THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN EMPIRICAL INVESTIGATION OF THE MULTIVARIATE MULTIPLE SAMPLE LOCATION PROBLEM

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

ALBERT E. (ERNEST) DORR

Oklahoma City, Oklahoma

1970

AN EMPIRICAL INVESTIGATION OF THE MULTIVARIATE

MULTIPLE SAMPLE LOCATION PROBLEM

APPROVED BY

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

I wish to express my gratitude to Dr. Edward N. Brandt, Jr. for suggesting the problem. Special appreciation is extended to Dr. Roy B. Deal, Jr. for his effort and ability in directing both the research and the writing of this dissertation.

A great personal debt is owed to Dr. John F. McCoy, whose constant encouragement and direction during a previous study were largely responsible for the present work. I am grateful to Dr. Paul S. Anderson, Jr., Dr. John C. Brixey, and Dr. Nabih R. Assal who so kindly served on the advisory and reading committees.

Computer time, without which this study would have been impossible, was generously provided by the Hospital Systems Department, University Hospital, University of Oklahoma Medical Center. Thanks are due to the personnel of Hospital Systems for their assistance.

I am grateful to Mrs. Rose Titsworth for her skill and assistance in the preparation of this manuscript.

TABLE OF CONTENTS

		Page
LIST OF TABLES		v
LIST OF ILLUSTRATIONS		vi
Chapter		
I.	INTRODUCTION	1
II.	METHOD OF INVESTIGATION	9
III.	PRELIMINARY INVESTIGATION	18
IV.	THE EXPERIMENTAL DESIGN	37
V.	DISCUSSION AND EVALUATION	44
VI.	SUMMARY	57
LIST OF REFERENCES		60
APPENDIXES		
1.	RANDU	63
2.	COMPUTER TIME REQUIREMENTS	65

LIST OF TABLES

Table		Page
1.	Empirical Frequency Distribution of U for Two Treatment Groups with Fifteen Observations per Group	32
2.	Empirical Frequency Distribution of U for Two Treatment Groups with Thirty Observations per Group	33
3.	Empirical Frequency Distribution of U for Four Treatment Groups with Fifteen Observations per Group	34
4.	Empirical Frequency Distribution of U for Four Treatment Groups with Thirty Observations per Group	35
5.	Population Dispersion and Correlation Matrices for Three Dimensions	39
6.	Population Dispersion and Correlation Matrices for Nine Dimensions	40
7.	Results for Two Treatment Groups with Multivariate Normal Distributions	46
8.	Results for Four Treatment Groups with Multivariate Normal Distributions	50
9.	Results for Two Treatment Groups with Multivariate Non-Normal Distributions	53
10.	Results for Four Treatment Groups with Multivariate Non-Normal Distributions	54

LIST OF ILLUSTRATIONS

Figure Page

1. The Empirical Frequency Distribution of the Sum of the Diagonal Elements of the Contingency Table for Two Treatment Groups with Fifteen Observations per Group..... 27

AN EMPIRICAL INVESTIGATION OF THE MULTIVARIATE MULTIPLE SAMPLE LOCATION PROBLEM

CHAPTER I

INTRODUCTION

The biomedical researcher is often concerned with determining whether several drugs, several methods of therapy, or several such treatments have different effects on living organisms with respect to a specified set of variables. An important scientific advance of this century has been the development of statistical tests which aid the researcher in his investigation of these treatment effects. If the researcher is willing to assume that his variables of interest have a multivariate normal distribution and that his observation vectors are independent, there are standard multivariate tests available. Anderson (1958) gives procedures for testing differences in mean vectors under the assumption of equality of dispersion (variance - covariance) matrices, for testing differences in dispersion matrices under the assumption of equality of mean vectors, and for simultaneously testing differences in either mean vectors or dispersion matrices. If significant differences are found in the third situation, it is not known whether the distributions are different with respect to mean vectors or to dispersion matrices.

These standard methods are somewhat restricted in their utility in an applied situation because of the assumptions required for their

mathematical derivation and a lack of knowledge as to their robustness properties. Ito and Schull (1964) and Holloway and Dunn (1967) have investigated the robustness of the T^2 test when the assumption of equality of the two dispersion matrices is violated. The literature does not seem to contain an investigation of the multivariate robustness problem for the multiple sample case or the non-normal case.

The researcher cannot always safely assume an underlying multivariate normal distribution; the most obvious case being where one or more of the variables are dichotomous, i.e., where the measurement, or observation, is the presence or absence of a given characteristic. Research is presently being directed toward the development of suitable non-parametric multivariate procedures to aid in the elimination of these problems.

Research in the area of non-parametric multivariate statistics has utilized various approaches as well as attacked different problems. Chung and Fraser (1958) have developed a randomization test in order to test the equality of mean vectors in the two-sample problem. Anderson (1966) has investigated the properties of statistically equivalent blocks, which were developed for use as tolerance limits, in order to test whether two distributions are identical and to classify an observation into one of two populations. Bhapkar (1966) has worked with some multivariate analogues of univariate rank-order statistics for his procedure to test the identity of several distributions. Murthy (1966) has used a univariate function $F_n(x)$ which is defined as the product of $1/n$ and the number of observations less than or equal to x among the n observations. This function was generalized and a non-parametric multivariate

procedure for estimating probability density functions was established. Koch and Sen (1968) have utilized ranks and a randomization model to test the equality of treatment effects in the mixed model, or repeated measurements, design. Koch and Sen have drawn heavily from mathematical developments made by Puri and Sen (1966) and Sen and Puri (1967). Dorr (1969) has investigated the use of a clustering technique to test the equality of mean vectors in the multiple sample case.

The following is a detailed discussion of Dorr's method and is presented because the present study is an extension of this work. Clustering is the partitioning of a set of multivariate observations into disjoint subsets, or clusters, where a cluster is a set of the observations which are close together in the multidimensional sample space, i.e., observations in the same cluster are close together in the sample space and observations in different clusters are relatively less close together. A clustering technique given by Ball and Hall (1965) was used to develop a test of the equality of mean vectors.

The first step in the procedure is to cluster the observations. Basically this consists of the determination of cluster centers which are scattered throughout the sample space and the assignment of observations to clusters determined by these cluster centers. The assignment is made by assigning each observation to the closest cluster. The closest cluster is that cluster for which the distance between the observation and the cluster center is a minimum. The distance between an observation and a cluster center is measured by $(X - Y_i)' S_P^{-1}(X - Y_i)$ where X is the p dimensional observation vector, Y_i is the p dimensional vector which is the cluster center for the i^{th} cluster, S_P is the

standard unbiased pooled estimate of the dispersion matrix, and p is the number of dimensions. S_p is computed from the equation:

$$S_p = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})(X_{ij} - \bar{X}_{i.})' \; / \; (\sum_{i=1}^{k} n_i - k)$$

where X_{ij} is the j^{th} p dimensional observation vector in the i^{th} treatment group, $\bar{X}_{i.}$ is the p dimensional sample mean vector for the i^{th} treatment group, n_i is the number of observation vectors in the i^{th} treatment group, and k is the number of treatment groups.

After each observation has been assigned to the closest cluster, each cluster center can be replaced by the mean vector of the observations in that cluster and the procedure repeated. Techniques are available for combining two or more clusters which are relatively close together and for splitting a single cluster into two clusters when the observations are too distant from one another to be considered one cluster.

After the observations have been clustered, a k x m contingency table is constructed where k is the number of treatment groups and m is the number of clusters found. The contingency table is such that the ij^{th} element is the number of observations from the i^{th} treatment group which are in the j^{th} cluster. The sum of the elements in the i^{th} row is the sample size for the i^{th} treatment group, and the sum of the elements in the j^{th} column is the number of observations which are in the j^{th} cluster.

The final step is to test the contingency table. This is done using the χ^2 test, and the null hypothesis of equal mean vectors is rejected at the α level if the sample χ^2 value is significant at the α level.

The rationale underlying the test will now be discussed. The clustering procedure forms a cluster from observations which are close together in the multidimensional sample space. If the assumption of equal dispersion matrices is made, two observations from the same treatment group are no more likely to lie close together than two observations from two different treatment groups which have equal mean vectors since the two observations come from the same distribution in both situations. Thus the assignment of observations to clusters should be independent of the treatment groups from which they come when the mean vectors are equal. On the other hand, two observations from the same treatment group would be expected to be closer together than two observations from two treatment groups with different mean vectors since the distance between the two mean vectors should contribute to the distance between the two observations in the latter case. Thus the assignment of observations to clusters should not be independent of the treatment groups when the treatment groups have different mean vectors. Therefore the contingency table discussed above is constructed and the χ^2 is used to test the independence of treatment groups and clusters. A significant lack of independence of treatment groups and clusters is taken to indicate a significant difference in location of the sampled observations from the treatment groups. It is not unreasonable to think that perhaps the method would also work for a wide variety of applied situations where the dispersion matrices are not equal.

There are difficulties associated with using the clustering procedure to partition the observations. The number and configuration of the clusters formed by the clustering algorithm are dependent upon

the values specified for the input parameters required by the algorithm. Specific values which seem to produce reasonable results for a wide variety of situations have been determined empirically, but it would be desirable to have a partitioning algorithm which is independent of input parameters. Also, the establishment of the original cluster centers is dependent upon the ordering of the observations within treatment groups and the ordering of the treatment groups. The determination of the original cluster centers is a critical step and most of the problems associated with the test procedure seem to result from improperly locating the original cluster centers.

The clustering procedure was developed to investigate data structure and not to test hypotheses. There is information available in the test of hypothesis situation which is not utilized by the clustering algorithm, namely knowledge as to which observations are in which treatment groups and the knowledge that the observations in a treatment group are from the same distribution and may be from the same distribution as the observations from one or more of the other treatment groups. It seems desirable to use this additional information to partition the observations. The obvious approach is to use each treatment group sample mean vector as an original cluster center.

If this approach is taken a simplified method for partitioning can be seen. Associate a cluster with each treatment group sample mean vector in such a manner that the observations in the cluster are those which are closer to that sample mean vector than to any other sample mean vector. The contingency table described above can then be constructed. However, the measure of distance and the test of the contingency table

used previously cannot be assumed to be appropriate. The distance measure is suspect since the distance is no longer a distance between an observation and a cluster center but between an observation and a sample mean vector. An observation is a component of one and only one mean vector, namely the mean vector for the treatment group containing the observation. Then, even under the null hypothesis of equal mean vectors, the observation would be expected to be closer to that mean vector than to any other. Also, the effect of using the sample dispersion matrix is unknown. The test of the contingency table must be modified in that a significant χ^2 is to be taken as a significant difference in mean vectors; but, a difference in mean vectors is not necessarily indicated by lack of independence of treatment groups and clusters. For example, if the significant χ^2 is due to large elements in the off diagonal positions, i.e., the observations in one treatment group are closer to the mean vector for another treatment group, one would certainly not wish to reject the hypothesis of equal mean vectors even though the χ^2 were significant.

This dissertation is concerned with the development of a method to test the multivariate multiple sample null hypothesis of no treatment differences against the alternate hypothesis that the treatment effects are not equal, i.e., H_o: $\mu_1 = \mu_2 = \ldots = \mu_k$, H_1: $\mu_i \neq \mu_j$ for some i and j, where μ_i is the p dimensional mean vector for the population receiving the i^{th} treatment, p is the number of dimensions or the number of variables per observation, and k is the number of treatments under consideration. The first method investigated consists of the construction and test of a k x k contingency table where the i^{th} row corresponds to the i^{th} treatment group, the j^{th} column corresponds to the sample

mean vector for the j^{th} treatment group, and the ij^{th} element is the number of observations in the i^{th} treatment group which are closer to the sample mean vector for the j^{th} treatment group than to any other sample mean vector. Difficulties in the use of the contingency table are encountered and the use of a continuous statistic is investigated.

