# Relationships between Two Methods for Dealing with Missing Data in Principal Component Analysis

Yoshio Takane, McGill University

Yuriko Oshima-Takane, McGill University

Published in *Behaviormetrika*, **30**, 2003, 145-154. The work reported in this paper has been supported by research grants from the Natural Sciences and Engineering Research Council of Canada to both authors. Requests for reprints should be sent to Yoshio Takane, Department of Psychology, 1205 Dr. Penfiled Avenue, Montréal, QC, H3A 1B1, Canada.

Mailing Address: Yoshio Takane
Department of Psychology, Mcgill Uinversity
1205 Dr. Penfield Ave., Montreal,
QC, H3A 1B1 CANADA
email: takane@takane2.psych.mcgill.ca

**Abstract**

Missing data arise in virtually all practical data analysis situations. The problem of how to deal with them presents a major challenge to many data analysts. A variety of methods have been proposed to deal with missing data. In this paper we discuss two such proposals for principal component analysis (PCA) and investigate their mutual relationships. One was proposed by Shibayama (1988) for test equating (the TE method), and the other is called missing-data-passive (MDP) approach in homogeneity analysis (Meulman, 1982). The two methods are shown to be essentially equivalent despite their different guises.

**Keywords:** test equating (TE), missing-data-passive (MDP), homogeneity criterion, (generalized) eigenvalue decomposition, (generalized) singular value decomposition, generalized prediction problem.

Missing data arise in virtually all practical data analysis situations. The problem of how to deal with them presents a major challenge to many data analysts. A variety of methods have been proposed to deal with missing data. In this paper we discuss two such proposals for principal component analysis (PCA) and investigate their mutual relationships. One is the TE method proposed by Shibayama (1988, 1995), and the other is called missing-data-passive (MDP) method in homogeneity analysis (Meulman, 1982). The two methods are shown to be essentially equivalent despite their different guises.

Shibayama's (1988) method was originally proposed for test equating (which we call the TE method in this paper). In university entrance examinations examinees often have options as to which examinations to take. Since not all examinees take all possible examinations, there will be missing data. Nonetheless, the examiners have to come up with a single set of scores, one for each examinee, that (at least partially) rank order all the examinees. The TE approach was initially devised for such situations. It has turned out (Takane, 1995, pp. 96-98) that this method is easily extensible to multiple sets of scores, yielding a PCA-like method for data with missing values.

The other method is called missing-data-passive (MDP) approach in homogeneity analysis of categorical data (e.g., Meulman, 1982). Such data are often incomplete due to non-responses to certain items by some respondents. In this approach, non-responses are coded as zeroes for all response categories in the items. Missing data then have "minimal" influence on the resultant solution. This approach is often implemented by supplying a weight matrix that indicates which elements of a data matrix are missing, and which are observed (e.g., Gabriel & Zamir, 1979), and then applying a weighted ho-

mogeneity analysis. A similar idea is used in PCA of numerical data (Gifi, 1990), where PCA is formulated as a kind of homogeneity analysis.

The TE method also uses a similar weighting scheme. As a result, both (TE and MDP) methods lead to analytic solutions, making the comparison between them relatively straightforward. The two methods are shown to be essentially equivalent, although there are subtle differences as well. We discuss both commonalities and differences between them.

# 1 Notations

We first summarize the notations we use:

$i$: index for cases; $i = 1, \ldots, n$, where $n$ is the total number of cases.

$j$: index for variables; $j = 1, \ldots, p$, where $p$ is the total number of variables.

$k$: index for components; $k = 1, \ldots, t$, where $t$ is the total number of components.

$X = [x_{ij}]$: the $n$ by $p$ matrix of raw data, whose $ij^{th}$ element (i.e., the element in the $i^{th}$ row and the $j^{th}$ column) is denoted by $x_{ij}$. The $i^{th}$ row vector of $X$ is denoted by $x_i'$, and the $j^{th}$ column vector is denoted by $x_j$.

