Grade correspondence analysis applied to contingency tables and questionnaire data

Wiesław Szczesny [1)]

[1)] Institute of Computer Science, Polish Academy of Science, Ordona 21, 01-237 Warsaw, Poland. Department of Econometric and Computer Science, Warsaw Agricultural University, Nowoursynowska 166, 02-787 Warsaw, Poland

Abstract:

An alternative approach methodology for Correspondence Analysis is presented. This approach, called Grade Correspondence Analysis (GCA), utilizes Spearman's rho to detect underlying associations and trends. Two examples are presented using: (1) a contingency table (Heuer's suicide data) with cause of death, gender, and age; and (2) a survey questionnaire (data matrix) concerning employment, personal economics, computer skills, and disability level of handicapped computer specialists in Poland. GCA uses a search strategy (multi-starts / random starts) to detect trends (not forced to be orthogonal) among rows and columns. (A similar strategy permits the determination of significance levels.) Results are discussed using measures of the "representativness" of the trends, as well as measures of their "regularity". Visualization of trends (as well as outlier trend detection) is via the concept of "overrepresentation" maps. Survey data may be measured on any non-negative scale. Meaningful disjoint aggregation (or division) of sub-populations and variables are possible. This paper is written for the practitioner and includes a "grade" concepts example in an appendix. There is also, however, an appendix with GCA theory relating to: grade distributions; local maxima of Spearman's rho and their representativness, regularity and regions of attraction; total positivity of order 2 (TP2); similarity measures; suitable "random references" for the determination of significance levels; and the application of GCA to non-negative data matrices.

Keywords: