制約条件付き多変量回帰分析における母集団差異の多次元表現モデル

稲垣　敦・松浦義行

Multidimensional representation model of population differences in restricted multivariate regression analysis, RPDMRA

Atsushi INAGAKI and Yoshiyuki MATSUURA

本研究では，制約条件付き母集団差多変量回帰分析，RPDMRA（Restricted Population Differences Multivariate Regression Analysis）のモデルと，こ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。この手法のモデルは，

$$\mathbf{Y}=\mathbf{XW}+\mathbf{E}$$
$$=\mathbf{XAB}+\mathbf{E}$$

で表され，

$$\mathbf{A}_s'\mathbf{A}_s=\mathbf{I}_r,$$
$$\mathbf{B}_i\mathbf{B}_i'=\mathbf{I}_r \text{ for all i},$$
$$m=n=r,$$

である。但し，

$$\mathbf{Y}=(\mathbf{Y}_1',\ \mathbf{Y}_2',\ \cdots,\ \mathbf{Y}_I')',$$
$$\mathbf{X}=\sum_{i=1}^{I}\mathbf{E}_{Ii}\otimes\mathbf{X}_i,$$
$$\mathbf{W}=(\mathbf{W}_1',\ \mathbf{W}_2',\ \cdots,\ \mathbf{W}_I')',$$
$$\mathbf{A}=\sum_{i=1}^{I}\mathbf{E}_{Ii}\otimes\mathbf{A}_i,$$
$$\mathbf{A}_i=\mathbf{A}_s \text{ for all i},$$
$$\mathbf{B}=(\mathbf{B}_1',\ \mathbf{B}_2',\ \cdots,\ \mathbf{B}_I')',$$
$$\mathbf{E}=(\mathbf{E}_1',\ \mathbf{E}_2',\ \cdots,\ \mathbf{E}_I')',$$

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×m）と独立変数行列（N×n），$\mathbf{A}_s$ は各母集団に共通な定常ウェイト行列（n×r），$\mathbf{B}_I$ は母集団固有の母集団差ウェイト行列（r×m），$\mathbf{E}_I$ は第 i 母集団における誤差行列（N×m）である。しかし，このモデルの適用においては，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。以上の点からわかるように，RPDMRA は，POPREG（POPulation differences REGression analsis；Inagaki, 1991）の従属変数を多変量化し，従属・独立変数と次元数の等化条件を加え，$\mathbf{A}_s$ と $\mathbf{B}_I$ の正規直交条件を挿入した場合に相当する。したがって，異なったグループや母集団からの標本というような複数組のデータに対応しており，その母集団差異を表現できるだけでなく，空間的に表現することにより，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，複数の時間や異なる条件からのデータにも適用可能であり，縦断的研究，及び多変量時系列データの分析に応用可能であると考えられる。さらに，ALSOS（Alternating Least Squares approach to Optimal Scaling；Young, 1981）を利用することにより，比率・間隔尺度で測定されたデータのみならず，順序・

名義尺度水準，およ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。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 ; ALS）を用いる。このアルゴリズムは，目的関数（誤差関数）を単調に減少させて局所最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分析を試みる必要がある。以上の特徴から，このモデルは体育学をはじめ，心理学，社会学など幅広い分野で応用可能であると考えられる。

Key words : Multivariate regression analysis, Population differecnes, RPDMRA

Introduction

This article presents the development of "Restricted population Differences Multivariate Regression Analysis" (RPDMRA) and its algorithm. This method can operate on N-sets of data from different populations or situations, and provide not only a linear multivariate regression equation, but also a multidimensional spatial representation of population differences under the assumption of the vector model. Moreover this method is useful instrument for longitudinal or time series research because it can be also applied to the data obtained at the different time points or occasions. The data can be defined at nominal, ordinal, interval or ratio levels of measurement[6], or can be mixture of two or more levels. As will be explained, RPDMRA model provides an optimal scale for each variable within the restrictions as to the measurement level and process. This scaling is optimal in the sense that correlation is maximized.

