1. INTRODUCTION

Canonical correlation analysis is a technique introduced by Hotelling (1936) to describe structural relationships between two sets of variables. It can be considered an extension of multiple regression, which describes a relationship of one variable to a set of variables. Just as analysis of variance and analysis of covariance are special cases of multiple regression, many multivariate procedures such as discrimination and multivariate analysis of variance are special cases of canonical correlation analysis.

Multivariate vectors are frequently incompletely observed in experimentation, and many techniques exist for using partial records in the estimation of population characteristics. We desire to extend estimation concepts which use partial records to canonical correlation. This research will consist primarily of two parts:

(a) A method for testing the significance of the Akaike information contained in the incomplete data vectors.

(b) A procedure to estimate canonical correlations that will utilize all data available, including partial observations.

Estimation of parameters of a multivariate normal distribution when data are incomplete has been discussed by many authors. However, none of these consider the estimation of canonical correlations using incomplete data sets. Most have found maximum likelihood estimates of parameters for various special cases. Wilks (1932) considers estimation for a bivariate normal population with missing data in both variables, as does Anderson (1957) who also indicates how to extend his results to a p-variate normal distribution for certain patterns of missing data. Srivastava and Zaatari (1973) use Monte Carlo simulation to compare four estimators of the covariance matrix of a bivariate normal distribution. Edgett (1956) gives maximum likelihood estimates of parameters of a trivariate normal distribution when observations on only one variable are missing. Lord (1955) and Mathai (1951) also find estimates of parameters of a trivariate normal distribution in other special cases.

Buck (1960) proposes using regression to estimate any missing observations and proceeding with classical full data estimation techniques. Beale and Little (1975) modify Buck's technique so that it gives maximum likelihood estimates when the population is normal. They also use Monte Carlo simulation to compare their method with others. Estimating regression coefficients using data with missing observations has also been studied. Edgett (1956) gives an example, and Afifi and Elashoff (1966) review and compare various missing value techniques in regression analysis.

Huseby, Schwertman and Allen (1980) give an algorithm that estimates the mean vectors and covariance matrix from several populations with a common covariance matrix when the data vectors have missing elements.

Hocking and Smith (1968) develop a method of estimating parameters of a p-variate normal distribution with zero mean vector in which the missing observations are not required to follow certain patterns. Their estimation technique can be summarized as follows:

(a) The data are divided into groups according to which variates are missing.

(b) Initial estimates of the parameters are obtained from that group of observations with no missing variates.

(c) These initial estimators are modified by adjoining, optimally, the information in the remaining groups in a sequential manner until all data is used.

For "nested" situations, the method yields estimates that are maximum likelihood, but in any case, they show that the estimates are consistent and asymptotically efficient. Hartley and Hocking (1971) use the same method to give parameter estimates when the mean vector for the normal distribution is unknown.

Hocking and Marx (1979) use the same method as Hocking and Smith (1968) to derive estimates, but their use of matrices simplifies the notation and gives the estimates in a form that is easily implemented on a computer. They also give exact, small sample, moments of the estimators for the case of two data groups. Their procedure will be used in our study.

Several attempts have been made to consolidate missing data estimation techniques into a more general framework. Two examples are Orchard and Woodbury's (1972) missing information principle (MIP) and Dempster, Laird and Rubin's (1977) expectation-maximization (EM) algorithm. Many of the previously discussed estimation techniques are special cases of one of these two algorithms.

2. Information and Incomplete Data

When some multivariate vectors are only partially observed in sampling, it is helpful to know whether these records contribute sufficiently to warrant the use of a missing data estimation scheme. Using data (both partial and complete records) resulting from sampling a multivariate normal population we will recall the reductions in variance of estimates made possible by including partial records, and also, we will derive a test of significance for the information in the partial records using Akaike's information criterion (Akaike 1974).

Let $\mathbf{x} = (x_1, \ldots, x_k)$ be a parametric
function unbiasedly estimated by \( t = t(x_1, \ldots, x_n) \), then the Cramer-Rao lower bound on the variance of \( t \) is
\[
\text{Var}(t) \geq \frac{\delta^2}{\sum_{i=1}^{n} \partial^2 g_i / \partial x_i^2},
\]
where \( \delta^2 = \{\partial^2 g_i / \partial x_i^2 \}_{i=1}^{n} \), and \( g_i \) is the likelihood function. Smith and Hocking (1969) give an algorithm to calculate \( U \) when sampling from a multivariate normal distribution and illustrate their technique in missing data situations.

For example, if \( n \) observations are taken from a bivariate normal population with zero mean vector and covariance matrix \( \Sigma = \begin{pmatrix} 0 & \rho \\ \rho & 1 \end{pmatrix} \), then the lower bounds on unbiased estimates of the elements of \( \Sigma \) are given by
\[
\begin{align*}
\text{MVB}(\sigma_{11}) &= 2\sigma_{11}^2/N \\
\text{MVB}(\sigma_{12}) &= (\sigma_{11}^2 + \sigma_{12}^2) / n_1 - 2(\sigma_{11}^2 / n_1)\sigma_{12}^2 \\
\text{MVB}(\sigma_{22}) &= 2\sigma_{22}^2 / n_1 - 2\sigma_{12}^2 / n_1 \sigma_{11}^2
\end{align*}
\]
where \( N = n_1 + n_2 \). Similarly, the minimum variance bound on unbiased estimates of the square of the correlation coefficient \( \rho \) is
\[
\frac{4\rho^2(1-\rho^2)^2 - 2n_2}{n_1^2} \leq 1 - 2(1-\rho^2)^2.
\]
Thus, the observance of the partial records lowers the variance bound even when the variance observed has no obvious relation to the parameter of interest (e.g. \( \rho \)).

When both mean and covariance matrices are unknown, the minimum variance bound can be extended to more than two groups of "nested" data, establishing at each stage whether the partial records contain significant Akaike information. This feature allows the user to determine whether or not to use or discard the partial records.

A direct method of establishing the statistical significance of the information contained in the partial records when sampling from a multivariate normal population can also be given. Let AIC be Akaike Information Criterion, where
\[
\text{AIC} = -2\ln L + 2k
\]
and \( k \) is the number of unconstrained parameters. Akaike (1974, 1979a, 1979b) and others have minimized the AIC to estimate the order of an autoregressive time series model and to estimate the order of a polynomial model.

We propose using the AIC to test whether or not the information contained in partial records is useful. This is done by comparing the AIC with parameter estimates using full records only to the AIC evaluated using combined estimates which utilize all data records, both full and partial. The combined estimates will be calculated according to an algorithm developed by Hocking and Marx (1979).

2.1 Unknown Mean Vector

To illustrate consider sampling from \( p \) variate normal distribution \( N_p(\mu, \Sigma) \) where \( \Sigma \) is known. Let \( N \) observations be taken, \( n \) of which are complete records and \( n-p \) are partial records from \( N_p(\mu, \Sigma) \), where \( q < p \), \( \mu = \mu_1 \), \( \Sigma = \Sigma_1 \). Thus, the observance of the partial records lowers the variance bound even when the variance observed has no obvious relation to the parameter of interest (e.g. \( \rho \)).

Thus, the AIC evaluated using the full data estimates only is
\[
\text{AIC}_1 = -2\ln L + n_1\ln|\Sigma_1| - \frac{1}{2}\text{tr}(E_1^{-1}) + \frac{1}{2}\text{tr}(E_2^{-1})
\]
and with the combined estimates the criterion becomes
\[
\text{AIC}_2 = -2\ln L + n_1\ln|\Sigma_1| - \frac{1}{2}\text{tr}(E_1^{-1}) + \frac{1}{2}\text{tr}(E_2^{-1})
\]
where \( E_1 \) and \( E_2 \) are elements of the variance bound even when the variate observed has no obvious relation to the parameter of interest (e.g. \( \rho \)).

Thus, \( Q_1 \) is a test statistic for significance of the added Akaike information contained in the \( n_2 \) partial records (as utilized by the combined estimate of \( \rho \)).

These results may be extended to more than two groups of "nested" data, establishing at each stage whether the partial records contain significant Akaike information. Also note that this test may be performed before \( \mu \) is actually calculated since \( Q_1 \) (a Mahalanobis distance) does not contain \( \mu \). This feature allows the user to determine whether or not to make use of the more complicated estimation algorithm or to simply ignore the partial records.

2.2 Unknown Mean Vector and Covariance Matrix

When both mean and covariance matrices are unknown in the sampling scheme described above, the joint log likelihood function may be written as
\[
\ln L = \ln L_1 + \ln L_2,
\]
where
\[
\begin{align*}
\ln L_1 &= -2n_1 - n_1\ln|\Sigma_1| - \frac{1}{2}\text{tr}(E_1^{-1}) \\
\ln L_2 &= -2n_2 - n_2\ln|\Sigma_2| - \frac{1}{2}\text{tr}(E_2^{-1})
\end{align*}
\]
and
\[
\begin{align*}
M_1 &= n_1(\Sigma_1 + (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T) \\
M_2 &= n_2(\Sigma_2 + (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T) + 2p
\end{align*}
\]
with \( \Sigma_1, \Sigma_2 \) being the usual maximum likelihood estimates from \( L_1, L_2 \). Hocking and Marx show that \( \Sigma \) and \( \Sigma_2 \) are maximized by \( \mu \) and that
\[ \bar{\mu} = \hat{\mu} + EZ \]
\[ \bar{\Sigma} = \bar{\Sigma}_1 - N\mathbf{B}\left(\Sigma_1 + \Sigma_2\right)\mathbf{B}'/n_2 + n_1\mathbf{BZZ}'/n_2, \]
where \( Z = \mathbf{B}_{\bar{\mu}} - \hat{\mu}_2 \) and \( B \) is as before. Note that \( B(\hat{\mu}_1, \bar{\Sigma}_1) = B(\hat{\mu}, \bar{\Sigma}) \).

Using the formulation of AIC from Section 2.1 and the fact that \( \hat{\mu} \) and \( \bar{\Sigma} \) are m.l.e. for \( \mu \) and \( \Sigma \), we have
\[ q_2 = AIC_{12} - AIC_1 \leq 0, \]
and will show that \( q_2 \) is a decreasing function of
\[ T_1^2 = Z'(D_{\Sigma_1}D_1 - \Sigma_1)^{-1}Z \]
and
\[ T_2^2 = tr(D_{\Sigma_1}D_1 - \Sigma_1)^{-1}Z Z'. \]
Note that \( T_1^2 \) and \( T_2^2 \) follow Hotelling's \( T^2 \) and generalized Hotelling's \( T^2 \) distributions, respectively (see Kshirsagar 1972).

To simplify the derivation let
\[ A_1 = tr(G_{t}^{-1}L_{t}^{-1}M_{t}) \].
Thus
\[ AIC = -2(c_1 + c_2) + n_1ln|\Sigma| + n_2ln|\mathbf{D}D'| \]
\[ + A_1 + A_2 + 2(p + \frac{1}{2}p(p+1)) \]
for the two group case. Let \( A_{1t}(-) \) be \( A_t \) evaluated with the m.l.e. estimates of \( \mu_t \) and \( \Sigma_t \), and let \( A_{2t}(-) \) be \( A_t \) evaluated with the full data estimates. Consider the following lemmas.

**Lemma 2.1** \( A_{1t}(\cdot) = n_1 p. \)

Proof: \( A_{1t}(\cdot) = tr\left\{n_1[\Sigma_1^{-1}c_1 + (\Sigma_2^{-1}c_2)G_{1}^{-1}L_{1}^{-1}G_{1}] \right\} \]
\[ = n_1 tr[I_q - (\mathbf{D}D')^{-1}Z Z'(\mathbf{D}D')^{-1}]/N \]
where \( Z = \bar{\mu}_2 - \bar{\mu}_1 \).

Proof: \( A_{1t}(\cdot) = tr\left\{n_1[\Sigma_1^{-1}c_1 + (\Sigma_2^{-1}c_2)G_{1}^{-1}L_{1}^{-1}G_{1}] \right\} \]
\[ = n_1 tr[I_q - (\mathbf{D}D')^{-1}Z Z'(\mathbf{D}D')^{-1}]/N \]
where \( Z = \bar{\mu}_2 - \bar{\mu}_1 \).

Proof: \( A_{1t}(\cdot) = tr\left\{n_1[\Sigma_1^{-1}c_1 + (\Sigma_2^{-1}c_2)G_{1}^{-1}L_{1}^{-1}G_{1}] \right\} \]
\[ = n_1 tr[I_q - (\mathbf{D}D')^{-1}Z Z'(\mathbf{D}D')^{-1}]/N \]
where \( Z = \bar{\mu}_2 - \bar{\mu}_1 \).

Proof: \( A_{2t}(\cdot) = n_1 tr[I_q - (\mathbf{D}D')^{-1}Z Z'(\mathbf{D}D')^{-1}]/N \)
\[ = n_1 tr[I_q - (\mathbf{D}D')^{-1}Z Z'(\mathbf{D}D')^{-1}]/N \]

Note that the calculation of \( W \) does not involve finding \( \bar{\mu} \) or \( \bar{\Sigma} \) first.

For the determinant terms of \( AIC_{12} - AIC_1 \), consider the following:

**Lemma 2.4** \[ |\bar{\Sigma}_{12} - \Sigma_1| = |n_1I_q + n_2(\mathbf{D}D')^{-1}Z Z'/N \]
\[ + n_1n_2(\mathbf{D}D')^{-1}Z Z'/N^2 | \]

Proof: From (2.6) we have

\[ A_1(\cdot) = n_1 tr\left\{\Sigma_1^{-1}c_1 + n_1BZZ'B'/n_2 \right\} + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]

Now \( B'B = n_2^2(\mathbf{D}D')^{-1}/N \), therefore
\[ A_1(\cdot) = n_1 tr\left\{(n_2-n_1)BZZ'B'/n_2 \right\} + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]

The term \( Z'(D_{\Sigma_1}D_1 - \Sigma_1)^{-1}Z \) measures the differences in the means of the two groups, and the term \( tr(D_{\Sigma_1}D_1 - \Sigma_1)^{-1}Z Z' \) measures the differences in the covariance matrices of the two groups. \( W \) may be rewritten as
\[ W = n_2 tr[I_q - (\mathbf{D}D')^{-1}Z Z' - (\mathbf{D}D')^{-1}Z Z'/N \]

Note that the calculation of \( W \) does not involve finding \( \bar{\mu} \) or \( \bar{\Sigma} \) first.

For the determinant terms of \( AIC_{12} - AIC_1 \), consider the following:

**Lemma 2.4** \[ |\bar{\Sigma}_{12} - \Sigma_1| = |n_1I_q + n_2(\mathbf{D}D')^{-1}Z Z'/N \]
\[ + n_1n_2(\mathbf{D}D')^{-1}Z Z'/N^2 | \]

Proof: From (2.6) we have

\[ A_1(\cdot) = n_1 tr\left\{\Sigma_1^{-1}c_1 + n_1BZZ'B'/n_2 \right\} + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]

Now \( B'B = n_2^2(\mathbf{D}D')^{-1}/N \), therefore
\[ A_1(\cdot) = n_1 tr\left\{(n_2-n_1)BZZ'B'/n_2 \right\} + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]

The term \( Z'(D_{\Sigma_1}D_1 - \Sigma_1)^{-1}Z \) measures the differences in the means of the two groups, and the term \( tr(D_{\Sigma_1}D_1 - \Sigma_1)^{-1}Z Z' \) measures the differences in the covariance matrices of the two groups. \( W \) may be rewritten as
\[ W = n_2 tr[I_q - (\mathbf{D}D')^{-1}Z Z' - (\mathbf{D}D')^{-1}Z Z'/N \]

Note that the calculation of \( W \) does not involve finding \( \bar{\mu} \) or \( \bar{\Sigma} \) first.

For the determinant terms of \( AIC_{12} - AIC_1 \), consider the following:

**Lemma 2.4** \[ |\bar{\Sigma}_{12} - \Sigma_1| = |n_1I_q + n_2(\mathbf{D}D')^{-1}Z Z'/N \]
\[ + n_1n_2(\mathbf{D}D')^{-1}/N^2 | \]

Proof: From (2.6) we have

\[ A_1(\cdot) = n_1 tr\left\{\Sigma_1^{-1}c_1 + n_1BZZ'B'/n_2 \right\} + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]

Now \( B'B = n_2^2(\mathbf{D}D')^{-1}/N \), therefore
\[ A_1(\cdot) = n_1 tr\left\{(n_2-n_1)BZZ'B'/n_2 \right\} + (n_2-n_1)BZZ'B'/n_2 \]
\[ + (n_2-n_1)BZZ'B'/n_2 \]

The term \( Z'(D_{\Sigma_1}D_1 - \Sigma_1)^{-1}Z \) measures the differences in the means of the two groups, and the term \( tr(D_{\Sigma_1}D_1 - \Sigma_1)^{-1}Z Z' \) measures the differences in the covariance matrices of the two groups. \( W \) may be rewritten as
\[ W = n_2 tr[I_q - (\mathbf{D}D')^{-1}Z Z' - (\mathbf{D}D')^{-1}Z Z'/N \]

Note that the calculation of \( W \) does not involve finding \( \bar{\mu} \) or \( \bar{\Sigma} \) first.
Thus the statistic \( Q_2 \) is a likelihood ratio type statistic. Large values of \( \text{tr} \left( \mathbf{D}_2^{-1}(\mathbf{ZZ}')^2 \right) \) would indicate added Akaike information in the partial data group.

3. Canonical Correlations

Let \( x \) and \( y \) be vectors of \( p \) and \( q \) random variables, respectively, with \( p \leq q \). Then there exist (see Kshirsagar 1972) linear transformations from \( x = (x_1, \ldots, x_p)' \) and \( y = (y_1, \ldots, y_q)' \) to \( u = (u_1, \ldots, u_p)' \) and \( \hat{v} = (v_1, \ldots, v_q)' \) such that:

(a) \( u_1, u_2, \ldots, u_p, v_1, \ldots, v_q \) have unit variances, \( \text{sd} u_i = 1, i = 1, \ldots, p \).

(b) \( u_1, u_2, \ldots, u_p, v_1, \ldots, v_q \) have unit variances, \( \text{sd} v_i = 1, i = 1, \ldots, q \).

The \( p \) correlations \( \rho_{ij} = \text{corr}(u_i, v_j), i = 1, \ldots, p \) are known as the canonical correlations between \( x \) and \( y \).

Calculation of the canonical correlations is as follows:

Let the variance-covariance matrix of \( x \) and \( y \) be \( \Sigma \), partitioned as

\[
\Sigma = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{12}^T & \Sigma_{22}
\end{bmatrix}
\] (3.1)

where \( \Sigma_{11} \) is the variance-covariance matrix of \( x \), \( \Sigma_{12} \) is the variance-covariance matrix of \( x \), \( \Sigma_{12} \) is the covariance matrix between \( x \) and \( y \). Suppose the rank of \( \Sigma_{12} \) is \( r \leq p \). Then the canonical correlation \( \rho_j \) is the positive square root of the \( r \)th largest solution to the determinental equation

\[
\det \left( \mathbf{I} - \rho_j^2 \Sigma_{11} \right) = 0 \quad (3.2)
\]

The canonical variables \( u \) and \( v \) are found by letting \( u_i = l_i x \) (\( i = 1, \ldots, p \)) and \( v_j = m_j y \) (\( j = 1, \ldots, q \)), where the vectors \( l_i \) and \( m_j \) satisfy

\[
\begin{align*}
+ \rho_j^2 l_i m_j & = 0, i = 1, \ldots, r \\
- \rho_j^2 l_i m_j & = 0, j = r+1, \ldots, q
\end{align*}
\] (3.3)

The vectors \( l_i \) and \( m_j \) are known as the canonical vectors. The entire covariance structure between \( x \) and \( y \) has been reduced to the \( r \) canonical correlations \( \rho_1, \ldots, \rho_r \) and the canonical vectors.

