293 136 pg \$2.75 1886 298829 20882 16.00 HAMDBOOK OF A TECHNIQUES CLASSIFICATION STATISTICAL by CARL F. KOSSACK DDC RARINA 1111 0 1964 EQUELY EQ DECIRA B Reproduction in whole or in part is permitted for any purpose of the United States Government. FURDUE UNIVERSITY Dafayette, Indiana 1964

PREFACE .

This handbook has been prepared under ONR Contract No. 1745-61-8479 and represents the accumulation of several years of experience and effort in the Classification Field by the author and his graduate students. An attempt has been made in this handbook to make available to the general public the fruits of not only this research effort, but the research effort of the many statisticians who have made contributions in this general technical area. Although credit will be given in the form of references to many of the principal contributors to the theory, it is recognized that to give each individual his proper recognition is not possible if one is to keep the handbook of practical use. The author apologizes in advance to individuals who are passed over in this regard as well as to those individuals whose techniques or theory has been included in the handbook with notational changes and perhaps even technical deficiencies in the brief descriptions used. TABLE OF CONTENTS

Preface	Page 11
Table of Contents	444
Glossary of Terms and Notational Conventions	
Introduction	17
	vi
A Brief History of Classification Theory	1
Chapter II	
The General Characteristics of Classification Techniques	7
Chapter III	
The Illustrative Example	14
Chapter IV	
Ine Classification Techniques	17
A. Non-Parametric Classification	18
B. Categorical Classification	31
C. General Parametric Classification	40
D. The Wald Classification Statistic	56
E. The Purdue Classification Statistic	70
F. The Anderson Classification Statistic	84
G. The Shaw Classification Statistic	103
Chapter V.	
The Hulti-Population Problem	114
Chapter VI	
Bibliography on Classification and Related Topics	117

GLOSSARY OF TERMS

(Notation Conventions)

In classification theory, the problem of notation becomes acute if one is to consider more than one theory or technique, since in general there are several variables to distinguish between as well as several populations. In fact, one has not only the sample observations from these populations to consider, but one must also represent a set of individuals requiring classification into these populations. In general, the general indices will be denoted by lower case letters while the range of the variable will be denoted by the corresponding upper case letter. Greek letters will generally denote parameter or population variables while Latin letters will denote sample estimates^{*} of the parameters.

Indices

For variables, p or q. Thus x_p , p = 1,2, --- P For populations k or \mathbf{x} . Thus π_k , k = 1, 2, --- KFor individuals, n or m Thus $n = 1,2, ---, N_k$

We therefore have the representations for observations:

x^(k)_{pn} = The value of the pth variable for the nth individual of the kth population

and

Z_{pm} = the value of the pth variable for the mth individual to be classified.

* In this handbook, maximum likelihood estimates will generally be used.

For distribution characteristics we have:

The arithmetic mean

 $\begin{array}{l} \label{eq:linear} \ensuremath{\mathcal{L}}(k) = \ensuremath{\text{the population mean of } X_p \ensuremath{\text{for population } \pi_k} \\ & m^{(k)}_p = \ensuremath{\text{an estimate of the population mean of } X_p \ensuremath{\text{for population } \pi_k} \end{array}$

The variance - covariance

 $\int f(k) = for p = q$ the variance of X_p for population π_k

for $p \neq q$ the covariance between X_p and X_q for population π_k

S(k) = the corresponding sample estimates

The inverse of the covariance matrix, general term

 $\mathcal{O}_{(k)}^{pq}$ = the p,q term of the inverse of the covariance matrix of population π_k

 $g_{(k)}^{q}$ = the p,q term of the inverse of the covariance matrix of estimates for population π_{k}

A vector quantity, $\underline{X} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_Q)$

A matrix

 $s_{pq}^{(1)} = \begin{pmatrix} s_{11}^{(1)} & s_{12}^{(1)} & \dots & s_{1Q}^{(1)} \\ s_{21}^{(1)} & \dots & \dots \\ \vdots & & & \\ s_{P1}^{(1)} & s_{PQ}^{(1)} \end{pmatrix}$

V

INTRODUCTION

Classification techniques deal with the problem of assigning one or more individuals to one of several possible groups or populations on the basis of a set of measurements observed on them. The most common forms of classification are restricted to the case of only two groups or populations and thus this handbook will emphasize this restricted case. The basis of the actual classification is sets of the same observations made on groups of individuals (samples) whose population class is known from previous experience.

For illustration we will consider an example which will be used to numerically demonstrate each of the techniques to be considered in the handbook. The illustration deals with the problem of admission of freshman applicants to an engineering curriculum of a college. The measurements to be used in making the decision are: $\underline{X} = a$ mathematics orientation test score (X_1) , an English orientation test score (X_2) and a general aptitude test score (X_3) . We are required on the basis of these three measures to classify the applicant into population of students who will be successful in their engineering studies (Population π_1) or into the population of unsuccessful students (Population π_2). Available to us are the set of these three measures for a group or sample of students who have been previously conitted but proved to be unsuccessful. The classification problem is to evolve a decision rule which will mathematically use the three measures (X_1, X_2, X_3) and the past experience to make the classification.

This handbook has been prepared to assist individuals who are interested in utilizing one of the known classification techniques and thus emphasis will be placed upon the mechanics involved in determining the actual classification rule and on its subsequent use. Since the effectiveness of any technique depends upon how well the underlying conditions of the actual problem fit the assumptions made in developing the classification technique, attention will also be

vi

given to the assumptions involved in each technique considered. Each technique is considered in a separate section and each section has been prepared so as to be independent of other technique sections. Therefore, statistical clerks or computer programmers can be given the task of applying an assigned technique without regard to techniques considered in earlier sections.

The first chapter deals with a history of classification theory and also contains the more formal definition of the problem. In Chapter II, the general characteristics of a classification technique are considered along with different concepts found in statistical decision theory that may be used in evolving any of the techniques. Each technique that is to be considered is listed along with the underlying assumptions associated with the technique. This enumeration may be of assistance to an individual in selecting the appropriate technique to be used in his particular problem. In Chapter III, the data associated with the illustrative example (engineering college admission) which will be used in each technique is given.

Chapter IV consists of separate sections for each technique being considered, while Chapter V deals with the procedures that may be employed in the case of more than two populations. Chapter VI contains a bibliography for individuals interested in obtaining further background on any of the techniques considered in the handbook along with references covering more general problems of discrimination.

Attention is directed to the Glossary of Terms or Notational Conventions that is immediate following the Table of Contents. Although every attempt has been made to keep the mathematical level of this handbook of an elementary nature, it is necessary for the user of the handbook to have a minimum appreciation of matrix algebra and the elements of statistics.

vii

CHAPTER I

A BRIEF HISTORY OF CLASSIFICATION THEORY

In its most general form the theory of classification deals with the problem of assigning one or more individuals to one of several possible groups of populations on the basis of a set of measurable characteristics observed on It can be considered as a special case of a statistical decision problem. them. Given (K 2 2) populations and one or more individuals which are known to belong to one of the populations, the problem is to make decisions on the basis of a set of measurements on the individuals, as to which population each of them belongs. For example, in biometric investigations, one may want to assign a skull found in archeological excavations to some dynastic period on the basis of anthropometric measurements on it. A taxonomist may want to classify a plant specimen into one of two species on the basis of measurements on its stems and leaves. Hanufactured articles may be accepted or rejected on the basis of certain measurements made to determine whether or not they conform to specifi cations. Personnel may be assigned to duties on the basis of their scores in a battery of tests given to each employee. Prospective students may be admitted to a college or not based on their scores in the entrance examination. These and many more are essentially problems in classification and in a somewhat more general sense - problems in discrimination.

Let us first give a statistical formulation of the classification problem. Each of the K different populations is assumed to be characterized by a distribution function $F_k(X_1,\ldots,X_p; \theta_1^k,\ldots,\theta_V^k)$ or equivalently by a probability density function (p.d.f.) $f_k(X_1,\ldots,X_p; \theta_1^k,\ldots,\theta_V^k)$ where X_1,\ldots,X_p are random variables, $\theta_1^k,\ldots,\theta_V^k$ are certain parameters and $k = 1,2,\ldots,K$. The form of F_k or f_k is assumed to be known, but the paremeters themselves may or may not be known, but if unknown, they can always be estimated from K available samples, one each from the K populations. Given observations $\underline{Z}_{m} = (Z_{1m}, \dots, Z_{pm})$, $m = 1, 2 \dots M$ on M individuals each of which is known "a priori" to belong to one and the same population, the problem is to devise a statistical method for classifying observations into the population to which they belong in the "best" possible manner.

The first recorded attempt to solve the classification problem statistically was made by Karl Pearson in 1921. In a paper¹⁾ in <u>Biometrika</u> by Miss M. L. Tildesley, Pearson introduced a Coefficient of Racial Likeness to serve as a measure of the "distance" between π_1 and π_2 . Since then, this measure has been used by many anthropologists of the biometric school for purposes of classifying skeletal remains. The Coefficient of Racial Likeness (C.R.L.) is defined in terms of the sample means, the sample variances and the sample sizes of the various characters involved.

Let
$$\underline{x}_{n}^{(1)} = (\underline{x}_{n}^{(1)}, \dots, \underline{x}_{p_{n}}^{(1)})$$
 $n = 1, 2, \dots, \underline{N}_{1}$ and $\underline{x}_{n}^{(2)} = (\underline{x}_{n}^{(2)}, \dots, \underline{x}_{p_{n}}^{(2)})$

 $n = 1, 2, \dots, N_2$ be the samples from populations π_1 and π_2 , then

C.R.L. = 1/P
$$\frac{N_1 N_2}{N_1 + N_2}$$
 $\stackrel{P}{p=1}$ $\frac{(\bar{X}_p^{(1)} - \bar{X}_p^{(2)})^2}{s_{pp}}$

where

$$\ddot{x}_{p}^{(k)} = 1/N_{k}$$
 $\sum_{n=1}^{N_{k}}$ $x_{pn}^{(k)} p = 1, 2, ..., P; k = 1, 2;$

and s is an estimate of the <u>common</u> variance of the pth character in the two populations and is given by

(1.2)
$$pp = \frac{1}{N_1 + N_2} \begin{bmatrix} N_1 \\ \sum_{n=1}^{N_1} (x_{pn}^{(1)} - \overline{x}_{pn}^{(1)})^2 + \sum_{n=1}^{N_2} (x_{pn}^{(2)} - \overline{x}_{pn}^{(2)})^2 \end{bmatrix}$$

It is clear from the definition of C.R.L. that the populations are assumed to have the same set of variances for all the characters, and the characters themselves are treated as uncorrelated. The coefficient itself may be used to measure the probability that the two samples are from one and the same population or in other words whether or not π_1 and π_2 are the same. In this sense C.R.L. is used as a

-2-

^{.. 1)} Tildesley, M. L., "A First Study of the Burmese Skull," <u>Biometrika</u>, Vol. 13 (1921), pp. 175-252.

test divergence rather than as a measure of divergence between two populations. In this handbook, we shall not consider this coefficient, but it has been mentioned here since it plays an important historical role.

It is really important to distinguish between a "test" of divergence and a "measure" of divergence. One of the early statisticians to draw a clear cut distinction between the two is P. C. Mahalanobis¹⁾ of the Calcutta School of Statistics. He introduced the concept of a "measure" of divergence between two populations as early as in 1925 in a presidential address to the Anthropological Section of the Indian Science Congress. In 1928, in a theoretical paper presented to the Indian Science Congress. In 1928, in a theoretical form of the generalized distance function D². When it was first proposed, the D² statistic was meant to be a measure estimating the divergence or the "distance" between any two populations. By applying a "test" of divergence to two populations, we can conclude in a statistical sense whether or not the populations π_1 and π_2 are identical. When the populations are distinct, we can use the measure of uivergence to find the extent to which π_1 differs from π_2 and, in fact, the form of the statistic gives insight into how the two populations actually diverge.

The classical form of D^2 involved only the population means, variances and covariances and provided a measure of divergence between two populations which have been accepted as distinct. The exact distribution of the classical form of D^2 involving sample means and population variances and covariances

 Mahalanobis, P. C., "On Tests and Measures of Group Divergence, Part I: Theoretical Formulae". <u>Journal and Proceedings, Royal</u> <u>Asiatic Society of Bengal</u>, New Series, Vol. 26 (1930), pp. 541-538.

-3-

was first obtained by R. C. Bose¹⁾ in 1936. The studentised form of D² involving only sample readings was first defined by Mahalanobis²⁾ in a note in the Proceedings of the National Institute of Sciences of India. Its distribution under very general conditions has been obtained by R. C. Bose and S. N. Roy in a series of papers in Sankhya.

In 1936 R. A. Fisher initiated a new approach to the problem of discrimination and classification with the introduction of linear discriminant function analysis. This approach led to a new method of deriving test criteria suitable to multiple variate situations. When two or more populations have been measured in several characteristics (x_1, x_2, \dots, x_p) special interest attaches to certain linear functions of these measurements by which the populations may be best discriminated. Fisher and others have shown that a set of multiple measurements may be used to provide a discriminant function, linear in observation, having the property that, better than any other linear function, it will discriminate between any chosen normally distributed classes such as taxonomic species, the sexes and so on. The principle behind the choice of a "discriminant function" is merely to reduce multivariate problems to univariate problems, and this process has been found extremely useful in multivariate analysis. The problem is reduced to that of a single variable by choosing a linear compound of the original variables and constructing a statistic suitable for the univariate consideration of this problem. In principle the discriminant function need not be linear but can cover any class of functions. In practice a linear function is nearly always chosen so as to avoid complex distribution problems.

-4-

Bose, R. C., "On the Exact Distribution of D²-Statistic", <u>Sankhya</u>, Vol. 2, (1936), pp. 143-154.

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

The first published application of discriminant function appears to be the work of Mildred Barnard¹⁾ on craniometry following the suggestion of R. A. Fisher. In 1936, Fisher²⁾ gave a further example in the use of multiple measurements in a taxonomic problem and explained the theory underlying the construction of a linear discriminant function. In his paper ³⁾ entitled "The Statistical Utilization of Multiple Measurements", Fisher showed the relation between his work and that of Hotelling⁴⁾ and Mahalanobis.

A classification technique that is closely related to that of Mahalanobis' D^2 and Fisher's linear discriminant function is that obtained by the late Abraham Wald. In a paper⁵⁾ in the <u>Annals of Mathematical Statistics</u> (1944) Wald made an important contribution by introducing another approach to the classification problem. He considered the specific problem of classifying a single P-variate observation into one of two P-variate normal populations π_1 and π_2 , it being given that the observation belongs to π_1 or to π_2 . The classification problem is reduced to a problem in testing hypotheses; testing the hypothesis H₁: that the observation belongs to π_1 against

1) Barnard, M. M., "The Secular Variations of Skull Characters in Four Series of Egyptian Skulls", <u>Annals of Eugenics</u>, Vol. 6, (1935), pp.352-371.

2) Fisher, R. A., "The Use of Multiple Measurements in Taxonomic Problems", <u>Annals of Eugenics</u>, Vol. 7, (1936), pp. 179-188.

3) Fisher, R. A., "The Statistical Utilization of Multiple Measurements", Annals of Eugenics, Vol. 8, (1936), pp. 376-386.

4) Hotelling, Harold, "The Generalization of Students' Ratio", Annals of Mathematical Statistics, Vol. 2, (1931) pp. 360-378.

5) Wald, A., "On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups", <u>Annals of Mathematical Statistics</u>, Vol. 15, (1944), pp. 145-162.

-5-

the alternative hypothesis H₂: that the observation belongs to π_2 . The Fundamental Lemma due to Neyman and Pearson provides a classification statistic to classify the observation in π_1 or π_2 in a manner that is "best", where the "best" manner of classification corresponds to the most powerful test of H1 against H2. In the first instance, the population parameters are assumed to be known, so that the classification statistic, depending only on these parameters, is exactly known. In general the parameters are, of course, not known but they are estimated from the two available samples, one each from π_1 and π_2 . The estimates of the parameters appearing in the classification statistic are obtained from the samples and are substituted in the statistic itself, obtaining another classification statistic depending on the sample values. In the case when π_1 and π_2 have a common covariance matrix, Wald has derived an asymptotic distribution of the classification statistic and an approximate distribution of a modified form of the classification statistic. Much work has been done along similar lines by T. W. Anderson of Columbia University. In an article in <u>Pavchometrika</u> 1), Anderson proposes a classification statistic that differs only slightly from that of Wald.

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

-6-

CHAPTER II

THE GENERAL CHARACTERISTICS OF A CLASSIFICATION TECHNIQUE

As was mentioned earlier, the consideration of classification will be restricted to the two population problem. In considering the techniques or theories that have been developed to solve the two population classification problem, two approaches can be identified. The first approach is associated with the likelihood ratio concept found in statistics and is thus directed towards estimation of a "likelihood" that the individual came from each of the two populations. These likelihood estimates generally depend upon the observed measures on this individual as well as parameter estimates made from the available sample from each population.

Formally one obtains the two quantities $f_1(z)$ and $f_2(z)$ where there are likelihood estimates and then the ratio

$$L(\underline{z}) = f_2(\underline{z})$$
$$\overline{f_1(\underline{z})}$$

is computed. The classification rule is then simply

"If L (z) > λ , classify the individual into π_2

and if L (z) $\leq \lambda$, classify the individual into π_1 The techniques using this approach differ only in the means of making the estimates $f_k(z)$.

The second approach to the classification problem involves the defining of a classification statistic (often evolved through some likelihood ratio concept) where the statistic depends upon the observations z_1 , z_2 , -- z_p made on the individual to be classified and certain statistics computed from the two available samples. The classification rule thus depends on the value of this statistic obtained from the observations. Formally we thus have as a statistic

 $C(\underline{z})$ ~ which is a numerical function of the observation \underline{z} . The classification rule is then

"If C (z) belongs to the interval I_0 , classify z into population π_2 , otherwise classify z into π_1 "

-7-

The techniques using this approach differ only in the form of $C(\underline{s})$ reat is to be used. Methods are, of course, required for the estimation of the coefficients that appear in the classification statistic $C(\underline{s})$. These methods make use of the numerical characteristics of the two samples available from past experience. In many cases, the rule reduces to that of $C(z) \ge \lambda$ and thus is parallel to the likelihood ratio cases.

Since a classification rule involves the designation of either the constant λ that appears in the likelihood ratio approach or the determination of the interval I that appears in the classification statistic approach, it is necessary to consider the several types of strategies or approaches that one may use in making such a designation. There are several strategies available, nearly all of which can be applied to any of the techniques that are considered in the Handbook. However, before considering these approaches, it is necessary to discuss the concept of operational effectiveness of any classification technique. It should be readily apparent that in making a two population classification decision there are two types of classification errors that may be made. One may classify an individual into population $2(\pi_2)$ when he really belongs to population 1 (π_1) or conversely one may classify an individual into π1 when he really belongs to π_2 . The basic measures of the operational effectiveness of any classification technique when applied to some particular problem are:

p (2/1) = the probability of classifying an individual into
$$\pi_2$$

when he really belongs in π_1

and

p (1/2) = the probability of classifying an individual into π_1 when he really belongs in π_2 .

-8-

Since all methods for determination of the classification "constant," λ or I₀, depend in some extent upon these two effectiveness measures, it is necessary to have available some procedure for estimation of these two probabilities.

In certain cases the classification statistic approach, the conditional probability distribution of the statistic can be determined from theoretical considerations. In these cases, the estimation of the two effectiveness probabilities can be made by using this distribution function. However, in many cases the conditional probability distribution function is not known or is very difficult to estimate. In such cases, one can use the Method of Empirical Estimation to make the estimations. This method simply applies the classification rule to each individual of the two available samples and thus generates an emperical distribution of the ratio or statistic which can be used to estimate the two probabilities.^{*}

To illustrate the procedure in a general way, let us denote by $H(\underline{Z})$ either the likelihood ratio estimate $R(\underline{Z})$ or the classification statistic $C(\underline{Z})$ depending on which concept is being used. Then if we let \underline{Z} take on in turn the values $\underline{X}_{1}^{(1)}$, $\underline{X}_{2}^{(1)}$, --, $\underline{X}_{N_{1}}^{(1)}$ and then $\underline{X}_{1}^{(2)}$, $\underline{X}_{2}^{(2)}$, --- $\underline{X}_{N_{2}}^{(2)}$.

we will generate two conditional frequency distributions for H, one for π_1 and one for π_2 . If these conditional frequency distributions are made into cummulative distributions, histograms may then be plotted and estimates of the frequency function smoothed in perhaps by eye if necessary. Such a graphical approach is illustrated in Figures 1 and 2 below. In each classification this emperical technique will be illustrated along with any theoretical approach if available.

* If the sample sizes are large, recourse to a digital computer may be necessary.

-9-





Let us now consider the possible strategies that can be used in determining the classification constant.

a) Single Error Control

In many situations the strategy that is appropriate is that of controlling one of the error probabilities since this requirement dominates the classi fication problem. That is, we wish to use "the constant that gives an effectiveness such that

$$p(2/1) < \alpha_0$$
 (say)

In this situation one needs to use that H_0 in his classification rule such that

$$\int_{H} f(H / \pi_1) dH$$

where f (H) is the conditional probability distribution of $H(\underline{Z})$. In the empirical estimation method H_0 is read directly off the curve developed as in Figure 1. See illustration on Figure 1.

b) Minimum Average Cost

If there are available estimates of the costs involved when either type classification error is made and also estimates of the a priori probability that an individual belongs to each of the populations, one may elect to use the so-called Bayes decision procedure and seek the rule which minimizes the average cost involved in making classifications.

Let $C(2/1) = Cost^*$ of making error of classifying an individual into π_2 if he belongs in π_1 .

C (1/2) = Cost^{*} of making an error of classifying an individual as from π_1 if he belongs to π_2 .

 q_1 = a priori probability that individual comes from π_1 .

 q_2 = a priori probability that individual comes from π_2 .

* This may be simply a factor and not necessarily actual dollar values.

Then for any rule we have the average cost as

 $C = q_1 p(2/1) C(2/1) + q_2 p(1/2) C(1/2)$

and the problem is to find the rule that minimizes this C. Since a rule determination can be reduced finally to the determination of the classification constant, the problem is to determine the constant so as to minimize C. In the emperical estimation method, this determination can be made from the probability frequency curves obtained as shown in Figures 1 and 2 through the use of some method of successive approximation. That is to select arbitrarily an H_0 , determine the corresponding C, and then successive ly change the selection of H_0 until the minimum is essentially reached.

c) Minimax Principle

As the name implies, this strategy seeks to use that rule which has a minimum for the maximum cost that may be realized in a classification. It is often used when a priori probabilities are not available.

However, it is not applicable to the two population case and hence its consideration will be delayed until the chapter on more than two populations.

