhibited in Figure 7.10 (Appendix C).

When the data are generated as specified in Case 11, then the Grid algorithm clearly outperforms all the other methods. The SQDF model has the second lowest misclassification rate (16.09%) which is about three times that of the Grid algorithm (5.51%). The MSDMG model and the FLDF method have the highest misclassification rates, 32.67% and 30.71%, respectively, as shown in Table 37 (Appendix A) and exhibited in Figure 7.11 (Appendix C).

CHAPTER V

THEORETICAL CHARACTERISTICS OF MATHEMATICAL PROGRAMMING MODELS

Overview

A number of theoretical results are presented in this chapter. These results identify certain characteristics and properties of the mathematical programming approaches to the multiple group classification problem. Theorem 1 identifies a sufficient condition on the range of values of the parameter M_1, and a necessary and sufficient condition on the range of values of the parameter M_2 so that the modified GSFC model will always yield feasible solutions. Lemma 1 and Lemma 2 identify an upper bound on the number of misclassifications possible that can be yielded by the GSFC and the GMFC models respectively. It is also shown that if this upper bound on the number of misclassifications is reached, then the solution may be unacceptable. Lemma 3 identifies the maximum number of misclassifications that can be yielded by either the GSFC model or the modified GSFC model when the minimum gap size is assigned the value of zero. Lemma 4 identifies the conditions under which the GSFC model will always yield an acceptable solution. Theorem 2 identifies conditions under which, adding a positive constant C, to the values of all the observations, will force the GSFC model to yield an acceptable solution with the number of misclassifications not exceeding a certain limit. Theorem 3 and Corollary 1 identify the maximum value possible of the objective function $\min \sum_{i=1}^{n} I_i$ under certain conditions. Theorem 4 shows that if the MSDMG model yields a degenerate solution in the three-group classification model, then the shifting constant a_0 does not fall strictly within a gap. The same theorem also gives the number of misclassifications in such case and computes an upper bound on the value of the objective function. Theorem 5 identifies the value of the shifting constant a_0 when the three-group MSDMG model yields a degenerate (unacceptable) solution and all group sizes are unequal. It also identifies a range of values for a_0 when not all group sizes are unequal and the MSDMG model yields a degenerate solution. Theorem 6 identifies the conditions under which the modified GSFC model will yield a solution with a minimum number of misclassifications not exceeding the minimum number of misclassifications yielded by the MIP3G model. Theorem 7 identifies the conditions under which the MIP3G model will yield a solution with a minimum number of misclassifications not exceeding the minimum number of misclassifications yielded by the modified GSFC model. Theorem 8 proves that, for appropriately chosen parameter values, the number of misclassifications yielded in the training sample by the GMFC model cannot exceed the number of misclassifications yielded by the GSFC model.

Theorems, Lemmas and Corollaries

Theorem 1:

If $M_1 \geq [(m-2)\acute{e} + (m-1)e]/2$ and $M_2 \geq m(\acute{e}+e)$, then the modified GSFC model will always have a feasible solution.

Proof:

Let the leftmost interval have lower endpoint 0 and upper endpoint $\acute{e}$. Let its adjacent interval have lower endpoint $\acute{e} + e$ and upper endpoint $2\acute{e} + e$. Continuing in the pattern, the rightmost interval will have lower endpoint $(m-1)(\acute{e}+e)$ and upper endpoint $m\acute{e} + (m-1)e$. Let $a_1=a_2=...=a_p=0$. Then each observation will be classified into the group that contains a_0. For $a_0 = [m\acute{e} + (m-1)e]/2$, the maximum misclassification deviation is $a_0 - \acute{e} = [m\acute{e} + (m-1)e]/2 - \acute{e} = [(m-2)\acute{e} + (m-1)e]/2$. Suppose that G_h is assigned the leftmost interval and G_g is assigned the rightmost interval. Thus, $J_{hg} = 1$ and $L_h - U_g + M_2 \geq e$ iff $-m\acute{e} - (m-1)e + M_2 \geq e$ iff $M_2 \geq m(\acute{e}+e)$. This condition on the values of the parameter M_2 is necessary and sufficient. However, the condition on the values of the parameter M_1, i.e., $M_1 \geq [(m-2)\acute{e} + (m-1)e]/2$ is only sufficient. It is not a necessary condition since, for example, a data set with completely separable groups may yield a solution with zero misclassifications when $M_1 = 0$.

Lemma 1:

The maximum number of misclassifications possible in the GSFC model is equal to $n - \max\{n_1, n_2, \ldots, n_m\}$ where $n = \sum_{i=1}^{m} n_i$.

Proof:

Let $a_1=a_2=\ldots=a_p=0$ (degenerate solution). Thus, every composite score $a_0 + \sum_{k=1}^{p} a_k X_k^{(i)}$ becomes equal to a_0 and every observation is classified to the same group. So the number of misclassifications will be minimized when $a_0 \in [L_h, U_h]$ where $n_h = \max\{n_1, n_2, \ldots, n_m\}$ is the size of group G_h. Thus, $\min \sum_{i=1}^{n} I_i = n - \max\{n_1, n_2, \ldots, n_m\}$. In Data Set 2 (Appendix B), $n_1=20$, $n_2=5$ and $n_3=4$. Both the GSFC model and the modified GSFC model yield solutions with $a_1=a_2=0$ and $\min \sum_{i=1}^{29} I_i = 9$ for this set of data. Thus the value $n - \max\{n_1, n_2, \ldots, n_m\}$ is the maximum number of misclassifications, not just an upper bound on the number of misclassifications.

Lemma 2:

The maximum number of misclassifications possible in the GMFC model is equal to $n - \max\{n_1, n_2, \ldots, n_m\}$ where $n = \sum_{i=1}^{m} n_i$.

Proof:

In the GMFC model, let $a_{jk} = 0$ for $j = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, p$ (degenerate solution). Also let $n_h = \max\{n_1, n_2, \ldots, n_m\}$.

Then $\min \sum_{i=1}^{n} I_i = n - \max\ \{n_1, n_2, \ldots, n_m\}$ for $a_{h0} - a_{j0} \geq e$ $(h,j = 1, 2, \ldots, m \quad j \neq h)$. In Data Set 2 (Appendix B), $n_1 = 20$, $n_2 = 5$ and $n_3 = 4$. The GMFC model yields a solution with $\min \sum_{i=1}^{29} I_i = 9$ for this set of data. Thus, $n - \max\{n_1, n_2, \ldots, n_m\}$ is the maximum number of misclassifications, not just an upper bound on the number of misclassifications.

Lemma 3:

If the parameter e(minimum gap size) is assigned the value 0 in the GSFC model or the modified GSFC model, then the maximum number of misclassifications possible is $n - \max\{T \mid T = n_g + n_h,\ \ g,h = 1, 2, \ldots, m \ g \neq h\}$.

Proof:

Let G_g and G_h be the groups with the largest and second largest number of misclassifications. If $a_1=a_2=\ldots=a_p=0$, then G_g and G_h will be assigned to adjacent intervals and a_0 will be equal to the value of the endpoint that the two groups have in common. Thus, all the observations in G_g and G_h will be correctly classified, whereas all the remaining observations will be misclassified. Thus,

$\min \sum_{i=1}^{n} I_i = n - \max\{T | T = n_g + n_h,\ g, h = 1, 2, \ldots, m\ \ g \neq h\}$. In Data Set 2 (Appendix B), $n_1 = 20$, $n_2 = 5$ and $n_3 = 4$. For $e = 0$, the GSFC model yields a solution with $\min \sum_{i=1}^{29} I_i = 4$ for this set of data. Thus, $n - \max\{T | T = n_g + n_h,$ $g, h = 1, 2, \ldots, m\ g \neq h\}$ is the maximum number of misclassifications for $e = 0$, not just an upper bound on the number of misclassifications.

Lemma 4:

The GSFC model will always yield an acceptable solution provided that $a_0 = 0$, $L_h \geq B_L$ for all h ($h = 1, 2, \ldots, m$) where B_L is a positive constant, and the value of the parameter M is chosen to be sufficiently large.

Proof:

Since $L_h \geq B_L$ for h ($h = 1, 2, \ldots, m$) and a_0, if $a_1 = a_2 = \ldots = a_p = 0$, then $\sum_{i=1}^{n} I_i = n$. However, the value of M can always be chosen large enough so that an acceptable solution yields $\sum_{i=1}^{n} I_i \leq n - 1$. Thus, an unacceptable solution will not occur unless $X_k^{(i)} = 0$ for all values of i and k ($i = 1, 2, \ldots, n\ \ k = 1, 2, \ldots, p$).

Theorem 2:

If $a_0 = 0$ and $L_h \geq B_L$ for every h ($h = 1, 2, \ldots, m$) where $B_L > 0$, then there exists a value of the parameter M and a positive constant C such that adding C to each of the observed $X_k^{(i)}$ ($i = 1, 2, \ldots, n$ $k = 1, 2, \ldots, p$) in the GSFC model will result in an acceptable solution with $\min \sum_{i=1}^{n} I_i \le n - \max\{n_1, n_2, \ldots, n_m\}$ where $n = \sum_{i=1}^{m} n_i$.

Proof:

If a_0 and $L_h \ge B_L$ for every h ($h = 1, 2, \ldots, m$) where $B_L > 0$, then the solution generated by the GSFC model is always acceptable according to Lemma 4. If a constant C is chosen so that $\tilde{X}_k^{(i)} = X_k^{(i)} + C > 0$ for all i and k ($i = 1, 2, \ldots, n$ $k = 1, 2, \ldots, p$), then positive coefficients a_k can be found such that all transformed observations are classified into a positive interval provided that the value of the parameter M is chosen to be sufficiently large. Thus, $\min \sum_{i=1}^{n} I_i \le n - \max\{n_1, n_2, \ldots, n_m\}$.

Theorem 3:

Let z be equal to the minimum number of misclassifications resulting from a GSFC model with minimum gap size $e > 0$ and assume that $\sum_{k=1}^{p} a_k \ne 0$. If in the GSFC model the minimum gap size is changed to 0, the constant a_0 is assigned the value of 0, and each L_i is bounded below by a positive constant, then there exists a positive constant C such that adding C to each of the observed $X_j^{(i)}$'s in this model will result in a solution with the minimum number of misclassifications no larger than z.

Proof:

Let $a_1, a_2, \ldots, a_p$ and $L_1, L_2, \ldots, L_m$ and $U_1, U_2, \ldots, U_m$ be values given by the original GSFC model with minimum gap size $e > 0$. If $\sum_{k=1}^{p} a_k < 0$, then there exists and alternative solution to the GSFC model with $\sum_{k=1}^{p} a_k > 0$. Without loss of generality, assume that $\sum_{k=1}^{p} a_k > 0$. Let B_L represent a positive lower bound for $L_1, L_2, \ldots, L_m$. Now, let C be chosen such that $\sum_{k=1}^{p} a_k C \geq s + B_L - (L_g - a_0)$ for each group G_g, where s is the maximum gap length between adjacent intervals in the solution to the GSFC model. Now for each correctly classified observation i, say in group G_g, in the GSFC model, we have $L_g \leq a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} \leq U_g$. this implies that

$$L_g - a_0 + \sum_{k=1}^{p} a_k C \leq \sum_{k=1}^{p} a_k (X_k^{(i)} + C) \leq U_g - a_0 + \sum_{k=1}^{p} a_k C.$$

Let

$$\bar{L}_g = \begin{cases} L_g + \sum_{k=1}^{p} a_k C - a_0 & \text{if group } G_g \text{ is associated with the leftmost interval} \\ \dfrac{L_g + U_h}{2} + \sum_{k=1}^{p} a_k C - a_0 & \text{if there exists a group } G_h \text{ such that } L_g \geq U_h \text{ and the interval associated with } G_h \text{ is adjacent to the interval associated with group } G_g \end{cases}$$

and let

$$\tilde{U}_g = \begin{cases} U_g + \sum_{k=1}^{p} a_k C - a_0 & \text{if group } G_g \text{ is associated with the rightmost interval} \\ \dfrac{L_h + U_g}{2} + \sum_{k=1}^{p} a_k C - a_0 & \text{if there exists a group } G_h \text{ such that } L_h \geq U_g \text{ and the interval associated with } G_h \text{ is adjacent to the interval associated with group } G_g \end{cases}$$

Note that for $\tilde{X}_k^{(i)} = X_k^{(i)} + C$, we have $\tilde{L}_g \leq \sum a_k \tilde{X}_k^{(i)} \leq \tilde{U}_g$ for each observation i that was correctly classified in the original model.

Note that

$$\tilde{L}_g \geq L_g - \frac{s}{2} + \sum_{k=1}^{p} a_k C - a_0$$

$$= \sum_{k=1}^{p} a_k C + L_g - a_0 - \frac{s}{2}$$

$$\geq B_L + s - \frac{s}{2}$$

$$\geq B_L$$

Thus the solution $a_1, a_2, \ldots, a_p, \tilde{L}_1, \tilde{L}_2, \ldots, \tilde{L}_m, \tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_m$, to the GSFC model with a gap size of 0, $a_0 = 0$, and each L_g bounded by a positive constant will yield no more than the number of misclassifications resulting from the original GSFC model with minimum gap size $e > 0$. Since the difference between discriminant scores does not change, there is no need to change the value of the constant M to have a solution with the number of misclassifications no larger than z.

Corollary 1:

Theorem 3 also holds if the assumption $\sum_{k=1}^{p} a_k \neq 0$ is replaced by $\sum_{k=1}^{p} a_k = 0$ and at least one $a_k \neq 0$ for $k = 1, 2, \ldots, p$.

Proof:

In the proof of Theorem 3, replace $\sum_{k=1}^{p} a_k C$ by $a_{k_0} C$ where $a_{k_0} \neq 0$ and also replace $\tilde{X}_k^{(i)}$ by $X_k^{(i)}$ if $k \neq k_0$ and $\tilde{X}_{k_0}^{(i)}$ by $X_{k_0}^{(i)} + C$. Instead of assuming that $\sum_{k=1}^{p} a_k > 0$, assume $a_k > 0$. Thus the result of the corollary follows.

Theorem 4:

If the MSDMG model yields a degenerate solution in the three-group classification problem, then the constant a_0 does not fall strictly within a gap and the number of misclassifications is equal to $n - \max\{n_1, n_2, n_3\}$. In this model, the value of the objective function does not exceed $e[n - \max\{n_1, n_2, n_3\}] + \acute{e} \min\{n_1, n_2, n_3\}$ where $\acute{e}$ is the minimum interval length and e is the minimum gap size.

Proof:

As the objective of the MSDMG model is the minimization of the sum of misclassification deviations, if the model yields a degenerate optimal solution, then the group with the largest number of observations will be assigned to the middle interval and the width of that interval will be $\acute{e}$. Furthermore, each of the gaps separating the intervals will be assigned a width of e. If the constant a_0 were to fall within a gap, it would fall within the gap between the two largest groups. Let d_1 be the distance of a_0 from the nearest endpoint of the middle interval. Then the value of the objective function is:

$$\begin{aligned} Z_1 &= (e-d_1)[n - \max\{n_1, n_2, n_3\} - \min\{n_1, n_2, n_3\}] \\ &\quad + d_1\max\{n_1, n_2, n_3\} + (d_1+\acute{e}+e)\min\{n_1, n_2, n_3\} \\ &= e[n - \max\{n_1, n_2, n_3\}] + \acute{e}\min\{n_1, n_2, n_3\} \\ &\quad + d_1[2\max\{n_1, n_2, n_3\} + 2\min\{n_1, n_2, n_3\} - n] \end{aligned}$$

$$\min_{d_1} Z_1 = e[n - \max\{n_1, n_2, n_3\}] + \acute{e}\min\{n_1, n_2, n_3\}$$

for $d_1 = 0$, i.e. a_0 does not fall strictly within a gap.

Therefore, for $d_1 = 0$ all the observations in the largest group will be correctly classified and the number of misclassifications will be $n - \max\{n_1, n_2, n_3\}$. In such case, the value of the objective function will be

$$\begin{aligned} & e[n - \max\{n_1, n_2, n_3\} - \min\{n_1, n_2, n_3\}] + (e+\acute{e})\min\{n_1, n_2, n_3\} \\ &\quad = e[n - \max\{n_1, n_2, n_3\}] + \acute{e}\min\{n_1, n_2, n_3\}. \end{aligned}$$

Thus $\max[\min\sum_{i=1}^{n} (d_{iu}+d_{il})] = e[n - \max\{n_1, n_2, n_3\}] + \acute{e}\min\{n_1, n_2, n_3\}]$

Theorem 5:

If the MSDMG model yields a degenerate solution in the three-group classification problem and all group sizes are unequal, then the constant a_0 falls on the endpoint of the middle interval that is closest to the interval assigned to the group with the second largest number of observations. If $n - \max\{n_1, n_2, n_3\} = 2\min\{n_1, n_2, n_3\}$ or if $n_1=n_2=n_3$, then $a_0 \in [L_h, U_h]$ where G_h is the group assigned to the middle interval.

Proof:

According to Theorem 4, a_0 will not fall strictly within a gap. Then a_0 will fall into the middle interval because the largest group is assigned to that interval. Let $d_2 + e$ be the distance between a_0 and the nearest endpoint of the interval assigned to the group with the second largest number of observations. Thus $0 \le d_2 \le é$ and the value of the objective function is:

$$Z_2 = (e+d_2)[n - \max\{n_1, n_2, n_3\} - \min\{n_1, n_2, n_3\}] + (é-d_2+e)\min\{n_1, n_2, n_3\}$$

$$= e[n - \max\{n_1, n_2, n_3\}] + é\min\{n_1, n_2, n_3\} + d_2[n - \max\{n_1, n_2, n_3\} - 2\min\{n_1, n_2, n_3\}]$$

$$\min_{d_2} Z_2 = e[n - \max\{n_1, n_2, n_3\}] + é\min\{n_1, n_2, n_3\}$$

for $d_2 = 0$ if

$$\max\{n_1, n_2, n_3\} > n - \max\{n_1, n_2, n_3\} - \min\{n_1, n_2, n_3\} > \min\{n_1, n_2, n_3\}$$

Therefore a_0 falls on the endpoint of the middle interval that is closest to the interval assigned to the group with the second largest number of observations. If $n - \max\{n_1, n_2, n_3\} = 2\min\{n_1, n_2, n_3\}$, or if $n_1=n_2=n_3$, then

$d_2[n - \max\{n_1, n_2, n_3\} - 2\min\{n_1, n_2, n_3\}] = 0$, regardless of the value of d_2 ($d_2 \leq é$). Therefore $a_0 \in [L_h, U_h]$, where G_h is the group assigned to the middle interval.

Theorem 6:

Any optimal solution of the MIP3G model is a feasible solution of the modified GSFC model if the values of the parameters e, é and M_1 are the same for both models and $M_2^a \geq 2M_2^b + 3e + é$, where M_2^a and M_2^b are the values of the parameter M_2 used in the modified GSFC model and the MIP3G model, respectively.

Proof:

Suppose that $K_2 = 1$ and $K_3 = 0$ in the MIP3G model. Then G_1 is the group assigned to the rightmost interval, G_2 is the group assigned to the leftmost interval and G_3 is assigned to the middle interval. Thus,

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i &\leq M_2^b + e + é \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i &\geq é + e \end{aligned}\right\} \quad \forall i \in G_1$$

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i &\leq é \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i &\geq -M_2^b - e \end{aligned}\right\} \quad \forall i \in G_2$$

$$\left.\begin{aligned} a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} - M_1 I_i &\leq é \\ a_0 + \sum_{k=1}^{p} a_k X_k^{(i)} + M_1 I_i &\geq 0 \end{aligned}\right\} \quad \forall i \in G_3$$

In the three-group modified GSFC model, let

$L_1 = e + é$, $U_1 = M_2^b + e + é$, $L_2 = -M_2^b - e$, $U_2 = -e$, $L_3 = 0$ and $U_3 = é$.

Then the first two constraints of the modified GSFC model become identical to the above constraints. Furthermore, the constraints $U_h - L_h \geq é$ are satisfied for the above values of L_h and U_h (h = 1, 2, 3). As $K_2=1$ and $K_3=0$, then $J_{12}=J_{13}=J_{32}=0$ and $J_{21}=J_{31}=J_{23}=1$. Thus,

$L_1 - U_2 \geq e$

$L_1 - U_3 \geq e$

$L_3 - U_2 \geq e$

$L_2 - U_1 + M_2^a \geq e$

$L_2 - U_3 + M_2^a \geq e$

$L_3 - U_1 + M_2^a \geq e$

The first three of the above six constraints are satisfied for $L_1 = e + é$, $U_1 = M_2^b + e + é$, $L_2 = -M_2^b - e$, $U_2 = -e$, $L_3 = 0$ and $U_3 = é$.

$L_2 - U_1 + M_2^a \geq e$ iff

$M_2^a \geq e + M_2^b + e + M_2^b + e + é$ iff

$M_2^a \geq 2M_2^b + 3e + é$

$L_2 - U_3 + M_2^a \geq e$ iff

$M_2^a \geq e + M_2^b + e + é$ iff

$M_2^a \geq M_2^b + 2e + é$

$L_3 - U_1 + M_2^a \geq e$ iff

$M_2^a \geq e + M_2^b + e + é$ iff

$M_2^a \geq M_2^b + 2e + é$

Therefore the last three of the above six constraints are satisfied for $M_2^a \geq 2M_2^b + 3e + é$. In a similar fashion it can be shown that the constraints of the modified GSFC model hold when $K_2 = 1$ and $K_3 = 1$. Thus, in the training sample, the number of misclassifications yielded by the modified GSFC model will never exceed the number of misclassifications yielded by the MIP3G model.

Theorem 7:

If $e = 0$ and the values assigned to parameters M_1 and é are the same in both the modified GSFC model and the MIP3G model, then any optimal solution of the modified GSFC model is a feasible solution of the MIP3G model if $M_2^b \geq M_2^a - 2é$.

