THE EFFECT OF CERTAIN MODIFICATIONS TO MATHEMATICAL PROGRAMMING MODELS FOR THE TWO-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

Pradit Wanarat, B.S., M.S., M.B.A.

Denton, Texas

May, 1994

THE EFFECT OF CERTAIN MODIFICATIONS TO MATHEMATICAL PROGRAMMING MODELS FOR THE TWO-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

Pradit Wanarat, B.S., M.S., M.B.A.
Denton, Texas

May, 1994

Wanarat, Pradit, The Effect of Certain Modifications to Mathematical Programming Models for the Two-Group Classification Problem. Doctor of Philosophy (Management Science), May, 1994, 193 pp., 58 tables, 23 illustrations, bibliography, 41 titles.

Traditional parametric statistical methods for solving the classification problem are based on certain assumptions. Innovative mathematical programming methods provide alternative approaches to the standard parametric discriminant procedures, when the underlying parametric assumptions are violated. For some data configurations, however, these mathematical programming models fail to provide the optimal classification rule.

This research examines certain modifications of the mathematical programming models to improve their classificatory performance. These modifications involve the inclusion of second-order terms and secondary goals in mathematical programming models. A Monte Carlo simulation study is conducted to investigate the performance of two standard parametric models and various mathematical programming models, including the MSD (minimize sum of deviations) model, the MIP (mixed integer programming) model and the hybrid linear programming model. Misclassification rates for the classification models are empirically

estimated on both training samples and validation (holdout) samples. Exact misclassification rates are determined from the estimated classification functions for some models. Several factors, such as sample size, covariance structure, distribution, and orientation of the data, are varied in the simulation study.

The results show that the modified mathematical programming models have potential for being very useful in situations in which violations of the usual parametric assumptions are severe. This study addresses certain issues in implementing mathematical programming approaches to the classification problem. For example, with some mathematical programming models, there are solutions that are not invariant under data translations or rotations. The study shows the usefulness of a general contaminated multivariate normal distribution in estimating misclassification probabilities. The study also illustrates that a wide range of values can be assigned to the measures of skewness and kurtosis when generating the contaminated normal distribution by using different parameter settings. The results of this study will assist practitioners in understanding and implementing improved versions of mathematical programming formulations and, thus, give them greater flexibility in choosing an appropriate classification model.

TABLE OF CONTENTS

LIST OF TABLES

Page

Page

Page

Page

LIST OF ILLUSTRATIONS

Page

CHAPTER I

INTRODUCTION

Overview of the Statistical Classification Problem

The statistical classification problem is a well-known problem in many areas of business applications, for example, as in differentiating between prospective buyers and nonbuyers, between successful employees and unsuccessful ones, or between promising new firms and those likely to fail. The intent of classification is to properly categorize or classify subjects or observations into two or more groups based on certain attributes or characteristics of the subjects to be classified.

Discriminant analysis is a statistical technique that uses the information available from a set of data to develop a rule or method for predicting to which group a new observation is most likely to belong based on the observed values of the observation's attribute variables. Discriminant analysis provides a powerful technique for examining differences between two or more groups of observations with regard to several attribute variables. For example, a credit manager may wish to classify previous holders of bank loans into two groups--payers or defaulters. For this situation, the credit manager may use several characteristics of the loan holders for attribute variables in the analysis. Some characteristics of interest might be size of the loan, income, liability, marital status, and credit history of the loan holder. All of these characteristics are measured at the time of the loan application. The analysis begins by finding a discriminant function which uses the measured values of the characteristics as input. This discriminant function will be used to identify potential payers or defaulters in the future. That is, the credit manager would measure these characteristics on future loan applicants and, by use of the discriminant function, identify applicants as either probable payers or defaulters.

The most commonly used methods for the classification problem are the parametric statistical methods. These traditional parametric statistical methods are based on certain assumptions, and these methods may not yield the optimal classification rule if the underlying assumptions are violated. Over the past thirteen years, the literature has increasingly recognized that a variety of standard statistical problems, such as discriminant analysis, can be examined and analyzed advantageously by using computer-intensive techniques from the field of optimization. Innovative mathematical programming methods provide alternative approaches to the standard parametric methods for the classification problem.

Some of the mathematical programming models have been found to compare favorably with the parametric models. For some data configurations, however, these mathematical programming models fail to provide the optimal classification rule. Furthermore, some of the mathematical programming models involve a large amount of computational effort, and there have been only limited simulation studies evaluating their classificatory performance.

This research examines certain modifications of the mathematical programming models in order to improve their classificatory performance. These modifications involve the inclusion of second-order terms in linear programming (LP) models and mixed integer programming (MIP) models, and the inclusion of secondary goals in MIP models. This study addresses certain issues in implementing mathematical programming approaches to the classification problem. For example, with some mathematical programming methods, there are solutions that are not invariant under data translations or rotations.

A Monte Carlo simulation study is performed to assess the performance of classification models. Two standard parametric models and various mathematical programming models are employed in this research study. Misclassification rates for various discriminant models are empirically estimated on both training samples and validation (holdout) samples. Also, exact misclassification rates are determined from the estimated classification functions for some models and from data configurations involving the contaminated normal distributions. Several factors, such as sample size, covariance structure, distribution, and orientation of the data, are varied in the simulation study. This study will assist decision-makers in understanding and implementing improved versions of mathematical programming formulations and, thus, give them greater flexibility in choosing an appropriate classification model.

An Application Comparing Different Classification Methods

An example illustrating the potential of the mathematical programming approaches to discriminant analysis is explained using a data set in Johnson and Wichern (1992). These authors presented an example using this data set to illustrate the standard discriminant analysis procedures to classify two groups of families in a city. In the example, a riding-mower manufacturer is interested in classifying families into one of two groups--G1: riding-mower owners, and G2: those without riding mowers (that is, nonowners). The classification is based on two attribute variables, x_1 = incomes and x_2 = lot size. Random samples of n_1 = 12 current owners and n_2 = 12 current nonowners yield the values in Table 1.

Table 1.--Data Set for Owners and Nonowners of Riding Mowers

G1: Riding-mower owners		G2: Nonowners	
x_1 (income in \$1000s)	x_2 (lot size in 1000 ft^2)	x_1 (income in \$1000s)	x_2 (lot size in 1000 ft^2)
64.8	21.6	52.8	20.8
61.5	20.8	64.8	17.2
60.0	18.4	43.2	20.4
87.0	23.6	84.0	17.6
101.1	19.2	49.2	17.6
108.0	17.6	59.4	16.0
82.8	22.4	66.0	18.4
85.5	16.8	47.4	16.4
69.0	20.0	33.0	18.8
93.0	20.8	75.0	19.6
51.0	22.0	51.0	14.0
81.0	20.0	63.0	14.8

Source: Johnson and Wichern, 1992, page 496.

Six classification models are used to analyze this data set. Fisher's linear discriminant function (LDF) and Smith's quadratic discriminant function (QDF) are used to represent the parametric statistical method. For the mathematical programming method, the minimize sum of deviations (MSD) model and the mixed integer programming (MIP) model are used in this example. MSD and MIP models are both linear classification models consisting of only first-order terms of the two attribute variables. Two second-order mathematical programming models, consisting of all first-order and second-order terms (5 variables), are also used to classify the data in the example. These second-order models are denoted by MSD5 and MIP5.

Table 2 shows results of the six classification models. If the LDF method is used to classify the data in this example, then 3 out of 24 observations will be misclassified. Specifically, one riding-mower owner will be classified as nonowner and two nonowners will be classified as riding-mower owners. If the QDF method is used, the same results will be obtained. That is, 3 out of the 24 observations will be classified incorrectly. For the mathematical programming methods, if the first-order MSD model is used, then 5 out of the 24 observations will be misclassified. Specifically, two riding-mower owners will be classified as nonowners and three nonowners will be classified as riding-mower owners. However, if the second-order MSD model is used, then the same results as the LDF and QDF methods will be obtained. If the first-order MIP model is used, then only two of the riding-mower owners will be misclassified as nonowners. However, if the second-order MIP model is used, then only 1 out of the 24 observations will be classified incorrectly. Specifically, only one riding-mower owner will be classified as nonowner but none of the nonowners will be misclassified.

It is interesting to note that the three misclassified observations by the second-order MSD method are also misclassified by both parametric methods, and that the only one misclassified observation by the second-order MIP method is also misclassified by all other methods. Clearly, from

Table 2.--Classification Results for the Data Set of Owners and Nonowners of Riding Mowers

Observations and Actual Group		Group into Which Models Classified Observations					
x_1	x_2	LDF	QDF	MSD	MSD5	MIP	MIP5
G1: Riding-mower owners							
64.8	21.6	1	1	1	1	1	1
61.5	20.8	1	1	1	1	1	1
60.0	18.4	(2)	(2)	(2)	(2)	(2)	(2)
87.0	23.6	1	1	1	1	1	1
101.1	19.2	1	1	1	1	1	1
108.0	17.6	1	1	1	1	1	1
82.8	22.4	1	1	1	1	1	1
85.5	16.8	1	1	(2)	1	(2)	1
69.0	20.0	1	1	1	1	1	1
93.0	20.8	1	1	1	1	1	1
51.0	22.0	1	1	1	1	1	1
81.0	20.0	1	1	1	1	1	1
G2: Nonowners							
52.8	20.8	2	2	(1)	2	2	2
64.8	17.2	2	2	2	2	2	2
43.2	20.4	2	2	2	2	2	2
84.0	17.6	(1)	(1)	(1)	(1)	2	2
49.2	17.6	2	2	2	2	2	2
59.4	16.0	2	2	2	2	2	2
66.0	18.4	2	2	2	2	2	2
47.4	16.4	2	2	2	2	2	2
75.0	19.6	(1)	(1)	(1)	(1)	2	2
33.0	18.8	2	2	2	2	2	2
51.0	14.0	2	2	2	2	2	2
63.0	14.8	2	2	2	2	2	2

Note: The misclassified observations are shown in parenthesis.

the results of this example, an appropriate mathematical programming method has the potential to effectively classify observations from certain data sets and, therefore, should be investigated.

Purpose, Problem, and Significance

Purpose of the Research Study

The purpose of this research is to analyze the performance of certain mathematical programming models for solving the statistical classification problem under certain modifications of these models. The research in this study investigates the appropriateness of the inclusion of second-order terms in LP models and in MIP models. The study also analyzes the effects of some existing and proposed secondary goals in MIP models on the classificatory performance of these models. The appropriateness of using contaminated normal data in simulation studies to generate different types of nonnormal data is also examined.

Problem Motivating the Research Study

The problem motivating this study is the lack of performance results for mathematical programming models proposed over the past decade to solve the discriminant problem. Although several Monte Carlo simulation studies have investigated the advantages and disadvantages of using

LP-based and MIP-based models, these simulation studies have not thoroughly explored certain modifications to these mathematical programming approaches for solving the discriminant problem. These simulation studies typically have not used higher-order terms in the classification models. One of the problems associated with MIP models is the possibility of numerous alternate optimal solutions. While these alternate solutions are all optimal on the training set of observations, they may each have different performance results on a validation set of observations. Some researchers have studied mathematical programming models with secondary goals, but they have not addressed the importance of the secondary goal.

Most of the simulation studies investigating mathematical programming approaches to the classification problem fail to use contaminated normal data, although normal and other nonnormal distributions are explored. Several simulation studies generate nonnormal data, using a simulation method in which the mean, variance, skewness, and kurtosis are specified, but the actual shape of the distribution of the data is not known. The range of values for the skewness and kurtosis of the contaminated normal distribution is not readily available for researchers desiring to use contaminated normal data in Monte Carlo simulation studies.

Significance of the Research Study

Advances in computer technology have spurred research in computer-intensive techniques such as solving statistical problems with mathematical programming models. The results of this research would allow practitioners to understand and implement improved versions of mathematical programming formulations for the discriminant problem by utilizing higher-order terms and appropriate secondary goals in certain mathematical programming models. These formulations have the potential for being very useful in situations in which violations of the usual parametric assumptions of discriminant analysis are severe.

Previous research studies of mathematical programming models primarily have investigated linear discriminant functions that included only first-order terms. It is easy to find data for which these first-order mathematical programming models fail to yield the optimal classification rule. Mathematical programming models that use all first-order terms and second-order terms of the attribute variables include all of the terms that are present in Smith's quadratic discriminant function. Thus, these mathematical programming formulations with first-order and second-order terms have the potential for being competitive with the quadratic method in problems requiring a classification function that is nonlinear in the attribute variables. The use of second-order terms in mathematical programming models would allow for greater flexibility in choosing an appropriate discriminant procedure.

The usefulness of various secondary goals proposed in the literature has not been adequately addressed. An appropriately selected secondary goal has the potential of improving the classificatory performance of the mathematical programming model on the validation samples. Understanding the types of configurations that may warrant the use of a certain secondary goal is important in utilizing the appropriate mathematical programming procedure.

The normal distribution contaminated with outliers is mentioned in the literature as being an important distribution to describe certain real-world data sets. Understanding the range of possible values for the skewness and kurtosis for these distributions will assist researchers in generating certain types of nonnormal distributions. An important advantage of using the contaminated normal distribution in a Monte Carlo simulation study investigating the performance of linear discriminant functions is that the exact misclassification rate of the estimated linear discriminant function can be found analytically, and, hence, the need for large validation samples is eliminated.

Organization of the Dissertation

This dissertation is organized into six chapters. Chapter 1 provides introductory material explaining an overview of the statistical problem, including a numerical example comparing different classification methods. Chapter 1 also contains the purpose, problem, and significance of the study. Chapter 2 provides a literature review of mathematical programming approaches for discriminant analysis, and it also includes research questions of the study. Chapter 3 contains the theoretical framework of the two-group classification problem and the proposed classification models used in the study. Chapter 4 provides experimental designs for the Monte Carlo simulations used in this study, including the selection of models and parameter settings of data configurations. Chapter 5 presents experimental results obtained from the simulations. Chapter 6 provides a summary of the findings, key assumptions, future directions, and major contributions of the research.

CHAPTER II

LITERATURE REVIEW

Overview of Previous Research

The classification problem in discriminant analysis is concerned with correctly classifying observations into well-defined groups or classes when group membership of these observations is either known or unknown (Huberty 1984). Applications of discriminant analysis extend to both business and scientific disciplines, including psychology (Huberty, Wisenbaker, and Smith 1987); economics (Sudarsanam and Toffler 1985); accounting (Welker 1974); and finance (Srinivason and Kim 1987).

Existing parametric statistical methods for solving the classification problem include Fisher's (1936) linear discriminant function (LDF) and Smith's (1947) quadratic discriminant function (QDF). Optimality for the LDF and QDF methods is based on the assumption that the attribute variables for each group follow a multivariate normal distribution, with equal and unequal variance-covariance structure across groups, respectively (Johnson and Wichern 1992). Alternative approaches for solving the classification problem have been researched in order to develop promising models that are robust to violations of these assumptions (Freed and Glover 1986).

Linear Programming Approaches

Linear programming approaches for solving the statistical classification problem have been given considerable attention since the introduction of LP-based models for the discriminant problem by Freed and Glover (1981) and Hand (1981). In many research studies involving LP models for discriminant analysis, the objective is to find a discriminant rule that is either optimal or competitive with the parametric approaches in correctly classifying observations from a set of new observations or from a representative validation sample (Glover, Keene, and Duea 1988). These new approaches are relatively easy for practitioners to implement.

In recent years, theoretical and empirical investigations of innovative discriminant analysis procedures have been an attempt to improve upon the classificatory performance of alternative discriminant procedures as opposed to the standard statistical discriminant procedures. Some studies have focused on the undesirable problems associated with mathematical programming models. Koehler (1989a and 1989b), Markowski and Markowski (1987), Rubin (1989 and 1991), and Glover (1990) have investigated problems that plagued certain mathematical programming models. These problems included formulations that obtained unbounded solutions, trivial solutions, and solutions that were not invariant under data translation or rotation. These problems have inspired numerous variations of mathematical programming formulations. Normalization constraints, such as those discussed in Glover, Keene, and Duea (1988) and Glover (1990), were introduced to overcome the undesirable problems associated with early mathematical programming formulations.

Mixed Integer Programming Approaches

Each LP-based model obtains a classification rule by optimizing an objective function that is a surrogate for minimizing the number of misclassifications. To directly minimize the number of misclassifications in the training sample, MIP models have been proposed. However, it does not always follow that these models will perform optimally on the validation sample. Because of the computationally intensive nature of these models, several researchers have proposed heuristic algorithms to make the MIP approach more computationally efficient. Koehler and Erenguc (1990), Banks and Abad (1991), and Rubin (1990a) have investigated heuristic algorithms that appear to yield good, albeit suboptimal, solutions to the MIP models. Loucopoulos (1993) investigated the performance of MIP models specifically designed for the multiple groups case. These MIP models tend to be particularly computer intensive.

Classificatory Performance of Models

Several recent studies have compared the classificatory performance of LP-based discriminant procedures with the performance of the standard statistical procedures. Experimental studies have been conducted by Mahmood and Lawrence (1987), Joachimsthaler and Stam (1988), Markowski and Markowski (1987), Freed and Glover (1986), and Rubin (1990b). The MSD (minimize sum of deviations) model of Freed and Glover (1986) was found to compare favorably with the existing discriminant procedures. However, Joachimsthaler and Stam (1988) concluded that relative differences in performance by linear programming formulations and standard statistical procedures are small, even under multivariate nonnormal conditions. An early study by Markowski and Markowski (1985) focused on limitations of the LP procedures. Studies such as Glover (1990) and Glover, Keene, and Duea (1988) later appeared to overcome these special limitations.

Rubin (1990b) found that Smith's quadratic procedure was superior to the fifteen linear programming models tested in his study when the data follow a multivariate normal distribution, with various parameter values for the means, variances, and correlations. This result is not totally surprising since the quadratic method allows for a nonlinear classification function. Silva and Stam (1994) conducted a simulation study using second-order terms in the hybrid and MSD models. They considered a large training samples of size 100 from exponentially distributed random variables. However, they did not consider the MIP model in their study. For the highly nonnormal data generated in their study, the hybrid model and the MSD model greatly benefitted from the second-order terms.

Several studies proposed the use of secondary goals in mathematical programming models (Freed and Glover 1981, Bajgier and Hill 1982, Glover 1990, and Rubin 1990a). Bajgier and Hill (1982) used an LP-based model, with the first goal of minimizing the deviations of the misclassified observations and the secondary goal of maximizing the deviations of the correctly classified observations from the cutoff value in the discriminant rule. This model is known as the OSD (optimize sum of distances) model. Bajgier and Hill (1982) also presented in their studies an MIP model with secondary goals. The first goal of their MIP model is to minimize the number of misclassified observations, while the secondary goals are to minimize the deviations of the misclassified observations and to maximize the deviations of the correctly classified observations. Rubin (1990a) used a secondary goal that maximized the minimum interior distance of the correctly classified observations and found promising results for the MIP model with this secondary goal in a limited simulation study.

Contaminated Normal Data

Several studies, such as Nath (1984), Hampel (1974), and Lee and Ord (1990), have considered the contaminated normal distribution to be useful in simulation studies. Nath (1984) pointed out that the contaminated normal distribution is of particular importance to researchers who wish to determine analytically the exact misclassification rate of a linear discriminant function for future observations. Thus, from the linear discriminant function estimated by using a training sample, an exact misclassification rate can be calculated without using large validation samples.

The contaminated normal distribution is widely accepted as realistic because a small proportion of outlying observations occurs even in good data sets. Especially in business-related problems, outlier-contaminated data are not uncommon (Mahmood and Lawrence 1987). Although the contaminated normal distribution is generally accepted as being an important nonnormal distribution, it has been used very little in Monte Carlo simulation studies that have investigated misclassification rates of mathematical programming approaches for solving the two-group discriminant problem.

In some published simulation studies, such as Freed and Glover (1986) and Rubin (1990b), only normally distributed data were used. Restricting the simulated data to normally distributed data eases the interpretation of the results as well as limits the complexity involved in generating multivariate data. Other studies, such as Joachimsthaler and Stam (1988), used a technique for generating nonnormal data with specified values for skewness and kurtosis. This technique for generating nonnormal data was presented by Vale and Maurelli (1983). However, there is no easy way to describe the generated data or the cumulative distribution function of the population. With contaminated normal data, the distribution of the data can be easily described.

Research Questions

Motivation for Research Question on Second-Order Term

The appropriateness of adding higher-order terms to mathematical programming models has not been thoroughly addressed (Silva and Stam 1994). In multiple regression analysis, it is well known that the independent variables used in a linear regression function may be first-order terms or higher-order terms (Draper and Smith 1981). The same approach may be used in discriminant analysis in which squared terms, crossproduct (interaction) terms, or higher-order terms are included to improve the classificatory performance of the models (Johnson and Wichern 1992).

Freed and Glover (1986) regarded Fisher's LDF procedure as an important benchmark of performance and showed that the MSD method with first-order terms performed competitively with the Fisher's LDF and the logistic models in a simulation study. In a more extensive simulation study in which the QDF procedure was included, Rubin (1990b) found that Smith's QDF procedure was superior to the fifteen LP-based models tested in his study when the data followed a multivariate normal distribution with various parameter values for the means, variances, and correlations. This result is not totally surprising since the quadratic method allows for second-order terms in the model, whereas the LP-based models include only first-order terms. Rubin (1990b, page 382) stated that "it is incumbent on researchers to include QDF as a benchmark when seeking situations in which the linear programming approaches would be advantageous." Rubin (1990b) also showed that the MSD method performed competitively with Fisher's LDF procedure and appeared to be one of the more promising LP-based models.

The procedure for implementing a mathematical programming model with all first-order and second-order terms present is similar to including second-order terms in a linear regression model. For example, let $Y_i = (a_{i1}, a_{i2}, \ldots, a_{ip})^T$ be the i^{th} observation with p attribute values. A first-order model for any of the LP-based procedures would simply use $\sum_{j=1}^{p} a_{ij}x_j$ as the discriminant score, with the weights x_j determined by the linear programming approach. A complete second-order model with all first-order terms would use the following discriminant score in the model, with the x_{jj}, x_j, and x_{hk} weights to be determined by the linear programming approach:

$$\sum_{j=1}^{p} a_{ij}^2 x_{jj} + \sum_{j=1}^{p} a_{ij} x_j + \sum\sum_{h>k} a_{ih} a_{ik} x_{hk}$$

It is important to note that the above discriminant score is linear in terms of the parameters (weights) to be estimated, although it is a second-order polynomial in terms of the attribute values. Silva and Stam (1994) used a second-order discriminant score in a simulation study that involved the LDF, QDF, hybrid, and MSD methods. However, their simulation study was restricted to exponentially distributed attribute values and training samples of size 100. Also, Silva and Stam (1994) found that including the crossproduct terms in the model appeared to improve the classificatory performance when correlation between attributes was present. However, it is not appropriate to extend this conclusion to situations in which other data configurations are used. Establishing conditions for translational and rotational invariance of LP-based model has been important in selecting desirable models (Freed and Glover 1986, Koehler and Erenguc 1990, Markowski and Markowski 1985). Silva and Stam (1994) did not address this issue in evaluating models with second-order terms. From the literature, it is clear that further research needs to address the following research question.

Research Question 1

> How do second-order terms in mathematical programming models affect the performance of certain two-group classification models for small to moderate training sample sizes and for normal and nonnormal data? Can the correlation structure of the data determine whether the crossproduct terms should be included in the models? Under what conditions are these models invariant with respect to translation and rotation of the data?

Motivation for Research Question on Secondary Goal

The hybrid model (Glover 1990) has several desirable goals. These goals require properly selected weights to prioritize the goals in the objective function of the formulation for the hybrid model. Silva and Stam (1994) found that the hybrid model performed competitively with the LDF and the QDF procedures when second-order terms were included in the model. Bajgier and Hill (1982) presented an MIP model with the goals of minimizing the deviations of the misclassified observations and maximizing the deviations of the correctly classified observations from the cutoff value in the discriminant rule. Since the MIP model is computationally intensive, particularly for large sample sizes, few simulation studies have included the model. Some extensive simulation studies, such as Rubin (1990b) and Joachimsthaler and Stam (1988), have excluded the MIP model because of the computational effort.

In recent simulation studies by Koehler and Erenguc (1990) and Stam and Jones (1990), the MIP model typically did not perform much better than the QDF or the LDF models on validation samples for configurations with normal and uniform distributions. Since the MIP model may have many alternative solutions that are optimal on the training samples (Bajgier and Hill 1982), it is possible that an appropriate secondary goal may improve the classificatory performance of the MIP model on the validation samples. The secondary goal would considerably limit the number of alternative solutions. Rubin (1990a) also used a secondary goal in his study. His secondary goal maximized the deviation between the cutoff value and the discriminant score of the closest correctly classified observation to the cutoff value. Another way to state this is to say that the secondary goal maximizes the minimum interior distance of correctly classified scores (Rubin 1990a). This very limited simulation study, which used only the normally distributed data, showed promising results for the MIP model with this goal.

One secondary goal that has not been investigated with MIP models is the goal of maximizing the separation between the discriminant scores of the centroid (mean vector of the attribute values) of each group. The theoretical motivation for using this secondary goal is the fact that Fisher's LDF method can be derived by maximizing the absolute difference $|w^T(\bar{a}_1 - \bar{a}_2)|$, where $\bar{a}_1$ and $\bar{a}_2$ are the mean vectors of the attribute values for group 1 and group 2, respectively, subject to the constraint $w^TSw = 1$, where S is the estimated variance-covariance matrix of the two populations (Morrison 1976). Now, $w^TSw = 1$ is nonlinear in the weights w_i of the w vector and thus cannot be used in the standard MIP formulation, which includes only linear constraints in the parameters that need to be estimated. One way to remedy this situation is to use a constraint on the range of the discriminant scores or a constraint on the range of values for the weights. These constraints would be surrogates for the constraint $w^TSw = 1$. The second research question addresses the issue of the importance of certain secondary goals in MIP models and is stated below.

Research Question 2

> Can the use of certain secondary goals improve the performance of MIP models for the two-group classification problem on small to moderate sample sizes?

Motivation for Research Question on Contaminated Normal Distribution

Several Monte Carlo simulation studies use nonnormal distributions to evaluate the robustness of various statistical procedures. Some studies have used distributions such as uniform, double exponential, lognormal, and discrete uniform to generate nonnormal data (Stam and Jones 1990; Nath, Jackson, and Jones 1992). These distributions are often the standard types of distributions used in simulation studies to represent distributions with various skewness and kurtosis values. However, not all real data correspond to the skewness and kurtosis values of these distributions. Fleishman (1978) generated nonnormal data by using a polynomial transformation and constructed a table of values for the skewness and kurtosis. This table could be used to select various skewness and kurtosis values for generating nonnormal data with a polynomial transformation. Vale and Maurelli (1983) observed that the shape of the generated data by Fleishman's method was difficult to understand and that both the exact probability density function and the cumulative distribution function were unknown.

Contaminated normal data is viewed as an important distribution in representing real-world data (Nath 1984, Hampel 1974, and Lee and Ord 1990). However, only Lee and Ord (1990) used the contaminated normal to evaluate LP-based models in a simulation study. Perhaps one reason that the contaminated normal distribution is not widely used in simulation studies evaluating LP-base models is that the range of possible values for the skewness and kurtosis is not readily available. Joachimsthaler and Stam (1988) used Fleishman's method and selected various values for the skewness and kurtosis from the table to generate nonnormal data.

One important motivation for considering the contaminated normal distribution as a nonnormal distribution in simulation studies with linear discriminant functions is that an exact misclassification rate can be calculated from an estimated linear discriminant function under the assumption of this distribution. Therefore, under this distribution, the need for validation samples can be eliminated when linear discriminant functions are being evaluated.

The following research question is important to researchers desiring to conduct a simulation study with nonnormal data, particularly if exact misclassification rates from estimated linear discriminant functions are desired.

Research Question 3

Since the contaminated normal distribution (mixture of two normals) can be used to assess the performance of linear discriminant functions without a validation sample, how appropriate is this distribution for a simulation study in generating nonnormal data with a variety of values for the skewness and kurtosis measures? In particular, what range of values for the measures of skewness and kurtosis can the contaminated normal distributions have by using different parameter settings for the mean, standard deviation, and contaminating fraction?

Summary of Research Questions

This research study investigates the effect of certain modifications of mathematical programming models for solving the statistical classification problem. A summary of the research questions to be answered in this research study is presented below.

Research Question 1

How do second-order terms in mathematical programming models affect the performance of certain two-group classification models for small to moderate training sample sizes and for normal and nonnormal data? Can the correlation structure of the data determine whether the crossproduct terms should be included in the models? Under what conditions are these models invariant with respect to translation and rotation of the data?

Research Question 2

Can the use of certain secondary goals improve the performance of MIP models for the two-group classification problem on small to moderate sample sizes?

Research Question 3

Since the contaminated normal distribution (mixture of two normals) can be used to assess the performance of linear discriminant functions without a validation sample, how appropriate is this distribution for a simulation study in generating nonnormal data with a variety of values for the skewness and kurtosis measures? In particular, what range of values for the measures of skewness and kurtosis can the contaminated normal distributions have by using different parameter settings for the mean, standard deviation, and contaminating fraction?

CHAPTER III

THEORETICAL FRAMEWORK

The goal of classification analysis is to describe, either graphically or algebraically, the differential features of objects (observations) from several known collections (populations) and to allocate new objects into two or more labeled classes (Johnson and Wichern 1992). Good classification procedures are constructed to achieve a high rate of correctly classifying observations under certain conditions. If one class or population has a greater likelihood of occurrence than the others, the classification rule should take this prior probability of occurrence into account. The cost of misclassification is another important consideration. The cost of misclassifying an observation from group 1 into group 2 may be greater than the cost of misclassifying an observation from group 2 into group 1. Most classification rules can be adapted to take into account the cost of misclassification as well as the prior probability of occurrence (Banks and Abad 1991).

The Two-Group Classification Problem

The two-group statistical classification problem may be more formally stated as follows. Let G_i, $i = 1, 2$ be two distinct populations. Assume that each object in G_i possesses a set of common characteristics or attributes defined by $Y = (a_1, a_2, \ldots, a_p)^T$, where the superscript T denotes the transpose of the vector and the subscript p denotes the number of attributes. The a_i's are assumed to be observable numerical entities. If an observation $Y = (a_1, a_2, \ldots, a_p)^T$ is randomly selected from the combined populations of G_1 and G_2, the statistical classification problem is to construct a decision rule that optimizes some criterion that is a surrogate for classification accuracy.

For many two-group discriminant models with linear discriminant functions, the resulting decision rule consists of an estimated vector of weights $X = (x_1, x_2, \ldots, x_p)^T$ and scalars C_1 and C_2, which are employed in the following fashion to classify an observation $Y = (a_1, a_2, \ldots, a_p)^T$: assign observation Y to group 1 if

$$Y^TX = \sum_{i=1}^{p} a_i x_i \leq C_1$$

and assign observation Y to group 2 if

$$Y^TX = \sum_{i=1}^{p} a_i x_i \geq C_2 \quad .$$

The observation Y is misclassified if the discriminant score Y^TX does not fall on the correct side of the cutoff value C_1 or C_2. For some classification models, C_1 and C_2 are set equal to each other. In such cases, the optimal decision rule provides a hyperplane that separates the groups with a minimum number of misclassifications.

However, other models allow for a "classification gap" by letting C_2 be greater than C_1.

General Classification Rules for Parametric Models

Classification rules for the parametric statistical models are based on the assumption that each group under consideration has a multivariate population density function $f_i(Y)$ for $i = 1, 2$ over the p measured variables. Furthermore, let the prior probability and the cost of misclassification be defined as follows:

p_1	is the prior probability of being from group 1,
p_2	is the prior probability of being from group 2,
$C(1\|2)$	is the cost of misclassification when an observation from group 2 is incorrectly classified as from group 1,
$C(2\|1)$	is the cost of misclassification when an observation from group 1 is incorrectly classified as from group 2.

The cost function can be written as follows:

$$C(i|j) = \begin{cases} > 0 & \text{if } i \neq j \quad \text{for } i, j = 1, 2 \\ = 0 & \text{if } i = j \quad \text{for } i, j = 1, 2. \end{cases}$$

The optimal classification rule is to assign an observation Y to group 1 if

$$\frac{f_1(Y)}{f_2(Y)} \geq \left[\frac{C(1|2)}{C(2|1)}\right]\left[\frac{p_2}{p_1}\right]$$

and to assign an observation Y to group 2 if

$$\frac{f_1(Y)}{f_2(Y)} < \left[\frac{C(1|2)}{C(2|1)}\right]\left[\frac{p_2}{p_1}\right] .$$

Now if the misclassification costs are equal and the prior probabilities are equal, then the optimal classification rule is to assign an observation Y to group 1 if

$$\frac{f_1(Y)}{f_2(Y)} \geq 1$$

and to assign an observation Y to group 2 if

$$\frac{f_1(Y)}{f_2(Y)} < 1 .$$

For the discriminant functions used in this study, all misclassification costs are assumed to be equal and all prior probabilities are assumed to be equal. Hence, these parameters (costs and prior probabilities) do not need to be assigned values in the discriminant functions.

Fisher's Linear Discriminant Function (LDF)

Fisher's (1936) linear discriminant function is designed to maximize the likelihood of a correct classification (minimize the probability of misclassification) when the groups have multivariate normal distributions with equal variance-covariance structures. If $f_i(Y)$ is the multivariate normal distribution with mean vector μ_i and variance-covariance matrix Σ_i for i = 1, 2 and $\Sigma_1 = \Sigma_2 = \Sigma$, then the optimal classification rule is to assign an observation Y to group 1 if

$$(\mu_1 - \mu_2)^T\Sigma^{-1}Y \quad - \quad ½(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 + \mu_2) \quad \geq \quad 0$$

and to assign an observation Y to group 2 if

$$(\mu_1 - \mu_2)^T\Sigma^{-1}Y \quad - \quad ½(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 + \mu_2) \quad < \quad 0.$$

In most practical situations, the population parameters are not known. If μ_1, μ_2, and Σ are replaced by their corresponding maximum likelihood sample estimators $\overline{Y}_1$, $\overline{Y}_2$, and S, then the optimal classification rule is to assign an observation Y to group 1 if

$$(\overline{Y}_1 - \overline{Y}_2)^TS^{-1}Y \quad - \quad ½(\overline{Y}_1 - \overline{Y}_2)^TS^{-1}(\overline{Y}_1 + \overline{Y}_2) \quad \geq \quad 0$$

and to assign an observation Y to group 2 if

$$(\overline{Y}_1 - \overline{Y}_2)^TS^{-1}Y \quad - \quad ½(\overline{Y}_1 - \overline{Y}_2)^TS^{-1}(\overline{Y}_1 + \overline{Y}_2) \quad < \quad 0.$$

<u>Smith's Quadratic Discriminant Function (QDF)</u>

Smith's (1947) quadratic discriminant function is designed to maximize the likelihood of a correct classification (minimize the probability of misclassification) when the groups have multivariate normal distributions with unequal variance-covariance structures. The QDF model includes first-order terms and second-order terms of the attribute variables. Using the same notation as in the LDF model, but here $\Sigma_1 \neq \Sigma_2$, the optimal classification rule is to assign an observation Y to group 1 if

$$(Y - \mu_2)^T\Sigma_2^{-1}(Y - \mu_2) - (Y - \mu_1)^T\Sigma_1^{-1}(Y - \mu_1) - \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) \geq 0$$

and to assign an observation Y to group 2 if

$$(Y - \mu_2)^T\Sigma_2^{-1}(Y - \mu_2) - (Y - \mu_1)^T\Sigma_1^{-1}(Y - \mu_1) - \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) < 0.$$

If μ_1, μ_2, Σ_1, and Σ_2 are replaced by their corresponding maximum likelihood sample estimators $\overline{Y}_1$, $\overline{Y}_2$, S_1, and S_2, then the optimal classification rule is to assign an observation Y to group 1 if

$$(Y - \overline{Y}_2)^T S_2^{-1}(Y - \overline{Y}_2) - (Y - \overline{Y}_1)^T S_1^{-1}(Y - \overline{Y}_1) - \ln\left(\frac{|S_1|}{|S_2|}\right) \geq 0$$

and to assign an observation Y to group 2 if

$$(Y - \overline{Y}_2)^T S_2^{-1}(Y - \overline{Y}_2) - (Y - \overline{Y}_1)^T S_1^{-1}(Y - \overline{Y}_1) - \ln\left(\frac{|S_1|}{|S_2|}\right) < 0.$$

Mathematical Programming Models

In general, the mathematical programming models for solving the two-group classification problem develop a hyperplane separating the two groups. The hyperplane is described by the equation

$$\sum_{j=1}^{p} x_j a_{ij} = C .$$

The a_{ij} variable represents the value of attribute j for observation i. The x_j and C variables represent the unknown attribute weights and the cutoff value, respectively.

Minimize Sum of Deviations Model

There is a plethora of variations on the minimize sum of deviations (MSD) model (Koehler and Erenguc 1990). The model presented in Ragsdale and Stam (1991) is selected. This model is similar to the original model suggested by Hand (1981). It does not require any normalization constraints such as that proposed by Freed and Glover (1986); Glover, Keene, and Duea (1988); and Glover (1990). Some of these normalization constraints have undesirable side effects, as illustrated by Koehler (1989a and 1989b). The objective of the MSD model is to minimize the sum of misclassification deviations. The criterion of minimizing the misclassification deviations is a surrogate for directly minimizing the number of misclassifications. The MSD model by Ragsdale and Stam (1991), however, does include a gap which separates the hyperplanes used for classification. Hand (1981) referred to the gap as a "safety margin." Koehler (1989a) showed that Hand's model does not have the undesirable side effects displayed by some other mathematical programming models.

