ASYMPTOTIC THEORY FOR COMMON PRINCIPAL COMPONENT ANALYSIS

BY

BERNHARD N. FLURY

TECHNICAL REPORT NO. 204
NOVEMBER 1984

PREPARED UNDER THE AUSPICIES
OF
NATIONAL SCIENCE FOUNDATION GRANT
2 FFZ 467
INGRAM OLKIN, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ASYMPTOTIC THEORY FOR COMMON PRINCIPAL COMPONENT ANALYSIS

By

Bernhard N. Flury*

Indiana University
Department of Mathematics
Bloomington, IN 47405

ABSTRACT

Under the Common Principal Component model \( k \) covariance matrices
\[ \Sigma_1, \ldots, \Sigma_k \] are simultaneously diagonalizable, i.e. there exists an
orthogonal matrix \( \beta \) such that \( \beta' \Sigma_i \beta = \Lambda_i \) is diagonal for \( i = 1, \ldots, k \).
In this article we give the asymptotic distribution of the maximum likelihood
estimates of \( \beta \) and \( \Lambda_i \). Using these results, we derive tests for (a)
equality of eigenvectors with a given set of orthonormal vectors, and
(b) redundancy of \( p-q \) (out of \( p \)) principal components. The likelihood-ratio test for simultaneous sphericity of \( p-q \) principal components in
\( k \) populations is derived, and some of the results are illustrated by a
biometrical example.

Keywords: maximum likelihood; covariance matrices; eigenvectors; eigenvalues.

* This work was done under Contract 882.008.0.82 of the Swiss National
Science Foundation at Stanford University. I wish to thank Professors
Ingram Olkin and T. W. Anderson.
1. INTRODUCTION

Common Principal Component Analysis (CPCA) is a generalization of Principal Component Analysis (PCA) to $k$ groups (Flury 1984). The key assumption is that the $p \times p$ covariance matrices $\Sigma_1, \ldots, \Sigma_k$ of $k$ populations can be diagonalized by the same orthogonal transformation, i.e. there exists an orthogonal matrix $\Phi$ such that

$$H_c : \Phi' \Sigma_i \Phi = \Lambda_i \text{ (diagonal)}; \quad i = 1, \ldots, k \quad (1.1)$$

holds. $H_c$ is called the hypothesis of common principal components (CPC's). Flury (1984) derives the normal theory maximum likelihood estimates of $\Phi$ and $\Lambda_i$ and gives numerical examples.

In the one sample case $k = 1$, CPC's reduce to ordinary principal components (PC's). In this case the ML-estimates of $\Phi$ and $\Lambda = \Lambda_1$ are the eigenvectors and eigenvalues of a Wishart matrix $S_1$. The asymptotic distribution theory for this situation has been developed by Girshick (1939), Lawley (1953, 1956) and Anderson (1963). The present paper gives essentially generalizations of results obtained by Anderson.

In one-group PCA, the eigenvectors $\Phi_i$ forming the orthogonal matrix $\Phi = (\Phi_1, \ldots, \Phi_p)$ are usually ordered according to the associated eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_p$. In CPCA no obvious fixed order of the columns of $\Phi$ need be given, since the rank order of the diagonal elements of the $\Lambda_i$ is not necessarily the same for all $\Lambda_i$. However, we can use some convention, e.g. that the columns of $\Phi$ be arranged according to the first group---such that $\Phi_1' \Sigma_1 \Phi_1 > \Phi_2' \Sigma_1 \Phi_2 > \ldots > \Phi_p' \Sigma \Phi_p$ (assuming that the $p$ characteristic
roots of $\xi_1$ are all distinct). This will enable us to speak about "first", "second" or "last" principal components also in the k-group case.

Tests for various hypotheses about $\xi$ and $\lambda$ in PCA have been proposed by Anderson (1963). We are going to construct similar tests for CPC. More specifically, we will treat the following problems:

1. Is the j-th eigenvector $\xi_j$ identical with a given (normalized) vector $\xi_j^0$? More generally, for q different eigenvectors $\xi_j, \xi_{j+1}, \ldots, \xi_{j+q-1}$, are they identical with q given (orthonormal) eigenvectors $\xi_j^0, \ldots, \xi_{j+q-1}^0$? This problem will be treated in Section 3.

2. As an associate editor handling the previous paper has pointed out, the most useful applications of CPC would probably be those in which some relatively small number q of rotated axes are sufficient to recover most of the variability in each of the k groups. It would therefore be helpful to have confidence intervals for sums of estimated eigenvalues and criteria for neglecting CPC's with small contributions. Solutions to these problems are given in Sections 4.1 and 4.2.

3. When PC's are interpreted, it is important to make sure that the roots $\lambda_j$ and $\lambda_h$ (say) are not identical, because otherwise the associated eigenvectors $\xi_j$ and $\xi_h$ are not uniquely defined. Similarly in CPC two eigenvectors $\xi_j$ and $\xi_h$ are uniquely defined if in at least one population the two associated eigenvalues are not identical. A likelihood ratio test dealing with this problem is given in Section 4.3.

We will from now on always assume that the matrices $\xi_1, \ldots, \xi_k$ are positive definite symmetric (p.d.s.). The diagonal elements of $\xi_j$ will be denoted
by \( \lambda_{ij} \), i.e. \( \Lambda = \text{diag}(\lambda_{i1}, \ldots, \lambda_{ip}) \) \((i = 1, \ldots, k)\). All inference will be based on \( k \) independent sample covariance matrices \( S_i \) with \( n_i \) degrees of freedom respectively, such that \( n_i S_i \) has the Wishart distribution \( W_p(n_i, \Sigma_i) \). The ML-estimates of \( \beta = (\beta_1, \ldots, \beta_p) \) and \( \Lambda_i \) are denoted by \( \hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p) \) and \( \hat{\Lambda}_i = \text{diag}(\hat{\lambda}_{i1}, \ldots, \hat{\lambda}_{ip}) \).