Proof:

Without loss of generality, one may consider only three group orderings of the modified GSFC model, say (G_1, G_2, G_3), (G_2, G_1, G_3) and (G_2, G_3, G_1), as the remaining three group orderings are mirror images of these orderings. Consider an optimal solution of the modified GSFC model where the ordering of the three intervals is (G_1, G_2, G_3). Let L_h^*, U_h^* $(h = 1, 2, 3)$, a_0^* and a_k^* $(k = 1, 2, ..., p)$ be the optimal solution values of L_h, U_h, a_0 and a_k respectively in the modified GSFC model. All the constraints of the modified GSFC model will still hold if L_h^*, U_h^* and a_k^* are replaced by $\tilde{L}_h = L_h^* \acute{e} / (U_2^* - L_2^*)$, $\tilde{U}_h = U_h^* \acute{e} / (U_2^* - L_2^*)$, $\tilde{a}_0 = a_0^* \acute{e} / (U_2^* - L_2^*)$ and $\tilde{a}_k = a_k^* \acute{e} / (U_2^* - L_2^*)$ for $(h = 1, 2, 3 \quad k = 1, 2, ..., p)$. Similarly, the constraints of the modified GSFC model will still hold if $\tilde{L}_h$, $\tilde{U}_h$, $\tilde{a}_0$, $\tilde{a}_k$ are replaced by $L_h = \tilde{L}_h - \tilde{L}_2$, $U_h = \tilde{U}_h - \tilde{U}_2$, $a_0 = \tilde{a}_0 - \tilde{L}_2$ and $a_k = \tilde{a}_k$ for $h = 1, 2, 3$ and $k = 1, 2, ..., p$. Now the middle interval $[L_2, U_2]$ in the modified GSFC model has width $\acute{e}$ and $L_2 = 0$. Hence the values $a_0, a_1, ..., a_p$ will provide a feasible solution to the MIP3G model for $e = 0$, $k_2 = 0$ and $k_3 = 1$ provided that M_2^b is sufficiently large. Since the length of the widest possible interval in the modified GSFC model is $M_2^a - 2\acute{e}$, setting $M_2^b \geq M_2^a - 2\acute{e}$, in the MIP3G model guarantees that the leftmost and rightmost intervals of the MIP3G will be long enough for a feasible solution. Therefore, the theorem holds when the ordering of the three intervals is (G_1, G_2, G_3). In a similar fashion, it can be shown that the theorem holds when the ordering of the three intervals is either (G_2, G_1, G_3) or (G_2, G_3, G_1).

Theorem 8:

Values for the constants M and e in the GMFC model can always be found such that the number of misclassifications for the GMFC model will be less than or equal to the number of misclassifications of the GSFC model with nonzero minimum gap size.

Proof:

Assume that we have a solution to the GSFC model. Furthermore, assume that $U_1 < U_2 < \ldots < U_m$. We can always relabel the groups so that this will be true. Also, without loss of generality, assume that gap $L_n - U_{n-1}$ for $n = 2, 3, \ldots, m$ is equal to the minimum possible gap, say e_0. An alternate solution can always be found such that each gap is equal. Now define $f(x) = a_0 + \sum_{i=1}^{p} a_i x_i$ where $x = (x_1, x_2, \ldots, x_p)$ and the a_i's are from the solution of the GSFC model. Let $\bar{e}$ be some small positive value to be used in the GMFC model for e. Now define $h_1(f(x)) = (\bar{e}/e_0)[U_1 + e_0 - f(x)]$. Note that for $f(x) \leq U_1 + e_0$, $h_1(f(x)) \geq 0$ and for $f(x) > U_1 + e_0$, $h_1(f(x)) < 0$. Define $h_2(f(x)) = (\bar{e}/e_0)(f(x) - U_1)$. Note that for $f(x) \leq U_1$, $h_2(f(x)) \leq 0$ and for $f(x) > U_1$, $h_2(f(x)) > 0$. Now define $h_n(f(x)) = b_n\ [f(x) - U_{n-1}]$ for $n = 3, 4, \ldots, m$ where $b_n = (b_{n-1}[U_{n-1} - U_{n-2} + e_0] + \bar{e})/e_0$ with $b_2 = \bar{e}/e_0$. Note that for $n = 2, 3, 4, \ldots, m$, we have for $f(x) \leq U_{n-1}$, $h_n(f(x)) \leq 0$ and for $f(x) > U_{n-1}$, $h_n(f(x)) > 0$. Therefore it is easy to see that if $n_2 > n_1$ and $f(x_0)$ is contained in the interval $[L_{n_1}, U_{n_1}]$ then for $n_1 > 1$, we have

$$h_{n_1}(f(x_0)) - h_{n_2}(f(x_0)) \geq h_{n_1}(f(x_0))$$
$$= b_{n_1}[f(x) - U_{n-1}]$$
$$\geq b_{n_1}[U_{n-1} + e_0 - U_{n-1}]$$
$$\geq \overline{e}$$

If $n_1 = 1$, then we have

$$h_{n_1}(f(x_0)) - h_{n_2}(f(x_0)) \geq h_{n_1}(f(x_0))$$
$$= (\overline{e}/e_0)[U_1 + e_0 - f(x)]$$
$$\geq (\overline{e}/e_0)(e_0)$$
$$= \overline{e}$$

Now we also want to show that the following statement is true: for $n_2 < n_1$ and $f(x_0)$ in the interval $[L_{n_1}, U_{n_1}]$, we have $h_{n_1}(f(x_0)) - h_{n_2}(f(x_0)) \geq \overline{e}$. For $n_1 = 2$, this statement follows because $h_1(f(x_0)) \leq 0$ and $h_2(f(x_0)) \geq \bar{e}$ for $f(x_0) \in [L_2, U_2]$. Next we show that for $f(x_0) \in [L_n, U_n]$ and $n \geq 3$,

$$h_n(f(x_0)) - h_{n-1}(f(x_0))$$
$$= b_n(f(x_0) - U_{n-1}) - b_{n-1}(f(x_0) - U_{n-2})$$
$$= (b_n - b_{n-1})(f(x_0)) - b_nU_{n-1} + b_{n-1}U_{n-2}$$
$$\geq (b_n - b_{n-1})(U_{n-1} + e_0) - b_nU_{n-1} + b_{n-1}U_{n-2}$$
$$= b_{n-1}(U_{n-2} - U_{n-1}) + e_0(b_n - b_{n-1})$$
$$= b_{n-1}(U_{n-2} - U_{n-1}) + b_{n-1}(U_{n-1} - U_{n-2}) + \overline{e}$$
$$= \overline{e}$$

Since $b_n > b_{n-1}$ for $n \geq 3$, it follows that for $n_1 \geq n \geq 3$ with $f(x_0) \in [L_{n_1}, U_{n_1}]$ that $h_n(f(x_0)) - h_{n-1}(f(x_0)) \geq \bar{e}$. Also, we have for $n_1 \geq n \geq 3$ with $f(x_0) \in [L_{n_1}, U_{n_1}]$ that $h_n(f(x_0)) \geq h_{n-1}(f(x_0))$. This implies that for $n_1 > n_2$ with $f(x_0) \in [L_{n_1}, U_{n_1}]$, that $h_{n_1}(f(x_0)) \geq h_{n_2}(f(x_0))$. Hence for $n_1 > n_2$ and $f(x_0) \in [L_{n_1}, U_{n_1}]$, we have the statement $h_{n_1}(f(x_0)) - h_{n_2}(f(x_0)) \geq h_{n_2+1}(f(x_0)) - h_{n_2}(f(x_0)) \geq \bar{e}$. Now the a_i coefficients for the linear discriminant function used for the j^{th} group in the GMFC model can be taken to be the coefficients of the function $h_j(x)$. The $\bar{e}$ value can be used for e and a large M can be taken so as to make the constraints feasible for the misclassified observations. Thus the number of misclassifications for the GMFC model will be less than or equal to the number of misclassifications for the GSFC model for suitably chosen constants.

CHAPTER VI

CONCLUSIONS

Research Questions Addressed

This dissertation addresses a number of research questions regarding the use of mathematical programming models for the three-group classification problem.

Research Question 1:

> How does the choice of parameter values affect the computational efficiency of the GSFC model for the three-group classification problem? What patterns in computational efficiency can be identified for various selections of distinct parameters for the maximum misclassification deviation and the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval?

In a simulation study, it was shown that the parameter M has a major effect on the CPU times and the number of integer iterations in the original three-group GSFC model. Specifically, it was shown that there is a strictly monotonic increase in the CPU times and in the number of integer iterations as M increases on generated normal data. Furthermore, it was shown that the computational efficiency of the GSFC model is improved by modifying this model to include distinct parameters for the maximum misclassification deviation and the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval. The magnitude of the parameter that limits the maximum deviation between the lower endpoint of the leftmost interval and the upper endpoint of the rightmost interval has minor effect on the CPU times and the number of integer iterations. However, the magnitude of the parameter denoting the maximum misclassification deviation has a major effect on the CPU times and the number of integer iterations.

Research Question 2:

> As three of the possible six group orderings in the three-group GSFC model are mirror images of the remaining orderings, can this fact be used in the construction of a computationally more efficient model with fewer binary variables?

A new model, the MIP3G, was proposed which considers only group orderings that are not mirror images of other considered group orderings. As a result, only three out of the six possible orderings are considered by the MIP3G model. Furthermore, the proposed model requires only two binary variables for the identification of group orderings, whereas both the GSFC and the modified GSFC model need six variables for group ordering identification. As shown in Chapter IV, the computational efficiency of the MIP3G is superior to that of the GSFC or the modified GSFC model. Theorem 6, in Chapter V, identifies the conditions under which the MIP3G model cannot yield a lower number of misclassifications than the modified GSFC model in the training sample. Theorem 7, in Chapter V, identifies the conditions under which the modified GSFC model cannot yield a lower number of misclassifications than the MIP3G model in the training sample.

Research Question 3:

> What is the classificatory performance of the mathematical programming models for the three-group classification problem relative to that of Fisher's linear discriminant function or Smith's quadratic discriminant function?

The simulation study presented in Chapter IV shows that proposed models have lower misclassification rates than the parametric methods in the holdout sample for certain distributions and configurations. However, when the population means are equidistant, the FLDF and the SQDF method outperform all the mathematical programming models tested.

Research Question 4:

> Can it be proved that the number of misclassifications yielded by the GMFC model in the training sample will not exceed the number of misclassifications yielded by the GSFC model?

Theorem 8 in Chapter V proves that the number of misclassifications yielded by the GMFC model in the training sample cannot exceed the number of misclassifications yielded by the GSFC model.

Research Question 5:

> Can a Two-phase method be identified that can improve the classificatory performance of the mathematical programming models?

The proposed Grid algorithm is a two-phase method. In the performed simulation, the Grid algorithm was shown to improve the classificatory performance of the MIP3G model in the holdout sample. Furthermore, it was shown that for certain configurations, the algorithm yielded lower misclassification rates than any of the other methods tested. Specifically, when data are generated as described in Case 10 and Case 11, the Grid algorithm outperforms the other methods in terms of classificatory accuracy on the holdout sample.

Research Question 6:

> Can the MSD model be extended for use in classification problems with more than two groups and how will its classificatory performance compare with that of the parametric methods?

The proposed MSDMG model is an extension of the MSD model that can be used for classificatory problems with more than two groups. The classificatory performance of the MSDMG model was found to be better than that of all the other methods tested when there are outlier observations in the leftmost and rightmost groups and all means are collinear. However, when the population means are equidistant or when the data are described in Case 10 and Case 11, the MSDMG model has much higher misclassification rates than the other methods.

Research Question 7:

> What anomalies, if any, are present in the mathematical programming models for the three-group classification problem?

Theorem 1, in Chapter V, identifies the conditions under which the GSFC model will always yield a feasible solution. In Lemma 1 and Lemma 2, in Chapter V, it is shown that a solution yielded by the GSFC model or the GMFC model may be unacceptable. Lemma 4 identifies the conditions under which the GSFC model cannot yield an unacceptable solution.

Limitations of this Study

The simulation results, presented in Chapter IV, show that the proposed mathematical programming models yield lower misclassification rates in the holdout sample for certain distributions and configurations. However, in several other configurations, the parametric procedures still appear to be robust to certain forms of non-normal data and thus the parametric procedures can outperform the mathematical programming models in those situations.

The MIP3G model has limitations on its usage. This model can only be used in a three-group classification problem. It cannot be easily extended for usage in problems with more than three groups like the MSDMG model or the modified GSFC model.

Another limitation of MIP models, in general, is the computational intensive nature of these models. For example, the CPU times for the modified GSFC model, and even the computationally more efficient MIP3G model, can be excessive, especially when the sample sizes are large. The same is true when the value assigned to the parameter limiting the maximum misclassification deviation is large. When the Grid algorithm is used, the discriminant problem is solved twice using either the modified GSFC model or the MIP3G model and thus the Grid may be impractical for some problems because of its computational intensity.

The Grid algorithm may fail to yield a solution when the data are generated from populations with collinear means. Furthermore, if the Grid algorithm is used with data for which the modified GSFC model or the MIP3G model yield zero misclassifications, then this algorithm may not provide any additional information for holdout classification.

The results on classificatory performance, presented in Chapter IV, are valid for the specific distributions and configurations used in the simulations. Results may be different for other distributions. Furthermore, it should be noted that, for a three-group problem, numerous configurations are possible.

Proposals for Further Study

The findings of this dissertation can be expanded by the investigation of a number of related research issues.

1. The simulation study performed in this dissertation was restricted to a limited number of distributions and configurations. The classificatory performance of the proposed mathematical programming models on holdout samples can be evaluated using other non-normal distributions and different orientations of the configurations.
2. This study was restricted to the classification problem with three groups. The classificatory performance of the modified GSFC model, the MSDMG model and the Grid algorithm can be compared to that of the parametric methods for classification problems with more than three groups.
3. In Chapter IV, the effect of the magnitude of parameter values on the computational efficiency of mathematical programming approaches was investigated. A study may be conducted to assess the effect of the choice of parameter values on the classificatory performance of

the proposed models in the holdout sample, particularly for small values of the M parameters.

4. This dissertation identified the secondary goal of maximizing the deviation between the projected means of the groups assigned to the leftmost and rightmost intervals. Other secondary goals may be identified and their effect on the classificatory performance of the modified GSFC and MIP3G models could be assessed.
5. Heuristics have been proposed by Banks and Abad (1991), Abad and Banks (1992), Rubin (1990a) and Koehler and Erenguc (1990) for improvement of the computational efficiency of models for the two-group classification problem. Heuristics may be proposed for the three-group problem and their classificatory performance compared with that of exact models.
6. Freed and Glover (1981b) proposed a pairwise comparisons approach for the classification problem with more than two groups. However, no simulation studies have been published on the classificatory performance of this method. Its classificatory performance should be compared with that of the parametric methods.
7. In the proposed Grid algorithm, the assignment role of the first discriminant prevails in case of a tie. Other methods for resolving ties may be identified that

will improve the classificatory performance of the algorithm.

Contribution to Managerial Decision-Making

Discriminant analysis is used in a variety of business problems like credit granting, assessment of rebate proneness etc. Typically, the decision makers have relied mainly on Fisher's linear discriminant function or Smith's quadratic discriminant function in order to classify observations into groups when faced with a three-group classification problem. The decision makers can easily implement the mathematical programming models in this dissertation by using one of the popular optimization packages such as SAS/OR, LINDO, MPSX, etc. The models and results in this dissertation will allow the decision makers to have greater flexibility in choosing an appropriate discriminant procedure, particularly when violations of the parametric assumptions are readily apparent.

This dissertation shows that when the data have certain shapes the proposed mathematical programming models have higher classificatory accuracy than their parametric counterparts. Furthermore, managers may find the classification process using the proposed mathematical programming models computationally more efficient than previous models. For example, the MIP3G model is much more practical from a computational standpoint than the originally proposed GSFC model. The experimental results in this dissertation can assist the decision maker in assessing the usefulness of the proposed mathematical programming models as alternative classification methods in situations where the parametric models do not perform optimally.

APPENDIX A

TABLES

Table 1

Effect of varying M on the average number of misclassifications, average CPU times(secs) and average number of integer iterations in the GSFC model for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

M	$\min \sum_{i=1}^{n} I_i$	CPU (secs)	Integer Iterations
25	10.72	270.76	2300.48
50	10.64	417.84	3445.80
75	10.60	511.44	4275.44
100	10.60	580.00	4990.28
200	10.60	701.64	6531.76
500	10.60	834.32	7705.44

Table 2

Effect of varying M_i in the modified GSFC model on the average number of misclassifications, average CPU times(secs) and average number of integer iterations for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

M_i	$\min \sum_{i=1}^{n} I_i$	CPU (secs)	Integer Iterations
3	13.40	33.20	383.80
5	10.92	57.40	604.80
10	10.60	115.12	1024.60
15	10.60	167.64	1397.56
25	10.60	265.00	2122.92
50	10.60	417.96	3441.96
75	10.60	523.56	4325.88
100	10.60	580.00	4980.28
1000	10.60	925.12	8575.40

Table 3

Effect of varying M_l in the modified GSFC model on the average number of misclassifications, average CPU times(secs) and average number of integer iterations for sample size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_l	$\min \sum_{i=1}^{n} I_i$	CPU (secs)	Integer Iterations
3	12.96	43.16	510.32
5	11.32	71.20	772.16
10	11.08	142.28	1248.88
15	11.08	201.88	1723.24
25	11.08	307.60	2621.96
50	11.08	485.16	4247.56
75	11.08	596.04	5282.00
100	11.08	683.16	6040.40
1000	11.08	1052.80	10121.52

Table 4

Effect of varying M_1 in the modified GSFC model on the average number of misclassifications, average CPU times(secs) and average number of integer iterations for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \quad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \quad \Sigma_2 = I \quad \Sigma_3 = I$$

M_1	$\min \sum_{i=1}^{n} I_i$	CPU (secs)	Integer Iterations
3	10.36	45.92	565.48
5	9.20	103.48	1184.40
10	9.04	195.52	2227.92
15	9.04	245.24	2805.96
25	9.04	295.64	3457.20
50	9.04	350.60	4131.40
75	9.04	373.64	4452.00
100	9.04	393.72	4650.48
1000	9.04	517.76	5771.72

Table 5

Effect of varying M_1 in the modified GSFC model on the average number of misclassifications, average CPU times(secs) and average number of integer iterations for samples of size n_1=n_2=n_3=25 generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	$\min \sum_{i=1}^{n} I_i$	CPU (secs)	Integer Iterations
3	10.04	45.64	560.68
5	9.00	95.08	1129.08
10	8.64	162.04	1869.36
15	8.64	204.04	2348.04
25	8.64	248.04	2885.64
50	8.64	302.48	3499.24
75	8.64	325.16	3766.52
100	8.64	346.40	3925.72
1000	8.64	436.72	4847.08

Table 6

Effect of varying the value assigned to the parameter M_2 in the modified GSFC model on the average number of misclassifications for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	M_2=20	M_2=40	M_2=75	M_2=100	M_2=250	M_2=400
3	10.44	10.04	10.04	10.04	10.04	10.04
5	9.40	9.00	9.00	9.00	9.00	9.00
10	9.28	8.64	8.64	8.64	8.64	8.64
15	9.28	8.64	8.64	8.64	8.64	8.64
25	9.28	8.64	8.64	8.64	8.64	8.64
50	9.28	8.64	8.64	8.64	8.64	8.64
75	9.28	8.64	8.64	8.64	8.64	8.64
100	9.28	8.64	8.64	8.64	8.64	8.64
1000	9.28	8.64	8.64	8.64	8.64	8.64

Table 7

Effect of varying the value assigned to the parameter M_2 in the modified GSFC model on the average CPU times(secs) for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	M_2=20	M_2=40	M_2=75	M_2=100	M_2=250	M_2=400
3	49.76	43.28	44.08	45.64	46.96	43.60
5	99.84	94.32	92.76	95.08	100.72	95.20
10	192.04	156.68	158.16	162.04	170.12	160.16
15	242.96	197.36	195.24	204.04	214.68	200.40
25	301.80	242.84	244.20	248.04	264.80	251.64
50	384.64	298.44	291.04	302.48	321.12	302.52
75	436.68	319.32	312.32	325.16	340.20	319.72
100	462.96	339.00	319.36	346.40	364.76	365.28
1000	659.92	467.88	427.88	436.72	430.40	455.84

Table 8

Effect of varying the value assigned to the parameter M_2 in the modified GSFC model on the average number of integer iterations for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	$M_2=20$	$M_2=40$	$M_2=75$	$M_2=100$	$M_2=250$	$M_2=400$
3	626.84	560.92	570.60	560.68	545.60	561.36
5	1160.28	1124.12	1131.36	1129.08	1112.16	1124.84
10	2102.68	1844.88	1864.68	1869.36	1846.44	1860.20
15	2626.80	2314.24	2345.28	2348.04	2319.04	2337.56
25	3216.20	2826.84	2890.92	2885.64	2854.60	2870.92
50	3992.00	3423.40	3485.36	3499.24	3465.48	3497.04
75	4457.08	3685.84	3722.48	3766.52	3736.16	3756.08
100	4701.40	3890.96	3880.12	3925.72	3905.60	3923.72
1000	6594.16	5103.96	4861.56	4847.08	4724.76	4776.68

Table 9

Effect of varying the value assigned to the parameter é in the modified GSFC model on the average number of misclassifications for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	é=5	é=2.5	é=1	é=.5
3	10.04	8.72	8.68	8.68
5	9.00	8.68	8.68	8.68
10	8.64	8.64	8.64	8.64
15	8.64	8.64	8.64	8.64
25	8.64	8.64	8.64	8.64
50	8.64	8.64	8.64	8.64
75	8.64	8.64	8.64	8.64
100	8.64	8.64	8.64	8.64
1000	8.64	8.64	8.64	8.64

Table 10

Effect of varying the value assigned to the parameter é in the modified GSFC model on the average CPU times(secs) for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	é=5	é=2.5	é=1	é=.5
3	45.64	111.04	198.12	262.72
5	95.08	162.24	253.68	298.40
10	162.04	230.48	296.88	338.24
15	204.04	262.92	320.16	358.08
25	248.04	296.08	352.76	375.08
50	302.48	339.56	379.32	393.88
75	325.16	354.80	392.60	402.32
100	346.40	365.44	407.72	409.08
1000	436.72	442.20	434.52	435.92

Table 11

Effect of varying the value assigned to the parameter é in the modified GSFC model on the average number of integer iterations for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\3.4641\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}-2\\0\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}2\\0\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	é=5	é=2.5	é=1	é=.5
3	560.68	1271.40	2285.36	3046.00
5	1129.08	1883.28	2873.04	3495.08
10	1869.36	2664.00	3453.88	3909.48
15	2348.04	3066.04	3720.32	4130.76
25	2885.64	3495.80	4031.32	4356.36
50	3499.24	3942.96	4337.44	4604.52
75	3766.52	4143.68	4489.68	4723.12
100	3925.72	4278.96	4581.44	4791.68
1000	4847.08	4951.36	4983.76	5043.84

Table 12

Effect of varying the value assigned to the parameter e in the modified GSFC model on the average number of misclassifications for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\3.4641\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}-2\\0\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}2\\0\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	e=1	e=.5	e=.1	e=.01	e=.001	e=0
3	16.85*	13.92	10.60	10.04	10.04	10.04
5	12.00	10.20	9.20	9.00	8.96	8.96
10	10.16	9.40	8.80	8.64	8.64	8.64
15	9.56	9.16	8.76	8.64	8.64	8.64
25	9.40	9.04	8.76	8.64	8.64	8.64
50	9.36	9.04	8.76	8.64	8.64	8.64
75	9.36	9.04	8.76	8.64	8.64	8.64
100	9.36	9.04	8.76	8.64	8.64	8.64
1000	9.36	9.04	8.76	8.64	8.64	8.64

* Average in 20 runs. Problem was infeasible in the remaining 5 runs for $M_1=3$, $M_2=100$, e=1 and é=5. The feasibility conditions identified in Theorem 1 are not satisfied for such parameter values.