The MSD model of Ragsdale and Stam (1991) is presented below. The training sample consists of n_i ($i = 1, 2$) observations from each of two groups for a total of $n = n_1 + n_2$ observations. The notation G_1 and G_2 will denote the sets of observations from group 1 and group 2, respectively.

Notation:

d_i denotes the external (undesirable) deviation of a misclassified observation's discriminant score from 0 or ε. For a correctly classified observation, d_i is equal to zero.

a_{ij} denotes the j^{th} attribute value for observation i.

x_j denotes the weight for attribute j.

x_0 denotes the constant term in the discriminant function.

ε denotes the minimum gap size separating the discriminant scores between the two groups.

p denotes the number of predictor variables (attributes).

MSD Formulation:

Minimize $\sum_{i \in G_1} d_i + \sum_{i \in G_2} d_i$

subject to

$$x_0 + \sum_{j=1}^{p} a_{ij}x_j - d_i \leq 0 \qquad i \epsilon G_1$$

$$x_0 + \sum_{j=1}^{p} a_{ij}x_j + d_i \geq \varepsilon \qquad i \epsilon G_2$$

where

x_j is a sign-unrestricted variable (j = 0, 1,..., p)

d_i is a nonnegative variable (i = 1, 2, ..., n)

ε is a small positive constant.

Mixed Integer Programming Model

The mixed integer programming (MIP) model used in this study is similar to that presented in Koehler and Erenguc (1990). By replacing the d_i's in the MSD model with binary variables I_i's and multiplying the I_i's by a large constant M in the constraints, it is easy to construct the MIP model. Using the same notation as in the MSD model, the MIP formulation is expressed next.

MIP formulation:

Minimize $\sum_{i \in G_1} I_i + \sum_{i \in G_2} I_i$

subject to

$$x_0 + \sum_{j=1}^{p} a_{ij} x_j - M I_i \leq 0 \qquad i \epsilon G_1$$

$$x_0 + \sum_{j=1}^{p} a_{ij} x_j + M I_i \geq \varepsilon \qquad i \epsilon G_2$$

where

x_j is a sign-unrestricted variable ($j = 0, 1, \ldots, p$)

I_i is a binary variable ($i = 1, 2, \ldots, n$)

ε is a small positive constant

M is a large positive constant.

In the above constraints, the M parameter can be interpreted as the maximum possible deviation that a misclassified observation can be from the gap. In choosing the values of M and ε, Koehler and Erenguc (1990, page 71) noted that "we rely on the standard maxim in mixed integer programming to choose M large enough and ε small enough."

Hybrid Model

The hybrid model was first introduced by Freed and Glover (1986). Unlike the objective function of the MSD model, which is only minimizing the external (undesirable) deviations, the objective function of the hybrid model simultaneously considers both minimizing the external deviations and maximizing the internal (desirable) deviations. Furthermore, the hybrid model also considers the maximum deviation of observations from the separating hyperplane.

Hybrid formulation:

Minimize $$h_0\alpha_0 + \sum_{i\in G} h_i\alpha_i - k_0\beta_0 - \sum_{i\in G} k_i\beta_i$$

subject to

$$\sum_{j=1}^{p} a_{ij}x_j - \alpha_0 - \alpha_i + \beta_0 + \beta_i = b \quad i\epsilon G_1$$

$$\sum_{j=1}^{p} a_{ij}x_j + \alpha_0 + \alpha_i - \beta_0 - \beta_i = b \quad i\epsilon G_2$$

$$-n_2 \sum_{i\in G_1} \sum_{j=1}^{p} a_{ij}x_j + n_1 \sum_{i\in G_2} \sum_{j=1}^{p} a_{ij}x_j = 1$$

where

α_0, α_i, β_0, and β_i are nonnegative variables

x_i's and b are sign-unrestricted variables

$G = G_1 \cup G_2$.

To give an interpretation of the objective function of the hybrid model, the α_i's can be considered as the misclassified (external) deviations, and the β_i's can be considered as the correctly classified (internal) deviations. The term α_0 can be interpreted as the maximum external deviation, whereas β_0 can be interpreted as the minimum internal deviation if the h_i and k_i weights in the objective function are very large relative to h_0 and k_0. The last constraint is a normalization used to prevent a degenerate (zero) solution. Glover (1990, page 772) remarked that this normalization "eliminates the previous distortions in the LP models and has attractive properties enabling it to obtain demonstrably superior solutions." In this study, $h_0 = 150$, $h_i = 2$, $k_0 = 80$, and $k_i = 1$ are selected. This choice of coefficient values is consistent with Glover's (1990) recommendations and with the parameters of the hybrid model in Silva and Stam (1994).

Second-Order Model Formulation

To form second-order mathematical programming formulations, the squared attribute values and the crossproduct values of all attributes need to be included as additional predictor variables. Note that the second-order terms in the MSD, MIP, and hybrid formulation still have constraints that are linear in the x parameters (coefficients of the discriminant rule). However, the constraints are obviously nonlinear in the attribute values. Since the second-order mathematical programming models have all of the terms present in the Smith's quadratic discriminant function, the second-order mathematical programming formulations have the potential of being competitive with the quadratic method in problems requiring a nonlinear classification function.

The following lemma and theorem are presented to establish that the MSD, MIP, and hybrid models with all first-order terms and second-order terms are translationally and rotationally invariant. Furthermore, the MIP model will not have more misclassifications on the training sample than the MSD, hybrid, or QDF methods if all first-order terms and second-order terms are included in the models.

Lemma 1

Any linear combination of second-order and first-order terms of $a_i = (a_{i1}, a_{i2}, \ldots, a_{ip})^T$ can be expressed as $a_i^T W a_i + a_i^T x$, where $W = (w_{hk})$ is a symmetric matrix and $x = (x_1, x_2, \ldots, x_p)^T$. The coefficients of the square terms are w_{hh}, the coefficients of the crossproduct terms are $2w_{hk}$, and the coefficients of the first-order terms are x_j.

Proof

A linear combination of the second-order and first-order terms of $a_i = (a_{i1}, a_{i2}, \ldots, a_{ip})^T$ can be written as

$$\sum_{j=1}^{p} a_{ij}^2 x_{jj} + \sum_{j=1}^{p} a_{ij} x_j + \sum_{h>k}\sum a_{ih} a_{ik} x_{hk} = \sum_{h\geq k}\sum a_{ih} a_{ik} x_{hk} + \sum_{j=1}^{p} a_{ij} x_j$$

$$= \sum_{h=1}^{p}\sum_{k=1}^{p} a_{ih} a_{ik} w_{hk} + \sum_{j=1}^{p} a_{ij} x_j$$

$$= a_i^T W a_i + a_i^T x$$

where $W = (w_{hk})$ and

$$w_{hk} = \begin{cases} x_{hk}/2 & \text{if } h > k \\ x_{kh}/2 & \text{if } h < k \\ x_{hk} & \text{if } h = k \end{cases}$$

From the equations above, the statement of the lemma readily follows.

Theorem 1

If all of the first-order terms and second-order terms are included in the MIP, MSD, and hybrid formulations, then

1. The MIP method will not have more misclassifications than the MSD, hybrid, or QDF methods on the training sample.

2. The MIP, MSD, and hybrid methods are rotationally and translationally invariant.

Proof

The first statement follows since the MIP procedure directly minimizes the total number of misclassifications on the training sample as seen by its objective function and since each of the MIP, MSD, hybrid, and QDF procedures is assumed in this theorem to contain all first-order and second-order terms. To show that the second statement holds, let P be an orthogonal matrix and let c be a vector of constants of length p. By Lemma 1, the discriminant score of observations for the MSD, MIP, and hybrid formulations can be written as $x_0 + a_1^T W a_1 + a_1^T x$.

Now consider both an orthogonal rotation P and a translation c of the a_1 vector. We have

$$x_0 + [P(a_1 + c)]^T W [P(a_1 + c)] + [P(a_1 + c)]^T x$$

$$= x_0 + a_1^T (P^T W P) a_1 + (Pc)^T W P a_1 + a_1^T (P^T W P) c$$

$$+ c^T (P^T W P) c + a_1^T P^T x + c^T P^T x$$

$$= \tilde{x}_0 + a_1^T \tilde{W} a_1 + a_1^T \tilde{x}$$

where

$$\tilde{x}_0 = x_0 + c^T (P^T W P) c + c^T P^T x$$

$$\tilde{W} = P^T W P$$

$$\tilde{x} = P^T x + 2P^T W P c.$$

We can see that $\tilde{x}_0 + a_1^T \tilde{W} a_1 + a_1^T \tilde{x}$ is still a linear combination of both the first-order and second-order terms of the values of the vector a_1. Thus, the statement of the theorem follows.

Note that if some of the second-order terms are missing, such as the crossproduct terms, then it is possible that the QDF procedure may produce fewer misclassifications than the MIP procedure on the training sample. Also note that if the crossproduct terms were missing from the second-order models for the MSD, MIP, and hybrid procedures, then these formulations would not be rotationally invariant.

MIP Models with Secondary Goals

Four MIP models that are used with secondary goals are investigated. The first and second MIP models have the secondary goal of maximizing the distance between the means of the discriminant scores for the two groups. These two models have not been previously investigated. The third and fourth MIP models are existing models that have not been thoroughly investigated under nonnormal configurations. The secondary goal of the third MIP model is used to maximize the minimum deviation of the correctly classified observations, whereas the secondary goal of the fourth MIP model is used to minimize the sum of all the deviations of the misclassified observations from the cutoff value in the discriminant rule. Because the motivation for including the secondary goal of maximizing the distance between the means of the discriminant score of attribute values is based on Fisher's method, it follows that this secondary goal may perhaps be more appropriate with only first-order terms in the MIP models. The four MIP models with secondary goals are presented next.

MIP 1: MIP model with a secondary goal of maximizing the distance between projected means (bounded scores).

$$\text{Minimize} \quad P_1\sum_{i=1}^{n} I_i \; - \; P_2\delta$$

subject to

$$\sum_{j=1}^{p} a_{ij}x_j - M_1I_i \quad \le \quad c - \varepsilon \qquad i\epsilon G_1$$

$$\sum_{j=1}^{p} a_{ij}x_j \quad \ge \quad c - M_2 \qquad i\epsilon G_1$$

$$\sum_{j=1}^{p} a_{ij}x_j + M_1I_i \quad \ge \quad c + \varepsilon \qquad i\epsilon G_2$$

$$\sum_{j=1}^{p} a_{ij}x_j \quad \le \quad c + M_2 \qquad i\epsilon G_2$$

$$\sum_{j=1}^{p} \bar{a}_j^{(2)} x_j - \sum_{j=1}^{p} \bar{a}_j^{(1)} x_j \quad \ge \quad \delta$$

where

P_1, P_2 are positive constants

I_i is a binary variable ($i = 1, 2, \ldots, n$)

x_j is a sign-unrestricted variable ($j = 1, 2, \ldots, p$)

M_1, M_2, and ε are positive constants

δ is a nonnegative variable

a_{ij} is the j^{th} attribute value for the i^{th} observation

$\bar{a}_j^{(i)}$ is the average value of the a_j's for group i

c is a sign-unrestricted variable.

MIP 2: MIP model with a secondary goal of maximizing the distance between projected means (bounded coefficients).

$$\text{Minimize} \quad P_1\sum_{i=1}^{n} I_i \; - \; P_2\delta$$

subject to

$$\sum_{j=1}^{p} a_{ij}x_j - MI_i \quad \le \quad c - \varepsilon \qquad i\in G_1$$

$$\sum_{j=1}^{p} a_{ij}x_j + MI_i \quad \ge \quad c + \varepsilon \qquad i\in G_2$$

$$\sum_{j=1}^{p} \overline{a}_j^{(2)} x_j - \sum_{j=1}^{p} \overline{a}_j^{(1)} x_j \quad \ge \quad \delta$$

$$-1 \le x_j \le 1$$

where

P_1, P_2 are positive constants

I_i is a binary variable ($i = 1, 2, \ldots, n$)

x_j is a sign-unrestricted variable ($j = 1, 2, \ldots, p$)

M and ε are positive constants

δ is a nonnegative variable

a_{ij} is the j^{th} attribute value for the i^{th} observation

$\overline{a}_j^{(i)}$ is the average values of a_j for group i

c is a sign-unrestricted variable.

MIP 3: MIP model with a secondary goal of maximizing the minimum internal deviation (bounded coefficients).

$$\text{Minimize} \quad P_1\sum_{i=1}^{n} I_i \; - \; P_2 d$$

subject to

$$\sum_{j=1}^{p} a_{ij}x_j + d - MI_i \le c - \varepsilon \qquad i \in G_1$$

$$\sum_{j=1}^{p} a_{ij}x_j - d + MI_i \ge c + \varepsilon \qquad i \in G_2$$

$$-1 \le x_j \le 1$$

where

P_1, P_2 are positive constants

I_i is a binary variable ($i = 1, 2, \ldots, n$)

x_j is a sign-unrestricted variable ($j = 1, 2, \ldots, p$)

M and ε are positive constants

d is a nonnegative variable

a_{ij} is the j^{th} attribute value for the i^{th} observation

c is a sign-unrestricted variable.

MIP 4: MIP model with a secondary goal of minimizing the sum of external deviations.

$$\text{Minimize} \quad P_1\sum_{i=1}^{n} I_i \; + \; P_2\sum_{i=1}^{n} d_i$$

subject to

$$\sum_{j=1}^{p} a_{ij}x_j - d_i \le c - \varepsilon \qquad i\epsilon G_1$$

$$\sum_{j=1}^{p} a_{ij}x_j + d_i \ge c + \varepsilon \qquad i\epsilon G_2$$

$$MI_i \ge d_i$$

where

P_1, P_2 are positive constants

I_i is a binary variable ($i = 1, 2, \ldots, n$)

x_j is a sign-unrestricted variable ($j = 1, 2, \ldots, p$)

M and ε are positive constants

d_i is a nonnegative variable ($i = 1, 2, \ldots, n$)

a_{ij} is the j^{th} attribute value for the i^{th} observation

c is a sign-unrestricted variable.

Contaminated Normal Distribution

Contaminated normal distribution is considered to be an important distribution in representing real-world data (Hampel 1974, Nath 1984, and Lee and Ord 1990). However, the contaminated normal distribution appears in only a few simulation studies evaluating LP-based models. The range of possible values for skewness and kurtosis measures does not appear to be readily available for this distribution.

The notation $CMN(\mu_1,\Sigma_1,\mu_2,\Sigma_2,\epsilon) = (1-\epsilon)N(\mu_1,\Sigma_1) + \epsilon N(\mu_2,\Sigma_2)$ will be used to denote the general contaminated multivariate normal distribution. The notation $N(\mu,\Sigma)$ represents the normal distribution with mean vector μ and variance-covariance matrix Σ. The $N(\mu_2,\Sigma_2)$ population can be interpreted as the contaminating part, and ϵ can be interpreted as the contaminating fraction of the data. Therefore, this general contaminated normal distribution can be viewed as a mixture of two normal populations. As the ϵ parameter becomes larger, the shapes of contaminated normal distribution are seen, not as one larger normal population with a small set of outliers, but rather as a mixture of two normally distributed populations. For $\epsilon = 0$ or 1, the contaminated multivariate normal distribution simply reduces to a multivariate normal distribution. Each of the parameters μ_1, Σ_1, μ_2, Σ_2, and ϵ plays a role in determining the skewness and kurtosis values of the distribution. This version of the contaminated normal distribution is more general than the distribution presented in Nath (1984) and Lee and Ord (1990). In their studies, μ_1 and μ_2 were selected to be equal, and, thus, the contaminated normal distribution was a symmetrical distribution and always had a value of zero for the skewness measure.

To show that any linear transformation of X_0, for X_0 from $CMN(\mu_1, \Sigma_1, \mu_2, \Sigma_2, \epsilon)$ is distributed as a contaminated univariate normal, consider the following equations where F is a cumulative distribution function, ℓ is a vector of constants, c is a constant, and Φ represents the standard normal cumulative distribution.

$$
\begin{aligned}
&P[\ell^T X_0 \le c] \\
&\quad = P[\ell^T X_0 \le c \text{ and } X_0 \text{ from } N(\mu_1, \Sigma_1) \\
&\qquad \text{or } \ell^T X_0 \le c \text{ and } X_0 \text{ from } N(\mu_2, \Sigma_2)] \\
&\quad = P[\ell^T X_0 \le c \mid X_0 \text{ from } N(\mu_1, \Sigma_1)] \cdot P[X_0 \text{ from } N(\mu_1, \Sigma_1)] \\
&\qquad + P[\ell^T X_0 \le c \mid X_0 \text{ from } N(\mu_2, \Sigma_2)] \cdot P[X_0 \text{ from } N(\mu_2, \Sigma_2)] \\
&\quad = (1 - \epsilon)\Phi[(c - \ell^T\mu_1)/(\ell^T\Sigma_1\ell)^{1/2}] + \epsilon\Phi[(c - \ell^T\mu_2)/(\ell^T\Sigma_2\ell)^{1/2}]
\end{aligned}
$$

Thus, $\ell^T X_0$ is distributed as a contaminated univariate normal distribution. An alternative proof could be provided using characteristic functions, as in Nath (1984). From the above equations, exact misclassification rate could easily be obtained for a given linear discriminant function. Therefore, under this distribution, the need for validation samples can be eliminated when a linear discriminant function is being evaluated.

It should be noted that the marginal distributions of the contaminated multivariate normal distribution are simply contaminated univariate normal distribution. Any random variable with a contaminated univariate normal distribution can be shifted and scaled so that its cumulative distribution function is

$$F(X) = (1-\epsilon)\Phi(X) + \epsilon\Phi((X-\mu)/\sigma) \ .$$

Using the technique given in Hogg and Craig (1978), the first, second, third, and fourth moments can be generated as the following:

$$E[X] = \epsilon\mu$$

$$E[X^2] = (1-\epsilon) + \epsilon(\sigma^2 + \mu^2)$$

$$E[X^3] = 3\epsilon\sigma^2\mu + \epsilon\mu^3$$

$$E[X^4] = 3(1-\epsilon) + 3\epsilon\sigma^4 + 6\epsilon\sigma^2\mu^2 + \epsilon\mu^4$$

Now let γ_1 and γ_2 be the notation for the skewness and kurtosis measures, respectively. Using the standard definitions for the measures of skewness and kurtosis, namely $E[(X-\mu)^3/\sigma^3]$ and $E[(X-\mu)^4/\sigma^4]$, γ_1 and γ_2 can be mathematically derived as

$$\gamma_1 = \frac{\epsilon\mu(1-\epsilon)(3\sigma^2 + \mu^2 - 2\mu^2\epsilon - 3)}{[1 - \epsilon + \epsilon\sigma^2 + \mu^2\epsilon(1 - \epsilon)]^{3/2}}$$

$$\gamma_2 = \frac{6\epsilon\mu^2\sigma^2(\epsilon^2-2\epsilon+1)+\epsilon\mu^4(1+6\epsilon^2-3\epsilon^2-4\epsilon)+6\mu^2\epsilon^2(1-\epsilon)+3(1-\epsilon)+3\epsilon\sigma^4}{[1 - \epsilon + \epsilon\sigma^2 + \mu^2\epsilon(1 - \epsilon)]^2}$$

From the above formulas, the pattern of possible values of skewness and kurtosis for various values of parameters μ, σ, and ϵ can be obtained. Also, the limiting values of the skewness and kurtosis measures when μ and/or σ approach infinity can be determined.

To understand the relationship between the values of the skewness and kurtosis measures, consider the following theorem.

Theorem 2

Let $\hat{\gamma}_1$ and $\hat{\gamma}_2$ be defined as the sample skewness measure and the sample kurtosis measure, respectively, as in Bickel and Doksum (1977). That is,

$\hat{\gamma}_1 = n^{1/2}\Sigma(X_i - \overline{X})^3/(\Sigma(X_i - \overline{X})^2)^{3/2}$, and

$\hat{\gamma}_2 = n\Sigma(X_i - \overline{X})^4/(\Sigma(X_i - \overline{X})^2)^2$, then

1. $\hat{\gamma}_1^2 \le \hat{\gamma}_2 - 1$

2. $\hat{\gamma}_2 \ge 1$

<u>Proof</u> Let

$$A = (X_1,\ X_2,\ \ldots,\ X_n)$$

$$B = ((X_1-\bar{X})^2,\ (X_2-\bar{X})^2,\ \ldots,\ (X_n-\bar{X})^2)$$

$$\bar{X} = \Sigma(X_i)/n$$

$$S_n = (\Sigma(X_i - \bar{X})^2/n)^{1/2}$$

$r_{A,B}$ = the sample correlation coefficient between A and B.

Note that

$$\Sigma[(X_i - \bar{X})^2 - S_n^2]^2$$

$$= \Sigma(X_i - \bar{X})^4 - nS_n^4$$

$$= [(\Sigma(X_i - \bar{X})^2)^2/n][n\Sigma(X_i - \bar{X})^4/(\Sigma(X_i - \bar{X})^2)^2] - nS_n^4$$

$$= nS_n^4\hat{\gamma}_2 - nS_n^4$$

$$= nS_n^4(\hat{\gamma}_2 - 1).$$

Hence,

$$r_{A,B} = \frac{\Sigma[(X_i - \bar{X})((X_i - \bar{X})^2 - S_n^2)]}{(nS_n^2)^{1/2}\ (\Sigma[(X_i - \bar{X})^2 - S_n^2]^2)^{1/2}}$$

$$= \frac{\Sigma(X_i - \bar{X})^3}{(nS_n^2)^{1/2}\ (nS_n^4(\hat{\gamma}_2 - 1))^{1/2}}$$

$$= \frac{nS_n^3\ \hat{\gamma}_1}{(nS_n^2)^{1/2}\ (nS_n^4(\hat{\gamma}_2 - 1))^{1/2}}$$

$$= \frac{\hat{\gamma}_1}{\sqrt{\hat{\gamma}_2 - 1}}$$

Since $r_{A,B} \le 1$, we have that

$$\frac{\hat{\gamma}_1}{\sqrt{\hat{\gamma}_2 - 1}} \le 1$$

or
$$\hat{\gamma}_1^2 \le \hat{\gamma}_2 - 1.$$

The second statement follows from the fact that $2\hat{\gamma}_1^2 \ge 0$, therefore $\hat{\gamma}_2 \ge 1$.

The above result does not appear to be readily available. It can be found in Devroye (1986), which used somewhat complicated Hankel determinants to prove it. However, the above proof shows that the result can readily follow from the sample correlation between A and B in the above theorem. This proof does not appear to be mentioned by many mathematical statistics books or simulation textbooks, such as Devroye (1986), Hogg and Craig (1978), and Bickel and Doksum (1977).

CHAPTER IV

SIMULATION DESIGNS

Simulation Designs for Models with Second-Order Terms

To determine how second-order terms in mathematical programming models affect their classificatory performance relative to the first-order models and the parametric statistical procedures, a Monte Carlo simulation study is conducted. Eleven classification models are used in this study and are listed in Table 3. The notations MSD5, MIP5, and HYB5 are used to denote the MSD, MIP, and hybrid procedures, respectively, with all of the squared, linear, and crossproduct terms in the models. The notations MSD4, MIP4, and HYB4 are used to denote the MSD, MIP, and hybrid procedures, respectively, with only the squared and linear terms (no crossproduct term) in the models. For the MSD, MIP, and hybrid procedures, which contain only the linear terms, the notations MSD2, MIP2, and HYB2, respectively, are used to indicate them. The notations LDF and QDF are used for the Fisher's linear discriminant function and the Smith's quadratic discriminant function, respectively.

Eight different data configurations are examined in the simulation study. The population distributions in the first six configurations are normally distributed, while the last two configurations contain nonnormal data. These data configurations are described in Table 4. Configurations 1A and 1B are the configurations in which a first-order (linear) classification rule would be optimal since the variance-covariance structures of the two populations are equal. The observations in configuration 1B are correlated, whereas the observations in configuration 1A are uncorrelated. For the other configurations, it is expected that a second-order (nonlinear) classification rule would be the classification rule of choice.

Configurations 1C and 1D are selected for examining the usefulness of the crossproduct terms in the mathematical programming models when correlation is present in the data. Configurations 1C and 1D contain interesting covariance structures. The crossproduct term for the QDF procedure with configuration 1C should not be needed because the off-diagonal terms of the matrix $M = \Sigma_1^{-1} - \Sigma_2^{-1}$ cancel out and, thus, M is a diagonal matrix (where Σ_1 and Σ_2 are the covariance matrices of the first and second populations, respectively; see Johnson and Wichern 1992, page 509). However, the crossproduct term for the QDF procedure with configuration 1D should be important since the off-diagonal terms of the matrix M are the only non-zero elements. The simulation study will show how important the crossproduct terms are in the mathematical programming models.

Configuration 1E is selected because there is no correlation among the variables for either group and because it is a configuration in which the QDF model should easily perform well. Configuration 1F consists of two normal populations with identical means, but the variance-covariance structure of one population is much larger than that of the other population. Configuration 1F can be viewed as one normal population contained inside another normal population. Any first-order linear discriminant function would be expected to perform poorly on a set of data from this configuration.

Configuration 1G is one of the two configurations that contain nonnormal data. The second population of configuration 1G consists of a normal population with mean vector $(2, 2)^T$ and two independent variables with each variance equals to one, but this population also contains a 15% contamination from a set of normally distributed outliers. The outlier group has mean vector $(-10, -10)^T$ and two independent variables with variances equal to 9.

Configuration 1H is the other configuration that contains nonnormal data. Consider a population in which the first attribute variable is uniformly distributed over the interval from 0.1 to 5.0 and the second attribute variable is uniformly distributed over the interval 0 to $1/a_1$ where a_1 is the value of the first attribute variable. Hence, the value of the second attribute variable is conditional on the value of the first attribute variable. Eighty percent of the first population's observations come from this distribution. The other 20% of observations come from the point (-4.60894, -4.60894). Now consider a population in which the first attribute variable is again uniformly distributed over the interval from 0.1 to 5.0, but the second attribute variable is uniformly distributed over the interval $1/a_1$ to $1/a_1 + 0.5$ where a_1 is the value of the first attribute variable. Note that the second attribute variable is dependent on the value of the first attribute variable. Eighty percent of the second population's observations come from this distribution. The other 20% of observations come from the point (4.195634, 4.195634). For group 1 and group 2, the points (-4.60894, -4.60894) and (4.195634, 4.195634) were selected to make the two attribute variables in each group uncorrelated.

Graphically, 80% of the values of the first population in configuration 1H can be thought of as falling under the curve $Y = 1/X$ on a two-dimensional graph with X being equal to values between 0.1 and 5.0, whereas 80% of the values of the second population fall above the curve $Y = 1/X$. The other 20% of the observations for configuration 1H come from a point for each group. Thus, 20% of the observations from each population can be considered outliers. While the distributions of the populations in configuration 1H are not commonly mentioned in the literature, they are included to gain some insight into the performance of various models on configurations that may include a mixture of continuous and discrete data. Also, the shape of this data will allow the correlation of the variables in each group to be zero. The simulation study can then assess the appropriateness of the crossproduct terms. In addition, these populations are highly nonnormal, and it is expected that the second-order mathematical programming models will perform well.

For each configuration in this simulation study, training sample sizes of $n_i = 25$, $i = 1, 2$, and $n_i = 50$, $i = 1, 2$ are used for each of the two groups. Validation sample sizes of 500 are used for each group, for a total of 1000 observations for each validation sample. In each simulation experiment, two attribute values are generated for each observation. The simulation study is performed by using the SAS statistical package (version 6.07) on the Solbourne 6/904 computer operating under UNIX at the Computing Center of the University of North Texas. All experimental conditions are replicated 100 times.

For each replication, the number of misclassified observations in both the training sample and the validation sample is determined. The mean and standard deviation of the number of misclassified observations are computed for 100 replications. Paired t-tests are used to indicate significant differences in classificatory performance among the models. A Bonferroni adjustment (Johnson and Wichern

1992) is used in finding the critical values of the test. Two models will be referred to as being significantly different if the paired t-test calculated from their misclassification rates shows a significant difference at the .05/55 significance level.

Simulation Designs for Models with Secondary Goals

In this section, four MIP models that include secondary goals are evaluated on normal and contaminated normal data. These data are used because they are important distributions representing real-world data. An additional advantage is that the exact misclassification rate on the estimated classification functions for these models can be calculated with these particular distributions, and, thus, large validation samples are not necessary. However, only linear (first-order) terms of the attribute variables can be used to easily obtain this exact misclassification rate. The objective of this section is to evaluate the added classificatory power that results from the secondary goals in the MIP models.

Four classification models are examined in a Monte Carlo simulation study to answer Research Question 2. These models are listed in Table 21 and are presented in the theoretical framework chapter of this dissertation. These models are labeled MIP1, MIP2, MIP3, and MIP4. Note that all of these models result in a classification function that is linear in terms of the attribute variables.

Fourteen different data configurations are used in this simulation study. These data configurations are described in Table 22. The population distributions of the data in the first three configurations are normally distributed, whereas those in the last eleven configurations are contaminated normal distributions. Some configurations contain contaminated normal data in only one of the two groups, while other configurations contain contaminated normal data in both groups. The last three data configurations contain contaminated normal populations with different values of skewness and kurtosis. Configuration 2L is designed to have low skewness (0.461) and high kurtosis (13.419). Configuration 2M, however, is designed to have moderate values of skewness (1.625) and kurtosis (7.612). Configuration 2N is designed to have low skewness (0.129) and very low kurtosis (2.214). Since the MIP2 and MIP3 models are not rotationally invariant, different orientations of the normal populations and contaminated normal populations are also considered in evaluating the variability in the classificatory performance of the MIP models. Contaminating fractions of 10%, 15%, and 20% are used on some data configurations.

There are two training sample sizes of $n_i = 20$, $i = 1, 2$, and $n_i = 40$, $i = 1, 2$ for each of the two groups in each data configuration. The training sample sizes are slightly less than the training sample sizes used in the simulation design described in the previous section. The smaller sample sizes in this section were chosen because of the computational intensiveness of MIP models. Validation samples are not used in this part of the study because exact misclassification rates can be directly determined from the estimated classification functions. The misclassification rates of the MIP models with secondary goals will all perform the same on the training samples because each model has the same first goal. In each simulation experiment, two independent variables are generated for each observation. The simulation study is performed by using the SAS statistical package (version 6.07) on the Solbourne 6/904 computer operating under UNIX at the Computing Center of the University of North Texas. All experimental conditions are replicated 200 times.

For each replication, the probability that a new observation will be misclassified (the estimate of the expected actual misclassification rate) is calculated. The mean and standard deviation of the estimated misclassification rates are computed on the 200 replications. Paired t-tests are used to indicate significant differences in classificatory performance among the models. A Bonferroni adjustment (Johnson and Wichern 1992) is used in finding the critical values of the test.

Two models will be referred to as being significantly different if the paired t-test calculated from their misclassification rates shows a significant difference at the .05/6 significance level.

CHAPTER V

EXPERIMENTAL RESULTS

Simulation Results for Models with Second-Order Terms

The results of a Monte Carlo simulation for models with second-order terms are presented in this section. These results will be used to answer Research Question 1. Tables 5 through 20 and Figures 1 through 8 contain the results from the simulation study.

Configuration 1A

For configuration 1A, the LDF model is expected to perform well since the two populations each have a normal distribution with equal variance-covariance structures. Thus, the squared and crossproduct terms should not be necessary for the mathematical programming models to perform well.

The results in Table 5 show that the LDF model has the lowest misclassification rate on the validation samples for both training samples of sizes 25 and 50 per group. The average misclassification rates on the validation samples of the LDF model are 8.36% and 8.13% for training samples of sizes 25 and 50 per group, respectively. However, all of the MSD and MIP models have a lower misclassification rate than both of the LDF and QDF models on the training samples for both training samples of sizes 25 and 50 per group. The QDF model performs almost as well as the LDF model.

As expected, the mathematical programming models with only linear terms (2 variables) outperform the mathematical programming models with crossproduct and squared terms on the validation samples for both training sample sizes. The addition of second-order terms decreases classificatory performance of the mathematical programming models, particularly for the case of 25 observations per training group. The mathematical programming models without the crossproduct term perform better than the second-order mathematical programming models with both the crossproduct and squared terms in the models.

The best mathematical programming models on the validation samples for training samples of sizes 25 and 50 per group of configuration 1A are HYB2 and MSD2, respectively. The average misclassification rate on the validation samples of the HYB2 model is 8.63% for training samples of size 25 per group. For the MSD2 model with training samples of size 50 per group, the average misclassification rate on the validation samples is 8.40%. These results are close to the results of the LDF model. The model that has the highest misclassification rate on the validation samples for this data configuration is MIP5. The average misclassification rates on the validation samples for the MIP5 are 14.66% and 11.64% for training samples of sizes 25 and 50 per group, respectively. However, the MIP5 model yields the lowest number of misclassified observations in the training samples. This occurs because the objective function of the MIP model is to directly minimize the number of misclassified observations and the MIP5 model contains all of the linear, squared, and crossproduct terms.

Table 13 shows the results of paired t-tests for the mean difference in classificatory performance of the models on validation samples for configuration 1A. The results reveal that the performance of the LDF model is significantly different from the performance of all other mathematical programming models with the Bonferroni adjustment to the family of 55 tests, thus using a significance level of .05/55. The results also reveal that the performance of the first-order MSD, MIP, and hybrid models is significantly different from the performance of the corresponding second-order MSD, MIP, and hybrid models, respectively, for configuration 1A.

Configuration 1B

Configuration 1B is another data configuration in which the variance-covariance structures of the two populations are equal. However, the observations within each population are correlated with the coefficient of correlation equals to

0.6. For this configuration, the LDF model is expected to perform optimally because all statistical assumptions are met. Hence, it is expected that the first-order mathematical programming models should outperform the second-order mathematical programming models.

The results in Table 6 show that the LDF model has the lowest average misclassification rate on the validation samples for both training samples of sizes 25 and 50 per group. The average misclassification rates on the validation samples for the LDF model are 4.98% and 4.74% for training samples of sizes 25 and 50 per group, respectively. However, the LDF model has an average misclassification rate on the training samples that is higher than those of the MSD, MIP, and QDF models. The QDF model performs almost as well as the LDF model. These results are similar to the results of configuration 1A. In configuration 1A, the standard deviation of the misclassification rate on the validation samples decreases for all models, except the three hybrid models, when the training sample size is increased from 25 to 50 per group. For configuration 1B, only the standard deviation for the HYB5 model increases for the misclassification rate on the validation sample when the training sample size increases from 25 to 50 per group.

For mathematical programming models, the models with only first-order terms outperform the corresponding models with the squared terms and crossproduct terms on the validation samples. The addition of squared and/or crossproduct terms decreases classificatory performance of the mathematical programming models, despite the superior performance of the second-order mathematical programming models on the training samples.

The best mathematical programming model on the validation samples for this data configuration is HYB2 model for training samples of size 25 per group. For training samples of size 50 per group, the best mathematical programming model is MSD2. The average misclassification rate on the validation samples for the HYB2 model is 5.18% for training samples of size 25 per group. The average misclassification rate on the validation samples of the MSD2 model is 5.10% for training samples of size 50 per group. These results are close to the results of the LDF model. The worst classification model on the validation samples for this data configuration is MIP5. The average misclassification rates on the validation sample of the MIP5 are 11.06% and 8.09% for training samples of sizes 25 and 50 per group, respectively.

From Table 14, the results of paired t-tests of the classificatory performance of the models on the validation samples reveal that only the HYB2 model is not significantly different from the LDF model for training samples of size 25 per group on configuration 1B. While the performances of the HYB2 and MSD2 on the validation samples are not significantly different for configuration 1A, they are significantly different for configuration 1B for training samples of size 25 per group. The results also reveal that the performance of the first-order mathematical programming models is significantly different from the performance of the corresponding second-order mathematical programming models for this data configuration.

Configuration 1C

Configuration 1C is a data configuration with unequal variance-covariance structures for the two populations. It is expected that the QDF model will perform optimally on data from these normally distributed populations. The variance-covariance structures of this data configuration are interesting in that the off-diagonal terms of matrix $\Sigma_1^{-1} - \Sigma_2^{-1}$ cancel out. Therefore, the crossproduct term for the QDF model is not expected to be needed. It is also expected that the second-order mathematical programming models without the crossproduct term will outperform the other corresponding mathematical programming models on the validation samples.

The results in Table 7 show that the QDF model has the best classification rate on the validation samples for this data configuration for both training samples of sizes 25 and 50 per group. The average misclassification rate on the validation samples of the QDF model are 6.88% and 6.47% for training samples of sizes 25 and 50 per group, respectively. As expected, the LDF model does not perform well on the validation samples for both training sample sizes. In fact, the LDF model has the highest misclassification rate on the validation samples for training samples of size 50 per group.

