3 Mardia, Kent, and Bibby's and Chang's Methods on Selection of Factors as Discriminant Variables

3.1 Factor Analysis

The observable random vector $\mathbf{X}$, with p components, has mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The orthogonal factor model with m common factors is given

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{LF} + \boldsymbol{\varepsilon}$$

where

$\boldsymbol{\mu}_i$: mean of variable i

$\boldsymbol{\varepsilon}_i$: ith specific factor

$\mathbf{F}_j$: jth common factor

l_{ij} : loading of the ith variable on the jth factor

The unobservable random vector $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ satisfy the following conditions:

$\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent,

$E(\mathbf{F}) = \mathbf{0}$, Cov $(\mathbf{F}) = \mathbf{I}$, and

$E(\boldsymbol{\varepsilon}) = \mathbf{0}$, Cov $(\boldsymbol{\varepsilon}) = \Psi$, where Ψ is a diagonal matrix.

Suppose $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ are eigenvalue-eigenvector pairs of $\mathbf{V}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, where $\mathbf{V}$ is the sample covariance matrix. If the factor loadings L are estimated by the principal component method, it is customary to compute factor scores using an unweighted least squares procedure. That is,

$$\mathbf{f}_j = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}}\mathbf{e}_1'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \frac{1}{\sqrt{\lambda_2}}\mathbf{e}_2'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \vdots \\ \frac{1}{\sqrt{\lambda_m}}\mathbf{e}_m'(\mathbf{x}_j - \bar{\mathbf{x}}) \end{pmatrix}, \quad j = 1, 2, \ldots, n.$$

Here, we just consider the factor scores obtained by the unweighted least squares method with the loadings estimated by the principal component method.

3.2 Selection of Factor Variables

In this section, we apply these two methods to select the best factor variables for further discrimination.

3.2.1 Mardia et al. Method

Suppose $\mathbf{x}$ is a p dimensional random variable with a mixture of two multi-normal distributions with means $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and a common covariance matrix $\boldsymbol{\Sigma}$. That is, a random sample $\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$; independently, a random sample $\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2} \sim N_p\ (\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_p$ are eigenvalues of $\boldsymbol{\Sigma}$ with corresponding eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Let $\mathbf{y}_{ij}$ be the principal component value of $\mathbf{x}_{ij}$, so we have $\mathbf{y}_{ij} = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p]'\mathbf{x}_{ij}$. The within group sums of squares matrix, $\mathbf{W} = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)'$. Define $\mathbf{d} = \bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2$ and $m = n_1 + n_2 - 2$. Then we have a test of null hypothesis $H_0 : \boldsymbol{\beta}_2 = 0$ using the sample Mahalanobis distances based on $\mathbf{y}_{ij}$:

$$D_p^2 = m\mathbf{d}'\mathbf{W}^{-1}\mathbf{d} \text{ and } D_k^2 = m\mathbf{d}_1'\mathbf{W}^{-1}\mathbf{d}_1.$$

This test uses the statistic

$$\eta_k = \frac{(m-p+1)}{(p-k)} c(D_p^2 - D_k^2)/(m + cD_k^2). \qquad \text{(see 2.1)}$$

Let $\mathbf{f}_{ij}$ be the factor score of $\mathbf{x}_{ij}$ obtained by the unweighted least squares method with the factor loadings estimated by the principal component method. Then we have

$$\begin{aligned} \mathbf{f}_j &= \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}}\mathbf{e}_1'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \frac{1}{\sqrt{\lambda_2}}\mathbf{e}_2'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \vdots \\ \frac{1}{\sqrt{\lambda_m}}\mathbf{e}_m'(\mathbf{x}_j - \bar{\mathbf{x}}) \end{pmatrix}, \quad j = 1, 2, \ldots, n. \\ &= \boldsymbol{\Lambda}^{-\frac{1}{2}}[\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p]'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ &= \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{y}_j - \boldsymbol{\Lambda}^{-\frac{1}{2}}\bar{\mathbf{y}} \end{aligned}$$

Where $\mathbf{\Lambda}^{-\frac{1}{2}} = \text{diag}(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \ldots, \frac{1}{\sqrt{\lambda_p}})$. Then the within group sums of squares matrix of $\mathbf{f}_j$,

$$\begin{aligned}\tilde{\mathbf{W}} &= \sum_{i=1}^{2}\sum_{j=1}^{n_i}(\mathbf{f}_{ij} - \bar{\mathbf{f}}_i)(\mathbf{f}_{ij} - \bar{\mathbf{f}}_i)' \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)'\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{W}\mathbf{\Lambda}^{-\frac{1}{2}}.\end{aligned}$$

