筑波大学体育科学系紀要 Bull. Inst. Health & Sport Sciences, Univ. of Tsukuba 16 : 145－151, 1993.

同時多変量回帰主成分分析

稲垣　敦・松浦義行

Simultaneous Multivariate Regression Principal Component Analysis, SMRPCA

Atsushi INAGAKI and Yoshiyuki MATSUURA

本研究では，同時多変量回帰主成分分析，SMRPCA（Simultaneous Multivariate Regression Principal Component Analysis）のモデルとこれをカノニカルデータに拡張したモデル及びこれらのモデルをデータに適合するための交互最小二乗アルゴリズムを提案する。SMRPCA のモデルは，

$$\mathbf{Y} = \tilde{\mathbf{X}}\tilde{\mathbf{A}}\mathbf{B} + \mathbf{E}$$
$$= \tilde{\mathbf{X}}\tilde{\mathbf{W}} + \mathbf{E}$$
$$= \tilde{\mathbf{F}}\mathbf{B} + \mathbf{E}$$

で表される。ここで

$$\mathbf{F}'\mathbf{F} = \mathrm{N}\mathbf{I}_t,$$
$$\mathbf{B}_i\mathbf{B}_i' = \mathbf{I}_t,$$

である。但し，

$$\mathbf{Y} = (\mathbf{Y}_1'\mathbf{Y}_2' \cdots \mathbf{Y}_I')',$$

$$\tilde{\mathbf{X}} = \sum_{i=1}^{I} \mathbf{E}_{ii} \otimes \mathbf{X}_i,$$

$$\tilde{\mathbf{F}} = \sum_{i=1}^{I} \mathbf{E}_{ii} \otimes \mathbf{F}_i,$$

$$\mathbf{F} = (\mathbf{F}_1'\mathbf{F}_2' \cdots \mathbf{F}_I')',$$
$$\tilde{\mathbf{A}} = \mathbf{I}_I \otimes \mathbf{A},$$
$$\mathbf{B} = (\mathbf{B}_1'\mathbf{B}_2' \cdots \mathbf{B}_I')',$$
$$\mathbf{E} = (\mathbf{E}_1'\mathbf{E}_2' \cdots \mathbf{E}_I')',$$

$$\mathrm{N} = \sum_{i=1}^{I} \mathrm{N}_i,$$

$$\mathrm{n} = \mathrm{m} = \mathrm{t}$$

である。上式で，$\mathbf{Y}_i$ と $\mathbf{X}_i$ は第 i 母集団の従属変数行列（$N_i \times m$）と独立変数行列（$N_i \times n$），$\mathbf{A}_s$ は各母集団に共通な成分得点係数行数（$n \times r$），$\mathbf{B}_i$ は各母集団における回帰係数行列（$r \times m$），$\mathbf{E}_i$ は第 i 母集団における残差行列（$N_i \times n$）である。しかし，実際には，各母集団からの標本が用いられるので，$\mathbf{A}_s$ と $\mathbf{B}_i$ はその推定量となる。このモデルでは，従属変数群の一次関数からなる直交成分（潜在因子）を仮定してこれを抽出し，この成分に母集団ごとに異なった回帰係数を与えて従属（基準）変数群を説明しようとするモデルである。したがって，回帰係数や成分得点をプロットすることにより母集団差異を空間的に表現でき，各母集団の特徴や差異の理解を助ける。また，データ解析のモデルとしては，複数の時点や異なる条件におけるデータにも適用可能なので，縦断的研究や多変量時系列データが得られる実験データの分析

に適用可能である。さらに，ALSOS（Alternating Least Squares approach to Optimal Scaling ; Young et al. , 1981）を利用することにより，比率・間隔尺度で推測されたデータのみならず，順序・名義尺度水準，及びこれらが混合したデータに適用できるように容易に拡張できる。以上の点から，このモデルは予測・推定よりもむしろ潜在的な線形構造を記述するのに有効であると考えられるので，体育学だけではなく，社会学や心理学などで利用可能であると考えられる。このモデルをデータに適用するには，交互最小二乗法（Alternating Least Square technique）を用いる。このアルゴリズムは，目的（誤差）関数を単調に減少させて極小に収束させることができるが，必ずしも最小に到達するとは限らないので，いくつかの初期値を用いて解析を試みる必要がある。

key words : multivariate regression analysis, principal component analysis, population differecnes, alternating least square technique

