Two-way Component Models of Fuzzy Data

Pierpaolo D'Urso
Department of SEGeS
University of Molise
durso@unimol.it

Paolo Giordani
Departiment of Statistics
University of Rome
paolo.giordani@uniroma1.it

Henk A.L. Kiers
Heymans Institute
University of Groningen
h.a.l.kiers@ppsw.rug.nl

Keywords: Fuzzy data, Principal Component models.

Abstract

In substantive applications, it frequently happens that observations cannot be described precisely but only approximately. In such cases, each observation is described by an interval of values, rather than by a single score. Such data are called fuzzy data. Notice that we are concerned with a kind of uncertainty which is different from randomness, and that is referred to as fuzziness. In fuzzy data sets instead of scores we have fuzzy numbers to denote each observation. In particular, an LR fuzzy number is denoted by $f \equiv (m,l,r)$, where m is the center of the fuzzy number, l and r are the left and right spreads, so that the full interval covered by a fuzzy number runs from m-l to m+r, and L and R are functions that give differential weights to the values in the intervals, respectively, to the left and to the right of the center. In many occasions, as well as for traditional single valued data, it is desirable to compress fuzzy data losing relevant information as little as possible. To summarize the data, we suggest to model the centers by a Principal Component model as follows:

$$\mathbf{M}^* = \mathbf{A}_M \mathbf{B}', \tag{1}$$

where $\mathbf{M}^*$ denotes the matrix with estimated centers, $\mathbf{B}$ contains the component loadings and $\mathbf{A}_M$ contains the component scores. The loadings express relations between variables in the sense that they define components that, for instance, relate to subgroups of variables. In the model we consider here, we assume that these relations between the variables have the same structure (e.g., the same grouping of variables) for the spreads as for the centers. Hence, we model the left and right spreads using the same loadings matrix as for the centers, as follows:

$$\mathbf{L}^* = \mathbf{A}_L \mathbf{B}', \quad \mathbf{R}^* = \mathbf{A}_R \mathbf{B}', \tag{2}$$

where $\mathbf{L}^*$ and $\mathbf{R}^*$ are respectively, the estimated left and right spreads, and $\mathbf{A}_L$ and $\mathbf{A}_R$ are the component score matrices for the left and right spreads. In this way, the model recovers a compromise structure between centers and spreads.

When the fuzzy data are symmetric ($l = r$), (2) should be replaced by

$$\mathbf{S}^* = \mathbf{A}_S \mathbf{B}', \tag{3}$$

where $\mathbf{S}^*$ denotes the estimated spreads and $\mathbf{A}_S$ is the component score matrix pertaining to the spreads as in Giordani and Kiers (2003).

The method will be illustrated by means of an application to real fuzzy data.

References

Giordani, P. and Kiers, H.A.L. (2003) Principal Component Analysis of symmetric fuzzy data. Computational Statistics and Data Analysis, in press.