ON THE PROBLEMS OF CONSTRUCTION AND STATISTICAL INFERENCE
ASSOCIATED WITH A GENERALIZATION OF CANONICAL VARIABLES

BY

ASHIS SEN GUPTA

TECHNICAL REPORT NO. 168
FEBRUARY 1982

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MCS 81-04262
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE PROBLEMS OF CONSTRUCTION AND STATISTICAL INFRINGEMENT
ASSOCIATED WITH A GENERALIZATION OF CANONICAL VARIABLES

BY

ASHIS SEN GUPTA

TECHNICAL REPORT NO. 168
FEBRUARY 1982 */

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MCS 81-04262 */

Ingram Olkin, Project Director

Also issued as Technical Report No. 52 under OFFICE OF NAVAL RESEARCH
Contract N00014-75-C-0442 (NR-042-034) Dept. of Statistics, Stanford University.
*/Technical Report No. 168 assigned in May of 1981 under previous
contract Number NSF MCS 78-07736.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
0. INTRODUCTION

Canonical correlation analysis plays an important role in applied research since it provides a method of creating meaningfully bivariate random variables from two groups of random variables, each consisting possibly of a large number of components. The main purpose of this study is to investigate approaches to deal with the problems of construction and statistical inference which are posed when one tries to extend the theory of canonical correlation to deal with more than two sets of variables.

Anderson (1958) had suggested an interesting constructional approach for extending the theory of canonical variables. Though Steel (1951) and Kettenring (1971) have attempted to construct generalized canonical variables based on Anderson's approach, their results seem formidable for practical use--as have been speculated by Anderson. This paper presents a simple solution to Anderson's problem after modifying it with a useful constraint. Also problems of statistical inference related to generalized canonical variables are formulated and solutions to some of them are presented in this paper. As an application, important practical problems similar to the one posed by Gnanadesikan (1977), p. 77 can be formulated as a hypothesis testing problem and appropriate tests can be performed.

From a review of the existing literature on generalizations of canonical correlations, it was found that all the approaches to construct generalized canonical variables, except that due to Vinograde (1950) have the two natural properties - (1) they all reduce to the classical method
when the number of groups is only two and (2) the criterion for selection
of generalized canonical variables optimizes some function of their
correlation matrix. No work seems to have been done on the problems
of statistical inference associated with the generalization of canonical
variables.

Section 2 introduces a new approach for the construction of
generalized canonical variables. The constraint of equi-correlation is
discussed in the light of a variety of results and applications while
the criterion of minimum generalized variance is shown to be of special
interest through its statistical interpretation and examples of
applications in a wide variety of fields. A brief formulation of the
problems of statistical inference to be investigated in this research is
also presented there.

In case of known population parameters, a derivation of the new
generalized canonical variables is given in Section 3. Statistical
estimation of these variables, when population parameters are unknown is
given next and the interesting theory for the singular case concludes
this section.

For the new generalized canonical variables, tests of hypotheses
concerning their equi-correlation coefficient and generalized variance
are of special interest. Using a large-sample approximation for the
distribution of the new variables in case of normal population, various
tests for the above hypotheses are proposed in Section 4. Some Properties
of the above tests and their relationship with the Likelihood Ratio Test
for correlation coefficient is also investigated in this section.

Tests for deciding whether the new generalized canonical variable
of a particular size is as good as the one with a larger number of
elements are given in Section 5. Also, in this section tests for comparing several new generalized canonical variables in some special cases are given using isotonic regression techniques.

Finally, some discussions and topics for future research are presented in Section 7.
1. DEVELOPMENT OF GENERALIZED CANONICAL VARIABLES

1.1 Genesis:

A random vector of interest, in many situations, poses both analytical and economical problems if it has too many components. The objective creation of a new random vector with a much smaller number of components from the original one with a large number of components can often be of great help in applied research. The first principal component in effect seeks to determine a linear combination of the components of a p-dimensional random vector which explains the maximum variance among a set of all possible standardized linear functions. This procedure has been extended by Gnanadesikan and Wilk (1969) to search for a non-linear combination, giving rise to non-linear first principal component. So, a p-dimensional vector can be represented by just a scalar variable. Very often, the components of the original vector can be meaningfully grouped into several disjoint subvectors. In the case of two groups, Hotelling (1936) proposed canonical variables, where the first canonical variable seeks a two-component vector, each element being a standardized linear combination of the corresponding group, such that the correlation between any two such standardized linear combinations is maximum. The main advantage of principal component and canonical correlation analysis is that they reduce the problem of dimensionality and further, meaningfully try to explain the variation in the set to a great extent. The method of
canonical correlation has been widely applied in economics, marketing and in most of the behavioral and social sciences.

Hotelling noted the need for the extension of his canonical variables to the case of several groups. The object here is the same as above. However, the criterion for the choice of the variables need to be carefully defined.

For the case of \( k \) groups, it is desirable that the criterion for the choice of the variates, should be some function of their correlation matrix. Various such criteria and hence generalizations of canonical variables have been proposed.

The case of a single variable per set has been studied by many authors, though from different viewpoints, e.g. Horst (1936), and McKeon (1966).

Kettenring (1971) has considered the algebraic derivation of the generalized canonical variables for more than two sets with several variables per set. An expository discussion of generalizations of canonical variables is given in Sen Gupta (1981c).
2. INTRODUCTION TO NEW GENERALIZED CANONICAL VARIABLES

2.1 Definitions:

In the previous section it was seen that various attempts to generalize canonical variables have been made. Though, no formal definition of GCV seems to exist, in the spirit of the above attempts the following definition can be given.

**Definition 1:** Suppose a random vector \( \mathbf{X} \) can be partitioned meaningfully into \( k \) disjoint groups or sub-vectors, \( X^{(1)}, \ldots, X^{(k)} \). The first GCV \( \mathbf{Y}^{(1)} \) with \( \mathbf{Y}^{(1)'} = [Y_1^{(1)}, \ldots, Y_k^{(1)}] = [f_1^{(1)}(X^{(1)}), \ldots, f_k^{(1)}(X^{(k)})] \), where \( f_{ij} \)s are real-valued functions, is a \( k \)-dimensional random variable, the components of which are chosen so as to optimize a criterion based on some function of their correlation matrix. The generalized higher order canonical variables \( \mathbf{Y}^{(2)}, \mathbf{Y}^{(3)}, \ldots \) are also \( k \)-dimensional random variables, the components of which are chosen so as to optimize the same criterion with some additional constraints imposed at each stage regarding the relationships among these variables.

Note, for all known methods \( f_{ij} \)s are linear functions. However, following the lines of non-linear principal component analysis, it is expected that non-linear GCVs can also be developed where the situation warrants the necessity of such a one.

In the present work, the search is for linear combinations
of the elements of each group, such that they are standardized and are equi-correlated. Formally, for the method to be presented here, we have

Definition 2: The first new GCV is the vector \( \mathbf{\bar{y}}^{(1)} = [y_1^{(1)}, \ldots, y_k^{(1)}] \)
\[ \begin{bmatrix} a_1^{(1)} \mathbf{x}_1^{(1)} & \ldots & a_k^{(1)} \mathbf{x}_k^{(1)} \end{bmatrix}, \]
where \( a_i^{(1)}, i=1,\ldots,k \) are chosen such that the \( y_i^{(1)}, i=1,\ldots,k \) are equi-correlated and the generalized variance of \( \mathbf{\bar{y}}^{(1)} \) is minimum. The \( s \)-th order new GCV, \( \mathbf{\bar{y}}^{(s)} \), \( s=2,3,\ldots \) has components \( y_i^{(s)} \), \( i=1,\ldots,k \); \( y_i^{(s)} = a_i^{(s)} \mathbf{x}_i \), where \( a_i^{(s)} \)'s are chosen so that \( y_i^{(s)} \) are equi-correlated, \( \mathbf{\bar{y}}^{(s)} \) is uncorrelated with \( \mathbf{\bar{y}}^{(r)} \), \( r<s \) and generalized variance of \( \mathbf{\bar{y}}^{(s)} \) is minimum.

2.2 The constraint of equi-correlation and the criterion of minimum generalized variance:

The motivations leading to seek such new GCVs are presented next.

Note that, though \( \mathbf{\bar{y}}^{(s)} \) has equi-correlated components, the original vector variable \( \mathbf{x} \) may have any arbitrary dispersion matrix.

(A) The equi-correlation model - the constraint: The equi-correlation model not only has wide applications and interesting properties but as we will see, leads to simplifications for the analysis of new GCVs.

(A.a) Examples and uses: The equi-correlation model has been greatly used in various fields, e.g. by Votaw (1950) in medical experiments, Seal (1964) in the study of basic patterns of growth of grasshoppers, Gupta and Panchapakesan (1969) in ranking and selection procedures, etc. Various problems of statistical inference have been researched using this model, e.g. in the same year (1968) in Biometrika by Geisser and Desu, Glesser and also Han. Because of its numerous applications, extensive tables for equi-correlated multivariate normal c.d.f. were published by Milton (1963).
(A.b) **Properties**: The equi-correlation structure possesses some interesting properties. Two of them are discussed below.

(i) In many tests of Multivariate Analysis, e.g. tests in MANOVA, Profile Analysis, Growth Curve Analysis etc. if sample size from each population is not at least \( p \), then in order that the mean square ratio in these Repeated Measurements Designs have exact \( F \)-distributions, the equi-correlation condition was thought to be necessary. However, Huynh and Feldt (1970) proved that this is a sufficient condition.

(ii) Another property follows from the following Lemma: For any correlation matrix \( R, R=(r_{ij}) \), \[ |R| < [1+(k-1)r][1-r] \]
with equality iff \( r_{ij} = \frac{\sum r_{ij}}{k(k-1)} = r \), for all \( i,j \) and where \( R \) is a \( k \times k \) matrix.


Note that using the constraint that the new GCVs are equi-correlated, we not only have great advantages from practical considerations, such as wide applications in a variety of fields and availability of tables for the probability integrals in case of multivariate normal population, but also from theoretical point of view, since now many tests in Multivariate Analysis can be performed on these new GCVs because they satisfy a sufficient condition for the validity of such test criteria.

