

T 1309.

A PROBABILISTIC METHOD OF  
RANKING UNDERGROUND EXPLORATION PROPOSALS  
IN CONDITIONS OF GEOLOGIC UNCERTAINTY

ARTHUR LAKES LIBRARY  
COLORADO SCHOOL OF MINES  
GOLDEN, COLORADO

by

Maurice G. F. de St. Jorre

ProQuest Number: 10781683

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10781683

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

A Thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science in mathematics.

Signed: M. G. F. de St. Jorre  
M.G.F. de St. Jorre

Golden, Colorado

Date: 11 May, 1970

ARTHUR LAKES LIBRARY  
COLORADO SCHOOL OF MINES  
GOLDEN, COLORADO

Approved: W. W. Whitman  
Dr. W. W. Whitman  
Thesis Advisor

Joseph R. Lee  
Dr. J. R. Lee  
Head, Department of  
Mathematics

Golden, Colorado

Date: May 11, 1970

ABSTRACT

A method was developed for using existing mine geological data to predict the true nature of unknown veins discovered by exploratory diamond drilling. The data consisted of the distributions of physical and mineralogical characteristics of known ore bearing and waste veins.

All available information was first considered to find out which vein characteristics were most useful in distinguishing between the two types of veins. These characteristics were then used to determine if all areas of the mine were reasonably homogeneous in nature. All veins whose characteristics appeared markedly different from the average mine population were eliminated.

Vein intersections of unknown nature were classified by discriminant and probabilistic analysis as being either ore bearing or waste veins. The probabilities of these classifications being incorrect were computed, and these were used to calculate decision rule probabilities. The unknown vein intersections were then ranked by the latter as to their likelihood of being ore or waste.

CONTENTS

	Page
INTRODUCTION . . . . .	1
Description of the Problem. . . . .	1
Objectives of the Study . . . . .	2
Review of Previous Work . . . . .	2
Basic Approach. . . . .	4
Analysis of the Problem . . . . .	5
Nature of Characteristic Data. . . . .	6
General Form of Analysis . . . . .	8
Selection of Significant Characteristics . . . . .	8
Division of the Mine Into Regions of Homogeneous Nature . . . . .	9
Characteristic Dependencies . . . . .	9
Classification of Unknown Data Points. . . . .	10
Decision Rule Formulation . . . . .	11
THEORETICAL ANALYSIS OF CHARACTERISTIC DATA. . . . .	12
Notation. . . . .	12
Introduction. . . . .	12
Characteristic Data Acquisition . . . . .	13
Selection of Significant Characteristics. . . . .	14
Determination of Homogeneity of Characteristic Data. . . . .	19
Determination of Dependency Between Characteristics . . . . .	31
Methods of Comparison of Unknown Data Points With Known Populations. . . . .	33
Discriminant Analysis. . . . .	34
Computation of Misclassification Probabilities. . . . .	41
Analysis of Joint Probability Distributions. . . . .	46
Decision Rule Formulation . . . . .	51

	Page
APPLICATION OF THEORETICAL RESULTS . . . . .	56
Description of the Galena Mine. . . . .	56
Characteristic Data . . . . .	58
First Data Set . . . . .	59
Second Data Set. . . . .	67
Selection of Significant Characteristics. . . . .	70
Determination of the Homogeneity of Characteristic Data . . . . .	74
Determination of Dependency Between Significant Characteristics . . . . .	77
Comparison of Unknown Data Points With Known Populations. . . . .	79
Discriminant Analysis. . . . .	79
First Data Set. . . . .	79
Second Data Set . . . . .	82
Analysis of Joint Probability Distributions. . . . .	87
First Data Set. . . . .	87
Second Data Set . . . . .	94
CONCLUSIONS AND RECOMMENDATIONS. . . . .	97
Conclusions . . . . .	97
Recommendations for Further Study . . . . .	101
Determination of Optimal Exploration Strategy . . . . .	101
Prediction of Most Suitable Area for Vein Search. . . . .	101
LITERATURE CITED . . . . .	102

LIST OF FIGURES

Figure	Page
1. Univariate discriminant misclassification probabilities . . . . .	45
2. Plan view of a typical level - Galena Mine. . . . .	57
3. Original histograms - Silver characteristic . . . . .	68
4. Original histograms - Copper characteristic . . . . .	69
5. Amended histograms - Silver characteristic . . . . .	71
6. Amended histograms - Copper characteristic . . . . .	72
7. Discriminant and probabilistic limits . . . . .	86

LIST OF TABLES

Table	Page
1. Transformed data from the ore sample of the second data set . . . . .	73
2. Dependencies between significant characteristics.	78
3. Data points of unknown origin . . . . .	80
4. Discriminant values, and distances of unknown data points from population centers . . . . .	83
5. Misclassification and decision rule probabilities. Discriminant analysis - First data set . .	84
6. Misclassification and decision rule probabilities. Discriminant analysis - Second data set. .	88
7. Significant characteristic distributions. . . . .	91
8. Misclassification and decision rule probabilities. Probabilistic analysis - First data set. .	93
9. Misclassification and decision rule probabilities. Probabilistic analysis - Second data set .	96
10. Comparison of the rankings of unknown veins obtained by different methods . . . . .	98

ACKNOWLEDGMENTS

This project was supported by a research assistantship provided by the Colorado School of Mines Office of Research Services.

I am indebted to Dr. W. W. Whitman, and Professor W. R. Astle of the Mathematics Department for their guidance and helpful suggestions in the correct mathematical formulation of the problem. I would like to thank Dr. J. Mather of the Mineral Economics Department, and Dr. R. E. D. Woolsey of the Mathematics Department for serving on my thesis committee. I would also like to thank Dr. A. R. Brown of the Computing Center for pointing out a method for the numerical integration of the bivariate normal distribution function.

The data from the Galena mine used in this study was kindly made available by the American Smelting and Refining Company, who also provided financial assistance to me for the purpose of visiting the mine. I would like to thank also Mr. S. N. Kesten of ASARCO, and his geological staff at the Galena mine for their assistance in the collection of this data.

Finally, I must express my gratitude to my wife, Dorothy, whose unfailing support and encouragement enabled me to complete this study.

## INTRODUCTION

### Description of the Problem

In certain underground mines, the economic mineralization occurs in an apparently erratic manner in fault planes, fissures, tension fractures and other\* similar geologic features. Ore deposits of this type do not usually exhibit any well defined geologic controls by which the position of other\*\* orebodies may be predicted.

If the mine has been in operation for some time there will probably be a large mass of information available about known orebodies. This may be in the form of characteristic data about their size, position, content, and mineralogy. This data is usually the only basis on which underground exploration decisions can be made.

The problem that is considered in this study is how such data can be used in an efficient manner in the search for new orebodies.

\*These will be called veins regardless of their origin or content.

\*\*An orebody is defined as a continuous section of a vein that can be mined and processed so as to produce a profit. A waste section of a vein is any part of the vein that is not an orebody.

### Objectives of the Study

In general, there are two major exploration problems in a mine of this type.

1. To predict the areas in which ore bearing veins are most likely to be found.

2. To predict which of several known, but unexplored veins have the greatest probability of containing an economic orebody.

This study will be principally concerned with finding a solution to the second of these problems. The solution sought is a general method of obtaining a decision rule on which underground exploration decisions can be based. Any such solution will necessarily be limited to mines operating in the type of geologic conditions already described.

### Review of Previous Work

Much of the work that has been done in this field has presupposed that little or no information is available about the area in which the search was taking place. These studies have primarily been concerned with the problem of locating a mineral deposit within a large geographic region rather than locating an orebody within the relatively small area of a mine.

Allais (1957, p. 284-347) considered the economic aspects of mineral exploration over a very large previously unexplored geographic area, using the distributions of

economic mineral deposits from intensively explored similar geologic regions of comparable size. This line of reasoning was extended by Harris (1968, p. 295-327) and Griffiths (1967, p. 189-209). The former used multivariate statistical analysis on geological characteristics measured within 20 mile square "cells" of known value in Utah, Arizona, and New Mexico to predict the expected value of "cells" in certain areas of Alaska.

Marshall (1964, p. 223-236) used a probabilistic approach to determine the optimal drilling pattern and depth of holes in the search for an orebody that is believed to exist within a certain region. This paper was oriented to surface exploration for massive deposits. It requires that certain probability distributions be known in order for the model to be effective.

Several geological studies into the use of multivariate data in classifying unknown samples have been made. Howd (1964, p. 207-222) and Botbol (1969, p. 1-17) both used a taxonomic approach to determine the true nature of an unknown multivariate characteristic data vector. These methods classify unknown data vectors by obtaining a measure of their "closeness" to a set of data vectors that all come from a known class. This measure is usually the value of some function of the unknown vector, and matrices formed from data vectors of known class.

A considerable amount of work has been done by the U.S.

Bureau of Mines in the statistical analysis of mine assay data, and in the formulation of probabilistic exploration decision models. Most of the latter are directed to the exploration for new deposits. One interesting method of statistical analysis (Koch, 1967, p. 1-64) applied to a vein type of deposit in Mexico, was to obtain trend surfaces for the various metal contents of the veins in three dimensions. The results were presented in the form of plane sections through the surfaces on each working level of the mine and parallel to each vein. The surfaces were used to predict the direction and probable extent of further mineralization along the extensions of the veins.

This method suffers from the drawback that it can only be used effectively if very large and complete sets of data on each characteristic are available. In addition, if strong dependencies exist between characteristics it would probably be necessary to use the trend surfaces of some transformation of the dependent characteristics. This would considerably complicate the analysis and subsequent interpretation. Thus its effectiveness is probably limited to situations in which only a small number of independent characteristics are considered significant.

#### Basic Approach

The basic approach used in this study was to find a subset of characteristics from the original set, that could

be used to distinguish ore bearing from waste veins. This subset of characteristics was used to describe the nature of ore bearing and waste veins, in the form of probability distributions. These descriptions were used to classify data points of unknown origin. Such classifications are not exact, but they can form the basis of a decision rule that can be used to evaluate proposed underground exploration.

The analysis of the problem was kept as general as possible within the limitations of the study. Characteristic data obtained from the Galena mine of the American Smelting and Refining Company in Idaho, was used throughout the study in an illustrative manner.

#### Analysis of the Problem

The reason the problem exists is primarily due to the quantity of information available. Data may have been gathered on dozens of characteristics at hundreds or even thousands of \* points. In its original state, it is virtually impossible to mentally assimilate such a mass of information and make logical decisions based on it.

Before proceeding into the analysis of the problem, the actual nature of the characteristic data must be carefully considered.

\*A data point is defined as the full cross section of the vein normal to strike and dip at some point in the study area.

### Nature of Characteristic Data

The actual set of characteristics initially chosen for use are a matter for geologic or economic determination at the mine under study. However, all characteristics that might have any relevance should be included, since the cost of gathering additional data will only marginally increase the cost of the study; whereas, the exclusion of a significant characteristic could destroy much of the value of the results.

There are four major questions that must be asked about the characteristic data.

1. What is the accuracy of the data, and is this accuracy reasonably uniform over time?

In most mines, data gathering techniques and personnel will change with time, so it is entirely possible that the variance of data gathered may have changed over time. As there is normally no way of checking if this has occurred, it must be assumed that the inherent variance of all data is the same regardless of when or how the information was obtained. Under this assumption, all observations of any characteristic will be considered equally reliable.

2. Are the measurements of characteristics taken at points within an orebody truly representative of that orebody?

If the data points are taken at random, and in sufficient numbers within the orebody, there is no reason to

suppose that these points will not represent a close approximation to the distribution of individual characteristics in that orebody. This will not be true for orebodies from which only a few data points are available.

3. Do all orebodies within the mine come from the same population?

The characteristic data should in some way be a measure of the original deposition controls that resulted in the formation of the orebodies. If these controls were not the same throughout the mine, it would be logical to expect that the characteristic distributions would also be different. In this situation, any distribution that is obtained by combining data from all parts of the mine would be meaningless. To avoid getting such spurious distributions, the data from each orebody, and for each characteristic must be tested to find if it is homogeneous.

4. Under the assumption that all orebodies come from the same population, does the number of observations taken from each orebody affect the nature of the overall characteristic distributions?

The question is basically whether or not the data points are obtained in a random fashion. If they are not, the distribution obtained is conditional on whatever non-random element entered into the sampling mechanism. The only obvious non-random element is that more data points are obtained from large orebodies than from small ones.

This does not invalidate the random assumption, however, because the expected number of data points taken from any orebody in truly random sampling would be proportional to its "size." Thus in the case of the Galena mine in which the veins are generally steeply dipping (60 degrees-90 degrees) and relatively narrow (4-ft. to 8-ft. wide) their "size" could be considered equal to their longitudinal section area taken parallel to the strike of the vein. If the number of data points taken from each orebody was roughly in proportion to its longitudinal section area, the assumption that the sampling was truly random could be made.

#### General Form of the Analysis of Characteristic Data

##### Selection of Significant Characteristics

The first stage of the analysis must determine whether there is a significant difference between groups of data points gathered from known orebodies and waste veins. Any characteristic that does not show this significant difference is obviously of little use for the classification of data points of unknown origin. Such characteristics can be discarded.

This stage should reduce the original characteristic set to a much smaller and more manageable subset. Because non-significant characteristics are rejected at this stage, there is no reason for excluding from the initial set any characteristic that could be conceivably useful.

The most efficient method of selecting the subset of significant characteristics is to carry out a multivariate hypothesis test on the data from the ore and waste samples, using as test statistics the differences between the means of respective characteristics. Using simultaneous confidence intervals about each of these statistics, the significant subset for any given level of  $\alpha$  can be found.

#### Division of the Mine Into Regions of Homogeneous Nature

The second stage of the analysis must determine if all the orebodies within the mine belong to the same population with respect to each significant characteristic.

This can be done using multivariate analysis of variance which can be used to test the hypothesis that only one population exists in the mine. If this hypothesis is rejected at any reasonable significance level, it is possible to obtain simultaneous confidence intervals about the mean of each characteristic of each orebody. These confidence intervals can be used to divide the mine into groups of apparently homogeneous orebodies.

All further analysis beyond the second stage must now be predicated on the assumption that each group is entirely independent of all other groups. In fact, each group must be treated as if it were a separate and independent mine.

#### Characteristic Dependencies

The third stage of the analysis is to check each group of orebodies to find if significant dependencies exist

between the characteristics. If such dependencies exist, they can seriously affect the analysis in later stages.

The standard contingency test can run on all pairs of characteristics. From the results of these tests, dependent sets can be grouped together.

#### Classification of Unknown Data Points

The fourth stage of the analysis is to obtain a method, or methods of comparing unknown data points with the data obtained from known ore bearing and barren veins. There are basically two ways in which this can be done.

(a) To obtain a numerical estimate of the distance in k-space of an unknown data point from the centers of the ore and waste data point clusters.

(b) To find the joint probability distribution of characteristic data obtained from ore bearing veins. This distribution is then used to find the probability that an unknown data point belongs to a population with the same distribution. A similar joint probability distribution can be obtained using characteristic data from waste veins. This can be used in the same way.

Either method will produce estimates of how close an unknown point is to the ore and waste populations. These estimates in themselves are somewhat nebulous, but they can be used to compare the relative value of all such unknown points so that only those with a high likelihood of coming from ore bearing veins are considered.

### Decision Rule Formulation

The fifth stage of the analysis is to convert the information obtained in the fourth stage into a form from which decisions can be made. It will not always be true that the best (closest to the ore bearing population) drill hole intersection will provide the highest expected return. The decision rule should be formulated so that it incorporates what current management objectives are, and it should also provide for sequential decisions to be made as further information becomes available.

THEORETICAL ANALYSIS OF CHARACTERISTIC DATA

Notation

As matrix notation will be used extensively in this section, the following conventions will be adopted.

1. Row vectors will be represented by underlined, primed lower case letters.
2. Column vectors will be represented by underlined lower case letters.
3. Matrices will be represented by underlined upper case letters.

Introduction

The analysis of the problem described in the preceding section can be conveniently broken down into six stages.

These are:

1. Characteristic data acquisition.
2. Selection of significant characteristics.
3. Determination of homogeneity of characteristic data.
4. Determination of dependency between characteristics.
5. Methods of comparison of unknown data points with known populations.
6. Decision rule formulation.

### Characteristic Data Acquisition

The geology of the mine area must be studied in detail to determine which set of measurable vein characteristics could possibly be of use in differentiating between ore bearing and waste veins. The actual number and type of characteristics used is irrelevant, as long as they can be measured both in veins of known nature, and in vein intersections of unknown nature.

As the whole study depends on the characteristic distributions obtained from the ore and waste vein samples, it is essential that these should be close approximations of the true distributions. This implies that the sample sizes should be sufficiently large that the addition of further observations will not change the form, or parameters of the individual characteristic distributions. The exact size of an adequate sample cannot be specified, but if it is large enough to exhibit a definite form of distribution, its parameters,  $\bar{x}$  and  $S^2$  should be good estimators of the true  $\mu$  and  $\sigma^2$ .

The actual acquisition of data should be relatively easy at most mines that have been operating for a number of years. Usually, assay information from drifts, raises and stopes is the only characteristic data immediately available in large quantities. However, if accurate geological plans of the mine workings have been kept, it should be possible to obtain geologic and geometric characteristics at a large

number of points without unduly adding to the cost of the study. All this data will almost certainly be randomly selected, since the actual point at which any observation is made is the outcome of a large number of random processes.

#### Selection of Significant Characteristics

Having selected a set of  $k$  characteristics, it is necessary to find the subset of them that can be used in classifying unknown data points. This is done by testing the null hypothesis that each characteristic has the same distribution for the data points sampled from both ore bearing and waste veins.

In multivariate hypothesis testing, a necessary condition is that the data vectors considered must have a multivariate normal distribution. In general, it is unlikely that this condition will be met for all characteristics. However, it is often possible to transform non-normal characteristics into an approximately normal form. The type of transformation required is largely a matter of trial and error, although it has been found (Hazen, 1967, p. 61-62) that many grade (oz silver/ton, percent copper, etc.) characteristics can be transformed by taking the logarithm of each observation.

Using this assumption of normality, the actual hypotheses made are,

$$H_0 : \underline{\mu}_1 = \underline{\mu}_2$$

$$H_1 : \underline{\mu}_1 \neq \underline{\mu}_2$$

where  $\underline{\mu}_1$  = vector of the true means of the k characteristics obtained from ore veins.

$\underline{\mu}_2$  = vector of the true means of the k characteristics obtained from waste veins.

Consider the case in which the k characteristics or variates (the two terms will be considered synonymous) are being tested, and  $N_1$  and  $N_2$  data points have been collected from the ore bearing and waste veins respectively. To test these hypotheses, a single test statistic of known distribution must be obtained from this data.

The most efficient statistic to use with multivariate data is known as Hotelling's  $T^2$ . This is a multivariate analog of Student's t statistic. A clear exposition of the derivation of this test statistic using Roy's union-intersection principle is given by Morrison (1967, p. 118-120).

Details of the results, and method of use of this test are given below.

Let  $\underline{x}_1'$  = vector of observed means of k-variate data obtained from ore bearing veins.

$\underline{x}_2'$  = vector of observed means of k-variate data obtained from waste veins.

$\underline{A}_1$  = matrix of sum of squares derived from k-variate data from ore bearing veins.

$\underline{A}_2$  = matrix of sum of squares derived from k-variate data from waste veins.

$\underline{S}$  = pooled covariance matrix of all data.

Note that a typical element of the  $\underline{A}_1$  (or  $\underline{A}_2$ ) matrix will be given by

$$a_{jrs} = \sum_{i=1}^{N_j} (x_{jri} - x_{jr.})(x_{jsi} - x_{js.}), \quad j = 1, 2$$

where  $x_{jri}$  =  $i^{\text{th}}$  observation on the  $r^{\text{th}}$  characteristic in the  $j^{\text{th}}$  group.

$x_{jr.}$  = mean of the  $N_j$  observations on the  $r^{\text{th}}$  characteristic in the  $j^{\text{th}}$  group.

The test statistic T is given by

$$T^2 = (\underline{x}_1 - \underline{x}_2)' \underline{S}^{-1} (\underline{x}_1 - \underline{x}_2) \cdot (N_1 N_2) / (N_1 + N_2) \quad (1)$$

where

$$\underline{S} = (\underline{A}_1 + \underline{A}_2) / (N_1 + N_2 - 2)$$

If the null hypothesis ( $H_0$ ) is true, the quantity

$$F = T^2 (N_1 + N_2 - k - 1) / (N_1 + N_2 - 2) \cdot k \quad (2)$$

will have an F distribution with  $k$  and  $(N_1 + N_2 - k - 1)$  degrees of freedom.

The test statistic shown in equation 2 can be easily computed once the matrix  $\underline{S}$  has been inverted. The criteria for the acceptance or rejection of the null hypothesis will be,

Accept  $H_0$  :  $F < F_{\alpha, k, (N_1 + N_2 - k - 1)}$

Reject  $H_0$  :  $F \geq F_{\alpha, k, (N_1 + N_2 - k - 1)}$

where the values of  $F_{\alpha, k, (N_1 N_2 - k - 1)}$  can be obtained from tables of the F distribution (Johnson and Leone, 1964a, p. 467-470).

If  $H_0$  is rejected at some predetermined level of  $\alpha$ , the next stage of the analysis would be to determine the subset of  $p$  characteristics that caused this difference. This is done by forming simultaneous confidence intervals about the vector  $(\underline{x}_1 - \underline{x}_2)$  at some confidence level  $\alpha$ , and rejecting all characteristics whose interval contains zero. For this to be effective, all such intervals must be simultaneous. This means that the  $k$  confidence intervals must be considered as a set, and the length of each member of this set must be such that for a given  $\alpha$ , each individual confidence interval statement will be true. This is done by expanding the size of the individual confidence intervals to cover the worst possible situation.

Consider a non null vector  $\underline{a}$ .

$$\text{Then } t^2(\underline{a}) \leq (\underline{x}_1 - \underline{x}_2 - \underline{\delta})' \underline{S}^{-1} (\underline{x}_1 - \underline{x}_2 - \underline{\delta}) (N_1 N_2) / (N_1 + N_2) \quad (3)$$

$$\text{where } \underline{\delta} = (\underline{\mu}_1 - \underline{\mu}_2)$$

since it can be proven (Morrison, 1967, p. 120) that the right hand side is the maximum possible value of  $t^2(\underline{a})$  for all  $\underline{a}$ . But the right hand side of equation 3 is

$T_{\alpha, k, (N_1 + N_2 - k - 2)}^2$  for some  $\alpha$ , from which the following probability statement can be made:

$$P \left[ \text{all } t^2(\underline{a}) \leq T_{\alpha, k, (N_1 + N_2 - k - 2)}^2 \right] = 1 - \alpha$$

This is equivalent to,

$$P \left[ \max_{\underline{a}} t^2(\underline{a}) \leq T_{\alpha, k, (N_1+N_2-k-2)}^2 \right] = 1 - \alpha \quad (4)$$

Now

$$t^2(\underline{a}) = \frac{(N_1 N_2) \{ \underline{a}'(\underline{x}_1 - \underline{x}_2 - \underline{\delta}) \}^2}{(N_1 + N_2) (\underline{a}' \underline{S} \underline{a})} \quad (5)$$

From which the simultaneous confidence interval can be obtained:

$$\begin{aligned} \underline{a}'(\underline{x}_1 - \underline{x}_2) - \sqrt{\left(\frac{N_1 + N_2}{N_1 N_2}\right) \underline{a}' \underline{S} \underline{a}} \cdot T_{\alpha, k, (N_1 + N_2 - k - 2)} &\leq \underline{a}' \underline{\delta} \\ &\leq \underline{a}'(\underline{x}_1 - \underline{x}_2) + \sqrt{\left(\frac{N_1 + N_2}{N_1 N_2}\right) \underline{a}' \underline{S} \underline{a}} \cdot T_{\alpha, k, (N_1 + N_2 - k - 2)} \end{aligned} \quad (6)$$

The theoretical value of  $T$  for given values of the parameters  $\alpha$ ,  $k$ ,  $N_1$ , and  $N_2$  can be calculated from equation 2. Using a value of  $T$ , a succession of confidence intervals for different  $\underline{a}$  vectors can be computed. The  $\underline{a}$  vectors of most interest are those with one in the  $i^{\text{th}}$  element and zeros elsewhere. These will give the confidence interval about the  $i^{\text{th}}$  element of the  $\underline{\delta}$  vector, which if it includes zero, will call for the rejection of the  $i^{\text{th}}$  characteristic at significance level  $\alpha$ .

By computing successive confidence intervals for each characteristic mean difference with decreasing values of  $\alpha$ , subsets of significant characteristics at different levels of  $\alpha$  are obtained. A critical  $\alpha$  can be selected, and all characteristics significant at this level will be used in further analysis.

In summary, the criteria for acceptance or rejection of a characteristic at some significance level  $\alpha$  are:

Accept characteristic  $i$  :  $U_{\alpha,i} < 0$

or :  $0 < L_{\alpha,i}$

Reject characteristic  $i$  :  $L_{\alpha,i} \leq 0 \leq U_{\alpha,i}$

where  $L_{\alpha,i}$  = lower simultaneous confidence interval bound of characteristic  $i$  at level  $\alpha$ .

$U_{\alpha,i}$  = upper simultaneous confidence interval bound of characteristic  $i$  at level  $\alpha$ .

Usually the subset of significant characteristics will be much smaller than the original set. This will reduce the dimension of the problem in subsequent analysis.

#### Determination of the Homogeneity of the Characteristic Data

If two or more distinct types of orebodies exist in the mine area, it would be useless to consider characteristic distributions formed from the sum of all observations.

As observed, the data points will be naturally grouped by orebodies, or by distinct geographical regions of the mine. It is important to find out if each of the  $p$  significant characteristics has the same distribution within all the natural groupings. If they do not, it is necessary to find which of the groups differ, so that groups with identical distributions can be classed together.

The statistical test required must carry out the following functions:

1. Allow for known inhomogeneities.
2. Determine if unknown inhomogeneities exist.
3. If inhomogeneities exist, it must determine which natural data groups cause them, so that these can be segregated into homogeneous populations.

A possible test to use is multivariate analysis of variance. In order to use this type of test, the assumption must be made that all the p-variate data vectors are independent and have a multivariate normal distribution. This in itself is a considerable limitation of the method, but can be overcome by careful selection of the data format and subsequent normalizing transformations.

The complexity of the method, and its inflexibility cause other limitations. These can be best explained using an example.