Table 13

Effect of varying the value assigned to the parameter e in the modified GSFC model on the average CPU times(secs) for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	e=1	e=.5	e=.1	e=.01	e=.001	e=0
3	49.75*	52.04	45.36	45.64	46.08	45.00
5	114.40	103.68	98.00	95.08	95.80	93.16
10	208.44	187.76	162.08	162.04	165.08	158.16
15	218.24	218.64	214.72	204.04	204.80	199.68
25	293.44	279.00	262.00	248.04	248.32	232.20
50	410.76	355.36	324.84	302.48	295.92	282.92
75	461.84	387.08	355.56	325.16	322.96	304.84
100	494.56	408.64	366.04	346.40	339.12	319.64
1000	646.76	537.40	494.48	436.72	434.64	404.80

* Average in 20 runs. Problem was infeasible in the remaining 5 runs for $M_1=3$, $M_2=100$, e=1 and é=5. The feasibility conditions identified in Theorem 1 are not satisfied for such parameter values.

Table 14

Effect of varying the value assigned to the parameter e in the modified GSFC model on the average number of integer iterations for samples of size $n_1=n_2=n_3=25$ generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	e=1	e=.5	e=.1	e=.01	e=.001	e=0
3	557.80*	589.44	548.88	560.68	587.88	580.60
5	1136.16	1066.44	1088.80	1129.08	1127.20	1122.80
10	1872.20	1857.12	1770.28	1869.36	1875.20	1869.28
15	2022.48	2207.92	2375.32	2348.04	2352.12	2339.60
25	2820.36	2829.92	2958.04	2885.64	2894.76	2862.40
50	4026.36	3697.60	3626.68	3499.24	3493.20	3453.68
75	4582.52	4108.12	3951.32	3766.52	3769.44	3711.40
100	4974.76	4335.44	4113.24	3925.72	3915.20	3875.08
1000	6395.28	5516.64	5109.04	4847.08	4829.60	4778.96

* Average in 20 runs. Problem was infeasible in the remaining 5 runs for $M_1=3$, $M_2=100$, e=1 and é=5. The feasibility conditions identified in Theorem 1 are not satisfied for such parameter values.

Table 15

Average number of misclassifications yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

M_1	Modified GSFC	MIP3G
3	13.40	13.40
5	10.92	10.92
10	10.60	10.64
15	10.60	10.64
25	10.60	10.64
50	10.60	10.64
75	10.60	10.64
100	10.60	10.64
1000	10.60	10.64

Table 16

Average CPU times(secs) yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

M_1	Modified GSFC	MIP3G
3	33.20	19.52
5	57.40	33.20
10	115.12	81.40
15	167.64	102.36
25	265.00	143.92
50	417.96	216.60
75	523.56	257.96
100	580.00	290.04
1000	925.12	447.32

Table 17

Average number of integer iterations yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

M_1	Modified GSFC	MIP3G
3	383.80	197.64
5	604.80	333.88
10	1024.60	545.72
15	1397.56	706.24
25	2122.92	1039.56
50	3441.96	1670.36
75	4325.88	2068.16
100	4990.28	2366.04
1000	8575.40	4057.88

Table 18

Average number of misclassifications yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	Modified GSFC	MIP3G
3	12.96	12.96
5	11.32	11.32
10	11.08	11.08
15	11.08	11.08
25	11.08	11.08
50	11.08	11.08
75	11.08	11.08
100	11.08	11.08
1000	11.08	11.08

Table 19

Average CPU times(secs) yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	Modified GSFC	MIP3G
3	43.16	28.32
5	71.20	39.12
10	142.28	86.80
15	201.88	111.16
25	307.60	157.44
50	485.16	234.16
75	596.04	282.16
100	683.16	318.76
1000	1052.80	473.20

Table 20

Average number of integer iterations yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44\,I \qquad \Sigma_3 = .64\,I$$

M_1	Modified GSFC	MIP3G
3	510.32	322.40
5	772.16	413.52
10	1248.88	635.88
15	1723.24	842.00
25	2621.96	1256.96
50	4247.56	2004.32
75	5282.00	2492.80
100	6040.40	2861.24
1000	10121.52	4679.80

Table 21

Average number of misclassifications yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

M_1	Modified GSFC	MIP3G
3	10.36	10.36
5	9.20	9.20
10	9.04	9.04
15	9.04	9.04
25	9.04	9.04
50	9.04	9.04
75	9.04	9.04
100	9.04	9.04
1000	9.04	9.04

Table 22

Average CPU times(secs) yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

M_1	Modified GSFC	MIP3G
3	45.92	20.24
5	103.48	45.00
10	195.52	83.12
15	245.24	104.12
25	295.64	123.88
50	350.60	150.40
75	373.64	160.60
100	393.72	168.08
1000	517.76	206.92

Table 23

Average number of integer iterations yielded by the modified GSFC and the MIP3G when samples of size n_1=n_2=n_3=25 are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

M_1	Modified GSFC	MIP3G
3	3565.48	298.24
5	1184.40	598.24
10	2227.92	1109.96
15	2805.96	1392.52
25	3457.20	1701.92
50	4131.40	2038.64
75	4452.00	2185.36
100	4650.48	2274.24
1000	5771.72	2754.72

Table 24

Average number of misclassifications yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	Modified GSFC	MIP3G
3	10.04	10.04
5	9.00	9.00
10	8.64	8.68
15	8.64	8.68
25	8.64	8.68
50	8.64	8.68
75	8.64	8.68
100	8.64	8.68
1000	8.64	8.68

Table 25

Average CPU times(secs) yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix}0\\3.4641\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}-2\\0\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}2\\0\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	Modified GSFC	MIP3G
3	45.64	21.04
5	95.08	43.92
10	162.04	73.16
15	204.04	92.00
25	248.04	109.84
50	302.48	131.40
75	325.16	143.60
100	346.40	149.64
1000	436.72	181.76

Table 26

Average number of integer iterations yielded by the modified GSFC and the MIP3G when samples of size $n_1=n_2=n_3=25$ are generated from multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 0 \\ 3.4641 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = 1.44I \qquad \Sigma_3 = .64I$$

M_1	Modified GSFC	MIP3G
3	560.68	301.60
5	1129.08	565.40
10	1869.36	942.68
15	2348.04	1184.32
25	2885.64	1450.28
50	3499.24	1741.00
75	3766.52	1869.12
100	3925.72	1951.48
1000	4847.08	2350.48

Table 27

Percentage misclassification rates in the training sample and the holdout sample for data generated from multivariate normal populations with collinear means and equal variances

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

Case 1

Method	Training Sample	Holdout Sample
GMFC	13.493	24.400
MIP3G	14.187	24.711
MSDMG	18.613	23.244
FLDF	19.147	21.822
SQDF	18.933	22.533

Table 28

Percentage misclassification rates in the training sample and the holdout sample for data generated from multivariate normal populations with collinear means and unequal variances

$$\mu_1 = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}2\\2\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}3\\3\end{bmatrix}$$

$$\Sigma_1 = .64I \qquad \Sigma_2 = .81I \qquad \Sigma_3 = 1.21I$$

Case 2

Method	Training Sample	Holdout Sample
GMFC	11.893	21.022
MIP3G	12.373	21.644
MSDMG	16.373	20.578
FLDF	16.960	19.956
SQDF	16.160	20.089

Table 29

Percentage misclassification rates in the training sample and the holdout sample for data generated from multivariate normal populations with equilateral (equidistant) means and equal variances

$$\mu_1 = \begin{bmatrix} 0 \\ 173.205 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -100 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 100 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = 5625\,I \qquad \Sigma_2 = 5625\,I \qquad \Sigma_3 = 5625\,I$$

Case 3

Method	Training Sample	Holdout Sample
GMFC	8.693	16.711
MIP3G	23.093	36.444
Grid	13.760	24.044
MSDMG	32.213	39.244
FLDF	13.973	14.711
SQDF	13.440	15.156

Table 30

Percentage misclassification rates in the training sample and the holdout sample for data generated from multivariate normal populations with equilateral (equidistant) means and unequal variances

$$\mu_1 = \begin{bmatrix} 0 \\ 173.205 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -100 \\ 0 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 100 \\ 0 \end{bmatrix}$$

$$\Sigma_1 = 6400I \qquad \Sigma_2 = 5625I \qquad \Sigma_3 = 4900I$$

Case 4

Method	Training Sample	Holdout Sample
GMFC	8.160	15.867
MIP3G	22.560	35.689
Grid	13.973	24.533
MSDMG	32.320	39.822
FLDF	13.813	14.578
SQDF	13.760	14.978

Table 31

Percentage misclassification rates in the training sample and the holdout sample for data generated from multivariate normal populations with a proportion p=.10 of outlier observations in the leftmost and rightmost populations where all five means are collinear

$$\mu_1 = \begin{bmatrix}13\\13\end{bmatrix} \qquad \mu_2 = \begin{bmatrix}15\\15\end{bmatrix} \qquad \mu_3 = \begin{bmatrix}17\\17\end{bmatrix}$$

$$\Sigma_1 = I \qquad \Sigma_2 = I \qquad \Sigma_3 = I$$

$$\mu_1^{(o)} = \begin{bmatrix}0\\0\end{bmatrix} \qquad \mu_3^{(o)} = \begin{bmatrix}30\\30\end{bmatrix}$$

$$\Sigma_1^{(o)} = 9I \qquad \Sigma_3^{(o)} = 9I$$

Case 5

Method	Training Sample	Holdout Sample
GMFC	5.120	12.089
MIP3G	7.787	12.044
MSDMG	9.973	10.311
FLDF	21.707	26.800
SQDF	12.960	16.667

Table 32

Percentage misclassification rates in the training sample and the holdout sample for data generated from continuous uniform populations with collinear means and equal variances

U_1 (x, y; 0, 10, 0, 10)

U_2 (x, y; 6, 10, 0, 10)

U_3 (x, y; 14, 24, 0, 10)

Case 6

Method	Training Sample	Holdout Sample
GMFC	10.720	19.156
MIP3G	14.507	24.311
MSDMG	20.907	23.333
FLDF	18.720	18.311
SQDF	17.280	17.244

Table 33

Percentage misclassification rates in the training sample and the holdout sample for data generated from continuous uniform populations with collinear means and unequal variances

U_1 (x, y; 0, 10, 0, 10)

U_2 (x, y; 8, 16, -4, 14)

U_3 (x, y; 14, 19, -6, 16)

Case 7

Method	Training Sample	Holdout Sample
GMFC	14.133	24.622
MIP3G	15.040	24.000
MSDMG	22.080	24.000
FLDF	23.253	23.733
SQDF	21.333	23.200

Table 34

Percentage misclassification rates in the training sample and the holdout sample for data generated from continuous uniform populations with equilateral (equidistant) means and equal variances

U_1 (x, y; 0, 10, 0, 10)

U_2 (x, y; 8, 18, 0, 10)

U_3 (x, y; 4, 14, 6.928, 16.928)

Case 8

Method	Training Sample	Holdout Sample
GMFC	5.227	13.111
MIP3G	19.307	31.733
Grid	11.040	21.111
MSDMG	26.400	32.533
FLDF	12.373	12.356
SQDF	10.187	11.689

Table 35

Percentage misclassification rates in the training sample and the holdout sample for data generated from continuous uniform populations with equilateral (equidistant) means and unequal variances

U_1 (x, y; 0, 10, 0, 10)

U_2 (x, y; 6, 24, 4, 6)

U_3 (x, y; 8, 12, 6.66, 20.66)

Case 9

Method	Training Sample	Holdout Sample
GMFC	7.147	14.622
MIP3G	15.787	26.622
Grid	11.893	21.867
MSDMG	20.213	26.667
FLDF	14.400	15.022
SQDF	11.200	11.956

Table 36

Percentage misclassification rates in the training sample and the holdout sample for data generated from the following:

Group 1: Multivariate normal populations with

$$\mu_1 = \begin{bmatrix} 3.25 \\ 0 \end{bmatrix} \text{ and } \Sigma_1 = 1.2I$$

Group 2: Multivariate normal population with

$$\mu_2 = \begin{bmatrix} 9.75 \\ 0 \end{bmatrix} \text{ and } \Sigma_2 = 1.2I$$

Group 3: Consists of three uniform populations

U_1 (x, y; 2.3, 11.5, 3.5, 4)

U_2 (x, y; 2.3, 11.5, -4, -3.5)

U_3 (x, y; -.5, .5, -2.5, 2.5)

where the proportions of observations in U_1, U_2, and U_3 are .275, .275 and .45, respectively.

Case 10

Method	Training Sample	Holdout Sample
GMFC	16.00	23.24
MIP3G	17.33	22.18
Grid	9.76	17.69
MSDMG	46.83	52.98
FLDF	35.68	40.71
SQDF	12.43	18.80

Table 37

Percentage misclassification rates in the training sample for data generated from the following:

Group 1: Uniform in the Y direction over the interval [-.5,.5] and U-shaped in the X direction with

$$f(x) = \begin{cases} \frac{3x^2}{128} & -4 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Group 2: Uniform in the X direction over the interval [-4,-3] and U-shaped in the Y direction with

$$f(y) = \begin{cases} \frac{3y^2}{128} & -4 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Group 3: Uniform in the X direction over the interval [3,4] and U-shaped in the Y direction with

$$f(y) = \begin{cases} \frac{3y^2}{128} & -4 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Case 11

Method	Training Sample	Holdout Sample
GMFC	18.88	24.27
MIP3G	19.20	22.89
Grid	2.88	5.51
MSDMG	32.69	32.67
FLDF	30.51	30.71
SQDF	15.89	16.09

APPENDIX B

DATA SETS

Data Set 1

MBA Admissions Data

Observation Number	Admitted		Not Admitted		Borderline	
	GPA	GMAT	GPA	GMAT	GPA	GMAT
1	2.96	596	2.54	446	2.86	494
2	3.14	473	2.43	425	2.85	496
3	3.22	482	2.20	474	3.14	419
4	3.29	527	2.36	531	3.28	371
5	3.69	505	2.57	542	2.89	447
6	3.46	693	2.35	406	3.15	313
7	3.03	626	2.51	412	3.50	402
8	3.19	663	2.51	458	2.89	485
9	3.63	447	2.36	399	2.80	444
10	3.59	588	2.36	482	3.13	416
11	3.30	563	2.66	420	3.01	471
12	3.40	553	2.68	414	2.79	490
13	3.50	572	2.48	533	2.89	431
14	3.78	591	2.46	509	2.91	446
15	3.44	692	2.63	504	2.75	546
16	3.48	528	2.44	336	2.73	467
17	3.47	552	2.13	408	3.12	463
18	3.35	520	2.41	469	3.08	440
19	3.39	543	2.55	538	3.03	419
20	3.28	523	2.31	505	3.00	509
21	3.21	530	2.41	489	3.03	438
22	3.58	564	2.19	411	3.05	399
23	3.33	565	2.35	321	2.85	483
24	3.40	431	2.60	394	3.01	453
25	3.38	605	2.55	528	3.03	414
26	3.26	664	2.72	399	3.04	446
27	3.60	609	2.85	381		
28	3.37	559	2.90	384		
29	3.80	521				
30	3.76	646				
31	3.24	467				

Data Set 2

Observation Number	Group 1	Group 2	Group 3
1	(-5,-3)	(21,22)	(26,26)
2	(8,7)	(23,23)	(26,27)
3	(10,9)	(23,23)	(18,18)
4	(15,14)	(15,14)	(8,7)
5	(11,9)	(16,15)	
6	(16,15)		
7	(17,16)		
8	(18,18)		
9	(15,13)		
10	(8,5)		
11	(18,17)		
12	(15,16)		
13	(17,15)		
14	(24,24)		
15	(24,25)		
16	(28,27)		
17	(40,40)		
18	(28,29)		
19	(30,34)		
20	(31,32)		

APPENDIX C

ILLUSTRATIONS

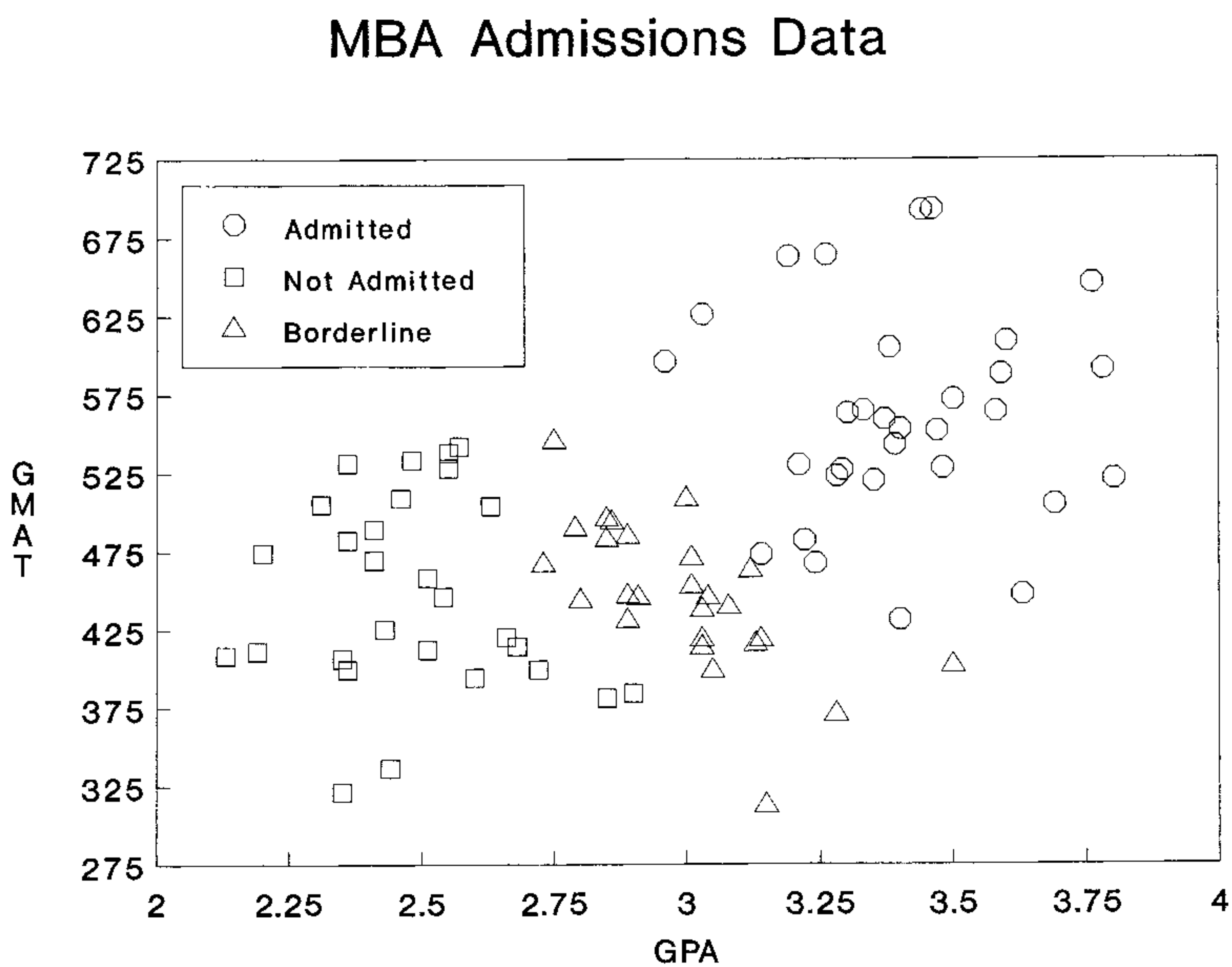

MBA Admissions Data
Admitted
Not Admitted
Borderline
GMAT
725
675
625
575
525
475
425
375
325
275
2
2.25
2.5
2.75
3
3.25
3.5
3.75
4
GPA