Further, we have the important advantage of working with a vector having a much smaller number of components as compared to the original one. Also, if the criterion is minimization of generalized variance, in view of property (ii) above, the method seems to utilize the minimax principle.

(B) **The generalized variance - the criterion.** The importance of generalized variance in Multivariate Analysis is well known. We review some interesting features about it below.
(B.a) The generalized variance as a measure of multi-dimensional scatter:

The generalized variance of a random vector variable is the determinant of the dispersion matrix of that variable. For the correlation matrix $R$, $|R|$ was termed 'Scatter coefficient' by Frisch in 1929. This definition of the generalized variance is a natural extension of the variance in the univariate case as a measure of dispersion.

Here we will be concerned with only one type of measures of dispersion, namely, the scatter of a random variable from a point of reference (and will thus disregard generalizations of other types like range etc.). Some of the following discussions are found in Mathai (1967).

If $X'$ is a random variable, $m$ is a fixed point of reference and $X=X'-m$, then a measure of dispersion for $X'$ from $m$ can be defined by the following metric $D$ satisfying the following axioms:

K1. $D(X) \geq 0$, $D(X) = 0$ iff $X = 0$ almost surely.

K2. $D(aX) = |a| D(X)$, where $a$ is a scalar quantity.

K3. $D(X+Y) \leq D(X) + D(Y)$ where $Y'$ is another random variable and $Y=Y'-m$.

K4. $D(X) = 1$ if $|X| = 1$ almost surely.

A $D(X)$ satisfying K1 to K4 can be called a measure of dispersion in $X$.

From statistical considerations two more desirable properties of $D(X)$ are

K5. $D(aX+b) = |a|^r D(X)$, where $a, b$ and $r (r > 0)$ are scalars.

K6. $D(X+Z) = D(X) + D(Z)$, where $Z$ is another random variable independent of $X$.

If $E$ denotes the operator mathematical expectation then an example of a measure of dispersion is given by $D(X) = [E|X|^r]^{1/r}$ for fixed $r \geq 1$.

$E|X|^2 = \sigma^2$, satisfies K5 and K6, where $E(X') = m$.

If $(X', Y')$ is a bivariate random variable, $(m_1, m_2)$ is a point
of reference and $(X,Y) = (X'-m_1, X'-m_2)$ then the joint dispersion $D(X,Y)$ in $(X,Y)$ can be defined by the following axioms:

L1. $D(X,Y) = D(Y,X)$.

L2. $D(aX,Y) = a D(X,Y)$ where $a$ is a scalar quantity.

L3. $D(X+Z,Y) = D(X,Y) + D(Z,Y)$ where $Z = Z' - m_3$ and $Z'$ is another random variable having a joint distribution with $(X,Y)$.

L4. $D(X,X) = D(X)$ where $D(X)$ is an univariate measure of dispersion.

A desirable property here would be

L5. $D[A(X,Y)+B] = |A|^2 D(X,Y)$, where $r(r>0)$ is a scalar, $A,B$ are matrices with scalar elements and $A(X,Y)+B$ is a non-singular transformation of $(X,Y)$.

The concept of covariance provides an example of a measure of joint dispersion.

Finally, consider a random vector $X$ with $k$ components. Let $d_{ij}$ denote a measure of joint dispersion in the $i^{th}$ and $j^{th}$ variables. So the matrix $D_2 = (d_{ij})$ can be taken as a multivariate measure of joint dispersion in the $k$ variables. Note that $X$ provides an example of such a measure. A measure of generalized dispersion or multi-dimensional scatter would be a scalar quantity arising as a multivariate analogue of a univariate measure of dispersion. Hence such a measure can be defined as any norm of the matrix, i.e. should satisfy

M1. $||D|| > 0$, $||D|| = 0$ iff $D$ is a null matrix.

M2. $||aD|| = |a| ||D||$ where $a$ is a scalar quantity.

M3. $||C+D|| < ||C||+||D||$ where $C$ is another matrix defined similarly.

M4. $||CD|| < ||C|| . ||D||$

A desirable property here would be

M5. $||AD(X)+B|| = |A| r ||D||$ where $r > 0$ is a scalar and $A$ is non-singular.
Hence taking $\Sigma$, the usual variance-covariance matrix, for $D$ any norm of $\Sigma$ can be taken as a measure of generalized dispersion. Largest characteristic root of $\Sigma$ is an example of such a measure.

Note first that it is reasonable to expect that any scalar function of $\Sigma$, used as a measure of multi-dimensional statistical scatter should take into consideration the magnitude of the correlations among the variables. Secondly, van der Vaart (1965) has shown that the expected volume of the simplex formed by the $k+1$ random points in $k$ dimensions or $k$ random points and the mean vector is equal to the generalized variance, $|\Sigma|$. This is a natural generalization of the fact that the expected distance between two points or one point and the mean is the variance in one dimension. Thirdly, if the probability that a random point will lie in an ellipsoid of a $k$-dimensional distribution per unit volume is large, then the population is well concentrated about the mean. Finally, for a multivariate normal population, MLE of $|\Sigma|$ is the sample generalized variance as in the univariate case and distributional properties of the sample generalized variance has a marked similarity with that of the sample variance, e.g. for a random sample of size $n$, $n_s^2$ is distributed as $\sigma^2\chi^2_{n-1}$ in the univariate case and $n^k_s^2 s^2 k^2 \chi^2_{n-1} \ldots \chi^2_{n-k}$ in the multivariate case with a vector variable of $k$ components, where $s^2 k$ and $\sigma^2 k$ are the sample and population generalized variances respectively.

From the above properties and the works of Mathai (1967) entitled 'Dispersion theory' and Wilks (1967) entitled 'Multidimensional statistical scatter' the term generalized variance for the determinant of the dispersion matrix and its use as a measure of generalized dispersion seems to be very well coined (though it is a semi-norm).
(B.b) **Applications of the generalized variance:** Some situations are cited where the generalized variance can be used. Next, some examples where the use of the generalized variance has been advocated are presented.

Mathai (1967) pointed out that various statistical problems of estimation and tests of hypotheses can be looked upon as the study of properly defined measures of dispersion. Using the work of van der Vaart (1965), let $Q^2 = |B|^2 / (k + 1)$, where $X_i$ is $k \times 1$, $B = \begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_{k+1} \end{bmatrix}$ for $i = 1, \ldots, k+1$.

If $X \sim F$, then $Q^2 \sim H(F)$, with mean as the generalized variance of $X$.

This is a very important and useful observation for testing purposes. Hence, any test for the equality of two location parameters can be used to provide a test for equal generalized variances. So, standard non-parametric tests for location can be used when the underlying distribution is unknown. Steyn (1978) has pointed out that testing the null hypothesis that the population mean vector $\mu_X$ of the multivariate normal distribution with mean $\mu_X$ and dispersion matrix $\Sigma$ remains constant during the sampling process against the alternative that the mean vector varies during the process, is equivalent to testing the null hypothesis that the generalized variance $= |\Sigma|$ against the alternative hypothesis that the generalized variance $= |\Sigma^*|$, where $\Sigma^* = (I + 2D/n)\Sigma$, $D = \sum_{r=1}^{n} X_r X_r^T / n$, $n$ is the sample size and $X_r \sim N(\mu_X, \Sigma)$, $r = 1, \ldots, n$.

Several practical examples are given below.

**Example 1.** (Press, 1972). A transistor produced by company $j$, $C_j$ is characterised by a vector of $k$ measurements, $X_j : k \times 1$. Let $X_j \sim N(\mu_j, \Sigma_j)$, $j = 1, \ldots, r$. A purchaser would want to select the company with the property that he minimizes his chance of receiving a product that is sometimes of
unacceptably poor quality, that is choose that j for which $|\Sigma_j| = \min_{1 \leq i \leq r} |\Sigma_i|$

Assuming a multivariate normal distribution, Gnanadesikan and Gupta (1970) and earlier Eaton (1967) have considered the problem of ranking the r underlying populations according to the magnitude of their generalized variances. Assuming that the loss function associated with the given ranking problem depends only on the r generalized variances $(\sigma_1^{2k}, \ldots, \sigma_r^{2k})$, Eaton has shown that the decision rule which ranks the populations according to the sample generalized variances $(s_1^{2k}, \ldots, s_r^{2k})$ possesses the following properties. It is (i) minimax within the class of all decision rules (ii) admissible within the class of decision rules which depend only on $(s_1^{2k}, \ldots, s_r^{2k})$ and (iii) is the uniformly best decision rule among the class of rules which depend only on $(s_1^{2k}, \ldots, s_r^{2k})$ and are invariant under permutations of $(s_1^{2k}, \ldots, s_r^{2k})$.

Example 2. Though the definition of optimality becomes a problem in the multivariate situation, various authors have suggested the use of generalized variance in determining the optimum allocation of sampling units. Ghosh (1958) and Arvanitis and Afonja (1971) have proposed to determine the optimum sample size for the j-th strata, in stratified random sampling by minimizing the generalized variance of the sample means.

Example 3. Goodman (1966) has advocated the use of of the generalized variance to compare the overall variability of the populations of maize. It is apparent that a system of classification based on the overall similarities among the races is necessary both for the optimum usage of the race collections in breeding programs and for the study of the evolution of maize. Knowledge of the origin of the races against which values of the generalized variance can be compared, motivates here the
generalized variance as a reasonable measure of the overall variability and its use as a method of classifying the races of the maize.

(C) Minimization of the generalized variance - the choice: New GCVs are selected such that their generalized variance is minimum at each stage. The choice of minimization stems from two important considerations. Firstly, in the case of 2 groups, it is clear that maximizing the correlation of the 2 linear functions corresponding to the 2 groups, is equivalent to minimizing the generalized variance of the two linear combinations. This property is extended to the case of k groups. Secondly, as has been illustrated in the various examples in sub-section (B) above, minimization of the generalized variance is widely used in many contexts and is of great practical use in various situations.