In some vein type deposits, a zoning effect is present, which causes certain vein characteristics to change considerably with increasing depth. These changes are merely the effect of the cooling of the mineralizing fluids as they rose upwards from their source, and are not due to an entirely different set of controls. In this case, the data groups are inhomogeneous with regard to depth, but it is still necessary to check if other inhomogeneities exist. To find if this were so, a two way crossed classification

model could be used.

For this model to be effective and easily used, there must be observations in each of the  $l \times q$  cells, where there are  $l$  zones and  $q$  natural data groups (orebodies). Moreover, the number of observation vectors in each cell must either be equal or proportional such that

$$\frac{N_{jt}}{N_{j't}} \text{ is independent of } t \text{ for all pairs } j \text{ and } j'$$

or  $\frac{N_{jt}}{N_{jt'}}$  is independent of  $j$  for all pairs  $t$  and  $t'$ ,

where  $N_{jt}$  = number of data vectors in the  $jt^{\text{th}}$  cell.

Since data points can only be taken where orebodies exist, it will be unlikely that every cell will have observations in it. For example, if the  $r^{\text{th}}$  orebody does not exist in the  $s^{\text{th}}$  zone, no information can be obtained about this part. If only one cell is empty in a given row or column, it is possible to obtain a good estimate of its value using a method given in Johnson and Leone (1964b, p. 71-72). The method given is for univariate data, but it can be directly extended for use with multivariate data.

Let the  $j^{\text{th}}$  cell be the only empty cell in the  $j^{\text{th}}$  row and  $t^{\text{th}}$  column, and  $\underline{x}_{jt}$  be the best estimate of the mean characteristic vector of this cell.

Then

$$\underline{x}_{jt.} = \frac{(\ell \underline{x}_{j..} + q \underline{x}_{.t.} - \underline{x}_{...})}{(\ell-1)(q-1)} \quad (7)$$

where  $\underline{x}_{j..}$  = characteristic mean vector of  $j^{\text{th}}$  zone  
 $\underline{x}_{.t.}$  = characteristic mean vector of  $t^{\text{th}}$  orebody  
 $\underline{x}_{...}$  = characteristic mean vector of all observations.

If more than one cell is empty in a row or column, it is possible to use an iterative procedure to estimate values for the missing characteristic vectors. It is apparent that if a large number of cells are empty, the results obtained from subsequent analysis will be of dubious value.

If the proportionality condition is violated, it can be handled using a method based on the general linear hypothesis given by Johnson and Leone (1967b, p. 66-68). This involves solving for a set of parameters that minimize certain sums of squares. It is fairly involved for the univariate case, and appears to be too unwieldly to use with multivariate data. It would seem preferable to meet the proportionality conditions by randomly discarding data vectors from cells that had an excess number. However, this method could not be used if it resulted in the loss of a large amount of information.

In summary, if one or more known inhomogeneities exist, although there are methods for handling the situation, the complexity of analysis, and required assumptions make the

possibility of obtaining meaningful results very small.

Because of this, the analysis will be restricted to the case in which no known inhomogeneities are present. This can be dealt with using a one way analysis of variance model. If zoning, or some other factor is present, each zone can be treated with this simple model. This of course will not give any information on interaction effects, but is probably the best that can be done in the situation.

Consider the case of a mine in which  $N_j$  observations of  $p$ -variate data have been made on the  $j^{\text{th}}$  orebody. The parametric model for such data is given by Johnson and Leone (1964b, p. 7) as:

$$x_{ijk} = A_i + B_{ij} + Z_{ijk} \quad (8)$$

where  $x_{ijk}$  =  $k^{\text{th}}$  observation of the  $i^{\text{th}}$  variate in  $j^{\text{th}}$  orebody.

$A_i$  = Average value of the  $i^{\text{th}}$  variate.

$B_{ij}$  = Deviation on the  $i^{\text{th}}$  variate due to the  $j^{\text{th}}$  orebody.

$Z_{ijk}$  = Residual error.

and  $\sum_j B_{ij} = 0$

The analogy between univariate and multivariate ANOVA is fairly close. Instead of obtaining the "within groups" and "between groups" sums of squares, matrices containing the same sums of squares and cross products are computed. For  $p$ -variate data, these will be non-singular  $p \times p$

matrices. Theoretically, it is possible for these matrices to be singular when two or more variates have a correlation coefficient of one, but if the variates are carefully selected, this situation will not occur. The test statistic used to check the null hypothesis that there is no difference between the orebodies is formed either from the ratio of the determinants, or from the determinantal equation roots of a function of the two matrices. The use of these test statistics is intuitively appealing since a determinant is a unique valuation of a non-singular matrix, and the ratio of two valuations is directly analagous to univariate ANOVA.

The statistical validity of these test statistics, and their distribution under the null hypothesis was originally established by Wilks (1962, p. 561-564) using the generalized likelihood ratio principle. A somewhat easier method to handle is given by Morrison (1967, p. 159-170), and is based on Roy's union-intersection principle.

The basis of this is to consider the set of all univariate null hypotheses dependent on some arbitrary non-null vector  $\underline{a}$ . For all such vectors, a test statistic  $F(\underline{a})$  can be obtained.

If  $F(\underline{a}) \leq F_{\beta, p, (N-q)}$  the univariate null hypothesis will be accepted at this level of  $\beta$ .

The original multivariate hypothesis will be accepted

if the multivariate test statistic falls within the intersection of all these univariate acceptance regions.

Although this intersection region cannot be directly found, an upper bound for its size is given by the maximum value of  $F(\underline{a})$ .

Thus, the multivariate acceptance region is equivalent to that defined by,

$$\max_{\underline{a}} F(\underline{a}) \leq F_{\beta, p, (N-q)}$$

where  $p$  = number of variates

$N$  = total number of observations

$q$  = number of orebodies.

The maximum value of  $F(\underline{a})$  can be shown (Roy, 1957) to be proportional to the greatest root of the determinantal equation:

$$|\underline{H} - \lambda \underline{E}| = 0 \quad (9)$$

where  $H$  = between groups sum of squares matrix

$E$  = within groups sum of squares matrix

$\lambda$  = eigenvalues of equation 9.

Equation 9 can be re-written as,

$$|\underline{H} \underline{E}^{-1} - \lambda \underline{I}| = 0 \quad (10)$$

so that the greatest eigenvalue can be extracted directly from the  $\underline{H}\underline{E}^{-1}$  matrix. However, the distribution of the eigenvalues of the matrix  $\underline{H}(\underline{H} + \underline{E})^{-1}$  under the null hypothesis are available, so that it is preferable to work with this statistic.

The actual application of this test is relatively simple, and is given below.

Typical elements of the  $\underline{H}$  and  $\underline{E}$  matrices can be obtained from the following formulae:

$$h_{uv} = \sum_{j=1}^q \frac{1}{N_j} \left( \sum_{k=1}^{N_j} x_{ujk} \right) \left( \sum_{k=1}^{N_j} x_{vjk} \right) - \frac{1}{N} \left( \sum_{j=1}^q \sum_{k=1}^{N_j} x_{ujk} \right) \left( \sum_{j=1}^q \sum_{k=1}^{N_j} x_{vjk} \right) \quad (11)$$

$$e_{uv} = \sum_{j=1}^q \sum_{k=1}^{N_j} x_{ujk} x_{vjk} - \sum_{j=1}^q \frac{1}{N_j} \left( \sum_{k=1}^{N_j} x_{ujk} \right) \left( \sum_{k=1}^{N_j} x_{vjk} \right) \quad (12)$$

where  $h_{uv}$  =  $uv^{\text{th}}$  element of  $\underline{H}$  matrix

$e_{uv}$  =  $uv^{\text{th}}$  element of  $\underline{E}$  matrix

$q$  = number of orebodies

$N_j$  = number of observations on the  $j^{\text{th}}$  orebody

$$N = \sum_{j=1}^q N_j$$

These matrices can be computed with little difficulty. The major part of the work comes from having to invert the  $(\underline{H} + \underline{E})$  matrix, product it with the H matrix, and finally extract the largest eigenvalue. Computer programs were written to perform these operations using standard techniques.

Let  $\theta$  = largest eigenvalue of the  $\underline{H}(\underline{H} + \underline{E})^{-1}$  matrix. Then  $\theta$  will have a distribution given by the Heck charts (Morrison, 1967, p. 312-319) with parameters,

$$s = \min (q - 1, p)$$

$$m = (|q - p - 1| - 1)/2$$

$$n = (N - q - p - 1)/2$$

Using the computed value of these parameters, and some level of  $\alpha$  (say 0.01), a critical value  $Z_{\alpha, s, m, n}$  can be read off the relevant chart.

$$\text{If } \theta < Z_{\alpha, s, m, n}$$

the null hypothesis that all the orebodies are the same cannot be rejected at this level of  $\alpha$ . Conversely, it will be rejected if the inequality is reversed.

The extension to a two way crossed classification model can be made without too much difficulty, but as mentioned earlier, it is unlikely that the necessary conditions for its application will ever be met in this type of physical situation.

If the null hypothesis is not rejected, it means that at significance level  $\alpha$ , the data from all orebodies can be considered homogeneous, and no further work is required in this stage of the analysis.

In the event that the null hypothesis is rejected, it is necessary to find out which particular orebodies caused this rejection. This can be done by obtaining simultaneous confidence intervals for the differences of the mean values of the  $i^{\text{th}}$  variate between each pair of orebodies.

The 100 (1 -  $\alpha$ ) percent confidence interval for any linear compound of the k-variate mean differences between the  $r^{\text{th}}$  and  $s^{\text{th}}$  orebodies is given by,

$$\begin{aligned} & \underline{a}'(\underline{x}_r - \underline{x}_s) - \sqrt{\left(\frac{N_r+N_s}{N_r N_s}\right) \underline{a}' \underline{E} \underline{a} \left(\frac{Z_\alpha}{1-Z_\alpha}\right)} \\ & \leq \underline{a}'(\underline{B}_r - \underline{B}_s) \\ & \leq \underline{a}'(\underline{x}_r - \underline{x}_s) + \sqrt{\left(\frac{N_r+N_s}{N_r N_s}\right) \underline{a}' \underline{E} \underline{a} \left(\frac{Z_\alpha}{1-Z_\alpha}\right)} \end{aligned} \quad (13)$$

where  $Z_\alpha = Z_{\alpha, s, m, n}$ , obtained from Heck charts

$\underline{a}$  = arbitrary non-null vector

$\underline{x}_r, \underline{x}_s$  = characteristic mean vectors for the  $r^{\text{th}}$  and  $s^{\text{th}}$  orebodies

$N_r, N_s$  = number of data vectors taken in the  $r^{\text{th}}$  and  $s^{\text{th}}$  orebodies

$\underline{E}$  = within groups sum of squares matrix

$\underline{B}_r, \underline{B}_s$  = vectors of deviations of the  $r^{\text{th}}$  and  $s^{\text{th}}$  orebody from the overall mean.

As in the case of multivariate hypothesis testing of means from two samples, the vector  $\underline{a}$  can be selected such that it can give confidence intervals about the mean differences for each characteristic. If any such interval contains zero, the hypothesis that  $B_{ir}$  and  $B_{is}$  are different must be rejected at this level of  $\alpha$  for the  $i^{\text{th}}$  characteristic.

It can be seen that if the number of orebodies ( $q$ ), and the number of characteristics ( $p$ ) is large, there will be a very large number of individual hypotheses to consider. There will in fact be a total of  $p\binom{q}{2}$ . This could be cumbersome to handle, and can be conveniently reduced by initially using an a vector consisting of all ones. This will give a confidence interval about the sum of all the characteristic mean differences between the  $r^{\text{th}}$  and  $s^{\text{th}}$  orebodies. If this interval includes zero, it is unlikely that there is a significant difference between these orebodies. If the null hypothesis is rejected, each individual characteristic confidence interval can be generated to find which characteristics are contributing to the difference.

In this way, all pairs showing a significant difference can be quickly isolated into similar groups by generating only  $\binom{q}{2}$  confidence intervals. It is not really essential to know what characteristics cause the difference between orebodies, but such information can be geologically valuable.

The results of this section can be summarized as follows:

1. Obtain  $\theta$  from H and E matrices.
2. If  $\theta \leq Z_{\alpha, s, m, n}$ , proceed with next stage of the analysis.

3. If  $\theta > Z_{\alpha, s, m, n}$ , form simultaneous confidence intervals for all  $\binom{q}{2}$  possible combinations of orebodies using a fixed  $\alpha$ , and an  $\underline{a}$  vector such that

$$\underline{a} = (1, 1, \dots, 1)$$

4. Group orebodies such that

$$L_{ij} \leq 0 \leq U_{ij}$$

where  $L_{ij}$  = lower bound on confidence interval of the  $j^{\text{th}}$  orebody pair in the  $i^{\text{th}}$  sub group.

$U_{ij}$  = upper bound on confidence interval of the  $j^{\text{th}}$  orebody pair in the  $i^{\text{th}}$  sub group.

The result of this series of tests will have been either to justify the assumption that all orebodies in the mine are of a homogeneous nature, or to subdivide the mine in groups that do have such a nature. If the latter is true, the original problem will have been split into several smaller problems, each of which must be treated in isolation from the others.

Before proceeding to the next stage, it would be advisable to re-run stage 1 of the analysis (selection of significant characteristics) if the original hypothesis of homogeneity was rejected. This test should be re-run on each of the new sub-groups using the original k-variate data from these groups. This is an essential precaution, for in the original test, the dissimilar populations taken

together may have masked certain significant characteristics.

In this study, it will be assumed that any division of the mine into separate geographic groups by analysis of ore bearing vein data will also hold for waste veins within that region. A similar analysis could be applied to waste vein data. This should either confirm the division made by the first analysis, or show no division at all. If it shows a different division of the mine from that given by the ore data, the whole study must be considered suspect.

#### Determination of Dependency

##### Between Characteristics

Up to this point in the analysis, no conditions have been made requiring the mutual independence of the set of characteristics. If probabilistic methods are to be used, it is of vital importance to determine if any dependencies exist between the characteristics.

The most widely used method for independence testing is the contingency table. The test statistic produced has an approximate Chi Square distribution if the ratio between the number of observations (N) and the number of possible level combinations is sufficiently large.

For a test on characteristics u and v, which have r and s possible discrete levels respectively, if

$$\frac{N}{r \cdot s} \geq 5$$

then the resulting test statistic will have an approximately Chi Square distribution.

Because it is not possible to construct a  $p$ -dimensional contingency table, all possible pairs of the  $p$  characteristics must be tested individually. It is assumed that the  $p$ -variate data has either been gathered in a discrete form, or that it has been partitioned into levels of equal size with respect to each characteristic. The technique will be briefly described below.

Let the data gathered on characteristics  $u$  and  $v$ , have  $r$  and  $s$  possible levels respectively. The contingency table will then have  $r \cdot s$  cells.

Let  $n_{ij}$  = number of observations in the  $ij^{\text{th}}$  cell

$$N_{i.} = \sum_{j=1}^s n_{ij}, \text{ the number of observations in the } i^{\text{th}} \text{ row}$$

$$N_{.j} = \sum_{i=1}^r n_{ij}, \text{ the number of observations in the } j^{\text{th}} \text{ row}$$

$$N_{..} = \sum_{i=1}^r N_{i.} = \sum_{j=1}^s N_{.j}, \text{ the total number of observations.}$$

Under the null hypothesis that the characteristics are mutually independent, the expected number of observations in the  $ij^{\text{th}}$  cell is given by,

$$E_{ij} = \frac{N_{i.} \cdot N_{.j}}{N_{..}}$$

Under this hypothesis, the following test statistic will have an approximately Chi Square distribution with  $(r - 1)(s - 1)$  degrees of freedom if the necessary conditions are met.

$$X^2 = \sum_{i=1}^r \sum_{j=1}^s (n_{ij} - E_{ij})^2 / E_{ij} \quad (14)$$

If  $X^2 < \chi^2_{\alpha, (r-1)(s-1)}$ , the hypothesis that characteristics  $u$  and  $v$  are independent will not be rejected.

If the size of  $N$  is too small to meet  $N/r \cdot s$  condition, the test can still be run, since a large value of  $X^2$  will indicate that dependency exists. However, a small value of  $X^2$  will not prove independence.

This test is not particularly satisfactory, but unless all the characteristic distributions are known to be normal, it is the only feasible one that can be used.

### Methods of Comparison of Unknown Data Points

#### With Known Populations

The following analysis is based on three major assumptions.

1. In the mine or section of the mine under consideration, all ore bearing veins come from the same population with respect to the significant characteristics.

2. All waste veins in the same region will come from the same population.

3. The characteristic distributions of ore and waste veins in this region will be significantly different.

There are basically two methods by which a p-variate data point of unknown origin can be assigned to one of two possible groups.

1. By determining the distance in p-space between the unknown point and the centers of the two populations with which it is being compared. This is generally known as discriminant analysis.

2. By finding the joint probability distributions of each of the known populations, and then assigning the unknown point to the one with which it has the greatest probability of belonging. This is really an extension of univariate hypothesis testing.

### Discriminant Analysis

The p-variate data from groups of known nature can be represented graphically as clusters of data points in p-space. Assuming that the dispersion of these points is not isotropic, there will be a number of orthogonal axes that will maximize the separation of the cluster centers.

Consider a linear transformation on the original set of p-variates giving a new variate  $y$ .

$$\text{Then } y = v_1x_1 + v_2x_2 \dots + v_px_p \quad (15)$$

where  $x_i = i$ th original measured variable

$$\underline{v}' = (v_1, v_2 \dots v_p)$$

In matrix notation, the transformed variable of the  $j^{\text{th}}$  observation on the  $t^{\text{th}}$  group is

$$y_{tj} = \underline{v}' \underline{x}_{tj}$$

In general,  $t$  can be any size, but in this particular study,  $t = 1, 2$  for ore bearing and waste veins respectively.

The vector  $\underline{v}$  that is sought must maximize the ratio

$$\lambda = \frac{\sum_{t=1}^2 n_t (y_{t.} - y_{..})^2}{\sum_{t=1}^2 \sum_{j=1}^{n_t} (y_{tj} - y_{t.})^2} \quad (16)$$

where  $n_t$  = number of observations taken on  $t^{\text{th}}$  group.

It can be seen that this ratio is merely the ratio of the "between groups" to "within groups" sum of squares in terms of the transformed variables. Any  $\underline{v}$  that maximizes  $\lambda$  must necessarily have selected the direction of maximum separation between the group centers.

This ratio can be written in terms of the original variables and the transformation vector  $\underline{v}$ .

$$\lambda = \frac{\underline{v}' \left\{ \sum_{t=1}^2 n_t (\underline{x}_{t.} - \underline{x}_{..})^2 \right\} \underline{v}}{\underline{v}' \left\{ \sum_{t=1}^2 \sum_{j=1}^{n_t} (\underline{x}_{tj} - \underline{x}_{t.})^2 \right\} \underline{v}} \quad (17)$$

Now  $(\underline{x}_{t.} - \underline{x}_{..})$  and  $(\underline{x}_{tj} - \underline{x}_{t.})$  are both  $p \times 1$  matrices, and their squares —  $(\underline{x}_{t.} - \underline{x}_{..})(\underline{x}_{t.} - \underline{x}_{..})'$  and  $(\underline{x}_{tj} - \underline{x}_{t.})(\underline{x}_{tj} - \underline{x}_{t.})'$  — are  $p \times p$  matrices. These can be replaced by  $\underline{B}_t$  and  $\underline{W}_{tj}$  respectively.

Then,

$$\lambda = \frac{\underline{v}' \left\{ \sum_{t=1}^2 n_t \underline{B}_t \right\} \underline{v}}{\underline{v}' \left\{ \sum_{t=1}^2 \sum_{j=1}^{n_t} \underline{W}_{tj} \right\} \underline{v}} \quad (18)$$

Since addition holds for matrices of the same dimensions,

$$\lambda = \frac{\underline{v}' \underline{B} \underline{v}}{\underline{v}' \underline{W} \underline{v}} \quad (19)$$

where  $\underline{B} = \sum_{t=1}^2 n_t \underline{B}_t$

$$\underline{W} = \sum_{t=1}^2 \sum_{j=1}^{n_t} \underline{W}_{tj}$$

The  $rs^{\text{th}}$  elements of these matrices are given by,

$$b_{rs} = \sum_{t=1}^2 n_t (x_{rt.} - x_{r..})(x_{st.} - x_{s..})$$

$$w_{rs} = \sum_{t=1}^2 \sum_{j=1}^{n_t} (x_{rtj} - x_{rt.})(x_{stj} - x_{st.})$$

This expression can be maximized by taking partial derivatives of  $\lambda$  with respect to each element of the  $\underline{v}$  vector, and setting the resultant equations equal to zero. A typical example of such an equation for bivariate data would be,

$$\frac{\partial \lambda}{\partial v_1} = \{2v_1 b_{11} + v_2 (b_{12} + b_{21})\} \{\underline{v}' \underline{W} \underline{v}\} - \{2v_1 w_{11} + v_2 (w_{12} + w_{21})\} \{\underline{v}' \underline{B} \underline{v}\} = 0 \quad (21)$$

(Note that with bivariate data,  $\underline{B}$  and  $\underline{W}$  would be 2 x 2 matrices and  $\underline{v}$  would be a 2 x 1 vector).

Since both the B and W matrices will always be symmetric, for all i and j

$$b_{ij} = b_{ji} \text{ and } w_{ij} = w_{ji}$$

Thus for the 2 x 2 example, the partial derivative equations will reduce to

$$(v_1 b_{11} + v_2 b_{12})(\underline{v}' \underline{W} \underline{v}) - (v_1 w_{11} + v_2 w_{12})(\underline{v}' \underline{B} \underline{v}) = 0$$

$$(v_1 b_{12} + v_2 b_{22})(\underline{v}' \underline{W} \underline{v}) - (v_1 w_{12} + v_2 w_{22})(\underline{v}' \underline{B} \underline{v}) = 0$$

These two equations can be combined in matrix form giving,

$$(\underline{B} \underline{v})(\underline{v}' \underline{W} \underline{v}) - (\underline{W} \underline{v})(\underline{v}' \underline{B} \underline{v}) = 0 \quad (21)$$

Since both  $(\underline{v}' \underline{W} \underline{v})$  and  $(\underline{v}' \underline{B} \underline{v})$  are scalars, it is permissible to divide through by the former to give,

$$(\underline{B} - \lambda \underline{W}) \underline{v} = 0 \quad (22)$$

From this it can be seen that  $\lambda$  is a root of the determinantal equation,

$$|\underline{W}^{-1} \underline{B} - \lambda \underline{I}| = 0 \quad (23)$$

and  $\underline{v}$  is its associated eigenvector. The largest of these will be called the principal eigenvector ( $\underline{v}_1$ ), and the principal eigenvalue ( $\lambda_1$ ). The number of non-zero eigenvalues in the  $\underline{W}^{-1} \underline{B}$  matrix will be equal to the matrix rank. This is given by Harris (1965, p. C21) to be the minimum of  $(t-1)$  and  $p$ .

In this study where only two populations (ore and waste)

are considered, the rank of the  $\underline{W}^{-1}\underline{B}$  matrix will always be one, no matter how many variates are used.

The extraction of these from the  $\underline{W}^{-1}\underline{B}$  matrix does not necessarily exhaust its possibilities for discrimination between the different populations if its rank is greater than one. Using the property of symmetric matrices that,

$$\sum_{i=1}^p \lambda_i = \text{Tr} \cdot (\underline{W}^{-1}\underline{B}) \quad (24)$$

if  $\lambda_1 / \text{Tr} \cdot \underline{W}^{-1}\underline{B} \rightarrow 1$  then  $\lambda_1$  will account for most of the scatter of the data points, and it will not be necessary to extract further eigenvalues and vectors from the matrix. If the ratio is significantly less than one, it may be necessary to extract at least one more eigenvalue and vector. The question arises as to how to judge whether the ratio is sufficiently close to one. Rao (1965, p. 370-378) gives a statistical method of determining when to stop. He showed that the function,

$$X^2 = \left\{ \sum_{t=1}^G kn_t - \frac{1}{2}(t + p) \right\} \{ \ln(1 + E_i) \} \quad (25)$$

had an approximate Chi Square distribution with  $(t + p - 2i)$  degrees of freedom,

where  $p$  = number of original variates

$t$  = number of groups

$n_t$  = number of observations made on  $t^{\text{th}}$  group

$E_i$  =  $i^{\text{th}}$  eigenvalue of the  $\underline{W}^{-1}\underline{B}$  matrix.

$\chi^2$  has the Chi Square distribution under the null hypothesis that there is no difference between the groups.

Thus, if

$$\chi^2 > \chi^2_{\alpha, (t+p-2i)}$$

the  $i^{\text{th}}$  discriminant function can be used. This test must be applied sequentially starting at the principal eigenvalue using some fixed  $\alpha$ . If the  $(i + 1)^{\text{th}}$  discriminant function is rejected, all subsequent functions must also be rejected.

The result of the foregoing transformation will have been to change the  $p$ -variate data into  $i$ -variate data, where in most cases  $i$  is less than three. The transformations are simply made by applying equation 15 for each significant  $\underline{v}_i$  vector to each data point.

Using this  $i$ -variate data, a mean vector and covariance matrix can be obtained for each population. The generalized distance in  $i$ -space separating a data point of unknown nature from each of the two population centers can be calculated from the following relation:

$$Z_{tj}^2 = (\underline{y}_j - \underline{y}_t)' D_t^{-1} (\underline{y}_j - \underline{y}_t) \quad (26)$$

where  $\underline{y}_t$  = mean discriminant vector of the  $t^{\text{th}}$  group

$\underline{y}_j$  = discriminant vector of the  $j^{\text{th}}$  unknown data point

$D_t^{-1}$  = inverse of the discriminant covariance matrix of the  $t^{\text{th}}$  group

$z_{tj}^2$  = squared distance between the center of the  $t^{\text{th}}$  group and the  $j^{\text{th}}$  unknown point.

It is possible to obtain an estimate of this distance in  $p$ -space in terms of the original  $p$  variables, without making the orthogonal transformation. There are several disadvantages to doing this. Although both methods should give the same classification to an unknown point, it is of great interest to know what is the probability of misclassification. This can only be done if the distribution of  $Z$  is known, and in practice, this means that unless it has a multivariate normal distribution with known parameters, it is unlikely that it can be found.

Good estimates of the distribution parameters can be obtained if large enough samples are taken from each population, but the normality condition is not as easy to satisfy. It is possible to use logarithmic or polynomial transformations on the original data to give it a more normal form. However, it has been shown that the discriminant variables will more closely satisfy the normality assumption than the original variables (Cooley and Lohnes, 1962), regardless of their original form. This smoothing to normality effect of the transformation is unlikely to be very pronounced unless there are a fairly large number of original variables used in forming the  $\underline{W}^{-1}\underline{B}$  matrix.

