THE COUNCIL OF THE SOCIETY

President
Pr. Marcel Rémon, Facultés universitaires ND de la Paix, Namur

Vice-President
Pr. Stefan Van Aelst, Universiteit Gent

Secretary of the Society
Pr. Gentiane Haesbroeck, Université de Liège

Treasurer
Pr. Roel Braekers, Universiteit Hasselt

Administrators
Pr. Gerda Claeskens, Katholieke Universiteit Leuven
Dr. Filip De Ridder, Janssen Pharmaceutica
Pr. Peter Goos, Universiteit Antwerpen
Pr. Tetyana Kadankova, Vrije Universiteit Brussel
Pr. Luc Lebrun, SPF Economie
Pr. Davy Paindaveine, Université Libre de Bruxelles
Pr. Christian Ritter, Université Catholique de Louvain
Pr. Herbert Thijs, Universiteit Hasselt
Pr. Ingrid Van Keilegom, Université Catholique de Louvain
Pr. Kristel Van Steen, Université de Liège

Website of the Society
www.sbs-bvs.be

B-Stat News editor
Sophie Vanbelle : sophie.vanbelle@maastrichtuniversity.nl
Herbert Thijs: herbert.thijs@uhasselt.be

Webmaster
Laurence Seidel : laurence.seidel@ulg.ac.be
# TABLE OF CONTENTS

- 20\textsuperscript{th} Annual Meeting of the Belgian Statistical Society .................................. 3
- Symposium on Causal Mediation Analysis ................................................................. 11
- Forthcoming statistical events .................................................................................... 12
- Recent PhD theses .................................................................................................... 13
- Job market .................................................................................................................. 21
The conference will be held at the Congress Centre of Liège while accommodation will be at the Alliance Hotel located just next to it (Esplanade de l’Europe 2, 4020 Liège).

The detailed scientific program is outlined below. It is worth mentioning that the social program includes:

- for the PhD students, a dinner on the fun fair of Liège on Wednesday evening,

- for every participant, a cocktail organised by a folkloric society during the poster session on Thursday evening

- and for all participants who want to join us on Friday afternoon, either a guided tour of Liège or a visit of the museum of the Maison de la Métallurgie et de l’Industrie de Liège.

The organising and scientific committees look forward to welcoming you in Liège.
Wednesday 24 October 2012
(Special afternoon devoted to PhD students, room Simenon, Palais des Congrès)

13:30 – 14:00 Welcome

14:00 – 15:00 Short course (part 1): Christian Ritter (UCL)
   From Chris’s bag of tricks: about automatic reporting, packaging, distributing, and archiving statistical research

15:00 – 16:30 Poster and quiz

16:30 – 17:00 Coffee break

17:00 – 18:00 Short course (part 2): Christian Ritter (UCL)

18:00 – 19:00 Quetelet Prizes: oral presentations

19:00 – Dinner on the Fun Fair of Liège
Thursday 25 October 2012

8:30 – 9:30  Registration and coffee

9:30 – 9:40  Welcome address by the President of the Belgian Statistical Society

9:40 – 10:30  Invited speaker I: J. de Uña-Álvarez (University of Vigo)
  Advances in survival data: parametric versus nonparametric truncation

10:30 – 11:00  Coffee break

11:00 – 12:20  Parallel sessions I and II: contributed papers

  Session I: Survival - censoring

  C. Heuchenne and G. Laurent (ULg)
  Error distribution estimation in nonparametric regression with right censored selection biased data

  S. Jaspers (UH), M. Aerts and G. Verbeke
  A new semi-parametric mixture model for interval censored data, with applications in the field of antimicrobial resistance

  M.A. Nicolae (UCL), H.J. van Houwelingen and H. Putter
  Dynamic pseudo-observations: a robust approach to dynamic prediction in competing risks

  A. Benoit (UCL), C. Legrand and W. Dewé
  Performances of Cox regression analysis in estimating influenza vaccine efficacy when sources of heterogeneity are omitted
**Session II: Robustness**