$Z = [z_{ij}]$: the $n$ by $p$ matrix of column-wise centered data, whose $ij^{th}$ element is denoted by $z_{ij}$. The $i^{th}$ row vector of $Z$ is denoted by $z_i'$, and the $j^{th}$ column vector by $z_j$. The centering is done with respect to only observed (non-missing) data; i.e., $z_j = (I_n - 1_n(1_n' D_{w_j} 1_n)^{-1} 1_n' D_{w_j}) x_j$,

where $I_n$, $1_n$, and $D_{w_j}$ are as will be defined below. (In both $X$ and $Z$, missing entries can be arbitrary.)

$W = [w_{ij}]$: the $n$ by $p$ matrix of weights whose $ij^{th}$ element, $w_{ij}$, indicates whether $x_{ij}$ is observed ($w_{ij} = 1$), or missing ($w_{ij} = 0$). The $i^{th}$ row vector of $W$ is denoted by $w_i'$, and the $j^{th}$ column vector is denoted by $w_j$.

$D_{x_i'}$: the diagonal matrix of order $p$ with the elements of $x_i'$ as its diagonal elements.

$D_{z_i'}$: This matrix is similar to $D_{x_i'}$, with $x_i'$ replaced by $z_i'$.

$D_{w_i'}$: the diagonal matrix of order $p$ with the elements of $w_i'$ as its diagonal elements.

$D_{w_j}$: the diagonal matrix of order $n$ with the elements of $w_j$ as its diagonal elements.

$D_W$: $= \sum_{j=1}^{p} D_{w_j}$.

$S = \text{diag}(s_{jj})$: the diagonal matrix of variations with $s_{jj} = z_j' D_{w_j} z_j$ as the $j^{th}$ diagonal element.

$V = [v_{jk}]$: a $p$ by $t$ matrix of weights (scaling parameters) applied to $x_{ij}$ to derive a matrix that can be re-scaled into the matrix of component scores in the TE method. This matrix is constrained to satisfy

$$V'SV = I_t. \tag{1}$$

The $j^{th}$ row vector of $V$ is denoted by $v_j'$.

$V_0 = [v_{0jk}]$: a $p$ by $t$ matrix of additive constants applied to $x_{ij}$ in the TE method. This matrix is constrained to satisfy

$$1_p' V_0 = 0_t'. \tag{2}$$

The $j^{th}$ row vector of $V_0$ is denoted by $v_{0j}'$.

$G$: an $n$ by $t$ matrix that can be scaled into component scores in the TE method. The $i^{th}$ row vector of $G$ is denoted by $g_i'$.

$U$: a $p$ by $t$ matrix of weights (scaling factors) applied to $z_{ij}$ to derive a matrix that can be re-scaled into the matrix of component scores in the MDP method (analogous to $V$ in the TE method). The $j^{th}$ row vector of $U$ is denoted by $u_j'$.

$U_0$: a $p$ by $t$ matrix of additive constants when the raw data are used in the MDP method. The $j^{th}$ row vector of $U_0$ is denoted by $u_{0j}'$.

$F$: an $n$ by $t$ matrix that can be scaled into component scores in the MDP method (analogous to $G$ in the TE method). Matrix $F$ is subject to the normalization restriction

$$F' D_W F = I_t. \tag{3}$$

$I_n$ and $I_p$: identity matrices of orders $n$ and $p$, respectively.

$1_n$, $1_p$, and $1_t$: $n$-, $p$-, and $t$-component vectors of ones, respectively.

${}_n0_t$ and ${}_p0_t$: $n$ by $t$ and $p$ by $t$ zero matrices, respectively.

$0_n$ and $0_t$: $n$- and $t$-component zero vectors, respectively.

Other symbols will be introduced as they become necessary.

# 2 The Two Methods

## 2.1 The TE method

In this section we discuss the TE method. This method was originally proposed for $t = 1$ (Shibayama, 1988, 1995). For the purpose of the present paper, however, we discuss the general case in which $t \geq 1$.