An additional benefit obtained from the discriminant transformation is that the discriminant functions will show how the original variables relate to the new variables, and could possibly give some geological insight about the nature of the new variables. The discriminant transformation is really a form of principal component analysis on the  $\underline{W}^{-1}\underline{B}$  matrix.

The results of this sub-section can be summarized as follows:

1. The significant eigenvalues ( $\lambda_i$ ) and eigenvectors ( $\underline{v}_i$ ) are obtained from the  $\underline{W}^{-1}\underline{B}$  matrix.
2. All data points from known groups, and from unknown intersections, are transformed into the new y variables using equation 15.
3. The mean vectors and covariance matrices for the known groups are computed in terms of the new variables.
4. The  $i^{\text{th}}$  dimensional distance between each unknown point and each of the group centers is computed. The unknown points are classified as belonging to the group whose center they are closest to.

#### Computation of Misclassification Probabilities

These are calculated using the transformed discriminant data, and are dependent on the following three assumptions:

1. The vector of  $\underline{y}_t$  is a good estimate of the mean vector  $\underline{\mu}_t$  of the  $t^{\text{th}}$  group.
2. The covariance matrix  $\underline{D}_t$  is a good estimate of the covariance matrix  $\underline{\Sigma}_t$  of the  $t^{\text{th}}$  group.
3. The  $t^{\text{th}}$  group has a multivariate normal distribution with parameters  $\underline{\mu}_t$  and  $\underline{\Sigma}_t$ .

The probabilities required in formulating a decision rule can be obtained from the misclassification probabilities using Bayes Theorem.

Let A = the event that the vein is ore bearing.

B = the event that the vein is waste.

$Y_j$  = the event that the  $j^{\text{th}}$  data point has values equal to  $\underline{y}_j$ .

Then the misclassification probabilities of the  $j^{\text{th}}$  unknown data point are,

$$P \left[ \underline{Y}_j \mid B \right]$$

if it was classified as ore when in reality the vein was waste, and

$$P \left[ \underline{Y}_j \mid A \right]$$

if it was classified as waste when in reality the vein was ore.

In the general case with  $i$ -variate discriminant data, there are two ways of calculating these probabilities. Both of these involve the use of the multivariate normal density function. This was given by Wilks (1961, p. 163-

166) as:

$$f_t(y_{1j}, \dots, y_{ij}) = \frac{e^{-\frac{1}{2}(Z_{tj}^2)}}{(2\pi)^{i/2} |D_t|^{1/2}} \quad (27)$$

where  $Z_{tj}^2 = (\underline{y}_j - \underline{y}_{t.})' D_t^{-1} (\underline{y}_j - \underline{y}_{t.})$

$\underline{y}_j = j^{\text{th}}$  discriminant data vector

$\underline{y}_{t.} =$  mean vector of  $t^{\text{th}}$  group

$|D_t| =$  determinant of the  $t^{\text{th}}$  group covariance matrix

$i =$  number of variates.

It is possible to obtain measures of the misclassification probabilities by considering only the densities of the various group distributions at the point  $J$ , the position of the  $j^{\text{th}}$  unknown data vector. These will be given by

$$P \left[ \underline{Y}_j \mid T \right] = f_t(y_{1j}, \dots, y_{ij}) dy_1, \dots, dy_i \quad (28)$$

where  $T =$  event that vein belongs to the  $t^{\text{th}}$  group

$dy_1, \dots, dy_i =$  infinitesimal multipliers used to give the density function mass at point  $J$ .

There will be a misclassification probability for each group.

This is not a very satisfactory approach since these probabilities will be very sensitive to errors in data observations (of the  $j$  vectors), and also to inaccuracies in the estimated distribution parameters.

A better method would be to use the point  $J$  to characterize the set of all other points into two subsets. The probability masses of each of these sets could then be computed for each group hypothesis. Although this is theoretically possible, the misclassification probability would be in the form,

$$P \left[ \overline{Y}_j \mid \overline{T} \right] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{y_{1j}} f_t(y_1, \dots, y_i) dy_1 \dots dy_i \quad (29)$$

where the hyper-plane through the point  $y_{1j}$ , orthogonal to all other axes, divides the set of all points into two subsets.

Equation 29 cannot be integrated to obtain an answer in closed form, and although it is possible to numerically integrate it using Hermite-Gauss integration, the accuracy of such results often leave much to be desired.

As this study only deals with univariate discriminant data, the multivariate case will not be considered. In the univariate case, the misclassification probabilities are:

$$P \left[ \overline{Y} \leq y_j \mid \overline{A} \right] = \int_{-\infty}^{y_j} f_A(y) dy \quad (30)$$

$$P \left[ \overline{Y} > y_j \mid \overline{B} \right] = \int_{y_j}^{\infty} f_B(y) dy \quad (31)$$

Figure 1 illustrates the use of this method. These probabilities can either be found from tables of the normal distribution function (Johnson and Leone, 1964a, p. 460-461),

UNIVARIATE DISCRIMINANT  
MISCLASSIFICATION PROBABILITIES

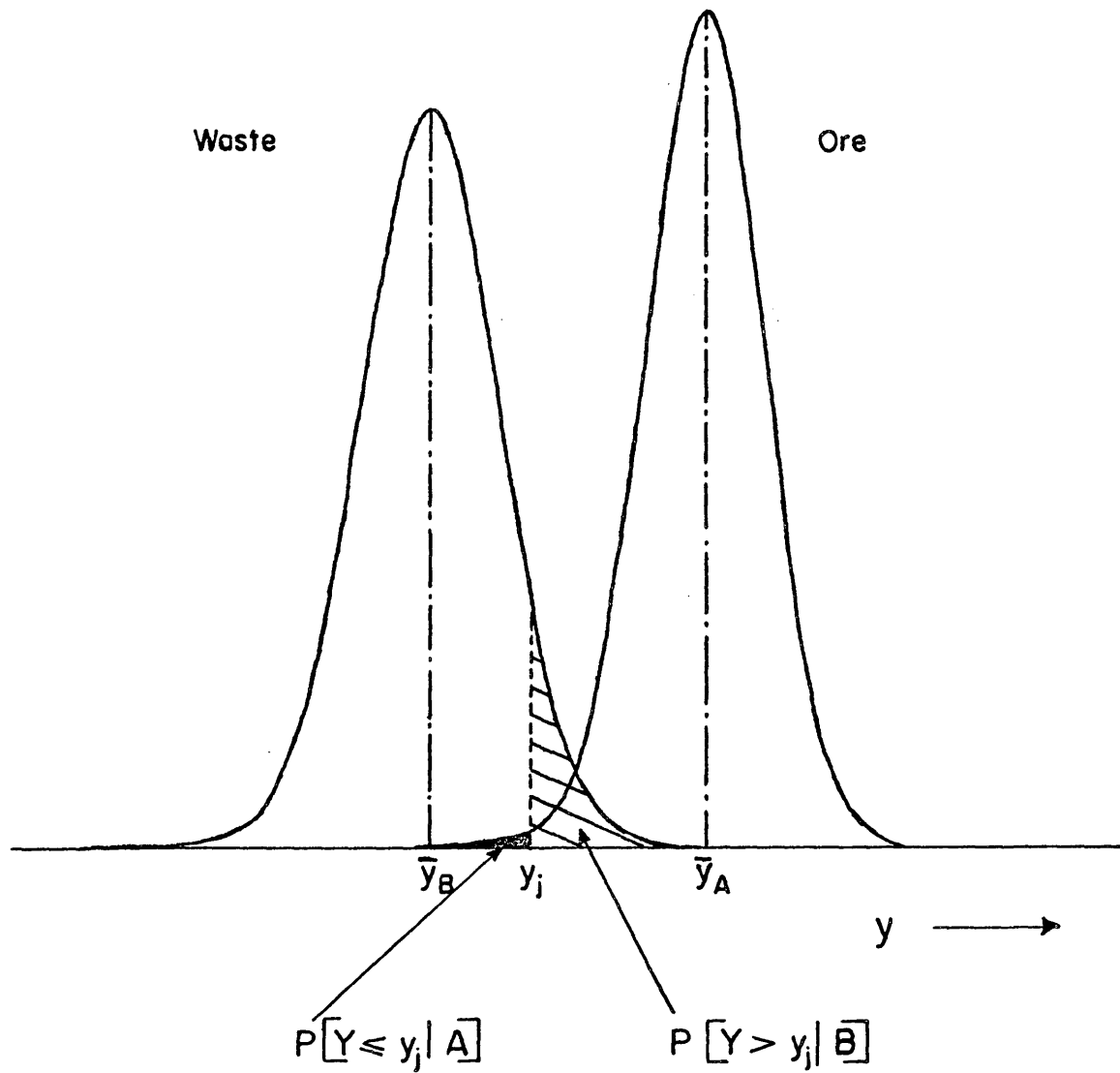


FIGURE 1

or by numerical integration of equations 30 and 31 over the relevant domain. All that is required for this are the discriminant distribution parameters ( $y_{t.}$ ,  $s_t^2$ ), and the discriminant value ( $y_j$ ) of the unknown data point.

The probabilities given by equations 30 and 31 can be used to compute the decision rule probabilities.

### Analysis of Joint Probability Distributions

The use of the joint probability distributions of the original variates to obtain the error probabilities is intuitively more appealing than using discriminant analysis. It has the severe limitation that these joint distributions can only be found if all variates are mutually independent. If the independence criterion is not met, the number of data points needed to obtain the joint distribution function will be prohibitively large.

If the characteristics are truly independent, very simple statements can be made about the probability of an unknown data point belonging to the ore bearing or waste vein groups.

Let there be  $p$  significant independent characteristics, and let the values of each of these be ordered such that

$$P_A \left[ X_i > x_i \right] > P_B \left[ X_i > x_i \right]$$

where  $X_i$  = the value of the  $i^{\text{th}}$  characteristic, and is a random variable

$x_i$  = value of the  $i^{\text{th}}$  characteristic at a point.

Then for any data vector  $\underline{x}$  from an unknown point J,

$$\begin{aligned} P \left[ X_j \mid B \right] &= P_B \left[ X_1 > x_1, X_2 > x_2, \dots, X_p > x_p \right] \\ &= P_B \left[ X_1 > x_1 \right] \cdot P_B \left[ X_2 > x_2 \right] \dots P_B \left[ X_p > x_p \right] \end{aligned} \quad (32)$$

Similarly,

$$P \left[ X_j \mid A \right] = P_A \left[ X_1 < x_1 \right] \cdot P_A \left[ X_2 < x_2 \right] \dots P_A \left[ X_p < x_p \right] \quad (33)$$

where  $X_j$  is the event described within square brackets on the right hand side.

By knowing each individual characteristic distribution from the ore bearing and waste groups, these probabilities can be easily found. These probabilities are in fact the best levels of significance at which the alternative hypotheses can be rejected. For classification, the hypothesis with the lowest error probability will be accepted, and the data point assigned to this group. These probabilities can also be used to obtain the decision rule probabilities.

In practice, the probability of being able to prove that all  $p$  significant characteristics are mutually independent is just about zero. However, it is possible that some of them will be independent, and other subsets show strong inter-related dependency, but these can still be independent from the rest. In this case, it is possible to construct a single new mutually independent variable

from each dependent subset.

This can either be done using regression analysis, or on a heuristic basis. From either method, a relationship between the new and old variables will be obtained, and from this, the distribution of the new variable can be computed. Regression analysis should probably be used if the dependent variables are highly correlated. If they are not, it would seem preferable to combine them on a heuristic basis that takes into account their nature, and relative geologic importance.

If the significant characteristics are mutually dependent, and if they each have a normal distribution, it is possible to compute the misclassification probabilities directly from the multivariate normal distribution function. This can be done by transforming the multivariate normal distribution integral into the Hermite-Gauss form, which can then be numerically integrated over the required domains. A bivariate example of this technique is shown below. Wilks (1962, p. 158-161) gave the bivariate normal distribution function as:

$$F(r_1, r_2) = \frac{1}{2\pi|D|^{1/2}} \int_{-\infty}^{r_1} \int_{-\infty}^{r_2} e^{-\frac{1}{2}Q(x_1, x_2)} dx_1 dx_2 \quad (34)$$

where  $|D|$  = determinant of the covariance matrix

$$Q(x_1, x_2) = (\underline{x} - \underline{\mu})' D^{-1} (\underline{x} - \underline{\mu})$$

Equation 34 cannot be integrated to give an answer in

closed form, but it can be transformed into a Hermite-Gauss integral, which can then be numerically integrated. This integral is given by:

$$\int_{-\infty}^{\infty} e^{-x^2} G(x) dx = \sum_{i=1}^n w_i G(X_i).$$

The transformation to this form was carried out in the following manner.

Let  $(\mu_1, \mu_2)_t$  = mean vector of the  $t^{\text{th}}$  population

$D_t$  = covariance matrix of the  $t^{\text{th}}$  population

$$= \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$$

$D_t^{-1}$  = inverse of  $D_t$

$$= \begin{pmatrix} d^{11} & d^{12} \\ d^{21} & d^{22} \end{pmatrix}$$

$$|D_t| = (d_{11} d_{22} - d_{12} d_{21}).$$

$$\text{Then, } F_t(r_1, r_2) = \frac{1}{2\pi |D_t|^{\frac{1}{2}}} \int_{-\infty}^{r_1} \int_{-\infty}^{r_2} e^{-\frac{1}{2} Q_t(x_1, x_2)} dx_1 dx_2 \quad (35)$$

where

$$Q_t(x_1, x_2) = (x_1 - \mu_1)^2 d^{11} + 2(x_1 - \mu_1)(x_2 - \mu_2) d^{21} + (x_2 - \mu_2)^2 d^{22}.$$

Making the transformations

$$x_1^2 = 1/2(x_1 - \mu_1)^2 d^{11}, \quad x_2^2 = 1/2(x_2 - \mu_2)^2 d^{22}$$

$$x_1 = (x_1 - \mu_1) \sqrt{\frac{d^{11}}{2}}, \quad x_2 = (x_2 - \mu_2) \sqrt{\frac{d^{22}}{2}}$$

$$dx_1 = \sqrt{\frac{2}{d^{11}}} \cdot dX_1 \quad , \quad dx_2 = \sqrt{\frac{2}{d^{22}}} \cdot dX_2$$

$$\begin{aligned} \text{and if } H(x_1, x_2) &= e^{-(x_1 - \mu_1)(x_2 - \mu_2)d^{21}} && \text{for } x_1 \leq r_1 \\ & && \text{and } x_2 \leq r_2 \\ &= 0 && \text{for } x_1 > r_1 \\ & && \text{and } x_2 > r_2 \end{aligned}$$

$$\begin{aligned} \text{then } G(X_1, X_2) &= e^{-X_1 X_2 (2d^{21} / \sqrt{d^{11} d^{22}})} \\ \text{for } X_1 &\leq (r_1 - \mu_1) \sqrt{\frac{d^{11}}{2}} \\ X_2 &\leq (r_2 - \mu_2) \sqrt{\frac{d^{22}}{2}} \end{aligned} \tag{35a}$$

$$\begin{aligned} \text{and } G(X_1, X_2) &= 0 \\ \text{for } X_1 &> (r_1 - \mu_1) \sqrt{\frac{d^{11}}{2}} \\ X_2 &> (r_2 - \mu_2) \sqrt{\frac{d^{22}}{2}} . \end{aligned} \tag{35b}$$

Now,

$$F_r(r_1, r_2) = \frac{1}{\pi(d^{11}d^{22}|D_t|)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-X_1^2} \int_{-\infty}^{\infty} e^{-X_2^2} G(X_1, X_2) dX_1 dX_2 \tag{36}$$

subject to conditions 35a and 35b.

This equation is equivalent to (35), but as it is in the Hermite-Gauss form it can be numerically integrated since,

$$F_t(r_1, r_2) = \frac{1}{\pi(d^{11}d^{22}|D_t|)^{\frac{1}{2}}} \sum_{i=1}^n \sum_{j=1}^n w_i w_j \cdot G(X_i, X_j) \tag{37}$$

where  $n$  = order of the polynomial used to compute the  $w_i$ 's.

If conditions 35a are violated for any  $x_i$ , the resultant value of

$$G(X_i, X_j) = 0.$$

The approximation becomes more exact as larger  $n$ 's are used. Values for the  $X$ 's and  $w$ 's were obtained from tables prepared by the National Bureau of Standards (1965, p. 924). A computer program was written to evaluate the misclassification probabilities of unknown data points using equation 37.

This method can be extended to multivariate data of dimension greater than two. It would appear inefficient to do this, since its accuracy is critically dependent on the exact normality of all the variate's distribution, a condition that is seldom met precisely. In addition, the dimension of the problem cannot be reduced as in discriminant analysis, which means that computational requirements will increase geometrically with the increase in number of variates.

#### Decision Rule Formulation

The classification analysis in the previous section will have assigned an unknown data point to one of the two groups, and hopefully, a reasonable estimate of the misclassification probabilities will have also been obtained.

The problem is to decide whether or not the unknown vein is worth further exploration effort. This decision must be made in consideration of current management objectives. These may be maximum expected return, minimum risk, minimum exploration expenditure, or some other aim.

To formulate a decision rule that will satisfy any one of these objectives, the following additional data is necessary:

1. The distributions of all known orebodies in terms of size (target area) and value (gross metal content).

2. Unit costs of development operations such as diamond drilling, drifting and raising.

In the simple case where the unknown vein is open on strike and dip, the mean of the distributions of vein size and value can be taken as the expected size and value of the unknown vein. If previous exploration has delimited the bounds of the unknown orebody, truncated forms of these distributions must be used.

It would be of great interest and use if the original characteristic data could be used to distinguish between orebodies of different sizes. If this could be done, three or more multivariate distributions would be obtained, one from the waste group, and the others from the different orebody size groups. The classification analysis would be the same, but multiple misclassification probabilities would be obtained. If such differences

exist, they would probably show up in the ANOVA test.

The misclassification probabilities cannot be used directly in the decision rule, but by using Bayes Theorem the decision rule probabilities can be found.

Let  $\underline{x}_j$  = characteristic vector of the  $j^{\text{th}}$  unknown data point

$\underline{X}$  = vector of characteristic random variables.

Then,

$$P[A | \underline{X} < \underline{x}_j] = \frac{P[\underline{X} < \underline{x}_j | A] \cdot P[A]}{P[\underline{X} < \underline{x}_j | A] \cdot P[A] + P[\underline{X} < \underline{x}_j | B] \cdot P[B]} \quad (38)$$

and

$$P[B | \underline{X} > \underline{x}_j] = \frac{P[\underline{X} > \underline{x}_j | B] \cdot P[B]}{P[\underline{X} > \underline{x}_j | A] \cdot P[A] + P[\underline{X} > \underline{x}_j | B] \cdot P[B]} \quad (39)$$

In most instances, the a priori probabilities  $P[A]$  and  $P[B]$  are either unknown, or only very hazy subjective estimates of them are available. In the Galena Mine study, it was not clear what could be used as even approximate values of these probabilities. To overcome this difficulty it was assumed that both these probabilities were equal to one-half. This assumption means that the resultant decision rule probabilities, and consequently the expected return on exploration investment, can only be used in a comparative manner.

Equations 38 and 39 now reduce to:

$$P[A|\underline{X} \leq \underline{x}_j] = \frac{P[\underline{X} \leq \underline{x}_j | A]}{P[\underline{X} \leq \underline{x}_j | A] + P[\underline{X} \leq \underline{x}_j | B]} \quad (40)$$

$$P[B|\underline{X} > \underline{x}_j] = \frac{P[\underline{X} > \underline{x}_j | B]}{P[\underline{X} > \underline{x}_j | A] + P[\underline{X} > \underline{x}_j | B]} \quad (41)$$

The decision rule probabilities are merely the normalized form of these conditional probabilities. These are given by:

$$P[A|J] = \frac{P[A|\underline{X} \leq \underline{x}_j]}{P[A|\underline{X} \leq \underline{x}_j] + P[B|\underline{X} > \underline{x}_j]} \quad (42)$$

$$P[B|J] = \frac{P[B|\underline{X} > \underline{x}_j]}{P[A|\underline{X} \leq \underline{x}_j] + P[B|\underline{X} > \underline{x}_j]} \quad (43)$$

where the event  $J$  describes the division of the set of all possible points by the vector  $\underline{x}_j$  into two subsets.

The expected orebody value,  $\bar{v}$  can be calculated from,

$$\bar{v} = \frac{U_v}{\sum_{t=L_v} t \cdot p_v(t)} \quad (44)$$

where:

$U_v$  = upper bound of orebody value,

$L_v$  = lower bound of orebody value,

$p_v(t)$  = probability mass function of orebody value,

$t$  =  $t^{\text{th}}$  value class of orebodies.

If a known bound on orebody size exists in any particular case,  $U_v$  must be replaced in equation 44 by this bound before  $\bar{v}$  is computed. This assumes that size and value are directly related, although this may not always be true.

The exploration costs ( $D_j$ ) associated with a proposal to explore the  $j^{\text{th}}$  unknown data point can be evaluated fairly exactly. The expected gross return from the exploration of the  $j^{\text{th}}$  unknown data point is given by:

$$E \left[ \text{GR}_j \right] = P \left[ A|J \right] \cdot \bar{v} - P \left[ B|J \right] \cdot D_j. \quad (45)$$

If mining costs are not the same in all parts of the mine,  $E \left[ \text{GR} \right]$  must be multiplied by some scaling factor, so that the expected returns from all exploration proposals can be directly compared.

Due to the uncertainty about the a priori probabilities, these expected returns cannot be considered to be a true measure of the actual return from exploration investment. Thus, if maximum expected return is the current management objective from exploration investment, the proposal that fulfills this criterion can be selected. The results obtained from equation 45 can also be used to select exploration proposals that meet other management objectives.

APPLICATION OF THEORETICAL RESULTS

Data obtained from an underground silver-copper mine in Idaho, was used to test the method outlined by the foregoing analysis.

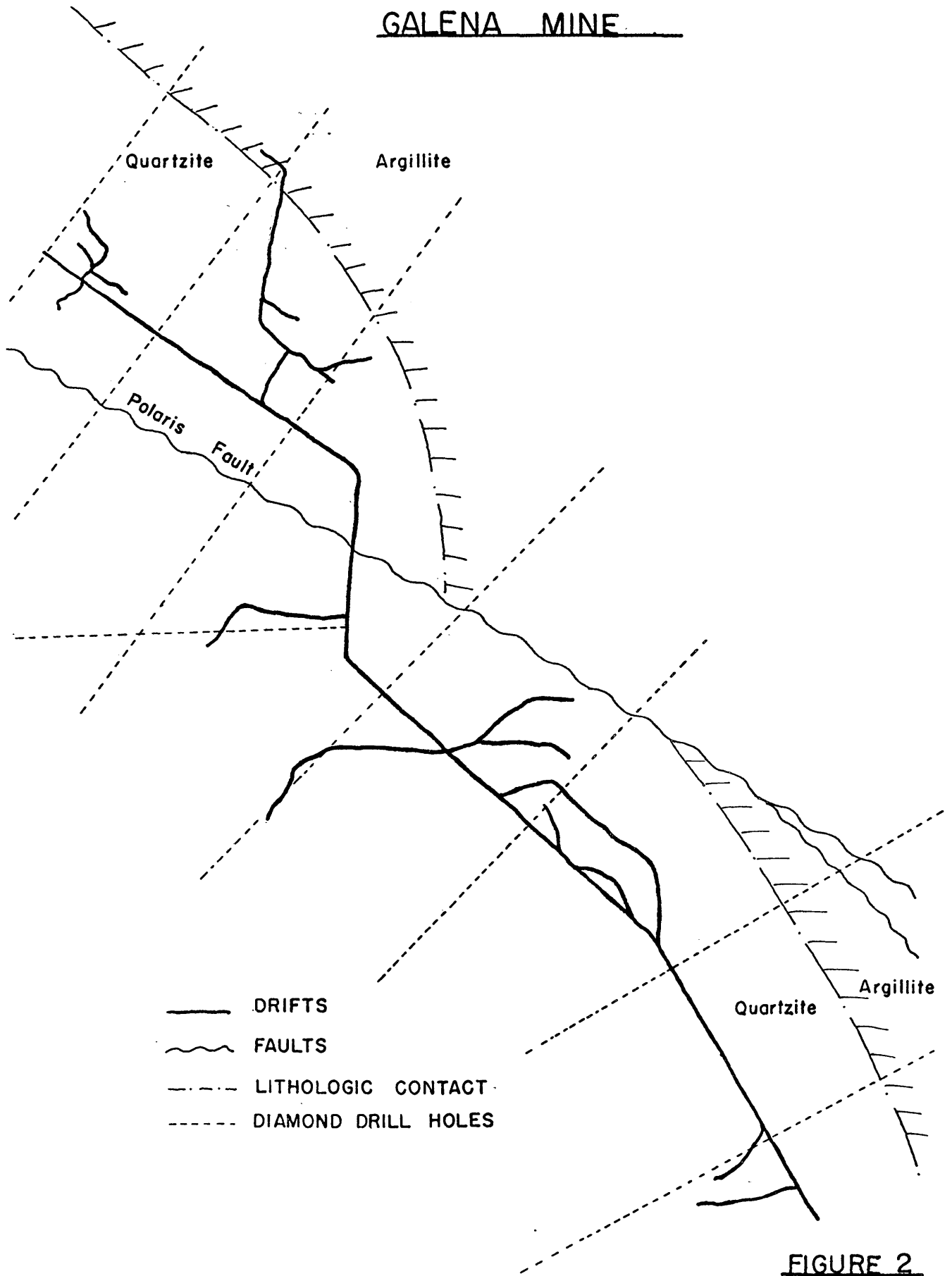
Description of the Galena Mine

The Galena Mine is situated close to the town of Wallace in northern Idaho. The principal economic metals mined are silver and copper, both of which occur mainly in the mineral tetrahedrite. Most of this mineralization is found in narrow veins intersecting bedded quartzites. These veins appear to have originally been tension fractures caused by a major fault (the Polaris fault) that runs through the mine area. This fault is thought to have been the major source of mineralization. A few economic veins have been found in argillites which underlie the quartzites. Figure 2 shows a generalized plan view of a typical level of the mine.

At present the mine workings extend about 10,000 ft laterally and 1,500 ft vertically. There are approximately 32 veins that are thought, or known to contain ore, and numerous other veins that apparently contain only waste.

