A GENERALIZATION OF PRINCIPAL COMPONENT ANALYSIS TO K GROUPS

by

Bernhard N. Flury*
Purdue University and University of Berne

Technical Report # 83-14

Department of Statistics
Purdue University

May 1983

* This work was done under contract no. 82.008.0.82 of the Swiss National Science Foundation. I wish to thank all those with whom I discussed the problems of this paper: Walter Gautschi, George McCabe, Robb Muirhead, K. C. S. Pillai, T. Ramalingam, Donald B. Rubin, Herman Rubin, and others.

Abstract:

This paper generalizes the method of principal components to so called "common principal components" in the following way: Assume the covariance matrices Σ_i ($i = 1,\ldots,k$) are simultaneously diagonalizable, that is, there is an orthogonal matrix β such that $\beta' \Sigma_i \beta$ is diagonal for $i = 1,\ldots,k$. Given samples from normal populations satisfying this condition on the covariance matrices, we derive the maximum likelihood estimates of Σ_i, $i = 1,\ldots,k$, and the log-likelihood-ratio statistic for testing the hypothesis of common principal axes. The solution is shown to have some favorable properties which do not depend on the normal assumptions. The method is illustrated by numerical examples. Applications to data reduction, multiple regression and nonlinear discriminant analysis are sketched.

Keywords: Maximum likelihood; Covariance matrices; Discriminant analysis

1. Introduction

Principal Component Analysis (PCA) is a well known and established technique of multivariate statistical analysis. It enjoys a solid theoretical foundation and possesses many optimal properties (Okamoto 1969, McCabe 1982). One important property of principal components (PC's) is that they can be considered as uncorrelated variables, obtained by an orthogonal rotation of the coordinate system. PCA is a one-sample method, and to the author's knowledge only two efforts have been made to generalize it to a two-sample method: Krzanowski (1979) compares the PC's of two different samples by computing the angles between the subspaces spanned by the first q PC's of each group. Flury (1983) gives a method of obtaining uncorrelated variables in two groups simultaneously, sacrificing the condition of orthogonality.

In practice we often deal with the situation of variables being measured on grouped objects, and the covariance structure may vary from group to group. Examples for this range from three species of Iris (Fisher 1936) over male and female turtles (Jolicoeur and Mosimann 1960) and human bones (Jolicoeur 1963) to real and forged bank notes (Flury and Riedwyl 1983). In all these cases, tests of significance suggest that the underlying population covariance matrices are not identical in all groups. Yet there may be a certain similarity between the covariance matrices of different groups, which could be used to improve estimation. One such similarity might be that the PC - transformation is the same in all populations. The mathematical formulation of this is the simultaneous diagonalizeability of k $p \times p$ - covariance matrices $\Sigma_1, \ldots, \Sigma_m^k$:

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

where β is an orthogonal $p \times p$ - matrix.

As will be shown in section 4 of this paper, assumption (1.1) may be quite reasonable in certain applications. As the columns of β can be viewed as the rotated coordinate axes, we will refer to condition (1.1) as "common principal axes", and the projections of the variables on the common principal axes will be called "common principal components (CPC's)". Note that, in contrast to the one-sample case, no obvious fixed order of the columns of β is given, since the rank order of the diagonal elements of the Λ_i is not necessarily the same for all Λ_i.

2. Maximum Likelihood Estimation of Common Principal Axes in k Normal Populations

Let the p-variate random vectors $X_i (i=1,\ldots,k)$ be independently distributed as $N_p(\mu_i, \Sigma_i)$, where $\mu_i \in \mathbb{R}^p$ and the Σ_i are positive definite and symmetric (p.d.s.). For samples of size $N_i = n_i + 1$, denote by $S_i (i = 1,\ldots k)$ the usual unbiased sample covariance matrices. Assume $\min_{1 \le i \le k} n_i \geq p$. Then the matrices $n_i S_i$ are independently distributed as $W_p(n_i, \Sigma_i)$ (Muirhead 1982, p. 85), and the common likelihood function of $\Sigma_1, \ldots, \Sigma_k$, given $S_1, \ldots, S_k$ is

$$L(\Sigma_1,\ldots,\Sigma_k) = C \times \prod_{i=1}^{k} \operatorname{etr}(-\frac{n_i}{2}\Sigma_i^{-1} S_i)\,|\Sigma_i|^{-n_i/2} , \quad (2.1)$$

where C is a constant which does not depend on the Σ_i, and etr denotes the exponential function of the trace. (Note that we could have started with the common likelihood function of the k normal samples, which would yield the same results as those obtained below, with n_i replaced by N_i.) Instead of maximizing (2.1) we can minimize the function

$$g(\Sigma_1,\ldots,\Sigma_k) = -2 \log L(\Sigma_1,\ldots,\Sigma_k) = \sum_{i=1}^{k} n_i(\log|\Sigma_i| + \operatorname{tr}\Sigma_i^{-1}S_i). \quad (2.2)$$

Let us now assume that (1.1) holds for an orthogonal matrix β. Let $\Lambda_i = \operatorname{diag}(\lambda_{i1},\ldots,\lambda_{ip})$, then

$$\log|\Sigma_i| = \sum_{j=1}^{p} \log \lambda_{ij} \qquad (i = 1,\ldots,k) \qquad (2.3)$$

and

$$\operatorname{tr}\Sigma_i^{-1}S_i = \operatorname{tr}(\beta\Lambda_i^{-1}\beta' S_i) = \operatorname{tr}(\Lambda_i^{-1}\beta' S_i \beta) \quad (i=1,\ldots k). \quad (2.4)$$

Let now $W^{(i)} = \beta' S_i \beta$ and β_j = the j-th column of β, then

$$\operatorname{tr}(\Lambda_i^{-1}W^{(i)}) = \sum_{j=1}^{p} w_{jj}^{(i)} / \lambda_{ij} \qquad (i=1,\ldots,k), \qquad (2.5)$$

where

$$w_{jj}^{(i)} = \beta_j' S_i \beta_j \qquad (i=1,\ldots,k ; j=1,\ldots,p) \qquad (2.6)$$

is the (j,j) - element of $W^{(i)}$. It follows that

$$\operatorname{tr}\Sigma_i^{-1} S_i = \sum_{j=1}^{p} \beta_j' S_i \beta_j / \lambda_{ij} \qquad (i = 1,\ldots,k) \qquad (2.7)$$

and therefore

$$g(\underset{\sim}{\Sigma}_1,\ldots,\underset{\sim}{\Sigma}_k) = g(\underset{\sim}{\beta}_1,\ldots,\underset{\sim}{\beta}_p, \lambda_{11},\ldots, \lambda_{1p}, \lambda_{21},\ldots, \lambda_{kp})$$
$$= \sum_{i=1}^{k} n_i[\sum_{j=1}^{p} (\log \lambda_{ij} + \underset{\sim}{\beta}_j'\underset{\sim}{S}_i\underset{\sim}{\beta}_j / \lambda_{ij})]. \qquad (2.8)$$

The function g is to be minimized under the restrictions

$$\underset{\sim}{\beta}_h' \underset{\sim}{\beta}_j = \begin{cases} 0 & \text{if } h \neq j \\ 1 & \text{if } h = j \end{cases} . \qquad (2.9)$$