**K. Vermeulen** (UG) and S. Vansteelandt  
*On estimation of nuisance working models in doubly robust estimators*

**P. Slock** (UG), S. Van Aelst and M. Salibian-Barrera  
*Robust estimation for heteroscedastic regression*

M. Hubert, P. Rousseeuw and **K. Vakili** (KUL)  
*The MCS estimator of location and scatter*

D. Paindaveine and **G. Van Bever** (ULB)  
*Nonparametrically consistent depth-based classifiers*

12:30 – 14:00  
**Lunch**

14:00 – 14:50  
*Invited speaker II: Claudia Czado* (Tech. Univ. München)  
*The world of vines*

14:50 – 15:50  
**Parallel sessions III and IV: contributed papers**

**Session III (part 1): Copula / Probabilistic index models**

I. Gijbels and **K. Herrmann** (KUL)  
*Portfolio value-at-risk and expected-shortfall in a copula setup*

**J. De Neve** (UG), O. thas, L. Clement and J.-P. Ottoy  
*A new kind of regression model*

**F. Zhang** (UG), S. Vansteelandt, J. De Neve and O. Thas  
*Probabilistic index mixed models for clustered data*
Session IV (part 1): Missing data

N. Sabbe (UG), O. Thas and J.-P. Ottoy
*Missing data in covariates for logistic Lasso: the EMLasso*

A.-F. Donneau (ULg)
*Factors influencing multiple imputation in longitudinal ordinal data*

E.N. Njagi (UH), G. Molenberghs, M.G. Kenward, G. Verbeke and D. Rizopoulos
*A Framework for characterizing missingness at random in generalized shared-parameter joint modeling framework for longitudinal and time-to-event data*

15:50 – 16:20  Coffee

16:20 – 17:00  Parallel sessions V and VI: contributed papers

Session III (part 2): Methodology

I. Gijbels and I. Vrinssen (KUL)
*Robust variable selection using the nonnegative garrote method*

J. Jaeger (ULg) and P. Lambert
*On the use of adhesion parameters to validate models specified using differential equations*

Session IV (part 2): Design

D. Akinç (KUL), M. Crabbe and M.Vandebroek
*Kulback-Leibler information based design criteria for the mixed logit choice model*

M. Ugille (KUL), M. Moeyaert, S.N. Beretvas, J. Ferron and W.Vanden Noortgate
*Meta-analysis of group-comparison and single-subject experimental designs*
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
</table>
| 17:00 – 17:50 | Invited speaker III (bENBIS): Dennis Fok (Erasmus Universiteit Rotterdam)  
Proactive churn management - Let sleeping dogs lie |
| 18:30    | General Assembly of the Belgian Statistical Society                  |
| 19:30 – 20:00 | Poster session                                                     |
| 20:00    | Conference dinner                                                   |
Friday 26 October 2012

9:00 – 9:50 Invited speaker IV (Quetelet Society): Torben Martinussen (Univ. Copenhagen) On some topic in survival analysis

9:50 – 10:50 Parallel sessions V and VI (part 1): contributed papers

Session V (part 1): Frailty models

S. Abrams (UH) and N. Hens Misspecification in frailty models: an application in infectious disease epidemiology

Cetinyürek, A. (ULg) and P. Lambert Flexible frailty model for interval-censored data using Bayesian P-splines

M. Munda (UCL), C. Legrand, P. Janssen and L. Duchateau Detecting time dependency in shared frailties using the mixed model methodology

Session VI (part 1): Methodology

Y. Zhang (UG) and O. Thas Constrained ordination analysis in the presence of zero inflation

E. Milanzi (UG), G. Molenberghs, A. Alonso, M.G. Kenward, A.A. Tsiatis, M. Davidian and G. Verbeke Estimation after a group sequential trial

D. Magis (ULg) Some formulas for the standard error of the weighted likelihood estimator of ability with small psychometric tests

10:50 – 11:20 Coffee

11:20 – 12:00 Parallel sessions V and VI (part 2): contributed papers
Session V (part 2): fMRI