Figure 1.1

MBA Admissions Data

Observations Misclassified by FLDF

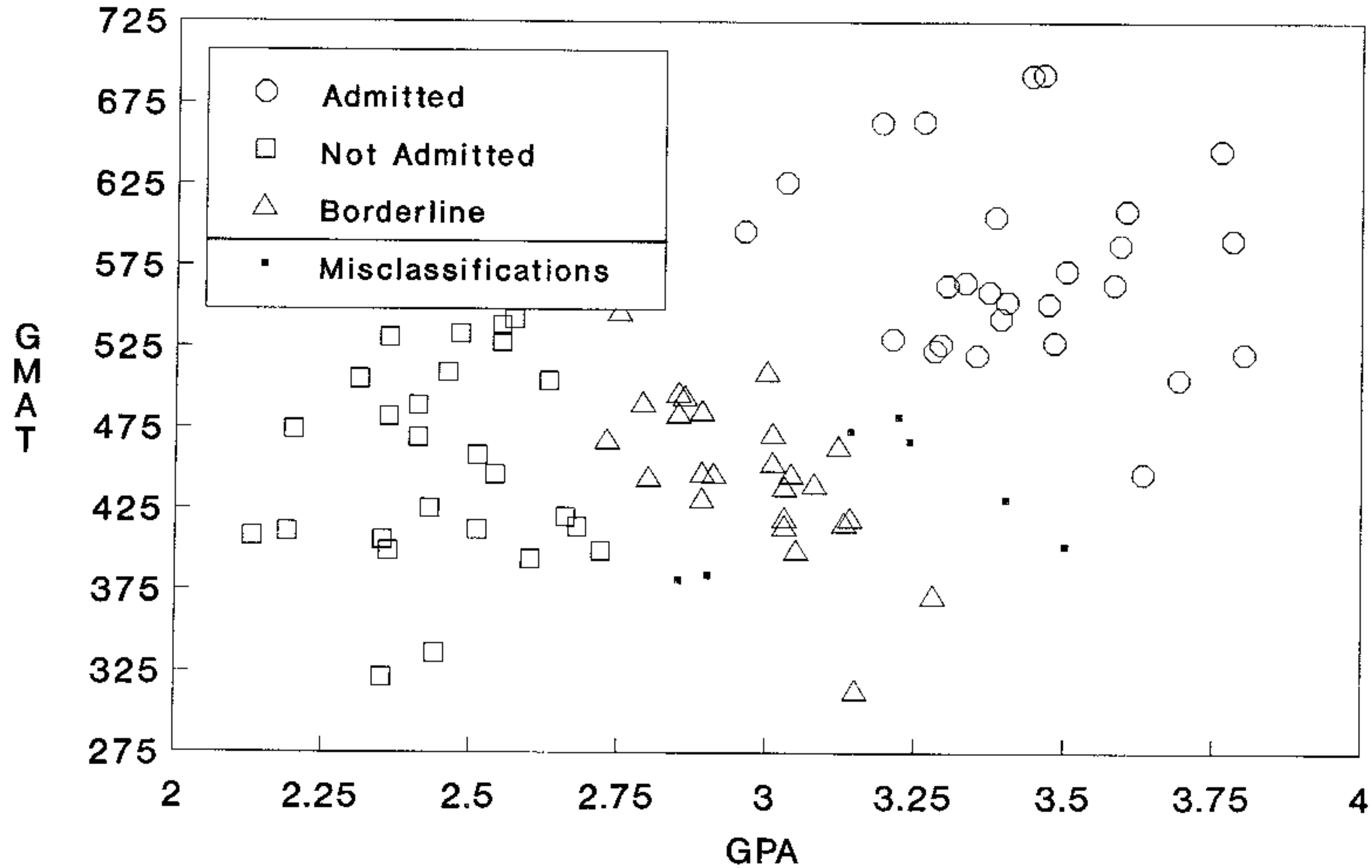


Figure 1.2

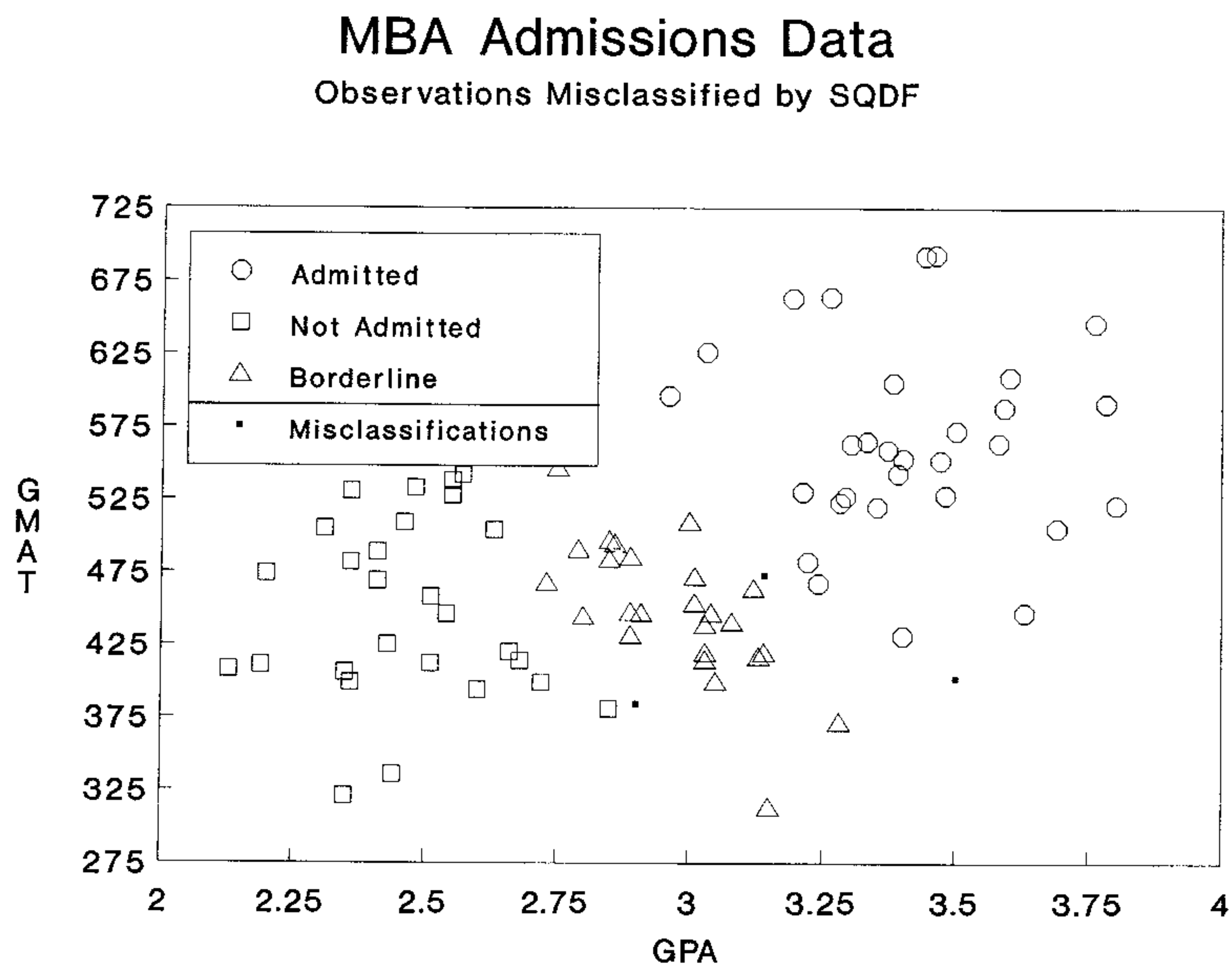

MBA Admissions Data
Observations Misclassified by SQDF
Admitted
Not Admitted
Borderline
Misclassifications
GMAT
725
675
625
575
525
475
425
375
325
275
GPA
2
2.25
2.5
2.75
3
3.25
3.5
3.75
4

Figure 1.3

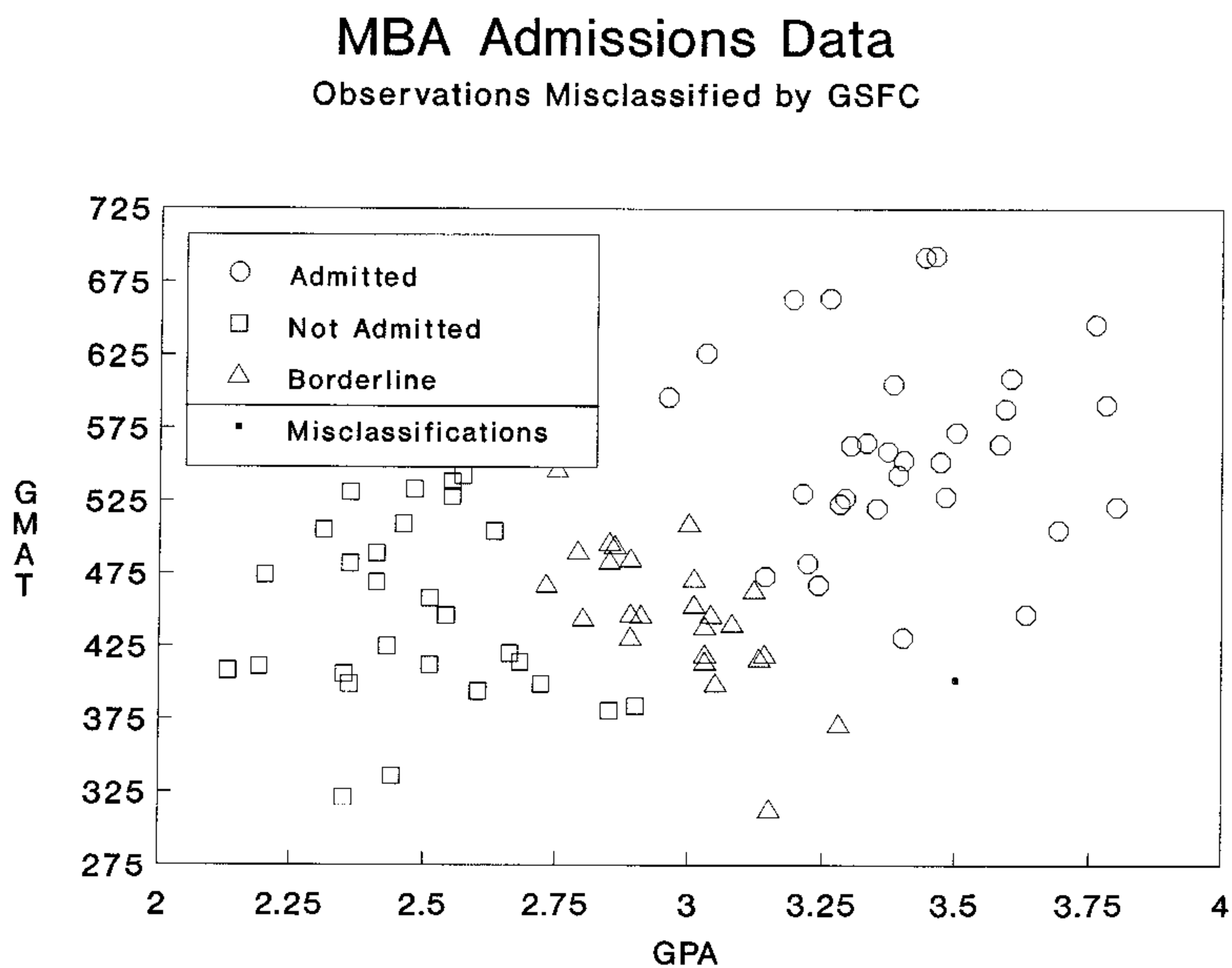
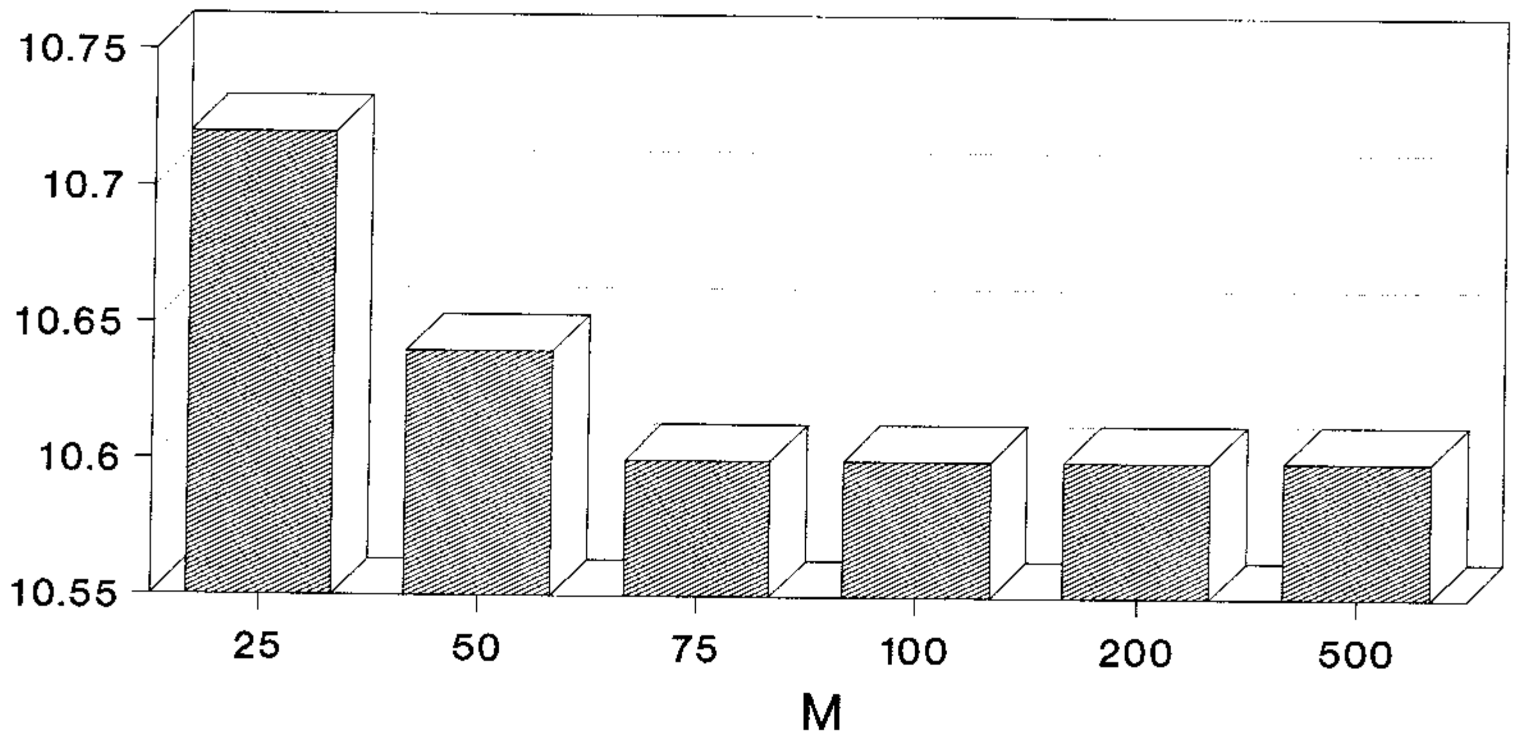

MBA Admissions Data
Observations Misclassified by GSFC
Admitted
Not Admitted
Borderline
Misclassifications
GMAT
725
675
625
575
525
475
425
375
325
275
2
2.25
2.5
2.75
3
3.25
3.5
3.75
4
GPA

Figure 1.4

VARYING PARAMETER M IN THE GSFC

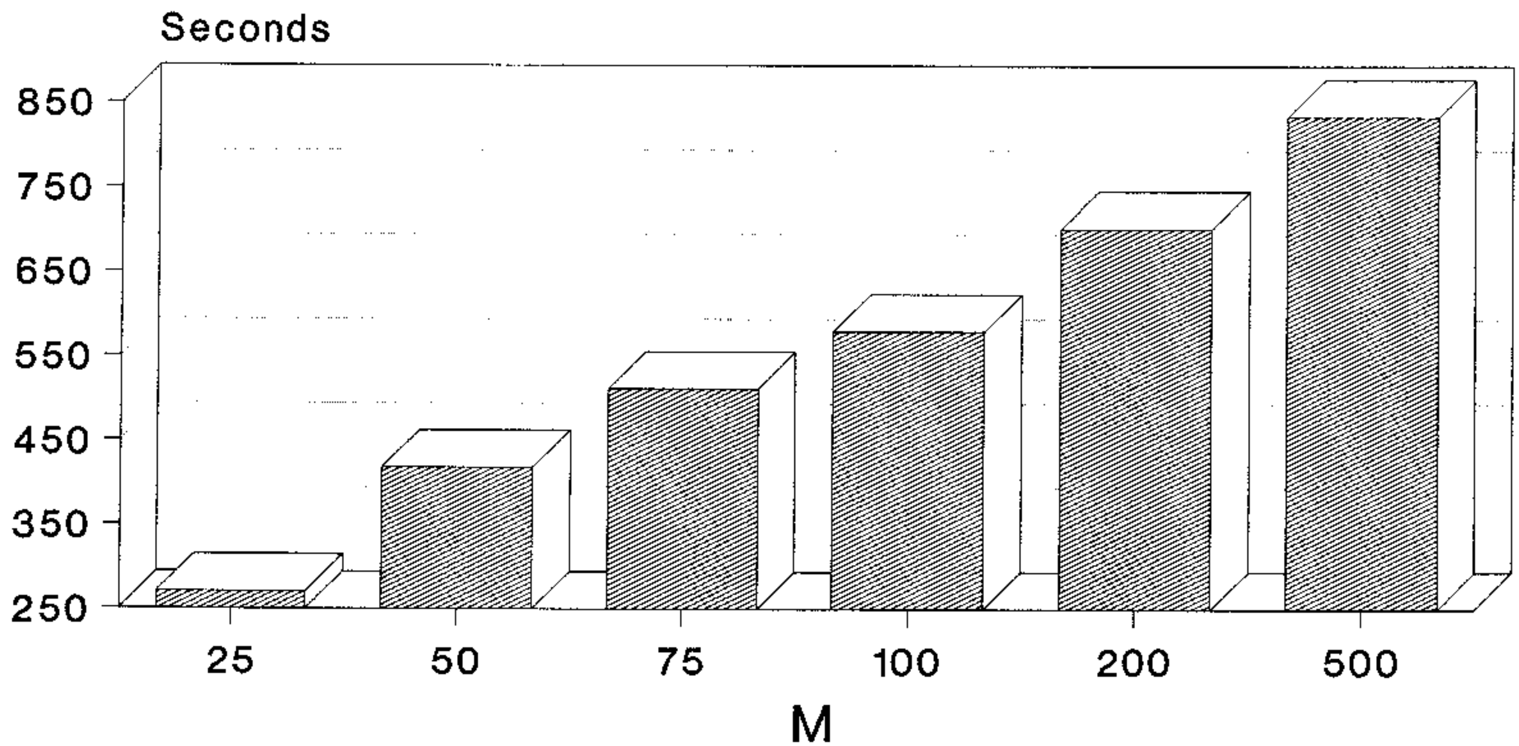
Effect of Magnitude of M on the Average Number of Misclassifications



Normal Populations with Collinear Means & Equal Variances

Figure 2.1

VARYING PARAMETER M IN THE GSFC
Effect of Magnitude of M
on the Average CPU Times

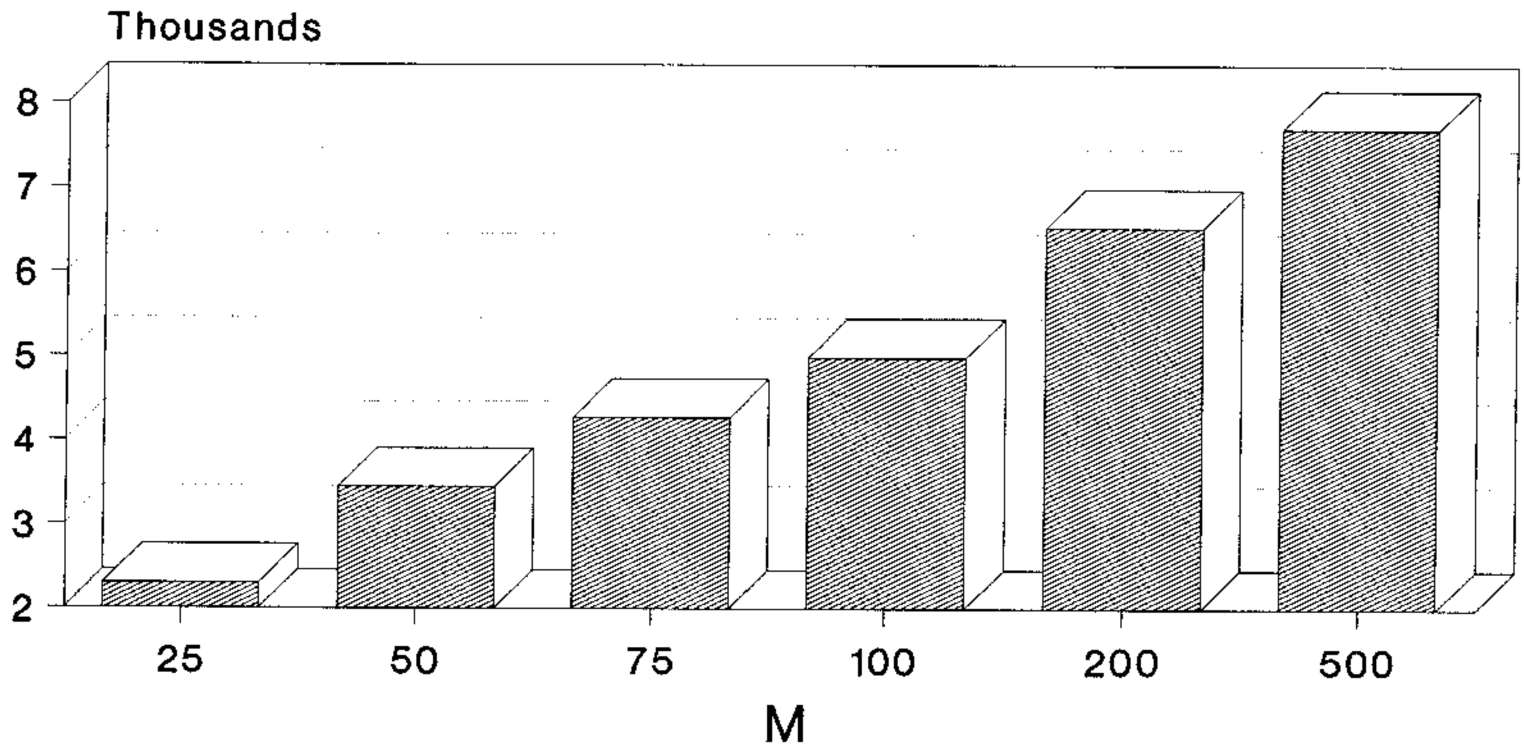


Normal Populations with
Collinear Means & Equal Variances

Figure 2.2

VARYING PARAMETER M IN THE GSFC

Effect of Magnitude of M on the Average Number of Integer Iterations

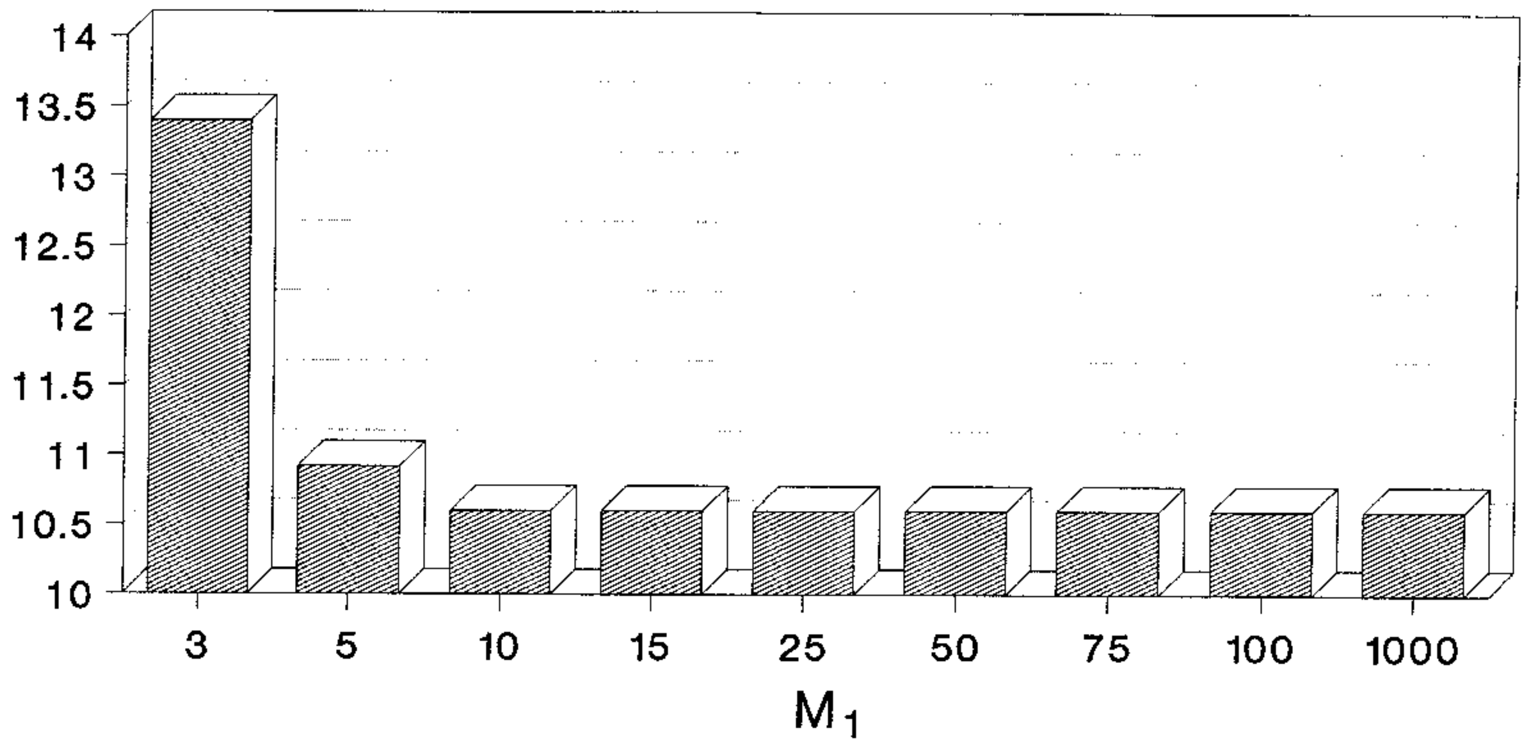


Normal Populations with
Collinear Means & Equal Variances

Figure 2.3

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average Number of Misclassifications

