# DISCRIMINANT ANALYSIS WITH PROPORTIONAL COVARIANCE STRUCTURE

by

ELDON YU-ZEN LI, B. OF COM., M.S.

A DISSERTATION

IN

BUSINESS ADMINISTRATION

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements of
the Degree of

DOCTOR OF BUSINESS ADMINISTRATION

Approved

Accepted

May, 1982

## ACKNOWLEDGEMENTS

I am deeply indebted to the committee chairman, Dr. Stephen C. Hora, for his continuous support and constructive direction of this dissertation. To Dr. James R. Burns and Dr. Gary D. Kelley, I would like to express my sincere appreciation for their helpful criticism and ideas. My appreciation is extended to Dr. James B. Wilcox and Dr. David E. Upton for their suggestions and editorial assistance. Finally, I am very grateful to my wife, Rebecca. Her patience, understanding, sacrifice, and secretarial aid helped me complete the preparation of this dissertation.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

Figure page

# Chapter I

## INTRODUCTION AND ORGANIZATION

The purpose of this dissertation is to examine the proportional quadratic discriminant function (PQDF). The PQDF is an optimal discriminant function for the two-population case with known k-variate normal distributions when the population covariance matrices are unequal, but proportional. The motivation for this research lies in the belief that the PQDF often will perform better than the linear discriminant function (LDF) and the quadratic discriminant function (QDF) when the population parameters must be estimated. The superiority of the estimated PQDF over the estimated QDF is found to hold even in regions of the parameter space where the QDF is optimal and the PQDF is nonoptimal.

In this research, methods have been derived for estimating the coefficients of the PQDF. Comparisons to the LDF and QDF are made using analytic asymptotic methods and Monte Carlo simulation.

## 1.1 Introduction

### 1.1.1 Overview

Since R. A. Fisher (1936) first developed the linear

discriminant function (LDF), the problem of multivariate discrimination has been of intense interest. The questions encountered in the study of discrimination functions are basically of five different types. These may be characterized as follows: (1) What information is available about the distribution of the discriminating variables? (2) Which classification rule should be used? (3) How well does the classification rule perform? (4) How robust is the rule to departures from the assumptions that are made? (5) What variables should be selected for use in the classification rule?

Various multivariate discriminant rules have been proposed. Some are distribution-free rules, but most are distribution-based in nature. Rules based on multivariate normal distribution theory such as Wald's (1944) and Anderson's (1951, 1958) have received the most attention. Two-group linear discrimination represents the prototype of a variety of discrimination problems and thus has been extensively studied. The most often used version of the linear classification rule for discriminating between two multivariate normal populations is given by:<1>

---

<1> Subscripts are denoted by an underscore throughout this dissertation. Vectors and matrices are shown in bold-face print except the coefficient of proportionality, **d**. Other terms that will be used in this research are defined in Appendix A.

$$(1.1)\quad L(X) = X'V^{-1}(u_1 - u_2) - 1/2\,\{(u_1 + u_2)'V^{-1}(u_1 - u_2)\} + \log p_1 C(2|1)/p_2 C(1|2),$$

where $L(X)$ is the optimal (assuming known parameters) linear discriminant function, log is the natural logarithm, $p_1$ and $p_2$ are *a priori* probabilities of populations 1 and 2, $C(i|j)$ $(i \neq j,\ i,j=1,2)$ is the cost of assigning an individual from population j into population i, $u_1$ and $u_2$ are the k X 1 mean vectors, and V is the k X k common covariance matrix. It is optimal to assign a randomly chosen individual X to population 1 if $L(X) \geq 0$, and to population 2 otherwise. (Wald 1944, Anderson 1951, 1958)

Smith (1947), Rao (1973), and Huberty (1975), among others, suggested that if the covariance structure is not constant (i.e., $V_1 \neq V_2$), a nonlinear rule which is typically the quadratic discriminant function (QDF) should be the choice. However, the performance of nonlinear discriminant functions has been frequently questioned. Gilbert (1969), Dunn (1971), Marks and Dunn (1974), Boullion, Odell, and Duran (1975), Van Ness and Simpson (1976), Wahl and Kronmal (1977), Huberty and Curry (1978), and Van Ness (1979) studied the performance of either the linear rule, the quadratic rule, or both. They reached similar conclusions concerning estimated discriminant functions, though in different ways. They concluded that given that the k-component vector X is normally distributed

with a priori probabilities $p\underline{1}=p\underline{2}=0.5$, $C(1|2)=C(2|1)=1$, mean vectors $u\underline{1}=(0,0,\ldots,0)$, $u\underline{2}'=(\Delta\sqrt{(1+d)/2},0,\ldots,0)$, covariance matrices $V\underline{1}=I$, $V\underline{2}=dI$, where $\Delta^2$ is the Mahalanobis's squared distance between the two means, I is a k X k identity matrix, and d is an arbitrary positive constant, the expected probability of misclassification is, in general, an increasing function of k (for fixed $\Delta^2$, d, and training sample size n), a decreasing function of $\Delta^2$, d, and n respectively (for other parameters being fixed). However, the mixed effect of these parameters is still ambiguous. Marks and Dunn (1974) also showed that the linear discriminant function is, in some instances, superior to the quadratic one. They attributed this result to either small n, small d, or large k.

All the above researchers treated PQDF as a special case of QDF. By assuming $V\underline{2}=dV\underline{1}$, the QDF described in (1.2) will, of course, become the PQDF in (1.3):

$$(1.2)\quad Q(X) = -1/2\ X'(V\underline{1}^{-1} - V\underline{2}^{-1})X + X'(V\underline{1}^{-1}u\underline{1} - V\underline{2}^{-1}u\underline{2})$$
$$-1/2\ (u\underline{1}'V\underline{1}^{-1}u\underline{1} - u\underline{2}'V\underline{2}^{-1}u\underline{2})$$
$$+1/2\ \log|V\underline{2}|/|V\underline{1}| + \log p\underline{1}C(2|1)/p\underline{2}C(1|2)\ ,$$

where Q(X) is the optimal (assuming known parameters) quadratic discriminant function and it is optimal to assign a random individual X to population 1 if $Q(X)\geq 0$, and to population 2 otherwise (Cooper 1963, Lachenbruch 1975).

After replacing $V\underline{2}$ with $dV\underline{1}$, the QDF becomes the PQDF:

$$(1.3)\quad P(X) = -1/2\ (1 - 1/d)X'V\underline{1}^{-1}X + (X'V\underline{1}^{-1}(u\underline{1} - 1/d\ u\underline{2}))$$

$$-1/2\ \{u\underline{1} + (1/\sqrt{d})u\underline{2}\}'V\underline{1}^{-1}\{u\underline{1} - (1/\sqrt{d})u\underline{2}\}$$

$$+1/2\ k(\log d) + \log p\underline{1}C(2|1)/p\underline{2}C(1|2)\ ,$$

where $P(X)$ is the optimal (assuming known parameters) proportional quadratic function and it is optimal to assign a random individual $X$ to population 1 if $P(X) \geq 0$, and to population 2 otherwise.

Consider the LDF, QDF and PQDF given respectively by Equations (1.1), (1.2), (1.3). The advantages of the PQDF over the LDF and QDF are easily understood when $p\underline{1}$, $p\underline{2}$, $C(1|2)$, and $C(2|1)$ are known and $u\underline{1}$, $u\underline{2}$, $V\underline{1}$, and $V\underline{2}$ must be estimated from the training sample. While the LDF requires $(k^2+5k)/2$ estimates for its parameters $u\underline{1}$, $u\underline{2}$, and $V$ ($V\underline{1}=V\underline{2}=V$), the QDF requires $k^2+3k$ estimates for $u\underline{1}$, $u\underline{2}$, $V\underline{1}$, and $V\underline{2}$, and the PQDF requires $(k^2+5k)/2 + 1$ estimates for $u\underline{1}$, $u\underline{2}$, $V$, and $d$ ($V\underline{1} = (1/d)V\underline{2} = V$).

Thus, by estimating one more parameter, $d$, than those required by the LDF, the PQDF can be applied in the situation of proportional covariances. Moreover, while both the QDF and PQDF are quadratic in $X$, the coefficients of the PQDF can be estimated more efficiently than those of the QDF because these coefficients depend upon $(k^2+k)/2 - 1$ fewer parameters. The increased efficiency of these estimates reduces the probable errors of misclassification to such an extent that even when the assumptions of the PQDF are not

met and those of the QDF are met, the PQDF estimated from training data may outperform the QDF estimated from the same data.

Han (1969, 1974) studied the distribution of the PQDF under known and unknown population parameters. With unknown parameters, he (1974) considered the case of two multivariate normal populations, $N(u_1, V)$ and $N(u_2, dV)$, and assumed that the mean vectors are unknown and the covariance matrices are partially known, i.e., either d is known or V is known. However, the performance of the PQDF compared to the LDF and QDF has not yet been studied. Further detailed investigation on the behavior of the PQDF will be conducted in this research.

### 1.1.2 The Problem of Discrimination

A discrimination problem arises when one wants to assign an element or observation of unknown origin to one of several predetermined, mutually exclusive, and collectively exhaustive classes or groups. The observed element is usually assumed to belong to exactly one of the several groups. Although there are rules that permit the strategy which leaves room for suspended judgment, these will not be treated here.

Some statisticians differentiate between classification and discrimination as the former is the identification of the category or group to which an individual or object

belongs on the basis of its observed characteristics, and the latter is the assignment of the individual into groups. That is, in discrimination, the existence of the groups is given; in classification, it is a matter to be determined (Kendall and Stuart 1975). More commonly, however, the terms "discrimination" and "classification" are used interchangeably. This practice will be followed here.

In discriminant analysis, one is uncertain of the element's true classification and can ascertain this true classification only at a cost, or perhaps not at all. The costs involved are of two kinds: one is the information cost which is the expense of obtaining additional information about the true classification, and another is the cost of misclassification itself. The cost of misclassification is the opportunity loss associated with assigning an element from population j to population i by using a chosen classification rule. This cost, denoted as $C(i|j)$, can be measured in terms of either dollars or utils. However, the cost of main concern is not the information cost but the misclassification cost. Considering all the possible ways of misclassification, one has the expected total misclassification cost denoted as TC and given by the following equation.

$$(1.4) \quad TC = \sum_{j=1}^{q} p_j \cdot r(j),$$

where

$$(1.5)\quad r(j) = \sum_{\substack{i=1 \\ i\neq j}}^{g} C(i|j) \cdot P(i|j), \quad j=1,2,\ldots,g,$$

$$(1.6)\quad P(i|j) = \int_{R\underline{i}} f(\mathbf{X}|j)\, d\mathbf{X}, \text{ for all } i\neq j,$$

and

$g$ = the number of groups,

$r(j)$ = the expected cost of misclassifying an element from population j into other populations,

$P(i|j)$ = the probability of misclassifying an observed vector $\mathbf{X}$ from population j into population i,

$R\underline{i}$ = the region in which observations are assigned to population i ($\Pi\underline{i}$),

$p_j$ = the <u>a priori</u> probability of a randomly chosen individual belonging to population j,

$f(\mathbf{X}|j)$ = the conditional probability of observing a vector $\mathbf{X}$ given that it is from population j.

The objective of discriminant analysis is to develop a strategy which minimizes TC, the expected total opportunity loss from misclassification (Anderson 1951, 1958), or, alternatively, minimizes the maximum expected individual misclassification costs given the others (Anderson 1977). When the costs of misclassification are all equal to one,

both procedures are equivalent to minimizing the total probability of misclassification (Welch 1939). The total probability of misclassification, also known as the error rate (ERR) (John 1961, Hills 1966), is denoted as T(R,f) and given below.

$$(1.7) \quad T(R,f) = \sum_{j=1}^{q} pi \cdot P(j),$$

where R refers to the classification regions, f refers to the presumed distribution of the observations that will be classified (Lachenbruch 1975), and

$$(1.8) \quad P(j) = \sum_{\substack{i=1 \\ i \neq j}}^{q} P(i|j), \quad j=1,2,\ldots,g,$$

where P(i|j) is defined as Equation (1.6).

Furthermore, the minimization of the maximum expected individual misclassification costs under equal C(i|j)'s is equivalent to minimizing the maximum probabilities of misclassification. The discriminant function derived by this procedure is also known as the "minimax rule" (Kendall 1957).

The first step in discriminant analysis is to identify a set of characteristics which will be used as variables in the discriminant function. These characteristics can be numerical measurements (continuous variables), or attributes (discrete variables), or both. In the case of unknown

population parameters, one needs to obtain samples with known origins (often called "training samples"). With these labeled objects and their observed characteristics, discriminant functions, either linear or non-linear composites of the variables, can be derived. Usually, the population parameters in the discriminant function are replaced by their corresponding estimates (Wald 1944, Day 1969).

Subsequently, some insignificant variables often can be eliminated using a variable selection procedure. Alternative variable selection procedures are based on either error rates (Habbema, et al. 1974, Hermans and Habbema 1976) or the t-statistic (Dixon and Massey 1957), the F-criterion (Dixon 1970, 1975, Nie, et al. 1975), the U-statistic derived from Wilks's Lambda (McCabe 1975, Rencher and Larson 1980), interdependence among the variables (Beale, Kendall and Mann 1967, Jolliffe 1972, 1973). The choice of a variable selection procedure depends upon what type of discrimination scheme has been used and the nature of the populations under study. After retaining those significant variables, the existing discriminant functions can be modified and applied to all the future observations.

The performance of a discriminant function is usually evaluated based upon the F-test (Rao 1946, 1950b, 1952, 1970), the chi-square test (Bartlett 1947a, 1947b), and the error rates (Toussaint 1974, Dillon 1979).

### 1.1.3 Geometric Interpretation of Two-group Discrimination

The geometric interpretation of discriminant analysis can be seen in the case of two groups (or populations) and two variates with the assistance of Figure 1. In this figure, the two sets of concentric ellipses represent the bivariate swarms for the two groups in an idealized form.

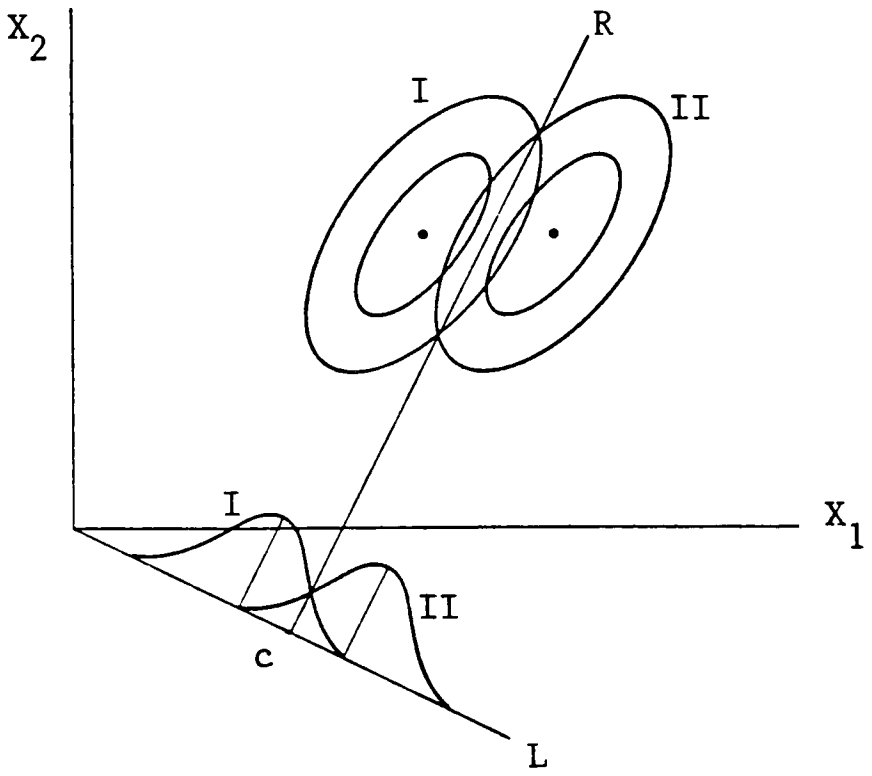



Figure 1: Two-dimensional Two-group Linear Discrimination

The two variates, X1 and X2, are moderately positively correlated. Each ellipse is the locus of points of equal density or frequency for a group. For instance, the outer ellipse for Group I might define the region within which 95 percent of Group I lies, and the inner ellipse, concentric with it, might define the region within which 80 percent of

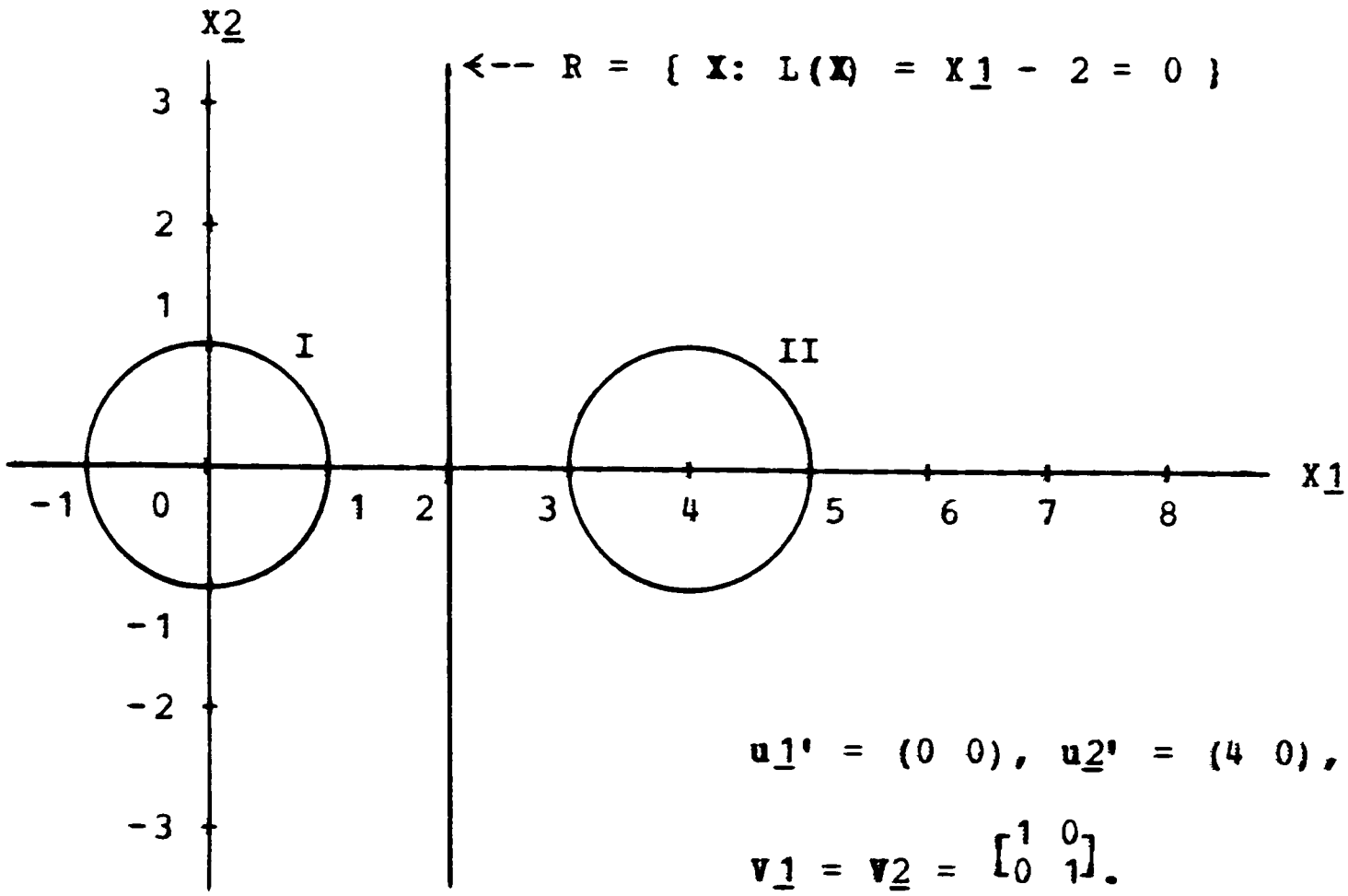
Group I lies. These ellipses are called "centours," for centile contours. The two points at which corresponding centours intersect determine a straight line, R. A second line, L, that goes through the origin is constructed perpendicular to line R. If the points in the two-dimensional space are projected onto L, the corresponding histogram of the canonical variable for the two populations are readily obtained. The overlap between the two groups will be smaller than for any other possible projection onto a line nonparallel to line L.

The discriminant function therefore transforms the individual observed values to a single discriminant score, and that score is the individual's location along line L. The point c where R intersects L divides the one-dimensional discriminant space into two regions, one indicating probable membership in Group I and the other region for Group II. The overlap of the two population distributions represents the total probability of misclassification. Notice that this diagram depends upon the equality of the two group dispersions. If either the variances or the covariances of $X_1$ and $X_2$, or both, were different for the two groups, the centours for the two groups would not have the same shape and/or orientation, and the boundary (line R) would not be a straight line. However, the sizes of the two populations do not have to be the same, only the dispersions of them (Cooley and Lohnes 1971).

Another geometric representation of a two-group optimal (known parameters) discriminant function $Dt(X)$ with two variables is shown in Figure 2. The boundary R is set up to obtain a maximum separation of the two populations, or a minimum expected total cost of misclassification. The aim of discrimination is to determine this boundary and classify each point in the space into one of two regions that correspond to the two populations. For the extension to the problem of multiple-group discrimination, one can treat the multiclass assignment as a sequence of pairwise discriminations when $C(i|i)=0$ and $C(i|j)=1$, for all $i \neq j$.

Figure 2(a) shows the optimal linear discriminant boundary with 2-variate normal distributions and $p1=p2=0.5$, $C(1|2)=C(2|1)=1$, $u1'=(0,0)$, $u2'=(4,0)$, and $V1=V2=I$, where I is a 2 X 2 identity matrix. Figure 2(b) shows the optimal quadratic discriminant boundary with the same conditions in Figure 2(a) except $V2= 2V1= 2I$. By substituting the above parameter values into Equations (1.1) and (1.2), respectively, the boundaries $R = \{ X: L(X) = 0 \}$ for Figure 2(a) and $R = \{ X: Q(X) = 0 \}$ for Figure 2(b) are obtained.

(a) Linear Discriminant Boundary




(b) Quadratic Discriminant Boundary

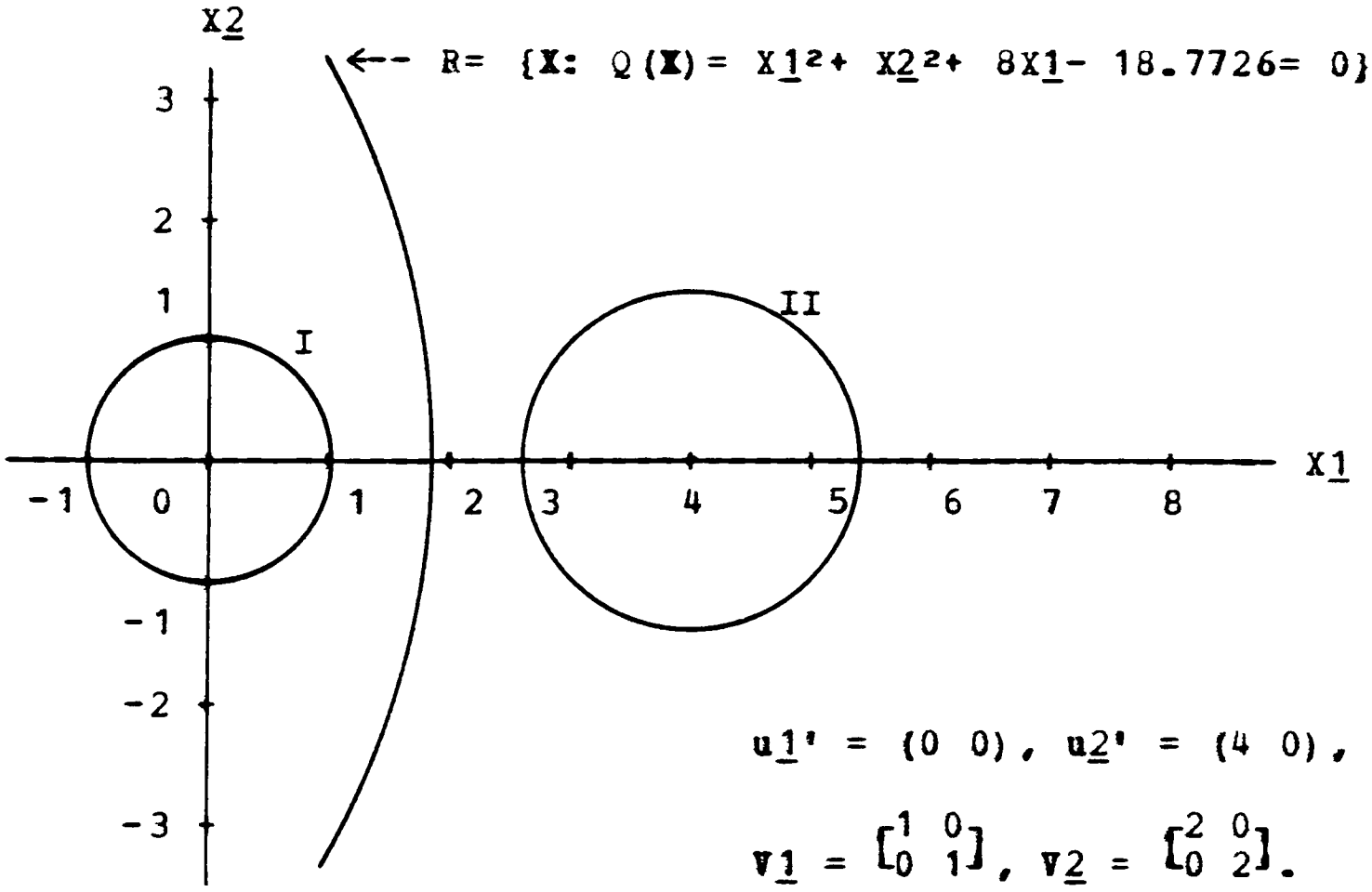



Figure 2: Linear and Quadratic Discriminant Boundaries

### 1.1.4 Historic Development of Discriminant Analysis

The idea of discriminating between multivariate populations can be traced back to the 1920's, and its development reflects the same broad phases as that of general statistical inference, namely, a Pearsonian phase followed by Fisherian, Neyman-Pearsonian, and Waldian phases (Hodges, 1950).

In the early work, the discrimination problem was not precisely formulated and it was often confounded with the problem of testing the equality of two or more distributions. Various test statistics which measured, in some sense, the divergence between two populations were proposed. Karl Pearson first proposed one such measure, termed as the "coefficient of racial likeness (CRL)" and denoted it by $C^2$, about 1920. However, the first work published on this measure was Tildesley's (1921), on Burmese skulls. Later, Pearson (1926) published a considerable amount of theoretical work on this CRL and suggested the following form for the coefficient when the variables are dependent and the two population dispersions are equal:

$$(1.9)\quad CRL = \frac{1}{k}\left\{\frac{n\underline{1}\ n\underline{2}}{n\underline{1}+n\underline{2}}\ (\bar{X}\underline{1}-\bar{X}\underline{2})'S^{-1}(\bar{X}\underline{1}-\bar{X}\underline{2})\right\} - 1,$$

where $\bar{X}\underline{i}$ is the k X 1 sample mean vector based on a sample of size $n\underline{i}$ from the population i, (i=1,2), and S is the pooled k X k sample covariance matrix, i.e., $S=(n\underline{1}+n\underline{2}-2)^{-1}\{(n\underline{1}-1)S\underline{1}+(n\underline{2}-1)S\underline{2}\}$, and

$$(1.10)\quad \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij},\ i=1,2,$$

$$(1.11)\quad S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)\ (X_{ij} - \bar{X}_i)',\ i=1,2,$$

where $X_{ij}$ is the $j$th k X 1 sample vector from the population i ($\Pi_i$).

The coefficient of racial likeness for the case of independent components was later modified by Morant (1928) and Mahalanobis (1927, 1930). Mahalanobis called his measure $D^2$ and further suggested (1930) some measures of divergence in variability, skewness, and kurtosis and studied their distributions. Subsequently, Mahalanobis (1936) gave the dependent variate versions of his $D^2$-statistic with the same dispersion within two populations in the classical and the studentized forms. Mahalanobis's $\Delta^2$ and $D^2$-statistics are given respectively in equations (1.12) and (1.13):

$$(1.12)\quad \Delta^2 = \frac{1}{k} \{(u_1 - u_2)' V^{-1} (u_1 - u_2)\},$$

$$(1.13)\quad D^2 = \frac{1}{k} \{(\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2)\},$$

where $\Delta^2$ and $D^2$ are, respectively, the generalized Mahalanobis's distance with known and unknown parameters, and

$$(1.14)\quad V = \frac{1}{n_1 + n_2}\,(\, n_1 V_1 + n_2 V_2 \,),$$

$$(1.15)\quad S = \frac{1}{n_1 + n_2 - 1}\,(\, n_1 S_1 + n_2 S_2 \,),$$

where $V_i$ is the $i$th population covariance matrix and

$$(1.16)\quad S_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)\,(X_{ij} - \bar{X}_i)', \quad i=1,2.$$

These generalized distances have been successfully applied to discrimination problems in many fields of study. Their distributions were studied by Bose (1936a, 1936b), Bose and Roy (1938), Bhattacharya and Narayan (1941), and recently by Davies (1972).

Bhattacharya and Narayan (1941) modified Mahalanobis's generalized distance for populations with unequal dispersions and gave the following different forms for S and $S_i$:

$$(1.17)\quad S = \frac{1}{n_1 + n_2 - 2}\,\{\, (n_1 - 1)S_1 + (n_2 - 1)S_2 \,\},$$

$$(1.18)\quad S_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)\,(X_{ij} - \bar{X}_i)', \quad i=1,2.$$

After Mahalanobis proposed his $D^2$-statistic, Hotelling (1931) generalized Student's t-statistic and suggested a test statistic $T^2$ which is a constant multiple of the studentized Mahalanobis's $D^2$:

$$(1.19)\quad T^2 = \frac{n_1 n_2}{n_1 + n_2} \cdot D^2.$$

The distributions of $T^2$ and $D^2$ in the null case are equivalent to the null distribution of $R^2$, i.e., the quantity

$$(1.20)\quad F = \frac{n_1 + n_2 - k - 1}{k(n_1 + n_2 - 2)} \cdot T^2$$

has the variance ratio F distribution with k and $n_1 + n_2 - k - 1$ degrees of freedom. This $T^2$ statistic was further discussed by Hotelling (1951) and Holloway and Dunn (1967).

The first clear statement of the problem of discrimination, and the first proposed solution to that problem were given by R. A. Fisher. In 1936, Fisher (1936) worked on the problem of classifying iris plants into one of the three species, iris setosa, iris versicolor, and iris virginica, by four measurements: sepal length, sepal width, petal length, and petal width. He reduced this 4-variate problem to a univariate one by considering an "optimum" linear combination of the four components. For a given linear composite Y of the four components, Fisher considered the ratio of the squared difference between the means of Y in the three species to the variance of Y within species and maximized this ratio in order to define the optimum linear combination. The coefficients of this optimum linear

composite were found to be proportional to a k x 1 vector $\beta$, such that

$$(1.21)\quad \beta = S^{-1}(\bar{X}_1 - \bar{X}_2).$$

He then used his univariate discrimination method with this linear composite as the random variable, i.e.:

$$(1.22)\quad Y = X'S^{-1}(\bar{X}_1 - \bar{X}_2),$$

and assigned the random variable Y to the population for which $|Y - \bar{Y}_i|$ is a minimum, i=1,2,3.

The next stage of development was influenced by Neyman and Pearson's fundamental works (1933a, 1933b, 1936) in the theory of statistical inference. Advancement proceeded with the development of decision theory. Welch (1939) derived the forms of Bayes rule and the minimax Bayes rule for discriminating between two known multivariate normal populations with the same covariance matrix. This case was also considered by Wald (1944) who followed a lemma of Neyman and Pearson (1936) and obtained the optimal linear classification rule given by Equation (1.1) above. He further proposed some heuristic rules for replacing the unknown parameters by their corresponding minimum variance unbiased estimates. Von Mises (1945) obtained the rule which maximizes the minimum probability of correct classification. Theoretical results on the discrimination problem in the framework of decision theory are given in the

book by Wald (1950) and in the paper by Wald and Wolfowitz (1951).

In a series of papers, Rao (1946, 1947a, 1947b, 1948, 1949a, 1949b, 1950a) examined the problem of discrimination into two or more populations following Neyman-Pearson's and Wald's approaches. He suggested a measure of distance between two populations and considered the possibility of withholding the decisions in the "doubtful" regions pending the availability of additional information. He later (1951a, 1951b, 1952, 1953) applied his approach to some classification problems.

The distribution of the LDF based on samples was studied by Wald (1944), Smith (1947), Sitgreaves (1952, 1961), Anderson (1958), Bowker and Sitgreaves (1961), Teichroew and Sitgreaves (1961), John (1959, 1960a, 1960b, 1962, 1964), Okamoto (1963, 1968), Kabe (1963), Memon and Okamoto (1970, 1971), and Anderson (1973). The distribution of the QDF was derived by Okamoto (1961) and Gilbert (1969). For the PQDF, the distribution was studied by Han (1969, 1974).

The problem of variable selection was examined by Cochran (1964), Dixon and Massey (1957), Weiner and Dunn (1966), Beale, Kendall and Mann (1967), Watanabe, et al. (1967), Urbakh (1971), Jolliffe (1972, 1973), Habbema, et al. (1974), Moran (1974), McCabe (1975), Hermans and Habbema (1976), Rencher and Larson (1980).

The problem of discrimination into two univariate normal populations with different variances was investigated by Cavalli (1945) and Penrose (1947). The multivariate analog of this problem was first resolved by Smith (1947). Smith (1947) suggested the use of the quadratic discriminant function obtained from the log likelihood ratio and compared LDF with QDF. Later, Lubin (1950) extended Smith's (1947) method to cases involving more than two populations. Bartlett and Please (1963) followed the work of Penrose (1947) and Smith (1947). They considered the case with $\underline{u}1=\underline{u}2=0$ and $V\underline{1}= (1-\rho\underline{1})I + \rho\underline{1}J$, $V\underline{2}= d(1-\rho\underline{2})I + \rho\underline{2}J$, where 0 is a k x k zero vector, J is a k X k matrix with all components equal to unity, d is some positive constant, and $\rho\underline{i}$ is the equal correlation among variables in population i. Cooper (1963) employed Anderson's (1958) quadratic form to extend the discriminant problem into several large classes of distributions. Bunke (1964) and Han (1968) followed Bartlett and Please's (1963) study and treated the case with unequal means. Han (1969, 1970, 1974) continued using the likelihood ratio procedure to study the distribution of discriminant functions when the covariance matrices are of either the proportional or circular type.

Kullback (1952, 1959), Clunies-Ross and Riffenburgh (1960), Riffenburgh and Clunies-Ross (1960), and Jennrich (1962) derived rules based on a linear function of X. Anderson and Bahadur (1962) obtained a "best linear" rule by

using a minimax procedure. Banerjee and Marcus (1965), following Anderson and Bahadur's (1962) work, gave the bounds of this minimax rule. Chaddha and Marcus (1968) analyzed, mainly by simulation, the behavior of some estimates of a divergence measure defined as $2(\underline{u}1-\underline{u}2)'\ (\underline{V}1+\underline{V}2)^{-1}\ (\underline{u}1-\underline{u}2)$. Recently, Eisenbeis and Avery (1972) used Fisher's iris data and showed that although the hypothesis of the equality of the dispersion matrices was rejected beyond any measurable level of significance, the use of the quadratic classification rule yielded identical results to those using the linear rule. They attribute this result to the large distance between group means.

Rao (1946, 1950b, 1952, 1970) provided an F-test and Bartlett (1947a, 1947b) gave a chi-square test for the significance of the discriminatory power. The estimation of classification error rates was studied by Smith (1947), Linhart (1959), Cochran and Hopkins (1961), John (1961), Highleyman (1962), Brailovskiy (1964), Brailovskiy and Lunts (1964), Frank, Massy and Morrison (1965), Lachenbruch (1965, 1967, 1968), Dunn and Varady (1966), Hills (1966), Lunts and Brailovskiy (1967), Lachenbruch and Mickey (1968), Toussaint (1969), Toussaint and Donaldson (1970), Dunn (1971), Fukunaga and Kessel (1971, 1972), Lainiotis and Park (1971), Sedransk and Okamoto (1971), Das Gupta (1972), Sorum (1971, 1972a, 1972b), Lissack and Fu (1972), Duda and Hart (1973), Lachenbruch, Sneeringer and Revo (1973), Toussaint and

Sharpe (1974), McLachlan (1974a, 1974b, 1975, 1976), Efron (1975), O'Neill (1975, 1980), Glick (1978), Hora and Hanna (1978), Streit (1979), Hora and Wilcox (1981). The effect of initial misclassification was examined by Lachenbruch (1966, 1974, 1979), Press (1968), and McLanchlan (1972).

### 1.1.5 Applications of Discrimination Techniques

It was R. A. Fisher (1936) who first applied the linear discriminant function to a botanical problem in 1936. Since then, various discriminant functions have been derived and employed in widely scattered fields such as taxonomy, botany, anthropology, archaeology, meteorology, geology, biometrics, biology, signal detection, remote sensing analysis, pattern recognition, medical diagnostics, psychology, education, sociology, marketing, personnel administration, accounting, banking, finance, etc. These applications have ranged from separating two species of flowers on the basis of petal length (Fisher 1936), or two series of Egyptian skulls on the basis of bone dimension (Barnard 1935), to distinguishing between good and poor salesmen (Wallace and Travers 1938), or good and poor loan risks in the consumer installment finance business (Durand 1941).

The following examples may delineate the wide variety of the applications of discrimination techniques.

(1) Consider three populations: the Brahmin, Artisan, and Korwa Castes of India. It is assumed that each of these three populations can be distinguished by four characteristics --- stature, sitting height, nasal depth, and nasal height --- of each member of the population. Based on sample observations of these characteristics, the problem is to classify an individual with observation vector X into one of the three populations. Rather than look at the characteristic observations individually, it may be easier to combine the sample data in some way to get a single score that can be used to solve this problem. Multiple-group linear discrimination can be applied to obtain the solution. (Rao 1948)

(2) A number of predictors are available at five weather stations in an area. These could include visibility, height of ceiling, east-west wind component, north-south wind component, total cloud coverage, changes in pressure in the last three hours, etc. On the basis of these measures, one wishes to predict what the ceiling will be at an airfield in two hours. Whether the forecasted condition of the field will be closed, low instrument, high instrument, low open, or high open, must be stated. This problem can be resolved by multiple-group discrimination. (Miller 1962)

(3) In the last year of secondary school, a student is given three tests: English, arithmetic, and form-relations.

On the basis of these test scores, the student is to be advised on a course of future study, and he has four choices available: engineering, building, art, or commerce. A multiple-group discriminant function can be used to classify this student. (Porebski 1966)

(4) The trace of an electrocardiogram is divided into 5-millisecond intervals, and a reading is made at these points. In addition, the length between the beginning and the ending of each peak, called the "QRS complex," is measured. On the basis of these measurements, a patient is to be diagnosed as a normal or abnormal person. This practice can be done by employing a polynomial discriminator to assign the patient to one of the two groups. (Specht 1967)

(5) In developing a certain rural area, a developer will consider the best strategy for this area to follow in its development. This problem may be resolved by determining which group the area falls into. For example, an area might be grouped as catering to recreation users, or attracting industry. Possible variables to consider include distance to the nearest city of population over 250,000, distance to the nearest major airport, percentage of land under forest, percentage of land under lakes, etc. (Bromley 1971)

(6) A study concerning the labor supply of married women teachers divides women into three employment groups:

full-time, part-time, and unemployed. Samples are taken and responses to variables such as wage of husband, full-time wage of wife, part-time wage of wife, household assets, ages of children, etc., are recorded. The assignments can be done by a multiple-group discriminant analysis. (Gramm 1973)

The aforementioned examples are only a small portion of the widespread applications of discriminant analysis. More references to the applications of these techniques are given in Sheth (1970), Lachenbruch (1975), Crask and Perreault (1977), and Eisenbeis (1977).

## 1.2 Statement of the Problem

The linear discriminant function (LDF) of Anderson (1958) described in Equation (1.1) has been widely accepted as a useful classification rule for discriminating between two multivariate normal populations. When the two competing populations are k-variate normal with known parameters and equal covariance matrices, the LDF, PQDF, and QDF are known to have the optimal property of minimizing the total probability of misclassification. When the covariance matrices are not equal and all the parameters are known, the optimal decision rule is a quadratic discriminant function (QDF). When all the parameters are known and the covariance matrices are unequal but proportional, the proportional quadratic discriminant function (PQDF) will also be an

optimal rule other than the QDF. The hierarchical relationship among these three discriminant functions can be expressed as LDF $\subset$ PQDF $\subset$ QDF.

In practice, the population parameters $p_1$, $p_2$, $u_1$, $u_2$, $V_1$, and $V_2$ are usually unknown and must be estimated from a training set, $\{(X_j,i),\ i=1,2,\ j=1,2,\ldots,n\}$, where i indicates the population from which $X_j$ comes. One common practice due to Wald (1944) and Day (1969) is to replace the unknown parameters in the discriminant function with their corresponding estimates: $n_1/N$, $n_2/N$, $\bar{X}_1$, $\bar{X}_2$, and $S_1$, and $S_2$, where $N= n_1+ n_2$, $n_i$ is the sample size, $\bar{X}_i$ is the k X 1 sample mean vector, and $S_i$ is the k X k sample covariance matrix of population i (i=1,2). However, the optimal property, in the sense of minimizing the total probability of misclassification for these three types of discriminant functions, will not necessarily hold when the parameters are estimated. Several researchers have studied and compared the performances of the LDF and QDF, both asymptotically and with finite sample sizes, but the comparisons of the PQDF to the LDF and QDF have never been made. In this dissertation, studies of the relative efficiency among LDF, QDF, and PQDF are confined to the case of two multivariate normal distributions. Methods for estimating the coefficient of proportionality in the PQDF are derived. Comparisons to the LDF and QDF are made through asymptotic analytic method and Monte Carlo simulation.

The primary problems imbedded in this research are five-fold. First, a method for the estimation of the coefficient of the proportionality, d, must be developed. Without an estimator of d, the performance of the estimated PQDF cannot be evaluated. Second, although the entire distribution of the estimated coefficient of proportionality is not needed for this research, the variance of this estimate has to be derived so as to derive the asymptotic error rate of the PQDF. Third, the estimation of the error rates using either the LDF, PQDF, or QDF plays an important role in this research. The asymptotic error rates (AERR) of these discriminant functions are to be calculated using analytic methods and numerical integration. Fourth, the performances of the LDF, QDF, and PQDF are to be compared. The relative efficiency (RE) will be used as the criterion for the comparisons of the performances among the three discriminant functions. The measure of the relative efficiency between two discriminant functions will be defined as the ratio of the deviations of their corresponding expected actual error rates from their optimal error rates. Fifth, the identification of regions in the parameter space where the PQDF outperforms the LDF and QDF is to be made. If the boundaries of the regions are ambiguous, this research will provide guidelines for delimiting these regions.

## 1.3 Purpose and Need for This Research

The reason for doing this research lies in the belief that the PQDF performs better than the LDF and QDF in a wide variety of situations, particularly when the available training sample sizes are small and the covariance matrices are not proportional.

All in all, the purpose of this research is three-fold: (1) to develop methods for estimating the coefficient of proportionality, (2) to measure both the asymptotic and small sample efficiency of the PQDF relative to the LDF and QDF, and (3) to define a region or regions in the parameter space where the PQDF outperforms, in the sense of minimizing the total probability of misclassification, the LDF and QDF. In addition, this research will provide guidelines for two-group normal discrimination in selecting the best discriminant function. From this, practitioners can accelerate the process of correct classification, use the available data effectively, select the best number of discriminating variables, and reduce the total probability of misclassification.

The need for this research becomes evident when one reviews the literature. Although several researchers have compared the behaviors of the LDF and QDF and provided some useful guidelines in assessing the likely agreement between the linear and quadratic rules, none have studied the behavior of the PQDF and its performance relative to the LDF

and QDF. The development of methods for estimating the error rates of the PQDF and identifying or delimiting the regions dominated by the PQDF will then begin to fill this void in the literature.

## 1.4 Summary and Dissertation Organization

In this chapter, the problem of discrimination was discussed as well as the efficiency in parameter estimation of the LDF, PQDF and QDF. The historical development of discriminant analysis was briefly reviewed. The problem studied in this research was explained, and finally, the purpose and need for this research was discussed.

A review of the pertinent literature is presented in Chapter II. Chapter III contains the experimental design, including the analytic method and the simulation study, for this research. Included in Chapter IV is an application of the method derived in this study. The final chapter, Chapter V, contains a summary of the dissertation, the conclusions derived from this research, and recommendations for future research.

# Chapter II

# REVIEW OF RELATED LITERATURE

## 2.1 The Linear Discriminant Function

It was Fisher (1936) who first proposed the linear function (LDF) to solve the problem of classifying iris plants into one of the three species: iris setosa, iris versicolor, and iris virginica, by four measurements: sepal length, sepal width, petal length, and petal width. However, the first person who adapted the assumption of normality and the likelihood ratio procedure was Abraham Wald (1944).

Wald (1944) followed the Neyman-Pearson Fundamental Lemma (Neyman and Pearson, 1936) and derived the LDF as follows: Let the k-variate random vector X be distributed as $N(\underline{u}1, V)$ in population 1 and $N(\underline{u}2, V)$ in population 2, and let $L(X|u,V)$ denote the joint probability density function of all variates at the sample point X. If $L(X|\underline{u}1,V)/L(X|\underline{u}2,V) \geq 1$ holds, the optimal rule is to assign X to population 1. After taking the logarithm, the left side of the above inequality can be transformed into:

$$(2.1) \quad W = X'V^{-1}(\underline{u}1 - \underline{u}2) - 1/2\ \{(\underline{u}1 + \underline{u}2)'V^{-1}(\underline{u}1 - \underline{u}2)\}.$$

Anderson (1951) followed Wald's (1944) study and considered the effect of a priori probabilities and misclassification costs. He gave a new version of Wald's inequality as follows: If

$$(2.2)\quad \frac{L(X|u_1,V)}{L(X|u_2,V)} \geq \frac{C(1|2)p_2}{C(2|1)p_1}$$

holds, the optimal rule is to assign X to population 1.

This version can be transformed into the Equation (1.1) in Section 1.1.1 by taking the logarithm of both sides. He further derived the distribution of Wald's W-statistic and extended the classification rule into the problem of more than two groups. Recently, Anderson (1973) continued his study and conducted an asymptotic extension of the distribution of the W-statistic.

## 2.2 The Quadratic Discriminant Function

Cavalli (1945) and Penrose (1947) undertook the studies in the problem of discrimination by considering two normal populations with different dispersions. However, their studies were limited to univariate problems.

It was Smith (1947) who first applied the log likelihood ratio procedure to derive the quadratic studentized form of the multivariate discriminant function. Later, Lubin (1950) extended Smith's (1947) method to multiple-group discriminant problems.

Following Penrose's (1947) concept and Smith's (1947) method, Bartlett and Please (1963) investigated the discrimination problem with zero mean differences. However, the most widely accepted version of the QDF, given in Equation (1.2), was not yet developed at that time.

The first version of the well-known quadratic discriminant function was given by Cooper (1963). Cooper (1963) applied Anderson's (1958) quadratic form to derive an optimal discriminant function for unequal covariance matrices. The decision rule for the multiple-group case was defined as follows:

In terms of a quadratic form

$$(2.3)\quad (X-u)'D(X-u) = \sum_{i=1}^{k}\sum_{j=1}^{k} d_{ij}(X_i-u_i)(X_j-u_j),$$

an unknown X is classified into one of g categories by assigning it to the $r$th category for which $F_r(X)$ is least, where

$$(2.4)\quad F_r(X) = (X-u_r)'D_r(X-u_r) + E_r \quad (r=1,2,\ldots,g),$$

and where $D_r$ is the covariance matrix of the $r$th group and $E_r$ is a constant.

For two-group classification, this discriminant function reduces to:

$$(2.5)\quad Q(X) = (X-u_2)'D_2(X-u_2) - (X-u_1)'D_1(X-u_1) + (E_2-E_1),$$

where $E\underline{1}$ and $E\underline{2}$ are some constants and it is optimal to assign a randomly selected individual X to population 1 if $Q(X) \geq 0$ holds, and to population 2 otherwise.

By letting $E\underline{1} = |D\underline{1}|$ and $E\underline{2} = |D\underline{2}|$, this decision rule is equivalent to that of Equation (1.2) in Section 1.1.1, apart from a constant.

## 2.3 The Proportional Quadratic Discriminant Function

The proportional quadratic discriminant function (PQDF) was first studied by Bartlett and Please (1963) for the case of zero mean differences. Bunke (1964) and Han (1968) followed Bartlett and Please's work and extended their studies to cases involving different mean vectors. However, the distribution of the PQDF was not studied until 1969.

Han (1969) derived the distribution of the discriminant function with known proportional covariance structures, V and dV (d> 1), for the case of two normal populations, $N(u\underline{1}, V)$ and $N(u\underline{2}, dV)$. Using the likelihood ratio procedure proposed by Welch (1939), the discriminant function he obtained, apart from some constant, is expressed as:

$$(2.6) \quad P(X) = (X - u\underline{1})' V^{-1} (X - u\underline{1}) - d^{-1} (X - u\underline{2})' V^{-1} (X - u\underline{2}).$$

Then he derived the distribution of P(X) for both known and unknown mean vectors. His procedure is briefly described as follows.

When $u\underline{1}$ and $u\underline{2}$ are known, without loss of generality, it can be assumed that $u\underline{1}=0$ and $u\underline{2}=\delta$, such that $\delta$ is a k X 1 arbitrary vector. Then P(X) will become

$$(2.7) \quad P(X) = X'V^{-1}X - d^{-1}(X-\delta)'V^{-1}(X-\delta).$$

With some algebra, P(X) can be written, apart from some constant, as

$$(2.8) \quad P(X) = (X+h\delta)'V^{-1}(X+h\delta),$$

where $h=1/(d-1)$. For the population in which X has a k-variate $N(0,V)$ distribution, P(X) is distributed as a non-central chi-square, denoted as $\chi'^2$, with k degrees of freedom and non-centrality parameter $h^2\Delta^2$, where $\Delta^2=\delta'V^{-1}\delta$ is the Mahalanobis's squared distance. For the other population in which X has a multivariate normal $N(\delta,dV)$ distribution, P(X) is distributed as $d\cdot\chi'^2$, where $\chi^2$ is the non-central chi-square with k degrees of freedom and non-centrality parameter $dh^2\Delta^2$.

When $u\underline{1}$ and $u\underline{2}$ are unknown, they are replaced by their corresponding maximum likelihood estimators $\bar{X}\underline{1}$ and $\bar{X}\underline{2}$. Thus Equation (2.6) becomes

$$(2.9) \quad P(X) = (X-\bar{X}\underline{1})'V^{-1}(X-\bar{X}\underline{1}) - d^{-1}(\bar{X}-X\underline{2})'V^{-1}(\bar{X}-X\underline{2}).$$

Han then used a linear transformation and Okamoto's (1963) procedure to derive the distribution of the P(X) given by Equation (2.9). Finally, the asymptotic expansion

for the cumulative distribution function (c.d.f.) of this PQDF was given in terms of $n_1$, $n_2$, d, h, $\Delta^2$, and G(X), the c.d.f. of a non-central chi-square variate X.

Recently, Han (1974) extended his study to the case of unknown mean vectors and partially known covariance matrices, i.e., either d or V is known. In the case of known d, the unknown parameters $u_1$, $u_2$, and V are estimated respectively by:

$$(2.10)\quad \bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i},$$

$$(2.11)\quad \bar{X}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} X_{2j},$$

and

$$(2.12)\quad S = \frac{1}{n_1 + n_2 - 2} \{ A_1 + \frac{1}{d} A_2 \},$$

where

$$(2.13)\quad A_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)', \quad i = 1, 2.$$

The estimators $\bar{X}_1$, $\bar{X}_2$, and $(n_1+n_2-2)S$ are independent of each other and distributed as $N(u_1, V/n_1)$, $N(u_2, dV/n_2)$, and $W(n_1+n_2-2, V)$, the Wishart distribution with degrees of freedom $n_1+n_2-2$ and covariance matrix V, respectively. With V estimated by S, Han obtained the discriminant function:

$$(2.14)\quad P(X) = \{X - \bar{X}\underline{1} + h(\bar{X}\underline{2} - \bar{X}\underline{1})\}' S^{-1} \{X - \bar{X}\underline{1} + h(\bar{X}\underline{2} - \bar{X}\underline{1})\}$$

$$- h(h+1)(\bar{X}\underline{1} - \bar{X}\underline{2})' S^{-1} (\bar{X}\underline{1} - \bar{X}\underline{2}),$$

where $h = 1/(d-1)$.

Then he employed a linear transformation and his (1969) procedure to derive the asymptotic expansion for the c.d.f. of the discriminant function in Equation (2.14), again, in terms of $n\underline{1}$, $n\underline{2}$, d, h, $\Delta^2$, and $G(X)$. He showed that the principal terms of the resulting cumulative distribution functions for both populations are of similar form, apart from some constant, to those in his previous study (1969).

For the case of known V and unknown d, the covariance matrix is completely specified under population $N(0,V)$. Estimators of $u\underline{1}$, $u\underline{2}$, and d are $\bar{X}\underline{1}$, $\bar{X}\underline{2}$, and $\hat{d}$, respectively, where

$$(2.15)\quad \hat{d} = \frac{1}{k(n\underline{2} - 1)} \sum_{i=1}^{n\underline{2}} (X\underline{2}i - \bar{X}\underline{2})' V^{-1} (X\underline{2}i - \bar{X}\underline{2}).$$

The distribution of $\hat{d}$ is $d\{k(n\underline{2}-1)\}^{-1}\chi^2$, where $\chi^2$ is the chi-square with $k(n\underline{2}-1)$ degrees of freedom and independent of $\bar{X}\underline{1}$ and $\bar{X}\underline{2}$. The analogous expression for the PQDF in Equation (2.14) is

$$(2.16)\quad P(X) = \{X - \bar{X}\underline{1} + \hat{h}(\bar{X}\underline{2} - \bar{X}\underline{1})\}' V^{-1} \{X - \bar{X}\underline{1} + \hat{h}(\bar{X}\underline{2} - \bar{X}\underline{1})\}$$

$$- \hat{h}(\hat{h}+1)(\bar{X}\underline{1} - \bar{X}\underline{2})' V^{-1} (\bar{X}\underline{1} - \bar{X}\underline{2}),$$

where $\hat{h}=1/(\hat{d}-1)$.

Again, he applied the same method used in deriving the asymptotic distribution of the discriminant function given by Equation (2.14) to this case. The asymptotic expansion for the c.d.f. of P(X) defined in Equation (2.16) was derived in terms of $n_1$, $n_2$, d, h, $\Delta^2$, and G(X). It was shown that the form of the asymptotic distribution of P(X) for one population is similar to those for the other population.

## 2.4 The Evaluation of Discriminant Functions

In the evaluation of a discriminant function five questions may arise:

(1) Are the observed between-group differences statistically significant?

(2) Is a subset of the variables sufficient for future classification? High dimensionality can make a discriminant function unwieldy and inefficient, thus dimension reduction is desirable for any type of analysis.

(3) Does the aggregate of the discriminant functions contribute significant information to the discrimination between any pair of groups? Moreover, is there any insignificant discriminant function that can be discarded?

(4) How well will the discriminant function perform on future samples? The performance of a discriminant function is commonly measured by the error rates (ERR). When the population parameters are unknown, various techniques are available for estimating the error rates.

(5) Is it possible to select the best discriminant function without investigating its performance? Several researchers have studied the relationship between the performance of a discriminant function and the parameters such as the number of variables, the training sample size, the distance between group means, etc. However, this research has been distribution-dependent.

### 2.4.1 Tests of Between-group Differences

The first problem of Section 2.4 can be solved by the Hotelling $T^2$ test if there are two groups (Hotelling 1931, 1951). This $T^2$ test performs the same function in multivariate analysis as "Student's t" does in univariate analysis. Consider the following composite hypothesis:

$$H\underline{0}: u\underline{1} = u\underline{2},$$

$$H\underline{1}: u\underline{1} \neq u\underline{2}.$$

The test statistic for this hypothesis test is:

$$(2.17)\quad F = \frac{n_1 + n_2 - k - 1}{(n_1 + n_2 - 2)k} \cdot T^2$$

$$= \frac{n_1 + n_2 - k - 1}{(n_1 + n_2 - 2)k} \cdot \frac{n_1 n_2}{n_1 + n_2} \cdot D^2.$$

This quantity F has the variance ratio F distribution with degrees of freedom k and $n_1+n_2-k-1$, thus the decision rule for the regular F test can be applied here.

The tests of group differences in the two or more groups are provided by Bartlett (1947a, 1947b) and Rao (1946, 1950b, 1952, 1970) using functions of Wilks's $\Lambda$ criterion (Wilks 1932), as the test statistic. Wilks's $\Lambda$ is given by

$$(2.18)\quad \Lambda = \frac{|W|}{|T|},$$

where T is the k X k matrix of total-sample sums-of-squares and cross-products (SSCP) and W is the k X k matrix of within group SSCP.

To test the pertinent composite hypothesis:

$H_0$: $u_1 = u_2 = \ldots = u_i = \ldots = u_g$,

$H_1$: Not all $u_i$ are equal,

Bartlett proposed a test statistic, V, which is a logarithmic function of Wilks's $\Lambda$:

$$(2.19)\quad V = -\{N - 1 - (k + g)/2\} \cdot \log \Lambda,$$

where k is the number of variables, g is the number of groups, and $N=n_1+n_2+\ldots+n_g$.

This function V was shown to be distributed approximately as a chi-square random variable with $k(g-1)$ degrees of freedom, provided $N-1-(k+g)/2$ is large. Schatzoff (1964, 1966) demonstrated that the chi-square approximation to Bartlett's V is reasonably good for moderately large sample sizes.

Another function of $\Lambda$, due to Rao (1946, 1950b, 1952, 1970), seems to offer a better approximate test than Bartlett's V. This function is:

$$(2.20)\quad Ra = \frac{1-\Lambda^{1/c}}{\Lambda^{1/c}} \cdot \frac{mc-k(g-1)/2+1}{k(g-1)},$$

where

$$m = N-1-(k+g)/2,$$

$$c = \sqrt{\frac{k^2(g-1)^2-4}{k^2+(g-1)^2-5}},$$

and $c=1$ if $k^2+(g-1)^2=5$.

This test statistic, Ra, was shown to have approximately an F distribution with degrees of freedom $k(g-1)$ and $mc-k(g-1)/2+1$.

### 2.4.2 Tests for Sufficiency of a Subset of Variables

Rao (1946, 1950b, 1952, 1970) developed procedures to examine the sufficiency of a subset of variables for two-group discrimination. A specified subset, $\{X\underline{1}, X\underline{2}, \ldots, X\underline{k'}\}$ for $k' \leq k$, is to be tested for sufficiency as discriminators against the full set, $\{X\underline{1}, X\underline{2}, \ldots, X\underline{k}\}$. Rao derived the statistic:

$$(2.21)\quad F = \frac{n\underline{1} + n\underline{2} - k - 1}{k - k'} \cdot \frac{C(D^2\underline{k} - D^2\underline{k'})}{1 + CD^2\underline{k'}},$$

where

$$(2.22)\quad C = \frac{n\underline{1}\ n\underline{2}}{(n\underline{1} + n\underline{2})(n\underline{1} + n\underline{2} - 2)}$$

and $D^2\underline{k}$ and $D^2\underline{k'}$ are Mahalanobis's $D^2$ statistics on the full set and subset of variables, respectively. This quantity F has a variance ratio F distribution with $k-k'$ and $n\underline{1}+n\underline{2}-k-1$ degrees of freedom. If $k'=k-1$, in which it is to be determined if a single specific variable has discriminating power or not, this F statistic with 1 and $n\underline{1}+n\underline{2}-k-1$ degrees of freedom is equivalent to a $t^2$ statistic. In this case, Rao's F-test can be transformed into a t-test.

### 2.4.3 A Test for Discriminatory Power

Fisher (1938) developed a significance test for an assigned discriminant function in the two-group case. If the assigned function is:

$$Y = \beta\underline{1}\ X\underline{1} + \beta\underline{2}\ X\underline{2} + \ldots + \beta\underline{k}\ X\underline{k},$$

then

$$D^2\underline{y} = \frac{(\bar{Y}\underline{1} - \bar{Y}\underline{2})^2}{V(Y)},$$

where $\bar{Y}\underline{1}$ and $\bar{Y}\underline{2}$ are the mean values of Y for the two groups and V(Y) is the variance of Y.

To test whether the assigned discriminant function contributes significant discriminatory power, the following hypothesis is of major interest:

$H\underline{0}$: $\beta\underline{1} = \beta\underline{2} = \ldots = \beta\underline{k} = 0$,

$H\underline{1}$: Not all $\beta j$ $(j=1,2,\ldots,k)$ equal 0.

The test statistic is:

$$(2.23)\quad U = \frac{1 + n\underline{1}\, n\underline{2}\, D^2/\{(n\underline{1}+n\underline{2})(n\underline{1}+n\underline{2}-2)\}}{1 + n\underline{1}\, n\underline{2}\, D^2\underline{y}/\{(n\underline{1}+n\underline{2})(n\underline{1}+n\underline{2}-2)\}} - 1,$$

where $D^2$ and $D^2\underline{y}$ are the Mahalanobis's $D^2$ statistics based on the variables X's and Y, respectively. For testing the above composite hypothesis, the quantity

$$(2.24)\quad Fi = \frac{n\underline{1} + n\underline{2} - k - 1}{k - 1} \cdot U$$

can be used. The statistic Fi has the variance ratio F distribution with k-1 and $n\underline{1}+n\underline{2}-k-1$ degrees of freedom, if $H\underline{0}$ is true.

For multiple-group problems, the Bartlett's test, described in Section 2.4.1 and given by Equation (2.19), is available (Bartlett 1947a).

By equating the reciprocals of both sides of Equation (2.18) and successively transforming, it is seen that

$$(2.25)\quad \Lambda^{-1} = |T|/|W| = |W^{-1}T| = |W^{-1}(W + B)| = |I + W^{-1}B| = (1 + \lambda_1)(1 + \lambda_2)\cdots(1 + \lambda_r),$$

where B is the k X k matrix of between group SSCP, $T=W + B$, $r \leq \min(k, g-1)$, $\lambda_i$ $(i=1,2,\ldots,r)$ is the nonzero eigenvalue of $W^{-1}B$, and $\lambda_1 > \lambda_2 > \ldots > \lambda_r$. Consequently, Bartlett's V statistic for testing the significance of an observed $\Lambda$ value, given in Equation (2.19), can be expressed as:

$$(2.26)\quad V = -\{N - 1 - (k + g)/2\} \log \Lambda = \{N - 1 - (k + g)/2\} \cdot \log\{(1 + \lambda_1)(1 + \lambda_2)\cdots(1 + \lambda_r)\} = \{N - 1 - (k + g)/2\} \cdot \log \prod_{j=1}^{r} (1 + \lambda_j) = \{N - 1 - (k + g)/2\} \cdot \sum_{j=1}^{r} \log (1 + \lambda_j),$$

where $r \leq \min(k, g-1)$ and $N = n_1 + n_2 + \ldots + n_g$.