Thus we wish to minimize the function

$$G(\underset{\sim}{\Sigma}_1,\ldots,\underset{\sim}{\Sigma}_k) = g(\underset{\sim}{\Sigma}_1,\ldots,\underset{\sim}{\Sigma}_k) - \sum_{h=1}^{p} \gamma_h(\underset{\sim}{\beta}_h'\underset{\sim}{\beta}_h - 1) - 2 \sum_{h<j}^{p} \gamma_{hj}\underset{\sim}{\beta}_h'\underset{\sim}{\beta}_j, \qquad (2.10)$$

where $\gamma_h(1 \leq h \leq p)$ and $\gamma_{hj}(1 \leq h < j \leq p)$ are $p(p + 1)/2$ Lagrange-multipliers. Taking the partial derivative with respect to γ_{ij} and setting this equal to zero yields immediately

$$\gamma_{ij} = \underset{\sim}{\beta}_j'\underset{\sim}{S}_i\underset{\sim}{\beta}_j \quad (i = 1,\ldots,k;\ j = 1,\ldots,p) \quad , \qquad (2.11)$$

and from (2.7) it follows that

$$\operatorname{tr} \underset{\sim}{\Sigma}_i^{-1}\underset{\sim}{S}_i = p \qquad (i = 1,\ldots,k) \quad . \qquad (2.12)$$

The vector of partial derivatives of G with respect to $\underset{\sim}{\beta}_j$, set equal to zero, is

$$\sum_{i=1}^{k} n_i\underset{\sim}{S}_i\underset{\sim}{\beta}_j / \lambda_{ij} - \sum_{\substack{h=1 \\ h\neq j}}^{p} \lambda_{jh}\underset{\sim}{\beta}_h - \gamma_j\underset{\sim}{\beta}_j = \underset{\sim}{0} \quad (j = 1,\ldots,p) \qquad (2.13)$$

where we put $\gamma_{jh} = \gamma_{hj}$ if $j < h$. Multiplying (2.13) from the left by $\underset{\sim}{\beta}_j'$ gives

$$\gamma_j = \sum_{i=1}^{k} n_i \qquad (j = 1,\ldots,p) \tag{2.14}$$

and thus

$$\sum_{i=1}^{k} n_i \underset{\sim}{S}_i \underset{\sim}{\beta}_j / \lambda_{ij} - \left(\sum_{i=1}^{k} n_i\right) \underset{\sim}{\beta}_j - \sum_{\substack{h=1\\h\neq j}}^{p} \gamma_{jh} \underset{\sim}{\beta}_h = \underset{\sim}{0} \quad (j = 1,\ldots,p). \tag{2.15}$$

Multiplying (2.15) from the left by $\underset{\sim}{\beta}_\ell'$ $(\ell \neq j)$ implies

$$\sum_{i=1}^{k} n_i \underset{\sim}{\beta}_\ell' \underset{\sim}{S}_j \underset{\sim}{\beta}_j / \lambda_{ij} = \gamma_{j\ell} \qquad (j = 1,\ldots,p;\ \ell \neq j). \tag{2.16}$$

By deriving (2.15) for $\underset{\sim}{\beta}_\ell$ and noting that $\underset{\sim}{\beta}_j' \underset{\sim}{S}_i \underset{\sim}{\beta}_\ell = \underset{\sim}{\beta}_\ell' \underset{\sim}{S}_i \underset{\sim}{\beta}_j$

it follows that

$$\sum_{i=1}^{k} n_i \underset{\sim}{\beta}_\ell' \underset{\sim}{S}_i \underset{\sim}{\beta}_j / \lambda_{i\ell} = \gamma_{j\ell} \qquad (\ell = 1,\ldots,p\ ;\ \ j \neq \ell) \tag{2.17}$$

and therefore, comparing (2.16) and (2.17),

$$\sum_{i=1}^{k} n_i \underset{\sim}{\beta}_\ell' \underset{\sim}{S}_i \underset{\sim}{\beta}_j / \lambda_{ij} = \sum_{i=1}^{k} n_i \underset{\sim}{\beta}_\ell' \underset{\sim}{S}_i \underset{\sim}{\beta}_j / \lambda_{i\ell} \quad (\ell \neq j). \tag{2.18}$$

This can be written as

$$\underset{\sim}{\beta}_\ell' \left(\sum_{i=1}^{k} n_i \frac{\lambda_{i\ell} - \lambda_{ij}}{\lambda_{i\ell}\lambda_{ij}} \underset{\sim}{S}_i \right) \underset{\sim}{\beta}_j = 0 \quad (\ell, j = 1,\ldots,p;\ \ell \neq j). \tag{2.19}$$

As the (ℓ,j)-th equation of (2.19) is the same as the (j,ℓ)-th, we have actually only $p(p-1)/2$ equations, say for $1 \leq \ell < j \leq p$. These have to be solved under the orthogonality conditions $\underset{\sim}{\beta}'\underset{\sim}{\beta} = \underset{\sim}{I}_p$ (2.9) and using (2.11). A numerical algorithm to accomplish this has been developed by Flury and Gautschi (1983); a short description of their so-called FG-algorithm is given in the appendix of this paper. Note that, in general, the solution of (2.19) is not unique, but the FG-algorithm converges rather fast to the solution which maximizes the likelihood.

Let us denote the maximizing solution by $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ and $\hat{\lambda}_{ij}(i = 1,\ldots,k;\ j = 1,\ldots,p)$, and put $\hat{\Lambda}_i = \text{diag}\,(\hat{\lambda}_{i1},\ldots,\hat{\lambda}_{ip})$ and $\hat{\Sigma}_i = \hat{\beta}\hat{\Lambda}_i\hat{\beta}'$ $(i = 1,\ldots,k)$. Then (using 2.12) the maximum of (2.1) is obtained as

$$L(\hat{\Sigma}_1,\ldots,\hat{\Sigma}_k) = C \times \prod_{i=1}^{k} \exp(-pn_i/2)|\hat{\Sigma}_i|^{-n_i/2}, \qquad (2.20)$$

while the unrestircted maximum can easily be seen to be equal to

$$L(S_1,\ldots,S_k) = C \times \prod_{i=1}^{k} \exp(-pn_i/2)\ |S_i|^{-n_i/2}. \qquad (2.21)$$

The log-likelihood-ratio statistic is therefore

$$X^2 = -\,2\log\frac{L(\hat{\Sigma}_1,\ldots,\hat{\Sigma}_k)}{L(S_1,\ldots,S_k)} = \sum_{i=1}^{k} n_i \log\frac{|\hat{\Sigma}_i|}{|S_i|}, \qquad (2.22)$$

and by the general theory of likelihood ratio tests (see Rao 1973, chapter 6) it follows that X^2 has asymptotically ($\min_{1\le i\le k} n_i \to \infty$) a chi square distribution with (k-1)p(p-1)/2 degrees of freedom, if the null hypothesis of identical principal axes in all k populations is true.

3. PROPERTIES OF COMMON PRINCIPAL COMPONENTS

Before we apply the method derived in section 2 to numerical examples, it may be useful to state some simple properties of the common principal components, which will illustrate their meaning and facilitate their correct application. Note that, from now on, we will refer to the "sample common principal components"

$$\underset{\sim}{U}_i = \hat{\underset{\sim}{\beta}}'\underset{\sim}{X}_i \qquad (i = 1,\ldots,k) \tag{3.1}$$

as CPC's, suppressing the prefix "sample".

Let us recall that the solution of the ML estimation problem consists of a orthogonal matrix $\hat{\underset{\sim}{\beta}} = (\hat{\underset{\sim}{\beta}}_1,\ldots,\hat{\underset{\sim}{\beta}}_p)$ and kp quadratic forms $\hat{\lambda}_{ij} = \hat{\underset{\sim}{\beta}}_j'\underset{\sim}{S}_i\hat{\underset{\sim}{\beta}}_j$. For the i-th sample, the $\hat{\lambda}_{ij}$ are the diagonal elements of the matrix

$$\underset{\sim}{F}_i = \hat{\underset{\sim}{\beta}}'\underset{\sim}{S}_i\hat{\underset{\sim}{\beta}} \qquad (i = 1,k). \tag{3.2}$$

Note that, in the one sample case, $\underset{\sim}{F}_i$ would be exactly diagonal, while this is in general not true in the k sample case. From (2.11) and (3.2) we have

$$\hat{\underset{\sim}{\Lambda}}_i = \text{diag}\,(\underset{\sim}{F}_i) \qquad (i = 1,\ldots,k) \quad . \tag{3.3}$$

$\hat{\underset{\sim}{\Lambda}}_i$ is the estimated covariance of the CPC's in population i and is, of course, diagonal, while $\underset{\sim}{F}_i$ is the sample covariance matrix of $\underset{\sim}{U}_i$. The importance of the $\underset{\sim}{F}_i$-matrices can now be seen from the fact that

$$|\underset{\sim}{S}_i| = |\underset{\sim}{F}_i| = \prod_{j=1}^{p} \ell_{ij} \qquad (i - 1,\ldots,k) \quad , \tag{3.4}$$

where the ℓ_{ij} are the eigenvalues of $\underset{\sim}{S}_i$ (and simultaneously those of $\underset{\sim}{F}_i$). Moreover, as

$$\hat{\underset{\sim}{\Sigma}}_i = \hat{\underset{\sim}{\beta}}\hat{\underset{\sim}{\Lambda}}_i\hat{\underset{\sim}{\beta}}' = \hat{\underset{\sim}{\beta}}(\text{diag}\,\underset{\sim}{F}_i)\hat{\underset{\sim}{\beta}}' \qquad (i - 1,\ldots,k) \quad , \tag{3.5}$$

we have

$$|\hat{\underset{\sim}{\Sigma}}_i| = |\text{diag}\,\underset{\sim}{F}_i| = \prod_{j=1}^{p} f_{jj}^{(i)} \qquad (i = 1,\ldots,k) \quad , \tag{3.6}$$

where $f_{jj}^{(i)}$ is the (j,j)-element of $\underset{\sim}{F}_i$. The statistic (2.22) can therefore be written as a function of the $\underset{\sim}{F}_i$ alone, namely

$$X^2 = \sum_{i=1}^{k} n_i \log \frac{|\mathrm{diag}\ \underset{\sim}{F}_i|}{|\underset{\sim}{F}_i|}$$

$$= \sum_{i=1}^{k} n_i \log \frac{\prod_{j=1}^{p} f_{jj}^{(i)}}{\prod_{j=1}^{p} \ell_{ij}}$$

$$= \sum_{i=1}^{k} n_i \sum_{j=1}^{p} (\log f_{jj}^{(i)} - \log \ell_{ij}) \quad . \tag{3.7}$$

This can be used to look at CPC's in the following distribution-free way: Clearly, the determinant of a p.d.s. matrix with given diagonal elements takes its maximum when the off-diagonal elements are all zero. Therefore, if we measure the "degree of diagonality" of a p.d.s. matrix $\underset{\sim}{F}_i$ by $d_i = \log (|\mathrm{diag}\ \underset{\sim}{F}_i| \ / \ |\underset{\sim}{F}_i|) \geq 0$, take k such matrices and sum their "degrees of diagonality", using weights n_i, the result is (3.7). It is also obvious from (3.7) that X^2 is zero if all the $\underset{\sim}{F}_i$ are exactly diagonal. X^2 is therefore a measure of "simultaneous diagonalizeability" of k p.d.s. matrices $\underset{\sim}{S}_i$. The CPC's can therefore be viewed as obtained by a simultaneous transformation yielding variables which are "as uncorrelated as possible".

To stress the importance of the $\underset{\sim}{F}_i$-matrices, we will compute them in the examples of section 4. Sometimes it may even be more convenient to interpret the correlation matrices

$$R_i = \hat{\Lambda}_i^{-1/2} \, F_i \, \hat{\Lambda}_i^{-1/2} \qquad (i = 1,\ldots,k) \quad . \tag{3.8}$$

These can easily provide information about violations of the assumption of common principal axes.

From the equation system (2.19) we can still gain a better understanding of CPC's. First, let us write it as

$$\beta_\ell' (n_1 A_{\ell j}^{(1)} + \ldots + n_k A_{\ell j}^{(k)}) \, \beta_j = 0 \qquad (1 \leq \ell < j \leq p), \tag{3.9}$$

where

$$A_{\ell j}^{(i)} = \frac{\lambda_{i\ell} - \lambda_{ij}}{\lambda_{i\ell} \, \lambda_{ij}} \, S_i \qquad (i = 1,\ldots,k). \tag{3.10}$$

Equation (3.9) remains unchanged if we replace one of the S_i (say S_1) by a proportional matrix cS_1 $(c > 0)$. This can be seen by denoting $S_1^* = cS_1$, then $\lambda_{1j}^* = c\lambda_{1j}$ $(j = 1,\ldots,p)$, and $(\lambda_{1\ell}^* - \lambda_{1j}^*) / \lambda_{1\ell}^* \lambda_{1j}^*$ $(\lambda_{1\ell} - \lambda_{1j}) / c \, \lambda_{1\ell} \lambda_{1j}$, leaving $A_j^{(1)}$ unchanged. This means that the CPC's are invariant under proportionality, where different constants of proportionality are admitted for the k groups.

Furthermore, we note that the weight of matrix S_i in (2.19) is $n_i(\lambda_{i\ell} - \lambda_{ij}) / \lambda_{i\ell} \lambda_{ij}$. Now, by (2.11), λ_{ij} is the variance of the linear combination $\beta_j' X_i$, and it becomes clear that the weight of matrix S_i in the (ℓ,j)-th equation is smaller, the closer the two variances λ_{ij} and $\lambda_{i\ell}$ are. If $\lambda_{ij} = \lambda_{i\ell}$, then S_i disappears from the (ℓ,j)-th equation. Of course this makes perfect sense, for it corresponds to sphericity of $\underset{\sim}{X}_i$ in the plane spanned by $\underset{\sim}{\beta}_j$ and $\underset{\sim}{\beta}_\ell$. In the extreme case $\underset{\sim}{S}_i = c\underset{\sim}{I}_p$, the i-th matrix will disappear completely from (2.19). We can therefore say that the overall influence of matrix $\underset{\sim}{S}_i$ is proportional to its degrees of freedom n_i, but also to its "deviation from sphericity".

As these considerations show, CPC's enjoy quite a few desireable properties and can easily be understood as a generalization of PCA to k groups. In the next section we will show that application of common principal component analysis (CPCA) to real data is not much more complicated than PCA, but can give very useful information about similarities and differences in the covariance structures of k populations.

4. APPLICATIONS

The practical use of CPCA can be demonstrated best if we apply it to some well-known examples in the multivariate literature.

Example 1: Fisher's Iris-data

These famous and most-abused data were first published by R. A. Fisher (1936) as an example of discriminant analysis. The four variables were

(1) sepal length

(2) sepal width

(3) petal length

(4) petal width,

measured on 3 species of Iris: versicolor, virginica and setosa. The sample covariance matrices, based on 49 degrees of freedom each, are shown in Table 1a. Part b of the same table displays the ML-estimates $\hat{\Sigma}_i$ under the restriction of common principal axes. Table 1c shows the estimates of the common principal axes, each column of $\hat{\beta}$ representing the coefficients for one component. The CPC's do not seem to have an obvious interpretation. Part d of Table 1 shows the variances $\hat{\lambda}_{ij}$ in all three groups, ordered according to the columns of $\hat{\beta}$. (Of course, this order is irrelevant and depends only on the initial approximation used in the FG-algorithm). In order to compare PC's and CPC's we give also the eigenvalues of the S_i (or F_i), again ordered in the same sense. If the hypotheses of common principal axes in all three populations is true, we would expect all the $\hat{\lambda}_{ij}$ to be close to the eigenvalues of S_i. This is the case for sample 1 (versicolor), while the differences are larger in sample 2 (virginica) and even worse in sample 3 (setosa). This impression is confirmed by the value $X^2 = 63.91$ of the statistic (2.22), which can be compared with quantiles of the chi square distribution with 12 degrees of freedom. At any reasonable level of significance we would conclude that the assumption of common principal axes does not hold, though the chi square approximation might still be rather poor for sample sizes of 50, and also non-normality might affect the exact significance level.

In order to learn more about the deviation of the data from the model of common principal axes, we can look at the F_i - and R_i -

matrices, given in table 1e. Note that the diagonals of the $\underset{\sim}{F}_i$ - matrices contain the variances of the CPC's. There is obviously a rather high correlation between the first and third CPC's in group 3 (setosa), which might explain the inadequacy of the assumption of common principal axes.

The results of the same analysis, this time performed with groups 1 (versicolor) and 2 (virginica) only, are displayed in table 2. The value of the statistic (2.22) is $X^2 = 13.46$, which lies between the 95-th and the 99-th percentile of the chi square distribution with 6 degrees of freedom. Though it seems doubtful whether the hypothesis of common principal axes should be accepted, we can note that the variances of the CPC's (diagonal of the $\underset{\sim}{F}_i$ - matrices) are now much closer to the eigenvalues of the $\underset{\sim}{S}_i$. Since the two $\underset{\sim}{F}_i$ - matrices are based on the same linear transformation, we can also look at the ratios of variances of the four CPC's, and we note that the largest ratio is 6.2186/5.1354 = 1.21 (for the third CPC), while the smallest ratio is obtained from the fourth CPC as 1.0119/4.5813 = .221. These two ratios are close to the extreme characteristic roots of $\underset{\sim}{S}_2^{-1}\,\underset{\sim}{S}_1$, as can be expected from corollary 1 of Flury (1983).

Example 2: Size and shape variation in the painted turtle.
In this famous example, due to Jolicoeur and Mosimann (1960), we use the logorithms of the original data, as proposed by Morrison (1976, p. 286). Table 3 displays the results of a CPCA, applied to $N_1 = 24$ male and $N_2 = 24$ female individuals of Chrysemys picta marginata, where the variables are

(1) log (carapace length)

(2) log (carapace width)

(3) log (carapace heighth) .

(Note that the covariance matrix of the logorithmically transformed data as given by Morrison (1976, p. 286) appears incorrect). As in the usual one-group PCA, the three CPC's can easily be interpreted as "size" (first) and "shape" (second, third) - variables. The eigenvalues and the variances of the CPC's are remarkably close (table 3d). The log-likelihood ratio statistic (2.22) is $X^2 = 7.93$ with 3 degrees of freedom; but since the two samples are rather small, this statistic should be interpreted with caution. Note that in the "size"-component the females vary considerably more than the males.

Example 3: Bone dimensions of the North American Marten.
Jolicoeur (1963) measured the variables

(1) log (length of the humerus)

(2) log (width of the humerus)

(3) log (length of the femur)

(4) log (width of the femur)

on $N_1 = 92$ male and $N_2 = 47$ female individuals of the species Martes americana. Principal component analyses performed separately on each group showed a similar pattern of the two arthogonal matrices. The CPCA of these data is summarized in table 4. The value of the log-likelihood ratio statistic is $X^2 = 8.34$, with 6 degrees of freedom. The hypothesis of common principal axes seems therefore quite plausible.

Comparing the coefficients of the CPC's with those given by Jolicoeur for each sex separately, we can see that his interpretation applies as well to our analysis, but it becomes simpler, since transformations are the same in both groups. Table 4d shows that the diagonal elements and eigenvalues of the $\underset{\sim}{F}_i$ are very close. The correlations between the CPC's, displayed in table 4e, confirm the impression that CPC's are justified in this example. Note that correlations in group 2 (females) tend to be larger in absolue value than in group 1 -- an effect of the unequality of sample sizes.

Example 4: Human bone measurements.

In the same paper as quoted in example 3, Jolicoeur gives sample covariance matrices of the two variables

(1) length of the femur

(2) width of the femur,

measured on $N_1 = 48$ male and $N_2 = 40$ female humans. This time, data were not transformed, and the original data were not available to the author of this paper. The results of a CPCA are summarized in table 5. The statistic (2.22) is $X^2 = .95$ with 1 degree of freedom. The correlation between the two CPC's is - .1055 for males and .1037 for females. Note that, comparing the variances of the two CPC's within each group, the ratio of the larger to the smaller variance is 27.01 for males and 39.66 for females. This shows that (using the terminology of Jolicoeur) the femur is more robust in males than in females.

Example 5: Real and forged bank notes

From Flury and Riedwyl (1983) we take the following variables, measured on real and forged Swiss bank notes:

(1) width of the bank note, measured on the left side

(2) width of the bank note, measured on the right side

(3) width of the lower margin

(4) width of the upper margin.

All measurements were in mm. The two samples consisted of $N_1 = 100$ real and $N_2 = 85$ forged notes which had probably all been produced by the same forgerer. The results of a CPCA on these data are displayed in table 6. The coefficients of the four CPC's, given in part c of table 6, show a rather interesting pattern, which can be compared with the one-group PCA performed by Flury and Riedwyl (1983, p. 120 ff) on the real notes, using two additional variables. The CPC's can roughly be interpreted as

(1) slant of the cut

(2) width of the print

(3) vertical position of the print

(4) size

These interpretations can be brought into relation with two independent phases of the production process (printing, cutting), which makes the assumption of common principal axes seem reasonable. The value of (2.22) is $X^2 = 12.04$, which is close to the 95-th percentile of the chi square distribution with 6 degrees of freedom. Again, it must be borne in mind that the chi square approximation might not be very accurate. Note that the diagonal elements and eigenvalues of the F_i - matrices agree very closely, so that the assumption of common principal axes seems at least a good approximation to reality.

This example shows also a desireable aspect of the method: since two eigenvalues of $\underset{\sim}{S}_2$ are rather close (.1024 and .1305 respectively), the associated principal components might be unstable in the sense of a nearly spherical distribution in the plane spanned by the two eigenvectors. However, this problem is taken care of by sample 1, where the corresponding variances differ much more. This illustrates a property of CPCA which was already mentioned at the end of section 3.

5. CONCLUSIONS

As the preceding examples show, CPCA may have quite useful applications. Some reasons for performing CPCA can be summarized as follows:

- reduction of parameter space: If the assumption of identical principal axes hold (and it might make sense to assume it in many examples, especially of a biological nature), it is obviously better to reduce the parameter space (ignoring mean vectors) from kp + kp(p-1)/2 to kp + p(p-1)/2 elements, and the estimates can be expected to have smaller variance than those of the unrestricted parameter space.

- data reduction: If the main purpose of a PCA consists of data reduction, and if in all k groups the smallest variances appear in the same CPC's, it might be useful to transform the data of all groups simultaneously to CPC's, and to discard the CPC's which have small variances in all k groups simultaneously.

- regression on PC's: A frequent application of PCA is to use the PC's instead of the measured variables as regressors in multiple regression. The obvious extension of this is to use CPC's if dummy-variables are to be included into the regression, the k groups being defined by the dummy-variables. However, in contrast to the one-group situation, the regressors will not be exactly uncorrelated in this case.

- non-linear discriminant analysis: Under the assumption of common principal axes the computation of Mahalanobis-distances for classification of observations could be much simplified: Let $\underset{\sim}{u} = (u_1, \ldots, u_p)'$ denote an observation and $\underset{\sim}{\bar{u}}^{(i)} = (\bar{u}_1^{(i)}, \ldots, \bar{u}_p^{(i)})'$ the mean vector of the i-th sample in the space of common principal axes, then the formula for computing the Mahalanobis-distance between $\underset{\sim}{u}$ and $\underset{\sim}{\bar{u}}^{(i)}$ becomes simply

$$D^2(\underset{\sim}{u}, \underset{\sim}{\bar{u}}^{(i)}) = \underset{\sim}{d}_i' \, \underset{\sim}{\hat{\Lambda}}_i^{-1} \, \underset{\sim}{d}_i = \sum_{j=1}^{p} d_{ij}^2 \, / \, \hat{\lambda}_{ij} \quad (i = 1, \ldots, k), \qquad (5.1)$$

where $\underset{\sim}{d}_i = (u_1 - \bar{u}_1^{(i)}, \ldots, u_p - \bar{u}_p^{(i)})'$, and $\underset{\sim}{\hat{\Lambda}}_i = \text{diag}\, \underset{\sim}{F}_i$ is defined in section 3. We hope to be able to show that this procedure improves the rate of correct classification, provided that the assumptions hold at least approximately.

APPENDIX: THE FG - ALGORITHM

This algorithm, developed by Flury and Gautschi (1983), solves the equation system

$$\mathbf{b}_\ell'\left(n_1 \frac{\lambda_{1\ell}-\lambda_{1j}}{\lambda_{1\ell}\,\lambda_{1j}}\, \mathbf{S}_1 + \dots + n_k \frac{\lambda_{k\ell}-\lambda_{kj}}{\lambda_{k\ell}\,\lambda_{kj}}\, \mathbf{S}_k\right) \mathbf{b}_j = 0$$

$$(1 \le \ell < j \le p) \qquad \text{(A.1)}$$

where $n_1,\dots,n_k$ are fixed positive numbers, and $\mathbf{S}_1,\dots,\mathbf{S}_k$ are fixed p.d.s. matrices of dimension p x p. (A.1) is to be solved under the restrictions

$$\lambda_{ij} = \mathbf{b}_j'\mathbf{S}_i\mathbf{b}_j \quad (i = 1,\dots,k;\ j = 1,\dots,p) \qquad \text{(A.2)}$$

and

$$\mathbf{b}_\ell'\mathbf{b}_j = \begin{cases} 0 & \ell \ne j \\ 1 & \ell = j \end{cases} \qquad \text{(A.3)}$$

that is, the solution $\mathbf{B} = (\mathbf{b}_1,\dots,\mathbf{b}_p)$ is a orthogonal matrix.
The FG-algorithm solves (A.1) by iteration on two levels:

- On the F-level, every pair $(\mathbf{b}_\ell,\mathbf{b}_j)$ of column vectors of $\mathbf{B}$ is rotated such that the (ℓ,j)-th equation of (A.1) is satisfied. One iteration on the F-level consists of rotations of all $p(p-1)/2$ pairs of vectors.

- On the G-level, (A.1) is solved iteratively for the two dimensional case, in which (A.1) consists of only one equation.

The F-algorithm

The key to this algorithm lies in the idea of pairwise adjustment of the columns of the orthogonal matrix $\underset{\sim}{B}$, similar to varimax (and other)-rotations in factor analysis (see, e.g., Weber 1974). The problem of pairwise adjustment is reduced to a 2-dimensional version of (A.1). The iterations of the algorithm yield a sequence of orthogonal matrices $\underset{\sim}{B}^{(0)}, \underset{\sim}{B}^{(1)},\ldots$

step F_0: define $\underset{\sim}{B}(p\times p) = (\underset{\sim}{b}_1,\ldots,\underset{\sim}{b}_p)$ as an initial approximation to the solution of (A.1), e.g. $\underset{\sim}{B} = \underset{\sim}{I}_p$. Put $f \leftarrow 0$.

step F_1: put $f \leftarrow f+1$ and $\underset{\sim}{B}^{(f)} \leftarrow \underset{\sim}{B}$

step F_2: repeat steps F_{21} to F_{24} for all pairs (ℓ,j), $1 \le \ell < j \le p$:

F_{21}: put $\underset{\sim}{G}(p\times 2) \leftarrow (\underset{\sim}{b}_\ell,\underset{\sim}{b}_j)$ and

$$\underset{\sim}{T}_i(2\times 2 \leftarrow \begin{pmatrix} b_\ell' S_i b_\ell & \underset{\sim}{b}_\ell' \underset{\sim}{S}_i \underset{\sim}{b}_j \\ \underset{\sim}{b}_j' \underset{\sim}{S}_i \underset{\sim}{b}_\ell & \underset{\sim}{b}_j' \underset{\sim}{S}_i \underset{\sim}{b}_j \end{pmatrix} \qquad (i = 1,\ldots,k)$$

($\underset{\sim}{T}_i$ is p.d.s.)

F_{22}: perform the G-algorithm on $(\underset{\sim}{T}_1,\ldots,\underset{\sim}{T}_k)$ to get an orthogonal

$$2\times 2 \text{ - matrix } \underset{\sim}{Q} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

$\underline{F_{23}}$: put $\underset{\sim}{G}^*(p\times 2) = (\underset{\sim}{b}^*_\ell, \underset{\sim}{b}^*_j) \leftarrow \underset{\sim}{G}\underset{\sim}{Q}$ (This corresponds to an orthogonal rotation of the two columns of $\underset{\sim}{G}$ by an angle φ).

$\underline{F_{24}}$: Replace (in the matrix $\underset{\sim}{B}$) $\underset{\sim}{b}_\ell$ by $\underset{\sim}{b}^*_\ell$, $\underset{\sim}{b}_j$ by $\underset{\sim}{b}^*_j$

$\underline{\text{step } F_3}$: If $\underset{\sim}{B}^{(f)}$ and $\underset{\sim}{B}$ are close enough (e.g. $||\underset{\sim}{B}^{(f)} - \underset{\sim}{B}|| < \varepsilon$) stop. Otherwise start next iteration at F_1.

A better initial approximation in step F_0 might be to take the eigenvectors of one of the $\underset{\sim}{S}_i$ (e.g. the one with largest n_i) as columns of $\underset{\sim}{B}$. In the examples of section 4, three to five iterations were required for $p = 4$ and $\varepsilon = .0001$. A six-dimensional example required 9 iterations for the same ε, using $\underset{\sim}{I}_6$ as an initial approximation for $\underset{\sim}{B}$.

