Comparison of Variability in SCA Maps Using the Procrustes Analysis

Woo Jung Yun and Yong-Seok Choi[1]

Abstract

Some multivariate analyses provide configurations for variables or objects in low dimensional space because we can see easily their relation. In particular, in simple correspondence analysis(SCA), we can obtain the various configurations which are called SCA Maps based on the algebraic algorithms. Moreover, it often occur the variability among them. Therefore, in this study, we will give a comparison of variability of SCA maps using the procrustes analysis which is a technique of comparing configurations in multidimensional scaling.

Keywords: Configuration, Multidimensional Analysis, , Simply Correspondence Analysis, Procrustes Analysis

1. Introduction

Multivariate analysis refers to all statistical methods that simultaneously analysis multiple measurements on each individual of object under investigation. The some techniques of multivariate analyses refer graphic representation for variables or objects in low dimensional space. We can obtain the configurations of various form by algebraic algorithms. Thus, it often necessary to compare one configuration of points in a Euclidean space with another, where there is a one-to-one mapping from one set of points to the other.

Until now, we saw the similarity of two configurations with the naked eye. Thus, in this study, we represent more objectively the similarity of two configurations by producing the numerical statistic for them using the procrustes analysis. The procrustes analysis provides a numerical comparison of two configurations obtained by moving one configuration that it aligns best with the other, after the innkeeper Procrustes, in Greek mythology, who would either stretch or lop off customers' limbs so they would fit his bed. Procrustes analysis is toriginally developed by Schönemann and Carroll (1970) and Gower (1971). Sibson(1978) gives a short review of procrustes analysis. In fact, the procrustes analysis has been applied mostly to only multidimensional scaling based on the Kruskal's stress. In addition, we will apply it to simple correspondence analysis(SCA).

In section 2, we introduce the algorithm of the procrustes analysis. In section 3, we introduce SCA simply and its algorithm. Next, we apply the procrustes analysis to SCA through an example.

1) Professor, Department of Statistics, Pusan National University, Busan 609-735, Korea

2. Procrustes Analysis

Suppose the $n \times p$ matrix C_k contains the coordinates for plotting with technique 1 and the $n \times p$ matrix C_l contains the coordinates from technique 2, where $q \le p$. We call the C_k and C_l configuration matrix. To determine how compatible the two configurations are, we move, say, the second configuration to match the first by shifting each point by the same amount and rotating or reflecting the configuration about the coordinate axes.

Let us translate by a vector **t** and multiply by an orthogonal matrix **R** so that the coordinates of the jth point $\mathbf{c}^l_j$ are transformed to $\mathbf{R}\mathbf{c}^l_j + \mathbf{t}$. Now the steps in procrustes analysis where configuration C_k is to be matched to configuration C_l are as follows:

Step 1. Subtract the mean vectors for the configurations from each of the respective points in order to have the centroids at the origin.

Step 2. Find the rotation matrix $R = VU'$ by SVD of $C_l'C_k = U\Lambda V'$

Step 3. Calculate the procrustes statistic

$$PA = tr(C_k'C_k) + tr(C_l'C_l) - 2tr(\Lambda)$$

Besides ***PA***, Cox and Cox (1994) and Sibson(1978) give the other procrustes statistic

$$\gamma = 1 - \frac{\{tr(C_lC_k'C_kC_l)^{1/2}\}^2}{tr(C_lC_l')tr(C_kC_k')}.$$

3. Application in Simple Correspondence Analysis

Correspondence analysis is a graphical procedure for representing associations in a table of frequencies or counts. Let $O = (o_{ij})$ be an $I \times J$ data matrix of unscaled frequencies or counts. We shall denote the overall total simply by $o_{++} = 1_I'O1_J$ where 1_I and 1_J are $I \times 1$, $J \times 1$ vectors with I and J ones respectively. Note that the correspondence matrix F is define as

$$F = (f_{ij}),\ f_{ij} = o_{ij}/o_{++},\quad i = 1,2,\ldots,I,\quad j = 1,2,\ldots,J$$

Next, by the singular value decomposition,

$$F = AD_uB' \tag{3.1}$$

where $D_r = diag(f_{1+},\ldots,f_{I+})$, $D_c = diag(f_{+1},\ldots,f_{+J})$, $A'D_r^{-1}A = B'D_c^{-1}B = I$ and $D_u = diag(1, \lambda_1, \ldots, \lambda_{J-1})$.

From (3.1), the coordinates of the row profiles and column profiles are

$$D_r^{-1}AD_u \text{ and } D_c^{-1}BD_u \tag{3.2}$$

This coordinates pairs of (3.2) give one configuration. In addition to (3.2), we can make three configurations more mainly used ($D_r^{-1}AD_u$ and $D_c^{-1}B$) , ($D_r^{-1}A$ and $D_c^{-1}BD_u$) and ($D_r^{-1}AD_u^{1/2}$ and $D_c^{-1}BD_u^{1/2}$). Let C_1, C_2 C_3 and C_4 are the configurations constructed by ($D_r^{-1}AD_u$ and $D_c^{-1}B$), ($D_r^{-1}AD_u$ and $D_c^{-1}BD_u$), ($D_r^{-1}A$ and $D_c^{-1}BD_u$), ($D_r^{-1}AD_u^{1/2}$ and $D_c^{-1}BD_u^{1/2}$) respectively. We call these configurations SCA maps.

Now, we will compare with four maps Figure 1, Figure 2, Figure 3 and Figure 4 which are coordinates pairs C_1, C_2 C_3 and C_4 respectively from SCA each other through the example. For this example, all the maps are comparatively similar. But Figure 1(C_1) and Figure 3(C_3) are different because row(preference) categories are in the middle of the plot in Figure 1(C_1), on the contrary, Figure 3(C_3) shows that column(scholastic ability) categories are in the middle of the plot. Table 1 is the result of procrustes analysis and shows how much they are similar each other numerically. C_2 and C_4 are extremely similar each other because procrustes statistic is $PA = .000$. The other hand, C_1 and C_3 are a little different from each other because of $PA = 21.121$.

Tablel. Procrustes statistics of SCA maps

	C_1	C_2	C_3
C_2	5.541		
C_3	21.121	6.530	
C_4	5.541	0.000	6.530

C_1 : ($D_r^{-1}AD_u$ and $D_c^{-1}B$) C_2 : ($D_r^{-1}AD_u$ and $D_c^{-1}BD_u$)

C_3 : ($D_r^{-1}A$ and $D_c^{-1}BD_u$) C_4 : ($D_r^{-1}AD_u^{1/2}$ and $D_c^{-1}BD_u^{1/2}$)

References

[1] Cox, T. F and Cox, M. A. A.(1994) *Multidimensional Scaling*, Chapman & Hall, London.

[2] Gower, J. C. (1971) "Statistical methods of comparing different multivariate analyses of the same data" in Mathematics in the Archaeological and Historical Sciences (F. R. Hodson, D. G. Kendall and P. Tautu, eds), 138-149, Edinburgh: University Press.

[3] Schonemann, P. H. and Carroll, R. M. (1970) "Fitting one matrix to another under choice of a central dilation and a rigid motion", *Psychometrika*, 35, 245-255

[4] Sibson, R. (1978) "Studies in the Robustness of Multidimensional Scaling" *Journal of the Royal Statistical Society(B)*, 40, 234-238.