And

$$\tilde{\mathbf{d}} = \bar{\mathbf{f}}_\mathbf{1} - \bar{\mathbf{f}}_\mathbf{2} = \mathbf{\Lambda}^{-\frac{1}{2}}(\bar{\mathbf{y}}_\mathbf{1} - \bar{\mathbf{y}}_\mathbf{2}) = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{d}.$$

Hence, a test of the null hypothesis $H_0 : \boldsymbol{\beta}_2 = 0$ using the sample Mahalanobis distances based on $\mathbf{f}_{ij}$:

$$\begin{aligned}\tilde{D_p}^2 &= m\tilde{\mathbf{d}}'\tilde{\mathbf{W}}^{-1}\tilde{\mathbf{d}} \\ &= m\mathbf{d}'\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{W}^{-1}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{d} \\ &= m\mathbf{d}'\mathbf{W}^{-1}\mathbf{d} \\ &= D_p^2.\end{aligned}$$

Similarly, we can get $\tilde{D_k}^2 = D_k^2$. Hence, we have the statistic

$$\begin{aligned}\tilde{\eta}_k &= \tfrac{(m-p+1)}{(p-k)}c(\tilde{D}_p^2 - \tilde{D}_k^2)/(m + c\tilde{D}_k^2) \\ &= \tfrac{(m-p+1)}{(p-k)}c(D_p^2 - D_k^2)/(m + cD_k^2) \\ &= \eta_k.\end{aligned}$$

That is, if we apply the method of Mardia et al. on principal component values and factor scores, we will get the same order of selection on principal components and factors.

3.2.2 Chang's Method

Let $\mathbf{y}$ be a p dimensional random variable with a mixture of two normal distributions with mean $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, mixing proportions of π and $1 - \pi$, respectively, and a common covariance matrix $\mathbf{\Sigma}$. That is, $\mathbf{y}_{11}, \mathbf{y}_{12}, \ldots, \mathbf{y}_{1n_1} \sim N_p\ (\boldsymbol{\mu}_1, \mathbf{\Sigma})$ and

$\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2n_2} \sim N_p\ (\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. Let $\bar{\mathbf{y}} = \dfrac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\mathbf{y}_{ij}$ and $\mathbf{x}_{ij} = \mathbf{y}_{ij} - \bar{\mathbf{y}}$. Let Δ denote the Mahalanobis between the two sub-populations $\mathbf{Y}_1$ and $\mathbf{Y}_2$, and $\mathbf{V}$ be the covariance matrix of $\mathbf{Y}$. Then, $\mathbf{V} = \pi(1-\pi)\mathbf{d}\mathbf{d}' + \boldsymbol{\Sigma}$ and $\Delta^2 = \mathbf{d}'\boldsymbol{\Sigma}^{-1}\mathbf{d}$, where $\mathbf{d}$=$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2$.

It is known that the effectiveness of using a set of variables can be measured by the Mahalanobis distance computed on the basis of these variables. Let $(\lambda_1, \mathbf{e}_1)$, $(\lambda_2, \mathbf{e}_2)$, ..., $(\lambda_p, \mathbf{e}_p)$ be the eigenvalue-eigenvector pairs of $\mathbf{V}$, where $\mathbf{e}_i'\mathbf{e}_j = 0$ if $i \neq j$ and $\mathbf{e}_i'\mathbf{e}_i = 1$. For a given $m \leq p$, let $B_m = (\frac{1}{\sqrt{\lambda_1}}\mathbf{e}_1, \frac{1}{\sqrt{\lambda_2}}\mathbf{e}_2, \ldots, \frac{1}{\sqrt{\lambda_m}}\mathbf{e}_m)$ and denote Δ_m as the distance between the population using $B_m'\mathbf{y}$. That is, Δ_m is the Mahalanobis distance based on factor scores obtained by the principal component method. Here, we do not assume that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$.

