



International Federation of Classification Societies Newsletter

President: Buck McMorris

Number 36

November 2008

Editor: André Hardy

The International Federation of Classification Societies, founded in 1985, is composed of:

Associação Portuguesa de Classificação e Análise de Dados, British Classification Society, Central American and Caribbean Society of Classification and Data Analysis, Classification Society of North America, Gesellschaft für Klassifikation, Irish Pattern Recognition and Classification Society, Japanese Classification Society, Korean Classification Society, Société Francophone de Classification, Società Italiana di Statistica, Statistical Society of Slovenia, Vereniging voor Ordinate en Classificatie and Section of Classification and Data Analysis of the Polish Statistical Society.

The IFCS is a non-profit, non-political scientific organization, the aims of which are to further classification research. Among other activities, the IFCS organises a biennial conference, and supports the "Journal of Classification" and the journal "Advances in Data Analysis and Classification (ADAC)".

IFCS Homepage:

<http://www.classification-society.org/>

Designed and maintained by David Dubin.

The website contains among others: the IFCS Constitution and By-Laws, the IFCS newsletters, and pointers to the websites of the member societies.

Contents

Words from the IFCS President
News from the IFCS Scientific Secretary
News about IFCS 2009
Awards and prizes conferred by IFCS Member Societies
News from the Classification Society
News from GfKI
News from SKAD
News from the SFC
News from the VOC
News from CLAD
News from JCS
Satellite Workshop of the IASC 2008 Conference
EGC 2009
12th International Facet Theory Conference
Short book review

Words from the IFCS President November 2008

This past July I attended the 32nd Annual Conference of GfKI that was held in Hamburg (a joint conference with the BCS and VOC). The conference organization was excellent and the scientific content of the program was superb. I left Hamburg with a feeling that our scientific community is flourishing with innovative ideas that continue to have impact in several disciplines. I know many of you feel that same way after one of your Society meetings. The upcoming March 2009 IFCS meeting in Dresden will meet or exceed the highest of expectations, so please plan to attend if at all possible. Of particular note (see the announcement in this Newsletter) is for the first time, in addition to the usual conference proceedings, special issues of both ADAC and JoC will feature papers based on presentations given at IFCS2009.

I wish you all Happy Holidays, and hope to see you in Dresden in March.

Buck McMorris
IFCS President
mcmorris@iit.edu

News from the IFCS Scientific Secretary – November 2008

Dear Colleagues and Friends of the Classification,

Here is a report on the activities run and the decisions taken by the IFCS Executive Committee (EC) and Council over the past few months:

1. Representatives of IFCS member societies on the IFCS Council

The IFCS Scientific Secretary got in touch with Presidents of the IFCS member societies whose representatives on the IFCS Council had their terms expired during the year 2008. Following these contacts, the representatives of the IFCS member societies on the IFCS Council are:

Helena Bacelar-Nicolau (CLAD)
Patrice Bertrand (SFC)
Rozenn Dahyot (IPRCS)
Anuska Ferligoj (SSS)
Krzysztof Jajuga (SKAD)
Berthold Lausen (GfKI)
Tae Rim Lee (KCS)
Boris Mirkin (BCS)
Fionn Murtagh (CS)
Akinori Okada (JCS)
Andrzej Sokolowski (SKAD)
Javier Trejos (SoCCCAD)
Alfred Ultsch (GfKI)
Maurizio Vichi (CLADAG-SIS)
Ron Wehrens (VOC)
Kazunori Yamaguchi (JCS)

2. IFCS new member societies

The EC members have been (and still are!) pursuing previously established contacts with groups from China, Spain and Greece who seem to be interested in submitting applications to eventually become new IFCS member societies. As it was already pointed out by the Executive Committee last year, the IFCS shall have the objective to broaden its audience and interests by addressing novel domains and groups of classification researchers outside the classical IFCS community.

3. Special issues of “Advances in Data Analysis and Classification” (ADAC) and “Journal of Classification” (JoC) from IFCS’09 in Dresden

The Executive Committee of the IFCS has endorsed the idea suggested by Willem Heiser about having special issues of JoC and ADAC devoted to papers that will be presented at the next IFCS conference in Dresden (13-18 March 2009). The call for submissions will be prepared by the editors Claus Weihs (IFCS’09 Proceedings), Hans Bock (ADAC) and Willem Heiser (JoC).

4. IFCS invited papers session at COMPSTAT 2008

The IFCS Scientific Secretary organized an invited papers session on behalf of IFCS at COMPSTAT 2008, the 18th International Conference on Computational Statistics in Porto. The theme of the session was “Models for Latent Class Detection” and took place on August 27th. The session was chaired by Alfredo Rizzi (CLADAG-SIS) and comprised the following presentations:

- "Clustering via Mixture Regression Models with Random Effects", Geoffrey J. McLachlan, S. K. Ng, Kui Wang
- "Latent classes of objects and variable selection", Giuliano Galimberti, Angela Montanari, Cinzia Viroli
- “Modelling Background Noise in Finite Mixtures of Generalized Linear Regression Models”, Friedrich Leisch

5. IFCS representative for COMPSTAT 2010

Henk Kiers (VOC) was appointed as the IFCS representative on the scientific programme committee of the COMPSTAT 2010 conference of the IASC (International Association for Statistical Computing), a section of the ISI (International Statistical Institute). COMPSTAT 2010 will take place in Paris on 23-27 August 2010 (www.compstat2010.fr).