The results for the mathematical programming models are somewhat surprising for the cases of 25 observations per training group. For training samples of size 25 per group, the MSD2, MIP2, and HYB5 are the best MSD, MIP, and hybrid classification models, respectively. This is surprising since it is expected that the MSD4, MIP4, and HYB4 models would be the classification models of choice for the MSD, MIP, and hybrid formulations, respectively. The best mathematical programming model for training samples of size 25 per group is the HYB5. The average misclassification rate on validation samples of the HYB5 model is 8.53% for training samples of size 25 per group. When the training sample size increases to 50 per group, the results are the same as what is expected. With training samples of size 50 per group, the second-order mathematical programming models without the crossproduct term outperform the corresponding second-order mathematical programming models with the crossproduct term and the corresponding first-order mathematical programming models. The best mathematical programming model for training samples of size 50 per group is the MSD4. The average misclassification rate on validation samples of the MSD4 model is 7.21% for training samples of size 50 per group. The models that have the highest misclassification rate on the validation samples for training samples of size 50 per group are the first-order mathematical programming models. This is expected because of the unequal variance-covariance structure of the two populations.

Table 15 shows the results of paired t-tests on the classificatory performance of the models on validation samples for configuration 1C. The results reveal that the performance of the QDF model is significantly different from the performance of all other models. Note that, at the Bonferroni significance level of .01/55 and training samples of size 50, the MSD4 model is significantly different from the MSD5 and MSD2 models, but the MIP4 and HYB4 models are not significantly different from their corresponding model with the crossproduct term and from their corresponding first-order model. However, at the Bonferroni significance level of .05/55, the HYB4 and HYB2 models are significantly different in performance for both training samples of sizes 25 and 50 per group.

Configuration 1D

Configuration 1D is another data configuration with unequal variance-covariance structures for the two populations. The QDF model should perform optimally on these normally distributed populations. Since the off-diagonal terms of the matrix $\Sigma_1^{-1} - \Sigma_2^{-1}$ are the only non-zero elements, the crossproduct term is an important term in the QDF model for this data configuration. Also, the second-order mathematical programming models with the crossproduct term should outperform the other corresponding mathematical programming models on the validation samples. The results in Table 8 show that the best performing model on the validation samples for this data configuration is the QDF model for both training samples of sizes 25 and 50 per group. The average misclassification rates on the validation samples of the QDF model are 5.92% and 5.58% for training samples of sizes 25 and 50 per group, respectively. Also, on the training samples, the misclassification rate of the QDF model is lower than those of the LDF and hybrid models for training samples of size 25 per group, and lower than those of the LDF, hybrid, and MSD2 models for training samples of size 50 per group.

The mathematical programming models yield unexpected results. The best MSD and MIP models are the first-order models for both training samples of sizes 25 and 50 per group. The best hybrid model is HYB4 for training samples of size 25 per group and is HYB2 for training samples of size 50 per group. These results are surprising because it is expected that the crossproduct term would be necessary for the optimal classification model. Perhaps the squared terms in the second-order mathematical programming models are overfitting the data and, thus, underperform the corresponding first-order models. The best mathematical programming model for training samples of size 25 per group is the HYB4, which has an average misclassification rate of 6.64%. When the training sample size increases to 50 per group, the best mathematical programming model shifts to the MSD2, which has an average misclassification rate of 6.12%. Note that the LDF model's misclassification rate on the validation samples is lower than all of the mathematical programming models except the HYB4 model for training samples of size 25 per group. It is also lower than all of the mathematical programming models except the MSD2 model for training samples of size 50 per group.

The paired t-tests in Table 16 show a significant difference between the QDF model and all other models. The table also shows that the performance of the MSD2 and MIP2 models is significantly different from the performance of their corresponding second-order models for both training sample sizes. The HYB2 model's performance is significantly different from the other hybrid models only for training samples of size 50 per group.

Configuration 1E

Configuration 1E is another data configuration in which the variance-covariance structures of the two populations are unequal. However, the variance-covariance structures of this data configuration are different from those of configuration 1C and configuration 1D in that the correlation between observations is zero. The first variance-covariance structure of this data configuration is in the form of an identity matrix while the second variance-covariance structure is four times that of the first one. However, configuration 1E is similar to configuration 1C in that the off-diagonal terms of the matrix $\Sigma_1^{-1} - \Sigma_2^{-1}$ are zero. Again, the QDF model should perform optimally on the normally distributed populations of configuration 1E. However, the crossproduct term for the QDF model should not be important. It is expected that the second-order mathematical programming models without the crossproduct term should outperform the other corresponding mathematical programming models.

The results in Table 9 show that the best performing model on the validation samples for this data configuration is the QDF model, as expected, for both training samples of sizes 25 and 50 per group. The average misclassification rates on the validation samples of the QDF model are 7.20% and 6.66% for training samples of sizes 25 and 50 per group, respectively. For training samples of size 25 per group, the optimal classification models of the MSD and MIP models are MSD2 and MIP2, respectively. These results are not expected since the second-order mathematical programming models should perform better than the first-order mathematical programming models for this data configuration. However, the HYB5 does perform better than the HYB2.

The best performing mathematical programming model with training samples of size 25 per group is the MSD2. The average misclassification rate of the MSD2 model for training samples of size 25 per group is 8.33%. For training samples of size 50 per group, the best performing MSD, MIP, and hybrid models are MSD4, MIP2, and HYB4, respectively. The best performing mathematical programming models with training samples of size 50 per group is MSD4, which has an average misclassification rate of 7.44%. However, at the Bonferroni significance level of .05/55, the MSD4 and MSD2 models, the MIP4 and MIP2 models, and the HYB4 and HYB2 models are all not significantly different for training samples of size 50 per group as indicated by the paired t-tests in Table 17.

Interestingly, for training samples of size 25 per group, the HYB4 model performs worse than the HYB5 and HYB2 models, and is significantly different in performance from the HYB5 and HYB2 models. The results in Table 17 also reveal that the QDF model performs better than all other models and its performance is significantly different from all other models.

Configuration 1F

For configuration 1F, it is expected that the QDF model would perform well, whereas the LDF model would perform poorly since the means of the two populations are equal but the variance-covariance structures are not equal. It is also expected that the second-order mathematical programming models without the crossproduct term would outperform other corresponding mathematical programming models.

The results in Table 10 show that the best performing model on the validation samples for this data configuration is the QDF model for both training samples of sizes 25 and 50 per group. The average misclassification rates on the validation samples of the QDF model are 5.82% and 5.25% for training samples of sizes 25 and 50 per group, respectively. As expected, the LDF model does not perform well at all for this data configuration. The average misclassification rates on the validation samples of the LDF model are 39.38% and 41.29% for training samples of sizes 25 and 50 per group, respectively. In fact, the LDF model has the highest misclassification rate on the validation samples of all the models for training samples of size 25 per group.

The high overlap of the two populations makes the MIP models impractical to compute for training samples of size

50 per group. This is the only experimental situation in which the MIP models are not assessed on 100 replications of the data. The mathematical programming models yield results according to expectations. The second-order mathematical programming models without the crossproduct term outperform the other corresponding mathematical programming models for both training sample sizes. All of the first-order mathematical programming models perform poorly. The MSD4 model has the lowest misclassification rate for the mathematical programming models for both training sample sizes. The average misclassification rates on the validation samples for the MSD4 model are 7.95% and 6.16% for training samples of sizes 25 and 50 per group, respectively.

The paired t-tests in Table 18 indicate that for training samples of size 25 per group, the HYB4 and HYB5 models, and the MSD4 and MSD5 models are not significantly different in performance. For training samples of size 50 per group, the MSD4 and MSD5 models are not significantly different in performance. As expected, the QDF model clearly outperforms all other models. However, the addition of second-order terms to the mathematical programming models greatly improves their classificatory performance over the first-order mathematical programming models.

Configuration 1G

Configuration 1G is a data configuration that has a normal population for one group and a contaminated normal population for the other group. The population of the second group contains 15% of its observations as outliers. It is expected that the nonnormality of this data set would weaken the classificatory performance of the QDF model. It is also expected that the second-order mathematical programming models would outperform the first-order mathematical programming models.

Table 11 shows that all of the first-order models perform rather poorly relative to the second-order models. The average misclassification rates on the validation samples of the QDF model are 13.32% and 12.18% for training samples of sizes 25 and 50 per group, respectively, while those of the LDF model are 41.31% and 42.66%, respectively. However, the QDF model is not the best performing classification model for this data configuration. The best performing models are MSD4 and MSD5 for training samples of sizes 25 and 50 per group, respectively. The average misclassification rate on the validation samples for the MSD4 model with training samples of size 25 per group is 10.09% and that for the MSD5 model with training samples of size 50 per group is 8.95%. The mathematical programming models are capable of outperforming the QDF model when the data set contains outlier. The paired t-tests in Table 19 indicate that the performances of many second-order mathematical programming models are significantly different from the performance of the QDF model, particularly for training samples of size 50 per group. As seen in Table 19, the performance of the following pairs of mathematical programming models are not significantly different: MSD4 and MSD5 models, MIP4 and MIP5 models, and HYB4 and HYB5 models.

Configuration 1H

Configuration 1H is another data configuration that contains nonnormal data. The populations of this data configuration consist of both discrete and continuous data. It is expected that the nonnormality of this data would weaken the classificatory performance of the QDF and LDF models. It is also expected that the second-order mathematical programming models should outperform the first-order mathematical programming models. Since this data configuration can be perfectly separated by equation XY = 1, it is expected that the crossproduct term would be significant to the mathematical programming models. This is an example of a data set with no correlation between the variables, but the crossproduct term is still expected to be significant for the classification models.

The results in Table 12 show that both the LDF and QDF models perform poorly for this data configuration. The average misclassification rates for both the LDF and QDF models are around 30% on both training samples and validation samples. The nonnormality of the data clearly weakens the classificatory performance of the two parametric statistical models. Configuration 1H is clearly an example of a data configuration where the second-order mathematical programming model can perform dramatically better than the QDF model. For mathematical programming models, the second-order models outperform the first-order models. As expected, the second-order mathematical programming models with the crossproduct term outperform the models without the crossproduct term. With the exception of the hybrid models for training samples of size 25 per group, the results in Table 20 indicate that the performances of the second-order mathematical programming models with the crossproduct term and those of the corresponding second-order models without the crossproduct term are significantly different.

The best performing mathematical programming model for training samples of size 25 per group is MSD5 which has an average misclassification rate of 5.54% on the validation samples. When the training sample size increases to 50 per group, the best performing mathematical programming model is still the MSD5 model, which has an average misclassification rate of 2.91% on the validation samples. However, Table 20 indicates that the MSD5 model and the MIP5 model do not have significantly different performance. The MSD5 and MIP5 models can perfectly classify observations in the training samples of the two populations because the groups can be separated by the equation XY = 1. This data configuration shows that the crossproduct term may be important for a classification model despite the fact that the variables for each population are uncorrelated.

Simulation Results for Models with Secondary Goals

The results of a Monte Carlo simulation for MIP models with secondary goals are presented in this section. These results will be used to answer Research Question 2. Tables 23 through 50 and Figures 9 through 22 contain the results from the simulation study.

Configuration 2A

Configuration 2A is a configuration that contains two normal populations with equal variance-covariance structures. The results in Table 23 show that the best performing MIP model for this data configuration is the MIP1 model for both training samples of sizes 20 and 40 per group. The average misclassification rates of the MIP1 model are 3.42% and 3.08% for training samples of sizes 20 and 40 per group, respectively. The secondary goal of maximizing the distance between projected means in the MIP1 model seems to be effective in reducing the number of misclassification when compared with other secondary goals.

However, with the same secondary goal but bounded coefficients, the MIP2 model performs poorly. The constraint of bounded coefficients decreases the classificatory performance of the MIP2 model. The average misclassification rates of the MIP2 model are 6.14% and 4.25% for training samples of sizes 20 and 40 per group, respectively. The classificatory performances of the MIP3 and MIP4 models are almost the same. Thus, for this data configuration, the performances of the MIP3 and MIP4 models show that either maximizing the minimum internal deviation or minimizing the sum of the external deviations as a secondary goal in an MIP model will yield similar results.

Table 37 shows the results of paired t-tests for the mean difference in classificatory performance of the models for configuration 2A. The results reveal that the performance of the MIP1 model is significantly different from the performance of the other MIP models with significance level of .05/6 for both training samples of sizes 20 and 40 per group. The MIP3 and MIP4 models are not significantly different in performance for training samples of size 20 per group.

Configuration 2B

Configuration 2B is the same as configuration 2A, except that the data are rotated 45 degrees. Note that the distance between the means of the two populations is still the same. The results in Table 24 show that the MIP1 model is still the classification model of choice among the MIP models for training samples of size 20 per group. The average misclassification rates of the MIP1 model are 3.29% and 3.03% for training samples of sizes 20 and 40 per group, respectively, which are very close to those for configuration 2A. The performance of MIP4 model for this configuration is also very close to that for configuration 2A. It is interesting to see the MIP2 and MIP3 models perform much better in this configuration than in configuration 2A. The reason for this is the fact that the MIP2 and MIP3 models are not rotationally invariant. For training samples of size 40 per group, the MIP2 model performs as well as the MIP1 model.

From Table 38, the results of paired t-tests reveal that neither the MIP1 and MIP3 models, nor the MIP2 and MIP3 models are significantly different in performance for both training samples of sizes 20 and 40 per group. The performance of the MIP1 model is significantly different from the performance of the MIP2 model for training samples of size 20 per group.

Configuration 2C

Configuration 2C also contains two normal populations. However, the variance-covariance structures of the two populations are not equal. The variance-covariance structure of the first population is four times larger than that of the second population. Among the MIP models, the MIP1 model yields the lowest misclassification rate. As shown in Table 25, the average misclassification rates of the MIP1 model are 16.74% and 16.18% for training samples of sizes 20 and 40 per group, respectively. However MIP2 model, which has the same secondary goal as the MIP1 model but with bounded coefficients constraint, does not perform well for this data configuration. The MIP3 model performs nearly as well as the MIP4 model for both training samples of sizes 20 and 40 per group.

The results of paired t-tests in Table 39 reveal that the performance of the MIP1 model is significantly different from the performance of the other MIP models for training samples of size 20 per group. For training samples of size 40 per group, the performance of the MIP1 model is significantly different from the performance of the MIP2 and MIP3 models. The performance of the MIP3 model is not significantly different from that of the MIP4 model for both sizes of training samples.

<u>Configuration 2D</u>

For configuration 2D, the first population contains normal data, whereas the second population contains nonnormal data. Ten percent of the observations in the second population are contaminated by another normally distributed group of data. The results in Table 26 show that, among the MIP models, the MIP3 model yields the lowest misclassification rate for this data configuration. The secondary goal of maximizing the minimum internal deviation in the MIP3 model works well for this nonnormal data. The average misclassification rates of the MIP3 model are 8.40% and 7.48% for training samples of sizes 20 and 40 per group, respectively. For training samples of size 40 per group, the MIP4 model performs nearly as well as MIP3 model.

The results of paired t-tests in Table 40 reveal that the performance of the MIP2 model is significantly different from the performance of the other models for both sizes of training samples. However, none of the pairs of the MIP1, MIP3, and MIP4 models show any significant difference in performance for both training sample sizes.

Configuration 2E

Configuration 2E is a configuration that results from a 45 degrees rotation of configuration 2D. Table 27 shows results of the classification models for configuration 2E. These results are similar to the results from configuration 2B, in that the MIP2 model performs significantly better in the rotated data. Among the MIP models with training samples of size 40 per group, the MIP2 model yields the lowest misclassification rate. The average misclassification rate of the MIP2 model is 7.38% for training samples of size 40 per group. For training samples of size 20 per group, the MIP2 model performs nearly as well as the MIP3 model. The average misclassification rate of the MIP2 model is 8.19%, whereas that of the MIP3 model is 8.16% for training samples of size 20 per group. From Table 41, the results of paired t-tests reveal that most of the performances of the four MIP models are not significantly different from each other. However, the MIP1 and MIP2 models for training samples of size 40 per group and the MIP3 and MIP4 models for both training sample sizes are each significantly different in performance.

Configuration 2F

Configuration 2F contains contaminated normal data for both populations. The contaminating fraction is 10% for both populations. The results in Table 28 show that the best performing MIP model for training samples of size 20 per group is the MIP1 model. The average misclassification rate of the MIP1 model is 5.47% for training samples of size 20 per group. Among the MIP models with training samples of size 40 per group, the MIP1, MIP3, and MIP4 models perform almost the same. The average misclassification rate of the MIP4 model is 4.61%, whereas those of the MIP1 and the MIP3 models are 4.65% and 4.69%, respectively, for training samples of size 40 per group. The MIP2 model performs poorly for this data configuration.

The results of paired t-tests in Table 42 indicate that the performance of the MIP2 model is significantly different from the performance of the MIP1, MIP3, and MIP4 models for both sizes of the training samples. The MIP1, MIP3, and MIP4 models are not significantly different in performance.

Configuration 2G

Configuration 2G is the configuration that results from a 45 degrees rotation of configuration 2F. Again, the results in Table 29 show a significant improvement of the MIP2 model with this rotated data. Among the MIP models with training samples of size 40 per group, the MIP2 model yields the lowest misclassification, which is 4.50%. For training samples of size 20 per group, the best performing MIP model is the MIP1 model, which yields an average misclassification rate of 5.18%. The results of the paired t-tests in Table 43 reveal that the performances of the MIP1, MIP2, and MIP3 models are not significantly different from each other for both training sample sizes. For training samples of size 20 per group, the performance of the MIP3 model is significantly different from that of the MIP4 model.

Configuration 2H

Configuration 2H is another configuration that contains contaminated normal data in both populations. The contaminating fraction is 10% for both populations. The results in Table 30 show that, among the MIP models, the MIP1 model yields the lowest misclassification rate for both sizes of the training samples. The average misclassification rates of the MIP1 model are 2.22% and 1.94% for training samples of sizes 20 and 40 per group, respectively. The MIP2 model performs poorly for this configuration. The MIP4 model performs nearly as well as the MIP3 model for training samples of size 40 per group. The results of paired t-tests in Table 44 indicate that the performances of all MIP models are significantly different from each other for both sizes of the training samples, except for the MIP3 and MIP4 models in the case of training samples of size 40 per group.

Configuration 2I

The data in configuration 2I are similar to those in configuration 2H, except that the contaminating fraction is increased to 20% for both populations. The results of this configuration are similar to those of configuration 2H. The results in Table 31 show that the MIP1 model is still the best among the MIP models for both sizes of the training samples. The average misclassification rates of the MIP1 model are 3.37% and 2.92% for training samples of sizes 20 and 40 per group, respectively. The MIP4 model performs nearly as well as the MIP3 model. As shown in Table 45, the results of the paired t-tests indicate that the performances of all the MIP models, except the MIP3 and MIP4 models, are significantly different from each other for both sizes of training samples.

Configuration 2J

For configuration 2J, the first population contains normal data, whereas the second population contains contaminated normal data. The contaminating fraction of the second population is 10%. The results in Table 32 show that, among the MIP models, the MIP3 model yields the lowest misclassification. However, the MIP4 model performs as well as the MIP3 model for training samples of size 40 per group. The average misclassification rates of the MIP3 model are 8.77% and 8.15% for training samples of sizes 20 and 40 per group, respectively. The results of the paired t-tests in Table 46 reveal that the performance of the MIP3 model is significantly different from the performance of the MIP1 and MIP2 models for both training samples of sizes 20 and 40 per group. However, the MIP3 model's performance is not significantly different from that of the MIP4 model.

Configuration 2K

Configuration 2K is similar to configuration 2J, except that the contaminating fraction of the second population is increased to 20%. The results for this configuration are similar to those for configuration 2J. Table 33 shows that, among the MIP models, the MIP3 model still has the lowest misclassification rate for both sizes of the training samples. The average misclassification rates of the MIP3 model are 13.59% and 12.99% for training samples of sizes 20 and 40 per group, respectively. The MIP4 model performs nearly as well as the MIP3 model. The results of paired t-tests in Table 47 indicate that the performance of the MIP3 model is significantly different from the performance of the MIP1 and MIP2 models for both training samples of sizes 20 and 40 per group. However, there is no significant difference in the performance between the MIP3 and MIP4 models.

Configuration 2L

Configuration 2L is the configuration chosen to have a low value of skewness and a high value of kurtosis. The values of skewness and kurtosis are chosen to be 0.461 and 13.419, respectively, for both populations. From these specified values of skewness and kurtosis, the means and variance-covariance structures of the two populations were obtained from the results generated on the contaminated normal distribution in the next section of this chapter and are presented in Table 22. The results in Table 34 show that, among the MIP models, the MIP3 and MIP4 models both yield an average misclassification rate of 13.44%, which is the lowest misclassification for training samples of size 20 per group. For training samples of size 40 per group, the best performing MIP model is the MIP1 model which has average misclassification rate of 12.40%. The MIP2 model does not perform well for this data configuration. The results of paired t-tests in Table 48 reveal that the performances of the MIP1, MIP3, and MIP4 models are not significantly different from each other for both sizes of the training samples.

Configuration 2M

Configuration 2M is the configuration chosen to have moderate values of skewness and kurtosis. The values of skewness and kurtosis are chosen to be 1.625 and 7.612, respectively, for both populations. From these specified values of skewness and kurtosis, the means and variance-covariance structures of the two populations were obtained from the results generated on the contaminated normal distribution in the next section of this chapter and are presented in Table 22. The results in Table 35 show that, among the MIP models, the MIP4 model yields the lowest misclassification for both sizes of the training samples. The average misclassification rates of the MIP4 model are 9.56% and 8.53% for training samples of sizes 20 and 40 per group, respectively. Table 49 shows that the results of paired t-tests that are similar to those of configuration 2L. The results indicate that the performances of the MIP1, MIP3, and MIP4 models are not significantly different from each other for both sizes of training samples.

Configuration 2N

Configuration 2N is the configuration chosen to have a low value of skewness and a very low value of kurtosis. The values of skewness and kurtosis are chosen to be 0.129 and 2.124, respectively, for both populations. From these specified values of skewness and kurtosis, the means and variance-covariance structures of the two populations were obtained from the results generated on the contaminated normal distribution in the next section of this chapter and are presented in Table 22. The results in Table 36 show that, among the MIP models, the MIP1 model yields the lowest misclassification rate for both sizes of the training samples. The average misclassification rates of the MIP1 model are 7.10% and 6.37% for training samples of sizes 20 and 40 per group, respectively. However, the results of paired t-tests in Table 50 show that the performances of the MIP3 and MIP4 models are not significantly different from that of the MIP1 model for training samples of size 20 per group.

Skewness and Kurtosis Measures for the Contaminated Normal Distribution

The general contaminated multivariate normal distribution can be written as

$$CMN(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \epsilon) = (1 - \epsilon)N(\mu_1, \sigma_1^2) + \epsilon N(\mu_2, \sigma_2^2).$$

The notation $N(\mu, \sigma^2)$ represents the normal distribution with mean μ and variance σ^2. The $N(\mu_2, \sigma_2^2)$ population can be interpreted as the contaminating part, and ϵ can be interpreted as the contaminating fraction of the data. The above distribution can be shifted and scaled (such that μ_1 = 0 and σ_1^2 = 1) so that the cumulative distribution function is

$$F(X) = (1-\epsilon)\Phi(X) + \epsilon\Phi((X-\mu)/\sigma) .$$

As shown in the theoretical framework chapter of this dissertation, the formulas for the skewness (γ_1) and the kurtosis (γ_2) measures can be mathematically derived as

$$\gamma_1 = \frac{\epsilon\mu(1-\epsilon)(3\sigma^2 + \mu^2 - 2\mu^2\epsilon - 3)}{[1 - \epsilon + \epsilon\sigma^2 + \mu^2\epsilon(1 - \epsilon)]^{3/2}}$$

$$\gamma_2 = \frac{6\epsilon\mu^2\sigma^2(\epsilon^2-2\epsilon+1)+\epsilon\mu^4(1+6\epsilon^2-3\epsilon^2-4\epsilon)+6\mu^2\epsilon^2(1-\epsilon)+3(1-\epsilon)+3\epsilon\sigma^4}{[1 - \epsilon + \epsilon\sigma^2 + \mu^2\epsilon(1 - \epsilon)]^2}$$

Now, if μ approaches infinity, the kurtosis measure would have a limiting value of

$$\frac{1 + 6\epsilon^2 - 3\epsilon^3 - 4\epsilon}{\epsilon(1-\epsilon)^2} = \frac{3\epsilon^2 - 3\epsilon + 1}{\epsilon(1-\epsilon)} = -3 + \frac{1}{\epsilon(1-\epsilon)}$$

If ϵ is equal to 0.5, then the limiting value of the kurtosis is 1 as the parameter μ approaches infinity. The value of one for the kurtosis is the smallest value that the kurtosis can have. Also note that as σ approaches infinity, the kurtosis has a limiting value of $3/\epsilon$. Thus, ϵ can be chosen to give any desired limiting value for the kurtosis.

For the measure of skewness, the limiting value as μ approaches infinity is

$$\frac{1 - 2\epsilon}{[\epsilon(1-\epsilon)]^{1/2}}$$

which is equal to zero for $\epsilon = .5$ and approaches infinity when ϵ becomes close to zero. Thus, there is a wide range of values that can be specified for the skewness and kurtosis measures in simulating contaminated normal data. Tables 51 through 58 contain various values of the skewness and kurtosis measures for ϵ = .01, .05, .10, .15, .20, .30, .40, and .50 with various settings of the parameters μ and σ.

Note that for the contaminated normal distribution with the contaminating fraction higher than 0.50, the distribution will be similar to the one with contaminating fraction equals to $1-\epsilon$. For example, if a contaminated normal distribution with contaminating fraction equals 0.60 and specific values of the skewness and kurtosis measures is desired, then one can simply use the table with ϵ = 0.40 to select the parameter settings for the distribution.

CHAPTER VI

CONCLUSIONS

Research Questions Addressed

This study has addressed three research questions regarding the effects of certain modifications to the mathematical programming models for solving the statistical classification problem and the appropriateness of using the contaminated normal distribution in Monte Carlo simulation studies.

Research Question 1

> How do second-order terms in mathematical programming models affect the performance of certain two-group classification models for small to moderate training sample sizes and for normal and nonnormal data? Can the correlation structure of the data determine whether the crossproduct terms should be included in the models? Under what conditions are these models invariant with respect to translation and rotation of the data?

From the results of simulation study, second-order terms in mathematical programming models can be very effective in correctly classifying observations for certain data configurations. For certain data configurations in which the data are highly nonnormal, including the second-order terms in mathematical programming models greatly improves the classification results over the first-order models and the Smith's quadratic discriminant method. Also, when the variance-covariance structures of the two populations are different, the second-order mathematical programming models can easily outperform the first-order models. However, particularly for a small sample size, it is possible for the first-order models to outperform the second-order models when the variance-covariance structures are only slightly different.

The correlation structure of the data can sometimes determine the need of the crossproduct term for mathematical programming models. If the sample size is moderate to large and the data are approximately normal, then the crossproduct term should not be included in the mathematical programming model for data configurations such that $\Sigma_1^{-1} - \Sigma_2^{-1}$ is strictly a diagonal matrix (where Σ_1 and Σ_2 are the variance-covariance matrices of the first and second populations, respectively). For a small sample size, second-order terms may reduce the classificatory performance of some mathematical programming models even if the variance-covariance matrices of the populations differ. For nonnormal data, the correlation structure may not determine the need for a crossproduct term. For example, the data may be uncorrelated but perfectly separable by the equation XY = constant, where X and Y are the two attribute variables. In this case, the crossproduct term can significantly improve the classificatory performance of the mathematical programming models despite the independence of the attribute variables. Figure 23 displays guideline for alternative mathematical programming models. To guarantee that the second-order mathematical programming models are both translationally invariant and rotationally invariant, all of the first-order and second-order terms must be included in the models. Omitting the crossproduct term, for example, may improve the performance of the model, but the model may not be optimal after a rotation.

Research Question 2

> Can the use of certain secondary goals improve the performance of MIP models for the two-group classification problem on small to moderate sample sizes?

The use of certain secondary goals can improve the performance of the MIP models. An appropriate secondary goal for an MIP model depends on the characteristic and orientation of the data. From the results of the simulation study, the secondary goal of maximizing the distance between the means of the discriminant scores is appropriate mostly for configurations in which both populations have the same distribution and the line between the two population means is approximately parallel to the horizontal axis. However, if this type of data configuration is rotated 45 degrees, then the same secondary goal with constraints to bound the coefficients would be more effective for the MIP model in classifying observations. For contaminated normal configurations with the two population distributions being very different, maximizing the minimum deviation of the correctly classified observations would be an appropriate secondary goal for the MIP model.

The secondary goal of minimizing the sum of all the misclassified observations' deviations is appropriate for the contaminated normal data with moderate values of skewness and kurtosis measures. However, if the contaminated normal data have low values for the skewness and kurtosis measures, then maximizing the distance between the means of the discriminant scores would be an appropriate secondary goal for the MIP model.

<u>Research Question 3</u>

> Since the contaminated normal distribution (mixture of two normals) can be used to assess the performance of linear discriminant functions without a validation sample, how appropriate is this distribution

for a simulation study in generating nonnormal data with a variety of values for the skewness and kurtosis measures? In particular, what range of values for the measures of skewness and kurtosis can the contaminated normal distributions have by using different parameter settings for the mean, standard deviation, and contaminating fraction?

This study shows the usefulness of a general contaminated multivariate normal distribution in estimating misclassification probabilities in a simulation study which investigates various classification models. The contaminated normal distribution is appropriate for a simulation study in generating nonnormal data. A wide range of values can be assigned to the measures of skewness and kurtosis when generating contaminated normal distribution by using different parameter settings for the mean (μ), standard deviation (σ), and contaminating fraction (ϵ).

The results on the contaminated normal distribution show that the limiting values of the skewness and kurtosis measures when μ approaches infinity are $(1-2\epsilon)/[\epsilon(1-\epsilon)]^{1/2}$ and $-3 + 1/\epsilon(1-\epsilon)$, respectively. Therefore, if ϵ equals 0.50 and μ approaches infinity, then the values of the skewness and kurtosis measures will approach 0 and 1, respectively. Note that the smallest value of the kurtosis measure for the contaminated normal distribution is 1.

However, the kurtosis measure will have a limiting value of $3/\epsilon$ as σ approaches infinity. When ϵ becomes close to zero and the value of μ is sufficiently large, the value of the skewness measure will approach infinity.

Tables illustrating various values of the skewness and kurtosis measures for the contaminated normal distribution with values of μ, σ, and ϵ help to identify contaminated normal distributions that approximate nonnormal distributions with certain skewness and kurtosis values. Thus, using the contaminated normal distribution in simulation studies allows for greater use of distributions that approximate certain real-world data sets with similar values for the measures of skewness and kurtosis.

Limitations and Key Assumptions

Limitations and keys assumptions of this study include the following:

1. Only the two-group classification problem is considered in this research study. It is common to find classification problems involving more than two groups. Although the extension of classification models to more than two groups is conceptually straightforward, different mathematical programming models would be needed. The results on the inclusion of second-order terms and the use of secondary goals may not be easily generalized to the multiple group discriminant problem.

2. The training sample is limited to small to moderate sizes (20 to 50 observations for each group). The simulation study does not compare the performance of the classification models with higher sample sizes. This is due to the computational intensiveness of the MIP procedures at higher sample sizes, particularly for the data in which the degree of overlap in the groups is large.

3. Only the MSD, MIP, and hybrid models of mathematical programming-based formulations are included in this study. Although these models are found to compare favorably with the parametric statistical models, other mathematical programming models and nonparametric models that have been presented in the literature have shown some potential for good classificatory performance under certain data configurations.

4. The simulation study is limited to only data configurations that are presented in the Simulation Designs chapter of this dissertation. The results may not necessarily extend to other data configurations. The simulation study includes mostly normal and contaminated normal data. Although this type of data represents real-world data, there are countless possibilities for data configurations.

5. This study considers only attribute variables with first-order and second-order terms. Some data configurations in which a nonlinear discriminant function is

the optimal classification rule may require terms that are perhaps higher than the second-order in the discriminant function for optimality.

6. The prior probability of an observation coming from either population is assumed to be equal. The cost of assigning an observation to one population when, in fact, it belongs to the other population, is considered to be equal for all observations.

Future Directions for Research

Many issues related to the study in this dissertation can be investigated in future research studies.

1. Although the results in this dissertation show benefits from inclusion of second-order terms in mathematical programming approaches to discriminant analysis for the two-group problem, the usefulness of second-order terms for the classification problem with more than two groups needs to be investigated.

2. This dissertation compares the classificatory performance of MIP models with four different secondary goals. There are other secondary goals that can be evaluated.

3. The sizes of training samples and the characteristics of data configurations other than the ones used in this dissertation can be explored in simulation studies.

4. The study comparing classificatory performance of the parametric statistical methods and the mathematical programming methods can be extended to classification problems with unequal prior probabilities and/or unequal costs of misclassification.

5. Further examination of other modifications to mathematical programming approaches may yield benefits to practitioners by having greater flexibility in choosing an appropriate model.

Major Contribution of the Research

The results from this study will assist practitioners and decision-makers in understanding and implementing improved versions of mathematical programming formulations and will give them greater flexibility in choosing appropriate models to solve the statistical classification problem. Previous simulation studies have shown that the MSD and MIP models can perform well in the presence of nonnormal data (Stam and Jones 1990). However, the inclusion of second-order terms of the attribute variables in these mathematical programming formulations gives these models the potential to be very competitive with Smith's quadratic discriminant method, which involves both first-order and second-order terms. The condition for rotational and translational invariance will help practitioners to understand the effect of omitting terms to obtain a parsimonious model.

The results of the simulation study reveal that the success exhibited by Rubin's (1990a) MIP model with a secondary goal in his limited simulation study is shared by MIP models with other secondary goals for certain data configurations. Some secondary goals may be appropriate with only certain types of data configurations. Not all of the MIP models with secondary goals are rotationally invariant. An appropriately selected secondary goal can improve the classificatory performance of the MIP model and make the model more competitive to both the parametric statistical procedures and the mathematical programming-based models.

The formulas for the measures of skewness and kurtosis for the general contaminated normal distribution were derived. For contaminated normal data, the measures of skewness and kurtosis are generally not available. However, the results in this dissertation show that a wide range of values for the measures of skewness and kurtosis are possible with contaminated normal distribution. These results make the contaminated normal distribution useful in simulating nonnormal data with various values of the skewness and kurtosis measures.

Managerial decision-makers can easily implement the mathematical programming models in this study by using a standard optimization computer package such as SAS/OR and LINDO. The results of this study allow the managerial decision-makers to use improved versions of mathematical programming formulations for the discriminant problem by utilizing second-order terms and appropriate secondary goals. When violations of the usual parametric assumptions are severe, these formulations provide alternative classification methods.