This model is characterized as follows :

1) Squared loss funtion-the proposed model is fitted to data in a sense of least square.
2) Alternating least squares technique-the proposed model is fitted to data with the alternating least squares technique.
3) Multidimensional spatial representation-the proposed model can represent both populations (groups, situatins, or occasions) and predictive variables as vectors in multidimensional space.
4) Time series data-the proposed method will be able to apply the time series and longitudinal data.
5) Level of measurement-the proposed model will be able to accomodate ratio, interval, ordinal and nominal data.

In the next section we will present a detailed account of RPDMRA model, emphasizing the characteristics of the model.

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, with a row vector denoted as transpose of a column vector($\mathbf{x}'$). We refer to a specific column vector of a matrix as $\mathbf{x}_j$, a specific element of a matrix as x_{ij}.

Let :

i =1, ···, I populations, occasions or conditions,
j =1, ···, n independent, predictive or explanatory variables,
k =1, ···, m dependent, predicted or criterion variables,
s =1, ···, r dimensions in a vector model contexts,
$\mathbf{Y}_i$ =matrix of dependent, predicted or criterion variables in the i-th population, occasion or condition(N×m), which is measured at interval measurement level, and is columnwisely normalized in each group,
$\mathbf{X}_i$ =matrix of independent, predictive or explanatory variables in the i-th popu-

lation, occasion or condition (N×n), which is measured at interval measurement level, and is columnwisely normalized in each group,

$\mathbf{W}_i$ = column vector of regression weights of $\mathbf{Y}_i$ on $\mathbf{X}_i$ (n×m),

$\mathbf{A}_s$ = matrix of stationary weight (n×r),

$\mathbf{B}_i$ = matrix of population differences weight of dimensions for the i-th population (r×m),

$\mathbf{E}_i$ = matrix of error or residual of the i-th population (N×m)

Using the above definitions, we can formulate the RPDMRA model by matrix form ;

$$\begin{aligned}\mathbf{Y} &= \mathbf{XW}+\mathbf{E}\\ &= \mathbf{XAB}+\mathbf{E},\end{aligned}$$

where

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

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

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

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

$$\mathbf{A}_i = \mathbf{A}_s \text{ for all i},$$

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

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

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 (ij) position, and with zeros elsewhere. This model is similar to POPREG[4]. Additionally we impose some constraints, that is,

$$\mathbf{A}_s'\mathbf{A}_s = \mathbf{I}_r,$$
$$\mathbf{B}_i\mathbf{B}_i' = \mathbf{I}_r \text{ for all i},$$
$$m = n = r.$$

Loss Function

In our case, as in many similar methods, we define a squared loss function. We then search for the best solution such that

$$f(\mathbf{A}_s, \mathbf{B}_1, \cdots, \mathbf{B}_I) = \mathrm{tr}[(\mathbf{Y}-\hat{\mathbf{Y}})'(\mathbf{Y}-\hat{\mathbf{Y}})]$$

is minimal, where $\hat{\mathbf{Y}}$ denotes Y estimated by the model. As mention before, the minimization has to be carried out under the constraints ;

$$\mathbf{A}_s'\mathbf{A}_s = \mathbf{I}_r,$$
$$\mathbf{B}_i\mathbf{B}_i' = \mathbf{I}_r \text{ for all i},$$
$$m = n = r$$

Estimation

To minimize loss fuction, we utilize alternating least squares technique (ALS). The ALS approach is related to the works of Wold[7], de Leeuw[1] 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. The algorithm is convergent, since all phases minimize a loss function.

In order to see how the ALS approach can be applied in the present context, let us return briefly to loss function :

$$\begin{aligned}&f(\mathbf{A}_s, \mathbf{B}_1, \cdots, \mathbf{B}_I)\\ &\quad = \sum_{i=1}^{I} \mathrm{tr}[(\mathbf{Y}_i-\mathbf{X}_i\mathbf{A}_s\mathbf{B}_i)'(\mathbf{Y}_i-\mathbf{X}_i\mathbf{A}_s\mathbf{B}_i)]\end{aligned}$$

Clearly the sets of parameters are here $\mathbf{A}_s$ and $\mathbf{B}_i$. Minimizing f over $\mathbf{A}_s$ holding $\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_I$ fixed is equivalent to solving one least squares problem and minimizing over $\mathbf{B}_i$ with $\mathbf{A}_s$ fixed is another. That we are in practice minimizing f does not prevent the problem from being ALS one.