2. ASYMPTOTIC DISTRIBUTION OF THE MAXIMUM LIKELIHOOD ESTIMATES

In this section we are using general properties of ML-estimates under regularity conditions; see e.g. Silvey (1975, chapters 4 and 7), Wilks (1944, chapter 6). In particular we will use the fact that the joint asymptotic distribution of the parameter estimates is multivariate normal, the covariance matrix being given by the inverse of the Fisher information matrix. Th log-likelihood function of the \( k \) samples, up to an additive constant, is given by

\[
g(\Lambda_1, \ldots, \Lambda_k, \beta | S_1, \ldots, S_k) = -\frac{1}{2} \sum_{i=1}^{k} n_i \left( \frac{p}{2} \log \lambda_{ii} + \beta_i S_i \hat{\beta}_i / \lambda_{ii} \right) \tag{2.1} \]

(Flury 1984, formula 2.5.) Assume that the \( \beta_j \) are well defined, i.e. for each pair \((j, \lambda)\) there is at least one \( i \in \{1, \ldots, k\} \) such that \( \lambda_{ij} \neq \lambda_{i\lambda} \). Let \( \lambda^{(i)} = (\lambda_{i1}, \ldots, \lambda_{ip}) \), \( s = p(p-1)/2 \), and denote by \( \bar{\beta}^* \) a vector composed of \( s \) functionally independent elements of \( \bar{\beta} \). Put \( n = n_1 + \ldots + n_k \) and \( r_i = n_i / n \) \((i = 1, \ldots, k)\). Then the information matrix is
\[
\begin{array}{c|cccc|c}
\lambda' & \lambda'(1) & \lambda'(2) & \cdots & \lambda'(k) & \beta^* \\
\hline
\lambda(1) & \frac{n}{2} r_1 \lambda_1^{-2} & 0 & \cdots & 0 & \\
\lambda(2) & 0 & \frac{n}{2} r_2 \lambda_2^{-2} & \cdots & 0 & \\
\vdots & \vdots & \vdots & \ddots & \vdots & \\
\lambda(k) & 0 & 0 & \cdots & \frac{n}{2} r_k \lambda_k^{-2} & \\
\hline
\beta^* & \mathbb{G} & \\
\end{array}
\]

where \( \mathbb{A} \) and \( \mathbb{G} \) are not yet determined.

Since \( \hat{\lambda}_{ij} = \hat{\beta}_j s_i \hat{\lambda}_i \) and \( \hat{\beta}_j \) is a consistent estimate of \( \beta_j \), we can use the asymptotic \((n_i \to \infty)\) normality of \( n_i s_i \) (Muirhead 1982, p. 19) to get the asymptotic univariate distribution of \( \hat{\lambda}_{ij} \) as

\[
\sqrt{n_i} (\hat{\lambda}_{ij} - \lambda_{ij}) \sim N(0, \lambda_{ij}^2).
\] (2.3)

From (2.2), the joint asymptotic distribution of \( (\hat{\lambda}'(1), \cdots, \hat{\lambda}'(k))' \) has covariance matrix

\[
\frac{1}{n} \mathbb{V}_\lambda = \left( \begin{array}{cccc}
\frac{n}{2} r_1 \lambda_1^{-2} & \cdots & 0 \\
\vdots & \ddots & \vdots & \\
0 & \cdots & \frac{n}{2} r_k \lambda_k^{-2} \\
\end{array} \right) - \mathbb{G}^T \mathbb{A}^{-1} \mathbb{G}^{-1}.
\] (2.4)

Since, by (2.3), the diagonal elements of \( \mathbb{V}_\lambda \) as \( \frac{2}{r_1} \lambda_{11}, \frac{2}{r_1} \lambda_{12}, \cdots, \frac{2}{r_k} \lambda_{kp} \), and \( \mathbb{A} \) is p.d.s., it follows that \( \mathbb{G} = 0 \). Thus we get
Theorem 1: The statistics \( \sqrt{n_i} (\hat{\lambda}_{ij} - \lambda_{ij}) \) are asymptotically
\[
\left( \min_{1 \leq i \leq k} n_i \to \infty \right) \text{ distributed as } N(0,2\lambda_{ij}^2),
\]
independent of each other and independent of the \( \hat{\beta}_j \).

The asymptotic distribution of \( \hat{\beta} \) requires more work. First, from the log-likelihood function (2.1) it is clear that the matrix \( \Lambda \) can be written as the sum of \( k \) matrices \( A_1, \ldots, A_k \), where \( A_i \) is associated with the \( i \)-th sample. In fact, \( A_i \) is exactly the same as if there were only one sample (namely the \( i \)-th one). In the one sample case, the eigenvectors of \( S_1 \) are ML-estimates for \( \beta \), and the asymptotic covariance matrix of the \( j \)-th eigenvector \( b_{ij} \) of \( S_1 \) can be written as (Anderson 1963, p. 130)
\[
\frac{1}{n_i} \Sigma_{i,j} = \frac{1}{n_i} \frac{\lambda_{ij}}{\lambda_{ih}} \sum_{h=1}^{p} \frac{\lambda_{ih}}{(\lambda_{ih} - \lambda_{ij})^2} \beta_h \beta_h', \quad (2.5)
\]
which is positive semidefinite of rank \( p-1 \).

One might now be tempted to attack the \( k \)-sample problem by defining
\[
g_{ij}(i) = \frac{1}{n_i} \frac{\lambda_{ij} \lambda_{ih}}{(\lambda_{ih} - \lambda_{ij})^2} \quad (h \neq j), \quad (2.6)
\]
Taking the Moore-Penrose inverse of \( \frac{1}{n_i} \Sigma_{i,j} \), adding these \( k \) "information matrices" over all \( k \) groups and taking again the Moore-Penrose inverse of the sum to get the asymptotic covariance matrix of \( \beta_j \) as
\[
\left( \sum_{i=1}^{k} n_i \Sigma_{i,j} \right)^+ = \left( n \sum_{i=1}^{k} \frac{1}{n} \sum_{h=1}^{p} g_{ij}(i)-1 \beta_h \beta_h' \right)^+
\]
\[
= \frac{1}{n} \sum_{h=1}^{p} \left( \sum_{i=1}^{k} g_{ij}(i)-1 \right) \beta_h \beta_h'. \quad (2.7)
\]
This "proof" is of course not correct, primarily because the eigenvectors of $S_i$ are (even asymptotically) not independent, and the covariance matrix of $\hat{\beta}_j$ can therefore not be obtained by inverting just the sub-matrix of the information matrix associated with $\beta_j$.

Surprisingly, the result (2.7) is nevertheless true, but its proof requires some more care. The key idea is to transform the information matrices $A_i$ to diagonal form. Denote by $V_i$ the asymptotic covariance matrix of $\sqrt{n_i} \text{vec} \hat{\beta} = \sqrt{n_i}(\hat{\beta}_1', \ldots, \hat{\beta}_p')'$ as obtained from the $i$-th sample alone. Following Anderson (1963, p. 130), we get

<table>
<thead>
<tr>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\beta}_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_1$</td>
<td>$\sum_{h=1}^{p} g_{1h} \hat{\beta}_h \hat{\beta}_1'$</td>
<td>$-g_{12} \hat{\beta}_2 \hat{\beta}_1'$</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>$-g_{21} \hat{\beta}_1 \hat{\beta}_2'$</td>
<td>$\sum_{h=1}^{p} g_{2h} \hat{\beta}_h \hat{\beta}_2'$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$\hat{\beta}_p$</td>
<td>$-g_{pi} \hat{\beta}_i \hat{\beta}_p'$</td>
<td>$-g_{pi} \hat{\beta}_i \hat{\beta}_p'$</td>
</tr>
</tbody>
</table>

Following Anderson (1963, p. 130), we get

$$V_i = \begin{pmatrix} V_{i1} & \cdots & V_{ip} \\ \vdots & \ddots & \vdots \\ V_{pi} & \cdots & V_{pp} \end{pmatrix}$$

The matrices $V_i$ have a remarkable property, namely

**Lemma 1:** The matrices $V_i$ ($i = 1, \ldots, k$) as given by (2.8) are simultaneously diagonalizable.

In other words, the property of having identical eigenvectors is inherited by the $V_i$ from the $S_i$ -- an intriguing fact! Lemma 1 is easily proved.
by showing that \( V_h V_i = V_i V_h \) for all pairs \((i, h)\), using the equivalence of simultaneous diagonalizability and commutativity under multiplication.

Thus there exists an orthogonal \( p^2 \times p^2 \)-matrix \( H = (H_1, H_2) \), where \( H_1 = (h_1, \ldots, h_s) \), \( s = p(p-1)/2 \), such that

\[
H'V_i H' = \begin{pmatrix} E_i & 0 \\ 0 & 0 \end{pmatrix} \quad (i = 1, \ldots, k) \tag{2.9}
\]

where \( E_i = \text{diag}(e_{i1}, \ldots, e_{is}) \), \( e_{ij} > 0 \). \( E_i \) has rank \( s \) because there are \( s \) functionally independent elements in \( \hat{g} \). For the transformed variables \( y = H_1' \text{vec} \hat{\beta} \) the information from the \( i \)-th sample is therefore

\[
A_i^* = n \text{diag}(e_{i1}^{-1}, e_{i2}^{-1}, \ldots, e_{is}^{-1})
\]

with

\[
e_{ij} = h_j' V_i h_j.
\]

The sum of these \( k \) information matrices is

\[
A^* = \sum_{i=1}^{k} A_i^* = n \text{diag} \left( \sum_{i=1}^{k} e_{i1}^{-1}, \ldots, \sum_{i=1}^{k} e_{is}^{-1} \right)
\]

\[
= n \text{diag}(e_1^{-1}, \ldots, e_s^{-1}) \tag{2.11}
\]

where \( e_j \) is the harmonic mean of \( e_{j1}, \ldots, e_{kj} \). The asymptotic covariance matrix of \( \hat{y} \) is therefore \( \text{diag}(e_1, \ldots, e_s)/n \). Transforming back to \( \hat{\beta} \), we get the asymptotic covariance matrix of \( \text{vec} \hat{\beta} \) as

\[
\frac{1}{n} V = \sum_{j=1}^{s} e_j h_j h_j'.
\]

\[
\tag{2.12}
\]
To establish the final result, we need now the explicit form of $H_1$. The $h_j$ are vectors of dimension $p^2$. If $\eta = (\eta'_1, \ldots, \eta'_p)'$ is a $p^2$-vector partitioned into $p$ vectors of dimension $p$, we will, for simplicity, refer to $\eta_j$ as the $j$-th position of $\eta$ (which corresponds to the scalar positions $(j-1)p+1$ thru $jp$).

**Lemma 2**: The (normalized characteristic vectors of $V_1$) are as follows:

1. $s = (s_1, \ldots, s_p)$ characteristic vectors (one for each pair of indices $j < k$) have $\beta_j/\sqrt{2}$ in position $j$ and $-\beta_j/\sqrt{2}$ in position $k$. All other positions are zero, and the associated eigenvalues are $2g_j^{(1)}$.

2. $s$ characteristic vectors (one for each pair of indices $j < k$) can be chosen to have $\beta_j/\sqrt{2}$ in position $j$ and $\beta_j/\sqrt{2}$ in position $k$. All other positions are zero, and the associated eigenvalues are zero.

3. $p$ characteristic vectors (one for each index $j$) can be chosen to have $\beta_j$ in position $j$ and zero elsewhere. The associated eigenvalues are zero.

The proof of Lemma 2 is straightforward and need not be given. Of course the eigenvectors in (2) and (3) are not uniquely defined, but the form given in the lemma is probably the simplest one. The eigenvectors in (1) form the matrix $H_1$.

Assuming that all $k$ matrices $\Sigma_i$ have $p$ distinct eigenvalues and that $\rho_i > 0$, the $g_j^{(i)}$ are all positive. From (2.10) it is now seen that $e_{jm} = g_j^{(i)}$ for some pair $(j, k)$. Thus, putting
\[ g_{j\ell} = \left( \sum_{i=1}^{k} g_{j\ell}^{(i)} \right)_{-1}^{-1} \]  

and writing \( h_{j\ell} \) for the eigenvector associated with the roots \( 2g_{j\ell}^{(i)} \), we get the asymptotic covariance matrix of \( \text{vec} \hat{\beta} \) as

\[ \Sigma = 2 \sum_{j \neq \ell} g_{j\ell} h_{j\ell} h_{j\ell}' . \]  

Writing this in terms of the \( \beta \)-vectors, using Lemma 2, we get therefore

**Theorem 2:** The asymptotic distribution of \( \sqrt{n} \text{vec}(\hat{\beta} - \beta) \) is normal with mean 0 and covariance matrix \( \Sigma \) given by

\[
\begin{array}{ccc}
\hat{\beta}_1' & \hat{\beta}_2' & \cdots & \hat{\beta}_p' \\
\sum_{h=1}^{P} g_{1h} \beta_{1h} & -g_{12} \beta_{12}' & \cdots & -g_{1p} \beta_{1p}' \\
-\Sigma_{21} & \sum_{h=1}^{P} g_{2h} \beta_{2h}' & \cdots & -g_{2p} \beta_{2p}' \\
\vdots & \vdots & \ddots & \vdots \\
-\Sigma_{p1} & -\Sigma_{p2} & \cdots & \sum_{h=1}^{P} g_{ph} \beta_{ph}' \\
\end{array}
\]  

where the \( g_{j\ell} \) are defined in (2.13), and the \( \beta_j \) are the (common) eigenvectors of the \( k \) matrices \( \Sigma_j \).

