AN EVALUATION OF KENDALL'S ORDER-STATISTIC METHOD OF DISCRIMINANT ANALYSIS AND RELATED STUDIES

by

Stewart F. Musket

Technical Report No. 112
Department of Statistics ONR Contract

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75222

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) SOUTHERN METHODIST UNIVERSITY	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP UNCLASSIFIED

3. REPORT TITLE

An Evaluation of Kendall's Order-statistic Method of Discriminant Analysis and Related Studies

4. DESCRIPTIVE NOTES (*Type of report and inclusive dates*)

Technical Report

5. AUTHOR(S) (*First name, middle initial, last name*)

Stewart F. Musket

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November 27, 1971	108	33

8a. CONTRACT OR GRANT NO N00014-68-A-0515 b. PROJECT NO NR 042-260 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 112 9b. OTHER REPORT NO(S) (*Any other numbers that may be assigned this report*)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research

13. ABSTRACT

Methods of nonparametric discriminant analysis, an example of which is Kendall's order-statistic method, are of interest because two of the assumptions of linear discriminant analysis, equality of variance-covariance matrices and multivariate normality, are not acceptable in many applications. A number of sampling experiments were preformed in order to evaluate Kendall's method and to compare it with linear discriminant analysis on the basis of the expected values of the probability of misclassification. Two p-variate (p=5) populations, Π_1 and Π_2, were considered and in most of the experiments the populations were assumed multivariate normal with variance-covariance matrices $\Sigma_1 = [(1-\rho_1)I + \rho_1 Epp]$ and $\Sigma_2 = \sigma^2[(1-\rho_2)I + \rho_2 Epp]$, where Epp is a pxp matrix of 1's. A few experiments were performed also with multivariate Cauchy, uniform, and lognormal distributions (all variables independent).

In some of the sampling experiments $\Sigma_1 = \sigma^2\Sigma_2$, that is, one variance-covariance matrix is a multiple of the other. In these cases the expected values of the probability of misclassification are compared with the theoretical probabilities of misclassification determined by Gilbert (Biometrics, 25, 1969, pp. 505-515).

Bartlett and Please (Biometrika, 50, 1963, pp. 17-21) have developed a method of linear discriminant analysis when there are zero mean differences between the two populations and $\Sigma_1 = [(1-\rho_1)I + \rho_1 Epp]$ and $\Sigma_2 = \sigma^2[(1-\rho_2)I + \rho_2 Epp]$. It is shown in this paper that the linear discriminant function which they obtain does not give equal probabilities of misclassification. A method for obtaining the linear discriminant function which does give equal prob. of misclassification is developed and applied.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED
Security Classification

AN EVALUATION OF KENDALL'S ORDER-STATISTIC METHOD OF DISCRIMINANT ANALYSIS AND RELATED STUDIES

by

Stewart F. Musket

Technical Report No. 112
Department of Statistics ONR Contract

November 27, 1971

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

LIST OF ILLUSTRATIONS

Figure | Page

CHAPTER I

INTRODUCTION

The primary purpose of this study is to compare Kendall's method with linear discriminant analysis in cases when both the assumptions of equal variance-covariance matrices and multivariate normality are valid, and in cases when either or both of these assumptions are invalid. The basis of comparison will be the probabilities of misclassification.

Consider two populations, Π_1 and Π_2, and suppose that samples of size n_1 and n_2, respectively, are available from each population. Let $f_i(\underline{x})$ denote the density function of the random vector in Π_i. It is frequently assumed that $f_1(\underline{x})$ and $f_2(\underline{x})$ are multivariate normal with means $\underline{\mu}_1$ and $\underline{\mu}_2$, respectively, and a common variance-covariance matrix Σ. The linear function which minimizes the probability of misclassification is

$$(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{x} - \tfrac{1}{2}(\underline{\mu}_2 - \underline{\mu}_1)'\Sigma^{-1}(\underline{\mu}_2 + \underline{\mu}_1), \tag{1}$$

the linear discriminant function, a form of which was first introduced by Fisher in 1936 (1).

When the parameters are estimated, the sample discriminant function is obtained:

$$D_S(\underline{x}) = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)'S^{-1}\underline{x} - \tfrac{1}{2}(\underline{\bar{x}}_2 - \underline{\bar{x}}_1)'S^{-1}(\underline{\bar{x}}_2 + \underline{\bar{x}}_1) \tag{2}$$

where

$$\underline{\bar{x}}_i = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} \underline{x}_{i\alpha}, \tag{3}$$

and

$$S = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{\alpha=1}^{n_i} (\underline{x}_{i\alpha} - \underline{\bar{x}}_i)(\underline{x}_{i\alpha} - \underline{\bar{x}}_i)' \tag{4}$$

Much work has been done in determining the probabilities of misclassification in linear discriminant analysis, particularly with respect to the sample discriminant function (2,3,4,5,6,7).

Methods of nonparametric discriminant analysis are of interest because, as noted above, the assumptions of multivariate normality and equal variance-covariance matrices necessary in linear discriminant analysis frequently are unacceptable. M. G. Kendall (8, 9) has suggested a method of nonparametric discriminant analysis, sometimes referring to it as the "order-statistic" method. In this method the variates are examined one at a time. Consider, for example, the ith variate. Referring to Figure 1, this method may be explained. The variate values from Π_1 are indicated by x's and the values from Π_2 by y's. Below .4 there are four values from Π_1 and none from Π_2. Above .85 there are three observations from Π_2 and none from Π_1. There are thus seven values outside the region of overlap. The lower and upper cutoff points are .4 and .85, respectively. All of the variates are examined in the same way and the variate having the largest number of values outside the region of overlap is selected as the first discrimination variate with the cutoff points as the discrimination cutoff points. All observations with values for that variate below the

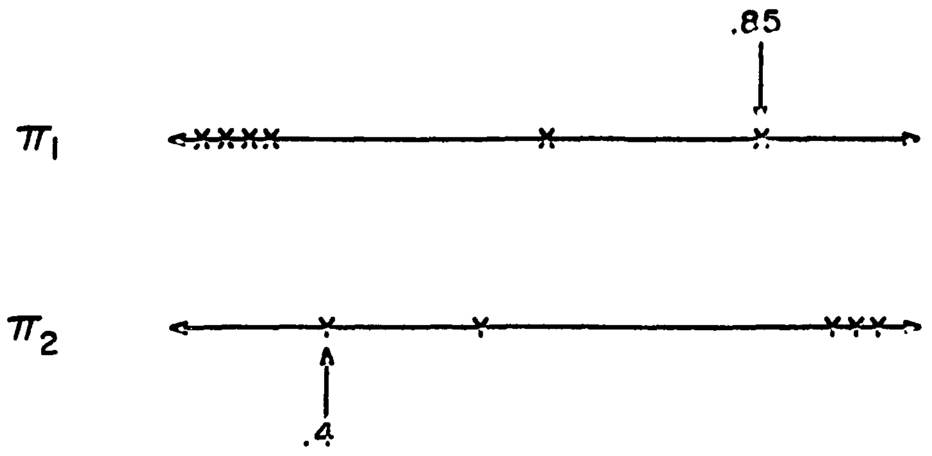

Figure 1. Illustrative Example of Kendall's Method

lower cutoff point and above the upper cutoff point are removed from further consideration; they have been classified. The procedure is continued with the remaining observations and the remaining variates. When the procedure is finished a set of classification rules will have been obtained. In this case Rule 1 would be as follows:

Rule 1	$x_i < .4$	assign to Π_1
	$x_i > .85$	assign to Π_2
	$.4 \leq x_i \leq .85$	see Rule 2

Early in the study the statistical literature was searched for examples of multivariate data which could be used to test the feasibility of Kendall's method. A total of seven examples was found and Kendall's method was used with each of these data sets.

In order to investigate the effect of unequal variance-covariance matrices, the following variance-covariance matrices were considered:

$$\Sigma_1 = (1 - \rho_1)I + \rho_1 Epp \tag{5}$$

$$\Sigma_2 = \sigma^2[(1 - \rho_2)I + \rho_2 Epp] \tag{6}$$

(Epp is a pxp matrix of 1's) $\quad 1 > \rho_i > -(p - 1)^{-1}$

These were chosen because they are not uncommon in biological and psychological work and may be good approximations in many other situations. Variance-covariance matrices of the form (5) and (6) have been considered in a number of studies concerned with discriminant analysis. In 1945 Beall (10) introduced an approximate method for calculating discriminant functions, assuming equality of covariances and variances, citing the earlier empirical evidence of Jackson (11) that this was not unreasonable. Later (1946-47), Penrose (12) developed the concept of size and shape components for the case $\Sigma_1 = \Sigma_2$ and $\rho_1 = \rho_2$. In 1963 Bartlett and Please (13) considered the general case of Σ_1 and Σ_2 given by (5) and (6) above with zero mean differences between the populations and applied the method to some measurements on twins. A Bayesian analysis of the same problem was given later by Geiser and Desu (14). Han (1968) (15) derived the discriminant function in the case of unequal mean vectors and later (1969) (16) studied the distribution of the discriminant function when $\rho_1 = \rho_2$.

Sampling experiments were performed using the variance-covariance matrices Σ_1 and Σ_2. Two p-variate normal populations, Π_1 and Π_2, were considered, with means $\underline{\mu}_1$ and $\underline{\mu}_2$ and variance-covariance matrices Σ_1

and Σ_2, respectively, in Π_1 and Π_2. In the experiments, ρ_1, ρ_2, σ^2 and $\underline{\mu}_2$ were varied; $\underline{\mu}_1$ was always set equal to the zero vector. The value of p used in all experiments was 5.

Initial samples of size 20 (sometimes 50 or 100) from populations Π_1 and Π_2 were generated and Kendall's discrimination rules were derived from the initial sample. These rules were then applied to samples of size 500 each from Π_1 and Π_2. The entire procedure was repeated 50 times, each time with new samples. The same procedure, using the same set of random numbers, was used with the linear discriminant function. The average probabilities of misclassification provided estimates of the expected value of the probability of misclassification when these discrimination procedures would be applied.

The results of the sampling experiments also provided the necessary data to compare the empirical probabilities of misclassification for linear discriminant analysis with the theoretical values obtained by Gilbert in the case $\Sigma_2 = d\Sigma_1$, that is, when one variance-covariance matrix is a multiple of the other.

Some sampling experiments have been done with multivariate non-normal distributions (all variables independent). The particular distributions considered, the Cauchy and the uniform, have been selected because of the difficulty of distinguishing between these distributions and the normal on the basis of a small sample. The lognormal distribution was considered also, serving as an example of an asymmetric distribution.

Another main purpose of this study is to develop a modified Bartlett and Please method. These authors have obtained a linear discriminant function in the case of zero-mean differences when the variance-covariance matrices are of the form already considered, i.e.,:

$$\Sigma_1 = (1 - \rho_1)I + \rho_1 Epp$$

$$\Sigma_2 = \sigma^2[(1 - \rho_2)I + \rho_2 Epp]$$

However, as A. Kshirsagar has noted, Bartlett and Please have not correctly obtained the cutoff point for the function which provides equal probabilities of misclassification. A procedure is developed in this thesis which does provide this cutoff point. The procedure is applied to the data considered by Bartlett and Please and the results compared. The modified Bartlett and Please method is compared also with Kendall's method in the case of zero mean differences.

Finally, a number of other nonparametric discrimination procedures are examined and compared with Kendall's method.

CHAPTER II

KENDALL'S NONPARAMETRIC DISCRIMINANT ANALYSIS METHOD

2.1 Description of Method

The simplest way to explain the order-statistic method is by considering an example. Kendall (8, 9) used the Iris data of Fisher. We will consider a multivariate example from geology. This example is given by Krumbein and Graybill (17) and is based on the work of Link (18). This example concerns discrimination between two carbonate subenvironments: clear, shallow water, and abundant algae water on the basis of two physicochemical variables and two measures of sedimentary texture. The data is given in Table 1. Here V_1 is the Eh below the interface, V_2 is the pH below the interface, V_3 is the phi mean diameter, and V_4 is the phi standard deviation.

Consider now Table 2, in which the data for group 1 (clear, shallow water) and group 2 (abundant algae water) have been combined and the measurements of each variable separately have been ordered from low to high. Consider V_4. In the range .97 to 1.67 there is overlap in the V_4 measurements of the two groups. However, below .97 all of the measurements are associated with group 1. Above 1.67 all of the measurements are associated with group 2. There are thus a total of 13 V_4 values outside the region of overlap - four associated with group 1 and nine associated with group 2. Examination of the data in Table 2 reveals that there are fewer observations lying outside the common

TABLE 1

MULTIVARIATE DATA FOR CARBONATE ENVIRONMENTS

	Clear, Shallow Water					Abundant Algae			
Obsn. Nbr.	V_1	V_2	V_3	V_4	Obsn. Nbr.	V_1	V_2	V_3	V_4
1	-261	7.56	0.82	1.30	1	48	7.92	1.68	1.08
2	110	4.44	2.31	0.94	2	- 76	7.97	2.17	0.97
3	83	4.30	2.51	0.56	3	-383	5.42	2.12	1.51
4	- 45	4.28	2.14	0.79	4	-225	4.89	1.37	1.78
5	-214	6.56	2.41	0.10	5	-193	4.60	1.70	1.60
6	0	7.08	0.13	1.57	6	-224	4.34	2.01	1.64
7	-158	5.53	2.38	1.01	7	-214	4.74	3.14	2.79
8	-107	5.86	1.93	1.13	8	-235	4.80	3.16	2.84
9	-264	7.22	1.90	1.20	9	-170	6.92	2.85	2.86
10	43	6.29	1.91	1.21	10	-213	6.10	3.52	2.72
11	104	5.65	0.78	1.41	11	-157	5.86	2.90	2.22
12	74	5.86	1.52	1.13	12	- 79	5.42	2.31	2.91
13	34	8.36	0.88	1.23	13	- 36	8.93	1.22	1.33
14	-200	4.86	1.93	1.55	14	-214	6.86	2.59	2.43
15	-158	5.19	1.72	1.67	15	-174	5.54	5.30	3.20

TABLE 2

DATA OF TABLE 1 ORDERED FROM LOW TO HIGH VALUES
OF EACH PARAMETER SEPARATELY

V_1		V_2		V_3		V_4	
Group 1	Group 2	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
	-383	4.28		.13		.10	
-264		4.30		.78		.56	
-261			4.34	.82		.79	
	-235	4.44		.88		.94	
	-225		4.60		1.22		.97
	-224		4.74		1.37	1.01	
-214			4.80	1.52			1.08
	-214	4.86			1.68	1.13	
	-214		4.89		1.70	1.13	
	-213	5.19		1.72		1.20	
-200			5.42	1.90		1.21	
	-193		5.42	1.91		1.23	
	-174	5.53		1.93		1.30	
	-170		5.54	1.93			1.33
-158		5.65			2.01	1.41	
-158			5.86		2.12		1.51
	-157	5.86		2.14		1.55	
-107		5.86			2.17	1.57	1.60
	- 79		6.10	2.31			1.64
	- 76	6.29			2.31	1.67	
- 45		6.56		2.38			1.78
	- 36		6.86	2.41			2.22
0			6.92	2.51			2.43
34		7.08			2.59		2.72
43		7.22			2.85		2.79
	48	7.56			2.90		2.84
74			7.92		3.14		2.86
83			7.97		3.16		2.91
104		8.36			3.52		3.20
110			8.93		5.30		

range of these variables than for V_4. This variable is then used for the first discrimination rule

1. $V_4 < .97$ assign to group 1 (4)

   $V_4 > 1.67$ assign to group 2 (9)

   $.97 \leq V_4 \leq 1.67$ see Rule 2 (17)

(The number in parenthesis is the number of observations for which the prior statement applies; e.g., there are four observations with $V_4 < .97$.)