The Monte Carlo technique is used to determine an appropriate measure of closeness and a test method, and to investigate the effectiveness of the test procedure in detecting differences in mean vectors, as measured by the probability of both Type I and Type II errors. The Monte Carlo study investigates the test effectiveness for different combinations of underlying distributions, sample sizes, numbers of variables, numbers of treatment groups, and positions of mean vectors for both homogeneous and heterogeneous dispersion matrices.

CHAPTER II

METHOD OF INVESTIGATION

The selection of a distance measure between an observation and a sample mean vector is of the utmost importance in the development of the test procedure. The most promising distance measures involve the use of a sample dispersion matrix in a quadratic form. With the commencement of the investigation it was seen that this type of distance measure has extremely complex mathematical properties. The complexity of the measure seems to defy a purely analytical study and it was decided to use empirical techniques to supplement the analytical aspects of the investigation.

The Monte Carlo technique is a scientific approach which utilizes random numbers or samples from known artificial populations. These populations are said to be artificial since they are constructed by the researcher solely for the purposes of the study. Since the sampled populations are known, the samples can be used to compare various techniques or decision functions, or to estimate physical constants or system parameters. The Monte Carlo approach is distinct from the more standard scientific method of utilizing samples from an unknown natural population to infer to that population. The technique has been used successfully in such fields as nuclear physics, numerical analysis, operations research, and statistics.

An extremely large number of random samples is usually necessary to insure adequate precision and the method is not feasible without the aid of a high speed electronic computer. Consequently, various methods have been devised to generate random numbers on high speed computers. Jansson (1966) gives salient characteristics of the methods currently used to generate random numbers.

The Monte Carlo method will be used to evaluate the proposed test with respect to the probability of Type I and Type II errors (α and β levels). The study will be conducted in a manner analogous to a factorial experiment where the factors of interest are distribution, equality of dispersion matrices, sample size, the number of treatment groups, and the separation of mean vectors. For a fixed level of each factor random vectors will be obtained from known populations. These random vectors will be tested at a specified α level using both the proposed test and the standard Multivariate Analysis of Variance (MVAOV) test as given by Cooley and Lohnes (1962).

Upon repeated replication of the procedure, a count is made of the number of times the null hypothesis is rejected using the proposed test and the MVAOV. Since the population mean vectors are known, the α or β level for the proposed test can be estimated, tested equal to a specified value, or tested equal to that of the MVAOV.

Three general cases will be investigated. First is the situation that the assumptions for the MVAOV are met, i.e., the vectors are independently sampled from multivariate normal distributions with a common dispersion matrix. The second case is that the vectors are independently sampled from multivariate normal distributions but the

dispersion matrices for the treatment groups are not all equal, i.e., there is heterogeneity of dispersion matrices. The third case is that the vectors are independently sampled from multivariate non-normal distributions with heterogeneous dispersion matrices. In this last case it is desired to investigate the situation where some of the variables are continuous, but not normal, and some of the variables are discrete.

A computer program has been written by the author in FORTRAN IV for the IBM 1800 which generates and tests random vectors, using both the proposed test and the MVAOV. The remainder of the chapter will be concerned with a description of the generation and characteristics of the random vectors.

Random vectors from a specific multivariate normal distribution are obtained from independent univariate normal samples, which are obtained, in turn, from the uniform distribution. First, 16384 independent samples from the uniform distribution on the interval (0,1), i.e., from U(0,1), were generated. These independent samples from the uniform were obtained using the subroutine RANDU, which is described in Appendix 1. These samples from the uniform were used to obtain 16384 samples from the standard univariate normal. The procedure for obtaining normals from uniforms is one given by Box and Muller (1958) and consists of using the transformation:

$$Y_1 = (-2 \ln X_1)^{1/2} \cos(2\Pi X_2)$$

$$Y_2 = (-2 \ln X_1)^{1/2} \sin(2\Pi X_2)$$

where X_1 and X_2 are independently distributed as U(0,1) and Y_1 and Y_2 are independently distributed as standard normals, i.e., as N(0,1).

The resulting 16384 random numbers were tested to insure a

reasonable approximation to a random sample from N(0,1). The testing was done by using the goodness of fit test as given by Steel and Torrie (1960). The goodness of fit test was applied 163 times; each application involved 100 random numbers and ten classification cells so that each χ^2 had nine degrees of freedom. Of the 163 tests, seven were significant at the five percent level. Thus the tests gave no reason to doubt that the numbers were a random sample from N(0,1).

A method given by Scheuer and Stoller (1962) is used to obtain a p dimensional random vector from $N(\mu,\Sigma)$. As the first step, p of the univariate samples from N(0,1) are considered to constitute the p components of the p dimensional vector Y, and Y is multiplied by the constant matrix C, where C is such that $CC' = \Sigma$. The p dimensional vector μ is added to CY and the resultant vector can be considered to be a random vector from $N(\mu,\Sigma)$. This follows from the fact that if a p dimensional random vector X is distributed $N(\phi,\Delta)$ then $CX + \delta$ is distributed $N(C\phi + \delta, C\Delta C')$ where C is a real non-singular constant p x p matrix and δ is a real p dimensional constant vector. This well known result from mathematical statistics is given as Theorem 3.22 by Graybill (1961). In this situation the dispersion matrix for Y is the p^{th} order identity matrix since the components are independently distributed with unit variance. The mean vector for Y is the p dimensional null vector, and $CY + \mu$ is distributed $N(\mu,\Sigma)$ since $CIC' = CC' = \Sigma$.

The matrix C can be determined uniquely if C is taken to be lower triangular with positive diagonal elements. The elements of C are determined recursively from the following equations:

$$c_{i1} = \frac{\sigma_{i1}}{(\sigma_{11})^{1/2}} \qquad 1 \leq i \leq p$$

$$c_{ii} = (\sigma_{ii} - \sum_{k=1}^{i-1} c_{ik}^2)^{1/2} \qquad 1 < i \leq p$$

$$c_{ij} = \frac{\sigma_{ij} - \sum_{k=1}^{j-1} c_{ik}\, c_{jk}}{c_{jj}} \qquad 1 < j < i \leq p$$

$$c_{ij} = 0 \qquad 1 \leq i < j \leq p$$

where c_{ij} is the ij^{th} element of C and σ_{ij} is the ij^{th} element of Σ.

The existence of C can be determined from the Gram-Schmidt orthogonalization process (Jacobson, 1953). The uniqueness of C can be seen from the equations given above.

Random vectors from a non-normal population are generated in a manner similar to that for the normal vectors. However, the non-normal vectors are obtained from independent observations from a skewed triangular distribution with zero mean and unit variance rather than from normal observations. The skewed triangular distribution was selected because unimodality and skewness are not uncommon for distributions found in applied research situations. The independent samples from the skewed triangular distribution were obtained from the same 16384 samples from U(0,1) that were used to generate the samples from N(0,1). The integral transform as given by Parzen (1960) was used to transform U(0,1) to a skewed triangular distribution. The transformation is given by:

$$Y = (2X)^{1/2} \qquad \text{for } 0 \leq X \leq \frac{2}{9}$$

$$Y = 3 - (7 - 7X)^{1/2} \qquad \text{for } \frac{2}{9} < X \leq 1$$

where X is distributed U(0,1). The probability density function for Y is given by:

$$f(y) = \begin{cases} y & 0 \le y \le \frac{2}{3} \\ -\frac{2}{7}(y-3) & \frac{2}{3} < y \le 3 \\ 0 & \text{elsewhere.} \end{cases}$$

Each sample had the mean subtracted and the result was divided by the standard deviation in order to standardize the distribution.

To obtain a non-normal p dimensional random vector, p of the univariate samples are considered to constitute the p components of the p dimensional vector Y. Y is multiplied by the real lower triangular constant matrix C, and the real p dimensional constant vector δ is added to the result. Y has the p dimensional null vector for its mean and the p^{th} order identity matrix for its dispersion matrix. Since $CY + \delta$ is a linear combination of the components of Y plus a translation, the mean of $CY + \delta$ is δ and the dispersion matrix is $CIC' = CC'$. The reader is referred to CHAPTER IV for a discussion of this property. The distribution of $CY + \delta$ can be derived for a specific C. However, in general this distribution is not a common distribution. Therefore, rather than deriving the distribution of $CY + \delta$, random vectors of this type were generated and the empirical results investigated. For the C's used in this study the empirical marginal distributions were always unimodal and skewed. Correlations between the components ranged from strong negative correlation to strong positive correlation. CHAPTER IV gives the population dispersion and correlation matrices used in the investigation. Random vectors generated in this manner are thought to be realistic since the marginals are unimodal and skewed, the correlations are not

all equal, and the variances are not all equal.

Next, some of the variables are categorized. For the three dimensional case, this is done by assigning a value of zero to all observations on the third dimension which are less than the mean of the treatment group population means for the third dimension and assigning a value of one to all observations on the third dimension which are greater than this mean. In the analysis the zeroes and ones are used as if they were from a continuous distribution.

For the nine dimensional case, the first five dimensions are left continuous. The sixth dimension is categorized into five states and the seventh dimension into three states. The eighth and ninth dimension are categorized into two states. In each case, a continuous observation is replaced by an integer which corresponds to the interval in which the continuous observation fell. The integers are then treated as if they were continuous in the analysis.

As stated above, the 16384 samples from U(0,1) were used to generate 16384 samples from N(0,1) and 16384 samples from the skewed triangular distributions. The investigation requires more than 16384 random numbers but these are all that are available using RANDU. There would probably be no objection to the limited number of random numbers available if one number could be selected at random from the 16384 each time a number is required. However, this is not feasible since it would require some method of obtaining random numbers to decide which random number to use. Therefore, some of the samples must be used sequentially.

The following method is used in an attempt to reduce the consequences of not having a sufficient supply of random numbers. The normal

and triangular random numbers have been placed in separate files on a magnetic disk from which they can be read as needed. For each series of experiments, RANDU is initialized, using a number selected from a random number table, and used to obtain one random number for each experiment in the series of experiments. As used here an experiment consists of obtaining a specified number of sample vectors from a specified number of known populations (not necessarily all different), analyzing the resultant vectors and noting significance or non-significance for both the proposed test and the MVAOV. A series of experiments consists of multiple replication of the experiment with the level of all factors (distribution, equality of dispersion matrices, etc.) held constant. The random number obtained from RANDU for every experiment is used to obtain a starting point for that experiment. The starting point ℓ is the random number modulo 16384 and will be an odd integer on the interval (0, 16384). If the number of p dimensional vectors needed for the experiment is n, the np random numbers are read sequentially, starting with the ℓ^{th} number in the appropriate file (normal or triangular), from the magnetic disk. The first p random numbers constitute the first vector, the second p numbers constitute the second vector, and so on. The first n_1 vectors constitute the sample for the first treatment group, where n_i is the number of observations for the i^{th} treatment group. The next n_2 vectors constitute the second treatment group, and so on through the k treatment groups. At this stage the vectors all come from the same population and must be transformed so that the treatment groups have the desired population mean vectors and dispersion matrices. This is accomplished by multiplying each vector in the i^{th} treatment group by C_i and adding μ_i

to the result. C_i is such that $C_i C_i' = \Sigma_i$, the desired population dispersion matrix for the i^{th} treatment group. Also, μ_i is the desired population mean vector for the i^{th} treatment group.

The method explained assumes that when the np random numbers used in one experiment overlap the np random numbers used in another experiment, the overlap is such that the random numbers go into different components and different treatment groups often enough in order not to invalidate the results. Since the number of dimensions, the number of vectors per treatment group, and the number of treatment groups are all being changed, the assumption is thought to be reasonable.