**S. Roels** (UG), T. Loeys and B. Moerkerke
*Bootstrapping fMRI data: a fully parametric approach versus a semi-parametric approach*

**J. Durnez** (UG), T. Nichols and B. Moerkerke
*Post-hoc power estimation for topological inference in fMRI*

Session VI (part 2): Nonparametric methods

**A. Daouia** (UCL)
*On kernel smoothing for extremal quantile regression*

**J.-F. Freyermuth** (KUL), F. Autin, G. Claeskens and R. von Sachs
*Hyperbolic wavelet thresholding rules: the curse of dimensionality through the maxiset approach*

12:00 – 12:50 **Invited speaker V (Biostat Section): Filip De Ridder**
(Johnson & Johnson Pharmaceutical R&D)
*Modelling & simulation: fact and fiction in drug development.*

12:50 Closing by the President of the Belgian Statistical Society

13:00 – 14:00 **Lunch**
SYMPOSIUM ON CAUSAL MEDIATION ANALYSIS

Ghent (Belgium), 28-29 January 2013

We are pleased to announce a symposium on “Causal Mediation Analysis” organized by the Center for Statistics of Ghent University (Belgium).

This meeting aims to bridge the gap between traditional mediation analysis building on the famous work of Baron and Kenny and state-of-the-art causal mediation analysis. The idea is to discuss recent developments between methodological researchers and to bring them to the wider research community in a non-technical way.

Registration and abstract submission

Registration for the symposium is now open. The number of participants is limited to 130. Participants are invited to submit an abstract for an oral presentation or a poster. The deadline for abstract submission is October 15th, 2012.

Registration and abstract submission should be done on-line on the website of the meeting.

Venue

Het Pand, Gent, Belgium

Scientific Committee

Beatrijs Moerkerke, Tom Loeys, Yves Rosseel, Stijn Vansteelandt, Tomasz Burzykowski, Francis Tuerlinckx

More information, registration and abstract submission

http://www.da.ugent.be/cvs
FORTHCOMING STATISTICAL EVENTS

October 15 – 19, 2012 – Frejus (France), Journées d’étude en statistique (SFdS)
More information:

October (24), 25 – 26, 2012 – Liège, Belgium, 20th Annual meeting of BSS

October 25 – 26, 2012 – Gent, Belgium, 15th Annual meeting of ESPACOMP
More information: www.espacomp.eu

Lixin Zhang. *Statistical methods for analysing serum protein electrophoresis data in External Quality assessment (EQA) programmes (2010)* – Promoter Pr. A. Albert (ULg)

Laboratory tests play a central role in clinical decision making. They are used to diagnose diseases, predict patient’s outcome, and monitor drug therapy or screen human populations for specific disorders. Daily, millions of tests results are produced by clinical laboratories nationwide, yielding a formidable amount of data to be interpreted. Healthcare authorities however have been increasingly concerned by the overall quality of the work done. Despite strict internal quality control procedures implemented by the laboratories, results obtained for the same sample in two distinct laboratories may sometimes differ markedly from each other. External Quality Assessment (EQA) schemes were precisely designed to control the analytical performance of clinical laboratories on a wider scale by an external agency. The goal of EQA is to ensure that test results are compatible regardless of which laboratory actually did the tests, in other terms to warrant laboratory comparability and hence optimal healthcare provision. EQA schemes consist of surveys in which all participating laboratories are requested to assay the same control specimen as in daily routine and return the test results to the EQA centre with detailed information about the assay techniques used. Results are then analysed statistically. The statistical analysis basically pursues three objectives: (1) to obtain robust estimates of the concentration of the analyte in the control sample (mean value) and of the between-laboratory variability (standard deviation); robustness is needed because EQA datasets often contain gross-errors and outliers which seriously affect the estimations; (2) to evaluate the quality performance of laboratories by highlighting “out-of-range” results and the corresponding “poor-performers”, on a short and long-term basis; (3) to assess and compare the analytical precision (coefficient of variation) of the techniques/equipments used by the participants.
For several decades, EQA programmes have been running nationwide to control most common biochemical analytes such as glucose, cholesterol or calcium, but also enzymes and hormones. More recently, serum protein electrophoresis was introduced in the national EQA programmes in addition to all tests already controlled for many years. Serum protein electrophoresis is a laboratory test, which unlike single tests, yields five fractions, albumin, α₁, α₂, β and γ globulins, respectively, which sum up to 100% and which ought to be interpreted jointly by multivariate methods. This thesis purposed to provide a theoretical and practical solution to the problem of analysing and interpreting electrophoretic data in the context of EQA programmes. The EQA datasets used to illustrate the statistical methodology developed in this work mostly originated from the French and Belgian EQA surveys carried out between 2004 and 2008.

The first part of this work addressed the problem of assessing laboratory performance from EQA electrophoretic data. The most simplistic approach would consist in applying the univariate statistical methods to each electrophoretic fraction separately. This method however ignores the fact that the fractions are (1) correlated and (2) linearly related (they sum up to 100%), yielding a singular covariance matrix. Therefore, the first multivariate approach envisaged consisted in deleting one electrophoretic fraction to avoid the singularity problem and to obtain a robust estimation of the mean vector and the covariance matrix by the minimum covariance determinant (MCD) technique. The Mahalanobis distance was then used to highlight “out-of-range” electrophoretic profiles and to detect the “poor-performing” laboratories. When applied to the EQA datasets, it turned out that this approach was not optimal, leading to spurious correlations and detecting too many “out-of-range” laboratories. Therefore, an improved and innovative approach was proposed in which an isometric log-ratio (ILR) transform was applied to electrophoretic profiles before analysing them statistically. The ILR transforms the five fractions into four dimensionless unconstrained variables. The MCD technique is then applied to obtain robust estimates of the mean vector and covariance matrix. The estimates are used to calculate Mahalanobis distances and uncover “out-of-range” laboratories. When applied to the EQA datasets, the second approach clearly superseded the first one by providing more realistic figures about the number of poor performing laboratories. Graphical representations of electrophoretic profiles were also investigated by means of the MCD z-score plot and the robust principal component analysis (PCA) star plot.
In the second part of the work, the author focussed on the assessment and comparison of the analytical precision of the electrophoretic techniques used by the EQA participants. This led to investigate multivariate extensions of the classical coefficient of variation (CV). Multivariate coefficients of variation were reviewed, including those proposed by Reyment (1960) and Van Valen (1974, 2005). Reyment was the first to propose a definition for the multivariate CV and to derive a formula for the standard error of the estimate. Van Valen suggested another definition applicable in all generality but which does not explicitly account for the correlated structure of the variables. We used an argument of Voinov and Nikulin (1996) to develop a multivariate CV based on the Mahalanobis distance and which is scale-invariant. We applied it to the 2004 French-Belgian EQA data and were able to rank the electrophoretic techniques by decreasing analytical precision. Unfortunately, this approach as well as Reyment’s CV require non-singular covariance matrices and hence are not really applicable for small sample size groups of laboratories. Therefore, we developed a novel and more general definition of the multivariate CV denoted \( CV_m \). The new method features many interesting theoretical properties: it is simple and easy to calculate, requiring no matrix inversion (unlike most other methods) but only quadratic forms. No restriction is imposed on the number of observations and on the number of variables of the problem. We applied the new multivariate CV to the EQA electrophoretic datasets and were able to demonstrate its feasibility in practice. It allowed us to rank the electrophoretic techniques even for groups of participants of very small size and to highlight the most precise ones, like the fully-automated capillary zone electrophoresis (CZE). We also applied the new method to EQA flow cytometry data and to a microarray dataset published in the literature (Golub et al., 1999), hence demonstrating its wide applicability. The thesis concluded with some personal theoretical improvements of Reyment’s CV which correct the original definition, its estimation and the corresponding standard error.