6. Chikio Hayashi Awards Program

The new IFCS Awards Committee for the CHA is made of:

Anuska Ferligoj (SSS) – Chair
David Banks (CS)
Paula Brito (CLAD)
Willem Heiser (VOC)
Fionn Murtagh (BCS)
Akinori Okada (JCS)
Maurizio Vichi (CLADAG-SIS)
Claus Weihs (GfKI)
Vincenzo Esposito Vinzi (ex officio as IFCS Secretary)

As already announced last year, the Guidelines and the Application Form for the Chikio Hayashi Awards Program were updated so as to characterize such an important Award as a real scientific award provided by the IFCS to young researchers. Among the new features of this award: candidates need to be authors of a paper to be presented at the IFCS Conference; it is compulsory that a candidate who has been selected for an award attends the IFCS conference for at least three days and presents a paper in a scientific session; the presentations of the CHA winners will be marked in the scientific program of the IFCS Conference as “Winner of the Chikio Hayashi Scientific Award”.

The deadline for submitting CHA applications is 30 November 2008. Details on the procedure for application and nomination, together with guidelines and application form, can be found at: www.classification-society.org/cha/

7. IFCS Finances

The financial report 2007 and the budget proposal 2008 were prepared by Iven Van Mechelen and discussed with the Executive Committee members. These documents were then sent to the Finance Committee that provided a positive audit report. Finally, the IFCS Council gave the final approval of this year’s financial documents.

The IFCS Council has approved Stanley L. Sclove (term till the beginning of 2012) as a new member of the Finance Committee. The Finance Committee is now made of: Akinori Okada (JCS) - Chair, Reinhold Decker (GfKI), Patrick Groenen (VOC), Stanley L. Sclove (CS).

8. IFCS Elections

At present, the IFCS Election committee is getting ready to run the election for the next IFCS Publication Officer and two Additional Members of the IFCS Council as the terms of the current ones will be all expiring at the beginning of year 2009. The results of this election will be most likely known by the end of January 2009.

Finally, please have a look at the IFCS website (www.classification-society.org) for the latest news on member societies, the journals officially supported by the IFCS (Journal of Classification – JoC, and Advances in Data Analysis and Classification - ADAC) and up-to-date information on the IFCS Council composition with contact details of member societies.

I take this opportunity to wish you and your beloved ones all the best for a Merry Christmas and a Happy New Year!

Vincenzo Esposito Vinzi
IFCS Scientific Secretary

News about ifcs2009

Nearly 290 abstracts were submitted for the conference.

[At ifcs2009 a new workshop is announced for Sunday, March 15, 2009:](#)

[Tasoulis/Adams/Rogers: Streaming Data Mining for Sensor Networks.](#)

For this workshop, deadline for Extended Abstract Submission is December 10, 2009.

At ifcs2009 the following tutorials are announced for Friday, March 13, 2009:

- Haasdonk: [Kernel Methods](#)
- Mucha: [Clustering Accompanied by Visualization and Validation](#)
- Schiffner/Szepannek/Ligges: [Local Classification Analysis in R](#)
- Strobl/Hothorn: [An Introduction to Machine Learning with Applications in R: Classification Trees, Random Forests and Boosting](#)
- Trautmann/Mehnen: [Multiobjective Optimization with Desirability Functions and Desirability Indices](#)
- Vinzi: [The PLS component-based approach to structural equation modeling: methodological foundations, recent developments and software applications](#)

For all tutorials, deadline for Registration is February 27, 2009.

For more information see the webpage of the conference:

<http://www.ifcs2009.de/>

C. Weihs (SPC chair)

Awards and prizes conferred by IFCS Member Societies

Since several years, various IFCS Member Societies as well as IFCS have installed prizes and awards for excellent scientific contributions, outstanding conference or Proceedings papers, excellent theses, etc., mostly for young researchers in the first stages of their career. Since members of IFCS Member Societies might be interested in information on such prizes, either for their own application or for suggesting candidates, I will list, in the following, the existing awards and specify briefly some details and links for further information.