CHAPTER III

PRELIMINARY INVESTIGATION

A standard measure of the distance between two multivariate normal populations is the generalized distance, Δ^2, which was first proposed by Mahalanobis (1930). Δ^2 is defined by:

$$\Delta^2 = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$$

where μ_1 and μ_2 are the p dimensional mean vectors, Σ is the p x p dispersion matrix (assumed common to the two populations), and p is the number of dimensions. Reyment (1962) has utilized the measure Δ_R^2, given by:

$$\Delta_R^2 = 2(\mu_1 - \mu_2)'(\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)$$

where Σ_1 and Σ_2 are the two dispersion matrices and are not assumed to be equal. It is easily seen that $\Delta^2 = \Delta_R^2$ when $\Sigma_1 = \Sigma_2$. It will be assumed in the sequel that Δ^2 and Δ_R^2 are meaningful measures of the distance between two multivariate populations even when the underlying distributions are non-normal. Under this assumption, it would be desirable for the expected value of the measure for the distance between an observation and a sample mean vector to increase as Δ^2 increases. Here, Δ^2 is the generalized distance between the population from which the one observation was selected and the population from which the observations contained in the sample mean vector were drawn. If the observation is in the i^{th} treatment group, Δ^2 is zero when the sample mean vector

under consideration is for the i^{th} treatment group and also zero when $\mu_i = \mu_j$, j=1, 2, ..., k, where the sample mean vector is for the j^{th} treatment group.

The intuitive approach is to use $(X_{ij} - \bar{X}_{m.})' S_P^{-1}(X_{ij} - \bar{X}_{m.})$ as the measure of distance between the j^{th} observation in the i^{th} treatment group and the sample mean vector for the m^{th} treatment group. X_{ij} is a p dimensional observation vector, $\bar{X}_{m.}$ is a p dimensional sample mean vector, and S_P is the standard unbiased pooled sample dispersion matrix given in CHAPTER I. The primary objection to this measure is that if i = m then X_{ij} has been used to compute $\bar{X}_{m.}$. This can be circumvented by removing X_{ij} from $\bar{X}_{i.}$, i.e., computing the mean vector for the i^{th} treatment group excluding X_{ij}. Let this new mean vector be designated $\bar{X}_{i.}^{ij}$ where the superscript designates the observation that has been deleted. Also, let $\bar{X}_{m.}^{ij}$ be the standard mean vector for the m^{th} treatment group when $i \neq m$, i.e., $\bar{X}_{m.}^{ij} = \bar{X}_{m.}$ when $i \neq m$. It is pointed out for computational convenience that by simple algebraic manipulation $(X_{ij} - \bar{X}_{i.}^{ij}) = n_i(X_{ij} - \bar{X}_{i.}) / (n_i - 1)$ where n_i is the sample size for the i^{th} treatment group. Then, the distance between the ij^{th} observation and the m^{th} sample mean vector can be taken to be D_{ijm}^2, where D_{ijm}^2 is given by:

$$D_{ijm}^2 = (X_{ij} - \bar{X}_{m.}^{ij})' S_P^{-1}(X_{ij} - \bar{X}_{m.}^{ij}) .$$

This measure is similar to the studentized D^2 statistic given by:

$$D^2 = (\bar{X}_{1.} - \bar{X}_{2.})' S_P^{-1}(\bar{X}_{1.} - \bar{X}_{2.}).$$

D^2 is used to estimate Δ^2 and its distribution has been derived by Bose and Roy (1938) who have assumed that the underlying distributions are

multivariate normal. The expected value of D^2 is given by Defrisse-Gussenhoven (1966) as

$$E(D^2) = \frac{n_1 + n_2 - 2}{n_1 + n_2 - p - 3} \left[\Delta^2 + \frac{p}{n_1} + \frac{p}{n_2} \right]$$

where n_1 and n_2 are the sample sizes for the two treatment groups.

The expected value of D^2 is of little help in ascertaining whether $E(D^2_{ijm}) = E(D^2_{iji})$ when $i \neq m$ but $\mu_i = \mu_m$. One problem is that D^2 is defined only when there are two treatments and another is that of determining whether an n is associated with the sample size used to estimate a mean vector or is associated with the estimation of Σ. For D^2, $\bar{X}_{1.}$ contains n_1 observations, $\bar{X}_{2.}$ contains n_2 observations, and the divisor for the sums of squares matrix is $n_1 + n_2 - 2$. However, for D^2_{ijm} and two treatment groups, X_{ij} has one observation and there are no observations available with which to compute deviations about this "mean vector". The derivation of the distribution of D^2_{ijm} would be extremely complex and is beyond the scope of the present study.

Mahalanobis (1936) uses a similar measure, D^2_M, given by:

$$D^2_M = (\bar{X}_{1.} - \bar{X}_{2.})' \Sigma^{-1} (\bar{X}_{1.} - \bar{X}_{2.}),$$

where it is assumed that there is a common dispersion matrix for the two populations. Assuming multivariate normal distributions, the expected value of D^2_M is:

$$E(D^2_M) = \Delta^2 + \frac{p}{n_1} + \frac{p}{n_2} .$$

If the S_p in D^2_{ijm} is replaced by Σ the result is a special case of D^2_M with n_1 equal to one since $\bar{X}_{1.}$ is the observation. In this special case the expected values are

$$\Delta^2 + p + \frac{p}{n_i - 1} \qquad \text{when } i = m$$

and

$$\Delta^2 + p + \frac{p}{n_m} \qquad \text{when } i \neq m.$$

This measure can be adjusted for the bias and the result has an expected value of Δ^2, both when $i = m$ and when $i \neq m$. Thus the measure has the desired characteristics.

In spite of having the desired characteristics, the measure is impractical since the researcher seldom knows Σ in an applied situation. For this reason S_p is first used instead of Σ. The measure to be investigated is $\tilde{D}^2_{ijm}$, given by:

$$\tilde{D}^2_{ijm} = (X_{ij} - \bar{X}^{ij}_{i.})' S_p^{-1} (X_{ij} - \bar{X}^{ij}_{i.}) - p - \frac{p}{n_i - 1} \qquad \text{when } i = m$$

and

$$\tilde{D}^2_{ijm} = (X_{ij} - \bar{X}_{m.})' S_p^{-1} (X_{ij} - \bar{X}_{m.}) - p - \frac{p}{n_m} \qquad \text{when } i \neq m.$$

It is pointed out that the use of $\tilde{D}^2_{ijm}$ is not defended mathematically since the expected value was derived assuming normality and knowledge of Σ and $\tilde{D}^2_{ijm}$ is to be used when neither of these assumptions applies. However, the invariance of expected values under changes in underlying distributions is not uncommon, examples being the expected values of the standard sample mean and variance. So the $\tilde{D}^2_{ijm}$ appears to be a reasonable measure to investigate.

$\tilde{D}^2_{ijm}$ can be used to partition the observations into k sets where the i^{th} set contains the observations closer to the sample mean vector for the i^{th} treatment group than to any other mean vector. Then the k x k contingency table given in CHAPTER I can be constructed. It is

assumed for the time being that $E(\tilde{D}^2_{ijm})$ does equal $E(\tilde{D}^2_{iji})$ when $\mu_i = \mu_m$. Then a reasonable approximation is that the probability that an observation is closest to the i^{th} sample mean vector, i = 1, 2, ..., k, is 1/k under the hypothesis of equal mean vectors since the expected value of the distance, adjusted for bias, to each sample mean vector is zero. Also reasonable is the assumption that the probability is greater than 1/k that an observation from the i^{th} treatment group is closer to the sample mean vector for the i^{th} treatment group when $\mu_i \neq \mu_m$ for some m. This follows from the fact that $E(\tilde{D}^2_{iji})$ is zero but $E(\tilde{D}^2_{ijm})$ is positive.

Under the preceding assumptions, the expected value for the sum of the diagonal elements in the contingency table is $(1/k) \sum_{i=1}^{k} n_i$ under the null hypothesis, since the i^{th} diagonal element is the number of observations from the i^{th} treatment group which are closer to the i^{th} sample mean vector than to any other sample mean vector. Also, the expected value for the sum of the diagonal elements is greater than $(1/k) \sum_{i=1}^{k} n_i$ when $\mu_i \neq \mu_j$ for one or more of the i and j. Although the expected value for the sum of diagonal elements is known, its distribution, which is required for a test of hypothesis, is not known unless another assumption is made. This assumption is that the distance measurements are independent. This assumption is perhaps the least tenable made thus far since it is known that the distances are not independent; it is only a question of whether the lack of independence is serious enough to invalidate the test procedure. However, proceeding under this assumption, the equality of mean vectors can be tested indirectly by testing whether the sum of the diagonal elements in the contingency table is greater than would be expected if the probability of an observation's

being on the diagonal is 1/k. The test is made using the binomial distribution. The null hypothesis of equal mean vectors is rejected with the rejection of the hypothesis that the probability is less than or equal to 1/k that a given observation appears on the diagonal.

In order to ascertain the validity of the given test, an empirical study was conducted. This study consisted of sampling multivariate normal vectors from known populations, using the techniques given in CHAPTER II, and testing the resultant vectors. At this stage primary interest was in the significance level, not the power level. For this reason, the population mean vectors for the treatment groups were equal. The population dispersion matrices for the treatment groups were those used in the main study and are given in CHAPTER IV. All tests were performed at the five percent significance level, and there were 75 independent experiments for each series of experiments, where a series of experiments is for a specific combination of the levels of the important factors.

This empirical study revealed that there was a bias in the method in spite of the correction factors employed in $\tilde{D}^2_{ijm}$. This bias was such that an observation had probability greater than 1/k of being on the diagonal of the contingency table, i.e., of being closer to the sample mean vector for the treatment group which it was in than to any other mean vector. This bias was made obvious by the fact that the total number of observations on the diagonal was always significantly greater than $(1/k) \cdot 75 \cdot N$ for a series of experiments. The value $(1/k) \cdot 75 \cdot N$ is the number of observations expected on the diagonal if there is no bias, where N is the number of observations required for

one experiment, i.e., $\sum_{i=1}^{k} n_i$, and 75 is the number of experiments in a series of experiments. In addition to the bias evident in each series of experiments, there was an increase in the observed error rate with an increase in the number of dimensions. For nine dimensions, the true error rate appeared to be close to 33 percent rather than the stated error rate of five percent.

$\tilde{D}^2_{ijm}$ is a special case of D^2_M if Σ is used instead of S_p. In order to determine the source of the problem, the sampled observation vectors were analyzed using Σ instead of S_p. In this situation, the error rate exceeded five percent but the procedure did not appear to be biased and the error rate did not increase with an increase in the number of dimensions.

Allais (1964) recommends that the number of observations used to estimate a dispersion matrix be greater than ten times the number of dimensions. It was thought that perhaps the trouble lay in an inadequate estimation of the dispersion matrices. The data were analyzed again using S_D instead of S_p, where S_D is S_p with all off diagonal elements set equal to zero. Thus, the measure utilized makes no use of the correlations. Also, the sample size required for adequate estimation is reduced since the covariances are not estimated. The results showed no bias, but the results of further investigation showed that the method was only about ten percent as powerful as the MVAOV when population dispersion matrices were not diagonal. Thus, a more efficient correction was sought.

Another possible cause of the problem was the manner in which S_p was computed and used. The deviation vector $(X_{ij} - \bar{X}_{i.})$ is a

component of S_p but $(X_{ij} - \bar{X}_{m.})$ is not a component of S_p when $i \neq m$. So there is a relationship between $(X_{ij} - \bar{X}_{i.})$ and S_p in the measure $\tilde{D}^2_{iji}$ which does not exist between $(X_{ij} - \bar{X}_{m.})$ and S_p. This difference in relationships does not exist if the total dispersion matrix, S_T, is used. S_T is computed from:

$$S_T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})(X_{ij} - \bar{X}_{..})' \; / \; (\sum_{i=1}^{k} n_i - 1)$$

where $\bar{X}_{..}$ is the mean vector for all observations. Under the null hypothesis of equal mean vectors, S_p and S_T are both unbiased estimates of Σ, so the use of S_T is not unreasonable.