The 13 cases discriminated by Rule 1 are then removed from further consideration. The data remaining is given in Table 3. V_3 is now the most discriminating variable, so Rule 2 becomes

2. $.97 \leq V_4 \leq 1.67$

   $V_3 < 1.22$ assign to group 1 (4)

   $V_3 > 2.17$ assign to group 1 (1)

   $1.22 \leq V_3 \leq 2.17$ see Rule 3 (12)

The remaining data is given in Table 4. For Rule 3,

3. $.97 \leq V_4 \leq 1.67$

   $1.22 \leq V_3 \leq 2.17$

   $V_2 < 4.86$ assign to group 2 (2)

   $V_2 > 7.22$ assign to group 2 (3)

   $4.86 \leq V_2 \leq 7.22$ see Rule 4 (7)

TABLE 3

TABLE 2 DATA REMAINING AFTER DISCRIMINATING WITH V_4

V_1		V_2		V_3	
Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
	-383		4.34	.13	
-264			4.60	.78	
-261		4.86		.82	
	-224	5.19		.88	
-200			5.42		1.22
	-193	5.53		1.52	
-158		5.65			1.68
-158		5.86			1.70
-107		5.86		1.72	
	- 76	6.29		1.90	
	- 36	7.08		1.91	
0		7.22		1.93	
34		7.56		1.93	
43			7.92		2.01
	48		7.97		2.12
74		8.36			2.17
104			8.93	2.38	

TABLE 4

TABLE 3 DATA REMAINING AFTER DISCRIMINATING WITH V_4 AND V_3

V_1		V_2	
Group 1	Group 2	Group 1	Group 2
	-383		4.34
-264			4.60
	-224	4.86	
-200		5.19	
	-193		5.42
-158		5.86	
-107		5.86	
	- 76	6.29	
	- 36	7.22	
43			7.92
	48		7.97
74			8.93

The remaining sample data is given below:

V_1	
Group 1	Group 2
	-383
-264	
-200	
-158	
-107	
43	
74	

So, finally, Rule 4 is

4. $.97 \leq V_4 \leq 1.67$

$1.22 \leq V_3 \leq 2.17$

$4.86 \leq V_2 \leq 7.22$

$V_1 < -264$ assign to group 2 (1)

$V_1 \geq -264$ assign to group 1 (6)

Residual group: 0

Thus it is seen that all of the 30 samples have been assigned correctly. Krumbein and Graybill in using linear discriminant analysis have found that 7 of the 30 samples were misclassified. There is, of course, concern here with the sampling variation. The set of rules derived from this particular example may perform poorly when applied to a new sample. This problem is examined in detail later.

2.2 Application to Some Additional Examples in the Statistical Literature

The statistical literature was examined for further examples of multivariate data which could be analyzed by Kendall's method. Cochran (19) had a convenient list of 12 numerical applications of linear discriminant analysis reported in the literature. Few of these papers were used, however, either because the data was not in a convenient form or else the required individual observation data was not listed. A total of seven examples, including the Fisher Iris data, were found finally; these are described in Table 5. One of the examples, that of Krumbein and Graybill, has already been considered in Section 2.1. In Table 6 there is a comparison of the results of

TABLE 5

VARIABLE AND POPULATION DESCRIPTIONS FOR DATA SETS USED WITH KENDALL'S METHOD

	AUTHOR	VARIABLES	POPULATIONS
1.	Fisher	Sepal and petal length and width of Iris	2 species of Iris
	a.		Versicolor and Virginica
	b.		Setosa and Virginica
	c.		Setosa and Versicolor
2.	Beall	4 psychological tests	Men and women
3.	Tintner	Length, amplitude, rate of change, etc., in price cycle	Consumers' and producers' goods
4.	Dempster	Renal blood pressure as a function of time	Control group and treated group of laboratory animals
5.	Krumbein & Graybill	Electrochemical measurements of water sample; grain size and sorting measurements	Water samples from two carbonate environments
6.	Mosteller & Tukey	Word frequency occurrence	Papers by Hamilton and Madison
7.	Beerstecher et al	Metabolic measurements	Alcoholic and nonalcoholic individuals

TABLE 6

COMPARISON OF RESULTS OF USTNG KENDALL'S METHOD AND LINEAR DISCRIMINANT ANALYSIS ON DATA SETS DESCRIBED IN TABLE 5

Author	Number Variates	Sample Size Π_1	Sample Size Π_2	LDA N(1,1)	LDA N(2,2)	Kendall N(1,1)	Kendall N(2,2)	Number Variates Used
1 a	4	50	50	49	48	44	43	4
b	4	50	50	50	50	50	50	1
c	4	50	50	50	50	50	50	1
2	4	32	32	28	28	22	20	4
3	4	9	10	8	9	9	10	3
4	9	12	7	12	7	12	7	2
5	4	15	15	14	9	14	14	2
6	5	11	11	11	11	11	11	2
7	62	8	4	*	*	8	4	2

* Not applicable

applying Kendall's order-statistic method and linear discriminant analysis to each sample set. Each of these examples, except the Iris data, are discussed later. The notation N(i,i) used in Table 6 is the number of observations from the ith population which were correctly assigned to the ith population. The last column of Table 6 gives the number of variates used in classification by Kendall's method.

These examples have been considered in order to examine the feasibility of applying Kendall's method in a wide variety of different situations.

Beall (10) - Four psychological tests were given to 32 men and 32 women [Table 7]. It is desired to find which test results differentiate between men and women. Kendall's method results in the following set of rules:

1.	$V_3 \geq 28$	assign to men	(18)
	$V_3 \leq 8$	assign to women	(3)
	$8 < V_3 < 28$	see Rule 2	(43)
2.	$8 < V_3 < 28$		
	$V_1 \geq 20$	assign to men	(1)
	$V_1 \leq 7$	assign to women	(2)
	$7 < V_1 < 20$	see Rule 3	(40)

TABLE 7

THE SCORES OF 32 MEN AND 32 WOMEN
ON FOUR PSYCHOLOGICAL TESTS.
DATA FROM BEALL (10).

Men				Women			
1	2	3	4	1	2	3	4
15	17	24	14	13	14	12	21
17	15	32	26	14	12	14	26
15	14	29	23	12	19	21	21
13	12	10	16	12	13	10	16
20	17	26	28	11	20	16	16
15	21	26	21	12	9	14	18
15	13	26	22	10	13	18	24
13	5	22	22	10	8	13	23
14	7	30	17	12	20	19	23
17	15	30	27	11	10	11	27
17	17	26	20	12	18	25	25
17	20	28	24	14	18	13	26
15	15	29	24	14	10	25	28
18	19	32	28	13	16	8	14
18	18	31	27	14	8	13	25
15	14	26	21	13	16	23	28
18	17	33	26	16	21	26	26
10	14	19	17	14	17	14	14
18	21	30	29	16	16	15	23
18	21	34	26	13	16	23	24
13	17	30	24	2	6	16	21
16	16	16	16	14	16	22	26
11	15	25	23	17	17	22	28
16	13	26	16	16	13	16	14
16	13	23	21	15	14	20	26
18	18	34	24	12	10	12	9
16	15	28	27	14	17	24	23
15	16	29	24	13	15	18	20
18	19	32	23	11	16	18	28
18	16	33	23	7	7	19	18
17	20	21	21	12	15	7	28
19	19	30	28	6	5	6	13

3. $8 < V_3 < 28$

$7 < V_1 < 20$

$V_4 \geq 24$	assign to women	(14)
$V_4 \leq 9$	assign to men	(1)
$9 < V_4 < 24$	see Rule 4	(25)

4. $8 < V_3 < 28$

$7 < V_1 < 20$

$9 < V_4 < 24$

$V_2 \geq 21$	assign to men	(1)
$V_2 \leq 5$	assign to men	(1)

Residual group: 23 (11 men + 12 women)

Dempster (20) - (Data from H. D. Sylwestrowicz of CIBA). [Table 8] This example concerns a type of data frequently found in pharmaceutical experimentation. Nine variables are measured on 19 animals. The nine variables are all measurements of renal blood pressure, but taken in intervals of 1/2 hour over four hours. The animals had been randomly divided into two groups of sizes 12 and 7. The first group was the control; the second received a specific drug treatment after the first of the nine measurements were taken. V_1 may then be considered a covariate. Kendall's method results in the following (non-unique) set of rules:

1.

$V_8 > 7$	assign to group 1 (control)	(11)
$V_8 < -8$	assign to group 2	(5)
$-8 \leq V_8 \leq 7$	see Rule 2	(3)

TABLE 8

MEASUREMENTS OF RENAL BLOOD PRESSURE TAKEN AT ONE-HALF HOUR INTERVALS ON TREATED AND UNTREATED ANIMALS. DATA FROM DEMPSTER (20)

Group	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
Control	17	27	17	17	25	25	25	15	17
	5	5	2	2	5	10	10	12	12
	20	20	20	20	18	17	17	17	15
	8	17	8	15	25	25	25	25	27
	22	22	20	20	15	12	18	13	12
	13	17	17	12	17	17	17	17	7
	35	23	25	23	28	27	42	42	30
	45	43	37	33	35	35	33	32	30
	2	5	2	- 5	- 7	-10	- 8	- 8	-18
	33	37	22	28	32	30	30	27	28
	25	35	22	28	28	30	28	25	22
	32	47	48	47	47	47	47	48	47
Treated	45	- 2	2	0	- 5	- 5	-10	-10	-12
	- 3	-27	-30	-33	-35	-35	-33	-33	-33
	32	17	12	12	7	2	2	7	7
	30	- 2	-10	-12	-12	-12	-12	-13	-13
	13	-20	-22	-22	-23	-27	-27	-28	-28
	20	18	2	-13	-18	-18	-22	-22	-23
	22	18	8	- 8	-10	- 8	- 7	- 2	0

2. $-8 \leq V_8 \leq 7$

$V_9 > 0$ assign to group 2 (2)

$V_9 < -18$ assign to group 1 (1)

Residual group: 0

Two points are worth making. First it should be noted that V_7 would provide as good a result as V_8. Secondly, although no use of V_1 as a covariate was made, the user of Kendall's method should realize that a covariate could be important. For example, a hypothetical case could arise in which all of the variable measurements overlapped considerably, but the difference between the subsequent measurements and the initial measurement was the key to discrimination. Kendall's method applied directly to the data in this case could result in poor results. Making the transformation of subtracting the initial measurement from the subsequent measurements (for example) could result in improved discrimination.

Tintner (21) [Table 9]- This concerns the problem of distinguishing between the prices of producers' goods and the prices of consumers' goods on the basis of certain measurements connected to their behavior during a business cycle. The data consists of the monthly wholesale prices of nine consumers' goods and ten producers' goods during the period 1860 - 1913. The seasonal and trend components had been removed by a moving average method. V_1 is the median length of the cycle in months. V_2 is the median percentage of the duration of cyclically rising prices relative to the total duration of the cycle. V_3 is the median cyclical amplitude expressed as a percentage of the trend. V_4 is the mean monthly rate of change in the cycle.

TABLE 9

CYCLICAL MEASUREMENTS OF THE PRICES OF CONSUMERS' AND PRODUCERS' GOODS. DATA FROM TINTNER (21).

	V_1	V_2	V_3	V_4
		Consumers' Goods		
Rice	72	50	8	0.5
Tea	66.5	48	15	1.0
Sugar	54	57	14	1.0
Flour	67	60	15	0.9
Coffee	44	57	14	0.3
Potatoes	41	52	18	1.9
Butter	34.5	50	4	0.5
Cheese	34.5	46	8.5	1.0
Beef	24	54	3	1.2
		Producers' Goods		
Gasoline	57	57	12.5	0.9
Lead	100	54	17	0.5
Pig Iron	100	32	16.5	0.7
Copper	96.5	65	20.5	0.9
Zinc	79	51	18	0.9
Tin	78.5	53	18	1.2
Rubber	48	50	21	1.6
Quicksilver	155	44	20.5	1.4
Copper Sheets	84	64	13	0.8
Iron Bars	105	35	17	1.8

1.	$V_1 < 48$	assign to consumers' goods	(5)
	$V_1 > 72$	assign to producers' goods	(8)
	$48 \leq V_1 \leq 72$	see Rule 2	(6)
2.	$48 \leq V_1 \leq 72$		
	$V_2 < 50$	assign to consumers' goods	(1)
	$V_2 > 57$	assign to consumers' goods	(1)
	$50 \leq V_2 \leq 57$	see Rule 3	(4)
3.	$48 \leq V_1 \leq 72$		
	$50 \leq V_2 \leq 57$		
	$V_3 < 12.5$	assign to consumers' goods	(1)
	$V_3 > 14$	assign to producers' goods	(1)
	$12.5 \leq V_3 \leq 14$	see Rule 4	
4.	$48 \leq V_1 \leq 72$		
	$50 \leq V_2 \leq 57$		
	$12.5 \leq V_3 \leq 14$		
	$V_4 \leq .9$	assign to producers' goods	(1)
	$V_4 > .9$	assign to consumers' goods	(1)

Residual group: 0

Mosteller and Tukey (22) - [Table 10]. This example concerns disputed authorship. There are a number of papers which were written by either Hamilton or Madison, and it is of some interest to be able to determine the correct author. This example is concerned with 11 papers

TABLE 10

RATES OF OCCURRENCE OF HIGH FREQUENCY WORDS IN SOME OF THE WRITINGS OF HAMILTON AND MADISON. DATA FROM MOSTELLER AND TUKEY (22).

"and" V_1	"in" V_2	"of" V_3	"the" V_4	"to" V_5
Hamilton				
16.1	35.3	63.9	93.3	38.4
32.2	24.5	78.2	110.0	31.4
24.3	23.5	64.7	90.8	42.3
18.0	27.2	59.6	86.8	35.9
20.6	26.9	61.4	83.6	39.5
21.8	17.4	73.1	90.4	35.6
27.9	23.1	61.9	85.4	41.3
28.5	26.1	71.3	74.5	33.3
28.9	20.9	56.9	82.7	44.9
21.3	25.0	60.4	82.2	47.7
18.5	30.7	72.7	109.3	36.6
Madison				
31.6	19.9	54.8	93.8	38.6
37.3	23.3	56.8	84.2	31.0
21.2	17.5	58.2	97.6	39.9
27.9	19.1	55.8	93.1	33.5
40.7	9.3	59.0	71.5	33.6
24.4	27.9	60.0	115.3	34.8
27.7	17.7	61.1	115.3	32.7
28.1	22.3	57.0	110.9	29.7
30.6	23.6	68.3	118.6	23.2
33.9	21.8	64.9	93.7	33.6
23.3	31.4	34.8	94.3	49.6

*All rates of occurrence of high frequency words are per thousand words of text.

mostly selected from the Federalist papers, of known authorship. The variables used for discriminating between the authors of the paper are certain high frequency words. This particular example was selected by Mosteller and Tukey to illustrate the application of the jackknife method in discriminant analysis.

1.	$V_3 > 68.3$	assign to Hamilton	(4)
	$V_3 \leq 56.8$	assign to Madison	(4)
	$56.8 < V_3 \leq 68.3$	see Rule 2	(14)
2.	$56.8 < V_3 \leq 68.3$		
	$V_5 > 39.9$	assign to Hamilton	(4)
	$V_5 \leq 35.9$	assign to Madison	(6)
	$35.9 < V_5 \leq 39.9$	see Rule 3	(4)
3.	$56.8 < V_3 \leq 68.3$		
	$35.9 < V_5 \leq 39.9$		
	$V_2 \geq 26.9$	assign to Hamilton	(3)
	$V_2 \leq 17.9$	assign to Madison	(1)

Residual group: 0

Beerstecher et al (23) (Because of the large amount of data it is not included here.) In this study 62 variables related to metabolic patterns were measured in 12 individuals over a period of one month. This was a preliminary study of the various traits of an alcoholic and nonalcoholic individual, the former during nondrinking periods, in order to discover differences worthy of further study later on. Using Kendall's method all 12 individuals were separated into the correct classes using two variates. The most discriminating variate was hippuric acid concentration in urine. Of the four alcoholics, three had concentrations above 17 units and of the eight nonalcoholics, seven had concentrations below 17 units. The remaining two subjects were correctly classified by a phagocytic index, although undoubtedly many other variates would have served as well in classifying these two remaining subjects.

In Beerstecher's study univariate t-tests were used to isolate important differentiating variates between the alcoholics and non-alcoholics. The hippuric acid urine concentration and the saliva sodium concentration had the largest t-values.

It is of course not surprising that Kendall's method provided compatible results since large t-values for a variate would indicate wide separation between the means of the samples from the two populations, causing it to be selected as a discriminating variate by Kendall's method.

This particular set of data was examined using Kendall's method in order to illustrate a potentially valuable use of this method as a screening technique for significant variates. Further consideration of this use is given in Section 6.1.

CHAPTER III

COMPARISON OF KENDALL'S METHOD WITH LINEAR DISCRIMINANT ANALYSIS

3.1 Method of Evaluation

The two assumptions of concern in linear discriminant analysis are equality of variance-covariance matrices and multivariate normality. Kendall's method is compared with linear discriminant analysis in cases where all of these assumptions are valid and in cases where either or both assumptions are invalid.

