

Analysis of an European Union election using principal component analysis

Paulo Canas Rodrigues^{1,3}, Ana T. Lima²
and Vera Jesus¹

¹*Nova University of Lisbon, Portugal,*

²*Polytechnical Institute of Coimbra, Portugal*

³*Agricultural University of Poznań, Poland*

Abstract

Many data sets present a high number of dimensions since the number of variables is also high. However, it is difficult for a human being to visualize more than three-dimensional spaces.

Principal Component Analysis (PCA) is a multivariate technique used to reduce multidimensional data sets to a lower number of dimensions. Moreover, PCA able us to identify variables inherent to the initial structure of a data set, with a physical meaning (Jolliffe, 2002).

This work aims boarding a practical point of view with an application to a real data set about the results of the European Parliament election in 2004. As pointed out by Katz and King (1999), multiparty data can be seen as Compositional Data (Aitchison, 1986). We present a comparative study between the results of PCA, Crude PCA and Logcontrast PCA (Aitchison, 1983; Kucera and Malmgren, 1998). From the results of these three approaches we point out that the one which produces more clearer results is the Logcontrast PCA.

Keywords

Compositional data, Principal component analysis, Crude PCA, Logcontrast PCA, Electoral data.

References

- Aitchison, J. (1986). *The Statistical analysis of compositional data*. London: Chapman and Hall.
- Aitchison, J. (1983). Principal component analysis of compositional data. *Biometrika* 70, 57–61.
- Jolliffe, I.T. (2002). *Principal component analysis*. New York: Springer.
- Katz, J.N. and G. King (1999). A statistical model for multiparty electoral data. *American Political Science Review* 93, 15–32.
- Kucera, M. and B.A. Malmgren (1998). Logratio transformation of compositional data - a resolution of the constant sum constraint. *Marine Micropaleontology* 34, 117–120.