IFCS: Chikio Hayashi Awards (CHA) Program

The IFCS provides awards to single persons in the framework of its CHA Program. Awards are given to promising researchers who will present a paper on classification, data analysis and related areas at an IFCS conference, and who are in the early stages of their professional careers, as a support for attending an IFCS conference. Candidates shall be under the age of 35 years at the time of the conference. The award is a cash payment to a candidate who attends the IFCS conference and may be used for travel, accommodation, and conference costs. Candidates must submit their CHA application to the chair of the Awards Committee before some deadline. - For details see the website <http://www.classification-society.org/cha/Hayashi07.pdf>, concerning IFCS2009 see the websites www.classification-society.org/cha/, www.ifcs2009.de, and www.gfkl.org

GfKI: GfKI Best paper Awards

There are three prizes with the purpose to encourage the presentation and submission of outstanding contributions for the GfKI conferences and their Proceedings. The *GfKI Method Award* is conferred for an outstanding theoretical or conceptual paper from any of the central methodological research fields of GfKI. The *GfKI Application Award* is conferred for an outstanding paper from a practical application field and more empirical papers. -- Moreover, there is the *GfKI Young Researcher Award* for the best paper that has been presented in the course of a 'Junior Session' at GfKI conferences, for candidates right before or after the doctoral dissertation (Ph.D.). The winners will get a document and some cash (500-1000 Euro). The prizes are conferred during the opening ceremony of the next GfKI conference where winners can present their results in a brief communication. - For details see the website www.gfkl.org (just click 'Awards') where you find also the list of winners in the past.

News from the Classification Society

Stanley L. Sclove, Secretary

CS: Classification Society Distinguished Dissertation Award

The Award will be for the best PhD (or approximately equivalent doctoral) dissertation nominated by an annual deadline. The theme is clustering, classification, related areas of data analysis, encompassing both associated theory and/or applications. Typical criteria for nomination for the Award include: (i) the most innovative or impressive work in theory/methodology, or the most innovative or well developed application(s); and (ii) the literature review has to be thorough. Nominations are to be received by 1 January each year, from the author (PhD), their advisor/supervisor, or other related person. In 2009 the Award winner will receive US\$500 in book vouchers from Chapman and Hall/CRC, will be invited for a presentation in a special plenary session at the next CS Annual Meeting where up to \$500 of the travel expenses will be covered by CS. – For details see the CS website <http://thames.cs.rhul.ac.uk/~fionn/classification-society/> where a shortlist of candidates is found as well.

SFC: Société Francophone de Classification

SFC is conferring an award, termed 'Prix Simon Régnier' each year. The award is designed for a young researcher (i.e., below 35 years) and conferred on the basis of a thesis (or/and his/her previous work and articles) on the fields of classification, data analysis, statistics etc. The price consists typically in a certain amount, the opportunity to present his/her work during the annual conference, exemption from fees and a one-year membership in SFC. Candidates must submit an application together with a CV, a maximum of 3 publications and, in case of a thesis, the corresponding evaluation reports. Deadlines and details are typically contained in the Call for Papers and the website of the annual conferences. – Moreover, for most conferences, SFC provides also stipends for young researchers (under 35 years) for covering travel costs for attending the conference. See also the SFC website <http://sfc.enst-bretagne.fr>

ELECTION 2008

The slate for the election at the end of this year is Carolyn Anderson and Nema Dean for the two positions on the Board of Directors and Stan Sclove for Secretary/Treasurer. Directors serve three-year terms. The Directors whose terms expire at the end of 2008 are Carolyn Anderson and Jim Rohlf. Rebecca Nugent, Sam Bates Prins, Padhraic Smyth, and Doug Steinley are continuing Directors .

Through 2009, Fionn Murtagh continues as President, Mel Janowitz as Past-President and Bill Shannon as President-Elect. Willem Heiser is Journal Editor, Mike Kurtz is Editor of the bibliographic Classification Literature Automated Search Service, and Frieder Koehn has assumed the Newsletter editorship. Dave Dubin and Fionn Murtagh have been sharing the Webmaster duties. Murtagh is currently the Representative to the IFCS. .

JOURNAL OF CLASSIFICATION

JoC is twenty-five years old! Issue No. 1 of Volume 25 (2008) has been published.

ANNUAL MEETING

The CSNA 2009 annual meeting will be held on June 11 - 13 at Washington University Medical School, St. Louis, Missouri, organized by Bill Shannon.

Further information will be available through <http://www.classification-society.org> .

OTHER

Stan Sclove has joined the Finance Committee of IFCS.

Classification Society Distinguished Dissertation Award Sam Prins, Chair, CS Dissertation Award Committee

The Classification Society (CS) invites nominations for the 2009 Distinguished Dissertation Award. Nominated PhD (or approximately equivalent doctoral) dissertations must be successfully completed in the 2008 calendar year in the themes of clustering, classification, related areas of data analysis, encompassing both associated theory and/or applications. In 2009 the Award totals US\$500 in book vouchers from Chapman and Hall/CRC and an invitation to present in a special plenary session at the CS Annual Meeting (with up to \$500 of the Award winner's travel expenses covered by CS). Nominations must be received by 1 January 2009. For more details on the Award including nomination and evaluation procedures please visit.