We must choose $\mathbf{A}$ and $\mathbf{B}_i$ to minimize f $(\mathbf{A}_s, \mathbf{B}_1, \cdots, \mathbf{B}_I)$ subject to $\mathbf{A}_s'\mathbf{A}_s = \mathbf{I}_r$, $\mathbf{B}_i\mathbf{B}_i' = \mathbf{I}_r$ for all i and $m = n = r$. We difine

$$\phi = \sum_{i=1}^{I} \mathrm{tr}[(\mathbf{Y}_i - \mathbf{X}_i\mathbf{A}_s\mathbf{B}_i)'(\mathbf{Y}_i - \mathbf{X}_i\mathbf{A}_s\mathbf{B}_i)] + \mathrm{tr}[(\mathbf{A}_s'\mathbf{A}_s - \mathbf{I}_r)\mathbf{L}] + \sum_{i=1}^{I} \mathrm{tr}[(\mathbf{B}_i\mathbf{B}_i' - \mathbf{I}_r)\mathbf{L}_i],$$

where $\mathbf{L}$ and $\mathbf{L}_i$ are $r \times r$ matrices of Lagrange multipliers. The partial derivative of ϕ with respect to $\mathbf{B}_i$ is relatively easy to obtain and setting to zero, and solving for the value of $\mathbf{B}_i$ which minimizes ϕ for given $\mathbf{A}_s$, results in the following expression for $\mathbf{B}_i$:

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

where

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

and where Δ_i denotes the diagonal matrix which elements are eigen values, and $\mathbf{V}_i$ and $\mathbf{W}_i$ are the matrices which consist of the corresponding eigen vectors. For $1=1, 2\cdots, I$, minimizing f over $\mathbf{B}_i$ while $\mathbf{A}_s$ is fixed is achieved with this procedure.

On the other hand, the poblem minimizing f as function of $\mathbf{A}_s$ subject to $\mathbf{A}_s'\mathbf{A}_s = \mathbf{I}_r$, for fixed $\mathbf{B}_i$, can be solved as follow. The partial derivative of ϕ with respect to $\mathbf{A}_s$ is also easy to obtain and setting to zero, and solving for the matrix of $\mathbf{A}_s$ which minimizes ϕ for given $\mathbf{B}_1, \cdots, \mathbf{B}_I$, results in the following expression for $\mathbf{A}_s$:

$$\mathbf{A}_s = \mathbf{V}\mathbf{W}',$$

where

$$[\sum_{i=1}^{I}(\mathbf{X}_i'\mathbf{Y}_i\mathbf{B}_i')][\sum_{i=1}^{I}(\mathbf{X}_i'\mathbf{Y}_i\mathbf{B}_i')]'\mathbf{V} = \mathbf{V}\Delta,$$

$$[\sum_{i=1}^{I}(\mathbf{X}_i'\mathbf{Y}_i\mathbf{B}_i')]'[\sum_{i=1}^{I}(\mathbf{X}_i'\mathbf{Y}_i\mathbf{B}_i')]\mathbf{W} = \mathbf{W}\Delta,$$

The ALS procedure presented here decreases function f monotonously and the convergence to a stationary point is guaranteed because each problemis solved in the least sqares sence. However it 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.

Starting Values

If there are many parameters, like RPDMRA, the number of iterations may be excessive in ALS procedure, but can be considerably decreased by the provision of good starting values. We provide the starting values which is obtained under some restrictions.