In summary, the statistical methodology developed in this thesis provides a comprehensive solution to the analysis of electrophoretic data collected in EQA surveys and entails some nice theoretical results.
Laboratory medicine has undergone a spectacular evolution in the past decades becoming of paramount importance for supporting daily diagnostic and therapeutic decisions. The increase of the volume of laboratory analyses has not gone without an emerging risk of measurement errors that may have far-reaching consequences, even on the patient’s life. External Quality Assessment (EQA) programmes established since many years in most countries and often running on an international level aim at going beyond the “internal quality control” procedures of every laboratory and hence at improving laboratory quality by inter-laboratory comparisons. An EQA survey generally consists in sending aliquots of the same sample to various laboratories for assaying selected tests. After finishing the assays, results are reported back to the EQA organizer. Subsequently these results are subject to a statistical analysis, which is performed globally on all the participants or by groups of participants using the same analytical technique. Finally, a report is sent to every participant that informs about the acceptability of the results that he/she produced.

This thesis, structured in five chapters, focuses on the External Quality Control of clinical laboratories by a critical analysis of existing methods and by developing new approaches to improve the current procedures.

The first chapter of this work emphasizes the evolution of the role of the clinical laboratory and EQA in overall quality improvement. After the report “To Err is Human: Building a Safer Health System”, numerous scientists became interested in investigating the frequency, source and impact of laboratory errors. The Total Testing Process (TTP) became recognized as the best framework to investigate laboratory errors. The three different phases of the TTP, respectively, the pre-analytical, analytical and post-analytical phases, were described in detail and the nature and frequency of errors in each phase explained. For each phase, possible improvements were stated and the role of EQA suggested. While today EQA mainly focuses on the assessment and improvement of the analytical phase, proposals were made to improve the role of EQA for assessing and improving pre-and post-analytical error as well, by using specific sample material and by automating the reporting of data and laboratory reports to the EQA participants. The basic principle of the comparison of results of a laboratory with those obtained by the other laboratories traditionally based on the calculation of
“z-scores” was critically reviewed. An in-depth simulation study comparing different techniques was carried out, shedding new light on the shortcomings and strong points of the different approaches. We concluded that robust techniques that eliminate outliers before calculating z-scores should be recommended.

The second chapter discussed the role of EQA as a tool to assess harmonization between methods. The role of EQA was described together with the pitfalls and current shortcomings for assessing harmonization. A major problem in assessing standardization between methods is the possible presence of matrix effects in control samples, in which a method-specific bias may appear. Several explanations for matrix effects were given and statistical techniques were described that assist EQA organizers to split up the data in homogeneous peer groups using multivariate statistics. The chapter also reviewed several techniques to be used in method comparison studies, and the preference for the use of orthogonal regression was expressed.

Chapter 3 introduced several evaluation techniques that combine information from different samples or parameters: variance and bias index scores, mean ranking scores, counts of z- and u-scores, and a long-term analytical coefficient of variation (CVa). Also, a new and original method was introduced that uses 3 steps to identify outliers in a first step, to find laboratories with exceeding variability in a second, and to identify laboratories with high bias in a third step. All techniques were evaluated and discussed by means of a data set in which accidental outliers, high variability and high bias were induced. In addition, the comparison between the different evaluation methods revealed that distinguishing between variability and bias is a tedious task, and that some long-term analysis methods lack robustness against outliers. Also, it was proven that evaluation techniques summarizing results of different parameters may hide useful information. Finally, the 3-step method was proposed as a method for discerning between errors produced in the pre-or post-analytical phase, and errors that arise from the analytical phase.

In Chapter 4, the 3-step method was applied to data obtained from the Belgian EQA. Data sets from alcohol, flow cytometry, lithium and semen analysis surveys were examined. The method was extended for applicability to heteroscedastic, i.e. unequal residual variability, regression models and proved to be of potential use in a wide range of surveys. For each of the surveys under consideration, a follow-up was made of the occurrence of
accidental mistakes, and the evolution of within-laboratory variability and bias for selected methods. Major conclusions were drawn from this work: a clear improvement of laboratory performance was attained over time, in particular a marked reduction of accidental mistakes.

