

# Nonparametric inference in categorical principal components analysis (CATPCA)

Mariëlle Linting	Patrick J.F. Groenen	Jacqueline J. Meulman
Data Theory Group	Econometric Institute	Data Theory Group
Leiden University	Erasmus University Rotterdam	Leiden University
The Netherlands	The Netherlands	The Netherlands
linting@fsw.leidenuniv.nl	groenen@few.eur.nl	meulman@fsw.leidenuniv.nl

**Keywords:** Principal components analysis, CATPCA, categorical data, bootstrap, permutation tests.

## Abstract

Categorical principal components analysis (CATPCA), available in the Categories module in SPSS (Meulman, Heiser, & SPSS, 1999), reduces data to a number of principal components by nonlinear transformation of the variables. Although CATPCA can be seen as an exploratory technique, it need not be deprived of confirmatory diagnostics. These diagnostics can be either in the form of stability measures for various estimated parameters, or in the form of  $p$ -values with respect to a particular null conjecture.

Linting, Groenen, and Meulman (2003) investigated the use of the bootstrap to establish and visualized the stability of the component loadings, object scores, and transformed variables from the CATPCA solution.

In the present paper, we will obtain additional confirmatory diagnostics by permutation tests. The basic idea of these tests is to compare an observed value for a particular goodness-of-fit statistic, for example an eigenvalue, to a distribution of this statistic obtained by well-defined random permutations of the data. We will consider various ways to generate the permutation distribution, for example by permuting objects within each variable independently (to establish the significance of the eigenvalue), and by permuting objects within a single variable, keeping the other variables fixed (to establish the significance of the variance accounted for with respect to the permuted variable).

We will demonstrate these procedures on data from the National Institute of Child Health and Human Development (NICHD Early Childhood Research Network, 1996).

## References

- Linting, M., Groenen, P.J.F., & Meulman (2003). Stability of Categorical Principal Components Analysis: An empirical study. Paper submitted for publication.
- Meulman, J.J., Heiser, W.J., & SPSS. (1999). *SPSS categories 10.0*. Chicago: SPSS.
- NICHD Early Childhood Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, 11, 269-306.