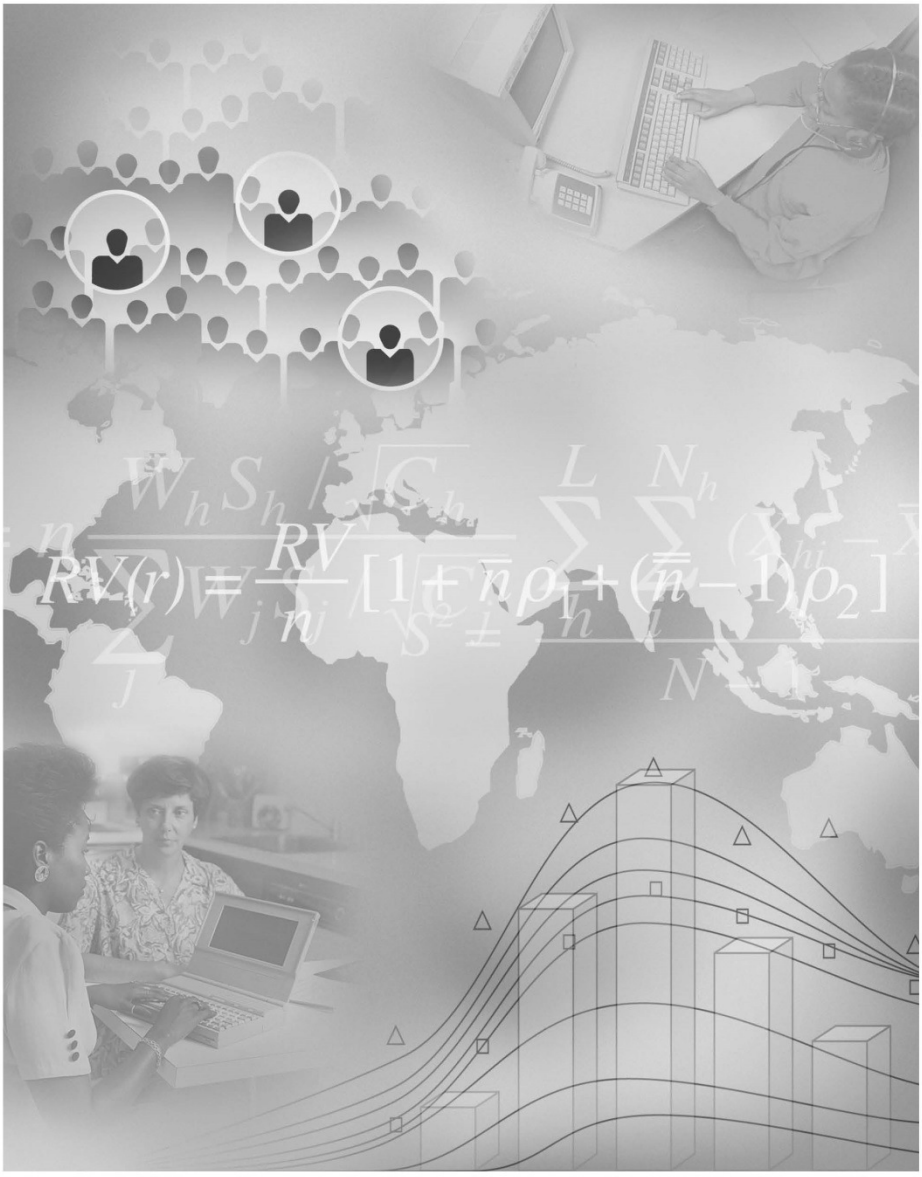
the Survey Statistician

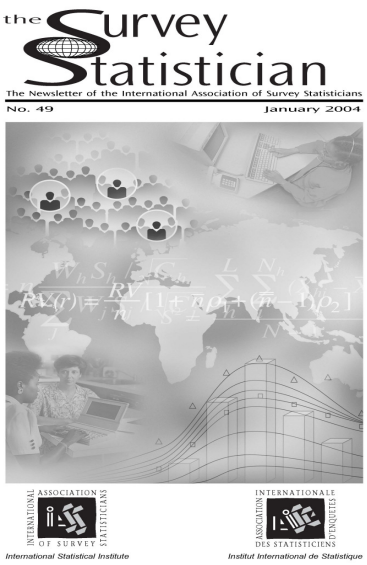
The Newsletter of the International Association of Survey Statisticians

No. 73 **January 2016**



International Statistical Institute

Institut International de Statistique



Editors
Natalie Shlomo and Eric Rancourt

Section Editors
Pierre Lavallée - Country Reports
Ken Copeland - Ask the Experts
Denise Silva - New and Emerging Methods
Natalie Shlomo - Book and Software Review

Production
Lori Young - Composition
Courtney Williamson - Circulation
Olivier Dupriez - Web site

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@statcan.gc.ca).

Letter from the Editors

The January 2016 issue contains articles of interest and important information regarding upcoming conferences, journal contents, updates from the IASS Executive and more. We hope you enjoy this issue, and would be happy to receive your feedback and comments on how we can make improvements.

In the *New and Emerging Methods* Section (edited by the Scientific Secretary Denise Silva), Piero Demetrio Falsori and Paolo Righi from ISTAT, Italy, have contributed an article titled: A flexible tool for defining optimal sampling designs. In the article, the authors proposes a flexible tool for defining the optimal inclusion probabilities in various survey contexts characterized by the need to disseminate survey estimates of prefixed accuracy, for a multiplicity of both variables and domains of interest.

In the *Ask the Experts* Section (edited by Ken Copeland), Craig McLaren and Kate Davies, from the ONS, United Kingdom, have provided a response to the question: Why do official surveys sometimes tell a different story from alternative sources? The authors provide case studies from the United Kingdom and summarize in the discussion: 'Different methodologies result in different characteristics to the outputs, but what is key is that the resulting indicators and outputs are interpreted appropriately'.

For the *Book and Software Review* Section, Natalie Shlomo from the University of Manchester contributes a review of the recently published book: Methodological Developments in Data Linkage, edited by Katie Harron, Harvey Goldstein and Chris Dibben, Wiley Series in Probability and Statistics, Chichester: John Wiley & Sons, 2016. The section concludes with some insight on future areas of research related to data linkage.

Please let Denise Silva (denise.silva@ibge.gov.br) know 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 (copeland-kennon@norc.org). If you are interested in writing a book or software review, please get in touch with Natalie Shlomo (Natalie.Shlomo@manchester.ac.uk).

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.Lavallee@Canada.ca) for his continuing efforts to obtain timely reports from the different countries. We ask all country representatives to please

share information on your country's current activities, applications, research and developments in survey methods.

In the newly added section of the newsletter *Contributions from Members,* we have included an article written by Zdenek Patak from Statistics Canada titled: Some thoughts on seasonal adjustment variances. If you would like to contribute brief articles or editorials to the newsletter for this new section, please send them directly to the editors of the newsletter, Eric Rancourt and Natalie Shlomo.

This issue of *The Survey Statistician* includes the first letter and updates from our IASS President, Steven Heeringa. In his letter, the President's warmly thanks his predecessor, Danny Pfeffermann as well as members of the IASS council who completed their term in 2015. The letter includes an overview of the recent accomplishments and recent activities of the association and focuses on the importance of promoting education, in particular through the next World Statistics Congress (WSC) and the various conferences supported by the IASS. This issue also features the first letter from our Scientific Secretary Denise Silva which includes the list of recently sponsored conferences and the ones to come in 2016.

In the *News and Announcement* section we have an article in honor of Dr. Mike Hidiroglou, who recently retired from Statistics Canada after a very productive 40-year career where he influenced two generations of methodologists. We wish Mike all the best on his well-deserved retirement and this new chapter in his life.

We also 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 http://isi.cbs.nl/iass/allUK.htm.

Eric Rancourt Eric.Rancourt@Canada.ca

Natalie Shlomo Natalie.Shlomo@manchester.ac.uk

Letter from the President

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS (IASS)

Dear IASS colleagues,

This is the first opportunity in my new role as IASS President to write to you on the status of our Association as well as ongoing activities and new initiatives that we are collectively undertaking in support of our international community of survey statisticians, methodologists and practitioners. In the interest of keeping your attention, I won't go into depth on all the ongoing activities and established resources of the IASS. Much of this is covered elsewhere in this issue of *The Survey Statistician* including the Report of our Scientific Secretary (Denise Silva) and the reports of the Country Representatives. I will highlight a few important "business" items and as well as key activities and initiatives that I believe are important to the maintenance and growth of our association.

Before turning to current business, I would like to say thank you to Danny Pfeffermann, 2013-2015 IASS President, for his leadership over the past two years and commitment to strengthening the global mission and reach of our association. The knowledge, experience and charisma that Danny brought to this role will be impossible to replicate but I can commit to carry forward the initiatives begun by Danny and Ray Chambers before him in the areas of communication, membership development, service to members and international statistical capacity building. I would also like to acknowledge the service of the IASS members who completed either their four year term as elected members of the IASS Council in 2015 (Christine Bycroft, Ka-Lin Chan, Olivier Dupriez, Natalie Shlomo, Marcel Viera, and Avaro Villabos) or a two year term as an officer of the IASS (Mick Couper, Jairo Arrow and Geoff Lee). Although their term in office has ended, many of these individuals remain active in service to IASS as editors, committee chairs and webmasters.

Since the origin of IASS in the 1970s, a major focus of our efforts in promoting education, networking and intellectual exchange within the international survey community has centered on the biennial World Statistics Congress (WSC) meetings. The 2017 World Statistics Congress will be held July 16-21, 2017 in Marrakech, Morocco and the IASS will again contribute in a major way to the Short Course Program, Paper Sessions and other elements of the Scientific Program at these meetings. Hopefully you have all received the call for Invited Paper Session (IPS) proposals either directly from the ISI Permanent Office or through the e-mail communication from our Scientific Secretary, Denise Silva. (A schedule of key dates for the WSC 2017 is included in a table at the end of my letter.) More information about the meetings and regular updates as the meeting dates draw closer can be found at the official web site, http://www.isi2017.org/. Marcel Viera (marcel.viera@ice.ufjf.br) is serving as the IASS representative to the WSC 2017 Scientific Program Committee (SPC). I am sure that Marcel would be happy to answer any questions that IASS members have concerning session proposals or the scientific program in general.

Although the start of the WSC 2017 is over 18 months away, planning activities will soon begin for the Short Course Program. ISI President, Pedro Silva (one of our own!), is currently in the process of forming the WSC 2017 Short Course Committee and shortly after that a call for course proposals will be issued. We will see that all members receive notification of this call through the IASS mail group that is securely maintained by the ISI permanent office. The IASS and its membership has a strong track record of organizing and teaching short courses in conjunction with the WSC meetings and I trust that many of you will again submit proposals and volunteer to teach a course on important topics in survey statistics and methodology.

Prior to the WSC 2017, the IASS 2017 Cochran-Hansen Prize (http://isi-iass.org/home/cochran-hansen-prize/) will again be awarded to a young statistician from a developing or transition country. This award is based on a paper competition that will refereed by a committee of IASS members appointed by the IASS officers. To be eligible for the competition, young statisticians will need to complete and submit their paper to the committee early in 2017. If you are an academic advisor, professional supervisor or mentor to a young statistician that you believe should submit their written work to this competition, I encourage you start planning the submission now. The Cochran-Hansen Prize not only enables the recipient to participate in the WSC and present their paper in a special session but also provides the young statistician with international recognition early in what we all hope will be a promising career in survey research.

The past two decades have produced a dramatic increase in the numbers of conferences, symposia, workshops and training activities devoted to survey methodology and survey statistics. Outside the biennial meetings of the World Statistics Congress, the IASS has rarely played the primary role as organizer of a major conference devoted to survey methodology and statistics. In lieu of mounting an effort to organize its own major conference, IASS has chosen each year to provide modest financial support to approximately 4 to 5 regional and international meetings that are specifically focused on survey statistics and survey methodology. Thanks to a policy instituted by Danny Pfeffermann when he was IASS President, whenever IASS contributes support to one of these meetings, IASS members can expect a discount on conference registration fees and organizers also agree to make IASS promotional material available to conference attendees. 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 (denise.silva@ibge.gov.br) or to me (sheering@umich.edu). The Report of the IASS Scientific Secretary included in this issue of TSS identifies the 2016 conferences and meetings that the officers have selected to receive IASS support. In addition, this issue also includes announcements for many more national, regional and international conferences that are primarily focused on survey methods or will include major sessions of relevance to those of us who work in the survey field.

I will mention that the IASS Committee that is charged with drafting a Strategic Plan has not gone dormant for the Northern hemisphere winter but our work has been delayed several months against the schedule that we published this past summer. You should expect to see a draft of the Strategic Plan for your review and comment during April of 2016.

I could go on at length about the many activities and initiatives of our Association or philosophize on the opportunities and challenges facing our profession but I have already broken my promise to keep this short and will stop now.

In closing, I wish you an enlightened, productive and healthy start to 2016.

Yours in the science and practice of survey research,

Steve Heeringa
IASS President

Report from the Scientific Secretary

The past semester was marked by the 60th World Statistics Congress in Rio de Janeiro. As reported by Mick Couper in July 2015, IASS organized or co-organized 13 invited paper sessions (for details see The Survey Statistician 72 from July 2015) and IASS members taught several one and two-day short courses. The IASS sponsored the two winners of the Cochran-Hansen Prize to present their papers at the WSC2015. The conference proceedings are now available at the conference website http://www.isi2015.org/.

IASS has also co-sponsored a WSC satellite meeting on Small Area Estimation in Santiago, Chile (see www.encuestas.uc.cl/sae2015), as well as the following conferences:

- ✓ The Fourth European Establishment Statistics Workshop- EESW15 hosted by Poznań University of Economics and Statistical Office in Poznań in Poland on 7-9 September 2015 (http://enbes.wikispaces.com/EESW15);
- ✓ The 4th Italian Conference on Survey Methodology ITACOSM 2015 – hosted by the University of Rome, Italy, on 24-26 June of 2015 (http://itacosm15.sta.uniroma1.it/node/65), and
- ✓ The 4th Baltic-Nordic Conference on Survey Statistics - BaNoCoSS2015 **-** held in Helsinki, Finland, on 25-28 August, 2015 (https://wiki.helsinki.fi/display/banocoss2015/4th+Baltic-Nordic+Conference+on+Survey+Statistics).

For this year, IASS is already committed to support:

- ✓ The Fifth International Conference on Establishment Surveys - ICES V – that will take place in Geneva, Switzerland, on 20-23 June 2016 (http://www.portal-stat.admin.ch/ices5/);
- ✓ The Baltic-Nordic-Ukrainian (BNU) Network on Survey Statistics will take place on 22-26 August 2016 in Kyiv, Ukraine. (http://wiki.helsinki.fi/display/BNU/Events), and
- ✓ The 9th French Colloquium in Survey Sampling to be held in Quebec, Canada on 11-14 October 2016 (http://sondages2016.sfds.asso.fr/).

Preparations for the 61th World Statistics Congress have already started. The 61st World Statistics Congress (WSC) of the ISI will be held from 16 to 21 July 2017 in Marrakech, Morocco (http://www.isi2017.org/). The Scientific Program Committee (SPC) for the 2017 WSC has already published the call for Invited Paper Session (IPS) proposals. Marcel Vieira (marcel.vieira@ice.ufjf.br) is our representative on the SPC.

Invited Paper Sessions constitute great opportunities to place survey statistics subjects in the spotlight of the conference and IASS counts on your ideas and inputs to make it happen. IASS needs your support to continue its relevant presence on the WSC agenda. The proposals should be submitted until 15 February 2016 using the form for IPS proposals available on the WSC website. When submitting a proposal, you should indicate that your proposed IPS is related to our Association then your proposal will be first reviewed by IASS under the guidance of our representative on the SPC. If you are considering submitting a proposal for an IPS, please notify Marcel Vieira and me so that we can track proposals from our membership before the deadline for submissions.

We are looking for your engagement and proposals regarding IASS activities and plans. Also, if you would like to indicate a subject or write an article to the *New and Emerging Methods* section of The Survey Statistician, please get in touch with me at denise.silva@ibge.gov.br or denisebritz@gmail.com.

Very best wishes,

Denise Silva

News and Announcements

61st IS World Statistics Congress 2017
Marrakech, Morocco

Invited Paper Sessions (IPS) and Special Topic Sessions (STS)

Call for Proposals

The Chair of the Scientific Programme Committee, Fabrizio Ruggeri, and the Chair of the Local Programme Committee, Mohamed Taamouti, invite the statistical community to present proposals for the Invited Paper Sessions (IPS) and Special Topic Sessions (STS).

Invited Paper Sessions

Invited Paper Sessions at the ISI World Statistics Congresses (WSC) serve to increase awareness about statistical research and to bring new research results to a broad audience. The 61st WSC of the ISI, to be held from 16 to 21 July 2017 in Marrakech, Morocco, will highlight the contributions that Statistics can make to the advancement of science and to human health and welfare across the globe. The WSC will host talks on a wide variety of topics, with the overall goal of presenting a balanced programme that provides a sense of the current state of the field. The WSC will feature state-of-the-art presentations on the various aspects of statistical work, including new theoretical findings in Probability and Statistics, advances in applied statistical methods and recent developments in the application of Statistics in areas of broad interest and importance.

The Scientific Programme Committee (SPC) for the 2017 WSC calls for proposals for Invited Paper Sessions (IPS). Invited Papers will be central to the success of the congress. The proposals should be submitted using the form for IPS proposals. The form will be available on the WSC website http://www.isi2017.org/index.php/isi2017/submission-system **starting 1 November 2015**. When submitting a proposal, you should indicate if your session could be related to one or more of the ISI Associations. If you specify any Associations, then

you will have to choose a primary one and your proposal will be first reviewed by that Association under the guidance of its representative on the SPC. For a list of SPC members and their emails, please visit http://isi2017.org/index.php/isi2017/committees/scientific-prog-committee.

Alternatively, if you think that your proposal needs to be considered outside of any ISI Association, please choose ISI as the primary association. Proposals approved by the Associations and the ISI will be reviewed by the SPC on a competitive basis. Upon submission, if the proposal is considered interesting but not accepted as an IPS, you will be asked if you agree to have your session considered as a Special Topic Session (see below).

To ensure full consideration, please submit your proposals by 15 February 2016. Each proposal should include a brief description and justification for the proposed session and a list of speakers and discussants who have agreed to participate. The selection criteria will take into account diversity, scientific quality and impact.
The ISI always tries to ensure that:

- The programme is of high quality, with sessions that emphasize novel ideas, approaches and/or applications to problems of importance;
- The sessions have a balance of organizers/speakers/discussants from around the world;
- The contents are clearly of interest to the ISI's and its Associations' members;
- The congress has a diversity of participants with respect to age, gender and specialization.