APPENDIX A

TABLES

Table 3.--Classification Models for Research Question 1

Models	Descriptions
1. MSD5	MSD with all linear, squared, and crossproduct terms (5 variables)
2. MSD4	MSD with linear and squared terms (4 variables)
3. MSD2	MSD with only linear terms (2 variables)
4. MIP5	MIP with all linear, squared, and crossproduct terms (5 variables)
5. MIP4	MIP with linear and squared terms (4 variables)
6. MIP2	MIP with only linear terms (2 variables)
7. HYB5	Hybrid with all linear, squared, and crossproduct terms (5 variables)
8. HYB4	Hybrid with linear and squared terms (4 variables)
9. HYB2	Hybrid with only linear terms (2 variables)
10. LDF	Fisher's Linear Discriminant Function
11. QDF	Smith's Quadratic Discriminant Function

Table 4.--Data Configurations for Research Question 1

Configuration	First Population		Second Population	
	Mean Vector	Covariance Matrix	Mean Vector	Covariance Matrix
1A	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
1B	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$
1C	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 4 \\ 4 & 20 \end{bmatrix}$	$\begin{bmatrix} 3.5 \\ 3.5 \end{bmatrix}$	$\begin{bmatrix} 4.47 & 4 \\ 4 & 4.47 \end{bmatrix}$
1D	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$
1E	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$
1F	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 49 & 0 \\ 0 & 49 \end{bmatrix}$
1G	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
			15% of observations from	
			$\begin{bmatrix} -10 \\ -10 \end{bmatrix}$	$\begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$
1H	a_1 from Uniform(0.1, 5.0)		a_1 from Uniform(0.1, 5.0)	
	a_2 from Uniform(0, $\frac{1}{a_1}$)		a_2 from Uniform($\frac{1}{a_1}$, $\frac{1}{a_1}$+.5)	
	20% of observations from		20% of observations from	
	$a_1 = a_2 = -4.60894$		$a_1 = a_2 = 4.195634$	

Table 5.--Percentages of Misclassified Observations for Training Samples of Sizes 25 and 50 Per Group for Configuration 1A

Method	$n_1 = n_2 = 25$				$n_1 = n_2 = 50$			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
MSD5	5.32	3.68	11.68	2.76	6.53	2.40	9.58	1.71
MSD4	5.76	3.77	10.43	2.08	6.66	2.30	9.10	1.29
MSD2	6.64	3.90	8.84	1.34	7.17	2.37	8.40	0.90
MIP5	3.16	2.38	14.66	4.11	4.36	1.74	11.64	2.23
MIP4	3.50	2.48	12.90	3.12	4.50	1.74	10.81	2.16
MIP2	4.28	2.57	9.86	2.30	5.11	1.87	9.14	1.22
HYB5	8.88	3.69	10.98	2.21	7.68	2.90	10.46	3.11
HYB4	8.42	3.37	10.11	1.96	7.50	2.46	10.42	2.36
HYB2	7.48	3.20	8.63	1.15	6.93	2.49	8.62	1.41
LDF	7.10	3.23	8.36	1.03	7.33	2.44	8.13	0.87
QDF	6.90	3.25	8.62	1.11	7.27	2.44	8.24	0.85

Table 6.--Percentages of Misclassified Observations for Training Samples of Sizes 25 and 50 Per Group for Configuration 1B

Method	$n_1 = n_2 = 25$				$n_1 = n_2 = 50$			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
MSD5	2.54	2.74	9.43	3.43	3.31	1.87	6.43	1.50
MSD4	2.86	2.89	8.48	3.00	3.77	1.96	5.78	1.33
MSD2	3.42	2.96	5.82	1.58	3.89	1.86	5.10	0.80
MIP5	1.48	1.57	11.06	3.27	2.00	1.24	8.09	2.18
MIP4	1.92	1.80	10.10	3.38	2.34	1.29	7.15	1.70
MIP2	2.12	1.86	6.70	2.14	2.62	1.40	5.64	1.17
HYB5	6.42	3.21	8.15	2.20	4.73	2.24	7.65	2.51
HYB4	6.84	3.14	8.06	2.07	4.89	2.02	7.20	1.90
HYB2	4.88	2.78	5.18	1.05	3.82	1.81	5.37	1.00
LDF	4.40	2.90	4.98	0.87	4.27	1.86	4.74	0.69
QDF	4.10	2.80	5.29	1.01	4.23	1.86	4.86	0.69

Table 7.--Percentages of Misclassified Observations for Training Samples of Sizes 25 and 50 Per Group for Configuration 1C

Method	$n_1 = n_2 = 25$				$n_1 = n_2 = 50$			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
MSD5	3.20	3.35	11.02	3.17	4.92	2.65	7.64	1.62
MSD4	3.68	3.74	9.53	2.50	5.23	2.71	7.21	1.24
MSD2	7.62	4.03	9.02	1.56	7.68	2.60	8.49	1.11
MIP5	1.72	1.78	12.40	3.15	2.90	1.57	9.38	1.96
MIP4	2.24	2.04	11.44	2.92	3.16	1.63	8.76	1.93
MIP2	5.90	3.33	10.74	2.68	6.19	2.29	9.15	1.71
HYB5	7.96	3.27	8.53	1.57	6.68	5.15	9.01	4.84
HYB4	8.24	3.19	8.74	1.50	6.47	4.67	8.84	4.69
HYB2	11.88	3.47	12.54	2.41	8.91	3.36	10.68	2.25
LDF	10.50	3.31	11.15	2.03	10.34	2.19	10.81	1.35
QDF	6.22	3.57	6.88	0.97	6.14	2.45	6.47	0.94

Table 8.--Percentages of Misclassified Observations for Training Samples of Sizes 25 and 50 Per Group for Configuration 1D

Method	$n_1 = n_2 = 25$				$n_1 = n_2 = 50$			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
MSD5	2.54	2.75	9.75	3.26	4.30	2.19	6.99	1.46
MSD4	3.40	3.01	8.34	2.33	4.60	2.13	6.58	1.16
MSD2	3.98	3.07	6.81	1.57	5.09	2.19	6.12	0.75
MIP5	1.60	1.58	11.97	3.50	2.78	1.42	8.97	2.14
MIP4	2.20	1.92	10.35	3.46	3.16	1.66	8.09	2.28
MIP2	2.78	2.11	7.66	1.96	3.81	1.84	6.81	1.22
HYB5	5.76	2.94	7.21	1.90	5.73	2.59	8.57	3.23
HYB4	5.68	2.94	6.64	1.33	5.75	2.78	8.12	2.63
HYB2	6.30	2.91	7.06	1.48	5.40	2.47	6.66	1.28
LDF	5.34	2.68	6.77	1.18	5.99	2.21	6.48	0.90
QDF	4.64	2.65	5.92	0.91	5.05	2.13	5.58	0.66

Table 9.--Percentages of Misclassified Observations for Training Samples of Sizes 25 and 50 Per Group for Configuration 1E

Method	$n_1 = n_2 = 25$				$n_1 = n_2 = 50$			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
MSD5	3.88	3.24	10.18	2.85	4.83	1.99	7.90	1.84
MSD4	4.34	3.42	9.06	2.30	4.97	2.05	7.44	1.40
MSD2	6.46	3.91	8.33	1.26	6.36	2.20	7.85	0.90
MIP5	2.34	1.84	12.45	3.57	3.06	1.35	9.52	2.38
MIP4	2.64	2.05	11.35	3.55	3.31	1.47	9.11	2.28
MIP2	4.30	2.61	9.43	2.12	4.64	1.68	8.55	1.40
HYB5	7.10	3.29	8.41	2.18	6.56	2.48	9.46	2.49
HYB4	9.18	3.75	10.72	2.69	6.29	2.43	8.76	2.05
HYB2	8.80	3.36	9.34	1.78	6.98	2.42	8.86	1.51
LDF	8.00	3.36	8.50	1.35	7.59	2.13	8.36	1.07
QDF	5.92	3.14	7.20	1.13	5.61	1.86	6.66	0.99

Table 10.--Percentages of Misclassified Observations for Training Samples of Sizes 25 and 50 Per Group for Configuration 1F

Method	$n_1 = n_2 = 25$				$n_1 = n_2 = 50$			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
MSD5	2.32	2.75	8.67	2.94	3.29	2.01	6.39	1.24
MSD4	2.66	2.98	7.95	2.50	3.56	2.42	6.16	1.53
MSD2	31.89	4.81	35.93	2.71	34.58	3.93	37.64	2.54
MIP5	1.76	2.07	10.58	3.70	*	*	*	*
MIP4	1.96	2.16	9.29	3.08	*	*	*	*
MIP2	21.56	3.00	32.80	1.89	*	*	*	*
HYB5	17.34	3.64	20.34	3.27	5.06	2.07	7.97	1.82
HYB4	17.76	3.54	19.89	3.36	4.72	2.44	7.30	1.85
HYB2	29.16	6.96	34.85	4.98	43.86	6.34	48.05	3.84
LDF	34.46	7.62	39.38	4.80	37.13	6.43	41.29	4.34
QDF	3.94	2.95	5.82	1.06	4.14	1.81	5.25	0.97

* Computationally too intensive to complete runs for this model.

Table 11.--Percentages of Misclassified Observations for Training Samples of Sizes 25 and 50 Per Group for Configuration 1G

Method	$n_1 = n_2 = 25$				$n_1 = n_2 = 50$			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
MSD5	4.48	3.62	12.06	3.30	5.97	2.43	8.95	1.37
MSD4	5.64	4.01	10.09	2.93	6.57	2.27	9.13	1.44
MSD2	25.16	12.80	30.64	13.94	26.28	12.17	29.65	12.85
MIP5	2.82	2.36	14.40	4.15	4.00	1.75	10.63	2.07
MIP4	3.58	2.59	13.63	3.20	4.43	1.82	10.25	1.70
MIP2	11.68	4.62	17.27	2.39	12.05	2.82	15.83	1.29
HYB5	8.56	3.95	11.97	2.55	7.64	2.76	10.36	1.75
HYB4	8.92	3.79	11.47	2.12	7.97	3.25	10.55	2.66
HYB2	30.80	19.67	36.93	20.57	38.98	21.21	42.48	20.75
LDF	34.66	12.15	41.31	12.65	38.50	11.60	42.66	11.49
QDF	10.16	5.41	13.32	4.27	11.12	4.10	12.18	2.58

Table 12.--Percentages of Misclassified Observations for Training Samples of Sizes 25 and 50 Per Group for Configuration 1H

Method	$n_1 = n_2 = 25$				$n_1 = n_2 = 50$			
	Training Sample		Validation Sample		Training Sample		Validation Sample	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
MSD5	0	0	5.54	2.83	0	0	2.91	1.47
MSD4	2.62	3.40	8.64	2.71	5.13	3.48	7.33	1.71
MSD2	11.44	7.47	14.49	6.33	11.50	5.75	12.96	5.18
MIP5	0	0	5.58	2.93	0	0	2.94	1.43
MIP4	1.14	1.31	8.25	2.74	2.03	1.38	6.71	1.74
MIP2	5.22	3.07	10.20	2.93	6.19	2.39	9.04	1.66
HYB5	10.80	4.93	15.36	6.86	3.72	4.41	6.45	4.70
HYB4	11.30	5.66	15.50	7.58	7.50	4.29	9.34	2.83
HYB2	17.08	7.18	20.53	9.01	13.33	5.39	14.83	4.27
LDF	27.48	5.87	30.89	4.30	28.14	4.39	29.98	3.48
QDF	29.44	7.11	31.46	5.23	30.97	5.08	32.20	3.90

Table 13.--Paired T-Tests of Mean Difference in Classification Performance on Validation Samples for Training Samples of Sizes 25 and 50 Per Group for Configuration 1A

Method	Method									
	MSD4	MSD2	MIP5	MIP4	MIP2	HYB5	HYB4	HYB2	LDF	QDF
MSD5										
n_i=25	5.99	11.49	-8.18	-3.51	5.92	2.11	5.24	10.44	12.21	11.75
n_i=50	3.72	8.24	-8.84	-5.93	2.73	-2.82	-3.52	5.19	8.82	8.54
MSD4										
n_i=25		8.68	-10.51	-8.35	2.23	-2.06	1.23	8.48	10.59	9.68
n_i=50		7.77	-11.84	-8.63	-0.26	-4.51	-6.30	3.29	8.37	8.29
MSD2										
n_i=25			-14.36	-13.48	-4.65	-9.30	-6.18	1.65	4.50	2.20
n_i=50			-14.66	-11.60	-6.42	-6.68	-8.95	-1.84	4.09	2.80
MIP5										
n_i=25				4.30	10.84	8.53	10.80	14.47	15.53	15.52
n_i=50				3.92	11.76	3.39	4.35	12.45	15.13	14.89
MIP4										
n_i=25					10.15	5.59	7.98	13.64	14.69	13.70
n_i=50					8.52	1.07	1.75	9.86	12.41	12.18
MIP2										
n_i=25						-4.14	-0.99	5.26	6.66	5.61
n_i=50						-4.07	-5.38	3.41	8.00	7.33
HYB5										
n_i=25							3.86	10.40	12.19	11.13
n_i=50							0.12	5.84	7.49	7.28
HYB4										
n_i=25								8.00	10.17	8.64
n_i=50								8.24	10.08	9.87
HYB2										
n_i=25									4.05	0.11
n_i=50									4.07	3.34
LDF										
n_i=25										-4.50
n_i=50										-3.43

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 55 tests, i.e., the individual computed t-value must be significant at α = .05/55, resulting in a critical value of 3.368. The above t-values can also be compared to the critical value of 3.815, which is the critical value obtained from the Bonferroni method with a significance level of .01/55.

Table 14.--Paired T-Tests of Mean Difference in Classification Performance on Validation Samples for Training Samples of Sizes 25 and 50 Per Group for Configuration 1B

Method	Method									
	MSD4	MSD2	MIP5	MIP4	MIP2	HYB5	HYB4	HYB2	LDF	QDF
MSD5										
n_i=25	3.02	11.35	-4.90	-1.68	8.10	3.71	3.97	12.35	13.53	13.10
n_i=50	6.64	10.31	-7.60	-4.23	5.30	-5.69	-4.58	6.99	11.87	11.31
MSD4										
n_i=25		10.66	-7.14	-5.10	5.90	0.99	1.32	11.39	12.49	11.66
n_i=50		6.30	-10.20	-9.12	1.15	-7.95	-9.65	3.25	8.89	8.05
MSD2										
n_i=25			-15.81	-13.59	-5.00	-8.87	-9.31	4.11	6.54	3.99
n_i=50			-14.38	-13.34	-5.36	-10.61	-12.00	-3.33	6.32	4.19
MIP5										
n_i=25				2.88	12.85	7.44	7.44	17.29	19.11	19.09
n_i=50				4.23	11.99	1.74	3.75	13.15	16.21	15.87
MIP4										
n_i=25					12.00	4.92	5.38	14.80	15.97	15.26
n_i=50					10.37	-1.84	-0.25	11.42	15.10	14.28
MIP2										
n_i=25						-5.10	-5.07	6.92	8.78	7.20
n_i=50						-8.06	-8.34	2.49	8.22	6.90
HYB5										
n_i=25							0.49	12.44	13.32	12.60
n_i=50							1.94	9.65	11.93	11.74
HYB4										
n_i=25								14.79	15.11	13.91
n_i=50								10.82	13.84	13.59
HYB2										
n_i=25									3.03	-1.14
n_i=50									6.78	5.76
LDF										
n_i=25										-5.73
n_i=50										-3.68

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 55 tests, i.e., the individual computed t-value must be significant at α = .05/55, resulting in a critical value of 3.368. The above t-values can also be compared to the critical value of 3.815, which is the critical value obtained from the Bonferroni method with a significance level of .01/55.

Table 15.--Paired T-Tests of Mean Difference in Classification Performance on Validation Samples for Training Samples of Sizes 25 and 50 Per Group for Configuration 1C

Method	Method									
	MSD4	MSD2	MIP5	MIP4	MIP2	HYB5	HYB4	HYB2	LDF	QDF
MSD5										
n_i=25	5.22	7.04	-3.87	-1.23	0.75	7.43	6.63	-3.84	-0.35	12.94
n_i=50	4.68	-5.73	-9.64	-5.58	-8.11	-2.87	-2.65	-12.28	-17.48	8.68
MSD4										
n_i=25		1.98	-8.75	-6.11	-3.39	3.98	2.89	-8.92	-5.61	11.36
n_i=50		-12.11	-12.99	-8.64	-12.20	-3.84	-3.66	-15.82	-24.56	7.83
MSD2										
n_i=25			-9.95	-8.80	-6.90	3.01	1.68	-13.08	-10.33	14.65
n_i=50			-4.91	-1.50	-4.71	-1.09	-0.75	-11.02	-19.89	27.66
MIP5										
n_i=25				2.97	4.09	11.77	10.47	-0.37	3.41	17.67
n_i=50				3.64	1.12	0.73	1.10	-5.06	-6.77	16.48
MIP4										
n_i=25					1.87	9.96	8.81	-3.18	0.87	15.92
n_i=50					-1.76	-0.49	-0.18	-7.50	-9.74	13.05
MIP2										
n_i=25						7.98	7.19	-5.07	-1.38	14.58
n_i=50						0.27	0.60	-7.59	-8.42	17.39
HYB5										
n_i=25							-1.90	-16.89	-15.68	12.20
n_i=50							0.26	-3.06	-3.71	5.22
HYB4										
n_i=25								-17.36	-14.29	12.89
n_i=50								-3.48	-4.15	5.23
HYB2										
n_i=25									7.25	22.64
n_i=50									-0.48	21.41
LDF										
n_i=25										21.56
n_i=50										36.14

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 55 tests, i.e., the individual computed t-value must be significant at α = .05/55, resulting in a critical value of 3.368. The above t-values can also be compared to the critical value of 3.815, which is the critical value obtained from the Bonferroni method with a significance level of .01/55.

Table 16.--Paired T-Tests of Mean Difference in Classification Performance on Validation Samples for Training Samples of Sizes 25 and 50 Per Group for Configuration 1D

Method	Method									
	MSD4	MSD2	MIP5	MIP4	MIP2	HYB5	HYB4	HYB2	LDF	QDF
MSD5										
n_i=25	4.41	8.83	-6.77	-1.43	6.14	7.04	9.08	7.44	8.83	12.47
n_i=50	3.21	6.18	-9.13	-4.86	1.03	-5.40	-4.20	1.90	3.37	10.08
MSD4										
n_i=25		7.76	-9.25	-5.97	3.11	4.27	7.05	4.52	6.24	11.36
n_i=50		4.27	-10.31	-7.63	-1.50	-6.01	-6.29	-0.60	0.70	8.76
MSD2										
n_i=25			-14.23	-10.42	-5.38	-1.91	0.95	-1.28	0.24	7.22
n_i=50			-13.27	-8.81	-5.70	-7.70	-7.76	-4.81	-3.97	8.36
MIP5										
n_i=25				4.18	12.31	12.57	14.58	12.90	14.21	17.70
n_i=50				3.44	9.47	1.22	2.70	10.28	10.95	16.14
MIP4										
n_i=25					7.54	8.18	9.92	9.09	10.09	13.21
n_i=50					5.48	-1.38	-0.15	6.53	6.56	10.99
MIP2										
n_i=25						1.94	4.87	2.65	4.53	9.91
n_i=50						-5.19	-4.68	0.99	2.39	10.35
HYB5										
n_i=25							3.83	0.61	1.99	6.88
n_i=50							1.14	5.81	6.37	9.50
HYB4										
n_i=25								-2.04	-0.74	5.48
n_i=50								6.51	6.07	9.71
HYB2										
n_i=25									3.66	7.61
n_i=50									1.31	9.09
LDF										
n_i=25										7.93
n_i=50										10.92

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 55 tests, i.e., the individual computed t-value must be significant at α = .05/55, resulting in a critical value of 3.368. The above t-values can also be compared to the critical value of 3.815, which is the critical value obtained from the Bonferroni method with a significance level of .01/55.

Table 17.--Paired T-Tests of Mean Difference in Classification Performance on Validation Samples for Training Samples of Sizes 25 and 50 Per Group for Configuration 1E

Method	Method									
	MSD4	MSD2	MIP5	MIP4	MIP2	HYB5	HYB4	HYB2	LDF	QDF
MSD5										
n_i=25	5.04	6.52	-6.71	-3.64	2.16	4.99	-1.39	2.65	5.53	11.20
n_i=50	4.85	0.28	-8.91	-6.62	-3.10	-7.28	-4.73	-4.66	-2.42	7.83
MSD4										
n_i=25		3.12	-9.20	-6.52	-1.23	2.03	-4.59	-1.09	2.29	8.34
n_i=50		-3.24	-10.40	-9.55	-6.47	-10.18	-8.21	-8.57	-6.33	6.95
MSD2										
n_i=25			-11.56	-8.66	-5.39	-0.40	-8.87	-5.36	-1.32	8.71
n_i=50			-7.20	-5.93	-5.82	-6.67	-4.54	-7.92	-5.86	16.31
MIP5										
n_i=25				3.43	7.75	9.61	3.92	8.14	10.64	14.69
n_i=50				2.15	3.77	0.27	3.37	2.69	4.38	13.19
MIP4										
n_i=25					5.00	7.49	1.53	5.16	7.73	12.08
n_i=50					2.56	-1.46	1.63	1.03	3.18	11.74
MIP2										
n_i=25						3.61	-3.92	0.36	3.94	9.41
n_i=50						-3.38	-0.91	-1.82	1.31	13.77
HYB5										
n_i=25							-9.18	-3.79	-0.40	5.58
n_i=50							3.42	2.58	4.19	12.32
HYB4										
n_i=25								6.25	9.45	13.90
n_i=50								-0.50	1.75	11.75
HYB2										
n_i=25									7.25	12.71
n_i=50									2.89	17.10
LDF										
n_i=25										10.20
n_i=50										16.76

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 55 tests, i.e., the individual computed t-value must be significant at α = .05/55, resulting in a critical value of 3.368. The above t-values can also be compared to the critical value of 3.815, which is the critical value obtained from the Bonferroni method with a significance level of .01/55.

Table 18.--Paired T-Tests of Mean Difference in Classification Performance on Validation Samples for Training Samples of Sizes 25 and 50 Per Group for Configuration 1F

Method	MSD4	MSD2	MIP5*	MIP4*	MIP2*	HYB5	HYB4	HYB2	LDF	QDF
MSD5										
n_i=25	2.91	-67.44	-5.51	-2.05	-70.11	-28.36	-29.30	-45.67	-55.07	9.75
n_i=50	1.55	-118.97				-9.01	-4.88	-102.20	-75.31	10.65
MSD4										
n_i=25		-79.17	-7.60	-5.02	-83.82	-31.85	-30.55	-49.05	-56.70	8.91
n_i=50		-107.84				-9.43	-5.85	-97.95	-74.76	6.31
MSD2										
n_i=25			55.92	69.33	9.55	44.89	44.20	1.95	-6.26	110.97
n_i=50						98.13	95.41	-27.69	-9.05	125.28
MIP5*										
n_i=25				4.18	-58.95	-21.27	-20.22	-38.24	-47.56	12.57
MIP4*										
n_i=25					-71.63	-26.00	-25.78	-44.35	-52.53	11.64
MIP2*										
n_i=25						34.04	33.29	-3.91	-13.04	132.19
HYB5										
n_i=25							1.97	-23.79	-32.19	42.13
n_i=50							4.80	-91.17	-67.89	16.29
HYB4										
n_i=25								-24.65	-32.89	39.62
n_i=50								-91.91	-70.67	11.88
HYB2										
n_i=25									-9.50	57.98
n_i=50									21.39	107.46
LDF										
n_i=25										69.35
n_i=50										80.06

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 55 tests, i.e., the individual computed t-value must be significant at α = .05/55, resulting in a critical value of 3.368. The above t-values can also be compared to the critical value of 3.815, which is the critical value obtained from the Bonferroni method with a significance level of .01/55.

* Computationally too intensive to complete runs for MIP method with training samples of size 50 per group.

Table 19.--Paired T-Tests of Mean Difference in Classification Performance on Validation Samples for Training Samples of Sizes 25 and 50 Per Group for Configuration 1G

Method	Method									
	MSD4	MSD2	MIP5	MIP4	MIP2	HYB5	HYB4	HYB2	LDF	QDF
MSD5										
n_i=25	3.20	-12.54	-6.69	-4.59	-15.64	0.26	1.65	-11.69	-21.41	-2.87
n_i=50	-1.33	-15.85	-8.14	-7.19	-45.60	-7.23	-5.67	-16.12	-29.06	-11.28
MSD4										
n_i=25		-13.57	-8.80	-8.91	-22.03	-3.23	-1.31	-12.36	-23.03	-5.20
n_i=50		-15.65	-7.02	-7.67	-40.45	-5.81	-5.21	-15.96	-28.62	-10.67
MSD2										
n_i=25			10.57	11.55	9.62	13.40	14.02	-6.13	-14.13	13.47
n_i=50			14.38	14.87	10.67	14.94	14.54	-10.87	-20.03	14.51
MIP5										
n_i=25				2.21	-6.84	6.86	7.29	-10.35	-18.87	2.26
n_i=50				1.85	-24.19	1.05	0.26	-15.19	-27.05	-4.45
MIP4										
n_i=25					-11.21	5.32	6.38	-10.91	-20.31	0.67
n_i=50					-29.87	-0.54	-1.12	-15.48	-27.77	-6.67
MIP2										
n_i=25						18.97	21.78	-9.56	-18.58	10.24
n_i=50						28.82	19.43	-12.83	-23.22	14.29
HYB5										
n_i=25							3.17	-12.20	-22.96	-3.40
n_i=50							-0.80	-15.49	-27.91	-7.24
HYB4										
n_i=25								-12.53	-23.89	-4.48
n_i=50								-15.19	-26.84	-5.12
HYB2										
n_i=25									-4.06	12.44
n_i=50									-0.17	15.37
LDF										
n_i=25										23.78
n_i=50										28.66

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 55 tests, i.e., the individual computed t-value must be significant at α = .05/55, resulting in a critical value of 3.368. The above t-values can also be compared to the critical value of 3.815, which is the critical value obtained from the Bonferroni method with a significance level of .01/55.

Table 20.--Paired T-Tests of Mean Difference in Classification Performance on Validation Samples for Training Samples of Sizes 25 and 50 Per Group for Configuration 1H

Method	MSD4	MSD2	MIP5	MIP4	MIP2	HYB5	HYB4	HYB2	LDF	QDF
MSD5										
n_i=25	-8.71	-12.37	-0.16	-7.58	-12.27	-13.71	-12.83	-16.23	-52.04	-40.69
n_i=50	-20.69	-17.90	-0.20	-17.87	-27.46	-7.71	-19.39	-25.32	-70.21	-64.09
MSD4										
n_i=25		-9.78	8.73	1.25	-3.96	-9.47	-9.01	-13.89	-51.86	-43.76
n_i=50		-12.32	20.50	3.29	-7.85	1.72	-9.95	-20.22	-62.70	-67.47
MSD2										
n_i=25			12.78	9.17	7.05	-0.90	-1.02	-6.55	-27.22	-25.26
n_i=50			17.95	12.20	7.76	8.51	7.19	-5.30	-33.26	-33.22
MIP5										
n_i=25				-8.00	-11.31	-13.25	-12.91	-15.43	-49.23	-40.89
n_i=50				-18.76	-29.25	-7.36	-19.48	-26.22	-71.11	-64.67
MIP4										
n_i=25					-4.91	-9.48	-8.88	-13.15	-43.27	-39.74
n_i=50					-10.63	0.54	-9.09	-20.18	-61.37	-60.68
MIP2										
n_i=25						-7.15	-6.55	-11.27	-40.98	-35.28
n_i=50						5.34	-0.97	-13.69	-56.83	-58.01
HYB5										
n_i=25							-0.27	-5.96	-21.41	-18.33
n_i=50							-4.71	-12.27	-39.58	-38.52
HYB4										
n_i=25								-5.52	-19.07	-16.41
n_i=50								-13.33	-50.46	-64.94
HYB2										
n_i=25									-13.61	-11.26
n_i=50									-31.68	-35.38
LDF										
n_i=25										-1.01
n_i=50										-4.81

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 55 tests, i.e., the individual computed t-value must be significant at α = .05/55, resulting in a critical value of 3.368. The above t-values can also be compared to the critical value of 3.815, which is the critical value obtained from the Bonferroni method with a significance level of .01/55.

Table 21.--Classification Models for Research Question 2

Models	Descriptions
1. MIP1	MIP with maximize distance between projected means (Bounded Scores)
2. MIP2	MIP with maximize distance between projected means (Bounded Coefficients)
3. MIP3	MIP with maximize the minimum internal deviation (Bounded Coefficients)
4. MIP4	MIP with minimize sum of the external deviations

Table 22.--Data Configurations for Research Question 2

Configuration	First Population: Mean Vector	First Population: Covariance Matrix	Second Population: Mean Vector	Second Population: Covariance Matrix
2A	$\begin{bmatrix} -2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2B	$\begin{bmatrix} -1.414 \\ 1.414 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.414 \\ -1.414 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2C	$\begin{bmatrix} -1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2D	$\begin{bmatrix} -2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
			10% of observations from	
			$\begin{bmatrix} -2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$
2E	$\begin{bmatrix} -1.414 \\ 1.414 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.414 \\ -1.414 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
			10% of observations from	
			$\begin{bmatrix} -1.414 \\ 1.414 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$
2F	$\begin{bmatrix} -2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
	10% of observations from		10% of observations from	
	$\begin{bmatrix} -2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$

Table 22.--Continued.

Configuration	First Population: Mean Vector	First Population: Covariance Matrix	Second Population: Mean Vector	Second Population: Covariance Matrix
2G	$\begin{bmatrix} -1.414 \\ 1.414 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.414 \\ -1.414 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
	10% of observations from		10% of observations from	
	$\begin{bmatrix} -1.414 \\ 1.414 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$	$\begin{bmatrix} 1.414 \\ -1.414 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$
2H	$\begin{bmatrix} -2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ -2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
	10% of observations from		10% of observations from	
	$\begin{bmatrix} -2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
2I	$\begin{bmatrix} -2 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ -2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
	20% of observations from		20% of observations from	
	$\begin{bmatrix} -2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
2J	$\begin{bmatrix} -4 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$
			10% of observations from	
			$\begin{bmatrix} -5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2K	$\begin{bmatrix} -4 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$
			20% of observations from	
			$\begin{bmatrix} -5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Table 22.--Continued.

Configuration	First Population: Mean Vector	First Population: Covariance Matrix	Second Population: Mean Vector	Second Population: Covariance Matrix
2L	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
	15% of observations from		15% of observations from	
	$\begin{bmatrix} -.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix}$	$\begin{bmatrix} 3.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix}$
2M	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
	20% of observations from		20% of observations from	
	$\begin{bmatrix} -2.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$	$\begin{bmatrix} 5.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$
2N	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
	20% of observations from		20% of observations from	
	$\begin{bmatrix} -2.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} .01 & 0 \\ 0 & .01 \end{bmatrix}$	$\begin{bmatrix} 5.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} .01 & 0 \\ 0 & .01 \end{bmatrix}$

Table 23.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2A

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	3.42	1.05	3.08	0.80
MIP2	6.14	2.66	4.25	1.70
MIP3	3.85	1.58	3.38	1.07
MIP4	3.94	1.82	3.26	0.99

Table 24.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2B

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	3.29	1.11	3.03	0.70
MIP2	3.81	2.02	3.02	0.89
MIP3	3.56	1.44	3.18	0.98
MIP4	3.86	1.61	3.24	0.95

Table 25.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2C

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	16.74	2.07	16.18	1.35
MIP2	18.40	2.71	16.91	1.90
MIP3	17.27	2.43	16.38	1.60
MIP4	17.15	2.39	16.34	1.50

Table 26.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2D

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	8.52	2.19	7.69	1.22
MIP2	9.89	2.46	8.39	1.67
MIP3	8.40	2.16	7.48	1.06
MIP4	8.52	2.28	7.50	1.02

Table 27.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2E

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	8.38	2.13	7.73	1.36
MIP2	8.19	2.08	7.38	1.40
MIP3	8.16	1.92	7.49	1.42
MIP4	8.37	1.93	7.54	1.42

Table 28.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2F

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	5.47	1.96	4.65	0.93
MIP2	7.72	2.61	5.44	1.56
MIP3	5.52	2.11	4.69	1.10
MIP4	5.57	1.13	4.61	0.97

Table 29.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2G

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	5.18	1.71	4.56	0.99
MIP2	5.35	2.08	4.50	1.17
MIP3	5.29	1.86	4.58	1.02
MIP4	5.50	1.85	4.66	1.04

Table 30.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2H

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	2.22	1.03	1.94	0.68
MIP2	3.20	1.57	2.60	1.07
MIP3	2.47	1.30	2.19	0.83
MIP4	2.74	1.79	2.29	0.93

Table 31.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2I

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	3.37	1.21	2.92	0.70
MIP2	4.58	1.87	3.49	1.07
MIP3	3.81	1.57	3.15	0.87
MIP4	3.90	1.65	3.13	0.91

Table 32.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2J

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	9.22	2.01	8.41	1.19
MIP2	11.11	2.70	9.01	1.65
MIP3	8.77	1.87	8.15	1.10
MIP4	8.89	2.24	8.15	1.13

Table 33.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2K

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	14.43	2.05	13.34	1.07
MIP2	15.99	2.78	13.99	1.49
MIP3	13.59	1.79	12.99	0.94
MIP4	13.64	2.10	13.03	1.02

Table 34.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2L

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	13.63	2.74	12.40	1.36
MIP2	14.90	3.01	13.10	1.93
MIP3	13.44	2.55	12.59	1.60
MIP4	13.44	2.57	12.52	1.55

Table 35.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2M

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	9.80	2.76	8.57	1.30
MIP2	11.10	3.10	9.15	1.68
MIP3	9.69	2.62	8.56	1.29
MIP4	9.56	2.47	8.53	1.26

Table 36.--Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2N

Method	$n_1 = n_2 = 20$		$n_1 = n_2 = 40$	
	Mean	STD	Mean	STD
MIP1	7.10	1.73	6.37	1.12
MIP2	8.81	2.68	7.42	1.60
MIP3	7.28	1.76	6.79	1.30
MIP4	7.17	1.57	6.67	1.25

Table 37.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2A

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	-14.14	-3.43	-3.63
$n_i = 40$	-9.79	-4.21	-2.92
MIP2			
$n_i = 20$		13.99	12.94
$n_i = 40$		8.24	9.06
MIP3			
$n_i = 20$			-1.22
$n_i = 40$			2.84

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at α = .05/6, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 38.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2B

Method	MIP2	MIP3	MIP4
MIP1			
n_i = 20	<u>-3.50</u>	-2.50	<u>-5.11</u>
n_i = 40	0.27	-2.51	<u>-3.54</u>
MIP2			
n_i = 20		1.64	-0.26
n_i = 40		-2.40	<u>-3.32</u>
MIP3			
n_i = 20			<u>-3.85</u>
n_i = 40			-1.62

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at α = .05/6, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 39.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2C

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	-9.72	-4.46	-3.25
$n_i = 40$	-7.56	-2.74	-2.44
MIP2			
$n_i = 20$		6.90	7.64
$n_i = 40$		6.55	6.56
MIP3			
$n_i = 20$			2.07
$n_i = 40$			1.21

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at $\alpha = .05/6$, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 40.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2D

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	<u>-7.64</u>	0.98	0.01
$n_i = 40$	<u>-6.31</u>	2.66	2.43
MIP2			
$n_i = 20$		<u>9.70</u>	<u>9.03</u>
$n_i = 40$		<u>8.14</u>	<u>8.16</u>
MIP3			
$n_i = 20$			-2.32
$n_i = 40$			-0.36

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at $\alpha = .05/6$, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 41.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2E

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	1.08	1.40	0.05
$n_i = 40$	<u>3.49</u>	2.28	1.82
MIP2			
$n_i = 20$		0.16	-1.17
$n_i = 40$		-1.21	-1.77
MIP3			
$n_i = 20$			<u>-3.54</u>
$n_i = 40$			<u>-2.67</u>

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at $\alpha = .05/6$, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 42.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2F

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
n_i = 20	<u>-11.36</u>	-0.36	-0.79
n_i = 40	<u>-7.55</u>	-0.61	0.59
MIP2			
n_i = 20		<u>13.06</u>	<u>12.24</u>
n_i = 40		<u>8.63</u>	<u>9.40</u>
MIP3			
n_i = 20			-0.54
n_i = 40			1.87

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at α = .05/6, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 43.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2G

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	-1.12	-0.72	-2.22
$n_i = 40$	0.79	-0.18	-1.22
MIP2			
$n_i = 20$		0.36	-0.87
$n_i = 40$		-0.96	-1.94
MIP3			
$n_i = 20$			-2.78
$n_i = 40$			-2.49

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at $\alpha = .05/6$, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 44.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2H

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	<u>-8.39</u>	<u>-3.06</u>	<u>-4.46</u>
$n_i = 40$	<u>-8.38</u>	<u>-4.47</u>	<u>-5.87</u>
MIP2			
$n_i = 20$		<u>5.53</u>	<u>2.97</u>
$n_i = 40$		<u>5.28</u>	<u>3.53</u>
MIP3			
$n_i = 20$			<u>-3.22</u>
$n_i = 40$			-2.61

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at $\alpha = .05/6$, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 45.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2I

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	<u>-8.87</u>	<u>-4.15</u>	<u>-4.86</u>
$n_i = 40$	<u>-7.99</u>	<u>-4.60</u>	<u>-3.24</u>
MIP2			
$n_i = 20$		<u>5.41</u>	<u>4.62</u>
$n_i = 40$		<u>5.17</u>	<u>4.79</u>
MIP3			
$n_i = 20$			-1.62
$n_i = 40$			0.57

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at α = .05/6, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 46.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2J

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	-9.89	3.37	2.17
$n_i = 40$	-5.86	3.21	3.15
MIP2			
$n_i = 20$		13.51	12.05
$n_i = 40$		8.89	8.90
MIP3			
$n_i = 20$			-1.90
$n_i = 40$			0.09

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at α = .05/6, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 47.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2K

Method	Method: MIP2	MIP3	MIP4
MIP1			
n_i = 20	<u>-8.64</u>	<u>5.78</u>	<u>5.05</u>
n_i = 40	<u>-5.97</u>	<u>4.19</u>	<u>3.56</u>
MIP2			
n_i = 20		<u>12.68</u>	<u>11.70</u>
n_i = 40		<u>11.08</u>	<u>10.65</u>
MIP3			
n_i = 20			-0.86
n_i = 40			-1.65

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at α = .05/6, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 48.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2L

Method	Method: MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	<u>-7.51</u>	1.34	1.28
$n_i = 40$	<u>-6.60</u>	-2.02	-1.34
MIP2			
$n_i = 20$		<u>9.50</u>	<u>9.32</u>
$n_i = 40$		<u>5.98</u>	<u>6.35</u>
MIP3			
$n_i = 20$			0.02
$n_i = 40$			1.80

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at $\alpha = .05/6$, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 49.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2M

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	<u>-7.00</u>	0.64	1.36
$n_i = 40$	<u>-5.23</u>	0.13	0.49
MIP2			
$n_i = 20$		<u>7.60</u>	<u>8.45</u>
$n_i = 40$		<u>5.54</u>	<u>6.21</u>
MIP3			
$n_i = 20$			1.59
$n_i = 40$			0.71

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at $\alpha = .05/6$, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 50.--Paired T-Tests of Mean Difference in Exact Misclassification Rates for Training Samples of Sizes 20 and 40 Per Group for Configuration 2N

Method	Method		
	MIP2	MIP3	MIP4
MIP1			
$n_i = 20$	-10.09	-1.57	-0.73
$n_i = 40$	-10.99	-5.98	-4.99
MIP2			
$n_i = 20$		10.10	10.00
$n_i = 40$		6.55	7.60
MIP3			
$n_i = 20$			1.59
$n_i = 40$			3.35

Note: A positive (negative) t-value in position (i,j) of the table indicates that the mean misclassification rate of method i is higher (lower) than that of method j. Underlined t-values indicate pairs of means that differ significantly after applying the Bonferroni adjustment to the family of 6 tests, i.e., the individual computed t-value must be significant at $\alpha = .05/6$, resulting in a critical value of 2.665. The above t-values can also be compared to the critical value of 3.187, which is the critical value obtained from the Bonferroni method with a significance level of .01/6.