<http://thames.cs.rhul.ac.uk/~fionn/classification-society/dissaward2009.html>

Samantha C. Bates Prins, Ph.D.

News from GfKL

Working Group AG-DANK.

The 30th Fall Meeting of the Working Group for Data Analysis and Numerical Classification (AG-DANK) of the GfKL took place at the Weierstrass Institute for Applied Analysis and Stochastics (WIAS) Berlin from Friday 14 to Saturday 15 Nov, hosted by Achim Mucha. The Meeting was attended by 17 participants, among them two from abroad. The nine talks dealt with spectral clustering, mixture distributions, multivariate visualization, pattern recognition, and applications to medical and market research and economic time series.

Three data sets had been offered for analysis, two synthetic and one archeometric. The genesis of the latter was presented in a talk. There was an animated discussion on the number of clusters of this data set. The synthetic data sets revealed again that not every method is appropriate for every data set.

A report with the talks and analyses will be published beginning of 2009 in the WIAS Series of Reports. Data sets and more details are available under the URL

www.uni-passau.de/ritter

Gunter Ritter

News from SKAD

(Section of Classification and Data Analysis of Polish Statistical Society)

The annual conference of SKAD took place on September 17-19 in Jastrzebia Gora. At the conference about 70 papers were presented. During the conference Board elections took place. The following persons were elected for the period 2009-2010:

Marek Walesiak – chairman

Krzysztof Jajuga – vice chairman

Eugeniusz Gatnar – secretary

Zdzislaw Hellwig, Joanicjusz Nazarko, Andrzej

Sokolowski, Waldemar Tarczynski, Kazimierz

Zajac – members

Krzysztof Jajuga

News from SFC

1. Elections

- Christiane GUINOT, Biometrics and Epidemiology Unit, CE.R.I.E.S., Neuilly sur Seine, France, has been elected as President of the SFC
- Jérôme SARACCO, University Montesquieu - Bordeaux IV GREThA - UMR 5113, Pessac Cedex, has been elected as Treasurer of the SFC

2. SFC'2008

The first joint-meeting of the Société Francophone de Classification (SFC) and the Group of Classification and Data Analysis (CLADAG) of the Italian Statistical Society took

place last June in Caserta (Italy) <http://www.sfc-cladag.unina2.it/>. All participants enjoyed the nice location of the meeting in the famous Royal Palace created by the King Charles III of Naples.

About 140 researchers and experts participated to the meeting and 106 papers were presented in 4 plenary, 4 semi-plenary, 8 invited and 16 contributed sessions. A PhD session was also organized as well as a session devoted to the « Simon Régnier Prize » awarded to a young researcher. The conference was preceded by a tutorial on Symbolic Data Analysis and the SODAS software.

At the end of the meeting, SFC and CLADAG decided to strengthen their links by concluding in the next future a formal agreement.

Bernard Fichet

SFC'2009

The 16th Francophone Classification Society (SFC) meeting will be held at Grenoble (France) on September 2-4 2009. The scientific program will include sessions with invited talks and contributed presentations.

The main topics of the conference are:

- Classification and discrimination
- Clustering and distance analysis
- Combinatorial methods
- Data mining
- Hierarchical and non hierarchical classification
- Image and signal analysis
- Mathematical and statistical approaches
- Neural networks and genetic algorithms
- Optimization approaches
- Representation and visualisation
- Similarities and dissimilarities
- Symbolic data analysis
- Trees, graphs and lattices
- Validation

Special emphasis will be laid on the following topics:

- Biostatistics, Bioinformatics
- Web-Mining
- Temporal and Spatial data analysis
- Structural data analysis (e.g. trees, graphs)

Textual-Mining
Data Stream

The first call for the conference will be announced in December 2008 (<http://sfc.enst-bretagne.fr/>).

Local Organizing Committee
Ahlame Douzal

SFC 2010

The 17th Francophone Classification Society (SFC) meeting will be held at the “Université de la Réunion”, La Réunion, June 2010.

Website: <http://tice2.univ-reunion.fr/sfc2010/>

Jean Diatta and Henri
Ralambondrainy

André HARDY,
SFC secretary

News from VOC

The autumn meeting of the VOC will take place Friday 28th November. The topic of the meeting is 'propensity scores'.

Those who would like to participate are welcome and are kindly requested to register, before November 24th, at the VOC website by using [this link](#). Participation is free.

The meeting will take place in room T 3-24 of the T-building at the Woudenstein location of the Erasmus University Rotterdam. For details on how to reach the T-building of the Woudenstein complex of the Erasmus University Rotterdam, see <http://www.eur.nl/adressen/wegwijzer/>.