Orebody sizes (in terms of longitudinal section area)

PLAN VIEW OF A TYPICAL LEVEL  
GALENA MINE



range from 400-sq ft to 1,400,000-sq ft. Vein widths vary from a few inches to 12-ft or more, but probably average about 5-ft.

Preliminary exploration in the mine is carried out by diamond drilling at regular intervals from long drifts, driven roughly parallel to the lithologic contact between the quartzites and argillites. The veins intersected in these holes form potential exploration targets.

#### Characteristic Data

An initial set of 24 vein characteristics was selected. Data had been gathered on these characteristics from 60 points in orebodies, and 21 points in waste veins for use in a classification analysis based on a method developed by Botbol (1969, p. 1-17). It was found that seven of the ore points were taken from veins that were no longer included in the mine's ore reserves. Consequently, these points were classed as waste, and the analysis was then carried out with sample sizes of 53 and 28 from the ore and waste populations respectively.

All this data was in binary form, which rendered it generally unsuitable for the type of analysis outlined earlier. However, all of these characteristics must possess continuous distributions, which could be found if data were gathered in a sufficiently detailed manner. These would not necessarily be normal distributions, but such data probably

could be transformed into an approximately normal form using transformations similar to those used by Harris (1968, p. 308).

The analysis of this characteristic data was used mainly for illustrative purposes, since its form was so far from normal that the results obtained from it cannot be given much weight.

Because the limitations of this set of data points were realized, a second set was collected on the only two characteristics for which information was readily available. Data points were selected at random from ore and waste veins, and silver and copper contents of the veins at these points were recorded. It was found that these characteristics had an approximate log normal distribution for both the ore and waste populations. Thus with this data set, the normality and random selection conditions were met, since all subsequent operations were performed on the log-transformed data.

#### First Data Set

The 24 characteristics and the results obtained from the 81 data points of this set are listed below. In all further discussion these characteristics will be referred to by number.

No. 1 - Distance of Vein Intersection from the Quartzite/  
Argillite Contact

This was measured as the shortest horizontal distance from the vein intersection to the contact.

<u>Code</u>	<u>Distance to Contact (ft)</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	0 - 100	3	5.6	4	14.3
2	100 - 200	7	13.3	3	10.7
3	200 - 300	11	20.8	4	14.3
4	300 - 400	11	20.8	4	14.3
5	400 - 500	4	7.6	4	14.3
6	500 - 600	8	15.1	3	10.7
7	600 - 700	4	7.6	1	3.6
8	700 - 800	2	3.6	2	7.1
9	> 800	3	5.6	3	10.7
Mean distance		440 ft.		450 ft.	
Coded mean		4.40		4.50	

No. 2 - Depth of intersection

The depth of each point was taken as the level from which it had come.

<u>Code</u>	<u>Depth (ft)</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	2400	0	0.0	0	0.0
2	2800	4	7.6	1	3.6
3	3000	5	9.4	2	7.1
4	3200	9	17.0	1	3.6
5	3400	0	0.0	1	3.6
6	3700	14	26.4	13	46.5
7	4000	6	11.3	4	14.2
8	4300	15	28.3	6	21.4
Mean depth		3,685 ft		3,775 ft	
Coded mean		5.75		6.11	

No. 3 - Vein dip and direction

Vein dip was recorded in eight classes as detailed below. This data must be considered modulo eight, as class one and class eight are actually adjacent.

<u>Code</u>	<u>Dip Angle &amp; Direction</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	0°-45°, S,SE,SW	1	1.9	0	0.0
2	45°-50°	4	7.5	0	0.0
3	60°-75°	22	41.5	16	57.2
4	75°-90°	22	41.5	9	32.1
5	75°-90°, N,NE,NW	3	5.7	3	10.7
6	60°-75°	1	1.9	0	0.0
7	45°-60°	0	0.0	0	0.0
8	0°-45°	0	0.0	0	0.0
Approx. mean		81°N,NE,NW		82°N,NE,NW	
Coded mean		--		--	

#### No. 4 - Vein Strike Azimuth

Vein strike was recorded in six classes, and must be considered modulo six, since class one is adjacent to class six. Strike azimuths in the second and third quadrants are identical to those in the fourth and first, respectively.

<u>Code</u>	<u>Strike Azimuth</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	0° - 30°	5	9.4	2	7.2
2	30° - 60°	11	20.8	5	17.8
3	60° - 90°	11	20.7	6	21.4
4	270° - 300°	24	45.3	12	42.8
5	300° - 330°	1	1.9	2	7.2
6	330° - 360°	1	1.9	1	3.6
Approx. mean		80°		82°	
Coded mean		--		--	

#### No. 5 - Silver Content of Intersection

The silver content of the intersection is calculated by multiplying the silver assay of the intersection by its width. The resultant number is in oz ft of silver/ton units.

Code	Silver Content (oz ft/ton)	Ore		Waste	
		Freq.	%	Freq.	%
1	0 - 60	8	15.1	27	96.4
2	60 - 100	10	18.9	1	3.6
3	100 - 150	14	26.4	0	0.0
4	150 - 250	12	22.7	0	0.0
5	250 - 375	4	7.5	0	0.0
6	> 375	5	9.4	0	0.0
Approx. mean		167		60.4	
Coded mean		3.17		1.04	

### No. 6 - Silver/Copper Ratio

The ratio is in the form oz silver/ton to % copper. It should be noted that the apparent discrepancy between the approximate and coded means is due to unequal class size.

Code	Silver/Copper Ratio (oz/ton, %)	Ore		Waste	
		Freq.	%	Freq.	%
1	0 - 15	4	7.5	9	32.2
2	15 - 20	5	9.4	0	0.0
3	20 - 30	17	32.1	4	14.3
4	30 - 50	20	37.8	5	17.8
5	50 - 80	5	9.4	4	14.3
6	> 80	2	3.8	6	21.4
Approx. mean		28		35	
Coded mean		3.43		3.46	

### No. 7 - Percentage Quartz in Vein

Code	Quartz in Vein (%)	Ore		Waste	
		Freq.	%	Freq.	%
1	0 - 10	19	35.9	4	14.3
2	10 - 25	26	49.0	13	46.5
3	25 - 50	5	9.5	5	17.8
4	> 50	3	6.6	6	21.4
Approx. mean		18%		29%	
Coded mean		1.85		2.46	

No. 8 - Distance of Vein Intersection from Polaris Fault

<u>Code</u>	<u>Distance</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	0-400	19	35.9	9	32.1
2	400-800	18	34.0	9	32.1
3	>800	16	30.1	10	35.8
Approx. mean		580 ft		610 ft	
Coded mean		1.94		2.04	

No. 9 - Wall Rock Type

<u>Code</u>	<u>Wall Rock Type</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Mainly argillite	2	3.8	1	3.6
2	Argillite-quartzite	5	9.4	4	14.3
3	Mainly quartzite	46	86.8	23	82.1
Coded mean		2.83		2.79	

No. 10 - Bedding Thickness of Wall Rock

<u>Code</u>	<u>Bedding Thickness</u> (in.)	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	>6	42	79.3	18	64.3
2	2-6	11	20.7	10	35.7
3	0-2	0	0.0	0	0.0
Coded mean		1.21		1.36	

No. 11 - Bedding/Vein Angle, Vertical Section

Characteristics 11 and 12 are measures of the closeness of the strike and dip of a vein to those of the bedding planes of the wall rock. As they are in fact the vertical and horizontal components of a single characteristic, they have also been combined as characteristic 12a.

<u>Code</u>	<u>Angle</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	0°-30°	14	26.4	6	21.4
2	30°-60°	27	51.0	16	57.2
3	60°-90°	12	22.6	6	21.4
Coded mean		1.96		2.00	

No. 12 - Bedding/Vein Angle, Horizontal Section

<u>Code</u>	<u>Angle</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	0°-30°	13	24.6	12	42.8
2	30°-60°	23	43.4	8	28.6
3	60°-90°	17	32.0	8	28.6
Coded mean		2.08		1.86	

No. 12a - Composite Bedding/Vein Angle

<u>Code</u>	<u>Code Product</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	1x1	1	1.9	1	3.6
2	1x2 or 2x1	11	20.8	7	24.1
3	1x3 or 3x1	14	26.4	9	32.1
4	2x2	12	22.6	6	21.4
5	2x3 or 3x2	15	28.3	5	17.8
6	3x3	0		0	
Coded mean		3.53		3.25	

No. 13 - Stringered Walls

This refers to the presence of small veinlets of quartz or other minerals in the wall rock on either side of the vein.

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	17	32.1	10	35.7
2	Present	36	67.9	18	64.3
Coded mean		1.68		1.64	

No. 14 - One or Both Walls with a Sharp, Frozen Contact

This refers to the nature of the interface between the vein material and the wall rock.

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	31	58.5	15	53.6
2	Present	22	41.5	13	46.4
Coded mean		1.36		1.46	

No. 15 - One or Both Walls Bounded by a Slip

This refers to the interface between the vein and the wall rock showing some evidence of faulting.

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	34	64.2	20	71.5
2	Present	19	35.8	8	28.5
Coded mean		1.36		1.29	

No. 16 - Vein or Wall Rock Brecciated

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	43	81.2	26	92.8
2	Present	10	18.8	2	7.2
Coded mean		1.19		1.07	

No. 17 - Tetrahedrite and/or Chalcopyrite in Wall Rocks

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	48	90.5	27	96.4
2	Present	5	9.5	1	3.6
Coded mean		1.09		1.04	

No. 18 - Percentage Wall Rock in Vein

<u>Code</u>	<u>Wall Rock in Vein</u> (%)	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	<25	49	92.5	26	92.8
2	>25	4	7.5	2	7.2
Coded mean		1.07		1.07	

No. 19 - Pyrite in Vein

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	19	35.9	9	32.1
2	Present	34	64.1	19	67.9
Coded mean		1.64		1.68	

No. 20 - Chalcopyrite in Vein

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	32	60.4	20	71.4
2	Present	21	39.6	8	28.6
Coded mean		1.40		1.29	

No. 21 - Galena in Vein

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	40	75.5	17	60.7
2	Present	13	24.5	11	39.3
Coded mean		1.24		1.39	

No. 22 - Arsenopyrite in Vein

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	48	90.6	26	92.8
2	Present	5	9.4	2	7.2
Coded mean		1.09		1.07	

No. 23 - Other Metallic Minerals in Vein

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	Absent	49	92.5	27	96.4
2	Present	4	7.5	1	3.6
Coded mean		1.08		1.04	

No. 24 - North or South of the Polaris Fault

<u>Code</u>	<u>Description</u>	<u>Ore</u>		<u>Waste</u>	
		<u>Freq.</u>	<u>%</u>	<u>Freq.</u>	<u>%</u>
1	North	14	26.4	6	21.4
2	South	39	73.6	22	78.6
Coded mean		1.74		1.79	

Second Data Set

A total of 659 and 279 data points were taken from ore bearing and waste veins respectively. These samples were large enough that it was reasonable to assume that the distributions obtained would be close approximations of the true distributions. The vein width, and the silver and copper grades were measured at each data point. The silver and copper contents of these intersections were calculated by:

$$\text{Silver content} = \text{Vein width} \times \text{oz silver/ton}$$

$$\text{Copper content} = \text{Vein width} \times \% \text{ copper}$$

The histograms for these quantities were positively skewed, but by logarithmically transforming each observation, an approximately normal form was obtained. These are shown in figures 3 and 4 for silver and copper respectively.

Although the transformed silver histograms are close to a normal shape, the copper histograms show a definite bimodal

ORIGINAL HISTOGRAMS-SILVER CHARACTERISTIC

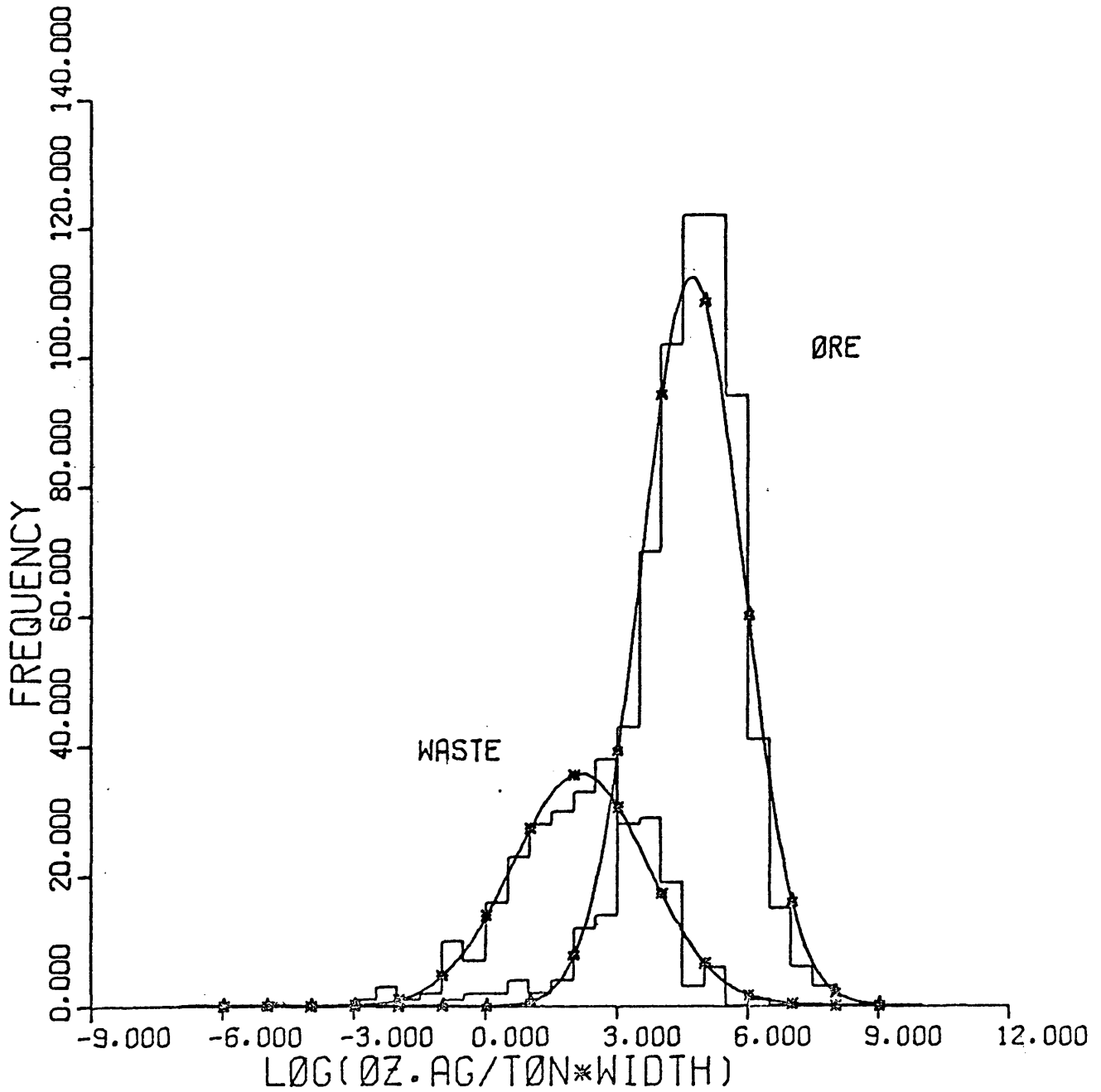


FIGURE 3

# ORIGINAL HISTOGRAMS-COPPER CHARACTERISTIC

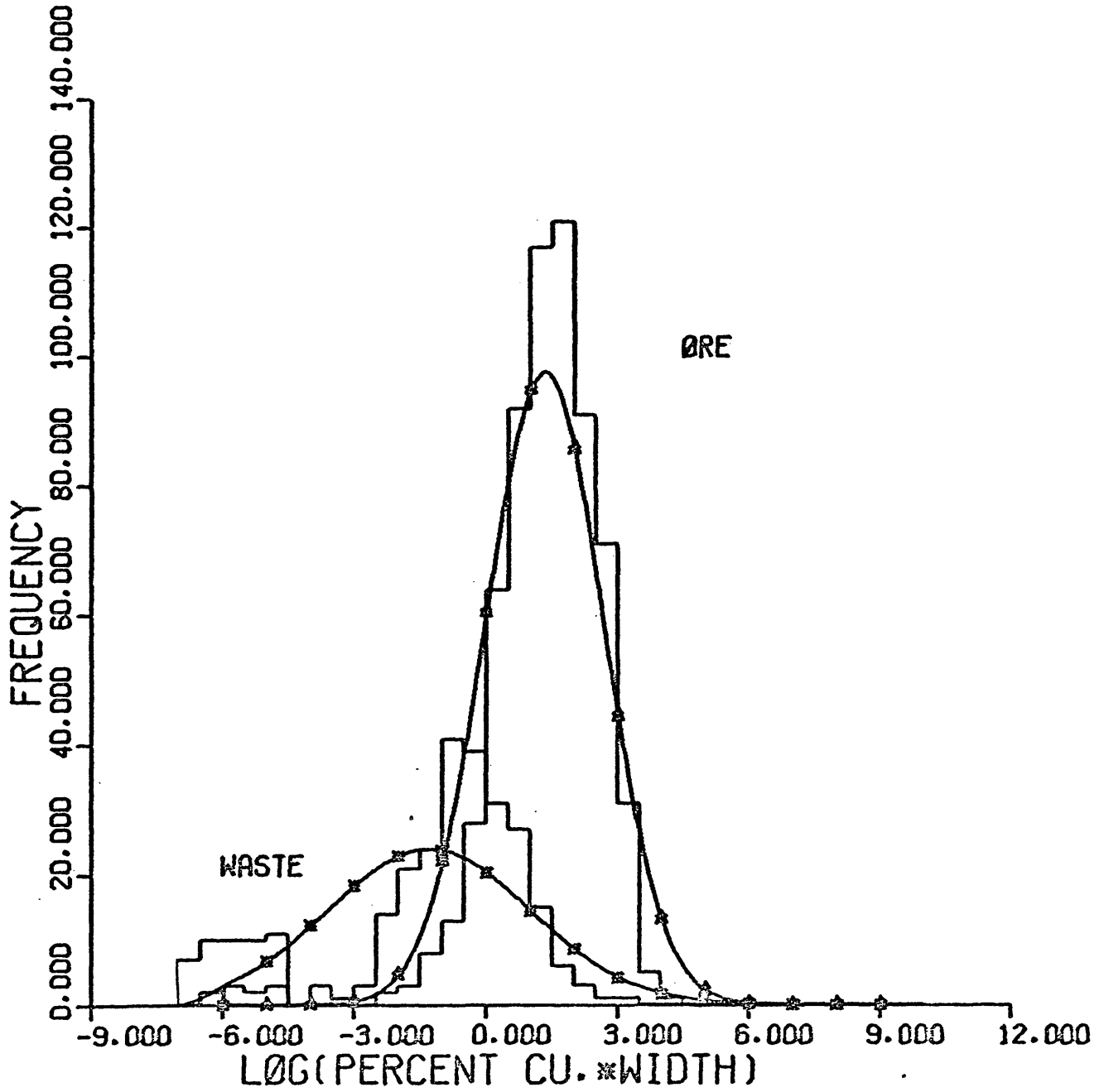


FIGURE 4

form. This was probably due to certain parts of the veins having been mineralized by fluids that had been stripped of their copper content, but still contained some silver.

These low copper data points were discarded as there was an insufficient number of them (11 and 47 for ore and waste respectively) to obtain a good estimate of their distribution. The amended histograms formed from the remaining 648 ore and 231 waste data points are shown in figures 5 and 6. It can be seen that a much better fit is obtained.

Table 1 shows the transformed amended means for the 28 ore bearing veins from which data was gathered.

#### Selection of Significant Characteristics

The stage 1 analysis (see p. 14-19) run on the 53 ore data points, and the 28 waste data points for the 24 characteristics gave the following results.

$$F = 3.3647$$

$$\text{and } F_{\alpha, k, (N_1 + N_2 - k - 1)} = 2.80 \quad (\text{from tables})$$

where

$$\alpha = 0.001$$

$$k = 24$$

$$(N_1 + N_2 - k - 1) = 56$$

Thus, the best level of significance at which the non-null hypothesis (that the ore and waste populations were different with respect to the 24 characteristics) could be rejected was less than 0.001. It is, therefore, extremely

AMENDED HISTOGRAMS-SILVER CHARACTERISTIC

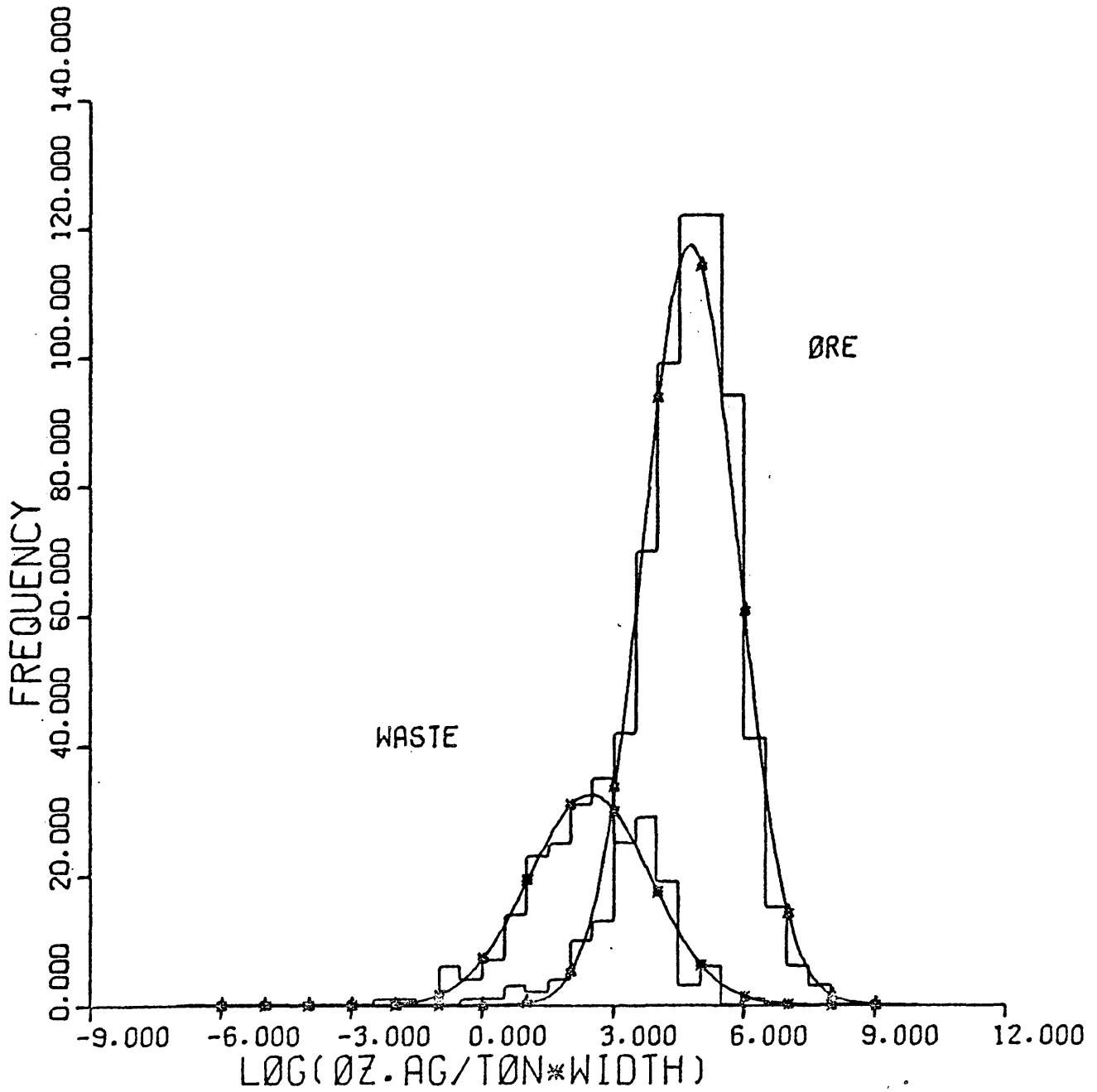


FIGURE 5

# AMENDED HISTOGRAMS-COPPER CHARACTERISTIC

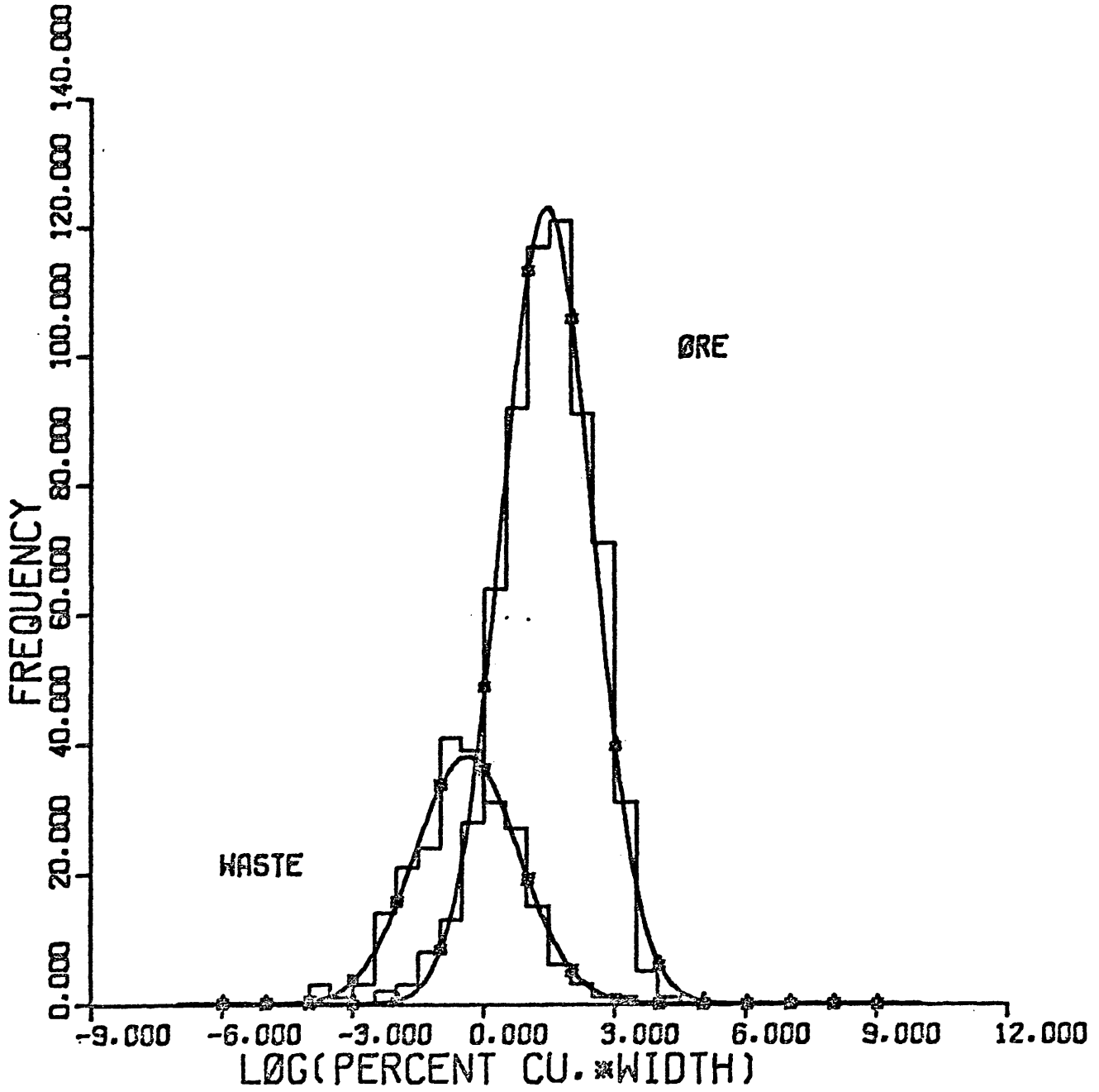


FIGURE 6

Table 1. Transformed data from the ore sample of the second data set.

