

The Newsletter of the International Association of Survey Statisticians

No. 75 January 2017















#### **Editors**

**Natalie Shlomo and Eric Rancourt** 

**Section Editors** 

Peter Wright Ken Copeland, Natalie Shlomo - Country Reports

- Ask the Experts

Frederic Vogel,

- New and Emerging Methods

Cristiano Ferraz

Monica Pratesi - Book & Software Review

Circulation /Production

Lori Young

**Olivier Dupriez** 

#### ISSN 2521-991X

The Survey Statistician is published twice a year by the International Association of Survey Statisticians and distributed to all its members. The Survey Statistician is also available on the IASS website at http://isi.cbs.nl/iass/alluk.htm.

Enquiries for membership in the Association or change of address for current members should be addressed to:

IASS Secretariat Membership Officer Margaret de Ruiter-Molloy International Statistical Institute P.O. Box 24070, 2490 AB The Hague, The Netherlands

Comments on the contents or suggestions for articles in *The Survey Statistician* should be sent via e-mail to the editors, Natalie Shlomo (Natalie.Shlomo@machester.ac.uk) or Eric Rancourt (Eric.Rancourt@Canada.ca).

#### In this Issue:

- 3 Letter from the Editors
- 5 Letter from the President
- 7 Report from the Scientific Secretary
- 9 News and Announcements

#### 17 Ask the Experts

How can we continue to protect the confidentiality of data subjects under open access initiatives?

#### 22 New and Emerging Methods

Handbook on Master Sampling Frames for Agricultural Statistics: Frame Development, Sample Design, and Estimation

#### 26 Book & Software Review

Analysis of Poverty Data by Small Area Estimation Monica Pratesi (Editor)

#### 31 Country Reports

Australia
Bosnia and Herzegovina
Canada
Israel
New Zealand
Ukraine

- 39 Upcoming Conferences and Workshops
- 46 In Other Journals
- 63 Welcome New Members
- 64 IASS Officers and Council Members
- 65 Institutional Members
- 66 Change of Address Form



#### Letter from the Editors

The January 2017 issue contains articles of interest and important information regarding upcoming conferences, contents of relevant journals, updates from the IASS Executive and more. We hope you enjoy this issue. Please send us your feedback and comments on how we can make improvements.

In the *New and Emerging Methods Section* (edited by the Scientific Secretary Denise Silva), Frederic Vogel and Cristiano Ferrez have contributed an article titled: Handbook on Master Sampling Frames for Agricultural Statistics: Frame Development, Sample Design and Estimation. In the article, the authors summarise the state-of-the-art on developing agricultural statistics based on a handbook that was published with the support of the Trust Fund of the Global Strategy funded by the UK's Department for International Development and the Bill & Melinda Gates foundation, the World Bank and the Joint Research Center of the European Union.

In the Ask the Experts Section (edited by Ken Copeland), Natalie Shlomo from the University of Manchester, has provided a response to the question: How can we continue to protect the confidentiality of data subjects under open access initiatives? The author presents some ways that agencies have been able to disseminate statistical data.

In the *Book Review Section* (edited by Eric Rancourt), Yong You from Statistics Canada have reviewed the book: Analysis of Poverty Data by Small Area Estimation (Ed. M. Pratesi), John Wiley & Sons, 2016.

Please contact Denise Silva (<a href="denise.silva@ibge.gov.br">denise.silva@ibge.gov.br</a>) if you would like to contribute an article to the New and Emerging Methods Section. If you have any questions which you would like to be answered by an expert, please send them to Ken Copeland (<a href="copeland-kennon@norc.org">copeland-kennon@norc.org</a>). You are also welcome to submit your own questions with an answer if you are aware of an important topic of interest. If you are interested in writing a book or software review, please contact Natalie Shlomo (<a href="Natalie.Shlomo@manchester.ac.uk">Natalie.Shlomo@manchester.ac.uk</a>). Finally, if you would like to contribute brief articles or editorials to the newsletter, please send them directly to the editors of the newsletter, Eric Rancourt and Natalie Shlomo.

The Country Report Section has always been a central feature of the IASS The Survey Statistician and we thank all the country representatives for their contribution and coordination of the reports. We also thank the editor of the section, Pierre Lavallée (Pierre.Lavallée@canada.ca) who sadly is retiring from Statistics Canada and hence this is his last issue as section editor of the newsletter. The new section editor is Peter Wright (peter.wright2@canada.ca) also of Statistics Canada and we welcome him to the editorial board. Peter will now be contacting all of the country representatives for their timely reports. Please contribute information on your

country's current activities, applications, research and developments in survey methods. This is of great interest to our IASS membership and the editorial board would like to see the number of country reports grow.

This issue of *The Survey Statistician* includes a letter and updates from our IASS President, Steven Heeringa and from our Scientific Secretary, Denise Silva. In addition, in the *News and Announcement* section we have announcements about nominations for the Waksberg award and the Cochran-Hansen prize. In addition, we include summaries of the highly successful Fifth International Conference on Establishment Surveys (ICES-V) that was held in June 2016 in Geneva, Switzerland and the 9th French Colloquium on Survey Sampling that was held in October 2016 in Quebec, Canada.

It is with deep sorrow that we have also included an *In Memoriam* for our distinguished colleague and friend, Professor Steve Fienberg, who sadly passed away last December. As Steve was so influential in developing the next generation of academics, the memorial is written by a collection of Steve's former students and mentees. He will be sorely missed.

We thank Lori Young from Statistics Canada for collating the advertisements of upcoming conferences and for preparing the tables of contents in the *In Other Journals* section. This is a very time-consuming and detailed task but the information she gathers is deeply appreciated by the members. We also thank Lori for her hard work in collating all the articles into this newsletter that you see before you.

Please take an active role in supporting the IASS newsletter by volunteering to contribute articles, book/software reviews and country reports and/or by making it known to friends and colleagues. We also ask IASS members to send in notifications about conferences and other important news items about their organizations or individual members.

The Survey Statistician is available for downloading from the IASS website at <a href="http://isi.cbs.nl/iass/allUK.htm">http://isi.cbs.nl/iass/allUK.htm</a>.

Eric Rancourt Eric.Rancourt@canada.ca

Natalie Shlomo Natalie.Shlomo@manchester.ac.uk

# INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS (IASS)

Dear IASS colleagues,

My mentor, Leslie Kish, once used the following play on words to communicate his thoughts on the role that our profession plays in society, "To err is human, to forgive divine; but to include errors in your design is statistical". Leslie's use of "..errors in design.." should not be interpreted as espousing sloppiness in our statistical work. Instead it refers to the fact that in statistical practice we acknowledge that the world that we to aim to study through survey research includes many sources of uncertainty. In a world where increased uncertainty—both statistical and not—poses new challenges in our day to day work, a mission of our association and its collective international membership must continue to be to promote and teach the scientific practice of survey research and to ensure that our designs, methods and inferences are both statistically efficient for their intended purpose and remain objective, transparent and relevant to the needs of our respective and mutual societies.

The 2017 World Statistics Congress (2017 WSC) will be held this July 16-21, 2017 in Marrakech, Morocco. IASS will be well-represented in the Scientific Program. With the strong support of Marcel Vieira (IASS Representative) and other members of the Scientific Program Committee (SPC), a total of 15 IASS sponsored Invited Paper Sessions (IPS) have been scheduled for the 2017 WSC. Many other session proposals from the IASS member community will be organized as special topic sessions. I am fortunate to be able to organize the IASS President's Invited Lecture IPS and have invited my long term colleague, Colm O'Muircheartaigh to deliver this special lecture. Colm has titled his talk, "Next Generation Data: Surveys in a Transformed Environment". I hope that many of you will be able to attend Colm's talk in person but if you cannot, look forward to reading his presentation paper in a future issue of *The Survey Statistician*. More information on the Scientific Program for the 2017 WSC can be found at the official web site, <a href="http://www.isi2017.org/">http://www.isi2017.org/</a>.

At the 2017 WSC, the IASS 2017 Cochran-Hansen Prize (<a href="http://isi-iass.org/home/cochran-hansen-prize/">http://isi-iass.org/home/cochran-hansen-prize/</a>) will again be awarded to a young statistician from a developing or transition country. If you are planning to compete for the Cochran-Hansen Prize, papers must be submitted to the committee chair, Monica Pratesi, <a href="mailto:monica.pratesi@unipi.it">monica.pratesi@unipi.it</a>, before February 15, 2017.

For 2017, IASS is providing modest financial support to approximately 4 to 5 regional and international meetings that are specifically focused on survey statistics and survey methodology. The Report of the IASS Scientific Secretary included in this issue of *TSS* provides more specific information on the 2017 conferences and meetings that the officers have selected to receive IASS support. If you are planning to attend one of the sponsored meetings, be sure to note that as an IASS member you will generally be eligible for a discount on your registration fees. If you are organizing a conference and would like to submit a request for IASS financial support please feel free to submit the

proposal to the Scientific Secretary (<u>denise.silva@ibge.gov.br</u>) or to me (<u>sheering@umich.edu</u>).

Beginning in February, IASS members will have the opportunity to vote for the officers who will comprise the 2017-2019 IASS Executive. Under the revised IASS statutes approved by the General Assembly in 2013, the election will be for the office of President-elect (for 2019) and four Vice-Presidents. The five individuals who are elected to office will join President-elect, Peter Lynn, to lead our association for 2017-2019. Peter will designate one of the elected VPs as the Scientific Secretary of the IASS and another VP to have the duties of Financial Officer. I wish to extend my thanks to Jairo Arrow (Chair) and the other members of the IASS 2017-2019 nominating committee (Marcel Vieira, Dalisay Maligalig, Imbi Traat, and Allyson Seyb) for their efforts and for assembling a highly qualified and diverse slate of candidates who have agreed to stand for election. If you are interested, the full report of the 2017-2019 Nominating Committee is available on the home page of the official IASS website, <a href="http://isi-iass.org/home/">http://isi-iass.org/home/</a>.

Before closing, I would like to repeat a suggestion that I included in my July 2016 letter. If each of us committed annually to introduce a student or junior colleague in our work place to the IASS (and in cases of financial need paying their membership fee) we would quickly begin to bring much-needed "youth" to our international association.

I hope that you will take the time to read this issue of *The Survey Statistician* including the report of our Scientific Secretary (Denise Silva), the reports of the Country Representatives and all the other regular content and special articles that the editors, Natalie Shlomo and Eric Rancourt, have worked very hard to assemble.

Yours in the science and practice of survey research,

Steve Heeringa IASS President sheering@umich.edu



#### Report from the Scientific Secretary

This report starts with the good news about IASS co-sponsored conferences for the period July-December 2016 and the ones that are already planned for 2017. In the last semester, IASS co-sponsored the **Baltic-Nordic-Ukrainian (BNU) Network Summer School on Survey Statistics** (<a href="http://wiki.helsinki.fi/display/BNU/Events">http://wiki.helsinki.fi/display/BNU/Events</a>) and the **9th French Colloquium in Survey Sampling** (<a href="http://sondages2016.sfds.asso.fr/">http://sondages2016.sfds.asso.fr/</a>). In the case of the French Colloquium, IASS also got funding from the ISI World Bank Trust Fund for Statistical Capacity Building (WB TFSCB) to financially support two young statisticians from developing countries, covering their expenses to attend the event (as informed in July). Detailed information about the events can be found at the conferences' websites, where visitors can download the proceedings of the BNU Summer School and presentations from the 9th French Colloquium.

For 2017, IASS is already committed to co-sponsor 5 events listed below in which IASS members are entitled to a reduced registration fee.

- **2017 Conference honouring Jon Rao on the occasion of his 80**<sup>th</sup> **birthday** from 24-27 May 2017, at Yun-An Auditorium in Kunming, China, hosted by the School of Mathematics and Statistics at Yunnan University, (http://www.raoconference2017.com/).
- 5th Italian Conference on Survey Methodology (ITACOSM 2017) that will take place at the Department of Statistical Sciences of the University of Bologna from 14 16 June 2017. The conference website is <a href="https://events.unibo.it/itacosm2017">https://events.unibo.it/itacosm2017</a>.
- **ISI Satellite meeting on Small Area Estimation (SAE 2017)** to be held in Paris from 10-12 July 2017 (<a href="http://sae2017.ensai.fr/">http://sae2017.ensai.fr/</a>). The conference is organized by Ensai (Ecole Nationale de la Statistique et de l'Analyse de l'Information), the CREST (Centre de Recherche en Economie et Statistique) and the ILB (Institute Louis Bachelier).
- Baltic-Nordic-Ukrainian Network Workshop on Survey Statistics 2017 that will take place in Vilnius, Lithuania, from 21-25 August 2017, (<a href="https://wiki.helsinki.fi/display/BNU/Events">https://wiki.helsinki.fi/display/BNU/Events</a>).
- **European Establishment Statistics Workshop (EESW) 2017** to be held in Southampton (UK) on August 30 September 1. For more information on EESW17 and the European Network for Better Establishment Statistics (ENBES), please visit <a href="https://www.enbes.org">www.enbes.org</a>.

In addition, IASS community will contribute to the success of the 61<sup>st</sup> World Statistics Congress – ISI2017 (<a href="http://www.isi2017.org/">http://www.isi2017.org/</a>) that will hold many sessions that were proposed by IASS members and then endorsed by our Association. We now call the IASS colleagues to promote the 2017 WSC, submitting contributed papers and attending the event. I hope to meet many of you in Marrakesh.

I would also like to raise your attention that submission for the Cochran-Hansen Prize 2017 (competition for young survey statisticians from developing and transitional countries) is still open. We strongly encourage the young, early career statisticians, to consider applying for this prestigious award and we ask our members to promote this prize competition. For information on how to submit a paper for consideration in the competition, eligibility rules and requirements for the paper, please visit the IASS website (http://isi-iass.org/home/cochran-hansen-prize/) or contact Denise Silva (denisebritz@gmail.com). The deadline for submission is February 15, 2017. All papers must be sent to the chair of the IASS 2017 C-H Prize Committee, Professor Monica Pratesi (monica.pratesi@unipi.it).

All members are invited to contribute to our newsletter by volunteering to send articles, book/software reviews and country reports to the next issue of *The Survey Statistician* to be published in July 2017. If you would like to suggest a subject, or write an article, to the *New and Emerging Methods* section, please get in touch with me at <a href="mailto:denisebritz@gmail.com">denisebritz@gmail.com</a> or <a href="mailto:denisebritz@gmail.com">denise.silva@ibge.gov.br</a>. This would not be a traditional scientific paper but an informative article of at most 8-10 pages, introducing the challenges, the methods, their uses and applications, and also the relevance of the subject for the development of survey methods. We are truly looking forward to get your views and contributions.

Please send a suggestion and share your knowledge with the IASS community.

Very best wishes Denise Silva

#### **News and Announcements**

#### Nominations Sought for the 2018 Waksberg Award

The journal *Survey Methodology* has established an annual invited paper series in honor of Joe Waksberg to recognize his contributions to survey methodology. Each year a prominent survey statistician is chosen to write a paper that reviews the development and current state of an important topic in the field of survey methodology. The paper reflects the mixture of theory and practice that characterized Joe Waksberg's work.

The recipient of the Waksberg Award will receive an honorarium and give the 2018 Waksberg Invited Address at the Statistics Canada Symposium to be held in the autumn of 2018. The paper will be published in a future issue of *Survey Methodology* (Targeted for December 2018).

The author of the 2018 Waksberg paper will be selected by a four-person committee appointed by *Survey Methodology* and the American Statistical Association. Nomination of individuals to be considered as authors or suggestions for topics should be sent before February 28, 2017 to the chair of the committee, Kirk Wolter wolter-kirk@norc.uchicago.edu.

Previous Waksberg Award honorees and their invited papers are:

2001 Gad Nathan, "Telesurvey Methodologies for Household Surveys-A Review and Some Thoughts for the Future." *Survey Methodology*, vol. 27, no. 1, pp.7-31.

2002 Wayne Fuller, "Regression Estimation for Survey Samples." *Survey Methodology,* vol. 28, no. 1, pp. 5-23.

2003 David Holt, "Methodological Issues in the Development and Use of Statistical Indicators for International Comparisons." *Survey Methodology*, vol. 29, no. 1, pp. 5-17.

2004 Norman Bradburn, "Understanding the Question-Answer Process." *Survey Methodology*, vol. 30, no. 1, pp. 5-15.

2005 J.N.K Rao, "Interplay Between Sample Survey Theory and Practice: An Appraisal." *Survey Methodology*, vol. 31, no. 2, pp. 117-138.

2006 Alastair Scott, "Population-Based Case Control Studies." *Survey Methodology,* vol. 32, no. 2, pp. 123-132.

2007 Carl-Erik Särndal, "The Calibration Approach in Survey Theory and Practice." *Survey Methodology*, vol. 33, no. 2, pp. 99-119.

2008 Mary Thompson, "International surveys: Motives and Methodologies." *Survey Methodology*, vol. 34, no. 2, pp. 131-141.

2009 Graham Kalton, "Methods for Oversampling Rare Subpopulations in Social Surveys." *Survey Methodology*, vol. 35, no. 2, pp. 125-141.

2010 Ivan Fellegi, "The organisation of statistical methodology and methodological research in national statistical offices." *Survey Methodology*, vol. 36, no. 2, pp. 123-130.

2011 Danny Pfeffermann. Modelling of complex survey data: Why is it a problem? How can we approach it? *Survey Methodology*, vol. 37, no. 2, pp. 116-136.

2012 Lars Lyberg, Manuscript topic "Survey Quality". *Survey Methodology*, vol. 38, no. 2, pp. 107-130.

2013 Ken Brewer, Manuscript topic "Three survey sampling controversies". *Survey Methodology*, vol. 39, no. 2, pp. 249-262.

2014 Connie Citro, "From Multiple Modes for Surveys to Multiple Data Sources for Estimates". *Survey Methodology*, vol. 40, no. 2, pp. 137-161.

2015 Robert M. Groves, Manuscript topic under consideration. To appear in the December 2016 issue of *Survey Methodology*.

2016 Don A. Dillman, Manuscript topic under consideration. To appear in the June 2017 issue of *Survey Methodology*.

2017 Don Rubin, Manuscript topic under consideration.



# Cochran-Hansen Prize 2017: Competition for Young Survey Statisticians From Developing and Transitional Countries

In celebration of its 25th anniversary in 1999, the International Association of Survey Statisticians (IASS) established the Cochran-Hansen Prize, which is awarded every two years for the best paper on survey research methods submitted by a young statistician from a developing or transition country. The Cochran-Hansen Prize consists of books and journal subscriptions in the value of € 500.

Participation in competition for the prize is restricted to young statisticians from developing and transition countries who are living in such countries and were born in 1987 or later.

A paper submitted for the competition must consist of original work which is either unpublished or has been published after January 1st 2016. A paper may be based on a university thesis and should be written in English. The deadline for submission of papers for the 2017 prize is 15 February 2017. All papers must be sent to the chair of the IASS 2017 Cochran-Hansen Prize Committee, Professor Monica Pratesi, email address; monica.pratesi@unipi.it.

Each submission must be accompanied by a cover letter, stating the author's year of birth, nationality and country of residence. The cover letter should also indicate if the paper submitted is based on a PhD thesis and, in the case of a joint paper, the contribution to the paper made by the prize competitor. The papers submitted will be reviewed by members of the Cochran-Hansen Prize Committee appointed by the IASS. The decision of the Committee will be final.

The prize winner will be invited to present his or her paper at the World Statistics Congress of the International Statistical Institute (ISI) to be held in Marrakech, Morocco, in 2017; http://www.isi2017.org. The IASS will provide the winner round trip economy class airfare and per diem for living expenses for the days that they participate in the 2017 WSC meetings in Marrakech.

For further information, please contact:

Professor Monica Pratesi Department of Statistics University of Pisa

Email: monica.pratesi@unipi.it

January 2017

#### **News and Announcements**

#### **ICES-V Highlights**

The Fifth International Conference on Establishment Surveys (ICES-V) was held for the first time in Europe, from 20-23 June 2016 in Geneva, Switzerland. The program included two keynote addresses, 16 invited sessions, 12 topic contributed sessions, and 20 contributed sessions, with a total of 189 presentations. In addition, the conference held a student contest, six short courses, seven introductory overview lectures, and four software demonstration sessions. ICES-V was well attended, with 324 persons attending from 43 countries, including many IASS members who took advantage of IASS sponsorship and received a discount on registration.

With so many sessions, it is difficult to summarize the program, but here are a few highlights. The two keynote speakers, Peter van de Ven of the Organisation for Economic Co-operation and Development and Connie Citro of the U.S. National Academies of Sciences, Engineering and Medicine, spoke on globalization, defining the statistical unit of measurement, the integration of administrative and survey data, and the use of multiple data sources – themes repeated in many of the conference sessions.

As establishment survey and statistical measurement moves into the future, the trend towards statistics on an international level led to discussions on statistical units, the structure of complex units, integrating statistics on a global scale, and breaking down stovepipes within and across national statistical institutes, countries, and continents. Another issue of great relevance covered multi-source and mixed-mode designs, including moving from fixed samples to efficient designs to obtain a representative response distribution, and targeting data collection efforts to strata that need more attention — with the added complexity of changing and heterogeneous business populations that can be difficult to reach. Respondent burden continued to be a key area of concern at ICES-V — covering a wide range of topics — the use of multiple data sources, the need for more auxiliary variables, imputation and estimation issues, as well as a focus on respondent relations and experiences.

The ICES-V introductory overview lectures complemented conference discussions with speakers on adaptive design, big data, administrative data, imputation, disclosure, and economic classification. Six short courses were held and were well attended, covering sampling, business survey data collection, record linkage, and estimation. Another highlight was the 15 software demonstrations, where participants could sit down with developers and view and discuss their work. They covered a wide range of common tasks, including sampling, questionnaire design, data collection and management,

processing, estimation, modeling, and geographic information systems. Importantly, thinking about the next generation of ICES contributors, the Program Committee developed and carried out a student contest to help create interest in establishment survey work and test technical skills and creativity.

The benefits of ICES include opportunities to catch up with colleagues, make new contacts, and learn about and discuss current topics and issues in the field of establishment surveys and statistics. New and experimental ideas were presented, as well as further development and progress on topics presented in past conferences. We encourage you to attend ICES-VI, the location and dates are not yet certain, but the conference is likely to be held in 2020. We thank the many sponsors for their support, as well as the Swiss Federal Statistical Office for hosting the conference.

#### The 9th French Colloquium on Survey Sampling

The 9th French Colloquium on Survey Sampling (Colloque francophone sur les sondages) was held on October 11-14, 2016, at the Université du Québec en Outaouais in Gatineau (Québec, Canada). The Colloquium attracted about 250 participants. During the Colloquium, 78 presentations were given in 27 sessions. The first day of the Colloquium, there was three workshops on the following topics: data editing; inferential approaches in survey sampling; and SAS procedures for survey sampling. The opening speech has been given by Yves Tillé (Université de Neuchâtel, Switzerland) on "Selecting samples adequately", and the closing speech was by Sylvie Michaud (Statistics Canada) on "Future challenges in methodology". The Waksberg Award has been given to Don Dillman (Washington State University) who talked on "The promises and perils of pushing respondents to the web in mixed-mode surveys". Five participants from developing countries received a bursary to participate to the Colloquium. The International Association of Survey Statisticians (IASS) financed two of these: Aziz Ilboudo (Burkina Faso) and Arouna Sow (Burkina Faso). Feedbacks from the participants were very good, which indicates that the 9th French Colloquium on Survey Sampling was a success.

For more information, please contact: Pierre Lavallée Pierre.Lavallée@canada.ca







Stephen Fienberg (Z"L) sadly passed away on Wednesday, December 14<sup>th</sup>, 2016 at the age of 74. His obituary mentioning his many acclaims, achievements and his amazing contribution to science and to public policies over his lifetime can be found here: <a href="http://www.cmu.edu/dietrich/news/news-stories/2016/december/stephen-fienberg-obituary.html">http://www.cmu.edu/dietrich/news/news-stories/2016/december/stephen-fienberg-obituary.html</a>

But we would like to focus on one aspect of his contribution – his development and nurturing of the next generation of dedicated statisticians and data scientists. Below are contributions from some of Steve's many mentees over the years:

Jerry Reiter: Steve Fienberg had an enormous influence on how I think about statistical science and a huge impact on my career. Steve's research is of course legendary; he made fundamental contributions to Bayesian inference, categorical data analysis, disclosure control, official statistics, and record linkage, to name key areas where our research interests overlapped. The statistical insights in these works are brilliant and far-reaching. They were especially important to me, directly inspiring much of my work including my first paper in JASA. The innovations in Steve's work target and respond to real-world, high impact problems in official statistics, resulting in leading edge methodology intended to make it into practice. His example as an academic statistician who improves the practice of official statistics is one that I strive to emulate. Although I was not one of Steve's students, I was the beneficiary of his mentorship. At conferences and workshops, and at reciprocating visits to CMU and Duke, Steve generously gave me many hours of his time, answering questions about research, official statistics, being a mentor to students, career options and paths, and life in general. I valued his advice dearly and followed it. I also benefited from Steve's mentorship more indirectly: I am the post-doc mentor to two of his Ph.D. students. I can say from first-hand experience that Steve's students are wonderful, both as scholars and as people. Of course, perhaps one should expect nothing less: they trained under Steve Fienberg, one of the greatest statisticians of all time.

**Natalie Shlomo:** I first met Steve as a (mature) PhD student in 2004 at the Hebrew University and the University of Southampton. My thesis topic was on Statistical Disclosure Control. I was subsequently invited to take part in a unique workshop held in Bertinoro, Italy in July 2005 organized by Steve. The aim was to bring together statisticians and computer scientists to disseminate research on confidentiality and privacy. At that time, both communities were working in this area of research separately. This was especially noteworthy in Bertinoro where researchers basically talked to others in their own community and there was little interaction. Steve changed all that. He brought differential privacy to the forefront of research in statistical disclosure control. He envisioned and established a unique statistical journal: The Journal of Privacy and

Confidentiality (JPC) where articles from both communities are published. The next joint meeting between statisticians and computer scientists was the workshop on Statistical and Learning-Theoretic Challenges in Data Privacy held at UCLA's Institute for Pure and Applied Mathematics in February 2010 which culminated in a special issue of the JPC in Volume 4, Issue 1 (2012) (See: http://repository.cmu.edu/jpc/vol4/iss1/ ). Finally, our most recent initiative was the Data Linkage and Anonymisation Programme which was held July through December 2016 at the Isaac Newton Institute of Mathematical Sciences at the University of Cambridge. Steve was to participate in the programme but alas his illness took the better of him during that time. In fact, Steve was to participate in all three programmes that were running at the Institute: Data Linkage and Anonymisation, Theoretical **Foundations** Statistical for Analysis and Probability and Statistics in Forensic Science which only goes to show the breadth and depth of his research activities and achievements. He was sorely missed. On a personal note, Steve marked my PhD dissertation, mentored me, provided me with endless advice, recommendation and promotion letters throughout my career. I had a wonderful visit to CMU during my sabbatical period in November 2011 spending quality time with him and Joyce. I can honestly credit Steve with where I am at today - a Professor in Social Statistics - at the University of Manchester having only entered into academia following the completion of my PhD in 2007. I will be forever grateful to him.

Mauricio Sadinle: I want to share the story of how I started interacting with Steve, because I think it shows how open and generous he was. Back in Colombia when I was an undergraduate student, I was involved in a project that had the goal of combining different sources of information to estimate the number of people who had been displaced due to the Colombian armed conflict. This is how I came across record linkage and capture-recapture methods. After learning the basics of these methodologies, I found that the literature didn't deal with some of the problems I was facing with my data. I realized that there was an author that kept coming up in the literature of these two areas: Stephen E. Fienberg. I decided to email Steve with my very rudimentary English of the time: "Dear Professor Fienberg, .." (Of course at the time I didn't know how famous and busy he was). I explained the problem I was working on and the difficulties that I was finding. To my surprise Steve was very prompt to reply, provided very helpful literature and references, and his email was quite extensive. Nevertheless, I realized that many of the issues that I was facing had not been fully addressed in the literature, so it seemed like this was an area that needed more research. Months later I emailed Steve saying that I wanted to do research on this, and that I wanted to "study a masters in capture-recapture estimation," to which he nicely replied that there wasn't such a thing and that I'd have to enroll in a PhD program. At the time probably not many of the things that I was writing made much sense, so I now realize how nice he was by taking me seriously and providing me guidance! Our email exchanges continued for around two years, until I finally enrolled in the PhD program at CMU, where I had the fortune of having him as my advisor. I will always be grateful for the trust that Steve put in me and for all the support and mentoring that he gave me.