Sampling experiments were done using the variance-covariance matrices

$$\Sigma_1 = (1-\rho_1)I+\rho_1 Epp$$
$$\Sigma_2 = \sigma^2[(1-\rho_2)I+\rho_2 Epp]$$

Two p-variate normal populations, Π_1 and Π_2, were considered, with means $\underline{\mu}_1$ and $\underline{\mu}_2$ and variance-covariance matrices Σ_1 and Σ_2 respectively. In the experiments ρ_1, ρ_2, σ^2 and $\underline{\mu}_2$ were varied; $\underline{\mu}_1$ was always equal to zero. The value of p used in all experiments was 5.

The particular set of parameters used are given in Table 11. A CDC 6400 computer which has a 60-bit word length was used in this study.

TABLE 11

PARAMETER VALUES USED IN SAMPLING EXPERIMENTS

ρ_1	ρ_2	σ^2	$\underline{\mu}_2$
.1	.1	1,2	$\underline{0}$*,$\underline{1}$,$\underline{2}$
.5	.5	1,2	$\underline{0}$,$\underline{1}$,$\underline{2}$
.1	.9	1,2	$\underline{0}$,$\underline{1}$,$\underline{2}$
.9	.1	1,2	$\underline{0}$,$\underline{1}$,$\underline{2}$
-.1	.9	2	$\underline{1}$,$\underline{2}$

*$\underline{0}' = (0,0,0,0,0)$

The method used to generate a sample vector with variance-covariance matrix Σ was to generate the vector

$$\underline{x}' = (x_1,x_2,x_3,x_4,x_5)$$

where x_i are independent uniform (0,1) variables. Using the inverse of the normal probability integral Φ^{-1}, the vector

$$\underline{y}' = (y_1,y_2,y_3,y_4,y_5)$$

was obtained, where $\Phi^{-1}(x_i) = y_i$. Thus

$$\underline{y} \sim MVN(\underline{0},I)$$

The variance-covariance matrix Σ_i was factored into the product of a lower triangular matrix and its transpose by a Crout factorization (24). Thus Σ is expressed in the form

$$\Sigma = TT'$$

Multiplying $\underline{y}$ by T produced the desired vector

$$T\underline{y} \sim MVN(\underline{0}, TT')$$

or

$$T\underline{y} \sim MVN(\underline{0}, \Sigma)$$

The algorithm for the Crout factorization is as follows:

$$t_{11} = \sigma_{11}^{\frac{1}{2}} \qquad \left(T = \left\{t_{ij}\right\}; \quad \Sigma = \left\{\sigma_{ij}\right\}\right) \tag{7}$$

$$t_{i1} = t_{11}^{-1}\sigma_{i1} \qquad i=2,3,\ldots p \tag{8}$$

If the preceding columns $k < j$ have been completed, the jth diagonal term is calculated by

$$t_{jj} = (\sigma_{ij} - \sum_{k=1}^{j-1} t_{jk}^{2})^{\frac{1}{2}} \tag{9}$$

If $j < p$, the elements below the diagonal are computed by the formula

$$t_{ij} = t_{jj}^{-1}(\sigma_{ij} - \sum_{k=1}^{j-1} t_{ik}t_{jk}) \qquad k = j+1,\ldots,p \tag{10}$$

As explained in the introduction, initial samples of size 20 (sometimes 50 or 100) from populations Π_1 and Π_2 were generated and Kendall's discrimination rules were derived from the initial sample. These rules were then applied to new samples of size 500 each from Π_1 and Π_2. This entire procedure was repeated 50 times, each time with new samples. The same procedure, using the same set of random numbers, was used with the linear discriminant function.

Consider one of the 50 repetitions of the experiment. Let F(i,j) denote the fraction of the initial sample from population j which was classified as population i, and $F(0,j)$ denote the fraction of the initial sample from j which was not classified. With Kendall's method $F(1,1)$, $F(2,2)$, $F(0,1)$, and $F(0,2)$ were calculated. Since no probabilities of misclassification were allowed, $F(1,2) = F(2,1) = 0$. With linear discriminant analysis, $F(1,1)$, $F(2,2)$, $F(1,2)$, and $F(2,1)$ were calculated using the resubstitution method. In linear discriminant analysis all of the samples are classified, so $F(0,1) = F(0,2) = 0$.

The set of rules derived from the initial sample was applied to a new sample of size 500 from each population. Let $FI(i,j)$ $(i=0,1,2; j=1,2)$ denote the fraction of this index sample classified as indicated. $FI(i,j)$ $(i=0,1,2;j=1,2)$ was calculated for Kendall's classification rules and $FI(i,j)$ $(i=1,2;j=1,2)$ was calculated for the linear discriminant function.

This was done for all 50 repetitions and the F's and FI's averaged to give estimates of the expected value of the probability of misclassification (and classification) based on the initial sample and the index sample, respectively. The average value of $F(i,j)$ $(i=0,1,2; j=1,2)$ is denoted by $\hat{P}(i,j)$; the average value of $FI(i,j)$ $(i=0,1,2; j=1,2)$ is denoted by $P^*(i,j)$.

$\hat{P}(i,j)$ $(i \neq j)$ is known to underestimate the expected value of the probability of misclassification of the sample discriminant function. It was calculated for comparison with the Kendall estimates from the initial sample.

The $P^*(i,j)$ values are of primary interest, since these measure the probabilities of misclassification when the classification rule or function is used. The $\hat{P}(i,j)$ values are of interest since in practice only the same sample used to derive the rule or function is available to judge the probability of misclassification when the rule or function is applied. It is important to determine the relationship between the $\hat{P}(i,j)$ and $P^*(i,j)$ values so that some judgment of the actual performance of the rule or function may be ascertained from the performance with the initial sample.

In some of the sampling experiments a principal components transformation was done. The principal components were estimated from the combined samples and the transformation then applied separately to the two samples. Kendall's method was then applied to the transformed data.

3.2 Testing the Random Number Generator

Since Kendall's method is based on the order-statistics of the distribution, it is this aspect of the random number generator which should be examined most carefully. In order to judge the quality of the random number generator, some tests were performed. In the first test 20 N(0,1) independent random numbers were generated and then ordered from low to high values. This was repeated 100,000 times, and the average value of each order-statistic calculated. These values were compared with the theoretical value of the order-statistics as tabled in Owen (25). The results are given in Table 12; the sampling experiment results agree quite well with the tabled values. In the second test use was made of the results of Gupta (26) who has calculated the percentage points of order-statistics from the normal distribution. Again, a sample of size 20 was generated from an N(0,1) distribution and this time the percentage points given by Gupta were used to calculate the number of samples which exceeded each percentage

TABLE 12

COMPARISON OF TABLED EXPECTED VALUES OF ORDER STATISTICS FROM N(0,1) DISTRIBUTION WITH THOSE OBTAINED FROM SAMPLING EXPERIMENT

Order Statistic	Tabled Value	Sampling Experiment Results	Order Statistic	Tabled Value	Sampling Experiment Results
1	-1.8675	-1.8678	11	0.0620	0.0613
2	-1.4076	-1.4091	12	0.1870	0.1861
3	-1.1309	-1.1314	13	0.3149	0.3148
4	-0.9210	-0.9187	14	0.4483	0.4483
5	-0.7454	-0.7448	15	0.5903	0.5910
6	-0.5903	-0.5891	16	0.7454	0.7455
7	-0.4483	-0.4478	17	0.9210	0.9206
8	-0.3149	-0.3146	18	1.1309	1.1289
9	-0.1870	-0.1872	19	1.4076	1.4074
10	-0.0620	-0.0617	20	1.8675	1.8701

point. This was done 10,000 times. The results are given in Table 13. Again, the two results agree very well.

3.3 Comparison in the case of multivariate normality, including consideration of unequal variance-covariance matrices

3.3.1 Summary and Analysis of Main Sampling Experiment Results

The results of the sampling experiments are given in Tables 14-18; K denotes Kendall's method; LDA, linear discriminant analysis; K(PC), Kendall's method with principal components transformation.

TABLE 13

COMPARISON OF THEORETICAL AND SAMPLING EXPERIMENT
α-LEVELS FOR 1ST, 10TH, 11TH, AND 20TH
ORDER STATISTICS OF SAMPLE OF
SIZE 20 FROM N(0,1)

Order Statistic	α-Level* (In Percent)	Experimental Results	Order Statistic	α-Level* (In Percent)	Experimental Results
1	50	49.7	11	50	49.7
	25	24.8		25	25.2
	10	10.1		10	10.0
	5	5.2		5	5.3
	1	1.1		1	1.0
10	50	50.3	20	50	50.0
	25	25.1		25	25.0
	10	10.1		10	10.1
	5	5.3		5	5.1
	1	1.1		1	1.0

* $\Pr\left[Z_{[K]} > Z^{*}_{[K]}(\alpha)\right] = \alpha$, where $Z_{[K]}$ is the Kth order-statistic (K=1,10,11,20) and $Z^{*}_{[K]}(\alpha)$ is the upper percentage point associated with α.

TABLE 14

SUMMARY OF SAMPLING EXPERIMENTS, $\rho_1 = \rho_2 = .1$

σ^2	$\underline{\mu}$	Type	$\hat{P}(1,1)$	$\hat{P}(2,2)$	P*(1,1)	P*(2,1)	P*(0,1)	P*(2,2)	P*(1,2)	P*(0,2)
1	1	K	.860	.885	.711	.213	.076	.721	.215	.065
		LDA	.865	.852	.805	.195	----	.790	.210	----
1	2	K	.998	1.000	.866	.114	.020	.860	.122	.018
		LDA	.979	.983	.958	.042	----	.958	.042	----
2	1	K	.739	.888	.558	.265	.177	.738	.205	.057
		LDA	.852	.779	.809	.191	----	.700	.300	----
2	2	K	.986	.995	.824	.142	.026	.839	.144	.017
		LDA	.975	.932	.953	.047	----	.884	.116	----

TABLE 15

SUMMARY OF SAMPLING EXPERIMENTS, $\rho_1 = \rho_2 = .5$

σ^2	$\underline{\mu}$	Type	$\hat{P}(1,1)$	$\hat{P}(2,2)$	P*(1,1)	P*(2,1)	P*(0,1)	P*(2,2)	P*(1,2)	P*(0,2)
1	1	K	.720	.723	.575	.214	.211	.617	.183	.201
		LDA	.781	.785	.691	.309	----	.709	.291	----
1	2	K	.979	.962	.849	.134	.017	.863	.116	.021
		LDA	.927	.921	.876	.124	----	.881	.119	----
2	1	K	.638	.796	.467	.264	.269	.672	.216	.111
		LDA	.768	.704	.706	.294	----	.620	.380	----
2	2	K	.929	.943	.752	.192	.056	.830	.150	.020
		LDA	.923	.844	.888	.112	----	.783	.217	----

TABLE 16

SUMMARY OF SAMPLING EXPERIMENTS, $\rho_1 = .9$, $\rho_2 = .1$

σ^2	μ	Type	$\hat{P}(1,1)$	$\hat{P}(2,2)$	P*(1,1)	P*(2,1)	P*(0,1)	P*(2,2)	P*(1,2)	P*(0,2)
1	0	K	.440	.690	.246	.283	.471	.458	.313	.229
		K(PC)	.886	.975	.546	.366	.088	.754	.231	.015
1	1	K	.710	.856	.640	.144	.216	.584	.317	.099
		K(PC)	.915	.985	.666	.269	.065	.692	.294	.015
		LDA	.750	.810	.737	.263	----	.671	.329	----
1	2	K	.949	.983	.896	.076	.028	.800	.179	.076
		K(PC)	.963	.990	.889	.087	.024	.758	.213	.029
		LDA	.884	.971	.861	.139	----	.910	.090	----
2	0	K	.487	.824	.205	.339	.456	.697	.186	.117
		K(PC)	.870	.990	.493	.393	.114	.868	.120	.012
2	1	K	.699	.896	.535	.199	.266	.668	.271	.061
		K(PC)	.915	.996	.647	.281	.072	.789	.195	.016
		LDA	.767	.790	.731	.269	----	.654	.346	----
2	2	K	.931	.988	.839	.116	.046	.793	.194	.014
		K(PC)	.919	.999	.768	.165	.067	.767	.214	.018
		LDA	.881	.924	.862	.138	----	.839	.161	----

TABLE 17

SUMMARY OF SAMPLING EXPERIMENTS, $\rho_1 = .1$, $\rho_2 = .9$

σ^2	μ	Type	$\hat{P}(1,1)$	$\hat{P}(2,2)$	P*(1,1)	P*(2,1)	P*(0,1)	P*(2,2)	P*(1,2)	P*(0,2)
1	0	K	.710	.400	.562	.204	.234	.175	.352	.474
		K(PC)	.965	.828	.858	.119	.023	.337	.539	.124
	1	K	.858	.739	.576	.323	.101	.658	.142	.200
		K(PC)	.990	.897	.725	.257	.018	.635	.300	.066
		LDA	.815	.748	.654	.346	----	.749	.251	----
	2	K	.985	.975	.809	.169	.022	.896	.085	.019
		K(PC)	1.000	.953	.784	.201	.015	.676	.101	.024
		LDA	.976	.883	.929	.071	----	.853	.147	----
2	0	K	.620	.534	.331	.376	.294	.415	.242	.343
		K(PC)	.926	.818	.748	.223	.039	.418	.479	.104
	1	K	.677	.658	.417	.366	.217	.608	.150	.242
		K(PC)	.954	.885	.617	.363	.021	.645	.298	.060
		LDA	.776	.690	.632	.368	----	.668	.332	----
	2	K	.913	.927	.688	.248	.064	.838	.117	.045
		K(PC)	.995	.955	.649	.338	.013	.833	.152	.016
		LDA	.946	.811	.886	.114	----	.768	.232	----

TABLE 18

SUMMARY OF SAMPLING EXPERIMENTS, $\rho_1 = -.1$, $\rho_2 = .9$

σ^2	μ	Type	$\hat{P}(1,1)$	$\hat{P}(2,2)$	P*(1,1)	P*(2,1)	P*(0,1)	P*(2,2)	P*(1,2)	P*(0,2)
2	1	K	.752	.654	.497	.317	.186	.607	.178	.215
		K(PC)	.977	.954	.566	.416	.018	.779	.197	.024
		LDA	.814	.715	.663	.337	----	.677	.323	----
	2	K	.944	.929	.701	.255	.045	.866	.103	.031
		K(PC)	.993	.981	.623	.367	.010	.881	.111	.008
		LDA	.976	.818	.908	.092	----	.780	.220	----

The values for $\sigma^2 = 1$, $\rho_1 = .1$, $\rho_2 = .9$, $\underline{\mu} = \underline{0}, \underline{1}, \underline{2}$ have been calculated already in Table 16 in the entries $\sigma^2 = 1$, $\rho_1 = .9$, $\rho_2 = .1$, $\underline{\mu} = \underline{0}, \underline{1}, \underline{2}$. The experiment has been repeated, however, to serve as a basis for judging the quality of the estimates. It can be seen from Table 19 that the estimates are stable for $\underline{\mu} = \underline{1}$, and $\underline{\mu} = \underline{2}$, but there is variation in the case of zero mean differences.