Because the successive discriminant functions are uncorrelated, the successive terms $\log(1+\lambda_j)$ in the last expression above are statistically independent (assuming multivariate normality of the origional k variables). As a result, the additive components of V are each approximately distributed as a chi-square variate. More specifically, the jth component of V, $V_j = \{N- 1- (k+ g)/2\} \log(1+ \lambda_j)$ is approximately a chi-square with k+g-2j degrees of freedom. Thus this V can be partitioned as follows:

TABLE 1

Decomposition of Bartlett's V Statistic

| Discriminant Function | Chi-square Test Statistic | Degrees of Freedom |
|---|---|---|
| First (j=1) | $\{N- 1- (k+g)/2\} \log(1+ \lambda_1)$ | (k+ g- 2) |
| Second (j=2) | $\{N- 1- (k+g)/2\} \log(1+ \lambda_2)$ | (k+ g- 4) |
| ⋮ | ⋮ | ⋮ |
| Last (j=r) | $\{N- 1- (k+g)/2\} \log(1+ \lambda_r)$ | (k+ g- 2r) |

The individual discriminant functions can now be tested, using the decision rule based on the chi-square distribution, for its contribution to the discriminatory power. Notice that Rao's Ra statistic, defined by Equation (2.20), does not have this decomposition feature.

Alternatively, when we cumulatively subtract V1, V2, and so forth from V, the remainder each time is also a chi-square variate. These successive remainders become appropriate statistics for testing whether the residual discrimination, after partialing out the first discriminant function, the first and the second discriminant functions, and so forth, is statistically significant. The successive test statistics and their corresponding degrees of freedom are summarized below:

TABLE 2

Bartlett's V Statistic for Residual Discriminant Functions

| Residual After Removing | Approximate Chi-square Statistic | Degrees of Freedom |
|---|---|---|
| First discriminant function | V- V1 | $k(g-1)-(k+g-2) = (k-1)(g-2)$ |
| First 2 discriminant functions | V- V1- V2 | $(k-1)(g-2)-(k+g-4) = (k-2)(g-3)$ |
| First 3 discriminant functions | V- V1- V2- V3 | $(k-2)(g-3)-(k+g-6) = (k-3)(g-4)$ |
| . | . | . |
| . | . | . |
| . | . | . |

When the residual, after removing the first m discriminant functions, becomes smaller than the prescribed fractile of the appropriate chi-square distribution, it is concluded that only the first m discriminant functions are significant

at that prescribed centile confidence level (Bartlett 1947b).

### 2.4.4 The Estimation of the Error Rates

In this section, for simplicity, only the case of two-group linear normal classification errors will be discussed. The error rates of interest, defined by Hills (1966) and Lachenbruch (1975), are as follows:

(1) The **optimal error rate**, $P\underline{1}(D\underline{t})$ --- the error rate obtained from the optimal discriminant function with known parameters:

$$(2.27)\quad T(R,f) = p\underline{1}\int_{R\underline{2}} f(\mathbf{X}|1)\,d\mathbf{X} + p\underline{2}\int_{R\underline{1}} f(\mathbf{X}|2)\,d\mathbf{X},$$

where R refers to the optimal classification regions, f refers to the presumed distribution of the observations that will be classified, $p\underline{i}$ $(i=1,2)$ is the a priori probability, and $f(\mathbf{X}|i)$ $(i=1,2)$ is the density function of $\mathbf{X}$ in population i.

Consider the case of equal a priori probabilities, equal misclassification costs, and equal dispersions. The optimal classification regions are

$$(2.28)\quad R\underline{1} = \{\mathbf{X}: D\underline{t}(\mathbf{X}) = (\mathbf{X} - 1/2\ (\mathbf{u}\underline{1}+\mathbf{u}\underline{2}))'\mathbf{V}^{-1}(\mathbf{u}\underline{1}-\mathbf{u}\underline{2}) > 0\},$$

$$R\underline{2} = \{\mathbf{X}: D\underline{t}(\mathbf{X}) = (\mathbf{X} - 1/2\ (\mathbf{u}\underline{1}+\mathbf{u}\underline{2}))'\mathbf{V}^{-1}(\mathbf{u}\underline{1}-\mathbf{u}\underline{2}) < 0\}.$$

Alternatively, the optimal error rate can be expressed as:

$$(2.29)\quad T(R,f) = p\underline{1}\ \Pr\{Dt(\mathbf{X}) < 0 \mid \Pi\,\underline{1}\} + p\underline{2}\ \Pr\{Dt(\mathbf{X}) > 0 \mid \Pi\,\underline{2}\}$$

$$= 1/2\ \Phi(-\Delta/2) + 1/2\ \Phi(-\Delta/2)$$

$$= \Phi(-\Delta/2),$$

where $\Phi$ is the cumulative standard normal distribution function and $\Delta$ is the Mahalanobis's distance.

(2) The **actual error rate,** $P\underline{3}(Ds)$ --- the error rate for the sample discriminant function as it will perform in future samples:

$$(2.30)\quad T(\hat{R},f) = p\underline{1} \int_{\hat{R}\underline{2}} f(\mathbf{X}\mid 1)\ d\mathbf{X} + p\underline{2} \int_{\hat{R}\underline{1}} f(\mathbf{X}\mid 2)\ d\mathbf{X}$$

$$= p\underline{1}\ \Pr\{Ds(\mathbf{X}) < 0 \mid \Pi\,\underline{1}\} + p\underline{2}\ \Pr\{Ds(\mathbf{X}) > 0 \mid \Pi\,\underline{2}\}$$

$$= 1/2\ \Phi\{-Ds(u\underline{1})/\sqrt{V(D)}\} + 1/2\ \Phi\{Ds(u\underline{2})/\sqrt{V(D)}\},$$

where

$$(2.31)\quad \hat{R}\underline{1} = \{\mathbf{X}:\ Ds(\mathbf{X}) = (\mathbf{X} - 1/2\ (\bar{\mathbf{X}}\underline{1}+\bar{\mathbf{X}}\underline{2}))'S^{-1}(\bar{\mathbf{X}}\underline{1}-\bar{\mathbf{X}}\underline{2}) > 0\},$$

$$\hat{R}\underline{2} = \{\mathbf{X}:\ Ds(\mathbf{X}) = (\mathbf{X} - 1/2\ (\bar{\mathbf{X}}\underline{1}+\bar{\mathbf{X}}\underline{2}))'S^{-1}(\bar{\mathbf{X}}\underline{1}-\bar{\mathbf{X}}\underline{2}) < 0\},$$

and

$$(2.32)\quad V(D) = (\bar{\mathbf{X}}\underline{1} - \bar{\mathbf{X}}\underline{2})'S^{-1}\ V\ S^{-1}(\bar{\mathbf{X}}\underline{1} - \bar{\mathbf{X}}\underline{2}).$$

(3) The **estimated actual error rate,** $P\underline{2}(Ds)$ --- the plug-in estimate of the actual error rate obtained by replacing the parameters in $f(\mathbf{X}\mid 1)$ and $f(\mathbf{X}\mid 2)$ with their corresponding sample estimates:

$$(2.33)\quad T(\hat{R},\hat{f}) = p\underline{1}\int_{\hat{R}\underline{2}} \hat{f}(X|1)\,dX + p\underline{2}\int_{\hat{R}\underline{1}} \hat{f}(X|2)\,dX$$

$$= p\underline{1}\ Pr\{D_{\underline{s}}(X) < 0 | \Pi\underline{1}\} + p\underline{2}\ Pr\{D_{\underline{s}}(X) > 0 | \Pi\underline{2}\}$$

$$= 1/2\ \Phi\{-D_{\underline{s}}(\bar{X}\underline{1})/\sqrt{S(D)}\} + 1/2\ \Phi\{D_{\underline{s}}(\bar{X}\underline{2})/\sqrt{S(D)}\},$$

where

$$(2.34)\quad D_{\underline{s}}(\bar{X}\underline{1}) = (\bar{X}\underline{1} - 1/2\ (\bar{X}\underline{1}+\bar{X}\underline{2}))'S^{-1}(\bar{X}\underline{1}-\bar{X}\underline{2}) = D^2/2,$$

$$D_{\underline{s}}(\bar{X}\underline{2}) = (\bar{X}\underline{2} - 1/2\ (\bar{X}\underline{1}+\bar{X}\underline{2}))'S^{-1}(\bar{X}\underline{1}-\bar{X}\underline{2}) = -D^2/2,$$

and

$$(2.35)\quad S(D) = (\bar{X}\underline{1} - \bar{X}\underline{2})'S^{-1}\ S\ S^{-1}(\bar{X}\underline{1} - \bar{X}\underline{2})$$

$$= (\bar{X}\underline{1} - \bar{X}\underline{2})'S^{-1}(\bar{X}\underline{1} - \bar{X}\underline{2})$$

$$= D^2.$$

Thus

$$(2.36)\quad T(\hat{R},\hat{f}) = 1/2\ \Phi(-D/2) + 1/2\ \Phi(-D/2) = \Phi(-D/2).$$

(4) The **expected actual error rate,** $P\underline{4}(D_{\underline{s}})$ --- the expected value of the actual error rate calculated over all possible training samples of size $n\underline{1}$ and $n\underline{2}$:

$$(2.37)\quad E\{T(\hat{R},f)\} = E\{p\underline{1}\int_{\hat{R}\underline{2}} f(X|1)\,dX + p\underline{2}\int_{\hat{R}\underline{1}} f(X|2)\,dX\}.$$

Several techniques are currently used to estimate errors of sample discriminant functions. The alternatives are basically of three types: (1) those using samples to

estimate error rates without any distributional information, (2) those using the assumption of normality, and (3) those using the holdout procedures. Each of several such methods, in two-group and equal dispersions case, is briefly described in what follows. The limitations and strengths of each method are noted. In those methods using the assumption of normality, equal a priori probabilities are assumed. Notice that most of these methods estimate the expected actual error rate.

(1) **Resubstitution Method** or the **R Method:** The training sample is classified by the estimated discriminant function. The $P(1|2)$ and $P(2|1)$ are then estimated by the sample proportion of the misclassified observations, i.e., $P(i|j)=n_{ij}/n_j$, where $n_{ij}$ $(i\neq j$ and $i,j=1,2)$ is the number of sample observations assigned to population i which actually came from population j (Smith 1947). The estimate $(n_{12}+n_{21})/(n_1+n_2)$, often called apparent error rate (APPERR), is consistent but may be badly biased and overly optimistic for small samples in particular.

(2) **Holdout Method** or the **H Method:** The sample is split into two mutually exclusive samples. One is used to estimate (or train) discriminant functions which are then employed to classify the other (or holdout) sample and estimate $P(1|2)$ and $P(2|1)$ with $n_{12}/n_2$ and $n_{21}/n_1$, respectively (Highleyman 1962, Kanal and Chandrasekaran 1971). These estimates are consistent and unbiased but

require large samples. This method obviously suffers from inefficiency in using the available data. Unless the sample size is large, the H method tends to give an overly pessimistic estimate of performance of the DF estimated using the entire sample.

(3) **Data Shuffling Method** or the **S Method:** The sample is randomly partitioned into two mutually exclusive samples of equal sizes for K times. The H method is applied here to the each pair of samples and thus K different estimates of error rate are obtained. These K estimates are then averaged to give a final estimate for the expected estimate of the error rate (Duda and Hart 1973). Although this method uses the data more efficiently than the H method, it still uses only half of the available data for training each time. Consequently, it also tends to give overly pessimistic results.

(4) **$D^2$-estimate Method** or the **D Method:** When the populations are normal and the parameters are known with equal covariances, $P(1|2)$ and $P(2|1)$ are given by

$$(2.38)\quad P(1|2) = P(2|1) = \Phi(-\sqrt{\Delta^2}/2),$$

where $\Phi$ is the cumulative standard normal distribution function and $\Delta^2 = (\underline{u}1-\underline{u}2)'\Psi^{-1}(\underline{u}1-\underline{u}2)$ is the Mahalanobis's squared distance. When the parameters are unknown, this D method replaces $\Delta^2$ with $D^2$, where

$$(2.39)\quad D^2 = (\bar{\underline{X}}1 - \bar{\underline{X}}2)'S^{-1}(\bar{\underline{X}}1 - \bar{\underline{X}}2)$$

(Wald 1944, Anderson 1951, 1958, Sitgreaves 1952).

The estimates of D are consistent but are biased upward. Hence the estimators of the P(1|2) and P(2|1) are consistent but may be badly biased downward, particularly for small samples.

(5) **Unbiased $D^2$-estimate Method or the D* Method:** This method simply substitutes an unbiased estimate of $\Delta$ for the biased estimate D in the D method. The unbiased estimator of $\Delta^2$, based on $D^2$, may be obtained from the non-central F distribution (Lachenbruch 1968). This estimate is given by

$$(2.40)\quad D^{*2} = \frac{n_1 + n_2 - k - 3}{n_1 + n_2 - 2} \cdot D^2 - \frac{(n_1 + n_2)\,k}{n_1 \cdot n_2}.$$

This method gives a consistent estimate of the error rate. Unfortunately, when $n_1$ and $n_2$ are small relative to k and $D^2$ is also small, $D^{*2}$ is frequently negative which precludes its use.

(6) The **DS Method:** Instead of using $D^{*2}$, one may construct an estimate of $\Delta^2$ using the quantity

$$(2.41)\quad Ds^2 = \frac{n_1 + n_2 - k - 3}{n_1 + n_2 - 2} \cdot D^2.$$

This estimator merely ignores the constant term in $D^{*2}$ given by Equation (2.40) above (Lachenbruch and Mickey 1968).

The DS method appears to give consistency, but is slightly optimistic of the error rate.

(7) The O Method: This method is based on a result of Okamoto (1963) who gave an asymptotic approximation for P(1|2) and P(2|1) in terms of $n\underline{1}$, $n\underline{2}$, k, and $\Delta^2$. By using $D^2$ as an estimate of $\Delta^2$ in Okamoto's expansions, this method allows us to have different estimates for P(1|2) and P(2|1), which we could not do with the D or DS methods. When the groups are close together and the sample sizes are small, this approximation is found to be poor. For large values of $D^2$, this method and the D method give similar results because the correction terms are proportional to $1/D^2$. However, if $D^2$ is moderate (say, $1 \leq D^2 \leq 10$), it is found to be better than the D method.

(8) The OS Method: Instead of substituting $D^2$ for $\Delta^2$ in the O method, this method uses $Ds^2$ as an estimate of $\Delta^2$. It also makes it possible to estimate P(1|2) and P(2|1), separately, and generally performs better than the O method.

(9) Leave-one-out Method or the U Method: In this method, one observation at a time is held out. The discriminant function is estimated using $n\underline{1}+n\underline{2}-1$ observations and the held out observation is then classified. If a held out observation is misclassified, the estimate of the error rate is set to be 1, and 0 otherwise. This procedure is repeated until all observations are classified, thus $n\underline{1}+n\underline{2}$ binary scores are taken. The final

estimate of the error rate is obtained by averaging the values of the $n_1+n_2$ binary estimates (Lunts and Brailovskiy 1964, Lachenbruch 1965).

In order to reduce the number of inversions required for the computations of the $n_1+n_2$ discriminant functions, Lachenbruch (1965) developed a method which requires only one explicit inversion, namely, the inversion of the sample covariance matrix based on the entire sample. An identity first given by Bartlett (1951) is used, i.e., if $B = A + cuv'$, then $B^{-1} = A^{-1} - \{(cA^{-1}uv'A^{-1})/(1 + cv'A^{-1}u)\}$, where B and A are k X k nonsingular matrices, u and v are k X 1 vectors, and c is an arbitrary scalar. Then the discriminant score for $X_j$ as computed without the sample point $X_j$ is given by:

$$(2.42)\quad D_s(X_j) = \{X_j - 1/2\,(\bar{X}_1 + \bar{X}_2 - (X_j - \bar{X}_\theta)/(n_\theta - 1))\}' S_j^{-1} \cdot \{\bar{X}_1 - \bar{X}_2 + (-1)^\theta (X_j - \bar{X}_\theta)/(n_\theta - 1)\},$$

where $\bar{X}_1$, $\bar{X}_2$ are the total sample mean vectors, $X_j$ is the holdout observation from the $\theta$th sample, $\bar{X}_\theta$ is the mean vector computed without $X_j$, $n_\theta$ is the sample size of the $\theta$th sample, and

$$(2.43)\quad S_j^{-1} = \frac{n_1 + n_2 - 3}{n_1 + n_2 - 2}\left\{S^{-1} + \frac{C_\theta\, S^{-1}(X_j - \bar{X}_\theta)(X_j - \bar{X}_\theta)' S^{-1}}{1 - C_\theta\,(X_j - \bar{X}_\theta)' S^{-1}(X_j - \bar{X}_\theta)}\right\},$$

where

$$(2.44)\quad C_\theta = \frac{n_\theta}{(n_\theta - 1)(n_1 + n_2 - 2)},$$

and S is the pooled total sample covariance matrix.

This method gives almost unbiased estimates for the expected actual error rate and the confidence intervals of P(1|2) and P(2|1). The U method avoids the sample size limitation associated with the H and S methods because of its efficient use of the available data. Furthermore, it has been found to perform better than the other methods aforementioned when approximate normality is questionable (e.g., when a large number of dichotomies are used as variables) and the sample size is small relative to the number of variables. However, this method has two disadvantages: (1) it does not give a minimum variance estimate, and (2) it requires excessive computation in the distribution-free case, unless the sample size is small.

(10) The $\bar{U}$ **Method:** This method combines the features of the U method and the use of the normal distribution. The scores of $n_1+n_2$ sample discriminant functions using the U method are found and then the mean and standard deviation, $u(D_i)$ and $s(D_i)$ $(i=1,2)$, of these discriminant scores within each group are used to estimate $P(2|1)$ by $\Phi\{-u(D_1)/\, s(D_1)\}$ and $P(1|2)$ by $\Phi\{u(D_2)/\, s(D_2)\}$ (Lachenbruch and Mickey 1968). These estimates are better than those obtained using the U method when normality can be assumed. If $D^2$ is small, or the training sample size is small relative to the number of parameters, this method performs better than the OS method.

(11) The $\Pi$ Method: Let h be an arbitrary integer such that $1 \leq h < N$ ($N=n_1+n_2$), $h/N \leq 1/2$, and $m=N/h$ is an integer. This method randomly partitions the whole sample into m disjoint subsets of equal sizes, holds one subset out at a time, estimates the discriminant function based on $n_1+n_2-h$ observations, and classifies the held out observations. This procedure is repeated m times and each time the proportion of misclassification is calculated. The resulting estimate of the error rate is obtained by averaging the values of m computed estimates (Toussaint 1969, Toussaint and Donaldson 1970). Notice that when h=1 and h=N/2, the $\Pi$ method reduces to the U method and H method. This method gives less biased results than the H method (depending respectively on the values of q, n, and k) and requires less computation than the U method. Because of these results, it is a method well suited to "medium-sized" data sets.

In summary, none of the above methods is uniformly best for all situations. The performances of these methods depend heavily upon the number of parameters, the training sample size, the distance between groups, and the distribution of the populations. Lachenbruch and Mickey (1968), Cochran (1968), Toussaint (1974), and Dillon (1979) have evaluated some or all of these methods. As might be expected, the R method and D method perform poorly. The U or $\bar{U}$ methods appear to be the best for small samples.

Overall, the OS method seems the best for normal data. However, the U, $\prod$, R, and H methods are distribution-free and most easily generalized to more than two group problems as well as the unequal dispersion cases (Eisenbeis and Avery 1972, Fukunaga and Kessel 1972). In this respect then, the U method would appear to be superior based upon current evidence. Other methods such as the triple cross-validation method, the leave-two-out method, the F method, and the posterior probability method, can be found in Norman (1965), Mosteller and Tukey (1968), Lissack and Fu (1972), Glick (1978), respectively.

### 2.4.5 The Relation of the Distribution Parameters to the Performance

Dunn (1971) investigated some expected values for the probabilities of correct classification in linear discriminant analysis. She used Monte Carlo simulation to obtain estimates of the unconditional probability of correct classification and the expected value of this estimate based on the calculated Mahalanobis's distance. The experimental variables were k= 2, 6, 10, 15, 20, 30, $\Delta/2$= 0, 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, $\underline{u}1$= $(-\Delta/2,0,\ldots,0)$, $\underline{u}2$= $(\Delta/2,0,\ldots,0)$, $\underline{V}1=\underline{V}2=I$, and $n\underline{1}=n\underline{2}=n$= 25, 50, 100 with corresponding replications r= 2000, 200, 200. Her results are presented in Table 3.

TABLE 3

Dunn's Estimate of the Unconditional Probability of Correct Classification Resulting from the Use of LDF*

| | | | $\Delta/2$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 0.25 | 0.50 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 | 3.50 |
| k | n | P1= | .5987 | .6915 | .8413 | .9332 | .9773 | .9938 | .9987 | .9998 |
| 2 | 25 | | .5779 | .6796 | .8339 | .9281 | .9746 | .9926 | .9982 | .9996 |
| | 50 | | .5892 | .6864 | .8378 | .9310 | .9760 | .9933 | .9985 | .9997 |
| | 100 | | .5946 | .6892 | .8395 | .9321 | .9766 | .9935 | .9986 | .9997 |
| 6 | 25 | | .5527 | .6527 | .8143 | .9153 | .9676 | .9894 | .9972 | .9993 |
| | 50 | | .5690 | .6713 | .8281 | .9249 | .9730 | .9921 | .9982 | .9996 |
| 10 | 25 | | .5428 | .6302 | .7940 | .9000 | .9586 | .9853 | .9955 | .9988 |
| | 50 | | .5576 | .6558 | .8177 | .9177 | .9694 | .9906 | .9976 | .9994 |
| 15 | 25 | | .5350 | .6109 | .7675 | .8788 | .9442 | .9778 | .9922 | .9976 |
| | 50 | | .5490 | .6418 | .8045 | .9102 | .9644 | .9884 | .9968 | .9993 |
| 20 | 25 | | .5282 | .5943 | .7427 | .8566 | .9283 | .9182 | .9870 | .9952 |
| | 50 | | .5446 | .6322 | .7938 | .9009 | .9585 | .9855 | .9958 | .9989 |
| 30 | 25 | | .5187 | .5681 | .6891 | .7984 | .8771 | .9283 | .9608 | .9799 |
| | 50 | | .5345 | .6113 | .7682 | .8803 | .9463 | .9793 | .9930 | .9980 |
| | 100 | | .5505 | .6428 | .8052 | .9097 | .9647 | .9886 | .9969 | .9993 |

Source: Olive J. Dunn (1971).

* This estimate equals one minus the estimate of the expected actual error rate.

It was shown that with other parameters being fixed, the actual unconditional probability of correct classification, P4, is a decreasing function of k, an increasing function of n, and an increasing function of $\Delta$. The difference between P4 and the optimal probability P1 is an increasing function of k. Dunn also compared her

estimate with Lachenbruch's (1968) estimate P4. She found that Lachenbruch's estimator P4 is a useful conservative approximation to the unconditional probability of correct classification. Lachenbruch's estimate is presented in Table 4.

TABLE 4

Lachenbruch's Estimate of the Unconditional Probability of Correct Classification Resulting from the Use of LDF*

| | | | Δ/2 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 0.25 | 0.50 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 | 3.50 |
| k | n | P1= | .5987 | .6915 | .8413 | .9332 | .9773 | .9938 | .9987 | .9998 |
| 2 | 25 | | .5749 | .6733 | .8285 | .9248 | .9729 | .9920 | .9981 | .9996 |
| | 50 | | .5848 | .6821 | .8331 | .9292 | .9752 | .9930 | .9984 | .9997 |
| | 100 | | .5911 | .6867 | .8383 | .9312 | .9763 | .9934 | .9985 | .9997 |
| 6 | 25 | | .5537 | .6476 | .8082 | .9112 | .9654 | .9888 | .9970 | .9993 |
| | 50 | | .5683 | .6673 | .8252 | .9231 | .9721 | .9917 | .9980 | .9996 |
| 10 | 25 | | .5425 | .6277 | .7877 | .8956 | .9561 | .9843 | .9953 | .9988 |
| | 50 | | .5580 | .6546 | .8153 | .9165 | .9686 | .9902 | .9975 | .9995 |
| 15 | 25 | | .5336 | .6078 | .7616 | .8735 | .9412 | .9762 | .9917 | .9975 |
| | 50 | | .5493 | .6410 | .8029 | .9077 | .9636 | .9880 | .9967 | .9993 |
| 20 | 25 | | .5272 | .5911 | .7349 | .8479 | .9216 | .9641 | .9854 | .9947 |
| | 50 | | .5430 | .6292 | .7904 | .8983 | .9579 | .9853 | .9957 | .9989 |
| 30 | 25 | | .5179 | .5626 | .6807 | .7808 | .8608 | .9174 | .9544 | .9765 |
| | 50 | | .5341 | .6095 | .7653 | .8774 | .9442 | .9780 | .9925 | .9978 |
| | 100 | | .5495 | .6416 | .8040 | .9087 | .9642 | .9885 | .9968 | .9993 |

Source: Olive J. Dunn (1971).

* This estimate equals one minus the estimate of the expected actual error rate.

Boullion, Odell and Duran (1975) studied the effect of the number of variates on the expected apparent error rate in linear discrimination. They used Monte Carlo simulation with the experimental variables k= 3, 6, 9, 12, $n_1 = n_2 = n$= 5, 10, 15, 20, 50, 100, r= 50, $\Delta$= 1, 2, 3, 4, 6, 10, 12, 20, 25, $u_1$= (0,0,...,0), $u_2$= ($\Delta$,0,0,0,), and $V_1 = V_2 = I$. Some of their results are presented in Table 5.

The relationship between the expected apparent error rate and k, n, and $\Delta$, respectively, was found to be similar to Dunn's (1971) result. They concluded that for small sample sizes, a subset of the variates can be chosen so as to yield better classification results than those obtained by using the full complement of variates.

TABLE 5

Estimates of the Expected Apparent Error Rates Using the LDF and the Resubstitution Method with 50 Replications

| $\Delta^2$ | k | n = 5 | 10 | 15 | 20 | 50 | 100 | Asym. |
|---|---|---|---|---|---|---|---|---|
| 1.0 | 3 | .3960 | .3672 | .3568 | .3432 | .3200 | .3110 | |
| | 6 | .4176 | .3824 | .3632 | .3520 | .3252 | .3150 | |
| | 9 | .4192 | .4052 | .3775 | .3556 | .3350 | .3208 | .3085 |
| | 12 | | .4084 | .3884 | .3960 | .3524 | .3290 | |
| 2.0 | 3 | .3292 | .2768 | .2724 | .2600 | .2444 | .2337 | |
| | 6 | .3324 | .3224 | .2940 | .2796 | .2552 | .2460 | |
| | 9 | .3416 | .3304 | .3176 | .3000 | .2572 | .2670 | .2389 |
| | 12 | | .3584 | .3224 | .3392 | .2820 | .2730 | |
| 3.0 | 3 | .2732 | .2336 | .2344 | .2116 | .2040 | .1875 | |
| | 6 | .2792 | .2763 | .2428 | .2300 | .2072 | .2020 | |
| | 9 | .2932 | .2888 | .2596 | .2464 | .2072 | .2050 | .1922 |
| | 12 | | .3088 | .2924 | .2884 | .2268 | .2050 | |
| 4.0 | 3 | .2396 | .2004 | .1868 | .1828 | .1672 | .1560 | |
| | 6 | .2452 | .2356 | .2084 | .1856 | .1724 | .1620 | |
| | 9 | .2700 | .2632 | .2352 | .2284 | .1728 | .1970 | .1587 |
| | 12 | | .2780 | .2490 | .2552 | .1916 | .2048 | |
| 6.0 | 3 | .1872 | .1472 | .1364 | .1228 | .1080 | .1040 | |
| | 6 | .2276 | .1652 | .1520 | .1440 | .1224 | .1102 | |
| | 9 | .2352 | .2156 | .1824 | .1572 | .1268 | .1290 | .1110 |
| | 12 | | .2260 | .1884 | .1844 | .1328 | .1522 | |
| 10.0 | 3 | .1216 | .0804 | .0744 | .0716 | .0604 | .0572 | |
| | 6 | .1288 | .1120 | .0864 | .0728 | .0644 | .0600 | |
| | 9 | .1540 | .1420 | .1104 | .1008 | .0728 | .0630 | .0571 |
| | 12 | | .1752 | .1352 | .1100 | .0748 | .0770 | |
| 12.0 | 3 | .0868 | .0556 | .0560 | .0460 | .0440 | .0414 | |
| | 6 | .0884 | .0916 | .0664 | .0580 | .0464 | .0540 | |
| | 9 | .1368 | .1196 | .0844 | .0656 | .0584 | .0540 | .0418 |
| | 12 | | .1580 | .0968 | .0858 | .0516 | .0632 | |
| 20.0 | 3 | .0460 | .0212 | .0192 | .0188 | .0156 | .0120 | |
| | 6 | .0528 | .0428 | .0268 | .0188 | .0156 | .0124 | |
| | 9 | .0768 | .0596 | .0308 | .0300 | .0176 | .0140 | .0126 |
| | 12 | | .0784 | .0332 | .0420 | .0196 | .0171 | |
| 25.0 | 3 | .0272 | .0140 | .0132 | .0104 | .0064 | .0040 | |
| | 6 | .0336 | .0248 | .0140 | .0112 | .0072 | .0054 | |
| | 9 | .0588 | .0436 | .0232 | .0152 | .0076 | .0060 | .0062 |
| | 12 | | .0508 | .0244 | .0256 | .0100 | .0060 | |

Source: T.L. Boullion, P.L. Odell and B.S. Duran (1975).

## 2.5 Asymptotic Distribution of Error Rates

Efron (1975) studied the efficiency of the logistic regression discrimination relative to maximum likelihood discrimination for the case of two multivariate normal distributions with a common covariance matrix. He developed asymptotic methods for both procedures based upon the expected deviation of the error rate using an estimated discriminant function from the error rate of the optimal function. He used the ratio of the expected deviations from the optimal error rate for the two procedures as the measure of the relative efficiency and concluded that the use of logistic regression discrimination could result in a substantial loss in efficiency.

Following Efron's work, O'Neill (1975) extended the case to non-normal distributions. He developed the asymptotic distribution of the error rates of classification rules for a general exponential family with either linear or quadratic exponents in variable **X**. He showed that the logistic regression estimates performed better than the maximum likelihood estimates over the range considered. This result was attributed to the sensitivity of the maximum likelihood estimates to the departure of the assumption of underlying distribution.

Recently, O'Neill (1980) followed Layard's (1972) Theorem (1.3) and generalized the asymptotic distribution of the error rates of an estimated classification rule to the

cases of arbitrary distributions. The general distribution of the error rates is described below:

Let $D(X,\theta)$ be the discriminant function with X, a k X 1 vector of variables, and $\theta$, a (k+1) X 1 vector of parameters. Let $ERR(\theta)$ denote the error rate of the classification rule with vector of known parameters $\theta$, and $R=\{X: D(X,\theta) = 0\}$ denote the discriminant boundary. Further, let $\nabla_{\underline{x}}$ be the vector partial differential operator with respect to X, and $|\nabla_{\underline{x}}|^2$ be the sum of squared elements in vector $\nabla_{\underline{x}}$. Now, suppose that t is an estimator of $\theta$ such that

$$(2.45)\quad t_n \xrightarrow{L} N_{k+1}(0, V/n)$$

or, equivalently,

$$(2.46)\quad \sqrt{n}\,(t_n - \theta) \xrightarrow{L} N_{k+1}(0, V).$$

Then

$$(2.47)\quad n\,\{ERR(t_n) - ERR(\theta)\} \xrightarrow{L} z'Bz,$$

where z is distributed as $N_{k+1}(0, V)$, and

$$(2.48)\quad B = \frac{1}{4} \int_R |\nabla_x D(X,\theta)|^{-1} \nabla_\theta D(X,\theta)\, \nabla'_\theta D(X,\theta)\, f(X)\, dm_R,$$

where $m_R$ is a Lebesque measure on R and

$$(2.49)\quad f(X) = p\underline{1}\, f(X|1) + p\underline{2}\, f(X|2).$$

The asymptotic error rate (AERR) of t is defined as:

$$(2.50)\quad AERR(t) = E(z'Bz).$$

If the moments converge correctly then this is the limit of the expected value of $n\{ERR(t)-ERR(\theta)\}$. Then

$$(2.51)\quad AERR(t) = \lim_{n\to\infty} E\{n\,(ERR(t)_n - ERR(\theta))\}$$

$$= E(z'Bz) = tr\{E(z'Bz)\}$$

$$= tr\{B\ E(zz')\} = tr(BV).$$

Thus, in order to obtain the asymptotic error rate, the scalar B and matrix V must be calculated.

O'Neill then compared logistic regression to maximum likelihood discrimination for some arbitrary distributions using

$$(2.52)\quad ARE = \frac{AERR(t_M)}{AERR(t_L)} = \frac{tr(B_M V_M)}{tr(B_L V_L)}$$

as the measure of asymptotic relative efficiency (ARE). Finally, he showed that the efficiency of logistic regression classification in some non-normal cases is low and suggested that maximum likelihood discrimination should always be used whenever possible.

## 2.6 Comparison of Linear and Quadratic Discriminant Rules

Gilbert (1969) investigated the effect of unegual covariance matrices on Fisher's linear discriminant function (LDF). The behavior of this LDF was compared with the optimal quadratic form assuming that the parameters of the two populations were known and one covariance matrix was a multiple (d) of the other. The calculations were carried out for d= 0.1, 0.2, 0.5, 1, 2, 5, 10, $\Delta^2$= 0,1,2,4,8 k= 1,2,6,10, and a priori probabilities p1= 1/2, 2/3, 5/6. The results showed that the QDF is theoretically no worse than the LDF and the reduction in optimal error rate is an increasing function of d (d≥ 1), an increasing function of k, and a decreasing dunction of $\Delta^2$. Her results are presented in Tables 6 and 7.

TABLE 6

The Optimal Error Rate Obtained by Using Fisher's Linear Discriminant Function with Equal A Priori Probabilities

| | | For all values of k | | | | |
|---|---|---|---|---|---|---|
| d | $\Delta^2$= | 0 | 1 | 2 | 4 | 8 |
| .1 | | .50 | .23 | .18 | .11 | .05 |
| .2 | | .50 | .26 | .20 | .13 | .06 |
| .5 | | .50 | .30 | .23 | .15 | .08 |
| 1.0 | | .50 | .31 | .24 | .16 | .08 |
| 2.0 | | .50 | .30 | .23 | .15 | .08 |
| 5.0 | | .50 | .26 | .20 | .13 | .06 |
| 10.0 | | .50 | .23 | .18 | .11 | .05 |

Source: Ethel S. Gilbert (1969).

TABLE 7

The Reduction in Optimal Error Rate Obtained by Using the QDF Rather Than Fisher's LDF

| | k = 1 | | | | | k = 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| d $\Delta^2$= | 0 | 1 | 2 | 4 | 8 | 0 | 1 | 2 | 4 | 8 |
| .1 | .25 | .03 | .01 | .02 | .01 | .35 | .10 | .07 | .04 | .03 |
| .2 | .18 | .01 | .01 | .01 | .01 | .26 | .07 | .05 | .03 | .02 |
| .5 | .08 | .00 | .00 | .00 | .01 | .12 | .07 | .01 | .01 | .01 |
| 1.0 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| 2.0 | .08 | .00 | .00 | .00 | .01 | .12 | .07 | .01 | .00 | .01 |
| 5.0 | .18 | .01 | .01 | .01 | .01 | .26 | .07 | .05 | .03 | .02 |
| 10.0 | .25 | .03 | .01 | .02 | .01 | .35 | .10 | .07 | .04 | .03 |
| | k = 6 | | | | | k = 10 | | | | |
| d $\Delta^2$= | 0 | 1 | 2 | 4 | 8 | 0 | 1 | 2 | 4 | 8 |
| .1 | .47 | .20 | .16 | .09 | .04 | .49 | .22 | .18 | .11 | .05 |
| .2 | .41 | .18 | .13 | .09 | .04 | .46 | .23 | .17 | .11 | .05 |
| .5 | .22 | .07 | .05 | .03 | .02 | .26 | .12 | .08 | .05 | .03 |
| 1.0 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| 2.0 | .22 | .07 | .05 | .03 | .02 | .26 | .12 | .08 | .05 | .03 |
| 5.0 | .41 | .18 | .13 | .09 | .04 | .46 | .23 | .17 | .11 | .05 |
| 10.0 | .47 | .20 | .16 | .09 | .04 | .49 | .22 | .18 | .11 | .05 |

Source: Ethel S. Gilbert (1969).

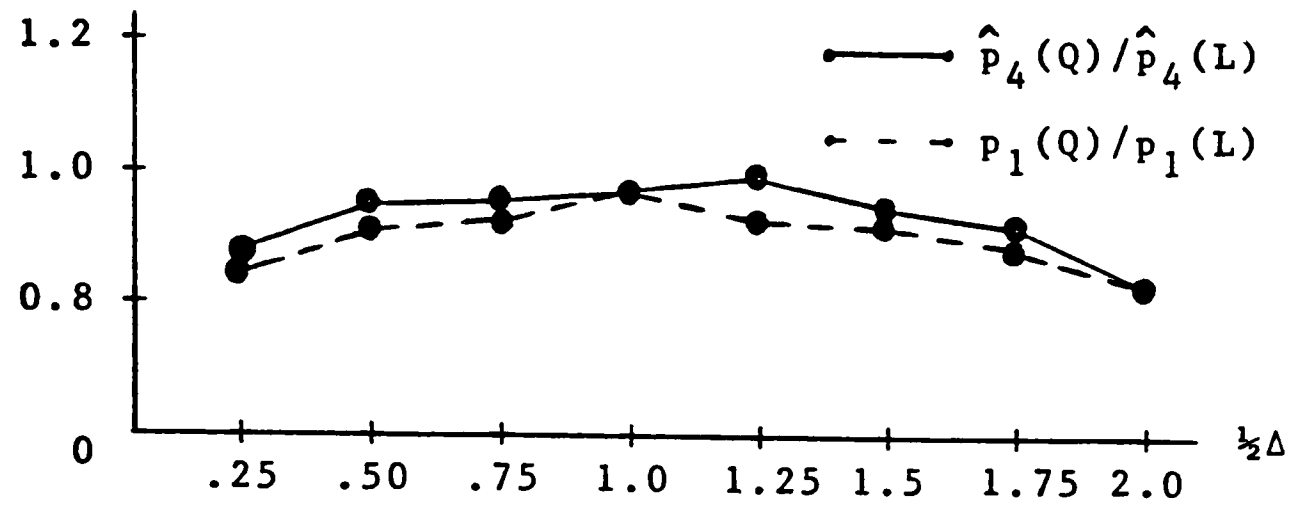
Marks and Dunn (1974) conducted a study of the efficiency of the QDF relative to the LDF with small samples. They used Monte Carlo simulation with parameters k= 2,4,6,8,10, d= 1,2,4,8,16,32,64, $n\underline{1}=n\underline{2}=25$, replications r=100, $p\underline{1}$= 0.25, 0.5, 0.75, $\Delta/2$= 0.5, 1, 1.5, 2, $u\underline{1}$= (0,0,...,0), $u\underline{2}$= ($\delta\underline{1}$,$\delta\underline{2}$,...,$\delta\underline{k}$), $V\underline{1}$=I, and $V\underline{2}$=D where D was a k X k diagonal matrix. Figures 4, 5, and 6 show the ratio of the error rate of the QDF to the LDF in terms of the expected actual error rate $P\underline{4}$ and the optimal error rate $P\underline{1}$ assuming that $p\underline{1}=p\underline{2}=0.5$, $u\underline{2}=(\Delta(1+\sqrt{d})/2,0,...,0)$,

and $V\underline{2}=dV\underline{1}$. It was shown that the LDF is superior to the QDF when k=2, $\Delta$=3.5, d≤ 4; or k=10, $\Delta$> 1, d=2; or k> 10, $\Delta$=3.5, d=4.

Wahl and Kronmal (1977) continued the work of Marks and Dunn (1974) and studied the effect of training sample size on the relative efficiency of the QDF to the LDF. They used k= 1,2,3,4,6,10,12, $0.25 \leq \Delta/2 \leq 2$ with an increment of 0.25, $n\underline{1}= n\underline{2}= 100$, $5\leq r\leq 25$, and d= 2,5,10 as the values of parameters. Their results assuming $u\underline{1}= (0,\ldots,0)$, $u\underline{2}=(\Delta\sqrt{(1+d)/2},0,\ldots,0)$, $V\underline{1}= I$, and $V\underline{2}= dI$ are presented in Figures 9 and 10.

It was shown that the QDF was superior to the LDF in all cases studied. They attributed this outcome to the increase in training sample size from 25 to 100. Thus, they concluded that when the dimension and the covariance differences are small, it makes little difference whether Fisher's linear or the quadratic function is used. However, when the dimension and the covariance differences are large, the QDF performs much better than the LDF provided the sample size is sufficiently large.

(a) k=2, d=2, p1=p2=0.5, n1=n2=25




(b) k=10, d=2, p1=p2=0.5, n1=n2=25

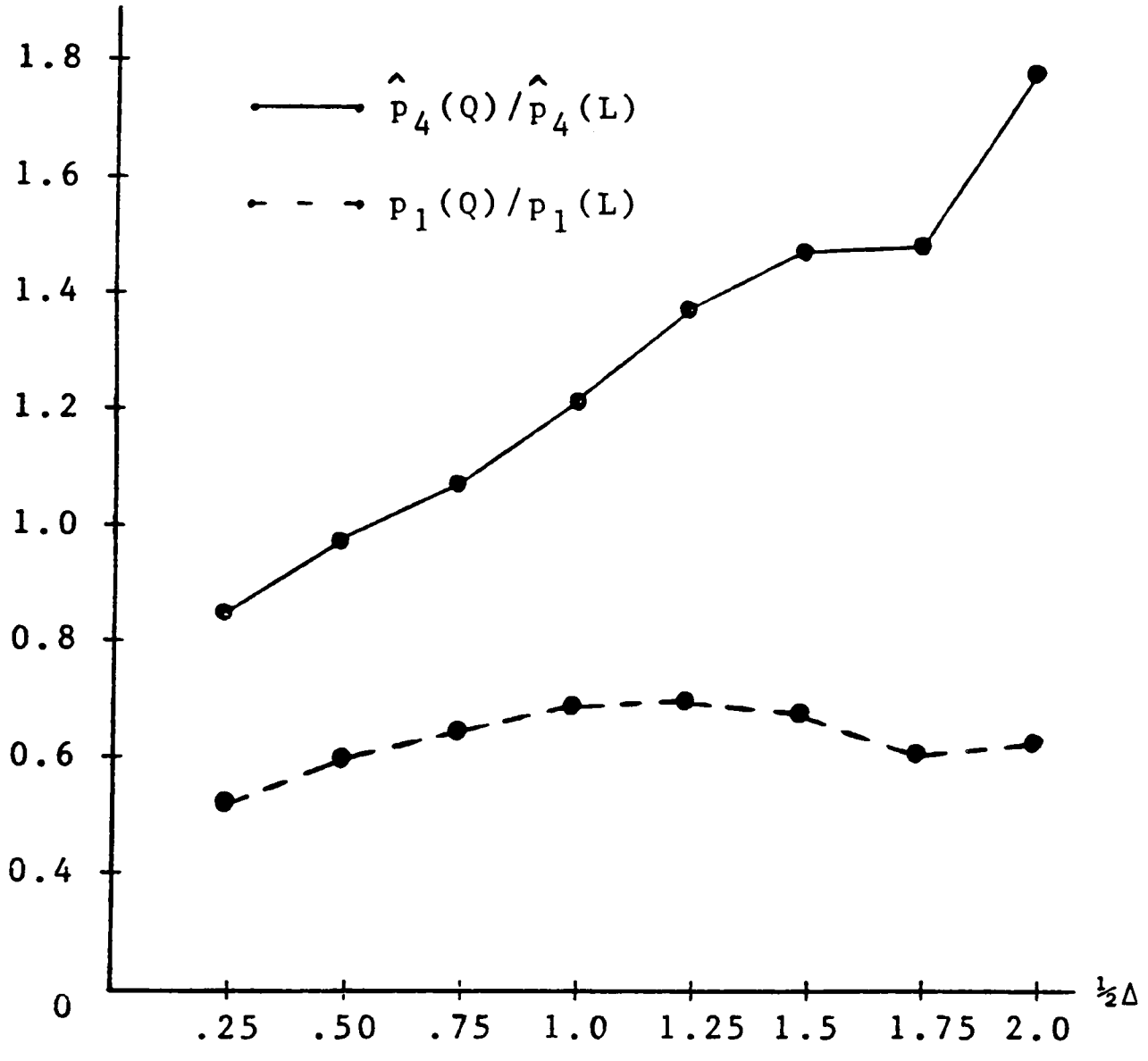
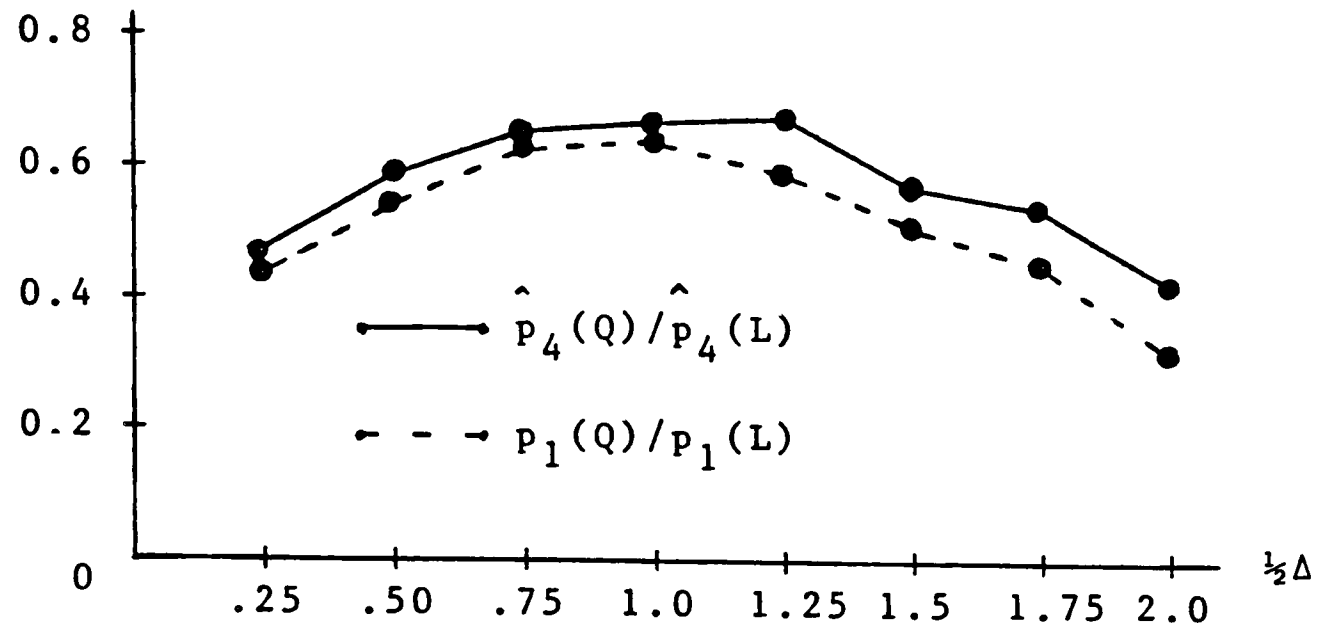



Figure 3: Ratios of Quadratic to Fisher's Error Rates with Different Group Distance (Small Sample Size)

(c) k=2, d=8, p1=p2=0.5, n1=n2=25




(d) k=10, d=8, p1=p2=0.5, n1=n2=25

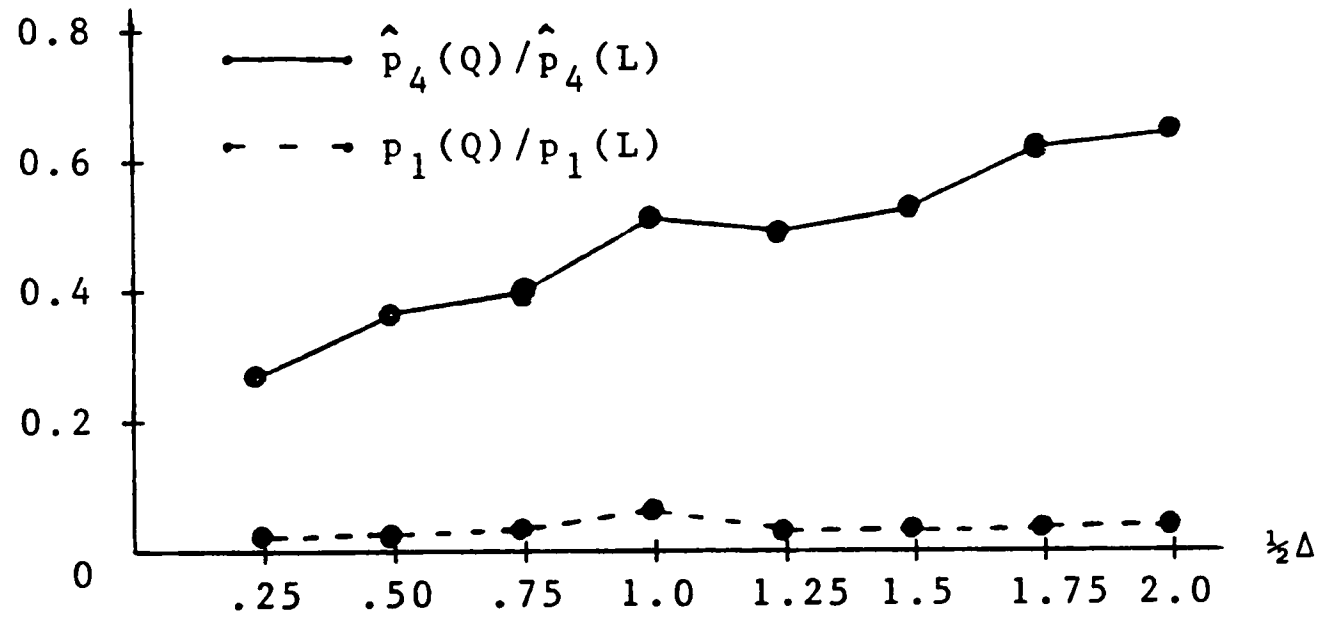
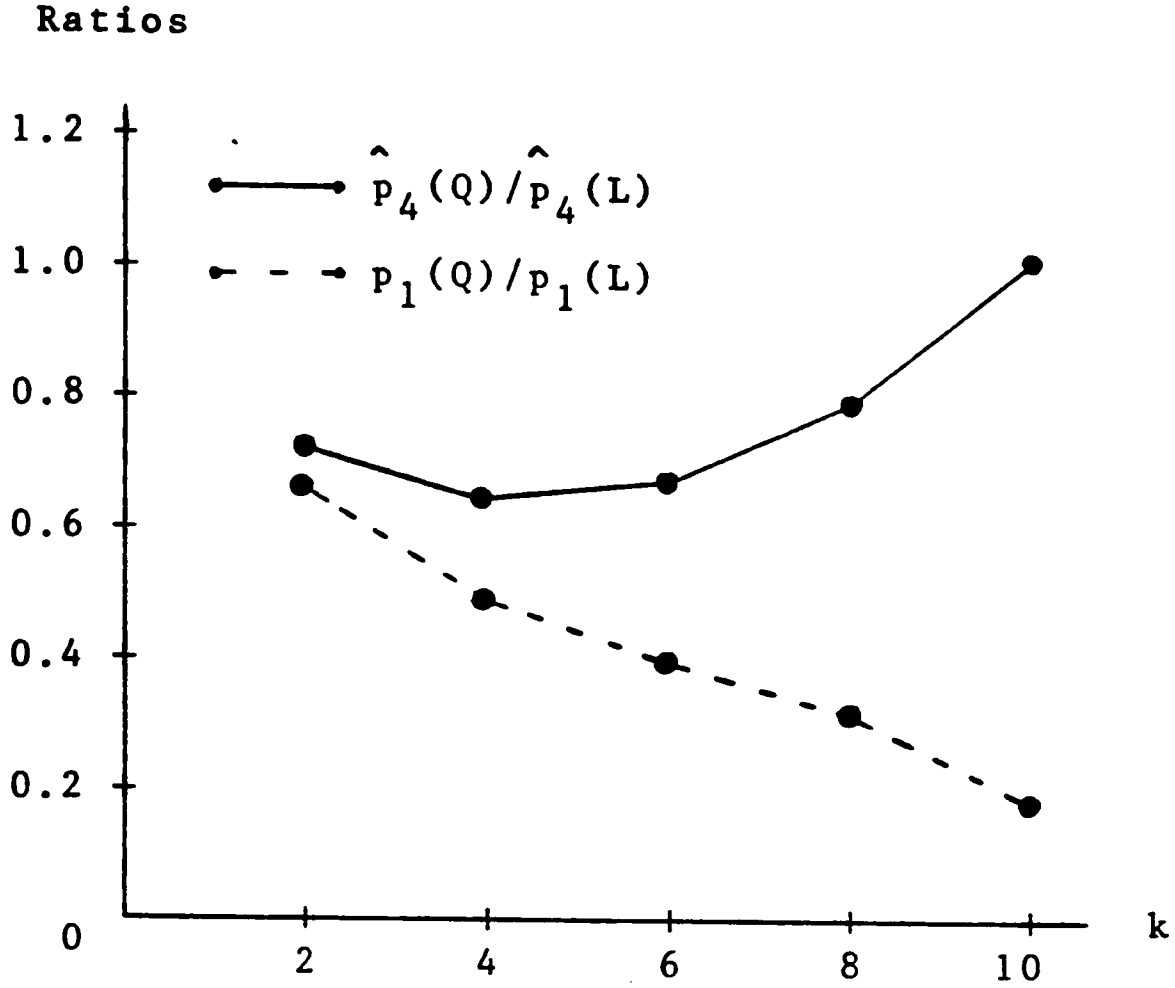



Figure 3: (Continued)

(a) $d=4$, $\Delta=3.5$, $p\underline{1}=p\underline{2}=0.5$, $n\underline{1}=n\underline{2}=25$




(b) $d=9$, $\Delta=3.5$, $p\underline{1}=p\underline{2}=0.5$, $n\underline{1}=n\underline{2}=25$

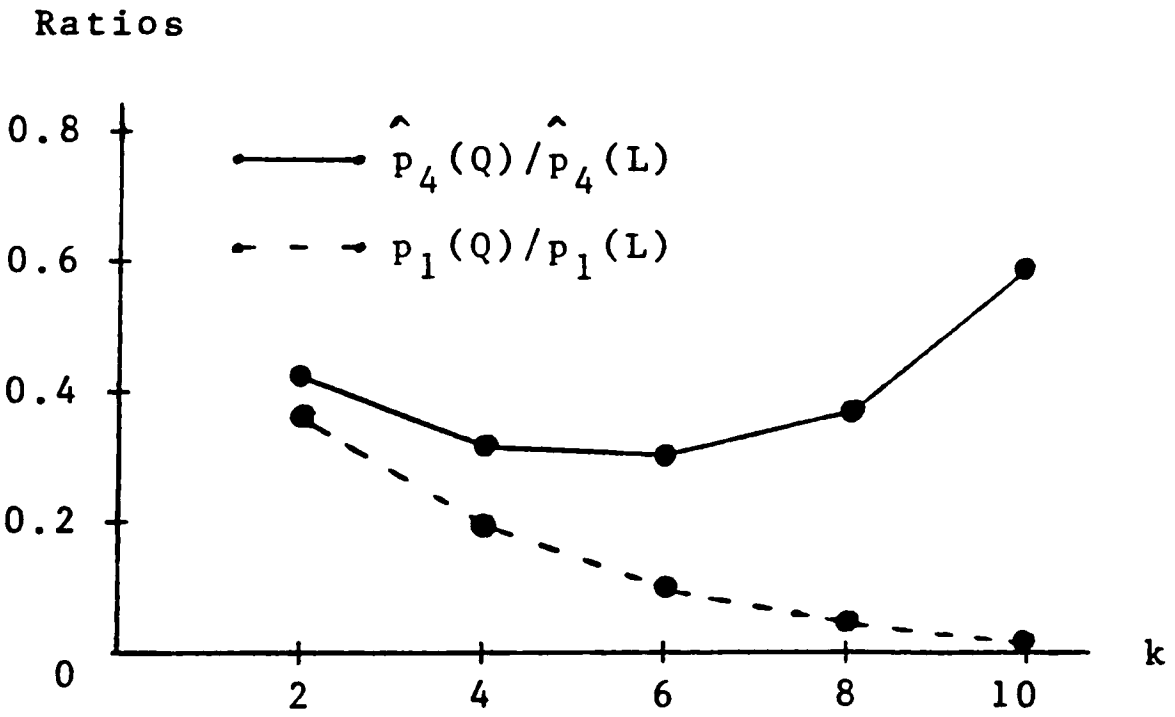
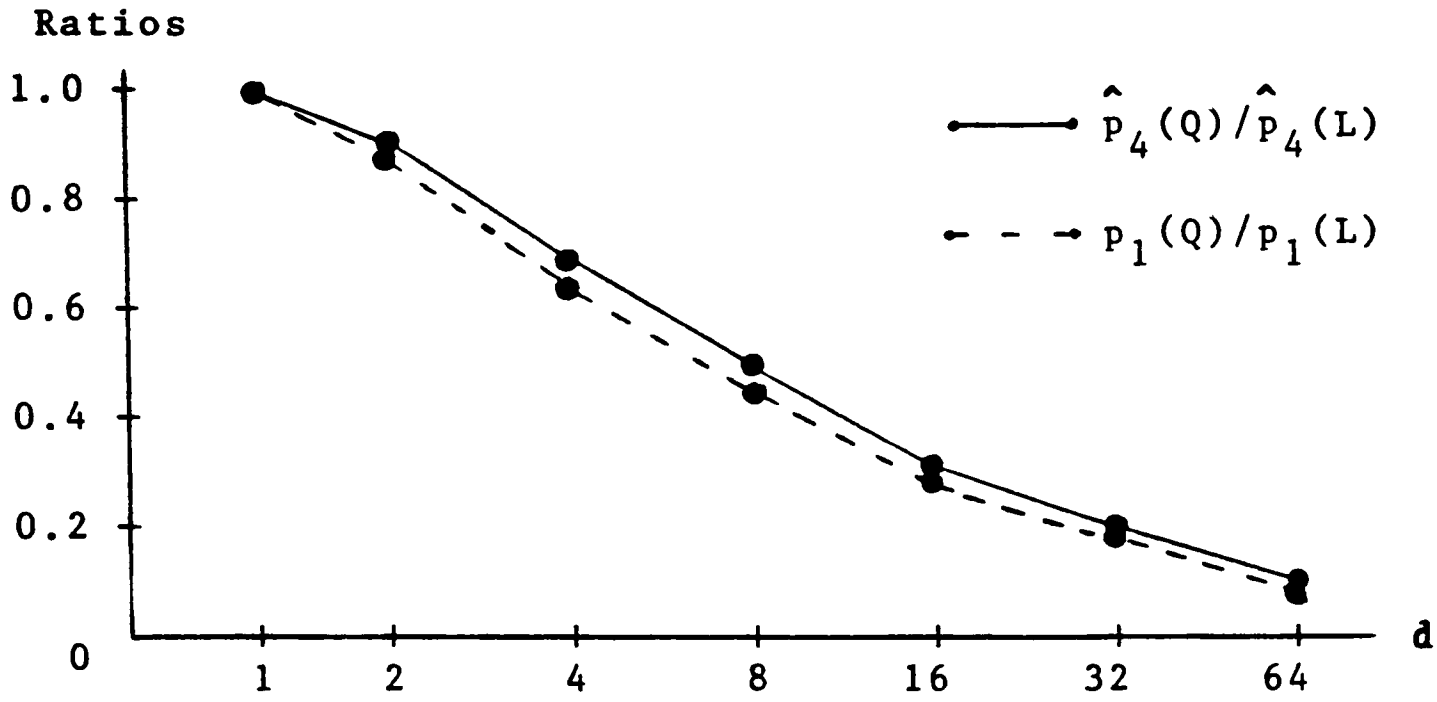



Figure 4: Ratios of Quadratic to Fisher's Error Rates with Different Number of Variables (Small Sample Size)