<u>Vein Code</u>	<u>Vein No.</u>	<u>Sample Size</u>	<u>Mean (Log Ag.)</u>	<u>Mean (Log Cu)</u>
1	Silver	202	5.04	1.78
2	North	36	4.45	0.88
3	7	24	4.63	0.95
4	13	6	3.66	1.58
5	38	13	5.16	1.28
6	67	9	4.58	0.83
7	79	14	4.67	1.31
8	85	10	4.43	0.82
9	74-91	32	3.84	0.98
10	92	24	5.25	1.61
11	97	9	4.55	1.41
12	100	13	4.94	0.86
13	102	6	5.78	1.49
14	104	42	5.27	2.08
15	119	6	5.01	0.91
16	120	14	4.26	0.98
17	122	3	5.47	1.43
18	131	15	4.13	0.67
20	134	46	4.50	1.29
23	174-177	4	4.90	1.05
24	178	6	3.78	0.27
26	284	3	4.37	1.50
27	285	19	4.47	1.19
28	288	12	4.59	1.60
29	290	48	4.93	1.94
30	306	25	4.39	1.40
31	94	3	3.51	-0.22
32	257	5	3.03	-1.08
Total		648	4.75	1.43

probable that a real difference exists between these populations.

Confidence intervals were formed about the characteristic mean differences between population at levels of  $\alpha$  varying from 0.05 to 0.001. It was found that at the largest value of  $\alpha$ , eight characteristics caused the difference, and at the lowest value of  $\alpha$ , there were still seven of these that were significantly different. These characteristics were:

<u>Characteristic No.</u>	<u>Description</u>	<u>Significant at <math>\alpha</math> level</u>
5	Silver content of vein	0.001
7	Percent quartz in vein	0.001
10	Bedding thickness of wall rock	0.001
16	Vein or wall rock brecciated	0.001
17	Tetrahedrite and/or chalcopyrite in wall rocks	0.001
20	Chalcopyrite in vein	0.005
21	Galena in vein	0.001
23	Other metallic minerals in vein	0.001

This test was not run on the second set of data, as inspection of the histograms formed from the silver and copper characteristics for the ore and waste populations showed that they were obviously different.

Determination of the Homogeneity  
of the Characteristic Data

The multivariate ANOVA test was not run on the first data set, as it contained insufficient observations. A one way bivariate ANOVA test (see p. 19-30) was run on the

second set of data taken from the ore bearing veins. The two variates were log (silver content) and log (copper content), and the 28 treatments were the physically distinct veins from which data points had been gathered. The number of points used from each vein is shown in table 1.

The results obtained were as follows:

$$H = \begin{pmatrix} 148.2 & 112.2 \\ 112.2 & 156.6 \end{pmatrix} \quad E = \begin{pmatrix} 635.5 & 523.1 \\ 523.1 & 556.5 \end{pmatrix}$$

$$(H + E) = \begin{pmatrix} 783.7 & 635.3 \\ 635.3 & 713.1 \end{pmatrix}$$

$$(H + E)^{-1} = 10^{-2} \cdot \begin{pmatrix} 0.459 & -0.409 \\ 0.409 & 0.505 \end{pmatrix}$$

$$H(H+E)^{-1} = \begin{pmatrix} 0.221 & -0.040 \\ -0.125 & 0.331 \end{pmatrix}$$

The largest eigenvalue ( $\theta$ ) of this matrix was,

$$\theta = 0.366.$$

The parameters of its distribution were

$$S = 2.0$$

$$M = 12.0$$

$$n = 308.5.$$

From the corresponding Heck chart,

$$Z_{\alpha, S, M, n} = 0.09$$

for  $\alpha = 0.01$ .

Since  $\theta > Z_{\alpha, S, M, n}$ ,

the hypothesis that all the orebodies were the same was rejected at the 0.01 significance level. Simultaneous confidence intervals were formed about the characteristic mean

differences for all orebody pairs. It was found that none of these were significantly different at the 0.01 level. This apparent contradiction was caused by the difference in the sensitivity of the two tests. The largest eigenvalue is extremely sensitive to the differences in the treatment parameters, whereas the simultaneous confidence intervals cover the worst possible combination of events, and are consequently insensitive. In this particular instance, the wide disparity in treatment sample size (from 3 to 207) increased this insensitivity to a point where real differences between treatment parameters did not show.

An attempt was made to overcome this difficulty on a semi-heuristic basis by reducing the value of  $Z_{\alpha}$  used in the formation of the confidence intervals, until some orebody pairs did show a significant difference. The orebodies that caused this difference were eliminated, and the ANOVA test was rerun. The new value of  $\theta$  was then checked to see if the null hypothesis was still being rejected.

It was found that the disparity in treatment sample size was causing the elimination of orebodies having large sample sizes, even though the differences between their means, and those of other orebodies were comparatively small. It would appear that to obtain meaningful results from this method, sample sizes must be of the same order so that the resultant confidence interval sizes would only reflect the

differences in treatment parameters.

In order to make some use of the results, all orebodies which had a low grade, and a small sample size were eliminated. This was done under the assumption that such orebodies were in fact waste. These orebodies were on vein nos. 13, 94, 178, 257, and 284. In addition, vein nos. 74-91 were eliminated, as the confidence interval test showed these veins to be different from the other veins. Their proximity to a zone of the mine containing veins with a high lead content may account for this difference. A total of 65 data points were eliminated. The ANOVA test was rerun without these 65 data points, but the  $\theta$  test statistic still showed that the null hypothesis was rejected at the 0.01 level.

The remaining 594 data points were used in all subsequent operations with the second data set from the ore population.

#### Determination of Dependency Between Significant Characteristics

Contingency tables were constructed for all  $\binom{7}{2}$  pairs of the seven characteristics that were found to be different at the 0.001 level of  $\alpha$  in the multivariate comparison of means test between the ore and waste populations. The results from these tests are listed in table 2.

Table 2. Dependencies Between Significant Characteristics.

<u>Characteristic Combinations</u>	<u>X<sup>2</sup></u>	<u>Degrees of Freedom</u>	<u>* Best <math>\alpha</math></u>
5- 7	7.91	15	>0.05
5-10	3.26	10	>0.05
5-16	2.77	5	>0.05
5-17	9.65	5	>0.05
5-21	15.44	5	0.01
5-23	6.73	5	>0.05
7-10	0.87	6	>0.05
7-16	3.54	3	>0.05
7-17	10.21	3	0.025
7-21	1.09	3	>0.05
7-23	3.01	3	>0.05
10-16	2.78	2	>0.05
10-17	1.46	2	>0.05
10-21	1.05	2	>0.05
10-23	0.05	2	>0.05
16-17	1.28	1	>0.05
16-21	0.20	1	>0.05
16-23	1.00	1	>0.05
17-21	0.71	1	>0.05
17-23	0.45	1	>0.05
21-23	1.52	1	>0.05

\*Refers to the best level of  $\alpha$  at which the hypothesis of independence could be rejected.

Only two pairs had a best level of significance less than 0.05, and even these pairs were not highly dependent. It was thought advisable to combine the characteristics of each dependent pair into a new independent random variable prior to computing the decision rule probabilities. The method of doing this is shown in the section on the analysis of the joint probability distributions (p. 87-90).

### Comparison of Unknown Data Points

#### With Known Populations

Vein intersections of unknown nature were selected from 17 diamond drill holes. This data is shown in table 3.

### Discriminant Analysis

#### First Data Set

Discriminant analysis was applied to the data from the seven significant characteristics using all 53 ore, and 28 waste data points using the technique shown on p. 34-40. The following results were obtained.

The  $\underline{W}^{-1}\underline{B}$  matrix was found to be singular with rank one as expected. It consequently had only one eigenvalue, which had a magnitude of 1.00152. The eigenvector associated with this eigenvalue was used to transform the original data into the discriminant form. The linear transformation equation was,

Table 3. Data Points of Unknown Origin.

Int. No.	D.D.H. No.	Dist. from collar (ft)	Data Set 1							Data Set 2	
			Characteristic Code No.							L.A.C.	L.C.C.
			5	7	10	16	17	21	23		
1	28-24	45	3	1	1	1	1	1	1	4.77	1.23
2	30-18	180	2	4	2	1	1	2	1	4.44	0.64
3	32-18	222	1	1	1	1	1	2	1	3.73	-0.78
4	34-02	65	3	2	2	1	2	1	1	4.71	0.84
5	34-06	90	1	3	1	1	2	1	1	3.61	0.33
6	37-13	160	1	2	1	1	2	1	1	3.96	-0.08
7	37-23	340	1	1	1	1	2	1	1	3.97	0.62
8	37-43	70	3	4	1	1	2	1	1	4.69	1.68
9	37-43	300	1	2	1	1	1	2	1	3.17	-0.29
10	37-45	570	6	1	2	1	1	1	1	6.11	2.62
11	40-01	215	1	1	1	1	2	1	1	3.84	-0.73
12	40-02	115	1	3	1	1	2	1	1	3.86	-0.08
13	40-32	420	2	1	2	1	2	1	1	4.24	0.13
14	40-38	320	4	2	1	1	2	2	1	5.46	1.97
15	43-04	410	5	1	1	2	2	1	1	5.95	1.57
16	43-16	380	1	3	1	1	1	1	1	4.08	-0.18
17	43-19	130	2	1	1	1	2	1	1	4.22	-0.18
18	43-19	155	4	1	1	2	2	2	1	5.29	0.94
19	43-29	160	3	3	1	1	1	1	1	4.92	1.15

D.D.H. No.: Diamond drill hole number  
 L.A.C. : Natural logarithm (vein width x mean silver assay)  
 L.C.C. : Natural logarithm (vein width x mean copper assay)

$$y_j = 0.442 x_{5j} - 0.263 x_{7j} - 0.338 x_{10j} + 0.517 x_{16j} \\ + 0.464 x_{17j} + 0.055 x_{21j} + 0.368 x_{23j}$$

where:

$y_j$  = discriminant value of the  $j^{\text{th}}$  data point,  
 $x_{5j}, \dots, x_{23j}$  = observed value of the 5<sup>th</sup>, ..., 23<sup>rd</sup>  
characteristics for the  $j^{\text{th}}$  data point.

It must be re-emphasized at this point that this example is shown for illustrative purposes only, as the results obtained are probably inaccurate due to the original data being in an unsuitable form.

The coefficients of this transformation equation indicate the relative importance of the characteristics in the discrimination between ore and waste veins. These coefficients appear to be of the correct magnitude and sign, with the possible exception of the coefficient of  $x_{21j}$ . There appears to be a definite correlation between the presence of lead and a vein being waste. Consequently, it would be expected that  $x_{21j}$  would have a negative coefficient. It is possible that the axis of maximum separation between the data point cluster centers is almost orthogonal to the lead characteristic axis. This would mean that this characteristic is of little use in discriminating between ore and waste when used in conjunction with the other six characteristics.

The single discriminant transformation was performed on the 53 ore, 28 waste, and 19 unknown data points. The discriminant means and variances for the first two categories were:

	<u>Mean</u>	<u>Variance</u>
Ore	2.0910	0.5009
Waste	0.8413	0.0937

The discriminant values for the 19 unknown data points, and their generalized one dimensional distances (obtained from equation 26) from the two population centers are given in table 4. The reason for the much larger distances associated with the waste population is that the variance of this population is very much smaller than that of the ore population. This in turn is mainly due to the uniformly low grade characteristic values of the waste sample data points.

The misclassification probabilities were computed using equations 30 and 31. These are shown together with the resultant decision rule probabilities in table 5.

#### Second Data Set

Discriminant analysis was applied to the data points remaining in the second set after the low copper points had been removed from the ore and waste samples, and all points from veins shown to be different in the ANOVA test had been eliminated from the ore sample. Points from these veins were

Table 4. Discriminant Values and Distances of Unknown Data Points from Population Centers.

First data set

<u>Intersection No.</u>	<u>y</u>	<u>z<sup>2</sup>(ore)</u>	<u>z<sup>2</sup>(waste)</u>
1	2.1274	0.003	17.66
2	0.6121	4.367	0.56
3	1.2991	1.252	2.24
4	1.9891	0.021	14.07
5	1.1808	1.654	1.23
6	1.4442	0.836	3.88
7	1.7075	0.294	8.01
8	1.8010	0.168	9.83
9	1.0357	2.224	0.40
10	3.1141	2.089	55.15
11	1.7075	0.294	8.01
12	1.1808	1.654	1.23
13	1.8106	0.157	10.03
14	2.8248	1.075	42.01
15	3.9914	7.210	105.96
16	0.2935	6.061	2.59
17	2.1493	0.007	18.27
18	3.6049	4.576	81.55
19	1.6006	0.480	6.16

$z^2$  = generalized one dimensional distance between data point and population center.

y = discriminant value of the unknown data point.

Table 5. Misclassification and Decision Rule Probabilities for Unknown Data Points.

Discriminant analysis - First data set

Intersection No.	$P[Y \leq y   A]$	$P[Y > y   B]$	$P[A   J]$	$P[B   J]$
1	0.520	0.00001	0.9999	0.0001
2	0.018	0.77308	0.1450	0.8550
3	0.132	0.06735	0.6320	0.3680
4	0.443	0.00009	0.9995	0.0005
5	0.099	0.13364	0.4429	0.5571
6	0.180	0.02442	0.8436	0.1564
7	0.294	0.00232	0.9858	0.0142
8	0.341	0.00086	0.9949	0.0051
9	0.068	0.26264	0.2774	0.7226
10	0.926	0.00000	1.0000	0.0000
11	0.294	0.00232	0.9858	0.0142
12	0.099	0.13364	0.4429	0.5571
13	0.346	0.00077	0.9955	0.0045
14	0.850	0.00000	1.0000	0.0000
15	0.996	0.00000	1.0000	0.0000
16	0.007	0.94624	0.1893	0.8107
17	0.532	0.00001	0.9999	0.0001
18	0.984	0.00000	1.0000	0.0000
19	0.244	0.00655	0.9583	0.0417

where:

A = event that vein is ore

B = event that vein is waste

y = discriminant value of unknown data point

J = event of the division of the set of all possible discriminant values into two subsets by value y.

not removed from the waste sample, as it appeared likely that the differences (with the possible exception of veins 74-91) arose because these veins were really waste. If this assumption was true, points taken from these veins were correctly classified in the waste sample, and consequently should not be removed.

As there were only two characteristics in the second data set, the discriminant transformation merely put the data in a suitable form from which the misclassification probabilities could be computed. It is possible to compute these probabilities directly using Hermite-Gauss integration (see p. 48-51) if the original dimension of the data is small. This latter method is computationally more efficient for bivariate data, but it does not appear as precise.

The misclassification probability masses computed by these two methods are different, since the limits for the integrals are set in a different manner. Figure 7 shows the reason for this difference. The discriminant limit (line DD) appears intuitively to be a better way of dividing the set of all possible points into two subsets.

The analysis was run with 594 points from 22 ore bearing veins, and 232 points from waste veins. The following results were obtained using the technique shown on p. 34-41.

# DISCRIMINANT AND PROBABILISTIC LIMITS

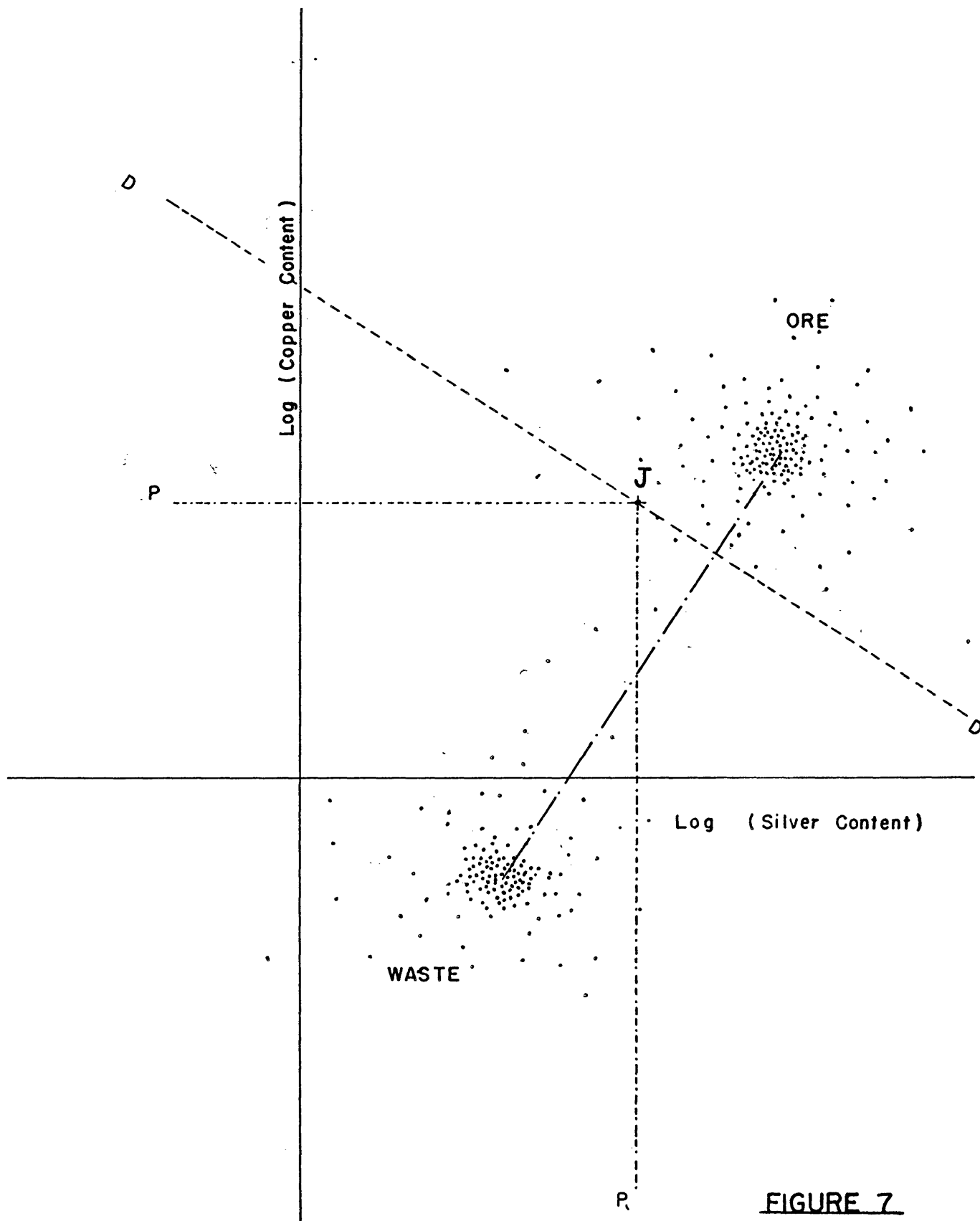


FIGURE 7

The  $W^{-1}B$  matrix was found to be singular with rank 1. Its eigenvalue was of magnitude 0.852, and it had an associated eigenvector (0.969, 0.247). This eigenvector was used to transform the second data set observations into new discriminant values.

$$y_j = 0.969 x_{1j} + 0.247 x_{2j}$$

where  $x_{1j} = \log$  (silver content) of the  $j^{\text{th}}$  data point

$x_{2j} = \log$  (copper content) of the  $j^{\text{th}}$  data point.

This discriminant transformation was performed on 594 ore, 232 waste, and 19 unknown data points. The discriminant means and variances for the first two categories are shown below.

	<u>Mean</u>	<u>Variance</u>
Ore	5.061	1.5721
Waste	2.279	2.5057

The misclassification probabilities for the 19 unknown data points were computed using equations 30 and 31. These, and the resultant decision rule probabilities are shown in table 6.

### Analysis of Joint Probability Distributions

#### First Data Set

The two pairs of dependent significant characteristics were combined in the following manner:

Table 6. Misclassification and Decision Rule Probabilities for Unknown Data Points.

Discriminant analysis - Second data set

Intersection No.	y	$P[Y \leq y   A]$	$P[Y > y   B]$	$P[A   J]$	$P[B   J]$
1	4.926	0.457	0.047	0.802	0.198
2	4.460	0.316	0.084	0.701	0.299
3	3.421	0.095	0.235	0.350	0.650
4	4.682	0.381	0.064	0.754	0.246
5	3.579	0.119	0.206	0.407	0.593
6	3.817	0.161	0.166	0.495	0.505
7	4.000	0.199	0.138	0.560	0.440
8	5.221	0.551	0.032	0.847	0.153
9	3.000	0.050	0.324	0.213	0.786
10	6.568	0.885	0.003	0.942	0.057
11	3.540	0.113	0.213	0.393	0.607
12	3.720	0.142	0.181	0.459	0.541
13	4.140	0.231	0.120	0.606	0.397
14	5.778	0.716	0.014	0.902	0.098
15	6.153	0.808	0.007	0.925	0.075
16	3.909	0.179	0.152	0.528	0.472
17	4.044	0.209	0.132	0.575	0.425
18	5.358	0.594	0.026	0.863	0.137
19	5.052	0.497	0.040	0.823	0.177

where: A = event that vein is ore

B = event that vein is waste

y = discriminant values of unknown data point

J = event that the set of all possible vectors Y is divided into two subsets by the discriminant vector y.

Pair (5-21): There appeared to be a definite negative correlation between increasing vein silver content and the presence of galena. A heuristic relationship was formed that reflected the relative importance of the characteristics, and ordered the coded values such that:

$$P_A \left[ \frac{X}{x_j} \right] > P_B \left[ \frac{X}{x_j} \right] \quad (42)$$

This relationship was,

$$(521) = (5) - 0.5 \times (21) \quad (43)$$

where:

(5) = coded value of characteristic 5

(21) = coded value of characteristic 21

(521) = equivalent coded value of composite characteristic 521.

This new characteristic had 12 possible values ranging from 0 to 5 1/2.

Pair (7-17): There was a weak positive correlation between the increasing percentage of quartz in a vein and the presence of chalcopyrite and/or tetrahedrite in the adjacent wall rock. The linking relationship used was:

$$(717) = - \left[ (7) + (17) \right] \quad (44)$$

where (7) = coded value of characteristic 7

(17) = coded value of characteristic 17

(717) = equivalent coded value of composite characteristic 717.

The negative sign was added so that the condition of equation 42 would be met. This new characteristic has five possible values ranging from -6 to -2.

The distributions of these five independent characteristics (nos. 521, 717, 10, 16 and 23) have to be in a form such that equations 32 and 33 of the preceding section can be applied. The cumulative frequency distributions of the composite characteristics were obtained from the ore and waste samples. As these were discrete step functions it was necessary to use these probabilities in the form,

$$P_A \left[ X \leq x_j \right]$$

and

$$P_B \left[ X \geq x_j \right].$$

The remaining three characteristics are in effect Bernoulli trials, and consequently it is convenient to replace the probabilities in equations 32 and 33 corresponding to these characteristics with,

$$P_A \left[ X = x_j \right]$$

and

$$P_B \left[ X = x_j \right].$$

The frequency distributions of the five characteristics obtained from the 53 ore, and 28 waste data points are given in table 7. It should be noted that for characteristic 521, small probabilities have been assigned to  $P_B \left[ X \geq x_j \right]$  for values of  $x_j \geq 2$ . This has been done because it is known

Table 7. Significant Characteristic Distributions.

Characteristic No.	x	Ore		Waste	
		Freq.	$P_A[X \leq x]$	Freq.	$P_B[X > x]$
521	0	5	0.094	10	1.0000
521	$\frac{1}{2}$	3	0.151	17	0.6430
521	1	1	0.170	0	0.0350
521	$1\frac{1}{2}$	9	0.339	1	0.0350
521	2	6	0.452	0	0.0040
521	$2\frac{1}{2}$	8	0.604	0	0.0035
521	3	0	0.604	0	0.0030
521	$3\frac{1}{2}$	12	0.830	0	0.0025
521	4	1	0.848	0	0.0020
521	$4\frac{1}{2}$	3	0.906	0	0.0015
521	5	0	0.906	0	0.0010
521	$5\frac{1}{2}$	5	1.000	0	0.0005
717	-6	1	0.019	1	1.000
717	-5	4	0.094	5	0.964
717	-4	3	0.151	5	0.785
717	-3	29	0.698	12	0.607
717	-2	16	1.000	5	0.178
		Freq.	$P_A[X=x]$	Freq.	$P_B[X=x]$
10	1	42	0.794	18	0.644
10	2	11	0.206	10	0.356
16	1	43	0.812	26	0.928
16	2	10	0.188	2	0.072
23	1	49	0.925	27	0.964
23	2	4	0.075	1	0.036

that the silver content of a waste vein intersection can exceed the quantity corresponding to  $x_j = 1\frac{1}{2}$ . This was the highest value of this characteristic recorded in the 28 data points taken from waste veins. To omit doing this would have implied that any intersection with a value of characteristic 521 greater than or equal to two, would be ore with probability one.

The misclassification and decision rule probabilities for the 19 unknown data points are shown in table 8. The method of calculating these is shown below using the intersection at 320 ft in D.D.H. 40-38.