Introduction

Investigating the relationships between variables is a favorite research activity of social scientist. They often want to explore the structure of a large body of data. To understand this organization, the data have to be condensed in one way or another, and the raw data have to be combined to form summary measures which are more easily comprehended. Among the most popular methods to achieve such condensation and sammarization are principal component analysis. Standard principal component analysis is applied when observations are available for many variables, and it is desired to condense these variables to a smaller amount of independent latent variables or components. In many research designs, observations on variables have been made under a few conditions, or at various points in time, or from some populations, etc. In such cases, the data can be classified by three kinds of modes, e. g., subjects, variables and conditions. Three-mode factor analysis by Tucker (6) has been developed to summarize and condense three-mode data.

This article presents the development of "simultaneous Multivariate Regression Principal Component Analysis (SMRPCA)". This model can operate on N-sets of data from different populations, different time points or different conditions, and can represent the differences using multidimensional vector space. Moreover, this method can determine the orthogonal components so as to the variance of dependent (or criterion) variables accounted for by the components underlying independent variables is maximized. This model is characterized as follows :

1) Principal components--the proposed method can determine components which can account for the most variance of dependent variables.
2) Multidimensional spatial representation--the proposed model can represent both populations (conditions or time points) and predictive variables as vectors in multidimensional space.
3) Time series data--the proposed methods will be able to apply the multivariate time series data.
4) Level of measurement--the proposed model can be applied to data which was measured at ratio, interval, ordinal, nominal level, and mixture of two or more measurement levels.
5) Alternating least squares technique--the proposed model is fitted to data using alternating least squares technique.

In the next section, we will present a detailed account of PRDMRA model, focussing on the mathematical formulation.

The Model

We use bold-face capital letters to represent matrices ($\mathbf{X}$) ; bold-face lower case letters for vector ($\mathbf{x}$) ; and regular lower case letters for scalars (x). Note that all vectors are assumed to be column vectors, so a row vector is denoted as transpose of a column vector ($\mathbf{x}'$). We refer to a specific column vector of a matrix as $\mathbf{x}_j$, and a specific element of a matrix as x_{ij}. For convenience, the notation used in this paper is presented as follows :

$i, =1,\cdots,$ I populations, conditions, or time points,

$j, =1,\cdots,$ n independent variables,

$k, =1,\cdots,$ m dependent variables,

$s, =1,\cdots,$ t components in a principal component analysis context,

N_i = number of samples from i-th population,

N = total number of samples $(=\Sigma N_i)$,

$\mathbf{Y}_i$ = matrix of dependent variable of i-th population, condition, or time point ($N_i \times m$), which is measured at more than interval measurement level, and is columnwisely standardized across populations,

$\mathbf{X}_i$ = matrix of independent variables of i-th population, condition, or time point ($N_i \times n$), which is measured at more than interval measurement level, and is columnwisely standardized across populations,

$\mathbf{F}_i$ = matrix of principal component scores of i-th population, condition or time point ($N_i \times n$), which is columnwisely standardized across populations,

$\mathbf{A}$ = matrix of component score coefficients ($n \times t$),

$\mathbf{B}_i$ = matrix of regression weight of $\mathbf{F}_i$ ($t \times m$),

$\mathbf{E}_i$ = matrix of residual of i-th population ($N_i \times m$).

Using the above definitions, we can represent SMRPCA model by matrix form ;

$$\begin{aligned}\mathbf{Y} &= \tilde{\mathbf{X}}\tilde{\mathbf{A}}\mathbf{B} + \mathbf{E},\\ &= \tilde{\mathbf{X}}\tilde{\mathbf{W}} + \mathbf{E},\\ &= \tilde{\mathbf{F}}\mathbf{B} + \mathbf{E},\end{aligned}$$