Special Topic Sessions

The 2017 WSC also has Special Topic Sessions (STS), to be selected by the Local Programme Committee (LPC). Proposals for STS can be submitted by individual members of the ISI and Associations, ISI Committees, or outside institutions and organisations. An STS usually consists of 4-5 papers and possibly a discussant invited by the organiser. The deadline for STS proposals is 1 August 2016, with submissions possible starting from 1 March 2016.

If you would like to know more about the STS, and/or are interested in arranging one, please contact the Local Programme Committee (LPC) via email: m.taamouti@bkam.ma. Competitive IPS Proposals that are not selected by the SPC due to limitation of timeslots or for other reasons may be considered by the LPC as STS proposals, provided this is agreeable to the organizers concerned.

Contributed Papers/Posters

Information about submitting Contributed Papers and Posters will be available on the ISI and WSC websites in the course of 2016. We anticipate that the submission period will be from 15 September 2016 to 1 February 2017.

For further information about the WSC, please visit http://www.isi2017.org/.

General questions about the scientific programme should be directed to Fabrizio Ruggeri at fabrizio@mi.imati.cnr.it. For questions about Special Topic Sessions, please contact Mohamed Taamouti at m.taamouti@bkam.ma.

For information about the ISI and its Associations, visit the ISI website.

Fabrizio Ruggeri
Chair of the Scientific Programme Committee

Mohamed Taamouti
Chair of the Local Programme Committee

A Fine Survey Statistician's Career: Michel A. Hidiroglou

J. Gambino and E. Rancourt

On August 31st, 2015, Dr. Michel A. Hidiroglou retired from Statistics Canada after an illustrious career. Dr. Hidiroglou obtained a PhD in statistics from Iowa State University in 1974 and spent the years since then making frequent and significant contributions to survey methodology, both at Statistics Canada and internationally. He has had a profound influence on the methods and software used in surveys, particularly business surveys. In addition, he has been a guide and mentor to many methodologists working on both research-oriented projects and on the practical application of innovative methods to surveys. He has also played an important international role in the advancement of statistical methods for surveys, both as an author and as the editor of the refereed journal *Survey Methodology*.

Dr. Hidiroglou, or "Mike" to his many friends, began his career at Statistics Canada in 1974. He rose quickly to the position of senior methodologist, and then chief and assistant director. In 2006, Mike became director of the new Statistical Research and

Innovation Division. A major highlight of Mike's career is his contribution, spanning several years, to the modernization of business surveys, including work on the Business Register, survey design, stratification, sampling and estimation. But this history hides Mike's many contributions beyond Statistics Canada. His expertise has been sought internationally, and he has met this demand by presenting numerous workshops on survey methods throughout the world, and also by sitting on advisory boards such as the prestigious U.S. Census Advisory Committee. Mike also worked at the Office for National Statistics in the U.K. for three years, heading the Survey Methods Division.

Mike's contributions to statistical methods for surveys have been many and varied. He has more than 50 papers published in refereed journals. While being sound mathematically, his research has always focused on tackling practical problems. The list of topics to which he contributed is too long to list here, but the highlights include Mike's contributions to stratification, sample allocation, double sampling, data editing, and dealing with outliers, small area estimation and estimation in general. His papers on these topics are often cited by statisticians throughout the world.

Mike was a central force in the development of generalized systems at Statistics Canada, particularly the Generalized Estimation System. The seeds for this work were sown when Mike was a PhD student in Iowa, where he developed an estimation system (PC-CARP) for complex surveys that was a breakthrough for the field. Unlike other estimation software at the time, Mike's system took into account the complex nature of surveys to produce correct estimates. At the other end of his career, Mike continues to contribute to the development of systems that use complex survey data to produce improved estimates. He has been the driving force behind the innovative Small Area Estimation system, now in prototype form, that will broaden the number and quality of estimates that can be produced by our surveys.

Throughout his career, Mike has devoted himself to pushing the boundaries of what can be accomplished in statistical surveys by being very active both on the practical and implementation side as well as on the theoretical and more mathematical side. In recognition of Mike's significant contributions to survey methodology, he was elected Fellow of the American Statistical Association in 1998.

Mike has been a highly-regarded manager of projects and people. In addition to managing a subdivision and then a division, Mike has headed the Methodology Research and Development Committee for many years, guiding the research done not only by his own staff but also by methodologists throughout Methodology Branch. In this role, Mike has set research priorities by successfully engaging subject-matter experts to ensure that their needs play an important role in decision making, another reflection of Mike's view that research should focus on practical problems. He often said that one should work on the problems rather than around them. That is why he liked to "push the math". Sometimes, after having been made aware of a statistical

issue in a survey, he would come back the next day with a full binder of derivations, with pages marked by the hours through the night!

Finally, Mike's influence on several generations of survey statisticians cannot be overstated, and this influence is not limited to researchers. It extends to the many people he supervised, both directly and indirectly, as well as to the very large number of his collaborators over the years. Throughout his career, young researchers have viewed Mike as a role model, particularly as someone who has successfully combined a career as a first-rate researcher with one as a manager of people and projects. Because of this, Mike's legacy goes beyond the practical results—new surveys, new methods, new systems—he has achieved over the decades. His legacy is also measured by the many people who were inspired by him and who, in turn, continue to contribute to the areas that Mike holds dear.

Ask the Experts

Why do official surveys sometimes tell a different story from alternative sources?

Craig McLaren and Kate Davies– United Kingdom, Office for National Statistics (ONS)
(Craig is currently Head of Methods and Systems for National Accounts within ONS and previously led the Retail Sales area and you can follow him on twitter@chmclaren; while Kate is currently Head of Construction Development within ONS and also previously headed Retail Sales and you can follow her on twitter@statskate)

1 Show me the numbers!

It seems that every minute of every day brings another set of data outputs for analysts and media organizations to pore over and generate their next set of headlines. Where these data come from and how they are estimated can often be missed in the rush to get out the next story as quickly as possible. When competing estimates are published from different sources, and those estimates are not consistent with each other, it can lead to questions around the methodology used, and ultimately the validity of the outputs.

This article considers the often asked question on why official estimates from a National Statistic Institution (NSI) can sometimes tell a different story to alternative data sources. We focus on a case study of Retail Sales within a United Kingdom context which we describe in more detail in the following sections.

2 Official statistics and alternative sources

It is important to note that outputs from NSIs and external surveys may not necessarily be directly comparable for a number of reasons, which can include differences in methodology of how the outputs are compiled, timeliness of when the outputs are available, the number of respondents, and the questions that are asked. For this reason, any comparison between NSI outputs and external survey data should be undertaken with care.

Alternative sources to official statistics will always exist and still have a role to play in analyzing available information. Examples of alternative sources are:

- Credit card companies (e.g. purchases via credit card can provide an indication of retail activity)
- Price data collected from supermarket scanners (e.g. used for alternative price indices)
- Business lobby groups (e.g. collecting information from their own membership which is then used to derive outputs based on their own members interests)

It should be acknowledged that alternative sources can help supplement official outputs for the following reasons:

- Timeliness – Official estimates can take time to compile due to the collection and processing of respondent information. Alternative sources often have fewer constraints on the compilation approach, which means the estimates can be published sooner. Consequently, alternative surveys can be valuable tools for providing an early indication of activity for the relevant industry or sector.

- Additional or alternative measures – External surveys can measure a variety of variables that are often not covered in official surveys or are not needed for any legislative requirement. For example, the collection of expectations or intentions to invest may be omitted from official surveys to minimize cost; but collected by alternative sources for different purposes.

3 Case study: Measuring Retail Sales in the United Kingdom

The Retail Sales Index (RSI) produced by ONS is a key economic indicator within the UK context, and is one of the mostly timed short-term measures of economic activity. It is usually produced within 2 to 3 weeks of the end of the reference period, e.g.

estimates for Retail Sales within the UK are based on sets of weekly periods, so estimates for the four weeks from 3 January to 30 January 2016 will be published 19 February 2016. The RSI is used to estimate consumer spending on retail goods and the output of the retail industry, both of which are used in the compilation of the UK National Accounts.

Within the UK, there are two prominent private industry surveys that can be compared with the ONS RSI outputs. These are the 'British Retail Consortium-KPMG United Kingdom Retail Sales Monitor' (BRC RSM) and the 'Confederation of British Industries Distributive Trades Survey' (CBI DTS). In the UK context, the BRC RSM publication is typically timed to be released a few days before the ONS RSI. This can lead to news articles which refer to both sets of outputs in a comparative way, particularly when they show different movements. For example, a recent headline quote was

- "Lies, dammed lies, and statistics: ONS' retail sales are, at best a partial truth and, at worse, misleading" was a July 2015 article written by Neil Saunders for the newspaper City A.M. in which he suggests that the ONS should publish an official figure that is more comparable with the alternative source from BRC.

In practice, we need to recognize that there are differences in the methodology for the compilation of estimates. As an example consider Table 1, which shows the sample size and total response of three different Retail Sales indicators within the UK. This shows that the external surveys have smaller sample sizes and fewer total responses than the official ONS data, although response rates from some external surveys are better, and BRC do not revise monthly data.

Table 1: Sample size differences between alternative Retail Sales indicators

	Sample size	Total response
ONS Retail Sales Index	95 per cent of total UK retail sales value (5000 companies)	93 per cent of total UK retail sales value (64 per cent response rate at the time of initial publication each month)
BRC-KPMG RSM	60 per cent of total UK retail sales value. See http://www.brc.org.uk/bis/default.asp?main_id=3	100 per cent response rate on a weekly basis across twelve product categories
CBI DTS	Between 850-900 companies, with an additional 800-900 approached every quarter. See http://www.cbi.org.uk/media/2182644/inside-cbi-business-surveys.pdf	A monthly average of 121 companies over 2013, of which 66 were in the retail sector

As well as differences in sample sizes and response rates, private sector surveys also employ different methodological approaches, which may provide further explanation of differences between the three series. The main methodological differences, although not an exhaustive list, between the RSI, RSM, and DTS are listed in Table 2. A more comprehensive list was produced by Palmer (2007).

A point that will be of interest to survey sample enthusiasts is the aspect of sample rotation. This is the practice of refreshing the sample each month to ensure that the coverage remains representative of the changing economy. Sample rotation also spreads out the burden of questionnaire filing on businesses. Within the ONS RSI sample survey, around 2500 businesses in the sample are designated as small firms and theoretically spend on average 15 months in the sample. A further 1600 are classed as medium sized firms, and theoretically spend an average of 27 months in the sample. In practice, the time actually spent in the sample can differ depending on an individual business circumstance. The remaining 900 businesses are the largest in Great Britain and are permanently in the RSI sample.

Table 2: Main methodological differences between alternative Retail Sales indicators

Criterion	ONS RSI	BRC RSM	CBI DTS
Sample review and regular update	Sample updated each month, in accordance with the profile of the British retail industry and updated UK wide business register	Sample related to BRC membership profile	Sample based on UK industry trade associations, CBI regional office contacts and publically available databases.
Enforcement	Statutory survey	Voluntary survey based on membership	Voluntary survey
Headline retail sales measures	Values and volumes, seasonally adjusted and non-seasonally adjusted	Changes in total, like-for-like and online sales values compared with a year ago, no seasonal adjustment	Balance statistic indicating the change in sales volume compared with a year ago, no seasonal adjustment
Weighting method	Survey data weighted by sales for each type of retailer	Survey data covering twelve product categories and six online categories weighted by household expenditure on each type of product	Survey data weighted dually, according to gross value added (GVA) of each firm's sector and employment size
Links: http://www.cbi.org.uk/media/2182644/inside-cbi-business-surveys.pdf. A balance statistic is the difference between the weighted percentage of companies responding that sales were "up" on a year ago, and those reporting that sales were "down".			

Users of outputs also need to consider which retail indicator to use to best understand changes in economic activity. For example, comparisons of 'like-for-like' sales are often cited for alternative sources, and are useful in assessing individual business performance. However, total sales as collected by ONS RSI provides a more appropriate measure of aggregate economic change as this reflects businesses openings and closures based on an updated business register, which themselves can give an indication of economic growth. The main question on the RSI questionnaire asks: "For the reporting period stated above, what was the value of the businesses retail turnover including VAT?" Additional questions consider the value of transactions which occurred using the Internet.

Where alternative surveys can add value is in the use of more subjective questions. For example, the CBI DTS survey offers a qualitative indication of the general direction of retail sales volumes compared with a year ago ("How do your sales and orders for this month and your expectations for next month compare with those in the same month a year ago? (a) Up (b) Same (c) Down (d) n/a"), while ONS official data are quantitative estimates that provide an accurate guide of the magnitude of movement.

Positive and negative replies to the CBI DTS are then expressed as percentages, and a weight is applied to each response (as outlined in the table above), to arrive at the total balance. Consequently, the final balance statistic may not always reflect the true magnitude of growth in retail sales, if, for example, a small proportion of the sample experience a marked change in volumes, but the remainder experience small changes in the opposing direction.

Research undertaken by Cunningham (1997) also suggests that balance statistics may not pick up small changes in output, as firms may choose to report the change in the volumes of sales as 'the same' for very small changes in output.

4 Discussion

What is important for one person isn't always the most important thing for someone else. This is why it's vital to offer different ways of looking at how things are changing in order to get the full picture, and having different indicators available facilitates this.

Different methodologies result in different characteristics to the outputs, but what is key is that the resulting indicators and outputs are interpreted appropriately. Some may suggest that it is better for only one set of headline figures to be presented, but different indicators will have the potential to be used in different ways, and by different types of user, including decision makers, analysts and the media.

Different user groups will always have different needs, which is why ONS presents its own estimates in more than one way so that the person using the statistics can choose the estimates and outputs that is right for them and their audience. When there is more than one indicator available to assess, the key is to ensure that the

official statistics compliment other available indicators and where they don't there are reasons available to explain why.

5 Acknowledgements

This article is based on previously published information articles from the Office for National Statistics, UK.

References:

Cunningham, A. (1997), Bank of England, Quarterly Bulletin, Vol. 3, Bank of England, London. Link: http://www.bankofengland.co.uk/archive/Documents/historicpubs/qb/1997/qb970306.pdf

Office for National Statistics (2014) Retail Sales Index (RSI) and external surveys. Link: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/retail-sales/comparability-of-rsi-sales-and-external-indicators-2012-13.pdf

Palmer, N. and Hynard J. (2007) Comparing the ONS's retail sales index with the BRC's retail sales monitor, Economic and Labour Market Review, Volume 1, Number 4, April 2007. Link: http://www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/no--4--april-2007/comparison-ons-s-retail-sales-index-with-the-brc-s-retail-sales-monitor.pdf

New and Emerging Methods

A Flexible Tool for Defining Optimal Sampling Designs

Piero Demetrio Falorsi
Paolo Righi

1. Introduction

The paper proposes a flexible tool for defining the optimal inclusion probabilities in various survey contexts characterized by the need to disseminate survey estimates of prefixed accuracy, for a multiplicity of both variables and domains of interest. The tool, mainly the algorithm for computing the inclusion probabilities, is suitable for the implementation of standard sampling designs such as stratified and PPS sampling designs, as well as non-conventional designs as incomplete stratified, controlled sampling or indirect sampling designs. Furthermore, the framework allows for dealing with more specific estimation problems (unit non-response, small area) in an integrated way.

The tool is useful for the sampling in official statistics surveys which commonly produce a large number of estimates relating to both different parameters of interest and highly detailed estimation domains or sub-populations. The parameter estimation is primarily performed defining a sampling strategy that is the couple sampling design and estimator. In particular, when the domain indicator variables are available for each sampling unit in the sampling frame, the survey sampling designer could attempt to select a sample in which the size for each domain is fixed. Thus, direct estimates can be obtained for each domain and sampling errors, at the domain level, would be controlled.

In a randomization approach to the inference, the design must envisage a random selection scheme, so that for each sample s of the reference population U, a selection probability is given. Customary selection schemes are one-stage or multi-stage sampling designs and simple (stratified or not) random sampling designs. When the probabilities are proportional to a given size measure, probability-proportional-to-size (PPS) sampling designs are implemented. Given the selection scheme, a sampling design (i.e., the selection probability distribution of the subpopulations of U) can be defined when the inclusion probabilities of the units are defined. For example, the Stratified Simple Random Sampling without replacement design (SSRS) is defined by the probabilities given by the ratio of the sample size to the population size at the stratum level, and by selecting a fixed number of units in each stratum (selection scheme).

The definition of the inclusion probabilities is a fundamental step when defining an efficient sampling strategy. This, given a sample size, must be aimed at keeping sampling errors to a minimum.

In the Stratified Simple Random Sampling without Replacement (SSRSWOR) designs, fixing the inclusion probabilities means defining the stratum's sample sizes; thus, the problem is denoted as a sample allocation (in the strata) or allocation problem. The literature on sampling has devoted much attention to sample allocation. When one target parameter is to be estimated for the overall population, the optimal allocation in stratified samplings can be performed (Cochran, 1977). In particular, the Neyman sample allocation minimizes the variance of the subject which depends upon a given budget (the cost-constraint optimization problem) or, reversing the problem, an allocation that minimizes costs can be performed, subject to a given sampling error threshold (the precision constraint optimization problem).

When more than one parameter and domain estimation must be estimated, the allocation is no longer optimal. If the strata are domains of interest, the Neyman allocation may cause certain strata to undergo great variation, due to the fact that the strata are not domains of estimation. The Neyman allocation is, hence, no longer optimal. A more suitable rough sample allocation may be the equal sample allocation, particularly efficient when estimating stratum (or specific domain) parameters. However, this may lead to a much greater variance of the estimator than that obtained from the Neyman allocation for the overall population. Hence, in several practical applications, a compromise allocation must be performed.

In multivariate cases, where more than one characteristic of each sampled unit must be measured, the optimal allocation for individual characteristics is of little practical use, unless the characteristics under study are highly correlated (an allocation that is optimal for one characteristic is generally far from optimal for others). The multidimensionality of the problem also leads to the definition of a compromise allocation method (Khan *et al*. 2010), with a loss of precision compared to the individual optimal allocations. The problem of defining compromise sampling designs has been addressed in some recent papers. Gonzalez and Eltinge (2010) present an interesting overview of the approaches for defining optimal sampling strategies. Several authors have discussed various criteria for obtaining a feasible compromise allocation as a solution to an optimization problem in which the concept of optimality differs from the cost-constraint or precision-constraint univariate and the uni-domain optimization problem (Kokan and Khan, 1967, Chromy, 1987; Bethel, 1989; Choudhry *et al*., 2012).

Falorsi and Righi (2015) propose a compromise allocation suitable for a wide range of sampling designs, offering, *de facto*, a generalized framework for dealing with the sample allocation.

In section 2 we describe the main elements characterizing the sampling framework. Section 3 develops some examples which show how the framework covers customary sampling designs. Section 4 illustrates the flexibility of the framework for dealing with issues such as nonresponse and the indirect sampling in the design phase (FAO, 2015). We refer to Falorsi and Righi (2008) and Falorsi *et al*. (2006) for further interesting topics, such as the small area problem which can be tackled using this framework. Section 5 wraps up the main features of the proposed framework.