Many topics in multivariate analysis can be expressed in terms of canonical analysis (see Kshirsagar 1972). For example, multivariate analysis of variance and covariance can be obtained by letting \( x \) be random variables and \( y \) be indicator variables. Discriminant analysis, and analysis of contingency tables can be obtained similarly.

If the population parameters are not known, a
Let $X$ and $Y$ be $pxN$ and $qxN$ matrices of observations from a random sample of size $N$ on $x$ and $y$, respectively. If $x$ and $y$ have a multivariate normal distribution, then the maximum likelihood estimates of $\Sigma_{11}$, $\Sigma_{22}$, and $\Sigma_{22}$ will be $\hat{\Sigma}_{11}$, $\hat{\Sigma}_{22}$, respectively, where

$$\begin{align*}
\hat{\Sigma}_{11} &= X(I - \frac{1}{N} E_{n} X') \\
\hat{\Sigma}_{12} &= X(I - \frac{1}{N} E_{n} Y') \\
\hat{\Sigma}_{22} &= Y(I - \frac{1}{N} E_{n} Y')
\end{align*}$$

Here, $E_{n}$ is an $nxm$ matrix of ones.

The sample canonical correlations, vectors and variables are then calculated by replacing $\Sigma_{11}$, $\Sigma_{22}$ and $\Sigma_{12}$ with $\hat{\Sigma}_{11}$, $\hat{\Sigma}_{22}$ and $\hat{\Sigma}_{12}$, respectively, in (3.2) and (3.3). The sample covariance matrix of $x$ and $y$ is positive definite with probability one, thus the rank of $\Sigma_{12}$ will be $p$.

There will then be $p$ non-zero sample canonical correlations, $r_1, r_2, \ldots, r_p$.

The joint distribution of the $p$ sample correlations has been derived under the following conditions:

(a) $Y$ is a fixed matrix,
(b) each column of $X$ has a $p$-variate normal distribution with common covariance matrix $\Sigma_{11}$.
(c) $E(X) = aE_{n} + bY$, where $a$ is a $px1$ vector and $b$ is a $pxq$ matrix, and
(d) $N > p + q + 1$.

The joint distribution of the $p$ sample correlations will be obtained under the following conditions:

- $Y$ is a fixed matrix,
- each column of $X$ has a $p$-variate normal distribution with common covariance matrix $\Sigma_{11}$,
- $E(X) = aE_{n} + bY$, where $a$ is a $px1$ vector and $b$ is a $pxq$ matrix, and
- $N > p + q + 1$.

Fisher (1939), Hau (1939) and Roy (1939) have independently derived the distribution of $r_1, \ldots, r_p$ with the above assumptions when $\beta = 0$, $m = p$, and $N$ is the null distribution. The non-null distribution, i.e., when $\beta \neq 0$, has been derived by Constantine (1963). Bartlett (1939) has derived a test for determining whether the last $p-s$ correlations are non-zero.

### 3.1 Incomplete Data

The techniques for estimation of canonical correlation coefficients will be extended to use multivariate vectors which are incompletely observed. Two estimation techniques which use partial data vectors will be given in this section and will be compared to the standard using only full vectors technique described above.

First, we apply the procedure developed by Hocking and Smith (1968) (hereafter denoted by HS) which modifies unbiased, full data estimates by adding weighted terms of zero expectation constructed from the set of sufficient statistics. Thus, correlation coefficients could be estimated from full data, then modified by adding weighted terms of zero expectation (e.g., the difference between two unbiased estimates of the first element of $\mu$, one being taken from full data only, the other the arithmetic mean of the corresponding element from partial records). The weights are chosen to minimize the variance of resulting estimates. Hocking and Smith (1968) show that these estimates are consistent and asymptotically efficient. Srivastava and Zaatari (1973) indicate that these estimates are not optimal for small samples and when no iteration takes place.

Hocking and Marx (1979) introduce an iterated and much improved version of the Hocking - Smith technique, and this procedure will also be applied to the estimation of canonical correlation coefficients in the presence of partial information. We choose to use the HS estimate ($\hat{E}$) of the covariance matrix $E$. The "$_c$" estimates will be substituted into equation (3.2), and the characteristic roots (canonical correlations) will be extracted.

To illustrate the properties of the three techniques several examples will be simulated.