Define a $p$ by $t$ matrix,

$$Y_i = D_{x_i'} V + V_0, \tag{4}$$

and the criterion

$$\phi_1(V, V_0, G) = \sum_{i=1}^{n} \mathrm{SS}(Y_i - 1_p g_i')_{D_{w_i'}}, \tag{5}$$

where $\mathrm{SS}(A)_B$, the sum of squares of $A$ under the metric matrix, $B$ (assumed non-negative definite), is defined to be $\mathrm{tr}(A'BA)$. We consider minimizing $\phi_1$ with respect to $V$, $V_0$, and $G$ subject to the constraints, (1) and (2). There are some indeterminacies in (5). Adding $1_p c'$ to $V_0$, where $c'$ is an arbitrary $t$-component row vector, can be offset by adding $c'$ to $g_i'$ for all $i$'s. Constraint (2) removes this indeterminacy. It should also be noticed that (5) can trivially be minimized by setting $V = {}_p0_t$, $V_0 = {}_p0_t$, and $G = {}_n0_t$. To avoid this trivial solution, we impose constraint (1).

We use a conditional minimization strategy to minimize $\phi_1$. We first minimize $\phi_1$ with respect to $G$ for given $V$ and $V_0$, then with respect to $V_0$ for given $V$, and finally with respect to $V$. This is based on the following relationship:

$$\min_{V,V_0,G} \phi_1(V, V_0, G) = \min_{V} [\min_{V_0|V} \{ \min_{G|V,V_0} \phi_1(V, V_0, G) \}]. \tag{6}$$

Minimizing $\phi_1(V, V_0, G)$ with respect to $G$ conditional on $V$ and $V_0$ leads to

$$\hat{g}_i' = (1_p' D_{w_i'} 1_p)^{-1} 1_p' D_{w_i'} Y_i, \tag{7}$$

for $i = 1, \ldots, n$. By putting this estimate of $g_i'$ into $\phi_1(V, V_0, G)$, we obtain

$$\min_{G|V,V_0} \phi_1(V, V_0, G) = \phi_1(V, V_0, \hat{G}) = \sum_i \mathrm{tr}(Y_i' D_{w_i'} Q_{1_p/D_{w_i'}} Y_i), \tag{8}$$

where $Q_{1_p/D_{w_i'}} = I_p - 1_p(1_p' D_{w_i'} 1_p)^{-1} 1_p' D_{w_i'}$. Let $C_i = D_{w_i'} Q_{1_p/D_{w_i'}}$. Then, (8) can be rewritten as

$$\phi_1(V, V_0, \hat{G}) = \mathrm{tr}(V' A_1 V) + 2\mathrm{tr}(V' A_2 V_0) + \mathrm{tr}(V_0' A_3 V_0), \tag{9}$$

where

$$A_1 = \sum_i D_{x_i'} C_i D_{x_i'}, \tag{10}$$

$$A_2 = \sum_i D_{x_i'} C_i, \tag{11}$$

and

$$A_3 = \sum_i C_i. \tag{12}$$

Minimizing $\phi_1(V, V_0, \hat{G})$ with respect to $V_0$ conditional on $V$ leads to

$$\hat{V}_0 = -A_3^+ A_2' V, \tag{13}$$

where $A_3^+$ indicates the Moore-Penrose inverse of $A_3$. Note that although constraint (2) was not explicitly imposed on $\hat{V}_0$, it can be readily verified that the use of $A_3^+$ (among all possible g-inverses of $A_3$) in (13) guarantees that it is satisfied. By putting this estimate of $V_0$ into $\phi_1(V, V_0, \hat{G})$, we obtain

$$\min_{V_0|V} \phi_1(V, V_0, \hat{G}) = \phi_1(V, \hat{V}_0, \hat{G}) = \mathrm{tr}(V' A V), \tag{14}$$