(a) $k=2,\ \Delta=3.5,\ p1=p2=0.5,\ n1=n2=25$




(b) $k=10,\ \Delta=3.5,\ p1=p2=0.5,\ n1=n2=25$

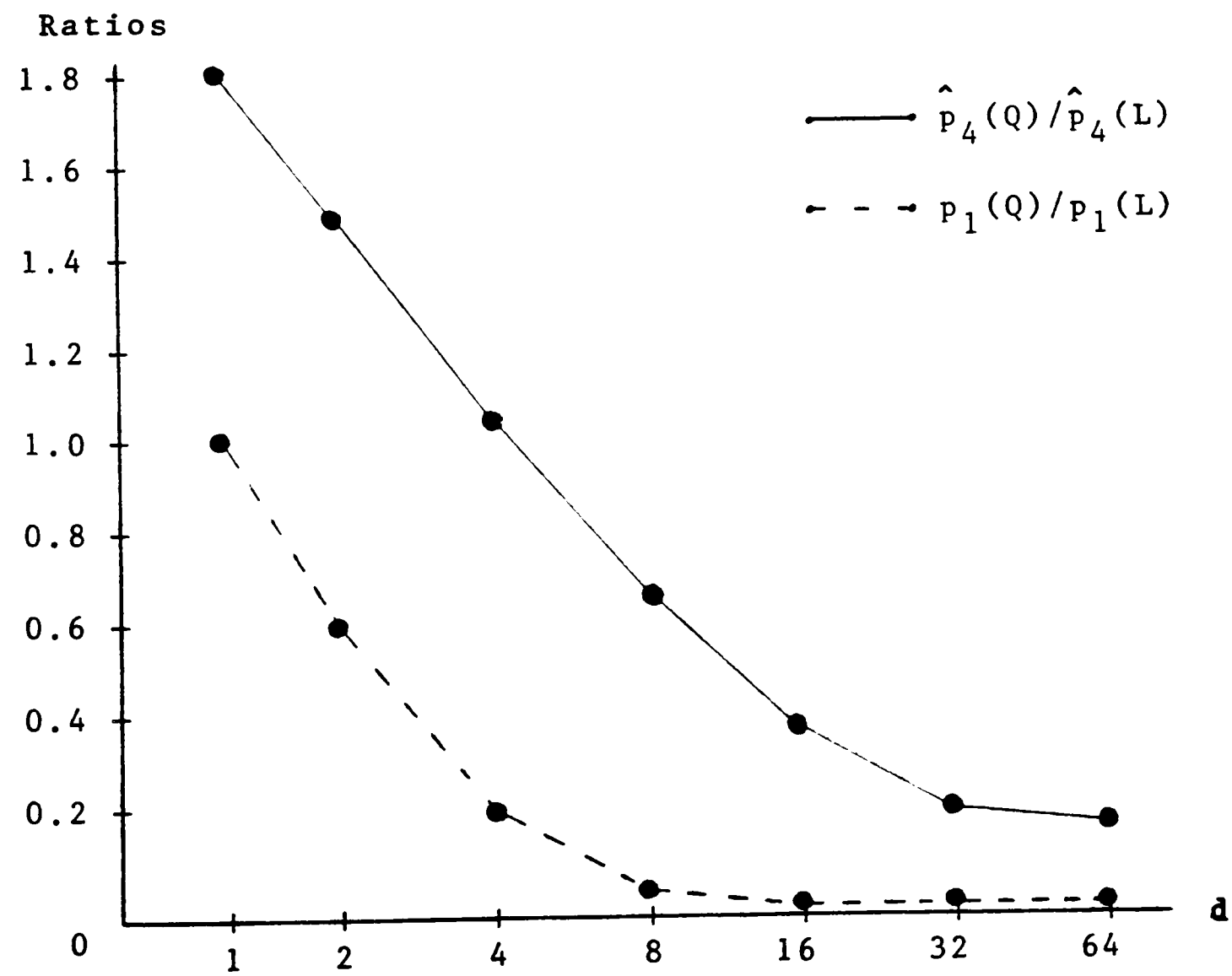
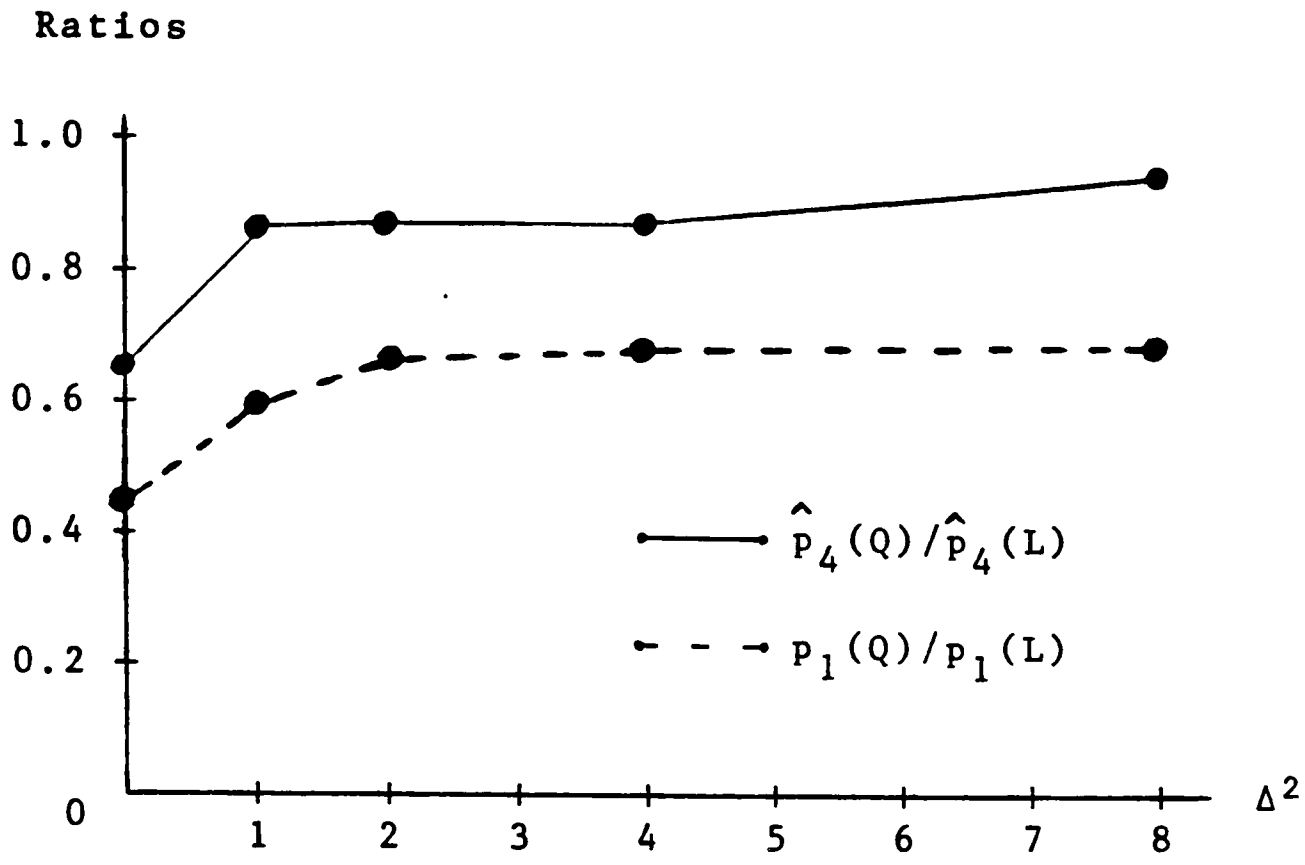



Figure 5: Ratios of Quadratic to Fisher's Error Rates with Different Proportionality (Small Sample Size)

(a) k=10, d=2, p1=p2=0.5, n1=n2=100




(b) k=10, d=10, p1=p2=0.5, n1=n2=100

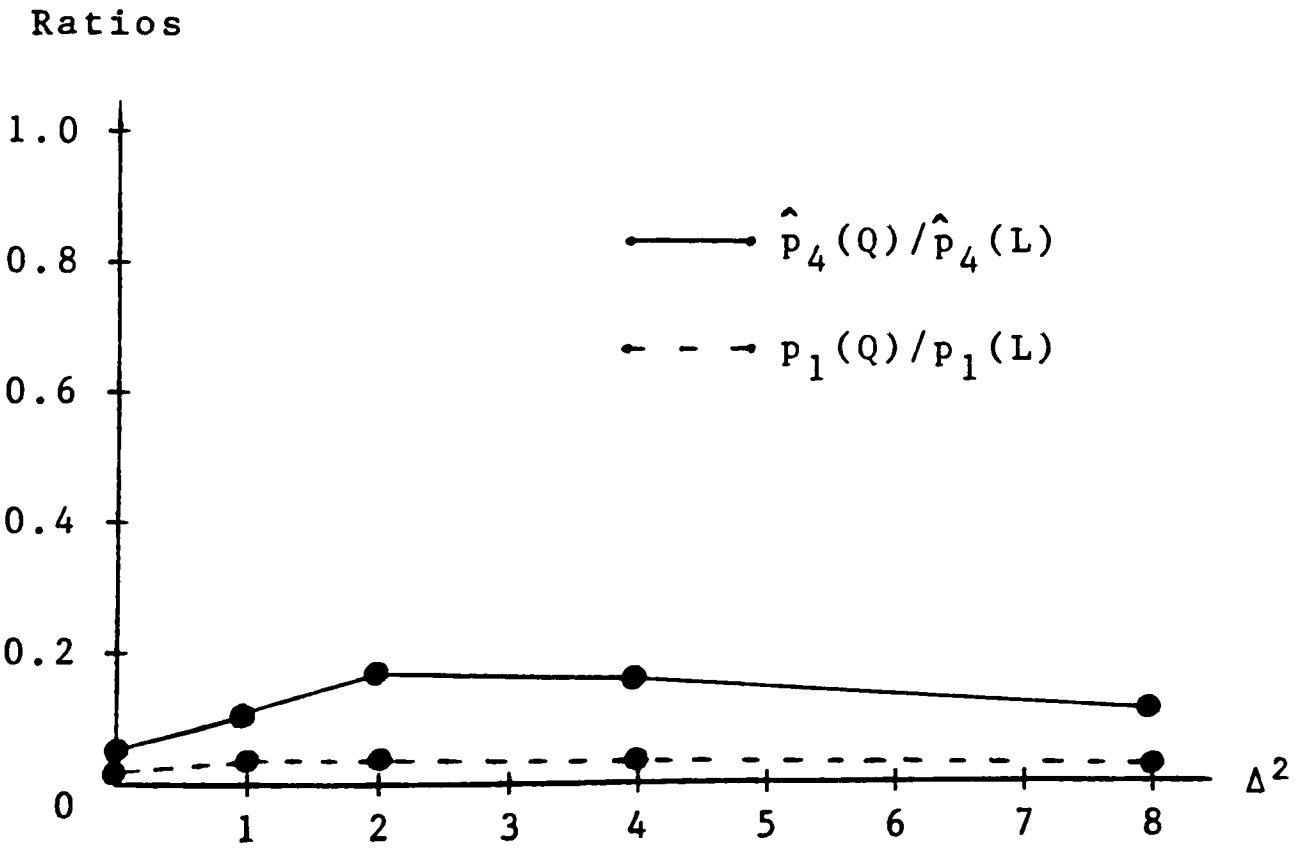
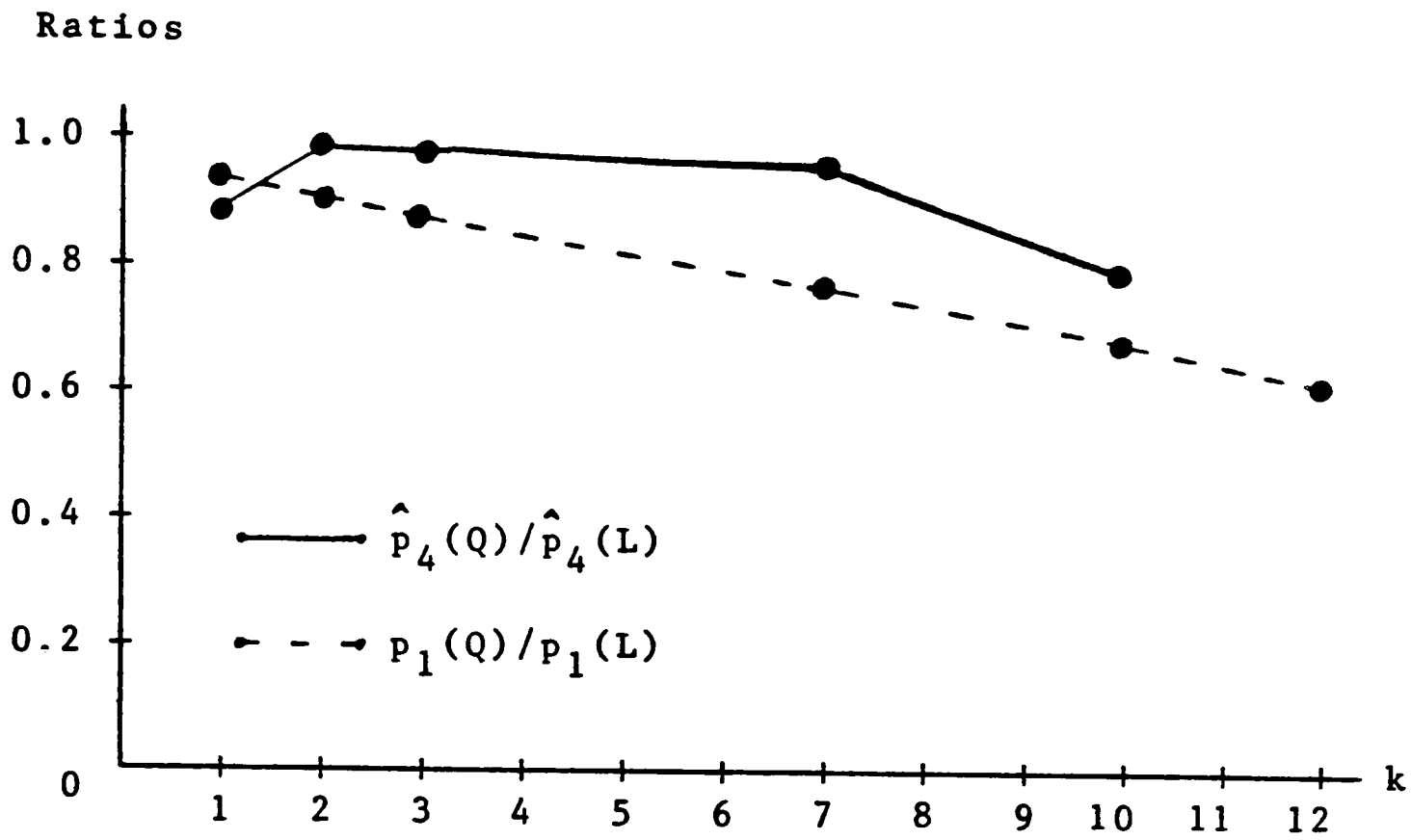



Figure 6: Ratios of Quadratic to Fisher's Error Rates with Different Group Distance (Large Sample Size)

(a) d=2, $\Delta^2=8$, $p1=p2=0.5$, $n1=n2=100$




(b) d=5, $\Delta^2=8$, $p1=p2=0.5$, $n1=n2=100$

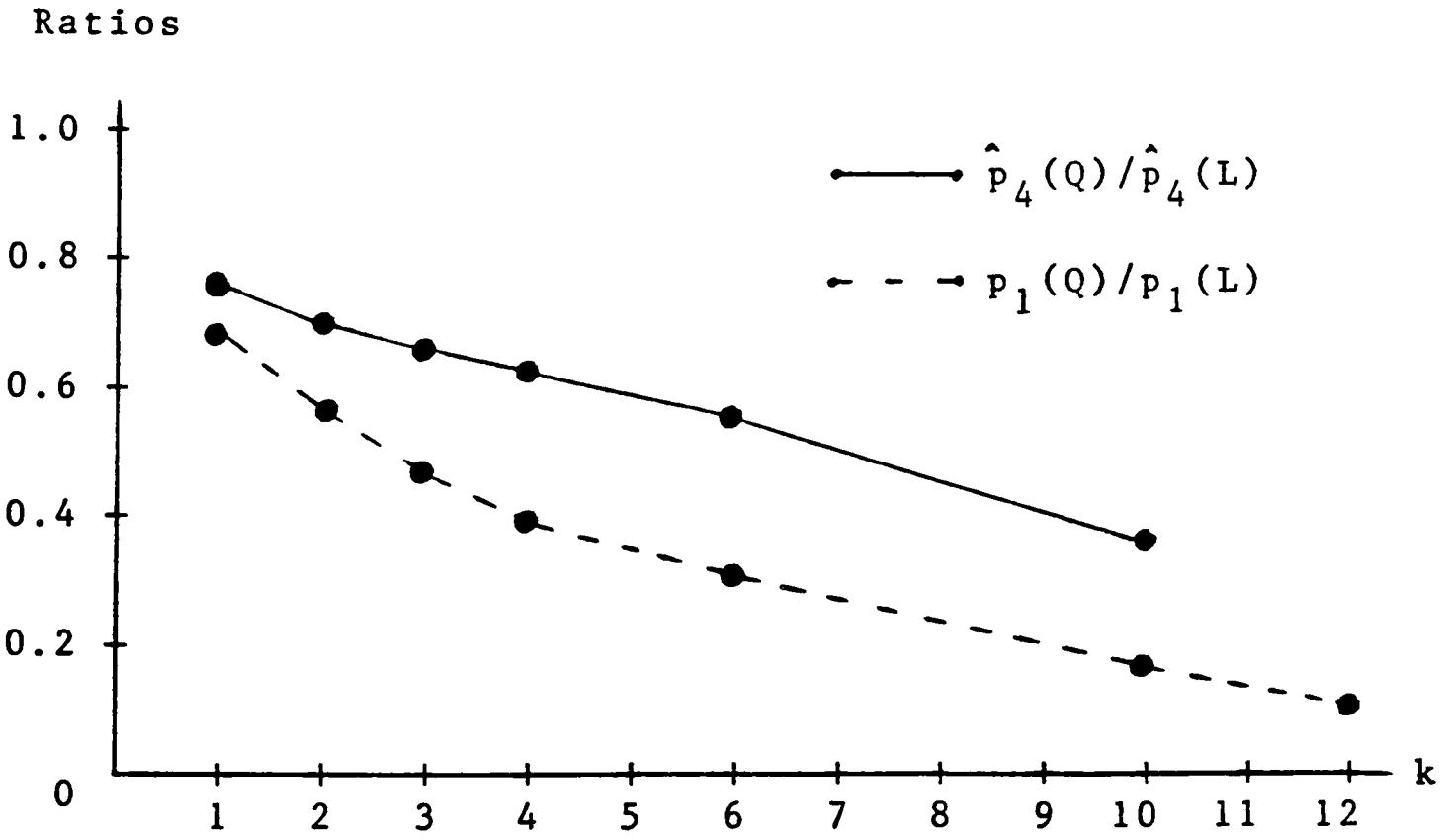



Figure 7: Ratios of Quadratic to Fisher's Error Rates with Different Number of Variables (Large Sample Size)

## 2.7 Discriminant Analysis in Business Literature

The adoption of discriminant analysis (DA) in business research has been rapid because investigators frequently are concerned with the nature and strength of the relationship between group membership (for example, different brand choices, or different market segments) and individual characteristics (such as demographic measures, or style measures). The widespread availability of easy-to-use discriminant analysis computer programs has greatly faciliated implementation of these DA techniques (Dixon 1970, 1975, Nie, et al. 1975). From the vast literature on business applications of discriminant analysis, the marketing and finance literature will be reviewed.

Applications of the DA techniques in marketing area can be found in Wallace and Travers (1938), Banks (1950), Harvey (1953), Evans (1959), King (1963), Claycamp (1965), Massy (1965), Pessemier, Burger and Tigert (1967), Brody and Cunningham (1968), Robertson and Kennedy (1968), Churchill, Ford and Ozanne (1970), Sheth (1970), Uhl, Andrus and Poulsen (1970), Wind (1970), Lessig and Tollefson (1971), Simkowitz and Monroe (1971), Frank and Talarzyk (1972), Ostlund (1972, 1974), Hustad and Pessemier (1973), Bao (1973), Darden and Reynolds (1974), Robertson and Rossiter (1974), Montgomery (1975), Shuchman and Riesz (1975), Utterback (1975), Etgar (1976), Riordan, Oliver and Donnelly (1977), Gitlow (1979), Huber and Holbrook (1979), etc.

In finance literature, the DA techniques were employed by Durand (1941), Tintner (1946), Blood and Baker (1958), Walter (1959), Beranek (1963), Myers and Forgy (1963), Mosteller and Wallace (1963), Peterson (1963), Smith (1965), Beaver (1966), Adelman and Morris (1968), Altman (1968, 1973), Carleton and Lerner (1969), Pogue and Soldofsky (1969), Horton (1970), Johnson (1970), Meyer and Pifer (1970), Orgler (1970), Haslem and Longbrake (1971), Simkowitz and Monroe (1971), Williams and Goodman (1971), Dake (1972), Dince and Fortson (1972), Edmister (1972), Eisenbeis and McCall (1972), Lane (1972), Bates (1973), Klemkosky and Petty (1973), Pinches and Mingo (1973, 1975), Awh and Waters (1974), Eisenbeis and Murphy (1974), Gilbert (1974, 1975), Katz (1974), Norgaard and Norgaard (1974), Stuhr and Wicklen (1974), Joy and Tollefson (1975, 1978), Schick and Verbrugge (1975), Sinkey (1975), Zumwalt (1975), Altman and Katz (1976), Eisenbeis (1977), Eisenbeis and Altman (1978), Scott (1978), Wiginton (1980), etc.

Among the extensive articles listed above, five out of those showed particularly effective use of discriminant analysis for the exploration of business phenomena. These five articles are surveyed here.

Evans (1959) considered three different factor sets and used the LDF to predict an automobile buyer's brand choice between Ford and Chevrolet. For the personality factors of the buyer, the variables he used were: achievement,

deference, exhibition, autonomy, affiliation, intraception, dominance, abasement, change, aggression, and heterosexuality. For the demographic and objective factors, the variables were: age of car, used over 10,000 miles per year, shopped before buying, owner smokes, homeowner or renter, three or more children at home, Catholic or not, Protestant or not, attend church more than once a month, Republican or not, Democrat or not, age, five or more years with same firm, and income. The combined psychological and objective factors were: owner smokes, homeownership, three or more children at home, Catholic or not, Protestant or not, Republican or not, Democrat or not, five or more years with same firm, deference, exhibition, autonomy, affiliation, and dominance.

He applied a linear discriminant function in this study and showed that the objective factors, overall, did a somewhat better job of discriminating between Ford and Chevrolet buyers than the other factors. Therefore, he concluded that the objective variables possess discriminatory power, a claim which is refuted by many motivation researchers, and pointed out that brand images may narrow the car market unnecessarily.

King (1963) treated a market expansion decision as a classification problem. The decision under study was to determine if a particular area should be classified as a penetrable or impenetrable market. King employed an

essentially classical approach and used test market data as sample information to estimate a linear discriminant function which minimizes the expected cost of error. The measurements which he deemed relevant were per capita sales of the product, rate of change of per capita personal income over time, market share of major competing secondary product, and total market share of all primary products.

Montgomery (1975) studied the factors or characteristics of a new product which had been accepted or rejected by supermarkets. He interviewed buyers and attendants at buying committee meetings to identify a list of potentially important variables considered in the decision to accept or reject a new product for distribution in a supermarket. This list was then used to develop a structured personal interview. The variables he selected were: promotion, company reputation, quality, newness, introductory allowances, competition, packaging, gross margin, advertising, private label, guarantee, distribution, broker, sales presentation, category volume, category growth, shelf space, and cost. Linear discriminant analysis was employed to discriminate between the groups that accepted and rejected the product. He further suggested that a large scale analysis would prove fruitful to manufacturers in reducing new product failure rates and to supermarkets in understanding the criteria which buyers seem to use in evaluating proposed new products.

Beranek (1963) considered the decision of whether or not credit should be granted to an applicant. The two relevant populations are composed of those who will default on a credit contract and those who will not. Beranek suggested that a linear discriminant function be used to classify credit applicants as acceptable or unacceptable. The lines of business which were identified by Beranek for potential application included major gasoline distributors, large department stores, finance and small loan companies, commercial and mutual savings banks, savings and loan associations, and life insurance companies. The measurements which he proposed as useful include the customer's past credit record, current asset position, leverage, magnitude and variability of gross proceed stream, future cash proceeds, length of time in current business, age of proprietor, whether renting or owning existing premises, length of time since last delinquent, and proportion of bills the customer discounted. Each line of business will, of course, have its unique set of measurements which are germane to the existing classification problem.

Klemkosky and Petty (1973) applied discriminant analysis to a study of the firm-unique factors responsible for affecting stock price variability. A random sample of 160 common stocks was selected and measures of variability (coefficients of variation) were computed for each stock.

Firm-unique investment data were collected on the stocks ranked in the upper and lower quartiles. The variables chosen to derive a linear discriminant function were debt/equity ratio, fixed charge coverge, percentage change in price of last 42 months, percentage change in price of last 6 months, earnings per share volatility, earnings per share growth, average price-earnings ratio, average dividend yield, average price per share, average shares outstanding, and annual turnover.

This discriminant function was then adjusted by the standard deviation of the associated variables so as to determine the relative contributions of each variable to the discriminatory power. He found that the two variables best able to differentiate between high and low volatility stocks were annual turnover and average price. It was concluded that the model has discriminatory power and is able to differentiate between high and low volatility stocks.

## 2.8 Contribution of This Research to the Literature

A review of the related literature reveals that several areas of the literature are severely lacking, especially in the behavior of the PQDF. In the research presented here, the coefficient of proportionality is estimated, the variance of the estimated coefficient of proportionality is derived, and the performance of the PQDF is examined. The comparisons of the PQDF to the LDF and the QDF are performed

using relative efficiency as the criterion. Guidelines for selecting the regions in the parameter space where the PQDF performs at least as well as the LDF and the QDF are provided.

Consequently, the major contribution of this research is the thorough investigation of the properties of the PQDF. Although Han (1974) studied the distribution of the PQDF with partially known covariance matrices, none of his studies treated both covariance matrices as random variables. No literature was found that studied the performance of the PQDF and compared it to those of the LDF and the QDF. Several authors have studied the relative performances between the LDF and the QDF. These include Gilbert (1969), Marks and Dunn (1974), Boullion, Odell, and Duran (1975), Van Ness and Simpson (1976), Wahl and Kronmal (1977), Huberty and Curry (1978), and Van Ness (1979). None of these authors, however, compared the performance of the PQDF to those of the LDF and QDF, nor did they examine the behavior of the PQDF.

A second contribution to the literature is the identification of the regions in the parameter space where the PQDF outperforms the LDF and QDF. A review of the related literature shows a void of this subject. The researchers noted above have provided some guidelines for practitioners in selecting the dominant rule between the LDF and the QDF according to the parameters of their own data

set. However, none of them has given any information about the domains in which the PQDF outperforms the others. This study will fill this void by delimiting the regions dominated by the PQDF and by providing some measures of the efficiency of the PQDF relative to the LDF and the QDF.

# Chapter III

# METHODOLOGY AND EXPERIMENTAL DESIGN

## 3.1 Experimental Design

The design of this research can be divided into two parts: the analytic method and the simulation study. Analytic method is used to investigate the properties of the PQDF and to compare, asymptotically, its performance to those of the LDF and QDF. Simulation is used to investigate and compare the three discriminant functions when the sample size is finite.

The problems studied were confined to two-group normal discrimination in which the costs of misclassification were equal (i.e., $C(1|2)=C(2|1)=1$). Modifications for unequal misclassification costs are trivial and the extension to multiple-group discrimination is straightforward. Further, the *a priori* probabilities were assumed to be equal in the simulation study to reduce the size of the experimental design.

### 3.1.1 Analytic Method

When the population parameters are known, the comparisons of the LDF, QDF, and PQDF can be made

analytically without much effort. Unfortunately, these parameters are usually unknown in practice and, therefore, must be estimated. It is necessary, then, to examine both the procedures for estimation and the properties of the estimator.

The estimation of the coefficient of proportionality was done using the maximum likelihood procedure under the assumption of unknown but proportional covariance matrices. The variance of this estimator was obtained from the inverse of Fisher's information matrix (Press 1972, Rao 1973, Kendall and Stuart 1979).

The asymptotic error rates of the estimated LDF, QDF, and PQDF were calculated using a procedure derived by Efron (1975) and O'Neill (1980). The performance of the estimated PQDF relative to the estimated LDF and QDF was compared using asymptotic relative efficiency (ARE), defined by Equation (2.52) in Section 2.5, as the criterion. The computed ARE's of PQDF to the LDF and the QDF were then tabulated and plotted against the population parameters. Plotted results are exhibited by Figures 8 through 21 in Appendix E.

The ARE is the limit of the ratio of the deviations of the actual error rate obtained by using each of two estimated discriminant functions from the optimal error rate. Thus if the optimal error rate is $P_1(D_t)$ and $P_4(L)$ and $P_4(P)$ are the expected actual error rates of the LDF and PQDF, the ARE of the PQDF relative to the LDF is given by

$$ARE(P/L) = \lim_{n\to\infty} \frac{E\{P_4(L) - P_1(D_t)\}}{E\{P_4(P) - P_1(D_t)\}} .$$

Alternatively, the ARE can be interpreted as a measure of the ratio of sample sizes required for each of the two procedures being compared to obtain the same expected error rate.

It is well-known in two-group normal discrimination that a random sample X can be assumed to be distributed as $N(u_1, V_1)$ in population 1 and $N(u_2, V_2)$ in population 2. If $u_1 \neq u_2$ and $V_1 \neq V_2$, this sample X can always be transformed linearly, without loss of generality, into a sample with $u_1' = (\delta_1, \delta_2, \ldots, \delta_k)$, $u_2' = (0, 0, \ldots, 0)$, $V_1 = I$, and $V_2 = D$, where $\delta_i$ $(i=1,2,\ldots,k)$ is an arbitrary constant, I is a k X k identity matrix and D a k X k diagonal matrix with diagonal elements $d' = (d_1, d_2, \ldots, d_k)$ where $d_i$ is a positive constant. The data which have been linearly transformed are invariant in terms of the probabilities of misclassification. The latter condition will be hereafter referred to as the "quadratic covariance structure" (QCS). When all the diagonal elements in d are equal, i.e., $d' = (d, d, \ldots, d)$, the mean vector $u_1$ can be further reduced to $(\Delta\sqrt{(1+d)/2}, 0, \ldots, 0)$ and the covariance matrix $V_2$ can be expressed as $V_2 = dI$. This condition is called the "proportional covariance structure" (PCS) here. Further, if $d=1$, then it is called the "linear covariance structure" (LCS).

Due to the conditions necessary for Theorem (1.3) of Layard (1972) and thus Theorem (1) of O'Neill (1980) to hold, studies cannot be conducted analytically using the QCS. To make comparisons, it is necessary that both estimated discriminant functions converge to that of the optimal discriminant function. Thus, in order to compare asymptotically the LDF and PQDF, the LCS must hold. However, the comparison of the PQDF and QDF may be made when either the LCS or PCS holds.

For simplicity, tractability, and understandability, only one-variate and two-variate cases were examined analytically in this study.

### 3.1.2 Simulation Study

In each experiment of the simulation study, a small sample was first generated. The size of the sample was determined by the training sample size n.

From each small training sample, the maximum likelihood estimates of $u_1$, $u_2$, $V_1$, and $V_2$ were computed, and thus the estimated LDF, QDF, and PQDF can be obtained along with the optimal DF. Based on each sample, two generalized likelihood ratio tests ($H_0$: $V_2=V_1$ against $H_1$: $V_2=dV_1$ and $H_0$: $V_2=dV_1$ against $H_1$: $V_2\neq dV_1$) were conducted in order to obtain the critical level, also called the "p-value," for both tests. These p-values can be a helpful clue for exploring the relationship between the covariance structures

and the performances of the discriminant functions. It should be apparent that the p-values are influenced by n, the training sample size, just as the estimates of the covariance matrices are affected by their corresponding sample sizes $n_i$ (i=1,2).

Subsequently, the coefficients of the optimal DF were obtained using the known covariance structure and those of the estimated DF were computed using the training sample. The optimal DF and the estimated LDF, PQDF, and QDF were applied individually to classify the small training sample. With the estimated DF, the percentage of misclassifications from classifying the small training sample is called the "apparent error rate." Moreover, the optimal DF and the estimated LDF, PQDF, and QDF were then used to classify a second very large sample, called "evaluation sample." The percentage of misclassifications from classifying the evaluation sample is called the "estimated actual error rate." With the optimal DF, the percentage of misclassifications from classifying the aggregate sample is called the "estimated optimal error rate." Experimental replication was performed to obtain the average of each error rate. These averages were then regarded as the "estimated expected error rates" and used to compare the performances of the LDF, QDF, and PQDF and to develop a procedure for the selection of the best discriminant function estimated from the training sample.

Finally, the estimated expected error rates were tabulated against the number of variables k, the Mahalanobis's squared distance between the two mean vectors $\Delta^2$, the training sample size n ($n_1=n_2=n$), and the covariance structures. The estimated relative efficiencies of the PQDF to the LDF and the QDF were used to compare the performance of the PQDF relative to the LDF and QDF. The regions dominated by the PQDF were found to be ambiguous. Therefore, guidelines rather than rules have been provided for the selection of a best DF.

## 3.2 Estimation of the Coefficient of Proportionality

Let the k-component random sample of size N ($N=n_1+n_2$) be distributed as $N(u_1,V)$ in population 1 and $N(u_2,dV)$ in population 2. The likelihood function of this sample is given by

$$(3.1)\quad L(u_1,u_2,V,d) = (2\pi)^{-(n_1+n_2)k/2} |V|^{-(n_1+n_2)/2} d^{-n_2k/2} \cdot e^{-1/2 \{\sum_{i=1}^{n_1} (X_i- u_1)'V^{-1}(X_i- u_1)\}} \cdot e^{-1/2 \{1/d \sum_{j=1}^{n_2} (X_j-u_2)'V^{-1}(X_j-u_2)\}} .$$

After taking the logarithm, this function can be transformed into the log likelihood function (log L) shown below:

$$(3.2)\quad \log L(X|u_1,u_2,V,d) = -k(n_1+n_2)/2 \log(2\pi) - (n_1+n_2)/2 \log|V| - kn_2/2 \log d - 1/2 \{ \sum_{i=1}^{n_1} (X_i-u_1)'V^{-1}(X_i-u_1) + 1/d \sum_{j=1}^{n_2} (X_j-u_2)'V^{-1}(X_j-u_2) \}.$$

It is well-known that the log L maximized with respect to $u_1$ and $u_2$ is

$$(3.3)\quad \log L(X|\bar{X}_1,\bar{X}_2,V,d) = -k(n_1+n_2)/2 \log(2\pi) - (n_1+n_2)/2 \log|V| - kn_2/2 \log d - 1/2 \{ \sum_{i=1}^{n_1} (X_i-\bar{X}_1)'V^{-1}(X_i-\bar{X}_1) + 1/d \sum_{j=1}^{n_2} (X_j-\bar{X}_2)'V^{-1}(X_j-\bar{X}_2) \}$$

$$= -k(n_1+n_2)/2 \log(2\pi) - (n_1+n_2)/2 \log|V| - kn_2/2 \log d - 1/2 \{ \operatorname{tr}(V^{-1}A_1) + 1/d \operatorname{tr}(V^{-1}A_2) \}$$

$$= -k(n_1+n_2)/2 \log(2\pi) - (n_1+n_2)/2 \log|V| - kn_2/2 \log d - 1/2 \operatorname{tr}[V^{-1}(A_1 + 1/d\ A_2)],$$

where $\bar{X}_1$, $\bar{X}_2$, $A_1$, and $A_2$ were defined in Equations (1.10)

and (2.13).

Differentiating log L with respect to V and d, and solving for the maximum likelihood estimates (MLE) generates the equations:

$$(3.4)\quad \hat{V} = S = (n\underline{1}+n\underline{2})^{-1}\{A\underline{1} + (1/\hat{d})\ A\underline{2}\},$$

$$(3.5)\quad kn\underline{2}\hat{d} = tr(\hat{V}^{-1}A\underline{2}) = tr(S^{-1}A\underline{2}).$$

Substituting (3.4) into (3.5) yields the equation:

$$(3.6)\quad kn\underline{2}/(n\underline{1}+n\underline{2}) = tr\{(\hat{d}\ A\underline{1}\ A\underline{2}^{-1} + I)^{-1}\}.$$

While an explicit solution for $\hat{d}$ is not possible, numerical evaluation using the Newton-Raphson procedure is straightforward.

The recurrence relation for the estimation of $\hat{d}$ is

$$(3.7)\quad \hat{d}_{i+1} = \hat{d}_i + \{\sum_{j=1}^{k} \delta_j(\hat{d}_i) - kn_2/(n_1+n_2)\} / \{\sum_{j=1}^{k} \lambda_j\ \delta_j^2(\hat{d}_i)\},$$

where $\lambda_j$, $j=1,2,..,k$ is the $j$th eigenvalue of $A\underline{1}A\underline{2}^{-1}$, $\delta_j(\hat{d}) = (\hat{d}\lambda_j + 1)^{-1}$ and $\hat{d}_i$ is the $i$th iterative estimate of d. A convenient initial estimate is provided by $\hat{d}_1 = (|A\underline{2}|/|A\underline{1}|)^{1/k}$. The MLE of V is then found immediately from Equation (3.4), once a satisfactory solution to Equation (3.6) is obtained.

The variance of $\hat{d}$ can be obtained using the inverse of

Fisher's information matrix, $I\underline{d}$. When all the other population parameters are known, it is given that

$$(3.8)\quad \mathrm{Var}(\hat{d}) = I\underline{d}^{-1},$$

where

$$(3.9)\quad I\underline{d} = E\left[\frac{\partial \log L}{\partial d}\right]^2 = -E\left[\frac{\partial^2 \log L}{\partial d^2}\right] = -\left[\frac{\partial^2 \log L}{\partial d^2}\right]_{\hat{\theta}=\theta}$$

(Rao 1973, Kendall and Stuart 1979).

After differentiating log L with respect to d, one has

$$(3.10)\quad \frac{\partial \log L}{\partial d} = -\frac{kn\underline{2}}{2d} + \frac{1}{2d^2}\,\mathrm{tr}(V^{-1}A\underline{2}),$$

$$(3.11)\quad \frac{\partial^2 \log L}{\partial d^2} = \frac{kn\underline{2}}{2d^2} - \frac{2}{2d^3}\,\mathrm{tr}(V^{-1}A\underline{2}),$$

$$(3.12)\quad \left[\frac{\partial^2 \log L}{\partial d^2}\right]_{\hat{\theta}=\theta} = \frac{kn\underline{2}}{2d^2} - \frac{1}{d^3}\,\mathrm{tr}(V^{-1}n\underline{2}dV),$$

$$= kn\underline{2}/2d^2 - 1/d^3\,\mathrm{tr}(n\underline{2}dI)$$

$$= kn\underline{2}/2d^2 - kn\underline{2}/d^2$$

$$= -\,kn\underline{2}/2d^2.$$

Thus, the variance of $\hat{d}$ is $\mathrm{Var}(\hat{d}) = 2d^2/kn\underline{2}$. Moreover, if the other parameters are unknown, the last expression in (3.12) is one of the elements of Fisher's information matrix. The procedure of deriving two-variate $\mathrm{Var}(\hat{d})$ is

presented in Appendix B.

## 3.3 Asymptotic Comparisons of the three Discriminant Functions

In this section, the asymptotic relative efficiency of the PQDF to the LDF and the QDF will be discussed. Derivation for the covariance matrix of the coefficients of the estimated PQDF with the PCS is presented in Appendix B. For the variance matrices of the coefficients of an estimated DF with other covariance structures, derivations are straightforward if one follows the procedure shown in Appendix B. Detailed calculations for the asymptotic error rates (AERR) of the estimated LDF, PQDF, and QDF with two-variate LCS or PCS are given in Appendix C.

### 3.3.1 One-variate Asymptotic Relative Efficiency

#### 3.3.1.1 With Linear Covariance Structure

The asymptotic error rates (AERR) of the LDF, PQDF, and the QDF with k=1 and a linear covariance structure (LCS) were found to be respectively as follows:

$$(3.13)\quad AERR(L) = (1/2p_2\Delta)\,\Phi(\Delta/2 + \lambda/\Delta) \cdot \{1 + \Delta^2/4 + \lambda^2/\Delta^2\,(1+2p_1p_2\Delta^2) + \lambda(2p_2-1)\},$$

$$(3.14)\quad AERR(P) = (1/2p_2\Delta)\,\Phi(\Delta/2 + \lambda/\Delta) \cdot \{\,3/2 + \lambda^2/\Delta^2\,(\,\lambda^2/2\Delta^2 + 2p_2\lambda - \lambda\,) + \Delta^2/4\,(\Delta^2/8 + 2p_2\lambda - \lambda\,) + 3\lambda^2/4\,\},$$

and AERR(Q)= AERR(P) where $\Phi$ is the cumulative standard normal distribution and $\lambda = \log(p\underline{1}/p\underline{2})$.

The asymptotic relative efficiency (ARE) of the PQDF to the LDF, ARE(P/L), is defined below:

(3.15) $$ARE(P/L) = AERR(L)/AERR(P)$$

$$= \{1 + \Delta^2/4 + \lambda^2/\Delta^2\ (1 + 2p\underline{1}p\underline{2}\Delta^2) + \lambda(2p\underline{2} - 1)\}$$
$$/\ \{\ 3/2 + \lambda^2/\Delta^2\ (\ \lambda^2/2\Delta^2 + 2p\underline{2}\lambda - \lambda\ )$$
$$+ \Delta^2/4\ (\Delta^2/8 + 2p\underline{2}\lambda - \lambda\ ) + 3\lambda^2/4\ \},$$

and ARE(Q/P)= AERR(P)/AERR(Q)= 1.

The greater ARE(P/L), the better the PQDF performs relative to the LDF, and <u>vice versa</u>. However, since the linear covariance structure (LCS) is assumed here, the ARE(P/L) will be less than one. This means that the PQDF performs worse than the LDF. The resulting plot is shown by Figure 8 in Appendix E.

### 3.3.1.2 With Proportional Covariance Structure

The necessary conditions of O'Neill's (1980) Theorem (1) preclude the derivation of AERR(L) with proportional covariance structure (PCS). The AERR(P) and the AERR(Q) are equal when k=1. The AERR(P) and the AERR(Q) are given below:

(3.16) $$AERR(P) = AERR(Q)$$

$$= (2p\underline{2}\sqrt{C})^{-1}\{E\underline{1}\underline{1}\{3/2 + p\underline{2}\Delta^4(1+d)^2/8\}$$
$$-E\underline{1}\underline{2}\{p\underline{2}\Delta^3(1+d)\sqrt{(1+d)/2}\}$$

$$+E\underline{2}\underline{2}\{3p\underline{2}\Delta^2(1+d)/2\}$$

$$-E\underline{2}\underline{3}\{2p\underline{2}\Delta\sqrt{(1+d)/2}\}$$

$$+E\underline{3}\underline{3}\{(p\underline{2}+p\underline{1}/d^2)/2\},$$

where

$$(3.17)\quad C = \Delta^2(1+d)/2d + (2\lambda + \log d)(1 - 1/d)$$

and $E\underline{i}\underline{j}$ is the $ij\underline{th}$ element of the 3 X 3 matrix E, where

$$(3.18)\quad E = \begin{bmatrix} E\underline{1}\underline{1} & E\underline{1}\underline{2} & E\underline{1}\underline{3} \\ E\underline{2}\underline{1} & E\underline{2}\underline{2} & E\underline{2}\underline{3} \\ E\underline{3}\underline{1} & E\underline{3}\underline{2} & E\underline{3}\underline{3} \end{bmatrix}$$

$$= vv'\,\Phi\{\{\Delta\sqrt{(1+d)/2} - d\sqrt{C}\}/(1-d)\}$$

$$+ ww'\,\Phi\{\{\Delta\sqrt{(1+d)/2} + d\sqrt{C}\}/(1-d)\}$$

and where

$$(3.19)\quad v = \begin{bmatrix} 1 \\ (\Delta\sqrt{(1+d)/2} - \sqrt{C})/(1 - 1/d) \\ \{\Delta^2(1+d)/2 + C - 2\Delta\sqrt{C(1+d)/2}\}/(1 - 1/d)^2 \end{bmatrix}$$

and

$$(3.20)\quad w = \begin{bmatrix} 1 \\ (\Delta\sqrt{(1+d)/2} + \sqrt{C})/(1 - 1/d) \\ \{\Delta^2(1+d)/2 + C + 2\Delta\sqrt{C(1+d)/2}\}/(1 - 1/d)^2 \end{bmatrix}.$$

Of course the ARE(Q/P) is equal to one since the two discriminant functions are equivalent.

### 3.3.2 Two-variate Asymptotic Relative Efficiency

#### 3.3.2.1 With Linear Covariance Structure

The AERR's of the LDF, PQDF, and QDF with LCS were found respectively to be as follows:

$$(3.21)\quad AERR(L) = (1/2p\underline{2}\Delta)\,\Phi(\Delta/2 + \lambda/\Delta)$$
$$\bullet\ \{2 + \Delta^2/4(1 + 4p\underline{1}p\underline{2}) + \lambda^2/\Delta^2\,(1 + 2p\underline{1}p\underline{2}\Delta^2) + \lambda(2p\underline{2} - 1)\},$$

$$(3.22)\quad AERR(P) = (1/2p\underline{2}\Delta)\,\Phi(\Delta/2 + \lambda/\Delta)$$
$$\bullet\ \{11/4 + \lambda^2/2\Delta^2\,(1 + \lambda^2/2\Delta^2 + 2p\underline{2}\lambda - \lambda) + \Delta^2/8\,(1 + \Delta^2/8 + 2p\underline{2}\lambda - \lambda) + p\underline{1}p\underline{2}(\Delta^2 + \lambda^2) + \lambda(3\lambda/8 + p\underline{2} - 1/2)\},$$

and

$$(3.23)\quad AERR(Q) = (1/2p\underline{2}\Delta)\,\Phi(\Delta/2 + \lambda/\Delta)$$
$$\bullet\ \{7/2 + \lambda^2/\Delta^2\,(1 + \lambda^2/2\Delta^2 + 2p\underline{2}\lambda - \lambda) + \Delta^2/4\,(1 + \Delta^2/8 + 2p\underline{2}\lambda - \lambda) + \lambda(3\lambda/4 + 2p\underline{2} - 1)\}.$$

Thus

$$(3.24)\quad ARE(P/L) = \{2 + \Delta^2/4(1 + 4p\underline{1}p\underline{2}) + \lambda^2/\Delta^2\,(1 + 2p\underline{1}p\underline{2}\Delta^2) + \lambda(2p\underline{2} - 1)\}$$
$$/\ \{11/4 + \lambda^2/2\Delta^2\,(1 + \lambda^2/2\Delta^2 + 2p\underline{2}\lambda - \lambda) + \Delta^2/8\,(1 + \Delta^2/8 + 2p\underline{2}\lambda - \lambda)$$

$$+p_1p_2(\Delta^2+\lambda^2) + \lambda(3\lambda/8 + p_2 - 1/2)\}$$

and

$$\begin{aligned}(3.25)\quad ARE(Q/P) = \{&11/4 + \lambda^2/2\Delta^2\,(1+\lambda^2/2\Delta^2 + 2p_2\lambda - \lambda\,) \\ &+\Delta^2/8\,(1+\Delta^2/8 + 2p_2\lambda - \lambda\,) \\ &+p_1p_2(\Delta^2+\lambda^2) + \lambda(3\lambda/8 + p_2 - 1/2)\} \\ /\,\{&7/2 + \lambda^2/\Delta^2\,(1+\lambda^2/2\Delta^2 + 2p_2\lambda - \lambda\,) \\ &+\Delta^2/4\,(1+\Delta^2/8 + 2p_2\lambda - \lambda\,) \\ &+\lambda(3\lambda/4 + 2p_2 - 1)\}.\end{aligned}$$

Three resulting plots are shown by Figures 9, 10, and 11 in Appendix E. It is seen that the plot shape of the two-variate ARE(P/L) with LCS is similar to that of the one-variate ARE(P/L) with LCS.

### 3.3.2.2 With Proportional Covariance Structure

The AERR's of the PQDF and QDF with PCS are given below.

$$\begin{aligned}(3.26)\quad AERR(P) = 1/2\,p_2\sqrt{C}\;\cdot \\ \{\,&E_{11}\{32 + (1+d)^2\Delta^4 p_2(1+p_1)\}/16 \\ &-E_{12}\{\Delta^3(1+d)\,p_2(1+p_1)\sqrt{(1+d)/2}\}/2 \\ &+E_{15}\{\Delta^2(1+d)\,p_2(3+3p_1-p_1/d)/4\} \\ &+E_{16}\{\Delta^2(1+d)\,p_2(1+p_1+p_1/d)/4\} \\ &-E_{25}\{\Delta\sqrt{(1+d)/2}\;p_2(1+p_1-p_1/d)\} \\ &-E_{34}\{2p_1(1-1/d)+(p_2+p_1/d)\}p_2\Delta\sqrt{(1+d)/2} \\ &+(2E_{44}+E_{55}+E_{66})\{p_2+p_1/d^2+p_1p_2(1-1/d)^2\}/4\}\end{aligned}$$

and

$$(3.27)\quad AERR(Q) = 1/2\,p\underline{2}\sqrt{C}\ \cdot$$

$$\{\ E\underline{11}\{2 + p\underline{2}\Delta^4(1+d)^2/8\}$$
$$- E\underline{12}\{p\underline{2}\Delta^3(1+d)\sqrt{(1+d)/2}\}$$
$$+ (3E\underline{15}+E\underline{16})\{p\underline{2}\Delta^2(1+d)/2\}$$
$$- (E\underline{25}+E\underline{34})\{p\underline{2}\Delta\sqrt{2(1+d)}\}$$
$$+ (2E\underline{44}+E\underline{55}+E\underline{66})(p\underline{2}+ p\underline{1}/d^2)/2\},$$

where

$$(3.28)\quad C = \Delta^2(1+d)/2d + 2(\log d+\lambda)(1- 1/d)$$

and $E\underline{ij}$ is the $ij\underline{th}$ element of the 6 X 6 matrix E, where

$$(3.29)\quad E = \begin{bmatrix} E\underline{11} & E\underline{12} & \cdot & \cdot & \cdot & E\underline{16} \\ E\underline{21} & E\underline{22} & \cdot & \cdot & \cdot & E\underline{26} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ E\underline{61} & E\underline{62} & \cdot & \cdot & \cdot & E\underline{66} \end{bmatrix}$$

$$= \int_{X_2} ww'\,\Phi\left(\frac{\Delta\sqrt{(1+d)/2} - d\sqrt{C\underline{x}}}{1-d}\right)\Phi(X_2)\,dX_2$$

$$+ \int_{X_2} ww'\,\Phi\left(\frac{\Delta\sqrt{(1+d)/2} + d\sqrt{C\underline{x}}}{1-d}\right)\Phi(X_2)\,dX_2$$

and where

$$(3.30)\quad C\underline{x} = C - (1- 1/d)^2 X\underline{2}^2,$$

$$(3.31)\quad \mathbf{v} = \begin{bmatrix} 1 \\ (\Delta\sqrt{(1+d)/2} - \sqrt{C\underline{x}})/(1-1/d) \\ X\underline{2} \\ (\Delta\sqrt{(1+d)/2} - \sqrt{C\underline{x}})\,X\underline{2}/(1-1/d) \\ \{\Delta^2(1+d)/2 + C\underline{x} - 2\Delta\sqrt{C\underline{x}(1+d)/2}\}/(1-1/d)^2 \\ X\underline{2}^2 \end{bmatrix}$$

and

$$(3.32)\quad \mathbf{w} = \begin{bmatrix} 1 \\ (\Delta\sqrt{(1+d)/2} + \sqrt{C\underline{x}})/(1-1/d) \\ X\underline{2} \\ (\Delta\sqrt{(1+d)/2} + \sqrt{C\underline{x}})\,X\underline{2}/(1-1/d) \\ \{\Delta^2(1+d)/2 + C\underline{x} + 2\Delta\sqrt{C\underline{x}(1+d)/2}\}/(1-1/d)^2 \\ X\underline{2}^2 \end{bmatrix}.$$

It is required for a real solution that $C\underline{x} \geq 0$, thus the integral limits are restricted by $X\underline{2}^2 \leq C(1-1/d)^{-2}$. If $C\underline{x} < 0$, the AERR is set to one. The exact solution for the E matrix is intractable, thus the Simpson's approximation rule and Newton's 3/8 rule were employed to evaluate E numerically using 21 computed data points. Notice that both AERR(P) and AERR(Q) require only 9 elements of the E matrix to be evaluated.

The ARE(Q/P) with PCS in this case is equal to AERR(P)/AERR(Q). Since the PQDF is the optimal discriminant function with the PCS, the ARE(Q/P) will be less than one.

Two selected tables of ARE's with the PCS and the LCS are shown in Tables 8 and 9. Various selected plots are shown by Figures 12 through 21 in Appendix E. Notice that in order to prevent blocking of the shape on these plots, the ARE in the region where $C\underline{x} < 0$ was set to zero while it is actually one. The reason that the ARE goes to one is attributed to the fact that one of their probability distributions is dominated by the other and the optimal rule is to classify every observation into one of the populations. Under this circumstance, one DF is as good as the other.

### 3.3.3 Interpretation and Conclusions

By reviewing the resulting tables and plots, the following conclusions can be made:

(1) With **linear covariance structure:**

(i) The ARE plots are all symmetric to $p\underline{1}=p\underline{2}=0.5$ along the axis of $p\underline{1}$ (see Figures 8 through 11 in Appendix E).

(ii) Both ARE(P/L) and ARE(Q/P) are decreasing functions of $\Delta$, if $\Delta \geq 4$.

(iii) ARE(Q/P) is an increasing function of $\min(p\underline{1}, p\underline{2})$.

(iv) The regions of the largest ARE(P/L) and ARE(Q/P) are all delimited by $2 \leq \Delta \leq 4$. These regions are of particular interest since differences in

performances of the three estimated DF's are small enough that selection of the best DF is difficult and requires scrutiny.

(2) With **proportional covariance structure:**

(i) Group distance and a priori probability tend to have more influence on ARE(Q/P) than the proportionality, d. This can be verified by the differences in the plot shapes (see Figures 12 through 21 in Appendix E).

(ii) When $p1=p2=0.5$ and $\Delta \geq 4$, the ARE(Q/P) is a decreasing function of $\Delta$ (see Figure 13 in Appendix E). This finding is in agreement with the conclusions drawn from the case using the LCS.

(iii) It can be seen from Figures 12, 13, and 14 in Appendix E that the region of the largest ARE(Q/P) is delimited by $2 \leq \Delta \leq 4$. This result suggests that further investigation on the estimated DF should be made within this region using finite samples.

(iv) The ARE(Q/P) is a decreasing function of d if $p1=p2=0.5$ and $2 \leq \Delta \leq 4$ (see Figure 13 in Appendix E). This suggests that only a small region, say $d \leq 4$, needs to be considered in the study using a finite sample.

TABLE 8

Asymptotic Relative Efficiency of the QDF to the PQDF with the PCS

| d | Δ | Prior Probability One 0.10 | 0.25 | 0.50 | 0.75 | 0.90 |
|---|---|---|---|---|---|---|
| 1.0 | 0.5 | .508225 | .571867 | .798082 | .571867 | .508225 |
| | 1.0 | .615490 | .768164 | .830579 | .768164 | .615490 |
| | 1.5 | .783658 | .854199 | .870199 | .854199 | .783658 |
| | 2.0 | .864807 | .896202 | .900000 | .896202 | .864807 |
| | 2.5 | .890594 | .907403 | .907833 | .907403 | .890594 |
| | 5.0 | .683945 | .725486 | .747599 | .725486 | .683945 |
| | 10.0 | .534912 | .560972 | .576246 | .560972 | .534912 |
| 1.5 | 0.5 | 1.000000 | 1.000000 | .783208 | .667615 | .586926 |
| | 1.0 | 1.000000 | .818808 | .846751 | .794695 | .715274 |
| | 1.5 | .756964 | .862577 | .872136 | .866905 | .828028 |
| | 2.0 | .856969 | .887213 | .896991 | .906821 | .885975 |
| | 2.5 | .878814 | .890189 | .903029 | .918080 | .905021 |
| | 5.0 | .681711 | .735168 | .773870 | .756984 | .705458 |
| 2.0 | 0.5 | 1.000000 | 1.000000 | .746785 | .691487 | .644969 |
| | 1.0 | 1.000000 | .797010 | .826425 | .819069 | .777201 |
| | 1.5 | .716927 | .879513 | .877042 | .883660 | .862177 |
| | 2.0 | .874167 | .884124 | .892792 | .915180 | .903905 |
| | 2.5 | .875029 | .873880 | .894710 | .922453 | .915078 |
| | 5.0 | .658224 | .703109 | .742135 | .735713 | .694524 |
| 3.0 | 0.5 | 1.000000 | .965464 | .752068 | .733099 | .711464 |
| | 1.0 | 1.000000 | .820834 | .772575 | .807581 | .813060 |
| | 1.5 | .849262 | .806168 | .836388 | .886485 | .895730 |
| | 2.0 | .838417 | .850949 | .876968 | .923767 | .928737 |
| | 2.5 | .866772 | .843978 | .878587 | .925375 | .928285 |
| | 5.0 | .639464 | .687172 | .734657 | .735360 | .693451 |
| 4.0 | 0.5 | 1.000000 | .909161 | .757671 | .766194 | .760396 |
| | 1.0 | 1.000000 | .809626 | .763591 | .807846 | .823465 |
| | 1.5 | .892737 | .772309 | .801902 | .869099 | .893592 |
| | 2.0 | .806032 | .796282 | .846597 | .914591 | .933835 |
| | 2.5 | .822442 | .807654 | .860248 | .921924 | .933713 |
| | 5.0 | .626384 | .676182 | .728723 | .733462 | .690789 |
| 6.0 | 0.5 | 1.000000 | .830150 | .763430 | .805125 | .821919 |
| | 1.0 | 1.000000 | .774370 | .762195 | .821132 | .845799 |
| | 1.5 | .874162 | .739432 | .777151 | .853067 | .884228 |
| | 2.0 | .775826 | .738639 | .805005 | .888348 | .919382 |
| | 2.5 | .750054 | .748771 | .824323 | .902961 | .925925 |
| | 5.0 | .609659 | .662050 | .720246 | .729382 | .685514 |
| 15.0 | 0.5 | .879194 | .722075 | .768550 | .854959 | .905137 |
| | 1.0 | .814215 | .704191 | .763461 | .852265 | .897798 |
| | 1.5 | .737613 | .686533 | .760234 | .851502 | .892202 |
| | 2.0 | .680488 | .676076 | .761353 | .852984 | .888311 |
| | 2.5 | .647660 | .672711 | .764590 | .851856 | .878084 |
| | 5.0 | .581987 | .638026 | .704054 | .719353 | .673084 |

TABLE 9

Asymptotic Relative Efficiency of the PQDF to the LDF with the LCS

| | | Prior Probability One | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.10 | | 0.25 | | 0.50 | |
| $\Delta$ | k= | 1 | 2 | 1 | 2 | 1 | 2 |
| 0.1 | | .004174 | .008330 | .016843 | .033387 | .668332 | .728098 |
| 0.3 | | .039848 | .077911 | .168766 | .300108 | .681552 | .734587 |
| 0.5 | | .123691 | .227785 | .509340 | .675200 | .707412 | .746996 |
| 1.0 | | .658484 | .772837 | .998196 | .876013 | .816327 | .796020 |
| 1.5 | | .999564 | .889306 | .951787 | .861450 | .942285 | .850838 |
| 2.0 | | .938577 | .864633 | .986293 | .884725 | 1.000000 | .888889 |
| 2.5 | | .964744 | .880375 | .992060 | .904400 | .941852 | .898476 |
| 5.0 | | .491139 | .715658 | .389739 | .677401 | .344725 | .662384 |
| 10.0 | | .091960 | .219917 | .086050 | .248464 | .082803 | .264631 |

## 3.4 Comparisons through Monte Carlo Simulation

### 3.4.1 Selection of Experimental Variables

The Monte Carlo simulation technique was applied to the finite sample investigation. The values of the parameters for the simulation study were selected using the analytic results and eight pilot runs. The asymptotic results suggest that the region delimited by $2 \leq \Delta \leq 4$ and $d \leq 4$ should be investigated under small sample condition. The pilot runs were made using $\Delta = 2, 4$ and $d = 4$ as the values of the parameters. While the training sample size was 30, four different combinations of evaluation sample size (N=500,1000) and experiment replication (r=200,100) were used. It was found that N=500 and r=200 give the smallest standard error of estimate.

The CPU time of each simulation run with k=2 took approximately eight minutes on the ITEL AS/6 computer. For k=4, it took almost sixteen minutes. Thus, higher dimensions will require more CPU time for each run. The required CPU time can be approximated by CPU=0.15r+0.0045rN when k=2.

The covariance structure used in the simulation study was parameterized to generate either the LCS, PQCS or the QCS. A matrix, Sk, was constructed to generate these three covariance structures. The Sk matrix is given by :

$$(3.33)\quad Sk=\begin{bmatrix} \left(\frac{1+s}{2}\right)^{\left(\frac{k-k}{k-1}-/\right)\frac{1}{2}} & & & 0 \\ & \ddots & & \\ & & \left(\frac{1+s}{2}\right)^{\left(\frac{k-2}{k-1}-/\right)\frac{1}{2}} & \\ & & & \left(\frac{1+s}{2}\right)^{\left(\frac{k-1}{k-1}-/\right)\frac{1}{2}} \\ 0 & & & \end{bmatrix},$$

where s is a skewness parameter.

The covariance matrix of the second population is then obtained from the following equation:

$$(3.34)\quad V= Sk\ dI\ Sk,$$

where I is a k X k identity matrix. The mean vector of population 1 was assumed to have equal elements for all experiments.

The flowchart of the computer simulation procedure is provided in Appendix F. The parameter values used were k=2,4, n=15,30, $\Delta$=1,4, d=1,2,4, and s=1,2,3. The evaluation sample size and the experiment replication were 500 and 200, respectively. The assumptions of equal sample size and equal a priori probabilities were made here so as to reduce the number of experimental combinations.

### 3.4.2 Generalized Likelihood Ratio Test of Proportionality

Based on the training sample, two generalized likelihood ratio tests were conducted. A p-value for each test was recorded. The generalized likelihood test is defined below.

To test

$H_0$: $V_1 = V_2$ (LCS)

$H_1$: $V_1 = (1/d)V_2$ (PCS)

and

$H_0$: $V_1 = (1/d)V_2$ (PCS)

$H_1$: $V_1 \neq (1/d)V_2$ (QCS),

the test statistics are respectively as follows:

(3.35) $t(L/P) = -2 \log \lambda\ (L/P)$

and

(3.36) $t(P/Q) = -2 \log \lambda\ (P/Q)$,

where

$$\lambda(L/P) = \frac{L(\bar{X}1, \bar{X}2, S)}{L(\bar{X}1, \bar{X}2, S, \hat{d})}$$

and

$$\lambda(P/Q) = \frac{L(\bar{X}1, \bar{X}2, S, \hat{d})}{L(\bar{X}1, \bar{X}2, S1, S2)},$$

and where $L(\bar{X}1, \bar{X}2, S)$, $L(\bar{X}1, \bar{X}2, S, \hat{d})$, and $L(\bar{X}1, \bar{X}2, S1, S2)$ are defined by Equations (D.5), (D.11), and (D.16) in Appendix D, respectively.

For the first test, the test statistic $t(L/P)$ has a chi-square distribution with one degree of freedom. The test statistic $t(P/Q)$, however, has a chi-square distribution with $k(k+1)/2 - 1$ degrees of freedom. The two p-values were found to have poor correlation with the error rates (less than 0.3).

### 3.4.3 Delimitation of the Dominant Regions

The estimated expected error rates resulting from each simulation run are provided in Appendix G. The variables shown in these tables should be interpreted as follows:

MEAN = Estimated expected value of the error rate,

STDERR = Standard error of estimate,

APPERL = Apparent error rate of the LDF,

APPERP = Apparent error rate of the PQDF,

APPERQ = Apparent error rate of the QDF,

ACTERL = Estimated actual error rate of the LDF,

ACTERP = Estimated actual error rate of the PQDF,

ACTERQ = Estimated actual error rate of the QDF,

P(L/P) = The p-value obtained from the GLR test for LCS vs. PCS,

P(P/Q) = The p-value obtained from the GLR test for PCS vs. QCS,

DELTAH = Computed group distance,

ALPHAH = Estimate of coefficient of proportionality, d.

