A HANDBOOK OF

STATISTICAL CLASSIFICATION TECHNIQUES

by

CARL F. KOSSACK

PURDUE UNIVERSITY
Lafayette, 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

	Page
Preface	ii
Table of Contents	iii
Glossary of Terms and Notational Conventions	iv
Introduction	vi
Chapter I A Brief History of Classification Theory	1
Chapter II The General Characteristics of Classification Techniques	7
Chapter III The Illustrative Example	14
Chapter IV The 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 Multi-Population Problem	114
Chapter VI Bibliography on Classification and Related Topics	117

G L O S S A R Y   O F   T E R M S

(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 ℓ. Thus π_k, $k = 1, 2, --- K$

For 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

$\mu_p^{(k)}$ = the population mean of X_p for population π_k

$m_p^{(k)}$ = an estimate of the population mean of X_p for population π_k

The variance - covariance

$\sigma_{pq}^{(k)}$ = for p = q the variance of X_p for population π_k

for p ≠ q the covariance between X_p and X_q for population π_k

$s_{pq}^{(k)}$ = the corresponding sample estimates

The inverse of the covariance matrix, general term

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

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

A vector quantity , $\underline{X} = (X_1, X_2, ---, X_Q)$

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

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 successful students and a set of the same three measures for a group or sample of students who have been previously admitted 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 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.

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 them. It can be considered as a special case of a statistical decision problem. Given ($K \geq 2$) 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. Manufactured 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 parameters 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 $\underline{Z}_m = (Z_{1m},\ldots,Z_{Pm})$, $m = 1, 2 \ldots M$ 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 paper[1)] in Biometrika 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 $\underline{X}^{(1)}_n = (X^{(1)}_{1n},\ldots,X^{(1)}_{Pn})$ $n = 1, 2,\ldots,N_1$ and $\underline{X}^{(2)}_n = (X^{(2)}_{1n},\ldots,X^{(2)}_{Pn})$ $n = 1, 2,\ldots,N_2$ be the samples from populations π_1 and π_2, then

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

where

$$\bar{X}^{(k)}_p = 1/N_k \sum_{n=1}^{N_k} X^{(k)}_{pn} \quad p = 1,2,\ldots,P;\; k = 1, 2;$$

and s_{pp} is an estimate of the common variance of the pth character in the two populations and is given by

$$(1.2) \qquad s_{pp} = \frac{1}{N_1 + N_2}\left[\sum_{n=1}^{N_1} (X^{(1)}_{pn} - \bar{X}^{(1)}_p)^2 + \sum_{n=1}^{N_2} (X^{(2)}_{pn} - \bar{X}^{(2)}_p)^2\right]$$

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

1) Tildesley, M. L., "A First Study of the Burmese Skull," Biometrika, Vol. 13 (1921), pp. 176-262.

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. Mahalanobis[1] 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, Mahalanobis proposed the classical form of the generalized distance function D^2. When it was first proposed, the D^2 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 divergence 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

---

1) Mahalanobis, P. C., "On Tests and Measures of Group Divergence, Part I: Theoretical Formulae". Journal and Proceedings, Royal Asiatic Society of Bengal, New Series, Vol. 26 (1930), pp. 541-588.

was first obtained by R. C. Bose[1] in 1936. The studentised form of D^2 involving only sample readings was first defined by Mahalanobis[2] 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 $(x_1, x_2 \ldots, x_p)$ 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.

---

1) Bose, R. C., "On the Exact Distribution of D^2-Statistic", Sankhya, Vol. 2, (1936), pp. 143-154.

2) 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 Barnard[1)] on craniometry following the suggestion of R. A. Fisher. In 1936, Fisher[2)] 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 Hotelling[4)] 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 paper[5)] in the Annals of Mathematical Statistics (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 H_1: that the observation belongs to π_1 against

---

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

2) Fisher, R. A., "The Use of Multiple Measurements in Taxonomic Problems", Annals of Eugenics, 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", Annals of Mathematical Statistics, Vol. 15, (1944), pp. 145-162.

the alternative hypothesis H_2: 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 H_1 against H_2. 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 Psychometrika [1)], Anderson proposes a classification statistic that differs only slightly from that of Wald.

---

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

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 $f_1(\underline{z})$ and $f_2(\underline{z})$ where there are likelihood estimates and then the ratio

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

is computed. The classification rule is then simply

"If $L(\underline{z}) > \lambda$, classify the individual into π_2

and if $L(\underline{z}) \leq \lambda$, classify the individual into π_1

The techniques using this approach differ only in the means of making the estimates $f_k(\underline{z})$.

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 $z_1, z_2, \text{--} z_p$ 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

$C(\underline{z})$ $\wedge$ which is a numerical function of the observation $\underline{z}$.

The classification rule is then

"If $C(\underline{z})$ belongs to the interval I_0, classify z into population π_2,

otherwise classify $\underline{z}$ into π_1"

The techniques using this approach differ only in the form of $C(\underline{z})$ that is to be used. Methods are, of course, required for the estimation of the coefficients that appear in the classification statistic $C(\underline{z})$. 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 $C(z) \geq \lambda$ 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_0 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 (π_2) when he really belongs to population 1 (π_1) or conversely one may classify an individual into π_1 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 π_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.

Since all methods for determination of the classification "constant," λ or I_0, 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 $\underline{X}_1^{(1)}, \underline{X}_2^{(1)}, \text{--}, \underline{X}_{N_1}^{(1)}$ and then $\underline{X}_1^{(2)}, \underline{X}_2^{(2)}, \text{---} \underline{X}_{N_2}^{(2)}$, 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.

Figure 1: Emperical Estimation of p(2|1) Sample From π_1.

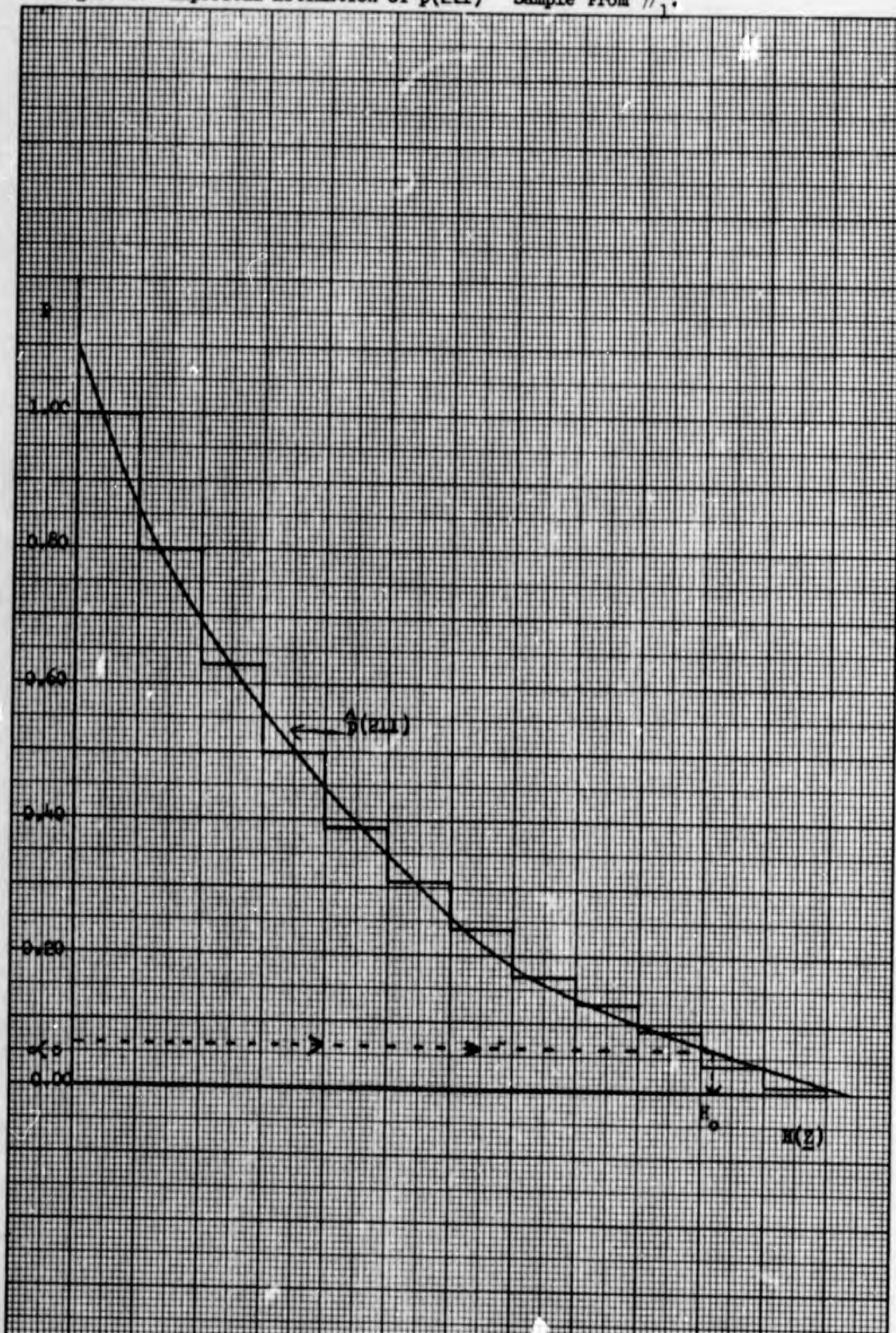
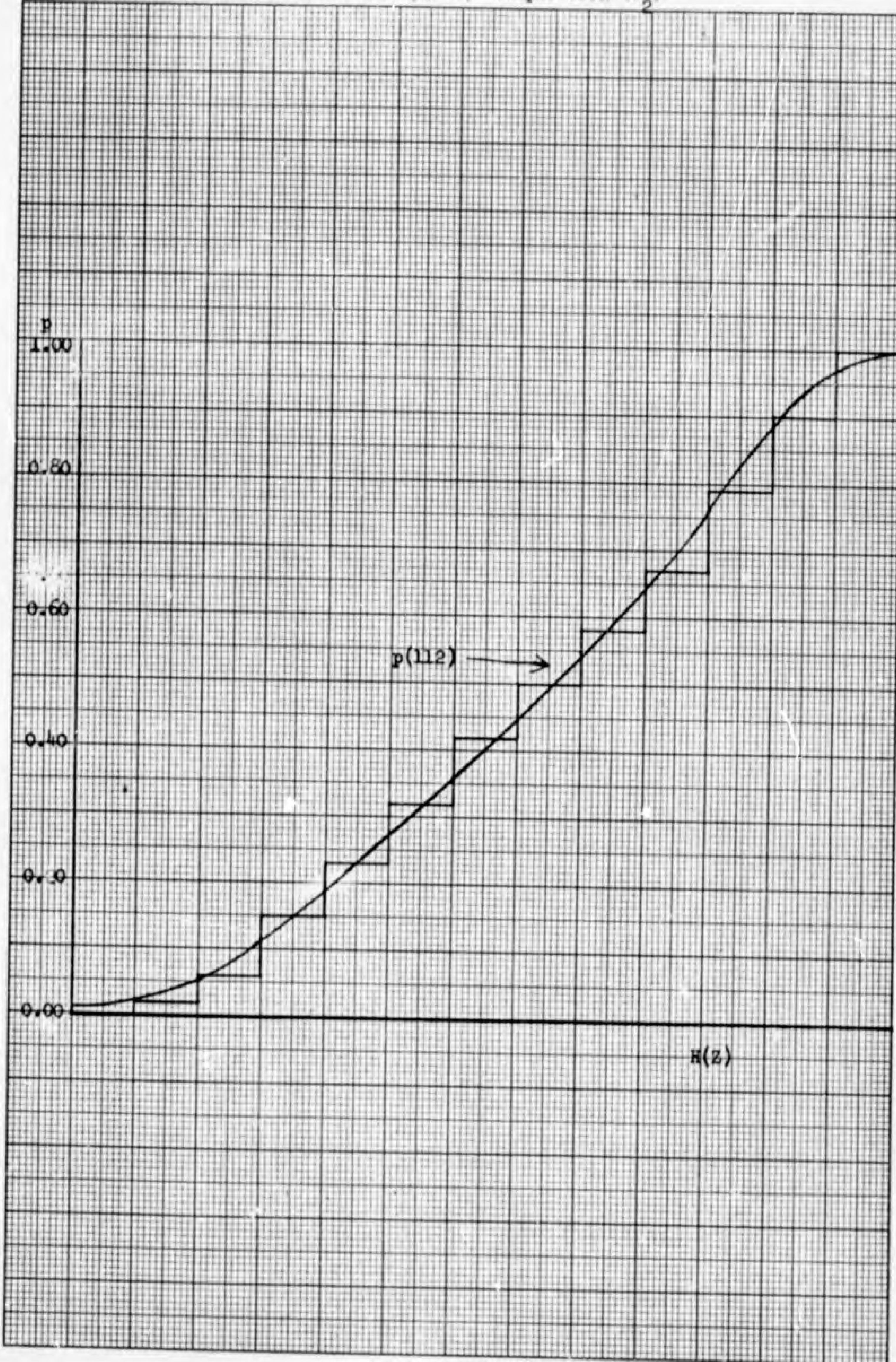


Figure 2: Emperical Estimation of p(112) Sample From π_2.



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 classification 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

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

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

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

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_o, determine the corresponding C, and then successively change the selection of H_o 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 Variables	Distribution	Paremater
A.	Non-Parametric	Measured	None	None
B.	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

CHAPTER 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 is available. Also available is a sample of nineteen successful 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 x_1	English Orientation Test Score x_2	General Aptitude Test Score x_3
1	22	35	15
2	32	45	7
3	53	48	10
4	46	36	14
5	38	54	13
6	39	53	13
7	21	24	16
8	64	59	30
9	59	45	19
10	57	36	10
11	64	60	19
12	55	44	16
13	35	46	16
14	50	55	19
15	35	48	20
16	41	37	19
17	30	26	11
18	46	40	26
N_1 = 19	40	53	18

Sample from Population π_2 (Successful Students)	Mathematics Orientation Test Score	English Orientation Test Score	General Aptitude Test Score
1	29	49	13
2	59	48	20
3	62	71	29
4	52	35	20
5	61	60	29
6	48	50	18
7	55	58	31
8	76	65	29
9	36	37	21
10	73	77	34
11	73	58	23
12	57	48	26
13	36	57	11
14	51	63	21
15	61	35	20
16	49	61	16
17	57	50	19
18	58	53	27
N_2 = 19	46	60	39

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

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 in the technique 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.

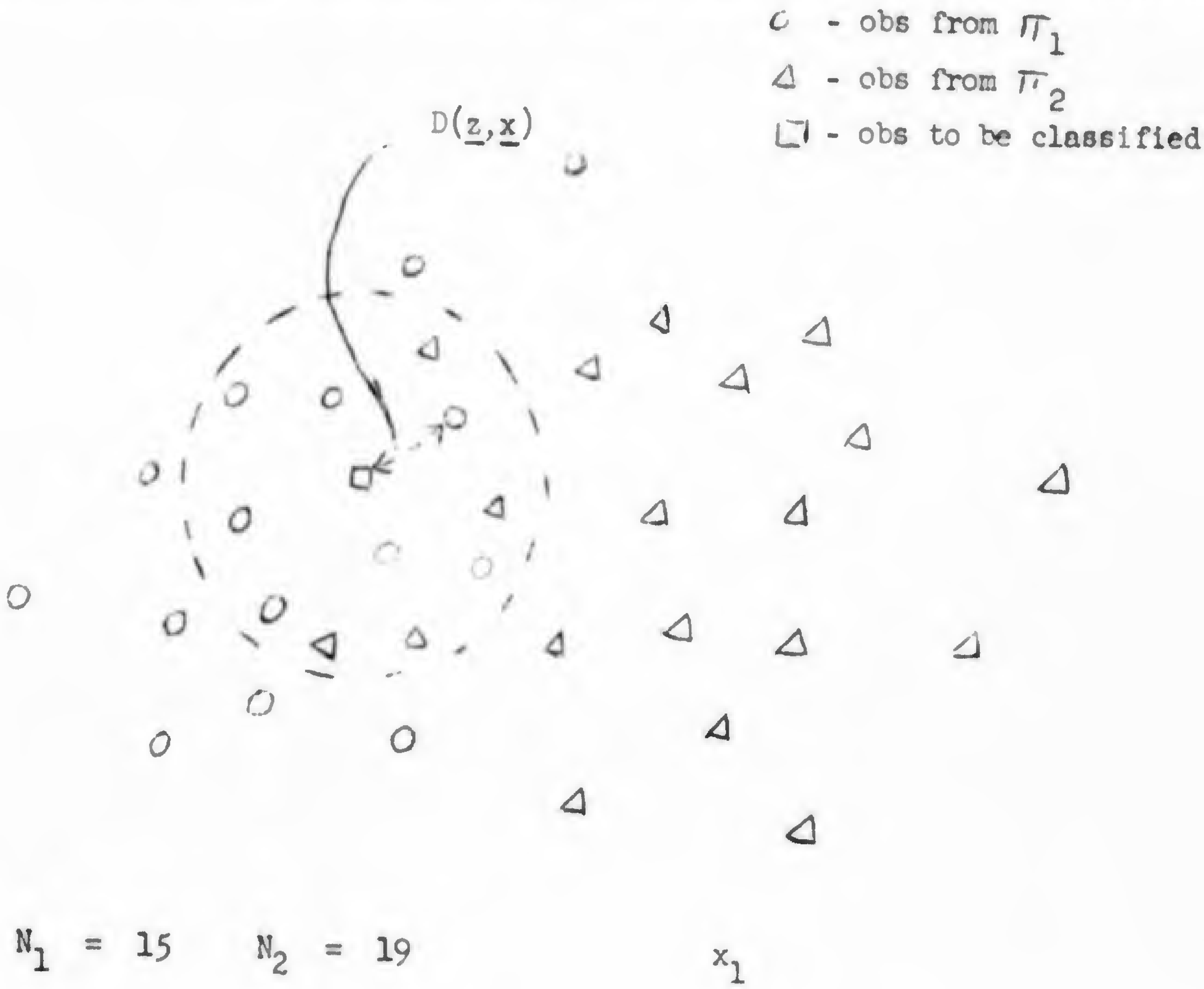
A. - Non Parametric Classification

General Description: 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, N_1 and N_2, 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.

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



$N_1 = 15 \quad N_2 = 19$

$M = 11 \;,\; n_1 = 7 \;,\; n_2 = 4$

$$\hat{f}_1(Z) = {}^{n_1}/N_1 = {}^{7}/15 = .47$$

$$\hat{f}_2(Z) = {}^{n_2}/N_2 = {}^{4}/19 = .21$$

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

Procedure

Step A. Selection of the Variables To Be Used: The Selection is restricted to measured variables, usually no transformation is performed.

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

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

$D(\underline{x}, \underline{x}^1)$ must be such that

$$D(\underline{x}, \underline{x}^1) > 0$$

$$D(\underline{x}, \underline{x}^1) = D(\underline{x}^1, \underline{x})$$

$$D(\underline{x}, x^1) + D(\underline{x}^1, \underline{x}^{11}) \geq D(\underline{x}, \underline{x}^{11})$$

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

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

B_3: Computation of $D(\underline{Z}, \underline{X})$ for all x 's from the samples from π_1 and π_2

B_4 Pool $D(\underline{Z}, \underline{X})$'s and order by increasing value. Select out first M. From these M's

B_5 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 Compute $f_1(\underline{Z}) = n_1/N_1$

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

B_7 Estimate of $R(\underline{Z}) = f_2(\underline{Z}) / f_1(\underline{Z})$

C. The Determination of the Classification Rule:

C_1: 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).