**Example 3.1:** When $p = q = 1$, the canonical correlation coefficient is the linear correlation coefficient. Data are $n_1$ observations on $x$ and $y$, $n_2$ observations on $y$ only. **Example 3.2:** When $p = 1, q = 2$, the canonical correlation coefficient between $x$ and $y = (y_1, y_2)$ is the multiple correlation coefficient. Data are $n_1$ observations on $x, y_1, \ldots, y_p$ and $n_2$ observations on $y_1$ and $y_2$. **Example 3.3:** Same as Example 3.2, but where data are $n_1$ observations on $x, y_1, y_2$ and $n_2$ observations on $x$ and $y_1$.

Random samples from populations with the $(p+q)$-variate $N(0, I)$ distribution are simulated for $p+q = 3$ with

$$
\Sigma = \begin{bmatrix}
1 & a & \cdots & a \\
a & 1 & \cdots & a \\
\vdots & \vdots & \ddots & \vdots \\
a & a & \cdots & 1
\end{bmatrix}
$$

Chan and Dunn (1972) report that, in simulating discrimination problems, the results of using this form for $\Sigma$ and those of using randomly chosen correlation matrices were similar. In Riggs (1981), the value of $a$ is $0.4, 0.5, \ldots, 0.9$, producing squared population canonical correlations in each of the examples ranging from about 0.20 to 0.85, and in each example, samples of size $N = 50$ are simulated, with $n$, the number of complete observations, taking on the values of 40, 30, 20, 10. Then $n = N - n$, is the number of incomplete vectors, all having the same pattern. We partially report his results.

The first simulation is of Example 3.1. The population canonical correlation is $\rho = a$. These results are in Tables 1-2. The second and third simulations are Examples 3.2 and 3.3, respectively. The single population canonical correlation is $\rho = 2a^2/(1 + a)$. The results of the $n$ observations on $y_1$ and $y_2$ are in Tables 3-6. In each of these three examples, the three es-
estimates are calculated. Since these estimates of canonical correlations are biased, the estimated bias and the mean-squared error (MSE) are reported. The tables give the results of 1000 simulations of each example.

Let $p = q = 2$, and the $n_1$ partial observations were made on $x_1, x_2$ and $Y_i$. There are two population correlations: $\gamma_{11} = \delta a/(1 + a)^2$ and $\gamma_{12} = 0$. These results are in Tables 7 - 8. Since $p = 2$ and there are two population canonical correlations, the results given are the square root of the sum of the squared biases, denoted $|\rho - \rho'|_2$, and the determinant and trace of the MSE matrix, $E(\rho - \rho')'(\rho - \rho')$.

### Table 1

<table>
<thead>
<tr>
<th>Population</th>
<th>Full Data</th>
<th>All Data</th>
<th>All Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$\rho^2$</td>
<td>Est. Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>.4</td>
<td>.16</td>
<td>.01346</td>
<td>.01116</td>
</tr>
<tr>
<td>.5</td>
<td>.25</td>
<td>.00592</td>
<td>.01258</td>
</tr>
<tr>
<td>.6</td>
<td>.36</td>
<td>.00549</td>
<td>.01416</td>
</tr>
<tr>
<td>.7</td>
<td>.49</td>
<td>-.00303</td>
<td>.01393</td>
</tr>
<tr>
<td>.8</td>
<td>.64</td>
<td>-.00154</td>
<td>.00866</td>
</tr>
<tr>
<td>.9</td>
<td>.81</td>
<td>-.00333</td>
<td>.00316</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Population</th>
<th>Full Data</th>
<th>All Data</th>
<th>All Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$\rho^2$</td>
<td>Est. Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>.4</td>
<td>.16</td>
<td>.02788</td>
<td>.02118</td>
</tr>
<tr>
<td>.5</td>
<td>.25</td>
<td>.01780</td>
<td>.02372</td>
</tr>
<tr>
<td>.6</td>
<td>.36</td>
<td>.01451</td>
<td>.02944</td>
</tr>
<tr>
<td>.7</td>
<td>.49</td>
<td>.00338</td>
<td>.02378</td>
</tr>
<tr>
<td>.8</td>
<td>.64</td>
<td>.00261</td>
<td>.01661</td>
</tr>
<tr>
<td>.9</td>
<td>.81</td>
<td>-.00180</td>
<td>.00519</td>
</tr>
</tbody>
</table>
### Table 3
Simulation of Example 3.2  
\( p = q = 1 \)  
\( n_1 = 40, n_2 = 10 \)

<table>
<thead>
<tr>
<th>( a )</th>
<th>( p^2 )</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>.4</td>
<td>.22857</td>
<td>.03661</td>
<td>.01344</td>
<td>.03860</td>
<td>.01358</td>
<td>.03799</td>
<td>.01344</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.33333</td>
<td>.02138</td>
<td>.01428</td>
<td>.02287</td>
<td>.01415</td>
<td>.02199</td>
<td>.01401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>.45000</td>
<td>.01654</td>
<td>.01256</td>
<td>.01956</td>
<td>.01210</td>
<td>.01836</td>
<td>.01196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>.57647</td>
<td>.01540</td>
<td>.00953</td>
<td>.01858</td>
<td>.00930</td>
<td>.01710</td>
<td>.00919</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>.71111</td>
<td>.00273</td>
<td>.00609</td>
<td>.00554</td>
<td>.00566</td>
<td>.00402</td>
<td>.00564</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.9</td>
<td>.85263</td>
<td>.00195</td>
<td>.00181</td>
<td>.00463</td>
<td>.00156</td>
<td>.00943</td>
<td>.00157</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4
Simulation of Example 3.2  
\( p = 1, q = 2 \)  
\( n_1 = 10, n_2 = 40 \)

<table>
<thead>
<tr>
<th>( a )</th>
<th>( p^2 )</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>.4</td>
<td>.22857</td>
<td>.13809</td>
<td>.06213</td>
<td>.17792</td>
<td>.08474</td>
<td>.15963</td>
<td>.07027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.33333</td>
<td>.09609</td>
<td>.05504</td>
<td>.13632</td>
<td>.07725</td>
<td>.11243</td>
<td>.05884</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>.45000</td>
<td>.07426</td>
<td>.04783</td>
<td>.12120</td>
<td>.06399</td>
<td>.09454</td>
<td>.04771</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>.57647</td>
<td>.05682</td>
<td>.03604</td>
<td>.10256</td>
<td>.04428</td>
<td>.07644</td>
<td>.03392</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>.71111</td>
<td>.01865</td>
<td>.02458</td>
<td>.06580</td>
<td>.02770</td>
<td>.03690</td>
<td>.01935</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.9</td>
<td>.85263</td>
<td>.00911</td>
<td>.00818</td>
<td>.05003</td>
<td>.01099</td>
<td>.02510</td>
<td>.00471</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5
Simulation of Example 3.3  
\( p = 1, q = 2 \)  
\( n_1 = 40, n_2 = 10 \)

<table>
<thead>
<tr>
<th>( a )</th>
<th>( p^2 )</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
<th>( \text{Est.} )</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>.4</td>
<td>.22857</td>
<td>.03174</td>
<td>.01404</td>
<td>.02644</td>
<td>.01249</td>
<td>.02936</td>
<td>.01232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.33333</td>
<td>.02514</td>
<td>.01359</td>
<td>.02335</td>
<td>.01181</td>
<td>.02477</td>
<td>.01170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>.45000</td>
<td>.01273</td>
<td>.01344</td>
<td>.01287</td>
<td>.01185</td>
<td>.01337</td>
<td>.01170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>.57647</td>
<td>.00575</td>
<td>.01030</td>
<td>.00962</td>
<td>.00903</td>
<td>.00939</td>
<td>.00889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>.71111</td>
<td>.00557</td>
<td>.00554</td>
<td>.00595</td>
<td>.00488</td>
<td>.00553</td>
<td>.00484</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.9</td>
<td>.85263</td>
<td>.00115</td>
<td>.00196</td>
<td>.00212</td>
<td>.00166</td>
<td>.00181</td>
<td>.00166</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6
Simulation of Example 3.3
p = 1, q = 2
n₁ = 10, n₂ = 40

<table>
<thead>
<tr>
<th>a</th>
<th>p²</th>
<th>Full Data</th>
<th></th>
<th></th>
<th>All Data</th>
<th></th>
<th></th>
<th>All Data</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Est. Bias</td>
<td>MSE</td>
<td></td>
<td>Est. Bias</td>
<td>MSE</td>
<td></td>
<td>Est. Bias</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>.4</td>
<td>.22857</td>
<td>.13567</td>
<td>.06176</td>
<td>.04660</td>
<td>.06338</td>
<td>.10519</td>
<td>.05072</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.33333</td>
<td>.10479</td>
<td>.05751</td>
<td>.03036</td>
<td>.06512</td>
<td>.07875</td>
<td>.03412</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>.45000</td>
<td>.07241</td>
<td>.04747</td>
<td>.04040</td>
<td>.04407</td>
<td>.06480</td>
<td>.02698</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>.57647</td>
<td>.03946</td>
<td>.03682</td>
<td>.04230</td>
<td>.03076</td>
<td>.04723</td>
<td>.02038</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>.71111</td>
<td>.02414</td>
<td>.02437</td>
<td>.03325</td>
<td>.02425</td>
<td>.03154</td>
<td>.01044</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.9</td>
<td>.85263</td>
<td>.00487</td>
<td>.00923</td>
<td>.02800</td>
<td>.01027</td>
<td>.01470</td>
<td>.00297</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7
Simulation of Example 3.4
p = 2, q = 2
n₁ = 40, n₂ = 10