The data were again analyzed, this time using S_T instead of S_p. The results indicated no bias and further work supported this finding. However, the error rate still exceeded five percent. The error rate did not increase with the number of dimensions. Since the use of S_T gave results quite similar to the results obtained when using Σ, the utilization of S_T was retained. The increased error rate was attributed to the lack of independence among the distance measurements. This lack of independence makes it difficult to use standard statistical tests since they usually require independence among the observations. Since standard statistical tests were of no help there were two alternatives: analytical or empirical determination of the distribution function for the number of observations on the diagonal of the contingency table. The latter approach was selected.

It was thought that the method would be feasible if the distribution function for the number of observations on the diagonal was invariant under changes in distribution, dispersion matrices, and the

number of dimensions even if it were not invariant under changes in the number of treatment groups, and sample sizes. Therefore, for each of several different numbers of treatment groups and sample sizes 2000 experiments were generated using procedures given in CHAPTER II. The treatment groups' population mean vectors and dispersion matrices were equal and the number of dimensions was one. In each case, the empirical frequency distribution for the number of observations on the diagonal was tabulated for the 2000 sample values. Then a five percent rejection region was constructed by assuming that the frequency distribution was a probability mass function.

Figure 1 gives the empirical frequency distribution of the number of observations on the diagonal for two treatment groups with fifteen observations per group. Empirical distributions obtained for two treatment groups with thirty observations per group, four treatment groups with fifteen observations per group, and four treatment groups with thirty observations per group were quite similar to Figure 1 and are not presented.

The use of the binomial test calls for rejection of the null hypothesis at the five percent level if there are twenty or more observations on the diagonal, but Figure 1 can be used to ascertain that the probability of twenty or more is approximately .10 when the null hypothesis is true. Thus it can be seen why an excessive error rate was encountered when the binomial test was used.

Next, series of experiments were conducted where the number of dimensions was greater than one and the dispersion matrices were not necessarily equal. The results indicated that the method was

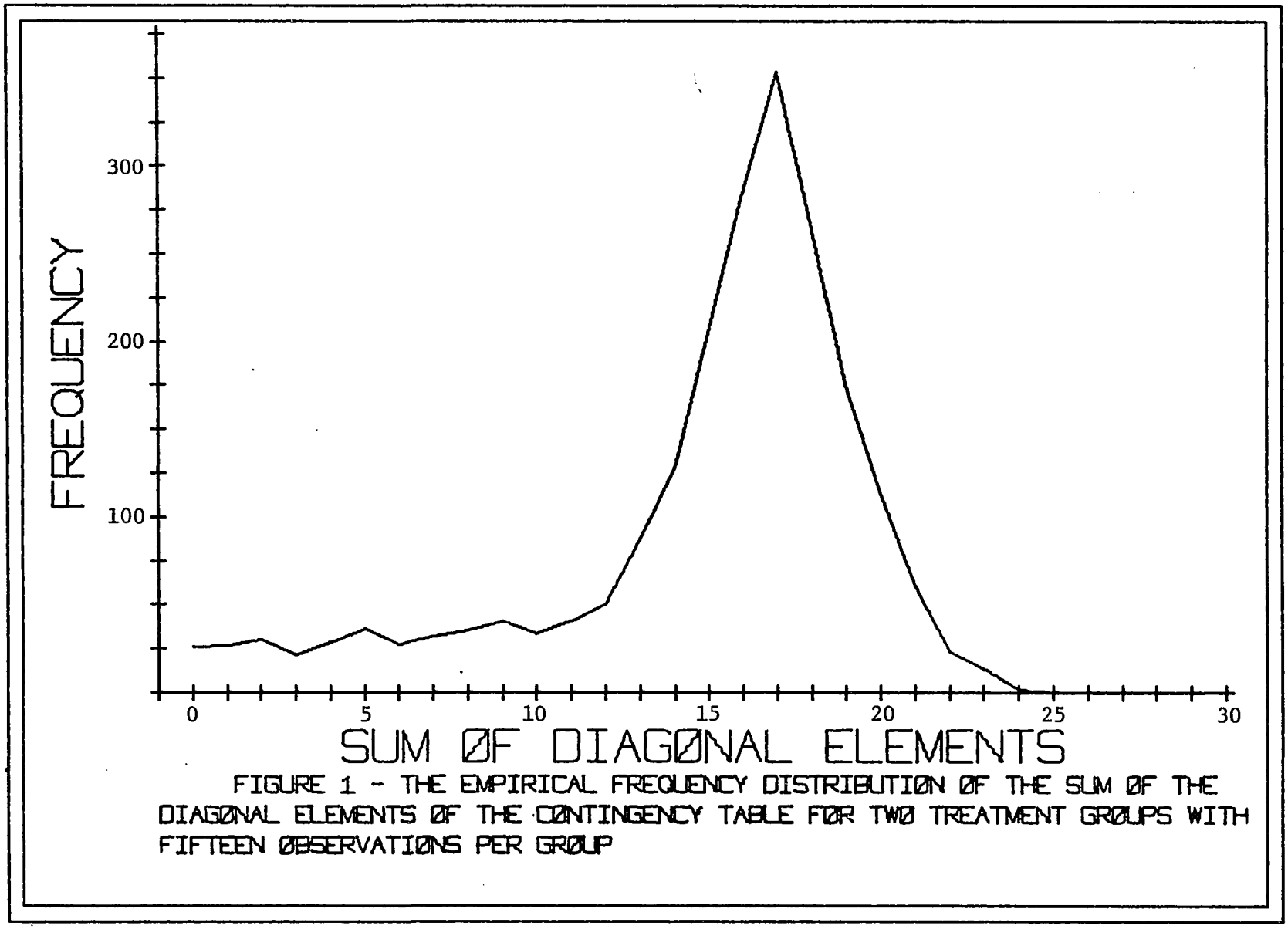

FIGURE 1 - THE EMPIRICAL FREQUENCY DISTRIBUTIØN ØF THE SUM ØF THE DIAGØNAL ELEMENTS ØF THE CØNTINGENCY TABLE FØR TWØ TREATMENT GRØUPS WITH FIFTEEN ØBSERVATIØNS PER GRØUP

satisfactory for two treatment groups but not always satisfactory for four treatment groups. The true α level appeared to be about five percent when the dispersion matrices were equal but to appreciably exceed five percent when the dispersion matrices were not equal for four groups. A possible explanation for the procedure's working for two groups but not for four is the manner in which the empirical distribution was obtained. The empirical distribution was for one dimensional observations. In a one dimensional space with four sample means, there is an unbounded region for both the smallest and largest means where an observation can be closest to the smallest or largest mean. The regions where an observation will be closest to one of the two intermediate means are both bounded. However, for two treatment groups both sample means have unbounded regions where an observation can be closer. For all dimensions greater than one, the regions are all unbounded for all numbers of treatment groups. Hence, for four treatment groups and one dimension the distribution for the number of diagonal elements could be different from the distribution for higher dimensions. This could result in a rejection of the null hypothesis more often than the stated α level even when the null hypothesis was true. This explanation does not take into account the fact that the method seemed to work when the dispersion matrices were equal. Nevertheless, a procedure was desired which would eliminate the bounded regions.

The solution selected involved the computation of only two distances for each observation no matter how many treatment groups were present. The distances associated with each observation are the distance to the sample mean vector of the treatment group which the

observation is in and the distance to the sample mean vector for all observations not in the same treatment group. The two distances to be associated with X_{ij} are given by:

$$\hat{D}^2_{1ij} = (X_{ij} - \bar{X}^{ij}_{i.})' S_T^{-1}(X_{ij} - \bar{X}^{ij}_{i.}) - p - \frac{p}{n_i - 1}$$

and

$$\hat{D}^2_{2ij} = (X_{ij} - \bar{X}^{i}_{..})' S_T^{-1}(X_{ij} - \bar{X}^{i}_{..}) - p - \frac{p}{\sum_{\substack{m=1 \\ m\neq i}}^{k} n_m}$$

where $\bar{X}^{i}_{..}$ is the sample mean vector for all observations not in the i^{th} treatment group. Next, a function ϕ is defined on the observation vectors such that

$$\phi(X_{ij}) = 1 \quad \text{when } \hat{D}^2_{2ij} - \hat{D}^2_{1ij} \geq 0$$

$$\phi(X_{ij}) = 0 \quad \text{when } \hat{D}^2_{2ij} - \hat{D}^2_{1ij} < 0.$$

Also T is defined to be the sum of the $\phi(X_{ij})$ over all observations, i.e.,

$$T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \phi(X_{ij}).$$

Then T can be used to test the null hypothesis of equal mean vectors once its distribution is known. Since T is identical to the previous test statistic studied when there are two treatment groups, its empirical frequency distribution had already been obtained. An empirical distribution of T for one dimensional observations and four treatment groups was generated and the test procedure was studied. The results indicated that the true α level was less than five percent; this was more noticeable when there were four groups. It appeared that the

distribution for T was different in one dimension than in the higher dimensions, so a standardization was determined.

If d_{ij} is taken to be the difference in the two distances for an observation, i.e., $d_{ij} = \hat{D}^2_{2ij} - \hat{D}^2_{1ij}$, then the expected value of d_{ij} should be zero under the null hypothesis and positive under the alternate hypothesis. At least the expected value of d_{ij} should increase with an increase in the separation of the two population mean vectors, even if there is a bias such that the expected value under the null hypothesis is not zero. The standard univariate test for this situation is Student's t; however, it cannot be assumed that the d_{ij} are normally or independently distributed. Also there is no assurance that the expected value of the mean difference is zero under the null hypothesis. A statistic U can be computed from the d_{ij} as Student's t and the distribution of U can be determined empirically. If

$$\bar{d} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} d_{ij}}{\sum_{i=1}^{k} n_i}$$

and

$$s_{\bar{d}} = \left[\sum_{i=1}^{k} \sum_{j=1}^{n_i} (d_{ij} - \bar{d})^2 \Big/ \left[\left(\sum_{i=1}^{k} n_i - 1 \right) \left(\sum_{i=1}^{k} n_i \right) \right] \right]^{1/2}, \text{ then}$$

$$U = \frac{\bar{d}}{s_{\bar{d}}} .$$

Four empirical frequency distributions of U were obtained where the null hypothesis of no difference in population mean vectors was known to hold. This was done by repeated sampling of known populations

using the techniques given in CHAPTER II and computing the U value for each sample. Each frequency distribution was obtained where the observation vectors were univariate and normally distributed. There were two thousand values of U computed for each frequency distribution.

The four empirical frequency distributions were for two treatment groups with fifteen observations per group, two treatment groups with thirty observations per group, four treatment groups with fifteen observations per group, and four treatment groups with thirty observations per group. These frequency distributions are given as Tables 1, 2, 3, and 4 respectively.

A one-tailed rejection region for the test is used because a negative $\bar{d}$ implies that an observation tends to lie closer to the sample mean vector for the other treatment groups than to the sample mean vector for its own treatment group and this cannot be attributed to differences in population mean vectors. However, a large positive $\bar{d}$ can be attributed to differences in population mean vectors.

In order to ascertain a five percent rejection region the critical value was determined such that five percent, or 100, of the observed U's were greater than the critical value. The critical value was never on the boundary of an interval so linear interpolation was used. This resulted in respective critical values of 1.48, 1.55, 1.22, and 1.18 for the four distributions. The use of U as a statistic was extensively investigated so a summary of the test procedure is given and the results of the investigation are presented in CHAPTER V.