Tables 20 and 21 summarize the probability of misclassification information for the Kendall method and LDA, respectively. $\hat{P}(i,j)$ is the probability, estimated from the initial sample, of assigning an observation from the jth population to the ith population. $\hat{P}_T$ is the average of $\hat{P}(2,1)$ and $\hat{P}(1,2)$; $\hat{P}_T(0)$ is the average of $\hat{P}(0,1)$ and $\hat{P}(0,2)$. $P_T^*(i,j)$ is the probability of assigning an observation from the jth population to the ith population when the classification rules derived from the initial sample are used. P_T^* is the average of $P_T^*(2,1)$ and $P_T^*(1,2)$. The entries are in order of increasing T^2 values for $\sigma^2 = 2$, and then for increasing T^2 values for $\sigma^2 = 1$, where T^2, the Mahalanobis distance, is

$$(\underline{\mu}_1 - \underline{\mu}_2)' \, \Sigma^{-1} \, (\underline{\mu}_1 - \underline{\mu}_2) \tag{11}$$

in the case of equal variance-covariance matrices. For Σ_1 and Σ_2 of the form considered in this study, the equation used for calculating T^2 will be derived:

$$\Sigma_1 = (1 - \rho_1)I + \rho_1 Epp$$

$$\Sigma_2 = \sigma^2[(1 - \rho_2)I + \rho_2 EPP]$$

$$\Sigma_1 + \Sigma_2 = [(1 + \sigma^2) - (\rho_1 + \sigma^2\rho_2)]I + (\rho_1 + \sigma^2\rho_2)Epp \tag{12}$$

TABLE 19

REPEATABILITY OF SAMPLING EXPERIMENT RESULTS

σ^2	μ	$\hat{P}(i,i)$	$\hat{P}(j,j)$	$P^*(i,i)$	$P^*(j,i)$	$P^*(0,i)$	$P^*(j,j)$	$P^*(i,j)$	$P^*(0,j)$
1	0	.440	.690	.246	.283	.471	.458	.313	.229
		.400	.710	.175	.352	.474	.562	.204	.234
	1	.710	.856	.640	.144	.216	.584	.317	.099
		.739	.858	.658	.142	.200	.576	.323	.101
	2	.949	.983	.896	.076	.028	.800	.179	.076
		.975	.985	.896	.085	.019	.809	.169	.022

For $\rho_1 = .9$, $\rho_2 = .1$: $i = 1$, $j = 2$

For $\rho_1 = .1$, $\rho_2 = .9$: $i = 2$, $j = 1$

TABLE 20

SUMMARY OF THE PROBABILITIES OF MISCLASSIFICATION USING KENDALL'S METHOD IN THE SAMPLING EXPERIMENTS

Case	ρ_1	ρ_2	σ^2	$\underline{\mu}$	T^2	$\hat{P}(0,1)$	$\hat{P}(0,2)$	$\hat{P}_T(0)$	$P^*(2,1)$	$P^*(1,2)$	P^*_T
1	.1	.9	2	$\underline{1}$	0.94	.323	.342	.333	.366	.150	.258
2	-.1	.9	2	$\underline{1}$	1.02	.248	.346	.297	.317	.178	.248
3	.5	.5	2	$\underline{1}$	1.12	.362	.204	.283	.264	.216	.240
4	.9	.1	2	$\underline{1}$	1.36	.301	.104	.203	.199	.271	.235
5	.1	.1	2	$\underline{1}$	2.38	.261	.112	.187	.265	.205	.235
6	.1	.9	2	$\underline{2}$	3.78	.087	.073	.080	.248	.117	.183
7	-.1	.9	2	$\underline{2}$	4.08	.056	.071	.064	.255	.103	.179
8	.5	.5	2	$\underline{2}$	4.44	.071	.057	.064	.192	.150	.171
9	.9	.1	2	$\underline{2}$	5.40	.069	.012	.041	.116	.194	.155
10	.1	.1	2	$\underline{2}$	9.52	.014	.005	.010	.142	.144	.143
11	.5	.5	1	$\underline{1}$	1.66	.280	.277	.279	.214	.183	.199
12	.9	.1	1	$\underline{1}$	1.66	.290	.144	[illegible]	.144	.317	.231
13	.1	.1	1	$\underline{1}$	3.58	.140	.115	[illegible]	.213	.215	.214
14	.5	.5	1	$\underline{2}$	6.66	.021	.038	.030	.134	.116	.125
15	.9	.1	1	$\underline{2}$	6.66	.051	.017	.034	.076	.179	.128
16	.1	.1	1	$\underline{2}$	14.28	.002	.000	.001	.114	.122	.118

TABLE 21

SUMMARY OF THE PROBABILITIES OF MISCLASSIFICATION USING LINEAR DISCRIMINANT ANALYSIS IN THE SAMPLING EXPERIMENTS

Case	ρ_1	ρ_2	σ^2	$\underline{\mu}$	T^2	$\hat{P}(2,1)$	$\hat{P}(1,2)$	$\hat{P}_T$	$P^*(2,1)$	$P^*(1,2)$	P^*_T
1	.1	.9	2	$\underline{1}$	0.94	.224	.310	.267	.368	.332	.350
2	-.1	.9	2	$\underline{1}$	1.02	.186	.285	.236	.337	.323	.330
3	.5	.5	2	$\underline{1}$	1.12	.232	.296	.264	.294	.380	.337
4	.9	.1	2	$\underline{1}$	1.36	.233	.210	.222	.269	.346	.308
5	.1	.1	2	$\underline{1}$	2.38	.148	.221	.185	.191	.300	.246
6	.1	.9	2	$\underline{2}$	3.78	.054	.189	.122	.114	.232	.173
7	-.1	.9	2	$\underline{2}$	4.08	.024	.182	.103	.092	.220	.156
8	.5	.5	2	$\underline{2}$	4.44	.077	.156	.117	.112	.217	.165
9	.9	.1	2	$\underline{2}$	5.40	.119	.076	.098	.138	.161	.150
10	.1	.1	2	$\underline{2}$	9.52	.025	.068	.047	.047	.116	.082
11	.5	.5	1	$\underline{1}$	1.66	.219	.215	.217	.309	.291	.300
12	.9	.1	1	$\underline{1}$	1.66	.250	.190	.220	.263	.329	.296
13	.1	.1	1	$\underline{1}$	3.58	.135	.148	.142	.193	.210	.203
14	.5	.5	1	$\underline{2}$	6.66	.073	.079	.076	.124	.119	.122
15	.9	.1	1	$\underline{2}$	6.66	.116	.029	.073	.139	.090	.115
16	.1	.1	1	$\underline{2}$	14.28	.021	.017	.019	.042	.042	.042

Let $\quad 1 + \sigma^2 = a \qquad \rho_1 + \sigma^2\rho_2 = b$ (13)

Then

$$[(a - b)I + bEpp]^{-1}$$

is determined using the relationship

$$[I + PQ]^{-1} = I - P(I + QP)^{-1}Q \tag{14}$$

Thus

$$[(a - b)I + bEpp]^{-1} = (a - b)^{-1}[I + \frac{b}{a-b}Epp]^{-1} \tag{15}$$

$$= \frac{1}{a-b}\left[I - \frac{b}{a-b}\begin{pmatrix}1\\1\\ \cdot\\ \cdot\\ \cdot\\1\end{pmatrix}\left(1 + \frac{pb}{a-b}\right)^{-1}(1,1,\ldots 1)\right] \tag{16}$$

$$= \frac{1}{a-b}\left[I - \frac{b}{a-b}\cdot\left(\frac{a-b+pb}{a-b}\right)Epp\right] \tag{17}$$

$$= \frac{1}{a-b}\left[I - \frac{b}{(a-b) + pb}Epp\right] \tag{18}$$

Let $\quad \dfrac{b}{(a-b) + pb} = c$ (19)

Then

$$\frac{T^2}{2} = \frac{1}{a-b}\,\underline{\mu}'[I - cEpp]\underline{\mu} \tag{20}$$

$$= \frac{\Sigma\mu_i^2 - c(\Sigma\mu_i)^2}{a-b} \tag{21}$$

Using equation (19),

$$\frac{T^2}{2} = \frac{1}{a-b}\left(\Sigma\mu_i^2 - \frac{b}{(a-b) + pb}(\Sigma\mu_i)^2\right) \tag{22}$$

Finally, using equation (13)

$$T^2 = \frac{2}{(1+\sigma^2) - (\rho_1+\sigma^2\rho_2)}\left\{\Sigma\mu^{\ 2} - \frac{(\rho_1+\sigma^2\rho_2)(\Sigma\mu_i)^2}{(1+\sigma^2) + (p-1)(\rho_1+\sigma^2\rho_2)}\right\} \tag{23}$$

With Kendall's method not all of the index sample will be classified. Thus P*(0,1) and P*(0,2) denotes the probability that an observation from Π_1 and Π_2, respectively, will not be classified. $P_T^*(0)$ denotes the average of P*(0,1) and P*(0,2). These values are summarized in Table 22. They must be taken into consideration in judging the performance of Kendall's method.

The relative performance of the Kendall order-statistic method and LDA (Linear Discriminant Analysis) may be compared by examining Table 23. One column of Table 23 is identified by $P_T^*(K)/P_T^*(LDA)$. This is the ratio of the average probability of misclassification of the Kendall method to that of LDA. The comparison is somewhat unfair, since not all of the index sample has been classified in using the Kendall method. A factor, f, has been introduced to allow comparison of the probabilities of misclassification only with respect to the portion of the sample actually classified. The factor f is the reciprocal of the

TABLE 22

SUMMARY OF PROBABILITIES THAT APPLICATION OF PREVIOUSLY DERIVED KENDALL RULES WILL NOT CLASSIFY A SAMPLE

Case	ρ_1	ρ_2	σ^2	$\underline{\mu}$	$P^*(0,1)$	$P^*(0,2)$	$P_T^*(0)$
1	.1	.9	2	$\underline{1}$	.217	.242	.230
2	-.1	.9	2	$\underline{1}$	.186	.215	.201
3	.5	.5	2	$\underline{1}$	.269	.111	.190
4	.9	.1	2	$\underline{1}$	.266	.061	.164
5	.1	.1	2	$\underline{1}$	.177	.057	.117
6	.1	.9	2	$\underline{2}$	.064	.045	.055
7	-.1	.9	2	$\underline{2}$	.045	.031	.038
8	.5	.5	2	$\underline{2}$	.056	.020	.038
9	.9	.1	2	$\underline{2}$	.046	.014	.030
10	.1	.1	2	$\underline{2}$	026	.017	.020
11	.5	.5	1	$\underline{1}$	.211	.201	.206
12	.9	.1	1	$\underline{1}$	.216	.099	.158
13	.1	.1	1	$\underline{1}$	.076	.065	.071
14	.5	.5	1	$\underline{2}$	.017	.021	.019
15	.9	.1	1	$\underline{2}$	.028	.076	.052
16	.1	.1	1	$\underline{2}$	.020	.018	.019

TABLE 23

COMPARISON OF THE PERFORMANCE OF KENDALL'S METHOD AND LINEAR DISCRIMINANT ANALYSIS WITH RESPECT TO THE PROBABILITIES OF MISCLASSIFICATION

Case	ρ_1	ρ_2	σ^2	$\underline{\mu}$	$\frac{P_T^*(K)}{P_T^*(LDA)}$	$f \cdot \frac{P_T^*(K)}{P_T^*(LDA)}$
1	.1	.9	2	$\underline{1}$	.737	.957
2	-.1	.9	2	$\underline{1}$	.752	.941
3	.5	.5	2	$\underline{1}$	.712	.879
4	.9	.1	2	$\underline{1}$	.763	.913
5	.1	.1	2	$\underline{1}$	.956	1.08
6	.1	.9	2	$\underline{2}$	1.06	1.12
7	-.1	.9	2	$\underline{2}$	1.15	1.19
8	.5	.5	2	$\underline{2}$	1.04	1.08
9	.9	.1	2	$\underline{2}$	1.03	1.07
10	.1	.1	2	$\underline{2}$	1.74	1.78
11	.5	.5	1	$\underline{1}$	0.663	.835
12	.9	.1	1	$\underline{1}$	0.780	.828
13	.1	.1	1	$\underline{1}$	1.05	1.14
14	.5	.5	1	$\underline{2}$	1.02	1.04
15	.9	.1	1	$\underline{2}$	1.11	1.17
16	.1	.1	1	$\underline{2}$	2.81	2.86

fraction of the index sample in Kendall's method which was classified. One of the columns, identified by $f \cdot [P_T^*(K)/P_T^*(LDA)]$, lists this ratio.

Considering the ratio $P_T^*(K)/P_T^*(LDA)$, in cases 1-6, 11 and 12, Kendall's method has a smaller P_T^* value than LDA. In cases 7-9, and 13-15, the two methods perform about the same. The Kendall method is definitely worse in cases 10 and 16. These are the two cases with the largest T^2 values, and with equal values of ρ_1 and ρ_2. Considering the ratio $f \cdot [P_T^*(K)/P_T^*(LDA)]$, similar results are obtained for all cases except for case 6 for which the Kendall method is better and cases 10 and 16 for which LDA is better.

3.2.2 Effect of Unequal Mean Components

In one sampling experiment, the Mahalanobis distance was kept constant and the components of the mean vector were varied. More specifically, in one sampling experiment already considered, $\rho_1 = .1$, $\rho_2 = .9$, $\sigma^2 = 2$, $\underline{\mu}' = (2,2,2,2,2)$, the Mahalanobis distance was 3.78. Components of the mean vector were chosen to be (0,0,1,1,x), where x was such that the distance was unchanged. The value of x was found to be 1.512. Examining Table 24 shows, as expected, that P_T^* does not change for linear discriminant analysis. However, P_T^* decreases substantially for Kendall's method, and this is only slightly indicated by the decrease in $\hat{P}(1,1)$ and $\hat{P}(2,2)$. The point of this single example is that the error rates can be strongly influenced by changes in the mean vector, even when the Mahalanobis distance is unchanged. To some extent this would be anticipated, since, as already noted, Kendall's method depends on the overlap of distributions more than on the distances between the means of the distributions.

TABLE 24

EFFECT OF UNEQUAL MEAN COMPONENTS

$\underline{\mu}' = (2, 2, 2, 2, 2)$;

$\underline{\mu}^{*\prime} = (0, 0, 1, 1, 1.512)$.

$\rho_1 = .1$, $\rho_2 = .9$, $\sigma^2 = 2$

Type	$\hat{P}(1,1)$	$\hat{P}(2,2)$	$P^*(1,1)$	$P^*(2,1)$	$P^*(0,1)$	$P^*(2,2)$	$P^*(1,2)$	$P^*(0,2)$	P^*_T
$K(\underline{\mu})$	.913	.927	.688	.248	.064	.838	.117	.045	.183
$K(\underline{\mu}^*)$	.897	.851	.560	.381	.059	.640	.261	.099	.321
$LDA(\underline{\mu})$	.946	.811	.886	.114	----	.768	.232	----	.173
$LDA(\underline{\mu}^*)$	.816	.917	.755	.245	----	.881	.119	----	.182

3.3.3 Effect of Allowed Probabilities of Misclassification

One problem with Kendall's method is that, especially with distributions of infinite range, one or more extreme valued observations from one distribution often will be mixed with the second distribution values and little separation of samples from the two distributions will be possible. Accepting these few outlying observations as allowed misclassifications, eliminating them from further consideration, and proceding with Kendall's method can produce a set of discrimination rules which result in an increase in the probabilities of classification in the index sample with only a limited increase in the probabilities of misclassification. That is, many of the observations unclassified before will now be classified, and most of these correctly.

There are many ways to introduce an allowed probability of misclassification. For example, a cumulative allowed probability of misclassification for each group could be specified. However, there is a problem in allocation. Suppose that for the samples of 20 from each of the two populations an allowed probability of misclassification of 0.1 is specified. Suppose also that low values of the first selected variate favor one population and high values those of another population. Then the cumulative allowed probability of misclassification could be used immediately, allowing two misclassifications for low values of the variate and two misclassifications for high values of the variate. However, better overall results may be obtained by using fewer than these four allowed misclassifications for the first selected variate, using some of these for increasing the classification quality of the next or subsequent variate. The computer logic for allocating the four allowed misclassifications so as to maximize the overall performance of the method could be worked out relatively easily and some consideration may be given to this at a later time. There would, of course, be a considerable increase in the computing time. For the particular set of experiments reported here, a very simple rule was used, one which is far from optimal. An allowed probability of 0.05 was allowed for misclassifications for low values of the first selected variate and 0.05 for the high values of this variate. If the selected variate was such that samples from one population had low values of this variate and samples from the other population had high values of this variate, then the allowed probability of misclassification for each group would be 0.05. However, if both high and low values of a variate were characteristic of a single population then the allowed probability of misclassification would in effect be 0.1, whereas there would be no allowed misclassification for the other group.

The effect of the allowed probability of misclassification was considered in several sampling experiments and the results are summarized in Table 25. Kendall's method is denoted by K, and Kendall's method with an allowed probability of misclassification is denoted by K(PA). The effect of increased sample size on the index sample probabilities obtained by Kendall's method can be seen for samples of size 20, 50, and 100 for the case $\sigma^2 = 1$, $\underline{\mu} = 1$, and $\rho_1 = \rho_2 = .5$, and for 20 and 100 in the case $\sigma^2 = 2$, $\underline{\mu} = \underline{2}$, $\rho_1 = .9$, $\rho_2 = .1$. Due to increased mixing of the distributions with the larger sample sizes, the portion of the sample classified decreases greatly, but with the compensation of reduced misclassifications of the index sample. If the primary consideration is with minimizing the misclassifications, this effect is of no concern. However, if a substantial portion of both

TABLE 25

EFFECT OF ALLOWED PROBABILITY OF MISCLASSIFICATION.
$PA_1 = PA_2 = 0.05$

σ^2	$\underline{\mu}$	ρ_1	ρ_2	Type	Sample Size	P*(1,1)	P*(2,1)	P*(0,1)	P*(2,2)	P*(1,2)	P*(0,2)
2	$\underline{1}$	.5	.5	K	20	.466	.288	.247	.683	.216	.101
				K(PA)	20	.535	.351	.114	.707	.253	.050
				K	50	.149	.097	.754	.496	.085	.419
				K(PA)	50	.257	.154	.590	.600	.133	.267
				K	100	.054	.052	.894	.401	.037	.562
				K(PA)	100	.186	.099	.716	.515	.107	.378
2	$\underline{2}$	.9	.1	K	20	.839	.116	.046	.793	.194	.014
				K(PA)	20	.861	.116	.023	.762	.220	.018
				K	100	.347	.018	.635	.795	.054	.151
				K(PA)	100	.733	.080	.180	.860	.124	.016

samples must be classified, even at the risk of increased misclassification, then a trade-off procedure between the portion of the sample classified and the portion of the sample misclassified is necessary. A simple method for controlling the probability of misclassification while trying to increase the sample classified was explained above. From Table 25 it can be seen that in the case $\rho_1 = \rho_2 = .5$, the effect of the allowed probability of misclassification is not too impressive, but in the case $\rho_1 = .9$, $\rho_2 = .1$ with a sample size of 100, $P^*(1,1)$ increases from .347 to .733 with relatively minor increases in the probabilities of misclassification.