It can be seen from Appendix G that:

(1) When d=1, the LDF is the best DF except one case where k=4, n=30, skewness=3, and $\Delta$=1 where the QDF is the best. This means that the QDF will outperform the LDF and the PQDF when the group distance is small, the skewness is high, and the sample size is large relative to the number of variables.

(2) When d=2, $\Delta$=4, and n=15, the LDF is more efficient than the QDF. This implies that the LDF will outperform the QDF if the sample size is small and the group distance is large, even when the population covariances are unequal.

(3) If $d \neq 1$ and skewness$\leq 2$, the PQDF is the best DF except when k=2, n=30, $\Delta$=1, and d=2,4 where QDF

is best. This indicates that the QDF will outperform the PQDF with sufficiently large sample and small group distance.

(4) When k=4 and d≠1, the PQDF is the best DF among the three. This verifies the belief that high dimensionality will increase the efficiency of the PQDF relative to the LDF and the QDF. However, if k=2, skewness=3, and d≠1, the QDF is the best DF except when n=15, $\triangle$=4, and d=2 where the LDF is the best. This suggests that when the sample size and the coefficient of proportionality are small but the group distance is large, the LDF may outperform the PQDF and the QDF even if the population covariances are nonproportional and highly skewed.

(5) If d≠1, the PQDF outperforms the LDF except when k=2, $\triangle$=4, n=15, skewness=3, and d=2 where the LDF is the best DF. This indicates that with small k and n, but large $\triangle$, the LDF may outperform the PQDF and the QDF even if the population covariances are unequal and nonproportional.

(6) The PQDF is the dominant DF when k=4 and d≠1. This finding suggests that the PQDF should be the chosen rule if the dimensionality is not small relative to the sample size, and the population covariances are unequal.

The number of times that either the LDF, the PQDF, or the QDF outperformed the others during 200 replications was recorded and tabulated in Table 10. It was found that the dominant regions are similar to those delimited above except that (1) when k=2, n=15, $\Delta$=1, d=4, and skewness=1,3, the dominant DF is the QDF, and (2) when k=2, n=15, $\Delta$=4, d=2, and skewness=2,3, the dominant DF is the LDF.

According to the estimated expected actual error rates provided in Appendix G, the relations of these error rates to an increasing population parameter can be summarized as follows:

(1) The rate of change of the estimated expected error rates resulting from the change of $\Delta$ is much higher than those from the changes of k, n, d, and the skewness parameter, s, in all cases examined.

(2) The actual error rates of all these three DF's are decreasing with n or $\Delta$ in all cases examined.

(3) The ACTERL is increasing with k in all cases, and (i) decreasing with d if k=2, except the cases where $\Delta$=1, s=3, and $\Delta$=4, n=15, s=3, (ii) increasing with d in all cases if k=4.

(4) The ACTERP is (i) decreasing with k if $\Delta$=1 and d$\neq$1, (ii) increasing with k in all cases if d=1, or if s=1, $\Delta$=4, and n=15, (iii) increasing with k if s=1, $\Delta$=4, and n=30, except one case where d=4, (iv) decreasing with d except one case where k=2, n=30, $\Delta$=4, and s=3.

TABLE 10

Occurrence of Best Discriminant Function during 200 Runs

| k | △ | n | Best | Proportionality d = 1 Skewness 1 | 1 Skewness 2 | 1 Skewness 3 | Proportionality d = 2 Skewness 1 | 2 Skewness 2 | 2 Skewness 3 | Proportionality d = 4 Skewness 1 | 4 Skewness 2 | 4 Skewness 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 15 | LDF | 131 | 110 | 81* | 41 | 36 | 16 | 2 | 0 | 1 |
| | | | PQDF | 48 | 52 | 44 | 113 | 95 | 69 | 146 | 116 | 70 |
| | | | QDF | 21 | 38 | 75 | 46 | 69 | 115 | 52 | 84 | 129 |
| | | 30 | LDF | 120 | 98 | 78 | 30 | 16 | 6 | 0 | 0 | 0 |
| | | | PQDF | 49 | 58 | 36 | 115 | 79 | 53 | 147 | 94 | 59 |
| | | | QDF | 31 | 44 | 86 | 55 | 105 | 141 | 53 | 106 | 141 |
| | 4 | 15 | LDF | 136 | 125 | 117 | 81 | 89 | 79 | 34 | 22 | 37 |
| | | | PQDF | 39 | 49 | 32 | 83 | 76 | 77 | 121 | 126 | 97 |
| | | | QDF | 25 | 26 | 51 | 36 | 35 | 44 | 45 | 52 | 66 |
| | | 30 | LDF | 117 | 116 | 120 | 58 | 58 | 68 | 18 | 15 | 11 |
| | | | PQDF | 45 | 44 | 37 | 104 | 104 | 78 | 138 | 131 | 113 |
| | | | QDF | 38 | 40 | 43 | 38 | 38 | 54 | 44 | 54 | 76 |
| 4 | 1 | 15 | LDF | 125 | 127 | 92 | 7 | 4 | 8 | 0 | 0 | 0 |
| | | | PQDF | 63 | 54 | 45 | 180 | 162 | 132 | 193 | 180 | 161 |
| | | | QDF | 12 | 19 | 63 | 13 | 34 | 60 | 7 | 20 | 39 |
| | | 30 | LDF | 131 | 108 | 69 | 1 | 2 | 1 | 0 | 0 | 0 |
| | | | PQDF | 58 | 58 | 23 | 181 | 164 | 104 | 186 | 149 | 120 |
| | | | QDF | 11 | 34 | 108 | 18 | 34 | 95 | 14 | 51 | 80 |
| | 4 | 15 | LDF | 139 | 128 | 135 | 63 | 56 | 63 | 27 | 25 | 27 |
| | | | PQDF | 33 | 50 | 33 | 114 | 122 | 104 | 143 | 132 | 136 |
| | | | QDF | 28 | 22 | 32 | 23 | 22 | 33 | 30 | 43 | 37 |
| | | 30 | LDF | 125 | 121 | 107 | 44 | 45 | 52 | 7 | 9 | 7 |
| | | | PQDF | 46 | 47 | 41 | 120 | 113 | 101 | 154 | 142 | 133 |
| | | | QDF | 29 | 32 | 52 | 36 | 42 | 47 | 39 | 49 | 60 |

* The number underscored indicates the best discriminant function among the three within each cell.

(5) The ACTERQ is decreasing with d in all cases examined, but (i) increasing with k if d=1 in all cases, (ii) increasing with k if $\Delta$=4, except two cases where n=30, d=4, and s≠1.

(6) The relations of these three error rates to s, shown in Table 11, are complicated. However, with d=4 and $\Delta$=1, the ACTERQ is decreasing with s.

TABLE 11

The Relation of the Estimated Expected Actual Error Rates to the Skewness

| Relation* of | ACTERL to s | | | ACTERP to s | | | ACTERQ to s | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | d | | | d | | | d | |
| k | $\Delta$ | n | 1 | 2 | 4 | 1 | 2 | 4 | 1 | 2 | 4 |
| 2 | 1 | 15 | A | V | + | A | - | - | - | - | - |
| | | 30 | V | + | + | V | A | - | - | - | - |
| | 4 | 15 | + | A | A | + | A | V | A | A | V |
| | | 30 | A | A | - | A | + | + | A | A | A |
| 4 | 1 | 15 | A | + | V | A | A | A | - | - | - |
| | | 30 | A | A | A | A | A | + | - | - | - |
| | 4 | 15 | - | A | A | V | V | + | A | V | + |
| | | 30 | + | V | - | + | + | + | A | - | V |

* "+" = the ACTERR is increasing with increasing skewness
"-" = the ACTERR is decreasing with increasing skewness
"A" = the ACTERR goes up and then down with skewness
"V" = the ACTERR goes down and then up with skewness

### 3.4.4 Relative Efficiency with Finite Sample Sizes

The relative efficiency of the PQDF to the LDF and QDF is defined as the ratio of the deviations of the expected actual error rates from the optimal error rate. Therefore, the relative efficiencies of the PQDF to the LDF and the QDF are defined, respectively, by:

$$(3.37)\quad RE(P/L) = \frac{E(ACTERL) - OPTERR}{E(ACTERP) - OPTERR}$$

and

$$(3.38)\quad RE(P/Q) = \frac{E(ACTERQ) - OPTERR}{E(ACTERP) - OPTERR.}$$

Since the expected actual error rates and the optimal error rate are not available here, their estimates presented in Appendix G were used to compare the relative efficiency (RE) of the PQDF to the LDF and QDF. The resulting estimated RE's are given in Table 12. Plots of estimated RE with skewness=1,3 are shown by Figures 22 through 37 in Appendix E. Notice that in these figures, "+" denotes a value of 1 for d, "x" a value of 2, and "*" a value of 4. The dominant regions delimited by these two RE's are identical to those delimited by the estimated expected actual error rates discussed in Section 3.4.3. The relations of the RE(P/L) and the RE(P/Q) to an increasing population parameter are exhibited by Tables 13 through 17.

It can be seen from Table 12 that:

TABLE 12

The Relative Efficiency of the PQDF to the LDF and the QDF with Finite Sample Sizes

| | | | Proportionality d | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | | | 2 | | | 4 | | |
| | | | Skewness | | | | | | | | |
| k | △ | n | R.E. | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 1 | 1 | 15 | (P/L) | 0.68 | 0.71 | 0.69 | 1.54 | 1.62 | 1.66 | 3.86 | 3.55 | 3.22 |
| | | | (P/Q) | 1.44 | 1.24 | 0.94 | 1.32 | 1.15 | 0.89 | 1.39 | 1.13 | 0.87 |
| | | 30 | (P/L) | 0.66 | 0.77 | 0.79 | 2.72 | 2.36 | 2.01 | 7.56 | 6.04 | 4.50 |
| | | | (P/Q) | 1.59 | 1.33 | 0.79 | 1.43 | 0.95 | 0.65 | 1.37 | 0.94 | 0.62 |
| 2 | 1 | 15 | (P/L) | 0.56 | 0.66 | 0.77 | 1.26 | 1.17 | 0.95 | 3.20 | 3.33 | 2.14 |
| | | | (P/Q) | 1.31 | 1.27 | 1.11 | 1.32 | 1.22 | 1.11 | 1.41 | 1.32 | 0.99 |
| | | 30 | (P/L) | 0.83 | 0.88 | 0.72 | 2.36 | 1.66 | 1.15 | 6.54 | 4.56 | 3.93 |
| | | | (P/Q) | 1.20 | 1.33 | 1.09 | 1.51 | 1.10 | 0.86 | 1.29 | 1.12 | 0.88 |
| 4 | 1 | 15 | (P/L) | 0.89 | 0.87 | 0.91 | 1.98 | 1.86 | 1.74 | 4.21 | 3.79 | 3.53 |
| | | | (P/Q) | 1.70 | 1.42 | 1.09 | 1.82 | 1.50 | 1.30 | 1.93 | 1.68 | 1.47 |
| | | 30 | (P/L) | 0.80 | 0.90 | 0.90 | 3.60 | 2.74 | 2.32 | 8.68 | 6.85 | 5.98 |
| | | | (P/Q) | 2.02 | 1.35 | 0.85 | 2.15 | 1.52 | 1.03 | 1.96 | 1.47 | 1.18 |
| | 4 | 15 | (P/L) | 0.80 | 0.87 | 0.82 | 1.31 | 1.41 | 1.25 | 2.95 | 2.55 | 2.26 |
| | | | (P/Q) | 1.77 | 2.01 | 1.63 | 1.84 | 1.81 | 1.68 | 2.01 | 1.75 | 1.65 |
| | | 30 | (P/L) | 0.92 | 0.93 | 0.88 | 2.19 | 1.95 | 1.74 | 6.88 | 5.22 | 4.92 |
| | | | (P/Q) | 1.71 | 1.62 | 1.17 | 1.60 | 1.44 | 1.24 | 1.94 | 1.50 | 1.33 |

(1) When d=1, the RE(P/L) is less than one in every case examined. If skewness=1, the RE(P/Q) is greater than one in every case. Both the RE(P/L) and the RE(P/Q) are greater than one in every case if k=4 and d≠1.

(2) When skewness=1 and other parameters being fixed, the RE(P/L) is (i) increasing with d in all cases examined,

(ii) increasing with n if $\Delta \neq 1$, (iii) decreasing with $\Delta$ if n=15, and (iv) increasing with k except two cases where n=15, d=4, $\Delta$=4 and n=30, d=2, $\Delta$=4. The RE(P/Q) is (i) increasing with k in all cases examined, (ii) increasing with $\Delta$ if n=15 and $k \neq 2$, (iii) decreasing with $\Delta$ if n=30, except one case where k=2 and d=2, (iv) increasing with n if $\Delta$=1, except one case where k=2 and d=4, (v) decreasing with n if $\Delta$=4, except one case where k=2 and d=2, and (vi) increasing with d if n=15, except one case where k=2 and $\Delta$=1.

(3) When skewness=2 and other parameters being fixed, the RE(P/L) is (i) increasing with either d or n in all cases examined, (ii) increasing with k except the case where n=15, $\Delta$=4, and d=4, and (iii) decreasing with $\Delta$ if $d \neq 1$. The RE(P/Q) is (i) increasing with k in all cases examined, (ii) increasing with $\Delta$ except the case where n=30, k=4, and d=2, (iii) decreasing with d if k=2 and $\Delta$=1, and (iv) decreasing with n except three cases where k=4, $\Delta$=1, d=2 and k=2, $\Delta$=1,4, d=1.

(4) When skewness=3 and other parameters being fixed, the RE(P/L) is (i) increasing with either k or d in all cases examined, (ii) increasing with n if $d \neq 1$, and (iii) decreasing with $\Delta$ except the case where n=15, k=2, and d=1. The RE(P/Q) is (i) increasing with either k or $\Delta$ and decreasing with n, in all cases examined, (ii) increasing with d if k=4, $\Delta \neq 4$, and $n \neq 15$, and (iii) decreasing with d if k=2, $n \neq 30$, and $\Delta \neq 4$.

The relations of the two RE's to the population parameters can be summarized as follows:

(1) The RE(P/L) is, most often, increasing with the number of variables, k, if the group distance is not large relative to the covariance structure. The RE(P/Q) is increasing with k in all cases examined. (See Table 13)

TABLE 13

The Relation of the Finite Sample Relative Efficiency to the Number of Variables

| Relation of* | | | Proportionality d | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | | | 2 | | | 4 | | |
| | | | Skewness | | | | | | | | |
| | △ | n | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| RE(P/L) to k | 1 | 15 | + | + | + | + | + | + | + | + | + |
| | | 30 | + | + | + | + | + | + | + | + | + |
| | 4 | 15 | + | + | + | + | + | + | - | - | + |
| | | 30 | + | + | + | - | + | + | + | + | + |
| RE(P/Q) to k | 1 | 15 | + | + | + | + | + | + | + | + | + |
| | | 30 | + | + | + | + | + | + | + | + | + |
| | 4 | 15 | + | + | + | + | + | + | + | + | + |
| | | 30 | + | + | + | + | + | + | + | + | + |

* "+" denotes that the RE is increasing with increasing k
"-" denotes that the RE is decreasing with increasing k

(2) The RE(P/L) is, in general, decreasing with the group distance, △, if the covariance structure is unegual and/or proportional. The RE(P/Q) is increasing with the group distance if the sample size is not large relative to the number of variables. (See Table 14)

TABLE 14

The Relation of the Finite Sample Relative Efficiency to the Group Distance

| Relation of* | | | Proportionality d | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | | | 2 | | | 4 | | |
| | | | Skewness | | | | | | | | |
| | k | n | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| RE(P/L) to △ | 2 | 15 | - | - | + | - | - | - | - | - | - |
| | | 30 | + | + | - | - | - | - | - | - | - |
| | 4 | 15 | - | + | - | - | - | - | - | - | - |
| | | 30 | + | + | - | - | - | - | - | - | - |
| RE(P/Q) to △ | 2 | 15 | - | + | + | + | + | + | + | + | + |
| | | 30 | - | - | + | + | + | + | - | + | + |
| | 4 | 15 | + | + | + | + | + | + | + | + | + |
| | | 30 | - | + | + | - | - | + | - | + | + |

* "+" denotes that the RE is increasing with increasing △
"-" denotes that the RE is decreasing with increasing △

(3) The RE(P/L) is increasing with the sample size, n, if the coefficient of proportionality is close to unity. The RE(P/Q) is decreasing with the sample size if the covariance structure is nonproportional and highly skewed. (See Table 15)

TABLE 15

The Relation of the Finite Sample Relative Efficiency to the Sample Size

| Relation of* | | | Proportionality d: 1 | | | 2 | | | 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Skewness | | | | | | | | |
| | k | △ | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| RE(P/L) to n | 2 | 1 | - | + | + | + | + | + | + | + | + |
| | | 4 | + | + | - | + | + | + | + | + | + |
| | 4 | 1 | - | + | - | + | + | + | + | + | + |
| | | 4 | + | + | + | + | + | + | + | + | + |
| RE(P/Q) to n | 2 | 1 | + | + | - | + | - | - | - | - | - |
| | | 4 | - | + | - | + | - | - | - | - | - |
| | 4 | 1 | + | - | - | + | + | - | + | - | - |
| | | 4 | - | - | - | - | - | - | - | - | - |

* "+" denotes that the RE is increasing with increasing n
"-" denotes that the RE is decreasing with increasing n

(4) The RE(P/L) is increasing with the coefficient of proportionality, d, in all cases examined. The relationship between the RE(P/Q) and d is vague. However, when the covariance structure is highly skewed, the RE(P/Q) seems to be decreasing (increasing) with d if the number of variables is small (large). (See Table 16)

(5) The RE(P/L) is decreasing with the skewness, s, if d is not near one and/or the sample size is not small relative to the number of variables. The RE(P/Q) is decreasing with s if the group distance is not large relative to the covariance structure. (See Table 17)

However, the combined effects of k, $\Delta$, n, d, and s are complex and interrelated. A simple conclusion about these effects cannot be drawn.

TABLE 16

The Relation of the Finite Sample Relative Efficiency to the Coefficient of Proportionality

| Relation of* | k | $\Delta$ | n | Skewness 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| RE(P/L) to d | 2 | 1 | 15 | + | + | + |
| | | | 30 | + | + | + |
| | | 4 | 15 | + | + | + |
| | | | 30 | + | + | + |
| | 4 | 1 | 15 | + | + | + |
| | | | 30 | + | + | + |
| | | 4 | 15 | + | + | + |
| | | | 30 | + | + | + |
| RE(P/Q) to d | 2 | 1 | 15 | V | - | - |
| | | | 30 | - | - | - |
| | | 4 | 15 | + | V | - |
| | | | 30 | A | V | V |
| | 4 | 1 | 15 | + | + | + |
| | | | 30 | A | A | + |
| | | 4 | 15 | + | - | A |
| | | | 30 | V | V | + |

* "+" denotes that the RE is increasing with increasing d
"-" denotes that the RE is decreasing with increasing d
"A" denotes that the RE goes up and then down with d
"V" denotes that the RE goes down and then up with d

TABLE 17

The Relation of the Finite Sample Relative Efficiency to the Skewness

| Relation of* | k | △ | n | Proportionality d: 1 | 2 | 4 |
|---|---|---|---|---|---|---|
| RE(P/L) to s | 2 | 1 | 15 | A | + | - |
| | | | 30 | + | - | - |
| | | 4 | 15 | + | - | A |
| | | | 30 | A | - | - |
| | 4 | 1 | 15 | V | - | - |
| | | | 30 | + | - | - |
| | | 4 | 15 | A | A | - |
| | | | 30 | A | - | - |
| RE(P/Q) to s | 2 | 1 | 15 | - | - | - |
| | | | 30 | - | - | - |
| | | 4 | 15 | - | - | - |
| | | | 30 | A | - | - |
| | 4 | 1 | 15 | - | - | - |
| | | | 30 | - | - | - |
| | | 4 | 15 | A | - | - |
| | | | 30 | - | - | - |

* "+" denotes that the RE is increasing with increasing s
"-" denotes that the RE is decreasing with increasing s
"A" denotes that the RE goes up and then down with s
"V" denotes that the RE goes down and then up with s

## 3.5 Assumptions and Limitations

Assumptions are necessary in research to reduce the complexity of the problems encountered and thus enable the research to be carried out. Limitations are always incurred by those assumptions which, in turn, cause a loss of generality.

The main assumptions upon which this research rests are listed below under three categories.

(1) **Initial assumptions:**

(i) There are only two populations from which the observed sample vectors come.

(ii) The costs of misclassification are all equal to one, i.e., $C(1|2)=C(2|1)=1$.

(2) **Distributional assumptions:**

(i) The distributions of the two populations are multivariate normal.

(ii) The *a priori* probabilities, $p_1$ and $p_2$, are known and equal in the simulation study, i.e., $p_1=p_2=1/2$.

(iii) The mean vector of population one contains equal elements in the simulation study.

(3) **Sampling assumptions:**

(i) The sample vectors are independent, identically distributed, and conditional on the population.

(ii) The sample sizes are equal, i.e., $n_1=n_2$.

(iii) It is known from which population a sample vector comes, i.e., every sample vector is initially correctly classified.

(iv) There are no missing values within any observed sample vector.

The limitations of this study are stated as follows:

(1) The aforementioned assumptions are made for the sake of research tractability. The two initial assumptions and the assumptions of equal a priori probabilities and equal-element mean vector can be relaxed with little additional effort required to demonstrate the applicability of this study. However, the violation of the other assumptions may cause the modification of the proposed study to be an arduous and elusive work.

(2) The performance of a discriminant function depends on the values of the underlying parameters, thus the regions dominated by the PQDF in the parameter space depend on the points or intervals of the parameters. In the simulation study, the parameters under consideration are: the number of variables k, the Mahalanobis's distance between the two mean vectors $\Delta$, the training sample size n ($n_1=n_2=n$), the coefficient of proportionality d, and the skewness of the second population covariance. Other parameters, such as a priori probability $p_i$ ($i=1,2$), the ratio of the misclassification costs $C(1|2)/C(2|1)$, etc., which may have some effects on the performance of the discriminant functions under study are assumed constant. The effects of these latter parameters remain unknown throughout the entire research.

(3) The boundaries of the regions dominated differently by the LDF, QDF and the PQDF were found to be ambiguous. Many researchers have made every effort to find

clear-cut boundaries for regions of the LDF and QDF without any success. This drawback, which may be attributed to the increased complexity of the combined parameter effects, cannot be rectified by this research.

# Chapter IV

## AN APPLICATION

In order to demonstrate the applicability of this research, an analysis of the pre-admission and academic achievement data of the medical doctoral program at Texas Tech University Health Sciences Center was conducted.

The data provided in Appendix H contains 130 sample points of medical students who will graduate in 1982 and 1983. Each sample observation was characterized by five numerical measurements: BCPM, OTHER, MCATBIO, MCATAVGV and CUMAVG. These five variables are defined as follows:

BCPM = Weighted average of the course grades of Biology, Chemistry, Physics, and Mathematics,

OTHER = Weighted average of all the course grades other than the BCPM,

MCATBIO = Score of the Biology section in the Medical College Admission Test (MCAT),

MCATAVG = Average score of all the section tests in the MCAT,

CUMAVG = Cumulative grade point average of the student's first four semesters at Texas Tech University Health Sciences Center.

The objective of the following analysis is to develop the best discriminant function (DF) so as to classify these students into two groups: high-performance and low-performance. This estimated DF can then be used as the future admission criterion.

The performance of each student was measured by the CUMAVG. The cutoff point was set to be CUMAVG=81.78 so as to divide students into two groups of equal sample sizes. Students who have CUMAVG≤81.78 were assigned to group 1, otherwise, to group 2. Three discriminant functions, the LDF, PQDF, and the QDF, were then derived using the variables BCPM, OTHER, MCATBIO, and MCATAVG. The a priori probabilities were assumed to be equal. Both the apparent error rates and the almost unbiased estimate of expected actual error rates given by the U-method were calculated. Some resulting statistics of this analysis are presented in Table 18.

Without knowing the resulting error rates, one might foresee the performance of the three DF's by examining Tables 10 and 12, and Figures 22 through 37. In this case, the number of variables and the training sample size are large, the computed distance is small, and the estimated coefficient of proportionality is less than one. Since it is invariant in terms of error rates to treat d as $d^{-1}$, the estimated d can be set to $(0.700073)^{-1}=1.4284$ here. It can be seen from Table 10 that the PQDF, under these conditions,

TABLE 18

An Application of the Three Discriminant Functions To Medical School Data (Four-variate)

| DF | Error Rates: Apparent Error Rate | Error Rates: Error Rate from U-method | Other Statistics |
|---|---|---|---|
| | | | D=DELTAH = 0.524671 |
| | | | $\hat{d}$=ALPHAH = 0.700073 |
| LDF | .384615 | .415385 | |
| | | | P(L/P) = 0.005850 |
| PQDF | .361538 | .392308 | |
| | | | P(P/Q) = 0.194534 |
| QDF | .384615 | .446154 | |

outperforms the LDF and the QDF in at least 52% (104/200) of the cases studied when d>1. Further, referring to Table 12 and Figures 24, 28, 32, and 36, it can be found that the PQDF has, in general, RE(P/L)>1.0 and RE(P/Q)>1.0 regardless the skewness.

The computed error rates showed that the PQDF is the best DF since it gives the lowest apparent error rate and estimated expected actual error. This finding is in agreement with the previous prediction. The conclusions drawn from the GLR tests also indicate that the PCS is the dominant covariance structure. However, the p-values of the GLR tests for LCS vs. PCS and PCS vs. QCS were found to be unreliable measures of the degree of proportionality in the simulation study. Thus, based upon only the two p-values, it cannot be absolutely certain that the PQDF will outperform the others. However, these GLR tests show that

the covariance structure is not the LCS but the PCS and thus they improve the accuracy of the above prediction.

Furthermore, the two most significant variables, BCPM and MCATBIO, were retained based upon the F-criterion (Dixon 1970, 1975, Nie, et al. 1975). A discriminant analysis using these two variables was then performed. It was found that the estimated PQDF remains the best estimated DF among the three according to the computed error rates exhibited in Table 19.

TABLE 19

An Application of the Three Discriminant Functions To Medical School Data (Two-variate)

| DF | Error Rates* | | Other Statistics |
|---|---|---|---|
| | Apparent Error Rate | Error Rate from U-method | |
| LDF | .376923 | .392308 | D=DELTAH = 0.471526 |
| PQDF | .346154 | .361538 | $\hat{d}$=ALPHAH = 0.835062 |
| QDF | .369231 | .407692 | P(L/P) = 0.304811 |
| | | | P(P/Q) = 0.913599 |

* Only BCPM and MCATBIO were used as the discriminating variables

The conclusion of this analysis is that the PQDF is the best discriminant function for this particular data set in terms of the error rate given by the U-method. This demonstration also shows the usefulness of the guidelines

provided by this research for predicting the best discriminant function.

# Chapter V

# SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

## 5.1 Summary and Conclusions

The purpose of this research is to examine the proportional quadratic discriminant function (PQDF) and to compare its performance with those of the linear discriminant function (LDF) and the quadratic discriminant function (QDF). The belief was verified that high dimensionality will increase the efficiency of the PQDF over the QDF and enable it to outperform the LDF and the QDF when the population parameters must be estimated.

The design of this research was divided into two parts: the analytic method and the simulation study. The analytic method used the maximum likelihood procedure and Newton-Raphson method to estimate d. The asymptotic error rates (AERR) of the estimated LDF, QDF, and PQDF were calculated in two-variate linear or proportional covariance structures (LCS or PCS). Then, the performances of the estimated PQDF relative to the estimated LDF and QDF were compared using the asymptotic relative efficiency (ARE). It was shown that:

(1) For the LCS, the ARE(P/L) and ARE(Q/P) are decreasing functions of the group distance ($\Delta$) for $\Delta \geq 4$ (see Figures 8 through 11 in Appendix E).

(2) For the PCS, the ARE(Q/P) is a decreasing function of $\Delta$ if $p\underline{1}=p\underline{2}=0.5$ and $\Delta \geq 4$ (see Figure 13 in Appendix E).

(3) The regions of the largest ARE(P/L) and ARE(Q/P) are all delimited by $2 \leq \Delta \leq 4$ for both covariance structures.

(4) The ARE(P/L) and the ARE(Q/P) are always less than one for both structures.

(5) For equal <u>a priori</u> probabilities and $2 \leq \Delta \leq 4$, the ARE(Q/P) with the PCS is a decreasing function of d.

(6) When k=1 and $\Delta=2$, the ARE(P/L) and ARE(Q/P) is 1.0.

The simulation study used k=2,4, $n\underline{1}=n\underline{2}=15,30$, $\Delta=1,4$, d=1,2,4, and skewness=1,2,3 as the values of the parameters. Estimated actual error rates (ACTERR) were computed using 500 sample observations. Expected values of estimated ACTERR were calculated using the average of 200 replications. Two generalized likelihood ratio tests (LCS vs. PCS, and PCS vs. QCS) were conducted so as to obtain the p-value for both tests. The PQDF was found to outperform both the LDF and the QDF, according to Table 12, when k=4 and $d \neq 1$. This finding suggests that the PQDF is superior to

the QDF if the number of discriminating variables to be used is not small relative to the sample size. However, the LDF performed better than the PQDF in all cases when d=1. This result suggests that if the estimated coefficient of proportionality is near one, the LDF should be used rather than the PQDF.

## 5.2 Recommendations

This research has investigated the properties and performance of the PQDF relative to the LDF and the QDF. The estimated PQDF was found to be a very useful DF since it outperforms the estimated QDF with finite sample sizes under many circumstances.

In practice, the population parameters are usually unknown and thus an estimated DF must be used. The findings aforementioned suggest that the PQDF should be considered as an alternative to the QDF when the population dispersions are found to be unequal.

Finally, the recommendations from this research are stated as follows:

(1) With high dimensionality, the difference in the performance of the PQDF and the QDF will be substantial if the training sample size is small. Unless the training sample is sufficiently large, the PQDF should usually be chosen when the population dispersions are unequal and the number of variables is large ($k \geq 4$).

(2) Increasing the training sample size or the group distance improves the performance of any one of the three discriminant functions discussed in this research. The relative efficiency of the PQDF to the LDF and the QDF relative to the PQDF will increase (decrease) along with increasing sample size (group distance) when the covariance structure is nonproportional and highly skewed. In other words, the performance of the PQDF (QDF) improves faster than that of the LDF (PQDF) with the sample size but slower with the group distance. The rates of change of an estimated DF's performance relative to the training sample size and the group distance should be investigated.

(3) It is recommended to practitioners that a method that gives an unbiased estimate of the expected actual error rate such as the U method should always be used. In this research, it was found that the QDF often gives a lower expected apparent error rate than the LDF and the PQDF since the apparent error rate is a biased estimate. This means that the QDF often fits the sample data better than the LDF and PQDF.

(4) Simple conclusions cannot be drawn about the relationship between the relative efficiency of the PQDF to the QDF and the coefficient of proportionality d. This finding confirms the result of the generalized likelihood ratio test which showed that the p-value of the test of the PCS vs. the QCS is not a good indicator of the relative

efficiency. Thus, the relation of the p-value to the covariance structure (d and skewness) should be further explored. Nevertheless, this research provides useful guidelines for practitioners in selecting the best estimated DF according to the number of variables, the training sample size, the group distance, and the covariance structure.

## LIST OF REFERENCES


Adelman, Irma and Cynthia Taft Morris, "Performance Criteria for Evaluating Economic Development Potential: An Operational Approach," Quarterly Journal of Economics, Vol. 82, No. 2, May 1968, pp. 260-280.

Altman, Edward I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," Journal of Finance, Vol. 23, No. 4, September 1968, pp. 589-609.

Altman, Edward I., "Predicting Railroad Bankruptcies in America," The Bell Journal of Economics and Management Science, Vol. 4, No. 1, Spring 1973, pp. 184-211.

Altman, Edward I. and Steven Katz, "Statistical Bond Rating Classification Using Financial and Accounting Data," Proceedings of the Conference on Topical Research in Accounting, Michael Schiff and George Sorter, eds., New York: Ross Institute of Accounting Research, The School of Business of New York University, 1976, pp. 205-239.

Anderson, T. W., "Classification by Multivariate Analysis," Psychometrika, Vol. 16, 1951, pp. 31-50.

Anderson, T. W., An Introduction to Multivariate Statistical Analysis, New York: John Wiley & Sons, Inc., 1958.

Anderson, T. W., "An Asymptotic Expansion of the Distribution of the 'Studentized' Classification Statistic W," Annals of Statistics, Vol. 1, 1973, pp. 964-972.

Anderson, T. W., "Classification and Discrimination," International Encyclopedia of Statistics, William H. Kruskal and Judith M. Tanur, eds., New York: The Free Press, 1977, pp. 628-635.

Anderson, T. W. and R. Bahadur, "Classification into Two Multivariate Normal Distributions with Different Covariance Matrices," Annals of Mathematical Statistics, Vol. 33, 1962, pp. 420-431.

Awh, R. Y. and D. Waters, "A Discriminant Analysis of Economic, Demographic, and Attitudinal Characteristics of Bank Charge-Card Holders: A Case Study," Journal of Finance, Vol. 29, No. 3, June 1974, pp. 973-980.

Banerjee, K. S. and L. F. Marcus, "Bounds in a Minimax Classification Procedure," Biometrika, Vol. 52, 1965, pp. 653-654.

Banks, Seymour, "The Relationships Between Preference and Purchase of Brands," Journal of Marketing, Vol. 15, October 1950, pp. 145-157.

Barnard, M. M., "The Secular Variations of Skull Characters in Four Series of Egyptian Skulls," Annuals of Eugenics, Vol. 6, December 1935, pp. 352-371.

Bartlett, M. S., "The Vector Representation of a Sample," Proceedings of the Cambridge Philosophical Society, Vol. 30, 1934, pp. 327-340.

Bartlett, M. S., "Multivariate Analysis," Journal of the Royal Statistical Society, Series B, Vol. 9, 1947a, pp. 176-197.

Bartlett, M. S., "The General Canonical Correlation Distribution," Annals of Mathematical Statistics, Vol. 18, 1947b, pp. 1-17.

Bartlett, M. S., "An Inverse Matrix Adjustment Arising in Discriminant Analysis," Annals of Mathematical Statistics, Vol. 22, 1951, pp. 107-111.

Bartlett, M. S. and N. W. Please, "Discrimination in the Case of Zero Mean Differences," Biometrika, Vol. 50, 1963, pp. 17-21.

Bass, Frank M. and W. Wayne Talarzyk, "An Attitude Model for the Study of Brand Preference," Journal of Marketing Research, Vol. 9, February 1972, pp. 93-96.

Bates, Timothy, "An Econometric Analysis of Lending to Black Businessmen," Review of Economics and Statistics, Vol. 55, August 1973, pp. 272-283.

Beale, E. M. L., M. G. Kendall and D. W. Mann, "The Discarding of Variables in Multivariate Analysis," Biometrika, Vol. 54, 1967, pp. 357-366.

Beaver, W. H., "Financial Ratios as Predictors of Failure," Journal of Accounting Research, Supplement to Vol. 4, 1966, pp. 71-111.

Beranek, W., Analysis for Financial Decisions, Homewood, Ill.: Richard D. Irwin, Inc., 1963.

Bhattacharya, D. P. and R. D. Narayan, "Moments of the $D^2$-statistic for Populations with Unequal Dispersions," Sankhya, Vol. 5, 1941, pp. 401-412.

Blood, D. M. and C. B. Baker, "Some Problems of Linear Discrimination," Journal of Farm Economics, Vol. 40, 1958, pp. 674-683.

Bose, R. C., "On the Exact Distribution and Moment Coefficients of the $D^2$-statistic," Sankhya, Vol. 2, 1936a, pp. 143-154.

Bose, R. C., "A Note on the Distribution of Differences in Mean Values of Two Samples Drawn From Two Multivariate Normally Distributed Populations, and the Definition of the $D^2$-statistic," Sankhya, Vol. 2, 1936b, pp. 379-384.

Bose, R. C. and S. N. Roy, "The Distribution of the Studentized $D^2$-statistic," Sankhya, Vol. 4, 1938, pp. 19-38.

Boullion, T. L., P. L. Odell and B. S. Duran, "Estimating the Probability of Misclassification and Variate Selection," Pattern Recognition, Vol. 7, 1975, pp. 139-145.

Bowker, A. H., "A Representation of Hotelling's $T^2$ and Anderson's Classification Statistic W in Terms of Simple Statistics," Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, Ingram Olkin, et al., eds., Stanford, Calif.: Stanford University Press, 1960, pp. 142-150.

Bowker, A. H. and R. Sitgreaves, "An Asymptotic Expansion for the Distribution of the W-Classification Statistic," Studies in Item Analysis and Prediction, H. Solomon, ed., Stanford, Calif.: Stanford University Press, 1961, pp. 293-310.

Box, G. E. P., "A General Distribution Theory for a Class of Likelihood Criteria," Biometrika, Vol. 36, 1949, pp. 317-346.

Brailovskiy, V. L., "An Object Recognition Algorithm with Many Parameters and Its Applications," Engineering Cybernetics, USSR, No. 2, 1964, pp. 22-30.

Brailovskiy, V. L. and A. L. Lunts, "The Multiparameter Recognition Problem and Its Solution," Engineering Cybernetics, USSR, No. 1, 1964, pp. 13-22.

Brody, Robert P. and Scott M. Cunningham, "Personality Variables and the Consumer Decision Process," Journal of Marketing Research, Vol. 5, February 1968, pp. 50-57.

Bromley, D. W., "The Use of Discriminant Analysis in Selecting Rural Development Strategies," American Journal of Agricultural Economics, No. 53, 1971, pp. 319-322.

Bucklin, Louis P., "Trade Area Boundaries: Some Issues in Theory and Methodology," Journal of Marketing Research, Vol. 8, February 1971, pp. 30-37.

Bunke, O., "Uber Optimale Verfahren der Diskriminazanalyse," Abhandlungen der Deutschen Akademie der Wissenschaften zu Berlin, Klasse fur Mathematik, Physik und Technik, Vol. 4, 1964, pp. 35-41.

Carleton, W. T. and E. M. Lerner, "Statistical Credit Scoring of Municipal Bonds," Journal of Money, Credit, and Banking, Vol. 1, No. 4, November 1969, pp. 750-764.

Cavalli, L. L., "Alumni Problemi della Analisi Biometrica di Popolazioni Naturali," Memorie dell'Istituto Italiano di Idrobiologia Dott, Vol. 2, 1945, pp. 301-323.

Chaddha, R. L. and L. F. Marcus, "An Empirical Comparison of Distance Statistics for Populations with Unequal Covariance Matrices," Biometrics, Vol. 24, 1968, p. 683-694.

Churchill, Gilbert A., Jr., Neil M. Ford and Urban B. Ozanne, "An Analysis of Price Aggressiveness in Gasoline Marketing," Journal of Marketing Research, Vol. 7, February 1970, pp. 36-42.

Claycamp, Henry J., "Characteristics of Thrift Deposit Owners," Journal of Marketing Research, Vol. 2, May 1965, pp. 163-170.

Clunies-Ross, C. W. and R. H. Riffenburgh, "Geometry and Linear Discrimination," Biometrika, Vol. 47, 1960, pp. 185-189.

Cochran, W. G., "On the Performance of the Linear Discriminant Function," Technometrics, Vol. 6, 1964, pp. 179-190.

Cochran, W. G. and C. Hopkins, "Some Classification Methods with Multivariate Qualitative Data," Biometrics, Vol. 17, 1961, pp. 10-32.

Cooley, William W. and Paul R. Lohnes, Multivariate Data Analysis, New York: John Wiley & Sons, Inc., 1971.

Cooper, Paul W., "Statistical Classification with Quadratic Forms," Biometrika, Vol. 50, 1963, pp. 439-448.

Crask, Melvin R. and William D. Perreault, Jr., "Validation of Discriminant Analysis in Marketing Research," Journal of Marketing Research, Vol. 14, February 1977, pp. 60-68.

Dake, J. L., "Comment: An Empirical Test of Financial Ratio Analysis," Journal of Financial and Quantitative Analysis, Vol. 7, No. 2, March 1972, pp. 1495-1497.

Darden, William R. and Fred D. Reynolds, "Backward Profiling of Male Innovators," Journal of Marketing Research, Vol. 11, February 1974, pp. 79-85.

Das Gupta, S., "Probability Inequalities and Error in Classification," Technical Report No. 190, School of Statistics, University of Minnesota, 1972.

Davies, M. G., "The Expectation of Mahalanobis' Generalized Distance," Annals of the Institute of Statistical Mathematics, Vol. 24, No. 1, 1972, pp. 111-125.

Day, N. E., "Linear and Quadratic Discrimination in Pattern Recognition," IEEE Transactions on Information Theory, Vol. IT-14, 1969, pp. 419-420.

Day, N. E. and D. F. Kerridge, "A General Maximum Likelihood Discriminant," Biometrics, Vol. 23, 1967, pp. 313-323.

Dillon, William R., "The Performance of the Linear Discriminant Function in Nonoptimal Situations and the Estimation of Classification Error Rates: A Review of Recent Findings," Journal of Marketing Research, Vol. 16, August 1979, pp. 370-381.

Dince, Robert R. and James C. Fortson, "The Use of Discriminant Analysis to Predict the Capital Adequacy of Commercial Banks," Journal of Bank Research, Vol. 3, No. 1, Spring 1972, pp. 54-62.

Dixon, W. J., ed., BMD: Biomedical Computer Programs, Berkeley and Los Angels: University of California Press, 1970.

Dixon, W. J., ed., BMDP: Biomedical Computer Programs, Berkeley and Los Angeles: University of California Press, 1975.

Dixon, W. J. and F. J. Massey, Jr., Introduction to Statistical Analysis, 2nd ed., New York: McGraw-Hill Book Company, 1957.

Duda, R. O. and P. E. Hart, Pattern Classification and Scene Analysis, New York: John Wiley & Sons, Inc., 1973.

Dunn, Olive Jean, "Some Expected Values for Probabilities of Correct Classification in Discriminant Analysis," Technometrics, Vol. 13, No. 2, May 1971, pp. 345-353.

Dunn, Olive Jean and P. V. Varady, "Probabilities of Correct Classification in Discriminant Analysis," Biometrics, Vol. 22, 1966, pp. 908-924.

Durand, D., Risk Elements in Consumer Installment Financing: Technical Edition, New York: National Bureau of Economic Research, 1941.

Edmister, R. O., "An Empirical Test of Financial Ratio Analysis for Small Business Prediction," Journal of Financial and Quantitative Analysis, Vol. 7, No. 2, March 1972, pp. 1477-1494.

Efron, Bradley, "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis," Journal of the American Statistical Association, Vol. 70, 1975, pp. 892-898.

Eisenbeis, Robert A., "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics," Journal of Finance, Vol. 32, No. 3, June 1977, pp. 875-900.

Eisenbeis, Robert A. and E. I. Altman, "Financial Applications of Discriminant Analysis: A Clarification," Journal of Financial and Quantitative Analysis, Vol. 33, No. 1, March 1978, pp. 185-195.

Eisenbeis, Robert A. and Robert B. Avery, Discriminant Analysis and Classification Procedures: Theory and Applications, Lexington, Mass.: D. C. Heath and Co., 1972.

Eisenbeis, Robert A. and Alan S. McCall, "Some Effects of Affiliations Among Mutual Savings and Commercial Banks," Journal of Finance, Vol. 27, No. 4, September 1972, pp. 865-877.

Eisenbeis, Robert A. and Neil B. Murphy, "Interest Rate Ceilings and Consumer Credit Rationing: A Multivariate Analysis of A Survey of Borrowers," Southern Economic Journal, Vol. 41, No. 1, July 1974, pp. 115-123.

Etgar, Michael, "The Effects of Administrative Control on Efficiency of Vertical Marketing Systems," Journal of Marketing Research, Vol. 13, February 1976, pp. 12-24.

Evans, Franklin B., "Psychological and Objective Factors in the Prediction of Brand Choice: Ford versus Chevrolet," Journal of Business, Vol. 32, October 1959, pp. 340-369.

Fisher, R. A., "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, Vol. 7, 1936, pp. 179-188.

Fisher, R. A., "The Statistical Utilization of Multiple Measurements," Annals of Eugenics, Vol. 8, 1938, pp. 376-386.

Frank, R. E., W. F. Massy and D. G. Morrison, "Bias in Multiple Discriminant Analysis," Journal of Marketing Research, Vol. 2, 1965, pp. 250-258.

Fukunaga, K. and D. Kessel, "Estimation of Classification Error," IEEE Transactions on Computers, Vol. C-20, 1971, pp. 1521-1527.

Fukunage, K. and D. Kessel, "Error Evaluation and Model Validation in Statistical Pattern Recognition," Technical Report TR-EE 72-23, School of Electrical Engineering, Purdue University, 1972.

Gilbert, Ethel S., "The Effect of Unequal Variance-Covariance Matrices on Fisher's Linear Discriminant Function," Biometrics, Vol. 25, September 1969, pp. 505-515.

Gilbert, Gary G., "Predicting De Novo Expansion in Bank Merger Cases," Journal of Finance, Vol. 29, No. 1, March 1974 , pp. 151-162.

Gilbert, Gary G., "An Analysis of Federal Regulatory Decisions on Market Extension Bank Merger," Journal of Money, Credit, and Banking, Vol. 7, No. 1, February 1975, pp. 81-92.

Gitlow, Howard S., "Discrimination Procedures for the Analysis of Nominally Scaled Data Sets," Journal of Marketing Research, Vol. 16, August 1979, pp. 387-393.

Glick, Ned, "Additive Estimators for Probabilities of Correct Classification," Pattern Recognition, Vol. 10, 1978, pp. 211-222.

Gramm, W. L., "The Labour Force Decision of Female Married Teachers: A Discriminant Analysis Approach," Review of Economics and Statistics, Vol. 55, 1973, pp. 341-348.

Habbema, J. D. F. and J. Hermans, "Selection of Variables in Discriminant Analysis by F-statistic and Error Rate," Technometrics, Vol. 19, 1977, pp. 487-493.

Habbema, J. D. F., J. Hermans and K. Van Den Broek, "A Stepwise Discriminant Analysis Program Using Density Estimation," Proceedings in Computational Statistics, Wien: Physica Verlag, 1974, pp. 101-110.

Han, Chien-Pai, **"A Note on Discrimination in the Case of Unequal Covariance Matrices,"** Biometrika, Vol. 55, 1968, pp. 586-587.

Han, Chien-Pai, **"Distribution of Discriminant Function When Covariance Matrices are Proportional,"** Annals of Mathematical Statistics, Vol. 40, No. 3, 1969, pp. 979-985.

Han, Chien-Pai, **"Distribution of Discriminant Function in Circular Models,"** Annals of the Institute of Statistical Mathematics, Vol. 22, 1970, pp. 117-125.

Han, Chien-Pai, **"Asymptotic Distribution of Discriminant Function When Covariance Matrices are Proportional and Unknown,"** Annals of the Institute of Statistical Mathematics, Vol. 26, No. 1, 1974, pp. 127-133.

Harter, H. L., **"On the Distribution of Wald's Classification Statistic,"** Annals of Mathematical Statistics, Vol. 22, 1951, pp. 58-67.

Harvey, John, **"The Content Characteristics of Best-Selling Novels,"** Public Opinion Quarterly, Vol. 17, Spring 1953, pp. 91-114.

Haslem, John A. and William A. Longbrake, **"A Discriminant Analysis of Commercial Bank Profitability,"** Quarterly Review of Economics and Business, Vol. 11, No. 3, 1971, pp. 39-46.

Hermans, J. and J. D. F. Habbema, **"Manual for the ALLOC Discriminant Analysis Programs,"** Technical Report, Department of Medical Statistics, University of Leiden, 1976.

Highleyman, W. H., **"The Design and Analysis of Pattern Recognition Experiments,"** Bell System Technical Journal, Vol. 41, March 1962, pp. 723-744.

Hills, M., **"Allocation Rules and Their Error Rates,"** Journal of the Royal Statistical Society, Series B, Vol. 28, 1966, p. 1-31.

Hodges, Joseph L., Jr., **"Survey of Discriminatory Analysis,"** Report No. 1, USAF School of Aviation Medicine, Texas: Randolpn Field, October, 1950.

Holloway, Lois N. and Olive J. Dunn, **"The Robustness of Hotelling's $T^2$,"** Journal of the American Statistical Association, Vol. 62, 1967, pp. 124-136.

Hora, Stephen C. and Michael Hanna, "Estimated Error Rates for Linear Predictive Discriminants," Proceedings of the ASA Section of the Business and Economic Statistics, Washington, D. C.: American Statistical Association, 1978, pp. 763-768.

Hora, Stephen C. and James Wilcox, "Estimation of Error Rates in Several Population Discriminant Analysis," Unpublished paper, College of Business Administration, Texas Tech University, Lubbock, Texas, 1981.

Horton, Joseph J., Jr., "Statistical Classification of Municipal Bonds," Journal of Bank Research, Vol. 1, No. 3, Autumn 1970, pp. 29-40.

Hotelling, H., "The Generalization of Student's Ratio," Annals of Mathematical Statistics, Vol. 2, 1931, pp. 360-378.

Hotelling, H., "A Generalized T Test and Measure of Multivariate Dispersion," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1951, pp. 23-41.

Huber, Joel and Morris B. Holbrook, "Using Attribute Ratings for Product Positioning: Some Distinctions Among Compositional Approaches," Journal of Marketing Research, Vol. 16, November 1979, pp. 507-516.

Huberty, Carl J., "Discriminant Analysis," Review of Educational Research, Vol. 45, 1975, pp. 543-598.

Huberty, Carl J. and Allen R. Curry, "Linear Versus Quadratic Multivariate Classification," Maltivariate Behavioral Research, Vol. 13, 1978, pp. 237-245.

Hustad, Thomad and Edgar Pessemier, "Will the Real Consumer Activist Please Stand Up: An Examination of Consumers' Opinions About Marketing Practices," Journal of Marketing Research, Vol. 10, August 1973, pp. 319-324.

Jennrich, Robert I., "Linear Discrimination in the Case of Unequal Covariance Matrices," Unpublished paper, Department of Biomathematics, University of California, Los Angeles, 1962.

John, S., "The Distribution of Wald's Classification Statistic When the Dispersion Matrix is Known," Sankhya, Vol. 21, 1959, pp. 371-376.

John, S., "On Some Classification Problems-I," Sankhya, Vol. 22, 1960, pp. 301-308.

John, S., "On Some Classification Statistics," Sankhya, Vol. 22, 1960, pp. 309-316.

John, S., "Errors in Discrimination," Annals of Mathematical Statistics, Vol. 32, 1961, pp. 1125-1144.

John, S., "On Classification by the Statistics R and Z," Annals of the Institute of Statistical Mathematics, Vol. 14, 1962, pp. 237-246.

John, S., "Further Results on Classification by W," Sankhya, Vol. 26, 1964, pp. 39-46.

Johnson, C. G., "Ration Analysis and the Prediction of Firm Failure," Journal of Finance, Vol. 25, No. 5, December 1970, pp. 1166-1168.

Jolliffe, I. T., "Discarding Variables in a Principal Component Analysis I: Artificial Data," Applied Statistics, Vol. 21, 1972, PP. 160-173.

Jolliffe, I. T., "Discarding Variables in a Principal Component Analysis II: Real Data," Applied Statistics, Vol. 22, 1973, pp. 21-23.

Joy, O. M. and J. O. Tollefson, "On the Financial Applications of Discriminant Analysis," Journal of Financial and Quantitative Analysis, Vol. 10, No. 5, December 1975, pp. 723-740.

Joy, O. M. and J. O. Tollefson, "Some Clarifying Comments on Discriminant Analysis," Journal of Financial and Quantitative Analysis, Vol. 33, No. 1, March 1978, pp. 197-200.

Kabe, D. G., "Some Results on the Distribution of Two Random Matrices Used in Classification Procedures," Annals of Mathematical Statistics, Vol. 34, 1963, pp. 181-185.

Katz, Steven, "The Price Adjustment Process of Bonds to Rating Reclassifications: A Test of Bond Market Efficiency," Journal of Finance, Vol. 29, No. 2, May 1974, pp. 551-559.

Kendall, Maurice G., A Course in Multivariate Analysis, New York: Hafner Publishing Company, 1957.

Kendall, Maurice G. and Alan Stuart, The Advanced Theory of Statistics, Vol. 3: Design and Analysis, and Time-Series, 3rd ed., London: Charles Griffin & Company Limited, 1975.

Kendall, Maurice G. and Alan Stuart, The Advanced Theory of Statistics, Vol. 2: Inference and Relationship, 4th ed., Riverside, New Jersey: Macmillan Publishing Co., Inc., 1979.

King, William R., "Marketing Expansion - A Statistical Analysis," Management Science, Vol. 9, July 1963, pp. 563-573.

Klemkosky, Robert C. and J. William Petty, "A Multivariate Analysis of Stock Price Variability," Journal of Business Research, Vol. 1, No. 1, Summer 1973, pp. 1-10.

Kullback, S., "An Application of Information Theory to Multivariate Analysis," Annals of Mathematical Statistics, Vol. 23, 1952, pp. 82-102.

Kullback, S., Information Theory and Statistics, New York: John Wiley & Sons, Inc., 1959.

Lachenbruch, Peter A., "Estimation of Error Rates in Discriminant Analysis," Unpublished doctoral dissertation, University of California at Los Angeles, 1965.

Lachenbruch, Peter A., "Discrimination Analysis When Initial Samples are Misclassified," Technometrics, Vol. 8, 1966, pp. 657-662.

Lachenbruch, Peter A., "An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis," Biometrics, Vol. 23, 1967, pp. 639-645.

Lachenbruch, Peter A., "On Expected Probabilities of Misclassification in Discriminant Analysis, Necessary Sample Size, and a Relation with the Multiple Correlation Coefficient," Biometrics, Vol. 24, December 1968, pp. 823-834.

Lachenbruch, Peter A., "Discriminant Analysis Where the Initial Samples are Misclassified. II: Non-Random Misclassification Models," Technometrics, Vol. 16, 1974, pp.419-424.

Lachenbruch, Peter A., Discriminant Analysis, New York: Hafner Press, 1975.

Lachenbruch, Peter A., "Note on Initial Misclassification Effect on the Quadratic Discriminant Function," Technometrics, Vol. 21, No. 1, February 1979, pp. 129-132.

Lachenbruch, Peter A. and M. R. Mickey, "Estimation of Error Rates in Discriminant Analysis," Technometrics, Vol. 10, 1968, pp. 1-11.

Lachenbruch, Peter A., C. Sneeringer and L. T. Revo, "Robustness of the Linear and Quadratic Discriminant Functions to Certain Types of Non-Normality," Communications in Statistics, Vol. 1, 1973, pp. 39-56.

Lainiotis, D. G. and S. K. Park, "Probability of Error Bounds," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-1, April 1971, pp. 175-178.

Lane, S., "Submarginal Credit Risk Classifications," Journal of Financial and Quantitative Analysis, Vol. 7, No. 1, January 1972, pp. 1379-1386.

Layard, Maxwell W. J., "Large Sample Tests for the Equality of Two Covariance Matrices," Annals of Mathematical Statistics, Vol. 43, 1972, pp. 123-141.

Lessig, V. Parker and John Tollefson, "Market Segmentation Through Numerical Taxonomy," Journal of Marketing Research, Vol. 8, November 1971, pp. 480-487.

Linhart, H., "Techniques for Discriminant Analysis with Discrete Variables," Metrika, Vol. 2, May 1959, pp. 138-149.

Lubin, A., "Linear and Non-Linear Discriminating Functions," British Journal of Psychology, Vol. 3, 1950, pp. 90-104.

Lunts, A. L. and V. L. Brailovskiy, "Evaluation of Attributes Obtained in Statistical Decision Rules," Engineering Cybernetics, USSR, No. 3, 1967, pp. 98-109.

Mahalanobis, P. C., "Analysis of Race Mixture in Bengal," Journal and Proceedings of Asiatic Society of Bengal, Vol. 23, No. 3, 1927, pp. 301-333.

Mahalanobis, P. C., "On Tests and Measurements of Group Divergence," Journal and Proceedings of Asiatic Society of Bengal, Vol. 26, 1930, pp. 541-588.

Mahalanobis, P. C., "On the Generalized Distance in Statistics," Proceedings of the National Institute of Sciences of India, Vol. 2, 1936, pp. 49-55.

Marks, Sidney and Olive Jean Dunn, "Discriminant Functions When Covariance Matrices are Unequal," Journal of the American Statistical Association, Vol. 69, No. 346, June 1974, pp. 555-559.

Massy, W. F., "Discriminant Analysis of Audience Characteristics," Journal of Advertising Research, Vol. 5, No. 1, March 1965, pp. 39-48.

McCabe, G. P., Jr., "Computations for Variable Selection in Discriminant Analysis," Technometrics, Vol. 17, 1975, pp. 103-109.

McLachlan, G. J., "Asymptotic Results for Discriminant Analysis When the Initial Samples are Misclassified," Technometrics, Vol. 14, 1972, pp. 415-422.

McLachlan, G. J., "Estimation of the Errors of Misclassification on the Criterion of Asymptotic Mean Square Error," Technometrics, Vol. 16, May 1974, pp. 255-256.

McLachlan, G. J., "An Asymptotic Unbiased Technique for Estimating the Error Rates in Discriminant Analysis," Biometrics, Vol. 30, June 1974, pp. 239-249.

McLachlan, G. J., "Confidence Intervals for the Conditional Probability of Misallocation in Discriminant Analysis," Biometrics, Vol. 31, March 1975, pp. 161-167.

McLachlan, G. J., "The Bias of the Apparent Error Rate in Discriminant Analysis," Biometrika, Vol. 63, No. 2, 1976, pp. 239-244.

Memon, A. Z. and M. Okamoto, "The Classification Statistic W in Covariate Discriminant Analysis," Annals of Mathematical Statistics, Vol. 41, 1970, pp. 1491-1499.

Memon, A. Z. and M. Okamoto, "Asymptotic Expansion of Distribution of Z-statistic in Discriminant Analysis," Journal of Multivariate Analysis, Vol. 1, 1971, pp. 294-307.

Meyer, Paul A. and Joward W. Pifer, "Prediction of Bank Failures," Journal of Finance, Vol. 25, No. 4, September 1970, pp. 853-868.

Michaelis, J., "Simulation Experiments with Multiple Group Linear and Quadratic Discriminant Analysis," Discriminant Analysis and Applications, T. Cacoullos, ed., 1973, pp. 225-239.

Miller, R. G., "Statistical Prediction by Discriminant Analysis," Meteorological Monographs, Vol. 4, No. 25, Boston, Mass: The American Meteorological Society, October 1962.

Montgomery, David B., "New Product Distribution: An Analysis of Supermarket Buyer Decisions," Journal of Marketing Research, Vol. 12, August 1975, pp. 255-264.

Moran, M. A., "The Performance of Fisher's Linear Discriminant Function With and Without Selection of Variables," Unpublished Ph.D. dissertation, University of Reading, 1974.

Morant, G. M., "A Preliminary Classification of European Races Based on Cranial Measurements," Biometrika, Series B, Vol. 20, 1928, pp. 301-375.

Morrison, Donald F., Multivariate Statistical Methods, 2nd ed., New York: McGraw-Hill, 1976.

Mosteller, F. and J. W. Tukey, "Data Analysis, Including Statistics," Revised Handbook of Social Psychology, Vol. 2, G. Lindzey and E. Aronson, eds., Reading, Mass.: Addison-Wesley, 1968, pp. 80-203.

Mosteller, F. and D. F. Wallace, "Inference in an Authorship Problem," Journal of the American Statistical Association, Vol. 58, No. 302, June 1963, pp. 275-309.

Myers, J. H. and E. W. Forgy, "The Development of Numerical Credit Evaluation Systems," Journal of the American Statistical Association, Vol. 58, September 1963, pp. 799-806.

Neyman, J. and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," Philosophical Transactions of the Royal Society of London, Series A, Vol. 231, 1933a, pp. 289-337.

Neyman, J. and E. S. Pearson, "On the Testing of Statistical Hypotheses in Relations to Probability A Priori," Proceedings of the Cambridge Philosophical Society, Vol. 9, 1933b, pp. 492-510.

Neyman, J. and E. S. Pearson, "Contributions to the Theory of Testing Statistical Hypothese, I," Statistical Research Memoirs, J. Neymand and E. S. Pearson, eds., London: Department of Statistics, University College London, Vol. 1, June 1936, pp. 1-37.

Nie, Norman H., C. Hadlai Hull, Jean G. Jenkins, Karin Steinbrenner and Dale H. Bent, eds., SPSS: Statistical Package for the Social Sciences, 2nd ed., New York: McGraw-Hill Book Company, 1975.

Norgaard, R. and C. Norgaard, "A Critical Examination of Share Repurchase," Financial Management, Vol. 3, No. 1, Spring 1974, pp. 44-50.

Norman, W. T., "Double-Split Cross-Validation: An Extension of Mosier's Design, Two Undesirable Alternatives, and Some Enigmatic Results," Journal of Applied Psychology, Vol. 49, 1965, pp. 348-357.

Okamoto, M., "Discrimination for Variance Matrices," Osaka Mathemathical Journal, Vol. 13, 1961, pp. 1-39.

Okamoto, M., "An Asymptotic Expansion for the Distribution of the Linear Discriminant Function," Annals of Mathematical Statistics, Vol. 34, 1963, pp. 1286-1301.

Okamoto, M., "Correction to: An Asymptotic Expansion for the Distribution of the Linear Discriminant Function," Annals of Mathematical Statistics, Vol. 39, 1968, pp. 1358-1380.

O'Neill, Terence J., "A Comparison of Logistic Regression and Maximum Likelihood Classification Methods," Technical Report No. 12, Department of Statistics, Stanford University, 1975.

O'Neill, Terence J., "Normal Discrimination with Unclassified Observations," Journal of the American Statistical Association, Vol. 73, 1978, pp. 821-826.

O'Neill, Terence J., "The General Distribution of the Error Rate of a Classification Procedure with Application to Logistic Regression Discrimination," Journal of the American Statistical Association, Vol. 75, No. 369, March 1980, pp. 154-160.

Orgler, Y. E., "A Credit Scoring Model for Consumer Loans," Journal of Money, Credit, and Banking, Vol. 2, No. 4, November 1970, pp. 435-445.

Ostlund, Lyman E., "Identifying Early Buyers," Journal of Advertising Research, Vol. 12, April 1972, pp. 25-30.

Ostlund, Lyman E., "Perceived Innovation Attributes as Predictors of Innovativeness," Journal of Consumer Research, Vol. 1, September 1974, pp. 23-29.

Pearson, K., "On the Coefficient of Racial Likeness," Biometrika, Vol. 18, 1926, pp. 105-117.

Penrose, L. S., "Some Notes on Discrimination," Annals of Eugenics, Vol. 13, 1947, pp. 228-237.

Pessemier, Edgar A., Philip C. Burger and Douglas J. Tigert, "Can New Product Buyers Be Identified?," Journal of Marketing Research, Vol. 4, November 1967, pp. 349-354.