2. The generalized sampling framework

The sample allocation procedure is undertaken to manage different concepts: (*i*) the target and planned domains, (*ii*) the superpopulation model for predicting the target variables, (*iii*) the expression of the variances of the estimates given by the selection scheme and (*iv*) the optimization problem to be solved.

2.1 Target and planned domains

Let *U* be the reference population of *N* and let $\boldsymbol{\pi}=(\pi_1,...,\pi_k,...,\pi_N)'$ be the *N*-vector of inclusion probabilities, being, π_k, and the probability of the *k-th* population unit. Denote by U_h (*h*=1, …, *H*) the subpopulation of size $N_h=\sum_{k\in U_h}\delta_{hk}$ where $\delta_{hk}=1$ if $k\in U_h$ and $\delta_{hk}=0$ otherwise. We focus on fixed size sampling designs which are those satisfying the condition $\sum_{k\in s}\boldsymbol{\delta}_k=\mathbf{n}$, (2.1.1)
where *s* is the selected sample, $\boldsymbol{\delta}_k=(\delta_{1k},...,\delta_{hk},...,\delta_{Hk})'$ and $\mathbf{n}=(n_1,...,n_h,...,n_H)'$ is the vector of integer numbers defining the sample sizes fixed at the design stage. Since the sample size n_h, corresponding to $U_h$does not vary among sample selections, the subpopulation U_h will be referred to as planned ***domain***. In our setting, the planned domains can overlap; therefore, the unit *k* may have more than one value $\delta_{hk}=1$ (for $h=1,...,H$). Let us suppose that the δ_{hk} values are known and available in the sampling frame for all the population units. We suppose that the *N*×*H* matrix $(\boldsymbol{\delta}_1,..,\boldsymbol{\delta}_k,...,\boldsymbol{\delta}_N)'$is non-singular.

Finally, let U_d (*d=1, …,D*) be an ***estimation*** or ***target*** *domain*, i.e. a generic sub-population of *U* with N_d elements, for which separate estimates must be calculated. Let us denote the domain membership indicator for unit *k by* γ_{dk} defined as $\gamma_{dk}=1$ if $k\in U_d$ and 0 otherwise. We assume that the γ_{dk} values are available in the sampling frame and more than one value γ_{dk} can be 1 for each unit *k*; therefore, the estimation domains can overlap. The relationship between the planned and target domain is exemplified in section 3.

2.2 Superpopulation model

Let y_{rk} denote the value of the *r*-th (*r =1, …, R*) variable of interest attached to the *k*-th population unit. The parameters of interest are the $D\times R$ domain totals

$$t_{(dr)}=\sum_{k\in U}y_{rk}\gamma_{dk} \qquad (r=1,\ldots, R;\ d=1,\ldots,D) \qquad (2.2.1)$$

and the related Horvitz-Thompson (HT) estimators are $\hat{t}_{(dr)}=\sum_{k\in s}y_{rk}\gamma_{dk}/\pi_k$.

Prior to sampling, the y_{rk} values are not known. It is, therefore necessary to either obtain some proxy values or to predict the y_{rk} values based on superpopulation models that exploit auxiliary information. The increasing availability of auxiliary information (deriving by integrating administrative registers and survey frames) facilitates the use of predictions. Under a model-based inference, the y_{rk} values are

assumed to be the realization of a superpopulation model *M*. The model we study has the following general form:

$$\begin{cases} y_{rk} = f_r(\mathbf{x}_k;\boldsymbol{\beta}_r) + u_{rk} \\ E_M(u_{rk}) = 0 \quad \forall k;\ E_M(u_{rk}^2) = \sigma_{rk}^2;\ \ E_M(u_{rk},u_{rl}) = 0\ \forall k \neq l \end{cases}, \tag{2.2.2}$$

where $\mathbf{x}_k$ is a vector of predictors (available in the sampling frame), $\boldsymbol{\beta}_r$ is a vector of regression coefficients, $f_r(\mathbf{x}_k;\boldsymbol{\beta}_r)$ is a known function, u_{rk} is the error term and $E_M(\cdot)$ denotes the expectation under the model. The parameters $\boldsymbol{\beta}_r$ and the variances σ_{rk}^2 are assumed to be known, although in practice they are usually estimated. The model (2.2.2) is variable-specific and different models for different variables may be used without creating any additional difficulty.

2.3 The selection scheme

Starting from the inclusion probabilities, the selection scheme completes the definition of the sampling design, i.e. the distribution *p* (.) of the samples *s* with $p(s) \geq 0$ being $\sum_{s \in U} p(s) = 1$. We restrict the sampling framework to the selection schemes by implementing the single-stage without replacement sampling design. The sample designs we consider belong to a general framework with varying inclusion probabilities, planned domains and target domains.

The class of such sampling designs is a specific case of balanced sampling under the randomization approach (Deville and Tillé, 2004).

Let $\mathbf{z}_k$ be a vector of auxiliary variables available for all $k \in U$. A sampling design *p*(*s*) is said to be balanced on the auxiliary variables if and only if it satisfies the following

balancing equation $$\sum_{k \in s} \frac{\mathbf{z}_k}{\pi_k} = \sum_{k \in U} \mathbf{z}_k \tag{2.3.1}$$

for each sample *s* so that *p*(*s*) >0 (Deville and Tillé, 2004). Depending on the auxiliary variables and the inclusion probabilities, the equation (2.3.1) can be exactly or approximately satisfied in each possible sample; therefore, a balanced sampling design does not always exist. By specifying

$$\mathbf{z}_k = \pi_k \boldsymbol{\delta}_k, \tag{2.3.2}$$

The equation (2.3.1) becomes $\sum_{k \in s} \boldsymbol{\delta}_k = \sum_{k \in U} \pi_k \boldsymbol{\delta}_k$. In this case, the balancing equation states that the sample size achieved in each subpopulation U_h is equal to the expected size. In different contexts, Ernst (1989) and Deville and Tillé (2004; p. 905 Section 7.3), have proved that, (*i*) with the specification (2.3.2) and (*ii*) if the vector of the expected sample sizes, given by $\mathbf{n} = \sum_{k \in U} \pi_k \boldsymbol{\delta}_k$, includes only integer numbers, a balanced sampling design always exists. Specification (2.3.2) defines sampling designs that guarantee the equation (2.1.1), upon which we wish to focus on. Deville and Tillé (2004, pp. 895 and 905), Deville and Tillé (2005, p. 577) have shown that several customary sampling designs may be considered as special cases of balanced sampling, by properly defining the vectors $\boldsymbol{\pi}$ and $\boldsymbol{\delta}_k$ of equation (2.3.2) Balanced samples may be drawn by means of the Cube method (Deville and Tillé, 2004). The Cube method satisfies the (2.3.1) exactly when (2.3.2) holds and **n** is a vector of integers. In the Simple Random Sampling without Replacement (SRSWOR)

designs and SSRSWOR cases, the standard sample selection methods can be used, as well as the Cube method. Deville and Tillé (2005) propose

$$\mathrm{E}_p(\hat{t}_{(dr)} - t_{(dr)})^2 \cong [N/(N-H)]\left[\sum_{k\in U}(1/\pi_k - 1)\eta^2_{(dr)k}\right] \quad (2.3.3)$$

As approximation of the variance for the HT estimator in the balanced sampling, where E_p denotes the sampling expectation under the sampling distribution, and $\eta_{(dr)k}$ are the regression residuals of $y_{rk}\ \gamma_{dk}$ with respect to the $\boldsymbol{\delta}_k$ variables, being $\eta_{(dr)k} = y_{rk}\ \gamma_{dk} - \pi_k \boldsymbol{\delta}'_k [\mathbf{A}(\boldsymbol{\pi})]^{-1} \sum_{j\in U} \pi_j\,(1/\pi_j - 1)\,\boldsymbol{\delta}_j\ y_{rk}\ \gamma_{dk}$ With $\mathbf{A}(\boldsymbol{\pi}) = \sum_{j\in U} \boldsymbol{\delta}_j \boldsymbol{\delta}'_j\, \pi_j (1-\pi_j)$.

2.4 Optimization problem

In our context, the variability of the estimates of $t_{(dr)}$ depends on the randomization framework and on the uncertainty of the y_{rk} values. A natural measure of uncertainty is the Anticipated Variance (*AV*) (Isaki and Fuller 1982). In our context, we can only define an Approximate *AV* (*AAV*) originated by the approximation (2.3.3):

$$AAV(\hat{t}_{(dr)}) = [N/(N-H)] \sum_{k\in U} (1/\pi_k - 1)\, E_M(\eta^2_{(dr)k}), \quad (2.4.1)$$

Where the terms $\eta^2_{(dr)k}$ in (2.3.3) are replaced by $E_M(\eta^2_{(dr)k})$. By defining $\tilde{y}_{rk} = f_r(\mathbf{x}_k; \boldsymbol{\beta}_r)$ the equation (2.4.1) may be reformulated as

$$AAV(\hat{t}_{(dr)}) = [N/(N-H)]\left[\sum_{k\in U}\frac{1}{\pi_k}(\tilde{y}^2_{rk} + \sigma^2_{rk})\gamma_{dk} - \sum_{k\in U}(\tilde{y}^2_{rk} + \sigma^2_{rk})\gamma_{dk} - AAV_{3(dr)}\right],$$

Where the third variance component, $AAV_{3(dr)}$ is an implicit function of $\boldsymbol{\pi}$ and is defined in Falorsi and Righi (2015).

Having defined the approximate anticipated variance, the vector of $\boldsymbol{\pi}$-values is determined by solving the following optimization problem:

$$\begin{cases} Min(\sum_{k\in U} \pi_k\, c_k) \\ AAV(\hat{t}_{(dr)}) \le \overline{V}_{(dr)} \quad (d = 1,...,D; r = 1,...,R), \\ 0 < \pi_k \le 1 \qquad\qquad (k = 1,...,N) \end{cases} \quad (2.4.2)$$

Where c_k is the cost for collecting information from unit *k* and $\overline{V}_{(dr)}$ is a fixed variance threshold corresponding to $\hat{t}_{(dr)}$. System (2.4.2) minimizes the expected cost, ensuring that the anticipated variances are bounded and that the inclusion probabilities lie between 0 and 1. If all the c_k values are constants equal to 1, then the problem (2.4.2) minimizes the sample size. We note that in problem (2.4.2) the variances σ^2_{rk} in $AAV(\hat{t}_{(dr)})$ are treated as known; in practice they must be estimated. Falorsi and Righi (2015) show an empirical evaluation on the sensitivity of the overall sample size with different estimated values of σ^2_{rk}. Finally, the authors describe the algorithm producing the solution of the problem (2.4.2).

Remark 2.1:

The solution defines equal inclusion probabilities for the units belonging to the same target domains with the same $\tilde{y}_{rk}$ and model uncertainty. This means that the computational effort for carrying out the algorithm can be extensively reduced when the superpopulation model is defined by means of categorical variables for obtaining homogeneous $\tilde{y}_{rk}$ by classes. In this way, the effective population size is the number of records in a given target domain with equal $\tilde{y}_{rk}$. On the other hand, for large populations with unit specific $\tilde{y}_{rk}$ values, the solution cannot be computed due to computational limits. An example of the use of a superpopulation model with categorical variables is the SSRSWOR design.

3. Examples

Let us now introduce an example from business surveys in order to illustrate how the optimal inclusion probabilities of several customary sampling designs may be obtained by properly defining the planned domains and the superpopulation models of the target variables. Thus, suppose that the survey estimates must be calculated separately considering three domain types: region (20 modalities), economic activity (2 modalities: goods and services) and enterprise size (3 modalities: small, medium and large enterprises). Hence, there are *D*=20+2+3=25 possible overlapping estimation domains.

3.1. Stratified simple random sampling without replacement

Planned domains: The single planned domain U_h is identified by a specific intersection of the categories of the estimation domains. In this case, $H = 20 \times 2 \times 3 = 120$ planned domains are defined and they represent a specific partition of *U*. The planned domains do not overlap: being $\sum_h \delta_{hk} = 1$, each planned domain can be viewed as a stratum.

Superpopulation model: The uniform stratum mean model is used. The model implies that the predicted values and model variances are $\tilde{y}_{rk} = \bar{Y}_{rh}$ and $\sigma^2_{rk} = \sigma^2_{hk}$ for all $k \in U_h$. The *AVV* becomes

$$AAV\,(\hat{t}_{(dr)}) = [N/(N-H)] \sum_{h \in H_d} \sigma^2_{rh}\, N_h\, (N_h / n_h - 1)\,, \qquad (3.1.1)$$

Where H_d is the set of planned domains included in U_d ? The above expression for the *HT* estimate in the SSRSWOR design is proven true when the number of strata *H* remains small compared to the overall population size *N*, and when the domain sizes N_h are large. Furthermore, all the units in the stratum have an equal inclusion probability $\pi_k = n_h / N_h$ for $k \in U_h$.

Note. The *Neyman* allocation, for which $n_h \approx n \sigma_{rh} N_h$, is obtained when: (*i*) the costs c_k are uniform, (*ii*) there is only one target variable (*r=R=1*) and (*iii*) the population *U* represents the only estimation domain. Furthermore, if the variances are constant

over the strata, that is $\sigma_{rh} = \overline{\sigma}_r$, then the proportional allocation is implemented, resulting in $n_h \approx N_h$.

3.2 Stratified sampling with varying probabilities without replacement

Planned domains: The planned domains are obtained as in the SSRSWOR case.

Superpopulation model: A general model is introduced with a model variance proportional to some size variable *x.* A customary model adopted in this case is the *Ratio Model,* in which $\tilde{y}_{rk} = x_k \overline{Y}_r / \overline{X}$ and $\sigma^2_{rk} = \sigma^2_r x_k$ where x_k is an auxiliary variable measuring the size of unit *k* , being $\overline{Y}_r$ and $\overline{X}$ the population means of the *r*-the target variable and of the auxiliary variable. Note that if the same conditions described for the Neyman allocation hold, the optimal inclusion probabilities are $\pi_k \approx \sqrt{x_k}$ (Särndal *et al.*, 1992, ch 12).

3.3 Incomplete multi-way Stratified Sampling design

Planned domains: The planned domains can be defined according to different options, two of which are described below. The planned domains overlap in all the options, so that $\sum_h \delta_{hk} > 1$.

Option 1. The planned domains U_h coincide with the estimation domains. Therefore, $H = D = 25$ and the $\boldsymbol{\delta}'_k$ are defined as vectors with three 1's, so that $\sum_h \delta_{hk} = 3$.

Option 2. The planned domains U_h are defined as (i) region by economic activity and (ii) economic activity by enterprise size; hence, $H = (20 \times 2) + (2 \times 3) = 46$ with $\sum_h \delta_{hk} = 2$.

Other Options. Other intermediate relationships among estimation and planned domains are possible.

In all the above mentioned options, the sample sizes are not controlled at the stratum level but are still controlled at an aggregated level while keeping the sampling accuracies under control. In Option 1 the sample sizes are controlled only for the estimation domains. In Option 2, the sample sizes are controlled for the subsets of two different partitions, defined by (i) region by economic activity and (ii) economic activity by enterprise size.

Superpopulation model: Different superpopulation models can be defined. A reasonable alternative is to define the uniform stratum mean model as for the SSRSWOR case. By means of this alternative, the units in the stratum have a uniform inclusion probability as in the SSRSWOR design, although the stratum sample sizes remain uncontrolled. The ratio model illustrated in section 3.2 can be adopted if a predictive *x* variable is available in the population frame.

4. Extensions

The framework allows for dealing with more specific estimation problems in an integrated way. In the following, we sketch some examples.

4.1 Design and model based mixed sampling strategies

If a given population partition, say *B*, defines a too large number of domains, the budget constraints may oblige to define too large prefixed sampling errors of the direct estimators of the domains of the partition *B* In this situation, it may be necessary to adopt a mixed estimation strategy in which an indirect small-area estimator is implemented by computing the estimates for the domains of the partition *B* so as to control the mean squared errors, while the estimates of the domains of other partitions are obtained thanks to the usual design based estimator. The basic idea is to compute modified η values which take into account the gain in efficiency by using a *small area estimator* with respect to the usual direct estimator. Falorsi and Righi (2008) shows how to adapt the algorithm for this case.

4.2 Total non-response

To make things simple, we describe this case by considering the SSRSWOR case described in section 3.1. Suppose that: (*i*) the phenomenon of nonresponse is substantially different among three different *response subgroups*, U_g (*g*=1,…, 3) given by the *legal status* of the enterprise (e.g. *private company, corporation and cooperative*).
In addition, suppose that (*i*) the response propensities, θ_k, are roughly constant for the units belonging to the subgroup U_g, that is $\theta_k \cong \theta_g$ for $k \in U_g$; (*ii*) when planning the sample design, a quite reliable estimate, say $\tilde{\theta}_g$, of the response propensity of the enterprises belonging to U_g may be obtained through previous surveys and (*iii*) the subgroup indicator variables are available in the sampling frame.
Note that the subgroups intersect the strata (as defined in section 3.1) and individuate a partition of the population *U* which differs from that given by the strata. The sampling practitioner would define a sampling plan controlling the sample sizes both at the stratum level (to deal with the sampling variance) and at the nonresponse subgroup level (to manage the non-response phenomenon in the sampling design phase). In the traditional setting, the only way to deal with the two somehow conflicting goals is to stratify the sampling units by region, economic activity, enterprise size and legal status, thus defining $H = 20 \times 2 \times 3 \times 3 = 360$ strata. Such a detailed stratification has a lot of *drawbacks* (see Falorsi, *et al.* 2006) while the *generalized framework* offers a flexible tool for dealing with the two objectives by defining 123 planned domains: the 120 strata (as described in section 2.1) plus the 3 *response* subgroups. In this setting, the estimator of the totals of interest can be calculated with the *direct weighting estimator* (Särndal, 1992, expression, (15.6.8)):

$$_{mr}\hat{t}_{(dr)} = \sum_{g=1}^{3} \frac{n_g}{m_g} \sum_{k \in s_g^*} y_{rk} \gamma_{dk} / \pi_k \qquad (r = 1,\ldots, R;\ d=1,\ldots, D), \qquad (4.2.1)$$

Where: s_g^* is the sample of respondents in U_g, being $n_g = \sum_{U_g} \pi_k$ the planned size of the sample s_g. Under the hypothesis that with estimator (4.2.1) the nonresponse bias becomes negligible and considering the response phenomenon as a second phase of sampling *,* the *AAV* of (4.2.1) may be computed by (Särndal and Lundström;

2005, page. 150): $AAV(_{mr}\hat{t}_{(dr)}) = AV_{sam} + AV_{NR}$ in which AV_{sam} is the anticipated variance of the *HT* estimator in the absence of nonresponse, given by (3.1.1) and AV_{NR} represents the additional part of variability due to the non-response phenomenon. Under the Response Homogeneity Group Model, the AV_{NR} is given by (Särndal, 1992, expression, (15.6.8))

$$AV_{NR} = E_p E_{\mathbf{n}(g)} \left(\sum_g n_g \ \frac{1-\tilde{\theta}_g}{\tilde{\theta}_g} S^2_{(dr)\hat{y}} \ \middle| s \right), \qquad (4.2.2)$$

where $E_{\mathbf{n}(g)}$ is the sampling expectation with respect to the distribution of n_g, given *s*, and $S^2_{(dr)\hat{y}}$ is the variance in s_g of $\tilde{y}_{rk}\gamma_{dk}/\pi_k$. Given that the n_g are planned in advance, by adopting the upward approximation to the variance of the *Poisson sampling,* the (4.2.2) may be approximated by:

$$AV_{NR} \cong \frac{1}{n_g} \sum_{k \in U_g} \left(\frac{1}{\pi_k} \left(\tilde{y}^2_{rk} + \sigma^2_{rk} \right) \gamma_{dk} \left(1 - \frac{1}{n_g}\right) \right) + \frac{1}{n_g^2} \sum_{k \in U_g} \tilde{y}^2_{rk} \gamma_{dk} - \frac{1}{n_g^2} \tilde{t}^2_{(dr)g}, \quad (4.2.3)$$

being $\tilde{t}_{(dr)g} = \sum_{k \in U_g} \tilde{y}_{rk} \gamma_{dk}$. On the basis of (4.2.3), the optimization problem (2.4.2) can be easily reformulated by considering the additional part of the variance deriving from the expected non response.

Thus, having reliable estimates of the response propensities $\tilde{\theta}_g$, it is possible to define the inclusion probabilities that individuate the minimum cost solution. This allows for respecting the accuracy constraints in presence of nonresponse.

4.3 Indirect Sampling

The techniques for the extension of the approach for defining an optimal sampling in the multivariate case where the variables of interest are related to different target populations are shown in FAO (2015) and in Falorsi, Lavallée and Righi (2016, ICAS conference, Rome 2016). In order to get insight in the underlying phenomena, the observation has to be carried out in an integrated way, implying that units of a given population and the related units of the other population must be observed jointly. The sampling procedure assumes a population U^A related to the interest population U^B, given that the sampling frame of U^A is available. Then, a sample on U^A is carried out and, by using the existing links between the two populations, the units of U^B are observed. Different scenarios related to the level of knowledge of the existing links are examined in the above papers. The starting scenario assumes that the links between the populations are known in the design phase; the second scenario assumes that the links between U^A and U^B are estimated in the design phase; in the third scenario no links between U^A and U^B are available, but auxiliary variables on U^A can provide useful information on U^B.

5. CONCLUSIONS

The paper proposes a tool for defining the optimal inclusion probabilities suitable in various survey contexts characterized by the need to disseminate survey estimates of prefixed accuracy, for a multiplicity of both variables and domains of interest. The basic assumption of the proposed tool is the knowledge of domain membership indicator variables, while the variables of interest are not known. The procedure is, thus, applied on the predicted values of the characteristics of interest via a superpopulation model, and the algorithm enables to consider the model uncertainty by using the Anticipated Variance as the measure of the estimators' precision. This is, generally, the limit of the standard algorithms for the sample allocation, in which the variables are assumed known.

This framework considers explicitly the use of a superpopulation model in the design phase and allows for the use, at least theoretically, of complex models. Note that, the search for optimal stratification should contemplate the use of a superpopulation model. The framework provides two elements to manage the model complexity in the sampling design, namely the concepts of planned and estimation domains. For instance, if the chosen superpopulation model produces homogeneous predicted values for subgroups of population units, we are defining sampling strata. A suitable sampling design fulfilling this model is the customary stratified sampling design which obliges to define strata as the planned domains, and the estimation domain as aggregation of strata. This approach has some serious drawbacks when strata are too small (Falorsi *et al.*, 2006) and can result in an inefficient sampling strategy. The proposed framework augments the degrees of freedom for defining the planned domains (section 3) and a possible choice is that the planned and target domain coincide.

In conclusion, the paper does not provide the algorithm. Falorsi and Righi (2015) describes the tool in detail, offers a simulation on real survey data to evaluate its performance and properties and gives proof of the convergence of the iterative process. Here we stress the algorithm is implemented in R language and will be available on the Italian National Statistical Institute website (http://www.istat.it/en/tools/methods-and-it-tools) as soon as possible.

References

Bethel, J. (1989). Sample Allocation in Multivariate Surveys, *Survey Methodology*, 15, 47-57.

Cochran, W.G. (1977). *Sampling Techniques*. Wiley. New York.

Choudhry, G. H., Rao, J. N. K., and Hidiroglou, M. A. (2012). On sample allocation for efficient domain estimation. *Survey Methodology*, 18, 23-29.

Chromy, J. (1987). Design Optimization with Multiple Objectives, *Proceedings of the Survey Research Methods Section. American Statistical Association*, 194-199.

Deville, J.-C. and Tillé, Y. (2004). Efficient Balanced Sampling: the Cube Method, *Biometrika*, 91, 893-912.

Deville, J.-C. and Tillé, Y. (2005). Variance approximation under balanced sampling, *Journal of Statistical Planning and Inference*, 128, 569-591.

Ernst, L. R. (1989). Further Applications of Linear Programming to Sampling Problems, *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 625-631.

Falorsi, P. D., Lavallée P., and Righi, P. (2016). Optimal Sampling for the Integrated Observation of Different Populations, ICAS Conference, Roma 2016.

Falorsi, P. D. and Righi, P. (2015). Generalized framework for defining the optimal inclusion probabilities of one-stage sampling designs for multivariate and multi-domain surveys, *Survey Methodology*, 41, 215-236.

Falorsi, P. D. and Righi, P. (2008). A Balanced Sampling Approach for Multi-way Stratification Designs for Small Area Estimation, *Survey Methodology*, 34, 223-234.

Falorsi, P. D., Orsini, D., Righi, P., (2006). Balanced and Coordinated Sampling Designs for Small Domain Estimation, *Statistics in Transition, 7, 1173-1198.*

FAO (2015). *Integrated Survey Framework*, Guidelines, Rome. http://www.gsars.org/wp-content/uploads/2015/05/ISF-Guidelines_12_05_2015-WEB.pdf. Accessed on 16 December 2015.

Gonzalez, J. M. and Eltinge, J. L. (2010). Optimal Survey Design: A Review, *Section on Survey Research Methods – JSM 2010*, October.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under a regression superpopulation model, *Journal of the American Statistical Association*, 77, 89-96.

Khan, M. G. M., Mati, T., and Ahsan, M. J. (2010). An optimal Multivariate stratified sampling design using auxiliary information: An integer solution using goal programming approach. *Journal of Official Statistics*, 26, 695-708.

Kokan, A. and Khan, S. (1967). Optimum allocation in multivariate surveys: An analytical solution. *Journal of the Royal Statistical Society*, *Series B*, 29, 115-125.

Särndal C.E., Lundström S. (2005). *Estimation in Surveys with Nonresponse*, Wiley, New York

Book and Software Review

Methodological Developments in Data Linkage, edited by Katie Harron, Harvey Goldstein and Chris Dibben, Wiley Series in Probability and Statistics, Chichester: John Wiley & Sons, 2016

Reviewed by: Natalie Shlomo, School of Social Sciences, University of Manchester

A recently published book by Wiley titled 'Methodological Developments in Data Linkage' edited by Katie Harron, Harvey Goldstein and Chris Dibben (Wiley Series in Probability and Statistics, 2016) contains 10 stimulating chapters on past, present and future of research into data linkage. The book focuses on current state-of-the-art aspects of data linkage where the aim is to identify same entities across datasets to produce a linked dataset. There is no doubt that with increasing availability of administrative and register data and more computing power, there is much to gain in merging records existing in single datasets to form a linked and enhanced dataset. In addition, to the state-of-the art, the reader is left with a good overview of open areas of future research.

The first and last chapters are written by the editors where the first chapter provides an overview of the contents of the remaining chapters and the final chapter a brief summary and areas of future research. In between, there are eight chapters written by experts in the area of data linkage as described below.

- Chapter 2: 'Probabilistic Linkage' is written by William E. Winkler who is an eminent researcher at the US Census Bureau and has contributed to the data linkage literature in the last two decades. The author provides a good overview of the Fellegi-Sunter model of record linkage, estimation of parameters without training data and other practical aspects of record linkage with respect to preparing the datasets, data cleaning, standardization and other pre-processing stages, string comparators and blocking. He also provides an overview of estimating error rates (false matches and false non-matches) which are needed to feed into measurement error models when analyzing linked data (see chapter 5).

- Chapter 3: 'The Data Linkage Environment' is written by Chris Dibben, Mark Elliot, Heather Gowans, Darren Lightfoot within the UK Administrative Data Research Network. This chapter presents considerations of anonymization, the security infrastructure and governance processes needed to ensure safe data linkage. The chapter provides models for data access and data linkage and also

contains case studies from Centers of Data Linkage from Canada, Australia and the United Kingdom. It contains important information for data archives wishing to develop data linkage capabilities.

- Chapter 4: 'Bias in Data Linkage Studies' is written by Megan Bohensky from the University of Melbourne in Australia. The chapter contains a description of types of linkage errors and how errors impact on research findings as well as provides a good overview of quality indicators. In addition, the author presents a meta-analysis of 13 studies out of which finds that 11 studies have some type of bias in the linkage results. Recommendations and implications for best practice are also provided.

- Chapter 5: 'Secondary Analysis of Linked Data' is written by Ray Chambers and Gunky Kim from the University of Wollongong in Australia. The chapter provides an excellent overview based on earlier published papers on how to compensate for linkage errors in statistical analysis of one-to-one linked data through a measurement error model. The authors propose a misclassification matrix based on an exchangeable linkage error model for those researchers analyzing the linked dataset as secondary data without access to matching probabilities. They then proceed to present theory on compensating for linkage error for regression analysis (linear, logistic) and also considers sample to register linkage, multiple linkages and weighting via calibration to deal with the non-linked records.

- Chapter 6: 'Record Linkage: a Missing Data Problem' is written by Harvey Goldstein and Katie Harron from the United Kingdom. The authors suggest that dealing with linkage errors under a one-to-one match scenario should have the same considerations as treating missing data. They differentiate between linkages that are unequivocal (deterministic link) vs equivocal (probabilistic link) and compare two approaches for analysis under linkage errors: multiple imputation based on the equivocal linkages (transformations are carried out on the data to assume a multivariate normal distribution) and a prior-informed imputation based on the equivocal linkages where the prior probabilities are obtained directly from the matching process. The authors note that these have the potential to reduce bias from linkage error whilst allowing for more correct inference by taking into account the variability arising from the linkage process. The authors provide an example from linking electronic healthcare data in the UK and a simulation study introducing non-random linkage errors.

- Chapter 7: 'Using Graph Databases to Manage Linked Data' is written by James M. Farrow from Australia. The author presents a graphical (network) representation of data linkage as a collection of nodes (the records) connected by edges (their relationships). By storing the data in a graph database, this allows new approaches to exploring the linked data. The author provides an easily accessible and understood comparison between flat databases and graph databases.

- Chapter 8: 'Large-scale Linkage for Total Populations in Official Statistics' is written by Owen Abbott, Peter Jones and Martin Ralphs from the United Kingdom Office for National Statistics. The authors discuss data linkage

technology for traditional censuses as well as alternative censuses based on population registers and other national population datasets. Data linkage is the key driver to the success of coverage adjustments for both a traditional and alternative census using a capture-recapture approach for estimating the population size that is missed based on an independent dataset such as a post-enumeration survey. Besides the United Kingdom, case studies are provided from Finland, The Netherlands, Poland and Germany. The chapter concludes with a description of the United Kingdom Beyond 2011 Program undertaken by the Office for National Statistics to replace the traditional census with an alternative census by combining national population datasets. The challenge is to deal with the linkage of anonymized datasets arising from different government agencies by encrypting the matching variables using hash functions. A trusted party has access to the strings for producing string comparators which are then fed into the linkage process.

- Chapter 9: 'Privacy-preserving Record Linkage' is written by Rainer Schnell from the University of Duisburg-Essen in Germany. The author describes the context of privacy preserving record linkage where the matching variables are encrypted using computer science encryption such as hash functions and bloom filters. A similarity measure such as the Dice Coefficient can be used to represent how close the matching variables are across the two files. Blocking is another important consideration in privacy preserving record linkage and there are several alternative methods that are listed. The author concludes with future areas of research and in particular the need to have a systematic 'proof of concept' to initiate wider use of these methods in data linkage enclaves and statistical agencies.

The book follows on from two previous books published on data linkage: Data Quality and Record Linkage Techniques by T.N. Herzog, F. Scheuren and W.E. Winkler published in 2007 and Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution and Duplicate Detection by P. Christian published in 2012. The difference with this book is that it is an edited volume containing chapters on a wide range of state-of-the-art topics in data linkage by a broad ensemble of authors. It is a credit to the editors who clearly have a well-rounded vision of current and future research in data linkage which has led to a good selection of topics covered and the identification of expert contributors. The chapters also cover relevant and current examples and case studies for researchers interesting in implementing data linkage in a variety of areas such as medical and public health research, official statistics, social statistics, policy and evaluation. The most important contribution of the book is that it produces the stimuli for continuing research into data linkage and identifies future research needs. These are summarized in the final chapter by the editors and I list some of them below with my own augmentation and interpretation as a reviewer and as a researcher in data linkage.

Quality Indicators:

One point that was highlighted many times throughout the book was the need for more quality indicators and other processing information from the agencies conducting data linkage for use in secondary analysis by researchers. On one end of the spectrum, the full range of matching probabilities (preferably in a vector format

where all matching probabilities of a record are provided for all other records in a block) would provide the misclassification matrix for the secondary analysis through regression modelling under linkage errors as described in Chapter 5 as well as inform the priors in the imputation approach described in Chapter 6. At this point in time, it seems unlikely that Data Linkage Centers will be releasing this information. On the other end of the spectrum, even if only quality indicators are released as outlined in Chapter 4, this may provide some information to improve on the Exchangeable Linkage Error Model for reproducing the misclassification matrix as described in Chapter 5 since it is clear from Chapter 4 that this model would likely not account for potential bias arising from the selectivity in data linkage where records not linked are related to the target variables of analyses. A seemingly middle approach would be to release the matching parameters since it may be possible under certain conditions of good 'separation' between matches and non-matches to simulate the matching probabilities and error rates through a bootstrap procedure as described in Winglee, Valliant and Scheuren (2005) or Chipperfield and Chambers (2015).

Informative linkage errors:
One important conclusion from the book is that forms of non-linkage and other linkage errors can be similarly categorized as the well - established mechanisms of non-response, specifically informative vs non-informative linkage errors. This is now a focus of intense research being undertaken through the work packages of the United Kingdom National Centre for Research Methods, specifically on informative missingness in bio-social research and biomarkers collected in longitudinal surveys (WP3) (See: **http://www.ncrm.ac.uk/research/**). One of the key drivers for missing links in biosocial research is due to the need to ask for consent from respondents to undergo physical and blood tests, many of whom refuse for reasons related to variables of interest. Therefore, strategies developed for dealing with informative missingness in statistical analysis are also useful in the context of analysis of linked data with linkage errors. These may include approaches such as propensity weighting (as proposed in Chapter 5, pp. 108), the use of selection models or measurement models in secondary analysis as well as other approaches. It seems likely that some form of calibration data may be needed to account for selectivity.

Privacy preserving record linkage:
As a relatively new area of research, privacy preserving record linkage is making the cross-over from the realm of computer science into the statistics community. There is still much work to be done to prove that the method is viable and 'fit for purpose'. Recent research at the University of Manchester in the framework of the recently completed EU 7th Framework Research Grant: Data without Boundaries, is to develop privacy preserving *probabilistic* record linkage. The aim is to develop a 'black box' for data linkage within and across data archives where data custodians would allow encryption of matching variables instead of simply deleting them from the files. Then, users can request a probabilistic data linkage based on the Fellegi-Sunter approach through an interface. Similarity scores are calculated between encrypted strings and the overall match score is then used to define bins for a multinomial EM Algorithm to estimate the matching probabilities. Smith and Shlomo, 2014 propose using a concatenated 1-bit minwise hashing on bigrams of strings and also propose

an estimate for the Jaccard Similarity Score to be used in the data linkage. As listed in Chapter 9 and from other research on privacy preserving record linkage, there are still many open questions for research, mainly how to link datasets that have undergone anonymization with coarsening and suppression of potential matching variables. As mentioned in Chapter 3, if there is a safe environment holding the datasets, it will be easier to convince data custodians to encrypt strings rather than to delete them from the files. Another problem is that clerical review cannot be undertaken under privacy preserving record linkage and hence the data linkage must be robust and allow for the calculation of potential linkage errors to feed into the statistical models of analysis as proposed in Chapter 5. Finally, how to encrypt ordinal or temporal variables and allow for similarity scores that are defined by the distance between numerical values remains an open question.

Investment in training:
It is clear that both researchers involved in carrying out data linkage as well as researchers involved in the analysis of secondary linked data need to undergo sufficient training. For those carrying out the data linkage, there is a need to understand the processes and release quality indicators and other information about the matching to allow researchers to take linkage errors into account in their analysis. For those researchers analyzing linked data, they must be well trained to use advanced methods of analysis that compensate for linkage (and other measurement) errors. This means investing in training away from the traditional frequentist approaches of data analysis and more emphasis on model based approaches to account for selectivity and informativeness.

References:

Chipperfield, J. O. and Chambers, R. (2005) Using the Bootstrap to Analyze Binary Data Obtained via Probabilistic Linkage. Journal of Official Statistics (in press)

Smith, D. and Shlomo, N. (2014) Record Linkage Approaches for Dynamic Database Integration: Privacy Preserving Probabilistic Record Linkage, Deliverable 11.1 for WP11, Data Without Boundaries. Available at: http://www.cmist.manchester.ac.uk/medialibrary/archive-publications/reports/2014-01-Data_without_Boundaries_Report.pdf

Winglee, M., Valliant, R. and Scheuren, F. (2005) A Case Study in Record Linkage. Survey Methodology, Vol. 31 (1), 3-12.

AUSTRALIA

Reporting: Anthony Russo

Winsorisation of rates for ABS Survey of Average Weekly Earnings

In business surveys, it is possible for a sample to include a small number of units with highly unusual values. These units are referred to as outliers. Their selection in the sample, combined with the application of the specified selection weights for these units, can mean that survey estimates appear implausible in relation to economic relationships or the commonly accepted view. In the Australian Bureau of Statistics (ABS), the two main methods of addressing outliers are surprise outliering and winsorisation.