Examining Tables 16, 17, and 18, the effect of the principal components transformation used with Kendall's method may be evaluated with respect to the index sample. When $\rho_1 = .9$, $\rho_2 = .1$, the transformation is of value when $\underline{\mu} = \underline{0}$. In this case $P^*(1,1) = .246$ and $P^*(2,2) = .458$ when no transformation is used, and $P^*(1,1) = .546$ and $P^*(2,2) = .754$ when the principal components transformation is used. $P^*(1,1)$ and $P^*(2,2)$ are not increased significantly when $\rho_1 = .9$, $\rho_2 = .1$, and $\underline{\mu} \neq \underline{0}$. When $\rho_1 = .1$, $\rho_2 = .9$, the only case for which the principal components transformation gives an improvement is when $\sigma^2 = 2$, $\underline{\mu} = \underline{0}$. In the remainder of the cases it reduces the performance considerably. When $\rho_1 = -.1$, $\rho_2 = .9$, the only case in which the transformation improves the performance of Kendall's method is for $\sigma^2 = 2$, $\underline{\mu} = \underline{1}$. It is worth noting that when the principal components transformation is used, almost all of the initial sample is correctly assigned, but the performance with the index sample tends to be very poor. For example, in the case $\rho_1 = .1$, $\rho_2 = .9$, $\sigma^2 = 2$, $\underline{\mu} = \underline{1}$, $\hat{P}(1,1) = .954$, $\hat{P}(2,2) = .885$, but $P^*(1,1) = .617$, and $P^*(2,2) = .645$.

3.4 Comparison in Cases When Multivariate Normality Does Not Apply

The analysis so far has allowed the possibility of unequal variance-covariance matrices but multivariate normality still has been assumed. Some sampling experiments have been done with non-normal distributions. The distributions considered, the Cauchy and the uniform, were selected because of the difficulty of distinguishing between these distributions and the normal on the basis of a small sampling from the distribution.

In the Cauchy sampling experiments, samples were selected from each of two populations, Π_1 and Π_2:

Π_1: $\underline{x}' = (x_1, x_2, x_3, x_4, x_5)$. All x_i's are independent and distributed as Cauchy random variables $(\Pi^{-1}(1 + x_i^2)^{-1})$

Π_2: $\underline{y}$ (px1) $\sim$ MVN($\underline{2}$,I), where $\underline{2}' = (2, 2, 2, 2, 2)$

In the uniform sampling experiments, samples were obtained from each of two populations Π_1 and Π_2:

Π_1: $\underline{x}' = (x_1, x_2, x_3, x_4, x_5)$. All x_i's are independent and distributed as uniform, U(-1, 1) random variables

Π_2: $\underline{y} \sim$ MVN($\underline{\mu}$,I). $\underline{\mu} = \underline{0}, \underline{1}, \underline{2}$.

In one experiment Π_1 was as above, but

Π_2: $\underline{y}' = (y_1, y_2, y_3, y_4, y_5)$. All y_i's are independent and distributed as U(0, 2) random variables.

The sampling experiments are described in more detail in Table 26. The results of using Kendall's method are given in Table 27, and the results of using LDA in Table 28. A comparison of the results is given in Table 29. In experiments 1, 2, and 3, Π_1 is Cauchy and Π_2 is multivariate normal. In experiment 1 there is a relatively large separation between the means. Kendall's method and LDA give comparable P_T^* values, but LDA gives the rather large $P^*(2,1)$ value of 0.31. In experiment two the sample size is increased from 20 to 100. Now only 2% of population two is classified using Kendall's method since with the larger sample size there is more mixing of the samples from the two populations. In experiment three there is an allowed probability of misclassification, $PA_1 = PA_2 = 0.05$, and more of the sample can be assigned.

Experiment one can serve as an example where Kendall's method would be preferred to LDA if the maximum probability of misclassification were of concern.

In experiments 4, 5, and 6, Π_1 is U(-1,1), and Π_2 is multivariate normal. In experiment four, LDA is, of course, not applicable, since there is zero mean differences between the populations. However, experiment four would indicate the superiority of Kendall's method for small mean differences. In experiments five and six the mean differences become increasingly large and LDA performs better. In experiment seven Π_1 is U(-1,1) and Π_2 is U(0,2). Both Kendall's method and LDA perform well, but LDA gives the lower probability of misclassification.

All of the distributions in the examples considered so far have been symmetrical. Two sampling experiments were done using

TABLE 26

DESCRIPTION OF SAMPLING EXPERIMENTS FOR NON-NORMAL DISTRIBUTIONS

C: Cauchy; U: Uniform; n_1, n_2: sample sizes; PA_1, PA_2: allowed probabilities of misclassification

Experiment Number	Description
1	$\Pi_1 = C$, $\Pi_2 = MVN(\underline{2}, I)$, $n_1 = n_2 = 20$
2	$\Pi_1 = C$, $\Pi_2 = MVN(\underline{2}, I)$, $n_1 = n_2 = 100$
3	$\Pi_1 = C$, $\Pi_2 = MVN(\underline{2}, I)$, $n_1 = n_2 = 100$; $PA_1 = PA_2 = .05$
4	$\Pi_1 = U(-1,1)$, $\Pi_2 = MVN(\underline{0}, I)$, $n_1 = n_2 = 20$
5	$\Pi_1 = U(-1,1)$, $\Pi_2 = MVN(\underline{1}, I)$, $n_1 = n_2 = 20$
6	$\Pi_1 = U(-1,1)$, $\Pi_2 = MVN(\underline{2}, I)$, $n_1 = n_2 = 20$
7	$\Pi_1 = U(-1,1)$, $\Pi_2 = U(0,2)$, $n_1 = n_2 = 20$

TABLE 27

SUMMARY OF RESULTS OF SAMPLING EXPERIMENTS WITH CAUCHY AND UNIFORM DISTRIBUTIONS USING KENDALL'S METHOD

Experiment Number	$\hat{P}(1,1)$	$\hat{P}(2,2)$	P*(1,1)	P*(2,1)	P*(0,1)	P*(2,2)	P*(1,2)	P*(0,2)
1	.999	.974	.838	.148	.013	.827	.137	.036
2	.732	.016	.733	.003	.264	.015	.020	.965
3	.854	.175	.840	.027	.133	.114	.080	.906
4	.671	.915	.341	.422	.237	.755	.201	.045
5	.987	.995	.850	.132	.017	.794	.197	.009
6	1.000	1.000	.915	.055	.040	.892	.089	.019
7	1.000	.994	.855	.123	.022	.857	.121	.022

TABLE 28

SUMMARY OF RESULTS OF SAMPLING EXPERIMENTS WITH CAUCHY AND UNIFORM DISTRIBUTIONS USING LINEAR DISCRIMINANT ANALYSIS

Experiment Number	$\hat{P}(1,1)$	$\hat{P}(2,2)$	P*(1,1)	P*(2,1)	P*(2,2)	P*(1,2)
1	.799	.968	.689	.311	.953	.047
2	.704	.959	.703	.297	.954	.046
4	.668	.629	.578	.422	.448	.552
5	.975	.901	.955	.045	.836	.164
6	1.000	.998	1.000	0.000	.980	.020
7	.985	.978	.964	.036	.969	.031

TABLE 29

COMPARISON OF THE RESULTS OF SAMPLING EXPERIMENTS WITH CAUCHY AND UNIFORM DISTRIBUTIONS USING KENDALL'S METHOD AND LINEAR DISCRIMINANT ANALYSIS

Experiment Number	Method	$P^*(2,1)$	$P^*(1,2)$	P^*_T
1	K	.148	.137	.143
	LDA	.311	.047	.179
2	K	.003	.020	.012
	LDA	.297	.046	.172
3	K	.027	.080	.049
4	K	.422	.201	.312
	LDA	.422	.552	.487
5	K	.132	.197	.165
	LDA	.045	.164	.105
6	K	.055	.089	.072
	LDA	.000	.020	.010
7	K	.123	.121	.122
	LDA	.036	.031	.034

distributions which were not symmetric. In experiments eight and nine, the populations were as follows:

Π_1: $\underline{x}' = (x_1, x_2, x_3, x_4, x_5)$: All x_i's are independent, lognormally distributed, $\ln x_i \sim N(0,1)$.

Π_2: Same as Π_1 except the lognormal distribution is shifted by μ, where $\mu = 1, 2$ in experiments eight and nine, respectively.

The results, summarized in Table 30 indicate the distinct superiority of Kendall's method in experiment eight. The two methods perform comparably in experiment nine when there is larger mean difference between the two populations.

TABLE 30

SAMPLING EXPERIMENTS WITH LOGNORMAL DISTRIBUTION
USING KENDALL'S METHOD AND LDA

Experiment Number	Type	$\hat{P}(1,1)$	$\hat{P}(2,2)$	$P^*(0,1)$	$P^*(1,1)$	$P^*(2,1)$	$P^*(0,2)$	$P^*(1,2)$	$P^*(2,2)$
8	K	.985	.976	.007	.814	.179	.042	.164	.794
	LDA	.813	.740	----	.748	.252	----	.316	.684
9	K	1.000	1.000	.030	.848	.122	.022	.046	.932
	LDA	.921	.959	----	.853	.147	----	.053	.947

CHAPTER IV

COMPARISON OF THE PROBABILITIES OF MISCLASSIFICATION FOR THE LINEAR DISCRIMINANT FUNCTION DETERMINED FROM THE SAMPLING EXPERIMENTS WITH THE THEORETICAL VALUES OBTAINED BY 1) GILBERT, 2) OKAMOTO, AND 3) LACHENBRUCH.

4.1 Comparison with Gilbert's Results

Gilbert (7) has considered the effect of unequal variance-covariance matrices on Fisher's linear discriminant function. She has calculated the probability of misclassification $\Pi P(2,1) + (1-\Pi)P(1,2)$ as a function of Π (the a priori probability that a sample comes from population 1), T^2, and d in the case $\Sigma_2 = d\Sigma_1$. No work was done, however, in determining the probability of misclassification when the population parameters are estimated. The results of some of the sampling experiments already considered in this thesis may be used to provide the expected value of the probability of misclassification when the population parameters are estimated.

Consider two populations, $\underline{x}_1 \sim MVN(\underline{\mu}_1, \Sigma_1)$ and $\underline{x}_2 \sim MVN(\underline{\mu}\ , d\Sigma_1)$. Choosing an orthogonal matrix P such that $P\Sigma P' = I$ and using the transformation $\underline{y}_i = P(\underline{x}_i - \underline{\mu}_i)$, the distributions may be expressed in the canonical form $\underline{x}_1 \sim MVN(\underline{0}, I)$, $\underline{y}_2 \sim MVN(\underline{\nu}, D)$, where $D = \mathrm{Diag}(d,d,\ldots,d)$. The total probability of misclassification is minimized by the rule which assigns an observation to population two whenever $\log [(1 - \Pi)f_2(\underline{x})/\Pi f_1(\underline{x})] > 0$ and to population one otherwise.

Using the linear discriminant function, an observation is classified as population two whenever

$$\sum_{i=1}^{p} \frac{\nu_i}{\Pi + (1 - \Pi)d}\, y_i > C \tag{24}$$

and to population one otherwise. The expression on the left side is distributed as

$$N(0, \sum_{i=1}^{p} \nu_i^2/[\Pi + (1 - \Pi)d]^2) \tag{25}$$

in population one and as

$$N(\Sigma\nu_i^2/[\Pi + (1 - \Pi)d], \sum_{i=1}^{p} \nu_i^2 d/[\Pi + (1 - \Pi)d]^2) \tag{26}$$

in population two. The Mahalanobis distance, T^2, is

$$\underline{\nu}'[\Pi I + (1 - \Pi)dI]^{-1}\underline{\nu} = \Sigma\nu_i^2/[\Pi + (1 - \Pi)d] \tag{27}$$

Hence, $P(2,1) = 1 - \Phi\left[([\Pi + (1 - \Pi)d]/T^2)^{\frac{1}{2}}C\right]$ (28)

and $P(1,2) = \Phi\left[([\Pi + (1 - \Pi)d]dT^2)^{\frac{1}{2}}(C - T^2)\right]$ (29)

where $\Phi(z^*) = P(Z < z^*)$ and $Z \sim N(0,1)$

The cutoff point C is chosen to minimize the total probability of misclassification. Gilbert finds that

$$C = \frac{1}{1 - d}\left\{T^2 \pm (dT^2)^{\frac{1}{2}}\left[T^2 + \frac{(d - 1)}{\Pi + (1 - \Pi)d}\left(\log d + 2 \log \frac{\Pi}{1 - \Pi}\right)\right]^{\frac{1}{2}}\right\}$$

$$\text{when } d \neq 1 \qquad (30)$$

$$C = \frac{1}{2}T^2 + \log [\Pi/(1 - \Pi)] \qquad \text{when } d = 1 \qquad (31)$$

Gilbert notes that the only instance when the optimal cutoff point is not given by equation (30) or (31) is when assigning all the observations to the same population yields a lower total probability of misclassification.

For the case in which we are interested,

$$\Sigma_1 = (1 - \rho)I + \rho Epp$$

$$\Sigma_2 = \sigma^2[(1 - \rho)I + \rho Epp]$$

and $d = \sigma^2$. T^2 has been previously calculated when $\underline{x}_1 \sim MVN(0,\Sigma_1)$, $\underline{x}_2 \sim MVN(\underline{\mu}_2,\Sigma_2)$, using equation (23). Since T^2 is invariant under linear transformations, this value may be used in equations (30) and (31) to calculate P(2,1) and P(1,2). For the analysis in this thesis, $\Pi = \frac{1}{2}$ always, so

$$P(2,1) = 1 - \Phi\left[([1 + \sigma^2]/2T^2)^{\frac{1}{2}}C\right] \qquad (32)$$

$$P(1,2) = \Phi\left[([1 + \sigma^2]/2\sigma^2T^2)^{\frac{1}{2}}(C - T^2)\right] \qquad (33)$$

where now

$$C = \frac{1}{1 - \sigma^2}\left\{T^2 \pm \sigma T\left[T^2 + \frac{2(\sigma^2 - 1)}{1 + \sigma^2}\left[\log \sigma^2\right]\right]^{\frac{1}{2}}\right\}$$

$$\sigma^2 \neq 1 \qquad (34)$$

$$C = \frac{1}{2}T^2 \qquad \sigma^2 = 1 \qquad (35)$$

Gilbert tables the values of the total probability of misclassification for selected values of Π, T^2, and d. The particular values of $\underline{\mu}$ chosen in this study resulted in T^2 values which were not tabled. Equations (32) and (33) were used to calculate P(2,1) and P(1,2) and the probability of misclassification,

$$P_T = \frac{1}{2}\left\{P(2,1) + P(1,2)\right\}. \qquad (36)$$

The results of these calculations are given in Table 31.

In Table 32 the theoretical values of the probabilities of misclassification for linear discriminant analysis are compared with the sampling experiment results.

From Table 32 it is seen that, as expected, $\hat{P}_T$ underestimates P_T, and in most cases $\hat{P}(2,1)$ underestimates P(2,1), and $\hat{P}(1,2)$ underestimates P(1,2). It should also be noted that $\hat{P}_T$ is a poor estimate of P_T^* except in the cases when the sample size is 100 (denoted by a prime in Table 32).