C_2: Determination of λ (the classification rule) according to the decision strategy to be used.

C_3: Statement of the Classification Rule: If $R(Z) > \lambda$, classify Z as belonging to π_2, if $R(Z) \leq \lambda$ classify Z as belonging to π_1

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	
		π_1	π_2
To Population	π_1	n_{11}	n_{12}
	π_2	n_{21}	n_{22}
Total		N_1	N_2

From this table we make the estimates

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

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

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

X_2 = 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

$$D\ (x\ ,\ x') = \sum_{p=1}^{P} \left| x_p - x'_p \right|$$

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

B_2: Decision as to number of points, M, to be included in neighborhood of $\underline{Z}$.

We will use M = 15 since $N_1 + N_2 = 38$

To illustrate steps B_3 to B_7

consider $X_{11}^{(1)} = 22$, $X_{21}^{(1)} = 35$, $X_{31}^{(1)} = 15$

B_3: Computation of the 38 D's from the given point to all the points in the sample

$$D_1^{(1)} = D(\underline{X}_1^{(1)}, \underline{X}_1^{(1)}) = |22 - 22| + |35 - 35| + |15 - 15| = 0$$

$$D_2^{(1)} = D(\underline{X}_2^{(1)}, \underline{X}_1^{(1)}) = |32 - 22| + |45 - 35| + |7 - 15| = 28$$

$$D_3^{(1)} = D(\underline{X}_3^{(1)}, \underline{X}_1^{(1)}) = |53 - 22| + |48 - 35| + |10 - 15| = 49$$

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

B_4: 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

B_5: Determine n_1 and n_2 by count of the results of step B_4

$n_1 = 12 \quad n_2 = 3$

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

$f_2(\underline{x}) = n_2/N_2 = 3/19 = .1578$

B_7 Compute $R_1^1 = \dfrac{f_2(\underline{x})}{f_1(\underline{x})} = \dfrac{.1578}{.6315} = .25$

C: The Determination of the Classification Rule

C_1: Determine the empirical distribution of R(x) by evaluating R(x) for each observation in the two samples. One must repeat Steps B_3 through B_7 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:

Table 1
Cumulative Empirical Frequency Distributions of R

R	π_1 f_1	π_1 Cum p	π_2 f_2	π_2 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
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.

Figure 1: Emperical Istimation of the Distribution of R given π_1.

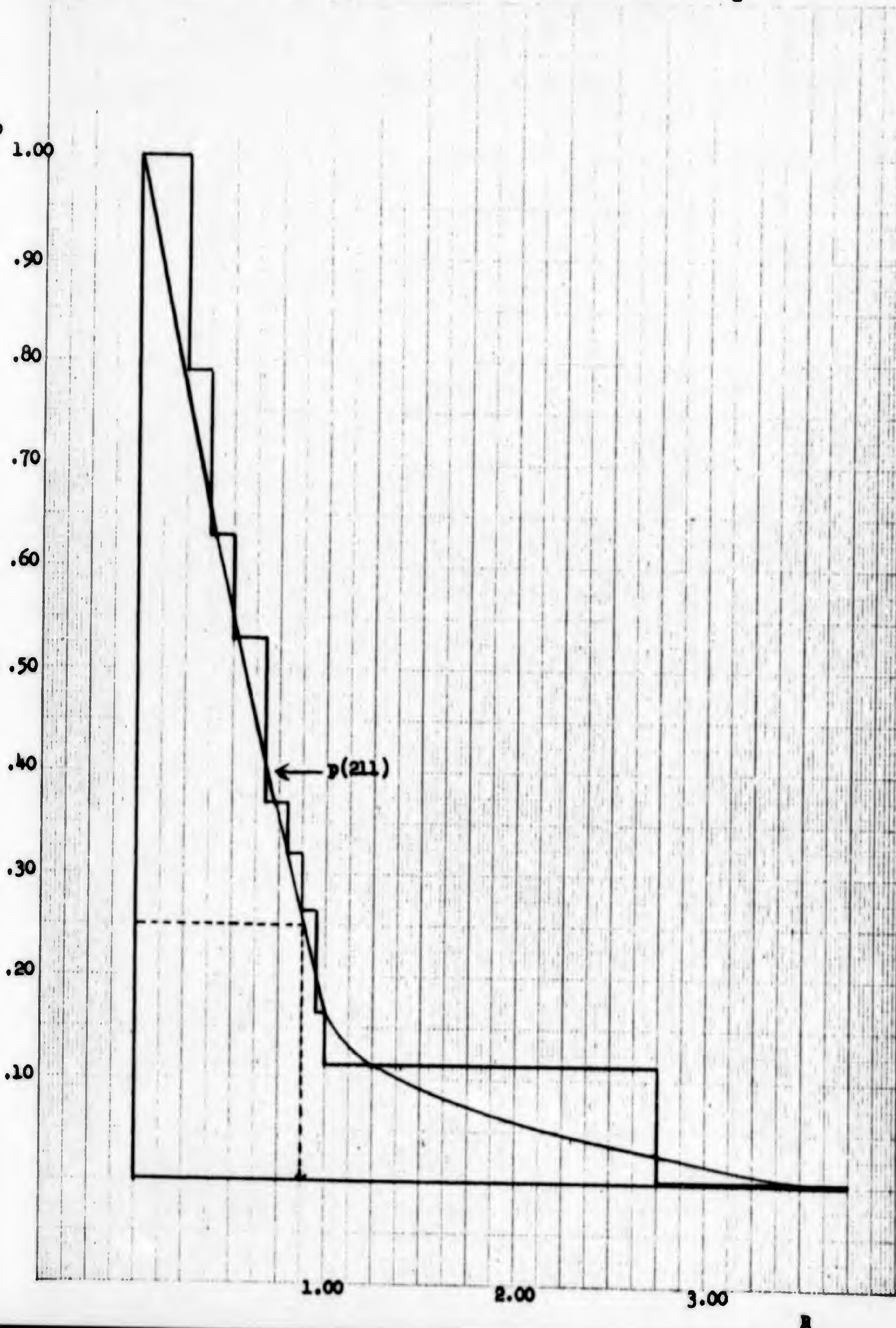
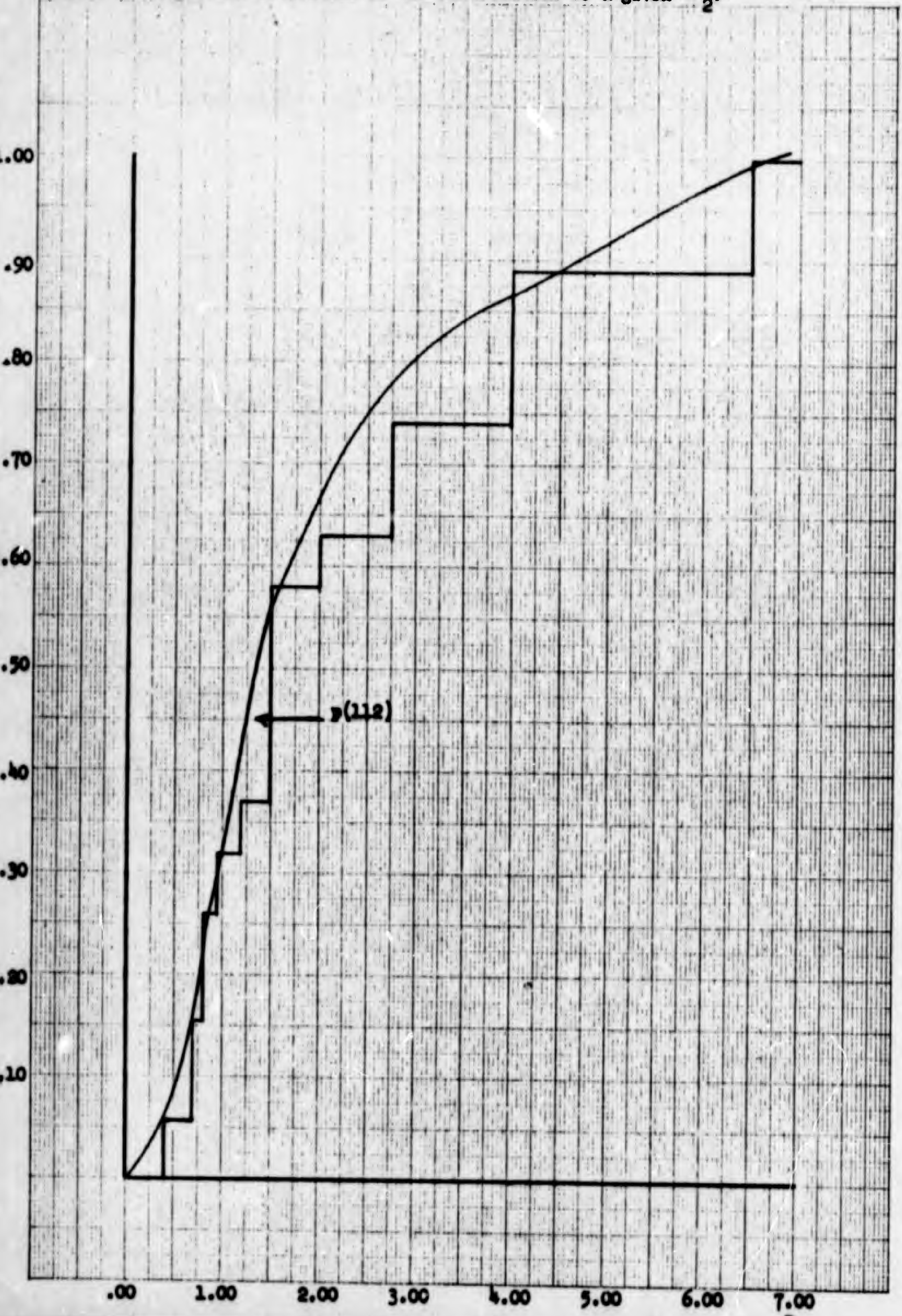


Figure 2: Emperical Estimation of the Distribution of R given π_2.



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.

λ	p(2/1)	p(1/2)	C
.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.

C_3: Statement of the Classification Rule:

If R ($\underline{Z}$) > 0.90 classify $\underline{Z}$ as belonging to π_2,

If R ($\underline{Z}$) $\leq$ 0.90 classify $\underline{Z}$ as belonging to π_1.

D. Measurement of Operational Effectiveness.

Using the rule of C_3 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

From Population

		$\pi 1$	$\pi 2$
To	π_1	14	5
Population	π_2	5	14
Total		19	19

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.

B_3: 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)

π_1 : 16, 16, 33, 14, 33, 23, 29, 71, 41, 25, 61, 33, 15, 41, 21, 15, 15, 30, 29

π_2 : 19, 45, 80, 27, 66, 35, 62, 78, 12, 92, 72, 49, 24, 53, 36, 44, 44, 56, 103

B_4: Pool and Select the 15th Smallest D's

π_1 : 14, 15, 15, 15, 16, 16, 21, 23, 23, 25, 29, 29

π_2 : 12, 19, 24

B_5: Determine n_1 and n_2

$n_1 = 12$, $n_2 = 3$

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

$f_1(\underline{Z}_1) = n_1/N_1 = 12/19 = 63$

$f_2(\underline{Z}_1) = n_2/N_2 - 3/19 = .16$

B_7: Compute $R(Z_1)$

$$R(Z_1) = \frac{f_2(\underline{Z}_1)}{f_1(\underline{Z}_1)} = \frac{.16}{.63} = .25$$

Apply Classification Rule

Since $R(\underline{Z}_1) = .25 < 0.90$, Z_1 is classified (correctly) as a π_1

For $\underline{Z}_2$, Z_{21} = .55, Z_{22} = 62, Z_{32} = 21

Following the same steps used for $\underline{Z}_1$, we have

B_3: Compute 38 $D(\underline{Z}_2, x)$'s

$D(\underline{Z}_2 , \underline{X}^{(1)}) = |22 - 55| + |35 - 62| + |15 - 21| = 66$ 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

B_4: Pool and select 15 smallest

π_1 : 13, 14, 21, 23, 23

π_2 : 5, 12, 14, 16, 16, 18, 19, 21, 22, 24

B_5: Determine n_1 and n_2

$n_1 = 5$, $n_2 = 10$

B_6: Compute $f_1(Z_2)$ and $f_2(Z_2)$

$f_1(Z_2) = {}^{n_1}/N_1 = 5/19 = .263$

$f_2(Z_2) = {}^{n_2}/N_2 = 10/19 = .526$

B_7: Compute $R(\underline{Z}_2)$

$$R(\underline{Z}_2) = \frac{f_2(\underline{Z}_2)}{f_1(\underline{Z}_2)} = \frac{.526}{.263} = 2.00$$

Apply Classification Rule:

Since $R(Z_2) = 2.00 > 0.90$, $\underline{Z}_2$ is classified (correctly) as π_2.

B. Categorial Classification

General Description: 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 (α_i, β_j, γ_k . . .). 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

Step A: 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.

A_1. 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

Division Points for Conversion of Measured Variables to Categorical

$$\bar{X} = \text{mean of pooled samples} = \sum_{n=1}^{N_1} X_n^{(1)} + \sum_{n=1}^{N_2} X_n^{(2)} \Big/ N_1 + N_2$$

$$S^2 = \frac{\sum_{n=1}^{N_1} (X_n^{(1)} - \bar{X}^{(1)})^2 + \sum_{n=1}^{N_2} (X_n^{(2)} - \bar{X}^{(2)})^2}{N_1 + N_2}$$

Number of Categories	Category	Range
2	1	$X \leq \bar{X}$
	2	$\bar{X} > X$
3	1	$X \leq \bar{X} - 0.6S$
	2	$X - 0.6S < X \leq \bar{X} + 0.6S$
	3	$\bar{X} + 0.6S < X$
4	1	$X \leq \bar{X} - S$
	2	$\bar{X} - S < X \leq \bar{X}$
	3	$\bar{X} < X \leq \bar{X} + S$
	4	$\bar{X} + S < X$
5	1	$X \leq \bar{X} - 1.2S$
	2	$\bar{X} - 1.2S < X \leq \bar{X} - 0.4S$
	3	$\bar{X} - 0.4S < X \leq \bar{X} + 0.4S$
	4	$\bar{X} + 0.4S < X \leq \bar{X} + 1.2S$
	5	$\bar{X} + 1.2S < X$
6	1	$X \leq \bar{X} - 1.4S$
	2	$\bar{X} - 1.4S < X \leq \bar{X} - 0.7S$
	3	$\bar{X} - 0.7S < X \leq \bar{X}$
	4	$\bar{X} < X \leq \bar{X} + 0.7S$
	5	$\bar{X} + 0.7S < X \leq \bar{X} + 1.4S$
	6	$\bar{X} + 1.4S < X$

A_2: 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

B_1. 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.

B_2. Tabulation of the Samples by Product Classes

Each population is kept separate, thus

Population		Product Categorical Class $(\alpha_i, \beta_j, \gamma_k)$		
π_1		$n_1(\alpha_i, \beta_j, \gamma_k)$		
π_2		$n_2(\alpha_i, \beta_j, \gamma_k)$		

B_3. Estimation of the Density Function for Each Product Class.

$$f_1(\alpha_i, \beta_j, \gamma_k) = n_1(\alpha_i, \beta_j, \gamma_k)/N_1$$
$$f_2(\alpha_i, \beta_j, \gamma_k) = n_2(\alpha_i, \beta_j, \gamma_k)/N_2$$

B_4. Computation of the Likelihood Ratio R for Each Product Class

$$R(\alpha_i, \beta_j, \gamma_k) = \frac{f_2(\alpha_i, \beta_j, \gamma_k)}{f_1(\alpha_i, \beta_j, \gamma_k)}$$

Step C. The Determination of the Classification Rule

C_1. Determine the empirical estimation of the conditional distributions of R using the R's generated in Step B_4 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 $p(i/j)$.

C_2: Determination of the λ (the classification constant) according to the decision strategy to be used.

C_3: Statement of the Classification Rule

The rule is then given by:

If $R(\underline{Z}) > \lambda$, classify $\underline{Z}$ as belonging to π_2; if $R(\underline{Z}) \leq \lambda$ classify $\underline{Z}$ as belonging to π_1. This rule can be applied to the R's of Step B_4 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 C_3 when applied to each product class $(\alpha_i, \beta_j, \gamma_k)$ 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 B_2. That is, if class $(\alpha_i, \beta_j, \gamma_k)$ is assigned by the rule to population π_2, then the individuals making up the $n_j(\alpha_i, \beta_j, \gamma_k)$ individuals in the class from the other population would have been erroneously classified. We can thus complete the table.

		From Population π_1	From Population π_2
To Population	π_1	n_{11}	n_{12}
	π_2	n_{21}	n_{22}
Total		N_1	N_2

From this table we make the estimates

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

$$\hat{P}(1/2) = n_{12}/N_2$$

E. The Application of the Classification Rule

For the individual $\underline{Z}$ to be classified, those variables that were converted to categorical (Step A_1) or for which the number of categories has been reduced (Step A_2) must be first considered. Then the product class for the individual is recognized and reference to the tabulation evolved in Step C_3 determines the classification.

Illustration of Procedure

Step A: Selection of 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

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

$$\bar{x}_1 = \frac{\sum x_{1n}^{(1)} + \sum x_{1n}^{(2)}}{N_1 + N_2} = \frac{827 + 1039}{38} = 49.1$$

$$\bar{x}_2 = \frac{\sum x_{2n}^{(1)} + \sum x_{2n}^{(2)}}{N_1 + N_2} = \frac{844 + 1035}{38} = 49.4$$

$$\bar{x}_3 = \frac{\sum x_{3n}^{(1)} + \sum x_{3n}^{(2)}}{N_1 + N_2} = \frac{311 + 446}{38} = 19.9$$

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

Population 1

Obs	X_1	X_2	X_3
1	L_1	L_2	L_3
2	L_1	L_2	L_3
3	H_1	L_2	L_3
4	L_1	L_2	L_3
5	L_1	H_2	L_3
6	L_1	H_2	L_3
7	L_1	L_2	L_3
8	H_1	H_2	H_3
9	H_1	L_2	L_3
10	H_1	L_2	L_3
11	H_1	H_2	L_3
12	H_1	L_2	L_3
13	L_1	L_2	L_3
14	H_1	H_2	L_3
15	L_1	L_2	H_3
16	L_1	L_2	L_3
17	L_1	L_2	L_3
18	L_1	L_2	H_3
19	L_1	H_2	L_3

Population 2

Obs	X_1	X_2	X_3
1	L_1	L_2	L_3
2	H_1	L_2	H_3
3	H_1	H_2	H_3
4	H_1	L_2	H_3
5	H_1	H_2	H_3
6	L_1	H_2	L_3
7	H_1	H_2	H_3
8	H_1	H_2	H_3
9	L_1	L_2	H_3
10	H_1	H_2	H_3
11	H_1	H_2	H_3
12	H_1	L_2	H_3
13	L_1	H_2	L_3
14	H_1	H_2	H_3
15	H_1	L_2	H_3
16	L_1	H_2	L_3
17	H_1	H_2	L_3
18	H_1	H_2	H_3
19	L_1	H_2	H_3

A2. Reduction of Number of Categories.