Surprise outliering reduces the design weight of an outlier unit to one, while adjusting the weights of the other units in the stratum upwards to compensate. While surprise outliering is relatively simple to implement, there are some issues with this method. These include:

- The identification of surprise outliers can be a subjective choice, and require a large amount of effort; and
- Setting a unit as a surprise outlier can have unexpected effects for average or rate estimates.

Winsorisation outliering identifies extreme reported values and replaces these with more reasonable values; design weights are left unchanged. This is equivalent to adjusting the weight to a value somewhere between 1 and the original selection weight. The identification of winsorised outliers depends on objective cut-offs calculated using historical survey data. Reported values that exceed a given cut-off will be modified to a more central value. The theory for winsorising level estimates is reasonably well established. However, the theory for directly winsorising rate estimates has not yet been developed.

An investigation was undertaken to assess whether winsorisation for rate estimates can be effectively accomplished by winsorising the components of the rate separately. Data from the Survey of Average Weekly Earnings (AWE) was used for this investigation. The AWE survey is a biannual collection that obtains data on employment and earnings from businesses in order to produce estimates of average

weekly earnings (total earnings divided by total employment). These estimates are produced for males, females and all persons, for particular classifications.

The approach taken for this investigation was to winsorise the main earnings and employment data items for AWE separately. Winsorisation was applied to historical AWE survey data (excluding the effect of surprise outliers), and the resulting estimates were compared with published estimates.

Results from the investigation were positive. In general, the impact on estimates when winsorisation was used instead of surprise outliering was non-significant, with most winsorised rate estimates lying well within the 95% confidence interval of the original published estimates. One benefit of winsorising the rate components has been that while we only target units with unusually large earnings or employment values, the winsorisation process will pick up units that have either an unusually high or unusually low earnings-to-employment ratio.

Winsorisation outliering of the main earnings and employment data items has been introduced as the primary method to treat outliers in AWE, and surprise outliering continues to be used for a small number of extreme values that may not be sufficiently moderated by the winsorisation method. Further refinement of the method used for AWE is planned in the near future to also account for some of the other data items such as overtime and non-fulltime earnings and employment.

For more information, please contact Lyndon Ang (lyndon.ang@abs.gov.au).

BOSNIA AND HERZEGOVINA

Reporting: Edin Šabanović

Pilot Survey on Income and Living Conditions 2015 (Pilot EU-SILC 2015)

The EU-SILC has become the main source of comparable multidimensional micro-data on income, poverty, social exclusion and living conditions at both household and individual level. This survey is carried out in all European Member States under European Regulation and is coordinated by Eurostat, which establishes essential rules that ensure the provision of timely results and the use of harmonized methodologies within all European Member States.

Until 2014, Bosnia and Herzegovina, as a potential-candidate-country for EU membership, did not have the opportunity to conduct Income and Living Conditions survey, although there was an urgent need for data on income and social exclusion. Living standard in Bosnia and Herzegovina is still measured within Household

Budget Survey and by using consumption expenditure instead of income and thus the methodology of poverty analysis is not fully harmonized to EU standards.

Based on the IPA Multi-beneficiary Program support, The Agency for Statistics of Bosnia and Herzegovina conducted the first Pilot Survey on Income and Living Conditions in 2015. The objectives of this survey were as follows:

- To develop a household survey able to deliver micro-data related to income and living conditions with a view to a gradual convergence towards the acquis (European Regulation (EC) No 1177/2003).
- To develop and harmonize methodology, guidelines and requirements which are compatible with the EU-SILC legal frameworks and practices.
- To carry out a first EU-SILC pilot survey, which should set grounds for the launch of a full-scale enquiry on continuous basis in the near future.
- To evaluate the current ILC production system through the production of quality reports gradually compatible with the acquis, and propose methods by which current system might be more aligned with European practices and ESS guidelines.

Data collection was performed by CAPI method, which was the first application of this technique in the statistical system of Bosnia and Herzegovina. The CAPI application was made in Blaise program by IT staff of the Agency for Statistics.

All objectives were fully reached and SPSS micro-data files and quality report were sent to Eurostat by 30th November 2015. Lessons learned from the pilot survey will be used to fine tune the methodology and the organizational issues of the full-scale ILC survey in 2017. To conduct the full-scale survey, there are two potential major risks: the absence of the sampling frame from the population census 2013 (whose results were not published yet) and inadequate budget for the survey.

For more information please contact Mr. Edin Šabanović, Assistant Director, Sector for Statistical Methodology, Standards, Planning, Quality and Coordination (e-mail: edin.sabanovic@bhas.ba).

CANADA

Reporting: Steve Matthews and Normand Laniel

Trend-Cycles along with Seasonally Adjusted Series

In October of 2015 Statistics Canada began to publish graphical information on trend-cycle movements for several monthly economic indicators. Estimates of the trend-cycle are presented along with the seasonally adjusted data in selected charts in *The Daily, Statistics Canada's primary vehicle for dissemination.* The seasonally adjusted estimates remain the main economic data, however the inclusion of trend-cycle information is intended to support the analysis and interpretation of the data. In order to help data users to analyze the new estimates, supporting documentation that provides detail on the methodology and discusses selected issues related to the use and interpretation of trend-cycle estimates has been added to Statistics Canada's website.

For more information, please contact Steve Matthews at the following address Steve.Matthews@canada.ca.

2016 Census of Population: Return of the mandatory long-form

The Government of Canada has reinstated the mandatory long-form for its 2016 Census of Population. It replaces the voluntary National Household Survey introduced in 2011. The goal of the 2016 Census is to produce quality data for special populations and at all levels of geography, including small municipalities. The sampling rate will be 1 in 4 households. This is smaller than the 1 in 3 of 2011 but larger than the 1 in 5 of 2006. The new sampling rate is directed at maintaining the precision of estimates should the 2016 response rate be lower than in 2006. The latter has the potential to materialize due to the risk that some of the population continues to view the long form Census as voluntary. Data collection will start in February for the northern areas and in May for the southern areas where the large majority of the population lives. With the use of a letter inviting to respond via Internet at the first wave of collection, it is expected that around 65% of households will respond electronically. The plan is to release the population counts in early 2017.

For more information, please contact Normand Laniel at normand.laniel@canada.ca.

ISRAEL

Reporting: Tom Caplan, Central Bureau of Statistics, Israel

Israel is happy to present this report of its new and continuing statistical developments. This report will focus on the census of population, some of the key newer surveys, as well as other areas of current developmental and methodological work. The latter include the Census of agriculture, Big Data, Meta Data and an examination of the possibility of adopting the open source R software as its principal program for statistical analysis and development.

The Israel Central Bureau of Statistics (ICBS) is studying the pros and cons of having the next population census as an integrated rolling census over a five year period. The 2008 census was already an integrated one, where administrative data was supplemented by surveys to evaluate and correct for over and under-coverage, and provide socio-economic data. As part of the study, the ICBS will conduct two experiment in 2017 and 2018 aiming to- (a) test its ability to evaluate the quality of the administrative census file in small statistical regions, based on statistical models, (b) test the feasibility of a rolling census, mainly the possibility of borrowing data over time and across space, and (c) test production processes, mainly: effective sampling based on the above models, the use of the internet as a mode of data collection, the use of tablet PC and the evaluation of the experiment results. At the same time, additional administrative files will be checked as substitutes for direct data collection and thereby enrich the census database.

The Israel Social Survey, while not new is basically changing every year. The survey, which began in 2002 and whose sample contains 7500 individuals aged 20 and over, serves as a platform for examining broad social and economic themes. The Survey is set up so that there is a core questionnaire that is asked every year, and a dedicated special questionnaire that examines in depth a new specific theme every year. the themes examined since 2002 include pension coverage and retirement, multidimensional measurement of welfare, non-compulsory education frameworks, participation in the labor force and attitudes to employment, caring for people with functional difficulties, welfare, satisfaction with government services, social mobility, religiosity, family life, health, lifetime learning, wellbeing of the population environment, social capital, public views concerning government services and citizen involvement

The Israel Longitudinal Survey of Families has now completed its third wave of data collection. The survey, a joint project of the ICBS, the Ministry of Finance, the Bank of Israel, the National Insurance Institute and the Ministry of Education, is carried out once a year on a permanent sample of 5000 households. A wide range of topics is asked in the survey including health, education, employment, income, personal

finances and wealth, and housing and properties. Now that three waves of interviews have been completed, planning is underway for publication of some initial longitudinal results.

Methodological work is underway for carrying out a new Census of Agriculture. The last Census of Agriculture in Israel was conducted in 1995, and since then no comprehensive picture of agriculture in Israel has been available. In 2015, a new frame of agricultural farms – has been built by merging six different administrative files, including the business register. The coverage of this frame is now being verified through intensive fieldwork. This frame will serve as the basis for the next Census of Agriculture that will be conducted in 2017.

As part of a world trend among National Statistics Institutes, the ICBS is examining the possible use of Big Data for national statistics purposes. To this end, a broad ranging Task Force was established to examine the parameters of Big Data and their applicability to National Statistics. The Task force has met with representatives of commercial enterprises, technical companies and academic experts, each of which presenting their point of view on the problems, possible solutions and potential benefits from the use of Big Data, as compared to traditional surveys. After these consultations we started examining in much detail the application of two areas of Big Data to the work of the ICBS: use of retail cash register files to facilitate and enhance the computation of the Consumer Price Index, and the use of files from Cell Phone companies for analyzing travel features and behavior. Work on these two themes is continuing.

Meta Data is critical to the understanding of the basis of official statistics. To this end, the National Statistician appointed a steering committee and implementation team with the aim of studying, developing and implementing a program of metadata for all the statistics programs at the ICBS. After researching and examining several options, the Bureau adopted the SIMS (Single Integrated Metadata Structure) standard for (See: http://ec.europa.eu/eurostat/documents/64157/4373903/03-Single-Integrated-Metadata-Structure-and-its-Technical-Manual.pdf/6013a162-e8e2-4a8a-8219-83e3318cbb39).

Meta data, and the implementation team has developed an electronic application for inputting and coding the metadata. At present, a pilot test of the system is being carried out in the Education Statistics branch. Once the pilot is evaluated, the next step will be the implementation of the system in all the ICBS programs.

Finally, the ICBS is in the process of examining the possibility of adopting the open source statistical software "R" as its primary software for statistical development and analysis, instead of SAS. Here again, we established a task force to examine the pros and cons of switching to R. The Task Force has produced a comprehensive report which is now under consideration by the senior management. As part of the deliberations, the ICBS is seeking information from other national statistics institutions who are carrying out a similar investigation or which have switched to R already.

LATVIA

Reporting: Mārtiņš Liberts

R package "vardpoor" for variance estimation

The Central Statistical Bureau of Latvia has developed and published R package "vardpoor" for variance estimation. R is a widely known software environment for statistical computing (https://www.r-project.org/). It allows users to write their own functions thus extending the capabilities of R. User written functions are distributed through packages. For the first time package "vardpoor" was published in 2014.

The package "vardpoor" is a variance estimation tool for the sample survey results. The possibilities of the package include:

- Calculation of the domain-specific study variables, if the sampling errors have to be estimated for population domains;
- Linearization of non-linear statistics, for example, ratio of two totals and many other;
- Calculation of regression residuals if survey weights are calibrated;
- Variance estimation with the ultimate cluster method (Hansen, Hurwitz and Madow,1953);
- Estimation of different other variance measures, for example, standard error, coefficient of variation, margin of error, confidence interval;
- Estimation of design effect where design effect is split into two components – sampling effect and calibration effect.

There is no limitation on sampling design used for a sample survey. The procedure is using ultimate cluster method for variance estimation which can be applied to many sampling designs. The aim of the tool is to provide a good balance between statistical accuracy for variance estimates and operational efficiency. The package is the main tool currently used for variance estimation in the Central Statistical Bureau of Latvia.

The stable version of the package is available on CRAN (Comprehensive R Archive Network) and the development version is available on GitHub (Web–based Git repository hosting service):

- Stable version: https://cran.r-project.org/web/packages/vardpoor/
- Development version: https://github.com/CSBLatvia/vardpoor

Comments, bugs and recommended improvements can be reported on corresponding GitHub page. The authors of the packages Juris Breidaks

<juris.breidaks@csb.gov.lv> and Mārtiņš Liberts <martins.liberts@csb.gov.lv> can be contacted for more detailed information.

MALAYSIA

Reporting: Mohd Uzir Mahidin

Economic Census (EC) and Household Income and Expenditure Survey (HIES) 2016 with the New Enhancement

The year 2016, Department of Statistics Malaysia (DOSM) to conduct Economic Census (EC) and Household Income and Expenditure Survey (HIES) throughout the nation. Census and survey were enormous statistical project that has been undertaken in order to produce very useful data for planning and implementation of national development. In 2015, Malaysia has implemented its biggest tax reform with the introduction of Goods and Services Tax (GST) on April 1 at 6%, primarily to replace the sales tax (10%) and service tax (6%). With the implementation of GST, Department of Statistics Malaysia would incorporate the GST element in the census and survey to study the impacts of GST on household as well as business operation. In view of this matters, department also reviewed the EC and HIES questionnaire accordingly by putting up new elements in line with the new national development.

The EC is conducted every 5 years, to provide periodic and comprehensive statistics on business establishments and activities. Besides the GST element, the EC 2016 has enhanced the questionnaire by introducing the competitiveness information of establishment. The competitiveness information contained 6 (six) module which consist of Information & Communication Technology (ICT) and e-Commerce; Access to Financing; Innovation; Marketing and Promotion; Imports and Exports of Goods and Services; and Environmental Compliance. Among the new key element and very important area was e-Commerce which could identify type of customers by categories of Business to Business (B2B); Business to Consumers (B2C); and Business to Government (B2G).

Occasionally the expenditure survey for household is carried out once in five years while the income survey is done twice over a period of five years. However since 2014, the HIES were carried out simultaneously and would be conducted again in 2016 to study the impact of GST to household which has been implemented in 1 April 2015. The HIES 2016 will provide the crucial input for the impact of GST to household and therefore the results of this survey will enable the Government to formulate policies and programs that would benefit the people. Three types of questionnaires will be used: the Income Survey, Household Expenditure and

Facilities; Household Income and Basic Amenities; and the Daily Record Household Spending. Household transaction in e-Commerce would also be introduced in HIES 2016 which will be useful in the compilation of Information and Communication Technology of Satellite Account (ICTSA).

Monitoring such big project and various types of survey needs a comprehensive and integrated system encompassing pre-collection, data collection, processing, and analysis and dissemination activities. Therefore DOSM moving well into the information age by implementing the National Enterprise-wide Statistical Systems (NEWSS) to monitor the economic census project. Through a central database, NEWSS will perform several functions, including:

- To standardize, consolidate and improve the existing system/application to support the strategic requirement and the operation of Department of Statistics;
- To simplify, improve and expedite the process of statistical data dissemination;
- To develop an integrated business process management that adheres to international statistical standards; and
- To build up a central repository to facilitate data sharing between the Department of Statistics and other government agencies.

Another modernization method in garnering data, DOSM has equipped the field enumerators with personal digital assistants (PDA) to expedite the collection on selected surveys. With the advancement of ICT, the web based data collection, e-survey was used in Monthly Manufacturing Survey and selected Services surveys.

NEW ZEALAND

Reporting: Charlie Dohrman and Anders Holmberg

Sharing insights on collection methodology of economic surveys and the census

Statistics New Zealand teams collaborated on developing an online mode for the Agricultural Production Survey (APS). In July, we gave farmers the chance to start filling out their survey forms online, once they received details in the post. We also tested elements of the modernized census collection approach, taking advantage of the similarities between the standard collection methodology for the APS and the methodology to be used for the 2018 Census.

As well as an online form for respondents, the APS provided a platform to undertake testing earlier in the development cycle than would otherwise have been possible, and an ongoing vehicle for census testing on respondent behavior in the rural subpopulation. Understanding the limitations that came from operational constraints of using a live production survey for census testing has helped us plan later tests. These tests include the March 2016 Census Test, which will use an experimental design based on what we learned from the APS.

Looking at the future of combining and integrating data

Statistics New Zealand took measures and explored ways to increase the value of, and ability to work on, integrated data. Here were some recent initiatives.

An infrastructure for administrative data

In October 2015 the New Zealand Government agreed that Statistics New Zealand should actively work towards a census that is based primarily on government administrative data, supported by the redevelopment of its household surveys. This will align the work on census transformation in New Zealand with work on the Integrated Data Infrastructure (IDI) (both were described in the previous issue of *The Survey Statistician*). The alignment will include an investigational statistics environment to explore the possibilities for creating a statistical register of individuals. This is one piece in pursuing the idea to build a system of statistical registers that cover populations of statistical objects and data standards about people, businesses, and land (we already have a business register). As we prepare for the population and dwelling census in 2018, ongoing work will provide infrastructures for data about real property and geography. More information about the progress of this work will follow next year.

A joint sprint with ABS exploring semantic linking on the Linked Employer-Employee Data.

In mid-November Statistics New Zealand and the Australian Bureau of Statistics joined up in Wellington to conduct a sprint project exploring semantic linking. A combined team worked intensively for four days and used semantic web techniques and methodology to answer questions about the impact of the Christchurch earthquakes on the city's labor market. Using our LEED data, the team created, queried, and visualized the results from a graph database using RDF (Resource Description Framework).

The work was done within our Administrative Data First project as a test on how the methods of combining data can fit in a future statistics-production environment. Although the visualization of the results had to be done afterwards, the overall result of the sprint was successful. The team gained a lot of experience on how the techniques and methodology can be used and what is required for further development. We will continue to work with the ABS on this topic. Our next step will be to try connect two or more seemingly disparate data sources.

Contact Grace Chiang at grace.chiang@stats.govt.nz or Christ Conran at chris.conran@abs.gov.au for information about the Statistics NZ–ABS sprint project.

See Census transformation in New Zealand on www.stats.govt.nz for further information or contact Tracey Savage at tracey.savage@stats.govt.nz.

PALESTINE

Reporting: Abdulhakeem Eideh

Dr. Abdulhakeem Eideh- Best paper award in the field of Sampling

The Indian Society of Agricultural Statistics has instituted prizes for the best papers in different fields published in the Journal of the Indian Society of Agricultural Statistics. Accordingly, the papers published in the Journals, Volume 66 (2012) and Volume 67 (201 3) have been evaluated for judging the best papers. The paper entitled Estimation and Prediction under Nonignorable Nonresponse via Response and Nonresponse Distributions by Abdulhakeem AH Eideh (Department of Mathematics, Al-Quds University, Palestine) published in Volume 66, o. 3, December, 2012, pp.359-380 has been selected for best paper award in the field of Sampling.