TABLE 31

THEORETICAL PROBABILITIES OF MISCLASSIFICATION

Experiment Number	ρ_1	ρ_2	σ^2	$\underline{\mu}$	T^2	P(2,1)	P(1,2)	P_T
3	.5	.5	2	$\underline{1}$	1.12	.187	.386	.287
5	.1	.1	2	$\underline{1}$	2.38	.152	.271	.212
8	.5	.5	2	$\underline{2}$	4.44	.104	.174	.139
10	.1	.1	2	$\underline{2}$	9.52	.046	.069	.058
11	.5	.5	1	$\underline{1}$	1.66	.261	.261	.261
13	.1	.1	1	$\underline{1}$	3.58	.173	.173	.173
14	.5	.5	1	$\underline{2}$	6.66	.109	.109	.109
16	.1	.1	1	$\underline{2}$	14.28	.029	.029	.029

TABLE 32

COMPARISON OF THE THEORETICAL PROBABILITIES OF MISCLASSIFICATION WITH THOSE OBTAINED FROM SAMPLING EXPERIMENTS

Experiment Number	$P(2,1)$	$P(1,2)$	P_T	$\hat{P}(2,1)$	$\hat{P}(1,2)$	$\hat{P}_T$	$P^*(2,1)$	$P^*(1,2)$	P^*_T
3	.187	.386	.287	.232	.296	.264	.294	.380	.337
5	.152	.271	.212	.148	.221	.185	.191	.300	.246
5'[†]	.152	.271	.212	.175	.247	.211	.179	.261	.220
8	.104	.174	.139	.077	.156	.117	.112	.217	.165
10	.046	.069	.058	.025	.068	.047	.047	.116	.082
11	.261	.261	.261	.219	.215	.217	.309	.291	.300
13	.173	.173	.173	.135	.148	.142	.195	.210	.203
13'	.173	.173	.173	.169	.175	.172	.173	.180	.177
14	.109	.109	.109	.073	.078	.076	.124	.119	.122
16	.029	.029	.029	.021	.017	.019	.042	.042	.042

[†]The prime denotes a repetition of the experiment with a sample size of 100.

4.2 Comparison of the sampling experiment results in the cases $\Sigma_1 = \Sigma_2$ with the results of Okamoto and with the results of Lachenbruch

It is interesting to compare the P_T^* values with those which may be obtained in a different way. Okamoto (27) has derived an asymptotic expansion for the distribution of the linear discriminant function statistic W (the distribution of the sample discriminant function). In the particular cases of interest in this thesis, the expansion is

$$\Pr\left\{W < 0 | \Pi_1\right\} = \Phi(-\frac{T}{2}) + \frac{a_1}{n_1} + \frac{a_2}{n_2} + \frac{a_3}{n}$$

$$+ \frac{b_{11}}{n_1^2} + \frac{b_{22}}{n_2^2} + \frac{b_{12}}{n_1 n_2} + \frac{b_{13}}{n_1 n} + \frac{b_{23}}{n_2 n} + \frac{b_{33}}{n^2} + O_3, \qquad (37)$$

where n_1 and n_2 are the sizes of samples from populations Π_1 and Π_2, respectively, and $n_1 + n_2 - 2 = n$. In Table 1 of his paper, Okamoto gives the values of the coefficients for a number of values of p, including the case of interest to us, p = 5, and for a number of T values, T = 1, 2, 3, 4, 6, 8. The particular T values in our case are not included in the Table, however. In order to obtain a highly accurate result, the coefficients should be calculated for these particular T values. Okamoto's expansion has been applied to the cases in Table 32 using, however, Okamoto's tabled coefficients for the T case closest to our particular T values. The results of the calculations are given in Table 33. It can be seen that the P_T^* values obtained in the sampling experiments and the P_T^* values calculated using Okamoto's asymptotic expansion are very close.

TABLE 33

COMPARISON OF P_T^* VALUES

Experiment Number	P_T^* Sampling Experiment	P_T^* Okamoto	P_T^* Lachenbruch
11	.300	.310	.302
13	.203	.198	.209
13'	.177	.180	.181
14	.122	.125	.127
16	.042	.037	.045

Lachenbruch (4) provides another way to calculate P_T^*. He considers the sample discriminant function

$$D_S(\underline{x}) = [\underline{x} - \tfrac{1}{2}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2)]'S^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2) \tag{38}$$

which is conditionally (on $\underline{\bar{x}}_1$, $\underline{\bar{x}}_2$, and S) normally distributed and has mean (in the kth group)

$$D_S(\underline{\mu}_k) = [\underline{\mu}_k - \tfrac{1}{2}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2)]'S^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2) \tag{39}$$

and variance (in either group)

$$V_D = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)'S^{-1}\Sigma S^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2) \tag{40}$$

He finds that

$$E[D_S(\underline{\mu}_k)] = \frac{n_1 + n_2 - 2}{2(n_1 + n_2 - p - 3)} \left(T^2(-1)^{k+1} - \frac{p(n_2 - n_1)}{n_1 n_2} \right) \tag{41}$$

$$E[V_D] = \left[T^2 + \frac{p(n_1 + n_2)}{n_1 n_2} \right] \cdot \frac{(n_1 + n_2 - 3)(n_1 + n_2 - 2)^2}{(n_1 + n_2 - p - 2)(n_1 + n_2 - p - 3)(n_1 + n_2 - p - 5)} \tag{42}$$

For n_1 and n_2 sufficiently large, the unconditional distribution is very close to normal, and

$$\overline{P}_1 = \Phi \left\{ E[-D_S(\underline{\mu}_1)] / [E(V_D)]^{\frac{1}{2}} \right\} \tag{43}$$

and $$\overline{P}_2 = \Phi \left\{ E[D_S(\underline{\mu}_2)] / [E(V_D)]^{\frac{1}{2}} \right\} \tag{44}$$

will supply approximate values for P_1 and P_2, where

$$P_1 = P(D_S(\underline{x}) < 0 \mid \underline{x} \varepsilon \Pi_1) \tag{45}$$

$$P_2 = P(D_S(\underline{x}) > 0 \mid \underline{x} \varepsilon \Pi_2) \tag{46}$$

Equations (41), (42), (43), and (44) were used to calculate P_T^* by Lachenbruch's method. The results are given in Table 33. The methods of Okamoto and Lachenbruch both agree well with each other and with the results of the sampling experiments.

CHAPTER V

THE MODIFIED BARTLETT AND PLEASE METHOD

5.1 Introduction

Bartlett and Please have obtained a linear discriminant function in the case of zero mean differences when the variance-covariance matrices are of the form considered in this thesis, i.e.,

$$\Sigma_1 = (1 - \rho_1)I + \rho_1 Epp$$

$$\Sigma_2 = \sigma^2[(1 - \rho_2)I + \rho_2 Epp]$$

In the case of zero mean differences it is of interest to compare Kendall's order-statistic method or variation thereof, with this Bartlett and Please method. However, the cutoff point which they obtain for equal probabilities of misclassification is shown to be incorrect. This chapter will be concerned with the development of a modified Bartlett and Please method which does provide equal probabilities of misclassification. In section 5.4 this modified Bartlett and Please method will be compared with Kendall's method.

5.2 Derivation of the Modified Bartlett and Please Method

Let us consider two populations

$$f_i(\underline{x}) \sim MVN(\underline{0}, \Sigma_i) \qquad (i = 1,2) \tag{47}$$

where

$$\Sigma_i = \sigma_i^2 \left| (1 - \rho_i)I + \rho_i Epp \right| \qquad (i = 1,2) \tag{48}$$

A discriminant function may be derived by considering the likelihood ratio principle:

$$\text{If } f_1(\underline{x})/f_2(\underline{x}) \geq \lambda \qquad \text{assign to } \Pi_1 \tag{49}$$

Expanding the ratio, using the logarithmic transformation,

$$\ln \left[\frac{\exp[-\frac{1}{2}\underline{x}'\Sigma_1^{-1}\underline{x}]}{(2\Pi)^{p/2}|\Sigma_1|^{\frac{1}{2}}} \middle/ \frac{\exp[-\frac{1}{2}\underline{x}'\Sigma_2^{-1}\underline{x}]}{(2\Pi)^{p/2}|\Sigma_2|^{\frac{1}{2}}} \right]$$

$$= \ln \frac{1}{|\Sigma_1|^{\frac{1}{2}}} - \frac{1}{2}\underline{x}'\Sigma_1^{-1}\underline{x} - \ln \frac{1}{|\Sigma_2|^{\frac{1}{2}}} + \frac{1}{2}\underline{x}'\Sigma_2^{-1}\underline{x} \geq \ln\lambda \tag{50}$$

$$\ln|\Sigma_1| + \underline{x}'\Sigma_1^{-1}\underline{x} - \ln|\Sigma_2| - \underline{x}'\Sigma_2^{-1}\underline{x} < -2\ln\lambda \tag{51}$$

When $\Sigma_1 = \Sigma_2$ it is a well-known result that $\lambda = 1$ provides equal probabilities of misclassification. Bartlett and Please have inadvertently assumed that $\lambda = 1$ also provides equal probabilities of misclassification when $\Sigma_1 \neq \Sigma_2$. When $\lambda = 1$,

$$\underline{x}'\Sigma_1^{-1}\underline{x} - \underline{x}'\Sigma_2^{-1}\underline{x} < -\ln|\Sigma_1| + \ln|\Sigma_2| = \ln(|\Sigma_2|/|\Sigma_1|) \tag{52}$$

In the particular case that we are considering, $\sigma_1^2 = 1$, $\sigma_2^2 = \sigma^2$, so

$$\text{if} \qquad \underline{x}'\Sigma_1^{-1}\underline{x} - \underline{x}'\Sigma_2^{-1}\underline{x} < p \ln \sigma^2 \quad \text{assign to } \Pi_1 \tag{53}$$

The discriminant function equation (53) may be written in the form

$$az_1 - bz_2 = K^* \tag{54}$$

where $\quad z_1 = \underline{x}'\underline{x} \qquad z_2 = (E_{1p}\underline{x})^2 \qquad E_{1p} = (1,1,\ldots,1)$

$$a = \frac{1}{1 - \rho_1} - \frac{1}{\sigma^2(1 - \rho_2)} \tag{55}$$

$$b = \frac{\rho_1}{1 - \rho_1} \cdot \frac{1}{1 + (p - 1)\rho_1} - \frac{\rho_2}{\sigma^2(1 - \rho_2)} \cdot \frac{1}{1 + (p - 1)\rho_2} \tag{56}$$

When $\rho_1 = \rho_2 = \rho$, the discriminant function becomes

$$Z_1 + \frac{\rho}{1 + (p - 1)\rho} Z_2 = \frac{(1 - \rho)p \ln \sigma^2}{1 - \frac{1}{\sigma^2}} \tag{57}$$

As noted earlier the cutoff point given by the right side of equation (57) does not provide equal probabilities of misclassification. Using this cutoff point, however, the Bartlett and Please rule is

$$\text{If } Z_1 - \frac{\rho}{1 + (p - 1)\rho} Z_2 \leq \frac{(1 - \rho)p \ln \sigma^2}{1 - \frac{1}{\sigma^2}} \quad \text{assign to } \Pi_1 \tag{58}$$

$$\text{If } Z_1 - \frac{\rho}{1 + (p - 1)\rho} Z_2 > \frac{(1 - \rho)p \ln \sigma^2}{1 - \frac{1}{\sigma^2}} \quad \text{assign to } \Pi_2 \tag{59}$$

A method for obtaining the cutoff point which does provide equal probabilities of misclassification now will be developed. It may be shown that

$$U = Z_1 - \frac{\rho}{1 + (p - 1)\rho} Z_2 \sim \sigma_i^2(1 - \rho)\chi_p^2. \tag{60}$$

Hence the rule becomes

$$\text{Assign } \underline{x} \text{ to } \Pi_1 \text{ when } U \leq K \tag{61}$$

$$\text{Assign } \underline{x} \text{ to } \Pi_2 \text{ when } U > K \tag{62}$$

with K to be determined so that there are equal probabilities of misclassification. The probabilities of misclassification are:

$$\alpha_1 = \int_{U \leq K} f_2(\underline{x})d\underline{x}; \quad \alpha_2 = \int_{U > K} f_1(\underline{x})d\underline{x} \tag{63}$$

For $\sigma_1^2 = 1$, $\sigma_2^2 = \sigma^2$, these become

$$\alpha_1 = \int_0^{K/1-\rho} \emptyset p(\omega)d\omega \text{ and } \alpha_2 = \int_{K/\sigma^2(1-\rho)}^{\infty} \emptyset p(\omega)d\omega, \tag{64}$$

respectively, where $\emptyset p \sim \chi^2(p)$.

The correct procedure for equal probabilities of misclassification would be to choose that value of K which makes the two integrals in (64) equal.

Now the function $az_1 - bz_2$ may be written as

$$\omega = a(1 - \rho_i)\sigma_i^2\, \chi_{p-1}^2 + (a - bp)\sigma_i^2 \left\{ 1 + (p - 1)\rho_i \right\} \chi_1^2 \tag{65}$$

and therefore,

$$E(\omega) = a(1 - \rho_i)\sigma_i^2(p - 1) + (a - bp)\sigma_i^2 \left\{ 1 + (p - 1)\rho_i \right\} = f_i \tag{66}$$

$$V(\omega) = 2a^2(1 - \rho_i)^2\sigma_i^4(p - 1) + 2(a - bp)^2\sigma_i^4 \left\{ 1 + (p - 1)\rho_i \right\}^2 = 2g_i \tag{67}$$

Then

$$\xi = \frac{f_i\omega}{g_i} \sim \chi^2\left(\frac{f_i^2}{g_i}\right)$$

$$\alpha_1 = \Pr(\omega \leq K \mid \underline{x} \text{ comes from } \Pi_2) = \int_0^{f_2K/g_2} \emptyset n_2(\xi)d\xi \tag{68}$$

$$\alpha_2 = \Pr(\omega > K|\underline{x} \text{ comes from } \Pi_1) = \int_{f_1K/g_1}^{\infty} \emptyset n_1(\xi)d\xi \tag{69}$$

The value of K is to be chosen so that the two integrals in (68) and (69) are equal. The procedure used to find the value of K was a Newton-Raphson method. To solve the equation $H(x) = 0$, a sequence of x values are calculated:

$$x_{i+1} = x_i - \frac{H(x_i)}{H'(x_i)} \tag{70}$$

When the difference between x_{i+1} and x_i becomes acceptably small, the x_{i+1} th value will be the solution of the equation. Here

$$H(K) = \int_{o}^{f_2K/g_2} \emptyset n_2(\xi)d\xi - \int_{f_1K/g_1}^{\infty} \emptyset n_1(\xi)d\xi \tag{71}$$

$$\frac{dH(K)}{dK} = h(K) = \frac{f_2}{g_2}\emptyset n_2\left(\frac{f_2K}{g_2}\right) + \frac{f_1}{g_1}\emptyset n_1\left(\frac{f_1K}{g_1}\right). \tag{72}$$

The parameters are estimated by the following procedure:

$$\text{Let } \sum_{r=1}^{n_i} Z_{2r} = A_i \qquad (i = 1,2) \tag{73}$$

$$\sum_{r=1}^{n_i} Z_{3r} = B_i \qquad (i = 1,2) \tag{74}$$

where

$$Z_{3r} = Z_{1r} - Z_{2r}/p \tag{75}$$

Then

$$E(A_i) = \sigma_i^2 p \left\{ 1 + (p - 1)\rho_i \right\} n_i \tag{76}$$

$$E(B_i) = \sigma_i^2 (1 - \rho_i) n_i (p - 1) \tag{77}$$

Hence

$$\frac{(p - 1)(A_i - B_i) - B_i}{p^2 B_i} \text{ estimates } \frac{\rho_i}{1 - \rho_i} \tag{78}$$

The estimated function is

$$\hat{a} Z_1 - \hat{b} Z_2 \tag{79}$$

where

$$\hat{a} = \frac{n_1(p - 1)}{B_1} - \frac{n_2(p - 1)}{B_2} \tag{80}$$

$$\hat{b} = \frac{(p - 1)(A_1 - B_1) - B_1}{p^2 B_1} \cdot \frac{p n_1}{A_1} - \frac{(p - 1)(A_2 - B_2) - B_2}{p^2 B_2} \tag{81}$$

$(p - 1)A_1 - pB_1 \quad n_1 \quad (p - 1)A_2 - pB_2 \quad n_2$

Lachenbruch (2) has developed an excellent method for determining the probabilities of misclassification in linear discriminant analysis. In this method, one observation is omitted from the sample used to calculate the discriminant function and the discriminant function is then used to classify the omitted observation. This is done in turn for each observation. The number of misclassifications provides a good indication of the probability of misclassification. Lachenbruch's method provides a much better estimate of the probability of misclassification than is provided by the resubstitution method, in which the entire sample is used to calculate the discriminant function and this function is in turn used to classify each of the observations. A method related to that of Lachenbruch is developed in this thesis and this method is compared with the resubstitution method. It is not strictly a Lachenbruch method since the cutoff points for equal probabilities of misclassification are not recalculated each time an observation is omitted from the sample. Let

$$g_r = \hat{a}_r Z_{1r} - \hat{b}_r Z_{2r} \tag{83}$$

be the estimated function omitting the rth observation from Π_1, where

$$\hat{a}_r = \frac{(n_1 - 1)(p - 1)}{(B_1 - Z_{3r})} - \frac{n_2(p - 1)}{B_2} \tag{84}$$

$$\hat{b}_r = \frac{(p - 1)(A_1 - Z_{2r}) - p(B_1 - Z_{3r})}{p(B_1 - Z_{3r})} \cdot \frac{(n_1 - 1)}{(A_1 - Z_{2r})}$$

$$- \frac{(p - 1)\, A_2 - pB_2}{pB_2} \cdot \frac{n_2}{A_2} \qquad (85)$$

Similarly, let

$$h_r = \hat{a}_r Z_{1r} - \hat{b}_r Z_{2r} \qquad (86)$$

be the estimated function omitting the rth observation from Π_2, where

$$\hat{a}_r = \frac{n_1(p - 1)}{B_1} - \frac{(n_2 - 1)(p - 1)}{(B_2 - Z_{3r})} \qquad (87)$$

$$\hat{b}_r = \frac{(p - 1)A_1 - pB_1}{pB_1} \cdot \frac{n_1}{A_1} \qquad (88)$$

$$- \frac{(p - 1)(A_2 - Z_{2r}) - p(B_2 - Z_{3r})}{p(B_2 - Z_{3r})} \cdot \frac{n_2 - 1}{(A_2 - Z_{2r})}$$

5.3 Application to an example considered by Bartlett and Please

Bartlett and Please considered the problem of discriminating between monozygotic and dizygotic pairs of twins. Ten variates were and both male and female sets of twins were considered. There were 15 samples each from the monozygotic and dizygotic male twins and 15 samples each from the monozygotic and dizygotic female twins.

