MULTIPLE COMPARISON METHODS AND CERTAIN DISTRIBUTIONS ARISING IN MULTIVARIATE STATISTICAL ANALYSIS

by

Om Parkash Bagai

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department of

MATHEMATICS

We accept this thesis as conforming to the standard required from candidates for the degree of Doctor of Philosophy.

Members of the Department of Mathematics

THE UNIVERSITY OF BRITISH COLUMBIA

April, 1960

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **Mathematics**

The University of British Columbia,
Vancouver 8, Canada.

Date **April 13, 1960**

GRADUATE STUDIES

Field of Study: Mathematics

Modern Algebra ....D. C. Murdoch

Mathematical Statistics ....S. W. Nash

Computation Methods ....F. M. Goodspeed

Real Variables ....P. S. Bullen

Probability and Statistics ....S. W. Nash

Other Studies:

Theoretical Mechanics ....W. Opechowski

Industrial Statistics ....T. I. Matuszewski

Directed Studies (Plant Science) ....V. C. Brink

Archive

The University of British Columbia

Faculty of Graduate Studies

PROGRAMME OF THE

FINAL ORAL EXAMINATION

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

of

OM PARKASH BAGAI

B.A. University of Panjab, India, 1948
M.A. University of Panjab, India, 1950

ROOM 225, BUCHANAN BUILDING

TUESDAY, APRIL 26, 1960 AT 9:30 A.M.

COMMITTEE IN CHARGE

Dean F. H. SOWARD: Chairman

S. W. NASH
S. A. JENNINGS
R. A. RESTREPO
B. N. MOYLS
W. OPECHOWSKI
V. C. BRINK
T. I. MATUSZEWSKI
C. WALDEN
A. D. MOORE

External Examiner: PROF. E. S. KEEPING
University of Alberta, Edmonton, Alta.

MULTIPLE COMPARISON METHODS AND CERTAIN DISTRIBUTIONS ARISING IN MULTIVARIATE STATISTICAL ANALYSIS

ABSTRACT

The problem of classifying multivariate normal populations into homogeneous clusters on the basis of random samples drawn from those populations is taken up. Three alternative methods have been suggested for this. One of them is explained fully with an illustrative example, and the tabular values for the corresponding statistic, used for the purpose, have been computed. In the case of the other two alternatives only the working procedure is discussed. Further, a new statistic R, 'the largest distance', is proposed in one of these two alternatives, and its distribution is determined for the bivariate case in the form of definite integrals.

Ignoring *a priori* probabilities, two alternative methods are suggested for assigning an arbitrary population to one or more clusters of populations, and are demonstrated by an illustrative example.

A method is discussed for finding confidence regions for the non-centrality parameters of the distributions of certain statistics used in multivariate analysis and this method is illustrated by an example.

The exact distribution of the determinant of the sum of products (S.P.) matrix is found (in series), both in the central and the non-central linear cases for particular values of the rank of the matrix. Further, these results have been made use of in finding the limiting distribution of the Wilks-Lawley statistic proposed for testing the null hypothesis of the equality of the mean vectors of any number of populations.

Six different statistics based on the roots of certain determinantal equations have been proposed for various tests of hypotheses arising in the problems of multivariate analysis of variance (Anova). Their distributions in the limited cases of two and three eigen roots have been found in the form of definite integrals. Also, the limiting distribution of Roy's statistics of the largest, an intermediate and the smallest eigen roots have been found by a simple, easy method of integration, which method is quite different from that of Nanda (1948).

Lastly, the distributions of the mean square and the mean product (M.P.) matrix have been approximated respectively in the univariate and multivariate cases of unequal sub-class numbers in the analysis of variance (Anova) of Model II.

PUBLICATION

Intermediate Algebra—Textbook for the Panjab University, India.

ABSTRACT

The problem of classifying multivariate normal populations into homogeneous clusters on the basis of random samples drawn from those populations is taken up. Three alternative methods have been suggested for this. One of them is explained fully with an illustrative example, and the tabular values for the corresponding statistic, used for the purpose, have been computed. In the case of the other two alternatives only the working procedure is discussed. Further, a new statistic R, 'the largest distance', is proposed in one of these two alternatives, and its distribution is determined for the bivariate case in the form of definite integrals.

Ignoring _a priori_ probabilities, two alternative methods are suggested for assigning an arbitrary population to one or more clusters of populations, and are demonstrated by an illustrative example.

A method is discussed for finding confidence regions for the non-centrality parameters of the distributions of certain statistics used in multivariate analysis and this method is also illustrated by an example.

The exact distribution of the determinant of the sum of products (S.P.) matrix is found (in series), both in the central and the non-central linear cases for particular values of the rank of the matrix. Further, these results have been made use of in finding the limiting distribution of the Wilks-Lawley statistic proposed for testing the null hypothesis of the equality of the mean vectors of any number of populations.

Six different statistics based on the roots of certain determinantal equations have been proposed for various tests of hypotheses arising in the problems of multivariate analysis of variance (Anova). Their distributions in the limited cases of two and three eigenroots have been found in the form of definite integrals. Also, the limiting distribution of the Roy's statistics of the largest, an intermediate and the smallest eigenroots have been found by a simple, easy method of integration, which method is quite different from that of Nanda (1948).

Lastly, the distributions of the mean square and the mean product (M.P.) matrix have been approximated respectively in the univariate and multivariate cases of unequal sub-class numbers in the analysis of variance (Anova) of Model II.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Stanley W. Nash of the Department of Mathematics for suggesting the topic of this thesis, and feels highly indebted to him for his invaluable advice during its preparation. His further thanks are due to Dr. S.A. Jennings and Dr. R.A. Restrepo for their assistance in preparation of the final manuscript.

He is pleased to thank the staff of the U.B.C. Computing Centre for their assistance in working on the problem for illustration and in the preparation of the tables of chi-square.

His thanks are also due to Mr. K.G. Fensom, Superintendent, Forest Products Laboratory, Canada, for allowing him to use the data for the illustrative example.

He is also pleased to acknowledge the support of the National Research Council of Canada which made this research possible.

CHAPTER ONE

INTRODUCTION

1.1 Test of Equality of Mean Vectors in the Case of Two p-variate Normal Populations

In sciences like anthropology, biology, and others, we often wish, on the basis of two p-variate samples drawn from two populations, to find whether the two populations, on a given probability level, are distinct or not. Karl Pearson (1921) gave a start to answering such a question by suggesting his well-known Coefficient of Racial Likeness (C.R.L.) to Tildesley (1921), and he himself discussed it in his paper in 1926. But this coefficient was found to be inadequate and was severely criticized by Mahalanobis and Morant as a measure of divergence. Mahalanobis (1925) modified C.R.L. and defined a measure of divergence D_2^2 the "Mahalanobis distance", both for classical and Studentized cases, as follows:

Given two p-variate samples of sizes N_1 and N_2 with observations X_{irh} $(i = 1, 2, \ldots, p;\ \ r = 1, 2;\ \ h = 1, 2, \ldots, N_r)$ drawn from two p-variate normal populations assumed to have the same covariance matrix Σ but different sets of means μ_{i1} and μ_{i2} $(i = 1, 2, \ldots, p)$, let $\bar{X}_{i1}$ and $\bar{X}_{i2}$ $(i = 1, 2, \ldots, p)$ respectively be the means of the ith trait from the two samples. If the covariance matrix (σ_{ij}) is known or has been computed on the basis of large samples, then, taking (σ^{ij})

the inverse of (σ_{ij}), the Mahalanobis distance in the classical case is defined as:

$$D_2^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij}(\bar{X}_{i1} - \bar{X}_{i2})(\bar{X}_{j1} - \bar{X}_{j2}) \qquad (1.1.1)$$

If (σ_{ij}) is not known, we estimate it from the samples and define the Studentized form as:

$$D_2^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} w^{ij}(\bar{X}_{i1} - \bar{X}_{i2})(\bar{X}_{j1} - \bar{X}_{j2}) \qquad (1.1.2)$$

$$\text{where } (N_1 + N_2 - 2)w_{ij} = \sum_{r=1}^{2} \sum_{h=1}^{N_r} (\bar{X}_{irh} - \bar{X}_{ir})(\bar{X}_{jrh} - \bar{X}_{jr})$$

and (w^{ij}) is the inverse of (w_{ij}).

Simultaneously Hotelling (1931) generalized Students' t to the multivariate case. We denote this by T_2^2. It was found to be identical (Roy and Bose 1938, Fisher 1938) in form to the Studentized D_2^2 except for a factor involving sample sizes, i.e.

$$T_2^2 = \frac{N_1 N_2}{N_1 + N_2} D_2^2 .$$

<u>Distributions of D_2^2 and T_2^2:</u>

$$\text{If } \Delta^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij}(\mu_{i1} - \mu_{i2})(\mu_{j1} - \mu_{j2}) \qquad (1.1.3)$$

be the measure of divergence between the populations, the distributions of (1.1.1.) and (1.1.2) for both central ($\Delta^2 = 0$) and non-central ($\Delta^2 \neq 0$) cases are known as stated below:

(i) In the Studentized case (Bose and Roy, 1938), under the null hypothesis $\mu_{i1} = \mu_{i2}$ ($i = 1, 2, \ldots, p$) or $\Delta^2 = 0$, the quantity

$$\frac{N_1 + N_2 - p - 1}{p} \times \frac{N_1 N_2}{N_1 + N_2} \times \frac{D_2^2}{N_1 + N_2 - 2}$$

is distributed as the central F-ratio with p and $(N_1 + N_2 - p - 1)$ degrees of freedom (D.F), while in the classical case, under the same null hypothesis,

$$\frac{N_1 N_2}{N_1 + N_2} D_2^2$$

is distributed as central chi-square with p D.F.

(ii) Again, in the Studentized case (Bose and Roy, 1938) for $\Delta^2 \neq 0$, the quantity

$$\frac{N_1 + N_2 - p - 1}{p} \quad \frac{N_1 N_2}{N_1 + N_2} \quad \frac{D_2^2}{N_1 + N_2 - 2}$$

is distributed as non-central F-ratio with p and $(N_1 + N_2 - p - 1)$ D.F. and parameter $\Delta^2 / (\frac{1}{N_1} + \frac{1}{N_2})$; while in the classical case, again for $\Delta^2 \neq 0$, $\frac{N_1 N_2}{N_1 + N_2} D_2^2$ is distributed as non-central chi-square with p D.F. and parameter $\Delta^2/(\frac{1}{N_1} + \frac{1}{N_2})$.

The distribution of T_2^2 in the central case was given by Hotelling (1931) and in the non-central case by Hsu (1938). These are identical to the distributions of Studentized D_2^2 except for the constant multiplier.

When the hypothesis of equality of mean vectors is rejected, the problem generally arises of giving confidence regions to the corresponding non-centrality parameter. We have attempted to answer this problem in Chapter Five, where we have taken simultaneously the case of two or any number of populations. We have first given the method and then, to demonstrate the method, we have presented an illustration.

1.2 Classification and Discrimination in the Case of k p-variate Normal Populations

Again in sciences like anthropology, biology and others, one is often faced with the problem of discrimination and classification. In the biological sciences we are concerned with specifying an individual as a member of one of the populations to which he can possibly belong, as when a taxonomist has to assign an organism to its proper species or sub-species or an anthropologist is faced with the problem of sexing a skull or jaw-bone. We are also faced with the problem of classification of the groups themselves into some significant system based on the configuration of the various characteristics, for example when 'a number of species or sub-species may have to be arrayed in hierarchical order showing the closeness of some and distinctiveness of the others'. In all such problems our first aim is to test whether the populations involved are distinct or not. Four statistics have been suggested for testing the hypothesis of equality of the mean vectors of the populations. We list them below:

Suppose we are given k p-variate normal populations, assumed to have the same covariance matrix Σ and distinct mean vectors $(\mu_{1r}, \mu_{2r}, \ldots, \mu_{pr})(r = 1, 2, \ldots, k)$. From these populations samples respectively of sizes $N_1, N_2, \ldots, N_k$ are drawn and observations X_{irh} $(i = 1, 2, \ldots, p;\ r = 1, 2, \ldots, k$ and $h = 1, 2, \ldots, N_r)$ are made. Let $W = (w_{ij})$ and $B = (b_{ij})$ be the within and between mean product (M.P.) matrices with respectively n_2 and n_1 D.F. where w_{ij} and b_{ij} are respectively defined as:

$$n_2 w_{ij} = \sum_{r=1}^{k} \sum_{h=1}^{N_r} (X_{irh} - \bar{X}_{ir})(X_{jrh} - \bar{X}_{jr}) \tag{1.2.1}$$

and $$n_1 b_{ij} = \sum_{r=1}^{k} N_r(\bar{X}_{ir} - \bar{X}_i)(\bar{X}_{jr} - \bar{X}_j) \tag{1.2.2}$$

where $$n_1 = k - 1 \text{ and } n_2 = \sum_{r=1}^{k} (N_r - 1) \tag{1.2.3}$$

(i) Hotelling's T_k^2-Statistic:

Hotelling (1947, 1950) gives a statistic T_k^2 to test the hypothesis of equality of k mean vectors and defines it in the classical case using a matrix (σ_{ij}) known or estimated on the basis of large samples as:

$$T_k^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij} \sum_{r=1}^{k} N_r(\bar{X}_{ir} - \bar{X}_i)(\bar{X}_{jr} - \bar{X}_j) \tag{1.2.4}$$

or $$T_k^2 = n_1 \operatorname{tr}\left[(\sigma^{ij})B\right] \tag{1.2.5}$$

where (σ^{ij}) is the inverse of (σ_{ij}) and $\bar{X}_i = \sum_{r=1}^{k} (N_r \bar{X}_{ir}) \Big/ \sum_{r=1}^{k} (N_r)$

The Studentized T_k^2 can be expressed in three different ways as follows:

$$T_k^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} w^{ij} \sum_{r=1}^{k} N_r(\bar{X}_{ir} - \bar{X}_i)(\bar{X}_{jr} - \bar{X}_j) \tag{1.2.6}$$

or $$T_k^2 = n_1 \operatorname{tr}(W^{-1}B) = n_2 \operatorname{tr}\left[(n_2 W)^{-1}(n_1 B)\right] \tag{1.2.7}$$

or $$T_k^2 = n_2 \sum_{i=1}^{\ell} (\phi_i) = n_2 \sum_{i=1}^{\ell} \left(\frac{\theta_i}{1 - \theta_i}\right) \tag{1.2.8}$$

where ϕ_i and θ_i are respectively the roots of the determinantal equations:

$$\left| n_1 B - \phi\, n_2 W \right| = 0 \tag{1.2.9}$$

and $$\left| n_1 B - \theta(n_1 B + n_2 W) \right| = 0 \qquad (1.2.10)$$

and where $$\ell = \text{Min.}(p, n_1) \qquad (1.2.11)$$

We have found another interesting expression of T_k^2 in terms of weighted Mahalanobis distances. It is given in the last section of Chapter Five. In Chapters Two and Four, we have made use of this statistic in forming clusters and in assigning an arbitrary population to one of the clusters.

The classical T_k^2 is known (Rao, 1952) to be distributed, under the null hypothesis, as central chi-square with $n_1 p$ D.F. In the case of non-centrality parameter $\tau_k^2 \neq 0$, the classical T_k^2 is non-central chi-square distributed with $n_1 p$ D.F., and the parameter τ_k^2 is defined as follows:

$$\tau_k^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij} \sum_{r=1}^{k} N_r(\mu_{ir} - \mu_i)(\mu_{jr} - \mu_j) \qquad (1.2.12)$$

where $$\mu_i = \sum_{r=1}^{k} (N_r \mu_{ir}) \Big/ \sum_{r=1}^{k} (N_r) \qquad (1.2.13)$$

The exact distribution of Studentized T_k^2 is not known in compact standard form. Ito (1956) has given, under the null hypothesis, its approximate formula as:

$$T_k^2 = \chi^2 + \frac{1}{2n_2} \left[\frac{p + n_1 + 1}{n_1 p + 2} \chi^4 + \cdots \right] + \cdots \qquad (1.2.14)$$

where χ^2 is central chi-square with $n_1 p$ D.F. The use of T_k^2 in multivariate analysis of variance (Anova) has been illustrated by Siotani (1958), who has constructed its tabular values for 5% and 1% significance levels for three or more dimensions.

(ii) Wilks Λ -Criterion:

Following the likelihood ratio method (Neyman and Pearson, 1928, 1931, and Pearson and Neyman, 1930), Wilks obtained a suitable extension of the univariate F-ratio in the form:

$$\Lambda = |n_2 W| \Big/ |n_2 W + n_1 B| \tag{1.2.15}$$

or alternatively as:

$$\Lambda = \prod_{i=1}^{\ell} (\rho_i) = \prod_{i=1}^{\ell} (1 - \theta_i) \tag{1.2.16}$$

where ρ_i and θ_i are respectively the roots of the determinantal equations

$$|n_2 W - \rho (n_2 W + n_1 B)| = 0 \tag{1.2.17}$$

and (1.2.10), where W and B are the usual mean products (M.P.) matrices.

Wilks (1932) and Nair (1939) have given the exact distribution of for $n_1 = 1, 2$ and any p, and for $p = 1, 2$ and any n_1 by comparing the moments of Λ with those of F-ratio. Bartlett (1934, 1938, 1947) suggested its useful approximation as follows:

$$-\left[(n_1 + n_2) - \tfrac{1}{2}(p + n_1 + 1)\right] \log_e \Lambda = \chi^2_{pn_1} + \frac{\gamma_2}{n_1^2}\left(\chi^2_{pn_1+4} - \chi^2_{pn}\right) + \cdots \qquad (1.2.18)$$

where $\gamma_2 = \frac{n_1 p}{48}(p^2 + n_1^2 - 5)$ and χ^2_f is central chi-square with f D.F.

We have made use of this approximate test in Chapter Two in testing for the over-all homogeneity of the species taken in the illustrative example.

More recently Bannerjee (1958) has been able to give the exact distribution of Λ in series form, but the tabular values are not yet available.

(iii) Wilks-Lawley U-statistic and Pillai's V-statistic:

There are two other statistics to test the homogeneity of k mean vectors due to Wilks-Lawley (1932, 1938) and Pillai (1954, 1956) defined respectively as:

$$U = |n_1 B| \Big/ |n_2 W + n_1 B| \qquad (1.2.19)$$

and

$$V = \mathrm{tr}\left[(n_2 W + n_1 B)^{-1}(n_1 B)\right] \qquad (1.2.20)$$

These can also be expressed respectively as follows:

$$U = \prod_{i=1}^{\ell}\left(\frac{\phi_i}{1+\phi_i}\right), \quad \text{or} = \prod_{i=1}^{\ell}(\theta_i) \qquad (1.2.21)$$

and

$$V = \sum_{i=1}^{\ell}\left(\frac{\phi_i}{1+\phi_i}\right), \quad \text{or} = \sum_{i=1}^{\ell}(\theta_i) \qquad (1.2.22)$$

where ϕ_i and θ_i are defined respectively as in (1.2.9) and (1.2.10).

These two statistics will be discussed further in Section 1.4.

When the hypothesis of the equality of mean vectors is rejected by the use of any of the above four statistics, three problems arise: (i) determining the confidence region for the population parameter corresponding to the statistic used to test the hypothesis of equality of mean vectors; (ii) to find groups or clusters of populations having like mean vectors; and (iii) to classify an arbitrary individual as belonging to one of the k normal populations, or an arbitrary population as belonging to one of the clusters.

We have dealt with the first problem in Chapter Five and have discussed the method of giving a confidence region to τ_k^2. Finally, we have demonstrated the method by taking a particular case with $k = 2$, $p = 4$, $n_1 = 4$, $n_2 = 29$.

For forming clusters of populations with like mean vectors, Rao (1948, 1955) and Tocher (1948) have given a subjective approach which is not based on probabilistic considerations. Working on the principle of minimum average distance, they have suggested a technique based on the criterion that 'any two groups belonging to the same cluster should at least on the average show a smaller D_2^2 than those belonging to different clusters'.

Rao's Graphical Approach

A graphical approach to the same problem has been given by Rao on the basis of significant discriminant scores or canonical variates. Since we have also made extensive use of significant discriminant scores in reducing T_k^2 and D_2^2 and likelihood functions to convenient and easy workable forms, we shall first discuss how Rao obtained these scores and then his graphical approach.

Rao (1952), like Fisher, takes the linear combinations $\ell_{i1}X_1 + \dots + \ell_{ip}X_p$ (i = 1, 2, ..., p) and maximizes the ratio:

$$\phi = \left[\sum_{i=1}^{p} \sum_{j=1}^{p} \ell_i \ell_j b_{ij}\right] \Big/ \left[\sum_{i=1}^{p} \sum_{j=1}^{p} \ell_i \ell_j w_{ij}\right] \qquad (1.2.23)$$

and gets the system of equations,

$$L(BW^{-1}) = \underset{\sim}{\Phi} L \qquad (1.2.24)$$

where $\underset{\sim}{\Phi}$ (p x p) is a diagonal matrix with diagonal elements ϕ_i (i = 1, 2, ..., p) and L(p x p) is the matrix of coefficients of the discriminant functions. Without losing generality we can suppose that $\phi_p \le \phi_{p-1} \le \dots \le \phi_2 \le \phi_1$ and test their significance by Bartlett's modified approximate formula (1.2.18) (Rao, 1952) given by:

$$\left[(n_1 + n_2) - \tfrac{1}{2}(p + n_1 + 1)\right] \log_e\left(1 + \frac{n_1}{n_2}\phi_i\right) \doteq \chi_i^2 \qquad (1.2.25)$$

where χ_i^2 is central chi-square with $(p + n_1 + 1) - 2i$ D.F.

By repeated use of formula (1.2.25), he gets a set of, say, p' ($\leq p$) significant eigenvalues and hence the corresponding p' significant discriminant functions. Placing their p' vectors of coefficients row-wise, so that the first row should correspond to the largest eigenvalue, the second to the second largest and so forth, he forms a matrix $K(p' \times p)$. Denoting further $\bar{X}^t(k \times p)$ as the matrix of k sample mean vectors, he gets the matrix $\bar{Y}^t(k \times p')$ of p' significant discriminant scores as:

$$\bar{X}^t K^t \qquad (1.2.26)$$

Note: To find $\emptyset$ $(p \times p)$ and $L(p \times p)$ as the solutions of (1.2.24), we can first symmetrize BW^{-1} by the procedure suggested by Nash and Jolicoeur (unpublished, 1959) which we have summarized in Appendix A and then apply the familiar technique due to Jacobi, which can be used on high speed computers.

Thus, knowing the significant discriminant scores, Rao then suggests plotting them in a space whose dimensionality is equal to the number of significant eigenvalues. If there are only two significant eigenvalues, there is no difficulty in having the plane representation of the points in which the closeness of the points (populations) with one another can be easily visualized. But it becomes difficult in the case of three or more eigenvalues. Rao

(1948) in such situations suggests having pair-wise plane representations of the points and then seeing (of course relying mostly on most significant scores) which of the populations lie close to one another.

In our discussion of the procedure for forming clusters in Chapter Two, we have sought a departure from Rao's and Tocher's subjective approach and have instead suggested two stages. Stage I is a sort of prediction by making use of Rao's graphical approach. In Stage II we give first our own definition of a cluster. Then we propose to correct the prediction by three alternative statistics where in each, unlike Rao and Tocher, we are able to attach probability to our decision. The first alternative has been discussed with an illustration in Chapter Two and the remaining two briefly in Chapter Three. Our working criteria for all three cases are multivariate analogues of previous criteria used in univariate analysis of variance (Anova) for forming clusters of like groups. The choice of the level of significance is that proposed by Duncan. Therefore we will discuss briefly such procedures for the univariate problem.

Some of the methods of forming clusters of like groups in univariate Anova are the following: Fisher's least significant difference test, the Student-Newman-Keuls' range test, and more recently Scheffé's multiple F-test, Tukey's test based on allowances and his gap-straggler and variance test, Duncan's multiple range and F-tests based on degrees of freedom, and further extensions by Sawkin, Kramer, Hartley, and Roy and

Bose. A detailed explanation of these procedures with illustrations is provided by Federer (1955). Since we have generalized Duncan's approach to the multivariate case, we give below briefly what he did. Duncan made a two-way attack on the problem - first by the multiple range test and second by the multiple F-test. To avoid duplication we will not give the description of his range test, since its procedure, except for significant ranges, is just the same as the Stage I of his multiple F-test.

Duncan's Level of Significance

Duncan's multiple range test is similar to the Student-Newman-Keuls' test and his multiple F-test similar to that of Scheffé. The only difference between Duncan and the others has been in the choice of a level of significance. He proposes that the level of significance should increase with the increase of the number of means in a group whereas others have kept the same pre-assigned level of significance as in the case of k-means. He justifies himself by arguing that any increase in the later levels would result in the increase of type II error and thus suggests that the r-mean ($r = 2, 3, \ldots, k$) significance level $\mathcal{L}_r$, for a pre-assigned $\mathcal{L}$, be

$$\mathcal{L}_r = 1 - (1 - \mathcal{L})^{r-1} \quad (1.2.27)$$

$$r = 2, 3, \ldots, k$$

where $(r - 1)$ is the number of independent comparisons which can be specified among the r means.

Duncan's Multiple F-test

Duncan, in this test procedure, has made use of both the range test and F-test by setting again the level of significance based on D.F. as described above. According to Federer, Duncan's test procedure can be set up in three stages of which we will give the first two - the second being the most important for our purpose:

Stage I: The first stage, as pointed out earlier, is in fact just the multiple range test but with different significant ranges. The procedure is as follows:

(i) Compute the quantities $R'_r = \sqrt{2(k - 1)F_{\mathcal{L}_r}(r - 1, f)}$ for $r = 1, 2, \ldots, k$, where, for a pre-assigned $\mathcal{L}$, $\mathcal{L}_r$ is defined as (1.2.27) and f is the D.F. associated with the pooled error variance $\hat{s}_{\bar{x}}$.

(ii) Compute the quantities $R_r = R'_r \; s_{\bar{x}} \quad (r = 1, 2, \ldots, k)$.

(iii) Compute the differences between the ranked means.

(iv) Finally, compare these differences of the ranked means with R_r $(r = 1, 2, \ldots, k)$ and determine the group of like means by following the criterion: "The differences between any two means in a set of k means is significant provided the range of each and every subset which contains the given means is significant".

<u>Stage II</u>: Stage II is the correction of the prediction made in Stage I. The procedure for correction is summarized as:

(i) Compute the sum of squares among the combinations of the means bracketed together in the prediction.

(ii) Compute the least significant sum of squares

$$\lambda_r = \frac{1}{2} R_r^2 \quad (r = 1, 2, \ldots, k), \text{ and}$$

(iii) Correct the predicted groups by following the criterion: "The difference between any two means in a set of $k_1 (\leq k)$ means is significant provided the variance of each and every subset which contains the given means is significant according to an $\mathcal{L}_r$-level F-test where r is the number of means in the set".

As pointed out earlier, the third problem that can arise after the hypothesis of equality of mean vectors is rejected is to classify an individual as belonging to one of the k distinct normal p-variate populations or a population as belonging to one of the clusters. Assuming *a priori* that the individual, with measurements $(X_1, X_2, \ldots, X_p)$, does belong to one of the k populations, Rao (1948) computes, where we ignore the *a priori* probabilities, the linear discriminant scores for the rth $(r = 1, 2, \ldots, k)$ population as

$$\hat{L}_r = \sum_{j=1}^{p} \left(\sum_{i=1}^{p} w^{ij} \bar{X}_{ir} \right) X_j - \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \left(w^{ij} \bar{X}_{ir} \bar{X}_{jr} \right) \qquad (1.2.28)$$

and then suggests assigning the individual to the sth population if $\hat{L}_s$ is greater than every other $\hat{L}_r$ for $r\ (\neq s) = 1, 2, \ldots, k$.

We have taken up, in Chapter Four, the problem of assigning a population known to belong a priori to one of the clusters and have suggested two alternative procedures - the first similar to the L-functions and the second based on the statistic T_k^2. Finally, an illustrative example is given to demonstrate the theory.

1.3 Generalized Variance and its Moments

Wilks (1932) defines the generalized variance to be the determinant of variances and covariances and considers it to be a measure of the spread of the observations. He then presents the hth moment of the generalized variance in the null case as follows:

If S be the sample variance-covariance matrix with n D.F. and $\Sigma'(p \times p) = E(nS)$, then the hth moment of $|A|$ $(= |nS|)$ in the central case is given by Wilks (1932):

$$E\left[|A|^h \right] = 2^{ph} \prod_{i\ 1}^{p} \Gamma\left(\frac{n+1-i}{2} + h\right) \Big/ \Gamma\left(\frac{n+1-i}{2}\right) \qquad (1.3.1)$$

Further, let k_i^2 $(i = 1, 2, \ldots, p)$ be the real and non-negative roots of the determinantal equation:

$$\left| T - k^2 \Sigma' \right| = 0 \tag{1.3.2}$$

where $T = \left\| \sum_{r=1}^{k} (\mu_{ir} - \mu_i)(\mu_{jr} - \mu_j) \right\|$ and $\mu_i = \frac{1}{k} \sum_{r=1}^{k} \mu_{ir}$.

Assuming now $k_i^2 = 0$ $(i = 2, 3, \ldots, p)$ and $k_1^2 \neq 0$, Anderson (1946) gives the h-th moment of $|A|$ in the non-central linear case as:

$$E\left[|A|^h\right] = 2^{ph} \exp(-\tfrac{1}{2} k_1^2) \prod_{i=1}^{p-1} \frac{\Gamma(\frac{n-i}{2} + h)}{\Gamma(\frac{n-i}{2})} \sum_{j=0}^{\infty} \frac{k_1^{2j}}{2^j j!} \frac{\Gamma(\frac{n}{2} + h + j)}{\Gamma(\frac{n}{2} + j)} \tag{1.3.3}$$

Making use of these moments we have found in Chapter Six the distribution of the determinant of the sum of products (S.P.) matrix A in the non-central linear case for some particular values of p, namely p = 2, 3, and 4.

1.4 Problem of Eigenroots of Certain Determinantal Equations

It is shown in Section (1.2) that, for testing the hypothesis of the equality of mean vectors of samples drawn from k p-variate normal populations, the four statistics (1.2.8), (1.2.16), (1.2.21) and (1.2.22) can all be expressed as functions of the roots of certain determinantal equations. There are two other tests of hypotheses due to Roy (1939) and Hotelling (1936) which also result in the roots of the same type of determinantal equations with, of course, the use of different matrices.

Roy's effort (1939) to seek a statistic to test the equality of dispersion matrices Σ_1 and Σ_2 of two p-variate normal populations finally led him, applying the same technique as Fisher's (1936), to test, instead of one, p Studentized statistics $\lambda_1, \lambda_2, \ldots, \lambda_p$ (all positive in this case) which are the p roots of the determinantal equation in λ :

$$\left| n_1W_1 - \lambda n_2W_2 \right| = 0 \tag{1.4.1}$$

or alternatively, by substituting $\theta_i = \dfrac{\lambda_i}{1+\lambda_i}$ (i=1,2,...,p), the roots of

$$\left| n_1W_1 - \theta(n_1W_1 + n_2W_2) \right| = 0 \tag{1.4.2}$$

where n_1W_1 and n_2W_2 are the S.P. matrices estimated from the respective samples.

To test the hypothesis of the independence of two sets of variates, such as p measurements of physical characteristics such as lengths and breadths of skulls and q measurements of mental characteristics such as scores on intelligence tests, Hotelling (1936) considered the determinantal equation of the roots $\theta_i (i = 1, 2, \ldots, p)$ and $(p \leq q)$ of

$$\left| W'_{pq} W'^{-1}_{qq} W'_{qp} - \theta W'_{pp} \right| = 0 \tag{1.4.3}$$

or $$\left| W'_{pq} W'^{-1}_{qq} W'_{qp} - \theta \left[(W'_{pp} - W'_{pq} W'^{-1}_{qq} W'_{qp}) + W'_{pq} W'^{-1}_{qq} W'_{qp} \right] \right| = 0 \tag{1.4.4}$$

Here $W'_{pq} W'^{-1}_{qq} W'_{qp}$ and W'_{pp} are independent S.P. matrices with q and

$(N - q - 1)$ D.F. and N is the size of the sample of individuals drawn from a $(p + q)$-variate normal population with covariance matrix Σ. Further W'_{pp} is the S.P. matrix of the sample observations on the p-set of variates, W'_{qq} that on the q-set and W'_{qq} that between the observations on the p-set and those on the q-set.

Thus in multivariate Anova (Pillai, 1954) the three tests of hypotheses above, i.e. I, "equality of two dispersion matrices", II, "equality of the p-dimensional mean vectors", and III, "the independence between a p-set and q-set of variates" depend, when the respective hypotheses to be tested are true, only on the roots θ_i or ϕ_i $(i = 1, 2, \ldots, \ell)$ respectively of the determinantal equations

$$|A - \theta(A + C)| = 0 \tag{1.4.5}$$

and

$$|A - \phi C| = 0 \tag{1.4.6}$$

where A and C are independent S.P. matrices based on sample observations with n_1 and n_2 D.F. respectively and can be defined differently for different hypotheses.

The common standard form (Nanda 1948, Roy 1957) of the joint distribution of the eigenroots of (1.4.5), under the respective hypotheses, is

$$c(m,m,\ell) \prod_{i=1}^{\ell} \theta_i^m (1 - \theta_i)^n \prod_{i=2}^{\ell} \prod_{j=1}^{i-1} (\theta_i - \theta_j) \prod_{i=1}^{\ell} d\,\theta_i \tag{1.4.7}$$

for $0 \le \theta_1 \le \theta_2 \le \ldots \le \theta \le 1$ and ℓ defined as in (1.2.11), where

$$c(m,m,\ell) = \frac{\pi^{\ell/2} \prod_{i=1}^{\ell} \Gamma\left(\frac{2m+2n+\ell+i+1}{2}\right)}{\prod_{i=1}^{\ell} \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma\left(\frac{2n+i+1}{2}\right) \Gamma\left(\frac{i}{2}\right)} \qquad (1.4.8)$$

ℓ, m, n can be different in the different situations defined below in (1.4.12) and (1.4.13).

The common standard form (Hsu 1939) of the joint distribution of the eigenroots of (1.4.6), under the respective hypotheses, is

$$c(m,n,\ell) \prod_{i=1}^{\ell} \phi_i^m (1+\phi_i)^{-(m+n+\ell+1)} \prod_{i=2}^{\ell} \prod_{j=1}^{i-1} (\phi_i - \phi_j) \prod_{i=1}^{\ell} d\,\phi_i \qquad (1.4.9)$$

for $0 \leq \phi_1 \leq \phi_2 \leq \dots \leq \phi_\ell < \infty$

where ℓ, $c(m,n,\ell)$ are defined respectively as in (1.2.11) and (1.4.8) and ℓ, m, n can be different in the different situations defined below in (1.4.12) and (1.4.13).

Finally, Nanda (1948) gives the limiting form of (1.4.7) by setting $\theta_i = \frac{c_i}{n}$ and then letting n tend to infinity. The limit is

$$K(\ell,m) \prod_{i=1}^{\ell} c_i \exp\left[-\sum_{i=1}^{\ell} c_i\right] \prod_{i=2}^{\ell} \prod_{j=1}^{i-1} (c_i - c_j) \prod_{i=1}^{\ell} d\,c_i \qquad (1.4.10)$$

where $K(\ell,m) = \pi^{\ell/2} \Big/ \prod_{i=1}^{\ell} \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma\left(\frac{i}{2}\right) \qquad (1.4.11)$

and ℓ is the same as in (1.211). Again ℓ, m assume different values defined below in different cases.

Finally, for the three tests of hypotheses I, II and III, we can sum up the values of ℓ, m, n for respective hypotheses as:

$$\text{I.}\quad \ell = p,\ m = \tfrac{1}{2}(n_1 - p - 1),\ n = \tfrac{1}{2}(n_2 - p - 1), \qquad (1.4.12)$$

$$\text{II.}\quad \text{If } p \leq n_1,\ \ell = p,\ m = \tfrac{1}{2}(n_1 - p - 1),\ n = \tfrac{1}{2}(n_2 - p - 1)$$

$$\text{If } p > n_1,\ \ell = n_1,\ m = \tfrac{1}{2}(p - n_1 - 1),\ n = \tfrac{1}{2}(n_2 - p - 1) \qquad (1.4.13)$$

III. Same as II

No great headway has been made so far in finding the distributions of the various statistics we have discussed. The exact or approximate distributions of two statistics T_k^2 and Λ have already been discussed in Section (1.2). Below is the brief account of the other statistics:

Roy (1943) proposed the statistics - largest, smallest or intermediate eigenroots of the determinantal equation (1.4.5) to test hypothesis I, II and III. Roy (1943) and Nanda (1948) have both worked out the distributions both for the limiting and non-limiting cases. Their tabular values have been given by Pillai (1957) for the cases $\ell = 2(1)5$, $m = 0(1)4$ and $n = 5$ to 1,000 both at 5% and 1% significance values.

Pillai (1954, 1955, 1959) has succeeded in giving an approximation to his statistic V defined in (1.2.22) and has been able to tabulate it for $\ell = 2(1)5$, $m = -.5(.5)5(5)80$ and $n = 5(5)80$. Nanda (1950) has also given its exact distribution for the special case when $m = 0$.

We have also been able to work out the exact distributions of various statistics for certain special cases in Chapter Seven. We have been able to give the distributions of all the statistics for the cases $\ell = 2, 3$ in the form of definite integrals which can be easily evaluated by some numerical method. The limiting distribution of Roy's statistics by another method of integration have been found and particular cases evaluated. Lastly the limiting distribution of Wilks-Lawley U-statistic for the cases $\ell = 2, 3$ and 4 has also been found.

1.5 Note on Analysis of Variance

Under both the Models I and II (Eisenhart) of Anova one is faced with two types of situations - firstly when the cell frequencies are equal and secondly when they are unequal. These cases are usually called balanced and unbalanced respectively.

Balanced Anova

For tests of significance in both univariate and multivariate balanced Anova of Model I and II and further for finding confidence regions again in both univariate and multivariate balanced Anova of Model I, there is not much difficulty. One can refer for such univariate problems to the various standard books, e.g. by Federer, Fisher, Anderson and Bancroft, Bennett and Franklin, Snedecor, Kempthorne and others, whereas for the multivariate problems sufficient material has been developed by Roy and Bose (1953), Roy (1955, 1956), Roy and Gnanadesikan (1959, I and II), Tukey (1949), Bartlett (1934, 1938, 1947), Kempthorne (1952), Rao (1948) and others.

The real difficulty arises in both univariate and multivariate problems when, in Model II, one is finding the confidence regions for the complex estimates (Satterthwaite, 1946) of the variance components, since in that case their corresponding distributions are not known. To overcome this difficulty in univariate problems various methods, approximate or otherwise, have been suggested. The more prominent amongst them are those due to Satterthwaite (1941, 1946), Brose (1950), Fisher (1935), Roy (1954a, 1954b, 1956), Roy and Bose (1953), Roy and Gnanadesikan (1957, 1959 I and II), Cornfield (1953), Ramachandran (1956) and Grayball, Morton and Godfrey (1956). Since we have made use of Satterthwaite's technique in our work in Chapter Eight, we briefly summarize what he did while finding the distribution of complex estimates:

<u>Satterthwaite's Procedure</u>

Let V_i be the mean squares independently distributed as $\lambda_i \chi_i^2$, where χ_i^2 is central chi-square with f_i D.F. The procedure is to approximate $\sum\limits_i (a_i V_i)$, a_i being constants, by $\chi_f^2 \; \frac{\sigma^2}{f}$, f being chosen so that the first two moments of the former are equal to those of the latter.

Therefore,

$$E\left[\sum_i (a_i V_i)\right] = \frac{\sigma^2}{f} E(\chi_f^2) = \frac{\sigma^2}{f} f = \sigma^2 \qquad (1.5.1)$$

and

$$E\left[\sum_i (a_i V_i) - E\left(\sum_i (a_i V_i)\right)\right]^2 = \frac{\sigma^4}{f^2}\, 2f = \frac{2\sigma^4}{f} \qquad (1.5.2)$$

From (1.5.1) and (1.5.2) we have:

$$f = 2\left[\sum_i E(a_i V_i)\right]^2 \Big/ \sum_i \left[a_i E(V_i - E(V_i)\right]^2 = \left[\sum_i (a_i \sigma_i^2)\right]^2 \Big/ \sum_i \left(\frac{a_i^2}{f_i} \sigma_i^4\right)$$

Since σ_i^2 are not known, he suggests to substitute for them their respective estimates and gets:

$$f \doteq \left[\sum_i (a_i V_i)\right]^2 \Big/ \sum_i \left(\frac{a_i^2}{f_i} V_i^2\right) \qquad (1.5.3)$$

It is again unfortunate that very little has been accomplished in analogous multivariate problems. Roy and Gnanadesikan (1959, I and III) have recently been able to give a lead, but their approach is under the very restrictive assumptions of $\Sigma_i(p \times p) = \sigma_i^2 \; \Sigma(p \times p)$, i.e. of proportional dispersion matrices, proposed usually (Federer, 1951) for certain types of genetical problems, where Σ_i is the covariance matrix due to the ith factor.

Unbalanced Anova

The problem is considerably complicated for both the cases of univariate and multivariate unbalanced Anova especially of Model II. In the univariate balanced case the mean squares were independent and distributed independently as chi-square but the situation now is worsened by the fact that the mean squares are not orthogonal and hence are not distributed as central chi-squares. They are in fact distributed (Anderson and Bancroft, 1952) as sums $\sum_r (\lambda_r \chi_r^2)$ where λ_r are functions of the variance components and the number of observations, and each χ_r^2 is a central chi-square with 1 D.F. Since the λ_r are distinct, we cannot apply the additive property of independent chi-squares to the sums $\sum_r (\lambda_r \chi_r^2)$.

Similarly for corresponding multivariate situations, the M.P. matrix is no longer distributed as a Wishart matrix but, as proved in Chapter Eight, is distributed as a sum $\sum_r (W_r)$ of independent Wishart matrices W_r, each distributed as $W\left[\Sigma_r, 1\right]$. If these Wishart matrices W_r had the common corresponding parameters, i.e. $\Sigma_1 = \Sigma_2 = \ldots = \Sigma$ (say), then there would be no problem. We could then simply use the additive property of independent Wishart matrices and would get another Wishart matrix.