Energy Consumption Survey in Palestine

The Palestinian Central Bureau of Statistics (PCBS) start conducting the survey on energy consumption in transport sector, under the **project** "Strengthening Statistical Capacity of Arab Countries in Producing Energy Statistics and Energy Consumption in Transport Sector Surveys". The overall objective of this project, funded by the Islamic Development Bank for one year, is to strengthen the capacity of National Statistical Offices (NSOs) in three member countries Egypt, Jordan, and Palestine, of the Economic and Social Commission for Western Asia (ESCWA) in improving the information on energy products consumption in the transport sector in order to assist governments in more effectively managing energy consumption in the countries.

Contributions from IASS Members

Some thoughts on seasonal adjustment variances

by Zdenek Patak, Statistics Canada

Statistical organizations conduct a variety of sub-annual economic and household surveys to produce a number of economic indicators. These indicators are used in a variety of ways: by government organizations to help develop public policy, by businesses to guide and support the implementation of successful business strategies, and by individuals to support/assist them in long term financial planning. Many of these indicators are a combination of long term trend, short term cyclical behavior, seasonal fluctuations and a degree of uncertainty.

To improve usefulness of time-series data collected from sub-annual surveys, survey data are adjusted for seasonal fluctuations. This note introduces the reader to (1) seasonal adjustment of time series, and (2) issues related to reporting reliability measures, such as variances of seasonally adjusted estimates. The producers of sample estimates have been reporting the corresponding variances since the early 1900's. The theory is well established and there is a large body of literature that deals with the subject. Seasonal adjustment has more recent beginnings.

The need for variances of seasonally adjusted data dates back to the 1960's. The "Gordon Commission" (President's Committee to Appraise Employment and Unemployment Statistics 1962), motivated the development of model-based seasonal adjustment methods (e.g., Burman 1980; Hillmer and Tiao 1982; Harvey 1989; and others) as well as nonparametric approaches, the most popular of which is the X11 method, pioneered by the US Bureau of the Census (Shiskin, Young, and Musgrave 1967).

Nonparametric methods are used by many statistical organizations (eg. Statistics Canada, US Bureau of the Census, and others). These organizations use a seasonal adjustment program called X-12-ARIMA that is based on a moving average approach with the X11 method at its core. The model-based approach is used by a number of countries of the European Union (and others) and it is the foundation of a computer software called SEATS (Gomez and Maravall 1997). Since many time series are produced by sample surveys, there is a need to incorporate sampling error in the computation of overall variance. A comparison of the two approaches suggests that the model-based method provides flexibility to adjust for the effects of sampling error while the X11 based method lacks such flexibility.

Despite these developments, questions still remain about how to compute variances for seasonally adjusted data. Bell (2005) presents both the nonparametric and model-based approaches and the ways in which they could be used to compute variances of seasonally adjusted estimates in the presence of sampling error. Tiller and Di Natale (2005) discuss the strengths and weaknesses of model-based

approaches versus its nonparametric counterpart based on moving averages. The remainder of this short note we will draw heavily on the last two references.

(a) Nonparametric Approach

The X11 approach to seasonal adjustment is based on the decomposition of a time series into three components, (i) trend (or trend-cycle), (ii) seasonal, and (iii) irregular,

$$y_t = T_t + S_t + I_t \qquad [1]$$

The first two components account for long- and short-term systematic variation in the series. The third component is a residual term that captures the remaining random variation. When the time series comes from a sample survey it may also have a sampling error component, e_t and equation (1) becomes

$$y_t = T_t + S_t + I_t + e_t \qquad [2]$$

We shall assume that the components of y_t are uncorrelated at all lags and leads.

The first three components are not directly observable and must be estimated to perform seasonal adjustment. The moving average approach uses a set of normalized weights, $\omega_S(B) = \sum_j \omega_{S,j} B^j$ (where B is the backshift operator [$By_t = y_{t-1}$]), over a long period of time to produce an estimate of the seasonal factor, $\hat{S}_t = \omega_S(B)y_t = \sum_j \omega_{S,j} y_{t-j}$. When the seasonal component is removed from the series, the remainder is called *seasonally adjusted* data, composed of a trend and a residual (irregular component). The moving averages are chosen to satisfy a criterion whereby when applied to the series, an unwanted component is approximately reduced to zero in the case of trend, and to a residual term with typically a negligible variance.

In the literature we often see T_t and I_t combined into what is called a *nonseasonal* component, N_t. A linear filter similar to the one used above for the computation of the seasonal factor can be applied to y_t to estimate N_t, i.e., $\widehat{N}_t = \omega_N(B)y_t = \sum_j \omega_{N,j} y_{t-j}$. So, seasonal adjustment can be viewed as either estimating the nonseasonal component, N_t, or estimating and removing S_t as in $A_t = y_t - S_t$. From the definitions of N_t and y_t, we define the estimated error in $\widehat{N}_t$ as

$$\hat{\varepsilon}_t^N = N_t - \omega_N(B)[S_t + N_t + e_t] = [1 - \omega_N(B)]N_t - \omega_N(B)S_t - \omega_N(B)e_t \qquad [3]$$

From the orthogonality of the components, the variance of $\hat{\varepsilon}_t^N$ is

$$Var(\hat{\varepsilon}_t^N) = Var\{[1 - \omega_N(B)]N_t\} + Var[\omega_N(B)S_t] + Var[\omega_N(B)e_t] \qquad [4]$$

Similarly, the error in using $\hat{S}_t$ to estimate S_t is

$$\hat{\varepsilon}_t^S = [1 - \omega_S(B)]S_t - \omega_S(B)N_t - \omega_S(B)e_t \qquad [5]$$

And the corresponding variance of this error is

$$Var(\hat{\varepsilon}_t^S) = Var\{[1 - \omega_S(B)]S_t\} + Var[\omega_S(B)N_t] + Var[\omega_S(B)e_t] \qquad [6]$$

From equations [4] and [6] we see that when there is no sampling error, $Var(\hat{\varepsilon}_t^N) = Var(\hat{\varepsilon}_t^S)$, given that the normalized weights $\omega_S(B)$ and $\omega_N(B)$ are constructed to satisfy $\omega_S(B) + \omega_N(B) = 1$ when there is no sampling error. When sampling error is present, $Var[\omega_N(B)e_t]$ is generally not equal to $Var[\omega_S(B)e_t]$, so $Var(\hat{\varepsilon}_t^N) \neq Var(\hat{\varepsilon}_t^S)$. The difference can be substantial as sampling error does not affect the extraction of the seasonal factors as much as it affects the smoothing of the nonseasonal component. As discussed in Bell (2005), $Var(\hat{\varepsilon}_t^N)$ can be much greater than $Var(\hat{\varepsilon}_t^S)$.

The moving average approach suffers from two major criticisms. One is the lack of standard statistical measures, such as standard errors. Government agencies sometimes publish the sampling error as a proxy for the variability of seasonally adjusted data. That is, however, just one component of the total variance. This may lead to the false impression that seasonal adjustment has no impact on the estimated variance of the adjusted series. Also, the absence of confidence intervals makes the analysis of month to month change more difficult.

A second major criticism is that the moving average approach is not tailored to the properties of a specific series. Many proponents of the method view this as strength as its being nonparametric makes it robust against departures from model assumptions. In a large scale production system, not having to verify individual model assumptions is more efficient. It is also desirable because it requires less involvement from an expert in time series analysis to attest to good model fit.

(b) Model-based approach

An alternative to the moving average approach is model-based seasonal adjustment. The model-based alternative specifies explicit statistical models of trend, seasonal, and irregular components,

$$y_t = s_t + m_t + u_t,$$

Where s_t is a seasonal component, m_t is a smooth trend-cycle, and u_t is an irregular component (assumed to be white noise). Each component is described by a general ARIMA model: $y_{it} = \mu + \frac{\theta(B)}{\phi(B)} a_{it}$ where y_{it} denotes s_t, m_t, u_t, respectively, and the numerator and denominator are moving average and autoregressive operators. The term a_{it} denotes a white-noise variable (also called innovation). Although the irregular component is white noise, it can also be presented in the form of an ARIMA model.

Model-based methods assume that time series and its components can be described by an econometric model. Hence, the model accuracy can be precisely evaluated on the basis of goodness-of-fit diagnostics. In theory, under the assumptions of the model, seasonal adjustment is “optimal” for the specific series. Moreover, the underlying assumptions can be verified. In particular, estimated variances can be computed and confidence intervals can be built around the estimates. In the presence of sampling error, another component, e_t can be added to the model. The sampling error is usually known (in some instances, assumptions regarding the error correlation structure may be necessary) from the underlying survey and may be directly available to the time series analyst.

Model-based methods are tailored to the specifics of a time series and may not universally hold across a class of series. The model assumptions have to be verified with the addition of new data points, and from time to time may break down. In some cases, a reliable model cannot be estimated. Clearly, both approaches have their strengths and weaknesses. Perhaps, the best practice could be to use both in a complementary fashion.

References:

Bell, William R. (2005), "Some Consideration of Seasonal Adjustment Variances", ASA Proceedings, November 2005.

Tiller, Richard, and Di Natale, Marisa (2005), "Model-based seasonally adjusted estimates and sampling error", Monthly Labour Review, September 2005, 27-37.

Burman, J. Peter (1980), "Seasonal Adjustment by Signal Extraction," Journal of the Royal Statistical Society Series A, 143, 321-337.

Gomez, Victor and Maravall, Agustin (1997), "Programs TRAMO and SEATS: Instructions for the User (Beta Version: June 1997)", Working Paper 97001, Ministerio de Economia y Hacienda, Direccion General de Analisis y Programacion Presupuestaria, Madrid, Spain.

Harvey, Andrew C. (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge, U. K., Cambridge University Press.

Hillmer, Steven C., and Tiao, George C. (1982), "An ARIMA-Model-Based Approach to Seasonal Adjustment," Journal of the American Statistical Association, 77, 63-70.

President's Committee to Appraise Employment and Unemployment Statistics (1962), Measuring Employment and Unemployment, U.S. Government Printing Office, Washington, DC.

Shiskin, J., Young, A. H., and Musgrave, J.C. (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper No. 15, U.S. Department of Commerce, Bureau of Economic Analysis.

Upcoming Conferences and Workshops

Statistics Canada 2016 International Methodology Symposium

Organized by: Statistics Canada
Where: the Palais des congrès de Gatineau (in Gatineau, Québec, Canada)
When: March 22 to 24, 2016
Homepage: http://www.statcan.gc.ca/eng/conferences/symposium2016/index

Growth in Statistical Information: Challenges and Benefits

Call for Contributed Papers

Statistics Canada's 2016 International Methodology Symposium will take place at the Palais des congrès de Gatineau (5 minutes from downtown Ottawa) from **March 22 to 24, 2016**.

The title of the Symposium is **"Growth in Statistical Information: Challenges and Benefits".** In recent years, the amount of data available for potential use in producing statistical information has grown by leaps and bounds. Terms such as Big Data, Data Science and Data Mining are becoming more and more common in the literature and the media. But what does it all mean for official statistics and what is the impact on how they are collected, compiled, analyzed and presented?

We are soliciting contributed papers that examine **methodological issues related to the sustained growth in statistical information**.

Topics may include the following:

- Administrative data
- Big Data
- Data on the Web
- Paradata
- Record linkage and statistical matching
- Data mining
- Disclosure Control
- Data warehousing
- Database systems
- Legal and operational access to Big Data
- Data quality
- Measurement errors and Estimation
- Analyses of large data set
- Non-standard analyses of data
- Representativeness
- Dissemination
- Visualization of multidimensional complex data

Please submit your proposal by email to STATCAN.Symposium2016-Symposium2016.STATCAN@statcan.gc.ca by **September 14, 2015**. It must include the following: title, an abstract of approximately 250 words (in English or French), three to six keywords and your full contact information.

Please visit our Web site regularly in order to get more detailed and updated information.
http://www.statcan.gc.ca/eng/conferences/symposium2016/index

The American Association for Public Opinion Research 71st Annual Conference

Where: Austin, Texas
When: May 12-15, 2016
Homepage: http://www.aapor.org/Conference/2016-Conference/Annual-Meeting-Home.aspx

AAPOR Annual Conference - The premier event for public opinion and survey research professionals

Plan to join us for the 71st Annual Conference, **May 12-15, 2016, in Austin, Texas**, themed, "Reshaping the Research Landscape: Public Opinion and Data Science." It promises to be an exceptional learning experience. Register in advance by April 23 and make plans to join your colleagues and learn at AAPOR 2016!

Important Dates

Full Panel Session Submission Deadline: **October 15, 2015**

Paper, Methodological Brief and Poster Submission Deadline: **November 13, 2015**

Registration opens: **February 11, 2016**
Early bird registration ends: **April 8, 2016**
Last day to register in advance: **April 23, 2016**
(*after which must register on site*)

Reserve hotel room: **Available Soon.**
David Dutwin, 2016 Conference Chair
Jennifer Dykema, 2016 Associate Conference Chair

Statistical Society of Canada
Société statistique du Canada

44th Annual Meeting of the Statistical Society of Canada
Where: St. Catharines, Ontario
When: May 29 - June 1, 2016
Website: http://www.ssc.ca/en/meetings/2016

The 44th Annual Meeting of the Statistical Society of Canada will be held at Brock University in St. Catharines, Ontario from Sunday May 29 to Wednesday June 1, 2016.

The Local Arrangements Chair is **Wai Kong (John) Yuen** of Brock University. The Program Chair is **Edward Susko** of Dalhousie University.

Call for Contributed Papers and Posters

Interested individuals are invited to submit abstracts for contributed 15-minute talks or poster presentations in statistics, probability, actuarial science or related areas. Submission must be made through the meeting website. The deadline for submissions is February 12, 2016.
Submissions must include the title of the presentation, the authors' names and affiliations, and an abstract in English or French of 100 words or less. The proposed presenter and the format of the presentation talk or poster should also be indicated. Students who submit an abstract should indicate whether they are eligible, and wish to be considered, for a Student Research Presentation Award. All presenters are required to register for the meeting at the time of abstract submission. The presenters are also responsible for their travel expenses to attend the meeting.

SSC 2016 Invited Sessions

Please see **SSC Liaison Vol 29 No 4, pp 6-10** for a first look at the invited sessions.

Workshops, Sunday, May 29, 2016

Q2016 European Conference on Quality in Official Statistics
Organized by: Eurostat and Instituto Nacional de Estadística - Spain
Where: Madrid, Spain
When: May 31 – June 3, 2016
Homepage: http://www.q2016.es

The National Statistical Institute of Spain (INE) and Eurostat are pleased to invite you to the European Conference on Quality in Official Statistics (Q2016) which will be held in the "Círculo de Bellas Artes" in Madrid, Spain, on 1-3 June 2016.

In addition, a series of short training courses will take place on the day previous to the start of the Conference (31 May 2016).

Since their creation in 2001, these Conferences have become an excellent framework to present and discuss the progress and development of quality in official statistics, as well as to exchange methods and good practices between experts in different areas (statistical offices, international organizations, researchers and academics).

The Conference aims to cover relevant and innovative topics on quality ranging from the challenges and the new paradigm of quality in an information and knowledge-driven society including big data and multi-source statistics, to governance and management aspects like the ones linked to the ESS Vision 2020 or the lessons learned from 2013-2015 peer reviews in the European Statistical System.

Furthermore, as an open forum of debate, it represents an opportunity to introduce innovation in the measurement and management of the quality in statistical domains and in specific statistical products.

The Conference website offers information on the conference venue, accommodation, short courses and leisure-time activities as well as deadlines for submission of papers. Registration is now available.

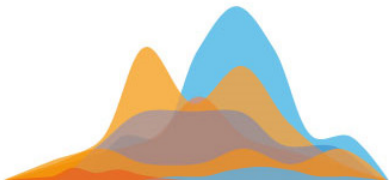
The Conference website offers information on the conference venue, accommodation, short courses and leisure-time activities as well as deadlines for submission of papers. Registration is now available.

If you have any questions, please contact us

Looking forward to seeing you in Madrid!

SIS2016 48th SCIENTIFIC MEETING OF THE ITALIAN STATISTICAL SOCIETY

UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI SCIENZE ECONOMICHE E STATISTICHE



http://meetings.sis-statistica.org/index.php/SIS2016/home

Organized by: Università degli Studi di Salerno - Campus universitario di Fisciano
Where: Fisciano (SA), Italy
When: June 8-10, 2016
Homepage: sis2016@unisa.it

SIS

The Italian Statistical Society was founded in 1939. Today there are about one thousand members including, among others scholars in statistical methodology, probability, social and economic statistics, bio-statistics and demography. The members of the Society are academics and scholars from public or private organizations.
The Italian Statistical Society (SIS) promotes every two years an international scientific meeting which focuses on methodological and applied statistical research.

SIS 2016 – OVERVIEW

The Conference will include plenary, specialized, solicited, contributed and poster sessions. Authors wishing to present a contributed paper are invited to submit an abstract and in case a short paper.

The Committee invites submissions of contributed papers/posters to be included in the program after acceptance. The poster .speed session. Format. These can be on any area of interest relevant to theoretical and applied statistics.

SUBMISSION OF CONTRIBUTED PAPERS AND POSTERS

Papers are submitted on-line. All submissions except invited talks, are subject to a blind refereeing process.

To ensure proper evaluation, short papers should also include a clear statement of the main results and conclusions and indicate the most important key references.

For more details visit the web page: http://meetings.sis-statIstlca.org/index.php/-SIS2016/home/about/submissions#onlineSubmissions

DEADLINE:
Abstract/Paper/Poster:
February 7, 2016 (Authors of the posters can submit only the abstract)

Authors Notification:
March 25, 2016

Final Version:
April 17, 2016

PUBLICATIONS:

a) The Book of Abstracts submitted through the on-line system will be printed for the meeting.
b) Proceedings: all the accepted papers and short papers will be published in the SIS2016 Proceedings (pen drive) with ISBN.

The Fifth International conference on Establishment Surveys (ICES-V)

Where: Geneva, Switzerland
When: June 20-23, 2016
Homepage: http://www.portal-stat.admin.ch/ices5/

The Fifth International Conference on Establishment Surveys (ICES-V) will be held in Geneva, Switzerland, on June 20-23, 2016. Continuing in the traditions of ICES-I to ICES-IV, ICES-V intends to explore new areas of establishment statistics as well as to reflect state-of-the-art at the time of holding the conference.

Situated for the first time in Europe, in the beautiful surroundings of the canton and city of Geneva, ICES-V is expected to be attractive to professionals and researchers in the area of statistics on businesses, farms and institutions throughout the world.

The conference is planned to include:

- Strong offering of short courses on different levels (introductory, intermediate, advanced)
- Introductory overview lectures on important and timely topics
- Selection of invited and contributed papers
- A keynote speaker and reception
- Poster sessions and software demonstrations

This site will be updated with new information as we progress in our steps towards the conference, so please do visit it occasionally. Alternatively, send an email to ices-v@bfs.admin.ch with the subject line "Please add to ICES list" to be kept abreast per email of events related to the ICES conference series.

Looking forward to welcoming you in Geneva, Switzerland

Boris Lorenc and Jean-Pierre Renfer, Conference Co-Chairs

CSDI Comparative Survey Design & Implementation

Second International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts (3MC)

Organized by: CSDI (Comparative Survey Design and Implementation)
Where: Chicago, Illinois
When: July 25 - 29, 2016
Homepage: https://www.csdiworkshop.org/

3MC International Conference 2016

As part of an ongoing effort to promote quality in multi-population surveys and to raise the level of methodological expertise in various applied fields of comparative survey research, the Second International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts (3MC 2016) will be held July 25 - 29, 2016 in Chicago.

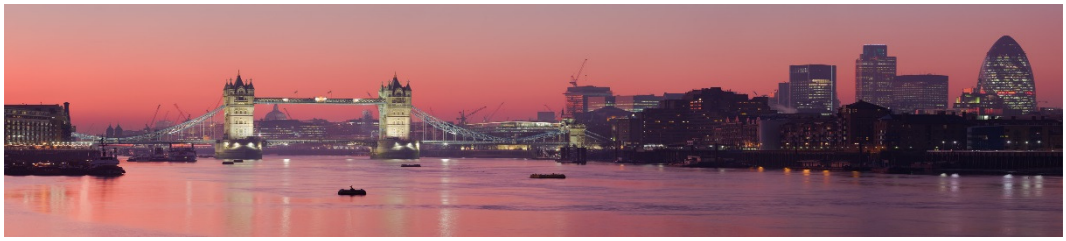
This conference will bring together researchers and survey practitioners concerned with survey methodology and practice in comparative contexts. It will provide a unique opportunity to discuss and present research that contributes to our understanding of survey needs and methods in multi-cultural, multi-national, and multi-lingual contexts. Conference contributions will help document current best practices and stimulate new ideas for further research and development.

We are pleased to announce that we are now accepting abstracts for individual presentations. Submit your abstract here.

For your convenience here is a list of important dates:

- October 7, 2015 - Call for presentations
- January 15, 2016 - Presentation abstracts due
- February 1, 2016 – Early bird registration opens
- March 1, 2016 – Registration opens
- May 21, 2016 – Late registration begins

2015 CSDI Workshop



We are pleased to announce the 2015 CSDI workshop will be held March 26th - March 28th at City University in London.

The main goal of CSDI is improve comparative survey design, implementation and related analysis. The workshop provides a forum and platform for researchers involved in research relevant for comparative survey methods.

This year's topics include:

- Comparability

- Questionnaire Development and Testing
- Translation, Adaptation and Assessment
- Measurement
- Data Collection Challenges
- Innovative Uses of Technology
- Paradata Use
- Comparative Standard Demographics
- Data Dissemination
- Harmonization
- Comparative Analyses

For your convenience here is a list of important dates:

- January 9, 2015 - Online registration opens
- March 1, 2015 - Online registration closes
- March 26 - March 28, 2015 - CSDI Workshop

Janet A. Harkness Student Paper Award

The World Association for Public Opinion Research (WAPOR) and the American Association for Public Opinion Research (AAPOR) jointly support Dr. Harkness's work by sponsoring the Janet A. Harkness Student Paper Award each year. This award recognizes "emerging young scholars in the study of multi-national/multi-cultural/multi-lingual survey research (aka 3M survey research) through support of the winner's participation in the WAPOR Conference and a cash prize." Please follow this link to donate to this effort in gratitude for Dr. Harkness's legacy in the field of cross-cultural, comparative survey research.

Dr. Janet A. Harkness initiated the International Workshop on Comparative Survey Design and Implementation and was one of the driving forces for the Cross-Cultural Survey Guidelines (http://ccsg.isr.umich.edu/). Dr. Harkness passed away in 2012. She inspired and influenced cross-cultural research and work through her steadfast conviction that resources must be made available to researchers and survey practitioners if we are to improve comparative survey research methods, dissemination and analysis.

http://wapor.org/janet-a-harkness-student-paper-award

The Joint Statistical Meetings (JSM) 2016
Where: Chicago, Illinois
When: July 30 - August 4, 2016
Homepage: https://www.amstat.org/meetings/jsm/2016/index.cfm

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians held in North America. It is held jointly with the:

- *American Statistical Association
- *International Biometric Society (ENAR and WNAR)
- *Institute of Mathematical Statistics
- *Statistical Society of Canada
- International Chinese Statistical Association
- International Indian Statistical Association
- Korean International Statistical Society
- International Society for Bayesian Analysis
- Royal Statistical Society
- International Statistical Institute

The 2016 Joint Statistical Meetings will be held July 30 to August 4 at McCormick Place, 2301 S. Lake Shore Drive, Chicago, IL 60616. Chicago offers a wide range of options for sharing time with friends and colleagues or sightseeing with family.

For information, contact meetings@amstat.org.

Small Area Estimation Conference 2016
Where: Maastricht, The Netherlands
When: August 17-19, 2016
Homepage: http://www.sae2016.nl/

Welcome to the website of the Small Area Estimation 2016 conference.

This conference is organized by Maastricht University School of Business and Economics and Statistics Netherlands.

The conference will be held in Maastricht in the faculty building of Maastricht University School of Business and Economics on **August 17-19, 2016**.

The following speakers already confirmed to deliver a presentation.

See also the provisional program for more details.

Keynote speakers:
- Prof. Dr. Thomas Louis, Department of Biostatistics, John Hopkins Bloomberg School of Public Health
- Prof. Dr. Jiming Jiang, Department of Statistics, University of California

Invited and special topic speakers:
- Dr. W. Bell, Census Bureau
- Dr. H.J. Boonstra, Statistics Netherlands
- Dr. H. Chandra, ICAR - Indian Agricultural Statistics Research Institute, New-Delhi
- Prof. Dr. G.S. Datta, Department of Statistics, University of Georgia
- Dr. S. Falorsi, Italian National Statistical Institute, Rome
- Prof. Dr. M. Ghosh, Department of Statistics, University of Florida
- Prof. Dr. S. Holan, Department of Statistics, University of Missouri, Colombia
- Prof. Dr. P. Lahiri, Joint program in Survey Methodology, University of Maryland, College Park, USA
- Prof. Dr. I. Molina, Department of Statistics, Universidad Carlos III de Madrid
- Prof. Dr. J. Opsomer, Department of Statistics, Colorado State University
- Prof. Dr. J.N.K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa
- Prof. Dr. J. Sunil Rao, Department of Public Health Sciences, University of Miami
- Dr. R. Steorts, Department of Statistical Science, Duke University

Royal Statistical Society 2016 International Conference

Where: University Place, Manchester, United Kingdom
When: September 5 - 8, 2016
Homepage:
http://www.rss.org.uk/RSS/Events/RSS_Conference/RSS_2016__International_Conference/RSS_2016_International_Conference.aspx?hkey=44c3b1e0-42b0-42e1-867f-a78b09121de

The RSS 2016 Conference will take place in Manchester – the European City of Science 2016.

Because of the range of topics presented and discussed and the breadth of the audience, this conference is the only one in the UK where all statisticians and users of data gather together. So it's the best forum to share information, network and learn from one another.

Now in its 24th year, the RSS conference has gained prestige for its focus on current statistical issues, how it fosters the exchange of ideas and information and the quality of its speakers. Plenary speakers in 2016 will include Christl Donnelly (Imperial College London) and Xiao-Li Meng (Harvard University).

As always we welcome submissions for talks or posters on any topic related to statistics and the use of data. The submission process will open in January 2016 with an initial deadline for talk submissions of 31 March.

The main conference program will be preceded by one-day training courses on Monday 5 September and in addition professional development workshops will run throughout the event.

For information about the many promotional opportunities available in conjunction with the conference please contact the conference manager.

Registration for the conference will open in March.

PRIVACY IN STATISTICAL DATABASES

Dubrovnik, Croatia, Sep. 14-16, 2016

Privacy in Statistical Databases 2016
Organized by: UNESCO Chair in Data Privacy
Where: Dubrovnik, Croatia
When: September 14-16, 2016
Homepage: http://unescoprivacychair.urv.cat/psd2016/

Privacy in statistical databases is about finding tradeoffs to the tension between the increasing societal and economical demand for accurate information and the legal and ethical obligation to protect the privacy of individuals and enterprises which are the respondents providing the statistical data. In the case of statistical databases, the motivation for respondent privacy is one of survival: statistical agencies or survey institutes cannot expect to collect accurate information from individual or corporate respondents unless these feel the privacy of their responses is guaranteed.
Beyond respondent privacy, there are two additional privacy dimensions to be considered: privacy for the data owners (organizations owning or gathering the data, who would not like to share the data they have collected at great expense) and privacy for the users (those who submit queries to the database and would like their analyses to stay private). Co-utility has shown to provide sustainable solutions to user privacy.

"Privacy in Statistical Databases 2016" (PSD 2016) is a conference sponsored and organized by the UNESCO Chair in Data Privacy with proceedings published by Springer-Verlag in Lecture Notes in Computer Science. Generous support by the Templeton World Charity Foundation is acknowledged. The purpose of PSD 2016 is to attract world-wide, high-level research in statistical database privacy.
PSD 2016 is a successor to PSD 2014 (Eivissa, Sep. 17-19, 2014), PSD 2012 (Palermo, Sep. 26-28, 2012), PSD 2010 (Corfu, Sep. 22-24, 2010), PSD 2008 (Istanbul, Sep. 24-26, 2008), PSD 2006 (Rome, Dec. 13-15, 2006) and PSD 2004 (Barcelona, June 9-11, 2004), all with proceedings published by Springer in LNCS 8744, LNCS 7556, LNCS 6344, LNCS 5262, LNCS 4302 and LNCS 3050, respectively. Those seven PSD conferences follow a tradition of high-quality technical conferences on SDC which started with "Statistical Data Protection-SDP'98", held in Lisbon in 1998 and with proceedings published by OPOCE, and continued with the AMRADS project SDC Workshop, held in Luxemburg in 2001 and with proceedings published in Springer LNCS 2316.

Like the aforementioned preceding conferences, PSD 2016 originates in Europe, but wishes to stay a worldwide event in database privacy and SDC. Thus, contributions and attendees from overseas are welcome.

9th French Colloquium on Survey Sampling
Organized by: Societe Francaise de Statistique
Where: l'Universite du Quebec en Outaouais, Quebec, Canada
When: October 11-14, 2016
Homepage: http://sondages2016.sfds.asso.fr/en/

The Ninth French Colloquium on Survey Sampling (*Colloque francophone sur les sondages*) will take place on **October 12-14, 2016,** on the main campus of the *Université du Québec en Outaouais* (UQO), in Gatineau (Canada). It will be preceded by training workshops on **October 11, 2016,** also on the main campus of UQO. This ninth Colloquium is organized by the *Société Française de Statistique* (SFdS) and its *Enquêtes, Modèles et Applications* group, and by UQO. UQO is considered a university in which the human aspect promotes learning, thought and creation, UQO is strong as a part of the *Université du Québec* network and can rely on the educational resources and shared services of the largest university network in Canada.

The Gatineau Colloquium will capitalize on two synergies, namely the synergy from the meeting of several continents and the synergy from the meeting of specialists from various communities and disciplines: statisticians and statistics users (for example, sociologists, demographers and political scientists) from academia, governments and the private sector.

Looking forward to see you!

ICAS VII The International Conference on Agricultural Statistics 2016
Organized by: Italian National Institute of Statistics (Istat) and Food and Agriculture
Where: Rome, Italy
When: October 26-28, 2016
Homepage: http://icas2016.istat.it/

Modernization of Agricultural Statistics in Support of the Sustainable Development Agenda

The Seventh International Conference on Agricultural Statistics (ICAS VII) will be held in Rome – Italy on 26-28 October 2016. ICAS VII is organized by the Italian National Institute of Statistics, in close collaboration with the Food and Agriculture Organization of the UN (FAO). The Conference focuses on bringing together research and best practices in the field of agriculture statistics, in response to the changing needs and opportunities for agricultural statistics.

ICAS VII convenes senior agricultural statisticians from all over the world. Most of them represent national statistical offices and ministries of agriculture, but the Conference is open to all producers, suppliers, trainers and users of agricultural statistics, such as economists, statisticians, agronomists, researchers, analysts and decision-makers from government entities, academia, development partners and international organizations

The Spirit of Official Statistics: Partnership and continuous innovation
Organized by: International Association of Official Statistics
Where: Abu Dhabi, United Arab Emirates
When: December 6-8, 2016
Homepage: http://www.iaos2016.ae/IAOS-2016-conference.php

About IAOS 2016 Conference

Hosted by the Emirate of Abu Dhabi from 6 to 8 December 2016, the 15th edition of IAOS Conference will provide a unique platform for the international community to share their knowledge and present their insights on innovations in statistics. We are committed to delivering a remarkable international event, which will surpass all expectations and boost the statistical knowledge in the region and the world.

The conference will be held at the multi-award winning venue, Abu Dhabi National Exhibition Centre (ADNEC), which offers a unique experience to conference participants.

The theme of the IAOS2016 conference is: "**The Spirit of Official Statistics: Partnership and continuous innovation**". During the conference, we shall explore what is and what will be the value of official statistics in light of emerging data sources and consider the various forms of potential and successful partnerships that exist. Particular attention will be brought to ways to innovate and modernize national statistical systems, having as a background the 2030 agenda among other things. Also, aspects of the practical applications of the fundamental principles of official statistics will be considered. Finally, statistical experiences and practices in the Gulf region will be highlighted.

The Programme will include keynote speakers, invited paper sessions, workshops, panels, special late breaking sessions and contributed paper sessions.

Submissions for IAOS contributed papers opens on February 1, 2016.

If you are interested in organizing an Invited Paper Session, please send your proposal to Eric Rancourt (eric.rancourt@canada.ca), Chair of the IAOS 2016 Programme Committee. Sessions will be selected/retained based on their quality and relevance to the conference themes. Efforts should be made to have presenters from different regions of the world and from different backgrounds.

For more detailed information on the conference themes, please see the IAOS 2016 web site at (http://www.iaos2016.ae) or contact the programme chair Eric Rancourt (eric.rancourt@canada.ca) directly.

Important Dates

- **Conference Dates**
  6-8 December 2016
- **Submission Deadline Invited Paper Sessions**
  29 January 2016 (extended)
- **Start Submission Contributed Papers**
  1 February 2016
- **Submission Deadline Contributed Papers**
  31 March 2016
- **Submission Deadline Young Statistician Prize 2016**
  31 January 2016

In Other Journals

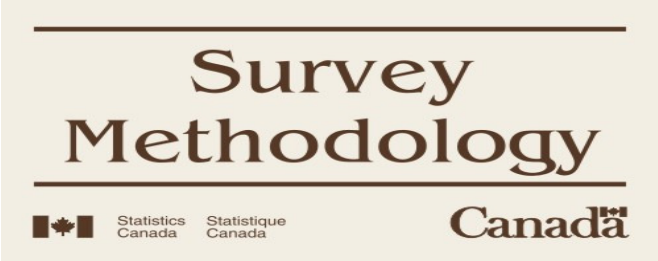
Journal of Survey Statistics and Methodology

VOLUME 3 / NUMBER 4 / DECEMBER 2015

http://jssam.oxfordjournals.org/content/3/4.toc.pdf

Survey Methodology

Geographic Oversampling for Race/Ethnicity Using Data from the 2010 U.S. Population Census
Sixia Chen and Graham Kalton



---

http://www.statcan.gc.ca/pub/12-001-x/12-001-x2015002-eng.htm

DECEMBER 2015, VOL 41, NO 2

Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design (Jan A. van den Brakel and Sabine Krieg)

Domain sample allocation within primary sampling units in designing domain-level equal probability selection methods (Avinash C. Singh and Rachel M. Harter)

A design effect measure for calibration weighting in single-stage samples (Kimberly A. Henry and Richard Valliant)

Model-based small area estimation under informative sampling (François Verret, J.N.K. Rao and Michael A. Hidiroglou)

Combining link-tracing sampling and cluster sampling to estimate the size of a hidden population in presence of heterogeneous link-probabilities (Martín H. Félix-Medina, Pedro E. Monjardin and Aida N. Aceves-Castro)

Model-assisted optimal allocation for planned domains using composite estimation (Wilford B. Molefe and Robert Graham Clark)

Optimum allocation for a dual-frame telephone survey (Kirk M. Wolter, Xian Tao, Robert Montgomery and Philip J. Smith)

Adaptive survey designs to minimize survey mode effects – a case study on the Dutch Labor Force Survey (Melania Calinescu and Barry Schouten)

Integer programming formulations applied to optimal allocation in stratified sampling
(José André de Moura Brito, Pedro Luis do Nascimento Silva, Gustavo Silva Semaan and Nelson Maculan)

Journal of Official Statistics

Volume 31, Issue 4 (Dec 2015)

http://www.degruyter.com/view/j/jos.2015.31.issue-4/issue-files/jos.2015.31.issue-4.xml

Letter to the Editor
Bijak, Jakub / Alberts, Isabel / Alho, Juha / Bryant, John / Buettner, Thomas / Falkingham, Jane / Forster, Jonathan J. / Gerland, Patrick / King, Thomas / Onorante, Luca / Keilman, Nico / O'Hagan, Anthony / Owens, Darragh / Raftery, Adrian / Ševčíková, Hana / Smith, Peter W.F.

Using Auxiliary Sample Frame Information for Optimum Sampling of Rare Populations
Barron, Martin / Davern, Michael / Montgomery, Robert / Tao, Xian / Wolter, Kirk M. / Zeng, Wei / Dorell, Christina / Black, Carla

Response Burden in Official Business Surveys: Measurement and Reduction Practices of National Statistical Institutes
Bavdaž, Mojca / Giesen, Deirdre / Černe, Simona Korenjak / Löfgren, Tora / Raymond-Blaess, Virginie

Statistical Estimators Using Jointly Administrative and Survey Data to Produce French Structural Business Statistics
Brion, Philippe / Gros, Emmanuel

First Impressions of Telephone Survey Interviewers
Broome, Jessica

Quarterly Regional GDP Flash Estimates by Means of Benchmarking and Chain Linking
Cuevas, Ángel / Quilis, Enrique M. / Espasa, Antoni

Coordination of Conditional Poisson Samples
Grafström, Anton / Matei, Alina

Cultural Variations in the Effect of Interview Privacy and the Need for Social Conformity on Reporting Sensitive Information
Mneimneh, Zeina M. / Tourangeau, Roger / Pennell, Beth-Ellen / Heeringa, Steven G. / Elliott, Michael R

Frameworks for Guiding the Development and Improvement of Population Statistics in the United Kingdom
Raymer, James / Rees, Phil / Blake, Ann

B-Graph Sampling to Estimate the Size of a Hidden Population
Spreen, Marinus / Bogaerts, Stefan

Quality Indicators for Statistical Disclosure Methods: A Case Study on the Structure of Earnings Survey
Templ, Matthias

Effects of Cluster Sizes on Variance Components in Two-Stage Sampling
Valliant, Richard / Dever, Jill A. / Kreuter, Frauke

On Proxy Variables and Categorical Data Fusion
Zhang, Li-Chun

Book Review: Online Panel Research: A Data Quality Perspective
Cornesse, Carina / Blom, Annelies G.

Book Review: Practical Tools for Designing and Weighting Survey Samples
Espejo, Mariano Ruiz

Book Review: Managing and Sharing Research Data: A Guide to Good Practice
Mulcahy, Timothy Michael

VOL 8, NO 6 (2015)
http://www.surveypractice.org/index.php/SurveyPractice/issue/view/65

Scientific Surveys Based on Incomplete Sampling Frames and High Rates of Nonresponse
Mansour Fahimi, Frances M Barlas, Randall K Thomas, Nicole Buttermore

The 2013 Census Test: Piloting Methods to Reduce 2020 Census Costs
Gina Walejko, Peter Miller

The Role of Automated SMS Text Messaging in Survey Research
Nina DePena Hoe, Heidi E Grunwald

Computers, Tablets, and Smart Phones: The Truth About Web-based Surveys
Patrick Merle, Sherice Gearhart, Clay Craig, Matthew Vandyke, Mary Elizabeth Brooks, Mehrnaz Rahimi

A Natural Experiment: Inadvertent Priming of Party Identification in a Split-Sample Survey
Marc D. Weiner

SRM

Survey Research Methods

Vol. 9, No. 3 (2015)
https://ojs.ub.uni-konstanz.de/srm/

Editorial
Ulrich Kohler

Identifying Pertinent Variables for Nonresponse Follow-Up Surveys Lessions Learned from 4 Cases in Switzerland
Caroline Vandenplas, Dominique Joye, Michèle Staehli, Alexandre Pollien

Comparing Coefficients of Nonlinear Multivariate Regression Models Bewteen Equations
Christoph Kern, Petra Stein

Does Correction for Measurement Error Have an Effect on the Structure of Basic Human Values?
Laur Lilleoja, Willem E. Saris

Revisiting "yes/no" "versus" check all that apply: Results from a mixed modes experiment
Gerry Nicolaas, Pamela Campanelli, Steven Hope, Annette Jäckle, Peter Lynn

Stable Relationships, Stable Participation? The Effects of Partnership Dissolution and Changes in Relationship Stability on Attrition in a Relationship and Family Panel
Bettina Müller, Laura Castiglioni

Essay: Sunday shopping – The case of three surveys
Jelke Bethlehem

Call for papers

SRM Special Issue: Uses of Geographic Information Systems Tools in Survey Data Collection and Analysis
Stephanie Eckman

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

VOL 31, NO 4 (2015)
http://content.iospress.com/journals/statistical-journal-of-the-iaos/31/4

Interview with Paul Cheung
Kirsten West

Introduction to the Janet Norwood Memorial Papers
Kirsten West

Tribute to Janet L. Norwood
Norwood, Peter

Remembering Janet L. Norwood
Stephen H. Norwood

A Legacy of Objectivity
Thomas J. Plewes

Remarks in Honor of Janet Norwood
Constance F. Citro

Mass appraisal at the census level: Israeli case
Fleishman, Larisa[*] | Gubman, Yury

Different methods to complete datasets used for capture-recapture estimation: Estimating the number of usual residents in the Netherlands
*Gerritse, Susanna C.[a; *] | Bakker, Bart F. M.[b] | van der Heijden, Peter G. M.[c]*

Adapting Labour Force Survey questions from interviewer-administered modes for web self-completion in a mixed-mode design
Betts, Peter[*] | Cubbon, Ben

Satisfaction with official statistics producers
*Steenvoorden, Tina[a] | Řvigelj, Tanja[b] | Bavdaž, Mojca[b; *]*

Statistical governance in the Latin American and the Caribbean Region: Achievements and challenges
*Martín-Guzmán, Pilar[a; *] | Aguilera, M.[b]*

On the interpretation of multi-year estimates of the American Community Survey as period estimates
*Nagaraja, Chaitra H.[a; *] | McElroy, Tucker[b]*

VOL 83, ISSUE 3 (December 2015)
http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1751-5823

A Conversation with John C. Gower
David R. Cox

The Development of Statistical Computing at Rothamsted
John C. Gower

An Interview with Jae C. Lee
Nicholas I. Fisher and Willem R. van Zwet

Approximate Bayesian Computation for a Class of Time Series Models
Ajay Jasra

Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics
Siu-Ming Tam and Frederic Clarke

TRANSACTIONS ON DATA PRIVACY
Foundations and Technologies
http://www.tdp.cat

Volume 8, Issue 3, December 2015
http://www.tdp.cat/issues11/vol08n02.php

A graph theoretic linkage attack on microdata in a metric space
Martin Kroll

Privacy in Data Publishing for Tailored Recommendation Scenarios
João M. Gonçalves, Diogo Gomes, Rui L. Aguiar

A model driven approach to data privacy verification in E-Health systems
Flora Amato, Francesco Moscato

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

June 2015, Vol 178 Issue 4 (October 2015)
http://onlinelibrary.wiley.com/doi/10.1111/rssa.2015.178.issue-4/issuetoc

Big data in social research
Natalie Shlomo[1] and Harvey Goldstein[2]

Statistics: a data science for the 21st century
Peter J. Diggle

Do household surveys give a coherent view of disability benefit targeting? : a multisurvey latent variable analysis for the older population in Great Britain
Ruth Hancock, Marcello Morciano, Stephen Pudney and Francesca Zantomio

Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility
Andrea Carriero, Todd E. Clark and Massimiliano Marcellino

The relationship between education and fertility in the presence of a time varying frailty component
Anna Gottard, Alessandra Mattei and Daniele Vignoli

Analyzing Baseball Data with R
Philip Pallmann

Simulating Nature: a Philosophical Study of Computer-simulation Uncertainties and Their Role in Climate Science and Policy Advice
Umut Okkan and Guül nan

Bayesian and Frequentist Regression Methods
Jonathan Gillard

Journal of the American Statistical Association

ASA AMERICAN STATISTICAL ASSOCIATION

Volume 110, Issue 511 (2015)
http://www.tandfonline.com/toc/uasa20/110/511

Semiparametric Bayesian Density Estimation with Disparate Data Sources: A Meta-Analysis of Global Childhood Undernutrition
Mariel M. Finucane, Christopher J. Paciorek, Gretchen A. Stevens & Majid Ezzati

Comment
Christopher K. Wikle & Scott H. Holan

Comment
Jim Hodges

Rejoinder
Mariel M. Finucane, Christopher J. Paciorek, Gretchen A. Stevens & Majid Ezzati

Stable Weights that Balance Covariates for Estimation with Incomplete Outcome Data
José R. Zubizarreta

An Integrated Bayesian Nonparametric Approach for Stochastic and Variability Orders in ROC Curve Estimation: An Application to Endometriosis Diagnosis
Beom Seuk Hwang & Zhen Chen

The Role of CPS Nonresponse in the Measurement of Poverty
Charles Hokayem, Christopher Bollinger & James P. Ziliak

Clustering High-Dimensional Landmark-Based Two-Dimensional Shape Data
Chao Huang, Martin Styner & Hongtu Zhu

Proper Use of Allele-Specific Expression Improves Statistical Power for cis-eQTL mapping with RNA-Seq Data
Yi-Juan Hu, Wei Sun, Jung-Ying Tzeng & Charles M. Perou

IsoDOT Detects Differential RNA-Isoform Expression/Usage With Respect to a Categorical or Continuous Covariate with High Sensitivity and Specificity
Wei Sun, Yufeng Liu, James J. Crowley, Ting-Huei Chen, Hua Zhou, Haitao Chu, Shunping Huang, Pei-Fen Kuan, Yuan Li, Darla Miller, Ginger Shaw, Yichao Wu, Vasyl Zhabotynsky, Leonard McMillan, Fei Zou, Patrick F. Sullivan & Fernando Pardo-Manuel De Villena

Simultaneous Edit-Imputation for Continuous Microdata
Hang J. Kim, Lawrence H. Cox, Alan F. Karr, Jerome P. Reiter & Quanli Wang

Smoothed Lexis Diagrams with Applications to Lung and Breast Cancer Trends in Taiwan
Li-Chu Chien, Yuh-Jenn Wu, Chao A. Hsiung, Lu-Hai Wang & I-Shou Chang

Theory and Methods:

Robust Estimation of Inverse Probability Weights for Marginal Structural Models
Kosuke Imai & Marc Ratkovic

Bias-Reduced Doubly Robust Estimation
Karel Vermeulen & Stijn Vansteelandt

Testing for Nodal Dependence in Relational Data Matrices
Alexander Volfovsky & Peter D. Hoff

Testing and Modeling Dependencies between a Network and Nodal Attributes
Bailey K. Fosdick & Peter D. Hoff

Blood Flow Velocity Field Estimation Via Spatial Regression With PDE Penalization
Laura Azzimonti, Laura M. Sangalli, Piercesare Secchi, Maurizio Domanin & Fabio Nobile

An Objective Approach to Prior Mass Functions for Discrete Parameter Spaces
C. Villa & S. G. Walker

Whittle Likelihood Estimation of Nonlinear Autoregressive Models with Moving Average Residuals
Tianhao Wang & Yingcun Xia

S-Estimators for Functional Principal Component Analysis
Graciela Boente & Matías Salibian-Barrera

Convergence Properties of a Sequential Regression Multiple Imputation Algorithm
Jian Zhu & Trivellore E. Raghunathan

Likelihood Inferences on Semiparametric Odds Ratio Model
Hua Yun Chen, Daniel E. Rader & Mingyao Li

The E-MS Algorithm: Model Selection with Incomplete Data
Jiming Jiang, Thuan Nguyen & J. Sunil Rao

Regression Analysis of Additive Hazards Model with Latent Variables
Deng Pan, Haijin He, Xinyuan Song & Liuquan Sun

Bandwidth Selection for High-Dimensional Covariance Matrix Estimation
Yumou Qiu & Song Xi Chen

Estimation of Multiple-Regime Threshold Autoregressive Models with Structural Breaks
Chun Yip Yau, Chong Man Tang & Thomas C. M. Lee

Analysis of the Proportional Hazards Model with Sparse Longitudinal Covariates
Hongyuan Cao, Mathew M. Churpek, Donglin Zeng & Jason P. Fine

Detection of Changes in Multivariate Time Series with Application to EEG Data
Claudia Kirch, Birte Muhsal & Hernando Ombao

The Empirical Distribution of a Large Number of Correlated Normal Variables
David Azriel & Armin Schwartzman

Likelihood Estimation for the INAR(p) Model by Saddlepoint Approximation
Xanthi Pedeli, Anthony C. Davison & Konstantinos Fokianos

Tracking Cross-Validated Estimates of Prediction Error as Studies Accumulate
Lo-Bin Chang & Donald Geman

An Equivalent Measure of Partial Correlation Coefficients for High-Dimensional Gaussian Graphical Models
Faming Liang, Qifan Song & Peihua Qiu

Localized Functional Principal Component Analysis
Kehui Chen & Jing Lei

A Regression Framework for Rank Tests Based on the Probabilistic Index Model
Jan De Neve & Olivier Thas

Model Estimation, Prediction, and Signal Extraction for Nonstationary Stock and Flow Time Series Observed at Mixed Frequencies
Tucker McElroy & Brian Monsell

Review :

Design and Analysis of the Randomized Response Technique
Graeme Blair, Kosuke Imai & Yang-Yang Zhou

BIOMETRIKA

Volume 102, issue 2 (2015)
http://biomet.oxfordjournals.org/content/current

Optimal multiple testing under a Gaussian prior on the effect sizes
DOBRIBAN, E., FORTNEY, K., KIM, S. K. and OWEN, A. B.

Strong control of the familywise error rate in observational studies that discover effect modification by exploratory methods
HSU, J. Y., ZUBIZARRETA, J. R., SMALL, D. S. and ROSENBAUM, P. R

Consistent testing for recurrent genomic aberrations
WALTER, V., WRIGHT, F. A. and NOBEL, A. B.

Direct estimation of the mean outcome on treatment when treatment assignment and discontinuation compete
LU, X. and JOHNSON, B. A.

Bayesian inference for partially observed stochastic differential equations driven by fractional Brownian motion
BESKOS, A., DUREAU, J. and KALOGEROPOULOS, K.

Shared kernel Bayesian screening
LOCK, E. F. and DUNSON, D. B.

Singular value shrinkage priors for Bayesian prediction
MATSUDA, T. and KOMAKI, F.

Efficient inference and simulation for elliptical Pareto processes
THIBAUD, E. and OPITZ, T.

Nonparametric methods for group testing data, taking dilution into account
DELAIGLE, A. and HALL, P.

A new specification of generalized linear models for categorical responses
PEYHARDI, J., TROTTIER, C. and GUÉDON, Y.

Diagnostic measures for the Cox regression model with missing covariates
ZHU, H., IBRAHIM, J. G. and CHEN, M.-H.

General weighted optimality of designed experiments
STALLINGS, J. W. and MORGAN, J. P.

Designing dose-finding studies with an active control for exponential families
DETTE, H., KETTELHAKE, K. and BRETZ, F.

MISCELLANEOUS:

Locally optimal designs for errors-in-variables models
KONSTANTINOU, M. and DETTE, H.

Space-filling properties of good lattice point sets
ZHOU, Y. and XU, H.

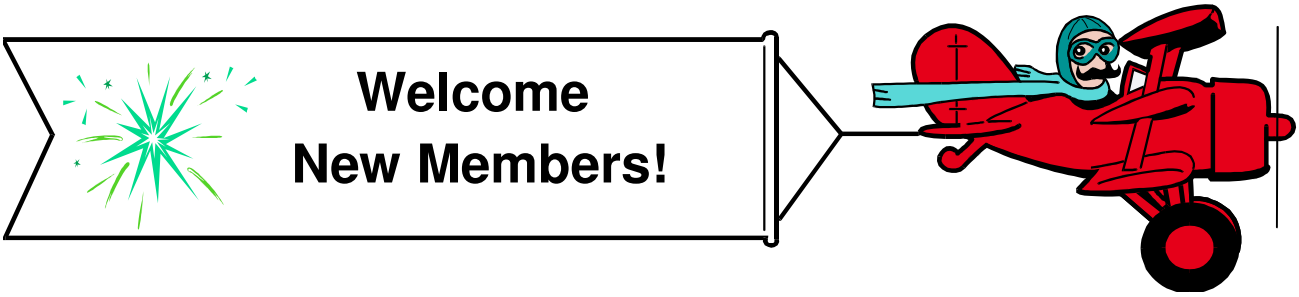
Optimal two-level choice designs for any number of choice sets
SINGH, R., CHAI, F.-S. and DAS, A.

Changepoint estimation: another look at multiple testing problems
CAO, H. and WU, W. B.

On the validity of the pairs bootstrap for lasso estimators
CAMPONOVO, L.

Score tests for association under response-dependent sampling designs for expensive covariates
DERKACH, A., LAWLESS, J. F. and SUN, L.

Clarifying missing at random and related definitions, and implications when coupled with exchangeability
MEALLI, F. and RUBIN, D. B.



We are very pleased to welcome the following new members!

New Members		Country
Dr.	Carlos Roberto Arieira	Brazil
Ms.	Aurélie Vanheuverzwyn	France
Prof.	Fulvia Mecatti	Italy
Prof.	Kyu Seong Kim	Korea, China
Dr.	Mohamed Salimi	Morocco
Ms.	Bouchra Bouziani	Morocco
Dr.	Kees Zeelenberg	Netherlands
Mr.	Nathanael Afolabi	Nigeria
Dr.	Snigdhansu Chatterjee	United States

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 (Estonic)	Imbi.traat@ut.ee
Scientific Secretary:	Denise Silva (Brazil)	denisebritz@gmail.com
Council Members (2013-2017)	Michael Brick (USA)	mikebrick@westat.com
	Daniela Cocchi (Italy)	daniela.cocchi@unibo.it
	Jack Gambino (Canada)	gambino@statcan.ca
	Risto Lehtonen (Finland)	risto.lehtonen@helsinki.fi
	Ralf Münnich (Germany)	muennich@uni-trier.de
	Jean Opsomer (USA)	jopsomer@stat.colostate.edu
(2015-2019)	Hukum Chandra (India)	Hchandra12@gmail.com
	Maria Govanna Ranalli (Italy)	Giovanna.ranalli@stat.unipg.it
	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 Programme Committee 2017	Marcel Vieira (Brazil)	Marcel.vieira@ice.ufjf.br
Chair of the committee for the Cochran Hansen prize award	Risto Lehtonen (Finland)	risto.lehtonen@helsinki.fi

Ex Officio Members

ISI Associate Director, Operation Manager / IASS Liaison	Shabani Mehta	s.mehta@cbs.nl
Transition Executive Director:	Catherine Meunier (France)	katherine.meunier@orange.fr
Director, Executive/Treasurer of the ISI	Ada van Krimpen (The Netherlands)	an.vankrimpen@cbs.nl
Finance Manager:	Michael Leeuwe	
IASS Webmaster:	Mehmood Asghar and Olivier Dupriez	odupriez@worldbank.org
ISI Membership Officer	Margaret de Ruiter-Molloy (The Netherlands)	m.deruitermolloy@cbs.nl

Institutional Members

2 International Organisations

AFRISTAT
EUROSTAT

15 Bureaus of Statistics

AUSTRALIA – AUSTRALIAN BUREAU OF STATISTICS
BRAZIL – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE)
CANADA – STATISTICS CANADA
DENMARK –DANMARKS STATISTIK
FINLAND – STATISTICS FINLAND
GERMANY –STATISTICHE BUNDESAMT
ITALY –INSTITUTO NAZIONALE DI STATISTICA (ISTAT)
KOREA, REPUBLIC OF – STATISTICS KOREA
LUXEMBOURG – EUROPEAN COMMISSION – EUROSTAT
MACAO – 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 – RESEARCH TRIANGLE INSTITUTE
UNITED STATES - SURVEY RESEARCH CENTER
UNITED STATES – NASS, RESEARCH & DEVELOPMENT DIVISION
UNITED STATES – NATIONAL CENTER FOR HEALTH STATISTICS
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 ***(please just indicate one)*****: _____**
May we list your e-mail address on the IASS web site?
Yes ☐ No ☐

Home address
Street: _____
City: _____
State/Province: _____ Zip/Postal code: _____
Country: _____
Telephone number: _____
Fax number: _____

Business address
Company: _____
Street: _____
City: _____
State/Province: _____ Zip/Postal code: _____
Country: _____
Telephone number and extension: _____
Fax number: _____

Please specify address to which your IASS correspondence should be sent:
Home ☐ Business ☐

Read
The Survey Statistician
online!

http://isi-iass.org/home/services/the-survey-statistician/