where

$$A = A_1 - A_2 A_3^+ A_2'. \tag{15}$$

Minimizing $\phi_1(V, \hat{V}_0, \hat{G})$ with respect to $V$ subject to (1) leads to the following generalized eigen-equation:

$$AV = SV\tilde{\Delta}^2, \tag{16}$$

where $\tilde{\Delta}^2$ is the diagonal matrix of order $t$ with the $t$ smallest eigenvalues of $A$ with respect to $S$ as its diagonal elements. This generalized eigen-equation can be solved by first solving the usual eigen-equation,

$$S^{-1/2} A S^{-1/2} V^* = V^* \tilde{\Delta}^2, \tag{17}$$

where $V^* = S^{1/2}V$. Here, $S^{1/2}$ and $S^{-1/2}$ are square root factors of $S$ and $S^{-1}$, respectively. The estimate of $V$ is then obtained by

$$\hat{V} = S^{-1/2} V^*. \tag{18}$$

Once $\hat{V}$ is obtained, $\hat{V}_0$ is obtained by (13) (by replacing $V$ by its estimate). Similarly, $\hat{g}_i'$ is obtained by (7) by replacing $V$ and $V_0$ by their estimates.

The method described above reduces to the usual PCA of a standardized data matrix, when there are no missing data. This can be shown as follows: We have $D_{w_i'} = I_p$, so that $C_i = I_p - 1_p 1_p'/p \equiv Q_p$. We have

$$A_1 = D_{h^2} - X'X/p,$$

where $D_{h^2} = \sum_i D_{x_i'}^2$,

$$A_2 = D_h - h1_p'/p,$$

where $h = \sum_i x_i$, and $D_h$ is the diagonal matrix with $h$ as its diagonal elements, and

$$A_3 = nQ_p.$$

Since $A_3^+ = Q_p/n$, we obtain $A_2 A_3^+ A_2' = (1/n)(D_h - h1_p'/p)Q_p(D_h - 1_p h'/p) = (1/n)(D_h^2 - hh'/p - hh'/p + h1_p'1_p h'/p^2) = (1/n)(D_h^2 - hh'/p)$. Thus, $A = D_{h^2} - X'X/p - D_h^2/n + hh'/np = (D_{h^2} - D_h^2/n) + (X'X - hh'/n)/p = S - Z'Z/p = S - C/p$, where $C = Z'Z$ (the sum of squares and cross-products matrix formed from the column-wise centered data matrix $Z$). Note that $S = \mathrm{diag}(C)$, and

$$S^{-1/2}(S - C/p)S^{-1/2} = I - R/p,$$

where $R$ is the correlation matrix. The eigenvectors of this matrix are identical to those of $R$, and the eigenvalues are equal to one minus the eigenvalues of $R$. The largest $t$ eigenvalues of $R$ thus correspond with the $t$ smallest eigenvalues of $S^{-1/2}(S - C/p)S^{-1/2}$.

## 2.2 The MDP method

In the missing-data-passive (MDP) approach in homogeneity analysis, we minimize

$$\phi_2(F, U) = \sum_{j=1}^{p} \mathrm{SS}(F - z_j u_j')_{D_{w_j}} \tag{19}$$

with respect to $U$ and $F$ subject to the normalization restriction, (3). In this approach the normalization restriction is placed on $F$ instead of $V$ to avoid the trivial solution (in which $U = {}_p0_t$, and $F = {}_n0_t$). We minimize $\phi_2$ by first minimizing it with respect to $U$ for given $F$, and then minimizing it

with respect to $F$ subject to (3). This minimization strategy is based on the following relationship:

$$\min_{F,U} \phi_2(F,U) = \min_{F}\{\min_{U|F} \phi_2(F,U)\}. \tag{20}$$

Minimizing $\phi_2(F,U)$ with respect to $U$ conditional on $F$ leads to

$$\hat{u}_j' = (z_j' D_{w_j} z_j)^{-1} z_j' D_{w_j} F, \tag{21}$$

for $j = 1, \ldots, p$. By putting this estimate of $U$ into $\phi_2(F,U)$, we obtain

$$\begin{aligned} \min_{U|F} \phi_2(F,U) &= \phi_2(F,\hat{U}) \\ &= \operatorname{tr}(F'(D_W - \sum_j D_{w_j} z_j (z_j' D_{w_j} z_j)^{-1} z_j' D_{w_j}) F). \end{aligned} \tag{22}$$