This confirms Assertion (2.7).
The following two points are worth noting:

1. If \( k = 1 \), Theorem 2 gives the asymptotic distribution of the characteristic vectors of a single Wishart matrix \( S_1 \). In fact, the similarity between (2.15) and the one-sample result (2.8) is striking - \( V \) can be obtained from \( V_1 \) by simply replacing the coefficients \( g_{ij}^{(1)} \) by their harmonic mean over all \( k \) groups.

2. The above proof of theorem 2 is based on the assumption that all \( k \) matrices \( S_1 \) have \( p \) distinct eigenvalues. However, since
\[
g_{ij}^{(1)-1} = r_1^2 \frac{1}{\lambda_{ij}} = 1 / \lambda_{ij},
\]
we can take \( g_{ij}^{(1)-1} = 0 \) if \( \lambda_{ij} = \lambda_{il} \).
In order for \( g_{ij} \) (2.13) to be defined, it suffices to have at least one \( S_1 \) with \( \lambda_{ij} \neq \lambda_{il} \). It is therefore assumed that Theorem 2 holds whenever CPC's are well defined.

3. AN ASYMPTOTIC TEST FOR \( q \) HYPOTHETICAL EIGENVECTORS

Using the asymptotic distribution theory, Anderson (1963, Appendix B) constructs a test for the hypothesis that the \( j \)-th eigenvector of \( S \) is identical with a specified vector \( \beta_j^0 \) (\( \beta_j^0 \beta_j^0 = 1 \)), under the assumption that this eigenvector corresponds to a root of multiplicity 1. In this section we are going to generalize Anderson's result in two ways by deriving an analogous test for \( q \) specified vectors (\( 1 \leq q \leq p \)) and \( k \) groups.
Without loss of generality we can order the CPC's such that the \( q \) vectors to be tested have the numbers 1 through \( q \). The null hypothesis is

\[
H_q : (\beta_1, \ldots, \beta_q) = (\beta_1^0, \ldots, \beta_q^0),
\]

where the \( \beta_j^0 \) are specified, mutually orthogonal and have unit length.
The test of \( H_q \) will be based on the asymptotic covariance matrix 
\[ \text{vec}(\hat{\beta}_1, \ldots, \hat{\beta}_q) \], that is, the upper left \( pq \times pq \) portion of \( \Psi \).
Call this submatrix \( \Psi(q) \). The eigenstructure of \( \Psi(q) \) is given by the 
following theorem, which is actually a generalization of Lemma 2. We are 
using the convention that the \( p \) scalar elements in positions \((j-1)p+1 \) 
through \( jp \) of a vector are referred to as \( j \)-th position.

**Theorem 3:** The upper left \( pq \times pq \) submatrix of \( \Psi \) has the following 
eigenvalues and eigenvalues:

1. \( \binom{q}{2} \) eigenvectors (one for each pair \( j, \ell \) with \( 1 \leq j < \ell \leq q \))
   have \( \hat{\beta}_j / \sqrt{2} \) in position \( j \) and \( -\hat{\beta}_\ell / \sqrt{2} \) in position \( \ell \). All 
   other positions are zero, and the associated roots are \( 2g_{j\ell} \).

2. \( (p-q)q \) eigenvectors (one for each combination of indices \( j, \ell \) 
such that \( 1 \leq j < q < \ell \leq p \)) have \( \hat{\beta}_j \) in position \( j \) and \( 0 \) 
in all other positions; the associated roots are \( g_{j\ell} \).

3. \( \binom{q}{2} \) eigenvectors (one for each pair of indices \( j, \ell \) such 
that \( 1 \leq j < \ell \leq q \)) can be chosen to have \( \hat{\beta}_j / \sqrt{2} \) in position 
\( j \), \( \hat{\beta}_\ell / \sqrt{2} \) in position \( \ell \) and zeros in all other positions. The 
associated roots are zero.

4. \( q \) eigenvectors (one for each \( j \) with \( 1 \leq j \leq q \)) can be 
chosen to have \( \hat{\beta}_j \) in position \( j \) and zeros elsewhere. The 
associated roots are zero.

The proof of Theorem 3 is easy and is therefore omitted. We see that 
if \( g_{j\ell} > 0 \) for \( 1 \leq j < q, 1 \leq \ell \leq p \), then \( \Psi(q) \) has rank 
\( t = \binom{q}{2} + q(p-q) \).
Let now $\phi$ denote a $t \times t$ diagonal matrix with diagonal elements equal to the nonzero roots of $V(q)$, i.e. $\phi = \text{diag}(\sqrt[2]{q_1}, \ldots, \sqrt[2]{q_p}, \sqrt[2]{q_{p+1}}, \ldots, \sqrt[2]{q_t})$, and let the columns of the $pq \times t$-matrix $\Gamma$ be given by the characteristic vectors associated with the nonzero roots. Putting $b_q = \text{vec}(\hat{\beta}_1, \ldots, \hat{\beta}_q) - \text{vec}(\beta_1, \ldots, \beta_q)$, the random vector $z = \sqrt{n} \phi^{-1/2} \Gamma' b_q$ has a limiting normal distribution with mean zero and covariance matrix $I_t$.