The following notation is used: X_{ij} is the j^{th} observation vector in the i^{th} group, $\bar{X}_{i.}^{ij}$ is the mean vector for the i^{th} group

TABLE 1

EMPIRICAL FREQUENCY DISTRIBUTION OF U FOR TWO TREATMENT GROUPS WITH FIFTEEN OBSERVATIONS PER GROUP

Interval		Frequency	Interval		Frequency
-6.00	-5.87	1	-1.23	-1.10	43
-5.87	-5.75	0	-1.10	-0.98	42
-5.75	-5.62	0	-0.98	-0.85	59
-5.62	-5.50	0	-0.85	-0.73	55
-5.50	-5.37	2	-0.73	-0.60	57
-5.37	-5.25	3	-0.60	-0.48	48
-5.25	-5.12	1	-0.48	-0.35	60
-5.12	-5.00	5	-0.35	-0.22	64
-5.00	-4.87	4	-0.22	-0.10	60
-4.87	-4.74	5	-0.10	0.02	53
-4.74	-4.62	4	0.02	0.14	66
-4.62	-4.49	4	0.14	0.27	60
-4.49	-4.37	13	0.27	0.39	58
-4.37	-4.24	21	0.39	0.52	54
-4.24	-4.12	21	0.52	0.64	46
-4.12	-3.99	15	0.64	0.77	55
-3.99	-3.87	30	0.77	0.90	46
-3.87	-3.74	24	0.90	1.02	42
-3.74	-3.61	19	1.02	1.15	42
-3.61	-3.49	29	1.15	1.27	38
-3.49	-3.36	30	1.27	1.40	20
-3.36	-3.24	22	1.40	1.52	27
-3.24	-3.11	19	1.52	1.65	18
-3.11	-2.99	34	1.65	1.77	12
-2.99	-2.86	28	1.77	1.90	15
-2.86	-2.74	25	1.90	2.03	9
-2.74	-2.61	23	2.03	2.15	6
-2.61	-2.48	27	2.15	2.28	3
-2.48	-2.36	36	2.28	2.40	12
-2.36	-2.23	31	2.40	2.53	3
-2.23	-2.11	41	2.53	2.65	4
-2.11	-1.98	34	2.65	2.78	4
-1.98	-1.86	41	2.78	2.90	7
-1.86	-1.73	46	2.90	3.03	2
-1.73	-1.61	43	3.03	3.16	2
-1.61	-1.48	29	3.16	3.28	1
-1.48	-1.35	50	3.28	3.41	0
-1.35	-1.23	46	3.41	3.53	1
Critical value is 1.48					

TABLE 2

EMPIRICAL FREQUENCY DISTRIBUTION OF U FOR TWO TREATMENT GROUPS WITH THIRTY OBSERVATIONS PER GROUP

Interval		Frequency	Interval		Frequency
-7.24	-7.10	2	-1.94	-1.80	35
-7.10	-6.96	1	-1.80	-1.66	37
-6.96	-6.82	0	-1.66	-1.52	37
-6.82	-6.68	5	-1.52	-1.38	39
-6.68	-6.54	2	-1.38	-1.24	38
-6.54	-6.40	3	-1.24	-1.10	60
-6.40	-6.26	3	-1.10	-0.97	55
-6.26	-6.12	5	-0.97	-0.83	48
-6.12	-5.98	9	-0.83	-0.69	64
-5.98	-5.84	7	-0.69	-0.55	51
-5.84	-5.70	9	-0.55	-0.41	51
-5.70	-5.56	15	-0.41	-0.27	68
-5.56	-5.42	14	-0.27	-0.13	69
-5.42	-5.29	12	-0.13	0.00	80
-5.29	-5.15	14	0.00	0.14	80
-5.15	-5.01	24	0.14	0.28	55
-5.01	-4.87	15	0.28	0.42	67
-4.87	-4.73	15	0.42	0.56	65
-4.73	-4.59	13	0.56	0.70	50
-4.59	-4.45	21	0.70	0.84	65
-4.45	-4.31	10	0.84	0.98	46
-4.31	-4.17	9	0.98	1.12	43
-4.17	-4.03	13	1.12	1.25	34
-4.03	-3.89	19	1.25	1.39	29
-3.89	-3.75	23	1.39	1.53	27
-3.75	-3.61	20	1.53	1.67	25
-3.61	-3.47	29	1.67	1.81	28
-3.47	-3.33	26	1.81	1.95	19
-3.33	-3.19	17	1.95	2.09	9
-3.19	-3.06	23	2.09	2.23	7
-3.06	-2.92	29	2.23	2.37	4
-2.92	-2.78	23	2.37	2.51	4
-2.78	-2.64	31	2.51	2.65	3
-2.64	-2.50	32	2.65	2.79	1
-2.50	-2.36	30	2.79	2.93	1
-2.36	-2.22	22	2.93	3.07	0
-2.22	-2.08	25	3.07	3.21	1
-2.08	-1.94	34	3.21	3.35	1
Critical value is 1.55					

TABLE 3

EMPIRICAL FREQUENCY DISTRIBUTION OF U FOR FOUR TREATMENT GROUPS WITH FIFTEEN OBSERVATIONS PER GROUP

Interval		Frequency	Interval		Frequency
-4.05	-3.97	2	-0.98	-0.90	55
-3.97	-3.89	1	-0.90	-0.82	60
-3.89	-3.81	0	-0.82	-0.74	43
-3.81	-3.73	2	-0.74	-0.66	55
-3.73	-3.65	2	-0.66	-0.58	37
-3.65	-3.57	2	-0.58	-0.50	60
-3.57	-3.49	0	-0.50	-0.42	59
-3.49	-3.41	5	-0.42	-0.34	64
-3.41	-3.33	2	-0.34	-0.26	62
-3.33	-3.24	3	-0.26	-0.18	64
-3.24	-3.16	4	-0.18	-0.10	55
-3.16	-3.08	3	-0.10	-0.01	57
-3.08	-3.00	7	-0.01	0.06	67
-3.00	-2.92	1	0.06	0.14	64
-2.92	-2.84	5	0.14	0.22	63
-2.84	-2.76	5	0.22	0.30	63
-2.76	-2.68	7	0.30	0.38	53
-2.68	-2.60	10	0.38	0.46	59
-2.60	-2.52	5	0.46	0.54	39
-2.52	-2.44	12	0.54	0.62	55
-2.44	-2.36	9	0.62	0.70	51
-2.36	-2.28	10	0.70	0.78	52
-2.28	-2.19	19	0.78	0.86	46
-2.19	-2.11	11	0.86	0.94	38
-2.11	-2.03	16	0.94	1.03	25
-2.03	-1.95	18	1.03	1.11	25
-1.95	-1.87	21	1.11	1.19	24
-1.87	-1.79	19	1.19	1.27	26
-1.79	-1.71	26	1.27	1.35	15
-1.71	-1.63	12	1.35	1.43	19
-1.63	-1.55	24	1.43	1.51	16
-1.55	-1.47	27	1.51	1.59	13
-1.47	-1.39	22	1.59	1.67	4
-1.39	-1.31	30	1.67	1.75	3
-1.31	-1.23	35	1.75	1.83	2
-1.23	-1.149	35	1.83	1.91	6
-1.14	-1.06	41	1.91	1.99	5
-1.06	-0.98	42	1.99	2.08	1
Critical value is 1.22					

TABLE 4

EMPIRICAL FREQUENCY DISTRIBUTION OF U FOR FOUR TREATMENT GROUPS WITH THIRTY OBSERVATIONS PER GROUP

Interval		Frequency	Interval		Frequency
-6.06	-5.95	1	-1.95	-1.84	18
-5.95	-5.85	0	-1.84	-1.73	26
-5.85	-5.74	0	-1.73	-1.62	24
-5.74	-5.63	0	-1.62	-1.51	32
-5.63	-5.52	0	-1.51	-1.41	38
-5.52	-5.41	0	-1.41	-1.30	39
-5.41	-5.30	0	-1.30	-1.19	56
-5.30	-5.20	0	-1.19	-1.08	49
-5.20	-5.09	0	-1.08	-0.97	52
-5.09	-4.98	0	-0.97	-0.86	68
-4.98	-4.87	0	-0.86	-0.76	49
-4.87	-4.76	0	-0.76	-0.65	77
-4.76	-4.65	0	-0.65	-0.54	61
-4.65	-4.55	1	-0.54	-0.43	68
-4.55	-4.44	1	-0.43	-0.32	76
-4.44	-4.33	3	-0.32	-0.21	90
-4.33	-4.22	1	-0.21	-0.11	64
-4.22	-4.11	1	-0.11	-0.00	77
-4.11	-4.00	3	-0.00	0.10	85
-4.00	-3.90	2	0.10	0.21	89
-3.90	-3.79	3	0.21	0.32	89
-3.79	-3.68	2	0.32	0.42	92
-3.68	-3.57	3	0.42	0.53	74
-3.57	-3.46	7	0.53	0.64	74
-3.46	-3.35	1	0.64	0.75	64
-3.35	-3.25	11	0.75	0.86	59
-3.25	-3.14	13	0.86	0.97	51
-3.14	-3.03	2	0.97	1.07	58
-3.03	-2.92	4	1.07	1.18	39
-2.92	-2.81	9	1.18	1.29	28
-2.81	-2.71	12	1.29	1.40	23
-2.71	-2.60	11	1.40	1.51	19
-2.60	-2.49	9	1.51	1.62	10
-2.49	-2.38	8	1.62	1.72	7
-2.38	-2.27	11	1.72	1.83	6
-2.27	-2.16	13	1.83	1.94	0
-2.16	-2.06	10	1.94	2.05	2
-2.06	-1.95	24	2.05	2.16	1
Critical value is 1.18					

excluding X_{ij}, $\bar{X}^{i}_{..}$ is the mean vector for all observations not in the i^{th} group, S_T is the sample total dispersion matrix, p is the number of dimensions, n_i is the number of observations in the i^{th} treatment group, and k is the number of treatment groups. First the two distances associated with X_{ij} are computed for all observations. These two distances are computed from:

$$\hat{D}^2_{1ij} = (X_{ij} - \bar{X}^{ij}_{i.})' S_T^{-1}(X_{ij} - \bar{X}^{ij}_{i.}) - p - \frac{p}{n_i - 1} \quad \text{and}$$

$$\hat{D}^2_{2ij} = (X_{ij} - \bar{X}^{i}_{..})' S_T^{-1}(X_{ij} - \bar{X}^{i}_{..}) - p - \frac{p}{k} \quad .$$

$$\left(\sum_{\substack{m=1 \\ m\neq i}} n_m\right)$$

Next the difference, d_{ij}, is computed for each observation where $d_{ij} = \hat{D}^2_{2ij} - \hat{D}^2_{1ij}$. Then the mean difference is divided by the sample standard error of the difference and the result is the test statistic, i.e., $U = \bar{d}/s_{\bar{d}}$. U is compared to the critical value given in the appropriate table. If U is greater than the critical value the null hypothesis of equal mean vectors is rejected at the five percent level.

CHAPTER IV

THE EXPERIMENTAL DESIGN

A sampling study was utilized to test the effectiveness of the test procedure presented in CHAPTER III. The sampling study consisted of the generation of random vectors from known multivariate populations and analysis of the results using both the proposed test and the MVAOV. Different combinations of the levels of the important factors were investigated. These important factors are distribution, equality of dispersion matrices, number of treatment groups, sample size, and separation of mean vectors. The present chapter is concerned with the experimental design of the sampling study.

Underlying distributions were either normal or non-normal. Salient characteristics of the non-normal distributions used were given in CHAPTER II. Basically, the non-normal distributions are characterized by a marked skewness for their continuous variables and having at least one discrete variable.