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

B. Selection of the Estimation Procedure for the Density Functions

B_1 Identification of the Product Classes

We have the following eight classes:

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

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

B_2 Tabulation of the Two Samples by Product Class

Product Class	Population 1 N_1	Population 2 N_2
$L_1L_2L_3$	7	1
$L_1L_2H_3$	2	1
$L_1H_2L_3$	3	3
$H_1L_2L_3$	4	0
$L_1H_2H_3$	0	1
$H_1L_2H_3$	0	4
$H_1H_2L_3$	2	1
$H_1H_2H_3$	1	8
	19	19

B_3 Estimation of the Density Functions for Each Product Class

Product Class	Population 1	Population 2
$L_1L_2L_3$	$f_1 = 7/19 = .368$	$f_2 = 1/19 = .053$
$L_1L_2H_3$	$f_1 = 2/19 = .105$	$f_2 = 1/19 = .053$
$L_1H_2L_3$	$f_1 = 3/19 = .158$	$f_2 = 3/19 = .158$
$H_1L_2L_3$	$f_1 = 4/19 = .211$	$f_2 = 0/19 = .000$
$L_1H_2H_3$	$f_1 = 0/19 = .000$	$f_2 = 1/19 = .053$
$H_1L_2H_3$	$f_1 = 0/19 = .000$	$f_2 = 4/19 = .211$
$H_1H_2L_3$	$f_1 = 2/19 = .105$	$f_2 = 1/19 = .053$
$H_1H_2H_3$	$f_1 = 1/19 = .053$	$f_2 = 8/19 = .421$

B_4 Computation of the Likelihood Ratio

Product Class	$R = f_2/f_1$
$L_1L_2L_3$	.053/.368 = .144
$L_1L_2H_3$	.053/.105 = .505
$L_1H_2L_3$	.158/.158 = 1.000
$H_1L_2L_3$	.000/.211 = .000
$L_1H_2H_3$	.053/.000 = ∞
$H_1L_2H_3$	.211/.000 = ∞
$H_1H_2L_3$	.053/.105 = .505
$H_1H_2H_3$	.421/.053 = 7.94

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	f_1	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
$L_1H_2H_3$	∞	0.000	0.00	0.053	0.79
$H_1H_2H_3$	7.94	0.053	0.05	0.421	0.74
$L_1H_2L_3$	1.00	0.158	0.21	0.158	0.32
$H_1H_2L_3$	.505	0.105	0.32	0.053	0.16
$L_1L_2H_3$	.505	0.105	0.43	0.053	0.11
$L_1L_2L_3$	.144	0.368	0.79	0.053	0.05
$H_1L_2L_3$	0	0.211	1.00	0.000	0.00

C_2 Use values of p (2/1) and p (1/2) of Step C_1 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 $L_1H_2L_3$ should be in π_2 for the Classification rule. The other four product classes are in π_1.

C_3 Statement of the Classification Rule:

If $\underline{Z}$ belongs 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 Population

		π_1	π_2
To Population	π_1	15	3
	π_2	4	16
Total		19	19

Thus $\hat{p}$ (2/1) = 4/19 = .211

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

E. The Application of the Classification Rule

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

Applying Step A_1 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," Biometrics, Vol. 3, No. 1 (1947).

Applying Step A to make categorical, we have

$\underline{Z}_2$: H_1, H_2, H_3

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 $f_1(\underline{x})$ and $f_2(\underline{x})$. The procedure is based upon the density ratio approach to classification. The best estimate for the parameters occurring in $f_1(\underline{x})$ and $f_2(\underline{x})$ are computed from the two available samples and the rule then involves

If $R(\underline{z}) = \dfrac{f_2(\underline{z})}{f_1(\underline{z})} \geq \lambda$, classify $\underline{z}$ 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

B_1 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:

$$f\,(x_1, x_2, ---, x_p, --x_p) = K \exp\left(\frac{-\sum_{p,q=1}^{P} C^{-pq}(X_p - \mu_p)(X_q - \mu_q)}{2\,|A_{pq}|} \right)$$

where

$$K = \frac{1}{(2\pi)^{P/2}\,|A_{pq}|^{1/2}}$$

and

$$(A_{pq}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1P} \\ \sigma_{21} & \cdot & & \\ \vdots & & \cdot & \\ \vdots & & & \cdot \\ \sigma_{P1} & & & \sigma_{PP} \end{pmatrix}$$

and

$$\sigma_{pq} = E\,(x_p - \mu_p)\,(x_q - \mu_q)$$

the population covariance

and

$$C^{pq} = \text{cofactor of } (A_{pq})$$
$$= p,q^{\text{th}} \text{ term of } (A_{pq})^{-1}$$

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 $x_1, x_2, ---, x_p, ..., x_p$ 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

$$u_1 = t_1$$

$$u_2 = t_2 - (t_2 \cdot z_1)\, z_1$$

$$- \; - \; - \; -$$

$$u_p = t_p - \sum_{p=1}^{p-1} (t_p \cdot z_p)\; z_p$$

and

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

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.

These equations are

$$Z_1 = \frac{X_1 - \bar{X}_1}{S_1}$$

$$Z_2 = \frac{\dfrac{X_2 - \bar{X}_2}{S_2} - r_{12} Z_1}{\sqrt{1 - r_{12}^2}}$$

$$Z_3 = \frac{\dfrac{X_3 - \bar{X}_3}{S_3} - \dfrac{(r_{23} - r_{12} r_{13})}{\sqrt{1 - r_{12}^2}} Z_2 - r_{13} Z_1}{\sqrt{1 - r_{13}^2 - \dfrac{(r_{23} - r_{12} r_{13})^2}{1 - r_{12}^2}}}$$

$$Z_4 = \left\{ \frac{X_4 - \bar{X}_4}{S_4} \cdot D_3 \sqrt{1 - r_{12}^2} - [(r_{34} - r_{13} r_{14})(1 - r_{12}^2) - (r_{23} - r_{12} r_{13})(r_{24} - r_{12} r_{14})] Z_3 - (r_{24} - r_{12} r_{14}) D_3 Z_2 - r_4 \sqrt{1 - r_{12}^2} D_3 Z_1 \right\} \cdot \frac{1}{D_4}$$

where

$$D_3 = \sqrt{1 - r_{13}^2 - r_{12}^2 - r_{23}^2 + 2 r_{12} r_{13} r_{23}}$$

and

$$D_4^2 = (1 - r_{12}^2)^2 (1 - r_{13}^2 - r_{14}^2 - r_{34}^2 + 2 r_{13} \cdot r_{14} \cdot r_{34}) - (1 - r_{12}^2) [(1 - r_{13}^2)(r_{24} - r_{12} r_{14})^2 + (1 - r_{24}^2)(r_{23} - r_{12} r_{13})^2] + 2 (r_{23} - r_{12} r_{13})(r_{24} - r_{12} r_{14})$$

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

B_2 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.

$$\bar{X}_p^{(k)} = \sum_{n=1}^{N_R} X_{pn} \Big/ N_R$$

$$s_p^{(R)} = \frac{1}{N_R}\sqrt{N_R \sum_{i=1}^{N_R} X_{pi}^{\,2} - \left(\sum_{i=1}^{N_R} X_{pi}\right)^2}$$

$$r_{pq}^{(R)} = \frac{\left[N_R \sum_{n=1}^{N_R} X_{pn} X_{qn} - \sum_{i=1}^{N_R} X_{pn} X_{qn}\right]}{\sqrt{\left[N_R \sum_{n=1}^{N_R} X_{pn}^2 - \left(\sum_{i=1}^{N_R} X_{pn}\right)^2\right]\left[N_R \sum_{n=1}^{N_R} X_{qn}^2 - \left(\sum_{n=1}^{N_R} X_{qn}\right)^2\right]}}$$

B_3 Determination of the coefficients in the transformation equations using estimates of step B_2.

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

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

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

B_4 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.

$$f_1\left(Z_p^{(1)}\right) = \frac{1}{\sqrt{2\pi}}\, e^{-Z_p^{(1)\,2}/2}$$

$$f_2\left(Z_p^{(2)}\right) = \frac{1}{\sqrt{2\pi}}\, e^{-Z_p^{(2)\,2}/2}$$

B_5 Evaluation of the multivariate density function.

We have the simple products

$$f_1(x_1, x_2, \text{--} x_p) = f_1(Z_1^{(1)}) \cdot f_1(Z_2^{(1)}) \cdots f_1(Z_p^{(1)})$$

and

$$f_2(x_1, x_2, \text{--} x_p) = f_2(Z_1^{(2)}) \cdot f_2(Z_2^{(2)}) \cdots f_2(Z_p^{(2)})$$

B_6 Computation of the Likelihood Ratio:

$$R(\underline{x}) = f_2(x_1, x_2, \text{--} x_p) \Big/ f_1(x_1, x_2, \text{--} x_p)$$

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_6 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.

C_3 Statement of the Classification Rule.

The rule then is given by

If $R(\underline{Z}) = \dfrac{f_2(Z_1, Z_2, \text{--} Z_p)}{f_1(Z_1, Z_2, \text{--} Z_p)} > \lambda$, 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.

		From Population	
		π_1	π_2
To Population	π_1	n_{11}	n_{12}
	π_2	n_{21}	n_{22}
	Total	N_1	N_2

From this table we make the estimates

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

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

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:

X_1 = Math Placement Grade

X_2 = English Placement Grade

X_3 = General Aptitude Test Score

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

B_1 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.

B_2 Estimation of Means Standard Deviation and Correlation Coefficents for Each Population from the Available Sample.

Computation of Sample Summations

Population 1		Population 2	
N_1	$= 19$	N_2	$= 19$
$\sum X_1$	$= 827$	$\sum X_1$	$= 1039$
$\sum X_2$	$= 844$	$\sum X_2$	$= 1035$
$\sum X_3$	$= 311$	$\sum X_3$	$= 446$
$\sum X_1^2$	$= 39033$	$\sum X_1^2$	$= 59547$
$\sum X_2^2$	$= 39388$	$\sum X_2^2$	$= 58599$
$\sum X_3^2$	$= 5641$	$\sum X_3^2$	$= 11408$
$\sum X_1X_2$	$= 38073$	$\sum X_1X_2$	$= 57661$
$\sum X_1X_3$	$= 14007$	$\sum X_1X_3$	$= 25250$
$\sum X_2X_3$	$= 14152$	$\sum X_2X_3$	$= 25003$

Population 1

$\bar{X}_1^{(1)} = \frac{827}{19} = 43.53$,

$\bar{X}_2^{(1)} = \frac{844}{19} = 44.42$,

$\bar{X}_3^{(1)} = \frac{311}{19} = 16.37$,

$s_1^{(1)} = \frac{1}{19}\sqrt{57698} = 12.64$,

$s_2^{(1)} = \frac{1}{19}\sqrt{36036} = 9.99$,

$s_3^{(1)} = \frac{1}{19}\sqrt{10458} = 4.05$

Population 2

$\bar{X}_1^{(2)} = \frac{1039}{19} = 54.68$

$\bar{X}_2^{(2)} = \frac{1035}{19} = 54.47$

$\bar{X}_3^{(2)} = \frac{446}{19} = 23.47$

$s_1^{(2)} = \frac{1}{19}\sqrt{53772} = 12.20$

$s_2^{(2)} = \frac{1}{19}\sqrt{44056} = 11.05$

$s_3^{(2)} = \frac{1}{19}\sqrt{17835} = 7.03$

$r_{12}^{(1)} = +0.557$ $r_{12}^{(2)} = +0.415$

$r_{13}^{(1)} = +0.483$ $r_{13}^{(2)} = +0.528$

$r_{23}^{(1)} = +0.438$ $r_{23}^{(2)} = +0.480$

B_3 Determination of the Coefficients in the Transformation Equations.

Population 1

$$t_1 = \frac{X_1 - 45.53}{12.64}, \quad t_2 = \frac{X_2 - 44.42}{9.99}, \quad t_3 = \frac{X_3 - 16.37}{4.05}$$

$$Z_1 = t_1$$

$$Z_2 = \frac{t_2 - 0.557\, Z_1}{0.831}$$

$$Z_3 = \frac{t_3 - \frac{0.169}{0.831}\, Z_2 - 0.483\, Z_1}{\sqrt{0.757 - \frac{0.0285}{0.690}}} = \frac{t_3 - 0.203\, Z_2 - 0.483\, Z_1}{0.850}$$

Population 2

$$t_1 = \frac{X_1 - 54.68}{12.20}, \quad t_2 = \frac{X_2 - 54.47}{11.05}, \quad t_3 = \frac{X_3 - 23.47}{7.03}$$

$$Z_1 = t_1$$

$$Z_2 = \frac{t_2 - 0.415\, Z_1}{0.910}$$

$$Z_3 = \frac{t_3 - \frac{0.261}{0.910}\, Z_2 - 0.528\, Z_1}{\sqrt{0.721 - \frac{0.068}{0.828}}} = \frac{t_3 - 0.287\, Z_2 - 0.528\, Z_1}{0.799}$$

B_4 Determination of the individual densities for each Z_p.

B_5 Evaluation of the multivariate density functions.

The following computational form* accomplishes the requirements of steps B_4 and B_5.

Population 1

Step	Operation
(1)	X_1
(2)	X_2
(3)	X_3
(4)	(1) - 43.53
(5)	(2) - 44.42
(6)	(3) - 16.37
(7)	(4) / 12.64
(8)	(5) / 9.99
(9)	(6) / 4.05
(10)	(7)
(11)	(8) - 0.557(7)
(12)	(11) / 0.831
(13)	(9) - 0.203 (12) - 0.483 (10)
(14)	(13) / 0.850
(15)	f_1 (10)
(16)	f_2 (12)
(17)	f_3 (14)
(18)	(15)(16)(17)

Population 2

Step	Operation
(1)	X_1
(2)	X_2
(3)	X_3
(4)	(1) - 54.68
(5)	(2) - 54.47
(6)	(3) - 23.47
(7)	(4) / 12.20
(8)	(5) / 11.05
(9)	(6) / 7.03
(10)	(7)
(11)	(8) - 0.415(7)
(12)	(11) / 0.910
(13)	(9) - 0.287 (12) - 0.528 (10)
(14)	(13) / 0.799
(15)	f_1 (10)
(16)	f_2 (12)
(17)	f_3 (14)
(18)	(15)(16)(17)

B_6 Compulation of the Likelihood Ratio R.

R is the quotient of step 18 for Population 2 by Population 1.

*The convention used is that (ν) means to use the value of step ν. f(Z) is the normal ordinate from table look up. For example: Table II, page 315: Hoel, Introduction to Mathematical Statistics, Wiley & Co.

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)	36	36
(3)	12	12
(4)	-9.53	-20.68
(5)	-8.42	-18.47
(6)	-4.37	-11.47
(7)	-0.754	-1.695
(8)	-0.843	-1.671
(9)	-1.079	-1.632
(10)	-0.754	-1.695
(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

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

Step C: The Determination of the Classification Rule.

C_1 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

Population 1		Population 2	
Observation	R	Observation	R
1	0.094	1	0.470
2	1.378	2	1.465
3	1.576	3	160.213
4	0.445	4	1.357
5	0.941	5	100.650
6	0.859	6	0.866
7	0.055	7	515.877
8	179.553	8	57.456
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	26.727
19	0.985	19	3×10^6

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_1	Ordered R	cum f_2
.055	1.00	.470	.05
.094	0.95	.790	.11
.134	0.89	.866	.16
.410	0.84	1.170	.21
.445	0.79	1.357	.26
.557	0.74	1.465	.32
.750	0.68	1.606	.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	26.727	.68
1.378	0.32	57.456	.74
1.536	0.26	100.650	.79
1.575	0.21	160.213	.84
1.824	0.16	515.877	.89
12.666	0.11	4192.712	.95
179.553	0.05	3×10^6	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).

Figure 1: Emperical Estimation of the Distribution of R given π_1.