2.3 Formulation of the problems:

Broadly speaking, the two distinct problems here are those of construction and statistical inference related to GCVs. The present work deals with a random vector which can be meaningfully grouped into k disjoint sub-vectors. For the problem of construction, in its complete generality, one would face at least three problems of optimization - (i) How to select the groups in the absence of a given grouping (ii) How to select the components of the GCVs and (iii) How to decide on the optimal stage of stopping for higher order GCVs. (i) can possibly be solved partly by Cluster Analysis. For (ii) some available methods have been reviewed above. In the case of more than 2 groups, for (iii), as Kettenring has remarked for some of the methods above the situation is 'somewhat arbitrary'. 
No work seems to be available for the problems of statistical inference related to the GCVs. Various problems would be interesting here. The distribution of GCVs needs to be derived. In case the form of the exact distribution is of not much use, suitable approximations and large sample distribution will be worth investigating. Several hypotheses develop here naturally from the definition of GCV and statistical tests are needed for these. Some such hypotheses are discussed next. (i) It would be natural to find out the extent of multidimensional statistical scatter explained by considering k groups. So, a test of hypothesis for a specific value of the generalized variance of the new GCV would be needed. Since one would want the new GCVs to have small scatter, the alternative hypothesis naturally should be of the less than type. So, we would test \( H_0: |\Sigma_{Y_j}| = \sigma_0^2 \) against \( H_1: |\Sigma_{Y_j}| < \sigma_0^2 \). (ii) Next, it would be interesting to see, if for the same dimension, regrouping of the variables in the original vector produces better result. For the new GCVs the related tests of hypotheses are - 
\[ H_{01}: |\Sigma_{Y1}| = |\Sigma_{Y2}| \] against \( H_{11}: |\Sigma_{Y1}| < |\Sigma_{Y2}| \); 
\[ H_{02}: |\Sigma_{Yj}|s \text{ are all equal} \] against \( H_{12}: |\Sigma_{Yj}|s \text{ are in a given order} \). (iii) Also, it is reasonable to explore the possibility of whether the consideration of k groups leads to any substantial gain as compared to less number of groups which may be explicitly stated a-priori or which may be constructed by reasonably regrouping the variables in a fewer number of groups or amalgamating some of the original groups. The advantage here would be, that, in terms of the criterion, a GCV with a fewer number of components may perform as good as, if not better, than the original one with a larger number of components. This consideration may be, as in (ii), extended to the case of more than two GCVs with the
possibility of some of them to have same number of components. Since the
criterion here must be comparable, the related tests of hypotheses become
\[ H_{01}: |\Sigma_{Y1}|^{1/k_1} = |\Sigma_{Y2}|^{1/k_2} \]
against \[ H_{11}: |\Sigma_{Y1}|^{1/k_1} < |\Sigma_{Y2}|^{1/k_2} \] \text{ and}
\[ H_{02}: |\Sigma_{Yj}|^{1/k_j} \] are all equal against \[ H_{12}: |\Sigma_{Yi}|^{1/k_i} < |\Sigma_{Yj}|^{1/k_j}, \ k_i < k_j \]
and a given ordering of generalized variances for new GCVs with
equal number of components.

Note that in (ii) \( Y_j \) referred to the new GCV obtained by the \( j \)-th mode of
grouping the original random vector, \( j = 1, 2, \ldots \) while in (iii) \( Y_j \)
referred to the new GCV with number of components \( k_j \), corresponding to
the \( j \)-th mode of grouping, \( j = 1, 2, \ldots \) An example where such a situation
may arise is provided by the following example [which corresponds to \( H_{01} \)
in (ii)].

Example: (Thurstone and Thurstone, 1941). Horst, Kettenring and Seal (1964)
discussed this example. Three (\( =k \)) sets of scores by several people on
three batteries of three tests each (group size = 3 for each group) were
obtained. The three tests in each battery were intended to measure,
respectively, the verbal, the numerical and the spatial abilities of the
persons tested. The correlation matrix of the original scores was given.
Gnanadesikan (1977) commented "An interesting alternative analysis in
this example would be to regroup the nine variables into three sets
corresponding to the three abilities measured rather than the three
batteries of tests." How profitable this alternative grouping would be
in terms of multidimensional scatter can be judged by the test for \( H_{01} \)
in (ii) above. Note that, since the \( Y_j \) s would be correlated, such tests
may not be very simple, even if the GCVs have a simple distribution.
3. DERIVATION OF THE NEW GENERALIZED CANONICAL VARIABLES

3.1 Notations

Let \( \mathbf{X}' = (X_1', \ldots, X_i', \ldots, X_k') \), \( X_i' = (X_{i1}, \ldots, X_{ip_i}) \), \( i = 1, \ldots, k \), be a random vector variable meaningfully partitioned into \( k \) disjoint sub-vectors. Let \( E(\mathbf{X}) \) denote the expectation of \( \mathbf{X} \) and since our interest lies only in the dispersion of \( \mathbf{X} \), assume \( E(\mathbf{X}) = 0 \). Further, let \( \Sigma(\mathbf{Z}) \) and \( |\Sigma(\mathbf{Z})| \) denote the dispersion matrix and the generalized variance of the random vector \( \mathbf{Z} \), respectively. Then,

\[
\Sigma(\mathbf{X}) = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\
\Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{k1} & \Sigma_{k2} & \cdots & \Sigma_{kk}
\end{bmatrix}
\begin{bmatrix}
P_1 \\
P_2 \\
\vdots \\
P_k
\end{bmatrix}
= \Sigma, \text{say}, \Sigma(\mathbf{X}_i) = \Sigma_{ii},
\]

Denote the \( s^{th} \) new generalized canonical variable by \( Y_i^{(s)'} = (Y_1^{(s)}, \ldots, Y_k^{(s)}) \), \( Y_i^{(s)} = a_i^{(s)}X_i' \), where \( a_i^{(s)} \) is the coefficient vector of \( X_i' \) at the \( s^{th} \) stage, \( i = 1, \ldots, k \). Let, \( a^{(s)} = (a_1^{(s)}, \ldots, a_k^{(s)}) \). Further, let \( \mathbf{a}' = (a_1', \ldots, a_k') \) denote the real but otherwise arbitrary coefficient vector which is used at the outset to derive the new generalized canonical variables \( Y \) at any stage and \( a_i \)'s satisfy the condition that \( E(Y) \) is equi-correlated, i.e.,
\[
\Sigma(Y) = \begin{bmatrix}
\alpha_1' \Sigma_{11} \alpha_1 & \alpha_1' \Sigma_{12} \alpha_2 & \cdots & \alpha_1' \Sigma_{1k} \alpha_k \\
\alpha_2' \Sigma_{21} \alpha_1 & \alpha_2' \Sigma_{22} \alpha_2 & \cdots & \alpha_2' \Sigma_{2k} \alpha_k \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_k' \Sigma_{k1} \alpha_1 & \alpha_k' \Sigma_{k2} \alpha_2 & \cdots & \alpha_k' \Sigma_{kk} \alpha_k
\end{bmatrix} = \begin{bmatrix}
1 & \rho & \cdots & \rho \\
\rho & 1 & \cdots & \rho \\
\vdots & \vdots & \ddots & \vdots \\
\rho & \rho & \cdots & 1
\end{bmatrix} = \Sigma_\rho \text{ say}
\]

where \( \rho \) denotes the equi-correlation coefficient for \( Y_i \)'s, \( i=1, \ldots, k \).

Also, \( \rho(s) \) will denote the \( s \)th new generalized canonical correlation coefficient.

3.2 Derivations of the new generalized canonical variables.

For the first new generalized canonical variable, we want to minimize \( |\Sigma_\rho| \) subject to scale conditions, that, for \( Y_i \)'s, variances are all unity and they are equi-correlated. We use the method of Lagrange's undetermined multiplier. Let,

\[
\phi = |\Sigma_\rho| + \sum_{i=1}^{k(k-1)} \lambda_i (\alpha_i' \Sigma_{ii} \alpha_i - 1) + \sum_{i \neq j} \Sigma \alpha_i \alpha_j (\Sigma \alpha_i' \Sigma_{ij} \alpha_j - 1)
\]

\[-k(k-1)\alpha_i' \Sigma_{ij} \alpha_j \]

The last term introduces the condition of equi-correlation of the \( Y_i \)'s and is justified by the result that, \( \bar{Z} = \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij} / mn = Z_{st} \) for all \( s, t \) iff \( Z_{ij} = Z_{kl} \) for all \( i, j, k, l \). The \( \lambda_i \)'s and \( \nu_{ij} \)'s are Lagrange's undetermined multipliers, with \( \nu_{ij} = \nu_{ji}, i \neq j \).

\[
\frac{\partial \phi}{\partial \alpha_t} = C(\rho) \left( \frac{2}{k(k-1)} \sum_{j(\neq t)} \Sigma \alpha_t' \alpha_j \right) + 2 \lambda_t \left( \frac{1}{k(k-1)} \sum_{i(\neq t)} \right)
\]

\[
+ 2(\Sigma \nu_{ij}) \sum_{j(\neq t)} \alpha_t' \alpha_j - 2k(k-1) \Sigma \nu_{ij} \alpha_t' \alpha_j = 0 \ldots \ldots (t.1)
\]
for $t = 1, \ldots, k$ and write $t+1 = k+1$ as 1, and where $C(\rho) = \partial |\Sigma_\rho| / \partial \rho$.

We also used $\rho = \Sigma \alpha_i \Sigma j \alpha_j / k(k-1)$ in the first term and the fact
that $\alpha_i \Sigma j \alpha_j = \alpha_j \Sigma i \alpha_i^{-1}$ for all $i, j = 1, \ldots, k$ in the last term.

Multiplying $(t \cdot 1)$ by $\alpha_1', \Sigma - k(k-1) \rho = 0$

or

$\{k(k-1)^{-2} (\lambda_t - \lambda_{t+1}) = \rho \Sigma \Sigma j (\alpha_j - \alpha_{t+1}) + (\alpha_{t+1} - \alpha_{t+1}) \}$ ...

A solution to the set of equations defined by $(t \cdot 2), t = 1, \ldots, k$

is given by

$\lambda_t = \lambda_{t+1}$ for all $t, t = 1, \ldots, k$. ...

Using these values in $(t \cdot 1)$ and $\lambda$ as the common value of the $\lambda_t$'s we

get the simpler equation, [since $(3.2.1)$ with $\alpha_{ij} = \alpha_{ji}$ gives $\lambda_{ij}$'s

all equal]

$C(\rho) \Sigma \Sigma j \alpha_j + 2 \frac{\lambda}{k(k-1)} (\Sigma t \alpha_t) = 0 ...

Multiplying the above equation by $\alpha_t'$ we get, using the conditions of

equi-correlation and variance of $Y_t = \alpha_t \Sigma t \alpha_t = 1, \lambda = -(k-1) \rho C(\rho)$.

Substituting this value of $\lambda$ in $(t \cdot 3)$ we finally get, assuming $C(\rho) \neq 0$.

$\Sigma \Sigma j \alpha_j - (k-1) \rho \Sigma t \alpha_t = 0 ...

$(t \cdot 4)$
In writing the above system of equations explicitly, we have

\[
\begin{bmatrix}
-(k-1)\rho \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\
\Sigma_{21} & -(k-1)\rho \Sigma_{22} & \cdots & \Sigma_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{k1} & \Sigma_{k2} & \cdots & -(k-1)\rho \Sigma_{kk}
\end{bmatrix}
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_k
\end{bmatrix} = 0 \quad \ldots (1)
\]

A non-trivial solution in \( \rho \) (which is necessary for a solution satisfying the condition that variances of \( Y_i \)'s are all unity) exists if and only if the coefficient matrix of the \( \alpha \) is singular. Writing this condition in a compact form we get,

\[
|\Sigma_{\od} - (k-1)\rho \Sigma_{\od\dagger}| = 0 \quad \text{or} \quad |\Sigma - \lambda^* \Sigma_{\od}| = 0 \quad \ldots (2)
\]

where

\[
\Sigma_{\od} = \begin{bmatrix}
0 & \Sigma_{12} & \cdots & \Sigma_{1k} \\
\Sigma_{21} & 0 & \cdots & \Sigma_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{k1} & \Sigma_{k2} & \cdots & 0
\end{bmatrix}, \quad \Sigma_{\od\dagger} = \begin{bmatrix}
\Sigma_{11} & 0 & \cdots & 0 \\
0 & \Sigma_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Sigma_{kk}
\end{bmatrix}
\]

and \( \lambda^* = 1 + (k-1)\rho \).

We may call \( \Sigma_{\od} \) a diagonal super-matrix (a term due to Horst) and \( \Sigma_{\od\dagger} \) an off-diagonal super-matrix with respect to \( \Sigma \).

**Note 1.** The solution in (2) coincides with that of MceKen's and hence to that of various other authors as pointed out in Chapter I.

The criterion and derivation were of course different here.

2. In order that the \( \rho \)'s obtained from (2) satisfy the condition of a correlation matrix, we must have that \(-1/(k-1) \leq \rho \leq 1\). Here \( \rho \leq 1 \).
by definition. Next, from Theorem 1 in Appendix we note that $\lambda^* \geq 0$, since $\Sigma$ is positive semi-definite and assume that $\Sigma_d$ is positive definite. [If $\Sigma_d$ is positive semi-definite, then at least one of the $l_{ii}$'s must be singular. The assumption reasonably requires that at least within each group we do not have any redundant variables - those which bear linear dependence with other elements in the same group]. Hence, $\rho \geq \frac{-1}{k-1}$.

We solve the determinantal equations in (2) for the $p$-roots in terms of $\lambda^*_i$'s and then by the one-one relationship, determine the corresponding $\rho_i$'s. Since our object is to minimize $|\Sigma_{\rho}|$ we choose that $\rho_i$ for which $|\Sigma_{\rho_i}|$ is minimum. Now,

$$\frac{\partial}{\partial \rho} |\Sigma_{\rho}| = -k(k-1)\rho(1-\rho)^{k-2} = C(\rho).$$

So, $|\Sigma_{\rho}|^+$ for $\rho < 0$ & $+\rho$ for $\rho > 0$.

Hence, we need to consider both positive and negative $\rho_i$'s and then choose the one say $\rho^{(1)}$ for which $|\Sigma_{\rho^{(1)}}|$ is minimum. The coefficient vector $\underline{\alpha}^{(1)}$ corresponding to $\rho^{(1)}$ is then determined from (1) and the conditions of equicorrelations and unit variances for the components of $\underline{X}^{(1)}$. Then, the first new GCV is given by $\underline{\Sigma}' = \underline{\Sigma}^{(1)}', = (\underline{\alpha}^{(1)}', \underline{x}_1, \ldots, \underline{x}_k', \underline{x}_k)$ where $\underline{\alpha}^{(1)}$ = ($\alpha_1^{(1)}', \ldots, \alpha_k^{(1)}'$) is partitioned in a fashion corresponding to that of $\underline{X}$.

3. We had previously assumed that $C(\rho) \neq 0$, which means $\rho \neq 0$ and $\rho \neq 1$. From (2), note that if $|\Sigma_{\rho d}| \neq 0$, then $\rho \neq 0$ and if $|\Sigma| \neq 0$ then $\rho \neq 1$ (and $\rho \neq \frac{-1}{k-1}$).
For higher order new generalized canonical variables, the constraints are next discussed and then the derivation is given below.

In case of canonical correlation analysis, i.e., $k=2$, $p_1 \leq p_2$ at the first stage one seeks the pair $(Y_1^{(1)}, Y_2^{(1)}) = Y^{(1)}$, with the maximum possible correlation. The procedure may be pursued further until $p_1$ pairs of canonical variables have been determined. At the $s^{th}$ stage, the $s^{th}$ pair of canonical variables $Y^{(s)}$ are found such that the corresponding $s^{th}$ canonical correlation, $p^{(s)}$ is the maximum correlation attainable satisfying the conditions,

$$\text{corr}(Y_1^{(s)}, Y_1^{(t)}) = \text{corr}(Y_2^{(s)}, Y_2^{(t)}) = \text{corr}(Y_1^{(s)}, Y_2^{(t)})$$

$$= \text{corr}(Y_1^{(t)}, Y_2^{(s)}) = 0, \ t = 1, \ldots, s-1.$$

Usually the first two conditions are sufficient, since they imply the latter ones and sometimes even only the first. Kettenring has considered in detail the above situations.

For GCVs, the same approach may be followed, introducing suitable restrictions such that the GCV for a particular stage are distinct from those of the previous stages. At the $(r+1)^{th}$ stage we invoke the restrictions,

$$\text{corr}(Y_1^{(r+1)}, Y_2^{(r+1)}) = 0, \ell = 1, \ldots, r; \ t, j = 1, \ldots, k$$

which is equivalent to the restrictions above for $k=2$, i.e., canonical variables. This type of restriction has been demonstrated by Kettenring to be equivalent to the restrictions

$$\text{corr}(F_1^{(r+1)}, F_j^{(r+1)}) = 0, \ell = 1, \ldots, r; t, j = 1, \ldots, k$$

where $F_t^{(r+1)}$ are the best fitting factors associated with the
(r+1)th stage fit of a k factor model like \( y^{(1)} = \sum_{j=1}^{k} \xi_j^{(1)} F_j^{(1)} + e_1 \)
where the \( \xi_j^{(1)} \) are arbitrary non-null vectors, the \( F_j^{(1)} \) are stan-
dardized random variables and \( e_1 \) is a vector of error variables -
using a criterion such as the one where the factors are determined
so that \( F_1^{(1)} \) is the most important, \( F_2^{(1)} \) the second most important
and so on. He has also pointed out interesting relationships of
the above restrictions with the MAXVAR and MINVAR procedures in a
factor analytic set-up.

At the (r+1)th stage using the above restriction, i.e.,
\[
\alpha_i^{(s)} \xi_j^{(s)} = 0 \quad l = 1, \ldots, r; i, j = 1, \ldots, k; \text{ or equivalently}
\]
\[
\alpha_i^{(s)} \xi_j^{(1)} = 0 = \alpha_i^{(1)} \xi_j^{(s)}
\]
we have,
\[
\phi_{r+1} = \phi + \sum_{ijs} \theta_{ijs} (\alpha_i^{(s)} \xi_j^{(s)} - \alpha_j^{(s)} \xi_i^{(s)})
\]
where \( \phi \) is the function defined in section 3.2. Hence, for \( t=1, \ldots, k \),
\[
\frac{\partial \phi_{r+1}}{\partial \alpha_t} = \sum_{ijs} \theta_{ijs} \xi_j^{(s)} \alpha_i^{(s)} \xi_j^{(s)} = 0 \quad (t.5)
\]
Pre-multiplying (t.5) by \( \alpha_t^{(l)} \), we get
\[
(\sum_{j=tj}^{tj} - \sum_{i=it}^{it}) \rho_l^{(l)} + \xi_{tl} = 0 \quad \ldots \quad (t.6)
\]
(t.6)-(t+1.6) \Rightarrow [\sum_{j=tj}^{tj} \theta_{t+1j}^{(l)} - \sum_{i=it}^{it} \theta_{it}^{(l)} \theta_{it}^{(l)}] \rho_l^{(l)} = \xi_{t+1l} - \xi_{tl}^{(l)} \quad t=1, \ldots, k
A solution to these are obtained by taking $\theta_{ts}$ all equal, $\zeta_{ts}$ all equal. But from (t.6) this means $\zeta_{ts} = 0$ for all $t,s,t=1,\ldots,k$, $s=1,\ldots,r$; and using these values for $\theta_{ts}$ and $\zeta_{ts}$ in (t.5) we get (t.1). Thus the previous arguments are applicable and we get (t.4) and (1). At the $(r+1)^{th}$ stage, we choose that root of $(2)$ say $\rho_{r+1}$ which gives the $(r+1)^{th}$ lowest value of $|\Sigma_{\rho}|$ and call it the $(r+1)^{th}$ new generalized canonical correlation and the variable $y^{(r+1)} = (\alpha_{1}^{(r+1)}, \ldots, \alpha_{k}^{(r+1)}, x^{r+1})$ obtained from a corresponding coefficient vector as the $(r+1)^{th}$ new generalized canonical variable.

