

Quantitative Methods of Standardization in Cluster Analysis: Finding Groups in Data

A.L. Nogueira¹, C.S. Munita^{2*}, A. F. Costa³, D. M.C. Gomes⁴ and R. Kipnis⁵

¹Coordenação e Licenciatura em Física, Universidade Federal de Sergipe, Sergipe, Brasil, 049400-000.

²Nuclear and Energy Research Institute, IPEN-CNEN/SP, Sao Paulo, Brazil, 05501-000.

³Museu Nacional, Universidade Federal de Rio de Janeiro, Rio de Janeiro, RJ, 020940-040.

⁴Museu Nacional, Universidade Federal de Rio de Janeiro, Rio de Janeiro, RJ, 020940-04.

⁵Scientia Consultoria Científica, São Paulo, SP, 05591-045.

(* Corresponding Author Email: camunita@ipen.br)

The archaeological study of ceramics using multi-elemental analytical techniques such as instrumental neutron activation analysis (INAA) is important in archaeology due to its potential to identify the raw materials used in their manufacture, and subsequently help to infer the degree of interaction among ancient communities. Several multivariate statistical methods are used in chemical composition data analysis, such as principal component analysis, cluster analysis and discriminant analysis. Pattern recognition approaches are divided into unsupervised learning and supervised learning. Cluster analysis is a technique for pattern recognition and is an unsupervised approach. When applying cluster analysis, raw data, or actual measurements, are not used directly. Thus, a problem that arises during cluster analysis involves the decision of whether or not to standardize the input variables before calculating measures of distance. The standardization of variables is necessary in cases where the measure of dissimilarity, such as the Euclidean distance, is sensitive to differences in the magnitudes or scales of the input variables. The aim of this paper is to assess the impact, and evaluate the usefulness of three standardization techniques in determining the number of clusters for a data set of 140 ceramic fragments from eight archaeological sites from the upper Madeira river, Rondônia, Brazil, in which Na, K, La, Sm, Yb, Lu, Sc, Cr, Fe, Co, Zn, Rb, Cs, Ce, Eu, Hf, Ta, and Th mass fractions were determined by INAA.