The program of the meeting is as follows:

10.00	Welcome & Coffee	
10.30	Saskia le Cessie	Propensity scores, an introduction
11.30	Fannie Cobben	Het gebruik van respons propensities in survey onderzoek
12.15	Lunch	
13.45	Arjan Blokland	The (collateral) effects of imprisonment
14.30	Edwin Martens	Preference for propensity scores when estimating an average treatment effect in case of a dichotomous outcome
15.15	Tea	
15.45	Stef van Buuren	Pooling outcomes after quintile stratification
16.20	Drinks	

Propensity scores, an introduction

Saskia le Cessie

Using propensity scores to deal with confounding has become very popular in recent years. By estimating the probability to receive a certain treatment (the propensity), one can adjust for observed imbalance between treatment groups. In this talk, the basic concepts of propensity scores are considered. We discussed in which situations propensity scores are useful. We also consider how propensity scores can be constructed; answering questions like whether all possible variables related to the treatment should be included in the score. Finally we compare different ways of using propensity scores: propensity matching, stratification, inverse probability weighting, and using the propensity score as covariate. We show that the different approaches can yield quite different results.

Saskia le Cessie is an associate professor at the department of Medical Statistics and Bio-informatics and at the department of Clinical Epidemiology of the Leiden University Medical

Center. Her research interests are in statistical methods for epidemiological research. She is a consultant for the Comprehensive Cancer Center West and a member of the scientific board of the Dutch Arthritis Association (Reumafonds). She has been an associate editor of Applied Statistics (JRSS-C) and has served in the Editorial Advisory Committee and the Council of the International Biometrical Society. She has been a co-author of over 150 publications in Medical and Statistical Journals.

Applying respons propensities in survey research

Fannie Cobben

Recently, in survey literature the introduction of propensity score method (Rosenbaum en Rubin, 1983) had induced applying estimated response probabilities. Rosenbaum en Rubin (1983) have introduced this method to estimate the effect(s) of medical interventions. Harris Interactive uses the propensity score method with estimated response probabilities to solve problems in volunteer internetpanels caused by undercoverage and self-selection. Herewith, the estimated response probabilities are considered as response propensities.

Use of (inverse) response probabilities to adjust for nonresponse has been proposed by Horvitz and Thompson. They suggest to adjust the inclusion probabilities of their wellknown Horvitz-Thompson estimator for the (selective) nonresponse. Bethlehem (1988) and Särndal et al. (1992) describe in which way the Horvitz-Thompson estimator can be adjusted for nonresponse by applying estimated response probabilities.

The lectures deals wit a couple of methods of applying response propensities to correct for nonresponse bias. Furthermore, these methods will be applied to the Continuous Research Living Situation (Permanent Onderzoek Leefsituatie) conducted in the year 2002. During the presentation, the differences between the pertinent methods will be discussed. In addition, the difference with the traditional method of linear weighing for correcting nonresponse will be outlined.

Fannie Cobben has studied econometrics at Free University, Amsterdam. Her specialisation is Statistical Econometrics. After having graduated,

in 2004 she started a job at CBS, Division Methodology and Quality, Voorburg. Under supervision of Jelke Bethlehem, she is writing a PhD-thesis about the analysis and correction of nonresponse in individual enquetes.

The (collateral) effects of imprisonment

Arjan Blokland

At the outset of the new millennium 2.5 million individuals are confined in prisons or jails across North America and Western Europe and in most countries rates are at or near all time highs.

A growing international literature has been attentive to the collateral consequences of the increased use of imprisonment. The potential irony of mass imprisonment is that, to the extent it has unintended adverse effects on life outcomes that are correlated with criminal offending, large-scale growth in the incarceration rate may actually exacerbate the crime problem over the long run by stigmatizing an ever larger class of individuals. Using data from the Netherlands-Based Criminal Career and Life-course Study the effect of first-time adult imprisonment on criminal recidivism and life circumstances in the years following the imprisonment was examined. Unadjusted comparisons of those imprisoned and those not imprisoned will be biased because imprisonment is not meted out randomly. Selection processes will tend to make the imprisoned group disproportionately crime prone compared to the not imprisoned group. In this study group-based trajectory modeling was combined with risk set matching to balance a variety of measurable indicators of criminal propensity.

Arjan Blokland (PhD) is researcher at the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) in Leiden and senior-researcher at Parnassia Addiction Research Centre (PARC) in The Hague. In 2006 he received a VENI-grant for his work on specialization in offending. He currently chairs the European Developmental and Life-course Criminology working group. His main area of research is life course criminology and focuses on the development of criminal careers, the influences of life course transitions on criminal behavior, drug use and crime and the (un)intended consequences of interventions.

Preference for propensity scores when estimating an average treatment effect in case of a dichotomous outcome

Edwin Martens

In observational studies with a dichotomous outcome, a multivariable logistic regression analysis is often used to adjust for confounding and estimate an adjusted treatment effect. This treatment effect is in general an overestimation of the treatment effect that is in most circumstances the intended one. The method of propensity scores on the other hand, will result in a treatment effect that is in general closer to the treatment effect that would have been found when the study was a randomized one. The larger the number of confounders or the larger the treatment effect, the more preferred is the method propensity scores over a multivariable logistic regression analysis.