3.3 Sample new generalized canonical variables:

Since in practice both $\mu$ and $\Sigma$, the population mean and dispersion matrix are unknown, we need to estimate new GCVs from a sample of independent observations from the distribution of $X$. To simplify matters we now need to assume that $X \sim_{d} \mathcal{N}(\mu, \Sigma)$. The procedure for non-normal populations may be pursued as done for canonical correlations, e.g. as in Rao (1973). Let $x^{u}, u = 1,\ldots,n$ be a random sample of $n$ observations from the population of $X$ and let $\overline{x}$ and $S$ be the sample mean and dispersion matrix. Corresponding to the partition of $X$, partition $S$. Then the MLE of $\Sigma_{ij}$ and $\Sigma$ are $S_{ij}$ and $S$ respectively. Now since $(\mu, (\Sigma, \Sigma), \Lambda, \Lambda)$, where $\Lambda = \text{diag}(\rho^{(1)}, \ldots, \rho^{(p)})$, $\Lambda = (\alpha^{(1)}, \ldots, \alpha^{(p)})$, is a single-valued function of $(\mu, \Sigma)$, it follows from Appendix A.2, that the MLE of $\rho^{(i)}$ is given by the corresponding characteristic root $r^{(i)}$, $i = 1,\ldots,p$, of $SS_{d}^{-1}$. Also the MLE of $\alpha^{(i)}$ is given by $\alpha^{(i)}$ satisfying $\alpha^{(i)} \alpha^{(i)'} = 1$, $(SS_{d}^{-1} - r^{(i)} I) \alpha^{(i)} = 0$. Hence we get,

**Theorem 1** The MLEs of the population non-zero new generalized canonical
correlations \( \rho^{(1)}, \ldots, \rho^{(q)} \), \( q \) being the number of non-zero roots, and the corresponding normalized coefficient vectors \( \overline{a}^{(1)}, \ldots, \overline{a}^{(q)} \), are given by respectively the ordered roots \( r^{(1)}, \ldots, r^{(q)} \) (i.e. \( r^{(i)} \) gives the \( i \)th minimum value of \( |\Sigma_r| \) among the values of \( r \), \( r = r^{(j)}, j = 1, \ldots, p \)) of the determinantal equation \( |S - (k-1)rS_d| = 0 \) and \( \overline{a}^{(1)}, \ldots, \overline{a}^{(q)} \), where \( a^{(i)} \)s satisfy \( (S-(k-1)r^{(i)}S_d) \overline{a}^{(i)} = 0 \), \( a^{(i)}u, S_{uv}a^{(i)} = 1 \), \( a^{(i)}u, S_{uv}a^{(i)} = r^{(i)}, u \neq v, u,v = 1, \ldots, k, \ i = 1, \ldots, q \).

Note that, though unlikely, some of the non-zero \( \rho^{(i)} \)s may be known to be equal. The estimates have to be modified then. Standard derivations of MLE for the canonical correlations (e.g. Giri (1977)), whose approach has been followed here) do not consider such cases. However, Anderson (1963) has given MLEs of characteristic roots in case of known multiple population roots for principal component analysis. The questions of under what situations some \( \rho^{(i)} \)s are equal and how to estimate them in those events seem to be interesting problems for further investigation. Also note that in the above derivation of MLEs we have used the Principle of Invariance of MLEs for non one-one functions advocated by Zehna (1966). This principle is stated in section A.2 of the Appendix.

3.4 The Singular Case:

Very often, when the vector variable under consideration has a large number of components, the dispersion matrix for the variable turns out to be singular. One is then faced with the problem of defining the generalized correlations and associated variables. In particular, for the generalized canonical correlations the following result overcomes that difficulty.
Theorem 4.4.1 The generalized canonical correlations, for the new method, are given by $\rho = (\rho^*-1)/(k-1)$ where $\rho^*$ are the non-zero roots of $|\Sigma_{k} \Sigma_{d}^- - \rho^* I| = 0$, $\Sigma_{d}^-$ being any g-inverse of $\Sigma_{d}'$.


Various other generalizations of canonical correlations are properly defined for the singular case by the application of the same Theorem 1 in Sen Gupta (1980).
4. STATISTICAL INference ASSOCIATED WITH A SINGLE SET OF
NEW GENERALIZED CANONICAL VARIABLES

4.1 Distribution of the new generalized canonical variables:

The exact distribution of even the canonical variables with the
special case of a multivariate normal population is complicated and is of
doubtful practical importance (see e.g. Kshirsagar, 1972). The situation
here is even more complex and so only the large sample distribution will
be studied here. The coefficient vector \( a_i^{(s)} \), induced MLE of \( a_i^{(s)} \), for \( X_i \)
with \( s = 1 \) will be of particular importance, \( i = 1, \ldots, k \). To simplify
matters, we further assume that the population is multivariate normal
because of the well known properties corresponding to linear functions of
the components and the generalized variance of a random vector having a
multivariate normal distribution. Since the estimated coefficient vector
\( \hat{a}_i^{(s)} \) is obtained by the principle of induced MLE, properties of such
estimators need to be considered. The asymptotic efficiency of induced
MLEs under certain regularity conditions is implicitly given in the proof
by Zacks (1971) in Theorem 5.4.2 and there it is shown that \( \hat{g}(\theta) \) can be
taken as an induced MLE for \( g(\theta) \) when \( g(\theta) \) is a non one-one function.
However, the case of the new GCVs is more complicated. We now have \( \theta \) as a
\( k(k-1)/2 \) - component vector instead of a scalar and \( g(\theta) \) in addition to
being non one-one can be scalar (for \( p \)) or a vector (for the coefficient
vector \( \alpha \). However the induced MLEs \( \hat{\rho}(s) \) and \( \hat{\alpha}(s) \) of \( \rho(s) \) and \( \alpha(s) \) respectively, are expected to be good approximations to corresponding population values under suitable comparisons. The validity of this approximation is shown below. We thus treat \( \hat{\alpha}(s) \), in large samples, as the true coefficient vector itself and consider it as a vector of constants and for clarity write it as \( \hat{\beta}(s) \). \( \hat{\beta}(1) \) will be written as \( \hat{\beta} \). Assuming that \( X \sim N_p(0, \Sigma) \), then \( Y' = (\beta_1x_1, \ldots, \beta_kx_k) = (y_1, \ldots, y_k) \sim N_k(0, \Sigma_{(1)}) \), in large samples.

Assume that non-zero roots are all distinct. Then from a theorem due to Hurwitz (see Appendix) it follows that there exists a \( \delta \) such that there will be exactly the same number of non-zero eigen values for \( A+E \) as for \( A \) when \( ||E|| < \delta \), where \( || \cdot || \) is a matrix-norm. Let \( \lambda' \) have the same rank among the eigen values of \( A+E \) as \( \lambda \) has among those of \( A \). To make them unique, normalize the corresponding eigen vectors such that the first non-zero component of each is positive.

Consider without loss of generality the eigen-value \( \lambda^*(1) \) of \( \Sigma_d^{-1} \) (corresponding to \( \rho(1) \)) and the corresponding MLE, say \( g \). By a standard theorem in complex analysis \( g \) is a continuous function of \( (S, S_d) \), because \( g \) is a continuous function of the coefficients obtained from the equation \( |S - gs_d| = 0 \) and in turn these coefficients are continuous functions of \( (S, S_d) \). The coefficient vector \( \alpha(1) \) corresponding to \( g \), obtained by continuous operations on \( g \) and elements of \( S \) and \( S_d \), is also a continuous function of \( (S, S_d) \). But \( (S, S_d)^P (\Sigma, \Sigma_d) \) and so \( \alpha(1)^P \alpha(1) \), i.e. \( \alpha(1) \) is a consistent estimator of \( \alpha(1) \) and so in large samples, the approximation considered above is meaningful.
4.2 Tests for equi-correlation coefficient $\rho^{(1)}$:

The value of $\rho^{(1)}$ by itself, may be of interest. It may also be compared with the first canonical correlation when a sub-division of $X$ into two groups also seems meaningful. Suppose $X_1, \ldots, X_n$ constitute an independent sample from $N_p(0, \Sigma)$ and let $Y_1$ be the sample first new GCV. Then from section 4.1, in large samples $Y_1 \sim N_k(0, \Sigma_\rho)$. Let $I$ and $E$ be the identity matrix and the matrix with all elements equal to unity. Then,

$$\Sigma_\rho = (1-\rho)I + \rho E; \quad \Sigma_\rho^{-1} = (1-\rho)^{-1}I - \rho(1-\rho)^{-1}(1+(k-1)\rho)^{-1}E = [c_{ij}],$$

where

$$c_{ii} = \frac{1+(k-2)\rho}{(1-\rho)(1+(k-1)\rho)} \quad \text{and} \quad c_{ij} = -\rho/(1-\rho)(1+(k-1)\rho), \quad i \neq j.$$

Hence the density function for non-singular $\Sigma_\rho$ can be written as,

$$f(Y_1; \rho) = \frac{1}{(2\pi)^{k/2}|\Sigma_\rho|^{1/2}} \exp \left[ -\frac{1}{2} \left( \frac{(\Sigma y_1^2)}{(1-\rho)} + \frac{(\Sigma y_1^2(-\rho))}{(1+(k-1)\rho)(1-\rho)} \right) \right].$$

The above representation is particularly useful because it shows that,

(i) there does not exist any one-dimensional sufficient statistic for $\rho$

(ii) $\Sigma (y_i - \bar{y})^2$ and $\bar{y} = \Sigma y_i/k$ are independent and

(iii) the part of the exponent within the second bracket is monotonically decreasing in $\rho$ with positive probability.