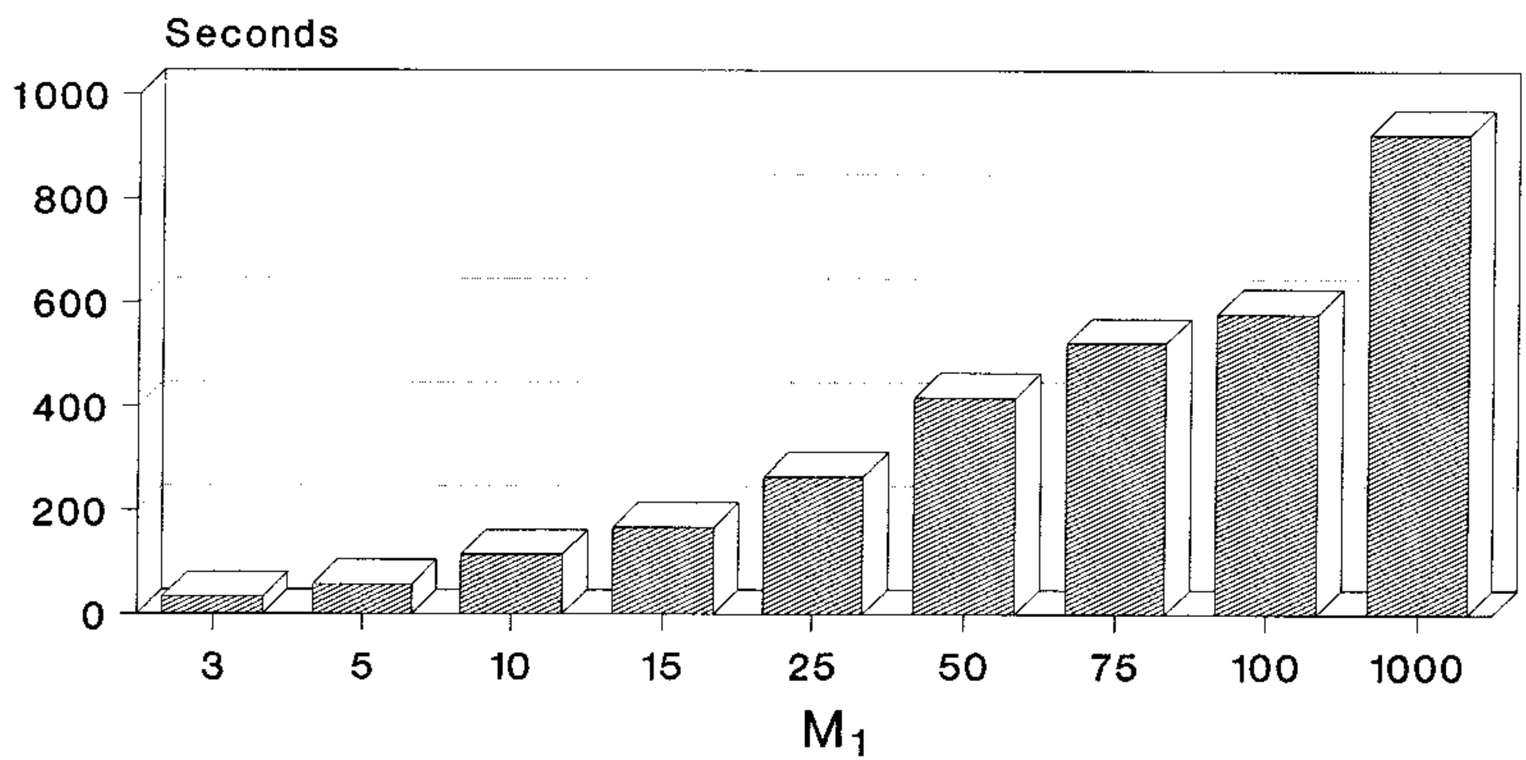


Normal Populations with
Collinear Means & Equal Variances

Figure 3.1

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average CPU Times

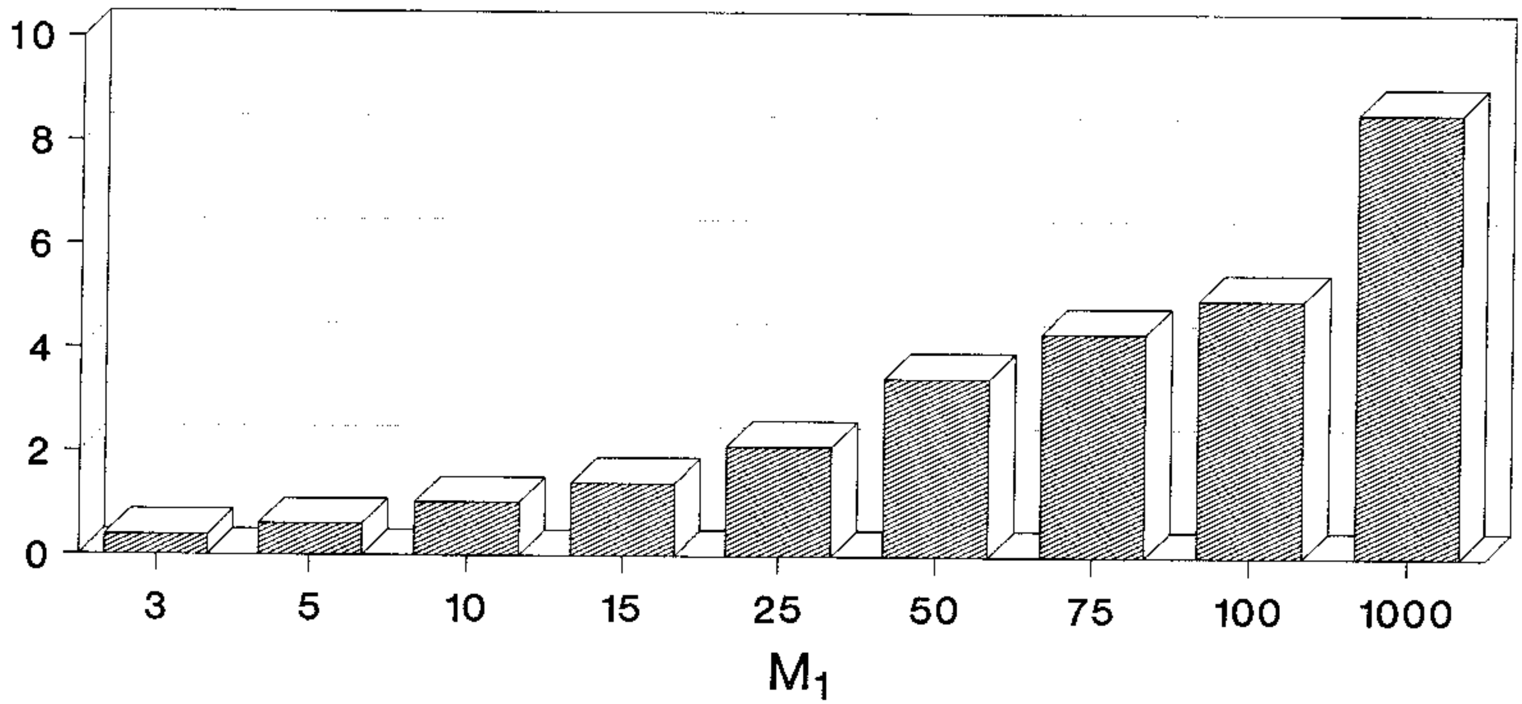


Normal Populations with
Collinear Means & Equal Variances

Figure 3.2

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average Number of Integer Iterations

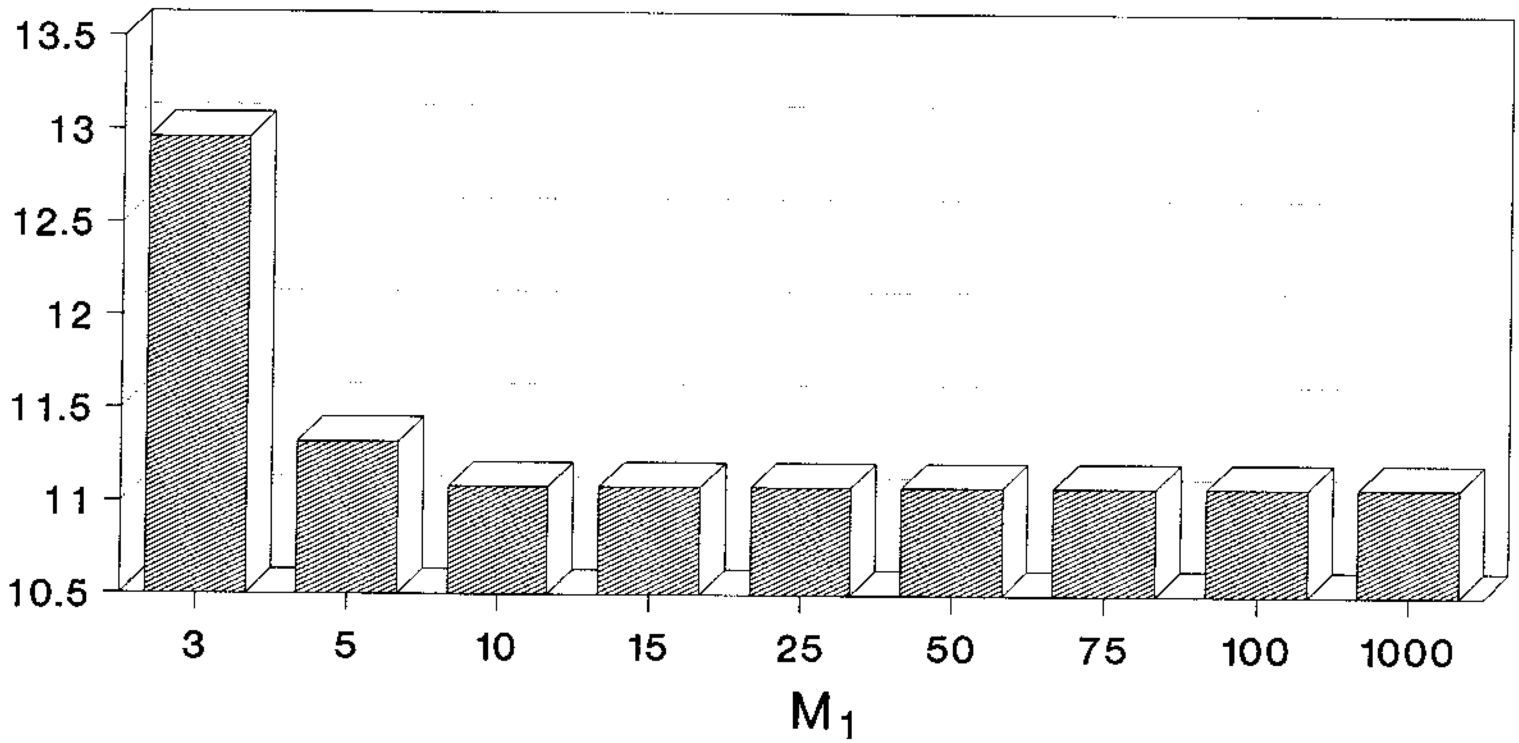


Normal Populations with
Collinear Means & Equal Variances

Figure 3.3

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average Number of Misclassifications

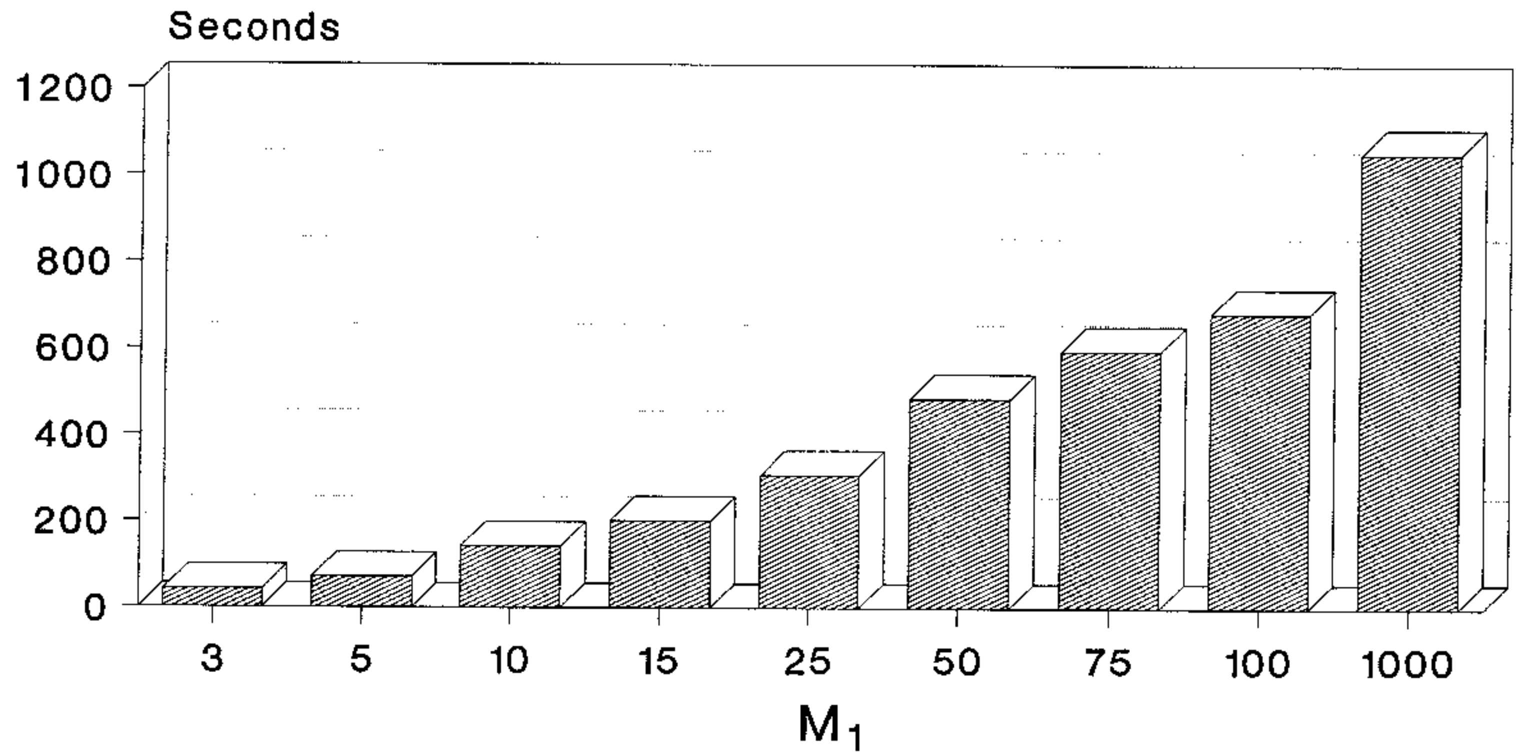


Normal Populations with Collinear Means & Unequal Variances

Figure 4.1

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average CPU Times

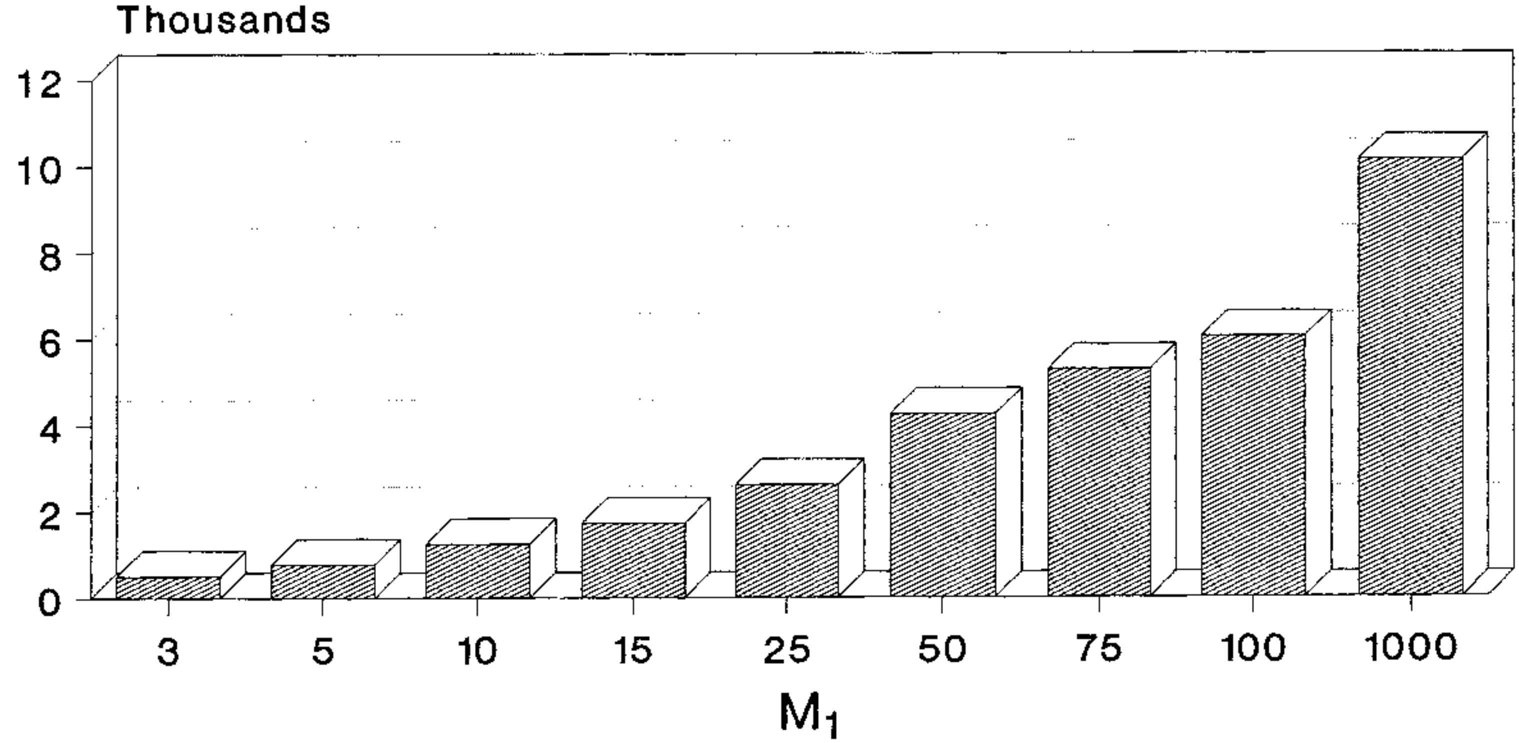


Normal Populations with Collinear Means & Unequal Variances

Figure 4.2

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average Number of Integer Iterations

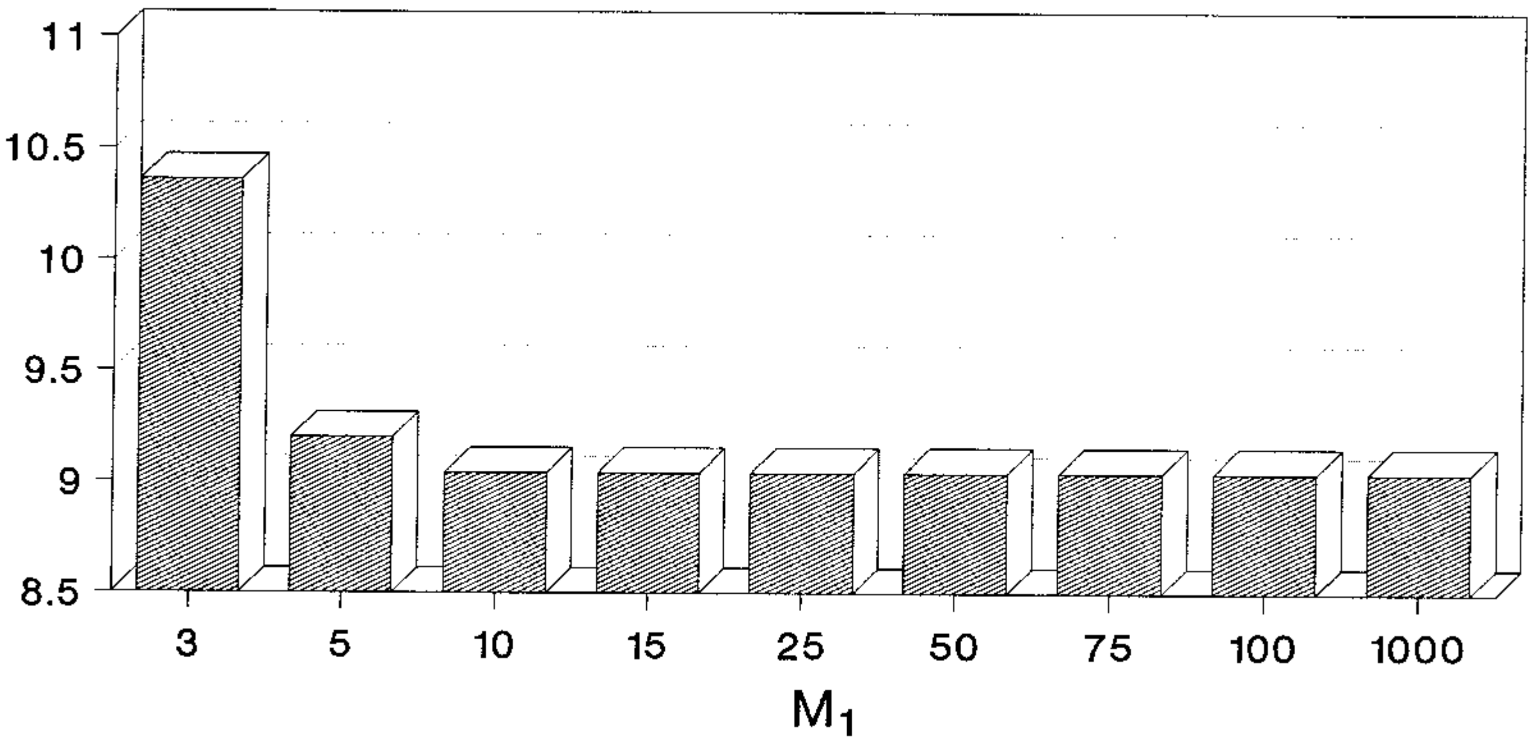


Normal Populations with
Collinear Means & Unequal Variances

Figure 4.3

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average Number of Misclassifications