SUMMARY OF CLASSIFICATION TECHNIQUES

	Name	Type of		
	Non De la	variables	Distribution	Paremater
A.	Non-Parametric	Measured	None	None
в.	Categorical	Mixed	None	None
C.	General Parametric	Various	Known	None
D.	Wald	Measured	Normal	Equal Covariance
E.	Purdue	Measured	Normal	None
F.	Anderson	Measured	Normal	None
G.	Shaw	Measured	Normal	Equal Means

- 13 -

- 14 -<u>CHAPTER</u> III The Illustrative Example

In order to facilitate the description of the several standard methods of classification to be considered, a simplified example will be used. This example is restricted to the two population case and in fact is limited to the three variables $x_1 = math$ orientation test score, $x_2 = English$ orientation test score, and $x_3 = general$ aptitude test score. The illustration deals with the problem of admission of freshman applicants to an engineering curriculum in college with population π_1 being made up of those students who would be unsuccessful in their course work and population π_2 consisting of the successful students. To evolve the classification rule, a sample of nineteen unsuccessful students. Two individuals are considered for classification, one from each population, and for checking we know from which population the individual really came. The following data summarizes the available information to be used in the various classifications.

A. Misclassification Cost Factors:

 $C(1/2) = 2 \quad C(2/1) = 1$

where C(1/2) is the cost associated with classifying an individual who is from population 2 into population 1. That is, failing to admit a student who would do successful work.

And, C(2/1) is the sort associated with classifying an individual who is from population 1 into population 2. That is, admitting a student who is unsuccessful.

B. A Priori Probabilities

 $q_1 = .75$ the probability of an applicant being from population π_1 $q_2 = .25$ the probability of an applicant being from population π_2 C. The Samples from Previous Experience

Sample from Population π_1 (Unsuccessful Students)	Mathematics Orientation Test Score	English Orientation Test Score	General Aptitude Test Score
	×1	×2	×3
1	22	35	16
2	32	45	1.7
3	53	49	10
4	46	36	.10
5	38	54	14
6	39	53	13
7	21	2/.	13
8	64	50	16
9	59	J7 1.5	30
10	57	45	19
11	64	30	10
12	55	00	19
13	35	44	16
14	50	40	16
15	35	>>	19
• 16	41	48	20
17	30	3/	19
18	50	26	11
N, = 19	40	40	26
1	40	53	18

.cs English	
on Orientation	General Aptitude
e Test Score	Test Score
49	13
48	20
71	20
35	20
60	20
50	18
58	31
65	29
37	21
77	34
58	23
48	26
57	11
63	21
35	20
61	16
50	19
53	27
60	39
	60

D. The Individuals (Applicants) To Be Classified

 $z_1 : x_{11} = 34$ $x_{21} = 36$ $x_{31} = 12$

Note: This was an unsuccessful student chosen at random from population π_1

 $z_2: x_{12} = 55 x_{22} = 62 x_{32} = 21$

Note: This was a successful student chosen at random from population π_2

- 17 -

CHAPTER IV

The Classification Techniques

In the sections that follow, each illustrating one of the techniques of classification, the following basic steps will be considered:

- A) The selection of the variables to be used in the classification. In this step, not only will the selection of the original variables be considered, but also the transformations applicable to the variables if needed.
- B) The selection of the estimation procedure. This step depends upon the method being used either for the likelihood ratio or classification statistic <u>in the technique</u> and may involve certain selections as to the form of these functions.
- C) Determination of the Classification Rule. This involves the determination of the classification constant as discussed in Chapter II.
- D) The Measurement of Operational Effectiveness. The empirical approach will always be considered and where applicable theoretical approaches will also be included.
- E) Application of the Classification Rule to the Unknown Observation. The method of application for practical use of the rule will be considered. Each section will contain first a brief general description of the technique along with the main references for the technique. The mechanics of applying the technique will then be considered in which the five basic steps outlined above are discussed in a general sense will be outlined.

The technique will then be illustrated using the college admission data given in Chapter III. The five basic steps will again be considered, but now as developed for the illustrative example.

A. - Non Parametric Classification

<u>General Description</u>: The non-parametric classification procedure is limited to measured variables. If one associates with the dimensional sample space some metric or distance, say $D(\underline{X}, \underline{X}^1)$, then a neighborhood can be defined about the observed point \underline{Z} by specifying the number of sampled points from all samples that are required to be within the neighborhood, where the distance a sampled point is from Z is measured by $D(\underline{Z}, \underline{X})$. Within this neighborhood, a certain number of π_1 sampled points and a certain number of π_2 sampled points will lie. These numbers, divided by their respective sample totals, N₁ and N₂, yield estimates of $f_1(z)$ and $f_2(z)$ from which the classification decision is made.

References: "Discriminatory Analysis: Non-parametric Discrimination: Consistency Properties" Evelyn Fix and J. L. Hodges, Jr., University of California Report No. 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas. February, 1951

> "Univariate Two-Population Distribution - Free Discrimination" David S. Stoller, JASA, Vol. 49, No. 268, pp. 770-777, Dec., 1954.

- 18 -

This procedure is best indicated by a two dimensional graphical illustration.

×1

$$N_1 = 15$$
 $N_2 = 19$
 $M = 11$, $n_1 = 7$, $n_2 = 4$
 $\hat{f}_1(2) = \frac{n_1}{N_1} = \frac{7}{15} = .47$
 $\hat{f}_2(2) = \frac{n_2}{N_2} = \frac{4}{19} = .21$

$$R(Z) = \frac{\hat{f}_2(Z)}{\hat{f}_1(Z)} = \frac{.21}{.47} = .45$$

- 19 -

Procedure

- Step A. <u>Selection of the Variables To Be Used</u>: The Selection is restricted to measured variables, usually no transformation is performed.
- Step B. <u>The Selection of the Estimation Procedure for the Density</u> <u>Functions</u>

 B_1 : Selection of the metric or distance $D(x, x^1)$

 $D(\underline{x} , \underline{x}^{1}) \text{ must be such that}$ $D(\underline{x} , \underline{x}^{1}) > 0$ $D(\underline{x} , \underline{x}^{1}) = D(\underline{x}^{1} , \underline{x})$ $D(\underline{x} , \underline{x}^{1}) + D(\underline{x}^{1} , \underline{x}^{11}) \ge D(\underline{x} , \underline{x}^{11})$

B2: Decision as to the number of points, M, to be included in neighborhood of Z.

It is suggested that M be somewhat less than $N_1 + N_2$ if $\frac{1}{2}$

 N_1 and N_2 are small (≤ 30) while if N_1 and N_2 are large, M should be taken significantly less than $N_1 + N_2$ $-\frac{2}{2}$. The problem is to get enough points in the neighborhood so as to make the estimation of f_1 (z) and f_2 (z) sensitive. If M is too small or too large relative to $N_1 + N_2$, the sensitivity of the estimate would be adversely affected.

B₃: Computation of $D(\underline{Z}, \underline{X})$ for all x 's from the samples from π_1 and π_2

- B₄ Pool $D(\underline{Z}, \underline{X})$'s and order by increasing value. Select out first M. From these M¹s
- B5 Count the number of D(Z,x)'s from population π_1 , call this n_1 count the D(Z, x)'s from population π_2 , call this number n_2 .

$$B_6 \quad \text{Compute } f_1(\underline{Z}) = n_1/N_1$$

and $f_2(\underline{Z}) = n_2/N_2$

B₇ Estimate of $R(\underline{Z}) = f_2(\underline{Z}) / f_1(\underline{Z})$

C. The Determination of the Classification Rule:

C1: Determine the empirical estimation of the conditional distribution of R by evaluating R for each observation in the two samples. Each set of R's is then tabulated into a cummulative frequency distribution. Usually a free hand smoothing of the frequency distribution is sufficient for estimating p(i/j).

C2: Determination of λ (the classification rule) according to the decision strategy to be used.

C3: Statement of the Classification Rule: If R(Z) > A, classify Z as belonging to π₂, if R(Z) ≤ A classify Z as belonging to π₁
 D. Measure the operating effectiveness of the classification rule.

From the values of R obtained in Step C, and the application of the decision rule using λ obtained above, complete the table

From Population

To	π,	π1	π2
Population	To	ⁿ 11	ⁿ 12
	~2	ⁿ 21	ⁿ 22
Total	· .,	Nl	N ₂

From this table we make the estimates

- 1) $\hat{p} (2/1) = n_{21} / N_1$ 11) $\hat{p} (1/2) - n_{12} / N_2$
- E. The Application of the Classification Rule to the Observations Requiring Classification

For each observation \underline{Z} , follow the steps given in B to obtain the estimate of $R(\underline{Z})$. Use the rule

If $R(Z) > \lambda$ classify as π_2

Illustration of Procedure

Step A: Selection of Variables To Be Used

We will use from the illustrative example:

 X_1 = Math placement grade

X₂ = English placement grade

 X_3 = General aptitude test score

Step B: The Selection of the Estimation Procedures

 B_1 : Selection of the metric or distance D(x, x')

We will use

$$D(x, x^{t}) = \sum_{p=1}^{p} |x_{p} - x^{t}_{p}|$$

(Attention should perhaps be given to the introduction of weights w_p in such a metric.)

- 22 -

 B_2 : Decision as to number of points, M, to be included in neighborhood of Z.

We will use M = 15 since N₁ + N₂ = 38 To illustrate steps B₃ to B₇ consider X₁₁⁽¹⁾ = 22, X₂₁⁽¹⁾ = 35, X₃₁⁽¹⁾ = 15

B₃: Computation of the 38 D's from the given point to all the points in the sample $D_{1}^{(1)} = D(X_{1}^{(1)}, \frac{X_{1}^{(1)}}{2}) = 22 - 22 + 35 - 35 + 15 - 15 = 0$ $D_{2}^{(1)} = D(X_{2}^{(1)}, \frac{X_{1}^{(1)}}{2}) = 32 - 22 + 45 - 35 + 7 - 15 = 28$ $D_{3}^{(1)} = D(X_{3}^{(1)}, \frac{X_{1}^{(1)}}{3}) = 53 - 22 + 48 - 35 + 10 - 15 = 49$

Similarly through the remainder of the 19 observations for the sample of π_1 and the 19 observations for the sample of π_2 Obtaining the 38 distances, $\underline{X}_1^{(1)}$ is from the sample points. π_1 : 0, 28, 49, 26, 37, 37, 13, 81, 51, 41, 71, 43, 25, 52, 31, 25, 21, 38, 39 π_2 : 23, 55, 90, 35, 78, 54, 72, 88, 22, 107, 82, 61, 40, 63, 44, 54, 54, 66, 73

- B₄: Pool, order, and select the M = 15 smallest distances keeping track of population, obtaining
- π_1 : 0, 28, 26, 37, 37, 13, 25, 31, 25, 21, 38, 39 π_2 : 23, 35, 22
- B₅: Determine n_1 and n_2 by count of the results of step B_4 $n_1 = 12$ $n_2 = 3$

 B_6 Compute $f_1(\underline{x}) = n_1/N_1 = 12/19 = .6315$

 $f_{2}(\underline{x}) = n_{2}/N_{2} = 3/19 = .1578$ B7 Compute $R_{1}^{1} = f(\underline{x}) = \frac{.1578}{f_{1}(\underline{x})} = .25$

- 23 -

C: The Determination of the Classification Rule

C₁: Determine the empirical distribution of R(x) by evaluating R(x) for each observation in the two samples. One must repeat Steps B₃ through B₇ for each of the 38 points obtaining 38 observed values of R, 19 from π_1 and 19 from π_2 .

These values are to be tabulated into two cumulative frequency distributions given in Table 1 below:

			Tante T			
	Cumulative	Empirical	Frequency	Distributions	of R	
R	f ₁	"1	Cum p.	f2	π2	Cum p
.25	4		1.00			0.00
. 36	3		0.79	2		0.11
.50	2		0.63			0.11
.61			0.53	1		0.16
. 67	3		0.53	2		0.26
.80	1		0.37			0.26
.88	1		0.32			0.26
.96	2		0.26	1		0.32
1.00	1		0.16			0.32
1.14			0.11	1		0.37
1.50			0.11	4		0.58
2.00			0.11	1		0.63
2.75			0.11	2		0.74
4.00	2		0.11	3		0.89
6.50			0.00	2		1.00
				10		

Total 19

19

The cumulative frequencies are then plotted as a frequency distribution and smoothed into an estimate of the frequency curve. This is done by eye. See Figure 1 and 2.

 C_2 : Determination of λ according to the decision strategy to be used.

Single Error Control a) Consider the requirement that p(2/1) = .25. That is that we do not want to have more than 25% of our Freshman class fail.

Emperical Istimation of the Distribution of R given 77 1. Figure 1:

p(211)

1.00

2.00

3.00

1.00

.90

.80

.70

.60

.50

.40

.30

.20

.10





- 27 -

Then from Figure 1

$$\lambda = R \left\{ p(2/1) = .25 \right\} = 0.90$$

b) Minimum Expected Loss:

Since
$$C = q_1 C(2/1) p(2/1) + q_2 C(1/2) p(1/2)$$
 we have
 $C = (.75) (1) p(2/1) + (.25) (2) p(1/2)$
 $= .75 p(2/1) + .50 p(1/2)$

Take a first value of λ , say $\lambda = .90$. Then using figures 1 and 2, p (2/1) = .25 and p (1/2) = .28 and hence

$$C = .75 (.25) + (.50)(.28) = .33$$

Setting up the following computation table, we approach the minimizing λ value numerically.

X	p(2/1)	p(1/2)	С
.90	.25	.28	. 33
1.00	.16	. 32	.28
.75	. 37	. 20	. 38
1.25	.11	.46	.31
1.10	.135	. 39	. 30

And we select $\lambda = 1.00$ as an approximation to the point of minimum expected cost.

In the continuation of this illustration and in all further illustrations, we will use the strategy a) of controlling a single error. C3: Statement of the Classification Rule:

If R (Z) > 0.90 classify Z as belonging to π_2 . If R (Z) \leq 0.90 classify Z as belonging to π_1 .

D. Measurement of Operational Effectiveness.

Using the rule of C₃ and the data in Table 1 of Section C, we have that 5 observations from π_1 would have been classified as π_2 , and 5 observations of π_2 would have been called π_1 's. Thus we have





and thus

 \hat{p} (2/1) = 5/19 = .26 \hat{p} (1/2) = 5/19 = .26 E.

Application of the Classification Rule to the Unknown Observation.

For \underline{Z}_1 : $Z_1 = 34$, $Z_2 = 36$, $Z_3 = 12$ Compute $R(Z_1)$ following Steps B_3 to B_7 .

B₃: Compute 38 D's

 $D_1(Z) = |22 - 34| + |35 - 36| + |15 - 12| = 16$ $D_2(Z) = |32 - 34| + |45 - 36| + |17 - 12| = 16$

Continuing through all 38 sample points obtaining the distributions of D (Z_1)

π₁ : 16, 16, 33, 14, 33, 23, 29, 71, 41, 25, 61, 33, 15, 41, 21, 15, 15, 30, 29 : 19, 45, 80, 27, 66, 35, 62, 78, 12, 92, 72, 49, 24, 53, 36, 44, 44, 56, 103 π2

B4: Pool and Select the 15th Smallest D's π_1 : 14, 15, 15, 15, 16, 16, 21, 23, 23, 25, 29, 29 π₂ : 12, 19, 24

B5: Determine n1 and n2

 $n_1 = 12$, $n_2 = 3$

 B_6 : Compute $f_1(\underline{Z})$ and $f_2(\underline{Z})$

 $f_1(Z_1) = n_1/N_1 = 12/19 = 63$

 $f_2(z_1) = n_2/N_2 - 3/19 = .16$

 B_7 : Compute $R(Z_1)$

$$R(Z_1) = \frac{f_2(Z_1)}{f_1(Z_1)} = \frac{.16}{.53} = .25$$

Apply Classification Rule

Since R (\underline{Z}_1) = .25 < 0.90, Z_1 is classified (correctly) as a π_1
For \underline{Z}_2 , $\overline{Z}_{21} = .55$, $\overline{Z}_{22} = 62$, $\overline{Z}_{32} = 21$ Following the same steps used for \underline{Z}_1 , we have B₃: Compute 38 $D(\underline{Z}_2, \mathbf{x})$'s $D(\underline{Z}_2, \underline{\mathbf{x}}^{(1)}) = |22 - 55| + |35 - 62| + |15 - 21| = 66$ etc. obtaining π_1 : 66, 54, 27, 42, 33, 33, 77, 21, 23, 39, 13, 23, 41, 14, 35, 41, 71, 36, 27 π_2 : 47, 19, 24, 31, 16, 22, 14, 32, 44, 46, 24, 21, 34, 5, 36, 12, 16, 18, 29

- B4: Pool and select 15 smallest π_1 : 13, 14, 21, 23, 23 π_2 : 5, 12, 14, 16, 16, 18, 19, 21, 22, 24
- **B5:** Determine n_1 and n_2 $n_1 = 5$, $n_2 = 10$
- B₆: Compute $f_1(Z_2)$ and $f_2(Z_2)$ $f_1(Z_2) = {}^{n_1/N_1} = 5/19 = .263$ $f_2(Z_2) = {}^{n_2/N_2} = 10/19 = .526$
- B₇: Compute $R(\underline{Z}_2)$ $R(\underline{Z}_2) = f_2(\underline{Z}_2)$ $f_1(\underline{Z}_2) = \frac{.526}{.263} = 2.00$

Apply Classification Rule:

Since $R(Z_2) = 2.00 > 0.90$, Z_2 is classified (correctly) as π_2 .

- 31 -

B. Categorial Classification

<u>General Description</u>: The categorial classification procedure assumes that each variable to be used is either of a categorial type or has been converted over to a categorical type. For the observation (x_1, x_2, \ldots, x_p) the product categories are determined, say $(a_i, \beta_j, Y_k \ldots)$. Thus if the categories for x_1 are, say, Black and White and the categories for X_2 are High, Average, and Low, the possible product category classes for the observation (x_1, x_2) are (Black, High), (Black, Average), (Black, Low) (White, High), (White, Average), (White, Low). For each of these product categorical classes, the number of observations

from the sample from population π_1 and from the sample from population π_2 are counted. These counts are used to estimate $f_2(x)$ and $f_1(x)$ by using the ratio of the number observed in the class to the total number in the sample. One then associates with each categorical class the ratio $R = f_2(x)/f_1(x)$ and thus the classification rule can be applied to each class. To classify an individual, all one needs to do then is to determine what class he belongs to and to observe which population the classification rule has associated with the class.

References: William G. Cochran and Carl E. Hopkins: "Some Problems in Multivariate Classification with Qualitative Data". Harvard University, ONR Research Project, 1959.

Procedure

<u>Step A</u>: Selection of the Variables To Be Used: Variables are restricted to categorical type; if measured variables are to be used, they must first be converted to categorical.

A1. Conversion to Categorical. The conversion can be done by simply dividing the range for the given variable up into a member of intervals and designating each interval as a category. For example, if only two categories are desired, one can simply use the mean of the pooled samples as a division point. Appropriate division points for more categories are given in the following table.

Table 2.1

- 32 -

Division Points for Conversion of Measured Variables to Categorical \overline{X} = mean of pooled samples = $\sum_{n=1}^{N_1} x_n^{(1)} + \sum_{n=1}^{N_2} x_n^{(2)} / \frac{N_1 + N_2}{N_1 + N_2}$

$$\frac{s^{2} = \sum_{n=1}^{n_{1}} (x_{n}^{(1)} - \overline{x}^{(1)})^{2} + \sum_{n=1}^{n_{2}} (x_{n}^{(2)} - \overline{x}^{(2)})^{2}}{N_{1} + N_{2}}$$

Number of Categories	Category	Range
2	1	$X \leq \overline{X}$
	2	$\overline{\mathbf{x}} > \mathbf{x}$
3	1	$x \leq \overline{x} - 0.$ (s
	2	$x = 0.6s < x \le \overline{x} + 0.6s$
	3	$\overline{\mathbf{x}}$ + 0.65 < \mathbf{x}
4	1	x < x - s
	2	$\overline{\mathbf{X}} - \mathbf{S} < \mathbf{X} \leq \overline{\mathbf{X}}$
	3	$\overline{\mathbf{X}}$ < $\mathbf{X} \leq \overline{\mathbf{X}} + \mathbf{S}$
	4	$\overline{\mathbf{x}} + \mathbf{s} < \mathbf{x}$
5	1	$X \leq \overline{X} - 1.25$
	2	$\bar{x} - 1.2s < x \leq \bar{x} - 0.45$
	3	$\bar{x} = 0.45 < x < \bar{x} + 0.47$
	4	$\bar{x} + 0.48 < x < \bar{x} + 1.23$
	5	\bar{x} + 1.28 < x
6	1	$X \leq \overline{X} - 1.43$
	2	$\bar{x} = 1.48 < x \leq \bar{x} = 0.73$
	3	$\overline{\mathbf{X}}$ - 0.79 < $\mathbf{X} \leq \overline{\mathbf{X}}$
	4	$\overline{\mathbf{x}}$ < $\mathbf{x} \leq \overline{\mathbf{x}} + 0.75$
	5	$\bar{x} + 0.75 < \bar{x} \leq \bar{x} + 1.45$
	5	$\bar{x} + 1.4s < x$

A₂: Reduction in Number of Categories.

Since the number of categories used for any variables should be small, it may be necessary to reduce the original number by combining or grouping categories.

B. The Selection of the Estimation Procedure for the Density Functions

B₁. Identification of each product categorical class.

It is well to recognize that the available samples will be spread out over these classes and thus the difficulty associated with small or non-occurring frequency in classes must be faced at this step in the procedure.

B2. Tabulation of the Samples by Product Classes

Each population is kept separate, thus

Product Categorical Class

Population	(a_i, β_j, γ_k)	
π1	$n_1 (a_i, \beta_j, \gamma_k)$	
π2	$n_2 (a_j, \beta_j, \gamma_k)$	
	2 i j k	

B_j. Estimation of the Density Function for Each Product Class. $f_1(a_i, \beta_j, \gamma_k) = n_1(a_i, \beta_j, \gamma_k) / N_1$ $f_2(a_i, \beta_j, \gamma_k) = n_2(a_i, \beta_j, \gamma_k) / N_2$

B₄. Computation of the Likelihood Ratio R for Each Product Class R (α_i , β_j , γ_k) = $\frac{f_2(\alpha_i, \beta_j, \gamma_k)}{f_1(\alpha_i, \beta_j, \gamma_k)}$

Step C. The Determination of the Classification Rule

C1. Determine the empirical estimation of the conditional distributions of R using the R's generated in Step B_{ij} above. Each set of R's are tabulated into a cummulative frequency distribution. Usually a free hand smoothing of the frequency graph is sufficient to produce a curve for estimating the smooth curve of p(i/j).

- 33 -

- C₂: Determination of the λ (the classification constant) according to the decision strategy to be used.
- C3: Statement of the Classification Rule

The rule is then given by:

If R (Z) > λ , classify Z as belonging to π_2 ; if R (Z) $\leq \lambda$ classify Z as belonging to π_1 . This rule can be applied to the R's of Step B₄ to determine for each product class the population assignment.