The  $\underline{x}_j$  vector for this intersection was

$$(3, -4, 1, 1, 1)$$

$$P_A \left[ \underline{X} < \underline{x}_j \right] = P_A \left[ X_{521} \leq 3 \right] \cdot P \left[ X_{717} \leq -4 \right] \cdot P \left[ X_{10} = 1 \right] \cdot P \left[ X_{16} = 1 \right] \cdot P \left[ X_{23} = 1 \right]$$

$$\begin{aligned} P_A \left[ \underline{X} < \underline{x}_j \right] &= 0.604 \times 0.151 \times 0.794 \times 0.812 \times 0.925 \\ &= 0.054. \end{aligned}$$

Similarly,

$$P_B \left[ \underline{X} \geq \underline{x}_j \right] = 0.00136.$$

Then, assuming the a priori probabilities are equal,

$$\begin{aligned} P \left[ A \mid \underline{X} \leq \underline{x}_j \right] &= \frac{0.0054}{0.054 + 1.0 - 0.00136} \\ &= 0.0513 \end{aligned}$$

$$\begin{aligned} P \left[ B \mid \underline{X} > \underline{x}_j \right] &= \frac{0.00136}{1.0 - 0.054 + 0.00136} \\ &= 0.001438. \end{aligned}$$

Table 8. Misclassification and Decision Rule Probabilities  
for Unknown Data Points.

Probabilistic analysis - First data set

Intersection No.	$\underline{x}$	$P[\underline{X} \leq \underline{x}   A]$	$P[\underline{X} > \underline{x}   B]$	$P[A   J]$	$P[B   J]$
1	$(2\frac{1}{2}, -2, 1, 1, 1)$	0.360	0.0004	0.9979	0.0021
2	$(1, -5, 2, 1, 1)$	0.002	0.0107	0.1896	0.8104
3	$(0, -2, 1, 1, 1)$	0.056	0.1025	0.3750	0.6250
4	$(2\frac{1}{2}, -4, 2, 1, 1)$	0.014	0.0009	0.9401	0.0599
5	$(\frac{1}{2}, -5, 1, 1, 1)$	0.008	0.3571	0.0468	0.9532
6	$(\frac{1}{2}, -4, 1, 1, 1)$	0.014	0.2908	0.0763	0.9237
7	$(\frac{1}{2}, -3, 1, 1, 1)$	0.063	0.2249	0.2793	0.7207
8	$(2\frac{1}{2}, -6, 1, 1, 1)$	0.007	0.0020	0.7707	0.2293
9	$(0, -3, 1, 1, 1)$	0.039	0.3497	0.1754	0.8246
10	$(5\frac{1}{2}, -2, 2, 1, 1)$	0.140	0.0001	0.9995	0.0005
11	$(\frac{1}{2}, -3, 1, 1, 1)$	0.063	0.2249	0.2793	0.7207
12	$(\frac{1}{2}, -5, 1, 1, 1)$	0.008	0.3571	0.0468	0.9532
13	$(1\frac{1}{2}, -3, 2, 1, 1)$	0.037	0.0068	0.8360	0.1640
14	$(3, -4, 1, 1, 1)$	0.054	0.0014	0.9730	0.0270
15	$(4\frac{1}{2}, -3, 1, 2, 1)$	0.087	0.0000	0.9994	0.0006
16	$(\frac{1}{2}, -4, 1, 1, 1)$	0.014	0.2908	0.0763	0.9237
17	$(1\frac{1}{2}, -3, 1, 1, 1)$	0.058	0.0001	0.9984	0.0016
18	$(3, -3, 1, 2, 1)$	0.141	0.0122	0.8990	0.1010
19	$(2\frac{1}{2}, -4, 1, 1, 1)$	0.054	0.0016	0.9687	0.0313

where: A = event that vein is ore

B = event that vein is waste

$\underline{x}$  = characteristic vector for unknown data point

J = event of the division of the set of all possible  
vectors  $\underline{X}$ , into two subsets by the vector  $\underline{x}$ .

Then using equations 40 and 41, the decision rule probabilities can be computed. This gives the decision rule probabilities as:

$$P \left[ \begin{array}{c|c} \text{A} & \text{J} \end{array} \right] = 0.973$$

$$P \left[ \begin{array}{c|c} \text{B} & \text{J} \end{array} \right] = 0.027.$$

In this case, if the exploration proposal was to follow the line of the diamond drill hole to the vein intersection, and then to explore the vein with 100 ft of drift, the expected gross return from the investment would be:

$$E \left[ \text{GR} \right] = \bar{V} \times 0.973 - (420 \times d) \times 0.027$$

where  $\bar{V}$  = expected gross value of orebody

$d$  = cost/ft of exploration drift.

#### Second Data Set

The second data set was in a suitable form for direct probabilistic analysis, since the distributions of both variates were closely normal in form. Although they were highly dependent, the bivariate normal distribution function for such variates is known, and probabilities can be computed directly from this function, as shown on p. 48-51. of the previous section.

Five values of  $n$ , ranging from 9 to 20, were used on each of the 19 unknown data points. The convergence obtained was not too satisfactory, but it was thought to be sufficiently good to make the results useable.

The misclassification and decision rule probabilities obtained using a value of  $n = 20$  in equation 48 are shown for the 19 unknown data points in table 9.

Table 9. Misclassification and Decision Rule Probabilities for Unknown Data Points.

Probabilistic analysis - Second data set

Intersection No.	$\underline{x}$	$P[\underline{X} \leq \underline{x}   A]$	$P[\underline{X} > \underline{x}   B]$	$P[A   J]$	$P[B   J]$
1	(4.77, 1.23)	0.328	0.037	0.831	0.169
2	(4.44, 0.64)	0.207	0.037	0.800	0.200
3	(3.73, -0.78)	0.011	0.143	0.093	0.907
4	(4.71, 0.84)	0.225	0.037	0.808	0.192
5	(3.61, 0.33)	0.090	0.186	0.370	0.630
6	(3.96, -0.08)	0.066	0.133	0.361	0.639
7	(3.97, 0.62)	0.169	0.109	0.578	0.422
8	(4.69, 1.68)	0.416	0.021	0.894	0.106
9	(3.17, 0.29)	0.024	0.250	0.134	0.866
10	(6.11, 2.62)	0.819	0.001	0.992	0.008
11	(3.84, -0.73)	0.011	0.143	0.093	0.907
12	(3.86, -0.08)	0.057	0.133	0.334	0.666
13	(4.24, 0.13)	0.066	0.109	0.396	0.604
14	(5.46, 1.97)	0.616	0.005	0.970	0.030
15	(5.95, 1.57)	0.499	0.002	0.985	0.015
16	(4.08, -0.18)	0.066	0.133	0.361	0.639
17	(4.22, -0.18)	0.066	0.133	0.361	0.639
18	(5.29, 0.94)	0.350	0.013	0.930	0.070
19	(4.92, 1.15)	0.328	0.037	0.831	0.169

where A = event that vein is ore

B = event that vein is waste

$\underline{x}$  = grade characteristic vector for unknown data point

J = event that the set of all possible vectors  $\underline{X}$  is divided into two subsets by the vector  $\underline{x}$ .

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The most unsatisfactory part of this method is the use of multivariate ANOVA (with Heck charts) to determine if the characteristic data belongs to more than one population. It may be more efficient in many situations to dispense with this test, and divide the study area into separate populations using only qualitative geological considerations.

Apart from this test, the method appears to be generally applicable if the characteristic information is available in a suitable form and quantity. The actual size of the samples required for reasonable estimates of the characteristic distributions will vary, but as a general rule at least 200 to 300 observations of each population should be obtained if much weight is to be placed on the results. The test sequence is reasonably easy to understand and apply. All data manipulation and computation can be done by computer, and memory requirements should not be excessive. The method as outlined can be applied directly to any mine with a decision problem similar to the Galena Mine.

Table 10 shows the rankings obtained from the decision rule probabilities computed by discriminant and probabilistic

Table 10. Comparison of the Rankings of Unknown Veins  
Obtained by Different Methods.

Rank No.	<u>Discriminant Analysis</u>		<u>Probabilistic Analysis</u>	
	<u>1st Data Set</u>	<u>2nd Data Set</u>	<u>1st Data Set</u>	<u>2nd Data Set</u>
1	15	10	10	10
2	18	15	15	15
3	10	14	17	14
4	14	18	1	18
5	17	8	14	8
6	1	19	19	1
7	4	1	4	19
8	13	4	18	4
9	8	2	13	2
10	7	13	8	7
11	11	17	3	13
12	19	7	7	5
13	6	16	11	6
14	3	6	2	16
15	5	12	9	17
16	12	5	6	12
17	9	11	16	9
18	16	3	5	3
19	2	9	12	11

- Note: 1. Rank numbers run from 1 (most likely to be ore) to 19 (most likely to be waste).
2. The numbers listed under the four headings are the vein intersection numbers as given in table 3.

analysis for the two data sets. Strictly, these rankings should come from the expected gross return of each exploration proposal computed as shown on page 55. The discriminant rankings obtained from the second data set are almost certainly more accurate than the corresponding probabilistic rankings, because the latter were obtained from numerical Hermite-Gauss integration which did not show good convergence.

The ranking differences between the first and second data sets reflect the presence of the non-grade characteristics in the first set. In using these rankings to select a target, either the second data set rankings should be used on their own, or a weighted sum of the first and second set rankings should be used to establish a new rank order. If the latter method is used, the second data set rankings should be given a heavier weight as results obtained from the first data set are suspect.

A possible relationship to get this new rank would be:

$$\begin{aligned} \text{(Weighted sum)} &= 0.20 \times (\text{1st data set rank}) \\ &+ 0.80 \times (\text{2nd data set rank}). \end{aligned}$$

These weighted sums would then be arranged in ascending order.

In any study in which the form of the original data satisfied all necessary conditions, only one ranking would result, and the above heuristic operation would not be needed. The prime requirement of a successful study of

this nature is that the method of data gathering and processing be set up in advance so that all necessary mathematical conditions are met. In addition, as the study progresses, the characteristic distributions should be continually updated as more information becomes available. In this way, the mathematical description of the ore and waste populations will become progressively more precise.

The final test of this method is whether or not it increases the success ratio of underground exploration consistently. If the probability of finding new veins does not decrease with increased underground exploration, this test can easily be made. Each exploration proposal carried through to completion is a Bernoulli trial, either ore is found in the vein, or it is not. Thus the number of successes in a given period will have a Binomial distribution.

$$P[X \geq x] = \sum_{r=x}^n \binom{n}{r} p^r (1-p)^{n-r}$$

where  $n$  = number of trials

$p$  = probability of success in a single trial

$r = 0, 1, \dots, n$ .

$x$  = number of successes.

An estimate of the  $p$  achieved in the past can be obtained simply by dividing the number of exploration successes by the number of completed exploration proposals.

If the success ratio changes with the introduction of the new method, a simple hypothesis test can be run to determine if this change is significant.

### Recommendations for Further Study

#### Determination of Optimal Exploration Strategy

Once a decision has been made to explore an unknown vein, there may be numerous alternative methods of obtaining further information about the target in order to determine its real nature. Each of these will have associated costs and returns. Given a reasonable amount of information about the veins of different natures, it should be possible to select the exploration strategy that minimized the expected loss (or maximizes the expected gain) of the proposal.

#### Prediction of Most Suitable Area for Vein Search

It should be possible to adapt the method used in this study for selecting areas within, or adjacent to the mine that have a greater probability of containing an ore bearing vein. This type of study would use regional characteristics instead of the localized characteristics of individual veins. Apart from this, the basis of the method would be the same. It may well be harder to define and use regional characteristics.

LITERATURE CITED

- Allais, M., 1957, Method of appraising economic prospects of mining exploration over large territories. Algerian Sahara case study: Management Sci., v. 3, no. 4, p. 285-347.
- Botbol, J. M., 1969, Characteristic analysis of base metal mining district mineralogy: AIME preprint no. 69-L-325, 17 p.
- Cooley, W. W., and Lohnes, P. R., 1962, Multivariate procedures for the behavioral sciences: New York, J. Wiley and Sons, Inc., 211 p.
- Griffiths, J. C., 1966, Exploration for natural resources: Operations Research, v. 14, no. 2, p. 189-209.
- Harris, D. P., 1965, Multivariate statistical analysis - a decision tool for mineral exploration: Short Course and Symposium on Computers and Computer Applications in Mining and Exploration Trans., Arizona Univ., Tucson, v. I, p. C1-C35.
- \_\_\_\_\_ 1968, Alaska's base and precious metals resources: A probabilistic regional appraisal: Colorado School of Mines Quart., v. 64, no. 3, p. 295-328.
- Hazen, S. W., 1967, Some statistical techniques for analyzing mine and mineral-deposit sample and assay data: U.S. Bur. Mines Bull. 621, 233 p.
- Howd, F. H., 1964, The taxonomy program - A computer technique for classifying geologic data: Colorado School of Mines Quart., v. 59, no. 4, p. 207-222.
- Johnson, N. L., and Leone, F. C., 1964a, Statistics and experimental design, v. I: New York, J. Wiley and Sons, Inc., 523 p.
- \_\_\_\_\_ and \_\_\_\_\_ 1964b, Statistics and experimental designs, v. II: New York, J. Wiley and Sons, Inc., 399 p.
- Koch, G. S. Jr., and Link, R. F., 1967, Geometry of metal distribution in five veins of the Fresnillo Mine, Zacatecas, Mexico: U.S. Bur. Mines Rept. Inv. 6919, 64 p.

- Link, R. F., and Koch, G. S. Jr., 1967, Linear discriminant analysis of multivariate assay and other mineral data: U.S. Bur. Mines Rept. Inv. 6898, 25 p.
- Marshall, K. T., 1964, A preliminary model for determining the optimum drilling pattern in locating and evaluating an orebody: Colorado School of Mines Quart., v. 59, no. 4, p. 223-236.
- Morrison, D. F., 1967, Multivariate statistical methods: New York, McGraw-Hill Book Co., 338 p.
- National Bureau of Standards, 1965, Handbook of mathematical functions: Washington, U.S. Government Printing Office, 1045 p.
- Rao, C. R., 1952, Advanced statistical methods in biometric research: New York, J. Wiley and Sons, Inc., 390 p.
- Wilks, S. S., 1962, Mathematical statistics: New York, J. Wiley and Sons, Inc., 644 p.

## DETERMINATION OF SIGNIFICANT CHARACTERISTICS

```

*****
DIMENSION EPPA(4)
DIMENSION COVO(24,24),COV(24,24),DELT(24),TD(24)
DIMENSION DA(65,48),PRD(24),COVI(24,24),TDA(24,2),COVS(24,24)
DIMENSION COCO(24,24),SIGMA(24),DIF(4,10),AB(4,10),X(10),Y(10)
  DIMENSION PRA(24,24),UP(24),SLP(24)
EQUIVALENCE (COVS(1,1),COV(1,1)), (COCO(1,1),COVQ(1,1))
  EQUIVALENCE (SIGMA(1),DELT(1))
EQUIVALENCE (DA(1),TD(1))
EQUIVALENCE (COCO(1,1),UP(1)), (COCO(1,2),SLP(1))
III=53
II=1
EN1=III
L=24
ENJ=L
KT=1
DO 3 I=II,III
3 READ (5,2) (DA(I,J),J=1,L)
2 FORMAT (24F1,0)
DO 910 J=1,III
DOT=DA(J,12)*DA(J,11)
DA(J,12)=DOT
910 CONTINUE
CALL MEAN (DA,TDA,KT,II,III,L)
CALL HISTO (DA,II,III,AB)
CALL COORD (AB,DA)
CALL COVAR (DA,TDA,II,III,KT,COVS,L)
III=28
DO 904 I=1,III
904 READ (5,905) (DA(I,J),J=1,L)
905 FORMAT (24F1,0)
KT=2
EN2=III
DO 911 J=1,III
DOT=DA(J,12)*DA(J,11)
DA(J,12)=DOT
911 CONTINUE
CALL MEAN (DA,TDA,KT,II,III,L)
CALL HISTO (DA,II,III,AB)
CALL COORD (AB,DA)
CALL COVAR (DA,TDA,II,III,KT,COVQ,L)
DO 906 J=1,L
DO 906 J=1,L
COVT=0.0
COVT=COVS(I,J)+COVQ(I,J)
COVS(I,J)=COVT/(EN1+EN2-2.0)
COVS(I,J)=COVT/(EN1+EN2)
906 CONTINUE
CALL CIVER (COVS,COVI,L)
DO 900 I=1,L
900 SIGMA(I)=SQRT(COVS(I,I))
DO 901 I=1,L
DO 901 J=1,L
COCO(I,J)=COVS(I,J)/(SIGMA(I)*SIGMA(J))

```

```

901 CONTINUE
  DO 902 I=1,L
    WRITE (6,903) (COCC(I,J),J=1,L)
903 FORMAT (2X,24F5.2)
902 CONTINUE
  DO700 I=1,L
    DEL=TDA(I,1)-TDA(I,2)
    DELT(I)=DEL
700 CONTINUE
  COEFF=(EN1*EN2)/(EN1+EN2)
  DO 701 I=1,L
    TOD=0.0
    DO701 J=1,L
      TOD=TOD+DELT(J)*COVI(I,J)
    TD(I)=TOD
701 CONTINUE
  DO 702 I=1,L
902 FADD=FADD+TD(I)*DELT(I)
  TSQ=FADD*COEFF
  FCOF=(EN1+EN2-ENJ-1.0)/((EN1+EN2-2.0)*ENJ)
  FST=FCOF*TSQ
  WRITE(6,703) FST,FCOF,TSQ
703 FORMAT (1X,3F20.8)
  READ (5,704) (EFFA(I),I=1,4)
704 FORMAT (4F3.2)
  DO 705 IK=1,4
    ATSQ=SQRT(EFFA(IK)/FCOF)
    DO 710 I=1,L
      UPC=DELT(I)+(COVS(I,I)/COEFF)*ATSQ
      UP(I)=UPC
      SLP=DELT(I)-(COVS(I,I)/COEFF)*ATSQ
      SLP(I)=SLP
710 CONTINUE
  DO 711 I=1,L
711 WRITE (6,712) SLP(I),DELT(I),UP(I),I
712 FORMAT (/,3F20.8,110)
705 CONTINUE
  STOP
  END
  SUBROUTINE MEAN(DA,TDA,KT,II,III,L)
  DIMENSION DA(60,24),TDA(24,2)
  WRITE (6,85)
85 FORMAT (//,12X,4HMEAN,/,12X,4(1H=),//)
  DIV=III
  DO 20 J=1,L
    DT=0.0
    DO 7 I=II,III
      DT=DT+DA(I,J)
    7 CONTINUE
    DDT=DT/DIV
    TDA(J,KT)=DDT
  WRITE (6,8) TDA(J,KT),J
  8 FORMAT (5X,F15.7,5X,I2)
  20 CONTINUE
  RETURN

```

```

END
SUBROUTINE HISTO (DA,II,III,AB)
DIMENSION AB(4,10),DA(65,24),KK(24,10)
WRITE (6,84)
84 FORMAT (20X,21HHISTOGRAM FREQUENCIES,/,20X,21(1H=),//)
J=1
82 DO 90 KE=1,10
90 KK(J,KE)=0
DO 80 I=II,III
IF (1.0=DA(I,J)) 50,51,52
50 IF (2.0=DA(I,J)) 53,54,70
53 IF (3.0=DA(I,J)) 56,57,70
56 IF (4.0=DA(I,J)) 59,60,70
59 IF (5.0=DA(I,J)) 62,63,70
62 IF (6.0=DA(I,J)) 65,66,70
65 IF (7.0=DA(I,J)) 68,69,70
68 IF (8.0=DA(I,J)) 100,72,70
100 IF (9.0=DA(I,J)) 70,101,70
51 KK(J,2)=KK(J,2)+1
GO TO 80
54 KK(J,3)=KK(J,3)+1
GO TO 80
57 KK(J,4)=KK(J,4)+1
GO TO 80
60 KK(J,5)=KK(J,5)+1
GO TO 80
63 KK(J,6)=KK(J,6)+1
GO TO 80
66 KK(J,7)=KK(J,7)+1
GO TO 80
69 KK(J,8)=KK(J,8)+1
GO TO 80
72 KK(J,9)=KK(J,9)+1
GO TO 80
101 KK(J,10)=KK(J,10)+1
GO TO 80
52 KK(J,1)=KK(J,1)+1
GO TO 80
70 WRITE (6,88)
88 FORMAT (5X,5HERROR)
80 CONTINUE
WRITE (6,81) (KK(J,KT),KT=1,10),J
81 FORMAT (5X,10I8,5X,I2,/)
J=J+1
IF (J=24) 82,82,83
83 DO 93 I=3,4
DO 93 J=1,10
ABT=KK(I,J)
AB(I,J) =ABT
93 CONTINUE
RETURN
END
SUBROUTINE COORD (AB,DA)
DIMENSION AB(4,10),DIF(4,10),DA(65,24),X(10),Y(10)
DO 203 J=1,10

```

```

    TEM1=AB(3,J)
    TEM2=AB(4,J)
    AB(3,J)=TEM2
    AB(4,J)=TEM1
203 CONTINUE
    DO 201 I=1,65
    TEMP1=DA(I,3)
    TEMP2=DA(I,4)
    DA(I,3)=TEMP2
    DA(I,4)=TEMP1
201 CONTINUE
    DO 150 I=3,4
    N=2*I
    PAR=N
    PI=3.1415926535897
    ANG=PI/PAR
    DO 120 J=1,N
    FJ=J
    XIN=ABS(COS(ANG))
    YIN=ABS(SIN(ANG))
    IF(PI/2.0=ANG) 151,152,152
151 IF(PI=ANG) 154,155,155
154 IF(1.5*PI=ANG)152,155,155
152 X(J)=XIN
    GO TO 156
155 X(J)=-XIN
156 IF(PI/2.0=ANG)157,158,158
157 IF(PI=ANG) 159,158,158
159 Y(J)=-YIN
    GO TO 191
158 Y(J)=YIN
191 ANG=ANG+2.0*PI/PAR
120 CONTINUE
    WRITE (6,123) (X(J),J=1,N),(Y(J),J=1,N)
123 FORMAT (8F10,5,/,8F10,5)
    NE=N+1
    ASUM=0.0
    DO 130 J=2,NE
130 ASUM=ASUM+AB(I,J)
    XMEAN=0.0
    JJ=0
    XMOM=0.0
    DO 131 J=2,NE
    JJ=J-1
131 XMOM=XMOM+X(JJ)*AB(I,J)
    XMEAN=XMOM/ASUM
    YMEAN=0.0
    YMOM=0.0
    JJ=0
    DO 132 J=2,NE
    JJ=J-1
132 YMOM=YMOM+Y(JJ)*AB(I,J)
    YMEAN=YMOM/ASUM
    WRITE (6,501) XMEAN,YMEAN,ASUM
501 FORMAT (3F20,6)

```

```

      DO 133 J=1,N
133  DIF(I,J)=SQRT((X(J)-XMEAN)**2 +(Y(J)-YMEAN)**2 )
      WRITE (6,134) (DIF(I,J),J=1,N)
134  FORMAT (8F10,5)
150  CONTINUE
      DO 139 I=3,4
          N=2*I
          DO 139 J=1,65
          DO 139 K=1,N
          FK=K
          IF(DA(J,I)=FK) 139,140,139
140  DA(J,I)=DIF(I,K)
139  CONTINUE
      DO 202 I=1,65
          TEMP1=DA(I,3)
          TEMP2=DA(I,4)
          DA(I,3)=TEMP2
          DA(I,4)=TEMP1
202  CONTINUE
      RETURN
      END
      SUBROUTINE COVAR (DA,TDA,II,III,KT,COV,L)
      DIMENSION DA(60,24),TDA(24,2),COV(24,24)
      WRITE (6,86)
86  FORMAT (//,60X,17HCOVARIANCE MARRIX,/,60X,,17(1H=)//)
      TAG4=TDA(4,KT)
      TAG3=TDA(3,KT)
      TDA(3,KT)=0,0
      TDA(4,KT)=0,0
      DO 10 J=1,L
      DO 10 K=1,L
          COVT=0,0
          DO 10I=II,III
              COVT=COVT+(DA(I,J)=TDA(J,KT))*(DA(I,K)=TDA(K,KT))
          COV(K,J)=COVT
10  CONTINUE
      DO 90 I=1,L
90  WRITE (6,91) (COV(I,J),J=1,L)
91  FORMAT (//,12F10,3,/,12F10,3,/)
      TDA(3,KT)=TAG3
      TDA(4,KT)=TAG4
      RETURN
      END
$ENTRY

```

NOTE See " Calculation of misclassification probabilities using discriminant analysis" for subroutine CIVER.