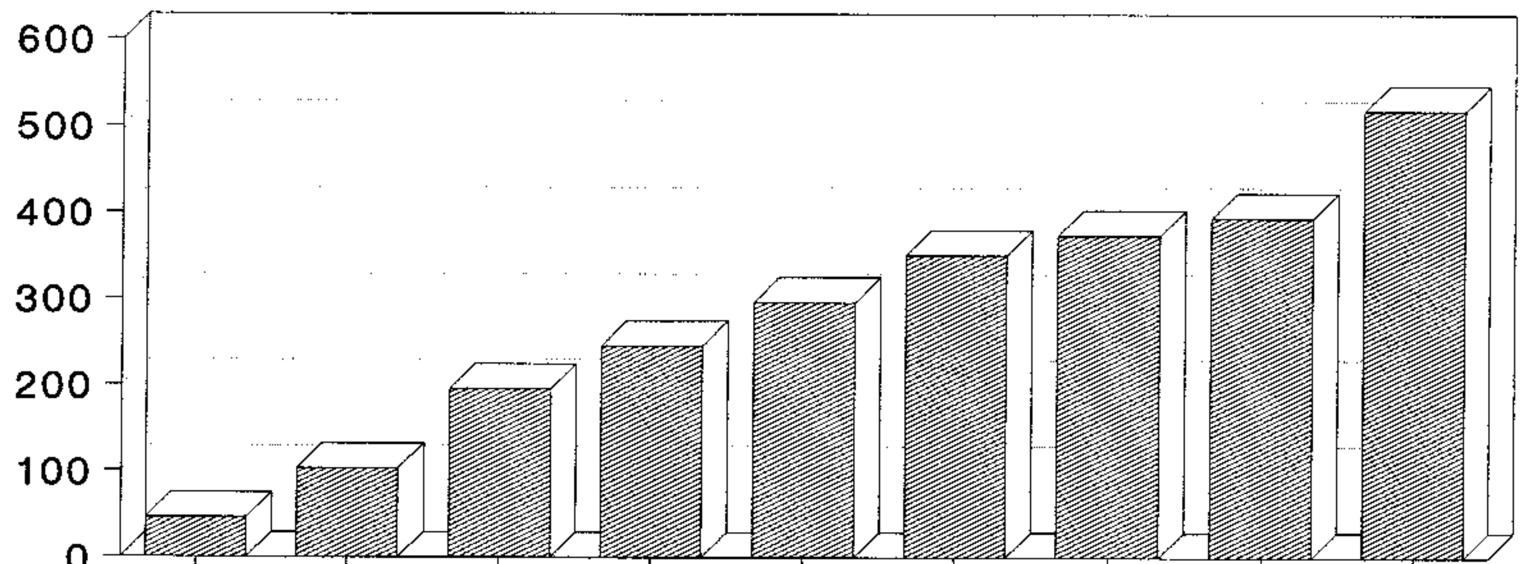


Normal Populations with
Equidistant Means & Equal Variances

Figure 5.1

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1
on the Average CPU Times

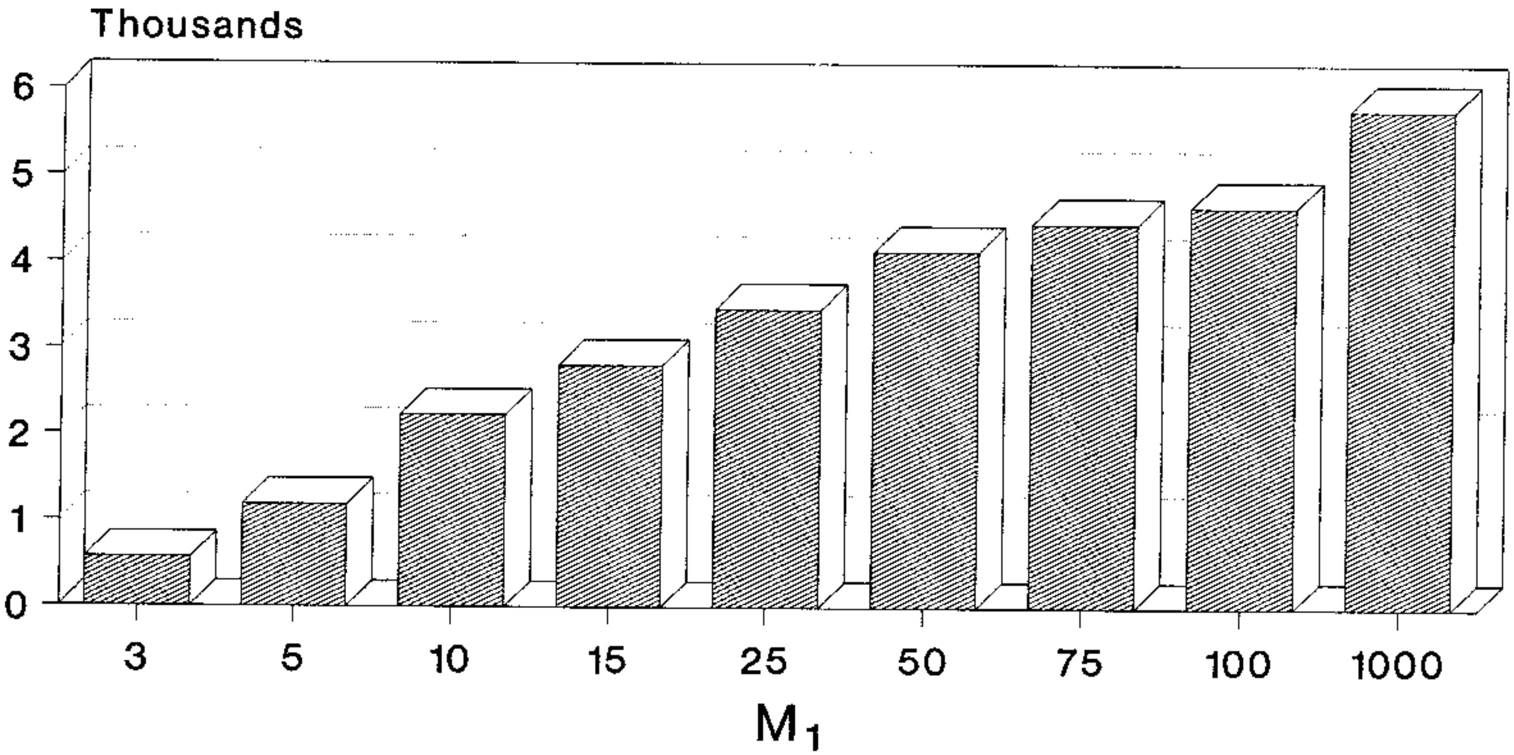


Normal Populations with
Equidistant Means & Equal Variances

Figure 5.2

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average Number of Integer Iterations

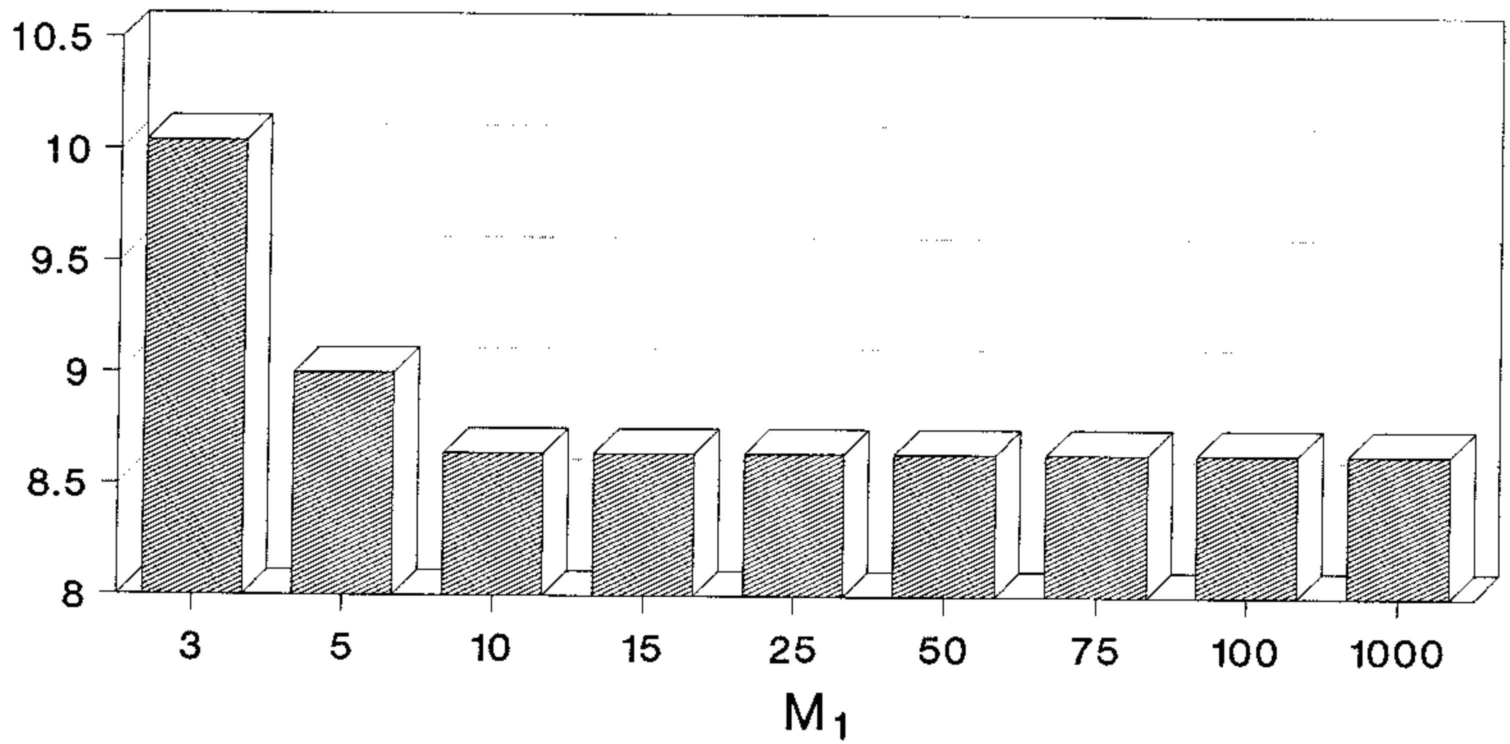


Normal Populations with Equidistant Means & Equal Variances

Figure 5.3

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average Number of Misclassifications

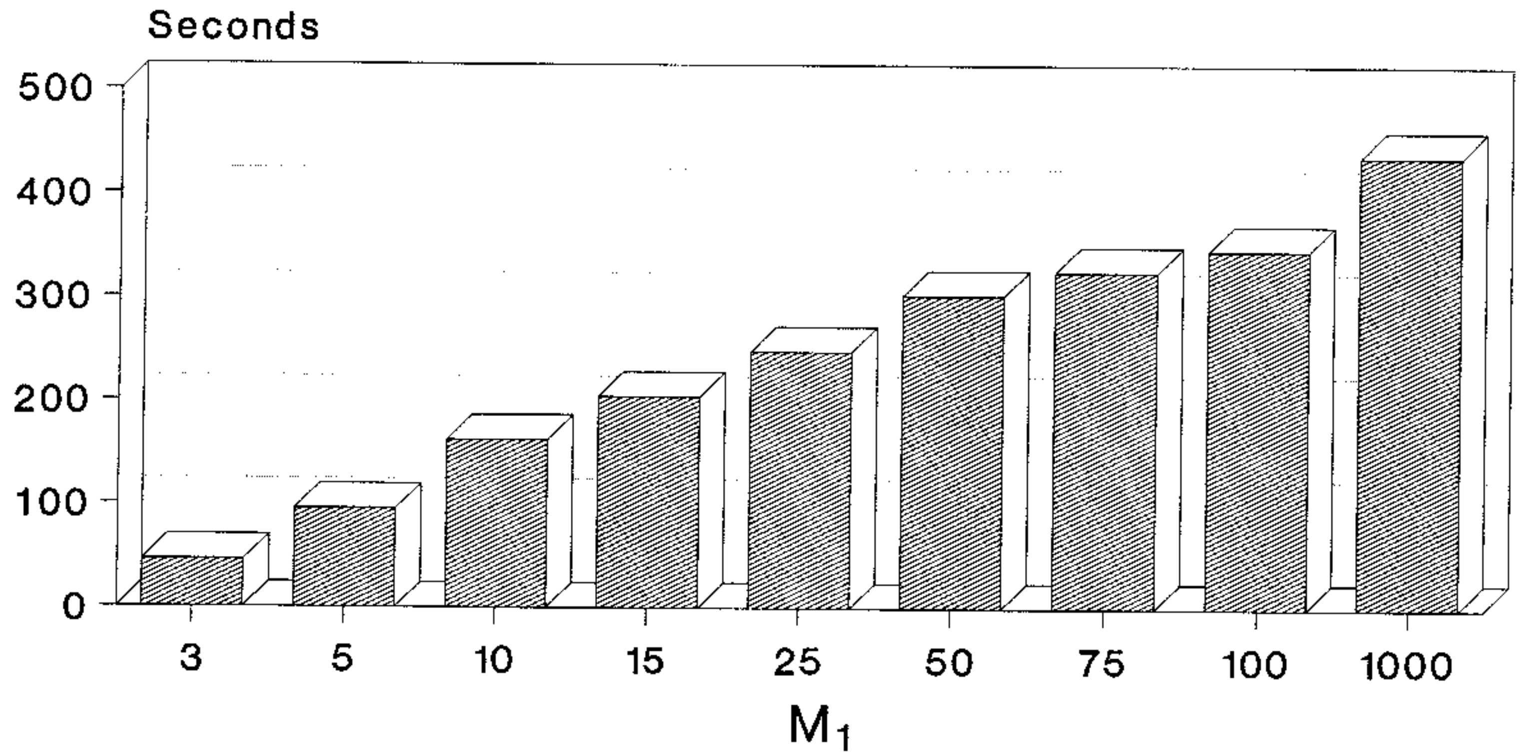


Normal Populations with
Equidistant Means & Unequal Variances

Figure 6.1

VARYING M_1 IN THE MODIFIED GSFC

Effect of Magnitude of M_1 on the Average CPU Times

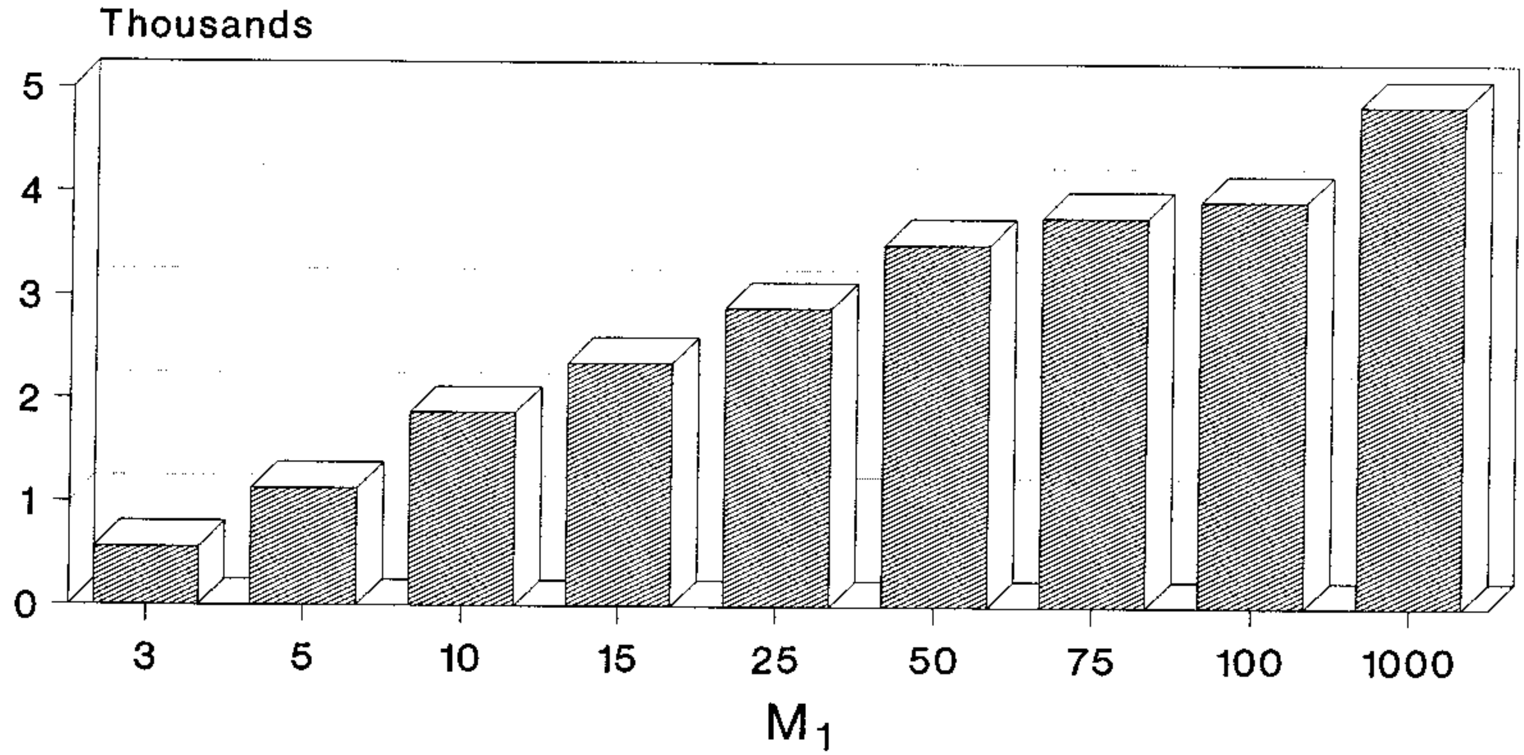


Normal Populations with Equidistant Means & Unequal Variances

Figure 6.2

VARYING M_1 IN THE MODIFIED GSFC

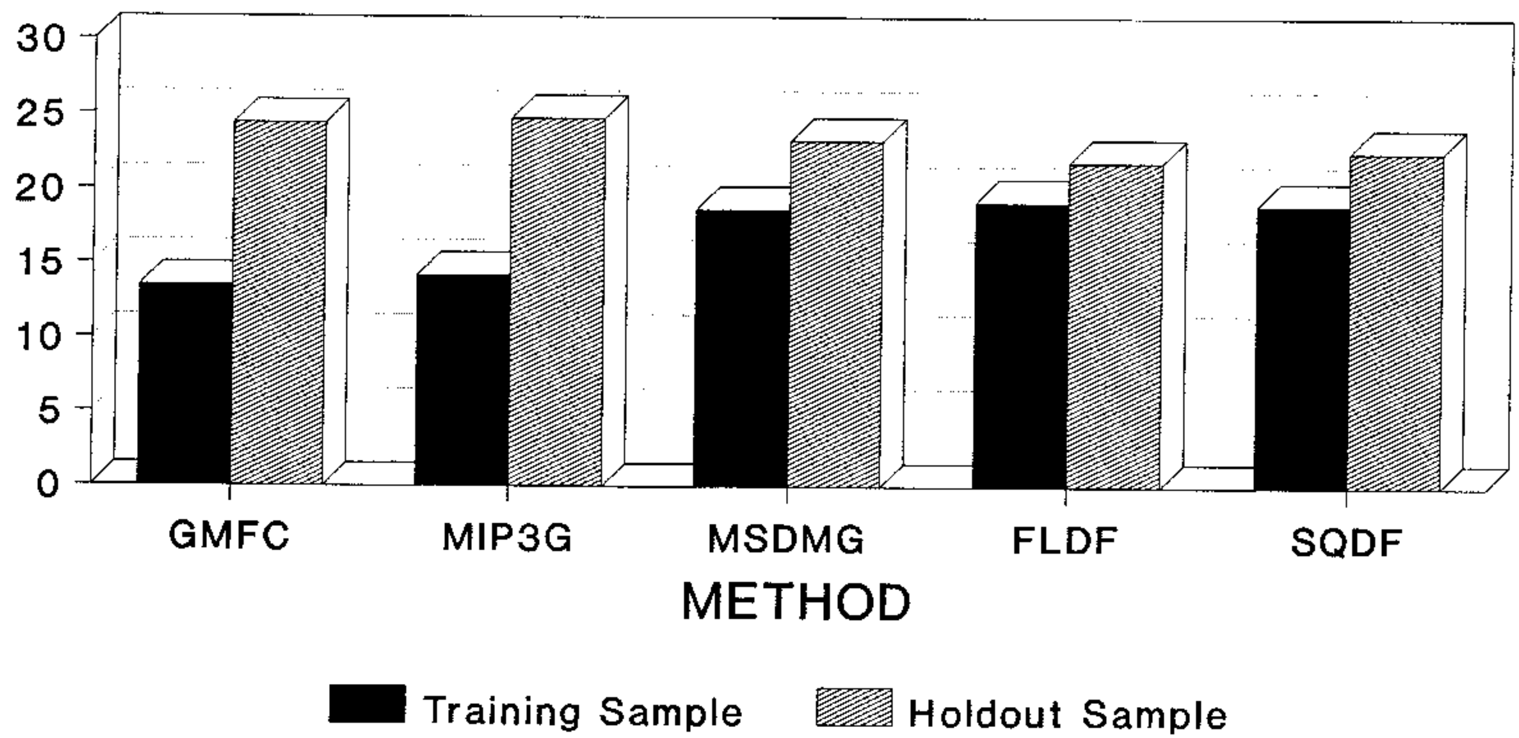
Effect of Magnitude of M_1 on the Average Number of Integer Iterations