D. Measurement of the Operational Effectiveness of the Classification Rule

Since the rule given in Step C₃ when applied to each product class $(a_i, \beta_j, \mathbf{y}_k)$ assigns all individuals of that classification into one or the other of the two populations, the frequencies of misclassification for the two samples can be read directly from the tabulations of B₂. That is, if class $(a_i, \beta_j, \mathbf{y}_k)$ is assigned by the rule to population π_2 , then the individuals making up the $n_j(a_i, \beta_j, \mathbf{y}_k)$ individuals in the class from the other population would have been erroneously classified. We can thus complete the table.



From this table we make the estimates

$$\hat{\mathbf{p}}$$
 (2/1) = $n_{21}/11_1$
 $\hat{\mathbf{p}}$ (1/2) = $n_{12}/12_2$

E. The Application of the Classification Rule

For the individual \underline{Z} to be classified, those variables that were converted to categorical (Step A₁) or for which the number of categories has been reduced (Step A₂) must be first considered. Then the product class for the individual is recognized and reference to the tabulation evolved in Step C₃ determines the classification.

Illustration of Procedure

Step A: Selection of Variables To Be Used.

We will use from the illustrative example

 X_1 = Nath placement grade

X₂ = English Placement Grade

X₃ = General Aptitude Test Score

A. Conversion to Categorical:

Since all three variables are measured, each must be converted to categorical. We elect to use simply two categories for each variable since the sample size is so small. The division point for each variable is then the pooled mean.

We compute

$$\bar{x}_{1} = \frac{\sum_{n} x_{1n}^{(1)} + \sum_{n} x_{2n}^{(2)}}{N_{1} + N_{2}} = \frac{327 + 1039}{38} = 49.1$$

$$\bar{x}_{2} = \frac{\sum_{n} x_{2n}^{(1)} + \sum_{n} x_{2n}^{(2)}}{N_{1} + N_{2}} = \frac{844 + 1035}{36} = 49.4$$

$$\bar{x}_{3} = \frac{\sum_{n} x_{3n}^{(1)} + \sum_{n} x_{2n}^{(2)}}{N_{1} + N_{2}} = \frac{311 + 446}{38} = 19.9$$

- 35 -

Let us denote the categories by L_p and H_p , p = 1, 2, 3. Thus if $X \le \overline{X}$ we have an L; otherwise an H category. This conversion when applied to the data given in Chapter III yields the following results:

	Po	pulation 1	L	•	F	Population 2	
Obs	x ₁	x ₂	x ₃	Obs	Х,	X	v
1	L1	La	La		1	2	3
2	L ₁	L2	La	2	L1	L ₂	L ₃
3	H ₁	La	L.	2	ⁿ 1	L ₂	н3
15	L	-2	-3	3	^H 1	H ₂	H ₃
5	-1	12	L3	L;	H ₁	L ₂	Ha
	-1	н2	L ₃	5	H ₁	H ₂	Ha
0	L ₁	H ₂	L ₃	6	L ₁	Ho	J.
7	L ₁	L ₂	L3	7	H1	Ho	На
3	н1	н2	H ₃	8	H1	Ho	
9	H1	L2	L3	9	L	7	"3
10	H1	L ₂	L	10		^L 2	Н3
11	H ₁	H ₂	La	10	ⁿ 1	H ₂	H ₃
12	Hı	La		11	H ₁	H ₂	H ₃
13	L	La	-3 L	12	H ₁	L ₂	H ₂
14	H1	Ha	~3	13	L ₁	H ₂	La
15	L	L	-3	14	H ₁	H ₂	-3 H-
16	1.	-2	"3	15	H ₁	L ₂	3 Ha
	-1	12	L ₃	16	L ₁	H ₂	Lo
17	L1	L ₂	L3	17	Ч,	Ho	-3
18	L ₁	L ₂	Ha	18	U.		^L 3
19	L	Ha	L		"1	H ₂	H ₃
		-2	-3	19	L ₁	H ₂	Ha

1.2. Reduction of Number of Categories.

This step is not used in the illustration since all the variables required conversion

- 36 -

- 37 -

B. Selection of the Estimation Procedure for the Density Functions

B1 Identification of the Product Classes

We have the following eight classes:

 $L_1L_2L_3$, $L_1L_2H_3$, $L_1H_2L_3$, $H_1L_2L_3$

 $L_1H_2H_3$, $H_1L_2H_3$, $H_1H_2L_3$, $H_1H_2H_3$

B₂ Tabulation of the Two Samples by Product Class

Product Class	Population 1 N1	Population 2 N ₂
L ₁ L ₂ L ₃	7	1
L ₁ L ₂ H ₃	2	1
L ₁ H ₂ L ₃	3	3
H ₁ L ₂ L ₃	۷,	0
L ₁ H ₂ H ₃	0	1
H ₁ L ₂ H ₃	0	4
H ₁ H ₂ L ₃	2	1
^H 1 ^H 2 ^H 3	1	3
	19	10

B3 Estimation of the Density Functions for Each Product Class

Product Class	Population 1	Population 2
L ₁ L ₂ L ₃	$f_1 = 7/19 = .368$	$f_2 = 1/19 = .053$
$L_1 L_2 H_3$	$f_1 = 2/19 = .105$	$f_2 = 1/19 = .053$
L ₁ H ₂ L ₃	$f_1 = 3/19 = .158$	$f_2 = 3/19 = .158$
H ₁ L ₂ L ₃	$f_1 = 4/19 = .211$	$f_2 = 0/19 = .000$
L ₁ H ₂ H ₃	$f_1 = 0/19 = .000$	$f_2 = 1/19 = .053$
H1L2H3	$f_1 = 0/19 = .000$	$f_2 = 4/19 = .211$
H1H2L3	$f_1 = 2/19 = .105$	$f_2 = 1/19 = .053$
H1H2H3	$f_1 = 1/19 = .053$	$f_2 = 3/19 = .421$

Computation of the Likelihood Ratio B4

Product Class	$R = f_2/f_1$
L ₁ L ₂ L ₃	.053/.368 = .144
L ₁ L ₂ H ₃	.053/.105 = .505
$L_1H_2L_3$.158/.158 = 1.000
H ₁ L ₂ L ₃	.000/.211 = .000
L ₁ H ₂ H ₃	.053/.000 = 00
H1L2H3	.211/.000 = 🏎
H1H2L3	.053/.105 = .505
^H 1 ^H 2 ^H 3	.421/.053 = 7.94

- 38 -

Step C: The Determination of the Classification Rule

 C_1 Order product classes by R value, and cummulate f_1 and f_2

Product Class	R	f1	p (2/1) Cum f ₁	f ₂	p (1/2) Cum f ₂
$H_1L_2H_3$	~	0.000	0.00	0.211	1.00
L ₁ H ₂ H ₃	-	0.000	0.00	0.053	0.79
H ₁ H ₂ H ₃	7.94	0.053	0.05	0.421	0.74
L ₁ H ₂ L ₃	1.00	0.158	0.21	0.158	0.32
H ₁ H ₂ L ₃	.505	0.105	0.32	0.053	0.16
L ₁ L ₂ H ₃	.505	0.105	0.43	0.053	0.11
L ₁ L ₂ L ₃	.144	0.363	0.79	0.053	0.05
H ₁ L ₂ L ₃	0	0.211	1.00	0.000	0.00

C₂ Use values of p (2/1) and p (1/2) of Step C₁ to determine λ value for chosen strategy.

To meet requirement that p (2/1) = .25 we see from the table given above that the four product classes $H_1L_2H_3$, $L_1H_2H_3$, $H_1H_2H_3$ and $L_1H_2I_3$ should be in π_2 for the Classification rule. The other four product classes are in π_1 .

C3 Statement of the Classification Rule:

If <u>Z</u> below to any of the classes $H_1H_2H_3$, $H_1L_2H_3$, $L_1H_2H_3$, and $L_1H_2L_3$ classify individual into π_2 . Otherwise, classify individual into π_1 .

D. Measurement of Operational Effectiveness

Using the rule of C_3 and the tabulation of B_2 , we have the result*

			From Populat	ion
		π1	π2	
ro Population	π1	15	3	
	π2	4	16	
Total		19	19	
Thus	3	p (2/1) = 4/	19 = .211	

E. The Application of the Classification Rule

Consider \underline{Z}_1 : $Z_{11} = 34$, $Z_{21} = 36$, $Z_{31} = 12$

 $\hat{p}(1/2) = 3/19 = .144$

Applying Step A1 to make categorical, we have

 \underline{Z}_1 : L_1, L_2, L_3

Referring to Classification Rule, \underline{Z}_1 is classified as π_1

Consider \underline{Z}_2 : $Z_{12} = 55$, $Z_{22} = 62$, $Z_{32} = 21$

*In cases of small frequencies such as in the illustration, this estimate is biased. One correction for this bias has been proposed by Bartlett: Bartlett, M.S., "The Use of Transformations," <u>Biometrics</u>, Vol. 3, No. 1 (1947). Applying Step A to make categorical, we have

 \underline{Z}_2 : H₁, H₂, H₃

Referring to Classification Rule, \underline{Z}_2 is classified as π_2 .

C. General Parametric Classification

General Description: A general procedure that can be followed in classification is available if one knows the parametric form for the density functions $f_1(x)$ and $f_2(x)$. The procedure is based upon the density ratio approach to classification. The best estimate for the parameters occurring in $f_1(x)$ and $f_2(x)$ are computed from the two available samples and the rule then involves

If R (z) =
$$f_2(z)$$

 $\frac{1}{f_1(z)} \ge \lambda$, classify z into π_2

The classification is accomplished for any \underline{z} by simply evaluating the density functions for \underline{z} using the best estimates of the parameters.

The multivariate normal density function is often used in the case of measured variables.

Reference: T. W. Anderson: "An Introduction to Multivariate Statistical Analysis" Wiley & Company, pp. 126-137.

Procedure

- Step A: Selection of the Variables To Be Used. The multivariate observation selected must be such that a joint density function can be exhibited.
- Step B: The Selection of the Estimation Procedure for the Density Function B₁ Selection of the Function To Be Used as the Density Function Methods for the selection of the density function either make use of previous experience as to the nature of the underlying multivariate determination or a study of the moments of the available samples.

Since the multivariate normal is so often encountered and, in fact, one often makes transformations on the original variables so as to achieve normally, we will restrict our consideration of estimation procedures to those associated with the multivariate normal distribution. The general form of the multi-variate normal can be written as:

$$f(x_1, x_2, \dots, x_p, \dots, x_p) = K \exp\left(-\frac{\sum_{p,q=1}^{P} (x_p - \mathcal{U}_p)(x_q - \mathcal{U}_p)}{2 | A_{pq} |}\right)$$

$$K = \frac{1}{(2\pi)^{p/2} + \Delta_{pq}} + \frac{1}{2}$$

and
$$\sigma_{11} \sigma_{12} \dots \sigma_{1P}$$

 (A_{pq})

and $\sigma_{p1} \dots \sigma_{pp}$

and $\sigma_{p1} = E(x_{p} - \mu_{p})(x_{p} - \mu_{p})$

where

the population covariance

and

$$C^{pq} = cofactor of (A_{pq})$$

= p,q th term of (A_{pq})⁻¹

In evaluating the multivariate normal density function, one may first perform a transformation on the original set of variables to a set of independently distributed variables. The following procedure accomplishes such a transformation.

Let x_1, x_2, \dots, x_p be the set of original variables. First transform to deviations from the mean, obtaining

$$t_p = x_p - \mu_p .$$

Then the transformation equation can be expressed as

and

$$z_p = u_p / \sqrt{\frac{(u_p \cdot u_p)}{N}}$$

where
$$(a_p \cdot b_q) = \sum_{n=1}^{N} a_{pn} \cdot b_{qn}$$

the common vector cross production notation.

If P is large, say greater than four, the computations given by the system of equation above for samples of any magnitude are such that they should be programmed for a digital computer. If P is less than or equal to four, the transformation can be expressed in terms of the moments of the sample and if used in this form, are more amenable to desk calculator computing. - 43 -

These equations are

W

$$\begin{split} \mathbf{z}_{1} &= \frac{\mathbf{x}_{1} - \overline{\mathbf{x}_{1}}}{\mathbf{s}_{1}} \\ \mathbf{z}_{2} &= \frac{\mathbf{x}_{2} - \overline{\mathbf{x}_{2}}}{\mathbf{s}_{2}} - \mathbf{r}_{12} \mathbf{z}_{1} \\ &= \frac{\mathbf{x}_{2} - \overline{\mathbf{x}_{2}}}{\sqrt{1 - \mathbf{r}_{12}^{2}}} \\ \mathbf{z}_{3} &= \frac{\mathbf{x}_{3} - \overline{\mathbf{x}_{3}}}{\mathbf{s}_{n}} - \frac{(\mathbf{r}_{23} - \mathbf{r}_{12} \mathbf{r}_{13})}{\sqrt{1 - \mathbf{r}_{12}^{2}}} \mathbf{z}_{2} - \mathbf{r}_{13} \mathbf{z}_{1} \\ &= \sqrt{1 - \mathbf{r}_{13}^{2}} - \frac{(\mathbf{r}_{23} - \mathbf{r}_{12} \mathbf{r}_{13})^{2}}{1 - \mathbf{r}_{12}^{2}} \\ \mathbf{z}_{4} &= \begin{cases} \mathbf{x}_{4} - \overline{\mathbf{x}_{4}} \\ \mathbf{x}_{5} \\ \mathbf{x}_{6} \\ \mathbf{x}_{7} \\ \mathbf{x}_$$

The steps that follow are those associated with the use of these latter formulas.

B2 Estimation of the means, standard deviation and correlation

coefficients for each population from the two available samples.

We prefer to use the maximum likelihood estimates.

$$\begin{split} \overline{\mathbf{x}}_{p}^{(k)} &= \sum_{n=1}^{N_{R}} x_{nn} / x_{R} \\ s_{p}^{(R)} &= \frac{1}{N_{R}} \sqrt{N_{R}} \sum_{L=1}^{N_{R}} x_{p1}^{2} - (\sum_{L=1}^{N_{R}} x_{p1})^{2} \\ r_{pq}^{(R)} &= \left[N_{R} \sum_{n=1}^{N_{R}} x_{ph} x_{qh} - \sum_{L=1}^{N_{R}} x_{ph} x_{qh}\right] \\ \sqrt{\left[N_{R} \sum_{n=1}^{N_{R}} x_{pn}^{2} - (\sum_{L=1}^{N_{R}} x_{ph})^{2}\right] \left[N_{R} \sum_{n=1}^{N_{R}} x_{qh}^{2} - (\sum_{n=1}^{N_{R}} x_{qn})^{2}\right]} \end{split}$$

B₃ Determination of the coefficients in the transformation equations using estimates of step B₂.

In applying the formula of step B_1 we will assume that each variate is first converted to a standarized variate by the transformation

$$t_p = \frac{X_p - \overline{X}_p}{S_p}$$

The formulae will then take these t's into the independent Z's.

B_L Determination of the individual densities for each X.

A table of ordinates for the univariate normal distribution in terms of standard variates can be used to determine the quantities.

$$f_{1}(z_{p}^{(1)}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{2}{p} \left(\frac{1}{2}\right)^{2}}$$
$$f_{2}(z_{p}^{(2)}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{2}{p} \left(\frac{2}{2}\right)^{2}}$$

B5 Evaluation of the multivariate density function.

We have the simple products

and

$$f_{1} (\mathbf{x}_{1}, \mathbf{x}_{2}, -\mathbf{x}_{p}) = f_{1} (Z_{1}^{(1)}) \cdot f_{1} (Z_{2}^{(1)}) \cdots f_{1} (Z_{p}^{(1)})$$

$$f_{2} (\mathbf{x}_{1}, \mathbf{x}_{2}, -\mathbf{x}_{p}) = f_{2} (Z_{1}^{(2)}) \cdot f_{2} (Z_{2}^{(2)}) \cdots f_{2} (Z_{p}^{(2)})$$

B5 Computation of the Likelihood Ratio:

$$R(\mathbf{x}) = \frac{f_2(x_1, x_2, -x_p)}{f_1(x_1, x_2, -x_p)}$$

Step C: The Determination of the Classification Rule.

- C_1 Determine the empirical estimation of the conditional distributions of R by using the R's generated through step B_5 above from each observation for each sample. These sets of R's are tabulated into a cumulative frequency distribution. Usually a free hand smoothing of the frequency graph is sufficient to produce a curve to be used in estimating values of p(i/j).
- C_2 Determination of the λ (the classification constant) according to the decision strategy being used.
- C3 Statement of the Classification Rule.

The rule then is given by

If $R(\underline{Z}) = \frac{f_2(Z_1, Z_2, --Z_p)}{f_1(Z_1, Z_2, --Z_p)} > \lambda$, classify \underline{Z} as belonging to π_2 , if $R(\underline{Z}) \leq \lambda$ classify \underline{Z} as belonging to π_1 .

Step D: Measurement of the Operational Effectiveness of the Classification Rule.

Using the R values determined for each sample observation in step C_1 , the classification rule of step C_3 can be applied yielding the totals in the following table.

- 46 -

From Population

		π1	"2
To Population	πı	ⁿ 11	ⁿ 12
to reputation	π ₂	ⁿ 21	n ₂₂
	Total	Nl	NZ

From this table we make the estimates

$$p(2/1) = n_{21} / N_1$$

 $p(1/2) = n_{12} / N_2$

Step E: The Application of the Classification Rule to the Observations Requiring Classification.

The computation steps given in B_4 , B_5 and B_6 are applied to each observation yielding a likelihood ratio R to which the classification rule of C_3 is applied.

Illustration of Procedure

Step A: Selection of the variables to be Used.

We will use from the illustrative example:

- X1 = Math Placement Grade
- X₂ = English Placement Grade
- X_3 = General Aptitude Test Score

Step B: The Selection of the Estimation Procedure for the Density Function.

B₁ Selection of the Function to be Used.

We will use the trivariate normal and will apply the second transformation approach using sample estimates of the population parameters. B2 Estimation of Means Standard Deviation and Correlation Coefficents

for Each Population from the Available Sample.

Computation of Sample Summations

Popul	Population 1			Popula	ati	on 2
Nl	=	19		N2	=	19
$\sum x_1$	11	827		Σx1	4000 0000	1039
$\sum x_2$	1	8444		Σ 1/2	1	1035
Σx_3	H	311		$\sum x_3$	11	445
$\sum x_1^2$	11	39033		$\sum x_1^2$	11	59547
$\sum x_2^2$	-000	39388		$\sum x_2^2$	-	58599
$\sum x_3^2$		5641		Σx_3^2	11	11408
$\sum_{x_1x_2}$	11	38073		$\Sigma_{x_1x_2}$	88	57661
$\sum x_1 x_3$	11	14007		$\sum_{x_1x_3}$	11	2 52 50
Ex2x3	1	14152		$\sum x_2 x_3$	=	2 5003

	Population 1					Por	pulation	2	
$\overline{x}_{1}^{(1)}$	$=\frac{827}{19}=43.53$,			x,	(2)	=	<u>1039</u> 19	8	54.68
ī ₂ (1)	$=\frac{844}{19}=44.42$,			x2	(2)		<u>1035</u> 19	11	54.47
ī, ⁽¹⁾	$=\frac{311}{19}=16.37$,			X	(2)	Altana Quadi	446	=	23.47
s1 ⁽¹⁾	= 1 19 1 57698	11	12.64.,	s1 ⁽²⁾	8	$\frac{1}{19}$	V 53772		= 12.20
s2 ⁽¹⁾	= 1 19 √ 36036		9.99 .	s ₂ ⁽²⁾	11	$\frac{1}{19}$	44055		= 11.05
s3 ⁽¹⁾	$=\frac{1}{19}\sqrt{10458}$	=	4.05	s ₃ (2)		1	V 17835	-	= 7.03

			- 48 -			
r ₁₂ ⁽¹⁾	=	+ 0.557		r ₁₂ (2)		+ 0.415
r ₁₃ ⁽¹⁾	=	+ 0.483		r ₁₃ (2)	=	+ 0.528
r ₂₃ ⁽¹⁾	11	+ 0.438		r ₂₃ ⁽²⁾	=	+ 0.480

B3 Determination of the Coefficients in the Transformation Equations.

Population 1

$$t_{1} = \frac{x_{1} - 45.53}{12.54}, \quad t_{2} = \frac{x_{2} - 44.42}{9.99}, \quad t_{3} = \frac{x_{3} - 16.37}{4.05}$$

$$z_{1} = t_{1}$$

$$z_{2} = \frac{t_{2} - 0.557}{0.831}$$

$$z_{3} = \frac{t_{3} - \frac{0.169}{0.831}}{\sqrt{0.757} - \frac{0.0285}{0.690}} = \frac{t_{3} - 0.203}{0.850}$$

$$\frac{Population 2}{0.850}$$

$$t_{1} = \frac{x_{1} - 54.68}{12.20}, \quad t_{2} = \frac{x_{2} - 54.47}{11.05}, \quad t_{3} = \frac{x_{3} - 23.47}{7.03}$$

$$z_{1} = t_{1}$$

$$z_{2} = \frac{t_{2} - 0.415}{0.910} \frac{z_{1}}{2} = 0.485 \frac{z_{1}}{10.05} = \frac{t_{3} - 0.203}{0.850} \frac{z_{2} - 0.483}{7.03}$$

 B_{μ} Determination of the individual densities for each Z_{p} .

B5 Evaluation of the multivariate density functions.

The following computational form* accomplishes the requirements of steps ${\rm B}_{\rm L}$ and ${\rm B}_{\rm S}.$

	Population 1		Population 2
Step (1)	Operation X1	Step (1)	Operation X ₁
(2)	x2	(5)	X ₂
(3)	x ₃	(3)	x _o
(4)	(1) - 43.53	(4)	(1) - 54.58
(5)	(2) - 44.42	(5)	(2) - 54.47
(5)	(3) - 16.37	(5)	(3) - 23.47
(7)	(4) / 12.64	(7)	(4) / 12.20
(8)	(5) / 9.99	(8)	(5) / 11.05
(9)	(6) / 4.05	(9)	(5) / 7.03
(10)	(7)	(10)	(7)
(11)	(8) - 0.557(7)	(11)	(8) = 0.415(7)
(12)	(11) / 0.831	(12)	(11)/0.910
(13)	(9) = 0.203 (12) = 0.483 (10)	(13)	(9) = 0.282 (12) = 0.522 (10)
(14)	(13) / 0.850	(14)	(13) / 0.799
(15)	f ₁ (10)	(15)	f ₁ (10)
(15)	f ₂ (12)	(16)	$f_{2}(12)$
(17)	f ₃ (14)	(17)	f_{2} (14)
(18)	(15)(16)(17)	(18)	(15) (15) (17)

B₅ Compulation of the Likelihood Ratio R.

R is the quotient of step 18 for Population 2 by Population 1. *The convention used is that $(\tau -)$ means to use the value of step r. f(2) is the normal ordinate from table look up. For example: Table II, page 315: Hoel, <u>Introduction to Mathematical Statistics</u>, Wiley & Co.