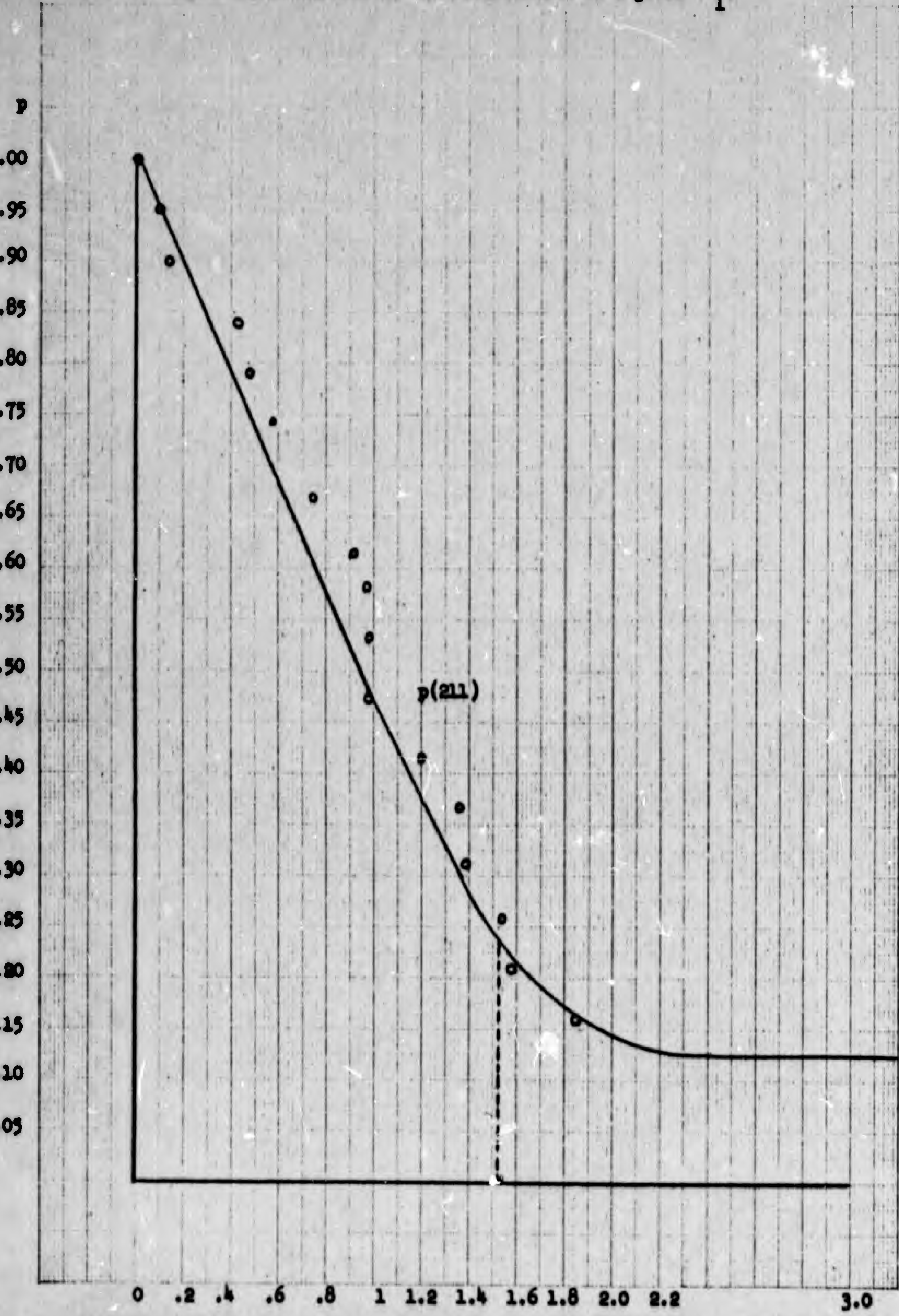
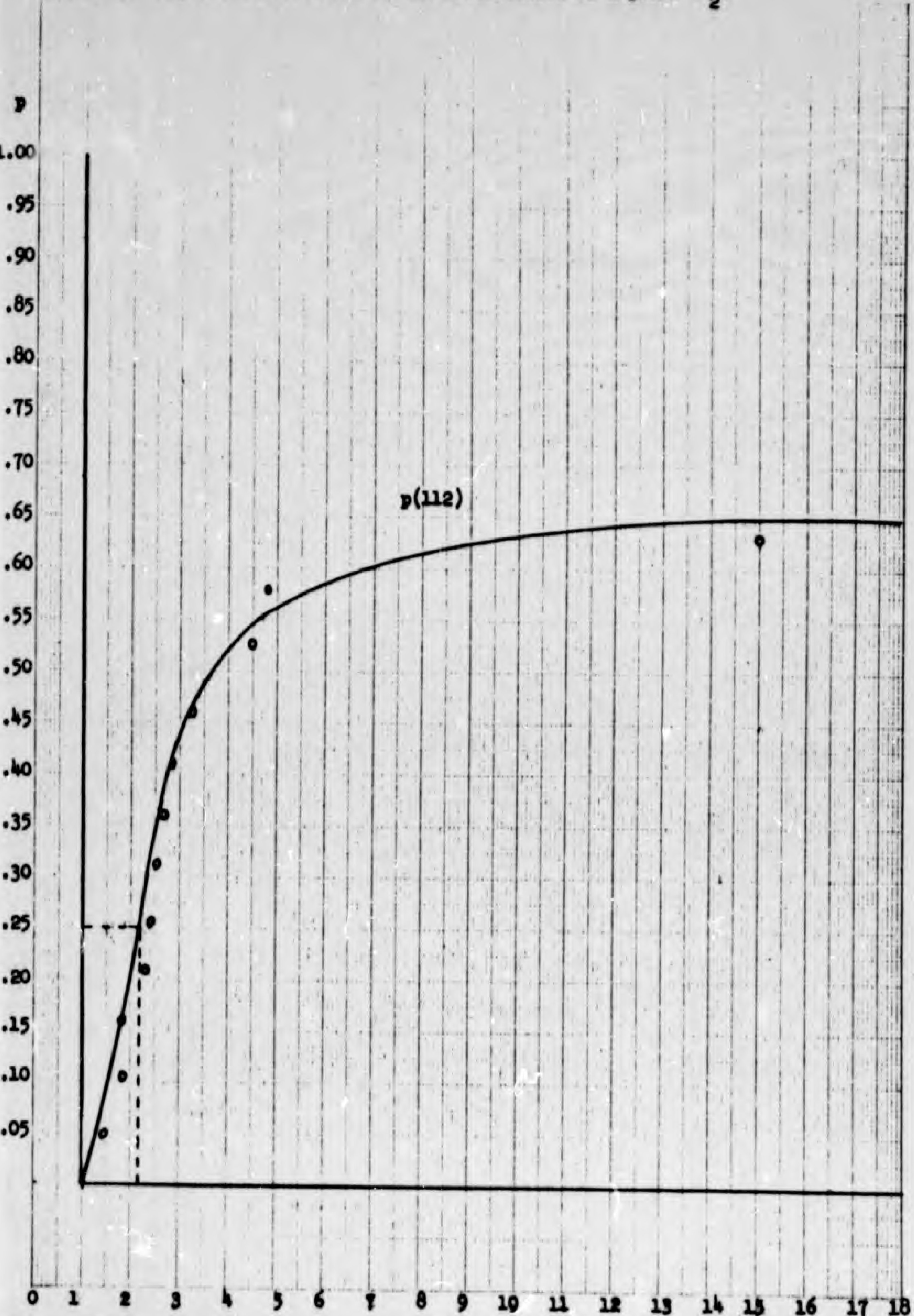


Figure 2: Emperical Estimation of the Distribution of R given π_2.



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

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

C_3: Statement of the Classification Rule.

If $R(Z_1, Z_2, Z_3) \geq 1.50$ / classify $\underline{Z}$ as belonging to π_2, if $R(Z_1, Z_2, Z_3) \leq 1.50$ 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 Population π_1	From Population π_2
To Population	π_1	14	6
	π_2	5	13
	Total	19	19

We obtain,

$$p(2/1) = 5/19 = .263$$

$$p(1/2) = 6/19 = .316$$

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

The computation form in step B_5 is applied to each observation.

Step	$\underline{Z}_1$		$\underline{Z}_2$	
	Population 1	2	1	2
(1)	34	34	55	55
(2)	36	36	62	62
(3)	12	12	21	21
(4)	-9.53	-20.68	+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	+0.026
(8)	-0.843	-1.671	+ 1.760	+0.681
(9)	-1.079	-1.632	+1.143	-0.351
(10)	-0.754	-1.695	+0.907	+0.026
(11)	-0.423	-0.968	+1.255	+0.670
(12)	-0.509	-1.064	+1.510	+0.736
(13)	-0.611	-0.432	+0.398	-0.576
(14)	-0.719	-0.541	+0.468	-0.721
(15)	0.3011	0.0941	0.2637	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	0.0373

$R_{(\underline{Z})}$: $.0074/0.0325 = 0.23$ $0.0373/0.0120 = 3.11$

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

$$W(Z) = \sum_{q=1}^{p} \sum_{p=1}^{p} S^{pq} \left(\bar{X}_q^{(2)} - \bar{X}_q^{(1)}\right) Z_p$$

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 $W(Z)$ 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, Annals of Mathematics Statistics, 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:

B_1: The sample moments computation.

$$\bar{X}_q^{(1)} = \sum_{n=1}^{N_1} X_{qn}^{(1)}/N_1$$

$$\bar{X}_q^{(2)} = \sum_{n=1}^{N_2} X_{qn}^{(2)}/N_1$$

$$S_{pq}^{(1)} = \frac{N_1 \sum_{n=1}^{N_1} X_{pn} X_{qn} - \sum_{n=1}^{N_1} X_{pn} \sum X_{qn}}{N_1^{\;2}}$$

$$S_{pq}^{(2)} = \frac{N_2 \sum_{n=1}^{N_2} X_{pn} X_{qn} - \sum_{n=1}^{N_2} X_{pn} \sum_{n=1}^{N_2} X_{qn}}{N_2^{\;2}}$$

$$S_{pq} = \frac{N_1 \; S_{pq}^{(1)} + N_2 \; S_{pq}^{(2)}}{N_1 + N_2}$$

B_2: Inversion of the Covariance Matrix.

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

B_3: Estimation of the Coefficients of the Wald Statistic.

In general we have

$$W(Z) = \sum_{p=1}^{P} \left(\sum_{q=1}^{P} S^{pq} \;(\bar{X}_q^{(2)} - \bar{X}_q^{(1)}) \right) Z_p$$

For P = 3 this can be expanded to read

$$W(\underline{Z}) = [S^{11} (\bar{X}_1^{(2)} - \bar{X}_1^{(1)}) + S^{12} (\bar{X}_2^{(2)} - \bar{X}_2^{(1)}) + S^{13} (\bar{X}_3^{(2)} - \bar{X}_3^{(1)})] Z_1$$
$$+ [S^{21} (\bar{X}_1^{(2)} - \bar{X}_1^{(1)}) + S^{22} (\bar{X}_2^{(2)} - \bar{X}_2^{(1)}) + S^{23} (\bar{X}_3^{(2)} - \bar{X}_3^{(1)})] Z_2$$
$$+ [S^{31} (\bar{X}_1^{(2)} - \bar{X}_1^{(1)}) + S^{32} (\bar{X}_2^{(2)} - \bar{X}_2^{(1)}) + S^{33} (\bar{X}_3^{(2)} - \bar{X}_3^{(1)})] Z_3$$

Step C: The Determination of the Classification Rule.

Empirical Estimation:

C_1: 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.

C_3: 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

and $$\alpha_2 = \sum_{p=1}^{P} \sum_{q=1}^{P} S^{pq} (\bar{X}_q^{(2)} - \bar{X}_q^{(1)}) \bar{X}_p^{(2)}$$

if $\underline{Z}$ belongs to π_2 and following the assumption, for either π_1 or π_2

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

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

We thus have for any λ (Classification Rule)

$$p_\lambda(2/1) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\lambda - a_1}{\sigma_W}}^{\infty} e^{-t^2/2} \, dt$$

and

$$p_\lambda(1/2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\lambda - a_2}{\sigma_W}} e^{-t^2/2} \, dt$$

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 Population π_1	From Population π_2
To Population	π_1	n_{11}	n_{12}
	π_2	n_{21}	n_{22}
	Total	N_1	N_2

From this table we can make the estimates

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

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

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

x_3 = General Aptitude Test Score

Step B: The Selection of the Estimation Procedure.

B_1: The Sample Moment Compulations

Summations

Population 1		Population 2	
N_1	= 19	N_1	= 19
$\sum x_1$	= 827	$\sum x_1$	= 1039
$\sum x_2$	= 844	$\sum x_2$	= 1035
$\sum x_3$	= 311	$\sum x_3$	= 446
$\sum x_1^2$	= 39033	$\sum x_1^2$	= 59647
$\sum x_2^2$	= 39388	$\sum x_2^2$	= 58699
$\sum x_3^2$	= 5641	$\sum x_3^2$	= 11408
$\sum x_1 x_2$	= 38073	$\sum x_1 x_2$	= 57661
$\sum x_1 x_3$	= 14007	$\sum x_1 x_3$	= 25250
$\sum x_2 x_3$	= 14152	$\sum x_2 x_3$	= 25003

Means

Population 1

$$\bar{x}_1^{(1)} = \frac{827}{19} = 43.53$$

$$\bar{x}_2^{(1)} = \frac{844}{19} = 44.42$$

$$\bar{x}_3^{(1)} = \frac{311}{19} = 16.37$$

Population 2

$$\bar{x}_1^{(2)} = \frac{1039}{19} = 54.68$$

$$\bar{x}_2^{(2)} = \frac{1035}{19} = 54.47$$

$$\bar{x}_3^{(2)} = \frac{446}{19} = 23.47$$

Pooled Variances and Covariances

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

$$S_{22} = \frac{\frac{(19)(36036)}{(19)^2} + \frac{(19)(44055)}{(19)^2}}{19 + 19} = 110.93$$

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

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

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

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

B_2: Inversion of the Covariance Matrix*

Evaluation of Covariance Determinant

$$\begin{vmatrix} 154.39 & 63.15 & 35.03 \\ 63.15 & 110.93 & 27.49 \\ 35.03 & 27.49 & 39.19 \end{vmatrix} = 383729$$

Determination of Cofactor Matrix

$$\begin{pmatrix} 3592 & -1512 & -2150 \\ -1512 & 4823 & -2032 \\ -2150 & -2032 & 13139 \end{pmatrix}$$

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.

Division of terms by covariance determinant

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

B_3: Estimation of the Coefficient of the Wald Statistic.

Computation of Mean Differences

$$\bar{x}_1^{(2)} - \bar{x}_1^{(1)} = +11.15$$

$$\bar{x}_2^{(2)} - \bar{x}_2^{(1)} = +10.05$$

$$\bar{x}_3^{(2)} - \bar{x}_3^{(1)} = +7.10$$

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(Z) For Sample Observations

Population 1 Obs.	W(Z)	Population 2 Obs.	W(Z)
1	4.029	1	4.576
2	3.708	2	6.173
3	4.749	3	8.425
4	4.546	4	5.416
5	5.025	5	7.908
6	5.006	6	5.733
7	3.639	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	6.888
13	4.974	13	4.855
14	6.135	14	6.773
15	5.573	15	5.641
16	5.103	16	5.996
17	3.316	17	6.086
18	6.254	18	7.264
19	5.668	19	8.807

Generation of the Cumulative Frequency Distributions for W

Population 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	.84
4.50 - 4.99	3	14	.74
5.00 - 5.49	4	11	.58
5.50 - 5.99	3	7	.37
6.00 - 6.99	3	4	.21
7.00 - 7.99	0	1	.05
8.00 - 8.99	1	1	.05

Population 2

Interval	f	cum f	cum p
4.50 - 5.49	4	4	.21
5.50 - 5.49	5	9	.47
6.50 - 6.99	2	11	.58
7.00 - 7.49	2	13	.68
7.50 - 7.99	2	15	.79
8.00 - 8.49	1	16	.84
8.50 - 8.99	2	18	.95
9.00 - 9.49	0	19	.95
9.50 - 9.99	1	19	1.00

Figure 1: Experical Estimation of the Distribution of W given π_1.

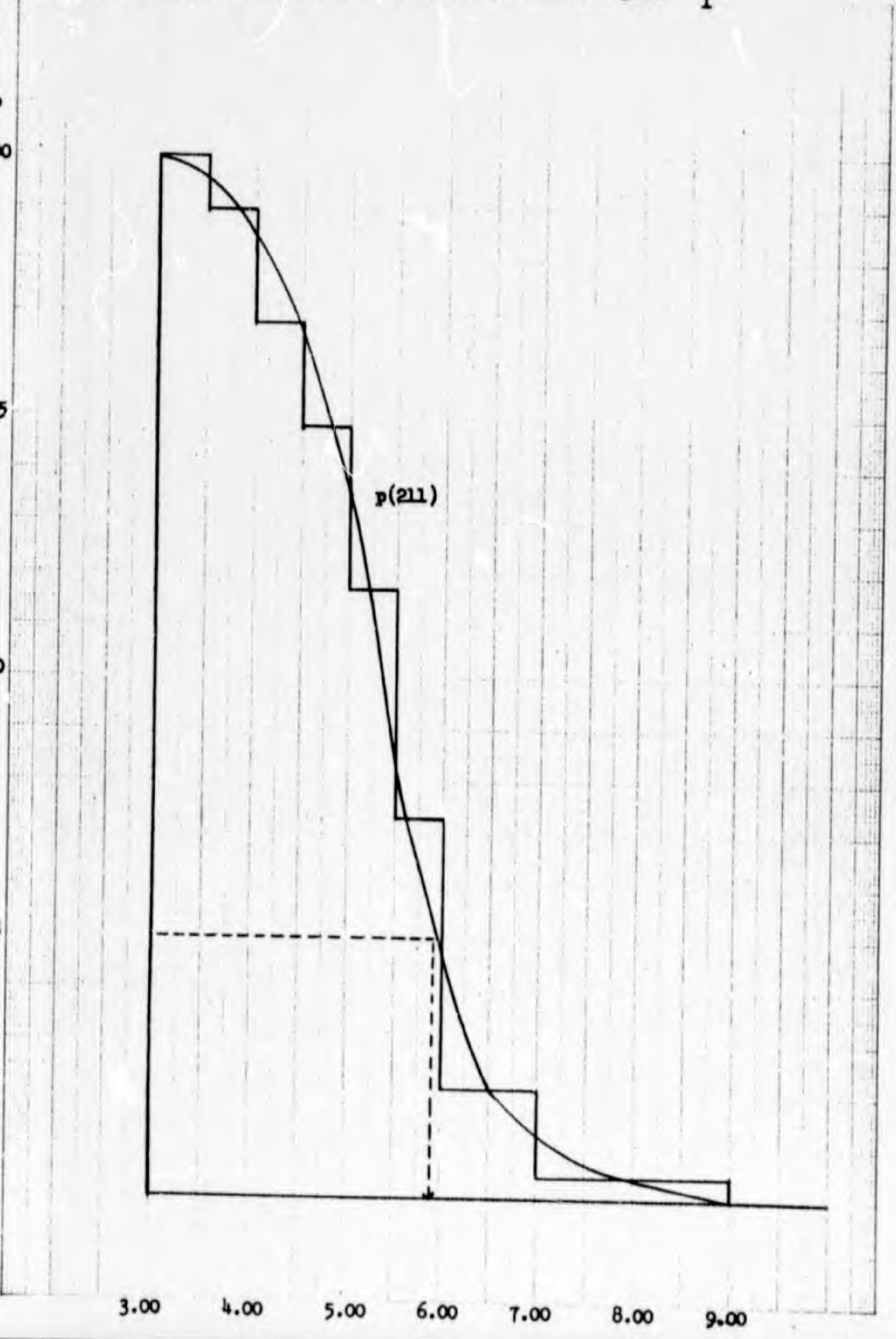
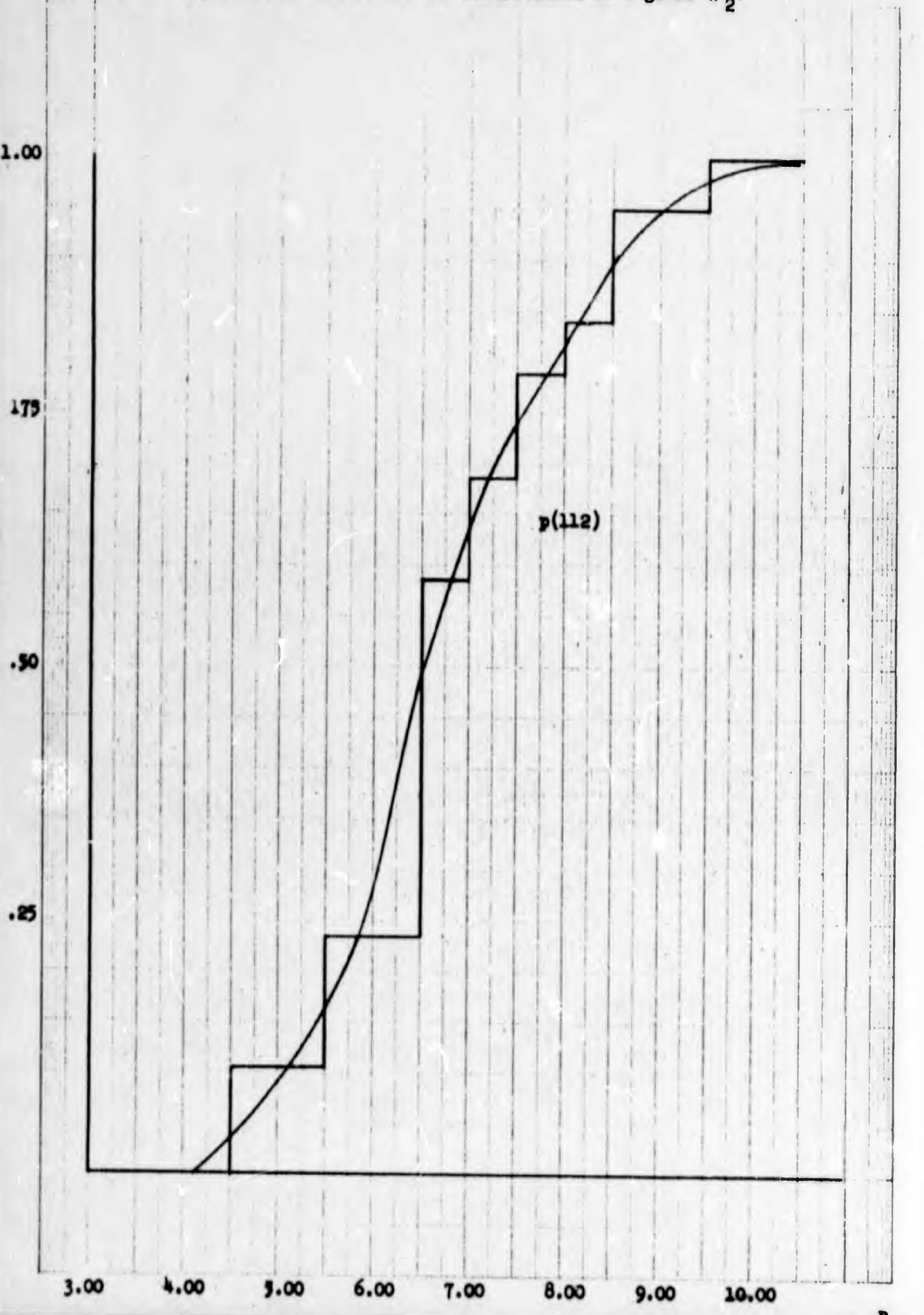


Figure 2: Emperical Estimation of the Distribution of W given π_2.