Peterson, D. E., A Quantitative Framework for Financial Management, Homewood, Ill.: Richard E. Irwin, Inc., 1963.

Pinches, George E. and Kent A. Mingo, "A Multivariate Analysis of Industrial Bond Ratings," Journal of Finance, Vol. 28, No. 1, March 1973, pp. 1-18.

Pinches, George E. and Kent A. Mingo, "The Role of Subordination and Industrial Bond Ratings," Journal of Finance, Vol. 30, No. 1, March 1975, pp. 201-206.

Pogue, Thomas F. and Robert M. Soldofsky, "What's in a Bond Rating," Journal of Financial and Quantitative Analysis, Vol. 4, No. 3, June 1969, pp. 201-228.

Porebski, O. R., "Discriminatory and Canonical Analysis of Technical College Data," British Journal of Mathematical and Statistical Psychology, Vol. 19, 1966. pp. 215-236.

Press, S. J., "Estimating from Misclassified Data," Journal of the American Statistical Association, Vol. 63, 1968, p. 123-133.

Press, S. J., Applied Multivariate Analysis, New York: Holt, Rinehart and Winston, Inc., 1972.

Rao, C. R., "Tests with Discriminant Functions in Multivariate Analysis," Sankhya, Vol. 7, 1946, pp. 407-413.

Rao, C. R., "The Problem of Classification and Distance Between Two Populations," Nature, Vol. 159, 1947a, pp. 30-31.

Rao, C. R., "A Statistical Criterion to Determine the Group to Which an Individual Belongs," Nature, Vol. 160, 1947b, pp. 835-836.

Rao, C. R., "The Utilization of Multiple Measurements in Problems of Biological Classification," Journal of the Royal Statistical Society, Series B, Vol. 10, 1948, pp. 159-203.

Rao, C. R., "On the Distance Between Two Populations," Sankhya, Vol. 9, 1949a, pp. 246-248.

Rao, C. R., "On Some Problems Arising out of Discrimination with Multiple Characters," Sankhya, Vol. 9, 1949b, pp. 343-366.

Rao, C. R., "Statistical Inference Applied to Classificatory Problems," Sankhya, Vol. 10, 1950a, pp. 229-256.

Rao, C. R., "A Note on the Distribution of $D^2_{p+q} - D^2_p$ and Some Computational Aspects of $D^2$ Statistic and Discriminant Function," Sankhya, Vol. 10, 1950b, pp. 257-268.

Rao, C. R., "Statistical Inference Applied to Classificatory Problems. II. The Problem of Selecting Individuals for Various Duties in a Specified Ratio," Sankhya, Vol. 11, 1951a, pp. 107-116.

Rao, C. R., "Statistical Inference Applied to Classificatory Problems. III. The Discriminant Function Approach in the Classification of Time Series," Sankhya, Vol. 11, 1951b, pp. 257-272.

Rao, C. R., Advanced Statistical Methods in Biometric Research, New York: John Wiley & Sons, Inc., 1952.

Rao, C. R., "Statistical Inference Applied to Classificatory Problems. Discriminant Function for Genetic Differentiation and Selection," Sankhya, Vol. 12, 1953, pp. 229-246.

Rao, C. R., "Inference on Discriminant Function Coefficients," Essays in Probabilities, R. C. Bose, et al., eds., Chapel Hill: University of North Carolina and Statistical Publishing Society, 1970, pp. 587-602.

Rao, C. Radhakrishna, Linear Statistical Inference and Its Applications, 2nd ed., New York: John Wiley & Sons, 1973.

Rao, Tanniru R., "Is Brand Loyalty a Criterion for Marketing Segmentation: Discriminant Analysis," Decision Sciences, Vol. 4, July 1973, pp. 395-404.

Rencher, Alvin C. and Steven F. Larson, "Bias in Wilks' $\Lambda$ in Stepwise Discriminant Analysis," Technometrics, Vol. 22, No. 3, August 1980, pp. 349-356.

Riffenburgh, Robert H. and Charles W. Clunies-Ross, "Linear Discriminant Analysis," Pacific Science, Vol. 14, 1960, pp. 251-256.

Riordan, Edward A., Richard L. Oliver and James H. Donnelly, Jr., "The Unsold Prospect: Dyadic and Attitudinal Determinants," Journal of Marketing Research, Vol. 14, November 1977, pp. 530-537.

Robertson, Thomas S. and John N. Kennedy, "Prediction of Consumer Innovations: Application of Multiple Discriminant Analysis," Journal of Marketing Research, Vol. 5, February 1968, pp. 64-69.

Robertson, Thomas S. and John Rossiter, "Children & Commercial Persuasion: An Attribution Theory Analysis," Journal of Consumer Research, Vol. 1, June 1974, pp. 13-20.

Sayre, James W., "The Distributions of the Actual Error Rates in Linear Discriminant Analysis," Journal of the American Statistical Association, Vol. 75, No. 369, March 1980, pp. 201-205.

Schatzoff, M., "Exact Distributions of Wilks' Likelihood Ratio Criterion and Comparisons with Competitive Tests," Unpublished doctoral dissertation, Harvard University, 1964.

Schatzoff, M., "Exact Distributions of Wilks' Likelihood Ratio Criterion," Biometrika, Vol. 53, 1966, pp. 347-358.

Schick, Richard A. and James A. Verbrugge, "An Analysis of Bank Price-Earnings Ratios," Journal of Bank Research, Vol. 6, No. 2, Summer 1975, pp. 140-149.

Scott, E., "On the Financial Applications of Discriminant Analysis: Comment," Journal of Financial and Quantitative Analysis, Vol. 33, No. 1, March 1978, pp. 201-205.

Sedransk, N. and M. Okamoto, "Estimation of the Probabilities of Misclassification for a Linear Discriminant Function in the Univariate Normal Case," Annals of Institute of Statistical Mathematics, Vol. 23, 1971, pp. 419-436.

Sheth, Jagdish N., "Measurement of Multidimensional Brand Loyalty of a Consumer," Journal of Marketing Research, Vol. 7, August 1970, pp. 348-354.

Shuchman, Abe and Peter C. Riesz, "Correlates of Persuasibility: The Crest Case," Journal of Marketing Research, Vol. 12, February 1975, pp. 7-11.

Simkowitz, Michael and Robert J. Monroe, "A Discriminant Analysis Function for Conglomerate Targets," Southern Journal of Business, Vol. 6, November 1971, pp. 1-16.

Sinkey, Joseph F., Jr., "A Multivariate Statistical Analysis of the Characteristics of Problem Banks," Journal of Finance, Vol. 10, No. 1, March 1975, pp. 21-36.

Sitgreaves, R., "On the Distribution of Two Random Matrices Used in Classification Procedures," Annals of Mathematical Statistics, Vol. 23, 1952, pp. 263-270.

Sitgreaves, R., "Some Results on the Distribution of the W-Classification Statistic," Studies in Item Analysis and Prediction, H. Solomon, ed., Stanford: Stanford University Press, 1961, pp. 241-251.

Smith, Cedric A. B., "Some Examples of Discrimination," Annals of Eugenics, Vol. 13, 1947, pp. 272-282.

Smith, H. F., "A Discriminant Function for Plant Selection," Annals of Eugenics, Vol. 7, November 1936, pp. 240-250.

Smith, Keith V., "Classification of Investment Securities Using Multiple Discriminant Analysis," Institute Paper No. 101, Institute for Research in the Behavioral, Economic and Management Sciences, Purdue University, 1965.

Sorum, M. J., "Estimating the Conditional Probabilities of Misclassification," Technometrics, Vol. 13, 1971, pp. 333-343.

Sorum, M. J., "Estimating the Expected and the Optimal Probabilities of Misclassification," Technometrics, Vol. 14, 1972a, pp. 935-943.

Sorum, M. J., "Three Probabilities of Misclassification," Technometrics, Vol. 14, 1972b, pp. 309-316.

Specht, D. F., "Vectorcardiographic Diagnosis Using the Polynomial Discriminant Method of Pattern Recognition," IEEE Transaction on Bio-Medical Engineering, Vol. BME-14, 1967, pp. 90-95.

Streit, Franz, "Multivariate Linear Discrimination When the Covariance Matrix is Unknown," South African Statistical Journal, Vol. 13, 1979, pp. 93-101.

Stuhr, David and Robert Van Wicklen, "Rating the Financial Condition of Banks: A Statistical Approach to Aid Bank Supervision," Monthly Review, Federal Reserve Bank of New York, Vol. 56, No. 9, September 1974, pp. 233-238.

Teichroew, D. and R. Sitgreaves, "Computation of an Empirical Sampling Distribution for W-Classification Statistic," Studies in Item Analysis and Prediction, H. Solomon, ed., Stanford: Stanford University Press, 1961, pp. 252-275.

Tildesley, M. L., "A First Study of the Burmese Skull," Biometrika, Vol. 13, 1921, pp. 247-251.

Tintner, G., "Some Applications of Multivariate Analysis to Economic Data," Journal of the American Statistical Association, Vol. 41, 1946, pp. 472-500.

Toussaint, Godfried T., "Machine Recognition of Independent and Contextually Constrained Countour-Traced Handprinted Characters," Unpublished M.A.S. thesis, University British Columbia, Vancouver, Canada, December 1969.

Toussaint, Godfried T., "Bibliography on Estimation of Misclassification," IEEE Transactions on Information Theory, Vol. IT-20, No. 4, July 1974, pp. 472-479.

Toussaint, Godfried T. and R. W. Donaldson, "Algorithms for Recognizing Contour-Traced Handprinted Characters," IEEE Transactions on Computers, Vol. C-19, June 1970, pp. 541-546.

Toussaint, Godfried T. and P. M. Sharpe, "An Efficient Method for Estimating the Probability of Misclassification Applied to a Problem in Medical Diagnosis," Computers in Biology and Medicine, Vol. 4, 1975, pp. 269-278.

Uhl, Kenneth, Roma Andrus and Lance Poulsen, "How Are Laggards Different? An Empirical Inquiry," Journal of Marketing Research, Vol. 7, February 1970, pp. 51-54.

Urbakh, V. Yu, "Linear Discriminant Analysis: Loss of Discriminating Power When a Variate is Omitted," Biometrics, Vol. 27, 1971, pp. 531-534.

Utterback, James M., "Successful Industrial Innovations: A Multivariate Analysis," Decision Sciences, Vol. 6, January 1975, pp. 65-77.

Van Ness, John W., "On the Effects of Dimension in Discriminant Analysis for Unequal Covariance Populations," Technometrics, Vol. 21, No. 1, February 1979, pp. 119-127.

Van Ness, John W. and Cary Simpson, "On the Effects of Dimension in Discriminant Analysis," Technometrics, Vol. 18, No. 2, May 1976, pp. 175-187.

Von Mises, R., "On the Classification of Observation Data into Distinct Groups," Annals of Mathematical Statistics, Vol. 16, 1945, pp. 68-73.

Wahl, Patricia W. and Richard A. Kronmal, "Discriminant Functions When Covariances are Unequal and Sample Sizes are Moderate," Biometrics, Vol. 33, 1977, pp. 479-484.

Wald, A., "On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups," Annals of Mathematical Statistics, Vol. 15, 1944, pp. 145-162.

Wald, A., Statistical Decision Functions, New York: John Wiley & Sons, 1950.

Wald, A. and J. Wolfowitz, "Characterization of the Minimal Complete Class of Decision Functions When the Number of Decisions and Distributions is Finite," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1951, pp. 149-157.

Wallace, N. and R. M. W. Travers, "A Psychometric Sociological Study of a Group of Speciality Salesmen," Annals of Eugenics, Vol. 8, May 1938, pp. 266-302.

Walter, J. E., "A Discriminant Function for Earnings Price Ratios of Industrial Corporations," Review of Economics and Statistics, Vol. 41, February 1959, pp. 44-52.

Watanabe, S., P. Lambert, et al., "Evaluation of Selection of Variables in Pattern Recognition," Computer and Information Sciences - II, Julius T. Tou, ed., New York: Academic Press, Inc., 1967, pp. 91-122.

Weiner, S. M. and O. J. Dunn, "Elimination of Variates in Linear Discrimination Problems," Biometrics, Vol. 22, June 1966, pp. 268-275.

Welch, B. L., "Note on Discriminant Functions," Biometrika, Vol. 31, 1939, pp. 218-220.

Wiginton, John C., "A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior," Journal of Financial and Quantitative Analysis, Vol. 15, No. 3, September 1980, pp. 757-770.

Wilks, Samuel S., "Certain Generalizations in the Analysis of Variance," Biometrika, Vol. 24, 1932, pp. 471-494.

Williams, W. H. and M. L. Goodman, "A Statistical Grouping of Corporations by Their Financial Characteristics," Journal of Financial and Quantitative Analysis, Vol. 6, No. 4, September 1971, pp. 1095-1104.

Wind, Yoram, "Industrial Source Loyalty," Journal of Marketing Research, Vol. 7, November 1970, pp. 450-457.

Zumwalt, J. K., "A Multivariate Analysis of the Preferred Stock Ratings of Electric Utilities," Journal of the Midwest Finance Association, April, 1975,

# APPENDIX A

## Glossary of Terms

| Term | Definition |
|---|---|
| d | Coefficient of Proportionality, |
| g | Number of populations, |
| k | Number of variables, |
| $\Lambda$ | Wilk's Lambda measure, |
| log | Natural logarithm, |
| $n\underline{i}$ | Sample size from population i, |
| n | Training sample size for all populations ($n\underline{1}=n\underline{2}=\ldots=n\underline{g}=n$), |
| s | Skewness parameter of a covariance structure, |
| X | k X 1 random vector element, |
| $X\underline{i}$ | k X 1 random vector element from population i, |
| $\Pi\underline{i}$ | Population i, |
| $u\underline{i}$ | k X 1 mean vector in $\Pi\underline{i}$, |
| $\bar{X}\underline{i}$ | k X 1 sample mean vector in $\Pi\underline{i}$, |
| $A\underline{i}=\sum_{j=1}^{n}(X\underline{i}j-\bar{X}\underline{i})(X\underline{i}j-\bar{X}\underline{i})'$ | k X k cross-product matrix of X in $\Pi\underline{i}$, |
| $V\underline{i}$ | k X k covariance matrix in $\Pi\underline{i}$, |
| $S\underline{i}$ | k X k sample covariance matrix in $\Pi\underline{i}$, |
| $\rho$ | Coefficient of correlation, |
| $\Delta^2=(u\underline{1}-u\underline{2})'V^{-1}(u\underline{1}-u\underline{2})$ | Mahalanobis's $\Delta^2$ distance (parameters known), |
| $D^2=(\bar{X}\underline{1}-\bar{X}\underline{2})'S^{-1}(\bar{X}\underline{1}-\bar{X}\underline{2})$ | Mahalanobis's $D^2$ distance (parameters unknown), |

| | |
|---|---|
| $Dt(X)$ | Optimal discriminant function or score (parameters known), |
| $Ds(X)$ | Sample discriminant function or score (parameters unknown), |
| $Lt(X)$ | $\{X-1/2\ (u1+u2)\}'V^{-1}(u1-u2)=$ Optimal linear discriminant function or score (parameters known), |
| $Ls(X)$ | $\{X-1/2\ (\bar{X}1+\bar{X}2)\}'S^{-1}(\bar{X}1-\bar{X}2)=$ Sample linear discriminant function or score (parameters unknown), |
| $P(X)$ | Proportional quadratic discriminant function or score, |
| $Q(X)$ | Quadratic discriminant function or score, |
| $\beta i$ | Coefficient of a discriminant function, |
| $\nabla x$ | Vector partial differential operator with respect to X, |
| $\|\nabla x\|^2$ | Sum of squared elements in vector $\nabla x$, |
| $pi$ | A priori probability that an observation comes from $\Pi i$, |
| $Ri$ | The region of the random vectors assigned to $\Pi i$, |
| $f(X\|i)$ | Density function of X in $\Pi i$, |
| $r(j)$ | Expected cost of misclassifying an observation from $\Pi j$, |
| $C(i\|j)$ | Cost of misclassifying an observation from $\Pi j$ into $\Pi i$, |
| $P(j)$ | Probability of misclassifying an observation from $\Pi j$, |
| $P(i\|j)$ | Probability of misclassifying an observation from $\Pi j$ into $\Pi i$, |
| $P1(Dt)$ | Optimal error rate, |

| | |
|---|---|
| $P\underline{2}(D\underline{s})$ | Estimated actual error rate, |
| $P\underline{3}(D\underline{s})$ | Actual error rate, |
| $P\underline{4}(D\underline{s})$ | Expected actual error rate, |
| $P\underline{5}(D\underline{s})$ | Expected value of estimated actual error rate, |
| $P\underline{6}(D\underline{s})$ | Apparent error rate, |
| $N(u,V)$ | Cumulative normal distribution function with mean vector u and covariance matrix V, |
| $\Phi$ | Cumulative standard normal distribution function, |
| $\chi^2(k)$ | Chi-square distribution function with k degrees of freedom, |
| $\chi'^2(k)$ | Non-central chi-square distribution function with k degrees of freedom, |
| $W(k,V)$ | Wishart distribution function with k degrees of freedom and covariance matrix V, |

# APPENDIX B

## The Variance Matrix of the Coefficients of the PQDF with Proportional Covariance Structure

Consider the two-variate normal case, let $u\underline{x}(1)$ and $u\underline{x}(2)$ be the means of variable 1 and 2 in population 1, $u\underline{y}(1)$ and $u\underline{y}(2)$ be the two means in population 2, d be the coefficient of proportionality, $V = \begin{bmatrix} V(11), & V(12) \\ V(21), & V(22) \end{bmatrix}$, and $W = \mathrm{Det}(V) = V(11)V(22) - V(12)V(21)$. The likelihood function of the sample with known parameters is given by

$$(B.1)\quad L(u\underline{x}(1), u\underline{x}(2), u\underline{y}(1), u\underline{y}(2), V(11), V(12), V(22), d)$$

$$= p\underline{1}^{n\underline{1}} (2\pi)^{-2n\underline{1}/2} |V|^{-n\underline{1}/2}$$

$$\cdot e^{-1/2 \{ \sum_{i=1}^{n\underline{1}} (X\underline{i} - u\underline{1})' V^{-1} (X\underline{i} - u\underline{1}) \}}$$

$$\cdot p\underline{2}^{n\underline{2}} (2\pi)^{-2n\underline{2}/2} |dV|^{-n\underline{2}/2}$$

$$\cdot e^{-1/2 \{ \sum_{j=1}^{n\underline{2}} (X\underline{j} - u\underline{2})' (dV)^{-1} (X\underline{j} - u\underline{2}) \}}$$

$$= p\underline{1}^{n\underline{1}} p\underline{2}^{n\underline{2}} (2\pi)^{-(n\underline{1}+n\underline{2})} |V|^{-(n\underline{1}+n\underline{2})/2} d^{-n\underline{2}}$$

$$\cdot e^{-1/2 \{ \sum_{i=1}^{n\underline{1}} (X\underline{i} - u\underline{1})' V^{-1} (X\underline{i} - u\underline{1}) \}}$$

$$\cdot e^{-1/2 \{ 1/d \sum_{j=1}^{n\underline{2}} (X\underline{j} - u\underline{2})' V^{-1} (X\underline{j} - u\underline{2}) \} \cdot}$$

Let $V^{-1} = \begin{bmatrix} W(11), & W(12) \\ W(21), & W(22) \end{bmatrix}$, then

$$(B.2)\quad L(u\underline{x}(1), u\underline{x}(2), u\underline{y}(1), u\underline{y}(2), V(11), V(12), V(22), d)$$

$$= p\underline{1}^{n\underline{1}}\, p\underline{2}^{n\underline{2}}\, (2\pi)^{-(n\underline{1}+n\underline{2})}\, d^{-n\underline{2}}$$

$$\cdot (W)^{-(n\underline{1}+n\underline{2})/2}$$

$$\cdot e^{-1/2 \left\{ \sum_{i=1}^{n\underline{1}} \begin{bmatrix} X\underline{1}-u\underline{x}(1) \\ X\underline{2}-u\underline{x}(2) \end{bmatrix}' \begin{bmatrix} W(11) & W(12) \\ W(21) & W(22) \end{bmatrix} \begin{bmatrix} X\underline{1}-u\underline{x}(1) \\ X\underline{2}-u\underline{x}(2) \end{bmatrix}\right.}$$

$$\left. + \left\{ 1/d \sum_{j=1}^{n\underline{2}} \begin{bmatrix} Y\underline{2}-u\underline{y}(2) \\ Y\underline{2}-u\underline{y}(2) \end{bmatrix}' \begin{bmatrix} W(11) & W(12) \\ W(21) & W(22) \end{bmatrix} \begin{bmatrix} Y\underline{2}-u\underline{y}(2) \\ Y\underline{2}-u\underline{y}(2) \end{bmatrix} \right\}.$$

The log likelihood function can be written as

$$(B.3)\quad \log L = n\underline{1} \log p\underline{1} + n\underline{2} \log p\underline{2}$$

$$- (n\underline{1}+n\underline{2}) \log 2\pi - n\underline{2} \log d$$

$$- (n\underline{1}+n\underline{2})/2 \log (W)$$

$$- \frac{1}{2} \sum_{i=1}^{n\underline{1}} \{ (X\underline{1}-u\underline{x}(1))^2 W(11) + (X\underline{2}-u\underline{x}(2))^2 W(22)$$

$$+ 2(X\underline{1}-u\underline{x}(1))(X\underline{2}-u\underline{x}(2)) W(12) \}$$

$$- \frac{1}{2d} \sum_{j=1}^{n\underline{2}} \{ (Y\underline{1}-u\underline{y}(1))^2 W(11) + (Y\underline{2}-u\underline{y}(2))^2 W(22)$$

$$+ 2(Y\underline{1}-u\underline{y}(1))(Y\underline{2}-u\underline{y}(2)) W(12) \}.$$

Since

$$W(11) = V(22)\, W^{-1},$$

$$W(22) = V(11)W^{-1},$$

$$W(12) = W(21) = -2\,V(12)W^{-1}$$

$$= -2\,V(21)W^{-1},$$

therefore,

$$\text{(B.4)}\quad \log L = n\underline{1}\log p\underline{1} + n\underline{2}\log p\underline{2}$$

$$- (n\underline{1}+n\underline{2})\log 2\pi - n\underline{2}\log d$$

$$- (n\underline{1}+n\underline{2})/2\log W$$

$$- \frac{1}{2}\sum_{i=1}^{n1}\{(X\underline{1}-u\underline{x}(1))^2V(22)W^{-1}$$

$$+ (X\underline{2}-u\underline{x}(2))^2V(11)W^{-1}$$

$$+ 2(X\underline{1}-u\underline{x}(1))(X\underline{2}-u\underline{x}(2))$$

$$\cdot(-2)V(12)W^{-1}\}$$

$$- \frac{1}{2d}\sum_{j=1}^{n2}\{(Y\underline{1}-u\underline{y}(1))^2V(22)W^{-1}$$

$$+ (Y\underline{2}-u\underline{y}(2))^2V(11)W^{-1}$$

$$+ 2(Y\underline{1}-u\underline{y}(1))(Y\underline{2}-u\underline{y}(2))$$

$$\cdot(-2)V(12)W^{-1}\}\ .$$

After partially differentiating the log likelihood function, one has the following results which had been simplified by using the setup conditions in the PCS, i.e., $N((\triangle\sqrt{(1+d)/2},0)',I)$ in population 1 and $N((0,0)',dI)$ in population 2.

$$\text{(B.5)}\quad \frac{\partial^2 \log L}{\partial p\underline{1}^2} = -\frac{n\underline{1}}{p\underline{1}^2} - \frac{n\underline{2}}{(1-p\underline{1})^2} = -\frac{n\underline{1}p\underline{2}^2 + n\underline{2}p\underline{1}^2}{p\underline{1}^2p\underline{2}^2},$$

$$(B.6)\quad \frac{\partial^2 \log L}{\partial u_{\underline{x}}(1)^2} = -\, n_{\underline{1}}\, W(11)\ ,$$

$$(B.7)\quad \frac{\partial^2 \log L}{\partial u_{\underline{x}}(2)^2} = -\, n_{\underline{1}}\, W(22)\ ,$$

$$(B.8)\quad \frac{\partial^2 \log L}{\partial u_{\underline{y}}(1)^2} = -\, \frac{n_{\underline{2}}}{d}\, W(11)\ ,$$

$$(B.9)\quad \frac{\partial^2 \log L}{\partial u_{\underline{y}}(2)^2} = -\, \frac{n_{\underline{2}}}{d}\, W(22)\ ,$$

$$(B.10)\quad \frac{\partial^2 \log L}{\partial V(11)^2} = \frac{(n_{\underline{1}}+n_{\underline{1}})\, V(22)^2}{2\, W^2}$$

$$- \{ \sum_{i=1}^{n_{\underline{1}}} (X_{\underline{1}}-u_{\underline{x}}(1))^2 + 1/d \sum_{j=1}^{n_{\underline{2}}} (Y_{\underline{1}}-u_{\underline{y}}(1))^2 \}$$

$$\cdot V(22)^3 W^{-3}$$

$$+2\{ \sum_{i=1}^{n_{\underline{1}}} (X_{\underline{1}}-u_{\underline{x}}(1))(X_{\underline{2}}-u_{\underline{x}}(2)) + \frac{1}{d} \sum_{j=1}^{n_{\underline{2}}} (Y_{\underline{1}}-u_{\underline{y}}(1))(Y_{\underline{2}}-u_{\underline{y}}(2)) \}$$

$$\cdot V(12)\, V(22)^2 W^{-3}$$

$$- \{ \sum_{i=1}^{n_{\underline{1}}} (X_{\underline{2}}-u_{\underline{x}}(2))^2 + 1/d \sum_{j=1}^{n_{\underline{2}}} (Y_{\underline{2}}-u_{\underline{y}}(2))^2 \}$$

$$\cdot V(12)\, V(21)\, V(22)\, W^{-3}\ ,$$

$$(B.11)\quad \frac{\partial^2 \log L}{\partial V(11)\, \partial V(12)} = \frac{-(n_{\underline{1}}+n_{\underline{1}})\, V(12)\, V(22)}{W^2}$$

$$+2\{ \sum_{i=1}^{n_{\underline{1}}} (X_{\underline{1}}-u_{\underline{x}}(1))^2 + 1/d \sum_{j=1}^{n_{\underline{2}}} (Y_{\underline{1}}-u_{\underline{y}}(1))^2 \}$$

$$\bullet V(12)\,V(22)^2 W^{-3}$$

$$- \{\sum_{i=1}^{n1} (X\underline{1}-u\underline{x}(1))\,(X\underline{2}-u\underline{x}(2)) + \frac{1}{d}\sum_{j=1}^{n2} (Y\underline{1}-u\underline{y}(1))\,(Y\underline{2}-u\underline{y}(2))\}$$

$$\bullet (3V(12)^2 + V(11)\,V(22))\,V(22)\,W^{-3}$$

$$+ \{\sum_{i=1}^{n1} (X\underline{2}-u\underline{x}(2))^2 + 1/d \sum_{j=1}^{n2} (Y\underline{2}-u\underline{y}(2))^2\}$$

$$\bullet V(12)\,(V(11)\,V(22)+V(12)^2)\,W^{-3}\ ,$$

$$(B.12)\quad \frac{\partial^2 \log L}{\partial V(11)\ \partial V(22)} = \frac{(n\underline{1}+n\underline{1})\,V(12)\,V(21)}{2\ W^2}$$

$$- \{\sum_{i=1}^{n1} (X\underline{1}-u\underline{x}(1))^2 + 1/d \sum_{j=1}^{n2} (Y\underline{1}-u\underline{y}(1))^2\}$$

$$\bullet V(22)\,V(12)^2 W^{-3}$$

$$+ \{\sum_{i=1}^{n1} (X\underline{1}-u\underline{x}(1))\,(X\underline{2}-u\underline{x}(2)) + \frac{1}{d}\sum_{j=1}^{n2} (Y\underline{1}-u\underline{y}(1))\,(Y\underline{2}-u\underline{y}(2))\}$$

$$\bullet V(12)\,\{V(11)\,V(22)+V(12)^2)\,W^{-3}$$

$$- \{\sum_{i=1}^{n1} (X\underline{2}-u\underline{x}(2))^2 + 1/d \sum_{j=1}^{n2} (Y\underline{2}-u\underline{y}(2))^2\}$$

$$\bullet V(11)\,V(12)\,V(21)\,W^{-3}\ ,$$

$$(B.13)\quad \frac{\partial^2 \log L}{\partial V(12)^2} = \frac{(n\underline{1}+n\underline{1})\,(V(11)\,V(22)+V(12)\,V(21))}{W^2}$$

$$- \{\sum_{i=1}^{n1} (X\underline{1}-u\underline{x}(1))^2 + 1/d \sum_{j=1}^{n2} (Y\underline{1}-u\underline{y}(1))^2\}$$

$$\bullet V(22)\,(V(11)\,V(22)+3V(12)^2)\,W^{-3}$$

$$+2\{\sum_{i=1}^{n1}(X\underline{1}-u\underline{x}(1))(X\underline{2}-u\underline{x}(2))+\frac{1}{d}\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))(Y\underline{2}-u\underline{y}(2))\}$$

$$\cdot V(12)(V(12)^2+3V(11)V(22))W^{-3}$$

$$-\{\sum_{i=1}^{n1}(X\underline{2}-u\underline{x}(2))^2+1/d\sum_{j=1}^{n2}(Y\underline{2}-u\underline{y}(2))^2\}$$

$$\cdot V(11)(V(11)V(22)+3V(12)^2)W^{-3},$$

$$(B.14)\quad \frac{\partial^2 \log L}{\partial V(12)\,\partial V(22)} = \frac{-(n\underline{1}+n\underline{1})V(11)V(12)}{W^2}$$

$$+\{\sum_{i=1}^{n1}(X\underline{1}-u\underline{x}(1))^2+1/d\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))^2\}$$

$$\cdot V(12)(V(11)V(22)+V(12)^2)W^{-3}$$

$$-\{\sum_{i=1}^{n1}(X\underline{1}-u\underline{x}(1))(X\underline{2}-u\underline{x}(2))+\frac{1}{d}\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))(Y\underline{2}-u\underline{y}(2))\}$$

$$\cdot(3V(12)^2+V(11)V(22))V(11)W^{-3}$$

$$+2\{\sum_{i=1}^{n1}(X\underline{2}-u\underline{x}(2))^2+1/d\sum_{j=1}^{n2}(Y\underline{2}-u\underline{y}(2))^2\}$$

$$\cdot V(11)^2V(12)W^{-3},$$

$$(B.15)\quad \frac{\partial^2 \log L}{\partial V(22)^2} = \frac{(n\underline{1}+n\underline{1})V(11)^2}{2\,W^2}$$

$$-\{\sum_{i=1}^{n1}(X\underline{1}-u\underline{x}(1))^2+1/d\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))^2\}$$

$$\cdot V(11)V(12)^2W^{-3}$$

$$+2\{\sum_{i=1}^{n1}(X\underline{1}-u\underline{x}(1))(X\underline{2}-u\underline{x}(2))+\frac{1}{d}\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))(Y\underline{2}-u\underline{y}(2))\}$$

$$\cdot V(11)^2V(12)W^{-3}$$

$$-\{\sum_{i=1}^{n1}(X\underline{2}-u\underline{x}(2))^2+1/d\sum_{j=1}^{n2}(Y\underline{2}-u\underline{y}(2))^2\}$$

$$\cdot V(11)^3W^{-3}\ ,$$

$$(B.16)\quad \frac{\partial^2 \log I}{\partial V(11)\ \partial d}=\frac{1}{2d^2W^2}\cdot$$

$$\{-\{\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))^2\}\cdot V(22)^2$$

$$+2\{\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))(Y\underline{2}-u\underline{y}(2))\}\cdot V(12)V(22)$$

$$-\{\sum_{j=1}^{n2}(Y\underline{2}-u\underline{y}(2))^2\}\cdot V(12)V(21)\}\ ,$$

$$(B.17)\quad \frac{\partial^2 \log L}{\partial V(12)\ \partial d}=\frac{1}{d^2W^2}\cdot$$

$$\{\{\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))^2\}\cdot V(12)V(22)$$

$$-\{\sum_{j=1}^{n2}(Y\underline{1}-u\underline{y}(1))(Y\underline{2}-u\underline{y}(2))\}$$

$$\cdot(V(11)V(22)+V(12)^2)$$

$$+\{\sum_{j=1}^{n2}(Y\underline{2}-u\underline{y}(2))^2\}\cdot V(11)V(12)\}\ ,$$

$$(B.18)\quad \frac{\partial^2 \log L}{\partial V(22)\ \partial d} = \frac{1}{2d^2\ W^2} \cdot$$

$$\{ - \{ \sum_{j=1}^{n2} (Y\underline{1}-u\underline{y}(1))^2 \} \cdot V(12)\ V(21)$$

$$+2 \{ \sum_{j=1}^{n2} (Y\underline{1}-u\underline{y}(1))\ (Y\underline{2}-u\underline{y}(2)) \} \cdot V(11)\ V(12)$$

$$- \{ \sum_{j=1}^{n2} (Y\underline{2}-u\underline{y}(2))^2 \} \cdot V(11)^2 \} ,$$

$$(B.19)\quad \frac{\partial^2 \log L}{\partial d^2} = \frac{n\underline{2}}{d^2} - \frac{1}{d^3\ W} \cdot$$

$$\{ \{ \sum_{j=1}^{n2} (Y\underline{1}-u\underline{y}(1))^2 \} \cdot V(22)$$

$$-2 \{ \sum_{j=1}^{n2} (Y\underline{1}-u\underline{y}(1))\ (Y\underline{2}-u\underline{y}(2)) \} \cdot V(12)$$

$$+ \{ \sum_{j=1}^{n2} (Y\underline{2}-u\underline{y}(2))^2 \} \cdot V(11) \} .$$

The results of the above second partial differentiations are then evaluated at the point where the estimated parameters equal their corresponding true parameters. One thus has

$$(B.20)\quad \left[\frac{\partial^2 \log L}{\partial p\underline{1}^2}\right]_{\hat{\theta}=\theta} = - \frac{n\underline{1}p\underline{2}^2 + n\underline{2}p\underline{1}^2}{p\underline{1}^2 p\underline{2}^2} ,$$

$$(B.21)\quad \left[\frac{\partial^2 \log L}{\partial u_{\underline{x}}(1)^2}\right]_{\hat{\theta}=\theta} = -\, n_{\underline{1}}\, W(11)\ ,$$

$$(B.22)\quad \left[\frac{\partial^2 \log L}{\partial u_{\underline{x}}(2)^2}\right]_{\hat{\theta}=\theta} = -\, n_{\underline{1}}\, W(22)\ ,$$

$$(B.23)\quad \left[\frac{\partial^2 \log L}{\partial u_{\underline{y}}(1)^2}\right]_{\hat{\theta}=\theta} = -\,\frac{n_{\underline{2}}}{d}\, W(11)\ ,$$

$$(B.24)\quad \left[\frac{\partial^2 \log L}{\partial u_{\underline{y}}(2)^2}\right]_{\hat{\theta}=\theta} = -\,\frac{n_{\underline{2}}}{d}\, W(22)\ ,$$

$$(B.25)\quad \left[\frac{\partial^2 \log L}{\partial V(11)^2}\right]_{\hat{\theta}=\theta} = \frac{(n_{\underline{1}}+n_{\underline{1}})\, V(22)^2}{2\, W^2}$$

$$-\ \{(n_{\underline{1}}+n_{\underline{2}})\, V(11)\}\ \cdot V(22)^3 W^{-3}$$

$$+2\,\{(n_{\underline{1}}+n_{\underline{2}})\, V(12)\}\ \cdot V(12)\, V(22)^2 W^{-3}$$

$$-\ \{(n_{\underline{1}}+n_{\underline{2}})\, V(22)\}\ \cdot V(12)\, V(21)\, V(22)\, W^{-3}\ ,$$

$$(B.26)\quad \left[\frac{\partial^2 \log L}{\partial V(11)\ \partial V(12)}\right]_{\hat{\theta}=\theta} = \frac{-(n_{\underline{1}}+n_{\underline{1}})\, V(12)\, V(22)}{W^2}$$

$$+2\,\{(n_{\underline{1}}+n_{\underline{2}})\, V(11)\}\ \cdot V(12)\, V(22)^2 W^{-3}$$

$$-\ \{(n_{\underline{1}}+n_{\underline{2}})\, V(12)\}\ \cdot (3V(12)^2+V(11)\, V(22))\, V(22)\, W^{-3}$$

$$+\ \{(n_{\underline{1}}+n_{\underline{2}})\, V(22)\}\ \cdot V(12)\ (V(11)\, V(22)+V(12)^2)\, W^{-3}\ ,$$

$$(B.27)\quad \left[\frac{\partial^2 \log L}{\partial V(11)\ \partial V(22)}\right]_{\hat{\theta}=\theta} = \frac{(n_{\underline{1}}+n_{\underline{1}})\, V(12)\, V(21)}{2\, W^2}$$

$$-\ \{(n_{\underline{1}}+n_{\underline{2}})\, V(11)\}\ \cdot V(22)\, V(12)^2 W^{-3}$$

$$+ \{(n\underline{1}+n\underline{2})V(12)\} \cdot V(12)\,(V(11)V(22)+V(12)^2)W^{-3}$$

$$- \{(n\underline{1}+n\underline{2})V(22)\} \cdot V(11)V(12)V(21)W^{-3} ,$$

$$\text{(B.28)} \quad \left[\frac{\partial^2 \log L}{\partial V(12)^2}\right]_{\hat{\theta}=\theta} = \frac{(n\underline{1}+n\underline{1})\,(V(11)V(22)+V(12)V(21))}{W^2}$$

$$- \{(n\underline{1}+n\underline{2})V(11)\} \cdot V(22)\,(V(11)V(22)+3V(12)^2)W^{-3}$$

$$+2\{(n\underline{1}+n\underline{2})V(12)\} \cdot V(12)\,(V(12)^2+3V(11)V(22))W^{-3}$$

$$- \{(n\underline{1}+n\underline{2})V(22)\} \cdot V(11)\,(V(11)V(22)+3V(12)^2)W^{-3} ,$$

$$\text{(B.29)} \quad \left[\frac{\partial^2 \log L}{\partial V(12)\ \partial V(22)}\right]_{\hat{\theta}=\theta} = \frac{-(n\underline{1}+n\underline{1})V(11)V(12)}{W^2}$$

$$+ \{(n\underline{1}+n\underline{2})V(11)\} \cdot V(12)\,(V(11)V(22)+V(12)^2)W^{-3}$$

$$- \{(n\underline{1}+n\underline{2})V(12)\} \cdot (3V(12)^2+V(11)V(22))V(11)W^{-3}$$

$$+2\{(n\underline{1}+n\underline{2})V(22)\} \cdot V(11)^2V(12)W^{-3} ,$$

$$\text{(B.30)} \quad \left[\frac{\partial^2 \log L}{\partial V(22)^2}\right]_{\hat{\theta}=\theta} = \frac{(n\underline{1}+n\underline{1})V(11)^2}{2\ W^2}$$

$$- \{(n\underline{1}+n\underline{2})V(11)\} \cdot V(11)V(12)^2W^{-3}$$

$$+2\{(n\underline{1}+n\underline{2})V(12)\} \cdot V(11)^2V(12)W^{-3}$$

$$- \{(n\underline{1}+n\underline{2})V(22)\} \cdot V(11)^3W^{-3} ,$$

$$\text{(B.31)} \quad \left[\frac{\partial^2 \log L}{\partial V(11)\ \partial d}\right]_{\hat{\theta}=\theta} = \frac{1}{2d^2\ W^2} \cdot$$

$$\{ - \{n\underline{2}\ dV(11)\} \cdot V(22)^2$$

$$+2\ \{n\underline{2}\ dV(12)\}\ \bullet V(12)V(22)$$

$$-\ \{n\underline{2}\ dV(22)\}\ \bullet V(12)V(21)\}\ ,$$

$$(B.32)\quad \left[\frac{\partial^2 \log L}{\partial V(12)\ \partial d}\right]_{\hat{\theta}=\theta} = \frac{1}{d^2\ W^2}\ \bullet$$

$$\{\quad \{n\underline{2}\ dV(11)\}\ \bullet V(12)V(22)$$

$$-\ \{n\underline{2}\ dV(12)\}\ \bullet (V(11)V(22)+V(12)^2)$$

$$+\ \{n\underline{2}\ dV(22)\}\ \bullet V(11)V(12)\quad\}\ ,$$

$$(B.33)\quad \left[\frac{\partial^2 \log L}{\partial V(22)\ \partial d}\right]_{\hat{\theta}=\theta} = \frac{1}{2d^2\ W^2}\ \bullet$$

$$\{\ -\ \{n\underline{2}\ dV(11)\}\ \bullet V(12)V(21)$$

$$+2\ \{n\underline{2}\ dV(12)\}\ \bullet V(11)V(12)$$

$$-\ \{n\underline{2}\ dV(22)\}\ \bullet V(11)^2\quad\}\ ,$$

$$(B.34)\quad \left[\frac{\partial^2 \log L}{\partial d^2}\right]_{\hat{\theta}=\theta} = \frac{n\underline{2}}{d^2} - \frac{1}{d^3\ W}\ \bullet$$

$$\{\quad \{n\underline{2}\ dV(11)\}\ \bullet V(22)$$

$$-2\ \{n\underline{2}\ dV(12)\}\ \bullet V(12)$$

$$+\ \{n\underline{2}\ dV(22)\}\ \bullet V(11)\quad\}\ .$$

Substituting $u\underline{x}(1)=\triangle\sqrt{(1+d)/2}$, $u\underline{x}(2)=u\underline{y}(1)=u\underline{y}(2)=0$, $V(11)=V(22)=1$, $V(12)=V(21)=0$, and $W=1$ into Equations (B.20) through (B.34) yields the following results.

$$\left[\frac{\partial^2 \log L}{\partial p\underline{1}^2}\right]_{\hat{\theta}=\theta} = -\frac{n\underline{1}p\underline{2}^2 + n\underline{2}p\underline{1}^2}{p\underline{1}^2 p\underline{2}^2} ,$$

$$\left[\frac{\partial^2 \log L}{\partial u\underline{x}(1)^2}\right]_{\hat{\theta}=\theta} = -n\underline{1} ,$$

$$\left[\frac{\partial^2 \log L}{\partial u\underline{x}(2)^2}\right]_{\hat{\theta}=\theta} = -n\underline{1} ,$$

$$\left[\frac{\partial^2 \log L}{\partial u\underline{y}(1)^2}\right]_{\hat{\theta}=\theta} = -\frac{n\underline{2}}{d} ,$$

$$\left[\frac{\partial^2 \log L}{\partial u\underline{y}(2)^2}\right]_{\hat{\theta}=\theta} = -\frac{n\underline{2}}{d} ,$$

$$\left[\frac{\partial^2 \log L}{\partial v(11)^2}\right]_{\hat{\theta}=\theta} = -\frac{(n\underline{1}+n\underline{2})}{2} ,$$

$$\left[\frac{\partial^2 \log L}{\partial v(11)\ \partial v(12)}\right]_{\hat{\theta}=\theta} = 0 ,$$

$$\left[\frac{\partial^2 \log L}{\partial v(11)\ \partial v(22)}\right]_{\hat{\theta}=\theta} = 0 ,$$

$$\left[\frac{\partial^2 \log L}{\partial v(12)^2}\right]_{\hat{\theta}=\theta} = -(n\underline{1}+n\underline{2}) ,$$

$$\left[\frac{\partial^2 \log L}{\partial v(12)\ \partial v(22)}\right]_{\hat{\theta}=\theta} = 0 ,$$

$$\left[\frac{\partial^2 \log L}{\partial v(22)^2}\right]_{\hat{\theta}=\theta} = \frac{-(n\underline{1}+n\underline{2})}{2} ,$$

$$\left[\frac{\partial^2 \log L}{\partial V(11)\ \partial d}\right]_{\hat{\theta}=\theta} = \frac{-n_2}{2d},$$

$$\left[\frac{\partial^2 \log L}{\partial V(12)\ \partial d}\right]_{\hat{\theta}=\theta} = 0,$$

$$\left[\frac{\partial^2 \log L}{\partial V(22)\ \partial d}\right]_{\hat{\theta}=\theta} = \frac{-n_2}{2d},$$

$$\left[\frac{\partial^2 \log L}{\partial d^2}\right]_{\hat{\theta}=\theta} = \frac{-n_2}{d^2},$$

Following Layard's (1972) Theorem (1.3) and Efron's (1975) Lemma (2), the joint limiting covariance matrix of $(\hat{p}_1, \hat{u}_x(1), \hat{u}_x(2), \hat{u}_y(1), \hat{u}_y(2), \hat{V}(11), \hat{V}(12), \hat{V}(22), \hat{d})$ can be found by multiplying the above equations with $-(n_1+n_2)^{-1}$, evaluating them at $n_1/(n_1+n_2)=p_1$ and $n_2/(n_1+n_2)=p_2$, and then taking inverse of the matrix. The resulting covariance matrix of the estimated parameters is given below:

(B.35) $V(\theta)=V(p_1, u_x(1), u_x(2), u_y(1), u_y(2), V(11), V(12), V(22), d)$

$$= \begin{bmatrix}
p_1p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/p_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & d/p_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & d/p_2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & (1+p_1)/p_1 & 0 & p_2/p_1 & -d/p_1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p_2/p_1 & 0 & (1+p_1)/p_1 & -d/p_1 \\
0 & 0 & 0 & 0 & 0 & -d/p_1 & 0 & -d/p_1 & d^2/p_1p_2
\end{bmatrix}.$$

The procedure of finding M matrix in Equation (3.8) of Efron's (1975) Lemma (2) is demonstrated below.

The optimal discriminant function with two-variate PCS is given by

$$(B.36)\quad Dt(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 ,$$

where

$$\beta_0 = \log p_1 - \log p_2 + \log d$$

$$-\frac{1}{2}\{(u_x^2(1) - \frac{1}{d}u_y^2(1))V(22)W^{-1} + (u_x^2(2) - \frac{1}{d}u_y^2(2))V(11)W^{-1}$$

$$-2(u_x(1)u_x(2) - \frac{1}{d}u_y(1)u_y(2))V(12)W^{-1}\},$$

$$\beta_1 = \{u_x(1) - \frac{1}{d}u_y(1)\}V(22)W^{-1} - \{u_x(2) - \frac{1}{d}u_y(2)\}V(12)W^{-1},$$

$$\beta_2 = \{u_x(2) - \frac{1}{d}u_y(2)\}V(11)W^{-1} - \{u_x(1) - \frac{1}{d}u_y(1)\}V(21)W^{-1},$$

$$\beta_3 = (1 - \frac{1}{d})V(12)W^{-1},$$

$$\beta_4 = -\frac{1}{2}(1 - \frac{1}{d})V(22)W^{-1},$$

$$\beta_5 = -\frac{1}{2}(1 - \frac{1}{d})V(11)W^{-1},$$

where $W = V(11)V(22) - V(12)V(21)$.

After partially differentiating $\beta\underline{0}$, $\beta\underline{1}$, ..., $\beta\underline{5}$ with respect to $p\underline{1}$, $u\underline{x}(1)$, $u\underline{x}(2)$, $u\underline{y}(1)$, $u\underline{y}(2)$, $V(11)$, $V(12)$, $V(22)$, and $d$, one has

$$(B.37)\quad M = \begin{bmatrix} \frac{\partial\beta\underline{0}}{\partial p\underline{1}} & \frac{\partial\beta\underline{0}}{\partial u\underline{x}(1)} & \frac{\partial\beta\underline{0}}{\partial u\underline{x}(2)} & \cdots & \frac{\partial\beta\underline{0}}{\partial V(11)} & \frac{\partial\beta\underline{0}}{\partial V(12)} & \frac{\partial\beta\underline{0}}{\partial V(22)} & \frac{\partial\beta\underline{0}}{\partial d} \\ \frac{\partial\beta\underline{1}}{\partial p\underline{1}} & \frac{\partial\beta\underline{1}}{\partial u\underline{x}(1)} & \frac{\partial\beta\underline{1}}{\partial u\underline{x}(2)} & \cdots & \frac{\partial\beta\underline{1}}{\partial V(11)} & \frac{\partial\beta\underline{1}}{\partial V(12)} & \frac{\partial\beta\underline{1}}{\partial V(22)} & \frac{\partial\beta\underline{1}}{\partial d} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial\beta\underline{5}}{\partial p\underline{1}} & \frac{\partial\beta\underline{5}}{\partial u\underline{x}(1)} & \frac{\partial\beta\underline{5}}{\partial u\underline{x}(2)} & \cdots & \frac{\partial\beta\underline{5}}{\partial V(11)} & \frac{\partial\beta\underline{5}}{\partial V(12)} & \frac{\partial\beta\underline{5}}{\partial V(22)} & \frac{\partial\beta\underline{5}}{\partial d} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{p\underline{1}p\underline{2}} & \frac{u\underline{x}(1)V(22)}{-W} & \frac{u\underline{x}(2)V(11)}{-W} & \frac{u\underline{y}(1)V(22)}{dW} & \frac{u\underline{y}(2)V(11)}{dW} \\ 0 & V(22)W^{-1} & -V(12)W^{-1} & -V(22)W^{-1}/d & V(12)W^{-1}/d \\ 0 & -V(12)W^{-1} & V(11)W^{-1} & V(12)W^{-1}/d & -V(11)W^{-1}/d \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{matrix} \frac{u\underline{x}(1)^2 - u\underline{y}(1)^2/d}{2V(22)^{-2}W^2} & \frac{u\underline{x}(1)u\underline{x}(2) - u\underline{y}(1)u\underline{y}(2)/d}{(W+2V(12)V(21))^{-1}W^2} \\ \frac{u\underline{y}(1)/d - u\underline{x}(1)}{V(22)^{-2}W^2} & \frac{u\underline{y}(2)/d - u\underline{x}(2)}{(W+2V(12)^2)^{-1}W^2} \\ \frac{u\underline{y}(2)/d - u\underline{x}(2)}{V(12)^{-2}W^2} & \frac{u\underline{y}(1)/d - u\underline{x}(1)}{(W+2V(12)^2)^{-1}W^2} \end{matrix}$$

$$
\begin{array}{ll}
\dfrac{V(12)V(22)}{(1/d-1)-1W^2} & \dfrac{W+2V(12)V(21)}{(1-1/d)-1W^2} \\
\dfrac{(1-1/d)V(22)^2}{2W^2} & \dfrac{(1/d-1)V(12)V(22)}{W^2} \\
\dfrac{(1-1/d)V(12)^2}{2W^2} & \dfrac{(1/d-1)V(11)V(12)}{W^2}
\end{array}
$$

$$
\begin{array}{ll}
\dfrac{(u\underline{x}(2)^2-u\underline{y}(2)^2/d)}{2V(11)-1W^2} & \dfrac{1}{d}/+\dfrac{u\underline{y}(1)^2V(22)+u\underline{y}(2)^2V(11)}{-2d^2W} \\
\dfrac{(u\underline{y}(1)/d-u\underline{x}(1)}{(V(12)V(21))-1W^2} & \dfrac{u\underline{y}(1)V(22)}{d^2W} \\
\dfrac{u\underline{y}(2)/d-u\underline{x}(2)}{V(11)-2W^2} & \dfrac{u\underline{y}(2)V(11)}{d^2W} \\
\dfrac{(1/d-1)V(11)V(12)}{W^2} & \dfrac{V(12)}{d^2W} \\
\dfrac{(1-1/d)V(12)V(21)}{2W^2} & \dfrac{-V(22)}{2d^2W} \\
\dfrac{(1-1/d)V(11)^2}{2W^2} & \dfrac{-V(11)}{2d^2W}
\end{array}
$$

where $W = V(11)V(22)-V(12)V(21)$.

Substituting the parameter values with two-variate PCS into (B.37) yields the following M matrix.

$$(B.38)\quad M=\begin{bmatrix} \frac{1}{p\underline{1}p\underline{2}} & -\Delta\sqrt{\frac{1+d}{2}} & 0 & 0 & 0 & \frac{(1+d)}{4\Delta^{-2}} & 0 & 0 & \frac{1}{d} \\ 0 & 1 & 0 & \frac{-1}{d} & 0 & -\Delta\sqrt{\frac{1+d}{2}} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \frac{-1}{d} & 0 & -\Delta\sqrt{\frac{1+d}{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-\frac{1}{d} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}\frac{1}{d}(1-/) & 0 & 0 & \frac{-1}{2d^2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}\frac{1}{d}(1-/) & \frac{-1}{2d^2} \end{bmatrix}.$$

Once both $V(\Theta)$ and M matrices have been found, the covariance matrix of the coefficients of an estimated PQDF, V(P), can be obtained immediately from the following equation:

$$(B.39)\quad V(P)= V_P = M\ V(\Theta)\ M'$$

$$= \frac{1}{p\underline{1}p\underline{2}} \begin{bmatrix} V\underline{11} & V\underline{12} & V\underline{13} & V\underline{14} & V\underline{15} & V\underline{16} \\ V\underline{21} & V\underline{22} & V\underline{23} & V\underline{24} & V\underline{25} & V\underline{26} \\ V\underline{31} & V\underline{32} & V\underline{33} & V\underline{34} & V\underline{35} & V\underline{36} \\ V\underline{41} & V\underline{42} & V\underline{43} & V\underline{44} & V\underline{45} & V\underline{46} \\ V\underline{51} & V\underline{52} & V\underline{53} & V\underline{54} & V\underline{55} & V\underline{56} \\ V\underline{61} & V\underline{62} & V\underline{63} & V\underline{64} & V\underline{65} & V\underline{65} \end{bmatrix},$$

where

$$V\underline{11} = 2 + \frac{(1+d)^2 \Delta^4 p\underline{2}(1+p\underline{1})}{16},$$

$$V\underline{22} = (p\underline{2} + \frac{p\underline{1}}{d}) + \frac{\Delta^2}{2}(1+d)\, p\underline{2}\,(1+p\underline{1}),$$

$$V\underline{33} = (p\underline{2} + \frac{p\underline{1}}{d}) + \frac{\Delta^2}{2}(1+d)\, p\underline{1}p\underline{2},$$

$$V\underline{44} = p\underline{1}p\underline{2}(1 - \frac{1}{d})^2,$$

$$V\underline{55} = V\underline{66} = \frac{1}{4}\{(p\underline{2} + \frac{p\underline{1}}{d^2}) + p\underline{1}p\underline{2}(1 - \frac{1}{d})^2\},$$

$$V\underline{12} = V\underline{21} = \frac{-\Delta^3}{4}(1+d)\, p\underline{2}(1+p\underline{1})\sqrt{\frac{1+d}{2}},$$

$$V\underline{15} = V\underline{51} = \frac{1}{2}\{\frac{\Delta^2}{4}(1+d)\, p\underline{2}\,(1 + p\underline{1} - \frac{p\underline{1}}{d}) - (p\underline{2} + \frac{p\underline{1}}{d})\},$$

$$V\underline{16} = V\underline{61} = \frac{1}{2}\{\frac{\Delta^2}{4}(1+d)\, p\underline{2} - 1\}\,(p\underline{2} + \frac{p\underline{1}}{d}),$$

$$V\underline{25} = V\underline{52} = \frac{-p\underline{2}\Delta}{2}\sqrt{\frac{1+d}{2}}(1 + p\underline{1} - \frac{p\underline{1}}{d}),$$

$$V\underline{26} = V\underline{62} = \frac{-p\underline{2}\Delta}{2}\sqrt{\frac{1+d}{2}}(p\underline{2} + \frac{p\underline{1}}{d}),$$

$$V\underline{34} = V\underline{43} = -p\underline{1}p\underline{1}(1-\frac{1}{d})\Delta\sqrt{\frac{1+d}{2}},$$

$$V\underline{56} = V\underline{65} = \frac{1}{4}\{(p\underline{2}+\frac{p\underline{1}}{d^2})-p\underline{1}p\underline{2}(1-\frac{1}{d})^2\},$$

and the rest of the elements in V(P) are all equal to zero.

# APPENDIX C

## Derivation of Asymptotic Error Rates with Two-variate LCS and PCS

Following Theorem (1) of O'Neill (1980), the asymptotic error rate of t is defined as:

$$(C.1)\quad AERR(t) = \lim_{n} E\{n\,(ERR(t) - ERR(\theta))\}$$

$$= E(z'Bz) = tr\{E(z'Bz)\}$$

$$= tr\{B\,E(zz')\} = tr(BV).$$

where z is distributed as $N_{k+1}(0, V)$, and

$$(C.2)\quad B = \frac{1}{4}\int_R |\nabla_x D(X,\theta)|^{-2} \nabla_\theta D(X,\theta)\ \nabla'_\theta D(X,\theta)\ f(X)\ dm_R,$$

where $m_R$ is a Lebesque measure on R and

$$(C.3)\quad f(X) = p_1 f(X|1) + p_2 f(X|2).$$

The derivation of the O'Neill's B matrix in two-variate case is shown as below.

(1) With Linear Covariance Structure (LCS): The optimal discriminant function with two-variate LCS can be written as

$$(C.4)\quad Dt(X) = L(X) = X'\Sigma^{-1}(\Delta,0)' - 1/2\ \{(\Delta,0)\Sigma^{-1}(\Delta,0)'\}$$

$$+\log(p_1/p_2)$$

$$= \Delta X_1 - 1/2\ \Delta^2 + \lambda,$$

where $\lambda = \log(p\underline{1}/p\underline{2})$. Then O'Neill's B matrix using an estimated LDF is given by

$$(C.5)\quad B_L = \frac{1}{4}\int_R \left[V_x\left(\lambda - \frac{\Delta^2}{2} + \Delta X_1\right)\right]^{-1} \cdot V_\beta(\beta_0 + \beta_1 X_1 + \beta_2 X_2)\, V'(\beta_0 + \beta_1 X_1 + \beta_2 X_2) f(X)\, dm$$

$$= \frac{1}{4}\int_R \Delta^{-1}\begin{bmatrix}1\\ X\underline{1}\\ X\underline{2}\end{bmatrix}(1, X_1, X_2)\{p_1 f(X|1) + p_2 f(X|2)\}\, dm$$

$$= \frac{1}{4\Delta}\int_R \begin{bmatrix}1 & X\underline{1} & X\underline{2}\\ X\underline{1} & X\underline{1}^2 & X\underline{1}X\underline{2}\\ X\underline{2} & X\underline{1}X\underline{2} & X\underline{2}^2\end{bmatrix}\left\{\frac{p\underline{1}}{2\pi} e^{-\frac{1}{2}\{(X_1-\Delta)^2+X_2^2\}} + \frac{p\underline{2}}{2\pi} e^{-\frac{1}{2}(X_1^2+X_2^2)}\right\} dm,$$

where $R=\{X|Dt(X)=0\}$. Thus on the discriminant boundary $R=\{X|\Delta X\underline{1} - 1/2\ \Delta^2+\lambda=0\}$, the value of $X\underline{1}$ is $X\underline{1}=\Delta/2 - \lambda/\Delta$. Substituting $X\underline{1}$ into (C.5), one has

$$(C.6)\quad B_L = \frac{1}{4\Delta}\int_{X_2} \begin{bmatrix}1 & X\underline{1} & X\underline{2}\\ X\underline{1} & X\underline{1}^2 & X\underline{1}X\underline{2}\\ X\underline{2} & X\underline{1}X\underline{2} & X\underline{2}^2\end{bmatrix}\left\{\frac{p\underline{1}}{2\pi} e^{-\frac{1}{2}\{(\frac{\Delta}{2}-\frac{\lambda}{\Delta}-\Delta)^2+X_2^2\}} + \frac{p\underline{2}}{2\pi} e^{-\frac{1}{2}\{(\frac{\Delta}{2}-\frac{\lambda}{\Delta})^2+X_2^2\}}\right\} dX_2.$$

$$= \frac{1}{4\Delta}\int_{X_2} \frac{1}{2\pi} e^{-\frac{1}{2}X_2^2}\left\{p_1 e^{-\frac{1}{2}(\frac{\Delta}{2}+\frac{\lambda}{\Delta})^2} + p_2 e^{-\frac{1}{2}(\frac{\Delta}{2}+\frac{-\lambda}{\Delta})^2}\right\} dX_2$$

Thus, with some algebra, one has

$$(C.7)\quad B_L = \frac{1}{4\Delta}\,\Phi\left(\frac{\Delta}{2}+\frac{\lambda}{\Delta}\right)\left(1+e^{-\lambda}e^{\lambda}\right)\int_{X_2}\begin{bmatrix}1 & \underline{X1} & \underline{X2}\\ \underline{X1} & \underline{X1}^2 & \underline{X1}\underline{X2}\\ \underline{X2} & \underline{X1}\underline{X2} & \underline{X2}^2\end{bmatrix}\Phi(X_2)\,dX_2 .$$

With $\underline{X1}$ being constant, the value of the integral with respect to $\underline{X2}$ can be regarded as the expected value of the 3 X 3 cross-product matrix of $\underline{X2}$ where $\underline{X2}$ has a standard normal distribution. Thus

$$(C.8)\quad B_L = \frac{p_{\underline{1}}}{2\Delta}\,\Phi\left(\frac{\Delta}{2}+\frac{\lambda}{\Delta}\right)\begin{bmatrix}1 & (\Delta/2-\lambda/\Delta) & 0\\ (\Delta/2-\lambda/\Delta) & (\Delta/2-\lambda/\Delta)^2 & 0\\ 0 & 0 & 1\end{bmatrix}.$$

Since the discriminant boundary of the estimated PQDF with LCS will converge asymptotically to $R=\{X \mid \Delta\,\underline{X1} - 1/2\,\Delta^2 + \lambda=0\}$, the B matrice using estimated PQDF and LDF are different only in the cross-product matrix of $\underline{X2}$, i.e.,

$$(C.9)\quad B_P = \frac{p_{\underline{1}}}{2\Delta}\,\Phi\left(\frac{\Delta}{2}+\frac{\lambda}{\Delta}\right)\int_{X_2}\begin{bmatrix}1\\ \underline{X1}\\ \underline{X2}\\ \underline{X1}\underline{X2}\\ \underline{X1}^2\\ \underline{X2}^2\end{bmatrix}\begin{bmatrix}1\\ \underline{X1}\\ \underline{X2}\\ \underline{X1}\underline{X2}\\ \underline{X1}^2\\ \underline{X2}^2\end{bmatrix}'\Phi(X_2)\,dX_2 .$$

The present problem is to compute the third and the fourth moments for $\underline{X2}$ which is a standard normal random variable. Following Kendall and Stuart (1977), the first and the third moment are both zero, the second moment is one, and the fourth moment is given by:

$$(C.10)\quad M(4) = E(X^4) = \frac{1}{2^2} \cdot \frac{4!}{2!} = 3 .$$

Thus, Equation (C.9) can be written as

$$(C.11)\quad B_P = \frac{P\underline{1}}{2\Delta}\, \Phi\left(\frac{\Delta}{2} + \frac{\lambda}{\Delta}\right) \begin{bmatrix} 1 & X\underline{1} & 0 & 0 & X\underline{1}^2 & 1 \\ X\underline{1} & X\underline{1}^2 & 0 & 0 & X\underline{1}^3 & X\underline{1} \\ 0 & 0 & 1 & X\underline{1} & 0 & 0 \\ 0 & 0 & X\underline{1} & X\underline{1}^2 & 0 & 0 \\ X\underline{1}^2 & X\underline{1}^3 & 0 & 0 & X\underline{1}^4 & X\underline{1}^2 \\ 1 & X\underline{1} & 0 & 0 & X\underline{1}^2 & 3 \end{bmatrix}$$

where $X\underline{1} = \Delta/2 - \lambda/\Delta$.