Table 51.--Values of Skewness and Kurtosis Measures for Various Settings of Mean (μ) and Standard Deviation (σ) with Contaminating Fraction (ϵ) = 0.01

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
0.5	0.1	-0.0136	3.0155	2.0	0.1	0.0180	2.9506
0.5	0.5	-0.0100	3.0064	2.0	0.5	0.0315	2.9916
0.5	1.0	0.0012	3.0006	2.0	1.0	0.0732	3.1379
0.5	1.5	0.0193	3.0633	2.0	1.5	0.1407	3.4395
0.5	2.0	0.0436	3.2923	2.0	2.0	0.2313	3.9745
0.5	2.5	0.0731	3.8047	2.0	2.5	0.3413	4.8362
0.5	3.0	0.1066	4.7220	2.0	3.0	0.4666	6.1213
0.5	3.5	0.1429	6.1558	2.0	3.5	0.6032	7.9177
0.5	4.0	0.1810	8.1960	2.0	4.0	0.7465	10.2955
0.5	4.5	0.2198	10.9039	2.0	4.5	0.8928	13.3006
0.5	5.0	0.2582	14.3083	2.0	5.0	1.0385	16.9519
0.5	5.5	0.2956	18.4067	2.0	5.5	1.1806	21.2419
0.5	6.0	0.3312	23.1687	2.0	6.0	1.3166	26.1390
1.0	0.1	-0.0197	2.9808	2.5	0.1	0.0724	3.0295
1.0	0.5	-0.0125	2.9824	2.5	0.5	0.0886	3.0968
1.0	1.0	0.0096	3.0091	2.5	1.0	0.1385	3.3226
1.0	1.5	0.0453	3.1229	2.5	1.5	0.2195	3.7493
1.0	2.0	0.0932	3.4173	2.5	2.0	0.3281	4.4448
1.0	2.5	0.1513	4.0043	2.5	2.5	0.4602	5.4902
1.0	3.0	0.2173	5.0000	2.5	3.0	0.6110	6.9690
1.0	3.5	0.2891	6.5110	2.5	3.5	0.7755	8.9570
1.0	4.0	0.3644	8.6230	2.5	4.0	0.9485	11.5137
1.0	4.5	0.4410	11.3939	2.5	4.5	1.1253	14.6769
1.0	5.0	0.5170	14.8506	2.5	5.0	1.3017	18.4600
1.0	5.5	0.5910	18.9896	2.5	5.5	1.4741	22.8524
1.0	6.0	0.6616	23.7801	2.5	6.0	1.6394	27.8221
1.5	0.1	-0.0112	2.9479	3.0	0.1	0.1550	3.2273
1.5	0.5	-0.0007	2.9666	3.0	0.5	0.1735	3.3232
1.5	1.0	0.0317	3.0451	3.0	1.0	0.2305	3.6359
1.5	1.5	0.0840	3.2403	3.0	1.5	0.3229	4.1994
1.5	2.0	0.1542	3.6388	3.0	2.0	0.4471	5.0707
1.5	2.5	0.2393	4.3448	3.0	2.5	0.5983	6.3174
1.5	3.0	0.3363	5.4656	3.0	3.0	0.7711	8.0090
1.5	3.5	0.4417	7.0999	3.0	3.5	0.9599	10.2079
1.5	4.0	0.5523	9.3267	3.0	4.0	1.1589	12.9620
1.5	4.5	0.6650	12.1986	3.0	4.5	1.3627	16.3000
1.5	5.0	0.7770	15.7392	3.0	5.0	1.5665	20.2291
1.5	5.5	0.8861	19.9431	3.0	5.5	1.7661	24.7353
1.5	6.0	0.9903	24.7795	3.0	6.0	1.9581	29.7858

Table 51.--Continued.

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
3.5	0.1	0.2672	3.5833	5.0	0.1	0.7741	5.8781
3.5	0.5	0.2876	3.7088	5.0	0.5	0.7976	6.0861
3.5	1.0	0.3503	4.1114	5.0	1.0	0.8704	6.7397
3.5	1.5	0.4522	4.8166	5.0	1.5	0.9887	7.8410
3.5	2.0	0.5892	5.8700	5.0	2.0	1.1484	9.4051
3.5	2.5	0.7563	7.3257	5.0	2.5	1.3442	11.4490
3.5	3.0	0.9476	9.2388	5.0	3.0	1.5696	13.9877
3.5	3.5	1.1569	11.6576	5.0	3.5	1.8180	17.0306
3.5	4.0	1.3780	14.6180	5.0	4.0	2.0824	20.5790
3.5	4.5	1.6051	18.1391	5.0	4.5	2.3561	24.6243
3.5	5.0	1.8327	22.2216	5.0	5.0	2.6330	29.1470
3.5	5.5	2.0562	26.8476	5.0	5.5	2.9074	34.1173
3.5	6.0	2.2718	31.9831	5.0	6.0	3.1748	39.4960
4.0	0.1	0.4089	4.1303	5.5	0.1	0.9918	7.0931
4.0	0.5	0.4307	4.2850	5.5	0.5	1.0158	7.3236
4.0	1.0	0.4980	4.7765	5.5	1.0	1.0897	8.0461
4.0	1.5	0.6073	5.6218	5.5	1.5	1.2100	9.2561
4.0	2.0	0.7544	6.8557	5.5	2.0	1.3727	10.9610
4.0	2.5	0.9341	8.5189	5.5	2.5	1.5724	13.1674
4.0	3.0	1.1402	10.6523	5.5	3.0	1.8028	15.8795
4.0	3.5	1.3662	13.2904	5.5	3.5	2.0574	19.0960
4.0	4.0	1.6056	16.4570	5.5	4.0	2.3291	22.8085
4.0	4.5	1.8520	20.1617	5.5	4.5	2.6112	27.0002
4.0	5.0	2.0997	24.3981	5.5	5.0	2.8975	31.6458
4.0	5.5	2.3436	29.1446	5.5	5.5	3.1823	36.7115
4.0	6.0	2.5796	34.3655	5.5	6.0	3.4607	42.1564
4.5	0.1	0.5787	4.8911	6.0	0.1	1.2283	8.5281
4.5	0.5	0.6016	5.0736	6.0	0.5	1.2523	8.7780
4.5	1.0	0.6722	5.6496	6.0	1.0	1.3266	9.5595
4.5	1.5	0.7868	6.6283	6.0	1.5	1.4475	10.8629
4.5	2.0	0.9415	8.0342	6.0	2.0	1.6112	12.6886
4.5	2.5	1.1306	9.8956	6.0	2.5	1.8126	15.0347
4.5	3.0	1.3481	12.2397	6.0	3.0	2.0454	17.8960
4.5	3.5	1.5870	15.0882	6.0	3.5	2.3032	21.2619
4.5	4.0	1.8407	18.4530	6.0	4.0	2.5792	25.1154
4.5	4.5	2.1026	22.3343	6.0	4.5	2.8666	29.4329
4.5	5.0	2.3667	26.7193	6.0	5.0	3.1593	34.1834
4.5	5.5	2.6276	31.5819	6.0	5.5	3.4515	39.3297
4.5	6.0	2.8809	36.8850	6.0	6.0	3.7382	44.8291

Table 52.--Values of Skewness and Kurtosis Measures for Various Settings of Mean (μ) and Standard Deviation (σ) with Contaminating Fraction (ϵ) = 0.05

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
0.5	0.1	-0.0691	3.0845	2.0	0.1	0.0491	2.7442
0.5	0.5	-0.0500	3.0360	2.0	0.5	0.1037	2.8901
0.5	1.0	0.0052	3.0021	2.0	1.0	0.2635	3.3837
0.5	1.5	0.0848	3.2642	2.0	1.5	0.4981	4.3058
0.5	2.0	0.1749	4.0941	2.0	2.0	0.7717	5.7311
0.5	2.5	0.2637	5.6271	2.0	2.5	1.0501	7.6724
0.5	3.0	0.3430	7.8335	2.0	3.0	1.3078	10.0691
0.5	3.5	0.4085	10.5681	2.0	3.5	1.5294	12.8074
0.5	4.0	0.4593	13.6393	2.0	4.0	1.7087	15.7527
0.5	4.5	0.4963	16.8634	2.0	4.5	1.8455	18.7770
0.5	5.0	0.5214	20.0921	2.0	5.0	1.9438	21.7755
0.5	5.5	0.5368	23.2196	2.0	5.5	2.0089	24.6710
0.5	6.0	0.5445	26.1791	2.0	6.0	2.0466	27.4130
1.0	0.1	-0.0986	2.9194	2.5	0.1	0.2263	2.9224
1.0	0.5	-0.0632	2.9233	2.5	0.5	0.2836	3.1289
1.0	1.0	0.0399	3.0310	2.5	1.0	0.4523	3.7888
1.0	1.5	0.1889	3.4685	2.5	1.5	0.7024	4.9228
1.0	2.0	0.3589	4.4546	2.5	2.0	0.9979	6.5437
1.0	2.5	0.5275	6.0932	2.5	2.5	1.3035	8.6220
1.0	3.0	0.6791	8.3483	2.5	3.0	1.5915	11.0822
1.0	3.5	0.8057	11.0841	2.5	3.5	1.8442	13.8169
1.0	4.0	0.9047	14.1247	2.5	4.0	2.0529	16.7089
1.0	4.5	0.9776	17.3008	2.5	4.5	2.2160	19.6495
1.0	5.0	1.0277	20.4749	2.5	5.0	2.3364	22.5499
1.0	5.5	1.0589	23.5480	2.5	5.5	2.4191	25.3447
1.0	6.0	1.0751	26.4569	2.5	6.0	2.4700	27.9906
1.5	0.1	-0.0619	2.7677	3.0	0.1	0.4519	3.3187
1.5	0.5	-0.0145	2.8419	3.0	0.5	0.5087	3.5692
1.5	1.0	0.1239	3.1403	3.0	1.0	0.6768	4.3500
1.5	1.5	0.3254	3.8161	3.0	1.5	0.9284	5.6392
1.5	2.0	0.5575	5.0167	3.0	2.0	1.2299	7.4038
1.5	2.5	0.7903	6.8020	3.0	2.5	1.5469	9.5818
1.5	3.0	1.0024	9.1255	3.0	3.0	1.8515	12.0842
1.5	3.5	1.1818	11.8631	3.0	3.5	2.1244	14.8070
1.5	4.0	1.3242	14.8595	3.0	4.0	2.3549	17.6453
1.5	4.5	1.4307	17.9655	3.0	4.5	2.5396	20.5058
1.5	5.0	1.5053	21.0588	3.0	5.0	2.6800	23.3128
1.5	5.5	1.5531	24.0508	3.0	5.5	2.7801	26.0112
1.5	6.0	1.5792	26.8839	3.0	6.0	2.8454	28.5644

Table 52.--Continued.

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
3.5	0.1	0.7060	3.9052	5.0	0.1	1.4837	6.2854
3.5	0.5	0.7601	4.1824	5.0	0.5	1.5256	6.5689
3.5	1.0	0.9213	5.0367	5.0	1.0	1.6517	7.4359
3.5	1.5	1.1649	6.4197	5.0	1.5	1.8472	8.8202
3.5	2.0	1.4607	8.2695	5.0	2.0	2.0934	10.6404
3.5	2.5	1.7771	10.5022	5.0	2.5	2.3689	12.7992
3.5	3.0	2.0870	13.0191	5.0	3.0	2.6534	15.1949
3.5	3.5	2.3705	15.7177	5.0	3.5	2.9295	17.7309
3.5	4.0	2.6157	18.5014	5.0	4.0	3.1842	20.3228
3.5	4.5	2.8172	21.2871	5.0	4.5	3.4091	22.9015
3.5	5.0	2.9748	24.0093	5.0	5.0	3.5998	25.4142
3.5	5.5	3.0914	26.6206	5.0	5.5	3.7549	27.8234
3.5	6.0	3.1716	29.0898	5.0	6.0	3.8756	30.1046
4.0	0.1	0.9708	4.6307	5.5	0.1	1.7178	7.1294
4.0	0.5	1.0212	4.9200	5.5	0.5	1.7556	7.4014
4.0	1.0	1.1718	5.8068	5.5	1.0	1.8697	8.2334
4.0	1.5	1.4016	7.2291	5.5	1.5	2.0480	9.5632
4.0	2.0	1.6843	9.1097	5.5	2.0	2.2746	11.3142
4.0	2.5	1.9916	11.3530	5.5	2.5	2.5315	13.3945
4.0	3.0	2.2983	13.8553	5.5	3.0	2.8008	15.7078
4.0	3.5	2.5847	16.5149	5.5	3.5	3.0666	18.1622
4.0	4.0	2.8381	19.2405	5.5	4.0	3.3166	20.6767
4.0	4.5	3.0517	21.9559	5.5	4.5	3.5420	23.1844
4.0	5.0	3.2235	24.6023	5.5	5.0	3.7375	25.6340
4.0	5.5	3.3551	27.1373	5.5	5.5	3.9009	27.9883
4.0	6.0	3.4498	29.5336	5.5	6.0	4.0321	30.2226
4.5	0.1	1.2330	5.4402	6.0	0.1	1.9328	7.9466
4.5	0.5	1.2791	5.7304	6.0	0.5	1.9668	8.2042
4.5	1.0	1.4176	6.6183	6.0	1.0	2.0698	8.9933
4.5	1.5	1.6308	8.0370	6.0	1.5	2.2318	10.2567
4.5	2.0	1.8961	9.9039	6.0	2.0	2.4395	11.9246
4.5	2.5	2.1890	12.1198	6.0	2.5	2.6777	13.9127
4.5	3.0	2.4865	14.5802	6.0	3.0	2.9308	16.1313
4.5	3.5	2.7699	17.1854	6.0	3.5	3.1846	18.4944
4.5	4.0	3.0262	19.8475	6.0	4.0	3.4275	20.9247
4.5	4.5	3.2473	22.4948	6.0	4.5	3.6506	23.3581
4.5	5.0	3.4302	25.0722	6.0	5.0	3.8484	25.7439
4.5	5.5	3.5747	27.5407	6.0	5.5	4.0177	28.0450
4.5	6.0	3.6831	29.8750	6.0	6.0	4.1575	30.2360

Table 53.--Values of Skewness and Kurtosis Measures for Various Settings of Mean (μ) and Standard Deviation (σ) with Contaminating Fraction (ϵ) = 0.10

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
0.5	0.1	-0.1405	3.1880	2.0	0.1	0.0292	2.5071
0.5	0.5	-0.1000	3.0818	2.0	0.5	0.1174	2.7083
0.5	1.0	0.0087	3.0025	2.0	1.0	0.3632	3.3581
0.5	1.5	0.1446	3.4249	2.0	1.5	0.6913	4.4712
0.5	2.0	0.2722	4.5761	2.0	2.0	1.0268	6.0035
0.5	2.5	0.3728	6.3454	2.0	2.5	1.3180	7.8340
0.5	3.0	0.4426	8.4634	2.0	3.0	1.5423	9.8086
0.5	3.5	0.4855	10.6737	2.0	3.5	1.6978	11.7890
0.5	4.0	0.5077	12.8024	2.0	4.0	1.7938	13.6769
0.5	4.5	0.5153	14.7560	2.0	4.5	1.8427	15.4155
0.5	5.0	0.5131	16.4985	2.0	5.0	1.8566	16.9808
0.5	5.5	0.5046	18.0271	2.0	5.5	1.8456	18.3698
0.5	6.0	0.4922	19.3562	2.0	6.0	1.8178	19.5918
1.0	0.1	-0.1980	2.8761	2.5	0.1	0.2580	2.6306
1.0	0.5	-0.1276	2.8731	2.5	0.5	0.3411	2.8843
1.0	1.0	0.0633	3.0348	2.5	1.0	0.5760	3.6624
1.0	1.5	0.3058	3.6796	2.5	1.5	0.8981	4.9012
1.0	2.0	0.5382	4.9499	2.5	2.0	1.2393	6.5017
1.0	2.5	0.7257	6.7387	2.5	2.5	1.5480	8.3318
1.0	3.0	0.8590	8.8166	2.5	3.0	1.7969	10.2557
1.0	3.5	0.9433	10.9640	2.5	3.5	1.9789	12.1606
1.0	4.0	0.9889	13.0284	2.5	4.0	2.0991	13.9679
1.0	4.5	1.0066	14.9259	2.5	4.5	2.1678	15.6324
1.0	5.0	1.0049	16.6230	2.5	5.0	2.1964	17.1349
1.0	5.5	0.9906	18.1164	2.5	5.5	2.1953	18.4732
1.0	6.0	0.9683	19.4187	2.5	6.0	2.1728	19.6557
1.5	0.1	-0.1363	2.5992	3.0	0.1	0.5103	2.9211
1.5	0.5	-0.0507	2.7109	3.0	0.5	0.5848	3.1958
1.5	1.0	0.1843	3.1449	3.0	1.0	0.7983	4.0236
1.5	1.5	0.4899	4.0478	3.0	1.5	1.0984	5.3063
1.5	2.0	0.7917	5.4609	3.0	2.0	1.4271	6.9189
1.5	2.5	1.0435	7.2739	3.0	2.5	1.7367	8.7238
1.5	3.0	1.2291	9.3006	3.0	3.0	1.9978	10.5949
1.5	3.5	1.3516	11.3652	3.0	3.5	2.1989	12.4338
1.5	4.0	1.4221	13.3429	3.0	4.0	2.3404	14.1740
1.5	4.5	1.4535	15.1633	3.0	4.5	2.4294	15.7775
1.5	5.0	1.4571	16.7970	3.0	5.0	2.4755	17.2284
1.5	5.5	1.4417	18.2406	3.0	5.5	2.4882	18.5252
1.5	6.0	1.4139	19.5048	3.0	6.0	2.4758	19.6755

Table 53.--Continued.

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
3.5	0.1	0.7587	3.3085	5.0	0.1	1.3701	4.5558
3.5	0.5	0.8238	3.5827	5.0	0.5	1.4119	4.7783
3.5	1.0	1.0126	4.4054	5.0	1.0	1.5361	5.4497
3.5	1.5	1.2839	5.6703	5.0	1.5	1.7237	6.4938
3.5	2.0	1.5903	7.2478	5.0	2.0	1.9510	7.8169
3.5	2.5	1.8897	9.0022	5.0	2.5	2.1934	9.3167
3.5	3.0	2.1531	10.8139	5.0	3.0	2.4293	10.8985
3.5	3.5	2.3659	12.5922	5.0	3.5	2.6432	12.4849
3.5	4.0	2.5246	14.2762	5.0	4.0	2.8254	14.0194
3.5	4.5	2.6326	15.8314	5.0	4.5	2.9720	15.4652
3.5	5.0	2.6970	17.2430	5.0	5.0	3.0827	16.8020
3.5	5.5	2.7259	18.5091	5.0	5.5	3.1599	18.0214
3.5	6.0	2.7270	19.6364	5.0	6.0	3.2075	19.1235
4.0	0.1	0.9880	3.7336	5.5	0.1	1.5236	4.9201
4.0	0.5	1.0443	3.9952	5.5	0.5	1.5597	5.1222
4.0	1.0	1.2090	4.7802	5.5	1.0	1.6679	5.7339
4.0	1.5	1.4503	5.9882	5.5	1.5	1.8332	6.6911
4.0	2.0	1.7303	7.4974	5.5	2.0	2.0370	7.9144
4.0	2.5	2.0131	9.1798	5.5	2.5	2.2590	9.3151
4.0	3.0	2.2716	10.9232	5.5	3.0	2.4808	10.8085
4.0	3.5	2.4896	12.6411	5.5	3.5	2.6877	12.3228
4.0	4.0	2.6606	14.2750	5.5	4.0	2.8701	13.8033
4.0	4.5	2.7850	15.7908	5.5	4.5	3.0225	15.2121
4.0	5.0	2.8670	17.1728	5.5	5.0	3.1433	16.5264
4.0	5.5	2.9129	18.4179	5.5	5.5	3.2333	17.7349
4.0	6.0	2.9293	19.5309	5.5	6.0	3.2949	18.8348
4.5	0.1	1.1921	4.1568	6.0	0.1	1.6552	5.2465
4.5	0.5	1.2406	4.3999	6.0	0.5	1.6867	5.4293
4.5	1.0	1.3836	5.1310	6.0	1.0	1.7813	5.9845
4.5	1.5	1.5967	6.2613	6.0	1.5	1.9274	6.8585
4.5	2.0	1.8497	7.6821	6.0	2.0	2.1101	7.9846
4.5	2.5	2.1127	9.2776	6.0	2.5	2.3128	9.2868
4.5	3.0	2.3614	10.9436	6.0	3.0	2.5199	10.6905
4.5	3.5	2.5793	12.5980	6.0	3.5	2.7181	12.1298
4.5	4.0	2.7581	14.1832	6.0	4.0	2.8979	13.5526
4.5	4.5	2.8953	15.6639	6.0	4.5	3.0532	14.9210
4.5	5.0	2.9930	17.0222	6.0	5.0	3.1813	16.2100
4.5	5.5	3.0554	18.2528	6.0	5.5	3.2816	17.4058
4.5	6.0	3.0876	19.3583	6.0	6.0	3.3554	18.5026

Table 54.--Values of Skewness and Kurtosis Measures for Various Settings of Mean (μ) and Standard Deviation (σ) with Contaminating Fraction (ϵ) = 0.15

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
0.5	0.1	-0.2146	3.3130	2.0	0.1	-0.0273	2.3169
0.5	0.5	-0.1501	3.1380	2.0	0.5	0.0849	2.5331
0.5	1.0	0.0106	3.0018	2.0	1.0	0.3848	3.2103
0.5	1.5	0.1858	3.5158	2.0	1.5	0.7552	4.3030
0.5	2.0	0.3242	4.7514	2.0	2.0	1.0966	5.6936
0.5	2.5	0.4137	6.3979	2.0	2.5	1.3583	7.2185
0.5	3.0	0.4622	8.1298	2.0	3.0	1.5319	8.7319
0.5	3.5	0.4823	9.7503	2.0	3.5	1.6301	10.1388
0.5	4.0	0.4844	11.1770	2.0	4.0	1.6718	11.3941
0.5	4.5	0.4759	12.3949	2.0	4.5	1.6743	12.4867
0.5	5.0	0.4616	13.4191	2.0	5.0	1.6512	13.4246
0.5	5.5	0.4443	14.2759	2.0	5.5	1.6122	14.2243
0.5	6.0	0.4259	14.9926	2.0	6.0	1.5640	14.9046
1.0	0.1	-0.2988	2.8693	2.5	0.1	0.2116	2.3493
1.0	0.5	-0.1933	2.8481	2.5	0.5	0.3098	2.6036
1.0	1.0	0.0745	3.0236	2.5	1.0	0.5790	3.3625
1.0	1.5	0.3763	3.7500	2.5	1.5	0.9265	4.5114
1.0	2.0	0.6242	5.0410	2.5	2.0	1.2658	5.9026
1.0	2.5	0.7914	6.6496	2.5	2.5	1.5440	7.3845
1.0	3.0	0.8869	8.3152	2.5	3.0	1.7433	8.8372
1.0	3.5	0.9300	9.8732	2.5	3.5	1.8684	10.1851
1.0	4.0	0.9387	11.2511	2.5	4.0	1.9332	11.3919
1.0	4.5	0.9263	12.4340	2.5	4.5	1.9532	12.4488
1.0	5.0	0.9018	13.4344	2.5	5.0	1.9417	13.3623
1.0	5.5	0.8707	14.2755	2.5	5.5	1.9093	14.1466
1.0	6.0	0.8368	14.9819	2.5	6.0	1.8637	14.8182
1.5	0.1	-0.2197	2.4859	3.0	0.1	0.4507	2.5071
1.5	0.5	-0.1014	2.6108	3.0	0.5	0.5336	2.7652
1.5	1.0	0.2063	3.0916	3.0	1.0	0.7657	3.5263
1.5	1.5	0.5689	4.0376	3.0	1.5	1.0774	4.6597
1.5	2.0	0.8835	5.3896	3.0	2.0	1.3979	6.0129
1.5	2.5	1.1090	6.9553	3.0	2.5	1.6774	7.4429
1.5	3.0	1.2472	8.5413	3.0	3.0	1.8923	8.8419
1.5	3.5	1.3170	10.0213	3.0	3.5	2.0398	10.1432
1.5	4.0	1.3391	11.3367	3.0	4.0	2.1278	11.3143
1.5	4.5	1.3303	12.4738	3.0	4.5	2.1685	12.3464
1.5	5.0	1.3025	13.4427	3.0	5.0	2.1736	13.2445
1.5	5.5	1.2639	14.2628	3.0	5.5	2.1534	14.0206
1.5	6.0	1.2196	14.9559	3.0	6.0	2.1159	14.6893

Table 54.--Continued.

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
3.5	0.1	0.6671	2.7213	5.0	0.1	1.1411	3.3585
3.5	0.5	0.7363	2.9652	5.0	0.5	1.1818	3.5360
3.5	1.0	0.9332	3.6851	5.0	1.0	1.3019	4.0679
3.5	1.5	1.2065	4.7587	5.0	1.5	1.4804	4.8839
3.5	2.0	1.5004	6.0446	5.0	2.0	1.6916	5.8983
3.5	2.5	1.7709	7.4105	5.0	2.5	1.9102	7.0223
3.5	3.0	1.9923	8.7559	5.0	3.0	2.1157	8.1785
3.5	3.5	2.1562	10.0169	5.0	3.5	2.2944	9.3086
3.5	4.0	2.2650	11.1608	5.0	4.0	2.4394	10.3743
3.5	4.5	2.3267	12.1769	5.0	4.5	2.5492	11.3542
3.5	5.0	2.3508	13.0676	5.0	5.0	2.6255	12.2398
3.5	5.5	2.3464	13.8425	5.0	5.5	2.6721	13.0307
3.5	6.0	2.3213	14.5141	5.0	6.0	2.6936	13.7319
4.0	0.1	0.8537	2.9477	5.5	0.1	1.2496	3.5303
4.0	0.5	0.9113	3.1703	5.5	0.5	1.2842	3.6879
4.0	1.0	1.0777	3.8300	5.5	1.0	1.3873	4.1623
4.0	1.5	1.3150	4.8219	5.5	1.5	1.5428	4.8967
4.0	2.0	1.5801	6.0227	5.5	2.0	1.7309	5.8213
4.0	2.5	1.8355	7.3137	5.5	2.5	1.9309	6.8609
4.0	3.0	2.0562	8.6011	5.5	3.0	2.1252	7.9471
4.0	3.5	2.2302	9.8222	5.5	3.5	2.3006	9.0255
4.0	4.0	2.3557	10.9421	5.5	4.0	2.4493	10.0576
4.0	4.5	2.4367	11.9467	5.5	4.5	2.5681	11.0197
4.0	5.0	2.4799	12.8350	5.5	5.0	2.6568	11.8999
4.0	5.5	2.4930	13.6138	5.5	5.5	2.7176	12.6946
4.0	6.0	2.4828	14.2932	5.5	6.0	2.7537	13.4058
4.5	0.1	1.0105	3.1634	6.0	0.1	1.3397	3.6795
4.5	0.5	1.0588	3.3630	6.0	0.5	1.3695	3.8195
4.5	1.0	1.1997	3.9577	6.0	1.0	1.4586	4.2427
4.5	1.5	1.4053	4.8608	6.0	1.5	1.5948	4.9032
4.5	2.0	1.6424	5.9688	6.0	2.0	1.7625	5.7441
4.5	2.5	1.8799	7.1778	6.0	2.5	1.9451	6.7023
4.5	3.0	2.0946	8.4017	6.0	3.0	2.1273	7.7184
4.5	3.5	2.2732	9.5796	6.0	3.5	2.2971	8.7422
4.5	4.0	2.4107	10.6743	6.0	4.0	2.4464	9.7365
4.5	4.5	2.5079	11.6678	6.0	4.5	2.5708	10.6762
4.5	5.0	2.5689	12.5554	6.0	5.0	2.6690	11.5467
4.5	5.5	2.5994	13.3403	6.0	5.5	2.7417	12.3416
4.5	6.0	2.6051	14.0303	6.0	6.0	2.7908	13.0601

Table 55.--Values of Skewness and Kurtosis Measures for Various Settings of Mean (μ) and Standard Deviation (σ) with Contaminating Fraction (ϵ) = 0.20

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
0.5	0.1	-0.2920	3.4631	2.0	0.1	-0.1053	2.1785
0.5	0.5	-0.2001	3.2050	2.0	0.5	0.0264	2.3894
0.5	1.0	0.0113	3.0004	2.0	1.0	0.3657	3.0381
0.5	1.5	0.2129	3.5591	2.0	1.5	0.7574	4.0449
0.5	2.0	0.3485	4.7670	2.0	2.0	1.0881	5.2589
0.5	2.5	0.4210	6.2019	2.0	2.5	1.3164	6.5141
0.5	3.0	0.4504	7.5731	2.0	3.0	1.4486	7.6920
0.5	3.5	0.4545	8.7622	2.0	3.5	1.5078	8.7343
0.5	4.0	0.4448	9.7494	2.0	4.0	1.5176	9.6263
0.5	4.5	0.4284	10.5544	2.0	4.5	1.4963	10.3764
0.5	5.0	0.4090	11.2079	2.0	5.0	1.4568	11.0022
0.5	5.5	0.3888	11.7395	2.0	5.5	1.4073	11.5234
0.5	6.0	0.3690	12.1743	2.0	6.0	1.3533	11.9584
1.0	0.1	-0.4019	2.8991	2.5	0.1	0.1290	2.1243
1.0	0.5	-0.2601	2.8475	2.5	0.5	0.2384	2.3630
1.0	1.0	0.0768	3.0048	2.5	1.0	0.5303	3.0625
1.0	1.5	0.4157	3.7426	2.5	1.5	0.8889	4.0864
1.0	2.0	0.6578	4.9545	2.5	2.0	1.2165	5.2737
1.0	2.5	0.7962	6.3293	2.5	2.5	1.4643	6.4808
1.0	3.0	0.8584	7.6385	2.5	3.0	1.6251	7.6119
1.0	3.5	0.8728	8.7822	2.5	3.5	1.7120	8.6194
1.0	4.0	0.8599	9.7404	2.5	4.0	1.7441	9.4900
1.0	4.5	0.8325	10.5284	2.5	4.5	1.7386	10.2297
1.0	5.0	0.7983	11.1728	2.5	5.0	1.7088	10.8532
1.0	5.5	0.7616	11.7001	2.5	5.5	1.6641	11.3771
1.0	6.0	0.7249	12.1336	2.5	6.0	1.6111	11.8179
1.5	0.1	-0.3104	2.4222	3.0	0.1	0.3475	2.1757
1.5	0.5	-0.1623	2.5427	3.0	0.5	0.4363	2.4089
1.5	1.0	0.2043	3.0175	3.0	1.0	0.6801	3.0871
1.5	1.5	0.5992	3.9268	3.0	1.5	0.9955	4.0708
1.5	2.0	0.9052	5.1450	3.0	2.0	1.3040	5.2064
1.5	2.5	1.0969	6.4549	3.0	2.5	1.5571	6.3632
1.5	3.0	1.1947	7.6933	3.0	3.0	1.7379	7.4549
1.5	3.5	1.2282	8.7828	3.0	3.5	1.8502	8.4368
1.5	4.0	1.2219	9.7057	3.0	4.0	1.9067	9.2946
1.5	4.5	1.1927	10.4731	3.0	4.5	1.9215	10.0311
1.5	5.0	1.1514	11.1068	3.0	5.0	1.9071	10.6580
1.5	5.5	1.1046	11.6298	3.0	5.5	1.8732	11.1896
1.5	6.0	1.0559	12.0628	3.0	6.0	1.8271	11.6402

Table 55.--Continued.

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
3.5	0.1	0.5344	2.2719	5.0	0.1	0.9146	2.5756
3.5	0.5	0.6063	2.4856	5.0	0.5	0.9550	2.7224
3.5	1.0	0.8082	3.1096	5.0	1.0	1.0733	3.1600
3.5	1.5	1.0808	4.0220	5.0	1.5	1.2470	3.8254
3.5	2.0	1.3631	5.0869	5.0	2.0	1.4488	4.6429
3.5	2.5	1.6110	6.1862	5.0	2.5	1.6531	5.5362
3.5	3.0	1.8029	7.2382	5.0	3.0	1.8401	6.4417
3.5	3.5	1.9354	8.1978	5.0	3.5	1.9978	7.3139
3.5	4.0	2.0148	9.0470	5.0	4.0	2.1213	8.1250
3.5	4.5	2.0514	9.7849	5.0	4.5	2.2106	8.8613
3.5	5.0	2.0556	10.4195	5.0	5.0	2.2687	9.5190
3.5	5.5	2.0366	10.9625	5.0	5.5	2.3000	10.1003
3.5	6.0	2.0016	11.4265	5.0	6.0	2.3094	10.6111
4.0	0.1	0.6883	2.3794	5.5	0.1	0.9968	2.6569
4.0	0.5	0.7470	2.5697	5.5	0.5	1.0309	2.7856
4.0	1.0	0.9147	3.1293	5.5	1.0	1.1317	3.1717
4.0	1.5	1.1489	3.9581	5.5	1.5	1.2823	3.7654
4.0	2.0	1.4030	4.9419	5.5	2.0	1.4619	4.5057
4.0	2.5	1.6391	5.9763	5.5	2.5	1.6495	5.3290
4.0	3.0	1.8346	6.9844	5.5	3.0	1.8278	6.1790
4.0	3.5	1.9811	7.9197	5.5	3.5	1.9848	7.0127
4.0	4.0	2.0798	8.7601	5.5	4.0	2.1144	7.8015
4.0	4.5	2.1369	9.5003	5.5	4.5	2.2143	8.5289
4.0	5.0	2.1605	10.1442	5.5	5.0	2.2858	9.1877
4.0	5.5	2.1584	10.7006	5.5	5.5	2.3316	9.7774
4.0	6.0	2.1373	11.1802	5.5	6.0	2.3554	10.3011
4.5	0.1	0.8134	2.4826	6.0	0.1	1.0640	2.7268
4.5	0.5	0.8618	2.6501	6.0	0.5	1.0931	2.8401
4.5	1.0	1.0020	3.1460	6.0	1.0	1.1798	3.1815
4.5	1.5	1.2033	3.8906	6.0	1.5	1.3112	3.7115
4.5	2.0	1.4301	4.7902	6.0	2.0	1.4712	4.3813
4.5	2.5	1.6510	5.7548	6.0	2.5	1.6428	5.1377
4.5	3.0	1.8443	6.7137	6.0	3.0	1.8110	5.9318
4.5	3.5	1.9988	7.6200	6.0	3.5	1.9647	6.7241
4.5	4.0	2.1122	8.4483	6.0	4.0	2.0969	7.4861
4.5	4.5	2.1872	9.1885	6.0	4.5	2.2042	8.1996
4.5	5.0	2.2291	9.8408	6.0	5.0	2.2862	8.8551
4.5	5.5	2.2442	10.4107	6.0	5.5	2.3443	9.4491
4.5	6.0	2.2383	10.9063	6.0	6.0	2.3811	9.9825

Table 56.--Values of Skewness and Kurtosis Measures for Various Settings of Mean (μ) and Standard Deviation (σ) with Contaminating Fraction (ϵ) = 0.30

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
0.5	0.1	-0.4589	3.8573	2.0	0.1	-0.3002	2.0541
0.5	0.5	-0.2999	3.3745	2.0	0.5	-0.1330	2.2212
0.5	1.0	0.0097	2.9969	2.0	1.0	0.2692	2.7420
0.5	1.5	0.2370	3.5587	2.0	1.5	0.6816	3.5362
0.5	2.0	0.3502	4.5856	2.0	2.0	0.9816	4.4445
0.5	2.5	0.3908	5.6105	2.0	2.5	1.1547	5.3216
0.5	3.0	0.3945	6.4669	2.0	3.0	1.2315	6.0913
0.5	3.5	0.3815	7.1396	2.0	3.5	1.2467	6.7336
0.5	4.0	0.3619	7.6589	2.0	4.0	1.2260	7.2571
0.5	4.5	0.3405	8.0601	2.0	4.5	1.1862	7.6801
0.5	5.0	0.3193	8.3727	2.0	5.0	1.1373	8.0218
0.5	5.5	0.2994	8.6190	2.0	5.5	1.0851	8.2991
0.5	6.0	0.2810	8.8156	2.0	6.0	1.0328	8.5257
1.0	0.1	-0.6187	3.0767	2.5	0.1	-0.0862	1.8593
1.0	0.5	-0.3974	2.9194	2.5	0.5	0.0435	2.0497
1.0	1.0	0.0631	2.9627	2.5	1.0	0.3732	2.6012
1.0	1.5	0.4367	3.6209	2.5	1.5	0.7448	3.3862
1.0	2.0	0.6441	4.6009	2.5	2.0	1.0486	4.2584
1.0	2.5	0.7297	5.5729	2.5	2.5	1.2500	5.1021
1.0	3.0	0.7470	6.3991	2.5	3.0	1.3600	5.8543
1.0	3.5	0.7305	7.0600	2.5	3.5	1.4031	6.4945
1.0	4.0	0.6987	7.5778	2.5	4.0	1.4025	7.0264
1.0	4.5	0.6615	7.9826	2.5	4.5	1.3753	7.4637
1.0	5.0	0.6234	8.3009	2.5	5.0	1.3331	7.8222
1.0	5.5	0.5867	8.5536	2.5	5.5	1.2834	8.1167
1.0	6.0	0.5524	8.7564	2.5	6.0	1.2306	8.3598
1.5	0.1	-0.5116	2.4344	3.0	0.1	0.0951	1.7662
1.5	0.5	-0.3052	2.5036	3.0	0.5	0.1955	1.9482
1.5	1.0	0.1587	2.8725	3.0	1.0	0.4616	2.4705
1.5	1.5	0.5833	3.6227	3.0	1.5	0.7849	3.2094
1.5	2.0	0.8534	4.5626	3.0	2.0	1.0759	4.0342
1.5	2.5	0.9858	5.4810	3.0	2.5	1.2921	4.8437
1.5	3.0	1.0293	6.2752	3.0	3.0	1.4291	5.5795
1.5	3.5	1.0227	6.9244	3.0	3.5	1.5005	6.2189
1.5	4.0	0.9906	7.4430	3.0	4.0	1.5241	6.7605
1.5	4.5	0.9469	7.8549	3.0	4.5	1.5153	7.2134
1.5	5.0	0.8991	8.1829	3.0	5.0	1.4860	7.5902
1.5	5.5	0.8513	8.4461	3.0	5.5	1.4445	7.9037
1.5	6.0	0.8052	8.6590	3.0	6.0	1.3964	8.1652

Table 56.--Continued.