Desiderata for population dispersion matrices were generality and applicability to applied situations. Two general cases were investigated: equal and not equal dispersion matrices. When the population dispersion matrices were equal, the common matrix was the p^{th} order identity matrix. However, the inference is not restricted to the identity matrix. This is due to a certain invariance property of the test

procedure which will now be shown. Let X be a continuous p dimensional random variable with mean equal to the null vector and dispersion matrix equal to the p^{th} order identity matrix. Then if we let Y = CX, where C is a real p x p lower triangular constant matrix, it is known from mathematical statistics that Y also has the null vector for its mean and has dispersion matrix CC'. This property can be easily shown using two basic properties of univariate random variables:

$$E(c_1 U + c_2 V) = c_1 E(U) + c_2 E(V)$$

and

$$VAR(c_1 U + c_2 V) = c_1^2 VAR(U) + c_2^2 VAR(V) + 2c_1c_2 COV(U, V).$$

Here, U and V are univariate random variables, c_1 and c_2 are scalar constants, E is the expected value operator, VAR is the variance operator, and COV is the covariance operator.

Thus a distribution with any desired dispersion matrix, Σ, can be obtained from X by determining C such that $CC' = \Sigma$. As given in CHAPTER II, C exists and is unique for Σ positive definite. The results for X and Y will be identical. This follows from the fact that

$$(Y_{ij} - \bar{Y}^{ij}_{m.})' S_{TY}^{-1}(Y_{ij} - \bar{Y}^{ij}_{m.}) = (CX_{ij} - C\bar{X}^{ij}_{m.})'[C S_{TX} C']^{-1}(CX_{ij} - C\bar{X}^{ij}_{m.}) =$$

$$(X_{ij} - \bar{X}^{ij}_{m.})' C' C'^{-1} S_{TX}^{-1} C^{-1} C(X_{ij} - \bar{X}^{ij}_{m.}) = (X_{ij} - \bar{X}^{ij}_{m.})S_{TX}^{-1}(X_{ij} - \bar{X}^{ij}_{m.}),$$

where S_{TX} and S_{TY} are the sample total dispersion matrices for observations sampled from X and Y respectively. Thus the distance between an observation and a mean vector is invariant under changes in the population dispersion matrix when there is homogeneity of dispersion matrices.

Tables 5 and 6 give the population dispersion matrices and correlation matrices that were used when the dispersion matrices were

TABLE 5

POPULATION DISPERSION AND CORRELATION MATRICES
FOR THREE DIMENSIONS

	Σ_1			R_1	
1.00	1.00	-3.00	1.00	0.44	-0.60
1.00	5.00	-3.00	0.44	1.00	-0.26
-3.00	-3.00	25.00	-0.60	-0.26	1.00
	Σ_2			R_2	
4.00	2.00	0.00	1.00	0.31	0.00
2.00	10.00	3.00	0.31	1.00	0.42
0.00	3.00	5.00	0.00	0.42	1.00
	Σ_3			R_3	
2.25	0.75	-3.00	1.00	0.31	-0.53
0.75	2.50	0.50	0.31	1.00	0.08
-3.00	0.50	14.00	-0.53	0.08	1.00
	Σ_4			R_4	
1.00	2.00	0.00	1.00	0.89	0.00
2.00	5.00	-1.00	0.89	1.00	-0.20
0.00	-1.00	5.00	0.00	-0.20	1.00

NOTE: Σ_1 and Σ_2 are used when there are two treatment groups.

TABLE 6

POPULATION DISPERSION AND CORRELATION MATRICES
FOR NINE DIMENSIONS

Σ_1

4.00	2.00	2.00	0.00	2.00	-2.00	0.00	0.00	-2.00
2.00	2.00	0.00	0.00	1.00	-1.00	1.00	0.00	0.00
2.00	0.00	3.00	1.00	1.00	-1.00	0.00	0.00	-1.00
0.00	0.00	1.00	2.00	-1.00	-1.00	1.00	1.00	0.00
2.00	1.00	1.00	-1.00	3.00	1.00	-1.00	-1.00	-1.00
-2.00	-1.00	-1.00	-1.00	1.00	4.00	-1.00	0.00	2.00
0.00	1.00	0.00	1.00	-1.00	-1.00	4.00	-1.00	3.00
0.00	0.00	0.00	1.00	-1.00	0.00	-1.00	4.00	0.00
-2.00	0.00	-1.00	0.00	-1.00	2.00	3.00	0.00	7.00

R_1

1.00	0.70	0.57	0.00	0.57	-0.50	0.00	0.00	-0.37
0.70	1.00	0.00	0.00	0.40	-0.35	0.35	0.00	0.00
0.57	0.00	1.00	0.40	0.33	-0.28	0.00	0.00	-0.21
0.00	0.00	0.40	1.00	-0.40	-0.35	0.35	0.35	0.00
0.57	0.40	0.33	-0.40	1.00	0.28	-0.28	-0.28	-0.21
-0.50	-0.35	-0.28	-0.35	0.28	1.00	-0.25	0.00	0.37
0.00	0.35	0.00	0.35	-0.28	-0.25	1.00	-0.25	0.56
0.00	0.00	0.00	0.35	-0.28	0.00	-0.25	1.00	0.00
-0.37	0.00	-0.21	0.00	-0.21	0.37	0.56	0.00	1.00

Σ_2

1.00	1.00	1.00	-1.00	0.00	1.00	-1.00	1.00	-1.00
1.00	2.00	0.00	0.00	0.00	0.00	0.00	1.00	-2.00
1.00	0.00	3.00	-3.00	1.00	2.00	-1.00	1.00	1.00
-1.00	0.00	-3.00	4.00	-1.00	-1.00	1.00	-1.00	0.00
0.00	0.00	1.00	-1.00	2.00	1.00	2.00	0.00	0.00
1.00	0.00	2.00	-1.00	1.00	5.00	-2.00	0.00	-1.00
-1.00	0.00	-1.00	1.00	2.00	-2.00	6.00	1.00	2.00
1.00	1.00	1.00	-1.00	0.00	0.00	1.00	4.00	1.00
-1.00	-2.00	1.00	0.00	0.00	-1.00	2.00	1.00	8.00

TABLE 6--Continued

R_2

1.00	0.70	0.57	-0.50	0.00	0.44	-0.40	0.50	-0.35
0.70	1.00	0.00	0.00	0.00	0.00	0.00	0.35	-0.50
0.57	0.00	1.00	-0.86	0.40	0.51	-0.23	0.28	0.20
-0.50	0.00	-0.86	1.00	-0.35	-0.22	0.20	-0.25	0.00
0.00	0.00	0.40	0.35	1.00	0.31	0.57	0.00	0.00
0.44	0.00	0.51	-0.22	0.31	1.00	-0.36	0.00	-0.15
-0.40	0.00	-0.23	0.20	0.57	-0.36	1.00	0.20	0.28
0.50	0.35	0.28	-0.25	0.00	0.00	0.20	1.00	0.17
-0.35	-0.50	0.20	0.00	0.00	-0.15	0.28	0.17	1.00

Σ_3

1.00	0.00	1.00	1.00	1.00	0.00	-1.00	1.00	-1.00
0.00	1.00	1.00	0.00	1.00	-1.00	0.00	1.00	0.00
1.00	1.00	3.00	1.00	3.00	-1.00	0.00	1.00	-2.00
1.00	0.00	1.00	2.00	0.00	-1.00	-2.00	0.00	0.00
1.00	1.00	3.00	0.00	5.00	0.00	1.00	3.00	-4.00
0.00	-1.00	-1.00	-1.00	0.00	3.00	0.00	0.00	-1.00
-1.00	0.00	0.00	-2.00	1.00	0.00	5.00	0.00	-2.00
1.00	1.00	1.00	0.00	3.00	0.00	0.00	7.00	-3.00
-1.00	0.00	-2.00	0.00	-4.00	-1.00	-2.00	-3.00	6.00

R_3

1.00	0.00	0.57	0.70	0.44	0.00	-0.44	0.37	-0.40
0.00	1.00	0.57	0.00	0.44	-0.57	0.00	0.37	0.00
0.57	0.57	1.00	0.40	0.77	-0.33	0.00	0.21	-0.47
0.70	0.00	0.40	1.00	0.00	-0.40	-0.63	0.00	0.00
0.44	0.44	0.77	0.00	1.00	0.00	0.20	0.50	-0.73
0.00	-0.57	-0.33	-0.40	0.00	1.00	0.00	0.00	-0.23
-0.44	0.00	0.00	-0.63	0.20	0.00	1.00	0.00	-0.36
0.37	0.37	0.21	0.00	0.50	0.00	0.00	1.00	-0.46
-0.40	0.00	-0.47	0.00	-0.73	-0.23	-0.36	-0.46	1.00

TABLE 6--Continued

Σ_4

1.00	1.00	-1.00	-1.00	0.00	0.00	1.00	0.00	1.00
1.00	2.00	-1.00	0.00	0.00	0.00	2.00	-1.00	1.00
-1.00	-1.00	2.00	0.00	0.00	0.00	-1.00	1.00	0.00
-1.00	0.00	0.00	4.00	0.00	0.00	0.00	-3.00	-3.00
0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	1.00
0.00	0.00	0.00	0.00	1.00	2.00	1.00	-1.00	0.00
1.00	2.00	-1.00	0.00	1.00	1.00	4.00	-1.00	2.00
0.00	-1.00	1.00	-3.00	0.00	-1.00	-1.00	5.00	2.00
1.00	1.00	0.00	-3.00	1.00	0.00	2.00	2.00	7.00

R_4

1.00	0.70	-0.70	-0.50	0.00	0.00	0.50	0.00	0.37
0.70	1.00	-0.50	0.00	0.00	0.00	0.70	-0.31	0.26
-0.70	-0.50	1.00	0.00	0.00	0.00	-0.35	0.31	0.00
-0.50	0.00	0.00	1.00	0.00	0.00	0.00	-0.67	-0.56
0.00	0.00	0.00	0.00	1.00	0.70	0.50	0.00	0.37
0.00	0.00	0.00	0.00	0.70	1.00	0.35	-0.31	0.00
0.50	0.70	-0.35	0.00	0.50	0.35	1.00	-0.22	0.37
0.00	-0.31	0.31	-0.67	0.00	-0.31	-0.22	1.00	0.33
0.37	0.26	0.00	-0.56	0.37	0.00	0.37	0.33	1.00

NOTE: Σ_1 and Σ_2 are used when there are two treatment groups.

not equal. The two tables give the four dispersion and four correlation matrices for three dimensions and nine dimensions, respectively. In both cases Σ_1 and Σ_2 were the population dispersion matrices when there were two treatment groups and heterogeneity of dispersion matrices. Of course, all four dispersion matrices were required when there were four treatment groups. These dispersion matrices were selected so that there would be unequal variances, both among variables within treatment group and among treatment groups for a given variable. Also, the correlations range from strong negative correlation to strong positive correlation.

The numbers of treatment groups investigated were two and four, and the sample sizes investigated were fifteen and thirty per treatment group.

Mean vectors were either equal or not equal. When population mean vectors were not equal an attempt was made to separate the mean vectors in such a manner that the power for the MVAOV was 75 to 95 percent. This attempt was made because a power of 75 to 95 percent is reasonable for an applied situation, and a comparison of the proposed test and the MVAOV is most meaningful in this range.

All combinations of the given levels of important factors were investigated except for combinations where the distribution is non-normal and dispersion matrices are equal. For each combination, a series of 75 experiments was conducted. CHAPTER V gives the results of the empirical investigation and a discussion of the test procedure.

CHAPTER V

DISCUSSION AND EVALUATION

The results from the empirical investigation for the various combinations of the factor levels are given in two by two contingency tables with the column headings F Significant (FS) and F Not Significant (FNS), and with row headings U Significant (US) and U Not Significant (UNS). The number in the FS column and US row is the number of experiments for which the test statistic for both the proposed test and the MVAOV was significant at the five percent level. The number in the FS column and UNS row is the number of experiments where the test statistic for the MVAOV was significant but the proposed test statistic was not significant. The number of experiments where the proposed test statistic was significant but the F was not is given in the FNS column and US row. The number in the FNS column and UNS row is the number of experiments where neither test statistic was significant at the five percent level.