Thus $z'z' = \text{vec}(\Gamma' \phi^{-1} \Gamma b_q)' \text{vec}(\Gamma' \phi^{-1} \Gamma b_q) = \Gamma' \phi^{-1} \Gamma' b_q b_q'$ has a limiting chi square distribution with $t$ degrees of freedom. The matrix $\Gamma' \phi^{-1} \Gamma'$ can be written as

\[
\Gamma' \phi^{-1} \Gamma' = \frac{1}{4} \begin{pmatrix}
\beta_2 \beta_2' & -\beta_2 \beta_1' & \cdots & 0 \\
-\beta_2 \beta_2' & \beta_2 \beta_1' & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{pmatrix}
\]

\[
+ \frac{1}{4} \begin{pmatrix}
\beta_3 \beta_3' & 0 & -\beta_3 \beta_1' & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
-\beta_1 \beta_3' & 0 & \beta_1 \beta_1' & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix}
+ \cdots
\]
\[
\begin{align*}
&+ \frac{1}{q} g_{q-1,q}^{-1} \begin{pmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & -\beta_{q-1}\beta_q \\
\end{pmatrix} \\
&+ \sum_{h=q+1}^{q+p-1} \begin{pmatrix}
\beta_{h+1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0 \\
\end{pmatrix} \begin{pmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0 \\
\end{pmatrix} \\
&+ \cdots + \sum_{h=q+1}^{q+p-1} \begin{pmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0 \\
\end{pmatrix} \begin{pmatrix}
\beta_{h+1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0 \\
\end{pmatrix} .
\end{align*}
\]

Using this, we get

\[
z'z = \begin{pmatrix}
\hat{\beta}_1 - \beta_1 \\
\vdots \\
\hat{\beta}_q - \beta_q \\
\end{pmatrix} \begin{pmatrix}
\Gamma \Phi^{-1} \Gamma' \\
\vdots \\
\Gamma \Phi^{-1} \Gamma' \\
\end{pmatrix} \begin{pmatrix}
\hat{\beta}_1 - \beta_1 \\
\vdots \\
\hat{\beta}_q - \beta_q \\
\end{pmatrix}
\]

\[
= \frac{1}{q} \sum_{j=1}^{q-1} \sum_{z=j+1}^{q} g_{j,z}^{-1} (\hat{\beta}_j - \beta_j)^2 + \sum_{j=1}^{q} \sum_{z=q+1}^{q+p} g_{j,z}^{-1} (\hat{\beta}_j - \beta_j)^2 .
\]

\[\text{(3.3)}\]
In practical applications the asymptotic covariance matrix \( \Psi(q) \) is not known, but can be consistently estimated by substituting the \( \hat{\lambda}_{ij} \) and \( \hat{\delta}_{j} \) for the respective parameters. This does not affect the validity of the asymptotic chi square approximation. Using \( \hat{\delta}_{j} \) instead of \( \delta_{j} \) and \( \hat{g}_{jj}^{-1} = \sum_{i=1}^{k} r_{i} (\hat{\lambda}_{ij} - \hat{\lambda}_{ij})/\hat{\lambda}_{ij} \hat{\lambda}_{ij} \) in (3.2) we get therefore

**Theorem 4:** Under \( H_{q} \) as defined in (3.1), the statistic

\[
X^{2}(H_{q}) = n \left[ \sum_{j=1}^{q-1} \sum_{k=j+1}^{q} \hat{\delta}_{j} \hat{\delta}_{k}^{-1} \left( \hat{\delta}_{j}^{0} - \hat{\delta}_{j} \right) \hat{\delta}_{k}^{0} \right]^{2} + \sum_{j=1}^{q} \sum_{k=q+1}^{p} \hat{g}_{jj}^{-1} (\hat{\delta}_{j}^{0})^{2}
\]  

(3.4)

is asymptotically distributed as chi square with \( q(p-(q+1)/2) \) degrees of freedom.

It is worth noting that (3.4) has an intuitively appealing interpretation. The first sum ranges over all pairs of eigenvectors fixed under \( H_{q} \). If \( H_{q} \) holds, we would expect \( \hat{\delta}_{j} \) and \( \delta_{j}^{0} \) to be nearly orthogonal for \( l \neq j \), and the cosines \( \hat{\delta}_{l} \hat{\delta}_{j}^{0} \) would be expected close to zero. Similarly, the second sum extends over the squared cosines between the hypothetical vectors \( \delta_{j}^{0} \) and the observed vectors \( \delta_{j} \) for the p-q eigenvectors not considered under \( H_{q} \). The \( \hat{g}_{jj}^{-1} \) serve as weights—again an appealing result, regarding the fact that \( \hat{g}_{jj}^{-1} \) is large if (at least in one group) \( \lambda_{ij} \) and \( \lambda_{ii} \) are far apart. It may also be noted that Theorem 4 does not require that the eigenvectors \( \delta_{q+1}^{0}, \ldots, \delta_{p}^{0} \) are well defined.

Two special cases deserve more attention.

**Case q = 1:** If only one hypothetical eigenvector \( \delta_{1}^{0} \) is specified, the first sum in (3.4) is empty, and the test statistic becomes
\[ x^2(H_1) = n \sum_{k=2}^{p} \frac{-1}{\lambda_{kk}^{0}} (\hat{\beta}_k^{0} \hat{\beta}_k')^2 \]

\[ = n \hat{\beta}_1^{0} \left( \sum_{i=1}^{k} \frac{\pi_i}{\lambda_{ii}^{0}} \left( \hat{\lambda}_{i1} \left( \hat{\Sigma}_1 - \lambda_{11}^{0} \hat{\beta}_1^{0} \hat{\beta}_1' \right) \right. \right. \]

\[ + \left. \hat{\lambda}_{11}^{0} \hat{\Sigma}_1 - \lambda_{11}^{0} \hat{\beta}_1^{0} \hat{\beta}_1' \right) - 2(\lambda_{11}^{0} \hat{\beta}_1^{0} \hat{\beta}_1') \]

\[ = n \sum_{i=1}^{k} \pi_i \left( \hat{\lambda}_{i1}^{0} \hat{\Sigma}_1 \hat{\beta}_1^{0} + \hat{\lambda}_{11}^{0} \hat{\Sigma}_1 \hat{\beta}_1^{0} - 2 \right), \quad (3.5) \]

where \( \hat{\Sigma}_1 = \sum_{j=1}^{p} \hat{\lambda}_{ij}^{0} \hat{\beta}_j^{0} \hat{\beta}_j' \) is the ML-estimate of \( \Sigma_1 \). The number of degrees of freedom associated with (3.5) is \( p-1 \). If \( k = 1 \), then (3.5) reduces to the well-known result given by Anderson (1963, p. 145). If we replace \( \hat{\Sigma}_1 \) by \( \Sigma_1 \) and \( \hat{\lambda}_{11} \) by \( \lambda_{11} \) (the first eigenvalue of \( \Sigma_1 \)), we get a test for the hypothesis \( H_1^* \) that \( \beta_1^{0} \) is the first principal component of \( \Sigma_1, \ldots, \Sigma_k \) without specifying the CPC-model. The test statistic

\[ x^2(H_1^*) = \sum_{i=1}^{k} \pi_i \left( \lambda_{i1}^{0} \Sigma_1 \hat{\beta}_1^{0} + \lambda_{11}^{0} \Sigma_1 \hat{\beta}_1^{0} - 2 \right) \]

\[ (3.6) \]

is merely the sum of \( k \) independent statistics of the form given by Anderson, and its asymptotic null distribution is chi square with \( k(p-1) \) degrees of freedom.