Normal Populations with
Equidistant Means & Unequal Variances

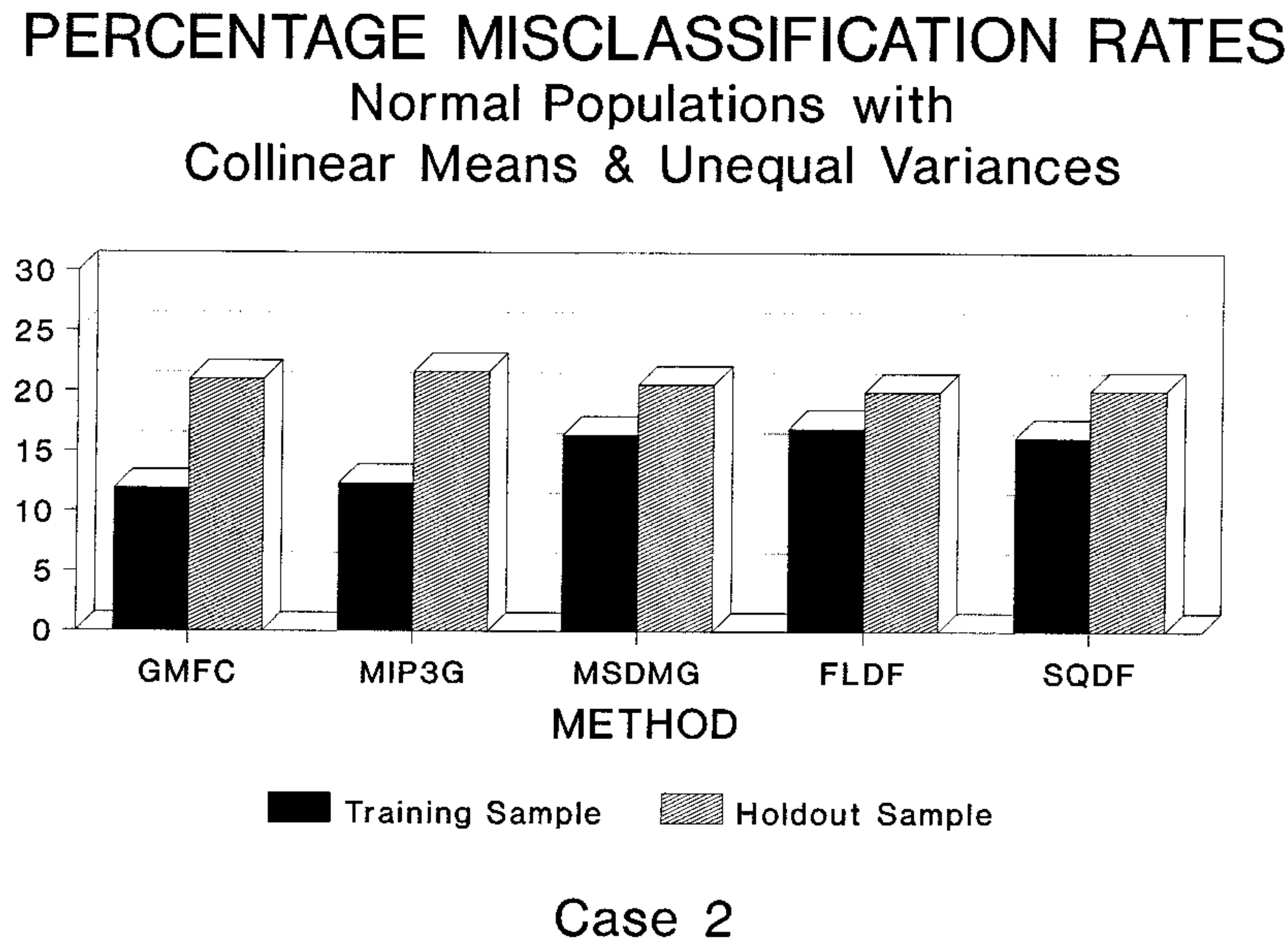
Figure 6.3

PERCENTAGE MISCLASSIFICATION RATES
Normal Populations with
Collinear Means & Equal Variances



Case 1

Figure 7.1



PERCENTAGE MISCLASSIFICATION RATES
Normal Populations with
Collinear Means & Unequal Variances
30
25
20
15
10
5
0
GMFC
MIP3G
MSDMG
FLDF
SQDF
METHOD
Training Sample
Holdout Sample
Case 2

Figure 7.2

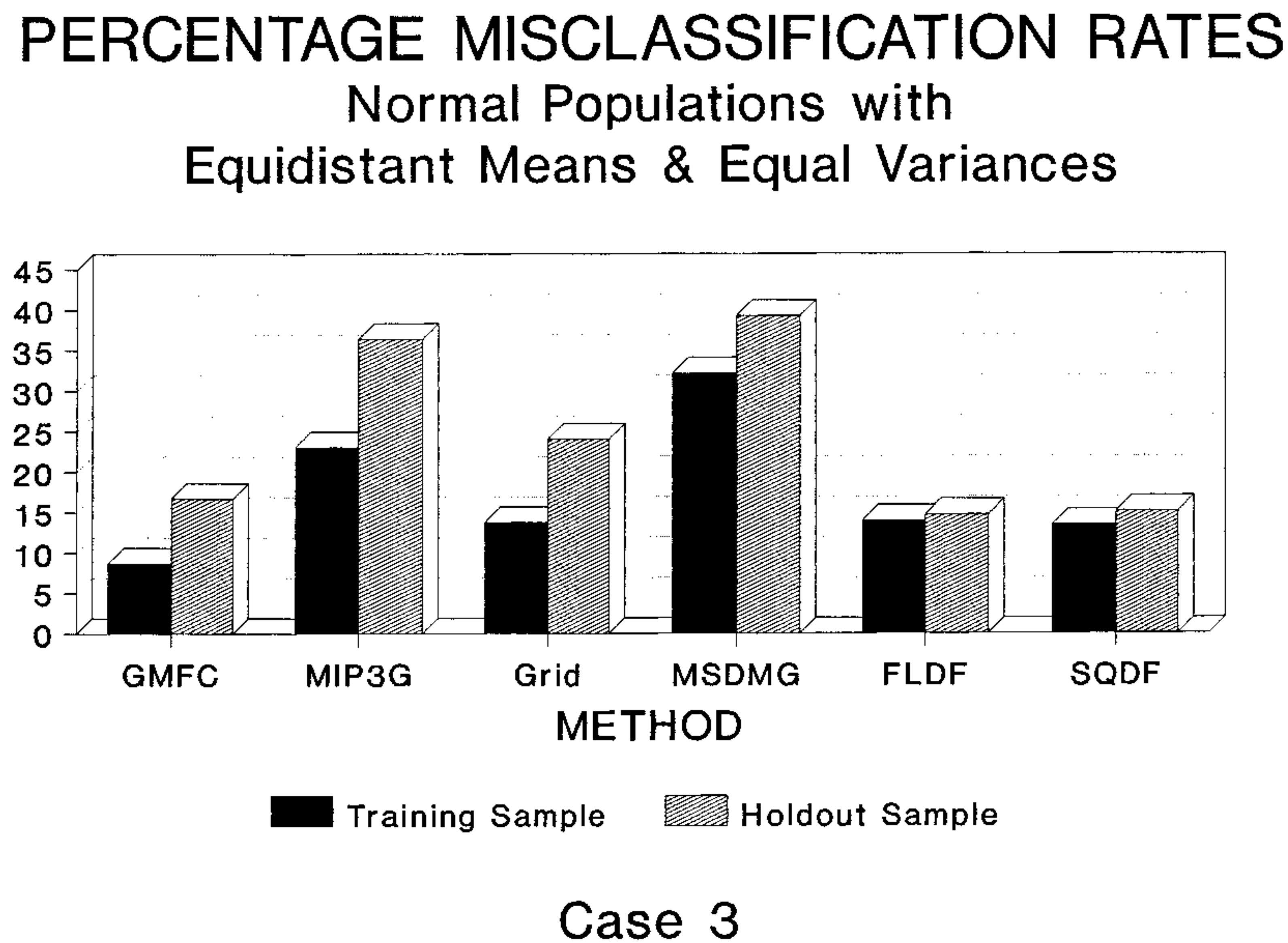

PERCENTAGE MISCLASSIFICATION RATES
Normal Populations with
Equidistant Means & Equal Variances
45
40
35
30
25
20
15
10
5
0
GMFC
MIP3G
Grid
MSDMG
FLDF
SQDF
METHOD
Training Sample
Holdout Sample
Case 3

Figure 7.3

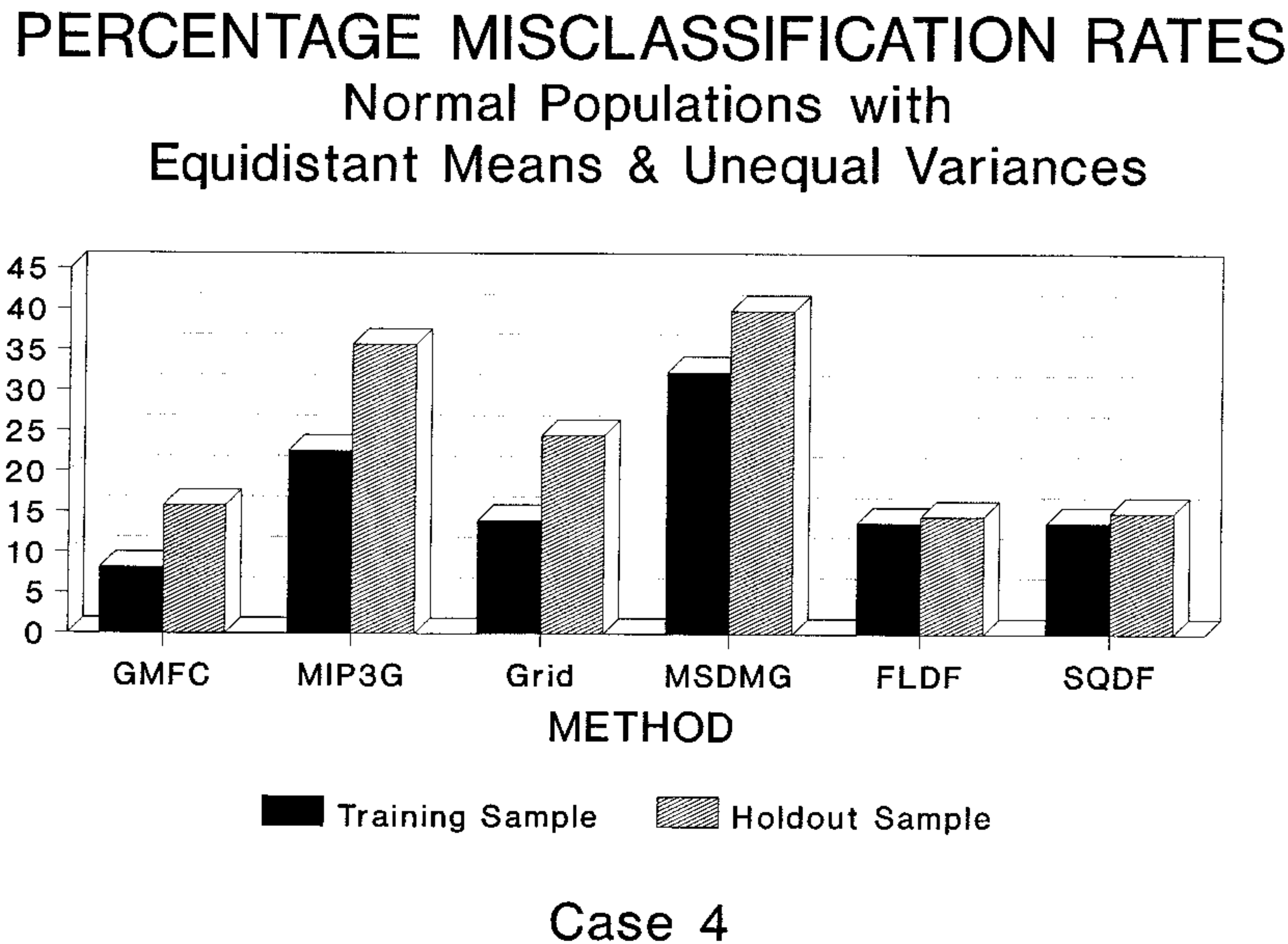
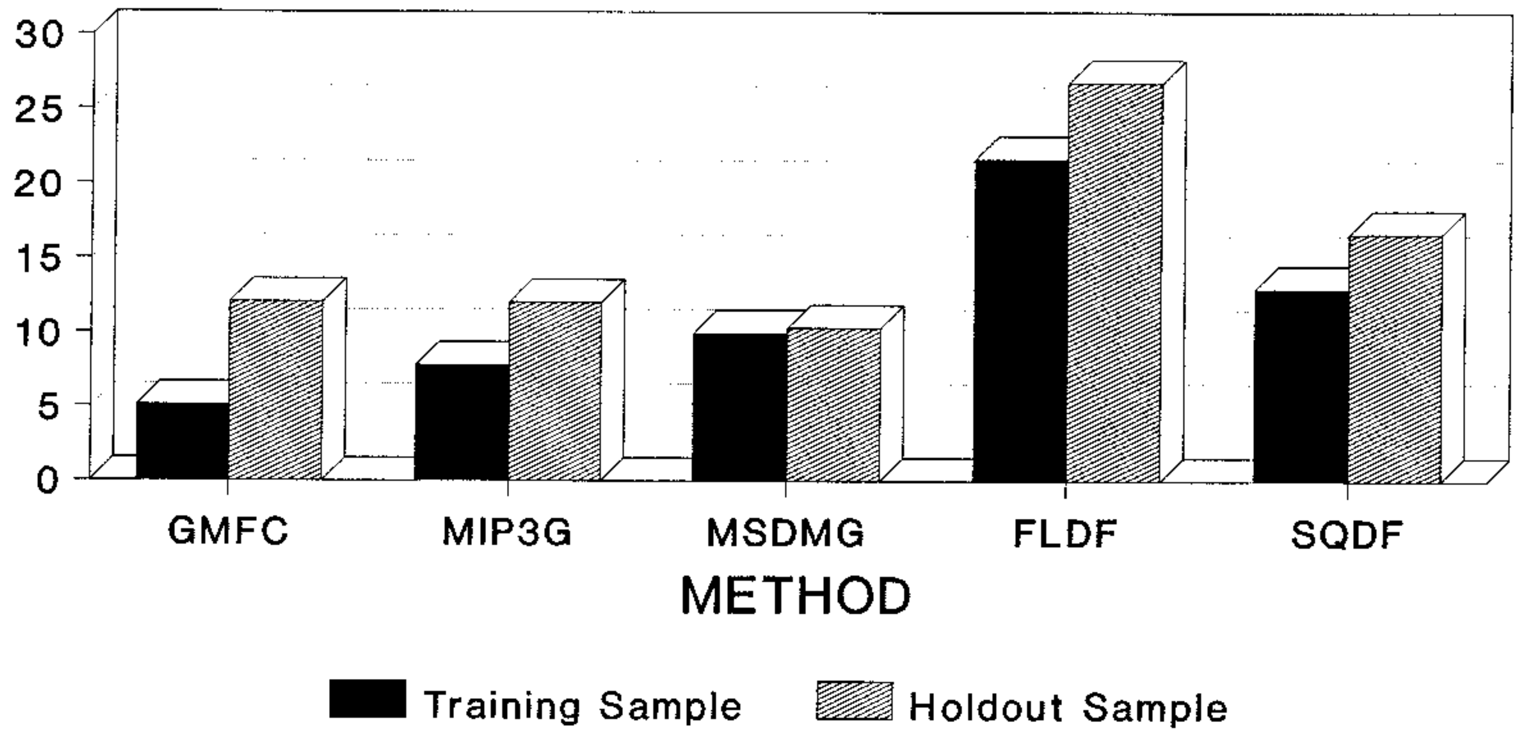

PERCENTAGE MISCLASSIFICATION RATES
Normal Populations with
Equidistant Means & Unequal Variances
45
40
35
30
25
20
15
10
5
0
GMFC
MIP3G
Grid
MSDMG
FLDF
SQDF
METHOD
Training Sample
Holdout Sample
Case 4

Figure 7.4

PERCENTAGE MISCLASSIFICATION RATES

Normal Populations with Outlier Observations (p=.10)



Case 5

Figure 7.5

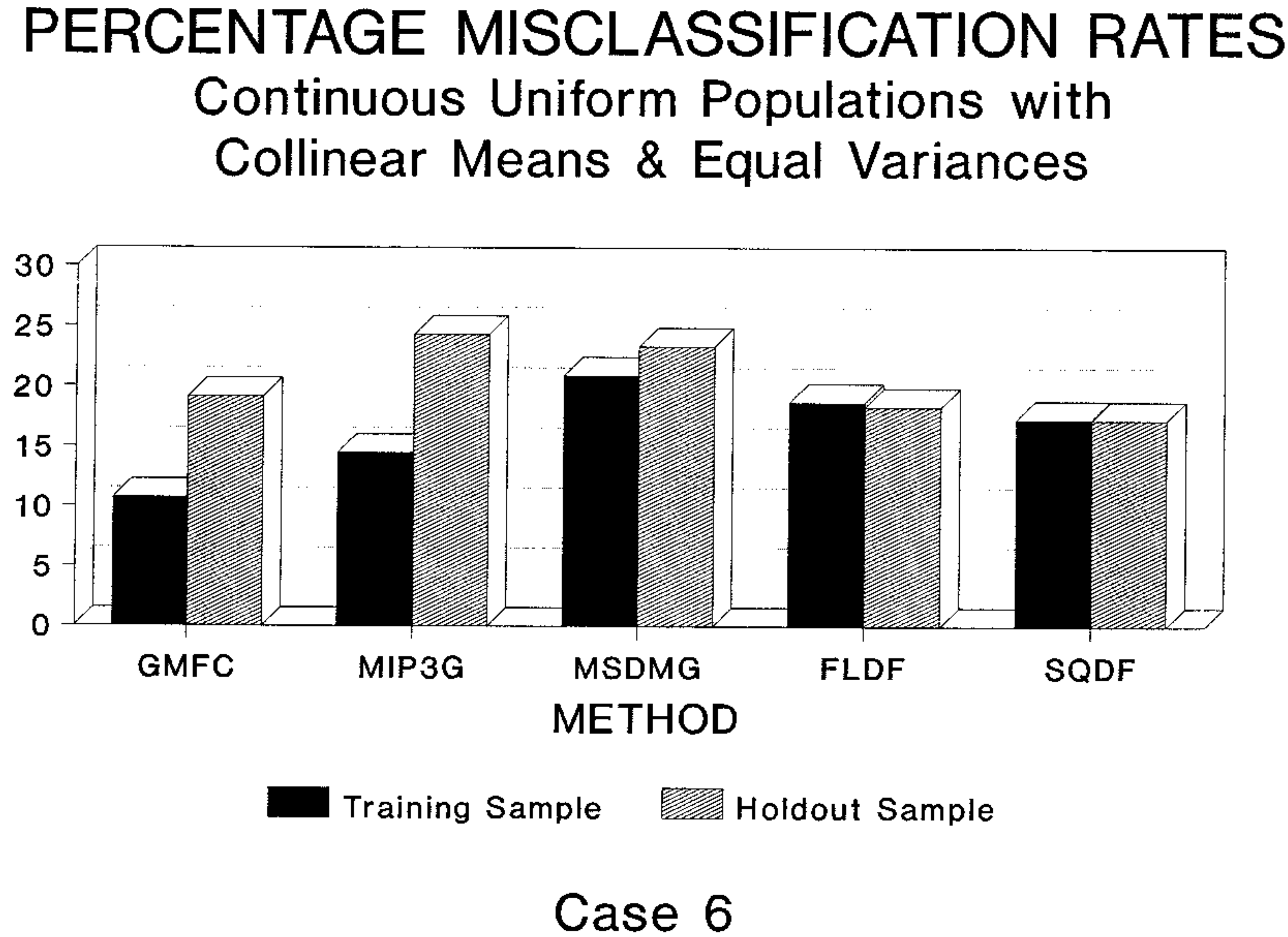
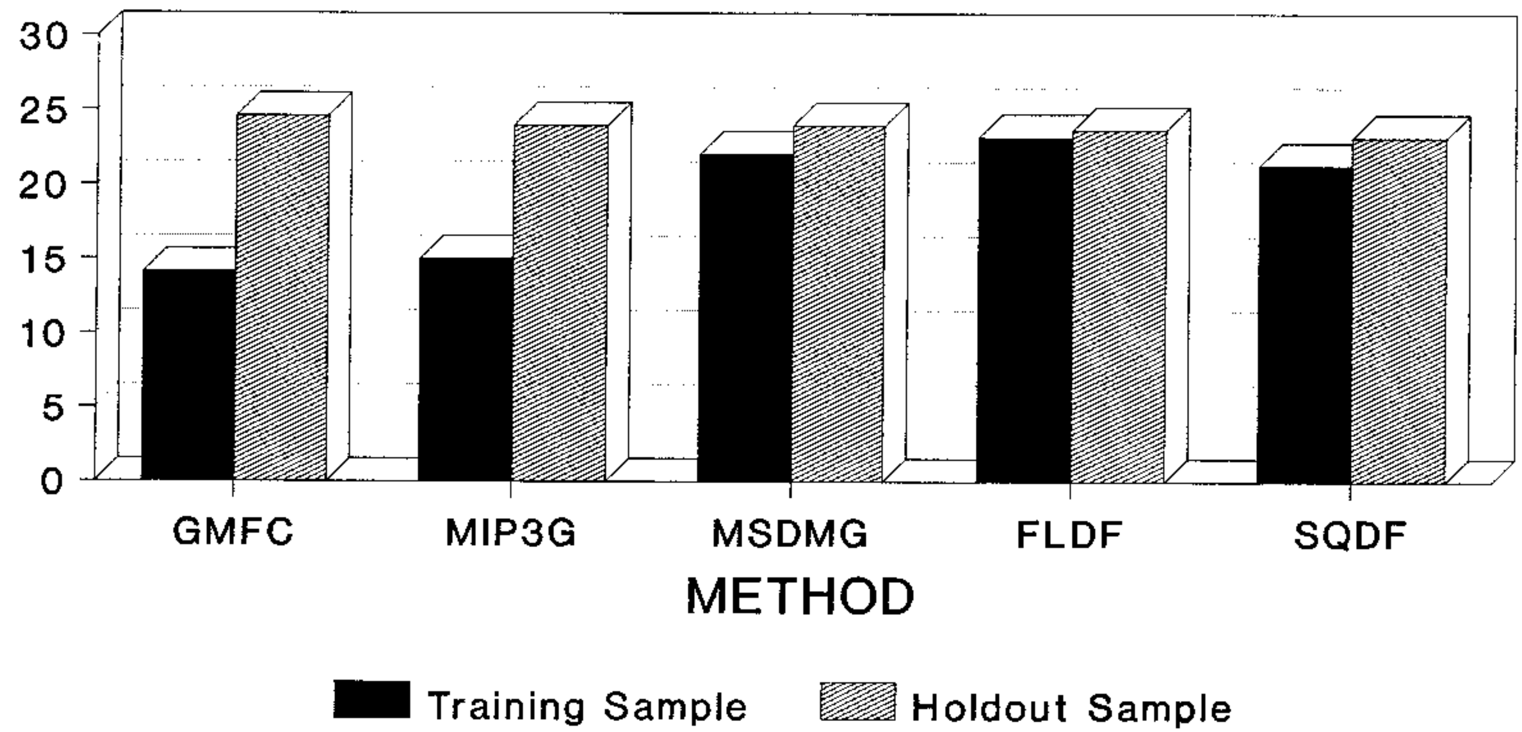