The B matrix using an estimated QDF, $B_Q$, is identical to $B_P$.

Let $B = B_P = B_Q$, the AERR(L), AERR(P), and AERR(Q) are then given by

$$(C.12)\quad AERR(L) = tr(B_L V_L),$$

$$(C.13)\quad AERR(P) = tr(B V_P),$$

and

$$(C.14)\quad AERR(Q) = tr(B V_Q),$$

where $V_L$, $V_P$, and $V_Q$ are the covariance matrices of the coefficients of the LDF, PQDF and the QDF, respectively. The procedure of deriving these matrices was discussed in Appendix B.

(2) With **Proportional Covariance Structure** (PCS): The optimal discriminant boundary with the PCS is

$$(C.15)\quad R = \{X \mid Dt(X) = \beta\underline{0}+\beta\underline{1}X\underline{1}+\beta\underline{2}X\underline{2}+\beta\underline{3}X\underline{1}X\underline{2}+\beta\underline{4}X\underline{1}^2+\beta\underline{5}X\underline{2}^2=0\}$$

$$= \{X \mid Dt(X) = \log d + \lambda - \Delta^2(1+d)/4 + \Delta\sqrt{(1+d)/2}\, X\underline{1} + (1-d)X\underline{1}^2/2d + (1-d)X\underline{1}^2/2d = 0\}.$$

Thus

$$\beta\underline{0} = \log d + \lambda - \Delta^2(1+d)/4,$$

$$\beta\underline{1} = \Delta\sqrt{(1+d)/2},$$

$$\beta\underline{2} = \beta\underline{3} = 0,$$

$$\beta\underline{4} = \beta\underline{5} = (1-d)/2d,$$

and

$$X_1^{(\pm)} = \{-\beta_1 \pm \sqrt{\beta_1^2 - 4\beta_4(\beta_0 + \beta_5 X_2^2)}\} / 2\beta_4 .$$

After substituting $X\underline{1}$ into (C.15) and following the previous procedure, the B matrix using an estimated PQDF is then given by the following:

$$(C.16)\quad B_P = \frac{1}{4}\int_R \{(\beta_1 + 2\beta_4 X_1)^2 + (2\beta_5 X_2)^2\}^{-1/2} \begin{bmatrix} 1 \\ X\underline{1} \\ X\underline{2} \\ X\underline{1}X\underline{2} \\ X\underline{1}^2 \\ X\underline{2}^2 \end{bmatrix} \begin{bmatrix} 1 \\ X\underline{1} \\ X\underline{2} \\ X\underline{1}X\underline{2} \\ X\underline{1}^2 \\ X\underline{2}^2 \end{bmatrix}' \cdot \{p_1 f(X \mid 1) + p_2 f(X \mid 2)\} dm_R .$$

Let $\underline{X}_{cp}$ denote the cross-product matrix, then

$$(C.17)\quad B_P = \frac{1}{4}\int_{X_2} \{\beta_1^2 - 4\beta_0\beta_4 - 4\beta_4\beta_5 X_2^2 + 4\beta_5^2 X_2^2\}^{-1/2} \cdot \underline{X}_{cp}\{p_1 f(X|1) + p_2 f(X|2)\}dX_2$$

$$= \frac{1}{4}\int_{X_2} \{\beta_1^2 - 4\beta_0\beta_4\}^{-1/2} \underline{X}_{cp}\{p_1 f(X|1) + p_2 f(X|2)\}dX_2$$

$$= \frac{1}{4}(\beta_1^2 - 4\beta_0\beta_4)^{-1/2}\int_{X_2} \underline{X}_{cp}\{p_1 f(X|1) + p_2 f(X|2)\}dX_2 .$$

Let $C = \beta_1^2 - 4\beta_0\beta_4$, then

$$(C.18)\quad B_P = \frac{1}{4} C^{-1/2}\int_{X_2} \{\underline{X}_{cp}\{\frac{p_1}{2\pi} e^{-1/2\{(X_1^{(+)} - \beta_1)^2 + X_2^2\}} + \frac{p_2}{2\pi d} e^{-\frac{1}{2d}\{(X_1^{(+)})^2 + X_2^2\}}\}$$

$$+ \underline{X}_{cp}\{\frac{p_1}{2\pi} e^{-1/2\{(X_1^{(-)} - \beta_1)^2 + X_2^2\}} + \frac{p_2}{2\pi d} e^{-\frac{1}{2d}\{(X_1^{(-)})^2 + X_2^2\}}\}\}.$$

where $\underline{X}_{cp}^{(\pm)}$ is the $\underline{X}_{cp}$ resulted by substituting $\underline{X}_1^{(\pm)}$ into the cross-product matrix.

With some algebra, Equation (C.18) can be reduced to the following:

$$(C.19)\quad B_P = \frac{p\underline{1}}{2}C^{-1/2}\Big\{\int_{X_2} \underline{x}_{cp}^{(+)}\Big\{\frac{1}{2\Pi}e^{-1/2[\beta_1^2-2\beta_4\beta_0-\beta_4\sqrt{C_x}(1+2\beta_4)+2\beta_1^2\beta_4(1+\beta_4)]}\Big\}dX_2$$

$$+\int_{X_2} \underline{x}_{cp}^{(-)}\Big\{\frac{1}{2\Pi}e^{-1/2[\beta_1^2-2\beta_4\beta_0+\beta_4\sqrt{C_x}(1+2\beta_4)+2\beta_1^2\beta_4(1+\beta_4)]}\Big\}dX_2\Big\},$$

where

$$(C.20)\quad C_x = \beta_1^2-4\beta_4(\beta_0+\beta_5X_2^2) = C-4\beta_4\beta_5X_2^2.$$

This equation can be written alternatively as:

$$(C.21)\quad B_P = \frac{p\underline{1}}{2}C^{-1/2}\Big\{\int_{X_2} ww'\,\Phi\Big(\frac{\beta\underline{1}-d\sqrt{C\underline{x}}}{1-d}\Big)\,\Phi(X_2)dX_2$$

$$+\int_{X_2} ww'\,\Phi\Big(\frac{\beta\underline{1}+d\sqrt{C\underline{x}}}{1-d}\Big)\,\Phi(X_2)dX_2\Big\},$$

where

$$(C.22)\quad \mathbf{v} = \begin{bmatrix} 1 \\ (-\beta\underline{1}+\sqrt{C\underline{x}})/2\beta\underline{4} \\ X\underline{2} \\ (-\beta\underline{1}+\sqrt{C\underline{x}})\,X\underline{2}/2\beta\underline{4} \\ (-\beta\underline{1}+\sqrt{C\underline{x}})^2/2\beta\underline{4}^2 \\ X\underline{2}^2 \end{bmatrix}$$

and

$$(C.23)\quad \mathbf{w} = \begin{bmatrix} 1 \\ (-\beta\underline{1}-\sqrt{C\underline{x}})/2\beta\underline{4} \\ X\underline{2} \\ (-\beta\underline{1}-\sqrt{C\underline{x}})\,X\underline{2}/2\beta\underline{4} \\ (-\beta\underline{1}-\sqrt{C\underline{x}})^2/2\beta\underline{4}^2 \\ X\underline{2}^2 \end{bmatrix}.$$

The problem of evaluating this B matrix is intractable, thus numerical integration methods such as Simpson's rule and Newton's method were applied here, using 21 data points.

The $B_Q$ matrix is equal to $B_P$ here since both DF's have the same cross-product matrices $\underline{Xcp}$. Thus the AERR(P) and AERR(Q) can be found using identical B Matrix ($B=B_P=B_Q$), i.e.:

$$(C.24)\quad AERR(P) = tr(B\mathbf{V}_P)$$

and

$$(C.25)\quad AERR(Q) = tr(B\mathbf{V}_Q)$$

where $\mathbf{V}_P$ and $\mathbf{V}_Q$ are the covariance matrices of the coefficients of the PQDF and the QDF.

# APPENDIX D

## Derivations of Maximum Likelihood Functions

(1) With Linear Covariance Structure (LCS), i.e., $V\underline{1}=V\underline{2}=V$:

The likelihood function of the sample with known parameters is given by

$$(D.1)\quad L(u\underline{1},u\underline{2},V) = (2\pi)^{-n\underline{1}k/2}\,|V|^{-n\underline{1}/2}$$

$$\cdot\, e^{-1/2\,\{\sum_{i=1}^{n\underline{1}}(X\underline{i}-u\underline{1})'V^{-1}(X\underline{i}-u\underline{1})\}}$$

$$\cdot\,(2\pi)^{-n\underline{2}k/2}\,|V|^{-n\underline{2}/2}$$

$$\cdot\, e^{-1/2\,\{\sum_{j=1}^{n\underline{2}}(X\underline{j}-u\underline{2})'V^{-1}(X\underline{j}-u\underline{2})\}}$$

$$= (2\pi)^{-(n\underline{1}+n\underline{2})k/2}\,|V|^{-(n\underline{1}+n\underline{2})/2}$$

$$\cdot\, e^{-1/2\,\{\sum_{i=1}^{n\underline{1}}(X\underline{i}-u\underline{1})'V^{-1}(X\underline{i}-u\underline{1})\}}$$

$$\cdot\, e^{-1/2\,\{\sum_{j=1}^{n\underline{2}}(X\underline{j}-u\underline{2})'V^{-1}(X\underline{j}-u\underline{2})\}}\,.$$

It is well known that the maximum likelihood estimates (MLE) of $u\underline{1}$, $u\underline{2}$, and $V$ with LCS are given by

$$(D.2)\quad \hat{u\underline{i}} = \bar{X\underline{i}} = \frac{1}{n\underline{i}}\sum_{j=1}^{n\underline{i}} X\underline{i}j,\ i=1,2,$$

$$(D.3)\quad \hat{V} = S = S_i = \frac{1}{(n_1+n_2)} (A_1 + A_2),\ i=1,2,$$

where $X_{ij}$ is the $j$th $k \times 1$ sample vector from the population $i$ ($\pi_i$), and

$$(D.4)\quad A_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)\ (X_{ij} - \bar{X}_i)',\ i=1,2.$$

Thus the maximum likelihood function derived from the sample can be written as

$$(D.5)\quad L(\bar{X}_1, \bar{X}_2, S) = (2\pi)^{-(n_1+n_2)k/2} |S|^{-(n_1+n_2)/2}$$

$$\cdot e^{-1/2 \{ \sum_{i=1}^{n_1} (X_i - \bar{X}_1)' S^{-1} (X_i - \bar{X}_1) \}}$$

$$\cdot e^{-1/2 \{ \sum_{j=1}^{n_2} (X_j - \bar{X}_2)' S^{-1} (X_j - \bar{X}_2) \}}.$$

Since

$$(D.6)\quad e^{-1/2 \{ \sum_{i=1}^{n_1} (X_i - \bar{X}_1)' S^{-1} (X_i - \bar{X}_1) + \sum_{j=1}^{n_2} (X_j - \bar{X}_2)' S^{-1} (X_j - \bar{X}_2) \}}$$

$$= e^{-1/2 \{ \mathrm{tr}(S^{-1}A_1) + \mathrm{tr}(S^{-1}A_2) \}}$$

$$= e^{-1/2\ \mathrm{trace}(S^{-1}A_1 + S^{-1}A_2)}$$

$$= e^{-1/2\ \mathrm{trace}\{S^{-1}(A_1 + A_2)\}}$$

$$= e^{-1/2\ \mathrm{trace}\{(n\underline{1}+n\underline{2})(A\underline{1}+A\underline{2})^{-1}(A\underline{1}+A\underline{2})\}}$$

$$= e^{-1/2\ \mathrm{trace}\{(n\underline{1}+n\underline{2})\cdot I\}}$$

$$= e^{-1/2\ k(n\underline{1}+n\underline{2})},$$

thus

$$(D.7)\quad L(\bar{X}\underline{1},\bar{X}\underline{2},S) = (2\pi)^{-(n\underline{1}+n\underline{2})k/2}\ |S|^{-(n\underline{1}+n\underline{2})/2} \cdot e^{-1/2\ k(n\underline{1}+n\underline{2})}.$$

(2) With Proportional Covariance Structure, PCS, $V\underline{1}=V\underline{2}/d=V$:

The likelihood function of the sample with known parameters is given by

$$(D.8)\quad L(u\underline{1},u\underline{2},V,d) = (2\pi)^{-n\underline{1}k/2}\ |V|^{-n\underline{1}/2} \cdot e^{-1/2\ \{\sum_{i=1}^{n1}(X\underline{i}-u\underline{1})'V^{-1}(X\underline{i}-u\underline{1})\}} \cdot (2\pi)^{-n\underline{2}k/2}\ |dV|^{-n\underline{2}/2} \cdot e^{-1/2\ \{\sum_{j=1}^{n2}(X\underline{j}-u\underline{2})'(dV)^{-1}(X\underline{j}-u\underline{2})\}}$$

$$= (2\pi)^{-(n\underline{1}+n\underline{2})k/2}\ |V|^{-(n\underline{1}+n\underline{2})/2}\ d^{-n\underline{2}k/2}$$

$$\cdot\, e^{-1/2 \left\{ \sum_{i=1}^{n1} (X\underline{i} - u\underline{1})' V^{-1} (X\underline{i} - u\underline{1}) \right\}}$$

$$\cdot\, e^{-1/2 \{1/d \sum_{j=1}^{n2} (X\underline{j} - u\underline{2})' V^{-1} (X\underline{j} - u\underline{2}) \}} .$$

The MLE's of $u\underline{1}$ and $u\underline{2}$ are the same as those in LCS. As for V and d, their MLE's were derived in Section 3.2 and given by Equations (3.4) and (3.5). Thus one can write the likelihood function maximized with respect to $u\underline{1}$, $u\underline{2}$, V, and d as

$$(D.9) \quad L(\bar{X}\underline{1}, \bar{X}\underline{2}, S, \hat{d}) = (2\pi)^{-(n\underline{1}+n\underline{2})k/2} |S|^{-(n\underline{1}+n\underline{2})/2} \hat{d}^{-n\underline{2}k/2}$$

$$\cdot\, e^{-1/2 \left\{ \sum_{i=1}^{n1} (X\underline{i} - \bar{X}\underline{1})' S^{-1} (X\underline{i} - \bar{X}\underline{1}) \right\}}$$

$$\cdot\, e^{-1/2 \{1/\hat{d} \sum_{j=1}^{n2} (X\underline{j} - \bar{X}\underline{2})' S^{-1} (X\underline{j} - \bar{X}\underline{2}) \}} .$$

Since

$$(D.10) \quad e^{-1/2 \{ \sum_{i=1}^{n1} (X\underline{i} - \bar{X}\underline{1})' S^{-1} (X\underline{i} - \bar{X}\underline{1}) + 1/\hat{d} \sum_{j=1}^{n2} (X\underline{j} - \bar{X}\underline{2})' S^{-1} (X\underline{j} - \bar{X}\underline{2}) \}}$$

$$= e^{-1/2 \{ tr(S^{-1} A\underline{1}) + 1/\hat{d}\, tr(S^{-1} A\underline{2}) \}}$$

$$= e^{-1/2\, trace\{ S^{-1} A\underline{1} + 1/\hat{d}\, (S^{-1} A\underline{2}) \}}$$

$$= e^{-1/2\, trace\{ S^{-1} (A\underline{1} + 1/\hat{d}\, A\underline{2}) \}}$$

$$= e^{-1/2 \text{ trace}\{(n\underline{1}+n\underline{2}) \ (A\underline{1}+1/\hat{d}\ A\underline{2})^{-1}(A\underline{1}+1/\hat{d}\ A\underline{2})\}}$$

$$= e^{-1/2 \text{ trace}\{(n\underline{1}+n\underline{2})\cdot I\}}$$

$$= e^{-1/2\ k(n\underline{1}+n\underline{2})},$$

thus

$$(D.11)\quad L(\bar{X}\underline{1},\bar{X}\underline{2},S,\hat{d}) = (2\pi)^{-(n\underline{1}+n\underline{2})k/2}\ |S|^{-(n\underline{1}+n\underline{2})/2}\ \hat{d}^{-n\underline{2}k/2} \cdot e^{-1/2\ k(n\underline{1}+n\underline{2})}.$$

(3) With Quadratic Covariance Structure (QCS), $V\underline{1} \neq (1/d)V\underline{2}$:

The likelihood function of the sample with known parameters is given by

$$(D.12)\quad L(u\underline{1},u\underline{2},V\underline{1},V\underline{2}) = (2\pi)^{-n\underline{1}k/2}\ |V\underline{1}|^{-n\underline{1}/2} \cdot e^{-1/2\ \{\sum_{i=1}^{n\underline{1}} (X\underline{i}-u\underline{1})'V\underline{1}^{-1}(X\underline{i}-u\underline{1})\}} \cdot (2\pi)^{-n\underline{2}k/2}\ |V\underline{2}|^{-n\underline{2}/2} \cdot e^{-1/2\ \{\sum_{j=1}^{n\underline{2}} (X\underline{j}-u\underline{2})'V\underline{2}^{-1}(X\underline{j}-u\underline{2})\}}$$

$$= (2\pi)^{-(n\underline{1}+n\underline{2})k/2}\ |V\underline{1}|^{-n\underline{1}/2}\ |V\underline{2}|^{-n\underline{2}/2}$$

$$\cdot\ e^{-1/2\ \{\sum_{i=1}^{n1}(\underline{X}i-\underline{u}1)'\underline{V}1^{-1}(\underline{X}i-\underline{u}1)\}}$$

$$\cdot\ e^{-1/2\ \{\sum_{j=1}^{n2}(\underline{X}j-\underline{u}2)'\underline{V}2^{-1}(\underline{X}j-\underline{u}2)\}}.$$

The MLE's of $\underline{u}1$ and $\underline{u}2$ are the same as those in LCS. The MLE's of $\underline{V}1$ and $\underline{V}2$ are defined as follows:

$$\text{(D.13)}\quad \hat{\underline{V}i} = \underline{S}i = \frac{1}{ni}\,\underline{A}i,\ i=1,2,$$

where $\underline{A}i$ was defined in Equation (D.4) and $\underline{X}ij$ is the $\underline{j}$th k X 1 sample vector from the population i.

Thus the maximum likelihood function derived from the sample can be written as

$$\text{(D.14)}\quad L(\bar{\underline{X}}1,\bar{\underline{X}}2,\underline{S}1,\underline{S}2) = (2\pi)^{-(n1+n2)k/2}\,|\underline{S}1|^{-n1/2}\,|\underline{S}2|^{-n2/2}$$

$$\cdot\ e^{-1/2\ \{\sum_{i=1}^{n1}(\underline{X}i-\bar{\underline{X}}1)'\underline{S}1^{-1}(\underline{X}i-\bar{\underline{X}}1)\}}$$

$$\cdot\ e^{-1/2\ \{\sum_{j=1}^{n2}(\underline{X}j-\bar{\underline{X}}2)'\underline{S}2^{-1}(\underline{X}j-\bar{\underline{X}}2)\}}.$$

Since

$$\text{(D.15)}\quad e^{-1/2\ \{\sum_{i=1}^{n1}(\underline{X}i-\bar{\underline{X}}1)'\underline{S}1^{-1}(\underline{X}i-\bar{\underline{X}}1)+\sum_{j=1}^{n2}(\underline{X}j-\bar{\underline{X}}2)'\underline{S}2^{-1}(\underline{X}j-\bar{\underline{X}}2)\}}$$

$$= e^{-1/2\ \{tr(\underline{S}1^{-1}\underline{A}1)\ +\ tr(\underline{S}2^{-1}\underline{A}2)\}}$$

$$= e^{-1/2\ \mathrm{trace}(S\underline{1}^{-1}A\underline{1} + S\underline{2}^{-1}A\underline{2})}$$

$$= e^{-1/2\ \mathrm{trace}\{n\underline{1}A\underline{1}^{-1}A\underline{1} + n\underline{2}A\underline{2}^{-1}A\underline{2}\}}$$

$$= e^{-1/2\ \mathrm{trace}\{n\underline{1}\cdot I + n\underline{2}\cdot I\}}$$

$$= e^{-1/2\ \mathrm{trace}\{(n\underline{1}+n\underline{2})\cdot I\}}$$

$$= e^{-1/2\ k(n\underline{1}+n\underline{2})},$$

thus

$$(D.16)\quad L(\bar{X}\underline{1}, \bar{X}\underline{2}, S\underline{1}, S\underline{2}) = (2\pi)^{-(n\underline{1}+n\underline{2})k/2} |S\underline{1}|^{-n\underline{1}/2} |S\underline{2}|^{-n\underline{2}/2}$$
$$\cdot e^{-1/2\ k(n\underline{1}+n\underline{2})}.$$

# APPENDIX E

## Plots of Asymptotic and Finite Sample Relative Efficiency

# ASYMPTOTIC RELATIVE EFFICIENCY

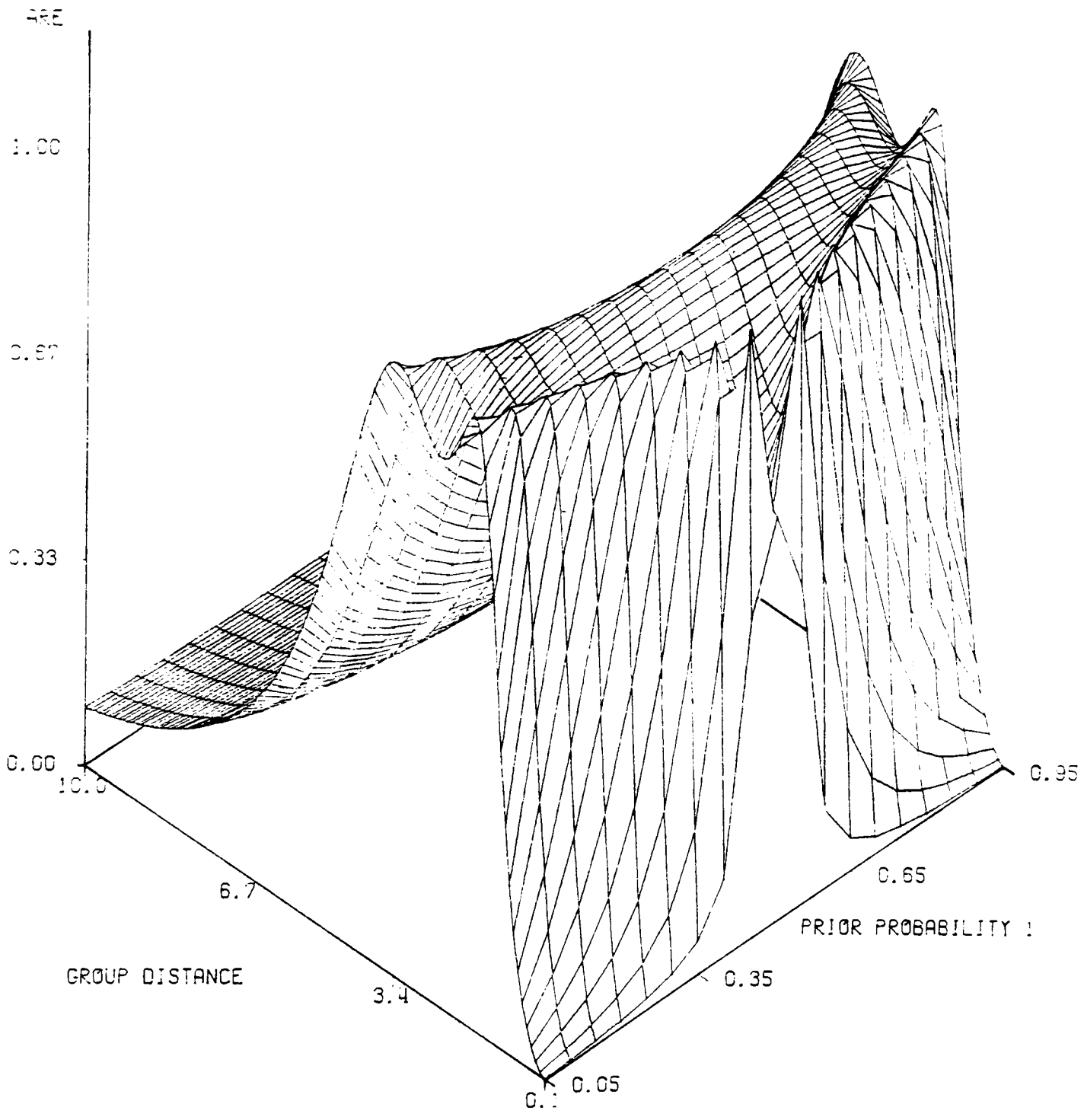



**Figure 8: Asymptotic Relative Efficiency of the PQDF to the LDF with One-variate LCS**

# ASYMPTOTIC RELATIVE EFFICIENCY

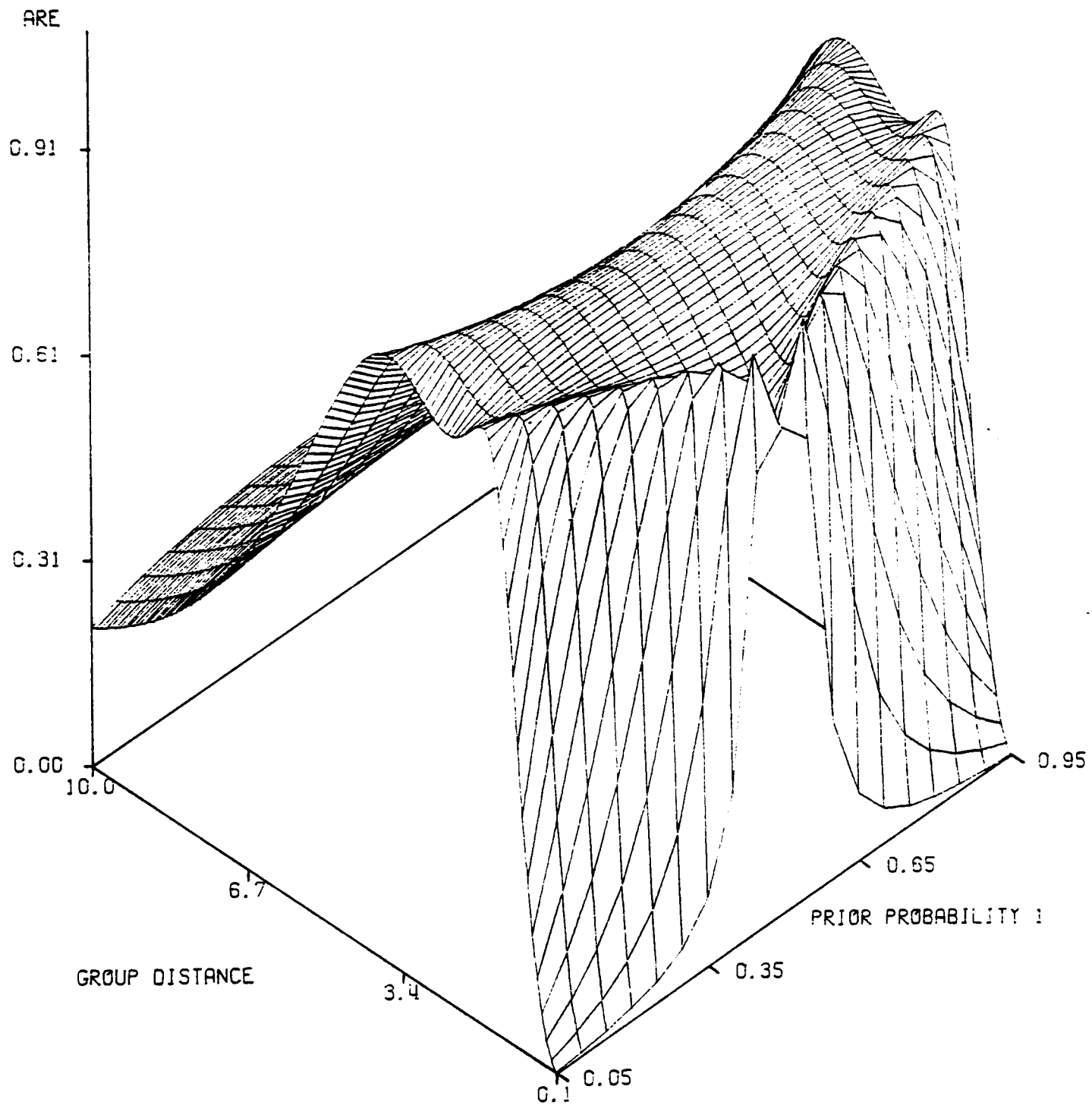



**Figure 9: Asymptotic Relative Efficiency of the PQDF to the LDF with Two-variate LCS**

# ASYMPTOTIC RELATIVE EFFICIENCY

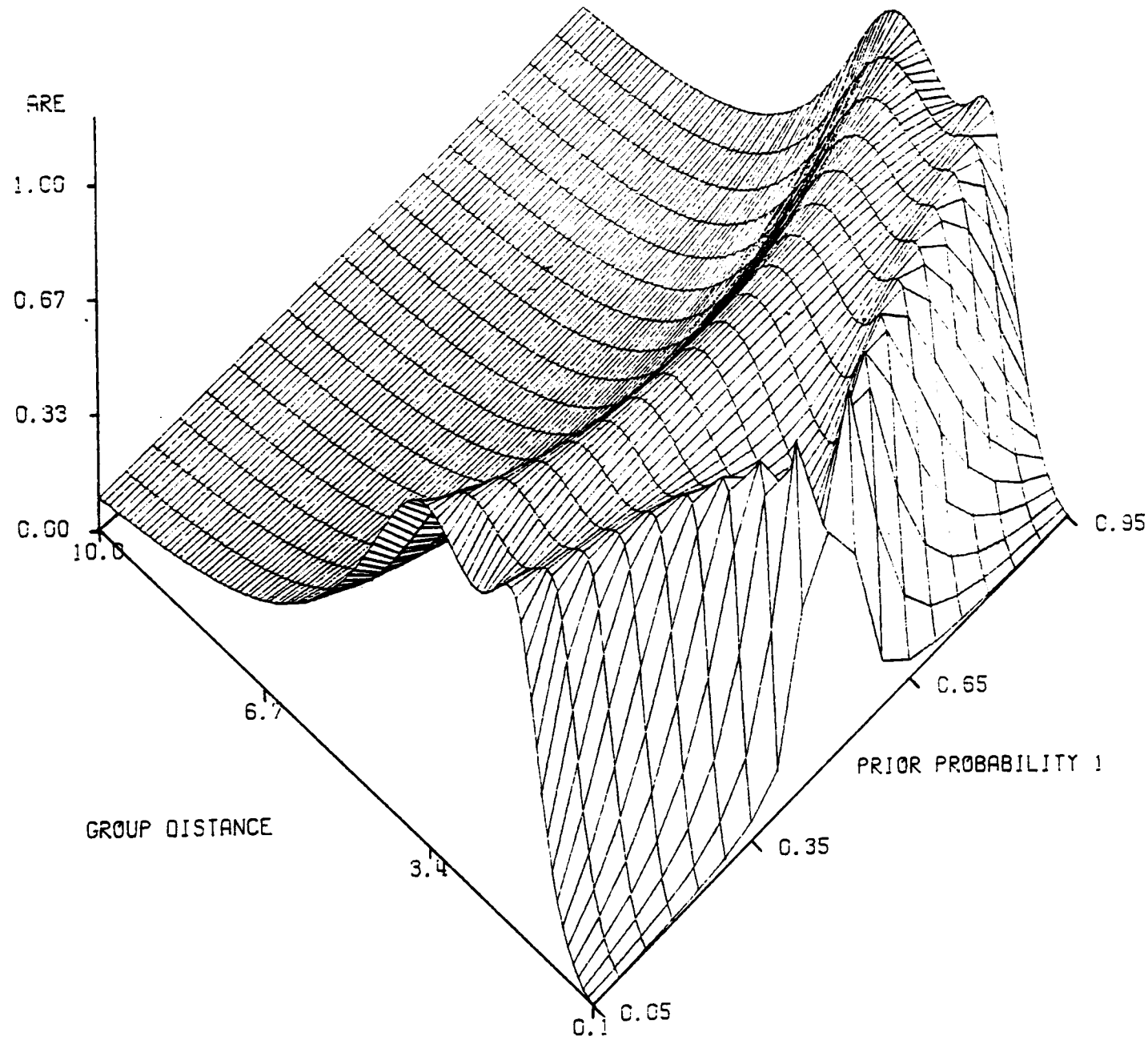



**Figure 10: Asymptotic Relative Efficiency of the PQDF to the LDF with Two-variate LCS (Tilted at a 45 Degree Angle)**

# ASYMPTOTIC RELATIVE EFFICIENCY

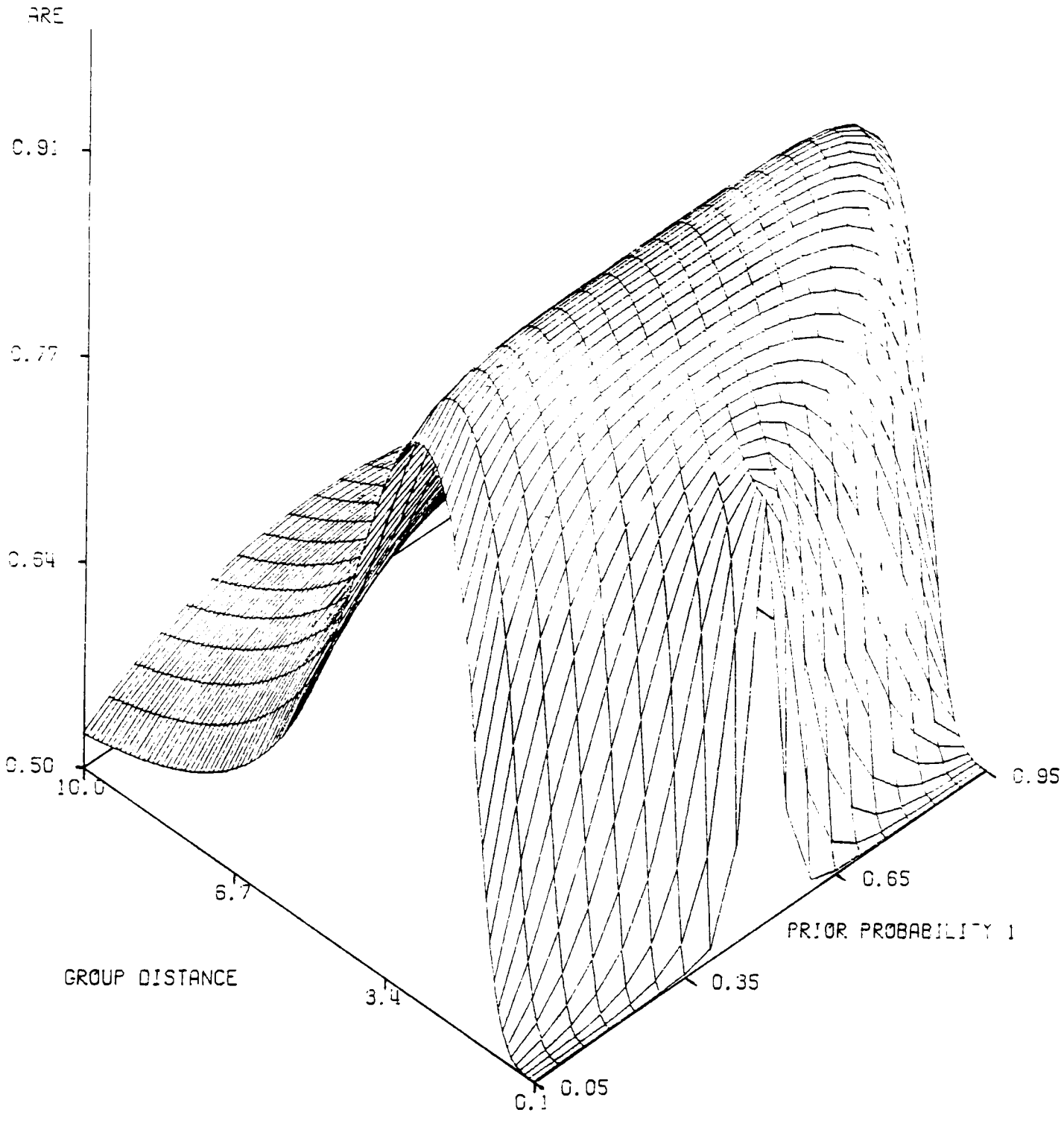



**Figure 11: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate LCS**

# ASYMPTOTIC RELATIVE EFFICIENCY

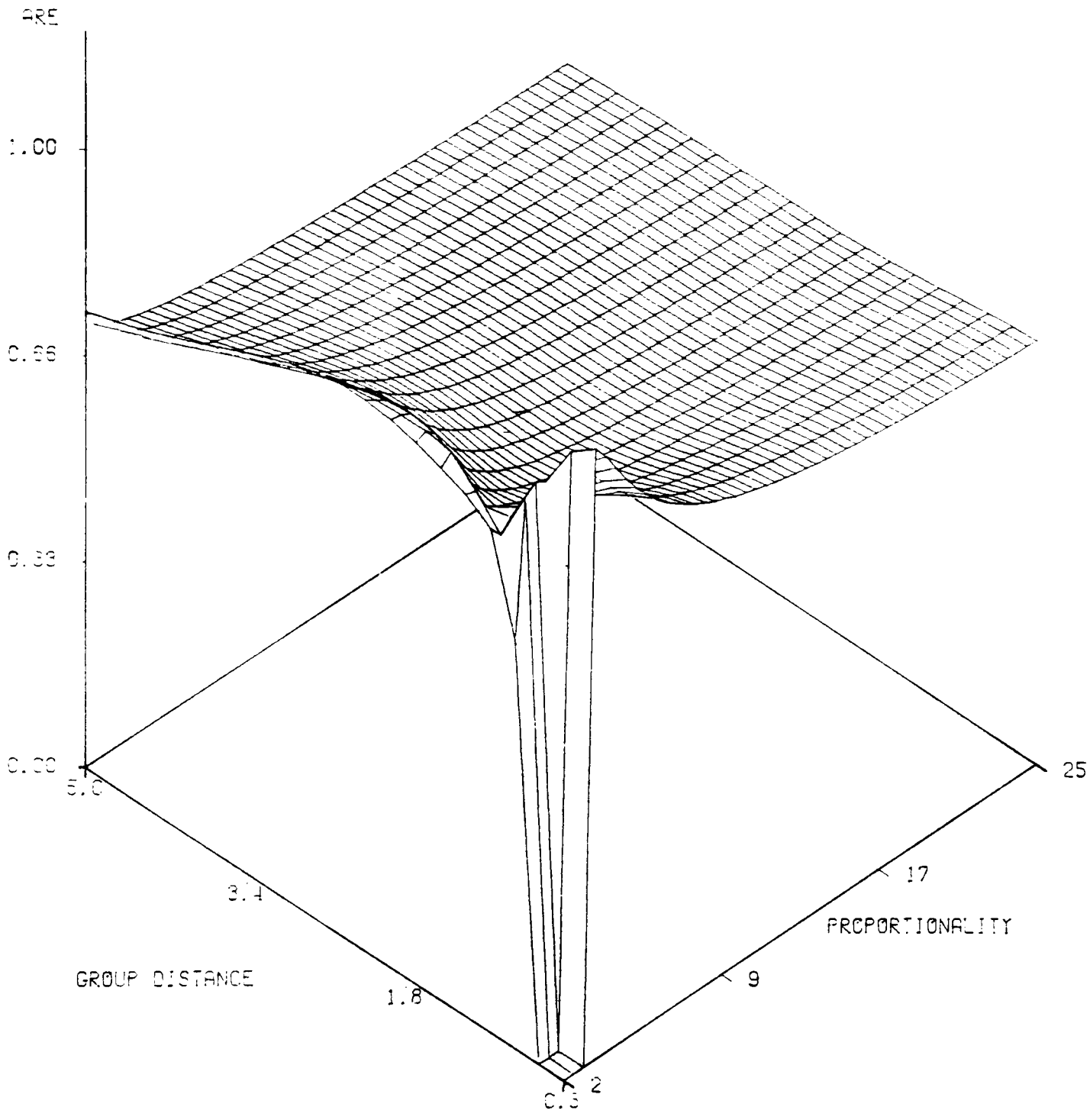



**Figure 12: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 0.25 as the Prior Probability One**

# ASYMPTOTIC RELATIVE EFFICIENCY

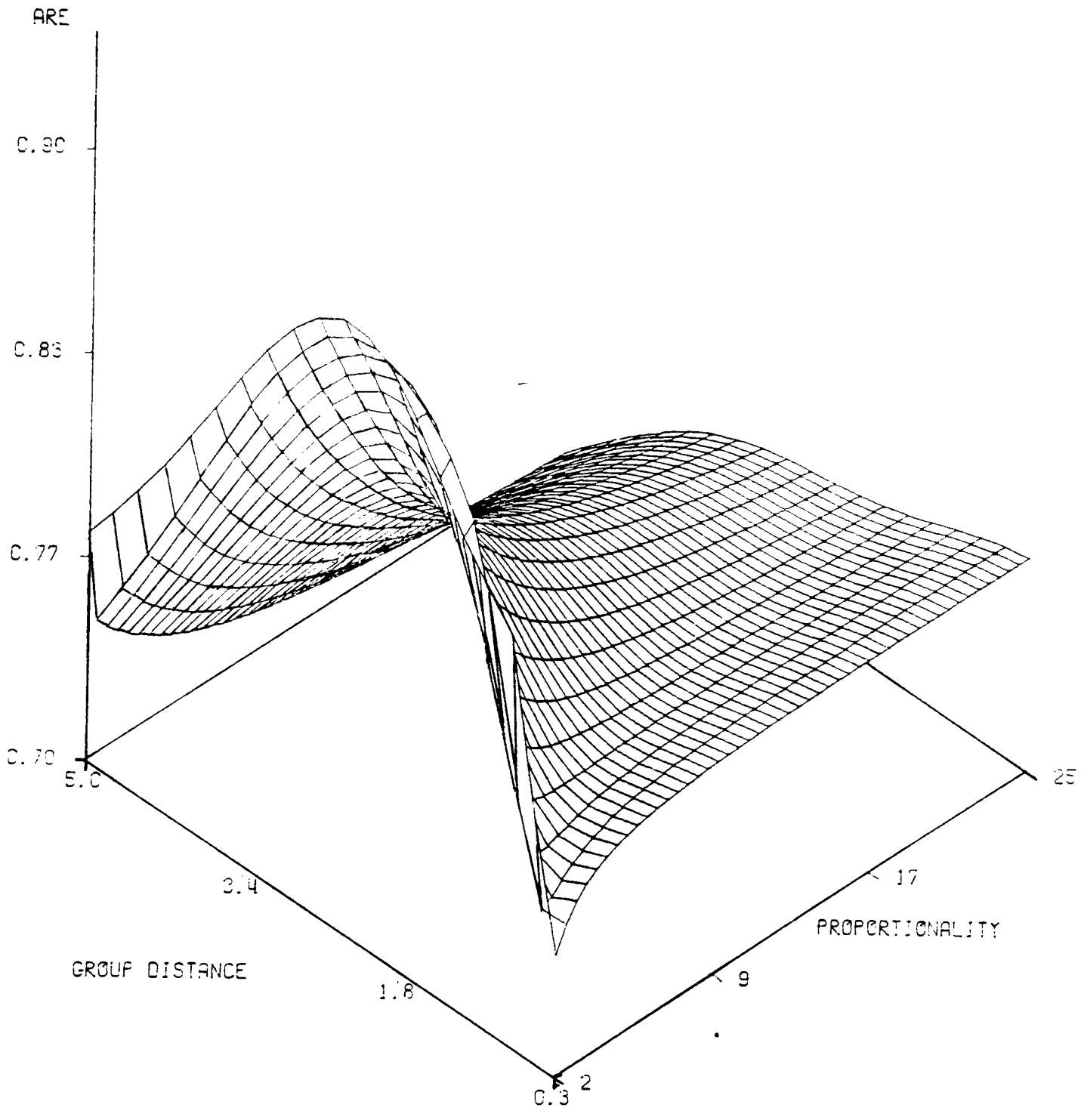



**Figure 13: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 0.5 as the Prior Probability One**

# ASYMPTOTIC RELATIVE EFFICIENCY

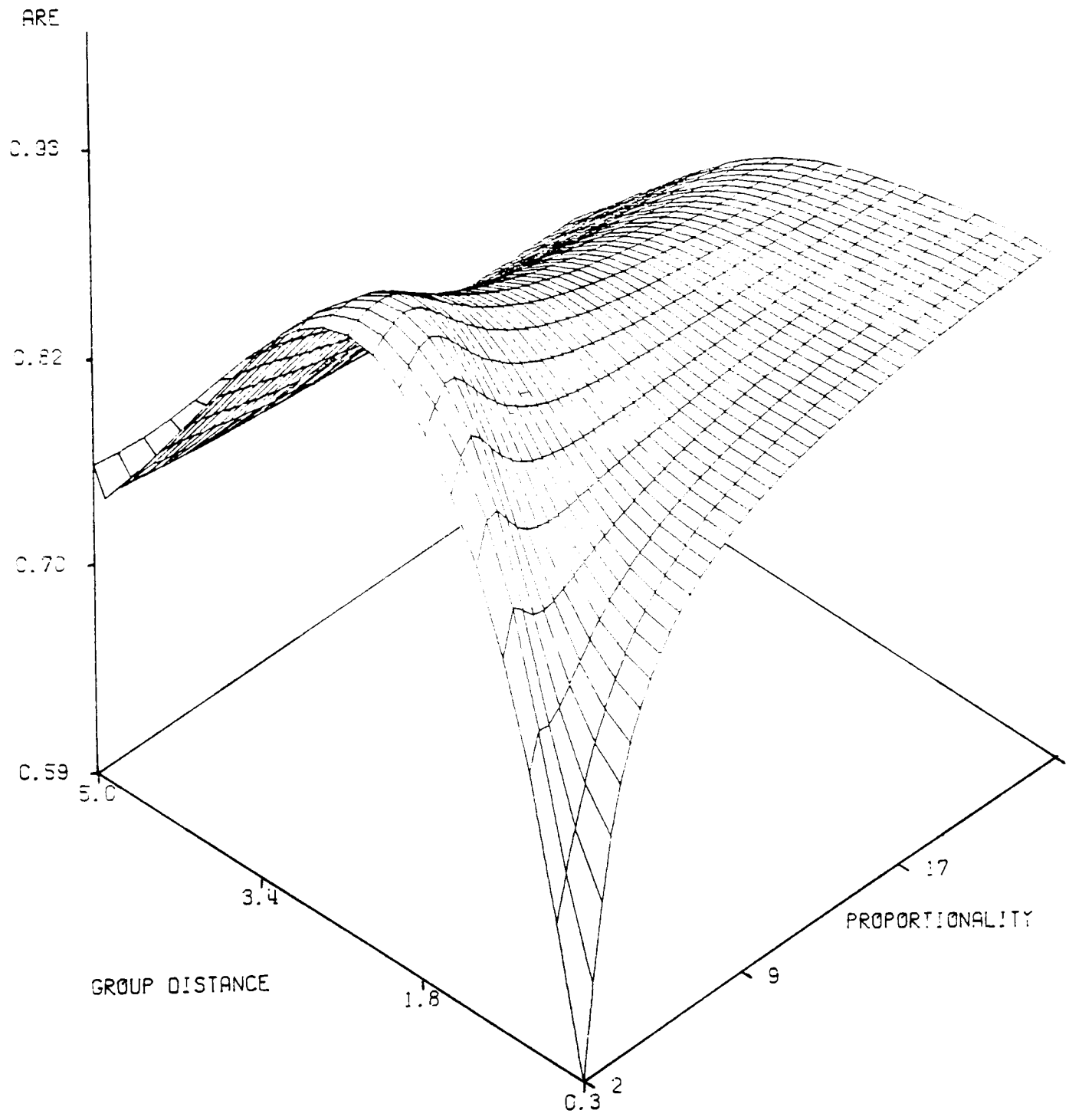



**Figure 14: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 0.75 as the Prior Probability One**

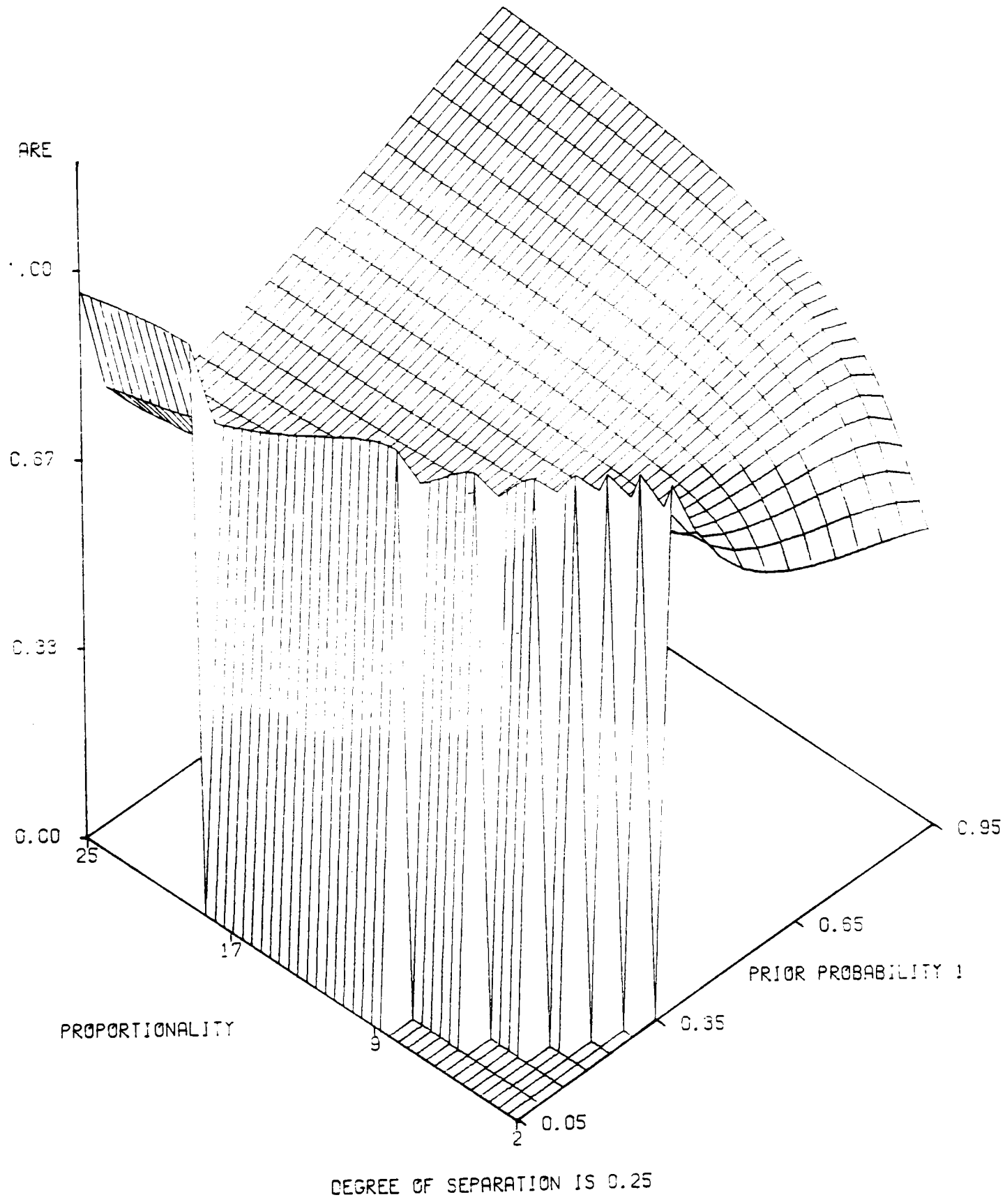



**Figure 15: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 0.25 as the Group Distance**

# ASYMPTOTIC RELATIVE EFFICIENCY

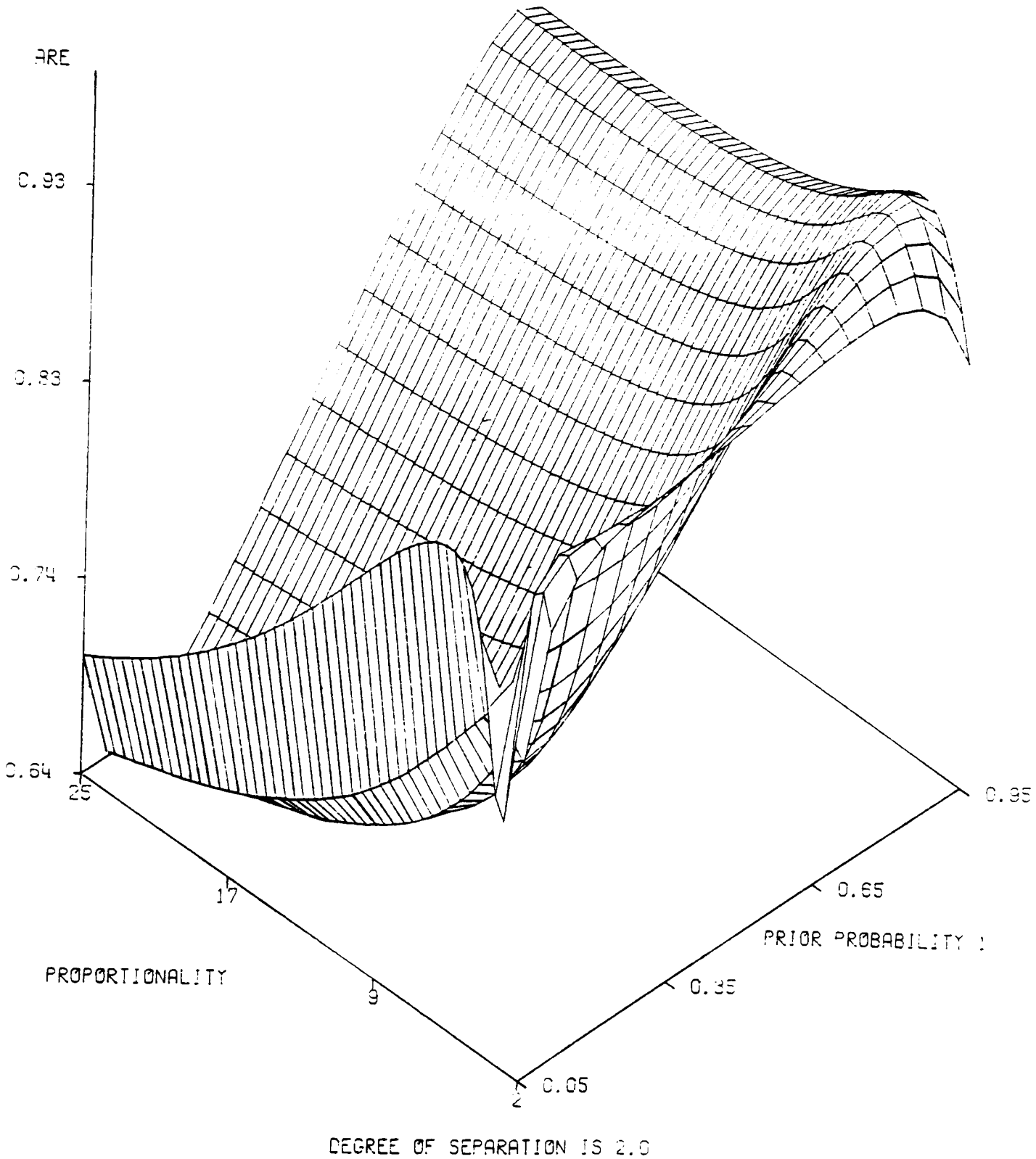



**Figure 16: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 2.0 as the Group Distance**

# ASYMPTOTIC RELATIVE EFFICIENCY

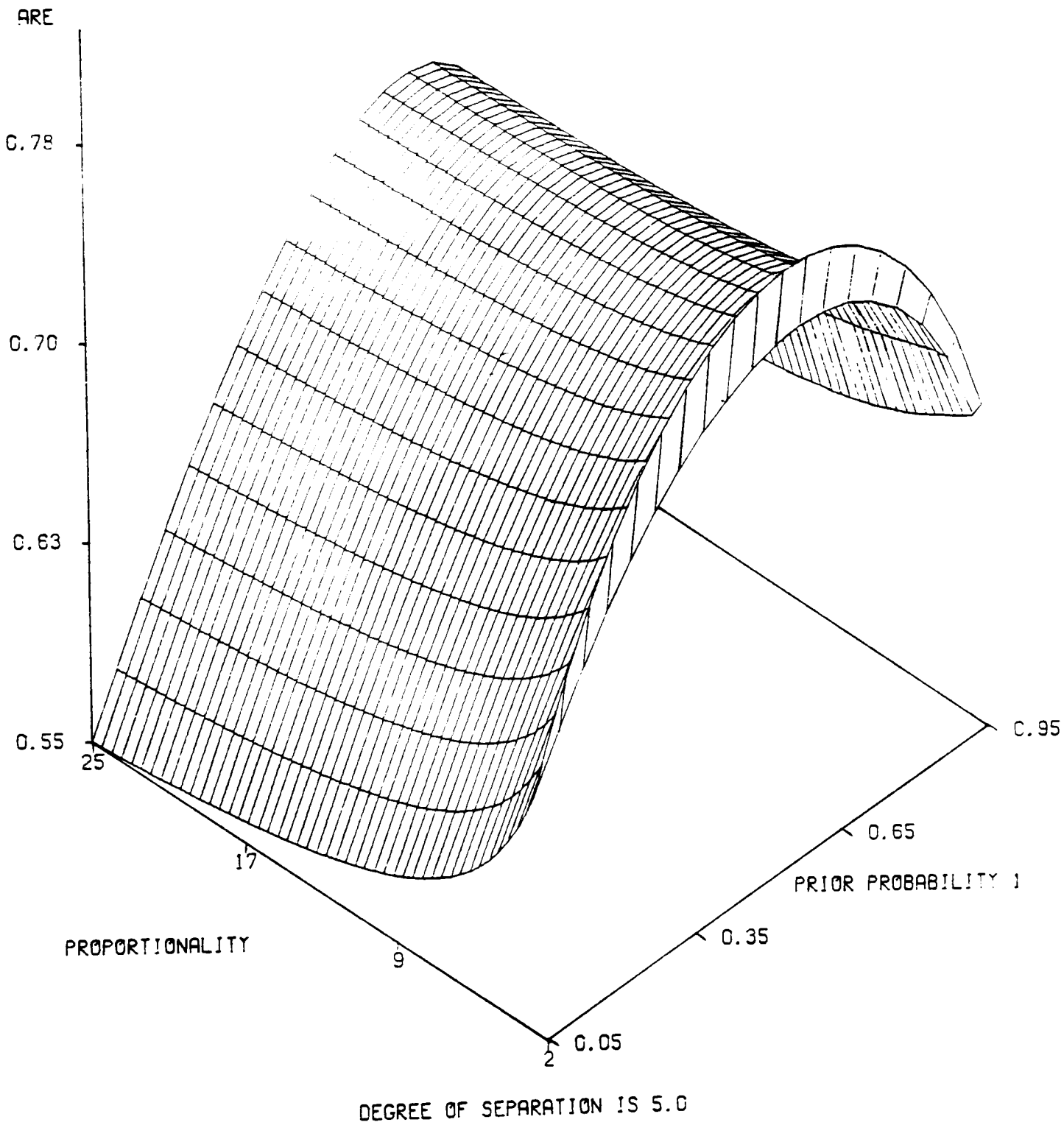



Figure 17: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 5.0 as the Group Distance