After 12 years of sociologic-economic research at the Erasmus University of Rotterdam, Edwin Martens worked from 2000 as a biostatistiaan at Utrecht University. In 2007 he finished his PhD on the methods of propensity scores and instrumental variables.

Pooling outcomes after quintile stratification

Stef van Buuren

Propensity score methods offer both theoretical and practical advantages over conventional regression techniques to control for bias in observational studies. Quintile stratification is a popular technique in which exposed and non-exposed subjects are divided into five homogeneous strata. Exposed and non-exposed are compared within each stratum, which leads to five results instead of one. The relevant literature pays surprising little attention to the problem how to aggregate these results into one overall estimate. I will outline pooling methods for differences in means and proportions and for the odds ratio, and illustrate these methods on real data.

Stef van Buuren develops and applies quantitative methods in medicine and social science, with an emphasis on childhood growth and incomplete

data. Van Buuren is professor of applied statistics in prevention at the University of Utrecht. More information can be found at <http://www.stefvanbuuren.nl>.

Hugo Duivenvoorden

News from CLAD

(Associação Portuguesa de Classificação e Análise de Dados, CLAD)

JOCLAD2009

XVI CLAD Annual Meeting

2 - 4 April 2009, Faro, Portugal

<http://www.fct.ualg.pt/joclad2009>

joclad2009@gmail.com

The XVI Annual Meeting of the **Associação Portuguesa de Classificação e Análise de Dados, CLAD, JOCLAD 2009**, will be held April 2 - 4 2009 at the University of Algarve in Faro.

JOCLAD 2009 is an opportunity to meet other researchers, make new contacts that will emphasize interdisciplinary research and interaction between theory and practise as well as to visit beautiful places in Algarve, the south cost of Portugal. This meeting is specially dedicated, but not only, to Health Care and Tourism thematic.

Important Dates:

- 14 February 2009: Deadline for extended abstracts submission
- 25 February 2009: Notification of acceptance
- 28 February 2009: Early Registration deadline
- 12 March 2009: Deadline for extended abstracts final version

Invited speakers include:

- Michael Greenacre, Universitat Pompeu Fabra, Barcelona
- Purificación Galindo Villardón, Departamento de Estadística, Universidad de Salamanca
- Miguel Angel Fajardo Caldera, Departamento de Estadística y Economía, Universidad de Extremadura

- Teresa Bago d'Uva, Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam

Further information about main topics, scientific program, short courses, and submission procedure instructions may be found on the conference website.

The authors of the best applied research papers presented at the Conference will be invited to submit full papers for publication in some special issues of collaborating journals, namely at CLAD's Journal.

Helena Bacelar-Nicolau

hbacelar@fpce.ul.pt

News from JCS

Information about 24th Annual Research Meeting of the Japanese Classification Society

The 24th Annual Research Meeting is in the planning stage to be held on March, 2009. The detailed announcement will be appeared on the JCS Web-page as <http://www.soc.nii.ac.jp/jcs/> in the future.

Yoshiro Yamamoto

The Satellite Workshop of the IASC 2008 Conference.

Theme: Statistical Modelling for Computer Security

December 1-3, 2008

Soongsil University, Seoul, Korea.

The conference is jointly organized by Soongsil University – Department of Statistics and Actuarial Science, Department of Computer Science, Graduate School, Korea Statistical Society – Statistical Computing Section, Classification Section

http://stat.inha.ac.kr/iasc2008_sws

Tae Rim Lee

9th EGC conference

Knowledge Extraction and Management
Strasbourg – 27-30 January 2009
<http://lsiit.u-strasbg.fr/egc09>

EGC 2009 is the meeting point for French speaking researchers, industrials and users from Databases, Machine Learning, Knowledge Representation, Knowledge Management, Statistics and Data Analysis communities.

Today, large masses of structured or semi-structured data are available in the information systems of private companies and on the web. These data are potentially rich of unheard resources, which could be exploited. To take advantage of them, we need methods and tools able to collect, represent, store and index data in order to integrate and classify them and then to extract relevant knowledge and finally to see the results of this extraction. To meet this expectation, many research projects are being developed in the fields of Knowledge Discovery in Data Bases and Knowledge Management.

The goal these days is to bring together, on the one hand researchers from related disciplines (Machine Learning, Statistics and Data Analysis, Information Systems and Databases, Knowledge engineering, etc.), and on the other hand engineers who are deploying extraction methods and Knowledge Management techniques, in order to contribute to the formation of a scientific community in the French speaking area world around this double issue of extraction and management of knowledge.

The EGC colloquium is composed of a scientific conference, where industrial and demo sessions are organized, a series of scientific workshops and tutorials.

The event aims to stimulate this community providing a place for debates and cross fertilization between academics, researchers and industrials. The objective is to determine the needs and the future challenges of this new booming discipline.