- 49 -

The computation is illustrated for a typical observation: $x_1 = 34$. $x_2 = 36, x_3 = 12.$

Step	Population 1	Population 2
(1)	34	34
(2)	35	35
(3)	12	12
(4)	-9.53	-20.58
(5)	-8.42	-18.47
(6)	-4.37	-11.47
(7)	-0.7.54	1.695
(8)	-0.843	-1.571
(9)	-1.079	-1.532
(10)	-0.754	-1.595
(11)	-0.423	-0.968
(12)	-0.509	-1.054
(13)	-0.611	-0.432
(14)	-0.719	-0.541
(15)	0.3011	0.0941
(16)	0.3503	0.2275
(17)	0.3079	0.3448
(18)	0.0325	0.0074

 $R(34, 36, 12) = \frac{0.0074}{0.0325} = 0.23$

The Determination of the Classification Rule. Step C:

Cl Determination of the empirical estimation of the conditional distributions of R.

Applying the computational form given above to the 38 observations in the two samples yields the following values of R.

Table 1

Empirica. Values of R

Populatio	on 1	Population 2		
Observation 1	R 0.094	Observation 1	R 0.470	
2	1.378	2	1.465	
3	1.576	3	160.213	
4	0.445	4	1.357	
5	0.941	5	100.550	
6	0.859	6	0.865	
7	0.055	7	515.877	
8	179.553	8	57.455	
9	1.208	9	0.790	
10	1.824	10	4192.712	
11	1.535	11	3.493	
12	0.750	12	14.931	
13	0.410	13	1.814	
14	1.352	14	3.762	
15	0.979	15	2.360	
16	0.557	16	1.606	
17	0.134	17	1.170	
18	12.666	18	25.727	
19	0.985	19	3 x 10 ⁶	

Order the observed R's* and compute cumulative percentage frequency.

*If a larger sample size were available, a frequency distribution would be more appropriate for this step.

Table 2

Cumulative Frequencies for R

Population 1		Population 2		
Ordered R	cum f _l	Ordered R	cum f2	
.055	1.00	.470	.05	
.094	0.95	.790	.11	
:134	0.89	.855	.16	
.410	0.84	1.170	.21	
.445	0.79	1.357	.25	
.557	0.74	1.465	. 32	
.750	0.58	1.505	.37	
.859	0.63	1.814	.42	
.941	0.58	2.360	.47	
.979	0.53	3.493	.53	
.985	0.47	3.762	. 58	
1.208	0.42	14.931	.63	
1.352	0.37	25.727	.68	
1.378	0.32	57.456	.74	
1.536	0.25	100.650	.79	
1.575	0.21	160.213	.84	
1.824	0.16	515.877	.89	
12.655	0.11	4192.712	.95	
179.553	0.05	3 x 10 ⁶	1.00	

These values are plotted in Figure 1 and 2 and a smoothed cumulative curve fitted free hand to the distribution to yield an empirical estimate of p(2/1) and p(1/2).







 C_2 : Determination of the value of λ (the classification constant) using decision strategy.

Since we want to control classification so that p(2/1) = .25, we use Figure 1 and have

 $\lambda = R(p(2/1) = .25) = 1.50$

C3: Statement of the Classification Rule.

olassify If $R(Z_1, Z_2, Z_3) \ge 1.50 / \underline{Z}$ as belonging to π_2 , if $R(Z_1, Z_2, Z_3) \le 1.50$ classify \underline{Z} as belong to π_1 .

Step D: Measurement of Operational Effectiveness.

Table 1 in Step C_1 enables one to readily apply the classification rule to the 38 sample observation yielding the results

From Pupulation

π	^π 2
14	6
5	13
19	19
	π ₁ 14 5 19

We obtain,

p(2/1) = 5/19 = .253p(1/2) = 6/19 = .316

Step E: The Application of the Classification Rule to the Observations Requiring Classification.

The computation form in step B5 is applied to each observation.

		-)• -		
Step	<u>Z</u>]	L	<u>Z</u> 2	
	Population 1	2	1	2
(1)	34	34	55	55 -
(2)	35	35	62	62
(3)	12	12	21	21
(4)	-9.53	-20.58	+11.47	+0.32
(5)	-8.42	-18.47	+17.58	+7.53
(6)	-4.37	-11.47	+ 4.63	-2.47
(7)	-0.754	- 1.695	+ 0.907	+0.025
(8)	-0.843	-1.671	+ 1.750	+0.581
(9)	-1.079	-1.632	+1.143	-0.351
(10)	-0.754	-1.695	+0.907	+0.025
(11)	-0.423	-0.968	+1.255	+0.570
(12)	-0.509	-1.054	+1.510	+0.735
(13)	-0.611	-0.432	+0.398	-0.576
(14)	-0.719	-0.541	+0.458	-0.721
(15)	0.3011	0.0941	C.2637	0.3988
(16)	0.3503	0.2275	0.1276	0.3034
(17)	0.3079	0.3448	0.3572	0.3079
(18)	0.0325	0.0074	0.0120	0.037

56

 $R_{(Z)}: .0074/0.0325 = 0.23$ 0.0373/0.0120 = 3.31

Since $R(\underline{Z}_1) < 1.50$, \underline{Z}_1 is classified as belonging to π_1 . Since $R(\underline{Z}_2) > 1.50$, \underline{Z}_2 is classified as belonging to π_2 .

D. The Wald Classification Statistic.

General Description:

The Wald Classification technique consists of using the multivariate normal assumption for the density function for each population. By making the further assumption that the two populations have the same covariance matrix, the likelihood ratio can be reduced to a simplified classification statistic. The statistic takes the form

$$W(\underline{Z}) = \sum_{q=1}^{P} \sum_{p=1}^{P} S^{pq} (\overline{\mathbf{x}}_{q}^{(2)} - \overline{\mathbf{x}}_{q}^{(1)}) Z_{p}$$

where S^{PQ} is obtained from the pooled estimates of the covariance.

The problem of determining the value of λ to use in the formulation of the classification rule can be considered theoretically since the statistic $W(\underline{2})$ is approximately normally distributed with known mean and standard deviation.

References: "On the Statistical Problem Arising in the Classification of an Individual into One of Two Groups," Abraham Wald, <u>Annals of Mathematics</u> <u>Statistics</u>, Vol. XV, no. 2, June, 1944.

Procedure

Step A: Selection of the Variables to be Used.

The variables used must be measured and have a joint distribution that can be approximated by the normal distribution with equal covariance matrices for the two populations. If this latter condition is not met, one often performs a transformation upon the original variables in order to bring about a closer approximation by the normal assumption.

Step B: The Selection of the Estimation Procedure.

Since the classification statistic involves sample estimates of population parameters, we elect to use the corresponding maximum likelihood estimates as follows:

B1: The sample moments computation.

$$\begin{aligned} \overline{\mathbf{x}}_{q}^{(1)} &= \sum_{n=1}^{N_{1}} x_{qn}^{(1)} / N_{1} \\ \overline{\mathbf{x}}_{q}^{(2)} &= \sum_{n=1}^{N_{2}} x_{qn}^{(2)} / N_{1} \\ s_{pq}^{(1)} &= \frac{N_{1}}{\sum_{n=1}^{N_{1}} x_{pn} x_{qn} - \sum_{n=1}^{N_{1}} x_{pn} \sum_{n} x_{qn}}{N_{1}^{2}} \\ s_{pq}^{(2)} &= \frac{N_{2}}{\sum_{n=1}^{N_{2}} x_{pn} x_{qn} - \sum_{n=1}^{N_{2}} x_{pn} \sum_{n=1}^{N_{2}} x_{qn}}{N_{2}^{2}} \end{aligned}$$

$$s_{pq} = \frac{N_1 s_{pq} + N_2 s_{pq}}{N_1 + N_2}$$

B2: Inversion of the Covariance Matrix.

$$\begin{pmatrix} s^{11} & s^{12} & \cdots & s^{1P} \\ s^{21} & & & \\ \vdots & & & \\ s^{P1} & & s^{PP} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1P} \\ s_{21} & & & \\ \vdots & & & \\ s_{P1} & & s_{PP} \end{pmatrix}^{-1}$$

B₃: Estimation of the Coefficients of the Wald Statistic. In general we have

$$W(\mathbf{z}) = \sum_{p=1}^{P} \left(\sum_{q=1}^{P} s^{pq} (\overline{x}_{q}^{(2)} - \overline{x}_{q}^{(1)}) \right) z_{p}$$

- 58 -

For P = 3 this can be expanded to read

$$W(\underline{Z}) = [s^{11} (\overline{x}_{1}^{(2)} - \overline{x}_{1}^{(1)}) + s^{12} (\overline{x}_{2}^{(2)} - \overline{x}_{2}^{(1)}) + s^{13} (\overline{x}_{3}^{(2)} - \overline{x}_{3}^{(1)})] z_{1} + [s^{21} (\overline{x}_{1}^{(2)} - \overline{x}_{1}^{(1)}) + s^{22} (\overline{x}_{2}^{(2)} - \overline{x}_{2}^{(1)}) + s^{23} (\overline{x}_{3}^{(2)} - \overline{x}_{3}^{(1)})] z_{2} + [s^{31} (\overline{x}_{1}^{(2)} - \overline{x}_{1}^{(1)}) + s^{32} (\overline{x}_{2}^{(2)} - \overline{x}_{2}^{(1)}) + s^{33} (\overline{x}_{3}^{(2)} - \overline{x}_{3}^{(1)})] z_{3}$$

Step C: The Determination of the Classification Rule.

Empirical Estimation:

- C1: Determine the empirical distribution of $W(\underline{Z})$ by evaluation $W(\underline{Z})$ for each observation in the two samples. These sets of W's are tabulated into a cumulative frequency distribution. Usually a free hand smoothing of the frequency graph is sufficient to produce a curve to be used in estimating values of p(i/j).
- C_2 : Determination of the λ (the classification constant) according to the decision strategy being used.
- C3: Statement of the Classification Rule:

The rule is then given by

If $W(\underline{Z}) > \lambda$, classify \underline{Z} as belonging to π_2 , if $W(\underline{Z}) \leq \lambda$, classify \underline{Z} as belonging to π_1 .

Theoretical Estimation:

The moments of W(Z) are:

The means
$$\alpha_1 = \sum_{p=1}^{P} \sum_{q=1}^{P} s^{pq} (\bar{x}_q^{(2)} - \bar{x}_q^{(1)}) \bar{x}_p^{(1)}$$

if \underline{Z} belongs to π_1

if <u>Z</u> belongs to π_2 and following the assumption, for either π_1 c: π_2

 $a_2 = \sum_{p=1}^{P} \sum_{q=1}^{P} s^{pq} (\bar{x}_q^{(2)} - \bar{x}_q^{(1)}) \bar{z}_p^{(2)}$

- 59 -

the variance of $W(\underline{Z})$ is

$$\sigma_{W}^{2_{2}} = \sum_{q=1}^{P} \sum_{p=1}^{P} s^{pq} (\bar{x}_{p}^{(2)} - \bar{x}_{p}^{(1)}) (\bar{x}_{q}^{(2)} - \bar{x}_{q}^{(1)})$$

We thus have for any λ (Classification Rule)

$$p_{\lambda}(2/1) = \frac{1}{\sqrt{2\pi}} \int_{\lambda - \alpha_{1}}^{\infty} e^{-t^{2}/2} dt$$
$$\frac{\lambda - \alpha_{1}}{\sigma_{W}}$$
$$p_{\lambda}(1/2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^{2}/2} dt$$

and

From these estimates of the probabilities of the two types of classification error one can determine the classification rule that corresponds to his decision strategy.

Step D: Measurement of the Operational Effectiveness of the Classification Rule. Empirical Estimation

Using the values of the classification statistic $W(\underline{Z})$ that were generated in step C_1 for the sample observations, one can apply the classification rule of step C_3 to each observation and obtain the totals for the following table.



From this table we can make the estimates

 $p(2/1) = n_{21}/N_1$ $p(1/2) = n_{12}/N_2$

Theoretical Estimation

If one elects to use the approximate distribution of the statistic $W(\underline{Z})$, we have the estimates of the probabilities of error from the integrals given in step C.

Step E: The Application of the Classification Rule to the Observations Requiring Classification.

The value of the statistic W(Z) as given in step B_3 is determined for the observation. Since the statistic is linear in \underline{Z} this evaluation is straight forward. Using this value one uses the classification rule of step C_3 to determine the appropriate classification for the individual.

Illustration of the Procedure

Step A: Selection of the Variables to be Used.

We will use from the illustrative example

 $x_1 = Math Placement Grade$

 x_2 = English Placement Grade

x₃ = General Aptitude Test Score

Step B: The Selection of the Estimation Procedure.

B1: The Sample Moment Compulations

Summations

Population 1

$$N_{1} = 19$$

$$\sum x_{1} = 827$$

$$\sum x_{2} = 844$$

$$\sum x_{3} = 311$$

$$\sum x_{1}^{2} = 39033$$

$$\sum x_{2}^{2} = 39388$$

$$\sum x_{3}^{2} = 5641$$

$$\sum x_{1}x_{2} = 38073$$

$$\sum x_{1}x_{3} = 14007$$

$$\sum x_{2}x_{3} = 14152$$

Means

Population 1

Population 2				
. N ₁	11	19		
$\sum x_1$	=	1039		
$\sum x_2$	=	1035		
$\sum x_3$		446		
$\sum_{x_1^2} x_1^2$	11	59547		
$\sum x_2^2$	=	58699		
$\sum x_3^2$		11408		
<u> </u>	=	57661		
$\sum x_1 x_3$	11	2 52 50		
L x2x3	н	25003		

Population 2

$\bar{x}_{1}^{(1)} =$	<u>827</u> 19		43.53	$\frac{x^{1}}{(5)} =$	<u>1039</u> 19	=	54.68
$\bar{x}_{2}^{(1)} =$	<u>844</u> 19	11	44.42	$\frac{1}{x_2}$ =	<u>1035</u> 19	-	54.47
$\bar{x}_{3}^{(1)} =$	<u>311</u> 19	н	16.37	$\frac{1}{x_3}$ =	<u>446</u> 19	11	23.47

Pooled Variances and Covariances

$$S_{11} = \frac{(19)(57698)}{(19)^2} + \frac{(19)(53772)}{(19)^2} = 154.39$$

$$S_{22} = \frac{(19)(36036)}{(19)^2} + \frac{(19)(44056)}{(19)^2} = 110.93$$

$$s_{33} = \frac{(19)(10458)}{(19)^2} + \frac{(19)(17836)}{(19)^2} = 39.19$$

$$19 + 19$$

$$s_{12} = \frac{(19)(25399)}{(19)^2} + \frac{(19)(20194)}{(19)^2} = 63.15$$

$$19 + 19$$

$$S_{13} = \frac{(19)(8936)}{(19)^2} + \frac{(19)(16356)}{(19)^2} = 35.03$$

19 + 19

1202/2000/2

$$S_{23} = \frac{(19)(6404)}{(19)^2} + \frac{(19)(13447)}{(19)^2} = 27.49$$

B2: Inversion of the Covariance Matrix*

Evaluation of Covariance Determinant

154.3963.1535.0363.15110.9327.49=35.0327.4939.19

Determination of Cofactor Matrix

3592	-1512	-2150
-1512	4823	-2032
-2150	-2032	13139

where 3592 = (110.93)(39.19) - (27.49)(27.49)

*Since the rank of the matrix is three the inversion will be made using the definition.

- 63 -

Division of terms by covariance determinant

$$(s^{pq}) = \begin{pmatrix} .00936 & -.00394 & -.00560 \\ -.00394 & .01257 & -.00530 \\ -.00560 & -.00530 & .03424 \end{pmatrix}$$

B3: Estimation of the Coefficient of the Wald Statistic.

Computation of Mean Differences

$$\begin{array}{c} x_{1} \\ x_{1} \\ x_{2} \\ x_{2} \\ x_{3} \\ x_{3} \\ x_{3} \end{array}^{(1)} = +11.15 \\ = +11.1$$

Determination of Coefficients

$$b_{1} = .00936(+11.15) - .00394(10.05) - .00560(7.10)$$

= +.0250
$$b_{2} = -.00394(11.15) + .01257(10.05) - .00530(7.10)$$

= +.0448
$$b_{3} = -.00560(11.15) - .00530(10.05) + .03424(7.10)$$

= .1274

Thus

 $W(\underline{Z}) = +.0250 Z_1 + .0448 Z_2 + .1274 Z_3$

Step C: The Determination of the Classification Rule.

Empirical Estimation

 C_1 : Determination of the Empirical Distributions of W(Z).

Evaluate W(Z) for each available sample observation. Thus

 $W(\underline{x}_{1}^{(1)}) = +.0250(25) +.0448(35) +.1274(15) = 4.029$ Repeating for the 38 observations yields the results.

Table: 3 W(2) For Sample Observations

- 65 -

Population 1		Population 2		
Dos.	W(Z)	Obs.	W(Z)	
1	4.029	1	4.575	
2	3,708	2	6.173	
3	4.749	3	8.425	
LL LL	4.546	4	5.415	
5	5.025	5	7.908	
6	5.006	5	5.733	
7	3.539	7	7.923	
8	8.065	8	8.507	
9	5.912	9	5.233	
10	4.312	10	9.606	
11	6.709	11	7.354	
12	5.385	12	5,888	
13	4.974	13	4.855	
14	5.135	14	5.773	
15	5.573	15	5.541	
16	5.103	16	5.995	
17	3.316	17	6.085	
18	6.254	18	7.264	
10	5,558	19	8.807	

Generation of the Cumulative Frequency Distributions for W

	Popula	tion 1	
Interval	f	cum f	cum p
3.00 - 3.49	1	19	1.00
3.50 - 3.99	2	18	.95
4.00 - 4.49	2	16	. 34
4.50 - 4.99	3	14	.74
5.00 - 5.49	4	11	. 58
5.50 - 5.99	3	7	.37
6.00 - 5.99	3	4	.21
7.00 - 7.99	Ó	1	.05
8 00 - 8 99	1	1	.05

	Popula	tion 2	
Interval	f	cum f	cum p
4.50 - 5.49	4	4	.21
5.50 - 5.49	5	9	.47
6.50 - 5.99	2	11	. 58
7.00 - 7.49	2	13	.68
7.50 - 7.99	2	15	.79
8.00 - 8.49	1	15	.84
8.50 - 8.99	2	18	.95
9.00 - 9.49	0	19	.95
9.50 - 9.99	1	19	1.00

1.00

P

.75

.50

.25






These distributions are plotted as in Figure 1 and 2 and a smooth cumulative curve filled free hand to the distribution to yield an empirical estimation of p(2/1) and p(1/2).

 C_2 : Determination of λ using the decision strategy.

Since we require a classification rule such that p(2/1) = .25, we use Figure 1 and have

 $\lambda = W(p(2/1) = .25) = 5.90$

C3: Statement of the Classification Rule.

If $W(Z_1, Z_2, Z_3) > 5.90$ classify \underline{Z} as belonging to π_2 , if $W(Z_1, Z_2, Z_3) \leq 5.90$ classify \underline{Z} as belonging to π_1 .

Theoretical Estimation

 C_1 : Determination of the Moments of W(Z)

$$a_{1} = .0250(43.53) + .0448(44.42) + .1274(16.37)$$

= 5.15
$$a_{2} = .0250(55.58) + .0448(54.47) + .1274(23.47)$$

= 6.80
$$\sigma^{2} = .0250(11.15) + .0448(10.05) + .1274(7.10)$$

= 1.63

 C_2 : Determination of λ using the decision strategy.

Since we require a classification rule such that p(2/1) = .25.

we have

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{\lambda - 5.16}{1.53}}^{\infty} e^{-t^2/2} dt = 0.25$$

From the tabulation of areas under the normal curve we have

$$\frac{\lambda - 5.15}{1.63} = 0.67$$

and $\lambda \cdot = 6.02$ (as compared to 5.90 determined from the empirical distribution).

Step D: Measurement of Operational Effectiveness.

Using Table 1 of step C_1 and the classification rule of C_3 , one can readily classify each of the sample observations and obtain the totals for the following table.

rrom ropulat	ion	
--------------	-----	--

		π	π2
То	πı	14	5
Population	π2	5	13
	Total	19	19

We obtain

$$p(2/1) = 5/19 = .26$$

 $p(1/2) = 6/19 = .32$

If we utilized the theoretical distribution, we would have obtained

$$p(2/1) = \sqrt{\frac{1}{\sqrt{2\pi}}} \int_{\frac{6}{1.28}}^{100} e^{-t^2/2} dt = 0.25 \text{ (as designed)}$$

and

$$p(1/2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{6.02 - 6.80}{1.28}} e^{-t^2/2} dt = .27$$

Step E: The Application of the Classification Rule to the Observations Requiring Classification.

We have

 $W(Z) = .0250 Z_1 + .0448 Z_2 + .1274 Z_3$

For \underline{Z}_1 : $\underline{Z}_1 = 34$, $\underline{Z}_2 = 35$, $\underline{Z}_3 = 12$

 $W(\underline{Z}_1) = .0250(34) + .0448(35) + .1274(12) = 3.992$

Applying the Classification Rule.

Since $W(\underline{Z}_1) \leq 5.90$, \underline{Z}_1 is classified as belonging to π_1 . For \underline{Z}_2 : $Z_1 = 55$, $Z_2 = 62$, $Z_3 = 21$

 $W(\underline{2}_{2}) = .0250(55) + .0448(52) + .1274(21) = 5.828$

Since $W(\underline{Z}_2) > 5.90$, \underline{Z}_2 is classified as belonging to π_2 .

E. The Purdue Classification Statistic

The Purdue University classification technique using the multivariate normal assumption for the density function of each population. Each population is considered to have its own covariance matrix and the likelihood ratio involved reduced to form the classification statistic. The statistic takes the form of a quadratic function of the 2's.

$$P(\underline{Z}) = \sum_{q=1}^{P} \sum_{p=1}^{P} [(s_{(1)}^{pq} - s_{(2)}^{pq}) Z_{p} Z_{q} - 2(s_{(1)}^{pq} \overline{x}_{p}^{(1)} - s_{(2)}^{pq} \overline{x}_{p}^{(2)}) Z_{q} + (s_{(1)}^{pq} \overline{x}_{p}^{(1)} \overline{x}_{q}^{(1)} - s_{(2)}^{pq} \overline{x}_{q}^{(2)} \overline{x}_{q}^{(2)})]$$

The distribution of $P(\underline{Z})$ can be approximated by an incomplete Gamma distribution. Since this distribution has been tabulated, this fact could be used to estimate p(2/1) and p(1/2), however, the approach is beyond the scope of this handbook.

Reference: "Multivariate Classification with Normal Alternatives," R. H. Shaw, Purdue University Thesis, Lafayette, Indiana, July, 1959. Step A: Selection of the Variables to be Used.