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
3.5	0.1	0.2393	1.7242	5.0	0.1	0.5082	1.7025
3.5	0.5	0.3180	1.8872	5.0	0.5	0.5502	1.8094
3.5	1.0	0.5334	2.3580	5.0	1.0	0.6720	2.1264
3.5	1.5	0.8106	3.0326	5.0	1.5	0.8467	2.6038
3.5	2.0	1.0801	3.7998	5.0	2.0	1.0435	3.1828
3.5	2.5	1.2996	4.5694	5.0	2.5	1.2352	3.8066
3.5	3.0	1.4553	5.2853	5.0	3.0	1.4033	4.4300
3.5	3.5	1.5513	5.9212	5.0	3.5	1.5384	5.0224
3.5	4.0	1.5991	6.4707	5.0	4.0	1.6385	5.5668
3.5	4.5	1.6113	6.9383	5.0	4.5	1.7060	6.0559
3.5	5.0	1.5986	7.3333	5.0	5.0	1.7455	6.4889
3.5	5.5	1.5695	7.6661	5.0	5.5	1.7623	6.8690
3.5	6.0	1.5301	7.9468	5.0	6.0	1.7614	7.2009
4.0	0.1	0.3518	1.7071	5.5	0.1	0.5627	1.7055
4.0	0.5	0.4146	1.8495	5.5	0.5	0.5979	1.7984
4.0	1.0	0.5905	2.2647	5.5	1.0	0.7010	2.0758
4.0	1.5	0.8275	2.8701	5.5	1.5	0.8522	2.4989
4.0	2.0	1.0723	3.5741	5.5	2.0	1.0280	3.0212
4.0	2.5	1.2868	4.2980	5.5	2.5	1.2059	3.5953
4.0	3.0	1.4529	4.9882	5.5	3.0	1.3690	4.1810
4.0	3.5	1.5677	5.6153	5.5	3.5	1.5072	4.7491
4.0	4.0	1.6372	6.1686	5.5	4.0	1.6162	5.2813
4.0	4.5	1.6701	6.6479	5.5	4.5	1.6962	5.7678
4.0	5.0	1.6756	7.0591	5.5	5.0	1.7497	6.2054
4.0	5.5	1.6613	7.4103	5.5	5.5	1.7807	6.5948
4.0	6.0	1.6336	7.7098	5.5	6.0	1.7930	6.9390
4.5	0.1	0.4395	1.7020	6.0	0.1	0.6063	1.7096
4.5	0.5	0.4904	1.8254	6.0	0.5	0.6362	1.7907
4.5	1.0	0.6359	2.1885	6.0	1.0	0.7245	2.0343
4.5	1.5	0.8388	2.7269	6.0	1.5	0.8563	2.4101
4.5	2.0	1.0589	3.3672	6.0	2.0	1.0133	2.8809
4.5	2.5	1.2634	4.0417	6.0	2.5	1.1774	3.4075
4.5	3.0	1.4330	4.7004	6.0	3.0	1.3335	3.9550
4.5	3.5	1.5606	5.3130	6.0	3.5	1.4714	4.4960
4.5	4.0	1.6476	5.8645	6.0	4.0	1.5858	5.0119
4.5	4.5	1.6993	6.3512	6.0	4.5	1.6750	5.4914
4.5	5.0	1.7226	6.7754	6.0	5.0	1.7400	5.9293
4.5	5.5	1.7241	7.1426	6.0	5.5	1.7833	6.3243
4.5	6.0	1.7095	7.4594	6.0	6.0	1.8080	6.6776

Table 57.--Values of Skewness and Kurtosis Measures for Various Settings of Mean (μ) and Standard Deviation (σ) with Contaminating Fraction (ϵ) = 0.40

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
0.5	0.1	-0.6476	4.4239	2.0	0.1	-0.5325	2.1315
0.5	0.5	-0.3985	3.5963	2.0	0.5	-0.3254	2.2203
0.5	1.0	0.0055	2.9941	2.0	1.0	0.1399	2.5602
0.5	1.5	0.2340	3.4965	2.0	1.5	0.5660	3.1447
0.5	2.0	0.3196	4.3097	2.0	2.0	0.8374	3.8258
0.5	2.5	0.3375	5.0246	2.0	2.5	0.9711	4.4683
0.5	3.0	0.3282	5.5706	2.0	3.0	1.0156	5.0133
0.5	3.5	0.3094	5.9737	2.0	3.5	1.0100	5.4537
0.5	4.0	0.2882	6.2717	2.0	4.0	0.9789	5.8028
0.5	4.5	0.2675	6.4948	2.0	4.5	0.9361	6.0787
0.5	5.0	0.2484	6.6647	2.0	5.0	0.8891	6.2977
0.5	5.5	0.2312	6.7963	2.0	5.5	0.8419	6.4728
0.5	6.0	0.2157	6.8999	2.0	6.0	0.7965	6.6143
1.0	0.1	-0.8574	3.4421	2.5	0.1	-0.3382	1.8250
1.0	0.5	-0.5399	3.0944	2.5	0.5	-0.1839	1.9525
1.0	1.0	0.0348	2.9313	2.5	1.0	0.1897	2.3400
1.0	1.5	0.4130	3.4556	2.5	1.5	0.5774	2.9167
1.0	2.0	0.5793	4.2158	2.5	2.0	0.8641	3.5665
1.0	2.5	0.6271	4.9050	2.5	2.5	1.0339	4.1895
1.0	3.0	0.6208	5.4490	2.5	3.0	1.1133	4.7345
1.0	3.5	0.5925	5.8609	2.5	3.5	1.1339	5.1888
1.0	4.0	0.5569	6.1710	2.5	4.0	1.1198	5.5588
1.0	4.5	0.5203	6.4063	2.5	4.5	1.0867	5.8578
1.0	5.0	0.4855	6.5873	2.5	5.0	1.0442	6.0995
1.0	5.5	0.4535	6.7286	2.5	5.5	0.9979	6.2956
1.0	6.0	0.4244	6.8405	2.5	6.0	0.9512	6.4559
1.5	0.1	-0.7414	2.6405	3.0	0.1	-0.1833	1.6369
1.5	0.5	-0.4693	2.5996	3.0	0.5	-0.0670	1.7661
1.5	1.0	0.0848	2.7746	3.0	1.0	0.2307	2.1434
1.5	1.5	0.5189	3.3365	3.0	1.5	0.5707	2.6873
1.5	2.0	0.7501	4.0509	3.0	2.0	0.8541	3.3000
1.5	2.5	0.8398	4.7142	3.0	2.5	1.0474	3.8999
1.5	3.0	0.8529	5.2579	3.0	3.0	1.1582	4.4404
1.5	3.5	0.8294	5.6830	3.0	3.5	1.2073	4.9042
1.5	4.0	0.7903	6.0111	3.0	4.0	1.2154	5.2922
1.5	4.5	0.7460	6.2648	3.0	4.5	1.1980	5.6128
1.5	5.0	0.7015	6.4628	3.0	5.0	1.1658	5.8767
1.5	5.5	0.6591	6.6191	3.0	5.5	1.1256	6.0942
1.5	6.0	0.6197	6.7440	3.0	6.0	1.0818	6.2743

Table 57.--Continued.

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
3.5	0.1	-0.0655	1.5164	5.0	0.1	0.1435	1.3394
3.5	0.5	0.0242	1.6349	5.0	0.5	0.1903	1.4185
3.5	1.0	0.2631	1.9792	5.0	1.0	0.3240	1.6531
3.5	1.5	0.5567	2.4769	5.0	1.5	0.5112	2.0067
3.5	2.0	0.8254	3.0461	5.0	2.0	0.7155	2.4360
3.5	2.5	1.0299	3.6173	5.0	2.5	0.9071	2.8991
3.5	3.0	1.1645	4.1467	5.0	3.0	1.0683	3.3622
3.5	3.5	1.2401	4.6140	5.0	3.5	1.1924	3.8025
3.5	4.0	1.2719	5.0148	5.0	4.0	1.2801	4.2071
3.5	4.5	1.2733	5.3535	5.0	4.5	1.3360	4.5707
3.5	5.0	1.2552	5.6375	5.0	5.0	1.3662	4.8926
3.5	5.5	1.2250	5.8754	5.0	5.5	1.3764	5.1751
3.5	6.0	1.1878	6.0749	5.0	6.0	1.3717	5.4218
4.0	0.1	0.0236	1.4359	5.5	0.1	0.1844	1.3094
4.0	0.5	0.0943	1.5404	5.5	0.5	0.2234	1.3782
4.0	1.0	0.2885	1.8460	5.5	1.0	0.3364	1.5837
4.0	1.5	0.5407	2.2934	5.5	1.5	0.4989	1.8973
4.0	2.0	0.7890	2.8155	5.5	2.0	0.6828	2.2847
4.0	2.5	0.9950	3.3530	5.5	2.5	0.8630	2.7107
4.0	3.0	1.1454	3.8649	5.5	3.0	1.0224	3.1458
4.0	3.5	1.2427	4.3289	5.5	3.5	1.1526	3.5681
4.0	4.0	1.2965	4.7368	5.5	4.0	1.2514	3.9641
4.0	4.5	1.3176	5.0889	5.5	4.5	1.3208	4.3265
4.0	5.0	1.3156	5.3897	5.5	5.0	1.3650	4.6527
4.0	5.5	1.2980	5.6456	5.5	5.5	1.3885	4.9433
4.0	6.0	1.2702	5.8631	5.5	6.0	1.3957	5.2003
4.5	0.1	0.0913	1.3798	6.0	0.1	0.2167	1.2866
4.5	0.5	0.1483	1.4708	6.0	0.5	0.2497	1.3467
4.5	1.0	0.3083	1.7390	6.0	1.0	0.3464	1.5273
4.5	1.5	0.5252	2.1375	6.0	1.5	0.4883	1.8059
4.5	2.0	0.7513	2.6123	6.0	2.0	0.6536	2.1552
4.5	2.5	0.9522	3.1130	6.0	2.5	0.8216	2.5460
4.5	3.0	1.1109	3.6022	6.0	3.0	0.9765	2.9525
4.5	3.5	1.2243	4.0569	6.0	3.5	1.1090	3.3546
4.5	4.0	1.2970	4.4658	6.0	4.0	1.2153	3.7386
4.5	4.5	1.3367	4.8261	6.0	4.5	1.2953	4.0960
4.5	5.0	1.3512	5.1396	6.0	5.0	1.3514	4.4228
4.5	5.5	1.3473	5.4105	6.0	5.5	1.3870	4.7181
4.5	6.0	1.3307	5.6439	6.0	6.0	1.4057	4.9827

Table 58.--Values of Skewness and Kurtosis Measures for Various Settings of Mean (μ) and Standard Deviation (σ) with Contaminating Fraction (ϵ) = 0.50

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
0.5	0.1	-0.8684	5.2582	2.0	0.1	-0.8043	2.4415
0.5	0.5	-0.4934	3.8760	2.0	0.5	-0.5431	2.4024
0.5	1.0	0.0000	2.9931	2.0	1.0	0.0000	2.5000
0.5	1.5	0.2138	3.4088	2.0	1.5	0.4409	2.8798
0.5	2.0	0.2743	4.0268	2.0	2.0	0.6872	3.3878
0.5	2.5	0.2780	4.5197	2.0	2.5	0.7917	3.8729
0.5	3.0	0.2634	4.8726	2.0	3.0	0.8165	4.2778
0.5	3.5	0.2439	5.1223	2.0	3.5	0.8015	4.5982
0.5	4.0	0.2245	5.3016	2.0	4.0	0.7684	4.8476
0.5	4.5	0.2066	5.4331	2.0	4.5	0.7285	5.0417
0.5	5.0	0.1906	5.5318	2.0	5.0	0.6872	5.1939
0.5	5.5	0.1765	5.6074	2.0	5.5	0.6473	5.3144
0.5	6.0	0.1641	5.6664	2.0	6.0	0.6097	5.4109
1.0	0.1	-1.1318	4.0703	2.5	0.1	-0.6244	2.0297
1.0	0.5	-0.6872	3.3878	2.5	0.5	-0.4347	2.0678
1.0	1.0	0.0000	2.9200	2.5	1.0	0.0000	2.2564
1.0	1.5	0.3651	3.2978	2.5	1.5	0.4118	2.6348
1.0	2.0	0.4934	3.8760	2.5	2.0	0.6870	3.1131
1.0	2.5	0.5162	4.3684	2.5	2.5	0.8332	3.5867
1.0	3.0	0.4988	4.7370	2.5	3.0	0.8923	4.0012
1.0	3.5	0.4681	5.0056	2.5	3.5	0.9004	4.3432
1.0	4.0	0.4347	5.2024	2.5	4.0	0.8811	4.6184
1.0	4.5	0.4026	5.3489	2.5	4.5	0.8483	4.8382
1.0	5.0	0.3732	5.4600	2.5	5.0	0.8098	5.0141
1.0	5.5	0.3468	5.5457	2.5	5.5	0.7697	5.1556
1.0	6.0	0.3233	5.6130	2.5	6.0	0.7303	5.2705
1.5	0.1	-1.0098	3.0897	3.0	0.1	-0.4871	1.7629
1.5	0.5	-0.6520	2.8504	3.0	0.5	-0.3462	1.8261
1.5	1.0	0.0000	2.7408	3.0	1.0	0.0000	2.0414
1.5	1.5	0.4347	3.1127	3.0	1.5	0.3687	2.4037
1.5	2.0	0.6297	3.6522	3.0	2.0	0.6520	2.8504
1.5	2.5	0.6893	4.1428	3.0	2.5	0.8295	3.3056
1.5	3.0	0.6860	4.5309	3.0	3.0	0.9221	3.7206
1.5	3.5	0.6568	4.8252	3.0	3.5	0.9574	4.0766
1.5	4.0	0.6185	5.0470	3.0	4.0	0.9575	4.3726
1.5	4.5	0.5787	5.2155	3.0	4.5	0.9375	4.6155
1.5	5.0	0.5406	5.3451	3.0	5.0	0.9068	4.8140
1.5	5.5	0.5053	5.4464	3.0	5.5	0.8708	4.9766
1.5	6.0	0.4731	5.5266	3.0	6.0	0.8332	5.1103

Table 58.--Continued.

μ	σ	Skew-ness	Kurto-sis	μ	σ	Skew-ness	Kurto-sis
3.5	0.1	-0.3857	1.5839	5.0	0.1	-0.2115	1.3040
3.5	0.5	-0.2780	1.6515	5.0	0.5	-0.1560	1.3560
3.5	1.0	0.0000	1.8634	5.0	1.0	0.0000	1.5137
3.5	1.5	0.3233	2.1996	5.0	1.5	0.2121	1.7591
3.5	2.0	0.6003	2.6119	5.0	2.0	0.4347	2.0678
3.5	2.5	0.7969	3.0428	5.0	2.5	0.6344	2.4108
3.5	3.0	0.9173	3.4499	5.0	3.0	0.7950	2.7620
3.5	3.5	0.9794	3.8116	5.0	3.5	0.9132	3.1013
3.5	4.0	1.0015	4.1219	5.0	4.0	0.9930	3.4165
3.5	4.5	0.9979	4.3833	5.0	4.5	1.0413	3.7016
3.5	5.0	0.9786	4.6017	5.0	5.0	1.0656	3.9550
3.5	5.5	0.9504	4.7837	5.0	5.5	1.0721	4.1777
3.5	6.0	0.9176	4.9357	5.0	6.0	1.0659	4.3723
4.0	0.1	-0.3106	1.4595	5.5	0.1	-0.1782	1.2538
4.0	0.5	-0.2262	1.5237	5.5	0.5	-0.1321	1.3000
4.0	1.0	0.0000	1.7200	5.5	1.0	0.0000	1.4399
4.0	1.5	0.2811	2.0257	5.5	1.5	0.1852	1.6588
4.0	2.0	0.5431	2.4024	5.5	2.0	0.3877	1.9370
4.0	2.5	0.7480	2.8052	5.5	2.5	0.5787	2.2513
4.0	3.0	0.8889	3.1975	5.5	3.0	0.7411	2.5794
4.0	3.5	0.9745	3.5574	5.5	3.5	0.8684	2.9033
4.0	4.0	1.0182	3.8752	5.5	4.0	0.9612	3.2107
4.0	4.5	1.0325	4.1498	5.5	4.5	1.0238	3.4944
4.0	5.0	1.0272	4.3841	5.5	5.0	1.0617	3.7512
4.0	5.5	1.0093	4.5830	5.5	5.5	1.0806	3.9807
4.0	6.0	0.9838	4.7516	5.5	6.0	1.0851	4.1842
4.5	0.1	-0.2543	1.3701	6.0	0.1	-0.1520	1.2150
4.5	0.5	-0.1866	1.4285	6.0	0.5	-0.1130	1.2559
4.5	1.0	0.0000	1.6054	6.0	1.0	0.0000	1.3800
4.5	1.5	0.2439	1.8801	6.0	1.5	0.1624	1.5754
4.5	2.0	0.4869	2.2218	6.0	2.0	0.3462	1.8261
4.5	2.5	0.6920	2.5947	6.0	2.5	0.5267	2.1133
4.5	3.0	0.8459	2.9678	6.0	3.0	0.6872	2.4184
4.5	3.5	0.9503	3.3197	6.0	3.5	0.8197	2.7252
4.5	4.0	1.0136	3.6387	6.0	4.0	0.9220	3.0220
4.5	4.5	1.0456	3.9210	6.0	4.5	0.9964	3.3010
4.5	5.0	1.0552	4.1670	6.0	5.0	1.0466	3.5579
4.5	5.5	1.0492	4.3796	6.0	5.5	1.0771	3.7910
4.5	6.0	1.0328	4.5625	6.0	6.0	1.0921	4.0007

APPENDIX B

ILLUSTRATIONS

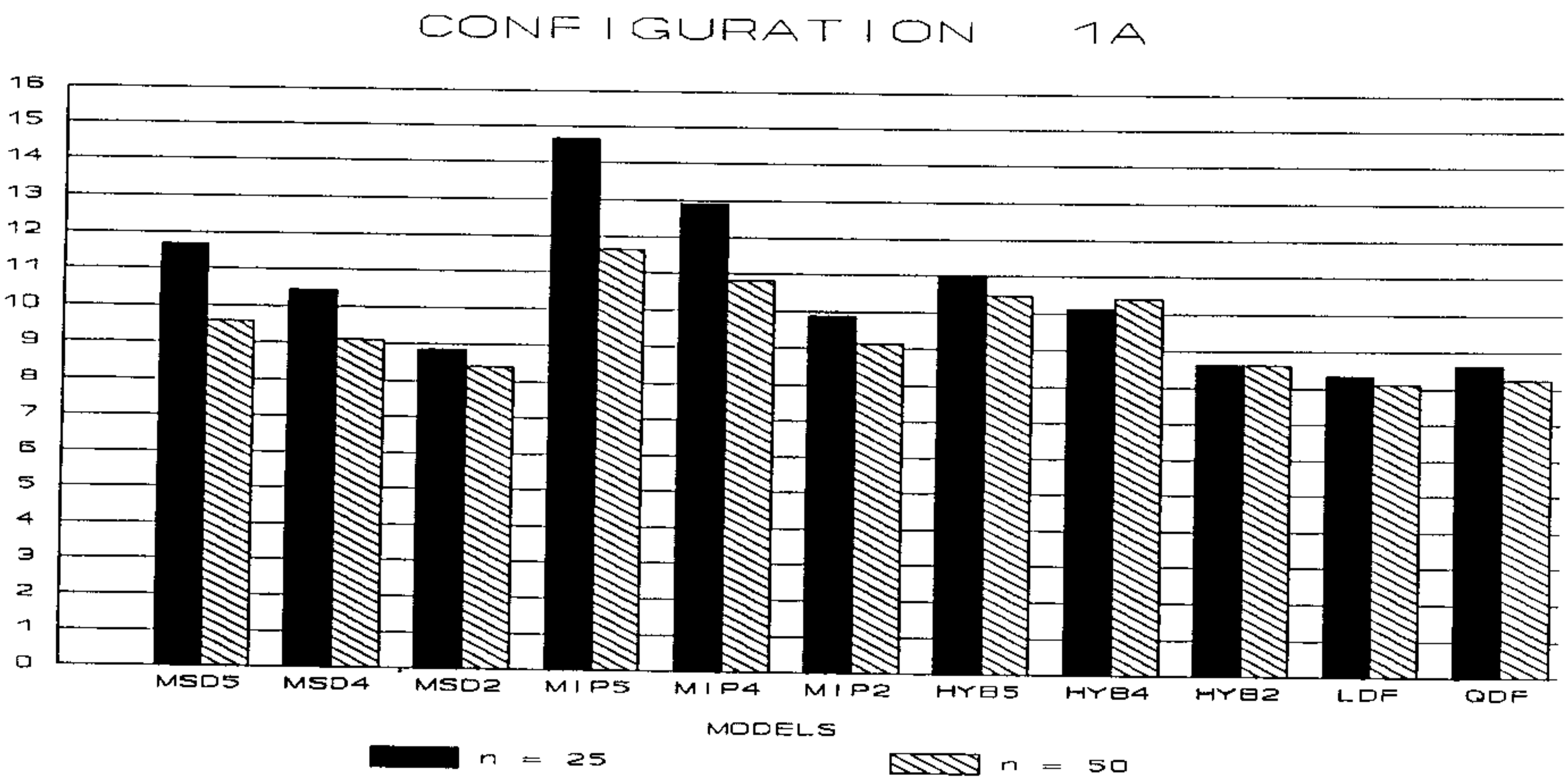

Fig. 1. Percentages of Misclassification on Validation Samples for Configuration 1A

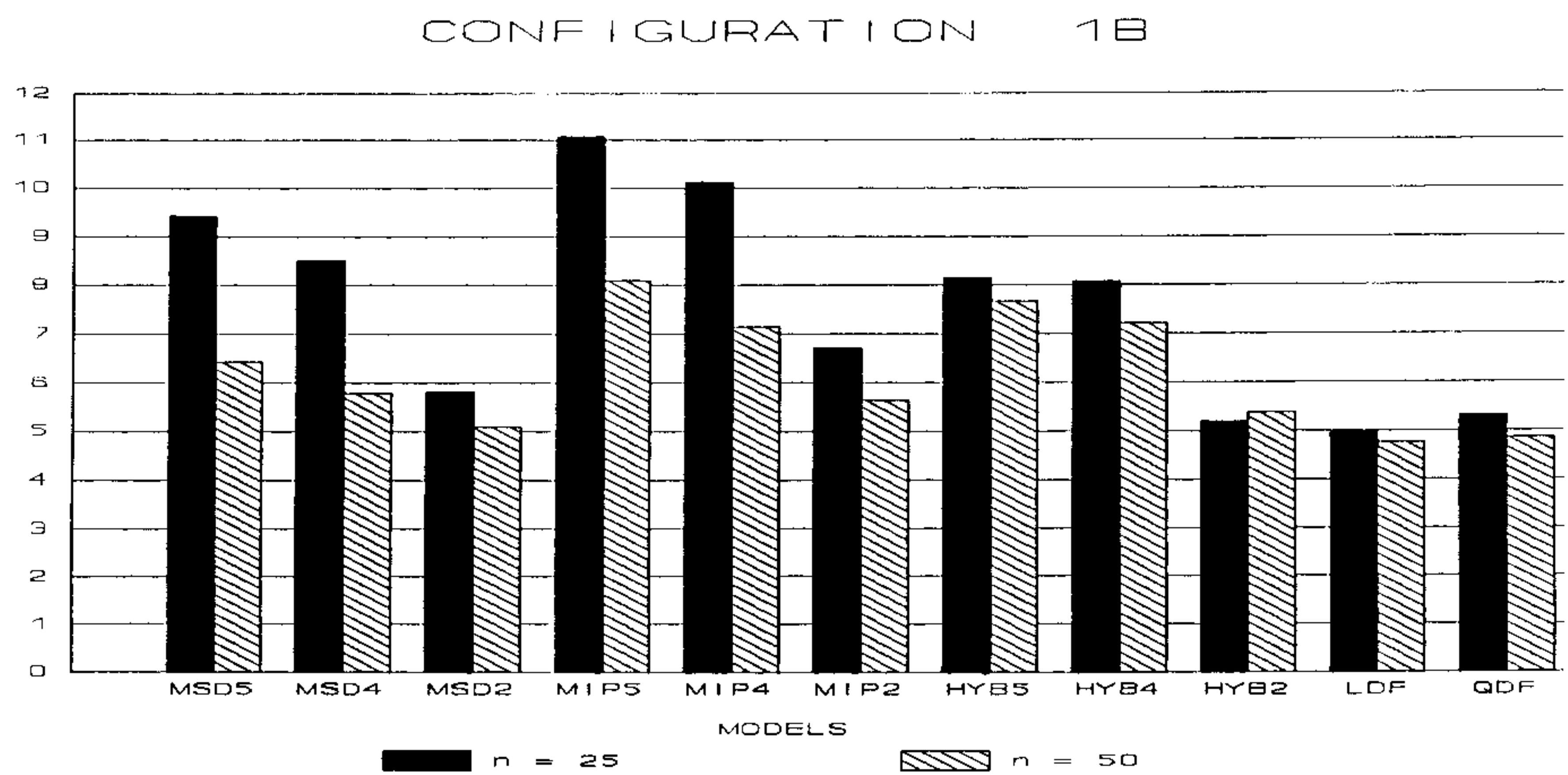

Fig. 2. Percentages of Misclassification on Validation Samples for Configuration 1B

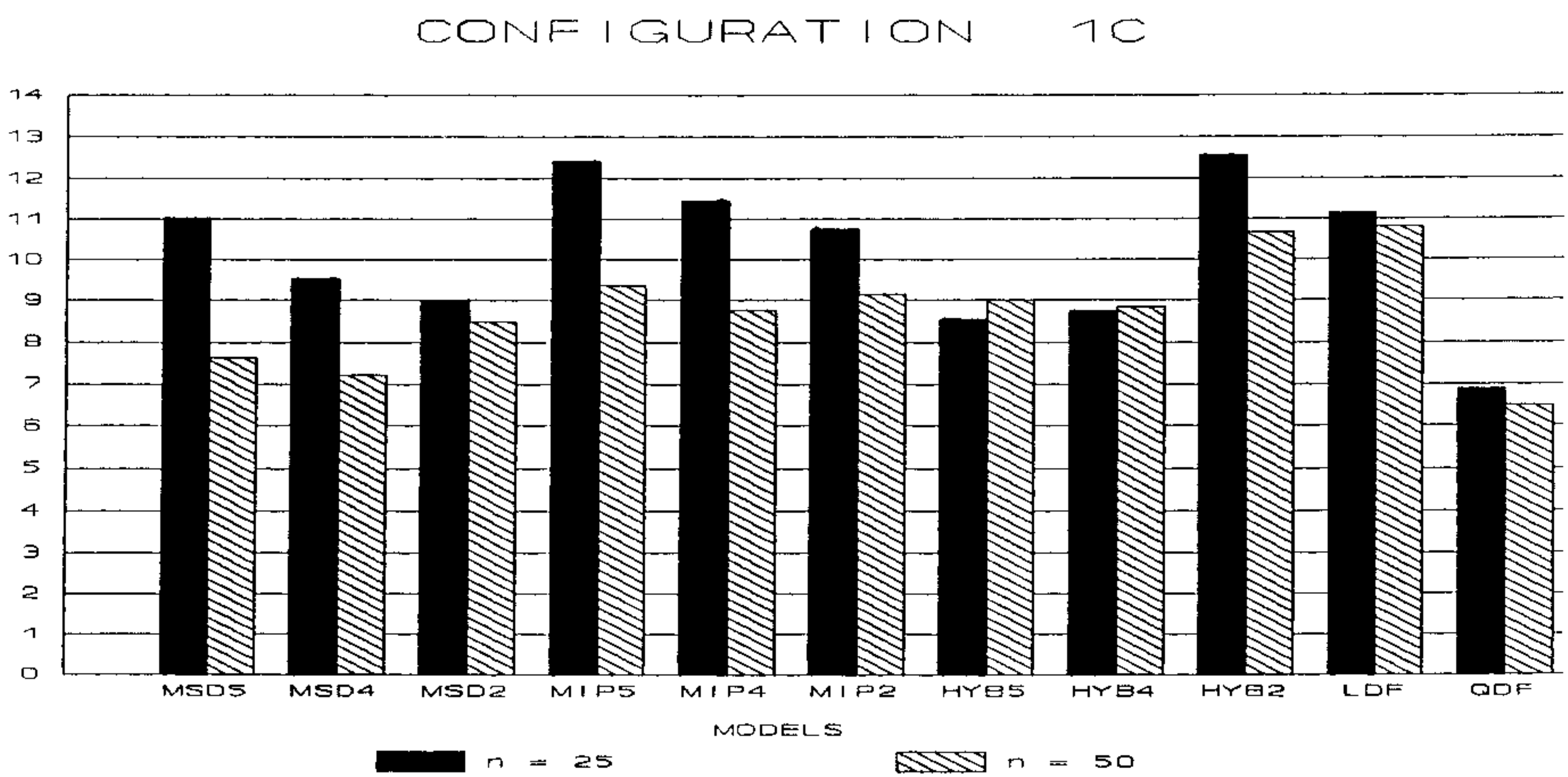

Fig. 3. Percentages of Misclassification on Validation Samples for Configuration 1C

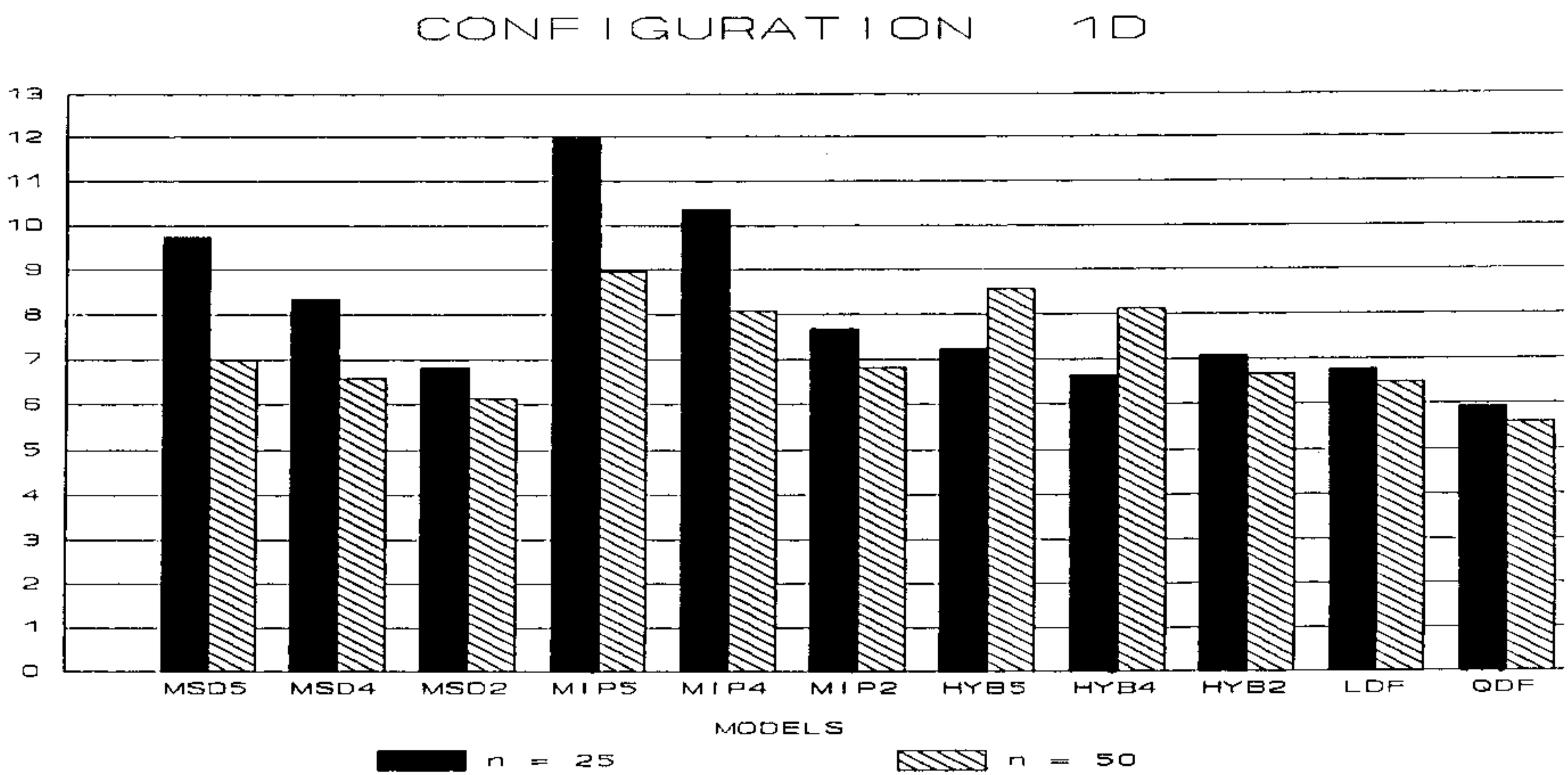

Fig. 4. Percentages of Misclassification on Validation Samples for Configuration 1D

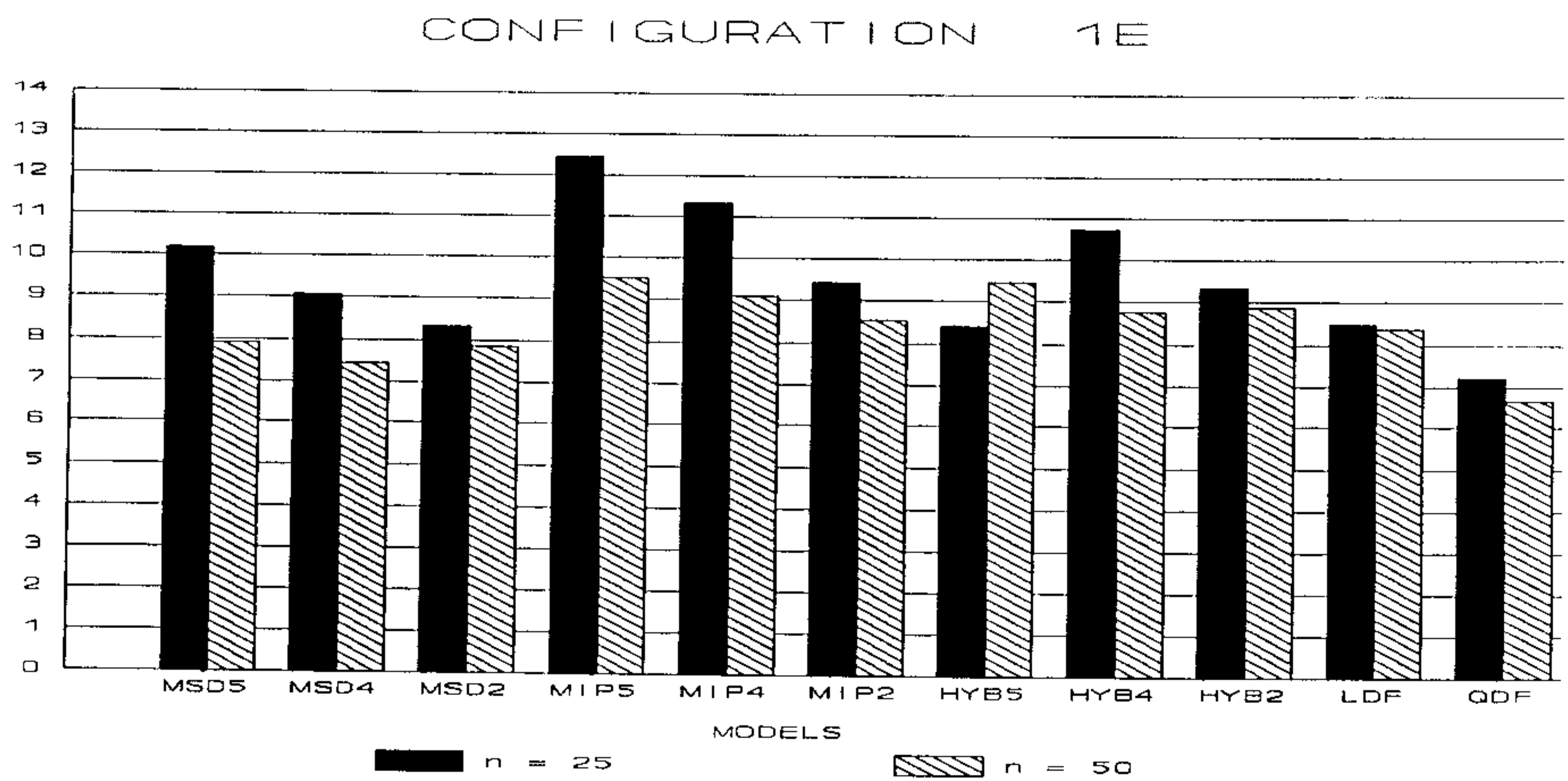

Fig. 5. Percentages of Misclassification on Validation Samples for Configuration 1E