DETERMINATION OF POPULATION HOMOGENEITY USING MULTIVARIATE  
ANALYSIS OF VARIANCE

\*\*\*\*\*

```

DIMENSION COVAR(2,2),COVRI(2,2)
DIMENSION AEU(10),XME(32,2),AVEC(10)
DIMENSION LOJ(50)
DIMENSION HUI(10,10),HEUI(10,10)
DIMENSION A(7,7),V(7,7),VR(7,7),AS(7,7),VV(7,7)
DIMENSION DIAG(7,7),VVI(7,7),DIG(7,7)
DIMENSION EIGV(7)
DIMENSION PRO(7),COV(7,7),COVI(7,7),DA(7,14),PRA(7,7)
DIMENSION VMEIG(7),AA(7,7),TA(49),TV(7)
DIMENSION TAAG(700,2),VOT(50)
DIMENSION HEU(10,10)
DIMENSION LOV(50),XSAT(2,50),GSAT(50),HU(10,10),EU(10,10)
DIMENSION AAG(700),ACU(700),W(700),AG(700),CU(700),I(90),LHS(41)
DIMENSION LHSC(61),LAHS(30),NAHS(30)
EQUIVALENCE (ACU(550),DIAG(1,1)),(ACU(600),VVI(1,1)),
2(ACU(650),DIG(1,1))
EQUIVALENCE (ACU(110),GSAT(1)),(ACU(200),VOT(1))
2,(ACU(260),LOV(1)),(ACU(300),A(1,1)),(ACU(350),V(1,1)),(ACU(400)
2,VR(1,1)),(ACU(450),AS(1,1)),(ACU(500),VV(1,1))
EQUIVALENCE (AG(1),TAAG(1,1)),(CU(1),TAAG(1,2))
EQUIVALENCE (W(1),XME(1,1)),(W(100),DA(1,1)),(W(300),HUI(1,1))
EQUIVALENCE (W(450),HEUI(1,1))
EQUIVALENCE (AAG(1),XSAT(1,1)),(AAG(200),HU(1,1)),(AAG(400),EU
2(1,1)),(ACU(1),HEU(1,1))
KTQ=279
KTQ=659
HECKA=0.0175
Q=28.
P=2.0
ITQ=1
KOJ=9
KAJ=41
E=2.718281828
PI=3.141592653589
KCJ=61
JU=30
KT=2
JT=32
READ(5,19) (I(LV),LV=1,16)
19 FORMAT (16I3)
READ(5,18) (I(LV),LV=17,32)
18 FORMAT (16I3)
DO 77 LV=1,JT
77 LOJ(LV)=I(LV)
32 READ(5,10) WT,AGT,CUT,K,LV
10 FORMAT (F2,1,F4,1,F3,1,I1,I2)
IF(CUT) 309,310,309
310 CUT=0.001
309 IF(LV)300,301,300
301 LZ=LZ+1
IF(LZ=5) 32,32,31
300 NU=NU+1

```

```

      II=I(LV)
      IF(K) 11,11,12
12   TW=TW*WT
      TAG=TAG+AGT*WT
      TCU=TCU+CUT*WT
      IF(K=KDJ)32,17,32
17   W(II)=TW
      IF(TAG)260,261,260
260  AAG(II)=ALOG(TAG)
261  IF(TCU)262,264,262
264  KOT=KOT+1
      WRITE(6,263) ITD,NU,KOT
263  FORMAT(3I20)
262  ACU(II)=ALOG(TCU)
      AG(II)=TAG/25.0
      CU(II)=TCU
      TW=0.0
      TAG=0.0
      TCU=0.0
      I(LV)=II+1
      GOTO 30
11   W(II)=WT
      AAG(II)=ALOG(AGT*WT)
      ACU(II)=ALOG(CUT*WT)
      AG(II)=AGT*WT/25.0
      CU(II)=CUT*WT
      I(LV)=II+1
30   ITD=ITD+1
      IF(LV=4)500,400,500
500  IF(LV=9)502,400,502
502  IF(LV=24)503,400,503
503  IF(LV=26)504,400,504
504  IF(LV=31)505,400,505
505  IF(LV=32)506,400,506
506  IF(ACU(II)+2.5) 400,400,274
400  WRITE(6,401) AAG(II),AG(II),ACU(II),CU(II),II,LV,NUC
401  FORMAT (/ ,4F15.6,3I10,/)
      AAG(II)=0.0
      ACU(II)=0.0
      NUC=NUC+1
274  IF(KTD=ITD)31,32,32
31   DO 150 J=1,KTD
150  WRITE (6,51) W(J),AG(J),CU(J),AAG(J),ACU(J),J
51   FORMAT (5F20.4,I10)
      DO 275 II=1,KTD
      AG(II)=0.0
      CU(II)=0.0
      TAAG(II,1)=AAG(II)
      TAAG(II,2)=ACU(II)
275  CONTINUE
      CSQ=0.0
      ASQ=0.0
      EKTD=KTD      =NUC
      TMA=0.0
      TMC=0.0

```

```

    DO 311 J=1,KTD
    TMA=TMA+AAG(J)
    TMC=TMC+ACU(J)
311 CONTINUE
    AMEAN=TMA/EKTD
    CMEAN=TMC/EKTD
    DO 312 J=1,KTD
    IF(AAG(J)) 402,312,402
402 ASQ=ASQ+(AAG(J)-AMEAN)**2
    CSQ=CSQ+(ACU(J)-CMEAN)**2
    ACSQ=ACSQ+(AAG(J)-AMEAN)*(ACU(J)-CMEAN)
312 CONTINUE
    AVAR=ASQ/(EKTD-1.0)
    CVAR=CSQ/(EKTD-1.0)
    ACVAR=ACSQ/(EKTD-1.0)
    SAVAR=SQRT(AVAR)
    SCVAR=SQRT(CVAR)
    COVAC=ACVAR/(SAVAR*SCVAR)
    WRITE(6,315) AMEAN,CMEAN,ASQ,CSQ,SAVAR,SCVAR
315 FORMAT (//,2F20.6,/,2F20.6,/,2F20.6,/)
    WRITE(6,79) ACSQ,ACVAR,COVAC
    79 FORMAT(3F20.6)
    COVAR(1,1)=AVAR
    COVAR(2,2)=CVAR
    COVAR(1,2)=ACVAR
    COVAR(2,1)=ACVAR
    CALL CIVER(COVAR,COVRI,KT,DET)
    DO 340 JJ=1,KT
340 WRITE(6,341) (COVAR(JJ,II),II=1,KT),(COVRI(JJ,II),II=1,KT)
341 FORMAT (//,2F20.6,10X,2F20.6)
    WRITE(6,200) NU,ITD
200 FORMAT (//,2I30)
    LOJ(4)=700
    LOJ(9)=700
    LOJ(24)=700
    LOJ(26)=700
    LOJ(31)=700
    LOJ(32)=700
    DO 74 LV=1,JT
    IF (LOJ(LV)=700) 74,75,74
    75 LOJ(LV)=0
    74 CONTINUE
    JG=0
    DO 42 LV=2,JT
    IF(LOJ(LV)) 43,73,43
43 LJ=LV-JG-1
    LOV(LJ)=LOJ(LV)-LOJ(LJ)
    JG=0
    GO TO 42
    73 JG=JG+1
    LOV(LV)=0
42 CONTINUE
    LOV(JT)=KTD-LOJ(JT)+1
    LOV(4)=0
    LOV(9)=0

```

```

LOV(24)=0
LOV(26)=0
LOV(30)=25
LOV(31)=0
LOV(32)=0
DO 63 LJ=1,JT
63 VOT(LJ)=LOV(LJ)
DO 76 LJ=1,JT
76 VOTI=VOTI+VOT(LJ)
DO 60 LJ=1,JT
60 WRITE(6,58) VOTI,VOT(LJ),LJ
58 FORMAT (2F20,6,I10)
DO 47 K=1,KT
DO 147 LV=1,JT
IF(LOV(LV))46,145,46
46 JIJ=LOJ(LV)+LOV(LV)=1
JIT=LOJ(LV)
XSA=0.0
XIT=LOV(LV)
DO 148 II=JIT,JIJ
IF(TAAG(II,K)) 460,461,460
461 XIT=XIT=1.0
460 XSA=XSA+TAAG(II,K)
XSAT(K,LV)=XSA
XME(LV,K)=XSA/XIT
148 CONTINUE
WRITE(6,700) XSAT(K,LV),XME(LV,K),LV,K
700 FORMAT (/ ,2F20,6,2I10)
VOT(LV)=XIT
GO TO 147
145 XSAT(K,LV)=0.0
147 CONTINUE
47 CONTINUE
DO 49 K=1,KT
GSA=0.0
DO 49 LV=1,JT
GSA=GSA+XSAT(K,LV)
GSAT(K)=GSA
WRITE(6,701)GSA,K
701 FORMAT (F20,6,I10)
49 CONTINUE
DO 52 K=1,KT
DO 52 IC=1,KT
HUV2=(GSAT(K)*GSAT(IC))/EKTO
HUV1=0.0
DO 55 LV=1,JT
IF(LOV(LV)) 55,55,71
71 HUV1=HUV1+(XSAT(K,LV)*XSAT(IC,LV))/VOT(LV)
55 CONTINUE
EUV1=0.0
DO 160 LV=1,JT
IF (LOJ(LV))64,160,64
64 JIT=LOJ(LV)
JIJ=LOJ(LV)+LOV(LV)=1
DO 50 II=JIT,JIJ

```

```

EUV1=EUV1+TAAG(II,K)*TAAG(II,IC)
50 CONTINUE
160 CONTINUE
HU(K,IC)=HUV1-HUV2
EU(K,IC)=EUV1-HUV1
HEU(K,IC)=HU(K,IC)+EU(K,IC)
52 CONTINUE
  DO 56 K=1,KT
56 WRITE(6,54) (HU(K,IC),IC=1,KT),(EU(K,IC),IC=1,KT)
54 FORMAT (2F20,6,20X,2F20,6)
  DO 181 K=1,KT
181 WRITE(6,182) (HEU(K,IC),IC=1,KT)
182 FORMAT (//,2F20,6)
  CALL CIVER (HEU,HEUI,KT,DET)
  DO 184 K=1,KT
  DO 184 IC=1,KT
  EGU=0.0
  DO 184 IT=1,KT
  EGU=EGU+HU(K,IT)*HEUI(IT,IC)
  HUI(K,IC)=EGU
  WRITE(6,702)EGU
702 FORMAT(F20,8)
184 CONTINUE
  CALL EIGEN (KT,HUI,VMEIG)
  HECK=HECKA/(1.0-HECKA)
  DO 820 J=1,JT
  DO 821 K=1,JT
  IF(J=K) 821,821,822
822 IF(LOV(J)) 820,820,823
823 IF(LOV(K)) 821,821,824
824 CONST=HECK*FLOAT(LOV(J)+LOV(K))/FLOAT(LOV(J)*LOV(K))
  DO 810 JU=1,KT
  AVEC(JU)=1.0
  DO 809 JL=1,KT
  IF(JU=JL) 808,809,808
808 AVEC(JL)=0.0
809 CONTINUE
  CALL CI(AVEC,XME,CONST,EU,ULP,UMP,J,K,KT,JU)
810 CONTINUE
821 CONTINUE
820 CONTINUE
  Q1=Q-1.0
  S=AMINI(Q1,P)
  EM=(ABS(Q-P-1.0)-1.0)/2.0
  EN=(EKTD-Q-P-1.0)/2.0
  WRITE(6,829) S,EM,EN,VMEIG(1)
829 FORMAT (//,3F10.1,F20,6)
  STOP
  END
  SUBROUTINE CIVER (COV,COVI,L,DET)
  DOUBLE PRECISION DA,DAT,DIK,DATT,DKL
  DOUBLE PRECISION DUM1,DUM2
  DIMENSION PRD(7),COV(7,7),COVI(7,7),DA(7,14),PRA(7,7)
  EQUIVALENCE (DA(1,1),PRA(1,1))
  DET=1.0

```

```

M=L*2
DO 300 I=1,L
DO 300 J=1,L
300 DA(I,J)=CQV(I,J)
DO 301 I=1,L
LLI=L+1
DO 301 J=LLI,M
KN=J-I
IF(KN=L)302,303,302
302 DA(I,J)=0,0
GO TO 301
303 DA(I,J)=1,0
301 CONTINUE
IP=1
JP=1
112 IK=1
DUM1=0,0
DUM2=0,0
IMAX=IP
II=IP+1
DO 201 I=II,L
IF (DA(IP,JP)=DA(I,JP))200,201,201
200 DUM2=DA(I,JP)
IF(DUM1=DUM2)202,201,201
202 IMAX=I
DUM1=DUM2
201 CONTINUE
IF(IMAX=IP)205,205,206
206 DO 204 J=JP,M
TEMP1=DA(IP,J)
TEMP2=DA(IMAX,J)
DA(IP,J)=TEMP2
DA(IMAX,J)=TEMP1
204 CONTINUE
SIGN=-1,0
GO TO 207
205 SIGN=1,0
207 DET=DET*SIGN*DA(IP,JP)
IF (DA(IP,JP))100,190,102
100 DO 103 J=1,M
103 DA(IP,J)=-DA(IP,J)
102 DID = DA(IP,JP)
DO 104 J=JP,M
DAT=DA(IP,J)/DID
DA(IP,J)=DAT
IF (DABS(DA(IP,J))=0,1D=06) 126,126,104
126 DA(IP,J)=0,0
104 CONTINUE
IK=IP+1
IF(IK=L-1) 110,113,113
110 SIG=-1,0
107 DIK=DA(IK,JP)
DO 109 J=JP,M
DATT=(SIG*DA(IP,J)*DIK)+DA(IK,J)
DA(IK,J)=DATT

```

```

      IF (DABS(DA(IK,J))-0,1D=06) 127,127,109
127 DA(IK,J)=0,0
109 CONTINUE
108 IK=IK+1
      IF(IK=L) 110,110,111
111 IP=IP+1
      JP=JP+1
      IF(JP=L) 112,205,113
113 LI=L
      LJ=L
58 KL=LI-1
56 SIG=-1,0
54 DKL=      DA(KL,LJ)
      DO 53 J=LJ,M
      DATT=(SIG*DA(LI,J)*DKL)+DA(KL,J)
      DA(KL,J)=DATT
      IF (DABS(DA(KL,J))-0,1D=06) 128,128,53
128 DA(KL,J)=0,0
53 CONTINUE
      KL=KL-1
      IF(KL=1) 55,56,56
55 LJ=LJ-1
      LI=LI-1
      IF(LI=1) 57,57,58
57 DO 304 I=1,L
      JLL=L+1
      DO 304 J=JLL,M
      KM=J-L
      COVI(I,KM)=DA(I,J)
304 CONTINUE
      DO 400 I=1,L
400 WRITE (6,401) (DA(I,J),J=1,M)
401 FORMAT (8D15,4,/,8D15,4,/,8D15,4,/,8D15,4,/,8D15,4,/,8D15,4,/)
      DO 305 I=1,L
      DO 305 J=1,L
      PRI=0,0
      DO 305 JJ=1,L
      PRI=PRI+COV(J,JJ)*COVI(JJ,I)
      PRA(J,I)=PRI
305 CONTINUE
      DO 306 I=1,L
306 WRITE (6,321) (PRA(I,J),J=1,L)
321 FORMAT (//,24F5,2,/)
      CN=1,0
      DO 210 I=1,L
      CNQR=0,0
      DO 290 J=1,L
290      CNQR=CNQR+(COV(I,J))**2
      CND=SQRT(CNQR)
      CN=CN*CND
210 CONTINUE
      DVI=DET/CN
      WRITE (6,212) DET,CN,DVI
212 FORMAT (//,3F20,8,/)
      GO TO 193

```

```

190 WRITE (6,191)
191 FORMAT (/,'15MATRIX SINGULAR')
192 RETURN
END
SUBROUTINE EIGEN(L,AA,VMEIG)
DIMENSION PRO(7),COV(7,7),COVI(7,7),DA(7,14),PRA(7,7)
DIMENSION A(7,7),V(7,7),VR(7,7),AS(7,7),VV(7,7)
DIMENSION DIAG(7,7),VVI(7,7),DIG(7,7)
DIMENSION EIGV(7)
DIMENSION VMEIG(7), AA(7,7),TA(49),TV(7)
EQUIVALENCE (AS(1,1),VVI(1,1)),(A(1,1),DIAG(1,1))
IK=1
TEST =1,OE=04
DO 54 I=1,L
DO 54 J=1,L
A(I,J)=AA(I,J)
54 CONTINUE
91 DO 100 I=1,L
DO 100 J=1,L
AS(I,J)=A(I,J)
100 CONTINUE
DO 8 I=1,L
V(I,IK)=A(I,I)
TV(I)=V(I,IK)
8 CONTINUE
WRITE(6,52) (V(I,IK),I=1,L)
52 FORMAT (7F15,3)
DO 95 I=1,L
IF(TV(1)-TV(I)) 94,95,95
94 TV(1)=TV(I)
95 CONTINUE
TMUX=TV(1)
DO 110 I=1,L
110 V(I,IK)=A(I,I)/TMUX
21 DO 99 I=1,L
DO 99 J=1,L
NP=(I-1)*L+J
TA(NP)=AS(I,J)
99 CONTINUE
NUM=NUM+1
NL=L*L
DO 97 J=1,NL
IF(TA(1)=TA(J))96,97,97
96 TA(1)=TA(J)
97 CONTINUE
TMAX=TA(1)
DO 11 K=1,L
DO 11 I=1,L
WRITE (6,88) AS(K,I),K,I,TMAX
88 FORMAT (F40.6,2I10,F20.3)
TEMP =0.0
DO 11 J=1,L
TEMP=TEMP+(AS(K,J)*AS(J,I))/(TMAX**2)
COV(K,I)=TEMP
11 CONTINUE

```

```

      DO 250 I=1,L
      DO 250 J=1,L
      AS(I,J)=CDV(I,J)
      CDV(I,J)=0,0
      WRITE(6,401) AS(I,J),CDV(I,J)
401  FORMAT (/ ,2F25,6)
250  CONTINUE
      IF (NUH=3) 21,21,360
360  DO 12 K=1,L
      VTEM=0,0
      DO 251 J=1,L
      VTEM=VTEM+V(J,IK)*AS(K,J)
251  CONTINUE
      VR(K,IK)=VTEM
      WRITE(6,402) VTEM
402  FORMAT (/ ,F30,6)
      12 CONTINUE
      VMAX=AMAX1(VR(1,IK),VR(2,IK),VR(3,IK),VR(4,IK),VR(5,IK),VR(6,IK),
      3 VR(7,IK))
      DO 13 J=1,L
      VRT=VR(J,IK)/VMAX
      VR(J,IK)=VRT
      13 CONTINUE
      WRITE (6,53) (VR(J,IK),J=1,L)
      53  FORMAT (7F15,3)
      NU=NU+1
      IFLAG=0
      DO 14 J=1,L
      VDUF=ABS(V(J,IK)-VR(J,IK))
      IF (VDUF=TEST) 14,14,16
      16  IFLAG=1
      14  CONTINUE
      18  DO 20 J=1,L
      20  V(J,IK)=VR(J,IK)
      IF(IFLAG) 17,15,17
      17  IF(NU=100) 21,21,19
      19  WRITE (6,40) NU
      40  FORMAT (I20,5X,16HND OF ITERATIONS,/)
      NUM=0
      SVLEN=0,0
      DO 22 J=1,L
      22  SVLEN=SVLEN+(V(J,IK))**2
      VLEN=SQRT(SVLEN)
      DO 23 I=1,L
      VT=V(I,IK)/VLEN
      VV(I,IK)=VT
      23  CONTINUE
C     CHECK EIGENVALUES NORMAL
      CVLEN=0,0
      DO 24 I=1,L
      24  CVLEN=CVLEN+(VV(I,IK))**2
      WRITE (6,25) IK,CVLEN,(VV(I,IK),I=1,L)
      25  FORMAT (/ ,15,F10,3,7F12,4)
C     FROM EIGENVECTOR FIND EIGENVALUE
      DO 240 I=1,L

```

```

EIG=0.0
DO 26 J=1,L
26 EIG=EIG+A(I,J)*VV(J,IK)
EIGV(I)=EIG/VV(I,IK)
WRITE (6,27) EIGV(I),I
27 FORMAT (/,'F30.9',I10,/)
240 CONTINUE
TEG=0.0
DO 28 I=1,L
28 TEG=TEG+EIGV(I)
ELL=L
VMEIG(IK)=TEG/ELL
FORM R1 MATRIX
DO 29 I=1,L
DO 29 J=1,L
RT=A(I,J)=VMEIG(IK)*VV(I,IK)*VV(J,IK)
A(I,J)=RT
29 CONTINUE
IK=IK+1
IF(IK=L) 91,91,30
30 DO 31 I=1,L
31 WRITE (6,32) VMEIG(I),I,(VV(I,J),J=1,L)
32 FORMAT (F15.5,I10,7F13,4)
CALL CIVER (VV,VVI,L,DET)
DO 33 K=1,L
DO 33 I=1,L
TUP=0.0
DO 33 J=1,L
TUP=TUP+VVI(I,J)*AA(J,K)
DIG(I,K)=TUP
33 CONTINUE
DO 36 K=1,L
DO 36 I=1,L
TIP=0.0
DO 36 J=1,L
TIP=TIP+DIG(I,J)*VV(J,K)
DIAG(I,K)=TIP
36 CONTINUE
DO 34 I=1,L
34 WRITE (6,35) (DIAG(I,J),J=1,L)
35 FORMAT (7F15,5)
19 RETURN
END
SUBROUTINE CI (AVEC,XME,CONST,EU,ULP,UMP,J,K,KT,JD)
DIMENSION AVEC(10),XME(32,2),EU(10,10),AEU(10)
UPI=0.0
DO 818 JJ=1,KT
818 UPI=UPI+AVEC(JJ)*(XME(J,JJ)-XME(K,JJ))
D=0.0
DO 817 IJ=1,KT
C=0.0
DO 816 JJ=1,KT
C=C+AVEC(JJ)*EU(JJ,IJ)
AEU(IJ)=C
816 CONTINUE

```

```
      D=D+AEU(IJ)*AVEC(IJ)
817  CONTINUE
      DIFF=SQRT(D*CONST)
      ULP=UPI-DIFF
      UMP=UPI+DIFF
      WRITE(6,815) ULP,UMP,J,K,JD
815  FORMAT(//,2F30,6,3I10)
      RETURN
      END
$ENTRY
```

```

C      DETERMINATION OF CHARACTERISTIC DEPENDENCIES USING CONTINGENCY
C      TABLES
C      *****
DIMENSION SUMR(9),SUMC(9),THBA(9,9),IND(24),BA(9,9),DA(65,24)
READ (5,199) (IND(I),I=1,24)
199  FORMAT(24I1)
      III=53
      III=28
      DO 196 I=1,III
197  READ(5,197) (DA(I,J),J=1,24)
      FORMAT (24F1,0)
198  CONTINUE
      DO 198 J=1,24
      KK=I+1
      DO 198 J=KK,24
      INI=IND(I)
      INJ=IND(J)
      DO 200 KJ=1,INI
      EJK=KJ
      DO 200 KJJ=1,INJ
      EJJ=KJJ
      BAT=0.0
      BA(KJ,KJJ)=0.0
      DO 200 L=1,60
201  IF(EJK=DA(L,I))200,201,200
202  IF(EJJ=DA(L,J))200,202,200
202  BAT=BAT+1.0
      BA(KJ,KJJ)=BAT
200  CONTINUE
      DO 204 KJ=1,INI
204  WRITE (6,205) (BA(KJ,KJJ),KJJ=1,INJ)
205  FORMAT (9F10,5,/)
      DO 206 KJ=1,INI
      SUML=0.0
      DO 206 KJJ=1,INJ
      SUML=SUML+BA(KJ,KJJ)
      SUMR(KJ)=SUML
206  CONTINUE
      DO 207 KJJ=1,INJ
      SUMI=0.0
      DO 207 KJ=1,INI
      SUMI=SUMI+BA(KJ,KJJ)
      SUMC(KJJ)=SUMI
207  CONTINUE
      TSUMR=0.0
      DO 208 KJ=1,INI
208  TSUMR=TSUMR+SUMR(KJ)
      TSUMC=0.0
      DO 209 KJJ=1,INJ
209  TSUMC=TSUMC+SUMC(KJJ)
      DIF=TSUMC-TSUMR
      WRITE(6,210) DIF,TSUMR,TSUMC
210  FORMAT (3F20,8)
      DO 211 KJ=1,INI
      DO 211 KJJ=1,INJ

```

```
      THB=SUMR(KJ)*SUMC(KJJ)/TSUMR
      THBA(KJ,KJJ)=THB
211  CONTINUE
      XSQ=0.0
      DO 212 KJ=1,INI
        DO 212 KJJ=1,INJ
212  XSQ=XSQ+((BA(KJ,KJJ)-THBA(KJ,KJJ))**2)/THBA(KJ,KJJ)
      IFJ=(IND(I)=1)*(IND(J)=1)
      WRITE (6,213) XSQ,I,J,IFJ
213  FORMAT (F20.8,3I10,///)
198  CONTINUE
      STOP
      END
$ENTRY
```

## CALCULATION OF MISCLASSIFICATION PROBABILITIES USING DISCRIMINANT ANALYSIS

\*\*\*\*\*

```

DIMENSION UNK(19,7), UNK2(19,2), IK(2), IIK(2), CHI(10), QRE(660,2),
2 WAST(280,2), TD(2), TW(2), G(2), BU(2,2), WU(2,2), DA(7,14), PRQ(7),
2 COV(7,7), PRA(7,7), COVI(7,7), WUI(2,2), BWUI(2,2), AA(7,7), EIGV(7),
2 VMEIG(7), VV(7,7), V(7,7), VR(7,7), TA(49), TV(7), A(7,7), DIG(7,7), DIAG
2(7,7), YUNK(19,2), YST(660,2), YORE(660,2), YWAST(280,2), YDA(2,2),
2 WCOVI(2,2), WCOVI(2,2), TEMP(19,2), DISQ(2,19), YDIF(19,2)
2 , YDIFD(19,2), YDIFW(19,2)
DIMENSION AS(7,7), VVI(7,7)
DIMENSION DEPRQ(19,2), WASPR(19), QREPR(19)
EQUIVALENCE (QRE(1,1), YST(1,1))
EQUIVALENCE (UNK(1,1), YUNK(1,1)), (UNK(1,3), TEMP(1,1)),
2 (UNK(1,5), YDIF(1,1))

```

L=7

NUNC=19

E=2.718281828

PI=3.141592653589

KT=2

DO 441 I=1, NUNC

441 READ(5,440) (UNK(I,J), J=1,L), (UNK2(I,K), K=1,KT)

440 FORMAT(7F1,0,2F4,2)

L=2

READ(5,96) (CHI(I), I=1,L)

96 FORMAT(7F4,2)

READ(5,102) (IK(J), J=1,KT)

102 FORMAT(2I3)

DO 472 IL=1,KT

KTQ=IK(IL)

IF(IL=1) 474,473,474

473 DO 471 II=1,KTQ

KD=KD+1

READ(5,470) QRE(II,1), QRE(II,2)

470 FORMAT(2F20,10)

IF(QRE(II,1)) 471,570,471

570 WRITE(6,571) KD

571 FORMAT(I20)

471 CONTINUE

GO TO 472

474 DO 479 II=1,KTQ

KW=KW+1

READ(5,476) (WAST(II,J), J=1,KT)

476 FORMAT(2F20,10)

IF(WAST(II,1)) 479,572,479

572 WRITE(6,573) KW

573 FORMAT(I20)

479 CONTINUE

472 CONTINUE

DO 110 IL=1,KT

III=IK(IL)