We have attempted, in Chapter Eight, to find the approximate distribution of mean squares or M.P. matrices. We have determined first the values of the above quoted quantities λ_r and Σ_r and then have applied Satterthwaite's technique in approximating the distributions of sums $\sum_r (\lambda_r \chi_r^2)$ and $\sum_r (W_r)$.

CHAPTER TWO

ANALOGUES OF DUNCAN'S PROCEDURE IN FORMING CLUSTERS IN MULTIVARIATE ANOVA

2.1 As already stated, we sometimes come across the following type of problem in anthropology and the biological sciences, namely this, certain multivariate populations are found to be distinct, and we want to find out which populations are most nearly alike and which are least alike. To do this, we propose to extend Duncan's procedure of the multiple comparisons' tests used in univariate Anova and to seek a departure from Rao's and Tocher's subjective approach. We give below first a different definition of the cluster and then, after clearing some preliminaries, suggest a procedure based on probabilistic considerations.

Definition of a cluster:

"A cluster of populations is a group of populations having the same vector mean."

2.2 Preliminaries and Procedure

Suppose we are given k p-variate normally distributed populations assumed to have the same dispersion matrix Σ. Let X_{irh} ($i = 1, 2, \ldots, p$; $r = 1, 2, \ldots, k$ and $h = 1, 2, \ldots, N_k$) be the observation of the ith trait on the hth individual from the rth sample of size N_r drawn from the rth population. Further, let B and W be the between and within independent S.P. matrices, with n_1 and n_2 D.F. respectively, computed on the basis of k p-variate samples defined respectively in (1.2.2) and 1.2.1.).

Suppose also the hypothesis of homogeneity of mean vectors of the populations has been rejected by the use of Wilks-Λ statistic (1.2.15) and Bartlett's approximation to its probability (1.2.18).

Knowing thus that the populations are heterogeneous, we proceed to form clusters. Before doing this we make the following preliminary remarks:

Since we have made frequent use of both Studentized D_2^2 and T_k^2, it would be appropriate to modify them to an easily workable form. To do this we derive first the significant discriminant scores discussed already in Section (1.2). We sum the matter up briefly in the following steps:

(a) Find, by the method given in Appendix A, a nonsingular matrix L(p x p) and the diagonal matrix $\underline{\emptyset}$(p x p) as the solution of (1.2.24).

(b) Test the significance of $\emptyset_i$ by the formula (1.2.25). Without losing generality suppose the first $p'(\leq p)$ of the p eigenroots are significant and the last $(p - p')$ are non-significant.

(c) Discard the last $(p - p')$ eigenroots, and hence the corresponding eigenvectors, because they in fact account for random variation.

(d) Obtain the matrix $K(p' \times p)$ of the eigenvectors whose first row corresponds to the largest eigenroot, its second to the second largest and so forth to the smallest one left, namely the p'th.

(e) Taking $\overline{X}^t(k \times p)$ to be the matrix of k sample mean vectors, using columns for characters and rows for sub-population samples, compute the matrix $\overline{Y}^t(k \times p')$, defined as in (1.2.26), which is the matrix of significant discriminant scores, and whose first column gives the discriminant score corresponding to the largest eigen value, the second column to the second largest, and so forth. With these scores, the Studentized statistics D_2^2 and T_k^2 reduce from (1.1.2) and (1.2.6) respectively to:

$$D_2^2 = \sum_{i=1}^{p'} (\overline{Y}_{i1} - \overline{Y}_{i2})^2 \tag{2.2.1}$$

and

$$T_k^2 = \sum_{i=1}^{p'} \sum_{r=1}^{k} N_r(\overline{Y}_{ir} - \overline{Y}_i)^2 \tag{2.2.2}$$

where

$$\overline{Y}_i = \sum_{r=1}^{k} (N_r \overline{Y}_{ir}) \Big/ \sum_{r=1}^{k} (N_r) \tag{2.2.3}$$

<u>Note</u>: The same technique works for the corresponding classical D_2^2.

<u>Statistic Used</u>

For testing the hypothesis of equality of the mean vectors involved in a cluster we suggest an analogue of Duncan's Stage 2 of the multiple F-test. He computed the variance of the means involved in a predicted group of like means and tested it against his least significant sums of squares with type I error based on D.F. In the multivariate situations as the analogue of his "variance of the means involved in a cluster" we propose an expression $T^2_{k_1}$, where $k_1 (\leq k)$ is the number of sample mean vectors of the populations involved in the predicted cluster. The distribution of $T^2_{k_1}$, under the null hypothesis, is known in the classical case to be central chi-square with $p(k_1-1)$ D.F. and in the Studentized case to be an asymptotic expression involving chi-squares as shown in (1.2.14), where again the D.F. for chi-square is $p(k_1-1)$.

<u>Note</u>: It should be noted that we have used p instead of p' for defining degrees of freedom, since (Rao, 1948) the effect of all p correlated variates has been taken care of by the discriminant scores.

<u>Level of Significance or Protection Level</u>

In selecting the level of significance or protection level we again propose to follow Duncan. In order to keep the two types of errors well balanced, we shall let the type I error increase with the increase in the number of populations in a cluster. Thus with k_1 ($\leq k$) populations in a cluster, for a pre-assigned significance level $\mathcal{L}$, we shall fix the level of significance to be:

$$\mathcal{L}_{k_1} = 1 - (1 - \mathcal{L})^{k_1 - 1} \qquad (2.2.4)$$

Preparation of Tables for the new Levels

Since the statistic T_k^2 involves a central chi-square for both the Studentized and classical cases, we need to modify the central chi-square tables for both 5% and 1% significance levels and also for different values of $k = 2(1)(20)$. To do it, we proceed as follows:

The table 1 below gives the various significance levels $1 - \gamma_k$ ($= \mathcal{L}_k$ or $= Q$) for $k = 2(1)(20)$, for pre-assigned significance levels 5% and 1%.

Again table 1 gives under the column X the normal variates X corresponding to each level of significance Q. X has been used in the computation of tabular values of chi-squares. To compute these X values, a linear interpolation formula:

$$X = X_0 + \frac{f(X) - f(X_0)}{f(X_1) - f(X_0)} (X_1 - X_0) \qquad (2.2.5)$$

has been used where X is the normal variate to be determined between the two known normal variates X_0 and X_1 and where also $f(X)$ ($= Q$) is a known quantity and $f(X_0)$ and $f(X_1)$, corresponding respectively to X_0 and X_1, are taken from table I of Hartley and Pearson, 1954.

Table 1

	5%			1%		
k	$\gamma_k = 1 - Q$	$1 - \gamma_k = Q$	X	$\gamma_k = 1 - Q$	$1 - \gamma_k = Q$	X
2	0.9500	0.05000000	1.64490	0.9900	0.0100	2.32630
3	0.9025	0.09750000	1.29600	0.9801	0.0199	2.05584
4	0.85737500	0.14262500	1.06860	0.970299	0.029701	1.88523
5	0.81450625	0.18549375	0.89450	0.96059601	0.03940399	1.75766
6	0.77378094	0.22621906	0.75136	0.95099005	0.04900995	1.65455
7	0.73509189	0.26490811	0.62830	0.94148015	0.05851985	1.56729
8	0.69833729	0.30166271	0.51960	0.93065349	0.06934651	1.48068
9	0.66342043	0.33657957	0.42180	0.92134695	0.07865305	1.41421
10	0.63024941	0.36975059	0.33250	0.91213348	0.08786652	1.35403
11	0.59873694	0.40126306	0.25008	0.90301215	0.09698785	1.29891
12	0.56880009	0.43119991	0.17330	0.89398202	0.10601798	1.24800
13	0.54036008	0.45963992	0.10140	0.88504220	0.11495780	1.20058
14	0.51334208	0.48665792	0.03350	0.87619178	0.12380822	1.15617
15	0.48767497	0.51232503	-0.03090	0.86742986	0.13257014	1.11434
16	0.46329122	0.53670878	-0.09220	0.85875556	0.14124444	1.07476
17	0.44012666	0.55987334	-0.15065	0.85016800	0.14983200	1.03717
18	0.41812033	0.58187967	-0.20670	0.84166632	0.15833368	1.00134
19	0.39721431	0.60278569	-0.26060	0.83324966	0.16675034	0.96710
20	0.37735359	0.62264641	-0.31244	0.82491716	0.17508284	0.93428

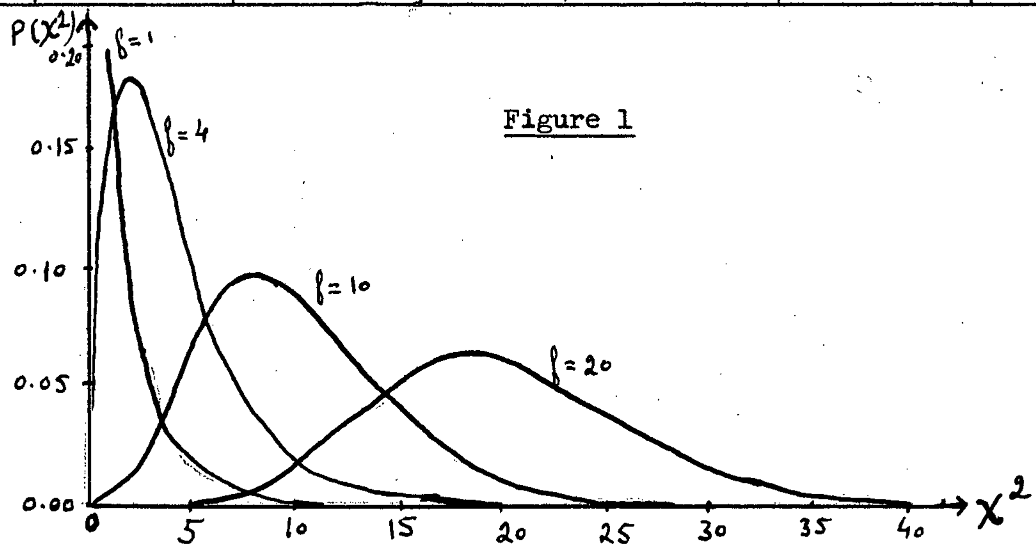


Figure 1

Further, the study of the behaviour of the chi-square curves (Fig. 1) for various degrees of freedom is very helpful.

From Fig. 1 it is obvious that with the increase of degrees of freedom, chi-square curves tend to be symmetrical while for the smaller degrees of freedom they lack symmetry. Thus direct linear interpolation of (1 - Q) values along with the corresponding chi-square values (especially for the smaller degrees of freedom) cannot be expected to lead us to accurate results. To keep the accuracy for the smaller degrees of freedom and also the uniformity of method, we have decided to use, instead of (1 - Q) values, the corresponding normal variates X shown in table 1.

Then, the Aitken's Iterative interpolation formula has been used to compute the tabular chi-square values. We give below a demonstration of the method for 3 D.F. against the normal value 2.055844. Then some of the values have been actually computed both by the use of (1 - Q) values and the corresponding X-variates and have been listed below in table 2. A brief glance over the table 2 will show that as the degrees of freedom increase, both methods lead approximately to the same result.

Demonstration of the Method

Let D.F. = 3, $\bar{X}$ = 2.055844 and χ^2 corresponding to $\bar{X}$ is to be found.

X	χ^2				$X - \bar{X}$
1.6449	7.81473				-0.410944
1.9600	9.3484	9.81489			-0.095844
2.3263	11.3449	9.94373	9.84860		0.270456
2.5758	12.8381	10.03228	9.84872	9.84846	0.519956

Thus adopting Aitken's iterative method for interpolation, the new chi-square values have been computed at various significance levels α_k for k = 2(1)20 and D.F. = 1(1)30(10)100 for pre-assigned α = .05 and .01. We record them for use in Appendices C and D respectively.

Table 2

D.F.	1 - Q	χ^2-corresponding to Q-values	X-normal variates	χ^2-corresponding to normal variates
3	.9801	9.71768	2.055644	9.84846
10	.9703	19.88597	1.885233	19.95269
25	.9703	39.90252	1.885233	39.92268

Finally, to find in the Studentized case the tabular T_k^2 values for any k, we have to use the formula (1.2.14) and substitute in it the newly computed chi-square values with n, p D.F. n_1 and n_2 are the degrees of freedom respectively for between and within independent covariance matrices and p is the number of characters. Since our illustration which is presented for demonstration concerns the studentized T_k^2, its tabular values needed for the purpose for $k = 2(1)5$, $n_1 = 1(1)4$, $p = 4$, and $n_2 = 29$ at 5% and 1% significance levels are tabulated approximately and presented below in table 3.

Table 3

$k=(n_1+1)$	D.F.$=p(k-1)$ $=n_1p$	$\chi^2(.05)_k$	$\chi^2(.01)_k$	$T^2(.05)_k$	$T^2(.01)_k$
2	4	9.4877	13.2767	12.1371	18.2030
3	8	13.4428	18.1825	16.7783	24.0936
4	12	17.1889	22.7746	21.7064	29.9100
5	16	20.8200	27.1912	25.6131	35.6187

Note: The tabular T_k^2 values have been computed on the assumption that terms involving the third and higher powers of $\frac{1}{n_2}$ are negligible. In fact they may affect the fourth significant figure.

2.3 The Proposed Stages for Forming Clusters

We propose two stages for the purpose. Stage I comprises three steps wherein we predict the possible clusters. Stage II then corrects the predictions on some probabilistic basis. So far three alternative methods have been proposed for Stage II. The first has been discussed with illustration in this very Chapter and the other two will be described in Chapter Three. The methods are as follows:

(i) The Duncan-Hotelling test.

(ii) The 'Extreme Distance from the Mean' - E-test.

(iii) The 'Largest Distance' - R-test.

Stage I: Prediction

Step 1: Compute $\binom{k}{2}$ Mahalanobis distances by the formula (2.2.1) between all the pairs of k populations and set up the table of distances, where the distances of each population from the remaining ones are arranged in order of increasing magnitude. Such a table (like Table 7) will help us to visualize which of the populations are closer to a particular one and which are farther away.

Step 2: Represent graphically the significant discriminant scores of each population. For $p' > 2$, they should be represented pair-wise on plane graph paper. Relying largely on the plane representations of the most significant discriminant scores, visualize which of the populations cluster together and which of them lie farther apart.

Step 3: Step 3 deals with the prediction of the clusters on the basis of the first two steps. Keeping in view the table of distances and the graphic plane representations, estimate roughly the 'would be' clusters - closeness being the only criterion for the populations to form a predicted cluster. The following two points are worth noting:

(i) That a wide range should be allowed to the clusters since giving a narrow range might result in the loss of a population lying actually in the cluster.

(ii) That overlappings should be allowed since sometimes one is uncertain as to whether to include one (or more) population(s) in one or the other cluster(s). In all such cases it is advisable to include the doubtful cases in all the neighbouring ones.

Stage II: Correction by the Duncan-Hotelling Test

No generality is lost if we explain the procedure for only one predicted cluster having k_1 populations in following steps:

(i) Compute the statistic $T^2_{k_1}$ by the formula (2.2.2).

(ii) Compare the computed $T^2_{k_1}$ with the tabular $T^2_{\mathcal{L}_{k_1}}$ where $\mathcal{L}_{k_1}$ is already defined as in (2.2.4).

(iii) If $T^2_{k_1}$ is less than or equal to $T^2_{\mathcal{L}_{k_1}}$, all the k_1 populations are concluded to form a cluster. Otherwise, split the k_1 populations into k_1 sets of $(k_1 - 1)$ populations each.

(iv) Compare the computed $T^2_{(k_1-1)}$ values for each of the k_1 sets with the tabular $T^2_{\alpha_{k_1-1}}$. Of these some may be significant and some may not be. Those non-significant will yield clusters with the corresponding number of populations involved in them. Those for which $T^2_{k_1-1}$ values are significant are further split into $(k_1 - 1)$ sets of $(k_1 - 2)$ populations each and their corresponding $T^2_{k_1-2}$ values are compared with the tabular $T^2_{\alpha(k_1-2)}$. In this way the process is continued till we arrive at the clusters of the type defined.

Thus the working criterion analogous to Duncan's can be presented as: "A group of k_1 populations will form a cluster if $T^2_{k_1}$ computed for the mean vectors of the k_1 populations is non-significant and also the T^2 of each and every set of populations of which the k_1 populations form a subset is significant according to α_r-level T^2_r-test for some pre-assigned α , where r is the number of populations in the set."

<u>Note</u>: The above procedure is for the Studentized case. In the classical case the procedure is the same except for the use of tabular chi-square values in place of T^2_r-values.

2.4 Demonstration of the Above Procedure by an Example

To demonstrate the theory we present below an example where the samples have been drawn on the basis of nested sampling:

Description of Data

Data has been taken from the 'Forest Products Laboratory Division, Forestry Branch, Department of Northern Affairs and National Resources, Vancouver, B.C., Canada'. Shipments of logs of various species of trees from various localities of Canada were received. The interest lies in comparing the species on the basis of static bending properties. For this purpose the following six measurements were taken at several locations in each tree:

X_1: Modulus of elasticity;

X_2: Work to the maximum limit;

X_3: Fibre strength at proportional limit;

X_4: Modulus of rupture;

X_5: Specific gravity at oven dry;

and X_6: Work to the proportional limit.

Note: While finding the values of the determinants of the S.P. matrices to be used for tests of significance, it was found that they came out to be zeros, which enabled us to conclude that the variables were functionally dependent. The fact was actually verified when the physical interpretation was known. The last two variables X_5 and X_6 were found to be functionally dependent on the first four X_1 , X_2 , X_3 , and X_4 . We thus discarded X_5 and X_6 and continued our work on the variables X_1, X_2, X_3, and X_4.

The species taken for the purpose are listed as follows:

(1) Yellow cedar, (2) Lodge pole pine, (3) Western larch, (4) Western yellow pine, (5) Western white pine, (6) Western white spruce, (7) Sitka spruce, (8) Amabilis fir, (9) Western hemlock, (10) Engelman spruce, (11) Western red cedar, (12) Coast mature Douglas fir, (13) Interior mature Douglas fir, and (14) Coast second growth Douglas fir.

Note: In what follows we will call each species by its corresponding number instead of specifying each time its name.

Description of the Model of Nested Sampling

We have the mixed model of nested sampling - with fixed species and random localities and locations on trees. Further, the number of localities and locations is not uniform in all cases.

Let $X_{ihjt\ell}$ be the observation of the ith character on the ℓ th location of the t-th tree belonging to the jth locality of the hth species. In place of observation $X_{ihjt\ell}$ we were provided with the means $\bar{X}_{ihjt.}$ along with the corresponding number of locations. The model for such data would be:

$$\underline{\bar{X}}_{hjt} = \underline{\mu} + \underline{\xi}_h + \underline{\eta}_{j(h)} + \underline{\delta}_{t(hj)} + \underline{\bar{e}}_{hjt} \tag{2.4.1}$$

where (1) $\underline{\bar{X}}_{hjt} \equiv (\bar{X}_{1hjt}, \ldots, \bar{X}_{4hjt})$ is a four dimensional mean vector of locations on the t-th tree from the jth locality of the hth species.

(2) $\underline{\mu}$ is the four dimensional mean vector of the populations and $\underline{\bar{X}}$ ....is the corresponding sample statistic.

(3) $\underline{\xi}_h$ is again the four dimensional hth species fixed effect, but for the sake of illustratbn we will take it as random, distributed normally with mean vector zero and covariance matrix Σ_{ξ}.

(4) $\underline{\eta}_{j(h)}$ is the four dimensional jth locality within hth species random effect, normally distributed with mean vector zero and covariance matrix Σ_{η} .

(5) $\underline{\delta}_{t(hj)}$ is the four dimensional t-th tree within hth species from the jth locality random effect, normally distributed with mean vector zero and covariance matrix Σ_{δ}

(6) $\underline{\bar{e}}_{hjt}$ is the four dimensional mean error vector of $\underline{e}_{hjt\ell}$ where each $\underline{e}_{hjt\ell}$ is random and normally distributed with mean vector zero and covariance matrix Σ_e

(7) Finally, $\underline{\xi}_h$, $\underline{\eta}_{j(h)}$ and $\underline{\delta}_{t(hj)}$ are independent and $E(\underline{\xi}_h) = E(\underline{\eta}_{j(h)}) = E(\underline{\delta}_{t(jh)}) = 0$.

Our model is just the analogue of the univariate model on nested sampling with unequal cell frequencies presented by Ganguli (1941). We follow his method for finding the coefficients of the expected M.P. matrices and end with the Table 4 of analysis of variance.

Table 4

Source of Variation	D.F.	S.P. Matrices	E(M.P. Matrices)
Species	13	A	$\Sigma_e + 13.381\,\Sigma_\delta + 81.27\,\Sigma_\eta + 246\,\Sigma_\varepsilon$
Localities within species	29	B	$\Sigma_e + 13.791\,\Sigma_\delta + 81.26\,\Sigma_\eta$
Trees within localities	217	C	$\Sigma_e + 13.372\,\Sigma_\delta$
Locations*	3248	D	Σ_e

* We do not have this row in our example since we have only the mean observations on each tree.

Here, $A = \left(\sum_h \left[n_{h\ldots}(\bar{X}_{i_1h\ldots} - \bar{X}_{i_1\ldots})(\bar{X}_{i_2h\ldots} - \bar{X}_{i_2\ldots})\right]\right)$

and $(\frac{A}{13}) = \begin{bmatrix} 10675527 & 38557 & 30971647 & 53851101 \\ 38557 & 305 & 156717 & 273320 \\ 30971647 & 156717 & 121780733 & 201012595 \\ 53851101 & 273320 & 201012595 & 343055522 \end{bmatrix}$;

$$B = \left(\sum_h \sum_j \left[n_{hj..}(\bar{X}_{i_1hj..} - \bar{X}_{i_1h\ldots})(\bar{X}_{i_2hj..} - \bar{X}_{i_2h\ldots})\right]\right)$$

and $(\frac{B}{29}) = \begin{bmatrix} 988308 & 1397 & 1936541 & 3167949 \\ 1397 & 21 & 6231 & 12721 \\ 1936541 & 6231 & 7821366 & 9469922 \\ 3167949 & 12721 & 9469922 & 15396656 \end{bmatrix}$;

and $C = \left(\sum_h \sum_j \sum_t n_{hjt.} (\bar{X}_{i_1 hjt.} - \bar{X}_{i_1 hj..})(\bar{X}_{i_2 hjt.} - \bar{X}_{i_2 hj..}) \right)$

and $(\frac{C}{217}) = \begin{bmatrix} 299438 & 558 & 593421 & 994326 \\ 558 & 7 & 1496 & 313 \\ 593421 & 1496 & 21011669 & 2575188 \\ 994325 & 313 & 2575188 & 4281234 \end{bmatrix}$.

Note: Referring back to Table 4 showing the analysis of variance, we notice that the corresponding coefficients in the formula for expected values are approximately equal. Thus we will treat it as a problem of nested sampling with equal numbers in the sub-classes and will proceed with the usual procedure of tests of significance.

To test the locality effect, Wilks' Λ -criterion was applied to the independent S.P. matrices B and C, with 29 and 217 D.F. respectively, and the locality effect was found to be significant by Bartlett's approximate test (1.2.18). Similarly the species effects were found to be significant upon taking the independent S.P. matrices A and B respectively with 13 and 29 D.F. From this we may conclude that the species are heterogeneous.

Start of the Problem

After concluding that the fourteen species are heterogeneous, we proceed to our main problem of forming clusters as follows:

We treat A and B respectively as the between and the within matrices with 13 and 29 D.F. and present below in table 5 the means of the characters of the species along with the corresponding sizes:

Table 5

Species No.	Size	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
1	264	1311	8.04	3664	6527
2	78	1285	5.35	2989	5657
3	158	1648	7.85	5002	8609
4	212	1137	5.45	3334	5718
5	324	1183	5.13	2877	4818
6	93	1113	5.76	2644	4831
7	380	1368	4.84	3078	5408
8	436	1341	5.57	2999	5460
9	200	1477	6.68	4150	6952
10	90	1251	5.36	3079	5662
11	207	1046	4.87	3102	5302
12	458	1650	6.97	4491	7548
13	348	1647	6.59	4099	7351
14	260	1583	7.41	4285	7697

We solve for $L(4 \times 4)$ and $\underline{\emptyset}(4 \times 4)$ the equation

$$L\left[\left(\frac{A}{13}\right)\left(\frac{B}{29}\right)^{-1}\right] = \underline{\emptyset}\, L \qquad (2.4.2)$$

by the method described in the Appendix A, and get:

$$L(4 \times 4) \equiv \begin{bmatrix} -0.001064336 & -0.158567182 & -0.000067004 & 0.000590678 \\ 0.001162664 & 0.369050519 & 0.000134634 & -0.000497864 \\ 0.001336923 & -0.069180685 & -0.000181461 & -0.000028873 \\ 0.000325045 & 0.023168191 & 0.000669918 & -0.000460445 \end{bmatrix}$$

and $$\bar{\phi}(4 \times 4) \equiv \begin{bmatrix} 25.94 & 0 & 0 & 0 \\ 0 & 11.84 & 0 & 0 \\ 0 & 0 & 5.65 & 0 \\ 0 & 0 & 0 & 1.65 \end{bmatrix}$$

Applying Bartlett's modified first approximation test (1.3.25) we test the significance of the eigenroots ϕ, i.e. of 25.94, 11.84, 5.65 and 1.65, and find 1.65 to be non-significant at the 5% level. Discarding thus the last row of L(4 x 4) which corresponds to 1.65, we get the matrix K(3 x 4). Now, if $\bar{X}^t(14 \times 4)$ be the matrix of mean vectors of species given in the last four columns of table 5, we get, by the formula (1.2.26) the matrix $\bar{Y}^t(14 \times 3)$ of significant discriminant scores which are presented below in Table 6 again, along with their corresponding sample sizes. (See Table 6, following page.)

Finally we compute the distances between the $\binom{14}{2}$ pairs of species of trees by the formula (2.2.1) and present them in Table 7 - called "Table of Distances", arranging the distances of each population from the remaining ones in order of increasing magnitude. (See Table 7, page 47.)

Also we plot these points pair-wise, i.e. $(\bar{Y}_1, \bar{Y}_2)$, $(\bar{Y}_1, \bar{Y}_3)$ and $(\bar{Y}_2, \bar{Y}_3)$ on the plane graphs which are shown respectively in Fig.2, Fig.3, and Fig.4.

Table 6

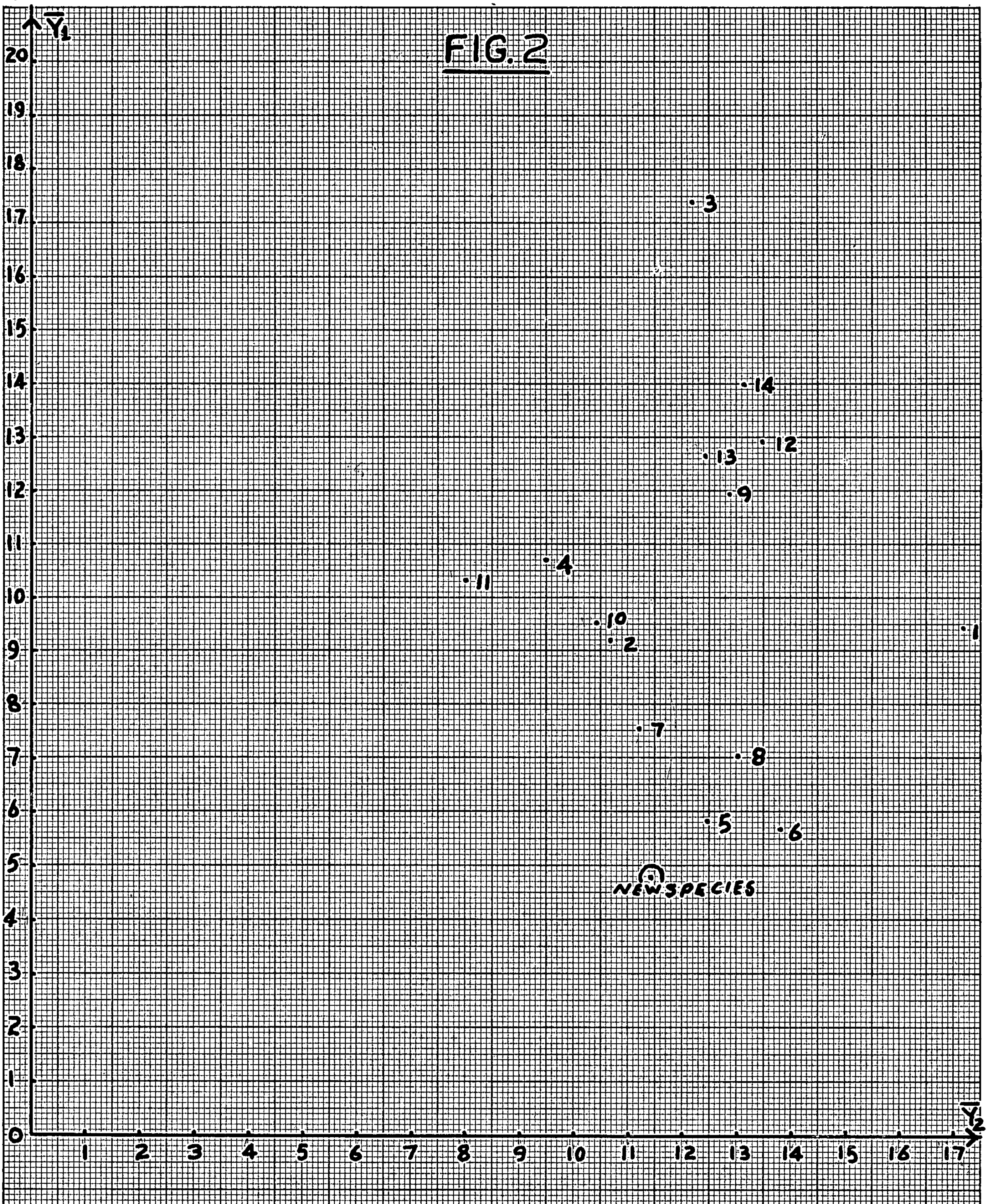
Species No.	Size	$\overline{Y}_1$	$\overline{Y}_2$	$\overline{Y}_3$
1	264	0.94597083	1.72039748	0.34593229
2	78	0.91725592	1.07290071	0.63864788
3	158	1.74328671	1.21889759	0.50048469
4	212	1.07183611	0.95381032	0.36949996
5	324	0.58531407	1.24622259	0.56758425
6	93	0.57211123	1.38532958	0.46747816
7	380	0.75515749	1.12082710	0.77524229
8	436	0.70890607	1.31124555	0.70355287
9	200	1.19390269	1.28747391	0.55733533
10	90	0.95036642	1.04299783	0.57671666
11	207	1.03365387	0.80245391	0.34345856
12	458	1.29139812	1.34851322	0.68878220
13	348	1.26791986	1.24270782	0.78926436
14	260	1.40109551	1.31631867	0.60461486

<u>Note</u>: The column under $\overline{Y}_1$ corresponds to the largest significant discriminant score, the column under $\overline{Y}_2$ to the second largest and that under $\overline{Y}_3$ to the third largest significant score.

Table 7 (Table of Distances)

1	2	3	4	5	6	7	8	9	10	11	12	13	14
6/ .2669	10/ .0058	14/ .1375	11/ .0250	6/ .0296	5/ .0296	8/ .0436	5/ .0380	12/ .0306	2/ .0058	4/ .0250	13/ .0119	12/ .0119	12/ .0202
9/ .2936	7/ .0472	12 .2565	10 .0657	8 .0380	8 .0800	2 .0472	7 .0436	14 .0460	4 .0657	10 .1193	14 .0202	14 .0573	9 .0460
8 .3516	8 .1048	9 .3098	2 .1106	7 .0877	7 .1983	10 .0836	6 .0800	13 .0613	7 .0836	2 .1739	9 .0306	9 .0613	13 .0573
12 .3752	4 .1106	13 .3100	9 .1616	2 .1453	2 .2427	5 .0877	2 .1048	10 .1195	11 .1193	9 .3067	2 .2185	2 .1745	3 .1375
5 .4041	9 .1292	4 .5384	7 .2929	10 .1747	1 .2669	6 .1983	10 .1464	2 .1292	9 .1195	13 .3475	10 .2223	10 .1860	10 .2698
14 .4374	5 .1453	10 .6655	14 .2951	4 .3614	10 .2723	9 .2675	9 .2572	4 .1616	8 .1464	7 .3654	3 .2565	7 .2780	2 .2946
2 .5078	11 .1739	11 .7018	13 .2984	9 .3722	9 .4043	13 .2780	13 .3245	8 .2572	5 .1747	5 .4483	4 .3060	4 .2984	4 .2951
10 .5122	13 .1745	2 .7232	12 .3060	1 .4041	4 .4457	4 .2929	12 .3409	7 .2675	13 .1860	14 .4676	8 .3409	3 .3100	1 .4374
13 .5284	12 .2185	1 .9112	5 .3614	11 .4483	12 .5677	12 .3470	1 .3516	1 .2936	12 .2223	12 .4840	7 .3470	8 .3245	11 .4674
7 .5802	6 .2427	7 1.0615	8 .3712	13 .5151	11 .5681	11 .3654	4 .3712	11 .3067	14 .2698	8 .4941	1 .3752	11 .3475	7 .4847
4 .6041	14 .2946	8 1.1198	6 .4457	12 .5237	13 .6080	14 .4847	14 .4890	3 .3098	6 .2723	6 .5681	11 .4840	5 .5151	8 .4890
11 .8504	1 .5078	5 1.3460	3 .5384	14 .6718	14 .7108	1 .5802	11 .4941	5 .3722	1 .5122	3 .7018	5 .5237	1 .5284	5 .6718
3 .9111	3 .7232	6 1.4003	1 .6041	3 1.3460	3 1.4003	3 1.0615	3 1.1198	6 .4043	3 .6655	1 .8504	6 .5677	6 .6080	6 .7108

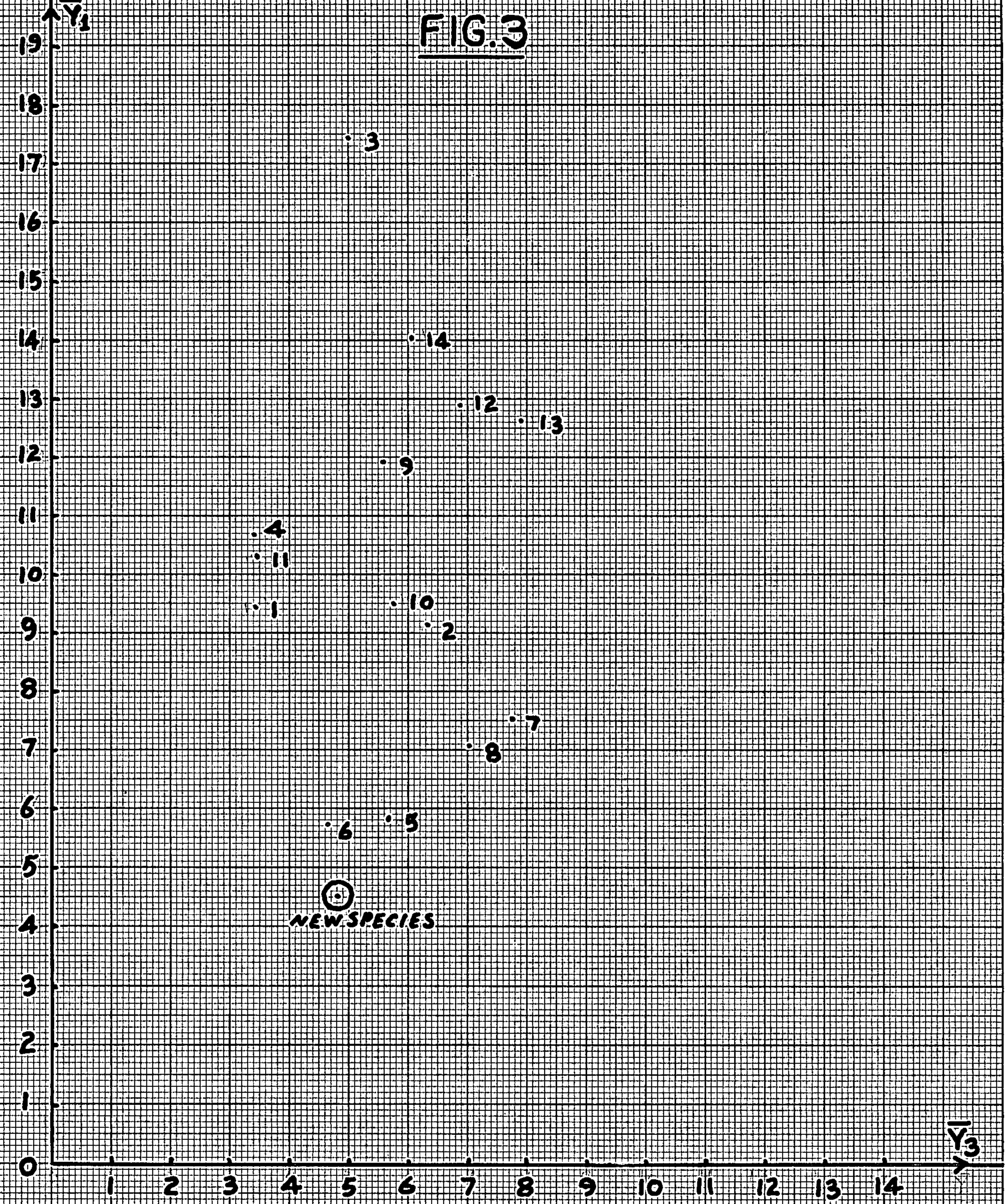
FIG. 2



SCALE : 0.1 UNIT = 1 Sq. ON BOTH AXES.

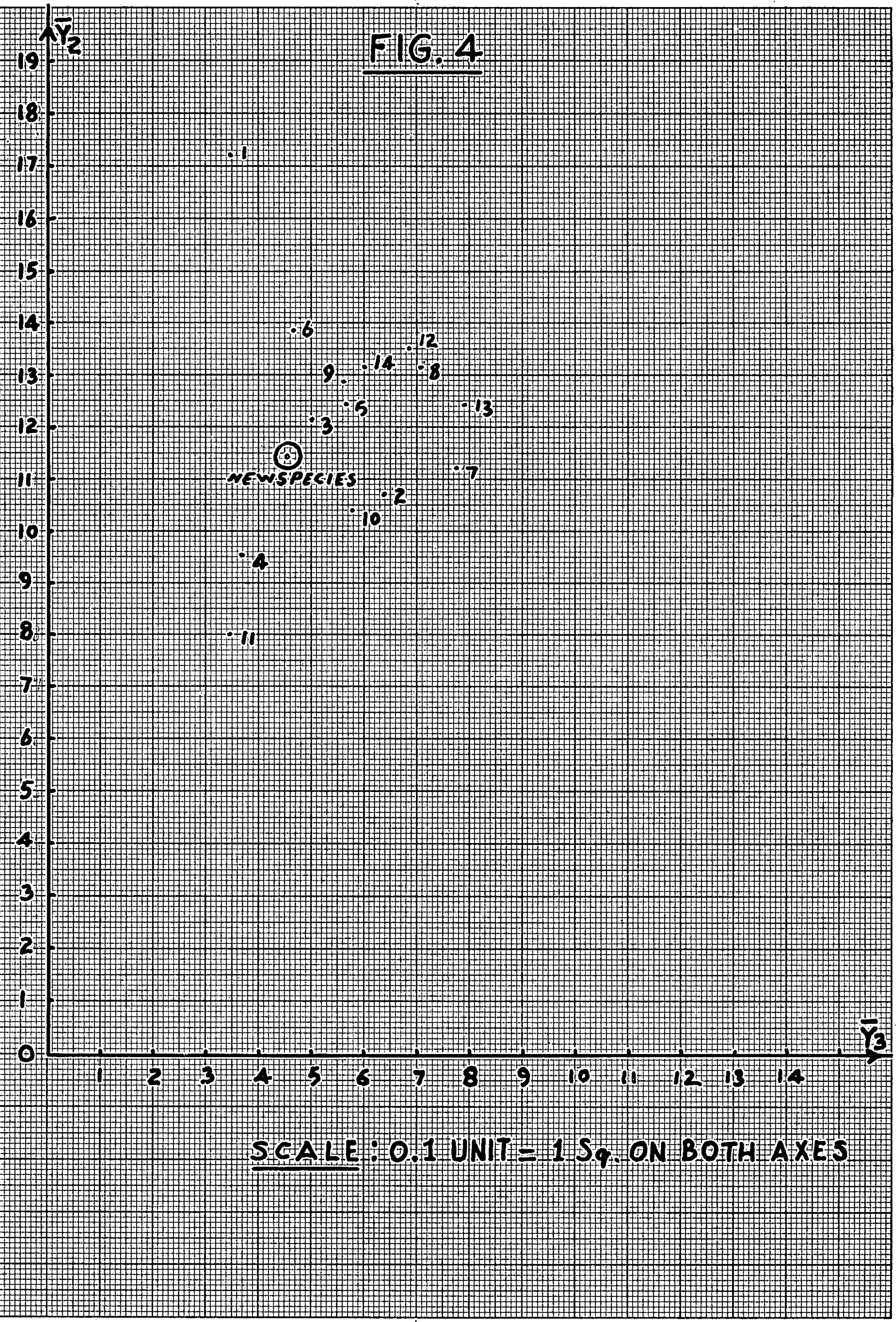
MOST SIGNIFICANT $\overline{Y}$'s

FIG.3



SCALE: 0.1 UNIT = 1 Sq. ON BOTH AXES.

FIG. 4



SCALE: 0.1 UNIT = 1 Sq. ON BOTH AXES

Forming of Clusters

Stage I:

In Stage I we predict the clusters, keeping before us Table 7 and Figures 2, 3, and 4. Relying on the plane representation of the most significant discriminant scores $\overline{Y}_1$ and $\overline{Y}_2$ and then following the criteria discussed in Step 3 of Stage I in Section (2.3), we predict the following clusters:

(i) 2, 5, 6, 7, and 8.

(ii) 2, 5, 7, 8, and 10.

(iii) 2, 4, 10, and 11.

