Stat. Meth. & Appl. (2007) 15:267–269
DOI 10.1007/s10260-006-0040-6

EDITORIAL

Special issue on robust multivariate analysis and classification

Marco Riani · Andrea Cerioli · Bruno Chiandotto

Published online: 21 December 2006
© Springer-Verlag 2006

It is now widely recognized that most classical statistical techniques are not resistant in the presence of outliers. Correspondingly, the development of highly robust and efficient statistical methods has become a goal of paramount importance in both theoretical and applied research. The demand for such methods has been driven by the increasing availability of data in almost any area of scientific research. These data sets are not only becoming larger in size, but also in complexity. The extraction of essential features and the discovery of structures and relations in complex data sets must not break down when atypical observations are present. In addition, there is a need for the development of effective diagnostics that can help to pinpoint these outliers. With many variables at hand, outlying observations can be hard to detect. Outliers need not necessarily be associated with "gross" contamination errors, but may instead contain valuable information. An example is the indication of the existence of several populations instead of one. While robust statistical methods and diagnostic tools are well established for studying data sets under simple univariate models, this is not the case for more complicated multivariate situations.

M. Riani (✉) · A. Cerioli
Department of Economics—Section of Statistics and Computing,
University of Parma, via Kennedy 6, 43100 Parma, Italy
e-mail: mriani@unipr.it

A. Cerioli
e-mail: andrea.cerioli@unipr.it

B. Chiandotto
Department of Statistics "G. Parenti", University of Florence,
Viale Morgagni 59, 50134 Firenze, Italy
e-mail: chiandot@ds.unifi.it

To promote research in these areas, from January 2004 to December 2006 the European Science Foundation supported the network "*Statistical Analysis of Complex Data with Robust and Related Statistical Methods* (SACD)". The Steering Committee of the SACD network was chaired by Prof. Christophe Croux (KU Leuven, Belgium) and included Prof. Frank Critchley (The Open University, UK), Prof. Peter Filzmoser (Vienna University of Technology, Austria), Prof. Ursula Gather (University of Dortmund, Germany), Prof. Hannu Oja (University of Tampere, Finland), Prof. Daniel Peña (Universidad Carlos III de Madrid, Spain), Prof. Ana Pires (Instituto Superior Técnico, Lisboa, Portugal), Prof. Marco Riani (University of Parma, Italy), Prof. Peter Rousseeuw (University of Antwerp, Belgium) and Prof. Stefan Van Aelst (Ghent University, Belgium). Among the network activities, the workshop "Robust Classification and Discrimination with High Dimensional Data" was held in Florence from January 25 to 28, 2006. The goal of the workshop was to provide a forum for new developments and applications of robust statistical methods for classification and discrimination, with emphasis on techniques suited to high-dimensional and complex data. A call for papers on these topics was circulated within several major statistical lists after the conference. As a consequence of the call for papers, 17 papers were submitted to *Statistical Methods and Applications*. All the papers were refereed following the usual peer-review process, which resulted in at least two reports for each manuscript. Six of the accepted papers are included in this special issue, together with a discussion paper sketching the state-of-the-art of robust statistics. The other accepted papers will appear in subsequent issues of *Statistical Methods and Applications*.

The first paper in this special issue, by Prof. Stephan Morgenthaler, surveys the development of robust statistical methods over the last few decades. It provides a history of robust thinking within the statistical community and some personal views on future directions, emphasizing the challenges provided by the analysis of large and complex data sets. One of these challenging topics, namely the development of robust methods for correlated observations, is covered in two of the subsequent research papers. Morgenthaler's survey is discussed, in seven distinct contributions, by leading experts in the field. All the discussants add further insights that can help a better understanding of the many facets of robust statistics, covering both foundational and technical aspects.

The research papers contained in this issue broadly cluster in two main areas. The first group of three papers addresses the fundamental issue of obtaining robust estimates of location and scale from multivariate data. The contribution by Boente, Critchley and Orellana deals with two estimators for the proportional principal component model: the plug-in estimator and the projection-pursuit estimator. The paper by Nevalainen, Larocque and Oja develops a multivariate weighted spatial median estimator for clustered data, taking account of the possible inter-subject correlation within the same cluster. Roelant and Van Aelst study a multivariate extension of the median which yields simultaneously a location estimate and a scatter estimate. The influence (or partial influence) function is an important tool for investigating the properties of such robust estimators.

Robust estimation of multivariate location and scale obviously has important implications also for the classification problems underlying the second group of research papers. In this area, Todorov makes use of the Minimum Covariance Determinant estimator to construct a robust version of the Wilks' lambda statistic useful for variable selection in linear discriminant analysis. The paper by Crosilla, Visintini and Sepic has a more applied flavour than the other contributions in this issue. It aims at developing a robust classification procedure for 3D objects surveyed with high-density laser scanning measurements, by combining nonparametric techniques with a block forward search approach that allows for possible spatial autocorrelation between laser measurements.

The paper by Gottard and Pacillo falls outside the two main clusters outlined above but well within the scope of this special issue. It focuses on robust analysis of the conditional independence structure of a multivariate distribution, by investigating the impact of different kinds of contamination on model selection procedures for graphical Gaussian models.

As a concluding remark, we are beholden to the Editor of *Statistical Methods and Applications*, Prof. Maurizio Vichi, for giving us the opportunity to collect together so many different and interesting ideas in the field of robust multivariate statistics. We would also like to thank the many distinguished scholars who have acted as referees for this special issue. Their careful and thoughtful work is gratefully acknowledged. Finally, thanks to all, both authors and referees, for bearing in patience our persistent reminders about deadlines: producing a special issue of a journal is not only a matter of good science but also of limited available time!