Automatic construction of graphical outputs of common multivariate analyses with a special reference to predictive biplots

M. Rui Alves[1,2] and **M. Beatriz Oliveira**[1]

[1] REQUIMTE, Serviço de Bromatologia, Faculdade de Farmácia, Universidade do Porto, R. Aníbal Cunha, 164, 4099-030 PORTO, Portugal

[2] ESTG / Instituto Politécnico de Viana do Castelo, Av. Atlântico, s/n, 4900–348 Viana do Castelo, Portugal, mruialves@estg.ipvc.pt

Predictive biplots [1] have several aspects which are very important in multivariate analyses, mainly because they enable an easier interpretation of multivariate analyses outputs by relating sample configurations to the initial, declared variables, without losing the modulation aspects so characteristic of these statistical methodologies.

The disadvantages of biplots are mostly related to software limitations causing difficulties in obtaining the final graphical solutions and to difficulties in deciding how relevant a given biplot axis is and how many plots are necessary to accurately describe available data.

The latter difficulties are also experienced by users of normal statistical methodologies when it comes to interpret multivariate outputs, i.e., how important initial variables are to explain latent variables and how many dimensions are necessary to describe data. One problem is that there is a huge degree of freedom in the way this can be done, and easy methods to help taking these decisions are very important, mainly for inexperienced statisticians.

This presentation follows our previous works on the subject [e.g., 2,3] and demonstrates how a decision on whether a biplot axis is or not drawn in two-way plots can be carried out automatically [4]. The method, based on the predictive power of variables and on a specially defined tolerance value, enables R to evaluate the interest of each of the initial variables and draw the predictive biplots automatically. Also, the final number of plots is also automatically decided by R. Since the method may be made fully automatic, inexperienced users can take profit of all the R facilities, carrying out multivariate analysis and final interpretations, and at the same time being protected from common over-fitting problems, difficulties in interpretations of multivariate outputs, etc.

The methods can be applied to several statistical methodologies, and examples are provided for principal components analysis and canonical variate analysis (in chemistry) and three-way Tucker-1 common loadings analysis (in the field of sensory analysis).

It is also shown that the method devised to produce biplots in an automatic way can be diverted to common outputs, enabling R to provide users of multivariate analyses with automatic interpretations of results, including the decision on the number of dimensions to retain and their respective interpretations.

[1] Gower, J.C.; Hand, D.J. Biplots. Chapman and Hall: London, 1996.

[2] Alves, M.R.; Oliveira, M.B. (2005). Monitorization of Consumer and Naïf Panels in the Sensory Evaluation of Two Types of Potato Chips by Predictive Biplots Applied to Generalized Procrustes and 3-way Tucker-1 Analysis. . *Journal of Chemometrics*, **19**: 564–574.

[3] Alves, M.R.; Oliveira, M.B. (2006). R algorithms for the calculation of markers to be used in the construction of predictive and interpolative biplot axes in routine multivariate analyses. User!2006 – The R user Conference, Wirtschaftsuniversität Wien.

[4] Alves, M.R.; Oliveira, M.B. A method for the production of predictive biplots applied to multivariate outputs with a complete R algorithm enabling fully automatic processes and user defined parameters. Submitted.