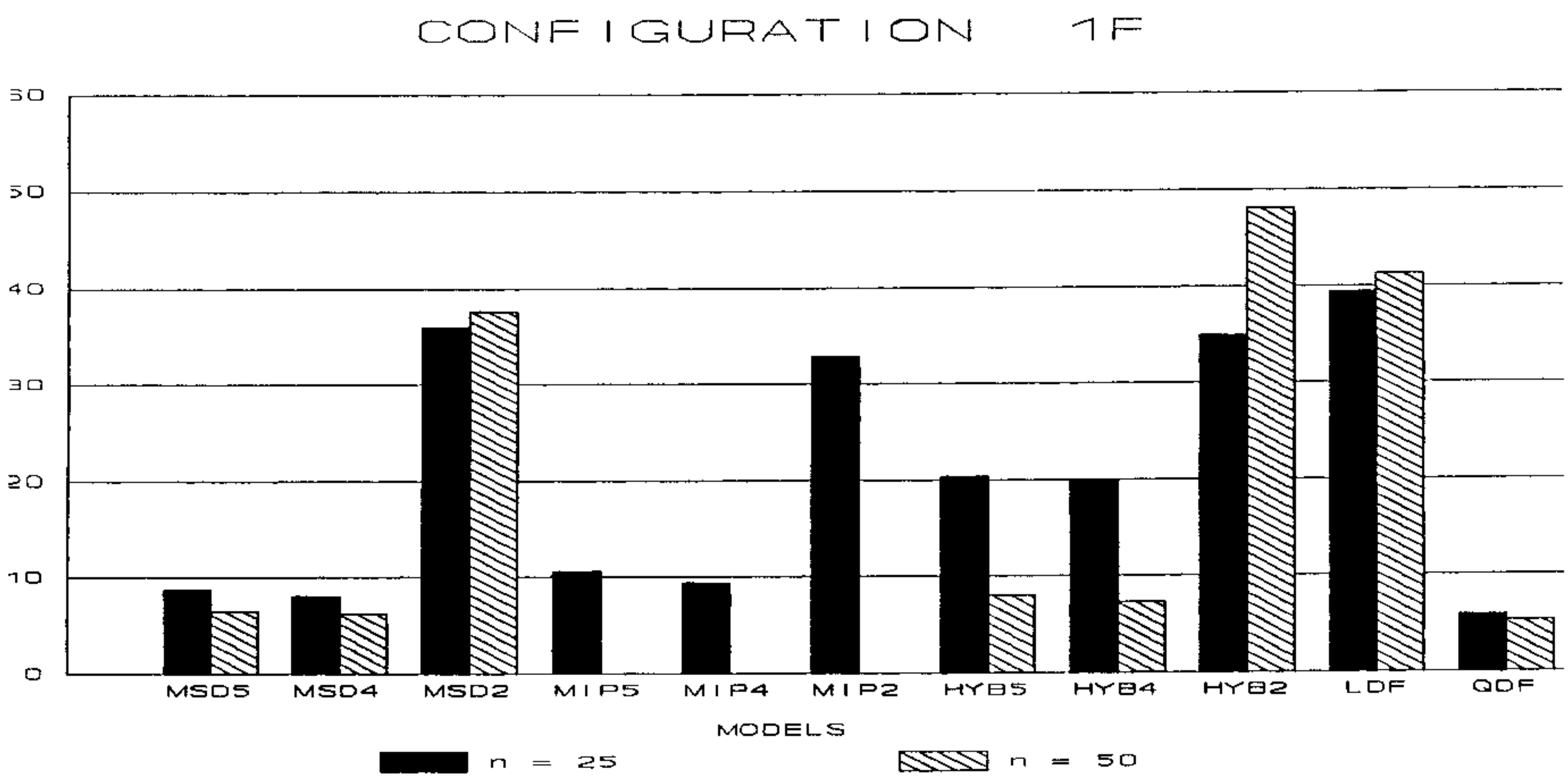

Fig. 6. Percentages of Misclassification on Validation Samples for Configuration 1F

Note: Computationally too intensive to complete runs for MIP models with n = 50.

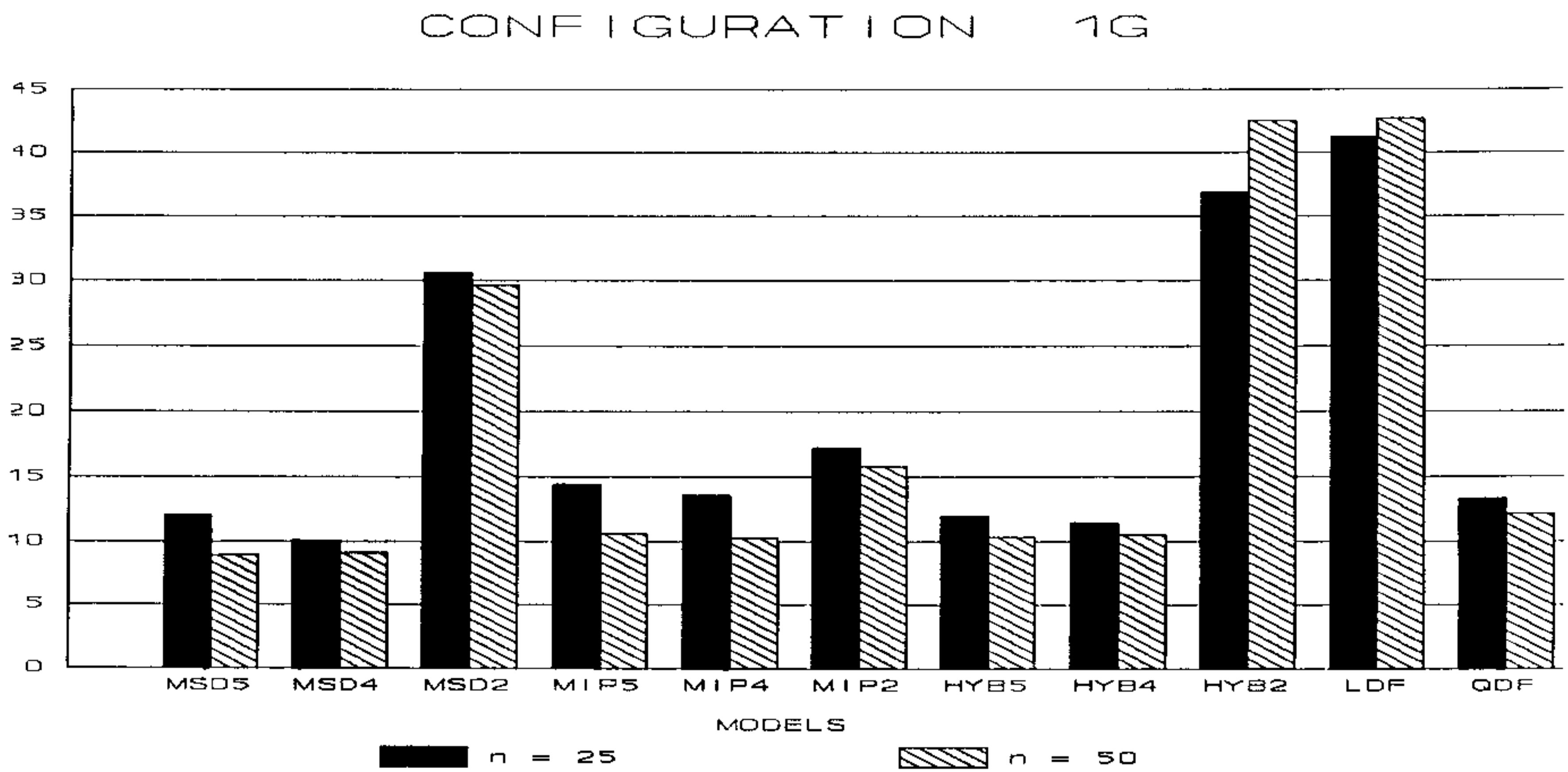

Fig. 7. Percentages of Misclassification on Validation Samples for Configuration 1G

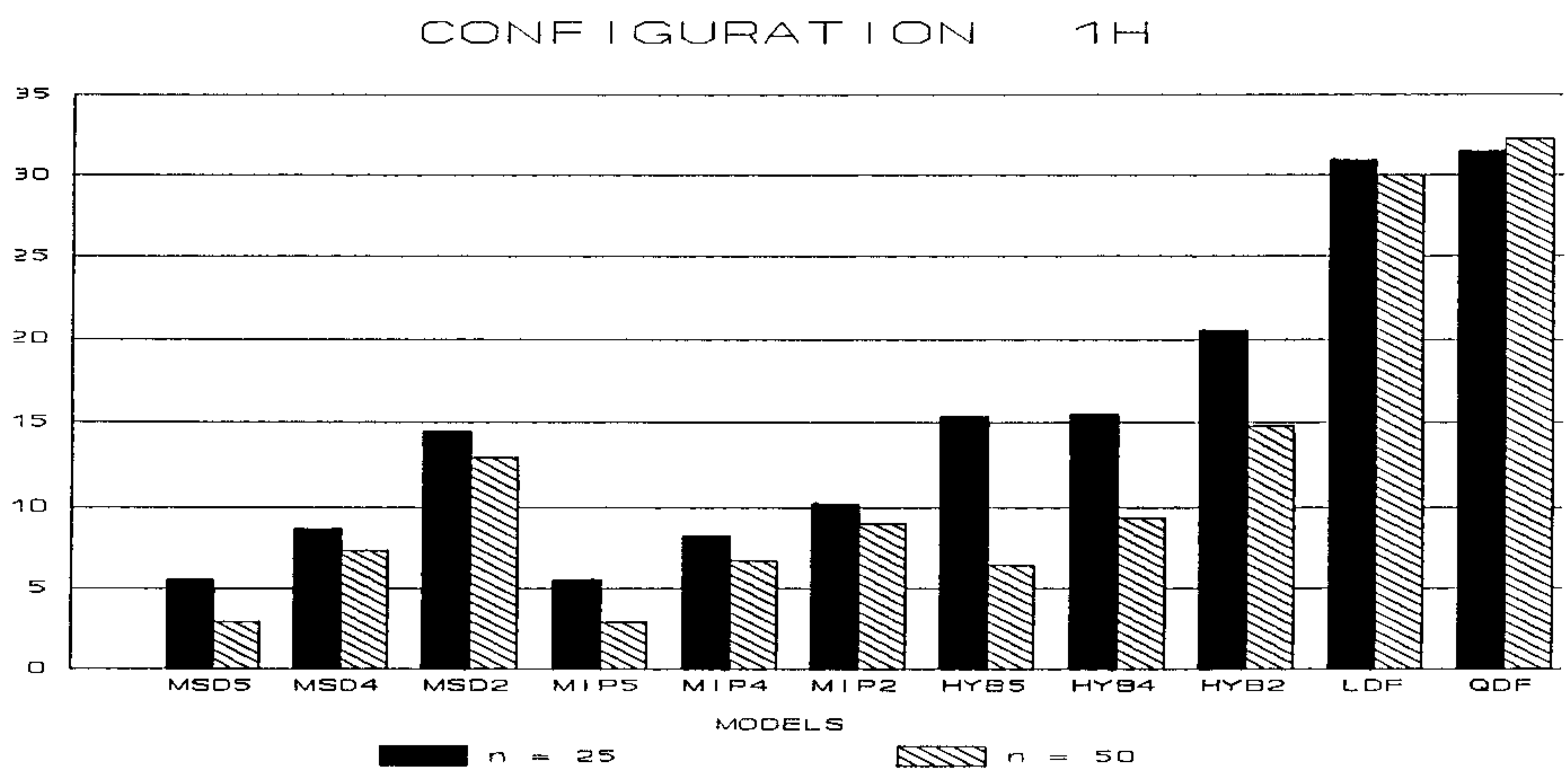

Fig. 8. Percentages of Misclassification on Validation Samples for Configuration 1H