CALL FOR PAPERS 12th International Facet Theory Conference

June 21-24, 2009, Jerusalem, ISRAEL

New horizons in theory construction and data analysis: Facet Theory (FT), Multidimensional Scaling (MDS), SSA, POSAC and related approaches

Organized by the FACET THEORY
ASSOCIATION, and
BAR ILAN UNIVERSITY

You are kindly invited to attend the 12th International Facet Theory Conference, to take place in Jerusalem from the 21st to the 24th of June, 2009. A course of introduction to FT will be on the first day.

Facet Theory is a systematic approach for coordinating theory and research. FT comprises the universe of observations, the population of respondents, and the range of observations. It stratifies these universes by facets and integrates the design by means of a mapping sentence which guides the construction of items and the formulation of *hypotheses*. Finally, particular *multivariate data analysis methods* (such as SSA, POSAC, MSA) have been developed to test these hypotheses.

The Aims of the Conference

- To present *theoretical advances* in Facet Theory, Mutidimensional Scaling (MDS) and related approaches.
- To review recent and innovative applications in diverse fields and to illustrate their contribution to cumulative science.
- To demonstrate *methodological advances* in multivariate analysis, Similarity Structure Analysis (SSA), and comparisons with other methodologies.

Deadline for Abstract submission – December, 10, 2008.

For further information see www.facet-theory.org

Looking forward to see you at the meeting,

Dov Elizur

Chair, Scientific Committee

E-mail: elizurd@bezeqint.net.

Short book reviews

Correspondence Analysis in Practice, Second Edition - Michael Greenacre,

Chapman & Hall/CRC, Series: Interdisciplinary Statistics Volume: 17

Price: US \$79.95, ISBN: 9781584886167, Publication Date: 5/7/2007, Number of Pages: 296

This book, written by the man who popularized correspondence analysis to english-speaking statisticians, is a complete rewriting of the first edition of 1993. Its main characteristic relies upon its original but very convenient presentation: each one of 25 chapters has exactly eight pages; they are self-contained with marginal notes and a summary at the end.

The author uses deliberately a non formalized style and a lot of worked examples with a large use of graphical displays. The theory is presented in an appendix of only 11 pages which is a *tour de force*.

There are 5 new chapters since the first edition; I found of particular interest: chapter 16 on multiway tables presenting in a very didactic way how to introduce the concept of interaction, chapter 17 and 21 on stacked tables and CA of submatrix, chapter 22 for the analysis of square tables which are common in social sciences (mobility), economics (input-output) and are decomposed into a symmetric and skew-symmetric parts.

M. Greenacre himself insists on the importance of the lengthy computational appendix (46 pages) devoted to the use of R, including the ca package which will be very useful for many users.

Historical details and a commented bibliography of main books are very useful since CA and especially MCA have been discovered and rediscovered by several authors. However dealing only with books without any reference to papers in journals prevents the reader to know more about the story of CA: eg that canonical correspondence analysis was first proposed by C.J.F. Ter Braak in ecology.

I strongly recommend this book to all data analysts from the beginners and students who will find a simple, complete but rigorous guide for performing correspondence analysis, but also to specialists who will be interested by some special topics which are not covered in other books, and to the teachers who will find many ideas for their own courses (like me).

The related website <http://www.carme-n.org/> provides the data sets used in the book, the R-code and some errata.

Gilbert Saporta
Conservatoire National

A Handbook of Statistical Analyses using R.

Brian S. Everitt and Torsten Hothorn.
Chapman & Hall/CRC, 2006.

It is a commonplace to say that the R system became, in these last years, a very widely used statistical software. Especially for researchers, because, with R, it is very easy to reproduce a piece of statistical analysis. All you have to do is to provide the data and your source code. The entire framework, the system and additional packages, are easily and freely downloadable. So, reproducibility of statistical analysis is accessible to every scientist, even for those working in developing countries. Then, R becoming the “lingua franca” of data analysis and statistical computing researchers.

However, when you have to start the learning of such a powerful piece of software, the only use of the official manuals and documentation can be arid and could be a discouraging task, even when

the quality of this documentation is good. This latter is what you need when you want some precision on a definition, a command's syntax, an option, a fine tuning, ...

Doing for R what Everitt's other Handbooks have done for S-plus, STAT, SPSS and SAS, this book can be used by ones who do not know any statistical system, those who want to switch from another system to R, and even by students who are learning statistics.

The first chapter begins with the very basics: download and install, find manuals and help, packages management. This point is directly used because a specific package (HSAUR) which contains all the datasets used in the book, is available on the Comprehensive R Archive Network (CRAN). Follows, the data import/export, the basic data manipulations and some simple, but classical graphs.

All the other fourteen chapters are devoted to one single thematic: simple inference, conditional inference, analysis of variance, multiple linear regression, logistic regression and generalised linear models, density estimation, recursive partitioning, survival analysis, analysing longitudinal data (two chapters), meta-analysis, principal component analysis, multidimensional scaling and cluster analysis.