PERCENTAGE MISCLASSIFICATION RATES
Continuous Uniform Populations with
Collinear Means & Equal Variances
30
25
20
15
10
5
0
GMFC
MIP3G
MSDMG
FLDF
SQDF
METHOD
Training Sample
Holdout Sample
Case 6

Figure 7.6

PERCENTAGE MISCLASSIFICATION RATES

Continuous Uniform Populations with Collinear Means & Unequal Variances

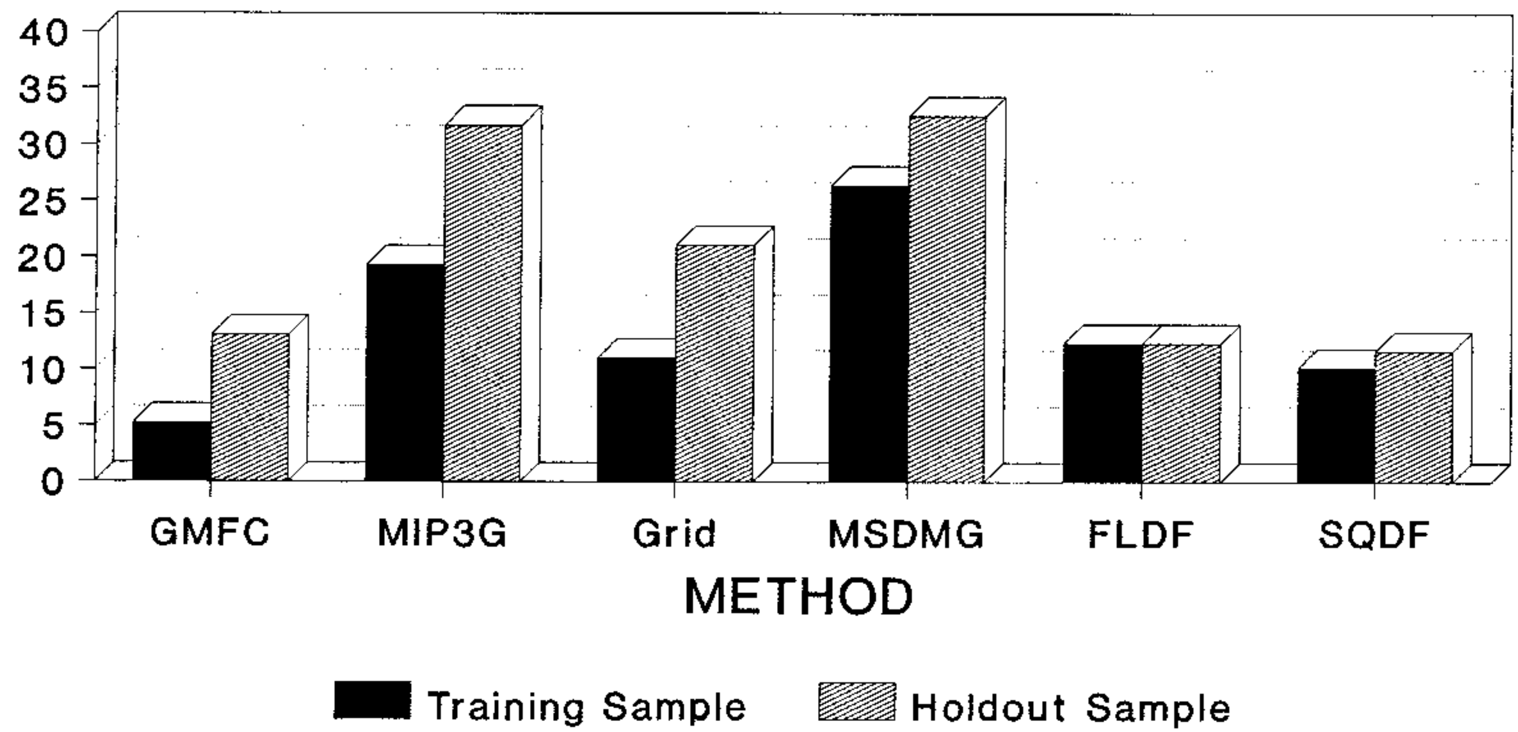


Case 7

Figure 7.7

PERCENTAGE MISCLASSIFICATION RATES

Continuous Uniform Populations with Equidistant Means & Equal Variances

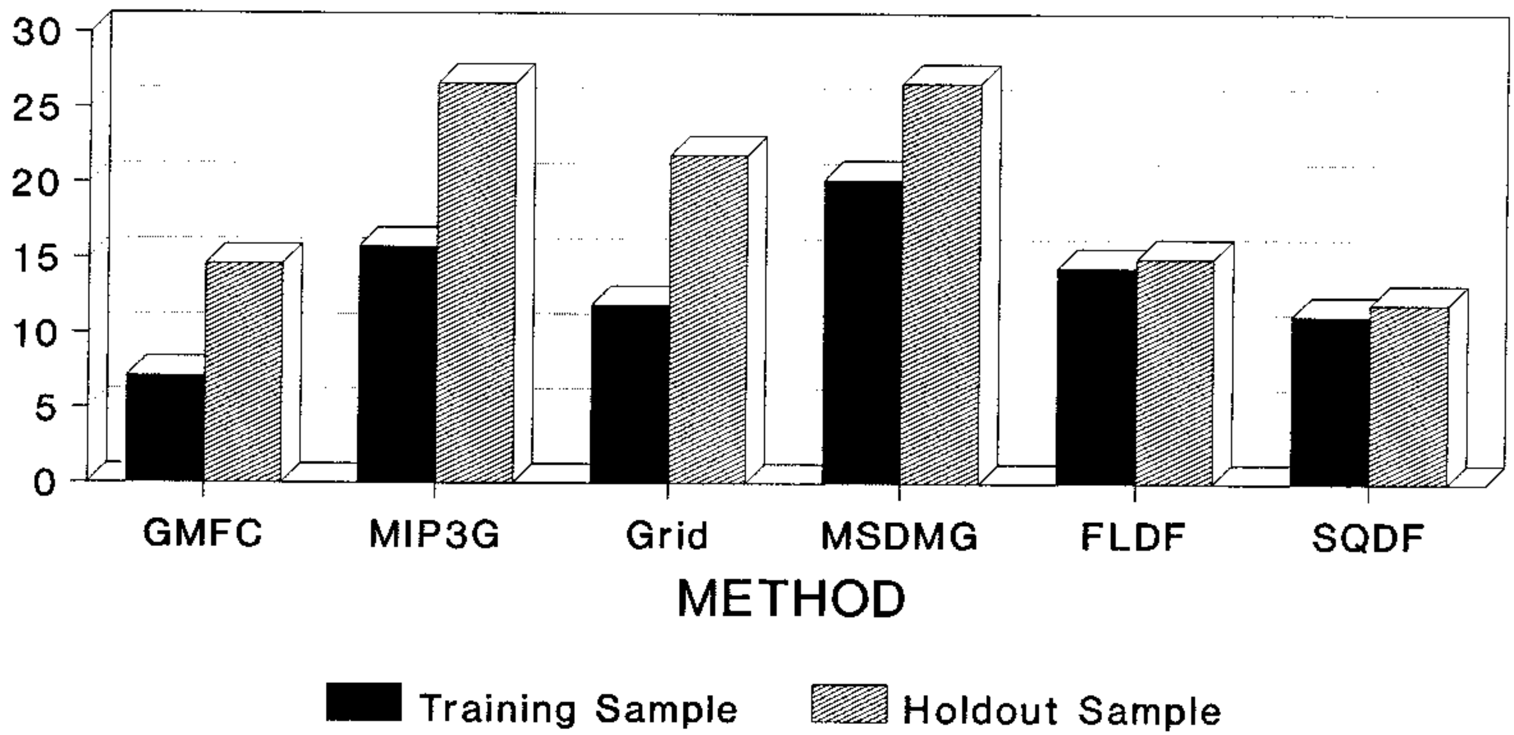


Case 8

Figure 7.8

PERCENTAGE MISCLASSIFICATION RATES

Continuous Uniform Populations with Equidistant Means & Unequal Variances

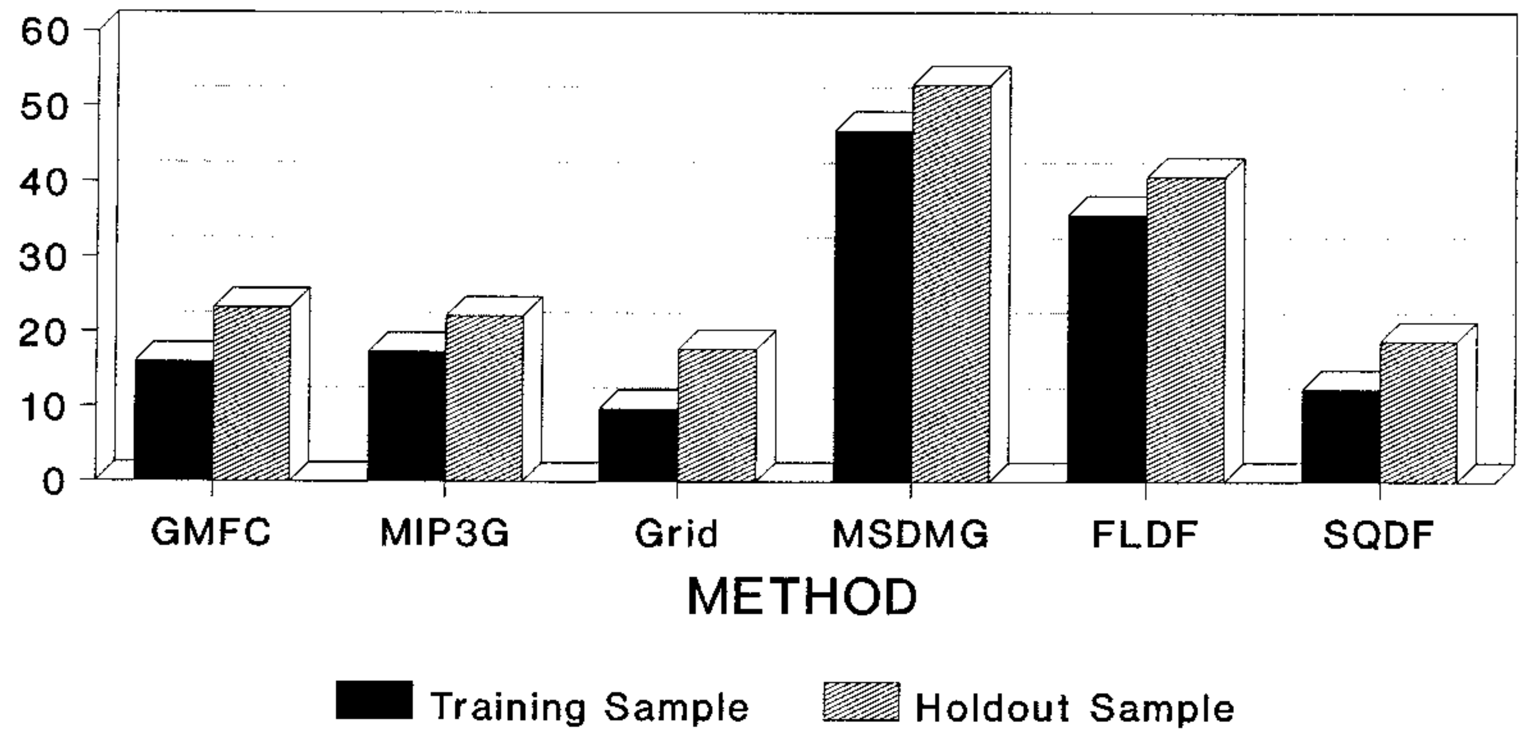


Case 9

Figure 7.9

PERCENTAGE MISCLASSIFICATION RATES

Normal Populations (G_1 & G_2) Partially Surrounded by Combination of Uniforms(G_3)

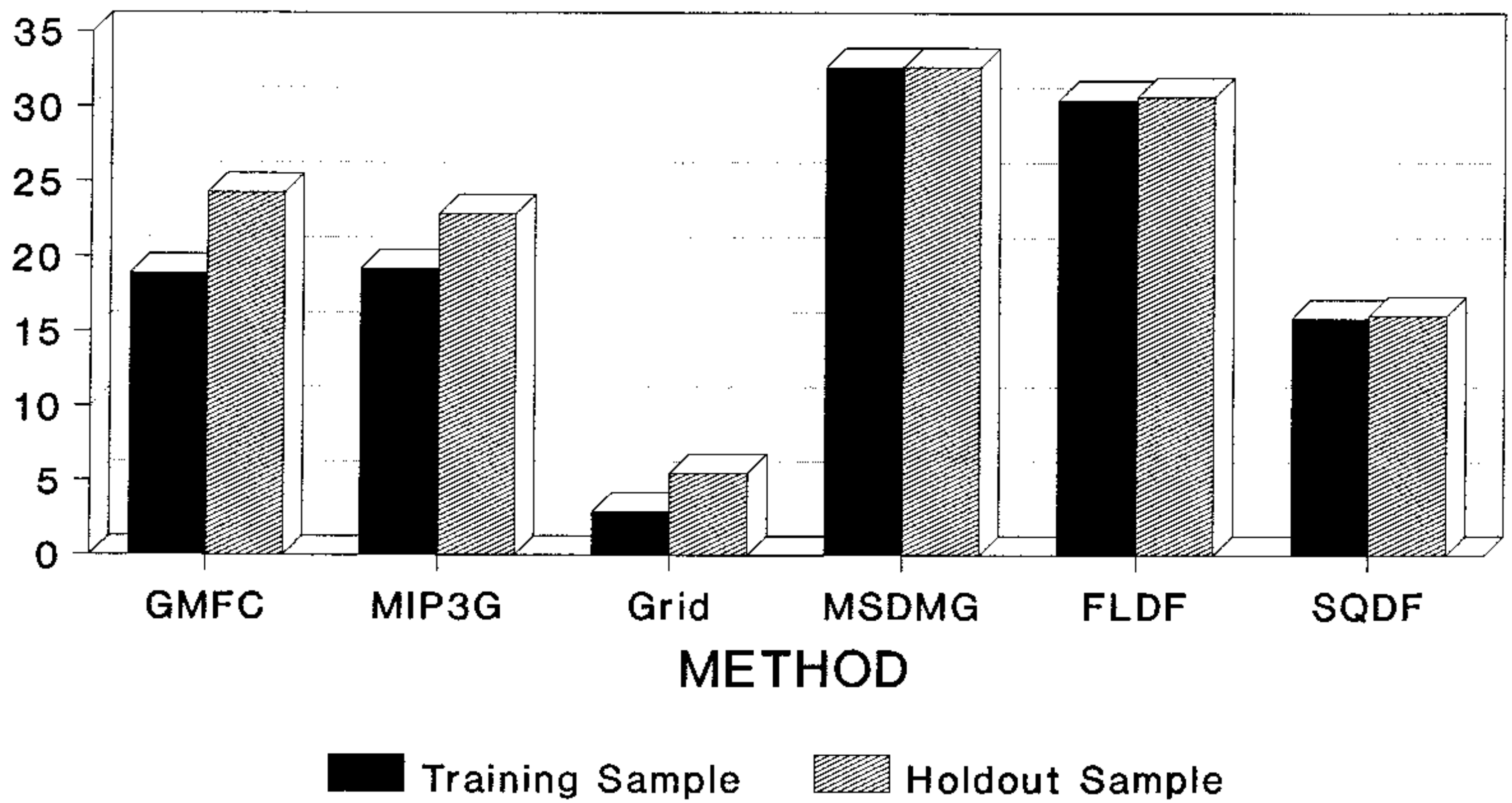


Case 10

Figure 7.10

PERCENTAGE MISCLASSIFICATION RATES

H-Shaped



Case 11

Figure 7.11

REFERENCES

Abad, P. L. and Banks, W. J. (1992). "New LP based heuristics for the classification problem," *European Journal of Operational Research,* 64, 1-13.

Bajgier, S. M. and Hill, A. V. (1982). "An experimental comparison of statistical and linear programming approaches to the discriminant problem," *Decision Sciences,* 13, 604-618.

Banks, W. J. and Abad, P. L. (1991). "An efficient optimal solution algorithm for the classification problem," *Decision Sciences,* 1991, 22, 1008-1023.

Bazaraa, M. and Jarvis, J. J. (1977). *Linear Programming and Network Flows,* New York, N.Y.: Wiley.

Bradley, J. W. (1982). "The insidious L-shaped distribution," *Bulletin of the Psychonomic Society,* 20, 85-88.

Choo, E. U. and Wedley, W. C. (1985) "Optimal criterion weights in repetitive multicriteria decision-making," *Journal of the Operational Research Society,* 36, 983-992.

Erenguc, S. S. (1990). "Survey of mathematical programming models and experimental results for linear discriminant analysis," *Managerial and Decision Economics*, 11, 215-225.

Fisher, R. A. (1936). "The use of multiple measurements in taxomonic problems," *Annals of Eugenics*, 7, 179-188.

Freed, N. and Glover, F. (1981a). "A linear programming approach to the discriminant problem," *Decision Sciences,* 12, 68-74.

Freed, N. and Glover, F. (1981b). "Simple but powerful goal programming formulations for the discriminant problem," *European Journal of Operational Research,* 7, 44-60.

Freed, N. and Glover, F. (1986a). "Evaluating alternative linear programming models to solve the two-group discriminant problem," *Decision Sciences,* 17, 151-162.

Freed, N. and Glover, F. (1986b). "Resolving certain difficulties and improving the classification power of LP discriminant analysis formulations," *Decision Sciences,* 17, 589-595.

Gehrlein, W. V. (1986). "General mathematical programming formulations for the statistical classification problem," *Operations Research Letters,* 5, 299-304.

Glover, F. (1988). "Improved linear programming models for discriminant analysis," *Decision Sciences,* 21, 771-785.

Glover, F., Keene, S., Duea, R. (1988). "A new class of models for the discriminant problem," *Decision Sciences,* 19, 269-280.

Hampel, F. R. (1974). "The influence curve and its role in robust estimation," *Journal of the American Statistical Association,* 69, 383-393.

Joachimsthaler, E. A., and Stam, A. (1988). "Four approaches to the classification problem in discriminant analysis: An experimental study," *Decision Sciences,* 19, 322-333.

Joachimsthaler, E. A., and Stam, A. (1990). "Mathematical programming approaches for the classification problem in two-group discriminant analysis," *Multivariate Behavioral Research,* 25, 427-454.

Johnson, R. A., and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*, (3rd edition), Englewood Cliffs, N. J.,: Prentice Hall.

Jolson, M. A., Weiner, J. L., Rosecky, R. B. (1987). "Correlates of rebate proneness," *Journal of Advertising Research,* February/March, 33-43.

Koehler, G. J. (1989a). "Characterization of unacceptable solutions in LP discriminant analysis," *Decision Sciences,* 20, 239-257.

Koehler, G. J. (1989b). "Unacceptable solutions and the hybrid model," *Decision Sciences,* 20, 844-848.

Koehler, G. J. (1990). "Mathematical programming models in discriminant analysis," *Managerial and Decision Economics,* 11, 227-234.

Koehler, G. J. and Erenguc, S. S. (1990). "Minimizing misclassifications in linear discriminant analysis," *Decision Sciences,* 21, 63-85.

Markland, R. F., and Sweigart, J. R. (1987). *Quantitative Methods: Applications to Managerial Decision Making,* New York, N. Y.: Wiley.

Markowski, C. A. and Markowski, E. P. (1987). "An experimental comparison of several approaches to the discriminant problem with both qualitative and quantitative variables," *European Journal of Operational Research,* 28, 74-78.

Markowski, E. P. and Markowski, C. A. (1985). "Some difficulties and improvements in applying linear programming formulations to the discriminant problem," *Decision Sciences,* 16, 237-247.

Rubin, P. A. (1989). "Evaluating the maximize minimum distance formulation of the linear discriminant problem," *European Journal of Operational Research,* 41, 240-248.

Rubin, P. A. (1990a). "Heuristic solution procedures for a mixed-integer programming model," *Managerial and Decision Economics,* 11, 255-266.

Rubin, P. A. (1990b). "A comparison of linear programming and parametric approaches to the two group discriminant problem," *Decision Sciences,* 21, 373-386.

Schrage, L. (1986). *Linear, Integer and Quadratic Programming with LINDO,* (3rd edition), Palo Alto, CA.: The Scientific Press.

Smith, C. (1947). "Some examples of discrimination," *Annals of Eugenics,* 13, 272-282.

Stam, A. (1989). "Solving the classification problem in discriminant analysis via linear and nonlinear programming methods," *Decision Sciences,* 20, 285-293.

Stam, A. (1990). "Extensions of mathematical programming-based classification rules: A multicriteria approach," *European Journal of Operational Research,* 48, 351-361.

Stam, A., and Joachimsthaler, E. A. (1989). "Solving the classification problem in discriminant analysis via linear and nonlinear programming methods," *Decision Sciences,* 20, 285-293.

Stam, A. and Joachimsthaler, E. A. (1990). "A comparison of a robust mixed integer approach to existing methods for establishing classification rules for the discriminant problem," *European Journal of Operational Research,* 46, 113-122.

Stam, A. and Jones, D. (1990). "Classification performance of mathematical programming techniques in discriminant analysis: Results for small and medium sample sizes," *Managerial and Decision Economics,* 11, 243-253.

Turban, E., and Meredith, J. R. (1988). *Fundamental of Management Science,* (4th edition), Plano, TX,: Science Business Publications.