The G-algorithm

This algorithm solves the equation

$$\underset{\sim}{q}_1' \left(n_1 \frac{\delta_{11}-\delta_{12}}{\delta_{11}\,\delta_{12}} \underset{\sim}{T}_1 + \ldots + n_k \frac{\delta_{k1}-\delta_{k2}}{\delta_{k1}\,\delta_{k2}} \underset{\sim}{T}_k\right) \underset{\sim}{q}_2 = 0 \quad \text{(A.4)}$$

where $\underset{\sim}{T}_1,\ldots,\underset{\sim}{T}_k$ are fixed p.d.s. 2×2-matrices, $n_i > 0$ are fixed constants,

$$\delta_{ij} = \underset{\sim}{q}_j'\underset{\sim}{T}_i\underset{\sim}{q}_j \qquad (i = 1,\ldots,k\ ;\ j = 1,2) \qquad \text{(A.5)}$$

and

$$\underset{\sim}{q}_1'\underset{\sim}{q}_1 = \underset{\sim}{q}_2'\underset{\sim}{q}_2 = 1 \quad ; \quad \underset{\sim}{q}_1'\underset{\sim}{q}_2 = 0 \qquad \text{(A.6)}$$

i.e. $\underset{\sim}{Q} = (\underset{\sim}{q}_1, \underset{\sim}{q}_2)$ is a orthogonal 2×2-matrix. The iterations of the algorithm yield a sequence of orthogonal matrices $\underset{\sim}{Q}^{(0)}, \underset{\sim}{Q}^{(1)}, \ldots$

<u>step G_0</u>: define $\underset{\sim}{Q}(2\times2)$ as an initial approximation to the solution of (A.4). (In later iterations on the F-level, $\underset{\sim}{Q} = \underset{\sim}{I}_p$ is the best choice). put $g \leftarrow 0$

<u>step G_1</u>: put $g \leftarrow g+1$, $\underset{\sim}{Q}^{(g)} \leftarrow \underset{\sim}{Q}$

<u>step G_2</u>: compute the δ_{ij} (A.5) using the current Q.

$$\text{Put } \underset{\sim}{A}(2\times2) \leftarrow n_1 \frac{\delta_{11}-\delta_{12}}{\delta_{11}\,\delta_{12}} \underset{\sim}{T}_1 + \ldots + n_k \frac{\delta_{k1}-\delta_{k2}}{\delta_{k1}\,\delta_{k2}} \underset{\sim}{T}_k$$

($\underset{\sim}{A}$ is symmetric)

<u>step G_3</u>: compute the (normalized) eigenvectors of $\underset{\sim}{A}$.

Put $\underset{\sim}{q}_1 \leftarrow$ first eigenvector of $\underset{\sim}{A}$, $\underset{\sim}{q}_2 \leftarrow$ second eigenvector of $\underset{\sim}{A}$.

<u>step G_4</u>: If $\underset{\sim}{Q}^{(g)}$ and $\underset{\sim}{Q}$ are close enough (e.g. $||\underset{\sim}{Q}^{(g)} - \underset{\sim}{Q}|| < \varepsilon$, possibly after permutation of the columns of $\underset{\sim}{Q}$ and/or multiplication of one or both columns with -1), stop. Otherwise start next iteration at G_1.

In our examples, up to 10 iterations on the G-level were required when the current matrix $\underset{\sim}{B}$ on the F-level was still far from the solution, but this number reduces drastically (to 1 or 2) as soon as $\underset{\sim}{B}$ is close to the solution.

References

FISHER, R. A. (1936), "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, VII, p4. 11, 179-188.

FLURY, B. (1983), "Some Relations Between the Comparison of Covariance Matrices and Principal Component Analysis," Computational Statistics and Data Analysis, 1,

FLURY, B. and GAUTSCHI, W. (1983), "The FG-algorithm," in preparation.

FLURY, B. and RIEDWYL, H. (1983), Angewandte multivariate Statistik, Stuttgart & New York: Verlag Gustav Fischer.

JOLICOEUR, P. (1963), "The Degree of Generality in Martes Americana," Growth, 27, 1-27.

JOLICOEUR, P. and MOSIMANN, J. E. (1960), "Size and Shape Variation in the Painted Turtle: A Principal Component Analysis," Growth, 24, 339-354.

KRZANOWSKI, W. J. (1979), "Between-Group Comparison of Principal Components," Journal of the American Statistical Association, 74, 703-707.

McCABE, G. P. (1982), "Principal Variables," Technical Report No. 82-3, Dept. of Statistics, Purdue University.

MORRISON, D. F. (1976), Multivariate Statistical Methods, New York: McGraw-Hill.

MUIRHEAD, R. J. (1982), Aspects of Multivariate Statistical Theory, New York: John Wiley.

OKAMOTO, M. (1969), "Optimality of Principal Components," in Multivariate Analysis - II, ed. P. R. Krishnaiah, New York: Academic Press, 673-685.

RAO, C. R. (1973), Linear Statistical Inference and its Applications, 2nd edition, New York: John Wiley.

WEBER, E. (1974), Einführung in die Faktorenanalyse, Stuttgart & New York: Verlag Gustav Fischer.

Table 1: Estimation of CPC's in Fisher's 1936 Iris data. The sample covariance matrices reported here were multiplied with 10^2.

Table 2: The same data as in table 1, but without group 3.

Table 3: Common Principal Component Analysis of Turtle Carapace Dimensions (transformed logarithmically). Sample Covariance Matrices Multiplied with 10^2.

Table 4: Bone Measurements on Martes Americana (data transformed logorithmically).

Table 5: Measurements on human femur.

Table 6: CPCA of real and forged bank notes.

Table 1

a) sample covariance matrices

Versicolor $(N_1 = 50)$

$$\underset{\sim}{S}_1 = \begin{pmatrix} 26.6433 & 8.5184 & 18.2898 & 5.5780 \\ 8.5184 & 9.8469 & 8.2653 & 4.1204 \\ 18.2898 & 8.2653 & 22.0816 & 7.3102 \\ 5.5780 & 4.1204 & 7.3102 & 3.9106 \end{pmatrix}$$

Virginica $(N_2 = 50)$

$$\underset{\sim}{S}_2 = \begin{pmatrix} 40.4343 & 9.3763 & 30.3290 & 4.9094 \\ 9.3763 & 10.4004 & 7.1380 & 4.7629 \\ 30.3290 & 7.1380 & 30.4588 & 4.8824 \\ 4.9094 & 4.7629 & 4.8824 & 7.5433 \end{pmatrix}$$

Setosa $(N_3 = 50)$

$$\underset{\sim}{S}_3 = \begin{pmatrix} 12.4249 & 9.9216 & 1.6355 & 1.0331 \\ 9.9216 & 14.3690 & 1.1698 & .9298 \\ 1.6355 & 1.1698 & 3.0159 & .6069 \\ 1.0331 & .9298 & .6069 & 1.1106 \end{pmatrix}$$

---

b) MLE's of population covariance matrices

$$\hat{\underset{\sim}{\Sigma}}_1 = \begin{pmatrix} 29.5860 & 7.3004 & 18.6600 & 4.6667 \\ 7.3004 & 7.4546 & 6.6121 & 2.8309 \\ 18.6600 & 6.6121 & 21.2145 & 6.2692 \\ 4.6667 & 2.8309 & 6.2692 & 3.2273 \end{pmatrix}$$

$$\hat{\underset{\sim}{\Sigma}}_2 = \begin{pmatrix} 40.6417 & 11.5005 & 27.8263 & 7.9275 \\ 11.5005 & 11.0588 & 8.8976 & 2.8603 \\ 27.8263 & 8.8976 & 29.6478 & 7.0677 \\ 7.9275 & 2.8603 & 7.0677 & 7.4885 \end{pmatrix}$$

$$\hat{\underset{\sim}{\Sigma}}_3 = \begin{pmatrix} 9.4477 & 3.5268 & 4.5255 & 1.2613 \\ 3.5268 & 10.2264 & -2.5687 & .2601 \\ 4.5255 & -2.5687 & 9.5669 & 2.1149 \\ 1.2613 & .2601 & 2.1149 & 1.6793 \end{pmatrix}$$

c) coefficients of common principal components

$$\hat{\beta} = \begin{pmatrix} .7367 & -.6471 & -.1640 & .1084 \\ .2468 & .4655 & -.8346 & -.1607 \\ .6047 & .5003 & .5221 & -.3338 \\ .1753 & .3382 & .0628 & .9225 \end{pmatrix}$$