Minimizing $\phi_2(F,\hat{U})$ with respect to $F$ subject to (3) is equivalent to maximizing

$$\phi_2^* = \operatorname{tr}(F'(\sum_j D_{w_j} P_{z_j/D_{w_j}})F) = \operatorname{tr}(F'PF) \tag{23}$$

subject to the same restriction, where

$$P = \sum_j D_{w_j} P_{z_j/D_{w_j}}, \tag{24}$$

and

$$P_{z_j/D_{w_j}} = z_j (z_j' D_{w_j} z_j)^{-1} z_j' D_{w_j}. \tag{25}$$

This amounts to $\text{GSVD}(D_W^{-1} Z^* S^{-1})_{D_W,S}$, the generalized singular value decomposition of $D_W^{-1} Z^* S^{-1}$ with row metric matrix $D_W$, and column metric matrix $S$, where

$$Z^* = \left[ D_{w_1} z_1, \ldots, D_{w_p} z_p \right]. \tag{26}$$

This GSVD, in turn, is obtained by the following procedure: Let $D_W^{-1/2} Z^* S^{-1/2} = \tilde{F} \Delta \tilde{U}'$ denote the usual singular value decomposition of

$D_W^{-1/2} Z^* S^{-1/2}$ (denoted by SVD($D_W^{-1/2} Z^* S^{-1/2}$)). Then, GSVD$(D_W^{-1} Z^* S^{-1})_{D_W,S}$ is obtained by

$$D_W^{-1} Z^* S^{-1} = D_W^{-1/2} \tilde{F} \Delta \tilde{U}' S^{-1/2} = \hat{F} \Delta U^{*\prime}, \tag{27}$$

where $\hat{F} = D_W^{-1/2} \tilde{F}$, and $U^* = S^{-1/2} \tilde{U}$. Matrix $U^*$ is related to the estimate of $U$ in (21) by $\hat{U} = U^* \Delta$. Matrix $\hat{F}$ satisfies $1_n' D_W \hat{F} = 0_t'$. This follows from the fact that the column-wise centered data are used in defining (19).

When there are no missing data, we have $D_{w_j} = I_n$ for all $j$'s, and $D_W = pI_n$. Then, GSVD$(D_W^{-1} Z^* S^{-1})_{D_W,S}$ reduces to GSVD$(Z^* S^{-1})_{I_n,S}$, which is essentially equivalent to PCA of the standardized data matrix, $Z^* S^{-1/2}$.

# 3 Relationships between the Two Methods

Both the TE method and the MDP approach use weights to indicate missing data. Both methods also use homogeneity criteria, although the homogeneity is measured case-wise (row-wise) in the TE method, while it is done variable-wise (column-wise) in the MDP method. This makes the two methods look more distinct than real. A similarity between them is more obvious, if we write the two criteria, (5) and (19), using element-wise notations:

$$\phi_1 = \sum_i \sum_{j,k} w_{ij} (x_{ij} v_{jk} + v_{0jk} - g_{ik})^2, \tag{28}$$

and

$$\phi_2 = \sum_j \sum_{i,k} w_{ij} (z_{ij} u_{jk} - f_{ik})^2. \tag{29}$$

In both cases, quantities associated with three subscripts, $i$, $j$, and $k$, say $y_{ijk}$ (i.e., $x_{ij} v_{jk} + v_{0jk}$ in the TE method, and $z_{ij} u_{jk}$ in the MDP method),