Each of these chapters has the same clear structure: first, an introduction which describes all the datasets used in for the thematic, included the bibliographic references for their first use; a brief, but concise recall of the statistical background needed follows; then the third part proceed to the analysis, given all source code, interpretation of the results and several graphs; each chapter is concluded with a brief summary and several additional exercises.

It is interesting to remark that (almost) each analysis is used to introduce a new type of graph, extending the control of the graphical tool R, for which this latter is highly deemed. In addition each graph is topped by the code used to generate it.

When I have to use a new function in R, one of the first things I am looking for, in the help is an understandable example and then, I read the complete syntax and details. And that's the way the authors have chosen: use reproducible examples to introduce concretely each type of

analysis. The reader who wants more details has then a sufficient knowledge to go deeper using the official documentation.

One thing is sure, if, in the future, I have to learn statistical software for which the authors have done a Handbook like this one, I will not hesitate at all to use it!

Etienne Cuvelier
Computer Science Faculty
University of Namur
Namur, Belgium

Applied Multiway Data Analysis.

Pieter M. Kroonenberg. Hoboken (NJ): John Wiley & Sons, Inc. 2008. (579 pages).

The term 'Multiway analysis' refers to component modeling of multiway data – data that can be organized in multidimensional boxes. For example, threeway data result when a number of individuals are repeatedly measured on a number of variables in a number of conditions. To have multiway data it is essential that the design is fully crossed. The usually complex relationships between individuals, variables and conditions can be disentangled with some kind of multiway component analysis. Such an analysis is not straightforward, because quite some nontrivial decisions have to be made, like the type of preprocessing of the observed data before analysis, and the specific model to use. The interpretation of the analysis results is usually not straightforward either, which is intrinsic to the complex data structures involved. However, commonly the researcher is rewarded with insights into the data that could not be obtained with other methods. The relatively unfamiliarity of multiway models is an extra barrier to use multiway analysis in practice. This book is an important step in making multiway analysis better accessible to a wider audience than the multiway research community itself.

As stated in the Preface 'The core of this book is concerned with practical issues in

applying multiway component techniques...'. Although this book is clearly guided by the practical issues, theory is certainly not avoided. This is a fortunate choice. A profound theoretical background is necessary to obtain a good understanding of the issues involved, and hence to make proper choices in performing a multiway analysis. The theory is carefully introduced, both at the mathematical and conceptual levels. Considerable effort has been put to make the theory understandable: The matter is often explained from different viewpoints, and supplemented with insightful graphs. I like this very much: it helps novices to obtain a good understanding; to readers who are familiar with multiway analysis it may shed a different light, as it did to me at certain places. When technical details are being short cut too much to the taste of the reader, as it was for me at some places, (s)he can resort to the original literature. The references are numerous and appropriate. As such, the book provides a very good overview of the available literature on threeway analysis, both on the theoretical and applied sides. The main application area of this book is the social sciences, which contrasts to the other book on multiway analysis, *Multi-way Analysis with Applications in the Chemical Sciences* (2004, by Smilde, Bro and Geladi). Furthermore, the latter is somewhat more technically oriented. Hence, both books nicely complement each other.

For absolute novices, the book may be somewhat discouraging. The general ins and outs of multiway analyses are discussed in Part I, on Data, Models, and Algorithms, and Part II, on Data handling, Model selection, and Interpretation, which take as much as 308 pages. This is quite some investment to get a feeling for multiway analysis. Furthermore, readers totally unfamiliar with multiway analysis may find it difficult to see the outline and to select the main points. Both problems could have been prevented by starting with an overview example, with references to coming chapters. As an alternative, introductory articles can partly serve this purpose, like the one of Kiers and Van Mechelen (2001, in *Psychological Methods*). Another problem for absolute novices may be the order of presentation. In particular, in the empirical examples in Chapter 7 some novel terms are used without a –short– explanation, or reference to a future section. This could have done rather easily, and is preferable over the alternative – adopting an Agatha Christie's style of ordering, and

postponing all empirical examples towards the end of the book.

Good things take time – and this holds for this book as well. The idea for the present book dates back to 1996. In my view, it was worth that time. Pieter Kroonenberg is one of the few with a profound knowledge of multiway analysis. It is meritorious that he took the effort to share his knowledge. It is to be hoped that a next edition will appear soon, to resolve some typo's and inaccuracies in the current edition, but above all because the book deserves a broad reading public.

Marieke Timmerman,
University of Groningen,
The Netherlands

Contributions for the coming issue of the IFCS newsletter can be sent to:

andre.hardy@fundp.ac.be

Text files are by preference in ascii or word, with pc format. For graphical materials, by preference GIF is used.

Editor:
André HARDY
Department of Mathematics
FUNDP - University of Namur
8, Rempart de la Vierge
B – 5000 Namur
Belgium
Tel : +32 81 72 49 05
Fax : +32 81 72 49 14