where

$$\mathbf{Y} = (\mathbf{Y}_1'\mathbf{Y}_2'\cdots\mathbf{Y}_I')',$$

$$\tilde{\mathbf{X}} = \sum_{i=1}^{I}\mathbf{E}_{ii}\otimes\mathbf{X}_i,$$

$$\tilde{\mathbf{F}} = \sum_{i=1}^{I}\mathbf{E}_{ii}\otimes\mathbf{F}_i,$$

$$\tilde{\mathbf{A}} = \mathbf{I}_I\otimes\mathbf{A},$$

$$\mathbf{B} = (\mathbf{B}_1'\mathbf{B}_2'\cdots\mathbf{B}_I')',$$

$$\mathbf{E} = (\mathbf{E}_1'\mathbf{E}_2'\cdots\mathbf{E}_I')',$$

where the notation $(\mathbf{X}\otimes\mathbf{Y})$ refers to right Kronecker product of matrices, $(\mathbf{X}\otimes\mathbf{Y}) = [x_{ij}\mathbf{Y}]$, and where $\mathbf{E}_{ij}$ denotes a matrix with the unit scalar in the (i,j) position, and with zeros elsewhere. Additionally, the equality constraints are imposed on the model,

$$\mathbf{F}'\mathbf{F} = N\mathbf{I}_n,$$

$$\mathbf{B}_i\mathbf{B}_i' = \mathbf{I}_t,$$

$$n = m = t,$$

where

$$\mathbf{F} = (\mathbf{F}_1'\mathbf{F}_2'\cdots\mathbf{F}_I')',$$

$$N = \sum_{i=1}^{I}N_i,$$

These constraints are required to solve the minimization problem which will be presented in the next section.

Estimation

As in many multivariate analyses, we define a squared loss function,

$$\begin{aligned}f(\mathbf{A},\ \mathbf{B}) &= \|\mathbf{Y}-\hat{\mathbf{Y}}\|^2,\\ &= \sum_{i=1}^{I}\|\mathbf{Y}_i-\hat{\mathbf{Y}}_i\|^2,\end{aligned}$$

where $\|\cdot\|$ denotes an Euclidean norm, and $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_i$ denote the estimated $\mathbf{Y}$ and $\mathbf{Y}_i$, respectively. We then search for the best solution such that f $(\mathbf{A},\ \mathbf{B})$ is minimal. As mentioned before, the minimization has to be carried out with the constraints,

$$\mathbf{F}'\mathbf{F} = N\mathbf{I}_n,$$

$$\mathbf{B}_i\mathbf{B}_i' = \mathbf{I}_t,$$

$$n = m = t.$$

To solve the problem, let us rewrite the loss function as follows ;

$$\begin{aligned}f(\mathbf{A},\ \mathbf{B}) &= \|\mathbf{Y}-\hat{\mathbf{X}}\hat{\mathbf{A}}\mathbf{B}\|^2.\\ &= \sum_{i=1}^{I}\|\mathbf{Y}_i-\mathbf{X}_i\mathbf{A}\mathbf{B}_i\|^2.\end{aligned}$$

We have to choose $\mathbf{A}$ and $\mathbf{B}_i$ $(i=1,2,\cdots,I)$ so as to minimize f $(\mathbf{A},\ \mathbf{B})$ subdect to $\mathbf{F}'\mathbf{F} = N\mathbf{I}_n$, $\mathbf{B}_i\mathbf{B}_i' = \mathbf{I}_t$, and $n = m = t$. Therefore we define

$$\phi = \sum_{i=1}^{I} [(\mathbf{Y}_i - \mathbf{X}_i\mathbf{A}\mathbf{B}_i)'(\mathbf{Y}_i - \mathbf{X}_i\mathbf{A}\mathbf{B}_i)]$$

$$+ \mathrm{tr}[(\mathbf{F}'\mathbf{F} - \mathrm{N}\mathbf{I}_n)\mathbf{L}] + \sum_{i=1}^{I} [\mathrm{tr}(\mathbf{B}_i\mathbf{B}_i' - \mathbf{I}_t)\mathbf{L}_i]$$

where $\mathbf{L}$ and $\mathbf{L}_i$ $(i=1,2,\cdots,I)$ are $t \times t$ matrices of Lagrange multipliers.

To minimize loss function, we use alternating least squares (ALS) technique. The ALS approach is related to the works of Wold (7), de Leeuw et al. (2) and Young (8). As is implied by the name, the essential feature of the ALS approach is that in solving optimization problems with more than one set of parameters, each set is estimated in turn by applying least squares procedures holding the other sets fixed. After all sets have been estimated once, the procedure is repeated until convergence is obtained. The algorithm is convergent since all phases minimize a loss function. In other to see how the ALS approach can be applied in this context, let us return briefly to loss function,

$$f(\mathbf{A}, \mathbf{B}) = \| \mathbf{Y} - \tilde{\mathbf{X}}\tilde{\mathbf{A}}\mathbf{B} \|^2.$$

Clearly the sets of parameters are here $\tilde{\mathbf{A}}$ (or $\mathbf{A}$) and $\mathbf{B}$. Minimizing f over $\mathbf{A}$ holding $\mathbf{B}$ fixed is one least squares problem, and minimizing f over $\mathbf{B}_i$ with $\mathbf{A}$, $\mathbf{B}_1$, $\cdots$, $\mathbf{B}_{i-1}$, $\mathbf{B}_{i+1}, \cdots$, and $\mathbf{B}_I$ fixed is the other. In practice, we have to use ALS approach to minimize f. From above explanation, readers could deduce a rough outline of the algorithm. First we choose an arbitrary $\mathbf{A}_{(0)}$, second compute a new $\mathbf{B}_{(1)}$ using $\mathbf{A}_{(0)}$, third compute $\mathbf{A}_{(1)}$ using $\mathbf{B}_{(1)}$, and iterate this cycle until convergence.

For $i=1,\cdots,I$, minimizing function f over $\mathbf{B}_i$ while matrix $B_1, \cdots, B_{i-1}, \cdots B_{i+1}, \cdots, B_I$ and $\mathbf{A}$ is fixed is achieved as follows. Taking the partial derivative of ϕ with respect to $\mathbf{B}_i$, and setting the partial derivative to zero gives the equation to be solved for $\mathbf{B}_i$. Solving for the value of $\mathbf{B}_i$ gives

$$\mathbf{B}_i = \mathbf{V}_i\mathbf{W}_i',$$

where

$$\mathbf{A}'\mathbf{X}_i'\mathbf{Y}_i\mathbf{Y}_i'\mathbf{X}_i\mathbf{A}\mathbf{V}_i = \mathbf{V}_i\Delta_i,$$

$$\mathbf{Y}_i'\mathbf{X}_i\mathbf{A}\mathbf{A}'\mathbf{X}_i'\mathbf{Y}_i\mathbf{W}_i = \mathbf{W}_i\Delta_i,$$

On the other hand, minimizing f over $\mathbf{A}$ while $\mathbf{B}_i$ are fixed can be done by differetiating ϕ partially with respect to $\mathbf{A}$ and setting the result equal to zero. This yields:

$$\mathbf{A} = \mathbf{R}^{-1}\left(\sum_{i=1}^{I} \mathbf{X}_i'\mathbf{Y}_i\mathbf{B}_i'\right)$$

$$\left[\left(\sum_{i=1}^{I} \mathbf{B}_i\mathbf{Y}_i'\mathbf{X}_i\right)\mathbf{R}^{-1}\left(\sum_{I=1}^{I} \mathbf{X}_i'\mathbf{Y}_i\mathbf{B}_i'\right)\right]^{-1/2},$$

where $\mathbf{R}$ is a correlation matrix,

$$\mathbf{R} = \mathrm{N}^{-1}\sum_{i=1}^{I} \mathbf{X}_i'\mathbf{X}_i.$$

The ALS procedure presented here decreases function f monotonously and the convergence to a stationary point is guaranteed because each problem is solved in the least squares sense.

The point at which the ALS process is initiated may be chosen in several ways. The nearer the starting point is to the solution, the smaller is the expected number of iterations required to reach the solution. The starting point suggested here is a general solution of standard principal component analtsis. The first step in obtaining starting values, $\mathbf{A}_{(0)}$, is to compute correlation matrix, $\mathbf{R}$. Second, we eigen-decompose $\mathbf{R}$,

$$\mathbf{R}\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}_t\Delta_t,$$

where $\boldsymbol{\Gamma}_t$ and Δ_t are diagonal matrix which elements are the t largest eigen values of $\mathbf{R}$, and the matrix consisted of corresponding eigenvectors, respectively. We then have

$$\mathbf{A}_{(0)} = \boldsymbol{\Gamma}_t\Delta_t^{-1/2}.$$

as a starting point for $\mathbf{A}$. However, the starting point can not be guaranteed that the global minimum will be attained. Therefore, it is suggested to run more than one analysis on the same data set with different starting values.