are made as close as possible to quantities with only two subscripts, $i$ and $k$, (i.e., $g_{ik}$ in the TE method, and $f_{ik}$ in the MDP method). That is, $y_{ijk}$ is made as homogeneous as possible across $j$. This is the essence of homogeneity criteria. The most conspicuous difference between the two criteria is in the preprocessing of data. Whereas in the TE method the raw data are used in defining the homogeneity criterion, the column-wise centered data are used in the MDP method. As will be shown, this difference has some differential consequence. A difference also exists in normalization convention. The normalization is imposed on $V$ (the component loadings side) in the TE method, while it is on $F$ (the component scores side) in the MDP approach. This difference, however, has only minor consequence.

## 3.1 Column-wise centered data in the TE method

Suppose we use the column-wise centered data and require $V_0 = 0$. (The latter seems justifiable, since this quantity is basically for discounting the differential means across variables in the raw data. No analogous quantity is used in the MDP method where the column-wise centered data are used in defining the homogeneity criterion.) Criterion (5) then becomes

$$\phi_1(V, G) = \sum_i \mathrm{SS}(Y_i - 1_p g_i')_{D_{w_i'}}, \tag{30}$$

where $Y_i = D_{z_i'} V$. The estimate of $g_i'$ is obtained by (7) as before, but because of $V_0 = 0$, the estimation of $V$ is much simplified. It reduces to the generalized eigenvalue decomposition of $A = A_1$ with respect to $S$, i.e., $A_1 V = SV\Delta$, which in turn reduces to the usual eigenvalue decomposition of $S^{-1/2} A_1 S^{-1/2}$. Since $A_1 = \sum_i D_{z_i'} C_i D_{z_i'} = \sum_i D_{z_i'}^2 D_{w_i'} - \sum_i D_{z_i'} D_{w_i'} 1_p (1_p' D_{w_i'} 1_p)^{-1} 1_p' D_{w_i'} D_{z_i'}$,

$S^{-1/2}A_1S^{-1/2} = I_p - S^{-1/2}Z^{*\prime}D_W^{-1}Z^*S^{-1/2}/p$, where $Z^*$ is as defined in (26). This matrix has the same set of eigenvectors as $S^{-1/2}Z^{*\prime}D_W^{-1}Z^*S^{-1/2}/p$. These eigenvectors also coincide with the right singular vectors, $\tilde{U}$, of $D_W^{-1/2}Z^*S^{-1/2}$ in the MDP method. Thus, $V = S^{-1/2}\tilde{U} = U^* = \hat{U}\Delta^{-1}$. This shows that the TE method gives solutions equivalent to the MDP method if the column-wise centered data are used in defining the criterion and if $V_0 = 0$ is enforced. This $V$, however, is different from that in the original formulation of the TE method described in Section 2.1. (This point will be clearer in the next section.)

What happens, however, if the column-wise centered data are used, but $V_0 = 0$ is not enforced, in the TE method? In this case the same estimates of $V$ and $G$ are obtained as in the original TE method in which the raw data are used. The difference between the column-wise centered data and the raw data is absorbed in the difference in the estimates of $V_0$. The estimate of $V_0$ is typically nonzero (unless it is enforced to be zero) even when the column-wise centered data are used. Thus, the critical difference seems to lie in whether or not we enforce $V_0 = 0$ when the column-wise centered data are used in the TE method.

## 3.2 Raw data in the MDP method

In the previous section we saw what would happen if we used the column-wise centered data and enforced $V_0 = 0$ in the TE method, which originally used the raw data in defining the criterion, (5). What about the reverse? That is, what will happen if we use the raw data in the MDP method, which originally used the column-wise centered data?