We want to test $H_0: \rho = \rho_0$ against $H_1: \rho < (>) \rho_0$ or against $H_2: \rho \neq \rho_0$.

**Likelihood ratio test:** For testing $H_0$ against $H_2$ the LRT is derived below. We also write $\rho = \rho^{(1)}$ for simplicity, as stated in the outset and also suppose the original sample of size $n$ is split into $m$ independent subsamples, giving rise to $m$ independent first new GCVs. The LRT can be performed even if $m = 1$. The likelihood function can be written from (1) above easily and differentiating this with respect to $\rho$ and equating the derivative to zero we have,
\[ g(\rho) = (k-1)k\rho(1-\rho)(1+(k-1)\rho) + \sum_j y_{ij}^2 \{1+(k-1)\rho\} + \sum_j \sum_i y_{ij}^2 \{1+(k-1)\rho^2\} = 0, \]

where \( y_{ij}' = (y_{1j}',...,y_{kj}') \) is the sample first new GCV obtained from the \( j \)th sub-sample, \( j = 1,...,m \). This is a cubic equation in \( \rho \), two of whose roots may be complex. Now, \( g(-1/(k-1)), g(1) \) and \( g(0) \) are all positive (with probability one). Thus, it is not very obvious from \( g(\rho) \) that an admissible solution of \( \rho \), say \( \hat{\rho} \), where \(-1/(k-1)<\hat{\rho}<1 \), will always exist. [However, in various cases admissible solutions may exist, e.g. if \( \sum y_{ij}^2 < 1/2 \), and \( \sum \sum y_{ij}^2 < 1/2 \) then \( g(-1/k) \) is negative. Hence, an admissible solution here lies in \((-1/(k-1),-1/k]\).] However, considering the likelihood function directly it can be shown that there exist at least one such admissible solution. Further, in cases of several admissible solutions (at most three), by principle of Maximum Likelihood, we choose as the MLE of \( \rho \), that which corresponds to the largest value of the Likelihood function and call it \( \hat{\rho} \). Thus we get the following

**Theorem 1**: Let \( Y \sim N_k(O,\Sigma_\rho) \). Then if \( Y_1,...,Y_m \) constitute an independent random sample from the above population, the Likelihood Ratio Test for testing \( H_0: \rho = \rho_0 \) against the alternative \( H_1: \rho \neq \rho_0 \), is given by

Reject \( H_0 \) iff

\[ \lambda = \frac{\sum_{\rho} y_{ij}^2}{\sum_{\rho_0} y_{ij}^2} \exp\left[ -\frac{1}{2} \{ a(f_1(\rho_0) - f_1(\hat{\rho})) + b(f_2(\rho_0) - f_2(\hat{\rho})) \} \right] < K, \text{a constant} \]

where \( \hat{\rho} \) is the MLE of \( \rho \), \( a = \sum_{\rho} y_{ij}^2 \), \( b = \sum_{\rho} (\sum_{\rho} y_{ij}^2)^2 \), \( f_1(\rho) = 1/(1-\rho) \), \( f_2(\rho) = -\rho/(1+(k-1)\rho)(1-\rho) \) and \( K \) is a constant to be determined so that the level of the test meets the specified value.

Under \( H_0 \), for large \( m \), \( -2 \ln \lambda \sim \chi_1^2 \) d.f.

The exact distribution of \( \hat{\rho} \) and the LRT seem quite complicated.

Due to the above difficulties in LRT a new test based on the Best Unbiased Estimator of \( \rho \) has been considered in Sen Gupta (1981b). The exact null and non-null distribution of the test statistic has been obtained in terms of Kummer's function and the test has been shown to
unbiased against one-sided alternatives. The distribution in (5.2.1) is also interesting because it constitutes a practical example of a member of Efron's curved exponential family. For detailed discussion the reader is referred to Sen Gupta (1981b).
4.3 Tests for a specified value of the generalized variance of the new
generalized canonical variables.

Since the criterion for optimization is the generalized variance, a test for generalized variance is needed here to judge the performance of the new GCVs. The Likelihood Ratio Test for generalized variance, for the special structure of the dispersion matrix here, admists of some simplifications. If, \(|\Sigma_\hat{\rho}| = \sigma_0^2 < 1\), then there are precisely two real distinct solutions, \(\rho_2 > 0 > \rho_1\), say.

Then (using MLE of \(\rho\) with \(m = 1\) in Section 4.2) we have, by an application of the results in Sen Gupta (1981b),

**Theorem 1.** The Likelihood Ratio Test for \(H_0: |\Sigma_\rho| = \sigma_0^2\) against
\(H_1: |\Sigma_\rho| \neq \sigma_0^2\) is given by,

\[
\text{Reject } H_0 \text{ iff } \lambda = \left(\frac{|\Sigma_\hat{\rho}|}{\sigma_0}\right)^{k/2} \exp\left\{-\frac{k}{2} \text{ Tr}\left(\Sigma_\hat{\rho}^{-1} - \Sigma_\rho^{-1}\right)Y^\prime Y\right\} < C
\]

where \(\hat{\rho}\) is MLE of \(\rho\), \(\rho^*\) is such that, \(f(Y_1, \rho^*) = \max_{\rho_1, \rho_2} f(Y_1; \rho)\) and \(C\) is a constant to be determined such that the test has the desired level.

If the original sample is split into \(m\) independent subsamples, so that \(Y_{1j}, j = 1, \ldots, m\) are new GCVs based on the \(j^{\text{th}}\) subsample, then,

\[
\text{Reject } H_0 \text{ iff } \lambda_m = \left(\frac{|\Sigma_\hat{\rho}|}{\sigma_0}\right)^{mk/2} \exp\left\{-\frac{k}{2} \sum_{j=1}^m \text{ Tr}\left(\Sigma_\hat{\rho}^{-1} - \Sigma_\rho^{-1}\right)Y_{1j}^\prime Y_{1j}\right\} < C'
\]

where \(\rho^{**}\) is such that \(\prod_{j=1}^m f(Y_{1j}, \rho^{**}) = \max_{\rho_1, \rho_2} \prod_{j=1}^m f(Y_{1j}, \rho)\), \(\hat{\rho}\) is the MLE of \(\rho\) and under \(H_0\), for large \(m\), \(-2 \ln \lambda_m \sim \chi^2_1\).

For other interesting tests for the generalized variance, \(|\Sigma_\rho|\) see Sen Gupta (1981b).
5. COMPARISON OF SEVERAL NEW GENERALIZED CANONICAL VARIABLES

5.1 Tests for equality of multidimensional scatter of two new GCVs:

As in Section 4 we will consider only the first new GCV. The problems that arise when the random variable \( X \) is grouped in two different ways are related to (a) same number of groups with same number of elements in each, (b) same number of groups with different number of elements and (c) different number of groups. Since (a) and (b) give rise to same number of components in the corresponding first new GCVs, for testing purposes they can be tackled in a similar manner.

A random sample of size \( n \), \( X_1, \ldots, X_n \), is taken from \( \mathbb{N}_p(Q, \Sigma) \). Let \( Y \) and \( Z \) be first new GCVs of order \( r \) and \( r-1 \) respectively. Under the approximations of Section 4.1 construct \( Y_1 = (\beta_Y^{*}, X_{11}^{*}, \ldots, \beta_Y^{*}, X_{1r}^{*}) \) and \( Z_1 = (\beta_Z^{*}, X_{11}^{**}, \ldots, \beta_Z^{*}, X_{1r-1}^{**}) \) for each \( X_1, i = 1, \ldots, n \) where \( X = (X_1^{*}, \ldots, X_r^{*}) \) and \( X^{**} = (X_1^{**}, \ldots, X_{r-1}^{**}) \) are the partitions of \( X \) into orders \( r \) and \( r-1 \) corresponding to \( Y \) and \( Z \) respectively. Then \( (Y_1, \ldots, Y_n) \) and \( (Z_1, \ldots, Z_n) \) can be considered, each, separately as a sample of GCVs corresponding to \( Y \) and \( Z \) respectively. But, \( Y_1 \) and \( Z_1 \) will still be dependent. This dependency can be avoided if a sub-sample \( X_1, \ldots, X_m \) of size \( m \) is used to estimate \( \Sigma \) and then obtain \( Y_{m}, i = 1, \ldots, m \), from this estimate and sub-sample and the remaining sample, \( X_j, j = m+1, \ldots, n \) is used to give an independent estimate of \( \Sigma \) and obtain \( Z_j, j = 1, \ldots, (n-m) \). A LRT can then be performed for the equality of the generalized variances of \( Y \) and \( Z \), using advantageously the independence of \( Y_{m} \) and \( Z_{j} \). The same method can be used for more than two new GCVs. However, this procedure will be quite inefficient because \( \Sigma \) will be estimated on small sample sizes and so will

33
of course be the coefficient vectors. Further, the number of $X_i$ and $Z_i$ will be greatly reduced. A modified method which uses the same number of GCVs $\tilde{X}_i$ and $\tilde{Z}_i$ as that of the $X_i$, namely $n$, and where $\Sigma$ is estimated using the entire sample of size $n$ is proposed below. The method is also applicable in the case of several new GCVs.

**Lemma 1:** If $X_1, \ldots, X_n$ are i.i.d. and $Y_1, \ldots, Y_n$ are i.i.d., $X_i$ is independent of $Y_j$, $i \neq j$, and $X_i$ and $Y_i$ have multivariate normal distributions, then $X = \Sigma a_i X_i$ and $Y = \Sigma b_i Y_i$ are independent if $\Sigma a_i b_i = 0$.

**Proof.** Cov $(X, Y) = \Sigma \sigma_i \sigma_j$, Cov $(X_i, Y_j) = \Sigma a_i b_j$. Thus, Cov $(X, Y) = \Sigma \Sigma a_i b_i = 0$ and hence the lemma follows from the normality of $X$ and $Y$.

Let $Y = \Sigma a_i Y_i$ and $Z = \Sigma b_i Z_i$. Let $s = 1, \ldots, r - 1$ and $j = 1, \ldots, n$.