**Case \( q = p \).** If all common eigenvectors \( \beta_1^{0}, \ldots, \beta_p^{0} \) of the \( \Sigma_1 \) are specified, the second sum in (3.4) is empty. Since \( \beta_p^{0} \) is completely determined by \( \beta_1^{0}, \ldots, \beta_{p-1}^{0} \), the hypotheses \( H_{p-1} \) and \( H_p \) are equivalent.
Interestingly, the two associated statistics $X^2_{p-1}$ and $X^2_p$ are in general not identical (unless $p = 2$), but the degrees of freedom are $p(p-1)/2$ for both statistics.

4. ASYMPTOTIC INFERENCE FOR EIGENVALUES

4.1 Confidence intervals for individual roots and sums of roots.

From Theorem 1 it is easy to construct an asymptotic confidence interval for any (distinct) root $\lambda_{ij}$. Denoting by $z_{\alpha/2}$ the upper $\alpha/2$ quantile of the standard normal distribution, the interval

$$
\left( \frac{\hat{\lambda}_{ij}}{1 + z_{\alpha/2}/\sqrt{N_i}}, \frac{\hat{\lambda}_{ij}}{1 - z_{\alpha/2}/\sqrt{N_i}} \right)
$$

(4.1)

contains the parameter $\lambda_{ij}$ (for large $N_i$) with probability approximately $1 - \alpha$. Of course $\alpha$ and $N_i$ must be such that the upper limit of (4.1) is positive.

In practical applications an investigator might wish to neglect CPC's with small variances. It is therefore useful to construct confidence intervals for sums of the $\lambda_{ij}$ within the same population. Typically we might look at the last $p-q$ eigenvalues of $\Sigma_i$ (assuming that the eigenvectors are arranged in some reasonable order). Let

$$
d_i = \sum_{j=q+1}^{p} \lambda_{ij}; \quad \hat{d}_i = \sum_{j=q+1}^{p} \hat{\lambda}_{ij}.
$$

(4.2)

Since the asymptotic variance of $\hat{d}_i$ is $2 \sum_{j=q+1}^{p} \lambda_{ij}^2/n_i$, it is not possible
as in (4.1) to solve explicitly for $d_i$. However, following Anderson
(1963, p. 133f), we can estimate this variance consistently and use

$$
\left( \frac{1}{\hat{d}_i + \left( \frac{2}{n_i} \sum_{j=q+1}^{p} \hat{\lambda}^2_{ij} \right)^{1/2} t_{\alpha/2}} \right) \quad \frac{1}{\hat{d}_i - \left( \frac{2}{n_i} \sum_{j=q+1}^{p} \hat{\lambda}^2_{ij} \right)^{1/2} t_{\alpha/2}}
$$

(4.3)

as a confidence interval containing $d_i$ with probability approximately
$1 - \alpha$. Here, $t_{\alpha/2}$ is the upper $\alpha/2$-quantile of the $t$-distribution
with $n_i$ degrees of freedom.

Similarly, one-sided confidence intervals can easily be constructed.

4.2 Neglecting common principal components with relatively small variances.

If the main purpose of a CPC is data reduction, it is useful to have
some criterion for discarding CPC's with relatively small variances. Let
$d_i$ and $\hat{d}_i$ be defined as in (4.2), and let

$$
c_i = \sum_{j=1}^{q} \hat{\lambda}_{ij}; \quad \hat{c}_i = \sum_{j=1}^{q} \hat{\lambda}_{ij},
$$

(4.4)

such that $c_i + d_i = \text{trace } \Sigma_i$, $\hat{c}_i + \hat{d}_i = \text{trace } \hat{\Sigma}_i$. Suppose that we wish
to discard the last $p-q$ CPC's in population $i$ if their relative
contribution to the trace of $\Sigma_i$ is not larger than a given fraction $f_0$
($0 < f_0 < 1$). Putting

$$
f_i = \frac{d_i}{\text{trace } \Sigma_i},
$$

(4.5)

the asymptotic distribution of $\sqrt{n_i} \left( (1-f_i) \hat{d}_i - f_i \hat{c}_i \right)$ is normal with mean
zero and variance \[ 2 \left[ f_0^2 \frac{q}{j=1} \lambda_{ij}^2 + (1-f_0)^2 \frac{p}{j=q+1} \lambda_{ij}^2 \right]. \] (The use of this criterion has been proposed by Anderson (1968, p. 115). Estimating this variance consistently by putting in the corresponding ML-estimates \( \hat{\lambda}_{ij} \) yields

\[
Z_i = \frac{\sqrt{n_i} \left[ (1-f_0) \hat{d}_i - f_0 \hat{d}_i \right]}{2 \left[ f_0^2 \frac{q}{j=1} \lambda_{ij}^2 + (1-f_0)^2 \frac{p}{j=q+1} \lambda_{ij}^2 \right]} \sim N(0,1) \quad (4.6)
\]

approximately for large \( n_i \) and under the hypothesis \( f_i = f_0 \). For testing the hypothesis that all \( f_i \) \((i = 1, \ldots, k)\) are less or equal \( f_0 \), a possible procedure is to reject the hypothesis if

\[
\max_{1 \leq i \leq k} Z_i > z_\beta
\]

with \( \beta = 1 - (1-\alpha)^{1/k} \),

where \( z_\beta \) is the upper \( \beta \)-quantile of the standard normal distribution. This test has asymptotic level \( \alpha \) if all \( f_i \) equal \( f_0 \).