DO 111 J=1,L

IF(IL=1) 112,112,113

112 TOR=0.0

DO 114 I=1,III

```

TOR=TOR+QRE(I,J)
TQ(J)=TOR
114 CONTINUE
GO TO 111
113 TWA=0.0
DO 115 I=1,III
TWA=TWA+WAST(I,J)
TW(J)=TWA
115 CONTINUE
111 CONTINUE
110 CONTINUE
DO 116 J=1,L
116 G(J)=TQ(J)+TW(J)
DO 117 I=1,L
DO 117 J=1,L
WUV1=0.0
III=MAX0(IK(1),IK(2))
DO 141 II=1,III
IF(II-IK(2)) 118,118,140
118 WUV1=WUV1+QRE(II,J)*QRE(II,I)+WAST(II,J)*WAST(II,I)
GO TO 141
140 WUV1=WUV1+QRE(II,J)*QRE(II,I)
141 CONTINUE
BUV1=(TW(I)*TW(J))/FLOAT(IK(2))+(TQ(I)*TQ(J))/FLOAT(IK(1))
BUV2=G(J)*G(I)/FLOAT(IK(1)+IK(2))
BU(I,J)=BUV1-BUV2
WU(I,J)=WUV1-BUV1
117 CONTINUE
DO 119 J=1,L
119 WRITE(6,120) (BU(I,J),J=1,L),(WU(I,J),J=1,L)
120 FORMAT (//,4X,7F8.2,10X,7F7.2)
CALL CIVER(WU,WUI,L,DET)
DO 121 I=1,L
DO 121 J=1,L
BWI=0.0
DO 121 II=1,L
BWI=BWI+WUI(I,II)*BU(II,J)
BWUI(I,J)=BWI
121 CONTINUE
DO 451 I=1,L
451 WRITE(6,452) (BWUI(I,J),J=1,L)
452 FORMAT (//,7F15.4)
DO 450 I=1,KT
450 IIK(I)=IK(I)
CALL EIGEN(L,BWUI,VMEIG,VV,KT,IIK,CHI,LT)
IF(LT=3) 477,478,478
477 DO 402 J=1,LT
DO 402 I=1,NUNC.
YD=0.0
DO 402 II=1,L
YD=YD+VV(II,J)*UNK2(I,II)
YUNK(I,J)=YD
WRITE(6,409)YUNK(I,J),I,J
409 FORMAT (//,F20.9,2I10)
402 CONTINUE

```

```

      DO 122 IL=1,KT
      III=IK(IL)
      DO 123 J=1,LT
      DO 128 I=1,III
      IF(IL=1) 124,124,125
124  YD=0,0
      DO 126 II=1,L
      YD=YD+VV(II,J)*ORE(I,II)
      YORE(I,J)=YD
126  CONTINUE
      GO TO 128
125  YW=0,0
      DO 127 II=1,L
      YW=YW+VV(II,J)*WAST(I,II)
      YWAST(I,J)=YW
127  CONTINUE
128  CONTINUE
123  CONTINUE
122  CONTINUE
      DO 132 IL=1,KT
      III=IK(IL)
      IF(IL=1) 133,133,134
133  CALL MEAN(YORE,YDA,IL,III,LT)
      GO TO 132
134  CALL MEAN(YWAST,YDA,IL,III,LT)
132  CONTINUE
      DO 135 IL=1,KT
      III=IK(IL)
      IF(IL=1) 136,136,137
136  CALL COVAR(YORE,YDA,III,IL,COV,LT)
      CALL CIVER(COV,OCOVI,LT,DET)
      CALL ZTRSO(YUNK,IL,YDA,YDIFO,NUNC,LT,OCOVI,DISQ)
C   HYPOTHESIS WASTE PROBABILITY OF BEING ORE
      DO 436 J=1,NUNC
      CALL PROBD(YDIFO,J,OCOVI,PI,E,DET,S,LT)
      OREPR(J)=1,0=S
      WRITE(6,435) OREPR(J),DISQ(IL,J),YDIFO(J,1),J
435  FORMAT(/,3F25,6,115)
436  CONTINUE
      GO TO 135
137  CALL COVAR(YWAST,YDA,III,IL,COV,LT)
      CALL CIVER(COV,WCOVI,LT,DET)
      CALL ZTRSO(YUNK,IL,YDA,YDIFW,NUNC,LT,WCOVI,DISQ)
C   HYPOTHESIS ORE PROBABILITY OF BEING WASTE
      DO 439 J=1,NUNC
      CALL PROBD(YDIFW,J,WCOVI,PI,E,DET,S,LT)
      WASPR(J)=S
      WRITE(6,438) WASPR(J),DISQ(IL,J),YDIFW(J,1),J
438  FORMAT(/,3F30,6,110)
439  CONTINUE
135  CONTINUE
      DO 530 I=1,NUNC
      DO 529 J=1,KT
      IF(J=1) 520,521,520
521  TEMPI=OREPR(I)/(OREPR(I)+1,0=WASPR(I))

```

```

TEMP2=WASPR(I)/(1.0-DREPR(I)+WASPR(I))
DEPRO(I,J)=TEMP1/(TEMP1+TEMP2)
GO TO 529
520 DEPRO(I,J)=1.0-DEPRO(I,1)
529 CONTINUE
WRITE(6,531) (DEPRO(I,J),J=1,KT)
531 FORMAT (//,2F30,8,2I20,/)
530 CONTINUE
478 STOP
END
SUBROUTINE MEAN(YST,YDA,IL,III,LT)
DIMENSION YDA(2,2),YST(660,2)
WRITE (6,85)
85 FORMAT (//,12X,4HMEAN,/,12X,4(1H-),/)
DIV=III
DO 20 J=1,LT
DT=0.0
DO 7 I=1,III
DT=DT+YST(I,J)
7 CONTINUE
DDT=DT/DIV
YDA(J,IL)=DDT
WRITE(6,8)YDA(J,IL),J,IL
8 FORMAT(F40,7,2I10)
20 CONTINUE
RETURN
END
SUBROUTINE CIVER(COV,COVI,L,DET)
DOUBLE PRECISION DA,DAT,DIK,DATT,DKL
DOUBLE PRECISION DUM1,DUM2
DIMENSIONPRO(7),COV(7,7),DA(7,14),PRA(7,7),COVI(7,7)
IF(L=1) 491,491,490
491 COVI(L,L)=1.0/COV(L,L)
GO TO 193
490 DET=1.0
M=L*2
DO 300 I=1,L
DO 300 J=1,L
300 DA(I,J) =COV(I,J)
DO 301 I=1,L
LLI=L+1
DO 301 J=LLI,M
KN=J-I
IF(KN=L) 302,303,302
302 DA(I,J)=0.0
GO TO 301
303 DA(I,J)=1.0
301 CONTINUE
IP=1
JP=1
112 IK=1
DUM1=0.0
DUM2=0.0
IMAX=IP
II=IP+1

```

```

      DO 201 I=II,L
      IF (DA(IP,JP)-DA(I,JP))200,201,201
200  DUM2=DA(I,JP)
      IF(DUM1=DUM2)202,201,201
202  IMAX=I
      DUM1=DUM2
201  CONTINUE
      IF(IMAX=IP)205,205,206
206  DO 204 J=JP,M
      TEMP1=DA(IP,J)
      TEMP2=DA(IMAX,J)
      DA(IP,J)=TEMP2
      DA(IMAX,J)=TEMP1
204  CONTINUE
      SIGN=-1.0
      GO TO 207
205  SIGN=1.0
207  DET=DET*SIGN*DA(IP,JP)
      IF (DA(IP,JP))100,190,102
100  DO 103 J=1,M
103  DA(IP,J)=-DA(IP,J)
102  DID = DA(IP,JP)
      DO 104 J=JP,M
      DAT=DA(IP,J)/DID
      DA(IP,J)=DAT
      IF (DABS(DA(IP,J))-0.1D-06) 126,126,104
126  DA(IP,J)=0.0
104  CONTINUE
      IK=IP+1
      IF(IK=L-1) 110,113,113
110  SIG=-1.0
107  DIK=DA(IK,JP)
      DO 109 J=JP,M
      DATT=(SIG*DA(IP,J)*DIK)+DA(IK,J)
      DA(IK,J)=DATT
      IF (DABS(DA(IK,J))-0.1D-06) 127,127,109
127  DA(IK,J)=0.0
109  CONTINUE
108  IK=IK+1
      IF(IK=L) 110,110,111
111  IP=IP+1
      JP=JP+1
      IF(JP=L)112,205,113
113  LI=L
      LJ=L
58  KL=LI-1
56  SIG=-1.0
54  DKL= DA(KL,LJ)
      DO 53 J=LJ,M
      DATT =(SIG*DA(LI,J)*DKL)+DA(KL,J)
      DA(KL,J) =DATT
      IF (DABS(DA(KL,J))-0.1D-06) 128,128,53
128  DA(KL,J)=0.0
53  CONTINUE
      KL=KL-1

```

```

      IF(KL=1) 55,56,56
55  LJ=LJ-1
      LI=LI-1
      IF(LI=1) 57,57,58
57  DO 304 I=1,L
      JLL=L+1
      DO 304 J=JLL,M
      KM=J-L
      COVI(I,KM)=DA(I,J)
304  CONTINUE
      DO 400 I=1,L
400  WRITE (6,401) (DA(I,J),J=1,M)
401  FORMAT (8D15,4,/,8D15,4,/,8D15,4,/,8D15,4,/,8D15,4,/,8D15,4,/)
      DO 305 I=1,L
      DO 305 J=1,L
      PRI=0.0
      DO 305 JJ=1,L
      PRI=PRI+COV (J,JJ)*COVI(JJ,I)
      PRA(J,I)=PRI
305  CONTINUE
      DO 306 I=1,L
306  WRITE (6,321) (PRA(I,J),J=1,L)
321  FORMAT (//,24F5,2,/)
      CN=1.0
      DO 210 I=1,L
      CNQR=0.0
      DO 290 J=1,L
290  CNQR=CNQR+(COV(I,J))**2
      CNQ=SQRT(CNQR)
      CN=CN*CNQ
210  CONTINUE
      DVI=DET/CN
      WRITE (6,212) DET,CN,DVI
212  FORMAT (//,3F39.17,/)
      GO TO 193
190  WRITE (6,191)
191  FORMAT (/,15HMATRIX SINGULAR)
193  RETURN
      END
      SUBROUTINE COVAR (YST,YDA,III,IL,COV,LT)
      DIMENSION YST(660,2),YDA(2,2),COV(2,2)
      WRITE (6,86)
86  FORMAT (//,60X,17HCOVARIANCE MATRIX,/,60X,17(1H-)//)
      DO 10 J=1,LT
      DO 10 K=1,LT
      COVT=0.0
      DO 10 I=1,III
      COVT=COVT+(YST(I,J)-YDA(J,IL))*(YST(I,K)-YDA(K,IL))
      COV(J,K)=COVT/FLOAT(III-1)
10  CONTINUE
      DO 90 I=1,LT
80  WRITE(6,91) (COV(I,J),J=1,LT)
91  FORMAT (//,7F15,4,/)
      RETURN
      END

```

```

SUBROUTINE EIGEN (L,AA,VMEIG,VV,KT,IJK,CHI,LT)
DIMENSION CHI(10),IJK(2)
DIMENSION AA(7,7),EIGV(7),VMEIG(7),VV(7,7),V(7,7),VR(7,7)
DIMENSION TA(49),TV(7),A(7,7),AS(7,7)
DIMENSION VVI(7,7),DIG(7,7),DIAG(7,7)
DIMENSION PRO(7),COV(7,7),DA(7,14),PRA(7,7),CDVI(7,7)
EQUIVALENCE (AS(1,1),VVI(1,1)),(A(1,1),DIAG(1,1))
EQUIVALENCE (V(1,1),DIG(1,1)),(VR(1,1),CDVI(1,1))
IJK=1
TEST =1,OE=04
DO 54 I=1,L
DO 54 J=1,L
A(I,J)=AA(I,J)
54 CONTINUE
91 DO 100 I=1,L
DO 100 J=1,L
AS(I,J)=A(I,J)
100 CONTINUE
DO 8 I=1,L
V(I,IK)=A(I,I)
TV(I)=V(I,IK)
8 CONTINUE
WRITE(6,52) (V(I,IK),I=1,L)
52 FORMAT (7F15,3)
DO 95 I=1,L
IF(TV(1)-TV(I)) 94,95,95
94 TV(1)=TV(I)
95 CONTINUE
TMUX=TV(1)
DO 110 I=1,L
110 V(I,IK)=A(I,I)/TMUX
21 DO 99 I=1,L
DO 99 J=1,L
NP=(I-1)*L+J
TA(NP)=AS(I,J)
99 CONTINUE
NUM=NUM+1
NL=L*L
DO 97 J=1,NL
IF(TA(1)-TA(J))96,97,97
96 TA(1)=TA(J)
97 CONTINUE
TMAX=TA(1)
DO 11 K=1,L
DO 11 I=1,L
WRITE (6,88) AS(K,I),K,I,TMAX
88 FORMAT (F40.6,2I10,F20,3)
TEMP =0.0
DO 11 J=1,L
TEMP=TEMP+(AS(K,J)*AS(J,I))/(TMAX**2)
COV(K,I)=TEMP
11 CONTINUE
DO 250 I=1,L
DO 250 J=1,L
AS(I,J)=COV(I,J)

```

```

      COV(I,J)=0,0
250 CONTINUE
      IF (NUM=3)21,21,360
360 DO 12 K=1,L
      VTEM=0,0
      DO 251 J=1,L
      VTEM=VTEM+V(J,IK)*AS(K,J)
251 CONTINUE
      VR(K,IK)=VTEM
      12 CONTINUE
      VMAX=AMAX1(VR(1,IK),VR(2,IK),VR(3,IK),VR(4,IK),VR(5,IK),VR(6,IK),
3 VR(7,IK))
      DO 13 J=1,L
      VRT=VR(J,IK)/VMAX
      VR(J,IK)=VRT
      13 CONTINUE
      WRITE (6,53) (VR(J,IK),J=1,L)
53 FORMAT (7F15,3)
      NU=NU+1
      IFLAG=0
      DO 14 J=1,L
      VDUF=ABS(V(J,IK)-VR(J,IK))
      IF (VDUF=TEST)14,14,16
      16 IFLAG=1
      14 CONTINUE
      18 DO 20 J=1,L
      20 V(J,IK)=VR(J,IK)
      IF(IFLAG)17,15,17
      17 IF(NU=100) 21,21,19
      15 WRITE (6,40) NU
      40 FORMAT (I20,5X,16HNO OF ITERATIONS,/)
      NUM=0
      SVLEN=0,0
      DO 22 J=1,L
      22 SVLEN=SVLEN+(V(J,IK))**2
      VLEN=SQRT(SVLEN)
      DO 23 I=1,L
      VT=V(I,IK)/VLEN
      VV(I,IK)=VT
      23 CONTINUE
C CHECK EIGENVALUES NORMAL
      CVLEN=0,0
      DO 24 I=1,L
      24 CVLEN=CVLEN+(VV(I,IK))**2
      WRITE (6,25) IK,CVLEN,(VV(I,IK),I=1,L)
      25 FORMAT (/,15,F10,3,7F12,4)
C FROM EIGENVECTOR FIND EIGENVALUE
      DO 240 I=1,L
      EIG=0,0
      DO 26 J=1,L
      26 EIG=EIG+A(I,J)*VV(J,IK)
      EIGV(I)=EIG/VV(I,IK)
      WRITE (6,27) EIGV(I),I
      27 FORMAT (/,F30,9,I10,/)
      240 CONTINUE

```

```

      TEG=0.0
      DO 28 I=1,L
28  TEG=TEG+EIGV(I)
      ELL=L
      VMEIG(IK)=TEG/ELL
      CALL RAD (VMEIG,L,KT,IK,IJK,XS)
      IF(XS=CHI(IK)) 19,117,117
C 117 LT=LT+1
      FORM R1 MATRIX
      DO 29 I=1,L
      DO 29 J=1,L
      RT=A(I,J)=VMEIG(IK)*VV(I,IK)*VV(J,IK)
      A(I,J)=RT
29  CONTINUE
      IK=IK+1
      IF(IK=L) 91,91,30
30  DO 31 I=1,L
31  WRITE (6,32) VMEIG(I),I,(VV(I,J),J=1,L)
32  FORMAT (F15.5,I10,7F13.4)
      CALL CIVER (VV,VVI,L,DET)
      DO 33 K=1,L
      DO 33 I=1,L
      TUP=0.0
      DO 33 J=1,L
      TUP=TUP+VVI(I,J)*AA(J,K)
      DIG(I,K)=TUP
33  CONTINUE
      DO 36 K=1,L
      DO 36 I=1,L
      TIP=0.0
      DO 36 J=1,L
      TIP=TIP+DIG(I,J)*VV(J,K)
      DIAC(I,K)=TIP
36  CONTINUE
      DO 34 I=1,L
34  WRITE (6,35) (DIAC(I,J),J=1,L)
35  FORMAT (7F15.5)
19  RETURN
      END
      SUBROUTINE ZTRSQ(YUNK,IL,YDA,YDIF,NUNC,LT,COVI,DISQ)
      DIMENSION YUNK(19,2),YDA(2,2),COVI(2,2),TEMP(19,2),DISQ(2,19),
2  YDIF(19,2)
      DO 404 I=1,NUNC
      DO 404 J=1,LT
      YDIF(I,J)=YUNK(I,J)=YDA(J,IL)
      WRITE (6,504) YDIF(I,J),YUNK(I,J),I,J
504  FORMAT (//,2F20.9,2I10)
404  CONTINUE
      DO 409 I=1,NUNC
      DO 407 II=1,LT
      TEM=0.0
      DO 407 J=1,LT
      TEM=TEM+YDIF(I,J)*COVI(J,II)
      TEMP(I,II)=TEM
407  CONTINUE

```

```

      TEM=0,0
      DO 408 J=1,LT
      TEM=TEM+TEMP(I,J)*YDIF(I,J)
      DISQ(IL,I)=TEM
408  CONTINUE
      WRITE(6,410) DISQ(IL,I),I,IL
410  FORMAT(//,F30,7,2I10)
409  CONTINUE
      RETURN
      END
      SUBROUTINE PROBD (YDIF,J,CQVI,PI,E,DET,S,LT)
      DIMENSION YDIF(19,7),CQVI(7,7)
      F(Y)=(ARCCOS(YDEE/Y)*ABS(Y))*(E**(-0,5*CQ*(Y**2)))/
2DENOM
      FF(X)=(E**(-0,5*CQ*(X**2)))/DENT
      CQ=CQVI(1,1)
      YDEE=YDIF(J,1)
      N=1000
      N1=N/2
      N2=N1-1
      S=0,0
      B=(15,0*SQRT(1,0/CQ))/1000,0
      IF(LT=1) 28,30,29
29  I=2
      DENOM=(2,0*PI)**(FLOAT(I)/2,0)*PI*SQRT(DET)
      S=F(YDIF(J,1))+4,0*F(YDIF(J,1)+B)
      DO 21 I=1,N2
21  S=S+2,0*F(YDIF(J,1)+B*FLOAT(I*2))
      +4,0*F(YDIF(J,1)+B*FLOAT(I*2+1))
      S=ABS(B)*(F(YDIF(J,1)+B*FLOAT(N))+S)/3,0
      GO TO 28
30  DENT=SQRT(2,0*PI/CQ)
      S=FF(YDIF(J,1))+4,0*FF(YDIF(J,1)+B)
      DO 22 I=1,N2
22  S=S+2,0*FF(YDIF(J,1)+B*FLOAT(I*2))+4,0*FF(YDIF(J,1)+B*FLOAT(I*2+1))
      +4,0*FF(YDIF(J,1)+B*FLOAT(N))+S)/3,0
28  RETURN
      END
      SUBROUTINE RAD (VMEIG,L,KT,IK, IIK,XS)
      DIMENSION VMEIG(7),IIK(2)
      XS=(FLOAT(IIK(1)+IIK(2))-0,5*FLOAT(L+KT))*ALOG(1,0+VMEIG(IK))
      LQDF=L+KT-IK*2
      WRITE(6,118)XS,LQDF,IK
118  FORMAT(//,F25,6,2I10)
      RETURN
      END
$ENTRY

```

```

C   CALCULATION OF MISCLASSIFICATION PROBABILITIES USING JOINT
C   DISTRIBUTION FUNCTIONS=FIRST DATA SET
C   ****
DIMENSION DEPRQ(20,2),PROBQ(20),PROBW(20),IP(20,5),PD(5,12),
2  PW(5,12),N(5)
   KT=2
   NUNC=19
   KE=5
   READ(5,25) (N(K),K=1,KE)
25  FORMAT (5I2)
   DO 22 I=1,KE
   II=N(I)
   DO 22 J=1,II
22  READ (5,21) PD(I,J),PW(I,J)
21  FORMAT (2F5,4)
   DO 23 I=1,NUNC
23  READ (5,24) (IP(I,J),J=1,KE)
24  FORMAT (I2,4I1)
   DO 26 I=1,NUNC
   CUM=1.0
   DO 26 J=1,KE
   JJ=IP(I,J)
   CUM=CUM*PD(J,JJ)
   PROBQ(I)=CUM
26  CONTINUE
   DO 27 I=1,NUNC
   CUM=1.0
   DO 27 J=1,KE
   JJ=IP(I,J)
   CUM=CUM*PW(J,JJ)
   PROBW(I)=CUM
27  CONTINUE
   DO 28 I=1,NUNC
   J=1
   TEMP1=PROBQ(I)/(PROBQ(I)+1.0-PROBW(I))
   TEMP2=PROBW(I)/(1.0-PROBQ(I)+PROBW(I))
   DEPRQ(I,J)=TEMP1/(TEMP1+TEMP2)
   J=2
   DEPRQ(I,J)=1.0-DEPRQ(I,1)
   WRITE (6,30) (DEPRQ(I,J),J=1,KT),I,J,PROBQ(I),PROBW(I)
30  FORMAT (//,2F25,9,2I10,2F18,6)
28  CONTINUE
   STOP
   END
$ENTRY

```

```

C      NUMERICAL INTERGRATION OF BIVARIATE NORMAL DISTRIBUTION FUNCTION
C      USING THE HERMITE GAUSS INTERGRAL
C      *****
DIMENSION DEPRQ(20,2)
DIMENSION UNK(19,2),UCQVI(2,2),WCQVI(2,2),UCQV(2,2),WCQV(2,2),
2N(5),X(20,5),W(20,5),NPRQB(19,5),WPRQB(19,5),RHO(2),YDA(2,2)
F(A,B,C)=E**(-A#B#C)
KE=5
NUNC=19
KT=2
E=2.718281828
PI=3.141592653589
DO 11 I=1,NUNC
11 READ(5,10) (UNK(I,J),J=1,KT)
10 FORMAT(2F4,2)
DO 8 JK=1,KT
8 READ(5,9) (YDA(J,JK),J=1,KT)
9 FORMAT(2F8,6)
READ(5,16) (N(K),K=1,KE)
16 FORMAT(5I2)
DO 14 K=1,KE
II=N(K)
DO 14 I=1,II
READ(5,15) X(I,K),W(I,K)
15 FORMAT(F8.6,E12,7)
14 CONTINUE
DO 12 I=1,KT
12 READ(5,13) (UCQVI(I,J),J=1,KT),(WCQVI(I,J),J=1,KT),
2(UCQV(I,J),J=1,KT),(WCQV(I,J),J=1,KT)
13 FORMAT(8F6,4)
DETO=UCQV(1,1)*UCQV(2,2)-UCQV(1,2)**2
CONS0=PI*SQRT(UCQVI(1,1)*UCQVI(2,2)*DETO)
DETW=WCQV(1,1)*WCQV(2,2)-WCQV(1,2)**2
CONSW=PI*SQRT(WCQVI(1,1)*WCQVI(2,2)*DETW)
DO 702 K=1,KE
II=N(K)
DO 702 I=1,II
WRITE(6,703) X(I,K),W(I,K)
703 FORMAT(//,2F30,27)
702 CONTINUE
DO 60 III=1,NUNC
DO 31 K=1,KE
II=N(K)
DO 30 JK=1,KT
IF(JK=1) 20,21,20
21 RHO(JK)=2.0*UCQVI(1,2)/SQRT(UCQVI(1,1)*UCQVI(2,2))
NUX=0
NUY=0
DO 22 I=1,II
J=1
24 XTEST= X(I,K)*SQRT(2.0/UCQVI(J,J))+YDA(J,JK)
IF(UNK(III,J)=XTEST) 70,26,26
26 NUX=NUX+1
70 J=2
23 YTEST= X(I,K)*SQRT(2.0/UCQVI(J,J))+YDA(J,JK)

```

```

IF(UNK(III,J)=YTEST) 22,27,27
27 NU Y=NUY+1
22 CONTINUE
WRITE(6,66)NUX,NUY
66 FORMAT (2I15)
PR=0.0
DO 29 I=1,NUX
DO 29 J=1,NUY
PR=PR+W(I,K)*W(J,K)*F(X(I,K),X(J,K),RHQ(JK))
29 CONTINUE
OPROB(III,K)=PR/CONSO
GO TO 30
20 RHQ(JK)=2.0*WCOVI(1,2)/SQRT(WCOVI(1,1)*WCOVI(2,2))
NUY=0
NUX=0
DO 42 I=1,II
J=1
44 XTEST= X(I,K)*SQRT(2.0/WCOVI(J,J))+YDA(J,JK)
IF(UNK(III,J)=XTEST)71,71,46
46 NUX=NUX+1
71 J=2
42 YTEST= X(I,K)*SQRT(2.0/WCOVI(J,J))+YDA(J,JK)
IF(UNK(III,J)=YTEST)42,42,47
47 NU Y=NUY+1
42 CONTINUE
WRITE(6,67)NUX,NUY
67 FORMAT (2I15)
PR=0.0
NUX=NUX+1
NUY=NUY+1
DO 48 I=NUX,II
DO 48 J=NUY,II
PR=PR+W(I,K)*W(J,K)*F(X(I,K),X(J,K),RHQ(JK))
48 CONTINUE
WPROB(III,K)=PR/CONSW
30 CONTINUE
31 CONTINUE
60 CONTINUE
JK=1
DO 90 III=1,NUNC
DO 90 K=1,KE
90 WRITE(6,92) OPROB(III,K),III,K,JK
50 FORMAT(//,F30,8,3I10)
JK=2
DO 91 III=1,NUNC
DO 91 K=1,KE
91 WRITE(6,50) WPROB(III,K),III,K,JK
32 FORMAT(//,F20,8,3I10)
DO 921 I=1,NUNC
K=5
J=1
TEMP1=OPROB(I,K)/(OPROB(I,K)+1.0+WPROB(I,K))
TEMP2=WPROB(I,K)/(1.0-OPROB(I,K)+WPROB(I,K))
OPROB(I,J)=TEMP1/(TEMP1+TEMP2)
J=2

```

```
DEPRD(I,J)=1,0-DEPRD(I,1)
WRITE(6,922) (DEPRD(I,J),J=1,KT),I
922 FORMAT (//,2F20,9,I10)
921 CONTINUE
STOP
END
$ENTRY
```