The criterion in this case becomes

$$\phi_2(F, U, U_0) = \sum_j \mathrm{SS}(F - x_j u_j' - 1_n u_{0j}')_{D_{w_j}}. \tag{31}$$

(We needed to introduce a quantity, $U_0$, which is analogous to $V_0$ in the TE method, to account for the differential means across variables.) As in (5), there is a trade-off between adding $1_n c'$ to $F$, where $c'$ is an arbitrary $t$-component row vector, and adding $c'$ to $u_{0j}'$ for all $j$'s. To remove this indeterminacy, we require

$$1_n' F = 0_t'. \tag{32}$$

Minimizing $\phi_2$ with respect to $u_{0j}$ for given $F$ and $u_j$ leads to

$$\hat{u}_{0j}' = (1_n' D_{w_j} 1_n)^{-1} 1_n' D_{w_j} (F - x_j u_j') \tag{33}$$

for $j = 1, \ldots, p$. Define $Q_{1_n/D_{w_j}} = I_n - 1_n (1_n' D_{w_j} 1_n)^{-1} 1_n' D_{w_j}$. Then,

$$\begin{aligned} \min_{U_0|F,U} \phi_2(F, U, U_0) &= \phi_2(F, U, \hat{U}_0) = \sum_j \mathrm{SS}(Q_{1_n/D_{w_j}}(F - x_j u_j'))_{D_{w_j}} \\ &= \sum_j \mathrm{SS}(F - x_j u_j')_{Q_j}, \end{aligned} \tag{34}$$

where $Q_j = D_{w_j} Q_{1_n/D_{w_j}}$.

Minimizing $\phi_2(F, U, \hat{U}_0)$ with respect to $u_j'$ for given $F$ leads to

$$\hat{u}_j' = (x_j' Q_j x_j)^{-1} x_j' Q_j F \tag{35}$$

for $j = 1, \ldots, p$. Then,

$$\min_{U|F} \phi_2(F, U, \hat{U}_0) = \phi_2(F, \hat{U}, \hat{U}_0) = \mathrm{tr}(F'(Q - \sum_j Q_j P_{x_j/Q_j}) F), \tag{36}$$

where $Q = \sum_j Q_j$, and $P_{x_j/Q_j} = x_j (x_j' Q_j x_j)^{-1} x_j' Q_j$.

Minimizing $\phi_2(F, \hat{U}, \hat{U}_0)$ with respect to $F$ subject to $F'QF = I_t$ is equivalent to maximizing $\mathrm{tr}(F'(\sum_j Q_j P_{x_j/Q_j})F)$ subject to the same restriction. This amounts to the GSVD of $Q^+ Z^* S^{-1}$ with the row metric matrix, $Q$, and the column metric matrix, $S$, where

$$Z^* = [Q_1 x_1, \ldots, Q_p x_p], \tag{37}$$

and $S = \mathrm{diag}(Z^{*\prime} Z^*)$. This $Z^*$ is the same $Z^*$ defined in (26). Note that $\sum_j Q_j P_{x_j/Q_j} = Z^* S^{-1} Z^{*\prime}$. Note that $Q_j x_j = D_{w_j} z_j$. Note also that $Q_j 1_n = 0_n$, and $Q 1_n = 0_n$. This guarantees that $\hat{F}$ obtained as the set of left singular vectors from $\mathrm{GSVD}(Q^+ Z^* S^{-1})_{Q,S}$ satisfies constraint (32).

This $\hat{F}$ is related to the estimate of $G$ obtained by the original TE method (i.e., by (7)) in a simple way. Specifically, $\hat{F} = Q_{1_n} \hat{G} \Delta^{-1}$, where $Q_{1_n} = I_n - 1_n 1_n'/n$ is the centering operator (of order $n$), and $\Delta$ is the diagonal matrix of singular values obtained in the above GSVD. However, this $\hat{F}$ is not related in a simple way to the estimate of $F$ obtained from the column-wise centered data under the implicit assumption that $U_0 = 0$ in Section 2.2. This can be seen by noting that the above $\hat{F}$ was obtained under the normalization restriction that $F'QF = I_t$, whereas the $\hat{F}$ in Section 2.2 was obtained under the restriction that $F' D_W F = I_t$. The critical difference arises from whether or not we enforce $U_0 = 0$.