# ASYMPTOTIC RELATIVE EFFICIENCY

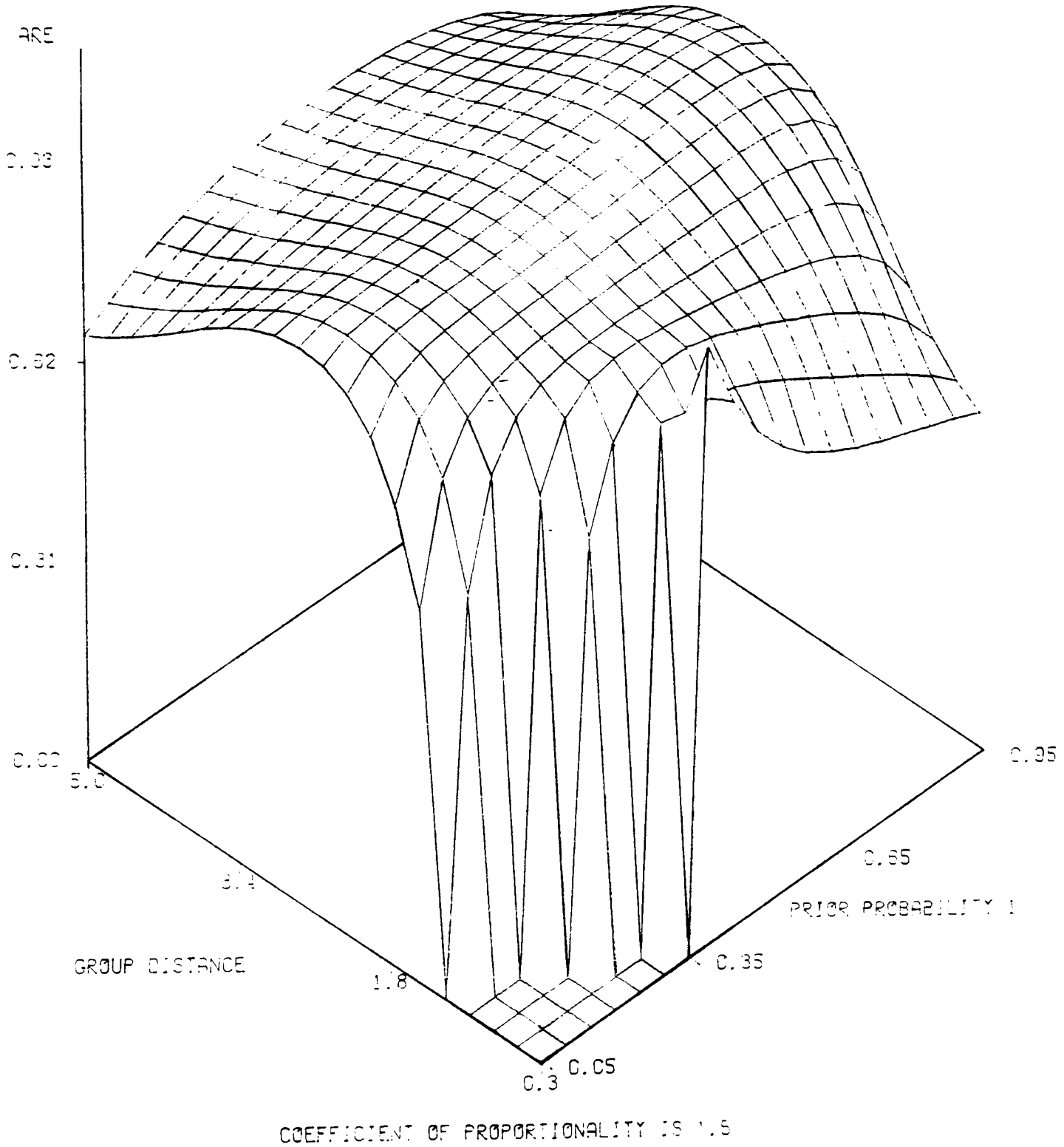



**Figure 18: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 1.5 as the Coefficient of Proportionality**

# ASYMPTOTIC RELATIVE EFFICIENCY

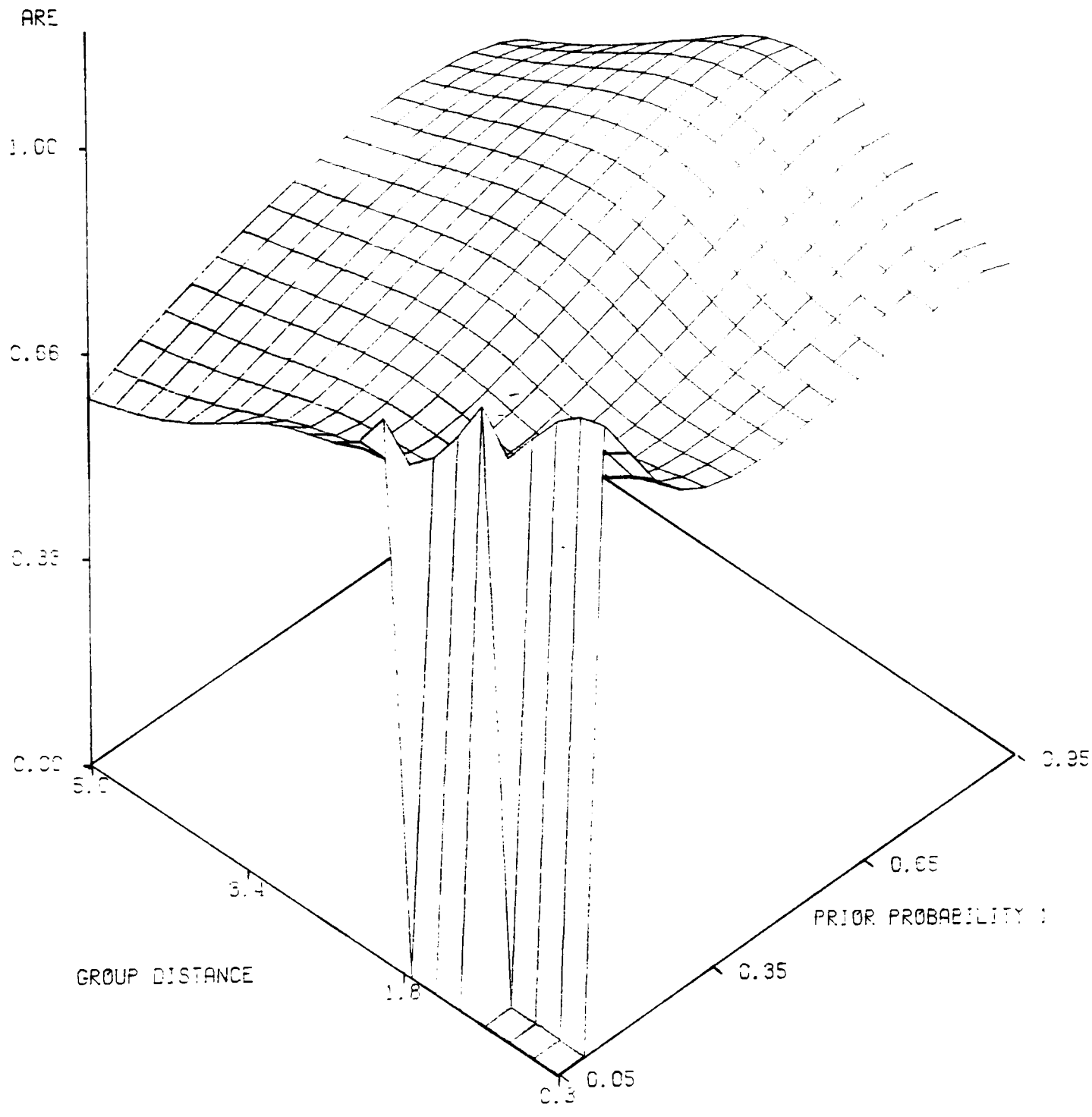



**Figure 19: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 6.0 as the Coefficient of Proportionality**

# ASYMPTOTIC RELATIVE EFFICIENCY

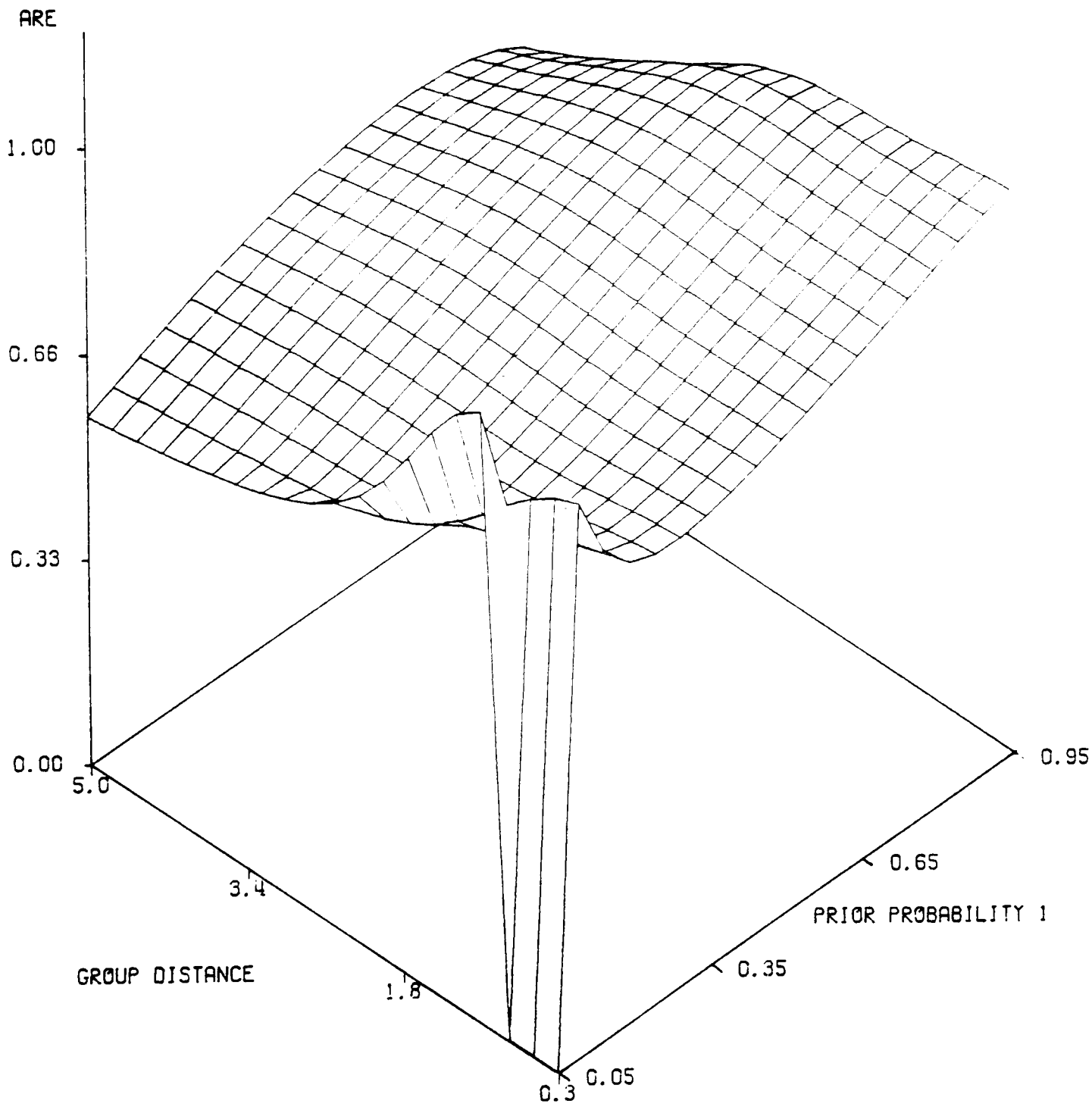



Figure 20: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 15.0 as the Coefficient of Proportionality

# ASYMPTOTIC RELATIVE EFFICIENCY

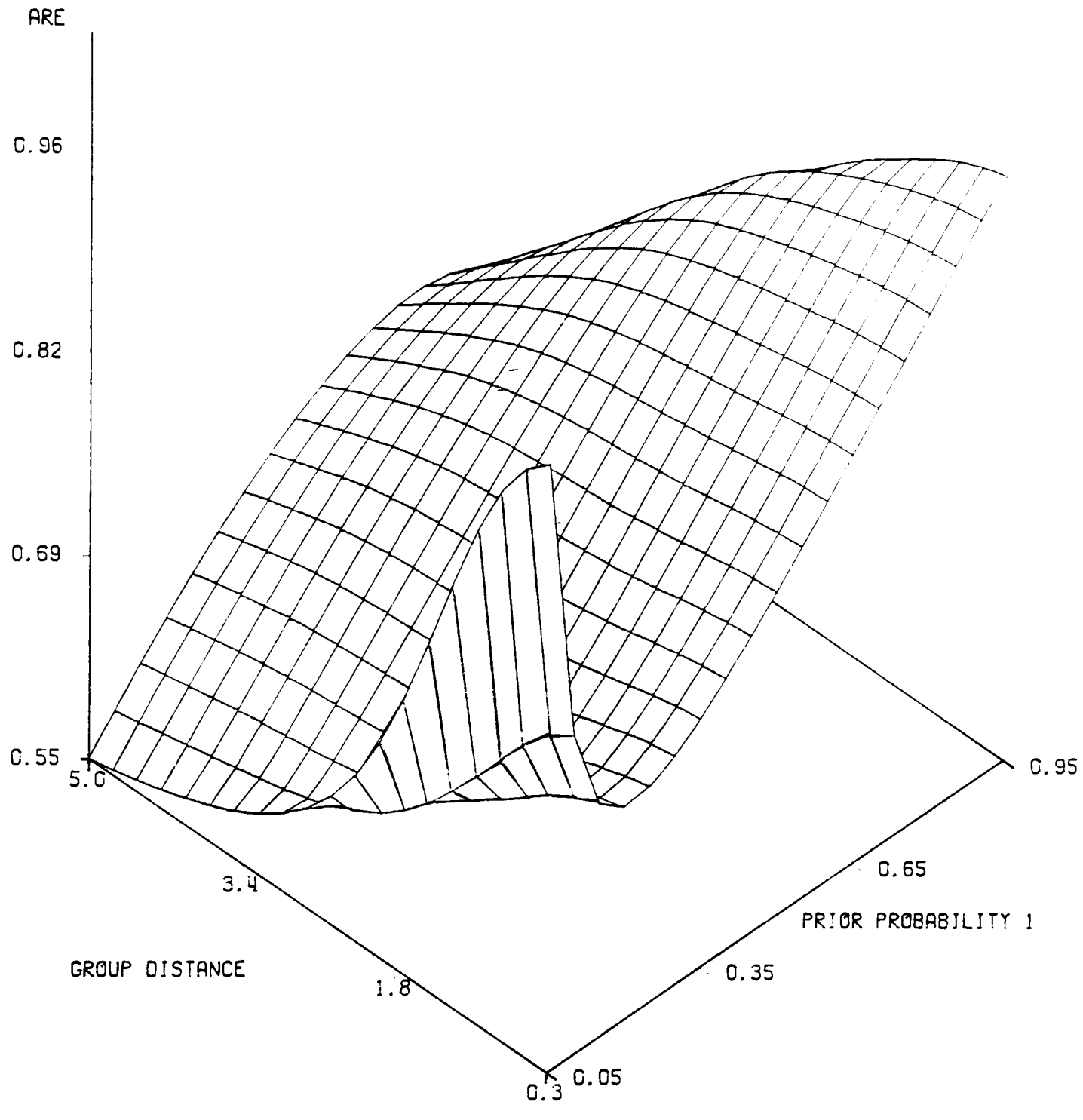



**Figure 21: Asymptotic Relative Efficiency of the QDF to the PQDF with Two-variate PCS and 25.0 as the Coefficient of Proportionality**

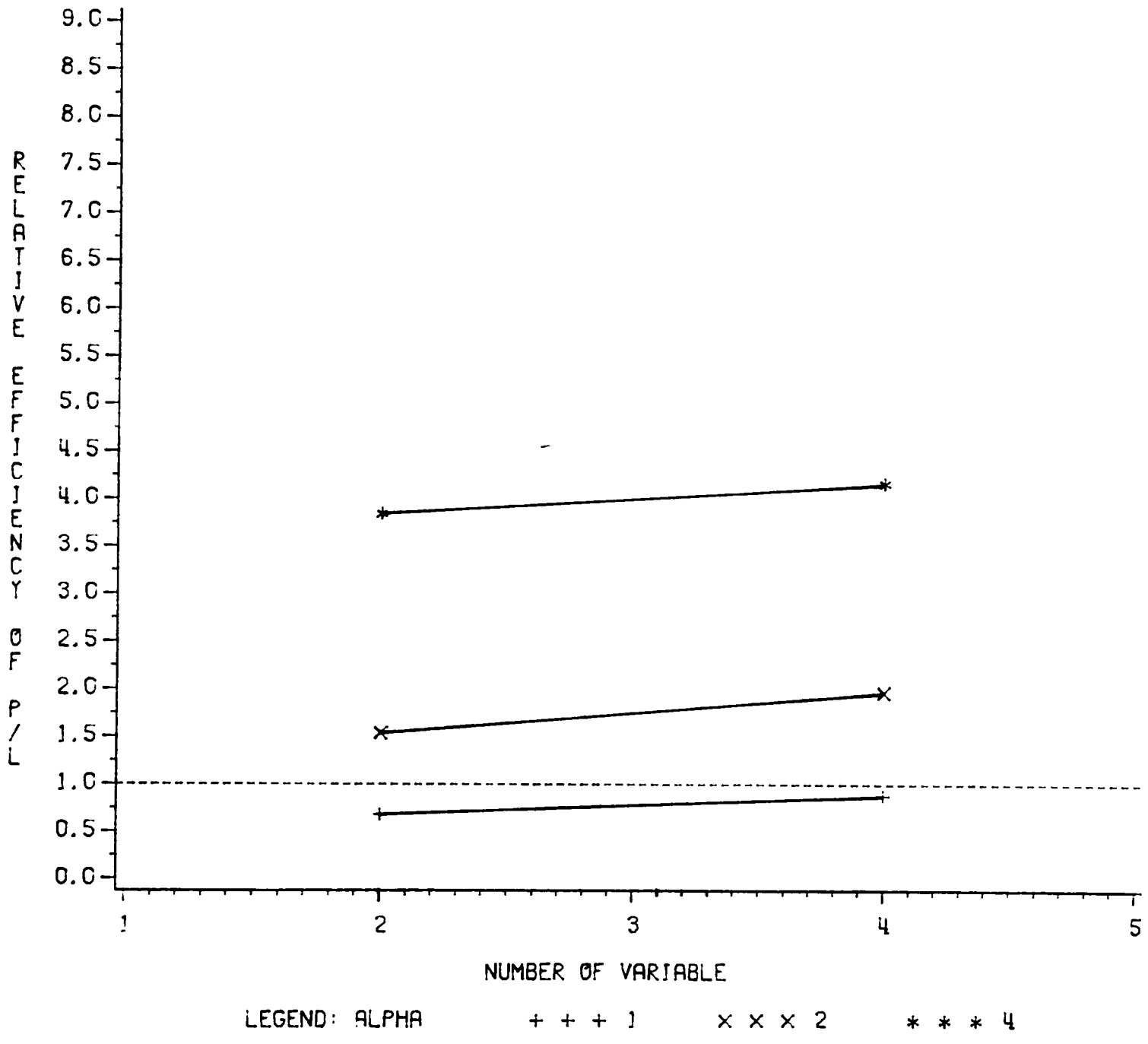



Figure 22: Relative Efficiency of the PQDF to the LDF with Skewness=1, $n_1=n_2=15$, and Distance=1

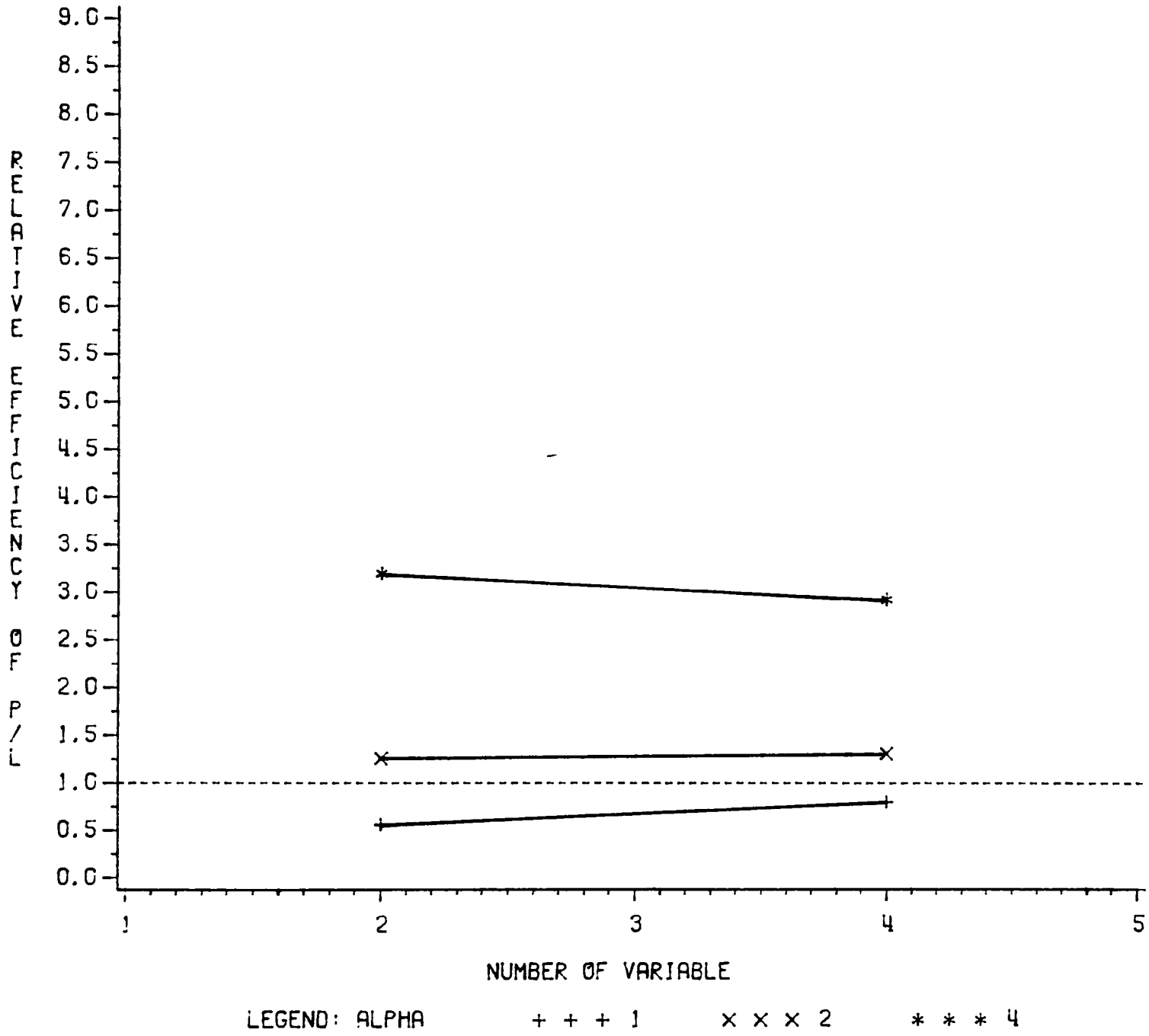



**Figure 23:** Relative Efficiency of the PQDF to the LDF with Skewness=1, $n_1=n_2=15$, and Distance=4

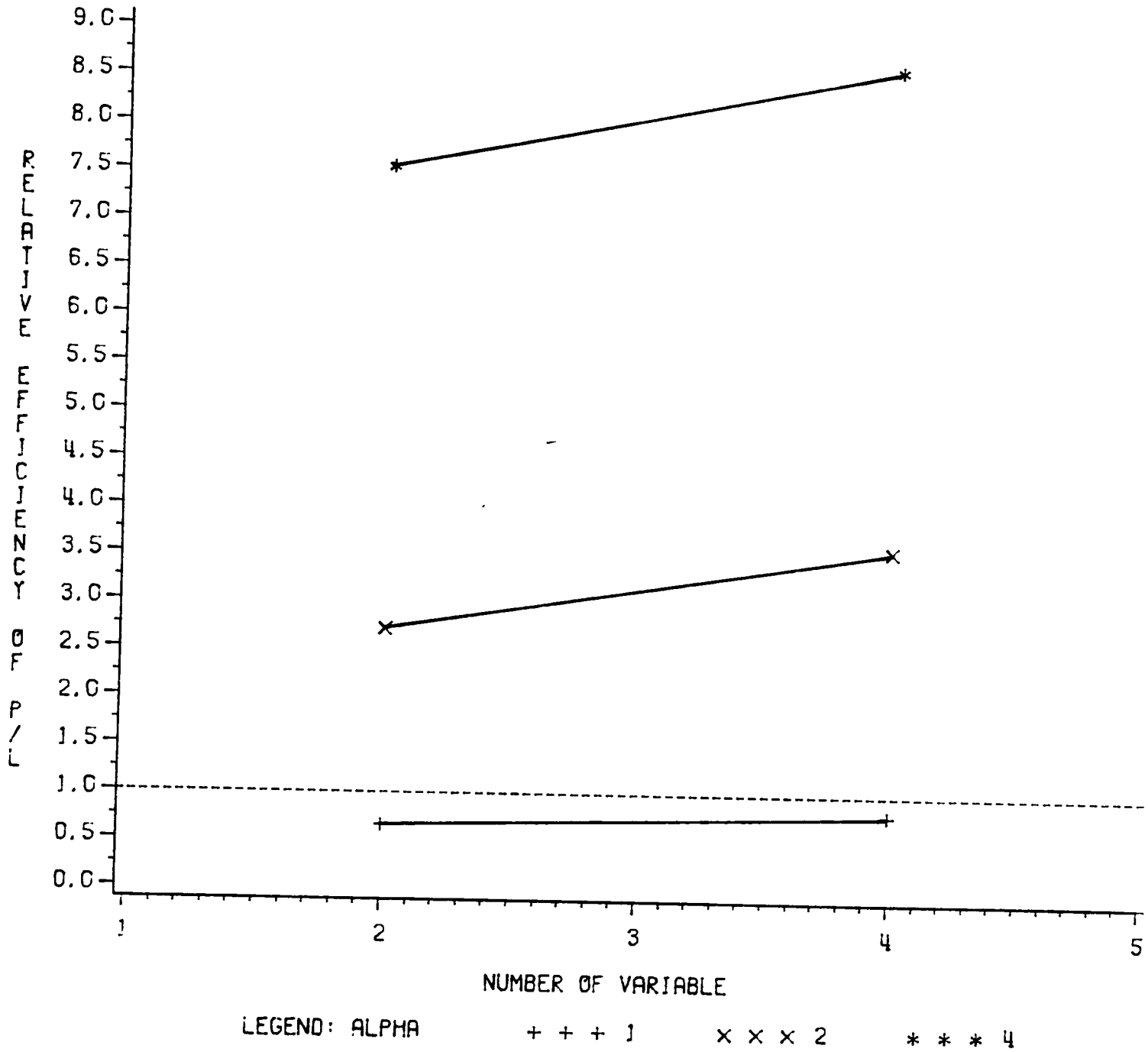



Figure 24: Relative Efficiency of the PQDF to the LDF with Skewness=1, $n_1=n_2=30$, and Distance=1

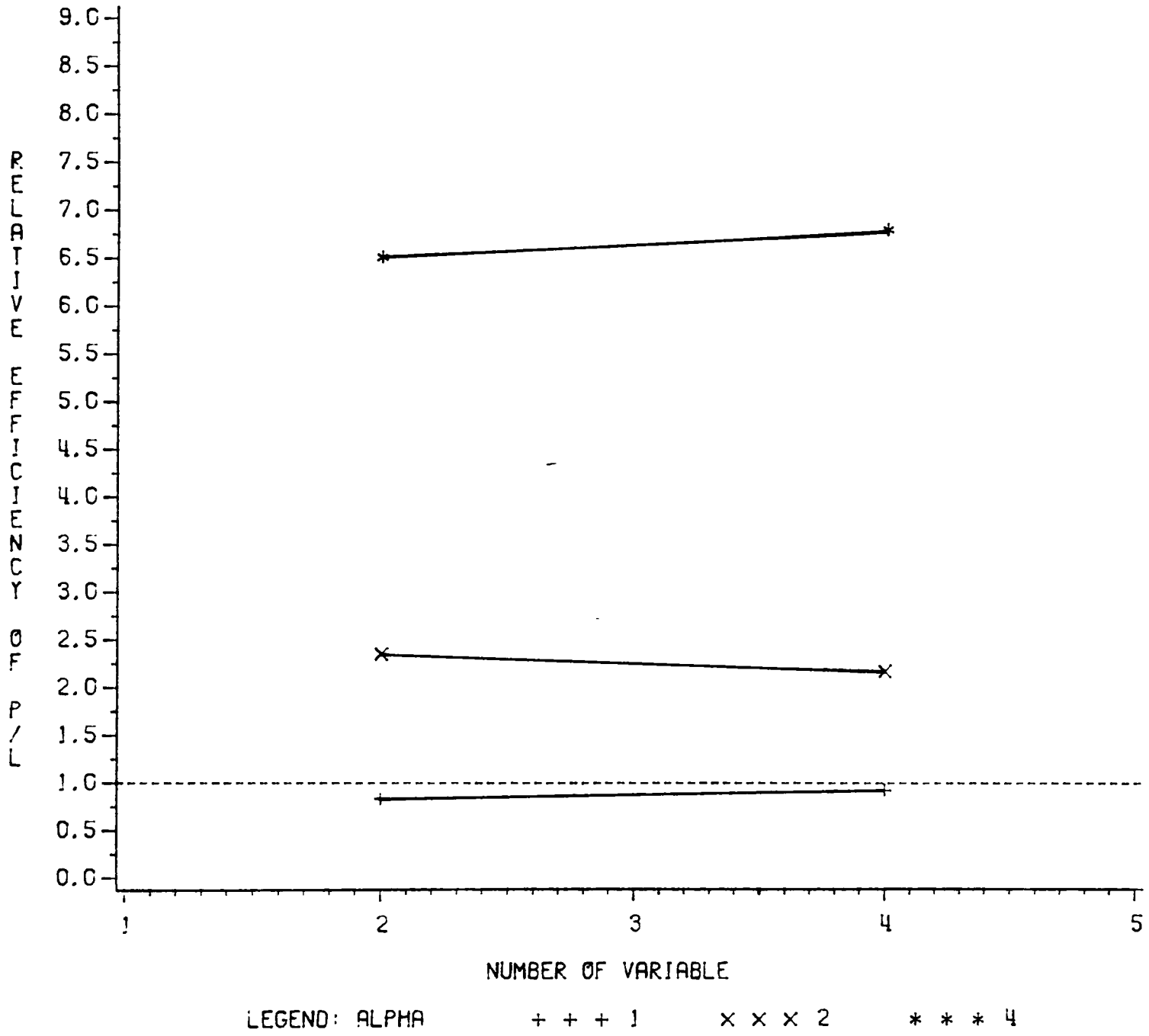



Figure 25: Relative Efficiency of the PQDF to the LDF with Skewness=1, $n_1=n_2=30$, and Distance=4

# RELATIVE EFFICIENCY WITH FINITE SAMPLE

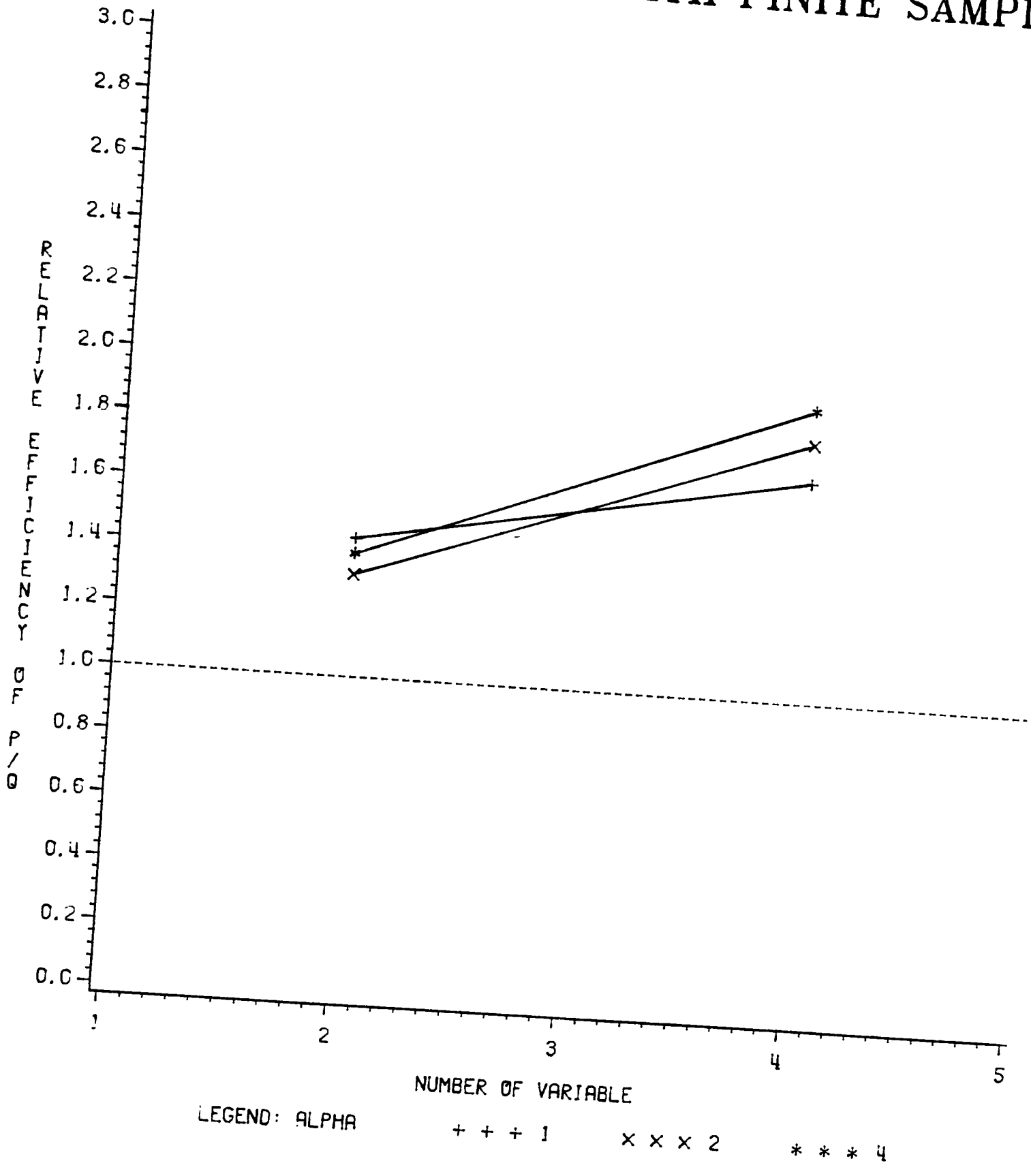



Figure 26: Relative Efficiency of the PQDF to the QDF with Skewness=1, $n_1=n_2=15$, and Distance=1

# RELATIVE EFFICIENCY WITH FINITE SAMPLE

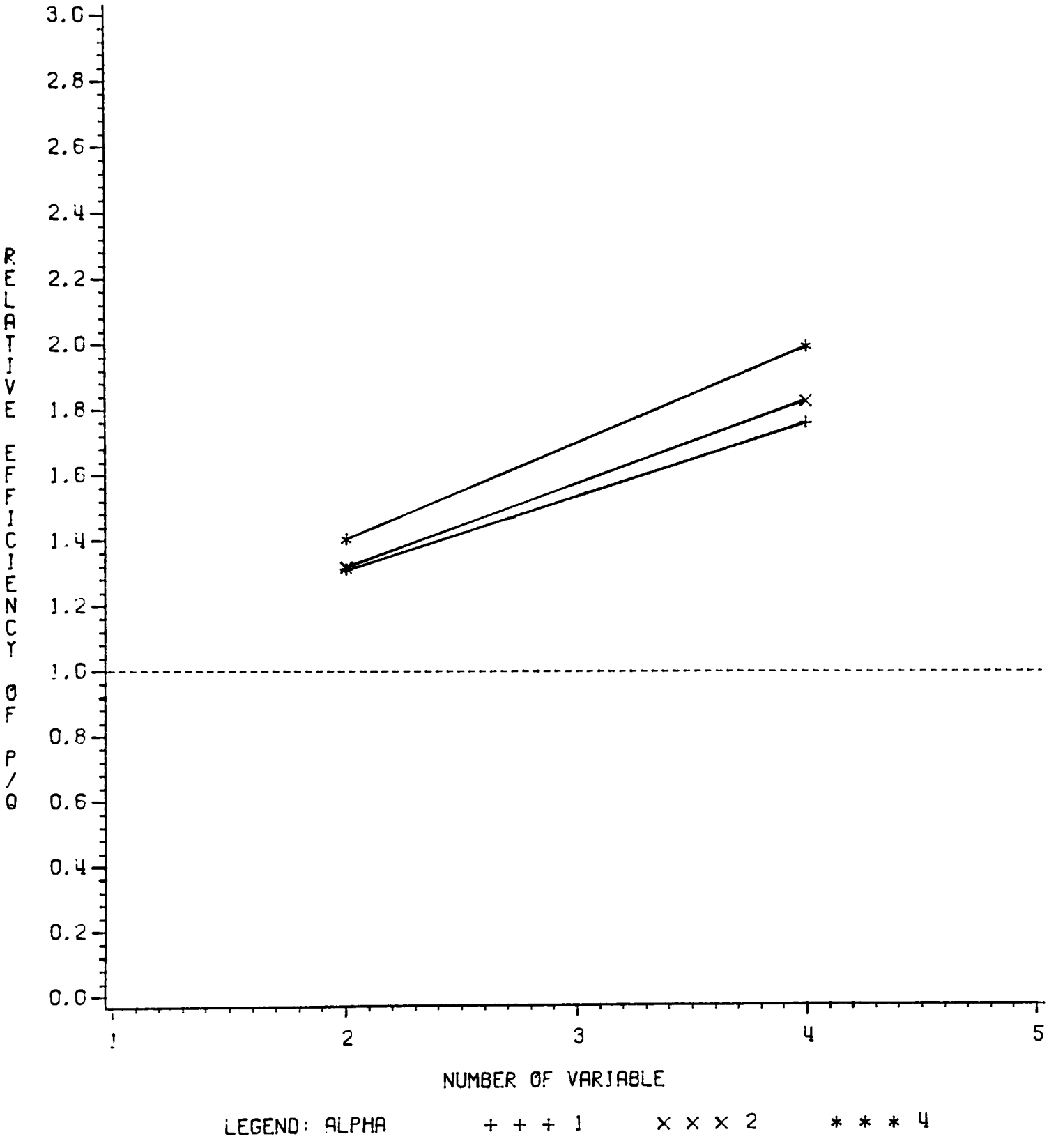



Figure 27: Relative Efficiency of the PQDF to the QDF with Skewness=1, $n_1=n_2=15$, and Distance=4

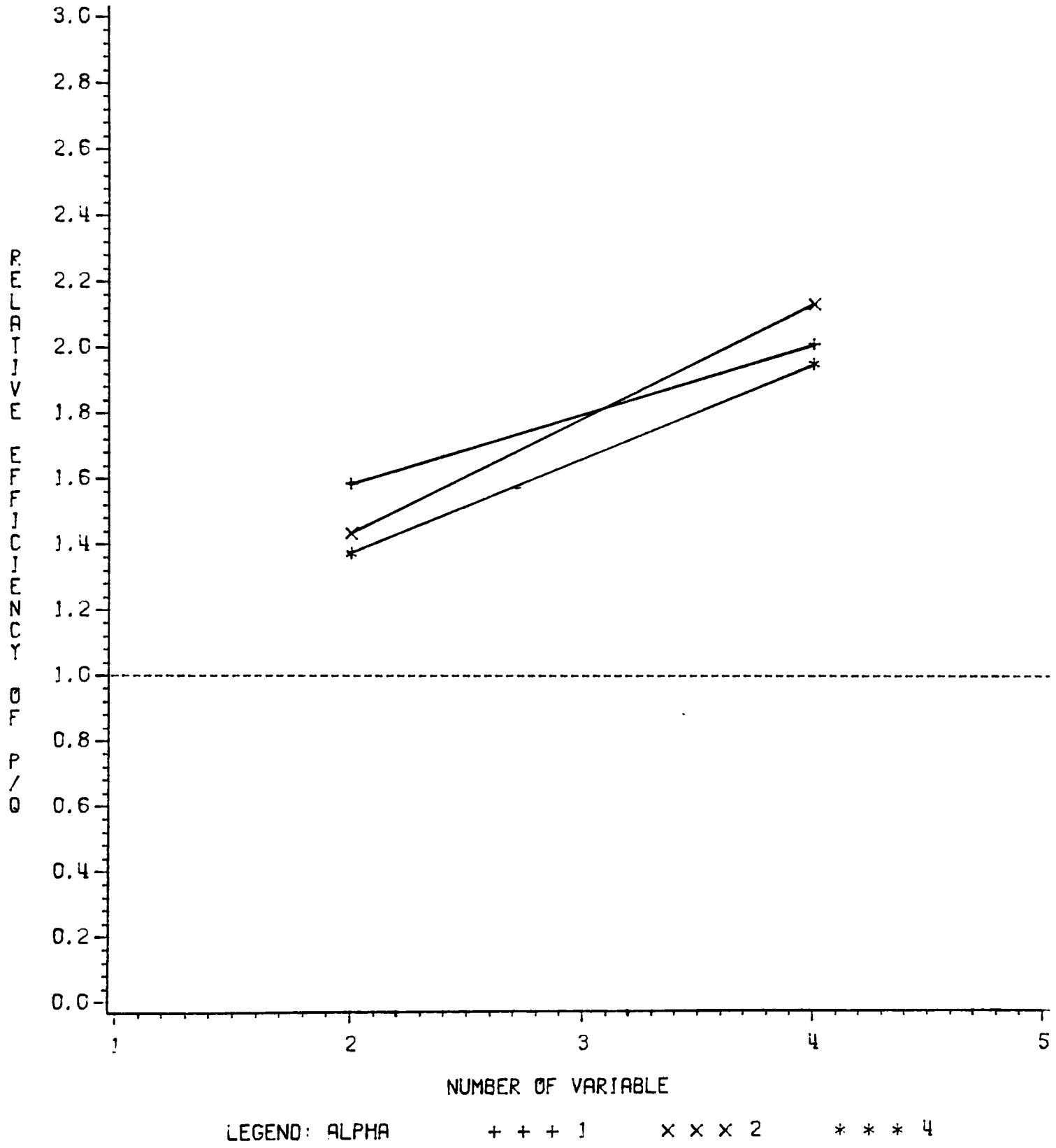



Figure 28: Relative Efficiency of the PQDF to the QDF with Skewness=1, $n_1=n_2=30$, and Distance=1

# RELATIVE EFFICIENCY WITH FINITE SAMPLE

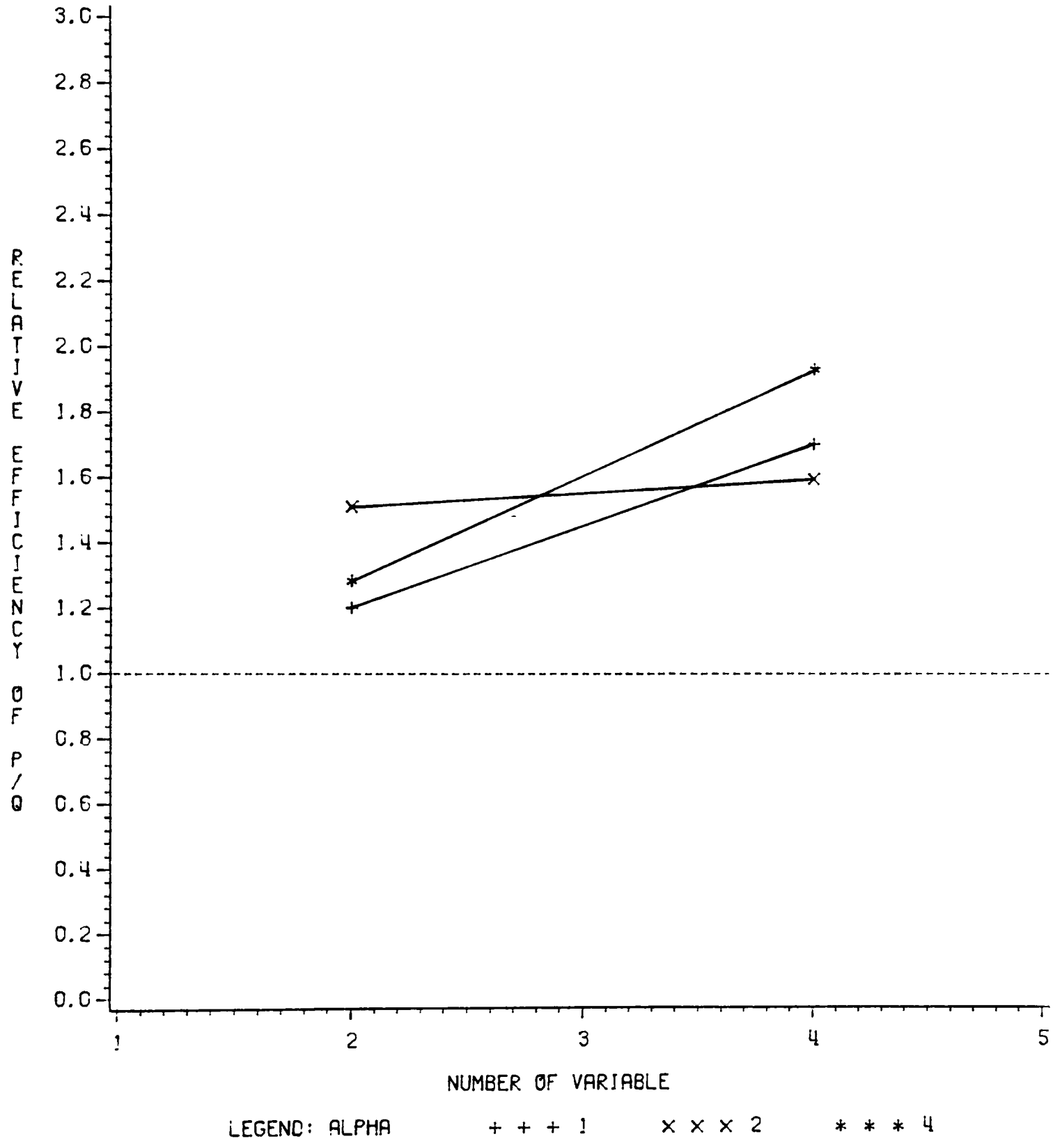



Figure 29: Relative Efficiency of the PQDF to the QDF with Skewness=1, $n_1=n_2=30$, and Distance=4

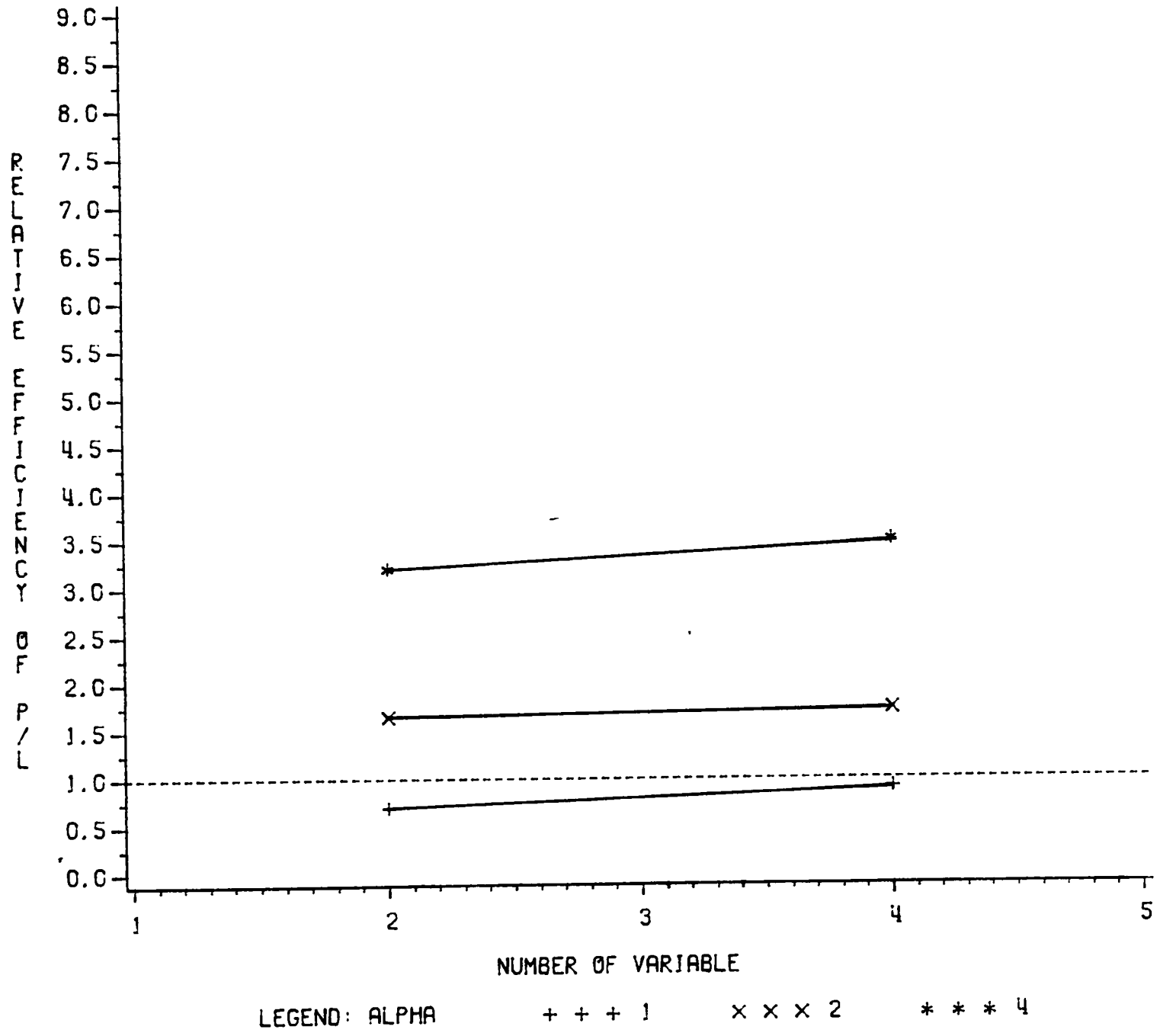



Figure 30: Relative Efficiency of the PQDF to the LDF with Skewness=3, $n_1=n_2=15$, and Distance=1

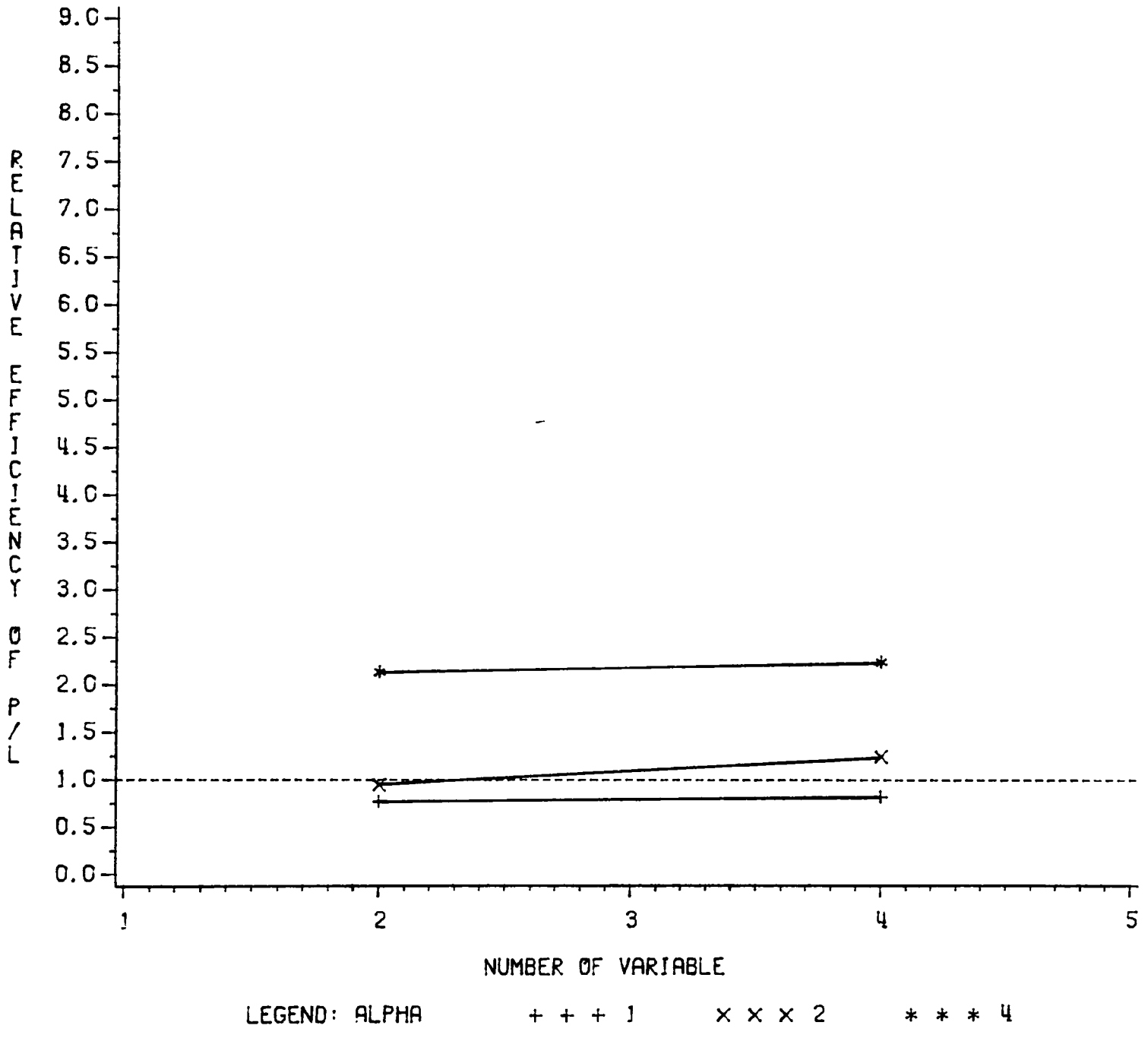



Figure 31: Relative Efficiency of the PQDF to the LDF with Skewness=3, $n_1=n_2=15$, and Distance=4

# RELATIVE EFFICIENCY WITH FINITE SAMPLE

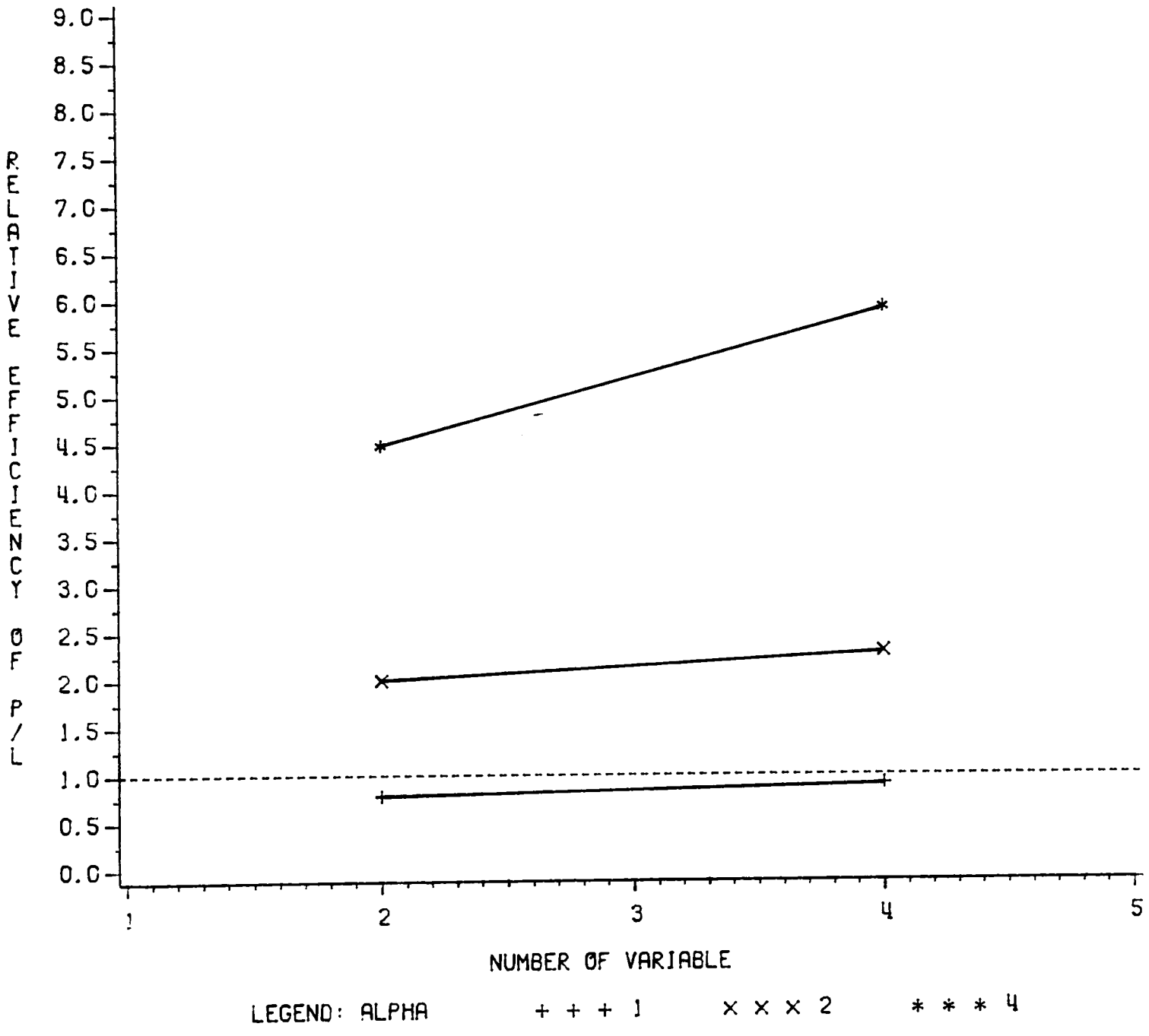



Figure 32: Relative Efficiency of the PQDF to the LDF with Skewness=3, $n_1=n_2=30$, and Distance=1

# RELATIVE EFFICIENCY WITH FINITE SAMPLE

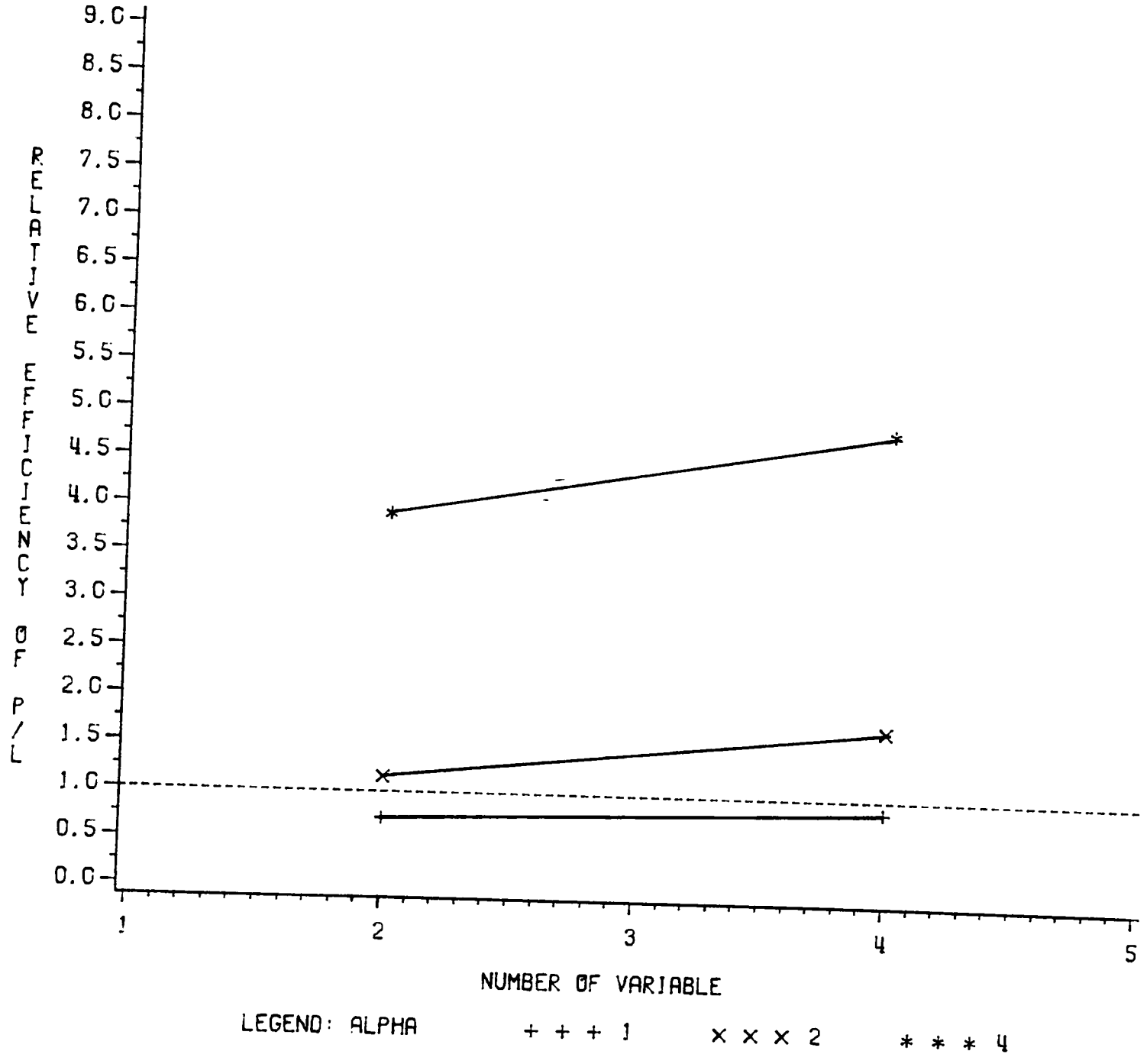



**Figure 33:** Relative Efficiency of the PQDF to the LDF with Skewness=3, $n_1=n_2=30$, and Distance=4

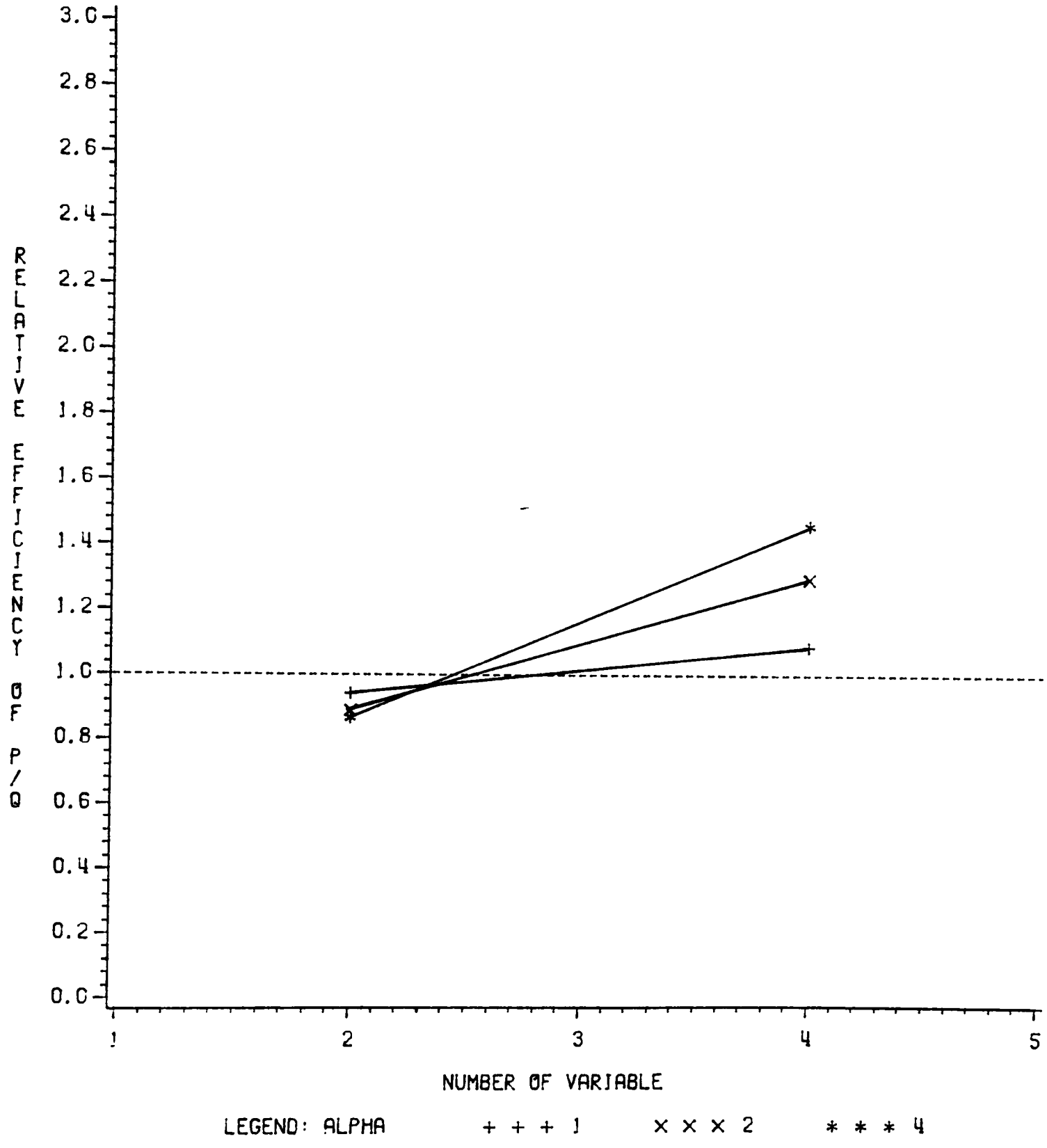



Figure 34: Relative Efficiency of the PQDF to the QDF with Skewness=3, $n_1=n_2=15$, and Distance=1

