

Workshop on:

CLUSTER ANALYSIS OF VARIABLES: CLV METHOD.

Organized by:

E. Vigneau, E.M. Qannari

Cluster analysis of variables is of paramount interest since it can serve several purposes among which we single out the selection of a subset of relevant variables or the investigation of the structure of a dataset. From this perspective, cluster analysis of variables can be seen as an alternative of Principal components Analysis followed by a rotation strategy to enhance the interpretation of the principal components.

We have designed *CLV* method (Cluster analysis of variables around Latent Variables) in the same vein as *VARCLUS* method, which is a popular method for cluster analysis of variables particularly among practitioners of *SAS* environment. However, in comparison to *VARCLUS*, *CLV* method is much simpler, more straightforward and more efficient. Furthermore, it offers a flexible framework which makes it versatile since, as it will be demonstrated, it can be adapted to various settings including the investigation of the relationships between two sets of variables. Indeed, *CLV* method is based on the optimization of simple criteria that aim at determining simultaneously clusters of variables and, associated with each cluster, a latent variable (or component) is determined. In practice, two kinds of clusters are of interest: (i) directional clusters where correlated variables are lumped together regardless of the sign of their correlation coefficients; (ii) local clusters where only those variables that are positively correlated are set in the same cluster.

Originally dedicated to the cluster analysis of quantitative variables, *CLV* method has been extended into various directions.

- Cluster analysis of variables taking account of external information on the individuals and/or the variables. In this case *CLV* analysis is carried out so that the group latent variables can be readily interpreted in terms of the external information. This situation is commonly encountered in the context of preference mapping where one might be interested in clustering a panel of consumers taking account of sensory properties of the products and/or the characteristics of the consumers (e.g. socio-demographic, usage and habits...).

- Cluster analysis of variables while discarding uninformative variables or atypical variables. In other words, this makes it possible to achieve a "cleansing" of the dataset at hand and a clustering of the remaining variables. For this purpose, two strategies are proposed. The first strategy consists in including a specific cluster (or

“trash” cluster) that contains those variables that are deemed to be linked to noise only. The second strategy consists in constraining the latent variables associated with the various clusters to be sparse in that sense that they are related to the relevant variables only. In practice, both strategies lead to very similar outcomes.

A package, available under *R* environment, named "*ClustVarLV*" offers an implementation of the *CLV* method and its various developments discussed above. Several illustrations of the strategies of analysis and the handling of the *R*-package will be presented. These illustrations mainly pertain to the field of sensory studies, but some datasets used for the illustration of the methods of analysis relate to the so-called “-omics” data.

References

- Vigneau E., Chen M., Qannari E.M. (in press). *ClustVarLV: An R Package for the Clustering of Variables around Latent Variables*. *RJournal*.
- Vigneau, E. and Chen, M. (2015). *ClustVarLV: Clustering of variables around Latent Variables*. *R package version 1.4.1*.
- Vigneau, E., Qannari, E. M., Navez, B. and Cottet, V. (2015). Segmentation of consumers in preference studies while setting aside atypical or irrelevant consumers. *Food Quality and Preference*, 54: 47-63.
- Vigneau, E., Charles, M. and Chen, M. (2014). External preference segmentation with additional information on consumers: a case study on apples. *Food Quality and Preference*, 32: 83-92.
- Vigneau, E., Qannari, E. M., Punter, P. H. and Knoop, S. (2001). Segmentation of a panel of consumers using clustering of variables around latent directions of preference. *Food Quality and Preference*, 12(5-7): 359-363.