7

Comparing Groups of Variables and Indscal Model

The previous chapters present detailed tools for analysing relationships between groups of variables, both from the point of view of the variables (between-group correlations) and the individuals (comparison of partial clouds). However, when there is a large amount of data (many groups, each with a lot of variables, and a large number of individuals), more general tools are required, at least at first, in order to answer questions such as:

- Given two groups of variables, can we consider them to be related? In other words: overall, are the two associated partial clouds similar?
- As is the case for representations of individuals and variables, is it possible to generate graphs in which each group is represented by a point, with the proximity of points j and h indicating a relation/similarity between groups j and h?

7.1 Cloud N_J of Groups of Variables

The raw data associated with a group j of variables makes up table X_j. Generally, there is no correspondence between the columns of the different X_j and it is not possible to compare the X_j tables directly.

The idea of comparing two groups of variables from their partial cloud (see Chapter 5) suggests representing a group by its matrix of between-individual distances. These matrices have the same dimensions from one group to another and their entries correspond pairwise: it is therefore possible to compare them directly.

From another point of view, it is legitimate to represent a cloud of individuals by the matrix of their scalar products (between individuals; denoted XX' in Section 1.5.1) with which it is associated: indeed, diagonalising this matrix makes it possible to represent the cloud of individuals perfectly on its principal axes, with the same weight attributed to each individual (see Section 1.5.3). These two matrices are closely linked: the matrix of scalar products can be obtained by performing a dual centring of the matrix of squared distances.

The notations are as follows: $\langle i, l \rangle$ indicates the scalar product between individuals i and l; $d\,(i, l)$ the distance between i and l; $d^2\,(i, .)$ (and, respectively,