(iv) 2, 4, 9, and 10.

(v) 9, 12, 13, and 14.

and (vi) 1 and 3 by themselves.

Stage II:

We now correct the above predicted clusters for each of which we have a tabular set up given below, and from them we obtain the corrected clusters.

Table 8

Populations involved	Computed T_k^2	D.F.	Tabular $T^2_{\alpha\ k}$ 5%	1%	Conclusion	Cluster
2,5,6,7,8	34.47	16	25.6131	35.6187	Significant	
2,5,6,8	19.89	12	21.7064	29.9100	Non-significant	2,5,6,8
2,5,6,7	41.43	"	"	"	Significant	
2,6,7,8	40.56	"	"	"	"	
2,5,7,8	34.21	"	"	"	Significant	
5,6,7,8	27.91	"	"	"	"	
2,5,7	20.50	8	16.7783	24.0936	Significant	
2,6,7	22.83	"	"	"	"	
2,7,8	13.61	"	"	"	Non-significant	2,7,8
5,6,7	23.37	"	"	"	Significant	
5,7,8	20.44	"	"	"	"	
6,7,8	19.21	"	"	"	"	
6,7	17.38	4	12.1371	18.2030	"	
5,7	15.35	4	"	"		

Table 9

Populations involved	Computed T^2_k	D.F.	Tabular T^2_k 5%	Tabular T^2_k 1%	Conclusion	Cluster
2,5,7,8,10	34.98	16	25.6131	35.6187	Significant	
2,5,7,10	27.62	12	21.7064	29.9100	"	
2,5,8,10	25.30	"	"	"	"	
5,7,8,10	30.15	"	"	"	"	
2,5,7,8	26.31	"	"	"	"	
2,7,8,10	21.37	"	"	"	Non-significant	2,7,8,10
2,5,7	20.50	8	16.7783	24.0936	Significant	
(*)2,5,8	17.79	"	"	"	Significant	2,5,8
2,5,10	18.90	"	"	"	"	
5,7,8	20.44	"	"	"	"	
5,7,10	23.62	"	"	"	"	
5,8,10	19.07	"	"	"	"	
5,7	15.35	4	12.1371	18.2030	"	
5,10	12.98	4	"	"	"	

Table 10

2,4,10,11	15.76	12	21.7064	29.9100	Non-significant	2,4,10,11

(*) We could exclude this from being considered because it already has been included in the bigger cluster (2,5,6,8).

Table 11

Populations involved	Computed T_k^2	D.F.	Tabular T_k^2 5%	Tabular T_k^2 1%	Conclusion	Cluster
9,12,13,14	16.62	12	21.7064	29.9100	Non-significant	9,12,13,14

Table 12

Populations involved	Computed T_k^2	D.F.	Tabular T_k^2 5%	Tabular T_k^2 1%	Conclusion	Cluster
2,4,9,10	24.37	12	21.7064	29.9100	Significant	
2,4,9	21.82	8	16.7783	24.0936	"	
(*) 2,4,10	8.20	"	"	"	Non-significant	2,4,10
2,9,10	11.45	"	"	"	"	2,9,10
4,9,10	20.42	"	"	"	Significant	
4,9	16.624	4	12.1371	18.2030	"	

(*) We could exclude this from being considered because it already has been included in the bigger cluster (2,4,10,11).

Thus, from tables 8 to 12, one concludes that the following are clusters:

(a) 2,5,6, and 8.
(b) 2,7,8, and 10.
(c) 2,9, and 10.
(d) 2,4,10 and 11.
(e) 9,12,13, and 14.
(f) 1, by itself.
(g) 3, by itself.

Further, it remains to prove that each and every set of populations of which these clusters form a subset is significant. To do this, we refer back to the Table 7 of distances and the Figs. 2, 3, and 4 and form the following bigger clusters by incorporating in the corrected clusters the populations lying closest to them:

(i) 2, 5, 6, 8, and 10.
(ii) 2, 4, 7, 8, and 10.
(iii) 2, 4, 7, 10 and 11.
(iv) 2, 4, 9, 10, and 11.
(v) 2, 9, 12, 13, and 14.
(vi) 3, 9, 12, 13, and 14.
(vii) 2, 9, 10, and 13.
(viii) 1 and 6.
(ix) 3 and 14.

We test the significance of these bigger clusters and, as shown in Table 13, find them all to be significant which confirms the conclusion made above.

Table 13

Populations involved	Computed T_k^2	D.F.	Tabular T_k^2 5%	Tabular T_k^2 1%	Conclusion	Cluster
2,5,6,8,10	31.02	16	25.6131	35.6187	Significant	
2,4,7,8,10	68.39	"	"	"	"	
2,4,7,10,11	68.07	"	"	"	"	
2,4,9,10,11	41.88	"	"	"	"	
2,9,12,13,14	30.85	"	"	"	"	
3,9,12,13,14	50.23	"	"	"	"	
2,9,10,13	26.63	12	21.7064	29.9100	"	
1,6	18.36	4	12.1371	18.2030	"	
3,14	14.71	4	"	"	"	

CHAPTER THREE

ANALOGUES OF DUNCAN'S PROCEDURE IN FORMING CLUSTERS IN MULTIVARIATE ANOVA (Contd.)

3.1 In section (2.3) we have proposed three alternative approaches to correct the predicted clusters where the first - called the Duncan-Hotelling test - has been explained quite at length with an illustrative example. Now we take up the remaining two - the 'Extreme Distance from the Mean' - E-test and the 'Largest Distance' - R-test. The exact distributions of both the statistics are not known. Siotani (1958) has found the approximate distribution of the E-statistic for the k p-variate normal populations and has computed the tabular values at 5% and 1% significance levels for some particular values of p. With Siotani's tabular values in hand we first discuss below the procedure for the E-test in Section (3.2). We then take up the R-statistic in Section (3.3) and discuss the working procedure. Lastly, in Section (3.4) we present the distribution of the R-statistic for the bivariate case in the form of definite integrals.

3.2 Procedure for the E-Statistic

The E-test is based on Mahalanobis' distance and Duncan's level of significance based on degrees of freedom.

Suppose again that the clusters have been predicted by following the procedure discussed in Stage I of Section (2.3). Without losing generality, we take up one of the predicted clusters containing k_1 populations and discuss the procedure for the E-test in the following steps:

(i) Compute the statistic E_i ($i = 1, 2, \ldots, k_1$), the Mahalanobis' distance between the mean vectors of the ith population and the grand mean vectors of the k_1 populations.

(ii) Without losing generality, let E_{k_1} be the largest of all the computed E_i $(i = 1, 2, \ldots, k_1)$.

(iii) Compare this E_{k_1} with tabular $E_{\mathcal{L}_{k_1}}$, where $\mathcal{L}_{k_1}$ is defined already in (2.2.6) and $\mathcal{L}$ is the pre-assigned significance level.

(iv) If E_{k_1} is less than or equal to $E_{\mathcal{L}_{k_1}}$, all the k_1 populations involved are concluded to form a cluster. Otherwise, split the k_1 populations into k_1 sets of (k_1-1) populations each.

(v) Compare the extreme distance of each set of (k_1-1) populations from their respective grand mean vectors with the tabular $E_{\mathcal{L}_{k_1-1}}$. Out of them some may be significant and some may not be. Those non-significant will yield clusters with the corresponding populations involved in them. Those, for which the extreme E's are significant, are further split into sets of (k_1-2) each and their corresponding extreme E's are them compared against the tabular value $E_{\mathcal{L}_{k_1-2}}$. In this way the process is continued till we arrive at the clusters of the type defined.

Thus a working criterion analogous to Duncan's can be stated as follows: 'A group of k_1 populations will form a cluster if the extreme distance E_{k_1} (assumed to be the largest amongst all the k_1 distances between the mean vectors of individual populations and their grand mean vector) is non-significant and if furthermore such extreme E's of each and every new set of populations of which the k_1 populations form a subset, is significant according to $\mathcal{L}_r$-level E-test for some pre-assigned $\mathcal{L}$, where r is the number of the populations in the set'.

Note: The exact distribution of the extreme classical distance was taken up by Mrs. Cuttle in her Master's thesis, 1956. She successfully solved the problem for three bivariate populations and gave the tabular values at some probability levels. We tried in vain to extend her procedure to four bivariate populations. The joint distribution of four distances came out in terms of elliptic funcions, whose further integration, in order to find the distribution of the extreme E amongst the four E's, was found to be quite involved.

3.3 Procedure for R-Statistic

Duncan's range test has already been explained in Section (1.2). We extend his procedure to the multivariate case. Suppose we have k p-variate normal populations having significantly different mean vectors. Suppose further that the clusters have been predicted by following the procedure discussed in Stage I of Section (2.3). In correcting these predicted clusters no generality is lost if we take up one cluster containing $k_1 (\leq k)$ populations. The procedure is described in detail in the following steps:

(i) Compute $\binom{k_1}{2}$ Mahalanobis distances R_{rs} $(r \neq s = 1, 2, \ldots, k_1)$ between the rth and sth populations.

(ii) Again, no generality is lost if we suppose that the distance R_{1k_1} between the first and the k_1th populations is the largest amongst $\binom{k_1}{2}$ distances.

(iii) Compare the computed R_{1k_1} with the tabular $R_{\alpha k_1}$, where α_{k_1} is already defined in (2.2.6) and α is a pre-assigned level of significance. If R_{1k_1} is less than or equal to $R_{\alpha k_1}$, all the k_1 populations involved are considered to form a cluster. Otherwise, split the set of k_1 populations into k_1 sets of (k_1-1) populations each.

(iv) Compare the largest distance of each set of (k_1-1) populations with the tabular $R_{\alpha_{k_1-1}}$. Out of them some may be significant and some may not be. Those non-significant will yield clusters with the populations involved in them. Those for which the largest distance is significant are further split into sets of (k_1-2) and their respective largest distances are then compared against their corresponding tabular values $R_{\mathcal{L}_{k_1-2}}$. In this way the process is continued till we arrive at the clusters of the type defined.

Thus the working criterion analogous to Duncan's can be summed up as follows: 'A group of k_1 populations will form a cluster if the distance (assumed to be the largest amongst all $\binom{k_1}{2}$ distances) between the first and the k_1th populations is non-significant and also the largest distance, amongst all possible distances between pairs of each and every new set of populations of which the k_1 populations form a subset, is significant according to $\mathcal{L}_r$-level R-test for some pre-assigned $\mathcal{L}$, where r is the number of populations in the set'.

There is no doubt that the test procedure set up above is completely analogous to what Duncan did in his multiple range test, but, in order to apply it, we need the distribution of the statistic R and hence the tabular values at $\mathcal{L}_r$-level for r populations. To overcome part of the difficulty we present below the simultaneous distribution of the distances involved in a predicted cluster in the case of bivariate populations. We have actually found the joint distribution for $k = 3, 4, 5$ populations and then have generalized it for any k. Lastly, we have also suggested the limits of integration to find the distributions of the individual largest distance i.e. of the statistic R. To find the tabular values one could apply any method of numerical integration.

3.4 The Distribution of the R-Statistic in the form of a Definite Integral

(a) Preliminaries and Notations

(i) Let $\bar{X}^t(k \times p)$ be the matrix of k mean vectors (columns for characters and rows for sub-population samples) of samples of sizes $N_1, N_2, \ldots, N_k$ respectively drawn independently from k p-variate normal populations.

Let the covariance matrix $(\mathcal{L}_{ij})$ be known or estimated on the basis of large samples.

Further, let the matrix $\bar{X}^t(k \times p)$ be transformed into another matrix $\bar{Y}^t(k \times p)$ by such an orthogonal transformation that the covariance matrix of $\bar{y}$'s is a diagonal matrix Λ $(p \times p)$ with elements λ_i $(i = 1, 2, \ldots, p)$. Without loss of generality we can assume that the true centroid of the distribution is $\mu_1 = \mu_2 = \ldots = \mu_p = 0$. The joint distribution of the $\bar{y}$'s is then:

$$f(\bar{y}_{11}, \ldots, \bar{y}_{p1}, \ldots, \bar{y}_{1k}, \ldots, \bar{y}_{pk}) \prod_{i=1}^{p} \prod_{r=1}^{k} d\bar{y}_{ir}$$

$$c_{pk} \exp\left(-\frac{1}{2} \sum_{r=1}^{k} N_r \sum_{i=1}^{p} \left(\frac{1}{\lambda_i} \bar{y}_{ir}^2\right)\right) \prod_{i=1}^{p} \prod_{r=1}^{k} d\bar{y}_{ir} \qquad (3.4.1)$$

where $$c_{pk} = \prod_{i=1}^{p} \prod_{r=1}^{k} \sqrt{\frac{N_r}{2\pi\lambda_i}} \qquad (3.4.2)$$

Now $$\sum_{r=1}^{k} N_r \sum_{i=1}^{p} \left(\frac{1}{\lambda_i}\,\bar{y}_{ir}^2\right) = \sum_{i=1}^{p} \frac{1}{\lambda_i} \sum_{r=1}^{k} N_r(\bar{y}_{ir} - \bar{y}_i)^2 + n_k \sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\bar{y}_i^2\right)$$

where $$n_k = \sum_{r=1}^{k} N_r \qquad (3.4.3)$$

Further, it is easy to prove that:

$$\sum_{r=1}^{k} N_r(\bar{y}_{ir} - \bar{y}_i)^2 = \frac{1}{n_k} \sum_{r=1}^{k-1} \sum_{s=r+1}^{k} N_r N_s (\bar{y}_{ir} - \bar{y}_{is})^2$$

Thus we have:

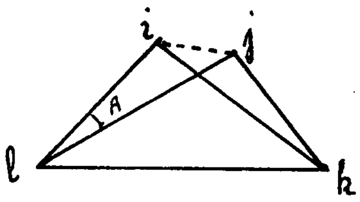
$$\sum_{r=1}^{k} N_r \sum_{i=1}^{p} \left(\frac{1}{\ }{}_i\,\bar{y}_{ir}^2\right) = \sum_{r=1}^{k-1} \sum_{s=r+1}^{k} R_{rs} + n_k \sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\,\bar{y}_i^2\right) \qquad (3.4.4)$$

where $$R_{rs} = \frac{N_r N_s}{n} \sum_{i=1}^{p} \frac{1}{\lambda_i}(\bar{y}_{ir} - \bar{y}_{is})^2 \qquad (3.4.5)$$

Thus the joint distribution (3.4.1) can be written as:

$$c_{pk} \exp\left(-\frac{1}{2} \sum_{r=1}^{k-1} \sum_{s=r+1}^{k} R_{rs} - \frac{1}{2} n_k \sum_{i=1}^{p} \left(\frac{1}{\lambda_i}\,\bar{y}_i^2\right)\right) \prod_{i=1}^{p} \prod_{r=1}^{k} d\bar{y}_{ir} \qquad (3.4.6)$$

(ii) From the quadrilateral joining points i, j, k and ℓ,



we can find the distance $\sqrt{R_{ij}}$ between i and j as follows:

$$R_{ij} = R_{i\ell} + R_{j\ell} - 2\sqrt{R_{i\ell}R_{j\ell}}\ \text{Cos}\ A \tag{3.4.7}$$

$$\text{where}\quad A = \text{Cos}^{-1}\frac{R_{i\ell} + R_{\ell k} - R_{ik}}{2\sqrt{R_{i\ell}R_{\ell k}}} - \text{Cos}^{-1}\frac{R_{\ell j} + R_{\ell k} - R_{jk}}{2\sqrt{R_{\ell j}R_{\ell k}}} \tag{3.4.8}$$

(iii) Frequently we shall have relations of the type:

$$ax + by = L$$

$$x^2 + y^2 = M$$

and

$$a^2 + b^2 = N \tag{3.4.9}$$

where we shall be required to find the value of:

$$bx - ay \tag{3.4.10}$$

Solving the first two equations of (3.4.9) we have

$$x = \frac{aL \pm b\sqrt{(a^2 + b^2)M - L^2}}{a^2 + b^2}$$

$$\text{and}\quad y = \frac{bL \mp a\sqrt{(a^2 + b^2)M - L^2}}{a^2 + b^2}$$

where we have placed the restriction that the signs before the square root in the expressions of x and y must be opposite. Therefore

$$bx - ay = \pm\sqrt{(a^2 + b^2)M - L^2} = \pm\sqrt{NM - L^2} \tag{3.4.11}$$

(iv) We shall frequently need the following:

$$\int_0^a \frac{dx}{\sqrt{ax - x^2}} = \pi \qquad (3.4.12)$$

(v) Lastly we give below the notations which are used quite frequently in what follows:

$$S_{ijk} \quad 2N_iN_jR_{ki}R_{kj} + 2N_iN_kR_{ji}R_{jk} + 2N_jN_kR_{ik}R_{ij} - N_i^2 R_{jk}^2 - N_j^2 R_{ik}^2 - N_k^2 R_{ij}^2 \qquad (3.4.13)$$

$$\bar{S}_{ijk} = 2R_{ki}R_{kj} + 2R_{ji}R_{jk} + 2R_{ik}R_{ij} - R_{jk}^2 - R_{ik}^2 - R_{ij}^2 \qquad (3.4.14)$$

$$S' = 2R'_{ki}R'_{kj} + 2R'_{ji}R'_{jk} + 2R'_{ik}R'_{ij} - R'^2_{jk} - R'^2_{ik} - R'^2_{ij} \qquad (3.4.15)$$

(b) <u>Distributions</u>

<u>Case I</u>: <u>For k = 3</u>

The joint distribution of $(\bar{y}_{11}, \bar{y}_{12}, \bar{y}_{13}; \bar{y}_{21}, \bar{y}_{22}, \bar{y}_{23})$ from (3.4.6.) is:

$$c_{23} \exp\left[-\frac{1}{2} \sum_{r=1}^{2} \sum_{s=r+1}^{3} R_{rs} - \frac{n_3}{2} \sum_{i=1}^{2} \frac{1}{\lambda_i} \bar{y}_i^2 \right] \prod_{i=1}^{2} \prod_{r=1}^{3} d\bar{y}_{ir} \qquad (3.4.16)$$

where,

from (3.4.2), $$c_{23} = \frac{N_1N_2N_3}{\left[2\pi\sqrt{\lambda_1 \lambda_2}\right]^3} \qquad (3.4.17)$$

and from (3.4.3), $$n_3 = N_1 + N_2 + N_3 \qquad (3.4.18)$$

Consider the orthogonal transformation:

$$u_1 = \frac{1}{\sqrt{3}}\left(\bar{y}_{11} + \bar{y}_{12} + \bar{y}_{13}\right) \qquad \text{and} \qquad u_2 = \frac{1}{\sqrt{3}}\left(\bar{y}_{21} + \bar{y}_{22} + \bar{y}_{23}\right)$$

$$v_{11} = \frac{1}{\sqrt{2}}\left(- \bar{y}_{11} + \bar{y}_{12}\right) \qquad v_{21} = \frac{1}{\sqrt{2}}\left(- \bar{y}_{21} + \bar{y}_{22}\right)$$

$$v_{12} = \frac{1}{\sqrt{6}}\left(- \bar{y}_{11} - \bar{y}_{12} + 2\bar{y}_{13}\right) \qquad v_{22} = \frac{1}{\sqrt{6}}\left(- \bar{y}_{21} - \bar{y}_{22} + 2\bar{y}_{23}\right) \tag{3.4.19}$$

whose inverse transformation is:

$$\bar{y}_{11} = \frac{u_1}{\sqrt{3}} - \frac{v_{11}}{\sqrt{2}} - \frac{v_{12}}{\sqrt{6}} \qquad \text{and} \qquad \bar{y}_{21} = \frac{u_2}{\sqrt{3}} - \frac{v_{21}}{\sqrt{2}} - \frac{v_{22}}{\sqrt{6}}$$

$$\bar{y}_{12} = \frac{u_1}{\sqrt{3}} + \frac{v_{11}}{\sqrt{2}} - \frac{v_{12}}{\sqrt{6}} \qquad \bar{y}_{22} = \frac{u_2}{\sqrt{3}} + \frac{v_{21}}{\sqrt{2}} - \frac{v_{22}}{\sqrt{6}}$$

$$\bar{y}_{13} = \frac{u_1}{\sqrt{3}} + \frac{2v_{12}}{\sqrt{6}} \qquad \bar{y}_{23} = \frac{u_2}{\sqrt{3}} + \frac{2v_{22}}{\sqrt{6}}$$

and from these and from (3.4.5), we have:

$$R_{12} = \frac{N_1N_2}{n_3}\left[\frac{1}{\lambda_1}\left(\frac{2v_{11}}{\sqrt{2}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{2v_{21}}{\sqrt{2}}\right)^2 \right]$$

$$R_{13} = \frac{N_1N_3}{n_3}\left[\frac{1}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} + \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} + \frac{3v_{22}}{\sqrt{6}}\right)^2 \right]$$

$$R_{23} = \frac{N_2N_3}{n_3}\left[\frac{1}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} - \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} - \frac{3v_{22}}{\sqrt{6}}\right)^2 \right] \tag{3.4.20}$$

The distribution now takes the form:

$$C_{23} \exp\left[- \frac{1}{2} \sum_{r=1}^{2} \sum_{s=r+1}^{3} R_{rs} - \frac{n_3}{2} \sum_{i=1}^{2} \frac{1}{\lambda_i} \frac{u_i^2}{3}\right] \prod_{i=1}^{2} \prod_{j=1}^{2} dv_{ij} \prod_{i=1}^{2} du_i \tag{3.4.21}$$

where R_{rs} are defined as in (3.4.20).

Integrating with respect to u_1 and u_2 both with the limits from $-\infty$ to ∞, we get the reduced form of (3.4.21) as:

$$\frac{6\pi\sqrt{\lambda_1\lambda_2}}{n_3} C_{23} \exp\left[-\frac{1}{2}\sum_{r=1}^{2}\sum_{s=r+1}^{3} R_{rs}\right]\prod_{i=1}^{2}\prod_{j=1}^{2} dv_{ij} \qquad (3.4.22)$$

Let $N = \frac{N_1N_2N_3}{n_3}$. Then we define R'_{12}, R'_{13} and R'_{23} as:

$$R'_{12} = N_3R_{12} = \frac{N}{\lambda_1}\left(\frac{2v_{11}}{\sqrt{2}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{2v_{21}}{\sqrt{2}}\right)^2$$

$$R'_{13} = N_2R_{13} = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} + \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} + \frac{3v_{22}}{\sqrt{6}}\right)^2$$

$$R'_{23} = N_1R_{23} = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} - \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} - \frac{3v_{22}}{\sqrt{6}}\right)^2 \qquad (3.4.23)$$

Further, to effect the change of variables from the v's to R's, we introduce a fourth R' defined by:

$$R' = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}}\right)^2 \qquad (3.4.24)$$

Finding first from (3.4.23) and (3.4.24) the Jacobian of the transformation, we conclude that:

$$\prod_{i=1}^{2}\prod_{j=1}^{2} dv_{ij} = \frac{dR'_{12}dR'_{13}dR'_{23}dR'}{\frac{288N^2}{\lambda_1\lambda_2}\left(\frac{N}{\lambda_1}\frac{v_{11}}{\sqrt{2}}\right)\left(\frac{N}{\lambda_2}\frac{v_{21}}{\sqrt{2}}\right)\left(\frac{N}{\lambda_1}\frac{v_{11}}{\sqrt{2}}\frac{N}{\lambda_2}\frac{v_{22}}{\sqrt{6}} - \frac{N}{\lambda_1}\frac{v_{12}}{\sqrt{6}}\frac{N}{\lambda_2}\frac{v_{21}}{\sqrt{2}}\right)}$$

Further, with the help of (3.4.9), (3.4.10), (3.4.11), (3.4.23) and (3.4.24), we obtain:

$$\prod_{i=1}^{2} \prod_{j=1}^{2} dv_{ij} = \frac{dR'_{12} dR'_{13} dR'_{23} dR'}{\frac{24N^2}{\lambda_1 \lambda_2} \sqrt{R'(\frac{R'_{12}}{4} - R')} \sqrt{S'_{123}}}$$

where S'_{123} is defined as in (3.4.15). Again using (3.4.23), we get

$$\prod_{i=1}^{2} \prod_{j=1}^{2} dv_{ij} = \frac{N_1 N_2 N_3}{24N^2} \frac{dR_{12} dR_{13} dR_{23} dR'}{\sqrt{R'(\frac{N_3}{4} R_{12} - R')} \sqrt{S_{123}}} \tag{3.4.25}$$

Using (3.4.12), (3.4.25) and the value $N = \frac{N_1 N_2 N_3}{n_3}$, the joint distribution (3.4.22) reduces, after integrating with respect to R' over the range from 0 to $\frac{N_3}{4} R_{12}$ as shown in (3.4.12), to:

$$(\tfrac{1}{2})^{3+1} (\tfrac{n_3}{2\pi})^{3-2} \frac{\exp\left[-\frac{1}{2}(R_{12} + R_{13} + R_{23})\right]}{\sqrt{S_{123}}} dR_{12} dR_{13} dR_{23} \tag{3.4.26}$$

which is the joint distribution of R_{12}, R_{13} and R_{23}. All these variates are always positive, and it is easy to check that they do not assume values outside the cone defined as $S_{123} \geq 0$. The distribution of $f(R_{12}, R_{13}, R_{23})$ is therefore always positive.

The Distribution of the Largest R_{rs}

Let us further restrict the problem by assuming the number of observations to be the same for all the three groups, i.e. $N_1 = N_2 = N_3 = N_0$, say. The joint distribution of (R_{12}, R_{13}, R_{23}) is

$$f(R_{12}, R_{13}, R_{23}) = \frac{3}{32\pi} \frac{\exp\left[-\frac{1}{2}(R_{12} + R_{13} + R_{23})\right]}{\sqrt{\bar{S}_{123}}} \qquad (3.4.27)$$

where now the variates R_{12}, R_{13}, and R_{23} do not assume values outside the cone defined by $\bar{S}_{123} \geq 0$.

We can assume without loss of generality that the variates have been ordered, say $0 \leq R_{23} \leq R_{13} \leq R_{12} < \infty$. The density of these ordered variates is $3!f(R_{12}, R_{13}, R_{23})$. Thus the probability $G(t)$, that $R_{12} \leq t$, is:

$$G(t) = \frac{3}{32\pi} \, 3! \iiint_V \frac{\exp\left[-\frac{1}{2}(R_{12} + R_{13} + R_{23})\right]}{\sqrt{\bar{S}_{123}}} \, dR_{12} dR_{13} dR_{23} \qquad (3.4.28)$$

where V is the region:

$$\begin{aligned}
(\sqrt{R_{12}} - \sqrt{R_{13}})^2 &\leq R_{23} \leq R_{13} \\
\tfrac{1}{4} R_{12} &\leq R_{13} \leq R_{12} \\
0 &\leq R_{12} \leq t \\
0 &\leq t < \infty
\end{aligned}$$

The procedure for its numerical integration has been given by Mrs. Cuttle and one can easily compute the values of t for known values of G(t).

Case II: For $k = 4$

The joint distribution of $(\bar{y}_{11}, \bar{y}_{12}, \bar{y}_{13}, \bar{y}_{14}; \bar{y}_{21}, \bar{y}_{22}, \bar{y}_{23}, \bar{y}_{24})$ from (3.4.6) is

$$C_{24} \exp\left[-\frac{1}{2} \sum_{r=1}^{3} \sum_{s=r+1}^{4} R_{rs} - \frac{n_4}{2} \sum_{i=1}^{2} \frac{1}{\lambda_i} \bar{y}_i^2 \right] \prod_{i=1}^{2} \prod_{r=1}^{4} d\bar{y}_{ir} \qquad (3.4.29)$$

where, from (3.4.2), $C_{24} = \dfrac{N_1 N_2 N_3 N_4}{\left(2\pi\sqrt{\lambda_1 \lambda_2}\right)^4}$ (3.4.30)

and from (3.4.3), $n_4 = N_1 + N_2 + N_3 + N_4$ (3.4.31)

Consider an orthogonal transformation of the type (3.4.19) whose inverse transformation we write as:

$$\bar{y}_{11} = \frac{u_1}{\sqrt{4}} - \frac{v_{11}}{\sqrt{2}} - \frac{v_{12}}{\sqrt{6}} - \frac{v_{13}}{\sqrt{12}} \quad \text{and} \quad \bar{y}_{21} = \frac{u_2}{\sqrt{4}} - \frac{v_{21}}{\sqrt{2}} - \frac{v_{22}}{\sqrt{6}} - \frac{v_{23}}{\sqrt{12}}$$

$$\bar{y}_{12} = \frac{u_1}{\sqrt{4}} + \frac{v_{11}}{\sqrt{2}} - \frac{v_{12}}{\sqrt{6}} - \frac{v_{13}}{\sqrt{12}} \qquad \bar{y}_{22} = \frac{u_2}{\sqrt{4}} + \frac{v_{21}}{\sqrt{2}} - \frac{v_{22}}{\sqrt{6}} - \frac{v_{23}}{\sqrt{12}}$$

$$\bar{y}_{13} = \frac{u_1}{\sqrt{4}} + \frac{2v_{12}}{\sqrt{6}} - \frac{v_{13}}{\sqrt{12}} \qquad \bar{y}_{23} = \frac{u_2}{\sqrt{4}} + \frac{2v_{22}}{\sqrt{6}} - \frac{v_{23}}{\sqrt{12}}$$

$$\bar{y}_{14} = \frac{u_1}{\sqrt{4}} + \frac{3v_{13}}{\sqrt{12}} \qquad \bar{y}_{24} = \frac{u_2}{\sqrt{4}} + \frac{3v_{23}}{\sqrt{12}}$$

With the help of these and (3.4.5) we have:

$$R_{12} = \frac{N_1N_2}{n_4}\left[\frac{1}{\lambda_1}\left(\frac{2v_{11}}{\sqrt{2}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{2v_{21}}{\sqrt{2}}\right)^2\right]$$

$$R_{13} = \frac{N_1N_3}{n_4}\left[\frac{1}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} + \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} + \frac{3v_{22}}{\sqrt{6}}\right)^2\right]$$

$$R_{14} = \frac{N_1N_4}{n_4}\left[\frac{1}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} + \frac{v_{12}}{\sqrt{6}} + \frac{4v_{13}}{\sqrt{12}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} + \frac{v_{22}}{\sqrt{6}} + \frac{4v_{23}}{\sqrt{12}}\right)^2\right]$$

$$R_{23} = \frac{N_2N_3}{n_4}\left[\frac{1}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} - \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} - \frac{3v_{22}}{\sqrt{6}}\right)^2\right]$$

$$R_{24} = \frac{N_2N_4}{n_4}\left[\frac{1}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} - \frac{v_{12}}{\sqrt{6}} - \frac{4v_{13}}{\sqrt{12}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} - \frac{v_{22}}{\sqrt{6}} - \frac{4v_{23}}{\sqrt{12}}\right)^2\right]$$

$$R_{34} = \frac{N_3N_4}{n_4}\left[\frac{1}{\lambda_1}\left(\frac{2v_{12}}{\sqrt{6}} - \frac{4v_{13}}{\sqrt{12}}\right)^2 + \frac{1}{\lambda_2}\left(\frac{2v_{22}}{\sqrt{6}} - \frac{4v_{23}}{\sqrt{12}}\right)^2\right] \qquad (3.4.32)$$

Making use of (3.4.32) and integrating with respect to u_1 and u_2 both extendings from $-\infty$ to ∞, the distribution (3.4.29) takes the form:

$$\frac{8\pi}{n_4}\sqrt{\lambda_1\lambda_2}\, C_{24} \exp\left[-\frac{1}{2}\sum_{r=1}^{3}\sum_{s=r+1}^{4} R_{rs}\right]\prod_{i=1}^{3}\prod_{j=1}^{3} dv_{ij} \qquad (3.4.33)$$

Let $N = \frac{N_1 N_2 N_3 N_4}{n_4}$. Then from (3.4.32) we have:

$$R'_{12} = N_3 N_4 R_{12} = \frac{N}{\lambda_1}\left(\frac{2v_{12}}{\sqrt{2}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{2v_{21}}{\sqrt{2}}\right)^2$$

$$R'_{13} = N_2 N_4 R_{13} = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} + \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} + \frac{3v_{22}}{\sqrt{6}}\right)^2$$

$$R'_{14} = N_2 N_3 R_{14} = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} + \frac{v_{12}}{\sqrt{6}} + \frac{4v_{13}}{\sqrt{12}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} + \frac{v_{22}}{\sqrt{6}} + \frac{4v_{23}}{\sqrt{12}}\right)^2$$

$$R'_{23} = N_1 N_4 R_{23} = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} - \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} - \frac{3v_{22}}{\sqrt{6}}\right)^2$$

$$R'_{24} = N_1 N_3 R_{24} = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} - \frac{v_{12}}{\sqrt{6}} - \frac{4v_{13}}{\sqrt{12}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} - \frac{v_{22}}{\sqrt{6}} - \frac{4v_{23}}{\sqrt{12}}\right)^2$$

$$R'_{34} = N_1 N_2 R_{34} = \frac{N}{\lambda_1}\left(\frac{2v_{12}}{\sqrt{6}} + \frac{4v_{13}}{\sqrt{12}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{2v_{22}}{\sqrt{6}} + \frac{4v_{23}}{\sqrt{12}}\right)^2 \qquad (3.4.34)$$

However, in changing from the v_{ij} to the R'_{ij}, we discover that the Jacobian of the transformation vanishes. In fact it should, since the quadrilateral is completely determined by taking two triangles standing on the same base or by taking any of the five out of six R' s. Thus, we do away with one of the six R' s (which can be done in 6 ways) and then, to complete the set of six R' s corresponding to six v's, we bring in another R', functionally independent of the five retained R' s. It is defined as:

$$R' = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\lambda 2}\right)^2 \qquad (3.4.35$$

Assuming R'_{34} to be the smallest, we do away with it and replace it by (3.4.35). Then, with the help of (3.4.34) where R'_{34} is left out and of (3.4.35), we find:

$$\prod_{i=1}^{3} \prod_{j=1}^{3} dv_{ij} = \frac{(\lambda_1\lambda_2)^{3/2}}{64N^3} \frac{dR' dR'_{12} dR'_{13} dR'_{14} dR'_{23} dR'_{24}}{\sqrt{R'(\frac{1}{4}R'_{12} - R')}\sqrt{S'_{124}}\sqrt{S'_{123}}} \tag{3.4.36}$$

Finally making use of (3.4.34), (3.4.36), (3.4.13), and (3.4.30), the distribution (3.4.33) reduces, after integrating out R' from 0 to $\frac{1}{4} R'_{12}$(or $\frac{N_2N_3}{4} R_{12}$), again as in (3.4.12), to

$$\binom{6}{1}\left(\frac{1}{2}\right)^{4+1}\left(\frac{n_4}{2\pi}\right)^{4-2} \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{3}\sum_{s=r+1}^{4} R_{rs}\right] dR_{12}dR_{13}dR_{14}dR_{23}dR_{24}}{\sqrt{S_{124}}\ \sqrt{S_{123}}} \tag{3.4.37}$$

where, by using (3.4.7) and (3.4.8), R_{34} is determined from the quadrilateral formed by the points (1,2,3,4) and is substituted in (3.4.37). Furthermore, the variates R_{12}, R_{13}, R_{23}, R_{14}, and R_{24} are all positive and it is easy to prove that R_{12}, R_{13}, and R_{23} do not assume values outside the cone $S_{123} \geq 0$ and that R_{12}, R_{14}, and R_{24} do not assume values outside the cone $S_{124} \geq 0$.

The Distribution of the Largest Distance:

Let us further restrict the problem by assuming that $N_1 = N_2 = N_3 = N_4 = N_0$, say. The joint distribution (3.4.37) then becomes:

$$\binom{6}{1}\left(\tfrac{1}{2}\right)^3 \frac{1}{\pi^2} \quad \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{3}\sum_{s=r+1}^{4} R_{rs}\right] dR_{12}dR_{13}dR_{14}dR_{23}dR_{24}}{\sqrt{\bar{S}_{124}}\ \sqrt{\bar{S}_{123}}} \tag{3.4.38}$$

where again the variates R_{12}, R_{13} and R_{23} do not assume values outside the cone $\bar{S}_{123} \geq 0$ and also R_{12}, R_{14} and R_{24} do not assume values outside the cone $\bar{S}_{124} \geq 0$. Furthermore the distribution of $f(R_{12}, R_{13}, R_{14}, R_{23}, R_{24})$ is always positive.

We can assume without loss of generality that R_{12} is the largest of the five R's and further that they are ordered as:

$$\begin{aligned} &0 \leq R_{23} \leq R_{13} \leq R_{12} < \infty \\ \text{and} \qquad &0 \leq R_{24} \leq R_{14} \leq R_{12} < \infty \end{aligned} \tag{3.4.39}$$

The density of the ordered variates is $5(2!)(2!)f(R_{12}, R_{13}, R_{23}, R_{14}, R_{24})$, and the probability $G(t)$ that $R_{12} \leq t$ is

$$G(t) = \binom{6}{1}(5)(2!)(2!)\left(\tfrac{1}{2}\right)^3 \frac{1}{\pi^2} \underset{V}{\int\!\!\int\!\!\int\!\!\int\!\!\int} \frac{\exp\left(-\frac{1}{2}\sum_{r=1}^{3}\sum_{s=r+1}^{4} R_{rs}\right) dR_{12}dR_{13}dR_{14}dR_{23}dR_{24}}{\sqrt{\bar{S}_{123}}\ \sqrt{\bar{S}_{124}}} \tag{3.4.40}$$

where V is the region:

$$\begin{aligned} (\sqrt{R_{12}} - \sqrt{R_{13}})^2 &\leq R_{23} \leq R_{13} \\ \tfrac{1}{4} R_{12} &\leq R_{13} \leq R_{12} \\ (\sqrt{R_{12}} - \sqrt{R_{14}})^2 &\leq R_{24} \leq R_{14} \end{aligned}$$

$$\tfrac{1}{4} R_{12} \leq R_{14} \leq R_{12}$$

$$0 \leq R_{12} \leq t$$

$$0 \leq t < \infty$$

$G(t)$ can be evaluated by some numerical method.

Case III: For $k = 5$

Following the similar steps as in Case II for $k = 4$ (from (3.4.29) to (3.4.32)) we finally obtain the distribution of R_{rs} ($s = (r+1)$ to 5, $r = 1$ to 4) as:

$$\frac{10\pi\sqrt{\lambda_1\lambda_2}}{n_5} \; C_{25} \; \exp\left[-\frac{1}{2}\sum_{r=1}^{4}\sum_{s=r+1}^{5} R_{rs}\right] \prod_{i=1}^{4}\prod_{j=1}^{4} dv_{ij} \quad (3.4.41)$$

where, from (3.4.2), $C_{25} = \dfrac{N_1N_2N_3N_4N_5}{(\lambda_1\lambda_2)^{5/2}} \; \dfrac{1}{(2\pi)^5}$ (3.4.42)

and from (3.4.3), $n_5 = N_1 + N_2 + N_3 + N_4 + N_5$ (3.4.43)

Letting $N = \dfrac{N_1N_2N_3N_4N_5}{n_5}$, we obtain as in (3.4.34) the following:

$$R'_{12} = N_3N_4N_5R_{12} = \frac{N}{\lambda_1}\left(\frac{2v_{11}}{\sqrt{2}}\right)^2 + \frac{N}{\lambda_1}\left(\frac{2v_{21}}{\sqrt{2}}\right)^2$$

$$R'_{13} = N_2N_4N_5R_{13} = \frac{N}{\lambda_1}\left(\frac{v_{11}}{\sqrt{2}} + \frac{3v_{12}}{\sqrt{6}}\right)^2 + \frac{N}{\lambda_2}\left(\frac{v_{21}}{\sqrt{2}} + \frac{3v_{22}}{\sqrt{6}}\right)^2$$

$$R'_{23} = N_1N_4N_5R_{23} = \frac{N}{\lambda_1}(\frac{v_{11}}{\sqrt{2}} - \frac{3v_{12}}{\sqrt{6}})^2 + \frac{N}{\lambda_2}(\frac{v_{21}}{\sqrt{2}} - \frac{3v_{22}}{\sqrt{6}})^2$$

$$R'_{14} = N_2N_3N_5R_{14} = \frac{N}{\lambda_1}(\frac{v_{11}}{\sqrt{2}} + \frac{v_{12}}{\sqrt{6}} + \frac{4v_{13}}{\sqrt{12}})^2 + \frac{N}{\lambda_2}(\frac{v_{21}}{\sqrt{2}} + \frac{v_{22}}{\sqrt{6}} + \frac{4v_{23}}{\sqrt{12}})^2$$

$$R'_{24} = N_1N_3N_5R_{24} = \frac{N}{\lambda_1}(\frac{v_{11}}{\sqrt{2}} - \frac{v_{12}}{\sqrt{6}} - \frac{4v_{13}}{\sqrt{12}})^2 + \frac{N}{\lambda_2}(\frac{v_{21}}{\sqrt{2}} - \frac{v_{22}}{\sqrt{6}} - \frac{4v_{23}}{\sqrt{12}})^2$$

$$R'_{15} = N_2N_3N_4R_{15} = \frac{N}{\lambda_1}(\frac{v_{11}}{\sqrt{2}} + \frac{v_{12}}{\sqrt{6}} + \frac{v_{13}}{\sqrt{12}} + \frac{5v_{14}}{\sqrt{20}})^2 + \frac{N}{\lambda_2}(\frac{v_{21}}{\sqrt{2}} + \frac{v_{22}}{\sqrt{6}} + \frac{v_{23}}{\sqrt{12}} + \frac{5v_{24}}{\sqrt{20}})^2$$

$$R'_{25} = N_1N_3N_4R_{25} = \frac{N}{\lambda_1}(\frac{v_{11}}{\sqrt{2}} - \frac{v_{12}}{\sqrt{6}} - \frac{v_{13}}{\sqrt{12}} - \frac{5v_{14}}{\sqrt{20}})^2 + \frac{N}{\lambda_2}(\frac{v_{21}}{\sqrt{2}} - \frac{v_{22}}{\sqrt{6}} - \frac{v_{23}}{\sqrt{12}} - \frac{5v_{24}}{\sqrt{20}})^2$$

$$R'_{34} = N_1N_2N_5R_{34} = \frac{N}{\lambda_1}(\frac{2v_{12}}{\sqrt{6}} - \frac{4v_{13}}{\sqrt{12}})^2 + \frac{N}{\lambda_2}(\frac{2v_{22}}{\sqrt{6}} - \frac{4v_{23}}{\sqrt{12}})^2$$

$$R'_{35} = N_1N_2N_4R_{35} = \frac{N}{\lambda_1}(\frac{2v_{12}}{\sqrt{6}} - \frac{v_{13}}{\sqrt{12}} - \frac{5v_{14}}{\sqrt{20}})^2 + \frac{N}{\lambda_2}(\frac{2v_{22}}{\sqrt{6}} - \frac{v_{23}}{\sqrt{12}} - \frac{5v_{24}}{\sqrt{20}})^2$$

$$R'_{45} = N_1N_2N_3R_{45} = \frac{N}{\lambda_1}(\frac{3v_{13}}{\sqrt{12}} - \frac{5v_{14}}{\sqrt{20}})^2 + \frac{N}{\lambda_2}(\frac{3v_{23}}{\sqrt{12}} - \frac{5v_{24}}{\sqrt{20}})^2 \qquad (3.4.44)$$

Again from the geometric representation of the five points, we see that seven of the R' s are independent and the remaining three can be found with the help of the known seven. So again we discard any three of the ten R' s (which can be done in $\binom{10}{3}$ ways) and then to complete a set of eight R' s corresponding to eight v's, we bring in another R', functionally independent of the remaining seven, defined as:

$$R' = \frac{N}{\lambda_1}(\frac{v_{11}}{\sqrt{2}})^2 \qquad (3.4.45)$$

Thus, assuming R'_{34}, R'_{35} and R'_{45} to be the smallest of the ten R's, we discard them and then with the remaining seven R' s and R' in (3.4.45), we conclude that:

$$\prod_{i=1}^{4} \prod_{j=1}^{4} dv_{ij} = \frac{(\lambda_1 \lambda_2)^2}{160N^4} \frac{dR'_{12}\ dR'_{13}\ dR'_{23}\ dR'_{14}\ dR'_{24}\ dR'_{15}\ dR'_{25}\ dR'}{\left[R'(\frac{1}{4}R'_{12} - R')\right]\frac{1}{2}\sqrt{S'_{123}}\sqrt{S'_{124}}\ \sqrt{S'_{125}}} \quad (3.4.46)$$

Making use of (3.4.44), (3.4.46), (3.4.13) and (3.4.42), we get the joint density of the seven R'_{rs} and R'. As was done in (3.4.12), we integrate out R', where $0 < R' \leq \frac{1}{4} R'_{12}$. This yields

$$\binom{10}{3}\left(\frac{1}{2}\right)^{5+1}\left(\frac{n_5}{2\pi}\right)^{5-2} \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{4}\sum_{s=r+1}^{5} R_{rs}\right] dR_{12}dR_{13}dR_{23}dR_{14}dR_{24}dR_{15}dR_{25}}{\sqrt{S_{123}}\sqrt{S_{124}}\sqrt{S_{125}}}$$

(3.4.47)

Here R_{34}, R_{35} and R_{45} are functions of the other R_{rs} and should be expressed in terms of these other R_{rs} in (3.4.47). R_{34}, R_{35} and R_{45} can be determined from the quadrilaterals formed by joining the sets of points (1,2,3,4), (1,2,3,5) and (1,2,4,5) respectively. The variates R_{12}, R_{13}, R_{23}, R_{14}, R_{24}, R_{15}, R_{25} are all positive, and the sets of variates (R_{12}, R_{13}, R_{23}), (R_{12}, R_{14}, R_{24}) and (R_{12}, R_{15}, R_{25}) do not assume values outside the cones $S_{123} \geq 0$, $S_{124} \geq 0$ and $S_{125} \geq 0$ respectively. Thus the density in (3.4.47) is always positive.