In Chapter 5, some graphical representations of EQA data were explored and a graphical representation of the 3-step method was described. The histogram, normal quantile plot and box plot were suggested for providing a quick online visual overview of EQA data. Other graphical representations that respond to specific questions were discussed, like Shewhart charts, cusum charts and graphical representations to combine variability and bias in one graph. The author also suggested the use of interactive graphs for improving feedback from the EQA organizers to the EQA participants by means of Scalable Vector Graphics. The latter approach was illustrated with web-accessible examples of long-term evaluation of z-scores and the results of the 3-step method for the data obtained in the Belgian EQA for alcohol determination in blood.

In brief, this work describes in a critical and constructive way current statistical methods used in EQA and proposes novel statistical and graphical techniques to help alleviating the future needs of External Quality Assessment programmes.
As mentioned in the title, the framework of this doctoral dissertation encompasses two different subjects: robust statistics on the one hand and classification and clustering techniques on the other hand.

Robust procedures try at the same time to emulate classical procedures and to produce results that are not unduly affected by contaminated observations or deviations from model assumptions. A basic example of robust estimator is the sample median that estimates the central tendency of a set of observations, as the sample mean does, but without being so vulnerable to changes in the observations. In order to measure the robustness of an estimator, tools have been developed. For example, the influence function measures the impact that infinitesimal contamination can have on the estimator while the breakdown point measures the amount of contamination that makes the estimator totally unreliable. Usually, the price to pay for protection against contamination is a loss in statistical efficiency.

Classification and clustering techniques try to find groups among observations. Grouping is one of the most basic abilities of living creatures; the simple fact of naming objects is already grouping. The main interest lies in the fact that the characteristics of a group as well as its differences from other groups can be used as a summary of the dataset. There exist a lot of techniques to construct objective partitions of the data. Classification procedures differ from clustering in at least one important point. A classification rule is set up on a dataset for which the memberships are known and this classification rule is then used to classify observations from another dataset. On the other hand, a clustering rule is set up without knowing the memberships of the observations. Nevertheless, classification and clustering procedures have a common ground; they are bound to misclassify some observations.

Firstly, the robustness and the efficiency of the error rate of the generalized and trimmed k-means clustering procedures are studied. It is shown that the error rate inherits the robustness properties of the procedure behind it. Moreover, the optimality (in the sense of reaching the smallest error rate) of these procedures is stated under balanced mixtures of homoscedastic and spherically symmetric distributions.
Then, a recent clustering procedure, the TCLUST procedure, is considered. Although this technique is assumed to be robust, no formal studies of its robustness have been carried out yet. It is shown that the classical level of robustness for procedures based on trimming, i.e. the trimming size, is reached when the dataset is suitable for a cluster analysis, i.e. when it is composed of several well separated clusters.

Finally, a robust procedure is introduced for classifying observations into ordered classes. Starting from the classical procedure, which is shown to be vulnerable to contamination, the idea behind the robust procedure is to downweight remote observations. Robustness and efficiency of this new method are studied and a diagnostic plot allowing to detect influential observations in a dataset is presented.
JOB MARKET

There is an extensive list with new job offers which would be time and space consuming to publish here. Please note that the complete list with further details is shown at our website:

http://www.sbs-bvs.be/
EDITORIAL NOTE

We would like to publish in this Newsletter any statistical matter such as:

- information about universities, institutes (1 to 3 pages);
- lists of recent publications and technical reports;
- abstracts of recent PhD theses;
- news of members;
- forthcoming statistical events and announcements;
- short papers about teaching methods in statistics, statistics in the industry, official statistics, etc.

Suggestions are welcome: please, contact us.

Suitable information for the next issue, prepared as (LA)TEX or WORD FILES, should reach the editors of the Newsletter BEFORE December 31, 2012, preferable by e-mail to:

sophie.vanbelle@maastrichtuniversity.nl or herbert.thijs@uhasselt.be

Any change of job, address, phone number, ...?

Please notify the Secretary of the Society:

Gentiane Haesbroeck
Université de Liège
Institut de mathématique
Sart Tilman B37
B-4000 Liège
g.haesbroeck@ulg.ac.be