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

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

C_3: 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 $W(Z)$

$$\alpha_1 = .0250(43.53) + .0448(44.42) + .1274(16.37)$$
$$= 5.16$$
$$\alpha_2 = .0250(56.68) + .0448(54.47) + .1274(23.47)$$
$$= 6.80$$
$$\sigma^2 = .0250(11.15) + .0448(10.05) + .1274(7.10)$$
$$= 1.63$$

C_2: Determination of λ using the decision strategy.

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

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{\lambda - 5.16}{1.63}}^{\infty} e^{-t^2/2} \, dt = 0.25$$

From the tabulation of areas under the normal curve we have

$$\frac{\lambda - 5.16}{1.63} = 0.67$$

and $\lambda^{*} = 6.02$ (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.

		From Population π_1	π_2
To	π_1	14	6
Population	π_2	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

$$p(2/1) = \frac{1}{\sqrt{2\pi}} \int_{\frac{6.02 - 5.17}{1.28}}^{\infty} e^{-t^2/2} \, dt = 0.25 \text{ (as designed)}$$

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

$$W(\underline{Z}_1) = .0250(34) + .0448(36) + .1274(12) = 3.992$$

Applying the Classification Rule.

Since $W(\underline{Z}_1) \leq 5.90$, $\underline{Z}_1$ is classified as belonging to π_1.

For $\underline{Z}_2$: $Z_1 = 55,\ Z_2 = 62,\ Z_3 = 21$

$$W(\underline{Z}_2) = .0250(55) + .0448(62) + .1274(21) = 6.828$$

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 Z's.

$$P(\underline{Z}) = \sum_{q=1}^{p} \sum_{p=1}^{p} [(S_{(1)}^{pq} - S_{(2)}^{pq}) Z_p Z_q - 2(S_{(1)}^{pq}\, \bar{X}_p^{(1)} - S_{(2)}^{pq}\, \bar{X}_p^{(2)})\, Z_q$$

$$+ \quad (S_{(1)}^{pq}\, \bar{X}_p^{(1)}\, \bar{X}_q^{(1)} - S_{(2)}^{pq}\, \bar{X}_p^{(2)}\, \bar{X}_q^{(2)})]$$

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.

B_1: The sample moment computation

$$\bar{X}_p^{(1)} = \sum_{n=1}^{N_1} X_{pn}^{(1)} / N_1$$

$$\bar{X}_p^{(2)} = \sum_{n=1}^{N_2} X_{pn}^{(2)} / N_2$$

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

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

B_2: Inversion of the covariance matrix.

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

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

B_3: **Estimation of the Coefficients of the Purdue Statistic.**

In general we have

$$P(\underline{Z}) = \sum_{q=1}^{P} \sum_{p=1}^{P} (S_{(1)}^{pq} - S_{(2)}^{pq}) \; Z_p Z_q$$

$$2 \sum_{q=1}^{P} \sum_{p=1}^{P} (S_{(1)}^{pq} \bar{X}_p^{(1)} - S_{(2)}^{pq} \bar{X}_p^{(2)}) Z_q$$

$$+ \sum \sum (S_{(1)}^{pq} \bar{X}_p^{(1)} \bar{X}_q^{(1)} - S_{(2)}^{pq} \bar{X}_p^{(2)} \bar{X}_q^{(2)})$$

For $P = 3$, this expands to the form

$$P(\underline{Z}) = (S_{(1)}^{11} - S_{(2)}^{11}) Z_1^2 + (S_{(1)}^{22} - S_{(2)}^{22}) Z_2^2 + (S_{(1)}^{33} - S_{(2)}^{33}) Z_3^2$$

$$+ 2 (S_{(1)}^{12} - S_{(2)}^{12}) Z_1 Z_2 + 2 (S_{(1)}^{13} - S_{(2)}^{13}) Z_1 Z_3 + 2 (S_{(1)}^{23} - S_{(2)}^{23}) Z_2 Z_3$$

$$- 2[(S_{(1)}^{11} \bar{X}_1^{(1)} - S_{(2)}^{11} \bar{X}_1^{(2)}) + (S_{(1)}^{12} \bar{X}_2^{(1)} - S_{(2)}^{12} \bar{X}_2^{(2)})$$

$$+ (S_{(1)}^{13} \bar{X}_3^{(1)} - S_{(2)}^{13} \bar{X}_3^{(2)})] Z_1$$

$$- 2[(S_{(1)}^{12}\bar{X}_1^{(1)} - S_{(2)}^{12}\bar{X}_1^{(2)}) + (S_{(1)}^{22}\bar{X}_2^{(1)} - S_{(2)}^{22}\bar{X}_2^{(2)}) + (S_{(1)}^{23}\bar{X}_3^{(1)} - S_{(2)}^{23}\bar{X}_3^{(2)})]\,Z_2$$

$$- 2[(S_{(1)}^{13}\bar{X}_1^{(1)} - S_{(2)}^{13}\bar{Y}_1^{(2)}) + (S_{(1)}^{23}\bar{X}_2^{(1)} - S_{(2)}^{23}\bar{X}_2^{(2)}) + (S_{(1)}^{33}\bar{X}_3^{(1)} - S_{(2)}^{33}\bar{X}_3^{(2)})]\,Z_3$$

$$+ (S_{(1)}^{11}\bar{X}_1^{(1)^2} - S_{(2)}^{11}\bar{X}_1^{(2)^2}) + (S_{(1)}^{22}\bar{X}_2^{(1)^2} - S_{(2)}^{22}X_{(2)}^{22}) + (S_{(1)}^{33}\bar{X}_3^{(1)^2} - S_{(2)}^{33}\bar{X}_3^{(2)^2})$$

$$+ 2(S_{(1)}^{12}\bar{X}_1^{(1)}\bar{X}_2^{(1)} - S_{(2)}^{12}\bar{X}_1^{(2)}\bar{X}_2^{(2)}) + 2(S_{(1)}^{13}\bar{X}_1^{(1)}\bar{X}_3^{(1)} - S_{(2)}^{13}\bar{X}_1^{(2)}\bar{X}_3^{(2)}) + 2(S_{(1)}^{23}\bar{X}_2^{(1)}\bar{X}_3^{(1)} - S_{(2)}^{23}\bar{X}_2^{(2)}\bar{X}_3^{(2)})$$

Step C: The Determination of the Classification Rule.

C_1: 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 $p(i/j)$.

C_2: Determination of the λ (the classification constant) according to the decision strategy to be used.

C_3: Statement of the Classification Rule.

The rule is then given by

"If $P(\underline{Z}) > \lambda$ 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 C_3 to each observation and obtain the totals for the following table.

From Population

		π_1	π_2
To	π_1	n_{11}	n_{12}
Population	π_2	n_{21}	n_{22}
	Total	N_1	N_2

From this table we can make the estimates

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

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

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

The value of the statistic $P(\underline{Z})$ 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

B_1: The sample moment computation.

The sample summations

Population 1		Population 2	
N_1	19	N_2	19
$\sum X_1$	827	$\sum X_1$	1039
$\sum X_2$	844	$\sum X_2$	1035
$\sum X_3$	311	$\sum X_3$	446
$\sum X_1^2$	39033	$\sum X_1^2$	59647
$\sum X_2^2$	39388	$\sum X_2^2$	58699
$\sum X_3^2$	5541	$\sum X_3^2$	11408
$\sum X_1X_2$	38073	$\sum X_1X_2$	57661
$\sum X_1X_3$	14007	$\sum X_1X_3$	25250
$\sum X_2X_3$	14152	$\sum X_2X_3$	25003

Estimation of Population Parameters

Population 1	Population 2
$\bar{X}_1^{(1)} = \frac{827}{19} = 43.53$	$\bar{X}_1^{(2)} = \frac{1039}{19} = 54.68$
$\bar{X}_2^{(1)} = \frac{844}{19} = 44.42$	$\bar{X}_2^{(2)} = \frac{1035}{19} = 54.47$
$\bar{X}_3^{(1)} = \frac{311}{19} = 16.37$	$\bar{X}_3^{(2)} = \frac{446}{19} = 23.47$
$s_{11}^{(1)} = \frac{57698}{361} = 159.83$	$s_{11}^{(2)} = \frac{53772}{361} = 148.95$
$s_{22}^{(1)} = \frac{36036}{361} = 99.82$	$s_{22}^{(2)} = \frac{44056}{361} = 122.04$
$s_{33}^{(1)} = \frac{10458}{361} = 28.97$	$s_{33}^{(2)} = \frac{17836}{361} = 49.41$
$s_{12}^{(1)} = \frac{35399}{361} = 70.36$	$s_{12}^{(2)} = \frac{20194}{361} = 55.94$
$s_{13}^{(1)} = \frac{8936}{361} = 24.75$	$s_{13}^{(2)} = \frac{16356}{361} = 45.31$

$s_{23}^{(1)} = \frac{5404}{351} = 17.74$ $\qquad$ $s_{23}^{(2)} = \frac{13447}{361} = 37.25$

B_2: Inversion* of the Covariance Matrices.

Evaluation of the Covariance Determinant

Population 1

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

Population 2

$$\begin{vmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{vmatrix} = 475155$$

Computation of the Cofactor Matrices

Population 1

$$\begin{pmatrix} 2577 & -1599 & -1222 \\ -1599 & 4018 & -1094 \\ -1222 & -1094 & 11004 \end{pmatrix}$$

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

Population 2

$$\begin{pmatrix} 46.42 & -1075 & -3446 \\ -10.75 & 5307 & -3014 \\ -34.45 & -3014 & 15049 \end{pmatrix}$$

*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

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

B_3: Estimation of the Coefficients.

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

$$(S_{(1)}^{pq} - S_{(2)}^{pq}) = \begin{pmatrix} .00958 & -.00594 & -.00454 \\ .00594 & .01493 & -.00407 \\ -.00454 & -.00407 & .04089 \end{pmatrix} - \begin{pmatrix} .00977 & -.00226 & -.00725 \\ -.00226 & .01117 & -.00634 \\ -.00725 & -.00634 & .03167 \end{pmatrix}$$

$$= \begin{pmatrix} -.00019 & -.00368 & .00271 \\ -.00368 & .00376 & .00227 \\ +.00271 & .00227 & .00922 \end{pmatrix}$$

Computation of linear term coefficients

$$\begin{aligned} b_1 = -2[& (.00958)(43.53) - (.00977)(54.68) \\ & + (-.00594)(44.42) - (-.00226)(54.47) \\ & + (-.00454)(16.37) - (-.00725)(23.47)] = -2(-.1621) \end{aligned}$$

similarly for b_2 and b_3.

Computation of the constant term

$$k = [(.00958)(43.53)^2 - (.00977)(54.68)^2] + \dots$$

$$+ [(-.00407)(54.47)(23.47) - (-.00634)(45.31)(37.25)]$$

$$= -8.333$$

Thus

$$P(\underline{Z}) = -.00019\, Z_1^2 + .00376\, Z_2^2 + .00922\, Z_3^2$$

$$- .00736\, Z_1 Z_2 + .00542\, Z_1 Z_3 + .00454\, Z_2 Z_3$$

$$+ .3242\, Z_1 - .0038\, Z_2 - .5788\, Z_3 - 8.333$$

Step C: Determination of the Classification Rule.

C_1: Determination of the values of the classification statistic for the sample observation from π_1 and π_2.

Thus

$$P(\underline{X}_1^{(1)}) = -.00019(25)^2 + .00376(35)^2 + .00922(15)^2$$

$$-.00736(25)(35) + .00542(25)(15) + .00454(35)(15)$$

$$+ .3242(25) - .0038(35) - .5788(15) - 8.333 = -4.922$$

Repeating for the 38 observations yields the results

Table P($\underline{Z}$) for Sample Observations

Population 1 Obs.	P($\underline{Z}$)	Population 2 Obs.	P($\underline{Z}$)
1	-4.922	1	-1.739
2	-2.264	2	0.637
3	-1.742	3	7.385
4	-1.791	4	0.016
5	-0.733	5	5.563
6	-0.921	6	-0.325
7	-6.579	7	6.195
8	6.087	8	5.962
9	0.324	9	-2.168
10	-0.980	10	11.268
11	0.781	11	2.309
12	-0.711	12	2.755
13	-1.811	13	-0.269
14	0.522	14	2.466
15	-0.839	15	1.398
16	-1.773	16	0.637
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

Interval		f	cum f	cum p
-6.99	-6.00	1	19	1.00
-5.99	-5.00	0	18	.95
-4.99	-4.00	2	18	.95
-3.99	-3.00	0	16	.84
-2.99	-2.00	1	16	.84
-1.99	-1.00	4	15	.79
-0.99	0.00	6	11	.58
+0.01	-1.00	4	5	.26
1.01	-2.00	0	1	.05
2.01	3.00	0	1	.05
3.01	4.00	0	1	.05
4.01	5.00	0	1	.05
5.01	6.00	0	1	.05
6.01	7.00	1	1	.05

Population 2

Interval		f	cum f	cum p
-2.99	-2.00	1	1	.05
-1.99	-1.00	1	2	.11
-0.99	0.00	2	4	.21
+0.01	1.00	4	8	.42
1.01	2.00	1	9	.47
2.01	3.00	3	12	.63
3.01	4.00	1	13	.68
4.01	5.00	0	13	.68
5.01	6.00	2	15	.79
6.01	7.00	1	16	.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).

C_2: 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

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

Figure 1: Emperical Estimation of the Distribution of P given π_1.

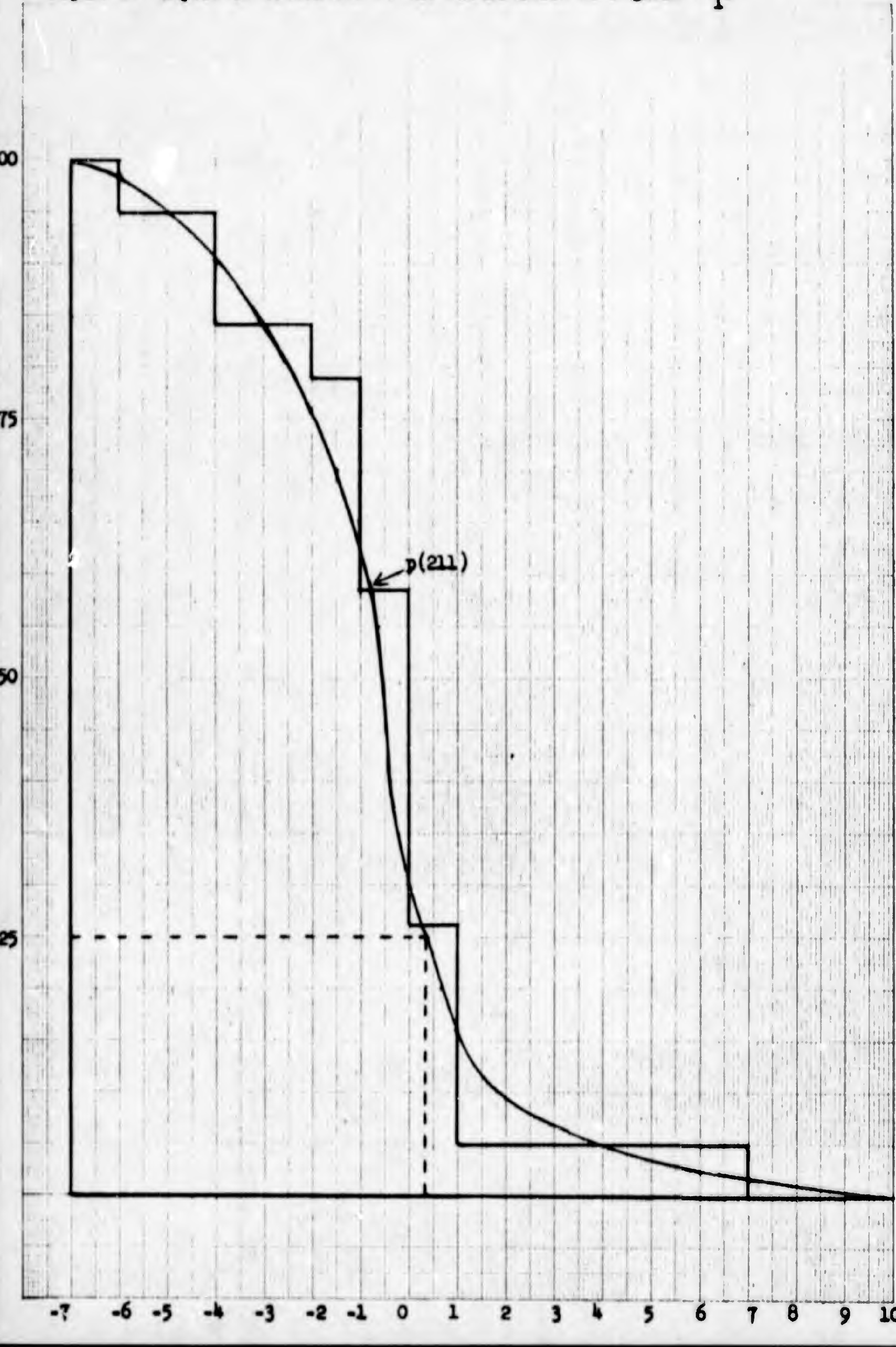
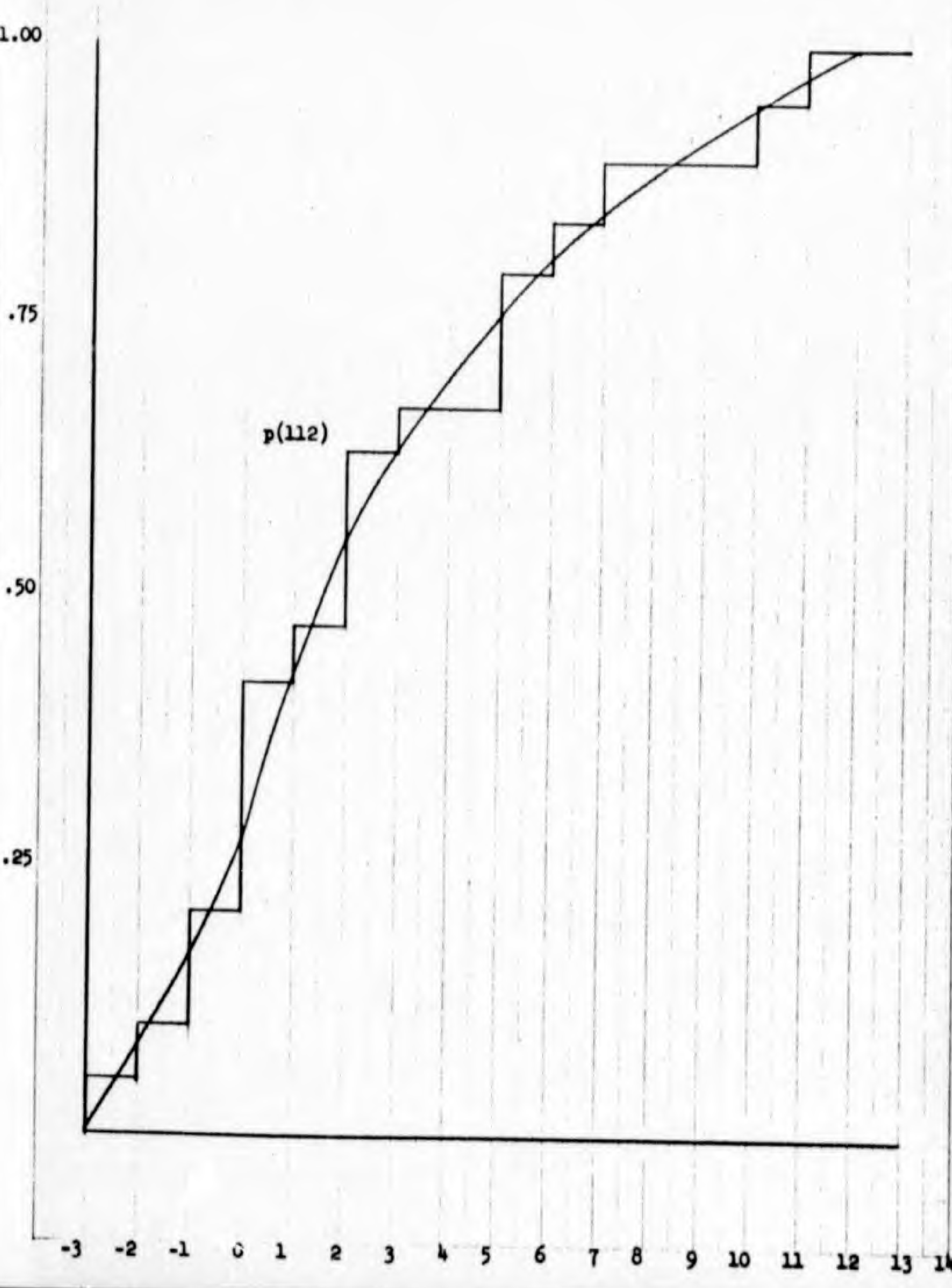


Figure 2: Emperical Estimation of the Distribution of P given π_2.