The Distribution of the Largest Distance

We again restrict the problem by assuming that $N_r = N_0$ ($r=1, 2, \ldots, 5$).

The joint distribution (3.4.47) reduces to:

$$\binom{10}{3}\left(\frac{1}{2}\right)^6\left(\frac{5}{2\pi}\right)^3 \quad \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{4}\sum_{s=r+1}^{5} R_{rs}\right]}{\sqrt{\bar{S}_{123}}\,\sqrt{\bar{S}_{124}}\,\sqrt{\bar{S}_{125}}} \, dR_{12}dR_{13}dR_{23}dR_{14}dR_{24}dR_{15}dR_{25} \qquad (3.4.48)$$

where again the sets of variates (R_{12}, R_{13}, R_{23}), (R_{12}, R_{14}, R_{24}) and (R_{12}, R_{15}, R_{25}) do not assume values outside the cones $\bar{S}_{123} \geq 0$, $\bar{S}_{124} \geq 0$ and $\bar{S}_{125} \geq 0$ respectively.

We can again assume without loss of generality that R_{12} is the largest of all the seven R's and further that they have been ordered as:

$$0 \leq R_{23} \leq R_{13} \leq R_{12} < \infty$$

$$0 \leq R_{24} \leq R_{14} \leq R_{12} < \infty$$

$$\text{and} \quad 0 \leq R_{25} \leq R_{15} \leq R_{12} < \infty$$

The density of the ordered variates is $7(2!)^3 f(R_{12}, R_{13}, R_{23}, R_{14}, R_{24}, R_{15}, R_{25})$, and the probability $G(t)$ that $R_{12} \leq t$ is:

$$G(t) = \binom{10}{3}\, 7(2!)^3\left(\frac{1}{2}\right)^6\left(\frac{5}{2\pi}\right)^3 \int \!\!\cdots\!\! \int_V \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{4}\sum_{s=r+1}^{5} R_{rs}\right] dR_{12}dR_{13}dR_{23}\cdots dR_{15}dR_{25}}{\sqrt{\bar{S}_{123}}\,\sqrt{\bar{S}_{124}}\,\sqrt{\bar{S}_{125}}} \qquad (3.4.49)$$

V is the region:

$$\text{where } (\sqrt{R_{12}} - \sqrt{R_{13}})^2 \leq R_{23} \leq R_{13}$$

$$\frac{1}{4} R_{12} \leq R_{13} \leq R_{12}$$

$$(\sqrt{R_{12}} - \sqrt{R_{14}})^2 \leq R_{24} \leq R_{14}$$

$$\frac{1}{4} R_{12} \leq R_{14} \leq R_{12}$$

$$(\sqrt{R_{12}} - \sqrt{R_{15}})^2 \leq R_{25} \leq R_{15}$$

$$\frac{1}{4} R_{12} \leq R_{15} \leq R_{12}$$

$$0 \leq R_{12} \leq t$$

$$0 \leq t < \infty \tag{3.4.50}$$

Generalization. For any k

An inspection of (3.4.26), (3.4.37) and (3.4.47) enables us to generalize the joint distribution of R's for any k - the number of bivariate normal populations. To start with we shall have $\binom{k}{2}$ R's from which (2k - 3) geometrically independent R's denoted by R_{12}, R_{13}, ... R_{1k}, R_{23}, R_{24}, ..., R_{2k} can arbitrarily be chosen to complete the k point figure. It should be noted that such a choice can be made in $\binom{\frac{k(k-1)}{2}}{2k-3}$ ways. The remainder $\left[\binom{k}{2} - (2k-3)\right]$ of the R's denoted by R_{34}, ..., R_{3k}; R_{45}, ..., R_{4k}; ..., ; $R_{(k-1)k}$ are again assumed to be the smallest and are discarded. Thus we conclude that the generalization of (3.4.26), (3.4.37) and (3.4.47) is the density

$$\binom{\frac{k(k-1)}{2}}{2k-3}\left(\frac{1}{2}\right)^{k+1}\left(\frac{n_k}{2\pi}\right)^{k-2} \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{k-1}\sum_{s=r+1}^{k} R_{rs}\right] dR_{12}dR_{13}dR_{23}\ldots dR_{1k}dR_{2k}}{\sqrt{S_{123}}\sqrt{S_{124}}\ldots\sqrt{S_{12k}}} \tag{3.4.51}$$

where $R_{34}, \ldots, R_{3k}; R_{45}, \ldots, R_{4k}; \ldots; R_{(k-1)k}$ can be determined as shown in (3.4.7) and (3.4.8), and where the line joining the points 1 and 2 is the common side of the quadrilaterals (1, 2, 3, 4), (1, 2, 3, 5), ..., (1, 2, 3, k); (1, 2, 4, 5) ...(1, 2, 4, k); ...; (1, 2, $\overline{k-1}$, k) respectively. Again the variates $R_{12}, R_{13}, \ldots, R_{1k}, R_{23}, \ldots, R_{2k}$ are all positive, and the sets of variates $(R_{12}, R_{13}, R_{23}), \ldots, (R_{12}, R_{1k}, R_{2k})$ do not assume values outside the cones $S_{123} \geq 0$, $S_{124} \geq 0, \ldots, S_{12k} \geq 0$ respectively.

The Distribution of the Largest Distribution

Assuming again the equality of sample sizes, that R_{12} is the largest and that the variates in each of the sets $(R_{12}, R_{13}, R_{23}), \ldots, (R_{12}, R_{1k}, R_{2k})$ are ordered as in the previous cases, we conclude finally the probability $G(t)$ that $R_{12} \leq t$ is

$$\binom{\frac{k(k-1)}{2}}{2k-3}\left(\frac{1}{2}\right)^{k+1}\left(\frac{k}{2\pi}\right)^{k-2}(2k-3)(2!)^{k-2}\int\cdots\int_V \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{k-1}\sum_{s=r+1}^{k} R_{rs}\right]dR_{12}dR_{13}\ldots dR_{2k}}{\sqrt{\bar{S}_{123}}\sqrt{\bar{S}_{124}}\ldots\sqrt{\bar{S}_{12k}}} \tag{3.4.52}$$

V is the region

$$
\begin{aligned}
\text{where}\quad (\sqrt{R_{12}} - \sqrt{R_{13}})^2 &\leq R_{23} \leq R_{13} \\
\tfrac{1}{4} R_{12} &\leq R_{13} \leq R_{12} \\
(\sqrt{R_{12}} - \sqrt{R_{14}})^2 &\leq R_{24} \leq R_{14} \\
\tfrac{1}{4} R_{12} &\leq R_{14} \leq R_{12} \\
&\text{-------------------------} \\
&\text{-------------------------} \\
(\sqrt{R_{12}} - \sqrt{R_{1k}})^2 &\leq R_{2k} \leq R_{1k} \\
\tfrac{1}{4} R_{12} &\leq R_{1k} \leq R_{12} \\
0 &\leq R_{12} \leq t \\
0 &\leq t < \infty
\end{aligned}
\qquad (3.4.53)
$$

CHAPTER FOUR

ASSIGNING A POPULATION TO ONE OF THE CLUSTERS

4.1 We propose a method for assigning any other individual or population to one of the clusters obtained by any of the methods described in Chapters Two and Three, where the prior fact is known that the individual or population being assigned belongs to one of the clusters. Two alternative approaches have been suggested, both of them being based on the assumption that the populations concerned are normally distributed. The first approach deals with the method of likelihood functions as already discussed in Section (1.3), and the second with the use of T^2 values. Finally an illustration is presented to demonstrate their use.

4.2 Since, by definition, all the populations included in the cluster have identical mean vectors, we can consider the cluster as one population whose mean vector is estimated to be the grand mean vector of that of the populations included in the cluster. Thus, if there are C clusters, we shall imagine them as C distinct populations with their estimated mean vectors as the grand means of those populations which are included in the respective clusters. Let the estimated mean vectors of the C (so-called) populations be given in matrix form as:

$$\bar{Z}^t(C \times p) \equiv \begin{bmatrix} \bar{Z}_{11}, & \bar{Z}_{21}, & \dots, & \bar{Z}_{p1} \\ \bar{Z}_{12}, & \bar{Z}_{22}, & \dots, & \bar{Z}_{p2} \\ \hline \bar{Z}_{1C}, & \bar{Z}_{2C}, & \dots, & \bar{Z}_{pC} \end{bmatrix} \tag{4.2.1}$$

To get the corresponding significant discriminant scores, we post-multiply $\bar{Z}^t(C \times p)$ by the matrix K^t (introduced in part (d) of Section (2.2) and obtain the corresponding matrix $\bar{U}^t$ defined as:

$$\bar{U}^t(C \times p') = (\bar{U}_{it})^t \qquad \begin{matrix} i = 1, 2, \dots, p' \\ t = 1, 2, \dots, C \end{matrix} \tag{4.2.2}$$

Further, if $\left[\bar{X}_1, \dots, \bar{X}_p\right]$ be the mean vector of the sample from another new population which we are trying to assign to one of the clusters, then the corresponding significant discriminant scores can be similarly obtained. We denote them by

$$(\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_{p'}) \tag{4.2.3}$$

4.3:

Discussion of Approaches

(a) Approach I: Use of $\hat{L}$-functions

Since the C so-called populations are normally distributed, we use Rao's procedure for assigning an arbitrary population to one of the multivariate normally distributed populations. We first compute $\hat{L}$-functions in the form already defined in Section (1.2) or in the form obtained below by the use of significant discriminant scores, namely

$$\hat{L}_t = \sum_{i=1}^{p'} \bar{U}_{it}\,\bar{Y}_i - \frac{1}{2}\sum_{i=1}^{p'} \bar{U}_{it}^2, \quad t=1,2,\dots,C. \tag{4.3.1}$$

Then, following Rao, we would assign, ignoring the a priori probabilities, the new population to the sth ($s \leq \mathcal{C}$) "so-called population" (or cluster) if

$$\hat{L}_s - \hat{L}_t > 0 \quad \text{for all } t = 1, 2, \ldots, s-1, s+1, \ldots, \mathcal{C}$$

(b) Approach II: Use of T^2-Statistic

In the previous method we have not been able to assign the probability to our decision. To achieve this aim we propose the following steps:

Step 1. Let the size of the sample drawn from the new normally distributed population be N and try including it in each of the clusters so that the number of populations involved in each cluster increases by one.

Step 2. Compute the statistic $T^2_{k_t+1}$ for $t = 1, 2, \ldots, \mathcal{C}$, and where k_t is the number of populations in the t-th cluster.

Step 3. Include the new population in the sth cluster if

(i) $T^2_{k_s+1} <$ all $T^2_{k_t+1}$ for $t(\neq s) = 1, 2, \ldots, s-1, s+1, \ldots, C$

and (ii) computed $T^2_{k_s+1} \leq$ tabular $T^2_{\alpha k_s+1}$

Note: Since we allow overlappings, we shall include the population in each cluster for which the computed T^2 is non-significant.

4.4 Illustration

To demonstrate the above approaches we continue with the illustration discussed in Chapter Two.

The B.C. Forest Laboratory obtained later a shipment of 7 trees of black cottonwood from some locality. To assign it to one of the clusters on the basis of its static bending property, the same four measurements X_1, X_2, X_3, and X_4 were taken on different locations of each tree, and the following results were obtained:

Sample Size	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
61	982	4.70	2287	4102

The corresponding significant discriminant scores are:

Sample Size	$\bar{Y}_1$	$\bar{Y}_2$	$\bar{Y}_3$
61	0.4794140	1.1417523	0.4540478

Demonstration of Approach I

Considering each cluster to be one population whose mean vector is estimated as the grand mean vector of the populations (species) involved in the corresponding cluster, we write below the mean vectors of each of the seven clusters by use of (4.2.1) and (4.2.2):

	Size	$\overline{U}_1$	$\overline{U}_2$	$\overline{U}_3$
(a)	931	0.66968541	1.27604842	0.62721413
(b)	984	0.76536770	1.19428189	0.71449228
(c)	368	1.08072219	1.17054579	0.58144678
(d)	587	1.01019297	0.92978089	0.42782876
(e)	1266	1.29464479	1.30553049	0.70800378
(f)	264	0.94597083	1.72039748	0.34593229
(g)	158	1.74328671	1.21889759	0.50048469

(4.4.2)

Note: These clusters (a) to (g) have been written in the same order as shown in the end of Chapter Two.

Using (4.3.1), (4.4.1) and (4.4.2) we obtain $\hat{L}$-functions as:

$\hat{L}_{(a)} = 0.82768514$ $\hat{L}_{(d)} = 0.69443164$

$\hat{L}_{(b)} = 0.79361765$ $\hat{L}_{(f)} = 0.58770051$

$\hat{L}_{(c)} = 0.72048219$ $\hat{L}_{(e)} = 0.49183864$

and $\hat{L}_{(g)} = 0.06075676$

Since $\hat{L}_{(a)}$ is greater than all the remaining $\hat{L}$-functions, we would assign the black cottonwood to the cluster (a) i.e. to (2, 5, 6, 8).

Demonstration of Approach II

Combining the new species of black cottonwood with each of the sets of populations already in clusters, we compute T^2-values by the formula (2.2.4) and obtain:

T_5^2 (for 2, 5, 6, 8 and new one) = 24.70

T_5^2 (for 2, 7, 8, 10 and new one) = 29.91

T_4^2 (for 2, 9, 10 and new one) = 30.94

T_5^2 (for 2, 4, 10, 11 and new one) = 34.39

T_2^2 (for 1 and new one) = 37.44

T_5^2 (for 9, 12, 13, 14 and new one) = 43.99

T_2^2 (for 3 and new one) = 77.66

Clearly T_5^2 (for 2, 5, 6, 8 and new species) is less than all the other computed T^2-values and also is the only one non-significant for 16 D.F. and for $\mathcal{L}$ = .05, since the corresponding tabular value is 26.7251. Hence the black cottonwood would naturally be assigned to the cluster of species 2, 5, 6 and 8.

<u>Remark</u>: We have plotted the point representing the new species 'black cottonwood' in Figures 2, 3 and 4. This graphical representation also shows that the new species is close to 2, 5, 6 and 8.

CHAPTER FIVE

DETERMINATION OF CONFIDENCE REGIONS FOR NON-CENTRALITY PARAMETERS CORRESPONDING TO D_2^2 AND T_k^2

and

ANOTHER EXPRESSION FOR T_k^2

5.1 In multivariate analysis of variance, when the hypothesis of the equality of mean vectors in the case of two or more populations is rejected, the need arises to set up confidence limits for the non-centrality parameters corresponding to the statistics used for tests of hypotheses. In Chapter One, Sections (1.1) and (1.2), we have considered using the statistics D_2^2 and T_k^2 for testing the hypotheses of equal mean vectors. Now we discuss the problem of setting up confidence regions for the corresponding non-centrality parameters Δ^2 and τ_k^2. Lastly we shall give another expression for T_k^2 in terms of the sum of weighted Mahalanobis distances.

5.2 Distributions of the Two Statistics in the Non-Central Case

The distribution of D_2^2, both for Studentized and classical cases, is summed up in Section (1.1) for the non-central case $\Delta^2 \neq 0$. As regards Studentized T_k^2, we do not have its exact distribution in compact known standard form even for the central case. The asymptotic expression of a percentage point of the central T_k^2-distribution in terms of corresponding percentage points of central chi-square with $p(k-1)$ D.F. has been given by Ito (1956); this we have already given in Section (1.2). We again write it below but in a different form suitable for our purpose as:

$$T_k^2 \doteq C_1\chi^2 + C_2(\chi^2)^2 + C_3(\chi^2)^3 + C_4(\chi^2)^4 \tag{5.2.1}$$

where $C_1 = 1 + \dfrac{p - n_1 + 1}{2n_2} + \dfrac{7p^2 + 12(1 - n_1)p + (7n_1^2 - 12n_1 + 1)}{24n_2^2}$

$$C_2 = \frac{p + n_1 + 1}{2n_2(n_1p + 2)} + \frac{13p^2 + 24p - 11n_1^2 + 7}{24n_2^2(n_1p + 2)}$$

$$C_3 = \frac{4n_1p^3 + 2(3n_1^2 + 3n_1 + 10)p^2 + 2(2n_1^3 + 3n_1^2 + 17n_1 + 18)p + 4(5n_1^2 + 9n_1 + 2)}{24n_2^2(n_1p + 2)^2(n_1p + 4)}$$

$$C_4 = \frac{6(p - 1)(p - 2)(n_1 + 1)(n_1 + 2)}{24n_2^2(n_1p + 2)^2(n_1p + 4)(n_1p + 6)} \tag{5.2.2}$$

$n_1 = k - 1$,

and n_2 is taken so large so that the cubes and higher powers of $\frac{1}{n_2}$ are negligible.

Although Ito considered only central T_k^2, there is no difficulty in deducing the approximate distribution of non-central T_k^2. If we go carefully through the procedure Ito (1956) followed in arriving at the distribution of central T_k^2, we can easily deduce the distribution for non-central T_k^2. We have only to replace the central chi-square by the non-central chi-square with the same degrees of freedom and non-centrality parameter τ_k^2 defined in Section 1.2. Thus we write for T_k^2, when $\tau_k^2 \neq 0$,

$$T_k^2 \doteq C_1 \chi'^2 + C_2(\chi'^2)^2 + C_3(\chi'^2)^3 + C_4(\chi'^2)^4 \qquad (5.2.3)$$

where χ'^2 is non-central chi-square with p(k-1) D.F. and parameter τ_k^2, and C_1, C_2, C_3, C_4 are defined above in (5.2.2). Further, in the classical case, T_k^2 is again χ'^2 distributed with p(k-1) D.F. and parameter τ_k^2.

5.3 Tabular Values of Non-Central F-Ratio and Chi-Square

The percentage points for both the non-central F-ratio and chi-square with appropriate degrees of freedom and non-centrality parameters are then needed for the above purpose and so we refer to the following:

Non-Central F-Ratio

Wishart (1932) and Tang (1938) have evaluated the probability integral for the non-central F-ratio. Patnaik (1949) has also computed the tables by an easier and approximate method by fitting an F-distribution with the exact first two moments of non-central F-ratio. Thus, for the use of tabular values at the required confidence level, any of the tables given by Wishart, Tang or Patnaik may be referred to.

Non-Central Chi-Square

Fisher (1931) and Garwood have each computed tables of the 5% significant points of non-central chi-square for 1 to 7 D.F. and $\tau_2^2 = \sqrt{\lambda} = 0(0.2)(5.0)$. Patnaik (1949) has also evaluated them by using various approximations to non-central chi-square, which are quite close to exact ones. Thus, for finding the confidence intervals, any of the available tables given by Fisher, Garwood, or Patnaik may be referred to.

5.4 Description of the Method Used for Confidence Regions

We now give the method for determining the confidence regions for either of the parameters Δ^2 or τ_k^2. Since the method used is the same for both, we shall take up only one statistic - Studentized T_k^2. We shall describe fully the procedure for this statistic, and the same technique can be made use of for the other also.

To do this we shall follow Mood's method (Art. 11.5) given for the functions not distributed independently of the **parameters.**

Let, for a pre-assigned $\mathcal{L}$, the confidence level be $100(1-\mathcal{L})\%$. Since, for a given value of $\tau_k^2 = \tau_{k(0)}^2$, the density of T_k^2, which is $g(T_k^2, \tau_{k(0)}^2)$ is completely specified, we can find numbers ϕ_1, ϕ_2, such that:

$$P_r\left[T_k^2 < \phi_1 / \tau_k^2 = \tau_{k(0)}^2 \right] = \int_0^{\phi_1} g\left[T_k^2, \tau_{k(0)}^2\right] dT_k^2 = 1$$

and $P_r\left[T_k^2 > \phi_2 \,/\, \tau_k^2 = \tau_{k(0)}^2\right] = \int_{\phi_2}^{\infty} g\left[T_k^2\,,\ \tau_{k(0)}^2\right] dT_k^2 = \mathcal{L}_2$ (5.4.1)

where $\mathcal{L}_1 + \mathcal{L}_2 = \mathcal{L}$ ($\mathcal{L}_1$, $\mathcal{L}_2$ are two predetermined numbers)

Similarly, for every value of τ_k^2, the pairs of numbers ϕ_1, ϕ_2 can be found which enable us to write ϕ_1, ϕ_2 as functions of τ_k^2 i.e. $\phi_1(\tau_k^2)$ and $\phi_2(\tau_k^2)$ respectively, and finally we state:

$$P_r\left[\phi_1(\tau_k^2) \leq \text{observed } T_k^2 \leq \phi_2(\tau_k^2)\right] = 1 - \mathcal{L} \qquad (5.4.2)$$

Writing $\phi_1(\tau_k^2) = T_k^2,\ \phi_2(\tau_k^2) = T_k^2,$ (5.4.3)

we invert them to obtain respectively:

$$\tau_k^2 = \Psi_1(T_k^2),\ \tau_k^2 = \Psi_2(T_k^2) \qquad (5.4.4)$$

and then rewrite (5.4.2) as:

$$P_r\left[\Psi_2(T_k^2) \leq \tau_k^2 \leq \Psi_1(T_k^2)\right] = 1 - \mathcal{L} \qquad (5.4.5)$$

which determines the region for τ_k^2 for a known value T_k^2 at $(1 - \mathcal{L})\%$ confidence.

Thus to compute the interval for $\tau_{k_1}^2$, corresponding to a known value of $T_{k_1}^2$, we refer meanwhile to the Fig. 5 below and explain the procedure as follows:

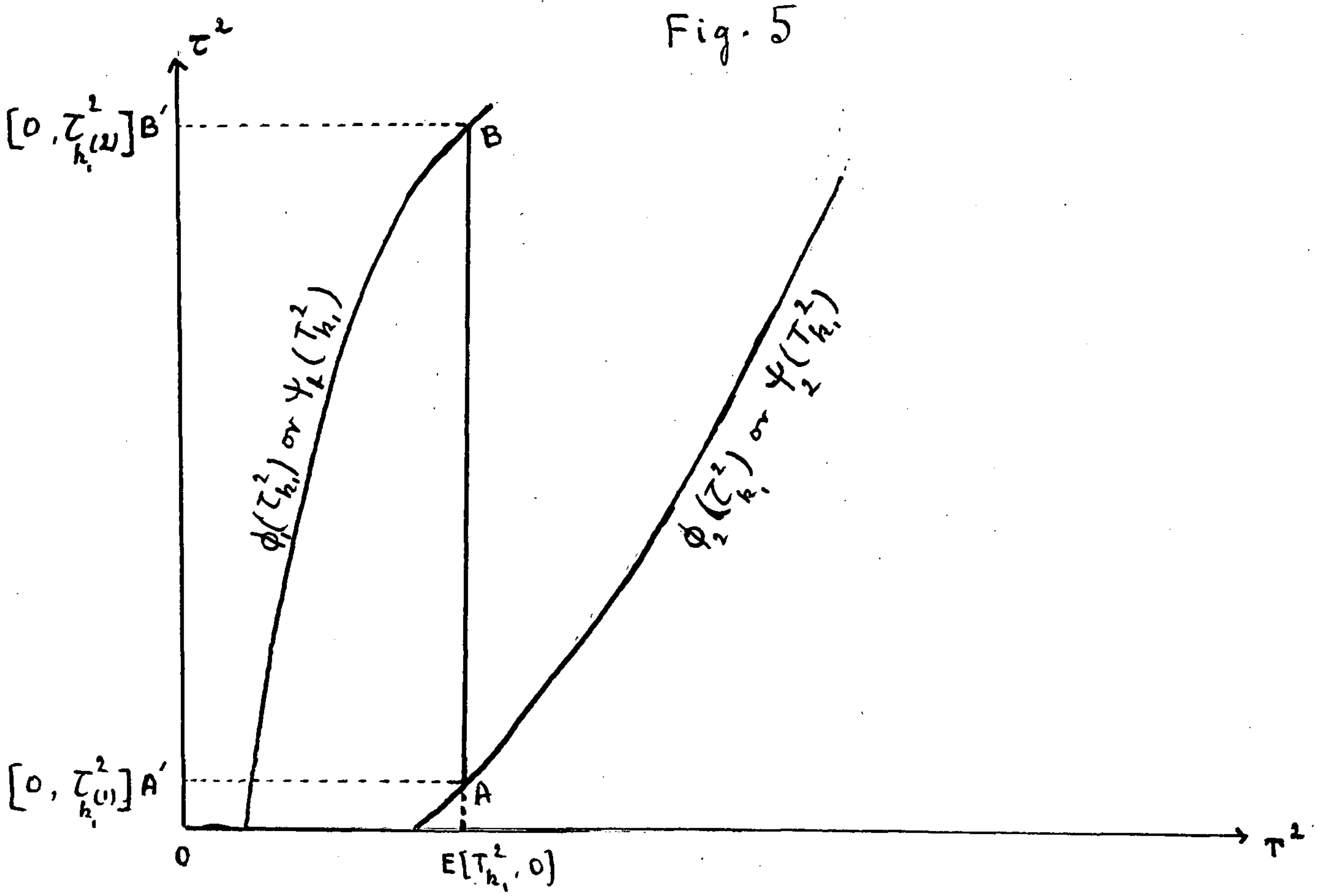
Suppose we have computed $T_{k_1}^2$ on the basis of k_1 populations. Through the point $E\left[T_{k_1}^2, 0\right]$ on T^2-axis, erect a perpendicular to the

T^2-axis and let it cut the curves $\psi_2(T^2_{k_1})$ and $\psi_1(T^2_{k_1})$ respectively at the points A and B. Take A' and B' respectively to be the images of A and B on the τ^2-axis. Then, if the distances of A' and B' from the origin are respectively $\tau^2_{k_1(1)}$ and $\tau^2_{k_1(2)}$, we have:

$$P_r\left[\tau^2_{k_1(1)} \le \tau^2_{k_1} \le \tau^2_{k_1(2)} \Big/ T^2 = T^2_{k_1}\right] = 1 - \mathcal{L} \quad (5.4.6)$$

which determines the region for a known value $T^2_{k_1}$ of T^2 with $100(1-\mathcal{L})\%$ confidence.

Fig. 5



5.5 Example

To make the procedure clearer we present below an example for $p = 4$, $n_1 = k-1 = 1$, $n_2 = 29$, and construct the 90% confidence region for τ_2^2 corresponding to known value of Studentized $T_2^2 = 25$.

Both lower and upper 5% significant points (Fisher and Garwood) of non-central chi-square for D.F. = 1(1)7 and $\sqrt{\lambda}$ $(= \sqrt{\tau_2^2}) = 0(0.2)5.0$ have long since been computed; but, since they were not immediately available to us, we have preferred to compute them by the approximate method suggested by Patnaik (1949), for $\lambda = \tau_2^2 = 0(2)36$ and D.F. $f = p(k-1) = 4$ as follows:

(i) We first select an appropriate percentage point of chi-square as tabled by Hartley and Pearson (1954) and use the 4-point Langrangian formula to get the same percentage point for chi-square with D.F. $= \left(f + \frac{\lambda^2}{f+2\lambda}\right)$, then multiply the result by $\rho = (1 + \frac{\lambda}{f+\lambda})$. The appropriate lower and upper 5% points obtained by the method are recorded respectively in the second and third columns of table 14.

(ii) Then we find the values of C_1, C_2, C_3, and C_4, defined in (5.2.2) for appropriate values of p, n_1 and n_2, which in our case are 1.07432, 0.0197, 0.000198 and 0.0000037 respectively.

(iii) Lastly, substituting the values, obtained above in steps (i) and (ii), in formula (5.2.3), we obtain the corresponding lower and upper 5% tabular values of Studentized T_2^2 and record them respectively in columns 4 and 5 of table 14.

Having obtained these lower and upper 5% tabular values of T_2^2, we plot them on the graph corresponding to respective values of i.e. τ_2^2, and obtain the two curves $\phi_1(\tau_2^2)$ and $\phi_2(\tau_2^2)$ as shown in Fig. 6.

Finally, to find 90% confidence region for computed $T_2^2 = 25$, we erect a perpendicular through the point E(25,0) on the T_2^2-axis and let it cut the curves $\phi_2(\tau_2^2)$ and $\phi_1(\tau_2^2)$ respectively at A and B. We then take A' and B' respectively the images of A and B on τ_2^2-axis. Reading their distances from the origin respectively to be 3.9 and 29.1 approximately, we conclude that:

$$P_r\left[\, 3.9 \leq \tau_2^2 \leq 29.1 \,\middle/\, T_2^2 = 25 \right] = .90 \tag{5.5.1}$$

which determines thus the region for a known value 25 of T_2^2 with 90% confidence.

<u>Note</u>: The non-centrality parameter τ_k^2 involves sample sizes. In order that the non-centrality parameter should contain population constants only, we have to resort to the assumption that the sample sizes are equal i.e. $N_1 = N_2 = \ldots N_k = N$ (say), in which case

$$\tau_k^2 = N \sum_i^p \sum_j \lambda^{ij} \sum_{r=1}^k (\mu_{ir} - \mu_i)(\mu_{jr} - \mu_j) \tag{5.5.2}$$

where $\mu_i = \left(\sum_{r=1}^k \mu_{ir} \right) \Big/ k$

or alternatively that $\tau_k^2 = N\,\mathcal{T}_k^2$ (5.5.3)

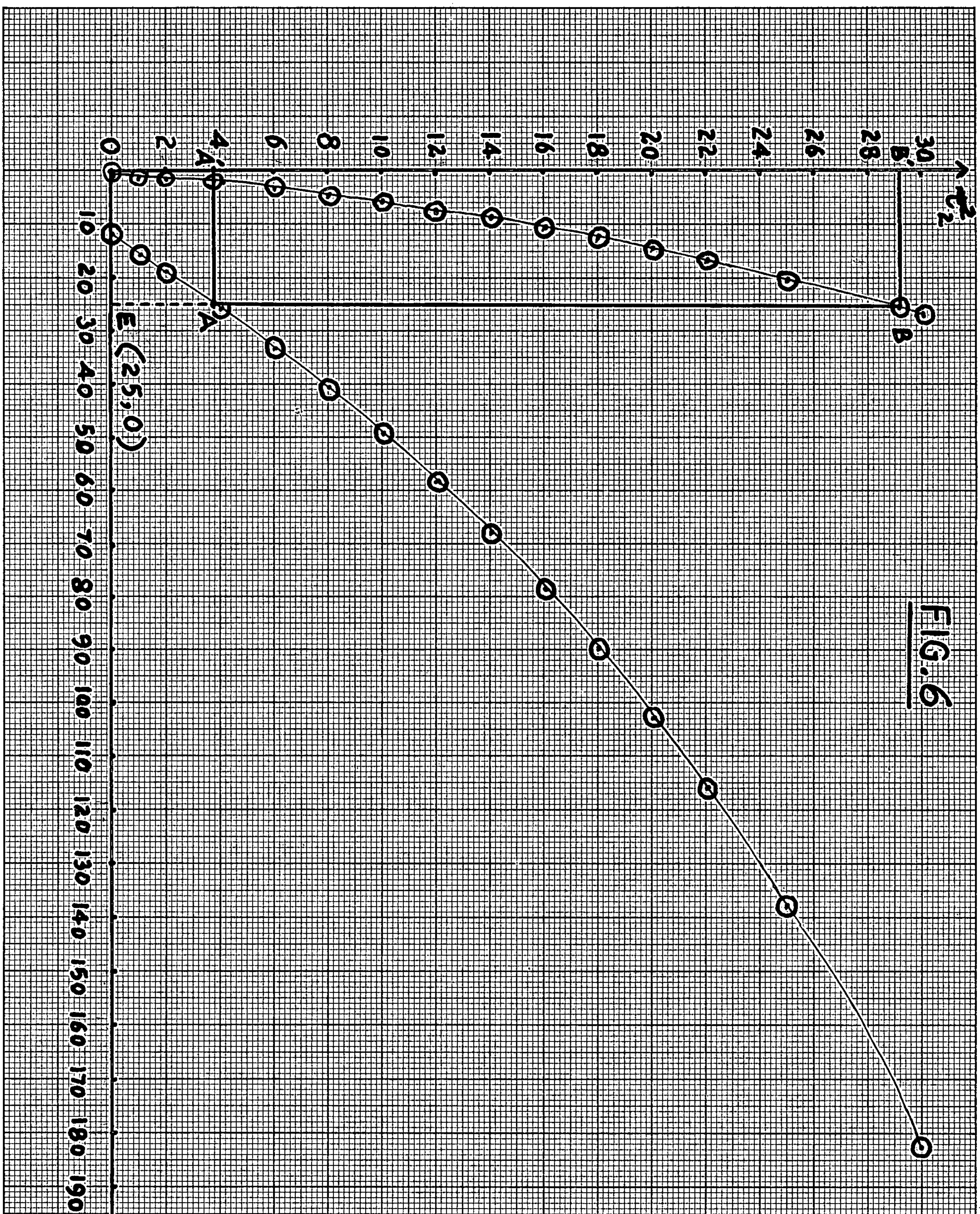
$$\text{where}\quad \mathcal{T}_k^2 = \sum_i^p \sum_j \mathcal{L}^{ij} \sum_{r=1}^k (\mu_{ir} - \mu_i)(\mu_{jr} - \mu_j) \qquad (5.5.4)$$

Thus if we suppose N = 15, say, we can deduce from (5.5.1) the following for $\mathcal{T}_2^2$ as:

$$P\left[\,0.26 \le \tau_2^2 \le 1.94 \,\middle/\, T_2^2 = 25\right] = .90 \qquad (5.5.5)$$

Table 14

λ or τ_2^2	5% chi-square values		5% T_2^2-values	
	Lower	Upper	Lower	Upper
0	0.71	9.49	0.773	12.168
1	0.93	11.72	1.005	15.676
2	1.24	13.72	1.362	19.090
4	1.77	17.31	1.965	25.859
6	2.83	20.77	3.202	33.275
8	3.80	24.08	4.379	41.302
10	4.85	26.97	5.698	49.146
12	5.98	29.93	7.176	58.079
14	7.15	32.85	8.770	67.878
16	8.36	35.69	10.493	78.440
18	9.64	38.44	12.396	89.731
20	10.91	41.29	14.375	102.635
22	12.24	43.96	16.547	115.935
25	14.26	47.94	20.053	138.136
30	17.77	54.55	26.792	182.128
36	22.12	62.24	36.432	246.443



0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
τ_2
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
T_2
A'
B'
A
B
E (25,0)

FIG.6

5.6 An Alternative Expression for T_k^2

We have already given three expressions of T_k^2 in Section 1.2. We now give below another expression as the sum of weighted Mahalanobis distances as:

$$T_k^2 = \sum_{1 \le r < s \le k}\sum \frac{N_r N_s}{\sum_{r=1}^{k} (N_r)} D^2_{(rs)} \qquad (5.6.1)$$

where $D^2_{(rs)}$ is the Mahalanobis distance between the rth and the sth populations. The statement (5.6.1) is proved as follows:

Consider the set of numbers u_r, v_r and the set of integers N_r $(r = 1, 2, \ldots, k)$.

Let $N = \sum_{r=1}^{k} N_r$ and $N\bar{u} = \sum_{r=1}^{k} N_r u_r$, $N\bar{v} = \sum_{r=1}^{k} N_r v_r$.

Let $S(u,v) = \frac{1}{N} \sum_{1 \le n < s \le k}\sum N_r N_s (u_r - u_s)(v_r - v_s)$

Then $S(u,v) = \frac{1}{2N} \sum_{n=1}^{k} \sum_{s=1}^{k} N_r N_s (u_r - u_s)(v_r - v_s)$

$$= \frac{1}{2N} \sum_{n=1}^{k} \sum_{s=1}^{k} N_r N_s \left\{(u_r - \bar{u}) - (u_s - \bar{u})\right\} \left\{(v_r - \bar{v}) - (v_s - \bar{v})\right\}$$

$$= \frac{1}{2N} \sum_{n=1}^{k} \sum_{s=1}^{k} N_r N_s \{(u_r-\bar{u})(v_r-\bar{v})-(u_r-\bar{u})(v_s-\bar{v})-(u_s-\bar{u})(v_r-\bar{v})$$

$$+(u_s-\bar{u})(v_s-\bar{v})\}$$

$$= \frac{1}{2N} \left\{ \sum_{r=1}^{k} N_r(u_r-\bar{u})(v_r-\bar{v})(\sum_{s=1}^{k} N_s) + (\sum_{r=1}^{k} N_r) \sum_{s=1}^{k} N_s(u_s-\bar{u})(v_s-\bar{v}) \right\} - 0$$

Thus $S(u,v) = \frac{1}{N} \sum_{1 \le r < s \le k} \sum N_r N_s (u_r-u_s)(v_r-v_s) = \sum_{r=1}^{k} N_r(u_r-\bar{u})(v_r-\bar{v})$

Now we apply this relationship, taking $u_r = \bar{x}_{ir}$, $v_r = \bar{x}_{jr}$, so that $\bar{u} = \bar{x}_i$, $\bar{v} = \bar{x}_j$. In the Studentized case

$$T_k^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} w^{ij} \sum_{r=1}^{k} N_r(\bar{x}_{ir}-\bar{x}_i)(\bar{x}_{jr}-\bar{x}_j)$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{p} w^{ij} \frac{1}{N} \sum_{1 \le n < s \le k} \sum N_r N_s (x_{ir}-x_{is})(x_{jr}-x_{js})$$

$$= \sum_{1 \le r < s \le k} \sum \frac{N_r N_s}{N} \sum_{i=1}^{p} \sum_{j=1}^{p} w^{ij} (\bar{x}_{ir}-\bar{x}_{is})(\bar{x}_{jr}-\bar{x}_{js})$$

$$= \sum_{1 \le r < s \le k} \sum \frac{N_r N_s}{N} D^2_{(rs)}$$

Let $T^2_{(rs)} = \frac{N_r N_s}{N_r + N_s} D^2_{(rs)}$ denote the corresponding Hotelling T_2^2.

$$\text{Then } T_k^2 = \sum_{1 \le r < s \le k}\sum \frac{N_r N_s}{N} D^2_{(rs)} = \sum_{1 \le r < s \le k}\sum \frac{N_r + N_s}{N} T^2_{(rs)}$$

The same argument works for classical T_k^2, which will be expressed in terms of classical $D^2_{(rs)}$ and $T^2_{(rs)}$, where σ^{ij} replaces w^{ij} throughout.

