



## Benchmarking cluster algorithms for ordinal survey data

Dominik Ernst

University of Natural Resources and Life Sciences, Vienna, Austria - Dominik.Ernst@boku.ac.at

Friedrich Leisch\*

University of Natural Resources and Life Sciences, Vienna, Austria - Friedrich.Leisch@boku.ac.at

In marketing research survey data of consumer preferences are often clustered in order to identify market segments. In many cases these data sets are on ordinal scale, e.g. ratings on scales from “very important” to “not important at all”. Most cluster algorithms on the other hand are designed for metric data, from the standard  $k$ -means algorithm to mixtures of Gaussians. We present model-based cluster algorithms for ordinal data and compare their performance on both artificial and real world data sets. Special emphasis is given to resampling methods for cluster validation like subsampling or bootstrapping. Resampling helps to assess the stability of single clusters and complete partitions as well as to identify the best number of clusters. Due to the increasing availability of parallel processing even on standard laptops and desktops these simulation-based approaches can now be used in everyday cluster analysis applications. All methods shown have been implemented in R and are freely available.

**Keywords:** cluster analysis, ordinal data, bootstrap, parallel processing, R