The sum of the two numbers in the first row is the number of experiments where the proposed U statistic was significant. Since there were 75 experiments in each series, this sum divided by 75 is an estimate of the α level for the proposed test when the mean vectors are equal and an estimate of the power for the proposed test when the mean vectors are not equal. The experiments within a series are statistically

independent and the binomial test can be used to test whether the true α exceeds the stated five percent level, where the event associated with the binomial is the significance or non-significance of the test statistic. If the true α level for the proposed test is the stated five percent, U can be expected to be significant 3.75 times. When U is significant eight or more times, a true α level of five percent is not accepted since this would occur by chance only three percent of the time if α were truly five percent. If U is significant seven times or less, a true α of five percent or less is not inconsistent with the data.

The sum of the numbers in the first column is the number of experiments where the F was significant at the five percent level. This sum divided by 75 is an estimate of the α level for the MVAOV when the mean vectors are equal and an estimate of the power when the mean vectors are not equal. A test of the true α level for the MVAOV can be performed in the same manner as for the U statistic.

It is also possible to test whether the proposed test and the MVAOV are equivalent by using the binomial test. This is done by comparing the lower left-hand number (FS-UNS) to the upper right-hand number (FNS-US) since these are the values where the two test procedures give different results. If the two test procedures are equivalent, the FNS-US value should be approximately equal to the FS-UNS value, since only chance should cause any difference in the values.

Table 7 gives the results for two treatment groups with multivariate normal distributions. The headings indicate the number of dimensions, the equality or non-equality of population mean vectors and dispersion matrices, and the sample sizes.

TABLE 7

RESULTS FOR TWO TREATMENT GROUPS WITH MULTIVARIATE NORMAL DISTRIBUTIONS

[a]

Three Dimensions, $\Sigma_1 = \Sigma_2$, $\mu_1 = \mu_2$

15 Observations/Group

	FS	FNS
US	5	1
UNS	0	69

30 Observations/Group

	FS	FNS
US	4	3
UNS	0	68

[b]

Three Dimensions, $\Sigma_1 \neq \Sigma_2$, $\mu_1 = \mu_2$

15 Observations/Group

	FS	FNS
US	1	3
UNS	0	71

30 Observations/Group

	FS	FNS
US	4	3
UNS	0	68

[c]

Three Dimensions, $\Sigma_1 = \Sigma_2$, $\mu_1 \neq \mu_2$

15 Observations/Group

	FS	FNS
US	62	7
UNS	0	6

30 Observations/Group

	FS	FNS
US	59	5
UNS	0	11

[d]

Three Dimensions, $\Sigma_1 \neq \Sigma_2$, $\mu_1 \neq \mu_2$

15 Observations/Group

	FS	FNS
US	60	3
UNS	0	12

30 Observations/Group

	FS	FNS
US	65	1
UNS	0	9

US - U Significant
UNS - U Not Significant
FS - F Significant
FNS - F Not Significant

TABLE 7--Continued

Nine Dimensions, $\Sigma_1 = \Sigma_2$, $\mu_1 = \mu_2$ [e]

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	3	3	US	3	5
UNS	0	69	UNS	0	67

Nine Dimensions, $\Sigma_1 \neq \Sigma_2$, $\mu_1 = \mu_2$ [f]

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	5	4	US	2	3
UNS	0	66	UNS	0	70

Nine Dimensions, $\Sigma_1 = \Sigma_2$, $\mu_1 \neq \mu_2$ [g]

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	68	3	US	63	6
UNS	0	4	UNS	0	6

Nine Dimensions, $\Sigma_1 \neq \Sigma_2$, $\mu_1 \neq \mu_2$ [h]

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	70	3	US	67	3
UNS	0	2	UNS	0	5

US - U Significant
UNS - U Not Significant
FS - F Significant
FNS - F Not Significant

The numbers in Table 7a for 15 observations per group are 5, 1, 0 and 69. Thus U was significant six times and the estimate of the true α is 6/75 = .08. This error rate is not significantly larger than five percent. The F was significant five times for an error rate of .07. The MVAOV and the proposed test gave the same results for 74 of the 75 experiments. Thus for this series there is little difference in the results for the two test procedures.

The proposed test statistic was significant seven times in Table 7a for 30 observations per group. The error rate is not significantly higher than five percent, but the fact that U was significant 13 times in the 150 experiments given in Table 7a indicates that perhaps the true α level for the proposed test does exceed five percent when there are three dimensions, homogeneity of dispersion matrices, and multivariate normal distributions.

The observed Type I error rate for the proposed test significantly exceeds the stated five percent rate for Table 7e with 30 observations per group and Table 7f with 15 observations per group (p = .03 and p = .01). Although only the two values significantly exceed the five percent rate in Table 7, the consistency with which the number of observed significant U's is near eight when the null hypothesis is true indicates a possibility that the true α level does in fact exceed the stated α. Also, U was significant more often than F for every series of experiments in Table 7. There is no important difference between the results where $\Sigma_1 = \Sigma_2$ and the results where $\Sigma_1 \neq \Sigma_2$. Based on the data in Table 7, the proposed test cannot be recommended when there are two treatment groups and multivariate normal distributions because of the

likelihood of an excessive Type I error rate.

There is no lack of robustness indicated for the MVAOV in Table 7. The true α level for the MVAOV seems to be about five percent even when the dispersion matrices are heterogeneous. This is supported by the fact that the overall observed Type I error rate is .04 for all series in Table 7 where the mean vectors are equal and the dispersion matrices are heterogeneous.

The results given in Table 8 are not so consistent as the results given in Table 7. The rejection rate for the proposed test is at times greater than that of the MVAOV and at other times smaller. The F statistic was significant six more times than the U in Table 8f with 15 observations per group. This could be taken as an indication that the proposed test has a smaller error rate than the MVAOV when there are four treatment groups and heterogeneity of dispersion matrices. However, the six and the zero are not significantly different, and the difference in the two test statistics is not so great for the same situation when there are 30 observations per group.

The error rate is significantly greater than five percent for both the proposed test and the MVAOV in Table 8b with 30 observations per group where both statistics were significant eight times. Inspection of the four series of experiments where the $\mu_i = \mu_j$ and $\Sigma_i \neq \Sigma_j$ reveals that in three of the four cases the F was significant six or more times. This lends some support to the hypothesis that the MVAOV is not so robust for four treatment groups as it is for two when dispersion matrices are heterogeneous. However, the quantity of the data is not sufficient to warrant a strong statement on this point.

TABLE 8

RESULTS FOR FOUR TREATMENT GROUPS WITH MULTIVARIATE NORMAL DISTRIBUTIONS

a

Three Dimensions, $\Sigma_i = \Sigma_j$, $\mu_i = \mu_j$

15 Observations/Group

	FS	FNS
US	3	0
UNS	0	72

30 Observations/Group

	FS	FNS
US	2	1
UNS	0	72

b

Three Dimensions, $\Sigma_i \neq \Sigma_j$, $\mu_i = \mu_j$

15 Observations/Group

	FS	FNS
US	3	0
UNS	0	72

30 Observations/Group

	FS	FNS
US	8	0
UNS	0	67

c

Three Dimensions, $\Sigma_i = \Sigma_j$, $\mu_i \neq \mu_j$

15 Observations/Group

	FS	FNS
US	63	3
UNS	0	9

30 Observations/Group

	FS	FNS
US	44	11
UNS	0	20

d

Three Dimensions, $\Sigma_i \neq \Sigma_j$, $\mu_i \neq \mu_j$

15 Observations/Group

	FS	FNS
US	70	2
UNS	0	3

30 Observations/Group

	FS	FNS
US	43	0
UNS	7	25

US - U Significant
UNS - U Not Significant
FS - F Significant
FNS - F Not Significant

TABLE 8--Continued

e

Nine Dimensions, $\Sigma_i = \Sigma_j$, $\mu_i = \mu_j$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	2	0	US	1	1
UNS	2	71	UNS	0	73

f

Nine Dimensions, $\Sigma_i \neq \Sigma_j$, $\mu_i = \mu_j$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	5	0	US	5	0
UNS	6	64	UNS	1	69

g

Nine Dimensions, $\Sigma_i = \Sigma_j$, $\mu_i \neq \mu_j$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	41	5	US	47	6
UNS	0	29	UNS	0	22

h

Nine Dimensions, $\Sigma_i \neq \Sigma_j$, $\mu_i \neq \mu_j$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	49	3	US	37	4
UNS	2	21	UNS	0	34

US - U Significant
UNS - U Not Significant
FS - F Significant
FNS - F Not Significant

The proposed test was significantly more powerful than the MVAOV for Table 8c with 30 observations per group ($p < .001$). There is some indication that the MVAOV is more powerful for Table 8d with 30 observations per group and that the proposed test is more powerful for Table 8g with 30 observations per group.

If one of the two test procedures were to be selected for all situations where there are four treatment groups based only on the information in Table 8, it would be the proposed test since a lower α level is strongly indicated once and a lower β level is strongly indicated twice whereas a lower α level for the MVAOV is never strongly indicated and a lower β level for the MVAOV is strongly indicated once. This implies neither a significant nor a meaningful difference, between the two test procedures, but only that the data lend slightly more support to the use of the proposed test than the MVAOV when there are four treatment groups and multivariate normal distributions.

The error rates for the proposed test are significantly larger than the stated five percent for both series of experiments in Table 9c ($p = .03$ and $p = .01$). The proposed test is significantly more powerful than the MVAOV for Table 9d with 15 observations per group ($p < .001$). The only series that even indicates a lack of robustness for the MVAOV is Table 9c with 30 observations per group where the F was significant six times. The interpretation of the results for two treatment groups with non-normal distributions is similar to that for the normal distributions: the possibility of an excessive Type I error rate for the proposed test and robustness of the MVAOV.

The data presented in Table 10 do not show any great difference

TABLE 9

RESULTS FOR TWO TREATMENT GROUPS WITH MULTIVARIATE NON-NORMAL DISTRIBUTIONS

[a] Three Dimensions, $\Sigma_1 \neq \Sigma_2$, $\mu_1 = \mu_2$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	0	1	US	4	1
UNS	0	74	UNS	0	70

[b] Three Dimensions, $\Sigma_1 \neq \Sigma_2$, $\mu_1 \neq \mu_2$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	61	5	US	69	0
UNS	0	9	UNS	0	6

[c] Nine Dimensions, $\Sigma_1 \neq \Sigma_2$, $\mu_1 = \mu_2$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	3	5	US	6	3
UNS	0	67	UNS	0	66

[d] Nine Dimensions, $\Sigma_1 \neq \Sigma_2$, $\mu_1 \neq \mu_2$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	52	12	US	62	4
UNS	0	11	UNS	0	9

US - U Significant
UNS - U Not Significant
FS - F Significant
FNS - F Not Significant

TABLE 10

RESULTS FOR FOUR TREATMENT GROUPS WITH MULTIVARIATE NON-NORMAL DISTRIBUTIONS

a

Three Dimensions, $\Sigma_i \neq \Sigma_j$, $\mu_i = \mu_j$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	7	0	US	6	0
UNS	4	64	UNS	0	69

b

Three Dimensions, $\Sigma_i \neq \Sigma_j$, $\mu_i \neq \mu_j$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	58	4	US	55	3
UNS	1	12	UNS	0	17

c

Nine Dimensions, $\Sigma_i \neq \Sigma_j$, $\mu_i = \mu_j$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	5	0	US	3	1
UNS	1	69	UNS	0	71

d

Nine Dimensions, $\Sigma_i \neq \Sigma_j$, $\mu_i \neq \mu_j$

15 Observations/Group			30 Observations/Group		
	FS	FNS		FS	FNS
US	64	2	US	66	5
UNS	0	9	UNS	0	4

US - U Significant
UNS - U Not Significant
FS - F Significant
FNS - F Not Significant

between the two test procedures. There is a slight tendency for the proposed test to have smaller Type I and Type II error rates. This slight tendency is due to the results in Table 10a with 15 observations per group and Table 10d with 30 observations per group.