The variables to be used must be measured and have a joint distribution that can be approximated by the multivariate normal distribution. If this condition is not met, one often performs transformations upon the original variables in order to bring about a closer approximation by the multivariate normal distribution function.

Step B: The Selection of the Estimation Procedure.

Since the classification statistic involves sample estimates of population parameters, we elect to use the corresponding maximum likelihood estimates in evaluating the statistic.

B1: The sample moment computation



$$s_{pq}^{(2)} = \frac{N_2 \sum_{n=1}^{2} x_{pn}^{(2)} x_{qn}^{(2)} - \sum_{n=1}^{2} x_{pn}^{(2)} \sum_{n=1}^{2} x_{qn}^{(2)}}{N_2^2}$$

B₂: Inversion of the covariance matrix.

- 71 -

$$\begin{pmatrix} s_{(1)}^{11} & s_{(1)}^{12} & s_{(1)}^{13} \\ s_{(1)}^{21} & s_{(1)}^{22} & s_{(1)}^{23} \\ s_{(1)}^{31} & s_{(1)}^{32} & s_{(1)}^{33} \end{pmatrix} = \begin{pmatrix} s_{11}^{(1)} & s_{12}^{(1)} & s_{13}^{(1)} \\ s_{21}^{(1)} & s_{22}^{(1)} & s_{23}^{(1)} \\ s_{31}^{(1)} & s_{32}^{(1)} & s_{33}^{(1)} \end{pmatrix}^{-1}$$

$$\begin{pmatrix} s_{(2)}^{11} & s_{(2)}^{12} & s_{(2)}^{13} \\ s_{(2)}^{21} & s_{(2)}^{22} & s_{(2)}^{23} \\ s_{(2)}^{31} & s_{(2)}^{32} & s_{(2)}^{33} \end{pmatrix} = \begin{pmatrix} s_{11}^{(2)} & s_{12}^{(2)} & s_{33}^{(1)} \\ s_{21}^{(2)} & s_{22}^{(2)} & s_{33}^{(2)} \\ s_{31}^{(2)} & s_{32}^{(2)} & s_{33}^{(2)} \end{pmatrix}^{-1}$$

B3: Estimation of the Coefficients of the Purdue Statistic.

In general we have

$$P(\underline{Z}) = \sum_{q=1}^{P} \sum_{p=1}^{P} (s_{(1)}^{pq} - s_{(2)}^{pq}) z_{p} z_{q}$$

$$2 \sum_{q=1}^{P} \sum_{p=1}^{P} (s_{(1)}^{pq} \overline{x}_{p}^{(1)} - s_{(2)}^{pq} \overline{x}_{p}^{(2)}) z_{q}$$

$$+ \sum_{q=1}^{P} \sum_{p=1}^{P} (s_{(1)}^{pq} \overline{x}_{p}^{(1)} - s_{(2)}^{pq} \overline{x}_{p}^{(2)}) z_{q}$$

For P = 3, this expands to the form

$$P(\underline{Z}) = (s_{(1)}^{11} - s_{(2)}^{11}) z_{1}^{2} + (s_{(1)}^{22} - s_{(2)}^{22}) z_{2}^{2} + (s_{(1)}^{33} - s_{(2)}^{33}) z_{3}^{2}$$

$$+ 2 (s_{(1)}^{12} - s_{(2)}^{12}) z_{1}z_{2} + 2 (s_{(1)}^{13} - s_{(2)}^{13})z_{1}z_{3} + 2(s_{(1)}^{23} - s_{(2)}^{23})z_{2}z_{3}$$

$$- 2[(s_{(1)}^{11} \overline{x}_{1}^{(1)} - s_{(2)}^{11} \overline{x}_{1}^{(2)}] + (s_{(1)}^{12} \overline{x}_{2}^{(1)} - s_{(2)}^{12} \overline{x}_{2}^{(2)})$$

$$+ (s_{(1)}^{13} \overline{x}_{3}^{(1)} - s_{(2)}^{13} \overline{x}_{3}^{(2)})] z_{1}$$

$$-73 - 2[(s_{(1)}^{12} \bar{x}_{1}^{(1)} - s_{(2)}^{12} \bar{x}_{1}^{(2)}) + (s_{(1)}^{22} \bar{x}_{2}^{(1)} - s_{(2)}^{22} \bar{x}_{2}^{(2)}) + (s_{(1)}^{23} \bar{x}_{3}^{(1)} - s_{(2)}^{23} \bar{x}_{3}^{(2)})] z_{2} - 2[(s_{(1)}^{13} \bar{x}_{1}^{(1)} - s_{(2)}^{13} \bar{y}_{1}^{(2)}) + (s_{(1)}^{23} \bar{x}_{2}^{(1)} - s_{(2)}^{23} \bar{x}_{2}^{(2)}) + (s_{(1)}^{33} \bar{x}_{3}^{(1)} - s_{(2)}^{23} \bar{x}_{2}^{(2)}) + (s_{(1)}^{33} \bar{x}_{3}^{(1)} - s_{(2)}^{23} \bar{x}_{3}^{(2)})] z_{3} + (s_{(1)}^{11} \bar{x}_{1}^{(1)^{2}} - s_{(2)}^{11} \bar{x}_{1}^{(2)^{2}}) + (s_{(1)}^{22} \bar{x}_{2}^{(1)^{2}} - s_{(2)}^{22} x_{(2)}^{22}) + (s_{(1)}^{33} \bar{x}_{3}^{(1)^{2}} - s_{(2)}^{33} \bar{x}_{3}^{(2)^{2}}) + (s_{(1)}^{33} \bar{x}_{3}^{(1)^{2}} - s_{(2)}^{33} \bar{x}_{3}^{(2)^{2}}) + 2(s_{(1)}^{12} \bar{x}_{1}^{(1)} \bar{x}_{3}^{(1)} - s_{(2)}^{13} \bar{x}_{1}^{(2)} \bar{x}_{3}^{(2)}) + 2(s_{(1)}^{23} \bar{x}_{2}^{(1)} \bar{x}_{3}^{(1)} - s_{(2)}^{23} \bar{x}_{2}^{(2)} \bar{x}_{3}^{(2)})$$

Step C: The Determination of the Classification Rule.

- C₁: Determine the empirical distribution of $P(\underline{Z})$ by evaluating $P(\underline{Z})$ for each observation in the two samples. These sets of P's are tabulated into a cumulative frequency distribution. Usually a free hand smoothing of the frequency graph is sufficient to produce a curve to be used in estimating values of p(i/j).
- C2: Determination of the λ (the classification constant) according to the decision strategy to be used.
- C₃: Statement of the Classification Rule. The rule is then given by ¹If $P(\underline{Z}) > \lambda$ classify \underline{Z} as belonging to π_2 , if $P(\underline{Z}) \leq \lambda$, classify \underline{Z} as belonging to π_1 ."

Step D: Measurement of the Operational Effectiveness of the Classification Rule.

Using the values of the classification statistic $P(\underline{Z})$ that were generated in step C_1 for the sample observations, one can apply the classification rule of C3 to each observation and obtain the totals for the following table.

		From Population		
-		π1	π2	
То	πı	nıı	n12	
Population	π2	ⁿ 21	n ₂₂	
	Total	Nl	N2	

From this table we can make the estimates

$$p(2/1) = {n \over 21/N_1}$$

 $p(1/2) = {n \over 12/N_2}$

Step E: The Application of the Classification Rule to the Observations Requiring Classification,

The value of the statistic $P(\underline{Z})$ as given in step B_3 is determined for the observation. Using this value one applies the classification rule of C_3 to determine the appropriate classification for the individual.

Illustration of the Procedure

Step A: Selection of the Variables to be Used.

We will use from the illustrative example

 $X_1 = Math Placement Grade$

 X_2 = English Placement Grade

 $X_3 =$ General Aptitude Test Score

Step B: The Selection of the Estimation Procedure

B1: The sample moment competation.

- 74 -

The sample summations

	Population 1					Popu	lat	lon 2	
	Nı		19				N ₂	н	19
Σ	xl	=	827			Σ	. x ₁	4000	1039
Σ	X2	=	844			Σ	X2	н	1035
Σ	x ₃	11	311			Σ	x ₃	11	446
Σ	x1 ²	1	39033			Σ	x1 ²	=	59647
Σ	x2 ²	=	39388			Σ	x22	11	58699
Σ	x3 ²	=	5541			Σ	x3 ²	=	11408
Σ	x1x2	н	38073			Σ	x ₁ x ₂	-	57661
Σ	x ₁ x ₃	-	14007			Σ	x1x3	=	2 52 50
Σ	X2X3	Contra Contra	14152			Σ	X2X3	18	2 5003

Estimation of Population Parameters

	Pop	ulation	1			P	opulat	10n	2
x ₁ ⁽¹⁾	11	$\frac{827}{19} =$	43	• 53	x1 ⁽²⁾	=	<u>1039</u> 19		54.68
x2 ⁽¹⁾	H	<u>844</u> 19 =	44	.42	x2 ⁽²⁾	=	1035 19	=	54.47
x ₃ ⁽¹⁾	=	<u>311</u> 19 =	15	.37	Ī3 ⁽²⁾		<u>446</u> 19	=	23.47
s ₁₁ ⁽¹⁾	-	<u>57698</u> 361	-	159.83	s (2) 11	=	53772 351	-	148.95
s ₂₂ ⁽¹⁾	N	<u>35036</u> 361	-	99.82	s ₂₂ ⁽²⁾	=	<u>44056</u> 361	=	122.04
s ₃₃ ⁽¹⁾	N	<u>10458</u> 361	14	28.97	s (2) 33	=	17835 361		49.41
s ₁₂ ⁽¹⁾		<u>35399</u> 361	83	70.36	s (2) 12	=	20194 361	-	55.94
s (1) 13	11	<u>8936</u> 351		24.75	s (2) 13	=	16355 361	- =	45.31

- 75 -

$$s_{23}^{(1)} = \frac{5404}{351} = 17.74$$
 $s_{23}^{(2)} = \frac{13447}{351} = 37.25$

- 76 -

B2: Inversion* of the Covariance Matrices.

Evaluation of the Covariance Determinant

Population 1

1	59.83	70.35	24.75		
1	70.35	99.82	17.74	Netto Ratio	259117
2	24.75	17.74	28.97		

Population 2

	45.31	55.94	148.95
= 47515	37.25	122.04	55.94
	49.41	37.25	45.31

Computation of the Cofactor Matrices

Population 1

1	2577	-1599	-1222
(-1599	4018	-1094
(-1222	-1094	11004

where 2577 = (122.04)(49.41) - (37.25)(37.25), etc.

Population 2

1	45.42	-1075	-3446
[-10.75	5307	-3014
	-34.45	-3014	15049

*Since the order of the matrix is three the inversion is made by use of the definition.

Division of terms by covariance determinant

Population 1

$$\begin{pmatrix} s_{(1)}^{11} & s_{(1)}^{12} & s_{(1)}^{13} \\ s_{(1)}^{21} & s_{(1)}^{22} & s_{(1)}^{23} \\ s_{(1)}^{31} & s_{(1)}^{32} & s_{(1)}^{33} \end{pmatrix} = \begin{pmatrix} +.00958 & -.00594 & -.00454 \\ -.00594 & .01493 & -.00407 \\ -.00454 & -.00407 & .04089 \end{pmatrix}$$

Population 2

$$\begin{pmatrix} s_{(2)}^{11} & s_{(2)}^{12} & s_{(2)}^{13} \\ s_{(2)}^{21} & s_{(2)}^{22} & s_{(2)}^{23} \\ s_{(2)}^{31} & s_{(2)}^{32} & s_{(2)}^{33} \end{pmatrix} = \begin{pmatrix} +.00977 & -.00226 & -.00725 \\ -.00226 & .01117 & -.00634 \\ -.00725 & -.00634 & .03167 \end{pmatrix}$$

B3: Estimation of the Coefficients.

Computation of matrix of differences, the coefficients of the quadratic terms.

$$(s_{(1)}^{pq} - s_{(2)}^{pq}) = \begin{pmatrix} .00958 & -.00594 & -.00454 \\ .00594 & .01493 & -.00407 \\ -.00454 & -.00407 & .04089 \end{pmatrix} - \begin{pmatrix} .00977 & -.00226 & -.00725 \\ -.00226 & .01117 & -.00634 \\ -.00725 & -.00534 & .03167 \end{pmatrix}$$
$$= \begin{pmatrix} -.00019 & -.00368 & .00271 \\ -.00368 & .00376 & .00227 \\ +.00271 & .00227 & .00922 \end{pmatrix}$$

Computation of linear term coefficients

$$b_{1} = -2[(.00958)(43.53) - (.00977)(54.68) + (-.00594)(44.42) - (-.00226)(54.47) + (-.00454)(15.37) - (-.00725)(23.47)] = -2(-.1621)$$

similarly for b2 and b3.

Computation of the constant term

$$\mathbf{k} = \left[(.00958)(43.53)^2 - (.00977)(54.68)^2 \right] + \dots + \left[(-.00407)(54.47)(23.47) - (-.00534)(45.31)(37.25) \right]$$

= - 8.333

Thus

Thus

$$P(\underline{Z}) = -.00019 \ z_1^2 + .00375 \ z_2^2 + .00922 \ z_3^2$$

$$-.00735 \ z_1^2 \ z_2 + .00542 \ z_1^2 \ z_3 + .00454 \ z_2^2 \ z_3$$

$$+.3242 \ z_1 - .0038 \ z_2 - .5788 \ z_3 - 8.333$$

Step C: Determination of the Classification Rule.

C₁: Determination of the values of the classification statistic for the sample observation from π_1 and π_2 .

$$P(\underline{X}_{1}^{(1)}) = -.00019(25)^{2} + .00376(35)^{2} + .00922(15)^{2}$$

-.00736(25)(35) + .00542(25)(15) + .00454(35)(15)
+ .3242(25) - .0038(35) - .5788(15) - 8.333 = - 4.922

Repeating for the 38 observations yields the results

-					3
ro	DU.	$\mathbf{1a}$	11	lor	1 1

Population 2

Obs. l	P(<u>Z</u>) -4.922	Cos. l	P(<u>Z</u>) -1.739
2	-2.254	2	0.537
3	-1.742	3	7.385
4	-1.791	4	0.015
5	-0.733	5	5.553
5	-0.921	5	-0.325
7	-5.579	7	5.195
8	5.087	8	5.952
9	0.324	9	-2.168
10	-0.980	10	11.258
11	0.781	11	2.309
12	-0.711	12	2.755
13	-1.811	13	-0.259
14	0.522	14	2.456
15	-0.839	15	1.398
16	-1.773	16	0.537
17	_4.240	17	0.277
18	0.888	18	3.645
19	-0.109	19	10.970

Generation of the Cumulative Frequency Distribution for P

Population 1

Inte	rval	f	cum f	CIIM D
-6.99	-5.00	1	19) 00
-5.99	-5.00	0	18	1.00
-4.99	-4.00	2	18	• 7)
-3.99	-3.00	0	16	• 7)
-2.99	-2.00	1	15	.04
-1.99	-1.00	4	15	.04
-0.99	0.00	5	11	• 17
+0.01	-1.00	24	5	•)0
1.01	-2.00	0	ĩ	. 2)
2.01	3.00	0	1	.05
3.91	4.00	0	1	.05
4.01	5.00	0	1	.05
5.01	5.00	0	1	.05
5.01	7.00	1	1	.05

Population 2

Interval		ſ	cum f	Cilm D
-2.99	-2.00	1	1	05
-1.99	-1.00	1	2	
-0.99	-0.10	2	4	21
+0.01	1.00	4	8	.42
1.01	2.00	1	9	.47
2.01	3.00	3	12	.53
3.01	4.00	ĩ	13	.68
4.01	5.00	0	13	.58
5.01	5.00	2	15	.79
5.01	. 7.00	1	15	.84
7.01	8.00	1	17	.89
8.01	9.00	0	17	.89
9.01	10.00	0	17	.89
10.01	11.00	1	18	.95
11.01	12.00	1	19	1.00

These distributions are plotted in Figures 1 and 2 and a smooth cumulative curve fitted free hand to the distribution to yield an empirical estimated p(2/1) and p(1/2).

C2: Determination of λ Using the Decision Strategy.

Since we require a classification rule such that p(2/1) = 0.25, we use Figure 1 and have

 $\lambda = P[p(2/1) = .25] = .40$

- 80 -







C3: Statement of the Classification Rule.

If $P(Z_1, Z_2, Z_3) > .40$ classify \underline{Z} as belonging to π_2 ; if $P(Z_1, Z_2, Z_3) \leq .40$ classify \underline{Z} as belonging to π_1 .

Step D: Measurement of Operational Effectiveness.

Using Table 1 of step C_1 and the classification rule of step C_3 , one can readily classify each of the sample observations and obtain the totals for the following table.

		From Population		
		"1	п2	
То	π1	15	6	
Population	π2	4	13	
	Total	19	19	

We obtain

p(2/1) = 4/19 = .21p(1/2) = 6/19 = .32

Step E: The Application of the Classification Rule to the Observations Requiring Classification.

We have

$$P(\underline{z}) = -.00019 \ z_1^2 + .00376 \ z_2^2 + .00922 \ z_3^2$$

-.00736 \ z_1^2 + .00543 \ z_1^2 + .00454 \ z_2^2 + .00454 \ z_2^2 + .3242 \ z_1 - .0038 \ z_2 - .5788 \ z_3 - 8.333

For \underline{Z}_1 : $Z_1 = 34$, $Z_2 = 36$, $Z_3 = 12$

$$P(\underline{Z}_{1}) = -.00019(34)^{2} + .00376(36)^{2} + .00922(12)^{2}$$

-.00736(34)(36) + .00543(34)(12) + .00454(36)(12)
+ .3242(34) - .0038(36) + .5788(12) - 8.333
$$P(Z_{2}) = -3.248$$

- 84 -

Applying the Classification Rule

Since $P(\underline{Z}_1) < .40$, \underline{Z}_1 is classified as a π_1 For \underline{Z}_2 : $Z_1 = 55$, $Z_2 = 62$, $Z_3 = 21$ $P(\underline{Z}_2) = -.00019(55)^2 + .00376(62)^2 + .00922(21)^2$ -.00736(55)(62) + .00543(55)(21) + .00454(62)(21)+ .3242(55) - .0038(62) + .5788(21) - 8.333

 $P(\underline{Z}_{2}) = 2.126$

Applying the Classification Rule

Since $P(\underline{Z}_2) > .40$, \underline{Z}_2 is classified as a π_2 .

F. The Anderson Classification Statistic

The Anderson classification technique makes use of the assumption of a multivariate normal distribution for each of the populations. However, no equal covariance matrix assumption is used.

The classification statistic is taken to be in a general linear form.

 $A(\underline{Z}) = a_1 Z_1 + a_2 Z_2 + \dots + a_p Z_p + b$

The coefficients are estimated from the two available samples in such a way that the probabilities of misclassification is a minimum among all such linear forms.

Since the statistic is linear in the Z's which are assumed to be normally distributed, the statistic itself will be normally distributed with its moments easily determined. Thus, one can use a theoretical approach to the determination of the classification rule and the measurement of its operational effectiveness.

Reference: "Classification into Two Multivariate Normal Distributions with Different Covariance Matrices," Anderson, T. W. and Bahodur, R. R., USAF School of Aviation Medicine, Report #10 (1951).

Procedure

Step A: Selection of the Variables to be Used.

The variables to be used must be measured and have a joint distribution that can be approximated by the multivariate normal distribution. If this condition is not met, one often performs transformations upon the original variables in order to bring about a close approximation by the multivariate normal distribution function.

Step B: The Selection of the Estimation Procedure.

B1: The Estimation of Population Parameters.

Since the classification statistic involves sample estimates of population parameters, we elect to use the corresponding maximum likelihood estimates in evaluating the statistic.

The sample moment computations

$$\bar{x}_{p}^{(1)} = \sum_{n=1}^{N_{1}} x_{pn/N_{1}}^{(1)}$$

$$\bar{x}_{p}^{(2)} = \sum_{n=1}^{N_{2}} x_{pn/N_{2}}^{(2)}$$



B₂: Estimation of r:

The values of \underline{a} and b that are used involve a quantity r that is the largest root of the following equation

$$s(\mathbf{r}) = \left\{ (\mathbf{s}_{(\mathbf{r})}^{\mathbf{pq}}) \cdot \underline{\delta} \right\}' (\mathbf{s}_{\mathbf{pq}}^{(\mathbf{r}^2)}) \left\{ (\mathbf{s}_{(\mathbf{r})}^{\mathbf{pq}}) \cdot \underline{\delta} \right\} = 0$$

$$(\mathbf{s}_{(\mathbf{r})}^{\mathbf{pq}}) = (\mathbf{s}_{\mathbf{pq}}^{(\mathbf{r})})^{-1} = (\mathbf{r} \ \mathbf{s}_{\mathbf{pq}}^{(1)} + (1 - \mathbf{r}) \ \mathbf{s}_{\mathbf{pq}}^{(2)})^{-1}$$

and

 $(s_{pq}^{(r^2)}) = (r^2 s_{pq}^{(1)} + (1 - r)^2 s_{pq}^{(2)})$

It should be noted that the expression within the brackets, $\{ \}$, are vectors whose expansion is best done numerically for each numerical case. This will be illustrated in the section showing a numerical application of the method. The form of the equation s(r) = 0 will usually require numerical methods for determining the roots.

$$\mu_{p} = \bar{\mathbf{x}}_{p}^{(2)} - \bar{\mathbf{x}}_{p}^{(1)} = \frac{\sum_{n=1}^{N_{2}} x_{pn}^{(2)}}{N_{2}} - \frac{\sum_{n=1}^{N_{1}} x_{pn}^{(1)}}{N_{1}}$$

and

$$s_{pq}^{(1)} = \frac{N_1 \sum_{n=1}^{N_1} x_{pn} x_{qn} - \sum_{n=1}^{N_1} x_{pn} \sum_{n=1}^{N_1} x_{qn}}{N_1^2}$$

$$s_{pq}^{(2)} = \frac{N_2}{\sum_{n=1}^{N_2} x_{pn}^{(2)} x_{qn}^2} - \sum_{n=1}^{N_2} \frac{x_{pn}^{(2)} \sum_{n=1}^{N_2} x_{qn}^{(2)}}{N_2^2}$$

B₃: Estimation of Coefficients a₁, a₂, ..., a_p. The coefficient a_p, p = 1, 2, ..., p is obtained from the equation

- 87 .

$$a_p = \sum_{q=1}^{p} s_{(r)}^{pq} \delta_q$$

where S^{pq}_r is the general term in the inverse of the weighted covariance matrix

$$(S_{pq}^{r}) = (r S_{pq}^{(1)} + (1 - r) S^{(2)})$$

and

$$S_q = \vec{x}_q^{(2)} - \vec{x}_q^{(1)}$$

B₁₁: Estimation of b

The constant b is given by the equation

$$= \frac{\sqrt{a'(s_{pq}^{(2)})a \cdot a'} \bar{x}^{(1)} + \sqrt{a's_{pq}^{(2)}a \cdot a'} \bar{x}^{2}}{\sqrt{a'(s_{pq}^{(2)})a} + \sqrt{a'(s_{pq}^{(1)})a}} + \sqrt{a'(s_{pq}^{(1)})a}$$

Step C: The Determination of the Classification Rule. Empirical Estimation

 C_1 : Determine the empirical distribution of $A(\underline{Z})$ by evaluating $A(\underline{Z})$ for each observation in the two samples. These sets of A's are tabulated into a cumulative frequency distribution. Usually a free-hand smoothing of the cumulative frequency graph is sufficient to produce a curve to be used in estimating values of p(i/j).