If we assume that the population differences weight matrices $\mathbf{B}_i$ are identical across all populations, the model can be written as

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

where

$$\tilde{\mathbf{W}} = \mathbf{A}_s\mathbf{B},$$
$$\tilde{\mathbf{X}} = (\mathbf{X}_1', \mathbf{X}_2', \cdots, \mathbf{X}_I')'.$$

In this case, squared loss function is minimized over $\tilde{\mathbf{W}}$ by choosing

$$\hat{\tilde{\mathbf{W}}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{Y}.$$

Therefore in order to obtain starting valne of $\mathbf{A}_s$, we decompose matrix $\tilde{\mathbf{W}}$ as

$$\hat{\tilde{\mathbf{W}}} = \mathbf{A}_s\mathbf{B}.$$

In other words, we factor $\mathbf{W}$ into principal components. However the usual factoring equation are not appropriate, since this matrix is an asymmetric. Therefore we utilize Singular-Value-Decomposition (SVD) to determine the initial $\mathbf{A}_s$ under the constraints that ;

$$\mathbf{A}_s'\mathbf{A}_s = \mathbf{I}_r,$$
$$\mathbf{B}\mathbf{B}' = \text{diagonal}.$$

The cross products matrix in the present analysis should always be computed between the variables on the shorter side of the approximated matrix, summing over the variables on the longer side. In our case, however, matrix $\mathbf{W}$ is square. So the starting value of $\mathbf{A}_s$ can be set up as

$\mathbf{A}_s = \mathbf{V}_r$,

where

$\hat{\hat{\mathbf{W}}}\hat{\hat{\mathbf{W}}}'\mathbf{V}_r = \mathbf{V}_r\Gamma_r$,

and where Γ_r and $\mathbf{V}_r$ denote the diagonal matrix which elements are the r largest eigen values and matrix consisted of the corresponding eigen vactors, respectively.

Identification

Before an attempt is made to estimate model parameters of this kind, the identification problem must be examined. The identification problem depend on the specification of fixed, free and constrained parameters. Under several specifications, each $\mathbf{A}_s$ and $\mathbf{B}_i$ generates one and only one $\mathbf{W}_i$, but it is well known that different $\mathbf{A}_s$ and $\mathbf{B}_i$ can generate the same $\mathbf{W}_i$. It shold be noted that if $\mathbf{A}_s$ is replaced by $\mathbf{A}_s\mathbf{T}^{-1}$ and $\mathbf{B}_i$ by $\mathbf{TB}_i$, where $\mathbf{T}$ is an arbitrary non-singuler matrix of order rxr, then $\mathbf{W}_i$ is unchanged. Since T has r^2 independent elements, this suggests that r^2 independent conditions shold be imposed on $\mathbf{A}_s$ or $\mathbf{B}_i$ to make these uniquely defined. However, when equality constraints over groups are taken into account, all the elements of the transformation matrix is not independent of each other and therefore a lesser number of conditions need to be imposed. It is hard to give further specific rule in the general case. In this method, $\mathbf{A}_s$ and $\mathbf{B}_i$ should be estimated without any constraints. If the unrotated dimensions is interpretable, then rotation is unnecessary. If not, some objective rotation can be tried as in the case of factor analysis.

Treatment of Missing Data

Missing data are allowed for in a manner which does not destroy the ALS property of the RPDMRA algorithm. If some observation is missing, then computation of the starting values is changed in a minor manner, that is, we simply estimate the optimal scaling observation as being the mean of the nomissing obsevations. Using these starting values, model parameters are estimated. Nextly the missing data points are reestimated in a regression fashon and then a new cycle of the iteration is started. In fact such procedures are standard within the ALS approach.

Examination of Number of Dimensions

As discribed before, the number of dimensions, r, were assumend to be equal to number of independent and dependent variables, since it is neccesary to insert the restriction for solving the minimizing problem under the restriction $\mathbf{A}_s'\mathbf{A}_s = \mathbf{I}_r$ and $\mathbf{B}_i\mathbf{B}_i' = \mathbf{I}_r$. However, in most multivariate analysis without external criterion like factor analysis, dimensional reduction is essential problem. Therefore if the reduction in number of dimension is necessary, analysis under the perfect restrictions should be carried out and then some of dimensions can be choose as in the case of factor analysis.

In practice, user must obtain several solutions in different number of dimensions and 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 paper[4].

Finally, we note that it can not be guaranteed the grobal minimum will be attained when the number of dimensions was not assumed equal to number of dependent and independent variables, because least square estimates for parameters are not obtaind with the procedure discribed in this section.