**Aleksandra Slavkovic**: I last saw Steve two months before his passing, at CMU's celebration of his transfer to emeritus status. I had a chance to speak about his contributions to the area of data privacy and confidentiality, which of course included a story about the Bertinoro workshop that Natalie Shlomo mentioned above. That story truly depicts the impact that Steve had on many of us, on statistics and other related fields, and if you are interested you can read more in the special issue of CHANCE from

four years ago ("Steve the Matchmaker: The Marriage of Statistics and Computer Science in the World of Data Privacy." CHANCE, Vol. 26, No. 4.). At the celebration there were many of Steve's friends and colleagues, former and current students and mentees, and many told short stories of how at one time or another Steve impacted their career and life, and these include even prominent statisticians such as Steve Stigler (CMU's first Fienberg Lecture), Alicia Cariguirry, Ed George, Larry Wasserman, to name only a few. For me personally, I became a statistician and academic, mostly due to Steve's influence. I took my very first statistics course with Steve at CMU in the spring of 1998, and shortly after that he recruited me for graduate studies at CMU. But more importantly, throughout our interactions at CMU, we had many discussions and he presented me with many wonderful opportunities that taught me how valuable statistics is, how wonderful it interplays with so many areas and disciplines and can address reallife problems, and I was hooked. Even after I began as a young faculty at Penn State Statistics Department in August 2004, we have continued working on shared research projects, editorial work, student advising, and he selflessly continued to be available as a friend and a mentor. Steve was simply put a great mentor. He tried to instill confidence in his students and advisees, to have them do their best and to make you believe in what you are doing can make a difference, like his work has. He himself said that was a big part of his legacy ("A Conversation with Stephen E. Fienberg" by M.Straf and J. Tanur, Stats.Sci., Vol. 28., No. 3, pp. 447-463, 2013). Many of us have become members of Steve and Joyce Fienberg extended family. I am very fortunate to have had Steve as my Ph.D. adviser, mentor, colleague and a dear friend. The void will not be filled but memories of him will leave on.

Bradley Malin: I met Steve as an undergraduate at Carnegie Mellon University, while I was serving as a research assistant to Latanya Sweeney. I was greatly interested in data privacy both from a technical and public policy perspective. As I read up on the field, I was introduced to all of the amazing research and application that transpired in the statistical disclosure control community, of which Steve was clearly one of the leaders. I was fortunate to be able to grab some of his time and discuss the differences and relationships between how the computer science and statistics community were pursuing data protection and was quite humbled by the fact that he spent so much time with me. I stayed at CMU and joined the graduate program in the Center for Automated Learning (now the Machine Learning Department), where I was fortunate to have Steve as my instructor for my intermediate statistics course. In addition to being exposed to his greatness in each lecture, I was further influenced by him to push the boundaries of research in data privacy. After finishing my doctorate, I was fortunate to remain in contact with Steve. He was a champion of my research and served as the editor for research I published on anonymization in the Proceedings of the National Academy of Sciences. We continued to cross paths at numerous meetings focusing on privacy, statistics, and computation, including the 2010 UCLA workshop mentioned by Natalie Shlomo and the Joint Statistical Meetings. Steve was always pushing me to be formal, yet practical, in my methodologies and viewpoints and I will truly miss him.

May his memory be a blessing (zichrono l'vracha)





# "How can we continue to protect the confidentiality of data subjects under open access initiatives?" By Natalie Shlomo, University of Manchester

#### 1. Introduction

Traditionally, statistical agencies release outputs in the form of microdata and tabular data. Microdata contain data from social surveys and tabular data contain either frequency counts, such as for census dissemination, or magnitude data typically arising from business surveys, e.g., total revenue. For each of these traditional outputs, there has been much research on how to quantify disclosure risk, optimal statistical disclosure control (SDC) methods, and how to assess the impact on data utility.

In traditional outputs, the two main disclosure risks are identity disclosure, where a statistical unit can be identified based on a set of identifying variables, and attribute disclosure, where new information can be learnt about an individual or a group of individuals. One disclosure risk that is often overlooked in traditional outputs is inferential disclosure. This disclosure risk has to do with learning new attributes with high probability. For example, a regression model with a very high predictive power may cause inferential disclosure. Even if an individual is not in the dataset, there would still be disclosure from this type of disclosure risk. Another example of inferential disclosure is disclosure by differencing when multiple releases are disseminated from one data source. For example, census tables can be differenced or manipulated to reveal individual units. For traditional hard-copy census tables, disclosure by differencing is controlled by having a fixed set of variables and categories which disallow differencing non-nested groups of individuals.

In the last decade, many statistical agencies have set up data enclaves on their premises. These are largely motivated by statistical agencies being unable to continue to meet demands for large amounts of data and still ensure the confidentiality of data subjects. The data enclave is a secure environment where researchers go on-site and gain access to confidential data. The secure servers have no connection to printers or the internet and only authorized users are allowed to access them. To minimize disclosure risk, no data is allowed to be removed from the enclave and researchers undergo training to understand the security rules. All

information flow is controlled and monitored. Any outputs to be taken out of the data enclave are dropped in a folder and manually checked by experienced confidentiality officers for disclosure risks. Examples of disclosure risks in outputs are small cell counts in tables, residual plots which may denote outliers, and Kernel density estimation with small band-widths.

The disadvantage of the data enclave is the need to travel, sometimes long distances, to access confidential data. In addition, access to the data enclave can be offset by a prolonged application process. In very recent years, some statistical agencies have been extending the concept of the data enclave to remote access through a 'virtual' data enclave. These 'virtual' data enclaves have been set up at organizations and Universities and allow users within the same group to interact with one another while working with the data. Users log on to secure servers to access the confidential data—and all activity is logged and audited at the keystroke level. Generally, the secure data lab must be approved by the agencies and outputs are reviewed remotely by confidentiality officers before being sent back to the researchers via a secure file transfer protocol site.

The question remains therefore on whether there are still opportunities to release statistical data under open access initiatives in the age of expanding data environments and Big Data which can be used to breach the confidentiality of data subjects. For that, we will need to look at the Computer Science literature on how they define and meet strict privacy guarantees.

#### 2. Disclosure Risk in Query Systems and Differential Privacy

The theory of differential privacy for protecting outputs in a remote query-based system has been widely presented in the computer science literature and it is closely related to the concept of inferential disclosure. See Dinur and Nissim, 2003, and Dwork, McSherry, Nissim, and Smith, 2006, for more details on differential privacy. Shlomo and Skinner, 2012, discuss differential privacy with respect to sampling and perturbation. A 'worst case' scenario is allowed for, in which the intruder has complete information about all the units in the database except for one unit of interest. Under this assumption, let x denote a cell value, taking possible values 1,...,k where the table is formed by cross-classifying all variables in the database. The population database is denoted by  $\mathbf{x}_{IJ} = (x_1, ..., x_N)$  where N denotes the size of the population  $U=\{1,...,N\}$ . We assume that sampling would be part of the SDC mechanism to protect the population database and that the intruder would not have response knowledge of who is in the sample. Let  $\tilde{x}_i$  denote the cell value of unit i in the database after an SDC method has been applied and define  $\tilde{f}_j = \sum_{i \in s} I(\tilde{x}_i = j)$ as the observed count in cell j where s denotes a sample drawn randomly from the population. We can view the released data as the vector of counts:  $\tilde{\mathbf{f}} = (\tilde{f}_1, \tilde{f}_2, ..., \tilde{f}_k)$ . Let  $\Pr(\tilde{\mathbf{f}} \mid \mathbf{x}_U)$  denote the probability of  $\tilde{\mathbf{f}}$  with respect to an SDC method, which includes sampling and/or perturbation, and where  $\mathbf{x}_{tt}$  is treated as fixed. In this framework, the definition of differential privacy is as follows:

Definition (Dwork et al, 2006):  $\varepsilon$  - differential privacy holds if:  $\max \left| \ln \left( \frac{\Pr[\tilde{\mathbf{f}} \mid \mathbf{x}_{U}^{(1)}]}{\Pr[\tilde{\mathbf{f}} \mid \mathbf{x}_{U}^{(2)}]} \right) \right| \le \varepsilon$ 

for some  $\varepsilon > 0$ , where the maximum is over all pairs  $(\mathbf{x}_{\scriptscriptstyle U}^{\scriptscriptstyle (1)},\mathbf{x}_{\scriptscriptstyle U}^{\scriptscriptstyle (2)})$ , which differ in only one element and across all possible values of  $\widetilde{\mathbf{f}}$ .

Differential privacy therefore aims to avoid inferential disclosure by ensuring that an intruder cannot make inference about a single unit when only one of its value is changed given that all other units in the populations are known. This definition would control for disclosure by differencing and highly predictive models. The solution to guarantee differential privacy in the computer science literature is by adding noise/perturbation to outputs of the queries under specific parameterizations. One advantage to differential privacy is that the parameters of the perturbation methods can be made public and hence researchers can use this information to carry out statistical analysis.

Shlomo and Skinner, 2012, show that under this definition sampling alone would not be differentially private although the rate of leakage would depend on the number of population unique in the vector of counts which could potentially be small and below the risk thresholds set by statistical agencies. Only perturbation where the SDC method has a non-zero chance of perturbing any record guarantees differential privacy. As in the Computer Science literature, adding noise to counts of a frequency table under specific parameterizations could be made differentially private but non-perturbative methods such as coarsening categories of the table would not.

#### 3. Open Access Initiatives

To see how differential privacy may be relevant for open access initiatives we look at three applications considered by statistical agencies for disseminating statistical outputs.

#### 3.1 Flexible Table Generating Servers

Driven by demand from policy makers and researchers for specialized and tailored tables from statistical data (particularly census data), some statistical agencies are considering the development of flexible table generating servers that allow users to define and generate their own tables. The United States Census Bureau and the Australian Bureau of Statistics have developed such servers for disseminating census tables, and the Israel Central Bureau of Statistics developed a server to disseminate tables of weighted counts from the Social Survey. Users access the servers via the internet and define their own table of interest from a set of pre-defined variables and categories typically from drop down lists.

When selecting SDC methods to apply to the output table, there are two approaches: apply SDC to the underlying data so that all tables generated in the server are deemed safe for dissemination (pre-tabular SDC), or produce tables directly from original data and apply the SDC method to the final tabular output (post-tabular SDC). Although sometimes a neater and less resource intensive approach for data from a single source, the pre-tabular approach is problematic since it will require a large amount of SDC methods which will have a compounding effect under aggregation

and reduce data utility. The post-tabular approach is motivated by the computer science definition of differential privacy as discussed in Section 2. When carrying out post-tabular perturbation methods, consistency across same cells generated in different tables is necessary to avoid the possibility of revealing information about the SDC method. This can be carried out by microdata keys. For each record in the microdata, a random number is defined which when combined with other records to form a cell of a table defines the seed for the perturbation. Records that are aggregated into same cells will always have the same seed and therefore a consistent perturbation (Fraser and Wooton, 2005, Shlomo and Young, 2008).

The design of remote table generating servers typically involves many ad-hoc preliminary SDC rules that can easily be programmed within the system to determine whether a table can be released. These SDC rules may include limiting the number of dimensions in the table, minimum population thresholds, and ensuring consistent and nested categories of variables to avoid disclosure by differencing.

#### 3.2 Remote Analysis Servers

A remote analysis server is an online system which accepts a query from the user, runs it within a secure environment on the underlying data and returns a confidentialized output without the need for human intervention to manually check the outputs for disclosure risks. The confidentialized output should follow the principles of differential privacy as a strict privacy guarantee.

Similar to flexible table generators, the queries are submitted through a remote interface and users do not have direct access to the data. The queries may include exploratory analysis, measures of association, regression models and statistical testing. The queries can be run on the original data or confidentialized data and may be restricted and audited depending on the level of required protection. O'Keefe and Good, 2008, describe regression modeling via a remote analysis server.

#### 3.3 Synthetic Data

In recent years, there has been a move to produce synthetic microdata as public-use files which preserve some of the statistical properties of the original microdata. The data elements are replaced with synthetic values sampled from an appropriate probability model. The model is fit to the original data to produce synthetic populations through a posterior predictive distribution similar to the theory of multiple imputation. Several samples are drawn from the population to take into account the uncertainty of the model and to obtain proper variance estimates. See Raghunathan, Reiter and Rubin, 2003, and Reiter, 2005, and references therein for more details on generating synthetic data. The synthetic data can be implemented on parts of the data so that a mixture of real and synthetic data is released. One application which is motivated by differential privacy is the US Census Bureau 'On the Map' available at: <a href="http://onthemap.ces.census.gov/">http://onthemap.ces.census.gov/</a>. It is a web-based mapping and reporting application that shows where workers are employed and where they live according to the Origin-Destination Employment Statistics. More information is given in Abowd and Vilhuber, 2008.

In practice it is very difficult to capture all conditional relationships between variables and within sub-populations. If models used in a statistical analysis are sub-models of the model used to generate data, then the analysis of multiple synthetic samples should give valid inferences. In addition, partially synthetic datasets may still have disclosure risks and need to be checked prior to dissemination.

For tabular data there are also techniques to develop synthetic magnitude tables arising from business statistics. Controlled tabular adjustment (CTA) carries out cell suppression and replaces the suppressed cells with imputed values that guarantee some statistical properties (Dandekar and Cox, 2002).

#### References:

Abowd, J.M. and Vilhuber, L., (2008). How Protective Are Synthetic Data? In *PSD'2008 Privacy in Statistical Databases*, (Eds. J.Domingo-Ferrer and Y. Saygin), Springer LNCS 5262, 239-246.

Dandekar, R.A. and Cox L. H. (2002). Synthetic Tabular Data: An Alternative to Complementary Cell Suppression. *Manuscript, Energy Information Administration*, U. S. Department of Energy.

Dinur, I. and Nissim, K. (2003). Revealing Information While Preserving Privacy. *PODS* 2003, 202-210.

Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography TCC* (eds. S. Halevi and R. Rabin). Heidelberg: Springer, LNCS 3876, 265-284.

Fraser, B. and Wooton, J. (2005). A Proposed Method for Confidentialising Tabular Output to Protect Against Differencing. *Joint UNECE/Eurostat work session on statistical data confidentiality*, Geneva, 9-11 November.

O'Keefe, C.M. and Good, N. (2008). A Remote analysis Server – What Does Regression Output Look Like? In *PSD'2008 Privacy in Statistical Databases*, (Eds. J.Domingo-Ferrer and Y. Saygin), Springer LNCS 5262, 270-283.

Raghunathan, T.E., Reiter, J. and Rubin, D. (2003). Multiple Imputation for Statistical Disclosure Limitation. *Journal of Official Statistics*, 19, No. 1, 1-16.

Reiter, J.P. (2005), Releasing Multiply Imputed, Synthetic Public-Use Microdata: An Illustration and Empirical Study. *Journal of the Royal Statistical Society*, A, Vol.168, No.1, 185-205.

Shlomo, N. and Skinner. C.J. (2012). Privacy Protection from Sampling and Perturbation in Survey Microdata. *Journal of Privacy and Confidentiality*, Vol. 4 (1).

Shlomo, N. and Young, C. (2008). Invariant Post-tabular Protection of Census Frequency Counts. In *PSD'2008 Privacy in Statistical Databases*, (Eds. J.Domingo-Ferrer and Y. Saygin), Springer LNCS 5262, 77-89.



#### **New and Emerging Methods**

# Handbook on Master Sampling Frames for Agricultural Statistics: Frame Development, Sample Design, and Estimation By Frederic Vogel and Cristiano Ferraz

The concept underlying the development of a handbook on master sampling frames for agricultural statistics has its roots in the Global Strategy to Improve Agricultural and Rural Statistics (World Bank 2011). The Global Strategy provided a ground-breaking effort to improve agricultural statistics that also has implications for other sectors in the national statistical systems.

The Global Strategy addressed the lack of direction regarding agricultural data requirements posed by the Millennium Development Goals and other emerging issues such as the use of food for biofuels, food security, and the environment. These emerging issues were coming during a time of a general decline in the overall quality and availability of agricultural statistics.

The Global Strategy was based on a thorough assessment of the data that users need and that are currently available. This data assessment led to the formulation of a conceptual framework that identified three dimensions of agriculture—economic, social, and environmental. These dimensions include the farm as the economic unit, the farm household as the social unit, and the land they occupy and use as the environmental unit.

The Global Strategy also addressed the issue that in most countries, agriculture is outside the national statistical system and often based on administrative data rather than statistical methodology. The need to consider the three dimensions of agriculture and the related farm, farm household, and land units pointed the need for the major pillar of the Global Strategy which was to integrate agriculture into the national statistical system. The Global Strategy then laid the foundation for a master sampling frame for agricultural and rural statistics and its use in an integrated survey system.

The Handbook on the Master Sampling Frame (MSF) provides the foundation for the selection of probability-based samples of farms, households, and land that enables the characteristics of the farm to be connected with those of the household, and both to the land cover and land use dimensions. The Handbook was prepared by a core team of five senior consultants and experts (see footnote below) and it is available free of charge.

Currently there are versions in English, Spanish and French that can be downloaded at: <a href="http://gsars.org/en/tag/mastersamplingframe/">http://gsars.org/en/tag/mastersamplingframe/</a>

In this short article the main ideas behind the publication are briefly introduced to provide an overview for those interested in the theme to go for a deeper reading of the Handbook.

#### **Basic Principles underlying the Master Sampling Frame**

The challenge in developing sampling frames for agriculture is that they must satisfy the needs of three different statistical units: the farm or agricultural holding, the farm or rural household, and the land the holdings and households use and occupy. In this scenario, specific frames could be used for targeting each type of sampling unit, with specific surveys being carried out separately (agricultural surveys, household surveys, and farm management surveys). However, if a single frame can be employed, this would save efforts, and improve efficiency for designing samples. A frame that allows for selecting different survey samples and accounting for different types of sampling units, is called a Master Sampling Frame. In agriculture, the three units (farm, household, and the land) must be linked so that household income, health and other measures may be compared to the farm's economic situation; and all of these to their impact on the environment and land use. Often there is a one-toone relationship between the agricultural holding, the household, and the land parcel making it easy to collect economic, social, and land use information from a single statistical unit. If these units are geo-referenced, the three units can also be associated with land cover and the environment. A challenge facing the development of MSF's occurs when there is not always a one-to-one link between the agricultural holding and the household. A challenge facing the development of MSF's occurs when there is not always a one-to-one link between the agricultural holding and the household. Therefore, one of the basic principles underlying the MSF is to employ geo-referencing methods for all statistical units.

Another principle is that there must be a careful linkage of the sampling units to reporting units so that there is complete coverage of the populations of interest, as well as a linking of the agricultural holding to the household and land dimensions. This listing of sampling units may consist of a list of the names of farm operators obtained through an agricultural census, of households from the population census, a list of commercial agricultural enterprises that are not associated with a household, or a list of area units defined geographically. The Handbook provides an in-depth view of steps to develop list sampling frames as well as area sampling frames. This is followed by considerable detail on the use of multiple frame sampling involving the joint use of two or more different sampling units to meet the coverage principle.

The Handbook is based on a long-term vision for how the MSF should be developed and used.

<sup>-----</sup>

<sup>\*</sup> Cristiano Ferraz, Javier Gallego, Miguel Galmes, Naman Keita, Frederic Vogel—also the technical editor. The publication was prepared with the support of the Trust Fund of the Global Strategy funded by the UK's Department for International Development and the Bill & Melinda Gates foundation. The World Bank and the Joint Research Center of the European Union also provided financial and technical support.

In order to capture the linkage between the dimensions of agriculture, the vision begins by defining how satellite imagery of the country's land mass be obtained that provides land cover by broad classifications. The vision continues by geo-referencing the Enumeration Areas from the population and agricultural censuses. Administrative areas such as villages, counties, etc., should also be geo-referenced. This would enable a better coordination of the difference censuses and other data collection efforts.

The Handbook begins by defining the data items to be measured. This is an issue often overlooked by textbooks and papers on sampling theory. Each data item can be represented by different populations from which data can be collected. These populations can be defined as listings of agricultural holdings, households, or blocks of land from which data are collected by means of a census or sample survey to provide measures that represent the data items.

The Handbook provides the framework to define the target population, the frame (or frames), the sampling unit and associated reporting units, how to determine the sample size, developing selection procedures, preparing the estimators and sampling error measures consistent with the sample design, and implementing statistical controls for detecting and correcting non-sampling errors.

The Handbook provides a summary of sample design considerations for developing an MSF. Basic concepts underlying the choices of sampling frame and estimation are defined and include defining the target population, sub populations, the sampled population, variables of interest, sampling unit, observation unit, and the reporting unit. These must be carefully defined before making decisions about the MSF and associated estimators. The Handbook also provides a review of major sampling designs or methods of sample selection and their main statistical properties as they relate to agriculture. Where possible, the efficiency if each is evaluated. The main purpose is to ensure that there is a close relationship between the development of the MSF and the sampling methods to be used and the choice of estimators.

The properties of the sampling designs used for agriculture are outlined and the relative strengths and weaknesses discussed. The use of PPS sampling vs stratification, for example, is discussed along with the roles of cluster sampling in two-stage designs. The Handbook provides considerable input to the role of cluster sampling in a multiple stage context. A common mistake made by countries is that in the effort to achieve cost efficiencies, multiple stage sampling is implemented without considering the impact on the resulting sampling variability. An overview of the variance components inherent in two-stage sample designs is also shown in an annex.

Next, the Handbook brings together the available technology available for agricultural statistics. Geographic Information Systems (GIS) are defined along with an in-depth presentation of their use, for example, to geo-reference elements in a list frame, and to develop GIS based administrative registers as a basis to define an area frame. Global Navigation Satellite Systems and Global Positioning Systems (GPS) are also described. Examples are provided on how GPS can be used to define a sampling frame and its use in the field work of running a survey. The role of remote sensing using satellite imagery is also presented with an overview of the main types of satellite images and their use for agriculture. The review of technology concludes with

a discussion of the role of aerial photographs and the use of drones (unmanned aerial vehicles).

The next two chapters deal with using list frames to build and use MSFs and provides guidelines on developing and using an area sampling frame. The chapter on the use of list frames defines how to use a population census to build an MSF and provides guidelines for choosing PSUs in a multiple-stage design. This is followed on guidelines on the use of agricultural censuses to build MSFs. The role of business registers is also discussed along with a discussion of coverage errors, errors due to misclassification, and duplication. The chapter continues with a comparison of the advantages and disadvantages of list frames and the problems of multiplicity and non-sampling errors and concludes with issues regarding how to maintain and update a list frame based MSF.

The area frame chapter provides an exhaustive review of guidelines on developing and using an area sampling frame for the MSF. The units in an area sampling include points, transects (lines of a certain length), or pieces of land called segments. These are carefully defined using examples based on satellite imagery and other sources and also a description of how each has actually been used by a country. The chapter then presents tools to improve the sampling efficiency of an area frame beginning with how stratification by land use can be used in a single or multiple stage design or multi-phase sampling. An important part of the chapter is defining the reporting units that can be attached to each sampling unit; that is the "open", "closed", and "weighted" methods. The estimators associated with different aspects of the area frame are also shown. The chapter concludes with a review of how to link area frames with census enumeration areas or administrative areas.

The chapters on area and list frames provide a detailed review of methodology on the development and use of list and area frames. Both have strengths and weaknesses. Lists of farms with associated auxiliary data on size measures can be more statistically efficient for sample survey purposes than other forms of sampling. However, they rapidly become out of date and are prone to under coverage. Area frames, while complete, are more suited for measuring characteristics of small farms and commodities widely distributed in the population. Sample sizes must be large enough to control sampling variability if there are large farms in the population. Multiple frame sampling was introduced as a method that builds off the strengths of individual frames with sampling flexibility for each frame.

The chapters on list and area frames are followed by a chapter on their joint use in multiple frame sample designs.

Multiple frame sampling involves the joint use of two or more sample frames. For agricultural purposes, this usually involves the joint use of area and list frames. The basic principles underlying multiple frame sampling are the same as those which apply to single frame sampling. Sampling units and associated reporting units must be defined. The sampling unit for an area frame is a segment of land or a point for which a reporting unit is formed. Rules of association are used to link the area sampling unit to the reporting unit, which is the farm or a portion of the farm. The sampling unit from the list frame is the name of the farm or – as common in most developing countries – the name of the landholder or the farm operator. The reporting unit is the holding associated with the name. The joint use of both frames generates

domains of estimation in a dual frame design. Based on such domains, major estimators available in the literature are reviewed and statistically introduced.

A conclusion is that the choice of estimator should be based on simplicity. The chapter overviews problems in application of multiple frame sampling by the requirement that the overlap between frames be identified. While the term multiple frame sampling implies more than two frames can be used, the complexities in determining overlap between frames rules in favor of limiting the choice to two frames. The general conclusion is that the list frame be kept as small as possible containing mostly large commercial farms and those producing important but rare items. The annex contains a summary of country experiences; several comment they are moving to multiple frame sampling.

The Handbook concludes with a summary of country experiences. These experiences illustrate the diversity of country situations and national factors that guided their choice of sampling frame(s) to develop.



#### **Book and Software Review**

# Analysis of Poverty Data by Small Area Estimation Monica Pratesi (Editor) John Wiley & Sons By Yong You, ICCSMD, Statistics Canada

A book titled "Analysis of Poverty Data by Small Area Estimation", edited by Monica Pratesi, and was recently published by Wiley. Monica Pratesi is Professor of Statistics at the University of Pisa, where she is holder of the Jean Monnet Chair "Small Area Methods for Monitoring of Poverty and Living Conditions in EU". This book contains 20 chapters which are written by different researchers and experts in the area of small area estimation (SAE) research and poverty data analysis. The book provides a comprehensive guide to implementing state-of-the-art SAE methods for poverty studies and poverty mapping.

The first chapter, Chapter 1. Introduction on measuring poverty at local level using small area estimation methods, written by Monica Pratesi and Nicola Salvati, is an introductory chapter. In this chapter the authors first provide definition of poverty indicators as well as the direct and indirect estimators of poverty indictors at small area level. Then the authors summarize some data-related and estimation-related problems faced for the estimation of poverty indictors. Finally the authors provide a short review of model-assisted and model-based SAE methods for the estimation of poverty indicators. From Chapter 2 to Chapter 20, there are 19 chapters which are divided into six major parts as described below.

Part I (Definition of indicators and data collection and integration methods) includes three chapters: Chapter 2. Regional and local poverty measures written by Achille Lemmi and Tomasz Panek; Chapter 3. Administrative and survey data collection integration written by Alessandra Coli, Paolo Consolino and Marcello D'Orazio; Chapter 4. Small area methods and administrative data integration written by Li-Chun Zhang and Caterina Giusti. In Chapter 2 the authors propose a methodology of poverty measurement at the regional level. The authors provide a good overview of the comprehensive multidimensional approach to poverty measurement which is focused not only on the current household income but also on the past incomes and accumulated assets measured in non-monetary terms. Chapter 3 presents methods to integrate data from different data sources including record linkage and statistical matching. The authors also provide some examples of application in well-being and poverty studies using data integration at local levels. In Chapter 4 the authors discuss SAE methods based on administrative data integration, where the settings may or may not involve survey data. Particularly, Section 4.2 presents register-based SAE without any survey data, and Section 4.3 presents SAE based on integration of sample survey and register data. The most common sources of error and their implications are described from a theoretical perspective, and several real-life datasets are used for illustration.

Part II (Impact of sampling design, weighting and variance estimation) includes three chapters: Chapter 5. Impact of sampling designs in small area estimation with applications to poverty measurement written by Jan Pablo Burgard, Ralf Munnich and Thomas Zimmermann; Chapter 6. Model-assisted methods for small area estimation of poverty indicators written by Risto Lehtonen and Ari Veijanen; Chapter 7. Variance estimation for cumulative and longitudinal poverty indicators from panel data at regional level written by Gianni Betti, Francesca Gagliardi and Vijay Verma. In Chapter 5 the authors investigate design-based and model-based small area methods for estimating the at-risk-of-poverty rate (ARPR). The authors consider different sampling design scenarios to elaborate the impact of sampling designs on poverty estimates and compare the estimation methods through a Monte Carlo study. In Chapter 6 the authors discuss design-based model-assisted methods for the estimation of indicators on inequality and poverty for small areas. Particularly they consider and compare various design-based estimators for Gini

index and model-assisted estimators for ARPR through simulation studies and empirical applications. Chapter 7 comprises nine sections to address the issue of estimating sampling error of cumulative and longitudinal poverty indicators. The authors introduce the most commonly used methods for cross-sectional variance estimation with an extension to cumulative and longitudinal poverty and inequality measures. An example is presented to illustrate the procedure to deal with additional difficulties caused by the lack of full information on sample structure in datasets such as EU-SILC.

Part III (Small area estimation modeling and robustness) includes three chapters: Chapter 8. Models in small area estimation when covariates are measured with error written by Serena Arima, Gauri S. Datta and Brunero Liseo; Chapter 9. Robust domain estimation of income-based inequality indicators written by Nikos Tzavidis and Stefano Marchetti; Chapter 10. Nonparametric regression methods for small area estimation written by M. Giovanna Ranalli, F. Jay Breidt and Jean D. Opsomer. Chapter 8 presents functional measurement error models for both area level and unit level models using frequentist and Bayesian methods. The authors describe empirical Bayes and hierarchical Bayes methods for structural measurement error models. The authors conclude with two real data applications to the US Census Bureau using Bayesian approach and EBLUP method. In Chapter 9 the authors present a small area estimation methodology based on Mquantile regression method to estimate inequality indicators. The authors also present a model-based simulation study and a real application to EU-SILC data to compare the proposed estimators. Chapter 10 is devoted to a review of nonparametric methods including nested error nonparametric unit level models using penalized splines and kernel methods. The authors also discuss other robust approaches such as M-quantile regression and model-based direct estimation (MBDE) approach. The chapter concludes with an application using data from the Living Standards Measurement Study in Albania.

Part IV (Spatio-temporal modeling of poverty) includes three chapters: Chaper 11. Area-level spatio-temporal small area estimation models written by María Dolores Esteban, Domingo Morales and Agustín Pérez; Chapter 12. Unit level spatio-temporal models written by Maria Chiara Pagliarella and Renato Salvatore; Chapter 13. Spatial information and geoadditive small area models written by Chiara Bocci and Alessandra Petrucci. In Chapter 11 the authors first review some spatial and temporal models as extensions of the Fay-Herriot model. Then the authors present a new area level model with MA (1) time correlation random effects. The EBLUP and MSE estimators are obtained and the authors also show how to apply the MA (1) model to obtain estimates of poverty proportions. Some simulations are presented to illustrate the introduced procedures. Chapter 12 is dedicated to the unit level spatio-temporal models in the context of EBLUP estimation. In this chapter the authors introduce unit level models, spatio-temporal

time-varying effects models and state space models with spatial structure. The corresponding EBLUP and MSE estimators are derived. The authors compare the different estimators with an application to household wealth study using the Italian EU-SILC data. Chapter 13 describes geoadditive models for skewed data after log transformation. The authors discuss two possible small area mean estimators based on EBLUP and MBDE. The authors present an example to estimate the average per-capita consumption expenditure at district level in Albania using the geoadditive SAE log-transformed model and the MBDE model-calibrated estimator.

Part V (Small area estimation of the distribution function of income and inequalities) includes four chapters: Chapter 14. Model-based direct estimation of a small area distribution function written by Hukum Chandra, Nicola Salvati and Ray Chambers; Chapter 15. Small area estimation for lognormal data written by Emily Berg, Hukum Chandra and Ray Chambers; Chapter 16. Bayesian beta regression models for the estimation of poverty and inequality parameters in small areas written by Enrico Fabrizi, Maria Rosaria Ferrante and Carlo Trivisano; Chapter 17. Empirical Bayes and hierarchical Bayes estimation of poverty measures for small areas written by Jon N. K. Rao and Isabel Molina. Chapter 14 describes the MBDE to estimate area-specific finite-population distribution of a response variable. The authors review the theory of MBDE and MSE estimation, and illustrate the use of MBDE to estimate the gender-specific distribution of equivalized income in three provinces of Italy. In Chapter 15 the authors focus on SAE methods for a unit level lognormal model. They provide a good review in Section 15.3 and compare three estimators including synthetic, model-based direct and empirical Bayes (EB) estimator. The authors present a simulation study to compare the estimators and to evaluate the robustness of the EB predictor. Chapter 16 presents Bayesian Beta regression models to estimate poverty and inequality parameters. The authors present different Beta regression models in details and corresponding data analysis for the estimation of the at-risk-of-poverty rate, the material deprivation rates and joint estimation of the at-risk-of-poverty rate and the Gini coefficient. Chapter 17 presents area level and unit level models to estimate Foster-Greer-Thorbeck (FGT) family of poverty measures. For the area level model, the authors describe the EBLUP and MSE estimation under the Fay-Herriot model. For the unit level model, the authors consider the nested error linear regression model and present the ELL (Elbers, Lanjouw and Lanjouw)/World Bank method, the EB and HB methods. The authors give a good overview of these methods and conclude with an example and remarks on the limitation of the proposed methods.

Part VI (Data analysis and applications) includes three chapters: Chapter 18. Small area estimation using both survey and census unit record data written by Stephen J. Haslett; Chapter 19. An overview of the U.S. Census Bureau's small area income and poverty estimates program written by William R. Bell, Wesley W.

Basel and Jerry J. Maples; Chapter 20. Poverty mapping for the Chilean comunas written by Carolina Casas-Cordero Valencia, Jenny Encina and Partha Lahiri. Chapter 18 presents a comprehensive overview of ELL method for small area estimation. Specifically the author describes ELL implementation process and ELL methodology in details, as well as the advantages, criticisms and disadvantages of ELL method. Chapter 19 presents an overview of the US Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program. The authors describe the US poverty measures, poverty data sources and poverty models with examples. The authors conclude the chapter with some discussion on current challenges and recent SAIPE research. Chapter 20 describes application of SAE methods implemented in Chile using the Casen 2009 data for Chilean poverty mapping. The authors present a description of the poverty measures and the data sources used in Chile, the data preparation process, the SAE models and estimation, and the results of poverty rate estimation.

The book is completed by an appendix chapter, Chapter 21. Appendix on software and codes used in the book, written by Antonella D'Agostino, Francesca Gagliardi and Laura Neri. In this final chapter the authors describe R and SAS software and scripts that are available on the book's website. The authors also briefly describe the EU-SILC survey and data. This appendix chapter is intended to provide guidance on how to use the R or SAS scripts for actually implementing the models and methods covered in the book.

As the book editor Monica Pratesi writes in the Preface, there exists no comprehensive source of information in the current literature regarding the use of SAE methods in poverty data analysis, and this book may serve to fill the gap. I think the book will be very useful to researchers and practitioners actively engaged in organizing, managing and conducting poverty studies and poverty mapping. The nice part about this book is that it not only covers theoretical SAE methods for poverty analysis, but also provides real applications and practical considerations in various scenarios and different countries. I think this book can also serve as a supplementary text in graduate seminar courses in survey sampling and SAE.



#### **AUSTRALIA**

**Reporting: Anthony Russo** 

## GLIDE – A new platform for visualising, exploring and analysing semantically linked data

Official statisticians are increasingly exploring the value of big data to answer the current and emerging information requirements of their users, especially when that data is linked with other datasets (including administrative and survey data) to provide large integrated pools of data. At the same time, advances in computing technology have provided new approaches to structure, store and dynamically visualise linked data to facilitate data exploration and aid analysis. In particular, the semantic web makes it easier to link data from multiple sources, retrieve it in web-useable forms, and analyse the network structure of interlinked connections.

The ABS (Australian Bureau of Statistics) has been exploring the intersection of these with its prototype GLIDE (Graphically Linked Information Discovery Environment) system, which accesses data stored in a semantic web format to create interactive web-based visualisations and support exploration and analysis of the linked data. This has been applied to two major semantically linked datasets – employer-employee data and freight data – to solve practical business problems.

GLEE (Graphically Linked Employer-Employee) data links survey and administrative data for firms and employees across several years. This has allowed study of the interconnections between employers and employees across time, including in specific geographic regions, which are easy to visualise within GLIDE. In particular, these connections can be used to assess firms which have entered/exited the economy in a given year, to determine whether these are genuine firm births/deaths, or the result of a merger/split/takeover. By considering where the employees of a firm worked in past and future years, as well as demographic information about the firm, a model can be built to distinguish births/deaths from other restructuring activities.

A joint project by ABS and BITRE (Bureau of Infrastructure, Transport and Regional Economics) aims to assess the technical feasibility of producing road freight statistics using Industry-sourced information (such as in-vehicle telematics and freight/logistics management systems) rather than traditional survey collection methods. Potential benefits include more detailed and up-to-date information and lower respondent burden. A number of freight industry operators have provided data

extracts for this project (either directly-sourced from freight transport providers or their telemetry service provider data streams) -- this data has been integrated along with information such as road network information and the location of key freight routes, truck stops and statistical geographic boundaries. A concept demonstrator has been developed using the GLIDE prototype -- preliminary assessment indicates that it is feasible to produce useful information for both industry and government, including for planning and policy decision-making. For example, the concept demonstrator can produce volumes and travel times along major corridors, and then drill down to identify congested freight network locations. It can also visualise the infrastructure used by freight vehicles, such as key freight routes and truck stops.

#### **BOSNIA AND HERZEGOVINA**

Reporting: Edin Šabanović

## 2013 Census of Population, Households and Dwellings: Final Results Disseminated

On 30<sup>th</sup> June 2016 the Agency for Statistics of Bosnia and Herzegovina released main results of the 2013 Census of Population, Households and Dwellings. According this results, in 2013 in Bosnia and Herzegovina lived 3,530,152 inhabitants within 1,155,736 households. The total number of dwellings registered in Census is 1,607,998. By the end of 2016 all census thematic books will be published, each of them will be related to specific domain of interest (demography, labour statistics, education, agriculture, dwellings, etc.). After more than two decades, data users have obtained a plenty of statistical data on socio-economic and demographic characteristics of the population.

## Post-enumeration Survey: A Measure of the Quality of the 2013 Census of Population, Households and Dwellings

In October of 2016 the Agency for Statistics of Bosnia and Herzegovina disseminated the results of the Post-enumeration Survey, as a measure of the quality of the 2013 Census of Population, Households and Dwellings. The quality of the Census was evaluated in terms of the coverage errors, as well as in terms of content errors. The net coverage error rate is -4.61%, while the majority of content quality indicators (net difference rate, index of inconstency, aggregate index of inconsistenca and rate of agreement) shown low content errors. Eurostat and International Monitoring Organisation have evaluated 2013 Census of Population, Households and Dwellings in Bosnia and Herzegovina as a statistical survey of good quality, which can be used for policy making, as well as for statistical purposes.

For more information, contact Edin Šabanović (<a href="mailto:edin.sabanovic@bhas.ba">edin.sabanovic@bhas.ba</a>), Sector for Statistical Methodology, Standard, Planning, Quality and Coordination, Agency for Statistics of Bosnia and Herzegovina.

#### **CANADA**

**Reporting: Normand Laniel** 

#### The 2016 Canadian Census of Population: Collection Results

The Canadian Census of Population collects demographic information from 3 in 4 households, via a short form, and demographic, social, economic and dwelling information from 1 in 4 households, via a long form, on persons residing in Canada. The Census is Canada's largest and most comprehensive data source for small areas.

The data collection operations for the 2016 Canadian Census of Population were successfully conducted from May to July 2016. Starting May 2, a bilingual letter was delivered to 82% of Canadian dwellings. This letter provided the required information so respondents could complete the questionnaire online. The letter also contained a toll-free number respondents could call to request a paper questionnaire. For the remaining dwellings (roughly 18%), questionnaires were either dropped off by enumerators (16%) or the questionnaires were completed by having enumerators conduct personal interviews (2%). For most dwellings (98% of the population), respondents were asked to complete the questionnaire for themselves and for members of their household online, or complete and then return a paper questionnaire in the mail. A toll-free number was also provided if a respondent wished to complete the questionnaire over the telephone. Also, reminder letters with or without questionnaires as well as in-person follow-ups were used for households that did not respond after the first delivery.

The willing participation of Canadians has enabled Statistics Canada to achieve an overall collection response rate of 98.4 per cent, this is higher than the rate obtained both the 2011 and 2006 Census cycles. Canadians' response to the 2016 long form was also outstanding. In 2016, the collection response rate for the long form was 97.8 per cent, the best ever recorded. The 2016 Census also stood out in two other regards: self-response and Internet response. Almost 88.8 percent of Canadian households completed their long or short form questionnaire without any assistance from Statistics Canada staff. This makes the Canadian Census the most efficient among traditional censuses conducted in the world. As for the Internet response, Canadians delivered a gold medal performance with an Internet response rate of 68.3 per cent, surpassing the ambitious initial objective of 65 per cent and setting yet again another world record. High rates of self-response and Internet response as well as overall response contribute to both the efficiency of data collection and exceptional data quality.

Statistics Canada will now turn its focus to providing high quality Census information. An accelerated release schedule is planned to provide Canadians with results in a more timely fashion than for any previous census. First results from the 2016 Census will be published on February 8, 2017, with the release of data on population and dwelling counts.

For more information, contact Normand Laniel, Assistant Director, Social Survey Methods Division, Statistics Canada at Normand.Laniel@Canada.ca.

#### **ISRAEL**

**Reporting: Tom Caplan** 

#### Israel Central Bureau of Statistics (ICBS)

We are pleased to report that there have been many developments in the statistics area in Israel since the last country report. We report here on some of them. The importance of statistics as essential to government decision making as well as in research and aspects of business and practically all aspects of life is recognized around the world. To that end the National Statistician was invited and appeared before the Government of Israel and he presented for 80 minutes highlights of the new statistics collected and produced in the last year. This presentation, the first of its kind, was well-received by the Ministers of the Government.

Preparations are underway for two pilot Population Census studies, one in 2017 and one in 2018. The two pilot studies will contribute significantly to the decision as to whether Israel will have a one year census or a rolling integrated census over five years. Another related major purpose to the pilot studies is to study the rate of success in the use of the internet as a preferred mode of response.

There are changes this year to the Household Expenditure Survey. Historically, the Survey has been based on families completing a two weeks diary. A decision has been made to change to a one week diary after careful experiments have shown that there will be no reduction in the quality of the data.

The Israeli Longitudinal Survey of Households is now getting ready for the field work for the fifth wave. For the first time, planning is underway to offer small financial incentives to the respondents.

October 2016 saw the launching of the second Twinning program-with Statistics Denmark. Following the success of the first Twinning program which began in 2013, this new Twinning Program has four components: 1- developing and enhancing Quality Management of Official Statistics in Israel; 2- developing procedures for enhancing access to micro data by researchers; 3- further development of infrastructure for agricultural statistics (specifically a farm register); 4- developing and adopting methodological and geo-spatial tools for improving the quality and efficiency of field surveys. The Twinning Program is funded by the European Union. In April 2015, the Government of Israel adopted a resolution obligating the Central Bureau of Statistics to publish well-being, sustainability, and national resilience indicators in the following domains: quality of employment; personal security; health; housing and infrastructure; education; higher education and skills; personal and social well-being; environment; civic engagement and governance; and material standard of living. It was also decided to develop two additional domains: information technology; and leisure, culture, and community. This defines a total of 11 well-being domains. For each domain 8 representative indicators were selected. As part of the implementation of the Government resolution, the ICBS will publish these indicators annually as of 2016.

We reported previously that the ICBS has launched a Bureau-wide Meta Data program. We adopted the SIMS (Single Integrated Meta Data Structure) as a standard and developed an electronic system for inputting Meta data for every statistical topic in the ICBS. We have now carried out a successful pilot applying the SIMS via the electronic system to 30 topics in Education Statistics. Based on what was learned from the pilot we now enter into a short phase of revising our system and then we will carry on implementing the program in other statistical programs.

The Israel Housing Price index is a very important statistic in Israel as it is in other countries. To that end, the Public Council on Statistics together with the ICBS have appointed an advisory committee, made up of academic specialists and experts from the public and private sectors. The task of the advisory committee is to examine all the issues related to computing the Housing Price Index.

These are some of the developments since the last report. There are considerably more and we will report on them in future reports.

#### **NEW ZEALAND**

**Reporting: Felibel Zabala** 

#### Developments in price measures new measures of inflation published

Statistics NZ has begun to publish new measures of inflation for groups of households – the household living-costs price indexes (HLPIs). These provide fresh insights into the experience of inflation for groups in our society (beneficiaries, superannuitants, income and expenditure quintiles, and Māori – the indigenous people of Aotearoa New Zealand). The series are designed to measure changes in the purchasing power of money, using a 'payment' conceptual approach. In contrast, the New Zealand CPI uses the 'acquisition' approach, which makes it best suited as a macroeconomic indicator for monetary policy.

HLPIs use 'democratic' expenditure weights to better reflect the inflation experience of a typical household in each household group. We believe this makes us the first country to adopt democratic weighting for official inflation measurement – our reasons and empirical analysis were tabled at the UNECE/ILO Meeting of the Group of Experts on CPIs.

See Household-group inflation: methods to combine expenditure patterns.

For more information see <u>Household living-costs price indexes: Background or contact Alan Bentley</u> or <u>info@stats.govt.nz</u>.

#### Big data used for measuring price change

Over the last few years, at Statistics NZ we've been very active in developing methods and processes for measuring price change from big data – such as scanner and online data.

After being early adopters of 'hedonics' for used cars in 2001, we developed and used similar model-based approaches to measuring price change in production for:

- mobile phones and TVs, in the overseas trade index from import data
- consumer electronics products (including TVs, computers, digital cameras), in the consumers price index <u>from scanner data.</u>

A big focus of our work in this area has been to develop and theoretically justify a regression-based approach called <u>the FEWS index.</u> This will produce non-revisable, quality-adjusted indexes even when there is no explicit information on product characteristics.

We recently signed up to purchase a year's worth of daily web-scraped online price data from US-based company <a href="PriceStats">PriceStats</a>, the commercial counterpart of <a href="MIT's Billion Prices Project">MIT's Billion Prices Project</a>. This data captures, in real-time, online prices from a wide range of different New Zealand retailers. It will enable us to do detailed research on the potential for enhancing and improving our current data collections and price measurement. Combining the online data with expenditure information from surveys or scanner data presents a rich opportunity for more-frequent and timely price indicators than are currently available.

Contact Frances Krsinich or info@stats.govt.nz for more information.

#### Integrated Data Infrastructure has a new spine

Statistics NZ's Integrated Data Infrastructure (IDI) has published the IDI prototype spine's creation and coverage, a paper explaining the creation of one specific dataset, called the 'spine', to which all other datasets link. Before 2015, the spine was created from only one source – Inland Revenue data. This paper describes the process to create the new spine – the IDI prototype spine.

For further information on our IDI, please contact <a href="mailto:Aaron Beck">Aaron Beck</a> or <a href="mailto:info@stats.govt.nz">info@stats.govt.nz</a>.

#### Coverage measurement aligns to a modernised census

The 2018 New Zealand Census of Population and Dwellings will use a modernised census model. We aim to reduce costs by mailing out internet response codes to 80 percent of potential respondents, to maximise online and self-response. However, the new model poses increased and different challenges for coverage, particularly for dwellings. The level of misclassification of dwellings (dwelling/non-dwelling, private/non-private dwelling) is expected to increase, which could lead to overcoverage of dwellings.

Previously, we optimised coverage measurement for assessing people coverage. This did not extend to dwelling over-coverage. To assess dwelling coverage in 2018, we need to adapt our existing coverage measurement and estimation methods, and

explore new methods. Avenues we are exploring include to:

- extend our coverage field operations to achieve better-quality dwelling enumeration
- over-sample the census non-contact sector to better target potential misclassification
- introduce a targeted dwelling-matching process
- investigate estimation methodologies adapted for over-coverage including an adapted dual system estimation approach in which a single coverage survey is used to adjust for both over- and under-coverage.

See <u>Census transformation in New Zealand</u> or contact <u>Abby Morgan</u> or info@stats.govt.nz for more details.

#### **UKRAINE**

The Baltic-Nordic-Ukrainian Summer School on Survey Statistics was held on August 22 – 26, 2016, in Kyiv, Ukraine.

This Summer School was organized by the Baltic-Nordic-Ukrainian Network on Survey Statistics (<a href="http://wiki.helsinki.fi/display/BNU/Home">http://wiki.helsinki.fi/display/BNU/Home</a>). The aim of the Network is long-term co-operation on education and research in survey statistics. Efforts are made to strengthen contacts between university teachers, students, researchers and practitioners. The Network also tries to promote interest among students and to motivate them to choose their professional career as survey statistician. The Network has grown continuously since 1996 to include partner universities in Belarus, Estonia, Finland, Latvia, Lithuania, Sweden and Ukraine. The Network activity was initiated, expended and guided during long time by Professor Gunnar Kulldorff (University of Umeå, Sweden) – the President of the International Statistical Institute in 1989–1991. We regret that he passed away in 2015 and miss the presence and energy of this outstanding person.

The aim of the Summer School was to promote interest and skills in the theory, methodology and practice of survey statistics, and to provide an opportunity for university teachers, research students and survey practitioners to share their knowledge and ideas, as well as to learn from experiences in other countries. The School started with the welcoming speeches by the Chair of Programme Committee Prof. Risto Lehtonen (University of Helsinki), Co-chair of Organizing Committee Prof. Yuliya Mishura (Taras Shevchenko National University of Kyiv) and special guest Vadym Pishcheiko, Advisor to Head of the State Statistics Service of Ukraine. The Programme Committee invited four keynote speakers. Jelke Bethlehem (University of Leiden) gave a series of lectures on "Estimation problems in web surveys", "Nonresponse problems in surveys: detection and correction" and "Use and abuse of graphs". Vassili Levenko (Statistics Estonia) told about "Preparing to conduct register-based population and housing census in Estonia in 2020". Kaija Ruotsalainen (Statistics Finland) presented "Register-based population census"

methodology in Finland", and Imbi Traat (University of Tartu) gave a lecture on "Sampling and estimation in surveys". There were also nine invited speakers, who delivered special lectures covering different topics of the theory and application of survey statistics and provided practical knowledge of working with R-packages: Danutė Krapavickaitė (Vilnius Gediminas Technical University), Alexander Kukush (Taras Shevchenko National University of Kyiv), Seppo Laaksonen (University of Helsinki), Thomas Laitila (University of Örebro), Risto Lehtonen (University of Helsinki), Natalja Lepik (University of Tartu), Mārtiņš Liberts (Statistics Latvia), Juris Breidaks (Statistics Latvia), Mykola Sydorov (Taras Shevchenko National University of Kyiv). There were 34 participants at the Summer School. Some of them presented contributed papers. All presentations were followed by discussions. The programme also contained a session with round-table discussion on challenges in teaching survey statistics in the era of big data.

We are very grateful to the International Association of Survey Statisticians for financial support, which made this event possible. The Summer School was extremely useful and important for all participants, especially for students from Ukraine.

For more information please visit the website of the Baltic-Nordic-Ukrainian Summer School on Survey Statistics at <a href="http://probability.univ.kiev.ua/school16">http://probability.univ.kiev.ua/school16</a>, or contact Olga Vasylyk @univ.kiev.ua.



# Upcoming Conferences and Workshops

The list below highlights events that have sessions or main subject related to areas such as survey methods, official statistics, data linkage and confidentiality. For a more wide-ranging list, please check the ISI Calendar of Events at <a href="https://www.isi-web.org/index.php/activities/calendar">https://www.isi-web.org/index.php/activities/calendar</a>



New Techniques and Technologies for Statistics (NTTS)

**Organized by:** Eurostat **Where:** Brussels, Belgium **When:** March 14-16, 2017

Homepage: <a href="https://ec.europa.eu/eurostat/cros/content/ntts-2017\_en">https://ec.europa.eu/eurostat/cros/content/ntts-2017\_en</a>

### **Inference from Non-probability Samples Conference of SRM, ESRA and ELIPSS**

Organized by: the European Survey Research Association (ESRA) and the Étude

Longitudinal par Internet Pour les Sciences Sociales (ELIPSS)

Where: Paris, France When: March 16 - 17, 2017

Homepage: https://ojs.ub.uni-konstanz.de/srm/article/view/6792/6113



Cross-cultural Survey Design & Implementation

Organized by: GESIS - Leibniz Institute for the Social Sciences and the Mannheim

Centre for European Social Research (MZES)

Where: Mannheim, Germany When: March 16 - 18, 2017

Homepage: https://www.csdiworkshop.org/







#### **International Conference on Official Statistics ICOS2017**

Organized by: School of Economics and Business of the University of Sarajevo and the

Institute for Statistics of the Federation of Bosnia and Herzegovina

Where: Sarajevo, Bosnia and Herzegovina

When: March 30 - 31, 2017

Homepage: http://www.icos2017.fzs.ba/



**AAPOR 72nd Annual Conference** 

Organized by: American Association for Public Opinion Research

Where: New Orleans, United States

When: May 18 - 21, 2017

Homepage: https://www.aapor.org/Conference-Events/Annual-Meeting.aspx



2017 Conference honoring Jon Rao on the occasion of his 80th birthday Organized by: School of Mathematics and Statistics at Yunnan University

Where: Kunming, China When: May 24 - 27, 2017

Homepage: http://www.raoconference2017.com/



Statistical Society of Canada 2017 ANNUAL MEETING

Organized by: University of Manitoba

Where: Winnipeg,

When: June 11 - 14, 2017

Homepage: https://ssc.ca/en/meeting/annual/2017

#### International Total Survey Error Workshop 2017

hosted by the Institute for Employment Research (IAB)

End: 14 Jun 2017



#### 2017 International Total Survey Error Workshop (ITSEW)

**Organized by:** Institute for Employment Research (IAB)

Where: Nuremberg, Germany When: June 12 - 14, 2017

Homepage: <a href="https://www.xing-events.com/itsew2017.html">https://www.xing-events.com/itsew2017.html</a>





#### 5th Italian Conference on Survey Methodology (ITACOSM 2017)

Organized by: Survey Sampling Group (S2G) of the Italian Statistical Society (SIS)

Where: Bologna, Italy When: June 14 - 16, 2017

Homepage: <a href="https://events.unibo.it/itacosm2017">https://events.unibo.it/itacosm2017</a>



#### ISI Satellite meeting on Small Area Estimation (SAE 2017)

Organized by: Ensai (Ecole Nationale de la Statistique et de l'Analyse de l'Information),

the CREST (Centre de Recherche en Economie et Statistique) and

the ILB (Institute Louis Bachelier)

Where: Paris, France When: July 10 - 12, 2017

Homepage: <a href="http://sae2017.ensai.fr/">http://sae2017.ensai.fr/</a>



#### 61st World Statistics Congress of the International Statistical Institute

Organized by: The High Commission for Planning – HCP

Where: Marrakech, Morocco When: July 16 – 21, 2017

Homepage: <a href="http://www.isi2017.org">http://www.isi2017.org</a>





The 7th Conference of the European Survey Research Association (ESRA)

Organized by: the University of Lisbon, jointly by the School of Economics and

Management (ISEG), the Centre for Research in Social Sciences and

Management (CSG) and the Institute for Social Sciences (ICS)

Where: Lisbon, Portugal When: July 17 – 21, 2017

Homepage: http://ec.europa.eu/eurostat/cros/content/ntts-2017



The Joint Statistical Meetings (JSM) 2017

Organized by: Several statistical societies and associations

Where: Baltimore, United States When: July 29 - August 3, 2017

Homepage: <a href="http://ww2.amstat.org/meetings/jsm/2017/">http://ww2.amstat.org/meetings/jsm/2017/</a>



### BALTIC-NORDIC-UKRAINIAN NETWORK ON SURVEY STATISTICS

Baltic-Nordic-Ukrainian Network Workshop on Survey Statistics 2017

**Organized by:** Baltic-Nordic-Ukrainian Network

Where: Vilnius, Lithuania When: August 21 - 25, 2017

Homepage: <a href="https://wiki.helsinki.fi/display/BNU/Events">https://wiki.helsinki.fi/display/BNU/Events</a>



Created and last modified by Boris Lorenc on 02 Nov, 2016

#### European Establishment Statistics Workshop (EESW) 2017

**Organized by:** European Network for Better Establishment Statistics (ENBES)

Where: Southampton, UK

When: August 30 - September 1, 2017

Homepage: <a href="http://www1.unece.org/stat/platform/display/ENBES/EESW17">http://www1.unece.org/stat/platform/display/ENBES/EESW17</a>



**Royal Statistical Society 2017 International Conference** 

Organized by: The Royal Statistical Society

Where: Glasgow, United Kingdom When: September 4 - 7, 2017

Homepage: http://www.rss.org.uk/conference2017



11th International Conference on Transport Survey Methods

Organized by: Interuniversitary Research Centre on Entreprise Networks, Logistics and

Transport (CIRRELT)

Where: Quebec, Canada

When: September 24 - 29, 2017

Homepage: http://www.hksts.org/isctsc.htm

# New Challenges for Statistical Software - The Use of R in Official Statistics - | The 5th International Conference

#### The Use of R in Official Statistics

Organized by: National Institute of Statistics of Romania, Bucharest University of

Economic Studies, Universiti Teknologi Mara, Ecological University of Bucharest, The University of Bucharest and Nicolae Titulescu University

of Bucharest

Where: Bucharest, Romania When: November 6 - 7, 2017

Homepage: http://r-project.ro/conference2017/



#### In Other Journals

### Journal of Survey Statistics and Methodology

#### Volume 4, Number 4 (December 2016)

https://jssam.oxfordjournals.org/content/current

#### **Survey Statistics**

The Pseudo-EBLUP Estimator for a Weighted Average with an Application to the Canadian Survey of Employment, Payrolls and Hours

Susana Rubin-Bleuer, Leon Jang, and Serge Godbout

#### Imputation Under Informative Sampling

Emily Berg, Jae-Kwang Kim, and Chris Skinner

#### Survey Methodology

#### Reflections on Surveys' Past and Future

Eleanor Singer

### Sampling, Data Collection, and Weighting Procedures for Address-Based Sample Surveys

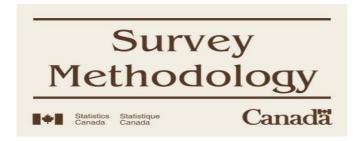
Michael P. Battaglia, Don A. Dillman, Martin R. Frankel, Rachel Harter, Trent D. Buskirk, Cameron B. McPhee, Jill M. DeMatteis, and Tracey Yancey

### **Estimation of Mode Effects in the Health and Retirement Study Using Measurement Models**

Alexandru Cernat, Mick P. Couper, and Mary Beth Ofstedal

#### Use of Replicates to Improve the Quality of Physical Measurements: Lessons Learned from a Psychometric Evaluation of an Adaptive Protocol in the National Children's Study

Michele F. Zimowski, Jack Moye, Melissa Heim Viox, Hildie Cohen, Bernard Dugoni, and Krishna Winfrey



#### Volume 42, Number 2 (December 2016)

http://www.statcan.gc.ca/pub/12-001-x/12-001-x2016002-eng.htm

#### Tests for evaluating nonresponse bias in surveys

Sharon L. Lohr, Minsun K. Riddles and David Morganstein

### Reducing the response imbalance: Is the accuracy of the survey estimates improved?

Carl-Erik Särndal, Kaur Lumiste and Imbi Traat

### Statistical inference based on judgment post-stratified samples in finite population

Omer Ozturk

### Adaptive rectangular sampling: An easy, incomplete, neighbourhood-free adaptive cluster sampling design

Bardia Panahbehagh

#### Unequal probability inverse sampling

Yves Tillé

#### A cautionary note on Clark Winsorization

Mary H. Mulry, Broderick E. Oliver, Stephen J. Kaputa and Katherine J. Thompson

#### **Short notes**

A few remarks on a small example by Jean-Claude Deville regarding nonignorable non-response

Yves Tillé

#### A note on the concept of invariance in two-phase sampling designs

Jean-François Beaumont and David Haziza

#### Corrigendum

#### Statistical matching using fractional imputation

Jae Kwang Kim, Emily Berg and Taesung Park



#### **Journal of Official Statistics**

#### Volume 32, Issue 4 (December 2016)

https://www.degruyter.com/view/j/jos.2016.32.issue-4/issue-files/jos.2016.32.issue-4.xml

JOS Special Section on The Role of Official Statistics in Statistical Capacity Building – Editorial

Ograjenšek, Irena

The Continuing Evolution of Official Statistics: Some Challenges and Opportunities

MacFeely, Steve

Helping Raise the Official Statistics Capability of Government Employees Forbes, Sharleen / Keegan, Alan

Statistical Capacity Building of Official Statisticians in Practice: Case of the Consumer Price Index

Deutsch. Tomi

Data-Mining Opportunities for Small and Medium Enterprises with Official Statistics in the UK

Coleman, Shirley Y.

From Quality to Information Quality in Official Statistics

Kenett, Ron S. / Shmueli, Galit

The Use of Official Statistics in Self-Selection Bias Modeling Dalla Valle, Luciana

Invited Commentary Special Section: The Role of Official Statistics in Statistical Capacity Building

Pullinger, John

Invited Commentary Special Section: Addressing the Needs of Official Statistics Users: The Case of Eurostat

De Smedt, Marleen

Measuring and Detecting Errors in Occupational Coding: an Analysis of SHARE Data

Belloni, Michele / Brugiavini, Agar / Meschi, Elena / Tijdens, Kea

Demographic Projections: User and Producer Experiences of Adopting a Stochastic Approach

Dunstan, Kim / Ball, Christopher

Small-Area Estimation with Zero-Inflated Data – a Simulation Study Krieg, Sabine / Boonstra, Harm Jan / Smeets, Marc

Dead or Alive? Dealing with Unknown Eligibility in Longitudinal Surveys Watson, Nicole

**Book Review** 

Web Survey Methodology

Herzing, Jessica M. E.

Improving Survey Methods: Lessons from Recent Research Olson, Kristen



**Volume 9, Number 5 (2016)** 

http://www.surveypractice.org/index.php/SurveyPractice/index

Experimenting with the Addressee Line in a Mail Survey of Hispanic Households Paul Lavrakas, Gerry Dirkz, Lisa Lusskin, Beth Ponce

Comparison of Lottery Incentive Amount on Click-Through Rates and Online Survey Response Rates in a Member Association Population

Ariel A. Finno, Simon Neubauer

Availability of Respondent Provided Financial Records to Supplement Self-Reports of Household Expenditures

Emily Geisen, Charles Lau, M. Christopher Stringer, Jennifer Edgar, Brandon Kopp, Ashley Richards

The Impact of Item Sequence Order on Local Item Dependence: An Item Response Theory Perspective

Kenneth D. Royal

The relationship between linkage refusal and selected health conditions of survey respondents

Judith Weissman, Jennifer D Parker, Donna M Miller, Eric A Miller, Renee M Gindi

#### Income Interpolation from Categories Using a Percentile-Constrained Inverse-CDF Approach

George Lance Couzens, Marcus Berzofsk, Kimberly Peterson

Practice-based considerations for using multi-stage survey design to reach special populations on Amazon's Mechanical Turk

Victoria Anne Springer, Peter J. Martini, Samuel C. Lindsey, I. Stephanie Vezich



#### **Survey Research Methods**

Volume 10, Number 3 (2016) https://ojs.ub.uni-konstanz.de/srm/index

#### **Table of Contents**

Sensitive Questions in Online Surveys: An Experimental Evaluation of Different Implementations of the Randomized Response Technique and the Crosswise Model Marc Höglinger, Ben Jann, Andreas Diekmann

Can we assess representativeness of cross-national surveys using the education variable?

Verena Ortmanns, Silke L. Schneider

Are Final Comments in Web Survey Panels Associated with Next-Wave Attrition? Cynthia McLauchlan, Matthias Schonlau

Evaluating the immediate and longer term impact of a refusal conversion strategy in a large scale longitudinal study

Lisa Calderwood, Ian Plewis, Sosthenes Ketende, Tarek Mostafa

**Longitudinal Wealth Data and Multiple Imputation** 

Christian Westermeier, Markus M. Grabka

Reliability and Stability of the Standard Fear of Crime Indicator in a National Panel Over 14 Years

Rainer Schnell, Marcel Noack



### Statistical Journal of the IAOS: Journal of the International Association for Official Statistics

Volume 32, Number 4 (2016)

http://content.iospress.com/journals/statistical-journal-of-the-iaos/32/4

#### **Editorial**

West, Kirsten

#### **Interview with Gemma Van Halderen**

Condon, Katherine M.

#### **Conversation with Monica Dashen**

Condon, Katherine M.

### Editorial Special Section on Integration of Statistical and Geospatial Information *Van Halderen, Gemma*

### Integrating statistical and geospatial information, cultures and professions: International developments and Australian experience

Van Halderen, Gemma; Minchin, Stuart; Brady, Martin; Scott, Greg

### jSTAT MAP: A new geostatistics web service for small area census data and its impact

Makita, Naoki

#### Geospatial information management in Europe - responding to the user needs

Haldorson, Marie; Zaccheddu, Pier-Giorgio; Fohgrub, Bianka; Petri, Ekkehard

#### Data integration to determine vulnerability to climate change

Bueno, Maria do Carmo Dias; D'Antona, Álvaro de Oliveira

# Spatial statistics on the web: A joint project between Finland's National Statistical Institute and National Mapping Agency to improve usability of spatial statistics

Tammilehto-Luode, Marja

### Integrated geography and statistical expertise at INEGI Ocampo, Rolando

### Integrating statistical and geospatial information in New Zealand *Morgan, Rochelle*

### Regional statistical data for Germany, Europe and the world *Schnorr-Baecker*, *Susanne*

Estimation from contaminated multi-source data based on latent class models Guarnera, Ugo; Varriale, Roberta

#### Statistical internationalism: From Quetelet's census uniformity to Kish's crossnational sample survey comparability

Michalopoulou, Catherine

### Risk management and the release of microdata: Balancing disclosure risks and data utility

Whiteley, Sonia; Skuja, Vitalis

#### **Enumeration via Internet - Estonian experience**

Beltadze, Diana

#### Analysing correspondence between administrative and survey data

Delden, Arnout van; Pannekoek, Jeroen; Banning, Reinder; Boer, Arjen de

### Improvements on the statistical quality practices at the Brazilian institute of geography and statistics

Zacharias, Maria Luiza Barcellos; Bianchini, Zélia Magalhães

#### **CORE:** A concrete implementation of the CSPA architecture

Bruno, Mauro; Duma, Rolando; Scannapieco, Monica; Silipo, Marco; Vaste, Giulia

#### Towards a market strategy for the official statistics industry

Rolland, Asle

### Ranked set sampling: As a cost-effective and more efficient data collection method

Sinha, Arun Kumar

### Big data and semantic technology: A future for data integration, exploration and visualisation

Harwood, Andrew; Mayer, Andreas

#### Web surveys for offline rural communities

Gichohi, Beatrice W.

#### Survey methodology for humanitarian intervention

Chu, Dung; Dashen, Monica; Selva, Marcia; Suchowski, Maria

#### Case study: Burmese healthy mothers and infants

Chu, Dung; Dashen, Monica; Selva, Marcia; Suchowski, Maria

### How to build capacity in statistics for providers of official statistics: A case study from Turkey

Gündoğan, Ünal; Gücümengil, Alper; Gedik, Gürdal

### Documenting the statistical editing process with process tables, a case study of the Luxembourg Financial Accounts

Roymans, Ingber

### The role of the conditional independence assumption in statistically matching income and consumption

Donatiello, Gabriella; D'Orazio, Marcello; Frattarola, Doriana; Rizzi, Antony; Scanu, Mauro; Spaziani, Mattia

### A user-friendly framework for metadata and microdata documentation based on international standards and the PCBS experience

Zeidan, Haitham; Greenwell, Geoffrey

#### The added value of micro-integration: Data on laid-off employees

van Rooijen, Johan; Bloemendal, Caroline; Krol, Nynke

### Deriving household indebtedness indicators by linking micro and macro balance sheet data

Kavonius, Ilja Kristian; Honkkila, Juha

### Non-extensive entropy econometrics and CES production models: Country case study

Bwanakare, Second

### IPUMS International: A review and future prospects of a unique global statistical cooperation programme

MacDonald, Alphonse L.

### International Group for Indigenous Health Measurement: Recommendations for best practice for estimation of Indigenous mortality

Coleman, Clare; Elias, Brenda; Lee, Vanessa; Smylie, Janet, Waldon, John; Hodge, Felicia Schanche; Ring, Ian

#### A conversation about implicit bias

Golbeck, Amanda L.; Ash, Arlene; Gray, Mary, Gumpertz, Marcia; Jewell, Nicholas P.; Kettenring, Jon R.; Singer, Judith D.; Gel, Yulia R.

Statistics as instruments for prosperous, transparent and democratic societies *Malaguerra*, *Carlo*; *MacDonald*, *Alphonse L*.



#### VOL 84, ISSUE 3 (December 2016)

http://onlinelibrary.wiley.com/doi/10.1111/insr.v84.1/issuetoc

### Posterior Inference in Bayesian Quantile Regression with Asymmetric Laplace Likelihood

Yunwen Yang, Huixia Judy Wang and Xuming He

### Discussions on Posterior Inference in Bayesian Quantile Regression with Asymmetric Laplace Likelihood

#### Discussion

Roger Koenker

#### **Discussion**

James P. Hobert and Kshitij Khare

#### **Discussion**

Lan Wang and Ben Sherwood

#### **Discussion**

Luke B. Smith

#### Discussion: Should a Working Model Actually Work?

Xiao-Li Meng

### Rejoinder to Posterior Inference in Bayesian Quantile Regression with Asymmetric Laplace Likelihood

#### Rejoinder

Yunwen Yang, Huixia Judy Wang and Xuming He

#### **Original Articles**

#### Statistical Inference, Learning and Models in Big Data

Beate Franke, Jean-François Plante, Ribana Roscher, En-shiun Annie Lee, Cathal Smyth, Armin Hatefi, Fuqi Chen, Einat Gil, Alexander Schwing, Alessandro Selvitella, Michael M. Hoffman, Roger Grosse, Dieter Hendricks and Nancy Reid

### Adversarial and Amiable Inference in Medical Diagnosis, Reliability and Survival Analysis

Nozer D. Singpurwalla, Barry C. Arnold, Joseph L. Gastwirth, Anna S. Gordon and Hon Keung Tony Ng

#### A Review on Modal Clustering

Giovanna Menardi

#### On Goodness of Fit for Operational Risk

Andrey Feuerverger

### **Lognormal Distributions and Geometric Averages of Symmetric Positive Definite Matrices**

Armin Schwartzman

#### H-likelihood Predictive Intervals for Unobservables

Youngjo Lee and Gwangsu Kim

#### Skysurveys, Light Curves and Statistical Challenges

G. Jogesh Babu and Ashish Mahabal

#### Implications of the Data Revolution for Statistics Education

Jim Ridgway

#### **Book Reviews**

### Estimation and Testing under Sparsity Sara van de Geer Springer, 2016 D.R. Cox

### Ordered Regression Models: Parallel, Partial, and Non-Parallel Alternatives Andrew S. Fullerton, Jun Xu

John H. Maindonald

### Handbook of Health Survey Methods Timothy P. Johnson Editor John Wiley and Sons, Inc., 2015

Steven G. Heeringa

### Generalised Principal Component Analysis Rene Vidal, Yi Ma, S. Shankar Sastry Springer, 2016

Lili Zhao

### **Branching Process Models of Cancer Richard Durrett Springer International Publishing, 2015**

Alexander Tsodikov

### **Exposure-Response Modelling: Methods and Practical Implementation Jixian Wang Chapman & Hall**

Reijo Sund

# TRANSACTIONS ON DATA PRIVACY

Foundations and Technologies <a href="http://www.tdp.cat">http://www.tdp.cat</a>

#### Volume 9, Issue 3, April 2016

http://www.tdp.cat/issues16/vol09n03.php

#### **TABLE OF CONTENTS**

DBMask: Fine-Grained Access Control on Encrypted Relational Databases Muhammad I Sarfraz, Mohamed Nabeel, Jianneng Cao, Elisa Bertino

Network Structure Release under Differential Privacy Hiep H. Nguyen, Abdessamad Imine, Michaël Rusinowitch

# Journal of Privacy and Confidentiality

Current Issue: Volume 7 (2015-2017), Issue 2 (2017) Special Issue on the Theory and Practice of Differential Privacy http://repository.cmu.edu/jpc/vol7/iss2/

#### **TABLE OF CONTENTS**

**Special Issue on the Theory and Practice of Differential Privacy** *Marco Gaboardi and Chris J. Skinner* 

#### Between Pure and Approximate Differential Privacy

Thomas Steinke and Jonathan Ullman

### Gradual Release of Sensitive Data under Differential Privacy Fragkiskos Koufogiannis, Shuo Han and George J. Pappas

**Dual Query: Practical private Query Release for High Dimensional Data** *Marco Gaboardi, Emilio Jesus Gallego Arias, Justin Hsu, Aaron Roth and Zhiwei Steven Wu* 

AnoA: A Framework for Analyzing Anonymous Communications Protocols

Michael Backes, Aniket Kate, Praveen Manoharan, Sebastian Meiser, and Esfandiar Mohammadi

#### **Heterogeneous Differential Privacy**

Mohammad Alaggan, Sébastien Gambs, and Anne-Marie Kermarrec

#### **Featherweight PINQ**

Hamid Ebadi and David Sands



#### Vol 180 Issue 1 (January 2017)

http://onlinelibrary.wiley.com/doi/10.1111/rssa.2017.180.issue-1/issuetoc

New statistics for old?—measuring the wellbeing of the UK

Paul Allin and David J. Hand

Taylor's power law and the statistical modelling of infectious disease surveillance data

Doyo Gragn Enki, Angela Noufaily, Paddy Farrington, Paul Garthwaite, Nick Andrews and Andre Charlett

Zero-inflated modelling for characterizing coverage errors of extracts from the US Census Bureau's *Master Address File* 

Derek S. Young, Andrew M. Raim and Nancy R. Johnson

Respondent-driven sampling bias induced by community structure and response rates in social networks

Luis E. C. Rocha, Anna E. Thorson, Renaud Lambiotte and Fredrik Liljeros

A space-time multivariate Bayesian model to analyse road traffic accidents by severity

Areti Boulieri, Silvia Liverani, Kees de Hoogh and Marta Blangiardo

Assessing the inequality of lifetime healthcare expenditures: a nearest neighbour resampling approach

Albert Wong, Hendriek Boshuizen, Johan Polder and José António Ferreira

Estimating the density of ethnic minorities and aged people in Berlin: multivariate kernel density estimation applied to sensitive georeferenced administrative data protected via measurement error

Marcus Groß, Ulrich Rendtel, Timo Schmid, Sebastian Schmon and Nikos Tzavidis

Combining Labour Force Survey data to estimate migration flows: the case of migration from Poland to the UK

Arkadiusz Wiśniowski

Adaptive and responsive survey designs: a review and assessment Roger Tourangeau, J. Michael Brick, Sharon Lohr and Jane Li

Unpacking the determinants of life satisfaction: a survey experiment Viola Angelini, Marco Bertoni and Luca Corazzini

Identifying subgroups of enhanced predictive accuracy from longitudinal biomarker data by using tree-based approaches: applications to fetal growth Jared C. Foster, Danping Liu, Paul S. Albert and Aiyi Liu

Multilevel structural equation models for longitudinal data where predictors are measured more frequently than outcomes: an application to the effects of stress on the cognitive function of nurses

Fiona Steele, Paul Clarke, George Leckie, Julia Allan and Derek Johnston

Florence Nightingale, William Farr and competing risks Jan Beyersmann and Christine Schrade

Stochastic block models for multiplex networks: an application to a multilevel network of researchers

Pierre Barbillon, Sophie Donnet, Emmanuel Lazega and Avner Bar-Hen

Exploring complete school effectiveness via quantile value added Garritt L. Page, Ernesto San Martín, Javiera Orellana and Jorge González



#### Journal of the American Statistical Association



Volume 111, Issue 516 (December 2016)
http://www.tandfonline.com/toc/uasa20/111/516?nav=tocList

**Applications and Case Studies** 

Improving and Evaluating Topic Models and Other Models of Text Edoardo M. Airoldi & Jonathan M. Bischof

#### Comments:

Matt Taddy Aleksandrina Goeva & Eric D. Kolaczyk David M. Blei

#### **Article**

Model Comparison and Assessment for Single Particle Tracking in Biological Fluids Martin Lysy, Natesh S. Pillai, David B. Hill, M. Gregory Forest, John W. R. Mellnik, Paula A. Vasquez & Scott A. McKinley

Hierarchical Feature Selection Incorporating Known and Novel Biological Information: Identifying Genomic Features Related to Prostate Cancer Recurrence Yize Zhao, Matthias Chung, Brent A. Johnson, Carlos S. Moreno & Qi Long

### Modeling the Evolution of Dynamic Brain Processes during an Associative Learning Experiment

Mark Fiecas & Hernando Ombao

### **Bayesian Methods for Nonignorable Dropout in Joint Models in Smoking Cessation Studies**

J. T. Gaskins, M. J. Daniels & B. H. Marcus

### Multiple Imputation of Missing Categorical and Continuous Values via Bayesian Mixture Models with Local Dependence

Jared S. Murray & Jerome P. Reiter

#### Low SNR in Diffusion MRI Models

Jörg Polzehl & Karsten Tabelow

#### **Localizing Temperature Risk**

Wolfgang Karl Härdle, Brenda López Cabrera, Ostap Okhrin & Weining Wang

#### **Theory and Methods**

#### **Comments**

Tianxi Cai & Lu Tian Jun Fan & Ming Yuan Alexander R. Luedtke & Mark J. van der Laan Elizabeth L. Ogburn Min Qian Michael Rosenblum

#### Rejoinder

Guanhua Chen, Donglin Zeng & Michael R. Kosorok

#### Smoothing Parameter and Model Selection for General Smooth Models

Simon N. Wood, Natalya Pya & Benjamin Säfken

#### Comment

Thomas Kneib Sonja Greven & Fabian Scheipl

#### Rejoinder

Simon N. Wood, Natalya Pya & Benjamin Säfken

#### Fast Bayesian Factor Analysis via Automatic Rotations to Sparsity

Veronika Ročková & Edward I. George

#### **Modeling Probability Forecasts via Information Diversity**

Ville A. Satopää, Robin Pemantle & Lyle H. Ungar

### Inference for Monotone Functions under Short- and Long-Range Dependence: Confidence Intervals and New Universal Limits

Pramita Bagchi, Moulinath Banerjee & Stilian A. Stoev

### Robust Improper Maximum Likelihood: Tuning, Computation, and a Comparison with Other Methods for Robust Gaussian Clustering

Pietro Coretto & Christian Hennig

#### **Theory and Methods**

#### A Bayesian Approach to Graphical Record Linkage and Deduplication

Rebecca C. Steorts, Rob Hall & Stephen E. Fienberg

#### Identifiability of Normal and Normal Mixture Models with Nonignorable Missing Data

Wang Miao, Peng Ding & Zhi Geng

#### **Testing the Predictor Effect on a Functional Response**

Valentin Patilea. César Sánchez-Sellero & Matthieu Saumard

#### On SURE-Type Double Shrinkage Estimation

Bing-Yi Jing, Zhouping Li, Guangming Pan & Wang Zhou

#### **Extremal Depth for Functional Data and Applications**

Naveen N. Narisetty & Vijayan N. Nair

#### Statistical Matching Analysis for Complex Survey Data with Applications

Pier Luigi Conti, Daniela Marella & Mauro Scanu

#### **Multitask Quantile Regression Under the Transnormal Model**

Jianqing Fan, Lingzhou Xue & Hui Zou

#### A Link Between the E-Value and the Robustness of Block Designs

J. D. Godolphin

#### Sparse Approximate Inference for Spatio-Temporal Point Process Models

Botond Cseke, Andrew Zammit-Mangion, Tom Heskes & Guido Sanguinetti

#### **Composite Robust Estimators for Linear Mixed Models**

Claudio Agostinelli & Víctor J. Yohai

### Optimal Model Averaging Estimation for Generalized Linear Models and Generalized Linear Mixed-Effects Models

Xinyu Zhang, Dalei Yu, Guohua Zou & Hua Liang

Bayesian Nonparametric Modeling of Higher Order Markov Chains Abhra Sarkar & David B. Dunson

Panel Data Models With Interactive Fixed Effects and Multiple Structural Breaks Degui Li, Junhui Qian & Liangjun Su

Sensitivity Analysis for Multiple Comparisons in Matched Observational Studies Through Quadratically Constrained Linear Programming Colin B. Fogarty & Dylan S. Small

Matching and Weighting With Functions of Error-Prone Covariates for Causal Inference

J. R. Lockwood & Daniel F. McCaffrey

### **BIOMETRIKA**

Volume 103, Issue 4 (December 2016)

https://academic.oup.com/biomet/issue

#### **Articles**

Some pioneers of modern statistical theory: a personal reflection  $\it D. R. Cox$ 

Replicates in high dimensions, with applications to latent variable graphical models

Kean Ming Tan; Yang Ning; Daniela M. Witten; Han Liu

Approximating fragmented functional data by segments of Markov chains A. Delaigle; P. Hall

Misclassified group-tested current status data L. C. Petito; N. P. Jewell

Optimal stratification in outcome prediction using baseline information Florence H. Yong; Lu Tian; Sheng Yu; Tianxi Cai; L. J. Wei

On inverse probability-weighted estimators in the presence of interference L. Liu; M. G. Hudgens; S. Becker-Dreps

#### In-sample forecasting with local linear survival densities

M. Hiabu; E. Mammen; M. D. Martinez-Miranda; J. P. Nielsen

#### Model selection with nonignorable nonresponse

Fang Fang; Jun Shao

#### Combining eigenvalues and variation of eigenvectors for order determination

Wei Luo; Bing Li

#### High-dimensional and banded vector autoregressions

Shaojun Guo; Yazhen Wang; Qiwei Yao

#### Nonnested model comparisons for time series

T. S. McElroy

#### Multivariate spatial covariance models: a conditional approach

Noel Cressie; Andrew Zammit-Mangion

#### The cylindrical K- function and Poisson line cluster point processes

Jesper Møller; Farzaneh Safavimanesh; Jakob Gulddahl Rasmussen

#### Default Bayesian analysis with global-local shrinkage priors

Anindya Bhadra; Jyotishka Datta; Nicholas G. Polson; Brandon Willard

#### Bayesian inference on quasi-sparse count data

Jyotishka Datta; David B. Dunson

#### Miscellanea

### Fast sampling with Gaussian scale mixture priors in high-dimensional regression

Anirban Bhattacharya; Antik Chakraborty; Bani K. Mallick

#### A conditional Lindley paradox in Bayesian linear models

Agniva Som; Christopher M Hans; Steven N MacEachern



### Welcome New Members!



We are very pleased to welcome the following new members!

Title	First name	Surname	Country
MR.	Ramiro Flores	Cruz	Argentina
MS	Isabela	Bertolini Coelho	Brazil
PROF	Alan Ricardo	da Silva	Brazil
MR.	Dramane	Bako	Burkina Faso
MR.	Abdoul Daouda Aziz	Ilboudo	Burkina Faso
MR.	Arouna	Sow	Burkina Faso
MR.	Qiao	Во	China
MS	Xiaoxiao	Chen	China
MS	Yanhong	Hao	China
MR.	Lin	Jia	China
MS	Yu	Jing	China
MS	Xia	Jing	China
MR.	Zheng	Yuebo	China
MR.	Andreas	Gieshoff	Germany
MS	Lucy Twumwaah	Afriyie	Ghana
MR.	Solomon	Owusu-Bempah	Ghana
MRS	Sigalit	Mazeh	Israel
DR.	Andrius	Ciginas	Lithuania
MR.	Tomas	Rudys	Lithuania
MS	Maria	Nanyemba	Namibia
MS	Frieda	Theofelus	Namibia
MR.	Surajudeen Oladimej	Abolade	Nigeria
MR.	Thanyani	Maremba	South Africa
DR.	Allen	Kabagenyi	Uganda
MS	Angela	Kiconco	Uganda United
MRS	Heather	Schroeder	States
MRS	María Eugenia	Riaño	Uruguay
MR.	Owen	Siyoto	Zambia

#### IASS Officers and Council Members

#### **Executive Officers**

President: Steve Heeringa (USA) sheering@isr.umich.edu **President-elect:** Peter Lynn (UK) Plynn@essex.ac.uk

Vice-Presidents: Monica Pratesi (Italy) Monica.pratesi@unipi.it

Imbi Traat (Estonia) Imbi.traat@ut.ee

**Scientific Secretary:** Denise Silva (Brazil) denisebritz@gmail.com

**Council Members** Michael Brick (USA) mikebrick@westat.com (2013-2017)

daniela.cocchi@unibo.it Daniela Cocchi (Italy) Jack Gambino (Canada) jack.gambino@Canada.ca Risto Lehtonen (Finland) risto.lehtonen@helsinki.fi Ralf Münnich (Germany) muennich@uni-trier.de jopsomer@stat.colostate.edu Jean Opsomer (USA)

(2015-2019)Hukum Chandra (India) Hchandra12@gmail.com

> Giovanna.ranalli@stat.unipg.it Maria Govanna Ranalli (Italy)

Timo Schmid (Germany) Timo.schmid@gmx.org David Steel (Australia) dstell@uow.edu.au Ineke Stoop (The Netherlands) i.stoop@scp.nl

Niko Tzavidis (UK) n.tzavidis@soton.ac.uk

**Chair of the Morocco IASS** 

**Program Committee 2017** 

Marcel Vieira (Brazil) Marcel.vieira@ice.ufjf.br

Chair of the committee for

the Cochran Hansen prize

award

risto.lehtonen@helsinki.fi Risto Lehtonen (Finland)

#### **Ex Officio Members**

ISI Associate Director,

**Operation Manager / IASS** s.mehta@cbs.nl Shabani Mehta

Liaison

**Transition Executive** 

**Director:** Catherine Meunier (France) katherine.meunier@orange.fr

Director, Executive/Treasurer Ada van Krimpen an.vankrimpen@cbs.nl

of the ISI (The Netherlands) **Finance Manager:** Michael Leeuwe

**IASS Webmaster:** Mehmood Asghar and Olivier Dupriez

**ISI Membership Officer** Margaret de Ruiter-Molloy

(The Netherlands)

odupriez@worldbank.org m.deruitermolloy@cbs.nl



## Institutional Members



#### **2 International Organisations**

AFRISTAT EUROSTAT

#### 23 Bureaus of Statistics

AUSTRALIA – AUSTRALIAN BUREAU OF STATISTICS

BRAZIL – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE)

CANADA - STATISTICS CANADA

**DENMARK** – STATISTICS DENMARK

FINLAND - STATISTICS FINLAND

**GERMANY** -STATISTICHES

**BUNDESAMT** 

ISRAEL- INTERNATIONAL REL. &

STATISTICAL COORDINATION

ITALY –INSTITUTO NAZIONALE DI STATISTICSA (ISTAT)

KOREA, REPUBLIC OF - STATISTICS KOREA

**LUXEMBOURG** – EUROPEAN COMMISSION – EUROSTAT

MACAO, SAR China - DIREÇCAO DOS SERVIÇOS DE

ESTATISTICA E CENSOS

MALI - AFRISTAT

**MAURITIUS** – STATISTICS MAURITIUS

MEXICO – INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI)

MEXICO - NUMÉRIKA-MEDICION Y ANALISIS ESTAD. AVANZADO, SC

**NEW ZEALAND - STATISTICS NEW ZEALAND** 

NORWAY - STATISTICS NORWAY

PORTUGAL - INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

**SWEDEN** - STATISTICS SWEDEN

**UNITED STATES - NATIONAL AGRICULTURE** 

STATISTICS SERVICE

UNITED STATES - NATIONAL CENTER FOR HEALTH

STATISTICS

United States – Research Triangle Institute
United States - Survey Research Center

**UNITED STATES** – WESTAT INC.

### INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

#### **CHANGE OF ADDRESS FORM**



If your home or business address has changed, please copy, complete, and mail this form to:

IASS Secretariat Membership Officer Margaret de Ruiter-Molloy International Statistical Institute P.O. Box 24070, 2490 AB The Hague, The Netherlands

	Name: Mr./Mrs./Miss/Ms.		First
name:			
E-mail address (nlease i	ust indicate one):		
	address on the IASS web sit		
	daress on the IA33 web sit	er	
Yes No			
Home address			
Street:			
City:			
Country:			
Fax number:			
Business address			
Company:			
Street:			
City:			
	extension:		
Please specify address to	to which your IASS corresp	ondence should be sent:	
Home Business		chaches chould be sent.	
nome Business	• 🗆		

# Read The Survey Statistician online!



http://isi-iass.org/home/services/thesurvey-statistician/