

# Hierarchical Clustering based on Principal Components

Guillaume Le Ray<sup>1</sup>, Quentin Molto <sup>1,2</sup>, François Husson<sup>2</sup>

1. Student from Agrocampus Ouest Majored in Applied Statistics.
2. Professor of Statistics from Agrocampus Ouest.

**Keywords:** Hierarchical clustering, Factorial Analysis, program, FactoMineR

The HCPC program for Hierarchical Clustering on Principal Components is dedicated to the Clustering especially after a Factorial Analysis. Factorial Analysis and clustering are complementary tools to explore data. Factorial Analysis enables to remove the last components, which means remove the noise and get the clustering more robust. The program structure is the following one: a Factorial Analysis to preserve the signal and delete the noise in the data, a Hierarchical Clustering, a choice of the number of clusters, a partition in clusters, a consolidation by K-means and a description of the clusters.

There is in a wide range of options, uses, results and graphical representations. It includes all the option of the function *agnes* (pachage *cluster*). As it is used in the Factorial Analysis the default distance is the *Euclidean* distance and the aggregation criterion for the Hierarchical clustering is the *Ward* indice. The program suggests a level of cutting calculated on the inertia gains between two partitions. You can choose either to follow this suggestion or to cut the tree on another level by clicking on the tree. The consolidation with *Kmeans* is optional to get more robust clusters. It returns a more optimal partition from the point of view of the inertia criterion. At this process, it adds interpretation tools like the descriptions of the clusters with *catdes* by the variables, the axes and description of the clusters by the individuals (the closest from the cluster barycentres and the most specific of each cluster).

The principal benefit of this function is the multiplicity of graphical representations. It is possible to plot the tree, factor maps with individuals colored by clusters, or both together, in 2D or 3D, where the tree is above the map. This function will be included in the *FactoMineR* package.

## References

Jérôme Pagès, Brigitte Escofier(2008). Analyse Factorielles Simples et Multiples 4ème édition. p 31-63