Up to this point, the discussion of SMRPCA has proceeded as if the number of dimensions, t, was equal to number of variables. In practice, however, dimensional condensation is essential feature of most multivariate analyses. A simple way to solve the problem is as follows. First, user must obtain several solutions in different dimensionalities which is smaller than number of variables, and second choose between them on the basis of three criteria : fit to the data, interpretability, and reproducibility. We will not discuss the details of each of them as they are the same as in our previous papers (4, 5). However, to decrease function f monotonously and to give the best solution which minimizes f can not be guaranteed using the procedure presented here since this procedure do not meet the requirement of SMRPCA.

Identification of Parameters

The problem of identificability is to specify what restrictions on the model are required to determine the parameters, uniquely. This problem can be solved by specification of fixed, free and constrained parameters. As well known, each $\mathbf{A}$ and $\mathbf{B}_i$ generates one and only one $\mathbf{W}_i$, but different $\mathbf{A}$ and $\mathbf{B}_i$ can generate the same $\mathbf{W}_i$--namely, if $\mathbf{A}$ is replaced by $\mathbf{AT}^{-1}$ and $\mathbf{B}_i$ by $\mathbf{TB}_i$, where $\mathbf{T}$ is an arbitrary non-singular matrix of order $t \times t$, then $\mathbf{W}_i$ is unchanged. Since $\mathbf{T}$ has t^2 independent elements, this suggests that t^2 independent conditions should be imposed on $\mathbf{A}$ or $\mathbf{B}_i$ to make these uniquely defined. However, It is hard to give a specific rule in the general case. In this method, therefore, no restrictions are imposed to specify the model parameters. If the unrotated components are interpretable, then rotation is unnecessary. If not, an objective rotation can be tried so that structure matrix,

$$\mathbf{S} = N^{-1} \sum_{i=1}^{I} \mathbf{X}'_i \mathbf{F}_i,$$

is simple as in the context of factor analysis.

Treatment of Missing Data

Up to now, we have assumed that the data were in a perfect form for data analysis. In practice, this assumption is often false. However, missing data is allowed for in a manner which does not destroy the ALS property of the SMRPCA algorithm. We can choose scores which minimize the loss function f. If some observation is missing, then computation of the starting values is changed in a minor manner, that is, we simply estimate the scores as being the mean of the nomissing obsevations. Using these starting values, model parameters are estimated. Next, the missing data points are reestimated in a regression fashion, and then a new cycle of the iteration is started.

Moreover, we can consider a practical alternative, which is to insert average of observed scores into missing responses. This approach is most popular one in multivariate statistical analysis. However, this approach simply ignores missing responses, and only the information of observed responses is used in optimization. This is conservative approach, compared with the first one, in the sense that it does not attempt to extract information from relationships between variables. Therefore, the approach would be suitable if all independent variables are relatively homogeneous in the statistical sense.

Treatment of Nonmetric Data

In this section, we present that SMRPCA can be easily extended to nenmetric data, i. e., ordinal and nominal measurement of data. Moreover, we assume two types of measurement process, i. e., discrete and continuous. For analysis designed for data having such a wide variety of measurement, Fisher's notion of optimal scaling (3) is useful. According to his notation, we wish to obtain the optimally scaled data which fit the model as well as possible in a least squares sense. In other words, we rescale the data so that loss function is minimized. The optimal scaling can be carried out with ALSOS (Alternating Least Squares technique to Optimal Scaking) proposed by Young et al. (9). Therefore, we simply impose "optimal scaling phase" bofore "model parameter estimation phase" mentioned earlier. We will not discuss the details of each optimal scaling technique as they are the same as in the de Leeuw's (1) and Young's paper (9).

