

P82

**COMPARATIVE STUDY OF HIERARCHICAL CLUSTERING**Pris.R. Carvalho<sup>a</sup>, C.S. Munita and A.L. Lapolli<sup>a</sup> [prii.ramos@gmail.com](mailto:prii.ramos@gmail.com)

Nuclear and Energy Research Institute, São Paulo, Brazil

In archaeological studies several analytical techniques are used to study the chemical and mineralogical composition of many materials of archaeological origin, generating a large data set. Thus, the multivariate statistical methods become indispensable for the interpretation of the results. These multivariate techniques, unsupervised and supervised, are accompanied by modern computational programs, which provide visualization and interpretation. Several methods have been used, such as cluster analysis, discriminant analysis, principal component analysis, among others. However, the most used is cluster analysis. The purpose of cluster analysis is to group the samples based on similarity or dissimilarity. The groups are determined in order to obtain homogeneity within the groups and heterogeneity between them. The literature presents many methods for partitioning of data set, and is difficult choose which is the most suitable, since the various combinations of methods based on different measures of dissimilarity can lead to different patterns of grouping and false interpretations. Nevertheless, little effort has been expended in evaluating these methods empirically using an archaeological data set. In this way, the objective of this work is make a comparative study of the different cluster analysis methods and to identify which is the most appropriate. For this, the study was carried out using a data set of the Archaeometric Studies Group from IPEN-CNEN/SP, in which 45 samples of ceramic fragments from three archaeological sites were analyzed by instrumental neutron activation analysis (INAA) which were determined the mass fraction of 13 elements (As, Ce, Cr, Eu, Fe, Hf, La, Na, Nd, Sc, Sm, Th, U). The methods used for this study were: single linkage, complete linkage, average linkage, centroid and Ward. The comparison was done using the cophenetic correlation coefficient and according these values the average linkage method obtained better results. A script of the statistical program R was created to obtain the cophenetic correlation coefficient. The purpose of this script is to facilitate the statistical study of researchers who do not have much familiarity with statistical programs. Therefore, the researcher can easily check which method is most appropriate for your data set.

---