The same argument works for the parameter τ_k^2.

$$\tau_k^2 = \sum_{i=1}^{p}\sum_{j=1}^{p}\sigma^{ij}\sum_{r=1}^{k} N_r(\mu_{ir}-\mu_i)(\mu_{jr}-\mu_j) = \sum_{1 \le r < s \le k}\sum \frac{N_r N_s}{N} \Delta^2_{(rs)} \text{ ,}$$

$$\text{where } \Delta^2_{(rs)} = \sum_{i=1}^{p}\sum_{j=1}^{p}\sigma^{ij} (\mu_{ir}-\mu_{is})(\mu_{jr}-\mu_{js})$$

Again the relation (5.6.1) can also be expressed in matrix form as:

$$\left[\sum_{r=1}^{k} N_r\right] T_k^2 \equiv [N_1, N_2, \ldots, N_k]\begin{bmatrix} 0 & D^2_{12} & D^2_{13} & \ldots & D^2_{1k} \\ D^2_{21} & 0 & D^2_{23} & \ldots & D^2_{2k} \\ \hline D^2_{k1} & D^2_{k2} & D^2_{k3} & \ldots & 0 \end{bmatrix}[N_1, N_2, \ldots, N_k]^t$$

(5.6.16)

CHAPTER SIX

DISTRIBUTION OF THE DETERMINANT OF THE S.P. MATRIX IN THE NON-CENTRAL LINEAR CASE FOR SOME VALUES OF p

6.1 Let k_1^2 be the non-centrality parameter for the linear case. Then the h-th moment of the determinant $|A|$, where A is the S.P. matrix with n D.F., is rewritten from (1.3.3) as follows:

$$E\left[|A|\right]^h = \left[\prod_{i=1}^{p-1} 2^h \frac{\Gamma(\frac{n-i}{2}+h)}{\Gamma(\frac{n-i}{2})}\right]\left[\exp(-\tfrac{1}{2}k_1^2)\sum_{j=0}^{\infty}\frac{k_1^{2j}\, 2^h\Gamma(\frac{n}{2}+j+h)}{2^j\, j!\,\Gamma(\frac{n}{2}+j)}\right] \tag{6.1.1}$$

The right hand side of (6.1.1) can be interpreted as follows:

(i) $\dfrac{2^h\Gamma(\frac{n-i}{2}+h)}{\Gamma(\frac{n-i}{2})}$ is the h-th moment of $f_i(u_i)$,

where $$f_i(u_i) \equiv \frac{1}{2^{\frac{n-i}{2}}\Gamma(\frac{n-i}{2})}\, u_i^{\frac{n-i}{2}-1}\exp(-\tfrac{1}{2}u_i) \qquad i=1,2,\ldots,p-1 \tag{6.1.2}$$

and (ii) $\exp(-\frac{1}{2}k_1^2)\displaystyle\sum_{j=0}^{\infty}\frac{k_1^{2j}2^h\,\Gamma(\frac{n}{2}+j+h)}{2^j\, j!\,\Gamma(\frac{n}{2}+j)}$ is the h-th

moment of $f_0(u_0)$, where

$$f_0(u_0) \equiv \sum_{j=0}^{\infty}\left[\exp(-\tfrac{1}{2}k_1^2)\frac{k_1^{2j}}{j!\,2^j}\left\{\frac{u_0^{\frac{1}{2}n+j-1}}{2^{\frac{1}{2}n+j}\Gamma(\frac{1}{2}n+j)}\exp(-\tfrac{1}{2}u_0)\right\}\right] \tag{6.1.3}$$

Thus: (i) from (6.1.2), $f_i(u_i)$, $i = 1, 2, \ldots, \overline{p-1}$, are central chi-squares, which we can take to be independently distributed with (n-i) D.F. for u_i;

and (ii) from (6.1.3), $f_0(u_o)$ is a non-central chi-square which we take again to be independently distributed with n D.F. and non-centrality parameter k_1^2.

Since the moment of a product of independent variables is the product of the moments of the variables, it follows that:

$$E\left[|A|^h\right] = E(u_0^h)\, E(u_1^h) \ldots E(u_{p-1}^h) = E\left[(u_0 u_1 u_2 \ldots u_{p-1})^h\right] .$$

Alternatively, therefore, the h-th moment of $|A|$ could be directly determined by multiplying respectively the h-th moments of independent $u_i (i = 0, 1, 2, \ldots, \overline{p-1})$ variates defined above from which one concludes that if one wants to determine the distribution of $|A|$, one can do so by finding the distribution of the product $(u_0 u_1 \ldots u_{p-1})$. Since $u_0, u_1, \ldots, u_{p-1}$ are independent, their joint distribution can be written down and the distribution of $(u_0 u_1 \ldots u_{p-1})$ can further be determined for $p \leq 2, 3,$ and 4 as follows:

The joint distribution of the independent variates u_i $(i = 0, 1, \ldots \overline{p-1})$ can be written as:

$$\left[\prod_{i=1}^{p-1} \frac{1}{\Gamma(\frac{n-i}{2})} \left(\frac{u_i}{2}\right)^{\frac{n-i}{2}-1} \exp(-\tfrac{1}{2} u_i)\, d\left(\frac{u_i}{2}\right)\right]\left[\frac{\exp(-\frac{1}{2} k_1^2)}{\Gamma(\frac{n}{2})} \left(\frac{u_0}{2}\right)^{\frac{1}{2}n-1}\right.$$

$$\left. \exp(-\tfrac{1}{2} u_0) \left(1 + \frac{1}{\frac{n}{2}} \left(\frac{u_0}{2} \cdot \frac{k_1^2}{2}\right) + \frac{1}{\frac{n}{2}(\frac{n}{2}+1)}\, \frac{1}{2!} \left(\frac{u_0}{2} \cdot \frac{k_1^2}{2}\right)^2 + \ldots\ d\left(\frac{u_0}{2}\right)\right)\right] \qquad (6.1.4)$$

where $0 \leq u_i < \infty$ $\qquad i = 0, 1, 2, \ldots, \overline{p-1}$

After a little manipulation and setting $n = 2m + p + 1$ $(p \leq n)$, the joint distribution of $u_i (i = 0, 1, 2, \ldots, \overline{p-1})$ becomes

$$\frac{1}{2^{p\left[m + \frac{p+3}{4}\right]}} \frac{u_{p-1}^{m}\, u_{p-2}^{m+\frac{1}{2}} \cdots u_2^{m+\frac{p-3}{2}}\, u_1^{m+\frac{p-2}{2}}\, u_0^{m+\frac{p-1}{2}}}{\Gamma(m+1)\,\Gamma(m+\frac{3}{2}) \cdots \Gamma(m+\frac{p+1}{2})} \exp\left(-\frac{1}{2}\sum_{i=0}^{p-1} u_i\right)$$

$$\exp(-\tfrac{1}{2}k_1^2)\left[1 + \frac{u_0}{1!}\,\frac{(k_1^2/2)}{2m+p+1} + \frac{u_0^2}{2!}\,\frac{(k_1^2/2)^2}{(2m+p+1)(2m+p+3)} + \cdots\right]\prod_{i=0}^{p-1} du_i \qquad (6.1.5)$$

where $0 \leq u_i < \infty$ $\qquad i = 0, 1, 2, \ldots, \overline{p-1}$.

6.2: Preliminaries

(i) We make use of Legendre's duplication formula for the gamma function, namely of

$$\Gamma(n + \tfrac{1}{2})\,\Gamma(n + 1) = \frac{\sqrt{\pi}\,\Gamma(2n + 1)}{2^{2n}} \qquad (6.2.1)$$

(ii) We list below the standard integrals, derived from various books of integral tables, of which frequent use has been made:

(a) Larsen's book of tables (p. 254) gives

$$\int_0^{\infty} \exp\left[-(x^2 + a^2 x^{-2})\right] dx = \frac{\sqrt{\pi}}{2} \exp(-2a) \qquad (6.2.2)$$

for $a \geqslant 0$.

(b) Bierens de Haan gives in his Table 98 (pp. 143-144) two integrals numbered (5) and (17) as follows:

$$\int_0^\infty a - \tfrac{1}{2} \quad \exp\left[-(px + qx^{-1})\right] dx = \left(\frac{q}{p}\right)^{\frac{a}{2}} \exp(-2\sqrt{pq}) \times$$

$$\times \sqrt{\frac{\pi}{p}} \sum_{n=0}^{\infty} \left[\frac{(a + 1 - n)^{2n/1}}{2^{n/2}(2\sqrt{pq})^n}\right] \tag{6.2.3}$$

and

$$\int_0^\infty x^{-a - \frac{1}{2}} \exp\left[-(px + qx^{-1})\right] dx = \left(\frac{p}{q}\right)^{\frac{a}{2}} \exp\left(-2\sqrt{pq}\right) \times$$

$$\times \sqrt{\frac{\pi}{p}} \sum_{n=0}^{\infty} \left[\frac{(a - n)^{2n/1}}{2^{n/2}(2\sqrt{pq})^n}\right] \tag{6.2.4}$$

<u>Note:</u> In both of these Kramp's notation is used, namely

$$x^{n/h} \equiv x(x + h)(x + 2h) \ldots (x + \overline{n-1}\, h)$$

(c) From Whittaker and Watson's book, we quote two integrals (p. 116, ex. 6 and p. 243, ex. 4):

$$\int_0^\infty \frac{\exp(-t) - \exp(-tz)}{t} dt = \int_0^\infty \frac{\exp(-t^{-1}) - \exp(-zt^{-1})}{t} dt = \log z \tag{6.2.5}$$

where the real part of z is positive;

$$\int_0^1 \frac{\exp(-u^{-1}) + \exp(-u) - 1}{u} du = \gamma = .5772157 \ldots \tag{6.2.6}$$

where γ is known as the Euler constant.

(iii) Evaluation of Certain Integrals by the Use of Differential Equations

(a) Evaluation of I, where

$$I = \int_0^{\infty} x \exp(-2x-2ax^{-1})dx. \qquad (6.2.7)$$

Setting $x = \frac{1}{2}u^{-1}$ and $b = 4a$ in I, one obtains the integral

$$K(b) = \frac{1}{4} \int_0^{\infty} u^{-3} \exp(-u^{-1}-bu)du \qquad (6.2.8)$$

If

$$k(u) = \frac{1}{4} u^{-3} \exp(-u^{-1}) \qquad (6.2.9)$$

then

$$\frac{1}{k} \frac{dk}{du} = (1-3u^2)u^{-2} \qquad (6.2.10)$$

and

$$K(b) = \int_0^{\infty} k(u)\exp(-bu)du \qquad (6.2.11)$$

The function K satisfies a differential equation of the form

$$(c_1+d_2b) \frac{d^2K}{db^2} + (c_1+d_1b) \frac{dK}{db} + (c_0+d_0b)K = 0, \qquad (6.2.12)$$

which after some simplification reduces to

$$b \frac{d^2K}{db^2} - \frac{dK}{db} - K = 0 \qquad (6.2.13)$$

Solving (6.2.13) by Frobenius method of series, we get:

$$K(b) = \left[A + B \log b\right]\left[-\frac{b^2}{2!1!} - \frac{b^3}{3!1!} - \frac{b^4}{4!2!} - \ldots\right]$$

$$+ B\left[1 - b + \frac{1}{2^2} b^2 + \frac{11}{3^2 2^2} b^3 \ldots\right] \tag{6.2.14}$$

To find A and B we proceed as follows:

Set b = 0 in (6.2.8) and in its derivative to get:

$$K(0) = -K'(0) = \frac{1}{4} \tag{6.2.15}$$

Now setting b = 0 in (6.2.14) and then using (6.2.15), we get:

$$B = \frac{1}{4} \tag{6.2.16}$$

However, the substitution of b = 0 in the derivative of K(b) defined in (6.2.14) does not help since by using (6.2.15):

$$-A = \underset{b \to 0}{\mathrm{Lt}} \frac{K'(b) + \frac{1}{4}\left[1 + b \log b + \log b\left(\frac{b^2}{2!2!} + \frac{b^3}{3!2!} + \ldots\right)\right] + O(b)}{b}$$

which is indeterminate.

Again, making use of L'Hospital's Rule,

$$-A - \frac{1}{4} = \underset{b \to 0}{\mathrm{Lt}} \left[K''(b) + \frac{1}{4} \log b\right]$$

$$\text{or } -4A - 1 = \underset{b \to 0}{\mathrm{Lt}} \left[\int_0^{\infty} \left[u^{-1} \exp\left[-(u^{-1} + bu)\right]\right] du + \log b\right]$$

$$= \underset{b \to 0}{\mathrm{Lt}} \left[\int_0^{\infty} \exp\left[-(u^{-1} + bu)\right] \frac{du}{u} + \int_0^{\infty} \frac{\exp(-u) - \exp(-bu)}{u}\, du\right]$$

(by using 6.2.5)

$$\approx \lim_{b\to 0}\left[\int_0^{\infty} \frac{\exp(-u^{-1} - bu) + \exp(-u) - \exp(-bu)}{u}\,du\right]$$

Since (i) $f(b,u) = \dfrac{\exp(-u^{-1} - bu) + \exp(-u) - \exp(-bu)}{u}$ is continuous on the right at $b = 0$ and (ii) for $0 \leq b \leq 1$

$$|f(u,b)| \leq \max\left[|f(u,0)|\ ,\ |f(u,1)|\right]$$

$$\leq \begin{cases} \dfrac{\exp(-u^{-1})}{u} + \dfrac{-\exp(-u)+1}{u} & \text{for } 0 \leq u \leq 1 \\[2ex] \dfrac{\exp(-u)}{u} + \dfrac{1 - \exp(-u^{-1})}{u} & \text{for } 1 \leq u < \infty \end{cases}$$

where each term in the last expression is integrable over the given interval, the order of limit and integration can be interchanged and one gets:

$$-4A - 1 = \int_0^{\infty} \frac{\exp(-u^{-1}) + \exp(-u) - 1}{u}\,du$$

$$= \int_0^{1} \frac{\exp(-u^{-1}) + \exp(-u) - 1}{u}\,du + \int_1^{\infty} \frac{\exp(-v^{-1}) + \exp(-v) - 1}{v}\,dv$$

Now setting $v = \frac{1}{u}$ in the second integral, we obtain by using (6.2.6)

$$-4A - 1 = 2\int_0^{1} \frac{\exp(-u^{-1}) + \exp(-u) - 1}{u}\,du = 2\gamma$$

$$A = \frac{1 + 2\gamma}{4} \qquad (6.2.17)$$

Finally from (6.2.7), (6.2.8), (6.2.14), (6.2.16) and (6.2.17), we get:

$$\int_0^\infty x \exp\left[-2(x + ax^{-1})\right] dx = \left[\frac{(1+2\gamma) - \log 4a}{4}\right]\left[\frac{(4a)^2}{2!0!} + \frac{(4a)^3}{3!1!} + \right.$$

$$\left. + \frac{(4a)^4}{4!2!} + \dots\right] + \frac{1}{4}\left[1 - (4a) + \frac{(4a)^2}{2^2} + \frac{11(4a)^3}{3^2\, 2^2} + \dots\right] \tag{6.2.18}$$

(b) Evaluation of $L_r(a) = 2 \int_0^\infty x^{2r+1} \exp(-x^2 - ax^{-1})dx$ (6.2.19)

for a real and positive and $r = 0, 1, 2, \dots$

The values of successive derivatives at $a = 0$ are:

$$L_r(0) = \Gamma(r + 1),\quad L_r'(0) = -\Gamma(r + \tfrac{1}{2}),\quad L_r''(0) = \Gamma(r)$$

$$L_r'''(0) = -\Gamma(r - \tfrac{1}{2}),\quad L_r^{iv}(0) = \Gamma(r - 1),\quad L_r^{v}(0) = \Gamma(r - \tfrac{3}{2}),\ \text{etc.} \tag{6.2.20}$$

Setting $x = u^{-1}$, we get from (6.2.19):

$$L_r(a) = 2 \int_0^\infty u^{-2r-3} \exp(-u^{-2} - au)du \tag{6.2.21}$$

Consider $\ell_r(u) = 2\, u^{-2r-3} \exp(-u^{-2})$ (6.2.22)

Its differential equation is:

$$\frac{1}{\ell_r}\,\frac{d\ell_r}{du} = \frac{2 - (2r + 3)u^2}{u} \tag{6.2.23}$$

Now $L_r(a) = \int_0^\infty \ell_r(u)\exp(-au)du$ is the solution of the differential equation:

$$(c_3 + d_3 a)\frac{d^3 L_r}{da^3} + (c_2 + d_2 a)\frac{d^2 L_r}{da^2} + (c_1 + d_1 a)\frac{dL_r}{da} + (c_0 + d_0 a)L_r = 0 \tag{6.2.24}$$

$$\text{if} \quad \int_0^\infty \left[(-c_3 u^3 + c_2 u^2 - c_1 u + c_0) + a(-d_3 u^3 + d_2 u^2 - d_1 u + d_0)\right] \ell_r(u)$$

$$\exp(-au)du \equiv 0$$

Proceeding as before, as in part (a), we obtain:

$$c_3 = 2d_3 \ , \quad c_2 = -2rd_3 \ , \quad d_3 \neq 0$$

and
$$c_1 = c_3 = d_0 = d_1 = d_2 = 0$$

Thus $L_r(a)$ satisfies the differential equation:

$$a \frac{d^3 L_r}{da^3} - 2r \frac{d^2 L_r}{da^2} + 2L_r = 0 \tag{6.2.25}$$

To solve (6.2.25) by Frobenius method of series, let:

$$L_r(a) = a^c(b_0 + b_1 a + b_2 a^2 + b_3 a^3 + \ldots) \tag{6.2.26}$$

Substituting it in (6.2.25), we obtain the following:

(i) from indicial equation, $c = 0,\ 1,\ 2(1+r)$ (6.2.27)

(ii) $b_1 = b_3 = b_5 = \ldots = b_{2n+1} = \ldots = 0$ (6.2.28)

and (iii) $b_2 = \dfrac{-2b_0}{(c+2)(c+1)(c-2r)}$

$$b_4 = \frac{2^2 b_0}{(c+4)(c+3)(c+2)(c+1)(c-2r)(c-2r+2)}$$

$$b_6 = \frac{-\,2^3 b_0}{(c+6)(c+5)\ldots(c+1)(c-2r)(c-2r+2)(c-2r+4)}$$

$$b_8 = \frac{2^4 b_0}{(c+8)(c+7)\ldots(c+1)(c-2r)(c-2r+2)(c-2r+4)(c-2r+6)}$$

$$\text{--------------------------------, etc.} \qquad (6.2.29)$$

Evaluation of $L_r(a)$ for Particular Values of r

(i) Setting r = 0, the differential equation (6.2.25) becomes

$$a \frac{d^3 L_0}{da^3} + 2L_0 = 0 \qquad (6.2.30)$$

Making use of results (6.2.26) to (6.2.30), we get:

$$L_0(a) = \left[A_0 + B_0 \log a\right]\left[- \frac{2a^2}{2!} + \frac{4a^4}{4!2} - \frac{8a^6}{6!2\ 4} + \ldots \right]$$

$$+ B_0\left[1 + \frac{2.3}{2^2\ 1^2} a^2 - \frac{4(124)}{4^2\ 3^2\ 2^4\ 1^2} a^4 + \ldots \right]$$

$$+ C_0\left[a - \frac{2}{3!} a^3 + \frac{4}{5!1\ 3} a^5 - \frac{8}{7!1\ 3\ 5} a^7 + \ldots\right] \qquad (6.2.31)$$

With the help of (6.2.20) and remembering that r = 0, we easily obtain from (6.2.31): $B_0 = \Gamma(1), \quad C_0 = - \Gamma(\frac{1}{2})$ (6.2.32)

To find A_0, we differentiate twice (6.2.31) with respect to a, and then, setting a = 0, we obtain:

$$- 2A_0 = \operatorname*{Lt}_{a \to 0} \left[L_0''(a) + 2 \log a\right]$$

$$= \operatorname*{Lt}_{a \to 0} \left[2 \int_0^\infty u^{-1} \exp(-u^{-2} - au)du + 2 \log a\right]$$

Setting $u = t^{-\frac{1}{2}}$ we obtain:

$$-2A_0 = \underset{a \to 0}{\mathrm{Lt}} \left[\int_0^\infty t^{-1} \exp(-t - at^{-\frac{1}{2}})dt + \log a^2 \right]$$

Finally, making use of (6.2.5) we get:

$$= \underset{a \to 0}{\mathrm{Lt}} \left[\int_0^\infty t^{-1} \left[\exp(-t - at^{-\frac{1}{2}}) + \exp(-t^{-1} - a^2 t^{-1}) \right] dt \right]$$

Again an interchange of limit and integration is possible, so we obtain:

$$-2A_0 = \int_0^\infty t^{-1} \left[\exp(-t) + \exp(-t^{-1}) - 1 \right] dt$$

Now proceeding as before in part (a), we get:

$$-2A_0 = 2\gamma, \quad \text{so } A_0 = -\gamma, \text{ the Euler constant.}$$

Thus

$$L_0(a) = (\gamma - \log a)\left(\frac{2a^2}{2!} - \frac{4a^4}{4!2} + \frac{8a^6}{6!2 \cdot 4} - \dots\right)$$

$$+ \left(1 + \frac{2\ 3}{2^2\ 1^2} a^2 - \frac{4(124)}{4^2 \cdot 3^2 \cdot 2^4 \cdot 1^2} a^4 + \dots\right)$$

$$-\sqrt{\pi}\left(a - \frac{2}{3!} a^3 + \frac{4}{5!1 \cdot 3} a^5 - \frac{8}{7!1 \cdot 3 \cdot 5} a^7 + \dots\right) \quad (6.2.33)$$

(ii) Setting $r = 1$, the differential equation (6.2.25) becomes:

$$a \frac{d^3 L_1}{da^3} - 2 \frac{d^2 L_1}{da^2} + 2L_1 = 0 \qquad (6.2.34)$$

Proceeding as above and similarly evaluating the constants with the help of (6.2.5), (6.2.6), (6.2.20) for $r = 1$, we get:

$$L_1(a) = \left[(\gamma + \tfrac{1}{2}) - \log a\right] \left(\frac{2^2}{4!2} a^4 - \frac{2^3}{6!2\cdot 2} a^6 + \frac{2^4}{8!\ 2\cdot 4\cdot 2} a^8 - \ldots\right)$$
$$+ \left(1 + \frac{1}{2} a^2 + \frac{19}{144} a^4 + \ldots\right)$$
$$- \frac{\sqrt{\pi}}{2} \left(a + \frac{2}{3!} a^3 - \frac{2^2}{5!} a^5 + \frac{2^3}{7!} a^7 - \ldots\right) \qquad (6.2.35)$$

(iii) For $r = 2$, the differential equation to be solved is:

$$a \frac{d^3 L_2}{da^3} - 4 \frac{d^2 L_2}{da^2} + 2L_2 = 0 \qquad (6.2.36)$$

Again with the help of (6.2.5), (6.2.6) and (6.2.20) for $r = 2$, the solution of (6.2.36) is

$$L_2(a) = \left[(\gamma + \tfrac{3}{2}) - \Gamma(3) \log a\right]\left[\frac{2^3}{6!2\cdot 4} a^6 - \frac{2^4}{8!(2\cdot 4)\ 2} a^8 + \ldots\right]$$
$$+ \Gamma(3)\left[1 + \frac{1}{4} a^2 + \frac{1}{48} a^4 + \frac{17}{6!10} a^6 + \ldots\right]$$
$$- \Gamma(\tfrac{5}{2}) \left[a + \frac{2}{3!3} a^3 + \frac{2^2}{5!1\cdot 3} a^5 + \ldots\right] \qquad (6.2.37)$$

6.3: Distribution of the Determinant of the S.P. Matrices A up to the Order 4 in the Non-Central Linear Case

Case 1: For $p = 2$, i.e. when A is of order 2 and is positive definite.

Substituting $p = 2$ in (6.1.5), the joint distribution of u_0 and u_1 is:

$$2^{-2(m+\frac{5}{4})} \frac{u_1^m u_0^{m+\frac{1}{2}}}{\Gamma(m+1)\,\Gamma(m+\frac{3}{2})} \exp\left[-\tfrac{1}{2}(u_0+u_1)\right] \exp\left(-\tfrac{1}{2}k_1^2\right)\Big[1 +$$

$$\frac{u_0}{1!}\,\frac{k_1^2/2}{2m+3} + \frac{u_0^2}{2!}\,\frac{(k_1^2/2)^2}{(2m+3)(2m+5)} + \dots\Big]\, du_0\, du_1 \qquad (6.3.1)$$

where $\quad 0 \le u_0,\ u_1 < \infty$

Set $\quad u_1u_0 = V_1^2\ ,\ u_0 = 2V_2^2 \qquad (6.3.2)$

so that $\quad du_0du_1 = 4V_1V_2^{-1}\, dV_1dV_2$

Making use of (6.2.1) and (6.3.2), the distribution (6.3.1) reduces to:

$$\frac{2}{\sqrt{\pi}}\ \frac{V_1^{2m+1}}{\Gamma(2m+2)} \exp\left(-\tfrac{1}{2}\ k_1^2\right) \exp\left(-\frac{V_1^2}{4V_2^2} - V_2^2\right)\Big[1 + \frac{V_2^2}{1!}\ \frac{k_1^2}{2m+3} +$$

$$+ \frac{V_2^4}{2!}\ \frac{k_1^4}{(2m+3)(2m+5)} + \dots\Big] dV_1dV_2 \qquad (6.3.3)$$

where $\quad 0 \le V_1,\ \ V_2 < \infty$

The distribution of $V_1 (= \sqrt{u_0u_1})$ is then:

$$\frac{2}{\sqrt{\pi}}\ \frac{\exp(-\frac{1}{2}k_1^2)\ \ V_1^{2m+1}\ dV_1}{\Gamma(2m+2)}\Big[\int_{V_2=0}^{\infty} \exp\left(-\frac{V_1^2}{4V_2^2} - V_2^2\right)\Big[1 + \frac{V_2^2}{1!}\ \frac{k_1^2}{2m+3} +$$

$$+ \frac{V_2^4}{2!}\ \frac{k_1^4}{(2m+3)(2m+5)} + \dots\Big] dV_2 \qquad (6.3.4)$$

where $\quad 0 \le V_1 < \infty$

Now using (6.2.2) $$\int_{V_2=0}^{\infty} \exp\left(-\frac{V_1^2}{4V_2^2} - V_2^2\right)dV_2 = \frac{\sqrt{\pi}}{2}\exp(-V_1) \quad (6.3.5)$$

For $r \neq 0$, $V_2^2 = t$, the integral $I_r = \int_{V_2=0}^{\infty} V_2^{2r}\exp\left(-\frac{V_1^2}{4V_2^2} - V_2^2\right)dV_2$

reduces to $I_r = \frac{1}{2}\int_0^{\infty} t^{r-\frac{1}{2}}\exp\left(-t - \frac{V_1^2}{4t}\right)dt$, and now using

(6.2.3) we have:

$$I_r = \frac{1}{2}\left(\frac{V_1}{2}\right)^r \exp(-V_1) \sum_{n=0}^{\infty} \frac{(r+1-n)^{2n/1}}{2^{n/2}\, V_1^n}$$

$$= \frac{\sqrt{\pi}}{2}\exp(-V_1)T_r \quad (6.3.6)$$

where $T_r = \left(\frac{V_1}{2}\right)^r \sum_{n=0}^{\infty} \frac{(r+1-n)^{2n/1}}{2^{n/2}\, V_1^n}$

Thus the distribution of $V_1 (= \sqrt{u_0 u_1})$ is

$$\frac{V_1^{2m+1}\exp(-V_1 - \frac{1}{2}k_1^2)}{\Gamma(2m+2)}\left[1 + \frac{T_1}{1!}\frac{k_1^2}{2m+3} + \frac{T_2}{2!}\frac{k_1^4}{(2m+3)(2m+5)} + \cdots\right]dV_1$$

(6.3.7)

where $0 \leq V_1 < \infty$ and $m = \frac{n-3}{2}$

<u>Note</u>: For $k_1^2 = 0$, and $m = \frac{n-3}{2}$, (6.3.7) becomes:

$$\frac{1}{\Gamma(n-1)} V_1^{n-2}\exp(-V_1)dV_1 \quad (6.3.8)$$

which is a gamma variate with parameter (n-1).

Case 2: For p = 3

Substituting p = 3 in (6.1.5), the joint distribution of u_0, u_1, u_2 is:

$$2^{-3(m+\frac{3}{2})} \frac{u_2^m u_1^{m+\frac{1}{2}} u_0^{m+1}}{\Gamma(m+1)\Gamma(m+\frac{3}{2})\Gamma(m+2)} \exp\left(-\frac{1}{2}\sum_{i=0}^{2} u_i - \frac{1}{2}k_1^2\right)$$

$$\left[1 + \frac{u_0}{1!}\frac{(k_1^2/2)}{2m+4} + \frac{u_0^2}{2!}\frac{(k_1^2/2)^2}{(2m+4)(2m+6)} + \dots\right] du_0 du_1 du_2$$

where $0 \le u_0, u_1, u_2 < \infty$ (6.3.9)

Setting $u_2u_1u_0 = V_1$, $u_1u_0^2 = V_2^2$, $u_0 = 2V_3^2$ (6.3.10)

so that $du_0du_1du_2 = 4V_2^{-1}V_3^{-1}\, dV_1dV_2dV_3$

and then making use of (6.2.1), we obtain the distribution of V_1 after a little manipulation as:

$$\frac{V_1^m \exp(-\frac{1}{2}k_1^2)}{\sqrt{\pi}\; 2^{m+\frac{1}{2}}} \frac{1}{\Gamma(m+1)\Gamma(2m+3)} \int_{V_3=0}^{\infty}\int_{V_2=0}^{\infty} \exp\left(-\frac{V_1V_3^2}{2} - \frac{V_2^2}{8V_3^4} - V_3^2\right)\left[1 + \right.$$

$$\left.\frac{V_3^2}{1!}\frac{k_1^2}{2m+4} + \frac{V_3^4}{2!}\frac{k_1^4}{(2m+4)(2m+6)} + \dots\right] dV_3dV_2dV_1 \qquad (6.3.11)$$

where $0 \le V_1 < \infty$

Making use of (6.2.2),

$$\int_0^{\infty} \exp\left(-\frac{V_1V_3^2}{V_2^2} - \frac{V_2^2}{8V_3^4}\right) dV_2 = \sqrt{2\pi}\; V_3^2 \exp\left(-\frac{1}{V_3}\sqrt{\frac{V_1}{2}}\right)$$

Then (6.3.11) reduces to:

$$\frac{V_1^m \exp(-\frac{1}{2}k_1^2)}{2^m \Gamma(m+1)\Gamma(2m+3)} \int_0^{\infty} \exp\left(-\frac{1}{V_3}\sqrt{\frac{V_1}{2}} + V_3^2\right)\left[V_3 + \frac{V_3^3}{1!}\frac{k_1^2}{2m+4} + \frac{V_3^5}{2!}\right.$$

$$\left.\frac{k_1^4}{(2m+4)(2m+6)} + \cdots\right] dV_3 dV_1 \tag{6.3.12}$$

Now making use of the integral (6.2.19) for $r = 0, 1, 2, \ldots$ given respectively in (6.2.33), (6.2.35), (6.2.37), etc., and remembering that a in (6.2.19) is equal to $\sqrt{\frac{V_1}{2}}$, the distribution of $V_1 (= u_0 u_1 u_2)$ is:

$$\frac{V_1^m \exp(-\frac{1}{2}k_1^2)}{2^{m+1}\Gamma(m+1)\Gamma(2m+3)}\left[L_0\left(\sqrt{\frac{V_1}{2}}\right) + \frac{k_1^2}{1!}\frac{L_1(\sqrt{\frac{V_1}{2}})}{2m+4} + \frac{k_1^4}{2!}\frac{L_2(\sqrt{\frac{V_1}{2}})}{(2m+4)(2m+6)} + \cdots\right. dV_1$$

where $\quad 0 \leq V_1 < \infty \quad$, and $m = \frac{n-4}{2}$ (6.3.13)

Note: Substituting $k_1 = 0$ and $m = \frac{n-4}{2}$, the distribution in the central case becomes:

$$\frac{V_1^{\frac{n-4}{2}}}{2^{\frac{n-2}{2}}\Gamma(\frac{n}{2}-1)\Gamma(n-1)} L_0\left(\sqrt{\frac{V_1}{2}}\right) \tag{6.3.14}$$

for $\quad 0 \leq V_1 < \infty$

where $L_0(\sqrt{\frac{V_1}{2}})$ is defined in (6.2.33)

Case 3: For $p = 4$:

Substituting $p = 4$ in (6.1.5), the joint distribution of u_0, u_1, u_2 and u_3 is:

$$2^{-4(m+\frac{7}{4})} \frac{u_3^m u_2^{m+\frac{1}{2}} u_1^{m+1} u_0^{m+\frac{3}{2}}}{\Gamma(m+1)\Gamma(m+\frac{3}{2})\Gamma(m+2)\Gamma(m+\frac{5}{2})} \exp\left(-\frac{1}{2}\sum_{i=0}^{3} u_i - \frac{1}{2}k_1^2\right) \times$$

$$\left[1 + \frac{u_0}{1!}\frac{(k_1^2/2)}{2m+5} + \frac{u_0^2}{2!}\frac{(k_1^2/2)^2}{(2m+5)(2m+7)} + \dots\right] du_0 du_1 du_2 du_3 \quad (6.3.15)$$

where $0 \le u_0,\ u_1,\ u_2,\ u_3 < \infty$

Setting $u_3u_2u_1u_0 = V_1$, $u_2u_1u_0 = 2V_2^2$, $u_1u_0 = V_3^2$, $u_0 = 2V_4^2$ (6.3.16)

so that $du_3du_2du_1du_0 = 8(V_2V_3V_4)^{-1}\, dV_1dV_2dV_3dV_4$

and also making use of (6.2.1), we obtain the distribution of V_1 afteralittle manipulation as follows:

$$\frac{2V_1^m \exp(-\frac{1}{2}k_1^2)dV_1}{\pi\Gamma(2m+2)\Gamma(2m+4)} \int_{V_4=0}^{\infty}\int_{V_3=0}^{\infty}\int_{V_2=0}^{\infty} \exp\left(-\frac{V_1^2}{4V_2^2} - \frac{V_2^2}{V_3^2} - \frac{V_3^2}{4V_4^2} - V_4^2\right)$$

$$\left[1 + \frac{V_4^2}{1!}\frac{k_1^2}{2m+5} + \frac{V_4^4}{2!}\frac{k_1^4}{(2m+5)(2m+7)} + \dots\right] dV_2dV_3dV_4 \quad (6.3.17)$$

where $0 \le V_1 < \infty$

Making use of (6.2.2), we integrate (6.3.17) first with respect to V_2 and obtain:

$$\frac{V_1^m \exp(-\frac{1}{2}k_1^2)\, dV_1}{\sqrt{\pi}\,\Gamma(2m+2)\Gamma(2m+4)} \int_{V_3=0}^{\infty} V_3 \exp\left(-\frac{\sqrt{V_1}}{V_3}\right) \int_{V_4=0}^{\infty} \exp\left(-\frac{V_3^2}{4V_4^2} - V_4^2\right) \times$$

$$\times\left(1 + \frac{V_4^2}{1!}\frac{k_1^2}{2m+5} + \frac{V_4^4}{2!}\frac{k_1^4}{(2m+5)(2m+7)} + \dots\right) dV_3dV_4 \quad (6.3.18)$$

where $0 \le V_1 < \infty$

To integrate with respect to V_4, we evaluate again the first integral as before by using (6.2.2), while in the others we set $V_4^2 = t$ and then, using (6.2.3), we obtain in place of (6.3.18):-

$$\frac{V_1^m \exp(-\frac{1}{2}k_1^2)\, dV_1}{2\,\Gamma(2m+2)\,\Gamma(2m+4)} \int_{V_3=0}^{\infty} V_3 \exp(-\frac{V_1}{V_3} - V_3)\left[1 + \frac{I_1}{1!}\ \frac{k_1^2}{2m+5} + \right.$$

$$\left. + \frac{I_2}{2!}\ \frac{k_1^4}{(2m+5)(2m+7)} + \ldots\right] dV_3 \qquad (6.3.19)$$

for $\qquad 0 \leq V_1 < \infty$

$$\text{where } I_r = (\frac{V_3}{2}) \sqrt{\pi} \exp(-V_3) \sum_{n=0}^{\infty} \frac{(r+1-n)^{2n/1}}{2^{n/2}\ V_3^m} \qquad (6.3.20)$$

Further to evaluate (6.3.19), we have to use either (6.2.3) or (6.2.4) for $p = 1$, $q = \sqrt{V_1}$ and suitable value of a. This determines the distribution of $V_1 (= u_0 u_1 u_2 u_3)$ where it should be remembered that $m = \frac{1}{2}(n-5)$.

<u>Note</u>: For the central case we set $k_1 = 0$ in (6.3.19) and then, making use of (6.2.18), we get the distribution of V_1 a

$$\frac{V_1^{\frac{n-5}{2}}\, dV_1}{\Gamma(n-3)\ \Gamma(n-1)} \left[\left(\frac{(1+2\gamma) - \log a}{2}\right)\left(\frac{a^2}{2!0!} + \frac{a^3}{3!1!} + \frac{a^4}{4!2!} + \ldots\right)\right.$$

$$\left. + \frac{1}{2}(1 - a + \frac{1}{2^2}a^2 + \frac{11}{3^2\ 2^2}\ a^3 + \ldots)\right] \qquad (6.3.21)$$

where $\qquad 0 \leq V_1 < \infty \qquad$ and $a = \sqrt{V_1}$

CHAPTER SEVEN

STATISTICS PROPOSED FOR VARIOUS TESTS OF HYPOTHESES I, II AND III AND THEIR DISTRIBUTIONS IN PARTICULAR CASES

7.1: We list below the statistics, based simultaneously on the roots of both the determinantal equations (1.4.5) and (1.4.6), which can be used to test the hypotheses I, II and III with the suitable use of independent S.P. matrices A and C:

(i) Roy's statistics of largest, smallest and intermediate eigenroots based on the determinantal equation (1.4.5). We can simultaneously propose to include that of the eigenroots based on the determinantal equation (1.4.6).

(ii) Hotelling's T_k^2-statistic defined as:

$$T_k^2 = n_2 tr(C^{-1}A) = n_2 \sum_{i=1}^{\ell} \left(\frac{\theta_i}{1 - \theta_i}\right) = n_2 \sum_{i=1} (\phi_i).$$

(iii) Wilks-Λ-statistic defined as:

$$= |C| \Big/ |A + C| = \prod_{i=1}^{\ell} (1 - \theta_i) = \prod_{i=1}^{\ell} (1 + \phi_i)^{-1}$$

(iv) The Wilks-Lawley U-statistic defined as:

$$U = |A| \Big/ |A + C| = \prod_{i=1}^{\ell} (\theta_i) = \sum_{i=1}^{\ell} \left(\frac{\phi_i}{1 + \phi_i}\right)$$

(v) Pillai's V-statistic defined as:

$$V = tr\left[(A + C)^{-1}A\right] = \sum_{i=1}^{\ell} (\theta_i) = \sum_{i=1}^{\ell} \left(\frac{\phi_i}{1 + \phi_i}\right)$$

(vi) We propose another statistic Y defined as:

$$Y = \frac{|A|}{|C|} = \prod_{i=1}^{\ell} \left(\frac{\theta_i}{1-\theta_i}\right) = \prod_{i=1}^{\ell} (\theta_i)$$

Of course, the distribution of any of the statistics, under the null hypothesis, can be found from either of the joint distributions (1.4.7) and (1.4.9); but it will be more convenient to use (1.4.9) for finding that of Λ, U, V, and either of the two for finding that of Roy's statistics.

We have taken in Section 7.2 the statistics T_k^2 and Y and have been able to give their distributions for $\ell = 2, 3$ in the form of definite integrals. Since the procedure is quite similar for the remaining statistics, we have only listed at the end of the Section 7.2 their respective distributions in the form of definite integrals, again for the cases $\ell = 2, 3$.

Nanda (1948) gives the joint limiting form of (1.4.7), which we have listed under (1.4.10). Following him, the joint limiting form of (1.4.9) is easily proved also to be (1.4.10) by setting $\phi_i = \frac{C_i}{n}$ in (1.4.9) and then letting n tend to infinity.

The fact that the limiting forms of both (1.4.7) and (1.4.9) are the same enables us to conclude that limiting distributions of the statistics Y and U will be the same and also that of T_k^2 and V except for the constant multiplier. The same can be said in the case of Roy's statistics.

In Sections (7.3) and (7.4) we have given another method, different from that of Nanda (1948b), of finding the limiting distributions of Roy's statistics. Further, to demonstrate the method of integration, we have solved some particular cases, giving various values to m, for ℓ = 2, 3, and 4.

Lastly, in Section 7.5, we have first found a new form suitable for finding the limiting distribution of U or Y. Since this form is quite similar to that already obtained in Chapter Six for finding the distribution of the determinant of S.P. matrix, we have only effected certain substitutions in the results obtained in Chapter Six and have been able to deduce the limiting distributions of U or Y for ℓ = 2, 3 and 4.

