

---

## Editorial

---

### Sofia Anastasiadou

Department of Early Childhood Education,  
University of Western Macedonia,  
GR-53100 Florina, Greece  
Email: anastasiadou@uowm.gr

### Giannoula Florou

Department of Accounting and Finance,  
Eastern Macedonia and Thrace Institute of Technology,  
GR-65404 Kavala, Greece  
Email: gflorou@teiemt.gr

### Angelos Markos\*

Department of Primary Education,  
Democritus University of Thrace,  
GR-68100 Nea Hili, Alexandroupoli, Greece  
Email: amarkos@eled.duth.gr  
\*Corresponding author

**Biographical notes:** Sofia Anastasiadou is a Professor at the Department of Early Childhood Education of University of Western Macedonia, Greece. Her research interests pertain to methodology and statistics in education and the social sciences. She has published several papers in international journals, conference proceedings and edited books.

Giannoula Florou is a Professor at the Department of Accounting and Finance of Eastern Macedonia and Thrace Institute of Technology, Greece, and the President of Greek Society of Data Analysis. Her research interests are in the field of business statistics, quantitative methods in finance, data analysis and clustering. She has considerable experience in management of European funded programs and has co-authored several papers in international journals and conference proceedings.

Angelos Markos is an Assistant Professor (tenured) of Data Analysis in the Social Sciences at the School of Education of Democritus University of Thrace, Greece. He is a board member of the Greek Society of Data Analysis. His research interests are in the field of multivariate data analysis, dimension reduction and clustering, particularly correspondence analysis and related methods. He has co-authored several papers in international journals, conference proceedings and edited books.

Correspondence analysis (CA) as a method of multivariate data visualisation has an almost 60-year history and has found use in almost every field of multivariate research. CA also has intimate connections with other multivariate areas, notably distance-based cluster analysis, principal component analysis, discriminant analysis, and analysis of variance. Because CA is applicable to the most basic of data types, namely categorical data, it can be applied to almost any mixed-type multivariate dataset, thanks to ingenious ways of recoding data to categorical scales. This special issue aims at providing a set of selected articles which describe theoretical and empirical results on CA and classification methods that are applied in realistic scenarios related to biological sciences, business research and educational research. All the papers composing this special issue have been carefully selected from the articles presented at the 9th PanHellenic Conference of Data Analysis with International Participation (PCDA9), a biennial conference for disseminating recent advances in the field of multivariate data analysis. The PCDA9 was held in Thessaloniki, Greece, on September 28–30, 2017.

An overview of the selected articles is given next:

- In ‘Comparison of hierarchical clustering methods for binary data from molecular markers’, Emmanouil D. Pratsinakis, Symela Ntoanidou, Alexios Polidoros, Christos Dordas, Panagiotis Madesis, Ilias Eleftherohorinos and George Menexes review the distance metrics and linkage criteria used for hierarchical clustering of binary data in the field of molecular markers. A total of 162 different hierarchical clustering approaches were compared using binary data from molecular markers applied to five populations of the wild mustard *Sinapis arvensis* species. Cluster solutions were validated using external criteria (geographical area and herbicides resistance). The results showed that the popular UPGMA linkage criterion combined with the squared Euclidean distance is not a panacea for clustering binary data derived from molecular markers. Thirty-six alternative hierarchical clustering approaches could be effectively used.
- In ‘Assessment of the awareness of Cypriot accounting firms level concerning cyber risk: an exploratory analysis’, Stratos Moschidis, Efstratios Livanis and Athanasios C. Thanopoulos employ multiple CA to investigate the level of awareness among accountancy firms about cyber-related risks. The data come from a survey of Cypriot professional accountants members of Institute of Certified Public Accountants of Cyprus (ICPAC). Results reveal that many market participants assume that cyber security is mainly an information technology issue rather than a cross-organisational risk management issue. The study implications with regard to the management of cyber risk and the devaluation of its financial dimensions are discussed.
- In ‘Sequential dimension reduction and clustering of mixed-type data’, Angelos Markos, Odysseas Moschidis and Theodore Chadjipantelis present and evaluate two-step strategies for dimension reduction and clustering of mixed-type data, i.e., datasets with a mixture of continuous and categorical variables. The first step involves the application of principal component analysis or CA on a suitably transformed matrix to retain as much variance as possible in as few dimensions as possible. In the second step, a partitioning or hierarchical clustering algorithm is applied to the object scores in the reduced space. The results of a benchmarking study showed that sequential dimension reduction and clustering outperforms

alternative methods when categorical variables are more informative than continuous with regard to the underlying cluster structure.

- In 'A comparative evaluation of dissimilarity-based and model-based clustering in science education research: the case of children's mental models of the Earth', Dimitrios Stamovlasis, Julie Vaiopoulou and George Papageorgiou employ dissimilarity-based clustering (partitioning around medoids) and model-based clustering (latent class analysis) to classify children response patterns with regard to their mental representation of the earth. The study aims to contribute to an ongoing debate in cognitive psychology and science education research between two antagonistic theories on the nature of children's knowledge, that is, the coherent versus fragmented knowledge hypothesis. Both clustering methods, despite their diverse theoretical underpinnings, converge to similar results providing empirical evidence in favour of the fragmented knowledge hypothesis.