<table>
<thead>
<tr>
<th>a</th>
<th>p²</th>
<th>Full Data</th>
<th></th>
<th></th>
<th>All Data</th>
<th></th>
<th></th>
<th>All Data</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Est. Bias</td>
<td>Det. MSEx10³</td>
<td>Trace</td>
<td>Est. Bias</td>
<td>Det. MSEx10³</td>
<td>Trace</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.4</td>
<td>.32653</td>
<td>.04883</td>
<td>.02471</td>
<td>.01650</td>
<td>.04477</td>
<td>.01906</td>
<td>.01483</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.44444</td>
<td>.04347</td>
<td>.02592</td>
<td>.01564</td>
<td>.04116</td>
<td>.02160</td>
<td>.01415</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>.56250</td>
<td>.03463</td>
<td>.01757</td>
<td>.01216</td>
<td>.03301</td>
<td>.01267</td>
<td>.01046</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>.67820</td>
<td>.02898</td>
<td>.01202</td>
<td>.00846</td>
<td>.02678</td>
<td>.01032</td>
<td>.00776</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>.79012</td>
<td>.02811</td>
<td>.00787</td>
<td>.00583</td>
<td>.02691</td>
<td>.00567</td>
<td>.00503</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.9</td>
<td>.89751</td>
<td>.02764</td>
<td>.00196</td>
<td>.00304</td>
<td>.02598</td>
<td>.00157</td>
<td>.00220</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8
Simulation of Example 3.4
p = 2, q = 2
n₁ = 20, n₂ = 30

<table>
<thead>
<tr>
<th>a</th>
<th>p²</th>
<th>Full Data</th>
<th></th>
<th></th>
<th>All Data</th>
<th></th>
<th></th>
<th>All Data</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Est. Bias</td>
<td>Det. MSEx10³</td>
<td>Trace</td>
<td>Est. Bias</td>
<td>Det. MSEx10³</td>
<td>Trace</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.4</td>
<td>.32653</td>
<td>.10331</td>
<td>.16013</td>
<td>.03897</td>
<td>.08391</td>
<td>.09417</td>
<td>.02835</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.44444</td>
<td>.07997</td>
<td>.15578</td>
<td>.03391</td>
<td>.06511</td>
<td>.08815</td>
<td>.02444</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>.56250</td>
<td>.06436</td>
<td>.11410</td>
<td>.02544</td>
<td>.05484</td>
<td>.06114</td>
<td>.01779</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>.67820</td>
<td>.05895</td>
<td>.08405</td>
<td>.02067</td>
<td>.05020</td>
<td>.04097</td>
<td>.01373</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>.79012</td>
<td>.05327</td>
<td>.04995</td>
<td>.01428</td>
<td>.04692</td>
<td>.02382</td>
<td>.00988</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.9</td>
<td>.89751</td>
<td>.05384</td>
<td>.01465</td>
<td>.00963</td>
<td>.04662</td>
<td>.00659</td>
<td>.00697</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The four examples of canonical correlation estimation simulated can be divided into two types. In the first two examples, the incomplete observation vectors contain information on only \( x \) or only \( y \). In the last two examples, the incomplete observation vectors contain information on all of \( x \) and on some of the variables in \( y \). In the first situation, the incomplete vectors contribute only to estimating the variance-covariance matrix of the represented vector variable. In the second situation, the incomplete vectors contribute to estimating \( \Sigma_{11} \) and also to a portion of \( \Sigma_{12} \).

In the first three examples, the MSE of HM is generally lower than that of HS, thus in the fourth example, only the HM estimates are computed. In examples 1 and 2, the HM estimates have lower MSE's than the full data estimates. The only exceptions are when the population correlation \( \rho \) is below about .25. For these simulations, the full data estimates have lower MSE's. When the true correlation between \( x \) and \( y \) is low, information on \( x \) or \( y \) alone is of little use in estimating that correlation.

In examples 3 and 4, where the partial information is on both \( x \) and \( y \), the HM estimates always have lower MSE's than the full data estimates. The partial information was useful even for low \( \rho \).

In all cases, as the percentage of incomplete vectors increases, the improvement in the MSE of the HM estimates over that of the full data estimates increases, especially for high \( \rho \).

Although the focus of this study is the variances or the MSE's of the estimators, it is of interest to look at the bias of each. In cases 1 and 2, the HM estimates generally have higher biases than the full data estimates. This is also true in case 3, except for lower values of \( \rho \). The bias of the HM estimates is slightly lower for the low \( \rho \). And in the case 4, the HM estimates generally have much lower biases than the full data only estimates.

In practically all situations, therefore, the estimates found by estimating \( \Sigma \) with the HM method, then using that estimate to calculate the HM have smaller MSE's than the estimates based on the complete data vectors only.

4. Summary
This paper gives a method for testing the significance of the information in partial records. The test will aid the user in deciding whether a rather complicated estimation procedure should be calculated. Additionally, we provide guidance in applying canonical analysis to multivariate data with missing elements.

3. Acknowledgements
The authors gratefully acknowledge financial support from the Center for Energy and Mineral Resources, Texas A&M University. Additionally, a portion of this research was supported by the Japanese Society for the Promotion of Science through a Visiting Scientist Fellowship which was awarded to the senior author.

REFERENCES