---

d) variances $\hat{\lambda}_{ij}$ of CPC's and eigenvalues of S_i

Versicolor:

$\hat{\lambda}_{1j}$ =	48.4602	7.4689	5.5394	1.0139
eigenvalues =	48.7874	7.2384	5.4776	.9790

Virginica:

$\hat{\lambda}_{2j}$ =	69.2235	6.7124	7.5367	5.3642
eigenvalues =	69.5255	5.2295	10.6552	3.4266

Setosa:

$\hat{\lambda}_{3j}$ =	14.6444	2.7526	12.5065	1.0169
eigenvalues =	23.6496	2.6796	3.6919	.9033

---

e) covariance and correlation matrices of CPC's

$$F_1 = \begin{pmatrix} 48.4602 & 3.4072 & -1.1931 & .7172 \\ 3.4072 & 7.4689 & -.3776 & .2049 \\ -1.1931 & -.3776 & 5.5394 & -.3278 \\ .7172 & .2049 & -.3278 & 1.0139 \end{pmatrix}$$

$$R_1 = \begin{pmatrix} 1.0000 & .1791 & -.0728 & .1023 \\ .1791 & 1.0000 & -.0587 & .0745 \\ -.0728 & -.0587 & 1.0000 & -.1383 \\ .1023 & .0745 & -.1383 & 1.0000 \end{pmatrix}$$

$$F_2 = \begin{pmatrix} 69.2235 & -1.6211 & 2.6003 & -2.9062 \\ -1.6211 & 6.7124 & -1.9278 & 2.3514 \\ 2.6003 & -1.9278 & 7.5367 & -2.2054 \\ -2.9062 & 2.3514 & -2.2054 & 5.3642 \end{pmatrix}$$

$$\underset{\sim}{R}_2 = \begin{pmatrix} 1.0000 & -.0752 & .1138 & -.1508 \\ -.0752 & 1.0000 & -.2710 & .3919 \\ .1138 & -.2710 & 1.0000 & -.3468 \\ -.1508 & .3919 & -.3468 & 1.0000 \end{pmatrix}$$

$$\underset{\sim}{F}_3 = \begin{pmatrix} 14.6444 & -.5682 & -9.9950 & -.2106 \\ -.5682 & 2.7526 & .0487 & -.4236 \\ -9.9950 & .0487 & 12.5065 & .4235 \\ -.2106 & -.4236 & .4235 & 1.0169 \end{pmatrix}$$

$$\underset{\sim}{R}_3 = \begin{pmatrix} 1.0000 & -.0895 & -.7385 & -.0546 \\ -.0895 & 1.0000 & .0083 & -.2532 \\ -.7385 & .0083 & 1.0000 & .1188 \\ -.0546 & -.2532 & .1188 & 1.0000 \end{pmatrix}$$

Table 2

a) MLE's of population covariance matrices

$$\hat{\underset{\sim}{\Sigma}}_1 = \begin{pmatrix} 28.1693 & 7.8487 & 19.1036 & 4.9047 \\ 7.8487 & 8.8832 & 6.1772 & 3.4523 \\ 19.1036 & 6.1772 & 21.9980 & 6.1257 \\ 4.9047 & 3.4523 & 6.1257 & 3.4319 \end{pmatrix}$$

$$\hat{\underset{\sim}{\Sigma}}_2 = \begin{pmatrix} 38.7745 & 10.4893 & 28.7304 & 7.8987 \\ 10.4893 & 13.1731 & 9.5169 & 4.3989 \\ 28.7304 & 9.5169 & 29.6661 & 7.1734 \\ 7.8987 & 4.3989 & 7.1734 & 7.2231 \end{pmatrix}$$

---

b) coefficients of CPC's

$$\hat{\underset{\sim}{\beta}} = \begin{pmatrix} .7206 & -.2914 & -.6159 & .1286 \\ .2545 & .9019 & -.1900 & -.2927 \\ .6188 & -.1186 & .7188 & -.2939 \\ .1817 & .2960 & .2607 & .9008 \end{pmatrix}$$

---

c) covariance matrices of CPC's

$$\underset{\sim}{F}_1 = \begin{pmatrix} 48.5836 & 2.7247 & .9370 & .4711 \\ 2.7247 & 6.6683 & .9278 & -.3659 \\ .9370 & .9287 & 6.2186 & -.0118 \\ .4711 & -.3659 & -.0118 & 1.0119 \end{pmatrix}$$

$$\underset{\sim}{F}_2 = \begin{pmatrix} 69.1434 & -4.1101 & -.7288 & -2.2362 \\ -4.1101 & 9.9766 & -.1066 & 2.5981 \\ -.7288 & -.1066 & 5.1354 & .4135 \\ -2.2362 & 2.5981 & .4135 & 4.5813 \end{pmatrix}$$

Table 3

a) sample covariance matrices

males $(N_1 = 24)$

$$\underset{\sim}{S}_1 = \begin{pmatrix} 1.1072 & .8019 & .8160 \\ .8019 & .6417 & .6005 \\ .8160 & .6005 & .6773 \end{pmatrix}$$

females $(N_2 = 24)$

$$\underset{\sim}{S}_2 = \begin{pmatrix} 2.6391 & 2.0124 & 2.5443 \\ 2.0124 & 1.6190 & 1.9782 \\ 2.5443 & 1.9782 & 2.5899 \end{pmatrix}$$

---

b) MLE's of population covariance matrices

$$\underset{\sim}{\hat{\Sigma}}_1 = \begin{pmatrix} .9778 & .7214 & .8509 \\ .7214 & .5935 & .6469 \\ .8509 & .6469 & .8548 \end{pmatrix}$$

$$\underset{\sim}{\hat{\Sigma}}_1 = \begin{pmatrix} 2.7911 & 2.0976 & 2.5120 \\ 2.0976 & 1.6619 & 1.9198 \\ 2.5120 & 1.9198 & 2.3950 \end{pmatrix}$$

---

c) coefficients of CPC's

$$\underset{\sim}{\hat{\beta}} = \begin{pmatrix} .6406 & -.6647 & -.3844 \\ .4905 & .7394 & -.4611 \\ .5907 & .1069 & .7998 \end{pmatrix}$$

---

d) variances of CPC's and eigenvalues of S_i

males:

$\hat{\lambda}_{1j}$ =	2.3148	.0385	.0729
eigenvalues =	2.3303	.0360	.0598

females:

$\hat{\lambda}_{2j}$ =	6.7135	.0538	.0807
eigenvalues =	6.7200	.0530	.0750

---

e) <u>covariance matrices of CPC's</u>

$$\underset{\sim}{F}_1 = \begin{pmatrix} 2.3148 & -.0483 & -.1811 \\ -.0483 & .0385 & -.0019 \\ -.1811 & -.0019 & .0729 \end{pmatrix}$$

$$\underset{\sim}{F}_2 = \begin{pmatrix} 6.7135 & .0670 & .1966 \\ .0670 & .0538 & .0037 \\ .1966 & .0037 & .0807 \end{pmatrix}$$