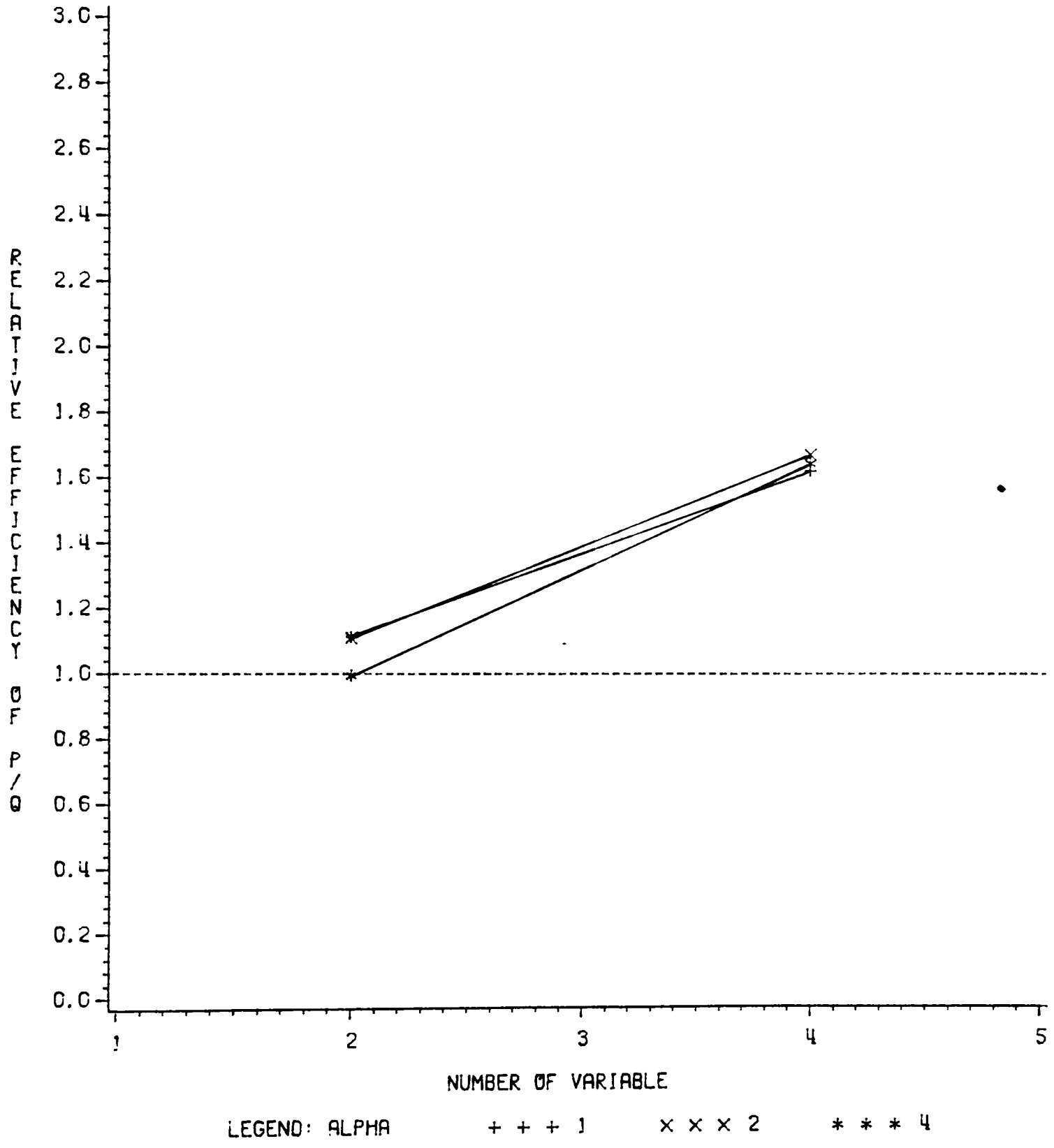



Figure 35: Relative Efficiency of the PQDF to the QDF with Skewness=3, $n_1=n_2=15$, and Distance=4

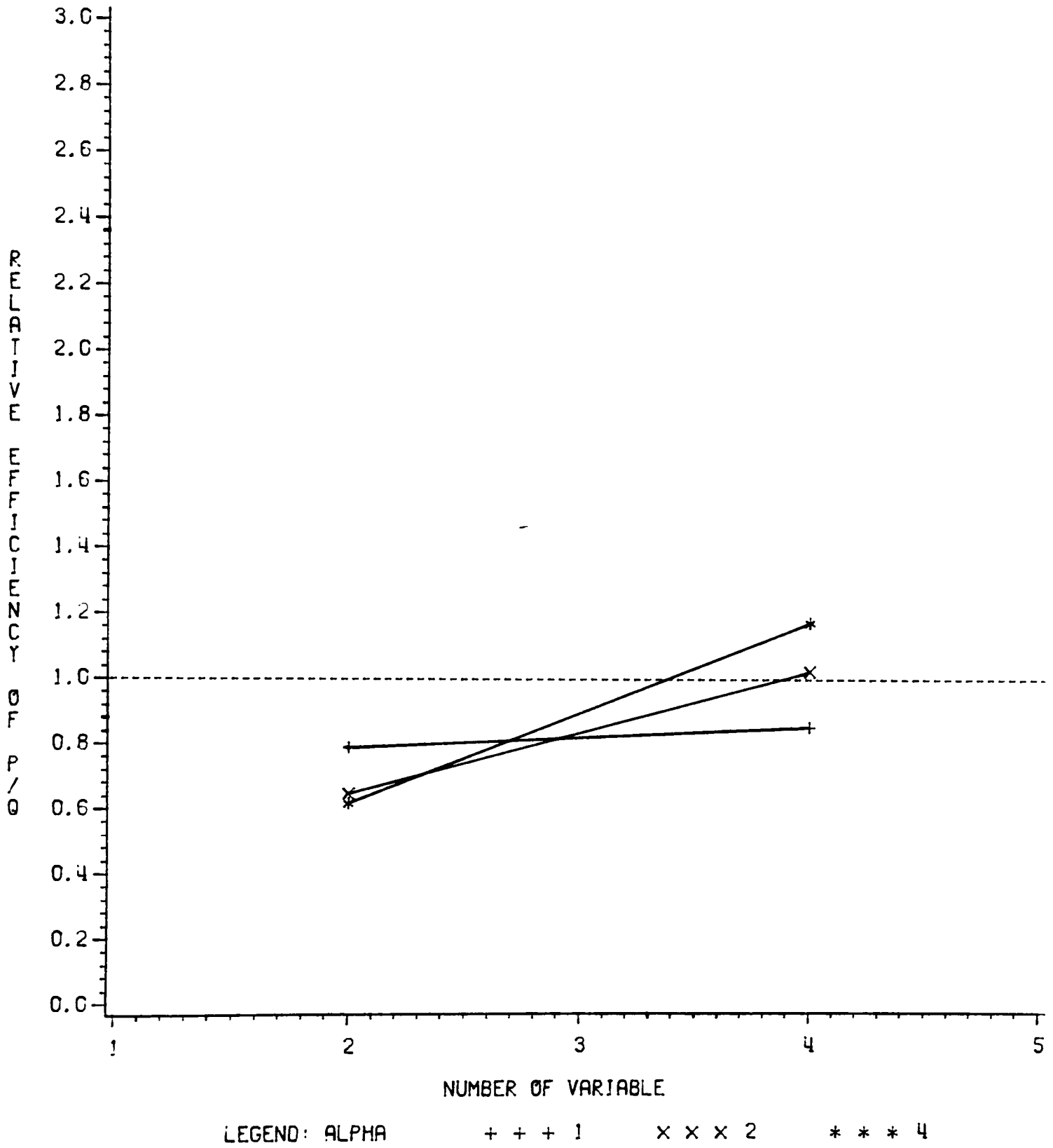



Figure 36: Relative Efficiency of the PQDF to the QDF with Skewness=3, $n_1=n_2=30$, and Distance=1

# RELATIVE EFFICIENCY WITH FINITE SAMPLE

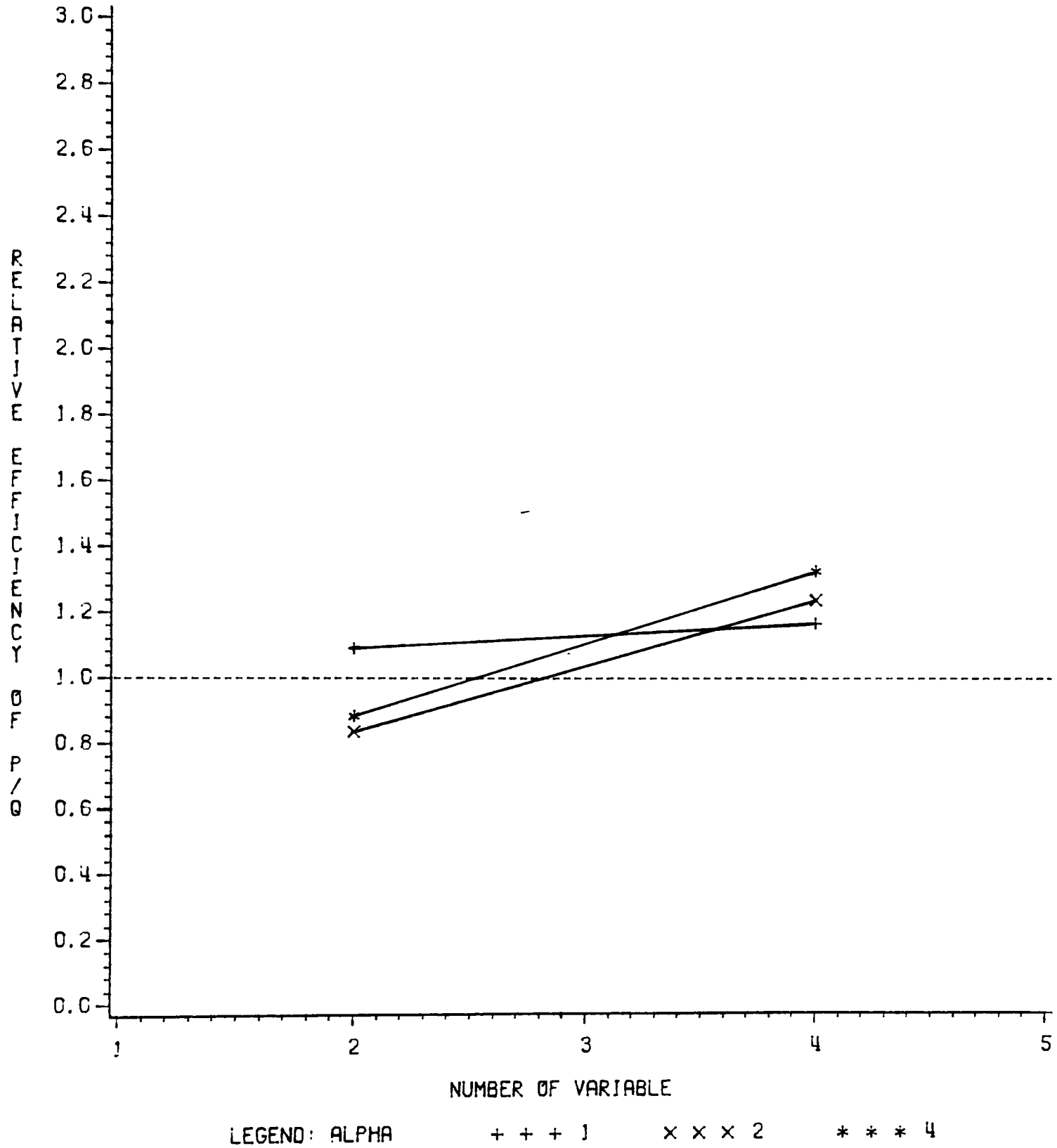
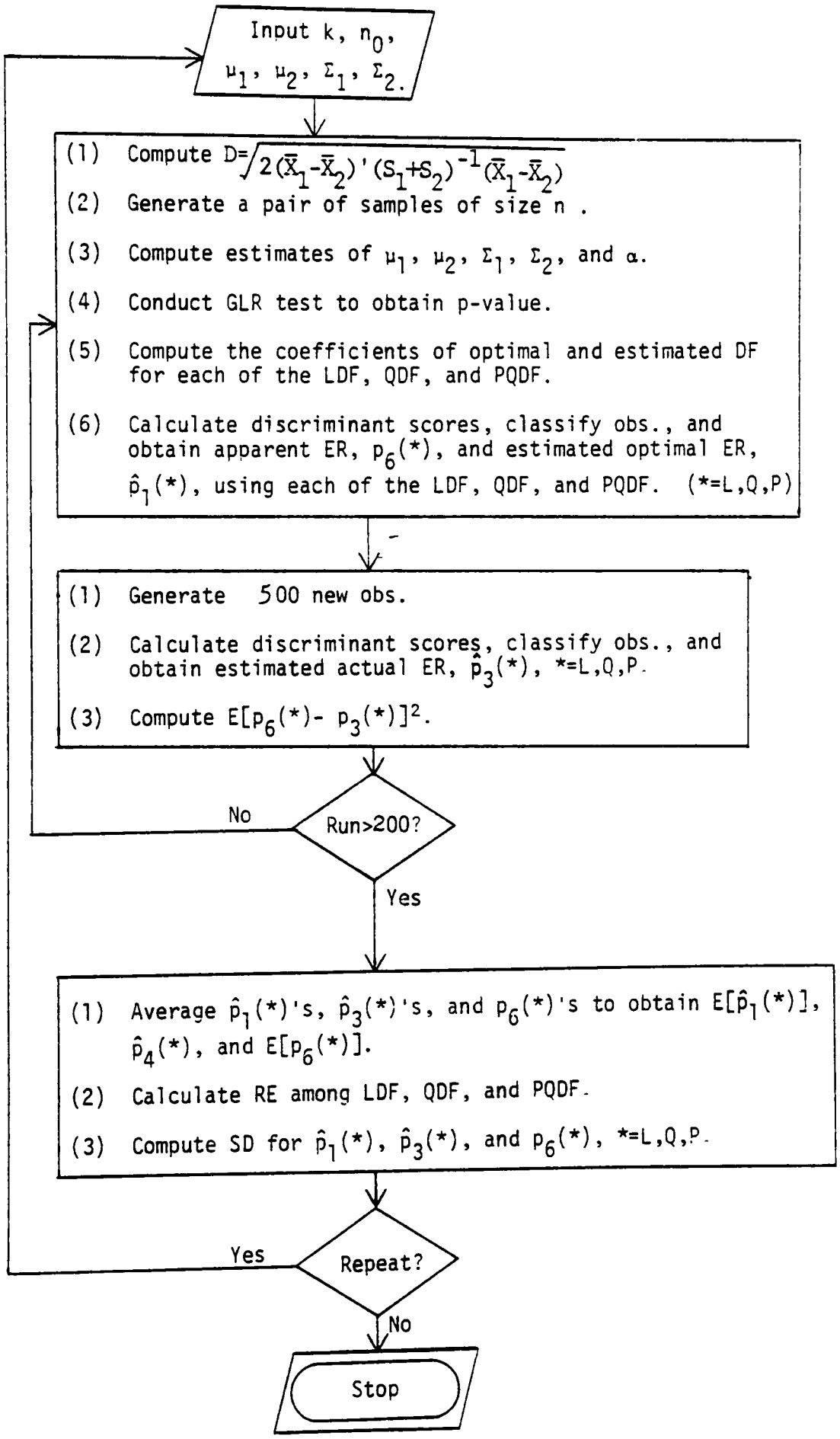



Figure 37: Relative Efficiency of the PQDF to the QDF with Skewness=3, $n_1=n_2=30$, and Distance=4

# APPENDIX F

## Flowchart for Monte Carlo Simulation



# APPENDIX G

## Tables of Results from Monte Carlo Simulation

TABLE 20

Simulation Results Using 200 Replications and Skewness=1

| Proportionality | d= | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|
| k n | △ SUBJECT | MEAN | STDERR | MEAN | STDERR | MEAN | STDERR |
| | APPERL | 0.3008 | 0.0062 | 0.2757 | 0.0061 | 0.2633 | 0.0058 |
| | APPERP | 0.2908 | 0.0060 | 0.2432 | 0.0054 | 0.1900 | 0.0050 |
| | APPERQ | 0.2827 | 0.0058 | 0.2313 | 0.0055 | 0.1812 | 0.0051 |
| | ACTERL | 0.3267 | 0.0023 | 0.3183 | 0.0022 | 0.2954 | 0.0020 |
| | ACTERP | 0.3359 | 0.0026 | 0.3037 | 0.0023 | 0.2316 | 0.0018 |
| 1 | ACTERQ | 0.3486 | 0.0027 | 0.3124 | 0.0023 | 0.2403 | 0.0019 |
| | OPTERR | 0.3068 | 0.0014 | 0.2766 | 0.0015 | 0.2092 | 0.0013 |
| | P(L/P) | 0.4837 | 0.0211 | 0.1695 | 0.0172 | 0.0114 | 0.0038 |
| | P(P/Q) | 0.4934 | 0.0211 | 0.4663 | 0.0211 | 0.4632 | 0.0208 |
| | DELTAH | 1.0746 | 0.0307 | 1.1710 | 0.0288 | 1.1299 | 0.0280 |
| 15 | ALPHAH | 1.0973 | 0.0306 | 2.1876 | 0.0609 | 4.4994 | 0.1286 |
| | APPERL | 0.0192 | 0.0019 | 0.0230 | 0.0018 | 0.0220 | 0.0018 |
| | APPERP | 0.0178 | 0.0019 | 0.0182 | 0.0017 | 0.0112 | 0.0014 |
| | APPERQ | 0.0163 | 0.0017 | 0.0173 | 0.0017 | 0.0093 | 0.0013 |
| | ACTERL | 0.0256 | 0.0006 | 0.0290 | 0.0008 | 0.0330 | 0.0009 |
| | ACTERP | 0.0286 | 0.0007 | 0.0272 | 0.0007 | 0.0203 | 0.0007 |
| 4 | ACTERQ | 0.0308 | 0.0008 | 0.0295 | 0.0009 | 0.0226 | 0.0009 |
| | OPTERR | 0.0217 | 0.0005 | 0.0201 | 0.0004 | 0.0145 | 0.0004 |
| | P(L/P) | 0.4943 | 0.0218 | 0.1783 | 0.0176 | 0.0115 | 0.0025 |
| | P(P/Q) | 0.4504 | 0.0198 | 0.4685 | 0.0202 | 0.4683 | 0.0215 |
| | DELTAH | 4.2139 | 0.0513 | 4.1650 | 0.0454 | 4.3325 | 0.0550 |
| 2 | ALPHAH | 1.0788 | 0.0350 | 2.2036 | 0.0637 | 4.2809 | 0.1237 |
| | APPERL | 0.3023 | 0.0038 | 0.2960 | 0.0042 | 0.2693 | 0.0044 |
| | APPERP | 0.3005 | 0.0037 | 0.2673 | 0.0040 | 0.2025 | 0.0036 |
| | APPERQ | 0.2962 | 0.0037 | 0.2643 | 0.0039 | 0.1993 | 0.0036 |
| | ACTERL | 0.3160 | 0.0017 | 0.3101 | 0.0018 | 0.2926 | 0.0020 |
| | ACTERP | 0.3198 | 0.0017 | 0.2892 | 0.0015 | 0.2219 | 0.0014 |
| 1 | ACTERQ | 0.3264 | 0.0021 | 0.2945 | 0.0018 | 0.2259 | 0.0014 |
| | OPTERR | 0.3085 | 0.0014 | 0.2771 | 0.0013 | 0.2111 | 0.0012 |
| | P(L/P) | 0.5098 | 0.0203 | 0.0435 | 0.0066 | 0.0003 | 0.0002 |
| | P(P/Q) | 0.5069 | 0.0206 | 0.5034 | 0.0216 | 0.4764 | 0.0207 |
| | DELTAH | 1.0459 | 0.0185 | 1.0499 | 0.0220 | 1.0680 | 0.0226 |
| 30 | ALPHAH | 1.0302 | 0.0192 | 2.1064 | 0.0360 | 4.1066 | 0.0733 |
| | APPERL | 0.0222 | 0.0014 | 0.0242 | 0.0013 | 0.0258 | 0.0013 |
| | APPERP | 0.0215 | 0.0014 | 0.0194 | 0.0013 | 0.0125 | 0.0010 |
| | APPERQ | 0.0216 | 0.0014 | 0.0191 | 0.0013 | 0.0120 | 0.0010 |
| | ACTERL | 0.0248 | 0.0005 | 0.0267 | 0.0006 | 0.0308 | 0.0007 |
| | ACTERP | 0.0251 | 0.0005 | 0.0230 | 0.0005 | 0.0167 | 0.0005 |
| 4 | ACTERQ | 0.0256 | 0.0006 | 0.0244 | 0.0006 | 0.0174 | 0.0005 |
| | OPTERR | 0.0228 | 0.0005 | 0.0202 | 0.0004 | 0.0141 | 0.0004 |
| | P(L/P) | 0.5443 | 0.0206 | 0.0646 | 0.0100 | 0.0002 | 0.0001 |
| | P(P/Q) | 0.5124 | 0.0204 | 0.4125 | 0.0200 | 0.5018 | 0.0206 |
| | DELTAH | 4.0729 | 0.0356 | 4.0740 | 0.0344 | 4.1354 | 0.0377 |
| | ALPHAH | 1.0020 | 0.0176 | 2.0731 | 0.0418 | 3.9643 | 0.0716 |

TABLE 20

(Continued)

| Proportionality d= | | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| k | n | △ SUBJECT | MEAN | STDERR | MEAN | STDERR | MEAN | STDERR |
| | | APPERL | 0.2432 | 0.0059 | 0.2468 | 0.0055 | 0.2362 | 0.0056 |
| | | APPERP | 0.2378 | 0.0055 | 0.1968 | 0.0052 | 0.1087 | 0.0041 |
| | | APPERQ | 0.1962 | 0.0052 | 0.1603 | 0.0048 | 0.0848 | 0.0039 |
| | | ACTERL | 0.3501 | 0.0033 | 0.3376 | 0.0027 | 0.3254 | 0.0031 |
| | | ACTERP | 0.3558 | 0.0032 | 0.2919 | 0.0025 | 0.1863 | 0.0018 |
| | 1 | ACTERQ | 0.3912 | 0.0032 | 0.3304 | 0.0029 | 0.2266 | 0.0027 |
| | | OPTERR | 0.3052 | 0.0014 | 0.2452 | 0.0013 | 0.1430 | 0.0010 |
| | | P(L/P) | 0.5053 | 0.0212 | 0.0860 | 0.0119 | 0.0004 | 0.0001 |
| | | P(P/Q) | 0.3277 | 0.0191 | 0.3722 | 0.0195 | 0.3649 | 0.0210 |
| | | DELTAH | 1.3566 | 0.0309 | 1.3234 | 0.0291 | 1.3623 | 0.0286 |
| | 15 | ALPHAH | 1.0587 | 0.0226 | 2.0522 | 0.0422 | 4.2380 | 0.0938 |
| | | APPERL | 0.0162 | 0.0018 | 0.0143 | 0.0016 | 0.0182 | 0.0017 |
| | | APPERP | 0.0162 | 0.0018 | 0.0102 | 0.0013 | 0.0068 | 0.0011 |
| | | APPERQ | 0.0113 | 0.0013 | 0.0070 | 0.0011 | 0.0055 | 0.0009 |
| | | ACTERL | 0.0342 | 0.0009 | 0.0357 | 0.0011 | 0.0409 | 0.0012 |
| | | ACTERP | 0.0370 | 0.0010 | 0.0317 | 0.0009 | 0.0210 | 0.0008 |
| | 4 | ACTERQ | 0.0479 | 0.0015 | 0.0426 | 0.0018 | 0.0313 | 0.0015 |
| | | OPTERR | 0.0229 | 0.0005 | 0.0187 | 0.0004 | 0.0108 | 0.0003 |
| | | P(L/P) | 0.4882 | 0.0219 | 0.0863 | 0.0134 | 0.0010 | 0.0006 |
| | | P(P/Q) | 0.3791 | 0.0205 | 0.3614 | 0.0203 | 0.3978 | 0.0209 |
| | | DELTAH | 4.3609 | 0.0502 | 4.5836 | 0.0532 | 4.5334 | 0.0628 |
| 4 | | ALPHAH | 1.0141 | 0.0226 | 2.0654 | 0.0410 | 4.0413 | 0.0809 |
| | | APPERL | 0.2819 | 0.0039 | 0.2776 | 0.0039 | 0.2597 | 0.0039 |
| | | APPERP | 0.2801 | 0.0038 | 0.2224 | 0.0040 | 0.1277 | 0.0033 |
| | | APPERQ | 0.2540 | 0.0037 | 0.2013 | 0.0035 | 0.1178 | 0.0032 |
| | | ACTERL | 0.3304 | 0.0021 | 0.3224 | 0.0018 | 0.3032 | 0.0021 |
| | | ACTERP | 0.3355 | 0.0020 | 0.2672 | 0.0017 | 0.1618 | 0.0013 |
| | 1 | ACTERQ | 0.3620 | 0.0025 | 0.2916 | 0.0020 | 0.1795 | 0.0016 |
| | | OPTERR | 0.3097 | 0.0015 | 0.2459 | 0.0013 | 0.1433 | 0.0011 |
| | | P(L/P) | 0.4643 | 0.0207 | 0.0112 | 0.0026 | 0.0000 | 0.0000 |
| | | P(P/Q) | 0.4429 | 0.0193 | 0.4385 | 0.0207 | 0.4585 | 0.0207 |
| | | DELTAH | 1.1423 | 0.0189 | 1.1491 | 0.0190 | 1.1413 | 0.0202 |
| | 30 | ALPHAH | 1.0123 | 0.0148 | 2.0292 | 0.0299 | 4.1532 | 0.0630 |
| | | APPERL | 0.0172 | 0.0012 | 0.0210 | 0.0013 | 0.0244 | 0.0013 |
| | | APPERP | 0.0169 | 0.0012 | 0.0173 | 0.0012 | 0.0095 | 0.0008 |
| | | APPERQ | 0.0147 | 0.0012 | 0.0158 | 0.0012 | 0.0085 | 0.0008 |
| | | ACTERL | 0.0265 | 0.0006 | 0.0296 | 0.0006 | 0.0352 | 0.0008 |
| | | ACTERP | 0.0269 | 0.0006 | 0.0236 | 0.0005 | 0.0144 | 0.0004 |
| | 4 | ACTERQ | 0.0306 | 0.0007 | 0.0266 | 0.0006 | 0.0177 | 0.0006 |
| | | OPTERR | 0.0218 | 0.0004 | 0.0185 | 0.0004 | 0.0108 | 0.0003 |
| | | P(L/P) | 0.5404 | 0.0200 | 0.0126 | 0.0035 | 0.0000 | 0.0000 |
| | | P(P/Q) | 0.4236 | 0.0216 | 0.4099 | 0.0204 | 0.4253 | 0.0210 |
| | | DELTAH | 4.2224 | 0.0347 | 4.2073 | 0.0362 | 4.1805 | 0.0377 |
| | | ALPHAH | 1.0198 | 0.0136 | 2.0432 | 0.0283 | 4.0969 | 0.0608 |

TABLE 21

Simulation Results Using 200 Replications and Skewness=2

| Proportionality d= | | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| k | n | △ SUBJECT | MEAN | STDERR | MEAN | STDERR | MEAN | STDERR |
| | | APPERL | 0.2863 | 0.0061 | 0.2843 | 0.0059 | 0.2602 | 0.0059 |
| | | APPERP | 0.2797 | 0.0060 | 0.2505 | 0.0055 | 0.1813 | 0.0045 |
| | | APPERQ | 0.2588 | 0.0056 | 0.2358 | 0.0054 | 0.1703 | 0.0047 |
| | | ACTERL | 0.3306 | 0.0028 | 0.3178 | 0.0023 | 0.3039 | 0.0021 |
| | | ACTERP | 0.3396 | 0.0030 | 0.2989 | 0.0022 | 0.2322 | 0.0018 |
| | 1 | ACTERQ | 0.3472 | 0.0029 | 0.3033 | 0.0023 | 0.2357 | 0.0020 |
| | | OPTERR | 0.3083 | 0.0014 | 0.2685 | 0.0015 | 0.2040 | 0.0012 |
| | | P(L/P) | 0.4831 | 0.0204 | 0.1853 | 0.0181 | 0.0069 | 0.0026 |
| | | P(P/Q) | 0.3178 | 0.0199 | 0.3171 | 0.0202 | 0.3272 | 0.0203 |
| | | DELTAH | 1.1248 | 0.0301 | 1.1310 | 0.0274 | 1.1921 | 0.0315 |
| | 15 | ALPHAH | 1.1547 | 0.0354 | 2.1953 | 0.0644 | 4.5710 | 0.1236 |
| | | APPERL | 0.0163 | 0.0016 | 0.0212 | 0.0018 | 0.0207 | 0.0018 |
| | | APPERP | 0.0145 | 0.0015 | 0.0155 | 0.0016 | 0.0095 | 0.0012 |
| | | APPERQ | 0.0137 | 0.0015 | 0.0148 | 0.0016 | 0.0091 | 0.0012 |
| | | ACTERL | 0.0267 | 0.0006 | 0.0293 | 0.0008 | 0.0331 | 0.0009 |
| | | ACTERP | 0.0293 | 0.0007 | 0.0281 | 0.0007 | 0.0199 | 0.0008 |
| | 4 | ACTERQ | 0.0313 | 0.0009 | 0.0297 | 0.0008 | 0.0217 | 0.0009 |
| | | OPTERR | 0.0217 | 0.0005 | 0.0208 | 0.0004 | 0.0142 | 0.0003 |
| | | P(L/P) | 0.4963 | 0.0205 | 0.1838 | 0.0169 | 0.0134 | 0.0034 |
| | | P(P/Q) | 0.3666 | 0.0214 | 0.3833 | 0.0210 | 0.3564 | 0.0201 |
| | | DELTAH | 4.2356 | 0.0457 | 4.1567 | 0.0524 | 4.2275 | 0.0506 |
| 2 | | ALPHAH | 1.0443 | 0.0323 | 2.1795 | 0.0662 | 4.1514 | 0.1330 |
| | | APPERL | 0.2974 | 0.0042 | 0.2936 | 0.0038 | 0.2878 | 0.0040 |
| | | APPERP | 0.2951 | 0.0040 | 0.2628 | 0.0035 | 0.2008 | 0.0036 |
| | | APPERQ | 0.2837 | 0.0037 | 0.2550 | 0.0037 | 0.1958 | 0.0037 |
| | | ACTERL | 0.3154 | 0.0016 | 0.3138 | 0.0017 | 0.3000 | 0.0020 |
| | | ACTERP | 0.3182 | 0.0016 | 0.2899 | 0.0017 | 0.2214 | 0.0014 |
| | 1 | ACTERQ | 0.3221 | 0.0017 | 0.2890 | 0.0017 | 0.2205 | 0.0014 |
| | | OPTERR | 0.3061 | 0.0013 | 0.2724 | 0.0013 | 0.2058 | 0.0012 |
| | | P(L/P) | 0.4941 | 0.0206 | 0.0612 | 0.0094 | 0.0002 | 0.0001 |
| | | P(P/Q) | 0.2378 | 0.0186 | 0.2279 | 0.0187 | 0.2794 | 0.0201 |
| | | DELTAH | 1.0726 | 0.0199 | 1.0591 | 0.0185 | 1.0353 | 0.0197 |
| | 30 | ALPHAH | 1.0472 | 0.0207 | 2.1058 | 0.0402 | 4.1430 | 0.0800 |
| | | APPERL | 0.0216 | 0.0013 | 0.0223 | 0.0015 | 0.0228 | 0.0012 |
| | | APPERP | 0.0212 | 0.0013 | 0.0193 | 0.0013 | 0.0129 | 0.0010 |
| | | APPERQ | 0.0203 | 0.0013 | 0.0178 | 0.0013 | 0.0119 | 0.0010 |
| | | ACTERL | 0.0255 | 0.0006 | 0.0271 | 0.0006 | 0.0292 | 0.0006 |
| | | ACTERP | 0.0259 | 0.0006 | 0.0244 | 0.0006 | 0.0177 | 0.0005 |
| | 4 | ACTERQ | 0.0269 | 0.0006 | 0.0248 | 0.0006 | 0.0181 | 0.0005 |
| | | OPTERR | 0.0226 | 0.0004 | 0.0204 | 0.0004 | 0.0144 | 0.0003 |
| | | P(L/P) | 0.5383 | 0.0192 | 0.0521 | 0.0071 | 0.0004 | 0.0002 |
| | | P(P/Q) | 0.2534 | 0.0191 | 0.2791 | 0.0200 | 0.2624 | 0.0183 |
| | | DELTAH | 4.0858 | 0.0333 | 4.1300 | 0.0355 | 4.1600 | 0.0351 |
| | | ALPHAH | 0.9838 | 0.0162 | 2.0487 | 0.0368 | 4.1607 | 0.0856 |

TABLE 21

(Continued)

| Proportionality d= | | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| k | n | △ SUBJECT | MEAN | STDERR | MEAN | STDERR | MEAN | STDERR |
| | | APPERL | 0.2630 | 0.0053 | 0.2480 | 0.0056 | 0.2455 | 0.0058 |
| | | APPERP | 0.2510 | 0.0052 | 0.1903 | 0.0055 | 0.1180 | 0.0042 |
| | | APPERQ | 0.1870 | 0.0049 | 0.1503 | 0.0047 | 0.0848 | 0.0037 |
| | | ACTERL | 0.3502 | 0.0026 | 0.3425 | 0.0028 | 0.3253 | 0.0028 |
| | | ACTERP | 0.3579 | 0.0028 | 0.2939 | 0.0025 | 0.1886 | 0.0020 |
| | 1 | ACTERQ | 0.3817 | 0.0027 | 0.3219 | 0.0028 | 0.2217 | 0.0027 |
| | | OPTERR | 0.3007 | 0.0013 | 0.2375 | 0.0013 | 0.1396 | 0.0011 |
| | | P(L/P) | 0.5104 | 0.0197 | 0.0892 | 0.0134 | 0.0003 | 0.0001 |
| | | P(P/Q) | 0.2651 | 0.0179 | 0.3158 | 0.0197 | 0.3062 | 0.0197 |
| | | DELTAH | 1.3083 | 0.0282 | 1.3625 | 0.0324 | 1.3077 | 0.0319 |
| | 15 | ALPHAH | 1.0528 | 0.0201 | 2.1168 | 0.0456 | 4.1123 | 0.0828 |
| | | APPERL | 0.0165 | 0.0018 | 0.0173 | 0.0018 | 0.0185 | 0.0016 |
| | | APPERP | 0.0153 | 0.0018 | 0.0120 | 0.0016 | 0.0062 | 0.0009 |
| | | APPERQ | 0.0117 | 0.0014 | 0.0102 | 0.0014 | 0.0057 | 0.0009 |
| | | ACTERL | 0.0337 | 0.0009 | 0.0367 | 0.0010 | 0.0408 | 0.0012 |
| | | ACTERP | 0.0353 | 0.0010 | 0.0314 | 0.0010 | 0.0226 | 0.0009 |
| | 4 | ACTERQ | 0.0484 | 0.0015 | 0.0418 | 0.0014 | 0.0314 | 0.0015 |
| | | OPTERR | 0.0223 | 0.0005 | 0.0188 | 0.0004 | 0.0109 | 0.0003 |
| | | P(L/P) | 0.4929 | 0.0203 | 0.1020 | 0.0130 | 0.0013 | 0.0006 |
| | | P(P/Q) | 0.3249 | 0.0206 | 0.2952 | 0.0182 | 0.3337 | 0.0195 |
| | | DELTAH | 4.3698 | 0.0562 | 4.4142 | 0.0570 | 4.5540 | 0.0693 |
| 4 | | ALPHAH | 1.0236 | 0.0212 | 1.9866 | 0.0397 | 4.3120 | 0.0948 |
| | | APPERL | 0.2790 | 0.0040 | 0.2802 | 0.0041 | 0.2603 | 0.0042 |
| | | APPERP | 0.2713 | 0.0040 | 0.2264 | 0.0037 | 0.1252 | 0.0030 |
| | | APPERQ | 0.2349 | 0.0038 | 0.2000 | 0.0037 | 0.1111 | 0.0031 |
| | | ACTERL | 0.3350 | 0.0019 | 0.3273 | 0.0019 | 0.3065 | 0.0021 |
| | | ACTERP | 0.3389 | 0.0021 | 0.2714 | 0.0016 | 0.1640 | 0.0012 |
| | 1 | ACTERQ | 0.3521 | 0.0022 | 0.2881 | 0.0019 | 0.1754 | 0.0015 |
| | | OPTERR | 0.3011 | 0.0013 | 0.2393 | 0.0012 | 0.1396 | 0.0010 |
| | | P(L/P) | 0.5118 | 0.0206 | 0.0072 | 0.0016 | 0.0000 | 0.0000 |
| | | P(P/Q) | 0.3051 | 0.0195 | 0.3073 | 0.0194 | 0.3250 | 0.0199 |
| | | DELTAH | 1.1708 | 0.0204 | 1.1245 | 0.0203 | 1.1794 | 0.0206 |
| | 30 | ALPHAH | 1.0060 | 0.0143 | 2.0439 | 0.0292 | 4.0740 | 0.0570 |
| | | APPERL | 0.0180 | 0.0012 | 0.0166 | 0.0011 | 0.0215 | 0.0012 |
| | | APPERP | 0.0179 | 0.0012 | 0.0129 | 0.0011 | 0.0081 | 0.0007 |
| | | APPERQ | 0.0162 | 0.0011 | 0.0119 | 0.0011 | 0.0073 | 0.0008 |
| | | ACTERL | 0.0274 | 0.0006 | 0.0295 | 0.0006 | 0.0347 | 0.0008 |
| | | ACTERP | 0.0277 | 0.0006 | 0.0241 | 0.0005 | 0.0151 | 0.0005 |
| | 4 | ACTERQ | 0.0309 | 0.0007 | 0.0266 | 0.0006 | 0.0175 | 0.0006 |
| | | OPTERR | 0.0226 | 0.0005 | 0.0185 | 0.0004 | 0.0105 | 0.0003 |
| | | P(L/P) | 0.5126 | 0.0200 | 0.0156 | 0.0053 | 0.0000 | 0.0000 |
| | | P(P/Q) | 0.3120 | 0.0200 | 0.3345 | 0.0209 | 0.3090 | 0.0201 |
| | | DELTAH | 4.2321 | 0.0341 | 4.2927 | 0.0372 | 4.2762 | 0.0356 |
| | | ALPHAH | 1.0092 | 0.0133 | 2.0449 | 0.0308 | 4.1187 | 0.0561 |

TABLE 22

Simulation Results Using 200 Replications and Skewness=3

| Proportionality d= | | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| k | n | △ SUBJECT | MEAN | STDERR | MEAN | STDERR | MEAN | STDERR |
| | | APPERL | 0.2860 | 0.0056 | 0.2905 | 0.0052 | 0.2845 | 0.0056 |
| | | APPERP | 0.2738 | 0.0055 | 0.2490 | 0.0053 | 0.1992 | 0.0052 |
| | | APPERQ | 0.2500 | 0.0054 | 0.2292 | 0.0054 | 0.1790 | 0.0047 |
| | | ACTERL | 0.3216 | 0.0021 | 0.3220 | 0.0021 | 0.3067 | 0.0022 |
| | | ACTERP | 0.3318 | 0.0028 | 0.2967 | 0.0022 | 0.2310 | 0.0017 |
| | 1 | ACTERQ | 0.3298 | 0.0025 | 0.2924 | 0.0024 | 0.2265 | 0.0019 |
| | | OPTERR | 0.2984 | 0.0014 | 0.2582 | 0.0013 | 0.1970 | 0.0011 |
| | | P(L/P) | 0.5062 | 0.0214 | 0.1881 | 0.0169 | 0.0099 | 0.0020 |
| | | P(P/Q) | 0.1884 | 0.0160 | 0.1776 | 0.0171 | 0.2031 | 0.0180 |
| | | DELTAH | 1.1155 | 0.0280 | 1.1327 | 0.0278 | 1.0918 | 0.0275 |
| | 15 | ALPHAH | 1.0563 | 0.0309 | 2.2234 | 0.0680 | 4.2165 | 0.1200 |
| | | APPERL | 0.0185 | 0.0016 | 0.0177 | 0.0017 | 0.0240 | 0.0019 |
| | | APPERP | 0.0178 | 0.0017 | 0.0155 | 0.0016 | 0.0120 | 0.0016 |
| | | APPERQ | 0.0153 | 0.0017 | 0.0128 | 0.0016 | 0.0085 | 0.0013 |
| | | ACTERL | 0.0282 | 0.0007 | 0.0268 | 0.0006 | 0.0305 | 0.0010 |
| | | ACTERP | 0.0300 | 0.0009 | 0.0272 | 0.0007 | 0.0223 | 0.0008 |
| | 4 | ACTERQ | 0.0309 | 0.0009 | 0.0280 | 0.0008 | 0.0222 | 0.0008 |
| | | OPTERR | 0.0223 | 0.0005 | 0.0199 | 0.0004 | 0.0150 | 0.0004 |
| | | P(L/P) | 0.5119 | 0.0191 | 0.1765 | 0.0171 | 0.0142 | 0.0035 |
| | | P(P/Q) | 0.1720 | 0.0159 | 0.1718 | 0.0157 | 0.2028 | 0.0180 |
| | | DELTAH | 4.2174 | 0.0463 | 4.2821 | 0.0543 | 4.1659 | 0.0515 |
| 2 | | ALPHAH | 1.0775 | 0.0307 | 2.2890 | 0.0684 | 4.4891 | 0.1344 |
| | | APPERL | 0.2990 | 0.0038 | 0.2956 | 0.0044 | 0.2842 | 0.0040 |
| | | APPERP | 0.2953 | 0.0040 | 0.2666 | 0.0042 | 0.2017 | 0.0037 |
| | | APPERQ | 0.2760 | 0.0037 | 0.2432 | 0.0041 | 0.1857 | 0.0032 |
| | | ACTERL | 0.3155 | 0.0016 | 0.3160 | 0.0020 | 0.3016 | 0.0017 |
| | | ACTERP | 0.3202 | 0.0018 | 0.2879 | 0.0017 | 0.2212 | 0.0015 |
| | 1 | ACTERQ | 0.3155 | 0.0018 | 0.2781 | 0.0017 | 0.2124 | 0.0015 |
| | | OPTERR | 0.2981 | 0.0014 | 0.2601 | 0.0014 | 0.1983 | 0.0012 |
| | | P(L/P) | 0.5079 | 0.0193 | 0.0699 | 0.0105 | 0.0002 | 0.0001 |
| | | P(P/Q) | 0.0734 | 0.0096 | 0.0767 | 0.0110 | 0.0870 | 0.0118 |
| | | DELTAH | 1.0560 | 0.0197 | 1.0638 | 0.0210 | 1.0774 | 0.0204 |
| | 30 | ALPHAH | 1.0309 | 0.0192 | 2.0642 | 0.0395 | 4.0613 | 0.0707 |
| | | APPERL | 0.0224 | 0.0014 | 0.0198 | 0.0013 | 0.0232 | 0.0013 |
| | | APPERP | 0.0224 | 0.0014 | 0.0182 | 0.0012 | 0.0135 | 0.0011 |
| | | APPERQ | 0.0210 | 0.0013 | 0.0170 | 0.0012 | 0.0125 | 0.0010 |
| | | ACTERL | 0.0244 | 0.0005 | 0.0262 | 0.0006 | 0.0279 | 0.0006 |
| | | ACTERP | 0.0253 | 0.0006 | 0.0255 | 0.0006 | 0.0184 | 0.0005 |
| | 4 | ACTERQ | 0.0256 | 0.0005 | 0.0247 | 0.0006 | 0.0181 | 0.0005 |
| | | OPTERR | 0.0219 | 0.0004 | 0.0208 | 0.0004 | 0.0152 | 0.0004 |
| | | P(L/P) | 0.5040 | 0.0199 | 0.0836 | 0.0107 | 0.0004 | 0.0002 |
| | | P(P/Q) | 0.0736 | 0.0096 | 0.0983 | 0.0126 | 0.0871 | 0.0123 |
| | | DELTAH | 4.0646 | 0.0322 | 4.1639 | 0.0353 | 4.1110 | 0.0353 |
| | | ALPHAH | 1.0794 | 0.0209 | 2.0481 | 0.0448 | 4.1174 | 0.0772 |

TABLE 22

(Continued)

| Proportionality d= | | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| k | n | △ SUBJECT | MEAN | STDERR | MEAN | STDERR | MEAN | STDERR |
| | | APPERL | 0.2550 | 0.0057 | 0.2560 | 0.0056 | 0.2398 | 0.0055 |
| | | APPERP | 0.2458 | 0.0056 | 0.2028 | 0.0056 | 0.1080 | 0.0040 |
| | | APPERQ | 0.1833 | 0.0050 | 0.1457 | 0.0046 | 0.0782 | 0.0035 |
| | | ACTERL | 0.3497 | 0.0026 | 0.3477 | 0.0025 | 0.3265 | 0.0029 |
| | | ACTERP | 0.3564 | 0.0029 | 0.2938 | 0.0024 | 0.1881 | 0.0020 |
| | 1 | ACTERQ | 0.3629 | 0.0028 | 0.3146 | 0.0031 | 0.2135 | 0.0027 |
| | | OPTERR | 0.2837 | 0.0015 | 0.2247 | 0.0013 | 0.1334 | 0.0010 |
| | | P(L/P) | 0.5072 | 0.0203 | 0.0875 | 0.0115 | 0.0004 | 0.0002 |
| | | P(P/Q) | 0.1967 | 0.0156 | 0.1891 | 0.0158 | 0.1826 | 0.0152 |
| | | DELTAH | 1.3204 | 0.0309 | 1.2754 | 0.0272 | 1.3440 | 0.0332 |
| | 15 | ALPHAH | 1.0706 | 0.0220 | 2.0483 | 0.0401 | 4.1916 | 0.0835 |
| | | APPERL | 0.0153 | 0.0016 | 0.0157 | 0.0017 | 0.0180 | 0.0017 |
| | | APPERP | 0.0148 | 0.0016 | 0.0100 | 0.0013 | 0.0052 | 0.0009 |
| | | APPERQ | 0.0105 | 0.0014 | 0.0075 | 0.0012 | 0.0047 | 0.0009 |
| | | ACTERL | 0.0333 | 0.0008 | 0.0361 | 0.0010 | 0.0392 | 0.0010 |
| | | ACTERP | 0.0358 | 0.0009 | 0.0325 | 0.0010 | 0.0232 | 0.0009 |
| | 4 | ACTERQ | 0.0447 | 0.0014 | 0.0422 | 0.0014 | 0.0315 | 0.0016 |
| | | OPTERR | 0.0217 | 0.0004 | 0.0182 | 0.0004 | 0.0104 | 0.0003 |
| | | P(L/P) | 0.5021 | 0.0205 | 0.1099 | 0.0138 | 0.0003 | 0.0001 |
| | | P(P/Q) | 0.2053 | 0.0171 | 0.1935 | 0.0152 | 0.2183 | 0.0169 |
| | | DELTAH | 4.3995 | 0.0572 | 4.4032 | 0.0565 | 4.3915 | 0.0561 |
| | 4 | ALPHAH | 1.0272 | 0.0227 | 2.0432 | 0.0457 | 4.2657 | 0.0898 |
| | | APPERL | 0.2851 | 0.0042 | 0.2848 | 0.0045 | 0.2645 | 0.0041 |
| | | APPERP | 0.2844 | 0.0043 | 0.2278 | 0.0038 | 0.1356 | 0.0035 |
| | | APPERQ | 0.2289 | 0.0038 | 0.1876 | 0.0034 | 0.1120 | 0.0030 |
| | | ACTERL | 0.3304 | 0.0018 | 0.3225 | 0.0018 | 0.3037 | 0.0020 |
| | | ACTERP | 0.3356 | 0.0019 | 0.2688 | 0.0017 | 0.1641 | 0.0012 |
| | 1 | ACTERQ | 0.3277 | 0.0021 | 0.2698 | 0.0019 | 0.1690 | 0.0014 |
| | | OPTERR | 0.2812 | 0.0013 | 0.2280 | 0.0013 | 0.1361 | 0.0009 |
| | | P(L/P) | 0.4964 | 0.0213 | 0.0172 | 0.0047 | 0.0000 | 0.0000 |
| | | P(P/Q) | 0.1238 | 0.0130 | 0.1337 | 0.0122 | 0.1247 | 0.0129 |
| | | DELTAH | 1.1222 | 0.0198 | 1.1213 | 0.0207 | 1.1607 | 0.0191 |
| | 30 | ALPHAH | 1.0188 | 0.0146 | 2.0392 | 0.0303 | 4.0059 | 0.0513 |
| | | APPERL | 0.0173 | 0.0012 | 0.0184 | 0.0013 | 0.0216 | 0.0013 |
| | | APPERP | 0.0168 | 0.0011 | 0.0160 | 0.0012 | 0.0093 | 0.0009 |
| | | APPERQ | 0.0148 | 0.0011 | 0.0136 | 0.0011 | 0.0073 | 0.0008 |
| | | ACTERL | 0.0281 | 0.0006 | 0.0296 | 0.0007 | 0.0343 | 0.0007 |
| | | ACTERP | 0.0290 | 0.0006 | 0.0248 | 0.0006 | 0.0160 | 0.0004 |
| | 4 | ACTERQ | 0.0304 | 0.0007 | 0.0264 | 0.0007 | 0.0176 | 0.0005 |
| | | OPTERR | 0.0212 | 0.0004 | 0.0183 | 0.0004 | 0.0114 | 0.0003 |
| | | P(L/P) | 0.4625 | 0.0209 | 0.0125 | 0.0040 | 0.0000 | 0.0000 |
| | | P(P/Q) | 0.1484 | 0.0132 | 0.1238 | 0.0135 | 0.1399 | 0.0145 |
| | | DELTAH | 4.2281 | 0.0380 | 4.2702 | 0.0397 | 4.2444 | 0.0369 |
| | | ALPHAH | 1.0340 | 0.0158 | 2.0501 | 0.0278 | 3.9855 | 0.0518 |

# APPENDIX H

## Pre-admission and Academic Achievement Data of the Medical Doctoral Program at Texas Tech University Health Science Center

| OBS | BCPM | OTHER | MCATBIO | MCATAVG | CUMAVG | GROUP |
|---|---|---|---|---|---|---|
| 1 | 3.49 | 3.80 | 10 | 8.33 | 71.34 | 1 |
| 2 | 3.76 | 3.80 | 9 | 7.67 | 72.51 | 1 |
| 3 | 3.50 | 3.62 | 7 | 7.67 | 74.08 | 1 |
| 4 | 2.86 | 3.40 | 9 | 9.33 | 74.09 | 1 |
| 5 | 2.40 | 2.85 | 7 | 7.83 | 74.83 | 1 |
| 6 | 3.95 | 3.91 | 7 | 6.50 | 74.84 | 1 |
| 7 | 3.34 | 3.30 | 7 | 6.50 | 74.94 | 1 |
| 8 | 2.59 | 3.31 | 7 | 6.33 | 75.36 | 1 |
| 9 | 3.68 | 6.90 | 6 | 7.67 | 76.76 | 1 |
| 10 | 3.16 | 3.91 | 11 | 9.50 | 76.86 | 1 |
| 11 | 3.13 | 3.74 | 7 | 6.83 | 77.01 | 1 |
| 12 | 2.53 | 3.22 | 6 | 5.33 | 77.08 | 1 |
| 13 | 3.44 | 3.73 | 6 | 8.00 | 77.14 | 1 |
| 14 | 3.32 | 3.79 | 7 | 6.67 | 77.29 | 1 |
| 15 | 3.94 | 3.73 | 10 | 7.83 | 77.59 | 1 |
| 16 | 2.69 | 3.52 | 10 | 9.33 | 77.72 | 1 |
| 17 | 3.08 | 3.70 | 7 | 7.33 | 77.84 | 1 |
| 18 | 3.66 | 3.25 | 10 | 8.67 | 78.31 | 1 |
| 19 | 3.31 | 3.20 | 9 | 9.33 | 78.58 | 1 |
| 20 | 3.82 | 3.47 | 9 | 9.00 | 78.58 | 1 |
| 21 | 3.53 | 3.87 | 6 | 8.00 | 78.61 | 1 |
| 22 | 3.52 | 4.00 | 8 | 7.17 | 78.61 | 1 |
| 23 | 3.57 | 3.74 | 9 | 7.67 | 78.62 | 1 |
| 24 | 3.48 | 3.91 | 11 | 8.67 | 78.64 | 1 |
| 25 | 3.90 | 3.74 | 6 | 8.83 | 78.68 | 1 |
| 26 | 3.49 | 3.85 | 9 | 8.17 | 78.68 | 1 |
| 27 | 3.26 | 3.50 | 9 | 9.17 | 78.86 | 1 |
| 28 | 3.21 | 3.90 | 10 | 8.17 | 78.90 | 1 |
| 29 | 3.31 | 3.65 | 11 | 8.83 | 78.96 | 1 |
| 30 | 3.20 | 2.98 | 11 | 9.67 | 79.12 | 1 |
| 31 | 3.49 | 3.92 | 8 | 9.33 | 79.12 | 1 |
| 32 | 3.40 | 3.52 | 7 | 8.33 | 79.16 | 1 |
| 33 | 3.71 | 3.78 | 9 | 7.33 | 79.28 | 1 |
| 34 | 2.90 | 3.19 | 9 | 9.00 | 79.29 | 1 |
| 35 | 3.23 | 3.39 | 12 | 9.83 | 79.36 | 1 |
| 36 | 3.46 | 3.24 | 9 | 8.67 | 79.37 | 1 |
| 37 | 3.46 | 3.90 | 7 | 8.50 | 79.43 | 1 |
| 38 | 3.27 | 3.08 | 10 | 8.67 | 79.53 | 1 |
| 39 | 3.38 | 3.64 | 7 | 8.00 | 79.53 | 1 |
| 40 | 2.87 | 3.58 | 10 | 8.83 | 79.80 | 1 |

| OBS | BCPM | OTHER | MCATBIO | MCATAVG | CUMAVG | GROUP |
|---|---|---|---|---|---|---|
| 41 | 3.06 | 3.54 | 9 | 9.67 | 79.91 | 1 |
| 42 | 3.06 | 3.36 | 9 | 7.83 | 79.92 | 1 |
| 43 | 2.98 | 3.14 | 7 | 6.67 | 79.96 | 1 |
| 44 | 3.47 | 2.86 | 7 | 8.67 | 80.16 | 1 |
| 45 | 3.22 | 3.40 | 8 | 8.50 | 80.22 | 1 |
| 46 | 2.98 | 3.71 | 7 | 7.17 | 80.22 | 1 |
| 47 | 3.51 | 3.62 | 8 | 7.67 | 80.22 | 1 |
| 48 | 2.87 | 3.17 | 9 | 9.50 | 80.26 | 1 |
| 49 | 3.05 | 3.75 | 6 | 9.83 | 80.42 | 1 |
| 50 | 2.90 | 3.38 | 9 | 7.67 | 80.57 | 1 |
| 51 | 3.05 | 3.50 | 7 | 7.50 | 80.58 | 1 |
| 52 | 4.00 | 4.00 | 10 | 9.83 | 80.61 | 1 |
| 53 | 3.48 | 3.55 | 10 | 8.17 | 80.61 | 1 |
| 54 | 3.38 | 3.85 | 11 | 7.83 | 80.82 | 1 |
| 55 | 3.17 | 3.33 | 6 | 8.67 | 81.09 | 1 |
| 56 | 3.84 | 4.00 | 6 | 8.50 | 81.13 | 1 |
| 57 | 2.92 | 2.87 | 10 | 10.67 | 81.18 | 1 |
| 58 | 3.80 | 4.00 | 7 | 7.50 | 81.21 | 1 |
| 59 | 3.03 | 3.23 | 7 | 8.83 | 81.26 | 1 |
| 60 | 3.91 | 3.89 | 7 | 10.00 | 81.34 | 1 |
| 61 | 3.76 | 4.00 | 6 | 8.67 | 81.39 | 1 |
| 62 | 3.37 | 3.64 | 9 | 7.67 | 81.46 | 1 |
| 63 | 3.40 | 3.13 | 6 | 7.50 | 81.55 | 1 |
| 64 | 2.84 | 2.92 | 11 | 10.00 | 81.72 | 1 |
| 65 | 3.22 | 3.65 | 9 | 8.17 | 81.74 | 1 |
| 66 | 3.28 | 3.72 | 8 | 8.50 | 81.96 | 2 |
| 67 | 3.51 | 3.94 | 10 | 8.33 | 81.97 | 2 |
| 68 | 3.13 | 3.74 | 9 | 8.50 | 82.08 | 2 |
| 69 | 3.17 | 3.82 | 10 | 8.33 | 82.20 | 2 |
| 70 | 3.76 | 3.56 | 7 | 7.83 | 82.20 | 2 |
| 71 | 3.80 | 3.94 | 8 | 8.17 | 82.29 | 2 |
| 72 | 2.57 | 2.91 | 9 | 9.00 | 82.56 | 2 |
| 73 | 3.23 | 3.53 | 9 | 9.00 | 82.68 | 2 |
| 74 | 3.40 | 3.54 | 6 | 6.50 | 82.68 | 2 |
| 75 | 3.29 | 3.51 | 9 | 8.17 | 82.70 | 2 |
| 76 | 3.01 | 3.84 | 10 | 8.17 | 82.75 | 2 |
| 77 | 3.06 | 2.91 | 10 | 10.67 | 82.82 | 2 |
| 78 | 3.53 | 3.77 | 8 | 7.50 | 82.83 | 2 |
| 79 | 3.62 | 2.65 | 9 | 9.67 | 82.95 | 2 |
| 80 | 3.47 | 3.66 | 9 | 9.00 | 83.00 | 2 |
| 81 | 3.10 | 3.51 | 10 | 10.67 | 83.03 | 2 |
| 82 | 3.83 | 3.95 | 9 | 8.17 | 83.05 | 2 |
| 83 | 3.30 | 3.65 | 10 | 8.50 | 83.16 | 2 |
| 84 | 2.91 | 2.00 | 6 | 6.17 | 83.16 | 2 |
| 85 | 3.47 | 3.84 | 8 | 8.00 | 83.26 | 2 |

| OBS | BCPM | OTHER | MCATBID | MCATAVG | CUMAVG | GROUP |
|---|---|---|---|---|---|---|
| 86 | 3.47 | 2.92 | 10 | 8.67 | 83.29 | 2 |
| 87 | 3.25 | 3.47 | 7 | 8.67 | 83.30 | 2 |
| 88 | 3.76 | 3.84 | 9 | 9.67 | 83.41 | 2 |
| 89 | 3.29 | 3.92 | 11 | 8.83 | 83.50 | 2 |
| 90 | 3.26 | 3.08 | 11 | 11.17 | 83.58 | 2 |
| 91 | 3.77 | 3.16 | 11 | 9.17 | 83.74 | 2 |
| 92 | 3.15 | 3.54 | 10 | 10.67 | 83.76 | 2 |
| 93 | 3.77 | 3.70 | 8 | 9.00 | 83.79 | 2 |
| 94 | 3.75 | 4.00 | 7 | 7.67 | 83.95 | 2 |
| 95 | 3.33 | 3.68 | 10 | 9.50 | 84.12 | 2 |
| 96 | 3.17 | 3.17 | 6 | 8.00 | 84.18 | 2 |
| 97 | 2.57 | 2.96 | 8 | 9.00 | 84.21 | 2 |
| 98 | 3.56 | 4.00 | 9 | 8.17 | 84.39 | 2 |
| 99 | 3.78 | 3.71 | 7 | 8.00 | 84.47 | 2 |
| 100 | 3.00 | 3.69 | 6 | 7.00 | 84.91 | 2 |
| 101 | 3.00 | 3.73 | 9 | 8.67 | 84.92 | 2 |
| 102 | 2.96 | 2.96 | 11 | 10.00 | 84.96 | 2 |
| 103 | 2.95 | 3.50 | 9 | 8.33 | 85.00 | 2 |
| 104 | 3.51 | 3.62 | 8 | 9.00 | 85.16 | 2 |
| 105 | 3.59 | 3.18 | 9 | 8.67 | 85.16 | 2 |
| 106 | 3.01 | 3.31 | 12 | 9.50 | 85.29 | 2 |
| 107 | 3.47 | 3.58 | 9 | 8.33 | 85.33 | 2 |
| 108 | 3.75 | 3.38 | 9 | 8.83 | 85.47 | 2 |
| 109 | 3.57 | 3.83 | 7 | 8.33 | 85.70 | 2 |
| 110 | 3.09 | 3.29 | 9 | 8.33 | 85.80 | 2 |
| 111 | 3.96 | 3.46 | 9 | 8.50 | 86.08 | 2 |
| 112 | 3.26 | 3.75 | 10 | 8.43 | 86.36 | 2 |
| 113 | 3.73 | 3.53 | 6 | 5.83 | 86.80 | 2 |
| 114 | 3.37 | 3.95 | 8 | 7.17 | 87.08 | 2 |
| 115 | 3.67 | 3.88 | 9 | 8.83 | 87.21 | 2 |
| 116 | 3.24 | 3.08 | 11 | 10.17 | 87.36 | 2 |
| 117 | 3.56 | 4.00 | 10 | 9.00 | 88.08 | 2 |
| 118 | 3.15 | 3.70 | 10 | 8.33 | 88.54 | 2 |
| 119 | 3.63 | 3.55 | 8 | 8.33 | 88.57 | 2 |
| 120 | 3.62 | 4.00 | 6 | 7.17 | 88.58 | 2 |
| 121 | 3.87 | 3.75 | 9 | 9.17 | 89.17 | 2 |
| 122 | 3.74 | 3.94 | 8 | 7.83 | 89.33 | 2 |
| 123 | 3.93 | 4.00 | 10 | 8.50 | 89.92 | 2 |
| 124 | 3.41 | 3.42 | 10 | 8.83 | 89.96 | 2 |
| 125 | 2.90 | 2.88 | 11 | 10.00 | 90.05 | 2 |
| 126 | 3.04 | 3.82 | 11 | 9.33 | 91.00 | 2 |
| 127 | 3.90 | 4.00 | 12 | 12.17 | 91.74 | 2 |
| 128 | 3.52 | 3.29 | 10 | 8.50 | 91.79 | 2 |
| 129 | 3.00 | 3.65 | 8 | 9.00 | 92.03 | 2 |
| 130 | 4.00 | 4.00 | 8 | 8.50 | 92.55 | 2 |