7.2: Distributions of the Statistics T_k^2 and Y for ℓ = 2, 3, and Further Results

Case I: For $\ell = 2$,

The joint distribution of ϕ_1 and ϕ_2 from (1.4.9) is

$$c(m, n, 2)(\phi_1\phi_2)^m \left[(1 + \phi_1)(1 + \phi_2)\right]^{-m-n-3}(\phi_2 - \phi_1)d\phi_1 d\phi_2 \qquad (7.2.1)$$

$$\text{for} \quad 0 \leq \phi_1 \leq \phi_2 < \infty$$

(i) For Y-statistic: Let,

$$\phi_1\phi_2 = u, \ (1 + \phi_1)(1 + \phi_2) = v \qquad (7.2.2)$$

so that $(\phi_2 - \phi_1)d\phi_1 d\phi_2 = du\, dv$, and the relation (7.2.1) becomes:

$$c(m, n, 2)u^m \, v^{-2(m+n+3)} \, du \, dv \tag{7.2.3}$$

Now the roots ϕ_1, ϕ_2 of the quadratic:

$$x^2 - (v - u - 1)x + u = 0 \tag{7.2.4}$$

are real if $\quad (v - u - 1)^2 \geq 4u$

i.e. if $\quad (1 + \sqrt{u})^2 \leq v$

Then the limits for v and u are given by:

$$(1 + \sqrt{u})^2 \leq v < \infty$$

$$0 \leq u < \infty \tag{7.2.5}$$

The distribution of $u(= \phi_1\phi_2)$ or Y is given by:

$$c(m, n, 2)u^m \, du \int_{v=(1+\sqrt{u})^2}^{\infty} v^{-m-n-3} \, dv$$

where $\quad 0 \leq u < \infty$

$$\text{or by} \quad \frac{2c(m, n, 2)}{m + n + 2} \quad \frac{(\sqrt{u})^{2m+1}}{(1+\sqrt{u})^{2(m+n+2)}} \, d(\sqrt{u}) \tag{7.2.6}$$

where $\quad 0 \leq u < \infty$

Further, for any test of hypothesis, we need to make two forms of substitutions:

If $p = 2(\leq n_1)$, $\quad m = \dfrac{n_1-3}{2}$, $\quad n = \dfrac{n_2-3}{2}$

If $n_1 = 2(\leq p)$, $\quad m = \dfrac{p-3}{2}$, $\quad n = \dfrac{n_2-3}{2}$ $\qquad$ (7.2.7)

Effecting these changes in (7.2.6), we have:

For $p = 2(\leq n_1)$, the distribution (7.2.6) reduces to

$$\frac{\Gamma(n_1 + n_2 - 2)}{\Gamma(n_1-1)\Gamma(n_2-1)} \frac{(\sqrt{u})^{(n_1-1)-1}}{(1+\sqrt{u})^{(n_1-1)+(n_2-1)}} \, d(\sqrt{u}) \qquad (7.2.8)$$

where $0 \leq u < \infty$

which states that $\sqrt{Y}$ $(= \sqrt{\phi_1 \phi_2})$ is distributed as F-ratio with $2(n_1-1)$ and $2(n_2-1)$ D.F.

For $n_1 = 2(\leq p)$ the distribution takes the form:

$$\frac{\Gamma(n_2)}{\Gamma(p-1)\Gamma(n_2-p+1)} \frac{(\sqrt{u})^{(p-1)-1}}{(1+\sqrt{u})^{n_2}} \, d(\sqrt{u}) \qquad (7.2.9)$$

where $0 \leq u < \infty$

which states that $\sqrt{Y}$ $(= \sqrt{\phi_1 \phi_2})$ is also F-distributed with $2(p-1)$, $2(n_2-1)$ D.F.

(ii) For T_k^2-statistic:

Considering now the change $(\phi_1 + \phi_2) = u$, $\phi_1 \phi_2 = v$ (7.2.10) and proceeding similarly as above, the joint distribution (7.2.1) becomes:

$$c(m,\ n,\ 2) v^m (1 + u + v)^{-m-n-3} \, du \, dv \qquad (7.2.11)$$

where $0 \leq v \leq \frac{u^2}{4}$

$0 \leq u < \infty$

Then the distribution of u is:

$$c(m,\ n,\ 2) \int_{v=0}^{\frac{1}{4}u^2} v^m (1 + u + v)^{-m-n-3} \, du \, dv \qquad (7.2.12)$$

where $0 \leq u < \infty$

Setting $v = (1 + u)V_0$, we get in place of (7.2.12), the distribution of u as:

$$c(m,\ n,\ 2)(1 + u)^{-n-2} \, du \int_{V_0=0}^{\frac{u^2}{4(1+u)}} V_0(1 + V_0)^{-m-n-3} \, dV_0 \qquad (7.2.13)$$

where $0 \leq u < \infty$

Again, effecting the changes in (7.2.13) as indicated above in (7.2.7) we have:

<u>For $p = 2(\leq n_1)$</u> the distribution of $u = T_k^2$ for two roots is:

$$\frac{1}{4} \frac{\Gamma(n_1\ n_2-1)}{\Gamma(n_1-1)\,\Gamma(n_2-1)} \frac{du}{(1+u)^{\frac{n_2\ 1}{2}}} \int_{V_0=0}^{\frac{u^2}{4(1+u)}} V_0^{\frac{n_1-3}{2}} (1+V_0)^{-\frac{n_1+n_2}{2}} \, dV_0 \qquad (7.2.14)$$

where $0 \leq u < \infty$

The integral involved is an incomplete beta function which can be easily evaluated.

<u>For $n_1 = 2(\leq p)$</u> the distribution of $u(= T_k^2)$ for two eigenroots is:

$$\frac{1}{4} \frac{\Gamma(n_2 + 1)}{\Gamma(p-1)\,\Gamma(n_2-p\ 1)} \frac{du}{(1+u)^{\frac{n_2-p+3}{2}}} \int_{V_0=0}^{\frac{u^2}{4(1+u)}} V_0^{\frac{p-3}{2}} (1+V_0)^{-\frac{n_2+2}{2}} \, dV_0 \qquad (7.2.15)$$

where $0 \leq u < \infty$

and again the integral involved is an incomplete beta-function which can be easily evaluated.

<u>Case II:</u> <u>For $\ell = 3$</u>

The joint distribution of ϕ_1, ϕ_2, ϕ_3 from (1.4.9) is

$$c(m,n,3)(\phi_1\phi_2\phi_3)^m \left[(1+\phi_1)(1+\phi_2)(1+\phi_3)\right]^{-m-n-4}$$

$$\prod_{i=2}^{3}\prod_{j=1}^{i-1}(\phi_i - \phi_j) \prod_{i=1}^{3} d\phi_i \tag{7.2.16}$$

for $0 \leq \phi_1 \leq \phi_2 \leq \phi_3 < \infty$

For finding the distributions of both the statistics Y and T_k^2 for three eigenroots, we effect the following changes:

$$\phi_1 + \phi_2 + \phi_3 = u, \quad \phi_1\phi_2 + \phi_1\phi_3 + \phi_2\phi_3 = v, \text{ and } \phi_1\phi_2\phi_3 = w \tag{7.2.17}$$

so that $(\phi_3-\phi_2)(\phi_2-\phi_1)(\phi_3-\phi_1)d\phi_1 d\phi_2 d\phi_3 = du\, dv\, dw$

Then (7.2.16) reduces to:

$$c(m,n,3)w^m(1+u+v+w)^{-m-n-4}\, du\, dv\, dw \tag{7.2.18}$$

where ϕ_1,ϕ_2,ϕ_3 are the roots of the cubic:

$$x^3 - ux^2 + vx - w = 0 \tag{7.2.19}$$

(i) <u>For Y-statistic:</u>

In order for the roots of the cubic (7.2.19) to be real and positive, we know, from the Appendix B, Form II, the limits on u, v, and w respectively must be the following:

$$
\begin{array}{lll}
0 \le w < \infty & \text{and} & 0 \le w < \infty \\
3w^{2/3} \le v \le 3w^{2/3}(1+\sqrt{3}) & & 3w^{2/3}(1+\sqrt{3}) \le v < \infty \\
\beta_3 \le u \le \beta_4 & & \beta_3' \le u \le \beta_4'
\end{array}
\tag{7.2.20}
$$

Thus the distribution of $w(= \phi_1\phi_2\phi_3) = Y$ from (7.2.18) and (7.2.20) is:

$$
c(m,n,3)w^m \int_v \int_u (1+u+v+w)^{-m-n-4}\, du\ dv\ dw \tag{7.2.21}
$$

where u, v, w are defined as in (7.2.20).

Effecting another change in (7.2.21) as follows:

$$
v = (1+w)V_1, \quad u = (1+w)(1+V_1)U_1 \tag{7.2.22}
$$

so that $du\,dv = (1+w)^2(1+V_1)\,dV_1\,dU_1$, we get in place of (7.2.21):

$$
c(m,n,3)\ \frac{w^m}{(1+w)^{m+n+2}} \int_{V_1} \frac{dV_1}{(1+V_1)^{m+n+3}} \int_{U_1} \frac{dU_1}{(1+U_1)^{m+n+4}}\ dw \tag{7.2.23}
$$

for

$$
\begin{array}{lll}
0 \le w < \infty & \text{and} & 0 \le w < \infty \\
\dfrac{3w^{2/3}}{1+w} \le V_1 \le \dfrac{3w^{2/3}(1+\sqrt{3})}{1+w} & & \dfrac{3w^{2/3}(1+\sqrt{3})}{1+w} \le V_1 < \infty \\
\dfrac{\beta_3}{(1+w)(1+V_1)} \le U_1 \le \dfrac{\beta_4}{(1+w)(1+V_1)} & & \dfrac{\beta_3}{(1+w)(1+V_1)} \le U_1 \le \dfrac{\beta_4}{(1+w)(1+V_1)}
\end{array}
$$

Further, for any test of hypothesis, we need to make following two kinds of changes for m, n in (7.2.23) as given below

$$\text{If } p = 3(\leq n_1), \quad m = \frac{n_1-4}{2}, \quad n = \frac{n_2-4}{2}$$

$$\text{and if } n_1 = 3(\leq p), \quad m = \frac{p-4}{2}, \quad n = \frac{n_2-4}{2} \tag{7.2.24}$$

(ii) For T_k^2-statistic:

In order that the roots of the cubic (7.2.19) be real and positive, we write down the conditions respectively for u, v and w, derived in Appendix B, Form I, as:

$$\text{(a)} \quad 0 \leq u < \infty$$

$$0 \leq v \leq \frac{u^2}{4}$$

$$\text{and} \quad 0 \leq w \leq \beta_2 \tag{7.2.25}$$

$$\text{and (b)} \quad 0 \leq u < \infty$$

$$\frac{1}{4}u^2 \leq v \leq \frac{1}{3}u^2$$

$$\beta_1 \leq w \leq \beta_2 \tag{7.2.26}$$

Thus the distribution of $u(= \phi_1 + \phi_2 + \phi_3) = T_k^2$ for 3 eigenroots, from (7.2.18) with the help of (7.2.25) and (7.2.26) is:

$$\text{(i)} \quad c(m,n,3) \int_v \int_w w^m(1+u+v+w)^{-m-n-4}\, dw\, dv\, du \tag{7.2.27}$$

$$\text{and (ii)} \quad c(m,n,3) \int_v \int_w w^m(1+u+v+w)^{-m-n-4}\, dw\, dv\, du \tag{7.2.28}$$

with limits in (i) and (ii) given by (a) and (b) above, respectively. Effecting another change for both (7.2.27) and (7.2.28) as:

$$v = (1 + u)V_2$$

$$w = (1 + u)(1 + V_2)U_2 \tag{7.2.29}$$

we get respectively as the distribution of $u = T_k^2$ for 3 eigen-roots:

$$(1)\ c(m,n,3)\ \frac{du}{(1+u)^{n+2}} \int_{V_2} \int_{U_2} \frac{U_2^m}{(1+V_2)^{n+3}\,(1+U_2)^{m+n+4}}\, dU_2 dV_2 \qquad (7.2.30)$$

where $0 \le u < \infty$

$$0 \le V_2 \le \frac{u^2}{4(1+u)}$$

$$0 \le U_2 \le \frac{\beta_2}{(1+u)(1+V_2)} \qquad (7.2.31)$$

and where v used in β_1 and β_2 is equal to $(1+u)V_2$, and

$$(2)\ c(m,n,3)\ \frac{du}{(1+u)^{n+2}} \int_{V_2} \int_{U_2} \frac{U_2^m}{(1+V_2)^{n+3}(1+U_2)^{m+n+4}}\, dU_2 dV_2 \qquad (7.2.32)$$

where $0 \le u < \infty$

$$\frac{u^2}{4(1+u)} \le V_2 \le \frac{u^2}{3(1+u)}$$

$$\frac{\beta_1}{(1+u)(1+V_2)} \le U_2 \le \frac{\beta_2}{(1+u)(1+V_2)} \qquad (7.2.33)$$

Finally, for any test of hypothesis, we need to make 2 types of changes as indicated in (7.2.24) for m and n in (7.2.30) and (7.2.32).

Distributions of Other Statistics

Since the method for the other statistics is quite similar to that used above, we give below only the final results.

Case I: For $\ell = 2$

(i) For U-statistic

The distribution of $u (= \theta_1 \theta_2$ or $U)$ for two eigenroots is:

$$\frac{2c(m,n,2)}{n+1} (\sqrt{u})^{2m+1} (1-\sqrt{u})^{2n+2} \, d(\sqrt{u}) \tag{7.2.34}$$

where $0 \le u \le 1$

(ii) For V-statistic

The distribution of $u (= \overline{\theta_1 + \theta_2}$ or $V^{(2)})$ is:

$$c(m,n,2)(1-u)^{m+n+1} \, du \int_{V_3} V_3^m (1+V_3)^n \, dV_3 \tag{7.2.35}$$

where $0 \le u \le 1$

$$0 \le V_3 \le \frac{u^2}{4(1+u)} \tag{7.2.36}$$

(iii) For Λ-statistic

The distribution of $u (= \overline{1-\theta_1}\;\overline{1-\theta_2})$ or Λ for two eigenroots is:

$$\frac{2c(m,n,2)}{m+1} (\sqrt{u})^{2n+1} (1-\sqrt{u})^{2m+2} \, d(\sqrt{u}) \tag{7.2.37}$$

where $0 \le u \le 1$

Further, for any test of hypothesis, the changes of the type indicated in (7.2.7) are possible.

Case II: For $\ell = 3$

(i) For U-statistic

The distribution of $w (= \theta_1 \theta_2 \theta_3$ or $= U$, three eigenvalues) is:

$$c(m,n,3)w^m(1-w)^{n+2} \int_{V_4} \int_{U_4} (1+V_4)^{n+1}(1-U_4)^n \, dV_4 \, dU_4 \, dw \qquad (7.2.38)$$

where $0 \le w \le 1$ and $0 \le w \le 1$

$$\frac{3w^{2/3}}{1-w} \le V_4 \le \frac{3w^{2/3}(1+\sqrt{3})}{1-w} \qquad \frac{3w^{2/3}(1+\sqrt{3})}{1-w} \le V_4 \le 1$$

$$\frac{\beta_3}{(1-w)(1+V_4)} \le U_4 \le \frac{\beta_4}{(1-w)(1+V_4)} \qquad \frac{\beta'_3}{(1-w)(1+V_4)} \le U_4 \le \frac{\beta'_4}{(1-w)(1+V_4)}$$

(7.2.39)

where β_3 and β_4 are defined in Appendix B, Form II and v used in them is equal to $(1-w)V_4$.

(ii) The distribution of $u(= \theta_1 + \theta_2 + \theta_3$ or V) is:

(1) $$c(m,n,3)(1-u)^{m+n+2} \, du \int_{V_5} \int_{U_5} U_5^m(1-U_5)^n(1+V_5)^{m+n+1} \, dU_5 dV_5 \qquad (7.2.40)$$

where $0 \le u \le 1$

$$0 \le V_5 \le \frac{u^2}{4(1-u)}$$

$$0 \le U_5 \le \frac{\beta_2}{(1-u)(1+V_5)} \qquad (7.2.41)$$

and (2):

$$c(m,n,3)(1-u)^{m+n+2} \, du \int_{V_5} \int_{U_5} U_5^m(1-U_5)^n(1+V_5)^{m+n+1} \, dU_5 dV_5 \qquad (7.2.42)$$

where $0 \le u \le 1$

$$\frac{u^2}{4(1-u)} \le V_5 \le \frac{u^2}{3(1-u)}$$

$$\frac{\beta_1}{(1-u)(1+V_5)} \leq U_5 \leq \frac{\beta_2}{(1-u)(1+V_5)} \tag{7.2.43}$$

where further β_1 and β_2 are defined in Appendix B, Form I, and v used in them is equal to $(1-u)V_5$.

(iii) The distribution of $w\left[= (1-\theta_1)(1-\theta_2)(1-\theta_3)\right]$ or Λ for three eigen-roots is:

$$c(m,n,3)w^n(1-w)^{m+2} \int_{V_6} \int_{U_6} (1+V_6)^{m+1}(1-U_6)^m \, dV_6 \, dU_6 \, dw \tag{7.2.44}$$

where $\quad 0 \leq w \leq 1 \quad$ and $\quad 0 \leq w \leq 1$

$$\frac{3w^{2/3}}{1-w} \leq V_6 \leq \frac{3w^{2/3}(1+\sqrt{3})}{1-w}, \qquad \frac{3w^{2/3}(1+\sqrt{3})}{1-w} \leq V_6 \leq 1$$

$$\frac{\beta_3}{(1-w)(1+V_6)} \leq U_6 \leq \frac{\beta_4}{(1-w)(1+V_6)}, \qquad \frac{\beta_3'}{(1-w)(1+V_6)} \leq U_6 \leq \frac{\beta_4'}{(1-w)(1+V_6)} \tag{7.2.45}$$

where again β_3 and β_4 are defined in Appendix B, Form II, and v used in them is equal to $(1-w)V_6$.

Finally for all these three parts, under any test of hypothesis, the suitable changes for m, n indicated in (7.2.24) can be effected.

7.3: Distribution of the Smallest Eigen-root in the Limiting Case:

The joint limiting distribution of the eigen-roots c_i of the determinantal equations (1.4.7) and (1.4.9) given in (1.4.10) is re-written as:

$$K(\ell,m) \prod_{i=1}^{\ell} c_i^m \exp(-c_i) \prod_{i=2}^{\ell}\prod_{j=1}^{i-1} (c_i - c_j) \prod_{i=1}^{\ell} dc_i \qquad (7.3.1)$$

$$0 \leq c_1 \leq c_2 \quad \dots \leq c_\ell < \infty \qquad \ell = \min.(p, n_1)$$

where $K(\ell,m) = \pi^{\ell/2} \Big/ \prod_{i=1}^{\ell} \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma\left(\frac{i}{2}\right)$ (7.3.2)

The distribution of the smallest eigenroot c_1 is:

$$P_r(c_1 \geq x) = K(\ell,m) \int_{c_1=x}^{\infty} \int_{c_2=c_1}^{\infty} \dots \int_{c_\ell=c_{\ell-1}}^{\infty} \prod_{i=1}^{\ell} c_i^m \exp(-c_i)$$

$$\prod_{i=2}^{\ell}\prod_{j=1}^{i-1} (c_i - c_j) \prod_{i=1}^{\ell} dc_i \qquad (7.3.3)$$

Set

$$c_\ell = c_1 u_1 u_2 \dots u_{\ell-2} u_{\ell-1}$$

$$c_{\ell-1} = c_1 u_1 u_2 \dots u_{\ell-2}$$

$$c_3 = c_1 u_1 u_2$$

$$c_2 = c_1 u_1 \qquad (7.3.4)$$

Then:

$$P_r(c_1 \geq x) = K(\ell,m) \int_{c_1=x}^{\infty} \int_{u_1=1}^{\infty} \dots \int_{u_{\ell-1}=1}^{\infty} \left[c_1^{\ell m + \frac{(\ell-1)(\ell-2)}{2}} \exp(-c_1) \right]$$

$$\left[u_1^{m(\ell-1) + \frac{(\ell-2)(\ell-1)}{2}} (u_2-1) \exp(-c_1 u_1) \right] \times$$

$$*\left[u_2^{m(\ell-2)+\frac{(\ell-3)(\ell)}{2}}(u_2-1)(u_1u_2-1)\exp(-c_1u_1u_2)\right]\cdots\left[u_{\ell-2}^{2m+2}(u_{\ell-2}-1)\right.$$

$$\left.(u_{\ell-2}u_{\ell-3}-1)\ldots(u_{\ell-2}u_{\ell-3}\ldots u_1-1)\exp(-c_1u_1u_2\ldots u_{\ell-2})\right]\left[u_{\ell-1}^{m}(u_{\ell-1}-1)\right.$$

$$\left.(u_{\ell-1}u_{\ell-2}-1)(u_{\ell-1}u_{\ell-2}u_{\ell-3}-1)\ldots(u_{\ell-1}u_{\ell-2}\ldots u_1-1)\exp(-c_1u_1u_2\ldots u_{\ell-1})\right]$$

$$du_{\ell-1}du_{\ell-2}\ldots du_1dc_1 \tag{7.3.5}$$

We have evaluated below the integrals for $\ell = 2, 3,$ and 4. The same method can be extended to any value of ℓ .

Symbols and Notations

(i) $$\mathcal{J}(n,a) = \int_1^\infty x^n \exp(-ax) = \frac{\exp(-a)}{a}\left[1 + \frac{n}{a} + \frac{n(n-1)}{a^2} + \ldots \frac{n!}{a^n}\right] \tag{7.3.6}$$

(ii) $$\mathcal{J}(n,p,q,r,\ldots;a) = \int_1^\infty x^n\exp(-ax)dx\left[\int_1^\infty y^p\exp(-axy)dy\right.$$
$$\left.\int_1^\infty z^q\exp(-axyz)dz(\ldots)\right] \tag{7.3.7}$$

(iii) $$\mathop{\mathrm{T}}_n^a\left(\mathcal{J}(p_1,p_2;ax) \pm \mathcal{J}(q_1,q_2;ax) \pm \ldots\right)$$
$$= \int_1^\infty x^n \exp(-ax)\left[\int_1^\infty y_1^{p_1} \exp(-axy_1)\left(\int_1^\infty y_2^{p_2} \exp(-axy_1y_2)dy_2\right)dy_1\right.$$
$$\left.\pm \int_1^\infty z_1^q \exp(-axz_1)\left(\int_1^\infty z_2^{q_2} \exp(-axz_1z_2)\,dz_2\,dz_1\right) \pm \ldots\right] dx \tag{7.3.8}$$

$$= \mathcal{J}(n,p_1,p_2;a) \pm \mathcal{J}(n,q_1,q_2;a) \pm \ldots \qquad (7.3.9)$$

and etc.

Case I: Substituting $\ell = 2$ in (7.3.5):

$$P_r(c_1 \geq x) = K(2,m) \int_{c_1=x}^{\infty} \int_{u_1=1}^{\infty} c_1^{2m+2} \exp(-c_1) \quad u_1^m(u_1-1)\exp(-c_1u_1)dc_1du_1 \qquad (7.3.10)$$

Making use of (6.2.1), (7.3.2) and after a little manipulation, we deduce from (7.3.10):

$$P_r(c_1 \geq x) = \frac{2^{2m+1}}{\Gamma(2m+2)} \int_{c_1=x}^{\infty} c_1^{2m+2} \exp(-c_1)\left[\mathcal{J}(m+1,c_1) - \mathcal{J}(m,c_1)\right]dc_1$$

Using (7.3.6) and simplifying, we get:

$$P_r(c_1 \geq x) = \frac{2^{2m+1}}{\Gamma(2m+2)} \int_{c_1=x}^{\infty} c_1^{2m} \exp(-2c_1)\left[1 + 2\frac{m}{c_1} + 3\frac{m(m-1)}{c_1^2} + \ldots + (m+1)\frac{m!}{c_1^m}\right] dc_1 \qquad (7.3.11)$$

which can be easily evaluated for successive substitutions of m=0, 1, 2, ...

Case II: For $\ell = 3$

Substituting $\ell = 3$ in (7.3.5), we get:

$$P_r(c_1 \geq x) = K(3,m) \int_{c_1=x}^{\infty}\int_{u_1=1}^{\infty}\int_{u_2=1}^{\infty} \left[c_1^{3m+5}\exp(-c_1)\right]\left[u_1^{2m+2}(u_1-1)\exp(-c_1u_1)\right]\left[u_2^m(u_2-1)(u_1u_2-1)\exp(-c_1 u_1u_2)\right] du_1du_2dc_1 \qquad (7.3.12)$$

Making use of (7.3.2) for $\ell = 3$ and then (6.2.1), we obtain from (7.3.12) after re-arrangement of terms:

$$P\,(c_1 \geqslant x) = \frac{2^{2m+3}}{\Gamma(m+1)\,\Gamma(2m+3)} \int_{c_1 = x}^{\infty} c_1^{3m+5} \exp(-c_1)$$

$$\times \Bigg[\mathop{T}_{2m+4}^{c_1} \Big\{ \mathcal{J}(m+2;\, u_1 c_1) - \mathcal{J}(m+1,\, u_1 c_1) \Big\}$$

$$+ \mathop{T}_{2m+3}^{c_1} \Big\{ \mathcal{J}(m; u_1 c_1) - \mathcal{J}(m\ 2,\, u_1 c_1) \Big\}$$

$$+ \mathop{T}_{2m\ 2}^{c_1} \Big\{ \mathcal{J}(m+1; u_1 c_1) - \mathcal{J}(m,\, u_1 c_1) \Big\} \Bigg]\, dc_1 \qquad (7.3.13)$$

Now we explain below how to make use of (7.3.13) to get the probabilities for particular values of m.

(i) <u>For m = 0</u>

$$P_r(c_1 \geqslant x) = 4 \int_{c_1 = x}^{\infty} c_1^{5} \exp(-c_1) \Bigg[\mathop{T}_{4}^{c_1} \Big\{ \mathcal{J}(2, u_1 c_1) - \mathcal{J}(1, u_1 c_1) \Big\}$$

$$+ \mathop{T}_{3}^{c_1} \Big\{ \mathcal{J}(0, u_1 c_1) - \mathcal{J}(2, u_1 c_1) \Big\} + \mathop{T}_{2}^{c_1} \Big\{ \mathcal{J}(1, u_1 c_1) - \mathcal{J}(0, u_1 c_1) \Big\} \Bigg]\, dc_1 \qquad (7.3.14)$$

Using (7.3.6) we obtain:

$$\mathcal{J}(2, u_1 c_1) - \mathcal{J}(1, u_1 c_1) = \frac{\exp(-u_1 c_1)}{u_1^2 c_1^2} \left(1 + \frac{2}{u_1 c_1}\right)$$

$$\mathcal{J}(0, u_1 c_1) - \mathcal{J}(2, u_1 c_1) = \frac{2\exp(-u_1 c_1)}{u_1^2 c_1^2} \left(1 + \frac{1}{u_1 c_1}\right)$$

$$\mathcal{J}(1,u_1c_1) - \mathcal{J}(0,u_1c_1) = \frac{\exp(-u_1c_1)}{u_1^2c_1^2} \tag{7.3.15}$$

Again using (7.3.6) and (7.3.8):

$$\mathop{T}_{4}^{c_1}\Big(\mathcal{J}(2,u_1c_1) - \mathcal{J}(1,u_1c_1)\Big) = \frac{\exp(-2c_1)}{2c_1^3}\left(1 + \frac{3}{c_1} + \frac{3}{2c_1^2}\right)$$

$$\mathop{T}_{3}^{c_1}\Big(\mathcal{J}(0,u_1c_1) - \mathcal{J}(2,u_1c_1)\Big) = -\frac{\exp(-2c_1)}{2c_1^3}\left(2 + \frac{3}{c_1}\right)$$

and $$\mathop{T}_{2}^{c_1}\Big(\mathcal{J}(1,u_1c_1) - \mathcal{J}(0,u_1c_1)\Big) = \frac{\exp(-2c_1)}{2c_1^3} \tag{7.3.16}$$

Substituting these in (7.3.14) and simplifying, we get:

$$P_r(c_1 \geqslant x) = 3\int_{c_1=x}^{\infty} \exp(-3c_1)dc_1 = \exp(-3x) \tag{7.3.17}$$

(ii) <u>For m = 1</u>

Substituting m = 1 in (7.3.13) and again following the steps of the type (7.3.15) and (7.3.16), we obtain:

$$P_r\big(c_1 \geqslant x\big) = \int_{c_1=x}^{\infty} (c_1^3 + 5c_1^2 + 5c_1)\exp(-3c_1)dc_1 \tag{7.3.18}$$
$$= \exp(-3x)\big[x^3 + 6x^2 + 9x + 3\big]\big/3$$

(iii) <u>For m = 2</u>

Substituting m = 2 in (7.3.13) and again following the steps of the type (7.3.15) and (7.3.16), we obtain:

$$P_r(c_1 \geqslant x) = \frac{1}{30} \int_{c_1=x}^{\infty} (2c_1^6 + 20c_1^5 + 80c_1^4 + 140c_1^3 + 105c_1^2)\exp(-3c_1)dc_1$$

$$= \exp(-3x)\left[2x^6 + 24x^5 + 120x^4 + 300x^3 + 405x^2 + 270x + 90\right] / 90 \quad \text{etc.} \qquad (7.3.19)$$

Case III: For $\ell = 4$

Substituting $\ell = 4$ in (7.3.5) we obtain:

$$P_r(c_1 \geqslant x) = K(4,m) \int_{c_1=x}^{\infty} \int_{u_1=1}^{\infty} \int_{u_2=1}^{\infty} \int_{u_3=1}^{\infty} \left[c_1^{4m+9} \exp(-c_1)\right] \times$$

$$\left[u_1^{3m+5}(u_1-1)\exp(-c_1u_1)\right]\left[u_2^{2m+2}(u_2-1)(u_2u_1-1)\exp(-c_1u_1u_2)\right]\left[u_3^{m}(u_3-1)\right.$$

$$\left.(u_3u_2-1)(u_3u_2u_1-1)\exp(-c_1u_1u_2u_3)\right] \, dc_1du_1du_2du_3 \qquad (7.3.20)$$

Making use of (6.2.1) and (7.3.2) for $\ell = 4$, and after re-arrangement of terms, we obtain from (7.3.20):

$$P_r(c_1 \geqslant x) = \frac{2^{4m+5}}{\Gamma(2m+2)\,\Gamma(2m+4)} \int_{c_1=x}^{\infty} c_1^{4m+9}\exp(-c_1) \left\{ \underset{3m+8}{\overset{c_1}{\Gamma}} \left[\underset{2m\ 4}{\overset{c_1u_1}{\Gamma}} \right.\right.$$

$$\left(\mathcal{J}(m+2,\, c_1u_1u_2) - \mathcal{J}(m+1,\, c_1u_1u_2)\right) + \underset{2m+5}{\overset{c_1u_1}{\Gamma}} \left(\mathcal{J}(m+1,\, c_1u_1u_2) - \right.$$

$$\left.\mathcal{J}(m+3, c_1u_1u_2)\right) + \underset{2m+6}{\overset{c_1u_1}{\Gamma}} \left.\left(\mathcal{J}(m+3, c_1u_1u_2) - \mathcal{J}(m+2, c_1u_1u_2)\right)\right] + \underset{3m+7}{\overset{c_1}{\Gamma}} \left[\underset{2m+3}{\overset{c_1u_1}{\Gamma}} \right.$$

$$\left(\mathcal{J}(m, c_1u_1u_2) - \mathcal{J}(m+2,\, c_1u_1u_2)\right) + \underset{2m+4}{\overset{c_1u_1}{\Gamma}} \left(\mathcal{J}(m+3, c_1u_1u_2) - \mathcal{J}(m, c_1u_1u_2)\right)$$

$$\left.\left. + \underset{2m+6}{\overset{c_1u_1}{\Gamma}} \left(\mathcal{J}(m\ 2, c_1u_1u_2) - \mathcal{J}(m\ 3, c_1u_1u_2)\right)\right]\right.$$

$$+\int_{3m+6}^{c_1}\Big\{\int_{2m+2}^{c_1u_1}\Big(\mathcal{J}(m+1,c_1u_1u_2) - \mathcal{J}(m,c_1u_1u_2)\Big) + \int_{2m+4}^{c_1u_1}\Big(\mathcal{J}(m,c_1u_1u_2) - \mathcal{J}(m+3,c_1u_1u_2)\Big) + \int_{2m+5}^{c_1u_1}\Big(\mathcal{J}(m+3,c_1u_1u_2) - \mathcal{J}(m+1,c_1u_1u_2)\Big)\Big\} + \int_{3m+5}^{c_1}\Big\{\int_{2m+2}^{c_1u_1}\Big(\mathcal{J}(m,c_1u_1u_2) - \mathcal{J}(m+1,c_1u_1u_2)\Big) + \int_{2m+3}^{c_1u_1}\Big(\mathcal{J}(m+2,c_1u_1u_2) - \mathcal{J}(m,c_1u_1u_2)\Big) + \int_{2m+4}^{c_1u_1}\Big(\mathcal{J}(m+1,c_1u_1u_2) - \mathcal{J}(m+2,c_1u_1u_2)\Big)\Big\}\Big]\,dc_1 \qquad (7.3.21)$$

Now we explain below how to make us of (7.3.21) to obtain the probabilities for a particular value of m.

(i) <u>For m = 0</u>

First we substitute m = 0 in (7.3.21), and then by using (7.3.6), we obtain the following:

$$\mathcal{J}(1,c_1u_1u_2) - \mathcal{J}(0,c_1u_1u_2) = \frac{\exp(-c_1u_1u_2)}{c_1^2u_1^2u_2^2}$$

$$\mathcal{J}(2,c_1u_1u_2) - \mathcal{J}(1,c_1u_1u_2) = \frac{2\exp(-c_1u_1u_2)}{c_1^2u_1^2u_2^2}\left(1 + \frac{1}{c_1u_1u_2}\right)$$

$$\mathcal{J}(2,c_1u_1u_2) - \mathcal{J}(1,c_1u_1u_2) = \frac{\exp(-c_1u_1u_2)}{c_1^2u_1^2u_2^2}\left(1 + \frac{2}{c_1u_1u_2}\right)$$

$$\mathcal{J}(3,c_1u_1u_2) - \mathcal{J}(0,c_1u_1u_2) = \frac{3\exp(-c_1u_1u_2)}{c_1^2u_1^2u_2^2}\left(1 + \frac{2}{c_1u_1u_2} + \frac{2}{c_1^2u_1^2u_2^2}\right)$$

$$\mathcal{J}(3,c_1u_1u_2) - \mathcal{J}(1,c_1u_1u_2) = \frac{2\exp(-c_1u_1u_2)}{c_1^2u_1^2u_2^2}\left(1 + \frac{3}{c_1u_1u_2} + \frac{3}{c_1^2u_1^2u_2^2}\right)$$

and

$$\mathcal{J}(3,c_1u_1u_2) - \mathcal{J}(2,c_1u_1u_2) = \frac{\exp(-c_1u_1u_2)}{c_1^2u_1^2u_2^2}\left(1 + \frac{4}{c_1u_1u_2} + \frac{6}{c_1^2u_1^2u_2^2}\right. \tag{7.3.22}$$

Again, using (7.3.6) and (7.3.8):

(A) $$\mathop{\mathrm{T}}\limits_{2}^{c_1u_1}\Big(\mathcal{J}(1,c_1u_1u_2) - \mathcal{J}(0,c_1u_1u_2)\Big) = \frac{\exp(-2c_1u_1)}{2c_1^3u_1^3}$$

(B) $$\mathop{\mathrm{T}}\limits_{3}^{c_1u_1}\Big(\mathcal{J}(2,c_1u_1u_2) - \mathcal{J}(0,c_1u_1u_2)\Big) = \frac{\exp(-2c_1u_1)}{c_1^3u_1^3}\left(1 + \frac{3}{2c_1u_1}\right)$$

(C) $$\mathop{\mathrm{T}}\limits_{4}^{c_1u_1}\Big(\mathcal{J}(2,c_1u_1u_2) - \mathcal{J}(1,c_1u_1u_2)\Big) = \frac{\exp(-2c_1u_1)}{2c_1^3u_1^3}\left(1 + \frac{3}{c_1u_1} + \frac{3}{2c_1^2u_1^2}\right)$$

(D) $$\mathop{\mathrm{T}}\limits_{4}^{c_1u_1}\Big(\mathcal{J}(3,c_1u_1u_2) - \mathcal{J}(0,c_1u_1u_2)\Big) = \frac{3\exp(-2c_1u_1)}{2c_1^3u_1^3}$$

$$\left(1 + \frac{3}{c_1u_1} + \frac{7}{2c_1^2u_1^2}\right)$$

(E) $$\mathop{\mathrm{T}}\limits_{5}^{c_1u_1}\Big(\mathcal{J}(3,c_1u_1u_2) - \mathcal{J}(1,c_1u_1u_2)\Big) = \frac{\exp(-2c_1u_1)}{c_1^3u_1^3}$$

$$\left(1 + \frac{9}{2c_1u_1} + \frac{15}{2c_1^2u_1^2} + \frac{15}{4c_1^3u_1^3}\right)$$

(F) $$\mathop{\mathrm{T}}\limits_{6}^{c_1u_1}\Big(\mathcal{J}(3,c_1u_1u_2) - \mathcal{J}(2,c_1u_1u_2)\Big) = \frac{\exp(-2c_1u_1)}{c_1^3u_1^3}$$

$$\left(1 + \frac{6}{c_1u_1} + \frac{15}{c_1^2u_1^2} + \frac{15}{c_1^3u_1^3} + \frac{15}{2c_1^4u_1^4}\right) \tag{7.3.23}$$

Finally, using (7.3.6), (7.3.8) and (7.3.23), we obtain:

$$\mathop{\mathrm{T}}_{5}^{c_1}\Big[-(A) + (B) - (C)\Big] = -\frac{\exp(-3c_1)}{4c_1^6}$$

$$\mathop{\mathrm{T}}_{6}^{c_1}\Big[(A) - (D) + (E)\Big] = \frac{3\exp(-3c_1)}{4c_1^6}\left(1 + \frac{2}{c_1}\right)$$

$$\mathop{\mathrm{T}}_{7}^{c_1}\Big[-(B) + (D) - (F)\Big] = -\frac{3\exp(-3c_1)}{4c_1^6}\left(1 + \frac{4}{c_1} + \frac{3}{c_1^2}\right)$$

$$\mathop{\mathrm{T}}_{8}^{c_1}\Big[(C) - (E) + (F)\Big] = \frac{\exp(-3c_1)}{4c_1^6}\left(1 + \frac{6}{c_1} + \frac{9}{c_1^2} + \frac{3}{c_1^3}\right) \qquad (7.3.24)$$

Hence from (7.3.24), (7.3.21) for m = 0, we get:

$$P_r(c_1 \leq x) = 4\int_{c_1=x}^{\infty} \exp(-4c_1)dc_1 = \exp(-4x) \qquad (7.3.25)$$

(ii) For m = 1,

Following the similar steps like (7.3.22), (7.3.23) and (7.3.24) for m = 1 in (7.3.21), we get:

$$P_r(c_1 \leq x) = \frac{4}{15}\int_{c_1=x}^{\infty} (30c_1 + 45c_1^2 + 18c_1^3 + 2c_1^4)\exp(-4c_1)dc_1$$

$$= \exp(-4x)\Big[2x^4 + 20x^3 + 60x^2 + 60x + 15\Big]/15 \qquad (7.3.26)$$

etc.

Generalization in the case of m = 0

We can make a generalization for $P_r(c_1 \geq x)$ in the case of m=0. Observing for m=0, the relations (7.3.11), (7.3.17) and (7.3.25), we can conclude for any ℓ that

$$P_r(c_1 \geq x) = \ell \int_{c_1=x}^{\infty} \exp(-c_1\ell)dc_1 = \exp(-x\ell) \qquad (7.3.27)$$

7.4: Limiting Distribution of the Largest Eigenroots

From (7.3.1), the distribution of the largest eigenvalues c_ℓ is:

$$P_r(c_\ell \geq x) = K(\ell,m) \int_{c_\ell=0}^{x} \int_{c_{\ell-1}=0}^{c_\ell} \cdots \int_{c_1=0}^{c_2} \prod_{i=1}^{\ell} c_i^m \exp(-c_i) \prod_{i>j=2}^{\ell} (c_i - c_j) \prod_{i=1}^{\ell} dc_i \qquad (7.4.1)$$

Set

$$\begin{aligned}
c_1 &= c_\ell u_1 u_2 \cdots u_{\ell-3} u_{\ell-2} u_{\ell-1} \\
c_2 &= c_\ell u_1 u_2 \cdots u_{\ell-3} u_{\ell-2} \\
&\text{-------------------------} \\
c_{\ell-2} &= c_\ell u_1 u_2 \\
c_{\ell-1} &= c_\ell u_1
\end{aligned} \qquad (7.4.2)$$

then the distribution (7.4.1) reduces to:

$$P_r(c_\ell \leq x) = K(\ell,m) \int_{c_\ell=0}^{x} \int_{u_1=0}^{1} \cdots \int_{u_{\ell-2}=0}^{1} \int_{u_{\ell-1}=0}^{1}$$

$$\left[c_\ell^{\,m+\frac{(\ell-1)(\ell-2)}{2}} \exp(-c_\ell) \right] \left[u_1^{\,m(\ell-1)+\frac{(\ell-2)(\ell-1)}{2}} (1-u_2)\exp(-c_\ell u_1) \right]$$

$$\left[u_2^{\,m(\ell-2)+\frac{(\ell-3)(\ell)}{2}} (1-u_2)(1-u_1u_2)\exp(-c_\ell u_1 u_2) \right] \cdots \left[u_{\ell-1}^{m} \right.$$

$$\left. (1-u_{\ell-1})(1-u_{\ell-1}u_{\ell-2}) \cdots (1-u_{\ell-1}u_{\ell-2}\cdots u_1)\exp(-c_\ell u_1 u_2 \cdots u_{\ell-1}) \right] dc_\ell$$

$$\prod_{i=1}^{\ell-1} du_i \qquad (7.4.3)$$

Here below we give the method for evaluating (7.4.3) for particular values of ℓ = 2, 3, 4 which can further be extended for any ℓ .

Symbols and Notations

(i) $$I(n,a) = \int_0^1 x^n \exp(-ax)dx = -\frac{\exp(-a)}{a}\left[1 + \frac{n}{a} + \frac{n(n-1)}{a^2} + \cdots + \frac{n!}{a^n}\right] + \frac{n!}{a^{n+1}} \qquad (7.4.4)$$

(ii) $$I(n,p,q,r,\ldots;a) = \int_0^1 x^n \exp(-ax)dx \left[\int_0^1 y^p \exp(-axy)dy \int_0^1 z^q \exp(-axyz)dz(\ldots) \right] \qquad (7.4.5)$$

(iii) $$\mathop{\perp}_{n}^{a} \left[I(p_1,p_2;ax) \pm I(q_1,q_2;ax) \pm \cdots \right] =$$

$$= \int_0^1 x^n \exp(-ax)\left[\int_0^1 y_1^{p_1} \exp(-axy_1) \left(\int_0^1 y_2^{p_2} \exp(-axy_1y_2) dy_2 \right) dy_1 \pm \right.$$

$$\left. \int_0^1 z_1^{q_1} \exp(-axz_1)\left(\int_0^1 z_2^{q_2} \exp(-axz_1z_2) dz_2 \right) dz_1 \pm \ldots \right] dx \qquad (7.4.6)$$

<u>Case I</u>: Substituting $\ell = 2$ in (7.4.3):

$$P_r(c_2 \leq x) = K(2,m) \int_{c_2=0}^{x} \int_{u_1=0}^{1} c_2^{2m+2} \exp(-c_2) \quad u_1^m(1-u_1)\exp(-c_2u_1)du_1dc_2 \qquad (7.4.7)$$

Making use of (6.2.1) and (7.3.2), (7.4.7) reduces to:

$$P_r(c_2 \leq x) = \frac{2^{2m+1}}{\Gamma(2m+2)} \int_{c_2=0}^{x} c_2^{2m+2} \exp(-c_2) \left[I(m,c_2) - I(m+1,c_2) \right] dc_2 \; .$$

Using (7.4.4) and simplifying:

$$P_r(c_2 \leq x) = \frac{2^{2m+1}}{\Gamma(2m+2)} \int_{c_2=0}^{x} c_2^{2m+2} \exp(-c_2) \left[\frac{\exp(-c_2)}{c_2} \left(1 + 2\frac{m}{c_2} + \right.\right.$$

$$\left.\left. 3\frac{m(m-1)}{c_2^2} + \ldots (m+1) \frac{m!}{c_2^m} \right) + \left(\frac{m!}{c_2^{m+1}} - \frac{(m+1)!}{c_2^{m+2}} \right) \right] dc_2 \qquad (7.4.8)$$

which can be easily evaluated for successive values of m.

<u>Case II</u>: <u>For $\ell = 3$</u>

Substituting $\ell = 3$ in (7.4.3), we have:

$$P_r(c_3 \leq x) = K(3,m) \int_{c_3=0}^{1} \int_{u_1=0}^{1} \int_{u_2=0}^{1} c_3^{3m+5} \exp(-c_3)$$

$$\left[u_1^{2m+2}(1-u_1)\exp(-c_3u_1)\right]\left[(u_2^{m}\,(1-u_2)(1-u_2u_1)\exp(-c_3u_1u_2)\right] dc_3du_1du_2$$

(7.4.9)

Using (7.3.2) for $\ell = 3$ and (6.2.1), we obtain from (7.4.9) after re-arrangement of terms:

$$P\,(c_3 \leq\ x)\quad \frac{2^{2m\;3}}{\Gamma(m+1)\,\Gamma(2m+3)} \int_{c_3=0}^{x} c_3^{3m\;5} \exp(-c_3)dc_3 \times$$

$$\left[\mathop{\perp}\limits_{2m+4}^{c_3}\Big(I(m+1,u_1c_3) - I(m+2,u_1c_3)\Big) + \mathop{\perp}\limits_{2m+3}^{c_3}\Big(I(m+2,u_1c_3) - I(m,u_1c_3)\Big)\right.$$

$$\left. + \mathop{\perp}\limits_{2m+2}^{c_3}\Big(I(m,u_1c_3) - I(m+1,u_1c_3)\Big)\right] \tag{7.4.10}$$

which can be easily evaluated for different values of m by repeated use of (7.4.4). In fact, we have to use the same steps as in (7.3.15) and (7.3.16), using repeatedly (7.4.4) instead of (7.3.6). Following this procedure we have computed probabilities for $m = 0$, 1 and 2. The results are as follows:

(i) For $m = 0$,

$$P_r(c_3 \leq x) = -3 \int_{c_3=0}^{x} \exp(-3c_3)dc_3 + 4 \int_{c_3=0}^{x} c_3^2 \exp(-2c_3)dc_3$$

$$+ \int_{c_3=0}^{x} (2c_3^2 - 6c_3 + 3)\exp(-c_3)dc_3 \tag{7.4.11}$$

(ii) <u>For m = 1,</u>

$$P_r(c_3 \le x) = -\int_{c_3=0}^{x} (c_3^3 + 5c_3^2 + 5c_3)\exp(-3c_3)dc_3$$

$$+ \frac{4}{3}\int_{c_3=0}^{x} c_3^4 \exp(-2c_3)dc_3$$

$$+ \int_{c_3=0}^{x} (c_3^3 - 5c_3^2 + 5c_3)\exp(-c_3)dc_3 \qquad (7.4.12)$$

(iii) <u>For m = 2</u>

$$P_r(c_3 \le x) = -\frac{1}{30}\int_{c_3=0}^{x} (2c_3^6 + 20c_3^5 + 80c_3^4 + 140c_3^3 + 105c_3^2)\exp(-3c_3)dc_3$$

$$+ \frac{8}{45}\int_{c_3=0}^{x} c_3^6 \exp(-2c_3)dc_3$$

$$+ \frac{1}{6}\int_{c_3=0}^{x} (2c_3^4 - 14c_3^3 + 21c_3^2)\exp(-c_3)dc_3 \qquad (7.4.13)$$

<u>Case III</u>: <u>For $\ell = 4$</u>

Substituting $\ell = 4$ in (7.4.3):

$$P_r(c_4 \le x) = K(4,m) \int_{c_4=0}^{x}\int_{u_1=0}^{1}\int_{u_2=0}^{1}\int_{u_3=0}^{1} \left[c_4^{4m+9}\exp(-c_4)\right]$$

$$\left[u_1^{3m}(1-u_1)\exp(-u_1c_4)\right]\left[u_2^{2m}(1-u_2)(1-u_1u_2)\exp(-c_4u_1u_2)\right]$$

$$\left[u_3^{m}(1-u_3)(1-u_3u_2)(1-u_3u_2u_1)\exp(-u_3u_2u_1c_4)\right] dc_4du_1du_2du_3 \qquad (7.4.14)$$

Using (6.2.1) and (7.3.2) for $\ell = 4$, we obtain after re-arrangement of terms:

$$
\begin{aligned}
P_r(c_4 \leq x) = {} & \frac{2^{4m+5}}{\Gamma(2m+2)\Gamma(2m+4)} \int_{c_4=0}^{x} c_4^{4m+9} \exp(-c_4) \Bigg[\mathop{\perp}\limits_{3m+8}^{c_4} \Bigg\{ \mathop{\perp}\limits_{2m+6}^{c_4u_1} \\
& \Big(\mathbf{I}(m+2, c_4u_1u_2) - \mathbf{I}(m+3, c_4u_1u_2)\Big) + \mathop{\perp}\limits_{2m+5}^{c_4u_1} \Big(\mathbf{I}(m+3, c_4u_1u_2) - \\
& \mathbf{I}(m+1, c_4u_1u_2)\Big) + \mathop{\perp}\limits_{2m+4}^{c_4u_1} \Big(\mathbf{I}(m+2, c_4u_1u_2) - \mathbf{I}(m+1, c_4u_1u_2)\Big)\Bigg\} + \\
& \mathop{\perp}\limits_{3m+7}^{c_4} \Bigg\{ \mathop{\perp}\limits_{2m+6}^{c_4u_1} \Big(\mathbf{I}(m+3, c_4u_1u_2) - \mathbf{I}(m+2, c_4u_1u_2)\Big) + \mathop{\perp}\limits_{2m+4}^{c_4u_1} \\
& \Big(\mathbf{I}(m, c_4u_1u_2) - \mathbf{I}(m+3, c_4u_1u_2)\Big) + \mathop{\perp}\limits_{2m+3}^{c_4u_1} \Big(\mathbf{I}(m+2, c_4u_1u_2) - \\
& \mathbf{I}(m, c_4u_1u_2)\Big)\Bigg\} + \mathop{\perp}\limits_{3m+6}^{c_4} \Bigg\{ \mathop{\perp}\limits_{2m+5}^{c_4u_1} \Big(\mathbf{I}(m+1, c_4u_1u_2) - \mathbf{I}(m+3, c_4u_1u_2)\Big) + \\
& \mathop{\perp}\limits_{2m+4}^{c_4u_1} \Big(\mathbf{I}(m+3, c_4u_1u_2) - \mathbf{I}(m, c_4u_1u_2)\Big) + \mathop{\perp}\limits_{2m+2}^{c_4u_1} \Big(\mathbf{I}(m, c_4u_1u_2) - \\
& \mathbf{I}(m+1, c_4u_1u_2)\Big)\Bigg\} + \mathop{\perp}\limits_{3m+5}^{c_4} \Bigg\{ \mathop{\perp}\limits_{2m+4}^{c_4u_1} \Big(\mathbf{I}(m+2, c_4u_1u_2) - \mathbf{I}(m+1, c_4u_1u_2)\Big) + \\
& \mathop{\perp}\limits_{2m+3}^{c_4u_1} \Big(\mathbf{I}(m, c_4u_1u_2) - \mathbf{I}(m+2, c_4u_1u_2)\Big) + \mathop{\perp}\limits_{2m+2}^{c_4u_1} \Big(\mathbf{I}(m+1, c_4u_1u_2) - \\
& \mathbf{I}(m, c_4u_1u_2)\Big)\Bigg\}\Bigg] \, dc_4
\end{aligned}
\tag{7.4.15}
$$

The procedure for evaluating (7.4.15) is the same as we used in Case III of Section (7.3) dealing with smallest eigen-roots. We have to follow the same steps as (7.3.22), (7.3.23) and (7.3.24) and have to make repeated use of (7.4.4).

7.5: Limiting Distribution of U or Y for ℓ = 2, 3 and 4.

The moment generating function of (1.4.10) is:

$$m(t) = \int_0^\infty \cdots \int_0^\infty K(\ell,m)(c_1 c_2 \ldots c_\ell)^m \exp\left[-\sum_{i=1}^{\ell} c_i + t\sum_{i=1}^{\ell} c_i\right] \prod_{i=2}^{\ell}\prod_{j=1}^{i-1}(c_i - c_j) \prod_{i=1}^{\ell} dc_i \qquad (7.5.1)$$

from which the h-th moment μ'_h about the origin is:

$$\mu'_h = \frac{K(\ell,m)}{K(\ell,m+h)} = \prod_{i=1}^{\ell} \frac{\Gamma\left(\frac{2m+i+1}{2}+h\right)}{\Gamma\left(\frac{2m+i+1}{2}\right)}$$

$$\text{or } \mu'_h = \frac{\Gamma\left(\frac{2m+2}{2}+h\right)}{\Gamma\left(\frac{2m+2}{2}\right)} \frac{\Gamma\left(\frac{2m+3}{2}+h\right)}{\Gamma\left(\frac{2m+3}{2}\right)} \cdots \frac{\Gamma\left(\frac{2m+\ell+1}{2}+h\right)}{\Gamma\left(\frac{2m+\ell+1}{2}\right)} \qquad (7.5.2)$$

This h-th moment shows that the moments of the limiting distribution of the product of the roots $(c_1 c_2 \ldots c_\ell)$ can also be determined from the following:

$$\frac{1}{\Gamma\left(\frac{2m+2}{2}\right)} v_1^{\frac{2m+2}{2}-1} \exp(-v_1)dv_1 \, \frac{1}{\Gamma\left(\frac{2m+3}{2}\right)} v_2^{\frac{2m+3}{2}-1} \exp(-v_2)dv_2 \ldots$$

$$\frac{1}{\Gamma\left(\frac{2m+\ell+1}{2}\right)} v_\ell^{\frac{2m+\ell+1}{2}-1} \exp(-v_\ell)dv_\ell$$

or from
$$\frac{\exp\left(-\sum_{i=1}^{\ell} v_i\right)}{\Gamma(m+1)\Gamma(m+\frac{3}{2})\ldots\Gamma(m+\frac{\ell+1}{2})} v_1^m v_2^{m+\frac{1}{2}} v_3^{m+1} \ldots v_\ell^{m+\frac{\ell-1}{2}} dv_1\ldots dv_\ell \tag{7.5.3}$$

where $\quad 0 \leq v_i < \infty, \quad i=1, 2, \ldots, \ell$

Case I: For $\ell = 2$

Substituting $\ell = 2$ in (7.5.3), we get the joint distribution of v_1 and v_2 as:

$$\frac{1}{\Gamma(m+1)\,\Gamma(m+\frac{3}{2})} v_1^m v_2^{m+\frac{1}{2}} \exp(-v_1-v_2)dv_1dv_2 \tag{7.5.4}$$

where $\quad 0 \leq v_i < \infty \quad i=1, 2,$

which is the same as (6.3.1) for $k_1^2 = 0$, $u_0 = 2v_2$, $u_1 = 2v_1$. Hence from (6.3.7), the distribution of $V_1 = 2\sqrt{v_1 v_2}$ or of $2\sqrt{c_1c_2}$ for $k_1^2 = 0$ is:

$$\frac{V_1^{2m+1} \exp(-V_1)}{\Gamma(2m+2)} dV_1 \;, \tag{7.5.5}$$

where $\quad 0 \leq V_1 < \infty$,

which is a gamma variate of parameter (2m+1). Further, for any test of hypothesis, we need to make the two types of substitutions for m. We proceed as follows:

(i): When $p = 2(\leq n_1)$, set $m = \frac{n_1-3}{2}$ in (7.5.5), and get the distribution of $V_1 (= 2\sqrt{c_1 c_2})$ as follows:

$$\frac{1}{\Gamma(n_1-1)} V_1^{n_1-2} \exp(-V_1) dV_1 , \tag{7.5.6}$$

where $0 \leq V_1 < \infty$,

which is a gamma variate with parameter (n_1-1).

(ii) When $n_1 = 2(\leq p)$, set $m = \frac{p-3}{2}$ in (7.5.5) and obtain:

$$\frac{1}{\Gamma(p-1)} V_1^{p-2} \exp(-V_1) dV_1 \tag{7.5.7}$$

$$0 \leq V_1 < \infty$$

which is a gamma variate with parameter (p-1).

Case II: For $\ell = 3$

Substituting $\ell = 3$ in (7.5.3), the joint distribution of v_1, v_2, and v_3 is:

$$\frac{\exp(-v_1-v_2-v_3)}{\Gamma(m+1)\Gamma(m+\frac{3}{2})\Gamma(m+2)} v_1^m v_2^{m+3/2} v_3^{m+1} dv_1 dv_2 dv_3 , \tag{7.5.8}$$

where $0 \leq v_i < \infty \qquad i=1,2,3,$

which is the same as (6.3.9) for $k_1^2 = 0$, $u_2 = 2v_1$, $u_1 = 2v_2$, $u_0 = 2v_3$. Hence from (6.3.13), the distribution of $V_1 (= 8v_1 v_2 v_3$ or $= 8c_1 c_2 c_3)$, where $k_1^2 = 0$, is:

$$\frac{V_1^m}{2^{m+1}\Gamma(m+1)\Gamma(2m+3)} L_0\left(\sqrt{\frac{V_1}{2}}\right) dV_1 \tag{7.5.9}$$

where $L_0(a)$, for $a = \sqrt{\frac{V_1}{2}}$, is defined in (6.2.33). Again, for any test of hypothesis, we need to make the following two types of substitutions in (7.5.9)

(i) if $p = 3(\leq n_1)$, set $m = \frac{n_1-4}{2}$ in (7.5.9)

(ii) if $n_1 = 3(\leq p)$, set $m = \frac{p-4}{2}$ in (7.5.9).

<u>Case III:</u> <u>For $\ell = 4$</u>

Substituting for $\ell = 4$ in (7.5.3), the joint distribution of v_1, v_2, v_3 and v_4 is:

$$\frac{\exp(-v_1-v_2-v_3-v_4)}{\Gamma(m+1)\Gamma(m+\frac{3}{2})\Gamma(m+2)\Gamma(m+\frac{5}{2})} v_1^m v_2^{m+\frac{1}{2}} v_3^{m+1} v_4^{m+3/2} \, dv_1 dv_2 dv_3 dv_4, \qquad (7.5.10)$$

where $\qquad 0 \leq v_i < \infty \qquad i=1,2,3,4,$

which is again of the type (6.4.15) for $k_1^2 = 0$, $u_3 = 2v_1$, $u_2 = 2v_2$, $u_1 = 2v_3$ and $u_0 = 2v_4$.

Hence from (6.3.19) for $k_1^2 = 0$, and making use of (6.2.18), the distribution of $V_1 (= 16v_1v_2v_3v_4$ or $16c_1c_2c_3c_4)$ is:

$$\frac{a^{2m}d(a^2)}{\Gamma(2m+2)\Gamma(2m+4)}\left[\left(\frac{(1+2\gamma)-\log a}{2}\right)\left(\frac{a^2}{2!1!}+\frac{a^3}{3!1!}+\frac{a^4}{4!2!}+\ldots\right)\right.$$

$$\left.+\frac{1}{2}\left(1-a+\frac{a^2}{2^2}+\frac{15a^3}{3^2\,2^2}+\ldots\right)\right] \qquad (7.5.11)$$

where $a = \sqrt{V_1}$ and $0 \leq V_1 < \infty$

Again, for any test of hypothesis, we need to make two types of changes for m in (7.2.11) as:

(i) For $p = 4(\le n_1)$, set $m = \frac{n_1 - 5}{2}$

(ii) For $n_1 = 4(\le p)$, set $m = \frac{p-5}{2}$

<u>Note</u>: For $\ell = 5, 6$; a similar method was applied but we were confronted with the following difficult integrals:

<u>For $\ell = 5$</u>

The integral in this case, is:

$$\frac{2^{4m+8}\, V_1^m dV_1}{\Gamma(m+1)\,\Gamma(2m+3)\,\Gamma(2m+5)} \int_{V_2=0}^{\infty}\int_{V_4=0}^{\infty} V_4 \exp\left(-\frac{V_1}{V_2^2} - \frac{2V_2}{V_4} - 2V_2\right) dV_2\, dV_4 \tag{7.5.12}$$

for $V_1 = c_1c_2c_3c_4c_5$ and $0 \le V < \infty$

<u>For $\ell = 6$</u>

The integral is:

$$\frac{2^{6m+11}\, V_1^m\, dV_1}{\Gamma(2m+2)\,\Gamma(2m+4)\,\Gamma(2m+6)} \int_{V_5=0}^{\infty}\int_{V_3=0}^{\infty} \exp\left(-2\frac{V_1}{V_3} - 2\frac{V_3}{V_5} - 2V_5\right) dV_3\, dV_5 \tag{7.5.13}$$

for $V_1 = c_1c_2 \ldots c_6$ and $0 \le V_1 < \infty$

CHAPTER EIGHT

APPROXIMATE DISTRIBUTIONS OF THE NON-ORTHOGONAL COMPLEX ESTIMATES

8.1: In the case of unequal sub-class numbers in Anova of Model II, we run into the difficulty of defining the distributions of the mean squares or the sum product (S.P.) matrices respectively in both the univariate and multivariate cases. In such situations, as pointed out earlier, for the univariate case the mean squares are distributed as sums like $\sum_r (\lambda_r \chi_r^2)$, where the λ_r are functions of the variance components and the number of observations, while each χ_r^2 is distributed as central chi-square with 1 D.F. Similarly, for the multivariate case the S.P. matrices, as proved below, are distributed as sums $\sum_r (W_r)$ of independent Wishart matrices with different parameter matrices Σ_r and one degree of freedom for each.

Thus, we try to approximate $\sum_r (\lambda_r \chi_r^2)$ and $\sum_r (W_r)$ respectively by $\lambda \chi^2$ and a Wishart matrix W with revised D.F. To do this we determine first what are the λ_r and Σ_r and then use Satterthwaite's technique to approximate the sums $\sum_r (\lambda_r \chi_r^2)$ and $\sum_r (W_r)$ respectively by $\lambda \chi^2$ and W to find the respective corresponding D.F.

For finding λ_r and Σ_r, we begin below with the multivariate case, from which the univariate case is deduced, and the corresponding D.F. determined for both.

8.2 Suppose N is greater than p or n. Observations $(X_{1\alpha}, X_{2\alpha}, \ldots X_{p\alpha}; z_{1\alpha}, z_{2\alpha}, \ldots z_{n\alpha})$ for $\alpha = 1, 2, \ldots, N$, are made on $(p + n)$ variables. The overall set of assumptions is Ω :

$$X_{i\alpha} = \beta_{i1} z_{1\alpha} + \ldots + \beta_{in} z_{n\alpha} + e_{i\alpha} \qquad (8.2.1)$$

for $i = 1, 2, \ldots, p$; $\alpha = 1, 2, \ldots, N$; and furthermore:

(i) the $z_{r\alpha}$ $(r = 1, 2, \ldots, n;\ \alpha = 1, 2, \ldots, N)$ are non-random, and the matrix $Z(n \times N) = (z_{r\alpha})$ is of rank n. (In the case we are most interested in, the case of Anova, some of the z's are zeros and the rest are ones)

(ii) the vectors $\underline{e}^t_\alpha \equiv (e_{1\alpha}, \ldots, e_{p\alpha})$ are independent and normally distributed with mean vectors 0 and error covariance matrix $\Sigma_e(p \times p)$, i.e. $E(e_{i\alpha}) = 0$ and $E(e_{i\alpha} \times e_{j\alpha}) = \sigma_{e,ij}$ for $i, j = 1, 2, \ldots, p$ and $\alpha = 1, 2, \ldots, N$, so that

$$\Sigma_e(p \times p) = (\sigma_{e,ij}) \qquad (8.2.2)$$

(iii) $-\infty < \beta_{ir} < +\infty$

Let us introduce some further notations as follows:

$$\beta\,(p \times n) \equiv (\beta_{ir}) \equiv [\underset{n_1}{\beta_1}, \underset{n'}{\beta_2}] \quad \text{for } n_1 + n' = n \qquad (8.2.3)$$

where n_1 and n' will be specified below.

$\underline{X}^t_\alpha (1 \times p) \equiv (X_{1\alpha}, X_{2\alpha}, \ldots, X_{p\alpha})$,

so that $\underline{X}(p \times N) \equiv (X_{i\alpha}) \equiv [\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_N] \qquad (8.2.4)$

$\underline{z}^t_\alpha (1 \times n) \equiv (z_{1\alpha}, z_{2\alpha}, \ldots, z_{n\alpha})$

so that $Z(n \times N) \equiv (z_{r\alpha}) \equiv [\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_N]$ (8.2.5)

$$A(n \times n) \equiv \begin{array}{cc} \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] & \begin{array}{l} n_1 \\ n' \end{array} \\ \begin{array}{cc} n_1 & n' \end{array} & \end{array} = ZZ^t \tag{8.2.6}$$

and

$$C(p \times n) \equiv \begin{array}{c} \left[\begin{array}{c|c} C_1 & C_2 \end{array}\right] \\ \begin{array}{cc} n_1 & n' \end{array} \end{array} = XZ^t \tag{8.2.7}$$

Under the overall set of assumptions Ω, the matrix B_{Ω} $(p \times n)$ which is the least square estimator of $\boldsymbol{\beta}$ $(p \times n)$ is

$$B_{\Omega} \quad (p \times n) = CA^{-1} \tag{8.2.8}$$

and hence the S.P. matrix Q_{Ω} (Anderson, pp. 181) is:

$$Q_{\Omega} = XX^t - B_{\Omega} A B_{\Omega}^t \tag{8.2.9}$$

If a hypothesis H specifies $\boldsymbol{\beta}_1(p \times n_1)$, then the distribution of Q_H, the S.P. matrix, due to deviations from hypothesis, depends on the nature of the $\boldsymbol{\beta}_{ir}$ $(i = 1, 2, \ldots, p;\ r = 1, 2, \ldots, n_1)$. The overall set of assumptions can be completed in two useful ways as follows:

(i) The columns of $\boldsymbol{\beta}_1$ are independent, normally distributed vectors with common covariance matrix $\Sigma_{\boldsymbol{\beta}}$ $(p \times p)$ of rank p, and are all quite independent of the columns of $\boldsymbol{\beta}_2(p \times n')$ and of $\underline{e}_{\alpha}$. $\boldsymbol{\beta}_2$ may be either random or constant.

(ii) $\boldsymbol{\beta}_1$ is constant.

Case (ii) is the usual regression problem considered in standard texts. In what follows we consider only case (i). We let w denote the subset of Ω for which the following hypothesis holds, H : $\Sigma_{\boldsymbol{\beta}} = 0$, which implies that $E(\boldsymbol{\beta}_1) = \boldsymbol{\beta}_{1,0}$, a matrix of constants.

Then $Q_w = (X - \beta_{1,0}Z_1)(X - \beta_{1,0}Z_1)^t - B_{2w}A_{22}B_{2w}^t$

$$= XX^t - B_{\Omega}\ A\ B_{\Omega}^t + (B_{1\Omega} - \beta_{1,0})A_{11.2}(B_{1\Omega} - \beta_{1,0})^t \tag{8.2.10}$$

where $B_{2w} = (C_2 - B_{1,0}A_{12})A_{22}^{-1}$ (8.2.11)

and $A_{11.2}(n_1 \times n_1) = A_{11} - A_{12}A_{22}^{-1}A_{21}$ (8.2.12)

Hence from (8.2.9) and (8.2.10), we obtain:

$$Q_H(p \times p) = Q_w - Q_{\Omega}$$

$$= (B_{1\Omega} - \beta_{1,0})A_{11.2}(B_{1\Omega} - \beta_{1,0})^t \tag{8.2.13}$$

Now there exists an orthogonal matrix U such that

$$UA_{11.2}U^t = \Gamma^2(n_1 \times n_1) \tag{8.2.14}$$

where (i) Γ^2 is a diagonal matrix with elements

γ_r^2 $(r = 1, 2, \ldots, n_1)$

and (ii) $A_{11.2}$, Γ^2 and U are all non-random and each of order $(n_1 \times n_1)$.

Therefore $Q_H(p \times p) = (B_{1\Omega} - \beta_{1,0})U^t\ \Gamma^2\ U(B_{1\Omega} - \beta_{1,0})^t$

$$= \left[(B_{1\Omega} - \beta_{1,0})U^t\right]\Gamma^2\left[(B_{1\Omega} - \beta_{1,0})U^t\right]^t \tag{8.2.15}$$

Setting $(B_{1\Omega} - \beta_{1,0})U^t = (D_{1\Omega} - \Delta_{1,0})$, (8.2.16) where $E(D_{1\Omega}) = \Delta_{1,0}$,

we obtain from (8.2.15) the following:

$$Q_H(p \times p) = (D_{1\Omega} - \Delta_{1,0}) \Gamma^2 (D_{1\Omega} - \Delta_{1,0})^t \tag{8.2.17}$$

Its (i,j)th element can be written as:

$$Q_{H,ij} = \sum_{r=1}^{n_1} \left[\gamma_r^2 \ (d_{ir} - \delta_{ir,0})(d_{jr} - \delta_{jr,0}) \right] \tag{8.2.18}$$

so that its expected value is:

$$E(Q_{H,ij}) = \sum_{r=1}^{n_1} (\sigma_{e,ij} + \gamma_r^2 \ \sigma_{\beta, ij}), \tag{8.2.19}$$

which enables us to write that:

$$Q_{H,ij} = \sum_{r=1}^{n_1} (\sigma_{e,ij} + \gamma_r^2 \ \sigma_{\beta, ij}) u_{ir} u_{jr}, \tag{8.2.20}$$

where the u_{ir} ($i = 1, 2, \ldots, p$ and $r = 1, 2, \ldots, n_1$) are normal variates with mean zero and variance 1. Therefore, for the univariate case, as also proved by Nash (1956), we obtain:

$$Q_H \text{ (for univariate case)} = \sum_{r=1}^{n_1} \left\{ \sigma_e^2 + \gamma_r^2 \ \sigma_\beta^2 \right\} \ u_r^2 \tag{8.2.21}$$

where u_r^2 are independent central chi-square variates each with one D.F. Thus the λ_r ($r = 1, 2, \ldots, n_1$), indicated earlier, are obtained. They are:

$$\lambda_r = \sigma_e^2 + \gamma_r^2 \sigma_\beta^2 \qquad r = 1, 2, \ldots, n_1 \tag{8.2.22}$$

Similarly from (8.2.20) we can deduce:

$$Q_H(p \times p) \equiv \sum_{r=1}^{n_1} \underline{u}_r(\Sigma_e + \gamma_r^2 \Sigma_\beta)\underline{u}_r^t = \sum_{r=1}^{n_1} W_r \tag{8.2.23}$$

where $W_r = \underline{u}_r(\Sigma_e + \gamma_r^2 \Sigma_\beta)\underline{u}_r^t$, $\quad r = 1, 2, \ldots, n_1$ $\qquad$ (8.2.24)

so that $\Sigma_r = E(W_r) = \Sigma_e + \gamma_r^2 \Sigma_\beta = \left(\sigma_{e,ij} + \gamma_r^2\ \sigma_{\beta,ij}\right)$

(8.2.25)

8.3: Approximate Distributions

(a) Univariate Case

Weconsider again the relation (8.2.21) and write:

$$Q_H(1 \times 1) = \sum_{r=1}^{n_1} (\sigma_e^2 + \gamma_r^2 \sigma_\beta^2)u_r^2 \tag{8.3.1}$$

Since the h-th cumulant of $(\ {}^2_e + {}^2_r\ {}^2)u_r^2$ is:

$$k_h = 2^{h-1}(h-1)!(\sigma_e^2 + \gamma_h^2 \sigma_\beta^2)^h, \tag{8.3.2}$$

it follows that the h-th cumulant of $Q_H(1 \times 1)$ is:

$$2^{h-1}(h-1)! \sum_{r=1}^{n_1} (\sigma_e^2 + \gamma_r^2 \sigma_\beta^2)^h. \tag{8.3.3}$$

Hence the first two moments of $Q_H(1 \times 1)$ about the origin are:

$$\mu_1' = \sum_{r=1}^{n_1} (\sigma_e^2 + \gamma_r^2 \sigma_\beta^2)$$

$$\mu_2' = 2 \sum_{r=1}^{n_1} (\sigma_e^2 + \gamma_r^2 \sigma_\beta^2)^2 \tag{8.3.4}$$

Following Satterthwaite, we approximate Q_H, defined in (8.3.1), by $\lambda\chi^2$ where χ^2 is a central chi-square with f D.F., so that the first two moments of $Q_H(1 \times 1)$ are respectively equal to those of $\lambda\chi^2$. Therefore, making use of (8.3.4), we obtain:

$$f\lambda = \sum_{r=1}^{n_1} (\sigma_e^2 + \gamma_r^2 \sigma_\beta^2) \tag{8.3.5}$$

$$2f\lambda^2 = 2 \sum_{r=1}^{n_1} (\sigma_e^2 + \gamma_r^2 \sigma_\beta^2)^2, \tag{8.3.6}$$

since $E(\chi^2) = f$ and $\text{var}(\chi^2) = 2f$.

Finally, from (8.3.5) and (8.3.6) we obtain:

$$f = \left[\sum_{r=1}^{n_1} (\sigma_e^2 + \gamma_r^2 \sigma_\beta^2)\right]^2 \Bigg/ \sum_{r=1}^{n_1} (\sigma_e^2 + \gamma_r^2 \sigma_\beta^2)^2 \tag{8.3.7}$$

Since σ_e^2 and σ_β^2 are not known, we shall substitute for them their respective estimates $\hat{\sigma}_e^2$ and $\hat{\sigma}_\beta^2$ and write:

$$f \doteq \left[\sum_{r=1}^{n_1} (\hat{\sigma}_e^2 + \gamma_r^2 \hat{\sigma}_\beta^2)\right]^2 \Bigg/ \sum_{r=1}^{n_1} (\hat{\sigma}_e^2 + \gamma_r^2 \hat{\sigma}_\beta^2)^2 \tag{8.3.8}$$

which thus determines the approximate distribution of a mean square in the unbalanced case to be $\lambda\chi^2$, where χ^2 is the new chi-square with estimated degrees of freedom defined in (8.3.8).

Note: In ordinary analysis of variance with a balanced design the eigenvalues are all equal, say γ^2. A balanced design occurs for example when the number of observations is the same in each sub-class. Then we obtain from (8.3.1) the simpler result:

$$Q_H(1 \times 1) = \sum_{r=1}^{n_1} (\sigma_e^2 + \gamma^2 \sigma_\beta^2) u_r^2 = (\sigma_e^2 + \gamma^2 \sigma_\beta^2) \sum_{r=1}^{n_1} u_r^2$$

or $$Q_H(1 \times 1) = (\sigma_e^2 + \gamma^2 \sigma_\beta^2) \chi^2_{n_1} \tag{8.3.9}$$

Thus, $Q_H(1 \times 1)$ in the balanced case is distributed as $\lambda \chi^2_{n_1}$ where $\lambda = \sigma_e^2 + \gamma^2 \sigma_\beta^2$.

(b) Multivariate Case

We approximate the sum $\sum_r (W_r)$ or $W_r\left[\Sigma_r, 1\right)$ $r = 1, 2, \ldots, n_1$ by a Wishart matrix $W\left[\Sigma, f\right]$ or order $(p \times p)$ where f is to be determined such that:

(i) The expected matrix of the approximating matrix is equal to that of the sum of the W_r;

(ii) The elements of the approximating Wishart matrix have an ellipsoid of concentration (Cramér 1946) whose volume is equal to the corresponding volume for the sum of the given Wishart matrices.

Condition (i) gives:

$$E(W) = \sum_r E(W_r)$$

i.e. $$f\Sigma = \Sigma_1 + \Sigma_2 + \ldots + \Sigma_{n_1} \tag{8.3.10}$$

Further, if $p^{(r)}\left[\binom{p+1}{2} \times \binom{p+1}{2}\right]$ be the matrix of the covariances of elements of W, and P that of W also of order $\left[\binom{p+1}{2} \times \binom{p+1}{2}\right]$, condition (ii) gives:

$$\text{Det.}(P) = \text{Det.}\left(\sum_r P^{(r)}\right) \tag{8.3.11}$$

Thus to find f', one should find from (8.3.10) by comparison the elements of Σ in terms of those of Σ_r and should substitute them in the left hand side of (8.3.11). For instance, in our case from (8.3.13), we have:

$$f' \; \sigma_{ij} = \sum_{r=1}^{n_1} (\sigma_{ij(r)}) \tag{8.3.12}$$

where $\sigma_{ij(r)} = \sigma_{e,ij} + \gamma_r^2 \; \sigma_{\beta,ij}; \;\; i,j = 1, 2, \ldots, p$

(8.3.13)

and, following T.W. Anderson (p. 161), from (8.3.11) we have:

$$\det.\left[f'(\sigma_{ik} \; \sigma_{j\ell} + \sigma_{i\ell} \; \sigma_{jk})\right] = \det.\left[\sum_{r=1}^{n_1} (\sigma_{ik(r)} \; \sigma_{j\ell(r)} + \sigma_{i\ell(r)} \; \sigma_{jk(r)})\right] \tag{8.3.14}$$

for i, j, k, ℓ = 1, 2, ..., p.

Finally, making use of (8.3.12) and (8.3.13), the relation (8.3.14) becomes:

$$\det.\left[\left(\sum_{r=1}^{n_1} \sigma_{ik(r)}\right)\left(\sum_{r=1}^{n_1} \sigma_{j\ell(r)}\right) + \left(\sum_{r=1}^{n_1} \sigma_{i\ell(r)}\right)\left(\sum_{r=1}^{n_1} \sigma_{jk(r)}\right)\right]$$

$$= f^{\frac{p(p+1)}{2}} \det.\left[\sum_{r=1}^{n_1} (\sigma_{ik(r)} \; \sigma_{j\ell(r)} + \sigma_{i\ell(r)} \; \sigma_{jk(r)})\right] \tag{8.3.15}$$

Again, since the σ's are not known, we substitute for them their respective estimates and then obtain appropriate degrees of freedom f, where it should be noted that the $\sigma_{ij(r)}$ are defined as in (8.3.13).

In this way the distribution of the S.P. matrix is approximated by the Wishart matrix with the estimated D.F. f and the estimated parameter matrix $\hat{\Sigma}$.

<u>Note</u>: In ordinary multivariate analysis of variance with a balanced design the eigenvalues are all equal, say γ^2. A balanced design occurs for example when the number of observations is the same in each sub-class. Then we obtain from (8.2.23) the simpler result that $Q_H(p \times p)$ is a Wishart matrix with the density

$$W\left[\Sigma_e + \gamma^2 \Sigma_{\beta}, \quad n_1\right]$$

APPENDIX A

EVALUATION OF THE EIGEN-VALUES AND EIGEN-VECTORS OF THE MATRIX BW^{-1}

We need to solve, for Γ^2(pxp) and L(pxp), the system of equations:

$$L(BW^{-1}) = \Gamma^2 L \tag{A-1}$$

where W and B denote symmetrical matrices, positive definite and at least positive semi-definite respectively, and Γ^2 denotes a diagonal matrix. Since the matrix (BW^{-1}) is non-symmetrical, the calculation of eigen-values and eigenvectors for this matrix is much more difficult than that for a symmetrical matrix. To solve the matrix equation (A-1), the step-by-step procedure due to Nash and Jolicoeur is as follows:

(i) Solve, for Λ^2 (diagonal) and U (orthogonal), the matrix equation:

$$UW = \Lambda^2 U \tag{A-2}$$

(ii) Obtain the matrix Λ^{-1}.

(iii) Compute the matrix product:

$$G = \Lambda^{-1} UBU^t \Lambda^{-1} \tag{A-3}$$

This matrix is theoretically symmetrical. If the computed matrix is not quite symmetrical due to round-off errors, symmetrize it by replacing g_{ij} and g_{ji} each by their arithmetic or geometric mean.

(iv) Solve for (diagonal) Γ^2(pxp) and orthogonal V(pxp), the matrix equation:

$$VG = \Gamma^2 V \tag{A-4}$$

(v) Obtain the matrix L of co-efficients of the discriminant function as follows:

$$L = V \Lambda^{-1} U \tag{A-5}$$

Thus both the matrices Γ^2(pxp) and L(pxp), the solutions of (A-1), are known respectively from (A-4) and (A-5).

APPENDIX B

FINDING BOUNDS FOR THE COEFFICIENTS OF CERTAIN CUBIC EQUATIONS

We take the cubic defined in (7.2.19) and re-write it:

$$x^3 - ux^2 + vx - w = 0. \tag{B-1}$$

We want to determine bounds for u, v, w so that the roots of equation (B-1) are real and positive.

Referring to any standard book on theory of equations, such as Burnside and Panton, the discriminant Δ of (B-1) is found to be

$$\Delta \equiv \left(w - \frac{uv}{3} + \frac{2}{27}u\right)^2 + \frac{4}{27}\left(v - \frac{u^2}{3}\right)^3 \tag{B-2}$$

or
$$27\Delta \equiv 4wu^3 - u^2v^2 - 18uvw + (27w^2 + 4v^3). \tag{B-3}$$

Furthermore the equation (B-1) has real and positive roots if Δ is negative, i.e., if

$$27\Delta \equiv 4wu^3 - u^2v^2 - 18uvw + (27w^2 + 4v^3) \leq 0 \tag{B-4}$$

and
$$v - \frac{u^2}{3} \leq 0 \qquad \text{i.e., if } \sqrt{3v} \leq u \tag{B-5}$$

Now we deduce the bounds for u, v, w from (B-4) and (B-5) in the following two forms:

Form I:

We re-write (B-4) as:

$$\Delta \equiv w^2 - \frac{2}{3}u\left(v - \frac{2}{9}u^2\right) + \frac{v^2}{27}(4v - u^2) \leq 0 \tag{B-6}$$

Solving $\Delta \equiv 0$ for w and making use of equation (B-5), the range for w is obtained as:

$$\text{Max}\begin{bmatrix}\frac{1}{3}u(v - \frac{2}{9}u^2) - \frac{2}{27}(u^2 - 3v)^{3/2} \\ 0\end{bmatrix} \le w \le \frac{1}{3}u(v - \frac{2}{9}u^2) + \frac{2}{27}(u^2 - 3v)^{3/2} \qquad \text{(B-7)}$$

Further, $\frac{1}{3}u(v - \frac{2}{9}u^2) - \frac{2}{27}(u^2 - 3v)^{3/2}$ is positive

if $\quad \left[9u(v - \frac{2}{9}u^2)\right]^2 - 4(u^2 - 3v)^3 \geq 0$

i.e. if $\qquad v \geq \frac{u^2}{4} \qquad$ (B-8)

Thus, from (B-5), (B-7) and (B-8), the following two parts of the bounds of u, v, w are:

(i)

$$0 \le u < \infty$$
$$0 \le v \le \frac{1}{4}u^2$$
$$0 \le w \le \beta_2$$

and (ii)

$$0 \le u < \infty$$
$$\frac{1}{4}u^2 \le v \le \frac{1}{3}u^2$$
$$\beta_1 \le w \le \beta_2 \qquad \text{(B-9)}$$

where β_1 and β_2 are defined as follows:

$$\beta_1 = \frac{u}{3}(v - \frac{2}{9}u^2) - \frac{2}{27}(u^2 - 3v)^{3/2}$$
$$\beta_2 = \frac{u}{3}(v - \frac{2}{9}u^2) + \frac{2}{27}(u^2 - 3v)^{3/2} \qquad \text{(B-10)}$$

<u>Note</u>: When the roots of (B-1) assume values between zero and one, we have to change the range $0 \le u < \infty$ in (B-9) into $0 \le u \le 1$, and the bounds for v and w remain unaffected.

Form II:

Applying Descartes' rule of signs to $\Delta = 0$ in (B-3), we conclude that the cubic in u, for known positive real values of v, w, has at most two positive roots and one negative root. Setting:

$$y = 4wu - \frac{v^2}{3} \tag{B-11}$$

in (B-4), we obtain:

$$f(y) \equiv y^3 - 3y(24vw^2 + \frac{v^4}{9}) + 2(216w^4 + 20v^3w^2 - \frac{v^6}{27}) \leq 0 \tag{B-12}$$

Now two cases arise:

Case I, $\quad 216w^4 + 20v^3w^2 - v^6/27$ is positive (B-13)

Case II, $\quad 216w^4 + 20v^3w^2 - v^6/27$ is negative (B-13')

Further, applying Descartes' rule of signs to (B-12) and using the case of (B-13), we conclude that the cubic in y, for known positive and real values of v, w, has again at most two positive roots and one negative root. Thus, for known real and positive values of v, w, the negative root of (B-12) shall correspond to the negative root of (B-4) and the two positive roots of (B-12) to the two positive roots of (B-4).

Similarly (B-13') enables us to conclude that the largest positive root of (B-4) corresponds to the only positive root of (B-12) and the smallest positive root of (B-4) corresponds to the largest negative root of (B-12).