The values of σ^2 and ρ_i estimated from the samples were as follows:

Case I	(Females)	$\hat{\sigma}^2 = 3.760$	$\hat{\rho}_1 = \hat{\rho}_2 = 0.160$
Case II	(Males)	$\hat{\sigma}^2 = 2.236$	$\hat{\rho}_1 = \hat{\rho}_2 = 0.223$

These two cases considered by Bartlett and Please were reexamined. Random samples with parameters equal to the estimated parameters of Bartlett and Please were generated. Experiments with total sample sizes of 30 and 200 were performed under a variety of conditions. All experiments were repeated 10 times for each set of parameters and set of conditions. The two general cases considered then were

$$\text{Case I:} \quad \sigma_1^2 = 1 \qquad \sigma_2^2 = 3.76$$
$$\rho_1 = \rho_2 = 0.160 \qquad p = 10$$

$$\text{Case II:} \quad \sigma_1^2 = 1 \qquad \sigma_2^2 = 2.236$$
$$\rho_1 = \rho_2 = 0.223 \qquad p = 10$$

Solving the equation $H(K) = 0$ yielded the following values of K with corresponding values of α.

$$\text{Case I:} \quad K = 12.38, \ \alpha_1 = \alpha_2 = 0.077$$
$$\text{Case II:} \quad K = 8.03, \ \alpha_1 = \alpha_2 = 0.176,$$

where K is the cutoff point for the discriminant function

$$aZ_1 - bZ_2 = K$$

Values of a and b were calculated from Equations (55) and (56).

Case I: $a = 0.87386$

$b = 0.05730$

Case II: $a = 0.74212$

$b = 0.05504$

The cutoff point, K*, in the Bartlett and Please discriminant function

$$aZ_1 - bZ_2 = K^*$$

is equal to $p \cdot \log_e \sigma^2$, so for Case I it is 13.24 and for Case II it is 8.60. Using these cutoff points, the true probability of misclassification may be calculated directly from the equations

$$\alpha_1 \doteq \int_0^{f_2K/g_2} \emptyset n_2(\xi)d\xi \qquad (89)$$

$$\alpha_2 \doteq \int_{f_1K/g_1}^{\infty} \emptyset n_1(\xi)d\xi \qquad (90)$$

These values were found to be as follows:

Case I: $\alpha_1 = 0.096$

$\alpha_2 = 0.054$

Case II: $\alpha_1 = 0.211$

$\alpha_2 = 0.136$

The values of α_1 and α_2 estimated by Bartlett and Please from their data were $\alpha_1 = 0.13$, $\alpha_2 = 0.00$ for Case I and $\alpha_1 = 0.47$, $\alpha_2 = 0.27$ for Case II.

The four discriminant functions,

Equal Probability Misclassification

Case I: $0.87386Z_1 - 0.05730Z_2 = 12.38$

Case II: $0.74212Z_1 - 0.05504Z_2 = 8.03$

Bartlett and Please Method

Case I: $0.87386Z_1 - 0.05730Z_2 = 13.24$

Case II: $0.74212Z_1 - 0.05504Z_2 = 8.60$

may be expressed in the form used in the Bartlett and Please article

Equal Probability Misclassification

$$\text{Case I:} \quad Z_1 - 0.06557Z_2 = 14.17$$

$$\text{Case II:} \quad Z_1 - 0.07417Z_2 = 10.82$$

Bartlett and Please Method

$$\text{Case I:} \quad Z_1 - 0.06557Z_2 = 15.15$$

$$\text{Case II:} \quad Z_1 - 0.07417Z_2 = 11.59$$

Using the calculated values of K and K*, the Lachenbruch method may be evaluated by comparing the α_1 and α_2 values obtained by this method with those estimated using equations (68) and (69). The resubstitution method was evaluated in a similar fashion also. In addition, both the Lachenbruch and resubstitution methods were used in conjunction with cutoff points K and K* which were estimated from the data. The results of all of these experiments are summarized in Tables 34 and 35 for Cases I and II, respectively. The symbols used in the tables are defined as follows:

$\tilde{\alpha}_i$: Calculated by Lachenbruch method

$\hat{\alpha}_i$: Calculated by Resubstitution method

$\tilde{\alpha}_i^*$: Calculated by Lachenbruch method with K estimated from sample

$\hat{\alpha}_i^*$: Calculated by Resubstitution method with K estimated from sample.

TABLE 34

SUMMARY OF SAMPLING EXPERIMENT RESULTS
WITH BARTLETT AND PLEASE AND
MODIFIED BARTLETT AND
PLEASE METHODS

Case I (σ^2 = 3.760)

K Theoretical

Type	N_i	α_1	α_2	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Eq. Pr.	15	.077	.077	.120	.087	.080	.087
Eq. Pr.	100	.077	.077	.086	.073	.083	.074
BP	15	.054	.096	.073	.100	.053	.100
BP	100	.054	.096	.064	.094	.056	.094

K Estimated

Type	N_i	$\tilde{\alpha}^*_1$	$\tilde{\alpha}^*_2$	$\hat{\alpha}^*_1$	$\hat{\alpha}^*_2$
Eq. Pr.	15	.093	.060	.053	.073
Eq. Pr.	100	.083	.072	.082	.079
BP	15	.073	.080	.047	.087
BP	100	.062	.093	.055	.094

TABLE 35

SUMMARY OF SAMPLING EXPERIMENT RESULTS
WITH BARTLETT AND PLEASE AND
MODIFIED BARTLETT AND
PLEASE METHODS

Case II (σ^2 = 2.362)

K Theoretical

Type	N_i	α_1	$\tilde{\alpha}_2$	$\tilde{\alpha}_1$	α_2	$\hat{\alpha}_1$	$\hat{\alpha}_2$
Eq. Pr.	15	.176	.176	.240	.133	.194	.153
Eq. Pr.	100	.176	.176	.172	.163	.166	.164
BP	15	.136	.211	.213	.174	.133	.200
BP	100	.136	.211	.145	.197	.139	.199

K Estimated

Type	N_i	$\tilde{\alpha}_1^*$	$\tilde{\alpha}_2^*$	$\hat{\alpha}_1^*$	$\hat{\alpha}_2^*$
Eq. Pr.	15	.233	.153	.167	.147
Eq. Pr.	100	.174	.157	.169	.157
BP	15	.180	.187	.127	.194
BP	100	.122	.187	.132	.187

Comparing α_i and $\hat{\alpha}_i$ it can be seen that the resubstitution method gives better results than the Lachenbruch method, probably because the estimation procedure used in the latter method is poor.

The differences between the performance of the various methods may be seen more clearly by comparing the values of

$$|\alpha_1 - [\alpha_1]| + |\alpha_2 - [\alpha_2]| \tag{91}$$

where $[\alpha_i]$ may be $\tilde{\alpha}_i$, $\hat{\alpha}_i$, $\tilde{\alpha}_i^*$, or $\hat{\alpha}_i^*$. The value of this expression is denoted by $\tilde{D}$, $\hat{D}$, $\tilde{D}^*$, or $\hat{D}^*$, depending on the $[\alpha_i]$ used.

The values of D are summarized in Tables 36 and 37. The superiority of the method of resubstitution is quite apparent for the smaller sample size, although the two methods provide comparable results for the larger sample size.

In the case of estimated cutoff points, the procedure used was not actually a Lachenbruch method, since the cutoff points were not recalculated for each sample after a particular observation had been deleted. This could have been done, but the amount of computation would have been increased considerably. For example, with a sample size of 200 the equation $H(K) = 0$ for the equal probability case would have to be solved 200 times. Although it is doubtful if this refinement would improve the results for the 30 sample size case, some further study will be devoted to it. However, the value of this method as a practical approach would be questionable even if some improvement resulted. This Lachenbruch method would be improved substantially only by impro ing the estimation procedure for $\hat{a}_r$ and $\hat{b}_r$.

TABLE 36

COMPARISON OF ERROR PROBABILITIES ASSOCIATED WITH BARTLETT AND PLEASE AND MODIFIED BARTLETT AND PLEASE METHODS

Case I (σ^2 = 3.760)

Type	N_i	$\tilde{D}$	$\hat{D}$	$\tilde{D}^*$	$\hat{D}^*$
Eq. Pr.	15	.053	.013	.033	.028
Eq. Pr.	100	.013	.009	.011	.007
BP	15	.023	.005	.035	.016
BP	100	.003	.004	.011	.003

Case II (σ^2 = 2.362)

Type	N_i	$\tilde{D}$	$\hat{D}$	$\tilde{D}^*$	$\hat{D}^*$
Eq. Pr.	15	.107	.041	.080	.038
Eq. Pr.	100	.017	.022	.021	.016
BP	15	.114	.014	.068	.026
BP	100	.023	.015	.036	.028

5.4 Comparison of the Modified Bartlett and Please Method with Kendall's Method

The Modified Bartlett and Please method was applied to some of the data sets obtained in the sampling experiments. The modified method was applied to each of the 50 sets of samples and the results averaged, just as has been done in all of the previous sampling experiments discussed so far. The results are given in Table 37. K, as before, refers to Kendall's method; K(PC), to Kendall's method with principal components, and MBP, to the modified Bartlett and Please method. The method of resubstitution was used to obtain $\hat{P}(1,1)$ and $\hat{P}(2,2)$ for the latter method.

The excellent performance of the modified Bartlett and Please method is evident from Table 37. The principal components transformation was applied in the case $\rho_1 = .9$, $\rho_2 = .1$, and, although the performance of Kendall's method was improved considerably, the modified Bartlett and Please method was still superior.

It is, of course, unfair to compare Kendall's method to the modified Bartlett and Please method in cases for which the latter method was designed. It would be interesting to compare Kendall's method to the modified Bartlett and Please method when there actually was a small mean difference.

TABLE 37

COMPARISON OF RESULTS USING THE MODIFIED BARTLETT AND PLEASE METHOD AND KENDALL'S METHOD. $\sigma^2 = 2$ IN ALL CASES

ρ_1	ρ_2	Type	$\hat{P}(1,1)$	$\hat{P}(2,2)$	$P^*(1,1)$	$P^*(2,1)$	$P^*(0,1)$	$P^*(2,2)$	$P^*(1,2)$	$P^*(0,2)$
.1	.1	K	.459	.727	.173	.429	.398	.656	.172	.172
		MBP	.715	.722	.674	.326	----	.702	.298	----
.5	.5	K	.452	.703	.158	.419	.422	.609	.190	.202
		MBP	.699	.698	.670	.330	----	.702	.298	----
.9	.1	K	.487	.824	.205	.339	.456	.697	.186	.117
		K(PC)	.870	.990	.493	.393	.114	.868	.120	.012
		MBP	.966	.947	----	.048	.952	.949	.051	----

CHAPTER VI

COMPARISON OF KENDALL'S METHOD WITH OTHER NONPARAMETRIC TECHNIQUES

6.1 Successive Screening

Feldman, Klein, and Honigfeld (28) have developed a discrimination method which is quite similar to Kendall's order-statistic method. Referring to their hypothetical example given in Table 38, their method can be explained. It is desired to separate out Group B. For scores of four and above, the ratio of Group B to Group A is 10/1; for three and above it is 25/10. The cutoff point for the parameter is that value for which the ratio of Group B to Group A is largest. Each parameter is examined in like fashion and that parameter giving the highest ratio is selected. All samples with that parameter value beyond the cutoff point are eliminated from further consideration, and the procedure is continued on the remaining samples using the remaining variables. The procedure stops when the maximum ratio at any stage falls below a specified limit.

In Kendall's method the cutoff point is chosen so that samples from only one of the populations are beyond the cutoff point or else, using the allowed probability of misclassification, that only a specified number of samples from the other distribution are allowed. In Kendall's method separation of samples from each of the two populations is equally important.

TABLE 38

HYPOTHETICAL DISTRIBUTION OF ITEM RATINGS FOR TWO DIAGNOSTIC GROUPS (FROM FELDMAN, ET AL (29))

	Ordinal Scale				
	1	2	3	4	Total
Group A	30	20	9	1	60
Group B	30	25	15	10	80

It would require only minor changes in Kendall's procedure to obtain that of Feldman, Klein, and Honigfeld. What would be the advantage of modifying Kendall's method in this way? In the case in which the variables are on an ordinal scale but only a small number of values are possible, such as in rating a personality trait or opinion on a scale from 0 through 10 in increments of 1, there may be considerable overlap in samples from the two populations; frequently there would even be complete overlap. In situations like this, modifying the Kendall order-statistic method to incorporate the Feldman, Klein, and Honigfeld ideas would be advisable. Since Kendall's method is more generally applicable than that of Feldman, Klein, and Honigfeld, and since only minor modifications of the computer program based on Kendall's method would be needed to incorporate this optional method, it would appear more advantageous to make this modification than to have two different methods available in the form of separate, but quite similar computer programs.

Feldman, Klein, and Honigfeld list eight advantages in the use of successive screening in medical work. These eight advantages are quoted here since the same claims may be made for Kendall's order-statistic method:

1. No restrictions are placed on the data distributions or the joint distributions.

2. The categories are polythetic, i.e., members of the same class need not have even a single trait in common, but must exhibit a minimum number of alternative trait set members. To some degree this resembles the ambiguity that occurs with keys formed by the summation of weighted items, but the successive screening method stipulates a specified minimum intensity for any trait to be of consequence in the discrimination.

3. Pathognomic signs may be easily recognized, but, in general, classification is made by sign pattern.

4. Certain classes may be 'ruled out' by certain traits.

5. Ordinal traits may have both extremes used for the same effect, i.e., a U-shaped relationship of classification to trait can be utilized.

6. The successive screening technique is essentially a counting procedure and does not utilize mathematical procedures that depend on interval or ratio scales.

7. The procedure makes easily inspected prima facie sense and does not involve such obscurities as suppressor variables.

8. The model systemitizes the sequential screening approach of clinical diagnosis but avoids the problem of making serious misclassifications through single measurement errors, by using multiple alternative traits at each decision point.

6.2 Henrichon and Fu Algorithm

There is another nonparametric discriminant analysis technique which has been well received by specialists in pattern recognition (29), although it has not been reported in the statistical literature. This is the Henrichon-Fu technique (30, 31) which will now be explained. Consider first the univariate case with two populations Π_1 and Π_2. Let $\underline{x} = (x_1, x_2, \ldots, x_{n_1})$ be a set of independent observations from Π_1 and

$\underline{y} = (y_1, y_2, \ldots, y_{n_2})$ be a set of independent observations from Π_2. The $\underline{x}$ and $\underline{y}$ observations are represented respectively by a) and b) in Figure 2 below.

