Spring Meeting of the VOC

Robust Methods

May 13, 2011

University of Antwerp, City Campus (R.225)

Program:

10.30	Registration and Coffee
10.45	Stefan Van Aelst: Concepts and methods in robust statistics
11.45	Andreas Alfons: Robust variable selection with application in the social sciences
12.30	Lunch
13.30	Marco Riani: The forward search: theory and data analysis
14.30	Wobbe Zijlstra: Robust Mokken scale analysis by means of the forward search algorithm
15.15	Tea + VOC annual member meeting
15.45	Jorn de Haan: Nonparametric transformations in the analysis of microarray data
16.30	Drinks

Registration details for the Spring Meeting:

Those who would like to participate are welcome and are kindly requested to register by sending an e-mail to meeting@voc.ac with subject 'Registration Spring Meeting 2011' and including your name and affiliation in the body of the e-mail. Participation is free, lunch is available for 10 Euros and must be requested upon registration. Registration deadline: May 10th.

Abstracts for the Spring Meeting

Stefan Van Aelst (Ghent University): Concepts and methods in robust statistics

We introduce the standard contamination model underlying most of the developments in robust statistics and discuss its properties and limitations. Standard measures of robustness are then introduced such as the influence function and the breakdown point. In this talk we focus on two key settings that form the basis of many statistical techniques, which are the linear regression and multivariate location and scatter models. We review some well-known classes of robust estimators such as M, S and MM-estimators and least trimmed residual distance estimators (LTS and MCD). For each of these classes we discuss their robustness properties as well as other important properties such as equivariance, consistency and efficiency. We then discuss outlier detection and robust inference methods. We illustrate these methods with applications in multivariate analysis, such as regression, principal components analysis, (multivariate) ANOVA, discriminant analysis and clustering.

Stefan Van Aelst *is associate professor of mathematical Statistics at the Faculty of Sciences of Ghent University. He obtained a PhD degree in mathematics from the University of Antwerp (2000). His research interests are robust methods, inference and model selection. He has published in many statistical journals, including Journal of the American Statistical Association, Annals of Statistics, Journal of the Royal Statistical Society B, and Statistical Science. Currently, he is associate editor of Computational Statistics and Data Analysis and Journal of Statistical Planning and Inference. Since 2011 he is Vice-President of the Belgian Statistical Society.*

Andreas Alfons (Katholieke Universiteit Leuven): Robust variable selection with application in the social sciences

Motivated by applications in the social sciences, a robust variable selection procedure has been developed. The procedure combines sequencing the most informative candidate predictors with a strategy to reduce the number of selected variables to a necessary minimum. The latter is crucial in the context of social sciences for better interpretability. In addition, strong dependencies among the regressor variables need to be eliminated such that the explanatory variables describe complementing effects. The fulfillment of these two requirements will therefore be called context-sensitivity and the resulting strategy for variable selection can be considered a tradeoff between quality of the model and interpretability. In practical applications, the proposed procedure led to highly interpretable models. Furthermore, the performance of the procedure is assessed by means of simulation. This simulation study verifies that primarily only variables with potentially new information are included in the resulting model.

Reference:

Alfons, A., Baaske, W.E., Filzmoser, P., Mader, W., & Wieser, R. (2011). Robust variable selection with application to quality of life research. *Statistical Methods & Applications, 20*(1), 65-82.

Andreas Alfons *is a postdoctoral research fellow at Katholieke Universiteit Leuven, Faculty of Business and Economics. He received a Master's degree in Applied Mathematics and a PhD degree in Statistics from Vienna University of Technology (Austria), where he also held a research assistant position at the Department of Statistics and Probability Theory. His research interests include robust statistics, multivariate data analysis, and the development of statistical software.*

Marco Riani (Parma University): The forward search: theory and data analysis

The Forward Search is a powerful general method, incorporating flexible data-driven trimming, for the detection of outliers and unsuspected structure in data and also for building robust models. Starting from small subsets of data, observations that are close to the fitted model are added to the observations used in parameter estimation. As this subset grows we monitor parameter estimates, test statistics and measures of fit such as residuals. This talk surveys theoretical and empirical development in the work on the Forward Search over the last decade and discusses similarities and differences with traditional robust estimators.

References:

Atkinson, A.C, & Riani, M. (2000). *Robust diagnostic regression analysis*. Springer Verlag, New York.