- C_2 : Determination of λ (the classification constant) according to the decision strategy being used.
- C₃: Statement of the Classification Rule. The rule is then given by "If $A(\underline{Z}) > \lambda$ classify \underline{Z} as belonging to π_2 , if $A(\underline{Z}) \leq \lambda$ classify \underline{Z} as belonging to π_1 .

Theoretical Estimation

 C_1 : Determination of the moments of $A(\underline{2})$.

The conditional means are given by the equations

$$\overline{A}_1 = \sum_{p=1}^{P} a_p \overline{X}_p^{(1)} + b$$
, given \underline{Z} belongs to π_1

$$\bar{A}_2 = \sum_{p=1}^{r} a_p \bar{X}_p^{(2)} + b$$
, given \underline{Z} belongs to π_2

The conditional variances are given by the equations

$$v_A^{(1)} = \sum_{q=1}^{P} \sum_{p=1}^{P} a_p a_q S_{pq}^{(1)}$$
, given Z belongs to π_1

$$v_A^{(2)} = \sum_{q=1}^{P} \sum_{p=1}^{P} a_{p,q} s_{pq}^{(2)}$$
, given Z belongs to π_2 .

 c_2 : Determination of λ the classification constant according to the decision strategy being used.

We have the relationships for any given λ

- 88 -

$$p(2/1) = \sqrt{\frac{1}{\sqrt{2\pi}}} \int_{\frac{\lambda - \overline{A_1}}{V_A}}^{\infty} e^{-t^{2/2}} dt$$

and

$$p(1/2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\lambda - \overline{A_2}}{V_A^{(2)}}} e^{-t^{2/2}} dt$$

These relationships can be used to determine the λ that corresponds to the decision strategy.

C3: Statement of the Classification Rule.

Step D: Measurement of the Operational Effectiveness of the Classification Rule.

Empirical Estimation.

Using the values of the classification statistic $A(\underline{Z})$ that were generated in step C_1 for the sample observations, one can apply the classification rule of step C_3 to each observation and obtain the totals for the following table.



- 89 -

From this table we make the estimates

$$p(2/1) = \frac{n_{21}}{N_{1}}$$

 $p(1/2) = \frac{n_{12}}{N_{2}}$

Theoretical Estimation:

From the normality approximation of the distribution of $A(\underline{Z})$ we have

$$p(2/1) = \frac{1}{\sqrt{2\pi}} \int \frac{\lambda - \bar{\lambda}_1}{\sqrt{\lambda} - \bar{\lambda}_1} e^{-t^{2/2}}$$
$$\frac{\lambda - \bar{\lambda}_1}{\sqrt{\lambda} - \bar{\lambda}_2} \int \frac{\lambda - \bar{\lambda}_2}{\sqrt{\lambda} - \sqrt{\lambda}} e^{-t^{2/2}}$$

Step E: The Application of the Classification Rule to the Observations Requiring Classification.

The value of the statistic $A(\underline{Z})$ as given in step B_2 and B_3 is determined for the observation. Using this value, one spplies the classification rule of C_3 to determine the appropriate classification for the individual.

Illustration of the Procedure

Step A: Selection of the Variables to be Used.

We will use from the illustrative example

- I = Math Placement grade
- X₂ = English Placement grade
- X3 = General Aptitude Test Score

- 90 -

Step B: The Selection of the Estimation Procedure.

B1: The sample moment computation.

The sample summations

	Population 1					Population				
	Nı		19				N2	=	19	
Σ	x _l	Ξ	827			Σ	x ₁	diago enigs	1039	
Σ	X2	81	844			Σ	X2	-	1035	
Σ	x3		311			Σ	x ₃	-	445	
Σ	x1 ²	11	39033			Σ	x1 ²	=	59547	
E	. x ₂ ²	=	39388			Σ	x22	н	58699	
Σ	x ₃ ²	1	5641			Σ	x32	11	11408	
Σ	x1x2	11	38073			Σ	x ₁ x ₂	600 500	57651	
5	x1x3	entes regu	14007			\sum	x1x3	=	2 52 50	
\sum	X ₂ X ₃		14152			Σ	x2x3		25003	

Estimation of Population Parameters

Population 1			Population 2							
x 1 ⁽¹⁾	-	<u>827</u> 19	H	43.53		$\overline{x}_{1}^{(2)}$	=	<u>1039</u> 19	11	54.68
Ī ₂ (1)	=	<u>844</u> 19		44.42		Ī2 ⁽²⁾	11	<u>1035</u> 19	11	54.47
x ₃ ⁽¹⁾	11	<u>311</u> 19	=	15.37		x ₃ ⁽²⁾	11	<u>446</u> 19	34	23.47
s ₁₁ ⁽¹⁾		<u>56798</u> 361	11	159.83		s ₁₁ ⁽²⁾	=	<u>53772</u> 361	11	148.95

S22 ⁽¹⁾	=	<u>36036</u> 361	=	99.82	s ₂₂ ⁽²⁾	=	440 56 351	=	122,04
s ₃₃ ⁽¹⁾	п	<u>104 58</u> 361		28.97	s33 ⁽²⁾	u	<u>17836</u> 351	=	49.41
s ₁₂ ⁽¹⁾		35399 351	=	70.35	s ₁₂ ⁽²⁾	п	20194 361		55.94
s13 ⁽¹⁾	=	<u>8935</u> 351	=	24.75	S13 ⁽²⁾	=	<u>16355</u> 361	н	45.31
s23 ⁽¹⁾		<u>5404</u> 351	=	17.74	s23 ⁽²⁾	=	<u>13447</u> 351	=	37.25

- 92 -

Estimation of r.

We will solve the equation s(r) = 0 numerically by isolating the positive root.

Consider r = 0, to determine $\mathbf{m}(0)$

(1) Evaluate (S_{pq}^{r}) for given r $(S_{pq}^{(r=0)}) = 0 \begin{pmatrix} 159.83 & 70.36 & 24.75 \\ 70.36 & 99.82 & 17.74 \\ 24.75 & 17.74 & 28.97 \end{pmatrix} + 1 \begin{pmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{pmatrix}$

$$= \begin{pmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{pmatrix}$$
(S. (\mathbf{r}))

(2) Invert (S (17)

148.95	55.94	45.31		
55.94	122.04	37.25		475155
45.31	37.25	49.41		
	148.95 55.94 45.31	148.9555.9455.94122.0445.3137.25	148.9555.9445.3155.94122.0437.2545.3137.2549.41	148.95 55.94 45.31 55.94 122.04 37.25 = 45.31 37.25 49.41

Determine matrix of cofactors

$$\begin{pmatrix} 4542 & -1076 & -3446 \\ -1075 & 5307 & -3014 \\ -3445 & -3014 & 15049 \end{pmatrix}$$

where 4542 = (122.04)(49.41) - (37.25)(37.25)

Divide cofactor matrix by
$$\begin{vmatrix} S_{pq}^{(r)} \\ pq \end{vmatrix}$$

 $(S_{(r)}^{pq}) = \begin{pmatrix} .00977 & -.00226 & -.00725 \\ -.00226 & .01117 & -.00534 \\ -.00725 & -.00634 & .03167 \end{pmatrix}$

(3) Compute vector of mean differences

$$\delta_{p} = \bar{x}_{p}^{(2)} - \bar{x}_{p}^{(1)}$$

 $\delta = 11.15, 10.05, 7.10$

where

$$\delta_1 = \overline{x}_1^{(2)} - \overline{x}_1^{(1)} = 54.68 - 43.53 = 11.15$$

etc.

(4) Compute the vector of sums of products (S_r^{pq}) $(\underline{5}' = \underline{B})(say)$

$$\begin{pmatrix} .00977 & -.00226 & -.00725 \\ -.00226 & .01117 & -.00634 \\ -.00725 & -.00634 & .03167 \end{pmatrix} \begin{pmatrix} 11, 15 \\ 10, 05 \\ 7, 10 \end{pmatrix} = \begin{pmatrix} .03475 \\ .04205 \\ .22486 \end{pmatrix}$$

(5) Evaluate $(s_{pq}^{(r^2)}) = r^2 (s_{pq}^{(1)}) + (1-r)^2 (s_{pq}^{(2)})$
For r=0
 $(s_{pq}^{(r^2)}) = \begin{pmatrix} 159.83 & 70.36 & 24.75 \\ 70.36 & 99.82 & 17.74 \\ 24.75 & 17.74 & 28.97 \end{pmatrix} + 1 \begin{pmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{pmatrix}$
$$= \begin{pmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{pmatrix}$$

(6) Compute the vectors of sums of products $(\underline{B}) \cdot (S_{pq}^{(r^2)}) = (\underline{C})$ say $(\underline{C}) = (.03475, .04205, .22486) \begin{pmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{pmatrix}$

= (-17.717, -15.452, -14.251)

where

-17.717 = (...3475)(148.95) + (..04205)(55.94) + (..22486)(45.31)(7) Multiply (<u>C</u>) (<u>B</u>) = s(r)

$$\mathbf{s}(0) = (-17.717, -15.452, -14.251) \begin{pmatrix} .03475 \\ .04205 \\ .22486 \end{pmatrix} = -4.47$$

Considering r=1, repeat steps (1) through (6) to obtain s(1)

s(1) = +2.09

Using linear interpolation to obtain estimate of r

 $r = 0 + \frac{s(0)}{s(0) - s(1)} = \frac{-4.47}{-4.47 - 2.09} = +.68$

Repeat procedure for r = +.7, r = +.6, etc., until root is determined to desired accuracy. We have in illustration

- **s**(.7) = +.46
- s(.6) = +.15

Since the root was not isolated, we use r = .5 obtaining

$$a(.5) = -.17$$

Thus by linear interpolation

$$r = +.5 + \frac{.17}{.15 + .17} = .553$$

We will use this value for r in continuing the illustration.

Compute $(S_{pq}^{(r)}) = r(S^{(1)}) + (1 - r) (S^{(2)})$

$$(s_{pq}^{(r)}) = 0.553 \begin{pmatrix} 159.83 & 70.36 & 24.75 \\ 70.36 & 99.82 & 17.74 \\ 24.75 & 17.74 & 28.97 \end{pmatrix} + 0.447 \begin{pmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{pmatrix}$$

$$= \begin{pmatrix} 154.97 & 53.91 & 33.94 \\ 63.91 & 109.75 & 26.46 \\ 33.94 & 26.45 & 38.11 \end{pmatrix}$$

- 95 -

Invert (S_{pq}^(r))

$$s_{pq}^{(r)} = \begin{array}{c} 154.97 & 63.91 & 33.94 \\ 63.91 & 109.75 & 26.46 \\ 33.94 & 26.46 & 38.16 \end{array}$$

Cofactor matrix of $(R_{pq}^{(r)})$

$$\begin{pmatrix} 3482 & -1538 & -2034 \\ -1538 & 4754 & -1931 \\ -2034 & -1931 & 12923 \end{pmatrix}$$

Division of cofactors by $\left| \begin{array}{c} R_{pq} \left(\mathbf{r} \right) \right|$

$$(s_{(r)}^{pq}) = \begin{pmatrix} .00935 & -.00413 & -.00546 \\ -.00413 & .01277 & -.00518 \\ .00546 & -.00518 & .03470 \end{pmatrix}$$

$$B_2: \text{ Solve for } a_p$$

$$a_1 = .00935(11.15) -.00413(10.05) -.00546(7.10)$$

$$= .0240$$

$$a_2 = -.00413(11.15) +.01277(10.05) -.00518(7.10)$$

$$= .0455$$

$$a_3 = -.00546(11.15) -.00518(10.05) +.03470(7.10)$$

= .1487

B₃: Solve for b Evaluate $\underline{a}' \cdot (\underline{s}_{pq}^{(1)})$

$$(.0240, .0455, .1487) \begin{pmatrix} 159.83 & 70.36 & 24.75 \\ 70.36 & 99.82 & 17.74 \\ 24.75 & 17.74 & 28.97 \end{pmatrix}$$
$$= (10.72 & 8.87 & 5.71)$$

Compute $[\underline{a}' (S_{pq}^{(1)})]\underline{a}$ (10.72)(.0240) + 8.87(.0455) + 5.71 (.1487) = 1.51 Compute $\sqrt{a'(S_{pq}^{(1)})a}$

 $\sqrt{1.51} = 1.23$ Evaluate <u>a'</u> · S_{pq}⁽²⁾ (.0240, .0455, .1487) = (12.85, 12.43, 10.13) Compute [<u>a'</u> (S_{pq}⁽²⁾)] a (12.85)(.0240) + (12.43)(.0455) + 10.13(.1487) = 2.38 Compute $\sqrt{\underline{a'}}$ (S_{pq}⁽²⁾ a

$$\sqrt{2.38} = 1.54$$

Compute $\underline{a}' \overline{\underline{X}}^{(1)}$

(10.72)(43.53) + (8.87)(44.42) + (5.71)(16.37) = 9.54Compute <u>a'x</u> (2)

(10.72)(54.58) = (8.87)(54.47) + (5.71)(23.47) = 12.03

Evaluate b

$$p = \frac{1.54(9.54) + 1.23(12.03)}{1.54 + 1.23} = 1055$$

Thus

$$A(Z) = .0240 Z_1 + .0455 Z_2 + .1487 Z_3 + 1065$$

Step C: The Determination of the Classification Rule.

Empirical Estimation

C₁: Determination of the values of the classification statistic for the sample observations from π_1 and π_2 . Thus: $\underline{X}_1^{(1)} = (22, 35, 15)$ $A(\underline{X}_1^{(1)}) = .0240(25) + .0455(35) + .1487(15) + 1065$ = 1069.351

The results of these computations are tabulated below:

Population 1

Population 2

Observation No.	А	Observation No.	
1	1059.351	observation No.	A
2	1068 856	1	1059.859
3	1000.150	2	1071.574
5	1059.943	3	1074.031
4	1059.824	4	1020 814
5	1070.302	5	1072 506
5	1070.281	5	1073.300
7	1058.975	2	1071.104
8	1073 581	6	1073.559
9	1071 280	0	1074.094
ló	1060 402	9	1070.570
11	10.09.493	10	1075.311
12	1072.091	11	1072.811
10	1070.701	12	1072 418
13	1070.312	13	1070 003
14	1071.528	14	1070.093
15	1070.998	15	1072.21)
16	1070-493	15	1071.030
17	1058 530	10	1071.331
18	1021 200	17	1071.458
10	1071.790	18	1072.818
▲ 7	1071.048	19	1074.533

These data are tabulated into cumulative frequency distributions.

- 98 -

Population 1

Interval	f	ciim f	
1058.0 - 1058	5 0	10	cum p
1058.5 - 1069.	.0 3	10	1.00
69.0 - 69.	5 2	17	1.00
69.5 - 70.	.0 2	7/1	. 84
70.0 - 70.	5 4	12	. 14
70.5 - 71.	.0 2	8	. 73
71.0 - 71.	5 2	5	.42
71.5 - 72.	S 0.	4	•)2
72.0 - 72.	5 1	2	.41
72.5 - 73.	.0 0	7	. 11
73.0 - 73.	5 0	ī	.05
73.5 - 74.	.0 1	ī	•05
74.0 - 74.	5 0	0	.00
74.5 - 75.	0 0	0	.00

Population 2

Interval		1	f	f cum f		
1059.5	-	1070.0	ī	7	cum p	
70.0	-	70.5	1	2	.05	
70.5	-	71.0	2	4	.11	
71.0	-	71.5	4	8	. 41	
71.5	-	72.0	1	9	-42	
72.0	-	72.5	S	11	- 47	
72.5	-	73.0	2	13	- 50	
73.0	-	73.5	0	13	- 50	
73.5		74.0	2	1.5	. 70	
74.0	-	74.5	2	17	• 7 7	
74.5	-	75.0	1	18	.09	
75.0	-	75.5	1	19	1 00	
75.5	-	75.0	0	19	1.00	

These distributions are plotted in Figures 1 and 2 and a smoothed cumulative curve fitted free hand to the distribution to yield an empirical estimation of p(2/1) and p(1/2).

 C_2 : Determination of λ the classification constant, according to the decision strategy being used. Since we require a classification rule such that p(2/1) = 0.25, we use Figure 1 and have

 $\lambda = A[p(2/1) = 0.25] = 1071.5$





C3: Statement of the Classification Rule.

If $A(\underline{Z}) > 1071.5$, classify \underline{Z} as belonging to π_2 , if $\Lambda(\underline{Z}) \leq 1071.5$ classify \underline{Z} as belonging to π_1 .

Theoretical Estimation

+

$$\begin{split} & c_{1}: \text{ Determination of the Moments of } A(\underline{2}) \\ & \overline{\lambda}_{1} = a_{1} \, \overline{\chi}_{1}^{(1)} + a_{2} \, \overline{\chi}_{2}^{(1)} + a_{3} \, \overline{\chi}_{3}^{(1)} + b \\ & = .0240(43.53) + .0455(44.42) + .1487(16.37) + 1065 \\ & = 1070.5 \\ & \overline{\lambda}_{2} = a_{1} \, \overline{\chi}_{1}^{(2)} + a_{2} \, \overline{\chi}_{2}^{(2)} + a_{3} \, \overline{\chi}_{3}^{(2)} + b \\ & = .0240(54.68) + .0455(54.47) + .1487(16.37) + 1065 \\ & = 1071.2 \\ & \overline{\nu}_{1} = a_{1}^{2} S_{11}^{(1)} + a_{2}^{2} S_{22}^{(1)} + a_{3}^{2} S_{33}^{(1)} + 2a_{1}a_{2}S_{12}^{(1)} + 2a_{1}a_{3}S_{13}^{(1)} + 2a_{2}a_{3}S_{23}^{(1)} \\ & = (.0240)^{2} (159.83) + (.0455)^{2} (99.82) + (.1487)^{2} (28.97) \\ & + 2(.0240) (.0455) (70.36) + 2(.0240) (.1487) (24.75) + 2(.0455) (.1487) (17.74) \\ & \nu_{A}^{(1)} = 1.51 \\ & \nu_{A}^{(2)} = a_{1}^{2} S_{11}^{(2)} + a_{2}^{2} S_{22}^{(2)} + a_{3}^{2} S_{33}^{(2)} + 2a_{1}a_{2}S_{12}^{(2)} + 2a_{1}a_{3}S_{13}^{(1)} + 2a_{2}a_{3}S_{23}^{(2)} \\ & = (.0240)^{2} (148.95) + (.0455)^{2} (122.04) + (.1487) (49.41) \\ & + 2(.0240) (.0455) (55.94) + 2(.0240) (.1487) (45.31) + 2(.0455) (.1487) (37.25) \\ & = 2.38 \end{split}$$

 $\mathtt{C}_2\colon$ Determination of λ , the classification constant, according to the decision strategy being used.

Since we require that p(2/1) = 0.25, we have

$$0.25 = \frac{1}{\sqrt{2\pi}} \int_{\lambda - \overline{\lambda_1}}^{1} e^{-t^{2/2}} dt$$
Thus,

$$\frac{\lambda - 1070.5}{\sqrt{1.51}} = 0.67$$

$$\lambda = (1.23)(0.67) + 1070.5 = 0.82 + 1070.5$$

$$= 1071.32$$

C3: Statement of the Classification Rule:

If $\Lambda(\underline{Z}) > 1071.32$ classify \underline{Z} as belonging to π_2 . if $\Lambda(\underline{Z}) \leq 1071.32$, classify \underline{Z} as belonging to π_1 . Step D: Measurement of Operational Effectiveness.

Using Table 1 of step C_1 (empirical) and the corresponding classification rule of step C_3 , one can readily classify each observation of the two samples and obtain the totals of the following table.

		From Population	
		<u>π</u> 1	π2
То	Π	15	8
Population	π2	4	11
	Total	19	19

Thus p(2/1) = 4/19 = .21and p(1/2) = 8/19 = .42

Step E: The Application of the Classification Rule to the Observations Requiring Classification.

We have

 $A(\underline{Z}) = .0240 \ Z_1 + .0455 \ Z_2 + .1437 \ Z_3 + 1065$ For $\underline{Z}_1: \ Z_1 = .0240 \ (.034) + .0455 \ (.036) + .1487 \ (.036) + .1065 = .1069.238$ Since $A(\underline{Z}_1) \le .0240 \ (.034) + .0455 \ (.036) + .1487 \ (.036) + .1065 = .1069.238$ Since $A(\underline{Z}_1) \le .071.5, \ \underline{Z}_1$ is classified as a π_1 For $\underline{Z}_2: \ Z_1 = .55, \ Z_2 = .62, \ Z_3 = .21$ $\Lambda(\underline{Z}_2) = .0240(55) + .0455(52) + .1487(21) + 1055 = 1072.264$ Since $\Lambda(\underline{Z}_2) > 1071.5$, \underline{Z}_2 is classified as a π_2 .

G. The Shaw Classification Statistic

The Shaw classification technique consists of using the multivariate normal assumption for the density function for each population. The vectors of means are assumed to be equal, with the differences between the covariance matrices of the two populations being available for yielding the classification rule. Simplication of the likelihood ratio under these assumptions yields the statistic in the form

$$S(\underline{Z}) = \sum_{q=1}^{p} \sum_{p=1}^{p} (S_{(2)}^{pq} - S_{(1)}^{pq}) Z_{p}^{2} Z_{q},$$

where S^{pq} is the p,q term in the inverse of the corresponding covariance matrix. Although the vector of means is assumed equal for the purpose of evolving the classification statistic, for computation of the covariance matrices, the individual population means as estimated from each sample are used.

The statistic S can be approximated by a \mathbf{X}^2 distribution thus enabling one to use this theoretical distribution to estimate the classification constant as well as the operational effectiveness of the statistic, however, the procedure is too involved for demonstration in this handbook. We will restrict our consideration to the empirical estimation approach to these problems.

Reference:

Shaw, R. H., The Multivariate Classification Statistic with Two Specified Normal Alternatives, Research Report, RC-412, IBM Research Center (1961).

- 104 -

Procedure

Step A: Selection of the Variables to be Used.

The variables to be used must be measured and have a joint distribution that can be approximated by the multivariate normal distribution. If this condition is not met, one often performs transformations upon the original variable in order to bring about a closer approximation.

Step B: The Selection of the Estimation Procedure.

B1: Estimation of the Population Covariances.

We chose to use the maximum likelihood estimates



B2: Inversion of the Covariance Matrices.

$$\begin{pmatrix} s_{11}^{(R)} & s_{12}^{(R)} & \cdots & s_{1p}^{(R)} \\ s_{21}^{(R)} & & & & \\ & & & & \\ & & & & \\ s_{p1}^{(R)} & & & s_{pp}^{(R)} \end{pmatrix}^{-1} = \begin{pmatrix} s_{(R)}^{(11)} & s_{(R)}^{(21)} & \cdots & s_{(R)}^{(P1)} \\ s_{(R)}^{(21)} & s_{(R)}^{(22)} & \cdots & s_{(R)}^{(P1)} \\ \vdots & & & \vdots & \vdots \\ \vdots & & & & \vdots & \vdots \\ s_{(R)}^{(P1)} & & s_{(R)}^{(P1)} & \cdots & s_{(R)}^{(P1)} \end{pmatrix}$$

B3: Determination of the Classification Statistic.

$$S(\underline{Z}) = \sum_{q=1}^{P} \sum_{p=1}^{P} (S_{(2)}^{pq} - S_{(1)}^{pq}) Z_{p} Z_{q}$$

Step C: The Determination of the Classification Rule.

- C_1 : Determine the empirical distribution of $S(\underline{Z})$ by evaluating $S(\underline{Z})$ for each observation in the two samples. These sets of $S(\underline{Z})$'s are tabulated into a cumulative frequency distribution. Usually, a free-hand smoothing of the cumulative frequency graph is sufficient to produce a curve to be used in estimating values of p(i/j).
 - C_2 : Determination of λ , the classification constant, according to the decision strategy being used.
 - C₃: Statement of the Classification Rule. The rule is given as If $S(\underline{Z}) > \lambda$, classify \underline{Z} as belonging to π_2 ; if $S(\underline{Z}) \leq \lambda$ classify \underline{Z} as belonging to π_1 .
- Step D: Measurement of the Operational Effectiveness of the Classification Rule.

Using the values of the classification statistic S(Z) that were generated in step C_1 for the sample observations, one can apply the classification rule of step C_3 to each observation and obtain the totals for the following tables.

		From Population	
		π1_	π2
То	π1	n ₁₁	n ₁₂
Population	π2	ⁿ 21	n ₂₂
	Total	N1	N ₂

From this table we make the estimates

$$p(2/1) = n_{21}/N_1$$

 $p(1/2) = n_{12}/N_2$

Step E: The Application of the Classification Rule to the Observations Requiring Classification.

The value of the statistic $S(\underline{Z})$ as given in step B_3 is determined for the observation. Using this value one applies the classification rule of step C_3 to determine the appropriate classification for the individual.

Illustration of the Procedures

Step A: Selection of the Variables to be Used.

We will use from the illustrative example

 X_1 = Math Placement Test Score

X₂ = English Placement Test Score

X3 = General Aptitude Test Score

Step B: The Selection of the Estimation Procedure.

B1: Estimation of the Population Covariances.

The Sample Summations

	Popul	ation	1	Popul	ation	2
	N ₁	11	19	NS	=	19
\geq	- ^x 1	=	827 <u></u>	x ₁	=	1039
Σ	x ₂	=	844	x2	=	1035
Σ	- X3	-	311 Σ	x3	=	446
Σ	x1 ²	11	39033 Σ	x1 ²	=	59647
Σ	- x ² ⁵	panint dina	39388	x22	=	58599
Σ	x ₃ ²	11	5541 <u></u>	x32	=	11408
Σ	X1X2	1	38073 S	X1X2	=	57651
\sum	x ₁ x ₃	1	14007	x ₁ x ₃	=	2 52 50
Σ	X2X3	dina) Ana	14152	x2x3	×	25003

Estimation of Population Covariances

s ₁₁ ⁽¹⁾	=	<u>56798</u> 19	11	159.83	s ₁₁ (2)	-	<u>53772</u> 19	=	148.95
s ₂₂ ⁽¹⁾	11	<u>36036</u> 19	11	99.82	s ₂₂ ⁽²⁾	=	<u>44056</u> 19	=	122.04
s (1) 33		<u>10458</u> 19	=	28.97	s ₃₃ ⁽²⁾	-	<u>17836</u> 19	=	49.41
s (1) 12		<u>35399</u> 19	-	70.35	s ₁₂ ⁽²⁾	н	<u>20194</u> 19		55.94
s ₁₃ ⁽¹⁾	=	<u>8935</u> 19	H	24.75	s ₁₃ ⁽²⁾	. 11	<u>15355</u> 19	=	45.31
s ₂₃ ⁽¹⁾	=	<u>5404</u> 19		17.74	s ₂₃ ⁽²⁾	-	<u>13447</u> 19	=	37.25

B₂: Inversion of the Covariance Matrices* Evaluation of covariance determinants

$$\begin{vmatrix} 159.83 & 70.36 & 24.75 \\ 70.36 & 99.82 & 17.74 \\ 24.75 & 17.74 & 28.97 \end{vmatrix} = 259117$$

$$\begin{vmatrix} 148.95 & 55.94 & 45.31 \\ 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 & 49.41 \end{vmatrix}$$

*Since the order of the matrices is 3, the definitions will be used in inverting the matrices.

Evaluation of Cofactors

$$(C_{pq}^{(1)}) = \begin{pmatrix} 2577 & -1599 & -1222 \\ -1599 & 4018 & -1094 \\ -1222 & -1094 & 11004 \end{pmatrix}$$

where 2577 = (99.82)(28.97) - (17.74)(17.74)

$$(C_{pq}^{(2)}) = \begin{pmatrix} 4542 & -1076 & -3446 \\ -1075 & 5307 & -3014 \\ -3446 & -3014 & 15049 \end{pmatrix}$$

Division of Cofactors by determinant

$$(s_{(1)}^{pq}) = \begin{pmatrix} +.00958 & -.00594 & -.00454 \\ -.00594 & +.01493 & -.00407 \\ -.00454 & -.00407 & +.04089 \end{pmatrix}$$

$$(s_{(2)}^{pq}) = \begin{pmatrix} +.00977 & -.00226 & -.00725 \\ -.00226 & +.01117 & -.00534 \\ -.00725 & -.00634 & +.03167 \end{pmatrix}$$

B₃: Determination of the Classification Statistic. Evaluation of $(S_{(2)}^{pq} - S_{(1)}^{pq})$ matrix

$$(s_{(2)}^{pq} - s_{(1)}^{pq}) = \begin{pmatrix} .00019 & .00368 & -.00271 \\ .00358 & -.00375 & -.00227 \\ -.00271 & -.00227 & -.00922 \end{pmatrix}$$

Determination of the coefficients of the statistic

$$S(\underline{Z}) = +.00019 \ z_1^2 \ .00376 \ z_2^2 \ -.0 \ .22 \ z_3^2 +.00736 \ z_1^2 \ -00542 \ z_1^2 \ .0454 \ z_2^2 \ z_3^2$$

Step C: The Determination of the Classification Rule.

C1: Determination of the values of the classification statistic for

the sample observations from
$$\pi_1$$
 and π_2 .
Thus $\underline{X}_1^{(1)} = (22,35,15)$
 $S(\underline{X}_1^{(1)}) = .00019(22)^2 - .00376(35)^2 - .00922(15)^2$
 $+ .00736(22)(35) - .00542(22)(15) - .00454(35)(15)$
 $= -5.093$

The results of these evaluations are tabulated below.

Population 1

Population 2

Observation No.	S(<u>Z</u>)	Observation No.	S(<u>Z</u>)
1	-5.093	1	-4.903
2	0.083	2	-1.500
3	4.621	3	-12.672
4	0.131	4	-3.200
5	-3.010	5	-11.133
5	-2.494	5	-3.054
7	-4.297	7	-14.860
8	-11,259	8	-6.688
9	-0.598	9	-5.789
10	5.201	10	-15.906
11	0.410	11	-0.508
12	0.781	12	-7.840
13	_4.611	13	-2.975
14	-3.881	14	-5.559
15	-7.905	15	-1.554
16	-4.405	16	-2.576
17	-0.833	17	-1.318
18	-9.508	18	-9.004
19	-5.875	19	-27.191

The data are tabulated into cumulative frequency distributions.

	Population	1	
Interval -11.999 to -11.000 -10.999 to -10.000 - 9.999 to - 9.000 - 8.999 to - 8.000 - 7.999 to - 7.000 - 5.999 to - 5.000 - 5.999 to - 5.000 - 4.999 to - 4.000 - 3.999 to - 3.000 - 2.999 to - 1.000 - 0.999 to - 1.000 + 0.001 to + 1.000 + 1.001 to + 2.000 + 3.001 to + 3.000 + 3.001 to + 4.000 + 4.001 to + 5.000	f 1 0 1 0 2 3 2 1 0 2 4 0 0 0 1 1	cum f 1 2 2 3 3 5 8 10 11 13 17 17 17 17 18 19	cum p .05 .05 .11 .11 .16 .16 .26 .42 .53 .58 .58 .58 .58 .58 .58 .58 .89 .89 .89 .89 .89 .89 .89 .89 .89
Interval -27.2 -15.999 to -15.000 -14.999 to -14.000 -13.999 to -13.000 -12.999 to -12.000 -11.999 to -11.000 -10.999 to -10.000 - 9.999 to - 9.000 - 8.999 to - 8.000 - 7.999 to - 7.000	<u>Population</u> f l l l 0 l l 0 l 1 0 l	2 cum f 19 18 17 16 16 15 14 14 14 13 13	cum p 1.00 .95 .89 .84 .84 .84 .79 .74 .74 .58 .68
$\begin{array}{r} - 0.999 \text{ to } - 0.000 \\ - 5.999 \text{ to } - 5.000 \\ - 4.999 \text{ to } - 4.000 \\ - 3.999 \text{ to } - 3.000 \\ - 2.999 \text{ to } - 3.000 \\ - 1.999 \text{ to } - 1.000 \\ - 0.999 \text{ to } - 0.000 \end{array}$	3 0 1 2 2 3 1	12 9 9 8 6 4	.63 .47 .47 .42 .32 .21

- 110 -

These distributions are plotted in Figure 1 and 2 and a smooth cumulative curve fitted free hand to the distribution to yield an empirical estimation of p(2/1) and p(1/2).





 $\lambda = F[p(2/1) = .25] = 5.60$

C₃: Statement of the Classification Rule. If $S(\underline{Z}) < -5.50$, classify \underline{Z} as belonging to π_2 ; if $S(\underline{Z}) \geq -5.60$, classify \underline{Z} as belonging to π_1 .

Step D: Measurement of Operational Effectiveness.

Using Table 1 of step C_1 and the classification rule of step C_3 , one can readily classify each observation of the two samples and obtain the totals for the following table.

From Pepulation

		π1	π2
To	πl	15	9
Population	π2	4	10
	Total	19	19

Thus p(2/1) = 4/19 = .21p(1/2) = 9/19 = .47

Step E: Application of the Classification Rule to the Observations Requiring Classification.

We have

$$s(\underline{z}) = .00019 z_1^2 - .00376 z_2^2 - .00922 z_3^2 + .00736 z_1^2 z_2^2 - .00542 z_1^2 z_3^2 - .00454 z_2^2 z_3$$

For \underline{Z}_1 : $Z_1 = 34$, $Z_2 = 35$, $Z_3 = 12$ $S(\underline{Z}_1) = .00019(34)^2 - .00376(35)^2 - .00922(12)^2$ + .00736(34)(36) - .00542(34)(12) - .00454(36)(12) = -1.145Since $S(\underline{Z}_1) > -5.60$ \underline{Z}_1 is classified as a π_1 For \underline{Z}_2 : $Z_1 = 55$, $Z_2 = 62$, $Z_3 = 21$ $S(\underline{Z}_2) = -5.018$

Since $S(Z_2) > -5.60$, Z_2 is classified as a π_2 .

Chapter V: The Multi-Population Problem

In the previous chapters, classification techniques have been described that could be applied to the problem of classifying an individual into one of two populations. In this chapter we will briefly consider methods of handling the classification problem when more than two populations are involved.

An obvious approach to this multi-population problem is to follow either a sequential classification in which the populations are paired and then a decision made as to the classification to be made for each pair in turn. Thus, if there are four populations involved, say π_1 , π_2 , π_3 and π_4 , the classification program would involve

- (1) The π_1 or π_2 classification, say $\pi(1)$
- (2) The π_3 or π_4 classification, cay $\pi_{(2)}$

(3) The $\pi_{(1)}$ or $\pi_{(2)}$ classification for the final decision. Or one could use the composite population approach creating, say in the four population situation, the composite populations $\pi_{(1)} = \pi_1 + \pi_2$ and $\pi_{(2)} = \pi_3 + \pi_4$. Then the sequence would be

- (1) The $\pi_{(1)}$ or $\pi_{(2)}$ classification
- (2) The π_1 or π_2 or π_3 or π_4 classification depending on the outcome of (1)

- 114 -

In either of these approaches, the process of pair creates a decision difficulty since one must determine how the original pairings should be made. It is apparent that the final outcome may be sensitive to what pairing is decided upon. This problem is somewhat alleviated if there is a natural ordering of the populations so that adjacent populations can be paired. In other cases a random pairing procedure must be used. In either case one encounters difficulty in determining the appropriate decision strategy to be followed for the sequencial classification decisions since the probabilities of misclassification are not independent of the proceeding decisions. Experience will assist a person in evolving a workable solution for particular types of classification problems.

If, however, one is using the minimum expected loss criteria and has available a priori probabilities for the populations then for a given \underline{Z} one can associate an expected loss if the individual is classified into the \underline{L} th population. Here

$$C_{f} = \sum_{\substack{m=1\\ m \neq k}}^{K} q_{m} p_{m} (Z) C(k/m)$$

The classification technique involves the computation of all K C_{μ} 's and classifying Z into the population with the smallest C_{μ} . This procedure is applicable to any classification technique that generates an estimation of p_m (Z) including nonparametric, categorical and the general parametric techniques discussed in the previous chapter.

It should be recognized that the problem of clssification is essentially that of choosing regions R_1, R_2, \ldots, R_K in the sample space such that if an observation falls in the region R_K we classify the individual as belonging to population π_K . If no priori probabilities can be assumed, the conditional expected loss is

$$r(l, R) = \sum_{k=1}^{K} C(k_{1}l') P\{k/l, R\}$$

where $C(\mathcal{L}/k)$ is the cost of classifying an individual into population $\pi_{\mathcal{L}}$ if it really belongs to population π_{k} and $P\left\{k/\mathcal{L}, R\right\}$ is the probability of making the classification given that the individual belongs to $\pi_{\mathcal{L}}$ for some classification rule R. If one elects to use the minimax criteria for selection of the R's, he must for each possible set of R's,

- 116 -

- (a) maximize r (\mathcal{L} , R) overall possible \mathcal{L} 's and
- (b) Choose R₁, R₂, ..., R_k so as to minimize this maximum expected loss.

Although the general solution for this approach is not known, in the special case where all errors of misclassification are equally costly and there is no gain or loss when the observation is classified correctly, we have some guidance as to how to select the R's from a theorem due to T. W. Anderson.

"If $p_k(\underline{Z})$ is the probability density of π_k (k=1, 2, ..., k) and if the costs of misclassification of an observation are constant, the regions of classification $R_1 R_2, \ldots, R_k$ that minimize the maximum expected loss (for observations from the given π_k) are such that the probability of correct classification are equal and, if \underline{Z} is in R_k , then

 $P_k(\underline{Z}) > M_k \mathcal{L} P \mathcal{L}(\underline{Z})$ where $k \neq \mathcal{L}$. The constants $M_k \mathcal{L}$ are determined so that the common probability of correct classification is maximized.ⁿ

Chapter VI

Bibliography* on Classification and Related Topics

- Adhikari, B. P. & D. D. Joshi, Distance-discrimination et resume exhaustif, <u>Publs, inst. stat. Univ. Paris</u>, vol. 5, (1955), Fasc. 2, 57-74.
- Anderson, T. W., Classification by multivariate analysis, <u>Psychometrika</u>, 16, (1951), 31-51.
- Anderson, T. W., <u>An Introduction to Multivariate Statistical Analysis</u>, Chapt. 5, John Wiley & Sons, Inc., New York, (1958).
- Anderson, T. W., On a problem of classification in multivariate normal distribution, A.F. Research Report, AF 41(657)-214, Columbia University.
- Anderson, T. W. & R. R. Bahadur, Classification into two multivariate normal distributions with different covariance matrices, USAF School of Aviation Medicine, Report #10, (1961).
- Armitage, P., Sequential analysis with more than two alternative hypotheses and its relation to discriminant function analysis, <u>J. Rov. Stat. Soc.</u>, B, 12, (1950), 137-144.
- Bahadur, R. R., On classification based on responses to n dichotomous items, USAF School of Aviation Medicine, Report 60-13, (1959)
- Barnard, M. M., The secular variations of skull characters in four series of Egyptian skulls, <u>Ann. Eugen</u>, 6, (1935), 352-371.
- Bartlett, M. S., A note on tests of significance in multivariate analysis, Proc. Camb. Phil. Soc., 35, 180-85.
- Bartlett, M. S., The standard errors of discriminant function coefficients, J. Roy. Stat. Soc. Suppl., 6, (1939), 169-173.
- Bartlett, M. S., Multivariate analysis, <u>J. Roy. Stat. Soc. Suppl.</u>, 9, (1947), 176-190.
- Bartlett, M. S., An inverse matrix adjustment arising in discriminant analysis, <u>Ann. Math. Stat.</u>, 22, (1951), 107-111.
- Baten, W. D., The discriminant function applied to spore measurements, <u>Mich. Acad. of Sci., Arts, and Letters</u>, 29, (1943), 3-7.
- Baten, W. D., The use of discriminant functions in comparing judges scores concerning potatoes, <u>J. Amer. Stat. Assoc.</u>, 40, (1945), 223-227.

*This bibliography was compiled by Harry O. Posten, I.B.M. Research Center.

- Baten, W. D. & C. C. DeWitt, Use of the discriminant function in the comparison of proximate coal analyses, <u>Indust. & Eng. Chem.</u>, Anal. Ed., 16, (1944), 32-34.
- Baten, W. D. & H. M. Hatcher, Distinguishing method differences by use of discriminant functions, J. of Exp. Ed., (March, 1944).
- Bauer, R. K., Experience with discriminant functions (abstract), Biometrics, 11, (1955), 242-243.
- Behrens, W. U., A note on discriminatory analysis (in German), <u>Biometrische</u> <u>Zeitschrift</u>, 1, (1959), 3-14.
- Ehattacharya 1., On discrimination and divergence, Proc. Indian Sci. Congress, part III, 13, (1942).
- Bhattacharya, A., On discrimination and divergence (abstract), <u>Sankyha</u>. (1943), 249.
- Bhattacharya, A., On a measure of divergence between two statistical populations defined by their probability distributions, <u>Bull, Calcutta</u> <u>Math. Soc.</u>, 35, (1943), 99-109.
- Bhattacharya, A., On a measure of divergence of two multinomial populations, Sankhya, 7, (1946), 401-406.
- Bhattacharya, B. C., An alternate method of derivation of the distribution of Mahalanobis D²-Statistic, <u>Bull. of the Calcutta Math Soc.</u>, 33, (1941), 87-92.
- Bhattacharya, D. &. & R. D. Narayan, Moments of the D²-statistic for populations with unequal dispersions, <u>Sankhya</u>, 5, (1941), 401-412.
- Birnbaum, A. & A. E. Maxwell, Classification statistics based on Baye's formula, <u>App. Stat.</u>, 9, (1959), 152.
- Blackith, R. E., A synthesis of multivariate technique to distinguish patterns of growth in grasshoppers, <u>Biometrics</u>, 16, (1960), 28-40.
- Bose, P. K., On recursion formulae, tables and Bessel function populations associated with the distribution of classical D²-statistic, <u>Sankhya</u>, 8, (1947), 235-248.
- Bose, P. K., Incomplete probability integral tables connected with studentised D²-statistic., <u>Bull. Calcutta Stat. Assoc</u>., 2, (1949), 131-137.
- Bose, P. K., Corrigenda (for tables of incomplete probability integral of classical D²-statistics by P. K. Bose in <u>Sankhya</u>, vol. 8), <u>Sankhya</u>, 11, (1951), 96.

- Bose, Purnendu & S. R. Rao, Numerical study of the limiting forms of the "D²" distribution (abstract), <u>Sankhya</u>, 6, (1943), 348.
- Bose, R. C., On the exact distribution and moment coefficients of the D²-statistic, <u>Sci. Cult.</u>, 1, (935), 205-206.
- Bose, R. C., On the exact distribution and moment coefficients of the D²-statistic, <u>Sankhva</u>, 2, (1936), 143-154.
- Bose, R. C., A note on the distribution of differences in mean values of two samples drawn from two multivariate normally distributed populations and the definition of the D-statistic, <u>Sankhya</u>, 2, (1936), 379-384.
- Bose, R. C., On the application of linear estimation to the general theory of the analysis of doubly classified hetergeneous materials, <u>Proc.</u> <u>Ind. Sc. Congress</u>, (1946).
- Bose, R. C., The fundamental theorem of linear estimation for correlated variates, <u>Proc. Ind. Sc. Congress</u>, (1946).
- Bose, R. C. & S. N. Roy, The distribution of the studentized D²-statistic, Proc. Ind. Stat. Conference, Calcutta, pp. 19-38, (1938).
- Bose, R. C. & S. N. Roy, The distribution of the studentized D²-statistic, <u>Sankhva</u>, 4, (1938), 19-38.
- Bose, S. N., On the complete moment-coefficients of the D²-statistic, Sankhya, 2, (1936), 385-396.
- Bose, S. N., On the moment coefficients of the D²-statistic and certain integral and differential equations connected with the multivariate normal population. Sankhya, 3, (1937), 105-124.
- Bowker, A. H., A representation of Hotelling's T² and Anderson's classification statistic W in terms of simple statistics, #12 in <u>Contributions</u> <u>to probability and statistics</u>, Essays in honor of Harold Hotelling, Stanford U. Press, 1950.
- Bowker, Albert & Rosedith Sitgreaves, An asymptotic expansion for the distribution function of the classification statistic W, Tech. report #53, Applied Math. & Stat. Labs., Stanford U., Dec. 15, 1959.
- Brier, G. W., R. G. Schoot & V. L. Simmons, The discriminant function applied to quality rating in sheep, <u>Proc. Amer. Soc. An. Prod.</u>, 1, (1940), 153-150.
- Brogden, Hubert E., A simple proof of a personnel classification theorem, <u>Psychom</u>., 19, (1954), 205-208.
- Brogden, Hubert E., Least squares estimates and optimal classification, <u>Psychom</u>., 20, (1955), 249-252.
- Brown, George W.; Discriminant functions, <u>Ann. Math. Stat.</u>, 18, (1947), 514-528.
- Brown, G. W., Basic principles for construction and application of discriminators (abstract), <u>Biometrics</u>, 5, (1949), 80.

- Brown, George, W., Basic principles for construction and application of discriminators, <u>J. Clin. Psychol.</u>, 6, (1950), 58-61.
- Eurt, Cyril, Statistical problems in the evaluation of army tests, <u>Psychom</u>., 9, (1944), 219-235.
- Cacoullos, Theophilos, Computing distances between Multivariate Populations I(abstract), <u>Ann. Math. Stat.</u>, 32, (1961), 917.
- Chernoff, H., A classification problem, Tech. report #33, Appl. Math. & Stat. Lab., Stanford U., (1955).
- Clunies-Ross, C. W. & R. H. Riffenburgh, Geometry and Linear Discrimination, <u>Biometrika</u>, 47. (1960), 185-189.
- Cochran, William G., On the performance of the linear discriminant function, Work assisted by contract with O.N.R., Navy Dept. (no date).
- Cochran, W. G. & C. J. Bliss, Discriminant functions with covariance. Ann. Math. Stat., 19, (1948), 151-176.
- Cochran, William G. & Carl E. Hopkins, Some classification problems with multivariate qualitative data, <u>Biometrics</u>, 17, (1961), 10-32.
- Cox, D. R. & L. Brandwood, On a discriminatory problem connected with the works of Plato, <u>J. Roy. Stat. Soc</u>., B, 21, (1959), 195-200.
- Cox, G. M. & W. P. Martin, Use of a discriminant function for differentiating soils with different Azotobacter populations, <u>Iowa State Col. J.</u> <u>Sci.</u>, 9, (1937), 323-331.
- Das, A. C., A note on the D²-statistic when the variances and covariances are known, <u>Sankhva</u>, 8, (1948), 372-374.
- Day, Bessie B. & Marion M. Sandomire, Use of the discriminant function for more than two groups, <u>J. Amer. Stat. Assoc</u>., 37, (1942), 451-472.
- Durand, D., Risk elements in consumer installment financing, Nat. Bur. of Econ. Res. Inc., Studies in consumer installment financing #8, (1941).
- Dwyer, Paul S., Solution of the personnel classification problem with the method of optimal regions, <u>Psychom.</u>, 19, (1954), 11-26.
- East, D. L. & L. Oschinsky, A comparison of serological and somatometrical methods used in differentiating between certain East African Racial Groups with special reference to D Analysis, <u>Sankhya</u>, 20, (1958), 31-68.
- Elfving, Gustav, Selection of item variables for prediction, USAF School of Aviation Medicine, Report 56-91, (1956).
- Elfving, Gustav, Further contributions to the theory of item selection, USAF School of Aviation Medicine, Report 57-97, (1957).

- Elfving, Gustav, Choice of Variables in Discrimination, USAF School of Aviation Medicine.
- Ellison, Bob E., A multivariate k-population classification problem, Ph.D. dissertation, U. of Chicago, (1960).
- Fisher, R. A., The use of multiple measurements in taxonomic problems, Ann. Eugen., 7, (1936), 179-188.
- Fisher, R. A., The statistical utilization of multiple measurements, <u>Ann. Eugen</u>., 8, (1938), 375-385.
- Fisher, R. A., The sampling distribution of some statistics obtained from non-linear equations, <u>Ann. Eugen</u>., 9, (1939), 238-249.
- Fisher, R. A., The precision of discriminant functions, <u>Ann. Eugen.</u>, 10, (1940), 422-429.
- Fisher, R. A., <u>Statistical Methods for Research Workers</u>, 12th Ed., p. 285-289, Hafner Pub. Co., Inc., New York, (1954).
- Fix, Evelyn, Factor analysis and discrimination, USAF School of Aviation Medicine, Report #2, (1950).
- Fix. Evelyn & J. L. Hodges, Jr., Discriminatory analysis, nonparametric discrimination: consistency properties, USAF School of Aviation Medicine, Proj. No. 21-49-004, Report #4, (1951).
- Fix, Evelyn & J. L. Hodges, Jr., Nonparametric discrimination: small sample performance, USAF School of Aviation Medicine, Report #11. (1951).
- Gales, Kathleen, Discriminant functions of socio-economic class, <u>Appl.</u> <u>Stat.</u>, 6, (1957), 123.
- Garrett, Henry E., The discriminant function and its use in psychology, <u>Psychom.</u>, 8, (1943), 55-79.
- Greenhouse, Samuel W., On the problem of discrimination between statistical populations, M. A. Thesis, George Washington U., (1954).
- Greenhouse, Samuel W., Information theory and the statistical problem of discrimination, Ph.D. dissertation, George Washington U., (1959).
- Harter, Harmon Leon, On the distribution of Wald's classification statistic, <u>Ann. Math. Stat.</u>, 22, (1951), 58.
- Hodges, J. L., Jr., Survey of Discriminatory Analysis, USAF School of Aviation Medicine, Report #1, (1955).
- Horst, P. & Smith, S., The discrimination of two racial samples, <u>Psychometrika</u>, 15, 271-90.
- Hoel, P. G. & R. P. Peterson, A solution to the problem of optimum classification, <u>Ann. Math. Stat.</u>, 20, (1949), 433-438.

- Hotelling, Harold, The generalization of student's ratio, Ann. Math. Stat., 2 (1931), 360-78.
- Hughes, H. M., Discriminatory of accident prone individuals, USAF School of Aviation Medicine, Report #3, (1950).
- Hsu, P. L., On the distribution of roots of certain determinantal equations, <u>Ann. of Eugenics</u>, 9, (1939), 250-58.
- John, S., On errors in discrimination (unpublished) 1959.
- John, S., The distribution of Wald's classification statistic when the dispersion matrix is known, <u>Sankhya</u>, 21, (1959), 371-375.
- John, S., On some classification problems I, Sankhva, 21, (1950), 301-308.
- John, S., On some classification statistics, Sankhya, 22, (1960), 309-316.
- Johns, M. V., An empirical Bayes approach to non-parametric classification, USAF School of Aviation Medicine, Report 60-16, (1959).
- Johnson, Palmer O., The quantification of qualitative data in discriminant analysis, J. Amer. Stat. Assoc., 45, (1950), 65-76.
- Kendall, M. G., <u>A Course in Multivariate Analysis</u>, Chapt. 9, Charles Griffin & Co., Limited, London (1957).
- Kossack, Carl F., On the mechanics of classification, <u>Ann. Math. Stat.</u>, 16, (1945), 95-97.
- Kossack, Carl F., Some techniques for simple classification, <u>Proc.</u> Berkeley Symp, on Math. Stat. and Prob., (1949), 345-352.
- Kullback, S., An application of information theory to multivariate analysis, <u>Ann. Mat. Stat.</u>, 23, (1952), 88-102.
- Kullback, S., An application of information theory to multivariate analysis II, <u>Ann. Math. Stat.</u>, 27, (1956), 122-145.
- Kullback, Solomon, <u>Information theory and statistics</u>. Chapt. 13, John Wiley & Sons, Inc., New York, (1959).
- Lehmann, E. L., On the theory of selection, USAF School of Aviation Medicine, Report #7, (1951).
- Linhart, H., Techniques for discriminant analysis with discrete variables, Metrika, 2, (1959), 138-149.
- Ivanovitch, B. V., Sur la descrimination des Ensembles Statesques, <u>Pub. de l'Inst de Stat</u>., Univ. de Paris, 13, (1954), 207-265.
- Lohnes, Paul L., A comparison of test space and discriminant space classification models, Ph.D. dissertation, Harvard U., (1950).

- Lubin, A., Linear and non-linear discriminating functions, <u>Brit. J. Psych</u>. (stat. sec.), 3, (1950), 90-104.
- MacDonald, D. K. C., Information theory and its application to taxonomy, J. Appl. Physics, 23 (1952), 529-531.
- Mahalanobis, P. C., Analysis of race mixture in Bengal, <u>J. Asiatic</u> <u>Soc. Bengal</u>, 23, (1927), 301-33.
- Mahalanobis, P. C., On the generalized distance in statistics, <u>Proc. Nat.</u> <u>Inst. of Science India</u>, (1936).
- Mahalanobis, P. C., On tests and measures of group divergence, <u>J. Asiatic</u> <u>Soc. of Bengal</u>, 25, (1930), p. 541-88.
- Mahalanobis, P. C., The application of statistical methods in physical anthropometry, <u>Sankhya</u>, 4, (1940), 594-598.
- Mahalanobi's, P. C., Historical note on the D²-statistic, <u>Sankhya</u>, 9, (1949), 237-240.
- Mallows, C. L., Sequential discrimination, Sankhya, 12, (1953), 321-330.
- Maung, Khint, Discriminant analysis of Tocker's eye-colour data for Scottish school children, <u>Ann. Eugen</u>, 11, (1941), 54-75.
- Mises, R. von, On the classification of observation data into distinct groups, <u>Ann. Math. Stat.</u>, 16, (1945), 68-73.
- Nair, K. R., Use of measurements of more than one biometric character regarding size of body parts for discriminating between phases in the case of six eye-striped specimens of the desert locust, <u>Indian J. of</u> <u>Etomology</u>, 14, (1952), 126.
- Nair, K. R. & H. K. Mukerji, Classification of natural and plantation teak (tectona grandis) grown at different localities of India and Burma with respect to physical and mechanical properties, <u>Sankhva</u>, 22, (1960), 1-20.
- Nanda, D. N., The standard errors of discriminant function coefficients in plant-breeding experiments, <u>J. Roy. Stat. Soc</u>., 11, (1949), 283-290.
- Nandi, H. K., On the power function of studentised D²-statistic, <u>Bull</u>. <u>Calcutta Math. Soc.</u>, 38, (1946), 79-84.
- Neyman, J. and Pearson, E. S., On the problem of R samples, <u>Bull</u>. <u>Academic Polonaise des Sciences et des Lettres</u>, Series A (1931), 460-81.

- Narain, R. D., Some results on discriminant functions, <u>J. Indian Soc</u>. <u>Agric, Stat.</u>, 2, (1949), 49-59.
- Navarro, Joseph A., The distribution of a classification statistic with specified bivariate normal alternative populations, Research Report RC-435, IBM Research Center, (1961).
- Pai, M. V., Comparisons of the methods of classification, Ph.D. Thesis, Purdue U., (1957).
- Panse, V. G., An application of the discriminant function for selection in poultry, J. Genetics (London), 47, (1946), 242.
- Pearson, Karl, On the coefficient of racial likeness, <u>Biometrika</u>, 18, (1926), 105-17.
- Penrose, L. S., Some notes on discrimination, <u>Ann. Eugen.</u>, 13, (1947), 228-237.
- Quenouille, M. H., A further note on discriminatory analysis, <u>Ann. Eugen.</u>, 15, (1949), 11-14.
- Quenouille, M. H., Note on the elimination of insignificant variates in discriminatory analysis, <u>Ann. Eugen</u>., 14, (1949), 305-308.
- Raiffa, Howard, Statistical decision theory approach to item selection for dichotomous tests and criterion variables, USAF School of Aviation Medicine, Report 56-139, (1957).
- Rao, C. R., Tests with discriminant functions in multivariate analysis, Sankhya, 7, (1946), 407.
- Rao, C. R., The problem of classification and distance between two populations, <u>Nature</u>, 159, (1947), 30-31.
- Rao, C. R., A statistical criterion to determine the group to which an individual belongs, <u>Nature</u>, 160, (1947), 835-836.
- Rao, C. R., On the significance of the additional information obtained by the inclusion of some extra variables in the discrimination of populations (letter), <u>Cur. Sci.</u>, 16, (1947), 216-217.
- Rao, C. R., Tests of significance in multivariate analysis, <u>Bionetrika</u>, 35, (1948), 58-79.
- Rao, C. R., The utilization of multiple measurements in problems of biological classification, <u>J. Roy. Stat. Soc</u>., B, 10, (1948), 159-193.
- Pearson, Karl, Note on standardization of method of using the coefficient of racial likeness, <u>Biometrika</u>, 20, (1928), 375-78.
- Rao, C. R., On the distance between two populations, <u>Sankhya</u>, 9, (1949), 245-248.

- Rao, C. R., On some problems arising out of discrimination with multiple characters, <u>Sankhya</u>, 9, (1949), 343.
- Rao, C. R., Statistical inference applied to classificatory problems, Sankhya, 10, (1950), 229-256.
- Rao, C. R., A note on the distribution of $D^2 = D^2$ and some computational aspects of D^2 statistics, <u>Sankhya</u>, 10, p+q = p(1950), 257-268.
- Rao, C. R., The problem of selecting individuals for various duties in a specified ratio, <u>Sankhva</u>, 11, (1951), 107-116.
- Rao, C. R., Statistical inference applied to classificatory problems, Part II: The problem of selecting individuals for various duties in a specified ratio, <u>Sarkhya</u>, 11, (1951), 107-116.
- Rao, C. R., The discriminant function approach in the classification of time series (Part III of statistical inference applied to classificatory problems), <u>Sankhya</u>, 11, (1951), 257-262.
- Rao, C. R., <u>Advanced statistical methods in Biometric research</u>, chapters 7-9, John Wiley & Sons, Inc., New York, (1952).
- Rao, C. R., Discriminant functions for genetic differentiation and selection (statistical inference applied to classificatory problems, part IV), <u>Sankhya</u>, 12, (1953), 229-246.
- Rao, C. R., On the use and interpretation of distance functions in statistics, Bull. Inst. Internat. Stat.: 34, (1954), 90-97.
- Rao, C. R., A general theory of discrimination when the information on alternative hypotheses is based on samples, <u>Ann. Math. Stat.</u>, 25. (1954), 651-670.
- Rao, C. R., Multivariate analysis: An indispensable statistical aid in applied research, Sankhya, 22, (1950), 317-338.
- Rao, C. R. and P. Slater, Multivariate analysis applied to differences between neurotic groups, <u>British J. Psych.</u>, Stat. Sec., 2, (1949), 17.
- Riffenburgh, R. H. & C. W. Clunies-Ross, Linear discriminant analysis, Pacif. Sci., 14, July, 1950.
- Rogers, David J. & T. T. Tanimoto, A computer program for classifying plants, <u>Science</u>, 132, (1950), 1115-1118.
- Romanovsky, V., On the criteria that two given samples belong to the same normal populations, Metron, 7 (1928) 3-46.
- Roy, S. N., A note on the distribution of the studentized D²-statistic, Sankhya, 4, (1939), 373-380.
- Roy, S. N., p-Statistics or some generalisations in analysis of variance appropriate to multivariate problems, Sankhya, 4, (1939), 381-395.

- Roy, S. N., Distribution of certain symmetric functions of p-statistics on the null hypothesis, <u>Sci. Cult.</u>, 5, (1940), 553.
- Roy, S. N., Distribution of p-statistics on the non-null hypothesis, <u>Sci</u>. <u>Cult</u>., 5, (1940), 562-563.
- Roy, S. N., The sampling distribution of p-statistics and certain allied statistics on the non-null hypothesis, <u>Sankhva</u>, 6, (1942), 15-34.
- Roy, S. N., Analysis of variance for multivariate normal populations: the sampling distribution of the requisite p-statistic on the null and non-null hypothesis, <u>Sankhya</u>, 6, (1942), 35-50.
- Roy, S. N., The individual sampling distribution of the maximum, the minimum and any intermediate one of the p-statistics on the null hypothesis, <u>Sankhya</u>, 7, (1945), 133-158.
- Roy, S. N., A note on multi-variate analysis of variance when the number of variates is greater than the number of linear hypotheses per character, <u>Sankhya</u>, 8, (1945), 53-66.
- Roy, S. N., Multivariate analysis of variance; the sampling distribution of the numerically largest of the p-statistics on the non-null hypothesis, <u>Sankhya</u>. 8, (1946), 15-52.
- Roy, S. N., & R. C. Bose; The use and distribution of the studentized D²statistic when the variances and covariances are based on K samples, Sankhya, 4, (1940), 535-542.
- Roy, S. N. & P. K. Bose: On the reduction formulae for the incomplete probability integral of the studentised D-statistics, <u>Sci. Cult.</u>, 5, (1940), 773-775.
- Sebestyen, George: Recognition of membership in classes, <u>IRE Trans</u>. on information theory, (January, 1961), 44-50.
- Shaw, R. H., The multivariate classification statistic with two specified normal alternatives, Research Report, RC-412, IBM Research Center, (1961).
- Shrivasta, M. P., On the D²-statistics, <u>Bull. Calcutta Math. Soc</u>., 33, (1941), 71-86.
- Sitgreaves, Rosedith, On the distribution of two random matrices used in classification procedures, Ann. Math. Stat., 23, (1952), 263-270.
- Sitgreaves, Rosedith, The principle of invariance applied to a problem of classification (abstract), <u>Ann. Math. Stat.</u>, 23, (1952), 473-474.
- Sitgreaves, Rosedith, Discriminating analysis in a non-classical case (abstract), <u>Biometrics</u>, 13, (1957), 545.
- Sitgreaves, Rosedith, Some results on the distribution of W-classification statistics, USAF School of Aviation Medicine, Report 58-3, (1958).

- Smith, C. A. B., Some examples of discrimination, <u>Ann. Eugen.</u>, 13, (1947), 272.
- Smith, H. F., A discriminant function for plant selection, <u>Ann. Eugen.</u>, 7, (1936), 240-250.
- Sobel, M. & A. Wald, A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution, <u>Ann. Math. Stat.</u>, 20, (1949), 502-522.
- Solomon, Herbert, Classification procedures based on dichotomous response vectors, #36 in <u>Contributions to probability and statistics</u>. Essays in honor of Harold Hotelling, Stanford U. Press. (1960).
- Stoller, David S., Univariate two-population distribution-free discrimination, J. Amer. Stat. Assoc., 49, (1954), 770-777.
- Tanimoto, T. T., An elementary methematical theory of classification and prediction, IBM Report, November 17, 1958, IBM Corp. New York, New York.
- Tanimoto, T. T. & Robert G. Loomis, A taxonomy program for the IBM 704: the IBM taxonomy application, M-A-5, IBM Applications Library, 590 Madison Avenue, New York 22, New York (1960).
- Thorndike, Robert L., The problem of classification of personnel, <u>Psychom</u>., 15, (1950), 215-235.
- Throckmorton, Thomas Neil, Structures of classification data, Ph.D. dissertation, Iowa State U., (1960).
- Tintner, Gerhard, Some applications of multivariate analysis to economic data, J. Amer. Stat. Assoc., 41, (1946), 472-500.
- Tintper, Gerhard, Some formal relations in multivariate analysis, <u>J. Roy</u>. <u>Stat. Soc.</u>, B. 12 (1950), 95-101.
- Travers, R. M. W., The use of a discriminant function in the treatment of Psychological group differences, <u>Psychom.</u>, 4, (1939), 25-32.
- Tyler, Fred. T., Some examples of multivariate analysis in educational and psychological research, <u>Psychom.</u>, 17, (1952), 289-296.
- Votaw, D. E., Jr., Methods of solving some personnel-classification problems, <u>Psychom.</u>, 17, (1952), 255-265.
- Votaw, D. E., Jr. & T. T. Dailey, Assignment of personnel to jobs, Research Bulletin 52-24, Air Training Command, Human Resources Research Center, Lackland Air Force Base, August, 1952.
- Wald, Abraham, On a statistical problem arising in the classification of an individual into one or two groups, <u>Ann. Math. Stat.</u>, 15, (1944), 145-162.

- Wallace, David, Classification without known-class data, Paper presented at 24th annual meeting Inst. Math. Stat. (Seattle), 1961.
- Wallace, N. & R. M. W. Travers, A psychometric sociolgia study of a group of specialty salesmen, <u>Ann. Eugen.</u>, 8, (1938), 265-302.
- Welch, B. L., Note on discriminant functions, <u>Biometrika</u>, 31, (1939), 218-220.
- Welch, Peter D. & Richard S. Wimpress, Two multivariate statistical computer programs and their application to vowel recognition, <u>J</u>. <u>Acoustical Soc. of Amer.</u>, 33, (1961), 425-434.
- Wesler, Oscar, A classification problem involving multinomials, <u>Ann.</u> <u>Math. Stat.</u>, 30, (1959), 128-133.
- Wilks, S. S., Multidimensional statistical scatter, #40 in <u>Contributions</u> <u>to probability and statistics</u>, Essays in honor of Harold Hotelling, Stanford U. Press, 1960.
- Wilks, S. S., <u>Mathematical statistics</u>, sections 18.7 and 18.8, John Wiley & Sons, Inc., New York, (1962).
- Williams, E. J., Some exact tests in multivariate analysis, <u>Biometrika</u>, 39, (1952), 17-31.
- Williams, E. J., Significance tests for discriminant functions and linear functional relationships, <u>Biometrika</u>, 42, (1955), 360-381.
- Wishart, John, Multivariate analysis, Appl. Stat., 4, (1955), 103-116.
- Yardi, M. R., A statistical approach to the problem of chronology of Shakespeare's plays, <u>Sankhva</u>, 7, (1945), 263-258.
- Yates, F. & M. J. R. Healy, Statistical methods in anthropology, <u>Nature</u>, 168, (1951), 1115-1118.
- (no author) Perfect discrimination as the number of variables increases. USAF School of Aviation Medicine, Report #5.
- (no author) On the simultaneous classification of several individuals, USAF School of Aviation Medicine, Report #6.

Additional References

- Baggaler, A. R., The relationship between scores obtained by Harvard freshmen on the Kuder preference record and their fields of concentration, <u>J. Ed. Psych.</u>, 38, 421-427, (1947).
- Birnbaum, Z. W. & D. G. Chapman, On optimum selections from multi-normal populations, <u>Ann. Math. Stat.</u>, 21, 443-447, (1950).
- Day, B. B., A suggested method for allocating logging costs to log sizes, J. For., 35, 69-71, (1937).
- Hazel, L. N., The genetic basis for constructing selection indexes, <u>Genetics</u>, 28, 479-490, (1943).
- Lehman, E. L., On families of admissible tests, <u>Ann. Math. Stat.</u>, 18, 97-104, (1947).
- Lorge, I., Two-group comparisons by multivariate analysis, <u>Amer. Ed.</u> <u>Res. Assoc. Off. Reprint</u>, 117-121, (1940).
- Martin, E. S., A study of an Egyptian series of mandibles with special reference to mathematical methods of sexing, <u>Biometrika</u>, 28, 149-178, (1935).
- Park, B. C. & B. B. Day, A simplified method for determining the condition of whitetail deer herds, U.S.D.A. Tech. Bull. #840, (1942).
- Paulson, E., A multiple decision procedure for certain problems in the analysis of variance, <u>Ann. Math. Stat</u>., 20, 95-98, (1949).
- Selover, R. B., A study of the sophomore testing program at the University of Minnesota - Part II, J. Appl. Psych., 455-467, (1942).
- Stevens, W. L., The standardization of rubber flexing tests, <u>India Rubber</u> <u>World</u>, 102, 36-41, (1940).
- Wald, A., An essentially complete class of admissible decision functions, Ann. Math. Stat., 18, 549-554, (1944).
- Wald, Abraham, Statistical Decision Functions, John Wiley & Sons, Inc., New York, (1955).