4.3 A likelihood ratio test for sphericity of \( p-q \) common principal components.

In PCA, the main motivation for testing for equality of \( p-q \) (out of \( p \)) characteristic roots stems from the model \( \xi = \psi + \sigma^2 I_p \), where \( \psi \) is positive semidefinite of rank \( q \). In this model the last \( p-q \) characteristic roots are all \( \sigma^2 \). In CPC, we can study the model \( \xi_i = \psi_i + \sigma_i^2 I_p \) \((i = 1, \ldots, k)\), where the \( \psi_i \) are simultaneously diagonalizable and of
rank q. Then the $\lambda_i$ satisfy $H_c$, and the last p-q CPC's are spherical, i.e.

$$H_s: \lambda_{i,q+1} = \ldots = \lambda_{ip} \quad (i = 1, \ldots, k). \quad (4.8)$$

We will refer to $H_s$ as "hypothesis of simultaneous sphericity of p-q CPC's".

It may be noted that the following derivation of the likelihood ratio test holds as well for any subset of CPC's, but for notational simplicity it is given in terms of the situation described above.

Putting $\lambda_{i,q+1} = \ldots = \lambda_{ip} = \lambda_i^\ast \quad (i = 1, \ldots, k)$, we get from (2.1)

$$-2g(A_1, \ldots, A_k, \beta|S_1, \ldots, S_k)$$

$$= \sum_{i=1}^{k} n_i \sum_{j=1}^{q} \left( \log \lambda_{ij} + \beta_j S_i \beta_j / \lambda_{ij} \right)$$

$$+ (p-q) \log \lambda_i^\ast + \left( \sum_{j=q+1}^{p} \beta_j S_i \beta_j / \lambda_i^\ast \right). \quad (4.9)$$

Using the same technique as Flury (1984) the likelihood equations are obtained as

$$\beta'_j \left( \sum_{i=1}^{k} n_i \frac{\lambda_{ij} - \lambda_{i,j}^\ast}{\lambda_{ij} \lambda_{i,j}^\ast} S_i \right) \beta_j = 0 \quad (1 \leq j \leq q)$$

$$\beta'_j \left( \sum_{i=1}^{k} n_i \frac{\lambda_{ij} - \lambda_{i,j}^\ast}{\lambda_{ij} \lambda_{i,j}^*} S_i \right) \beta_j = 0 \quad (1 \leq j \leq q \leq p).$$
\[ \lambda_{ij} = \beta_i' S_i \beta_j \quad (i = 1, \ldots, k; \ j = 1, \ldots, q) \]

\[ \lambda_i^* = \left( \sum_{j=q+1}^{p} \frac{\beta_i' S_i \beta_j}{(p-q)} \right) \quad (i = 1, \ldots, k) \]  

(4.10)

with the orthogonality restrictions \( \beta_i' \beta_j = 0 \) \( (i \neq j) \), \( \beta_i' \beta_i = 1 \).

Equation system (4.10) can be solved using a modified version of the FG-algorithm (Flury and Gautschi 1984).

In contrast to the unrestricted CPC-model the vectors \( \tilde{\beta}_{q+1}, \ldots, \tilde{\beta}_p \) are not uniquely determined by the likelihood equations. In fact, only the subspace spanned by \( \tilde{\beta}_{q+1}, \ldots, \tilde{\beta}_p \) is determined. Let us denote by \( \tilde{\xi}_1, \ldots, \tilde{\xi}_q, \tilde{\beta}_{q+1}, \ldots, \tilde{\beta}_p \) a set of orthonormal vectors solving (4.10), then the same maximum of the likelihood is obtained if we replace \( (\tilde{\beta}_{q+1}, \ldots, \tilde{\beta}_p) \) by \( (\tilde{\beta}_{q+1}, \ldots, \tilde{\beta}_p) H \), where \( H \) is an arbitrary orthogonal matrix of dimension \( (p-q) \times (p-q) \). With \( \tilde{\lambda}_{ij}^\dagger \) and \( \lambda_i^* \) denoting the ML-estimates of the eigenvalues, the log likelihood ratio statistic for \( H_s \) can be written as

\[ X^2_s = \sum_{i=1}^{k} \sum_{j=1}^{q} \frac{\lambda_i^* \lambda_i^\dagger \lambda_{ij}^* \lambda_{ij}^\dagger}{(p-q)} \]  

(4.11)

where the \( \lambda_{ij}^\dagger \) are the ML-estimates for the ordinary (unrestricted) CPC-model. Under \( H_s \), the number of parameters determining \( \xi_1, \ldots, \xi_k \) is \( q(2p-q-1)/2 + k(q+1) \), compared with \( p(p-1)/2 + kp \) parameters for the ordinary CPC-model (see, e.g., Mardia, Kent and Bibby (1979, p. 235f) for a discussion of this problem in the one-group situation). Thus the null distribution of (4.11) is asymptotically chi square with \( (p-q-1)(p-q+2k)/2 \) degrees of freedom.
It may be noted that, unless \( k = 1 \), the ML-estimates \( \hat{\beta}_j \) and \( \hat{\lambda}_{ij} \) for \( j \leq q \) are not identical with \( \tilde{\beta}_j \) and \( \tilde{\lambda}_{ij} \). However, we can approximate \( X_s^2 \) without computing the restricted solution by replacing \( \tilde{\lambda}_{ij} \) by \( \hat{\lambda}_{ij} \) (\( j = 1, \ldots, q \)) and \( \hat{\lambda}_1^* \) by \( \hat{\lambda}_1^* = (\hat{\lambda}_{1,q+1} + \ldots + \hat{\lambda}_{1,p})/(p-q) \). This yields

\[
X_s^2(\text{approx}) = \sum_{i=1}^{k} n_i \log \frac{(\hat{\lambda}_1^*)^{p-q}}{P_{j=q+1}^{j=p} \hat{\lambda}_{ij}}.
\]

Since, under \( H_s \), the likelihood is maximized for the \( \tilde{\lambda} \)'s, we have always \( X_s^2(\text{approx}) \geq X_s^2 \). Thus the approximate statistic can be used to accept \( H_s \), but not to reject it. Similar tests can be constructed for the hypothesis of sphericity in \( k' \) groups (\( 1 \leq k' < k \)), i.e. for

\[
H_s': \lambda_{1,q+1} = \ldots = \lambda_{1,p} \quad (i = 1, \ldots, k').
\]

Again the resulting likelihood equations can be solved using appropriate modifications of the PG-algorithm, but we are not going into the details for this case which seems of minor practical importance.