The Type I error rate for the MVAOV was significantly larger than five percent for Table 10a with 15 observations per group ($p < .001$). Although the results were not significant there is an indication that the MVAOV has a true error rate greater than five percent for Table 10a with 30 observations per group and for Table 10c with 15 observations per group. Thus it appears that the MVAOV may not be so robust for four treatment groups as it is for two when the distributions are non-normal and there is heterogeneity of dispersion matrices.

Some support of the validity of the data is provided by the fact that for the 600 experiments where the null hypothesis was true and the assumptions for the MVAOV were met, the F was significant 23 times. The 23 is not significantly different from 30 which is the number expected if the true α is five percent. Since it is known that the true α for the MVAOV is five percent when the assumptions are met, it can be assumed that the populations specified are quite similar to the populations actually sampled.

The general impressions gained from the empirical study are that the proposed method should not be used when there are two treatment groups, the proposed method may be safely used when there are four treatment groups, and the MVAOV may not be so robust for four treatment groups as for two.

The prohibition against the use of the proposed test when there

are two treatment groups is due to the possibility of an excessive Type I error rate for both normal and non-normal distributions. When there are four treatment groups the α and β levels of the proposed test appear no greater than the levels of the MVAOV when the assumptions for the MVAOV are met, and the observed Type I and Type II error rates are usually slightly lower for the proposed test than for the MVAOV when the assumptions for the MVAOV are not met.

The lack of robustness of the MVAOV for four treatment groups is indicated by the fact that when the null hypothesis is true and the assumptions for the MVAOV are not met, the MVAOV rejection rate is often large enough to be near the five percent critical value.

The proposed test procedure appears to be satisfactory for four treatment groups, but further investigation of the test statistic is required before the method can be considered practical. Perhaps a detailed investigation of the distance measure utilized would reveal modifications which would result in a test procedure satisfactory for two treatment groups. A determination of the effect of unequal sample sizes would be beneficial for use in applied situations. Also of use would be methods which are satisfactory for partitioning the U when significant differences are found, in order to ascertain which treatment groups are responsible for the differences. Perhaps procedures could be found which would be used to determine those variables of importance in detecting differences in mean vectors.

CHAPTER VI

SUMMARY

This dissertation is concerned with the development and evaluation of a multivariate test of location for multiple samples. Various measures of the distances between an observation and the sample mean vectors, which are used to assign observations to subsets, and procedures for testing the resultant subsets are investigated using Monte Carlo techniques. Difficulties in the use of the subsets are encountered and a continuous statistic, U, is selected for investigation.

U is obtained by first associating two distances with each observation. The first distance is between the observation and the sample mean vector for that observation's treatment group, and the second distance is between the observation and the sample mean vector for all observations not in the same treatment group as the observation. Then a difference is associated with each observation, this difference being the difference of the two distances. U is the mean difference divided by the standard error of the mean difference.

The distribution of U under the null hypothesis of equal mean vectors is determined empirically, and the test procedure is evaluated with respect to Type I and Type II errors for different combinations of distributions, numbers of treatment groups, sample sizes, numbers of dimensions, and separations of population mean vectors for both

homogeneous and heterogeneous dispersion matrices. This evaluation is based upon an empirical study, and the Multivariate Analysis of Variance (MVAOV) is used as a control test procedure.

The results from the empirical investigation indicate that the true α level for the proposed test exceeds the stated α level when there are two treatment groups, both for multivariate normal and multivariate non-normal distributions. For this reason the proposed test is not recommended when there are two treatment groups.

Although the results are not conclusive, the data lend some support to the use of the proposed test rather than the MVAOV when there are four treatment groups. The α and β levels for the proposed test appear to be no higher than those for the MVAOV when the assumptions for the MVAOV are met. The observed Type I and Type II error rates for the proposed test are usually slightly lower than those of the MVAOV when the assumptions of normality and homogeneity of dispersion matrices do not hold.

No strong statement is warranted by the data, but there is some indication that the MVAOV is not so robust for four treatment groups as for two.

Although the proposed test appears to be satisfactory when there are four treatment groups, the evidence for preferring the proposed test to the MVAOV is very slight and further investigation is required before the test can be considered practical. More information on the properties of the distance measure utilized could reveal modifications which would result in a test suitable for two treatment groups and superior to the MVAOV when its assumptions are not met. Mathematical investigation of

the distance measure is hampered by the complexity of the quadratic form involved, and it may be necessary to rely on empirical techniques. In spite of the complexity of the distance measure, the results of the present work seem to warrant continued research on the measure.

LIST OF REFERENCES

Allais, D. C. 1964 The selection of measurements for prediction. Stanford Elec. Lab., System Theory Div. Report, TR6103-9 (AD 456 770), Stanford, California

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

Anderson, T. W. 1966 Some nonparametric multivariate procedures based on statistically equivalent blocks. Multivariate Analysis, P. R. Krishnaiah, ed. Academic Press, New York, 5-28.

Ball, G. H. and Hall, D. J. 1965 ISODATA, A novel method of data analysis and pattern classification. Stanford Research Institute, Menlo Park.

Bhapkar, V. P. 1966 Some nonparametric tests for the multivariate several sample location problem. In Multivariate Analysis, P. R. Krishnaiah, ed. Academic Press, New York, 29-42.

Bose, R. C_2 and Roy, S. N. 1938 The distribution of the studentised D^2-statistic. Sankhya 4: 19-38.

Box, G. P. and Muller, M. E. 1958 A note on the generation of random normal deviates. Annals of Mathematical Statistics 29: 610-611.

Chung, J. H. and Fraser, D. A. S. 1958 Randomization tests for a multivariate two-sample problem. Journal of the American Statistical Association 53: 729-735.

Cooley, W. W. and Lohnes, P. R. 1962 Multivariate Procedures for the Behavioral Sciences. John Wiley and Sons, Inc., New York.

Defrisse-Gussenhovern, E. 1966 A masculinity-femininity scale based on a discriminant function. Acta Genetica et Statistica Medica 16: 198-208.

Dorr, A. E. 1969 An empirical investigation of a clustering technique for testing the equality of mean vectors. M. S. Thesis, University of Oklahoma Medical Center, Oklahoma City.

Graybill, F. A. 1961 An Introduction to Linear Statistical Models, Volume 1. McGraw-Hill Book Company, Inc., New York.

Holloway, L. N. and Dunn, O. J. 1967 The robustness of Hotelling's T^2. Journal of the American Statistical Association 62: 124-136.

International Business Machine Corporation 1967 IBM Application Program, 1130 Scientific Subroutine Package (1130-CM-02X) Programmer's Manual, White Plains, New York.

Ito, K. and Schull, W. J. 1964 On the robustness of the T_o^2 test in multivariate analysis of variance when variance-covariance matrices are not equal. Biometrika 51: 71-82.

Jacobson, N. 1953 Lectures in Abstract Algebra. D. Van Nostrand Company, Inc., New York, 174-176.

Jansson, B. 1966 Random Number Generators. Victor Pettersons Bokindustri Aktiebolag, Stockholm.

Koch, G. G. and Sen, P. K. 1968 Some aspects of the statistical analysis of the mixed model. Biometrics 24: 27-47.

Mahalanobis, P. C. 1930 On tests and measures of group divergence. J. and Proc. Asiat. Soc. Beng. 26: 541-588.

Mahalanobis, P. C. 1936 On the generalized distance in statistics. Proc. Nat. Inst. Sci. India 2: 49-55.

Murthy, V. K. 1966 Nonparametric estimation of multivariate densities with applications. In Multivariate Analysis, P. R. Krishnaiah, ed. Academic Press, New York, 43-58.

Parzen, E. 1960 Modern Probability Theory and Its Applications. John Wiley and Sons, Inc., New York.

Puri, M. L. and Sen, P. K. 1966 On a class of multivariate multisample rank order tests. Sankhya, Series A 28: 353-376.

Reyment, R. A. 1962 Observations on homogeneity of covariance matrices in paleontologic biometry. Biometrics 18: 1-11.

Scheuer, E. M. and Stoller, D. S. 1962 On the generation of normal random vectors. Technometrics 4: 278-281.

Sen, P. K. and Puri, M. L. 1967 On the theory of rank order tests for location in the multivariate one sample problem. Annals of Mathematical Statistics 38: 1216-1228.

Steel, R. G. D. and Torrie, J. H. 1960 Principles and Procedures of Statistics. McGraw-Hill Book Company, Inc., New York.

APPENDIXES

APPENDIX 1

RANDU

RANDU is an IBM supplied subroutine and is contained in the 1130 Scientific Subroutine Package (1967). This subroutine generates random numbers from the uniform distribution on the interval (0, 1), i.e., from U(0, 1). RANDU is machine specific and is applicable to the IBM 1800 and the IBM 1130. Random numbers from U(0, 1) are obtained by generating odd random integers on the interval (0, 32768) and dividing by 32768. Only the generation of the odd integers will be considered.

This random number generator produces a series of random numbers from the recurrence relation

$$X_{n+1} = 899 \cdot X_n \pmod{2^{15}}$$

where X_i is the i^{th} random number in the series and X_o is an odd integer on the interval (1, 32767) which is supplied by the user. Jansson (1966) refers to this type of random number generator as a multiplicative generator and gives its properties.

The maximum period of this generator is 2^{13} or 8192, i.e., after 8192 random numbers the numbers begin repeating themselves. Furthermore, the generator produces numbers of the type $8\nu + 1$ and $8\nu + 3$ ($\nu = 0, 1, \ldots, 2^{12} - 1$) when $X_o \equiv 1, 3, 9,$ or 11 (mod 16) and numbers of the type $8\nu + 5$ and $8\nu + 7$ ($\nu = 0, 1, \ldots, 2^{12} - 1$) when $X_o \equiv 5, 7, 13,$ or 15 (mod 16).

In effect, there are only two sequences each containing 8192 numbers, with a fixed order. An X_o determines only one of the two sequences and the starting point in the sequence.

APPENDIX 2

COMPUTER TIME REQUIREMENTS

The following information is presented so that one interested in conducting an empirical investigation similar to the present study may have an approximation as to the computer time required. The times given are those that were required using FORTRAN IV programs and an IBM 1800 with a four micro-second access time. All work was done in extended precision. There were three principal divisions for the computations performed: generation of the multivariate vectors, analysis using the MVAOV, and analysis using the proposed test.

The transformed univariate random numbers were generated and placed on a magnetic disk prior to the investigation so this time is not reflected in the times given. The generation of vectors consists of reading the univariate random number from magnetic disk and multiplying each vector thus obtained by the appropriate C matrix and adding the appropriate μ vector to the result. The generation of the multivariate vectors appeared to require a small portion of the total time required, perhaps five percent.

The primary time requirements for the MVAOV analysis were for computation of sample mean vectors, dispersion matrices, and determinants of pooled and total dispersion matrices. The MVAOV analysis appeared to require about ten percent of the total time.

Mean vectors and inverted matrices utilized by the proposed test had been computed during the execution of the MVAOV; still about 85 percent of the total time was for the proposed test.

The times required for one experiment when there were three dimensions and two treatment groups were .007 hours and .009 hours for sample sizes of fifteen and thirty per group. For nine dimensions these times were .032 hours and .056 hours. When there were four treatment groups, .012 hours and .019 hours were required for three dimensions, and .060 hours and .107 hours for nine dimensions.

So that the reader is not misled by the smallness of the numbers, it is pointed out that 8.04 hours of computer time were required for one series of seventy-five experiments when there were nine dimensions, four treatment groups, and thirty observations per group.