<u>Step 1</u>. (Figure 2,c) - Combine the $\underline{x}$ and $\underline{y}$ observations and order the set $\underline{x}\cup\underline{y}$ according to increasing numerical value. Partition the set $\underline{x}\cup\underline{y}$ so that there are K samples in each cell. In Figure 2, K = 5.

<u>Step 2</u>. (Figure 2, d and e) - Let $C_i (i = 1,2)$ be the cost of misclassifying an observation from $\Pi_i (i = 1,2)$ and let C_0 be the cost of not classifying an observation. In Figure 2, $C_1 = C_2 = 6$, $C_0 = 1$. Count the number of x's and y's in each cell and assign to Π_1, Π_2, or Π_0 (unclassified) according to the following:

If

$$\underset{i}{\text{Min}} \left\{ \sum_{\substack{j=1 \\ j \neq i}}^{2} C_j \text{ (no. of samples from } \Pi_j) \right\} < C_0 K$$

then assign to Π_i (i=1,2).

Otherwise assign to Π_0.

Then combine adjacent cells of the same class.

<u>Step 3</u>. Adjust the cell boundaries by perturbing them a maximum of K/2 samples in either direction and locate the boundary at the point giving best classification (Figure 2, f).

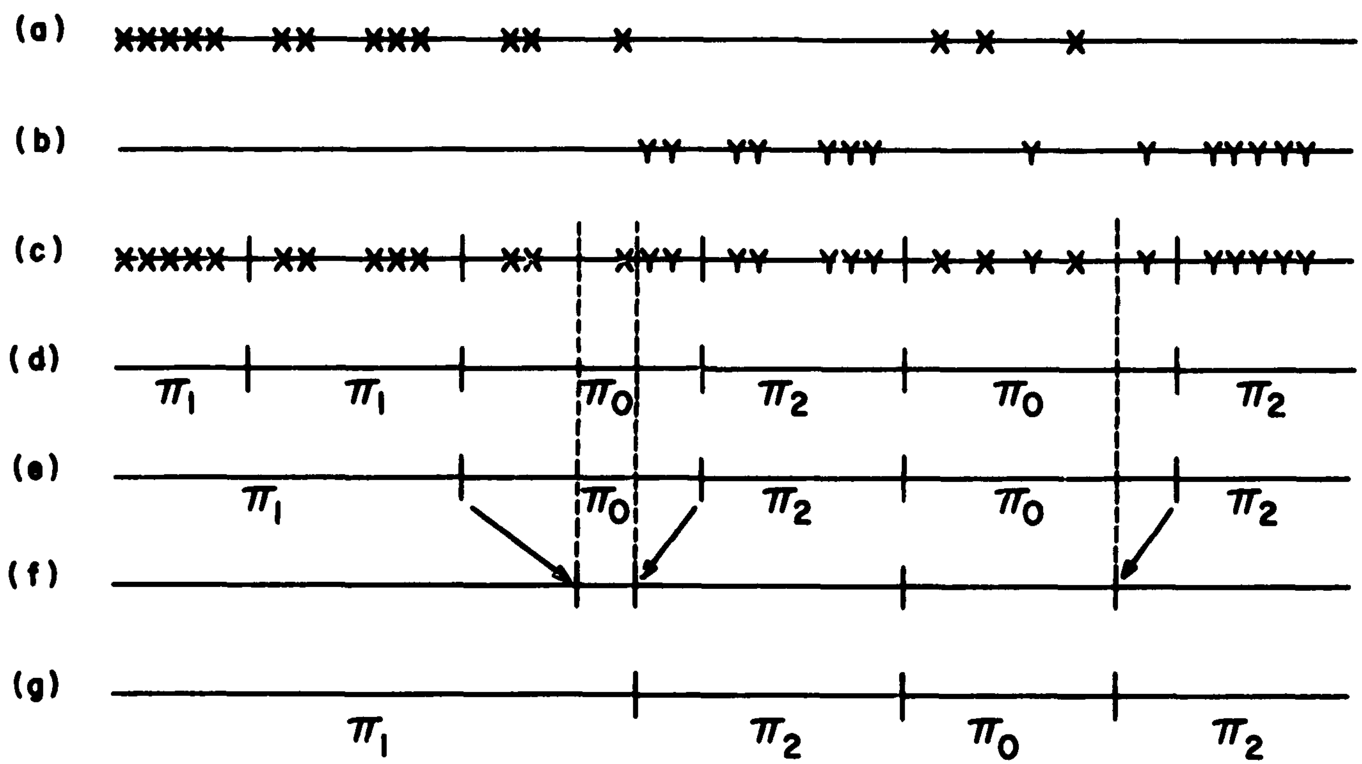

Figure 2. Illustration of Basic Henrichon-Fu Algorithm

Step 4. For any remaining cell with less than K/2 samples, dissolve the cell, placing the samples in the cell into adjacent cells in such a way that there is the smallest increase in misclassification (*Figure* 2, *g*).

Step 5. Repeat Step 2. For this final partition compute the empirical classification statistic, or Score:

$$\text{Score} = \sum_{i=1}^{2} C_i \text{ (No. of samples misclassified from } \Pi_i) \qquad (92)$$

$$+ \; C_0 \text{ (No. of unclassified samples)}$$

The procedure is extended to the multidimensional case by calculating a score for each variate and selecting the variate with the lowest score first. After the space has been partitioned with this first variate, the resulting cells are further partitioned using the variate with the second lowest score. This procedure is continued until there are no new cells formed.

The Henrichon-Fu algorithm is similar to the Kendall algorithm. In the Kendall method observations once classified using a particular variable are removed from further consideration in examining any other variables. In the Henrichon-Fu algorithm the observations are not explicitly removed from further consideration once they are classified as Π_1 or Π_2, but this is indeed the effect. Only Π_0 cells resulting from the use of a variable are further partitioned when the next variables are considered.

If we consider these two techniques applied to two unimodal distributions, then they are quite similar. Consider the Kendall method with no initial misclassifications allowed. The Henrichon-Fu algorithm with C_1 and C_2 very much larger than C_o would cause all cells with at least one observation differing from the rest to be labelled Π_o, producing the same effect. For any selections of allowed probabilities of misclassification with the Kendall method the same results could be produced with the Henrichon-Fu algorithm using suitably chosen values of C_o, C_1, C_2 and K.

The superiority of the Henrichon-Fu algorithm becomes evident when multimodal distributions are considered. Referring to Figure 3, it can be seen that Kendall's method would be useless but the Henrichon-Fu method would be appropriate. Another point in favor of the Henrichon-Fu algorithm is that it is readily usable with more than two populations, but the Kendall method in the present form is usable only with two populations.

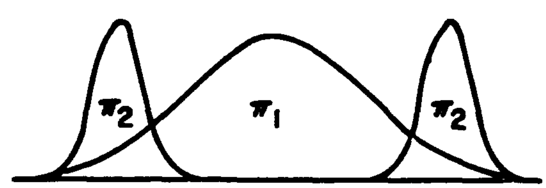

Figure 3. Hypothetical Case When Kendall's Method Would Not Perform As Well As the Henrichon-Fu Method

The Kendall method could still be better in the unimodal, two-population case, since a poor choice of K with the Henrichon-Fu algorithm could produce poor results. Of course the data could be examined prior to use of the Henrichon-Fu algorithm in order to select the best value of K. However, if there are many variates, each requiring possibly a different value of K, the task could become difficult. There is another important reason why the Kendall method would still be preferred. This is the fact that Kendall's method is easily used in routine discrimination, even without the use of a digital computer.

In summary, for discriminating between two unimodal distributions, the Kendall method may still be better, but for many other cases the Henrichon-Fu algorithm promises to be a superior method. Regardless of the advantages of the Henrichon-Fu algorithm in more sophisticated situations, the Kendall method still has the distinct advantage of being capable of being used without recourse to the digital computer.

Even when the Henrichon-Fu algorithm is applicable it would be interesting to consider the use of an initial Kendall method prior to application of the Henrichon-Fu technique.

Further study of the Henrichon-Fu algorithm and its possible modifications and a more complete comparison of the two algorithms are planned for future research.

6.3 Nearest Neighbor and Related Methods

Fix and Hodges (32) have considered a nearest neighbor method in which a point to be allocated is assigned to the class of the nearest classified point in the p-dimensional space. Variations of this method have considered allocating a point to the class of the majority of the nearest points. The latter method is referred to as the k-nearest neighbor decision rule.

Kendall (9) criticizes the nearest neighbor methods for generating impossibly complicated classification regions when there is considerable mixing of the samples from the different populations, and this is the situation of most importance.

Pelto (33) has developed a method which he calls adaptive nonparametric classification. This method estimates probability densities by counting known points observed within a hypersphere around the point to be classified. The radius of the hypersphere is fixed to minimize the expected loss of the decision rule.

CHAPTER VII

CONCLUSIONS

7.1 Summary of Results

1. Kendall's order-statistic method is a promising technique in discriminant analysis. In cases where the LDF, linear discriminant function, is appropriate, Kendall's method does not give much larger error rates than are obtained by use of the LDF. In cases of multivariate random variables with symmetric distributions such as the Cauchy or the uniform (both with independent random variables), or multivariate normal random variables with unequal variance-covariance matrices, Kendall's method gives lower error rates than those obtained by use of the LDF for populations with small mean differences. It gives comparable error rates to those obtained by use of the LDF for populations with larger mean differences. However, in most of these cases there is a relatively large portion of the index sample which will not be classified. The multivariate lognormal (independent variables) has been considered as a representative case of a multivariate random variable with a distribution which is not symmetric. In this case Kendall's method provides much lower error rates than are obtained by use of the LDF and most of the index sample is classified.

2. Use of an allowed probability of misclassification in the initial sample can greatly increase the portion of the index sample classified, while raising the error rates only slightly.

3. Maintaining the Mahalanobis distance but reducing the overlap of the component variates may sharply increase the error rate using Kendall's method.

4. Error rates for linear discriminant functions estimated from samples are compared with the theoretical error rates for a number of cases of multivariate normality with unequal variance-covariance matrices.

5. A modified Bartlett and Please method has been developed which provides equal probabilities of misclassification. This has been applied in a number of cases.

7.2 Future Work

Some interesting questions for future research would be

1. Extension of Kendall's method to more than two populations. If A, B, and C denote three populations, one possible approach would be to find which is most easily discriminated, A from B and C, B from A and C, or C from A and B. That separation is then carried out. Then discrimination could be tried between one of the two groups remaining and the other group combined with the residual group from the first separation.

2. Investigation of the effect of allowed probability of misclassification in Kendall's method. More work needs to be done on the trade-off between increasing the acceptable misclassification level and decreasing the portion of the sample unclassified.

3. Investigation of the effect of differences in sample sizes in Kendall's method. This could be an important factor, particularly if one sample size is much larger than the other.

4. Investigation of the effect of non-normal multivariate populations on Kendall's method. This work is important in order to develop a measure of the probabilities of misclassification to be expected.

5. Investigation of the effect of unequal mean components on Kendall's method, i.e., $\underline{\mu}' = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$, μ_i not necessarily equal to μ_j $(i \neq j)$. In the one example considered in the study, maintaining the Mahalanobis distance but changing the mean components strongly affected the result.

6. Investigation of the effect of different values of p other than 5. With smaller values of p discrimination will be reduced, but the question is how much? With larger values of p, Kendall's method, like linear discriminant analysis, may classify the initial sample only too well, greatly overestimating the probabilities of classification and underestimating the probabilities of misclassification.

7. Consideration of the effect of the a priori probability that a sample comes from a particular population. In this study equal a priori probabilities have been assumed.

8. Further investigation of the Henrichon and Fu algorithm and comparison with Kendall's method. As mentioned in the study, this may be an improvement over Kendall's method. Professor Fu kindly has supplied a copy of the latest version of his computer program.

9. Calculation of the probabilities of misclassification in using Kendall's methods by means of nonparametric statistics such as is done by Henrichon and Fu in their consideration of generalized tolerance limits.

10. Study of the adequacy of the representation of

$$\sigma^2[(1 - \rho)I + \rho Epp]$$

for variance-covariance matrices in general.

11. Improvement of the estimation technique in the modified Bartlett and Please method.

LIST OF REFERENCES

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

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

3. Lachenbruch, P. A., "Estimation of Error Rates in Discriminant Analysis," Technometrics, Vol. 16, 1968, pp. 1-11.

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

5. Hills, M., "Allocation Rules and Their Error Rates," (With Discussion), J. Royal Stat. Soc. London, Vol. B28, 1966, pp. 1-31.

6. Dunn, O. J., and Varady, P. D., "Probabilities of Correct Classification in Discriminant Analysis," Biometrics, Vol. 22, 1966, pp. 908-924.

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

8. Kendall, M. G., and Stuart, Alan, The Advanced Theory of Statistics, Hafner, New York, 1968, Vol. III, pp. 332-336.

9. Kendall, M. G., "Discrimination and Classification," Multivariate Analysis, Edited by P. R. Krishnaiah, Academic Press, New York, 1966, pp. 165-184.

10. Beall, G., "Approximate Methods in Calculating Discriminant Functions," Psychometrika, Vol. 10, 1945, pp. 205-217.

11. Jackson, R. W. B., "Approximate Multiple Regression Weights," J. Exp. Education, Vol. 11, 1943, pp. 221-225.

12. Penrose, L. S., "Some Notes on Discrimination," Ann. Eugenics, Vol. 13, 1946-47, pp. 228-237.

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

14. Geisser, S., and Desu, M. M., "Predictive Zero-Mean Uniform Discrimination," Biometrika, Vol. 55, 1968, pp. 519-524.

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

16. Han, Chien-Pai, "Distribution of Discriminant Function When Covariance Matrices are Proportional," Annals Math. Stat., Vol. 40, 1969, pp. 979-985.

17. Krumbein, W. C., and Graybill, F. A., An Introduction to Statistical Models in Geology, McGraw-Hill, New York, 1965, pp. 362-367.

18. Link, A. J., A Physico-Chemical and Textural Study of Carbonate Sedimentation in a Lagonal Environment, M. A. Thesis (Unpublished), 1964, Northwestern University.

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

20. Dempster, A. F., Elements of Continuous Multivariate Analysis, Addison-Wesley, Reading, Mass., 1969, pp. 213-219.

21. Tintner, G., "Some Applications of Multivariate Analysis to Economic Data," Jour. Amer. Stat. Assoc., Vol. 41, 1946, pp. 472-500.

22. Mosteller, F., and Tukey, J. W., "Data Analysis, Including Statistics," The Handbook of Social Psychology, Edited by Lindzey, G., and Aaronson, E., Addison-Wesley, Reading, Mass., 1968, Vol II, pp. 147-160.

23. Beerstecher et al, "Biochemical Individuality, V. Explorations with Respect to the Metabolic Patterns of Compulsive Drinkers," Arch. Biochem., Vol. 29, 1950, pp. 27-40.

24. Franklin, Joel N., "Numerical Simulation of Stationary and Non-Stationary Gaussian Random Processes," SIAM Review, Vol. 7, 1965, pp. 68-80.

25. Owen, D. B., Handbook of Statistical Tables, Addison-Wesley, Reading, Mass., 1962, pp. 151-154.

26. Gupta, Shanti, S., "Percentage Points and Modes of Order Statistics from the Normal Distribution," Ann. Math. Stat., Vol. 32, 1961, pp. 888-893.

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

28. Feldman, S., Klein, D. F., and Honigfeld, G., "A Comparison of Successive Screening and Discriminant Function Techniques in Medical Taxonomy," Biometrics, Vol. 25, 1969, pp. 725-734.

29. Course entitled "Computer-Oriented Techniques in Pattern Recognition," taught by Dr. W. S. Meisel, University of Southern California, July 20-24, 1970.

30. Henrichon, E. G., Jr., and Fu, King-Sun, "A Nonparametric Partitioning Procedure for Pattern Classification," IEEE Trans. on Computers, Vol. C-18, 1969, pp. 614-624.

31. Henrichon, E. G., Jr., "On Nonparametric Methods for Pattern Recognition," Unpublished Ph. D. Dissertation, 1970, Purdue University.

32. Fix, E., and Hodges, J. L., Jr., "Nonparametric Discrimination: Consistency Properties," USAF School of Aviation Medicine, Randolph Field, Texas, Proj. No. 21-49-004, Report 4, Feb. 1961.

33. Pelto, C. R., "Adaptive Nonparametric Classification," Technometrics, Vol. 11, 1969, pp. 775-792.