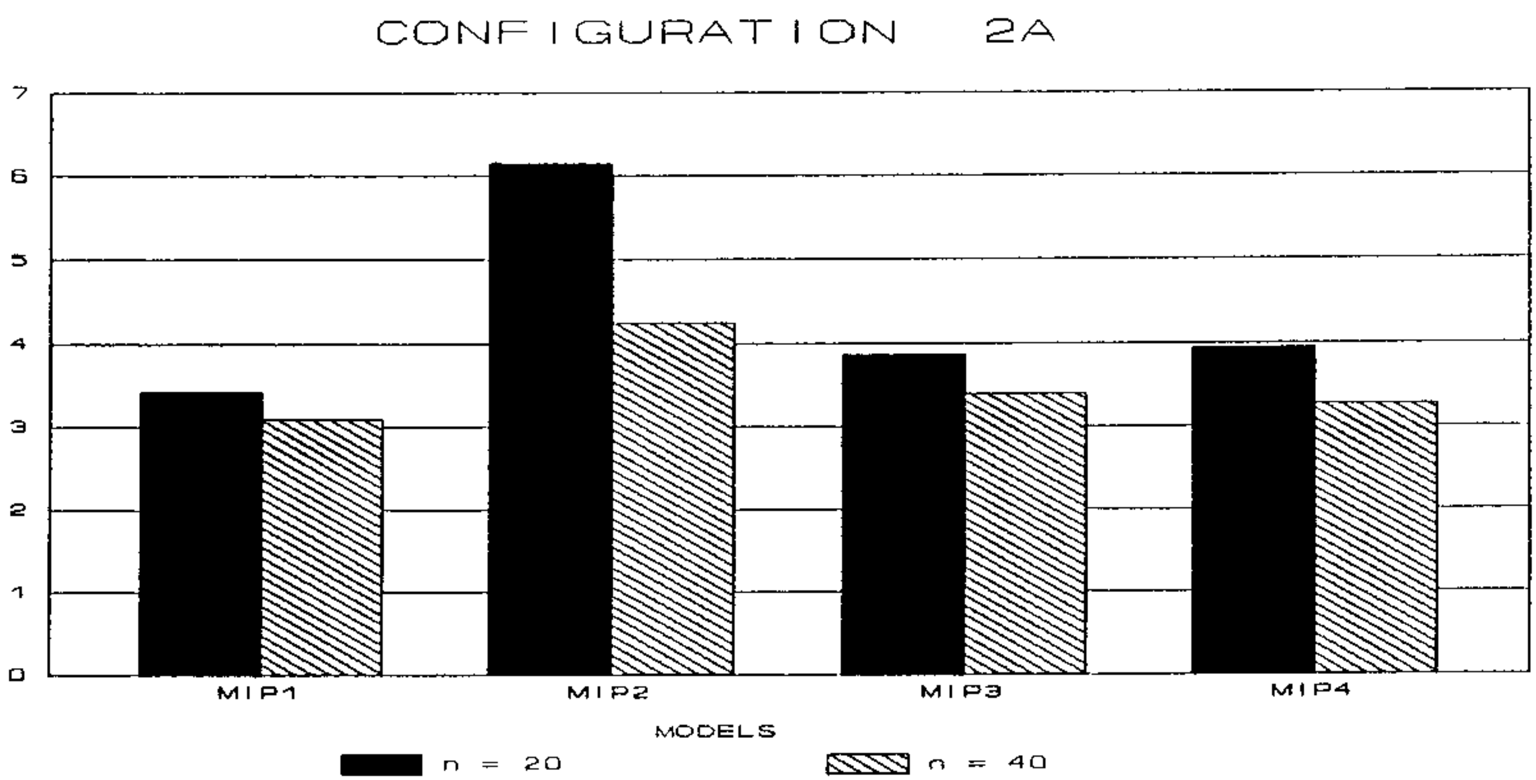

Fig. 9. Percentages of Exact Misclassification for Configuration 2A

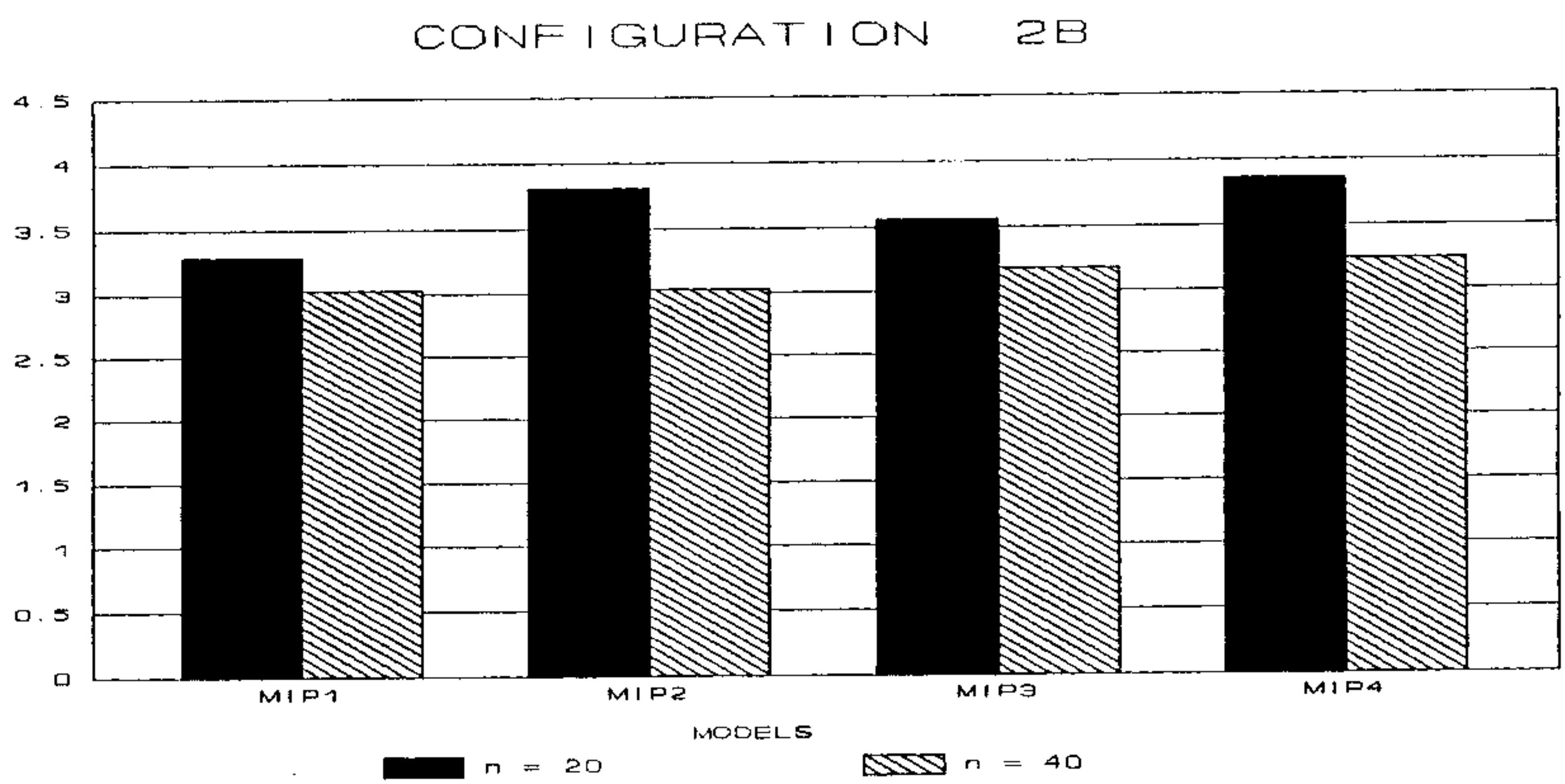

Fig. 10. Percentages of Exact Misclassification for Configuration 2B

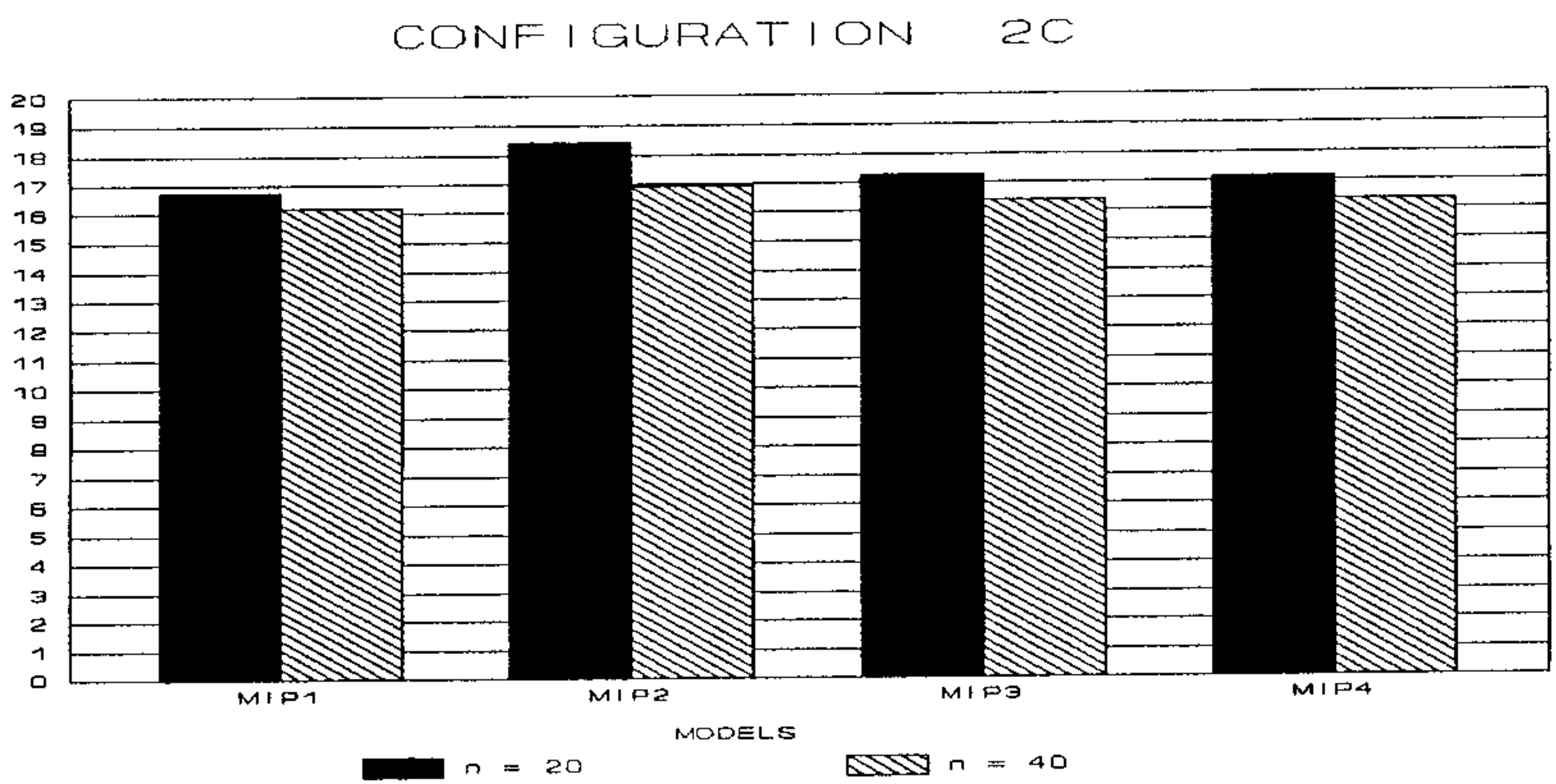

Fig. 11. Percentages of Exact Misclassification for Configuration 2C

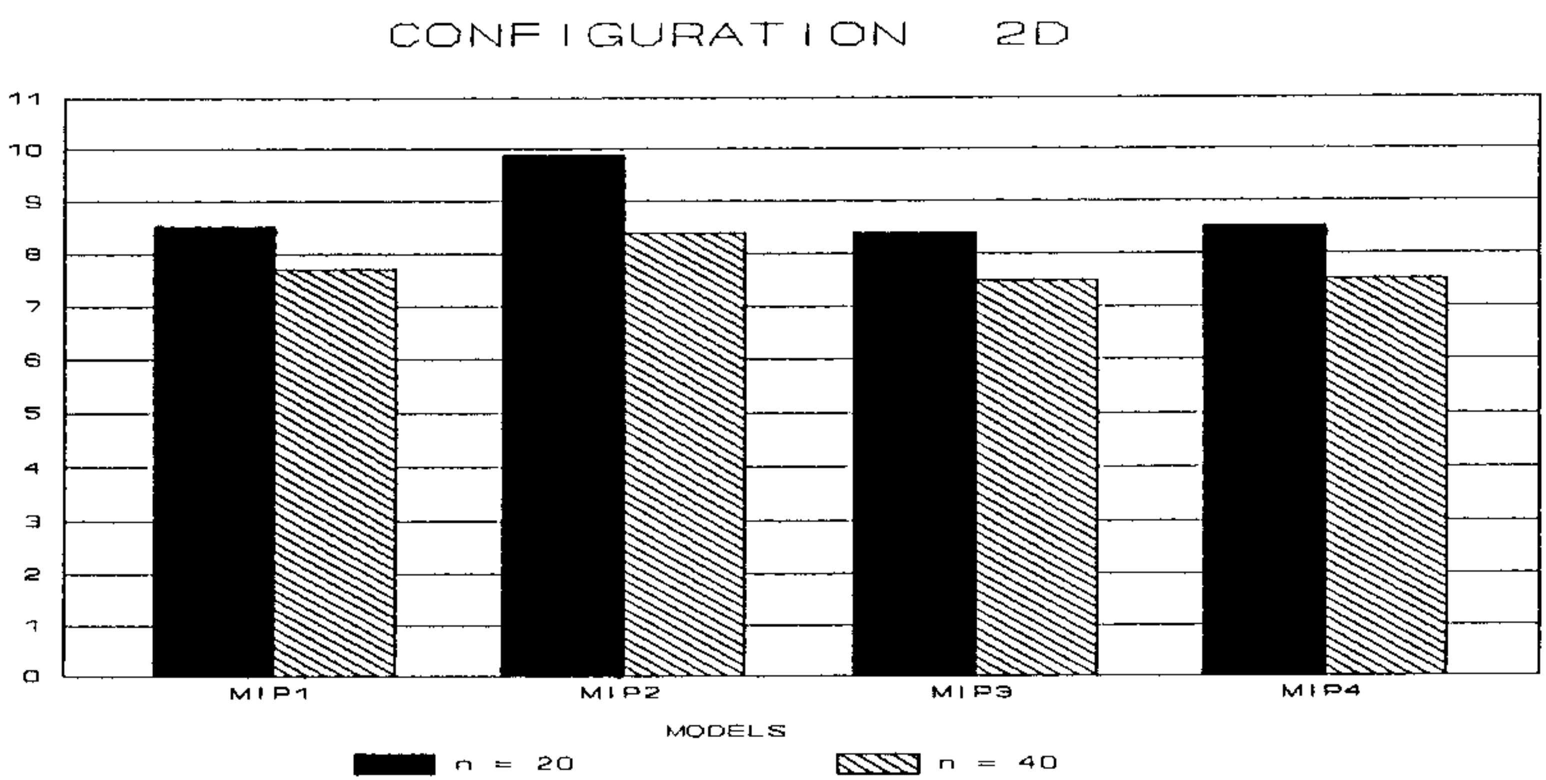

Fig. 12. Percentages of Exact Misclassification for Configuration 2D

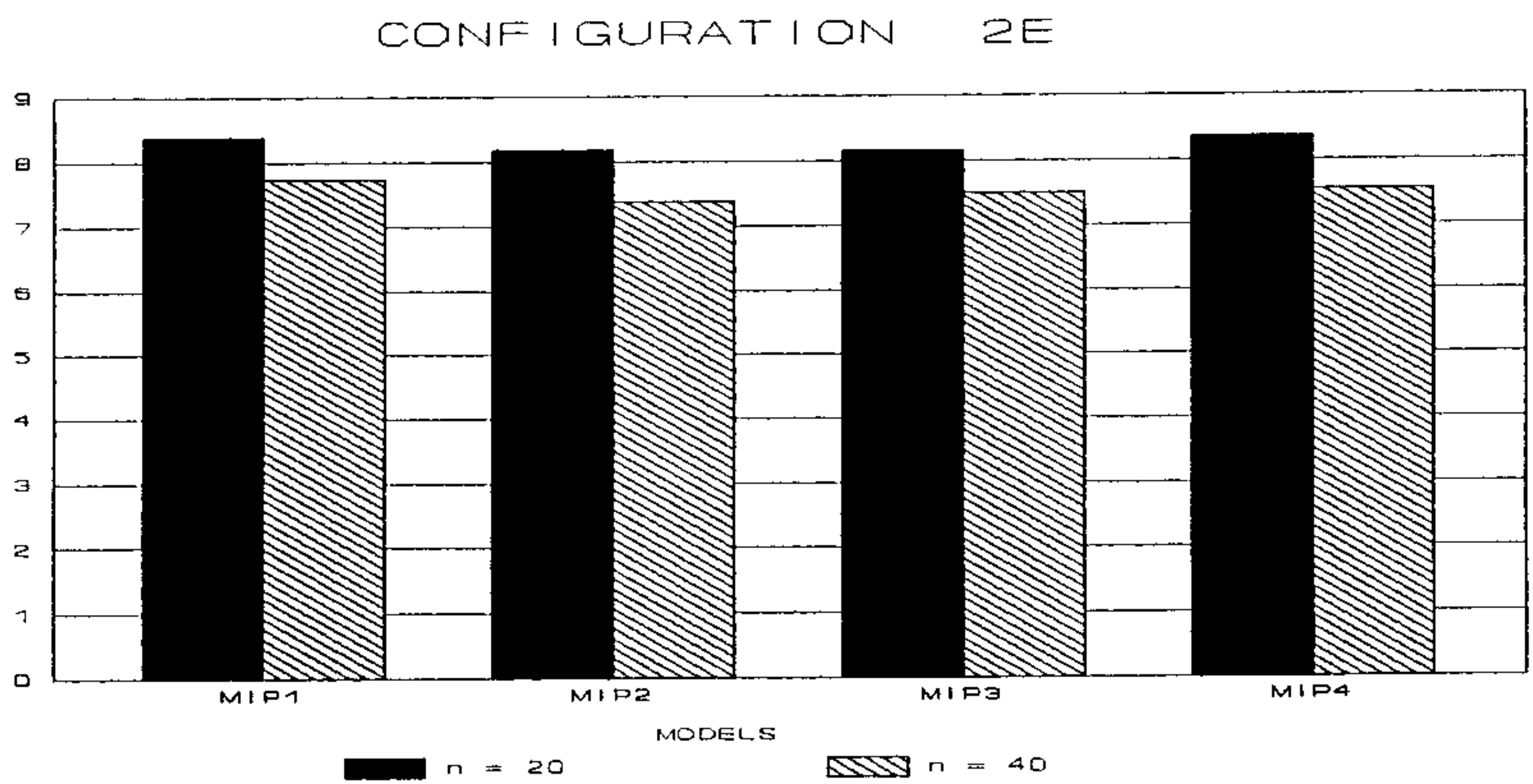

Fig. 13. Percentages of Exact Misclassification for Configuration 2E

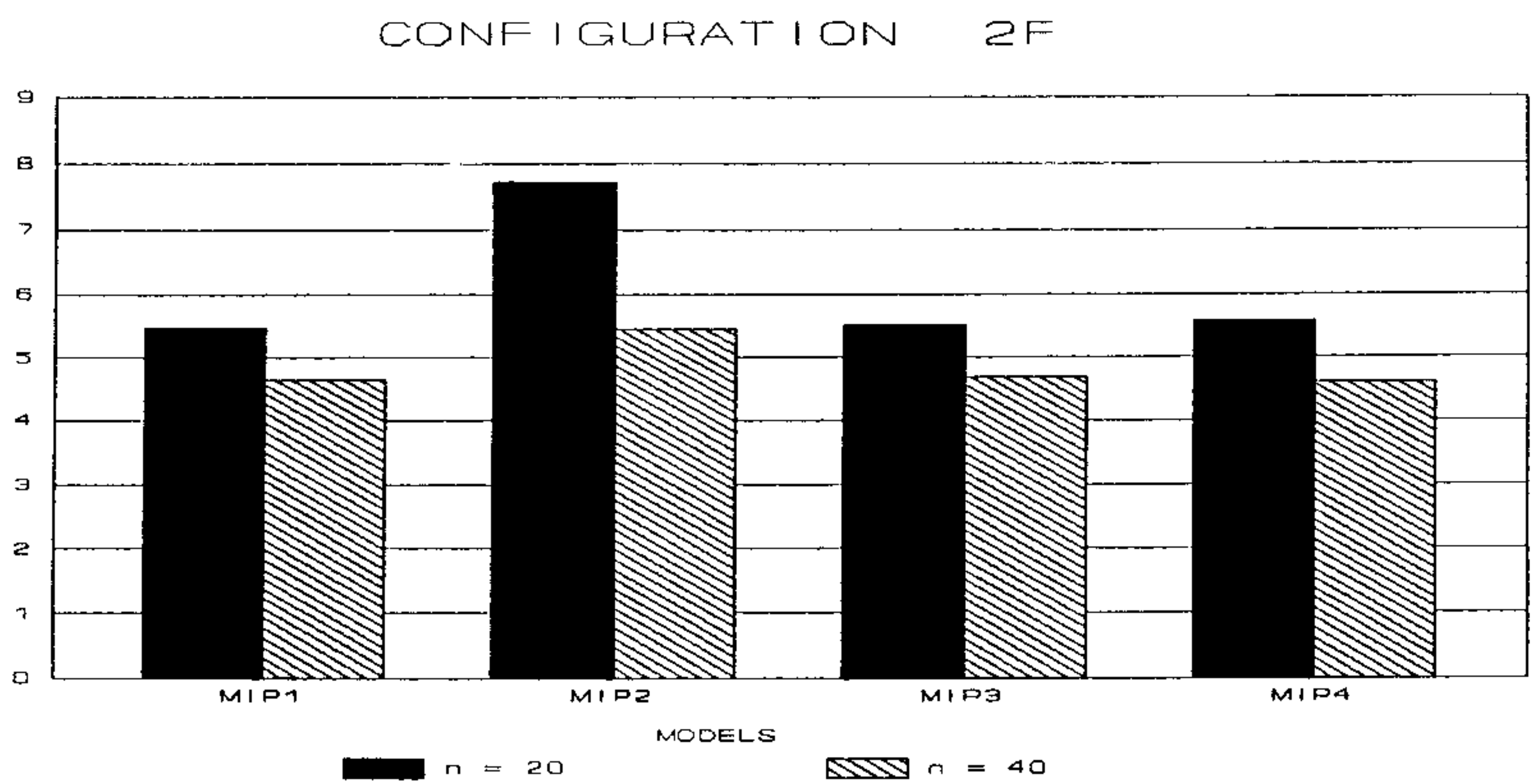

Fig. 14. Percentages of Exact Misclassification for Configuration 2F

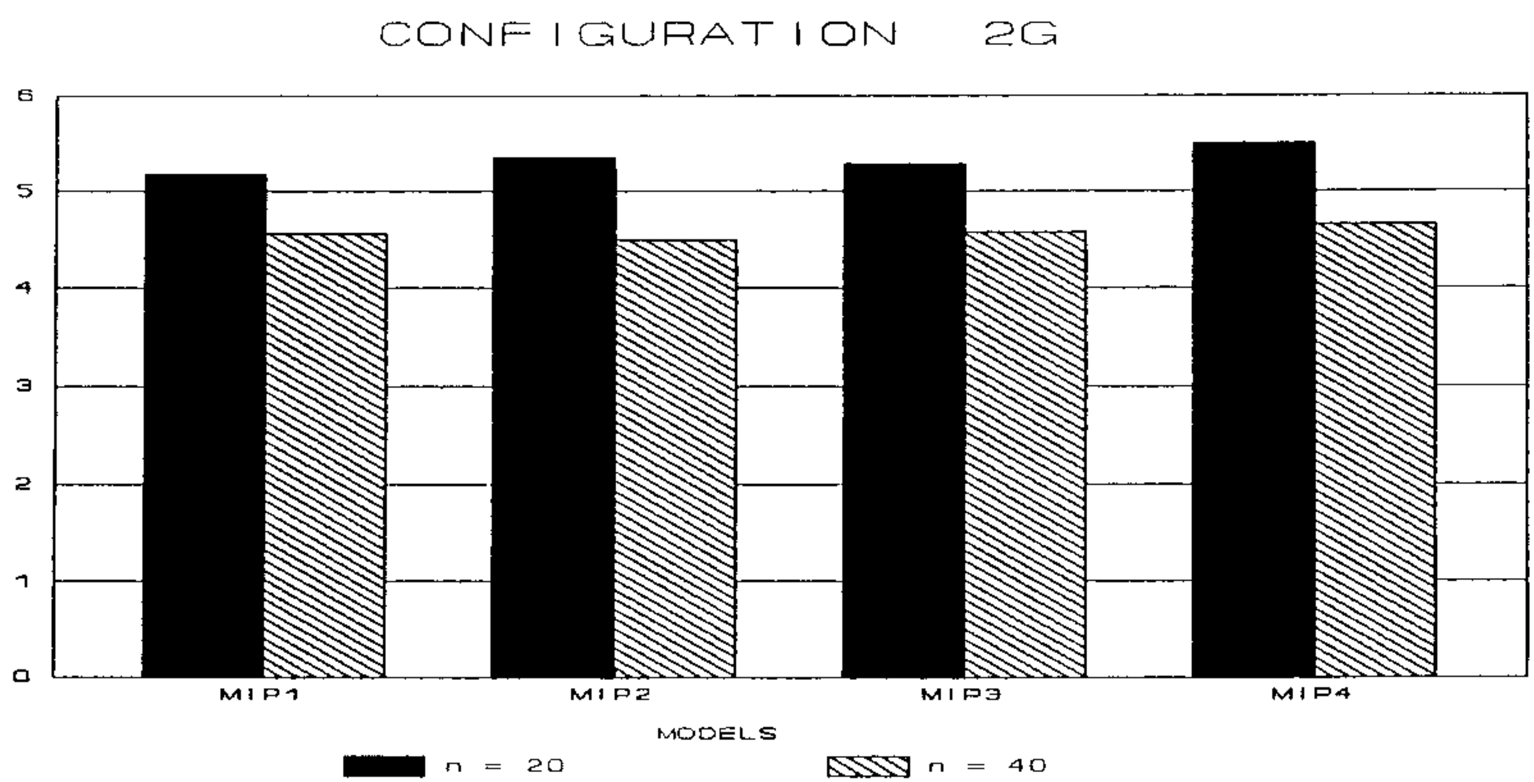

Fig. 15. Percentages of Exact Misclassification for Configuration 2G

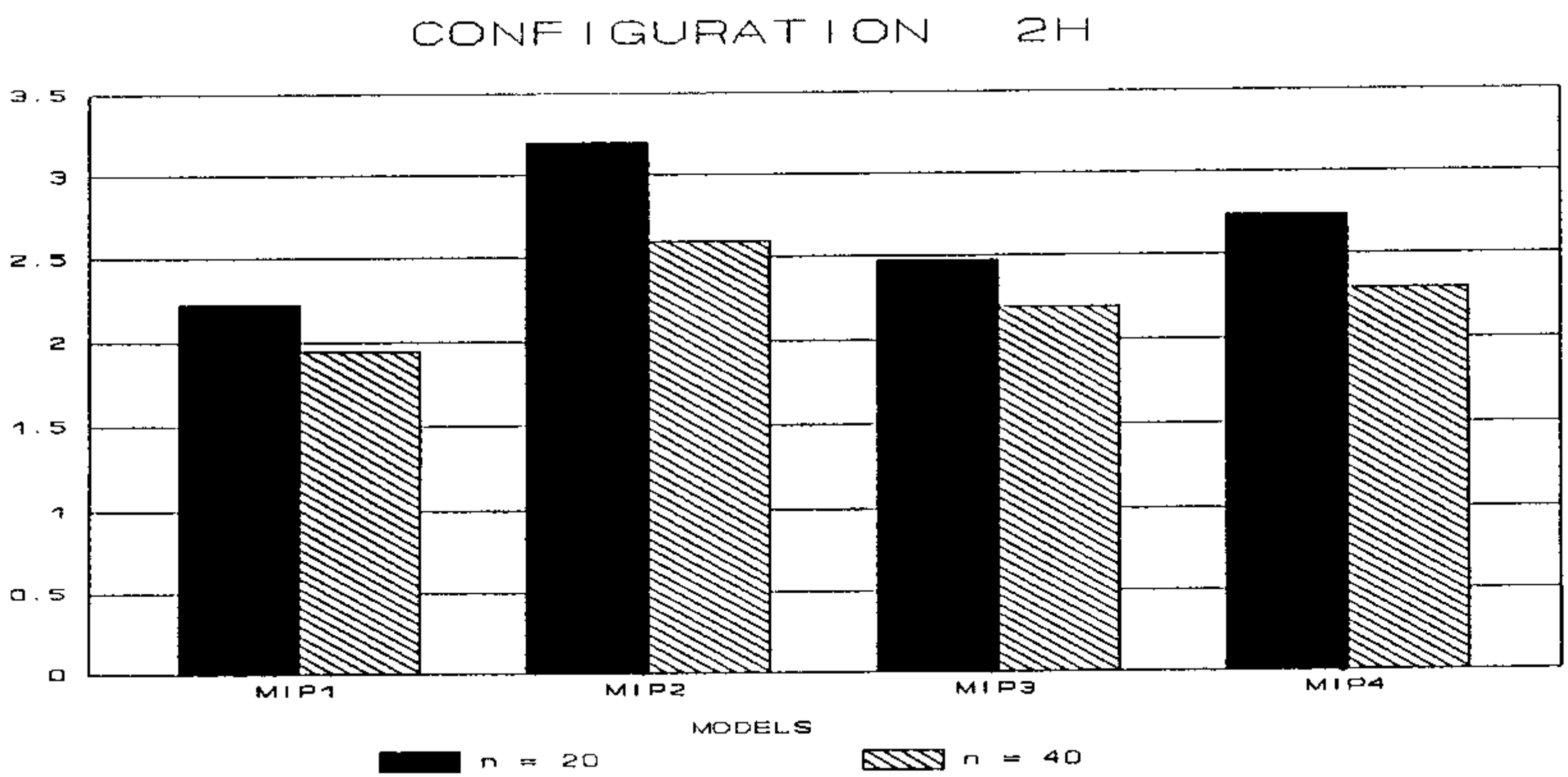

Fig. 16. Percentages of Exact Misclassification for Configuration 2H

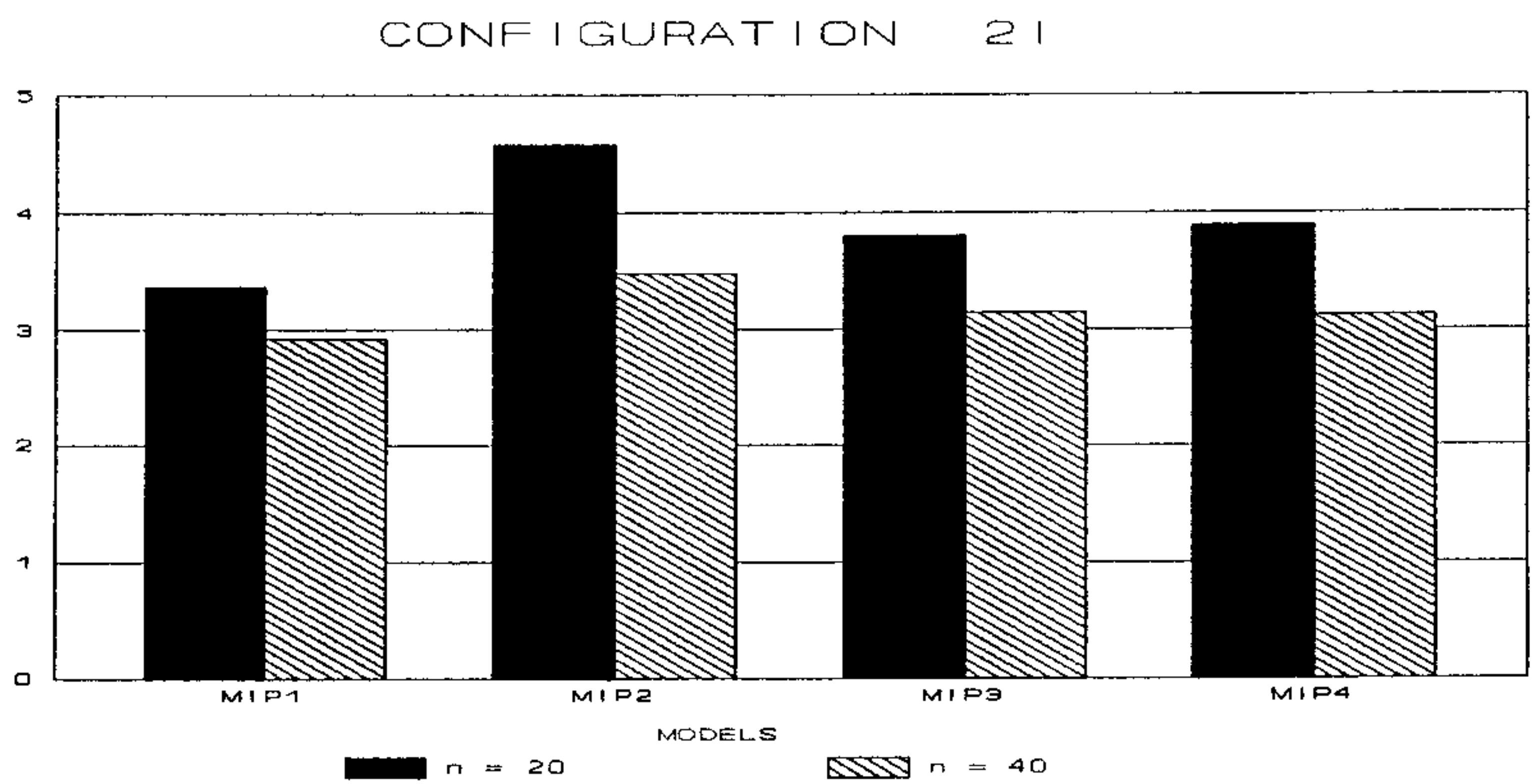

Fig. 17. Percentages of Exact Misclassification for Configuration 2I

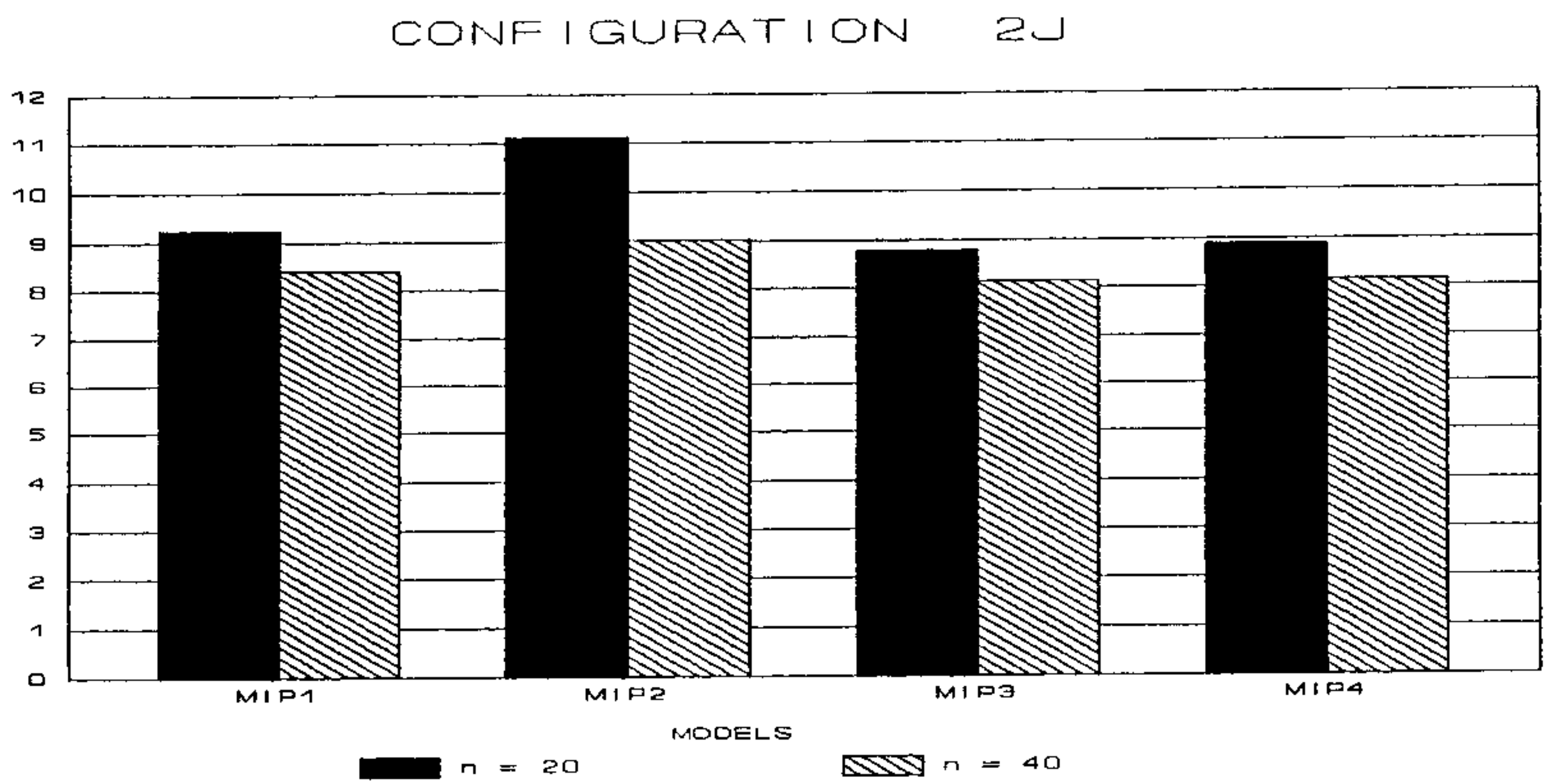

Fig. 18. Percentages of Exact Misclassification for Configuration 2J

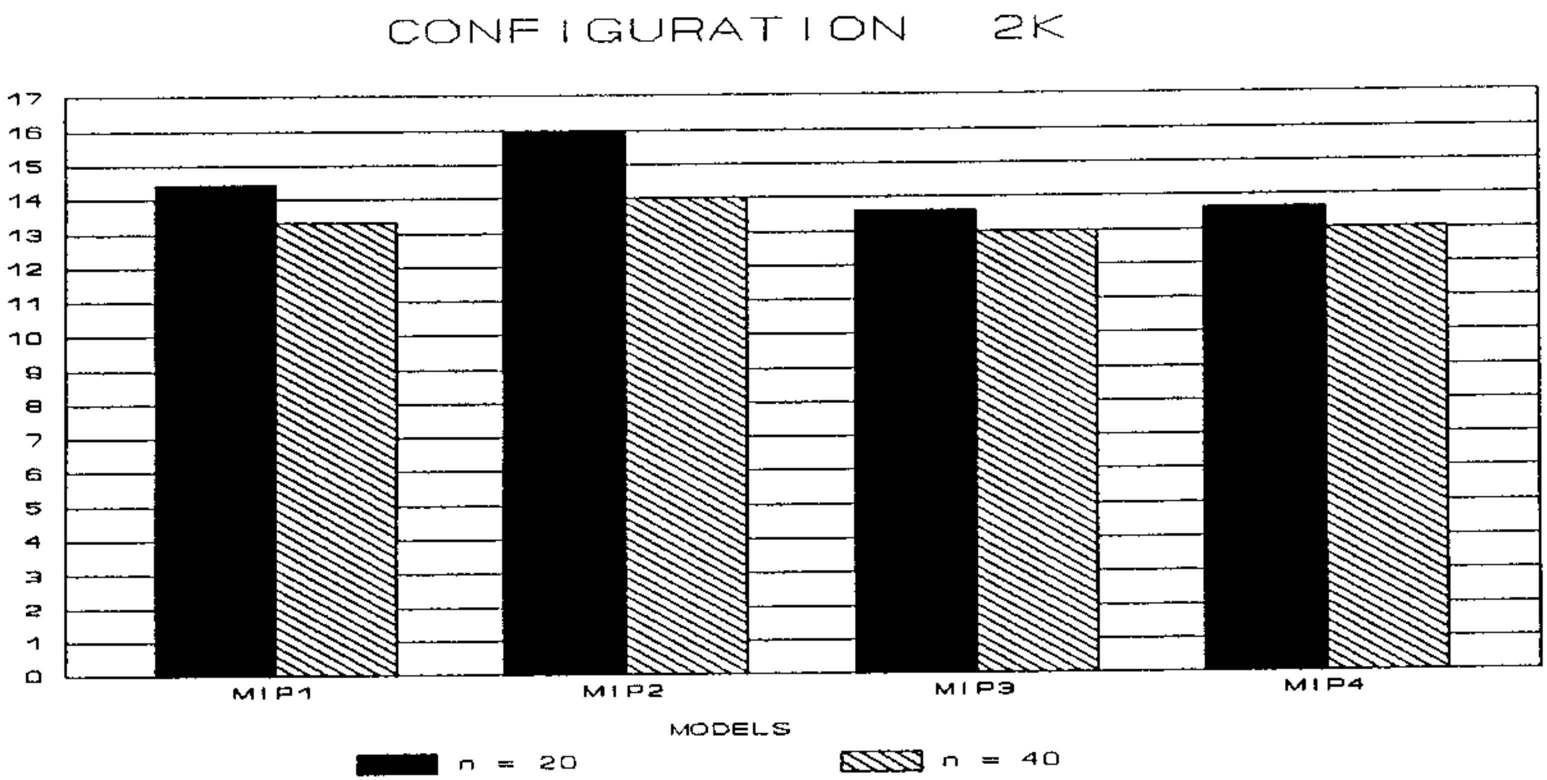

Fig. 19. Percentages of Exact Misclassification for Configuration 2K

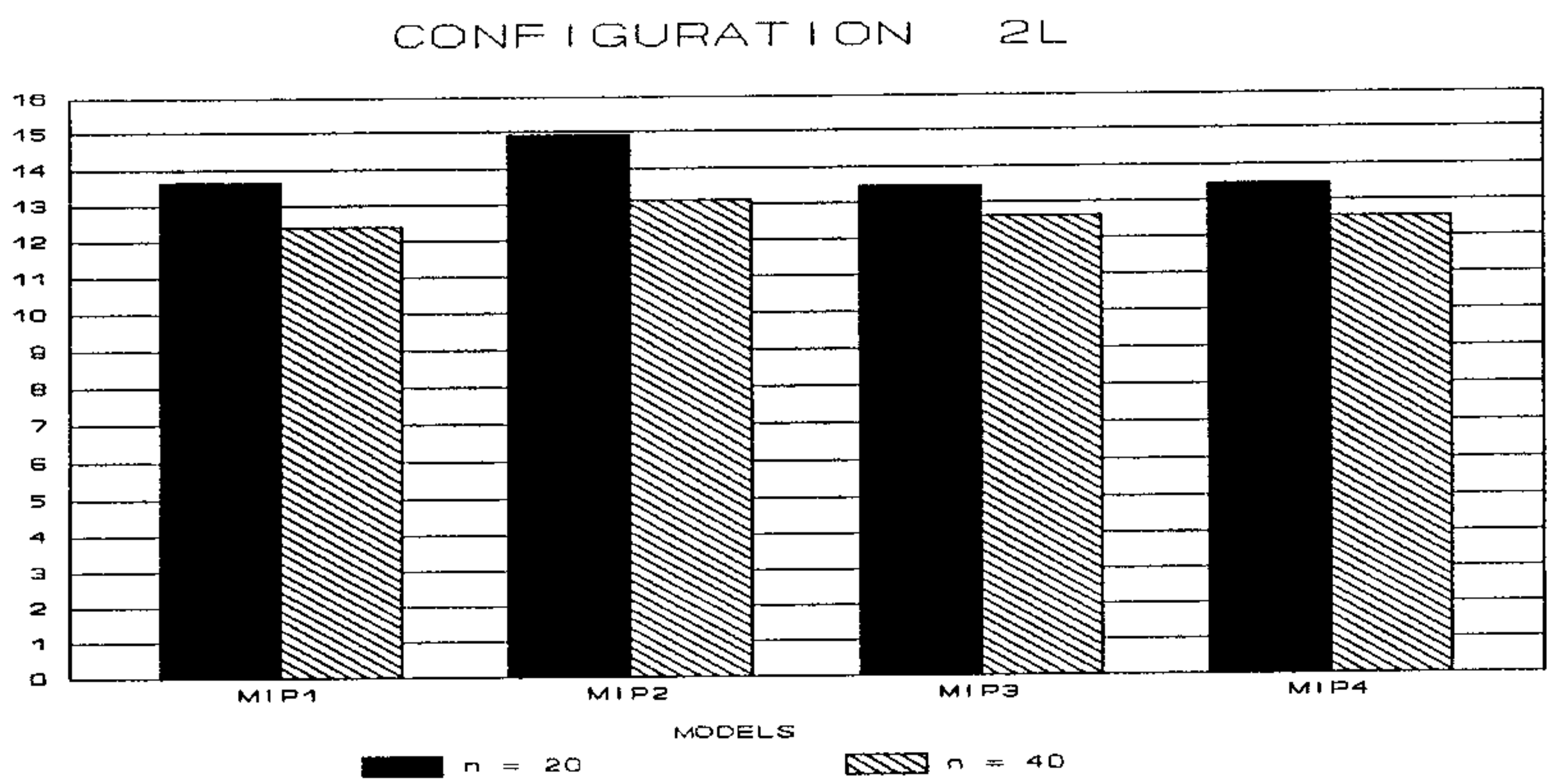

Fig. 20. Percentages of Exact Misclassification for Configuration 2L

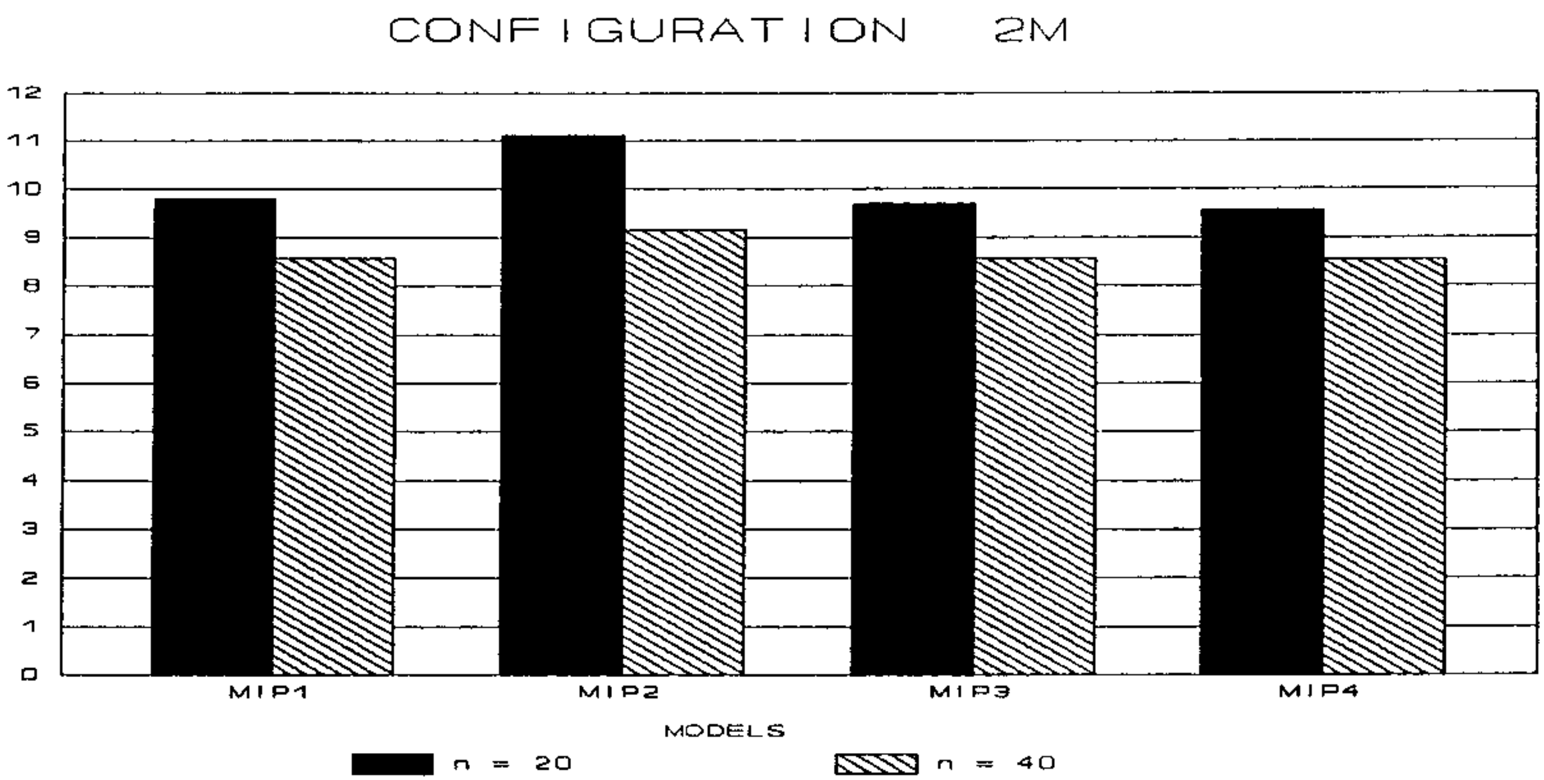

Fig. 21. Percentages of Exact Misclassification for Configuration 2M

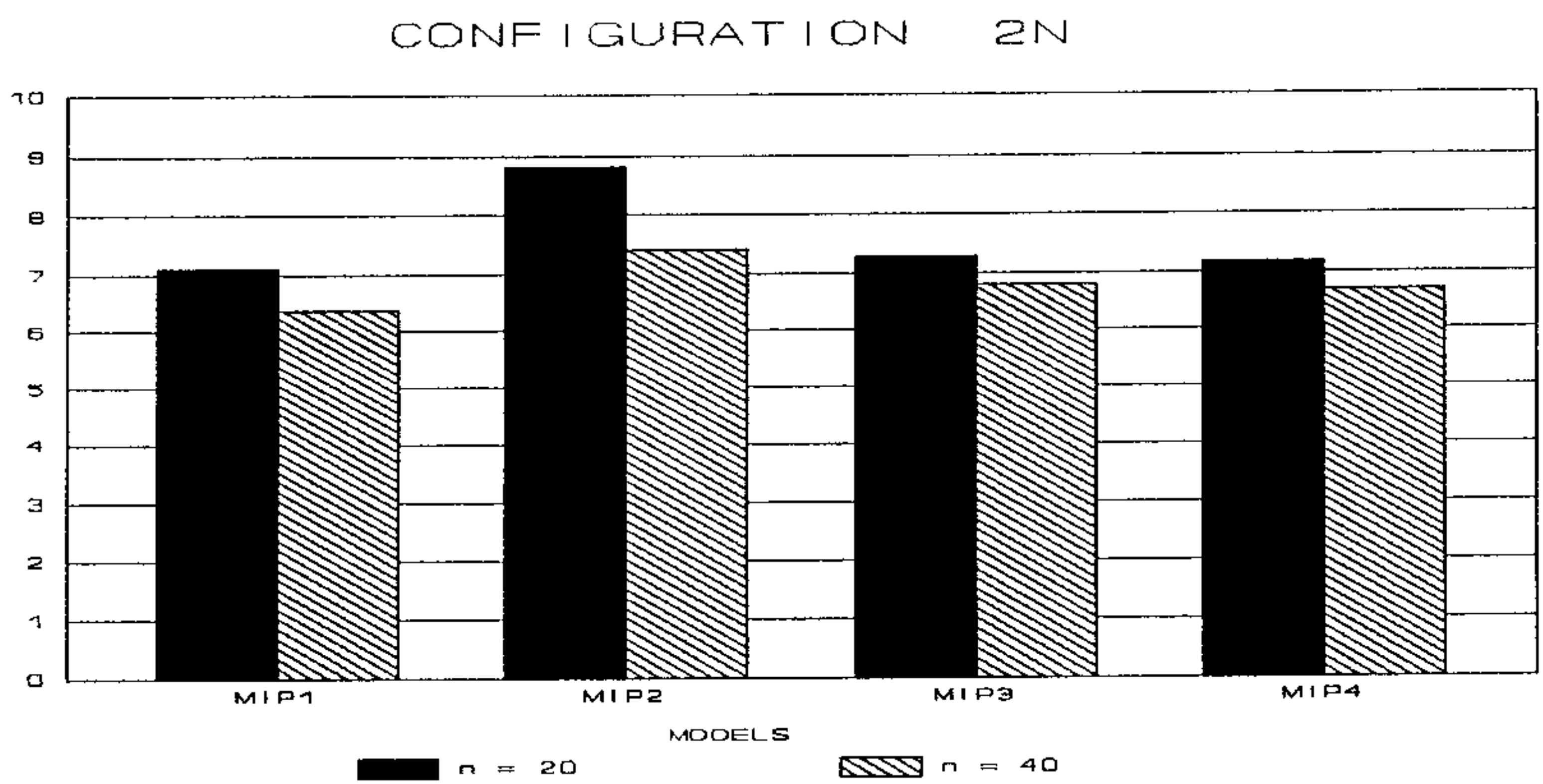

Fig. 22. Percentages of Exact Misclassification for Configuration 2N

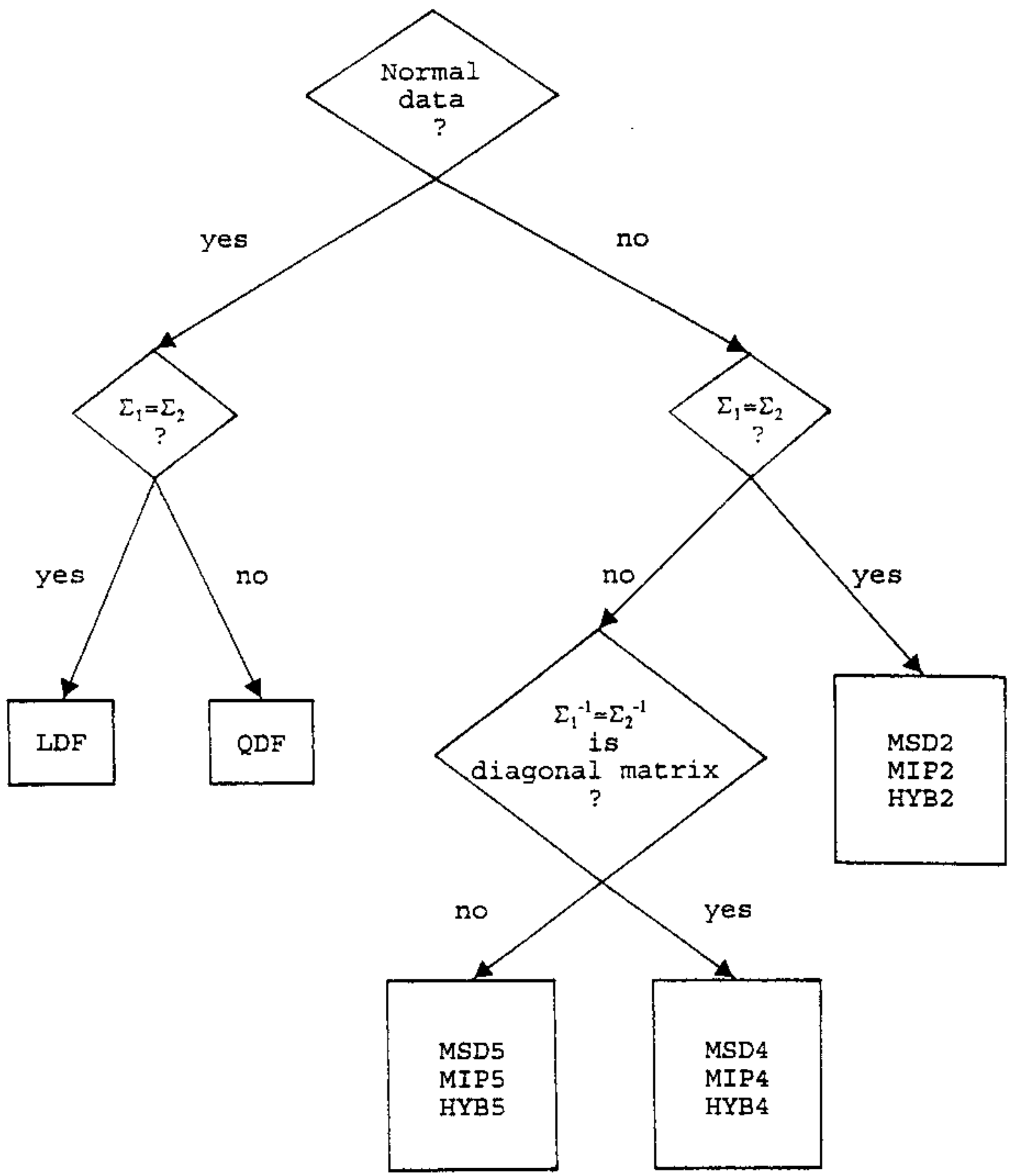

Fig. 23. Guideline for Alternative Mathematical Programming Models.

REFERENCE LIST

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

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

Bickel, P. and K. Doksum. 1977. *Mathematical statistics: basic idea and selected topics.* San Francisco: Holden-Day.

Devroye, L. 1986. *Non-uniform random variate generation.* New York: Spring-Verlag.

Draper, N. and H. Smith. 1981. *Applied regression analysis.* New York: John Wiley.

Fisher, R. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7:179-188.

Fleishman, A. 1978. A method for simulating nonnormal distributions. *Psychometrika* 43:521-532.

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

________. 1986. Resolving certain difficulties and improving the classification power of LP discriminant analysis formulations. *Decision Sciences* 17:589-595.

Glover, F. 1990. Improved linear programming models for discriminant analysis. *Decision Sciences* 21:771-785.

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

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

Hand, D. 1981. *Discrimination and classification.* New York: John Wiley.

Hogg, V. and A. Craig. 1978. *Introduction to mathematical statistics.* New York: Macmillan.

Huberty, C., J. Wisenbaker, and J. Smith. 1987. Assessing predictive accuracy in discriminant analysis. *Multivariate Behavioral Research* 22:307-329.

Huberty, C. 1984. Issues in the use and interpretation of discriminant analysis. *Psychological Bulletin* 95:156-171.

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

Johnson, R. and D. Wichern. 1992. *Applied multivariate statistical analysis.* New Jersey: Prentice Hall.

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

________. 1989b. Unacceptable solutions and the hybrid discriminant model. *Decision Sciences* 20:844-848.

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

Lee, C. and J. Ord. 1990. Discriminant analysis using least absolute deviations. *Decision Sciences* 21:86-96.

Loucopoulos, C. 1993. *Mathematical programming approaches to the three-group classification problem.* Doctoral Dissertation, University of North Texas.

Mahmood, M. and E. Lawrence. 1987. A performance analysis of parametric and nonparametric discriminant approaches to business decision making. *Decision Sciences* 18:308-326.

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

Markowski, C. and E. Markowski. 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.

Morrison, D. 1976. *Multivariate statistical methods.* New York: McGraw-Hill.

Nath, R. 1984. Estimation of misclassification probabilities in the linear programming approaches to the two group discriminant problem. *Decision Science* 21:373-386.

Nath, R., W. Jackson, and T. Jones. 1992. A comparison of the classical and the linear programming approaches to the classification problem in discriminant analysis. *Journal of Statistical Computation and Simulation* 41:73-93.

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

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

________. 1990a. Heuristic solution procedures for a mixed-integer programming discriminant model. *Managerial and Decision Economics* 11:255-266.

________. 1990b. A comparison of linear programming and parametric approaches to the two group discriminant problem. *Decision Sciences* 21:373-386.

________. 1991. Separation failure in linear programming discriminant models. *Decision Sciences* 22:519-535.

Silva, A. and A. Stam. 1994. Second order mathematical programming formulations for discriminant analysis. *European Journal of Operational Research* (Forthcoming).

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

Srinivason, V. and Y. Kim. 1987. Credit granting: A comparative analysis of classification procedures". *Journal of Finance* 42:665-683.

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

Sudarsanam, P. and R. Toffler. 1985. Industrial classification in UK capital markets: a test of economic homogeneity. *Applied Economics* 17:291-308.

Vale, C. and V. Maurelli. 1983. Simulating multivariate nonnormal distributions. *Psychometrika* 48:465-471.

Welker, R. 1974. Discriminant and classification analysis as an aid to employee selection. *Accounting Review* 49:514-523.