C_3: 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 π_1	From Population π_2
To Population	π_1	15	6
	π_2	4	13
	Total	19	19

We obtain

$$p(2/1) = 4/19 = .21$$

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

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

We have

$$P(\underline{Z}) = -.00019\, Z_1^2 + .00376\, Z_2^2 + .00922\, Z_3^2 -.00736\, Z_1Z_2 + .00543\, Z_1Z_3 + .00454\, Z_2Z_3 + .3242\, Z_1 - .0038\, Z_2 - .5788\, Z_3 - 8.333$$

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

$$P(\underline{Z}_1) = -.00019(34)^2 + .00376(36)^2 + .00922(12)^2$$

$$-.00736(34)(36) + .00543(34)(12) + .00454(36)(12)$$

$$+ .3242(34) - .0038(36) + .5788(12) - 8.333$$

$$P(\underline{Z}_1) = -3.248$$

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.

$$A(\underline{Z}) = a_1Z_1 + a_2Z_2 + \ldots + a_pZ_p + b$$

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 (1961).

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.

B_1: 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

$$\bar{X}_p^{(1)} = \sum_{n=1}^{N_1} X_{pn}^{(1)}/N_1$$

$$\bar{X}_p^{(2)} = \sum_{n=1}^{N_2} X_{pn}^{(2)}/N_2$$

$$S_{pq}^{(1)} = \frac{N_1 \sum_{n=1}^{N_1} X_{pn}^{(1)} X_{qn}^{(1)} - \sum_{n=1}^{N_1} X_{pn}^{(1)} \sum_{n=1}^{N_1} X_{qn}^{(1)}}{N_1^{\,2}}$$

$$S_{pq}^{(2)} \quad \frac{N_2 \sum_{n=1}^{N_2} X_{pn}^{(2)} X_{qn}^{(2)} - \sum_{n=1}^{N_2} X_{pn}^{(2)} \sum_{n=1}^{N_2} X_{qn}^{(2)}}{N_2^{\,2}}$$

B_2: 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

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

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

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.

$$\mu_p = \bar{X}_p^{(2)} - \bar{X}_p^{(1)} = \frac{\sum_{n=1}^{N_2} X_{pn}^{(2)}}{N_2} - \frac{\sum_{n=1}^{N_1} X_{pn}^{(1)}}{N_1}$$

and

$$S_{pq}^{(1)} = \frac{N_1 \sum_{n=1}^{N_1} X_{pn} X_{qn} - \sum_{n=1}^{N_1} X_{pn} \sum_{n=1}^{N_1} X_{qn}}{N_1^{\,2}}$$

$$S_{pq}^{(2)} = \frac{N_2 \sum_{n=1}^{N_2} X_{pn}^{(2)} X_{qn}^{(2)} - \sum_{n=1}^{N_2} X_{pn}^{(2)} \sum_{n=1}^{N_2} X_{qn}^{(2)}}{N_2^{\,2}}$$

B_3: Estimation of Coefficients $a_1, a_2, \ldots, a_p$.

The coefficient a_p, $p = 1, 2, \ldots, p$ is obtained from the equation

$$a_p = \sum_{q=1}^{P} S_{(r)}^{pq} \quad \delta_q$$

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

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

and

$$\delta_q = \bar{X}_q^{(2)} - \bar{X}_q^{(1)}$$

B_4: Estimation of b

The constant b is given by the equation

$$b = \frac{\sqrt{a' (S_{pq}^{(2)}) \, a} \cdot a' \ \bar{X}^{(1)} + \sqrt{a' \ S_{pq}^{(2)} \ a} \cdot a' \ \bar{X}^{2}}{\sqrt{a' (S_{pq}^{(2)}) \, a} + \sqrt{a' (S_{pq}^{(1)}) \, a}}$$

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.

C_3: 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 $A(\underline{Z})$.

The conditional means are given by the equations

$$\bar{A}_1 = \sum_{p=1}^{P} a_p \bar{X}_p^{(1)} + b \text{, given } \underline{Z} \text{ belongs to } \pi_1$$

$$\bar{A}_2 = \sum_{p=1}^{P} a_p \bar{X}_p^{(2)} + b \text{, given } \underline{Z} \text{ belongs to } \pi_2$$

The conditional variances are given by the equations

$$V_A^{(1)} = \sum_{q=1}^{P} \sum_{p=1}^{P} a_p a_q S_{pq}^{(1)} \text{, given } \underline{Z} \text{ belongs to } \pi_1$$

$$V_A^{(2)} = \sum_{q=1}^{P} \sum_{p=1}^{P} a_p a_q S_{pq}^{(2)} \text{, given } \underline{Z} \text{ belongs to } \pi_2 .$$

C_2: Determination of λ the classification constant according to the decision strategy being used.

We have the relationships for any given λ

$$p(2/1) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\lambda - \bar{A}_1}{V_A^{(1)}}}^{\infty} e^{-t^2/2} \, dt$$

and

$$p(1/2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\lambda - \bar{A}_2}{V_A^{(2)}}} e^{-t^2/2} \, dt$$

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

C_3: 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.

		From Population π_1	From Population π_2
To Population	π_1	n_{11}	n_{12}
	π_2	n_{21}	n_{22}
		N_1	N_2

From this table we make the estimates

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

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

Theoretical Estimation:

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

$$p(2/1) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\lambda - \bar{A}_1}{V_A^{(1)}}}^{\infty} e^{-t^2/2}$$

$$p(1/2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\lambda - \bar{A}_2}{V_A^{(2)}}} e^{-t^2/2}$$

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 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.

B_1: The sample moment computation.

The sample summations

Population 1		Population 2	
N_1	= 19	N_2	= 19
$\sum X_1$	= 827	$\sum X_1$	= 1039
$\sum X_2$	= 844	$\sum X_2$	= 1035
$\sum X_3$	= 311	$\sum X_3$	= 446
$\sum X_1^2$	= 39033	$\sum X_1^2$	= 59647
$\sum X_2^2$	= 39388	$\sum X_2^2$	= 58699
$\sum X_3^2$	= 5641	$\sum X_3^2$	= 11408
$\sum X_1X_2$	= 38073	$\sum X_1X_2$	= 57661
$\sum X_1X_3$	= 14007	$\sum X_1X_3$	= 25250
$\sum X_2X_3$	= 14152	$\sum X_2X_3$	= 25003

Estimation of Population Parameters

Population 1	Population 2
$\bar{X}_1^{(1)} = \frac{827}{19} = 43.53$	$\bar{X}_1^{(2)} = \frac{1039}{19} = 54.68$
$\bar{X}_2^{(1)} = \frac{844}{19} = 44.42$	$\bar{X}_2^{(2)} = \frac{1035}{19} = 54.47$
$\bar{X}_3^{(1)} = \frac{311}{19} = 16.37$	$\bar{X}_3^{(2)} = \frac{446}{19} = 23.47$
$S_{11}^{(1)} = \frac{56798}{361} = 159.83$	$S_{11}^{(2)} = \frac{53772}{361} = 148.95$

$S_{22}^{(1)} = \frac{36036}{361} = 99.82$ $S_{22}^{(2)} = \frac{44056}{361} = 122.04$

$S_{33}^{(1)} = \frac{10458}{361} = 28.97$ $S_{33}^{(2)} = \frac{17836}{361} = 49.41$

$S_{12}^{(1)} = \frac{25399}{361} = 70.35$ $S_{12}^{(2)} = \frac{20194}{361} = 55.94$

$S_{13}^{(1)} = \frac{8935}{361} = 24.75$ $S_{13}^{(2)} = \frac{16356}{361} = 45.31$

$S_{23}^{(1)} = \frac{6404}{361} = 17.74$ $S_{23}^{(2)} = \frac{13447}{361} = 37.25$

Estimation of r.

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

Consider $r = 0$, to determine $s(0)$

(1) Evaluate (S_{pq}^{r}) for given r

$$(S_{pq}^{(r=0)}) = 0\begin{pmatrix} 159.83 & 70.35 & 24.75 \\ 70.35 & 99.82 & 17.74 \\ 24.75 & 17.74 & 28.97 \end{pmatrix} + 1\begin{pmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{pmatrix}$$

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

(2) Invert $(S_{pq}^{(r)})$

Determine

$$\left|S_{pq}^{(r)}\right| = \begin{vmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{vmatrix} = 475155$$

Determine matrix of cofactors

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

where $4642 = (122.04)(49.41) - (37.25)(37.25)$

Divide cofactor matrix by $\left| S_{pq}^{(r)} \right|$

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

(3) Compute vector of mean differences

$$\delta_p = \bar{X}_p^{(2)} - \bar{X}_p^{(1)}$$

$$\underset{\sim}{\delta} = 11.15, \ 10.05, \ 7.10$$

where

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

etc.

(4) Compute the vector of sums of products $(S_r^{pq})(\underset{\sim}{\delta}' = \underset{\sim}{B})$(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

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

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

(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

$$-17.717 = (.03475)(148.95) + (.04205)(55.94) + (.22486)(45.31)$$

(7) Multiply $(\underline{C})\,(\underline{B}) = s(r)$

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

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

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

Thus by linear interpolation

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

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

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

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

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

Invert $(S_{pq}^{(r)})$

$$\left| S_{pq}^{(r)} \right| = \begin{vmatrix} 154.97 & 63.91 & 33.94 \\ 63.91 & 109.75 & 26.46 \\ 33.94 & 26.46 & 38.16 \end{vmatrix} = 372380$$

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

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

Division of cofactors by $\left| R_{pq}^{(r)} \right|$

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

B_2: Solve for a_p

$$a_1 = .00935(11.15)\ -.00413(10.05)\ -.00546(7.10)$$
$$= .0240$$
$$a_2 = -.00413(11.15)\ +.01277(10.05)\ -.00518(7.10)$$
$$= .0455$$

$$a_3 = -.00546(11.15) \; -.00518(10.05) \; +.03470(7.10)$$
$$= .1487$$

B_3: Solve for b

Evaluate $\underline{a}' \cdot (S_{pq}^{(1)})$

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

$$= (10.72, \; 8.87, \; 5.71)$$

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 $\underline{a}' \cdot S_{pq}^{(2)}$

$$(.0240, \; .0455, \; .1487)\begin{pmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{pmatrix}$$

$$= (12.86, \; 12.43, \; 10.13)$$

Compute $[\underline{a}' (S_{pq}^{(2)})]\, a$

$$(12.86)(.0240) + (12.43)(.0455) + 10.13(.1487) = 2.38$$

Compute $\sqrt{\underline{a}' \; (S_{pq}^{(2)} \; a}$

$$\sqrt{2.38} = 1.54$$

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

$$(10.72)(43.53) + (8.87)(44.42) + (5.71)(16.37) = 9.54$$

Compute $\underline{a}'\underline{\bar{X}}^{(2)}$

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

Evaluate b

$$b = \frac{1.54(9.54) + 1.23(12.03)}{1.54 + 1.23} = 1065$$

Thus

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

Step C: The Determination of the Classification Rule.

Empirical Estimation

C_1: 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 Observation No.	A	Population 2 Observation No.	A
1	1069.351	1	1069.859
2	1068.856	2	1071.574
3	1069.943	3	1074.031
4	1069.824	4	1070.814
5	1070.302	5	1073.506
6	1070.281	6	1071.104
7	1068.975	7	1073.569
8	1073.681	8	1074.094
9	1071.289	9	1070.670
10	1069.493	10	1075.311
11	1072.091	11	1072.811
12	1070.701	12	1072.418
13	1070.312	13	1070.093
14	1071.528	14	1072.213
15	1070.998	15	1071.030
16	1070.493	16	1071.331
17	1068.539	17	1071.468
18	1071.790	18	1072.818
19	1071.048	19	1074.633

These data are tabulated into cumulative frequency distributions.

Population 1

Interval	f	cum f	cum p
1068.0 - 1068.5	0	19	1.00
1068.5 - 1069.0	3	19	1.00
69.0 - 69.5	2	16	.84
69.5 - 70.0	2	14	.74
70.0 - 70.5	4	12	.63
70.5 - 71.0	2	8	.42
71.0 - 71.5	2	6	.32
71.5 - 72.0	2	4	.21
72.0 - 72.5	1	2	.11
72.5 - 73.0	0	1	.05
73.0 - 73.5	0	1	.05
73.5 - 74.0	1	1	.05
74.0 - 74.5	0	0	.00
74.5 - 75.0	0	0	.00

Population 2

Interval	f	cum f	cum p
1069.5 - 1070.0	1	1	.05
70.0 - 70.5	1	2	.11
70.5 - 71.0	2	4	.21
71.0 - 71.5	4	8	.42
71.5 - 72.0	1	9	.47
72.0 - 72.5	2	11	.58
72.5 - 73.0	2	13	.68
73.0 - 73.5	0	13	.68
73.5 - 74.0	2	15	.79
74.0 - 74.5	2	17	.89
74.5 - 75.0	1	18	.95
75.0 - 75.5	1	19	1.00
75.5 - 76.0	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

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

Figure 1: Emperical Estimation of the Distribution of A given TT.

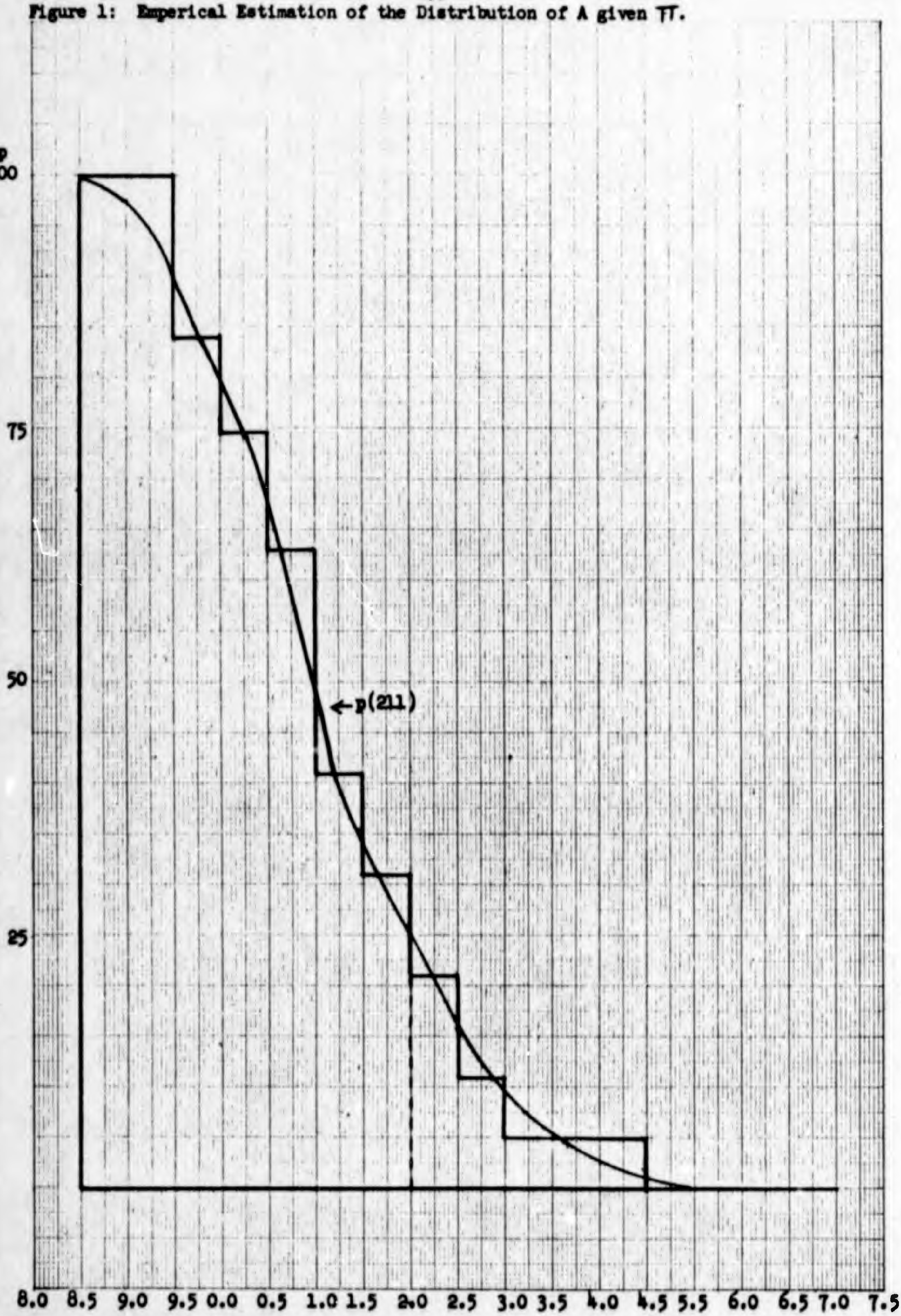
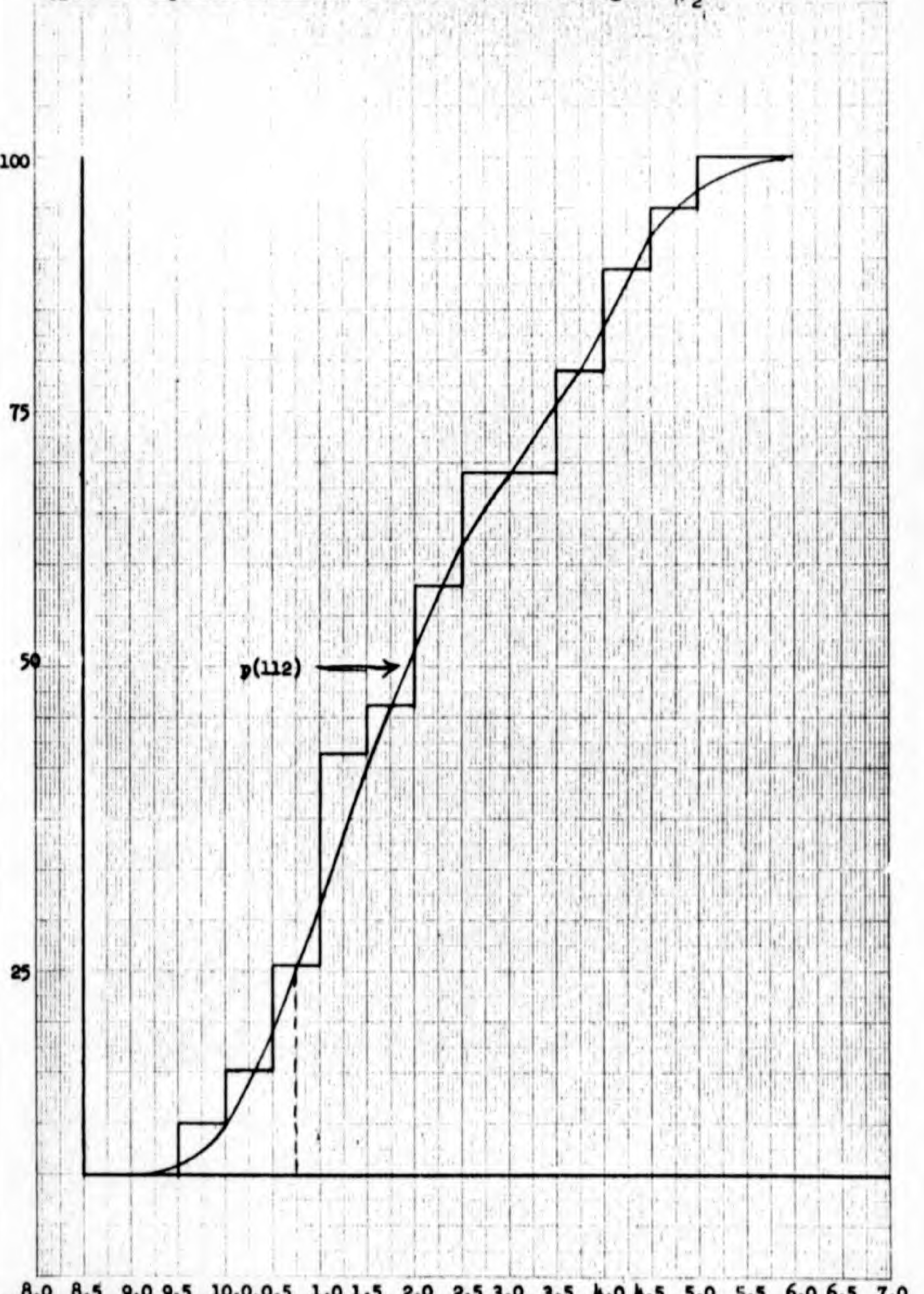


Figure 2: Emperical Estimation of the Distribution of A given Π_2.



C_3: Statement of the Classification Rule.

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

Theoretical Estimation

C_1: Determination of the Moments of $A(\underline{Z})$

$$\bar{A}_1 = a_1 \bar{X}_1^{(1)} + a_2 \bar{X}_2^{(1)} + a_3 \bar{X}_3^{(1)} + b$$

$$= .0240(43.53) + .0455(44.42) + .1487(16.37) + 1065$$

$$= 1070.5$$

$$\bar{A}_2 = a_1 \bar{X}_1^{(2)} + a_2 \bar{X}_2^{(2)} + a_3 \bar{X}_3^{(2)} + b$$

$$= .0240(54.68) + .0455(54.47) + .1487(16.37) + 1065$$

$$= 1071.2$$

$$\bar{V}_1 = a_1^2 S_{11}^{(1)} + a_2^2 S_{22}^{(1)} + a_3^2 S_{33}^{(1)} + 2a_1a_2S_{12}^{(1)} + 2a_1a_3S_{13}^{(1)} + 2a_2a_3S_{23}^{(1)}$$

$$= (.0240)^2(159.83) + (.0455)^2(99.82) + (.1487)^2(28.97)$$

$$+ 2(.0240)(.0455)(70.36) + 2(.0240)(.1487)(24.75) + 2(.0455)(.1487)(17.74)$$

$$V_A^{(1)} = 1.51$$

$$V_A^{(2)} = a_1^2 S_{11}^{(2)} + a_2^2 S_{22}^{(2)} + a_3^2 S_{33}^{(2)} + 2a_1a_2S_{12}^{(2)} + 2a_1a_3S_{13}^{(2)} + 2a_2a_3S_{23}^{(2)}$$

$$= (.0240)^2(148.95) + (.0455)^2(122.04) + (.1487)(49.41)$$

$$+ 2(.0240)(.0455)(55.94) + 2(.0240)(.1487)(45.31) + 2(.0455)(.1487)(37.25)$$

$$= 2.38$$

C_2: Determination of λ, the classification constant, according to the decision strategy being used.

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

$$0.25 = \frac{1}{\sqrt{2\pi}} \int_{\frac{\lambda - \bar{A}_1}{\sqrt{V_A^{(2)}}}}^{\infty} e^{-t^2/2}\, dt$$

Thus,

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

$$\lambda = (1.23)(0.67) + 1070.5 = 0.82 + 1070.5$$
$$= 1071.32$$

C_3: 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 π_1	π_2
To	π_1	15	8
Population	π_2	4	11
	Total	19	19

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

and $p(1/2) = 8/19 = .42$

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

We have

$$\Lambda(\underline{Z}) = .0240\ Z_1 + .0455\ Z_2 + .1437\ Z_3 + 1065$$

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

$$\Lambda(\underline{Z}_1) = .0240(34) + .0455(36) + .1487(12) + 1065 = 1069.238$$

Since $\Lambda(\underline{Z}_1) \leq 1071.5$, $\underline{Z}_1$ is classified as a π_1

For $\underline{Z}_2$: $Z_1 = 55$, $Z_2 = 62$, $Z_3 = 21$

$$\Lambda(\underline{Z}_2) = .0240(55) + .0455(62) + .1487(21) + 1065 = 1072.264$$

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

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

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 χ^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).

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.

B_1: Estimation of the Population Covariances.

We chose to use the maximum likelihood estimates

$$S_{pq}^{(1)} = \frac{N_1 \sum_{n=1}^{N_1} X_{np}^{(1)} X_{nq}^{(1)} - \sum_{n=1}^{N_1} X_{np} \sum_{n=1}^{N_1} X_{nq}}{N_1^2}$$

$$S_{pq}(2) = \frac{N_2 \sum_{n=1}^{N_2} X_{np}^{(2)} X_{nq}^{(2)} - \sum_{n=1}^{N_2} X_{np} \sum_{n=1}^{N_2} X_{nq}}{N_2^2}$$

B_2: Inversion of the Covariance Matrices.

$$\begin{pmatrix} S_{11}^{(R)} & S_{12}^{(R)} & \cdots & S_{1P}^{(R)} \\ S_{21}^{(R)} & \cdot & & \\ \vdots & & \ddots & \\ S_{P1}^{(R)} & & \cdot & S_{PP}^{(R)} \end{pmatrix}^{-1} = \begin{pmatrix} S_{(R)}^{11} & S_{(R)}^{21} & \cdots & S_{(R)}^{P1} \\ S_{(R)}^{21} & S_{(R)}^{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ S_{(R)}^{P1} & \cdots & \cdots & S_{(R)}^{PP} \end{pmatrix}$$

B_3: 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.

C_3: 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 π_1	π_2
To	π_1	n_{11}	n_{12}
Population	π_2	n_{21}	n_{22}
	Total	N_1	N_2

From this table we make the estimates

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

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

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

X_2 = English Placement Test Score

X_3 = General Aptitude Test Score

Step B: The Selection of the Estimation Procedure.

B_1: Estimation of the Population Covariances.

<u>The Sample Summations</u>

Population 1		Population 2	
N_1	19	N_2	19
$\sum X_1$	827	$\sum X_1$	1039
$\sum X_2$	844	$\sum X_2$	1035
$\sum X_3$	311	$\sum X_3$	446
$\sum X_1^2$	39033	$\sum X_1^2$	59647
$\sum X_2^2$	39388	$\sum X_2^2$	58699
$\sum X_3^2$	5641	$\sum X_3^2$	11408
$\sum X_1X_2$	38073	$\sum X_1X_2$	57661
$\sum X_1X_3$	14007	$\sum X_1X_3$	25250
$\sum X_2X_3$	14152	$\sum X_2X_3$	25003

Estimation of Population Covariances

$$s_{11}^{(1)} = \frac{56798}{19} = 159.83 \qquad s_{11}^{(2)} = \frac{53772}{19} = 148.95$$

$$s_{22}^{(1)} = \frac{36036}{19} = 99.82 \qquad s_{22}^{(2)} = \frac{44056}{19} = 122.04$$

$$s_{33}^{(1)} = \frac{10458}{19} = 28.97 \qquad s_{33}^{(2)} = \frac{17836}{19} = 49.41$$

$$s_{12}^{(1)} = \frac{35399}{19} = 70.36 \qquad s_{12}^{(2)} = \frac{20194}{19} = 55.94$$

$$s_{13}^{(1)} = \frac{8936}{19} = 24.75 \qquad s_{13}^{(2)} = \frac{16356}{19} = 45.31$$

$$s_{23}^{(1)} = \frac{6404}{19} = 17.74 \qquad s_{23}^{(2)} = \frac{13447}{19} = 37.25$$

B_2: Inversion of the Covariance Matrices*

Evaluation of covariance determinants

$$\left| s_{pq}^{(1)} \right| = \begin{vmatrix} 159.83 & 70.36 & 24.75 \\ 70.36 & 99.82 & 17.74 \\ 24.75 & 17.74 & 28.97 \end{vmatrix} = 269117$$

$$\left| s_{pq}^{(2)} \right| = \begin{vmatrix} 148.95 & 55.94 & 45.31 \\ 55.94 & 122.04 & 37.25 \\ 45.31 & 37.25 & 49.41 \end{vmatrix} = 475155$$

*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 $2577 = (99.82)(28.97) - (17.74)(17.74)$

$$(C_{pq}^{(2)}) = \begin{pmatrix} 4642 & -1076 & -3446 \\ -1076 & 5307 & -3014 \\ -3446 & -3014 & 15049 \end{pmatrix}$$

Division of Cofactors by determinant

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

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

B_3: Determination of the Classification Statistic.

Evaluation of $(S_{(2)}^{pq} - S_{(1)}^{pq})$ matrix

$$(S_{(2)}^{pq} - S_{(1)}^{pq}) = \begin{pmatrix} .00019 & .00368 & -.00271 \\ .00368 & -.00376 & -.00227 \\ -.00271 & -.00227 & -.00922 \end{pmatrix}$$

Determination of the coefficientsof the statistic

$$S(\underline{Z}) = +.00019\, Z_1^2 \quad .00376\, Z_2^2 \quad -.00922\, Z_3^2$$
$$+.00736\, Z_1Z_2 \quad -00542\, Z_1Z_3 \quad .00454\, Z_2Z_3$$

Step C: The Determination of the Classification Rule.

C_1: Determination of the values of the classification statistic for the sample observations from π_1 and π_2.

Thus $\underline{X}_1^{(1)} = (22, 35, 15)$

$$S(\underline{X}_1^{(1)}) = .00019(22)^2 - .00376(35)^2 - .00922(15)^2$$
$$+ .00736(22)(35) - .00542(22)(15) - .00454(35)(15)$$
$$= -5.093$$

The results of these evaluations are tabulated below.

Population 1		Population 2	
Observation No.	$S(\underline{Z})$	Observation No.	$S(\underline{Z})$
1	-5.093	1	-4.903
2	0.083	2	-1.600
3	4.621	3	-12.672
4	0.131	4	-3.200
5	-3.010	5	-11.133
6	-2.494	6	-3.054
7	-4.297	7	-14.860
8	-11.259	8	-6.688
9	-0.698	9	-6.789
10	5.201	10	-15.906
11	0.410	11	-0.508
12	0.781	12	-7.840
13	-4.611	13	-2.976
14	-3.881	14	-6.659
15	-7.906	15	-1.664
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	f	cum f	cum p
-11.999 to -11.000	1	1	.05
-10.999 to -10.000	0	1	.05
- 9.999 to - 9.000	1	2	.11
- 8.999 to - 8.000	0	2	.11
- 7.999 to - 7.000	1	3	.16
- 6.999 to - 6.000	0	3	.16
- 5.999 to - 5.000	2	5	.26
- 4.999 to - 4.000	3	8	.42
- 3.999 to - 3.000	2	10	.53
- 2.999 to - 2.000	1	11	.58
- 1.999 to - 1.000	0	11	.58
- 0.999 to 0.000	2	13	.68
+ 0.001 to + 1.000	4	17	.89
+ 1.001 to + 2.000	0	17	.89
+ 2.001 to + 3.000	0	17	.89
+ 3.001 to + 4.000	0	17	.89
+ 4.001 to + 5.000	1	18	.95
+ 5.001 to + 6.000	1	19	1.00

Population 2

Interval	f	cum f	cum p
-27.2	1	19	1.00
-15.999 to -15.000	1	18	.95
-14.999 to -14.000	1	17	.89
-13.999 to -13.000	0	16	.84
-12.999 to -12.000	1	16	.84
-11.999 to -11.000	1	15	.79
-10.999 to -10.000	0	14	.74
- 9.999 to - 9.000	1	14	.74
- 8.999 to - 8.000	0	13	.68
- 7.999 to - 7.000	1	13	.68
- 6.999 to - 6.000	3	12	.63
- 5.999 to - 5.000	0	9	.47
- 4.999 to - 4.000	1	9	.47
- 3.999 to - 3.000	2	8	.42
- 2.999 to - 2.000	2	6	.32
- 1.999 to - 1.000	3	4	.21
- 0.999 to - 0.000	1	1	.05

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).

Figure 1: Emperical Estimation of the Distribution of S given π_1.

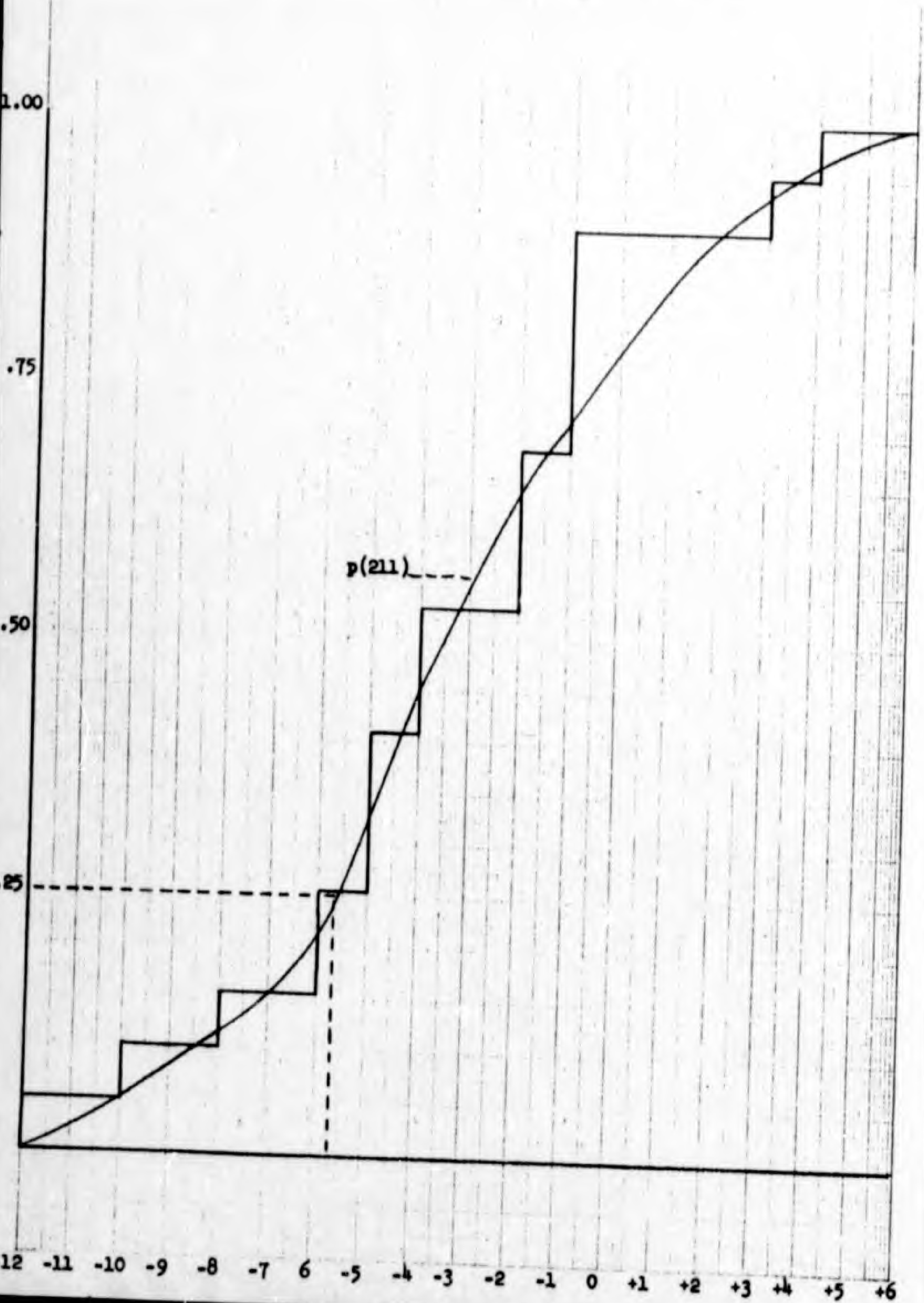
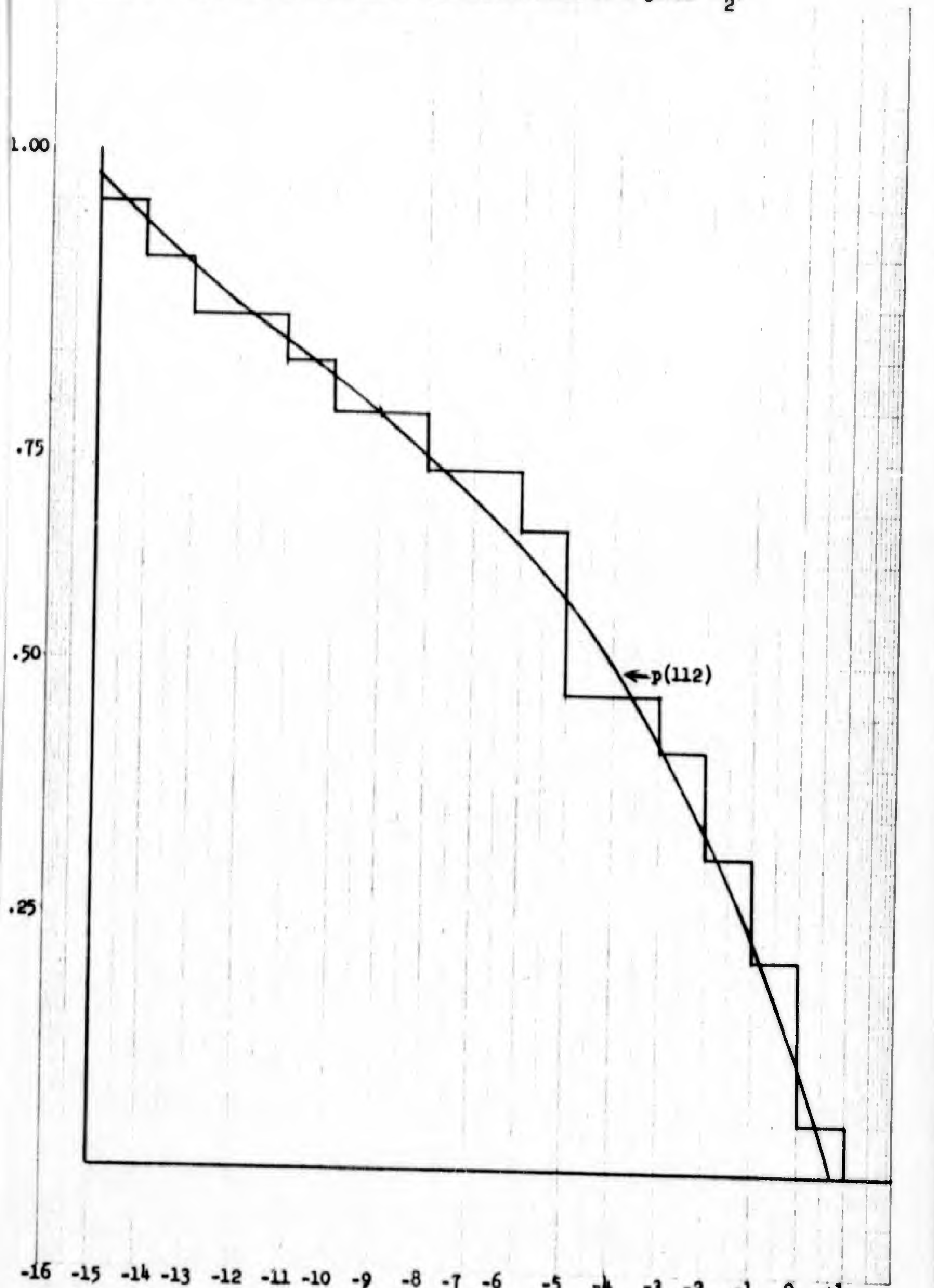