Choose $a_i b_j$ such that $\Sigma a_i b_j = 0$. Also if $Y \sim N_r(0, \Sigma)$ and $Z \sim N_r(0, \Sigma)$ then, Var $(Y) = r(1+\rho_r)$ and Var $(Z) = (r-1)(1+r-2\rho_r)$. If we have chosen $a_i b_j$ such that $\Sigma a_i^2 = \Sigma b_i^2 = 1$. Call the resulting variables $Y^*$ and $Z^*$. Using the approximations of section 4.1 and by Lemma 1, in large samples $Y^*$ and $Z^*$ are independently distributed.

We want to test $H_0: \sigma_1 = \sigma_2$ where $\sigma_1 = c^\frac{1}{\rho_r}$ and $\sigma_2 = c^\frac{1}{(r-1)}$ against $H_1: \sigma_1 \neq \sigma_2$ or $H_2: \sigma_1 \neq \sigma_2$. For testing $H_0$ against $H_2$, we use LRT. Differentiating the Likelihood function with respect to $\rho_r$ and $\rho_{r-1}$ and equating the derivatives to zero, we have,

$y^{**2} = r(1+r-1\rho_r)$ and $z^{**2} = (r-1)(1+r-2\rho_{r-1})$ (estimate of variances)

Then, under $\Omega$, maximum of the Likelihood function $L(y^{**}, z^{**})$ is $1/(y^{**} z^{**})$.

Under $H_0$, we have to maximize the Likelihood function and then consider,

$L(y^{**}, z^{**}) = L(y^{**}, z^{**}) + \lambda(\sigma_1 - \sigma_2)$, where $\lambda$ is Lagrange's multiplier.

The solution seems to need numerical or iterative methods. With these solutions, let Max $L^{**}$ be $f(y^{**}, z^{**})$ a function of $y^{**}$ and $z^{**}$.
Reject $H_0$ iff $\eta = f(y^{**}, z^{**})/(y^{**} z^{**}) < \eta_o$, a constant.

Under $H_0$, $-2 \ln \eta$ is approximately distributed as a $\chi^2$ variable with 1 d.f.

So far, except the conditions (i) $\sum a_i^2 = \sum b_i^2 = 1$ and (ii) $\sum a_i b_i = 0$, $a_i$, $b_i$s are arbitrary. Since the statistic in LRT looks formidable, the power function of the above test may be difficult to compute. However, we may try to improve on the estimators, $\hat{\rho}_i$, $i = r-1, r-2$, obtained above when $L(y^{**}, z^{**})$ was maximized under $\Omega$, are unbiased estimates. We will try to find minimum variance unbiased estimators among the above class. This amounts to minimizing $\sum a_i^4$ and in view of condition (i) above, subject to $\sum a_i^2 = 1$. The solution is easily found to be $a_i^2 = 1/n$, $i = 1, \ldots, n$. 
5.2 Tests for equality of multidimensional scatter of more than two new

generalized canonical variables:

Here we will consider separately GCVs with same and different number

of components, because some simplifications are available in the former case.

(a) GCVs with same number of components and with equi-correlations all \( > 0 \) or

all \( < 0 \): Consider the case of \( s \) (say, even) GCVs with same number of com-

ponents \( r \) and equi-correlation coefficients \( \rho_i \), \( i = 1, \ldots, s \) all \( > 0 \) or all \( < 0 \).

Take \( s/2 \) independent sub-samples of same size and construct \((y_{11}^{**}, y_{22}^{**}), \ldots, (y_{s-1}^{**}, y_{ss}^{**})\)

as mentioned in section 5.1. Thus \( y_{**}^{i} \)'s are independent normal variables. Then,

the Likelihood function is given by \( L \), where, writing \( y_{i}^{*} = y_{i}^{**} \),

\[
\ln L = C - \frac{1}{2} \sum \ln(r(1+r-r\rho_i)) + y_i^2/(r(1+r-r\rho_i)), \quad \text{where } C = (2\pi)^{s/2}.
\]

Then, in the notations of Barlow et al (1972) letting,

\[
X = \{1, \ldots, s\}, \quad g(i) = y_i^2, \quad f(i) = \{r(1+r-r\rho_i)\} \quad \text{and } w(i) = 1, \quad i = 1, \ldots, s
\]

we have the

Lemma 1. The Isotonic regression \( g^* \) maximizes \( L \) under \( H_1 \): The generalized

variance of \( i^{\text{th}} \) GCV is greater than that of the \( j^{\text{th}} \) GCV, \( i > j \).

Proof: Note first that since equi-correlations are all \( < 0 \) (all \( > 0 \)) and

ordering of the generalized variances of the new GCVs gives the same (reverse)

ordering of the corresponding equi-correlation coefficients. With the above

values of \( X, g(i), f(i) \) and \( w(i) \), we use Theorem 1.10 of Barlow et al quoted

in our Appendix and the relevant notations. Let \( \Phi(u) = -\ln u \). To complete

the proof it suffices to note that maximizing \( \ln L \) is equivalent to minimizing

\( \Sigma[\log f(i) + \{g(i)/f(i)\}] \) both subject to \( f \) isotonic, since the first and last

1 terms of \( \Delta_{g} \) do not involve \( f \). (When \( \rho_i > 0 \) \( \forall i \), isotonicity is w.r.t. reverse of

the natural order on \( X \) and for computational purposes a formula is

\[
g_i^* = \min_{s<1} \max_{t>1} \sum_{j=s}^{t} y_{ij}^2.
\]

When \( \rho_i < 0 \) \( \forall i \), \( \min \) and \( \max \) are interchanged in the above formula.)
Using the fact that under $H_0$ Max $L$ is obtained by simply replacing $1 + (r-1)\rho_1$ by $\nu_1^2$ in $L$ and Lemma 1, the LRT statistic $\lambda$ can be easily found. However, even under $H_0$ the distribution of the statistic seems to be complicated. In large sample, $-2 \ln \lambda \sim \chi^2$ (see Barlow p. 198).

(b) GCVs with different number of components: In view of the above difficulties even in the special cases as in (a), we propose to use the approximation to the distribution of the generalized variance by Gamma distribution by Hoel (1937). Barlow et al. (p. 198) has considered a similar case and with a slight modification it follows that the LRT statistic $-2 \ln \lambda \sim \chi^2$. We reject the hypothesis of equal generalized variances in favor of a given ordered relationship among them, if $-2 \ln \lambda$ exceeds the upper $\alpha$ per-cent point of the distribution of $\chi^2$. Some tables are given in Barlow et al. For this test, however, see comments in Section 7.
6. DISCUSSIONS AND AN EXAMPLE

In this chapter new GCVs are found for the classical example due to Thurstone and Thurstone mentioned in Section 2.3 and studied by Horst, Kettenring, Seal and Gnanadesikan. The correlation matrix, $R$, for the 9 variables we consider here was given by,

$$
R = \begin{pmatrix}
.636 & .126 & .059 & .626 & .195 & .059 \\
I & -.021 & .633 & .049 & .035 & .459 & .129 \\
.016 & .157 & .521 & .048 & .238 & .426 \\
I & .039 & .532 & .190 \\
.067 & .258 & .299 & I
\end{pmatrix}
$$

where $I$ is the $3 \times 3$ identity matrix.

Here $k = 3$ and $p = 9$. Since $R_d^{-1} = I$ (the $9 \times 9$ identity matrix) we need only to solve $|R - \hat{\lambda}^* I| = 0$, i.e. determine the characteristic roots, say $\hat{\lambda}_i^*$, of $R$ and then find $\rho_i = (\hat{\lambda}_i^* - 1)/2$, $i = 1, \ldots, 9$. The coefficient vector for the $s^{th}$ new GCV is the characteristic vector corresponding to $\hat{\lambda}^*(s)$, say, where the corresponding equi-correlation coefficient $\rho(s)$ is the one which gives the $s^{th}$ minimum value among $(1 + 2\rho_i)(1 - \rho_i)^2$, $i = 1, \ldots, 9$; $s = 1, \ldots, 9$. 

38
Table 1 exhibits the order of the first four new GCVs and the corresponding generalized variances. We note that (1) The new GCCs are computed quite easily e.g. \( r^{(1)} = .745 \). For the GENVAR method and several others discussed by Steel and Kettenring, however, both the GCCs and the coefficients of the GCVs for each stage need to be computed through extensive iterative methods. (2) For a negative value of \( r \), i.e. \( r = -.385 \), we have the third new GCV and (3) Since \(|\Sigma_r|\) for the third new GCV is as high as .45, it seems that the higher order new GCVs are redundant from practical considerations.

