Group Average Models

John C, Gower
The Open University
Walton Hall
Department of Statistics
Milton Keynes, MK7 6AA, U.K.

Abstract

In univariate methods, the average or mean is a basic summary statistic; in multidimensional psychometric methods the corresponding statistic is the group average. The group-average manifests itself in disparate ways in different models that, at first sight, seem to have little in common. The relationships between several formulations of Generalised Procrustes Analysis, STATIS, INDSCAL, PINDIS, multidimensional scaling, and generalisations of canonical correlation analysis will be explored. Close examination shows that, in addition to the concept of a group average, these models have other features in common.

We consider K data matrices $\mathbf{X}_k$ (k = 1,2,…,K). The equivalence of the generalised Procrustes criterion of minimising $\sum_{i<j}^{K} \| \mathbf{X}_i\mathbf{T}_i - \mathbf{X}_j\mathbf{T}_j \|$ to the minimisation of $\sum_{i<j}^{K} \| \mathbf{X}_i\mathbf{T}_i - \mathbf{G} \|$, where $\mathbf{G}$ = $\frac{1}{K}\sum_{i<j}^{K} \mathbf{X}_i\mathbf{T}_i$, the group-average, is only part of the story. Different transformation matrices $\mathbf{T}_k$ induce different methods; different constraints, both of the weak and strong types, have important effects. The use of diagonal anisotropic scaling matrices $\mathbf{S}_k$ and whether they are applied to $\mathbf{G}$ or to $\mathbf{T}_k$ and if so whether as pre- or post- multiplying matrices (or both), is also important and is a central feature of PINDIS. The matrix $\mathbf{Z}_{ij} = \{trace(\mathbf{T}_i'\mathbf{X}_i'\mathbf{X}_j\mathbf{T}_j)\}$ links anisotropic scaling in certain forms of Procrustes analysis with STATIS. Canonical correlation analysis has several forms, but one of the more interesting expresses the problem in the Generalised Procrustes form, with size constraints on the transformation matrices; the resulting group-average has optimal correlational interpretations. Common components analysis is more remote but in some forms may be set up as a reduced rank matching problem with a consequent group-average. Many models expressed in terms of data matrices $\mathbf{X}_k$ have analogous formulations in terms of inner product matrices $\mathbf{B}_k$ = $\mathbf{X}_k\mathbf{T}_k\mathbf{T}_k'\mathbf{X}_k'$ and distance matrices $\mathbf{D}_k$ and hence the links with multidimensional scaling and INDSCAL.

Unification and, I hope, a better understanding is achieved by concentrating on the common features of this group of methods; however, important differences remain.