Extension to Canonical model

Finally, we present extention of SMRPCA to canonical data. In this case, the model is changed as follows :

$$\tilde{\mathbf{Y}}\tilde{\mathbf{A}}_Y\mathbf{B}_Y = \tilde{\mathbf{X}}\tilde{\mathbf{A}}_X\mathbf{B}_X + \mathbf{E},$$

that is,

$$\tilde{\mathbf{Y}}\mathbf{W}_Y = \tilde{\mathbf{X}}\mathbf{W}_X + \mathbf{E},$$

or

$$\tilde{\mathbf{F}}_Y\mathbf{B}_Y = \tilde{\mathbf{F}}_X\mathbf{B}_X + \mathbf{E},$$

where,

$$\tilde{\mathbf{X}} = \sum_{i=1}^{I} \mathbf{E}_{ii}\otimes\mathbf{X}_i,$$

$$\tilde{\mathbf{Y}} = \sum_{i=1}^{I} \mathbf{E}_{ii}\otimes\mathbf{Y}_i,$$

$$\tilde{\mathbf{F}}_X = \sum_{i=1}^{I} \mathbf{E}_{ii}\otimes\mathbf{F}_{Xi},$$

$$\tilde{\mathbf{F}}_Y = \sum_{i=1}^{I} \mathbf{E}_{ii}\otimes\mathbf{F}_{Yi},$$

$$\tilde{\mathbf{A}}_X = \mathbf{I}_I\otimes\mathbf{A}_X$$
$$\tilde{\mathbf{A}}_Y = \mathbf{I}_I\otimes\mathbf{A}_Y$$
$$\mathbf{B}_X = (\mathbf{B}_{X1}'\mathbf{B}_{X2}'\cdots\mathbf{B}_{XI}')',$$
$$\mathbf{B}_Y = (\mathbf{B}_{Y1}'\mathbf{B}_{Y2}'\cdots\mathbf{B}_{YI}')',$$
$$\mathbf{E} = (\mathbf{E}_1'\mathbf{E}_2'\cdots\mathbf{E}_I')'.$$

As in the case of SMRPCA, the equality constraints, that is,

$$\mathbf{F}_X'\mathbf{F}_X = N\mathbf{I}_n,$$
$$\mathbf{F}_Y'\mathbf{F}_Y = N\mathbf{I}_n,$$
$$\mathbf{B}_{Xi}\mathbf{B}_{Xi}' = \mathbf{I}_t,$$
$$\mathbf{B}_{Yi}\mathbf{B}_{Yi}' = \mathbf{I}_t,$$
$$n = m = t,$$

are required to solve the minimization problem. The loss function to be minimize is difined as follows :

$$g(\mathbf{A}_X, \mathbf{A}_Y, \mathbf{B}_X, \mathbf{B}_Y) = \sum_{i=1}^{I} \| \mathbf{Y}_i\mathbf{A}_Y\mathbf{B}_{Yi} - \mathbf{X}_i\mathbf{A}_X\mathbf{B}_{Xi} \|^2.$$

A careful reader will detect that a problem of minimizing g can be solved in the way similar to SMRPCA. We use ALS procedure, which consists of four phases for estimating $\mathbf{A}_X$, $\mathbf{A}_Y$, $\mathbf{B}_X$, and $\mathbf{B}_Y$. Computation of each estimation phase is summarized as follows :

$$\mathbf{A}_X = \mathbf{R}_{XX}^{-1}\left(\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{Y}_i\mathbf{A}_Y\mathbf{B}_{Yi}\mathbf{B}_{Xi}'\right)$$

$$\left[\left(\sum_{i=1}^{I}\mathbf{B}_{Xi}\mathbf{B}_{Yi}'\mathbf{A}_Y'\mathbf{Y}_i'\mathbf{X}_i\right)\mathbf{R}_{XX}^{-1}\left(\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{Y}_i\mathbf{A}_Y\mathbf{B}_{Yi}\mathbf{B}_{Xi}'\right)\right]^{-1/2},$$

$$\mathbf{A}_Y = \mathbf{R}_{YY}^{-1}\left(\sum_{i=1}^{I}\mathbf{Y}_i'\mathbf{X}_i\mathbf{A}_X\mathbf{B}_{Xi}\mathbf{B}_{Yi}'\right)$$

$$\left[\left(\sum_{i=1}^{I}\mathbf{B}_{Yi}\mathbf{B}_{Xi}'\mathbf{A}_X'\mathbf{X}_i'\mathbf{Y}_i\right)\mathbf{R}_{YY}^{-1}\left(\sum_{i=1}^{I}\mathbf{Y}_i'\mathbf{X}_i\mathbf{A}_X\mathbf{B}_{Xi}\mathbf{B}_{Yi}'\right)\right]^{-1/2},$$

$$\mathbf{B}_{Xi} = \mathbf{V}_{Xi}\mathbf{W}_{Xi}',$$
$$\mathbf{B}_{Yi} = \mathbf{V}_{Yi}\mathbf{W}_{Yi}',$$