Atkinson, A.C., Riani M., & Cerioli, A. (2004). *Exploring multivariate data with the forward search.* Springer Verlag, New York.

Atkinson, A.C., Riani, M., & Cerioli, A. (2010). The forward search: Theory and data analysis (with discussion). *Journal of the Korean Statistical Society, 39*, 117-134.

Riani, M., Atkinson, A.C., & Cerioli, A. (2009). Finding an unknown number of multivariate outliers. *Journal of The Royal Statistical Society, Series B Statistical Methodology, 71,* 447-466.

Marco Riani, *after receiving his PhD in Statistics in 1995 from the University of Florence, joined the Faculty of Economics at Parma University as postdoctoral fellow. In 1997 he won the prize for the best Italian PhD thesis in Statistics. He is currently Full Professor of Statistics in the University of Parma. He is author or co-author of more than 80 publications, 30 of which have appeared in international Journals of statistics and 2 books published by Springer Verlag New York. He is co-editor of a book published by Springer Verlag Berlin and guest editor of some special thematic issues of International Journals of Statistics. In 2009 he was the chairman of the scientific committee of the international conference in robust statistics (ICORS). He is currently involved in research projects in the area of robust statistics. At present he is the Italian coordinator of a joint project Italy-Spain involving the analysis of complex data using robust methods, and principal investigator of a project financed by the Italian Ministry of Education involving a network of local research units.*

Wobbe Zijlstra (Tilburg University & CoRPS): Robust Mokken scale analysis by means of the forward search algorithm

Exploratory Mokken scale analysis (MSA) is a popular method for identifying scales from larger sets of items. As with any statistical method, in MSA the presence of outliers in the data may result in biased results and wrong conclusions. The forward search algorithm is a robust diagnostic method for outlier detection, which we adapt here to identify outliers in MSA. This adaptation involves choices with respect to the algorithm's objective function, selection of items from samples without outliers, and scalability criteria to be used in the forward search algorithm. The application of the adapted forward search algorithm for MSA is demonstrated using real data. Recommendations are given for its use in practical scale analysis.

Reference:
Zijlstra, W.P., van der Ark, L.A., & Sijtsma, K. (2011). Robust Mokken scale analysis by means of the forward search algorithm for outlier detection. *Multivariate Behavioral Research, 46,* 58-89.

Wobbe Zijlstra *studied Psychology at University of Groningen (2004). He obtained his PhD at Tilburg University (2009) on the topic "outlier detection in questionnaire data for attribute measurement". Currently he is a lecturer at the Department of Methodology and Statistics at Tilburg University and he works as research consultant / methodologist at CoRPS (Center of Research on Psychology in Somatic Diseases) at Tilburg University. At CoRPS he helps medical psychology researchers on their research designs and statistical analyses (e.g., longitudinal data, missing data, power analysis, and sampling).*

Jorn de Haan (Genetwister Technologies, Wageningen): Nonparametric transformations in the analysis of microarray data.

Analysis of Variance (ANOVA) can be used to separate the effects of different factors in a data set. Typical examples for gene expression or microarray data are the factors time and treatment. This separation can improve the interpretability of the results. However, the main effects and interactions, calculated in ANOVA, can be heavily influenced by outliers, large numbers of non-expressed genes with noise, and the heavy-tailedness of the distribution of expression values. Robust methods are less affected by these and will improve the analysis.
In this presentation I will show work done on several methods of transformation to perform robust nonparametric ANOVA. The methods are applied to a large multi-treatment time series dataset.

Jorn de Haan *studied medical biology at the Vrije Universiteit (VU) Amsterdam. He obtained his Phd at the Radboud University Nijmegen in 2011 with the thesis "Improvements in the analysis of microarray data". In 2007 he started a postdoc position at TNO quality of life in Zeist while finishing his PhD. A large part of his work there, consisted of taking part in the Nutri Genomics Consortium (NGC). Relations between diet and health were investigated in the NGC with data rich techniques like microarrays, proteomics and metabolomics. In 2009 Jorn joined Genetwister Technologies in Wageningen. There he became research manager bioinformatics and biostatistics. At Genetwister Jorn is involved in the analysis of large microarray studies and Next Generation Sequencing data, to find answers to biological questions in a wide range of plant species.The interests of Jorn are in the scientific field where biology meets multivariate statistics, where new approaches are needed to get more knowledge from large datasets.*