Figure 2: Emperical Estimation of the Distribution of S given π_2.



C_2: 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

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

C_3: Statement of the Classification Rule.

If $S(\underline{Z}) < -5.60$, 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 Population	
		π_1	π_2
To	π_1	15	9
Population	π_2	4	10
	Total	19	19

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

$p(1/2) = 9/19 = .47$

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

We have

$$S(\underline{Z}) = .00019\, Z_1^2 - .00376\, Z_2^2 - .00922\, Z_3^2 + .00736\, Z_1 Z_2 - .00542\, Z_1 Z_3 - .00454\, Z_2 Z_3$$

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

$$S(\underline{Z}_1) = .00019(34)^2 - .00376(36)^2 - .00922(12)^2$$
$$+ .00736(34)(36) - .00542(34)(12) - .00454(36)(12)$$
$$= -1.145$$

Since $S(Z_1) > -5.60$ $\underline{Z}_1$ 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 $S(Z_2) > -5.60$, $\underline{Z}_2$ 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, say $\pi_{(2)}$

(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)

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 proceding 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 ℓ th population. Here

$$C_{\ell} = \sum_{\substack{\ell = 1 \\ \ell \neq \ell}}^{K} q_m \, p_m(Z) \; C(\ell/m)$$

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

$$r(\ell, R) = \sum_{k=1}^{K} C(k_1 \ell') \; P\left\{ k/\ell, R \right\}$$

where $C(\ell/k)$ is the cost of classifying an individual into population π_ℓ if it really belongs to population π_k and $P\left\{ k/\ell, R \right\}$ is the probability of making the classification given that the individual belongs to π_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,

(a) maximize $r(\ell, R)$ overall possible ℓ's and

(b) Choose $R_1, R_2, \ldots, R_k$ 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 $(k=1, 2, \ldots, k)$ 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

$$p_k(\underline{Z}) > M_{k\ell}\, p_\ell(Z) \quad \text{where } k \neq \ell.$$

The constants $M_{k\ell}$ are determined so that the common probability of correct classification is maximized."

Chapter VI

Bibliography* on Classification and Related Topics

Adhikari, B. P. & D. D. Joshi, Distance-discrimination et resume exhaustif, Publs. inst. stat. Univ. Paris, vol. 5, (1956), Fasc. 2, 57-74.

Anderson, T. W., Classification by multivariate analysis, Psychometrika, 16, (1951), 31-51.

Anderson, T. W., An Introduction to Multivariate Statistical Analysis, Chapt. 6, 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, J. Roy. Stat. Soc., 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, Ann. Eugen, 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, J. Roy. Stat. Soc. Suppl., 9, (1947), 176-190.

Bartlett, M. S., An inverse matrix adjustment arising in discriminant analysis, Ann. Math. Stat., 22, (1951), 107-111.

Baten, W. D., The discriminant function applied to spore measurements, Mich. Acad. of Sci., Arts, and Letters, 29, (1943), 3-7.

Baten, W. D., The use of discriminant functions in comparing judges scores concerning potatoes, J. Amer. Stat. Assoc., 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, Indust. & Eng. Chem., 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), Biometrische Zeitschrift, 1, (1959), 3-14.

Bhattacharya A., On discrimination and divergence, Proc. Indian Sci. Congress, part III, 13, (1942).

Bhattacharya, A., On discrimination and divergence (abstract), Sankyha, (1943), 249.

Bhattacharya, A., On a measure of divergence between two statistical populations defined by their probability distributions, Bull. Calcutta Math. Soc., 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 D^2-Statistic, Bull. of the Calcutta Math Soc., 33, (1941), 87-92.

Bhattacharya, D. &. & R. D. Narayan, Moments of the D^2-statistic for populations with unequal dispersions, Sankhya, 5, (1941), 401-412.

Birnbaum, A. & A. E. Maxwell, Classification statistics based on Baye's formula, App. Stat., 9, (1959), 152.

Blackith, R. E., A synthesis of multivariate technique to distinguish patterns of growth in grasshoppers, Biometrics, 16, (1960), 28-40.

Bose, P. K., On recursion formulae, tables and Bessel function populations associated with the distribution of classical D^2-statistic, Sankhya, 8, (1947), 235-248.

Bose, P. K., Incomplete probability integral tables connected with studentised D^2-statistic., Bull. Calcutta Stat. Assoc., 2, (1949), 131-137.

Bose, P. K., Corrigenda (for tables of incomplete probability integral of classical D^2-statistics - by P. K. Bose in Sankhya, vol. 8), Sankhya, 11, (1951), 96.

Bose, Purnendu & S. R. Rao, Numerical study of the limiting forms of the "D^2" distribution (abstract), Sankhya, 6, (1943), 348.

Bose, R. C., On the exact distribution and moment coefficients of the D^2-statistic, Sci. Cult., 1, (1935), 205-206.

Bose, R. C., On the exact distribution and moment coefficients of the D^2-statistic, Sankhya, 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^2-statistic, Sankhya, 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, Proc. Ind. Sc. Congress, (1946).

Bose, R. C., The fundamental theorem of linear estimation for correlated variates, Proc. Ind. Sc. Congress, (1946).

Bose, R. C. & S. N. Roy, The distribution of the studentized D^2-statistic, Proc. Ind. Stat. Conference, Calcutta, pp. 19-38, (1938).

Bose, R. C. & S. N. Roy, The distribution of the studentized D^2-statistic, Sankhya, 4, (1938), 19-38.

Bose, S. N., On the complete moment-coefficients of the D^2-statistic, Sankhya, 2, (1936), 385-396.

Bose, S. N., On the moment coefficients of the D^2-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 T^2 and Anderson's classification statistic W in terms of simple statistics, #12 in Contributions to probability and statistics, Essays in honor of Harold Hotelling, Stanford U. Press, 1960.

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, Proc. Amer. Soc. An. Prod., 1, (1940), 153-160.

Brogden, Hubert E., A simple proof of a personnel classification theorem, Psychom., 19, (1954), 205-208.

Brogden, Hubert E., Least squares estimates and optimal classification, Psychom., 20, (1955), 249-252.

Brown, George W., Discriminant functions, Ann. Math. Stat., 18, (1947), 514-528.

Brown, G. W., Basic principles for construction and application of discriminators (abstract), Biometrics, 5, (1949), 80.

Brown, George, W., Basic principles for construction and application of discriminators, J. Clin. Psychol., 6, (1950), 58-61.

Burt, Cyril, Statistical problems in the evaluation of army tests, Psychom., 9, (1944), 219-235.

Cacoullos, Theophilos, Computing distances between Multivariate Populations I(abstract), Ann. Math. Stat., 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, Biometrika, 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. I. 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, Biometrics, 17, (1961), 10-32.

Cox, D. R. & L. Brandwood, On a discriminatory problem connected with the works of Plato, J. Roy. Stat. Soc., B, 21, (1959), 195-200.

Cox, G. M. & W. P. Martin, Use of a discriminant function for differentiating soils with different Azotobacter populations, Iowa State Col. J. Sci., 9, (1937), 323-331.

Das, A. C., A note on the D^2-statistic when the variances and covariances are known, Sankhya, 8, (1948), 372-374.

Day, Bessie B. & Marion M. Sandomire, Use of the discriminant function for more than two groups, J. Amer. Stat. Assoc., 37, (1942), 461-472.

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, Psychom., 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^2 Analysis, Sankhya, 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, Ann. Eugen., 8, (1938), 376-386.

Fisher, R. A., The sampling distribution of some statistics obtained from non-linear equations, Ann. Eugen., 9, (1939), 238-249.

Fisher, R. A., The precision of discriminant functions, Ann. Eugen., 10, (1940), 422-429.

Fisher, R. A., Statistical Methods for Research Workers, 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, Appl. Stat., 6, (1957), 123.

Garrett, Henry E., The discriminant function and its use in psychology, Psychom., 8, (1943), 65-79.

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, Ann. Math. Stat., 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, Psychometrika, 15, 271-90.

Hoel, P. G. & R. P. Peterson, A solution to the problem of optimum classification, Ann. Math. Stat., 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, *Ann. of Eugenics*, 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, *Sankhya*, 21, (1959), 371-376.

John, S., On some classification problems I, *Sankhya*, 21, (1960), 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., *A Course in Multivariate Analysis*, Chapt. 9, Charles Griffin & Co., Limited, London (1957).

Kossack, Carl F., On the mechanics of classification, *Ann. Math. Stat.*, 16, (1945), 95-97.

Kossack, Carl F., Some techniques for simple classification, *Proc. Berkeley Symp. on Math. Stat. and Prob.*, (1949), 345-352.

Kullback, S., An application of information theory to multivariate analysis, *Ann. Math. Stat.*, 23, (1952), 88-102.

Kullback, S., An application of information theory to multivariate analysis II, *Ann. Math. Stat.*, 27, (1956), 122-145.

Kullback, Solomon, *Information theory and statistics*, 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, *Pub. de l'Inst de Stat.*, 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., (1960).

Lord, Fredrick M., Notes on a problem of multiple classification, Psychom., 17, (1952), 297-304.

Lubin, A., Linear and non-linear discriminating functions, Brit. J. Psych. (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, J. Asiatic Soc. Bengal, 23, (1927), 301-33.

Mahalanobis, P. C., On the generalized distance in statistics, Proc. Nat. Inst. of Science India, (1936).

Mahalanobis, P. C., On tests and measures of group divergence, J. Asiatic Soc. of Bengal, 26, (1930), p. 541-88.

Mahalanobis, P. C., The application of statistical methods in physical anthropometry, Sankhya, 4, (1940), 594-598.

Mahalanobis, P. C., Historical note on the D^2-statistic, Sankhya, 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, Ann. Eugen, 11, (1941), 64-76.

Mises, R. von, On the classification of observation data into distinct groups, Ann. Math. Stat., 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, Indian J. of Etomology, 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, Sankhya, 22, (1960), 1-20.

Nanda, D. N., The standard errors of discriminant function coefficients in plant-breeding experiments, J. Roy. Stat. Soc., 11, (1949), 283-290.

Nandi, H. K., On the power function of studentised D^2-statistic, Bull. Calcutta Math. Soc., 38, (1946), 79-84.

Neyman, J. and Pearson, E. S., On the problem of R samples, Bull. Academic Polonaise des Sciences et des Lettres, Series A (1931), 460-81.

Narain, R. D., Some results on discriminant functions, J. Indian Soc. Agric. Stat., 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, Biometrika, 18, (1926), 105-17.

Penrose, L. S., Some notes on discrimination, Ann. Eugen., 13, (1947), 228-237.

Quenouille, M. H., A further note on discriminatory analysis, Ann. Eugen., 15, (1949), 11-14.

Quenouille, M. H., Note on the elimination of insignificant variates in discriminatory analysis, Ann. Eugen., 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, Nature, 159, (1947), 30-31.

Rao, C. R., A statistical criterion to determine the group to which an individual belongs, Nature, 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), Cur. Sci., 16, (1947), 216-217.

Rao, C. R., Tests of significance in multivariate analysis, Biometrika, 35, (1948), 58-79.

Rao, C. R., The utilization of multiple measurements in problems of biological classification, J. Roy. Stat. Soc., B, 10, (1948), 159-193.

Pearson, Karl, Note on standardization of method of using the coefficient of racial likeness, Biometrika, 20, (1928), 376-78.

Rao, C. R., On the distance between two populations, Sankhya, 9, (1949), 246-248.

Rao, C. R., On some problems arising out of discrimination with multiple characters, Sankhya, 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 $D^2_{p+q} - D^2_p$ and some computational aspects of D^2 statistics, Sankhya, 10, (1950), 257-268.

Rao, C. R., The problem of selecting individuals for various duties in a specified ratio, Sankhya, 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, Sankhya, 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), Sankhya, 11, (1951), 257-262.

Rao, C. R., Advanced statistical methods in Biometric research, 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), Sankhya, 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, Ann. Math. Stat., 25, (1954), 651-670.

Rao, C. R., Multivariate analysis: An indispensable statistical aid in applied research, Sankhya, 22, (1960), 317-338.

Rao, C. R. and P. Slater, Multivariate analysis applied to differences between neurotic groups, British J. Psych., Stat. Sec., 2, (1949), 17.

Riffenburgh, R. H. & C. W. Clunies-Ross, Linear discriminant analysis, Pacif. Sci., 14, July, 1960.

Rogers, David J. & T. T. Tanimoto, A computer program for classifying plants, Science, 132, (1960), 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 D^2-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), 381-396.

Roy, S. N., Distribution of certain symmetric functions of p-statistics on the null hypothesis, Sci. Cult., 5, (1940), 563.

Roy, S. N., Distribution of p-statistics on the non-null hypothesis, Sci. Cult., 5, (1940), 562-563.

Roy, S. N., The sampling distribution of p-statistics and certain allied statistics on the non-null hypothesis, Sankhya, 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, Sankhya, 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, Sankhya, 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, Sankhya, 8, (1946), 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, Sankhya, 8, (1946), 15-52.

Roy, S. N., & R. C. Bose, The use and distribution of the studentized D^2-statistic 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^2-statistics, Sci. Cult., 5, (1940), 773-775.

Sebestyen, George, Recognition of membership in classes, IRE Trans. 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 D^2-statistics, Bull. Calcutta Math. Soc., 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), Ann. Math. Stat., 23, (1952), 473-474.

Sitgreaves, Rosedith, Discriminating analysis in a non-classical case (abstract), Biometrics, 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, Ann. Eugen., 13, (1947), 272.

Smith, H. F., A discriminant function for plant selection, Ann. Eugen., 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, Ann. Math. Stat., 20, (1949), 502-522.

Solomon, Herbert, Classification procedures based on dichotomous response vectors, #36 in Contributions to probability and statistics, 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-6, IBM Applications Library, 590 Madison Avenue, New York 22, New York (1960).

Thorndike, Robert L., The problem of classification of personnel, Psychom., 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.

Tintner, Gerhard, Some formal relations in multivariate analysis, J. Roy. Stat. Soc., B, 12 (1950), 95-101.

Travers, R. M. W., The use of a discriminant function in the treatment of Psychological group differences, Psychom., 4, (1939), 25-32.

Tyler, Fred. T., Some examples of multivariate analysis in educational and psychological research, Psychom., 17, (1952), 289-296.

Votaw, D. E., Jr., Methods of solving some personnel-classification problems, Psychom., 17, (1952), 255-266.

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, Ann. Math. Stat., 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, Ann. Eugen., 8, (1938), 266-302.

Welch, B. L., Note on discriminant functions, Biometrika, 31, (1939), 218-220.

Welch, Peter D. & Richard S. Wimpress, Two multivariate statistical computer programs and their application to vowel recognition, J. Acoustical Soc. of Amer., 33, (1961), 425-434.

Wesler, Oscar, A classification problem involving multinomials, Ann. Math. Stat., 30, (1959), 128-133.

Wilks, S. S., Multidimensional statistical scatter, #40 in Contributions to probability and statistics, Essays in honor of Harold Hotelling, Stanford U. Press, 1960.

Wilks, S. S., Mathematical statistics, sections 18.7 and 18.8, John Wiley & Sons, Inc., New York, (1962).

Williams, E. J., Some exact tests in multivariate analysis, Biometrika, 39, (1952), 17-31.

Williams, E. J., Significance tests for discriminant functions and linear functional relationships, Biometrika, 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, Sankhya, 7, (1945), 263-268.

Yates, F. & M. J. R. Healy, Statistical methods in anthropology, Nature, 168, (1951), 1116-1118.

(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

Baggaley, A. R., The relationship between scores obtained by Harvard freshmen on the Kuder preference record and their fields of concentration, J. Ed. Psych., 38, 421-427, (1947).

Birnbaum, Z. W. & D. G. Chapman, On optimum selections from multi-normal populations, Ann. Math. Stat., 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, Genetics, 28, 479-490, (1943).

Lehman, E. L., On families of admissible tests, Ann. Math. Stat., 18, 97-104, (1947).

Lorge, I., Two-group comparisons by multivariate analysis, Amer. Ed. Res. Assoc. Off. Reprint, 117-121, (1940).

Martin, E. S., A study of an Egyptian series of mandibles with special reference to mathematical methods of sexing, Biometrika, 28, 149-178, (1936).

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, Ann. Math. Stat., 20, 95-98, (1949).

Selover, R. B., A study of the sophomore testing program at the University of Minnesota - Part II, J. Appl. Psych., 456-467, (1942).

Stevens, W. L., The standardization of rubber flexing tests, India Rubber World, 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).