where

$$\mathbf{A}_X'\mathbf{X}_i'\mathbf{Y}_i\mathbf{A}_Y\mathbf{B}_{Yi}\mathbf{B}_{Yi}'\mathbf{A}_Y'\mathbf{Y}_i'\mathbf{X}_i\mathbf{A}_X\mathbf{V}_{Xi} = \mathbf{V}_{Xi}\boldsymbol{\Delta}_{Xi},$$
$$\mathbf{B}_{Yi}'\mathbf{A}_{Yi}'\mathbf{Y}_i'\mathbf{X}_i\mathbf{A}_X\mathbf{A}_X'\mathbf{X}_i'\mathbf{Y}_i\mathbf{A}_Y\mathbf{B}_{Yi}\mathbf{W}_{Xi} = \mathbf{W}_{Xi}\boldsymbol{\Delta}_{Xi},$$
$$\mathbf{A}_Y'\mathbf{Y}_i'\mathbf{X}_i\mathbf{A}_X\mathbf{B}_{Xi}\mathbf{B}_{Xi}'\mathbf{A}_X'\mathbf{X}_i'\mathbf{Y}_i\mathbf{A}_Y\mathbf{V}_{Yi} = \mathbf{V}_{Yi}\boldsymbol{\Delta}_{Yi},$$
$$\mathbf{B}_X'\mathbf{A}_{Xi}'\mathbf{X}_i'\mathbf{Y}_i\mathbf{A}_Y\mathbf{A}_Y'\mathbf{Y}_i'\mathbf{X}_i\mathbf{A}_X\mathbf{B}_{Xi}\mathbf{W}_{Yi} = \mathbf{W}_{Yi}\boldsymbol{\Delta}_{Yi},$$

$$\mathbf{R}_{XX} = N^{-1}\sum_{i=1}^{I}(\mathbf{X}_i'\mathbf{X}_i),$$

$$\mathbf{R}_{YY} = N^{-1}\sum_{i=1}^{I}(\mathbf{Y}_i'\mathbf{Y}_i),$$

and where

$$\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{1} = \mathbf{0},$$

$$\sum_{i=1}^{I}\mathbf{Y}_i'\mathbf{1} = \mathbf{0},$$

$$\mathrm{Vec}[\mathrm{Diag}(N^{-1}\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{X}_i)] = \mathbf{1},$$

$$\mathrm{Vec}[\mathrm{Diag}(N^{-1}\sum_{i=1}^{I}\mathbf{Y}_i'\mathbf{Y}_i)] = \mathbf{1},$$

References

1) de Leeuw J (1969) : The linear nonmetric model (report RN003-69). University of Leiden, The Netherlands.

2) de Leeuw J, Young FW, and Takane Y (1976) : Additive structure in qualitative data : An alternating least squares method with optimal scaling feature. Psychometrika 41 : 471-503.

3) Fisher RA (1946) : Statistical methods for research workers (10th ed.). Oliver and Boyd, Edinburgh.

4) Inagaki A (1992) : Population differences multivariate regression analysis, PDMRA. Circular of measurement and evaluation Division, Japanese Society of Physical Education 53 : 153-160.

5) Inagaki A, and Matsuura Y (1992) : Multidimensional representation model of population differences in restricted regression analysis, PRDMRA. Bulletin of Health and Sport Sciences, University of Tsukuba 15 : 213-219.

6) Tucker LR (1963) : Implications of factor analysis of three-way matrices for measurement of change. (Ed.) CW Harris, (In) Problems in measuring change. University of Wisconsin Press, Madison, pp. 122-137.

7) Wold H, and Lyttkens E (1969) : Nonlinear iterative partial least squares (NIPALS) estimation procedure. Bulletin of the International statisical Institute 43 : 29-47.

8) Young FM (1972) : A model for polynominal conjoint analysis algorithms. (Eds.) Shepard RN, AK Romney, and S Nerlove, (In) Multidimensional scaling : Theory and applications in the behavior-sciences. Academic Press, New York.

9) Young FW, de Leeuw J, and Takane Y (1976) : Regression with qualitative and quantitative variables : An alternating least squares method with optimal scaling feature. Psychometrika 41 : 505-529.