5. APPLICATIONS

In this section some of the preceding theory is illustrated by a numerical example. The data used have been published by Jolicoeur and Mosimann (1960) and have served as an example of PCA in various textbooks...
(e.g. Morrison 1976; Mardia, Kent and Bibby 1979). The main appeal of this example is its simplicity—the data are only 3-dimensional, yet illustrate the purpose of CPCA clearly, which outweighs the disadvantage of rather small sample sizes.

Table 1a displays covariance matrices $S_{ij}$ of samples of $n_1 + 1 = 24$ male and $n_2 + 1 = 24$ female individuals of the species *Chrysemys picta marginata* (painted turtle). The variables (1) log (carapace length); (2) log (carapace width); (3) log (carapace height). The logarithms are used instead of the measured variables because of their relationship to allometry; see Morrison (1976, p. 295). Table 1b shows the eigenvalues of the $S_{ij}$ and the ML-estimates $\hat{\lambda}_{ij}$. The value of the chi square statistic for $H_c$ (Flury 1984) is $\chi^2 = 7.93$ with 3 degrees of freedom, which is close to the 95 percent quantile of the asymptotic null distribution of the criterion. Regarding the relatively small sizes it may be reasonable to assume that $H_c$ holds.

Table 1c shows the estimated eigenvectors $\hat{\beta}_j$ and the estimated standard errors of their coefficients. The standard errors were obtained from the main diagonal of the sample analog of (2.15). It is obvious that $\hat{\beta}_1$ has stable coefficients, while $\hat{\beta}_2$ and $\hat{\beta}_3$ seem rather poorly defined.

The hypothesis of allometric growth of an organism (Jolicoeur 1963) implies that the first principal component of the covariance matrix of the logarithms of the measured dimensions is $\beta'_1 = (1, \ldots, 1)/\sqrt{p}$. Let us therefore test the hypothesis $H_1$ (3.1) for $\beta_1^0 = (1,1,1)'/\sqrt{3}$. The statistic $\chi^2(H_1)$ is obtained from (3.4) or (3.5) as $\chi^2(H_1) = 46.17$ with 2 degrees of freedom. At any reasonable level $\alpha$ we would therefore conclude that the allometric model does not hold in this case.
Finally, let us see whether the second and third CPC's are well defined, i.e. let us test the hypothesis of simultaneous sphericity of the second and third CPC's. The null hypothesis is $H_5 : \lambda_{i2} = \lambda_{i3}$ ($i = 1, 2$). Without computing the ML-estimates under $H_5$, we can easily calculate the approximation (4.12) from the values displayed in Table 1b. The resulting statistic is $X^2_s(\text{approx}) = 3.24$. Since this is far below the 95%-quantile of the chi square distribution with 3 degrees of freedom, we conclude that $H_5$ is tenable. Taking into consideration the relative smallness of the second and third roots in both groups, it is thus reasonable to think of the three shell dimensions as distributed about a single principal axis ("size") and two minor axes containing merely measurement errors, the main axis having the same orientation in space for both male and female turtles.

6. CONCLUSIONS

In this paper we have shown that asymptotic theory can be used for inference on CPC-models. The methods given in sections 3 and 4 reflect merely the author's opinion about which hypotheses might be important in practice. Other hypotheses and restrictions of the model can easily be formulated; we might for instance be interested in a model where some of the eigenvalues of $\Sigma_1$ and $\Sigma_h$ ($h \neq i$) are identical. Tests for such hypotheses could be constructed either by the likelihood ratio method or using the asymptotic results of Section 2.

One open problem deserves perhaps to be investigated: Suppose that we are interested only in the first $q$ (out of $p$) CPC's and wish to neglect
the last \( p-q \) components. Then we would actually not care whether
the \( \Sigma_i \) have all eigenvectors in common—it would be sufficient to know
that \( \beta_1, \ldots, \beta_q \) are common to \( \Sigma_1, \ldots, \Sigma_k \). This could be called a
partial CPC-model.

Obviously a partial CPC model may hold even when the ordinary CPC-
model is wrong, and the test for \( H_C \) may in some situations reject the
hypotheses although the first \( q \) eigenvectors are common to all matrices
\( \Sigma_1, \ldots, \Sigma_k \). To my knowledge no work has been done on this problem so far,
but the diagnostic tools illustrated in example 1 of Flury (1984) might be
used to fit a partial CPC-model. Future research will perhaps lead to a
more satisfactory solution.
TABLE 1: Common Principal Component Analysis of Turtle Carapace Dimensions, Transformed Logarithmically

a) Sample covariance matrices (multiplied by $10^2$)

<table>
<thead>
<tr>
<th></th>
<th>males ($n_1 = 23$)</th>
<th>females ($n_2 = 23$)</th>
</tr>
</thead>
</table>
| $S_1$ | \[
\begin{pmatrix} 1.1072 & .8019 & .8160 \\ .8019 & .6417 & .6005 \\ .8160 & .6005 & .6773 \end{pmatrix}
\] | $S_2$ = \[
\begin{pmatrix} 2.6391 & 2.0124 & 2.5443 \\ 2.0124 & 1.6190 & 1.9782 \\ 2.5443 & 1.9782 & 2.5899 \end{pmatrix}
\] |

b) Variances of CPC's and Eigenvalues of $S_1$

<table>
<thead>
<tr>
<th></th>
<th>males:</th>
<th>females:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\lambda}_{1j}$</td>
<td>2.3148</td>
<td>6.7135</td>
</tr>
<tr>
<td>$\hat{\lambda}_{2j}$</td>
<td>0.0729</td>
<td>0.0807</td>
</tr>
<tr>
<td>eigenvalues</td>
<td>2.3303</td>
<td>0.0598</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0360</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0538</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0530</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\beta}_3$</th>
</tr>
</thead>
</table>
| males:| \[
\begin{pmatrix} .6406 \\ .4905 \\ .5907 \end{pmatrix}
\] | \[
\begin{pmatrix} -.3844 \\ -.4611 \\ .7998 \end{pmatrix}
\] | \[
\begin{pmatrix} .1811 \\ .2012 \\ .0313 \end{pmatrix}
\] |
| females: | (0.0128) | (0.1811) | (0.1050) |
|         | (0.0146) | (0.2012) | (0.1256) |
|         | (0.0159) | (0.0313) | (0.2176) |
|         | (0.0159) | (0.0313) | (0.2176) |
|         | (0.0159) | (0.0313) | (0.2176) |
REFERENCES


26