Extension to Nonmetric Data

Finaly, we emphasize that RPDMRA model can be easily extended to nonmetric 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 opti-

mal 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 sence. In other words, we rescale the data so that multiple correlation is maximized. The optimal scaling can be carried out with ALSOS (Alternating Least Squares technique to Optimal Scaling) proposed by Young[10].

In order to handle nonmetric data, we impose the concept of "measurement transformation",

$$\mathbf{Y}_{ij}^{*} = gy_{ij}(\mathbf{Y}_{ij}) \quad (i=1, \cdots, I\ ;\ j=1, \cdots, m),$$
$$\mathbf{X}_{ij} = gx_{i}(\mathbf{X}_{ij}) \quad (i=1, \cdots, I\ ;\ j=1, \cdots, n),$$

where gx_{ij} and gy_{ij} are so-called "measurement transformations", and are subject to restrains by the measurement level and process of their variables. The restrains were discussed in Young[9]. As mentioned before, we determine $\mathbf{X}_{ij}^{*}$ and $\mathbf{Y}_{ij}^{*}$ so that the loss function is minimzed. Thus we refer to $\mathbf{X}_{ij}^{*}$ and $\mathbf{Y}_{ij}^{*}$ as optimally scaled observations (data). According to this, the model can be rewritten as ;

$$\mathbf{Y}^{*} = \mathbf{X}^{*}\mathbf{A}\mathbf{B} + \mathbf{E},$$

where

$$\mathbf{Y}^{*} = (\mathbf{Y}_{1}^{*\prime}, \mathbf{Y}_{2}^{*\prime}, \cdots, \mathbf{Y}_{I}^{*\prime})',$$
$$\mathbf{X}^{*} = \sum_{i=1}^{I} \mathbf{E}_{ii} \otimes \mathbf{X}_{i}^{*}.$$

For the numerical data, the optimal scaling phase is skipped. For the ordinal data, "Kruskal's least squares monotonic regression"[5] can be used. In this case, "the primary approach to tie" is chosen for contiuous-ordinal data, whereas "the secondary approach to tie" is chosen for discrete-ordinal data. For the discrete-nominal data, optimally scaled data are category means. Finally, for the continuouous-nominal data we assume it to be pseudo-continus-ordinal data to determine the optimally scaled data. We will not discuss the details of each optimal scaling technique as they are the same as in the de Leeuw's[2] and Young's paper[9].

In the case where raw data are measured on ordinaly scale, we simply assume that the raw data are measured on interval scale and then set up starting values by our method explained in previous section. Using this starting values, we can easily investigate whether our assumption concerning measurement levels are correct or not. On the other hand, we must assign arbitrary values to the observation categories when variable is assumed to be nominal.

A Computer Program

A program, RPDMRA, was developed for computing the solution. It is written in single precision of FORTRAN77 for all real variables. The number of variables and dimensions can range up 10, and the number of samples to be used in each group can range up 100. Printed output includes the title ; the job parameters ; the stationary weight matirx $\mathbf{A}_s$; the population differences weight matrices $\mathbf{B}_i$ (i=1, ···, I) ; the convergent process of loss function ; two-dimensional scatterplot of $\mathbf{A}_s$; two-dimensional scatterplot of $\mathbf{B}_i$.

References

1) de Leeuw J (1969) : The linear nonmetric model (report RN003-69). University of Leiden, 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 and Matsuura Y (1991) : A multidimensional represntation model of population differences in regression analysis. Bulletin of Institute of Health and Sport sciences, University of Tsukuba 14 : 211-222.

5) Kruskal JB (1964) : Nonmetric multidimensional scaling : A numerical method. Psychometrika 29 : 28-42.

6) Stevens SS (1951) : Mathematics, measurement, and psychophysics. (Ed.) Stevens SS (In) Handbook of Experimental Psychology. Wiley, New York.

7) Wold H and Lyttkens E (1969) : Nonlinear

iterative partial Least squares(NIPALS) estimation procedure. Bull ISI 43 : 29-47.

8) Young FM (1972) : A model for polynominal conjoint analysis algorithms. (Eds.) Shepard RN, Romney AK and Nerlove S (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.

10) Young FW (1981) : Quantitative analysis of qualitative data. Psychometrika 46 : 357-388.