To find the bounds on u, v, w, in both these cases we proceed as follows:-

Case I

To find the bounds for y, v, w for f(y) in (B-12), we first draw its curve and from its shape we conclude what the bounds for y are:

Consider:

$$f(y) = y^3 - 3y(24vw^2 + \frac{v^4}{9}) + 2(216w^4 + 20v^3w^2 - \frac{v^6}{27}) \qquad \text{(B-14)}$$

(i) When $y = 0$, $f(y) > 0$.

(ii) By Descartes rule of signs f(y) has at most two positive roots and one negative root.

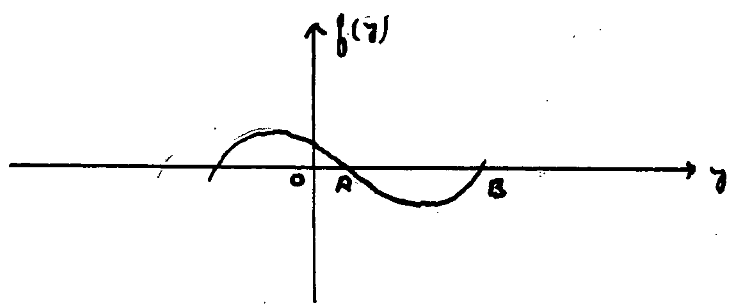
(iii) Finding $f'(y)$ and $f''(y)$, we conclude the following:

(a) $y = \sqrt{24vw^2 + \frac{v^4}{9}}$ gives a minimum of f(y).

(b) $y = -\sqrt{24vw^2 + \frac{v^4}{9}}$ gives a maximum of f(y).

and (c) $y = 0$ is a point of inflection.

Thus the shape of the curve (B-14) is as shown below.



Hence, in order that f(y) be negative for real and positive values of y, y must take the values from A to B, i.e.:

$$OA \leq y \leq OB$$

and from (B-11), $\frac{1}{4w}(OA + \frac{v^2}{3}) \leq u \leq \frac{1}{4w}(OB + \frac{v^2}{3})$ (B-15)

Thus to have positive zeros of (B-14), i.e. OA and OB, we follow Todhunter or Burnside and Panton and write the zeros of (B-14) as:

$$2(24vw^2 + \frac{v^4}{9}) \cos \frac{\phi}{3},$$

$$-2(24vw^2 + \frac{v^4}{9}) \cos \frac{\pi + \phi}{3},$$

$$-2(24vw^2 + \frac{v^4}{9}) \cos \frac{\pi - \phi}{3}, \tag{B-16}$$

where ϕ is defined by the relation:

$$\tan \phi = - \frac{8(\frac{v^3}{3} - 9w^2)^{3/2} w}{2(216w^4 + 20v^3w^2 - \frac{v^6}{27})} \tag{B-17}$$

and where a real value of ϕ is possible,

if $\frac{v^3}{3} > 9w^2$

i.e. if $3w^{2/3} \leq v$ (B-18)

Since tan ϕ in (B-17) is negative, ϕ will be an obtuse angle which will make the first two roots of (B-16) positives and the last negative. Therefore, OA and OB are obtained as:

$$OA = 2(24vw^2 + \frac{v^4}{9}) \cos \frac{\phi}{3},$$

and $$OB = -2(24vw^2 + \frac{v^4}{9}) \cos \frac{\pi + \phi}{3}, \tag{B-19}$$

which enable us to write the reduced form of (B-15) as follows:

$$\frac{1}{4w}\left[2\left(24vw^2+\frac{v^4}{9}\right)^{\frac{1}{2}}\cos\frac{\phi}{3}+\frac{v^2}{3}\right] \leq u \leq \left[\frac{1}{4w}\left(-2\left(24vw^2+\frac{v^4}{9}\right)^{\frac{1}{2}}\right.\right.$$
$$\left.\left.\cdot\cos\left(\frac{\pi+\phi}{3}\right)+\frac{v^2}{3}\right)\right]$$

or, making use of (B-5), we get:

$$\beta_3 \leq u \leq \beta_4 \tag{B-20}$$

where $\beta_3 = \mathrm{Max}\left[\sqrt{3}v, \frac{1}{4w}\left(2\left(24vw^2+\frac{v^4}{9}\right)^{\frac{1}{2}}\cos\frac{\phi}{3}+\frac{v^2}{3}\right)\right]$

and $$\beta_4 = \frac{1}{4}\left[-2\left(24vw^2+\frac{v^4}{9}\right)\cos\left(\frac{\pi+\phi}{3}\right)+\frac{v^2}{3}\right]\frac{1}{w} \tag{B-21}$$

Further, $216w^4 + 20v^3w^2 - \frac{v^6}{27}$ is positive, if

$$27w^2(10-6\sqrt{3}) \leq v^3 \leq 27w^2(10+6\sqrt{3})$$

i.e. if $\quad 3w^{2/3}(1-\sqrt{3}) \leq v \leq 3w^{2/3}(1+\sqrt{3})$

But v cannot be negative. Therefore

$$0 \leq v \leq 3w^{2/3}(1+\sqrt{3}) \tag{B-22}$$

Finally, from (B-5), (B-20), (B-21) and (B-22), we obtain:

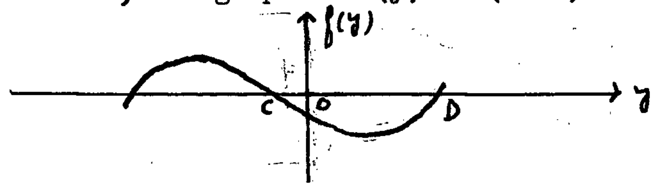
$$0 \leq w \leq \infty$$
$$3w^{2/3} \leq v \leq 3w^{2/3}(1+\sqrt{3})$$
$$\beta_3 \leq u \leq \beta_4 \tag{B-23}$$

where β_3 and β_4 are defined in (B-21).

Note: When the roots of (B-1) assume values between zero and one, we have to change the range $0 \leq w \leq \infty$ in (B-23) to $0 \leq w \leq 1$, and the ranges for v and u remain unchanged.

Case II

Proceeding as before, the graph of f(y) in (B-12) is as follows:



Hence, in order that f(y) be negative for real and positive values of y, y must take the values from C to D and thus from (B-11),

$$\frac{1}{4w}(CO+\frac{v^2}{3}) \leq u \leq \frac{1}{4w}(OD+\frac{v^2}{3}) \tag{B-24}$$

To find CO and OD, we proceed as in Case I, and conclude that the same relations as (B-16), (B-17) and (B-18) hold. Further, since tan ϕ in (B-17) is positive, ϕ will be an acute angle. Noting this fact and equation (B-13'), we can easily obtain the bounds on u, v, w, for this case also. Finally the bounds obtained for both of the cases are written down as follows:

$$\begin{array}{lll} 0 \leq w < \infty & \text{and} & 0 \leq w < \infty \\ 3w^{2/3} \leq v \leq 3w^{2/3}(1+\sqrt{3}) & & 3w^{2/3}(1+\sqrt{3}) \leq v < \infty \\ \beta_3 \leq u \leq \beta_4 & & \beta_3' \leq u \leq \beta_4' \end{array} \tag{B-25}$$

where

$$\beta_3' = \text{Max}\left\{\frac{1}{4w}\left[-2(24vw^2+\frac{v^4}{9})^{1/2}\cos\frac{\pi+\phi_1}{3}+\frac{v^2}{3}\right],\ \sqrt{3v}\right\}$$

and

$$\beta_4' = \frac{1}{4w}\left[2(24vw^2+\frac{v^4}{9})^{1/2}\cos\frac{\phi_1}{3}+\frac{v^2}{3}\right] \tag{B-26}$$

where ϕ_1 is the supplement of ϕ used in Case I.

Appendix C

Upper 100 α_h Percentage Points of χ^2
$\alpha = .05$ and $\alpha_h = 1-(1-\alpha)^{k-1}$.

D.F \ k	2	3	4	5	6	7	8
1	3.84146	2.74605	2.15011	1.75364	1.46490	1.24301	1.06704
2	5.99147	4.65590	3.89530	3.36943	2.97266	2.65677	2.39686
3	7.81473	6.30923	5.43433	4.81934	4.34847	3.96832	3.65123
4	9.48773	7.84315	6.87534	6.18829	5.65765	5.22577	4.86276
5	11.0705	9.30510	8.25704	7.50787	6.92580	6.44941	6.04700
6	12.5916	10.7179	9.59808	8.79351	8.16561	7.64970	7.21215
7	14.0671	12.0944	10.9092	10.0504	9.38445	8.83250	8.36304
8	15.5073	13.4428	12.1970	11.2951	10.5869	10.0016	9.50270
9	16.9190	14.7685	13.4660	12.5205	11.7762	11.1599	10.6334
10	18.3070	16.2966	14.7194	13.7328	12.9546	12.3091	11.7567
11	19.6751	17.3664	15.9598	14.9342	14.1238	13.4504	12.8735
12	21.0261	18.6438	17.1889	16.1261	15.2850	14.5852	13.8950
13	22.3621	19.9092	18.4080	17.3096	16.4392	15.7142	15.0915
14	23.6848	21.1643	19.6187	18.4861	17.5875	16.8382	16.1940
15	24.9958	22.4100	20.8215	19.6560	18.7302	17.9575	17.2927
16	26.2962	23.6473	22.0173	20.8110	19.8680	19.0727	18.3879
17	27.5871	24.8710	23.2070	21.9789	21.0015	20.1844	19.4802
18	28.8693	26.0999	24.3909	23.1330	22.1310	21.2927	20.5697
19	30.1435	27.3164	25.5699	24.2827	23.2568	22.3978	21.6567
20	31.4104	28.5272	26.7436	25.4283	24.3792	23.5002	22.7413
21	32.6705	29.7325	27.9131	26.5703	25.4985	24.5999	23.8238
22	33.9244	30.9330	29.0785	27.7088	26.6149	25.6973	24.9043

APPENDIX C

Upper I00 α_k Percentage Points of χ^2

$\alpha = .05$ and $\alpha_k = I-(I-\alpha)^{k-1}$ – Contd.

D.F \ k	2	3	4	5	6	7	8
23	35.I725	32.I287	30.240I	28.844I	27.7285	26.7923	25.9829
24	36.4I5I	33.3202	3I.3982	29.9765	28.8397	27.8854	27.0598
25	37.6525	34.5076	32.5524	3I.I060	29.9486	28.9764	28.I35I
26	38.8852	35.69II	33.7042	32.2328	3I.055I	30.0654	29.2087
27	40.II33	36.87I2	34.8528	33.3573	32.I596	3I.I530	30.28II
28	4I.3372	38.0479	35.9986	34.4793	33.2622	32.2387	3I.3520
29	42.5569	39.22I4	37.I4I8	35.5992	34.3630	33.3230	32.42I7
30	43.7729	40.39I2	38.2824	36.7I70	35.4620	34.4057	33.490I
40	55.7585	5I.958I	49.5762	47.8005	46.37I7	45.I660	44.II89
50	67.5048	63.3355	60.7II0	58.7506	57.I704	55.8345	54.67I7
60	79.08I9	74.579I	7I.7368	69.6095	67.8920	66.4380	65.I708
70	90.53I2	85.7220	82.6795	80.3988	78.5550	76.9924	75.6293
80	I0I.879	96.7848	93.5562	9I.I329	89.I7I7	87.5082	86.0559
90	II3.I45	I07.783	I04.379	I0I.822	99.7505	97.9922	96.4562
I00	I24.342	II8.726	II5.I57	II2.473	II0.298	I08.450	I06.834
X	I.6449	I.2960	I.0686	0.8945	0.75I4	0.6283	0.5I96

Appendix C

Upper IOO α_k Percentage Points of χ^2

α=.05 and $\alpha_k = \mathrm{I}-(\mathrm{I}-\alpha)^{k-1}$ - Contd.

D.F \ k	9	10	11	12	13	14	15
I	0.92376	0.80490	0.70494	0.6I989	0.54904	0.48403	0.42935
2	2.I7786	I.98989	I.82634	I.68238	I.55476	I.94048	I.3376I
3	3.38059	3.I4536	2.938I5	2.75354	2.58790	2.43980	2.30I02
4	4.55064	4.27748	4.03522	3.8I794	3.62I7I	3.44275	3.27864
5	5.6993I	5.39360	5.I2I26	4.87595	4.65347	4.44973	4.262I3
6	6.8328I	6.498I6	6.I9909	5.92886	5.68304	5.45726	5.24877
7	7.95493	7.59396	7.27058	6.97769	6.7I064	6.46483	6.23734
8	9.06802	8.68277	8.33697	8.023I7	7.73656	7.47226	7.22724
9	I0.I739	9.765I0	9.39925	9.06594	8.76I05	8.47948	8.2I8I0
I0	II.2739	I0.8446	I0.458I	I0.I064	9.78433	9.4865I	9.20970
II	I2.3686	II.9I9I	II.5I40	II.I449	I0.8065	I0.4933	I0.20I9
I2	I3.4586	I2.990I	I2.5672	I2.I8I6	II.8276	II.4998	II.I945
I3	I4.5454	I4.0582	I3.6I83	I3.2I68	I2.848I	I2.5062	I2.I876
I4	I5.6285	I5.I237	I4.6674	I4.2507	I3.8677	I3.5I23	I3.I8I0
I5	I6.7085	I6.I867	I5.7I47	I5.2834	I4.8866	I4.5I83	I4.I747
I6	I7.7858	I7.2475	I6.7604	I6.3I49	I5.9049	I5.524I	I5.I686
I7	I8.8607	I8.3065	I7.8046	I7.3454	I6.9226	I6.5297	I6.I627
I8	I9.9333	I9.3637	I8.8476	I8.4752	I7.9399	I7.5353	I7.I572
I9	2I.0038	20.4I92	I9.8893	I9.4040	I8.9567	I8.5406	I8.I5I7
20	22.0725	2I.4737	20.9300	20.432I	I9.9730	I9.5459	I9.I464
2I	23.I395	22.526I	2I.9696	2I.4595	20.9890	20.55I0	20.I4I3
22	24.2048	23.5775	23.0082	22.4862	22.0045	2I.5560	2I.I363

Appendix C

Upper I00 α_k Percentage Points of χ^2

α = .05 and $\alpha_k = I-(I-\alpha)^{k-1}$ - Contd.

D.F \ k	9	10	11	12	13	14	
23	25.2686	24.6278	24.0460	23.5I23	23.0I98	22.56I0	22.56I0
24	26.33I0	25.6770	25.0830	24.5379	24.0346	23.5658	
25	27.392I	26.725I	26.II9I	25.5629	25.0493	24.5705	
26	28.45I9	27.7722	27.I545	26.5874	26.0635	25.575I	
27	29.5I06	28.8I84	28.I893	27.6II4	27.0775	26.5797	
28	30.568I	29.8638	29.2234	28.6350	28.09I3	27.584I	
29	3I.6247	30.9084	30.2570	29.6583	29.I049	28.5886	
30	32.6803	3I.9522	3I.2898	30.68I0	30.II8I	29.5929	
40	43.I906	42.3544	4I.5922	40.8904	40.2404	39.6329	
50	53.6393	52.7080	5I.8580	5I.0745	50.3479	49.6680	
60	64.0444	63.0273	62.0982	6I.2408	60.445I	59.7000	
70	74.4I66	73.3207	72.3I87	7I.3935	70.5344	69.7293	
80	84.7629	83.5936	82.524I	8I.5358	80.6I75	79.7567	
90	95.0879	93.8498	92.7I66	9I.6692	90.6956	89.7824	
I00	I05.395	I04.092	I02.898	I0I.795	I00.769	99.8067	
X	0.42I8	0.3325	0.250I	0.I733	0.I0I4	0.0335	

Appendix C

Upper $100\alpha_h$ Percentage Points of χ^2

$\alpha = .05$ and $\alpha_h = 1-(1-\alpha)^{h-1}$ - Contd.

D.F \ h	15	16	17	18	19	20
I	0.42935	0.38I7I	0.34003	0.30333	0.27092	0.242283
2	I.3376I	I.24457	I.I60I7	I.083I0	I.0I246	0.947680
3	2.30I02	2.I7582	2.06087	I.95466	I.856I6	I.76474
4	3.27864	3.I2745	2.98779	2.85793	2.73677	2.62363
5.	4.262I3	4.0886I	3.92769	3.77748	3.63679	3.5049I
6	5.24877	5.05538	4.87553	4.90720	4.549I0	4.4005I
7	6.23734	6.02588	5.82882	5.64400	5.47005	5.30624
8	7.22724	6.999I0	6.786I3	6.58607	6.39750	6.2I964
9	8.2I8I0	7.97438	7.74657	7.53229	7.33005	7.I3904
I0	9.20970	8.95I30	8.70953	8.48I86	8.26674	8.06337
II	I0.20I9	9.92960	9.67458	9.43422	9.2069I	8.99I82
I2	II.I945	I0.9090	I0.64I4	I0.3889	I0.I500	9.92378
I3	I2.I896	II.8895	II.6098	II.3458	II.0957	I0.8588
I4	I3.I8I0	I2.8707	I2.5794	I2.3043	I2.0436	II.7964
I5	I4.I747	I3.8527	I3.5503	I3.2645	I2.9935	I2.7364
I6	I5.I686	I4.8353	I4.5222	I4.2260	I3.945I	I3.6785
I7	I6.I627	I5.8I85	I5.4950	I5.I888	I4.8983	I4.6225
I8	I7.I572	I6.8024	I6.4687	I6.I529	I5.853I	I5.5683
I9	I8.I5I7	I7.7866	I7.443I	I7.II79	I6.809I	I6.5I56
20	I9.I464	I8.7II3	I8.4I83	I8.0839	I7.7663	I7.4644
2I	20.I4I3	I9.7564	I9.394I	I9.0508	I8.7246	I8.4I44
22	2I.I363	20.74I9	20.3705	20.0I85	I9.6839	I9.3657

Appendix C

Upper IOO α_h Percentage Points of χ^2

α = .05 and α_h = $I-(I-\alpha)^{h-1}$ - Contd.

D.F \ h	15	16	17	18	19	20
23	22.I3I5	2I.7278	2I.3475	20.987I	20.6443	20.3I82
24	23.I267	22.7I39	22.3250	2I.9562	2I.6055	2I.27I7
25	24.I22I	23.7005	23.3030	22.926I	22.5675	22.2262
26	25.II75	24.6872	24.28I4	23.8965	23.5303	23.I8I7
27	26.II3I	25.6742	25.2603	24.8676	24.4939	24.I380
28	27.I089	26.66I6	26.2397	25.8394	25.4583	25.0953
29	28.I047	27.649I	27.2I94	26.8II5	26.4232	26.0532
30	29.I004	28.6368	28.0I6I	27.784I	27.3886	27.0II9
40	39.0623	38.5244	38.0I6I	37.5330	37.0722	36.6327
50	49.0289	48.4257	47.8552	47.3I23	46.7943	46.2996
60	58.9990	58.3369	57.7I02	57.II35	56.5435	55.9990
70	68.97I4	68.255I	67.5767	66.9305	66.3I30	65.7227
80	78.9460	78.I994	77.453I	76.7602	76.099I	75.4662
90	88.922I	88.I082	87.3368	86.60I4	85.898I	85.2253
I00	98.9994	98.0409	97.2270	96.4507	95.708I	99.9976
X	-0.0309	-0.0922	-0.I507	-0.2067	-0.2606	-0.3I24

Appendix D

Upper IOO α_k Percentage Points of χ^2

$\alpha = .0I$ and $\alpha_k = I-(I-\alpha)^{k-1}$ - C

D.F \ k	2	3	4	5	6	7	8
I	6.63490	5.42074	4.72647	4.24329	3.87492	3.57869	3.2984I
2	9.2I034	7.834I6	7.03308	6.46767	6.03I36	5.67655	5.337I6
3	II.3449	9.84847	8.96928	8.36409	7.85922	7.46260	7.08I30
4	I3.2767	II.6797	I0.7359	I0.06I4	9.5360I	9.I0502	8.68933
5	I5.0863	I3.4008	I2.3998	II.6825	II.I22I	I0.66I2	I0.2I56
6	I6.8II9	I5.0464	I3.994I	I3.2382	I2.646I	I2.I584	II.6860
7	I8.4753	I6.6362	I5.5368	I4.7453	I4.I243	I3.6I20	I3.II5I
8	20.0902	I8.I825	I7.0394	I6.2I48	I5.567I	I5.03I9	I4.5I23
9	2I.6660	I9.6938	I8.5095	I7.6540	I6.98II	I6.4246	I5.8836
I0	23.2093	2I.I759	I9.9527	I9.0680	I8.37I4	I7.7946	I7.2335
II	24.7250	22.6336	2I.3734	20.4608	I9.74I5	I9.I457	I8.5655
I2	26.2I70	24.070I	22.7746	2I.8354	2I.0945	20.4804	I9.8820
I3	27.6883	25.488I	24.I586	23.I940	22.4325	2I.8008	2I.I849
I4	29.I4I3	26.8897	25.5277	24.5385	23.757I	23.I085	22.4758
I5	30.5779	28.2768	26.8832	25.8705	25.0699	24.4050	23.7562
I6	3I.9999	29.6509	28.2269	27.I9I2	26.3720	25.69I4	25.0269
I7	33.4087	3I.0I3I	29.5595	28.50I7	27.6646	26.9687	26.289I
I8	34.8053	32.3646	30.8823	29.8030	28.9485	28.2378	27.5439
I9	36.I908	33.706I	32.I960	3I.0957	30.2242	29.4992	28.7905

Appendix D

Upper I00 α_k Percentage Points of χ^2

α=.0I and α_k= I-(I-α)$^{k-1}$ Contd.

D.F \ k	2	3	4	5	6	7	8
20	37.5662	35.0387	33.50I4	32.3807	3I.4927	30.7536	30.03I0
2I	38.932I	36.3628	34.7990	33.6584	32.7543	32.00I6	3I.2654
22	40.2894	37.6792	36.0895	34.9295	34.0097	33.2436	32.4942
23	4I.6384	38.9884	37.3734	36.I944	35.2592	34.480I	33.7I77
24	42.9798	40.2907	38.6509	37.4535	36.5032	35.7II4	34.9363
25	44.3I4I	4I.5867	39.9227	38.7070	37.7420	36.9377	36.I502
26	45.64I7	42.8768	4I.I890	39.9555	38.976I	38.I595	37.3598
27	46.9630	44.I6II	42.4500	4I.I990	40.2055	39.3769	38.5653
28	48.2782	45.4402	43.7062	42.438I	4I.4307	40.4903	39.7670
29	49.5879	46.7I44	44.9678	43.6728	42.65I7	4I.7997	40.9650
30	50.8922	47.9838	46.205I	44.9035	43.8689	43.0056	42.I595
40	63.6907	60.4606	58.4783	57.0242	55.866I	54.898I	53.9480
50	76.I539	72.6399	70.478I	68.8895	67.6226	66.5625	65.5208
60	88.3794	84.6085	82.2843	80.574I	79.2088	78.0655	76.94II
70	I00.425	96.4I79	93.9443	92.I223	90.6666	89.4467	88.2462
80	II2.329	I08.I02	I05.48I	I03.563	I02.022	I00.73I	99.4598
90	I24.II6	II9.682	II6.939	II4.9I5	II3.296	III.978	II0.600
I00	I35.807	I3I.I77	I28.3I0	I26.I93	I24.500	I23.078	I2I.678
X	2.32630	2.05584	I.88523	I.75766	I.65455	I.56729	I.48068

Appendix D

Upper I00 α_k Percentage Points of

α = .0I and $\alpha_k = I-(I-\alpha)^{k-1}$ Contd.

D.F \ k	9	10	11	12	13	14
I	3.09254	2.9I302	2.75429	2.6I264	2.48494	2.36892
2	5.08536	4.86380	4.66620	4.48823	4.32639	4.I78I7
3	6.79703	6.54588	6.32I02	6.II768	5.93209	5.76I54
4	8.37850	8.903I8	7.85608	7.63206	7.42720	7.23849
5	9.88I7I	9.58543	9.3I905	9.07720	8.8556I	8.65I23
6	II.33I4	II.0I64	I0.7327	I0.4749	I0.2384	I0.0200
7	I2.74I7	I2.4095	I2.II0I	II.8377	II.5875	II.3563
8	I4.I2I3	I3.7732	I3.4593	I3.I733	I2.9I06	I2.6676
9	I5.4763	I5.II33	I4.7857	I4.487I	I4.2I25	I3.9584
I0	I6.8I07	I6.4336	I6.093I	I5.7827	I5.4969	I5.2324
II	I8.I280	I7.7376	I7.3849	I7.0630	I6.7669	I6.4924
I2	I9.4304	I9.0274	I8.6629	I8.3303	I8.0240	I7.740I
I3	20.7I99	20.3046	I9.9290	I9.5860	I9.2700	I8.977I
I4	2I.9979	2I.57I0	2I.I846	20.83I7	20.5065	20.2049
I5	23.2658	22.8276	22.4309	22.0684	2I.7343	2I.4243
I6	24.5245	24.0753	23.6686	23.2969	22.954I	22.636I
I7	25.7750	25.3I52	24.8989	24.5I82	24.I67I	23.84I2
I8	27.0I80	26.5480	26.I222	25.7328	25.3736	25.0402
I9	28.2542	27.7742	27.3393	26.94I5	26.5743	26.2335

Appendix D

Upper I00 α_k Percentage Points of χ^2

α = .0I and $\alpha_k = I-(I-\alpha)^{k-1}$ Contd.

D.F \ k	9	10	11	12	13	14
20	29.4840	28.9943	28.5505	28.I444	27.7697	27.42I6
2I	30.7079	30.2088	29.7563	29.3422	28.9600	28.6050
22	3I.9265	3I.4I8I	30.9572	30.5353	30.I457	29.7839
23	33.I400	32.6226	32.I534	3I.7238	3I.327I	30.9586
24	34.3489	33.8227	34.3453	32.9083	32.5046	32.I295
25	35.5532	35.0I84	34.533I	34.0886	33.6783	33.2968
26	36.7535	36.2I0I	35.7I7I	35.2655	34.8483	34.4606
27	37.9499	37.3982	36.8976	36.4390	36.0I52	35.62I3
28	39.I425	38.5827	38.0746	37.609I	37.I789	36.7790
29	40.3345	39.7639	39.2485	38.7762	38.3398	37.9339
30	4I.5I75	40.94I9	40.4I93	39.9403	39.4977	39.0860
40	53.226I	52.578I	5I.989I	5I.4493	50.9499	50.485I
50	64.7286	64.0I69	63.3696	62.7754	62.2253	6I.7I30
60	76.0853	75.3I6I	74.6I60	73.973I	73.3776	72.8228
70	87.332I	86.5099	85.76I4	85.0736	84.4365	83.8426
80	98.49I4	97.620I	96.8267	96.0972	95.42I3	94.79I0
90	I09.580	I08.663	I07.827	I07.058	I06.3458	I05.68I
I00	I20.6I0	II9.648	II8.772	II7.967	II7.220	II6.523
X	I.4I42I	I.35403	I.2989I	I.24800	I.20058	I.I56I7

Appendix D

Upper IOO α_k Percentage of χ^2

α = .0I and $\alpha_k = \mathrm{I}-(\mathrm{I}-\alpha)^{k-1}$ - Contd.

D.F. \ k	15	16	17	18	19	20
I	2.26275	2.I6508	2.07477	I.99093	I.9I282	I.83979
2	4.04I49	3.9I479	3.79678	3.68642	3.58286	3.48535
3	5.60372	5.45695	5.3I98I	5.I9II6	5.07008	4.95573
4	7.06353	6.90049	6.74787	6.60442	6.469I7	6.34I2I
5	8.46I46	8.28438	8.II839	7.962I8	7.8I472	7.67502
6	9.8I699	9.62737	9.44946	9.28I87	9.I235I	8.97336
7	II.I4I3	I0.9402	I0.75I4	I0.5735	I0.4052	I0.2455
8	I2.44I3	I2.2297	I2.0308	II.8433	II.6758	II.4973
9	I3.72I8	I3.5002	I3.2920	I3.0954	I2.9093	I2.7326
I0	I4.9859	I4.755I	I4.5380	I4.3330	I4.I388	I3.9543
II	I6.2365	I5.9969	I5.77I3	I5.5583	I5.3565	I5.I646
I2	I7.4757	I7.2273	I6.9937	I6.7730	I6.5639	I6.3650
I3	I8.7039	I8.4477	I8.2064	I7.9784	I7.7623	I7.5567
I4	I9.9235	I9.6595	I9.4I09	I9.I806	I8.953I	I8.74I0
I5	2I.I349	20.8635	20.6078	20.3660	20.I366	I9.9I84
I6	22.339I	22.0605	2I.7979	2I.5496	2I.3I40	2I.0897
I7	23.5368	23.25I2	22.9820	22.7273	22.4857	22.2557
I8	24.7286	24.4362	24.I606	23.8998	23.6522	23.4I66
I9	25.9I50	25.6I60	25.334I	25.0673	24.8I4I	24.5729

Appendix D

Upper IOO $\mathcal{L}_k$ Percentage of χ^2

$\mathcal{L}$=.OI and $\mathcal{L}_k$=I-(I-$\mathcal{L}$)$^{k-1}$ - Contd.

D.F. \ k	15	16	17	18	19	20
20	27.0964	26.79I0	26.5029	26.2303	25.97I5	25.725I
2I	28.273I	27.96I4	27.6675	27.3892	27.I250	26.8733
22	29.4456	29.I278	28.8280	28.5442	28.2748	28.0I80
23	30.6I40	30.2902	29.9848	29.6956	29.42I0	29.I593
24	3I.7788	3I.4492	3I.I383	30.8438	30.564I	30.2975
25	32.9400	32.6047	32.2884	3I.9887	3I.704I	3I.4328
26	34.0979	33.7570	33.4353	33.I306	32.84II	32.5652
27	35.2528	34.9065	34.5796	34.2699	33.9757	33.6952
28	36.4049	36.053I	35.72II	35.4065	35.I076	34.8227
29	37.554I	37.I97I	36.860I	36.5407	36.2373	35.9479
30	38.7007	38.3386	37.9966	37.6726	37.3646	37.0709
40	50.0496	49.6398	49.2525	48.8853	48.5390	48.2026
50	6I.2330	60.78II	60.3539	59.9486	59.5629	59.I948
60	72.3027	7I.8I29	7I.3496	70.9I00	70.49I5	70.09I9
70	83.2856	82.7609	82.2646	8I.7933	8I.3447	80.9I6I
80	94.I998	93.6427	93.II56	92.6I50	92.I383	9I.6829
90	I05.058	I04.470	I03.9I4	I03.386	I02.883	I02.403
I00	II5.869	II5.2530	II4.670	II4.II5	II3.587	II3.083
X	I.II433	I.07476	I.037I7	I.00I34	I.967II	0.93428

BIBLIOGRAPHY

Anderson, T.W. (1946) "The non-central Wishart distribution and certain problems of multivariate statistics." Ann. of Math. Stat. 17, 409-431.
(1957) An Introduction to Multivariate Statistical Analysis. New York, John Wiley and Sons.

Anderson, R.L. and Bancroft, T.A. (1952) Statistical Theory in Research. New York, McGraw-Hill Book Co.

Bannerjee, D.P. (1958) "On the exact distribution of a test in multivariate analysis." J. Roy. Stat. Soc., Ser. B, 20, 108-110.

Bartlett, M.S. (1934) "The vector representation of sample." Proc. Camb. Phil. Soc. 30, 327-340.
(1938) "Further aspects of the theory of multiple regression." Proc. Camb. Phil. Soc. 34, 33-40.
(1947) "Multivariate analysis." Supp. J. Roy. Stat. Soc. 9, 176-197.

Bierens deHaan, D. (1939) Nouvelles Tables d'Intégrales Définies. Ed. of 1867 - corrected. New York, G.E. Stechert and Co.

Bose, R.C. (1936) "On the exact distribution and moment coefficients of the D^2 - statistic." Sankhyā 2, 143-154.

Bose, R.C. and Roy, S.N. (1938) "The distribution of the Studentized D^2 - statistic." Sankhyā 4, 19-38.

Bross, I. (1950) "Fiducial intervals for variance components." Biometrics 6, 136-144.

Burnside, W.S. and Panton, A.W. (1904) Theory of Equations. Dublin, Dublin University Press Series, Vol. 1.

Cramér, H. (1946) Mathematical Methods of Statistics. Princeton, Princeton University Press.

Crump, S.L. (1945) "The estimation of variance components in analysis of variance." Biometrics 1, 7-11.

Cuttle, Yvonne (1956) "The distribution of extreme Mahalanobis distance from the sample mean." Master's Thesis, University of British Columbia, Vancouver.

Duncan, D.B. (1951) "A significance test for difference between ranked treatments in an analysis of variance." Virginia Journal of Sc. 2, 171-189.
(1953) "Multiple range tests and the multiple comparisons test." (Preliminary report)." Biometrics 9, 262, Abstract 220.
(1955) "Multiple range and multiple F tests." Biometrics 11, 1-42.
(1957) "Multiple range tests for correlated and heteroscedastic means." Biometrics 13, 164-176.

Eisenhart, C. (1947)"The assumptions underlying the analysis of variance." Biometrics 3, 1-21.

Federer, W.T. (1955) Experimental Design. New York, The Macmillan Co.

Fisher, R.A. (1928) "The general sampling distributions of the multiple correlation coefficient." Proc. Roy. Soc. London, Ser. A, 121, 654-673.
(1931) "Introduction" in Mathematical Tables of the British Association for the Advancement of Science. Vol. 1. Cambridge, Cambridge University Press. p. 26.
(1935) "The fiducial argument of variance components." Ann. Eug. 6, 391-398.
(1936) "The use of multiple measurements in taxonomic problems." Ann. Eug. 7, 179-188.
(1938) "The statistical utilization of multiple measurements." Ann. Eug. 8, 376-386.
(1939) "The sampling distribution of some statistics obtained from non-linear equations." Ann. Eug. 9, 238-249.

Ganguli, M. (1941) "A note on nested sampling." Sankhyā 5, 449-452.

Garwood, F. "Unpublished table of the lower 5% points of the noncentral chi-square."

Girshick, M.A. (1939) "On the sampling theory of roots of determinantal equations." Ann. Math. Stat. 10, 203-224.

Grayball, F.A. , Martin, F. and Godfrey, G. (1956) "Confidence intervals for variance ratios specifying genetic heritability." Biometrics 12, 99-109.

Hartley, H. O. and Pearson, E. S. (1954) Biometrika Tables for Statisticians. Vol. 1, Cambridge University Press.

Hotelling, H. (1931) "The generalization of Student's ratio." Ann. Math. Stat. 2. 36-378.
(1936) "Relations between two sets of variates." Biometrika 28, 321-377.
(1950) "A generalized T test and measure of multivariate dispersion." Proc. of Second Berkeley Symposium on Math. Stat. and Prob. 23-41. Berkeley, University of California Press.

Hsu, P.L. (1938) "Notes on Hotelling's generalized $T_{\bullet}^{2}$." Ann. Math. Stat. 9, 231-243.
(1939) "On the distribution of roots of certain determinantal equations." Ann. Eug. 9, 250-258.
(1940) "On the limiting distribution of roots of a determinantal equation." J. London Math. Soc. 16, 183-194.

Ito, K. (1956) "Asymptotic formulae for the distribution of Hotelling's generalized $T_{\bullet}^{2}$ - statistic." Ann. Math. Stat. 27, 1091-1105.

Keuls, M. (1952) "The use of the Studentized range in connection with an analysis of variance." Euphytica 1, 112-122.

Kramer, C.Y. (1955) "On the analysis of variance of a two-way classification with unequal sub-class numbers." Biometrics 11, 441-452.
(1956) "Extension of multiple measurements to group means with unequal numbers of replications." Biometrics 12, 307-310.
(1957) "Extension of multiple reange tests to group correlated adjusted means." Biometrics 13, 13-18.

Larsen, H.D. (1948). Rinehart Mathematical Tables Formulas and Curves. New York, Rinehart and Co. Inc., publishers.

Lawley, D.N. (1938a) "A generalization of Fisher's z test." Biometrika 30, 180-187.
(1938b) "Tests of significance for the latent roots of covariance and correlation matrices." Biometrika 43, 128-136.

Mahalanobis, P.C. (1927) "Analysis of race-mixtures in Bengal." J. Asiat. Soc. Bengal, 23, 301-333.
(1936) "On the generalized distance in statistics." Pro. Nat. Inst. Sci. India, 2, 49-55.

Mahalanobis, P.C., Majumdar, D.M. and Rao, C.R. (1949) "Anthropometric survey of the United Provinces, 1941. A statistical study." Sankhyā 9, 89-324.

Mood, A.M. (1950) Introduction to the Theory of Statistics. New York, McGraw-Hill Book Co.
(1951) "On the distribution of the characteristic roots of normal second-moment matrices." Ann. Math. Stat. 22, 266-273.

Morant, G.M. (1923) "Afirst study of the Tibetan skull." Biometrika 14, 193-260.
(1924) "A study of certain oriental series of crania including the Nepalese and Tibetan series in the British Museum (Natural History)." Biometrika 16, 1-105.
(1926) "Studies of Palaeolithic man. 1. The Chancelade skull and its relation to the modern Eskimo skull." Ann. Eug. 1. 257-276.

Nair, U.S. (1939) "The application of moment functions in the study of distribution laws in Statistics." Biometrika 30, 274-294.

Nanda, D.N. (1948) "Distribution of a root of a determinantal equation." Ann. Math. Stat. 19, 47-57
(1948) "Limiting distributions of a root of a determinantal equation." Ann. Math. Stat. 19,340-350.
(1950) "Distribution of the sum of roots of a determinantal equation under a certain condition." Ann. Math. 21, 432-439.

Nash, S.W. (1956) "Contributions to the theory of experiments with many treatments." Berkeley, University of California Press.

Nash, S.W. and Jolicoeur, P. (1959) "Calculating discriminant functions." Unpublished.

Newman, D. (1939) "The distribution of the range in samples from a normal population expressed in terms of an independent estimate of standard deviation." Biometrika 31, 20-30.

Patnaik, P.B. (1949) "The non-central chi-square and F distributions and their applications." Biometrika 36, 202-232.

Pearson, K. (1926) "On the coefficient of racial likeness." Biometrika18, 105-117.
(1928) "The application of the coefficient of the racial likeness to test the character of samples." Biometrika 20B, 294-300.

Pillai, K.C.S. (1953) "On the distribution of the sum of the roots of a determinantal equation." Abstract, Ann. Math. Stat. 24, 495.
(1954) "On the distribution of Hotelling's generalized T - test." Abstract, Ann. Math. Stat. 25, 412.
(1954) "On some distribution problems in multivariate analysis." Mimeo. Series No. 88. Institute of Statistics, University of North Carolina.
(1955) "Some new criteria in multivariate analysis." Ann. Math. Stat. 26, 117-121.
(1956) "On the distribution of the largest or the smallest root of a matrix in multivariate analysis." Biometrika 43, 122-127.
(1957) Concise Tables for Statistics. Manila, The Statistical Center, University of Phillipines.

Rao, C.R. (1946) "Tests with discriminant functions in multivariate analysis." Sankhyā 7, 407-414.
(1948) "The utilization of multiple measurements in problems of biological classification." J. Roy. Stat. Soc. B. 10, 159-203.
(1952) Advanced Statistical Methods in Biometric Research. New York, John Wiley and Sons.

Roy, S.N. (1939) "p - statistics or some generalizations in analysis of variance appropriate to multivariate problems." Sankhyā 4, 381-396.

(1942a) "Analysis of variance for multivariate normal populations - the sampling distribution of the requisite p-statistics on the null and non-null hypotheses." Sankhyā 6, 35-50.

(1942b) "The sampling distribution of p-statistics and certain allied statistics on the non-null hypothesis." Sankhyā 6, 15-34.

(1945) "The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null hypothesis." Sankhyā 7, 133-158.

Roy, S.N. and Bose, R.C. (1953) "Simultaneous confidence interval estimation." Ann. Math. Stat. 24, 513-536.

Roy, S.N. (1954) "Some further results in simultaneous confidence interval estimation." Ann. Math. Stat. 25, 725-776.

(1956) "Anote on some further results in simultaneous confidence interval estimation." Ann. Math. Stat. 27, 856-858.

(1957) Some Aspects of Multivariate Ananlysis. New York, John Wiley and Sons. Calcutta, Indian Statistical Institute.

Roy, S.N. and Gnanadesikan, R. (1957) "Further contributions to multivariate confidence bounds." Biometrika 44, 399-410.

(1959a) "Some contributions to Anova in one or more dimensions. I." Ann. Math. Stat. 30, 304-317.

(1959b) "Some contributions to Anova in one or more dimensions. II." Ann. Math. Stat. 30, 318-340.

Satterthwaite, F.E. (1941) "Synthesis of variance." Psychometrika 6, 309-316.

(1946) "An approximate distribution of estimates of variance components." Biometrics 1, 110-114.

Scheffé, H. (1953) "Amethod of judging all contrasts in the analysis of variance." Biometrika 40, 87-104.

Siotani, M. (1958) "Note on the utilization of the generalized Student ratio in the analysis of variance of dispersion." Tokyo, Annals of the Institute of Statistical Math., 9, 157-171.

(1959) "The extreme value of the generalized distance of the individual points in the multivariate normal samples." Tokyo, Annals of the Institute of Statistical Math., 10, 183-208.

Tang, P.C. (1938) "The power function of theanalysis of variance tests with tables and illustrations of their use." Stat. Res. Mem. 2, 126-149 and 8 pages of tables.

Tildesley, N.L. (1921) "A first study of the Burmese skull." Biometrika 13, 176-262.

Tukey, J.W. (1949) "Comparing individual means in the analysis of variance." Biometrics 5, 99-114.
(1951) "Quick and dirty methods in statistics." Part II. Simple analysis for standard designs. Proc. Fifth Annual Convention, Am. Soc. for Quality Control, 189-197.
(1952) "Allowances for various types of error rates." Unpublished invited address, Blacksburgh, Virginia meeting of the Institute of Math. Stat.
(1953) "The problem of multiple comparisons." Unpublished dittoed notes. Princeton University, Princeton, N.J. 396 pages.

Whittaker, E.T. and Watson, G.N, (1927) A Course of Modern Analysis. 4th. ed. Cambridge, Cambridge University Press.

Wilks, S.S. (1932) "Certain generalizations in the analysis of variance." Biometrika 26, 471-494.
(1950) Mathematical Statistics. Princeton, N.J., Princeton University Press.