Proposition

$$\Delta_m^2 = \sum_{i=1}^{m}\frac{(\mathbf{e}_i'\mathbf{d})^2}{\lambda_i}/(1-\pi(1-\pi)\sum_{i=1}^{m}\frac{(\mathbf{e}_i'\mathbf{d})^2}{\lambda_i}).$$

In particular, for $m = 1$,

$$\Delta_1^2 = \frac{(\mathbf{e}_1'\mathbf{d})^2}{\lambda_1}/(1-\pi(1-\pi)\frac{(\mathbf{e}_1'\mathbf{d})^2}{\lambda_1}).$$

Proof. The within population covariance matrix of $B_m'\mathbf{Y}$ is clearly $B_m'\boldsymbol{\Sigma}B_m$. The squared distance between the two sub-populations computed on the basis of $B_m'\mathbf{y}$ is therefore given by

$$\begin{aligned}\Delta_m^2 &= (B_m'\mathbf{d})'(B_m'\boldsymbol{\Sigma}B_m)^{-1}(B_m'\mathbf{d})\\ &= \mathbf{d}'B_m(B_m'(\mathbf{V}-\pi(1-\pi)\mathbf{d}\mathbf{d}')B_m)^{-1}B_m'\mathbf{d}.\end{aligned}$$

Now, replace $\mathbf{V}$ by its spectral representation $\mathbf{V} = \sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i'$, we have

$$
\begin{aligned}
\Delta_m^2 =& \mathbf{d}' B_m [(\frac{1}{\sqrt{\lambda_1}}\mathbf{e}_1, \frac{1}{\sqrt{\lambda_2}}\mathbf{e}_2, \ldots, \frac{1}{\sqrt{\lambda_m}}\mathbf{e}_m)'(\sum_{j=1}^{p} \lambda_j \mathbf{e}_j \mathbf{e}_j' - \pi(1-\pi)\mathbf{d}\mathbf{d}') \\
& (\frac{1}{\sqrt{\lambda_1}}\mathbf{e}_1, \frac{1}{\sqrt{\lambda_2}}\mathbf{e}_2, \ldots, \frac{1}{\sqrt{\lambda_m}}\mathbf{e}_m)]^{-1} B_m' \mathbf{d} \\
=& \mathbf{d}' B_m [\mathbf{I}_m - \pi(1-\pi) B_m' \mathbf{d}\mathbf{d}' B_m]^{-1} B_m' \mathbf{d}
\end{aligned}
$$

Let Λ be a diagonal matrix, then by the result of Bartlett(1951):

$$
\begin{aligned}
& [\Lambda - \pi(1-\pi) B_m' \mathbf{d}\mathbf{d}' B_m]^{-1} \\
=& \Lambda^{-1} + (\pi(1-\pi)\Lambda^{-1} B_m' \mathbf{d}\mathbf{d}' B_m \Lambda^{-1})/(1 - \pi(1-\pi)\mathbf{d}' B_m \Lambda^{-1} B_m' \mathbf{d})
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\Delta_m^2 &= \mathbf{d}' B_m I_m^{-1} B_m' \mathbf{d} + [\pi(1-\pi)(\mathbf{d}' B_m I_m^{-1} B_m' \mathbf{d})^2 / (1 - \pi(1-\pi)\mathbf{d}' B_m I_m^{-1} B_m' \mathbf{d})] \\
&= [(\mathbf{d}' B_m B_m' \mathbf{d}) - (\pi(1-\pi) - \pi(1-\pi))(\mathbf{d}' B_m B_m' \mathbf{d})^2] / (1 - \pi(1-\pi)\mathbf{d}' B_m B_m' \mathbf{d}) \\
&= \mathbf{d}' B_m B_m' \mathbf{d} / (1 - \pi(1-\pi)\mathbf{d}' B_m B_m' \mathbf{d}) \\
&= \sum_{i=1}^{m} \frac{(\mathbf{e}_i' \mathbf{d})^2}{\lambda_i} / [1 - \pi(1-\pi) \sum_{i=1}^{m} \frac{(\mathbf{e}_i' \mathbf{d})^2}{\lambda_i}]
\end{aligned}
$$

❑

Here, $B_m \mathbf{Y}$ is the factor score obtained by the unweighted least square procedure when the loadings are estimated by the principal component method. Thus, by this proposition, we have the Mahalanobis distance computed by the ith principal component and the ith factor are the same. Therefore, if we use the ith principal component and the ith factor (obtained by the principal component method) as discriminant variable, we will have the same classification result.