The computational aspect of the new GCVs is quite interesting. At the first stage, for the coefficients \( a^{(1)}_i \), one needs to solve a simultaneous system of non-linear equations

\[
\begin{align*}
& a^{(1)}_i S^{-1}_{ii} a^{(1)}_i = 1, \\
& a^{(1)}_i S^{-1}_{ij} a^{(1)}_j = r^{(1)}, \quad i \neq j, i, j = 1, \ldots, k
\end{align*}
\]

where \( r^{(1)} \) has already been obtained (as above) from (2) of p.20. \( S \) can be first transformed to a (block) correlation matrix \( R \), as done by Horst for our above example. The conditions then become

\[
\begin{align*}
& b^{(1)}_i b^{(1)}_i = 1, \\
& b^{(1)}_i R_{ij} b^{(1)}_j = r^{(1)}, \\
& (b^{(1)}_i = a^{(1)}_i S^{-1/2}_{ii}, i \neq j, i, j = 1, \ldots, k.
\end{align*}
\]

Further, a polar transformation on \( b^{(1)}_i \) reduces by one the number of unknown variables in each set. The usual Gauss-Seidel method which determines one variable from each equation needs to be modified for our purpose. Suppose an initial value \( b^{(1)}_i \) of \( b^{(1)}_i \) is given. Usually a good initial value is

\[
b^{(1)}_i = \frac{b^{(1)}_i}{d^{(1)}_i}, \quad i = 1, \ldots, k
\]

where \( \bar{b}^{(1)}_i = (\bar{b}^{(1)}_1, \ldots, \bar{b}^{(1)}_k) \) is the eigen vector of \( R \) corresponding to \( r^{(1)} \). We solve for \( b^{(1)}_2 \) from

\[
b^{(1)}_2 R_{21} b^{(1)}_1 = r^{(1)}
\]

giving \( b^{(1)}_2 \), and so on till finally for \( b^{(1)}_k \)

\[
b^{(1)}_k R_{k1} b^{(1)}_1 = r^{(1)}
\]

giving \( b^{(1)}_k \). The \( b^{(1)}_1, \ldots, b^{(1)}_k \) are standardized to have unit
length, yielding the coefficient vector at the end of the first iteration as, using the same notation, \( \hat{b}_i^{(1)} \) from \( b_i^{(1)} \). The procedure is then further iterated. Since \( r^{(1)} \) is known previously, the iteration can be terminated at the \( n \)-th stage if

\[
\sum_{i=1}^{k} \left| \frac{b_i^{(1)}}{n-i+1} - \frac{b_i^{(1)}}{n-i+1} - r^{(1)} \right| < \varepsilon
\]

for some pre-assigned \( \varepsilon \) and the suffix \( k+1 \) is replaced by \( l \). It is reasonable to make a preliminary polar transformation on \( b_i^{(1)} \), each separately for the \( i \)-th set, \( i=1, \ldots, k \), as suggested earlier. The entire procedure outlined above can then be performed in terms of the transformed variables. The ZXMIN subroutine in IMSL can then be advantageously exploited to solve for \( b_i^{(1)} \). This procedure is applicable to coefficients for higher stage new GCVs also, \( b_2^{(2)}, b_3^{(3)}, \ldots \) — only the equations defining the additional constraints at the corresponding stages have now to be also solved simultaneously.

For our example, a single iteration of the above procedure yielded

\[
\begin{align*}
\hat{b}_1^{(1)} &= (.642691, \ .619972, \ .450091) \\
\hat{b}_2^{(1)} &= (.543281, \ .733599, \ .408263) \\
\hat{b}_3^{(1)} &= (.668885, \ .663320, \ .335559) \quad \text{and}
\end{align*}
\]

\[
\hat{b}_1^{(1)} R_2 b_2^{(1)} = .745, \quad \hat{b}_2^{(1)} R_3 b_3^{(1)} = .748, \quad \hat{b}_3^{(1)} R_1 \hat{b}_1^{(1)} = .749,
\]

the value of \( r^{(1)} \) being .745 as determined at the outset.

It is seen from this example that unlike several of the more reasonable methods studied in details by Kettenring, the new GCVs are quite straightforward and easy to compute. We have to merely find the characteristic roots and the standardized eigen vectors of a given matrix. From the
computational point of view the gain is significant. Also, we have already seen from Sections 4 and 5 several advantages of new GCVs from the point of view of statistical inference.

| \( \lambda^* \) | \( r(s) \) | \( |\Sigma r(s)| \) | new GCV |
|----------------|--------|-----------------|---------|
| 2.48985        | .745   | .16             | first stage |
| 2.16422        | .582   | .33             | second stage |
| .23548         | -.385  | .45             | third stage |
| 1.61986        | .310   | .77             | fourth stage |
7. TOPICS FOR FUTURE RESEARCH

The tests based on the approximation to the distribution of the generalized variance by a gamma distribution may, with further research, be possibly improved upon since this approximation does not take into account the additional information of equi-correlation structure of the dispersion matrix. This leads to the estimation and exact distributional problems associated with the generalized variance corresponding to an equi-correlated dispersion matrix. Hence, new directions need to be sought for these problems and once the solutions are obtained, problems of statistical inference presented in this report can be considered with these modifications. A simulation study comparing the modified methods with the present one will be worth attempting. Properties of the proposed tests may also be considered in more details.

Relationships of GCVs with such important fields as Time Series Analysis, Regression Analysis, Prediction Theory, MANOVA, Discriminant Analysis, and Scaling and Factor Analysis are topics for future research. These are expected to achieve simplifications in terms of cost and analysis. The case with nominal variables (used extensively in Bio-Medical sciences) and a Ranking and Selection Procedure based on GCVs are interesting problems for further investigations.

Statistical inference associated with the previous GCVs and generalized variances in general, are important topics for future research.
For LRTs for generalized variances and their associated properties see Sen Gupta (1981c). Such tests are applicable to generalized canonical variables obtained by minimizing the generalized variance, of general structure, as considered by Anderson (1958), Kettenring (1971), Steel (1951) and others.

Acknowledgements

It's a pleasure to acknowledge the helpful discussions I had with Professor C.R. Rao during this research. I am thankful to Professor T.W. Anderson and Professor R. C. Srivastava for their kind encouragements. Finally, I am grateful to Professor K. S. Lii for the computations required in the example.
APPENDIX

A.1 Theorem 1: Let $A$ be positive definite and $B$ be positive semi-definite. Then the roots of $B$ in the metric of $A$ are non-negative (where the roots of the equation $|\lambda A - B| = 0$ are called the characteristic roots of $B$ in the metric of $A$). If $B$ is also positive definite, then such roots are positive.


A.2 Theorem 2: (Principle of Invariance of MLE) Let $\theta \in \Omega$ (an interval in a k-dimensional Euclidean space) and let $L(\theta)$ denote the Likelihood function - a mapping from $\Omega$ to the real line $\mathbb{R}$. Assume that the MLE $\hat{\theta}$ of $\theta$ exists so that $\hat{\theta} \in \Omega$ and $L(\hat{\theta}) \geq L(\theta)$ for all $\theta \in \Omega$. Let $f$ be an arbitrary transformation mapping $\Omega$ to $\Omega^*$ (an interval in an r-dimensional Euclidean space, $1 \leq r \leq k$). Then $f(\hat{\theta})$ is a maximum induced likelihood estimator of $f(\theta)$.


A.3 Theorem 3: (Theorem of Hurwitz) Let $\{f_n(x)\}$ be a sequence of analytic functions regular in a region $G$, and let this sequence be uniformly convergent in every closed subset of $G$. Suppose the analytic function

$$\lim_{n \to \infty} f_n(x) = f(x)$$

does not vanish identically. Then if $x = a$ is a zero of $f(x)$ of order $k$, a neighbourhood $|x - a| < \delta$ of $x = a$ and a number $N$ exist such that if $n > N$, $f_n(x)$ has exactly $k$ zeros in $|x - a| < \delta$.

Proof: See page 22 of Szego (1939).

A.4 Definition (Isotonic regression): Let $X$ be the finite set $\{x_1, \ldots, x_k\}$ with the simple order $x_1 \leq x_2 \leq \cdots \leq x_k$. A real valued function $f$ on $X$ is isotonic if $x, y \in X$ and $x \preceq y$ imply $f(x) \leq f(y)$. Let $g$ be a given function on
X and w a given positive function on X. An isotonic function $g^*$ on X is an isotonic regression of g with weights w with respect to the simple ordering $x_1 \preceq x_2 \preceq \ldots \preceq x_k$ if it minimizes in the class of isotonic function f on X the sum $\sum_{x \in X} [g(x) - f(x)]^2 w(x)$. When the weight function and the simple ordering are understood, we call $g^*$ simply an isotonic regression of g.

Note: A binary relation "\preceq" on X establishes a 'simple order' on X if

1. it is reflexive: $x \preceq x$ for $x \in X$;
2. it is transitive: $x, y, z \in X$, $x \preceq y$, $y \preceq z$ imply $x \preceq z$;
3. it is antisymmetric: $x, y \in X$, $x \preceq y$, $y \preceq x$ imply $x = y$;
4. every two elements are comparable: $x, y \in X$ implies either $x \preceq y$ or $y \preceq x$.

Algorithms for isotonic regression: Barlow et al (1972) have discussed in details an algorithm called the Pool-Adjacent-Violators algorithm for finding the isotonic regression $g^*$. They have also discussed a scheme in order to program this algorithm for a computer (see page 72 of Barlow et al). Using essentially this scheme Kruskal (1964) has written a program to carry it out as a part of a large program.

Theorem 4: (Theorem 1.10 of Barlow et al). For $\phi$ convex, let

$$\Delta_\phi(g(x), f(x)) \equiv \Delta(g, f) = \phi(g) - \phi(f) - (g - f)\phi'(f)$$

where $\phi'$ is the derivative of $\phi$ at $f$, or if it does not have a derivative at $f$, $\phi'(f)$ denotes any number between the left and the right derivative at $f$. If f is isotonic on X and if the range of f is in I then

$$\sum_{x \in X} [g(x), f(x)] w(x) \geq \sum_{x \in X} [g^*(x), f(x)] w(x) + \sum_{x \in X} \Delta(g^*(x), f(x)) w(x).$$

Consequently $g^*$ minimizes $\sum_{x \in X} [g(x), f(x)] w(x)$ in the class of isotonic f with range in I and maximizes $\sum_{x \in X} [\phi[f(x)] + [g(x) - f(x)]\phi'(f(x))] w(x)$. 
The minimizing (maximizing) function is unique if \( \phi \) is strictly convex.

A corollary which is useful for our purpose is

**Corollary:** Let \( \psi_1, \ldots, \psi_p \) be arbitrary real valued functions and let \( h_1, \ldots, h_m \) be isotonic functions on \( X \). Then \( g^* \) minimizes \( \sum_x \Delta [g(x), f(x)] w(x) \) in the class of isotonic functions \( f \) with range in \( I \) satisfying any or all of the side conditions

\[
\sum_x [g(x) - f(x)] \psi_j [f(x)] w(x) = 0, \quad j = 1, \ldots, p
\]

\[
\sum_x f(x) h_j(x) w(x) \geq \sum x g(x) h_j(x) w(x), \quad j = 1, \ldots, m.
\]

**Proof:** See corollary in page 42 of Barlow et al.
BIBLIOGRAPHY


Horst, P. (1936): "Obtaining a composite measure from a number of different measures of the same attribute," Psychometrika 1, 53-60.