# 4 Discussion

We have seen some interesting relationships between two methods for dealing with missing data in PCA, one proposed by Shibayama (1988) for test

equating (the TE method), and the other called missing-data-passive approach (the MDP method) in homogeneity analysis (Gifi, 1990). They are shown to be essentially equivalent, although there are subtle differences as well. We have found a way to make one method to work exactly like the other. The difference between the two methods is due to whether the raw data or the column-wise data are used in defining the homogeneity criterion, and when the latter is used, whether or not we enforce $V_0 = 0$ in the TE method, or $U_0 = 0$ in the MDP method. When the raw data are used, enforcing $V_0 = 0$ or $U_0 = 0$ is clearly unjustifiable. It is assumed justifiable when the column-wise centered data are used in the MDP method. However, is it really justifiable? When the data are centered in the presence of missing data, it is done with respect to the means of the observed portions of the data, and if those means reflect the means to be obtained when the data were complete, this assumption is perhaps justifiable. Such may be the case, when the missing data occur at random. However, this may not be true in general. In entrance examinations, for example, applicants tend to choose the subjects which they are good at, so that the means of the observed data could be substantially higher than those that would have been obtained if all the examinees took the examinations. In such cases it is better to allow for nonzero $V_0$ (or $U_0$) whether or not the data are column-wise centered.

It should be pointed out that there are other approaches to missing data in PCA. Gabriel and Zamir (1979), and Kiers (1997), for example, proposed to minimize

$$\phi_3 = \sum_{i,j} w_{ij}(x_{ij} - \hat{x}_{ij})^2 \tag{38}$$

subject to the restriction that $\text{rank}([\hat{x}_{ij}]) = t$. In the terminology of Gifi

(1990), however, this minimization problem belongs to the "join" loss problem as opposed to the "meet" loss problem, of which homogeneity analysis is a special case. Both of the two methods we discussed in this paper belong to the latter approach. While there is a simple relationship between the join loss and the meet loss when there are no missing data (Gifi, 1990, p. 168), no such relationship exists when there are missing data. The methods based on the join loss require iterative solutions. That the two methods discussed in this paper lead to analytic solutions can be considered as an advantage over those methods that require iterative solutions.

Missing data are often considered nuisance in data analysis. In a wider perspective, however, missing data problems present interesting perspectives to statistics. They can be viewed as generalized forms of prediction problems. In regression analysis the variable on which missing data exist is always the dependent variable (assuming that no other missing data exist), and we predict values on the dependent variable. In the general missing data problems, missing data can occur on any variable or variables. Variables on which missing data occur can also vary across different observation units. Our task is to predict values of missing data based on the information from observed data. This problem is called by different names in different disciplines: It is called data imputation problem for missing data in statistics (Little & Rubin, 1987), matrix completion problem in mathematics (matrix algebra), pattern completion problem in neural network literature, etc.

## References

Gabriel, K. R. & Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights, *Technometrics*, **21**, 489-498.

Kiers, H. A. L. (1997). Weighted least squares fitting using iterative ordinary least squares algorithms, *Psychometrika*, **62**, 251-266.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: Wiley.

Little, R. A. & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Meulman, J. (1982). *Homogeneity analysis of incomplete data*. Leiden: DSWO Press.

Shibayama, T. (1988). *Kessokuchi o fukumu tesuto sukoa no tahenryōkaiseki* (Multivariate analysis of test scores with missing data). Unpublished Doctoral Dissertation, Faculty of Education, University of Tokyo (in Japanese).

Shibayama, T. (1995). A linear composite method for test scores with missing values, *Niigata daigaku kyōikugakubu kiyō* (Memoirs of the Faculty of Education, Niigata University), **36**, 445-455.

Takane, Y. (1995). *Seiyakutsuki shuseibunbunsekihō* (Constrained principal component analysis). Tokyo: Asakurashoten, (in Japanese).