Table 4

a) sample covariance matrices

males (N_1 = 92)

$$\underset{\sim}{S}_1 = \begin{pmatrix} 1.1544 & .9109 & 1.0330 & .7993 \\ .9109 & 2.0381 & .7056 & 1.4083 \\ 1.0330 & .7056 & 1.2100 & .7958 \\ .7993 & 1.4083 & .7958 & 2.0277 \end{pmatrix}$$

females (N_2 = 47)

$$\underset{\sim}{S}_2 = \begin{pmatrix} .9617 & .2806 & .9841 & .6775 \\ .2806 & 1.8475 & .3129 & 1.2960 \\ .9841 & .3129 & 1.2804 & .7923 \\ .6775 & 1.2960 & .7923 & 1.7819 \end{pmatrix}$$

---

b) MLE's of population covariance matrices

$$\hat{\underset{\sim}{\Sigma}}_1 = \begin{pmatrix} 1.0709 & .8014 & .9883 & .8461 \\ .8014 & 2.0413 & .6642 & 1.3798 \\ .9883 & .6642 & 1.1938 & .9137 \\ .8461 & 1.3798 & .9137 & 2.1241 \end{pmatrix}$$

$$\hat{\underset{\sim}{\Sigma}}_2 = \begin{pmatrix} 1.0566 & .4326 & 1.0137 & .6820 \\ .4326 & 1.8615 & .2895 & 1.3034 \\ 1.0137 & .2895 & 1.2826 & .6850 \\ .6820 & 1.3034 & .6850 & 1.6708 \end{pmatrix}$$

---

c) coefficients of CPC's

$$\hat{\underset{\sim}{\beta}} = \begin{pmatrix} .7288 & .3914 & .4864 & -.2811 \\ -.1408 & .5662 & -.5757 & -.5729 \\ -.6637 & .3941 & .6306 & -.0810 \\ .0920 & .6090 & -.1855 & .7656 \end{pmatrix}$$

---

d) variances of the CPC's and eigenvalues of S_i

males:

$\hat{\lambda}_{1j}$ =	.1228	4.5419	1.0811	.6844
eigenvalues =	.1209	4.5482	1.1163	.6447

females:

$\hat{\lambda}_{2j}$ =	.1359	3.7641	1.5987	.3727
eigenvalues =	.1240	3.7749	1.6047	.3679

e) correlation matrices of CPC's

$$\underset{\sim}{R}_1 = \begin{pmatrix} 1.0000 & .0762 & -.0792 & -.0394 \\ .0762 & 1.0000 & .0010 & -.0831 \\ -.0792 & .0010 & 1.0000 & -.1432 \\ -.0394 & -.0831 & -.1432 & 1.0000 \end{pmatrix}$$

$$\underset{\sim}{R}_2 = \begin{pmatrix} 1.0000 & -.1758 & .1314 & .1437 \\ -.1758 & 1.0000 & -.0034 & .1257 \\ .1314 & -.0034 & 1.0000 & .0823 \\ .1437 & .1257 & .0823 & 1.0000 \end{pmatrix}$$

Table 5

a) sample covariance matrices

males (N_1 = 48)

$$\underset{\sim}{S}_1 = \begin{pmatrix} 408.1280 & 35.7910 \\ 35.7910 & 18.3100 \end{pmatrix}$$

females (N_2 = 40)

$$\underset{\sim}{S}_2 = \begin{pmatrix} 356.4590 & 44.9850 \\ 44.9850 & 14.8560 \end{pmatrix}$$

---

b) MLE's of population covariance matrices

$$\hat{\underset{\sim}{\Sigma}}_1 = \begin{pmatrix} 406.2748 & 43.9343 \\ 43.9343 & 20.1632 \end{pmatrix}$$

$$\hat{\underset{\sim}{\Sigma}}_2 = \begin{pmatrix} 357.7821 & 39.1709 \\ 39.1709 & 13.5329 \end{pmatrix}$$

---

c) coefficients of CPC's

$$\hat{\underset{\sim}{\beta}} = \begin{pmatrix} .9937 & -.1116 \\ .1116 & .9937 \end{pmatrix}$$

---

d) variances of the CP's and eigenvalues of S_i

males

$\hat{\lambda}_{1j}$	=	411.2108	15.2272
eigenvalues	=	411.3869	15.0511

females

$\hat{\lambda}_{2j}$	=	362.1830	9.1320
eigenvalues	=	362.2837	9.0313

Table 6

a) sample covariance matrices

real notes ($N_1 = 100$)

$$\underset{\sim}{S}_1 = \begin{pmatrix} .1326 & .0859 & .0567 & .0491 \\ .0859 & .1263 & .0582 & .0306 \\ .0567 & .0582 & .4132 & -.2635 \\ .0491 & .0306 & -.2635 & .4212 \end{pmatrix}$$

forged notes ($N_2 = 85$)

$$\underset{\sim}{S}_2 = \begin{pmatrix} .0641 & .0489 & .0289 & -.0130 \\ .0489 & .0940 & -.0109 & .0071 \\ .0289 & -.0109 & .7242 & -.4330 \\ -.0130 & .0071 & -.4330 & .4039 \end{pmatrix}$$

---

b) MLE's of population covariance matrices

$$\hat{\underset{\sim}{\Sigma}}_1 = \begin{pmatrix} .1253 & .0849 & .0640 & .0425 \\ .0849 & .1329 & .0507 & .0435 \\ .0640 & .0507 & .4674 & -.2512 \\ .0425 & .0435 & -.2512 & .3677 \end{pmatrix}$$

$$\hat{\underset{\sim}{\Sigma}}_2 = \begin{pmatrix} .0677 & .0461 & .0404 & -.0189 \\ .0461 & .0836 & .0170 & -.0202 \\ .0404 & .0170 & .6641 & -.4399 \\ -.0189 & -.0202 & -.4399 & .4708 \end{pmatrix}$$

---

c) coefficients of CPC's

$$\hat{\underset{\sim}{\beta}} = \begin{pmatrix} .7664 & .3140 & .0469 & .5585 \\ -.6297 & .5390 & .0299 & .5586 \\ -.0921 & -.5133 & .7783 & .3497 \\ -.0874 & -.5895 & -.6254 & .5037 \end{pmatrix}$$

d) variances of CPC's and eigenvalues of S_i

real notes:

$\hat{\lambda}_{1j}$ = .0431 .0865 .6750 .2887

eigenvalues = .0426 .0827 .6815 .2865

forged notes:

$\hat{\lambda}_{2j}$ = .0272 .1163 1.0207 .1220

eigenvalues = .0265 .1024 1.0269 .1305

---

e) correlation matrices of CPC's

$$\underset{\sim}{R}_1 = \begin{pmatrix} 1.0000 & .0165 & -.0524 & .0820 \\ .0165 & 1.0000 & .1971 & -.0032 \\ -.0524 & .1971 & 1.0000 & -.0711 \\ .0820 & -.0032 & -.0711 & 1.0000 \end{pmatrix}$$

$$\underset{\sim}{R}_2 = \begin{pmatrix} 1.0000 & -.0087 & .0384 & -.1292 \\ -.0087 & 1.0000 & -.2155 & .0999 \\ .0384 & -.2155 & 1.0000 & .0288 \\ -.1292 & .0999 & .0288 & 1.0000 \end{pmatrix}$$
