Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

CORRESPONDENCE ANALYSIS AND SERIATION

BY ROUH-JANE CHOU
National Tsing Hua University

The first nontrivial correspondence analysis (CA) solution of a two-way contingency table gives scores of the row and column categories so that the correlation between the row variable and the column variable is maximized. Hence it is natural to order the categories by the scores. In this paper, the appropriateness of this technique is investigated. Sufficient conditions are given. Sampling theories, when the data fail to satisfy the conditions because of the presence of random errors, are studied. As to whether one should use one or two CA solutions for the ordering, simulation study is used to show their difference.

1. Introduction. Ordering things in time has been the interest of archaeologists since a century ago when Petri tried to seriate chronologically some 900 graves by means of the numbers of various potteries in them. Mathematicians had developed methods of seriation since then. Among them, D.G. Kendall had made significant contributions in the sixties and seventies. Kendall (1963) proposed a model for the graves and made statistical inference about the model. In Kendall (1971a), a measure, called common content, of similarity between graves was suggested. An algorithm for achieving a right ordering of the graves satisfying the condition of being pre-Q was also given. Then in Kendall (1971b), the method of multidimensional scaling was introduced to seriate objects, given their similarity matrix.

Recently, the method of correspondence analysis (CA) for contingency tables has gained more and more attention, especially in its use of ordering objects (Greenacre (1984), Hill (1974) and Schriever (1986)). The idea is to order the rows by the elements of the first non-trivial eigenvector of a matrix obtained from the matrix of frequencies or proportions. Though this method looks suitable intuitively, a mathematical justification, however, is needed.

In this paper, we study the appropriateness of the first non-trivial correspondence analysis solution as a rule for seriating the rows of a Q-matrix. In Section 2, method of CA, ideas of Q-matrix and total positivity (TP) are

---

AMS 1991 Subject Classifications: Primary 62H20.

Key words and phrases: Contingency table, correspondence analysis, total positivity.